Uncertainty quantification in unfolding elementary particle spectra at the Large Hadron Collider

THÈSE N^{O} 7118 (2016)

PRÉSENTÉE LE 22 JUILLET 2016
À LA FACULTÉ DES SCIENCES DE BASE
CHAIRE DE STATISTIQUE MATHÉMATIQUE
PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Mikael Johan KUUSELA

acceptée sur proposition du jury:

Prof. F. Nobile, président du jury
Prof. V. Panaretos, directeur de thèse
Prof. Ph. Stark, rapporteur
Prof. S. Wood, rapporteur
Prof. A. Davison, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

“If everybody tells you it’s possible,
then you are not dreaming big enough.”

— Bertrand Piccard and André Borschberg,
while crossing the Pacific Ocean
on a solar-powered aircraft

To my parents

Acknowledgements

I would first and foremost like to sincerely thank my advisor Victor Panaretos for the opportunity of carrying out this work as a member of the Chair of Mathematical Statistics at EPFL and for his tireless support and advice during these past four years. The opportunity of doing a PhD that combines statistical research with an intriguing applied problem from CERN was truly a dream come true for me and these years have indeed been very exciting and intellectually satisfying. I believe that during this work we have been able to make important contributions to both statistical methodology and to data analysis at the LHC and all this would not have been possible without Victor's continuous support.

Throughout this work, I have served as a Statistics Consultant on unfolding issues for the Statistics Committee of the CMS experiment at CERN. This interaction has significantly improved my understanding of the statistical issues that are pertinent to LHC data analysis and I hope that my contributions to the committee's work have improved the way unfolding is carried out at the LHC. I would like to thank the Statistics Committee members, and especially Olaf Behnke, Bob Cousins, Tommaso Dorigo, Louis Lyons and Igor Volobouev, for stimulating discussions, which have often provided me with new insights and research directions. I am also grateful to Volker Blobel, Mikko Voutilainen and Günter Zech for very helpful discussions about unfolding.

During autumn 2014, I was visiting Philip Stark at the Department of Statistics at the University of California, Berkeley. I am extremely grateful to Philip for agreeing to host me and to Victor for providing the opportunity for this visit. I would especially like to thank Philip for taking the time to discuss strict bounds confidence intervals with me. It was these discussions and the support of both Philip and Victor that eventually led to the work described in Chapter 7 of this thesis.

During this work, I have had the great opportunity of discussing research questions with many outstanding statisticians. Besides Victor and Philip, I would especially like to thank Anthony Davison, Peter Green, George Michailidis and Simon Wood for discussions that were very helpful for this work.

I would also like to thank Fabio Nobile, Victor Panaretos, Anthony Davison, Philip Stark and Simon Wood for agreeing to evaluate this thesis as members of my thesis committee and for providing insightful feedback on this work. I am also grateful to my colleagues Anirvan Chakraborty, Pavol Guričan, Valentina Masarotto and Yoav Zemel for providing comments on selected parts of this thesis and to Marie-Hélène Descary for helping with the French translation of the abstract.

I would like to gratefully acknowledge the Swiss National Science Foundation for partially funding this project. I am also grateful to EPFL for providing such a supportive, stimulating and international working environment and to CUSO for the organizing the fantastic statistics summer and winter schools up on the Swiss Alps.

I would also like to thank all my friends in Lausanne, in Helsinki and in the Bay Area for all the memorable moments and fun experiences we have shared over the years. I am also grateful to the current and former SMAT, STAT and STAP group members for this nice time that we have spent together at EPFL. I would especially like to thank Yoav for being such a great officemate during these past 4.5 years and for helping me out with many mathematical questions.

Finally, I would like to thank my parents for all their support during my studies which culminate in this thesis.

Lausanne, July 10, 2016 M. K.

Abstract

This thesis studies statistical inference in the high energy physics unfolding problem, which is an ill-posed inverse problem arising in data analysis at the Large Hadron Collider (LHC) at CERN. Any measurement made at the LHC is smeared by the finite resolution of the particle detectors and the goal in unfolding is to use these smeared measurements to make nonparametric inferences about the underlying particle spectrum. Mathematically the problem consists in inferring the intensity function of an indirectly observed Poisson point process.

Rigorous uncertainty quantification of the unfolded spectrum is of central importance to particle physicists. The problem is typically solved by first forming a regularized point estimator in the unfolded space and then using the variability of this estimator to form frequentist confidence intervals. Such confidence intervals, however, underestimate the uncertainty, since they neglect the bias that is used to regularize the problem. We demonstrate that, as a result, conventional statistical techniques as well as the methods that are presently used at the LHC yield confidence intervals which may suffer from severe undercoverage in realistic unfolding scenarios.

We propose two complementary ways of addressing this issue. The first approach applies to situations where the unfolded spectrum is expected to be a smooth function and consists in using an iterative bias-correction technique for debiasing the unfolded point estimator obtained using a roughness penalty. We demonstrate that basing the uncertainties on the variability of the bias-corrected point estimator provides significantly improved coverage with only a modest increase in the length of the confidence intervals, even when the amount of bias-correction is chosen in a data-driven way. We compare the iterative bias-correction to an alternative debiasing technique based on undersmoothing and find that, in several situations, bias-correction provides shorter confidence intervals than undersmoothing. The new methodology is applied to unfolding the Z boson invariant mass spectrum measured in the CMS experiment at the LHC.

The second approach exploits the fact that a significant portion of LHC particle spectra are known to have a steeply falling shape. A physically justified way of regularizing such spectra is to impose shape constraints in the form of positivity, monotonicity and convexity. Moreover, when the shape constraints are applied to an unfolded confidence set, one can regularize the length of the confidence intervals without sacrificing coverage. More specifically, we form shape-constrained confidence intervals by considering all those spectra that satisfy the shape constraints and fit the smeared data within a given confidence level. This enables us to derive regularized unfolded uncertainties which have by construction guaranteed simultaneous finite-sample coverage, provided that the true spectrum satisfies the shape constraints. The uncertainties are conservative, but still usefully tight. The method is demonstrated using simulations designed to mimic unfolding the inclusive jet transverse momentum spectrum at the LHC.

Keywords: bias-variance trade-off, deconvolution, empirical Bayes, finite-sample coverage, high energy physics, iterative bias-correction, Poisson inverse problem, shape-constrained inference, strict bounds confidence intervals, undersmoothing.

Résumé

Cette thèse traite de l'inférence statistique dans le cas du problème d'*unfolding* en physique des hautes énergies, qui est un problème inverse mal posé apparaissant lors de l'analyse des données au Grand collisionneur de hadrons (Large Hadron Collider (LHC) en anglais) au CERN. Toutes les mesures prises au LHC sont perturbées à cause de la résolution finie des détecteurs de particules et l'objectif du *unfolding* est d'utiliser ces mesures corrompues afin de faire de l'inférence non-paramétrique sur le spectre sous-jacent des particules. D'un point de vue mathématique, le problème consiste à inférer la fonction d'intensité d'un processus de Poisson ponctuel observé indirectement.

La quantification rigoureuse de l'incertitude du spectre *unfoldé* est d'une importance majeure pour les physiciens. Le problème est typiquement résolu en formant tout d'abord un estimateur ponctuel régularisé dans l'espace *unfoldé* et en utilisant par la suite la variabilité de cet estimateur afin de former des intervalles de confiance fréquentistes. Cependant, de tels intervalles sous-estiment l'incertitude puisqu'il néglige le biais introduit lors de la régularisation du problème. Nous démontrons qu'il en résulte que les techniques conventionnelles en statistique ainsi que les méthodes qui sont présentement utilisées au LHC produisent des intervalles de confiance qui peuvent présentés de sérieux problèmes de sous-couverture pour des scénarios réalistes d'*unfolding*.

Nous proposons deux façons complémentaires afin d'aborder ce problème. La première approche s'applique aux situations pour lesquelles on s'attend à ce que le spectre soit une fonction lisse et elle utilise une technique itérative de correction du biais afin de réduire le biais de l'estimateur ponctuel *unfoldé* obtenu en pénalisant la rugosité. Nous démontrons qu'en basant les incertitudes sur la variabilité de l'estimateur ponctuel qui a été corrigé, la couverture est significativement améliorée, et il en découle seulement une augmentation modeste de la longueur des intervalles, et ce, même lorsque la quantité de correction du biais est choisie en fonction des données. Nous comparons les intervalles de confiance obtenus avec la méthode itérative de correction du biais à une autre méthode de réduction du biais basée sur le sous-lissage, et nous trouvons que dans plusieurs situations, la correction du biais produit des intervalles de confiance plus courts que ceux obtenus par l'autre méthode. La nouvelle méthodologie est appliquée afin de faire l'*unfolding* du spectre de masse invariante du boson Z mesuré à l'expérience CMS au LHC.

La seconde approche exploite le fait qu'une portion significative des spectra de particules du LHC sont connus pour décroître de façon très raide. Une manière de régulariser de tels spectra, justifiée par la physique, est d'imposer des contraintes sur leur forme telles que la positivité, la monotonocité et la convexité. De plus, lorsque les contraintes de forme sont appliquées à un ensemble de confiance *unfoldé*, il est possible de régulariser la longueur des intervalles de confiance sans pour autant sacrifier la probabilité de couverture. Plus précisément, nous formons des intervalles de confiance en considérant tous les spectra qui satisfont les contraintes de forme et qui ajustent les données pour un niveau de confiance donné. Ceci nous permet d'obtenir des incertitudes régularisés qui ont par construction une probabilité de couverture simultanée garantie pour des échantillons de taille finie, à condition que le vrai spectre satisfasse les contraintes de forme. Les incertitudes sont conservatrices, mais elles ont tout de même une longueur qui est utile en pratique. La méthode est démontrée à l'aide de simulations qui ont été conçues afin d'imiter l'*unfolding* du spectre de quantité de mouvement transversale des jets inclusifs au LHC.

Mots-clés : Bayes empirique, compromis biais-variance, correction itérative du biais, couverture pour des échantillons de taille finie, déconvolution, inférence sous des contraintes de forme, intervalles de confiance à bornes strictes, physique des hautes énergies, problème inverse de Poisson, sous-lissage.

Contents

Acknowledgements **i**

Abstract **iii**

Résumé **v**

Notation and abbreviations **xi**

1 Introduction **1**

2 Unfolding at the Large Hadron Collider **7**
2.1 Overview of LHC experiments 7
2.2 Statistics at the LHC 9
2.3 Unfolding in LHC data analysis 11

3 Preliminaries **15**
3.1 Forward model for unfolding 15
3.2 Discretization 17
3.2.1 Semi-discrete formulation 18
3.2.2 Spline discretization 18
3.2.3 Histogram discretization 20
3.3 Uncertainty quantification using frequentist confidence intervals 22
3.3.1 Confidence intervals in unfolding 22
3.3.2 Interpretation and use of frequentist confidence intervals 24
3.4 Test scenarios 26
3.4.1 Test setup 1: Two peaks on a uniform background 26
3.4.2 Test setup 2: Inclusive jet transverse momentum spectrum 27

4 Current unfolding methods **31**
4.1 Existing unfolding methodology 31
4.1.1 Overview 31
4.1.2 Tikhonov regularization 33
4.1.3 D'Agostini iteration 35
4.1.4 Other methods 37
4.2 Undercoverage of the unfolded confidence intervals 37

5 Point estimation of smooth spectra 41

- 5.1 Regularization using a smoothness prior . . . 41
  - 5.1.1 Proper smoothness prior using boundary conditions . . . 42
- 5.2 Regularized point estimators . . . 44
  - 5.2.1 Posterior mean estimation . . . 44
  - 5.2.2 Maximum a posteriori and penalized maximum likelihood estimation . 45
  - 5.2.3 Positivity-constrained Gaussian approximation . . . 46
  - 5.2.4 Unconstrained Gaussian approximation . . . 47
- 5.3 Choice of the regularization strength . . . 48
  - 5.3.1 Empirical Bayes . . . 48
  - 5.3.2 Hierarchical Bayes . . . 51
  - 5.3.3 Weighted cross-validation . . . 52
  - 5.3.4 Other methods . . . 53
- 5.4 Simulation study . . . 54
  - 5.4.1 Convergence studies for empirical Bayes . . . 54
  - 5.4.2 Comparison of empirical Bayes and hierarchical Bayes . . . 57
  - 5.4.3 Comparison of empirical Bayes and cross-validation . . . 59

6 Uncertainty quantification for smooth spectra 63

- 6.1 Conventional interval estimates . . . 64
  - 6.1.1 Empirical Bayes credible intervals . . . 64
  - 6.1.2 Hierarchical Bayes credible intervals . . . 65
  - 6.1.3 Gaussian confidence intervals and bootstrap resampling . . . 65
  - 6.1.4 Bootstrap percentile intervals . . . 67
  - 6.1.5 Basic bootstrap intervals . . . 68
  - 6.1.6 Other bootstrap intervals . . . 69
- 6.2 Iteratively bias-corrected percentile intervals . . . 69
  - 6.2.1 Iterative bootstrap bias-correction . . . 70
  - 6.2.2 Percentile intervals from the bias-corrected point estimator . . . 71
- 6.3 Comparison of iterative bias-correction with conventional interval estimates . 73
- 6.4 Iterative bias-correction for linear estimators and Gaussian data . . . 80
  - 6.4.1 Iterative bias-correction for linear estimators . . . 80
  - 6.4.2 Coverage for Gaussian observations . . . 84
  - 6.4.3 Data-driven iteratively bias-corrected intervals . . . 86
  - 6.4.4 Data-driven undersmoothed intervals . . . 88
  - 6.4.5 Simulation study . . . 90
- 6.5 Unfolding with data-driven bias-correction . . . 99
- 6.6 Real data demonstration: Z boson invariant mass spectrum . . . 104
  - 6.6.1 Description of the data . . . 104
  - 6.6.2 Unfolding setup and results . . . 106

7 Shape-constrained uncertainty quantification for steeply falling spectra 109
7.1 Regularization of steeply falling spectra using shape constraints 110
7.2 Outline of the strict bounds construction 111
7.3 Unfolding with shape-constrained strict bounds 114
  7.3.1 Confidence set in the smeared space 114
  7.3.2 Strict bounds via duality 115
  7.3.3 Constraints of the dual program 118
  7.3.4 Conservative discretization of the dual constraints 121
  7.3.5 Summary 128
7.4 Demonstration: Inclusive jet transverse momentum spectrum 128

8 Concluding remarks 133
8.1 Discussion and outlook 133
8.2 Conclusions 136

Appendices 139

A Technical details 139
A.1 Single-component Metropolis-Hastings sampler for unfolding 139
A.2 Implementation details for shape-constrained strict bounds 141

B Full simulation results 143
B.1 Full simulation results for Section 6.3 143
B.2 Full simulation results for Section 6.4.5 153
B.3 Full simulation results for Section 6.5 166

Bibliography 177

Curriculum Vitae 187

Notation and abbreviations

Notation

$\mathbf{1}_A(\cdot)$	indicator function of the set A
$\lVert\cdot\rVert_p^{\boldsymbol{w}}$	weighted ℓ^p-norm, $\lVert\boldsymbol{x}\rVert_p^{\boldsymbol{w}} = \lVert\mathrm{diag}(\boldsymbol{w})\boldsymbol{x}\rVert_p$
$\mathrm{bias}(\hat{\boldsymbol{\theta}})$	bias of the estimator $\hat{\boldsymbol{\theta}}$, $\mathrm{bias}(\hat{\boldsymbol{\theta}}) = \mathrm{E}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$
$C^k(E)$	the space of functions from E to $\mathbb{R}$ with k continuous derivatives
$\mathrm{cov}(\boldsymbol{X})$	covariance matrix of the random vector $\boldsymbol{X}$
$\delta_x(\cdot)$	Dirac measure at x
$\mathrm{diag}(\boldsymbol{a})$	diagonal square matrix whose diagonal elements are given by the vector $\boldsymbol{a}$
$\mathrm{E}_{\boldsymbol{\beta}}(\boldsymbol{X})$	expectation of random vector $\boldsymbol{X}$ under parameter $\boldsymbol{\beta}$
$\boldsymbol{I}_{n\times n}$	identity matrix of size $n \times n$
$K^\dagger$	Moore–Penrose pseudoinverse of the linear operator K
$\mathcal{K}^{-1}(A)$	preimage of the set A under mapping $\mathcal{K}$
$\mathbb{N}_0^d$	d-dimensional natural numbers including 0
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$
$\phi(\cdot)$	standard normal probability density function
$\Phi(\cdot)$	standard normal cumulative distribution function
$\mathrm{P}_{\boldsymbol{\beta}}(A)$	probability of event A under parameter $\boldsymbol{\beta}$
$\mathrm{Poisson}(\boldsymbol{\lambda})$	d-variate probability distribution whose components are independent and Poisson distributed with parameters λ_i, $i = 1, \ldots, d$
$\mathbb{R}_+^d$	positive d-dimensional real numbers, $\mathbb{R}_+^d = \left\{\boldsymbol{x} \in \mathbb{R}^d : x_i \geq 0,\ i = 1, \ldots, d\right\}$
$\mathbb{R}^I$	Cartesian exponentiation of $\mathbb{R}$ with respect to index set I, $\mathbb{R}^I = \prod_{i\in I}\mathbb{R}$
$\mathrm{SE}(\hat{\theta})$	standard error of the estimator $\hat{\theta}$, $\mathrm{SE}(\hat{\theta}) = \left(\mathrm{var}(\hat{\theta})\right)^{1/2}$
$\sigma_{\max}(\boldsymbol{A})$	the largest singular value of the matrix $\boldsymbol{A}$
V^*	algebraic dual space of the vector space V
$\mathrm{var}(X)$	variance of the random variable X
$X \sim P_X$	the random variable X follows the distribution P_X
$X_n \xrightarrow{\text{a.s.}} X$	the sequence of random variables X_n converges almost surely to X
$X \overset{a}{\sim} P_X$	the random variable X follows approximately the distribution P_X
$X^{(1)}, \ldots, X^{(n)} \overset{\text{i.i.d.}}{\sim} P_X$	the random variables $X^{(1)}, \ldots, X^{(n)}$ are independent and identically distributed with distribution P_X
z_α	α-quantile of the $N(0,1)$ distribution

Abbreviations

BC	bias-correction
CMS	Compact Muon Solenoid (experiment)
CV	cross-validation
EB	empirical Bayes
ECAL	electromagnetic calorimeter
EM	expectation-maximization (algorithm/iteration)
HB	hierarchical Bayes
HCAL	hadron calorimeter
HEP	high energy physics
i.i.d.	independent and identically distributed
ISE	integrated squared error
LHC	Large Hadron Collider
MAP	maximum a posteriori (estimator)
MC	Monte Carlo
MCEM	Monte Carlo expectation-maximization (algorithm/iteration)
MCMC	Markov chain Monte Carlo
MISE	mean integrated squared error
MLE	maximum likelihood estimator
MMLE	marginal maximum likelihood estimator
PMLE	penalized maximum likelihood estimator
SE	standard error
SQP	sequential quadratic programming
SVD	singular value decomposition
UQ	uncertainty quantification
US	undersmoothing

1 Introduction

The Large Hadron Collider (LHC) at CERN, the European Organization for Nuclear Research, is the world's largest and most powerful particle accelerator. Located near Geneva, Switzerland, this machine was built to study some of the most profound questions about the physical universe, such as the origin of mass, the nature of dark matter and the asymmetry between ordinary matter and antimatter. This is done by colliding two beams of protons, traveling at speeds very close to the speed of light, and then studying the particles that emerge from these collisions. The collision events are recorded using massive underground particle detectors, resulting in data streams of the order of 100 terabytes per second. These data are then analyzed using a variety of statistical techniques with the aim of obtaining new insight into the physical world. The LHC datasets are not only enormously large, but also have complex internal structure. As a result, the statistical analysis of these data poses major, and sometimes unique, computational and methodological challenges.

This thesis focuses on a particular data analysis task arising at the LHC called the *unfolding problem* (Prosper and Lyons, 2011; Cowan, 1998; Blobel, 2013; Zech, 2016). Let X be some physical quantity of interest studied at the LHC experiments. This could, for example, be the momentum, invariant mass or production angle of particles. Whenever we make a measurement of X, the limited resolution of the particle detectors causes us to observe a noisy version Y of this quantity. The noisy value Y is called a *smeared* or *folded* version of X. The smearing causes the observed, detector-level spectrum of Y to be a blurred version of the true, particle-level spectrum of X. The unfolding problem is then to use observations from the smeared spectrum to make inferences about the actual physical spectrum of X. This is illustrated in Figure 1.1.

The main challenge in unfolding is that it is an ill-posed statistical inverse problem (Kaipio and Somersalo, 2005; Engl et al., 2000). This means that even though the mapping K from the true space into the smeared space is well-behaved, the inverse mapping K^{-1} (or more generally the pseudoinverse $K^{\dagger}$) is unstable with respect to statistical fluctuations in the smeared observations. In other words, within the statistical uncertainties, the smeared data can be explained by the actual physical solution but also by a large family of wildly oscillating

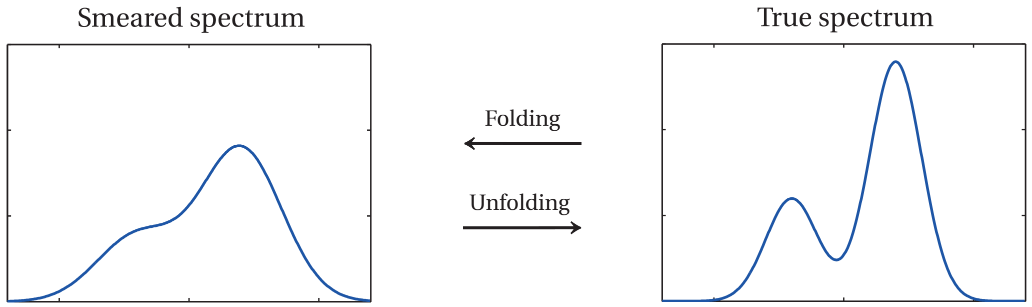


Figure 1.1: The finite resolution of particle detectors causes the observed particle spectra to be smeared versions of the true physical spectra. The unfolding problem is to use the smeared spectrum to make inferences about the true physical spectrum.

unphysical solutions, and the smeared observations alone cannot distinguish among these alternatives.

Physical solutions can nevertheless be obtained by introducing additional a priori information about physically plausible spectra. This is called *regularization* and statistically corresponds to reducing the variance of the unfolded point estimator by introducing a small amount of bias. Two complementary ways of regularizing the unfolding problem are studied in this work. In the first one, we use the knowledge that most physical spectra are smooth functions. We do this by employing a roughness penalty that discourages large oscillations and encourages physically more plausible smooth solutions. In the second one, we rule out unphysical solutions by imposing physically motived shape constraints in the form of positivity, monotonicity and convexity. Such shape constraints are particularly well-suited to unfolding so-called *steeply falling spectra*, which decay rapidly over many orders of magnitude and which are commonplace in LHC data analysis.

Mathematically, both the true observations X and the smeared observations Y can be modeled as realizations of two interrelated Poisson point processes. The particle spectra then correspond to the intensity functions of these processes. Denoting the true intensity by f and the smeared intensity by g, the two are related by $g = Kf$, where the mapping K represents the response of the particle detector. As such, unfolding is an example of a Poisson inverse problem (Antoniadis and Bigot, 2006; Reiss, 1993) and closely related to deconvolution in optics and image reconstruction in positron emission tomography.

A feature that however distinguishes unfolding from many other inverse problems is the need for rigorous uncertainty quantification in the unfolded space. Indeed, a simple point estimate of the unfolded spectrum is of little use to a particle physicist who wishes to use the spectrum to, for example, test theory predictions, compare and combine different experiments and extract further theoretical parameters. To properly carry out these tasks requires the ability to quantify the uncertainty of the unfolded spectrum. Indeed, one could say that in most physics analyses the main scientific output are the uncertainties and the point estimate is only of secondary interest. In this work, we understand uncertainty quantification as the task of constructing frequentist confidence intervals in the unfolded space and providing appropriate statistical methodology for doing this is the main goal of this thesis. Such intervals can then be used as the statistical starting point for the above mentioned more complex inferential tasks.

It turns out that uncertainty quantification in unfolding is a highly nontrivial problem that touches upon several key research topics in contemporary statistics, including post-selection inference (Berk et al., 2013), construction of adaptive nonparametric confidence intervals (Low, 1997) and uncertainty quantification in penalized regression problems, such as lasso regression (Javanmard and Montanari, 2014) and spline smoothing (Ruppert et al., 2003, Chapter 6). The main issue is that most existing uncertainty quantification techniques build the confidence intervals based on the variability of the regularized point estimator. This underestimates the uncertainty and results in confidence intervals that may suffer from serious undercoverage, since the bias that is used to regularize the problem is ignored. But one cannot simply widen the confidence intervals by an amount corresponding to the bias since this amount is unknown. Neither can one simply remove all the bias since this would unregularize the problem and lead to unacceptably large uncertainties.

Given this background, the main contributions of this thesis can be stated as follows:

1. We demonstrate that both existing uncertainty quantification techniques used in LHC data analysis as well as standard frequentist and Bayesian approaches can severely underestimate the uncertainty in the unfolded space, leading to major frequentist undercoverage in realistic unfolding scenarios (Sections 4.2 and 6.3). This happens because existing techniques fail to adequately account for the bias.
2. We propose using an iterative bias-correction technique (Kuk, 1995; Goldstein, 1996) for reducing the bias of roughness-penalized point estimators in order to obtain improved uncertainty quantification in the unfolded space (Section 6.2). We demonstrate that the resulting confidence intervals can attain nearly nominal coverage with only a modest increase in interval length (Sections 6.3, 6.4.5 and 6.5). The coverage is improved as the bias is reduced, but the small amount of residual bias that remains is enough to regularize the interval length.
3. We compare the iterative bias-correction to an alternative debiasing technique based on undersmoothing and find that, in several realistic situations, the bias-corrected confidence intervals are shorter than the undersmoothed intervals (Sections 6.4.5 and 6.5).
4. We propose a novel data-driven procedure for choosing the amount of debiasing in uncertainty quantification (Section 6.4). The method is applicable to both bias-corrected and undersmoothed confidence intervals. Even when the amount of debiasing is chosen in a data-driven way, both methods attain nearly the desired target coverage, but bias-correction produces shorter confidence intervals than undersmoothing. Both debiased intervals are orders of magnitude shorter than unregularized intervals and only moderately longer than non-bias-corrected intervals, which tend to suffer from major undercoverage (Sections 6.4.5 and 6.5).

5. We propose a novel way of regularizing the problem of unfolding steeply falling spectra using shape constraints. Following the *strict bounds* methodology of Stark (1992), we apply the shape constraints directly to an unregularized unfolded confidence set. This enables us to derive regularized unfolded confidence intervals with *guaranteed simultaneous finite-sample frequentist coverage*, provided that the true spectrum satisfies the shape constraints (Chapter 7).

Secondary contributions of this work include formulation of the unfolding problem using Poisson point processes (Section 3.1), statistical description of the existing unfolding techniques (Section 4.1), the proposal of using empirical Bayes estimation to choose the regularization strength (Section 5.3.1), comparison of empirical Bayes and cross-validation for choosing the regularization strength (Section 5.4.3), extension of the strict bounds methodology of Stark (1992) to Poisson noise (Section 7.3), introduction of a new way of imposing and discretizing monotonicity and convexity constraints in strict bounds (Section 7.3) and the proposal of using the single-component Metropolis–Hastings algorithm of Saquib et al. (1998) to sample unfolded posterior distributions (Section A.1).

The rest of this thesis is structured as follows. Chapter 2 provides background on the LHC and the role that unfolding plays in the analysis of LHC data. Chapter 3 sets the stage for the rest of the thesis: it provides the appropriate statistical model for unfolding using Poisson point processes, explains how we discretize the problem, formalizes how the uncertainty is quantified using frequentist confidence intervals and introduces the unfolding test scenarios that are used in the simulation studies through this work. Chapter 4 provides an overview of the statistical methodology that is currently used to solve the unfolding problem at the LHC and demonstrates that the existing methods can suffer from severe undercoverage. Chapters 5 and 6 investigate in detail uncertainty quantification when the true spectrum is expected to be a smooth function. We follow the standard practice of first constructing roughness penalized point estimates and then using these point estimates to construct confidence intervals. Chapter 5 explains how these point estimates are formed and focuses in particular on the choice of the regularization strength. Chapter 6 then explains how to use the point estimates to construct confidence intervals. We review a number of standard frequentist and Bayesian constructions and demonstrate that all these approaches can suffer from severe undercoverage when applied to the unfolding problem. We then introduce the iteratively bias-corrected confidence intervals and demonstrate that they yield significantly improved coverage performance. This chapter also provides the comparison between bias-corrected and undersmoothed intervals and applies the new methodology to unfolding the Z boson invariant mass spectrum measured in the CMS experiment at the LHC. The shape-constrained strict bounds confidence intervals are studied in Chapter 7, where we also apply the resulting methodology to a simulated dataset designed to mimic unfolding the inclusive jet transverse momentum spectrum in the CMS experiment. Chapter 8 closes the thesis with discussion and conclusions. The appendices provide further technical details and simulation results.

Out of the main contributions of this work listed above, items 1, 2 and 5 appear in the journal article Kuusela and Panaretos (2015) and in the preprint Kuusela and Stark (2016), and the present text is partially based on those papers. In particular, Chapters 5 and 6 are an extended version of Kuusela and Panaretos (2015) and Chapter 7 is based on Kuusela and Stark (2016). Also Sections 2.1 and 3.1 in particular borrow heavily from Kuusela and Panaretos (2015), while Chapter 4 extends the results of Kuusela and Stark (2016). Some of the secondary contributions have also partially appeared in the author's MSc thesis Kuusela (2012).

All the simulation studies presented in this thesis were implemented by the author in MATLAB R2014a.

2 Unfolding at the Large Hadron Collider

This chapter puts this work into context by providing background on the LHC experiments and on the role that unfolding plays in the statistical analysis of their data. We first describe in Section 2.1 how LHC experiments detect particles and other physical phenomena. We then provide in Section 2.2 an overview of the statistical methods that are used in LHC data analysis, while the role of unfolding is described in Section 2.3.

2.1 Overview of LHC experiments

The Large Hadron Collider (LHC) is a circular proton-proton collider located in a 27 km long underground tunnel at CERN near Geneva, Switzerland. With a center-of-mass energy of 13 TeV (the electron volt, eV, is the customary unit of energy in particle physics, $1\text{ eV} \approx 1.6 \cdot 10^{-19}$ J), the LHC is the world's most powerful particle accelerator. The protons are accelerated in bunches of billions of particles and bunches moving in opposite directions are made to collide inside four large particle detectors called ALICE, ATLAS, CMS and LHCb. The bunches collide every 25 ns, resulting in some 40 million collisions per second in each detector, out of which the few hundred most interesting ones are stored for further analysis. This results in datasets of the size of roughly 30 petabytes per year, which are stored and analyzed using a global network of supercomputing centers.

Out of the four experiments, ATLAS and CMS are multipurpose detectors capable of performing a large variety of physics analyses ranging from the discovery of the Higgs boson to precision studies of quantum chromodynamics. The other two detectors, ALICE and LHCb, specialize in studies of lead-ion collisions and b-hadrons, respectively. In this section, we provide an overview of the CMS experiment, but similar principles also apply to ATLAS and, to some extent, to other high energy physics experiments.

The CMS experiment (CMS Collaboration, 2008), an acronym for Compact Muon Solenoid, is situated in an underground cavern along the LHC ring near the village of Cessy, France. The detector, weighing a total of 12 500 tons, has a cylindrical shape with a diameter of 14.6 m

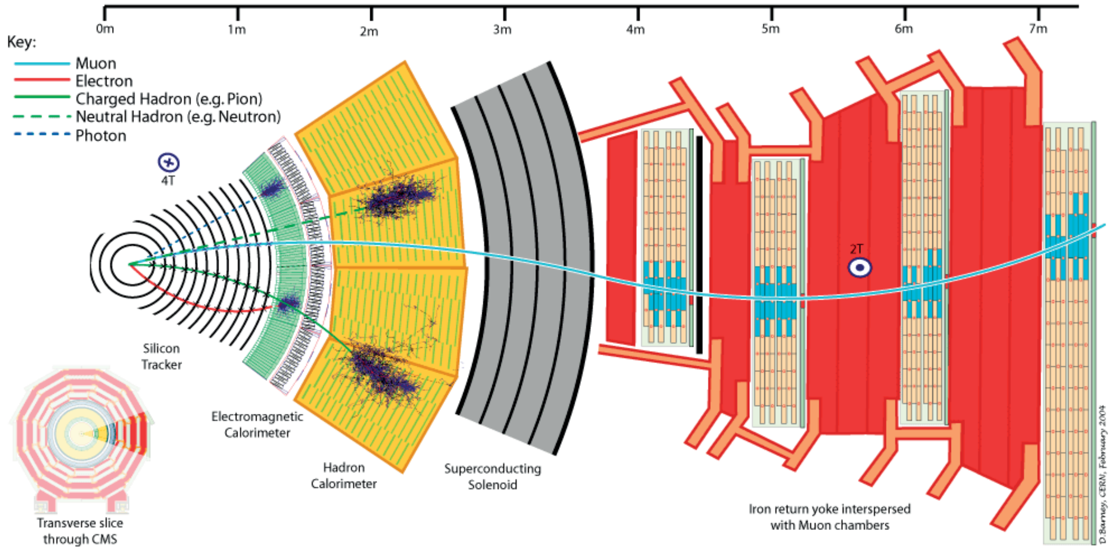

0m
1m
2m
3m
4m
5m
6m
7m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
Transverse slice through CMS
D.Barney, CERN, February 2004

Figure 2.1: Illustration of the detection of particles at the CMS experiment (Barney, 2004). Each type of particle leaves its characteristic trace in the various subdetectors. This enables the identification of different particles and the measurement of their energies and trajectories. Copyright: CERN, for the benefit of the CMS Collaboration.

and a length of 21.6 m. The construction, operation and data analysis of the experiment is conducted by an international collaboration of over 4000 scientists, engineers and technicians.

When two protons collide at the center of CMS, their energy is transformed into matter in the form of new particles. A small fraction of these particles are exotic, short-lived particles, such as the Higgs boson or the top quark. Such exotic particles are the focal point of the scientific interest of the high energy physics community. They decay almost instantly into more familiar, stable particles, such as electrons, muons or photons. Using various subdetectors, the energies and trajectories of these particles can be recorded in order to study the properties and interactions of the exotic particles created in the collisions.

The layout of the CMS detector is illustrated in Figure 2.1. The detector is immersed in a 3.8 T magnetic field created using a superconducting solenoid magnet. This magnetic field bends the trajectory of any charged particle traversing the detector. This enables the measurement of the particle's momentum: the larger the momentum, the less the particle's trajectory is bent.

CMS consists of three layers of subdetectors: the tracker, the calorimeters and the muon detectors. The innermost detector is the silicon tracker, which consists of an inner layer of pixel detectors and an outer layer of microstrip detectors. When a charged particle passes through these semiconducting detectors, it leaves behind electron-hole pairs and hence creates an electric signal. These signals are combined into a particle track using a Kalman filter in order to reconstruct the trajectory of the particle.

The next layer of detectors are the calorimeters, which are devices for measuring energies of particles. The CMS calorimeter system is divided into an electromagnetic calorimeter (ECAL) and a hadron calorimeter (HCAL). Both of these devices are based on the same general principle: they are made of extremely dense materials with the aim of stopping the incoming particles. In the process, a portion of the energy of these particles is converted into light in a scintillating material and the amount of light, which depends on the energy of the particle, is measured using photodetectors inside the calorimeters. The ECAL measures the energy of particles that interact mostly via the electromagnetic interaction, in other words, electrons, positrons and photons. The HCAL, on the other hand, measures the energies of hadrons, that is, particles composed of quarks. These include, for example, protons, neutrons and pions. The HCAL is also instrumental in measuring the energies of jets, which are collimated streams of hadrons produced by quarks and gluons, and in detecting the so-called missing transverse energy, an energy imbalance caused by non-interacting particles, such as neutrinos, escaping the detector.

The outermost layer of CMS consists of muon detectors, whose task is to identify muons and measure their momenta. Accurate detection of muons was of central importance in the design of CMS since muons provide a clean signature for many exciting new physics processes. This is because there is a very low probability for other particles, with the exception of the non-interacting neutrinos, to pass through the CMS calorimeter system. For example, the four-muon decay mode played an important role in the discovery of the Higgs boson (CMS Collaboration, 2012b).

The information of all CMS subdetectors is combined (CMS Collaboration, 2009) to identify the stable particles (muons, electrons, positrons, photons and various types of hadrons) produced in each collision event; see Figure 2.1. For example, a muon will leave a track in both the silicon tracker and the muon chamber, while a photon produces a signal in the ECAL without an associated track in the tracker. The information about these individual particles is then used to reconstruct higher-level physics objects, such as jets or missing transverse energy.

2.2 Statistics at the LHC

This section gives a brief overview of statistical techniques that are used in LHC data analysis. Cowan (1998) and Cranmer (2015) provide accessible introductions to statistical data analysis in high energy physics, while Behnke et al. (2013) provides a more in-depth treatment, including various more technical aspects. Good sources of information are also the proceedings of the PHYSTAT conference series Prosper et al. (2008) and Prosper and Lyons (2011). ATLAS and CMS Collaborations (2011) is a key reference on the statistical methodology that was used in the search for the Higgs boson at the LHC; see also van Dyk (2014).

Physics analyses at the LHC can be broadly speaking divided into two categories which we shall call *discovery analyses* and *measurement analyses*. In the first case, one is performing a search for some new particle or phenomenon and the aim is to use LHC data to establish whether such new physics exists or not. Examples of such analyses include the discovery of the Higgs boson (ATLAS Collaboration, 2012a; CMS Collaboration, 2012b) and searches for dark matter and extra dimensions (CMS Collaboration, 2012a). In measurement analyses, on the other hand, one is interested in studying in detail the properties of some physical phenomenon whose existence has already been previously established. Examples include studies of the properties of the W boson (ATLAS Collaboration, 2012b) and the top quark (CMS Collaboration, 2012d), to name just a few. Somewhat different statistical methods are used in discovery and measurement analyses, but, irrespective of the analysis type, the statistical methods that are used at the LHC fall predominantly under the frequentist paradigm of statistics (Lyons, 2013).

The main statistical tools for discovery analyses are hypothesis tests and one-sided confidence intervals. To illustrate this, let us consider the following simplified model of a single-channel discovery analysis:

$$y \sim \mathrm{Poisson}(b + \mu s), \tag{2.1}$$

where y is the event count in the experiment, $b \geq 0$ is the expected number of background events from known physics processes, $s > 0$ is the expected number of signal events from the new physics process and $\mu \geq 0$ is a signal strength modifier. Hence, $\mu = 0$ corresponds to the absence of new physics and $\mu = 1$ to the predicted strength of the new signal. Then the search for the new phenomenon boils down to testing the null hypothesis $H_0 : \mu = 0$ against the alternative $H_1 : \mu = 1$. The p-value of this test is typically converted to the scale of Gaussian standard deviations (σ) and, by convention, one calls a p-value corresponding to 3σ *evidence* for the new phenomenon, while 5σ constitutes a *discovery* of the new phenomenon. Most discovery analyses also form a one-sided 95 % confidence interval $[0, \overline{\mu}]$ for μ, where $\overline{\mu}$ is called an *upper limit* for μ. If the 95 % upper limit $\overline{\mu} < 1$, then one concludes that the phenomenon can be *excluded* at 95 % confidence level.

In reality, the statistical models that are used in discovery analyses involve several search channels. Moreover, both the background prediction b and the signal strength s depend on some nuisance parameters $\boldsymbol{\xi}$, $b = b(\boldsymbol{\xi})$ and $s = s(\boldsymbol{\xi})$, and the proper handling of these nuisance parameters in the hypothesis test and in the upper limit is one of the key statistical challenges in this type of analyses. Another challenge concerns multiple testing, or the *look-elsewhere-effect* (Gross and Vitells, 2010). For example, in the search for the Higgs boson, the hypothesis test was performed separately for each value m_{H} of the Higgs mass and both non-multiplicity-corrected *local p-values* and multiplicity-corrected *global p-values* were reported. In the case of most LHC discovery analyses, the sensitivity of the search (i.e., the power of the hypothesis test) is optimized by employing machine learning techniques (Bhat, 2011), typically in the form of either neural networks or boosted decision trees.

In measurement analyses, the appropriate statistical tools are point estimators and confidence intervals and sets. Most of these analyses are performed in a parametric mode, where some theory-motivated parametric model is fitted to LHC data. The statistical model is typically of the form

$$\boldsymbol{y} \sim \text{Poisson}\big(\boldsymbol{\mu}(\boldsymbol{\theta})\big), \tag{2.2}$$

where $\boldsymbol{y} \in \mathbb{N}_0^n$ are binned observations (see Section 3.2) and the Poisson mean $\boldsymbol{\mu} \in \mathbb{R}_+^n$ is parameterized by some parameter $\boldsymbol{\theta} \in \mathbb{R}^p$, $p \ll n$, which is of scientific interest. The Poisson distribution follows from the fact that most LHC data can be modeled using Poisson point processes (see Section 3.1). The task is then to both estimate $\boldsymbol{\theta}$ and to quantify its uncertainty. In most cases, $\boldsymbol{\theta}$ can be estimated simply using the maximum likelihood estimator (MLE) and confidence intervals and sets can be obtained from standard MLE asymptotics. A simple extension of this procedure, where the maximum likelihood fit is performed using unbinned data, is also commonly used at the LHC.

Measurement analyses are also where unfolding is employed. In essence, unfolding can be seen as a generalization of problem (2.2) to a situation where the physical parameter-of-interest is no longer finite-dimensional. In that case, the problem becomes ill-posed and maximum likelihood fits and asymptotics are no longer viable. As we will see, the estimation problem needs to be regularized and this makes uncertainty quantification extremely challenging. We next proceed to explain why unfolding is needed in LHC measurement analyses in the first place and the rest of this thesis is then devoted to providing methods for solving the problem, with a particular emphasis on how to quantify the uncertainty of the unfolded solution.

2.3 Unfolding in LHC data analysis

The need for unfolding stems from the fact that any quantity measured at the LHC detectors is corrupted by stochastic noise. Let X be the true physical value of some quantity observed in the detector. Then the actual recorded value of this quantity, say Y, is a slightly perturbed version of X. That is, the conditional distribution of Y given X, $p(Y|X)$, is supported on a continuum of values around X. For instance, let $\mathscr{E}$ be the true energy of an electron hitting the CMS ECAL. Then the observed value of this energy follows to a good approximation the Gaussian distribution $\mathcal{N}\big(\mathscr{E}, \sigma^2(\mathscr{E})\big)$, where the variance satisfies (CMS Collaboration, 2008)

$$\left(\frac{\sigma(\mathscr{E})}{\mathscr{E}}\right)^2 = \left(\frac{S}{\sqrt{\mathscr{E}}}\right)^2 + \left(\frac{N}{\mathscr{E}}\right)^2 + C^2, \tag{2.3}$$

where S, N and C are fixed constants.

The effect of this smearing is that the observed spectrum of particles is “blurred” with respect to the true physical spectrum. This becomes an issue for many measurement analyses, where the scientific goal is to infer this true spectrum and, as a result, the analyses need to use unfolding to “undo” the smearing. This would be a simple task if a physics-driven parametric model were available for the true spectrum, but unfortunately in most measurement analyses parametric assumptions are either considered too strong or no well-justified parametric model exists and regularized nonparametric inferences are needed instead.

Unfolding is used annually by dozens of LHC measurement analyses. Examples include studies of the characteristics of jets (CMS Collaboration, 2012c, 2013b) as well as the properties of the W boson (ATLAS Collaboration, 2012b) and the top quark (CMS Collaboration, 2012d, 2013c), to name just a few. Unfolding has recently also been used to measure differential cross sections associated with the Higgs boson (CMS Collaboration, 2016). In the early phases of this work, we used the CERN Document Server (https://cds.cern.ch/) to survey all papers published by the CMS Collaboration in year 2012 in order to understand how often unfolding is used in LHC data analysis. That year, CMS published a total of 103 papers out of which 16 made direct use of unfolding. Many more also indirectly relied on unfolded results for example through *parton distribution functions* (Forte and Watt, 2013), which are estimated using simultaneous fits to several unfolded spectra (NNPDF Collaboration, 2015). We expect similar statistics to also hold for the other three LHC experiment.

A pertinent question is whether unfolding, and the challenges related to it, could be avoided altogether by simply performing the analysis in the smeared space (Lyons, 2011). The answer to this question depends on what the ultimate goal of the analysis is and how the results are expected to be used. Unfolding is needed at least when the purpose of the analysis is one of the following:

- **Comparison of experiments with different responses:** The only direct way of comparing the spectra measured in two or more experiments with different resolution functions is to compare the unfolded spectra.
- **Input to a subsequent analysis:** Certain tasks, such as the estimation of parton distributions functions or the fine-tuning of Monte Carlo event generators, are typically easiest to carry out using unfolded input spectra.
- **Comparison with future theories:** When the spectra are reported in the unfolded space, a theorist can directly use them to compare with any new theoretical predictions which might not have existed at the time of the original measurement. (This justification is controversial since alternatively one could publish the detector response and the theorist could use it to smear their new predictions. One counterargument is that it is not straightforward to publish and communicate all the systematic uncertainties affecting the detector response.)
- **Exploratory data analysis:** The unfolded spectrum could reveal features and structure in the data which are not considered in any of the existing theoretical models.

These considerations ultimately boil down to the question of what is the best way of communicating the information contained in LHC data so that the format is most useful for further scientific use. Arguably, the ideal solution would be to use the unfolded space, but existing unfolding techniques (see Chapter 4) do not provide a reliable way of doing this. The aim of this work is to contribute towards improving the situation by providing unfolding methodology with better statistical performance and guarantees.

3 Preliminaries

This chapter lays the foundations for the rest of this work. We first introduce and justify in Section 3.1 the appropriate forward model for unfolding based on indirectly observed Poisson point processes. Section 3.2 then explains different ways of discretizing this model. Section 3.3 defines the various forms of frequentist confidence intervals that we aim to construct in this work and explains why such intervals yield useful uncertainty quantification. We close the chapter by introducing in Section 3.4 the unfolding scenarios that are used in the simulation studies throughout this thesis.

3.1 Forward model for unfolding

Most data in experimental high energy physics can be modeled as a Poisson point process (Cranmer, 2015). Let $E \subseteq \mathbb{R}$ be a compact interval and V a space of (sufficiently regular) functions on E. Then the random measure M on state space E is a *Poisson point process* (Reiss, 1993) with a positive intensity function $f \in V$ if and only if:

(i) $M(B) \sim \mathrm{Poisson}(\lambda(B))$ with $\lambda(B) = \int_B f(s)\,\mathrm{d}s$ for every Borel set $B \subseteq E$;

(ii) $M(B_1), \ldots, M(B_n)$ are independent for pairwise disjoint Borel sets $B_i \subseteq E$, $i = 1, \ldots, n$.

In other words, the number of points $M(B)$ observed in the set $B \subseteq E$ follows a Poisson distribution with mean $\int_B f(s)\,\mathrm{d}s$ and point counts in disjoint sets are independent random variables.

The appropriate statistical model for unfolding in an *indirectly observed Poisson point process*, where the Poisson process M represents the true, particle-level spectrum of events. The smeared, detector-level spectrum is represented by another Poisson process N with state space $F \subseteq \mathbb{R}$. We assume that F is a compact interval and denote by W the space of (sufficiently regular) functions on F. Let $g \in W$ be the intensity function of N. The intensities f and g are related by a bounded linear operator $K : V \to W$ so that $g = Kf$. In what follows, we assume K

to be a Fredholm integral operator, that is,

$$g(t) = (Kf)(t) = \int_E k(t,s) f(s)\,\mathrm{d}s, \tag{3.1}$$

with a sufficiently regular integration kernel k. Throughout this work, we assume that k is known, although in practice there is usually some amount of uncertainty associated with it; see the discussion in Chapter 8. The unfolding problem is then to make inferences about the true intensity f given a single observation of the smeared Poisson process N.

This *Poisson inverse problem* (Antoniadis and Bigot, 2006; Reiss, 1993) is ill-posed in the sense that, for most realistic forward operators K, it is possible for true intensities f which are very far apart in V to map into smeared intensities g which are very close to each other in W. From the perspective of point estimation, this means that the naïve approach of first estimating g using, for example, a kernel density estimate $\hat{g}$ and then estimating f using $\hat{f} = K^\dagger \hat{g}$, where $K^\dagger$ is the pseudoinverse of K, is unstable with respect to the Poisson fluctuations of $\hat{g}$. Indeed, in most cases, the pseudoinverse $K^\dagger$ is an unbounded, and hence discontinuous, linear operator (Engl et al., 2000). From the perspective of uncertainty quantification, the ill-posedness means that the set of those solutions $f \in V$ which are consistent with our smeared observations need not be bounded, resulting in possibly arbitrarily large uncertainties.

To better understand the physical justification and meaning of this model, let us consider the unfolding problem at the point level. Denoting by X_i the true, particle-level observations, the Poisson point process M can be written as $M = \sum_{i=1}^{\tau} \delta_{X_i}$ (Reiss, 1993, Section 1.2), where δ_{X_i} is a Dirac measure at $X_i \in E$ and $\tau, X_1, X_2, \ldots$ are independent random variables such that $\tau \sim \mathrm{Poisson}(\lambda(E))$ and the X_i are identically distributed with probability density $f(\cdot)/\lambda(E)$, where $\lambda(E) = \int_E f(s)\,\mathrm{d}s$. In the case of LHC data, the total number of points τ can be assumed to be Poisson distributed by the law of rare events and the observations $X_1, X_2, \ldots$ are independent since individual collision events are independent.

When the particle corresponding to X_i passes through the detector, the first thing that can happen is that it might not be observed at all because of the limited efficiency and acceptance of the device. Mathematically, this corresponds to *thinning* of the Poisson process (Reiss, 1993, Section 2.4). Let $Z_i \in \{0,1\}$ be a variable indicating whether the point X_i is observed ($Z_i = 1$) or not ($Z_i = 0$). We assume that $\tau, (X_1, Z_1), (X_2, Z_2), \ldots$ are independent and that the pairs (X_i, Z_i) are identically distributed. Then the thinned true process is given by $M^* = \sum_{i=1}^{\tau} Z_i \delta_{X_i} = \sum_{i=1}^{\xi} \delta_{X_i^*}$, where $\xi = \sum_{i=1}^{\tau} Z_i$ and the X_i^* are the true points with $Z_i = 1$. The thinned process M^* is a Poisson point process with intensity function $f^*(s) = \varepsilon(s) f(s)$, where $\varepsilon(s) = P(Z_i = 1 | X_i = s)$ is the efficiency of the detector for a true observation at $s \in E$.

For each observed point $X_i^* \in E$, the detector measures a noisy value $Y_i \in F$. We assume the noise to be such that the the smeared observations Y_i are independent and identically distributed. The probability density of Y_i is given by

$$p(Y_i = t) = \int_E p(Y_i = t, X_i^* = s)\,\mathrm{d}s = \int_E p(Y_i = t | X_i^* = s) p(X_i^* = s)\,\mathrm{d}s. \tag{3.2}$$

It follows from this that the smeared observations Y_i constitute a Poisson point process $N = \sum_{i=1}^{\xi} \delta_{Y_i}$, whose intensity function g is given by

$$g(t) = \int_E p(Y_i = t | X_i^* = s) \varepsilon(s) f(s)\,\mathrm{d}s. \tag{3.3}$$

We hence identify that the kernel k in Equation (3.1) is given by

$$k(t,s) = p(Y_i = t | X_i^* = s) \varepsilon(s). \tag{3.4}$$

By the definition of X_i^* and $\varepsilon(s)$, this can also be written as

$$k(t,s) = p(Y_i = t | X_i = s, X_i \text{ observed}) P(X_i \text{ observed} | X_i = s). \tag{3.5}$$

Notice that if we are in the special case where $k(t,s) = k(t-s)$, then unfolding corresponds to a deconvolution problem (Meister, 2009) for Poisson point process observations.

3.2 Discretization

LHC data are typically analyzed in a binned form where the Poisson point processes are discretized using a histogram. This is done both for convenience and for computational reasons. Indeed, in many analyses, there may be billions of observations and treating them individually would not be computationally feasible.

In the case of unfolding, it is natural to carry out the histogram discretization for the smeared space F, and this is indeed done in most unfolding techniques, including the ones studied in this work. Let $\{F_i\}_{i=1}^n$ be a partition of the smeared space F consisting of n histograms bins of the form

$$F_i = \begin{cases} [F_{i,\mathrm{min}}, F_{i,\mathrm{max}}), & i = 1, \ldots, n-1, \\ [F_{i,\mathrm{min}}, F_{i,\mathrm{max}}], & i = n, \end{cases} \tag{3.6}$$

where $F_{i,\mathrm{max}} = F_{i+1,\mathrm{min}}$, $i = 1, \ldots, n-1$. Let us furthermore denote by y_i the number of smeared observations falling on interval F_i, that is, $y_i = N(F_i)$, $i = 1, \ldots, n$. Since N is a Poisson point process, it follows that the histogram of smeared observations $\boldsymbol{y} = \left[y_1, \ldots, y_n\right]^{\mathrm{T}}$ consists of independent and Poisson distributed event counts. The expected value of these counts is given by $\boldsymbol{\mu} = \left[\int_{F_1} g(t)\,\mathrm{d}t, \ldots, \int_{F_n} g(t)\,\mathrm{d}t\right]^{\mathrm{T}}$. Using Equation (3.1), we find that the components of the smeared mean $\boldsymbol{\mu}$ are related to the true intensity f by

$$\mu_i = \int_{F_i} g(t)\,\mathrm{d}t = \int_{F_i} \int_E k(t,s) f(s)\,\mathrm{d}s\,\mathrm{d}t = \int_E k_i(s) f(s)\,\mathrm{d}s, \quad i = 1, \ldots, n, \tag{3.7}$$

where $k_i(s) = \int_{F_i} k(t,s)\,\mathrm{d}t$. We assume here that k and f are sufficiently regular so that the order of integration can be changed. Notice that at this stage we have incurred a loss of information by replacing the individual observations by the bin counts y_i, but we have not made any implicit assumptions regarding the true intensity f.

The situation is less straightforward with regard to the discretization of the true space E. The problem is that relating the smeared mean $\boldsymbol{\mu}$ directly to a histogram discretization of the true process M requires making assumptions about the shape of f inside the true bins. The following sections describe some options that are available for treating the true space E.

3.2.1 Semi-discrete formulation

The most straightforward option is to simply not discretize the problem further. That is, we consider a semi-discrete problem formulation where the smeared space is discretized using a histogram, but the true intensity f is treated as an object of the infinite-dimensional function space V. The semi-discrete forward mapping from f to $\boldsymbol{\mu}$ is then given by

$$\mathcal{K} : V \to \mathbb{R}^n,\ f \mapsto \boldsymbol{\mu} = \left[K_1 f, \ldots, K_n f\right]^{\mathrm{T}}, \tag{3.8}$$

where the functionals K_i are given by Equation (3.7),

$$K_i : V \to \mathbb{R},\ f \mapsto \mu_i = \int_E k_i(s) f(s)\,\mathrm{d}s. \tag{3.9}$$

The corresponding statistical model is

$$\boldsymbol{y} \sim \text{Poisson}(\boldsymbol{\mu}), \text{ with } \boldsymbol{\mu} = \mathcal{K} f, \tag{3.10}$$

and the notation is taken to imply that the components of $\boldsymbol{y}$ are independent. Our task is then to make non-parametric inferences regarding f under this model. We use this semi-discrete formulation in Chapter 7, when we construct shape-constrained confidence intervals for functionals of f.

3.2.2 Spline discretization

A common approach in non-parametric regression problems is to effectively turn the infinite-dimensional problem (3.10) into a finite-dimensional one by considering a basis expansion of f in the form $f(s) = \sum_{j=1}^{p} \beta_j \phi_j(s)$, where $\{\phi_j\}_{j=1}^{p}$ is a finite dictionary of basis functions. Since the true spectra f are typically expected to be smooth functions, a particularly attractive representation for f is provided by using splines (de Boor, 2001; Schumaker, 2007; Wahba, 1990). Let $E_{\min} = s_0 < s_1 < s_2 < \cdots < s_L < s_{L+1} = E_{\max}$ be a sequence of $L+2$ knots in the true space $E = [E_{\min}, E_{\max}]$. Then an order-m spline with knots s_i, $i = 0, \ldots, L+1$, is a piecewise polynomial, whose restriction to each interval $[s_i, s_{i+1})$, $i = 0, \ldots, L$, is an order-m polynomial (that is, a polynomial of degree $m-1$) and which has $m-2$ continuous derivatives at each

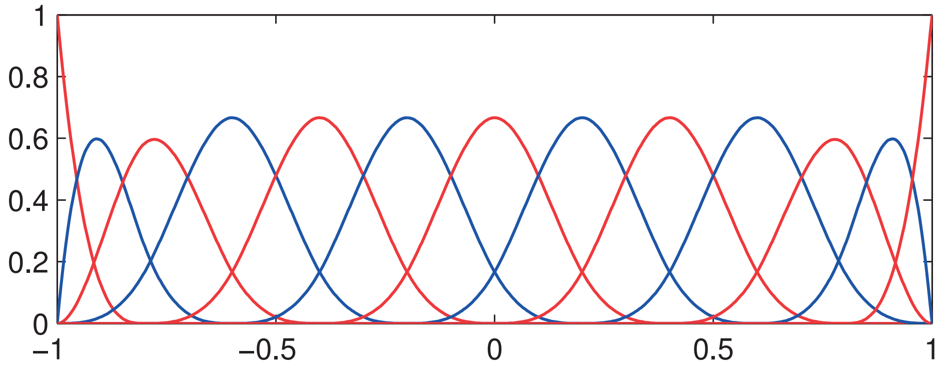


Figure 3.1: B-spline basis with $L = 9$ interior knots on the interval $[-1,1]$.

interior knot s_i, $i = 1,\ldots,L$. An order-m spline with L interior knots can be written as a linear combination of $p = L + m$ basis functions. In this work, we employ the spline discretization in Chapters 5 and 6, where we focus on order-4 cubic splines which consist of third degree polynomials and are twice continuously differentiable.

Various basis systems $\{\phi_j\}_{j=1}^{p}$ can be used to express splines of arbitrary order. We focus on the *B-spline basis* B_j, $j = 1,\ldots,p$, which consists of spline basis functions of minimal local support (de Boor, 2001). This basis has good numerical properties and is also conceptually simple in the sense that each basis coefficient β_j controls the value of f only locally on the support of B_j. Figure 3.1 illustrates the B-spline basis. We use the MATLAB Curve Fitting Toolbox (Mathworks, 2014a) to construct, evaluate and perform basic operations on B-splines. These algorithms are based on recursive use of lower-order B-spline basis functions; see de Boor (2001).

When we plug the B-spline basis expansion $f(s) = \sum_{j=1}^{p} \beta_j B_j(s)$ into Equation (3.7), we find that the smeared means are given by

$$\mu_i = \sum_{j=1}^{p} \left(\int_E k_i(s) B_j(s) \, \mathrm{d}s \right) \beta_j = \sum_{j=1}^{p} K_{i,j} \beta_j, \quad i = 1,\ldots,n, \tag{3.11}$$

where we have denoted

$$K_{i,j} = \int_E k_i(s) B_j(s) \, \mathrm{d}s, \quad i = 1,\ldots,n, \quad j = 1,\ldots,p. \tag{3.12}$$

Hence unfolding reduces to inferring the spline coefficients $\boldsymbol{\beta}$ in the Poisson regression problem

$$\boldsymbol{y} \sim \text{Poisson}(\boldsymbol{\mu}), \text{ with } \boldsymbol{\mu} = \boldsymbol{K}\boldsymbol{\beta}. \tag{3.13}$$

This forward model is correct if the true intensity f is indeed a spline with the knot sequence $\{s_i\}_{i=0}^{L+1}$. For other smooth functions, the model is only approximately correct and the severity of the approximation error depends on how well f can be represented using a spline. If f

is non-smooth, then other types of basis expansions may be more appropriate than splines. For example, if f is known to contain sharp spikes, then a wavelet basis expansion could be considered. However, such situations seem to be rare in HEP unfolding applications.

Since any Poisson intensity function is known to be positive, we would also like to constrain the B-spline $f(s) = \sum_{j=1}^{p} \beta_j B_j(s)$ to be a positive function. This can be done by requiring that

$$\boldsymbol{\beta} \in \mathbb{R}_{+}^{p} = \left\{ \boldsymbol{x} \in \mathbb{R}^{p} : x_i \geq 0,\, i = 1, \ldots, p \right\}. \tag{3.14}$$

Since the B-spline basis functions B_j are positive, this is a sufficient condition for the positivity of f. However, this is not in general a necessary condition for the positivity of f (except for order-1 and order-2 B-splines). Restricting ourselves to the family of B-splines induced by the constraint $\boldsymbol{\beta} \in \mathbb{R}_{+}^{p}$ hence entails a reduction in the family of spline functions available to us in comparison to all positive splines (de Boor and Daniel, 1974), but, given the shape of the B-spline basis functions (see Figure 3.1), this reduction is unlikely to be restrictive in practice.

Also other qualitative shape constraints can be easily imposed on the B-spline basis (Pya and Wood, 2015). For example, a sufficient condition for f to be decreasing is

$$\beta_{j+1} - \beta_j \leq 0, \quad j = 1, \ldots, p-1. \tag{3.15}$$

Similarly,

$$\beta_{j+2} - 2\beta_{j+1} + \beta_j \geq 0, \quad j = 1, \ldots, p-2 \tag{3.16}$$

is a sufficient condition for f to be convex, provided that the knots are uniformly spaced.

O'Sullivan (1986, 1988) was one of the first authors to employ regularized B-splines in statistical applications. The approach was later popularized by Eilers and Marx (1996) in the form of P-splines where a difference penalty is applied to the basis coefficients β_j. The use of B-splines in the HEP unfolding problem was pioneered by Blobel (1985). Also the recent contributions by Dembinski and Roth (2013) and Milke et al. (2013) use splines to discretize the unfolding problem. However, none of these techniques is presently widely used in LHC experiments.

3.2.3 Histogram discretization

The discretization that is used in most LHC unfolding analyses directly relates the smeared mean histogram $\boldsymbol{\mu}$ to a histogram discretization $\boldsymbol{\lambda}$ of the true Poisson point process M. Let $\{E_j\}_{j=1}^{p}$ be a partition of the true space E consisting of p histogram bins of the form

$$E_j = \begin{cases} [E_{j,\min}, E_{j,\max}), & j = 1, \ldots, p-1, \\ [E_{j,\min}, E_{j,\max}], & j = p, \end{cases} \tag{3.17}$$

with $E_{j,\text{max}} = E_{j+1,\text{min}}$, $k = 1,\ldots,p-1$. The histogram of expected number of true events within each true bin E_j is then given by

$$\boldsymbol{\lambda} = \left[\int_{E_1} f(s)\,\mathrm{d}s,\ldots,\int_{E_p} f(s)\,\mathrm{d}s\right]^{\mathrm{T}}. \tag{3.18}$$

The elements of $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are related by (Cowan, 1998, Chapter 11)

$$\mu_i = \int_E k_i(s)f(s)\,\mathrm{d}s = \sum_{j=1}^{p}\int_{E_j} k_i(s)f(s)\,\mathrm{d}s = \sum_{j=1}^{p}\frac{\int_{E_j} k_i(s)f(s)\,\mathrm{d}s}{\int_{E_j} f(s)\,\mathrm{d}s}\lambda_j = \sum_{j=1}^{p} K_{i,j}\lambda_j, \tag{3.19}$$

where the elements of the *smearing* or *response matrix* $\boldsymbol{K}$ are given by

$$K_{i,j} = \frac{\int_{E_j} k_i(s)f(s)\,\mathrm{d}s}{\int_{E_j} f(s)\,\mathrm{d}s}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,p. \tag{3.20}$$

Equivalently, the (i,j)th element of $\boldsymbol{K}$ is the probability that an event in the true bin E_j propagates to the smeared bin F_i (Kuusela, 2012, Proposition 2.11),

$$K_{i,j} = P(Y \in F_i | X \in E_j), \tag{3.21}$$

where X is a true particle-level event and Y is the corresponding smeared detector-level event. The corresponding statistical model then becomes

$$\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{\mu}), \text{ with } \boldsymbol{\mu} = \boldsymbol{K}\boldsymbol{\lambda}, \tag{3.22}$$

and the goal of unfolding is to make inferences about $\boldsymbol{\lambda}$ given a realization of $\boldsymbol{y}$.

Notice that each element $K_{i,j}$ depends on the shape of f within the true bin E_j. Intuitively, the distribution of true events within E_j has an effect on the probability of finding these events within the smeared bin F_i. The dependence on f will cancel out only if f is constant within E_j, in which case $K_{i,j} = \frac{1}{E_{j,\text{max}}-E_{j,\text{min}}}\int_{E_j} k_i(s)\,\mathrm{d}s$ (notice that this corresponds to taking f to be an order-1 spline with the knots placed at the boundaries between the true bins; see Section 3.2.2). The more f deviates from a constant within E_j, the larger its effect on $K_{i,j}$. This dependence is particularly strong for steeply falling spectra, where f may change by several orders of magnitude within E_j.

In real LHC analyses, f is obviously unknown. Most analyses address this by replacing f in Equation (3.20) by an ansatz f^{MC} obtained using a Monte Carlo (MC) event generator, and this is also the approach we follow in Chapter 4 when we study existing unfolding techniques. In other words, unfolding is performed using using the response matrix $\boldsymbol{K}^{\text{MC}}$ with elements

$$K_{i,j}^{\text{MC}} = \frac{\int_{E_j} k_i(s)f^{\text{MC}}(s)\,\mathrm{d}s}{\int_{E_j} f^{\text{MC}}(s)\,\mathrm{d}s}, \quad i = 1,\ldots,n, \quad j = 1,\ldots,p, \tag{3.23}$$

instead of the actual response matrix $\boldsymbol{K}$. The resulting statistical model $\boldsymbol{y} \sim \text{Poisson}(\boldsymbol{K}^{\text{MC}}\boldsymbol{\lambda})$ is correct only if f^{MC} is proportional to f within each true bin E_j. When this is not the case, the model is only approximately correct, and the more the shape of f^{MC} deviates from that of f within the true bins E_j, the larger the modeling error. The situation can be alleviated by using smaller true bins E_j, but the drawback is that this increases the dimension of $\boldsymbol{\lambda}$ and hence increases the ill-posedness of the problem.

It is sometimes argued that the unfolding problem should be regularized by increasing the size of the true bins E_j (Zech, 2011). In the present formulation, this only superficially circumvents the ill-posedness of the problem since with wide bins the Monte Carlo dependence of $\boldsymbol{K}^{\text{MC}}$ becomes increasingly acute. However, in the formulation that we develop in Chapter 7, the true bins can be made arbitrarily large, while avoiding the Monte Carlo dependence altogether.

3.3 Uncertainty quantification using frequentist confidence intervals

A central goal in this work is to quantify our uncertainty regarding the intensity function f using frequentist confidence intervals. The aim of this section is to define what we mean by this (Section 3.3.1) and to explain how to interpret and use the resulting intervals (Section 3.3.2).

3.3.1 Confidence intervals in unfolding

Let $\theta = Hf$ be some real-valued feature of f that we are interested in, with the functional $H : V \to \mathbb{R}$ relating f to the quantity of interest θ. Unless explicitly stated, H need not be linear. Let $\alpha = (0,1)$ and let $\boldsymbol{y} \sim \text{Poisson}(\mathscr{K} f)$, where $\mathscr{K}$ is the semi-discrete forward mapping from Section 3.2.1. Then a *frequentist confidence interval* for θ with confidence level (or coverage probability) $1-\alpha$ is a random interval $\left[\underline{\theta}(\boldsymbol{y}), \overline{\theta}(\boldsymbol{y})\right]$ that satisfies

$$\mathrm{P}_f\left(\theta \in \left[\underline{\theta}(\boldsymbol{y}), \overline{\theta}(\boldsymbol{y})\right]\right) \geq 1-\alpha, \quad \forall f \in V. \tag{3.24}$$

In other words, under repeated sampling of $\boldsymbol{y}$, the random intervals $\left[\underline{\theta}(\boldsymbol{y}), \overline{\theta}(\boldsymbol{y})\right]$ must cover the true value of θ at least $100 \times (1-\alpha)\,\%$ of the time, and *this must hold true for any true intensity* f. Constructing confidence intervals for a given fixed f is trivial—the challenging part is to provide a construction where (3.24) is satisfied for any f in V, or at least in a large subset of V. (Indeed, the author has on several occasions witnessed a situation where a Monte Carlo ansatz f^{MC} is used to construct confidence intervals and the coverage is then checked for data $\boldsymbol{y}$ generated by *this same* spectrum f^{MC}, when the check should be done for one or more spectra that are different from f^{MC}. Such error can be surprisingly difficult to spot in complex data analysis situations.)

In practice, we are often interested in multiple features of f. Let $\{\theta_i\}_{i \in I}$ be a collection of quantities of interest, where $\theta_i = H_i f$ with $H_i : V \to \mathbb{R}$ for each $i \in I$. Then the confidence statement can either be pointwise or simultaneous with respect to the index set I. More specifically, we call a collection of random intervals $\left[\underline{\theta}_i(\boldsymbol{y}),\overline{\theta}_i(\boldsymbol{y})\right]$, $i \in I$, *pointwise confidence intervals* for $\{\theta_i\}_{i\in I}$ with confidence level $1-\alpha$ if

$$\mathrm{P}_f\left(\theta_i \in \left[\underline{\theta}_i(\boldsymbol{y}),\overline{\theta}_i(\boldsymbol{y})\right]\right) \geq 1-\alpha, \quad \forall i \in I, \quad \forall f \in V. \tag{3.25}$$

The intervals $\left[\underline{\theta}_i(\boldsymbol{y}),\overline{\theta}_i(\boldsymbol{y})\right]$, $i \in I$, are called *simultaneous confidence intervals* for $\{\theta_i\}_{i\in I}$ with confidence level $1-\alpha$ if

$$\mathrm{P}_f\left(\theta_i \in \left[\underline{\theta}_i(\boldsymbol{y}),\overline{\theta}_i(\boldsymbol{y})\right], \forall i \in I\right) \geq 1-\alpha, \quad \forall f \in V. \tag{3.26}$$

In the latter case, the $100 \times (1-\alpha)\,\%$ coverage probability concerns the whole collection of intervals $\left[\underline{\theta}_i(\boldsymbol{y}),\overline{\theta}_i(\boldsymbol{y})\right]$, $i \in I$, simultaneously, while in the former case the coverage statement is made for each index $i \in I$ separately.

It is sometimes more natural to view a collection of $1-\alpha$ simultaneous confidence intervals for $\{\theta_i\}_{i\in I}$ as the Cartesian product $\prod_{i\in I}\left[\underline{\theta}_i(\boldsymbol{y}),\overline{\theta}_i(\boldsymbol{y})\right]$ in which case we call this subset of $\mathbb{R}^I$ a $1-\alpha$ *simultaneous confidence set* or, when these is no risk of confusion, simply a $1-\alpha$ *confidence set* for $\{\theta_i\}_{i\in I} \in \mathbb{R}^I$. A confidence set may also be a more general subset of $\mathbb{R}^I$. More specifically, a random set $C(\boldsymbol{y}) \subseteq \mathbb{R}^I$ is a $1-\alpha$ *confidence set* for $\{\theta_i\}_{i\in I}$ if it holds that $\mathrm{P}_f\left(\{\theta_i\}_{i\in I} \in C(\boldsymbol{y})\right) \geq 1-\alpha$, $\forall f \in V$.

If the index set I is finite, then pointwise intervals can always be conservatively adjusted to satisfy a simultaneous confidence statement. Namely, let $\left[\underline{\theta}_i(\boldsymbol{y}),\overline{\theta}_i(\boldsymbol{y})\right]$, $i = 1,\dots,p$, be a collection of $1-\alpha'$ pointwise confidence intervals for $\{\theta_i\}_{i=1}^p$. Then

$$\mathrm{P}_f\left(\underline{\theta}_1(\boldsymbol{y}) \leq \theta_1 \leq \overline{\theta}_1(\boldsymbol{y}),\dots,\underline{\theta}_p(\boldsymbol{y}) \leq \theta_p \leq \overline{\theta}_p(\boldsymbol{y})\right) \geq 1-\sum_{i=1}^p\left(1-\mathrm{P}_f\left(\underline{\theta}_i(\boldsymbol{y}) \leq \theta_i \leq \overline{\theta}_i(\boldsymbol{y})\right)\right) \tag{3.27}$$

$$\geq 1-\sum_{i=1}^p \alpha' = 1-p\alpha', \tag{3.28}$$

where we have used the inequality $\mathrm{P}\left(\bigcap_i A_i\right) \geq 1-\sum_i\left(1-\mathrm{P}(A_i)\right)$. We hence see that if we construct p pointwise confidence intervals at confidence level $1-\frac{\alpha}{p}$, then their simultaneous coverage probability is at least $1-\alpha$. This is called *Bonferroni correction* of the pointwise confidence intervals.

The functionals that we specifically consider in this work are either point evaluators, $H_s f = f(s)$, $s \in E$, or integrals over bins, $H_i f = \int_{E_i} f(s)\,\mathrm{d}s$, $i = 1,\dots,p$, where the bins E_i are of the form (3.17). In the former case, we are interested in $1-\alpha$ *pointwise confidence bands* for f, that is, collections of random intervals $\left[\underline{f}(s;\boldsymbol{y}),\bar{f}(s;\boldsymbol{y})\right]$, $s \in E$, that satisfy

$$\mathrm{P}_f\left(\underline{f}(s;\boldsymbol{y}) \leq f(s) \leq \bar{f}(s;\boldsymbol{y})\right) \geq 1-\alpha, \quad \forall s \in E, \quad \forall f \in V. \tag{3.29}$$

This is the setup we study in Chapter 6. (In practice, the confidence bands are computed on a grid of s values, but this grid can be made arbitrarily dense.) The latter case corresponds to forming confidence intervals $\left[\underline{\lambda}_i(\boldsymbol{y}), \overline{\lambda}_i(\boldsymbol{y})\right]$, $i = 1, \dots, p$, for the expected true histogram $\boldsymbol{\lambda} \in \mathbb{R}^p$ defined in Equation (3.18), and this is the setup we study in Chapters 4 and 7. We consider these intervals both in their pointwise (or rather, binwise) form,

$$\mathrm{P}_f\left(\lambda_i \in \left[\underline{\lambda}_i(\boldsymbol{y}), \overline{\lambda}_i(\boldsymbol{y})\right]\right) \geq 1 - \alpha, \quad \forall i \in 1, \dots, p, \quad \forall f \in V, \tag{3.30}$$

and in their simultaneous form,

$$\mathrm{P}_f\left(\boldsymbol{\lambda} \in \left[\underline{\lambda}_1(\boldsymbol{y}), \overline{\lambda}_1(\boldsymbol{y})\right] \times \cdots \times \left[\underline{\lambda}_p(\boldsymbol{y}), \overline{\lambda}_p(\boldsymbol{y})\right]\right) \geq 1 - \alpha, \quad \forall f \in V. \tag{3.31}$$

In the rest of this work, we mostly write confidence intervals for θ in the form $\left[\underline{\theta}, \overline{\theta}\right]$ omitting the $\boldsymbol{y}$ dependence. It is however important to keep in mind that, in the frequentist definition of a confidence interval (3.24), it is the interval $\left[\underline{\theta}, \overline{\theta}\right]$ that is random and the quantity of interest θ is fixed. This is in contrast with a *Bayesian credible interval* satisfying $\mathrm{P}\left(\theta \in \left[\underline{\theta}, \overline{\theta}\right] \middle| \boldsymbol{y}\right) \geq 1 - \alpha$, where θ is random and the interval $\left[\underline{\theta}, \overline{\theta}\right]$ is fixed.

3.3.2 Interpretation and use of frequentist confidence intervals

We now discuss the interpretation of frequentist confidence intervals and explain why they are useful in carrying out a number of statistical tasks. We treat this question generically without particular reference to the unfolding problem. We focus on the case of univariate confidence intervals, but the discussion can be easily generalized to confidence sets.

Let θ_i be some quantity of interest and let $\left[\underline{\theta}_i, \overline{\theta}_i\right]$ be a $1 - \alpha$ confidence interval for θ_i. For simplicity, we assume that the coverage probability of the interval is exactly $1 - \alpha$, that is, $\mathrm{P}\left(\theta_i \in \left[\underline{\theta}_i, \overline{\theta}_i\right]\right) = 1 - \alpha$. Let us assume that we make n independent repetitions of the measurement for θ_i and let us denote the confidence interval of the jth repetition by $\left[\underline{\theta}_{i,j}, \overline{\theta}_{i,j}\right]$. Then the interpretation that is most often given to frequentist confidence intervals is that, when n is large enough, $100 \times (1 - \alpha)\,\%$ of the intervals $\left[\underline{\theta}_{i,j}, \overline{\theta}_{i,j}\right]$, $j = 1, \dots, n$, will cover the true value of θ_i.

However, in practice, most scientific experiments are only done once. For example, if θ_i is the mass of the Higgs boson, then for the foreseeable future, we only have one particle accelerator where this value can be measured. For this reason, an interpretation that the author finds more attractive is to consider the implications of frequentist coverage for multiple quantities of interest. Let $\{\theta_i\}_{i=1}^m$ be a collection of m quantities of scientific interest. In the case of high energy physics, these could, for example, be the mass of the top quark, the branching ratio of the $Z \to l^+ l^-$ decay mode, the value of the strong coupling constant, the value of the weak mixing angle and so forth. If the measurements of each θ_i are independent and if m is large enough, then frequentist coverage implies that $100 \times (1 - \alpha)\,\%$ of the intervals $\left[\underline{\theta}_{i,1}, \overline{\theta}_{i,1}\right]$, $i = 1, \dots, m$, will cover their corresponding true values θ_i, $i = 1, \dots, m$. In other words, if we make a large number of independent measurements of *different* quantities of interest θ_i, then $100 \times (1-\alpha)$ % of our inferences are correct in the sense that the resulting confidence interval covers the corresponding true value of θ_i. For example, if all the intervals published by LHC experiments were 95 % confidence intervals and if all the measurements were in mutually exclusive channels, then, in the long run, 95 % of these results would be correct in the sense that the published interval covers the actual physical value of the corresponding quantity. (In practice the situation is more complicated as systematic uncertainties introduce correlations between measurements.) This interpretation of frequentist confidence intervals is advocated, for example, by Wasserman (2004, Section 6.3.2).

More rigorously, let $Z_{i,j} = \mathbf{1}\{\theta_i \in [\underline{\theta}_{i,j}, \overline{\theta}_{i,j}]\}$ be the indicator variable showing whether the interval $[\underline{\theta}_{i,j}, \overline{\theta}_{i,j}]$ covers θ_i. Then, for every i, j, we have that $Z_{i,j} \sim \text{Bernoulli}(1-\alpha)$ and, assuming that both the different measurements and the repetitions of the same measurement are independent, the $Z_{i,j}$ are i.i.d. random variables with mean $\mu = \mathrm{E}(Z_{i,j}) = 1-\alpha$. Then, by the strong law of large numbers,

$$\frac{\#\{Z_{i,j} = 1\}_{j=1}^{n}}{n} = \frac{1}{n}\sum_{j=1}^{n} Z_{i,j} \xrightarrow{\text{a.s.}} \mu = 1-\alpha, \quad \text{when } n \to \infty. \tag{3.32}$$

That is, the fraction of intervals that cover θ_i when the ith measurement is repeated n times converges almost surely to $1-\alpha$ as $n \to \infty$. This corresponds to the conventional interpretation of frequentist confidence intervals. But we also have that

$$\frac{\#\{Z_{i,1} = 1\}_{i=1}^{m}}{m} = \frac{1}{m}\sum_{i=1}^{m} Z_{i,1} \xrightarrow{\text{a.s.}} \mu = 1-\alpha, \quad \text{when } m \to \infty. \tag{3.33}$$

That is, as the number of measurements m grows, the fraction of intervals $[\underline{\theta}_{i,1}, \overline{\theta}_{i,1}]$, $i = 1, \ldots, m$, that cover their corresponding quantity of interest θ_i, $i = 1, \ldots, m$, converges almost surely to $1-\alpha$. This leads to the alternative interpretation of the intervals via coverage in independent measurements of unrelated quantities of interest. This is illustrated in Figure 3.2.

A frequentist $1-\alpha$ confidence interval for θ_i is useful for carrying out a number of statistical tasks. By the duality between confidence intervals and hypothesis tests, a test that rejects the null hypothesis $H_0 : \theta_i = \theta_{i,0}$ when $\theta_{i,0} \notin [\underline{\theta}_i, \overline{\theta}_i]$ has significance level α (e.g., Casella and Berger, 2002, Theorem 9.2.2). In other words, we can perform a hypothesis test by simply checking whether the null value $\theta_{i,0}$ is contained within the interval $[\underline{\theta}_i, \overline{\theta}_i]$. Two or more independent confidence intervals for the same quantity of interest θ_i can also be combined by making an appropriate multiplicity correction and then considering the intersection of the multiplicity-corrected intervals. For example, a simple calculation shows that if $[\underline{\theta}_{i,1}, \overline{\theta}_{i,1}]$ and $[\underline{\theta}_{i,2}, \overline{\theta}_{i,2}]$ are two independent confidence intervals for θ_i with confidence level $\sqrt{1-\alpha}$, then their intersection is a $1-\alpha$ confidence interval for θ_i. Furthermore, confidence intervals and more generally confidence sets with appropriate multiplicity corrections can be propagated to further analyses using a construction similar to the one outlined later in Section 7.2. To enable

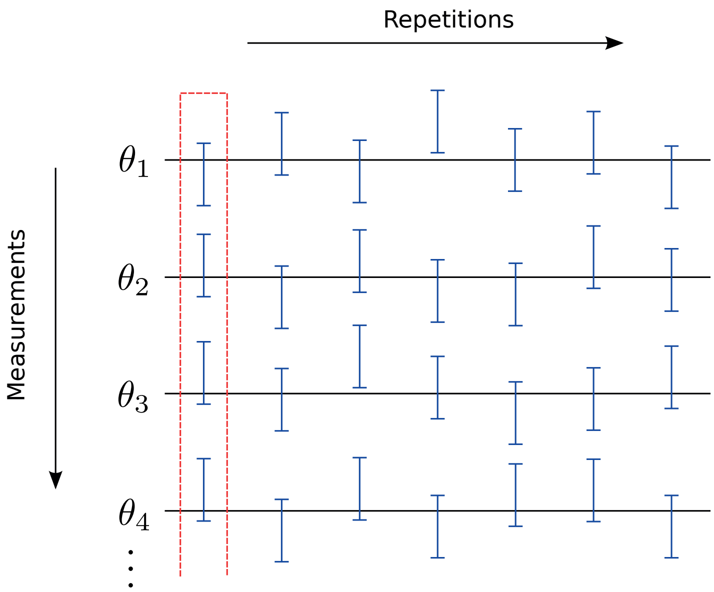


Figure 3.2: Frequentist confidence intervals imply coverage not only for independent repetitions of the same measurement, but also for independent measurements of unrelated quantities of interest $\{\theta_i\}$.

this kind of use, it can be helpful to publish confidence intervals at all possible confidence levels in the form of a *confidence distribution* (Xie and Singh, 2013).

3.4 Test scenarios

We close this chapter by introducing the two unfolding scenarios that will be primarily used in the simulation studies throughout this thesis. The first corresponds to a case where it is reasonable to assume the spectrum to be a smooth function and is primarily used in Chapters 5 and 6. The second corresponds to a situation where the spectrum is most naturally regularized using shape constraints and is primarily studied in Chapter 7. Both cases are also used in Chapter 4 to demonstrate the shortcomings of existing unfolding methods.

3.4.1 Test setup 1: Two peaks on a uniform background

The first test case we consider consists of two Gaussian peaks on top of a uniform background. Such an intensity resembles situations where invariant mass peaks are observed on top of some background events. More specifically, we let the intensity of the true process M be

$$f(s) = \lambda_{\text{tot}} \left\{ \pi_1 \mathcal{N}(s|-2,1) + \pi_2 \mathcal{N}(s|2,1) + \pi_3 \frac{1}{|E|} \right\}, \quad s \in E. \tag{3.34}$$

where $\lambda_{\text{tot}} = \mathrm{E}(\tau) = \int_E f(s)\,\mathrm{d}s > 0$ is the expected number of true observations and $|E|$ denotes the length of the interval E. The mixing proportions π_i sum to one and are set to $\pi_1 = 0.2$,

$\pi_2 = 0.5$ and $\pi_3 = 0.3$. We consider the sample sizes $\lambda_{\mathrm{tot}} = 1\,000$, $\lambda_{\mathrm{tot}} = 10\,000$ and $\lambda_{\mathrm{tot}} = 50\,000$, which we call the small, medium and large sample size cases, respectively. We take both the true space E and the smeared space F to be the interval $[-7, 7]$. We assume the smeared intensity g to be given by the Gaussian convolution

$$g(t) = (Kf)(t) = \int_E \mathcal{N}(t - s|0, 1) f(s)\,\mathrm{d}s, \quad t \in F. \tag{3.35}$$

This corresponds to a situation where the true points X_i are smeared by additive standard Gaussian noise, that is, $Y_i = X_i + \varepsilon_i$, where X_i and ε_i are independent and $\varepsilon_i \sim N(0, 1)$. Points that are smeared beyond the boundaries of F are discarded from the analysis. The resulting true intensity f and smeared intensity g in the case of the medium sample size are shown in Figure 3.3(a). Notice that in classical deconvolution theory (Meister, 2009) this setup belongs to the most difficult class of deconvolution problems since the Gaussian noise has a supersmooth probability density function.

We discretize the smeared space F using a histogram of $n = 40$ equal-width bins. For the experiments of Chapter 4, we discretize the true space E as described in Section 3.2.3 using a histogram of $p = 30$ equal-width bins. In Chapters 5 and 6, we use the spline discretization described in Section 3.2.2. More specifically, we discretize the true space E using cubic B-splines with $L = 26$ uniformly placed interior knots corresponding to $p = L + 4 = 30$ basis coefficients.

3.4.2 Test setup 2: Inclusive jet transverse momentum spectrum

We use the inclusive jet transverse momentum spectrum (CMS Collaboration, 2011, 2013b) as an example of a steeply falling particle spectrum. *A jet* is a collimated stream of energetic particles. Observing a jet in an LHC particle detector signifies that a quark or a gluon was created in the proton-proton collision. By the theory of quantum chromodynamics, quarks and gluons cannot exist as isolated objects. They will hence create other particles from the vacuum around them in a process called *hadronization*, and the end result of this process is a stream of particles in the form of a jet. The inclusive jet transverse momentum spectrum is the average number of jets as a function of their transverse momentum p_{T}, that is, their momentum in the direction perpendicular to the proton beam. The transverse momentum p_{T} is measured in units of electron volts (eV). Measurement of this spectrum constitutes an important test of the Standard Model of particle physics and provides constraints on the free parameters of the theory.

Our aim is to provide a simulation setup that mimics unfolding the inclusive jet p_{T} spectrum in the CMS experiment (CMS Collaboration, 2008) at the LHC. We generate the data using the particle-level intensity function

$$f(p_{\mathrm{T}}) = LN_0 \left(\frac{p_{\mathrm{T}}}{\mathrm{GeV}}\right)^{-\alpha} \left(1 - \frac{2}{\sqrt{s}} p_{\mathrm{T}}\right)^{\beta} e^{-\gamma/p_{\mathrm{T}}}, \quad 0 < p_{\mathrm{T}} \le \frac{\sqrt{s}}{2}, \tag{3.36}$$

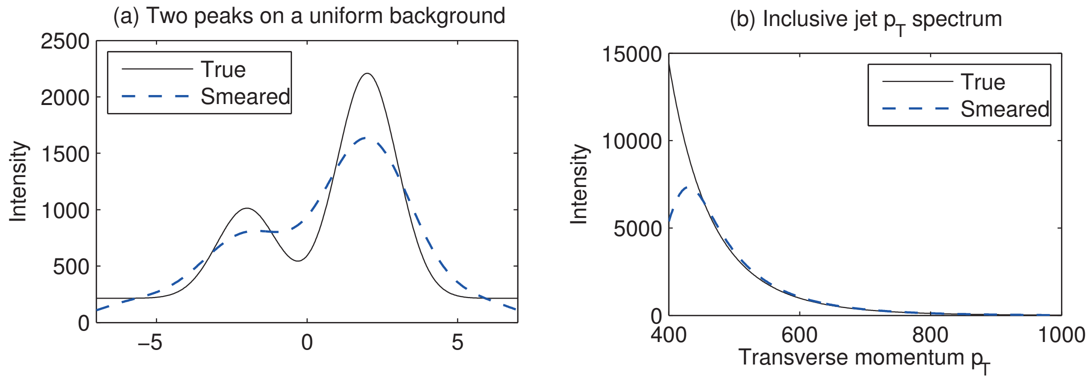


Figure 3.3: Illustration of the unfolding scenarios that we consider in our simulation studies. Figure (a) shows the intensity functions for the two peaks on a uniform background test setup and Figure (b) for the inclusive jet transverse momentum spectrum. The first case serves as an example of a spectrum that is naturally regularized using a roughness penalty, while the second case is an example of a steeply falling spectrum, which can be regularized using shape constraints.

where $L > 0$ is the integrated luminosity (a measure of the amount of collisions produced in the accelerator, measured in units of inverse barns, b^{-1}), $\sqrt{s}$ is the center-of-mass energy of the proton-proton collisions and N_0, α, β and γ are positive parameters. This parameterization was used in early inclusive jet analyses at the LHC (CMS Collaboration, 2011) and is motivated by physical considerations. For example, the $\left(1 - \frac{2}{\sqrt{s}} p_\mathrm{T}\right)^\beta$ term corresponds to a kinematic cut-off at $p_\mathrm{T} = \frac{\sqrt{s}}{2}$.

Let p'_T denote the smeared transverse momentum. When the jets are reconstructed using calorimeter information, the smearing can be modeled as additive Gaussian noise, $p'_\mathrm{T} = p_\mathrm{T} + \varepsilon$, where $\varepsilon | p_\mathrm{T} \sim \mathcal{N}\left(0, \sigma(p_\mathrm{T})^2\right)$ with the variance $\sigma(p_\mathrm{T})^2$ satisfying (cf. Equation (2.3))

$$\left(\frac{\sigma(p_\mathrm{T})}{p_\mathrm{T}}\right)^2 = \left(\frac{N}{p_\mathrm{T}}\right)^2 + \left(\frac{S}{\sqrt{p_\mathrm{T}}}\right)^2 + C^2, \tag{3.37}$$

where N, S, and C are fixed positive constants (CMS Collaboration, 2010). The smeared intensity is then given by the convolution

$$g(p'_\mathrm{T}) = \int_E \mathcal{N}\left(p'_\mathrm{T} - p_\mathrm{T} \middle| 0, \sigma(p_\mathrm{T})^2\right) f(p_\mathrm{T})\,\mathrm{d}p_\mathrm{T}, \quad p'_\mathrm{T} \in F. \tag{3.38}$$

Hence the forward kernel is given by $k(p'_\mathrm{T}, p_\mathrm{T}) = N\left(p'_\mathrm{T} - p_\mathrm{T} \middle| 0, \sigma(p_\mathrm{T})^2\right)$ and the unfolding problem becomes a heteroscedastic Gaussian deconvolution problem.

At the center-of-mass energy $\sqrt{s} = 7$ TeV and in the central part of the CMS detector, realistic values for the parameters of $f(p_\mathrm{T})$ are given by $N_0 = 10^{17}$ fb/GeV, $\alpha = 5$, $\beta = 10$ and $\gamma = 10$ GeV and for the parameters of $\sigma(p_\mathrm{T})^2$ by $N = 1$ GeV, $S = 1$ GeV$^{1/2}$ and $C = 0.05$ (M. Voutilainen, personal communication, 2012). We furthermore set $L = 5.1\ \text{fb}^{-1}$, which corresponds to the size of the CMS 7 TeV dataset collected in 2011. We let the true and smeared spaces be $E = F = [400\ \text{GeV}, 1000\ \text{GeV}]$ and partition both spaces into $n = p = 30$ equal-width bins of the form (3.6) and (3.17). The resulting intensity functions are illustrated in Figure 3.3(b).

4 Current unfolding methods

This chapter provides an overview of the statistical methods that are presently used at the LHC for solving the unfolding problem. The focus is on the methods that are implemented in the RooUnfold software framework (Adye, 2011), which is the unfolding software that is most commonly used by LHC analyses. We explain in Section 4.1 that the problem is currently regularized by biasing the solution towards a Monte Carlo ansatz of the unknown intensity function f. The uncertainty of the solution is then quantified by using the variance of the regularized point estimator, ignoring the regularization bias. We demonstrate in Section 4.2 that this may lead to serious underestimation of the uncertainty, resulting in confidence intervals whose coverage probability can be much smaller than expected.

An accessible introduction to the statistical methodology used in HEP unfolding is given in Cowan (1998, Chapter 11). Blobel (2013) and Zech (2016) provide thorough reviews of current unfolding methodology and literature. Many statistical issues pertinent to unfolding at the LHC are also discussed in the proceedings of the unfolding workshop at the PHYSTAT 2011 conference (Prosper and Lyons, 2011).

4.1 Existing unfolding methodology

4.1.1 Overview

The unfolding methods that are presently used in LHC data analysis consist predominantly of those methods that are provided in the RooUnfold (Adye, 2011) software framework. These methods are:

(i) Bin-by-bin correction factors;

(ii) Matrix inversion;

(iii) SVD variant of Tikhonov regularization (Höcker and Kartvelishvili, 1996);

(iv) TUnfold variant of Tikhonov regularization (Schmitt, 2012);

(v) D'Agostini iteration with early stopping (D'Agostini, 1995).

All these methods use the smeared histogram $\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{\mu})$ to estimate $\boldsymbol{\lambda}$, the histogram version of the true spectum f, defined by Equation (3.18).

The *bin-by-bin correction factor method* (Cowan, 1998, Section 11.3) relates $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, which are assumed to share the same binning, by estimating the binwise ratios $C_i = \lambda_i / \mu_i$ using a Monte Carlo event generator and a detector simulation software. Unfolding is then performed with the help of these Monte Carlo correction factors $C_i^{\mathrm{MC}} = \lambda_i^{\mathrm{MC}} / \mu_i^{\mathrm{MC}}$ by estimating the true histogram using $\hat{\lambda}_i = C_i^{\mathrm{MC}} y_i$. This approach gives correct results only if the Monte Carlo simulation is performed using the correct, but obviously unknown, true spectrum. As such, the results have an extremely strong dependence on the MC ansatz, which makes it practically impossible to rigorously quantify the uncertainty associated with an unfolded spectrum obtained using bin-by-bin corrections. This is by now well-recognized by the LHC experiments and the method has been largely phased out in LHC data analysis, at least within the CMS collaboration. Historically the method has however been extensively used for example at the HERA experiments at DESY, at the Tevatron experiments at Fermilab and also in early LHC data analysis at CERN.

The other methods provided in RooUnfold relate $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ using the full response matrix $\boldsymbol{K}^{\mathrm{MC}}$ as described in Section 3.2.3. This yields the statistical model $\boldsymbol{y} \sim \mathrm{Poisson}\big(\boldsymbol{K}^{\mathrm{MC}}\boldsymbol{\lambda}\big)$. Also here the response matrix depends on the MC ansatz, but the dependence is not as strong as in the bin-by-bin method, which effectively replaces the full response matrix by a diagonal approximation. For brevity, we omit the superscript in $\boldsymbol{K}^{\mathrm{MC}}$ in the rest of this chapter, but it is worth keeping in mind that when histogram discretization is used the forward model is only approximate.

The *matrix inversion method* estimates $\boldsymbol{\lambda}$ using the relation $\boldsymbol{\mu} = \boldsymbol{K}\boldsymbol{\lambda}$ by simply plugging in $\boldsymbol{y}$ for $\boldsymbol{\mu}$ and then inverting the response matrix, that is, $\hat{\boldsymbol{\lambda}} = \boldsymbol{K}^{-1}\boldsymbol{y}$. This approach obviously requires that $\boldsymbol{K}$ be an invertible square matrix. Furthermore, since there is no regularization, the approach gives sensible results only if $\boldsymbol{K}$ is well-conditioned. This happens if the width of the forward kernels $k_i(s)$ is small in comparison to the size of the true bins E_j, in which case the response matrix becomes almost diagonal. If this is the case for true bins that are small enough so that the MC dependence of $\boldsymbol{K}$ can be safely neglected, then the unfolding problem is trivially solved using the matrix inverse. However, in most cases, the response matrix $\boldsymbol{K}$ is badly ill-conditioned, in which case the solution corresponding to the matrix inverse contains large, unphysical oscillations and some form of regularization is needed.

Two types of regularized unfolding are implemented in RooUnfold and widely used in LHC data analysis. The first approach regularizes the problem with the help of a Tikhonov-type penalty term which biases the unfolded solution $\hat{\boldsymbol{\lambda}}$ towards $\boldsymbol{\lambda}^{\mathrm{MC}}$, a Monte Carlo prediction of the true histogram,

$$\boldsymbol{\lambda}^{\mathrm{MC}} = \left[\int_{E_1} f^{\mathrm{MC}}(s)\,\mathrm{d}s, \ldots, \int_{E_p} f^{\mathrm{MC}}(s)\,\mathrm{d}s \right]^{\mathrm{T}}. \tag{4.1}$$

The second approach uses an expectation-maximization iteration to find the maximum likelihood estimator (MLE) of $\boldsymbol{\lambda}$ and regularizes the problem by stopping the iteration prematurely before convergence to the MLE. The iteration is started from $\boldsymbol{\lambda}^{\mathrm{MC}}$ and hence the early stopping introduces a bias towards the MC prediction. The following sections describe these approaches in greater detail.

4.1.2 Tikhonov regularization

The Tikhonov regularized techniques (Tikhonov, 1963; Phillips, 1962) estimate $\boldsymbol{\lambda}$ by solving

$$\hat{\boldsymbol{\lambda}} = \operatorname*{arg\,min}_{\boldsymbol{\lambda} \in \mathbb{R}^p} \left((\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\lambda})^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\lambda}) + 2\delta P(\boldsymbol{\lambda}) \right), \tag{4.2}$$

where $\hat{\boldsymbol{\Sigma}} = \mathrm{diag}(\boldsymbol{y}_+)$, with $y_{+,i} = \max(1, y_i)$, is an estimate of the covariance $\boldsymbol{\Sigma}$ of $\boldsymbol{y}$, $\delta > 0$ is a regularization parameter and $P(\boldsymbol{\lambda})$ is a penalty term that regularizes the otherwise ill-posed problem by penalizing physically implausible solutions. Here the first term is a Gaussian approximation to the Poisson likelihood under the model $\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{K}\boldsymbol{\lambda})$ and, as such, Tikhonov regularization can be regarded as approximate penalized maximum likelihood estimation; see Section 5.2. The regularization strength δ controls the relative importance of the data-fit term $(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\lambda})^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\lambda})$ and the penalty term $P(\boldsymbol{\lambda})$. Various data-driven procedures for selecting δ have been proposed; see Sections 4.2 and 5.3. We also note that there does not appear to be a standard way of handling zero event counts in the estimation of the covariance $\boldsymbol{\Sigma}$. In this work, we simply replace the zero counts by ones, but this detail has little to no impact on our results as we always consider situations where the probability of obtaining a zero count is very small.

Two variants of Tikhonov regularization, which differ in their choice of $P(\boldsymbol{\lambda})$, are commonly used at the LHC. In the *singular value decomposition* (SVD) *variant* (Höcker and Kartvelishvili, 1996), the penalty term is given by

$$P(\boldsymbol{\lambda}) = \left\| \mathbf{L} \begin{bmatrix} \lambda_1 / \lambda_1^{\mathrm{MC}} \\ \lambda_2 / \lambda_2^{\mathrm{MC}} \\ \vdots \\ \lambda_p / \lambda_p^{\mathrm{MC}} \end{bmatrix} \right\|_2^2, \tag{4.3}$$

where $\boldsymbol{\lambda}^{\mathrm{MC}} = \left(\lambda_j^{\mathrm{MC}}\right)_{j=1}^{p}$ is the Monte Carlo prediction of $\boldsymbol{\lambda}$ given by Equation (4.1) and the matrix $\boldsymbol{L}$ is given by

$$\boldsymbol{L} = \begin{bmatrix} -1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -1 \end{bmatrix}. \tag{4.4}$$

This corresponds to a discretized second derivative with reflexive boundary conditions (Hansen, 2010). In other words, the method favors solutions where the second derivative of the binwise ratio of $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^{\mathrm{MC}}$ is small. The method is named after the singular value decomposition following the numerical procedure used by Höcker and Kartvelishvili (1996), but the use of the SVD is by no means essential in the method—once δ and the form of $P(\boldsymbol{\lambda})$ are fixed, the same unfolded solution can be obtained using various numerical procedures.

The other form of Tikhonov regularization available in RooUnfold, and also as a standalone software package, is the TUnfold *variant* (Schmitt, 2012). In this case, the penalty term is of the form

$$P(\boldsymbol{\lambda}) = \|\boldsymbol{L}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0)\|_2^2, \tag{4.5}$$

where, by default, $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}^{\mathrm{MC}}$ and various choices are available for $\boldsymbol{L}$, including, but not limited to, the identity matrix, the discretized first derivative and the discretized second derivative. In other words, by default, both implementations of Tikhonov regularization introduce a bias towards a Monte Carlo ansatz of $\boldsymbol{\lambda}$ and this bias can be sizable even in cases where the MC dependence of the response matrix $\boldsymbol{K}$ is small.

With a calculation similar to Sections 5.2.3 and 5.2.4, one can easily show that, for the SVD penalty (4.3), the solution of (4.2) is given by

$$\hat{\boldsymbol{\lambda}} = \left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\delta\tilde{\boldsymbol{L}}^{\mathrm{T}}\tilde{\boldsymbol{L}}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y}, \tag{4.6}$$

where $\tilde{\boldsymbol{L}} = \boldsymbol{L}\,\mathrm{diag}(\boldsymbol{\lambda}^{\mathrm{MC}})^{-1}$. Similarly, for the TUnfold penalty (4.5), the estimator is given by

$$\hat{\boldsymbol{\lambda}} = \left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\delta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\right)^{-1}\left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y} + 2\delta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{\lambda}_0\right). \tag{4.7}$$

Notice that both estimators are available in closed form and, if we ignore the data-dependence of $\hat{\boldsymbol{\Sigma}}$ (and possibly also the data-dependence of an estimator of δ), the SVD solution is a linear transformation of $\boldsymbol{y}$ and the TUnfold solution an affine transformation of $\boldsymbol{y}$.

Both methods quantify the uncertainty of $\hat{\boldsymbol{\lambda}}$ using $1-\alpha$ Gaussian confidence intervals (see

Section 6.1.3) given by

$$\left[\hat{\lambda}_j - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}\left(\hat{\lambda}_j\right)}, \hat{\lambda}_j + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}\left(\hat{\lambda}_j\right)}\right], \quad j = 1, \ldots, p, \tag{4.8}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ standard normal quantile and $\widehat{\mathrm{var}}\left(\hat{\lambda}_j\right)$ is the estimated variance of $\hat{\lambda}_j$. The intervals are typically computed at 68.3 % confidence level, that is, with $z_{1-\alpha/2} = 1$. Using its linearity, the covariance of the SVD estimator can be estimated with

$$\widehat{\mathrm{cov}}\left(\hat{\boldsymbol{\lambda}}\right) = \left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\delta\tilde{\boldsymbol{L}}^{\mathrm{T}}\tilde{\boldsymbol{L}}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K}\left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\delta\tilde{\boldsymbol{L}}^{\mathrm{T}}\tilde{\boldsymbol{L}}\right)^{-1}. \tag{4.9}$$

Similarly, for TUNFOLD, we can estimate

$$\widehat{\mathrm{cov}}\left(\hat{\boldsymbol{\lambda}}\right) = \left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\delta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K}\left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\delta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\right)^{-1}. \tag{4.10}$$

The variances in Equation (4.8) are then estimated using the diagonal elements of the covariance matrix, $\widehat{\mathrm{var}}\left(\hat{\lambda}_j\right) = \widehat{\mathrm{cov}}\left(\hat{\boldsymbol{\lambda}}\right)_{j,j}$. Notice that in SVD the MC prediction $\boldsymbol{\lambda}^{\mathrm{MC}}$ affects both the length and the location of the confidence intervals, while in TUNFOLD it only affects the location of the intervals.

4.1.3 D'Agostini iteration

The *D'Agostini method* (D'Agostini, 1995) solves the unfolding problem iteratively. Given a starting point $\boldsymbol{\lambda}^{(0)} > \boldsymbol{0}$, the $(t+1)$th step of the iteration is

$$\lambda_j^{(t+1)} = \frac{\lambda_j^{(t)}}{\sum_{i=1}^n K_{i,j}} \sum_{i=1}^n \frac{K_{i,j} y_i}{\sum_{k=1}^p K_{i,k}\lambda_k^{(t)}}. \tag{4.11}$$

This method can be derived as an expectation-maximization (EM) iteration (Dempster et al., 1977) for finding the MLE of $\boldsymbol{\lambda}$ in the Poisson regression problem $\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{K\lambda})$, $\boldsymbol{\lambda} \geq \boldsymbol{0}$. The derivation is given, for example, in McLachlan and Krishnan (2008, Section 2.5) or Kuusela (2012, Section 4.1.2); see also Section 5.3.1 for a brief description of the EM algorithm. The standard EM convergence theorems of Wu (1983) are not directly applicable to the D'Agostini iteration, but Vardi et al. (1985, Theorem A.1) show that the iteration (4.11) converges to an MLE of $\boldsymbol{\lambda}$, that is, $\boldsymbol{\lambda}^{(t)} \xrightarrow{t\to\infty} \hat{\boldsymbol{\lambda}}_{\mathrm{MLE}}$, where $\hat{\boldsymbol{\lambda}}_{\mathrm{MLE}} \in \mathbb{R}_+^p$ is a global maximizer of the likelihood function

$$L(\boldsymbol{\lambda}; \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\lambda}) = \prod_{i=1}^n \frac{\left(\sum_{j=1}^p K_{i,j}\lambda_j\right)^{y_i}}{y_i!} e^{-\sum_{j=1}^p K_{i,j}\lambda_j}, \quad \boldsymbol{\lambda} \in \mathbb{R}_+^p. \tag{4.12}$$

The convergence result holds even when the MLE is not unique (that is, when $\boldsymbol{K}$ does not have full column rank).

Regularization is achieved by stopping the iteration (4.11) prematurely before convergence to the MLE. Denoting the number of iterations by T, the regularized point estimator is $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^{(T)}$. The effect of this is to create a bias towards $\boldsymbol{\lambda}^{(0)}$, the starting point of the iteration, and the number of iterations T serves as a regularization strength controlling the bias-variance tradeoff of the solution. In the RooUnfold implementation of the method, the starting point is the MC prediction, $\boldsymbol{\lambda}^{(0)} = \boldsymbol{\lambda}^{\mathrm{MC}}$, where $\boldsymbol{\lambda}^{\mathrm{MC}}$ is given by Equation (4.1).

RooUnfold quantifies the uncertainty in the D'Agostini iteration using the Gaussian confidence intervals (4.8) with $z_{1-\alpha/2} = 1$ and with the variance $\widehat{\mathrm{var}}\big(\hat{\lambda}_j\big)$ estimated using error propagation. That is, by using a linearized approximation of the iteration (4.11), the covariance of $\boldsymbol{\lambda}^{(t+1)}$ can be estimated (Cowan, 1998, Section 1.6) using

$$\widehat{\mathrm{cov}}\big(\boldsymbol{\lambda}^{(t+1)}\big) = \boldsymbol{J}^{(t+1)}\,\hat{\boldsymbol{\Sigma}}\,(\boldsymbol{J}^{(t+1)})^{\mathrm{T}}, \tag{4.13}$$

where $\boldsymbol{J}^{(t+1)}$ is the Jacobian of $\boldsymbol{\lambda}^{(t+1)}$ evaluated at $\boldsymbol{y}$ and $\hat{\boldsymbol{\Sigma}} = \mathrm{diag}(\boldsymbol{y}_+)$ is an estimate of the covariance of $\boldsymbol{y}$. Denote $\varepsilon_j = \sum_{i=1}^{n} K_{i,j}$ and $M_{i,j}^{(t)} = \frac{\lambda_j^{(t)}}{\varepsilon_j} \frac{K_{i,j}}{\sum_{k=1}^{p} K_{i,k}\lambda_k^{(t)}}$. Then the elements of the Jacobian are given by (Adye, 2011)

$$J_{j,i}^{(t+1)} = \frac{\partial \lambda_j^{(t+1)}}{\partial y_i} = M_{i,j}^{(t)} + \frac{\lambda_j^{(t+1)}}{\lambda_j^{(t)}} J_{j,i}^{(t)} - \sum_{k=1}^{p} \sum_{l=1}^{n} y_l \frac{\varepsilon_k}{\lambda_k^{(t)}} M_{l,j}^{(t)} M_{l,k}^{(t)} J_{k,i}^{(t)}, \tag{4.14}$$

where $J_{j,i}^{(0)} = 0$ for all j, i. The variances are then estimated using the diagonal elements of the covariance, $\widehat{\mathrm{var}}\big(\hat{\lambda}_j\big) = \widehat{\mathrm{cov}}\big(\hat{\boldsymbol{\lambda}}\big)_{j,j} = \widehat{\mathrm{cov}}\big(\boldsymbol{\lambda}^{(T)}\big)_{j,j}$. As usual, this estimate ignores the data-dependence of T. It should also be mentioned that the error propagation procedure in D'Agostini (1995, Section 4) omits the last two terms in Equation (4.14) and hence underestimates the variance of $\hat{\boldsymbol{\lambda}}$ (Adye, 2011). This has been corrected in RooUnfold, which calculates the variance using the full Jacobian (4.14).

The iteration (4.11) has been discovered independently in various fields where the experimental data consist of Poisson counts. The first authors to derive the method seem to have been Richardson (1972) in optics and Lucy (1974) in astronomy. In these fields, the iteration is typically called the *Richardson–Lucy algorithm.* Shepp and Vardi (1982), Lange and Carson (1984) and Vardi et al. (1985) apply the method to image reconstruction in positron emission tomography. In HEP, the method was popularized by the work of D'Agostini (1995), but it was already studied earlier by Kondor (1983) and Mülthei and Schorr (1987a,b, 1989). See also the recent review by Zech (2013). In the HEP community, the D'Agostini iteration is sometimes called "Bayesian" unfolding since D'Agostini (1995) derives the iteration using a repeated application of Bayes' theorem. However, this does not correspond to the way statisticians usually understand Bayesian inference. The author's position is that the iteration is best understood as a fully frequentist technique for finding the (regularized) MLE of $\boldsymbol{\lambda}$ and should not be called "Bayesian".

4.1.4 Other methods

Several other methods, which we review here briefly, have also been proposed for solving the HEP unfolding problem. RUN (Blobel, 1985, 1996) is a penalized maximum likelihood method (see Section 5.2.2) with a roughness penalty and B-spline discretization of the unfolded space. TRUEE (Milke et al., 2013) is a C++ implemenation of RUN. ARU (Dembinski and Roth, 2013) is similar to RUN, but uses an entropy penalty instead of a second derivative penalty. Volobouev (2015) proposes an approach similar to Silverman et al. (1990) where the EM iteration is smoothed on each step of the algorithm. Smoothing is also considered in the original paper by D'Agostini (1995), but by default ROOUNFOLD does not smooth the EM iteration (Adye, 2011). Choudalakis (2012) proposes to solve the problem using Bayesian inference, but does not address the choice of the regularization strength.

4.2 Undercoverage of the unfolded confidence intervals

The two most common regularized unfolding methods in current use at the LHC are the SVD variant of Tikhonov regularization (Höcker and Kartvelishvili, 1996) and the D'Agostini iteration (D'Agostini, 1995) with early stopping. As explained in Section 4.1 both of these methods are regularized by biasing the solution towards a Monte Carlo prediction $\boldsymbol{\lambda}^{\mathrm{MC}}$ of the true histogram $\boldsymbol{\lambda}$. In this section, we demonstrate that if the uncertainty of the unfolded solution is quantified using the Gaussian confidence intervals of Equation (4.8), which is essentially what is implemented in current unfolding software (Adye, 2011; Schmitt, 2012), then the resulting confidence intervals may suffer from serious undercoverage. This happens because the confidence intervals ignore the bias from the regularization and discretization of the problem. In our simulations, we use fairly small bins in the true space and hence the first effect is the dominant one. With larger bins, the discretization bias would also contribute to the undercoverage.

We study the coverage of the methods using 1 000 independent replications of the two unfolding scenarios described in Section 3.4. In each case, we take the MC ansatz f^{MC}, which is used to compute the smearing matrix in Equation (3.23) and the MC prediction $\boldsymbol{\lambda}^{\mathrm{MC}}$ in Equation (4.1), to be a slightly perturbed version of the unknown true spectrum f. For the two peaks on a uniform background test case (Section 3.4.1), we consider the medium sample size $\lambda_{\mathrm{tot}} = 10\,000$ and use the MC ansatz

$$f^{\mathrm{MC}}(s) = \lambda_{\mathrm{tot}} \left\{ \pi_1 \mathcal{N}\left(s \,\middle|\, -2, 1.1^2\right) + \pi_2 \mathcal{N}\left(s \,\middle|\, 2, 0.9^2\right) + \pi_3 \frac{1}{|E|} \right\}, \tag{4.15}$$

where λ_{tot} and π_i, $i = 1, \ldots, 3$, are set to their correct values. In other words, we assume that the MC event generator predicts a spectrum where the left peak is slightly wider and the right peak slightly narrower than in f. For the inclusive jet transverse momentum spectrum (Section 3.4.2), we assume that the MC spectrum is given by Equation (3.36) with the parameter values $N_0 = 5.5 \cdot 10^{19}$ fb/GeV, $\alpha = 6$ and $\beta = 12$, with the rest of the parameters set to their correct values. In other words, we take the MC spectrum to fall off slightly faster than f in both the power-law and the energy cut-off terms. The value of N_0, which has little effect on the results, is chosen so that the overall scale of f^{MC} is similar to f.

The coverage probability of the unfolded confidence intervals depends strongly on the regularization strength, which in the SVD method is controlled by the parameter δ and in the D'Agostini method by the number of iterations T. Figure 4.1 shows the empirical simultaneous coverage of Bonferroni-corrected 95 % Gaussian confidence intervals as a function of the regularization strength for the two methods and for the two peaks on a uniform background test setup. (We plot the Bonferroni-corrected simultaneous coverage instead of the binwise coverage in order to be able visualize the effects on a single graph. If the binwise coverage is correct, it implies that the Bonferroni-corrected simultaneous coverage is correct; see Section 3.3.1.) With SVD unfolding, weak regularization yields nominal coverage, but, as the regularization strength is increased, the coverage eventually drops to zero. With D'Agostini iteration, the effect is qualitatively similar, with the difference that even with weak regularization (i.e., with a large number of iterations) the empirical coverage appears to be slightly below the nominal value of 95 %. The most likely explanation for this is that the linearization in Equation (4.13) becomes inaccurate with a large number of iterations. Figure 4.2 shows the same experiments for the inclusive jet transverse momentum spectrum. Also in this case the coverage varies between zero and the nominal value depending on the choice of the regularization strength (here also D'Agostini seems to eventually attain nominal coverage).

An obvious question to ask is where along the coverage curves of Figures 4.1 and 4.2 do typical unfolding results lie? Unfortunately it is impossible to tell as the answer depends on the quality of the MC predictions f^{MC} and the way the regularization strength is chosen. There currently exists no standard way of choosing the regularization strength in LHC data analysis and many analyses make the choice using various non-standard heuristics. For example, ROOUNFOLD documentation (Adye, 2011) recommends simply using four iterations for the D'Agostini method, and many LHC analyses indeed seem to follow this convention. There is, however, no principled reason for using four iterations which in our simulations would result in serious undercoverage for the two peaks on a uniform background test setup and *zero* simultaneous coverage for the inclusive jet transverse momentum spectrum; see Figures 4.1(b) and 4.2(b).

To obtain an idea of the coverage performance when the regularization strength is chosen in a principled data-driven way, we study the empirical coverage of the methods with the regularization strength chosen using weighted cross-validation as described in Section 5.3.3. Figures 4.3 and 4.4 show the binwise coverage of the resulting 95 % Gaussian confidence intervals (without multiplicity correction) and the simultaneous coverage of the corresponding Bonferroni-corrected intervals. The former figure is for the two peaks on a uniform background test setup and the latter for the inclusive jet transverse momentum spectrum (in both cases, the D'Agostini runs where the cross-validation score was still decreasing after 20 000 iterations were discarded from the analysis; there were 4 such runs in the first case and 3 in the second case). We see that in both test cases and for both unfolding methods the confidence intervals

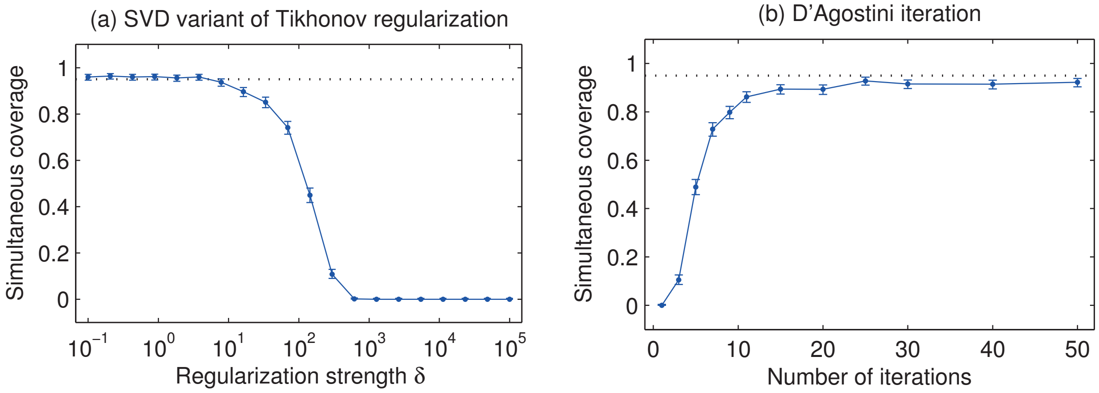


Figure 4.1: Empirical simultaneous coverage of Bonferroni-corrected 95 % Gaussian confidence intervals as a function of the regularization strength for the two peaks on a uniform background test setup. Figure (a) shows the coverage for the SVD variant of Tikhonov regularization and Figure (b) for the D'Agostini iteration. The error bars are the 95 % Clopper–Pearson intervals and the nominal confidence level is shown by the horizontal dotted line. With strong regularization, both methods undercover substantially.

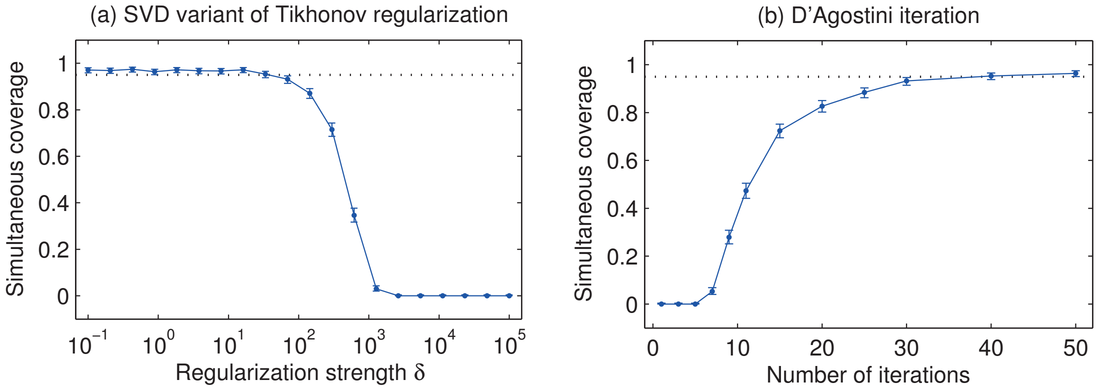


Figure 4.2: Same as Figure 4.1, but for the inclusive jet transverse momentum spectrum.

suffer from substantial undercoverage. In the two-peak case, the undercoverage is the worst around the larger peak on the right, while, in the case of the steeply falling jet transverse momentum spectrum, the undercoverage occurs mostly in the tails of the spectrum. By varying the Monte Carlo prediction f^{MC}, one can easily obtain even further reductions in the coverage.

Some caution should however be exercised in interpreting these findings in the context of existing LHC unfolding results. Namely, the current LHC practice is to treat the Monte Carlo model dependence as a systematic uncertainty of the unfolded result. A common way to take this uncertainty into account is to compute the unfolded histograms using two or more Monte Carlo event generators and to use the observed differences as an estimate of the systematic

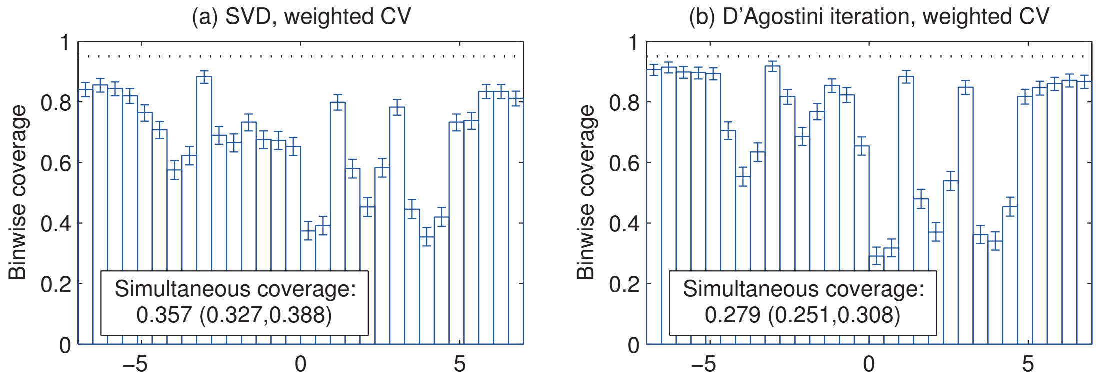


Figure 4.3: Empirical binwise and simultaneous coverage of the 95 % Gaussian confidence intervals for the two peaks on a uniform background test setup when the regularization strength is chosen using weighted cross-validation. Figure (a) shows the coverage for the SVD variant of Tikhonov regularization and Figure (b) for the D'Agostini iteration. The simultaneous coverage is given for Bonferroni-corrected intervals. The uncertainties are the 95 % Clopper–Pearson intervals and the nominal confidence level is shown by the horizontal dotted line. Both methods suffer from severe undercoverage.

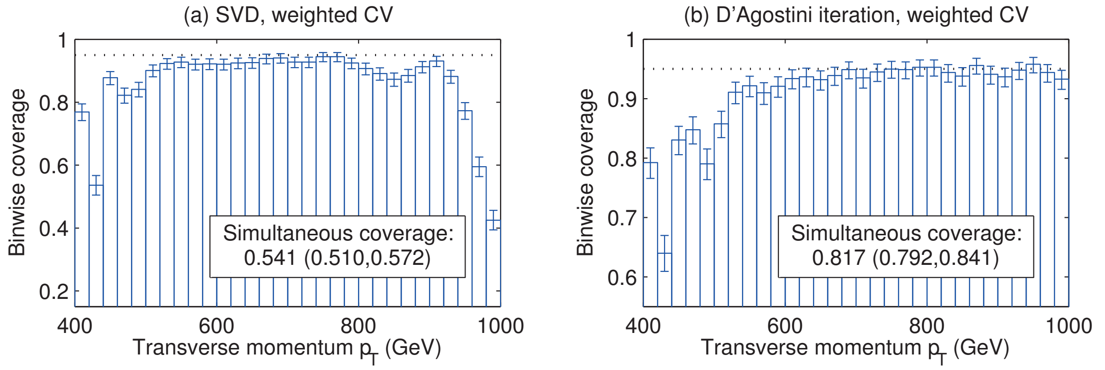


Figure 4.4: Same as Figure 4.3, but for the inclusive jet transverse momentum spectrum.

uncertainty. The success of this approach obviously depends on how well the Monte Carlo models represent the space of plausible truths and on whether the true f is in some sense "bracketed" by these models.

We conclude that the coverage of the existing unfolding methods is a delicate function of the Monte Carlo prediction f^{MC}, the size of the true bins E_j, the choice of the regularization strength and the way these factors are taken into account as systematic uncertainties. With some luck, the coverage may be close to the nominal value, but there is ample room for error. In general, it seems very difficult, if not impossible, to provide rigorous statistical guarantees for these methods, especially when the regularization relies so heavily on a specific ansatz of the unknown intensity function f.

5 Point estimation of smooth spectra

The construction of confidence intervals for smooth spectra typically proceeds in two steps: one first forms a regularized point estimate of the spectrum and then uses the variability of this estimator to quantify the uncertainty of the solution. In order to do this, we need to make the following choices:

(i) Which regularized point estimator $\hat{f}$ to use?

(ii) How to choose the regularization strength?

(iii) How to use the variability of $\hat{f}$ to form confidence intervals?

In this chapter, we focus on the first two questions. We first explain in Section 5.1 how a priori information about smoothness can be introduced using a Bayesian smoothness prior. We then present in Section 5.2 various Bayesian and frequentist unfolded point estimators and draw connections between the two paradigms. In Section 5.3, we focus on the crucial question concerning the choice of the regularization strength and present empirical Bayes, hierarchical Bayes and cross-validation techniques for making this choice. We argue in favor of empirical Bayes and support our argument with a simulation study presented in Section 5.4.

5.1 Regularization using a smoothness prior

Let us consider the spline discretization of the unfolding problem (see Section 3.2.2) leading to the forward model

$$\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{K}\boldsymbol{\beta}), \tag{5.1}$$

where the spline coefficients $\boldsymbol{\beta} \in \mathbb{R}_+^p$ in order to enforce the positivity constraint and the elements of $\boldsymbol{K}$ are given by

$$K_{i,j} = \int_{F_i} \int_E k(t,s) B_j(s) \,\mathrm{d}s \,\mathrm{d}t, \quad i = 1, \ldots, n, \quad j = 1, \ldots, p. \tag{5.2}$$

In the Bayesian framework, inferences concerning $\boldsymbol{\beta}$ are based on the *posterior distribution*, whose density is given by

$$p(\boldsymbol{\beta}|\boldsymbol{y},\delta) = \frac{p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta}|\delta)}{p(\boldsymbol{y}|\delta)} = \frac{p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta}|\delta)}{\int_{\mathbb{R}_+^p} p(\boldsymbol{y}|\boldsymbol{\beta}')p(\boldsymbol{\beta}'|\delta)\,\mathrm{d}\boldsymbol{\beta}'}, \quad \boldsymbol{\beta} \in \mathbb{R}_+^p. \tag{5.3}$$

Here $p(\boldsymbol{y}|\boldsymbol{\beta})$ is the Poisson likelihood function given by

$$p(\boldsymbol{y}|\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\left(\sum_{j=1}^{p} K_{i,j}\beta_j\right)^{y_i}}{y_i!} e^{-\sum_{j=1}^{p} K_{i,j}\beta_j}, \quad \boldsymbol{\beta} \in \mathbb{R}_+^p, \tag{5.4}$$

and $p(\boldsymbol{\beta}|\delta)$ is the density of the *prior distribution* which is used to regularize the otherwise ill-posed problem and which may depend on some hyperparameter δ. The denominator $p(\boldsymbol{y}|\delta)$ is the *marginal density* of the smeared observations $\boldsymbol{y}$. Seen as a function of δ, it may also be understood as the likelihood of the hyperparameter δ in which case it is called the *marginal likelihood function*.

When we have a priori information that the true intensity f should be smooth, the prior $p(\boldsymbol{\beta}|\delta)$ should be chosen to reflect this. In this work, we consider the Gaussian smoothness prior

$$p(\boldsymbol{\beta}|\delta) \propto \exp\left(-\delta\|f''\|_2^2\right) = \exp\left(-\delta\int_E \{f''(s)\}^2\,\mathrm{d}s\right) \tag{5.5}$$

$$= \exp\left(-\delta\int_E \left\{\left(\sum_{j=1}^{p}\beta_j B_j(s)\right)''\right\}^2 \mathrm{d}s\right) \tag{5.6}$$

$$= \exp\left(-\delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}\right), \quad \boldsymbol{\beta} \in \mathbb{R}_+^p, \quad \delta > 0, \tag{5.7}$$

where $\boldsymbol{\Omega}$ is a $p \times p$ matrix whose elements are given by $\Omega_{i,j} = \int_E B_i''(s)B_j''(s)\,\mathrm{d}s$. The probability mass of this prior is concentrated on positive intensities whose curvature $\|f''\|_2^2$ is small. In other words, the prior favors smooth solutions.

The hyperparameter δ controls the concentration of the prior distribution and serves as a regularization strength that controls the trade-off between smooth solutions and fidelity to the data. For large values of δ, the prior is concentrated near the origin resulting in strong regularization, while, for small values of δ, the prior becomes nearly flat on $\mathbb{R}_+^p$ which corresponds to weak regularization. The choice of an appropriate value of δ is explored in Section 5.3.

5.1.1 Proper smoothness prior using boundary conditions

The prior defined in Equation (5.7) is improper. Indeed, the rank of the matrix $\boldsymbol{\Omega}$ is $p-2$ corresponding to the two-dimensional null space of the second derivative operator. Take for example $\boldsymbol{\beta} = a\mathbf{1}$ with $a > 0$. Then

$$p(\boldsymbol{\beta}|\delta) \propto \exp\left(-\delta \int_E \left\{\left(a \sum_{j=1}^{p} B_j(s)\right)''\right\}^2 \mathrm{d}s\right) = \exp(0) = 1, \quad \forall a > 0, \tag{5.8}$$

where we have used the property that for B-splines $\sum_{j=1}^{p} B_j(s) = 1, \forall s \in E$. In other words, the prior density is constant in the direction of the vector $\mathbf{1}$ and cannot be normalized in this direction.

In principle, this is not an issue, since the posterior density $p(\boldsymbol{\beta}|\boldsymbol{y},\delta)$ would still be proper. Indeed, Kuusela (2012, Proposition 5.3) showed the following result:

Proposition 5.1. *Consider the likelihood $p(\boldsymbol{y}|\boldsymbol{\beta})$ given by* (5.4) *and assume that $K_{i,j} > 0, \forall i,j$. Let $q(\boldsymbol{\beta}|\delta) \geq 0$ be an unnormalized, possibly improper prior density for $\boldsymbol{\beta}$ depending on hyperparameter δ. Then the unnormalized posterior $q(\boldsymbol{\beta}|\boldsymbol{y},\delta) = p(\boldsymbol{y}|\boldsymbol{\beta})q(\boldsymbol{\beta}|\delta)$ is normalizable, that is, $\int_{\mathbb{R}_+^p} q(\boldsymbol{\beta}|\boldsymbol{y},\delta)\,\mathrm{d}\boldsymbol{\beta} \in (0,\infty)$, if the prior density function is bounded and its support is a subset of $\mathbb{R}_+^p$ of strictly positive Lebesgue measure.*

The problem, however, is that, for an improper prior, it becomes difficult to interpret $p(\boldsymbol{y}|\delta)$ as a marginal likelihood. Indeed, if $p(\boldsymbol{\beta}|\delta)$ is improper, we have

$$\sum_{\boldsymbol{y}\in\mathbb{N}_0^n} p(\boldsymbol{y}|\delta) = \sum_{\boldsymbol{y}\in\mathbb{N}_0^n} \int_{\mathbb{R}_+^p} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta}|\delta)\,\mathrm{d}\boldsymbol{\beta} \tag{5.9}$$

$$= \int_{\mathbb{R}_+^p} \sum_{\boldsymbol{y}\in\mathbb{N}_0^n} p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta}|\delta)\,\mathrm{d}\boldsymbol{\beta} = \int_{\mathbb{R}_+^p} p(\boldsymbol{\beta}|\delta)\,\mathrm{d}\boldsymbol{\beta} = \infty \tag{5.10}$$

and $p(\boldsymbol{y}|\delta)$ is not a proper probability density. The interchange of summation and integration is allowed here by the monotone convergence theorem since the summands $p(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta}|\delta)$ are positive for each $\boldsymbol{y} \in \mathbb{N}_0^n$. As we are particularly interested in empirical Bayes procedures that rely on the marginal likelihood $p(\boldsymbol{y}|\delta)$, we prefer to use a smoothness prior that can be properly normalized in order to avoid any undesired complications.

To obtain a normalizable smoothness prior, we replace the matrix $\boldsymbol{\Omega}$ in Equation (5.7) by an augmented version $\boldsymbol{\Omega}_{\mathrm{A}}$, whose elements are given by

$$\Omega_{\mathrm{A},i,j} = \begin{cases} \Omega_{i,j} + \gamma_{\mathrm{L}}, & \text{if } i = j = 1, \\ \Omega_{i,j} + \gamma_{\mathrm{R}}, & \text{if } i = j = p, \\ \Omega_{i,j}, & \text{otherwise,} \end{cases} \tag{5.11}$$

where $\gamma_{\mathrm{L}}, \gamma_{\mathrm{R}} > 0$ are fixed constants. In other words, we add the constant γ_{L} to the first element on the diagonal and the constant γ_{R} to the last element on the diagonal. The augmented matrix $\boldsymbol{\Omega}_{\mathrm{A}}$ is positive definite and hence defines a proper prior density.

This change can be justified by imposing *Aristotelian boundary conditions* (Calvetti et al.,

2006) on f. The idea is to condition on the boundary values $f(E_{\min})$ and $f(E_{\max})$ and then introduce additional hyperpriors for these values. Since $f(E_{\min}) = \beta_1 B_1(E_{\min})$ and $f(E_{\max}) = \beta_p B_p(E_{\max})$, we can equivalently condition on β_1 and β_p. Assuming that β_1 and β_p are independent in the hyperprior, the smoothness prior becomes

$$p(\boldsymbol{\beta}|\delta) = p(\beta_2,\dots,\beta_{p-1}|\beta_1,\beta_p,\delta)\,p(\beta_1|\delta)\,p(\beta_p|\delta), \quad \boldsymbol{\beta} \in \mathbb{R}^p_+, \tag{5.12}$$

with $p(\beta_2,\dots,\beta_{p-1}|\beta_1,\beta_p,\delta) \propto \exp\left(-\delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}\right)$. It is convenient to take the hyperpriors to be the truncated Gaussians

$$p(\beta_1|\delta) \propto \exp\left(-\delta\gamma_{\mathrm{L}}\beta_1^2\right), \quad \beta_1 \geq 0, \tag{5.13}$$

$$p(\beta_p|\delta) \propto \exp\left(-\delta\gamma_{\mathrm{R}}\beta_p^2\right), \quad \beta_p \geq 0. \tag{5.14}$$

The prior is then given by

$$p(\boldsymbol{\beta}|\delta) \propto \exp\left(-\delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}\right), \quad \boldsymbol{\beta} \in \mathbb{R}^p_+, \tag{5.15}$$

where the elements of $\boldsymbol{\Omega}_{\mathrm{A}}$ are given by Equation (5.11).

In addition to defining a proper smoothness prior, the Aristotelian boundary conditions provide further regularization on the boundaries. Namely, Equations (5.13) and (5.14) can be understood to penalize large values of f on the boundaries. As a result, the construction helps to keep the variance of the unfolded intensity under control near the boundaries.

5.2 Regularized point estimators

In this section, we introduce various regularized point estimators motivated by the Bayesian framework described in Section 5.1. We present the estimators in decreasing order of computational complexity. The price to pay for this increasing computational tractability is a decreasing fidelity to the original Bayesian formulation. In the author's experience, one can also expect the quality of the unfolded point estimates to roughly follow the order in which the estimators are presented here, with the posterior mean typically giving the best point estimation performance. However, for reasonable sample sizes and for a fixed regularization strength δ, the differences between these estimators tend to be rather small in practice.

5.2.1 Posterior mean estimation

The most common Bayesian point estimate is the *posterior mean*

$$\hat{\boldsymbol{\beta}}_{\mathrm{PM}} = \mathrm{E}\left(\boldsymbol{\beta}|\boldsymbol{y},\delta\right) = \int_{\mathbb{R}^p_+} \boldsymbol{\beta}\, p(\boldsymbol{\beta}|\boldsymbol{y},\delta)\,\mathrm{d}\boldsymbol{\beta}. \tag{5.16}$$

The use of the posterior mean can be justified using a decision-theoretic argument: namely, the posterior mean minimizes the posterior expected loss under the ℓ^2 loss function $L(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}) = \|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|_2^2$ (Bernardo and Smith, 2000, Proposition 5.2). Heuristically, it is a point estimate that corresponds to the center of mass of the posterior distribution.

In practice, the posterior mean cannot be obtained in closed form for the unfolding problem. The problem is that, for the Poisson likelihood function (5.4), the denominator of Bayes' rule in Equation (5.3) and the integral in Equation (5.16) are intractable high-dimensional integrals. A numerical approximation can nevertheless be obtained using Markov chain Monte Carlo (MCMC) sampling (Robert and Casella, 2004), where the main idea is to construct a Markov chain on $\mathbb{R}_+^p$ whose equilibrium distribution is the posterior $p(\boldsymbol{\beta}|\boldsymbol{y}, \delta)$. An approximation of the posterior mean can then be obtained as the empirical mean of a realization of this Markov chain. Under regularity conditions for the MCMC sampler, one can show that the empirical mean converges almost surely to the true posterior mean as the length of the chain tends to infinity (Tierney, 1994, Theorem 3).

It turns out that the basic MCMC samplers are not well-suited for the unfolding problem: the Gibbs sampler is computationally inefficient since no fast algorithms exist for sampling from the non-standard full posterior conditionals; and, for the multivariate Metropolis–Hastings sampler, it is difficult to find good proposal distributions since the posterior tends to have very different scales for different components of $\boldsymbol{\beta}$. To overcome these difficulties, we adopt the single-component Metropolis–Hastings (also known as the Metropolis-within-Gibbs) sampler of Saquib et al. (1998, Section III.C). Denoting $\boldsymbol{\beta}_{-k} = \left[\beta_1, \ldots, \beta_{k-1}, \beta_{k+1}, \ldots, \beta_p\right]^{\mathrm{T}}$, the sampler replaces the full posterior conditional of the Gibbs sampler $p(\beta_k|\boldsymbol{\beta}_{-k}, \boldsymbol{y}, \delta)$ with a more tractable approximation (Gilks et al., 1996; Gilks, 1996). One then samples from this approximate full posterior conditional and corrects for the approximation error using a Metropolis–Hastings acceptance step. In our case, we approximate the full posterior conditionals using either truncated Gaussians or the exponential distribution. The details of the sampler are given in Appendix A.1.

5.2.2 Maximum a posteriori and penalized maximum likelihood estimation

Another common Bayesian point estimator is the *maximum a posteriori* (MAP) *estimator*

$$\hat{\boldsymbol{\beta}}_{\mathrm{MAP}} = \underset{\boldsymbol{\beta} \in \mathbb{R}_+^p}{\arg\max}\; p(\boldsymbol{\beta}|\boldsymbol{y}, \delta) \tag{5.17}$$

corresponding to the mode of the posterior density. The decision-theoretic justification for the MAP estimator is that it arises as the limit of decisions minimizing the posterior expected loss for the 0-1 loss function $L_\varepsilon(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}) = \mathbf{1}\left\{\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\|_2 > \varepsilon\right\}$ as $\varepsilon \to 0$ (Bernardo and Smith, 2000, Proposition 5.2). Heuristically, the MAP estimator is sensitive to the height of the posterior probability density and it can differ substantially from the center of posterior probability mass given by the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ in cases where the posterior is multimodal or skewed.

Notice that

$$\hat{\boldsymbol{\beta}}_{\mathrm{MAP}} = \underset{\boldsymbol{\beta}\in\mathbb{R}_+^p}{\arg\max}\, \log p(\boldsymbol{\beta}|\boldsymbol{y},\delta) = \underset{\boldsymbol{\beta}\in\mathbb{R}_+^p}{\arg\max}\, \big(\log p(\boldsymbol{y}|\boldsymbol{\beta}) + \log p(\boldsymbol{\beta}|\delta)\big) \tag{5.18}$$

$$= \underset{\boldsymbol{\beta}\in\mathbb{R}_+^p}{\arg\max}\, \big(l(\boldsymbol{\beta};\boldsymbol{y}) - \delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}\big) = \hat{\boldsymbol{\beta}}_{\mathrm{PMLE}}, \tag{5.19}$$

where $l(\boldsymbol{\beta};\boldsymbol{y}) = \log p(\boldsymbol{y}|\boldsymbol{\beta})$ is the log-likelihood of $\boldsymbol{\beta}$. This shows that the MAP estimator $\hat{\boldsymbol{\beta}}_{\mathrm{MAP}}$ with the Gaussian smoothness prior coincides with the penalized maximum likelihood estimator (PMLE) $\hat{\boldsymbol{\beta}}_{\mathrm{PMLE}}$ with the roughness penalty $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}$. In other words, the Bayesian and frequentist solutions of the problems are the same. This shows that in ill-posed inverse problems the two paradigms are closely related and a Bayesian prior can be understood as a frequentist penalty term, and vice versa.

The MAP/PMLE solution is easier to compute than the posterior mean since obtaining the estimator involves solving a non-linear optimization problem instead of computing a high-dimensional integral. The estimator can be computed using standard non-linear programming techniques (see, e.g., Nocedal and Wright (2006) and Bazaraa et al. (2006)) or using the problem-specific one-step-late expectation-maximization algorithm (Green, 1990a,b).

5.2.3 Positivity-constrained Gaussian approximation

A further computational simplification can be obtained by using a Gaussian approximation to the Poisson likelihood in the MAP/PMLE solution. For large intensities, the smeared histogram $\boldsymbol{y}$ is approximately Gaussian

$$\boldsymbol{y} \overset{a}{\sim} N(\boldsymbol{K}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{5.20}$$

where the covariance is given by $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{K}\boldsymbol{\beta})$. In practice, we estimate the covariance using $\hat{\boldsymbol{\Sigma}} = \mathrm{diag}(\boldsymbol{y}_+)$, where $y_{+,i} = \max(1, y_i)$ in order to guarantee that $\hat{\boldsymbol{\Sigma}}$ is positive definite.

The log-likelihood function can hence be approximated using

$$l(\boldsymbol{\beta};\boldsymbol{y}) = \log p(\boldsymbol{y}|\boldsymbol{\beta}) \approx -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\beta})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\beta}) + C, \quad \boldsymbol{\beta} \in \mathbb{R}_+^p, \tag{5.21}$$

where C is a constant that does not depend on $\boldsymbol{\beta}$. Plugging this into Equation (5.19) yields the Gaussian approximated point estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{G}_+} = \underset{\boldsymbol{\beta}\in\mathbb{R}_+^p}{\arg\max}\, \left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\beta})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\beta}) - \delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}\right) \tag{5.22}$$

$$= \underset{\boldsymbol{\beta}\in\mathbb{R}_+^p}{\arg\min}\, \left((\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\beta})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y} - \boldsymbol{K}\boldsymbol{\beta}) + 2\delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}\right). \tag{5.23}$$

Let $\hat{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$ and $\boldsymbol{\Omega}_{\mathrm{A}} = \boldsymbol{N}^{\mathrm{T}}\boldsymbol{N}$ be the Cholesky factorizations of $\hat{\boldsymbol{\Sigma}}^{-1}$ and $\boldsymbol{\Omega}_{\mathrm{A}}$. The estimator can then be rewritten as

$$\hat{\boldsymbol{\beta}}_{\mathrm{G}_+} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p_+}{\arg\min} \left(\|\boldsymbol{MK\beta} - \boldsymbol{My}\|_2^2 + \|\sqrt{2\delta}\boldsymbol{N\beta}\|_2^2 \right) \tag{5.24}$$

$$= \underset{\boldsymbol{\beta}\in\mathbb{R}^p_+}{\arg\min} \left\| \begin{bmatrix} \boldsymbol{MK} \\ \sqrt{2\delta}\boldsymbol{N} \end{bmatrix} \boldsymbol{\beta} - \begin{bmatrix} \boldsymbol{My} \\ \boldsymbol{0} \end{bmatrix} \right\|_2^2 . \tag{5.25}$$

This is a positivity-constrained least-squares problem. The problem cannot be solved in closed form, but there exist efficient algorithms for obtaining the solution iteratively in a finite number of steps. In this work, we compute the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ using the MATLAB function `lsqnonneg`, which uses the NNLS algorithm described in Lawson and Hanson (1995, p. 161).

5.2.4 Unconstrained Gaussian approximation

A further simplification can be obtained by dropping the positivity constraint in Equation (5.23). The resulting estimator is

$$\hat{\boldsymbol{\beta}}_{\mathrm{G}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \left((\boldsymbol{y} - \boldsymbol{K\beta})^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{y} - \boldsymbol{K\beta}) + 2\delta \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\Omega}_{\mathrm{A}} \boldsymbol{\beta} \right). \tag{5.26}$$

We recognize that this form corresponds to the well-known techniques of Tikhonov regularization (Tikhonov, 1963; Phillips, 1962) and ridge regression (Hoerl and Kennard, 1970). The optimization problem can again be rewritten as

$$\hat{\boldsymbol{\beta}}_{\mathrm{G}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \left\| \begin{bmatrix} \boldsymbol{MK} \\ \sqrt{2\delta}\boldsymbol{N} \end{bmatrix} \boldsymbol{\beta} - \begin{bmatrix} \boldsymbol{My} \\ \boldsymbol{0} \end{bmatrix} \right\|_2^2 , \tag{5.27}$$

which is a standard unconstrained least-squares problem. Since the matrix $\begin{bmatrix} \boldsymbol{MK} \\ \sqrt{2\delta}\boldsymbol{N} \end{bmatrix}$ has full column rank, the solution is given by

$$\hat{\boldsymbol{\beta}}_{\mathrm{G}} = \left(\begin{bmatrix} \boldsymbol{MK} \\ \sqrt{2\delta}\boldsymbol{N} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{MK} \\ \sqrt{2\delta}\boldsymbol{N} \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{MK} \\ \sqrt{2\delta}\boldsymbol{N} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{My} \\ \boldsymbol{0} \end{bmatrix} \tag{5.28}$$

$$= \left(\boldsymbol{K}^{\mathrm{T}} \boldsymbol{M}^{\mathrm{T}} \boldsymbol{MK} + 2\delta \boldsymbol{N}^{\mathrm{T}} \boldsymbol{N} \right)^{-1} \boldsymbol{K}^{\mathrm{T}} \boldsymbol{M}^{\mathrm{T}} \boldsymbol{My} \tag{5.29}$$

$$= \left(\boldsymbol{K}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{K} + 2\delta \boldsymbol{\Omega}_{\mathrm{A}} \right)^{-1} \boldsymbol{K}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{y}. \tag{5.30}$$

The matrix operations needed to compute $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ are many orders of magnitude faster than running the MCMC to obtain $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$. If we ignore the data-dependence of $\hat{\boldsymbol{\Sigma}}$ (and possibly also the data-dependence of an estimator of δ), the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ has the added benefit of being a linear function of $\boldsymbol{y}$, while the other estimators considered here are nonlinear in $\boldsymbol{y}$.

5.3 Choice of the regularization strength

The regularization strength δ controls the bias-variance trade-off of the unfolded estimator and has a major impact on the solution. The appropriate choice of this parameter is of central importance in order to obtain reasonable solutions. In this section, we introduce various data-driven ways of choosing δ motivated by both Bayesian and frequentist ideas. The focus will be on choices that primarily aim to provide good point estimation performance.

5.3.1 Empirical Bayes

We first consider empirical Bayes (EB) estimation of the hyperparameter δ. Carlin and Louis (2009, Chapter 5) provides an extensive introduction to empirical Bayes methods, while for example Wood (2011) and Ruppert et al. (2003, Section 5.2) study empirical Bayes selection of the regularization strength in semi- and non-parametric regression models. This approach can be understood as a hybrid between the frequentist and Bayesian paradigms. The main idea is to use the marginal likelihood $L(\delta; \boldsymbol{y}) = p(\boldsymbol{y}|\delta)$ to perform maximum likelihood estimation of δ. In other words, we seek to estimate δ using the *marginal maximum likelihood estimator* (MMLE)

$$\hat{\delta}_{\text{MMLE}} = \operatorname*{arg\,max}_{\delta>0} L(\delta; \boldsymbol{y}) = \operatorname*{arg\,max}_{\delta>0} p(\boldsymbol{y}|\delta) = \operatorname*{arg\,max}_{\delta>0} \int_{\mathbb{R}_+^p} p(\boldsymbol{y}|\boldsymbol{\beta}) p(\boldsymbol{\beta}|\delta) \, \mathrm{d}\boldsymbol{\beta}. \tag{5.31}$$

The estimated regularization strength $\hat{\delta}_{\text{MMLE}}$ is then plugged into any of the point estimators described in Section 5.2 to obtain the unfolded solution. Notice that, by the bijective equivariance property of maximum likelihood estimators (see, e.g., Panaretos (2016, Proposition 3.17)), the empirical Bayes approach is invariant to bijective transformations of the hyperparameter δ.

The main challenge in this construction is the computation of $\hat{\delta}_{\text{MMLE}}$. Since the marginal likelihood is given by an intractable integral, one cannot directly evaluate, let alone maximize, the objective function in Equation (5.31). One could consider Monte Carlo integration in the form of

$$p(\boldsymbol{y}|\delta) \approx \frac{1}{S} \sum_{s=1}^{S} p(\boldsymbol{y}|\boldsymbol{\beta}^{(s)}), \quad \boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(S)} \overset{\text{i.i.d.}}{\sim} p(\boldsymbol{\beta}|\delta), \tag{5.32}$$

but this does not work well in practice since most of the $\boldsymbol{\beta}^{(s)}$'s fall on regions of the parameter space where the likelihood $p(\boldsymbol{y}|\boldsymbol{\beta}^{(s)})$ is numerically zero.

These issues can be circumvented using the *expectation-maximization* (EM) *algorithm* (Dempster et al., 1977; McLachlan and Krishnan, 2008) to find the MMLE. The EM algorithm is a widely applicable technique for computing maximum likelihood estimates. Let $L(\boldsymbol{\theta}; \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta})$ be a likelihood function that we seek to maximize and assume that for algorithmic or numerical reasons this is difficult to accomplish. Let $\boldsymbol{z}$ be some unobserved latent variables chosen in such a way that the likelihood function $L(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta})$ can be easily maximized. Using standard EM terminology, we call $\boldsymbol{y}$ the *incomplete data* and $(\boldsymbol{y}, \boldsymbol{z})$ the *complete data*. The algorithm then alternates between the *expectation step* (E-step) and the *maximization step* (M-step). In the E-step, one computes the conditional expectation of the complete-data log-likelihood given the observations $\boldsymbol{y}$ and the current iterate $\boldsymbol{\theta}^{(t)}$,

$$Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) = \mathrm{E}\big(l(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{z})\big|\boldsymbol{y},\boldsymbol{\theta}^{(t)}\big) = \mathrm{E}\big(\log p(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\theta})\big|\boldsymbol{y},\boldsymbol{\theta}^{(t)}\big), \tag{5.33}$$

where $l(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) = \log L(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) = \log p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta})$. In the subsequent M-step, one then obtains the next iterate by maximizing this conditional expectation with respect to the parameter $\boldsymbol{\theta}$,

$$\boldsymbol{\theta}^{(t+1)} = \operatorname*{arg\,max}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}). \tag{5.34}$$

The resulting iteration is guaranteed to increase the incomplete-data likelihood, that is, $L(\boldsymbol{\theta}^{(t+1)}; \boldsymbol{y}) \geq L(\boldsymbol{\theta}^{(t)}; \boldsymbol{y})$ for all $t = 0, 1, 2, \ldots$ (Dempster et al., 1977, Theorem 1). Under further regularity conditions, it can be shown that the sequence of iterates $\boldsymbol{\theta}^{(t)}$, $t = 0, 1, 2, \ldots$, converges to a stationary point of the incomplete-data likelihood (Wu, 1983).

In our case, the incomplete-data likelihood is given by the marginal likelihood $L(\delta; \boldsymbol{y}) = p(\boldsymbol{y}|\delta)$ and we take $(\boldsymbol{y}, \boldsymbol{\beta})$ to be the complete data. The complete-data log-likelihood is then given by

$$l(\delta;\boldsymbol{y},\boldsymbol{\beta}) = \log p(\boldsymbol{y},\boldsymbol{\beta}|\delta) = \log p(\boldsymbol{y}|\boldsymbol{\beta}) + \log p(\boldsymbol{\beta}|\delta). \tag{5.35}$$

On the E-step, we need to evaluate the conditional expectation

$$\begin{aligned} Q(\delta;\delta^{(t)}) = \mathrm{E}\big(l(\delta;\boldsymbol{y},\boldsymbol{\beta})\big|\boldsymbol{y},\delta^{(t)}\big) &= \mathrm{E}\big(\log p(\boldsymbol{y},\boldsymbol{\beta}|\delta)\big|\boldsymbol{y},\delta^{(t)}\big) && (5.36)\\ &= \mathrm{E}\big(\log p(\boldsymbol{\beta}|\delta)\big|\boldsymbol{y},\delta^{(t)}\big) + \mathrm{const}, && (5.37) \end{aligned}$$

where the constant does not depend on δ. Notice that here the expectation is taken over the posterior $p(\boldsymbol{\beta}|\boldsymbol{y}, \delta^{(t)})$. On the M-step, we then maximize this expectation with respect to δ to obtain the next iterate,

$$\delta^{(t+1)} = \operatorname*{arg\,max}_{\delta>0} Q(\delta;\delta^{(t)}) = \operatorname*{arg\,max}_{\delta>0} \mathrm{E}\big(\log p(\boldsymbol{\beta}|\delta)\big|\boldsymbol{y},\delta^{(t)}\big). \tag{5.38}$$

Here the E-step still involves an intractable integral

$$\mathrm{E}\big(\log p(\boldsymbol{\beta}|\delta)\big|\boldsymbol{y},\delta^{(t)}\big) = \int_{\mathbb{R}_+^p} p(\boldsymbol{\beta}|\boldsymbol{y},\delta^{(t)}) \log p(\boldsymbol{\beta}|\delta)\,\mathrm{d}\boldsymbol{\beta}, \tag{5.39}$$

which this time around can be computed using Monte Carlo integration. We simply need to sample $\{\boldsymbol{\beta}^{(s)}\}_{s=1}^{S}$ from the posterior $p(\boldsymbol{\beta}|\boldsymbol{y}, \delta^{(t)})$ using the single-component Metropolis–Hastings algorithm described in Appendix A.1 and the value of the integral is then approximately given by the empirical mean

$$\mathrm{E}\big(\log p(\boldsymbol{\beta}|\delta)\big|\boldsymbol{y},\delta^{(t)}\big) \approx \frac{1}{S}\sum_{s=1}^{S}\log p(\boldsymbol{\beta}^{(s)}|\delta), \quad \boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(S)} \sim p(\boldsymbol{\beta}|\boldsymbol{y},\delta^{(t)}). \tag{5.40}$$

Monte Carlo integration is better behaved here than in Equation (5.32) since the sample is from the posterior instead of the prior and hence most of the $\boldsymbol{\beta}^{(s)}$'s lie within the bulk of prior density $p(\boldsymbol{\beta}|\delta)$. The logarithm also helps to stabilize the computations. Since we have replaced the E-step by a Monte Carlo approximation, the resulting iteration is typically called a *Monte Carlo expectation-maximization* (MCEM) *algorithm* (Wei and Tanner, 1990; Casella, 2001). A similar algorithm has been used in tomographic image reconstruction by Geman and McClure (1985, 1987) and Saquib et al. (1998).

To summarize, the MCEM iteration for finding the MMLE is given by:

E-step: Sample $\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(S)}$ from the posterior $p(\boldsymbol{\beta}|\boldsymbol{y},\delta^{(t)})$ and compute

$$\widetilde{Q}(\delta;\delta^{(t)}) = \frac{1}{S}\sum_{s=1}^{S}\log p(\boldsymbol{\beta}^{(s)}|\delta). \tag{5.41}$$

M-step: Set $\delta^{(t+1)} = \arg\max_{\delta>0}\widetilde{Q}(\delta;\delta^{(t)})$.

This algorithm admits a rather intuitive interpretation: In the E-step, the posterior sample $\{\boldsymbol{\beta}^{(s)}\}_{s=1}^{S}$ summarizes our current best understanding of $\boldsymbol{\beta}$ given the hyperparameter value $\delta^{(t)}$. This sample is then plugged into the prior and in the M-step the hyperparameter is updated so that the prior matches the posterior sample as well as possible.

For the Gaussian smoothness prior given by Equation (5.15), the M-step is available in closed form. The prior density is given by

$$p(\boldsymbol{\beta}|\delta) = C(\delta)\exp(-\delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}), \quad \boldsymbol{\beta}\in\mathbb{R}_{+}^{p}, \tag{5.42}$$

where $C(\delta) = \delta^{p/2}\big/\int_{\mathbb{R}_{+}^{p}}\exp(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta}$ is a normalization constant. Hence the log-prior is

$$\log p(\boldsymbol{\beta}|\delta) = \frac{p}{2}\log\delta - \delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta} + \mathrm{const}, \tag{5.43}$$

where the constant does not depend on δ. Substituting this into Equation (5.41), we find that the maximizer on the M-step is given by

$$\delta^{(t+1)} = \frac{1}{\frac{2}{pS}\sum_{s=1}^{S}(\boldsymbol{\beta}^{(s)})^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}^{(s)}}. \tag{5.44}$$

Algorithm 1 summarizes the resulting MCEM iteration for finding the marginal maximum likelihood estimator $\hat{\delta}_{\mathrm{MMLE}}$. To facilitate the convergence of the MCMC sampler, we start the Markov chain from the posterior mean of the previous iteration. Devising a good stopping rule for the MCEM iteration is non-trivial because of the Monte Carlo error associated with

Algorithm 1 MCEM algorithm for finding the MMLE

```
Input:
  $\boldsymbol{y}$ — Smeared data
  $\delta^{(0)} > 0$ — Initial guess
  $N_{\mathrm{EM}}$ — Number of MCEM iterations
  $S$ — Size of the MCMC sample
  $\boldsymbol{\beta}_{\mathrm{init}}$ — Starting point for the MCMC sampler
Output:
  $\hat{\delta}_{\mathrm{MMLE}}$ — MMLE of the hyperparameter $\delta$

  Set $\bar{\boldsymbol{\beta}} = \boldsymbol{\beta}_{\mathrm{init}}$;
  for $t = 0$ to $N_{\mathrm{EM}} - 1$ do
    Sample $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(S)} \sim p(\boldsymbol{\beta}|\boldsymbol{y}, \delta^{(t)})$ starting from $\bar{\boldsymbol{\beta}}$ using the single-component Metropolis–Hastings sampler described in Appendix A.1;
    Set $\delta^{(t+1)} = 1/\left(\frac{2}{pS}\sum_{s=1}^{S}(\boldsymbol{\beta}^{(s)})^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}^{(s)}\right)$;
    Compute $\bar{\boldsymbol{\beta}} = \frac{1}{S}\sum_{s=1}^{S}\boldsymbol{\beta}^{(s)}$;
  end for
  return $\hat{\delta}_{\mathrm{MMLE}} = \delta^{(N_{\mathrm{EM}})}$;
```

the iterates $\delta^{(t)}$ (Booth and Hobert, 1999). Because of this, we run the algorithm for a fixed number of iterations N_{EM} and verify the convergence graphically.

5.3.2 Hierarchical Bayes

The fully Bayesian way of handling the unknown regularization strength δ is to use a Bayesian hierarchical model. This hierarchical Bayes (HB) approach places a hyperprior $p(\delta)$ on the hyperparameter δ and then performs Bayesian inference for both $\boldsymbol{\beta}$ and δ using the joint posterior

$$p(\boldsymbol{\beta}, \delta|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\beta})\,p(\boldsymbol{\beta}|\delta)\,p(\delta)}{p(\boldsymbol{y})}. \tag{5.45}$$

As such, this approach does not simply consider a single value of δ but instead a distribution of probable values of δ. The marginal posterior of the spline coefficients $\boldsymbol{\beta}$ is obtained by integrating out the hyperparameter

$$p(\boldsymbol{\beta}|\boldsymbol{y}) = \int_{\mathbb{R}_+} p(\boldsymbol{\beta}, \delta|\boldsymbol{y})\,\mathrm{d}\delta. \tag{5.46}$$

We then use the mean of the marginal posterior $\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{y})$ as a point estimator of $\boldsymbol{\beta}$. This is again in practice computed as the empirical mean of an MCMC sample from the posterior.

Notice that the marginal posterior (5.46) corresponds to simply using $p(\boldsymbol{\beta}) = \int_{\mathbb{R}_+} p(\boldsymbol{\beta}|\delta)\,p(\delta)\,\mathrm{d}\delta$ as the prior in the original Bayes' rule (5.3). So hierarchical Bayes can be understood as weighting the contributions of the smoothness priors $p(\boldsymbol{\beta}|\delta)$ for different δ with the weights given by the hyperprior $p(\delta)$. When seen from this perspective, the choice of δ is entirely driven by the hyperprior $p(\delta)$ instead of the observations $\boldsymbol{y}$. The main difficulty in hierarchical Bayes indeed concerns the selection of the hyperprior and the approach is known to be sensitive to this non-trivial choice (Gelman, 2006), an observation confirmed by our simulations in Section 5.4.2.

In this work, we consider hyperpriors of the form

$$p(\delta) \propto \mathbf{1}_{[L,\infty)}(\delta)\,\delta^{a-1}e^{-b\delta}, \tag{5.47}$$

where a, b and L are parameters chosen in such a way that the density can be normalized. This includes as special cases the Pareto$(-a, L)$ distribution ($a < 0$, $b = 0$ and $L > 0$) and the Gamma(a, b) distribution ($a > 0$, $b > 0$ and $L = 0$). The hyperprior family (5.47) is conditionally conjugate. Indeed, the full posterior conditional for δ is given by

$$p(\delta|\boldsymbol{\beta},\boldsymbol{y}) = p(\delta|\boldsymbol{\beta}) \propto p(\boldsymbol{\beta}|\delta)\,p(\delta) \tag{5.48}$$

$$\propto \mathbf{1}_{[L,\infty)}(\delta)\,\delta^{p/2+a-1}\exp\left(-(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta}+b)\delta\right), \tag{5.49}$$

which has the same form as Equation (5.47). When $p/2 + a > 0$ and $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta} + b > 0$, $p(\delta|\boldsymbol{\beta},\boldsymbol{y})$ is the Gamma$(p/2 + a, \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}_{\mathrm{A}}\boldsymbol{\beta} + b)$ distribution truncated to the interval $[L,\infty)$. This makes it straightforward to incorporate sampling from the joint posterior (5.45) into the single-component Metropolis–Hastings sampler. We loop over all the unknowns and for each β_k use the sampler described in Appendix A.1 given the current value of δ, while for δ we simply sample from the truncated Gamma distribution (5.49) given the current value of $\boldsymbol{\beta}$.

5.3.3 Weighted cross-validation

The most common frequentist technique for choosing the regularization strength δ is to use cross-validation (CV) (Stone, 1974). The approach is based on choosing δ to minimize the prediction error

$$\mathrm{E}\left((\boldsymbol{y}^* - \hat{\boldsymbol{\mu}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}^* - \hat{\boldsymbol{\mu}})\right) = \mathrm{E}\left(\sum_{i=1}^{p}\frac{(y_i^* - \hat{\mu}_i)^2}{\mathrm{var}(y_i^*)}\right), \tag{5.50}$$

where $\boldsymbol{y}^*$ is a new smeared observation, $\hat{\boldsymbol{\mu}} = \boldsymbol{K}\hat{\boldsymbol{\beta}}$ is an estimate of the smeared mean histogram $\boldsymbol{\mu} = \boldsymbol{K}\boldsymbol{\beta}$ and $\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{y}^*) = \mathrm{cov}(\boldsymbol{y})$ is the diagonal covariance matrix of the smeared observations. Here $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\delta)$ may be any of the unfolded point estimators discussed in Section 5.2. The covariance is included in the prediction error to account for the heteroscedasticity of the problem.

The prediction error (5.50) can be estimated using *weighted leave-one-out cross-validation*

(Green and Silverman, 1994, Section 3.5.3)

$$\mathrm{CV}(\delta) = \sum_{i=1}^{n} \frac{\left(y_i - \hat{\mu}_i^{-i}\right)^2}{y_{+,i}}, \tag{5.51}$$

where $\hat{\mu}_i^{-i}$ is an estimate of the ith smeared bin obtained without using the smeared observation y_i in that bin and the denominator $y_{+,i} = \max(1, y_i)$ is an estimate of the variance. More specifically, $\hat{\mu}_i^{-i} = \boldsymbol{k}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}^{-i}$, where $\boldsymbol{k}_i^{\mathrm{T}}$ is the ith row of $\boldsymbol{K}$ and $\hat{\boldsymbol{\beta}}^{-i}$ is the unfolded solution obtained using $\boldsymbol{K}^{-i} = [\boldsymbol{k}_1, \ldots, \boldsymbol{k}_{i-1}, \boldsymbol{k}_{i+1}, \ldots, \boldsymbol{k}_n]^{\mathrm{T}}$ as the smearing matrix, $\boldsymbol{y}^{-i} = [y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n]^{\mathrm{T}}$ as the smeared data and δ as the regularization strength. The CV estimate of δ is then the minimizer of the CV criterion (5.51), $\hat{\delta}_{\mathrm{CV}} = \arg\min_{\delta>0} \mathrm{CV}(\delta)$.

Notice that cross-validation chooses δ to optimize the prediction error in the smeared space and not the estimation error in the unfolded space. As such, it will not necessarily guarantee optimal unfolding performance. Wood (2011) suggests that (generalized) cross-validation may be more variable and more susceptible to overfitting than empirical Bayes. This happens because the optimum of the CV objective function tends to be less pronounced than that of the EB objective. Reiss and Ogden (2009) provide a theoretical comparison of the two methods and conclude that empirical Bayes tends to be more stable and less prone to issues with local optima. Our simulation results in Section 5.4.3 support these findings.

Evaluation of the cross-validation criterion (5.51) requires computing n point estimates for each value of δ. Hence, minimization of this function is in practice too slow for the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ and the MAP/PMLE estimator $\hat{\boldsymbol{\beta}}_{\mathrm{MAP}}/\hat{\boldsymbol{\beta}}_{\mathrm{PMLE}}$. In our simulations in Section 5.4.3, we use the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ to compute the CV criterion. An even faster solution would be provided by the unconstrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$, since, for estimators of the form (5.30), one can show (Green and Silverman, 1994, Section 3.5.3) that the CV criterion is given by

$$\mathrm{CV}(\delta) = \sum_{i=1}^{n} \frac{1}{y_{+,i}} \left(\frac{y_i - \hat{\mu}_i}{1 - H_{i,i}} \right)^2, \tag{5.52}$$

where $\boldsymbol{H} = \boldsymbol{K}\left(\boldsymbol{K}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{K} + 2\delta \boldsymbol{\Omega}_{\mathrm{A}}\right)^{-1} \boldsymbol{K}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1}$. Notice that in this case only one point estimate needs to be computed for each value of δ.

5.3.4 Other methods

In this work, we focus on empirical Bayes, hierarchical Bayes and cross-validation techniques for choosing δ. Nevertheless, several other methods, some of which we mention here for completeness, have also been proposed in the literature. The `TUnfold` software (Schmitt, 2012) implements the L-curve technique of Hansen (1992) as well as a heuristic that aims to minimize the correlations in the unfolded space. Volobouev (2015) proposes to use the Akaike information criterion (Akaike, 1973, 1974) with a correction for the finite sample size.

Other potential methods include the Morozov discrepancy principle (Morozov, 1966) and goodness-of-fit testing in the smeared space (Veklerov and Llacer, 1987).

5.4 Simulation study

In this section, we perform simulation studies comparing empirical Bayes, hierarchical Bayes and cross-validation in the two peaks on a uniform background test setup described in Section 3.4.1.

5.4.1 Convergence studies for empirical Bayes

We first verify the convergence of the MCEM iteration. In all the experiments described in this thesis, we start the MCEM iteration from $\delta^{(0)} = 1 \cdot 10^{-5}$ and run it for $N_{\mathrm{EM}} = 30$ iterations. We use the single-component Metropolis–Hastings sampler to generate $S = 1\,000$ post-burn-in observations from the posterior. As the starting point of the sampler, we use the positivity-constrained least-squares spline fit to the smeared data without unfolding. That is, $\boldsymbol{\beta}_{\mathrm{init}} = \min_{\boldsymbol{\beta} \geq \mathbf{0}} \|\tilde{\boldsymbol{K}}\boldsymbol{\beta} - \boldsymbol{y}\|_2^2$, where $\tilde{\boldsymbol{K}}$ is given by Equation (5.2) with $k(t,s) = \delta_0(t-s)$ with δ_0 denoting the Dirac delta function. Unless otherwise indicated, the boundary hyperparameters are set to $\gamma_{\mathrm{L}} = \gamma_{\mathrm{R}} = 5$.

Figure 5.1(a) illustrates the convergence of the MCEM iteration for the different sample sizes. We find that in each case 30 iterations is sufficient for the convergence of the algorithm and that the convergence appears to be the faster the larger the sample size. The Monte Carlo variation of the estimates is also small. For the small, medium and large sample sizes, the algorithm converged to the hyperparameter estimates $\hat{\delta}_{\mathrm{MMLE}} = 2.2 \cdot 10^{-4}$, $\hat{\delta}_{\mathrm{MMLE}} = 1.2 \cdot 10^{-6}$ and $\hat{\delta}_{\mathrm{MMLE}} = 4.1 \cdot 10^{-8}$, respectively.

Figure 5.2 shows a realization of the unfolded intensities $\hat{f}$ when the spline coefficients are estimated using the mean $\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{y}, \hat{\delta}_{\mathrm{MMLE}})$ of the empirical Bayes posterior $p(\boldsymbol{\beta}|\boldsymbol{y}, \hat{\delta}_{\mathrm{MMLE}})$, obtained by plugging the estimated hyperparameter $\hat{\delta}_{\mathrm{MMLE}}$ into Equation (5.3). We find that in each case the unfolded point estimator captures the two-peak shape of the true intensity. Unsurprisingly, the quality of the point estimate improves with increasing sample size. We also note that the estimates appear to be biased downwards near the larger peak. A central theme in Chapter 6 will be the development of confidence bands around these solutions and accommodating this bias will play a key role in our discussion.

To further study how empirical Bayes point estimation behaves as a function of the sample size, we repeat the experiment on a logarithmic grid of sample sizes ranging from $\lambda_{\mathrm{tot}} = 5\,000$ up to $\lambda_{\mathrm{tot}} = 100\,000$. For each sample size, we unfold 200 independent realizations of the smeared data $\boldsymbol{y}$ and estimate the mean integrated squared error (MISE) of $\hat{f}$ as the sample

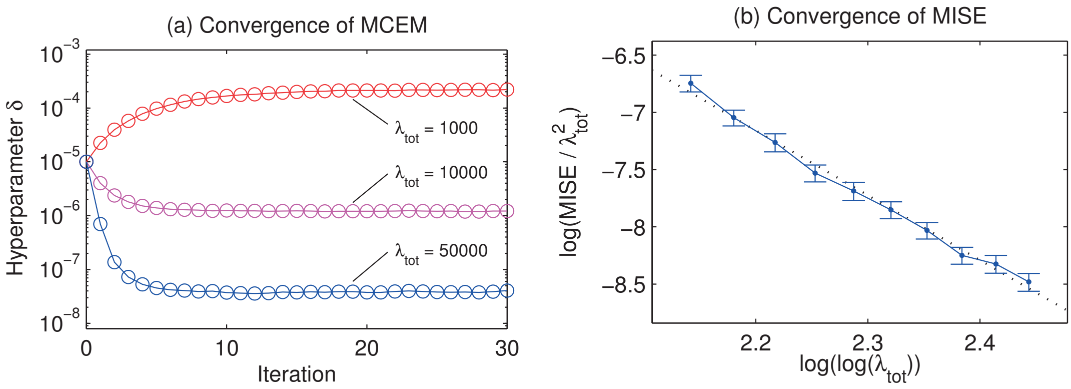


Figure 5.1: Convergence studies for empirical Bayes unfolding in the two peaks on a uniform background test setup. Figure (a) illustrates the convergence of the Monte Carlo expectation-maximization (MCEM) iteration and shows that the algorithm converges faster for larger sample sizes. Figure (b) shows the convergence of the mean integrated squared error (MISE) as the expected sample size λ_{tot} grows. The error bars are 95 % confidence intervals and the dotted straight line is a least-squares fit to the convergence curve.

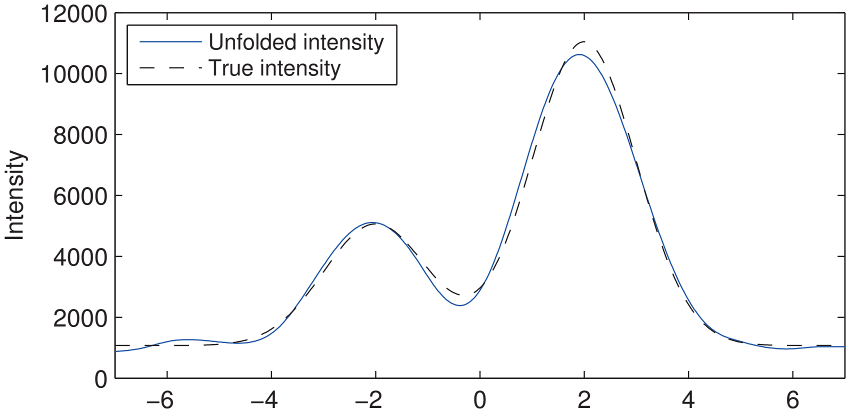
mean of the integrated squared errors

$$\text{ISE} = \int_E \big(\hat{f}(s) - f(s)\big)^2 \, \mathrm{d}s. \tag{5.53}$$

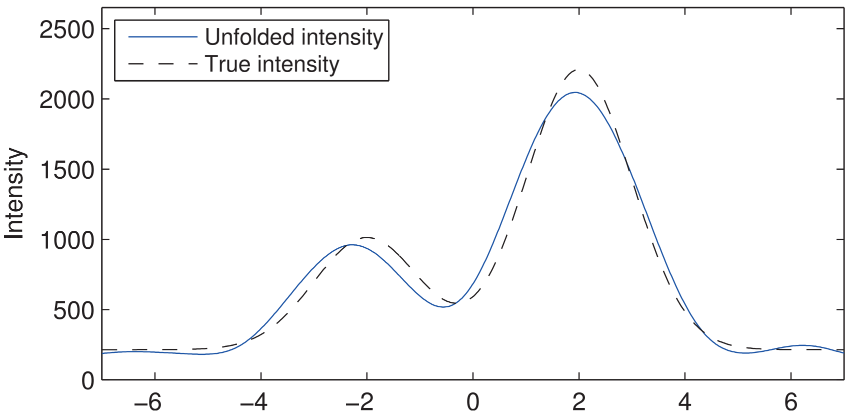
As $\lambda_{\text{tot}} \to \infty$, one would expect the MISE to diverge, but $\text{MISE}/\lambda_{\text{tot}}^2$ should converge to zero, and this is indeed what we observe in Figure 5.1(b).

In the classical problem of deconvolving a density function smeared by Gaussian noise, the optimal convergence rate of the MISE is of the order $(\log n)^{-k}$ (Meister, 2009), where n is the number of i.i.d. smeared observations and $k > 0$ depends on the smoothness of the true density. Our setup differs slightly from the classical one in the sense that we observe a realization of a smeared Poisson point process and try to estimate the intensity function of the corresponding true process. We also perform the estimation on a compact interval, which introduces boundary effects near the end points of the interval. Nevertheless, one could conjecture that $\text{MISE}/\lambda_{\text{tot}}^2$ converges at the rate $(\log \lambda_{\text{tot}})^{-k}$, in which case the values in Figure 5.1(b) should fall on a straight line with slope $-k$. This indeed appears to approximately be the case: the $\text{MISE}/\lambda_{\text{tot}}^2$ values seem to follow fairly well the line with slope $-k = -5.67$, which is also shown in the figure. However, in a more careful inspection, it appears that the convergence curve may have a slightly convex shape. There are two potential explanations for this: either the beginning of the curve is not yet in the asymptotic regime or the convergence rate is slightly slower than $(\log \lambda_{\text{tot}})^{-k}$. If the rate is indeed slower than expected, this might be due to the fact that we have kept the dimension of the spline basis fixed when increasing λ_{tot}. As a result, the discretization error from the spline fit should eventually slow down the convergence rate.

(a) EB posterior mean, λ_{tot} = 50000



(b) EB posterior mean, λ_{tot} = 10000



(c) EB posterior mean, λ_{tot} = 1000

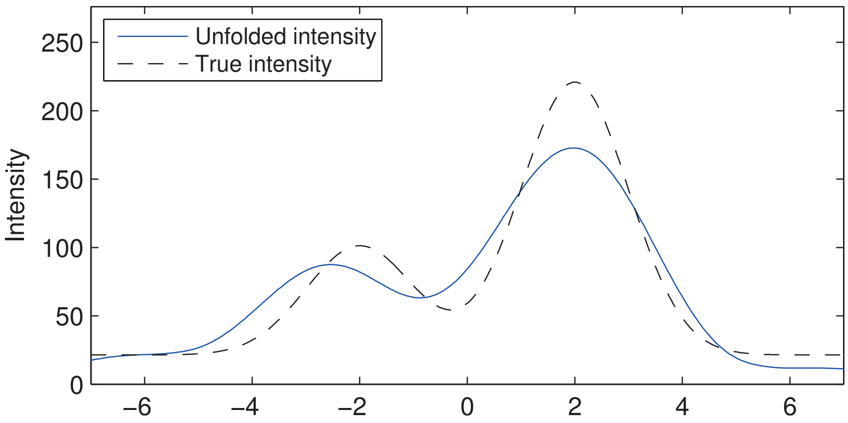


Figure 5.2: Unfolded intensities $\hat{f}$ in the two peaks on a uniform background test setup, when the regularization strength is estimated using the marginal maximum likelihood estimator $\hat{\delta}_{\mathrm{MMLE}}$ and the spline coefficients are estimated using the mean of the resulting empirical Bayes posterior $\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{y},\hat{\delta}_{\mathrm{MMLE}})$.

5.4.2 Comparison of empirical Bayes and hierarchical Bayes

In this section, we compare the performance of empirical Bayes (EB) and hierarchical Bayes (HB) in point estimation of f. For hierarchical Bayes, the point estimate is the mean $\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{y})$ of the marginal posterior (5.46), and, for empirical Bayes, we use the mean of the empirical Bayes posterior $\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{y},\hat{\delta}_{\mathrm{MMLE}})$. The performance of the point estimates is quantified using the integrated squared error (ISE) defined by Equation (5.53).

We consider hierarchical Bayes with the following uninformative hyperpriors:

(a) $\mathrm{Pareto}(1,10^{-10})$;

(b) $\mathrm{Pareto}(1/2,10^{-10})$;

(c) $\mathrm{Gamma}(0.001,0.001)$;

(d) $\mathrm{Gamma}(1,0.001)$.

All of these are nearly flat for some transformation of δ. Indeed, hyperprior (a) is nearly uniform for $1/\delta$, (b) for $1/\sqrt{\delta}$, (c) for $\log(\delta)$ and (d) for δ. Hyperprior (c) is often used in the literature; see, for example, Browne and Draper (2006), Ruppert et al. (2003, Section 16.3) and Young and Smith (2005, Section 3.8). Gelman (2006) however argues that (b) should provide better results. Hyperprior (a) is studied, for example, by Browne and Draper (2006). One could also naïvely expect (d) to be a sensible choice, as it is nearly flat for the untransformed hyperparameter δ itself. Notice that all of these hyperpriors are proper probability densities guaranteeing that the joint posterior (5.45) is also proper.

We study the performance of the methods for 1 000 repeated observations of the two peaks on a uniform background test setup (Section 3.4.1). The hierarchical Bayes MCMC sampler was started from $(\boldsymbol{\beta}_{\mathrm{init}},\delta^{(0)})$. The remaining parameters were set to the same values as in Section 5.4.1.

Figure 5.3 shows boxplots of the pairwise relative ISE differences $(\mathrm{ISE}_{\mathrm{HB},i}-\mathrm{ISE}_{\mathrm{EB},i})/\mathrm{ISE}_{\mathrm{EB},i}$ between the two methods for the different hyperpriors. Positive values in the figure indicate that HB incurred a larger error than EB. When tested using the Wilcoxon signed-rank test, the differences between the methods for sample sizes $\lambda_{\mathrm{tot}} = 1\,000$ and $\lambda_{\mathrm{tot}} = 10\,000$ are statistically significant at any reasonable significance level, except for hyperprior (c) with $\lambda_{\mathrm{tot}} = 10\,000$, which is only significant at the 2 % level (two-sided p-value 0.011). For sample size $\lambda_{\mathrm{tot}} = 50\,000$, the differences are statistically significant at the 5 % level for hyperpriors (a)–(c), but not for hyperprior (d), where the two-sided p-value is 0.50.

Figure 5.3 enables us to make a number of observations. Firstly, there are fairly large differences between the different hyperpriors, especially when there is only a limited amount of data available. For $\lambda_{\mathrm{tot}} = 1\,000$, the median performance of HB ranges from 17 % better to 30 % worse than EB, depending on the hyperprior used. The performance is generally better for hyperpriors that favor small values of δ. In particular, hyperprior (d) tends to regularize too

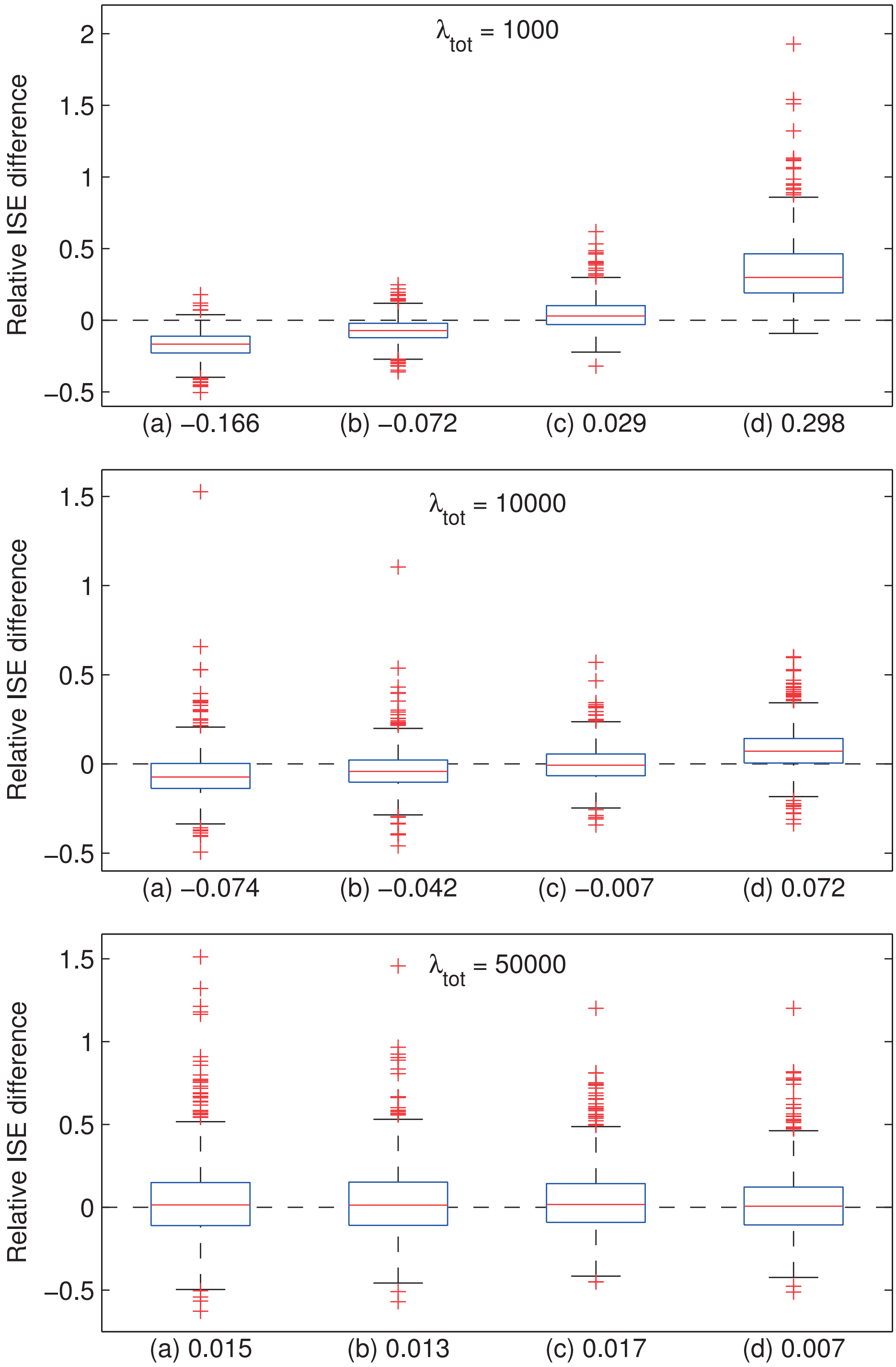


Figure 5.3: Relative pairwise integrated squared error differences between empirical Bayes and hierarchical Bayes with the hyperpriors (a) Pareto$(1, 10^{-10})$, (b) Pareto$(1/2, 10^{-10})$, (c) Gamma$(0.001, 0.001)$ and (d) Gamma$(1, 0.001)$. The numbers below the boxplots show the median relative differences. Positive values indicate that the error of hierarchical Bayes was larger than that of empirical Bayes.

strongly. Secondly, as the sample size grows, the performance of the methods becomes increasingly similar. For $\lambda_{\mathrm{tot}} = 50\,000$, there is little difference between the different hyperpriors and between EB and HB. Thirdly, EB is in each case competitive with HB and the superiority of the two methods depends on which HB hyperprior is used. The performance of EB is similar to hyperprior (c) which seems to be the most common choice in HB literature (see the beginning of this section). Most importantly, EB achieves this performance without making any extra distributional assumptions about δ.

We conclude from this that, especially for small sample sizes, the performance of HB is indeed sensitive to the choice of the hyperprior, while EB achieves comparable performance without the need to make this choice. Furthermore, as discussed in Section 5.3.2, it is difficult to interpret HB as making a data-driven choice of δ. Instead, the choice is fully driven by the hyperprior $p(\delta)$ and some hyperpriors are better suited for each particular situation that others. For example, there are no guarantees that hyperprior (a) would always lead to best performance. On the other hand, the choice of δ in EB is fully data-driven. For these reasons, we prefer to primarily focus on EB instead of HB techniques in the reminder of this work.

5.4.3 Comparison of empirical Bayes and cross-validation

In this section, we compare empirical Bayes and cross-validation by unfolding 1 000 repeated observations of the two peaks on a uniform background test setup (Section 3.4.1). We estimate the spline coefficients using the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ (Section 5.2.3) with the regularization strength δ chosen using either weighted cross-validation (Section 5.3.3) or marginal maximum likelihood estimation (Section 5.3.1). The CV criterion (5.51) is minimized by performing a grid search on $\log\delta$ between $[-35, 0]$.

Figure 5.4 shows boxplots of the integrated squared error for the two methods and for the different sample sizes. We observe that the median performance of CV is slightly better for $\lambda_{\mathrm{tot}} = 1\,000$, but, for larger sample sizes, EB leads to better median performance. More strikingly, we find that the variability of the CV estimates is much larger for each sample size and that there is in particular a long tail of very poor CV estimates (notice that the plot is on a log scale; on a linear scale, the tail is so long that the boxplots are barely visible). To obtain further insight into this evident instability of the CV estimates, Figure 5.5 shows the estimated hyperparameter values $\hat{\delta}$ for the two methods. We see that the CV estimates are much more variable than the EB estimates and that there is in particular a long tail of small hyperparameter values leading to the large ISE values seen in Figure 5.4. In other words, CV tends to overfit the data in a non-negligible number of cases. For each sample size, there is also a cluster of estimates at $\log(\hat{\delta}_{\mathrm{CV}}) = -35$, which was the lower bound of the search grid. In those cases, the minimum of the CV criterion is most likely situated at even smaller values of δ. The EB estimates, on the other hand, appear to be well-clustered around a reasonable value of δ without the heavy tail of the CV estimates.

These findings are fully consistent with those of Wood (2011, see in particular Figure 1), who

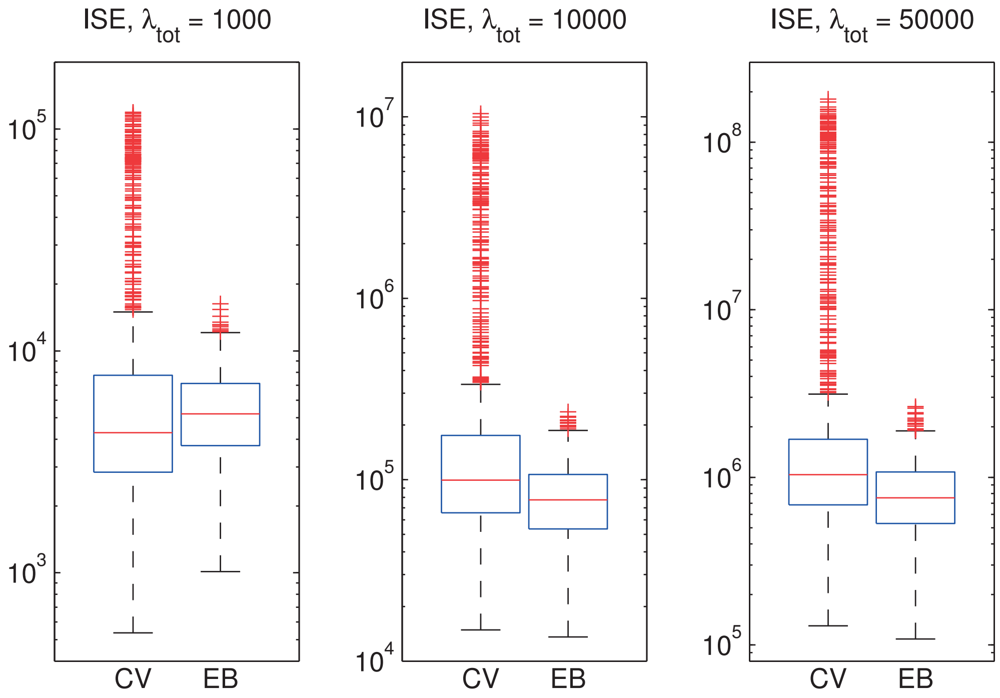
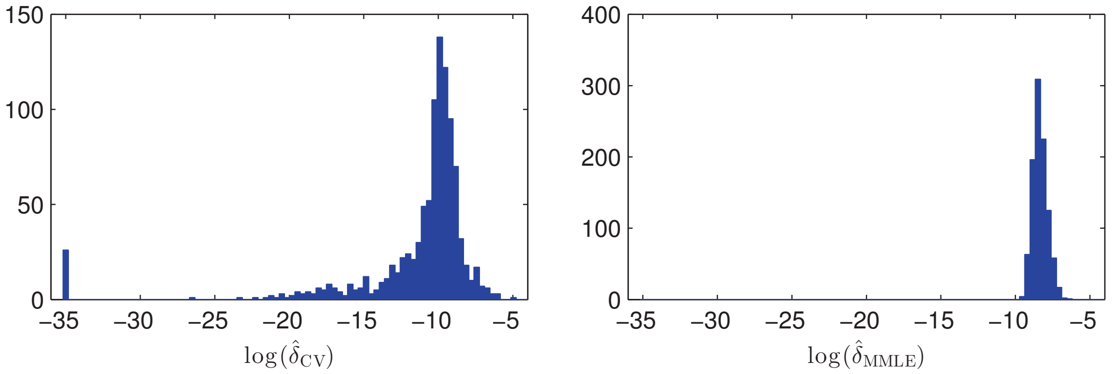


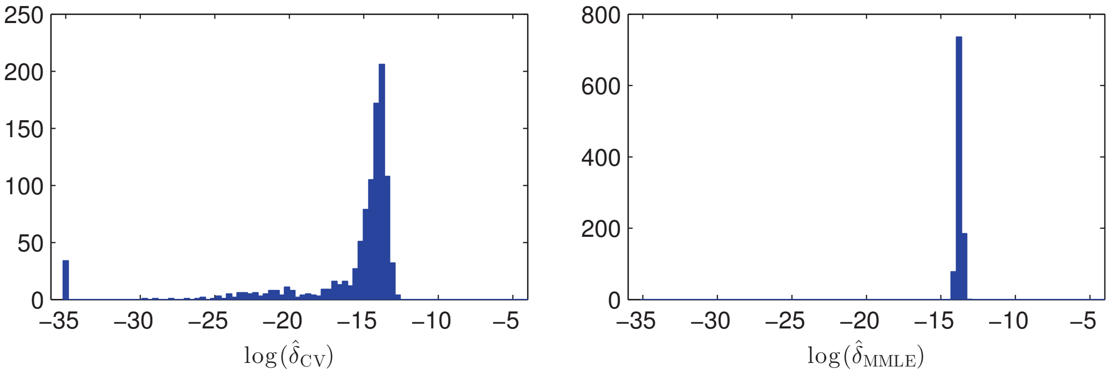
Figure 5.4: Integrated squared errors for weighted cross-validation and empirical Bayes in the two peaks on a uniform background test setup, with the spline coefficients estimated using the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$.

compares the CV and EB objective functions and finds that EB tends to have a much more pronounced optimum, leading to increased protection against overfitting and less variable estimates. We are not able to directly compare the objective functions since the marginal likelihood $p(\boldsymbol{y}|\delta)$ cannot be evaluated for our model, but Figure 5.5 indicates that also in our case the EB objective is better-behaved than the CV objective. Notice also that we are not able to guarantee that the MCEM iteration converges to a global optimum of the marginal likelihood, but, even if some of the estimates were local optima, the algorithm as a whole appears to be much more stable than cross-validation. For these reasons, we use empirical Bayes instead of cross-validation to choose δ in the remainder of this work.



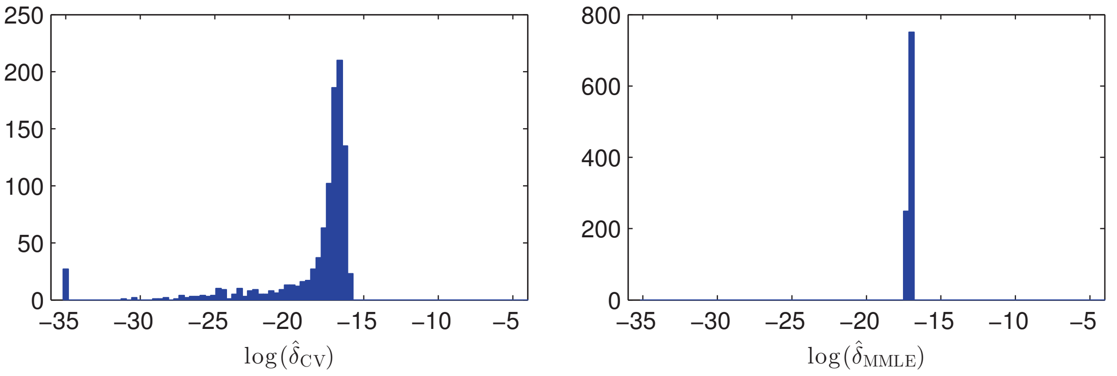
(a) CV, $\lambda_{\text{tot}} = 1000$

(b) EB, $\lambda_{\text{tot}} = 1000$



(c) CV, $\lambda_{\text{tot}} = 10000$

(d) EB, $\lambda_{\text{tot}} = 10000$



(e) CV, $\lambda_{\text{tot}} = 50000$

(f) EB, $\lambda_{\text{tot}} = 50000$

Figure 5.5: Estimated regularization strengths $\hat{\delta}$ for weighted cross-validation (CV) and empirical Bayes (EB) in the two peaks on a uniform background test setup.

6 Uncertainty quantification for smooth spectra

In this chapter, we study unfolded uncertainty quantification when then true spectrum f is known to be a smooth function. In other words, we aim to construct confidence bands around unfolded point estimates, such as the ones shown in Figure 5.2. When constructing interval estimates, we use the regularized point estimation techniques developed in Chapter 5 as our starting point.

Our focus in this chapter is on $1-\alpha$ pointwise confidence bands. That is, we aim to form a collection of random intervals $\left[\underline{f}(s;\boldsymbol{y}), \bar{f}(s;\boldsymbol{y})\right]$ which, for any $s \in E$, $\alpha \in (0,1)$ and intensity function $f \in V$, satisfy $\mathrm{P}_f\left(\underline{f}(s;\boldsymbol{y}) \leq f(s) \leq \bar{f}(s;\boldsymbol{y})\right) \geq 1-\alpha$. In practice, it is very difficult to satisfy this inequality for all f and all s, unless we have strong a priori information about the smoothness of f (Low, 1997; see also Section 7.2). So we can at best hope to form confidence bands whose coverage probability does not fall much below $1-\alpha$.

We first describe in Section 6.1 a number of conventional interval estimates that are often treated in the non-parametric regression literature. We then introduce in Section 6.2 an iterative technique for constructing bias-corrected confidence intervals. We use a simulation study in Section 6.3 to demonstrate that, due to the inherent bias of the regularized point estimates, the conventional interval estimates can suffer from drastic undercoverage, while the iteratively bias-corrected intervals yield close-to-nominal coverage with only a modest increase in interval length. We then analyze in Section 6.4 a simplified version of the unfolding problem, where the point estimates are linear functions of the smeared data and the noise is Gaussian, and use this setup to develop a data-driven way of choosing the number of bias-correction iterations. In that section, we also compare the iterative bias-correction to an alternative debiasing method based on undersmoothing and find that, at least in the test cases we have looked at, bias-correction yields shorter intervals than undersmoothing. Section 6.5 explains how the data-driven bias-correction can be applied to the full unfolding problem. These techniques are then applied in Section 6.6 to unfolding the Z boson invariant mass spectrum as measured in the CMS experiment at the Large Hadron Collider.

6.1 Conventional interval estimates

In this section, we introduce various standard interval estimates that are often treated in the literature. We will later demonstrate in Section 6.3 that all these intervals tend to suffer from major undercoverage in the unfolding problem. This happens because the intervals fail to properly account for the bias that is introduced to regularize the ill-posed problem.

6.1.1 Empirical Bayes credible intervals

For each $s \in E$, the empirical Bayes posterior $p(\boldsymbol{\beta}|\boldsymbol{y}, \hat{\delta}_{\mathrm{MMLE}})$ induces a posterior distribution for $f(s)$ through the transformation

$$f(s) = \sum_{j=1}^{p} \beta_j B_j(s). \tag{6.1}$$

The $1-\alpha$ *empirical Bayes credible interval* for $f(s)$ is then simply the interval whose endpoints are the $\alpha/2$ and $1-\alpha/2$ quantiles of this posterior. The MCMC sample $\{\boldsymbol{\beta}^{(s)}\}_{s=1}^{S}$ can be trivially transformed to a sample from the posterior of $f(s)$ by applying the transformation (6.1) to each posterior observation $\boldsymbol{\beta}^{(s)}$. The credible interval is then computed using the empirical quantiles of this posterior sample.

The empirical Bayes credible intervals are wider than purely frequentist constructions, such as those presented in Sections 6.1.3–6.1.5, which are based on the variability of the point estimator $\hat{\boldsymbol{\beta}}$. This is because the intervals include an extra component of uncertainty corresponding to the prior $p(\boldsymbol{\beta}|\hat{\delta}_{\mathrm{MMLE}})$. One can argue that this extra width accounts for the bias of $\hat{\boldsymbol{\beta}}$. Indeed, when there is no smearing, the noise is Gaussian and there are no constraints on $\boldsymbol{\beta}$, one can show (Ruppert et al., 2003, Section 6.4; Hastie and Tibshirani, 1990, Section 3.8.1) that the empirical Bayes intervals replace the standard error of $\hat{f}(s)$ by

$$\sqrt{\mathrm{var}\left(\hat{f}(s)\right) + \mathrm{E}\left(\mathrm{bias}^2_{\boldsymbol{\beta}}\left(\hat{f}(s)\right) \big| \hat{\delta}_{\mathrm{MMLE}}\right)}, \tag{6.2}$$

where the expectation is taken with respect to the prior $p(\boldsymbol{\beta}|\hat{\delta}_{\mathrm{MMLE}})$. Here the second term quantifies the "average bias" of $\hat{f}(s)$. The end result is that, subject to conditions on the choice of the regularization parameter, one expects the average coverage probability to be close to the nominal value (Nychka, 1988; Wahba, 1983):

$$\frac{1}{m+1} \sum_{i=1}^{m+1} \mathrm{P}_f\left(\underline{f}(s_i;\boldsymbol{y}) \le f(s_i) \le \bar{f}(s_i;\boldsymbol{y})\right) \approx 1-\alpha, \tag{6.3}$$

where $\{s_i\}_{i=1}^{m+1}$ is a grid of values in the true space. For these reasons, the empirical Bayes intervals are often studied in the non-parametric regression literature; see, for example, Ruppert et al. (2003, Section 6.4), Wood (2006a) and Marra and Wood (2012).

This discussion highlights an obvious limitation of the empirical Bayes intervals. Namely, average coverage of the type of Equation (6.3) does not give guarantees about pointwise coverage. Indeed, in practice, the intervals tend to undercover in areas of non-negligible bias and overcover for the rest of the function (Ruppert and Carroll, 2000), a finding supported by our simulation studies in Section 6.3.

6.1.2 Hierarchical Bayes credible intervals

The fully Bayesian solution to uncertainty quantification in unfolding is to use *hierarchical Bayes credible intervals*. These are defined the same way as the empirical Bayes credible intervals of Section 6.1.1, except that now one considers the intervals induced by the marginal posterior $p(\boldsymbol{\beta}|\boldsymbol{y})$ given in Equation (5.46).

The hierarchical Bayes intervals are often argued to improve upon the empirical Bayes intervals by also taking into account the uncertainty concerning the choice of the hyperparameter δ (Ruppert et al., 2003, Sections 6.4 and 16.3). However, in practice, the performance of the intervals depends on the choice of the hyperprior $p(\delta)$ and, as we will see in Section 6.3, the intervals suffer from the same issues concerning the bias as the empirical Bayes intervals.

Hierarchical Bayes intervals have been studied in the literature for several models that are related to unfolding. Examples include Weir (1997) with applications in tomography and Fahrmeir et al. (2004), who compare HB and EB intervals in generalized additive models and find their coverage performance to be roughly the same.

6.1.3 Gaussian confidence intervals and bootstrap resampling

Let $\hat{\boldsymbol{\beta}}$ be any of the point estimators discussed in Section 5.2. Most frequentist uncertainty quantification techniques are based on using the variability of $\hat{\boldsymbol{\beta}}$ as a measure of its uncertainty. Assume that $\hat{\boldsymbol{\beta}}$ is approximately unbiased and Gaussian, that is, $\hat{\boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}(\boldsymbol{\beta}, \boldsymbol{V})$, for some covariance matrix $\boldsymbol{V}$. Then $\hat{f}(s) = \sum_{j=1}^{p} \hat{\beta}_j B_j(s) = \boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$, where $\boldsymbol{c} = [B_1(s), \ldots, B_p(s)]^{\mathrm{T}}$, is also approximately unbiased and Gaussian, $\hat{f}(s) \overset{a}{\sim} \mathcal{N}(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}, \boldsymbol{c}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{c})$, with $f(s) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$. Hence we have the approximate pivot

$$\frac{\hat{f}(s) - f(s)}{\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{c}}} \overset{a}{\sim} \mathcal{N}(0,1). \tag{6.4}$$

Inversion of this pivot leads to the confidence interval

$$\begin{aligned} &\left[\hat{f}(s) - z_{1-\alpha/2}\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{c}},\, \hat{f}(s) + z_{1-\alpha/2}\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{c}}\right] & (6.5)\\ =& \left[\hat{f}(s) - z_{1-\alpha/2}\sqrt{\mathrm{var}\big(\hat{f}(s)\big)},\, \hat{f}(s) + z_{1-\alpha/2}\sqrt{\mathrm{var}\big(\hat{f}(s)\big)}\right], & (6.6)\end{aligned}$$

with approximate coverage probability $1-\alpha$,

$$\mathrm{P}_f\left(\hat{f}(s) - z_{1-\alpha/2}\sqrt{\mathrm{var}\big(\hat{f}(s)\big)} \le f(s) \le \hat{f}(s) + z_{1-\alpha/2}\sqrt{\mathrm{var}\big(\hat{f}(s)\big)}\right) \approx 1-\alpha. \tag{6.7}$$

This shows that in some cases it can indeed be reasonable to base the uncertainties on the variability of a point estimator.

There are however a number of reasons why this approach might fail for unfolding. Firstly, even if $\hat{\boldsymbol{\beta}}$ was approximately unbiased and Gaussian, the covariance $\boldsymbol{V}$ and hence the variance $\mathrm{var}\big(\hat{f}(s)\big)$ is usually unknown. Secondly, due to the Poisson noise and the positivity constraint, the distribution of $\hat{\boldsymbol{\beta}}$ tends to be asymmetric, especially for components that correspond to areas where the true intensity $f(s)$ is small. Thirdly, and most importantly, regularization makes the estimator $\hat{\boldsymbol{\beta}}$ biased and this bias is not taken into account by the Gaussian interval (6.6).

These issues can be addressed using *bootstrap resampling* (Efron and Tibshirani, 1993; Davison and Hinkley, 1997). Let us first consider the unknown variance $\mathrm{var}\big(\hat{f}(s)\big)$ and postpone the discussion of asymmetry and bias to the following sections. The basic idea is the following: Imagine we had a sample of point estimators $\hat{f}(s)$. Then we could easily estimate the variance of $\hat{f}(s)$ using the sample variance and replace $\mathrm{var}\big(\hat{f}(s)\big)$ in (6.6) by this estimate. Obviously such sample is not available to us, but, if we knew $\boldsymbol{\mu} = \boldsymbol{K}\boldsymbol{\beta}$, we could generate one by resampling new observations $\boldsymbol{y}^*$ from the $\boldsymbol{y}^* \sim \mathrm{Poisson}(\boldsymbol{\mu})$ distribution. The key idea in (parametric) bootstrapping is to replace the unknown $\boldsymbol{\mu}$ by an estimate $\hat{\boldsymbol{\mu}}$ and then carry out the resampling from the $\boldsymbol{y}^* \sim \mathrm{Poisson}(\hat{\boldsymbol{\mu}})$ distribution. In our case, a reasonable estimator of $\boldsymbol{\mu}$ is the MLE $\hat{\boldsymbol{\mu}} = \boldsymbol{y}$. High energy physicists typically call bootstrap resamples “toy experiments” or “toy data”.

To be more specific, bootstrap estimation of the variance of $\hat{f}(s)$ proceeds as follows:

1. Resample $\boldsymbol{y}^{*(1)},\ldots,\boldsymbol{y}^{*(R_{\mathrm{UQ}})} \overset{\text{i.i.d.}}{\sim} \mathrm{Poisson}(\hat{\boldsymbol{\mu}})$.
2. For each $r = 1,\ldots,R_{\mathrm{UQ}}$, compute the estimator $\hat{f}^{*(r)}(s)$ corresponding to $\boldsymbol{y}^{*(r)}$.
3. Estimate $\mathrm{var}\big(\hat{f}(s)\big)$ using the sample variance of $\big\{\hat{f}^{*(r)}(s)\big\}_{r=1}^{R_{\mathrm{UQ}}}$. Denote the resulting estimate by $\widehat{\mathrm{var}}^*\big(\hat{f}(s)\big)$.

The bootstrap version of the Gaussian confidence interval (6.6) is then obtained by replacing $\mathrm{var}\big(\hat{f}(s)\big)$ with the bootstrap variance estimate $\widehat{\mathrm{var}}^*\big(\hat{f}(s)\big)$. The resulting interval is

$$\left[\hat{f}(s) - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}^*\big(\hat{f}(s)\big)},\, \hat{f}(s) + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}^*\big(\hat{f}(s)\big)}\right]. \tag{6.8}$$

For computational reasons, we keep the regularization strength δ fixed to its original value $\hat{\delta} = \hat{\delta}(\boldsymbol{y})$ when computing the resampled estimator $\hat{f}^{*(r)}(s)$. The effect of this is that the bootstrap sample $\big\{\hat{f}^{*(r)}(s)\big\}_{r=1}^{R_{\mathrm{UQ}}}$ does not include the variability caused by the data-driven choice of δ.

Notice that for the unconstrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ of Section 5.2.4 we can obtain an estimate of $\mathrm{var}\left(\hat{f}(s)\right)$ without using the bootstrap. Ignoring the data-dependence of $\hat{\boldsymbol{\Sigma}}$ and $\hat{\delta}$, the covariance of $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ is given by

$$\boldsymbol{V} = \mathrm{cov}\left(\hat{\boldsymbol{\beta}}_{\mathrm{G}}\right) = \left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\hat{\delta}\boldsymbol{\Omega}_{\mathrm{A}}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\,\mathrm{cov}(\boldsymbol{y})\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K}\left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\hat{\delta}\boldsymbol{\Omega}_{\mathrm{A}}\right)^{-1}. \tag{6.9}$$

We can hence estimate this using

$$\hat{\boldsymbol{V}} = \widehat{\mathrm{cov}}\left(\hat{\boldsymbol{\beta}}_{\mathrm{G}}\right) = \left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\hat{\delta}\boldsymbol{\Omega}_{\mathrm{A}}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K}\left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\hat{\delta}\boldsymbol{\Omega}_{\mathrm{A}}\right)^{-1}, \tag{6.10}$$

which leads us to estimate the variance of $\hat{f}(s)$ in (6.6) using $\widehat{\mathrm{var}}\left(\hat{f}(s)\right) = \boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{V}}\boldsymbol{c}$.

6.1.4 Bootstrap percentile intervals

Bootstrap percentile intervals provide a way of handling asymmetries in $\hat{f}(s)$ by considering approximate $1-\alpha$ confidence intervals of the form $\left[\hat{f}^*_{\alpha/2}(s), \hat{f}^*_{1-\alpha/2}(s)\right]$. Here $\hat{f}^*_{\alpha/2}(s)$ and $\hat{f}^*_{1-\alpha/2}(s)$ are the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of the bootstrap sample $\left\{\hat{f}^{*(r)}(s)\right\}_{r=1}^{R_{\mathrm{UQ}}}$ obtained using the procedure described in Section 6.1.3.

The method can be motivated (Davison and Hinkley, 1997, Section 5.3.1) as follows: Suppose $\hat{f}(s)$ has an asymmetric sampling distribution around $f(s)$ and assume that there exists a transformation $\phi = m(f(s))$, with m increasing, such that the sampling distribution of $\hat{\phi} = m(\hat{f}(s))$ is symmetric around ϕ. In other words, the distribution of $\hat{\phi}-\phi$ is symmetric around zero. Denoting the $\alpha/2$ and $1-\alpha/2$ quantiles of this distribution by $q_{\alpha/2}$ and $q_{1-\alpha/2}$, we have

$$\begin{aligned}
1-\alpha &= \mathrm{P}_f\left(q_{\alpha/2} \le \hat{\phi}-\phi \le q_{1-\alpha/2}\right) && (6.11)\\
&= \mathrm{P}_f\left(-q_{1-\alpha/2} \le \phi-\hat{\phi} \le -q_{\alpha/2}\right) && (6.12)\\
&= \mathrm{P}_f\left(q_{\alpha/2} \le \phi-\hat{\phi} \le q_{1-\alpha/2}\right) && (6.13)\\
&= \mathrm{P}_f\left(\hat{\phi}+q_{\alpha/2} \le \phi \le \hat{\phi}+q_{1-\alpha/2}\right), && (6.14)
\end{aligned}$$

where the third equality follows from the symmetry of the distribution. We do not know the quantiles $q_{\alpha/2}$ and $q_{1-\alpha/2}$, but we can estimate them using the quantiles of $\hat{\phi}^*-\hat{\phi}$, where $\hat{\phi}^*$ is a bootstrap replication of $\hat{\phi}$. In other words, $q_{\alpha/2} \approx \hat{\phi}^*_{\alpha/2}-\hat{\phi}$ and $q_{1-\alpha/2} \approx \hat{\phi}^*_{1-\alpha/2}-\hat{\phi}$, where $\hat{\phi}^*_{\alpha/2}$ and $\hat{\phi}^*_{1-\alpha/2}$ are the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of the bootstrap sample $\left\{\hat{\phi}^{*(r)}\right\}_{r=1}^{R_{\mathrm{UQ}}}$. Substituting these into Equation (6.14), we find

$$\begin{aligned}
1-\alpha &\approx \mathrm{P}_f\left(\hat{\phi}^*_{\alpha/2} \le \phi \le \hat{\phi}^*_{1-\alpha/2}\right) && (6.15)\\
&= \mathrm{P}_f\left(m^{-1}(\hat{\phi}^*_{\alpha/2}) \le m^{-1}(\phi) \le m^{-1}(\hat{\phi}^*_{1-\alpha/2})\right) && (6.16)\\
&= \mathrm{P}_f\left(\hat{f}^*_{\alpha/2}(s) \le f(s) \le \hat{f}^*_{1-\alpha/2}(s)\right). && (6.17)
\end{aligned}$$

As a result, $\left[\hat{f}^*_{\alpha/2}(s), \hat{f}^*_{1-\alpha/2}(s)\right]$ provides an approximate $1-\alpha$ confidence interval for $f(s)$.

Notice that the final interval does not require us to know the transformation m.

The main problem with this approach is that the required transformation m may not exist. In particular, this approach cannot be used to correct for the bias in $\hat{f}(s)$. To see why, assume that the distribution of $\hat{f}(s)$ is centered around $f(s)+b$, where b is the bias. One could then try to use the transformation $\phi = m(f(s)) = f(s) - b$, but this would center $\hat{\phi}$ around $f(s)$ instead of the required value ϕ.

6.1.5 Basic bootstrap intervals

Basic bootstrap intervals quantify the uncertainty of $\hat{f}(s)$ using approximate $1-\alpha$ confidence intervals of the form $\left[2\hat{f}(s) - \hat{f}^*_{1-\alpha/2}(s), 2\hat{f}(s) - \hat{f}^*_{\alpha/2}(s)\right]$. These intervals can be used to partially account for the bias of $\hat{f}(s)$.

The basic interval is based on estimating the quantiles of $\hat{f}(s) - f(s)$ using the bootstrap (Davison and Hinkley, 1997, Section 5.2). Let $q_{\alpha/2}$ and $q_{1-\alpha/2}$ be the $\alpha/2$ and $1-\alpha/2$ quantiles of $\hat{f}(s) - f(s)$. Then

$$\begin{aligned} 1-\alpha &= \mathrm{P}_f\left(q_{\alpha/2} \le \hat{f}(s) - f(s) \le q_{1-\alpha/2}\right) && (6.18)\\ &= \mathrm{P}_f\left(\hat{f}(s) - q_{1-\alpha/2} \le f(s) \le \hat{f}(s) - q_{\alpha/2}\right). && (6.19) \end{aligned}$$

The bootstrap estimates of the quantiles are given by $q_{\alpha/2} \approx \hat{f}^*_{\alpha/2}(s) - \hat{f}(s)$ and $q_{1-\alpha/2} \approx \hat{f}^*_{1-\alpha/2}(s) - \hat{f}(s)$. Substituting these into Equation (6.19), yields

$$\mathrm{P}_f\left(2\hat{f}(s) - \hat{f}^*_{1-\alpha/2}(s) \le f(s) \le 2\hat{f}(s) - \hat{f}^*_{\alpha/2}(s)\right) \approx 1-\alpha. \tag{6.20}$$

Hence $\left[2\hat{f}(s) - \hat{f}^*_{1-\alpha/2}(s), 2\hat{f}(s) - \hat{f}^*_{\alpha/2}(s)\right]$ is an approximate $1-\alpha$ confidence interval for $f(s)$.

To enable this approach to probe the bias of $\hat{f}(s)$, we need to employ a bootstrap sampling scheme which is different from the one introduced in Section 6.1.3. Since the smeared data $\boldsymbol{y}$ can be parameterized using either the smeared mean $\boldsymbol{\mu}$ or the spline coefficients $\boldsymbol{\beta}$, there are at least two reasonable ways of obtaining the bootstrap replicates $\boldsymbol{y}^*$. In Sections 6.1.3 and 6.1.4, we used $\boldsymbol{y}^* \sim \mathrm{Poisson}(\hat{\boldsymbol{\mu}})$ with $\hat{\boldsymbol{\mu}} = \boldsymbol{y}$, but an alternative procedure would be to use $\boldsymbol{y}^* \sim \mathrm{Poisson}(\boldsymbol{K}\hat{\boldsymbol{\beta}})$. It is easy to see that if the former procedure was used, then the distribution of $\hat{f}^*(s) - \hat{f}(s)$ would not capture the bias of $\hat{f}(s)$, but, with the latter procedure, the bias is partially included. In an attempt to account for the bias, we will hence construct the basic intervals using the sampling scheme $\boldsymbol{y}^* \sim \mathrm{Poisson}(\boldsymbol{K}\hat{\boldsymbol{\beta}})$, although we will see in Section 6.3 that in practice the performance of the resulting intervals leaves much room for improvement.

6.1.6 Other bootstrap intervals

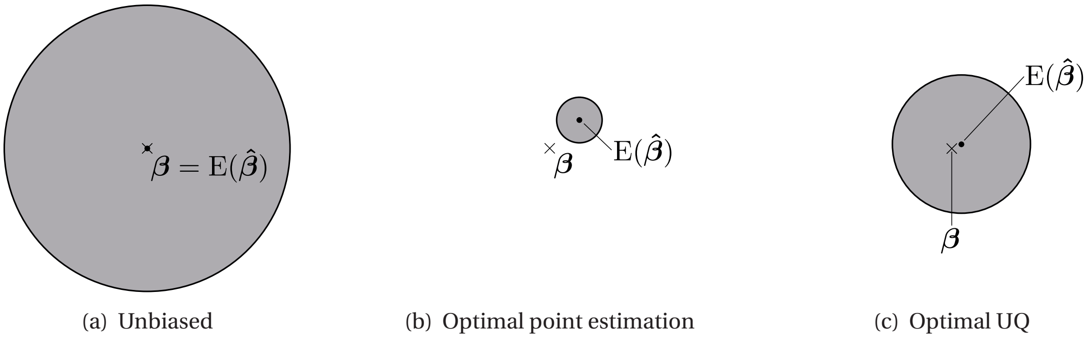
Several methods have been proposed for improving the performance of bootstrap confidence intervals. We mention in particular the BC_a intervals of Efron (1987), with BC_a standing for bias-corrected and accelerated. The BC_a intervals modify the quantiles used in percentile intervals to account for the bias of $\hat{f}(s)$ and the dependence of $\mathrm{var}\big(\hat{f}(s)\big)$ on $f(s)$. We considered using BC_a intervals to improve upon the basic and percentile intervals, but in practice we found that a better strategy is to directly correct the point estimator $\hat{\boldsymbol{\beta}}$ for the bias instead of correcting the bootstrap percentiles; see Section 6.2. The reason is that this allows us to form confidence intervals that also account for the uncertainty of the bias-correction. Furthermore, in the presence of a significant bias, the BC_a intervals would be formed using quantiles that are very far in the tails of the bootstrap distribution and hence difficult to estimate accurately using simulations.

A rather different approach consists in changing the way the bootstrap resampling is performed. Beran (1995), for example, studies the problem of constructing a confidence ball for $\boldsymbol{\mu} \in \mathbb{R}^n$ in the model $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{I})$, with the ball centered at the James–Stein estimator $\hat{\boldsymbol{\mu}}_{\mathrm{JS}} = \big(1-(n-2)/\|\boldsymbol{y}\|_2^2\big)\,\boldsymbol{y}$ and its radius calibrated using the bootstrap to yield the desired coverage probability. He shows that the calibration should not be done using $\boldsymbol{y}^* \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_{\mathrm{JS}}, \boldsymbol{I})$ or $\boldsymbol{y}^* \sim \mathcal{N}(\boldsymbol{y}, \boldsymbol{I})$. Instead, the resampling scheme should be $\boldsymbol{y}^* \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_{\mathrm{CL}}, \boldsymbol{I})$, where $\hat{\boldsymbol{\mu}}_{\mathrm{CL}}$ is a modified James–Stein estimator, with the property that $\|\hat{\boldsymbol{\mu}}_{\mathrm{CL}}\|_2$ estimates $\|\boldsymbol{\mu}\|_2$ well. In our case, this would correspond to centering the confidence intervals around $\hat{f}$, but calibrating their length based on some modification of $\hat{\boldsymbol{\beta}}$. However, it is not immediately clear how $\hat{\boldsymbol{\beta}}$ should be modified to obtain accurate bootstrap calibration and the location of the resulting intervals would in any case be suboptimal because of the bias in $\hat{f}$.

6.2 Iteratively bias-corrected percentile intervals

When the variability of a regularized point estimator $\hat{\boldsymbol{\beta}}$ is used to construct confidence intervals, the intervals either ignore or only partially account for the bias of the estimator. This is the case for all the intervals discussed in Section 6.1 and we will see in Section 6.3 that this translates into poor coverage performance. In this section, we propose solving this problem by iteratively bias-correcting $\hat{\boldsymbol{\beta}}$ and then basing the intervals on the variability of the bias-corrected estimator $\hat{\boldsymbol{\beta}}_{\mathrm{BC}}$ instead of the original estimator $\hat{\boldsymbol{\beta}}$. This approach has analogies with the work of Javanmard and Montanari (2014), who use debiasing to quantify the uncertainty in ℓ^1-regularized lasso regression.

At first it may seem counterintuitive that debiasing the point estimator $\hat{\boldsymbol{\beta}}$ could yield improved uncertainty quantification—the bias is after all needed to regularize the ill-posed problem and reducing the bias should increase the variance of the estimator. However, as illustrated in Figure 6.1, *the optimal bias-variance trade-off is different for point estimation and uncertainty quantification.* Each iteration of the bias-correction procedure will shift the balance towards


$\boldsymbol{\beta} = \mathrm{E}(\hat{\boldsymbol{\beta}})$
$\boldsymbol{\beta}$
$\mathrm{E}(\hat{\boldsymbol{\beta}})$
$\mathrm{E}(\hat{\boldsymbol{\beta}})$
$\boldsymbol{\beta}$

(a) Unbiased (b) Optimal point estimation (c) Optimal UQ

Figure 6.1: Bias-variance trade-off for point estimation and uncertainty quantification. The unbiased estimator in Figure (a) has a large variability illustrated by the shaded region and needs to be regularized by introducing a bias. Optimal bias-variance trade-off for point estimation tends to look like Figure (b), where the variability is not a good measure of uncertainty. Figure (c) illustrates a better bias-variance trade-off for uncertainty quantification where the balance is shifted towards the direction of less bias and more variance. Now the variability provides a good measure of uncertainty with better coverage than in Figure (b) but smaller size than in Figure (a).

the direction of less bias and more variance and by doing so will improve the coverage of the intervals at the expense of increased interval length. Indeed, for uncertainty quantification, there is a trade-off between the coverage and the length of the confidence intervals. The results of this chapter however indicate that, by stopping the bias-correction iteration early enough, it is possible to balance the coverage-length trade-off in such a way that the intervals have nearly nominal coverage with only a modest increase in interval length. In other words, the iterative bias-correction is able to remove so much of the bias that the coverage is close to the nominal value, but the small amount of residual bias that remains is enough to regularize the interval length. A similar phenomenon has been observed by Javanmard and Montanari (2014) in debiased lasso regression.

6.2.1 Iterative bootstrap bias-correction

Our debiasing procedure is based on a repeated application of the bootstrap bias-correction. Kuk (1995) and Goldstein (1996) have used a similar iterative bias-correction technique in generalized linear mixed models with an emphasis on point estimation, while Cornillon et al. (2013, 2014) employ the approach to improve the predictive performance of linear nonparametric smoothers. We are however not aware of previous applications of the procedure to uncertainty quantification in ill-posed inverse problems.

Let $\hat{\boldsymbol{\beta}}$ be any of the unfolded point estimators considered in Section 5.2. The bias of $\hat{\boldsymbol{\beta}}$ is given by $\mathrm{bias}(\hat{\boldsymbol{\beta}}) = \mathrm{E}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}$, where $\mathrm{E}_{\boldsymbol{\beta}}(\cdot)$ denotes the expectation when the observations $\boldsymbol{y}$ follow the $\mathrm{Poisson}(\boldsymbol{K}\boldsymbol{\beta})$ distribution. The conventional bootstrap bias estimate (Davison and Hinkley,

1997) replaces the unknown parameter $\boldsymbol{\beta}$ by its estimate $\hat{\boldsymbol{\beta}}$. Denoting the observed value of $\hat{\boldsymbol{\beta}}$ by $\hat{\boldsymbol{\beta}}^{(0)}$, the bootstrap bias estimate is given by

$$\widehat{\text{bias}}^{(0)}(\hat{\boldsymbol{\beta}}) = \mathrm{E}_{\hat{\boldsymbol{\beta}}^{(0)}}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}}^{(0)}, \tag{6.21}$$

where in practice the expectation is estimated using simulations. The bias-corrected estimator is then obtained by subtracting the estimated bias,

$$\hat{\boldsymbol{\beta}}^{(1)} = \hat{\boldsymbol{\beta}}^{(0)} - \widehat{\text{bias}}^{(0)}(\hat{\boldsymbol{\beta}}). \tag{6.22}$$

This is not a perfectly unbiased estimator since in Equation (6.21) the unknown parameter $\boldsymbol{\beta}$ was replaced by the estimate $\hat{\boldsymbol{\beta}}^{(0)}$. But since $\hat{\boldsymbol{\beta}}^{(1)}$ should be less biased than $\hat{\boldsymbol{\beta}}^{(0)}$, we could hope to obtain a better estimate of the bias by replacing $\hat{\boldsymbol{\beta}}^{(0)}$ in Equation (6.21) by $\hat{\boldsymbol{\beta}}^{(1)}$. This leads to a new bias estimate

$$\widehat{\text{bias}}^{(1)}(\hat{\boldsymbol{\beta}}) = \mathrm{E}_{\hat{\boldsymbol{\beta}}^{(1)}}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}}^{(1)}, \tag{6.23}$$

with the corresponding bias-corrected point estimate

$$\hat{\boldsymbol{\beta}}^{(2)} = \hat{\boldsymbol{\beta}}^{(0)} - \widehat{\text{bias}}^{(1)}(\hat{\boldsymbol{\beta}}). \tag{6.24}$$

The same logic leads us then to replace $\hat{\boldsymbol{\beta}}^{(1)}$ in Equation (6.23) by $\hat{\boldsymbol{\beta}}^{(2)}$ and so on. The result is the following *iterative bootstrap bias-correction* procedure:

1. Estimate the bias: $\widehat{\text{bias}}^{(t)}(\hat{\boldsymbol{\beta}}) = \mathrm{E}_{\hat{\boldsymbol{\beta}}^{(t)}}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}}^{(t)}$.
2. Compute the bias-corrected estimate: $\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(0)} - \widehat{\text{bias}}^{(t)}(\hat{\boldsymbol{\beta}})$.

For nonlinear estimators $\hat{\boldsymbol{\beta}}$, the expectation $\mathrm{E}_{\hat{\boldsymbol{\beta}}^{(t)}}(\hat{\boldsymbol{\beta}})$ will in practice have to be computed by resampling R_{BC} i.i.d. observations from the Poisson$(\boldsymbol{K}\hat{\boldsymbol{\beta}}^{(t)})$ distribution. For computational reasons, we do not resample the regularization strength δ when computing the expectation but instead keep it fixed to its original value $\hat{\delta}$. We also enforce the positivity constraint by setting any negative entries to zero. The resulting bias-correction procedure is summarized in Algorithm 2. Notice that the procedure is fully generic in the sense that it does not depend on the specific forms of the point estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\delta}$. The choice of the number of bias-correction iterations N_{BC} is discussed in Sections 6.4.3 and 6.5.

6.2.2 Percentile intervals from the bias-corrected point estimator

Let $\hat{\boldsymbol{\beta}}_{\mathrm{BC}}$ denote the bias-corrected estimate of the spline coefficients $\boldsymbol{\beta}$ obtained using N_{BC} bias-correction iterations and let $\hat{f}_{\mathrm{BC}}(s) = \sum_{j=1}^{p} \hat{\beta}_{\mathrm{BC},j} B_j(s)$ be the corresponding intensity function estimate. Since $\hat{f}_{\mathrm{BC}}$ should be less biased than $\hat{f}$, we can use the techniques presented in Sections 6.1.3–6.1.5 with $\hat{f}_{\mathrm{BC}}$ to obtain improved confidence intervals. Since in most cases $\hat{f}_{\mathrm{BC}}$ is not be available in closed form, we need to resort to a double bootstrap: We first resample

Algorithm 2 Iterative bootstrap bias-correction

```
Input:
  $\hat{\boldsymbol{\beta}}^{(0)}$ — Observed value of the estimator $\hat{\boldsymbol{\beta}}$
  $\hat{\delta}$ — Estimated value of the regularization strength $\delta$
  $N_{\mathrm{BC}}$ — Number of bias-correction iterations
  $R_{\mathrm{BC}}$ — Size of the bootstrap sample
Output:
  $\hat{\boldsymbol{\beta}}_{\mathrm{BC}}$ — Bias-corrected point estimate

  for $t = 0$ to $N_{\mathrm{BC}} - 1$ do
      Sample $\boldsymbol{y}^{*(1)}, \boldsymbol{y}^{*(2)}, \ldots, \boldsymbol{y}^{*(R_{\mathrm{BC}})} \overset{\text{i.i.d.}}{\sim} \mathrm{Poisson}(\boldsymbol{K}\hat{\boldsymbol{\beta}}^{(t)})$;
      For each $r = 1, \ldots, R_{\mathrm{BC}}$, compute the estimator $\hat{\boldsymbol{\beta}}^{*(r)}$ using $\boldsymbol{y}^{*(r)}$ as the smeared data and
      $\hat{\delta}$ as the regularization strength;
      Compute $\widehat{\mathrm{bias}}^{*(t)}(\hat{\boldsymbol{\beta}}) = \frac{1}{R_{\mathrm{BC}}} \sum_{r=1}^{R_{\mathrm{BC}}} \hat{\boldsymbol{\beta}}^{*(r)} - \hat{\boldsymbol{\beta}}^{(t)}$;
      Set $\hat{\boldsymbol{\beta}}'^{(t+1)} = \hat{\boldsymbol{\beta}}^{(0)} - \widehat{\mathrm{bias}}^{*(t)}(\hat{\boldsymbol{\beta}})$;
      Set $\hat{\boldsymbol{\beta}}^{(t+1)} = \max\left(\hat{\boldsymbol{\beta}}'^{(t+1)}, \boldsymbol{0}\right)$, where the maximum is taken element-wise;
  end for
  return $\hat{\boldsymbol{\beta}}_{\mathrm{BC}} = \hat{\boldsymbol{\beta}}^{(N_{\mathrm{BC}})}$;
```

new i.i.d. observations $\boldsymbol{y}^{*(r)}, r = 1, \ldots, R_{\mathrm{UQ}}$, and compute the corresponding regularized point estimates $\hat{\boldsymbol{\beta}}^{*(r)}$. We then run the iterative bootstrap bias-correction (Algorithm 2) for each of these to obtain a bootstrap sample of bias-corrected spline coefficients $\hat{\boldsymbol{\beta}}_{\mathrm{BC}}^{*(r)}$ along with the corresponding intensity function estimates $\hat{f}_{\mathrm{BC}}^{*(r)}$. We can then use one of the bootstrap interval estimates discussed in Sections 6.1.3–6.1.5 to construct an approximate confidence interval for $f(s)$ based on $\{\hat{f}_{\mathrm{BC}}^{*(r)}(s)\}_{r=1}^{R_{\mathrm{UQ}}}$. We again keep δ fixed to its original estimated value $\hat{\delta}$ throughout this procedure.

The simplest approach would be to use the bootstrap version of the Gaussian confidence interval given in Equation (6.8). However, preliminary simulations showed that the Gaussian intervals suffer from slight undercoverage in areas where the true intensity $f(s)$ is small. This is due to the skewness of the sampling distribution of $\hat{f}_{\mathrm{BC}}(s)$ caused by the positivity constraint. As explained in Section 6.1.4, this kind of asymmetries can be taken into account using the $1-\alpha$ bootstrap percentile intervals $\left[\hat{f}^*_{\mathrm{BC},\alpha/2}(s), \hat{f}^*_{\mathrm{BC},1-\alpha/2}(s)\right]$, where $\hat{f}^*_{\mathrm{BC},\alpha/2}(s)$ and $\hat{f}^*_{\mathrm{BC},1-\alpha/2}(s)$ denote the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of $\{\hat{f}_{\mathrm{BC}}^{*(r)}(s)\}_{r=1}^{R_{\mathrm{UQ}}}$. The percentile intervals were indeed found to yield better coverage performance than the Gaussian intervals and, as a result, we will primarily use them to quantify the uncertainty of f. The resulting uncertainty quantification procedure is summarized in Algorithm 3.

Algorithm 3 Iteratively bias-corrected percentile intervals

```
Input:
  $\boldsymbol{y}$ — Smeared data
  $\hat{\delta}$ — Estimated value of the regularization strength $\delta$
  $N_{\mathrm{BC}}$ — Number of bias-correction iterations
  $R_{\mathrm{BC}}$ — Size of the bootstrap sample for bias-correction
  $R_{\mathrm{UQ}}$ — Size of the bootstrap sample for uncertainty quantification
Output:
  $[\underline{f}(s), \bar{f}(s)]$ — Iteratively bias-corrected $1-\alpha$ percentile interval for $f(s)$

  Sample $\boldsymbol{y}^{*(1)}, \boldsymbol{y}^{*(2)}, \ldots, \boldsymbol{y}^{*(R_{\mathrm{UQ}})} \overset{\text{i.i.d.}}{\sim} \mathrm{Poisson}(\hat{\boldsymbol{\mu}})$ with $\hat{\boldsymbol{\mu}} = \boldsymbol{y}$;
  for $r = 1$ to $R_{\mathrm{UQ}}$ do
      Compute the estimator $\hat{\boldsymbol{\beta}}^{*(r)}$ using $\boldsymbol{y}^{*(r)}$ as the smeared data and $\hat{\delta}$ as the regularization strength;
      Compute $\hat{\boldsymbol{\beta}}_{\mathrm{BC}}^{*(r)}$ using Algorithm 2 with $N_{\mathrm{BC}}$ bias-correction iterations, $R_{\mathrm{BC}}$ bootstrap replicates and regularization strength $\hat{\delta}$;
      Compute $\hat{f}_{\mathrm{BC}}^{*(r)}(s) = \sum_{j=1}^{p} \hat{\beta}_{\mathrm{BC},j}^{*(r)} B_j(s)$;
  end for
  return $[\underline{f}(s), \bar{f}(s)] = [\hat{f}^{*}_{\mathrm{BC},\alpha/2}(s), \hat{f}^{*}_{\mathrm{BC},1-\alpha/2}(s)]$;
```

6.3 Comparison of iterative bias-correction with conventional interval estimates

In this section, we perform a simulation study comparing the following unfolded confidence intervals:

(i) Iteratively bias-corrected percentile intervals (Section 6.2.2) induced by either the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ (Section 5.2.1) or the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ (Section 5.2.3);

(ii) Bootstrap percentile intervals (Section 6.1.4) induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$;

(iii) Basic bootstrap intervals (Section 6.1.5) induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$;

(iv) Empirical Bayes credible intervals (Section 6.1.1);

(v) Hierarchical Bayes credible intervals (Secction 6.1.2) for the four hyperpriors studied in Section 5.4.2.

We compare the methods using the two peaks on a uniform background test setup of Section 3.4.1. For methods (i)–(iv), the regularization strength δ is chosen using the MMLE as described in Section 5.3.1. The boundary hyperparameters as well as the parameters of the MCMC sampler and the MCEM iteration are set as in Sections 5.4.1 and 5.4.2. The number of bias-correction iterations is set to $N_{\mathrm{BC}} = 15$, $N_{\mathrm{BC}} = 10$ and $N_{\mathrm{BC}} = 5$ for the small, medium and large sample size cases, respectively. In method (i), the bias estimates are obtained using

$R_{\mathrm{BC}} = 10$ bootstrap observations and the bootstrap confidence intervals in methods (i)–(iii) are formed using a bootstrap sample of size $R_{\mathrm{UQ}} = 200$. All the intervals are formed at 95 % pointwise confidence level and all coverage studies are performed using 1 000 independent replications.

Figure 6.2 shows the iteratively bias-corrected percentile intervals induced by the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ for the different sample sizes. We observe that the bias-corrected point estimator $\hat{f}_{\mathrm{BC}}$ is able to capture the two peaks of the true intensity f better than the posterior mean. The price to pay for this is an increased wiggliness of $\hat{f}_{\mathrm{BC}}$, especially near the boundaries of the true space E. The percentile intervals induced by $\hat{f}_{\mathrm{BC}}$ also capture the shape of f well and, for these particular realizations, cover the true value of f throughout the spectrum.

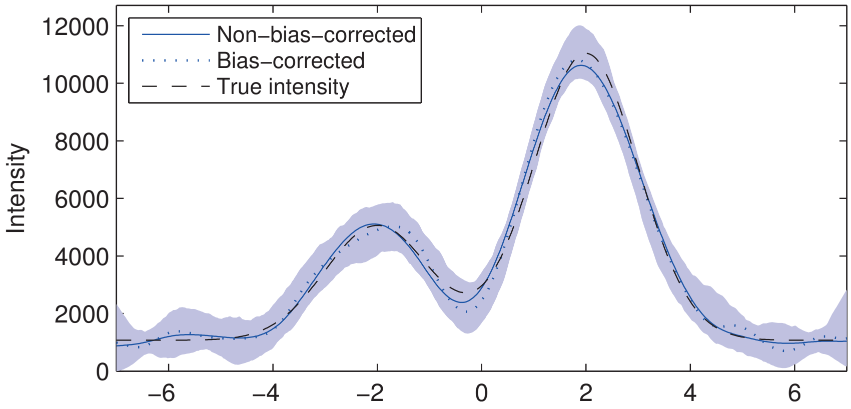
Depending on the number of bias-correction iterations used, computing the intervals shown in Figure 6.2 took 14–39 hours on a quad-core 2.7 GHz Intel Core i5 processor, with the outer bootstrap loop parallelized to the four cores of the setup. This is too long to perform a meaningful coverage study for these intervals. In order to get a handle on the coverage probability of the iteratively bias-corrected intervals, we use a computational simplification, where we replace the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ with the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$, whose computation is orders of magnitude faster than running the MCMC for the posterior mean.

The iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ are shown in Figure 6.3. By comparing these intervals with those of Figure 6.2, we conclude that there is little difference between using $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ to form the intervals. The main difference is that the intervals constructed using $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ appear to be somewhat more wiggly near the tails of f, which is also where the Gaussian approximation deviates most from the underlying Poisson distribution. Given the overall similarity of the two intervals, we expect the coverage performance of the intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ to be similar to those induced by $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$.

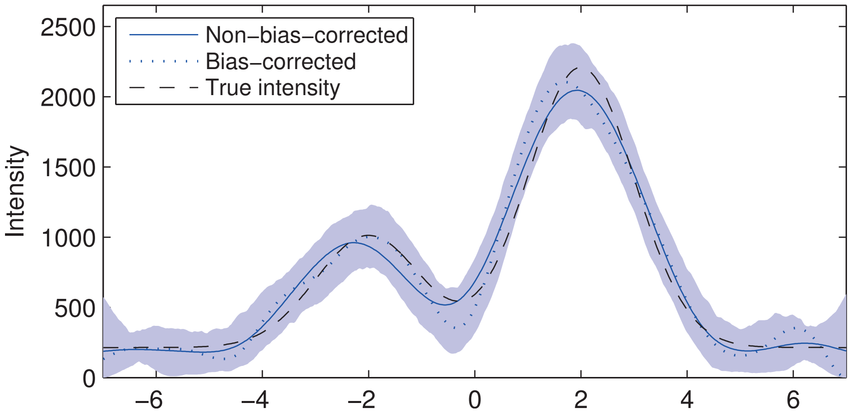
Figure 6.4(a) compares the empirical coverage of the iteratively bias-corrected percentile intervals based on $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ to the alternative methods in the case of the medium sample size. We find that the coverage of the bias-corrected intervals is close to the nominal value 95 % throughout the spectrum. The alternative bootstrap methods, on the other hand, suffer from severe undercoverage. The non-bias-corrected percentile intervals in particular fail to cover the truth at the larger peak at $s = 2$, where their empirical coverage is 0.327 (0.298, 0.357) (95 % Clopper–Pearson interval). The Bayesian constructions, on the other hand, overcover for most of the spectrum, but, in the area around $s = 2$, where the bias is the largest, also these intervals suffer from undercoverage. Interestingly, the coverage of the bias-corrected intervals does not seems to suffer from the fact that the regularization strength δ was kept fixed to $\hat{\delta}_{\mathrm{MMLE}}$ throughout the bootstrap computations.

To get further insight into the iterative bias-correction, we show in Figure 6.4(b) the empirical coverage of the bias-corrected intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ for various choices of the number of bias-correction iterations N_{BC}. Notice that for $N_{\mathrm{BC}} = 0$ these are simply the standard percentile intervals. For one bias-correction iteration, we already observe a major improvement in the

(a) Bias−corrected percentile intervals, posterior mean, λ_{tot} = 50000



(b) Bias−corrected percentile intervals, posterior mean, λ_{tot} = 10000



(c) Bias−corrected percentile intervals, posterior mean, λ_{tot} = 1000

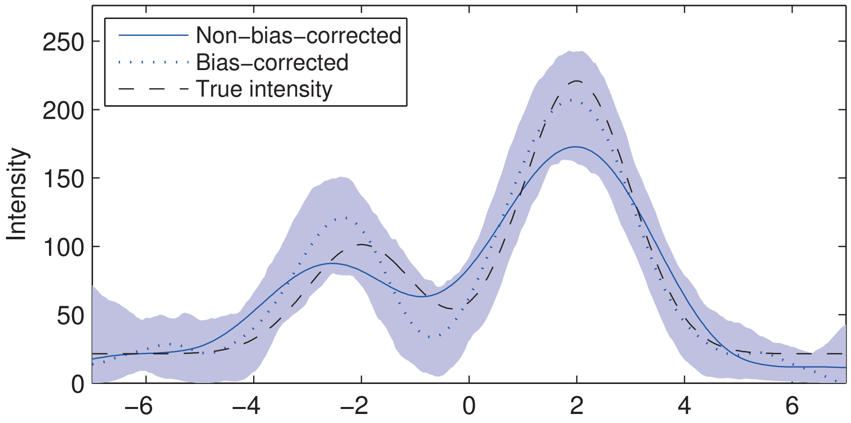
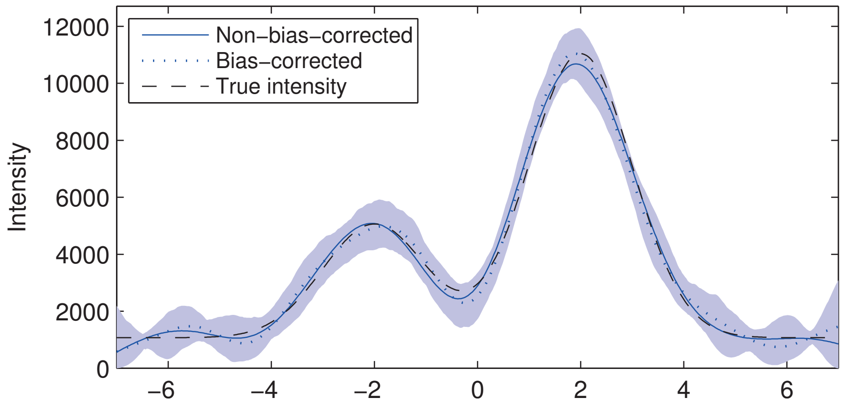
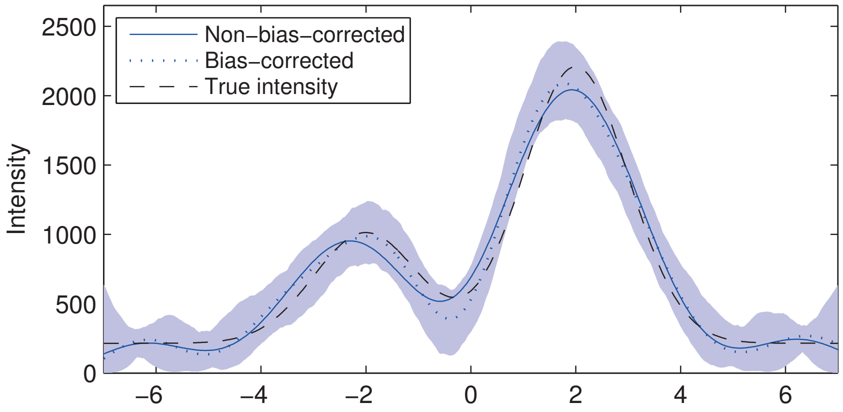


Figure 6.2: Iteratively bias-corrected percentile intervals induced by the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$, and the corresponding point estimates, in the two peaks on a uniform background test setup. The regularization strength is chosen using the MMLE and the number of bias-correction iterations is (a) 5, (b) 10 and (c) 15, depending on the sample size λ_{tot}. The intervals are formed for 95 % nominal pointwise coverage.

(a) Bias−corrected percentile intervals, Gaussian approx., λ_{tot} = 50000



(b) Bias−corrected percentile intervals, Gaussian approx., λ_{tot} = 10000



(c) Bias−corrected percentile intervals, Gaussian approx., λ_{tot} = 1000

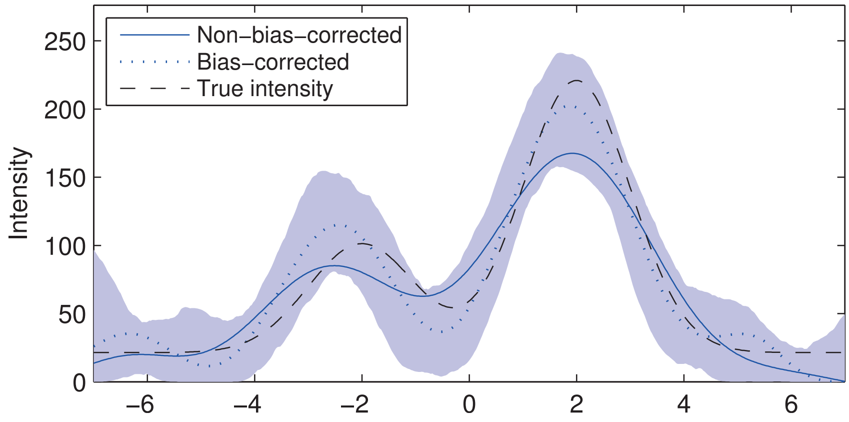


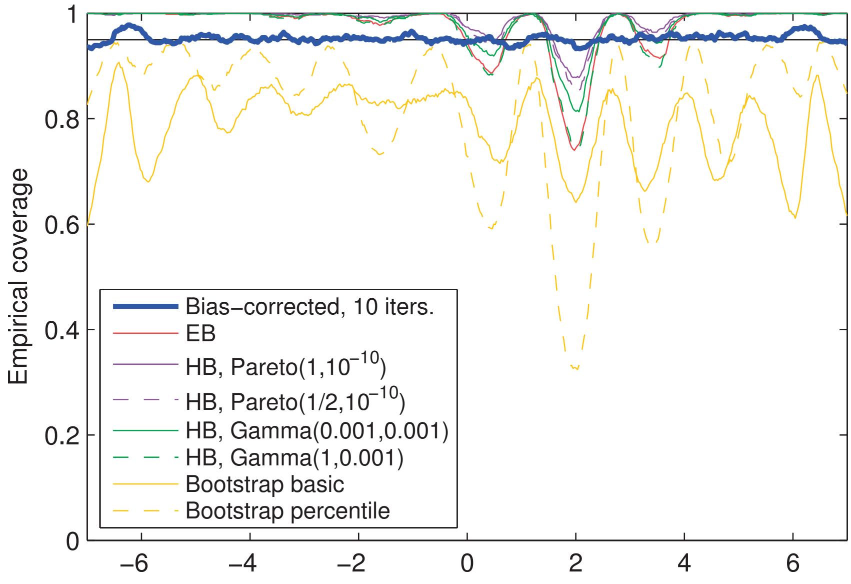
Figure 6.3: Iteratively bias-corrected percentile intervals induced by the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$, and the corresponding point estimates, in the two peaks on a uniform background test setup. The regularization strength is chosen using the MMLE and the number of bias-correction iterations is (a) 5, (b) 10 and (c) 15, depending on the sample size λ_{tot}. The intervals are formed for 95 % nominal pointwise coverage.

coverage, with more iterations further increasing the coverage. To obtain nearly nominal coverage, $N_{\mathrm{BC}} = 10$ or more iterations is needed in this particular case. The price to pay for this improved coverage is an increase in the length and wiggliness of the intervals. This is illustrated in Figure 6.5, where a realization of the bias-corrected intervals is shown for different amounts of bias-correction iterations. Crucially, we observe that, even with a fairly large number of bias-correction iterations, the interval length is only modestly increased, while, at the same time, there is a significant improvement in the coverage performance. For example, at $N_{\mathrm{BC}} = 10$, where the coverage is close to the nominal value, the bias-corrected intervals are only moderately longer than the standard percentile intervals, which suffer from significant undercoverage.

These effects become much more pronounced when the sample size is reduced. Figure 6.6 shows the empirical coverage of the different methods for $\lambda_{\mathrm{tot}} = 1\,000$. We observe that in this case both the Bayesian intervals and the standard bootstrap intervals suffer from drastic undercoverage. All these methods fail in particular at the larger peak at $s = 2$, where the empirical coverage of the conventional methods is less than 50 %, with non-bias-corrected percentile intervals in particular having close to zero coverage. The bias-corrected percentile intervals with $N_{\mathrm{BC}} = 15$ bias-correction iterations also have undercoverage at $s = 2$, but their empirical coverage 0.863 (0.840, 0.884) is still significantly better than that of the alternative methods. Moreover, with more bias-correction iterations, the coverage of the bias-corrected intervals would be further improved. With $N_{\mathrm{BC}} = 50$ iterations, the coverage would be close to the nominal value, but obviously the interval lengths will also be longer, although still fairly reasonable for this particular realization (see Figures B.7(b) and B.9 in the appendix).

Figure 6.7, which shows a realization of the different intervals for $\lambda_{\mathrm{tot}} = 1\,000$, helps us better understand the observed coverage performance of each interval type (the hierarchical Bayes intervals for the $\mathrm{Pareto}(1/2, 10^{-10})$ and $\mathrm{Gamma}(0.001, 0.001)$ hyperpriors, which are not shown in the figure, lie between the two hierarchical Bayes intervals that are shown). The percentile intervals, which only quantify the variability of $\hat{f}$, fail to cover since they completely ignore the sizeable bias in $\hat{f}$. The width of the basic intervals is similar to that of the percentile intervals, but they are shifted to account for the bias. However, the effect is not large enough to yield adequate coverage performance. The empirical Bayes credible intervals are wider than the percentile intervals and, as explained in Section 6.1.1, this extra width can be understood to partially accommodate the bias, but clearly even wider intervals would be needed to fully account for the bias. For hierarchical Bayes, the location and width of the intervals depends on the choice of the hyperprior. The weaker the regularization implied by the hyperprior, the better the coverage and the longer the intervals. However, even for the $\mathrm{Pareto}(1, 10^{-10})$ hyperprior, the intervals are not wide enough to fully accommodate the bias. Finally, the bias-corrected intervals account for the bias much better than the other methods, but are also longer and more wiggly than the alternative constructions. Out of the intervals shown, the bias-corrected intervals provide the most realistic measure of the unfolded uncertainty, while the other methods underestimate the uncertainty.

(a) Comparison of coverage performance, $\lambda_{\text{tot}} = 10000$



(b) Effect of bias−correction on coverage performance, $\lambda_{\text{tot}} = 10000$

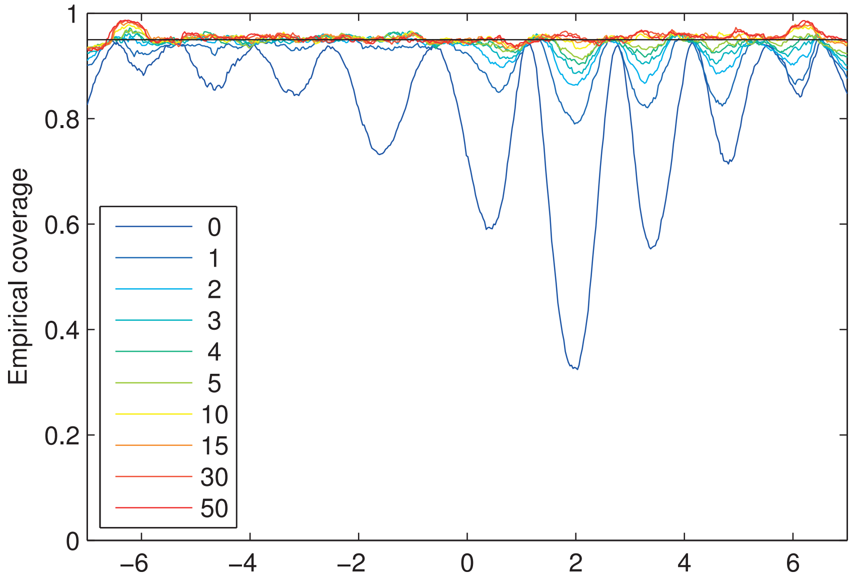


Figure 6.4: Coverage studies in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\text{tot}} = 10\,000$. Figure (a) compares the empirical coverage of the iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ to the alternative empirical Bayes (EB), hierarchical Bayes (HB) and bootstrap intervals. The number of bias-correction iterations is set to 10 and the regularization strength is chosen using the MMLE, expect for HB where four different uninformative hyperpriors are considered. Figure (b) shows the coverage of the bias-corrected intervals when the number of bias-correction iterations is varied between 0 and 50. All the intervals are for 95 % nominal pointwise coverage shown by the horizontal line.

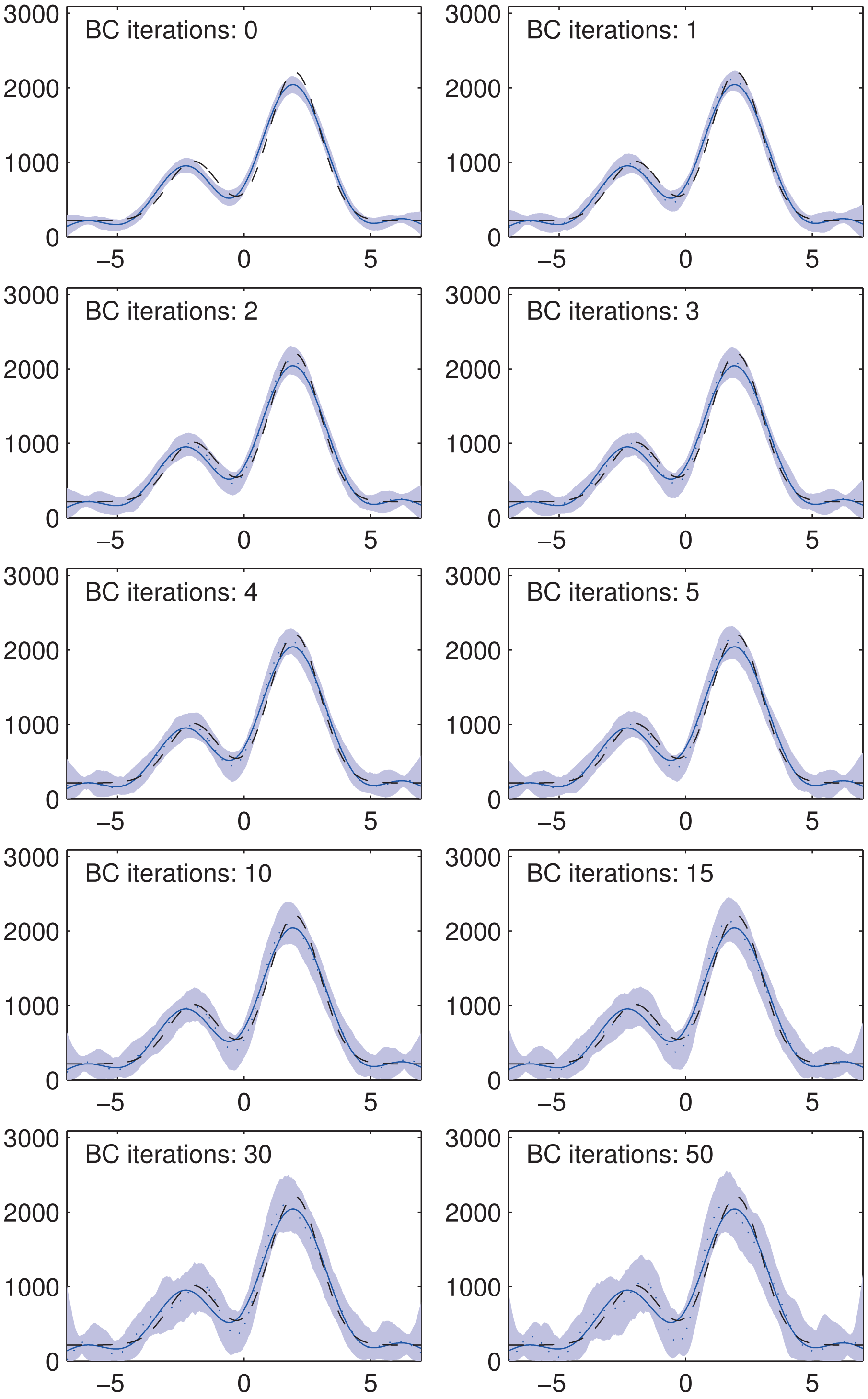


Figure 6.5: Iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ in the two peaks on a uniform background test setup with $\lambda_{\mathrm{tot}} = 10\,000$ as the number of bias-correction iterations is varied. The true intensity is shown by the dashed line, the point estimate corresponding to $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ by the solid line and the bias-corrected point estimate by the dotted line. The intervals have 95 % nominal pointwise coverage.

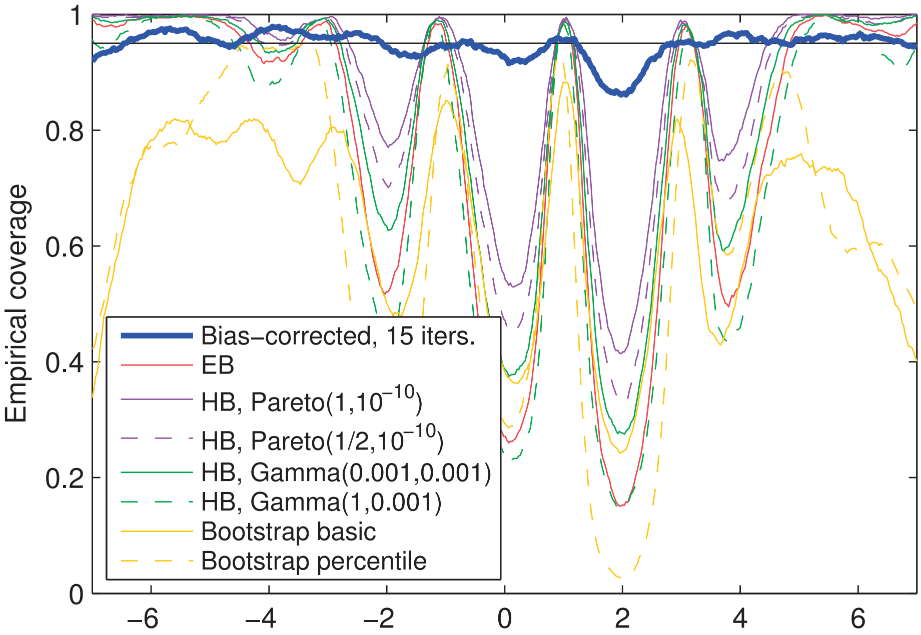


Figure 6.6: Same as Figure 6.4(a), but for sample size $\lambda_{\text{tot}} = 1\,000$ and $N_{\text{BC}} = 15$ bias-correction iterations.

The full results of the simulations carried out in this section are provided in Appendix B.1. In particular, the appendix shows that, in the large sample size case with $\lambda_{\text{tot}} = 50\,000$, the iteratively bias-corrected intervals again attain close-to-nominal coverage, while the conventional bootstrap intervals still undercover. The Bayesian intervals also continue to overcover, but in this case also their coverage at $s = 2$ is at or above the nominal value of 95 %.

6.4 Iterative bias-correction for linear estimators and Gaussian data

To gain further insight into iteratively bias-corrected uncertainty quantification, we study in this section an analytically more tractable simplification of the problem where the data $\boldsymbol{y}$ follow a Gaussian distribution and the estimator $\hat{\boldsymbol{\beta}}$ is a linear function of $\boldsymbol{y}$. In this case, we can write down in closed form both the bias-corrected estimators $\hat{\boldsymbol{\beta}}^{(t)}$ and the coverage probability of the associated Gaussian confidence intervals. This leads us to propose a data-driven way of choosing the number of bias-correction iterations N_{BC}. We also consider undersmoothing as an alternative to bias-correction and present some empirical evidence suggesting that bias-correction tends to provide shorter confidence intervals than undersmoothing.

6.4.1 Iterative bias-correction for linear estimators

In this section, we investigate the iterative bias-correction procedure described in Section 6.2.1 in the case where the estimator $\hat{\boldsymbol{\beta}}$ is a linear function of the data $\boldsymbol{y}$, that is, $\hat{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{y}$, for some $p \times n$ matrix $\boldsymbol{A}$. We also assume that $n \geq p$ and that the expectation of $\boldsymbol{y}$ is a linear function

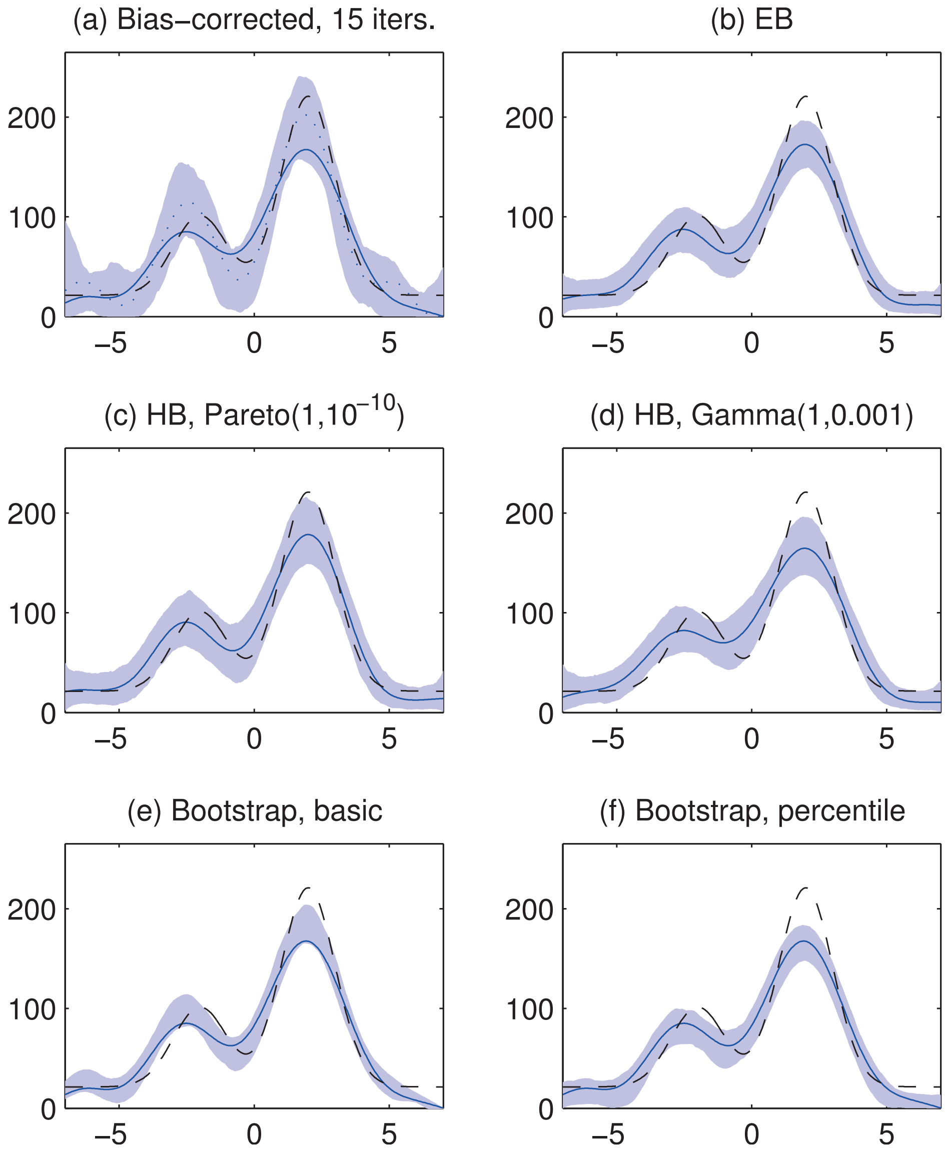


Figure 6.7: A single realization of (a) the iteratively bias-corrected percentile intervals with 15 bias-correction iterations, (b) the empirical Bayes (EB) credible intervals, (c)–(d) the hierarchical Bayes (HB) credible intervals for the two extremal hyperpriors, (e) the basic bootstrap intervals and (f) the standard bootstrap percentile intervals in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\mathrm{tot}} = 1\,000$. Intervals (a), (e) and (f) are based on the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$. Also shown are the corresponding point estimates $\hat{f}$ (solid lines) and the true intensity f (dashed lines). In Figure (a), also the bias-corrected point estimate $\hat{f}_{\mathrm{BC}}$ (dotted line) is shown. All the intervals are formed for 95 % nominal pointwise coverage.

of the unknown $\boldsymbol{\beta}$, that is, $\mathrm{E}_{\boldsymbol{\beta}}(\boldsymbol{y}) = \boldsymbol{K}\boldsymbol{\beta}$, for some $n \times p$ matrix $\boldsymbol{K}$. This is the case for example in the Poisson regression model (5.1) that pertains to unfolding. Besides the linearity of the expectation, we do not make further distributional assumptions regarding $\boldsymbol{y}$ in the present section.

The bias estimate on the tth step of the iteration is given by

$$\widehat{\mathrm{bias}}^{(t)}(\hat{\boldsymbol{\beta}}) = \mathrm{E}_{\hat{\boldsymbol{\beta}}^{(t)}}(\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\beta}}^{(t)}. \tag{6.25}$$

Under the stated assumptions, this becomes

$$\widehat{\mathrm{bias}}^{(t)}(\hat{\boldsymbol{\beta}}) = \boldsymbol{A}\mathrm{E}_{\hat{\boldsymbol{\beta}}^{(t)}}(\boldsymbol{y}) - \hat{\boldsymbol{\beta}}^{(t)} = \boldsymbol{A}\boldsymbol{K}\hat{\boldsymbol{\beta}}^{(t)} - \hat{\boldsymbol{\beta}}^{(t)} = (\boldsymbol{A}\boldsymbol{K} - \boldsymbol{I})\hat{\boldsymbol{\beta}}^{(t)}. \tag{6.26}$$

Notice that the bias estimate is available in closed form without the need for bootstrap resampling. It follows that the corresponding bias-corrected estimator is given by

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(0)} - \widehat{\mathrm{bias}}^{(t)}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}^{(0)} - (\boldsymbol{A}\boldsymbol{K} - \boldsymbol{I})\hat{\boldsymbol{\beta}}^{(t)}. \tag{6.27}$$

Let $\boldsymbol{J}^{(t)}$, $t = 0, 1, 2, \ldots$, be the sequence of $p \times p$ matrices defined by

$$\boldsymbol{J}^{(0)} = \boldsymbol{I}, \tag{6.28}$$

$$\boldsymbol{J}^{(t)} = \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{K})\boldsymbol{J}^{(t-1)}, \quad t = 1, 2, \ldots \tag{6.29}$$

We proceed to show that $\hat{\boldsymbol{\beta}}^{(t)} = \boldsymbol{J}^{(t)}\hat{\boldsymbol{\beta}}^{(0)}$ for every $t = 0, 1, 2, \ldots$ The claim obviously holds for $\hat{\boldsymbol{\beta}}^{(0)}$. Now assume that $\hat{\boldsymbol{\beta}}^{(t)} = \boldsymbol{J}^{(t)}\hat{\boldsymbol{\beta}}^{(0)}$. Then

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(0)} - (\boldsymbol{A}\boldsymbol{K} - \boldsymbol{I})\boldsymbol{J}^{(t)}\hat{\boldsymbol{\beta}}^{(0)} = \big(\boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{K})\boldsymbol{J}^{(t)}\big)\hat{\boldsymbol{\beta}}^{(0)} = \boldsymbol{J}^{(t+1)}\hat{\boldsymbol{\beta}}^{(0)}, \tag{6.30}$$

and the claim follows by induction. Dropping the superscript in $\hat{\boldsymbol{\beta}}^{(0)}$, we conclude that the bias-correction iteration is given by

$$\hat{\boldsymbol{\beta}}^{(t)} = \boldsymbol{J}^{(t)}\hat{\boldsymbol{\beta}}, \; t = 0, 1, 2, \ldots, \quad \text{with} \quad \boldsymbol{J}^{(0)} = \boldsymbol{I}, \quad \boldsymbol{J}^{(t)} = \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{K})\boldsymbol{J}^{(t-1)}, \tag{6.31}$$

where $\hat{\boldsymbol{\beta}}$ is the original non-bias-corrected point estimate of $\boldsymbol{\beta}$.

We have the following result:

Proposition 6.1. *Assume that* $\boldsymbol{A}\boldsymbol{K} \in \mathbb{R}^{p \times p}$ *is invertible and consider the iteration* $\boldsymbol{J}^{(t)} = T\big(\boldsymbol{J}^{(t-1)}\big)$, $t = 1, 2, \ldots$, *defined by Equation* (6.29). *Then it holds that:*

(a) $\boldsymbol{J} = (\boldsymbol{A}\boldsymbol{K})^{-1}$ *is the unique fixed point of the iteration;*

(b) If $\|\boldsymbol{I} - \boldsymbol{A}\boldsymbol{K}\| < 1$, *for some matrix norm* $\|\cdot\|$ *that satisfies* $\|\boldsymbol{M}\boldsymbol{N}\| \leq \|\boldsymbol{M}\|\|\boldsymbol{N}\|$, *then* $\boldsymbol{J}^{(t)} \xrightarrow{t\to\infty} \boldsymbol{J}$ *in the metric induced by* $\|\cdot\|$.

Proof.

(a) Any fixed point must satisfy $\boldsymbol{J} = T(\boldsymbol{J}) = \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{AK})\boldsymbol{J}$. That is, $\boldsymbol{AKJ} = \boldsymbol{I}$, whose unique solution is $\boldsymbol{J} = (\boldsymbol{AK})^{-1}$.

(b) For all $p \times p$ matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, we have

$$\begin{aligned} \|T(\boldsymbol{X}) - T(\boldsymbol{Y})\| &= \|\boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{AK})\boldsymbol{X} - \boldsymbol{I} - (\boldsymbol{I} - \boldsymbol{AK})\boldsymbol{Y}\| && (6.32)\\ &= \|(\boldsymbol{I} - \boldsymbol{AK})(\boldsymbol{X} - \boldsymbol{Y})\| \le \|\boldsymbol{I} - \boldsymbol{AK}\|\|\boldsymbol{X} - \boldsymbol{Y}\|. && (6.33) \end{aligned}$$

Hence, T is a contraction if $\|\boldsymbol{I} - \boldsymbol{AK}\| < 1$. The result then follows from the Banach fixed-point theorem (see, e.g., Theorem 5.1-2 and Corollary 5.1-3 in Kreyszig (1978)).

□

The bias-corrected point estimate $\hat{\boldsymbol{\beta}}_{\mathrm{BC}} = \boldsymbol{J}\hat{\boldsymbol{\beta}} = (\boldsymbol{AK})^{-1}\hat{\boldsymbol{\beta}} = (\boldsymbol{AK})^{-1}\boldsymbol{Ay}$ corresponding to the fixed point $\boldsymbol{J}$ is indeed unbiased:

$$\mathrm{E}_{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\beta}}_{\mathrm{BC}}\right) = (\boldsymbol{AK})^{-1}\boldsymbol{A}\mathrm{E}_{\boldsymbol{\beta}}(\boldsymbol{y}) = (\boldsymbol{AK})^{-1}\boldsymbol{AK\beta} = \boldsymbol{\beta}. \tag{6.34}$$

Hence, if $\|\boldsymbol{I} - \boldsymbol{AK}\| < 1$, the bias-correction iteration for linear estimators converges to the unbiased estimator $\hat{\boldsymbol{\beta}}_{\mathrm{BC}} = (\boldsymbol{AK})^{-1}\boldsymbol{Ay}$. If $\boldsymbol{A}$ and $\boldsymbol{K}$ are both invertible square matrices, then $\hat{\boldsymbol{\beta}}_{\mathrm{BC}} = \boldsymbol{K}^{-1}\boldsymbol{y}$.

The condition $\|\boldsymbol{I} - \boldsymbol{AK}\| < 1$ can be interpreted as a requirement that $\boldsymbol{A}$ be sufficiently close to a left inverse of $\boldsymbol{K}$. The condition is satisfied for example for Tikhonov regularization (Tikhonov, 1963; Phillips, 1962) and ridge regression (Hoerl and Kennard, 1970), with

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + 2\delta\boldsymbol{I}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{y}. \tag{6.35}$$

To see this, let $\boldsymbol{K} = \boldsymbol{USV}^{\mathrm{T}}$ be the singular value decomposition of $\boldsymbol{K}$ with $\boldsymbol{S} = \begin{bmatrix}\tilde{\boldsymbol{S}}\\ \boldsymbol{0}\end{bmatrix} \in \mathbb{R}^{n\times p}$, where $\tilde{\boldsymbol{S}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{p\times p}$ contains the singular values $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_p \ge 0$ of $\boldsymbol{K}$. Then

$$\begin{aligned} \boldsymbol{AK} &= \left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + 2\delta\boldsymbol{I}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} && (6.36)\\ &= \left(\boldsymbol{V}\tilde{\boldsymbol{S}}^2\boldsymbol{V}^{\mathrm{T}} + 2\delta\boldsymbol{I}\right)^{-1}\boldsymbol{V}\tilde{\boldsymbol{S}}^2\boldsymbol{V}^{\mathrm{T}} && (6.37)\\ &= \boldsymbol{V}\left(\tilde{\boldsymbol{S}}^2 + 2\delta\boldsymbol{I}\right)^{-1}\tilde{\boldsymbol{S}}^2\boldsymbol{V}^{\mathrm{T}} && (6.38) \end{aligned}$$

and hence

$$\boldsymbol{I} - \boldsymbol{AK} = \boldsymbol{V}\left(\boldsymbol{I} - \left(\tilde{\boldsymbol{S}}^2 + 2\delta\boldsymbol{I}\right)^{-1}\tilde{\boldsymbol{S}}^2\right)\boldsymbol{V}^{\mathrm{T}}. \tag{6.39}$$

As a result, the singular values of $\boldsymbol{I} - \boldsymbol{AK}$ are given by $1 - \frac{\sigma_i^2}{\sigma_i^2 + 2\delta} = \frac{2\delta}{\sigma_i^2 + 2\delta}$, $i = 1, \ldots, p$, and the

2-norm of $\boldsymbol{I}-\boldsymbol{AK}$ is

$$\|\boldsymbol{I}-\boldsymbol{AK}\|_2 = \sigma_{\max}(\boldsymbol{I}-\boldsymbol{AK}) = \frac{2\delta}{\sigma_p^2+2\delta} \leq 1, \tag{6.40}$$

where the inequality is strict if $\sigma_p > 0$, that is, if $\boldsymbol{K}$ has full rank. Proposition 6.1 then shows that the bias-correction iteration converges to the estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{BC}} = (\boldsymbol{AK})^{-1}\boldsymbol{Ay} = \left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K}\right)^{-1}\left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K}+2\delta\boldsymbol{I}\right)\left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K}+2\delta\boldsymbol{I}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{y} = \left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{y}, \tag{6.41}$$

provided that $\boldsymbol{K}$ has full rank. Notice that this is simply the least-squares estimator of $\boldsymbol{\beta}$.

6.4.2 Coverage for Gaussian observations

When the observations $\boldsymbol{y}$ follow a Gaussian distribution and the estimator $\hat{\boldsymbol{\beta}}$ is a linear function of $\boldsymbol{y}$, we can write down in closed form the coverage probability of the Gaussian confidence intervals described in Section 6.1.3. This is given by the following result:

Proposition 6.2. *Assume $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ is a known covariance matrix and $\boldsymbol{K} \in \mathbb{R}^{n\times p}$. Let $\hat{\boldsymbol{\beta}} = \boldsymbol{Ay}$ with $\boldsymbol{A} \in \mathbb{R}^{p\times n}$ be a linear estimator of $\boldsymbol{\beta}$ and let $\hat{\theta} = \boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$ be the corresponding estimator of the quantity of interest $\theta = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$. Then the confidence interval*

$$[\underline{\theta}, \overline{\theta}] = \left[\hat{\theta} - z_{1-\alpha/2}\sqrt{\mathrm{var}\left(\hat{\theta}\right)}, \hat{\theta} + z_{1-\alpha/2}\sqrt{\mathrm{var}\left(\hat{\theta}\right)}\right] \tag{6.42}$$

$$= \left[\hat{\theta} - z_{1-\alpha/2}\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A\Sigma A}^{\mathrm{T}}\boldsymbol{c}}, \hat{\theta} + z_{1-\alpha/2}\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A\Sigma A}^{\mathrm{T}}\boldsymbol{c}}\right] \tag{6.43}$$

has coverage probability

$$\mathrm{P}_{\boldsymbol{\beta}}\left(\underline{\theta} \leq \theta \leq \overline{\theta}\right) = \Phi\left(\frac{\mathrm{bias}\left(\hat{\theta}\right)}{\mathrm{SE}\left(\hat{\theta}\right)} + z_{1-\alpha/2}\right) - \Phi\left(\frac{\mathrm{bias}\left(\hat{\theta}\right)}{\mathrm{SE}\left(\hat{\theta}\right)} + z_{\alpha/2}\right), \tag{6.44}$$

where $\mathrm{bias}\left(\hat{\theta}\right) = \mathrm{E}_{\boldsymbol{\beta}}\left(\hat{\theta}\right) - \theta = \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{AK}-\boldsymbol{I})\boldsymbol{\beta}$ *is the bias of* $\hat{\theta}$, $\mathrm{SE}\left(\hat{\theta}\right) = \sqrt{\mathrm{var}\left(\hat{\theta}\right)} = \sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A\Sigma A}^{\mathrm{T}}\boldsymbol{c}}$ *is the standard error of $\hat{\theta}$ and Φ is the standard normal cumulative distribution function.*

Proof. The sampling distribution of the estimator $\hat{\theta} = \boldsymbol{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}} = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Ay} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c})^{\mathrm{T}}\boldsymbol{y}$ is given by

$$\hat{\theta} \sim \mathcal{N}\left((\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c})^{\mathrm{T}}\boldsymbol{K\beta}, (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c})^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c})\right) = \mathcal{N}\left(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{AK\beta}, \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A\Sigma A}^{\mathrm{T}}\boldsymbol{c}\right). \tag{6.45}$$

Hence the bias and the standard error of $\hat{\theta}$ are given by

$$\mathrm{bias}\left(\hat{\theta}\right) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{AK\beta} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta} = \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{AK}-\boldsymbol{I})\boldsymbol{\beta}, \quad \mathrm{SE}\left(\hat{\theta}\right) = \sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A\Sigma A}^{\mathrm{T}}\boldsymbol{c}}. \tag{6.46}$$

The coverage probability is

$$\mathrm{P}_{\boldsymbol{\beta}}\left(\theta \in [\underline{\theta}, \overline{\theta}]\right) = 1 - \mathrm{P}_{\boldsymbol{\beta}}\left(\theta \notin [\underline{\theta}, \overline{\theta}]\right) = 1 - \mathrm{P}_{\boldsymbol{\beta}}\left(\theta \leq \underline{\theta}\right) - \mathrm{P}_{\boldsymbol{\beta}}\left(\theta \geq \overline{\theta}\right). \tag{6.47}$$

Here the latter term is given by

$$\begin{aligned}
\mathrm{P}_{\boldsymbol{\beta}}\left(\theta \geq \overline{\theta}\right) &= \mathrm{P}_{\boldsymbol{\beta}}\left(\theta \geq \hat{\theta} + z_{1-\alpha/2}\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}\right) && (6.48)\\
&= \mathrm{P}_{\boldsymbol{\beta}}\left(\frac{\hat{\theta} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{K}\boldsymbol{\beta}}{\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}} \leq \frac{\theta - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{K}\boldsymbol{\beta}}{\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}} - z_{1-\alpha/2}\right) && (6.49)\\
&= \mathrm{P}_{\boldsymbol{\beta}}\left(\frac{\hat{\theta} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{K}\boldsymbol{\beta}}{\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}} \leq -\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} - z_{1-\alpha/2}\right) && (6.50)\\
&= \Phi\left(-\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} - z_{1-\alpha/2}\right), && (6.51)
\end{aligned}$$

where we have used the fact that $\frac{\hat{\theta} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{K}\boldsymbol{\beta}}{\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}} \sim \mathcal{N}(0,1)$.

An analogous calculation shows that

$$\mathrm{P}_{\boldsymbol{\beta}}\left(\theta \leq \underline{\theta}\right) = \mathrm{P}_{\boldsymbol{\beta}}\left(\theta \leq \hat{\theta} - z_{1-\alpha/2}\sqrt{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}\right) = 1 - \Phi\left(-\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} - z_{\alpha/2}\right). \tag{6.52}$$

Hence

$$\begin{aligned}
\mathrm{P}_{\boldsymbol{\beta}}\left(\theta \in [\underline{\theta}, \overline{\theta}]\right) &= \Phi\left(-\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} - z_{\alpha/2}\right) - \Phi\left(-\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} - z_{1-\alpha/2}\right) && (6.53)\\
&= \Phi\left(\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} + z_{1-\alpha/2}\right) - \Phi\left(\frac{\mathrm{bias}(\hat{\theta})}{\mathrm{SE}(\hat{\theta})} + z_{\alpha/2}\right). && (6.54)
\end{aligned}$$

□

Notice that this result holds for any linear point estimator $\hat{\boldsymbol{\beta}}$, including the iterative bias-correction described in Section 6.4.1. In our case, the quantity of interest $\theta = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$ is the point evaluator for a B-spline, $\theta = f(s) = \sum_{j=1}^{p} \beta_j B_j(s) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{\beta}$, with $\boldsymbol{c} = \left[B_1(s), \ldots, B_p(s)\right]^{\mathrm{T}}$.

We see from Equation (6.44) that the coverage is a function of the ratio of the bias and the standard error of $\hat{\theta}$. Denoting this ratio by $\gamma = \mathrm{bias}(\hat{\theta})/\mathrm{SE}(\hat{\theta})$, the coverage is given by

$$C(\gamma) = \Phi\left(\gamma + z_{1-\alpha/2}\right) - \Phi\left(\gamma + z_{\alpha/2}\right). \tag{6.55}$$

For $\gamma = 0$, that is, for zero bias, we obviously obtain the nominal coverage,

$$C(0) = \Phi\left(z_{1-\alpha/2}\right) - \Phi\left(z_{\alpha/2}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha. \tag{6.56}$$

The coverage is symmetric around $\gamma = 0$,

$$\begin{aligned}
C(-\gamma) &= \Phi\left(-\gamma + z_{1-\alpha/2}\right) - \Phi\left(-\gamma + z_{\alpha/2}\right) && (6.57)\\
&= \Phi\left(-\gamma - z_{\alpha/2}\right) - \Phi\left(-\gamma - z_{1-\alpha/2}\right) && (6.58)\\
&= \Phi\left(\gamma + z_{1-\alpha/2}\right) - \Phi\left(\gamma + z_{\alpha/2}\right) = C(\gamma). && (6.59)
\end{aligned}$$

This shows that there is no difference in the coverage performance between positive and negative biases.

The derivative of $C(\gamma)$ is given by

$$C'(\gamma) = \Phi'\left(\gamma + z_{1-\alpha/2}\right) - \Phi'\left(\gamma + z_{\alpha/2}\right) = \phi\left(\gamma + z_{1-\alpha/2}\right) - \phi\left(\gamma + z_{\alpha/2}\right), \tag{6.60}$$

where ϕ is the standard normal probability density function. By the symmetry of ϕ, the sign of $C'(\gamma)$ is given by

$$C'(\gamma) \begin{cases} > 0, & \text{if } \gamma < 0, \\ = 0, & \text{if } \gamma = 0, \\ < 0, & \text{if } \gamma > 0. \end{cases} \tag{6.61}$$

Since $C(\gamma)$ is continuous, this shows that the maximum coverage is attained for $\gamma = 0$ and that the coverage reduces monotonically from the nominal value $1-\alpha$ as $|\gamma|$ grows. That is, for any non-zero amount of bias, the coverage will be strictly less than $1-\alpha$. Our key empirical observation is that in ill-posed inverse problems there are biased estimators $\hat{\boldsymbol{\beta}}$ such that the coverage of the interval (6.43) is only slightly below $1-\alpha$, but the interval length is still orders of magnitude shorter than that of an unbiased estimator.

6.4.3 Data-driven iteratively bias-corrected intervals

We have seen in Section 6.4.1 that for linear estimators $\hat{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{y}$ the bias-correction iteration is given by $\hat{\boldsymbol{\beta}}^{(t)} = \boldsymbol{J}^{(t)}\hat{\boldsymbol{\beta}} = \boldsymbol{J}^{(t)}\boldsymbol{A}\boldsymbol{y}$, with $\boldsymbol{J}^{(t)}$ given by Equation (6.31). Since $\hat{\boldsymbol{\beta}}^{(t)}$ is linear in $\boldsymbol{y}$, the coverage of the Gaussian intervals (6.43) induced by $\hat{\boldsymbol{\beta}}^{(t)}$ is given by Equation (6.44), provided that $\boldsymbol{y}$ is Gaussian. We also know by Section 6.4.2 that in order to obtain nominal $1-\alpha$ coverage we would need to run the iteration until convergence to an unbiased estimator, which would inflate the interval length. At the same time, we have seen in Section 6.3 that, as long as the iteration is stopped sufficiently early, the iteratively bias-corrected intervals can yield coverage which is close to the nominal value while maintaining reasonable interval length.

Putting these observations together, we propose the following strategy for choosing the number of bias-correction iterations N_{BC}:

> At each iteration t, use Equation (6.44) to estimate for all quantities of interest $\theta_i = \boldsymbol{c}_i^{\mathrm{T}}\boldsymbol{\beta}$ the coverage $C_i^{(t)}$ of the Gaussian intervals (6.43) induced by $\hat{\boldsymbol{\beta}}^{(t)}$ and stop the iteration once the smallest estimated coverage $\hat{C}^{(t)} = \min_i \hat{C}_i^{(t)}$ is greater than $1-\alpha-\varepsilon$, for some positive tolerance ε.

In our case, the quantities of interest $\{\theta_i\}$ are the point evaluators $\theta_i = f(s_i) = \sum_{j=1}^{p} \beta_j B_j(s_i) = \boldsymbol{c}_i^{\mathrm{T}}\boldsymbol{\beta}$, with $\boldsymbol{c}_i = \left[B_1(s_i), \ldots, B_p(s_i)\right]^{\mathrm{T}}$, on a fine grid of values $\{s_i\}_{i=1}^{m+1}$ in the true space.

The proposed criterion means that we concede that we are not able to obtain exactly the nominal coverage $1-\alpha$ using the intervals (6.43), but we instead try to guarantee that, for all practical purposes, the coverage is close enough to $1-\alpha$. For example, if 95 % intervals have in reality 94 % coverage, this would be considered satisfactory in most situations.

Notice also that the criterion does not necessarily lead to undercoverage. For example, if the aim is to obtain intervals with 95 % coverage, one could use 99 % intervals with $\varepsilon = 0.04$ to obtain 95 % intervals, provided that one can reliably estimate the coverage probability.

One could also argue that the proposed criterion simply turns the problem of choosing N_{BC} into the problem of choosing ε. While this is certainly a valid argument, it should be emphasized that ε gives a direct handle on the coverage probability of the intervals. After fixing ε, one can expect to obtain confidence intervals with coverage close to $1-\alpha-\varepsilon$, while fixing N_{BC} does not have such a clear meaning.

The main challenge with the proposed approach concerns the fact that the coverage probability (6.44) depends on

$$\mathrm{bias}\big(\hat{\theta}_i^{(t)}\big) = \boldsymbol{c}_i^{\mathrm{T}}(\boldsymbol{J}^{(t)}\boldsymbol{A}\boldsymbol{K}-\boldsymbol{I})\boldsymbol{\beta}, \tag{6.62}$$

where $\hat{\theta}_i^{(t)} = \boldsymbol{c}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}^{(t)}$ is the estimator of $\theta_i = \boldsymbol{c}_i^{\mathrm{T}}\boldsymbol{\beta}$ induced by $\hat{\boldsymbol{\beta}}^{(t)}$. This bias is obviously unknown and needs to be estimated. A first idea would be to simply estimate the bias by plugging in the non-bias-corrected estimate $\hat{\boldsymbol{\beta}}^{(0)}$. However, as we have seen, this tends to underestimate the bias. As a result, the coverage is overestimated, the bias-correction iteration is stopped too early and the actual coverage is less than $1-\alpha-\varepsilon$. A second idea would be to plug in the current bias-corrected estimate $\hat{\boldsymbol{\beta}}^{(t)}$ and to estimate the bias of $\hat{\theta}_i^{(t)}$ using

$$\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(t)}\big) = \boldsymbol{c}_i^{\mathrm{T}}(\boldsymbol{J}^{(t)}\boldsymbol{A}\boldsymbol{K}-\boldsymbol{I})\hat{\boldsymbol{\beta}}^{(t)}. \tag{6.63}$$

The problem with this approach is that, as the bias-correction iteration proceeds, the bias estimates and hence the coverage estimates become increasingly noisy. The problem is aggravated by the fact that we take a minimum of the estimated coverages over the grid $\{s_i\}_{i=1}^{m+1}$. Indeed, the estimated coverages based on (6.63) often increase for the first few bias-correction iterations and then suddenly start to decrease, while the actual coverage continues to improve. This leads us to propose the following procedure: We estimate the coverage based on (6.63) and check if the estimated coverage increased in comparison to the previous iteration. If the estimated coverage decreased, we conclude that the bias estimate has become too noisy to be reliable and estimate the bias by plugging in the previous estimate of $\boldsymbol{\beta}$ instead. Denoting this previous estimate by $\hat{\boldsymbol{\beta}}^{(T)}$, we then estimate the bias from that point onwards using

$$\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(t)}\big) = \boldsymbol{c}_i^{\mathrm{T}}(\boldsymbol{J}^{(t)}\boldsymbol{A}\boldsymbol{K}-\boldsymbol{I})\hat{\boldsymbol{\beta}}^{(T)}, \quad t > T, \tag{6.64}$$

keeping the estimate of $\boldsymbol{\beta}$ fixed to $\hat{\boldsymbol{\beta}}^{(T)}$.

The resulting procedure for choosing the number of bias-correction iterations N_{BC} is summarized in Algorithm 4. The data-driven bias-corrected intervals with target coverage $1-\alpha-\varepsilon$ are then those induced by $\hat{\boldsymbol{\beta}}_{\mathrm{BC}} = \hat{\boldsymbol{\beta}}^{(\hat{N}_{\mathrm{BC}})}$, where $\hat{N}_{\mathrm{BC}}$ is the estimated number of bias-correction iterations. Section 6.4.5 demonstrates that this approach performs reasonably well in practice.

6.4.4 Data-driven undersmoothed intervals

An alternative approach for addressing the issue of bias in nonparametric confidence intervals is to employ *undersmoothing* (Hall, 1992). Let $\hat{\boldsymbol{\beta}}_{\delta}$ be a regularized point estimator of $\boldsymbol{\beta}$ depending on regularization strength δ. Undersmoothing consists in debiasing the estimator by choosing δ to be smaller than the value that leads to optimal point estimation performance. As in the case of iterative bias-correction, the confidence intervals induced by the variability of the undersmoothed point estimator $\hat{\boldsymbol{\beta}}_{\mathrm{US}}$ will have improved coverage at the cost of increased interval length.

Although undersmoothing has been extensively studied from a theoretical perspective, there exist few, if any, data-driven ways of choosing the amount of undersmoothing (Hall and Horowitz, 2013). Notice, however, that the data-driven procedure we introduced in Section 6.4.3 can also be adapted to undersmoothing. Namely, let $\hat{\boldsymbol{\beta}}_{\delta} = \boldsymbol{A}_{\delta}\boldsymbol{y}$ be a linear estimator of $\boldsymbol{\beta}$, where the matrix $\boldsymbol{A}_{\delta}$ depends on the regularization strength δ. Then Proposition 6.2 can be used to write down for each δ the coverage probability of the Gaussian confidence intervals (6.43) induced by $\hat{\boldsymbol{\beta}}_{\delta}$. We can then estimate this coverage probability as explained in Section 6.4.3 and choose the largest δ for which all estimated coverage probabilities are greater than $1-\alpha-\varepsilon$ for some tolerance $\varepsilon > 0$. Denote this value by $\hat{\delta}_{\mathrm{US}}$. The data-driven undersmoothed confidence intervals are then those induced by $\hat{\boldsymbol{\beta}}_{\mathrm{US}} = \hat{\boldsymbol{\beta}}_{\hat{\delta}_{\mathrm{US}}}$.

We compare in Section 6.4.5 the performance of undersmoothing and iterative bias-correction in a number of deconvolution scenarios. We find that, at least in the situations that we have investigated, iterative bias-correction provides better performance in the sense that, for a given coverage probability, the iteratively bias-corrected intervals are shorter than the undersmoothed intervals. Interestingly, our conclusions are opposite to those of Hall (1992), who recommends undersmoothing instead of bias-correction in the context of kernel density estimation. Due to a number of differences between our setup and that of Hall (1992), it is rather difficult to pinpoint the exact reason for our differing findings, but one likely explanation is that the intervals used by Hall (1992) do not account for the variability of the bias estimate, while our intervals take this effect into account.

Algorithm 4 Data-driven choice of the number of bias-correction iterations

```
Input:
    $\boldsymbol{y}$ — Smeared data following the $\mathcal{N}(\boldsymbol{K\beta}, \boldsymbol{\Sigma})$ distribution
    $\boldsymbol{A}$ — The mapping from the data $\boldsymbol{y}$ to the non-bias-corrected estimator $\hat{\boldsymbol{\beta}}$, i.e., $\hat{\boldsymbol{\beta}} = \boldsymbol{Ay}$
    $1 - \alpha \in (0,1)$ — Nominal coverage
    $\varepsilon > 0$ — Tolerance
Output:
    $\hat{N}_{\mathrm{BC}}$ — Estimated number of bias-correction iterations for target coverage $1 - \alpha - \varepsilon$

    Set $\hat{\boldsymbol{\beta}}^{(0)} = \boldsymbol{Ay}$;
    Estimate the minimum coverage
        $$\hat{C}^{(0)} = \min_i \left[ \Phi\left( \frac{\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(0)}\big)}{\mathrm{SE}\big(\hat{\theta}_i^{(0)}\big)} + z_{1-\alpha/2} \right) - \Phi\left( \frac{\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(0)}\big)}{\mathrm{SE}\big(\hat{\theta}_i^{(0)}\big)} + z_{\alpha/2} \right) \right],$$
    where $\mathrm{SE}\big(\hat{\theta}_i^{(0)}\big) = \sqrt{\boldsymbol{c}_i^{\mathrm{T}} \boldsymbol{A\Sigma A}^{\mathrm{T}} \boldsymbol{c}_i}$ and $\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(0)}\big) = \boldsymbol{c}_i^{\mathrm{T}} (\boldsymbol{AK} - \boldsymbol{I}) \hat{\boldsymbol{\beta}}^{(0)}$;
    Set $\boldsymbol{J}^{(0)} = \boldsymbol{I}$;
    Set $t = 0$;
    Set FIXED = FALSE;
    while $\hat{C}^{(t)} < 1 - \alpha - \varepsilon$ do        ▷ Iterate until estimated coverage larger than $1 - \alpha - \varepsilon$
        Set $t \leftarrow t + 1$;
        Set $\boldsymbol{J}^{(t)} = \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{AK}) \boldsymbol{J}^{(t-1)}$;
        Set $\hat{\boldsymbol{\beta}}^{(t)} = \boldsymbol{J}^{(t)} \boldsymbol{Ay}$;
        Estimate the minimum coverage
            $$\hat{C}^{(t)} = \min_i \left[ \Phi\left( \frac{\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(t)}\big)}{\mathrm{SE}\big(\hat{\theta}_i^{(t)}\big)} + z_{1-\alpha/2} \right) - \Phi\left( \frac{\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(t)}\big)}{\mathrm{SE}\big(\hat{\theta}_i^{(t)}\big)} + z_{\alpha/2} \right) \right],$$
        where $\mathrm{SE}\big(\hat{\theta}_i^{(t)}\big) = \sqrt{\boldsymbol{c}_i^{\mathrm{T}} \boldsymbol{J}^{(t)} \boldsymbol{A\Sigma} \big(\boldsymbol{J}^{(t)} \boldsymbol{A}\big)^{\mathrm{T}} \boldsymbol{c}_i}$ and
            $$\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(t)}\big) = \begin{cases} \boldsymbol{c}_i^{\mathrm{T}} (\boldsymbol{J}^{(t)} \boldsymbol{AK} - \boldsymbol{I}) \hat{\boldsymbol{\beta}}^{(t)}, & \text{if FIXED = FALSE}, \\ \boldsymbol{c}_i^{\mathrm{T}} (\boldsymbol{J}^{(t)} \boldsymbol{AK} - \boldsymbol{I}) \hat{\boldsymbol{\beta}}^{(T)}, & \text{if FIXED = TRUE}; \end{cases}$$
        if $\hat{C}^{(t)} < \hat{C}^{(t-1)}$ and FIXED = FALSE then        ▷ Estimated coverage decreased
            Set FIXED = TRUE;
            Set $\hat{\boldsymbol{\beta}}^{(T)} = \hat{\boldsymbol{\beta}}^{(t-1)}$;
            Re-estimate the minimum coverage
                $$\hat{C}^{(t)} = \min_i \left[ \Phi\left( \frac{\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(t)}\big)}{\mathrm{SE}\big(\hat{\theta}_i^{(t)}\big)} + z_{1-\alpha/2} \right) - \Phi\left( \frac{\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(t)}\big)}{\mathrm{SE}\big(\hat{\theta}_i^{(t)}\big)} + z_{\alpha/2} \right) \right],$$
            where $\mathrm{SE}\big(\hat{\theta}_i^{(t)}\big) = \sqrt{\boldsymbol{c}_i^{\mathrm{T}} \boldsymbol{J}^{(t)} \boldsymbol{A\Sigma} \big(\boldsymbol{J}^{(t)} \boldsymbol{A}\big)^{\mathrm{T}} \boldsymbol{c}_i}$ and $\widehat{\mathrm{bias}}\big(\hat{\theta}_i^{(t)}\big) = \boldsymbol{c}_i^{\mathrm{T}} (\boldsymbol{J}^{(t)} \boldsymbol{AK} - \boldsymbol{I}) \hat{\boldsymbol{\beta}}^{(T)}$;
        end if
    end while
    return $\hat{N}_{\mathrm{BC}} = t$;
```

6.4.5 Simulation study

6.4.5.1 Experiment setup

We demonstrate the ideas presented in Sections 6.4.1–6.4.4 using a deconvolution setup, where we have n independent observations from the model

$$y_i = g(t_i) + e_i, \quad e_i \sim \mathcal{N}(0, \sigma^2), \tag{6.65}$$

where

$$g(t) = \int_E k(t,s) f(s)\,\mathrm{d}s, \quad t \in F, \tag{6.66}$$

with $k(t,s) = k(t-s) = \mathcal{N}(t-s|0,1)$. We set $E = F = [-7,7]$ and let the observation points t_i consist of a uniform grid of $n = 40$ points. We also assume that f vanishes on the boundary of E, that is, $f(-7) = f(7) = 0$, and that f can be can be represented using a basis of $p = 30$ cubic B-splines, $f(s) = \sum_{j=1}^{p} \beta_j B_j(s)$. A basis that imposes the boundary constraints can be obtained by constructing the usual B-spline basis for $L = 28$ uniformly placed interior knots and dropping the first and the last basis functions. We then have that

$$g(t_i) = \int_E k(t_i,s) f(s)\,\mathrm{d}s = \sum_{j=1}^{p} \beta_j \int_E k(t_i,s) B_j(s)\,\mathrm{d}s = \sum_{j=1}^{p} K_{i,j}\beta_j, \tag{6.67}$$

with $K_{i,j} = \int_E k(t_i,s) B_j(s)\,\mathrm{d}s$. This leads to the model

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{K\beta}, \sigma^2 \boldsymbol{I}), \tag{6.68}$$

where the noise level $\sigma > 0$ is assumed to be known.

We consider the noise levels $\sigma = 0.001$, 0.005 and 0.025 and the following two true functions f:

$$f_1(s) = \mathcal{N}(s|0,1); \tag{6.69}$$

$$f_2(s) = \pi_1 \mathcal{N}(s|-2,1) + \pi_2 \mathcal{N}(s|2,1), \tag{6.70}$$

with $\pi_1 = 0.3$ and $\pi_2 = 0.7$. The first consists of a single Gaussian peak and the second is a Gaussian mixture model with two peaks. Notice that f_2 is similar to the test setup of Section 3.4.1, but without the uniform background component. These functions are not strictly speaking zero at $s = -7$ and $s = 7$, but for all practical purposes we can assume this to be the case and restrict our attention to estimators that satisfy these boundary constraints.

This setup simplifies the actual unfolding problem in the following ways:

1. The observations $\boldsymbol{y}$ follow a Gaussian distribution instead of a Poisson distribution, which significantly simplifies the problem as most formulas that were not available for Poisson noise can now be written down in closed form;

2. The variance of $\boldsymbol{y}$ is assumed to be homoscedastic and known, which simplifies the construction of both point and interval estimators;
3. There is no positivity constraint, which further simplifies the construction and computation of the estimators;
4. The value of f at the boundaries is assumed to be known, which leads to a trivial boundary condition;
5. The forward mapping is not integrated over bins in the smeared space.

We later demonstrate in Section 6.5 that, despite these simplifications, the conclusions of the present section carry over to the full unfolding problem.

6.4.5.2 Estimators

As in Section 5.1, we regularize the problem using the smoothness prior

$$p(\boldsymbol{\beta}|\delta) \propto \exp\left(-\delta\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\beta}\right), \quad \boldsymbol{\beta} \in \mathbb{R}^p, \quad \delta > 0, \tag{6.71}$$

where $\Omega_{i,j} = \int_E B_i''(s)B_j''(s)\,\mathrm{d}s$. When the boundary values are fixed, $\boldsymbol{\Omega}$ has full rank p and there is no need to use the augmented form (5.11). It is convenient to reparameterize the problem in terms of $\gamma = 2\delta\sigma^2$. The prior then becomes $p(\boldsymbol{\beta}|\gamma) \sim N\left(\boldsymbol{0}, \left(\frac{\gamma}{\sigma^2}\boldsymbol{\Omega}\right)^{-1}\right)$. When both the prior and the likelihood are Gaussian and there is no positivity constraint, the posterior is also Gaussian and is given by

$$p(\boldsymbol{\beta}|\boldsymbol{y},\gamma) \sim \mathcal{N}\left(\left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + \gamma\boldsymbol{\Omega}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{y}, \sigma^2\left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + \gamma\boldsymbol{\Omega}\right)^{-1}\right). \tag{6.72}$$

In this case, the posterior mean, the maximum a posteriori estimator and the penalized maximum likelihood estimator all coincide and are given by

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + \gamma\boldsymbol{\Omega}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{y}. \tag{6.73}$$

Notice that this is linear in $\boldsymbol{y}$.

The marginal distribution of $\boldsymbol{y}$ is also Gaussian and is given by

$$p(\boldsymbol{y}|\gamma) \sim \mathcal{N}\left(\boldsymbol{0}, \sigma^2\left(\boldsymbol{I} + \frac{1}{\gamma}\boldsymbol{K}\boldsymbol{\Omega}^{-1}\boldsymbol{K}^{\mathrm{T}}\right)\right). \tag{6.74}$$

As a result, the marginal log-likelihood of γ is available in closed form and is given by

$$l(\gamma;\boldsymbol{y}) = \log p(\boldsymbol{y}|\gamma) = -\frac{1}{2}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\gamma}\boldsymbol{K}\boldsymbol{\Omega}^{-1}\boldsymbol{K}^{\mathrm{T}}\right) + \frac{1}{\sigma^2}\boldsymbol{y}^{\mathrm{T}}\left(\boldsymbol{I} + \frac{1}{\gamma}\boldsymbol{K}\boldsymbol{\Omega}^{-1}\boldsymbol{K}^{\mathrm{T}}\right)^{-1}\boldsymbol{y}\right] + \text{const}, \tag{6.75}$$

where the constant does not depend on γ. We again estimate the regularization parameter γ

using the marginal maximum likelihood estimator (see Section 5.3.1),

$$\hat{\gamma}_{\mathrm{MMLE}} = \operatorname*{arg\,max}_{\gamma>0} \; l(\gamma; \boldsymbol{y}) = \operatorname*{arg\,max}_{\gamma>0} \log p(\boldsymbol{y}|\gamma). \tag{6.76}$$

The maximizer cannot be solved in closed form, but can be easily found numerically by performing a grid search in $\log \gamma$.

6.4.5.3 Results

Let us first focus on the one-peak function f_1 and the intermediate noise level $\sigma = 0.005$. We start by computing the point estimator $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + \hat{\gamma}_{\mathrm{MMLE}}\boldsymbol{\Omega}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{y}$ corresponding to the MMLE. We then apply iterative bias-correction and undersmoothing to $\hat{\boldsymbol{\beta}}$, form the corresponding 95 % Gaussian confidence intervals (6.43) for $f(s)$ and compute their actual coverage probability (ignoring the variability of $\hat{\gamma}_{\mathrm{MMLE}}$) using Equation (6.44). Figure 6.8 shows the coverage probability as a function of the interval length at $s = 0$, that is, at the top of the peak, where the coverage problems are the most severe. This figure enables us to make a number of key observations. Firstly, the coverage-length curves for both debiasing methods have a strongly concave shape. In other words, one can obtain significant improvements in coverage with only a modest increase in interval length. For example, within the range shown in the figure, the coverage improves from the MMLE value 0.42 to nearly the nominal value 0.95, but the interval length increases only by a factor of three. To put this in perspective, the unregularized intervals with $\gamma = 0$ have nominal coverage but length 19476, which is six orders of magnitude larger than the length of the MMLE intervals 0.030. Secondly, bias-correction appears to be more powerful than undersmoothing in the sense that, for a given coverage probability, the bias-corrected intervals are shorter than the undersmoothed intervals. To the author's best knowledge, this phenomenon, which occurs in all of our test cases, has not been observed before.

We next investigate the data-driven choice of the number of bias-correction iterations and the amount of undersmoothing. We use the tolerance $\varepsilon = 0.01$. That is, we aim to use the nominal 95 % intervals to obtain 94 % coverage. Figure 6.9 shows the minimum estimated bias-corrected coverage over 500 grid points for the different methods discussed in Section 6.4.3. The methods differ in the way they estimate the bias that is needed to estimate the coverage. The curve labeled "$\hat{\boldsymbol{\beta}}^{(0)}$" estimates the bias by plugging in for $\boldsymbol{\beta}$ the non-bias-corrected estimate; the curve labeled "$\hat{\boldsymbol{\beta}}^{(t)}$" estimates the bias by always plugging in for $\boldsymbol{\beta}$ the current bias-corrected estimate; and the curve labeled "$\hat{\boldsymbol{\beta}}^{(T)}$" estimates the bias by plugging in for $\boldsymbol{\beta}$ the current bias-corrected estimate until the estimated coverage starts to decrease at which point the estimate of $\boldsymbol{\beta}$ is fixed to the previous value $\hat{\boldsymbol{\beta}}^{(T)}$. The actual coverage is shown by the curve labeled "$\boldsymbol{\beta}$". We then use each of these methods to estimate the number of bias-correction iterations N_{BC} needed to obtain the target coverage $1 - \alpha - \varepsilon = 0.94$. As discussed in Section 6.4.3, the first method overestimates the coverage and chooses $\hat{N}_{\mathrm{BC}} = 1$ iteration, when at least 7 iterations would be needed to obtain the target coverage. The second method

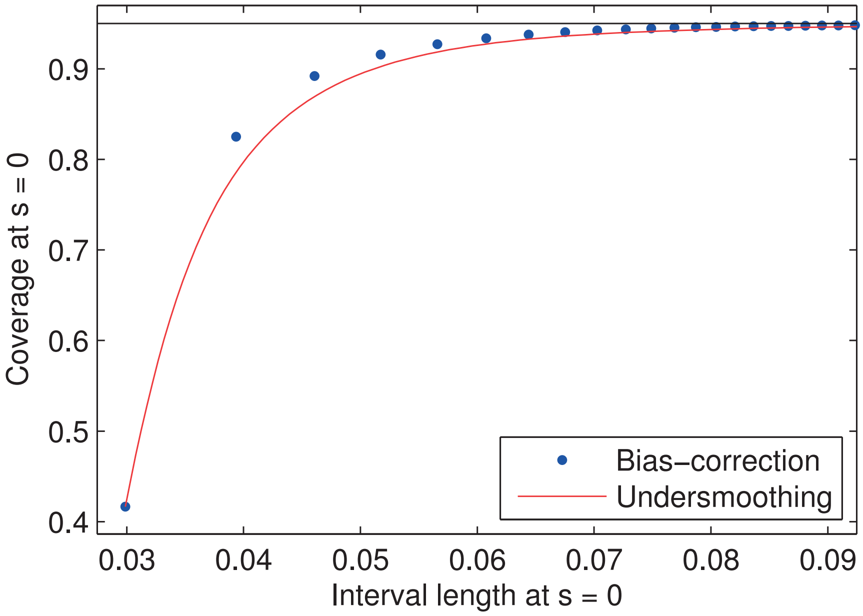


Figure 6.8: Coverage-length trade-off for iterative bias-correction and undersmoothing for the one-peak function f_1 and noise level $\sigma = 0.005$. Both methods start from the non-bias-corrected point estimator where the regularization strength is chosen using the MMLE and then debias this estimator to obtain improved coverage.

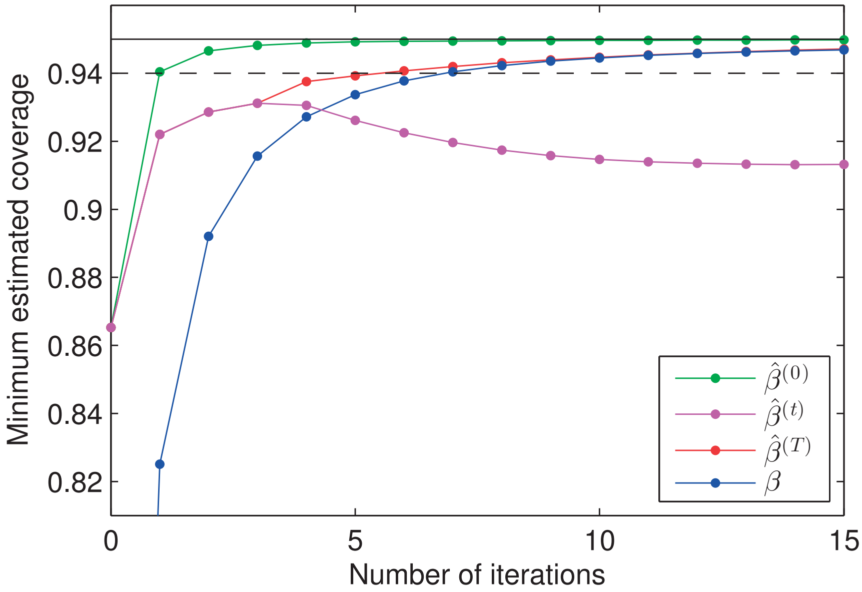


Figure 6.9: Estimated coverage as a function of bias-correction iterations for the one-peak function f_1 and noise level $\sigma = 0.005$. The nominal coverage is shown by the horizontal solid line and the target coverage by the horizontal dashed line. The number of iterations is chosen based on the point where the estimated coverage exceeds the target coverage. See the text for a description of the different estimates. The actual coverage is shown by the blue curve.

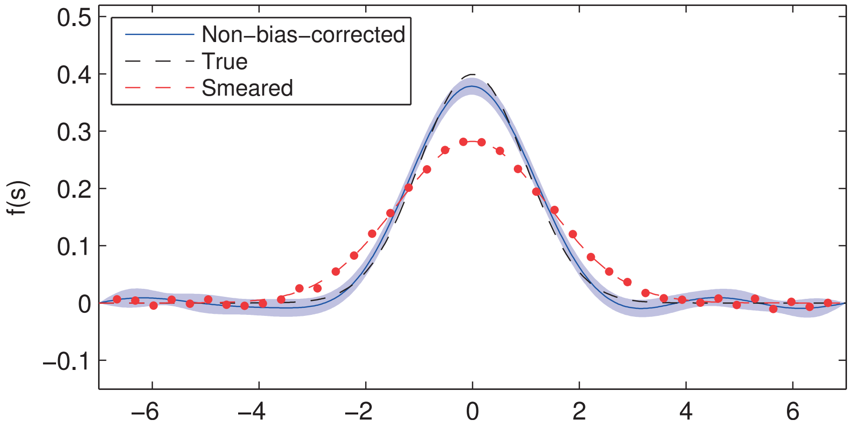
fails to achieve the target coverage since after 3 iterations the bias estimate becomes too noisy and the estimated coverage starts to decrease. The third method chooses $\hat{N}_{\mathrm{BC}} = 6$ iterations, which is one iteration short of the optimal value. (For this particular realization, the procedure chose a slightly unconservative number of iterations. However, on average, the procedure tends to be more on the conservative side for this noise level.) These results look qualitatively similar for undersmoothing and the other test scenarios—although for a very large amount of noise, the third method also underestimates the number of iterations needed to obtain the target coverage. In the rest of our simulations, we use the third method to choose the amount of debiasing.

Figure 6.10 shows a realization of the non-bias-corrected intervals, the iteratively bias-corrected intervals and the undersmoothed intervals with a data-driven choice of the amount of bias-correction and undersmoothing. The debiased intervals are wider than the non-bias-corrected ones, but only with a modest amount. We also see that the undersmoothed intervals are slightly wider and more wiggly than the iteratively bias-corrected intervals, which is consistent with our observations in Figure 6.8.

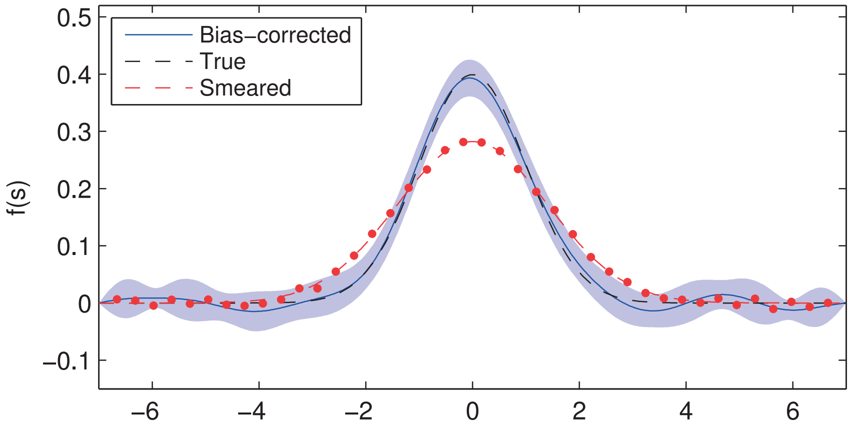
The empirical coverage and mean length of the different intervals under repeated sampling of 1 000 independent observations is shown in Figure 6.11. The curves labeled “data” are fully data-driven and the curves labeled “oracle” estimate γ with the MMLE, but use the true value of $\boldsymbol{\beta}$ to estimate the required amount of bias-correction and undersmoothing. The figure also shows the coverage and length of the non-bias-corrected intervals when γ is chosen using the MMLE or by minimizing the mean integrated squared error, which again requires oracle knowledge of $\boldsymbol{\beta}$. The debiased oracle intervals obviously achieve the target coverage 94 % and this is effectively also the case for the data-driven debiased intervals. The data-driven intervals are slightly longer than the oracle intervals, which can be attributed to the fact that the data-driven procedure occasionally selects a conservative amount of bias-correction and undersmoothing. We also observe that the undersmoothed intervals are consistently longer than the bias-corrected ones, even though both intervals have similar coverage performance. This difference is present also in the oracle intervals and is hence not an artifact of the plug-in estimate of the coverage. We also see that both non-bias-corrected intervals have shorter length but suffer from severe undercoverage. Notice that the undercoverage of the intervals induced by the MMLE is not due to a failure of $\hat{\boldsymbol{\beta}}$ as a point estimator. Indeed, the MISE optimal point estimator has even worse coverage performance than the one corresponding to the MMLE. This shows in a striking way that optimality in point estimation is different from optimality in interval estimation.

These results are summarized in Table 6.1, which shows that the mean length of the data-driven iteratively bias-corrected intervals is approximately 13 % shorter than the mean length of the undersmoothed intervals. The table also includes the results for the low noise level $\sigma = 0.001$ and the high noise level $\sigma = 0.025$ situations. With $\sigma = 0.001$, the data-driven debiased intervals attain the 94 % target coverage and there is little difference between iterative bias-correction and undersmoothing. With $\sigma = 0.025$, the data-driven debiased intervals fall slightly

(a) Non–bias–corrected intervals, σ = 0.005



(b) Iteratively bias–corrected intervals, σ = 0.005



(c) Undersmoothed intervals, σ = 0.005

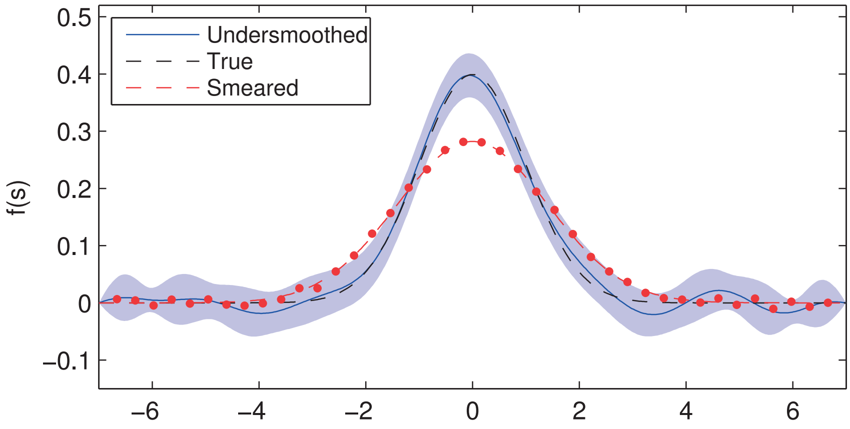
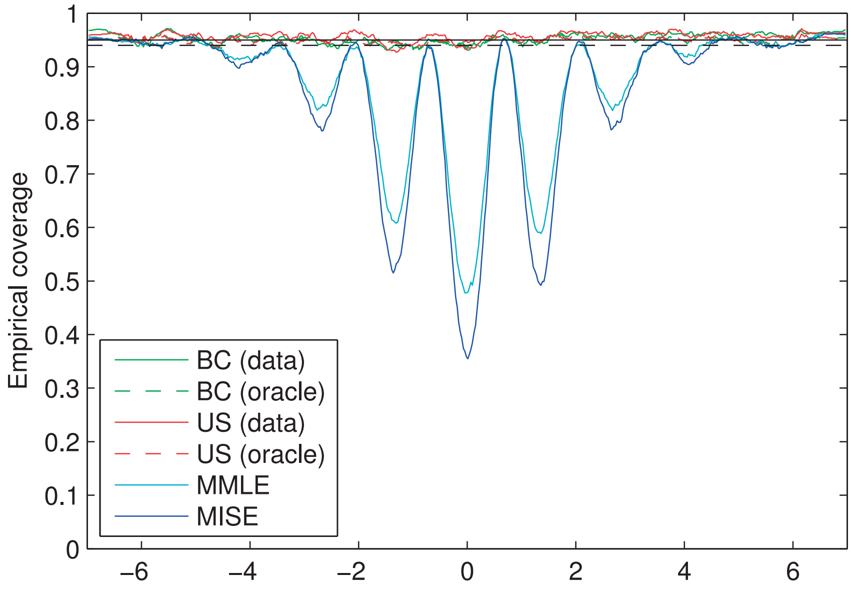


Figure 6.10: Data-driven confidence intervals for the one-peak function f_1 and noise level $\sigma = 0.005$. The regularization strength is chosen using the MMLE and the amount of bias-correction and undersmoothing using the procedure described in Sections 6.4.3 and 6.4.4. The intervals have 95 % nominal coverage and 94 % target coverage.

(a) Empirical coverage (one peak, σ = 0.005)



(b) Mean interval length (one peak, σ = 0.005)

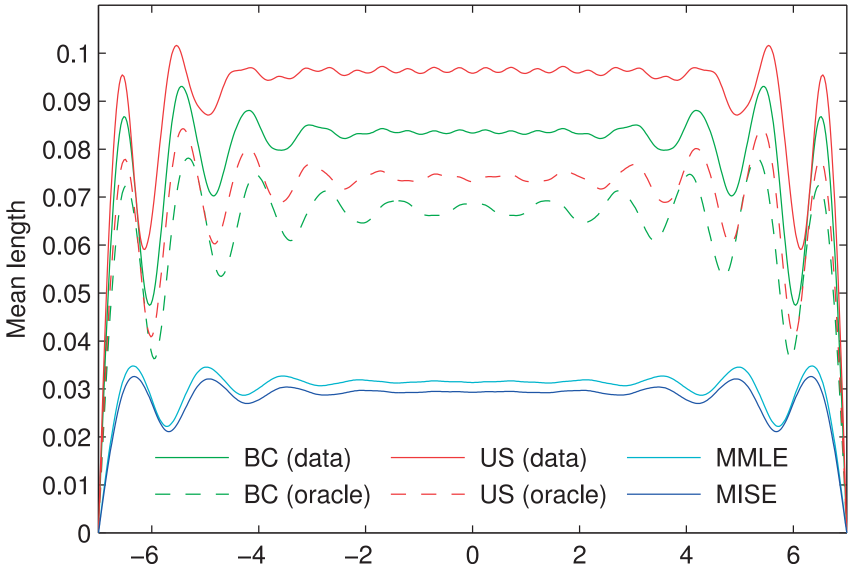


Figure 6.11: Empirical coverage and mean length of the data-driven bias-corrected (BC), undersmoothed (US) and non-bias-corrected intervals for the one-peak function f_1 and noise level $\sigma = 0.005$. The curves labeled "data" are fully data-driven and the curves labeled "oracle" use knowledge of f_1 to choose the amount of debiasing. The non-bias-corrected results are given for both the MMLE choice of the regularization strength as well as for the choice that minimizes the MISE. The intervals have 95 % nominal coverage and 94 % target coverage.

Table 6.1: Empirical coverage at $s = 0$ (the location of the peak) and mean interval length (averaged over both s and the sampling variability) for the one-peak function f_1. The results are shown for bias-corrected (BC), undersmoothed (US), non-bias-corrected (MMLE & MISE) and unregularized intervals with 95 % nominal coverage and 94 % target coverage. The debiased intervals are either fully data-driven or use oracle knowledge of f_1 to choose the amount of debiasing. The uncertainties given in the parentheses are 95 % confidence intervals.

Noise level	Method	Coverage at $s = 0$	Mean length
$\sigma = 0.001$	BC (data)	0.936 *(0.919, 0.950)*	0.047 *(0.045, 0.049)*
	BC (oracle)	0.939 *(0.922, 0.953)*	0.028 *(0.028, 0.028)*
	US (data)	0.947 *(0.931, 0.960)*	0.047 *(0.044, 0.049)*
	US (oracle)	0.942 *(0.926, 0.956)*	0.029 *(0.029, 0.029)*
	MMLE	0.817 *(0.792, 0.841)*	0.017 *(0.017, 0.017)*
	MISE	0.402 *(0.371, 0.433)*	0.011
	Unregularized	0.956 *(0.941, 0.968)*	8063
$\sigma = 0.005$	BC (data)	0.932 *(0.915, 0.947)*	0.079 *(0.077, 0.081)*
	BC (oracle)	0.937 *(0.920, 0.951)*	0.064 *(0.064, 0.064)*
	US (data)	0.933 *(0.916, 0.948)*	0.091 *(0.087, 0.095)*
	US (oracle)	0.949 *(0.933, 0.962)*	0.070 *(0.070, 0.070)*
	MMLE	0.478 *(0.447, 0.509)*	0.030 *(0.030, 0.030)*
	MISE	0.359 *(0.329, 0.390)*	0.028
	Unregularized	0.952 *(0.937, 0.964)*	40316
$\sigma = 0.025$	BC (data)	0.865 *(0.842, 0.886)*	0.132 *(0.129, 0.134)*
	BC (oracle)	0.939 *(0.922, 0.953)*	0.155 *(0.155, 0.155)*
	US (data)	0.881 *(0.859, 0.900)*	0.171 *(0.165, 0.177)*
	US (oracle)	0.945 *(0.929, 0.958)*	0.171 *(0.171, 0.171)*
	MMLE	0.193 *(0.169, 0.219)*	0.062 *(0.062, 0.062)*
	MISE	0.374 *(0.344, 0.405)*	0.072
	Unregularized	0.946 *(0.930, 0.959)*	201580

short of the target coverage, but perform still significantly better than the non-bias-corrected intervals induced by the MMLE, which for this noise level fail catastrophically. In each case, the regularized intervals are several orders of magnitude shorter than the unregularized ones.

Table 6.2 shows the same results for the two-peak function f_2. The general conclusions from these results are similar to those of the one-peak situation. We notice however that in this case at noise level $\sigma = 0.001$ the bias-corrected intervals seem to be slightly longer than the undersmoothed ones, while in all other test cases the undersmoothed intervals have been longer. This perhaps suggests that in low-noise situations, where little debiasing is needed, the full bias-correction iterations are too coarse and undersmoothing benefits from the possibility of making finer adjustments. Nevertheless, on a coverage-length plot similar to Figure 6.8, bias-correction still lies above undersmoothing.

Table 6.2: Empirical coverage at $s = 2$ (the location of the larger peak) and mean interval length (averaged over both s and the sampling variability) for the two-peak function f_2. The results are shown for bias-corrected (BC), undersmoothed (US), non-bias-corrected (MMLE & MISE) and unregularized intervals with 95 % nominal coverage and 94 % target coverage. The debiased intervals are either fully data-driven or use oracle knowledge of f_2 to choose the amount of debiasing. The uncertainties given in the parentheses are 95 % confidence intervals.

Noise level	Method	Coverage at $s = 2$	Mean length
$\sigma = 0.001$	BC (data)	0.960 *(0.946, 0.971)*	0.040 *(0.038, 0.041)*
	BC (oracle)	0.951 *(0.936, 0.964)*	0.026 *(0.026, 0.026)*
	US (data)	0.958 *(0.944, 0.970)*	0.038 *(0.036, 0.040)*
	US (oracle)	0.947 *(0.931, 0.960)*	0.025 *(0.025, 0.025)*
	MMLE	0.845 *(0.821, 0.867)*	0.014 *(0.014, 0.014)*
	MISE	0.235 *(0.209, 0.263)*	0.009
	Unregularized	0.942 *(0.926, 0.956)*	8063
$\sigma = 0.005$	BC (data)	0.934 *(0.917, 0.949)*	0.065 *(0.064, 0.067)*
	BC (oracle)	0.931 *(0.913, 0.946)*	0.053 *(0.053, 0.053)*
	US (data)	0.923 *(0.905, 0.939)*	0.076 *(0.073, 0.078)*
	US (oracle)	0.933 *(0.916, 0.948)*	0.058 *(0.057, 0.058)*
	MMLE	0.612 *(0.581, 0.642)*	0.027 *(0.027, 0.027)*
	MISE	0.539 *(0.508, 0.570)*	0.025
	Unregularized	0.963 *(0.949, 0.974)*	40316
$\sigma = 0.025$	BC (data)	0.859 *(0.836, 0.880)*	0.112 *(0.110, 0.114)*
	BC (oracle)	0.941 *(0.925, 0.955)*	0.129 *(0.129, 0.129)*
	US (data)	0.852 *(0.828, 0.873)*	0.149 *(0.144, 0.154)*
	US (oracle)	0.943 *(0.927, 0.957)*	0.139 *(0.139, 0.139)*
	MMLE	0.132 *(0.112, 0.155)*	0.051 *(0.051, 0.051)*
	MISE	0.582 *(0.551, 0.613)*	0.070
	Unregularized	0.939 *(0.922, 0.953)*	201580

These same experiments were repeated with the smearing kernel in Equation (6.66) replaced by the Laplace distribution with zero mean and unit variance, $k(t,s) = k(t-s) = \mathrm{Laplace}(t-s|0,1/\sqrt{2})$. The conclusions from the Laplace experiments were analogous to those of the Gaussian experiments. The main difference was that the Laplace smearing seems less sensitive to noise than the Gaussian smearing. For example, at $\sigma = 0.025$, the empirical coverage was almost at the target value 94 % for both debiasing methods and both test scenarios. This difference is hardly surprising since the Laplace distribution is an example of an ordinary smooth density and as such should lead to more tractable deconvolution problems than the supersmooth Gaussian distribution (Meister, 2009).

For the case of Gaussian smearing, the full simulation results for the different scenarios are given in Appendix B.2.

6.5 Unfolding with data-driven bias-correction

In this section, we demonstrate how to apply the ideas presented in Section 6.4 to the full unfolding problem. One could in principle use the bootstrap to calibrate the iteratively bias-corrected percentile intervals based on either the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ or the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ to yield $1-\alpha-\varepsilon$ coverage. This would be done by resampling new observations $\boldsymbol{y}^*$ and using these to check the coverage of the intervals for different numbers of bias-correction iterations N_{BC}. However, this would mean three layers of bootstrap resampling: one for estimating the bias, one for constructing the intervals and one for calibrating the coverage of these intervals. While this is in principle doable, the procedure would in practice be computationally extremely demanding. For this reason, we opt for a computationally more tractable approach based on the unconstrained Gaussian approximation described in Section 5.2.4.

Throughout this section, we use the MMLE as described in Section 5.3.1 to estimate the regularization strength δ. If we ignore the data-dependence of $\hat{\boldsymbol{\Sigma}}$ and $\hat{\delta}_{\mathrm{MMLE}}$, the unconstrained Gaussian estimator is linear in $\boldsymbol{y}$, $\hat{\boldsymbol{\beta}}_{\mathrm{G}} = \boldsymbol{A}\boldsymbol{y}$, with $\boldsymbol{A} = \left(\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{K} + 2\hat{\delta}_{\mathrm{MMLE}}\boldsymbol{\Omega}_{\mathrm{A}}\right)^{-1}\boldsymbol{K}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}$. We then know by Section 6.4.1, that the corresponding iteratively bias-corrected estimator is given by $\hat{\boldsymbol{\beta}}^{(t)} = \boldsymbol{J}^{(t)}\hat{\boldsymbol{\beta}}_{\mathrm{G}} = \boldsymbol{J}^{(t)}\boldsymbol{A}\boldsymbol{y}$, where $\boldsymbol{J}^{(t)} = \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{A}\boldsymbol{K})\boldsymbol{J}^{(t-1)}$ and $\boldsymbol{J}^{(0)} = \boldsymbol{I}$. The $1-\alpha$ iteratively bias-corrected Gaussian confidence intervals for $\theta_i = f(s_i) = \sum_{j=1}^{p} \beta_j B_j(s_i) = \boldsymbol{c}_i^{\mathrm{T}}\boldsymbol{\beta}$ induced by this estimator are

$$[\underline{\theta}_i, \overline{\theta}_i] = \left[\hat{\theta}_i - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}(\hat{\theta}_i)}, \hat{\theta}_i + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{var}}(\hat{\theta}_i)}\right] \tag{6.77}$$

$$= \left[\hat{\theta}_i - z_{1-\alpha/2}\sqrt{\boldsymbol{c}_i^{\mathrm{T}}\boldsymbol{J}^{(t)}\boldsymbol{A}\hat{\boldsymbol{\Sigma}}\left(\boldsymbol{J}^{(t)}\boldsymbol{A}\right)^{\mathrm{T}}\boldsymbol{c}_i}, \hat{\theta}_i + z_{1-\alpha/2}\sqrt{\boldsymbol{c}_i^{\mathrm{T}}\boldsymbol{J}^{(t)}\boldsymbol{A}\hat{\boldsymbol{\Sigma}}\left(\boldsymbol{J}^{(t)}\boldsymbol{A}\right)^{\mathrm{T}}\boldsymbol{c}_i}\right], \tag{6.78}$$

where $\hat{\theta}_i = \boldsymbol{c}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}^{(t)}$. If we assume that the sample size is large enough that the distribution of $\boldsymbol{y}$ is close enough to the Gaussian $N(\boldsymbol{K}\boldsymbol{\beta}, \hat{\boldsymbol{\Sigma}})$ and again ignore the data-dependence of $\hat{\boldsymbol{\Sigma}}$ and $\hat{\delta}_{\mathrm{MMLE}}$, then Proposition 6.2 can be used to write down the coverage of the interval (6.78). We can then use Algorithm 4 to estimate this coverage and choose the number of bias-correction iterations N_{BC} to yield the target coverage $1-\alpha-\varepsilon$. An analogous approach can also be implemented for data-driven undersmoothed Gaussian intervals as described in Section 6.4.4.

We demonstrate the performance of these data-driven debiased intervals by unfolding the two peaks on a uniform background test setup described in Section 3.4.1 and also studied in Sections 5.4 and 6.4.1. The MMLE of the regularization strength was obtained as in Section 5.4.1 and the boundary hyperparameters were also set to the same values as in that section. The number of bias-correction iterations was chosen to give 94 % target coverage for the nominal 95 % intervals. The estimated number of bias-correction iterations was $\hat{N}_{\mathrm{BC}} = 16$, $\hat{N}_{\mathrm{BC}} = 16$ and $\hat{N}_{\mathrm{BC}} = 12$ for the small, medium and large sample size cases, respectively (the same number of iterations for the small and medium sample size is a coincidence; on average a larger number of iterations is taken for the small sample size than for the medium sample size). For undersmoothing, the respective estimated regularization strengths were

Table 6.3: Empirical coverage at $s = 2$ (the location of the larger peak) and mean interval length (averaged over both s and the sampling variability) in unfolding the two peaks on a uniform background test setup using data-driven iteratively bias-corrected, undersmoothed and non-bias-corrected Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. The regularization strength was chosen using the MMLE and the intervals have 94 % pointwise target coverage and 95 % nominal coverage. The uncertainties in the parentheses are 95 % confidence intervals.

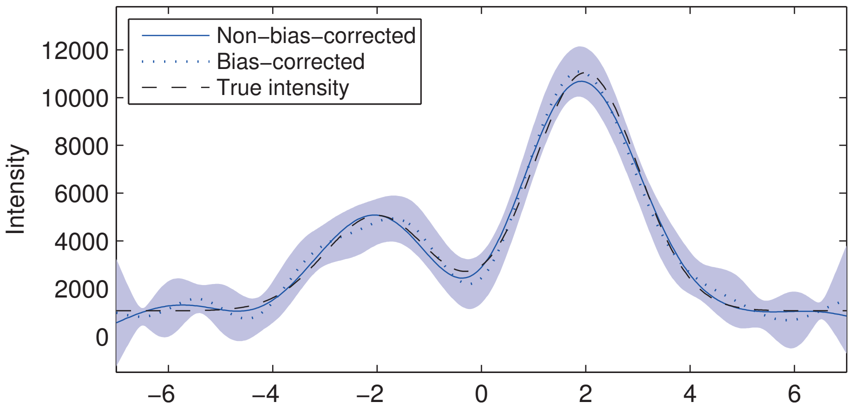
Sample size	Method	Coverage at $s = 2$	Mean length
$\lambda_{\mathrm{tot}} = 50\,000$	Bias-corrected	0.937 *(0.920, 0.951)*	2111 *(2041, 2181)*
	Undersmoothed	0.924 *(0.906, 0.940)*	2291 *(2189, 2393)*
	Non-bias-corrected	0.557 *(0.526, 0.588)*	737 *(736, 739)*
$\lambda_{\mathrm{tot}} = 10\,000$	Bias-corrected	0.918 *(0.899, 0.934)*	500 *(489, 510)*
	Undersmoothed	0.919 *(0.900, 0.935)*	633 *(605, 661)*
	Non-bias-corrected	0.342 *(0.313, 0.372)*	198 *(197, 198)*
$\lambda_{\mathrm{tot}} = 1\,000$	Bias-corrected	0.805 *(0.779, 0.829)*	68.9 *(67.6, 70.2)*
	Undersmoothed	0.855 *(0.832, 0.876)*	99.5 *(95.9, 103.1)*
	Non-bias-corrected	0.047 *(0.035, 0.062)*	29.2 *(29.0, 29.3)*

$\hat{\delta}_{\mathrm{US}} = 3.0 \cdot 10^{-6}$, $\hat{\delta}_{\mathrm{US}} = 3.4 \cdot 10^{-8}$ and $\hat{\delta}_{\mathrm{US}} = 2.1 \cdot 10^{-9}$, which should be compared with the MMLE values $\hat{\delta}_{\mathrm{MMLE}} = 2.2 \cdot 10^{-4}$, $\hat{\delta}_{\mathrm{MMLE}} = 1.2 \cdot 10^{-6}$ and $\hat{\delta}_{\mathrm{MMLE}} = 4.1 \cdot 10^{-8}$.

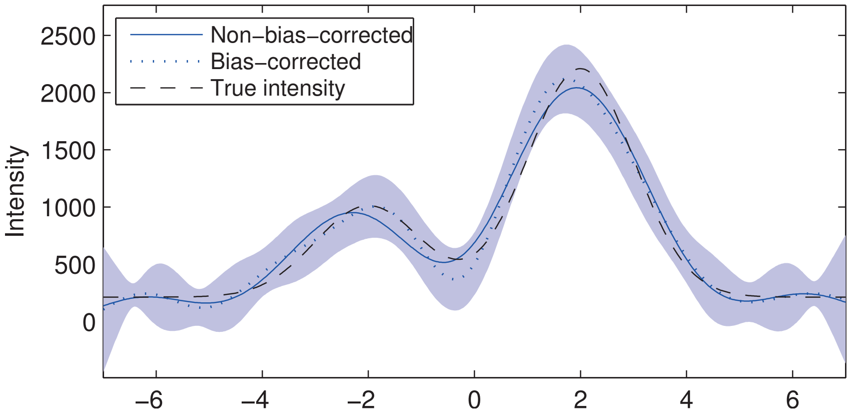
The resulting data-driven iteratively bias-corrected Gaussian intervals are shown in Figure 6.12. The undersmoothed intervals are given in Appendix B.3, which also provides the Gaussian intervals induced by the non-bias-corrected estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. We find that the data-driven procedure has reduced the bias of the estimators enough to provide improved uncertainty estimates especially at the larger peak, while still managing to maintain the length of the intervals within reasonable limits. The confidence bands are perhaps slightly more wiggly than desired near the tails of the intensity, which can be partially attributed to the lack of the positivity constraint in $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ as well as to the weak boundary constraint employed here.

Figure 6.13 shows the empirical coverage and mean length of these intervals for 1 000 repeated observations and sample size $\lambda_{\mathrm{tot}} = 10\,000$. Analogous plots for the other sample sizes are given in Appendix B.3 and a summary of these results is given in Table 6.3. The conclusions that emerge from these results are similar to those of Section 6.4.5. The coverage of the data-driven debiased intervals is close to the 94 % target coverage, except for the small sample size, where the intervals have slight undercoverage. In all cases, the coverage of the debiased intervals is significantly better than that of the non-bias-corrected intervals, which suffer from drastic undercoverage. As in Section 6.4.5, the bias-corrected intervals are shorter than the undersmoothed ones. For example, at the medium sample size, the mean length (averaged over both s and the sampling variability) of the bias-corrected intervals is 21 % smaller than that of the undersmoothed intervals. As a result, we conclude that the bias-corrected

(a) Iteratively bias−corrected Gaussian intervals, λ_{tot} = 50000



(b) Iteratively bias−corrected Gaussian intervals, λ_{tot} = 10000



(c) Iteratively bias−corrected Gaussian intervals, λ_{tot} = 1000

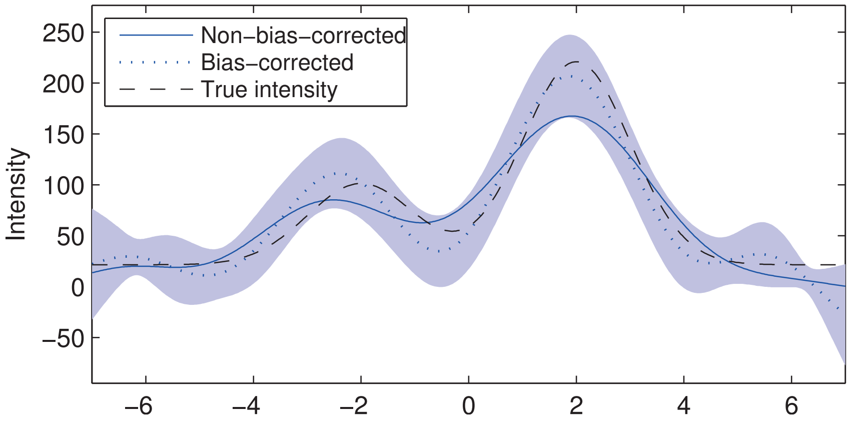
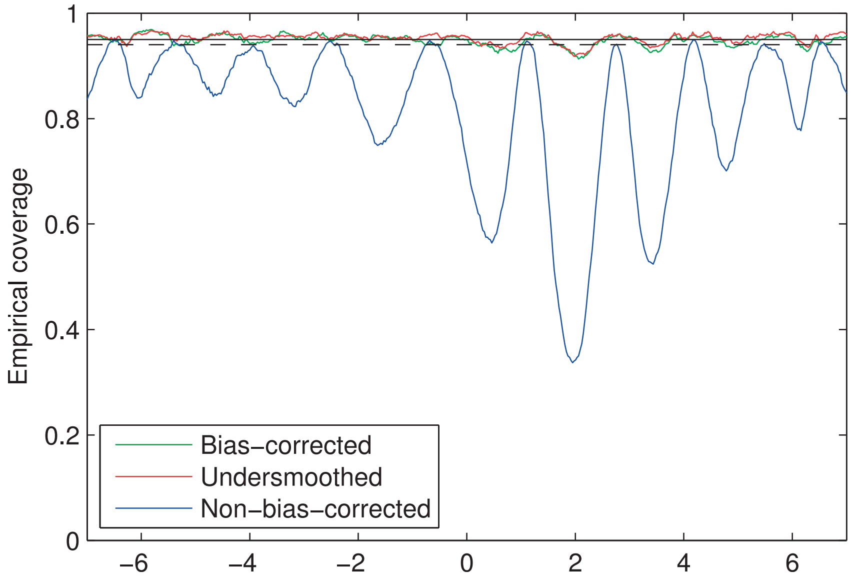


Figure 6.12: Unfolding of the two peaks on a uniform background test setup using data-driven iteratively bias-corrected Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. The intervals have 94 % pointwise target coverage and 95 % nominal coverage. The regularization strength was chosen using the MMLE.

(a) Empirical coverage, λ_{tot} = 10000



(b) Mean interval length, λ_{tot} = 10000

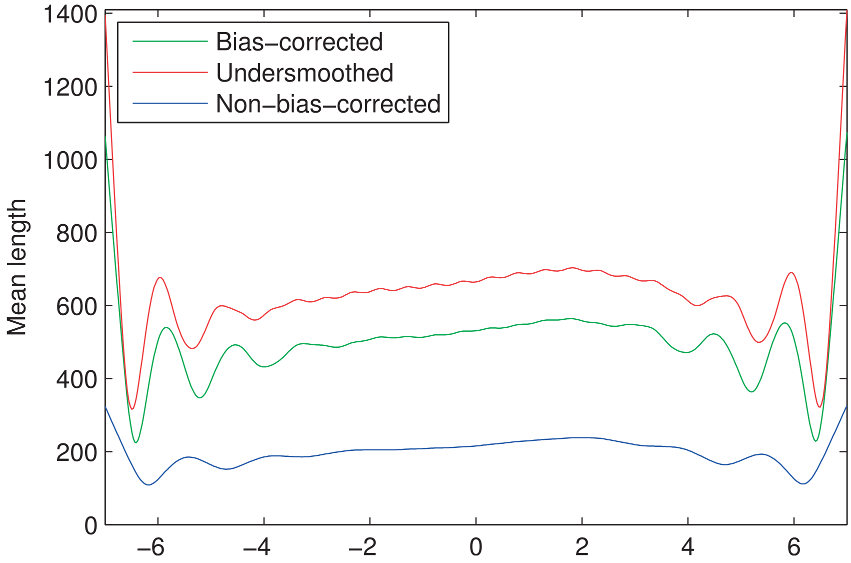


Figure 6.13: Empirical coverage and mean interval length in unfolding the two peaks on a uniform background test setup using data-driven iteratively bias-corrected, undersmoothed and non-bias-corrected Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. The sample size was $\lambda_{\mathrm{tot}} = 10\,000$ and the regularization strength chosen using the MMLE. The intervals have 94 % pointwise target coverage (dashed horizontal line) and 95 % nominal coverage (solid horizontal line).

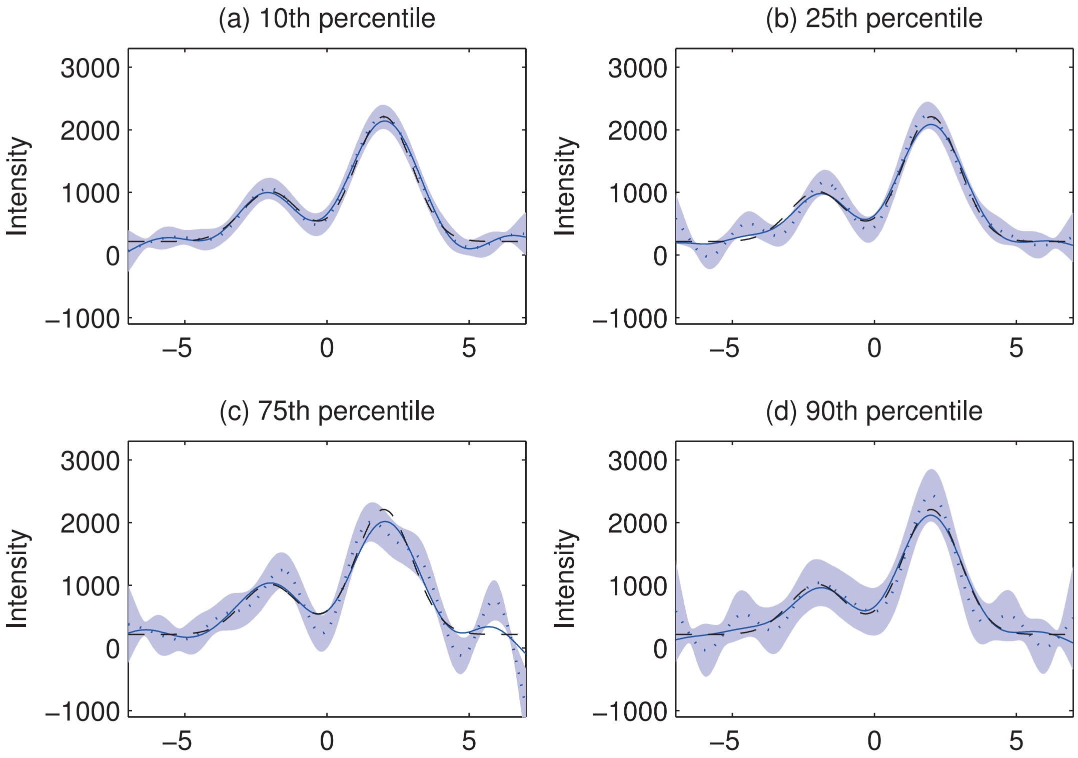


Figure 6.14: Illustration of the variability of the data-driven iteratively bias-corrected Gaussian confidence intervals in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\text{tot}} = 10\,000$. The panels show realizations of the intervals for given percentiles of the sampling distribution of interval lengths (averaged over s). The intervals have 94 % pointwise target coverage and 95 % nominal coverage. The figures also show the true intensity (dashed line), the non-bias-corrected point estimate (solid line) and the bias-corrected point estimate (dotted line).

intervals should be preferred over the undersmoothed ones for sample sizes $\lambda_{\text{tot}} = 10\,000$ and $\lambda_{\text{tot}} = 50\,000$. For the small sample size $\lambda_{\text{tot}} = 1\,000$, the situation however is not as clear-cut. Namely, the bias-corrected intervals are shorter but also seem to undercover more than the undersmoothed intervals. This is most likely an anomaly caused by the data-based estimate of the coverage—for an oracle coverage, we would expect by Section 6.4.5 that both methods attain the 94 % target coverage, with the bias-corrected intervals still remaining shorter than the undersmoothed ones. Another difference in comparison to Section 6.4.5 is that there is a significant increase in the interval length near the boundaries of the true space E. This happens because the boundary constraint here is weaker than that of Section 6.4.5 and could be addressed by including stronger a prior information about the behavior of f near the boundary, if such information is available.

It should be noted that there is some amount of variation in the length and wiggliness of the debiased confidence intervals. The realizations shown in Figure 6.12 are such that the interval length (averaged over s) is close to the mean interval length (averaged over both s

and the sampling variability). To get a better sense of the sampling variability of the intervals, Figure 6.14 shows realizations of the data-driven bias-corrected intervals for given percentiles of the sampling distribution of interval lengths (averaged over s) for sample size $\lambda_{\mathrm{tot}} = 10\,000$. Analogous plots for the other sample sizes are given in Appendix B.3. We conclude that even in the more severe cases the length and wiggliness of the intervals remains tolerable.

6.6 Real data demonstration: Z boson invariant mass spectrum

6.6.1 Description of the data

In this section, we demonstrate the uncertainty quantification methodology developed in the present chapter by unfolding the Z boson invariant mass spectrum measured in the CMS experiment at the LHC. The Z boson is the carrier of the weak interaction and is produced in large quantities at the LHC. Once created, it decays almost instantaneously to other types of particles. Here we consider in particular the decay mode to a positron and an electron, $Z \to e^+e^-$. We use the smeared invariant mass spectrum published in CMS Collaboration (2013a). These data were originally collected to calibrate the CMS electromagnetic calorimeter, but also constitute an excellent testbed for unfolding since the true intensity f underlying these data is known with great accuracy from previous experiments and theoretical considerations.

The electron and the positron from the Z decay pass first through the CMS silicon tracker after which their energies $\mathscr{E}_i$, $i = 1, 2$, are measured in the ECAL; see Section 2.1. Using the information from these two detectors, one can compute the *invariant mass* W of the electron-positron system,

$$W^2 = (\mathscr{E}_1 + \mathscr{E}_2)^2 - \|\boldsymbol{p}_1 + \boldsymbol{p}_2\|_2^2, \tag{6.79}$$

where $\boldsymbol{p}_i$, $i = 1, 2$, are the momenta of the two particles and the equation is written in natural units where the speed of light $c = 1$. Since $\|\boldsymbol{p}_i\|_2^2 = \mathscr{E}_i^2 - m_e^2$, where m_e is the rest mass of the electron, one can calculate the invariant mass W using only the energy deposits $\mathscr{E}_i$ and the opening angle between the two tracks in the tracker.

The invariant mass W is conserved in particle decays. It is also Lorentz invariant, that is, it has the same value in every frame of reference. This implies that the invariant mass of the Z boson, which is simply its rest mass m, is equal to the invariant mass of the electron-positron system, $W = m$. As a result, the invariant mass spectrum of the electron-positron pair is directly also the mass spectrum of the Z boson that produced the pair.

As a result of the time-energy uncertainty principle, the Z boson does not have a unique rest mass m. Instead, the mass is known to follow the Cauchy (or Breit–Wigner) distribution,

$$p(m) = \frac{1}{2\pi} \frac{\Gamma}{(m - m_Z)^2 + \frac{\Gamma^2}{4}}, \tag{6.80}$$

where $m_Z = 91.1876$ GeV is the mode of the distribution (sometimes simply called *the* mass of the Z boson) and $\Gamma = 2.4952$ GeV is the full width of the distribution at half maximum (Beringer et al., Particle Data Group, 2012). Since the contribution of background processes to the electron-positron channel near the Z peak is negligible (CMS Collaboration, 2013a), the true intensity $f(m)$ is proportional to $p(m)$ and we can validate the unfolded results by comparing the shape of the unfolded spectrum to that of $p(m)$.

The Z boson invariant mass m is smeared by the noise in the energy measurements $\mathscr{E}_i$. To a first approximation, the detector response can be modeled as a convolution with a fixed-width Gaussian kernel. However, a more realistic response can be obtained by replacing the left tail of the kernel by a more slowly decaying function in order to account for energy losses in the ECAL. The resulting response function is the so-called *Crystal Ball function* (Oreglia, 1980; CMS Collaboration, 2013a),

$$\mathrm{CB}(m|\Delta m, \sigma^2, \alpha, \gamma) = \begin{cases} Ce^{-\frac{(m-\Delta m)^2}{2\sigma^2}}, & \frac{m-\Delta m}{\sigma} > -\alpha, \\ C\left(\frac{\gamma}{\alpha}\right)^{\gamma} e^{-\frac{\alpha^2}{2}} \left(\frac{\gamma}{\alpha} - \alpha - \frac{m-\Delta m}{\sigma}\right)^{-\gamma}, & \frac{m-\Delta m}{\sigma} \le -\alpha, \end{cases} \tag{6.81}$$

where $\sigma, \alpha > 0$, $\gamma > 1$ and C is a normalization constant chosen so that the function is a probability density. The Crystal Ball function is a Gaussian density with mean Δm and variance σ^2, where the left tail is replaced by a power law. The parameter α controls the location of the transition from exponential decay into power-law decay and γ gives the decay rate of the power-law tail. The integration kernel in Equation (3.1) is then given by

$$k(t, s) = k(t - s) = \mathrm{CB}(t - s|\Delta m, \sigma^2, \alpha, \gamma), \tag{6.82}$$

corresponding to a convolution with the Crystal Ball function (6.81).

The specific dataset we use is a digitized version of the lower left hand plot of Figure 11 in CMS Collaboration (2013a). These data have center-of-mass energy 7 TeV and correspond to an integrated luminosity of 4.98 fb^{-1} collected by the CMS experiment in 2011. The dataset consists of 67 778 electron-positron events in 100 equal-width bins with smeared invariant masses between 65 GeV and 115 GeV. Details of the event selection are given in CMS Collaboration (2013a) and the references therein.

We estimate the parameters of the Crystal Ball response by dividing the dataset into two independent samples by drawing a binomial random variable independently for each bin with the number of trials equal to the observed bin contents. The bins of the resulting two smeared histograms are marginally independent and Poisson distributed. Each event had a 70 % probability of belonging to the histogram $\boldsymbol{y}$ used in the unfolding demonstration; the rest of the smeared events were used to estimate the Crystal Ball parameters.

We estimated the Crystal Ball parameters $(\Delta m, \sigma^2, \alpha, \gamma)$ using maximum likelihood on the full invariant mass range 65–115 GeV. To do this, we assumed that the true intensity $f(m)$ is proportional to the Breit–Wigner distribution (6.80) and also estimated the unknown proportionality constant as part of the maximum likelihood fit. The likelihood function was obtained by substituting Equations (6.80) and (6.82) into Equation (3.7). The maximum likelihood estimates of the Crystal Ball parameters were

$$(\Delta\hat{m}, \hat{\sigma}^2, \hat{\alpha}, \hat{\gamma}) = (0.56\ \text{GeV}, (1.01\ \text{GeV})^2, 1.95, 1.40). \tag{6.83}$$

In other words, the smeared events are on average shifted to the right by 0.56 GeV and have an experimental resolution of approximately 1 GeV. As a cross-check of the fit, the expected smeared histogram corresponding to the estimated Crystal Ball response was compared to the smeared observations and the two were found to be in good agreement.

The procedure described here for estimating the forward mapping is similar to what would be used to estimate the detector response from test beam measurements, but is unrealistic in the sense that it requires at least partial knowledge of the unknown intensity f. A more realistic alternative approach would be to use Monte Carlo detector simulations. That is, one would generate simulated $Z \rightarrow e^+e^-$ events and propagate the true invariant masses X_i^{MC} through a computer model of the CMS detector to obtain the corresponding smeared invariant masses Y_i^{MC}. One would then use the pairs $(X_i^{\mathrm{MC}}, Y_i^{\mathrm{MC}})$ to obtain an estimate of the smearing kernel $k(t,s)$, which could also be nonparametric. However, due to the unavailability of the required MC sample, we did not pursue such more complex estimates of the forward mapping in the present work.

6.6.2 Unfolding setup and results

We unfold the Z boson invariant mass spectrum using the $n = 30$ bins on the interval $F = [82.5\ \text{GeV}, 97.5\ \text{GeV}]$. The resulting subsampled smeared histogram $\boldsymbol{y}$ had 42 475 electron-positron events. We let the true space be $E = [81.5\ \text{GeV}, 98.5\ \text{GeV}]$, which extends F by approximately $1\hat{\sigma}$ in order to account for events that are smeared into the observed interval from outside its boundaries. The true space E was discretized using cubic B-splines with $L = 34$ uniformly placed interior knots corresponding to $p = 38$ unknown spline coefficients. This overparameterization with $p > n$ was found to improve the mixing the MCMC sampler. The condition number of the resulting design matrix $\boldsymbol{K}$ was $\mathrm{cond}(\boldsymbol{K}) \approx 8.1 \cdot 10^3$, indicating that the problem is severely ill-conditioned. We used the values $\gamma_{\mathrm{L}} = \gamma_{\mathrm{R}} = 50$ for the boundary hyperparameters.

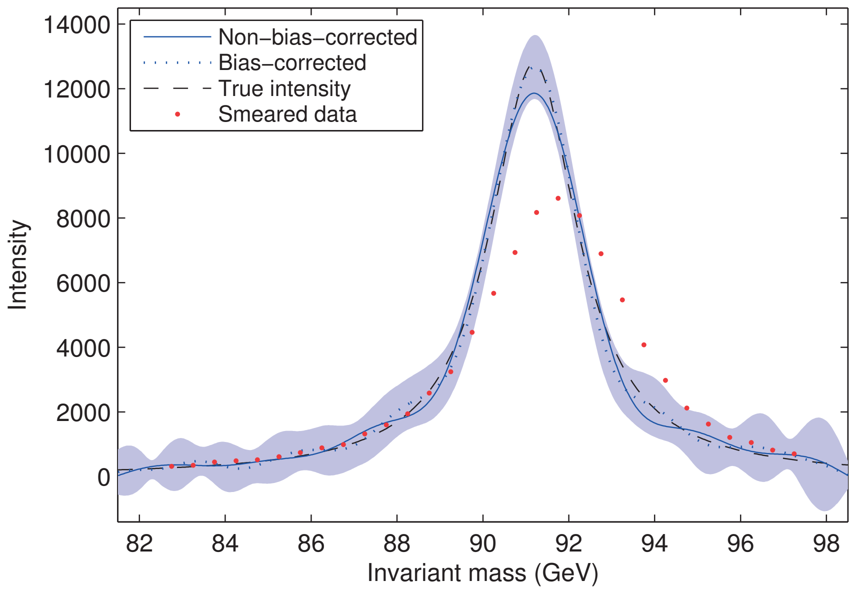
We first used the MCEM iteration described in Algorithm 1 to obtain the MMLE of the regularization strength δ. The parameters of the algorithm were set to the same values as in Section 5.4.1. The algorithm converged in approximately 10 iterations to the estimate $\hat{\delta}_{\mathrm{MMLE}} = 7.0 \cdot 10^{-8}$ with little Monte Carlo variation. We then used Algorithm 4 to obtain an estimate of the number of bias-correction iterations needed to obtain 94 % target coverage using the nominal 95 % Gaussian intervals. The choice of the number of bias-correction iterations was based on the unconstrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ as described in Section 6.5. The estimated number of bias-correction iterations was $\hat{N}_{\mathrm{BC}} = 14$.

Figure 6.15(a) shows the nominal 95 % Gaussian intervals induced by the bias-corrected estimator $\hat{\boldsymbol{\beta}}_{\mathrm{BC}} = \boldsymbol{J}^{(14)}\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. By Sections 6.4 and 6.5, we expect these intervals to have close to 94 % pointwise coverage and we indeed observe that the intervals cover the true intensity f across the whole spectrum. (The proportionality constant of f was obtained using the maximum likelihood fit described in Section 6.6.1.) The length of the bias-corrected intervals seems very reasonable given the ill-posedness of the problem, although the intervals are a bit more wiggly than desired at the tails of the mass peak, but this is the price that one has to pay in order to obtain good coverage performance at the peak itself. Notice also how the bias-corrected point estimator $\hat{f}_{\mathrm{BC}}$ improves upon the non-bias-corrected estimator at the peak, where the original estimator clearly underestimates the size of the peak.

We also used Algorithm 3 with $\hat{N}_{\mathrm{BC}} = 14$ iterations to compute the 95 % iteratively bias-corrected percentile intervals based on the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ (see Section 5.2.1). The regularization strength was set to $\hat{\delta}_{\mathrm{MMLE}}$ and the same bootstrap sample sizes were used as in Section 6.3. The resulting intervals, which are shown in Figure 6.15(b), are fairly similar to the iteratively bias-corrected Gaussian intervals and also cover the true intensity f across the whole spectrum. However, the percentile intervals are less wiggly at the tails of the Z peak. This difference is presumably due to the fact that the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$ imposes the positivity constraint while the Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ does not. At the same time the intervals are more rugged than the Gaussian intervals which is due to the finite size of the bootstrap and MCMC samples. Strictly speaking, we are not able to make claims about the coverage of these intervals based on our previous experiments. However, given the similarity of the percentile intervals and the Gaussian intervals especially at the top of the Z peak one could conjecture that the coverage of these intervals is also close to the target coverage 94 %.

Notice that both of these intervals neglect the uncertainty concerning the data-based determination of the smearing kernel. While it might be possible to incorporate this uncertainty into the bootstrap-based intervals, we consider detailed studies of this effect to be beyond the scope of the present work. In the particular case of the Z boson mass spectrum, we expect the size of this effect to be small in comparison to the overall uncertainty since a fairly large amount of data was used to constraint only 5 free parameters (the 4 parameters of the Crystal Ball kernel and the unknown proportionality constant).

(a) Iteratively bias−corrected Gaussian intervals



(b) Iteratively bias−corrected percentile intervals

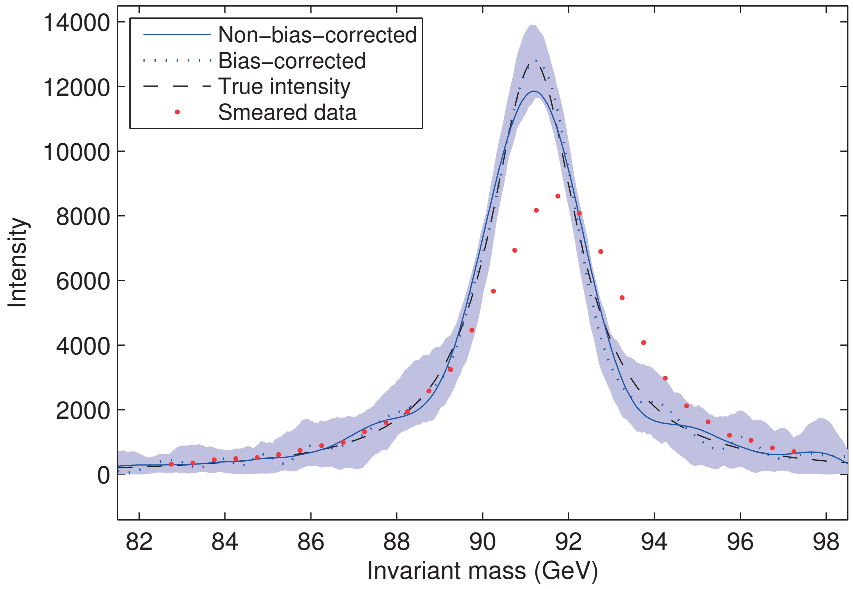


Figure 6.15: Data-driven iteratively bias-corrected confidence intervals in unfolding the Z boson invariant mass spectrum measured in the CMS experiment at the LHC. Figure (a) shows the Gaussian intervals based on the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$ and Figure (b) the percentile intervals based on the posterior mean $\hat{\boldsymbol{\beta}}_{\mathrm{PM}}$. The number of bias-correction iterations was calibrated to yield 94 % pointwise target coverage for the Gaussian intervals. The nominal coverage was 95 % and the regularization strength chosen using the MMLE.

7 Shape-constrained uncertainty quantification for steeply falling spectra

In this chapter, we study unfolded uncertainty quantification in situations where the true spectrum f is known to satisfy qualitative shape constraints in the form of positivity, monotonicity and convexity. In particular, we construct regularized confidence intervals by exploiting the fact that a large fraction of LHC spectra are known to be *steeply falling*. Such spectra are positive, decreasing and, in most cases, also convex.

Typical examples of steeply falling spectra are the energy and transverse momentum spectra of particle interactions. Recent LHC analyses that involve unfolding such spectra include the measurement of the differential cross section of jets (CMS Collaboration, 2013b), top quark pairs (CMS Collaboration, 2013c), the W boson (ATLAS Collaboration, 2012b) and the Higgs boson (CMS Collaboration, 2016). More precise measurement of these and other steeply falling particle spectra is the subject of several ongoing physics analyses at the LHC.

We form shape-constrained confidence intervals using the *strict bounds construction* described in Stark (1992). In other words, we form the confidence intervals by considering all those true intensities f that satisfy the shape constraints and fit the smeared observations $\boldsymbol{y}$ within a given confidence level. This enables us to form confidence intervals for functionals of f with *guaranteed simultaneous frequentist finite-sample coverage*. To the best of our knowledge, this construction is the first one to yield usefully tight unfolded confidence intervals with rigorous coverage guarantees.

We first explain in Section 7.1 how shape constraints can be used to regularize the unfolding of steeply falling spectra. We then give in Section 7.2 an outline of the strict bounds construction, while Section 7.3 provides details of the construction for shape-constrained unfolding. We demonstrate the resulting confidence intervals in Section 7.4 by unfolding the steeply falling inclusive jet transverse momentum spectrum described in Section 3.4.2.

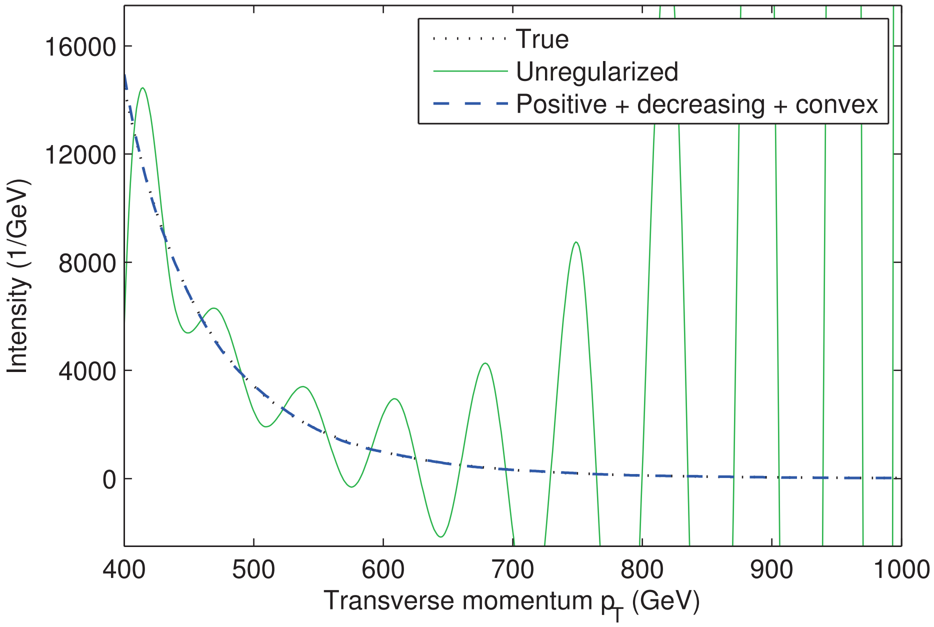


Figure 7.1: Illustration of using shape constraints to unfold a steeply falling particle spectrum. The unregularized solution (here a maximum likelihood spline fit) exhibits unphysical oscillations, but when the solution is constrained to be positive, decreasing and convex, it becomes almost indistinguishable from the true solution.

7.1 Regularization of steeply falling spectra using shape constraints

A priori shape information provides strong, physically justified regularization for the otherwise ill-posed unfolding problem. This is illustrated in Figure 7.1, which shows the inclusive jet transverse momentum spectrum of Section 3.4.2 unfolded with and without positivity, monotonicity and convexity constraints. The curves in the figure are maximum likelihood estimates for the spline discretization with $L = 16$ uniformly placed interior knots and the shape constraints were applied by imposing Equations (3.14)–(3.16) on the spline coefficients. The unregularized solution exhibits large unphysical oscillations, while the shape-constrained solution can be barely distinguished from the true spectrum f. For increasing dimension of the spline basis, the oscillations in the unregularized solution become even larger, but the shape-constrained solution remains well-behaved. The spline basis is used here only for illustration purposes; the shape-constrained uncertainty quantification method developed in this chapter does not require f to be a spline.

There is an extensive literature on using shape-constraints to regularize non-parametric point estimates; see, e.g., Robertson et al. (1988) and Groeneboom and Jongbloed (2014). The seminal work on this topic is by Grenander (1956), who derived the nonparametric maximum likelihood estimator of a density subject to a monotonicity constraint. However, most of this work has focused on point estimation of density or regression functions without measurement error and using shape constraints to regularize deconvolution-type problems has received only limited attention, with contributions by Wahba (1982), Stark (1992), Carroll et al. (2011) and Pflug and Wets (2013). Especially the case of constructing shape-constrained confidence intervals for indirect observations, which is our main topic of interest in the present chapter, has been rarely treated in the literature, with the exception of Stark (1992), who provides a general prescription for constructing shape-constrained finite-sample confidence intervals. A similar construction is also sketched by Rust and Burrus (1972).

To the author's best knowledge, monotonicity and convexity constraints have not been previously used to regularize unfolding of steeply falling particle spectra in high energy physics. There exists a related line of work in nuclear spectroscopy, where Burrus (1965), Burrus and Verbinski (1969) and O'Leary and Rust (1986) use a construction similar to ours to construct positivity-constrained unfolded confidence intervals. However, these authors consider situations where the true spectrum contains one or more peaks, and hence monotonicity or convexity constraints would not be appropriate. This is in contrast with differential cross section measurements at the LHC, where the spectra are typically known to be decreasing and usually also convex.

7.2 Outline of the strict bounds construction

We form the shape-constrained confidence intervals using the strict bounds construction of Stark (1992). This is a generic way of forming confidence intervals with guaranteed simultaneous coverage for a set of functionals of f given the a priori information that $f \in C$, where $C \subseteq V$. The a priori information we have in mind is that C consists of those functions that satisfy the desired shape constraints, but, for the present section, C may be any subset of V.

Recall from Section 3.2.1 the semi-discrete statistical model

$$\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{\mu}), \text{ with } \boldsymbol{\mu} = \mathcal{K} f, \tag{7.1}$$

where $\mathcal{K} : V \to \mathbb{R}^n,\ f \mapsto \left[K_1 f, \ldots, K_n f\right]^{\mathrm{T}}$, with the forward functionals K_j, $j = 1, \ldots, n$, given by Equation (3.9). Given this model, our aim in this section is to form simultaneous confidence intervals for the vector

$$\boldsymbol{\lambda} = \left[H_1 f, \ldots, H_p f\right]^{\mathrm{T}}, \tag{7.2}$$

where $H_k : V \to \mathbb{R}, f \mapsto H_k f$, $k = 1, \ldots, p$, are functionals corresponding to features of f that are of scientific interest. We are particularly interested in functionals H_k that correspond to the binned means of the true Poisson point process M,

$$H_k : V \to \mathbb{R}, f \mapsto \int_{E_k} f(s)\,\mathrm{d}s, \tag{7.3}$$

where the true bins E_k are as in Equation (3.17). The construction of the present section,

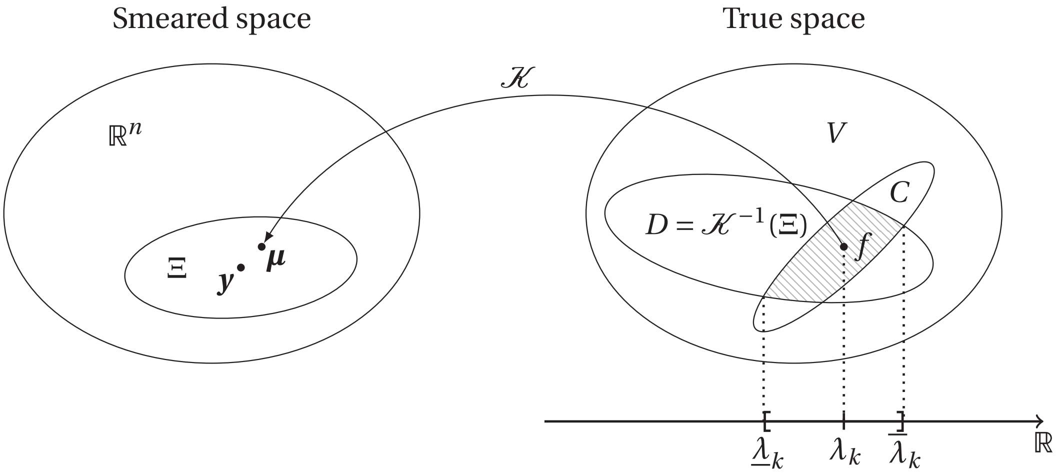


Figure 7.2: Illustration of the strict bounds construction (Stark, 1992). The set Ξ is a $1-\alpha$ simultaneous confidence set for $\boldsymbol{\mu}$ based on the smeared data $\boldsymbol{y}$. Its preimage D is a $1-\alpha$ confidence set for f. If $f \in C$, then $C \cap D$ is a regularized $1-\alpha$ confidence set for f. The extremal values of the functionals of interest over $C \cap D$ yield the strict bound confidence intervals $[\underline{\lambda}_k, \overline{\lambda}_k]$, $k = 1, \ldots, p$, which have conservative $1-\alpha$ simultaneous coverage for the quantities of interest λ_k, $k = 1, \ldots, p$.

however, is valid for any functionals H_k, including, for example, point evaluators $H_k f = f(s_k)$, derivatives $H_k f = f'(s_k)$ or nonlinear functionals, such as the location of the mode of f.

The strict bounds construction of confidence intervals for $\boldsymbol{\lambda}$ is illustrated in Figure 7.2. The construction proceeds as follows:

1. For $\alpha \in (0,1)$, we first form a $1-\alpha$ simultaneous confidence set for the smeared mean $\boldsymbol{\mu}$ based on the smeared data $\boldsymbol{y}$. Let us denote this set by Ξ.

2. We then look at $\mathscr{K}^{-1}(\Xi)$, the preimage of Ξ under the forward mapping $\mathscr{K}$. This is a $1-\alpha$ confidence set for the true intensity f in the unfolded space. Let us denote this set by D, that is, $D = \mathscr{K}^{-1}(\Xi)$.

3. We then regularize the confidence set D by intersecting it with the a priori constraints $C \subseteq V$. Provided that the actual intensity function f satisfies these constraints, the set $C \cap D$ is a $1-\alpha$ confidence set for f.

4. We then look at the extremal values of the functionals of interest H_k over $C \cap D$. That is, for each $k = 1, \ldots, p$, we compute

$$\underline{\lambda}_k = \inf_{f \in C \cap D} H_k f \quad \text{and} \quad \overline{\lambda}_k = \sup_{f \in C \cap D} H_k f. \tag{7.4}$$

   It follows that the set $[\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]$ is a $1-\alpha$ simultaneous confidence set for $\boldsymbol{\lambda} = \left[H_1 f, \ldots, H_p f\right]^{\mathrm{T}}$.

More formally, we have the following result:

Theorem 7.1 (Stark (1992))**.** *Let* $\boldsymbol{\mu} = \mathscr{K} f$ *for* $\mathscr{K} : V \to \mathbb{R}^n$ *and let* $\Xi \subseteq \mathbb{R}^n$ *be a random set satisfying* $\mathrm{P}_f\left(\boldsymbol{\mu} \in \Xi\right) \geq 1-\alpha, \forall f \in V$. *Denote* $D = \mathscr{K}^{-1}(\Xi)$ *and let*

$$\underline{\lambda}_k = \inf_{f \in C \cap D} H_k f \quad \text{and} \quad \overline{\lambda}_k = \sup_{f \in C \cap D} H_k f \quad \text{for} \quad k = 1, \ldots, p, \tag{7.5}$$

where H_k *are functionals on* V *and* $C \subseteq V$ *is such that* $f \in C$. *Then*

$$\mathrm{P}_f\left(\boldsymbol{\lambda} \in [\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]\right) \geq 1-\alpha, \forall f \in C, \tag{7.6}$$

where $\boldsymbol{\lambda} = \left[H_1 f, \ldots, H_p f\right]^{\mathrm{T}}$.

Proof. By definition, $f \in \mathscr{K}^{-1}(\Xi) \Leftrightarrow \mathscr{K} f \in \Xi \Leftrightarrow \boldsymbol{\mu} \in \Xi$. Hence $\mathrm{P}_f(\boldsymbol{\mu} \in \Xi) = \mathrm{P}_f(f \in \mathscr{K}^{-1}(\Xi)) = \mathrm{P}_f(f \in D)$. Since $\mathrm{P}_f(\boldsymbol{\mu} \in \Xi) \geq 1-\alpha, \forall f \in V$, we also have that $\mathrm{P}_f(f \in D) \geq 1-\alpha, \forall f \in V$. Since $C \subseteq V$,

$$\begin{aligned} \mathrm{P}_f(f \in D) \geq 1-\alpha, \forall f \in V &\Rightarrow \mathrm{P}_f(f \in D) \geq 1-\alpha, \forall f \in C & (7.7)\\ &\Rightarrow \mathrm{P}_f\left(\{f \in C\} \cap \{f \in D\}\right) \geq 1-\alpha, \forall f \in C & (7.8)\\ &\Rightarrow \mathrm{P}_f(f \in C \cap D) \geq 1-\alpha, \forall f \in C. & (7.9) \end{aligned}$$

We also have

$$\begin{aligned} f \in C \cap D &\Rightarrow \inf_{f' \in C \cap D} H_k f' \leq H_k f \leq \sup_{f' \in C \cap D} H_k f', \forall k & (7.10)\\ &\Rightarrow \underline{\lambda}_k \leq \lambda_k \leq \overline{\lambda}_k, \forall k \Rightarrow \boldsymbol{\lambda} \in [\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]. & (7.11) \end{aligned}$$

Hence

$$\mathrm{P}_f\left(\boldsymbol{\lambda} \in [\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]\right) \geq \mathrm{P}_f(f \in C \cap D) \tag{7.12}$$

and by Equation (7.9) we have that $\mathrm{P}_f\left(\boldsymbol{\lambda} \in [\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]\right) \geq 1-\alpha, \forall f \in C$. □

Notice that as long as it is known by physical considerations that $f \in C$, then $[\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]$ is a regularized confidence set with *guaranteed simultaneous finite-sample coverage*. However, the inequality in Equation (7.12) is generally strict and, as a result, the coverage probability of $[\underline{\lambda}_1, \overline{\lambda}_1] \times \cdots \times [\underline{\lambda}_p, \overline{\lambda}_p]$ is usually strictly greater than $1-\alpha$. In other words, the construction is conservative in the sense that the resulting confidence set may have a non-negligible amount of overcoverage, but it cannot undercover for any $f \in C$.

Notice also that this construction is fully generic in the sense that it does not depend on the specific form of the constraints C or the functionals H_k. In the rest of this chapter, we use the construction to form shape-constrained confidence intervals, but in principle the approach could also be applied with a smoothness constraint of the form $C = \{f \in V : \|f''\|_2^2 \le \tau\}$. However, the coverage would only be guaranteed if the upper bound τ is known, which is usually not the case.

With the strict bounds construction, the problem of forming confidence intervals for $\boldsymbol{\lambda}$ reduces to solving the optimization problems $\inf_{f \in C \cap D} H_k f$ and $\sup_{f \in C \cap D} H_k f$ for $k = 1, \ldots, p$. Since

$$\sup_{f \in C \cap D} H_k f = -\inf_{f \in C \cap D} -H_k f, \tag{7.13}$$

we can without loss of generality focus on the minimization problem. Solving the minimization problem $\inf_{f \in C \cap D} H_k f$ is non-trivial since it involves an infinite-dimensional unknown subject to an infinite set of constraints. We can nevertheless follow the approach of Stark (1992) to find a conservative solution for this problem. Namely, we use Fenchel duality to turn the infinite-dimensional minimization problem into a semi-infinite maximization problem with an n-dimensional unknown and an infinite set of constraints. We then discretize the constraints in such a way that any feasible point of the resulting finite-dimensional maximization problem is guaranteed to yield a lower bound for $\inf_{f \in C \cap D} H_k f$. In other words, any such feasible point provides a conservative confidence bound, with confidence level at least $1 - \alpha$.

7.3 Unfolding with shape-constrained strict bounds

In this section, we explain in detail the construction of the shape-constrained strict bounds for the unfolding problem. In doing so, we extend the existing methodology of Stark (1992) to handle Poisson noise and present a novel way of imposing and discretizing the monotonicity and convexity constraints. We first explain in Section 7.3.1 how to use Garwood intervals to construct the smeared confidence set Ξ. In Section 7.3.2, we use Fenchel duality to turn the infinite-dimensional primal program $\inf_{f \in C \cap D} H_k f$ into a semi-infinite dual program. Section 7.3.3 derives the explicit form of the dual constraints for different shape constraints and Section 7.3.4 explains how these constraints can be discretized conservatively so that the confidence level is maintained for the resulting finite-dimensional optimization problems. The construction is summarized as a theorem in Section 7.3.5.

From this point onwards, we assume that the functionals H_k are linear and that the constraint set C is convex. Further assumptions will be stated as we proceed.

7.3.1 Confidence set in the smeared space

The first step of the strict bounds construction is to form the confidence set Ξ for the smeared mean $\boldsymbol{\mu}$ under the model $\boldsymbol{y} \sim \mathrm{Poisson}(\boldsymbol{\mu})$. This is straightforward since the components of $\boldsymbol{y}$ are independent. For each $j = 1, \ldots, n$ and for $\alpha' \in (0,1)$, let $[\underline{\mu}_{j,\alpha'}, \overline{\mu}_{j,\alpha'}]$ be a $1 - \alpha'$ confidence interval for μ_j and assume that the interval depends on y_j only. Then

$$\mathrm{P}_f\left(\underline{\mu}_{1,\alpha'} \leq \mu_1 \leq \overline{\mu}_{1,\alpha'}, \ldots, \underline{\mu}_{n,\alpha'} \leq \mu_n \leq \overline{\mu}_{n,\alpha'}\right) = \prod_{j=1}^{n} \mathrm{P}_f\left(\underline{\mu}_{j,\alpha'} \leq \mu_j \leq \overline{\mu}_{j,\alpha'}\right) \geq (1-\alpha')^n. \quad (7.14)$$

Hence, by setting $\alpha' = 1-(1-\alpha)^{1/n}$ with $\alpha \in (0,1)$, we have that $\Xi = \left[\underline{\mu}_{1,\alpha'}, \overline{\mu}_{1,\alpha'}\right] \times \cdots \times \left[\underline{\mu}_{n,\alpha'}, \overline{\mu}_{n,\alpha'}\right]$ is a $1-\alpha$ simultaneous confidence set for $\boldsymbol{\mu}$, that is, $\mathrm{P}_f\left(\boldsymbol{\mu} \in \Xi\right) \geq 1-\alpha$.

We use the Garwood construction (Garwood, 1936) to form the binwise intervals $\left[\underline{\mu}_{j,\alpha'}, \overline{\mu}_{j,\alpha'}\right]$. For each bin with a strictly positive event count, the $1-\alpha'$ Garwood intervals are

$$\underline{\mu}_{j,\alpha'} = \frac{1}{2} F_{\chi^2}^{-1}\left(\frac{\alpha'}{2}; 2y_j\right) \quad \text{and} \quad \overline{\mu}_{j,\alpha'} = \frac{1}{2} F_{\chi^2}^{-1}\left(1-\frac{\alpha'}{2}; 2(y_j+1)\right), \quad (7.15)$$

where $F_{\chi^2}^{-1}(\cdot\,; k)$ is the quantile function (i.e., the inverse cumulative distribution function) of the χ^2 distribution with k degrees of freedom. If $y_j = 0$, the upper bound is given by Equation (7.15), but the lower bound is zero, $\underline{\mu}_{j,\alpha'} = 0$. The Garwood intervals have guaranteed $1-\alpha'$ confidence level, but, due to the discreteness of the Poisson distribution, the actual coverage probability is strictly greater than $1-\alpha'$ for any finite μ_j (see Heinrich (2003, Section 6) for a plot of the coverage probability of these intervals as a function of μ_j).

For the following sections, it will be convenient to write the hyperrectangle Ξ using its center point $\tilde{\boldsymbol{y}}$, that is, $\Xi = \left\{\tilde{\boldsymbol{y}} + \boldsymbol{\xi} \in \mathbb{R}^n : \|\mathrm{diag}(\boldsymbol{l})^{-1}\boldsymbol{\xi}\|_\infty \leq 1\right\}$, where, for each $j = 1, \ldots, n$, $\tilde{y}_j = \left(\underline{\mu}_{j,\alpha'} + \overline{\mu}_{j,\alpha'}\right)/2$ and $l_j = \left(\overline{\mu}_{j,\alpha'} - \underline{\mu}_{j,\alpha'}\right)/2$.

7.3.2 Strict bounds via duality

Our next task is to find a way to compute a conservative value for the lower bound $\inf_{f \in C \cap D} H_k f$. We follow the approach of Stark (1992) and solve the problem using *Fenchel duality* (Luenberger, 1969, Section 7.12). For the primal problem

$$v(\mathscr{P}) = \inf_{f \in C \cap D} H_k f, \quad (7.16)$$

the Fenchel dual is given by

$$v(\mathscr{D}) = \sup_{f^* \in C^* \cap D^*} \left\{ \inf_{f \in D} f^*[f] + \inf_{f \in C} (H_k - f^*)[f] \right\}, \quad (7.17)$$

where $C^* = \left\{f^* \in V^* : \inf_{f \in C}(H_k - f^*)[f] > -\infty\right\}$, $D^* = \left\{f^* \in V^* : \inf_{f \in D} f^*[f] > -\infty\right\}$ and V^* is the algebraic dual space of V, that is, the set of all linear functionals on V. Here the supremum over an empty set is defined to be $-\infty$. Notice that the Fenchel dual divides the dependence on C and D into two separate terms.

As shown for example in Stark (1992, Section 5), $v(\mathscr{P})$ and $v(\mathscr{D})$ satisfy weak duality:

Theorem 7.2 (Weak Fenchel duality)**.** *Let the primal problem $v(\mathscr{P})$ and the dual problem $v(\mathscr{D})$ be as in Equations* (7.16) *and* (7.17)*. Then $v(\mathscr{P}) \geq v(\mathscr{D})$.*

Proof. For any $f^* \in V^*$,

$$\inf_{f\in C\cap D} H_k f = \inf_{f\in C\cap D} \{f^*[f] + (H_k - f^*)[f]\} \tag{7.18}$$

$$\geq \inf_{f\in C\cap D} f^*[f] + \inf_{f\in C\cap D} (H_k - f^*)[f] \tag{7.19}$$

$$\geq \inf_{f\in D} f^*[f] + \inf_{f\in C} (H_k - f^*)[f]. \tag{7.20}$$

We have hence established the inequality

$$v(\mathscr{P}) = \inf_{f\in C\cap D} H_k f \geq \sup_{f^*\in V^*} \left\{ \inf_{f\in D} f^*[f] + \inf_{f\in C} (H_k - f^*)[f] \right\} \tag{7.21}$$

$$= \sup_{f^*\in C^*\cap D^*} \left\{ \inf_{f\in D} f^*[f] + \inf_{f\in C} (H_k - f^*)[f] \right\} = v(\mathscr{D}). \tag{7.22}$$

□

Weak duality suffices to guarantee that the solution of dual problem (7.17) provides a conservative confidence bound. However, under further technical regularity conditions, detailed in Luenberger (1969, Section 7.12) and Stark (1992, Section 10.1), one can also establish strong duality $v(\mathscr{P}) = v(\mathscr{D})$, in which case there is no slack from solving the dual problem instead of the primal.

It turns out that the Fenchel dual (7.17) can be written using a finite-dimensional unknown. Namely, by Stark (1992, Section 5) and Backus (1970), the set D^* consists of those functionals that are linear combinations of the forward functionals K_j,

$$D^* = \{f^* \in V^* : f^* = \boldsymbol{\nu}\cdot\mathscr{K}, \boldsymbol{\nu} \in \mathbb{R}^n\}, \tag{7.23}$$

where $\boldsymbol{\nu}\cdot\mathscr{K} = \sum_{j=1}^n \nu_j K_j$. The dual problem hence becomes

$$v(\mathscr{D}) = \sup_{\boldsymbol{\nu}\in\mathbb{R}^n : \boldsymbol{\nu}\cdot\mathscr{K}\in C^*} \left\{ \inf_{f\in D} (\boldsymbol{\nu}\cdot\mathscr{K})[f] + \inf_{f\in C} (H_k - \boldsymbol{\nu}\cdot\mathscr{K})[f] \right\}. \tag{7.24}$$

By a simple modification of the argument given in Stark (1992, Section 5), we find that the first term in (7.24) can be expressed in closed form:

Lemma 7.3. *Let $D = \mathscr{K}^{-1}(\Xi)$, where $\Xi = \{\tilde{\boldsymbol{y}} + \boldsymbol{\xi} \in \mathbb{R}^n : \|\mathrm{diag}(\boldsymbol{l})^{-1}\boldsymbol{\xi}\|_\infty \leq 1\}$. Then*

$$\inf_{f\in D} (\boldsymbol{\nu}\cdot\mathscr{K})[f] \geq \boldsymbol{\nu}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \|\boldsymbol{\nu}\|_1^{\boldsymbol{l}}, \tag{7.25}$$

where $\|\boldsymbol{\nu}\|_1^{\boldsymbol{l}} = \|\mathrm{diag}(\boldsymbol{l})\boldsymbol{\nu}\|_1$ is the weighted ℓ^1-norm. If the forward functionals $\{K_j\}_{j=1}^n$ are linearly independent, then (7.25) *holds with equality.*

Proof. When $f \in D$,

$$(\boldsymbol{v} \cdot \mathscr{K})[f] = \boldsymbol{v}^{\mathrm{T}}(\tilde{\boldsymbol{y}} + \boldsymbol{\xi}) \geq \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - |\boldsymbol{v}^{\mathrm{T}}\boldsymbol{\xi}| = \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - |(\mathrm{diag}(\boldsymbol{l})\boldsymbol{v})^{\mathrm{T}}(\mathrm{diag}(\boldsymbol{l})^{-1}\boldsymbol{\xi})| \tag{7.26}$$

$$\geq \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \|\mathrm{diag}(\boldsymbol{l})\boldsymbol{v}\|_1 \|\mathrm{diag}(\boldsymbol{l})^{-1}\boldsymbol{\xi}\|_\infty \geq \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \|\boldsymbol{v}\|_1^{\boldsymbol{l}}. \tag{7.27}$$

Hence,

$$\inf_{f \in D} (\boldsymbol{v} \cdot \mathscr{K})[f] \geq \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \|\boldsymbol{v}\|_1^{\boldsymbol{l}}. \tag{7.28}$$

To show that this lower bound is sharp, we employ the result of Stark (1992, Appendix A) which shows that if $\{K_j\}_{j=1}^n$ are linearly independent, then there exists functions $\{\tilde{f}_i\}_{i=1}^n$ in V such that $K_j[\tilde{f}_i] = \delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker delta, that is, $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$.

We shall show that $\tilde{f} = \sum_{j=1}^n \beta_j \tilde{f}_j$, where $\beta_j = \tilde{y}_j - \frac{v_j}{|v_j|} l_j$, attains the lower bound in (7.28). We have

$$\mathscr{K}[\tilde{f}] = \mathscr{K}\Big[\sum_{j=1}^n \beta_j \tilde{f}_j\Big] = \sum_{j=1}^n \beta_j \mathscr{K}[\tilde{f}_j] = \sum_{j=1}^n \beta_j \boldsymbol{e}_j = \sum_{j=1}^n \left(\tilde{y}_j - \frac{v_j}{|v_j|} l_j\right) \boldsymbol{e}_j = \tilde{\boldsymbol{y}} + \boldsymbol{u}, \tag{7.29}$$

where $\boldsymbol{u} = -\sum_{j=1}^n \frac{v_j}{|v_j|} l_j \boldsymbol{e}_j$ and $\boldsymbol{e}_j \in \mathbb{R}^n$ has 1 in the jth position and 0 elsewhere. Since $\|\mathrm{diag}(\boldsymbol{l})^{-1}\boldsymbol{u}\|_\infty = 1$, we have that $\mathscr{K}[\tilde{f}] = \tilde{\boldsymbol{y}} + \boldsymbol{u} \in \Xi$ and hence $\tilde{f} \in D$. Moreover,

$$(\boldsymbol{v} \cdot \mathscr{K})[\tilde{f}] = \boldsymbol{v}^{\mathrm{T}}(\mathscr{K}[\tilde{f}]) = \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \sum_{j=1}^n \frac{v_j^2}{|v_j|} l_j = \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \sum_{j=1}^n |v_j| l_j = \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \|\boldsymbol{v}\|_1^{\boldsymbol{l}}, \tag{7.30}$$

and hence $\tilde{f}$ attains the lower bound and the bound is sharp. □

We have hence established the inequality

$$\inf_{f \in C \cap D} H_k f \geq \sup_{\boldsymbol{v} \in \mathbb{R}^n : \boldsymbol{v} \cdot \mathscr{K} \in C^*} \left\{ \boldsymbol{v}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \|\boldsymbol{v}\|_1^{\boldsymbol{l}} + \inf_{f \in C} (H_k - \boldsymbol{v} \cdot \mathscr{K})[f] \right\}, \tag{7.31}$$

which holds as an equality under regularity conditions. If the inequality is strict, the right-hand side still yields a valid conservative bound.

We next characterize the set C^* under the assumption that C is a *convex cone*, that is, C is convex and satisfies that if $f \in C$, then $\gamma f \in C$ for all $\gamma \geq 0$. This is satisfied for all the shape constraints that we consider. We then have by Stark (1992, Section 6.2) the following result:

Lemma 7.4. *Let C be a convex cone. Then the set $C^* = \{f^* \in V^* : \inf_{f \in C}(H_k - f^*)[f] > -\infty\}$ is equivalently given by*

$$C^* = \{f^* \in V^* : \inf_{f \in C}(H_k - f^*)[f] = 0\} = \{f^* \in V^* : (H_k - f^*)[f] \geq 0, \forall f \in C\}. \tag{7.32}$$

The second equality in Equation (7.32) is trivial and the first one follows by noting that if for any $f \in C$, $(H_k - f^*)[f] < 0$, then $(H_k - f^*)[\gamma f] = \gamma(H_k - f^*)[f]$ can be made arbitrarily small by taking $\gamma \to \infty$.

Hence, when C is a convex cone, the dual problem in Equation (7.31) simplifies further and is given by

$$\sup_{\boldsymbol{\nu} \in \mathbb{R}^n} \left\{ \boldsymbol{\nu}^{\mathrm{T}} \tilde{\boldsymbol{y}} - \|\boldsymbol{\nu}\|_1^{\boldsymbol{l}} \right\} \quad \text{subject to} \quad (H_k - \boldsymbol{\nu} \cdot \mathscr{K})[f] \geq 0,\ \forall f \in C. \tag{7.33}$$

This is a semi-infinite program with an n-dimensional unknown and an infinite set of constraints.

7.3.3 Constraints of the dual program

In this section, we rewrite the dual constraint

$$(H_k - \boldsymbol{\nu} \cdot \mathscr{K})[f] \geq 0,\ \forall f \in C, \tag{7.34}$$

in an equivalent form that does not involve f. We do this for functionals H_k of the form (7.3) and for the following shape constraints:

(P) f is positive, $C = \{f \in V : f(s) \geq 0, \forall s \in E\}$;

(D) f is positive and decreasing, $C = \{f \in V : f(s) \geq 0 \wedge f'(s) \leq 0, \forall s \in E\}$;

(C) f is positive, decreasing and convex, $C = \{f \in V : f(s) \geq 0 \wedge f'(s) \leq 0 \wedge f''(s) \geq 0, \forall s \in E\}$.

The positivity constraint (P) holds for any Poisson intensity function f, while the monotonicity constraint (D) and the convexity constraint (C) correspond to shapes that are typically expected for steeply falling particle spectra.

Let the functional H_k be given by Equation (7.3). Then the left-hand side of the dual constraint (7.34) can be rewritten as

$$(H_k - \boldsymbol{\nu} \cdot \mathscr{K})[f] = H_k f - \sum_{j=1}^{n} \nu_j K_j f = \int_{E_k} f(s)\,\mathrm{d}s - \sum_{j=1}^{n} \nu_j \int_E k_j(s) f(s)\,\mathrm{d}s \tag{7.35}$$

$$= \int_E \left(\mathbf{1}_{E_k}(s) - \sum_{j=1}^{n} \nu_j k_j(s) \right) f(s)\,\mathrm{d}s = \int_E h_k(s) f(s)\,\mathrm{d}s, \tag{7.36}$$

where $\mathbf{1}_{E_k}$ is the indicator function of E_k and we have denoted

$$h_k(s) = \mathbf{1}_{E_k}(s) - \sum_{j=1}^{n} \nu_j k_j(s). \tag{7.37}$$

The dual constraint hence becomes $\int_E h_k(s) f(s)\,\mathrm{d}s \geq 0$, $\forall f \in C$.

We make the following assumptions concerning the function space V and the integration kernels k_j in Equation (3.9):

(A1) V consists of twice continuously differentiable functions on E, $V = C^2(E)$;

(A2) The forward kernels k_j are continuous on E, $k_j \in C(E)$, $j = 1, \ldots, n$.

Notice that assumption (A2) implies that the functions h_k are right-continuous on E. We also mention that assumption (A1) can be relaxed, at least for the positivity constraint (P) and the monotonicity constraint (D), but for simplicity we prefer to work with (A1). We then have the following result:

Lemma 7.5. *Assume (A1) and (A2). Then, for the shape constraints (P), (D) and (C), the dual constraint* $\int_E h_k(s) f(s)\,\mathrm{d}s \geq 0,\ \forall f \in C$, *can be equivalently written as*

$$\text{(P)}\quad h_k(s) \geq 0,\ \forall s \in E; \tag{7.38}$$

$$\text{(D)}\quad \int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s' \geq 0,\ \forall s \in E; \tag{7.39}$$

$$\text{(C)}\quad \int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s' \geq 0,\ \forall s \in E \ \wedge\ \int_E h_k(s)\,\mathrm{d}s \geq 0. \tag{7.40}$$

Proof. The result for the positivity constraint (P) follows directly. For (D) and (C), the proof employs integration by parts:

(D) We need to show

$$\int_E h_k(s) f(s)\,\mathrm{d}s \geq 0,\ \forall f \in C \quad \Leftrightarrow \quad \int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s' \geq 0,\ \forall s \in E, \tag{7.41}$$

where C corresponds to the monotonicity constraint (D). Integration by parts gives

$$\int_E h_k(s) f(s)\,\mathrm{d}s = \int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s' f(s)\bigg|_{E_{\min}}^{E_{\max}} - \int_E \int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s' f'(s)\,\mathrm{d}s \tag{7.42}$$

$$= \int_E h_k(s)\,\mathrm{d}s\, f(E_{\max}) - \int_E \int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s' f'(s)\,\mathrm{d}s. \tag{7.43}$$

It is clear from this form that the right-hand side of Equation (7.41) implies the left-hand side. To show the converse, assume that $\int_{E_{\min}}^{s^*} h_k(s')\,\mathrm{d}s' < 0$ for some s^* in the interior of E. Then, by the continuity of the integral, $\int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s' < 0$ for all $s \in (s^* - \delta, s^* + \delta)$ for some $\delta > 0$. Let us consider a function $d \in C$ which is a strictly positive constant on the interval $[E_{\min}, s^* - \delta]$ and zero on $[s^* + \delta, E_{\max}]$. Substituting d into Equation (7.43) yields

$$\int_E h_k(s) d(s)\,\mathrm{d}s = -\int_{s^*-\delta}^{s^*+\delta} \int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s' d'(s)\,\mathrm{d}s < 0, \tag{7.44}$$

which is a contradiction. Hence $\int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s' \geq 0$ for all s in the interior of E and, by the continuity of the integral, for all $s \in E$.

(C) We need to show

$$\int_E h_k(s) f(s)\,\mathrm{d}s \geq 0,\ \forall f \in C \tag{7.45}$$

$$\Leftrightarrow \quad \int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s' \geq 0,\ \forall s \in E \quad \wedge \quad \int_E h_k(s)\,\mathrm{d}s \geq 0, \tag{7.46}$$

where C corresponds to the convexity constraint (C). A second application of integration by parts in Equation (7.43) gives

$$\int_E h_k(s) f(s)\,\mathrm{d}s = \int_E h_k(s)\,\mathrm{d}s\, f(E_{\max}) - \int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s'\, f'(s) \Big|_{E_{\min}}^{E_{\max}} \tag{7.47}$$

$$+ \int_E \int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s'\, f''(s)\,\mathrm{d}s \tag{7.48}$$

$$= \int_E h_k(s)\,\mathrm{d}s\, f(E_{\max}) - \int_{E_{\min}}^{E_{\max}} \int_{E_{\min}}^{s} h_k(s')\,\mathrm{d}s'\,\mathrm{d}s\, f'(E_{\max}) \tag{7.49}$$

$$+ \int_E \int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s'\, f''(s)\,\mathrm{d}s. \tag{7.50}$$

We see from this expression that the right-hand side of Equation (7.45) implies the left-hand side. To show the converse, take $d \in C$ such that $d(s) = a > 0$ for all $s \in E$. Substituting this into the left-hand side implies that we must have $\int_E h_k(s)\,\mathrm{d}s \geq 0$. Let us then assume that $\int_{E_{\min}}^{s^*} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s' < 0$ for some s^* in the interior of E. By the continuity of the integral, it follows that $\int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s' < 0$ for all $s \in (s^* - \delta, s^* + \delta)$ for some $\delta > 0$. Let us consider a function $d \in C$ which is linear and strictly decreasing on the interval $[E_{\min}, s^* - \delta]$ and zero on $[s^* + \delta, E_{\max}]$. Substituting d into Equation (7.50) gives

$$\int_E h_k(s) d(s)\,\mathrm{d}s = \int_{s^*-\delta}^{s^*+\delta} \int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s'\, d''(s)\,\mathrm{d}s < 0, \tag{7.51}$$

which is a contradiction. Hence $\int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} h_k(s'')\,\mathrm{d}s''\,\mathrm{d}s' \geq 0$ for all s in the interior of E and, by the continuity of the integral, for all $s \in E$.

□

By substituting h_k from Equation (7.37) to the results of Lemma 7.5, we find that the constraints on $\boldsymbol{\nu}$ in the dual problem $\sup_{\boldsymbol{\nu}\in\mathbb{R}^n} \left\{\boldsymbol{\nu}^{\mathrm{T}} \tilde{\boldsymbol{y}} - \|\boldsymbol{\nu}\|_1^{\boldsymbol{l}}\right\}$ are given by

$$\text{(P)} \quad \sum_{j=1}^{n} \nu_j k_j(s) \leq L_k^{\mathrm{P}}(s),\ \forall s \in E; \tag{7.52}$$

$$\text{(D)} \quad \sum_{j=1}^{n} \nu_j k_j^*(s) \leq L_k^{\mathrm{D}}(s),\ \forall s \in E; \tag{7.53}$$

$$\text{(C)} \quad \sum_{j=1}^{n} \nu_j k_j^{**}(s) \leq L_k^{\mathrm{C}}(s),\ \forall s \in E \ \wedge\ \sum_{j=1}^{n} \nu_j k_j^*(E_{\max}) \leq E_{k,\max} - E_{k,\min}, \tag{7.54}$$

where

$$k_j^*(s) = \int_{E_{\min}}^{s} k_j(s')\,\mathrm{d}s', \tag{7.55}$$

$$k_j^{**}(s) = \int_{E_{\min}}^{s} \int_{E_{\min}}^{s'} k_j(s'')\,\mathrm{d}s''\,\mathrm{d}s', \tag{7.56}$$

and the functions L_k^{P}, L_k^{D} and L_k^{C} on the right-hand side are

$$L_k^{\mathrm{P}}(s) = \mathbf{1}_{E_k}(s), \tag{7.57}$$

$$L_k^{\mathrm{D}}(s) = \begin{cases} 0, \ s < E_{k,\min}, \\ s - E_{k,\min}, \ E_{k,\min} \le s < E_{k,\max}, \\ E_{k,\max} - E_{k,\min}, \ s \ge E_{k,\max}, \end{cases} \tag{7.58}$$

$$L_k^{\mathrm{C}}(s) = \begin{cases} 0, \ s < E_{k,\min}, \\ \frac{1}{2}(s - E_{k,\min})^2, \ E_{k,\min} \le s < E_{k,\max}, \\ \frac{1}{2}(E_{k,\max} - E_{k,\min})^2 + (E_{k,\max} - E_{k,\min})(s - E_{k,\max}), \ s \ge E_{k,\max}. \end{cases} \tag{7.59}$$

Here L_k^{P} is piecewise constant, L_k^{D} consists of two constant parts connected by a linear part and L_k^{C} has a constant and a linear part connected by a quadratic part.

We have hence derived explicit expressions for the constraints of the semi-infinite dual program corresponding to the lower bound $\underline{\lambda}_k = \inf_{f \in C \cap D} H_k f$. We can follow similar reasoning to find the dual program corresponding to the upper bound $\overline{\lambda}_k = \sup_{f \in C \cap D} H_k f = -\inf_{f \in C \cap D} -H_k f$. The end result is that $\overline{\lambda}_k$ is bounded from above by the solution (or any feasible point) of $\inf_{\boldsymbol{\nu} \in \mathbb{R}^n} -\{\boldsymbol{\nu}^{\mathrm{T}} \tilde{\boldsymbol{y}} - \|\boldsymbol{\nu}\|_1^{\boldsymbol{l}}\}$ subject to the constraints

$$\text{(P)} \quad \sum_{j=1}^{n} \nu_j k_j(s) \le -L_k^{\mathrm{P}}(s), \ \forall s \in E; \tag{7.60}$$

$$\text{(D)} \quad \sum_{j=1}^{n} \nu_j k_j^*(s) \le -L_k^{\mathrm{D}}(s), \ \forall s \in E; \tag{7.61}$$

$$\text{(C)} \quad \sum_{j=1}^{n} \nu_j k_j^{**}(s) \le -L_k^{\mathrm{C}}(s), \ \forall s \in E \ \wedge \ \sum_{j=1}^{n} \nu_j k_j^*(E_{\max}) \le E_{k,\min} - E_{k,\max}. \tag{7.62}$$

Notice that the constraints (7.60)–(7.62) and (7.52)–(7.54) differ only in the sign of the right-hand side.

7.3.4 Conservative discretization of the dual constraints

We now have explicit expressions for the semi-infinite dual programs corresponding to the lower bound $\underline{\lambda}_k$ and the upper bound $\overline{\lambda}_k$. However, these still cannot be easily solved on a computer due to the infinite set of constraints. In this section, we discretize these constraints in such a way that the confidence level is preserved. The resulting dual programs have a finite-dimensional unknown and a finite number of constraints and can be hence solved using standard numerical optimization tools.

In order to discretize the constraints, let $s_1 < s_2 < \cdots < s_m < s_{m+1}$ be a grid on E consisting of $m+1$ grid points, with $m \gg p$. We assume that $s_1 = E_{\min}$, $s_{m+1} = E_{\max}$ and that there is a grid point at each boundary between the true bins $\{E_k\}_{k=1}^{p}$. A first idea would be to simply impose the constraints on the grid points $\{s_i\}_{i=1}^{m+1}$. For example, for the positivity constraint (P), this would correspond to requiring that

$$\sum_{j=1}^{n} \nu_j k_j(s_i) \leq \pm L_k^{\mathrm{P}}(s_i), \quad i = 1, \ldots, m+1. \tag{7.63}$$

However, this does not guarantee that the confidence level is preserved. Indeed, the discretized feasible set would be larger than the original feasible set and the resulting confidence intervals could be too short.

To guarantee the confidence level, we need to discretize the constraints in such a way that the discretized feasible set is a subset of the original feasible set. This requires making sure that the constraints are also satisfied between the grid points $\{s_i\}_{i=1}^{m+1}$. We do this in the following sections by finding on each interval $[s_i, s_{i+1})$ a convenient upper bound for the left-hand side of the constraint relations (7.52)–(7.54) and (7.60)–(7.62) and then constraining this upper bound to be below the right-hand size functions $\pm L_k^{\mathrm{P}}$, $\pm L_k^{\mathrm{D}}$ or $\pm L_k^{\mathrm{C}}$.

7.3.4.1 Positive intensities

Let us first consider the positivity constraint (P) and the lower bound $\underline{\lambda}_k$, where the dual constraint is given by Equation (7.52). For each $j = 1, \ldots, n$, let us write $\nu_j = \nu_j^+ - \nu_j^-$ with $\nu_j^+, \nu_j^- \geq 0$. Then, for every $s \in [s_i, s_{i+1})$, we have the upper bound

$$\sum_{j=1}^{n} \nu_j k_j(s) = \sum_{j=1}^{n} \nu_j^+ k_j(s) - \sum_{j=1}^{n} \nu_j^- k_j(s) \tag{7.64}$$

$$\leq \sum_{j=1}^{n} \nu_j^+ \sup_{\xi \in [s_i, s_{i+1})} k_j(\xi) - \sum_{j=1}^{n} \nu_j^- \inf_{\xi \in [s_i, s_{i+1})} k_j(\xi) \tag{7.65}$$

$$= \sum_{j=1}^{n} \nu_j^+ \overline{\rho}_{i,j} - \sum_{j=1}^{n} \nu_j^- \underline{\rho}_{i,j}, \tag{7.66}$$

where we have denoted $\overline{\rho}_{i,j} = \sup_{\xi \in [s_i, s_{i+1})} k_j(\xi)$ and $\overline{\rho}_{i,j} = \inf_{\xi \in [s_i, s_{i+1})} k_j(\xi)$. This bounds the left-hand side of Equation (7.52) by a constant with respect to s on the interval $[s_i, s_{i+1})$. Since the right-hand side L_k^{P} is also constant on $[s_i, s_{i+1})$, we simply need to enforce $\sum_{j=1}^{n} \nu_j^+ \overline{\rho}_{i,j} - \sum_{j=1}^{n} \nu_j^- \underline{\rho}_{i,j} \leq L_k^{\mathrm{P}}(s_i)$, for $i = 1, \ldots, m$, to obtain a conservative discretization of (7.52). This construction is illustrated in Figure 7.3.

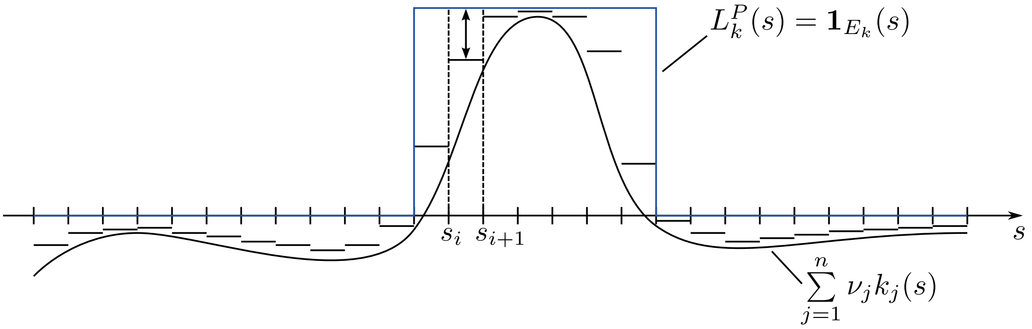

$L_k^P(s) = \mathbf{1}_{E_k}(s)$
s_i
s_{i+1}
s
$\sum_{j=1}^{n} \nu_j k_j(s)$

Figure 7.3: Illustration of the conservative discretization of the constraint $\sum_{j=1}^{n} \nu_j k_j(s) \le L_k^{\mathrm{P}}(s)$, for all $s \in E$, using a constant upper bound on each interval $[s_i, s_{i+1})$.

Let us arrange the scalars $\overline{\rho}_{i,j}$ and $\underline{\rho}_{i,j}$ into the matrix

$$\boldsymbol{A} = \begin{bmatrix} \overline{\rho}_{1,1} & \cdots & \overline{\rho}_{1,n} & -\underline{\rho}_{1,1} & \cdots & -\underline{\rho}_{1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \overline{\rho}_{m,1} & \cdots & \overline{\rho}_{m,n} & -\underline{\rho}_{m,1} & \cdots & -\underline{\rho}_{m,n} \end{bmatrix}. \tag{7.67}$$

Denote

$$\tilde{\boldsymbol{\nu}} = \begin{bmatrix} \boldsymbol{\nu}^+ \\ \boldsymbol{\nu}^- \end{bmatrix} \quad \text{with} \quad \boldsymbol{\nu}^+ = \begin{bmatrix} \nu_1^+ \\ \vdots \\ \nu_n^+ \end{bmatrix} \quad \text{and} \quad \boldsymbol{\nu}^- = \begin{bmatrix} \nu_1^- \\ \vdots \\ \nu_n^- \end{bmatrix}, \tag{7.68}$$

and let $\boldsymbol{b}_k^{\mathrm{P}} \in \mathbb{R}^m$ be the vector with components $b_{k,i}^{\mathrm{P}} = L_k^{\mathrm{P}}(s_i) = \mathbf{1}_{E_k}(s_i)$, $i = 1, \ldots, m$. Then the discretized dual constraint can be simply written as $\boldsymbol{A}\tilde{\boldsymbol{\nu}} \le \boldsymbol{b}_k^{\mathrm{P}}$.

Since $\boldsymbol{\nu} = \boldsymbol{\nu}^+ - \boldsymbol{\nu}^- = \boldsymbol{D}\tilde{\boldsymbol{\nu}}$, where $\boldsymbol{D} = \begin{bmatrix} \boldsymbol{I}_{n\times n} & -\boldsymbol{I}_{n\times n} \end{bmatrix}$, and

$$\|\boldsymbol{\nu}\|_1^{\boldsymbol{l}} = \sum_{j=1}^{n} l_j |\nu_j| \le \sum_{j=1}^{n} l_j (\nu_j^+ + \nu_j^-) = \tilde{\boldsymbol{l}}^{\mathrm{T}} \tilde{\boldsymbol{\nu}} \quad \text{with} \quad \tilde{\boldsymbol{l}} = \begin{bmatrix} \boldsymbol{l} \\ \boldsymbol{l} \end{bmatrix}, \tag{7.69}$$

we conclude that any feasible point of the linear program

$$\begin{aligned} \sup_{\tilde{\boldsymbol{\nu}} \in \mathbb{R}^{2n}} \quad & (\boldsymbol{D}^{\mathrm{T}} \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{l}})^{\mathrm{T}} \tilde{\boldsymbol{\nu}} \\ \text{subject to} \quad & \boldsymbol{A}\tilde{\boldsymbol{\nu}} \le \boldsymbol{b}_k^{\mathrm{P}}, \\ & \tilde{\boldsymbol{\nu}} \ge \boldsymbol{0}, \end{aligned} \tag{7.70}$$

gives a conservative lower bound for λ_k subject to the positivity constraint (P). Similarly, any feasible point of the linear program

$$\begin{aligned} \inf_{\tilde{\boldsymbol{\nu}} \in \mathbb{R}^{2n}} \quad & -(\boldsymbol{D}^{\mathrm{T}} \tilde{\boldsymbol{y}} - \tilde{\boldsymbol{l}})^{\mathrm{T}} \tilde{\boldsymbol{\nu}} \\ \text{subject to} \quad & \boldsymbol{A}\tilde{\boldsymbol{\nu}} \le -\boldsymbol{b}_k^{\mathrm{P}}, \\ & \tilde{\boldsymbol{\nu}} \ge \boldsymbol{0}, \end{aligned} \tag{7.71}$$

yields a conservative upper bound. Notice that, in order to compute the positivity-constrained confidence intervals, we simply need to solve for each bin two $2n$-dimensional linear programs subject to $m+2n$ inequality constraints.

7.3.4.2 Decreasing intensities

The dual constraints (7.53) and (7.61) corresponding to the monotonicity constraint (D) can be discretized using an approach similar to Section 7.3.4.1. However, since now the right-hand side $\pm L_k^{\mathrm{D}}$ can vary within the intervals $[s_i, s_{i+1})$, a constant upper bound similar to Equation (7.66) would be too strong. The strategy we follow instead is to employ a first-order Taylor expansion of k_j^* in order to obtain a linear upper bound.

For any $s \in [s_i, s_{i+1})$, we have

$$k_j^*(s) = k_j^*(s_i) + (k_j^*)'(\xi_j)(s - s_i) = k_j^*(s_i) + k_j(\xi_j)(s - s_i), \quad \xi_j \in [s_i, s). \tag{7.72}$$

This gives the bound

$$\sum_{j=1}^{n} \nu_j k_j^*(s) = \sum_{j=1}^{n} \nu_j k_j^*(s_i) + \sum_{j=1}^{n} \nu_j^+ k_j(\xi_j)(s - s_i) - \sum_{j=1}^{n} \nu_j^- k_j(\xi_j)(s - s_i) \tag{7.73}$$

$$\leq \sum_{j=1}^{n} \nu_j k_j^*(s_i) + \sum_{j=1}^{n} \nu_j^+ \sup_{\xi \in [s_i, s_{i+1})} k_j(\xi)(s - s_i) - \sum_{j=1}^{n} \nu_j^- \inf_{\xi \in [s_i, s_{i+1})} k_j(\xi)(s - s_i) \tag{7.74}$$

$$= \sum_{j=1}^{n} \nu_j k_j^*(s_i) + \sum_{j=1}^{n} \nu_j^+ \overline{\rho}_{i,j}(s - s_i) - \sum_{j=1}^{n} \nu_j^- \underline{\rho}_{i,j}(s - s_i). \tag{7.75}$$

In other words, we have established a linear upper bound for $\sum_{j=1}^{n} \nu_j k_j^*(s)$ on $[s_i, s_{i+1})$.

Since L_k^{D} is also linear on each interval $[s_i, s_{i+1})$, we can simply enforce the constraint at the endpoints of the interval. By the continuity of L_k^{D}, we require, for each $i = 1, \ldots, m$, that

$$\begin{cases} \sum_{j=1}^{n} \nu_j k_j^*(s_i) \leq \pm L_k^{\mathrm{D}}(s_i), \\ \sum_{j=1}^{n} \nu_j k_j^*(s_i) + \sum_{j=1}^{n} \nu_j^+ \overline{\rho}_{i,j} \delta_i - \sum_{j=1}^{n} \nu_j^- \underline{\rho}_{i,j} \delta_i \leq \pm L_k^{\mathrm{D}}(s_{i+1}), \end{cases} \tag{7.76}$$

where $\delta_i = s_{i+1} - s_i$. In fact, since $\sum_{j=1}^{n} \nu_j k_j^*(s)$ is continuous, the first inequality in Equation (7.76) is redundant and it suffices to simply enforce the second one.

Let $\boldsymbol{\Delta} = \mathrm{diag}\left(\{\delta_i\}_{i=1}^m\right)$, let $\boldsymbol{K}^*$ denote the $m \times n$ matrix with elements $K_{i,j}^* = k_j^*(s_i)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, and let $\boldsymbol{b}_k^{\mathrm{D}} \in \mathbb{R}^m$ be the vector with elements $b_{k,i}^{\mathrm{D}} = L_k^{\mathrm{D}}(s_{i+1})$, $i = 1, \ldots, m$. Then $(\boldsymbol{K}^* \boldsymbol{D} + \boldsymbol{\Delta} \boldsymbol{A})\tilde{\boldsymbol{\nu}} \leq \pm \boldsymbol{b}_k^{\mathrm{D}}$ gives a conservative discretization of the dual constraints (7.53) and (7.61), where the matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ are as in Section 7.3.4.1.

We hence conclude that any feasible point of the linear program

$$\begin{aligned} \sup_{\tilde{\boldsymbol{v}}\in\mathbb{R}^{2n}} \quad & (\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{l}})^{\mathrm{T}}\tilde{\boldsymbol{v}} \\ \text{subject to} \quad & (\boldsymbol{K}^*\boldsymbol{D} + \boldsymbol{\Delta A})\tilde{\boldsymbol{v}} \leq \boldsymbol{b}_k^{\mathrm{D}}, \\ & \tilde{\boldsymbol{v}} \geq \boldsymbol{0}, \end{aligned} \tag{7.77}$$

yields a conservative lower bound for λ_k subject to the monotonicity constraint (D). Similarly, a conservative upper bound is given by any feasible point of

$$\begin{aligned} \inf_{\tilde{\boldsymbol{v}}\in\mathbb{R}^{2n}} \quad & -(\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{l}})^{\mathrm{T}}\tilde{\boldsymbol{v}} \\ \text{subject to} \quad & (\boldsymbol{K}^*\boldsymbol{D} + \boldsymbol{\Delta A})\tilde{\boldsymbol{v}} \leq -\boldsymbol{b}_k^{\mathrm{D}}, \\ & \tilde{\boldsymbol{v}} \geq \boldsymbol{0}. \end{aligned} \tag{7.78}$$

Notice that computing the confidence interval again reduces to simply solving two linear programs with $2n$ unknowns and $m + 2n$ inequality constraints.

7.3.4.3 Convex intensities

In the case of the convexity constraint (C), the right-hand side $\pm L_k^{\mathrm{C}}$ of the dual constraints (7.54) and (7.62) is a piecewise quadratic function. As such, the appropriate way of bounding the left-hand side is to use a quadratic upper bound, which can be obtained by employing a second-order Taylor expansion of k_j^{**}.

For any $s \in [s_i, s_{i+1})$, we have

$$k_j^{**}(s) = k_j^{**}(s_i) + (k_j^{**})'(s_i)(s - s_i) + \frac{1}{2}(k_j^{**})''(\xi_j)(s - s_i)^2 \tag{7.79}$$

$$= k_j^{**}(s_i) + k_j^{*}(s_i)(s - s_i) + \frac{1}{2}k_j(\xi_j)(s - s_i)^2, \quad \xi_j \in [s_i, s). \tag{7.80}$$

This yields the bound

$$\sum_{j=1}^{n} \nu_j k_j^{**}(s) = \sum_{j=1}^{n} \nu_j k_j^{**}(s_i) + \sum_{j=1}^{n} \nu_j k_j^{*}(s_i)(s - s_i) + \frac{1}{2}\sum_{j=1}^{n} \nu_j k_j(\xi_j)(s - s_i)^2 \tag{7.81}$$

$$\leq \sum_{j=1}^{n} \nu_j k_j^{**}(s_i) + \sum_{j=1}^{n} \nu_j k_j^{*}(s_i)(s - s_i) + \frac{1}{2}\sum_{j=1}^{n} \left(\nu_j^{+}\overline{\rho}_{i,j} - \nu_j^{-}\underline{\rho}_{i,j}\right)(s - s_i)^2, \tag{7.82}$$

where the inequality is obtained as in Equation (7.75). We have hence established a quadratic upper bound for $\sum_{j=1}^{n} \nu_j k_j^{**}(s)$ on $[s_i, s_{i+1})$ and we need to ensure that this parabola lies below $\pm L_k^{\mathrm{C}}$ for every $s \in [s_i, s_{i+1})$. In other words, we need to require that

$$\pm L_k^{\mathrm{C}}(s) - \sum_{j=1}^{n} \nu_j k_j^{**}(s_i) - \sum_{j=1}^{n} \nu_j k_j^{*}(s_i)(s - s_i) - \frac{1}{2}\sum_{j=1}^{n} \left(\nu_j^{+}\overline{\rho}_{i,j} - \nu_j^{-}\underline{\rho}_{i,j}\right)(s - s_i)^2 \geq 0, \tag{7.83}$$

for all $s \in [s_i, s_{i+1})$. Since L_k^{C} is quadratic on $[s_i, s_{i+1})$, the left-hand side in (7.83) is a parabola and we need to make sure that this parabola is positive on the interval $[s_i, s_{i+1})$.

Let $a_{i,k}s^2 + b_{i,k}s + c_{i,k}$ be the parabola corresponding to the left-hand side of (7.83) and let $s_{i,k}^* = -b_{i,k}/(2a_{i,k})$ be the s-coordinate of its vertex. Then $a_{i,k}s^2 + b_{i,k}s + c_{i,k} \geq 0$, $\forall s \in [s_i, s_{i+1})$, is equivalent to requiring that

$$\begin{cases} a_{i,k}s_i^2 + b_{i,k}s_i + c_{i,k} \geq 0, \\ a_{i,k}s_{i+1}^2 + b_{i,k}s_{i+1} + c_{i,k} \geq 0, \\ \mathbf{1}_{(s_i,s_{i+1})}\big(s_{i,k}^*\big)\big(a_{i,k}\big(s_{i,k}^*\big)^2 + b_{i,k}s_{i,k}^* + c_{i,k}\big) \geq 0. \end{cases} \tag{7.84}$$

Here the first two conditions guarantee that the endpoints of the parabola lie above the s-axis, while the last condition ensures that the vertex is above the s-axis when it is located on the interval (s_i, s_{i+1}). As before, by the continuity of $\sum_{j=1}^n \nu_j k_j^{**}(s)$ and $L_k^{\mathrm{C}}(s)$, the first condition is redundant and can be dropped.

Since $s_{i,k}^*$ depends nonlinearly on $\tilde{\boldsymbol{\nu}}$, the conservatively discretized dual program cannot be expressed as a linear program. Nevertheless, any feasible point of the program

$$\begin{aligned} \sup_{\tilde{\boldsymbol{\nu}} \in \mathbb{R}^{2n}} \quad & (\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{l}})^{\mathrm{T}}\tilde{\boldsymbol{\nu}} \\ \text{subject to} \quad & a_{i,k}s_{i+1}^2 + b_{i,k}s_{i+1} + c_{i,k} \geq 0,\ i = 1, \ldots, m, \\ & \mathbf{1}_{(s_i,s_{i+1})}(s_{i,k}^*)(a_{i,k}(s_{i,k}^*)^2 + b_{i,k}s_{i,k}^* + c_{i,k}) \geq 0,\ i = 1, \ldots, m, \\ & -\sum_{j=1}^n (\boldsymbol{D}\tilde{\boldsymbol{\nu}})_j\, k_j^*(E_{\max}) \geq E_{k,\min} - E_{k,\max}, \\ & \tilde{\boldsymbol{\nu}} \geq \mathbf{0}, \end{aligned} \tag{7.85}$$

yields a conservative lower bound for λ_k subject to the convexity constraint (C). Here $\boldsymbol{D}$ is as in Section 7.3.4.1, $s_{i,k}^* = -\frac{b_{i,k}}{2a_{i,k}}$ and the coefficients $a_{i,k}$, $b_{i,k}$ and $c_{i,k}$, which depend on $\tilde{\boldsymbol{\nu}}$, are given by

$$a_{i,k} = \begin{cases} -A_i,\ s_i < E_{k,\min}, \\ -A_i + \frac{1}{2},\ E_{k,\min} \leq s_i < E_{k,\max}, \\ -A_i,\ s_i \geq E_{k,\max}, \end{cases} \tag{7.86}$$

$$b_{i,k} = \begin{cases} 2A_i s_i - B_i,\ s_i < E_{k,\min}, \\ 2A_i s_i - B_i - E_{k,\min},\ E_{k,\min} \leq s_i < E_{k,\max}, \\ 2A_i s_i - B_i + E_{k,\max} - E_{k,\min},\ s_i \geq E_{k,\max}, \end{cases} \tag{7.87}$$

$$c_{i,k} = \begin{cases} -A_i s_i^2 + B_i s_i - C_i,\ s_i < E_{k,\min}, \\ -A_i s_i^2 + B_i s_i - C_i + \frac{1}{2}E_{k,\min}^2,\ E_{k,\min} \leq s_i < E_{k,\max}, \\ -A_i s_i^2 + B_i s_i - C_i - \frac{1}{2}E_{k,\max}^2 + \frac{1}{2}E_{k,\min}^2,\ s_i \geq E_{k,\max}, \end{cases} \tag{7.88}$$

where

$$A_i = \frac{1}{2}\sum_{j=1}^{n}\left(\nu_j^+\overline{\rho}_{i,j} - \nu_j^-\underline{\rho}_{i,j}\right), \tag{7.89}$$

$$B_i = \sum_{j=1}^{n}(\nu_j^+ - \nu_j^-)k_j^*(s_i), \tag{7.90}$$

$$C_i = \sum_{j=1}^{n}(\nu_j^+ - \nu_j^-)k_j^{**}(s_i). \tag{7.91}$$

Similarly, a conservative upper bound is given by any feasible point of

$$\begin{aligned}
\inf_{\tilde{\boldsymbol{\nu}}\in\mathbb{R}^{2n}} \quad & -(\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{l}})^{\mathrm{T}}\tilde{\boldsymbol{\nu}} \\
\text{subject to} \quad & a_{i,k}s_{i+1}^2 + b_{i,k}s_{i+1} + c_{i,k} \geq 0,\ i = 1,\ldots,m, \\
& \mathbf{1}_{(s_i,s_{i+1})}(s_{i,k}^*)(a_{i,k}(s_{i,k}^*)^2 + b_{i,k}s_{i,k}^* + c_{i,k}) \geq 0,\ i = 1,\ldots,m, \\
& -\sum_{j=1}^{n}(\boldsymbol{D}\tilde{\boldsymbol{\nu}})_j\, k_j^*(E_{\max}) \geq E_{k,\max} - E_{k,\min}, \\
& \tilde{\boldsymbol{\nu}} \geq \mathbf{0},
\end{aligned} \tag{7.92}$$

where the coefficients are

$$a_{i,k} = \begin{cases} -A_i,\ s_i < E_{k,\min}, \\ -A_i - \frac{1}{2},\ E_{k,\min} \leq s_i < E_{k,\max}, \\ -A_i,\ s_i \geq E_{k,\max}, \end{cases} \tag{7.93}$$

$$b_{i,k} = \begin{cases} 2A_is_i - B_i,\ s_i < E_{k,\min}, \\ 2A_is_i - B_i + E_{k,\min},\ E_{k,\min} \leq s_i < E_{k,\max}, \\ 2A_is_i - B_i - E_{k,\max} + E_{k,\min},\ s_i \geq E_{k,\max}, \end{cases} \tag{7.94}$$

$$c_{i,k} = \begin{cases} -A_is_i^2 + B_is_i - C_i,\ s_i < E_{k,\min}, \\ -A_is_i^2 + B_is_i - C_i - \frac{1}{2}E_{k,\min}^2,\ E_{k,\min} \leq s_i < E_{k,\max}, \\ -A_is_i^2 + B_is_i - C_i + \frac{1}{2}E_{k,\max}^2 - \frac{1}{2}E_{k,\min}^2,\ s_i \geq E_{k,\max}, \end{cases} \tag{7.95}$$

and A_i, B_i and C_i are given by Equations (7.89)–(7.91).

These programs have a linear objective function, but the constraints are nonlinear. The programs can nevertheless be solved using standard nonlinear programming methods, as long as some care is taken when choosing the algorithm and its starting point; see Appendix A.2 for details. Notice also that, since any feasible point of these programs gives a conservative bound, we do not necessarily need to find a global optimum. Instead, it suffices to find a good enough feasible point.

7.3.5 Summary

Sections 7.3.1–7.3.4 along with Theorem 7.1 can be summarized as follows:

Theorem 7.6. *Let* $\boldsymbol{y} \sim \mathrm{Poisson}(\mathcal{K} f)$ *with* $\mathcal{K} : V \to \mathbb{R}^n,\ f \mapsto \left[K_1 f, \ldots, K_n f\right]^{\mathrm{T}}$, *where* $V = C^2(E)$ *with* $E \subset \mathbb{R}$ *a compact interval. Let the forward functionals be given by*

$$K_j : V \to \mathbb{R},\ f \mapsto \int_E k_j(s) f(s)\,\mathrm{d}s, \tag{7.96}$$

where $k_j \in C(E)$, *and let the quantity of interest be* $\boldsymbol{\lambda} = \left[\int_{E_1} f(s)\,\mathrm{d}s, \ldots, \int_{E_p} f(s)\,\mathrm{d}s\right]^{\mathrm{T}}$, *where* $\{E_j\}_{j=1}^{p}$ *is a binning of* E *of the form* (3.17)*. For* $\alpha \in (0,1)$, *set* $\alpha' = 1 - (1-\alpha)^{1/n}$ *and, for each* $j = 1, \ldots, n$, *let*

$$\underline{\mu}_{j,\alpha'} = \frac{1}{2} F_{\chi^2}^{-1}\left(\frac{\alpha'}{2}; 2y_j\right) \quad \text{and} \quad \overline{\mu}_{j,\alpha'} = \frac{1}{2} F_{\chi^2}^{-1}\left(1 - \frac{\alpha'}{2}; 2(y_j+1)\right). \tag{7.97}$$

Let $\tilde{\boldsymbol{y}}$ *and* $\boldsymbol{l}$ *be vectors in* $\mathbb{R}^n$ *with components* $\tilde{y}_j = \left(\underline{\mu}_{j,\alpha'} + \overline{\mu}_{j,\alpha'}\right)/2$ *and* $l_j = \left(\overline{\mu}_{j,\alpha'} - \underline{\mu}_{j,\alpha'}\right)/2$, *and denote* $\boldsymbol{D} = \left[\boldsymbol{I}_{n\times n} \ \ -\boldsymbol{I}_{n\times n}\right]$ *and* $\tilde{\boldsymbol{l}} = \left[\boldsymbol{l}^{\mathrm{T}}\ \boldsymbol{l}^{\mathrm{T}}\right]^{\mathrm{T}}$*. Let* $\underline{\lambda}(\tilde{\boldsymbol{v}}_k) = (\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{l}})^{\mathrm{T}}\tilde{\boldsymbol{v}}_k$ *and* $\overline{\lambda}(\tilde{\boldsymbol{v}}_k) = -(\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{l}})^{\mathrm{T}}\tilde{\boldsymbol{v}}_k$ *and, for each* $k = 1, \ldots, p$, *let* $\tilde{\boldsymbol{v}}_{k,\mathrm{P}}^{\mathrm{lb}}$, $\tilde{\boldsymbol{v}}_{k,\mathrm{P}}^{\mathrm{ub}}$, $\tilde{\boldsymbol{v}}_{k,\mathrm{D}}^{\mathrm{lb}}$, $\tilde{\boldsymbol{v}}_{k,\mathrm{D}}^{\mathrm{ub}}$, $\tilde{\boldsymbol{v}}_{k,\mathrm{C}}^{\mathrm{lb}}$ *and* $\tilde{\boldsymbol{v}}_{k,\mathrm{C}}^{\mathrm{ub}}$ *be feasible points of the programs* (7.70), (7.71), (7.77), (7.78), (7.85) *and* (7.92)*, respectively. Then*

(P) $\mathrm{P}_f\left(\boldsymbol{\lambda} \in \left[\underline{\lambda}(\tilde{\boldsymbol{v}}_{1,\mathrm{P}}^{\mathrm{lb}}), \overline{\lambda}(\tilde{\boldsymbol{v}}_{1,\mathrm{P}}^{\mathrm{ub}})\right] \times \cdots \times \left[\underline{\lambda}(\tilde{\boldsymbol{v}}_{p,\mathrm{P}}^{\mathrm{lb}}), \overline{\lambda}(\tilde{\boldsymbol{v}}_{p,\mathrm{P}}^{\mathrm{ub}})\right]\right) \geq 1 - \alpha,\ \forall f \in C$, *where* $C = \left\{f \in V : f(s) \geq 0, \forall s \in E\right\}$;

(D) $\mathrm{P}_f\left(\boldsymbol{\lambda} \in \left[\underline{\lambda}(\tilde{\boldsymbol{v}}_{1,\mathrm{D}}^{\mathrm{lb}}), \overline{\lambda}(\tilde{\boldsymbol{v}}_{1,\mathrm{D}}^{\mathrm{ub}})\right] \times \cdots \times \left[\underline{\lambda}(\tilde{\boldsymbol{v}}_{p,\mathrm{D}}^{\mathrm{lb}}), \overline{\lambda}(\tilde{\boldsymbol{v}}_{p,\mathrm{D}}^{\mathrm{ub}})\right]\right) \geq 1 - \alpha,\ \forall f \in C$, *where* $C = \left\{f \in V : f(s) \geq 0 \wedge f'(s) \leq 0, \forall s \in E\right\}$;

(C) $\mathrm{P}_f\left(\boldsymbol{\lambda} \in \left[\underline{\lambda}(\tilde{\boldsymbol{v}}_{1,\mathrm{C}}^{\mathrm{lb}}), \overline{\lambda}(\tilde{\boldsymbol{v}}_{1,\mathrm{C}}^{\mathrm{ub}})\right] \times \cdots \times \left[\underline{\lambda}(\tilde{\boldsymbol{v}}_{p,\mathrm{C}}^{\mathrm{lb}}), \overline{\lambda}(\tilde{\boldsymbol{v}}_{p,\mathrm{C}}^{\mathrm{ub}})\right]\right) \geq 1 - \alpha,\ \forall f \in C$, *where* $C = \left\{f \in V : f(s) \geq 0 \wedge f'(s) \leq 0 \wedge f''(s) \geq 0, \forall s \in E\right\}$.

Notice that the theorem states that the resulting confidence intervals have *guaranteed finite-sample simultaneous coverage*, provided that f satisfies the stated shape constraints. The theorem is written for any feasible point of the conservatively discretized dual programs since, especially for nonlinear programming, there is no guarantee that the output of a numerical algorithm is a global optimum. But obviously in practice we wish to find as good a feasible point as possible, leading to intervals with the least amount of slack. We next proceed to demonstrate that, even though these intervals are conservative, they can still yield usefully tight confidence bounds in a realistic unfolding scenario.

7.4 Demonstration: Inclusive jet transverse momentum spectrum

We demonstrate the shape-constrained strict bounds by unfolding the inclusive jet transverse momentum spectrum described in Section 3.4.2 and also studied in Section 4.2. Here the true intensity f obviously satisfies the positivity constraint (P) for all values of the transverse momentum p_T. It is also decreasing for $p_T \gtrsim 2.0$ and convex for $p_T \gtrsim 2.8$. In other words, the monotonicity constraint (D) and the convexity constraint (C) are satisfied for intermediate and large p_T values, which are the main focus of inclusive jet analyses at the LHC (CMS Collaboration, 2013b). Even without reference to a particular parameterization of the jet p_T spectrum, physical considerations lead us to expect that the spectrum is decreasing for intermediate and large p_T values and convex at least for intermediate p_T values.

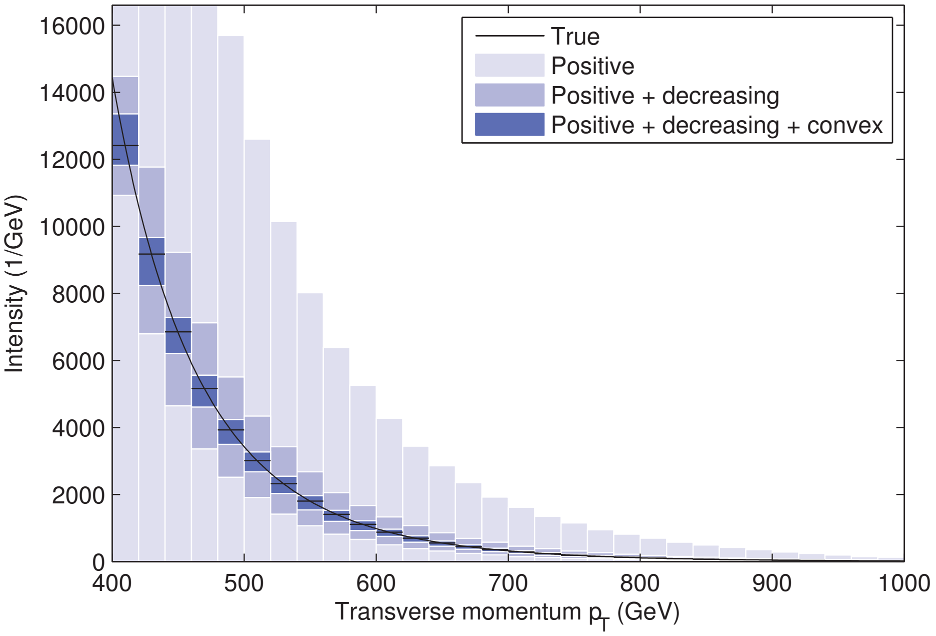
When computing the strict bounds, we discretize the dual constraints using $m+1 = 10p+1$ uniformly spaced grid points, which corresponds to subdividing each true bin E_k into 10 sub-bins. The intervals are computed at 95 % simultaneous confidence level. The remaining parameters are as described in Section 3.4.2. All the experiments were implemented in MATLAB R2014a; see Appendix A.2 for further details on the implementation.

Figure 7.4 shows the 95 % shape-constrained strict bounds for the inclusive jet p_T spectrum. The true value of $\boldsymbol{\lambda}$ is shown by the horizontal lines. To enable comparison with the true intensity f, the binned quantities are converted to the intensity scale by dividing them by the bin width. The results are shown both on the linear scale and the log scale. We see that the confidence intervals cover $\boldsymbol{\lambda}$ in every bin. Notice in particular that the inferences are well-calibrated also in the tail of the spectrum, even though the intensity varies over three orders of magnitude. Furthermore, the shape constraints have a major impact on the length of the intervals. With only the positivity constraint, the intervals are fairly wide (but presumably still orders of magnitude shorter than unregularized intervals) and the lower bound is zero at every bin. But, with the stronger monotonicity and convexity constraints, the intervals become significantly tighter, leading to sharper inferences.

Figure 7.5 shows the dual constraints $\pm L_{10}^{\mathrm{P}}(s)$, $\pm L_{10}^{\mathrm{D}}(s)$ and $\pm L_{10}^{\mathrm{C}}(s)$ (see Equations (7.57)–(7.59)) and the corresponding optimal solutions $\sum_{j=1}^{n} \nu_j k_j(s)$, $\sum_{j=1}^{n} \nu_j k_j^*(s)$ and $\sum_{j=1}^{n} \nu_j k_j^{**}(s)$ at the 10th true bin. We see that, despite the conservative discretization, the optimal solutions can be very close to the constraints $\pm L_{10}^{\mathrm{P}}(s)$, $\pm L_{10}^{\mathrm{D}}(s)$ and $\pm L_{10}^{\mathrm{C}}(s)$. For the positivity-constrained lower bound, the optimal solution is to have $\boldsymbol{\nu} = \mathbf{0}$, which is consistent with the lower bound $\underline{\lambda}_{10} = 0$. We also compared the conservatively discretized intervals with the potentially unconservative intervals obtained using the naïve discretization where the dual constraint is only imposed on the grid points $\{s_i\}_{i=1}^{m+1}$ as in Equation (7.63). We found that the conservatively discretized intervals were not much wider then the unconservative ones. For the monotonicity and convexity constraints, the length difference was less than 1 % in most bins. For the bin where the difference was the largest, the conservatively discretized intervals were 13.2 %, 2.4 % and 2.0 % longer for the positivity, monotonicity and convexity constraints, respectively. We hence conclude that the conservative discretization enables us to guarantee the confidence level without excessively increasing the interval length.

By construction, the simultaneous coverage probability of the confidence intervals in Figure 7.4 is at least 95 %. To verify that this is indeed the case, we computed the intervals for 1 000 independent observations. We found that, for each replication, the intervals covered $\boldsymbol{\lambda}$ for all

(a) Inclusive jet p_T spectrum, linear scale



(b) Inclusive jet p_T spectrum, log scale

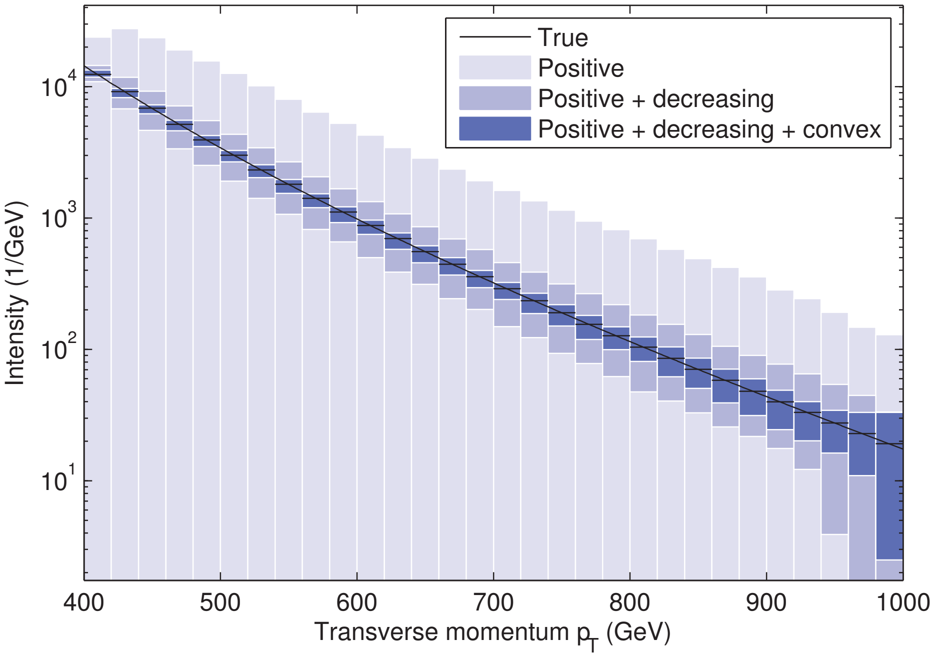
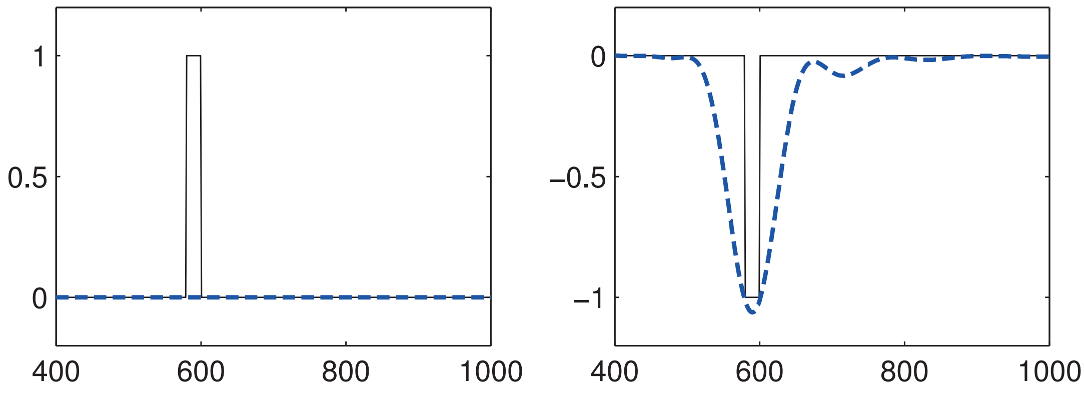
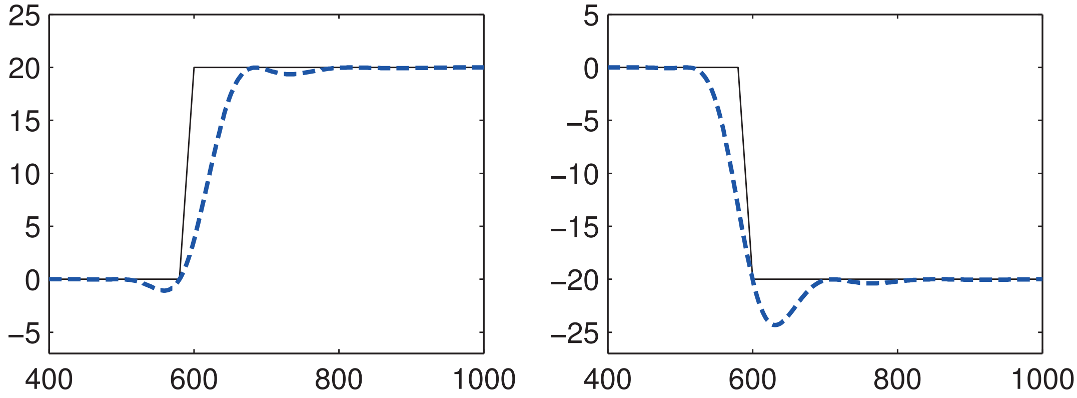


Figure 7.4: Shape-constrained uncertainty quantification in unfolding the inclusive jet transverse momentum spectrum. Figure (a) shows the 95 % shape-constrained strict bounds on a linear scale and Figure (b) the same intervals on a log scale. These intervals have by construction guaranteed finite-sample simultaneous coverage.



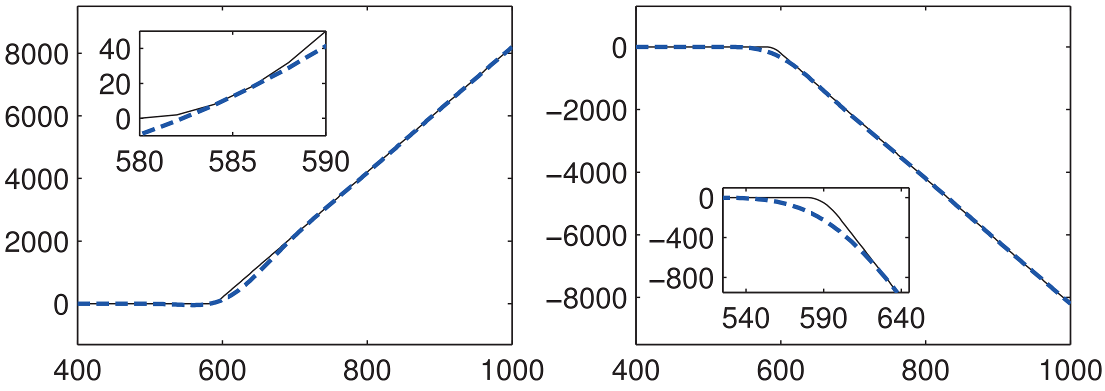
(a) Positive, lower bound

(b) Positive, upper bound



(c) Decreasing, lower bound

(d) Decreasing, upper bound



(e) Convex, lower bound

(f) Convex, upper bound

Figure 7.5: The dual constraints $\pm L_{10}^{\mathrm{P}}$, $\pm L_{10}^{\mathrm{D}}$ and $\pm L_{10}^{\mathrm{C}}$ (solid lines) and the corresponding optimal solutions (dashed lines) for the 10th true bin and for the different shape constraints. The insets in Figures (e) and (f) show the quadratic part of the constraint in greater detail.

the shape constraints. In other words, the empirical coverage of the intervals is 100 % in this particular example and the 95 % Clopper–Pearson interval for the actual coverage probability is $[0.996, 1.000]$. This confirms that the 95 % confidence level is indeed attained, but also shows that the intervals are conservative in the sense that the actual coverage probability is much greater than 95 %. This is in stark constrast with the methods that are presently used at the LHC, which suffer from severe undercoverage in this problem; see Section 4.2.

8 Concluding remarks

We close with a discussion in Section 8.1, which is followed by a summary of our conclusions in Section 8.2.

8.1 Discussion and outlook

In this section, we discuss various aspects of this work, with a particular emphasis on areas where future work is needed and on directions to which the present work can be extended:

Estimation and uncertainty of the forward operator:

> Throughout this work, we have assumed that the forward operator K is known, while in reality it is usually estimated using either simulations or auxiliary measurements and is hence uncertain. This raises two interesting methodological questions: Firstly, how should one estimate and quantify the uncertainty of K? And secondly, how should one incorporate these estimates into the unfolding procedure? When a trustworthy parametric model is available for K, which was the case when we considered unfolding the inclusive jet p_T spectrum (Section 3.4.2) or the Z boson invariant mass spectrum (Section 6.6), then standard parametric techniques can be used to estimate and quantify the uncertainty of K, and it should be feasible to incorporate the resulting uncertainties into the bootstrap procedures of Chapter 6 and into the strict bounds construction of Chapter 7. A significantly more challenging situation arises when nonparametric estimates of K are needed. In this case, estimation of K essentially becomes a nonparametric quantile regression problem and one is faced with the task of quantifying the uncertainty of the quantiles and incorporating this uncertainty into either the bootstrap or the strict bounds confidence intervals.

Further analysis of the iterative bias-correction and the coverage-length trade-off:

> The empirical performance of the iteratively bias-corrected confidence intervals raises several interesting theoretical and methodological questions. For example, in all but one of our test cases (the minor exception being the two-peak function of Section 6.4.5 with

noise level $\sigma = 0.001$), we have observed that the bias-corrected intervals are shorter than the undersmoothed intervals, but it would be desirable to be able to establish analytically what conditions need to be satisfied for this to happen and what factors influence the size of the difference. This obviously also raises the question of whether there are other generic methods that yield even shorter intervals for the same amount coverage. In other words, is there a way to be above the bias-correction curve in a coverage-length plot, such as the one in Figure 6.8, and would it be possible to derive a theoretical upper bound for such a curve?

Iterative bias-correction in other applications:

The iteratively bias-corrected bootstrap confidence intervals introduced in Section 6.2 are fully generic in the sense that the basic construction is applicable to any point estimator or noise model. The only requirement is the ability to sample from the underlying statistical model. As such, it would be interesting to study how the approach performs in other situations where uncertainty quantification is hampered by the bias. In particular, it would be interesting to study whether the approach can be applied to ℓ^1-regularized lasso regression and how it compares to existing debiasing techniques, such as Javanmard and Montanari (2014), in that situation. Other potential applications include, for example, scatterplot smoothing (Ruppert et al., 2003), generalized additive models (Wood, 2006b) and Gaussian processes (Rasmussen and Williams, 2006). Notice also that the iterative bias-correction is more widely applicable than undersmoothing, since the latter requires a distinct regularization parameter that controls the size of the bias, while the iterative bias-correction is also applicable to situations where no such parameter can easily be identified.

Adaptive roughness penalties:

In some sense, the confidence intervals derived using a simple roughness penalty fail to attain appropriate coverage because a penalty term of the form $\|f''\|_2^2$ cannot accommodate both large and small curvature at the same time. In this work, we solved the problem by debiasing the point estimates, but a potentially viable alternative would be to employ a spatially adaptive penalty term (e.g., Ruppert and Carroll, 2000; Pintore et al., 2006). Such penalties would enable the amount of smoothing to adapt locally to the shape of the intensity function. The drawback, however, is that typically the adaptation is controlled by a high-dimensional regularization parameter whose data-driven choice is even more challenging than the choice of the single regularization parameter associated with the roughness penalty.

Overcoverage of the strict bounds intervals:

The shape-constrained strict bounds of Chapter 7 are conservative in the sense that their coverage probability may be much larger than the nominal value. The main reason for this is the way the infinite-dimensional confidence set $C \cap D$ (see Section 7.2) is turned into a finite number of confidence intervals. In essence, $C \cap D$ is first mapped through $\mathscr{H} : V \to \mathbb{R}^p, f \mapsto [H_1 f, \ldots, H_p f]^{\mathrm{T}}$ and then the resulting set $\mathscr{H}(C \cap D)$ is bounded by the smallest possible ℓ^∞-ball. Here the main source of slack is presumably that the geometry of ℓ^∞-balls might not be well-suited for bounding $\mathscr{H}(C \cap D)$. There are two possibilities for reducing this slack. The first possibility (Stark, 1992, Section 10.2) is to tune the geometry of Ξ so that the geometry of $\mathscr{H}(C \cap D)$ is better suited for bounding with an ℓ^∞-ball. The second possibility, which to the author's best knowledge has not been proposed before, is to bound $\mathscr{H}(C \cap D)$ with some other set, instead of an ℓ^∞-ball, that can be represented and communicated using a finite collection of numbers. For example, if the geometry of $\mathscr{H}(C \cap D)$ resembles a hyperellipsoid, then an efficient way of bounding it would be to use a weighted ℓ^2-ball. Such approach would resemble the way uncertainties are quantified and communicated in classical well-posed Gaussian regression problems.

Other types of shape constraints:

In Chapter 7, we considered shape constraints in the form of positivity, monotonicity and convexity. While the positivity constraint is satisfied for any Poisson intensity function, not all particle spectra satisfy the monotonicity and convexity constraints. These would, for example, not be the appropriate shape constraints for the Z boson invariant mass peak of Section 6.6 or for the intensity function of Section 3.4.1, which consists of two peaks on a uniform background. A natural way of regularizing these situations would be to use a unimodality constraint in the first case and a bimodality constraint in the second case. More generally, it would be useful to generalize the methodology of Chapter 7 to k-modal intensities (Hengartner and Stark, 1995). Another useful generalization would be to consider intensities that have a concave part and a convex part with an unknown changepoint between the two. Such shapes would be able to handle, for example, the full inclusive jet p_T spectrum of Equation (3.36) without having to focus only on the steeply falling tail.

Regularization using wide bins:

As explained in Section 3.2.3, one cannot simply regularize the conventional unfolding techniques by increasing the size of the true bins as this would increase the dependence of the response matrix on the shape of the Monte Carlo prediction f^MC inside the bins. However, with the strict bounds intervals of Chapter 7, there is no such restriction. In particular, one can use arbitrarily wide bins with the positivity constraint to obtain regularized unfolded confidence intervals that have guaranteed coverage *without making any assumptions about* f (except that it is regular enough so that all the integrals are well-defined). This approach would be similar in spirit to that of Burrus (1965). The drawback, of course, is that, by using wide bins, the resolution with which we probe the function f decreases and we might end up missing important features of the true solution.

Use of unfolded confidence intervals:

By the duality between hypothesis tests and confidence intervals, the uncertainties that we have derived in Chapters 6 and 7 immediately yield a hypothesis test in the unfolded space. More specifically, the pointwise confidence bands of Chapter 6 can be used to test a theory prediction at any single point $s \in E$ and, with a multiple testing correction, also at a finite number of points. The simultaneous confidence intervals of Chapter 7, on the other hand, directly yield a test for the whole theory prediction without further multiplicity corrections. For example, the 95 % strict bounds of Figure 7.4 can be used to perform a test of a theory prediction at 5 % significance level by simply overlaying the prediction on the confidence envelope and verifying whether it is inside the envelope at each bin.

As noted in Section 3.3.2, unfolded confidence intervals can also serve as the basis for more complex inferential tasks. For example, one could envisage using the strict bounds construction of Section 7.2 with an identity smearing operator to combine two or more multiplicity-corrected unfolded spectra. With an appropriate choice of the forward operator, the construction can also be used to extract further physical parameters from one or more unfolded measurements. A particularly pertinent topic for future research would be to develop methodology for extracting rigorous parton distribution function uncertainties (NNPDF Collaboration, 2015) from several unfolded spectra.

8.2 Conclusions

We have shown that unfolded confidence intervals can suffer from serious undercoverage in realistic unfolding scenarios. This applies to both standard frequentist and Bayesian constructions as well as to the methods that are currently used in LHC data analysis.

This does not happen due to some fault in the regularized point estimators. Instead, optimality in uncertainty quantification is fundamentally different from optimality in point estimation and methods that aim to achieve optimal point estimation do not necessarily yield good uncertainty quantification performance.

There exist at least two ways of obtaining improved unfolded uncertainty quantification. The first is to debias the regularized point estimators, which can be achieved by either using iterative bias-correction or through undersmoothing. We have compared the two approaches and found that, in several situations, bias-correction yields shorter confidence intervals than undersmoothing. In both methods, the amount of debiasing can be chosen in a data-driven way in order to approximately reach a given target coverage. For reasonable sample sizes and noise levels, the debiased intervals can yield nearly nominal coverage with only a modest increase in interval length.

The second possibility is to impose quantitative shape constraints. If such constraints are applied directly to an unregularized unfolded confidence set, one can derive usefully tight unfolded confidence intervals with rigorous finite-sample coverage guarantees. We have provided a conservative way of doing this, but possibilities exist for reducing the slack of the intervals.

We conclude that if physically justified shape information is available, then one should use the shape-constrained intervals with their strong coverage guarantees. When no such information is available, debiasing methods are still applicable and provide a way of obtaining much more accurate uncertainty quantification than conventional techniques.

A Technical details

A.1 Single-component Metropolis-Hastings sampler for unfolding

In this section, we provide a description of the single-component Metropolis–Hastings sampler that we use for sampling from the posterior $p(\boldsymbol{\beta}|\boldsymbol{y},\delta)$ in Chapters 5 and 6. This MCMC sampler was originally proposed by Saquib et al. (1998, Section III.C) in the context of tomographic image reconstruction.

For the posterior given by Equation (5.3), the logarithm of the kth full posterior conditional is given by

$$\log p(\beta_k|\boldsymbol{\beta}_{-k},\boldsymbol{y},\delta) = \sum_{i=1}^{n} y_i \log\left(\sum_{j=1}^{p} K_{i,j}\beta_j\right) - \sum_{i=1}^{n}\sum_{j=1}^{p} K_{i,j}\beta_j \tag{A.1}$$

$$- \delta \sum_{i=1}^{p}\sum_{j=1}^{p} \Omega_{A,i,j}\beta_i\beta_j + \mathrm{const} := f(\beta_k,\boldsymbol{\beta}_{-k}), \tag{A.2}$$

where $\boldsymbol{\beta}_{-k} = \left[\beta_1,\ldots,\beta_{k-1},\beta_{k+1},\ldots,\beta_p\right]^{\mathrm{T}}$ and the constant does not depend on β_k. In the standard Gibbs sampler, one would sample a new point β_k^* from this full conditional given the current position of the Markov chain $\boldsymbol{\beta}$. Unfortunately, sampling from this univariate density is difficult because of the log-term. We hence take a second-order Taylor expansion of the log-term in (A.2) around β_k, the current value of the kth component, to find

$$f(\beta_k^*,\boldsymbol{\beta}_{-k}) \approx d_{1,k}(\beta_k^* - \beta_k) + \frac{d_{2,k}}{2}(\beta_k^* - \beta_k)^2 \tag{A.3}$$

$$- \delta\left(\Omega_{\mathrm{A},k,k}(\beta_k^*)^2 + 2\sum_{i\neq k}\Omega_{\mathrm{A},i,k}\beta_i\beta_k^*\right) + \mathrm{const} := g(\beta_k^*,\boldsymbol{\beta}), \tag{A.4}$$

where

$$d_{1,k} = -\sum_{i=1}^{n} K_{i,k}\left(1 - \frac{y_i}{\mu_i}\right), \quad d_{2,k} = -\sum_{i=1}^{n} y_i\left(\frac{K_{i,k}}{\mu_i}\right)^2, \tag{A.5}$$

with $\boldsymbol{\mu} = \boldsymbol{K\beta}$. The approximate full conditional corresponding to $g(\beta_k^*, \boldsymbol{\beta})$ is a univariate Gaussian $\mathcal{N}(m_k, \sigma_k^2)$ with mean

$$m_k = \frac{d_{1,k} - d_{2,k}\beta_k - 2\delta\sum_{i\neq k}\Omega_{\mathrm{A},i,k}\beta_i}{2\delta\Omega_{\mathrm{A},k,k} - d_{2,k}} \tag{A.6}$$

and variance

$$\sigma_k^2 = \frac{1}{2\delta\Omega_{\mathrm{A},k,k} - d_{2,k}}. \tag{A.7}$$

If $m_k \geq 0$, we sample the new point β_k^* from $\mathcal{N}(m_k, \sigma_k^2)$ truncated to the non-negative real line. On the other hand, if $m_k < 0$, we sample β_k^* from an exponential distribution $\mathrm{Exp}(\lambda_k)$ satisfying the condition

$$\frac{\partial}{\partial\beta_k^*}\log p(\beta_k^*|\boldsymbol{\beta})\Big|_{\beta_k^*=0} = \frac{\partial}{\partial\beta_k^*}g(\beta_k^*, \boldsymbol{\beta})\Big|_{\beta_k^*=0}. \tag{A.8}$$

This gives

$$\lambda_k = -d_{1,k} + d_{2,k}\beta_k + 2\delta\sum_{i\neq k}\Omega_{\mathrm{A},i,k}\beta_i. \tag{A.9}$$

We use the exponential distribution since rejection sampling from the positive Gaussian tail with $m_k \ll 0$ would be computationally demanding. To summarize, the new point β_k^* is sampled from the approximate kth full posterior conditional given by

$$p(\beta_k^*|\boldsymbol{\beta}) = \begin{cases} \mathcal{N}_+(m_k, \sigma_k^2), & \text{if } m_k \geq 0, \\ \mathrm{Exp}(\lambda_\mathrm{k}), & \text{if } m_k < 0, \end{cases} \tag{A.10}$$

where $\beta_k^* \geq 0$ and $\mathcal{N}_+(m_k, \sigma_k^2)$ denotes the truncation of $\mathcal{N}(m_k, \sigma_k^2)$ to $[0, \infty)$.

To correct for the use of the approximate full conditional (A.10) instead of the actual full conditional (A.2), we need to perform a Metropolis–Hastings acceptance step for the proposed move to β_k^*. Let us denote $p(\beta_k^*|\boldsymbol{\beta}) = q(\beta_k^*, \beta_k, \boldsymbol{\beta}_{-k})$ and $p(\boldsymbol{\beta}|\boldsymbol{y}, \delta) = h(\beta_k, \boldsymbol{\beta}_{-k})$. Then the probability of accepting β_k^* is given by (Gilks et al., 1996)

$$a(\beta_k^*, \boldsymbol{\beta}) = \min\left\{1, \frac{h(\beta_k^*, \boldsymbol{\beta}_{-k})q(\beta_k, \beta_k^*, \boldsymbol{\beta}_{-k})}{h(\beta_k, \boldsymbol{\beta}_{-k})q(\beta_k^*, \beta_k, \boldsymbol{\beta}_{-k})}\right\}. \tag{A.11}$$

If β_k^* is rejected, the Markov chain remains at its current position $\boldsymbol{\beta}$.

The resulting single-component Metropolis–Hastings sampler is summarized in Algorithm 5. Note that the only free parameters in this MCMC sampler are the sample size S and the starting point $\boldsymbol{\beta}^{(1)}$. In particular, there is no tuning parameter to adjust the step size of the sampler. Instead, the step sizes are automatically adapted to the scale of the full conditionals in the same way as in the traditional Gibbs sampler.

Algorithm 5 Single-component Metropolis–Hastings sampler

```
Input:
  $\boldsymbol{\beta}^{(1)}$ — Starting point
  $S$ — Size of the MCMC sample
Output:
  $\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(S)}$ — MCMC sample from $p(\boldsymbol{\beta}|\boldsymbol{y}, \delta)$

  Set $\boldsymbol{\beta} = \boldsymbol{\beta}^{(1)}$;
  for $s = 2$ to $S$ do
      for $k = 1$ to $p$ do
          Sample $\beta_k^*$ from $p(\beta_k^*|\boldsymbol{\beta})$ given by Equation (A.10);
          Compute the acceptance probability $a(\beta_k^*, \boldsymbol{\beta})$ given by Equation (A.11);
          Sample $U \sim \mathrm{Unif}(0, 1)$;
          if $U \leq a(\beta_k^*, \boldsymbol{\beta})$ then
              Set $\beta_k = \beta_k^*$;
          end if
      end for
      Set $\boldsymbol{\beta}^{(s)} = \boldsymbol{\beta}$;
  end for
  return $\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(S)}$;
```

A.2 Implementation details for shape-constrained strict bounds

The optimization problems yielding the shape-constrained strict bounds involve a relatively high-dimensional solution space, numerical values at very different scales and fairly complicated constraints. As a result, some care is needed in their numerical solution, including verifying the validity of the optimization algorithms' output.

For the positivity and monotonicity constraints, where the bounds can be found by linear programming, we use the interior-point linear program solver as implemented in the `linprog` function of the MATLAB Optimization Toolbox (Mathworks, 2014b). To find the convexity-constrained bounds, we use the sequential quadratic programming (SQP) algorithm as implemented in the `fmincon` function of the same toolbox.

The optimization problems described in Section 7.3.4 tend to suffer from numerical instabilities when the solver explores large values of $\tilde{\boldsymbol{v}}$. We address this issue by introducing an upper bound for $\tilde{\boldsymbol{v}}$. That is, for each j, we replace the constraint $\tilde{v}_j \geq 0$ with the constraint $0 \leq \tilde{v}_j \leq M$, where M is chosen to be large enough so that the upper bound is not active at the optimal solution. (Notice that even if the upper bound was active, the solution of the modified problem would still be a valid conservative confidence bound since the restricted feasible set is a subset of the original feasible set.) We found that imposing the upper bound significantly improves the stability of the numerical solvers. In the experiments of this thesis, M is set to 30 for the positivity constraint, 15 for the monotonicity constraint and 10 for the convexity constraint.

The solutions found by the optimization algorithms can violate the constraints within a preset numerical tolerance. This could make the confidence bounds optimistic rather than conservative. To ensure that this does not happen, we verify the feasibility of the solutions returned by the optimization algorithms. In case a solution is infeasible, we iteratively scale $\boldsymbol{\nu}^+$ down and $\boldsymbol{\nu}^-$ up until it becomes feasible. Typically only a limited amount of fine-tuning of this kind was required to obtain a feasible point.

The SQP algorithm needs to be initialized with a good feasible point. To find one, we first solve the linear program corresponding to the unconservative discretization (see the beginning of Section 7.3.4)

$$\sum_{j=1}^{n} \nu_j k_j^{**}(s_i) \leq \pm L_k^{\mathrm{C}}(s_i), \quad i = 1, \ldots, m+1. \tag{A.12}$$

We then scale the solution as described above to make it feasible for the conservative discretization and the result is used as the starting point for SQP.

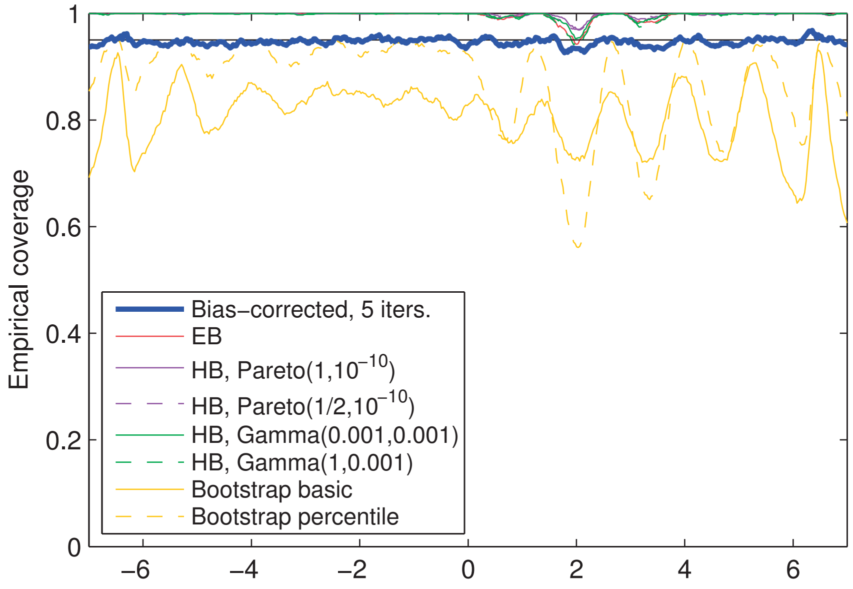
The implementation described here generally works robustly for unfolding the inclusive jet spectrum of Section 3.4.2, but occasionally the algorithms return a suboptimal feasible point. This maintains conservative coverage, but adjusting the tuning parameters of the algorithms might help finding a better feasible point. For the lower bound, a feasible point can always be found using $\tilde{\boldsymbol{\nu}} = \mathbf{0}$ (yielding the trivial bound of zero), while, for the upper bound, it may happen that the algorithms do not find a feasible point, in which case the bound should be set to $+\infty$.

B Full simulation results

B.1 Full simulation results for Section 6.3

This section provides the full results of the simulation study of Section 6.3. More specifically, the following pages include coverage studies (the analogue of Figure 6.4) and observed intervals (the analogues of Figures 6.5 and 6.7) for each of the three sample sizes $\lambda_{\text{tot}} = 1\,000$, 10 000 and 50 000. The results are given in the order of decreasing sample size.

(a) Comparison of coverage performance, $\lambda_{tot} = 50000$



(b) Effect of bias−correction on coverage performance, $\lambda_{tot} = 50000$

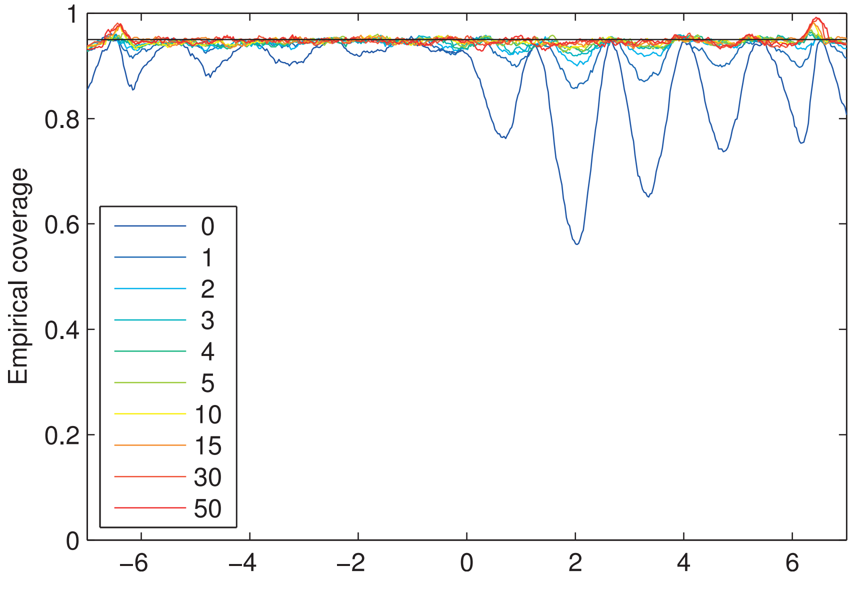


Figure B.1: Coverage studies in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\text{tot}} = 50\,000$. Figure (a) compares the empirical coverage of the iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{G_+}$ to the alternative empirical Bayes (EB), hierarchical Bayes (HB) and bootstrap intervals. The number of bias-correction iterations was set to 5 and the regularization strength chosen using the MMLE, expect for HB where four different uninformative hyperpriors were considered. Figure (b) shows the coverage of the bias-corrected intervals when the number of bias-correction iterations is varied between 0 and 50. All the intervals are for 95 % nominal pointwise coverage shown by the horizontal line.

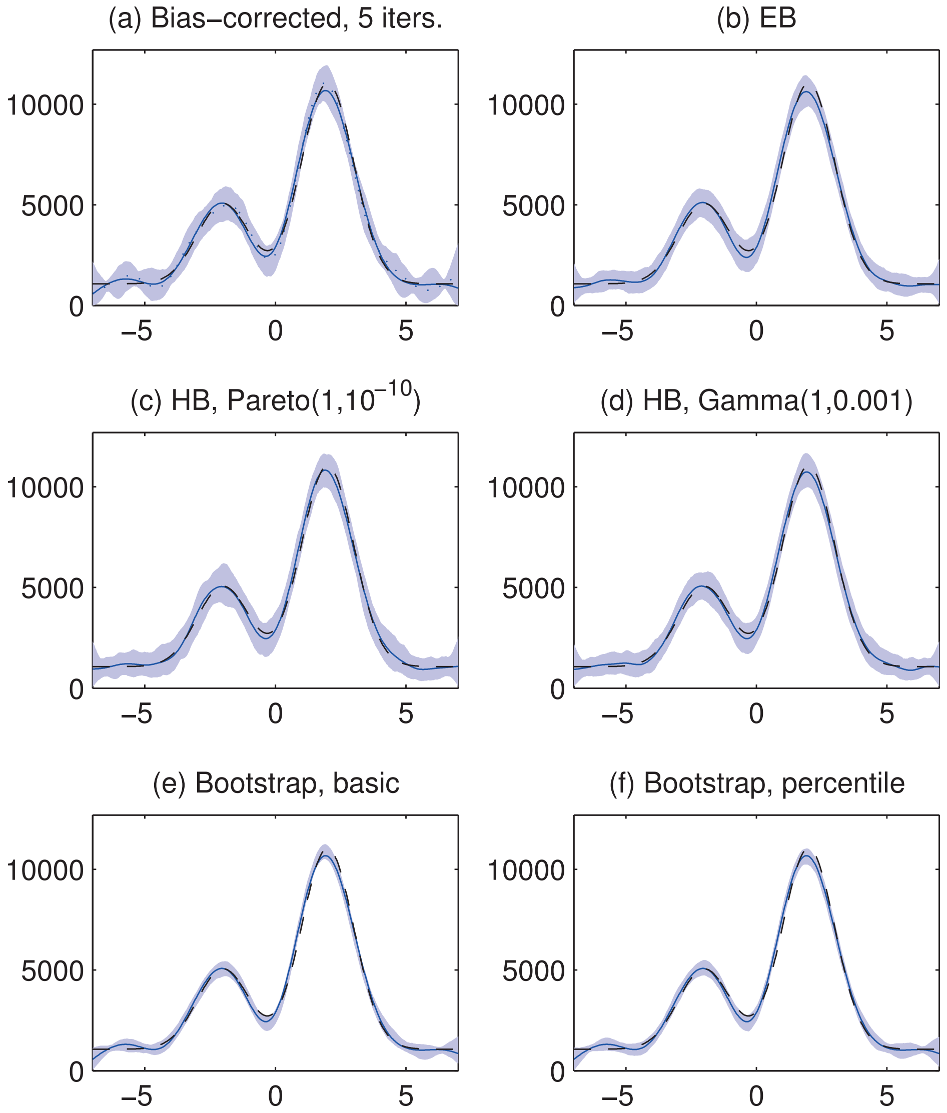


Figure B.2: A single realization of (a) the iteratively bias-corrected percentile intervals with 5 bias-correction iterations, (b) the empirical Bayes (EB) credible intervals, (c)–(d) the hierarchical Bayes (HB) credible intervals for the two extremal hyperpriors, (e) the basic bootstrap intervals and (f) the standard bootstrap percentile intervals in unfolding the two peaks on a uniform background test setup with $\lambda_{\text{tot}} = 50\,000$. Intervals (a), (e) and (f) are induced by the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$. Also shown are the corresponding point estimates $\hat{f}$ (solid lines) and the true intensity f (dashed lines). In Figure (a), also the bias-corrected point estimate $\hat{f}_{\mathrm{BC}}$ (dotted line) is given. All the intervals are formed for 95 % nominal pointwise coverage.

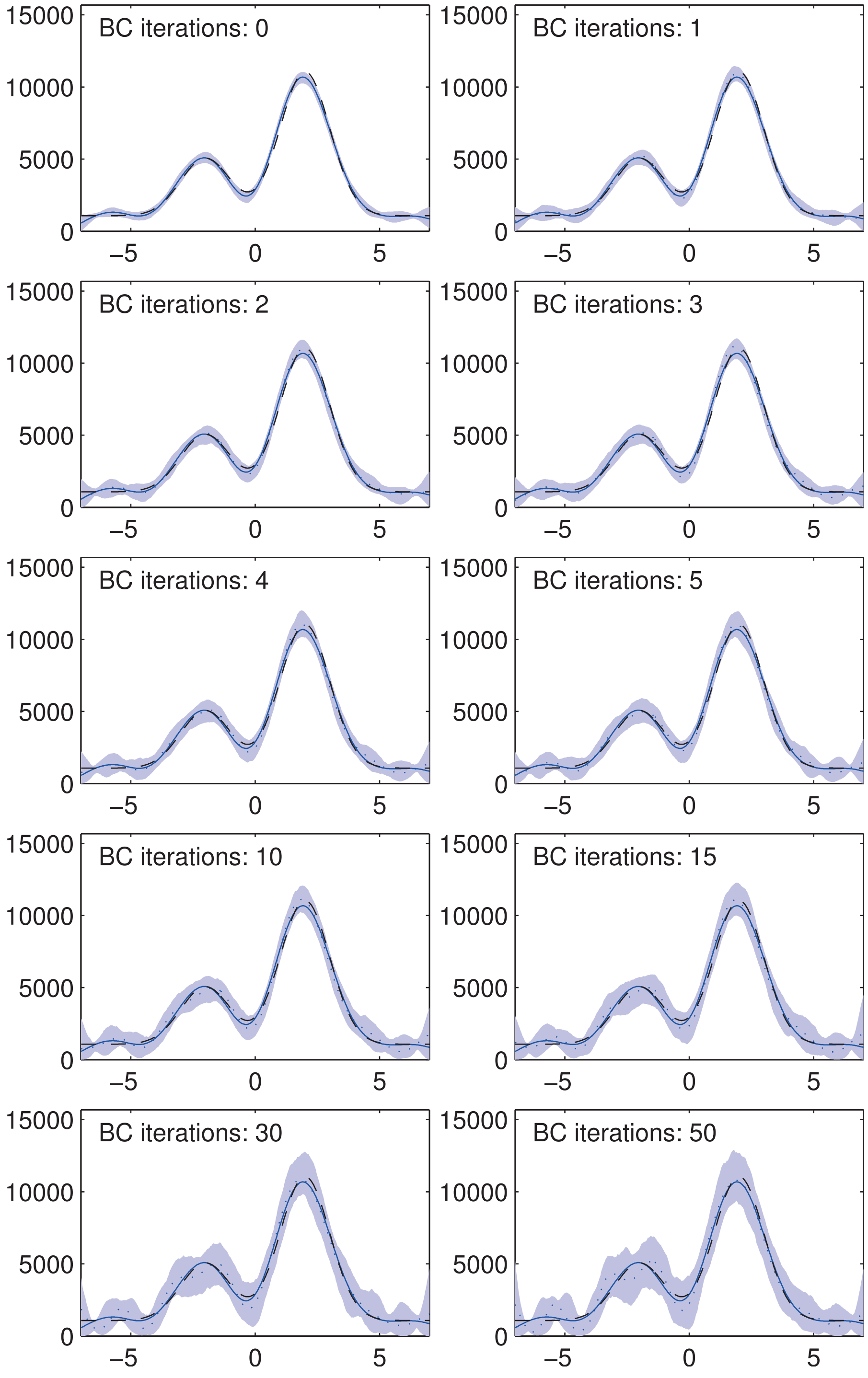
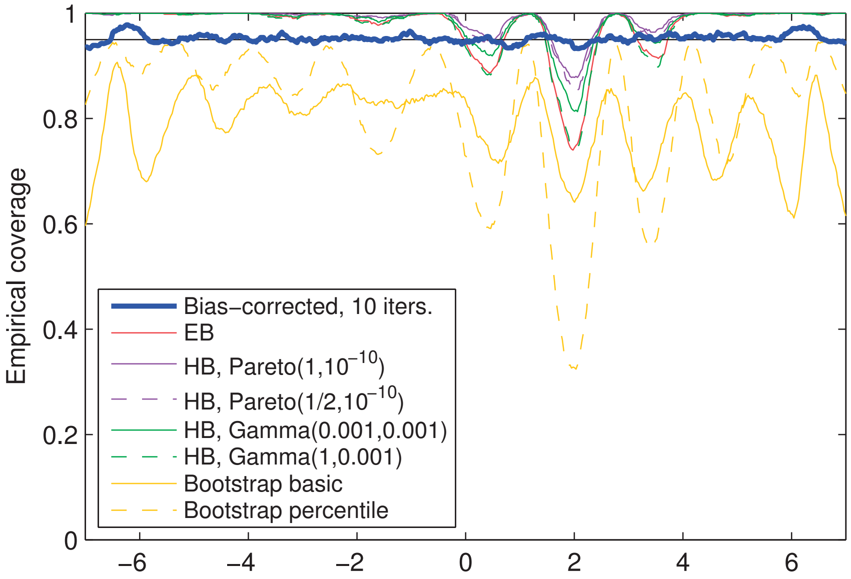


Figure B.3: Iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ in the two peaks on a uniform background test setup with $\lambda_{\mathrm{tot}} = 50\,000$ as the number of bias-correction iterations is varied. The true intensity is shown by the dashed line, the point estimate corresponding to $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ by the solid line and the bias-corrected point estimate by the dotted line. The intervals have 95 % nominal pointwise coverage.

(a) Comparison of coverage performance, λ_{tot} = 10000



(b) Effect of bias−correction on coverage performance, λ_{tot} = 10000

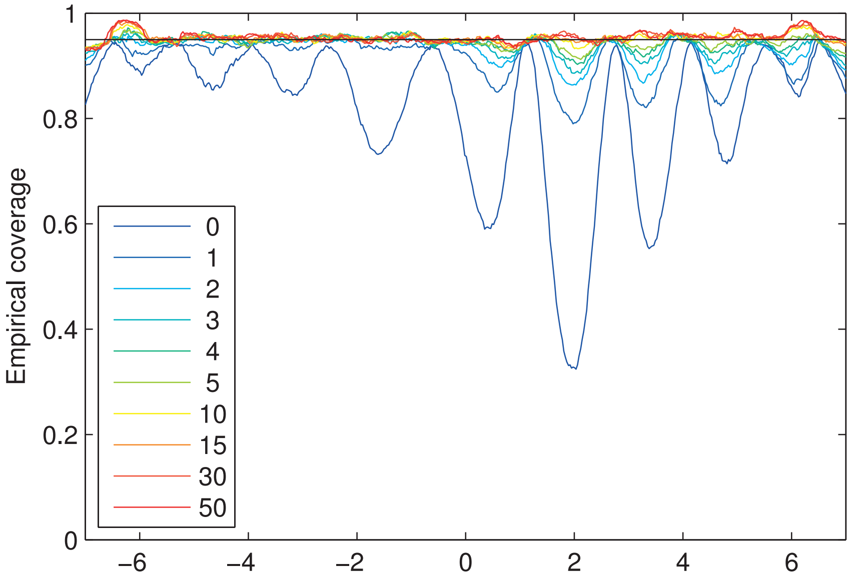


Figure B.4: Coverage studies in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\text{tot}} = 10\,000$. Figure (a) compares the empirical coverage of the iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ to the alternative empirical Bayes (EB), hierarchical Bayes (HB) and bootstrap intervals. The number of bias-correction iterations was set to 10 and the regularization strength chosen using the MMLE, expect for HB where four different uninformative hyperpriors were considered. Figure (b) shows the coverage of the bias-corrected intervals when the number of bias-correction iterations is varied between 0 and 50. All the intervals are for 95 % nominal pointwise coverage shown by the horizontal line. This figure is the same as Figure 6.4 in the main text.

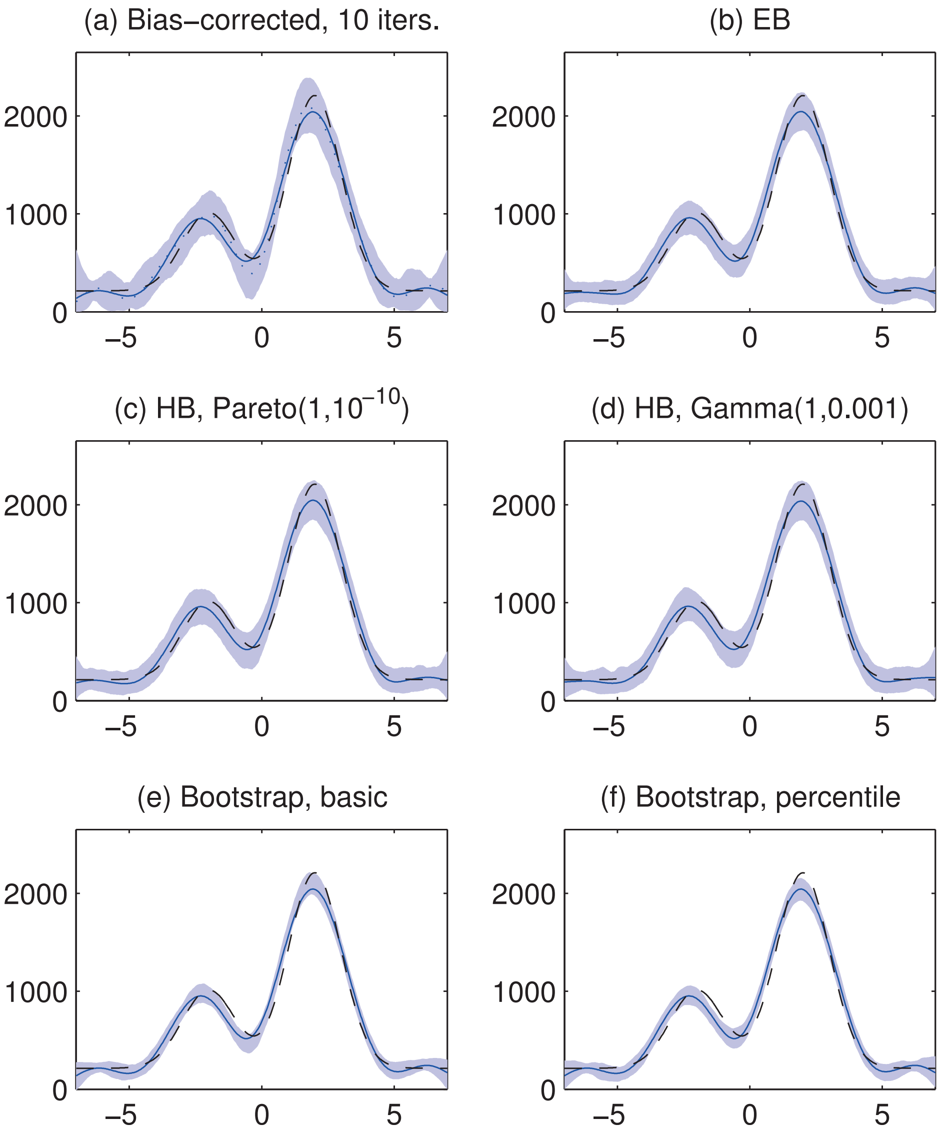


Figure B.5: A single realization of (a) the iteratively bias-corrected percentile intervals with 10 bias-correction iterations, (b) the empirical Bayes (EB) credible intervals, (c)–(d) the hierarchical Bayes (HB) credible intervals for the two extremal hyperpriors, (e) the basic bootstrap intervals and (f) the standard bootstrap percentile intervals in unfolding the two peaks on a uniform background test setup with $\lambda_{\text{tot}} = 10\,000$. Intervals (a), (e) and (f) are induced by the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\text{G}_+}$. Also shown are the corresponding point estimates $\hat{f}$ (solid lines) and the true intensity f (dashed lines). In Figure (a), also the bias-corrected point estimate $\hat{f}_{\text{BC}}$ (dotted line) is given. All the intervals are formed for 95 % nominal pointwise coverage.

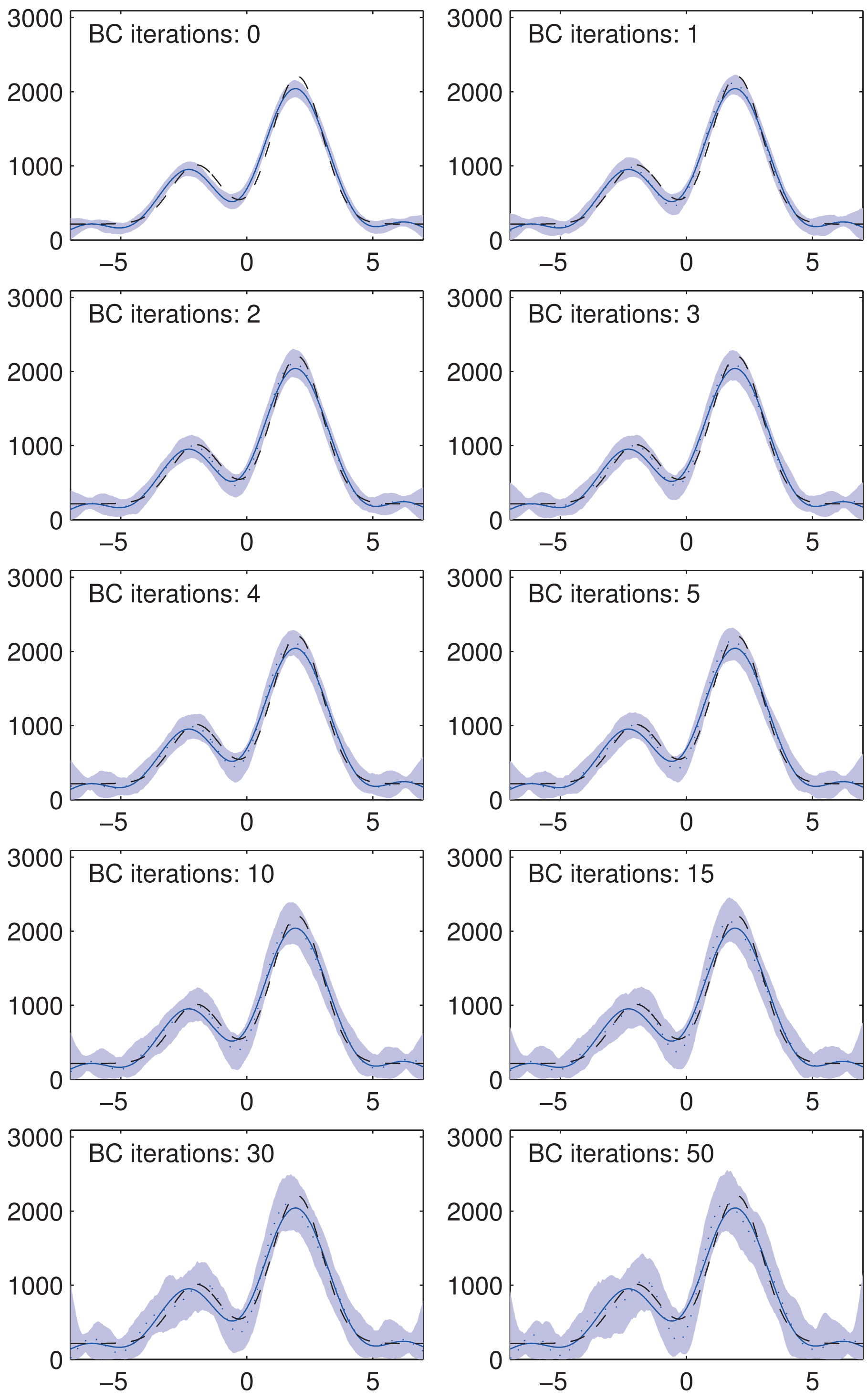
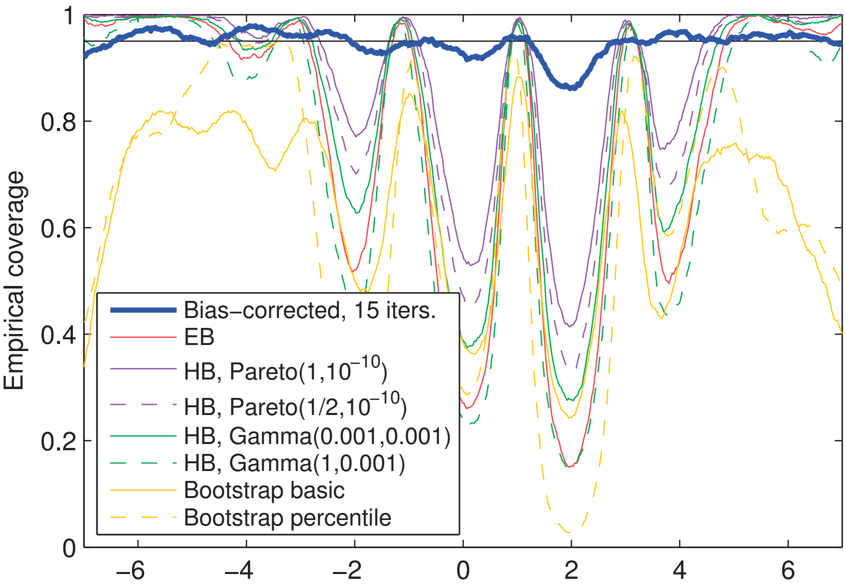


Figure B.6: Iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ in the two peaks on a uniform background test setup with $\lambda_{\mathrm{tot}} = 10\,000$ as the number of bias-correction iterations is varied. The true intensity is shown by the dashed line, the point estimate corresponding to $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ by the solid line and the bias-corrected point estimate by the dotted line. The intervals have 95 % nominal pointwise coverage. This figure is the same as Figure 6.5 in the main text.

(a) Comparison of coverage performance, λ_{tot} = 1000



(b) Effect of bias−correction on coverage performance, λ_{tot} = 1000

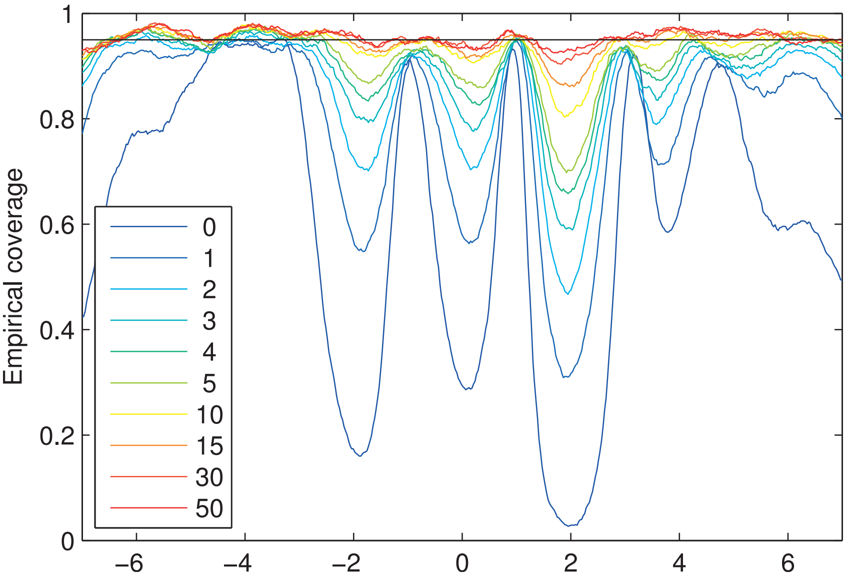


Figure B.7: Coverage studies in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\mathrm{tot}} = 1\,000$. Figure (a) compares the empirical coverage of the iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$ to the alternative empirical Bayes (EB), hierarchical Bayes (HB) and bootstrap intervals. The number of bias-correction iterations was set to 15 and the regularization strength chosen using the MMLE, expect for HB where four different uninformative hyperpriors were considered. Figure (b) shows the coverage of the bias-corrected intervals when the number of bias-correction iterations is varied between 0 and 50. All the intervals are for 95 % nominal pointwise coverage shown by the horizontal line. Figure (a) is the same as Figure 6.6 in the main text.

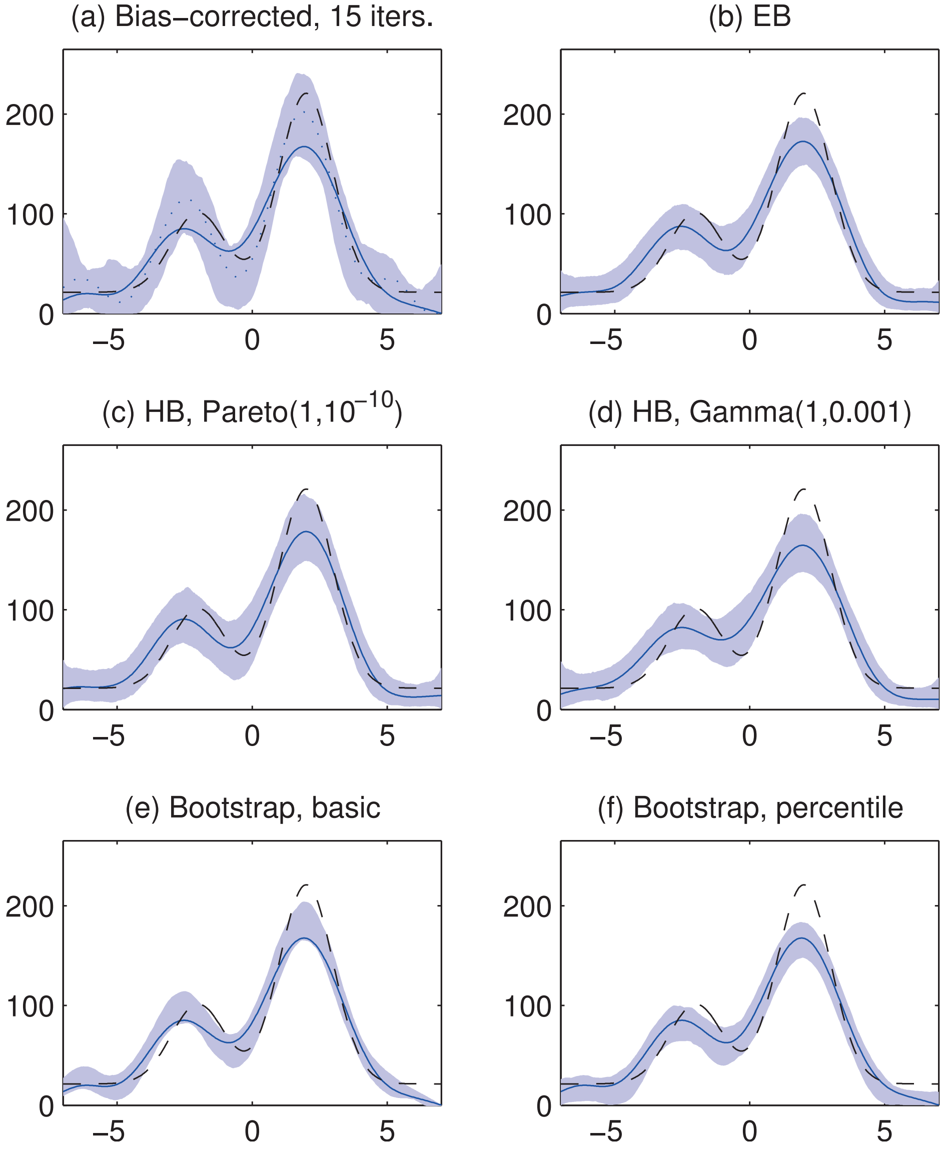


Figure B.8: A single realization of (a) the iteratively bias-corrected percentile intervals with 15 bias-correction iterations, (b) the empirical Bayes (EB) credible intervals, (c)–(d) the hierarchical Bayes (HB) credible intervals for the two extremal hyperpriors, (e) the basic bootstrap intervals and (f) the standard bootstrap percentile intervals in unfolding the two peaks on a uniform background test setup with $\lambda_{\mathrm{tot}} = 1\,000$. Intervals (a), (e) and (f) are induced by the positivity-constrained Gaussian approximation $\hat{\boldsymbol{\beta}}_{\mathrm{G}_+}$. Also shown are the corresponding point estimates $\hat{f}$ (solid lines) and the true intensity f (dashed lines). In Figure (a), also the bias-corrected point estimate $\hat{f}_{\mathrm{BC}}$ (dotted line) is given. All the intervals are formed for 95 % nominal pointwise coverage. This figure is the same as Figure 6.7 in the main text.

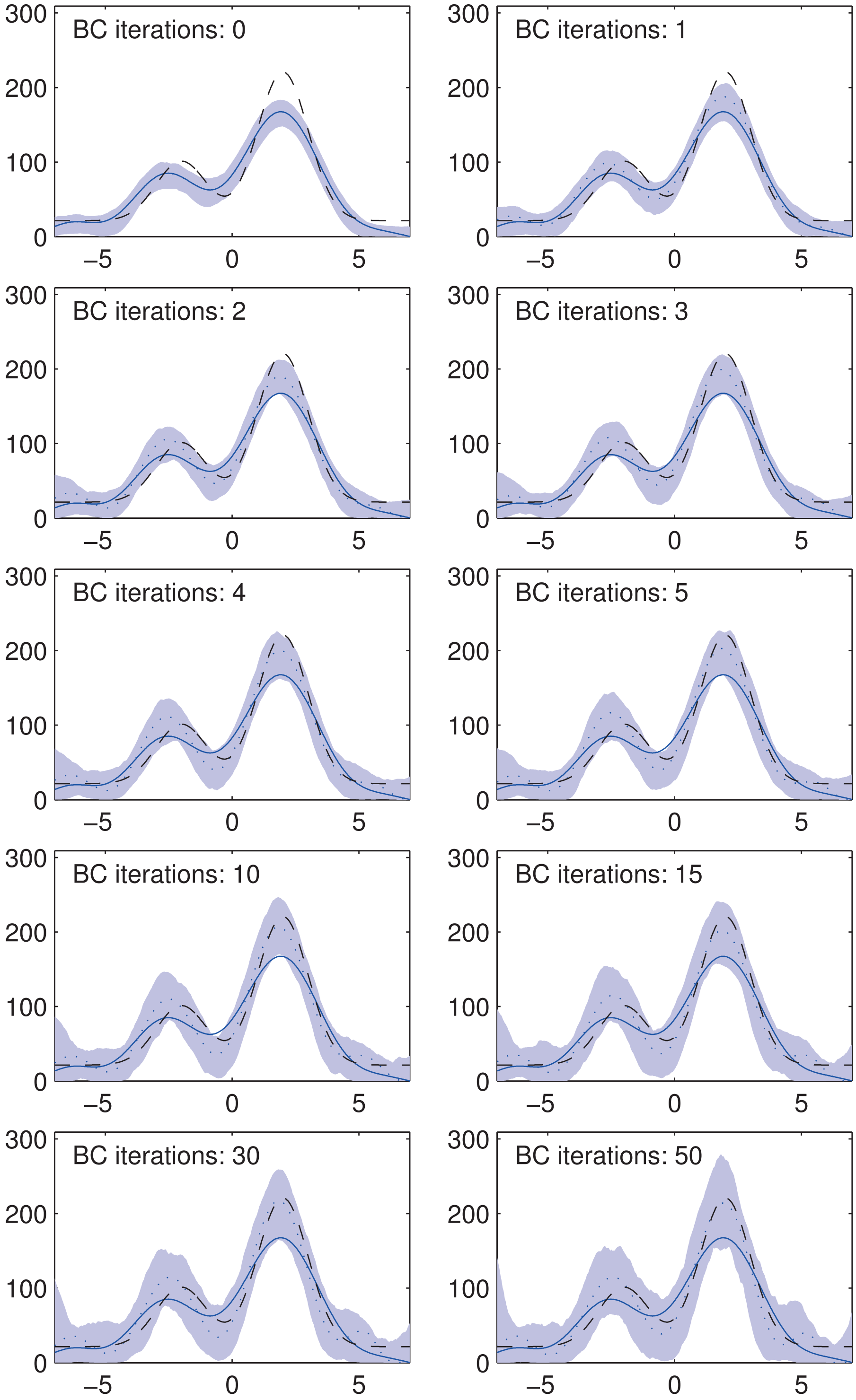
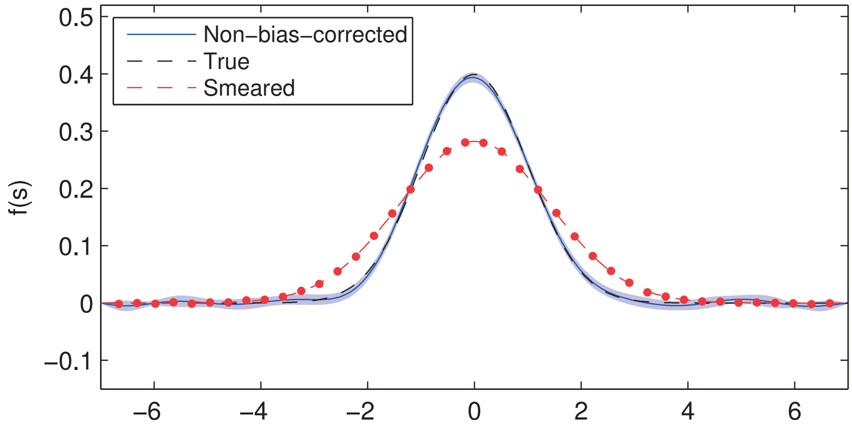
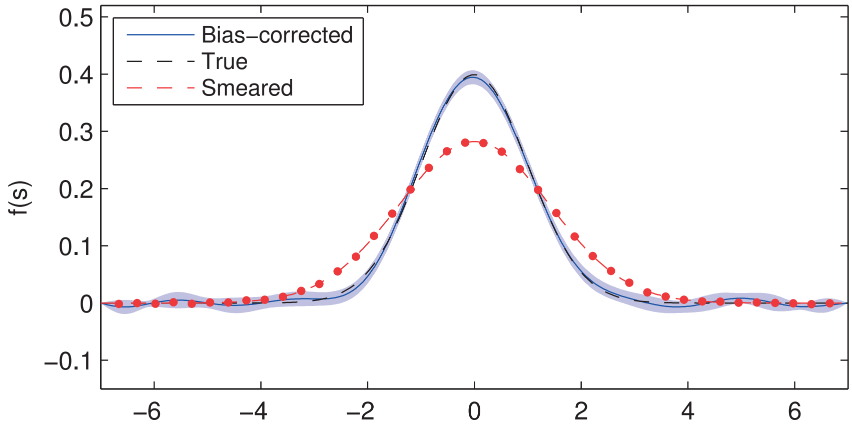


Figure B.9: Iteratively bias-corrected percentile intervals induced by $\hat{\boldsymbol{\beta}}_{G_+}$ in the two peaks on a uniform background test setup with $\lambda_{\text{tot}} = 1\,000$ as the number of bias-correction iterations is varied. The true intensity is shown by the dashed line, the point estimate corresponding to $\hat{\boldsymbol{\beta}}_{G_+}$ by the solid line and the bias-corrected point estimate by the dotted line. The intervals have 95 % nominal pointwise coverage.

B.2 Full simulation results for Section 6.4.5

This section gives the full results for the simulation study of Section 6.4.5. In particular, the following pages include the equivalents of Figures 6.10 and 6.11 for the one-peak and two-peak test cases and for the various sample sizes. The results are given first for the one-peak function f_1 of Equation (6.69) and then for the two-peak function f_2 of Equation (6.70). For each test case, the results are given in the order of increasing noise level.


(a) Non–bias–corrected intervals, σ = 0.001


(b) Iteratively bias–corrected intervals, σ = 0.001

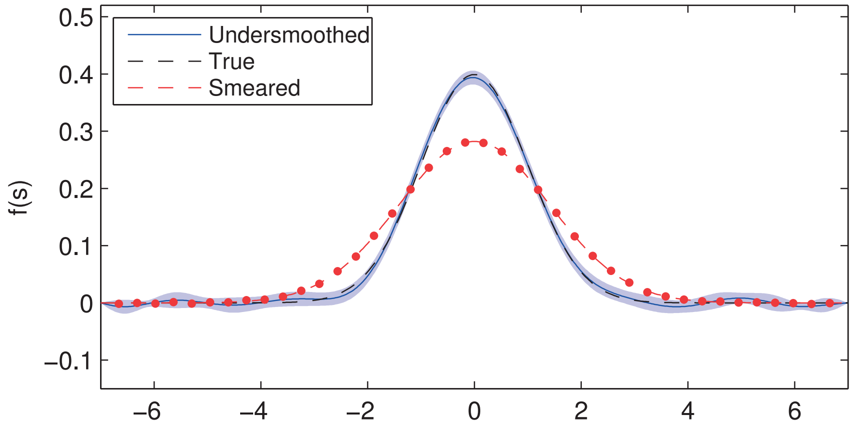
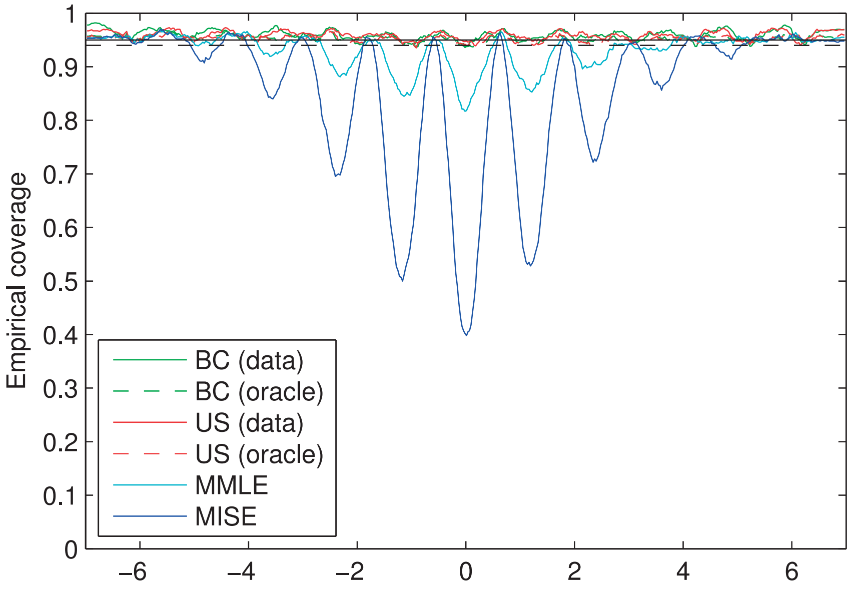

(c) Undersmoothed intervals, σ = 0.001

Figure B.10: Data-driven confidence intervals for the one-peak function f_1 and noise level $\sigma = 0.001$. The regularization strength is chosen using the MMLE and the amount of bias-correction and undersmoothing using the procedure described in Sections 6.4.3 and 6.4.4. The intervals have 95 % nominal coverage and 94 % target coverage.

(a) Empirical coverage (one peak, σ = 0.001)



(b) Mean interval length (one peak, σ = 0.001)

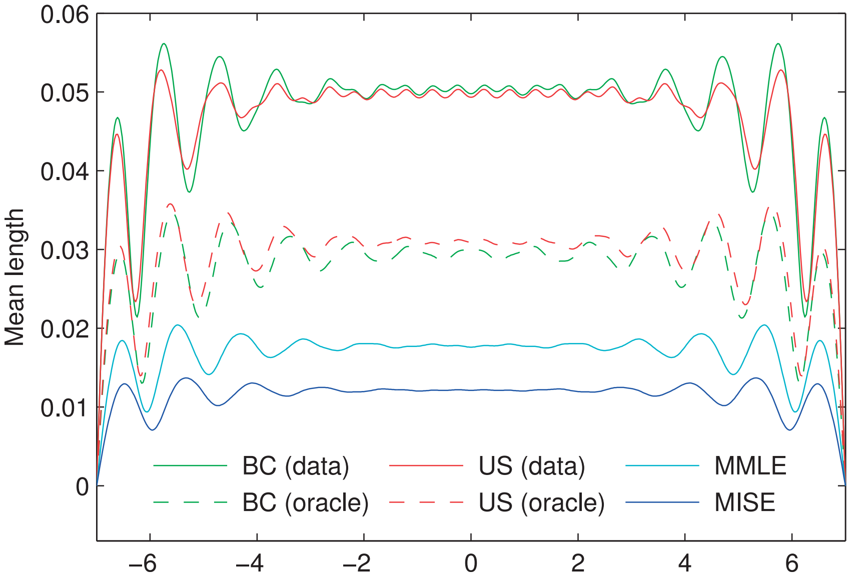
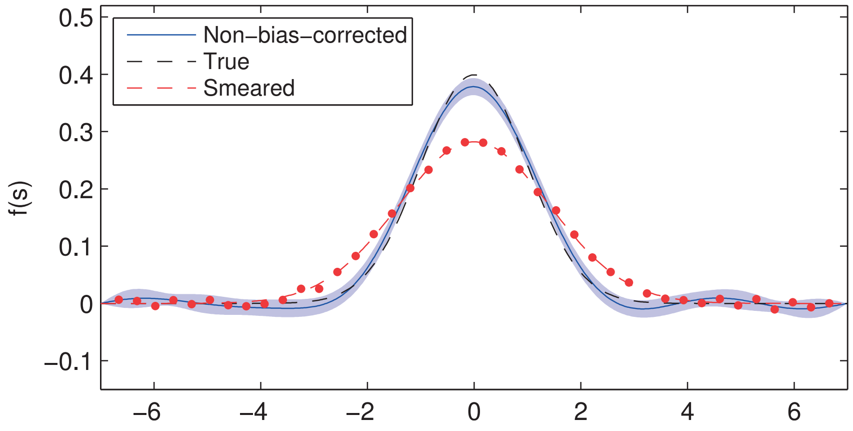
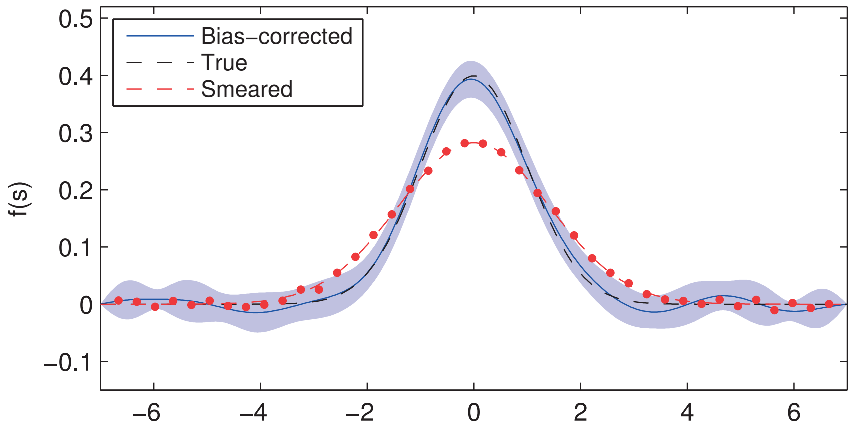


Figure B.11: Empirical coverage and mean length of the data-driven bias-corrected (BC), undersmoothed (US) and non-bias-corrected intervals for the one-peak function f_1 and noise level $\sigma = 0.001$. The curves labeled "data" are fully data-driven and the curves labeled "oracle" use knowledge of f_1 to choose the amount of debiasing. The non-bias-corrected results are given for both the MMLE choice of the regularization strength as well as for the choice that minimizes the MISE. The intervals have 95 % nominal coverage and 94 % target coverage.

(a) Non–bias–corrected intervals, σ = 0.005



(b) Iteratively bias–corrected intervals, σ = 0.005



(c) Undersmoothed intervals, σ = 0.005

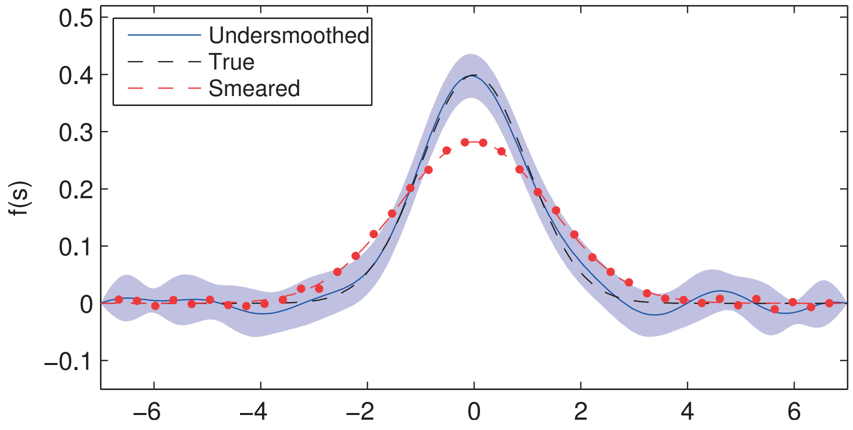
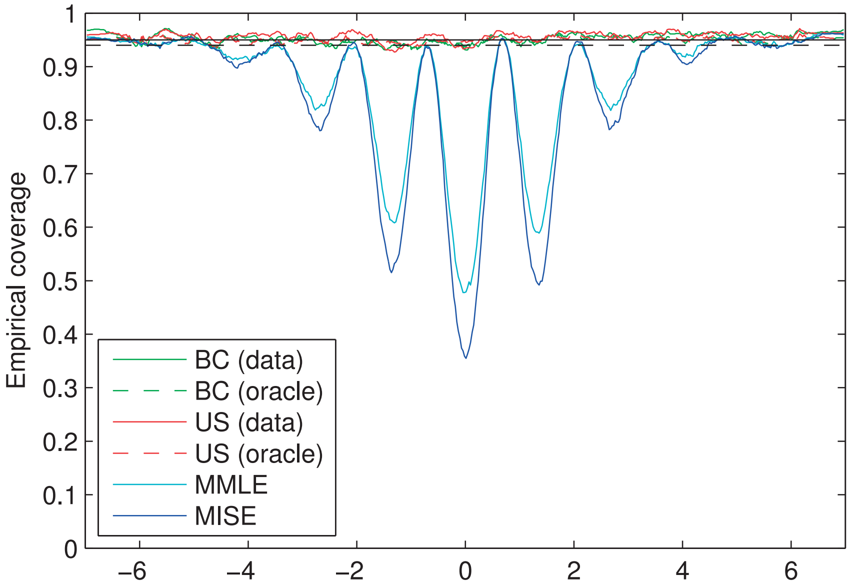


Figure B.12: Data-driven confidence intervals for the one-peak function f_1 and noise level $\sigma = 0.005$. The regularization strength is chosen using the MMLE and the amount of bias-correction and undersmoothing using the procedure described in Sections 6.4.3 and 6.4.4. The intervals have 95 % nominal coverage and 94 % target coverage. This figure is the same as Figure 6.10 in the main text.

(a) Empirical coverage (one peak, σ = 0.005)



(b) Mean interval length (one peak, σ = 0.005)

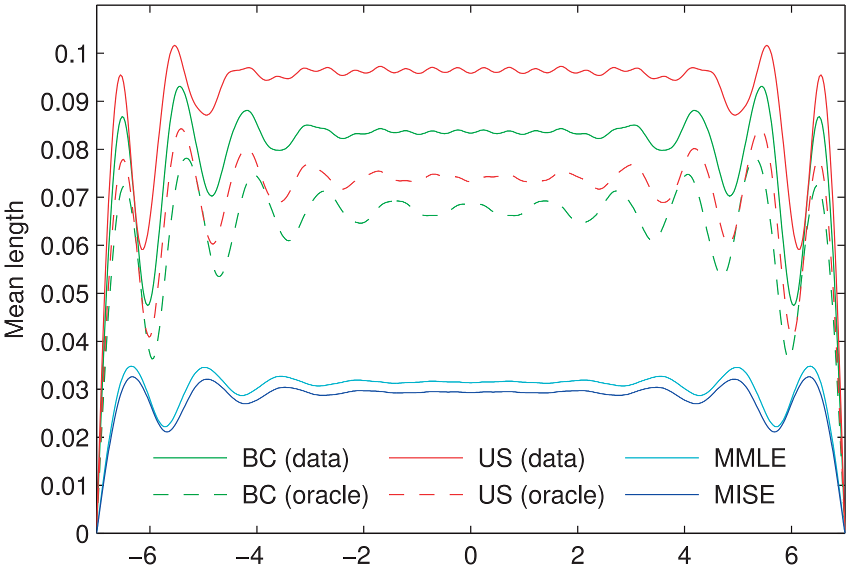
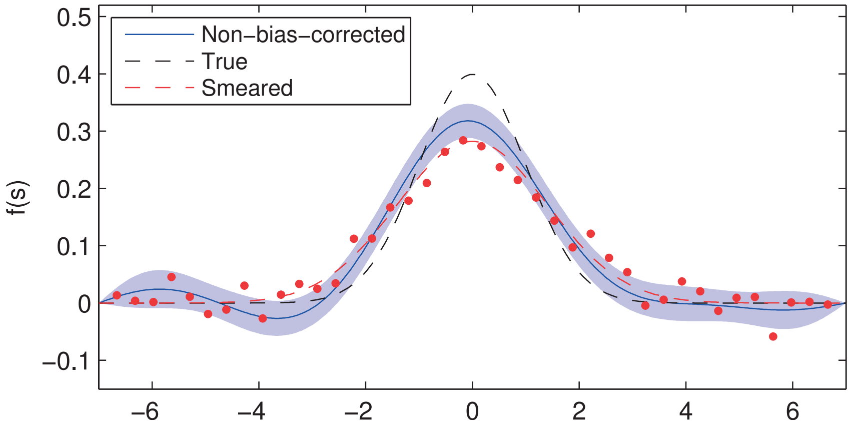
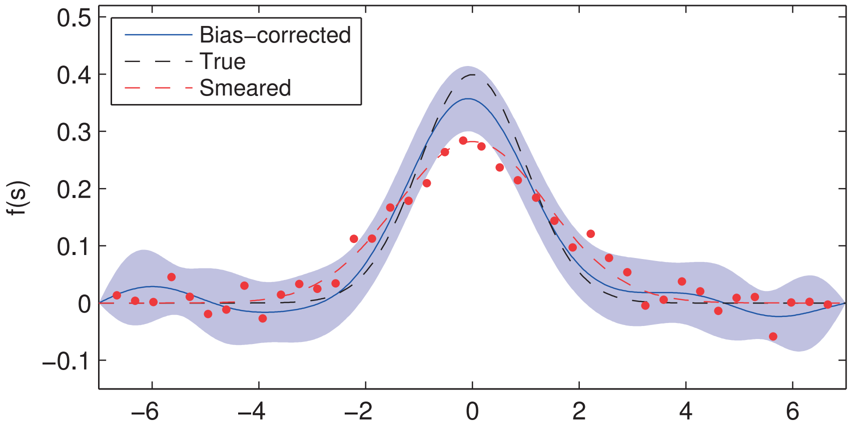


Figure B.13: Empirical coverage and mean length of the data-driven bias-corrected (BC), undersmoothed (US) and non-bias-corrected intervals for the one-peak function f_1 and noise level $\sigma = 0.005$. The curves labeled “data” are fully data-driven and the curves labeled “oracle” use knowledge of f_1 to choose the amount of debiasing. The non-bias-corrected results are given for both the MMLE choice of the regularization strength as well as for the choice that minimizes the MISE. The intervals have 95 % nominal coverage and 94 % target coverage. This figure is the same as Figure 6.11 in the main text.

(a) Non–bias–corrected intervals, σ = 0.025



(b) Iteratively bias–corrected intervals, σ = 0.025



(c) Undersmoothed intervals, σ = 0.025

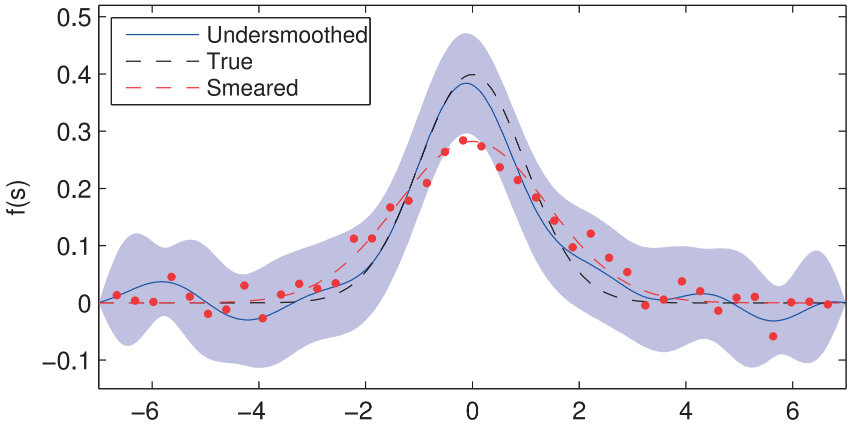
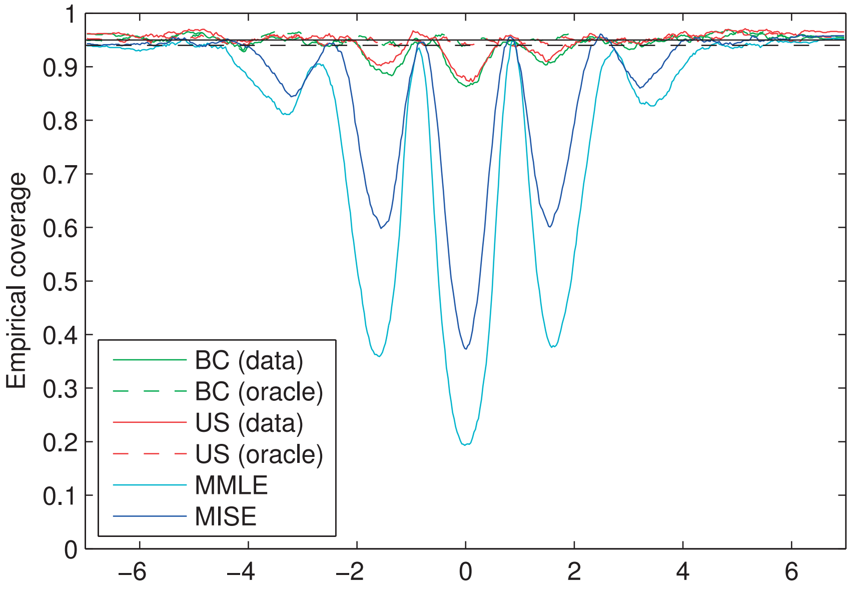


Figure B.14: Data-driven confidence intervals for the one-peak function f_1 and noise level $\sigma = 0.025$. The regularization strength is chosen using the MMLE and the amount of bias-correction and undersmoothing using the procedure described in Sections 6.4.3 and 6.4.4. The intervals have 95 % nominal coverage and 94 % target coverage.

(a) Empirical coverage (one peak, σ = 0.025)



(b) Mean interval length (one peak, σ = 0.025)

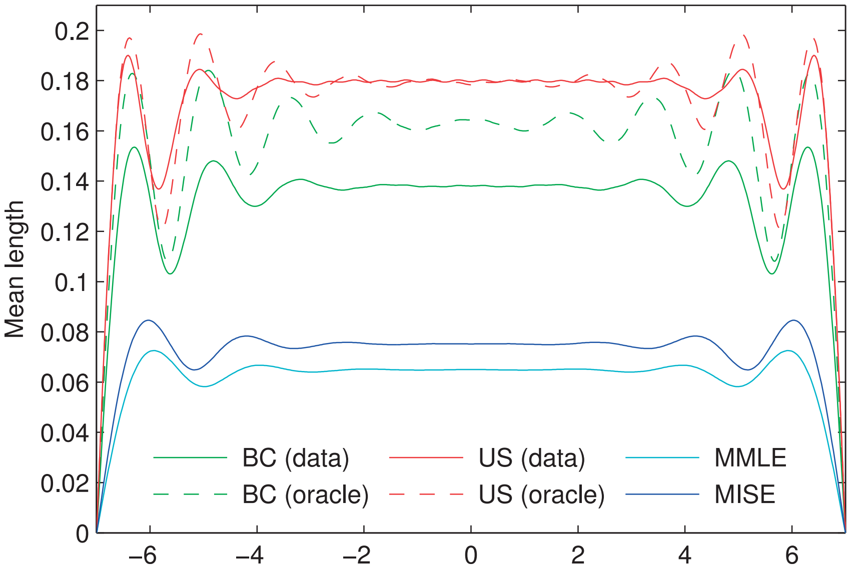
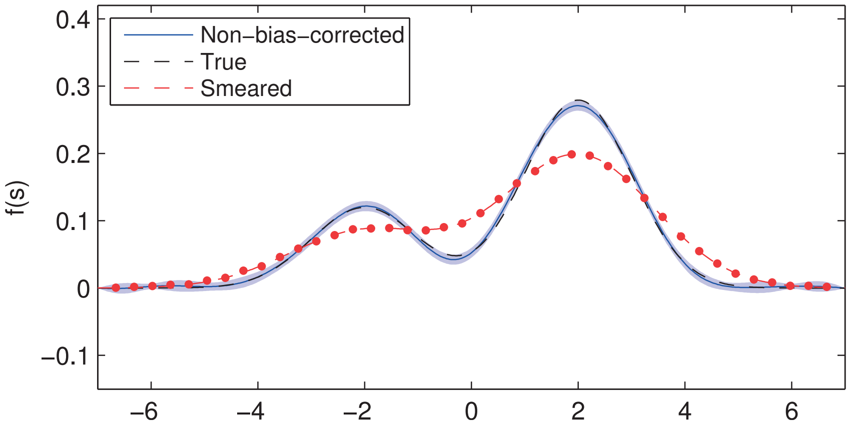
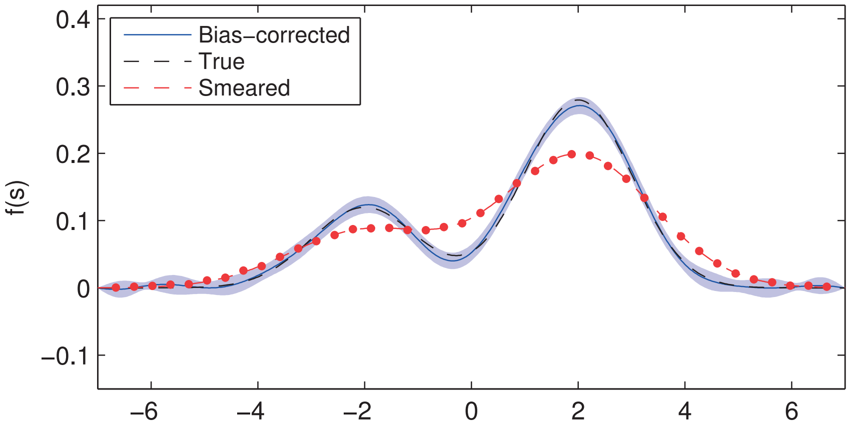


Figure B.15: Empirical coverage and mean length of the data-driven bias-corrected (BC), undersmoothed (US) and non-bias-corrected intervals for the one-peak function f_1 and noise level $\sigma = 0.025$. The curves labeled “data” are fully data-driven and the curves labeled “oracle” use knowledge of f_1 to choose the amount of debiasing. The non-bias-corrected results are given for both the MMLE choice of the regularization strength as well as for the choice that minimizes the MISE. The intervals have 95 % nominal coverage and 94 % target coverage.

(a) Non–bias–corrected intervals, σ = 0.001



(b) Iteratively bias–corrected intervals, σ = 0.001



(c) Undersmoothed intervals, σ = 0.001

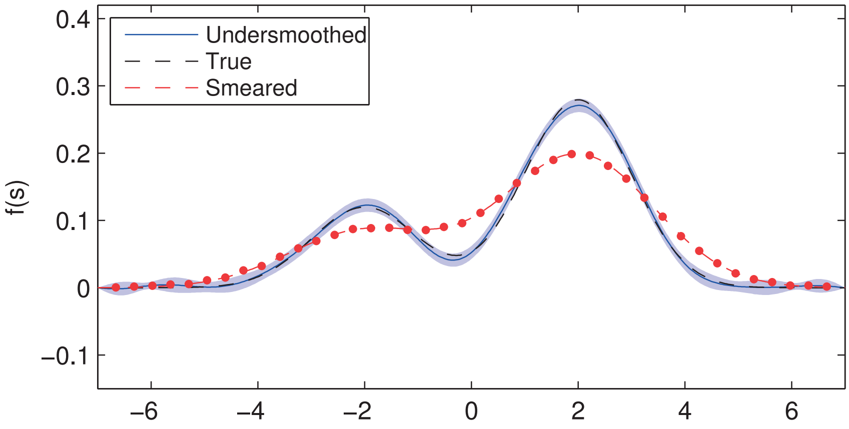
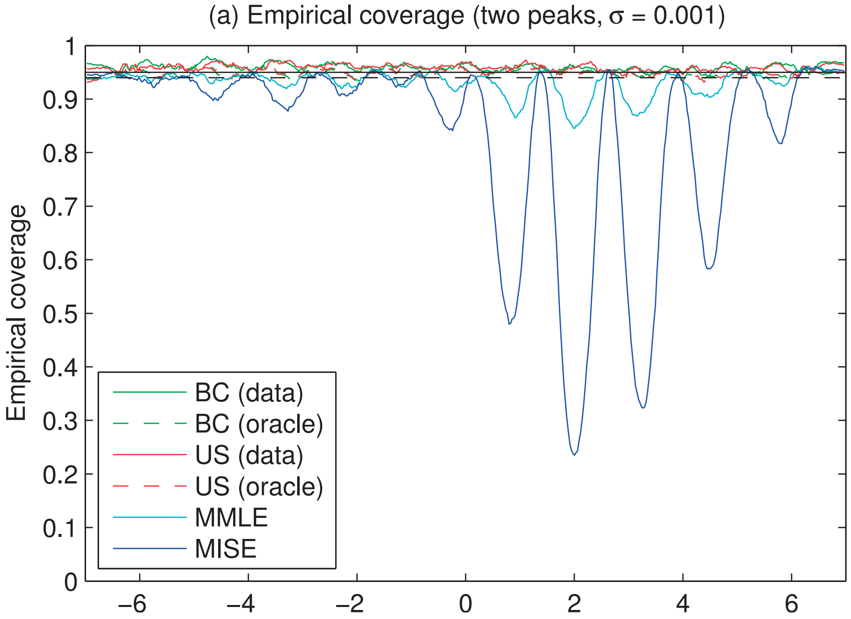


Figure B.16: Data-driven confidence intervals for the two-peak function f_2 and noise level $\sigma = 0.001$. The regularization strength is chosen using the MMLE and the amount of bias-correction and undersmoothing using the procedure described in Sections 6.4.3 and 6.4.4. The intervals have 95 % nominal coverage and 94 % target coverage.


(a) Empirical coverage (two peaks, σ = 0.001)

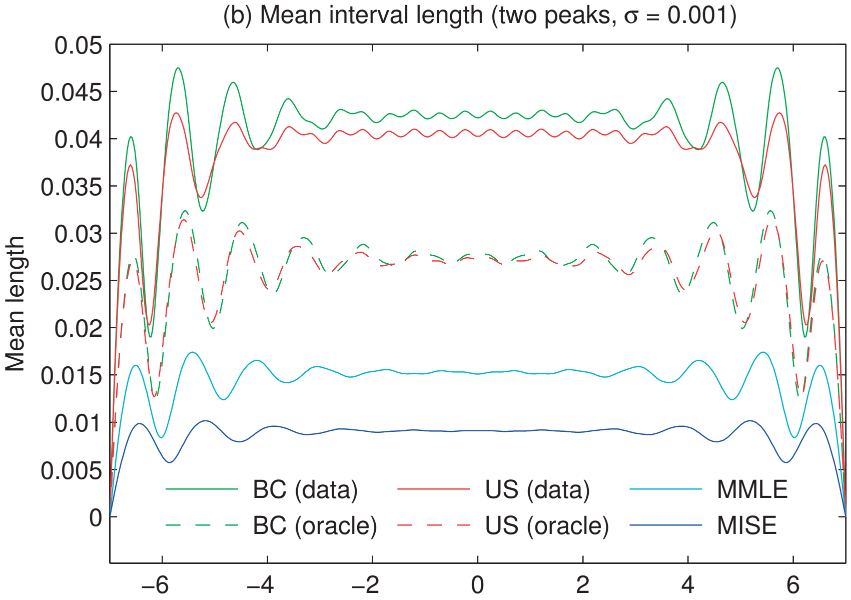
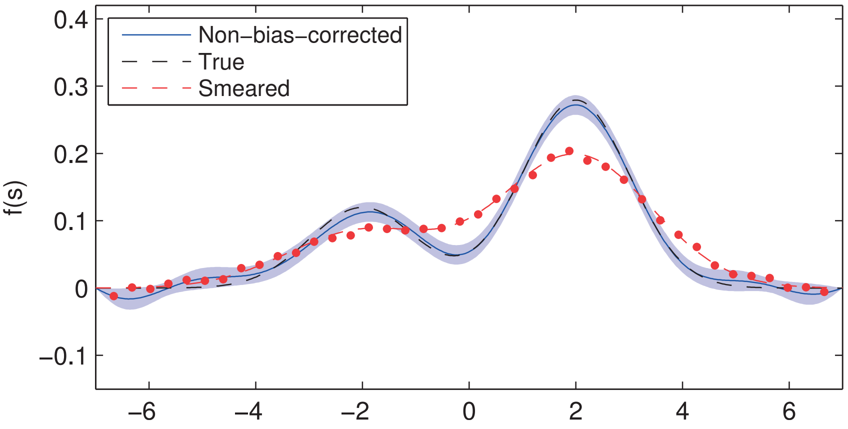
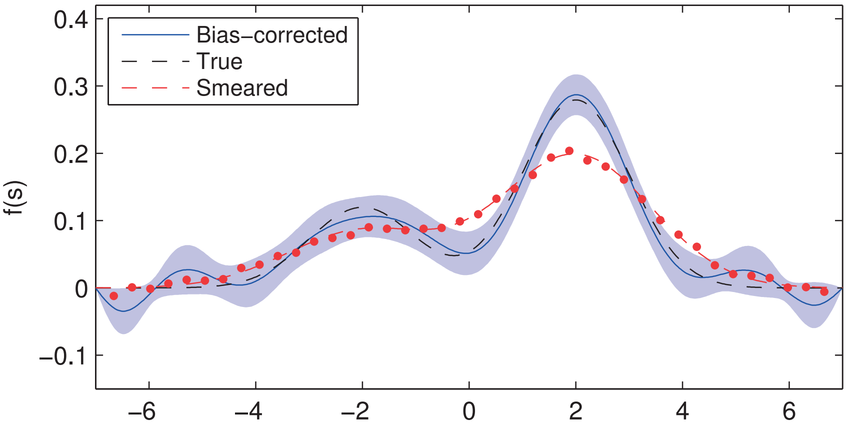

(b) Mean interval length (two peaks, σ = 0.001)

Figure B.17: Empirical coverage and mean length of the data-driven bias-corrected (BC), undersmoothed (US) and non-bias-corrected intervals for the two-peak function f_2 and noise level $\sigma = 0.001$. The curves labeled "data" are fully data-driven and the curves labeled "oracle" use knowledge of f_2 to choose the amount of debiasing. The non-bias-corrected results are given for both the MMLE choice of the regularization strength as well as for the choice that minimizes the MISE. The intervals have 95 % nominal coverage and 94 % target coverage.


(a) Non–bias–corrected intervals, σ = 0.005


(b) Iteratively bias–corrected intervals, σ = 0.005

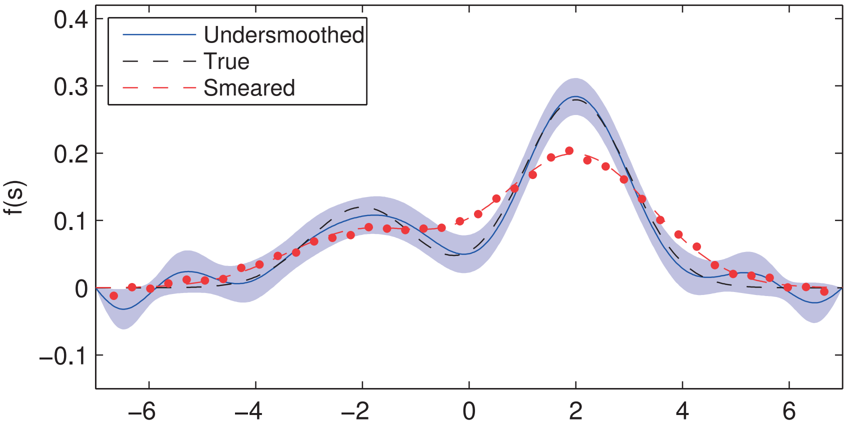
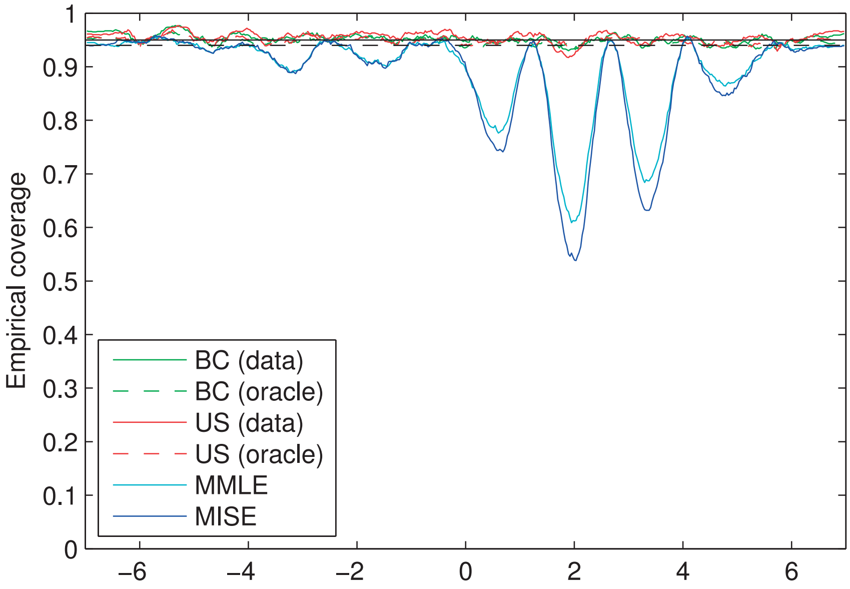

(c) Undersmoothed intervals, σ = 0.005

Figure B.18: Data-driven confidence intervals for the two-peak function f_2 and noise level $\sigma = 0.005$. The regularization strength is chosen using the MMLE and the amount of bias-correction and undersmoothing using the procedure described in Sections 6.4.3 and 6.4.4. The intervals have 95 % nominal coverage and 94 % target coverage.

(a) Empirical coverage (two peaks, σ = 0.005)



(b) Mean interval length (two peaks, σ = 0.005)

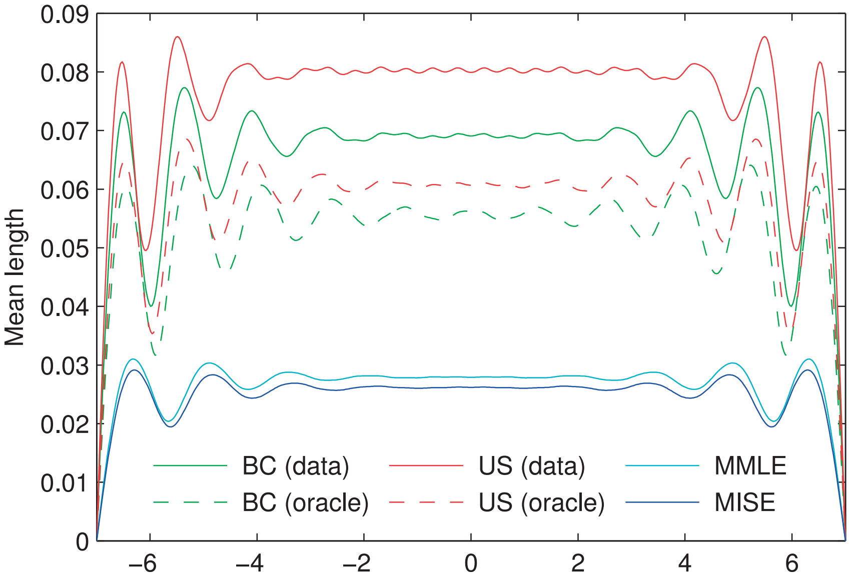
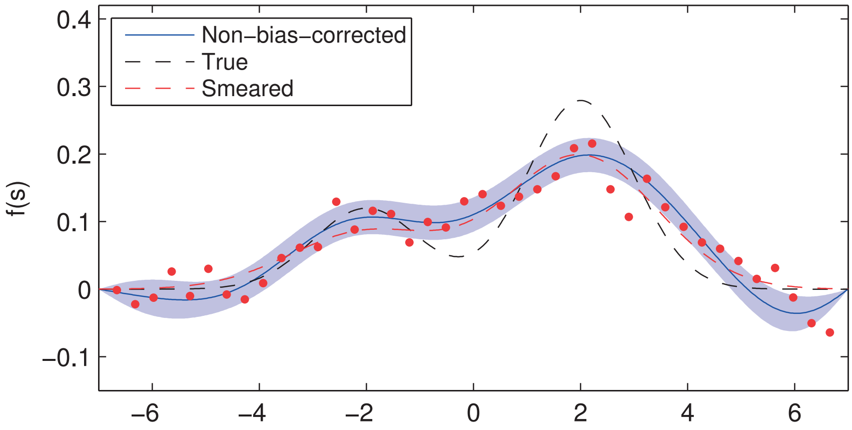
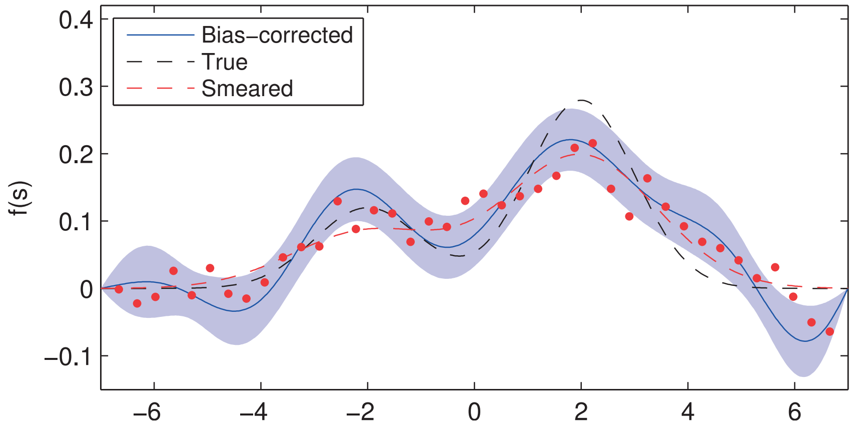


Figure B.19: Empirical coverage and mean length of the data-driven bias-corrected (BC), undersmoothed (US) and non-bias-corrected intervals for the two-peak function f_2 and noise level $\sigma = 0.005$. The curves labeled "data" are fully data-driven and the curves labeled "oracle" use knowledge of f_2 to choose the amount of debiasing. The non-bias-corrected results are given for both the MMLE choice of the regularization strength as well as for the choice that minimizes the MISE. The intervals have 95 % nominal coverage and 94 % target coverage.

(a) Non–bias–corrected intervals, σ = 0.025



(b) Iteratively bias–corrected intervals, σ = 0.025



(c) Undersmoothed intervals, σ = 0.025

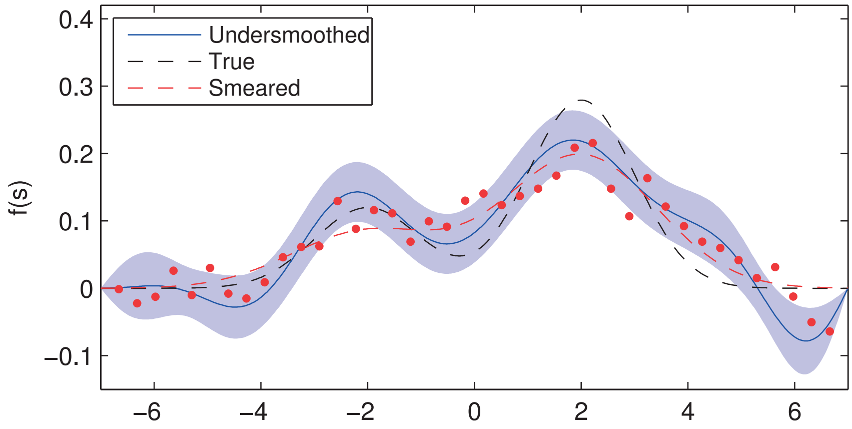
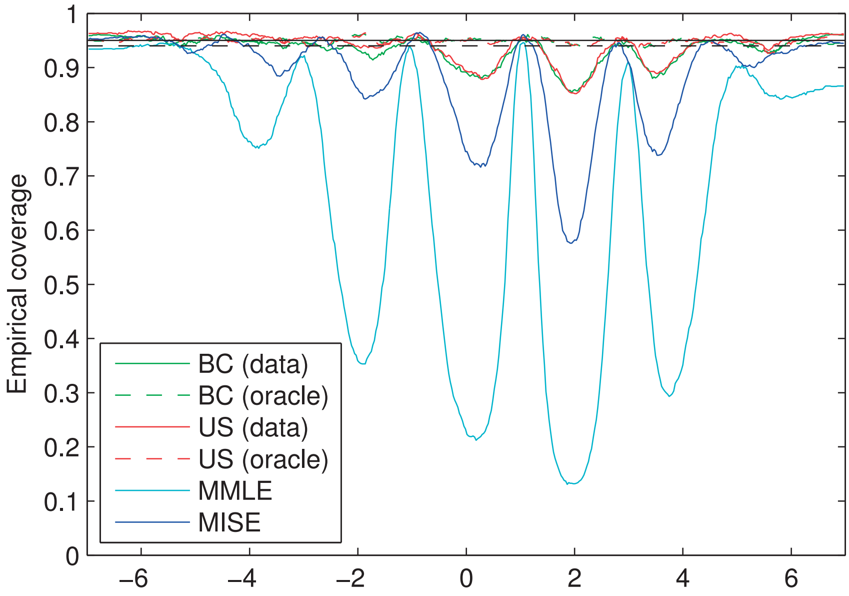


Figure B.20: Data-driven confidence intervals for the two-peak function f_2 and noise level $\sigma = 0.025$. The regularization strength is chosen using the MMLE and the amount of bias-correction and undersmoothing using the procedure described in Sections 6.4.3 and 6.4.4. The intervals have 95 % nominal coverage and 94 % target coverage.

(a) Empirical coverage (two peaks, σ = 0.025)



(b) Mean interval length (two peaks, σ = 0.025)

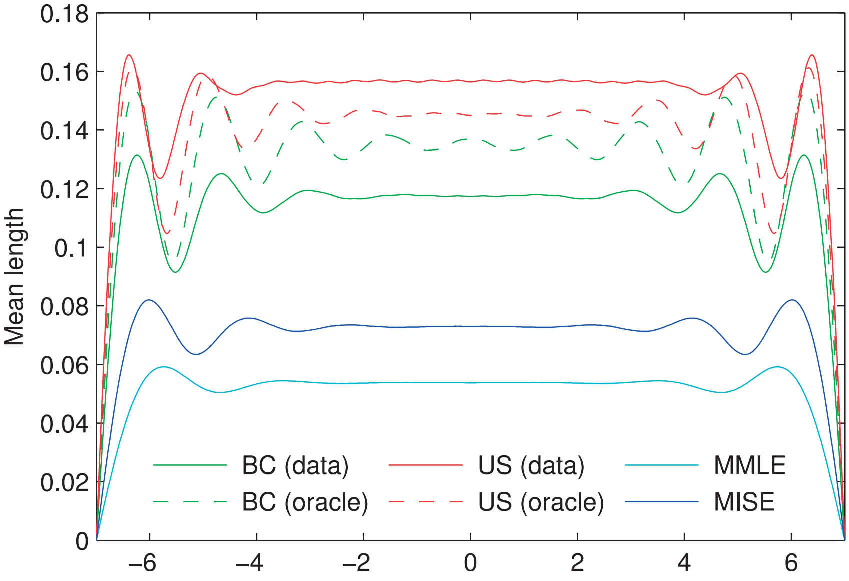
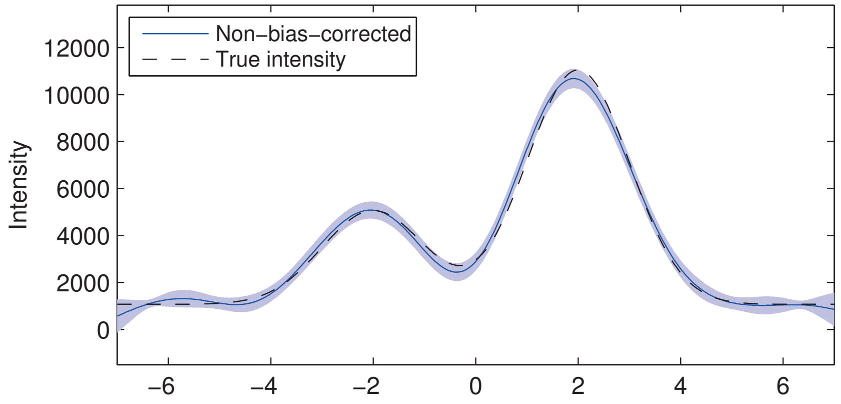


Figure B.21: Empirical coverage and mean length of the data-driven bias-corrected (BC), undersmoothed (US) and non-bias-corrected intervals for the two-peak function f_2 and noise level $\sigma = 0.025$. The curves labeled "data" are fully data-driven and the curves labeled "oracle" use knowledge of f_2 to choose the amount of debiasing. The non-bias-corrected results are given for both the MMLE choice of the regularization strength as well as for the choice that minimizes the MISE. The intervals have 95 % nominal coverage and 94 % target coverage.

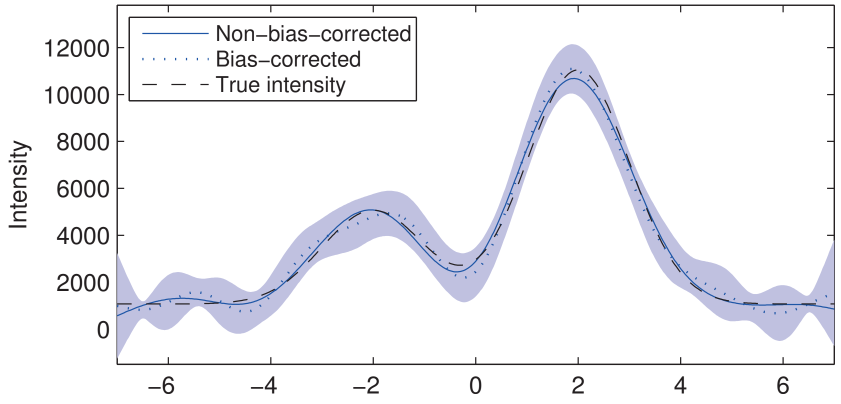
B.3 Full simulation results for Section 6.5

This section gives the full results of the simulation study of Section 6.5. In particular, the following pages include a realization of the different intervals as well as the equivalents of Figures 6.13 and 6.14 for the three sample sizes $\lambda_{\text{tot}} = 1\,000$, 10 000 and 50 000. The results are given in the order of decreasing sample size.

(a) Non–bias–corrected Gaussian intervals, λ_{tot} = 50000



(b) Iteratively bias–corrected Gaussian intervals, λ_{tot} = 50000



(c) Undersmoothed Gaussian intervals, λ_{tot} = 50000

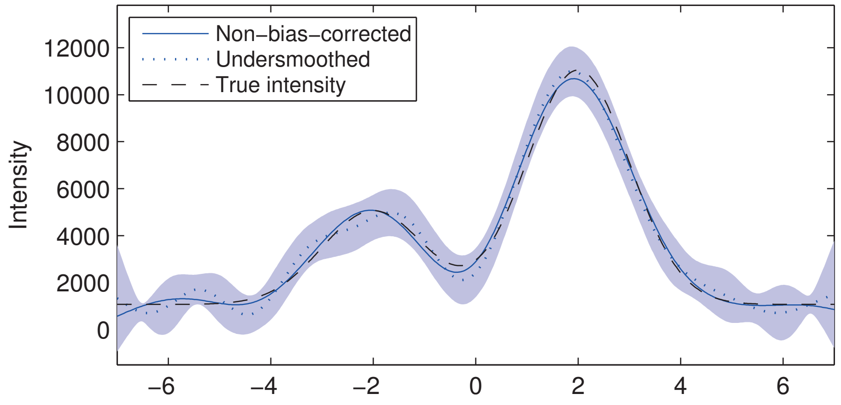
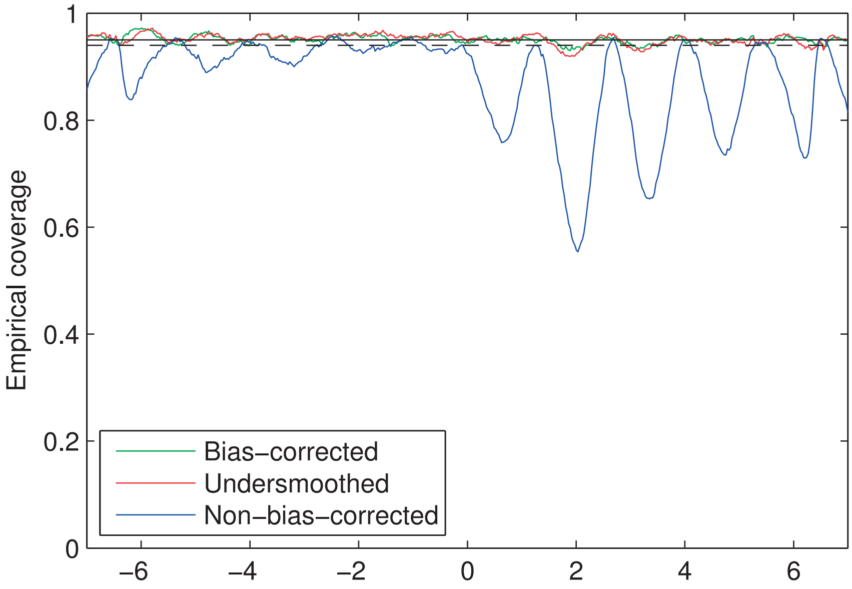


Figure B.22: Unfolding of the two peaks on a uniform background test setup using Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. The figure shows (a) the non-bias-corrected intervals, (b) the data-driven iteratively bias-corrected intervals and (c) the data-driven undersmoothed intervals. The amount of bias-correction and undersmoothing is calibrated to give 94 % pointwise target coverage for intervals with 95 % nominal coverage. The sample size was $\lambda_{\mathrm{tot}} = 50\,000$ and the regularization strength chosen using the MMLE.

(a) Empirical coverage, λ_{tot} = 50000



(b) Mean interval length, λ_{tot} = 50000

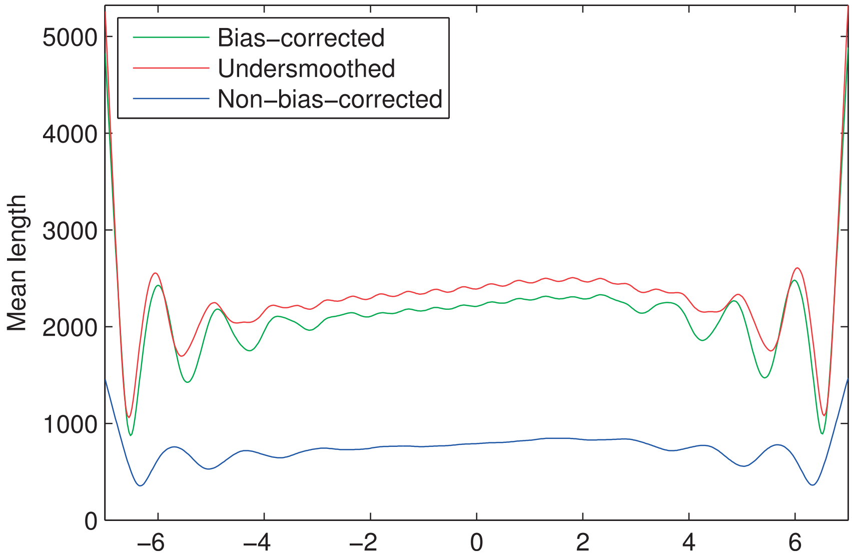


Figure B.23: Empirical coverage and mean interval length in unfolding the two peaks on a uniform background test setup using data-driven iteratively bias-corrected, undersmoothed and non-bias-corrected Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. The sample size was $\lambda_{\mathrm{tot}} = 50\,000$ and the regularization strength chosen using the MMLE. The intervals have 94 % pointwise target coverage (dashed horizontal line) and 95 % nominal coverage (solid horizontal line).

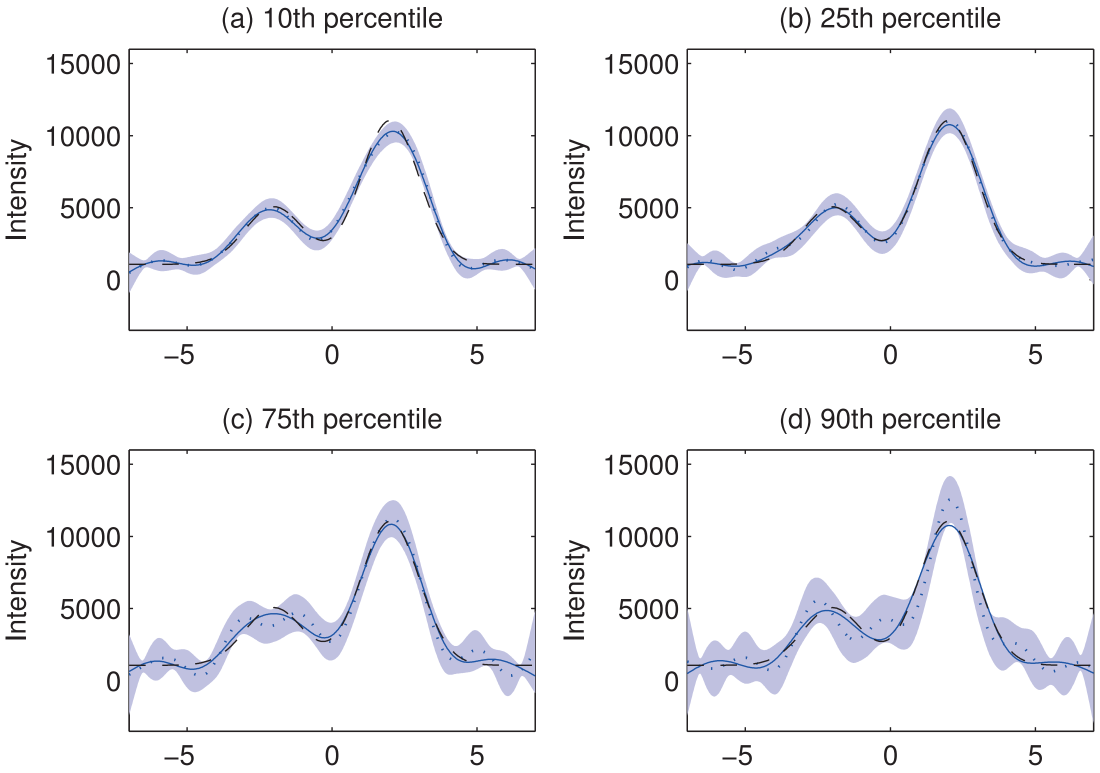
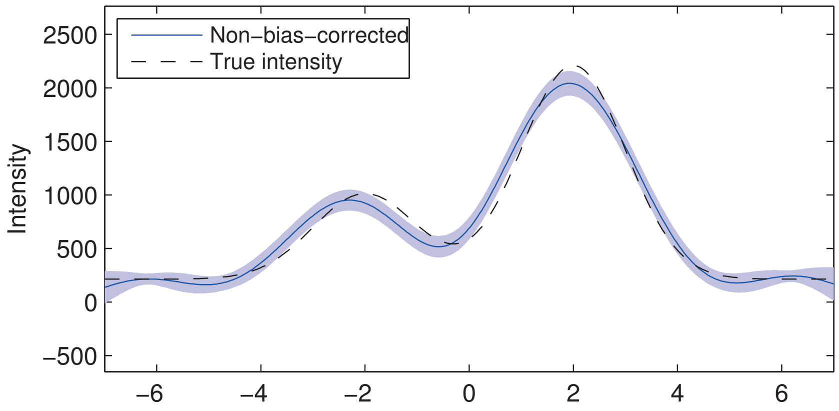
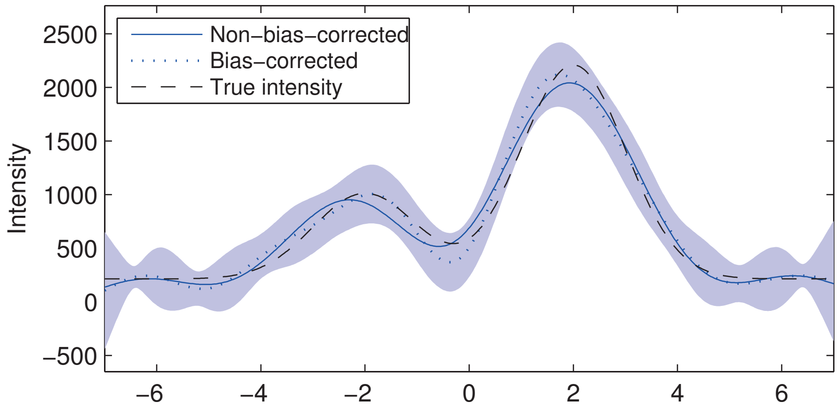


Figure B.24: Illustration of the variability of the data-driven iteratively bias-corrected Gaussian confidence intervals in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\text{tot}} = 50\,000$. The panels show realizations of the intervals for given percentiles of the sampling distribution of interval lengths (averaged over s). The intervals have 94 % pointwise target coverage and 95 % nominal coverage. The figures also show the true intensity (dashed line), the non-bias-corrected point estimate (solid line) and the bias-corrected point estimate (dotted line).

(a) Non−bias−corrected Gaussian intervals, $\lambda_{\text{tot}} = 10000$



(b) Iteratively bias−corrected Gaussian intervals, $\lambda_{\text{tot}} = 10000$



(c) Undersmoothed Gaussian intervals, $\lambda_{\text{tot}} = 10000$

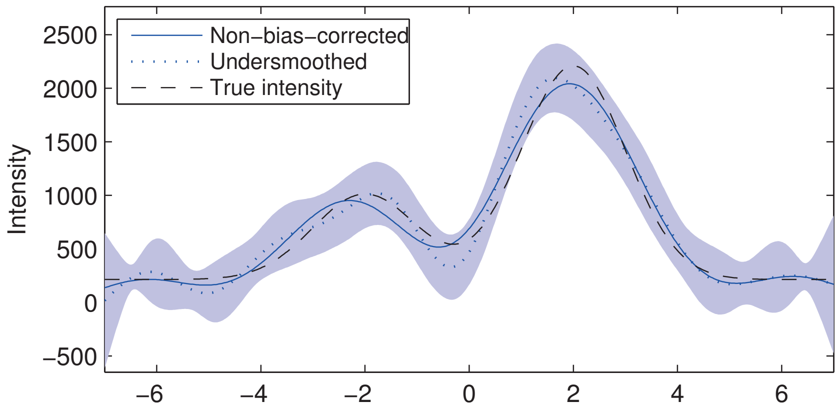
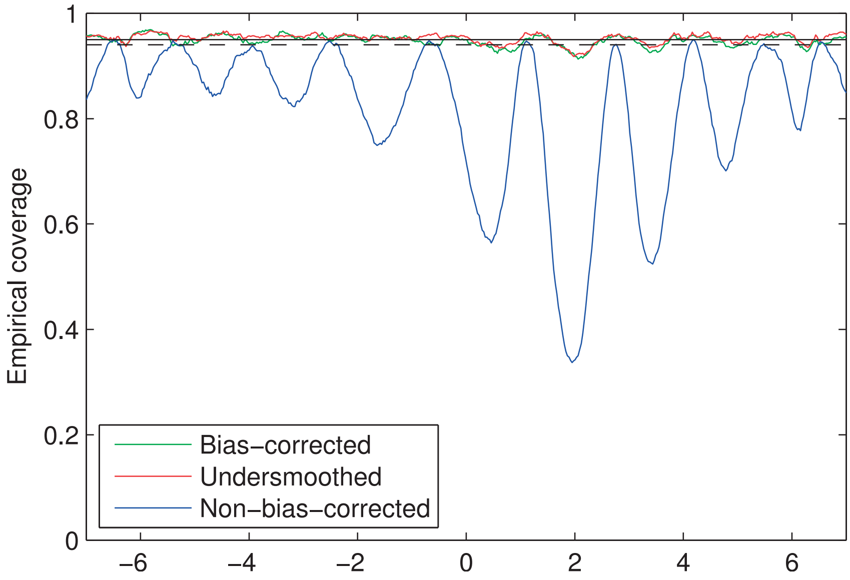


Figure B.25: Unfolding of the two peaks on a uniform background test setup using Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\text{G}}$. The figure shows (a) the non-bias-corrected intervals, (b) the data-driven iteratively bias-corrected intervals and (c) the data-driven undersmoothed intervals. The amount of bias-correction and undersmoothing is calibrated to give 94 % pointwise target coverage for intervals with 95 % nominal coverage. The sample size was $\lambda_{\text{tot}} = 10\,000$ and the regularization strength chosen using the MMLE.

(a) Empirical coverage, λ_{tot} = 10000



(b) Mean interval length, λ_{tot} = 10000

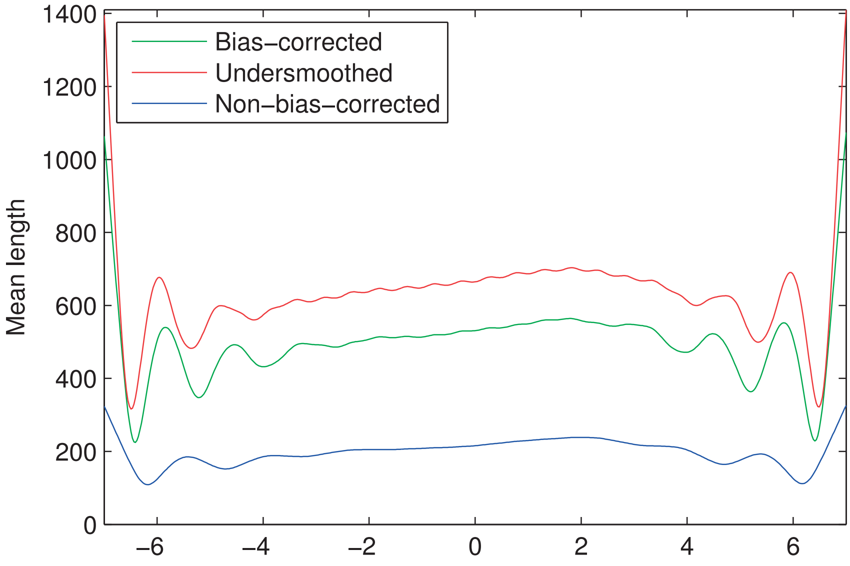


Figure B.26: Empirical coverage and mean interval length in unfolding the two peaks on a uniform background test setup using data-driven iteratively bias-corrected, undersmoothed and non-bias-corrected Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. The sample size was $\lambda_{\mathrm{tot}} = 10\,000$ and the regularization strength chosen using the MMLE. The intervals have 94 % pointwise target coverage (dashed horizontal line) and 95 % nominal coverage (solid horizontal line). This figure is the same as Figure 6.13 in the main text.

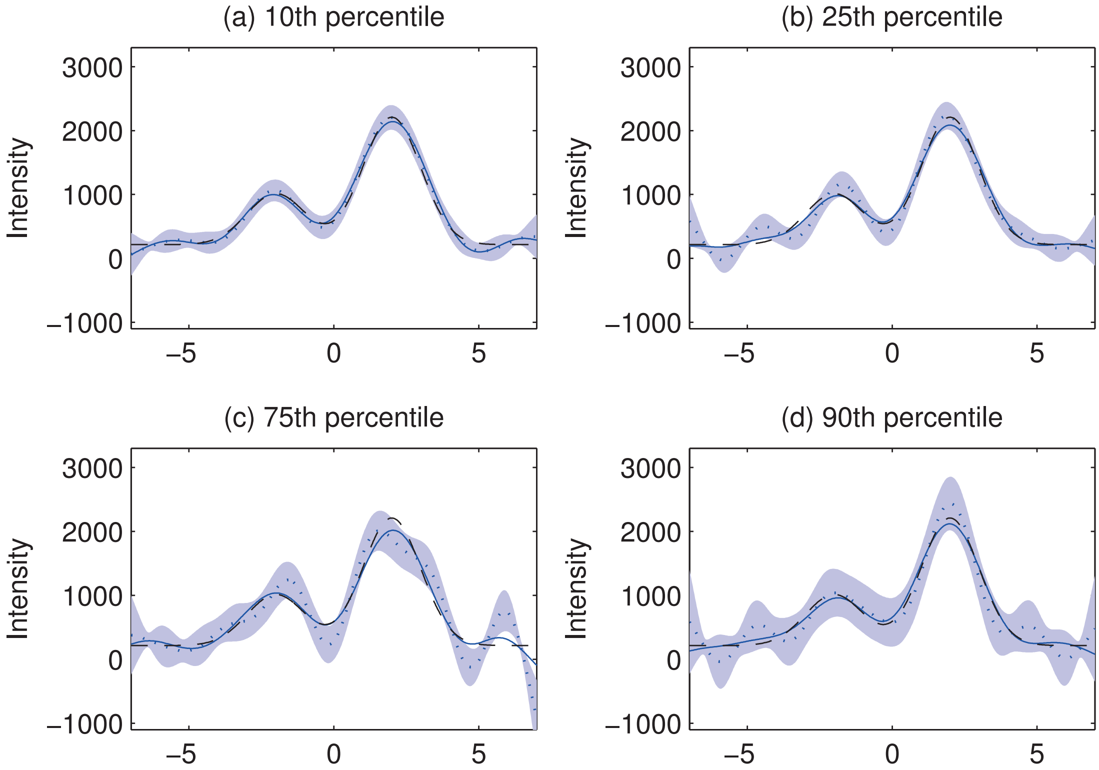
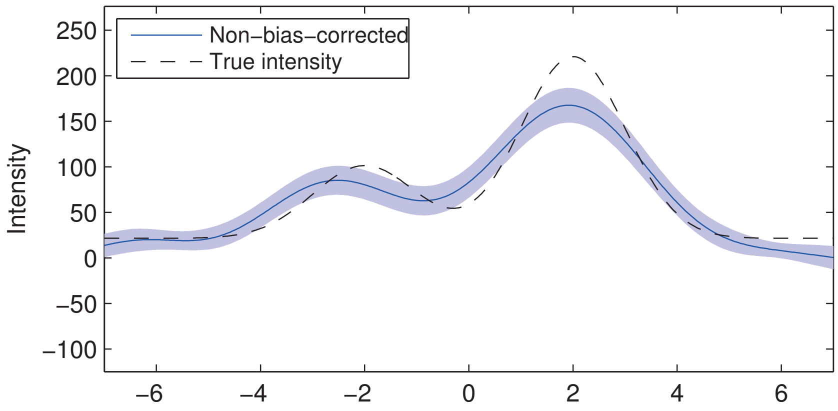
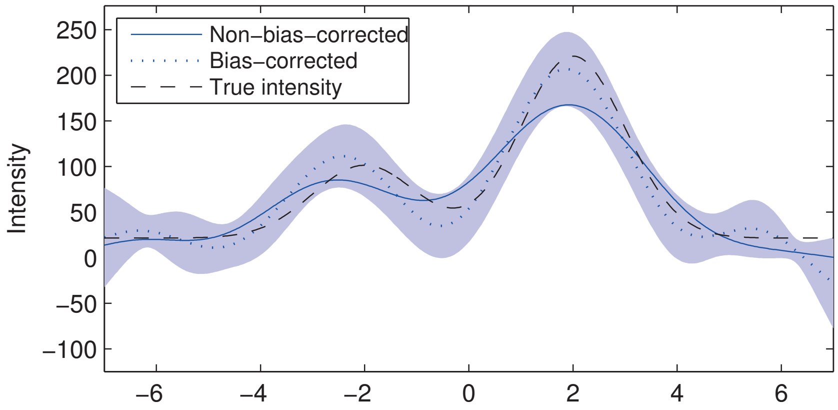


Figure B.27: Illustration of the variability of the data-driven iteratively bias-corrected Gaussian confidence intervals in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\text{tot}} = 10\,000$. The panels show realizations of the intervals for given percentiles of the sampling distribution of interval lengths (averaged over s). The intervals have 94 % pointwise target coverage and 95 % nominal coverage. The figures also show the true intensity (dashed line), the non-bias-corrected point estimate (solid line) and the bias-corrected point estimate (dotted line). This figure is the same as Figure 6.14 in the main text.

(a) Non−bias−corrected Gaussian intervals, λ_{tot} = 1000



(b) Iteratively bias−corrected Gaussian intervals, λ_{tot} = 1000



(c) Undersmoothed Gaussian intervals, λ_{tot} = 1000

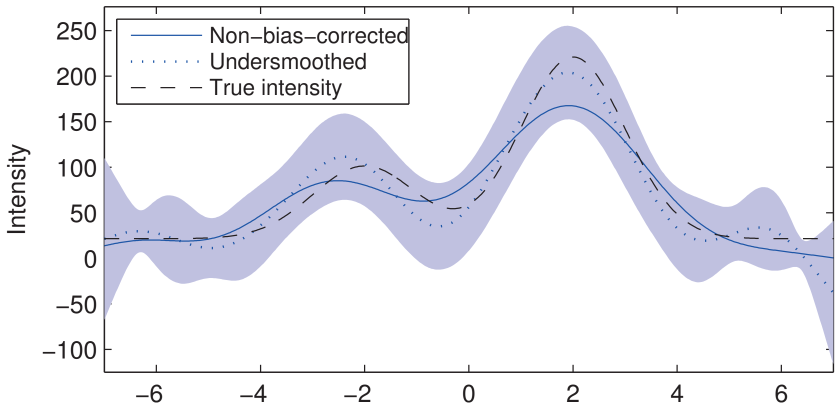
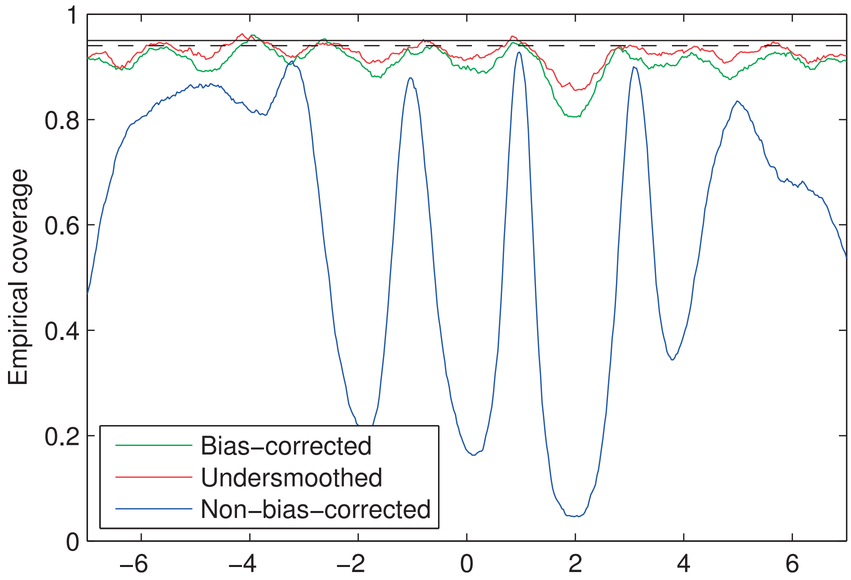


Figure B.28: Unfolding of the two peaks on a uniform background test setup using Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_{\mathrm{G}}$. The figure shows (a) the non-bias-corrected intervals, (b) the data-driven iteratively bias-corrected intervals and (c) the data-driven undersmoothed intervals. The amount of bias-correction and undersmoothing is calibrated to give 94 % pointwise target coverage for intervals with 95 % nominal coverage. The sample size was $\lambda_{\mathrm{tot}} = 1\,000$ and the regularization strength chosen using the MMLE.

(a) Empirical coverage, λ_{tot} = 1000



(b) Mean interval length, λ_{tot} = 1000

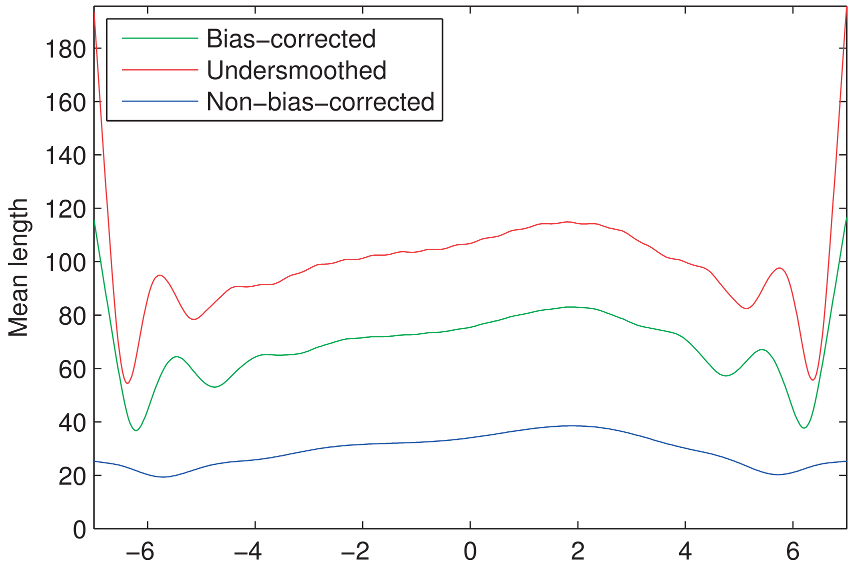


Figure B.29: Empirical coverage and mean interval length in unfolding the two peaks on a uniform background test setup using data-driven iteratively bias-corrected, undersmoothed and non-bias-corrected Gaussian confidence intervals induced by the unconstrained Gaussian estimator $\hat{\boldsymbol{\beta}}_G$. The sample size was $\lambda_{tot} = 1\,000$ and the regularization strength chosen using the MMLE. The intervals have 94 % pointwise target coverage (dashed horizontal line) and 95 % nominal coverage (solid horizontal line).

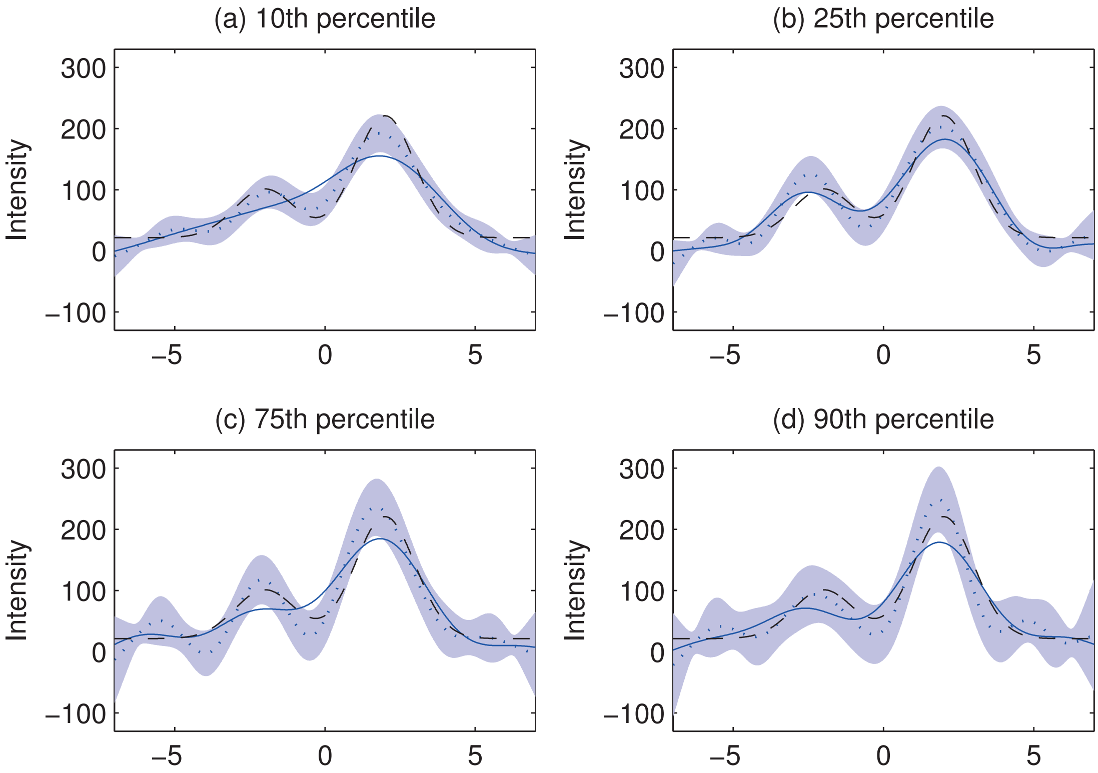


Figure B.30: Illustration of the variability of the data-driven iteratively bias-corrected Gaussian confidence intervals in unfolding the two peaks on a uniform background test setup with sample size $\lambda_{\mathrm{tot}} = 1\,000$. The panels show realizations of the intervals for given percentiles of the sampling distribution of interval lengths (averaged over s). The intervals have 94 % pointwise target coverage and 95 % nominal coverage. The figures also show the true intensity (dashed line), the non-bias-corrected point estimate (solid line) and the bias-corrected point estimate (dotted line).

Bibliography

T. Adye. Unfolding algorithms and tests using RooUnfold. In H. B. Prosper and L. Lyons, editors, *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, CERN-2011-006, pages 313–318, CERN, Geneva, Switzerland, 17–20 January 2011.

H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Czáki, editors, *Second International Symposium on Information Theory*, pages 267–281, Akademiai Kiadó, Budapest, 1973.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

A. Antoniadis and J. Bigot. Poisson inverse problems. *The Annals of Statistics*, 34(5):2132–2158, 2006.

ATLAS and CMS Collaborations. Procedure for the LHC Higgs boson search combination in summer 2011. ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005, 2011.

ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012a.

ATLAS Collaboration. Measurement of the transverse momentum distribution of W bosons in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Physical Review D*, 85:012005, 2012b.

G. Backus. Inference from inadequate and inaccurate data, I. *Proceedings of the National Academy of Sciences*, 65(1):1–7, 1970.

D. Barney. CMS-doc-4172. https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=4172, 2004. Retrieved 21.1.2014.

M. Bazaraa, H. Sherali, and C. Shetty. *Nonlinear Programming: Theory and Algorithms*. Wiley-Interscience, 3rd edition, 2006.

O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius, editors. *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*. Wiley, 2013.

R. Beran. Stein confidence sets and the bootstrap. *Statistica Sinica*, 5:109–127, 1995.

J. Beringer et al. Review of particle physics. *Physical Review D*, 86:010001, Particle Data Group, 2012.

R. Berk, L. Brown, A. Buja, K. Zhang, and L. Zhao. Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837, 2013.

J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley & Sons, 2000.

P. C. Bhat. Multivariate analysis methods in particle physics. *Annual Review of Nuclear and Particle Science*, 61:281–309, 2011.

V. Blobel. Unfolding methods in high-energy physics experiments. In *CERN Yellow Report 85-09*, pages 88–127, 1985.

V. Blobel. The RUN manual: Regularized unfolding for high-energy physics experiments. OPAL Technical Note TN361, 1996.

V. Blobel. Unfolding. In O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius, editors, *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*, pages 187–225. Wiley, 2013.

J. G. Booth and J. P. Hobert. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(1):265–285, 1999.

W. J. Browne and D. Draper. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1(3):473–514, 2006.

W. R. Burrus. Utilization of a priori information by means of mathematical programming in the statistical interpretation of measured distributions, June 1965. Oak Ridge National Laboratory, ORNL-3743.

W. R. Burrus and V. V. Verbinski. Fast-neutron spectroscopy with thick organic scintillators. *Nuclear Instruments and Methods*, 67(2):181–196, 1969.

D. Calvetti, J. P. Kaipio, and E. Someralo. Aristotelian prior boundary conditions. *International Journal of Mathematics and Computer Science*, 1:63–81, 2006.

B. P. Carlin and T. A. Louis. *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC, 3rd edition, 2009.

R. J. Carroll, A. Delaigle, and P. Hall. Testing and estimating shape-constrained nonparametric density and regression in the presence of measurement error. *Journal of the American Statistical Association*, 106(493):191–202, 2011.

G. Casella. Empirical Bayes Gibbs sampling. *Biostatistics*, 2(4):485–500, 2001.

G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, 2nd edition, 2002.

G. Choudalakis. Fully Bayesian unfolding. arXiv:1201.4612v4 [physics.data-an], 2012.

CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3: S08004, 2008.

CMS Collaboration. Particle-flow event reconstruction in CMS and performance for jets, taus, and $E_{\mathrm{T}}^{\mathrm{miss}}$. CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

CMS Collaboration. Measurement of the inclusive jet cross section in pp collisions at 7 TeV, 2010. CMS-PAS-QCD-10-011, available at http://cds.cern.ch/record/1280682.

CMS Collaboration. Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 7$ TeV. *Physical Review Letters*, 107:132001, 2011.

CMS Collaboration. Search for dark matter and large extra dimensions in monojet events in pp collisions at $\sqrt{s} = 7$ TeV. *Journal of High Energy Physics*, 09:094, 2012a.

CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012b.

CMS Collaboration. Shape, transverse size, and charged-hadron multiplicity of jets in pp collisions at $\sqrt{s} = 7$ TeV. *Journal of High Energy Physics*, 06:160, 2012c.

CMS Collaboration. Measurement of the charge asymmetry in top-quark pair production in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Physics Letters B*, 709(1–2):28–49, 2012d.

CMS Collaboration. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 8(09):P09009, 2013a.

CMS Collaboration. Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector. *Physical Review D*, 87:112002, 2013b.

CMS Collaboration. Measurement of differential top-quark-pair production cross sections in pp collisions at $\sqrt{s} = 7$ TeV. *The European Physical Journal C*, 73:2339, 2013c.

CMS Collaboration. Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 8$ TeV. *The European Physical Journal C*, 76:13, 2016.

P.-A. Cornillon, N. Hengartner, N. Jegou, and E. Matzner-Løber. Iterative bias reduction: a comparative study. *Statistics and Computing*, 23(6):777–791, 2013.

P.-A. Cornillon, N. W. Hengartner, and E. Matzner-Løber. Recursive bias estimation for multivariate regression smoothers. *ESAIM: Probability and Statistics*, 18:483–502, 2014.

G. Cowan. *Statistical Data Analysis*. Oxford University Press, 1998.

K. Cranmer. Practical statistics for the LHC. arXiv:1503.07622v1 [physics.data-an], earlier versions published in CERN Yellow Reports CERN-2014-003 and CERN-2015-001, 2015.

G. D'Agostini. A multidimensional unfolding method based on Bayes' theorem. *Nuclear Instruments and Methods A*, 362:487–498, 1995.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press, 1997.

C. de Boor. *A Practical Guide to Splines*. Springer, revised edition, 2001.

C. de Boor and J. W. Daniel. Splines with nonnegative B-spline coefficients. *Mathematics of Computation*, 28(126):565–568, 1974.

H. P. Dembinski and M. Roth. An algorithm for automatic unfolding of one-dimensional data distributions. *Nuclear Instruments and Methods in Physics Research A*, 729:410–416, 2013.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1): 1–38, 1977.

B. Efron. Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185, 1987.

B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.

P. H. C. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–102, 1996.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, 2000.

L. Fahrmeir, T. Kneib, and S. Lang. Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica*, 14(3):731–761, 2004.

S. Forte and G. Watt. Progress in the determination of the partonic structure of the proton. *Annual Review of Nuclear and Particle Science*, 63:291–328, 2013.

F. Garwood. Fiducial limits for the Poisson distribution. *Biometrika*, 28(3/4):437–442, 1936.

A. Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1(3):515–534, 2006.

S. Geman and D. E. McClure. Bayesian image analysis: An application to single photon emission tomography. In *Proceedings of the American Statistical Association, Statistical Computing Section*, pages 12–18, 1985.

S. Geman and D. E. McClure. Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, LII(4):5–21, 1987.

W. R. Gilks. Full conditional distributions. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 75–88. Chapman & Hall, 1996.

W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Introducing Markov chain Monte Carlo. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 1–19. Chapman & Hall, 1996.

H. Goldstein. Consistent estimators for multilevel generalized linear models using an iterated bootstrap. *Multilevel Modelling Newsletter*, 8(1):3–6, 1996.

P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, 9(1):84–93, 1990a.

P. J. Green. On use of the EM for penalized likelihood estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3):443–452, 1990b.

P. J. Green and B. W. Silverman. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman & Hall, 1994.

U. Grenander. On the theory of mortality measurement: Part II. *Scandinavian Actuarial Journal*, 1956(2):125–153, 1956.

P. Groeneboom and G. Jongbloed. *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics*. Cambridge University Press, 2014.

E. Gross and O. Vitells. Trial factors for the look elsewhere effect in high energy physics. *European Physical Journal C*, 70(1–2):525–530, 2010.

P. Hall. Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics*, 20(2):675–694, 1992.

P. Hall and J. Horowitz. A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, 41(4):1892–1921, 2013.

P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, 1992.

P. C. Hansen. *Discrete Inverse Problems: Insight and Algorithms*. SIAM, 2010.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. Chapman & Hall, 1990.

J. G. Heinrich. Coverage of error bars for Poisson data. CDF Public Note CDF/MEMO/STATISTICS/PUBLIC/6438, 2003.

N. W. Hengartner and P. B. Stark. Finite-sample confidence envelopes for shape-restricted densities. *The Annals of Statistics*, 23(2):525–550, 1995.

A. Höcker and V. Kartvelishvili. SVD approach to data unfolding. *Nuclear Instruments and Methods in Physics Research A*, 372:469–481, 1996.

A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

A. Javanmard and A. Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, 15(Oct):2869–2909, 2014.

J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2005.

A. Kondor. Method of convergent weights – An iterative procedure for solving Fredholm's integral equations of the first kind. *Nuclear Instruments and Methods*, 216:177–181, 1983.

E. Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons, 1978.

A. Y. C. Kuk. Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(2):395–407, 1995.

M. Kuusela. Statistical issues in unfolding methods for high energy physics. Master's thesis, Aalto University, 2012.

M. Kuusela and V. M. Panaretos. Statistical unfolding of elementary particle spectra: Empirical Bayes estimation and bias-corrected uncertainty quantification. *The Annals of Applied Statistics*, 9(3):1671–1705, 2015.

M. Kuusela and P. B. Stark. Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. arXiv:1512.00905v3 [stat.AP], submitted, 2016.

K. Lange and R. Carson. EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2):306–316, 1984.

C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. SIAM, 1995.

M. G. Low. On nonparametric confidence intervals. *The Annals of Statistics*, 25(6):2547–2554, 1997.

L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astronomical Journal*, 79(6):745–754, 1974.

D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969.

L. Lyons. Unfolding: Introduction. In H. B. Prosper and L. Lyons, editors, *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, CERN-2011-006, pages 225–228, CERN, Geneva, Switzerland, 17–20 January 2011.

L. Lyons. Bayes and frequentism: A particle physicist's perspective. *Contemporary Physics*, 54 (1):1–16, 2013.

G. Marra and S. N. Wood. Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39(1):53–74, 2012.

Mathworks. *Curve Fitting Toolbox: User's Guide*, 2014a. Version 3.4.1.

Mathworks. *Optimization Toolbox: User's Guide*, 2014b. Version 7.0.

G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley-Interscience, 2nd edition, 2008.

A. Meister. *Deconvolution Problems in Nonparametric Statistics*. Springer, 2009.

N. Milke, M. Doert, S. Klepser, D. Mazin, V. Blobel, and W. Rhode. Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics. *Nuclear Instruments and Methods in Physics Research A*, 697:133–147, 2013.

V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Mathematics Doklady*, 7:414–417, 1966.

H. N. Mülthei and B. Schorr. On an iterative method for a class of integral equations of the first kind. *Mathematical Methods in the Applied Sciences*, 9:137–168, 1987a.

H. N. Mülthei and B. Schorr. On an iterative method for the unfolding of spectra. *Nuclear Instruments and Methods in Physics Research A*, 257:371–377, 1987b.

H. N. Mülthei and B. Schorr. On properties of the iterative maximum likelihood reconstruction method. *Mathematical Methods in the Applied Sciences*, 11:331–342, 1989.

NNPDF Collaboration. Parton distributions for the LHC run II. *Journal of High Energy Physics*, 1504:040, 2015.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2nd edition, 2006.

D. Nychka. Confidence intervals for smoothing splines. *Journal of the American Statistical Association*, 83(404):1134–1143, 1988.

D. P. O'Leary and B. W. Rust. Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems. *SIAM Journal on Scientific and Statistical Computing*, 7(2):473–489, 1986.

M. J. Oreglia. *A study of the reactions* $\psi' \rightarrow \gamma\gamma\psi$. PhD thesis, Stanford University, 1980.

F. O'Sullivan. A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1(4): 505–527, 1986.

F. O'Sullivan. Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on Scientific and Statistical Computing*, 9(2):363–379, 1988.

V. M. Panaretos. *Statistics for Mathematicians: A rigorous first course*. Birkhäuser, 2016.

G. C. Pflug and R. J.-B. Wets. Shape-restricted nonparametric regression with overall noisy measurements. *Journal of Nonparametric Statistics*, 25(2):323–338, 2013.

D. L. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *Journal of the ACM*, 9(1):84–97, 1962.

A. Pintore, P. Speckman, and C. C. Holmes. Spatially adaptive smoothing splines. *Biometrika*, 93(1):113–125, 2006.

H. B. Prosper and L. Lyons, editors. *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, CERN-2011-006, CERN, Geneva, Switzerland, 2011.

H. B. Prosper, L. Lyons, and A. D. Roeck, editors. *Proceedings of the PHYSTAT LHC Workshop on Statistical Issues for LHC Physics*, CERN–2008–001, CERN, Geneva, Switzerland, 2008.

N. Pya and S. N. Wood. Shape constrained additive models. *Statistics and Computing*, 25(3): 543–559, 2015.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

P. T. Reiss and R. T. Ogden. Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71(2): 505–523, 2009.

R.-D. Reiss. *A Course on Point Processes*. Springer-Verlag, 1993.

W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, 1972.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2004.

T. Robertson, F. T. Wright, and R. L. Dykstra. *Order Restricted Statistical Inference*. John Wiley & Sons, 1988.

D. Ruppert and R. J. Carroll. Spatially-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics*, 42(2):205–223, 2000.

D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, 2003.

B. W. Rust and W. R. Burrus. *Mathematical Programming and the Numerical Solution of Linear Equations*. American Elsevier, 1972.

S. S. Saquib, C. A. Bouman, and K. Sauer. ML parameter estimation for Markov random fields with applications to Bayesian tomography. *IEEE Transactions on Image Processing*, 7(7): 1029–1044, 1998.

S. Schmitt. TUnfold, an algorithm for correcting migration effects in high energy physics. *Journal of Instrumentation*, 7:T10003, 2012.

L. L. Schumaker. *Spline Function: Basic Theory*. Cambridge University Press, 3rd edition, 2007.

L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982.

B. W. Silverman, M. C. Jones, J. D. Wilson, and D. W. Nychka. A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(2):271–324, 1990.

P. B. Stark. Inference in infinite-dimensional inverse problems: Discretization and duality. *Journal of Geophysical Research*, 97(B10):14055–14082, 1992.

M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36:111–147, 1974.

L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4): 1701–1728, 1994.

A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics: Doklady*, 4:1035–1038, 1963.

D. A. van Dyk. The role of statistics in the discovery of a Higgs boson. *Annual Review of Statistics and Its Application*, 1:41–59, 2014.

Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *Journal of the American Statistical Association*, 80(389):8–20, 1985.

E. Veklerov and J. Llacer. Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Transactions on Medical Imaging*, 6(4):313–319, 1987.

I. Volobouev. On the expectation-maximization unfolding with smoothing. arXiv:1408.6500v2 [physics.data-an], 2015.

G. Wahba. Constrained regularization for ill posed linear operator equations, with applications in meteorology and medicine. In S. S. Gupta and J. O. Berger, editors, *Statistical Decision Theory and Related Topics III*, volume 2, pages 383–418. Academic Press, 1982.

G. Wahba. Bayesian “confidence intervals” for the cross-validated smoothing spline. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(1):133–150, 1983.

G. Wahba. *Spline Models for Observational Data*. SIAM, 1990.

L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

G. C. G. Wei and M. A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704, 1990.

I. S. Weir. Fully Bayesian reconstructions from single-photon emission computed tomography data. *Journal of the American Statistical Association*, 92(437):49–60, 1997.

S. N. Wood. On confidence intervals for generalized additive models based on penalized regression splines. *Australian and New Zealand Journal of Statistics*, 48(4):445–464, 2006a.

S. N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, 2006b.

S. N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 73(1):3–36, 2011.

C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1): 95–103, 1983.

M. Xie and K. Singh. Confidence distribution, the frequentist distribution estimator of a parameter: A review (with discussion). *International Statistical Review*, 81(1):3–39, 2013.

G. A. Young and R. L. Smith. *Essentials of Statistical Inference*. Cambridge University Press, 2005.

G. Zech. Regularization and error assignment to unfolded distributions. In H. B. Prosper and L. Lyons, editors, *Proceedings of the PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, CERN-2011-006, pages 252–259, CERN, Geneva, Switzerland, 17–20 January 2011.

G. Zech. Iterative unfolding with the Richardson–Lucy algorithm. *Nuclear Instruments and Methods in Physics Research A*, 716:1–9, 2013.

G. Zech. Analysis of distorted measurements: Parameter estimation and unfolding. Preliminary version of an upcoming report on unfolding techniques, 2016.

CURRICULUM VITAE

July 10, 2016

Mikael Kuusela

EPFL SB MATHAA SMAT
MA B1 493 (Bâtiment MA)
Station 8
CH-1015 Lausanne
Switzerland

`mikael.kuusela@epfl.ch`
+41 21 69 35531
http://smat.epfl.ch/~kuusela/
Born on February 19, 1987 in Helsinki, Finland
Finnish nationality

EDUCATION

07/2016 **PhD in Statistics, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland**

- PhD advisor: Prof. Victor M. Panaretos
- Thesis title: *Uncertainty quantification in unfolding elementary particle spectra at the Large Hadron Collider*

08/2012 **MSc in Engineering Physics and Mathematics, Aalto University (formerly Helsinki University of Technology), Helsinki, Finland**

- Major: Mathematics, Minor: Particle and Astrophysics
- Thesis title: *Statistical Issues in Unfolding Methods for High Energy Physics*
- Erasmus exchange: EPFL, Lausanne, Switzerland, 08/2010–07/2011
- Degree awarded with distinction

06/2010 **BSc in Engineering Physics and Mathematics, Aalto University, Helsinki, Finland**

- Major: Systems Sciences, Minor: Computational Science and Engineering
- Thesis title: *Algorithms for Variational Learning of Mixture of Gaussians*

06/2006 **Matriculation Examination, Ressu Upper Secondary School, Helsinki, Finland**

- Final examination: Laudatur (top 5% nationwide) in all 6 subjects

RESEARCH INTERESTS

Statistics in the physical sciences, uncertainty quantification, statistical inverse problems, shape constraints, nonparametric statistics, semi-supervised learning, computational statistics

PROFESSIONAL EXPERIENCE

09/2012 – Present **Doctoral assistant, Institute of Mathematics, EPFL, Lausanne, Switzerland**
Estimation and inference in ill-posed inverse problems motivated by the high energy physics unfolding problem

01/2012 – Present **Statistics consultant, CERN, Geneva, Switzerland**
Statistics consultant on unfolding issues for the Statistics Committee of the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC)

10/2014 – 12/2014 **Visiting student researcher, Department of Statistics, University of California, Berkeley, CA, USA**
Shape-constrained uncertainty quantification in the high energy physics unfolding problem, hosted by Prof. Philip B. Stark

01/2012 – 07/2012 **Research assistant, Institute of Mathematics, EPFL, Lausanne, Switzerland and CERN, Geneva, Switzerland**
Empirical Bayes estimation in the high energy physics unfolding problem

05/2008 – 12/2011 **Research assistant, Department of Information and Computer Science, Aalto University, Helsinki, Finland**
Statistical machine learning algorithms with an emphasis on the Riemannian conjugate gradient descent in variational Bayesian learning

07/2011 – 09/2011 **Visiting researcher, Fermilab, Batavia, IL, USA**
Neural network classifiers for Higgs boson searches at the Tevatron

05/2010 – 08/2010 **Visiting researcher, CERN, Geneva, Switzerland**
Semi-supervised anomaly detection for new physics searches

06/2009 – 08/2009 **Summer student, CERN, Geneva, Switzerland**
Multi-class classification in analysis of diffractive physics at the LHC

PUBLICATIONS

In preparation

M. Kuusela and V. M. Panaretos. Iteratively bias-corrected uncertainty quantification for ill-posed inverse problems. In preparation, 2016.

Submitted

M. Kuusela and P. B. Stark. Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. arXiv:1512.00905 [stat.AP], submitted, 2016.

Published

M. Kuusela and V. M. Panaretos. Statistical unfolding of elementary particle spectra: Empirical Bayes estimation and bias-corrected uncertainty quantification. *The Annals of Applied Statistics*, 9(3):1671–1705, 2015.

T. Vatanen, **M. Kuusela**, E. Malmi, T. Raiko, T. Aaltonen, and Y. Nagai. Semi-supervised detection of collective anomalies with an application in high energy particle physics. In *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2012*, pages 1–8, Brisbane, Australia, June 2012.

M. Kuusela, T. Vatanen, E. Malmi, T. Raiko, T. Aaltonen, and Y. Nagai. Semi-supervised anomaly detection – towards model-independent searches of new physics. *Journal of Physics: Conference Series*, 368:012032, 2012.

M. Kuusela, E. Malmi, R. Orava, and T. Vatanen. Soft classification of diffractive interactions at the LHC. *AIP Conference Proceedings*, 1350(1):111–114, 2011.

A. Honkela, T. Raiko, **M. Kuusela**, M. Tornio, and J. Karhunen. Approximate Riemannian conjugate gradient learning for fixed-form variational Bayes. *Journal of Machine Learning Research,* 11(Nov):3235–3268, 2010.

M. Kuusela, J. W. Lämsä, E. Malmi, P. Mehtälä, and R. Orava. Multivariate techniques for identifying diffractive interactions at the LHC. *International Journal of Modern Physics A*, 25(8):1615–1647, 2010.

M. Kuusela, T. Raiko, A. Honkela, and J. Karhunen. A gradient-based algorithm competitive with variational Bayesian EM for mixture of Gaussians. In *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2009*, pages 1688–1695, Atlanta, Georgia, June 2009.

TEACHING EXPERIENCE

Teaching assistant

- *Linear Models* (regression models for 3rd year mathematicians, lectured by Prof. Victor M. Panaretos / Prof. Anthony C. Davison), EPFL, autumn 2013/2014/2015
- *Statistics for Mathematicians* (introductory statistics for 2nd year mathematicians, lectured by Prof. Victor M. Panaretos), EPFL, spring 2014/2015
- *Mathématiques générales I / II* (introductory mathematics for 1st year biologists and pharmacists, lectured by Dr. Darlene R. Goldstein), University of Lausanne, autumn 2012 / spring 2013

Supervision of semester projects

- Alexandre Mösching (MSc student in Mathematics), *Bayesian Regression Using Linear Basis Function Models,* EPFL, spring 2015
- Matthieu Simeoni (MSc student in Mathematics), *Statistical Inference in Positron Emission Tomography*, EPFL, spring 2014
- Louis Larmonier (BSc student in Mathematics), *Statistical Methods for Image Deblurring*, EPFL, spring 2013
- Alexandre David (MSc student in Computer Science), *Variational Bayes Inference for Bernoulli Mixture Models,* EPFL, autumn 2012

SELECTED TALKS (Electronic versions available at http://smat.epfl.ch/~kuusela/)

Invited talks

- Shape-Constrained Uncertainty Quantification in Unfolding Elementary Particle Spectra at the Large Hadron Collider, invited research talk, *EPSRC Workshop on Inverse Problems*, Edinburgh, UK, March 17, 2016
- Shape-Constrained Uncertainty Quantification in Unfolding Elementary Particle Spectra at the Large Hadron Collider, invited statistics seminar, University of Padova, Italy, December 4, 2015
- Machine Learning at CERN: The Current State-of-the-Art, the Kaggle Challenge and Some Open Problems, invited lunch talk, Kaggle Inc., San Francisco, CA, USA, November 18, 2014
- Empirical Bayes Unfolding of Elementary Particle Spectra at the Large Hadron Collider, invited statistics seminar, University of Bristol, UK, June 13, 2014
- Multivariate Techniques for Identifying Diffractive Interactions in Hadron-Hadron Colliders, invited high energy physics seminar, Iowa State University, Ames, IA, USA, August 30, 2011

Invited minicourses

- Introduction to Unfolding in High Energy Physics, ETH Zurich, Switzerland, July 15, 2014

Contributed conference talks

- Shape-Constrained Uncertainty Quantification in Unfolding Elementary Particle Spectra at the Large Hadron Collider, *3rd Conference of the International Society for Nonparametric Statistics, ISNPS III*, Avignon, France, June 16, 2016
- Shape-Constrained Uncertainty Quantification in Unfolding Elementary Particle Spectra at the Large Hadron Collider, *SIAM Conference on Uncertainty Quantification, UQ16*, Lausanne, Switzerland, April 5, 2016
- Statistical Methods for Unfolding Elementary Particle Spectra at the Large Hadron Collider, *Applied Inverse Problems 2015*, Helsinki, Finland, May 29, 2015
- Semi-Supervised Anomaly Detection – Towards Model-Independent Searches of New Physics, *14th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, ACAT 2011*, London, UK, September 6, 2011
- Multivariate Techniques for Identifying Diffractive Interactions at the LHC, *Diffraction 2010: International Workshop on Diffraction in High Energy Physics*, Otranto, Italy, September 15, 2010

Other

- Unfolding: Statistical Issues and Recommendations, *CMS Top Quark Physics Unfolding Workshop*, CERN, Geneva, Switzerland, May 13, 2016
- Unfolding: Statistical Issues, Recommendations and Latest Developments, *Topical CMS Statistics Committee Workshop*, CERN, Geneva, Switzerland, December 11, 2015
- A Statistician's Perspective on Unfolding, *CMS Standard Model Physics Unfolding Workshop*, CERN, Geneva, Switzerland, May 26, 2014
- The Statistics Behind the Discovery of the Higgs Boson, lunch talk, EPFL, Lausanne, Switzerland, April 10, 2014

AWARDS AND HONORS

- EPFL School of Basic Sciences Teaching Award, September 22, 2015
- Scholarship for excellence in studies, Prof. E. J. Nyström Foundation, January 27, 2012
- Scholarship for excellence in studies, Joint Student Foundation, December 14, 2011
- Valedictorian of the Finnish Air Force Reserve Officer Course 82, January 26, 2007
- Honorable Mention, *37th International Physics Olympiad, IPhO 2006*, Singapore, July 16, 2006
- Technology Industries of Finland award for full points in the Marticulation Examination in Advanced Mathematics, June 3, 2006

EXTERNAL AFFILIATIONS

- Member of the CMS experiment at CERN
- Member of the “*Advanced Multivariate Analysis for New Physics Searches at the LHC”* (AMVA4NewPhysics) research network

LANGUAGE SKILLS (CEFR level)

Finnish: mother tongue
English: fluent (C2, TOEFL iBT 120/120)
French: intermediate (B1)
Swedish: intermediate (B2)

OTHER

- Served as a referee for:
  - The Annals of Applied Statistics
  - Nuclear Instruments and Methods in Physics Research A
- Coauthor (with O. Behnke and I. Volobouev) of the internal unfolding recommendations for the CMS collaboration at CERN, February 15, 2016
- Featured in a promotional video for the EPFL Doctoral School (https://youtu.be/7JE3LExwI7Y)
- Military service: Second Lieutenant, Finnish Air Force, 07/2006 – 07/2007
- Hobbies: traveling, downhill skiing, hiking