# The importance of being random – and how to cope with it

Gunnar Flötteröd, Yu Chen

March 19, 2010

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# Outline

How to capture uncertainty in a model

How to reduce the uncertainty of model outputs

How to reduce the uncertainty of model inputs

Summary

# Outline

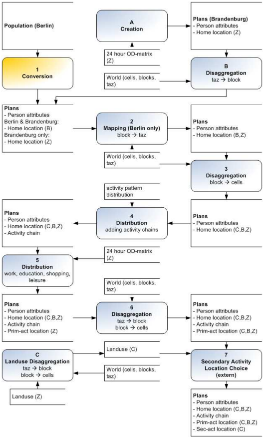
How to capture uncertainty in a model

How to reduce the uncertainty of model outputs

How to reduce the uncertainty of model inputs

Summary

# Random numbers everywhere




M. Balmer, Dissertation, ETHZ 2007

Listing 2.3: Fair vehicle movement at the intersections as in Cetin [24].

```
// Move vehicles across intersections:
 for all nodes
   while there are still eligible links
      Select an eligible link randomly proportional to capacity
      Mark link as non-eligible
      while vehicle has arrived at end of link
        AND vehicle can be moved according to capacity
        AND there is space on destination link
          move vehicle from source link to destination link
      end while
   end while  // eligible links
 end for  // all nodes
```

D. Strippgen, Dissertation, TUB 2009

**ExpBetaPlanSelector** selects a random plan according to a logit model: [11]

$$p_j = \frac{e^{\beta \cdot s_j}}{\sum_i e^{\beta \cdot s_i}} \tag{4.5}$$

where $p_j$ is the probability for plan $j$ to be selected and $s_j$ its current score. $\beta$ is a sensitivity parameter, set to 2.

M. Rieser, Dissertation, TUB 2010

# Justification

- there is an input of our model, say $X$, we are uncertain about
  - we model this uncertainty by assuming a distribution $f_X(x)$
  - we simulate this uncertainty by drawing realizations from $f_X(x)$
- this results in a random output of our model, say $Z = h(X)$
- almost any question about the model can now be ...

... phrased as $\mathrm{E}\{Z\} = \int h(x) f_X(x) dx = \,?$

... answer by $\mathrm{E}\{Z\} \approx \frac{1}{R} \sum_{r=1}^{R} h(x^r) \qquad x^r \sim f_X,\ r = 1 \dots R$

# Implications

- random input $\Rightarrow$ random output

  **a random simulation output represents uncertainty**

- identify uncertainty in the output
  - optimal: look at many simulation runs
  - at least: look at many relaxed iterations
- model uncertainty in the output
  - supplement averages with variances
  - make histograms, do statistical tests, ...

# Outline

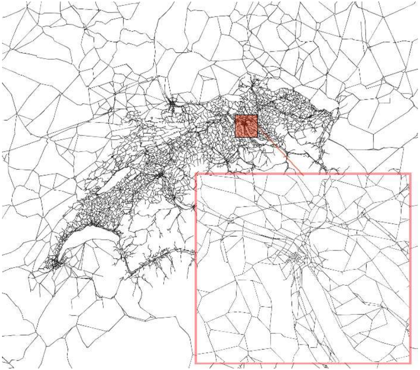
# Basic idea

- the more we know, the less uncertain we are
  - uncertainty = randomness
  - knowledge = data

**additional data reduces uncertainty in model outputs**

- that data must be related to the model outputs
- here: count cars to reduce uncertainty about travel behavior

# Zurich scenario



Grether et al., Report 08-10, TUB 2008

- configuration
  - network with 60 492 links and 24 180 nodes
  - 187 484 travelers
  - hourly vehicle counts from 161 sensors
- calibrate
  - route choice
  - departure time choice
  - mode choice

for every single agent

# Combine model output with additional data

- **model output:** simulated travel behavior is uncertain
  - $V_n(i)$ is utility of travel plan $i$ as perceived by driver $n$
  - $P_n(i) \sim \exp(V_n(i))$ is respective plan choice probability
- **additional data:** reduce uncertainty using traffic counts
  - $y_{ak}$ is traffic count on link $a$ in time step $k$
  - $\sigma^2_{ak}$ is variance of counting error
- making some assumptions and applying some math

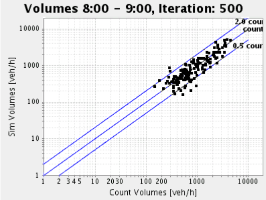
$$P_n(i|\{y_{ak}\}_{ak}) \sim \exp\left(V_n(i) + \sum_{ak \in i} \frac{y_{ak} - q_{ak}}{\sigma^2_{ak}}\right)$$

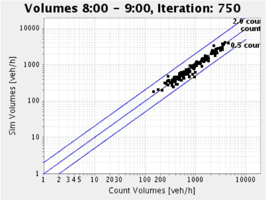
  - $q_{ak}$ is simulated flow on link $a$ in time step $k$
  - increases utility of more plausible plans

# Results, qualitatively

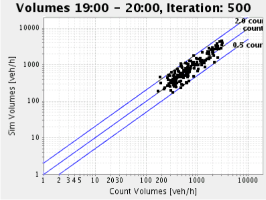
plain simulation | with calibration

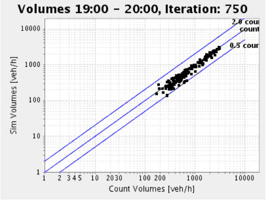
morning




evening



# Results, quantitatively

| | reproduction $(\cdot)^2$ error | validation $(\cdot)^2$ error | comp. time until stationarity |
|---|---|---|---|
| plain simulation | 103.6 | 103.6 | $18^1/_2$ h (500 it) |
| calibrated simulation | 20.9 | 75.1 | $20^1/_4$ h (500 it) |
| relative difference | - 80 % | - 28 % | + 9 % |

- 10-fold cross-validation
- negligible computational overhead
- very stable results

# Discussion

- predictive power of adjusted plan choice distributions
  - good within a day: plans apply to the whole day
  - poor beyond this: plans are not (yet) linked across days
- what structural (long-term) information can we get out of this?
  - essentially, we change the alternative specific constants (ASC)

$$P_n(i|\{y_{ak}\}_{ak}) \sim \exp\left(V_n(i) + \sum_{ak \in i} \frac{y_{ak} - q_{ak}}{\sigma^2_{ak}}\right)$$

  - ASC of a plan is sum over ASC components per link
  - questionable but possible: predict based on fixed link-ASC

# Outline

How to capture uncertainty in a model

How to reduce the uncertainty of model outputs

How to reduce the uncertainty of model inputs

Summary

# Basic idea

- the more we know, the less uncertain we are
  - uncertainty = randomness
  - knowledge = data

**additional data reduces uncertainty in model inputs**

- that data must be related to the model inputs
- here: count cars to reduce uncertainty about mode choice model

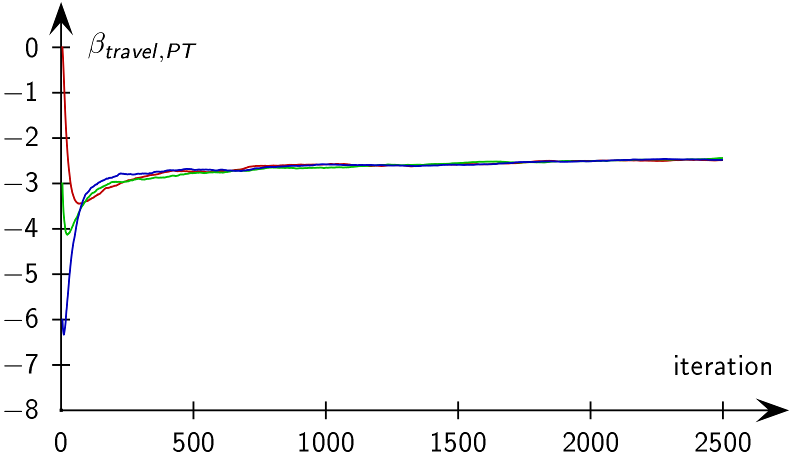
# Scenario description

- same Zurich scenario as before
- utility function for logit plan choice

$$\begin{aligned} V(\text{car-plan}) &= 2.0 \cdot (\beta_{travel,car}\, t_{travel} + \beta_{act}\, t_{act}) \\ V(\text{PT-plan}) &= 2.0 \cdot (\beta_{travel,PT}\, t_{travel} + \beta_{act}\, t_{act}) \end{aligned}$$

  where $\beta_{travel,car} = -6\,\mathrm{h}^{-1}$ and $\beta_{act} = +6\,\mathrm{h}^{-1}$
- maximum likelihood estimation of $\beta_{travel,PT}$
  - have *closed-form* approximations of gradient & Hessian
  - use Newton-Rhapson algorithm with MSA-like step size
  - do one parameter update per iteration of the simulation

# Parameter evolution over iterations



# Discussion

- final estimate $\beta_{travel,PT} = -2.47$
  - deviation between different runs $\sim 10^{-2}$
  - square root of inverse negative Hessian $\sim 10^{-3}$
- measures of fit
  - null log-likelihood $\sim -60.2$
  - final log-likelihood $\sim -54.6$
- criticism
  - many assumptions and approximations
  - simulation noise everywhere
  - a single parameter hardly explains the data

# Outline

How to capture uncertainty in a model

How to reduce the uncertainty of model outputs

How to reduce the uncertainty of model inputs

Summary

# Summary

1. appropriate input randomness reveals uncertainty in the results
   - fixing an input means to be perfectly sure about it
   - looking at only one output realization ignores its uncertainty
2. additional data helps to reduce this uncertainty
   - for both model inputs (parameters) and outputs
   - this talk only considers traffic counts
   - new data sources: vehicle identification, smart phones, ...
3. it is conceptually & computationally feasible to actually do this
   - there are some theoretical results by now
   - free software: `transp-or2.epfl.ch/cadyts/`

# Summary

1. appropriate input randomness reveals uncertainty in the results
   - fixing an input means to be perfectly sure about it
   - looking at only one output realization ignores its uncertainty
2. additional data helps to reduce this uncertainty
   - for both model inputs (parameters) and outputs
   - this talk only considers traffic counts
   - new data sources: vehicle identification, smart phones, ...
3. it is conceptually & computationally feasible to actually do this
   - there are some theoretical results by now
   - free software: `transp-or2.epfl.ch/cadyts/`

**Thank you for your attention.**

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE