

CCDFs in MOOCs

Tao C. Lee

LCAV, EPFL
TAM Talk, Lausanne, Switzerland
tao.lee@epfl.ch

June 2, 2014



Digital Signal Processing

Paolo Prandoni and Martin Vetterli

Learn the fundamentals of digital signal processing theory and discover the myriad ways DSP makes everyday life more productive and fun.





Digital Signal Processing

Paolo Prandoni and Martin Vetterli

Learn the fundamentals of digital signal processing theory and discover the myriad ways DSP makes everyday life more productive and fun.



More than 90,000 learners since spring '13 ...



Digital Signal Processing

Paolo Prandoni and Martin Vetterli

Learn the fundamentals of digital signal processing theory and discover the myriad ways DSP makes everyday life more productive and fun.



More than 90,000 learners since spring '13 ...

How to study learner dynamics?

- Complementary Cumulative Distribution Functions (CCDFs)

- Complementary Cumulative Distribution Functions (CCDFs)
- Final grades CCDFs in MOOCs

- Complementary Cumulative Distribution Functions (CCDFs)
- Final grades CCDFs in MOOCs
- Learner subpopulation analysis

- Complementary Cumulative Distribution Functions (CCDFs)
- Final grades CCDFs in MOOCs
- Learner subpopulation analysis
- CCDFs by learner subpopulations

- Complementary Cumulative Distribution Functions (CCDFs)
- Final grades CCDFs in MOOCs
- Learner subpopulation analysis
- CCDFs by learner subpopulations
- Contributions & take-home messages

Massive Open Online Courses (MOOCs)

- History: 2012, the Year of the MOOC

Massive Open Online Courses (MOOCs)

- History: 2012, the Year of the MOOC
- Players: Coursera, edX, Udacity, ...

Massive Open Online Courses (MOOCs)

- History: 2012, the Year of the MOOC
- Players: Coursera, edX, Udacity, ...
- Visibility: > 6 million learners on Coursera, early 2014

Massive Open Online Courses (MOOCs)

- History: 2012, the Year of the MOOC
- Players: Coursera, edX, Udacity, ...
- Visibility: > 6 million learners on Coursera, early 2014

MOOC datasets

Data source	Description
Database dumps	Quiz submits & grades, forum posts & comments

Massive Open Online Courses (MOOCs)

- History: 2012, the Year of the MOOC
- Players: Coursera, edX, Udacity, ...
- Visibility: > 6 million learners on Coursera, early 2014

MOOC datasets

Data source	Description
Database dumps	Quiz submits & grades, forum posts & comments
Clickstreams	Time-stamped video, quiz, wiki, forum clicks

Massive Open Online Courses (MOOCs)

- History: 2012, the Year of the MOOC
- Players: Coursera, edX, Udacity, ...
- Visibility: > 6 million learners on Coursera, early 2014

MOOC datasets

Data source	Description
Database dumps	Quiz submits & grades, forum posts & comments
Clickstreams	Time-stamped video, quiz, wiki, forum clicks
Other data	Survey, quiz responses and personal information

Survey data

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

Survey data

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

- Around 20% of learners responded

Survey data

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

- Around 20% of learners responded
- Age

Survey data

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

- Around 20% of learners responded
- Age
- Motivation

Survey data

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

- Around 20% of learners responded
- Age
- Motivation
- Occupation

Survey data

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

- Around 20% of learners responded
- Age
- Motivation
- Occupation
- Gender

Survey data

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

- Around 20% of learners responded
- Age
- Motivation
- Occupation
- Gender
- Region

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

- Around 20% of learners responded
- Age
- Motivation
- Occupation
- Gender
- Region
- Education background

Survey data	Description (participated / total learners)
Spring'13	9181 / 48401 (19%)
Fall'13	6319 / 35318 (18%)

- Around 20% of learners responded
- Age
- Motivation
- Occupation
- Gender
- Region
- Education background
- And more ...

Definition of CCDFs

Definition:

$$CCDF(x) = 1 - \sum_{j=1}^{x-1} p(j) = \sum_{j=x}^{\infty} p(j)$$

Definition of CCDFs

Definition:

$$CCDF(x) = 1 - \sum_{j=1}^{x-1} p(j) = \sum_{j=x}^{\infty} p(j)$$

The derivative:

Definition of CCDFs

Definition:

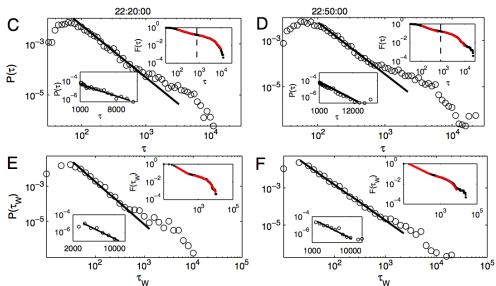
$$CCDF(x) = 1 - \sum_{j=1}^{x-1} p(j) = \sum_{j=x}^{\infty} p(j)$$

The derivative:

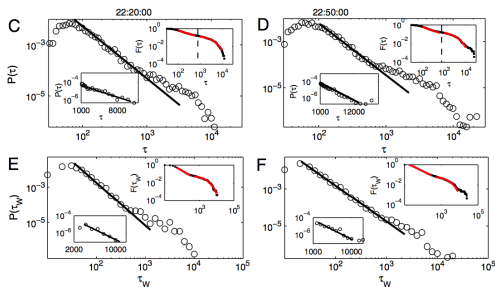
$$-\frac{d(CCDF(x))}{dx} = p(x)$$

CCDFs in Human Dynamics

Evidence of a bimodal distribution in human communications [Wu'10]



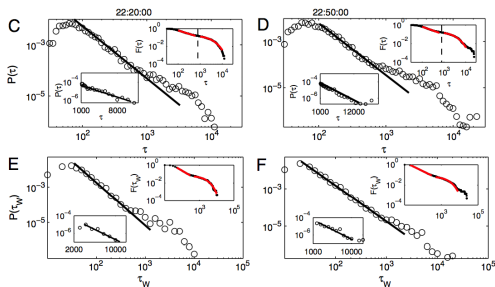
Evidence of a bimodal distribution in human communications [Wu'10]



- Interval distribution of Short Messages (SMs)

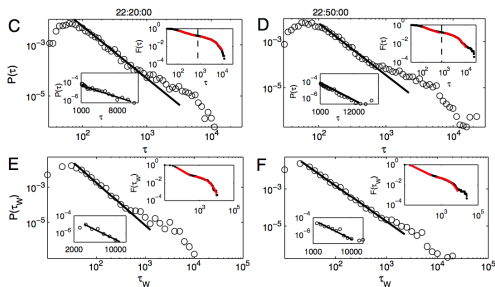
CCDFs in Human Dynamics

Evidence of a bimodal distribution in human communications [Wu'10]



- Interval distribution of Short Messages (SMs)
- Light-tail: short response time

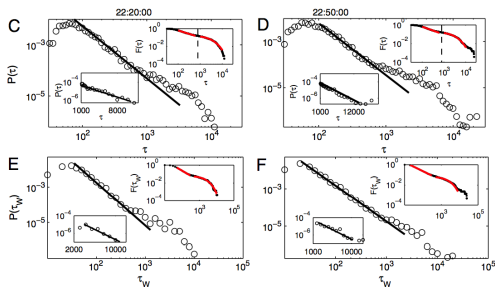
Evidence of a bimodal distribution in human communications [Wu'10]



- Interval distribution of Short Messages (SMs)
- Light-tail: short response time
- Heavy-tail: long response time

CCDFs in Human Dynamics

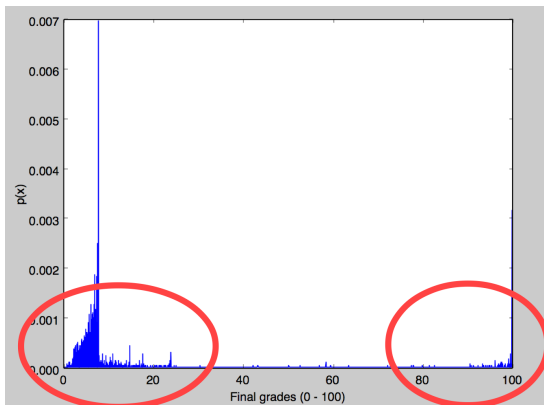
Evidence of a bimodal distribution in human communications [Wu'10]



- Interval distribution of Short Messages (SMs)
- Light-tail: short response time
- Heavy-tail: long response time
- Bimodal distribution shown in log-log scale

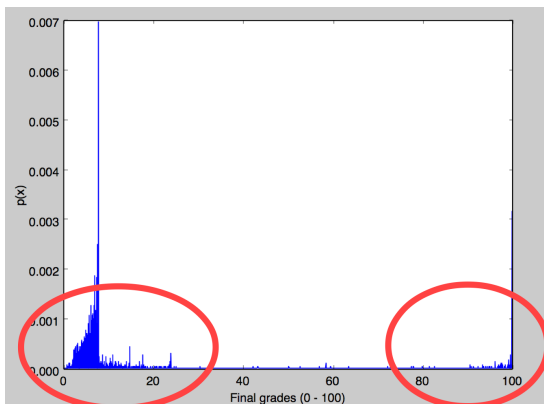
Final grades distribution

Normalize final grades distribution (Spring'13) is also bimodal!



Final grades distribution

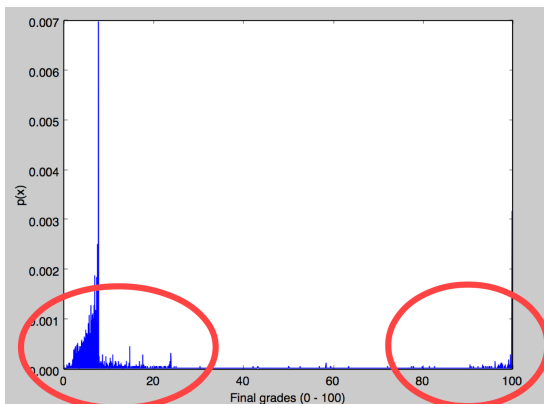
Normalize final grades distribution (Spring'13) is also bimodal!



- Many students on two extremes

Final grades distribution

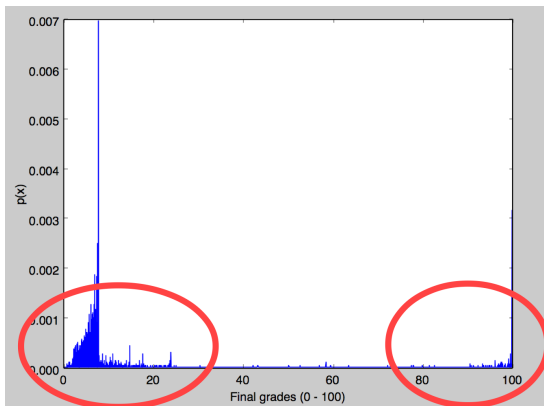
Normalize final grades distribution (Spring'13) is also bimodal!



- Many students on two extremes
- Histogram is noisy & difficult to model

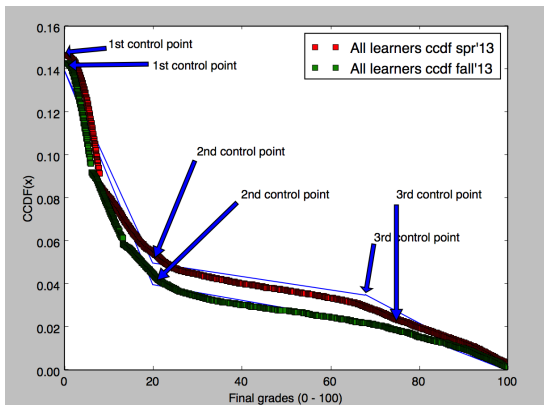
Final grades distribution

Normalize final grades distribution (Spring'13) is also bimodal!



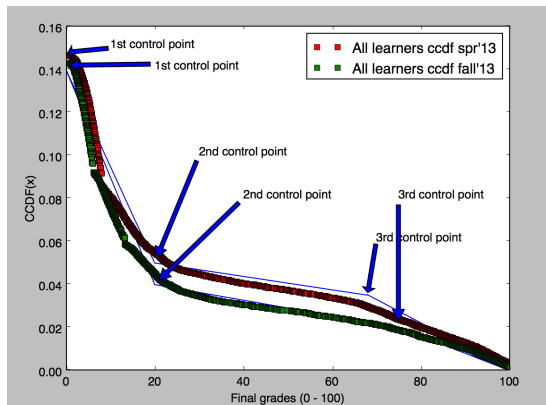
- Many students on two extremes
- Histogram is noisy & difficult to model
- Let's use CCDFs!

Final grades CCDFs



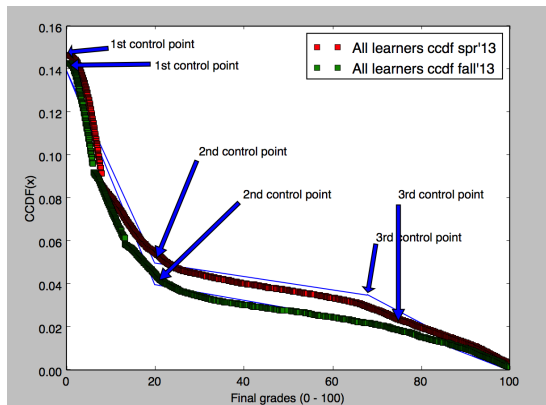
- 20 = 4 quizzes full marks

Final grades CCDFs



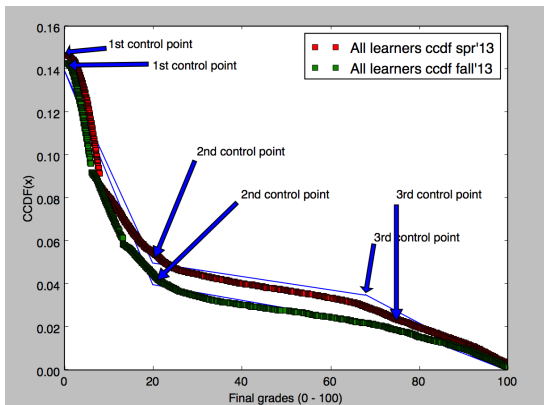
- 20 = 4 quizzes full marks
- > 40 = pass

Final grades CCDFs



- 20 = 4 quizzes full marks
- > 40 = pass
- > 90 = distinction

Final grades CCDFs



- 20 = 4 quizzes full marks
- > 40 = pass
- > 90 = distinction

What makes the CCDFs different?

Learner subpopulation analysis

Spring/Fall'13 semesters

- Age distribution

Learner subpopulation analysis

Spring/Fall'13 semesters

- Age distribution
- Motivation distribution

Learner subpopulation analysis

Spring/Fall'13 semesters

- Age distribution
- Motivation distribution
- Occupation distribution

Learner subpopulation analysis

Spring/Fall'13 semesters

- Age distribution
- Motivation distribution
- Occupation distribution
- Gender distribution

Spring/Fall'13 semesters

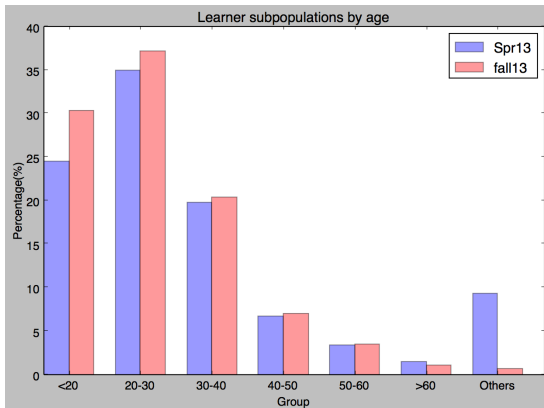
- Age distribution
- Motivation distribution
- Occupation distribution
- Gender distribution
- Region distribution

Spring/Fall'13 semesters

- Age distribution
- Motivation distribution
- Occupation distribution
- Gender distribution
- Region distribution
- Education background distribution

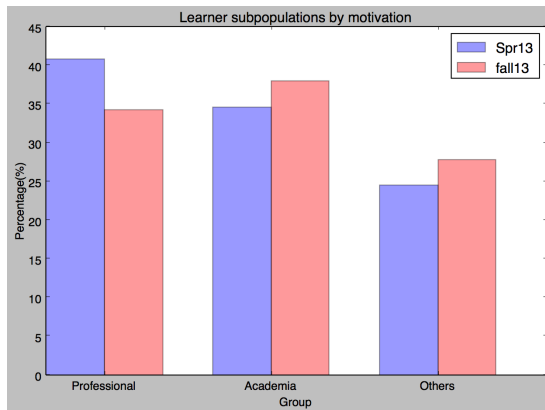
Learners by age

More young learners in the fall semester ...



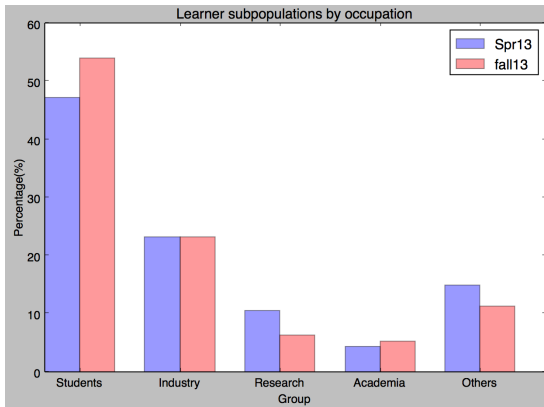
Learners by motivation

More academia learners in the fall semester ...



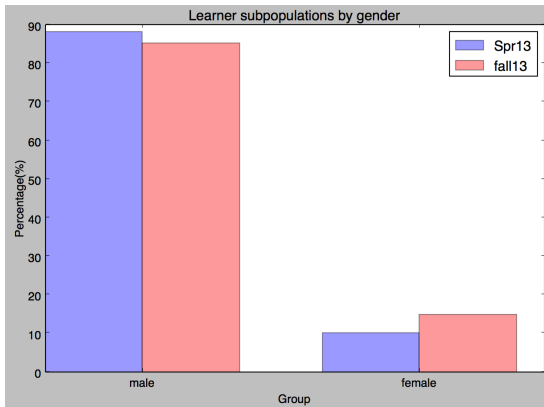
Learners by occupation

More student learners in the fall semester ...



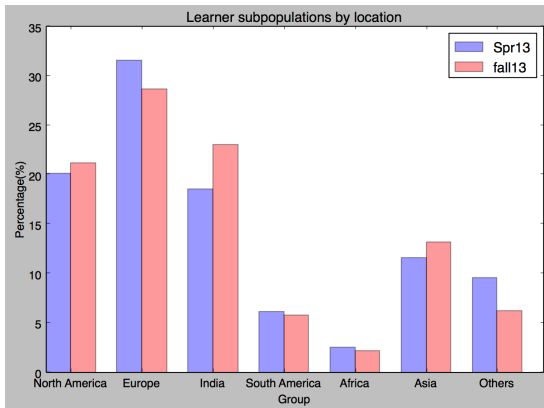
Learners by gender

More female learners in the fall semester ...



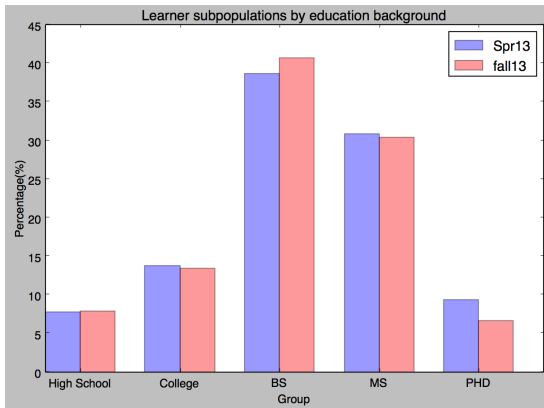
Learners by region

More Asian learners in the fall semester ...



Learners by education background

More college learners in the fall semester ...



Reaching out

- $\Delta(\text{student learners}) > 0$

Reaching out

- $\Delta(\text{student learners}) > 0$
- $\Delta(\text{European learners}) < 0$

Reaching out

- $\Delta(\text{student learners}) > 0$
- $\Delta(\text{European learners}) < 0$
- $\Delta(\text{female learners}) > 0$

Final grades CCDFs by learner subpopulations

- Age below vs. above 30

Final grades CCDFs by learner subpopulations

- Age below vs. above 30
- Professional vs. academia

Final grades CCDFs by learner subpopulations

- Age below vs. above 30
- Professional vs. academia
- Students vs. non-students

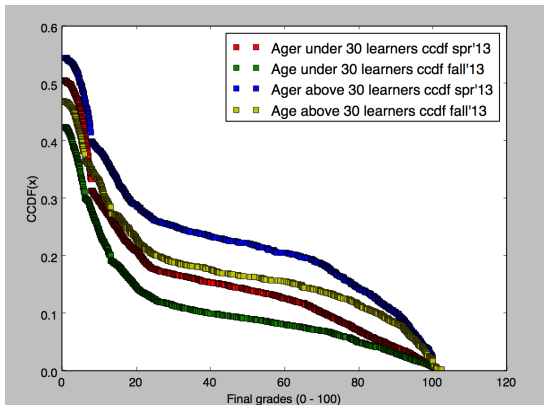
Final grades CCDFs by learner subpopulations

- Age below vs. above 30
- Professional vs. academia
- Students vs. non-students
- Female vs. male

Final grades CCDFs by learner subpopulations

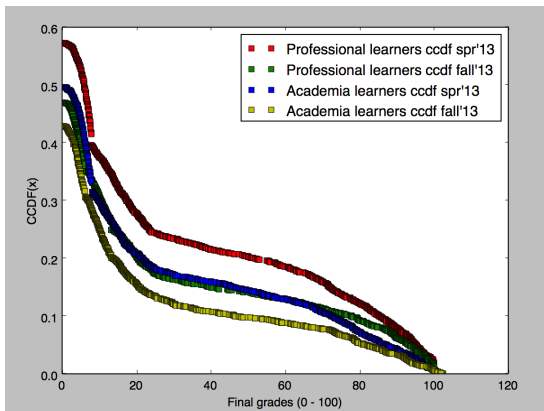
- Age below vs. above 30
- Professional vs. academia
- Students vs. non-students
- Female vs. male
- European vs. others

Older learners achieved better performance ...



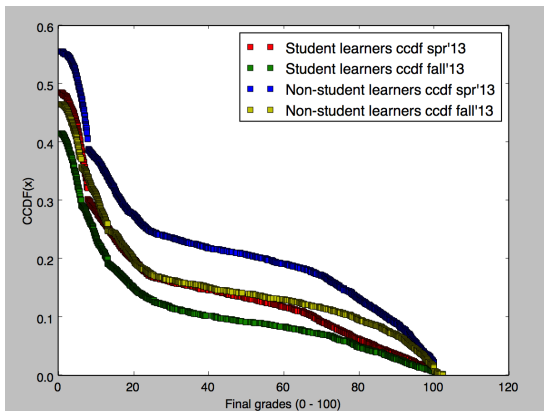
CCDFs by motivation

Professional learners achieved better performance ...



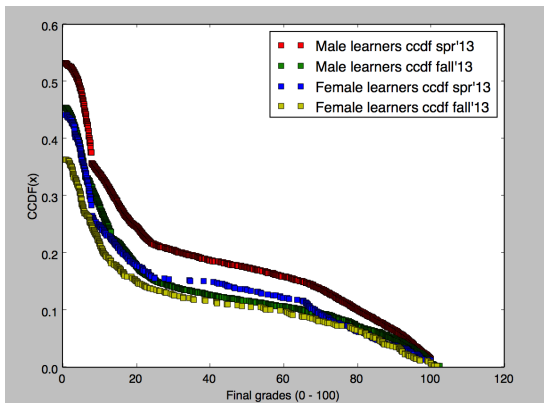
CCDFs by occupation

Non-student learners achieved better performance ...



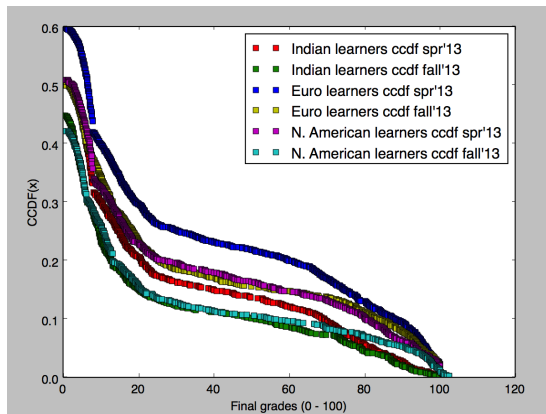
CCDFs by gender

No gender gap ...



CCDFs by region

European learners achieved better performance ...



Some results

- $CCDF_{\text{non-student learners}} > CCDF_{\text{student learners}}$

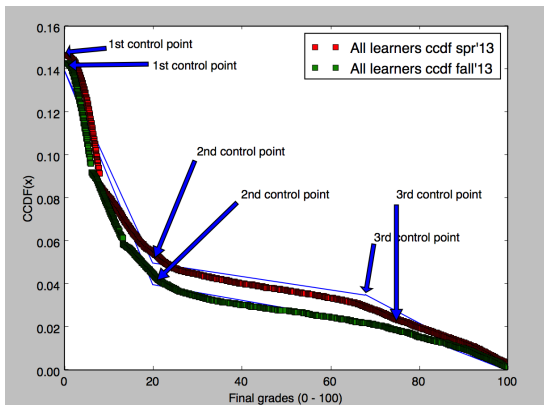
Some results

- $CCDF_{\text{non-student learners}} > CCDF_{\text{student learners}}$
- $CCDF_{\text{European learners}} > CCDF_{\text{Other region learners}}$

Some results

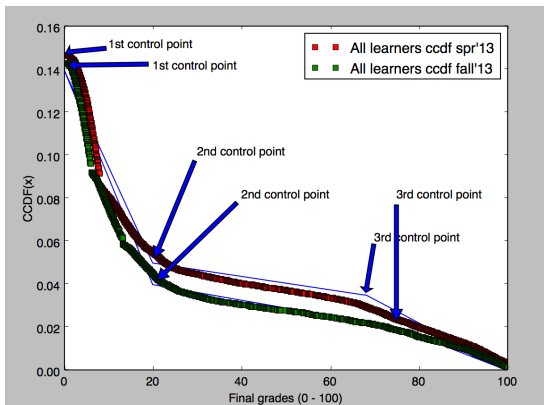
- $CCDF_{\text{non-student learners}} > CCDF_{\text{student learners}}$
- $CCDF_{\text{European learners}} > CCDF_{\text{Other region learners}}$
- $CCDF_{\text{female learners}} \sim CCDF_{\text{male learners}}$

Explaining final grades CCDFs



- $\Delta(\text{student learners}) > 0 \Rightarrow \Delta CCDF < 0$

Explaining final grades CCDFs



- $\Delta(\text{student learners}) > 0 \Rightarrow \Delta CCDF < 0$
- $\Delta(\text{European learners}) < 0 \Rightarrow \Delta CCDF < 0$

CCDFs in MOOCs

- A tool to visualize learner performance

CCDFs in MOOCs

- A tool to visualize learner performance
- A step towards learner performance modeling

CCDFs in MOOCs

- A tool to visualize learner performance
- A step towards learner performance modeling
- By learner subpopulations

Take-home messages

Grades distribution

- Bimodal

Take-home messages

Grades distribution

- Bimodal
- Non-student, aged, professional learners perform better

Grades distribution

- Bimodal
- Non-student, aged, professional learners perform better
- EU learners perform better

Take-home messages

Grades distribution

- Bimodal
- Non-student, aged, professional learners perform better
- EU learners perform better

Reaching out

- Should attract more student learners

Grades distribution

- Bimodal
- Non-student, aged, professional learners perform better
- EU learners perform better

Reaching out

- Should attract more student learners
- Should attract more Non-EU learners

Thank you, questions please.

tao.lee@epfl.ch