CCDFs in MOOCs

Tao C. Lee

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

June 2, 2014
Digital Signal Processing MOOC at Coursera

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

How to study learner dynamics?
Outline

- Complementary Cumulative Distribution Functions (CCDFs)
Outline

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

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

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

- 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
Background

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

<table>
<thead>
<tr>
<th>Data source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database dumps</td>
<td>Quiz submits &amp; grades, forum posts &amp; comments</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Data source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database dumps</td>
<td>Quiz submits &amp; grades, forum posts &amp; comments</td>
</tr>
<tr>
<td>Clickstreams</td>
<td>Time-stamped video, quiz, wiki, forum clicks</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Data source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database dumps</td>
<td>Quiz submits &amp; grades, forum posts &amp; comments</td>
</tr>
<tr>
<td>Clickstreams</td>
<td>Time-stamped video, quiz, wiki, forum clicks</td>
</tr>
<tr>
<td>Other data</td>
<td>Survey, quiz responses and personal information</td>
</tr>
<tr>
<td>Survey data</td>
<td>Description (participated / total learners)</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>Spring’13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>
## Survey data

<table>
<thead>
<tr>
<th>Survey data</th>
<th>Description (participated / total learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring'13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>

- Around 20% of learners responded
### Survey data

<table>
<thead>
<tr>
<th>Survey data</th>
<th>Description (participated / total learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring’13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>

- Around 20% of learners responded
- Age
Survey data

<table>
<thead>
<tr>
<th>Survey data</th>
<th>Description (participated / total learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring'13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>

- Around 20% of learners responded
- Age
- Motivation
## Survey data

<table>
<thead>
<tr>
<th>Survey data</th>
<th>Description (participated / total learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring'13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>

- Around 20% of learners responded
- Age
- Motivation
- Occupation
## Survey data

<table>
<thead>
<tr>
<th>Survey data</th>
<th>Description (participated / total learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring'13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Survey data</th>
<th>Description (participated / total learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring'13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Survey data</th>
<th>Description (participated / total learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring’13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Survey data</th>
<th>Description (participated / total learners)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring’13</td>
<td>9181 / 48401 (19%)</td>
</tr>
<tr>
<td>Fall’13</td>
<td>6319 / 35318 (18%)</td>
</tr>
</tbody>
</table>

- 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]
CCDFs in Human Dynamics

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

- Interval distribution of Short Messages (SMs)
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!
20 = 4 quizzes full marks
Final grades CCDFs

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

- $20 = 4$ quizzes full marks
- $> 40 = \text{pass}$
- $> 90 = \text{distinction}$

What makes the CCDFs different?
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
Learner subpopulation analysis

Spring/Fall’13 semesters

- Age distribution
- Motivation distribution
- Occupation distribution
- Gender distribution
- Region distribution
Learner subpopulation analysis

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 ...
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 ...
More Asian learners in the fall semester ...
More college learners in the fall semester ...
Remarks

Reaching out

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

Reaching out

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

Reaching out

- $\Delta$(student learners) $> 0$
- $\Delta$(European learners) $< 0$
- $\Delta$(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
CCDFs by age

Older learners achieved better performance ...
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 ...
Remarks

Some results

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

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

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
Key contributions

CCDFs in MOOCs

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

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
Take-home messages

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
Take-home messages

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