Circadian Cycles and Work Under Pressure: A Stochastic Process Model for E-learning Population Dynamics

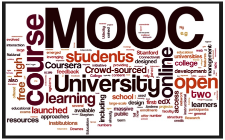
César Ojeda
Cesar.Ali.Ojeda.Marin@iais.fraunhofer.de

Fraunhofer Institute for
Intelligent Analysis and Information Systems IAIS

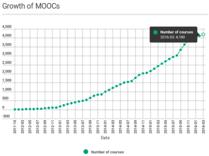
June 12, 2017

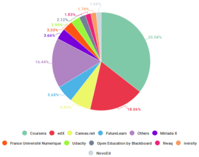
Fraunhofer
IAIS

Introduction



- A solution to the fast evolving work market: on-line education





- Growth, like that associated with Coursera, is *faster than Facebook*, specifically in terms of having a user growth rate greater than 2,000%.
- Growth from roughly 160,000 learners at one university in 2011 to 35,000,000 learners at 570 universities and twelve providers in 2015.
- 2016 Numbers: 58 million students, 700 Universities, 6850 courses.

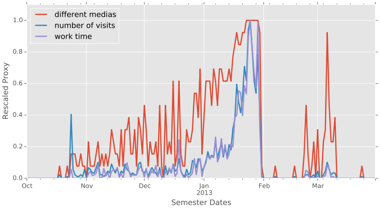
Problem Statement

To define dynamical models of user behavior.
Can we simulate how users visit a given platform in time?

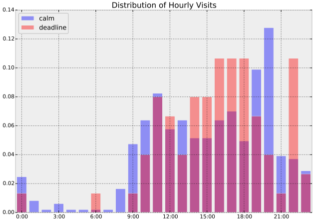
Use Case: A University Platform

In total, the system provides access to 1,147 different online lectures from 115 different courses. Our data set captures a total of 186,658 anonymized, time-stamped visits from 30,497 different IP addresses covering the four semesters in the time from April 2012 to March 2014.

Empirical Behavior



(a) Number of visits in time



(b) Hourly use of the system

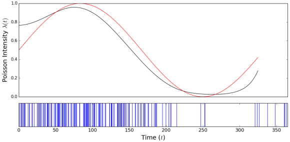
Model Requirements

1. circadian cycles of human activity
2. the stochastic nature of aggregated behaviors of many different users
3. the tendency of students to learn most of the material close to the deadline
4. the short term behavior of a session of study

Poisson Process

$$Pr\{\tilde{N}^A(t,t+\delta)=1\} = \lambda(t)\delta+o(\delta)$$
$$Pr\{\tilde{N}^A(t,t+\delta)=0\} = 1-\lambda(t)\delta+o(\delta)$$
$$Pr\{\tilde{N}^A(t,t+\delta)\geq 2\} = o(\delta),$$

Number of arrivals at time t ($N^A(t)$) to be equal to n_A for a Poisson process with intensity $x(t)$:



$$Pr\{N^A(t)=n_A\} = \frac{m(t)^{n_A}\exp -m(t)}{n_A!}$$

In order to define our model, we have to find the correct functional form for the intensity: $\lambda(t) = V_d(t)P_d(t)$

Procastination Reaction

How the students react to the deadline:

$$\frac{d}{dt}V_d(t) = \left(\frac{\alpha}{\beta}\right)\left(\frac{t}{\beta}\right)^{\alpha}\Big[T_e - V_d(t)\Big] \quad (1)$$

$$g(t) = \left(\frac{t}{\beta}\right)^{\alpha} \quad (2)$$

$$V_d(t) = \begin{cases} T_e - (T_e - V_0)e^{-(\frac{t}{\beta})^{\alpha}} & \text{if } t < t_d \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

and determine θ_{pr} as those parameters that minimizes the following error

$$D(t, \theta_{pr}) = \sum_{\tau=1}^{T_d} (V_d^d(\tau) - \hat{V}_d^d(\tau)[\theta_{pr}])^2 \quad (4)$$

Piece Wiese Model

For a course with many deadlines, we simply incorporate many peaks:

$$V_d(t) = T_a - (T_a - V_0)e^{-(\frac{t}{t_a - \delta})^{\alpha}} \tag{5}$$

The peak values are obtained directly from the data:

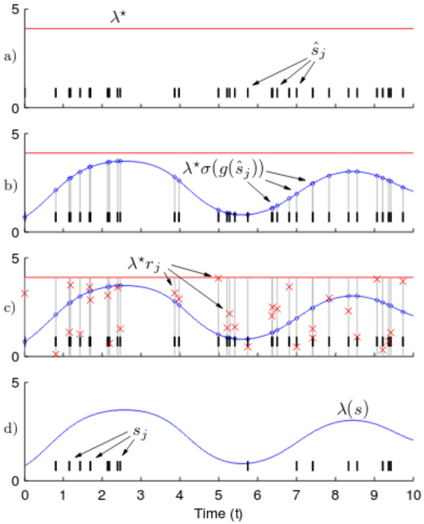
$$Gamma(T|\rho, \nu) = \frac{\rho^{\nu}}{\Gamma(\nu)} T^{\nu-1} e^{-\rho T} \tag{6}$$

Thinning Algorithm

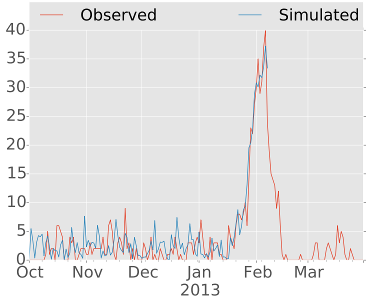
Algorithm 1 Generating a Cox Process

Require: Semester τ_s, Distribution Parameters θ_{pw}, E

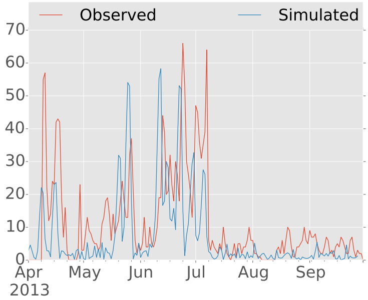
$\{\hat{t}_i\}_{i=1}^{E} \sim \text{Uniform}(\tau_s)$
$T_a \sim P(T|\theta)$
$U \sim \emptyset$
for $i \leftarrow 1, ..., E$ **do**
    $u_i \sim \text{Uniform}(0, 1)$
    $r_i \sim V(t_i|T_a)$
    if $u_i < r_i$ **then**
        $U \leftarrow U \cup \hat{t}_i$
    end if
end for
return U



Results

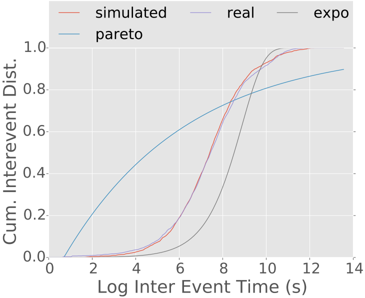


(a) Large Population Course. One deadline

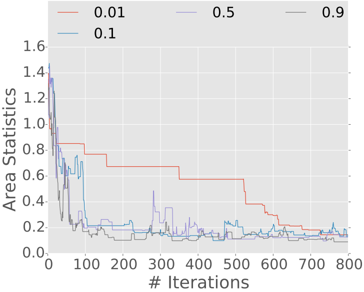


(b) Many Deadline Course. Piecewise model

Results



(a) cumulative distribution function



(b) training epochs

Training Results

Table: Area- and Kolmogorov-Smirnoff statistics for randomly selected courses in our data set.

Course Name	Area Statistic	KS Divergence
Computer Science	0.038	0.016
Databases	0.053	0.014
U.S.-American Literature	0.054	0.019
School Studies	0.13	0.051
Multivariable Calculus	0.075	0.026

Marks

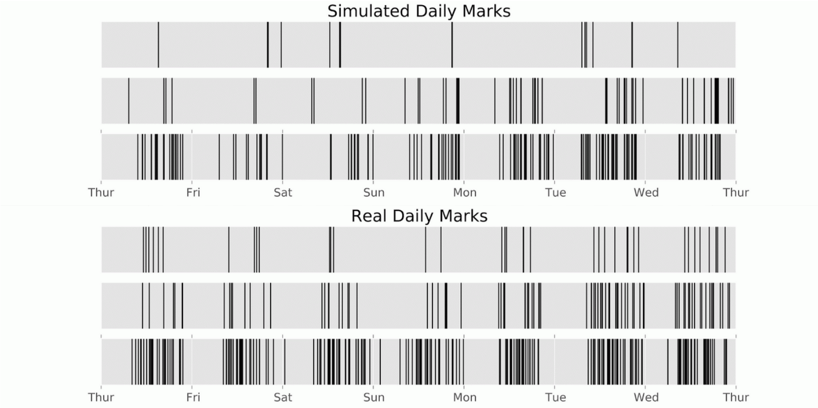


Figure: Real distribution of visits over a period three weeks before an examination deadline related to one of the courses in our data set (lower panel), and simulated point process of visits using the Cox process discussed in the text. Note that idle periods reflect reduced activities over night.

Conclusions

- We created models of human activity for an e-learning platform.
- We incorporated periodicity as well as patterns of procastination and reaction.
- We were able to reproduce the behavior of the inter event distribution.