Measuring Student Engagement During Collaboration

Peter F. Halpin

Joint work with Alina von Davier

NEW YORK UNIVERSITY

Outline

Part 1: An overview of temporal point processes[1]

Step 1 Defining and detecting dependence in event times

Step 2 Assessing temporal clustering

Step 3 Parametric modeling (with Hawkes processes)

[1] Halpin & von Davier (under review). Modeling collaboration using point processes.

Outline

Part 1: An overview of temporal point processes[1]

- Step 1 Defining and detecting dependence in event times
- Step 2 Assessing temporal clustering
- Step 3 Parametric modeling (with Hawkes processes)

Part 2: Part 2: Applying Hawkes processes to assessments involving collaboration[2]

- Collaborative performance assessments as temporally complex
- Interpretation of response intensity
- Review of results

[1] Halpin & von Davier (under review). Modeling collaboration using point processes.

[2] Halpin & von Davier 2013, Hao, & Lui (under review). Measuring student engagement during collaboration. Journal of Educational Measurement.

Outline

Part 1: An overview of temporal point processes[1]

Step 1 Defining and detecting dependence in event times

Step 2 Assessing temporal clustering

Step 3 Parametric modeling (with Hawkes processes)

Part 2: Part 2: Applying Hawkes processes to assessments involving collaboration[2]

- Collaborative performance assessments as temporally complex
- Interpretation of response intensity
- Review of results

Part 3: Data analysis

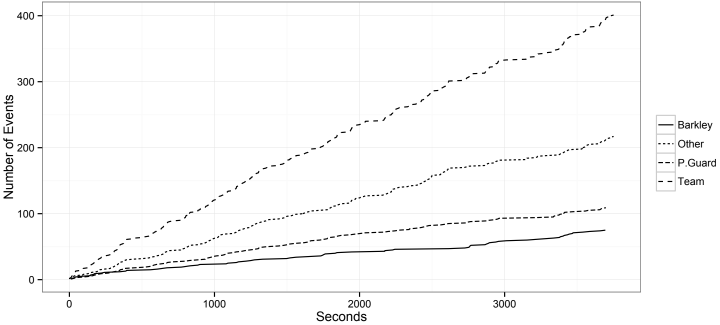
- Email example from Halpin & De Boeck, 2013[3]

[1] Halpin & von Davier (under review). Modeling collaboration using point processes.

[2] Halpin & von Davier 2013, Hao, & Lui (under review). Measuring student engagement during collaboration. Journal of Educational Measurement.

[3] Halpin & De Boeck (2013). Modelling dyadic interaction with Hawkes processes. Psychometrika, 78(4), 793?814.

Event times



Cumulative event count as a function of time for each player unit and the entire team

Step 1: Dependence among event times

- Let $X = (X_1, X_2, ..., X_N)$ and $Y = (Y_1, Y_2, ..., Y_M)$ denote two sequences of event times
  - e.g., the event times for Barkley begin $(14.00, 19.33, 44.08, \dots)$

- The mutual information of X and Y is

$$I_{XY} \equiv E_{X,Y} \, \ln\left(\frac{f(X,Y)}{f_X(X) f_Y(Y)}\right)$$

- where f is the joint probability density function, f_X and f_Y the marginals, E_U denotes expectation over the distribution of U

Some useful properties[4]

$$I_{XY} \equiv E_{X,Y} \ \ln\left(\frac{f(X,Y)}{f_X(X)f_Y(Y)}\right)$$

1) $I_{XY} = 0$ if and only if X and Y are statistically independent; otherwise $I_{XY} > 0$
2) I_{XY} makes mild assumptions about the kind of relationship between X and Y, and in particular the relationship can be non-linear
3) When $I_{XY} = 0$, it sample estimate has a known sampling distribution; this provides confidence bounds on the hypothesis of "no team interaction"

[4] Brillinger (2004). Some data analyses using mutual information. *Brazilian Journal of Probability and Statistics*, *18*, 163?182.

Some useful properties[4]

$$I_{XY} \equiv E_{X,Y} \ \ln\left(\frac{f(X,Y)}{f_X(X)f_Y(Y)}\right)$$

4) I_{XY} can be generalized to more than 2 sequences of event times via "superposition"
   - $Y = Y_1 + Y_2$ denotes "superposition" of Y_1 and Y_2,
   - Just the ordered sequence of all the event times in Y_1 and Y_2
   - Apply definition of I_{XY} as usual
   - Useful to let X denote a single team unit and Y denote all remaining team units

[4] Brillinger (2004). Some data analyses using mutual information. *Brazilian Journal of Probability and Statistics*, *18*, 163?182.

Some useful properties[4]

$$I_{XY} \equiv E_{X,Y} \ \ln\left(\frac{f(X,Y)}{f_X(X)f_Y(Y)}\right)$$

5) I_{XY} readily incorporates historical dependence

- $Y = X - a$, where $a > 0$ denotes a fixed constant subtracted from each element of X
- $I_{XY} = I_{XX-a}$ denotes the dependence of an event stream on its own past at lag a (like autocorrelation)
- I_{XY-a} denotes the dependence of X on the past of Y, at lag a (like crosscorrelation)
- Useful to treat as function of a

[4] Brillinger (2004). Some data analyses using mutual information. *Brazilian Journal of Probability and Statistics, 18,* 163?182.

"Rough" plug-in estimator

- Discretize the time interval $[0, T]$, e.g., into $k = 1, ..., K$ bins of size $\delta = T/K$.
- Recode X as K realizations of a random variable U, with $u_k = 1$ if an X-event happens in the interval $z_k = [\delta(k-1),\ \delta k)$ and $u_k = 0$ otherwise
- Similarly recode Y using the random variable W with realizations w_k

$$I_{XY} \approx I_{UW} = \sum_{i,j \in \{0,1\}} p_{ij}\ \ln\left(\frac{p_{ij}}{p_{i+}\ p_{+j}}\right)$$

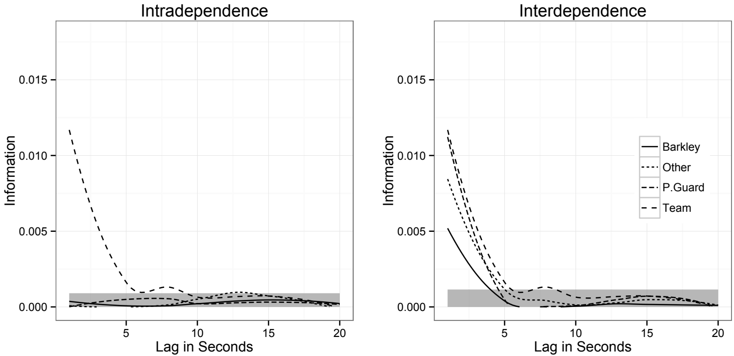
- $p_{ij} = \mathrm{Prob}\{U = i, W = j\}$
- p_{i+} and p_{+j} are the marginals
- When U and W are independent, the sampling distribution of I_{UW} is proportional to that of a chi-square statistic on one degree of freedom

"Rough" plug-in estimator

$$I_{XY} \approx I_{UW} = \sum_{i,j \in \{0,1\}} p_{ij} \, \ln\left(\frac{p_{ij}}{p_{i+}\, p_{+j}}\right)$$

- Main limitations of this approach to estimation
  1. Choosing number of intervals K – less is faster, more is better approximation
  2. Getting the p_{ij} – data are not i.i.d if $I_{XY} > 0$.
     - Alternative to using $\hat{p}_{ij} = \sum_k u_k \, w_k / K$?
- But useful for "data mining"

$\hat{I}_{UW}$ for basketball data

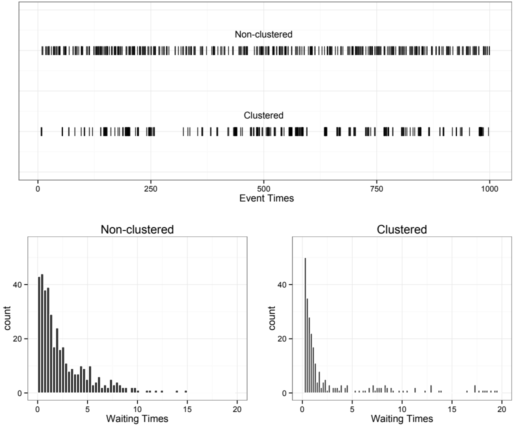


Sample intra- and interdependence functions for the entire team and the team units, as a function of time lag ($\delta = 1$ second)

Step 2: Characterizing the dependence

- What kind of dependence?
  - "Is the correlation positive or negative"
  - "Are the event times clustered or not"
- Important for choosing an appropriate model
- Empirical that clustering is a prevalent characteristic of human interaction
  - Barabási, 2005; Crane et al, 2008; Min, 2011; Matsubara, 2012
  - "Bursts"; "heavy-tailed waiting time distributions"; "power law distributions"

Clustered vs non-clustered event times



Event times and waiting times for clustered and non-clustered data (synthetic example, $N_i \approx 300$)

Clustering: how to test it

Clustering is defined as over-dispersion relative to the homogeneous Poisson process

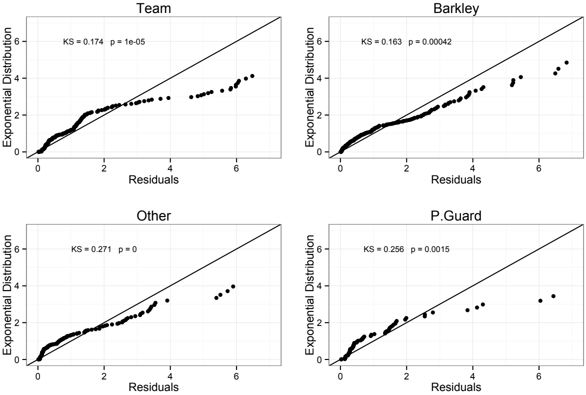
- "White noise"; no historical dependence

A general result for testing the goodness of fit between a theoretical point process and a given event stream:

- The time change theorem (Daley & Vera-Jones, 2003)
- The residuals of a correctly specified model are Poisson with rate $= 1$

QQ plots of the residualized waiting times against the exponential distribution

Testing for clustering: Basketball data



Quantile-quantile plots of residual waiting times from the Poisson process against the exponential distribution with rate of one. Insets give Kolmogorov-Smirnoff (KS) test and its two-sided p-value.

Step 3: Modeling

- All three team units exhibit clustering, which is a plausible source of cross-dependence

Step 3: Modeling

- All three team units exhibit clustering, which is a plausible source of cross-dependence

  A general model for clustered data is the Hawkes process (Hawkes, 1971; Hawkes and Oakes,
- 1974).
  - This is a nice model for dyadic data!

Hawkes processes

- Seen a lot of applications, general parametric models continue to pose problems
  - Ill-conditioning of and multiple roots in the likelihood function
  - Has lead to simplified assumptions about dependence and lag (exponential decay)

Hawkes processes

- Seen a lot of applications, general parametric models continue to pose problems
  - Ill-conditioning of and multiple roots in the likelihood function
  - Has lead to simplified assumptions about dependence and lag (exponential decay)
- Recent work on EM algorithm (e.,g, Veen & Schoenberg 2008) and Bayes (Rasmussen, 2012)
  - An EM algorithm for a multivariate Hawkes process with general decay functions (Halpin & De Boeck, 2013; Halpin 2013)

Hawkes process: intuitive explanation

- Each event stream is actually a mixture of 3 types of events
  - Spontaneous events: do not depend on previous events

Hawkes process: intuitive explanation

- Each event stream is actually a mixture of 3 types of events
  - Spontaneous events: do not depend on previous events
  - Self-response events: depend on their own past (auto-dependence)

Hawkes process: intuitive explanation

- Each event stream is actually a mixture of 3 types of events
  - Spontaneous events: do not depend on previous events
  - Self-response events: depend on their own past (auto-dependence)
  - Other-response events: depend on the past of another event stream (cross-dependence)

Hawkes process: intuitive explanation

- Each event stream is actually a mixture of 3 types of events
  - Spontaneous events: do not depend on previous events
  - Self-response events: depend on their own past (auto-dependence)
  - Other-response events: depend on the past of another event stream (cross-dependence)
- Hawkes process “guesses” which events are of which type based only on the timing of the events

Point Processes: formal representation

Point Processes: formal representation

Modern approaches to point processes are based on the conditional intensity (rate, hazard) function:

$$\lambda(t) = \lim_{\Delta \downarrow 0} \frac{E[N\{(t, t+\Delta]\} \mid H_t]}{\Delta}$$

- N is a counting measure defined on $\mathbb{R}+$ (finite, integer-valued, random measure)
- $E[N\{(a, b)\}]$ is the expected number of points falling in (a, b).
- H_t denotes the history of the process previous to time t (all events $t_1, t_2, \cdots, t_k$ up to and including time t)
- Extends immediately to the multivariate case – each margin of N then gives then corresponds to a specific type of event, e.g., of each person on a team

Point processes: formal representation

- When the process is orderly (only 1 event can happen at time):

$$\lambda(t) = \lim_{\Delta \downarrow 0} \frac{\mathrm{Prob}\left\{\text{event in } [t, t+\Delta) \mid H_t\right\}}{\Delta}$$

  - $\lambda(t)\Delta$ is approximately the Bernoulli probability of an event happening in $(t, t+\Delta]$, conditional on H_t – the *instantaneous probability* of an event at time t.
  - Mechanism for explaining historical dependence

Hawkes process: formal representation

- Using Hawkes process (in the bivariate case)

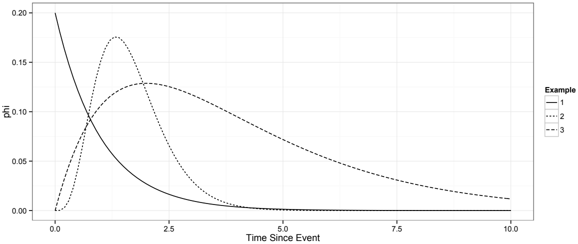
$$\lambda_i(t) = \mu_i + \sum_{t_{ik}<t} \phi_{ii}(t - t_{ik}) + \sum_{t_{jk}<t} \phi_{ij}(t - t_{jk})$$

for $i, j \in \{1, 2\}$ denoting the margins of N, $t_{ik} \in \{1, \ldots, N_i\}$ denoting event times and

$$\begin{array}{ll} \mu_i > 0; & \int_0^\infty \phi_{ij}(u) du \leq 1; \\ \phi_{ij}(u) \geq 0,\ u \geq 0; & \phi_{ij}(u) = 0,\ u < 0 \end{array}$$

- Can write $\phi_{ij}(u) = \alpha_{ij} \times f_{ij}(u)$, where f_{ij} is a density on $\mathbb{R}+$ and $0 < \alpha_{ij} < 1$.

Hawkes process: example response functions with gamma kernel



Hawkes process: estimation

Incomplete data log-likelihood (univariate case):

$$l(\theta) = \sum_i \left(\sum_{t \in \tau_i} \log(\lambda_i(t)) - \Lambda_i(T) \right)$$

where $\tau_i = \{t_{11} \dots t_{1N_i}\}$, $\Lambda(T) = \int_0^T \lambda(s) ds$, and $(0, T]$ is the observation period.

Problem: log of a weighted sum of densities, leads to ill-conditioning, local maxima

Hawkes process: estimation

Incomplete data log-likelihood (univariate case):

$$l(\theta) = \sum_i \left(\sum_{t \in \tau_i} \log(\lambda_i(t)) - \Lambda_i(T) \right)$$

where $\tau_i = \{t_{11} \ldots t_{1N_i}\}$, $\Lambda(T) = \int_0^T \lambda(s) ds$, and $(0, T]$ is the observation period.

Problem: log of a weighted sum of densities, leads to ill-conditioning, local maxima

Solution: Hawkes & Oakes (1974) – also Rasmussen (2012), Veen & Schoenberg(2008), Halpin & De Boeck (2013)

Hawkes process: estimation

Missing data:

$$z_i = \{z_{ioo}, z_{ii1}, \ldots, z_{iiN_i}, z_{ij1}, \ldots, z_{ijN_j}\}$$

$Z_i = (Z_{i1}, \ldots, Z_{iN_i})$ and write $Z_{ik} = z$ if $t_{ik} \in z$ for $z \in z_i$

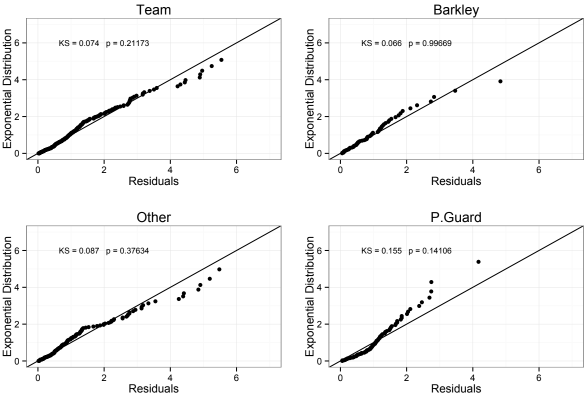
Q function:

$$Q_i(\theta \mid \theta^{(n)}) = \left(\sum_{z \in z_i} \left(\sum_{k=1}^{N_i} \log(\lambda_z(t_{ik})) \times \mathrm{Prob}(Z_{ik} = z \mid X_i, \theta_i^{(n)}) - \Lambda_z(T) \right) \right)$$

Posteriors:

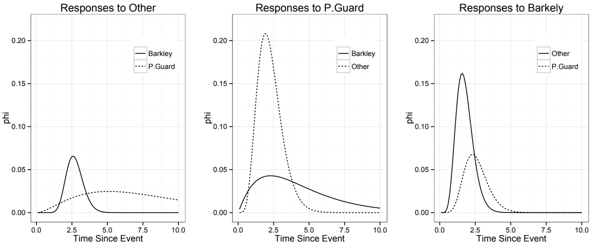
$$\mathrm{Prob}(Z_{ik} = z \mid X_i, \theta_i^{(n)}) = \frac{\lambda_z^{(n)}(t_{ik})}{\sum_{m \in z_i} \lambda_m^{(n)}(t_{ik})}$$

Hawkes process: goodness of fit



Quantile-quantile plots of residual waiting times from the Hawkes process against the exponential distribution with rate of one. Insets give Kolmogorov-Smirnoff (KS) test and its two-sided p-value

Hawkes process: estimated response functions



Estimated response functions, $\hat{\alpha}_{ij} \times f(u; \hat{\xi}_{ij})$, for the basketball data

Part 2: Applying Hawkes processes to assessments involving collaboration

- Let $\mathbf{X}_T = \{X_T, X_{T-1}, \ldots, X_1\}$ denote a sequence of random variables describing the actions of a student during the completion of a task
  - e.g., X_t could be a dichotomous variable indicating whether or not an individual has sent a chat message at time index t.
  - e.g., t might index the sequence of items on a conventional multiple choice assessment, in which case X_t would denote the item responses and $\mathbf{X}_T$ a response pattern
  - Call X_t task components

Simple tasks

- Define a (temporally) simple task: $p(\mathbf{X}_T) = \prod_t p(\mathbf{X}_t)$
  - All tasks obtained by permuting the order of the components have the same joint distribution

Simple tasks

- Define a (temporally) simple task: $p(\mathbf{X}_T) = \prod_t p(\mathbf{X}_t)$
  - All tasks obtained by permuting the order of the components have the same joint distribution

$$p(\mathbf{X}_T \mid \theta) = \prod_{t=1}^{T} p(X_t \mid \theta)$$

  - Local independence = conditionally simple
  - The process of collaboration is clearly not simple – the order of turns in an interaction cannot be rearranged like the items on a multiple choice test

Complex tasks

- Define a (temporally) complex task: $p(X_t \mid \mathbf{X}_{t-1}) \neq p(X_t)$ for some $t > 1$

- Can measure complexity with KL divergence
  - $p(\mathbf{X}_T) = \prod_{t=1}^{T} p(X_t \mid \mathbf{X}_{t-1})$
  - $q(\mathbf{X}_T) = \prod_{t=1}^{T} p(X_t)$

$$D[p(\mathbf{X}_T) \mid\mid q(\mathbf{X}_T)] \equiv E_p\left[\ln \frac{p(\mathbf{X}_T)}{q(\mathbf{X}_T)}\right]$$

Complex collaborative tasks

- Let $\mathbf{X} = \{\mathbf{X}_{1\,T_1}, \mathbf{X}_{2\,T_2}, \ldots, \mathbf{X}_{J\,T_J}\}$ denote a J-dimensional sequence of random variables,
- $\mathbf{X}_{j\,T_j} = \{X_{j\,T_j}, X_{j\,T_j-1}, \ldots, X_{j\,1}\}$ denoting the sequence of T_j actions of student $j = 1, \ldots J$

$$
\begin{aligned}
D[p(\mathbf{X}) \,||\, q(\mathbf{X})] &= E_p\left[\ln \frac{p(\mathbf{X})}{q(\mathbf{X})}\right] \\
&= E_p\left[\ln \frac{p(\mathbf{X})}{q(\mathbf{X})} + \ln \prod_{j=1}^{J} \frac{p(\mathbf{X}_{j\,T_j})}{p(\mathbf{X}_{j\,T_j})}\right] \\
&= E_p\left[\ln \frac{p(\mathbf{X})}{\prod_{j=1}^{J} p(\mathbf{X}_{j\,T_j})} + \ln \prod_{j=1}^{J} \frac{p(\mathbf{X}_{j\,T_j})}{q(\mathbf{X}_{j\,T_j})}\right] \\
&= D[p(\mathbf{X}) \,||\, \prod_{j=1}^{J} p(\mathbf{X}_{j\,T_j})] + \sum_{j=1}^{J} D[p(\mathbf{X}_{j\,T_j}) \,||\, q(\mathbf{X}_{j\,T_j})]
\end{aligned}
$$

Summarizing complex collaborative tasks

- Dependence among J collaborators during the completion of a complex task can be factored into $J + 1$ parts
- Only one of these parts, $D[p(\mathbf{X}) \mid\mid \prod_{j=1}^{J} p(\mathbf{X}_{j\,T_j})]$ depends on the joint distribution of $\mathbf{X}$
- It is equal to zero only if the actions of the J students are independent
- Describes the portion of overall task complexity that is due to interactions among students, or "inter-individual complexity"

So what?

Proposition 1[5]: Let $(0, T]$ denote the observation period and let $\mathbf{X}_j = \{t_{1j}, t_{2j}, \ldots t_{n_j\, j}\}$, $j = 1, \ldots, J$, denote the event times of the j-th margin of a point process. Assume that the data generating distribution of $\mathbf{X} = \{\mathbf{X}_1, ...\mathbf{X}_J\}$ is a Hawkes process as specified in equations 4 through 7. Then $D[p(\mathbf{X}) \;||\; \prod_j p(\mathbf{X}_j)] = 0$ if and only if the matrix $A = \{\alpha_{jk}\}$ is diagonal.

- If the Hawkes process is correctly specified, all of the inter-individual complexity in a collaborative task is "explained by" the response intensity parameters

[5] Halpin et al. (under review)

Interpretation of response intensity parameters

- Halpin et al. (under review) suggest the response intensity parameter as a measure of engagement of student j with k

$$\alpha_{jk} > \bar{n}_{jk}/n_k$$

- $\bar{n}_{jk} \equiv \sum_{i=1}^{n_k} E[N_{jk}(t_{ik}, T]]$ is the expected total number of responses made by student j to student k is (inferred from model)
- n_k is the number of actions of student k (observed)
- Lower bound is tight in practice; not necessary for computations, just useful for interpretation

Team-level engagement

- Aggregating to team (dyad) level

$$\alpha \equiv \frac{\alpha_{12} n_2 + \alpha_{21} n_1}{n_1 + n_2}$$

- Interpretation: the proportion of all group members' actions, $n_1 + n_2$, that were responded to by any other member during a collaboration
- See paper for more details, including initial results on SEs of α_{jk}

What are collaborative process data?

- Ideally a richly detailed recording of the sequence of actions taken by each team member during the completion of a task
  - ATC21S collaborative problem solving prototype items[6]
  - CPS frame[7]
  - OpenEdx CPSX[8]

[6] http://www.atc21s.org/uploads/3/7/0/0/37007163/pd_module_3_nonadmin.pdf

[7] In alpha at Computational Psychometrics lab at ETS

[8] github.com/ybergner/cpsx

What are collaborative process data?

- Ideally a richly detailed recording of the sequence of actions taken by each team member during the completion of a task
  - ATC21S collaborative problem solving prototype items[6]
  - CPS frame[7]
  - OpenEdx CPSX[8]
- Focus today: chat messages sent between online collaborators

[6] http://www.atc21s.org/uploads/3/7/0/0/37007163/pd_module_3_nonadmin.pdf

[7] In alpha at Computational Psychometrics lab at ETS

[8] github.com/ybergner/cpsx

Two perspectives on the analysis of chat / email / etc.

- Text-based analysis of strategy and sentiment
  - Howley, Mayfield, & Rosè, 2013; Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2015
- Time series analysis of sending times
  - Barabàsi, 2005; Ebel, Mielsch, & Bornholdt, 2002; Halpin & De Boeck, 2013

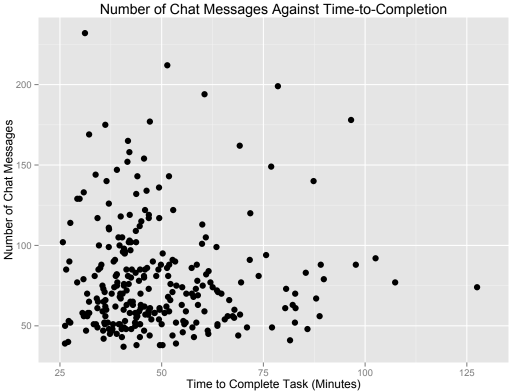
Example: Tetralogue

- A simulation-based science game with an embedded assessment recently developed at ETS (Hao, Liu, von Davier, & Kyllonen, 2015)
  1. Dyads work together to learn and make predictions about volcano activity
  2. At various points in the simulation, the students are asked to individually submit their responses to an assessment item without discussing the item.
  3. Following submission of responses from both students, they are invited to discuss the question and their answers.
  4. Lastly, they are given an opportunity to revise their responses to the item, with the final answers counting towards the team's score.

Example: Full sample

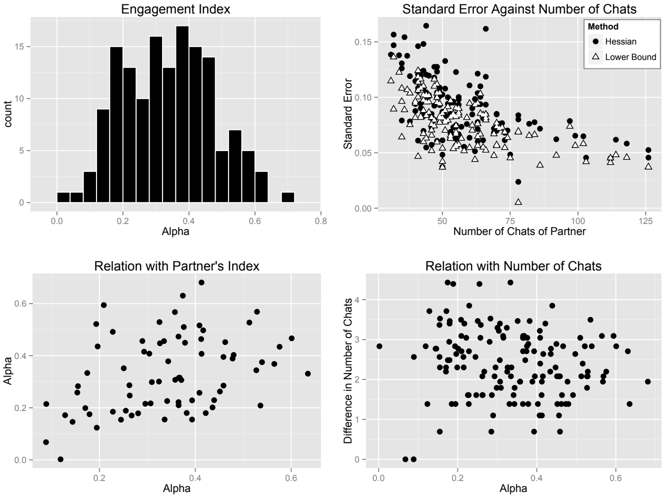
- 286 dyads solicited via AMT and randomly paired (based on arrival in queue)
- Median reported age was 31.5 years
- 52.5% reported that they were female
- 79.2% reported that they were White.
- Additionally, all participants were required to
  - Have an IP address located in the United States
  - Self-identify as speaking English as their primary language
  - Self-identify as having at least one year of college education

Example: Analysis sample



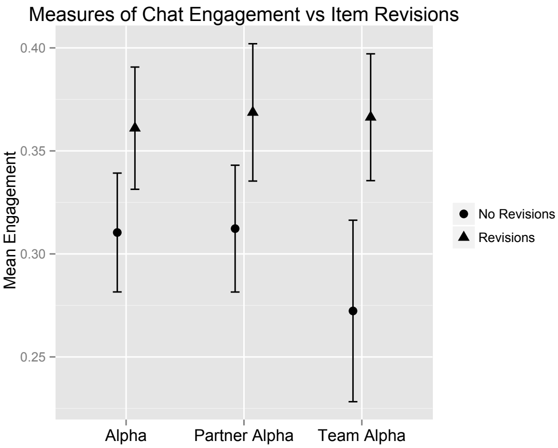
Omitted dyads whose total number of messages was less than 85, yielding a final sample size of 90 teams. While more chat messages would be preferable, the present example suffices to illustrate the methodology.

Example: Estimating chat engagement



Alpha denotes the estimated response intensities from Equation alpha2. Hessian denotes standard errors obtained via the Hessian of the log-likelihood. Lower Bound denotes Inequality (**??**). Difference in Number of Chats was scaled using the log of the absolute value of the difference.

Example: Relation with revision on embedded assessment



Comparison of mean levels of engagement indices for individuals who either did or did not revise at least one response after discussion with their partners. Alpha denotes the estimated response intensities from Equation alpha2; Partner's Alpha denotes the partner's response intensity; Team Alpha denotes the team-level index in Equation alpha4. For the latter, the data are reported for dyads, not individuals, and no revisions means that both individuals on the team made no revisions. Error bars are 95% confidence intervals on the means.

Example: Relation with revision on embedded assessment

Table 1: Summary of group differences.

Index	Group	Mean	SD	N	Hedges' g	r
Alpha	No Revisions	0.31	0.13	82	–	
Alpha	Revisions	0.36	0.10	66	0.40	.20
Partner's Alpha	No Revisions	0.31	0.14	82	–	
Partner's Alpha	Revisions	0.37	0.14	66	0.44	.21
Team Alpha	No Revisions	0.27	0.11	26	–	
Team Alpha	Revisions	0.37	0.13	48	0.84	.38

Note: Alpha denotes the estimated response intensities from Equation alpha2; Partner's Alpha denotes the engagement index of the individual's partner; Team Alpha denotes the team-level index in Equation alpha4. Hedges' g used the correction factor described by Hedges Hedges1981 and r denotes the point-biserial correlation.

Summary of collaborative processes

- Hawkes processes are a feasible model for process data obtained on collaborative tasks
- Resulting measures of chat engagement are meaningfully related to task performance
- Future modeling work
  - Random effects models for simultaneous estimation over multiple groups
  - Inclusion of model parameters describing task characteristics
  - Analytic expressions for standard errors of model parameters
  - Methods for improving optimization with relatively small numbers of events
  - Integration with text-based analyses (e.g., using marks / time-varying covariates)

What's next

- Integration of task design, outcomes, processes, ... and theory!!

Contact: peter.halpin@nyu.edu

Collaborators: Yoav Bergner, ETS; Jacqueline Gutman, NYU

Support: This research was funded by a postdoctoral fellowship from the Spencer Foundation and an Education Technology grant from NYU Steinhardt.