An Autocorrelated Mixture Model for Sequences of Response Time Data

Peter F. Craigmile

Department of Statistics, OSU

Quantitative Studies in Consumer Behavior Seminar

Tuesday 9 May 2006

This is joint work with Mario Peruggia and Trisha Van Zandt

`http://www.stat.ohio-state.edu/~pfc/`

Introduction

- Response time (RT) is a ubiquitous measure of human performance.
- Has been used as a window on psychological processes for almost two centuries.
- Forms the foundation for most work in cognitive psychology.
  - Used to formulate theories of brain function and cognitive processing; to evaluate training regimens, user interface design, vehicle operation, and task design and to evaluate medical conditions, especially schizophrenia, learning disorders, and other psychological disorders.
- Efforts to model RT, or, more fundamentally, the processes that give rise to certain patterns of RT, have been the focus of attention for much of experimental psychology for the last 50 years (see Luce (1986) for a comprehensive review).

IID approaches to modeling RT

- Most common approach:
  - Derive a hypothesis about how a summary statistic changes in an experiment.

    e.g., mean RT should decrease with increased practice (Heathcote et al. 2000).
- Less common to formulate distributional hypotheses about the entire sample of RTs.
  - For example, if response activation is controlled by neural events occurring according to a Poisson process, then the distribution of RTs should be gamma.
  - Common marginal distributions used to describe RTs include: gamma, Weibull, inverse normal, and the "ex-Gaussian".
- Other theoretical approaches are based on the minima of first passage time distributions (Ratcliff and Smith 2004).

Non-IID approaches

- Other approaches embrace the notion that sequences of RTs are neither independent nor identically distributed.
- Work investigating sequential effects, priming, inhibition of return, task-switching, and so forth is directed toward understanding how information about previous stimuli and responses influences processing of later stimuli and facilitates or inhibits responses to those stimuli (Jones et al. 2006; Meeter and Olivers 2006; Stewart et al. 2005).
- Gilden and others emphasize dependence by modeling long-range effects across sequences of trials, or $1/f$ (pink) noise.

  (Gilden 1997; Gilden 2001; Pressing and Jolley-Rogers 1997; Thornton and Gilden 2005; Van Orden et al. 2003; Wagenmakers et al. 2004; Farrell et al. 2005)

Extreme observations

- Little agreement on how to handle extreme observations.
- Since RT data are positively skewed, some tail values are to be expected from the process that generated the data.
- Other extreme observations in the right tail are due to other factors.
  - Examples: unscheduled rests in the middle of a block of trials, lapses in attention and intrusions of the subject's physiological system (sneezing, itching, etc.).
- Extreme observations in the lower tail are harder to detect:
  - Example: fast RTs are commonly attributed to rapid guessing or error (twitches).
- Modeling such data of are interest in both psychology and statistics (Barnett and Lewis 1994; Ratcliff 1993; Ulrich and Miller 1994; Van Selst and Jolicoeur 1994; Belin and Rubin 1995).

Bayesian approaches

- Attractive because allow the flexibility to specify RT distributions that depend on varying experimental conditions, typically via hierarchical regression models for parameters parameterizing the experimental conditions.
- Examples:
  - Rouder, Sun, Speckman, Lu, and Zhou (2003) assume that the observed RTs are conditionally independent, ignoring sequential dependencies.
  - Peruggia, Van Zandt, and Chen (2002) attempt to capture sequential dependencies through an autoregressive structure for the log of the scale parameter of the RT distribution.

Our aims

- Suitably **detrended** sequence of log RT sequences can be described with a **mixture** likelihood.
  - Detrending is crucial, because it removes smooth changes in RT levels due to learning effects, fatigue, etc., distinguishing them from more localized dependencies.
- The mixture model incorporates the following:
  1. A Gaussian autoregressive component captures local dependencies.
  2. Two exponential components model the two tails of the marginal log RT distribution.
- Provides a framework for performing RT analysis that does not compromise model realism and autocorrelation structure.
- Can be used to build **sensible** Bayes hierarchical models which are not based entirely on consideration of convenience (namely, conditional conjugacy).

Example response time data

- RT data come from Wagenmakers, Farrell, and Ratcliff (2004)

  `http://users.fmg.uva.nl/ewagenmakers/fnoise/noisedat.html`.
- Their motivation: To study **autocorrelation** structure across long sequences of trials.
- Experimental setup:
  - Six subjects.
  - A fixed set of stimuli (the numerals 1 to 9, except 5).
  - The numerals were presented on a monitor.
  - Each of three tasks required subject to make a keypress response.

The three tasks

- Three different tasks:
  1. Simple detection: respond as soon as any stimulus was presented.
  2. Choice response: respond with one key when the numeral was greater than 5 and another key when the numeral was less than 5.
  3. Time estimation task: respond with a keypress one second after the numeral was presented, regardless of the numeral.
- Experiment designed so stimulus sequences were exactly the same regardless of which task a subject was to perform.

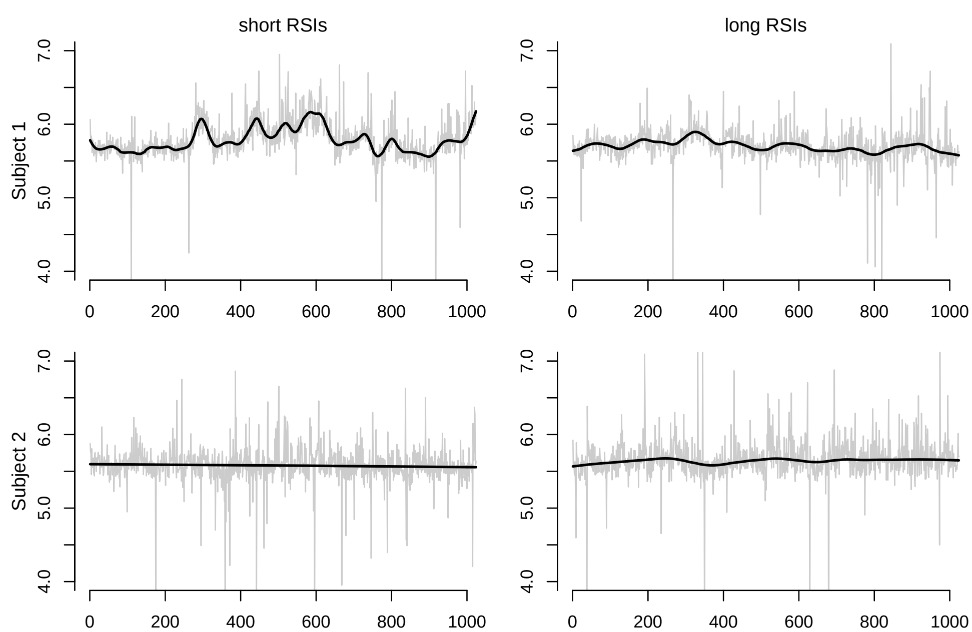
Response time versus response-stimulus interval

- An important manipulation in the study:
  - **The response-stimulus interval** or RSI is the interval between a response and the presentation of the subsequent stimulus.
  - To prevent anticipatory responses (extremely fast RTs), RSIs were uniformly distributed between 550ms and 950ms in **"short"** conditions, and between 1150ms and 1550ms in **"long"** conditions.
- Each subject provided 6 RT sequences: one sequence for each task in each RSI condition.
  - For each sequence there were 1048 successive trials collected without a break.
  - First 24 'practice' responses were discarded.
- For the moment, we focus only on the **simple detection** experiment.

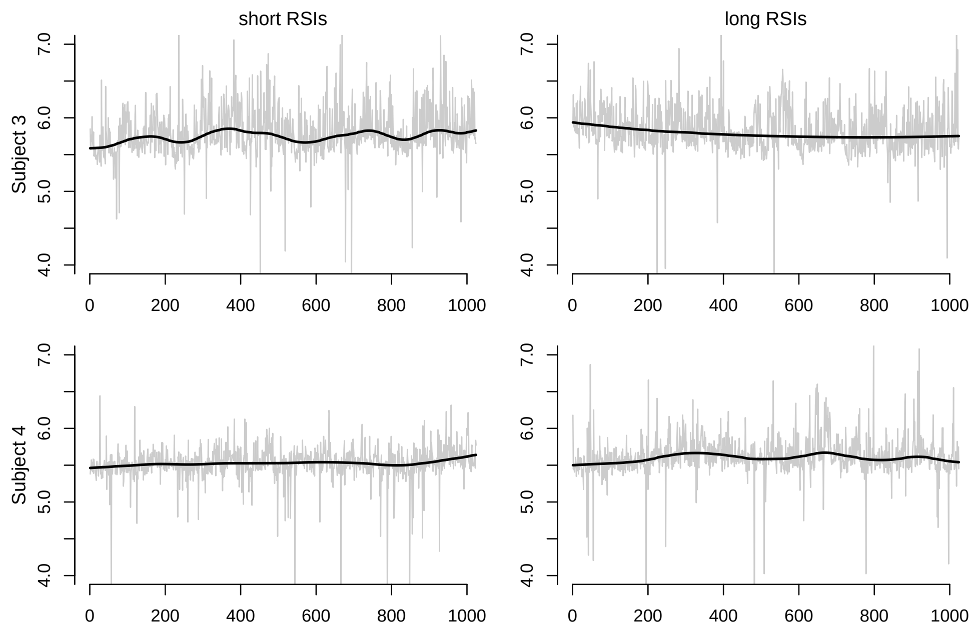
Exploratory RT analysis

- Remember that here and elsewhere, the "time" series is really a "trial" series.
- The abscissa of the plots is the trial number (from 1 to 1024) and the ordinate is the log RT for each trial.
  - The log transformation allow us to examine the RT sequences for potential extreme observations in **both** tails of the distribution and to assess the extent of serial dependence in each sequence.
- Evidence of trends (smooth changes of the level of the process over time), and both short- and long-range serial dependence.
  - Trends occur for a variety of reasons, including learning (negative trends) or fatigue (positive trends). Fluctuations in attentional state or task readiness may result in a fixed shift in the mean of the process.

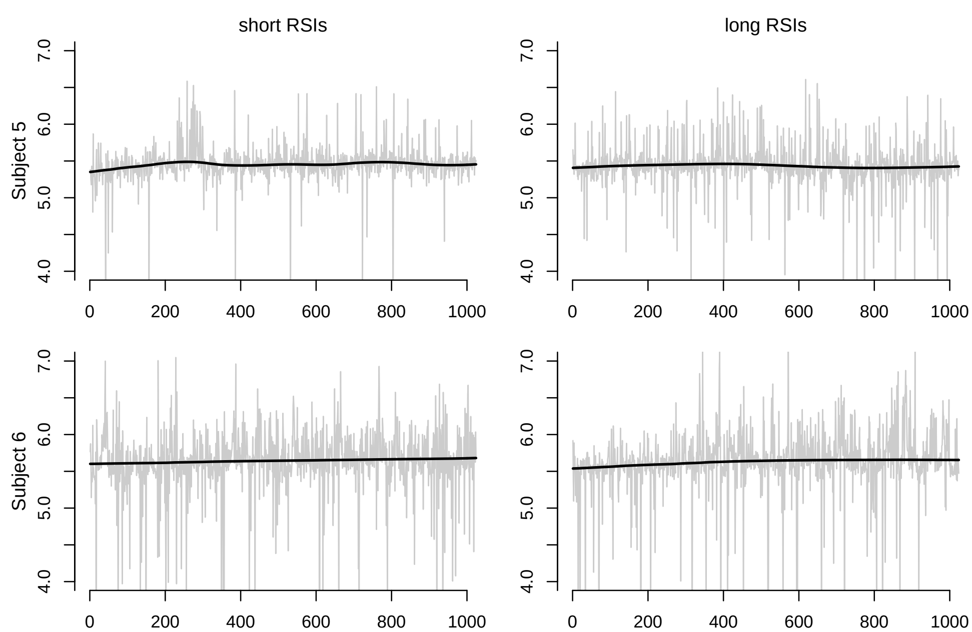
Subjects 1 and 2



Subjects 3 and 4



Subjects 5 and 6



Detrending the log RT series

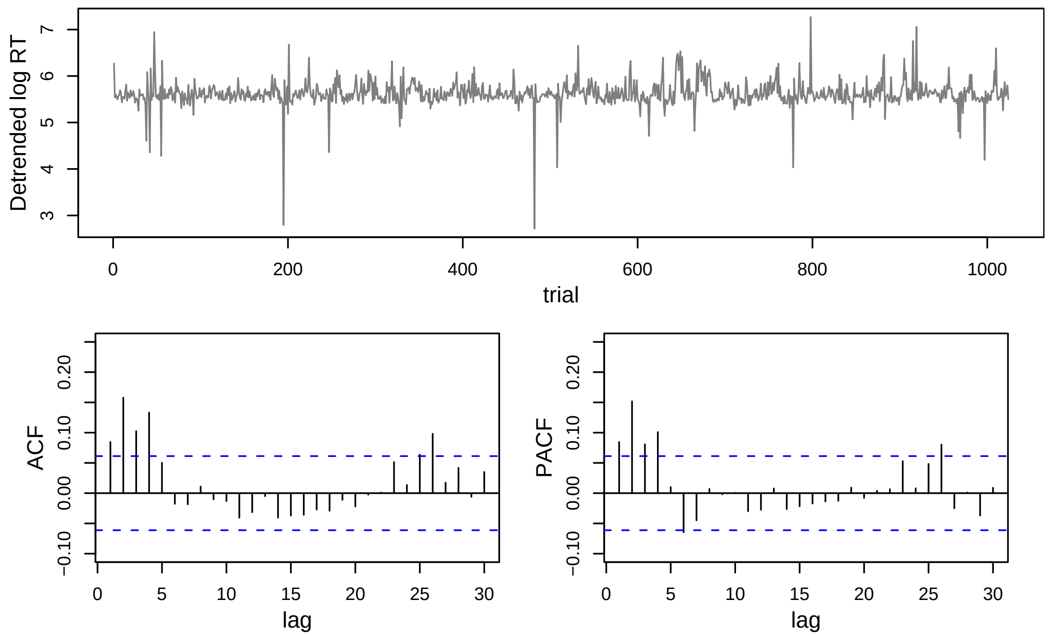
- Remember our data is non-Gaussian and dependent.
- Without special care, standard methods to estimate trend will **oversmooth**.
- Our approach:

1. We first converted each log RT sequence to normal scores.
2. We then smoothed the normal score data using the method of Wang (1998) as implemented in the ASSIST R package.
   - Cubic smoothing splines are fit to Gaussian data and the dependence structure is modeled with an AR(1) process.
3. Using interpolation we converted the smoothed trend back to the original log RT scale.

Describing the trends

- Subject 1's trends are erratic for both conditions.
- Other subjects have more steady trends.
- Most subjects show (oscillatory) increasing trend in both conditions, suggesting some tiring by the end of the trial sequence.
- Only subject 3's responses in the long RSI condition steadily and markedly decrease.

Motivating example: subject 4



Dependence and extreme observations in the detrended log RT series

- Serial dependencies appear as "clustering" of data points in the series.
  - Short-range effects are quantified by autocorrelation and partial autocorrelations over short lags. Can be caused by carry-over effects from immediately preceding trials, either due to particular stimuli or to particular responses.
  - Long-range dependencies and/or unmodeled trends are characterized by **slowly** decaying autocorrelation and partial autocorrelation sequences. May reflect the formation of mental representation.
- There are extremely long (slow) and extremely short (fast) RTs.
  - These observations would be considered outliers based on a model that specifies a Gaussian marginal distribution for the data.

Problems with interpreting dependence

- Given the extreme observations observed, some caution should be taken in interpreting the autocorrelation and partial correlation function.
- For example:
  - Large additive outliers in an ARMA process will damp down the autocorrelation estimates (Peña et al. 2001).
  - Similarly, for our mixture model, the autocorrelation function (ACF) for the process will be damped down by the presence of the independent additive exponential components.
- Also, because the data are not Gaussian, the variability and bias of the sample ACF increases as the extreme observations become more prominent.
- Trimming (removing) or winsorizing (truncating) these observations is not a satisfactory solution, because this operation can also affect the estimation of the ACF.

A Monte Carlo experiment

- Let $\{X_t\}$ be the AR(1) process defined by

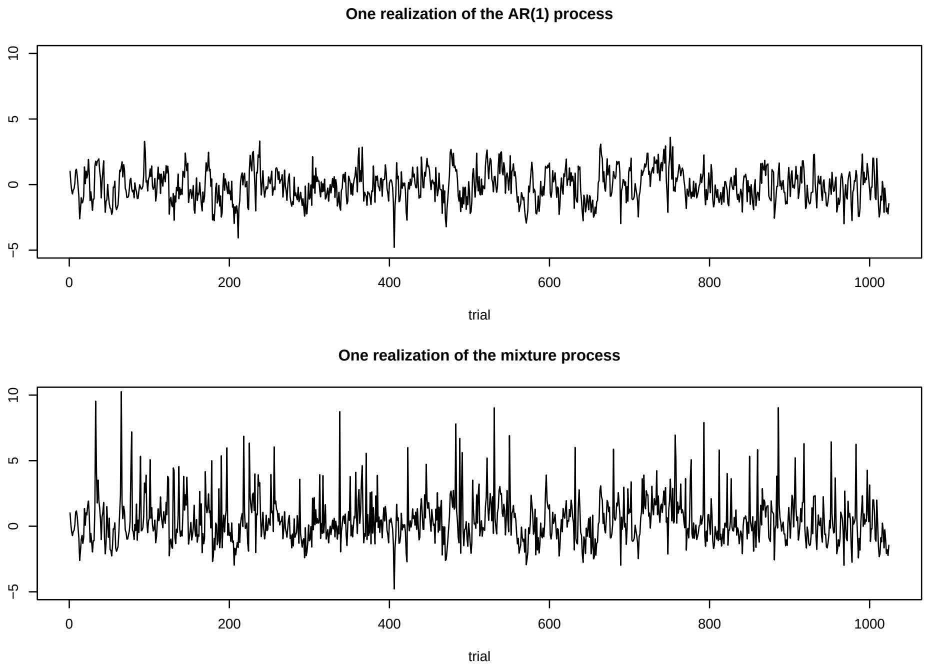
$$X_t = 0.6X_{t-1} + U_t.$$

  where $\{U_t\}$ is an *IID* sequence of N(0, 1) RVs.
- Define the mixture process, $\{W_t\}$, by

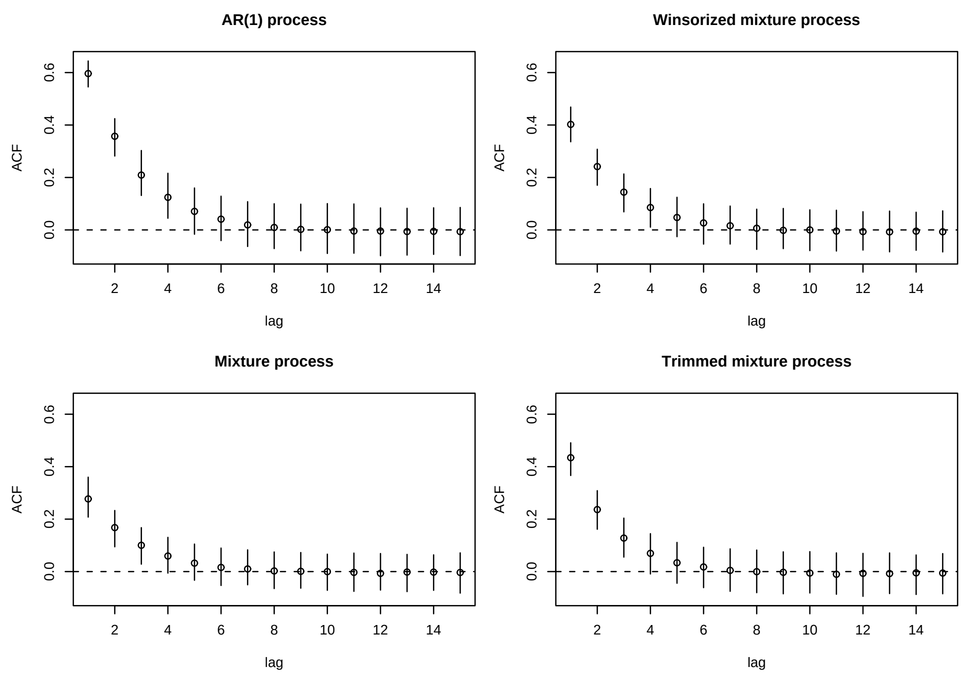
$$W_t = X_t + \delta_t Y_t,$$

  where $\{\delta_t\}$ is an *IID* sequence of Bernoulli(0.2) RVs and $\{Y_t\}$ is an *IID* sequence of exponential RVs with mean two.
- Assume $\{X_t\}$, $\{\delta_t\}$, and $\{Y_t\}$ are mutually independent.
- We generate 500 realizations of $\{X_t\}$ and $\{W_t\}$.

Example realizations



Estimating the autocorrelation function



Conclusions of the experiment

- The ACF is damped down relative to the ACF for the AR(1) process.
- Two methods are commonly used to improve the estimate of the ACF.
  - Winsorizing: truncate values that exceed some threshold before calculating the ACF.
  - Trimming: remove those trials that exceed a given threshold and then calculate the ACF of this reduced series. This disturbs stationarity of the process, making interpretation of the ACF difficult.
- While trimming and winsorizing both slightly reduce the damping effect of the additive extreme values upon the sample ACF, further experiments showed that changing the percentile cutoff value did not allow us to ever approach the ACF for the AR(1) process.
- We conclude that any method of analysis should model both the dependence and extreme observations jointly, rather than trimming or winsorizing.

The glue for our model: the ex-Gaussian distribution

- Suppose $X \sim N(\mu, \sigma^2)$, independent of $Y \sim \text{Exp}(\lambda)$ (with $E(Y) = 1/\lambda$).
  - Then $W = X + Y$ has a *ex-Gaussian* distribution with parameters μ, σ^2 and λ.
- The original motivation for the ex-Gaussian distribution is due to Hohle (1965).
  - He reasoned that the Gaussian component models the peripheral processing time, and the exponential component models the cognitive processing time.
  - This is no longer an accepted theory in the RT literature.
- However, the ex-Gaussian model sees wide use as a description of RT data because it is very flexible and can adequately describe extremely large observations.

  (Andrews and Heathcote 2001; Gottlob 2004; Heathcote et al. 1991; Penner-Wilger et al. 2002; Ratcliff and Murdock 1976; Reber et al. 1997).

Remarks on the ex-Gaussian model

- The density for W is a convolution of an exponential and a Gaussian density.
- So think of W as arising either from:

  1. a shifted exponential distribution with a normally distributed shift, or
  2. a normal distribution with an exponentially distributed mean.

- The distribution of W puts positive mass on the negative values – not a model for positively valued data, unless data is transformed or weight given to negative values can be considered negligible.
- We will extend the ex-Gaussian model to allow for extreme observations in the negative direction as well.

Our Bayesian mixture model

- Let $\{W_t\}_{t=1}^{1024}$ be the detrended log RT values for each subject with a short or long RSI.
- The prototype model for W_t is

$$W_t = \begin{cases} X_t & \text{with prob. } p_X \\ X_t + Y_t & \text{with prob. } p_Y \\ X_t - Z_t & \text{with prob. } p_Z. \end{cases}$$

- Here $\{X_t\}$ is a hidden AR(1) process, that models the serial dependence in the RTs.
- The possible occurrence of extreme observations is modeled by the two independent sequences of exponentially-distributed RVs $\{Y_t\}$ and $\{Z_t\}$, with means $1/\lambda_Y$ and $1/\lambda_Z$, respectively.

The hidden time series model

- Define $\{X_t\}$ by

$$X_1 - \mu = U_1,$$
$$X_t - \mu = \phi(X_{t-1} - \mu) + U_t, \;\; t = 2, \ldots, 1024,$$

where U_1 is a $\mathrm{N}(0,\, \sigma_1^2)$ RV and $\{U_t\}_{t=2}^{1024}$ is an *IID* sequence of $\mathrm{N}(0,\, \sigma^2)$ RVs.

- $\{X_t\}$ conditional on its parameters, is **not** a stationary time series model.

Prior specifications

- Assume all priors are independent, and any constants are known.
- Priors for $\{X_t\}$:

$$\phi \sim \mathrm{N}(\mu_\phi, \sigma_\phi^2), \quad \mu \sim \mathrm{N}(\eta, \sigma_\mu^2),$$
$$1/\sigma_1^2 \sim \mathrm{Gamma}(\alpha_{\sigma_1^2}, \beta_{\sigma_1^2}), \quad 1/\sigma^2 \sim \mathrm{Gamma}(\alpha_{\sigma^2}, \beta_{\sigma^2}),$$

(Parametrized so $\mathrm{Gamma}(\alpha, \beta)$ RV has mean α/β).

- Priors for $\{Y_t\}$ and $\{Z_t\}$:

$$\lambda_Y \sim \mathrm{Gamma}(\alpha_Y, \beta_Y), \lambda_Z \sim \mathrm{Gamma}(\alpha_Z, \beta_Z).$$

- Prior for the mixture probabilities:

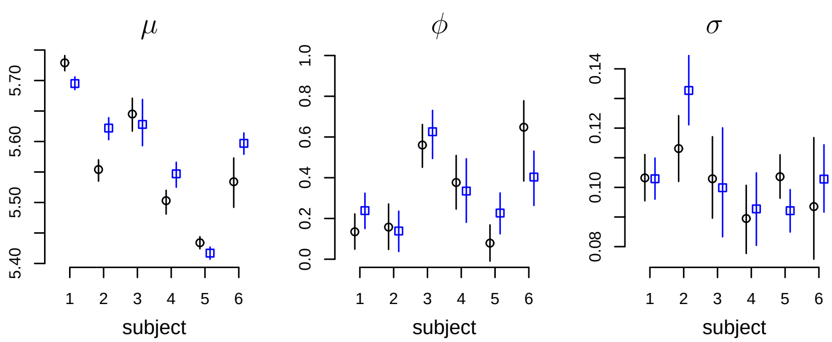
$$\mathbf{p} = (p_X, p_Y, p_Z) \sim \mathrm{Dirichlet}(\gamma_X, \gamma_Y, \gamma_Z).$$

(Dirichlet distribution has density proportional to $p_X^{\gamma_X - 1} p_Y^{\gamma_Y - 1} p_Z^{\gamma_Z - 1}$ over the unit simplex).

Results

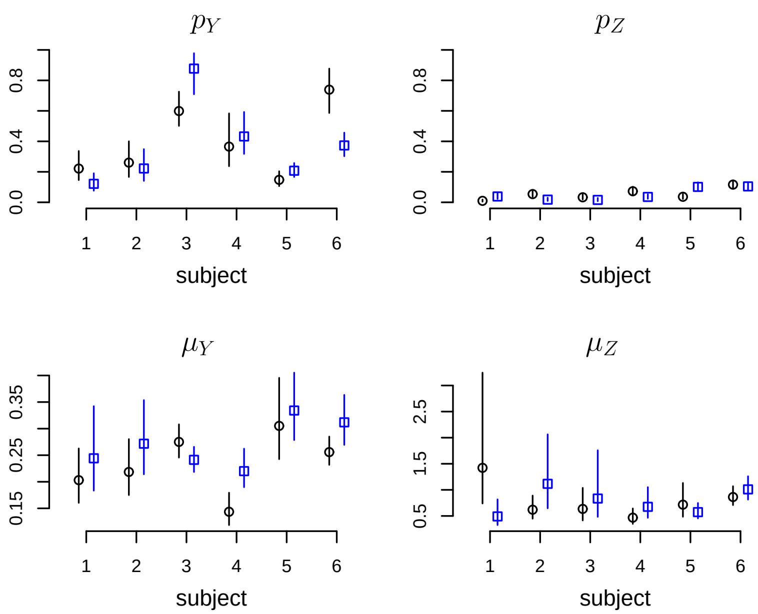
- We used `WinBUGS` (Spiegelhalter, Thomas, Best, and Lunn 2003) to fit the models by Markov Chain Monte Carlo (MCMC) simulation.
  - We post-processed the output in `R` using the `RBugs` package.
- For each subject and either the long or short RSIs experiment:
  - We discarded a burn-in of 12,500 iterations.
  - We then ran each chain for a further 125,000 iterations.
  - We thinned out the chains by subsampling every 25th iteration, to leave 5,000 equally spaced iterations for analysis.
- We assessed convergence of the model using trace plots and autocorrelation plots.

Posterior summaries for time series parameters



- Agreement between the posterior parameters for the long and short RSI experiments.
  - Suggests that the nature of the serial correlation in RTs more closely related to **subject specific** characteristics than to the experimental manipulations.
- Strength of autocorrelation (posterior of ϕ), ranges from weak for subjects 1, 2 and 5 to stronger for subjects 3, 4, and 6.
  - Carry-over effects are larger for some subjects than for others.

Posterior summaries for extreme distribution parameters



Posterior summaries for extreme distribution parameters (cont.)

- Overall, good agreement within subjects for long and short RSI experiments.
- Large between subject variability for the probability of a fast extreme component (p_y), with estimates that range from close to 0 to close to 1.
- The estimated probabilities of a slow extreme component are much less variable and typically smaller in size, ranging between zero and 0.15 for all subjects.
- Expected values of the fast components are generally larger for the long RSI experiments than for the short RSI experiments.
- Slow extreme components, while occurring less frequently, have expected values that are typically larger than those of the fast extreme components.

Bivariate summaries

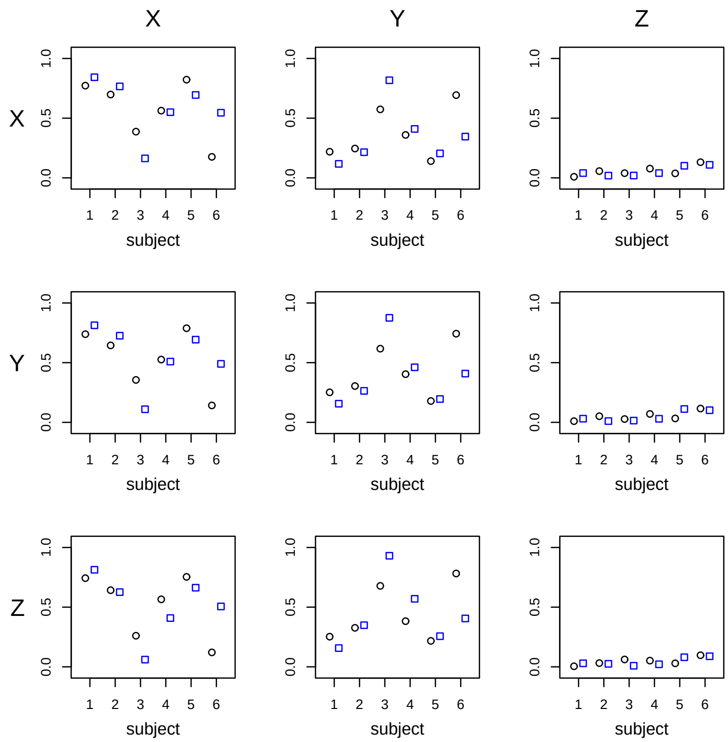
- Used Spearman's rank correlation as a measure of association between posterior parameters (avoids need to make distributional assumptions about the posterior draws).

		Subject					
Variables	RSI	1	2	3	4	5	6
μ and ϕ	Short	**-0.15**	0.02	**-0.09**	**-0.30**	-0.07	**-0.32**
μ and ϕ	Long	-0.04	-0.01	**-0.26**	**-0.34**	0.01	-0.02
μ and μ_y	Short	**0.48**	**0.53**	**0.22**	**0.52**	**0.29**	**0.20**
μ and μ_y	Long	**0.33**	**0.54**	**-0.09**	**0.51**	**0.21**	**0.30**
μ and μ_z	Short	-0.02	-0.01	**-0.10**	-0.01	**-0.09**	**-0.10**
μ and μ_z	Long	**-0.13**	0.03	0.03	**0.05**	**-0.14**	-0.03

Bivariate summaries (cont.)

- General tendency for the correlation between the mean and autoregressive parameter of the AR(1) component to be negative.
  - This would reflect a direct link between a tendency to respond more slowly and the weakening of carry over effects from one trial to the next.
- Mean of the AR(1) component is, with one exception, positively correlated with the mean of the fast component, and, generally, negatively correlated with the mean of the slow component.
  - Interpretation is far from straightforward because the posterior probabilities of the various components also play a role.

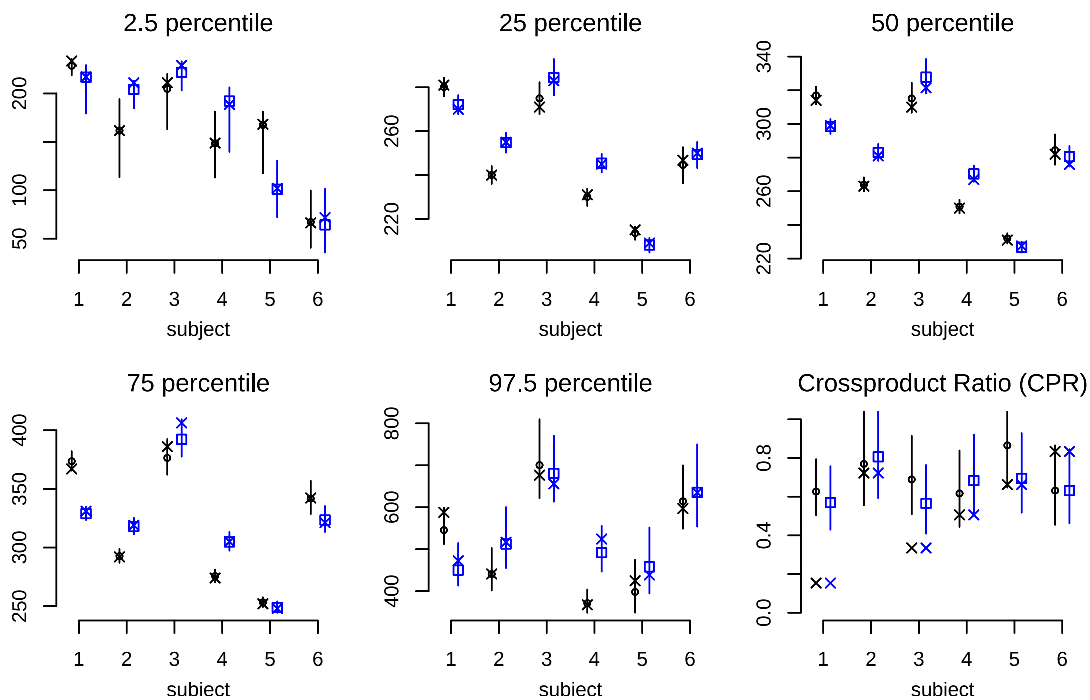
Summarizing the transitions between "states"



Model validation via Monte Carlo simulation

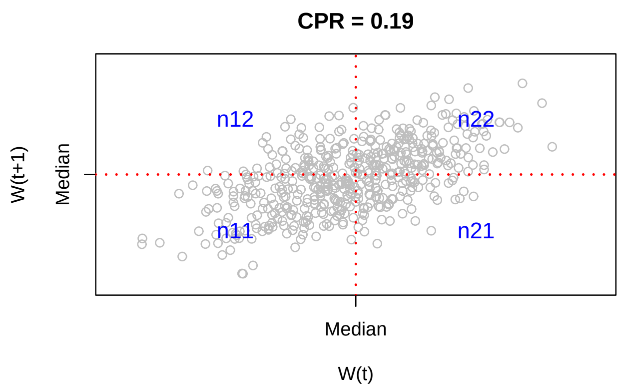
- Is the behavior of RT series simulated from the posterior predictive distributions of the fitted models compatible with the statistical properties of the original RT sequences?
- We assess agreement of marginal distribution and first order serial dependence properties.
- For RT sequence for each subject and RSI:
  - We have 5,000 sets of parameter draws from the posterior distribution.
  - Conditional on each of these sets of parameter draws, we simulate a sequence of de-trended log RT values from our model.
  - Add in the estimate trend and transform from log to original scale.
  - This gives us 5,000 simulated posterior predictive RT sequences.

Marginal and CPR summaries



The cross-product ratio (CPR)

$$\text{CPR} = \frac{n_{12}\, n_{21}}{n_{11}\, n_{22}}$$



- Stronger direct positive association between successive trials would result in more points falling in the upper right and lower left quadrant and produce lower CPR values.
- Satisfactory agreement between observed and the simulated CPR values for most subjects.
- Subject 1 and, to a lesser extent, subject 3 are exceptions, having observed CPR values that are lower than the CPR values predicted by the models.

Discussion

- We have introduced mixture time series models for RT data that are simple to understand and captures a number of typical features of RT data, including **serial dependencies** and both **fast and slow extreme** observations.
- Model can be conveniently fit by MCMC methods, using publicly available software (`WinBUGS`, `R`, and the `RBugs` package).
- Using Monte Carlo simulations, we were able to verify that synthetic data generated from the posterior predictive distributions of the fitted models behaved, to a large extent, like the observed data with regard to both their marginal distributional properties and the nature of their first order serial dependencies.

Where next?

- This model can be used as a starting point for the systematic investigation of effects of experimental factors that influence both the marginal RT distributions and dependencies in a sequence of trials.
  - Incorporate experimental conditions (fixed effects) and subject specific terms (random effects reflecting individual differences) in a hierarchical Bayesian model.
  - Model trend within the Bayesian structure.
  - Explore different time series dependencies.
  - Use flexible parametric families of distributions (Gamma, Weibull, etc.) to model the extreme observations.
  - If necessary, consider for example a Markov structure for the evolution of extreme observations in the mixture.

References

Andrews, S. and A. Heathcote (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory and Cognition 27*, 514–544.

Barnett, V. and T. Lewis (1994). *Outliers in Statistical Data* (3rd ed.). New York: John Wiley & Sons.

Belin, T. R. and D. B. Rubin (1995). The analysis of repeated-measures data on schizophrenic reaction times using mixture models. *Statistics in Medicine 14*, 747–768.

Farrell, S., E.-J. Wagenmakers, and R. Ratcliff (2005). ARFIMA time series modeling of serial correlations in human performance. Submitted for publication.

Gilden, D. L. (1997). Fluctuations in the time required for elementary decisions. *Psychological Science 8*, 296–301.

Gilden, D. L. (2001). Cognitive emissions of $1/f$ noise. *Psychological Review 108*, 33–56.

Gottlob, L. R. (2004). Location cuing and response time distributions in visual attention. *Perception and Psychophysics 66*, 1293–1302.

Heathcote, A., S. Brown, and D. J. K. Mewhort (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review 7*, 185–207.

Heathcote, A., S. J. Popiel, and D. J. Mewhort (1991). Analysis of response time distributions: An example using the stroop task. *Psychological Bulletin 109*, 340–347.

Hohle, R. H. (1965). Inferred components of reaction time as a function of foreperiod duration. *Journal of Experimental Psychology 69*, 382–386.

Jones, M., B. C. Love, and W. T. Maddox (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition 32*, 316–332.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

Meeter, M. and C. N. L. Olivers (2006). Intertrial priming stemming from ambiguity: A new account of priming in visual search. *Visual Cognition 13*, 202–222.

Peña, D., D. Pena, G. C. Tiao, and R. S. Tsay (2001). *A Course in Time Series Analysis*. John Wiley & Sons.

Penner-Wilger, M., C. Leth-Steensen, and J.-A. LeFevre (2002). Decomposing the problem-size effect: A comparison of response time distributions across cultures. *Memory and Cognition 30*, 1160–1167.

Peruggia, P., T. Van Zandt, and M. Chen (2002). Was it a car or a cat I saw? An analysis of response times for word recognition. In *Case Studies in Bayesian Statistics*, Volume 6, pp. 319–334. New York: Springer–Verlag.

Pressing, J. and G. Jolley-Rogers (1997). Spectral properties of human cognition and skill. *Biological Cybernetics 76*, 339–347.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin 114*, 510–532.

Ratcliff, R. and B. B. Murdock, Jr. (1976). Retrieval processes in recognition memory. *Psychological Review 83*, 190–214.

Ratcliff, R. and P. L. Smith (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review 111*, 333–367.

Reber, P. J., P. Alvarez, and L. R. Squire (1997). Reaction time distributions across normal forgetting: Searching for markers of memory. *Learning and Memory 4*, 284–290.

Rouder, J., D. Sun, P. Speckman, J. Lu, and D. Zhou (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika 68*, 589–606.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn (2003). *WinBUGS User Manual, Version 1.4*. Cambridge, UK: MRC Biostatistics Unit.

Stewart, N., G. D. A. Brown, and N. Chater (2005). Absolute identification by relative judgment. *Psychological Review 112*, 881–911.

Thornton, T. L. and D. L. Gilden (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin and Review 12*, 409–441.

Ulrich, R. and J. Miller (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General 123*, 34–80.

Van Orden, G. C., J. G. Holden, and M. T. Turvey (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General 132*, 331–350.

Van Selst, M. and P. Jolicoeur (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology 47*, 631–650.

Wagenmakers, E.-J., S. Farrell, and R. Ratcliff (2004). Estimation and interpretation of $1/f$ noise in human cognition. *Psychonomic Bulletin & Review*.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association 93*, 341–348.