Latent variable models for discriminating trend, dependence, and tail structure in response time data

Peter F. Craigmile

Department of Statistics, The Ohio State University

http://www.stat.osu.edu/~pfc/


Fri 10 Sep 2010

Joint with Mario Peruggia (Statistics) and Trisha Van Zandt (Psychology).

Supported by the National Science Foundation under award numbers BCS-0738059, DMS-0604963, DMS-0605052, SES-0214574 and SES-0437251.
Michael . . .

- Is a statistician and a psychologist.
- Is a valuable member of our department.
- Is integral in teaching multivariate statistics to our students.
- I enjoy every conversation I have with him.
- I wish we would see him more.
- Recently he has been eager to SPAM . . .
SPAM (Statistics, Psychology and Marketing)
Modeling Processes

One can use most statistical models, including the familiar multiple regression models in psychology, in highly atheoretical ways, that is, solely for forecasting and prediction. Such pragmatic applications are commonplace in economics, engineering, and environmental science. In these cases, virtually no theoretical meaning need be ascribed to the parameters of the model. What matters is how accurately the model forecasts or predicts, not how well it “squares” with some theory-based set of expectations. This orientation used to be prevalent in psychology, especially among devotees of multiple regression, but many psychologists are uncomfortable with such an atheoretical stance at this point in the history of the discipline. Whether it is in the context of an elaborate structural model or simply a stepwise regression analysis, some “interpretation” of model parameters tends to be undertaken, especially so if theories concerning processes are being developed and refined.
A central premise on which we will anchor the remainder of this discussion is that “process” involves patterns of changes that are defined across variables and organized over time [see, e.g., Nesselroade & Molenaar (2003) for additional discussion of this and related points]. In physics, the label *dynamic* typically pertains to a force or forces that produce motion. In accordance with this usage, we will use the label *dynamic* for a force or forces driving intraindividual change. The variables that are actually being changed will be referred to as the *process* variables. Thus, a model for a dynamic, psychological process should include (a) variables representing the forces that cause change, (b) variables on which the changes (outcomes) are actually manifested, and (c) parameters that define, at least to some extent, a temporal flow in the relationships among and between the two kinds of variables. Thus, the actual values of the model parameters are highly salient to understanding and communicating the nature of process information.
Response time (RT) – a measure of human performance

- 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.
  - Employed as a basis to evaluate training regimens, user interface design, vehicle operation, and task design.
  - Can be used to evaluate medical conditions, especially schizophrenia, learning disorders, and other psychological disorders.
Efforts to model processes than give rise to patterns in RT

• Has been the focus of attention for the last 50 years.
  
  – See Luce [1986] for a review.

• Two strategies for modeling:

  1. Sequences of RTs are independent and identically distributed (IID).
     “the statis/equilibrium perspective” [Browne and Nesselroade, 2005]

  2. Discard independence and model the dependence in sequences of RTs.
     “intravidual variability and change” [Browne and Nesselroade, 2005]
IID approaches to modeling RT

• Most commonly, derive a hypothesis about how a 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 entire sample of RTs.

• Common marginal RT distributions: gamma, Weibull, inverse Gaussian, and the ex-Gaussian.

• Other theoretical approaches are based on the minima of first passage time distributions [Ratcliff and Smith, 2004] – see later.
Dependence in RTs

• Embraces the idea that sequences of RTs are neither independent nor IID.

1. Previous stimuli and responses influences processing of later stimuli and facilitates or inhibits responses to those stimuli. e.g., sequential effects, priming, inhibition of return, task-switching.

[Brown et al., 2008, Jones et al., 2006, Meeter and Olivers, 2006, Stewart et al., 2005, Steinhauser and Hübner, 2009].

2. Gilden and others emphasize dependence by modeling long-range effects across sequences of trials, or $1/f$ (pink) noise.


• Common theme – model sequences of RTs using time series processes.
Extreme observations

- Little agreement on how to handle extreme observations.

- But, modeling such data of are interest in both psychology and statistics.


- Since RT data are positively skewed, some tail values are to be expected from the process that generated the data.
Extreme observations, cont.

• More extreme observations in the upper tail are due to other factors; e.g.,
  – Unscheduled rests in the middle of a block of trials.
  – Lapses in attention.
  – Intrusions of the subject’s physiological system (sneezing, itching, etc.).

• Extreme observations in the lower tail are harder to detect:
  – Ex: fast RTs are commonly attributed to rapid guessing or error (twitches).
Hierarchical model approaches

- Attractive because allow the flexibility to specify:

  RT distributions that depend on varying experimental conditions.

- Typically parameterized via hierarchical regression models for parameters parameterizing the experimental conditions. Examples:

  - Rouder et al. [2003] assume that the observed RTs are conditionally independent, ignoring sequential dependencies.
  - Peruggia et al. [2002] attempt to capture sequential dependencies through an autoregressive structure for the log of the scale parameter of the RT distribution.
Craigmile, Peruggia and Van Zandt (To appear, Psychometrika)

- RT sequences can be modeled as a mixture of latent components incorporating:
  1. A trend term to capture smooth changes in RT levels due to learning effects, fatigue, etc..
  2. A time series component that captures local dependencies.
  3. Components that model the upper and lower tail behaviour.

- Provides a framework for performing RT analysis that does not compromise model realism and autocorrelation structure.

- Can be used to build sensible hierarchical models which are not based entirely on consideration of convenience (namely, conditional conjugacy).
A dataset from Wagenmakers et al. [2004]

- A “simple RT” study with 6 subjects, to investigate autocorrelation structure across long sequences of trials.

  - A randomly generated response-stimulus interval or RSI is used to prevent anticipatory responses (extremely fast RTs).

  - Two different RSIs conditions:

    Short RSIs: $\sim U(550\text{ms}, 950\text{ms})$.

    Long RSIs: $\sim U(1150\text{ms}, 1550\text{ms})$. 
A dataset from Wagenmakers et al. [2004]

The subject sees the symbol on the screen and presses the ’/’ key

The time to respond is the first RT
A dataset from Wagenmakers et al. [2004]

The subject then sees nothing for a randomly generated response-stimulus interval (RSI)
A dataset from Wagenmakers et al. [2004]

The subject sees the symbol again and presses the '/' key.

The time to respond now is the second RT.
A dataset from Wagenmakers et al. [2004]

Then again nothing for another randomly generated response-stimulus interval (RSI)
A dataset from Wagenmakers et al. [2004]

We continue in the same way until we collect 1024+24 trials per subject and RSI condition.

The first 24 trials are discarded.
Exploratory RT analysis

- We use a “trial” series plot of the log transformed RTs to investigate for:

1. potential trends (smooth changes of the level of the process over time);
2. potential extreme observations in both tails of the distribution;
3. to assess the extent of serial dependence in each sequence.
Subjects 1, 2, and 3

subject 1, short RSI

subject 2, short RSI

subject 3, short RSI

subject 1, long RSI

subject 2, long RSI

subject 3, long RSI
Subjects 4, 5, and 6

subject 4, short RSI

subject 4, long RSI

subject 5, short RSI

subject 5, long RSI

subject 6, short RSI

subject 6, long RSI
Trends in the log RT series

- **Trends** occur for a variety of reasons including, over long scales,
  - learning (negative trends) or
  - fatigue (positive trends).

- Also, fluctuations in attentional state or task readiness may result in shifts in the mean of the process over medium scales.

- Remember our data is non-Gaussian and dependent.
  - Without special care, standard methods to estimate trend will **over-smooth**.
  - Craigmile et al. [2010] discuss the issue of detrending RT sequences.
A wavelet-based approach to estimate the trend

- **Wavelets** are “small waves”.

- A wavelet transformation decomposes a time series into:
  1. **averages** over a specified (typically long) time scale, and
  2. **changes of averages** over a range of shorter time scales.

- It is a time-scale (some say frequency) decomposition.

- We obtain a **naive** estimate of trend by zeroing out the changes of averages over the shortest time scales.

[Brillinger, 1994, Craigmile et al., 2004]
Trends for subjects 1, 2, and 3

subject 1, short RSI

subject 1, long RSI

subject 2, short RSI

subject 2, long RSI

subject 3, short RSI

subject 3, long RSI
Trends for subjects 4, 5, and 6

subject 4, short RSI

subject 4, long RSI

subject 5, short RSI

subject 5, long RSI

subject 6, short RSI

subject 6, long RSI

Trial number
A Bayesian hierarchical model for the simple RT data

• Assume log RTs are realizations of a wavelet-based model for trend plus a noise process which follows a correlated mixture of Gaussian and ex-Gaussian distributions.

• The model can account for sequential dependencies and allows for the possible occurrence of short and long responses.

• We assume a hierarchical structure to combine information across subjects and conditions.
A model for log RTs

• The $t$th log RT for subject $s$ under RSI condition $c$ ($c = 0$: short RSI; $c = 1$: long RSI) follows

$$R_{s,c,t} = \mu_{s,c,t} + S_{s,c,t}.$$ 

• We model the trend $\mu_{s,c,t}$ using a slightly modified, but standard, wavelet regression model with coefficients depending on subject and RSI condition. [See e.g., Müller and Vidakovic, 1999]

– Within subject/condition: The trend borrows strength over different scales.
– Between subjects/conditions: We borrow strength knowing that condition is nested within subject.
The model for the “irregular noise process” \( \{ S_{s,c,t} \} \)

\[
S_{s,c,t} = \begin{cases} 
X_{s,c,t} & \text{with probability } p_{X,s,c} \\
X_{s,c,t} + Y_{s,c,t} & \text{with probability } p_{Y,s,c} \\
X_{s,c,t} - Z_{s,c,t} & \text{with probability } p_{Z,s,c},
\end{cases}
\]

- \( \{ X_{s,c,t} \} \) is a hidden stationary autoregressive process of order one

[Read a discussion of stationarity in Du Toit and Browne, 2007].

- The possible occurrence of extreme observations is modeled by \( \{ Y_{s,c,t} \} \) and \( \{ Z_{s,c,t} \} \), which, conditional on their means \( 1/\lambda_{Y,s,c} \) and \( 1/\lambda_{Z,s,c} \), are independent sequences of exponential random variables.

  - Build models for the means that learn across subjects and RSI conditions.

- Nothing surprising for prior distributions.
Results via MCMC: fitted trends for subjects 1, 2, and 3
Results: fitted trends for subjects 4, 5, and 6
Posterior summaries for the AR(1) process

\[
\begin{align*}
\phi_{s,c} &\quad \sigma_{s,c}
\end{align*}
\]
Posterior summaries for the tail parameters

$p_{Y, s, c}$

$p_{Z, s, c}$

$1/\lambda_{Y, s, c}$

$1/\lambda_{Z, s, c}$
Learning across subjects and conditions

\[ \alpha_V \]

\[ \gamma_V \]

\[ \alpha_Y \]

\[ \gamma_Y \]

\[ \alpha_Z \]

\[ \gamma_Z \]
Model validation

• We considered two predictive methods:

1. Compare the marginal and higher order serial dependence properties of synthetic data generated from the posterior predictive distribution with those of the observed data.

2. Use a modified prequential approach which compares the empirical distribution of the generalized residuals (the values of omitted observations evaluated at the estimated predictive CDFs) to a uniform distribution.

• Our model fits better compared to simplified versions of the model.
Moving on: Interesting RT experimental designs

- We have started to collect our own data.

- We obtain RT trials (time series) as part of an experimental design.

  [Not a new idea; see, e.g., Brillinger, 1973]
A first experiment

- Similar to Wagenmakers et al. [2004], except we use non-random RSIs.
- For each subject we run our experiment over six days.
- For each day we use three different RSI values (200, 400 and 600ms).
- We randomize order of RSI values over days using a latin square design.
- For each condition we collect 128 RT trials with small symbols on the screen, followed by 128 further trials with larger symbols.
Days 1, 2, and 3
Days 4, 5, and 6
Alternative process models

“Other theoretical approaches are based on the minima of first passage time distributions [Ratcliff and Smith, 2004]”

• For \( \{V_s\} \) a standard Brownian motion, consider the stochastic process \( \{W_s\} \) defined by \( W_0 = 0 \) with

\[
W_s = \nu t + \sigma V_s,
\]

where \( \nu > 0 \) is the drift parameter and \( \sigma > 0 \) is a scale parameter.

• This defines a Brownian motion with drift.
The inverse Gaussian distribution

• Assuming that $\sigma = 1$ the time for $\{W_s\}$ to hit a threshold $\alpha$ is distributed according to an inverse Gaussian distribution with parameters $\alpha/\nu$ and $\alpha^2$.

• Letting the random variable $X$ denote this time for one RT trial:

  - The mean is $E(X) = \alpha/\nu \equiv \mu$.
  - The variance is $\text{var}(X) = \mu^3/\alpha^2 = \alpha/\nu^3$. 

Process modeling

- Question:
  - Is the inverse Gaussian distribution “good” for explaining RT data \( \{X_t\} \)?

- Answer: Early studies indicate maybe!
  
  1. Need to consider the threshold \( \{\alpha_t\} \) and the drift \( \{\nu_t\} \) as stochastic processes.
  
  2. We still need a mixture component to handle the extremes.

- Plenty of work still to do!
Discussion

- How do we build models for time series collected in experiments?

- Non-Gaussian time series model are still a fertile area of research.
  
  - Many distributions have, at the least, interesting mean-variance relationships which make for interesting time series behaviors.

- Inference and model validation is much more involved than for Gaussian processes.

THANK YOU!
References


