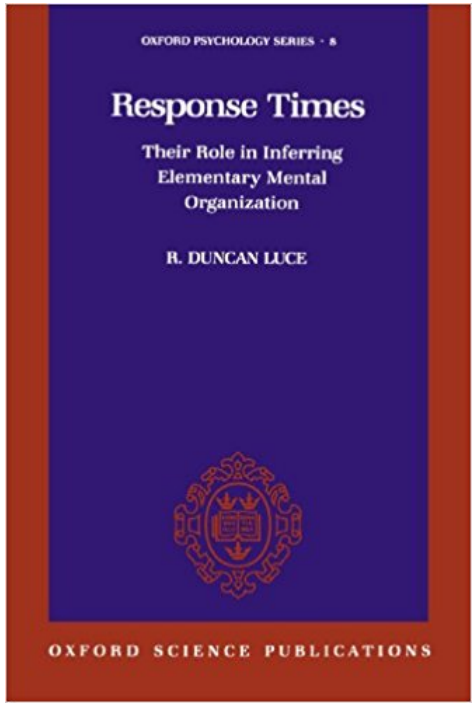
A hierarchical Bayesian shifted Wald model with censoring

Thomas J. Faulkenberry

Tarleton State University

Virtual MathPsych 2023

Response times are a classic measure of cognitive processing



Preface

For almost as long as I have been doing research, response times have struck me as a fascinating, albeit tricky, source of information about how the mind is organized. Whenever I teach mathematical psychology or psychophysics, I include a dose of response times along with the repeated admonition that we surely do not understand a choice process very thoroughly until we can account for the time required for it to be carried out. When I came to Harvard in 1976 I offered, for the first time, a seminar-course on the subject (in style more a course, in size more a seminar), first with David Green and later alone. It was only then that I felt a need for a more systematic mastery of the field, and in academic 80–81 when I had a sabbatical leave, the Guggenheim Foundation agreed to support my self-education and the beginnings of this book.

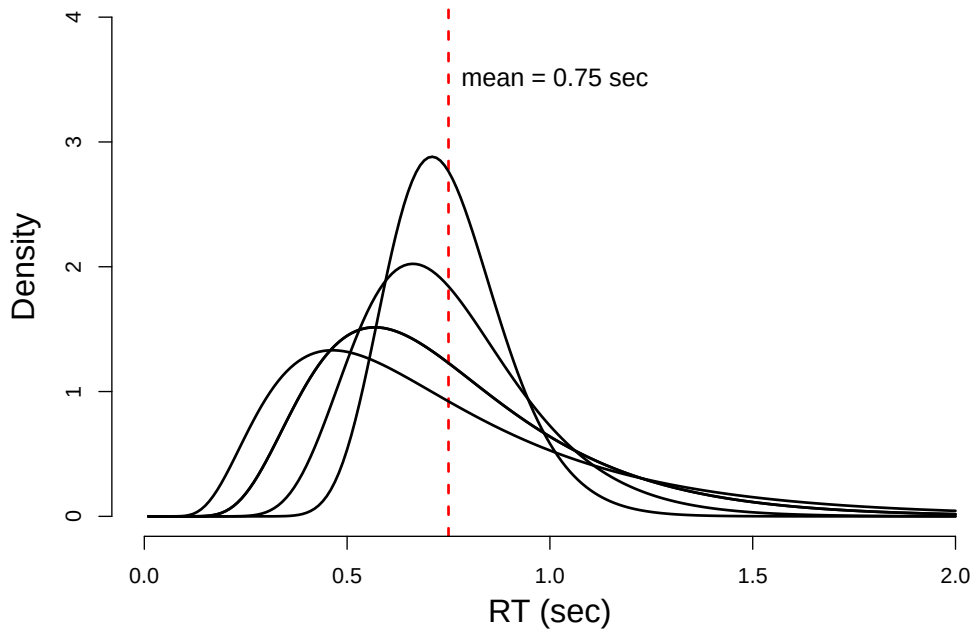
In cognition, RTs inform our understanding of latent processes involved in various decision-making tasks.

Classical technique:

1. have people do a bunch of trials – measure RT on each trial
2. find mean RT for each experimental condition / person
3. test for differences in means (e.g., ANOVA)

This method is lossy. . . it collapses each person's RT distribution to a single number

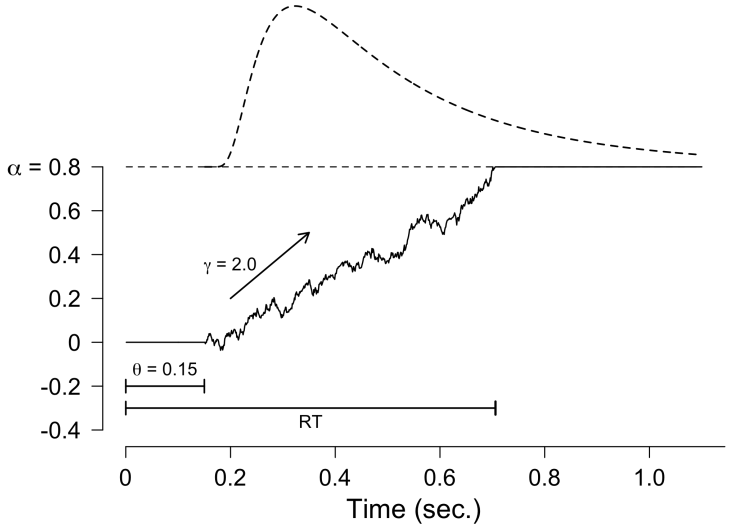
The mean does not uniquely identify the distribution – same mean, different behavioral signatures



Need a way to capture the distribution of response times.

- Gold standard – diffusion model
- Problem – estimating diffusion parameters requires a sufficient number of valid error RTs
- . . . but in some tasks, errors are quite rare.

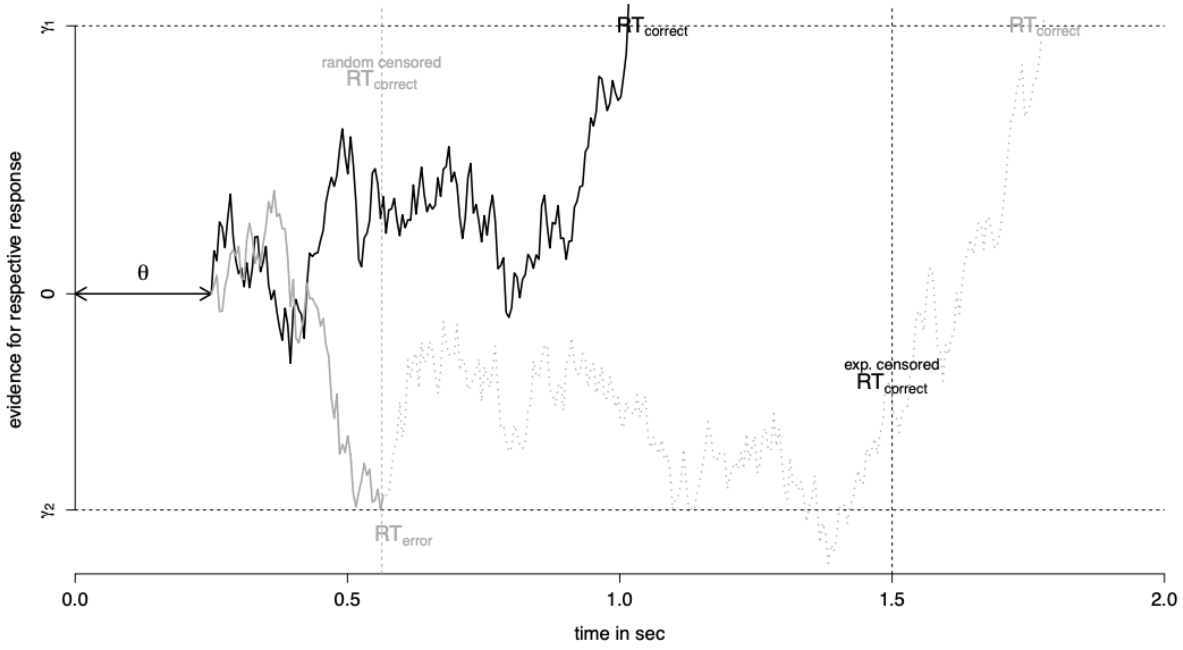
The shifted Wald model provides one way to model RTs in high accuracy tasks, but. . .



- it is a single boundary accumulator, so errors are ignored
- parameters do not uniquely correspond to diffusion parameters (Matzke & Wagenmakers, 2009)

Goal – augment the shifted Wald model so that it can account for (a small number of) error trials.

- use a censored shifted Wald model (Miller et al., 2018)



How to build a likelihood for a censored SW model:

- if a trial is correct, then the likelihood follows directly from the density function:

$$\begin{aligned}\mathcal{L}(\gamma, \alpha, \theta \mid x) &= f_{SW}(x \mid \gamma, \alpha, \theta) \\ &= \frac{\alpha}{\sqrt{2\pi(x-\theta)^3}} \cdot \exp\left(-\frac{(\alpha - \gamma(x-\theta))^2}{2(x-\theta)}\right)\end{aligned}$$

How to build a likelihood for a censored SW model:

- if a trial is incorrect, then we consider that the accumulator toward the *correct* boundary is still ongoing, and the recorded RT is randomly censored at some point prior.

- in this case, the likelihood comes from the survival function:

$$\begin{aligned}\mathcal{L}(\gamma, \alpha, \theta \mid x) &= S_{SW}(x \mid \gamma, \alpha, \theta) \\ &= 1 - F_{SW}(x \mid \gamma, \alpha, \theta) \\ &= 1 - \Phi\left(\frac{\alpha(x-\theta) - \gamma}{\sqrt{x-\theta}}\right) - e^{2\gamma\alpha}\Phi\left(-\frac{\alpha(x-\theta) + \gamma}{\sqrt{x-\theta}}\right)\end{aligned}$$

So, for a collection of N observed RTs $\mathbf{x} = \{x_i\}$, we have:

$$\mathcal{L}(\gamma, \alpha, \theta \mid \mathbf{x}) = \prod_{i=1}^{N} f_{SW}(x_i \mid \gamma, \alpha, \theta)^{1-d_i} \cdot \prod_{i=1}^{N} S_{SW}(x_i \mid \gamma, \alpha, \theta)^{d_i}$$

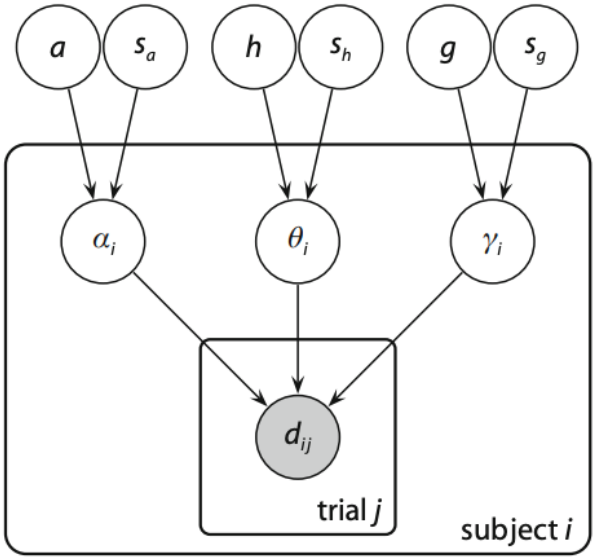
where d_i is a censoring indicator:

$$d_i = \begin{cases} 0 & \text{if correct on trial } i \\ 1 & \text{if incorrect on trial } i \end{cases}$$

This gives a log likelihood of

$$\begin{aligned}
\log \mathcal{L}(\gamma, \alpha, \theta \mid \mathbf{x}) &= \log \left(\prod_{i=1}^{N} f_{SW}(x_i \mid \gamma, \alpha, \theta)^{1-d_i} \cdot \prod_{i=1}^{N} S_{SW}(x_i \mid \gamma, \alpha, \theta)^{d_i} \right) \\
&= \sum_{i=1}^{N} \log f_{SW}(x_i \mid \gamma, \alpha, \theta)^{1-d_i} + \sum_{i=1}^{N} \log S_{SW}(x_i \mid \gamma, \alpha, \theta)^{d_i} \\
&= \sum_{i=1}^{N} \log(1-d_i) f_{SW}(x_i \mid \gamma, \alpha, \theta) + \sum_{i=1}^{N} d_i \log S_{SW}(x_i \mid \gamma, \alpha, \theta)
\end{aligned}$$

Miller et al. (2018) used maximum likelihood estimation to show that the censored shifted Wald was able to accurately recover diffusion model parameters. What about a hierarchical Bayesian approach (i.e., Matzke & Wagenmakers, 2009)?



$a \sim \text{Uniform}(0.67, 2.35)$
$s_a \sim \text{Uniform}(0, 0.485)$
$h \sim \text{Uniform}(0, 0.82)$
$s_h \sim \text{Uniform}(0, 0.237)$
$g \sim \text{Uniform}(0.85, 7.43)$
$s_g \sim \text{Uniform}(0, 1.899)$

$\alpha_i \sim \text{Gaussian}(a, s_a^2)$
$\theta_i \sim \text{Gaussian}(h, s_h^2)$
$\gamma_i \sim \text{Gaussian}(g, s_g^2)$

$d_{ij} \sim \text{Shifted Wald}(\alpha_i, \theta_i, \gamma_i)$

Thank you!

- slides available at `https://tomfaulkenberry.github.io`
- Twitter/Mastodon: `@tomfaulkenberry@mathstodon.xyz`
- Email: `faulkenberry@tarleton.edu`