Parameter estimation and simulation for one-choice Ratcliff diffusion model

Amiya Patanaik
Nanyang Technological University
School of Computer Engineering
Singapore
amiya1@e.ntu.edu.sg

Vitali Zagorodnov
Nanyang Technological University
School of Computer Engineering
Singapore
zvitali@ntu.edu.sg

C.K. Kwoh
Nanyang Technological University
School of Computer Engineering
Singapore
asckkwoh@ntu.edu.sg

ABSTRACT
One-choice reaction-time (RT) tasks are very popular in cognitive psychology to access and study cognitive and behavioral performance. The Ratcliff diffusion model is a powerful cognitive model based on the concept of evidence accumulation, a well-established phenomenon. Initially proposed for two-choice RT tasks, the model was recently adapted to one-choice RT tasks. The wide spread use of the model is restricted due to complexity involved in estimating the model parameters from data. Furthermore the current method for simulating the model is prohibitively slow. We propose improvements in the simulation process that considerably improves runtime. We also derive a closed form solution for the model and propose a maximum likelihood estimator to determine the parameters.

Categories and Subject Descriptors
G.3 [**Mathematics of Computing**]: Probability and Statistics – *Distribution functions, Stochastic processes.*

Keywords
Ratcliff diffusion model, one-choice reaction-time, parameter estimation, simulation.

1. INTRODUCTION
Mental chronometry, i.e. the use of reaction-time (RT) in perceptual motor tasks to infer the content, duration and temporal sequencing of cognitive operations is one of the most prevalent approaches in cognitive and experimental psychology. Considerable research has gone into developing methods for analyzing RT data. To draw inferences about mental processes researches initially relied upon summery statistics like mean, median and standard deviation of RT. It was soon realized that these measures fail to capture the complete RT distribution. RT distributions are notoriously skewed; ignoring the RT distribution shape can have detrimental effect on the analysis. The general trend has been to move away from the use of less informative summery statistics to distributional approaches. The Ratcliff diffusion model [1] has been especially successful as it not only provides a good model fit but also the model parameters are strongly tied to the underlying neuro-cognitive process [2]. It is a model of perceptual decision making which allows a fine-grained analysis of the cognitive processes underlying simple RT tasks and provides unambiguous interpretations of the parameters [3]. It assumes that perceptual decision-making follows the accumulation of evidence over a period of time, a principle supported by multiple experiments [4-7]. The model was initially proposed for a two-choice decision task [8] and has been applied successfully to experiments in many different fields. Despite its advantages, the adoption of the model in the research community had been slow due to considerable mathematical complexities in simulating and estimating the parameters of the model. Vandekerckhove *et al.* [9] suggested a general method for performing two-choice diffusion model analysis on experimental data to address the problem. The model was recently adapted to one-choice RT task [10]. While the single choice RT diffusion model is considerably simpler than the two-choice RT model; the estimation and simulation process described by Ratcliff *et al.* [10] can be considerably improved and simplified. We believe this will help researchers adopt and leverage a powerful model of perceptual decision making for single choice RT experiments.

SAC'14, March 24-28, 2014, Gyeongju, Korea.

http://dx.doi.org/10.1145/2554850.2554872

2. THE RATCLIFF MODEL
Simple perceptual decision-making can be thought of as involving a decision and a non-decision component. The non-decision component comprises time spent encoding sensory input (pre-decision time) and time taken to execute a decision once it is taken (post-decision time) see Figure 1a. Critically, the evidence accumulation stage within the decision component can be modeled mathematically as a diffusion process component (see [11] for details). In the Ratcliff diffusion model, the non-decision time is assumed to vary from trial to trial according to a uniform distribution with mean T_{er} and width S_t. The decision time is modeled using a single boundary diffusion process with a drift. Evidence is accumulated from starting point of 0 until it reaches the boundary at a. The drift parameter is also allowed to vary across trials according to a normal distribution with mean ξ and standard deviation η (Figure 1b.).

The five parameters of the model $\Theta = [a, T_{er}, S_t, \xi, \eta]$ do not uniquely identify the model [10]. For example the boundary parameter a can be scaled by equally scaling the drift parameters (ξ and η) without affecting the RT distribution. Ratcliff *et al.* [10] recommend using invariant parameter ratios ($\xi/\eta, a/\xi$) instead of the original parameters. Or the boundary parameter a can be normalized to a fixed value and only the rest four appropriately scaled parameters can be reported.

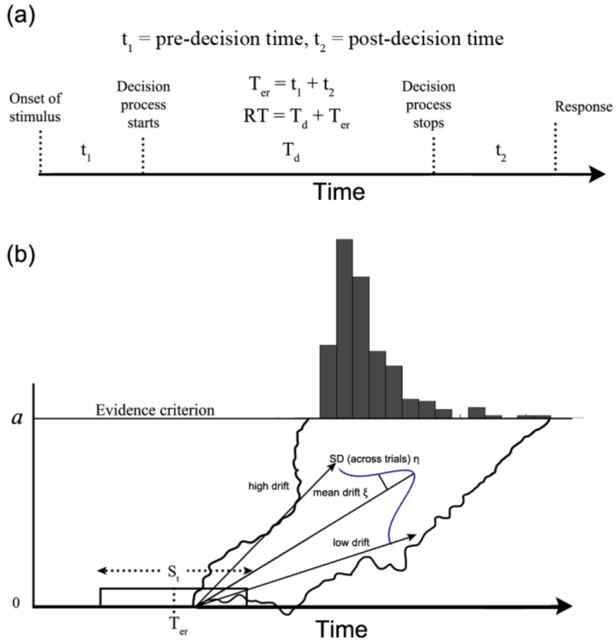

Figure 1: (a) Components of reaction time (RT). (b) Diffusion model for one-choice RT task. Note that while drift is symmetrically distributed about ξ, drift angle is not! It is shown symmetrical here for representational purposes

3. SIMULATION AND ESTIMATION

3.1 Existing methods

To simulate the model, the drift and non-decision time can be sampled from their respective distributions. The diffusion process is stochastic in nature and can be simulated using a random walk approximation with very small displacements Δ and small time intervals τ. The process stops when the process reaches the boundary at a. At every step the probability of making a positive move (i.e. $+\Delta$) is p and a negative move ($-\Delta$) is $1-p$. If the following conditions are satisfied:

$$\Delta = \sigma\sqrt{\tau} \;\; and \;\; p = \frac{1}{2}\left(1 + \frac{\mu\sqrt{\tau}}{\sigma}\right) \qquad (1)$$

then the random walk converges to a diffusion process with drift μ and variance σ^2 under the limiting condition $\tau \to 0$ (see Cox and Miller, page 205 [12]). Note that σ^2 is volatility of the diffusion process which is different from inter-trail variance in drift η^2. It is another scaling parameter, the exact value of which can be ignored. We have fixed the value of σ at 0.1 for all our analysis, which is a value used by previous researchers [8, 9]. The accuracy of the simulation depends on the value of τ which has a trade off with speed.

The parameter estimation processes described by Ratcliff *et al.* [10] relies on simulation based on random walk approximation and leaves several free parameters that can be altered depending on the data. To estimate the parameters Θ of the model, RT distribution is simulated and compared with RT data obtained from experiments using a χ^2 goodness of fit. The process is started with an initial guess for the parameters using a Markov chain Monte Carlo maximum likelihood estimate [13] and utilized an iterative algorithm based on simplex minimization routine [14] to successively improve the model fit. For the simulation process 20,000 RTs per distribution are obtained. A step size of 0.5 ms for the random walk approximation is chosen. To fit the model to data, 0.05, 0.1, … ,0.95 quantiles of the RT distribution is obtained. These quantile RTs are used to find the proportion of responses in the RT distribution from the model lying between the data quantiles which are multiplied by the number of observations to get the expected values (E). The observed values (O) are simply 0.05 multiplied by the number of observations. A χ^2 statistic is then computed as $\sum(O-E)^2/E$. The estimation process is summarized in Fig 2.

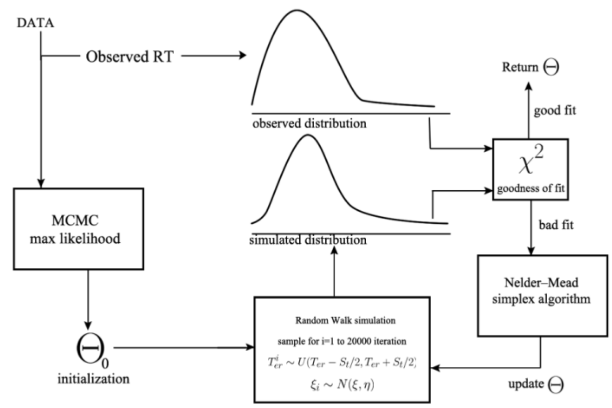

Figure 2: Parameter Estimation Process described by Ratcliff *et al*. [10]

3.2 Improvements in simulation

Ratcliff *et al.* [10] suggested using random walk approximation to diffusion process because of lack of explicit mathematical solution to RT distribution with negative drift rates (which results from across-trial variability η). It is known [15] that a diffusion process starting from origin with drift μ and volatility σ^2 with process termination boundary at a has a distribution:

$$f(t|a,\mu) = \frac{a}{\sigma\sqrt{2\pi t^3}} e^{\left\{-\frac{(a-\mu t)^2}{2\sigma^2 t}\right\}}, t > 0 \qquad (2)$$

Which is an inverse Gaussian distribution. If drift is positive, then the process is guaranteed to terminate. On the other hand if drift is zero or negative ($\mu \leq 0$) then the process may never terminate and the distribution (2) becomes defective. In other words for negative or zero drift value, the process has a finite probability of never terminating and the point mass at infinity is:

$$P(T = \infty) = 1 - e^{\frac{2\xi}{\sigma^2}}, \xi \leq 0 \qquad (3)$$

Fast and efficient method of sampling from the inverse Gaussian distribution (2) exists [16]. Therefore samples can be generated much more efficiently and accurately when drift is positive without relying on random walk approximation. For most experiments the proportion of negative or zero drift is very small. By relying on the inverse Gaussian distribution for positive drifts and falling back on random walk approximation otherwise can significantly improve simulation speed without compromising accuracy.

3.3 Improvements in estimation

The complete distribution for the model is a compound probability distribution which can be derived from (2). The resulting distributions are defective, but nonetheless can be used to get a maximum likelihood estimate of the parameters. Starting from standard diffusion process with drift μ and volatility σ^2 originating at 0 with process termination boundary at a, the distribution of RT is $f(t)$. Now assuming $\mu \sim N(\xi, \eta)$, the resulting compound distribution is given by:

$$g(t|a,\xi,\eta) = \int_{-\infty}^{\infty} f(t|\mu)p(\mu)d\mu, t > 0 \tag{4}$$

$$g(t|a,\xi,\eta) = \int_{-\infty}^{\infty} \frac{a}{\sigma\sqrt{2\pi t^3}} e^{\left\{-\frac{(a-\mu t)^2}{2\sigma^2 t}\right\}} \left(\frac{1}{\sqrt{2\pi\eta^2}}\right) e^{-\frac{(\mu-\xi)^2}{2\eta^2}} d\mu,\ t > 0 \tag{5}$$

The integration can be done by manipulating the quadratic forms in the exponents to yield:

$$g(t|a,\xi,\eta) = \frac{a}{\sqrt{2\pi t^3(\eta^2 t + \sigma^2)}} e^{\left\{-\frac{(a-\xi t)^2}{2t(\eta^2 t+\sigma^2)}\right\}}, t > 0 \tag{6}$$

$g(t|a,\xi,\eta)$ can be taken to be 0 for $t \leq 0$. Then the cumulative distribution function $G(t|\xi,\eta)$ is given by:

$$G(t|a,\xi,\eta) = \begin{cases} 1 - \Phi\left(\frac{a-\xi t}{\sqrt{t^2\eta^2+\sigma^2 t}}\right) + \\ e^{\frac{2a\xi+2a^2\eta^2}{\sigma^2}} \Phi\left(-\frac{a\sigma^2+2at\eta^2+\sigma^2\xi t}{\sigma^2\sqrt{t^2\eta^2+\sigma^2 t}}\right), t > 0 \\ 0, t \leq 0 \end{cases} \tag{7}$$

Where Φ is the cumulative distribution function of the standard normal distribution. In the Ratcliff model the distribution (6) is shifted because of the non-decision time. Let the shift in distribution be t_0. Then (6) can be appropriately modified to:

$$g(t|a,\xi,\eta,t_0) = \begin{cases} \frac{a}{\sqrt{2\pi t_k^3(\eta^2 t_k+\sigma^2)}} e^{\left\{-\frac{(a-\xi t_k)^2}{2t(\eta^2 t_k+\sigma^2)}\right\}}, t_k = t - t_0, t_k > 0 \\ 0, t_k \leq 0 \end{cases} \tag{8}$$

As per the model $t_0 \sim U\left(T_{er} - \frac{S_t}{2}, T_{er} + \frac{S_t}{2}\right)$, the resulting compound distribution will be:

$$r(t|\Theta) = \int_{T_{er}-\frac{S_t}{2}}^{T_{er}+\frac{S_t}{2}} g(t|a,\xi,\eta,t_0)p(t_0)dt_0 \tag{9}$$

$$r(t|\Theta) = \left(\frac{1}{S_t}\right) \int_{t-T_{er}-\frac{S_t}{2}}^{t-T_{er}+\frac{S_t}{2}} g(t_k|a,\xi,\eta)dt_k \tag{10}$$

$$r(t|\Theta) = \left(\frac{1}{S_t}\right)\left(\int_{-\infty}^{t-T_{er}+\frac{S_t}{2}} g(t_k|a,\xi,\eta)dt_k - \int_{-\infty}^{t-T_{er}-\frac{S_t}{2}} g(t_k|a,\xi,\eta)dt_k\right) \tag{11}$$

$$r(t|\Theta) = \left(\frac{1}{S_t}\right)\left\{G\left(t - T_{er} + \frac{S_t}{2}|a,\xi,\eta\right) - G\left(t - T_{er} - \frac{S_t}{2}|a,\xi,\eta\right)\right\} \tag{12}$$

Equation (12) captures the defective probability density function for the one-choice Ratcliff diffusion model. Instead of relying on simulation, the model parameter (Θ) can now be estimated using a maximum likelihood estimate (MLE). Given a set of observation $\{t_1, t_2, \dots, t_n\}$ the log likelihood of parameter vector Θ can be computed as:

$$\ln \mathcal{L}(\Theta|t_1,\dots,t_n) = \sum_{i=1}^{n} \ln\left(r(t_i|\Theta)\right) \tag{13}$$

It is quite common to have censored data in one-choice RT task experiments due to limitation of time. If the experiment is censored at $t = t^c$, and there are m censored observations along with n uncensored observations the log likelihood will be modified to:

$$\text{Ln}\,\mathcal{L}(\Theta|t_1,\dots,t_n,t_1^c,\dots,t_m^c) = \sum_{i=1}^{n} \ln\left(r(t_i|\Theta)\right) + m\ln\left(1 - R(t_i^c|\Theta)\right) \tag{14}$$

The minimum negative log likelihood is estimated using interior-point algorithm [17]. $R(t_i^c|\Theta)$is computed using trapezoidal numerical integration. For starting the estimation process the boundary a is chosen randomly. Initial value of mean drift ξ^0 and mean non-decision time T_{er}^0 is computed as:

$$\xi^0 = \sqrt[3]{\frac{a\sigma^2}{\sigma_{RT}^2}} \quad ; \quad T_{er}^0 = \mu_{RT} - \frac{a}{\xi^0} \tag{15}$$

here μ_{RT} is the mean and σ_{RT}^2 is the variance of observed RTs. These expressions are obtained by assuming a specific condition in the model where $\eta = 0, S_t = 0$ (eq. (8)). The starting values of η and S_t can be chosen as a suitable fraction of ξ^0 and T_{er}^0 respectively based on domain knowledge. We refer to the proposed estimator as MLE based estimator.

4. COMPARISON

4.1 Methods

The proposed simulation method which is a combination of inverse Gaussian and random walk simulation (from here on referred to as mixed simulation method) is compared with previously proposed method by Ratcliff *et al.* [10] relying solely on random walk approximation (from here on referred to as random walk simulation). Since, the modified simulation proposed partly relies on exact sampling method it is guaranteed to be more accurate than previously proposed method based on random walk approximation only. The estimators are compared both in terms of speed and efficiency. Both estimators are checked for any bias in estimation. To carry out the analysis we used six standard parameter sets as shown in Table 1.

Table 1. Standard parameter sets

Set	a	$T_{er}(ms)$	$S_t(ms)$	ξ	η
1	0.09	150	30	1.1	0.20
2		170	40	1.0	0.25
3		190	50	0.9	0.30
4		210	60	0.8	0.35
5		230	70	0.7	0.40
6		250	80	0.6	0.45

For simulation 20,000 RT samples were generated for each set using both simulation methods. For the estimation analysis we chose all possible combinations of drift and variability in drift (ξ and η) for each set from the table. This allows 36 combinations of ξ and η for each set in Table 1, i.e. a total of 216 estimations. The values for table I are arrived at based on observed estimated parameter values for 135 subjects on a Psychomotor Vigilance Task (PVT) before and after sleep deprivation. The PVT [18] is a one-choice RT task that normally runs for 10 minutes, and the subject is shown a display terminal on which a light appears randomly and a RT counter starts running to which the subject has to respond as soon as possible via a hand held switch. The interval after which a visual stimulus is shown is known as Inter-Stimulus Interval (ISI). In the standard PVT, ISI is distributed uniformly between 2 to 10 seconds. A 10 minute PVT typically results in ≈ 95 responses. Set 6 represents a poorly performing subject while set 1 represents a very fast subject. To make the estimation process realistic, the RTs are generated by simulating a PVT. The proposed Ratcliff diffusion model simulator can be very easily adapted to simulate a PVT. Similar to a real PVT generated RTs are censored at $t = 10\ sec$. Three sets of estimation are carried to represent 10 minute, 20 minute and 30 minute PVT session. The analysis is carried out on a 3GHz Intel Core 2 Quad CPU with 4 GB of RAM. The algorithms are implemented in Matlab R2010b and are vectorized for performance.

4.2 Comparison of simulation methods

For the simulation, the step size for random walk approximation is fixed at $\tau = 0.5\ ms$. The simulations were censored at $t = 30\ sec$. Fig. 3 shows the comparison between the two simulators for the 6 sets.

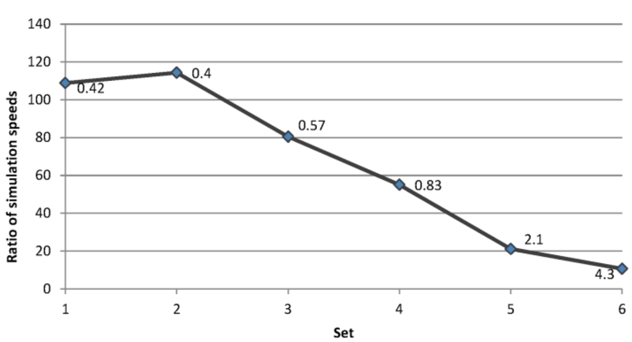

Figure 3: Ratio of simulation speeds of random walk simulation to mixed simulation method. The numbers shown are actual simulation speed of the proposed mixed simulation method in seconds.

For higher values of drift ξ and lower values of drift variability η (set 1) the proportions of negative drifts in the simulation is very small. In such situations the proposed simulator is two orders of magnitude faster than relying solely on random walk approximation. Even for parameters with low value of drift and high value of variability (set 6) where the proportion of negative drift in the simulation is relatively high the proposed method is still an order of magnitude faster.

4.3 Comparison of estimation methods

For carrying out the χ^2 based estimator as proposed by Ratcliff *et al.* we used the modified simulator. This is because the random walk approximation based simulator was prohibitively slow for carrying out estimations ($\approx 25\ min/estimation$). The estimation speed for both is shown in Table 2. Table 3 summarizes the mean and standard deviation of error in estimation. Where error in estimation e is defined as $e = \Theta - \widehat{\Theta}$. The estimates are unbiased based on a one sample t-test at significance level $\alpha = 0.05$. All parameters are scaled for a normalized boundary value $a = 0.09$.

Table 2: Estimation speed

Set	Total estimation time (for 216 estimations)	
	χ^2 based estimator	MLE based estimator
10 Min. PVT	12 Hours	12 Minute
20 Min. PVT	12 Hours	11 Minute
30 Min. PVT	10 Hours	10 Minute

Table 3: Estimation Efficiency

Set	Parameter	χ^2 based estimator	MLE based estimator
10 Min. PVT	T_{er}(ms)	-1.17 ± 10.8	-0.8 ± 15
	S_t (ms)	-1.2 ± 18.8	0.9 ± 15.5
	ξ	-0.009 ± 0.128	-0.003 ± 0.142
	η	-0.005 ± 0.089	-0.003 ± 0.093
20 Min. PVT	T_{er}(ms)	-0.94 ± 9.8	1.4 ± 14
	S_t (ms)	-1.2 ± 18	0.9 ± 13.4
	ξ	-0.008 ± 0.102	-0.001 ± 0.124
	η	-0.006 ± 0.085	-0.008 ± 0.065
30 Min. PVT	T_{er}(ms)	-0.25 ± 8.8	0.33 ± 13.5
	S_t (ms)	1.4 ± 18.5	1.0 ± 11.9
	ξ	-0.007 ± 0.091	0.0002 ± 0.09
	η	-0.006 ± 0.08	0.003 ± 0.06

The efficiency of both the estimators is similar, although the proposed simulator is an order of magnitude faster and much simpler to implement.

5. CONCLUSIONS

The one choice Ratcliff diffusion model is a powerful model of perceptual decision making. The availability of simple simulation and estimation methods is essential for the clinical community to fully leverage the model for real world applications. We brought to notice the simple observation that the simulation time can be considerably improved by switching between sampling from inverse-Gaussian distribution for positive drifts and falling back on random walk simulation for negative or zero drifts. The novelty of the paper lies in deriving a closed form solution for the model which did not exist earlier in literature. The closed form solution for the model can help gain insights into one choice perceptual decision making and help design more efficient and fast estimators. We used the maximum likelihood estimator, which is a well-established method for parameter estimation to show that the estimated parameters based on the close form solution are in close agreement with previous method based on data fitting. The estimation speed of the MLE based estimator is at-least an order of magnitude faster while maintaining the same level of efficiency compared to previously proposed method. The performance of the estimator is not tested in the presence of noise and contaminant reaction times (see [19] for details). Therefore, when applied to clinical data we recommend using the proposed closed form solution based MLE in conjunction with χ^2 based estimator to fine tune the estimated parameters. This will allow for both quick and reliable parameter estimates.

6. ACKNOWLEDGMENTS

We would like to thank Dr. Michael Chee and other members of Cognitive Neuroscience Laboratory – Duke NUS graduate medical School, Singapore for providing the data used to estimate model parameters based on which the analysis were carried out as well as for comments and suggestions.

7. REFERENCES

[1] R. Ratcliff, "A theory of memory retrieval," Psychological review, vol. 85, p. 59, 1978.

[2] A. Voss, K. Rothermund, and J. Voss, "Interpreting the parameters of the diffusion model: An empirical validation," Memory & Cognition, vol. 32, pp. 1206-1220, 2004.

[3] R. Ratcliff, "A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data," Psychonomic Bulletin & Review, vol. 9, pp. 278-291, 2002.

[4] H. R. Heekeren, S. Marrett, and L. G. Ungerleider, "The neural systems that mediate human perceptual decision making," Nature Reviews Neuroscience, vol. 9, pp. 467-479, 2008.

[5] D. Vickers, "Evidence for an accumulator model of psychophysical discrimination," Ergonomics, vol. 13, pp. 37-58, 1970.

[6] E. J. Ploran, S. M. Nelson, K. Velanova, D. I. Donaldson, S. E. Petersen, and M. E. Wheeler, "Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI," The Journal of Neuroscience, vol. 27, pp. 11912-11924, 2007.

[7] M. K. van Vugt, P. Simen, L. E. Nystrom, P. Holmes, and J. D. Cohen, "EEG oscillations reveal neural correlates of evidence accumulation," Frontiers in Neuroscience, vol. 6, 2012.

[8] R. Ratcliff and J. N. Rouder, "Modeling response times for two-choice decisions," Psychological Science, vol. 9, pp. 347-356, 1998.

[9] J. Vandekerckhove and F. Tuerlinckx, "Fitting the Ratcliff diffusion model to experimental data," Psychonomic Bulletin & Review, vol. 14, pp. 1011-1026, 2007.

[10] R. Ratcliff and H. P. Van Dongen, "Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation," Proceedings of the National Academy of Sciences, vol. 108, pp. 11285-11290, 2011.

[11] J. I. Gold and M. N. Shadlen, "The neural basis of decision making," Annu. Rev. Neurosci., vol. 30, pp. 535-574, 2007.

[12] D. D. R. Cox and H. D. Miller, The theory of stochastic processes vol. 134: CRC Press, 1977.

[13] C. J. Geyer, Markov chain Monte Carlo maximum likelihood: Defense Technical Information Center, 1992.

[14] J. A. Nelder and R. Mead, "A simplex method for function minimization," Comput J, vol. 7, pp. 308-313, 1965.

[15] J. Folks and R. Chhikara, "The inverse Gaussian distribution and its statistical application--a review," Journal of the Royal Statistical Society. Series B (Methodological), pp. 263-289, 1978.

[16] J. R. Michael, W. R. Schucany, and R. W. Haas, "Generating random variates using transformations with multiple roots," The American Statistician, vol. 30, pp. 88-90, 1976.

[17] G. B. Dantzig, Linear programming and extensions: Princeton university press, 1998.

[18] D. F. Dinges and J. W. Powell, "Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations," Behavior Research Methods, Instruments, & Computers, vol. 17, pp. 652-655, 1985.

[19] Ratcliff, Roger, and Francis Tuerlinckx. "Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability." Psychonomic bulletin & review, vol. 9, pp. 438-481, 2002.