Generalising the Drift Rate Distribution for Linear Ballistic Accumulators

Andrew Terry[1], A.A.J. Marley[2], Avinash Barnwal[3], E.-J. Wagenmakers[3], Andrew Heathcote[1], Scott D. Brown[1]

1. School of Psychology University of Newcastle Callaghan NSW 2308 Australia

2. Department of Psychology University of Victoria British Columbia Canada and *Institute for Choice, University of South Australia.*

3. Department of Psychology University van Amsterdam The Netherlands

Abstract

The linear ballistic accumulator model is a theory of decision-making that has been used to analyze data from human and animal experiments. It represents decisions as a race between independent evidence accumulators, and has proven successful in a form assuming a normal distribution for accumulation ("drift") rates. However, this assumption has some limitations, including the corollary that some decision times are negative or undefined. We show that various drift rate distributions with strictly positive support can be substituted for the normal distribution without loss of analytic tractability, provided the candidate distribution has a closed-form expression for its mean when truncated to a closed interval. We illustrate the approach by developing three new linear ballistic accumulation variants, in which the normal distribution for drift rates is replaced by either the lognormal, Fréchet, or gamma distribution. We compare some properties of these new variants to the original normal-rate model.

Keywords: response time; evidence accumulation; accumulator model; decision making

The linear ballistic accumulator model (LBA: Brown and Heathcote, 2008) is an evidence accumulation model for simple decision-making, which has been applied to a wide range of data from human and animal experiments. The LBA assumes that decisions are made by separate independent accumulators, each of which gathers evidence in favour of a different choice outcome, with the first accumulator to reach a threshold deciding the response. Figure 1 illustrates a typical LBA accumulator, with a decision threshold (dotted line) and an accumulation process (rising arrow). Figure 1 also shows the simplicity of the LBA model, with constant linear accumulation, and allowing just two sources of variability. The shaded rectangle indicates random variability in the starting point of the evidence accumulation process, and the bell curve indicates random variability in the rate of evidence accumulation. Both of these sources of randomness operate independently from decision to decision, and independently between accumulators corresponding to different choices. Using just these two sources of variability, the LBA model accounts for the variability observed in decision-making data across a wide range of experimental paradigms.

Following similar assumptions for the diffusion model (Ratcliff and Rouder, 1998), the LBA model has a uniform distribution for the starting points of evidence accumulation, and a normal distribution for the speed of accumulation ("drift rates"). These assumptions allowed the development of simple, closed-form expressions for both the probability density function (PDF) and the cumulative distribution function (CDF) for the finishing times of the accumulation process (Brown and Heathcote, 2008). This mathematical tractability is an important feature of the LBA model. For example, it makes efficient estimation easy using a wide variety of optimisation techniques and statistical approaches (Donkin et al., 2009a; Turner et al., 2013).

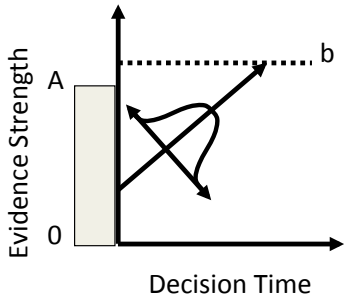

Figure 1: Schematic LBA accumulator. As decision time grows (abscissa) evidence is accumulated (ordinate), with an example accumulation trajectory shown by the rising arrow. Typically, several independent accumulators would race in parallel, representing the different choices, with the decision triggered by the first to reach threshold (dotted line). Variability in decision-making is modelled by randomness in the start point of evidence accumulation and in the accumulation rate. These are conventionally assumed to follow uniform and normal distributions, respectively.

The assumption of a normal distribution for drift rates in the LBA model means that on some trials it is possible that the sampled drift rates for all accumulators will be negative, so a decision is never made. Brown and Heathcote (2008) found that such cases were extremely rare in practice when fitting to a range of empirical

data sets. However, here we address this potential problem by developing a general mathematical approach that maintains the mathematical tractability of the LBA model while allowing for various, strictly positive, drift rate distributions.

In the following, we first develop a general approach for working with a class of drift rate distributions in the LBA model. We then illustrate this method using three distributions from this class. This is followed by an investigation of the ability of the three new LBA model variants, as well as a variant of the original LBA, to account for seminal data reported by Wagenmakers et al. (2008). We also investigate similarities between the LBA model variants, by fitting each variant to synthetic data generated by the other variants. The three new variants all assume drift rate distributions which are strictly positive. Therefore, to make the comparison more precise, rather than compare against the original LBA (with normally distributed drift rates) we employ a slight modification: we assume that the normal distribution of drift rates is truncated to positive-only values. To keep this clear, we refer to this variant as the "truncated normal LBA". The truncated normal LBA has simple analytic solutions, and has been shown to be almost identical, in practice, to the conventional LBA (e.g., Heathcote and Love, 2012).

Analytical derivation of PDF for arbitrary drift rate distribution

For the purposes of a very wide variety of applications, it is sufficient to know the density and cumulative distribution functions (PDF and CDF, respectively) for the finishing times of a single linear ballistic accumulator. For example, with these two expressions, the joint density over response time and choice can be written via standard independent-race equations, for a large range of decision models. These models include simple races for N-alternative forced choice, as well as more complex architectures involving logical AND and OR stopping rules (Brown and Heathcote, 2008; Eidels et al., 2010).

Brown and Heathcote (2008) derived the CDF for the linear ballistic accumulator model with normally-distributed drift rates by working directly with the expression for the normal distribution's density function. At one point in their analysis, one of the terms in the expression for the CDF is related to a truncated mean of the drift rate distribution. It is this observation that motivates our current work, and allows the development of a more general approach.

Consider a single linear ballistic accumulator, with uniformly distributed starting points across trials. Our approach to the problem involves conditioning on a particular sample of the drift rate, which we call u. Conditioning this way allows for calculation of the finishing time in the obvious manner. Of course, these rates cannot be observed in practice, so we then integrate over the distribution of start points, to remove the conditionality.

Suppose, without loss of generality, that the uniform distribution of start points is on the interval $[0, A]$, and that the response threshold is at $b \geq A$; therefore, the distribution of distances from starting point to threshold is also uniform, on the interval $[b - A, b]$. Suppose also that drift rates are distributed across trials according to a strictly positive distribution with density g and cumulative distribution function G. Let $\boldsymbol{P}$ be the random variable

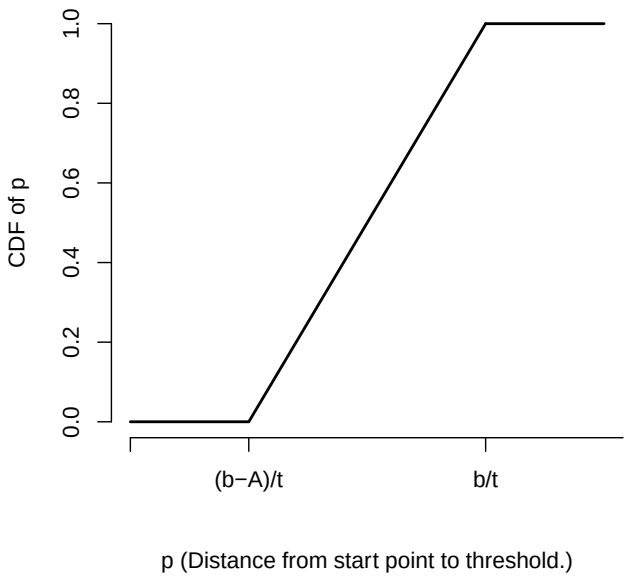

Figure 2: Cumulative distribution function (CDF) of the distance, p, between the starting point of evidence accumulation and the threshold, which makes clear the reason for the three-branch structure of the equation for the CDF for finishing times. The distribution is uniform on the interval $[\frac{b-A}{t}, \frac{A}{t}]$, so its CDF is: zero for values smaller than $\frac{b-A}{t}$; one for values larger than $\frac{A}{t}$; and increases linearly between those points.

representing the distance to threshold on some trial, with a uniform distribution on the interval $[b - A, b]$, and let $\boldsymbol{U}$ be the random variable for the drift rate, with distribution G and density g. The time to reach threshold is then simply distance to be traveled (i.e. the sample from $\boldsymbol{P}$) divided by rate of travel (i.e. the sample from $\boldsymbol{U}$). Thus, the cumulative distribution function for finishing times of this accumulator at time t, say $F(t)$, is given by:

$$\begin{aligned} F(t) &= prob(\frac{\boldsymbol{P}}{\boldsymbol{U}} \leq t) \\ &= prob(\boldsymbol{P} \leq \boldsymbol{U}t) \end{aligned}$$

To obtain the probabilities associated with the random variable $\boldsymbol{U}$, we integrate over samples from this distribution, say u, with respect to the density function, $g(u)$, which has support on the positive real line:

$$F(t) = \int_0^\infty prob[\boldsymbol{P} \leq ut]g(u)du \tag{1}$$

Since $\boldsymbol{P}$ has a uniform distribution, it has a three-piece linear CDF, as shown in Figure 2. The CDF gives the probability required in Equation 1, by expansion into three terms: it is zero whenever $u < \frac{b-A}{t}$; it is linear in u, whenever $\frac{b-A}{t} < u < \frac{b}{t}$; and it is one whenever $u > \frac{b}{t}$. The first of these three branches is zero, and so is

dropped below. The second linear branch can be expanded into a term with integrand $g(u)$ and another term with integrand $ug(u)$. The third branch also gives a term with integrand $g(u)$. Together, this gives:

$$F(t) = \frac{A-b}{A}\int_{\frac{b-A}{t}}^{\frac{b}{t}} g(u)du + \frac{t}{A}\int_{\frac{b-A}{t}}^{\frac{b}{t}} ug(u)du + \int_{\frac{b}{t}}^{\infty} g(u)du.$$

The terms with integrands of $g(u)$ require only evaluation of G, by definition of the cumulative distribution function. However, the term with integrand $ug(u)$ is related to the mean of the drift rate distribution when truncated to the interval represented by that term's limits of integration. For presentation reasons, it is helpful to add the normalizing constant corresponding to the mass of the distribution in the interval $[\frac{b-A}{t}, \frac{A}{t}]$. With the normalizing constant included, and replacing the integrals in the other terms with the distribution function for G:

$$F(t) = \frac{A-b}{A}\left[G\left(\frac{b}{t}\right) - G\left(\frac{b-A}{t}\right)\right] + \frac{t\left[G\left(\frac{b}{t}\right) - G\left(\frac{b-A}{t}\right)\right]}{A}\int_{\frac{b-A}{t}}^{\frac{b}{t}} \frac{ug(u)du}{G(\frac{b}{t}) - G(\frac{b-A}{t})} + 1 - G\left(\frac{b}{t}\right)$$

Let $Z(t)$ represent the mean of the distribution g after truncation to the above interval, that is:

$$Z(t) = \frac{1}{G(\frac{b}{t}) - G(\frac{b-A}{t})}\int_{\frac{b-A}{t}}^{\frac{b}{t}} ug(u)du$$

Then, after some simplification:

$$F(t) = 1 + \left(\frac{tZ(t)-b}{A}\right)G\left(\frac{b}{t}\right) + \left(\frac{b-A-tZ(t)}{A}\right)G\left(\frac{b-A}{t}\right) \quad (2)$$

The density function for the finishing times of this linear ballistic accumulator is found by differentiation of Equation 2 with respect to t. This requires $Z(t)$ to be differentiable at all $t > 0$; we denote its derivative at t by $Z'(t)$. Then recalling that, since G is a CDF, $\frac{d}{du}G(u) = g(u)$, this gives:

$$f(t) = \left(\frac{Z(t)+tZ'(t)}{A}\right)\left[G\left(\frac{b}{t}\right) - G\left(\frac{b-A}{t}\right)\right] + \left(\frac{tZ(t)-b}{A}\right)g\left(\frac{b}{t}\right) + \left(\frac{b-A-tZ(t)}{A}\right)g\left(\frac{b-A}{t}\right). \quad (3)$$

Equations 2 and 3 can be used to provide closed-form expressions for the CDF and PDF for a linear ballistic accumulator model with any strictly positive distribution for drift rates, provided that the drift rate distribution has a closed form expression for its truncated mean ($Z(t)$, above), and that this expression can be easily differentiated with respect to t ($Z'(t)$ above).

Three New Example Drift Rate Distributions

We illustrate the above method with three new examples of strictly positive drift rate distributions: the gamma, Fréchet and lognormal distributions. The question of which is the best distribution to use for drift rates is very complex, and beyond the scope of this work. Nevertheless, we provide some discussion of the different considerations in this debate, in the concluding sections.

Gamma Distributed Drift Rates

The gamma distribution is an interesting case partly because it can approximate the normal distribution under some parameter settings. Given the success of the traditional LBA model (with normally distributed drift rates) in fitting data, this suggests that a gamma-LBA model might be similarly successful.

Suppose that drift rates follow a gamma distribution with parameters α and β for shape and scale, respectively. Then $Z_\Gamma(t)$ from Equations 2 and 3 is the mean of a gamma distribution restricted to the interval $[\frac{b-A}{t}, \frac{b}{t}]$. Coffey and Muller (2000) provide expressions for this mean, which lead to:

$$Z_\Gamma(t) = \frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)}\left[\frac{G\left(\frac{b}{t};\alpha+1,\beta\right)-G\left(\frac{b-A}{t};\alpha+1,\beta\right)}{G\left(\frac{b}{t};\alpha,\beta\right)-G\left(\frac{b-A}{t};\alpha,\beta\right)}\right] \quad (4)$$

Here, $G(x;a,b)$ represents the cumulative distribution function of the gamma distribution evaluated at x, with shape parameter a and scale parameter b (see Appendix A). Throughout, we use $\Gamma(x)$ to represent the gamma function and $\Gamma(x,a)$ to represent its generalisation to the lower incomplete gamma function (also both specified in Appendix A). The derivative with respect to time of Equation 4, $Z'_\Gamma(t)$, is easy to calculate but also cumbersome - see Appendix A.

Fréchet Distributed Drift Rates

Let drift rates be distributed according to a Fréchet distribution with scale and shape parameters μ and α, respectively, and let $Fr(x;.\alpha,\mu)$ represent the corresponding cumulative distribution function. Then Nadarajah (2009) provides an expression for the truncated mean, which leads to:

$$Z_{Fr}(t) = \left[\frac{\Gamma\left(1-\frac{1}{\alpha},(\frac{\mu b}{t})^{-\alpha}\right)-\Gamma\left(1-\frac{1}{\alpha},(\frac{\mu(b-A)}{t})^{-\alpha}\right)}{\mu\left(Fr\left(\frac{b}{t};\mu,\alpha\right)-Fr\left(\frac{b-A}{t};\mu,\alpha\right)\right)}\right] \quad (5)$$

Once again, see Appendix A for the derivative with respect to time.

An interesting consequence of assuming a Fréchet distribution for drift rates is that, in the absence of trial-to-trial variability in the starting point of evidence accumulation (i.e., with parameter $A = 0$), Fréchet-distributed

drift rates give rise to Gumbel distributions for time-to-threshold. This provides interesting links to models used in the study of discrete choice in various applied areas (Hawkins et al., 2014; Colonius and Marley, 2014).

Lognormal LBA

Heathcote and Love (2012) investigated a simplified linear ballistic accumulator model, the Lognormal Race Model, where both the drift rate and the start point to threshold distributions were lognormal, and so the distribution of threshold-crossing times for an accumulator is also lognormal. Here, we outline the case where the starting point distribution remains uniform and the threshold a constant, but the drift rates are distributed lognormally with parameters μ and σ for the underlying normal distribution. Bebu and Mathew (2009) provide expressions for the moments of the truncated lognormal distribution. The truncated mean takes the form:

$$Z_{LN}(t) \quad = \quad \exp\left(\mu + \frac{\sigma^2}{2}\right)\left[\frac{\Phi\left(\log(\frac{b}{t}); \mu + \sigma^2, \sigma\right) - \Phi\left(\log(\frac{b-A}{t}); \mu + \sigma^2, \sigma\right)}{\Phi\left(\log(\frac{b}{t}); \mu, \sigma\right) - \Phi\left(\log(\frac{b-A}{t}); \mu, \sigma\right)}\right] \tag{6}$$

Here, $\Phi(x; \mu, \sigma)$ is the normal cumulative density function evaluated at x with mean μ and standard deviation σ. The derivative with respect to time is given in the Appendix A.

Similarities and Differences Between Four LBA Variants

The three new LBA variants developed above have quite different properties for their drift rate distributions. The new variants have positively skewed drift rate distributions, and those distributions also differ in how quickly their right tails approach zero. We investigated the relationships between these new variants, and truncated normal LBA, in a model recovery exercise. This exercise involved generating synthetic data from each of the three new variants and from the truncated normal LBA, and then fitting those data with each of the four LBA variants in turn. The results revealed both similarities and differences among variants.

To determine the values to use when generating data, we began with parameter settings for the truncated normal LBA which were typical of parameters estimated from standard psychophysical experiments. These parameters are given in the column headed *Truncated Normal* in Table 1. To determine reasonable parameter settings for the three new LBA variants, we generated a very large sample of synthetic data from the truncated normal LBA using the parameters in Table 1, and found the maximum-likelihood estimates for the parameters of each of the three new model variants by fitting those data. These parameters were rounded off, and are reported in the left three columns of Table 1. There were a total of eight free parameters for each LBA variant, reflecting an experiment investigating two-alternative forced choices in two conditions, one emphasising decision speed, the other emphasising decision accuracy (as in, for example: Forstmann et al., 2008; Ratcliff and Rouder, 1998). These parameters were:

- The width of the starting-point distribution (A).
- Non-decision time (t_0), separately for the speed-emphasis condition and the accuracy-emphasis condition.
- The distance between the upper bound of the starting point distribution and the response threshold ($b - A$), separately for the speed-emphasis and accuracy emphasis conditions.
- Two drift rate distribution location parameters, for the distributions of drift rates in the accumulators corresponding to the correct and incorrect responses (v_c and v_e, respectively).
- A drift rate distribution scale parameter (s_c) only for the distribution of drift rates in the accumulator corresponding to the correct response.

The parameter corresponding to s_c in the accumulator representing the incorrect response was fixed arbitrarily at $s = 1$, to satisfy a scaling property of the models (Donkin et al., 2009b). Namely, if the parameters of the drift rate distribution are adjusted such that the drift rate distribution is "scaled" by a constant factor, then the other parameters of the model can be adjusted accordingly to compensate. This situation leaves the model making identical predictions from very different parameters, which can be problematic in data analysis. For example, in the conventional LBA model with normally distributed drift rates, suppose the mean and standard deviation of the drift rate distribution were both doubled (which has the effect of doubling all predicted drift rates). Then, if the start point distribution (A) and decision threshold (b) are also both doubled, the predictions of the model are unchanged. This occurs for the obvious reason: the rate growth in the accumulator has doubled, but the distance to travel has also doubled, so the finishing time is the same. This indeterminacy is avoided by arbitrarily fixing one of the parameters.

In our analyses of the truncated normal LBA model, we fixed the standard deviation of the drift rate distribution for the accumulator corresponding to the incorrect response to 1.0. We took a similar approach, fixing the rate distribution's scale parameter to 1.0, for both the lognormal and Fréchet variants. Because of the multiplicative properties of the lognormal distribution, changes to the mean of the underlying normal distribution (μ) result in "scaling" of the drift rate distribution, so we fixed this parameter to 1.0 for the accumulator corresponding to the incorrect response. The Fréchet distribution has a similar property attached to its scale parameter (also called μ), so we also fixed this parameter to 1.0 for the accumulator corresponding to the incorrect response. We took a different approach for the gamma distribution, where we fixed the shape parameter of the drift rate distribution in the accumulator corresponding to the correct response to 1.0. This was because preliminary explorations found that the gamma scale parameter had very similar effects to the normal mean parameter, whereas the gamma shape parameter had very similar effects to the normal scale parameter.

For each model variant, we generated 20,000 synthetic decisions using the above parameters. Our use of such a large sample was intended to eliminate variability due to sampling error, allowing more precise characterization of the differences between the models' predictions due to their different drift rate distributions alone.

Table 1: Parameters used to generate synthetic data. Note that the same t_0 parameters were used for all distributions. The v parameter rows corresponds to the normal mean, gamma scale (*beta*) , Fréchet shape (α) and lognormal standard deviation (*sigma*) parameters, and the s parameter row corresponds to the normal standard deviation, gamma shape (*alpha*), Fréchet scale (*mu*) and lognormal mean (*mu*) parameters.

	Drift Rate Distribution			
Parameter	Truncated Normal	Fréchet	Gamma	Lognormal
t_0 (speed)		0.15		
t_0 (accuracy)		0.20		
$b - A$ (speed)	0.20	0.57	0.05	1.20
$b - A$ (accuracy)	1.00	1.14	1.00	2.70
v_c	3.00	1.86	4.20	1.60
v_e	1.50	2.50	2.50	0.50
s_c	0.75	2.25	0.70	0.60
A	3.00	1.00	6.00	2.80

We fit the synthetic data set generated by each of the four LBA variants with all of the LBA variants. We used maximum-likelihood estimation with algorithms described in detail by Donkin et al. (2009a). Figure 3 shows cumulative distribution functions jointly over correct and incorrect responses (black and grey, respectively) for both speed-emphasis and accuracy-emphasis conditions (left and right pairs of curves in each panel). Columns and rows indicate the LBA model variant used to generate and fit the synthetic data, respectively. To illustrate with an example, the lower-left panel shows the results when synthetic data were generated from an LBA model variant where drift rates are sampled from a normal distribution truncated to positive values, and then fit using an LBA model variant with a lognormal distribution for drift rates. The left-most set of black circles in that panel show the $10^{th}, 30^{th}, 50^{th}, 70^{th}$, and 90^{th} percentiles of the correct responses from the speed-emphasis condition of the synthetic data, averaged over participants. These percentiles are plotted against the probability of jointly observing a correct response in data from that condition, and the associated response time falling in the bottom $10^{th}, 30^{th}, 50^{th}, 70^{th}$, and 90^{th} of the data from that condition. The solid grey and black lines overlaid illustrate the same quantities calculated from the maximum-likelihood parameter estimates of the fitting model.

The panels along the main diagonal of Figure 3 show almost perfect agreement between the synthetic data (symbols) and posterior predictive data (lines). This is to be expected, showing that, as for the traditional LBA, in large samples maximum-likelihood estimation is able to recover parameters values for the variants. We investigate the accuracy of the recovered parameters in these self-fits below. The off-diagonal panels illustrate that the LBA model does not appear to suffer from undue flexibility; rather, its predictions appear quite tightly constrained. This is apparent in the inability of some model variants to enable a close fit to synthetic data generated by a different variant. For example, the truncated normal LBA provided a quite poor fit to data generated from the lognormal LBA and by the Fréchet LBA.

An important property of the conventional LBA model is its ability to support accurate parameter estimation. We tested the new LBA variants on this ability, by examining the four cases where the same LBA model variant was used to both generate and fit synthetic data – i.e., the cases shown in the main diagonal of Figure 3. Table 2 shows the absolute difference between the data-generating parameter values and their maximum-likelihood

estimates, expressed as percentages. It is clear that the three new LBA model variants developed above all support excellent model recovery, at least for these parameter values and in large samples.

Table 2: Absolute bias in recovered parameters (in percent, rounded to nearest integer).

	Drift Rate Distribution			
Parameter	Truncated Normal	Fréchet	Gamma	Lognormal
t_0 (speed)	1	1	0	1
t_0 (accuracy)	4	1	2	2
$b - A$ (speed)	2	0	2	3
$b - A$ (accuracy)	3	2	5	1
v_c	0	0	2	0
v_e	1	1	2	1
s_c	0	0	2	0
A	0	1	1	2

Fits to Real Data

A final test for the three new LBA model variants was to account for real data. Our aim here was not to falsify any particular variant, because that decision is probably not best made on the basis of a single data set. Rather, we aimed to establish whether the new LBA model variants were capable of fitting real data to approximately the same degree as the conventional LBA, thus validating their suitability for future investigation. For testing, we used data from a lexical decision task reported by, Wagenmakers et al. (2008) (their Experiment 2), in which eight participants each classified 1,920 letter strings as either valid or invalid words (words vs. non-words). There were two within-subject manipulations of interest. Firstly, half of the blocks of trials contained three times as many word as non-word stimuli, with the other half of blocks having three times as many non-word as word stimuli. Secondly, the type of word stimuli varied randomly from trial to trial in three classes – high frequency (common) words, low frequency (uncommon) words, and very low frequency (very rare) words. Higher frequency words are easier for participants to classify correctly, leading to higher accuracy rates and shorter response times. Wagenmakers et al. used this experiment to investigate criterion setting using the diffusion model, which would be expected to be selectively influenced by the proportion of words and non-words in each block.

We fit each model to these data using maximum likelihood estimation following the methods outlined by Donkin et al. (2009a). Heathcote and Love (2012) used the same methods to compare their Lognormal Race Model to the truncated normal LBA, and to the conventional LBA. They found the conventional and truncated normal LBA models fit about equally well to data from the first experiment reported by Wagenmakers et al. (2008), and both fit a little better than the Lognormal Race Model. Here we also used the truncated normal LBA – which once again fit about as well as the conventional LBA – so that none of the four models being compared allowed undefined response times.

For each participant, we estimated a single non-decision time parameter (t_0) and assumed a selective influence of the stimulus manipulation (i.e., non-words and the various types of words) on only the two rate

parameters. However, we otherwise used a quite flexible parameterisation for the models in order to see if they could capture fine details in the data, such as small but theoretically important effects on the relative speeds of correct and incorrect responses. Starting point variability (A) and the threshold (b) parameters could differ both between word and non-word accumulators and between the proportion conditions, allowing for differences in response bias. Similarly, we let both rate parameters vary with proportion condition and with accumulator (i.e., a different values for the accumulator that matches the stimulus and the one that mismatches).

We explored two solutions to the scaling issue: fixing either the scale parameter of the rate distributions or the mean threshold parameter, in both cases for just one accumulator in one condition. Differences in estimation performance (i.e., larger maximised likelihood values and less variable parameter estimates) reflected the pairings found in the last section; overall there was slightly better performance in terms of speed of convergence when fixing the scale parameter for the truncated normal and gamma variants, and slightly better performance for the lognormal and Fréchet variants when fixing the threshold parameter. We report goodness of fit results for each variant based on the best overall scaling solution for that variant. The best overall fit, as quantified by maximised log-likelihood values summed over participants (L), was given by the gamma (L = 12,710) followed by the positive LBA (L = 12,511) and lognormal (L = 12,341), with the Fréchet noticeably worse (L = 10,373).

Figure 4 shows that all four model variants provided a good account of the lexical decision accuracy, with all predicted values falling within the 95% confidence intervals except for one case each for the truncated normal and gamma models. Figure 5 quantifies the description of RT distribution for correct responses by displaying estimates of the middle of the distribution (i.e., the median RT or 50^{th} percentile) and its fast (10^{th} percentile) and slow (90^{th} percentile) tails. Again the fit is quite good, with no predicted values falling outside the confidence intervals except for the Fréchet 10^{th} percentile, which is underestimated for all word stimuli in the 25%-word condition and for non-words in the 75%-word condition.

Figure 6 enables comparison of correct and error RT, displaying the median for each. The fits to error RT are noticeably worse than those to correct RT, which is largely attributable the relative infrequency of error responses, particularly for high-frequency words, and for non-words in the 25% condition. The attendant variability is reflected in large 95% confidence intervals for these points, and overall all models have at most one or two predicted points falling outside the 95% confidence intervals for errors. Importantly, all models capture the general "crossover" pattern, caused by errors being faster than correct responses for the rarer stimulus type (i.e., words in the 25% condition and non-words in the 75% condition) and slower than correct responses for the more common stimulus type (i.e., non-words in the 25% condition and words in the 75% condition), with the worst quantitative accounts provided by the truncated normal and Fréchet variants for non-words with 75% words. As noted by Wagenmakers et al. (2004), the crossover pattern follows from the geometry of the evidence accumulation process. For example, in the 75% word condition, where there is a bias towards word responses, and hence a smaller distance to the word than non-word threshold at the start of accumulation, correct "word" responses (which terminate on

the closer word boundary) are faster than incorrect "word" responses (which terminate on the more distant nonword boundary). This reverses in the 25% word condition where the bias is towards non-words.

Discussion

The linear ballistic accumulator model of Brown and Heathcote (2008) assumes that simple decisions are made by evidence accumulators racing towards a threshold. The LBA model makes a key simplifying assumption, that the accumulation of evidence is linear and deterministic. This simplification allows for simple, closed-form, expressions for the probability density and cumulative distribution functions of the time taken to reach threshold.

Instead of allowing variability in the evidence accumulation process within a trial, as assumed – with only a few exceptions (e.g., Grice, 1972) – by earlier evidence accumulation models, the LBA model assumes only decision-to-decision variability in the rate of evidence accumulation and in the amount of evidence accumulated prior to stimulus onset. In the LBA model the distribution assumed for drift rates was normal, so that some decisions could sample negative drift rates, potentially leading to indeterminate finishing times for evidence accumulation. Whether or not this is a conceptual problem for the model probably depends on taste, and in practice it has not proven problematic, but, nevertheless, it is worth addressing. Here, we outlined a more general mathematical treatment for replacing the normal distribution of drift rates with strictly positive distributions, while still providing closed-form expressions for the density and cumulative distribution functions. Our approach requires only that the candidate drift rate distribution itself has closed-form expressions for its density and distribution functions, as well as a differentiable expression for its mean when truncated to a closed interval.

We illustrated this approach using three new candidate distributions for drift rates: the Fréchet, gamma and lognormal distributions. These distributions, along with the truncated normal LBA, were all used to generate synthetic data. The four distributions lead to quite different distributions of drift rates, as shown in the left column of Figure 7. Even so, the strict constraints imposed by the structure of the LBA model means that the RT distributions predicted by these drift rates are broadly similar. For example, the CDFs from the four LBA variants (second-to-right column of Figure 7) are difficult to tell apart. Even the hazard functions (right column) are only qualitatively different in the tails, where the Fréchet distribution's hazard function does not decrease. We do not have great confidence that differences in the hazard functions could help with empirically distinguishing the model variants, for two reasons. Firstly, informal investigation of the parameter space revealed that the qualitative patterns of hazard functions in Figure 7 was not invariant. For some parameter settings, for example, the hazard function of the Fréchet LBA showed the same increasing-then-decreasing shape as the other model variants. Secondly, in applications, the tails of the hazard functions are calculated from a very small proportion of the data. Such calculations are inherently noisy, and have proven inconclusive in the past (e.g. Luce, 1986).

Fits of these four LBA variants showed that all models were able to account for real data (Wagenmakers et al., 2008, Experiment 2) adequately, but certainly not identically. Fits of the four models to data generated by

another of the models also showed similarities and differences among the models, probably representing both the relative flexibility of the models and perhaps differences in the tails of the distributions. For example, Figure 3 shows that the truncated normal LBA cannot adequately fit data generated by the Fréchet LBA. The lognormal LBA is so constrained that it could not adequately fit data by any of the three other variants, at least for the particular set of parameters investigated.

Clearly, further work is required to assess the generality of these findings, both with respect to real data and model mimicry [1]. In order to facilitate such work, and the broader use of the new LBA variants, we have added their CDF and PDF equations to the package *rtdists* for the open-source statistical language R (Singmann et al., 2015). This package includes help sections, instructions, and simple examples of how to use the routines. It is available for download from `https://cran.r-project.org/web/packages/rtdists/`.

Consideration of the variability assumptions in decision models is particularly apposite at the moment, given recent attention to these assumptions, and to questions about the flexibility and falsifiability of decision-making models. Jones and Dzhafarov (2014) analyzed a new set of quite different models, in which the drift rate distributions were allowed to be arbitrarily complex, and to vary arbitrarily between conditions. Unsurprisingly, these models are unfalsifiable; they can fit any pattern of data. Our results do not speak to that conclusion, because we consider only models constrained in the usual manner, with parametric assumptions on drift rates. However, our results speak against the broader implications levelled by Jones and Dzhafarov, that the variability assumptions of accumulator models make them difficult to tell apart. The results in Figure 3 demonstrate that there exist patterns of results that each LBA model variant cannot accommodate (see also Heathcote et al., in press). Even more importantly, these patterns of results are not outlandish or unrealistic, but look a lot like typical data from decision-making experiments. Secondly, our fits to real data show that the three new LBA model variants all perform adequately, which contradicts Jones and Dzhafarov's implication that the particular assumption of a normal distribution was key in the LBA model's success.

More generally, our approach illustrates a tractable way to investigate the properties conferred on the LBA model by different choices of drift rate distribution. By comparing new variants that differ only in their distributions of drift rates, the consequences of each choice of drift rate distribution can be examined. This approach runs counter to claims made by Jones and Dzhafarov (2014) regarding the inability to separate out the effects of different assumptions. In particular, Jones and Dzhafarov suggested that assumptions about the architecture of the LBA model (such as linear evidence accumulation) could not be separated from assumptions about the distribution of drift rates. Our approach provides another method for investigating exactly those comparisons.

[1] The degree of mimicry is likely to vary in different regions of parameter space. A reviewer suggested quantifying similarity between distributions using the Kullback-Liebler distance metric. We agree that this is a very useful way of summarising findings from the sort of large-scale study that would be necessary to thoroughly investigate model mimicry.

Which is the best distribution for drift rates?

Because of its broad scope, answering the question of which is the best parametric distribution to assume for drift rates, is beyond what could be achieved here. Our goal was instead to set out a method by which alternative drift rate distributions could be investigated. Deciding which distribution is best is a multi-faceted problem, because of the many different ways in which we might define "best". One way in which a distribution might be better than another is that it might lead to better fits to empirical data. This, however, requires fitting a wide variety of data sets from different experimental paradigms. The process of fitting many different data sets is the best way to validate and test a new model, enabling one to uncover unforeseen problems, and to identify situations in which the model does and does not fit well. This comparison will also require careful attention to issues of model selection. Even though all of the candidate drift rate distributions we have considered have the same number of parameters, they will certainly differ in their functional form complexity, which makes model selection difficult. Even more, they will differ in complexity depending on how the drift rate distributions' parameters are constrained across different conditions or groups in an experiment. Given these considerations, an appropriate model selection metric must take into account complexity in a sophisticated manner; for example metrics based on predictive performance such as the Bayes factor or cross-validation.

A second way in which one drift rate distribution may be better than another is that it may be generated by psychologically meaningful assumptions. That is, there may exist cognitive theories which give rise to a particular parametric form for drift rates. For example, the lognormal distribution we have considered here might be motivated by cascaded processing stages (Heathcote and Love, 2012). Alternatively, the Fréchet distribution might be motivated by the consideration of Luce-style choice models (Colonius and Marley, 2014). The methods we outline here provide a template to help others incorporate into the LBA new drift rate distributions motivated by different theoretical considerations. Similar approaches have previously been taken with categorisation and absolute identification models (Nosofsky and Palmeri, 1997; Brown et al., 2008).

A third measure of superiority for the drift rate distributions is based on their statistical properties. A primary use of the LBA model is in the measurement of cognitive effects, for example between experimental conditions (Rae et al., 2014) or between different populations (Ho et al., 2014). It is important for such applications that the model is able to accurately recover the data-generating parameters. While this has been demonstrated for the conventional LBA, our analyses suggest that the new variants investigated here may perform even better in this sense. However, further investigations are required to compare the variants in more realistic sample sizes. Such investigations could also address the issue of model mimicry in more detail, determining what sample sizes are required to differentiate among variants in different parameter regions.

A problem related to the choice between different drift rate distributions is the precise parameterisation used for each model variant. In the truncated normal LBA model the drift rate distribution was characterised by a mean parameter and a standard deviation parameter. This parameterisation has worked well in many situations,

and allows for a separation of effects on location and scale of the drift rate distributions. The new proposed drift rate distributions all have two parameters each, but do not all have a natural location-and-scale parametersiation. Additionally, each new distribution has at least two different parameterisations that are relatively widely used (such as as the rate vs. mean parameterisation for the gamma distribution). It seems likely that identifying the best parameterisation for any new distribution will require large scale empirical investigations.

References

Bebu, I., Mathew, T., 2009. Confidence intervals for limited moments and truncated moments in normal and lognormal models. Statistics and Probability Letters 79, 375–380.

Brown, S. D., Heathcote, A. J., 2008. The simplest complete model of choice reaction time: Linear ballistic accumulation. Cognitive Psychology 57, 153–178.

Brown, S. D., Marley, A., Donkin, C., Heathcote, A. J., 2008. An integrated model of choices and response times in absolute identification. Psychological Review 115 (2), 396–425.

Coffey, C. S., Muller, K. E., 2000. Properties of doubly-truncated gamma variables. Communications in Statistics—Theory and Methods 29 (4), 851–857.

Colonius, H., Marley, A., 2014. Decision and choice: Random utility models of choice and response time. In: Wright, J. (Ed.), International Encyclopedia of the Social and Behavioral Sciences, 2nd Edition. Elsevier.

Donkin, C., Averell, L., Brown, S., Heathcote, A. J., 2009a. Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. Journal of Mathematical Psychology 41 (4), 1095–1110.

Donkin, C., Brown, S. D., Heathcote, A. J., 2009b. The over-constraint of response time models: Rethinking the scaling problem. Psychonomic Bulletin & Review 16, 1129–1135.

Eidels, A., Donkin, C., Brown, S. D., Heathcote, A., 2010. Converging measures of workload capacity. Psychonomic Bulletin & Review.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., Wagenmakers, E.-J., 2008. Striatum and pre–SMA facilitate decision–making under time pressure. Proceedings of the National Academy of Sciences 105, 17538–17542.

Grice, G. R., 1972. Application of a variable criterion model to auditory reaction time as a function fo the type of catch trial. Perception & Psychophysics 12, 103–107.

Hawkins, G., Marley, A., Heathcote, A. J., Flynn, T., Louviere, J., Brown, S., 2014. Integrating cognitive process and descriptive models of attitudes and preferences. cognitive science pdf download the fitting code. Cognitive Science 38, 701–735.

Heathcote, A. J., Brown, S. D., Wagenmakers, E.-J., in press. The falsifiability of actual decision-making models psychological review. Psychological Review.

Heathcote, A. J., Love, J., 2012. Linear deterministic accumulator models of simple choice. Frontiers in Cognitive Science 3, 292.

Ho, T. C., Yang, G., Wu, J., Cassey, P., Brown, S. D., Hoang, N., Chan, M., Connolly, C. G., Henje-Blom, E., Duncan, L. G., Chesney, M. A., Paulus, M. P., Max, J. E., Patel, R., N., S. A., Yang, T. T., Feb 2014. Functional connectivity of negative emotional processing in adolescent depression. Journal of Affective Disorders 155, 65–74. URL http://dx.doi.org/10.1016/j.jad.2013.10.025

Jones, M., Dzhafarov, E. N., 2014. Unfalsifiability and mutual translatability of major modelling schemes for choice reaction time. Psychological Review 121, 1–32.

Luce, R. D., 1986. Response Times. Oxford University Press, New York.

Nadarajah, S., 2009. Some truncated distributions. Acta Applied Mathematics 106 (105–123).

Nosofsky, R. M., Palmeri, T. J., 1997. An exemplar–based random walk model of speeded classification. Psychological Review 104, 266–300.

Rae, B., Heathcote, A. J., Donkin, C., Averell, L., Brown, S. D., 2014. The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. Journal of Experimental Psychology: Learning, Memory, and Cognition 40, 1226–1243.

Ratcliff, R., Rouder, J. N., 1998. Modeling response times for two–choice decisions. Psychological Science 9, 347–356.

Singmann, H., Gretton, M., Brown, S., Heathcote, A., 2015. rtdists: Distribution functions for accumulator models in r. In: Paper presented at the 45th annual meeting of the Society for Computers in Psychology (SCiP). Chicago.

Turner, B. M., Sederberg, P., Brown, S. D., Steyvers, M., 2013. A note on efficiently sampling from distributions with correlated dimensions. Psychological Methods 18 (3), 368–384.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., McKoon, G., 2008. A diffusion model account of criterion shifts in the lexical decision task. Journal of Memory and Language 58, 140–159.

Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., Zeelenberg, R., 2004. A model for evidence accumulation in the lexical decision task. Cognitive Psychology 48, 332–367.

Acknowledgements

This research has been supported by Natural Science and Engineering Research Council Discovery Grant 8124-98 to the University of Victoria for Marley, and by Australian Research Council grants FT120100244, DP12102907 and DP110100234 to the University of Newcastle for Brown and Heathcote. The work was carried out, in part, while Marley was a Distinguished Professor (part-time) at the Institute for Choice, University of South Australia Business School.

Appendix A

Details for Gamma Distribution

The gamma distribution function at x, with shape parameter α and scale parameter β is:

$$G(x;\alpha,\beta) = \frac{\gamma(\alpha, \frac{x}{\beta})}{\Gamma(\alpha)}$$

Here, γ is the lower incomplete gamma function (i.e., $\gamma(x,s) = \int_0^s u^{x-1}e^{-u}du$) and Γ is the standard gamma function, which is just $\gamma(x,\infty)$. Equation 4 gives $Z(t)$ for gamma-distributed drift rates. Its derivative with respect to time is:

$$\begin{aligned}
\frac{d}{dt}\left(Z_\Gamma(t)\right) &= \frac{-b\Gamma(\alpha+1)}{t^2\beta\Gamma(\alpha)}\Bigg[\left(g\left(\frac{b}{t};\alpha+1,\beta\right) - g\left(\frac{b-A}{t};\alpha+1,\beta\right)\right)\left(\Gamma\left(\frac{b}{t};\alpha,\beta\right) - \Gamma\left(\frac{b-A}{t};\alpha,\beta\right)\right) \\
&- \left(g\left(\frac{b}{t};\alpha,\beta\right) - g\left(\frac{b-A}{t};\alpha,\beta\right)\right)\left(\Gamma\left(\frac{b}{t};\alpha+1,\beta\right) - \Gamma\left(\frac{b-A}{t};\alpha+1,\beta\right)\right)\Bigg]\left(\Gamma\left(\frac{b}{t};\alpha,\beta\right)\right. \\
&- \left.\Gamma\left(\frac{b-A}{t};\alpha,\beta\right)\right)^{-2}
\end{aligned}$$

Details for Fréchet Distribution

The cumulative distribution function for the Fréchet distribution, with shape parameter α and scale parameter μ is:

$$p(x;\alpha,\mu) = \exp\left(\left(-\frac{x}{\mu}\right)^{-\alpha}\right)$$

Equation 5 gives $Z(t)$ for Fréchet-distributed drift rates. Its derivative with respect to time is:

$$\begin{aligned}
\frac{d}{dt}\left(Z_{Fr}(t)\right) &= \Bigg[\left(-\alpha\exp\left(-\left(\frac{\mu b}{t}\right)^{-\alpha+1}\right)\frac{\left(\frac{\mu b}{t}\right)^{-\alpha+1}}{t^{-\alpha+2}} + \alpha\exp\left(-\left(\frac{\mu(b-A)}{t}\right)^{-\alpha+1}\right)\frac{\left(\frac{\mu(b-A)}{t}\right)^{-\alpha+1}}{t^{-\alpha+2}}\right)\left(Fr\left(\frac{b}{t};\mu,\alpha\right)\right. \\
&- \left.Fr\left(\frac{b-A}{t};\mu,\alpha\right)\right) + \frac{b}{t^2}\left(g\left(\frac{b}{t};\mu,\alpha\right) - g\left(\frac{b-A}{t};\mu,\alpha\right)\right)\left(\Gamma\left(1-\frac{1}{\alpha},(\frac{\mu b}{t})^{-\alpha}\right)\right. \\
&- \left.\Gamma\left(1-\frac{1}{\alpha},(\frac{\mu(b-A)}{t})^{-\alpha}\right)\right)\Bigg]\left(Fr\left(\frac{b}{t};\mu,\alpha\right) - Fr\left(\frac{b-A}{t};\mu,\alpha\right)\right)^{-2}
\end{aligned}$$

Details for Lognormal Distribution

The cumulative distribution function for the lognormal distribution, with underlying mean μ and standard deviation σ is:

$$p(x;\mu,\sigma) = \Phi(log(x);\mu,\sigma)$$

As throughout, $\Phi(.;\mu,\sigma)$ indicates the cumulative distribution function of a normal Equation 6 gives $Z(t)$ for lognormally-distributed drift rates. Its derivative with respect to time is:

$$\begin{aligned}
\frac{d}{dt}\left(Z_{LN}(t)\right) &= \frac{-1}{t\sigma}\Bigg[\left(\phi\left(\frac{\log(\frac{b}{t})-\mu-\sigma^2}{\sigma}\right)+\phi\left(\frac{\log(\frac{b-A}{t})-\mu-\sigma^2}{\sigma}\right)\right)\left(\Phi\left(\frac{\log(\frac{b}{t})-\mu}{\sigma}\right)+\Phi\left(\frac{\log(\frac{b-A}{t})-\mu}{\sigma}\right)\right) \\
&- \left(\phi\left(\frac{\log(\frac{b}{t})-\mu}{\sigma}\right)+\phi\left(\frac{\log(\frac{b-A}{t})-\mu}{\sigma}\right)\right)\left(\Phi\left(\frac{\log(\frac{b}{t})-\mu-\sigma^2}{\sigma}\right)\right. \\
&+ \left.\Phi\left(\frac{\log(\frac{b-A}{t})-\mu-\sigma^2}{\sigma}\right)\right)\Bigg]\left(\Phi\left(\frac{\log(\frac{b}{t})-\mu}{\sigma}\right)+\Phi\left(\frac{\log(\frac{b-A}{t})-\mu}{\sigma}\right)\right)^{-2}\exp\left(\mu+\frac{\sigma^2}{2}\right)
\end{aligned}$$

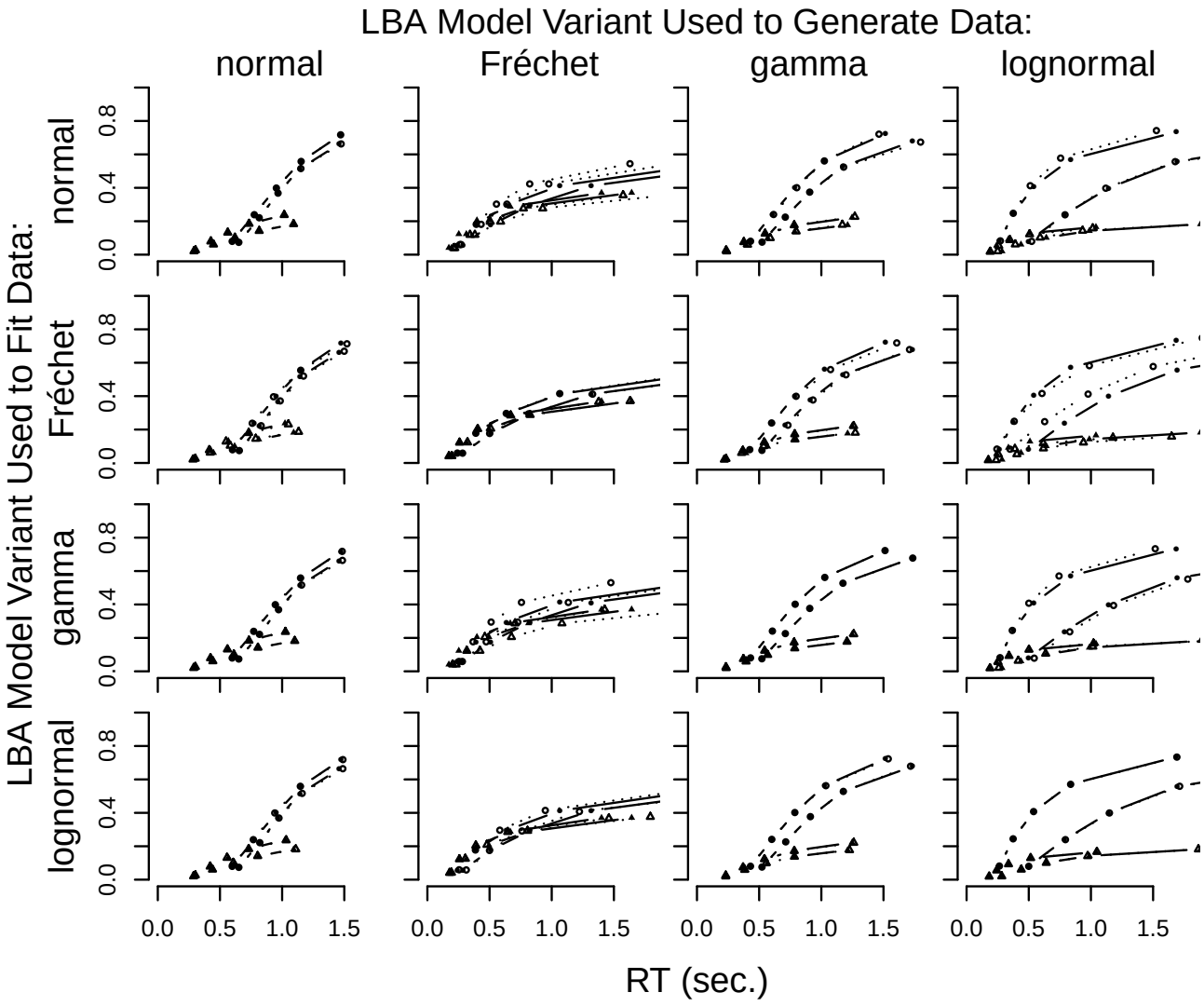

Figure 3: Results of cross-fitting the model variants. Columns indicate which of the four LBA model variants was used to generate synthetic data (name of drift rate distribution for each variant given at top of column). Likewise, rows indicate which variant was used to fit those synthetic data. In each panel, the leftmost pair of solid lines and filled symbols show the joint cumulative distribution (plotted by five percentiles: 10, 30, 50, 70, and 90) over correct and incorrect responses for the simulated speed-emphasis condition, and the rightmost pair show the same for the simulated accuracy-emphasis condition. Percentiles for correct responses are shown by circles, and for errors by triangles. Overlaid open symbols and dotted lines show the model fits.

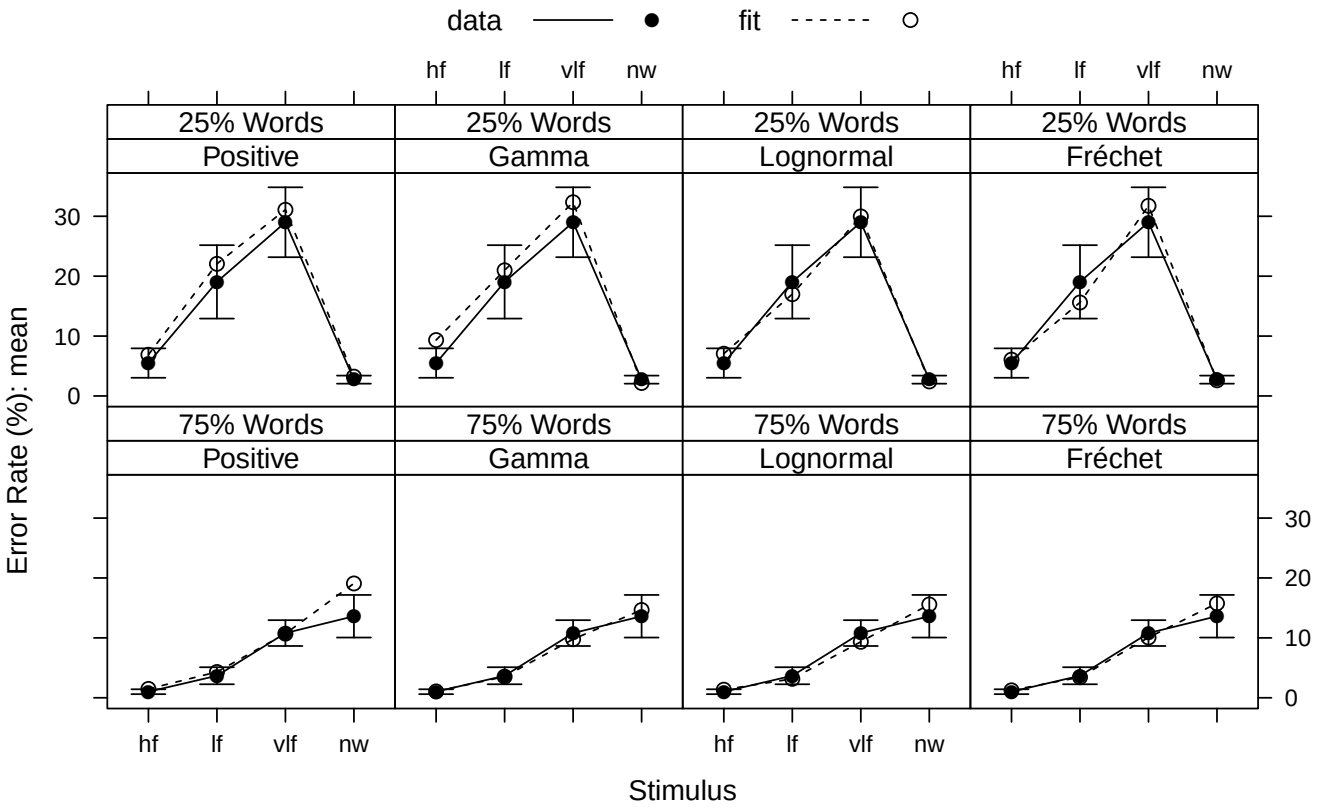

Figure 4: Lexical decision error rates and corresponding model fits, both averaged over participants, as a function of stimulus (nw = non-word, for words hf = high frequency, lf = low frequency, vlf = very-low frequency) and proportion of word stimuli. Error bars indicated 95% confidence intervals.

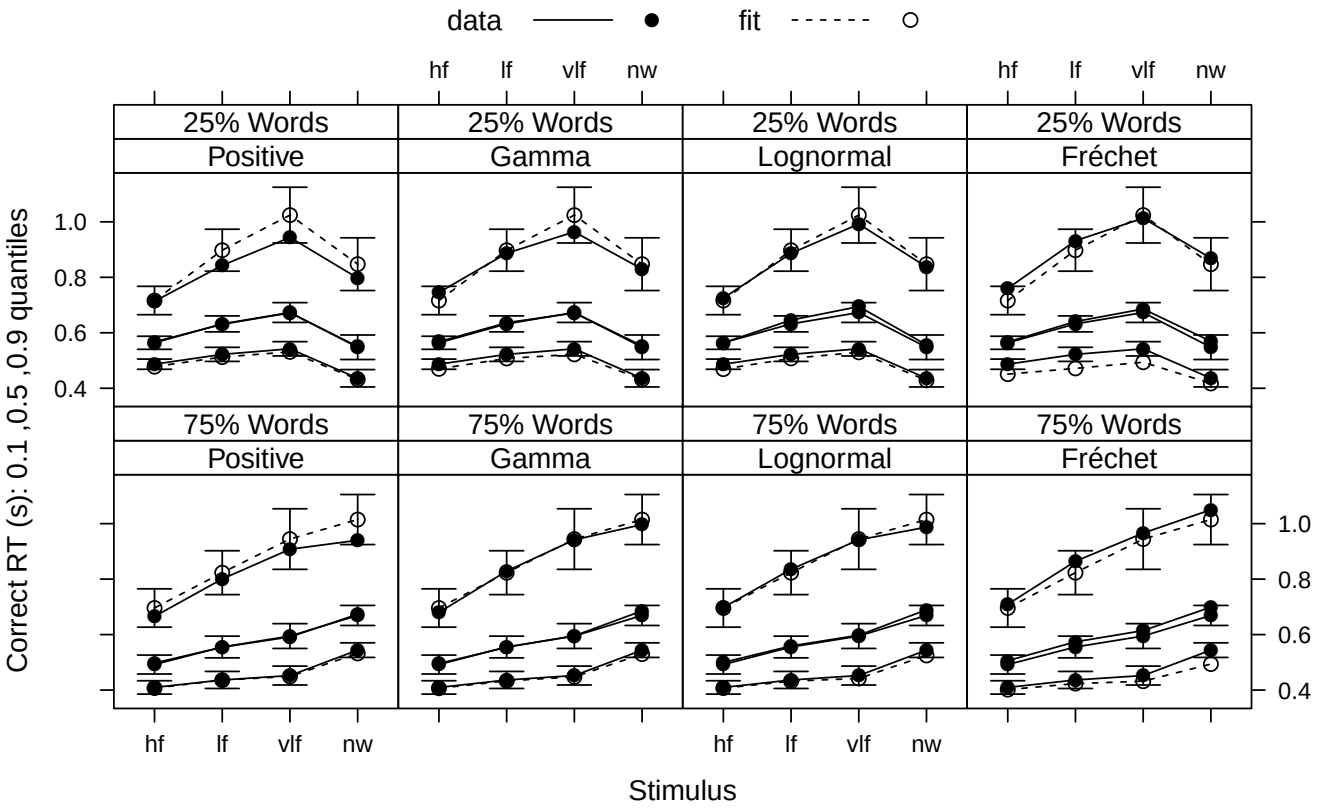

Figure 5: Lexical decision RT quantiles (10^{th}, 50^{th} and 90^{th} percentiles) for correct responses and corresponding model fits, both averaged over participants, as a function of stimulus (nw = non-word, for words hf = high frequency, lf = low frequency, vlf = very-low frequency) and proportion of word stimuli. Error bars indicate 95% confidence intervals.

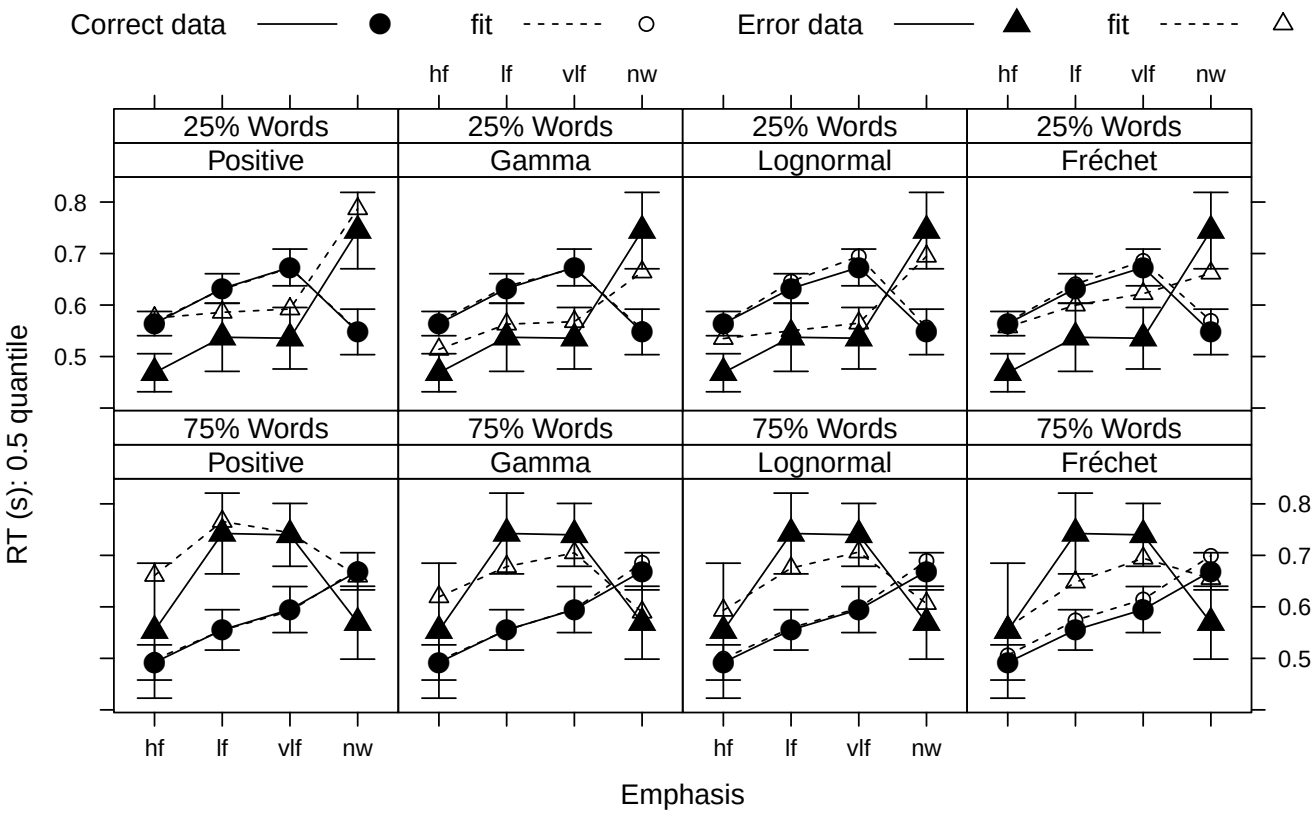

Figure 6: Lexical decision median RT for correct and error responses and corresponding model fits, both averaged over participants, as a function of stimulus (nw = non-word, for words hf = high frequency, lf = low frequency, vlf = very-low frequency) and proportion of word stimuli. Error bars indicate 95% confidence intervals.

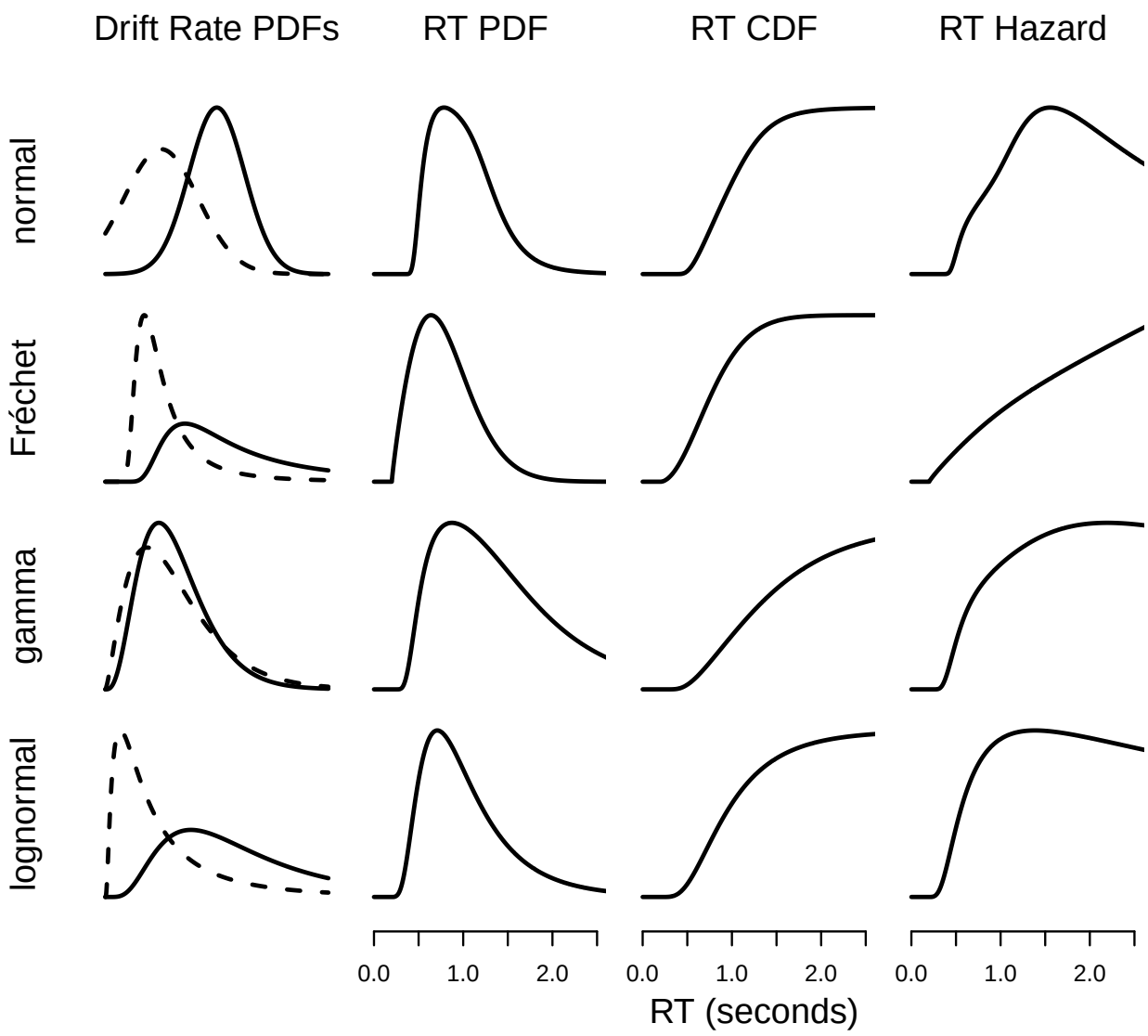

Figure 7: Each of the four LBA variants is shown on a different row. The left column shows the distributions of drift rates for two accumulators which race to produce RT distributions. The other three columns show the following properties of those RT distributions, from left-to-right: the PDF, the CDF, and the hazard function (i.e. $\frac{PDF}{1-CDF}$). The parameters used to generate this plot match the parameters used for the main simulation study.