Reaction Times and the Ex-Gaussian Distribution: When is it Appropriate?

Saul Sternberg [1]
University of Pennsylvania

Fitting of the ex-Gaussian distribution to reaction times and drawing conclusions from its estimated parameter values is becoming increasingly popular. The purpose of this note is to show that the ex-Gaussian distribution is inappropriate, given assumptions that are often made about the processes underlying reaction times. If so, the estimated parameters may be misleading, and the inferences from them incorrect.

Christie & Luce (1956) and McGill (1963) suggested that the reaction time (RT) can be thought of as the sum (convolution) of two stochastically independent random variables, one of which is exponentially distributed, corresponding to the durations of two successive processes, or *stages*. Hohle (1965) suggested that the other random variable was Gaussian, and introduced the ex-Gaussian distribution, a convolution of exponential and Gaussian distributions, as a description of RT distributions, and applied it to the data from a simple RT experiment. He argued that the duration of the decision process might be exponentially distributed, and that because the residual latency might represent the sum of the durations of several operations with stochastically independent durations, it might be approximately Gaussian.

The mathematical tractability of the exponential distribution is appealing, but to what sort of process might it apply? It may describe the waiting times between initiations of independent telephone calls in a large parallel network, or the times between decays among the innumerable atoms in a lump of radioactive material. However, because of the no-memory property unique to the exponential distribution it cannot describe the duration of a mental process that accomplishes something, unless all the work is done in an instant. No matter how much time has elapsed, as long as processing is not yet complete, the expected remaining time needed to complete it is unchanged. (The hazard function associated with the exponential distribution is flat.) Thus, at any time point, either no work or all the work has been accomplished. The work must therefore be accomplished instantaneously. Instead, a plausible processing-time distribution must surely be one whose hazard function increases during at least some time interval.

The ex-Gaussian distribution[2] was promoted by Ratcliff & Murdock (1976) (who called it the "convolution model", and showed that it compared favorably to two other distributions for their study-test data), by Ratcliff (1978, 1979), who used it for several memory retrieval paradigms, and by Hockley (1984), who used it for memory search and visual search data with set and display sizes from 3 to 6, as well as for other paradigms.

---

1. Thanks to Frank Norman for helpful suggestions, especially regarding the implications of Eq. 4.

2. The name "ex-Gaussian" appears to have been coined by Burbeck & Luce (1982).

It was applied to recency discrimination RTs by Muter (1979) and Hacker (1980); since then it has been used by many other investigators.

Hohle's hope that the exponential and Gaussian components would describe the durations of functionally distinct stages of which the RT is the sum, so that different experimental factors would be associated consistently with the two stages, has not been realized (Schwarz, 2001; Matzke & Wagenmakers, 2009, especially Table 1). Nonetheless, the ex-Gaussian distribution has become increasingly popular[3] as a description of RT distributions from a wide range of experimental paradigms, and estimates of its parameters have been used to draw conclusions about underlying mechanisms.

It is therefore of interest to ask whether the kinds of processes whose durations can be described by the ex-Gaussian distribution are constrained in some way. One purpose of this note is to show that such constraints exist, for sequential processes (stages) with stochastically independent durations.[4]

Reaction Times for Iterated Processes

Assume a process that generates reaction times RT_s that contain a base time (T_b) plus s sequential comparison times (T_c), as in some search models in which the number of elements in a display a memory set is s, and s comparisons must be made:

$$RT_s = T_b + T_{c1} + T_{c2} + \cdots + T_{cs} \quad . \tag{1}$$

Suppose that these component times are mutually stochastically independent. Can the ex-Gaussian distribution be used to describe the resulting reaction-time distributions $\{RT_s\}$ for a set of different s-values?

Two set sizes or display sizes: Consider first the case of two different s-values, $s2 > s1$ and the component of RT_s for $s > 1$ that contains the sum of two or more comparison times. Because sums of independent exponential variables are neither exponential nor Gaussian nor ex-Gaussian (and because sums of Gaussian variables are Gaussian), it is the base-time distribution that must be exponential and the comparison-time distribution Gaussian. This implies that the difference between the distributions of RT_{s1} and RT_{s2} should be captured by one or both of the Gaussian parameters μ and σ, and that there should be no difference in the exponential parameter τ.

Contrary to this expectation, the exponential parameter of fitted ex-Gaussian distributions has been found to change with changes in set size in visual search of small sets (sizes 3 - 6, Hockley, 1984, Exp. 1, negative responses) and larger sets (sizes 3 - 18,

---

3. In PsycInfo, of the 108 papers published through 2013 whose abstracts mention "ex-Gaussian", 84% were published after 1999. Of the 153 citations of Hohle (1965) through 2013, 53% are in papers published after 1999.

4. That stage models with stochastically independent durations imply constraints on RT distributions surprised me. In another report (Sternberg, 2014) I show that such models are inconsistent with the invariance of the shapes of RT distributions across conditions or groups.

Palmer et al., 2011, positive and negative responses), and in memory search of small sets (sizes 3 - 6, Hockley, 1984, Exp. 1, positive and negative responses). This could mean that the model described by Eq. 1 is incorrect (as argued by Hockley for his memory-search data but not for visual search) or that the ex-Gaussian distribution is inappropriate.

Three or more set sizes or display sizes: Next, consider the case of three or more different s-values. The cumulant of order r of the ex-Gaussian distribution is the sum of the rth cumulants of the exponential and Gaussian distributions. For set size s, let the exponential parameter for s comparisons be τ_s, and let the Gaussian parameters be μ_s and σ_s. The first four cumulants of the exponential distribution are τ_s, $\tau_s{}^2$, $2\tau_s{}^3$, and $6\tau_s{}^4$. The first four cumulants of the Gaussian distribution are μ_s, $\sigma_s{}^2$, 0, and 0. It follows that the first four cumulants of the ex-Gaussian reaction-time RT_s distributions are the sums: $\kappa_{1s} = \tau_s + \mu_s$, $\kappa_{2s} = \tau_s{}^2 + \sigma_s{}^2$, $\kappa_{3s} = 2\tau_s{}^3$, and $\kappa_{4s} = 6\tau_s{}^4$.

If we assume identical comparison-time distributions as well as stochastic independence, all four of these cumulants must be linear in s. However, it follows from the above that $\kappa_{4s} = 3\kappa_{3s}{}^{4/3}$. Because κ_{4s} is a power function of κ_{3s} with exponent not equal to 1.0, they cannot both be linear in s. It follows that the ex-Gaussian distribution cannot describe the RT distributions in this situation for more than two values of s.

Reaction Times in an $m \times n$ Factorial Experiment With Selective Influence

Because an $m \times n$ experiment can be regarded as a concatenation of 2×2 experiments, we can limit ourselves to factors with two levels, without loss of generality.

Suppose a stage model with two stages, **A** and **B**, whose durations are stochastically independent, and two factors, F_j and G_k, each with two levels, $j = 1, 2$; $k = 1, 2$; that influence the stage durations selectively, so that $T_A = T_A(F_j) = T_{Aj}$, $T_B = T_B(G_k) = T_{Bk}$, and the total RT is $T_{jk} = T_{Aj} + T_{Bk}$. Consider a 2×2 factorial experiment with the four resulting conditions.

Then, because convolution is associative and commutative, $T_{11} + T_{22}$ has the same distribution as $T_{12} + T_{21}$, namely, the convolution of the distributions of T_{A1}, T_{A2}, T_{B1}, and T_{B2}:

$$T_{11} * T_{22} = T_{12} * T_{21} \quad , \tag{2}$$

where * represents convolution.[5] It follows that

$$mgf_{11} \times mgf_{22} = mgf_{12} \times mgf_{21} \quad , \tag{3}$$

where mgf_{jk} denotes the moment generating function of T_{jk}.

Now, suppose that all of the T_{jk} are ex-Gaussian. What constraints if any does Eq. 3 impose on the parameters of the four ex-Gaussian distributions? There are three such constraints: (1) No more than one of the two factors can influence the exponential parameter. (2) The Gaussian parameters $\{\mu_{jk}\}$ must be additive. (3) The Gaussian

5. For a more formal proof, see Roberts and Sternberg (1993, p. 646).

parameters $\{\sigma^2_{jk}\}$ must be additive. A proof follows:

Let the means and variances of the Gaussian distributions be $\{\mu_{jk}\}$ and $\{\sigma^2_{jk}\}$ and the means of the exponential distributions be $\{\tau_{jk}\}$. The ex-Gaussian moment generating function is the product of the Gaussian mgf, $\exp(\mu t + \sigma^2 t^2/2)$, and the exponential mgf, $1/(1 - \tau t)$; this product is $\exp(\mu t + \sigma^2 t^2/2)/(1 - \tau t)$.

It follows from Eq. 3 that:

$$\frac{\exp(\mu_{11} t + \sigma^2_{11} t^2/2)}{(1 - \tau_{11} t)} \times \frac{\exp(\mu_{22} t + \sigma^2_{22} t^2/2)}{(1 - \tau_{22} t)} = \frac{\exp(\mu_{12} t + \sigma^2_{12} t^2/2)}{(1 - \tau_{12} t)} \times \frac{\exp(\mu_{21} t + \sigma^2_{21} t^2/2)}{(1 - \tau_{21} t)} . \quad (4)$$

The left-hand side of Eq. 4 goes to infinity when either $t = 1/\tau_{11}$ or $t = 1/\tau_{22}$. The right-hand side goes to infinity when either $t = 1/\tau_{12}$ or $t = 1/\tau_{21}$. Because the two sides must go to infinity for the same values of t, one of the following three conditions must obtain:

(a) All the τ_{jk} are equal, which means that neither F nor G influences τ.

(b) $\tau_{11} = \tau_{21} \neq \tau_{22} = \tau_{12}$, which means that factor F has no influence on τ.

(c) $\tau_{11} = \tau_{12} \neq \tau_{22} = \tau_{21}$, which means that factor G has no influence on τ.

There is thus a contradiction between both factors influencing τ and a stage model with stochastically independent durations. Furthermore, given any one of the three conditions above, the products of the denominators on the two sides of Eq. 4 are equal, which means that we can ignore them, and that

$$\exp\,[\,(\mu_{11} + \mu_{22})t + (\sigma^2_{11} + \sigma^2_{22})t^2/2\,] = \exp\,[\,(\mu_{12} + \mu_{21})t + (\sigma^2_{12} + \sigma^2_{21})t^2/2\,]\,, \quad (5)$$

or

$$(\mu_{11} + \mu_{22})t + (\sigma^2_{11} + \sigma^2_{22})t^2/2 = (\mu_{12} + \mu_{21})t + (\sigma^2_{12} + \sigma^2_{21})t^2/2\,, \quad (6)$$

or

$$\mu_{11} + \mu_{22} = \mu_{12} + \mu_{21}\,, \quad (7)$$

and

$$\sigma^2_{11} + \sigma^2_{22} = \sigma^2_{12} + \sigma^2_{21}\,. \quad (8)$$

Thus, whichever of the three conditions (a), (b), or (c) applies, the Gaussian means and variances must be separately additive.

I know of two reported $m \times n$ factorial experiments in which the effects of two factors on mean RT are additive, consistent with a two-stage model with selective influence, and in which ex-Gaussian distributions were fitted to the data from the mn conditions. In a 2×2 factorial experiment, Plourde & Besner (1997) varied word frequency (*WF*) and stimulus quality (*SQ*) in a lexical decision task. Both *WF* and *SQ* influenced τ, violating constraint (1). Otherwise the constraints were satisfied: Both factors influenced μ_{ij}, and their effects were additive; and neither *WF*, *SQ*, nor their interaction influenced σ^2. In a 2×4 factorial experiment, Leth-Steenson (2009) varied foreperiod (*FP*, two levels) and numerical proximity (*NP*, four levels) in a number-comparison task.[6] Both *FP* and *NP* influenced τ, violating constraint (1). In this experiment, both of the other constraints were also violated: there were interactions of the effects of *FP* and *NP* on both μ and σ^2.

How Well Does It Fit?

Does success of the ex-Gaussian distribution argue against the assumptions (sequential processes with stochastically independent durations) that make this distribution unlikely? We need to ask how successful it has actually been. In a sample of 15 of the papers published during 2012 and 2013 in which estimated parameters from fits of the ex-Gaussian distribution to RT data were reported, I found only two that reported statistical tests of goodness of fit. Perhaps investigators believe that its success for some experiments implies that it is appropriate for others. But not all RT distributions are created equal.

The earlier papers are more helpful in this respect. Hohle (1965) reported a summary of the χ^2 tests of goodness of fit for the 32 RT distributions of 100 observations from four subjects in each of eight conditions in a simple-RT experiment, indicating satisfactory fits, and displayed plots of the observed and fitted cumulative distribution functions for eight of the distributions. In experiments on memory search, visual search, and recency judgment, Hockley (1984) found that the ex-Gaussian distribution fit well in the proponderance of cases. Whereas Ratcliff and Murdock (1976) and Ratcliff (1979) reported no tests, they did provide values of χ^2 and degrees of freedom for fits of the ex-Gaussian distribution to a large subset of their distributions of RTs for correct responses in study-test experiments, which permit performing such tests.[7] Their findings argue emphatically against the ex-Gaussian distribution: the median p-value is 0.0001, and only 10 of the 72 p-values are 0.05 or greater. [8]

The plots of 43 distributions in two of these papers (figures 9 and 13 in Ratcliff & Murdock, 1976 for individual subjects, for which χ^2 values are reported, and figures 11 and 15 in Ratcliff, 1978, for group distributions, for which they are not reported) suggest systematic features of the deviations between the fitted and observed distributions: In majorities of these distributions, the fitted mode is below the observed mode, and the fitted proportions in a region to the right of the mode are above those observed.[9]

---

6. In this study the ex-Gaussian distribution was fitted to 16 data sets from each of 25 subjects, but each set contained only 16 or fewer observations.

7. These distributions include twelve for each of the four subjects in Experiment 1 and eight for subject 3 in Experiment 4 in Ratcliff & Murdock (1976), one for each of the four subjects in Experiment 2 in that study, reported in Ratcliff (1979), and one for each of the four subjects in Experiment 1 of Ratcliff (1978), also reported in Ratcliff (1979).

9. One way to facilitate detection of such systematic deviations, which may occur even when sets of individual distributions do not fail goodness-of-fit tests, is exemplified by Fig. 26.3 in Roberts & Sternberg (1993), who compared distributions by examining the across-subject mean differences between cumulative distribution functions, along with their standard errors. In plots of the observed and fitted cumulative distributions themselves, such as those of Hohle (1965), it may be difficult to see small deviations.

Implications

Investigators are drawing conclusions from estimates of the parameters of ex-Gaussian distributions fitted to RT data, and the effects of experimental factors on these parameters. The arguments above show that this distribution may be inappropriate, given assumptions that are often made about the processes underlying reaction times, in particular, that they include sequential operations (stages) whose durations are stochastically independent. This conclusion is supported by the fact that it often fits poorly. If so, the estimated parameters may be misleading, and the inferences from them incorrect.

References

Burbeck, S. L. & Luce, R. D. (1982) Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics, 32,* 117-133.

Christie, L. S. & Luce, R. D. (1956) Decision structure and time relations in simple choice behavior. *Bulletin of Mathematical Biophysics, 18,* 89-112.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6,* 651-675.

Hockley, W. E. (1984) Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 598-615.

Hockley, W. E. (2008). Memory search: A matter of time. In H.L. Roediger, III (Ed.) *Cognitive Psychology of Memory*. Vol. 2 of *Learning and Memory: A Comprehensive Reference* (J. Byrne, Ed.). Oxford: Elsevier. Pp. 417-444.

Hockley, W. E., & Corballis, M. C. (1982). Tests of serial scanning in item recognition. *Canadian Journal of Psychology 36,* 189-212.

Hohle, R. H. (1965) Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology, 69,* 382-386.

Matzke, D. & Wagenmakers, E.-J. (2009) *Psychonomic Bulletin & Review, 16,* 798-817.

McGill, W. J. (1963) Stochastic latency mechanisms. In R. D. Luce, R.R. Bush, & E. Galanter (Eds.) *Handbook of mathematical psychology, Vol. 1.* New York: Wiley. Pp. 309-360.

---

8. Ratcliff & Murdock (1976) suggest that the poor fits result from inhomogeneity over conditions and practice (p. 199), conclude that "the convolution is adequate" (p. 201), and say that "the convolution . . . (where the fit is reasonable) provides an excellent summary of reaction time distributions. . ." (p. 202). The interpretation of these remarks varies widely: After reporting that goodness of fit tests indicated significant differences for 35 of their 96 RT distributions in their Exp. 1, Hockley & Corballis (1982, p. 199) say that "Although the proportion of significant chi-square values might seem high, Ratcliff and Murdock (1976) point out that chi-square values are almost inevitably inflated by systematic trends, such as practice effects, in the data. We therefore consider the fits to be reasonably satisfactory." For their Exp. 2, they report that goodness of fit decreased systematically with set size, with the percentage of significant differences ranging from 22% to 78%. On the other hand, referring to Ratcliff & Murdock (1976) and Hohle (1965), McElree & Dosher (1993, p. 298), who report no measures of goodness of fit for their data, say that "RT distributions of recognition judgments appear to be extremely well fit by the convolution of a Gaussian and an exponential distribution . . .". And Hockley (2008) claims that "Ratcliff and Murdock (1976) showed that the ex-Gaussian distribution provides a very good description of observed RT distributions for a number of recognition memory phenomena."

McElree, B. & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, *122*, 291-315.

Muter, P. (1979). Response latencies in discriminations of recency. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 160-169.

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 37,* 58-71

Ratcliff, R. (1978) A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

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

Roberts, S. & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.) *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive Neuroscience.* Cambridge, MA : M.I.T. Press. Pp. 611-653.

Schwarz, W. (2001) The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, & Computers, 33,* 457-469.

Sternberg, S. (2014). *Sequential processes and the shapes of reaction-time distributions.* Retrieved from http://www.psych.upenn.edu/˜saul/RTshapes.stages.pdf