Running head: SURVIVAL AND LATENCIES

Are divergence point analyses suitable for response time data?

Pablo Gomez (1), Javier Breithaupt (2), Manuel Perea (2, 3), Jeff Rouder (4)

(1) DePaul University, Chicago IL, USA

(2) Universitat de València, Valencia, Spain

(3) BCBL, Basque Center on Cognition, Brain, and Language, San Sebastián, Spain

(4) University of Missouri, Columbia MO, USA

Pablo Gomez

Department of Psychology

DePaul University

Chicago, Illinois

60614

e-mail: pgomez1@depaul.edu

Abstract

Estimating the time course of the influence of different factors in human performance is one of the principal topics of research in cognitive psychology/neuroscience. Over the past decades, researchers have proposed several methods to tackle this question using latency data. Here we examined a recently proposed procedure that employs survival analyses on latency data to provide "precise estimates" of the timing of the first discernible influence of a given factor on performance (e.g., word frequency on lexical access). A number of articles have used this method in recent years, and hence an exploration of its strengths and its potential weaknesses is in order. Unfortunately our analysis revealed that the method has conceptual flaws, and it might lead researchers into believing that they are obtaining a measurement of a processing components when in fact they are obtaining a nonsensical measurement.

Are divergence point analyses suitable for response time data?

Perhaps the most common cognitive psychology experiment is one in which participants are presented with stimuli that vary in a dimension of theoretical interest (e.g., length, word frequency, etc.). The stimulus elicits a response, and researchers measure latencies to make inferences about hypothesized underlying cognitive processes. This form of mental chronometry is widely used in the analyses of data from a wide range of experimental paradigms such as choice tasks, naming, eye-tracking, and many others.

Although the most popular model of analysis are tests of mean latencies, the shortcomings of focusing only on means to infer cognitive processes are well known (Balota & Yap, 2011; Heathcote, Popiel, & Mewhort, 1991; Ratcliff, 1979, Rouder, Lu, Speckman, Sun, & Jiang, 2005); indeed, theory development benefits from exploring distributional properties of latency measurements. To take advantage of the distributional information of latency data, some researchers use methods that are based on fitting functional forms like the ex-Gaussian or the Weibull distributions (see Heathcote et al., 1991), while other researchers use methods that are based on process models like the diffusion model for choice response times (Ratcliff, 1978) or the EZ-reader model for eye fixation durations during reading (Reichle, Pollatsek, Fisher, & Rayner, 1998).

There are several desired properties for any general method of analysis. These properties include (a) a method that uses distributional properties, (b) a lack of detailed assumptions about the underlying process, (c) a lack of specific assumptions the parametric form, (d) broad applicability in many paradigms, and (e) interpretability in theoretically meaningful terms. In this note, we explore a method, the *divergence point analysis*, a method some consider to have all of properties outlined above. The proponents of this method (Sheridan, 2013; see also Ando, Matsuki, Sheridan, & Jared, 2015; Reingold & Sheridan, 2014; Reingold, Reichle, Glaholt, & Sheridan, 2012; Sheridan & Reingold,

2013; Sheridan, Rayner, & Reingold, 2013) aim to estimate the onset of the influence of a given variable on the basis of latency data; this onset is referred to as the *divergence point*.

Divergence point analysis utilizes survival functions in its estimation procedure; hence, a basic description of survival functions is helpful Let T be a random variable that denotes the response time and let $F(t)$ be its cumulative distribution function (CDF). The CDF denote the probability that the value of T is smaller than some value t, i.e., $F(t) = Pr(T \leq t)$. The survival function is the complement probability: $S(t) = Pr(T > t) = 1 - F(t)$, this is the probability that a response occurs later than some value t; hence at $t = 0$, $S(0) = 1$ and at $t = \infty$, $S(\infty) = 0$. There have been attempts to use survival functions in latency analyses. Notably, Van Zandt (2002) examined several of these procedures and concluded that serious analyses of this type, "would use samples of at least a few hundred observations" (p. 482). Along similar lines, Houpt and Townsend (2010, 2011) discussed a rather sophisticated method of survival analysis that involves experimental methods and a non-trivial algorithm termed survivor interaction contrasts (SIC).

One of the appeals of the divergence point method is that it might overcome the limitations of traditional survival function analyses by using a computationally intensive bootstrapping procedure. The divergent point method is a computational approach to assessing where two survivor functions first diverge. Each survivor function represents an experimental condition, say generically, experimental and control. Processing is thought to be unaffected by the manipulation before the divergent time point and affected thereafter. The time point then is the time in processing where the manipulation first has influence. Understanding how these divergent time points depend on manipulations then provides a valuable insight into processing.

An example from Sheridan, 2013 is provided in Figure 1. The top panel shows vincentized RT distributions and the goal is to assess the earliest point these distributions

differ. The corresponding survivor functions are shown in the bottom panel. The thick horizontal bar shows time points where the survival functions differ significantly as determined by a bootstrap test. The earliest such point, the vertical line, is the divergent point. It the earliest point where the manipulation is said to affect processing.

Although this method might seem promising and useful for researchers interested in exploring the time course of a given empirical effect, there is a fundamental conceptual flaw in the foundation of the method.

Estimating divergence points in latency distributions is conceptually flawed

The *divergence point* analysis rests on the notion that processing in two conditions proceeds in the same way and at the rate until a point in time (i.e., the divergence point). To show that this claim is hopelessly flawed, we considered perhaps the simplest case. Suppose processing was mediated by two stages which operated in serial manner. The finishing time of the first stage is a normal with mean μ and standard deviation σ; the finishing time of the second stage is an exponential with scale τ (the latencies would follow an exgaussian distribution). The manipulation of interest is assumed to affect only the later exponential stage, and it corresponds to a lengthening of scale τ for one of the conditions. In this setup a divergence point might be expected, perhaps at μ. Figure 2 shows the density, CDF, and survival functions for a hypothetical control and experimental conditions under these assumptions. Perhaps counter intuitively, the CDFs have no common points to diverge from. Instead, the distribution with the smaller exponential scale is faster everywhere. Obviously, if the manipulation of interest was in the the mean μ of the first stage the same conclusion holds as shown in Figure 3. In both cases, the relationship is one of strict stochastic dominance. If one is to talk about a divergence point at all, it would be at $-\infty$.

This flaw in divergence point analyses is easy to understand as a an error in logic. The claim is based on a moment-to-moment view of processing on a single trial. Yet, RT distributions are the collection of several trials, and the properties of the distribution do not reflect the properties of any one specific trial. The flaw is an example of distortions from aggregation which has been well known since Estes (1956). As an exercise we generated distributions with a true divergence point (525 ms) and realized that we needed to utilize a somewhat contrived procedure to generate these distributions: we assumed that latencies under 525 ms in one of the conditions would be substituted by a random number from an exponential distribution with scale = 121 to which we added a constant of 525; these distributions are shown in Figure 4. We then attempted to think of processing architectures that could yield a true divergence point of this type. We considered diffusion models, counter models, race models, additive factor and stage insertion models. These models all fail to produce a true divergence point. In all cases, there is stochastic dominance, meaning that the divergence happens at the bottom or minimum of the distributions. We were able to specify one case, a mixture model where the components do not overlap at all. Such a model is highly artificial and unrealistic in any setting. It remains an open challenge, one that we suspect cannot be met, to find a plausible process model that predicts a non-minima divergence point.

The method provides nonsensical results

What happens when the divergence-point analysis is applied to a realistic example? Well, the algorithm guarantees a real-valued answer so long as the distributions differ somewhere (and they do in the examples presented below). In small sample sizes, the smaller differences at the bottom of the distribution cannot be detected and divergence points in the middle of the distribution might be expected. But in larger sample sizes, the estimate will shift radically downward. In the large sample limit, the estimate will shift

downward without bound (at least theoretically, although in practice the divergence point is bounded by the shortest latencies).

As an illustration, we generated data from an ex-Gaussian distribution assuming that the experimental effect was either in the μ or τ parameters. We manipulated the number of hypothetical trials per condition and then applied the bootstrapping method to estimate the divergence point.

In the first simulation, we generated data in which the difference between the two conditions was an effect on μ. The data for the baseline condition was generated from an ex-Gaussian distribution with $\mu = 541$, $\sigma = 68$, and $\tau = 115$. We generated data for three simulated experimental conditions by changing μ to 561, 581 and 621 ($\Delta\mu = 20, 40, 80$) There is, therefore, stochastic dominance of the baseline condition relatively to all of these other conditions, and the true divergence point is at the starting point of the distributions. The results from this simulation are not encouraging for the method, as the estimation of the divergence point is highly biased by the number of trials per condition ($n = 20, 30, 50, 100, 250, 500, 1000$). Figure 5 shows the average divergence point for each of the parameter combinations (μ and n) across 1000 simulations. As a consequence of increased statistical power due to larger sample size at the trial level, the larger the number of trials per condition, the shorter the estimated divergence point (i.e., there is a statistical bias dependent on sample size). For example, with $\Delta\mu = 80$, and $n = 50$, the divergence point is about 100ms higher than for $n = 1000$. In fact, when the number of trials is below 100, the different conditions are indistinguishable from each other.

In the second simulation, we generated latency data in which the difference between the two conditions was an effect on τ. The data for the baseline condition was the same as in the previous simulations: it was generated from an ex-Gaussian distribution with $\mu = 541$, $\sigma = 68$, and $\tau = 115$. We generated data for three simulated experimental conditions by changing τ to 135, 155 and 195 ($\Delta\tau = 20, 40, 80$). As shown in Figure 2,

changes in τ produce stochastic dominance of the baseline condition and the true divergence point is at the starting point of the distributions. The results from this simulation are very similar to those from the first simulation: The estimation of the divergence point is severely biased by the number of trials per condition (n = 20, 30, 50, 100, 250 or 500). Figure 6 shows the average divergence point for each of the parameter combinations (τ and n) across 1000 simulations.

Conclusion

Although the goals of the divergence point method are worth pursuing, our analysis revealed serious shortcomings on the conceptual foundation of the procedure: Latency measurements tend to exhibit stochastic dominance between experimental conditions, and hence the divergence point would be at the leading edge of the latency distribution regardless of other distributional differences. Furthermore, if the method is applied to data, an estimate of the divergence point will be provided by the method. This estimate will be affected mostly by the number of observations. In short, our exploration of the method forces us to conclude that it is not advisable to utilize this divergence point method when analyzing latency data.

References

Ando, E., Matsuki, K., Sheridan, H., & Jared, D. (2015). The locus of Katakana-English masked phonological priming effects. *Bilingualism: Language and Cognition, 18,* 101-117. doi: 10.1017/S1366728914000121

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science, 20,* 160-166. doi: 10.1177/0963721411408885.

Estes, W. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 52*, 134-140. doi: 10.1037/h0045156.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 340-347. doi: 10.1037/0033-2909.109.2.340

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59-108. doi: 10.1037/0033-295X.85.2.59

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446-461. doi: 10.1037/0033-2909.86.3.446

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125 -157. doi: 10.1037/0033-295x.105.1.125

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology, 65,* 177-206. doi: 10.1016/j.cogpsych.2012.03.001

Reingold, E. M., & Sheridan, H. (2014). Estimating the divergence point: a novel distributional analysis procedure for determining the onset of the influence of experimental variables. *Frontiers in Psychology, 5,* 1432. doi:10.3389/fpsyg.2014.01432

Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical

model for estimating response time distributions. *Psychonomic Bulletin & Review, 12,* 195-223. doi:10.3758/BF03257252

Sheridan, H. (2013). The Time-course of Lexical Influences on Fixation Durations during Reading: Evidence from Distributional Analyses. Doctoral Dissertation, University of Toronto. Retrieved from https://tspace.library.utoronto.ca/bitstream/1807/35996/6/Sheridan_Heathe_201306_PhD_thesis.pdf

Sheridan, H., & Reingold, E. M. (2013). Distinct stages of word identification during reading: Evidence from eye movements. *Journal of Vision, 13,* 511-511. doi:10.1167/13.9.511

Sheridan, H., Rayner, K., & Reingold, E. M. (2013). Unsegmented text delays word identification: Evidence from a survival analysis of fixation durations. *Visual Cognition, 21,* 38-60. doi:10.1080/13506285.2013.767296

Houpt, J. W., & Townsend J. T. (2010). The statistical properties of the survivor interaction contrast. *Journal of Mathematical Psychology, 54,* 446-453. doi:10.1016/j.jmp.2010.06.006

Houpt, J. W., &. Townsend, J. T. (2011). An extension of SIC predictions to the Wiener coactive model, *Journal of Mathematical Psychology, 55,* 267-270. doi:dx.doi.org/10.1016/j.jmp.2011.02.002.

Van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology* (3rd ed., pp. 461-516). San Diego, CA: Academic Press.

Vincent, S. B. (1912). The function of vibrissae in the behavior of the white rat. *Behavioral Monographs, 1* (Whole No. 5).

Author Note

The research reported in this article has been partially supported by Grant PSI2014-53444-P from the Spanish Ministry of Economy and Competitiveness. Pablo Gomez was the recipient of the *Convitat* grant at the Universitat de Vàlencia.

Figure Captions

Figure 1. The figure (taken from Sheridan, 2013, p. 27) shows the distributions of first-fixation duration on target words in the low and high predictability conditions in the top panel, and the survival curves in the bottom panel. The row of points at the top of the survival curves indicates the time bins with a significant difference between the low and high predictability curves using the method being examined in the present note

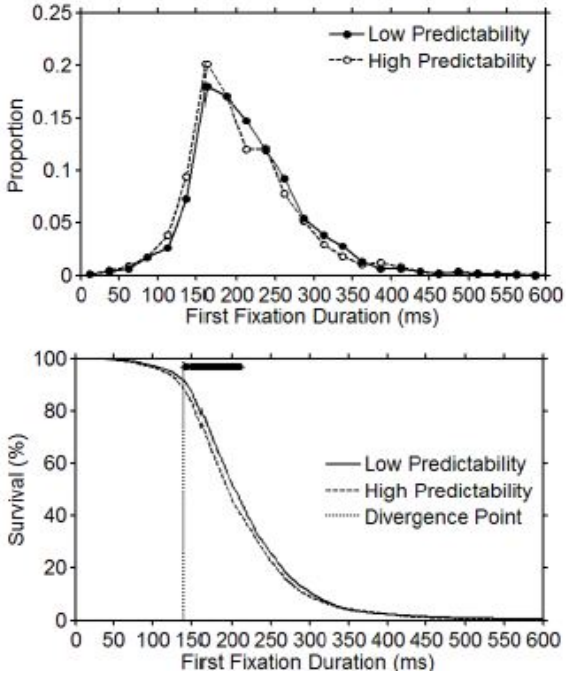
Figure 2. The figure shows the density, cumulative density function and survival function of two distributions. The difference between the two distributions (the simulated effect) is in the rate (τ) parameter of exponential component.

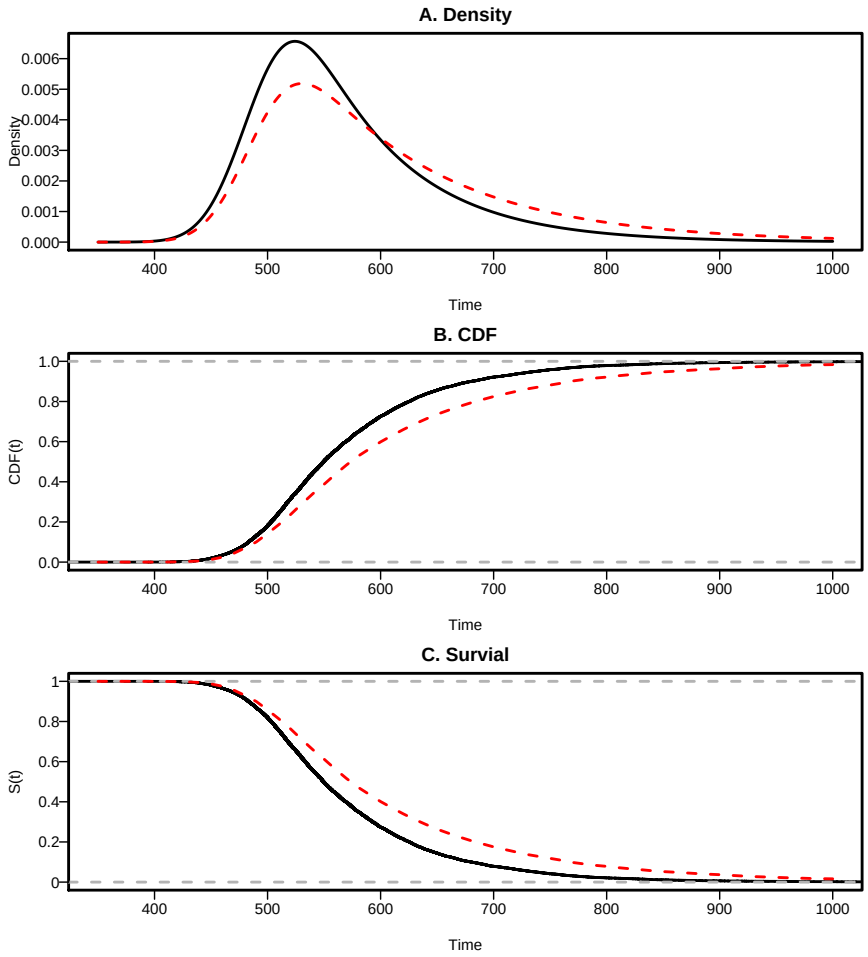
Figure 3. The figure shows the density, cumulative density function and survival function of two distributions. The difference between the two distributions (the simulated effect) is in the rate (μ) parameter of Gaussian component.

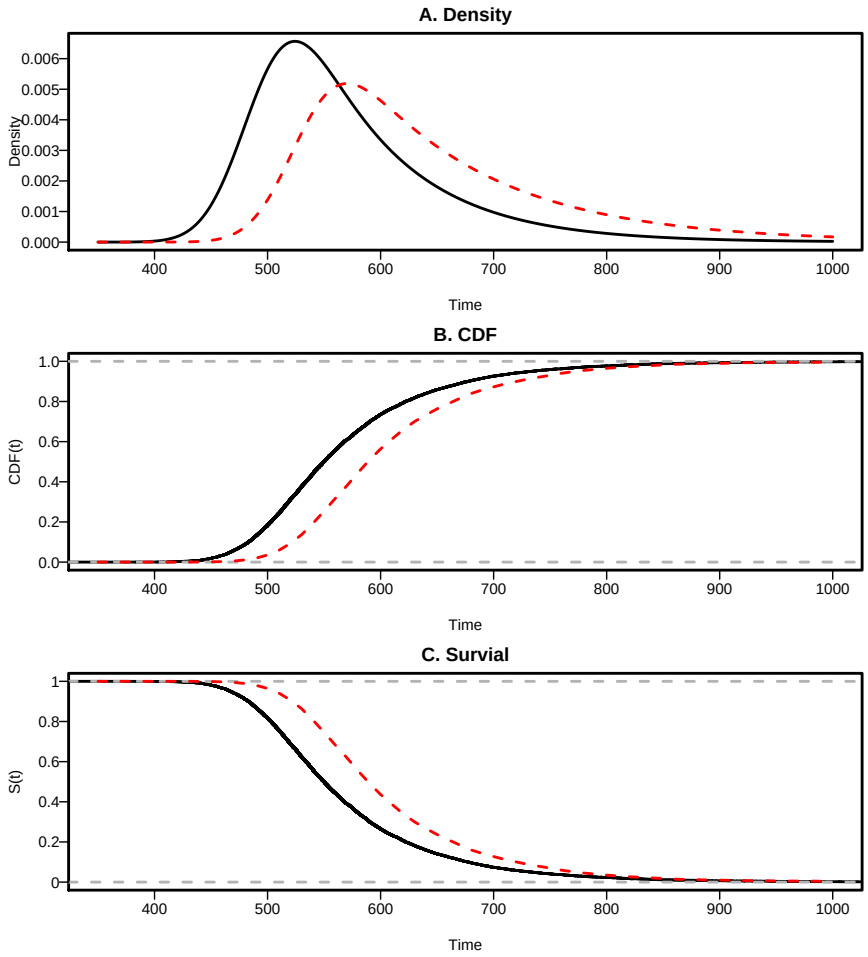
Figure 4. The figure shows the density, cumulative density function and survival function of two distributions in which there is a divergence point (525 ms; the location of the vertical lines in Panels B and C)

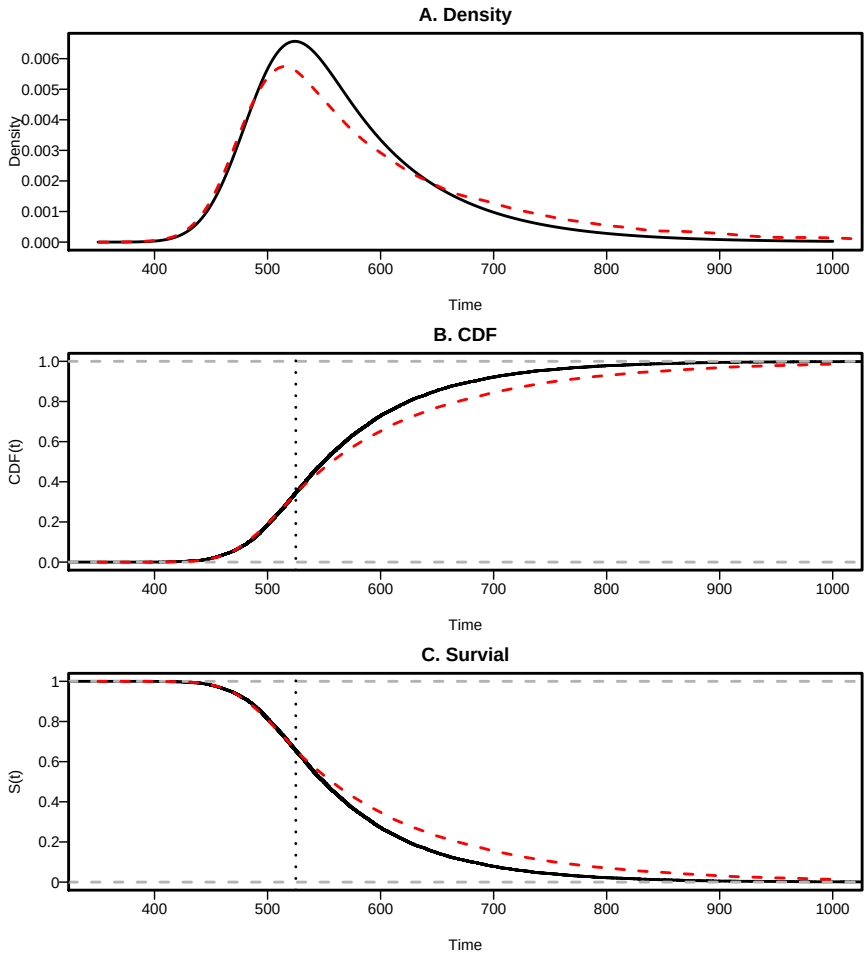
Figure 5. The figure shows average divergence point for simulated data using ex-Gaussian distributions with effects on μ. The points represent the size of the effect in μ (2:20 ms; 4: 40 ms; 8: 80 ms).

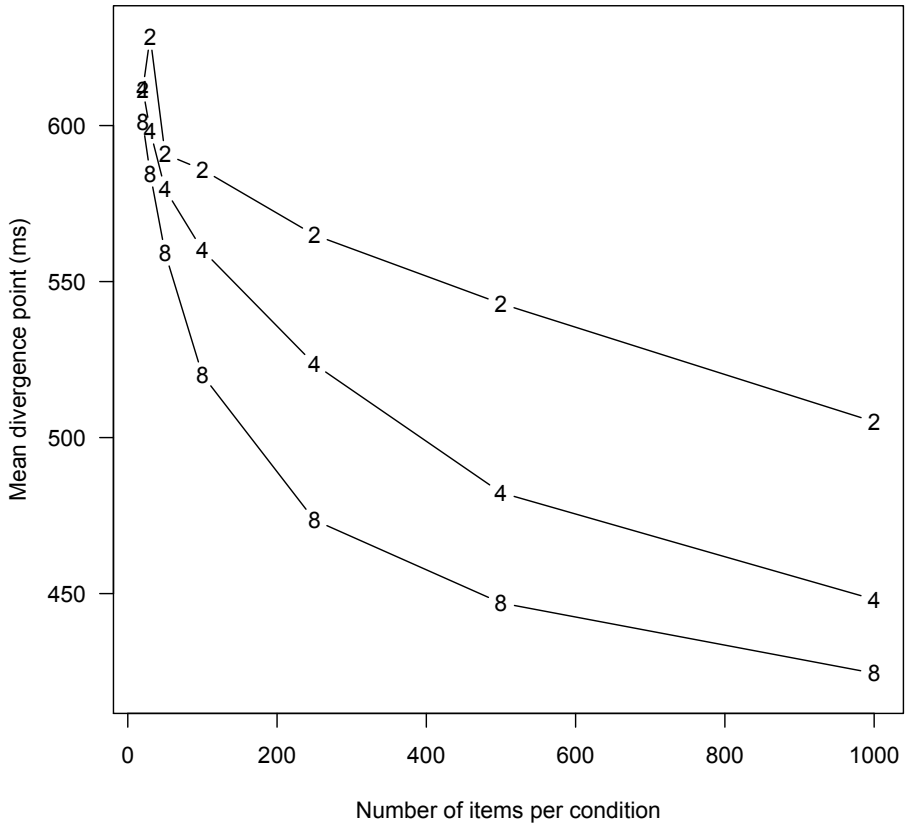
Figure 6. The figure shows average divergence point for simulated data using ex-Gaussian distributions with effects on τ. The points represent the size of the effect in τ (2:20 ms; 4: 40 ms; 8: 80 ms).

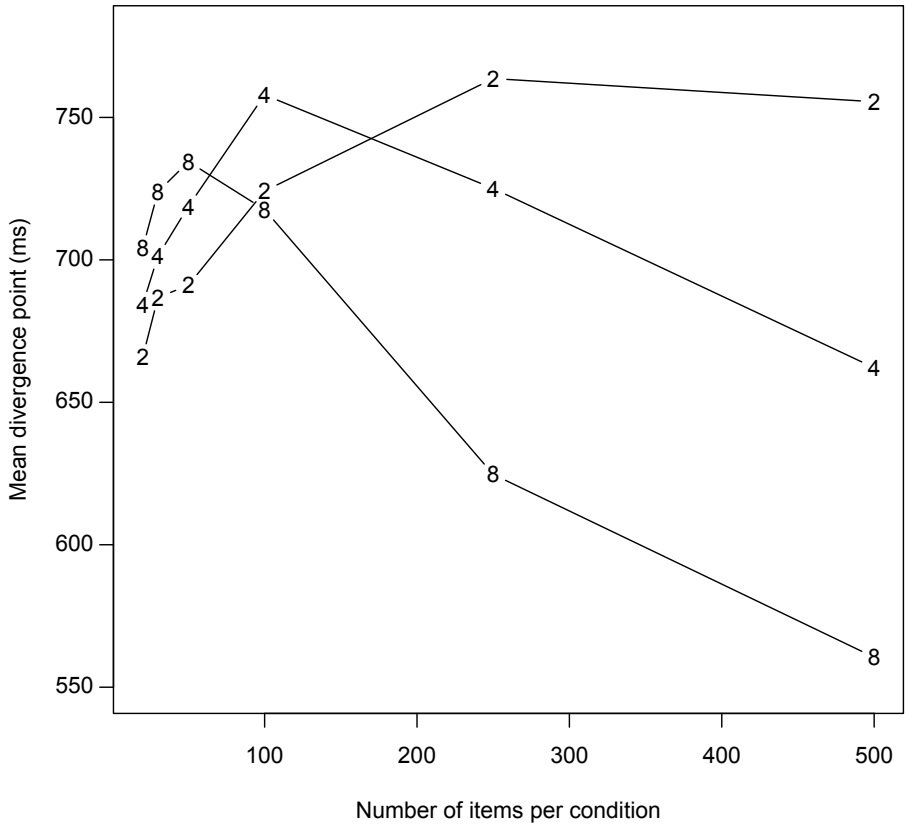

Low Predictability
High Predictability
Proportion
0.25
0.2
0.15
0.1
0.05
0
0 50 100 150 200 250 300 350 400 450 500 550 600
First Fixation Duration (ms)
Low Predictability
High Predictability
Divergence Point
Survival (%)
100
80
60
40
20
0
0 50 100 150 200 250 300 350 400 450 500 550 600
First Fixation Duration (ms)


A. Density
Density
0.006
0.005
0.004
0.003
0.002
0.001
0.000
400
500
600
700
800
900
1000
Time
B. CDF
CDF(t)
1.0
0.8
0.6
0.4
0.2
0.0
400
500
600
700
800
900
1000
Time
C. Survial
S(t)
1
0.8
0.6
0.4
0.2
0
400
500
600
700
800
900
1000
Time


A. Density
Density
0.006
0.005
0.004
0.003
0.002
0.001
0.000
400
500
600
700
800
900
1000
Time
B. CDF
CDF(t)
1.0
0.8
0.6
0.4
0.2
0.0
400
500
600
700
800
900
1000
Time
C. Survial
S(t)
1
0.8
0.6
0.4
0.2
0
400
500
600
700
800
900
1000
Time


A. Density
Density
0.006
0.005
0.004
0.003
0.002
0.001
0.000
400
500
600
700
800
900
1000
Time
B. CDF
CDF(t)
1.0
0.8
0.6
0.4
0.2
0.0
400
500
600
700
800
900
1000
Time
C. Survial
S(t)
1
0.8
0.6
0.4
0.2
0
400
500
600
700
800
900
1000
Time


Mean divergence point (ms)
600
550
500
450
0
200
400
600
800
1000
Number of items per condition


Mean divergence point (ms)
750
700
650
600
550
100
200
300
400
500
Number of items per condition