How to Fit a Response Time Distribution

Trisha Van Zandt

The Johns Hopkins University

Running Head: RT Distributions

Address correspondence to:
Trisha Van Zandt
Department of Psychology
The Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218

Abstract

Among the most valuable tools in behavioral science is statistically fitting mathematical models of cognition to data, response time distributions in particular. However, techniques for fitting distributions vary widely and little is known about the efficacy of different techniques. In this article, we assessed several fitting techniques by simulating six widely cited models of response time and using the fitting procedures to recover model parameters. The techniques include the maximization of likelihood and least-squares fits of the theoretical distributions to different empirical estimates of the simulated distributions. A running example was used to illustrate the different estimation and fitting procedures. The simulation studies revealed that empirical density estimates are biased even for very large sample sizes. Some fitting techniques yielded more accurate and less variable parameter estimates than others. Methods that involved least-squares fits to density estimates generally yielded very poor parameter estimates.

How to Fit a Response Time Distribution

The importance of considering the entire response time (RT) distribution in testing formal models of cognition is now widely appreciated. Fitting a model to mean RT alone can mask important details of the data that examination of the entire distribution would reveal, such as the behavior of fast and slow responses across the conditions of an experiment (e.g., Heathcote, Popiel & Mewhort, 1991), the extent of facilitation between perceptual channels (Miller, 1982), and the effects of practice on RT quantiles (Logan, 1992). Techniques for testing hypotheses based on the RT distribution have been developed (Townsend, 1990). In addition, the RT distribution provides an important meeting ground between theory and data; the ability of a model to predict the observed shape of the RT distribution is seen as a critical test of that model (Luce, 1986).

Many models state explicitly the characteristics of RT by specifying it as a random variable. All of the information about a random variable is contained in its probability density function (density, for short) or cumulative distribution function (or CDF).[1] The density represents the likelihood that an RT is observed within some arbitrarily small window of time, whereas the CDF represents the probability that an RT is less than or equal to some specific time. Most models of RT predict CDFs that are ogival: monotonic, nondecreasing S-shaped functions that begin at zero and asymptote at one. The RT densities predicted by most models are, in contrast, bell-shaped: nonmonotonic, usually positively skewed and unimodal. Despite their differences, the density and CDFs are mathematically equivalent ways of stating the properties of a random variable.

The shape of the density often provides clues to the kind of random variable and therefore potentially the candidate processes underlying the execution of a particular RT task. To investigate density shape from a

sample of RTs requires that those RTs be used as the basis of an estimate of the density. As we will demonstrate in this paper, some density estimation techniques more accurately recover the true shape of the density than others. In addition, it is often a goal of analysis to fit or estimate the parameters of a model, which also may require an estimate of the density or CDF. The model-fitting strategy used will determine the accuracy of the estimated parameter values. Although a number of computer programs are now available to assist researchers in plotting and performing fits of models to RT distributions (Cousineau & Larochelle, 1997; Dawson, 1988; Heathcote, 1996), no detailed assessment of the techniques most commonly used in density and parameter estimation currently exists in the psychological literature, especially in the context of the most likely models to be examined in such estimations. The purpose of this article is to provide such an evaluation.

Two separate but related estimation issues must to be addressed. First is the purely descriptive problem of accurately estimating the shape of the distribution from which a sample was derived, without making assumptions about the functional form of the population. This problem has been the focus of most of the work in the statistical area of nonparametric density estimation (Devroye, 1987; Silverman, 1986; Tapia & Thompson, 1978). Second is the analytical problem of accurately estimating the parameters of a presumed distribution, or, simply, fitting a model to the sample. To estimate these parameters, we may need to make use of a density or CDF estimate, the same estimate used to investigate the shape of the distribution, as we describe in some detail below.

In what follows, we first discuss the different ways to estimate the shape of an RT distribution (either the density or the CDF) and the properties of each estimator. Using a small sample, we illustrate the techniques by which

densities and CDFs are estimated. We then present a simulation study, in which the quality of each estimator is explored. We explain how models are fit to data, and then fit models to the simulated data to test the accuracy of different parameter estimation techniques. The study had two goals: first, to determine which of the nonparametric density estimators most popular in psychological research are the most accurate; second, to determine how best to estimate the parameters of a model using these density estimates. We also explore the effects of sample size on the accuracy of the density and parameter estimates. The results demonstrate that certain density estimators, including the popular Vincent estimator (Ratcliff, 1979), are highly biased estimators of the true density function, and that model fits to density estimates (rather than maximum likelihood estimates or fits to estimates of the CDF) often fail to recover accurate parameter values.

Density and CDF Estimators

There is a large statistical literature concerning various parametric and nonparametric density function estimators (see, e.g., Tapia & Thompson, 1978 for a historical review). Parametric density estimators make assumptions about the functional form of the empirical density, and the estimate is constructed by finding the best parameters for the density function given the data. For RT data, the most common parametric estimate of the density is the ex-Gaussian, the convolution of a normal and an exponential random variable. This distribution has been shown to fit empirical RT distributions well, and so it has been used to characterize RT densities (Hockley, 1984; Juhel, 1993; Rohrer & Wixted, 1994; Spieler, Balota & Faust, 1996; Wixted & Rohrer, 1993), even when the ex-Gaussian is not generated by the model of interest (Heathcote et al., 1991; Ratcliff, 1978;

Ratcliff & Murdock, 1976). Indeed, there is good evidence that many RT distributions are not distributed as ex-Gaussians, at least for weak signals (Burbeck & Luce, 1982; Green & Luce, 1971, see Luce, 1986 for a review).

In contrast, nonparametric estimators use the characteristics of the data to arrive at an optimal shape without making assumptions about a particular functional form. There are many ways to estimate the density function nonparametrically. These include histograms, kernel estimates, nearest neighbor estimates and orthogonal series estimates, among others. We are most interested in histogram and kernel estimators, the two types of estimates most commonly found in RT literature (e.g., Ashby, Tein & Balakrishnan, 1993; Green & Smith, 1982; Tolhurst, 1975).

For the following discussion, it will be useful to establish some notation. The time taken to make a response is indicated by the random variable T. Let the data observed be represented by the random vector $\mathbf{T} = \{T_1, T_2, \ldots, T_N\}$, where N is the total number of observations, and T_i is the i^{th} RT observed. We call $\mathbf{T}$ an independent and identically distributed (*iid*) sample: each T_i is assumed to be statistically independent from every other T_j, and all the T_is are generated by exactly the same process. This process produces RTs with CDF $F_T(t)$ and density function $f_T(t)$, and the data $\mathbf{T}$ are a sample from this distribution. The goal is to estimate F_T and/or f_T using the sample $\mathbf{T}$. An estimate of either F_T or f_T will be denoted as $\hat{F}_T$ or $\hat{f}_T$.

If the observations in $\mathbf{T}$ are ordered from smallest to largest, we can examine the properties of $T_{(i)}$, the i^{th} largest observation or *order statistic* in the sample. Let the order statistics of the data be represented by the vector $\{T_{(1)}, T_{(2)}, \ldots, T_{(M)}\}$, and the number of times each statistic is observed in the vector $\mathbf{T}$ by $\{n_1, n_2, \ldots, n_M\}$, where $\sum_{i=1}^{M} n_i = N$. To construct an estimate it may be necessary to specify an array of points

$\mathbf{t} = \{t_0, t_1, t_2, \ldots, t_r\}$ distributed over the range of T. The number of observations m_i in the half-open interval $[t_{i-1}, t_i)$ (where the square bracket and the parenthesis indicate that the interval includes t_{i-1} but excludes t_i) will be denoted by the vector $\mathbf{m} = \{m_1, \ldots, m_r\}$, where $\sum_{i=1}^{r} m_i = N$.

Estimates of a parameter or a function are random variables. They may vary widely over different samples. A sample mean $\overline{X}$ is a good example of an estimate; it estimates the mean μ of the distribution from which the sample was taken. Different samples from the same distribution will yield different values of $\overline{X}$. For our purposes, we must define two important properties of estimators. First, an estimator is unbiased if, on average, it equals the value of the parameter or function that it estimates. That is,

$$E[\hat{f}] = f,$$

where the notation $E[X]$ represents the expected value (mean) of the variable X. The sample mean is an unbiased estimator: $E[\overline{X}] = \mu$. This means that the average of all the sample means computed from all possible samples of size N drawn from a distribution with mean μ will equal μ. Second, an estimator is consistent if it "converges in probability" to the parameter or function that it estimates. In other words, the probability that the estimate differs from the thing estimated becomes zero as the sample size N grows:

$$\lim_{N \to \infty} P(|f - \hat{f}| > \epsilon) = 0$$

for any positive number ϵ, no matter how small. The sample mean is also a consistent estimator of the population mean μ. If an estimator is consistent, we can be confident that the error in the estimate is getting smaller as N gets larger.

Another important characteristic of an estimator is whether it maximizes the likelihood of observing the data. The probability of the data is the joint probability of sampling the entire set of observations from T_1 to T_N:

$P(\mathbf{T}) = P(T_1 \cap T_2 \cap \cdots \cap T_N)$. Because the observations are assumed to be independent from each other, the probability of the sample is therefore $P(T_1)P(T_2)\cdots P(T_N) = \prod_{i=1}^{N} P(T_i)$, the product of the probabilities of each separate observation. For a particular set of parameter estimates $\hat{\theta} = \{\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n\}$, the likelihood of the sample $\mathbf{T}$ is given by

$$L(\hat{\theta}) = \prod_{i=1}^{N} f_T(T_i; \hat{\theta}), \tag{1}$$

where $f_T(t; \hat{\theta})$ is the density function of the model being fit to the data, which replaces the $P(T)$ notation used above. The function L is a measure of how likely a particular data set is for a particular set of parameters $\hat{\theta}$. If $\hat{\theta}$ is chosen such that $L(\hat{\theta})$ is as large as possible, then $\hat{\theta}$ is called a maximum likelihood estimate of the true parameters $\theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. Maximum likelihood estimates do not necessarily exist, and they are not necessarily unique; under general conditions, they are consistent (Bickel & Doksum, 1977).

Other important characteristics of estimators warrant concern, such as their relative efficiency (their variance relative to the variance of any other estimator of the same parameter or function) and sufficiency (whether or not the estimate can be improved by considering some other aspect of the data). We will concern ourselves with unbiasedness and consistency primarily. Although unbiased estimators are usually preferable to biased ones, the use of a biased, consistent estimator may be warranted if it is more efficient than an unbiased estimator. For kernel estimators, which we present below, there are only small differences in efficiency across the different alternatives. Interested readers should consult Silverman (1986) for a basic treatment of these issues, or Devroye (1987) or Tapia and Thompson (1978) for a more advanced discussion.

We will now present the computations for estimating CDF and density

functions. To motivate this discussion, we use a working example in the form of a small ($N = 50$) data set sampled from a Weibull distribution (see Table 1). We first discuss the CDF, and then go on to the density functions.

The CDF

The CDF $F_T(t)$ has a natural nonparametric estimator in the cumulative relative frequency. The cumulative relative frequency is given by

$$\hat{F}_T(t) = \begin{cases} 0 & t < T_{(1)} \\ \frac{1}{N}\sum_{i=1}^{k} n_i & T_{(k)} \leq t < T_{(k+1)} \\ 1 & T_{(M)} \leq t \end{cases} , \tag{2}$$

and is computed for our Weibull sample and shown in Table 1, column 3.

Constructing this estimate is straightforward, and likely familiar to most readers. Suppose, for instance, we wish to estimate the value of $F_T(t)$ for $t = 756$. Using Table 1, we see that 756 lies between the 19^{th} and 20^{th} order statistics. Therefore, from Equation 2,

$$\hat{F}_T(756) = \frac{1}{50}\sum_{i=1}^{19} n_i = .40.$$

The true value of $F_T(756)$ is .396, using the parameters given in Table 3 and the expression for the Weibull CDF given in the Appendix.

The cumulative relative frequency estimator of the CDF (Equation 2) has a number of desirable properties. It is unbiased and consistent, and it is asymptotically normal (Bickel & Doksum, 1977). That is, for every value t, the distribution of $\hat{F}_T(t)$ will approach a normal distribution with mean $\mu = F_T(t)$ (the value of the true CDF at the point t) and standard deviation σ that we will not present here. The cumulative relative frequency estimate of F_T therefore is well-behaved and well-understood.

For model-fitting and graphical purposes, we need to determine the values of t for which $\hat{F}_T(t)$ is to be computed. We could examine all points

within a finite range, e.g., 500 ms, 501 ms, 502 ms, ..., 1200 ms. Or, we could examine all points that were observed in a sample, $\{T_{(1)}, T_{(2)}, \ldots, T_{(M)}\}$. Nothing prevents us from using a large number of points except computational expense. A more frugal approach is to select a small number of ts for which $F(t)$ is to be estimated across the range of the sample. We can do this by selecting a fixed step size (e.g., 50 ms) and origin (e.g., $t_0 = 500$ ms) as we do for the fixed-width histogram (discussed below). Or, equivalently, we could select ts based on the density of the sampled times, using percentiles (Logan, 1992; Van Zandt, Colonius & Proctor, 2000). Because of the direct relationship between percentiles and several of the density estimators, we chose to use percentiles. Also, by basing the estimated points on percentiles, we insure that we obtain estimates evenly over values of $\hat{F}_T(t)$, and so the shape of $F_T(t)$ is accurately estimated. The shape of the fixed-step CDF estimate will not necessarily be as accurate, because if too few or poorly located points are used it is possible to skip over most of the CDF. Using the percentiles has the added benefit of being automatic: there is no need to investigate each data set to determine the appropriate origin and step size, which undoubtedly differs across subjects and conditions.

For the percentile-based CDF estimate, the quantiles of the sample are computed that correspond to the desired points on the CDF estimate. We chose the deciles for the distribution estimates in this paper, plus the 0^{th} and 100^{th} percentiles. Using the data in Table 1 and averaging over times with a common percentile rank, the deciles are computed and shown in Table 2. Consider, for example, the 30^{th} percentile. All times t between 732 and 735 could be the 30^{th} percentile, i.e., $P(RT < t) = .30$. Therefore, we will estimate the 30^{th} percentile using the midpoint between 732 and 735:

$$\frac{732 + 735}{2} = 733.5.$$

The true value of the 30^{th} percentile is 734.774, using the parameters given in

Table 3.

The estimated CDF can be viewed by plotting the computed percentiles on the x-axis and the corresponding percentile ranks on the y-axis. In Figure 1, the points on the y-axis are always the same (0, .1, .2, etc.), but the corresponding points on the x-axis are taken from Table 2. The points on the x-axis will vary from sample to sample. The solid line is the true CDF computed using the parameters in Table 3 for the Weibull distribution.

The Density Function

We focus our attention next on histogram and kernel estimates, the two types of density estimates most commonly found in RT studies. It should be noted that there are other density estimators available, most notably the naive or Rosenblatt estimator, nearest-neighbor estimators, and spline and orthogonal series estimators (see Silverman, 1986 or Tapia & Thompson, 1978 for discussion of these alternatives), which we will not discuss because they have not previously been used in a psychological context. Further, none of these alternatives have significant advantages over the estimators we discuss here.

Histogram Estimates

We will examine two types of histogram estimators. The first is the standard *fixed-width* histogram, in which observations are combined into bins of equal width, and the second is the more general *variable-width* histogram, in which bins are not necessarily equally spaced. The array of points $\mathbf{t} = \{t_0, t_1, \ldots, t_r\}$ defines the bin boundaries of the histogram, and the height of the bars depict density, such that the bar areas are equal to the proportion of observations that fall within each region.[2] For example, if a bin stretches from 500 ms to 550 ms, and 25% of the observations fall in that bin, the height of the bar would be 1/200 ($.25/(550 - 500) = 1/200$).

Using $\mathbf{t} = \{t_0, t_1, \ldots, t_r\}$, where t_0 represents the origin of the histogram or the smallest value of T to be considered, we may write the general histogram estimate as

$$\hat{f}_T(t) = \frac{1}{N} \sum_{i=0}^{N} \frac{I\{T_i, [t_{k-1}, t_k)\}}{t_k - t_{k-1}} \text{ for } t \in [t_{k-1}, t_k), \tag{3}$$

where $I\{T_i, [t_{k-1}, t_k)\}$ is the indicator function

$$I\{T_i, [t_{k-1}, t_k)\} = \begin{cases} 1 & \text{if } T_i \in [t_{k-1}, t_k) \\ 0 & \text{otherwise} \end{cases} .$$

Equation 3 is a rather obtuse way of stating that the height of the histogram estimate is proportional to the number of observations within the bin covered by the bar. The indicator function allows us to count the number of observations that fall in the same bin as the selected value of t ($[t_k, t_{k+1})$). Each T_i that is in the same bin as t causes the indicator function to add 1 to the sum in Equation 3. For t contained within the kth interval $[t_{k-1}, t_k)$, the height of the density estimate (histogram) at that point equals $m_k/[N \cdot (t_k - t_{k-1})]$. Writing the histogram estimate as in Equation 3 makes obvious its close relationship to the kernel estimate, as we will see below.

For the fixed-width histogram, the points $\mathbf{t} = \{t_0, t_1, \ldots, t_r\}$ are separated by a constant distance h_N, so that $t_{k+1} = t_k + h_N$. The parameter h_N is called a *smoothing parameter*: the larger it is, the smoother the final density estimate will be. We selected an origin $t_0 = 500\ ms$ and a step size $h_N = 50\ ms$ for all the histogram estimates regardless of sample size. The upper and lower limits of each bin are therefore defined by $\mathbf{t} = \{500, 550, 600, 650, \ldots\}$, and the histogram bars are centered on $\{525, 575, 625, \ldots\}$. Returning to our example (Table 1), the number of observations within each bin is $\mathbf{m} = \{0, 0, 0, 6, 12, \ldots\}$. The first bar of nonzero height will be centered on 675 ms. There are $m_4 = 6$ observations that occur in the fourth interval $[650, 700)$. Using Equation 3,

$\hat{f}(t) = \frac{6}{(50)(50)} = .0024$ for all $t \in [650, 700)$. The next bar, centered on 725 ms, has height $\hat{f}(t) = \frac{12}{(50)(50)} = .0048$ for all $t \in [700, 750)$. Continuing in this way, we obtain the fixed width histogram estimate shown in Figure 2, panel A. Superimposed on the estimate is the density for the Weibull distribution from which the observations were sampled.

Along with the fixed-width histogram estimator, we will examine two variable-width histogram estimators. One we will call the "decile" estimate, which defines 11 boundary points: t_1, t_2, ..., t_9 are the 9 deciles (10%, 20%, ..., 90%) computed from the sample in the same way as for the CDF, plus the minimum ($t_0 = T_{(0)}$) and maximum ($t_{10} = T_{(M)}$) observed value of T. The other is the better-known "Vincent" histogram, introduced by Ratcliff (1979), in which the boundary points are also defined by ten (different) quantiles (approximately 5%, 15%, ..., 95%), and the minimum and maximum are not used in the histogram. The strength of the Vincent estimator is that it allows for distributional averaging over a group of subjects or conditions, and this is the context in which it is typically used (e.g., Balota & Spieler, 1999; Spieler et al., 1996). We will use the Vincent histogram as a nonparametric density estimator without averaging.

For the decile estimate, the deciles shown in Table 2 are used as the real upper and lower limits of the histogram bars. The height of each bar is chosen so that the area of the bar is equal to .10, the relative frequency of observations falling within each region. The bar is then centered at the midpoint of the upper and lower limit. For example, the first bar of the estimate will extend from 673.0 ms to 697.5 ms, a distance of 24.5 ms, and it will be centered on 685.25 ms. The estimated density, the height of the bar, is $\hat{f}(t) = \frac{.10}{24.5} = .0041$ for all $t \in [673.0, 697.5)$. Similarly, the next bar extends from 697.5 ms to 712.0 ms, a width of 14.5 ms, and is centered on 704.75 ms. The height of the bar is $\hat{f}(t) = \frac{.10}{14.5} = .0069$ for all

$t \in [697.5, 712.0)$. Continuing in this way, we obtain the decile histogram estimate shown in Figure 2, panel B.

The Vincent estimate is obtained in a manner very similar to the decile estimate. First, however, the "vincentiles"[3] must be computed. Following Vincent's procedure (e.g., Ratcliff, 1979), each of the observations given in Table 1 are duplicated 10 times (one duplication for each desired vincentile). Then, the first 50 (N) duplicated observations are averaged to give the first vincentile. The number so obtained –

$$\frac{1}{50}(10 \cdot 673 + 10 \cdot 676 + 10 \cdot 680 + 10 \cdot 687 + 10 \cdot 697) = 682.6\ ms$$

– is supposed to estimate the 5^{th} percentile.[4] The next 50 observations are averaged to give 705.4 ms, an estimate of the 15^{th} percentile, and so on, to obtain the vincentiles shown in Table 2.

To construct the Vincent estimate, histogram bars of equal area (.10, in our case) are drawn with upper and lower limits given by the vincentiles. So, the first bar of the estimate has a height of $.10/(705.4 - 682.6) = .0044$. This bar will be centered at $(682.6\ ms + 705.4\ ms)/2 = 694\ ms$, the midpoint of the interval between the first and second vincentile, corresponding roughly to the 10^{th} percentile. Using the vincentiles in Table 2, we construct 9 bars encompassing 90% of the sample. The upper and lower 5% of the data are not represented in the density estimate, although they contributed to the computation of the vincentiles. The Vincent estimate is shown in Figure 2, panel C.

Gaussian Kernel Estimator

The fourth density estimator we will consider is the Gaussian kernel estimate (Parzen, 1962), given by

$$\hat{f}_T(t) = \frac{1}{N h_N} \sum_{i=1}^{N} \phi\left(\frac{T_i - t}{h_N}\right). \tag{4}$$

The function ϕ is the standard Gaussian density:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

For any value t, the estimate of the density at t is constructed by centering a Gaussian density over t. Each observation in the sample contributes an amount to the estimate equal to the height of the Gaussian curve at the observation's location. To compute the smoothing parameter h_N, we used Silverman's (1986) method:

$$h_N = \frac{0.3}{N^{.2}} \min\left(s, \frac{q_3 - q_1}{1.349}\right),$$

where q_1, q_3, and s are the first and third quartiles and standard deviation of the sample, respectively.

For the sample shown in Table 1, $h_N = 18.0442$. Consider the estimated density for $t = 690$ ms. From Table 1, there is an observation of 673 ms. This observation contributes $\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\left(\frac{673-t}{h_N}\right)^2) = .26$ to the estimate. The next observation, 676 ms, contributes $\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\left(\frac{676-t}{h_N}\right)^2) = .30$. We continue in this way throughout all the observations in the sample, and compute $\hat{f}(690) = \frac{1}{Nh_N}(.26 + .30 + \ldots) = .0039$ (see Equation 4). To determine the estimated density at the point 720 ms the Gaussian is shifted to 720 ms and the process is repeated, giving $\hat{f}(720) = .0050$. The Gaussian kernel density estimate for the sample in Table 1 is shown in Figure 2, panel D, for 60 equally-spaced points along the time axis.

Like the CDF estimate, the number of points at which the Gaussian kernel estimate is computed can be very large or small. The only concerns in deciding which points to use are the computational expense and the ability of the estimate to represent the true density. We selected the number of points (60) relatively arbitrarily, although it served our purposes well. For visualizing density functions, the number of points used doesn't matter much, except that enough points be used. There may well be some tradeoffs,

however, if too many or too few points are used in model-fitting attempts, discussed below.

Properties of the Estimators

In Figure 2 it is clear that different density estimators can sometimes lead to very different ideas about the properties of a distribution. The fixed-width histogram estimate in panel A shows only a hint of bimodality, but the Gaussian kernel estimate in panel D shows two clear modes, and potentially a third in the upper tail. We have already presented the properties that make an estimator desirable – unbiasedness and consistency – and we will now discuss these properties in the context of nonparametric density estimators.

The histogram and Gaussian kernel estimators are general weight function estimators (Silverman, 1986). This means that each can be written as

$$\hat{f}(t) = \frac{1}{N} \sum_{i=1}^{N} w(T_i, t),$$

where the weight function $w(x, t)$ is positive and integrates to 1 over t. This insures that the estimate $\hat{f}_T$ is itself a probability density. Under general conditions on the function w, which are satisfied for the estimators we are examining, we can state some general properties of the histogram and kernel estimators. Note that the critical parameter of the estimators is h_N, explicit in the Gaussian kernel and fixed-width histogram but not in the variable-width histograms. The parameter h_N determines the extent to which the estimate is smoothed; for example, in the Gaussian kernel estimator h_N is the standard deviation of a Gaussian distribution. The larger h_N is, the smoother the estimate will be. For the histograms, h_N determines the width of the bins.

Unlike the CDF estimator $\hat{F}(t)$, the density estimators are not usually unbiased or consistent. If $\lim_{N\to\infty} h_N = 0$ then the estimator $\hat{f}_T$ will be

asymptotically unbiased. This means that if N is large enough, the average of the estimate will equal the true density: $\lim_{N\to\infty} E[\hat{f}_T] = f_T$. If, as well, $\lim_{N\to\infty} Nh_N = \infty$, then $\hat{f}_T$ is consistent (Parzen, 1962). The important thing to note is that the asymptotic behavior of $\hat{f}_T$ is not determined by N except through changes in h_N. If h_N is not changing appropriately with N, increasing the sample size will not always help to decrease the bias of the estimate.

This fact warrants repeating. Usually one may be assured of the robustness of common statistical procedures when the sample size is large enough. Means become normal, variance ratios become distributed like $F(df_1, df_2)$, and so forth. The comfortable habit of increasing N to safeguard against unknown violations of statistical assumptions will not improve the accuracy of a density estimator if the smoothing parameter of that estimator is chosen poorly. Furthermore, the extent of bias (and hence the optimal h_N) depends on the underlying and presumably unknown density function f_T. The accurate estimation of a density function is not, therefore, a trivial matter.

Of the three density estimators discussed, only the Gaussian kernel estimator has the benefit of not only satisfying the conditions for asymptotic unbiasedness and consistency, but the mean integrated squared error (the continuous version of the sum of squared errors) decreases faster than the histogram estimators with increases in N. That is, it is more efficient than any of the histogram estimators. It also has the benefit of being automatic; the parameter h_N can be selected according to the statistics of the sample. Similarly, the variable-width Vincent and decile histograms are automatic; the bin boundaries are determined by the statistics of the sample. This is not true of the fixed-width histogram, where the origin and bin width may need to be adjusted based on the characteristics of a particular sample. The

importance of the existence of an automatic strategy for selecting the smoothing parameter should not be underestimated. Consider, for example, how one might select the parameter h_N if (as in the present case) one had several dozen densities to estimate. In Figure 2, for instance, should h_N be made relatively small to reflect the multimodel nature of the sample, or should it be made relatively large to reflect the unimodal nature of the population? How do we decide when we don't know f_T?

Because it is asymptotically unbiased, consistent and efficient, the Gaussian kernel estimator should more accurately recover the true underlying density than the histogram estimators. Furthermore, any parameter estimation technique dependent on a density estimate will be less accurate than a parameter estimation technique dependent on the unbiased and consistent CDF estimate. Because the extent of bias and the rate with which bias decreases with N (if at all) depends on the shape of the true density function f_T, there will also be large, model-specific effects on accuracy of estimation using different techniques. In the rest of this article, we seek to answer the following three questions: 1) Which density estimate provides the most accurate recovery of the shape of the underlying density? 2) Can accurate parameter estimation techniques be based on any of these density estimates? 3) What is the minimum sample size necessary to achieve a reasonable degree of accuracy in either the density or the parameter estimates?

Methods

RT data from six models were generated. The equations for the densities and CDFs for each model are given in the Appendix. Each distribution arises from a well-established cognitive model. The diffusion model (Ratcliff, 1978)

has been used to explain RT and accuracy in a wide range of choice response experiments. The ex-Gaussian model was originally proposed to represent a two-stage cognitive process, in which the time to make a decision was exponentially distributed (Hohle, 1965; or normally distributed, McGill, 1963) and the sum of the remaining perceptual and response processing times was normally distributed (Hohle, 1965; or exponentially distributed, McGill, 1963). The gamma model represents the processing time for a number of serially arranged exponential processing stages (McClelland, 1979; McGill, 1963). The Poisson race model represents the accumulation of information on parallel perceptual (or response) channels (Pike, 1973; Townsend & Ashby, 1983; Van Zandt et al., 2000). The Wald distribution arises from a diffusion process with one absorbing boundary, and has been used to explain simple (detection) behavior (Emerson, 1970; Wald, 1947), as well as to describe the firing pattern of simple neurons (Rudd, 1996). Finally, the Weibull model describes RT (under certain conditions, see Colonius, 1995) when a large number of parallel processes compete to produce a response (Logan, 1988, 1992, 1995).

For each of the six models, 500 samples of size $N = 50$, 100, 500, 1,000, and 10,000 were generated (a total of $500 \times 5 = 2500$ samples for each model). For each of these samples, nonparametric distribution and density estimates were obtained as described above (5 nonparametric estimators $\times 2500$ samples $= 12,500$ estimates for each model). Then, using these distribution and density estimates, the original parameters of the models were recovered using seven parameter-estimation techniques (7×2500 samples = 17,500 recovered parameter sets for each model). We examined first the nonparametric distribution and density estimates, to determine how well each estimator recovered the true shape of the distribution or density from which the sample was derived (see Figures 4-13). Then, after fitting

each model back to the samples that each model generated, we examined how well the original parameters were recovered (see Figures 15-20). We will now describe the specific methods used in each step of this procedure.

Simulations

The models were simulated by sampling randomly from the appropriate distributions, with parameters constrained so that each simulated RT distribution had a mean $\mu = 800$ and variance $\sigma^2 = 100^2$, approximately what might be expected if RTs were measured in a choice response task. The parameter values for each model, plus a base time component, are given in Table 3. The skewness for each distribution varied between 1.0 and approximately 2.0, with the gamma having the least skew (1.0) and the diffusion having the most (approximately 1.7).

The workhorse of the simulations was a uniform random number generator presented by Press et al. ("ran3": Press, Teukolsky, Vetterling & Flannery, 1992), which is based on Knuth (1981). Gaussian random numbers were generated using the Box-Muller method (Knuth, 1981), by way of the routine also provided by Press et al. All the models were simulated by exploiting known relationships between the desired variable and the available uniform and Gaussian random variables. In the simplest cases, we used the method of transformation (Press et al., 1992).

Some unique problems arose with the analyses of the race and diffusion models that made it necessary to treat these models somewhat differently from the others. For these models, we needed to obtain accuracies as well as RT data from the simulations. The race model and the diffusion model both generate correct and incorrect responses, but only the RTs for correct responses were fit. For the race model, correct and incorrect RTs were generated naturally as the minimum statistic of two gammas; occasionally

the "incorrect" gamma was the smallest and so determined the RT. Accuracy for the race model was therefore determined by the number of times the correct gamma was faster than the incorrect gamma. For the diffusion, incorrect responses were generated according to a binomial process before each RT was simulated. A uniform random number was generated, and if that number was less than the probability of a correct response (as calculated from the parameters of the model), a correct RT was generated. Otherwise, an incorrect RT was generated. For both the race model and the diffusion model, the simulations proceeded until the desired sample size of correct RTs was obtained. In this way, we obtained both a sample of RTs and an accuracy for each sample.

For each model, 500 independent samples of sizes 50, 100, 500, 1,000, and 10,000 were generated (approximately 35×10^6 samples total, well within the cycle length of the random number generator). Observations were not duplicated within larger samples; e.g., the observations that made up the sample of size 50 were not the first 50 observations in the sample of size 10,000. Density and CDF estimates were computed for each model and each sample size.

Nonparametric Estimates

Nonparametric CDF and density estimates were obtained exactly as described in the previous section. For the fixed-width histogram estimates, observations were placed into bins 50 ms wide, with the first real lower limit at 500 ms. The height of each bar was equal to the relative frequency of observations within the bar divided by 50 ms (the width of the bar; see Equation 3). The quantile-based histograms were estimated in two ways, as described above. First, the deciles were estimated and used as the real upper and lower limits of ten histogram bars. Second, the procedure advocated by

Ratcliff (1979) was used. The major difference between the two techniques is the inclusion of the minimum and maximum in the decile estimate. The upper and lower 5% of the data are not shown in the Vincent estimate, meaning that only 90% of the total probability mass is represented in nine bars. The different techniques for computing quantiles and vincentiles result different bin boundaries. The Gaussian kernel estimator was computed according to Silverman (1986), as described above. The Gaussian estimates were calculated at 60 equally spaced points along the range of each sample.

The empirical CDF was estimated by the same deciles that were used to construct the decile histogram. The 0^{th} and 10^{th} (minimum and maximum) deciles were also included in the CDF estimate.

Fitting the Models

We used two basic techniques for fitting the models to each sample. The first technique was maximum likelihood estimation. The second technique was the method of least squares, which required the nonparametric distribution and density estimates. In all the applications in this paper, we fit each model to the data sets that were generated by it; we never fit an incorrect model. Our goal was to determine how well the parameters of the models could be recovered when those parameters were known. Most of these models can very easily be made to mimic each other (Van Zandt & Ratcliff, 1995; Van Zandt et al., 2000), and so we did not attempt to fit models to samples that they did not generate. Because the true model is known and the correct parameters are used to initiate the fits, our analyses represent a best-case scenario. We will now illustrate the maximum likelihood and least-squares techniques, returning once again to our example to make matters clear.

Maximum Likelihood

The goal of maximum likelihood estimation is to find estimates of the model parameters such that the probability of the sample is as great as possible. Maximum-likelihood estimation procedures do not require that a nonparametric estimate of the density or CDFs be computed from the data before a model is fit to the data. In essence, maximum-likelihood estimation is parametric in that a specific density must be selected to compute the likelihood $L(\hat{\theta})$ (see Equation 1). The parameters that maximize $L(\hat{\theta})$ can then be used (by plugging them back into the expression of the density) to generate a parametric density estimate, if desired.

After the model to be fit has been selected (the Weibull, for our example), an initial set of parameters must be chosen to begin to fit the model. For our fits, we always began with the correct parameters, those that were used to generate the data. The starting values for the Weibull fits were $\xi = 652.72$, $\alpha = 163.15$, and $k = 1.5$, or $\hat{\theta} = \{652.72, 163.15, 1.5\}$. The likelihood for these values is computed using Equation 1. For computational ease, we can work instead with the log-likelihood function, given by

$$\ln L(\hat{\theta}) = \sum_{i=1}^{N} \ln f_T(T_i; \hat{\theta}). \quad (5)$$

The parameters that maximize the log-likelihood function will also maximize the likelihood function. Each observation is entered into the density function, along with $\hat{\theta}$, to compute the log likelihood. Using Table 1, the parameter starting values, and the Weibull density from the Appendix,

$$\begin{aligned}
\ln L(\{652.72, 163.15, 1.5\}) &= \sum_{i=1}^{N} \ln f_T(T_i; \hat{\theta}) \\
&= \sum_{i=1}^{N} \ln \left\{ \frac{1.5}{163.15} \left(\frac{T_i - 652.72}{163.15} \right)^{1.5-1} e^{\left(\frac{T_i - 652.72}{163.15}\right)^{1.5}} \right\} \\
&= -5.69 + -5.61 + \cdots + -.97 \\
&= -286.61.
\end{aligned}$$

The parameters are then systematically adjusted in an attempt to make $\ln L(\hat{\theta})$ larger. (See Wickens, 1982, for a more general treatment of maximum likelihood in this context.)

The log-likelihood function for the Weibull distribution is a surface in four-dimensional space. Figure 3, panel A shows the log-likelihood surface produced by holding $\hat{k}$ fixed and varying $\hat{\xi}$ and $\hat{\alpha}$ for the sample in Table 1. Maximum likelihood estimation yields the values of $\hat{\xi}$, $\hat{\alpha}$ and $\hat{k}$ that give the largest likelihood, or the highest point on the surface shown in Figure 3. For our sample, the highest point on the surface is given by $\hat{\xi} = 668.89$, $\hat{\alpha} = 130.81$ and $\hat{k} = 1.47$.

In this example, and in this paper, maximum likelihood has been applied to the sample as a whole. There are instances where maximum likelihood has been applied to fit a model to a nonparametric distribution estimate (e.g., Heathcote, 1996). In these situations, the likelihood function is usually changed to

$$L(\hat{\theta}) = \prod_{i=0}^{r} f(t_i; \hat{\theta}),$$

where the likelihood is computed for a fixed number of estimated points t_i rather than the observations T_i. This is useful for the variable-width histograms when the estimate is actually an average obtained over subjects, as in Vincent averaging (Ratcliff, 1979). However, it is not recommended for single samples, because it requires that the sample be replaced with a small number of sample statistics to compute the likelihood (10, if the t_is represent 10 vincentiles, for example), resulting in a poorer estimate of the parameters than if the entire sample had been used.[5] Maximum likelihood also will not be of use for fixed-width estimates, such as the fixed-width histogram or the kernel estimator, because maximum likelihood depends on the way that observations are scattered. If all the t_is are equally spaced, every sample will look exactly the same regardless of the distribution from which it came.

There are other ways that likelihood might be computed from these estimates, but they are more complicated and rarely used. For these reasons, we did not attempt maximum likelihood estimates using nonparametric density estimates; least-squares estimates were used instead.

Least-Squares Estimation

The goal of least-squares estimation is to choose parameters such that the sum of squared error (SSE) between the model and the data is as small as possible. In our analyses, the data is a nonparametric CDF or density estimate, and the model is the functional form of the CDF or density. For a particular set of parameters $\hat{\theta} = \{\theta_1, \ldots, \theta_n\}$:

$$SSE(\hat{\theta}) = \sum_{i=1}^{r} (\hat{f}_T(t_i) - f_T(t_i; \hat{\theta}))^2,$$

where $\hat{f}_T$ is the empirical estimate of the density (or CDF), $f_T(x; \hat{\theta})$ is the theoretical density (or CDF) being fit, and $\mathbf{t} = \{t_0, t_1, t_2, \ldots, t_r\}$ is a fixed number of points spanning the range of the sample. Least-squares estimation therefore requires that a density or CDF estimate be used to recover the model parameters; rather than using the entire sample, the sample is summarized by the estimated height of the density at only a few points.

As for maximum likelihood, a model and starting values for its parameters must be selected. Also, a nonparametric estimate to which the model is to be fit must be selected. Then for some number of points on the nonparametric estimate, the corresponding predicted value of the model density is computed. The difference between the theoretical and estimated density is computed, squared, and added to the SSE. More specifically, using the same starting values as before, the true values of the Weibull parameters, we can compute $SSE(\hat{\theta})$ for the fit of the Weibull to the Gaussian kernel estimate (shown in Figure 2, panel D). The first point estimated by the

Gaussian kernel is $t = 667.30\ ms$. The height of the estimate at this point is .0018. Using the parameters given in Table 3 and the Weibull density given in the Appendix, the height of the curve at $t = 667.30\ ms$ is $\frac{1.5}{163.15}\left(\frac{667.30-652.72}{163.15}\right)^{1.5-1}\exp\{-\left(\frac{667.30-652.72}{163.15}\right)^{1.5}\} = .0027$. This point therefore contributes $(.0018 - .0027)^2 = 1 \times 10^{-6}$ to the SSE. Continuing in this way over all the points in the density estimate yields $SSE(\hat{\theta}) = 3.2 \times 10^{-5}$.

Just like the log-likelihood function, $SSE(\hat{\theta})$ for the Weibull is a surface in four-dimensional space. Figure 3, panel B shows the $SSE(\hat{\theta})$ surface produced by holding $\hat{k}$ fixed and varying $\hat{\xi}$ and $\hat{\alpha}$. The best fit of the model is given by those parameters that make $SSE(\hat{\theta})$ as small as possible. For our sample, those values are $\hat{\xi} = 603.64$, $\hat{\alpha} = 195.65$ and $\hat{k} = 2.39$. Notice that even though the model is the same and the starting values of the parameters are the same, the final recovered values of the parameters are very different for maximum likelihood and least-squares estimation.

Minimizing $SSE(\hat{\theta})$ for a model that is nonlinear in its parameters (as all of our models are) is called *nonlinear regression*. If the model has been correctly specified, and the error at each point to be fit is normally distributed with equal variance, then estimating θ, the true parameter vector, by minimizing $SSE(\hat{\theta})$ is desirable for a number of reasons. The resulting estimated set of parameters $\hat{\theta}$ is asymptotically normally distributed and unbiased. The estimate $\hat{\theta}$ has minimum variance, and maximizes the likelihood of the sample for the model being fit, i.e., minimizing $SSE(\hat{\theta})$ will give the same results as maximum likelihood estimation. However, the normal, equal variance errors assumption likely does not hold in our case, which raises a number of concerns about performing ordinary least-squares minimization as we have done.

The first concern is the homogeneity of variance of the errors. For fitting densities, we should not expect the variances to be equal; errors should be

larger at the tails of the density than at the mode because of the higher frequency of observations at the mode. The second concern is the correlation between the errors: are the errors uncorrelated with each other, or, if not, are they stationary, or do they depend on the point at which the function is estimated? In practice, it is impossible to answer these questions because the theoretical form of the density or CDF is not known. Therefore, we cannot compute the variance structure of the errors. With independent, heteroscedastic errors, there are a number of techniques that can be used, including weighted least-squares and linearization, that can convert the nonlinear, heteroscedastic regression problem into a linear, homoscedastic problem and so guarantee unbiased, maximum likelihood parameter estimates. However, it is likely also that the errors are correlated and nonstationary, in which case there is no straightforward technique to reduce the problem to the linear homogeneous case. As a result, the best approach is an ordinary (unweighted) least-squares minimization, with the understanding that standard inferential statistics on the resulting parameters will require more complicated analyses (Gallant, 1987).

Our results show that the ordinary, unweighted SSE minimization will result in reasonably good parameter values if the density/CDF estimates used in the minimization are appropriate. But, it should be noted that more sophisticated nonlinear regression techniques are available, and, should they be used, they might improve the ability to recover parameters using other estimates.

Typical Practice

It is not always easy to determine from a Methods or Results section what procedure people have used to fit a model to data. There are potentially many features of a parameter estimation technique, and it is frequently the

case that some features are glossed over for the sake of brevity. At this point we will discuss who has used which techniques and what techniques are used by the most popular model-fitting routines.

By far the most common fitting technique is maximum likelihood (Hockley, 1984; Ratcliff, 1978; Ratcliff & Murdock, 1976). One routine available to perform MLE of ex-Gaussian and Weibull parameters is that of Cousineau and Larochelle (1997). Many people also use maximum likelihood on the vincentiles of a sample, whether averaged across subjects or conditions or not (Balota & Spieler, 1999; Heathcote et al., 1991; Ratcliff, 1979, 1988; Spieler et al., 1996). Heathcote's RTSYS routine (Heathcote, 1996) has been frequently cited and used to perform fits to vincentiles (e.g., Leth-Steenson, King Elbaz & Douglas, 2000; Madden et al., 1999; Plourde & Besner, 1997; Smith & Mewhort, 1998). As we noted above, MLE fits to vincentiles are less accurate than MLE fits using the whole sample, even when the sample is small. RTSYS by default performs MLE fits to vincentiles, but will perform MLE fits to the whole sample if the number of vincentiles specified is equal to the sample size. However, when performing MLE fits to an average (i.e., Vincentized) distribution, the only option available is to use the vincentiles.

Many people also fit models by minimizing SSE between either the theoretical and empirical density functions (Blough, 1988; McElree, 1998; McElree & Dosher, 1993) or CDFs (Logan, 1992; Van Zandt et al., 2000). Some fits have been performed by minimizing the Pearson χ^2 goodness-of-fit statistic, which is a variant of the SSE minimization performed on the CDF (Nosofsky & Palmeri, 1997; Possamai, 1991; Smith & Vickers, 1988; Strayer & Kramer, 1994; Van Zandt et al., 2000). The χ^2 function to be minimized is constructed by placing the observed RTs into bins $[0, t_1), [t_1, t_2), \ldots, [t_{r-1}, t_r), [t_r, \infty)$, typically selected so that the proportion of RTs falling into each bin is equal ($O_i = 1/(r+1)$ for each bin i). The

expected proportions of observations in each bin equals $E_1 = F(t_1;\hat{\theta}), E_2 = F(t_2;\hat{\theta}) - F(t_1;\hat{\theta}), \ldots, E_r = F(t_r;\hat{\theta}) - F(t_{r-1};\hat{\theta}), E_{r+1} = 1 - F(t_r;\hat{\theta})$, for the theorized CDF F. The function minimized in the fits is then

$$\chi^2(\hat{\theta}) = N \sum_{i=1}^{r+1} \frac{(O_i - E_i)^2}{E_i}$$

for a sample of size N. Dawson's (1988) routine fits the ex-Gaussian distribution by minimizing $\chi^2(\hat{\theta})$, where observations have been binned using vincentiles.

Once a strategy for fitting a model has been selected, a method for finding parameters that minimize or maximize the desired function must also be selected. We turn now to the issue of optimization, or finding the minimum or maximum point on a surface.

Optimization Techniques

There are a number of ways to perform a nonlinear sum of squares minimization (or likelihood maximization), the most well-studied being perhaps Gauss-Newton (Gallant, 1987). Unfortunately, many minimization methods (including Gauss-Newton) depend on the quality of the initial guess for $\hat{\theta}$ and how well-behaved the functions $\mathrm{SSE}(\hat{\theta})$ and $\ln L(\hat{\theta})$ are. (A well-behaved function is one that is fairly smooth and does not have too many local minima.) One technique that we have found to be very reliable with poorly-behaved functions and inaccurate starting values is that of an iterative simplex method. A simplex algorithm (Nelder & Mead, 1965) is a routine that attempts to find global minima (see also Dawson, 1988 and Heathcote, 1996). It is so named because it constructs a simplex (an n-dimensional polygon with $n + 1$ vertices) in the n-dimensional space defined by the n parameters composing $\hat{\theta}$. A triangle in two-dimensional space is an example of a simplex. The value of $SSE(\hat{\theta})$ or $-\ln L(\hat{\theta})$ is

computed at each corner of the simplex, and the simplex reorients itself around the corner with the lowest value. This process is repeated, until the simplex is squeezed into a single point, the value of $\hat{\theta}$ at the minimum. (Note that minimizing $-\ln L(\hat{\theta})$ is equivalent to maximizing $\ln L(\hat{\theta})$.)

In practice, a "global" minimum is global only when no smaller minimum can be found. On surfaces as irregular as the ones typically produced by fitting RT data, there is almost always another minimum lurking beyond the region of the parameter space covered by a single iteration of the simplex. Therefore, the simplex algorithm iterates many times, each time the previous best value of $\hat{\theta}$ is used as the starting point and the corners of the new simplex are constructed by perturbing each element of $\hat{\theta}$ by a normal random variable. With each successive iteration, the variance of the perturbations is increased, causing larger and larger simplexes to be constructed. In this way, the iterative process acts much like simulated annealing. After settling into a local minimum, the perturbation process "shakes" the parameters harder and harder in an attempt to cover the entire parameter space. For most of the model fits, 40 iterations were used. Occasionally, fewer iterations were used for the purposes of speed. For instance, the diffusion fits occasionally required as long as an hour for 5-10 iterations. When the iterations seemed very slow, only 20 were performed. The SSE and likelihood functions usually changed very little after 5-10 iterations.

One advantage that we had in these fits is that we knew what the true parameters were (so Gauss-Newton probably would have performed quite well). One of the greatest challenges in model fitting is the selection of good starting values for any minimization process. It is often the case that a nonlinear fitting problem's dependence on good starting values is so great that the true parameters need to be approximately known before attempting to find them. Not knowing the true parameter values, or using very poor

starting values, can result in settling into a local minimum or a complete failure of the minimization routine to converge. At the very least, using poor starting values increases the variance of the parameter estimate. In the fits that we performed, the starting values of the parameters were always the actual values of the parameters that were used to generate the data. If there were no error in the data, the minimization routine would be unable to find any better parameters and would stop. However, there is considerable error in the simulated data, and in the resulting density and CDF estimates, and so the parameters that minimize SSE or maximize $\ln L$ are never exactly equal to the original values. Occasionally, they aren't even close to the original values. This tendency for error to skew the values of the parameters is caused by the fitting algorithm's attempts to accommodate even fluctuations due to error.

In application, we are forced to use inaccurate starting values when we attempt to fit a model to a set of RT data generated by an unknown process. It is not possible to determine if the fluctuations observed in a particular density or CDF estimate are due to error or to the process. Therefore, the results that we present in this paper represent the optimal or best-case situation. There is no error variance due to poor starting values, and any variance in the recovered parameters we know must be due to error in the data (as opposed to a potentially misspecified model). Our results can therefore speak directly to the influence of different density and CDF estimators on the variance of the parameters. It must be kept in mind, however, that misspecifying the starting values will increase the variance of the estimates, and may increase the bias of the estimates as well.

For each fit to each model, using each sample size and density/CDF estimation technique, we recovered parameter values by minimizing SSE (or maximizing χ^2), or maximizing $\ln L$. This was done 500 times, so that we

could estimate the variance of both the density/CDF estimates and of the parameters themselves.

Results

We begin by exploring the accuracy of the different density and CDF estimators; that is, how well each estimator recovered the true shape of the density or CDF that was the basis of the simulation. Then we will present the results of the parameter estimations.

Density and CDF Estimates

For each of the 500 simulated data sets for each model at each sample size, the estimated densities and CDFs were computed. These estimates were averaged across simulations. For the fixed-width histogram and the Gaussian kernel estimators, the mean and variance of the height of each bar were computed. For the variable-width histograms, the means and variances of the height of each bar and each bar midpoint were computed. For the CDF estimator, the mean and variance of the percentiles were computed. These means and variances were used to construct Figures 4-13. Note that bars around mean density and CDF estimates are standard deviations and not standard errors. Bias, the difference between the mean estimate and the true density, is plotted separately with the standard error of the mean estimate around each point. The proportion of estimated points that show no significant bias (as indicated by z-tests performed at each estimated point) are given in Table 4.

Histogram estimates

The averaged estimate is shown in Figures 4 and 5 for the decile-based and Vincent variable-width histograms. Because both the limits of the histogram bars and the heights of the bars vary across data sets, the standard deviations were calculated for both. The standard deviations are plotted as the major and minor axes of an "error ellipse" around the average (midpoint and height) point for each bar.[6] A number of shortcomings of these estimates are immediately clear. One concern is the fact that the variable-width histograms are severely biased for smaller sample sizes. This is evident for sample sizes of 50 and 100, where the average height of the curve is quite consistently higher than the actual density function. As sample size increases, the extent of bias decreases, but is still evident for samples as large as $N = 1000$. Although it is difficult to see in Figures 4 and 5, there are also patterns of bias that do not disappear as the sample size becomes extremely large. Another concern is the extent of variation around the mean estimate, which does not disappear until $N = 1000$. The potential for obtaining a highly inaccurate estimate is quite great for all but the largest sample sizes.

Figures 6 and 7 show the bias for each estimate at each sample size for each model for the decile and Vincent estimates, respectively. The last bar of the decile histograms (centered on the 95^{th} percentile) is, on average, considerably higher than the actual height of the density, as indicated by the significant positive bias in Figure 6, and is widely variable. This effect, as well as the overall overestimation of the density, is probably due to the dependence of the 95^{th} percentile on the maximum statistic (i.e., the 100^{th} percentile or slowest RT). Because it is difficult to estimate this accurately (each curve extends to positive infinity), the last bar of the histogram is also inaccurate. This effect can be appreciated by noticing that not only does the bias for the last point decrease as N increases, but it also shifts

systematically to the right.

The Vincent histograms did not show this problem as severely, because the uppermost 5% of the data were not included in the plot. However, whereas the decile-based variable-width histograms had difficulty with the last bar of the histogram, the Vincent estimates had difficulty with the first bar. For each model, as sample size increased, the Vincent estimate tended to underestimated the height of the density at the first point. This was especially prominent for the diffusion model, in which the Vincent estimate was poor for the first two points of the estimate at even the largest sample size. Figure 7 shows the negative bias at the first point for each model. As for the decile-based histograms, this may again be due to the difficulty of estimating extrema: the minimum, in this case, is difficult to estimate accurately. However, failing to observe a minimum that was small enough should have resulted in an overestimation of the density rather than the observed underestimation. For example, in the decile-based estimates, there is another bias that seems to be specific to the ex-Gaussian model. The first point of the histogram overestimates the true height of the density. Accurate estimation of the minimum would be a greater concern for the ex-Gaussian, which takes on values to negative infinity, than for any of the other models.

Figures 6 and 7 show that the variable-width histograms are consistent estimators of the density function: as sample size increases, the bias decreases. However, they are inefficient: both the decile and Vincent estimators are highly biased, and significant bias persists up through samples of size $N = 10,000$. As shown in Table 4, even when N equals 10,000, most of the estimated points on the density function shown in Figures 4 and 5 deviated significantly from the true density function. Overall, the Vincent histograms were less variable than the decile-based histograms and showed fewer significant deviations (as assessed by z-tests at each point) than the

decile-based histograms at intermediate sample sizes.

One might conclude from Figures 4 and 5 that the observed bias is not so bad because the curve for the true density runs through the standard deviation ellipses around the mean estimate (see, for example, the estimates for $N = 50$ in Figures 4 and 5). This conclusion would be wrong. The ellipses do not represent the standard error of the mean, which was computed for the bias plots, but the standard deviation of the height of the estimate at each point. If the plots were constructed using standard errors, the radii of the ellipses would be reduced by a factor of $\sqrt{500}$: the radius of the largest ellipse pictured for the $N = 50$ estimates would be reduced approximately to the radii of the ellipses for the $N = 10,000$ estimates. The standard errors associated with the height of the estimates are shown around the bias points in Figures 6 and 7. The extent of variation in the estimates and the persistent bias indicate that not only is an estimate likely to be highly inaccurate, but also that parameter estimates based on fits to these estimates are going to be inaccurate, and that the inaccuracies in the parameters will be a recurring problem for even very large sample sizes.

The average estimates for the fixed-width histograms are presented in Figure 8 and their bias is presented in Figure 9. From these figures, it is clear that, like the variable-width histograms, the fixed-width histogram is highly variable and significantly biased at all sample sizes. Although the extent of variation and bias is less for the fixed-width than the variable-width histograms, the fixed-width histogram is not consistent: bias persists through the largest sample size and, unlike the other histograms, the extent of bias seems unaffected by increases in the sample size. The major problem with this estimator is located at the mode of the distributions, where the density is underestimated for all models at all sample sizes. The points in Figure 9 can be directly compared to the points in Figure 8; the large negative

differences at the early points on the bias curves correspond to the mode or the points immediately preceding the mode.

As shown in Table 4, the number of significant deviations between the mean fixed-width histogram estimate and the true density function actually increased with sample size. The reason for this was explained above. Because the smoothing parameter h_N does not decrease with increases in N, the estimator is not guaranteed to be either asymptotically unbiased or consistent. The increase in significant deviations from the curve with increases in N reflects the fact that the bin boundary distance ($h_N = 50\ ms$) is inappropriate for larger samples (although it does fairly well for smaller samples).

Gaussian kernel estimates

The average estimates for the Gaussian kernel estimators are presented in Figure 10, and their bias is presented in Figure 11. This estimate, too, is very biased and highly variable at smaller sample sizes, and tends to underestimate the mode as the fixed-width histogram estimates do. However, the extent of bias attenuates quickly with sample size and seems to have all but disappeared by $N = 1000$ (except for the diffusion model), although significant bias remains at $N = 10,000$. Table 4 shows that the proportion of nonsignificant differences is uniformly high across sample sizes, unlike the histogram estimates, making this estimator the most accurate of the density estimators we have examined.

CDF estimates

The average CDFs are shown in Figure 12, and their bias is presented in Figure 13. There is very little variance in the estimates even at the smallest sample size. The estimates are relatively unbiased at all sample sizes, and

there are no model-specific failures of the estimates. Variance seems to be largest at the higher deciles (the 90^{th}, in particular), but negligible at the smallest deciles even when $N = 50$. Table 4 gives the proportion of nonsignificant differences between the mean quantiles and the true quantiles. Unlike the other estimators, there is little effect on the number of significant differences with increasing sample size, as most of the differences are nonsignificant even at $N = 50$. (The minimum and maximum percentiles were not included in the bias analysis.)

Summary

The results of these analyses demonstrate that the most commonly used density estimates are highly variable and biased and that this bias persists even for very large sample sizes. Of the nonparametric density estimates, the Gaussian kernel estimator suffers the least from bias and model-specific inaccuracies. The CDF estimates are relatively unbiased for all sample sizes and have very small variance. From these results, we may speculate that the parameter values derived from least-squares fits to nonparametric density estimates would be less accurate than those derived from least-squares fits to the CDF or from maximum-likelihood procedures. We turn now to an analysis of the recovered parameter values.

Parameter Recovery

Each model was fit to the data from each of its own simulations (thus there were 500 fits for each model to each density or distribution estimate). For least-squares fits, each of the 500 density and CDF estimates computed in the previous analysis were used. Maximum likelihood and χ^2 minimization was also used for each of the 500 simulations at each sample size for each model. The recovered parameters that minimized the SSE or χ^2 or

maximized the likelihood were then averaged across simulations for each estimation technique. One important characteristic of the parameters is that, within a model, the parameters are usually correlated to some degree. So, for example, the three parameters of the gamma distribution tend to be highly correlated. The mean RT is determined by the base time T_0 and the number of exponential waiting times K. If K is very large, T_0 will have to decrease to compensate. Also, if K is very large, the rate λ may have to increase. Therefore, variances are correlated across the parameters and it is difficult to determine the precise reasons for variance differences in parameters for different models. In an attempt to make comparisons easier, similar parameters in the models (e.g., T_0) are plotted on identically scaled axes.

A few words are in order concerning the fits to the diffusion and race models. Recall that these models predict not only the speed of a single response, but the accuracy of that response and the speed of the alternative (incorrect) response. There are several approaches that might be taken to fit the data from these simulations. One is to ignore the alternative response and to fit only the RT distributions for the "correct" response, i.e., $f(t|\text{correct})$. However, the race and diffusion models are extremely flexible, and there is a range of parameter values that can approximate a density $f(t;\theta)$ if accuracy and the incorrect response times are neglected. For example, three sets of parameters were recovered from fits to three different simulations of the diffusion model with $N = 10,000$. The parameters yielded three different accuracies for the simulation (.14, .46 and .85, versus the true accuracy of .83). The (theoretical) CDFs and densities for these three sets of parameters are shown (as broken lines) in Figure 14, along with the true CDF and density function (the distribution that produced the data; solid lines). The different curves are virtually indistinguishable, even though the parameters are quite different and yield very different accuracies. Parameter

recovery was quite poor for least-squares fits to all of the estimators computed for the diffusion model when only correct RTs were fit.

Another approach could involve fitting not only the correct but also the incorrect RT distributions. In practice, there are rarely as many incorrect as correct RTs, and so the two distributions would have to be unequally weighted in any optimal fitting routine. The instability of the estimates for the incorrect distribution could potentially skew the parameters recovered for the correct distribution. Therefore, although the incorrect RTs do help constrain the parameters, it is not clear that the constraints are beneficial over and above the constraint provided by accuracy alone (as we did here). One useful approach would be to fit the models to RTs collected across several experimental conditions, in which it might be assumed that a subset of the parameters (say, those corresponding to response thresholds or bias) is constant, and the rest (all those corresponding to stimulus quality) vary. This approach seems to promote accurate recovery of the parameters (Van Zandt et al., 2000).

For the fits reported here, we constrained the densities and CDFs by the accuracy of each simulation. This was done by fitting the joint RT density $f(t, \text{correct})$ to the joint empirical estimates instead of fitting the conditional RT density $f(t|\text{correct})$ to the conditional empirical estimates. The area under the joint RT density $f(t, \text{correct})$ must equal the accuracy.[7] Fitting the joint RT densities and distributions therefore implicitly fits the accuracy as well. All the diffusion and race density and CDF estimates were multiplied by the observed proportions of correct responses ($N/(N+n)$, where n is the number of incorrect responses and N is the sample size; there were always N correct responses), to obtain estimates of the joint densities and CDFs. The joint density and CDFs of the diffusion and race model were fit to these modified estimates to recover the parameters. As we will show, the resulting

parameters were comparable (in terms of bias and variability) to the parameters recovered for the single-response models.

Occasionally, for certain models, the parameters that minimized the SSE were wildly wrong. Such fits usually happened at the smaller sample sizes, and were specific to particular density estimates. This was a problem especially for the Wald distribution, where the best-fitting values of μ and λ were prone to become very large, which was offset by very small values of T_0. Such estimates were removed from the analysis under the assumption that they would be considered unacceptable in practice (and perhaps a different set of starting values would be used in another fitting attempt). For the Wald distribution, fits that gave values of μ or λ greater than 100,000 were not included in these analyses. The total number of fits (out of 500) deleted for the Wald distribution were 87 and 35 for $N = 50$ and 100 for the decile-based histograms, 58 and 20 for $N = 50$ and 100 for the Vincent histograms, 14 and 1 for $N = 50$ and 100 for the fixed-width histograms, and 13 and 3 for $N = 50$ and 100 for the Gaussian kernel estimates. For the diffusion model, fits that gave values of a greater than 2,000 were not included in these analyses. The total number of fits deleted for the diffusion model were 42, 18 and 1 for $N = 50$, 100 and 500 for the decile-based histograms, 27 and 12 for $N = 50$ and 100 for the Vincent histograms, and 8 and 1 for $N = 50$ and 100 for the Gaussian kernel estimates. Note that although these fits were deleted from the parameter averages, the density estimates on which they were based were included in the density estimate analyses above.

The diffusion. In Figure 15, the four parameters for the diffusion model are shown for each sample size and estimation technique. Fits to the histogram estimates yielded the poorest parameter estimates, with high bias and very large variance for the variable-width histograms and significant bias in

parameters T_0 and ξ for the fixed-width histogram at $N = 10,000$. Fits to the CDFs yielded the best outcome with very little bias and small variance, and recovered the parameters very accurately even for $N = 50$. The maximum likelihood parameters are excellent at all sample sizes, except for the parameter ξ. Unfortunately, the maximum-likelihood technique recovered biased values of ξ even at the largest sample size, and the variance of the estimate did not significantly decrease with sample size. The Gaussian kernel estimator and χ^2 fits, although they performed better than the histogram estimator fits, did not perform as well as the CDF estimator fits, producing biased and variable parameter estimates at all sample sizes (see, e.g., ξ for $N = 10,000$ using the Gaussian kernel and a and z using χ^2 minimization).

The ex-Gaussian. The parameters for the ex-Gaussian model are shown in Figure 16. Again, the nonparametric density estimate fits performed poorly - especially the Vincent estimate fits, which yielded highly biased σ and τ estimates with very large variance even for samples of size $N = 10,000$. The fact that the Vincent estimate fits perform so poorly for the ex-Gaussian distribution is interesting, considering that for applications in which the ex-Gaussian is fit to data using least-squares, the Vincent estimate is used (Dawson, 1988; McElree, 1998; McElree & Dosher, 1993). The maximum likelihood estimates and the fits to the CDF accurately recovered the parameter values even at the smallest sample sizes, showing very little bias and smaller variance than the density estimates. Although the Gaussian kernel estimate and χ^2 minimization fits outperformed the histogram estimates, they did not do as well as the CDF estimate fits. Overall, the maximum likelihood parameter estimates are the least biased and least variable for the ex-Gaussian model.

The gamma. The three parameters for the gamma distribution are shown in Figure 17. There is considerable variance in these estimates for all six techniques until $N = 500$. The gamma distribution is an example of a distribution, similar to the example provided by the diffusion model above, where a single distribution can be well fit by a large number of quite disparate sets of parameter values. In many of the fits in the smaller sample sizes, the shape parameter K reached the value 20 (the largest permitted magnitude; values of K greater than 20 did not significantly change the shape of the distribution, which was at that point approximately normal). Large values of K produced large values of λ. For fits using the variable-width histogram estimators, K was maximized in approximately 10% of the fits for samples of size $N = 100$. For fits using all other estimators, including the fixed-width histogram, this occurred in less than 1% of the fits. For the decile-based variable-width histogram fits, there is bias evident in the parameter estimates even for the largest sample size. The maximum likelihood estimate, CDF estimate, and χ^2 minimization fits are performing comparably, although the CDF estimates are more biased at the smallest sample size, and slightly more variable at all sample sizes. The best parameter recovery for the gamma distribution is accomplished by the maximum likelihood estimates.

The race. The recovered parameters for the race model are shown in Figure 18. All parameters are plotted on the same scale as for the gamma distribution. It is immediately obvious for this model that, although the CDF and maximum-likelihood estimates seem to perform better than any of the others overall, all estimates had considerable difficulty recovering the parameters λ_I and K_I, the parameters determining the behavior of the losing counter in the race. Furthermore, there is no clear improvement with these

estimates with increases in sample size. Even though the fits were constrained by the probability of counter C winning the race, this constraint was not enough to force accurate recovery of the parameters for counter I. The behavior of counter I had only a modest effect on the overall RT distribution. Even at the largest sample size, only the maximum-likelihood estimates produced low bias for these parameters, but they are still more variable than the parameters recovered for other models at this sample size. For small sample sizes, the CDF fits were less variable and biased than any of the other estimates, but for larger samples the maximum-likelihood estimates recovered the parameters best.

The Wald. Estimates for the parameters of the Wald distribution are shown in Figure 19. The histogram estimates again performed poorly, with large biases at small sample sizes that attenuate but persist through the largest sample size. The maximum-likelihood, χ^2 and CDF estimates were unbiased, and the maximum likelihood estimates were of minimum variance, and quite accurate with sample sizes as small as $N = 100$. It should be noted that the parameter λ is highly variable, and often reached very large values which were offset by very small values of T_0. Therefore, the relatively small error bars around the maximum likelihood and CDF estimates for the smaller sample sizes represent considerable variance, in comparison to the variance in parameters of other models. The Vincent histogram estimate thus experienced considerable difficulty accurately estimating λ even when $N = 1000$. It is also biased to overestimate T_0 at $N = 10,000$. Overall, the maximum likelihood and CDF estimates recovered the parameters best and with comparable accuracy, with the maximum likelihood estimate having least variance.

The Weibull. Shown in Figure 20 are the estimates for the Weibull distribution. The CDF and χ^2 estimates were unbiased at all sample sizes, with small variance. The maximum-likelihood estimates were slightly biased at the smaller sample sizes but this bias effect disappeared for samples of size $N = 500$ and larger. The variable-width histogram estimates performed quite poorly at the smaller samples, but by $N = 500$ all the estimators accurately recovered the parameters.

Goodness of fit. An important aspect of parameter estimation that we have not yet discussed is the determination of whether or not a model fits the data well. In practice, after parameters for a model are estimated, a goodness-of-fit statistic like χ^2 is computed for the fit model. Although there are many goodness-of-fit statistics from which one could choose (e.g., Chechile, 1998; Read & Cressie, 1988), and many ways that model fits can be compared across different models (Gallant, 1987; Myung, 1999; Golden, 1999), we will concentrate on χ^2, the most commonly reported goodness-of-fit statistic. First, it is desirable that the model-fitting technique that one chooses produces low χ^2 values when the model is correctly specified. Unfortunately, although χ^2 is by far the most popular goodness-of-fit statistic, it is very powerful at larger sample sizes. As N grows, χ^2 often indicates significant deviations from a model when in fact none exist, as we will soon see.

For the estimation techniques discussed above, we calculated the χ^2 goodness-of-fit statistic for each fit to each model. Using the deciles (plus the minimum and maximum) for each data set, ten bins were defined, each containing 10% of the observations in that data set. The expected frequency of observations within those bins was computed by integrating each theoretical density between the bin boundaries. Successive bins were

collapsed together to obtain at least five observations for both observed and theoretical frequencies (so the χ^2 statistics shown here are not the same as the values of $\chi^2(\hat{\theta})$ minimized in fitting). For small samples, this occasionally resulted in negative degrees of freedom. In this situation, the data set was not included in the χ^2 calculation. We computed the proportion of χ^2 tests that indicated significant deviations from the model using an α-level of .05. Because the models that were fit were in fact the models that generated the data, we might have expected to see approximately $\alpha = .05$ of the fits with significant deviations. Unfortunately this was not the case. The proportion of significant χ^2 statistics are shown in Table 5 for each model and technique. The footnotes indicate the proportion of simulations included (out of 500) after eliminating those with unacceptable parameter values and/or negative degrees of freedom.

The table shows that the extent to which an estimation technique results in good fits depends on the model being fit. The diffusion and race models produced large numbers of significant deviations even though the model being fit was the true model. Fits to the variable-width histograms also show large proportions of poorly fit models. The other techniques, using the fixed-width or kernel estimators, or maximum likelihood or fits to the CDFs, have approximately the expected number of poor fits for an α-level of .05. However, as N becomes very large, all models and all techniques produced higher than expected failure rates for the models. This is because of the sensitivity of χ^2 to sample size; small variations due to noise at large sample sizes are exaggerated by a factor N in the computation of the χ^2 statistic. It should be noted that, with the exception of the diffusion and the race model, the maximum likelihood estimates and least-squares fits to the CDF produced the lowest model failure rate at larger sample sizes.

General Discussion

We have examined two aspects of a number of techniques for estimating densities and CDFs. The first was how well each technique recovered the true shape of the density or CDF. The second was how well each technique, when embedded in a nonlinear least-squares fitting routine, recovered the original parameters of the models.

All of the histogram estimators were highly biased, especially at small sample sizes. While the bias generally declined with sample size, some of these estimators were still significantly biased even when sample size was very large ($N = 10,000$). The extent of bias for the various estimators was dependent on the shape of the density. For example, fixed- and variable-width histogram estimates of the diffusion model and ex-Gaussian distributions were noticeably biased even when sample sizes were very large. The Gaussian kernel estimators were less biased than the histogram estimates, but of comparable variance.

In contrast to the density estimates, CDF estimates showed little bias and had small variance even at the smallest sample sizes, and there were no obvious model-dependent effects on bias or variance for the CDF estimates. It should be noted that the decile-based histogram is the density estimate that would be derived from differencing (taking the derivative of) the CDF estimate, and the decile estimate was one of the most biased and variable of the density estimates. Therefore, the consistency of the CDF estimate is not transferred to a consistent density estimate by differencing (a notoriously inefficient way of estimating the derivative of a function).

Parameter estimates derived from minimizing the SSE between the density and CDF estimates and the model density and CDFs were evaluated according to their bias and variance. All the density estimates produced biased parameter estimates for small sample sizes, and for some models, this

bias persisted even until the sample size was very large. The histogram estimates were particularly poor for recovering accurate parameters. At smaller sample sizes, the histogram estimates occasionally permitted $SSE(\hat{\theta})$ to be minimized by highly inappropriate parameter values. The maximum likelihood estimation technique was, overall, the best at recovering unbiased, minimal variance parameter estimates, followed closely by least-squares fits to the CDF estimate. The maximum likelihood and CDF fits also resulted in the lowest model failure rate as measured by the χ^2 goodness-of-fit statistic.

For the parameter estimates, there were strong model-dependent effects on bias and variability. For example, the race model parameters were not well recovered by any of the estimation techniques. This result contrasts with results presented by Van Zandt et al. (2000) showing that race model parameters could be accurately estimated if a number of experimental conditions were fit simultaneously. The contrast between their results and the present findings suggests a general principle for all such curve-fitting exercises: all data should be used in the fit to the greatest extent possible (including error RTs, if sufficient are present) to avoid erroneous estimates.

Parametric Density Estimates

These results demonstrate that the two techniques that do not allow direct inspection of the shape of the density function, the maximum likelihood and CDF fits, are best at recovering the parameter values for the six RT models examined. The shape of the density function is more intuitively diagnostic of the type of random variable that gave rise to it than is that of the CDF, which looks pretty much the same for all random variables. This result suggests that one way to examine the density function might be to estimate it parametrically, by generating the theoretical density using the parameters estimated by maximum likelihood or fits to the CDFs.

To explore this approach, we computed the density functions for each simulation at each sample size from the parameters found by maximum likelihood and least-squares fits to CDFs. The average density functions are shown in Figures 21 and 22. The two techniques have comparable variance, although variance is slightly smaller for the maximum-likelihood curves (Figure 21) than for the CDF fits (Figure 22). The maximum likelihood estimates have much smaller bias than any other density estimate for all models except the diffusion and the Weibull, even at the smallest sample size. By $N = 100$ observations, the variance of the estimates are small in comparison to the nonparametric estimates, except for the portions of the densities that are changing very rapidly.

Table 4 shows that there are fewer nonsignificant differences between the estimate and the true density function for the MLE and CDF techniques, and that the rate at which bias decreases depends strongly on the model being fit. When $N = 500$, the proportion of nonsignificant differences for the ex-Gaussian, gamma, and Wald models is equivalent to the proportion observed for the Gaussian kernel estimator, but the kernel estimator was more accurate for the rest. Overall, the parameters from CDF fits gave more accurate density estimates than did the MLE parameters (except for the race model) but these estimates are not generally as accurate as the kernel estimates.

These findings suggest that, for graphical purposes when an explicit model is being fit, density estimates could be constructed parametrically using the parameters recovered from maximum likelihood or least-squares CDF fits. These density estimates will be more accurate than the histogram estimates. However, especially for sample sizes less than $N = 500$, the Gaussian kernel estimator will still be less biased for some models (compare Figure 11 with Figures 23 and 24). The density estimates derived from the

CDF fits for the race model were quite poor, for example. Also, these parametric estimates of the density might smooth over systematic deviations in the data that nonparametric estimates would reveal, such as the presence of bimodality.

The Vincent Estimator

The Vincent estimator is probably the most popular nonparametric density estimator in use today (e.g., Balota & Spieler, 1999; Blanco & Alvarez, 1994; Hockley, 1984; McElree, 1998; McElree & Dosher, 1993; Ratcliff, 1978, 1979; Ratcliff & Murdock, 1976; Reber, Alvarez & Squire, 1997). Because of its popularity, it was particularly important to us to determine how well this technique recovered the true shape of the density. As it turned out, the Vincent estimator was the poorest of the estimators we investigated. Note that the Vincent estimator is most useful as a vehicle for obtaining an average distribution shape across different subjects or conditions in an experiment. We have not investigated the effects of averaging in this paper. However, it should be clear that any weakness in the method that arises before averaging will be carried over to the estimate obtained after averaging.

One interesting finding was the interaction between the Vincent estimator and the ex-Gaussian distribution. The variable-width histogram estimates were generally very poor for recovering the ex-Gaussian parameters if the estimates were used for least-squares fits (see Figure 16). Occasionally, the ex-Gaussian is estimated by fitting the Vincent histogram estimate - the resulting estimate is then twice removed from the raw data (Dawson, 1988; McElree, 1998; McElree & Dosher, 1993). The Vincent estimate is an inaccurate estimator of the density, and ex-Gaussian parameters are recovered poorly from it. We therefore investigated the implications of using the ex-Gaussian as a density estimate based on fits to the Vincent estimate

to determine how badly error is compounded by using both procedures.

We first investigated whether the parametric ex-Gaussian density estimate, computed from parameters recovered from least-squares fits to the Vincent estimate, accurately recovered the true ex-Gaussian density. We generated the density estimates from the ex-Gaussian parameters derived from the Vincent fits (shown in Figure 16) and averaged those estimates for the plots shown in Figure 25 (top panel). The critical comparison is between these plots, where the ex-Gaussian densities themselves are averaged, and the Vincent histogram estimates (Figure 5, second row). Compared to the Vincent estimates, there is a serious bias for the ex-Gaussian densities across the range of the distribution, one aspect of which is the underestimation (negative bias) of most of the density tails at all sample sizes (see Figure 26, top panel versus Figure 7, second row). Furthermore, the variance of the ex-Gaussian estimate declines only slightly with sample size. Compared to the estimates of the densities derived from maximum likelihood or kernel estimation, these are quite poor estimates indeed.

Next, we investigated the extent to which accurate densities could be recovered when the true density was not ex-Gaussian. Using MLE and least-squares fits to the Vincent histograms, we computed the best-fitting ex-Gaussian parameters for the simulations of the gamma model. These ex-Gaussian parameters were used to generate parametric estimates of the gamma densities: a "SSE" estimate for the ex-Gaussian fit by least squares and a "MLE" estimate for the ex-Gaussian fit by MLE. We then fit the gamma density back to both ex-Gaussian density estimates using least-squares. This analysis simulates the use of the ex-Gaussian as a parametric density estimate when the ex-Gaussian is not necessarily the model of interest (Heathcote et al., 1991; Ratcliff, 1978; Ratcliff & Murdock, 1976; but our fitting procedures are not the same as those used in these

references). The average ex-Gaussian density estimates are shown in Figure 26 (bottom two panels); there is a consistent bias to overestimate the true mode of the gamma distribution and to underestimate the tails at all sample sizes. Variance around the SSE estimate does not decrease with sample size, nor does bias. Table 4 shows the proportion of nonsignificant deviations between the average SSE estimate and the true density at each sample size. The accuracy of the estimate does not improve with increases in N, and overall accuracy is nowhere near as good as the accuracy of the kernel estimates. The MLE estimates are more biased than the SSE estimates, but the variance around the MLE estimates is smaller and decreases as N increases.

The gamma parameters recovered from the least-squares fits to the SSE and MLE ex-Gaussian estimates are shown in Figure 27. The parameters derived from fits to the SSE estimates are biased at all sample sizes and the magnitude of bias and variance does not decrease with sample size. Although the MLE estimates were more biased than the SSE estimates, the parameters recovered from fits to the MLE estimates were less variable and less biased than those recovered from fits to the SSE ex-Gaussian estimates. This is probably due to the smaller variance of the MLE estimates. However, the parameters recovered from the MLE estimates are still not as accurate as those recovered using MLE directly or least-squares fits to the CDF without using the ex-Gaussian as a parametric estimate of the density.

The ex-Gaussian has been advocated as a general description of the unimodal, positively skewed RT distribution, and it has been extremely useful in that regard. Ratcliff and others (Heathcote et al., 1991; Hockley, 1984; Ratcliff, 1978; Ratcliff, 1979; Ratcliff & Murdock, 1976) used the ex-Gaussian for data-smoothing: it is a very flexible density function and is well able to fit a wide range of unimodal, positively-skewed curves. This

ability, however, does not constitute evidence that RTs are distributed as true ex-Gaussians, although converging evidence might be found to support such a claim (Rohrer & Wixted, 1994). However, some researchers now use the "ex-Gaussian nature of latency distributions" not only as a data summary device, but also to tease apart what are believed to be two independent cognitive processes (Hockley & Murdock, 1987; Madden et al., 1999; Plourde & Besner, 1997; Possamai, 1991), one reflected in the μ and σ^2 parameters and the other in the τ parameter. This argument depends on several assumptions, the most important being the appropriateness of the ex-Gaussian distribution to characterize RTs. Even if the ex-Gaussian were the correct model, our results indicate that the ex-Gaussian parameters can be difficult to recover unless the sample size is fairly large. Therefore the behavior of the μ and τ parameters across experimental conditions should be interpreted cautiously.

Conclusions and Recommendations

We conclude this paper with a number of simple suggestions for data analysis. With respect to visualizing the density function, if an explicit model is not to be fit to the data, the density should be estimated using a Gaussian kernel estimator. The parameters of the estimate can be calculated automatically, and the error in the estimate decreases faster (with N) than the error in any of the other density estimators that we examined. The variable-width histogram estimators (i.e., the decile and Vincent histograms) should be avoided because they are biased even for large sample sizes, and their large variances preclude their use for small sample sizes.

If a formal model is to be fit to the data, the parameters of the model should be estimated by the maximum likelihood technique or by

least-squares fits to the CDF. If a density estimate is then required to visualize the data, the parameters recovered from either of these fits can be used to generate a parametric estimate of the density function only with caution; the accuracy of the density estimates so computed will depend on the model. Although the accuracy of these parametric estimates can be comparable to that of the Gaussian kernel estimate, the Gaussian kernel estimator may be a better choice for some models. After parameter recovery, if goodness-of-fit is measured with χ^2, the results of the χ^2 test, if significant, should not be taken too seriously. For some models, even small sample sizes will produce large (significant) χ^2 statistics.

Finally, the number of observations required to obtain accurate estimates of the density and model parameters depends not only on the process that generated the data but also the model being fit to the data and the technique used to fit it. If direct fits to the CDF or maximum likelihood estimation is used, reasonable accuracy can be obtained with samples of $N = 100$ or smaller when the model is correctly specified. However, the underlying process is not observable (i.e., the model is not likely to be correctly specified), and so it will not be possible to determine the degree to which an estimator might fail. *Caveat emptor.*

References

Ashby, F. G., Tein, J.-Y., & Balakrishnan, J. D. (1993). Response time distributions in memory scanning. *Journal of Mathematical Psychology*, *37*, 526–555.

Balota, D. A. & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*, 32–55.

Bickel, P. J. & Doksum, K. A. (1977). *Mathematical statistics*. San Francisco: Holden-Day.

Blanco, M. J. & Alvarez, A. A. (1994). Psychometric intelligence and visual focused attention: Relationships in nonsearch tasks. *Intelligence*, *18*, 77–106.

Blough, D. (1988). Quantitative relations between visual-search speed and target-distractor similarity. *Perception and Psychophysics*, *43*, 57–71.

Bloxom, B. (1984). Estimating response time hazard functions: An exposition and extension. *Journal of Mathematical Psychology*, *28*, 401–420.

Bloxom, B. (1985). A constrained spline estimator of a hazard function. *Psychometrika*, *50*, 301–321.

Burbeck, S. L. & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception and Psychophysics*, *32*, 117–132.

Chechile, R. A. (1998). Reexamining the goodness-of-fit problem for interval-scale scores. *Behavioral Research Methods, Instruments, & Computers*, *30*, 227–231.

Colonius, H. (1995). The instance theory of automaticity: Why the Weibull? *Psychological Review*, *102*, 744–750.

Cousineau, D. & Larochelle, S. (1997). PASTIS: A program for curve and distribution analyses. *Behavioral Research Methods, Instruments, & Computers*, *29*, 542–548.

Dawson, M. R. (1988). Fitting the ex-gaussian equation to reaction time distributions. *Behavioral Research Methods, Instruments, & Computers*, *20*, 54–57.

Devroye, L. (1987). *A course in density estimation*. Boston: Birkhäuser.

Emerson, P. L. (1970). Simple reaction time with Markovian evolution of Gaussian discriminal processes. *Psychometrika*, *35*, 99–109.

Feller, W. (1968). *An introduction to probability theory and its applications*, volume 1. John Wiley: New York.

Gallant, A. R. (1987). *Nonlinear statistical models*. New York: Wiley.

Golden, R. M. (1999). Statistical tests for comparing possibly misspecified and non-nested models. *Journal of Mathematical Psychology*, *43*, xxx–xxx.

Green, D. M. & Luce, R. D. (1971). Detection of auditory signals presented at random times: III. *Perception and Psychophysics*, *9*, 257–268.

Green, D. M. & Smith, A. F. (1982). Detection of auditory signals occurring at random times: Intensity and duration. *Perception and Psychophysics*, *31*, 117–127.

Heathcote, A. (1996). Rtsys: A DOS application for the analysis of reaction time data. *Behavioral Research Methods, Instruments, & Computers*, *28*, 427–445.

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.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *6*, 598–615.

Hockley, W. E. & Murdock, B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, *94*, 341–358.

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

Juhel, J. (1993). Should we take the shape of reaction time distributions into account when studying the relationship between rt and psychometric intelligence? *Personality and Individual Differences*, *15*, 357–360.

Knuth, D. E. (1981). *Seminumerical algorithms*. Reading, MA: Addison-Wesley.

Leth-Steenson, C., King Elbaz, Z., & Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of adhd children: A response time distributional approach. *Acta Psychologica*, *74*, xxx–xxx.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 883–914.

Logan, G. D. (1995). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review*, *102*, 751–756.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Madden, D., Gottlob, L., Denny, L., Turkington, T., Provenzale, J., Hawk, T., & Coleman, R. (1999). Aging and recognition memory: Changes in regional cerebral blood flow associated with components of reaction time distributions. *Journal of Cognitive Neuroscience*, *11*, 511–520.

McClelland, J. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language*, *37*, 225–252.

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

McGill, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce & R. R. Bush (Eds.), *Handbook of mathematical psychology*, volume 1 (pp. 309–360). New York: Wiley Press.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.

Myung, I. J. (1999). Importance of complexity in model selection. *Journal of Mathematical Psychology*, *43*, xxx–xxx.

Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.

Nosofsky, R. M. & Palmeri, T. (1997). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception and Psychophysics*, *59*, 1027–1048.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, *33*, 1065–1076.

Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, *80*, 53–68.

Plourde, C. & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology*, *51*, 181–194.

Possamai, C. (1991). A responding hand effect in a simple-RT precueing experiment: Evidence for a late locus of facilitation. *Acta Psychologica*, *77*, 47–63.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing* (2 ed.). New York: Cambridge University Press.

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

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, *95*, 238–255.

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

Read, T. R. C. & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. Springer-Verlag: New York.

Reber, P. J., Alvarez, P., & Squire, L. R. (1997). Reaction time distributions across normal forgetting: Searching for markers of memory. *Learning and Memory*, *4*, 284–290.

Rohrer, D. & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory and Cognition*, *22*, 511–524.

Rudd, M. E. (1996). A neural timing model of visual threshold. *Journal of Mathematical Psychology*, *40*, 1–29.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Smith, D. & Mewhort, D. (1998). The distribution of latencies constrains theories of decision time: A test of the random-walk model using numeric comparison. *Australian Journal of Psychology*, *50*, 149–156.

Smith, P. L. & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, *32*, 135–168.

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 461–479.

Strayer, D. L. & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 318–341.

Tapia, R. A. & Thompson, J. R. (1978). *Nonparametric probability density estimation*. Baltimore, MD: Johns Hopkins University Press.

Tolhurst, D. J. (1975). Reaction times in the detection of gratings by human observers: A probabilistic mechanism. *Vision Research*, *15*, 1143–1149.

Townsend, J. T. (1990). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, *108*, 551–567.

Townsend, J. T. & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin and Review*, *7*, xxx–xxx.

Van Zandt, T. & Ratcliff, R. (1995). Statistical mimicking of reaction time distributions: Mixtures and parameter variability. *Psychonomic Bulletin and Review*, *2*, 20–54.

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: W.H. Freeman.

Wixted, J. T. & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1024–1039.

Appendix

The models simulated in this paper are described by the following density and CDFs. Interested readers might also consult Luce (1986) for additional details of the models. All models but the ex-Gaussian and the Weibull have a base time parameter T_0, which shifts the distribution along the positive axis by T_0. This parameter is represented by specifying the density and CDFs as $f(t+T_0)$ and $F(t+T_0)$, respectively, to make the expressions more compact. To obtain the expression for, say, $f(t)$, simply replace the variable t on the right hand side of each equation with $t-T_0$.

The diffusion model

With four parameters, T_0, a, z, and ξ (the drift variance parameter σ^2 is unobservable and was set equal to 0.1), the density of the time to absorb at 0 starting at z is given by

$$f(t+T_0) = \frac{1}{P}\frac{e^{-\frac{(t\xi+2z)\xi}{2\sigma^2}}}{\sqrt{2\pi\sigma^2 t^3}}\sum_{k=-\infty}^{\infty}(z+2ka)e^{-\frac{(z+2ka)^2}{2t\sigma^2}}.$$

This is an alternate form of the density given by Feller (1968) that converges for values of t close to zero. The variable P is the probability of reaching 0 before reaching a:

$$P = \frac{e^{-2\frac{\xi a}{\sigma^2}} - e^{-2\frac{\xi z}{\sigma^2}}}{1-e^{-2\frac{\xi a}{\sigma^2}}}.$$

The density $f(t+T_0)$ can be integrated to give the CDF:

$$\begin{aligned} F(t+T_0) \;=\; & \frac{-e^{-z\xi/\sigma^2}}{2P}\sum_{k=-\infty}^{\infty}\left(\mathrm{sgn}(k)\; e^{|\xi|(z+2ka)/\sigma^2}\{\mathrm{erf}\left(\frac{z+2ka+|\xi|t}{\sqrt{2t\sigma^2}}\right)-1\}+\right. \\ & \left.\mathrm{sgn}(k)\; e^{-|\xi|(z+2ka)/\sigma^2}\{\mathrm{erf}\left(\frac{z+2ka-|\xi|t}{\sqrt{2t\sigma^2}}\right)-1\}\right), \end{aligned}$$

where $\mathrm{erf}(x)$ is the error function

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du,$$

and the sign function

$$\text{sgn}(k) = \begin{cases} -1 & \text{if } k < 0 \\ 1 & \text{if } k \geq 0 \end{cases}.$$

The ex-Gaussian distribution

The ex-Gaussian distribution is the convolution of a normal (Gaussian) random variable, with mean μ and variance σ^2, with an exponential random variable with mean τ. The base time parameter T_0 is unobservable, as it simply shifts the parameter μ. The density of the ex-Gaussian is

$$f(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}+\frac{\mu}{\tau}+\frac{\sigma^2}{2\tau^2}} \Phi\left(\frac{t-\mu-\sigma^2/\tau}{\sigma}\right),$$

where Φ is the standard normal CDF. The ex-Gaussian density can be integrated to yield the CDF

$$F(t) = -e^{-\frac{t}{\tau}+\frac{\mu}{\tau}+\frac{\sigma^2}{2\tau^2}} \Phi\left(\frac{t-\mu-\sigma^2/\tau}{\sigma}\right) + e^{\frac{\mu}{\tau}+\frac{\sigma^2}{2\tau^2}} \Phi\left(\frac{-\mu-\sigma^2/\tau}{\sigma}\right) + \Phi\left(\frac{t-\mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right).$$

The gamma distribution

A gamma random variable is the sum of K exponential variables each with mean $1/\lambda$. The density function is

$$f(t+T_0) = \frac{\lambda(\lambda t)^{K-1}}{(K-1)!} e^{-\lambda t},$$

and the CDF is

$$F(t+T_0) = 1 - \sum_{k=0}^{K-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

The Poisson race model

The density function of the race model is the density of the minimum of two gamma random variables T_I and T_C, one representing the correct response (C) and the other representing the incorrect response (I). There are five

parameters: T_0, λ_C, λ_I, K_C, and K_I. The density is conditioned on the response being correct:

$$f(t+T_0) = \frac{1}{P}\frac{\lambda_C(\lambda_C t)^{K_C-1}}{(K_C-1)!}e^{-\lambda_C t}\sum_{j=0}^{K_I-1}\frac{(\lambda_I t)^j}{j!}e^{-\lambda_I t},$$

which, when integrated, yields

$$F(t+T_0) = 1 - \frac{e^{-t(\lambda_C+\lambda_I)}}{P}\left(\frac{\lambda_C}{\lambda_C+\lambda_I}\right)^{K_C} \times \sum_{i=0}^{K_I-1}\binom{K_C+i-1}{i}\left(\frac{\lambda_I}{\lambda_C+\lambda_I}\right)^i \sum_{j=0}^{K_C+i-1}\frac{[t(\lambda_C+\lambda_I)]^j}{j!}.$$

The factor P is the probability that the correct response was given:

$$P = \sum_{i=0}^{K_I-1}\binom{K_C+i-1}{i}\left(\frac{\lambda_C}{\lambda_C+\lambda_I}\right)^{K_C}\left(\frac{\lambda_I}{\lambda_C+\lambda_I}\right)^i.$$

The Wald distribution

The Wald (or inverse normal) distribution with parameters μ and λ arises from a diffusion process with infinite upper bound a. The density of the time to absorption at 0 is

$$f(t+T_0) = \sqrt{\frac{\lambda}{2\pi t^3}}e^{-\frac{\lambda}{2\mu^2 t}(t-\mu)^2}.$$

The integral of this density can be computed to give

$$F(t+T_0) = e^{\frac{2\lambda}{\mu}}\left[1-\Phi\left(\sqrt{\frac{\lambda}{t}}\,\frac{\mu+t}{\mu}\right)\right] + \left[1-\Phi\left(\sqrt{\frac{\lambda}{t}}\,\frac{\mu-t}{\mu}\right)\right].$$

The relationships between the Wald parameters and the diffusion parameters is that $\lambda = z^2/\sigma^2$ and $\mu = -z/\xi$, where z is the starting point of the process, σ^2 is the drift variance, and ξ is the drift rate.

The Weibull distribution

The Weibull distribution is the limiting distribution of the minimum statistic. Under general conditions, the value of the smallest of N random

variables tends to the Weibull distribution as N gets large. The density of the Weibull distribution is

$$f(t) = \frac{k}{\alpha}\left(\frac{t-\xi}{\alpha}\right)^{k-1} e^{-\left(\frac{t-\xi}{\alpha}\right)^k}$$

for $t > \xi$, and the CDF is

$$F(t) = 1 - e^{-\left(\frac{t-\xi}{\alpha}\right)^k}.$$

The parameter ξ is a base time parameter.

Author Note

Portions of this article were presented at the 31st Annual Meeting of the Society for Mathematical Psychology, August 1998, Vanderbilt University and at the 39th Annual Meeting of the Psychonomic Society, November 1998, Dallas, TX. The author gratefully acknowledges the many contributions to this project by Steven Yantis. Thanks also are due Howard Egeth, Andrew Heathcote, and John Wixted for many helpful comments that greatly improved the paper. The project was funded by NSF #SBR-9702291.

Footnotes

[1]A third way of characterizing a random variable is by way of its *hazard function*. One benefit of hazard function analyses is that, in the absence of parameter variability, the hazard function is more diagnostic of a particular model than either the density or the CDF. Unfortunately, hazard function estimation can be quite tricky (Bloxom, 1984, 1985; Luce, 1986). Also, if parameter variations are present in a data set, hazard functions become less diagnostic (Van Zandt & Ratcliff, 1995). We will not consider hazard function estimates in this paper for these reasons, and also because hazard functions are not typically used for parameter estimation.

[2]Note that the histogram density estimator is different from a frequency histogram, in which the height of each bar is equal to the proportion of observations within the bar interval.

[3]Thanks to Andrew Heathcote for providing this useful nomenclature.

[4]It is not strictly true that vincentiles are evenly spaced and centered over the desired quantiles. To see this, suppose a small sample is obtained: $\{1, 3, 3, 6, 7\}$. We wish to compute 3 evenly-spaced quantiles (the 16.67^{th}, 50^{th}, and 83.33^{th} percentiles). Using linear interpolation, these quantiles are .83, 3, and 6.17, respectively. However, using Vincent's procedure, they are 1.8, 5.4, and 6.6. Again using linear interpolation, these quantiles correspond to the 28^{th}, 72^{th}, and 92^{th} percentile ranks, respectively; they are not the expected quantiles and they are not evenly spaced. We investigated the behavior of vincentiles for several distributional forms, including the gamma and the Weibull. In general, the average vincentiles were in the neighborhood of the desired quantiles, but were biased. The vincentiles tended to be pulled in the direction of a distribution's skew, as demonstrated for the small sample above.

[5]To see that this must be true, using MLE we estimated ex-Gaussian

parameters using both the entire sample and using only 10 vincentiles computed for the same sample. For 500 samples of each sample size, the parameters estimated from the vincentiles were significantly poorer than those estimated from the entire sample, even when the sample was very large. We then computed 10 exact quantiles (with percentile ranks of 5%, 15%, ..., 95%) for the ex-Gaussian with parameters $\mu = 705.13$, $\sigma = 31.62$ and $\tau = 94.87$. Using MLE on the 10 exact quantiles (and using the true parameters as starting points), the recovered parameters were $\mu = 697.07$, $\sigma = 28.51$ and $\tau = 97.43$. Therefore, the 10 quantiles, even though they were exact, were not sufficient to recover the original parameters exactly. Replacing the exact quantiles in the MLE routine with sample statistics (such as vincentiles) only makes the parameter estimates less accurate.

[6]The ellipses are meant to give an indication of the variance in bar placement as well as bar height. They are not meant to represent the two-dimensional variance around the height of the density. Because the height and width of the bars are negatively correlated, a true two-dimensional equiprobable ellipse would be slanted to the left. The correlation between the midpoint of the bar and its height is not perfect, however, and the height and width of each bar on the histogram is correlated with the height and width of every other bar on the histogram. Neither is the variance around the mean likely to be symmetric. Computing the true equiprobable ellipse is therefore not a trivial problem.

[7]The distribution of the finishing time (when the process ends regardless of the outcome) for both the race and the diffusion has a density function of the form

$$f(t) = f(t, \text{correct}) + f(t, \text{incorrect}).$$

In other words, the time that the process ends can be due either to the time to make a correct or an incorrect response. Usually we concentrate only on

the correct RTs, or $f(t|\text{correct})$. By the definition of a conditional probability,

$$f(t|\text{correct}) = \frac{f(t, \text{correct})}{P(\text{correct})},$$

and therefore,

$$f(t, \text{correct}) = f(t|\text{correct})P(\text{correct}).$$

Because the integral of the conditional density $f(t|\text{correct})$ over all t must equal one,

$$P(\text{correct}) = \int_0^\infty f(t, \text{correct})dt.$$

Fitting $f(t, \text{correct})$ must then implicitly incorporate the probability of a correct response.

Table 1. Order statistics ($T_{(j)}$) for a sample of 50 observations drawn from a Weibull distribution with mean $\mu = 800$ and variance $\sigma^2 = 100^2$. Each order statistic occurred n_j times in the sample. The cumulative relative frequency $F_j = \sum_{i=1}^{j} n_j/50$.

j	$T_{(j)}$	n_j	F_j	j	$T_{(j)}$	n_j	F_j	j	$T_{(j)}$	n_j	F_j
1	673	1	.02	17	747	1	.36	33	820	2	.72
2	676	1	.04	18	750	1	.38	34	825	1	.74
3	680	1	.06	19	753	1	.40	35	828	1	.76
4	687	1	.08	20	758	1	.42	36	833	1	.78
5	697	1	.10	21	766	2	.46	37	836	1	.80
6	698	1	.12	22	771	1	.48	38	852	1	.82
7	703	1	.14	23	778	1	.50	39	858	1	.84
8	704	1	.16	24	780	1	.52	40	865	1	.86
9	710	1	.18	25	796	1	.54	41	868	1	.88
10	712	2	.22	26	799	1	.56	42	871	1	.90
11	718	1	.24	27	804	1	.58	43	886	1	.92
12	721	1	.26	28	805	1	.60	44	910	1	.94
13	725	1	.28	29	806	1	.62	45	966	1	.96
14	732	1	.30	30	809	1	.64	46	998	1	.98
15	735	1	.32	31	810	1	.66	47	1016	1	1.00
16	740	1	.34	32	816	1	.68				

Table 2. Percentiles computed from the sample in Table 1 using linear interpolation (Deciles) and Vincent's procedure (Vincentiles).

Percentile	Deciles	Vincentiles
0	673.0	
5		682.6
10	697.5	
15		705.4
20	712.0	
25		721.6
30	733.5	
35		745.0
40	755.5	
45		767.8
50	779.0	
55		796.8
60	805.5	
65		812.2
70	820.0	
75		828.4
80	844.0	
85		862.8
90	878.5	
95		955.2
100	1016.0	

Table 3. The parameter values used in the simulations of each model. For the Weibull, the parameter ξ is a base time parameter. Parameter values were selected so that each model predicted a mean RT of 800 ms and RT variance of 100^2 ms^2.

Model	Parameters				
Diffusion	$T_0 = 672.36$	$a = 7.85$	$z = 3.93$	$\xi = -.02$	
Ex-Gaussian		$\mu = 705.13$	$\sigma^2 = 1000$	$\tau^2 = 9000$	
Gamma	$T_0 = 600$	$\lambda = .02$	$K = 4$		
Race	$T_0 = 655.14$	$\lambda_C = .0126$	$\lambda_I = .0032$	$K_C = 2$	$K_I = 2$
Wald	$T_0 = 600$	$\mu = 200$	$\lambda = 800$		
Weibull	$\xi = 652.72$	$\alpha = 163.15$	$k = 1.5$		

Table 4. Proportion of nonsignificant differences (by z-tests, $\alpha = .05$) between the mean estimates for each estimator and the true density. The number of points at which the density was estimated is given in parentheses next to each estimator. The "Distribution Density" is the density estimator derived from fits to the CDF.

		Model					
Estimator	N	Diffusion	ex-Gaussian	Gamma	Race	Wald	Weibull
	50	.00	.00	.00	.00	.00	.00
	100	.00	.00	.00	.00	.00	.00
Decile	500	.20	.10	.10	.10	.10	.10
(10)	1000	.30	.40	.50	.40	.30	.50
	10000	.30	.20	.50	.40	.30	.50
	50	.00	.00	.11	.00	.00	.00
	100	.11	.11	.11	.11	.00	.22
Vincent	500	.44	.33	.67	.56	.44	.67
(9)	1000	.44	.56	.78	.67	.67	.67
	10000	.67	.33	.00	.22	.67	.00
	50	.75	.75	.75	.89	.88	.94
	100	.88	.56	.75	.81	.81	.94
Fixed	500	.63	.63	.69	.69	.81	.69
(16)	1000	.63	.50	.56	.63	.69	.56
	10000	.19	.19	.25	.31	.25	.25
	50	.73	.73	.70	.70	.80	.63
	100	.83	.60	.67	.77	.60	.73
Kernel	500	.67	.63	.70	.67	.77	.67
(30)	1000	.83	.70	.63	.83	.77	.70
	10000	.83	.73	.70	.77	.80	.83
	50	.56	1.00	.89	.67	1.00	.78
	100	1.00	.78	.67	.78	.78	.89
Distribution	500	.89	.89	1.00	1.00	1.00	1.00
(9)	1000	1.00	1.00	1.00	1.00	.89	.67
	10000	1.00	.89	.89	.89	1.00	1.00

Table 4 continued.

		Model					
Estimator	N	Diffusion	ex-Gaussian	Gamma	Race	Wald	Weibull
	50	.27	.39	.24	.21	.21	.18
	100	.27	.48	.52	.24	.36	.21
MLE	500	.33	.82	.73	.36	.88	.27
(33)	1000	.27	.91	.52	.48	.70	.24
	10000	.64	.91	.82	.30	.97	.88
	50	.21	.33	.30	.15	.27	.18
Distribution	100	.15	.52	.52	.06	.52	.24
Density	500	.21	.82	.76	.09	.73	.30
(33)	1000	.55	.91	.88	.09	.76	.24
	10000	.97	.91	.85	.09	.97	.88

Table 5. Proportion of significant χ^2 statistics for each fitting technique and sample size N. Significance was determined using an α-level of .05. Degrees of freedom varied depending on how cells were collapsed to obtain expected bin frequencies of at least 5. Footnotes indicate the proportion of fits for which χ^2 could be calculated.

		Fitting Technique						
Model	*N*	Decile	Vincent	Hist	Kernel	MLE	Dist	Chi
	50	.65[2]	.74[2]	.46[1]	.31[1]	.20[1]	.32[1]	.53[1]
	100	.60[1]	.66[1]	.50	.19[1]	.06[1]	.09	.32
Diffusion	500	.70[1]	.84	.99	.54	.18[1]	.19	.27
	1000	.65	.84	1.0	.54	.11	.15	.16
	10000	.85	.98	1.0	.90	.39	.40	.36
	50	.40[1]	.35[1]	.12	.11	.11	.05[1]	.11[1]
	100	.38[1]	.33[1]	.08	.08	.08	.04	.05[1]
Ex-Gaussian	500	.40[1]	.79	.15	.11	.13	.08	.07
	1000	.60	.90[1]	.20	.12	.13	.08	.07
	10000	1.0	1.0	.91	.29	.31	.20	.17
	50	.25[1]	.19	.09	.07	.09	.07	.10
	100	.22	.12	.05	.05	.04	.04	.04
Gamma	500	.23	.31	.13	.12	.08	.08	.08
	1000	.21	.32	.15	.11	.10	.10	.10
	10000	.55	.65	.60	.28	.26	.26	.22
	50	.60[3]	.60[3]	.55[3]	.42[3]	.36[3]	.43[3]	.51[3]
	100	.61	.61	.42[1]	.30	.19[1]	.32	.32
Race	500	.65	.79	.60	.58	.25[1]	.55	.52
	1000	.74	.87	.67	.63	.24[1]	.58	.59
	10000	.98	1.0	1.0	.93	.44[1]	.73	.73

Table 5 continued.

		Fitting Technique						
Model	*N*	Decile	Vincent	Hist	Kernel	MLE	Dist	Chi
	50	.35[1]	.29[1]	.12[1]	.09[1]	.09[1]	.05	.09[1]
	100	.31	.30	.09	.07	.05	.05	.07
Wald	500	.34	.68	.19	.12	.09	.08	.09
	1000	.30	.75	.20	.13	.09	.08	.09
	10000	.90	.95	.64	.27	.22	.20	.17
	50	.37[1]	.39	.15	.16	.13[1]	.06	.15[1]
	100	.40	.44	.11	.11	.07	.04	.07
Weibull	500	.49	.75	.58	.28	.12	.12)	.21
	1000	.65	.79	.75	.28	.12	.13	.24
	10000	.99	.87	1.0	.84	.21	.39	.43

[1]Less than 100%, [2]Less than 80%, [3]Less than 40%

Table 6. Proportion of significant differences ($\alpha = .05$) between the mean estimates for the ex-Gaussian parametric estimators and the true density. The densities were estimated at 37 equally-spaced points. The parameters for the ex-Gaussian data were estimated by least-squares fits to the Vincent estimator. The parameters for the gamma data were estimated by least-squares fits of the ex-Gaussian to the Vincent estimator (SSE) and by MLE.

	Sample Size				
Model	50	100	500	1000	1000
ex-Gaussian	.76	.64	.73	.76	.79
Gamma (SSE)	.79	.82	.85	.79	.82
Gamma (MLE)	.92	.95	1.00	1.00	1.00

Figure Captions

1. Decile estimate (points) of the CDF for the data presented in Table 1. The curve drawn through the estimates shows the true CDF for the distribution from which the data were sampled.

2. Fixed-width histogram (panel A), decile histogram (panel B), Vincent histogram (panel C), and Gaussian kernel (panel D) density estimates for the data presented in Table 1. The curve drawn through the estimates shows the true density function of the distribution from which the data were sampled.

3. Log likelihood (panel A) and sum of squared error (panel B) for fits of the Weibull model to the data presented in Table 1. The Weibull parameters are ξ and α, and k is fixed in these figures.

4. Average decile histogram estimates with standard deviation ellipses indicating plus or minus one standard deviation in the height and width of the estimate. Estimates are calculated for each model at each sample size. The true density is drawn on each plot using a dotted line.

5. Average Vincent histogram estimates with standard deviation ellipses indicating plus or minus one standard deviation in the height and width of the estimate. Estimates are calculated for each model at each sample size. The true density is drawn on each plot using a dashed line.

6. Average bias (average estimate minus true density) for decile histograms with error bars indicating plus and minus one standard error around the average. Bias is multiplied by 10^4 and presented for each model at each sample size.

7. Average bias (average estimate minus true density) for Vincent histograms with error bars indicating plus and minus one standard error around the average. Bias is multiplied by 10^4 and presented for each model at each sample size.

8. Average fixed-width histogram estimates with bars showing plus or minus

one standard deviation in the height of the estimate. Estimates are calculated for each model at each sample size. The true density is drawn on each plot using a dashed line.

9. Average bias (average estimate minus true density) for fixed-width histograms with error bars indicating plus and minus one standard error around the average. Bias is multiplied by 10^4 and presented for each model at each sample size.

10. Average Gaussian kernel estimates with bars showing plus or minus one standard deviation in the height of the estimate. Estimates are calculated for each model at each sample size. The true density is drawn on each plot using a dashed line.

11. Average bias (average estimate minus true density) for Gaussian kernel estimates with error bars indicating plus and minus one standard error around the average. Bias is multiplied by 10^4 and presented for each model at each sample size.

12. Average CDF estimates with bars showing plus or minus one standard deviation in the placement of each decile. Estimates are calculated for each model at each sample size. The true CDF is drawn on each plot using a dashed line.

13. Average bias (average estimate minus true decile) for CDF estimates (deciles) with error bars indicating plus and minus one standard error around the average. Bias is presented for each model at each sample size.

14. Three diffusion CDFs and densities for three different sets of parameters producing very different accuracy levels. The true density/CDF is also plotted, but is difficult to see. The three parameter sets (T_0, z, a, ξ), recovered by fitting the true model to a sample of size $N = 10,000$ without regard for accuracy and using 3 different sets of starting values, were (673.31,3.94,8.17,.02), (673.52,3.75,6.95,.00), and (672.15,3.89,7.90,-.02).

15. Average recovered parameters for the diffusion model with each technique at each sample size. Error bars are plus and minus one standard deviation. The true value of each parameter is shown as the solid horizontal line.

16. Average recovered parameters for the ex-Gaussian model with each technique at each sample size. Error bars are plus and minus one standard deviation. The true value of each parameter is shown as the solid horizontal line.

17. Average recovered parameters for the gamma model with each technique at each sample size. Error bars are plus and minus one standard deviation. The true value of each parameter is shown as the solid horizontal line.

18. Average recovered parameters for the race model with each technique at each sample size. Error bars are plus or minus one standard deviation. The true value of each parameter is shown as the solid horizontal line.

19. Average recovered parameters for the Wald model with each technique at each sample size. Error bars are plus and minus one standard deviation. The true value of each parameter is shown as the solid horizontal line.

20. Average recovered parameters for the Weibull model with each technique at each sample size. Error bars are plus and minus one standard deviation. The true value of each parameter is shown as the solid horizontal line.

21. Average density estimates derived from maximum likelihood estimates with bars showing plus or minus one standard deviation in the height of the estimate. Estimates are calculated for each model at each sample size. The true density is drawn on each plot using a dashed line.

22. Average density estimates derived from CDF estimates with bars showing plus or minus one standard deviation in the height of the estimate. Estimates are calculated for each model at each sample size. The true density is drawn on each plot using a dashed line.

23. Average bias (average estimate minus true density) for MLE estimates

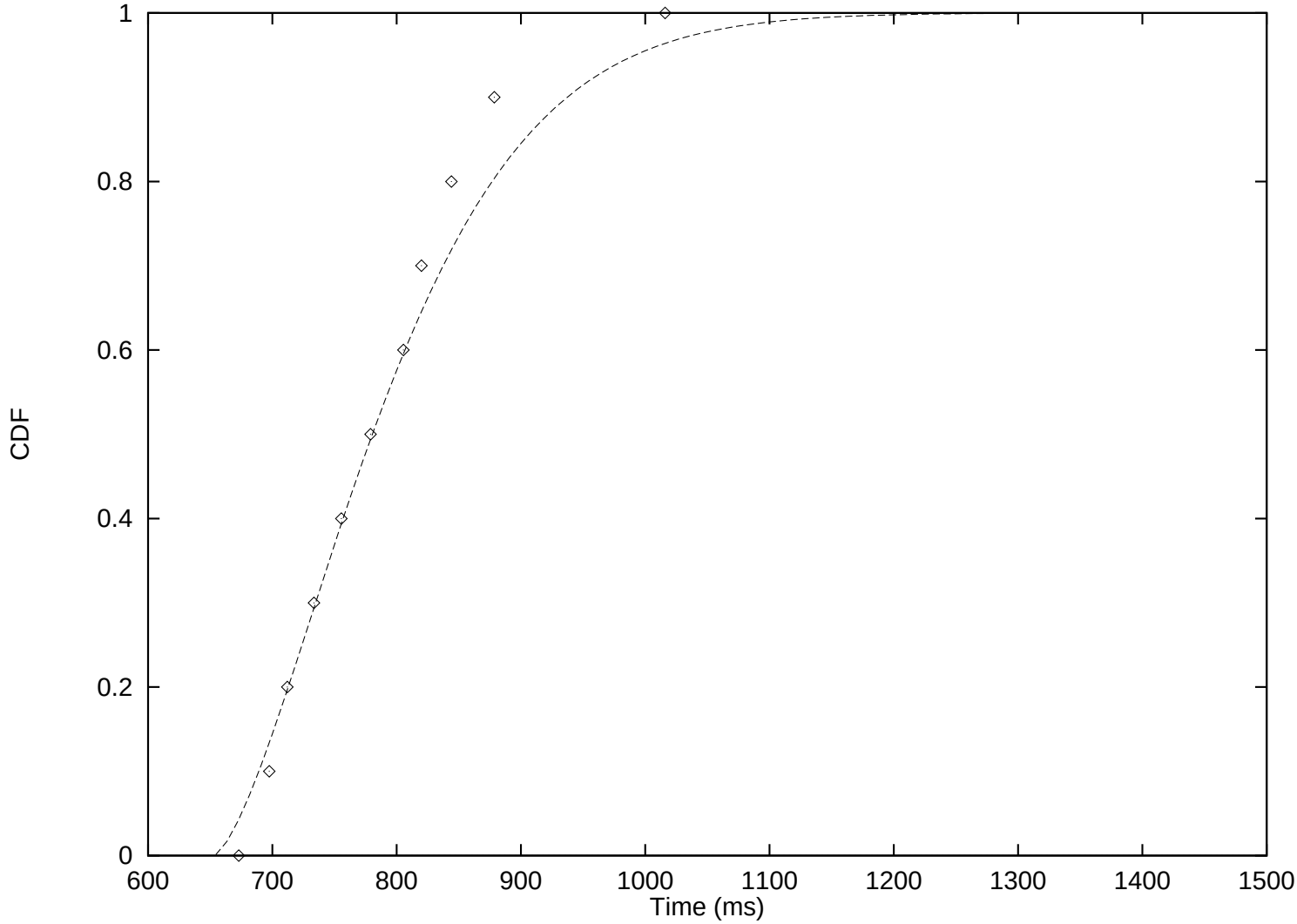
with error bars indicating plus and minus one standard error around the average. Bias is multiplied by 10^4 and presented for each model at each sample size.

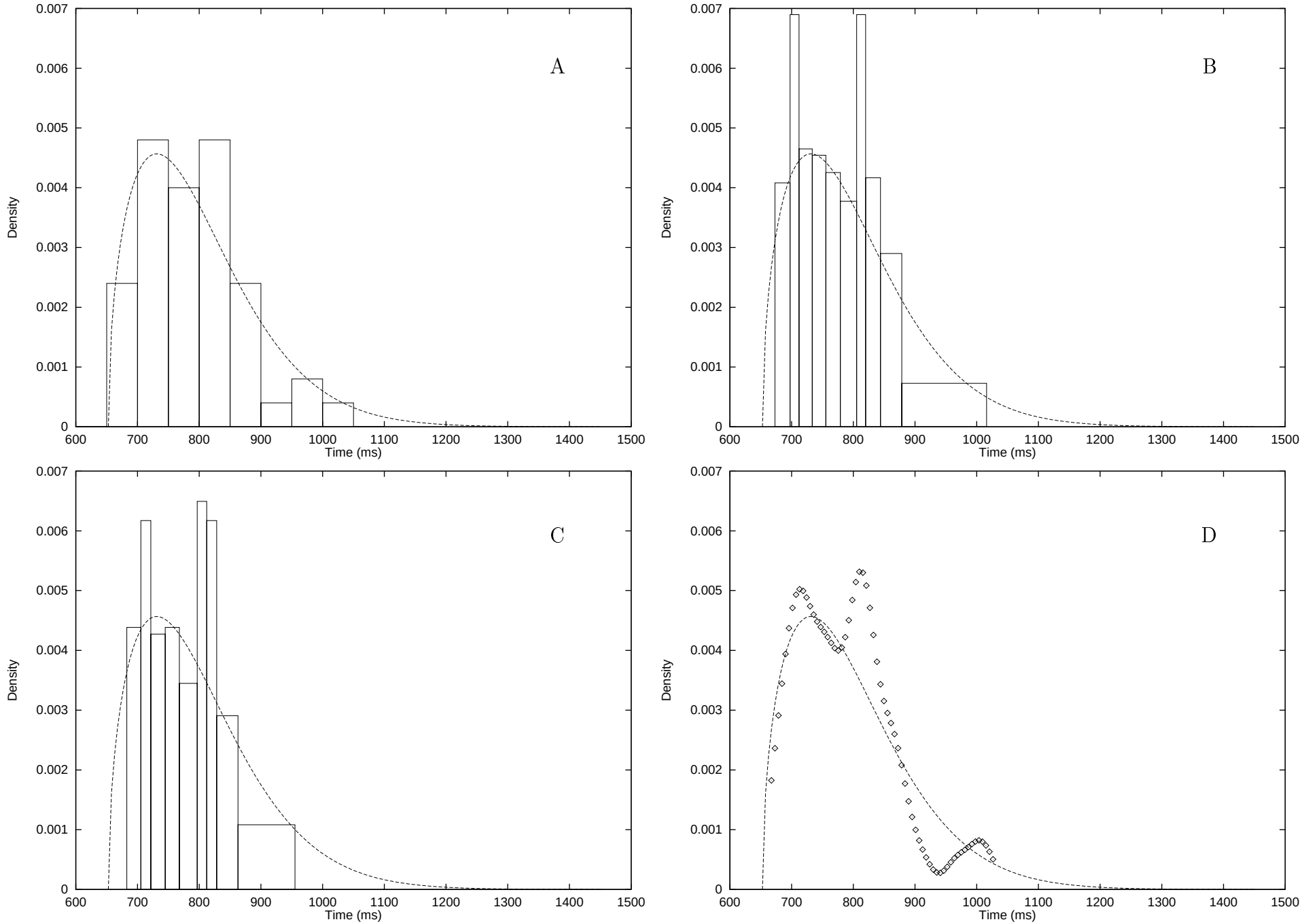
24. Average bias (average estimate minus true density) for distribution estimates with error bars indicating plus and minus one standard error around the average. Bias is multiplied by 10^4 and presented for each model at each sample size.

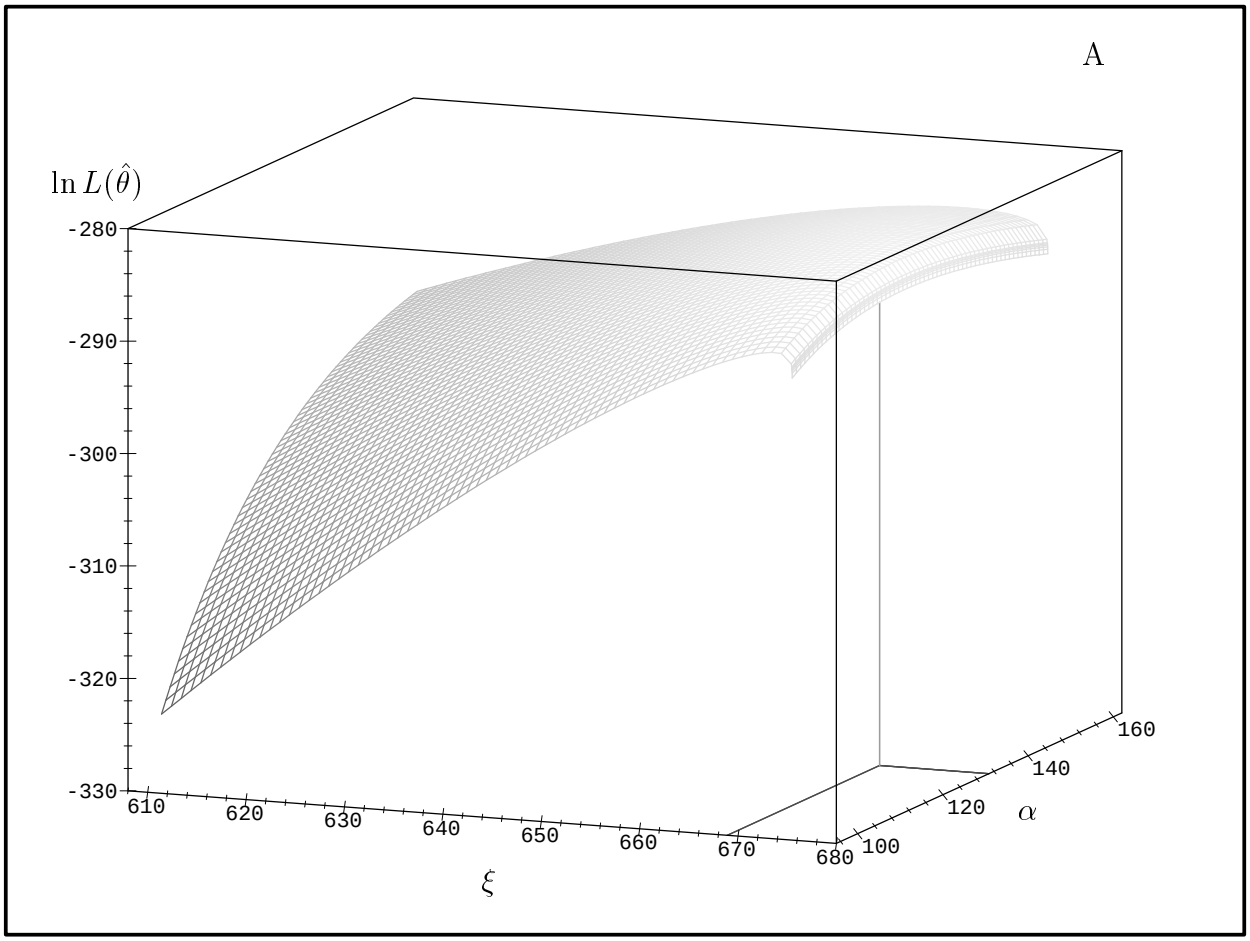
25. Average ex-Gaussian (top panel) and gamma density estimates (bottom two panels) using the ex-Gaussian as a parametric density estimate. For the ex-Gaussian and gamma (SSE) estimates, ex-Gaussian parameters were estimated from least-squares fits to the Vincent histogram estimates. For the gamma (MLE) estimates, MLE was used to fit the ex-Gaussian to the gamma data. The bars show plus or minus one standard deviation in the height of the estimate. Estimates are calculated for each model at each sample size. The true density is drawn on each plot using a dashed line.

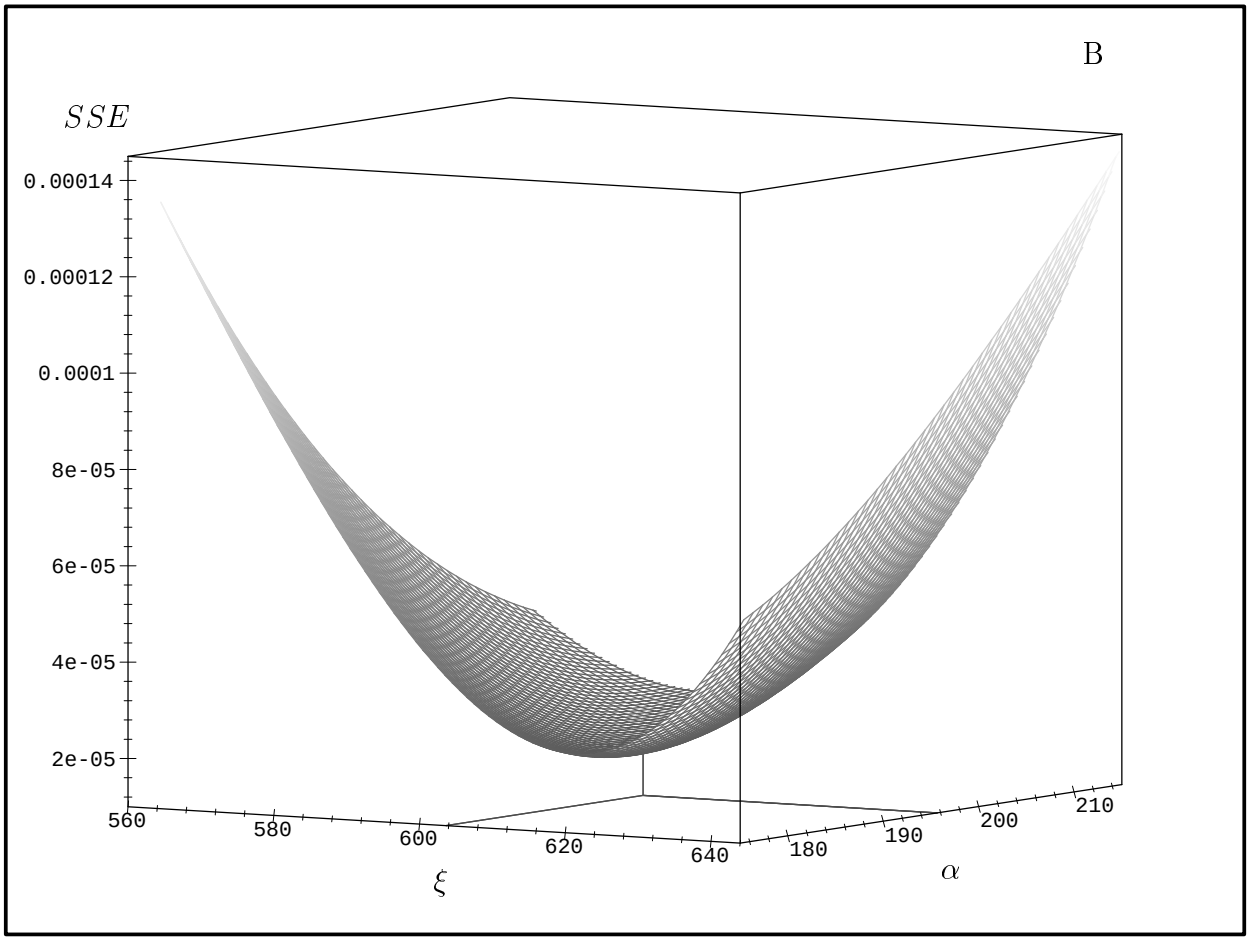
26. Average bias (average estimate minus true density) for parametric ex-Gaussian density estimates for ex-Gaussian data (top panel) and gamma data (bottom two panels). Error bars indicate plus and minus one standard error around the average. Bias is multiplied by 10^4.

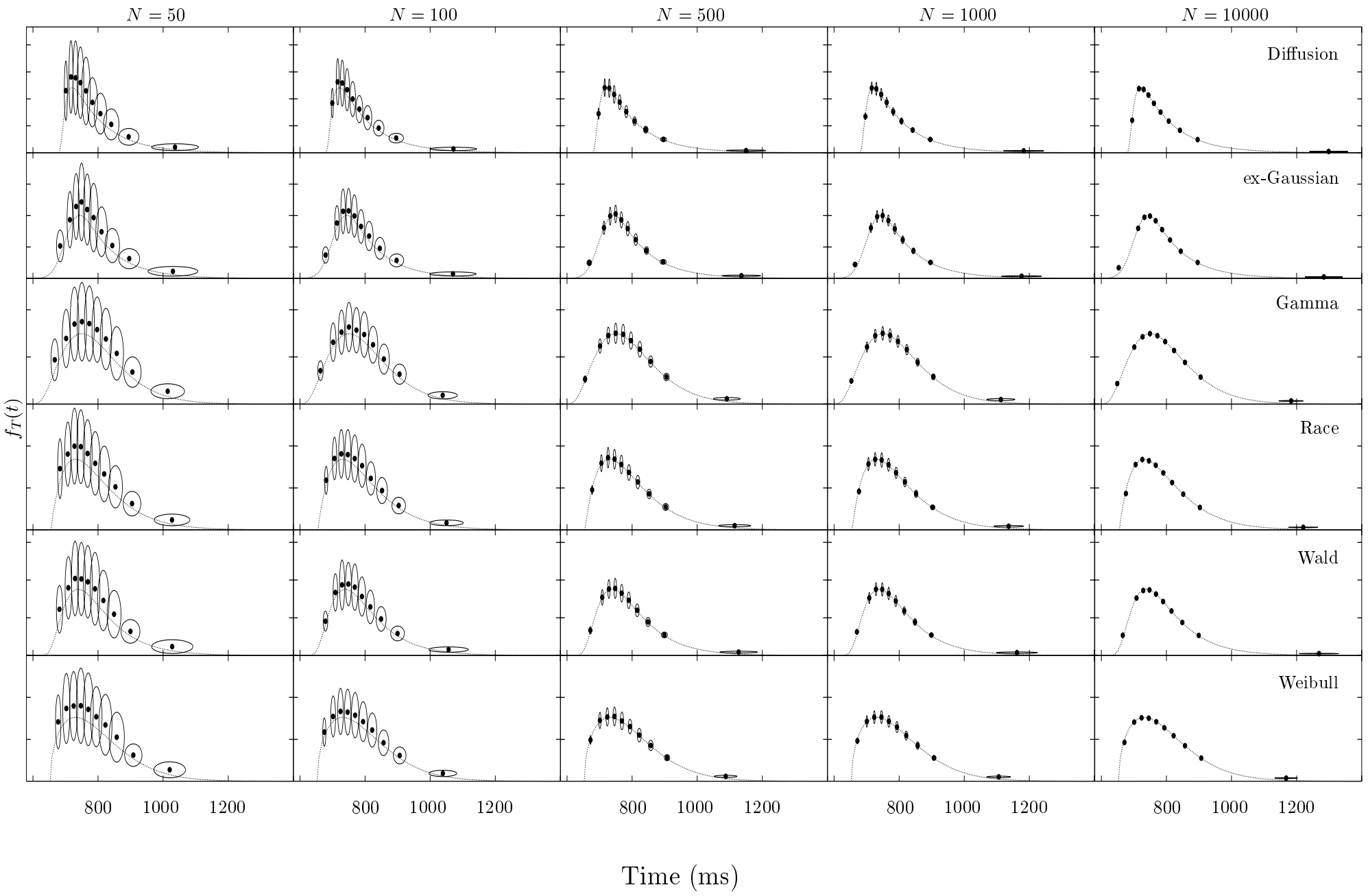
27. Average gamma parameter estimates recovered from least-squares fits of the gamma model to the parametric ex-Gaussian density estimates (SSE and MLE). Error bars are plus and minus one standard deviation. The true value of each parameter is indicated by the solid horizontal line.

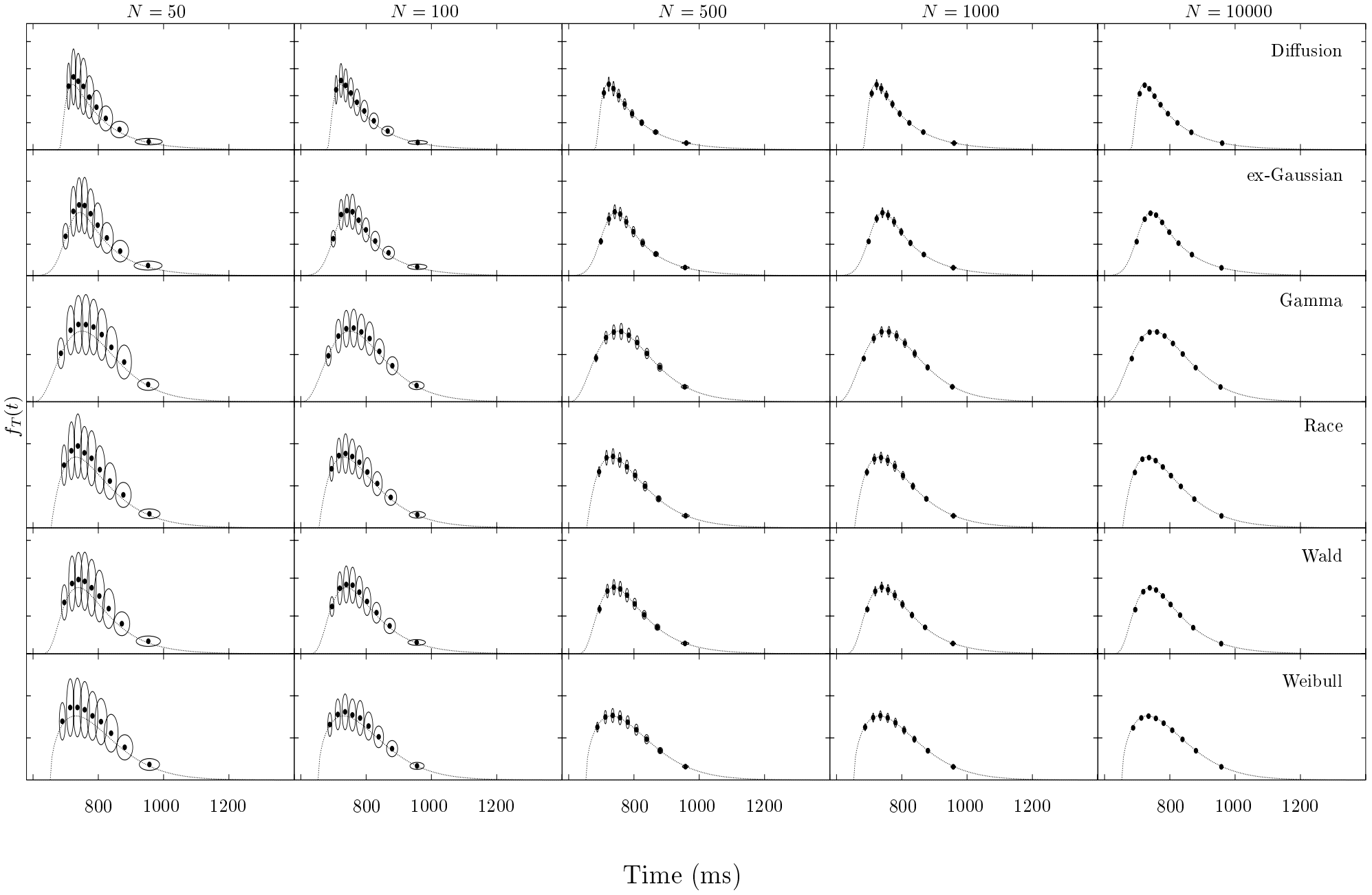

CDF
Time (ms)
0
0.2
0.4
0.6
0.8
1
600
700
800
900
1000
1100
1200
1300
1400
1500

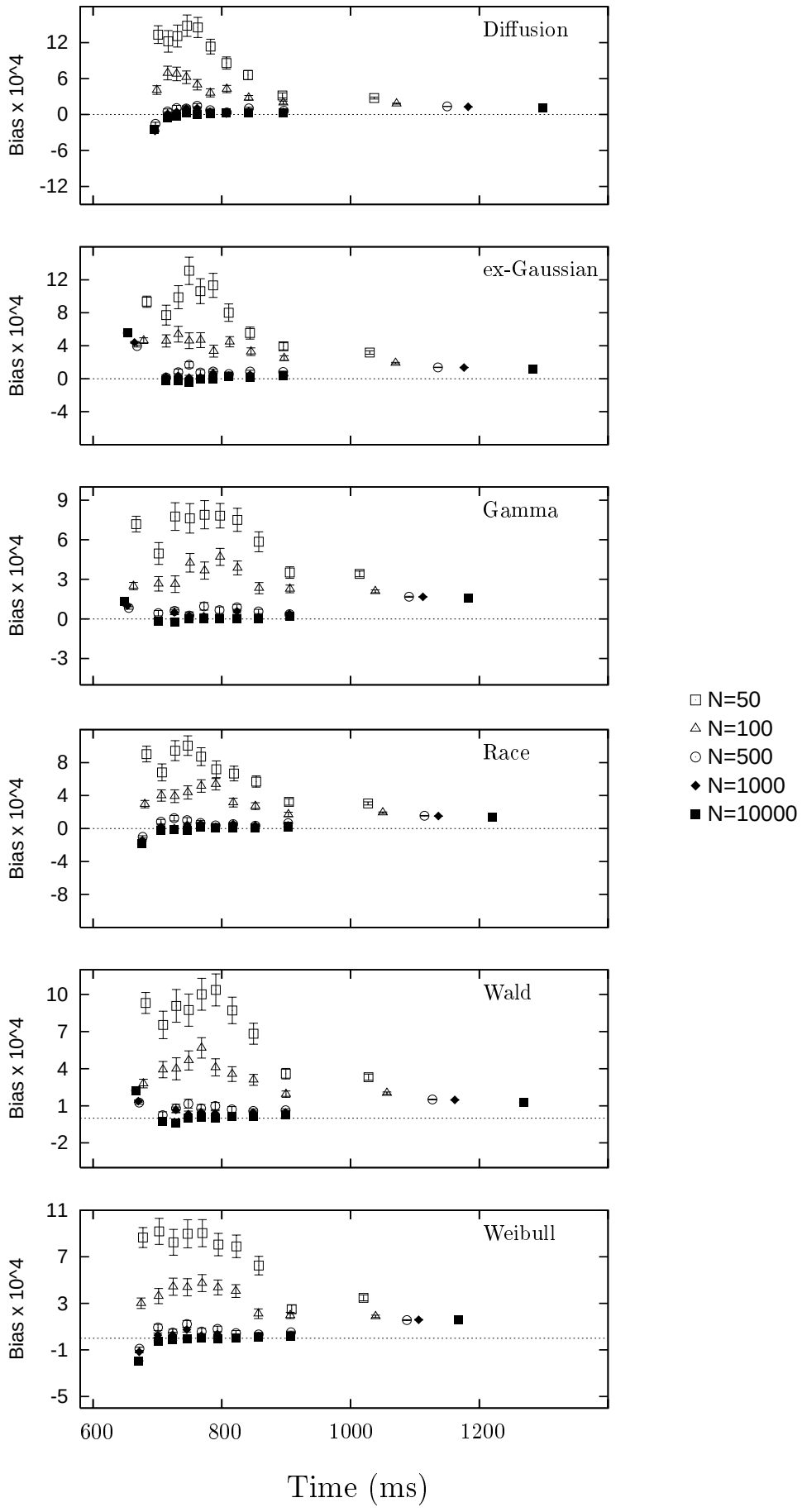

A
B
C
D
Density
Time (ms)
0
0.001
0.002
0.003
0.004
0.005
0.006
0.007
600
700
800
900
1000
1100
1200
1300
1400
1500

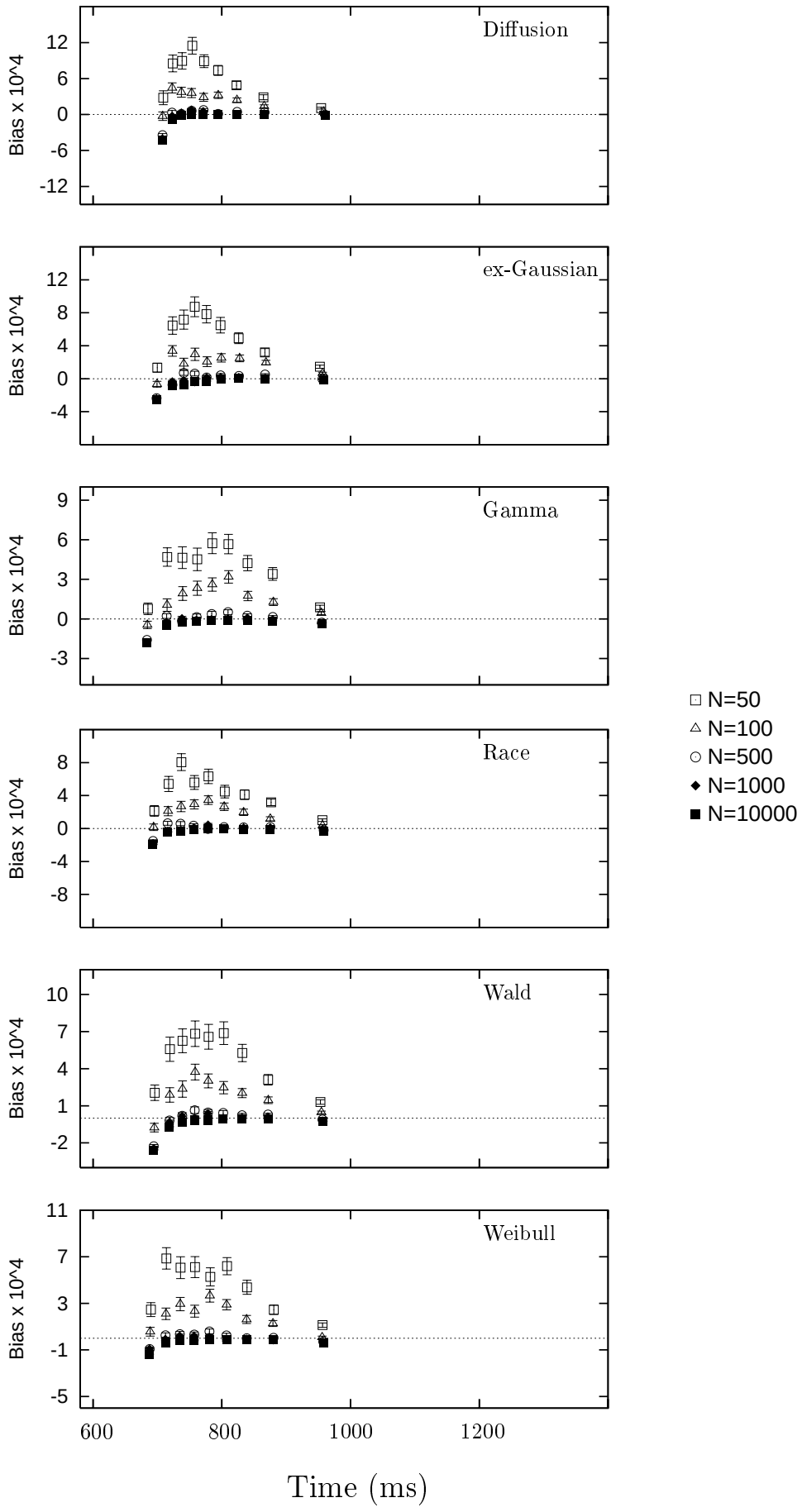

A
$\ln L(\hat{\theta})$
-280
-290
-300
-310
-320
-330
610
620
630
640
650
660
670
680
ξ
100
120
140
160
α

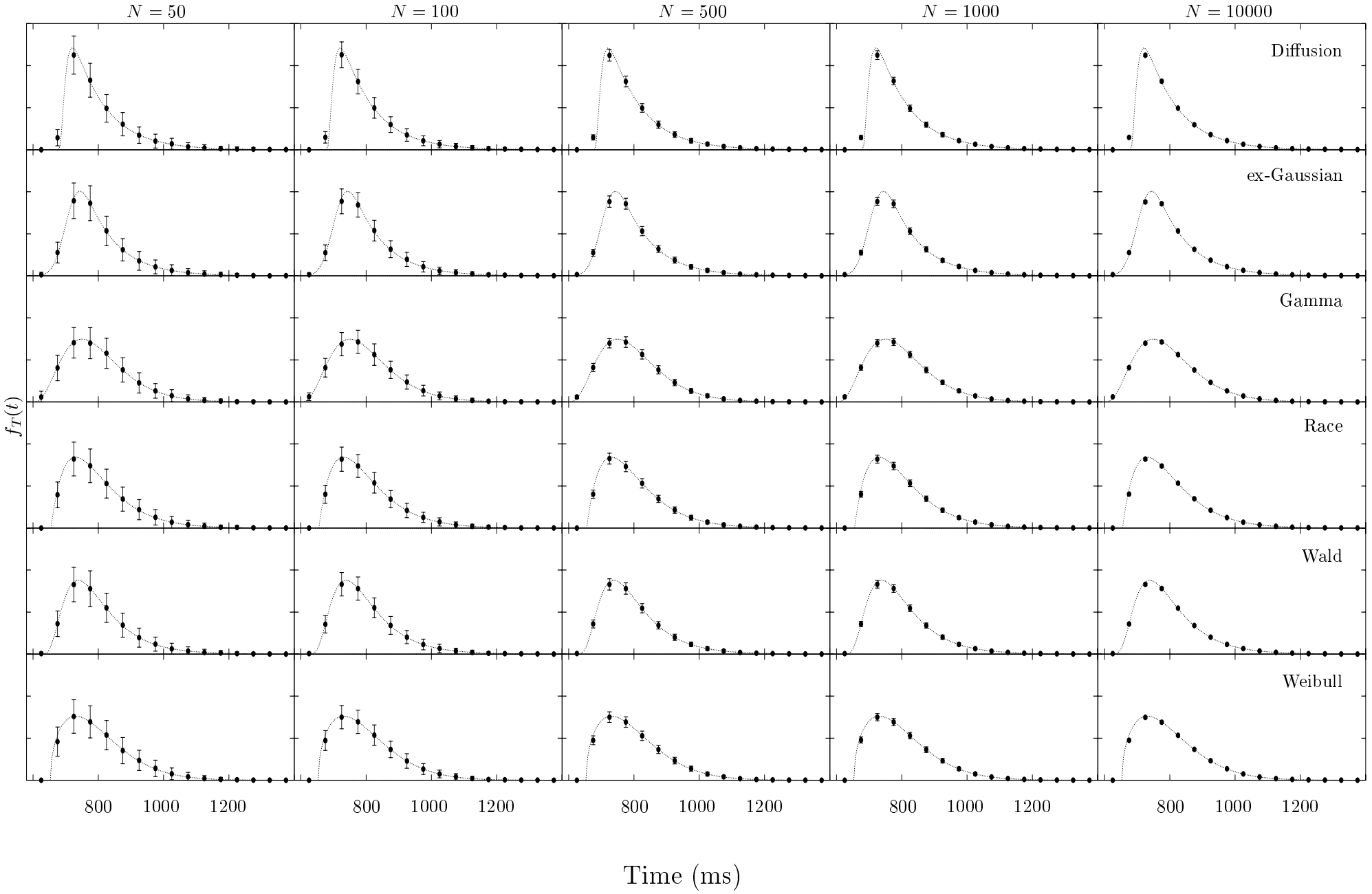

B
SE
0.00014
0.00012
0.0001
8e-05
6e-05
4e-05
2e-05
560
580
600
620
640
ξ
180
190
200
210
α

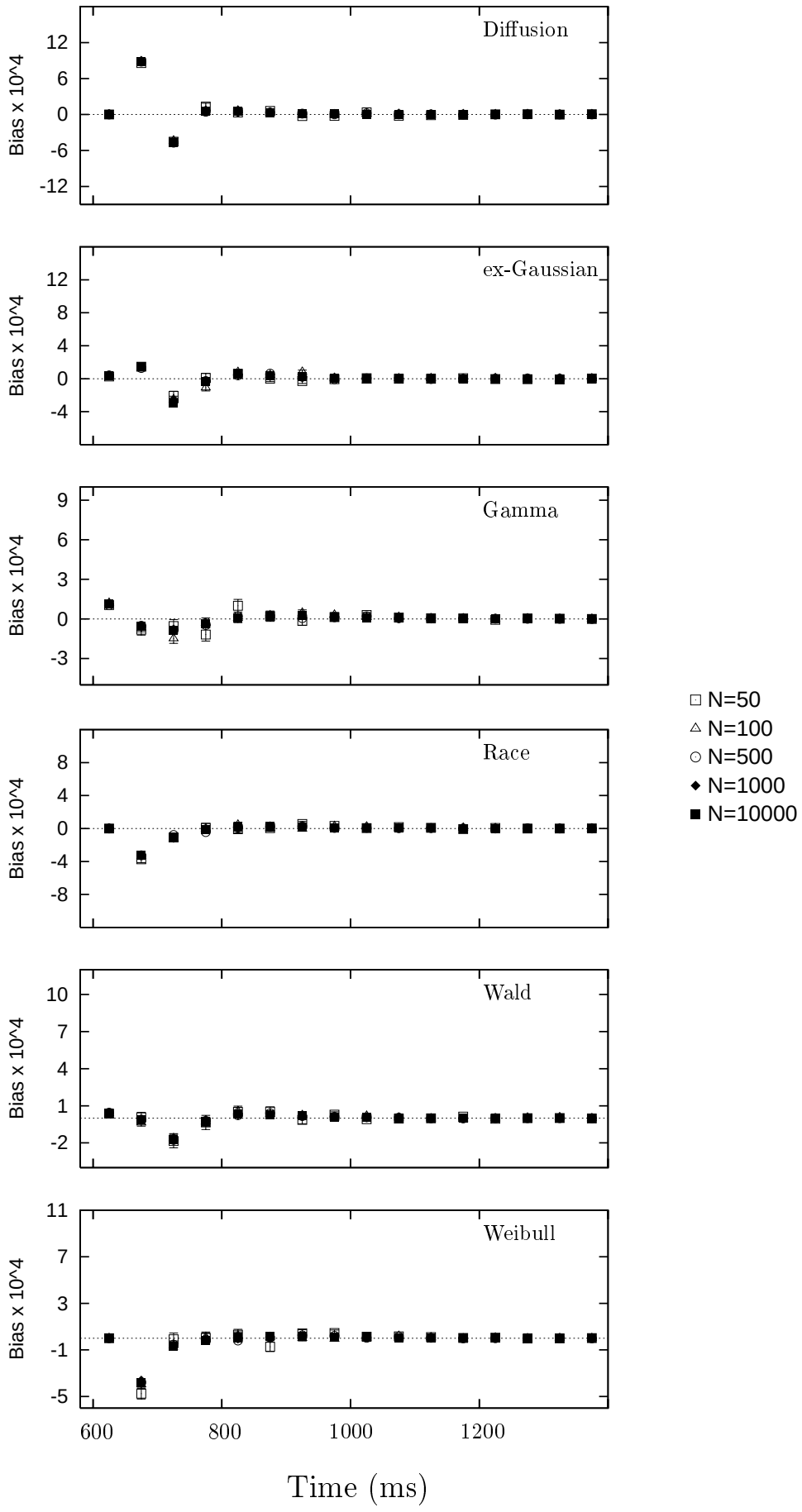

$N = 50$
$N = 100$
$N = 500$
$N = 1000$
$N = 10000$
Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
$f_T(t)$
800
1000
1200
Time (ms)

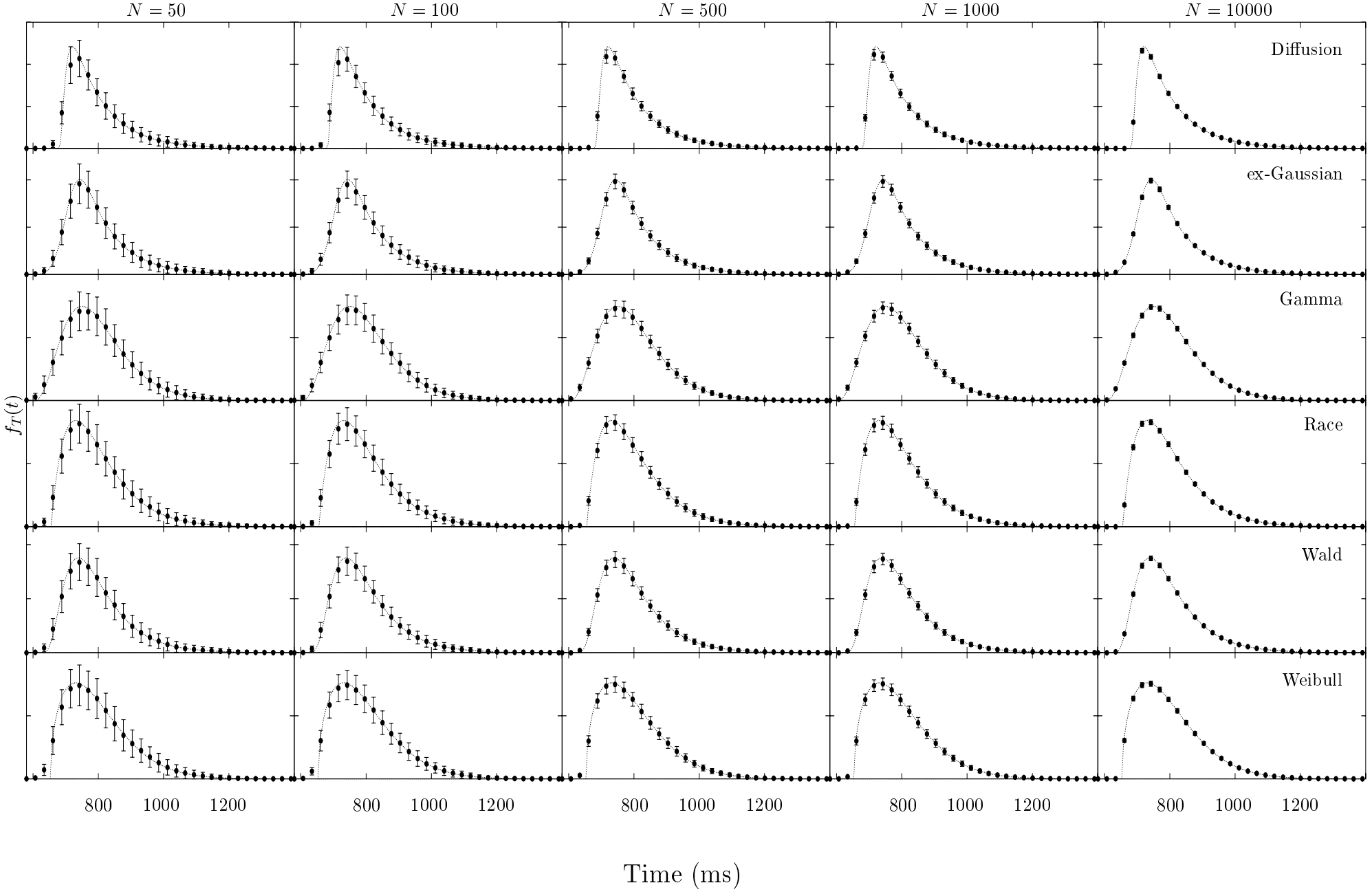

$N = 50$
$N = 100$
$N = 500$
$N = 1000$
$N = 10000$
Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
$f_T(t)$
800
1000
1200
Time (ms)

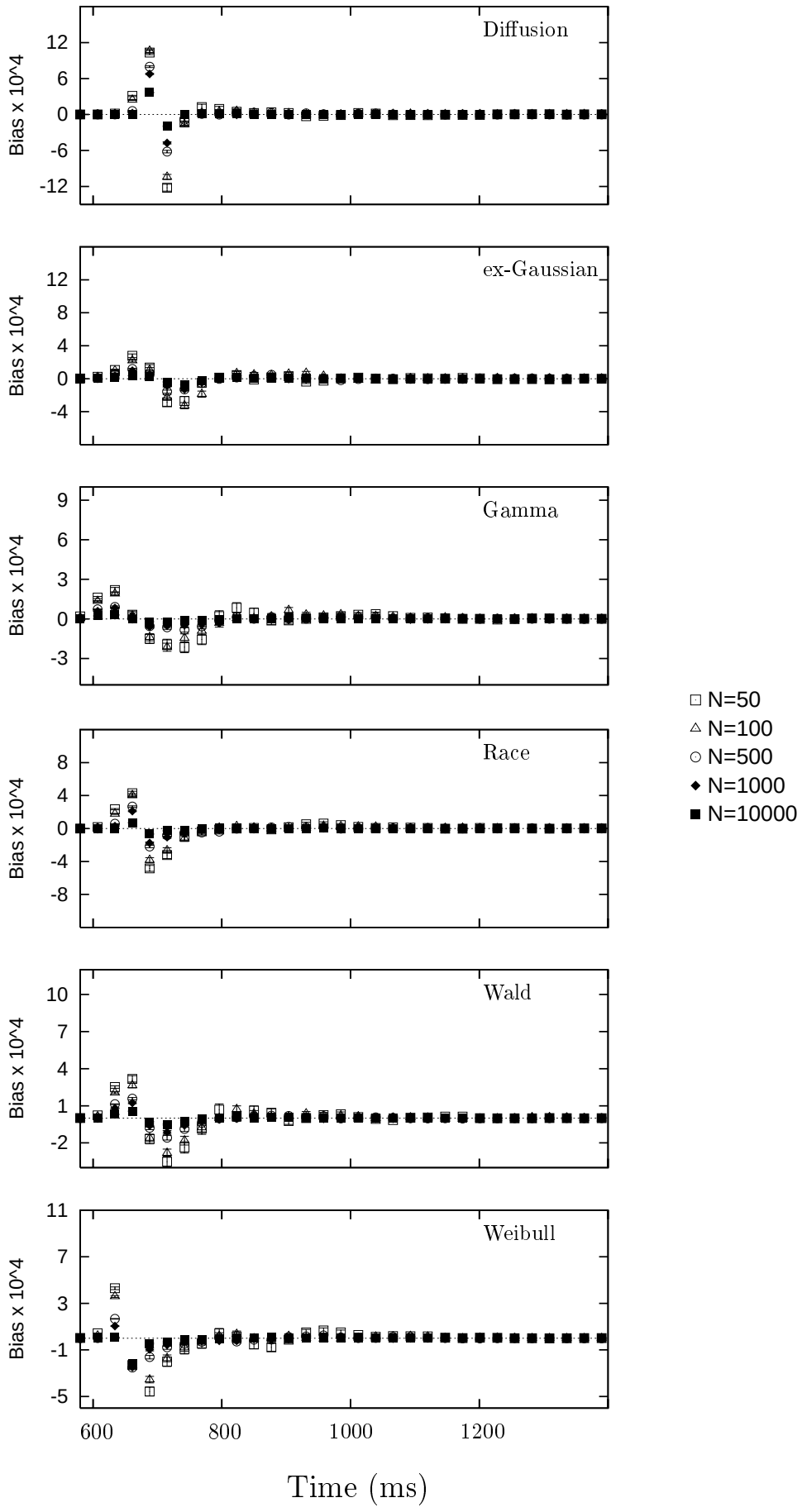

Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
Bias x 10^4
Time (ms)
600
800
1000
1200
N=50
N=100
N=500
N=1000
N=10000

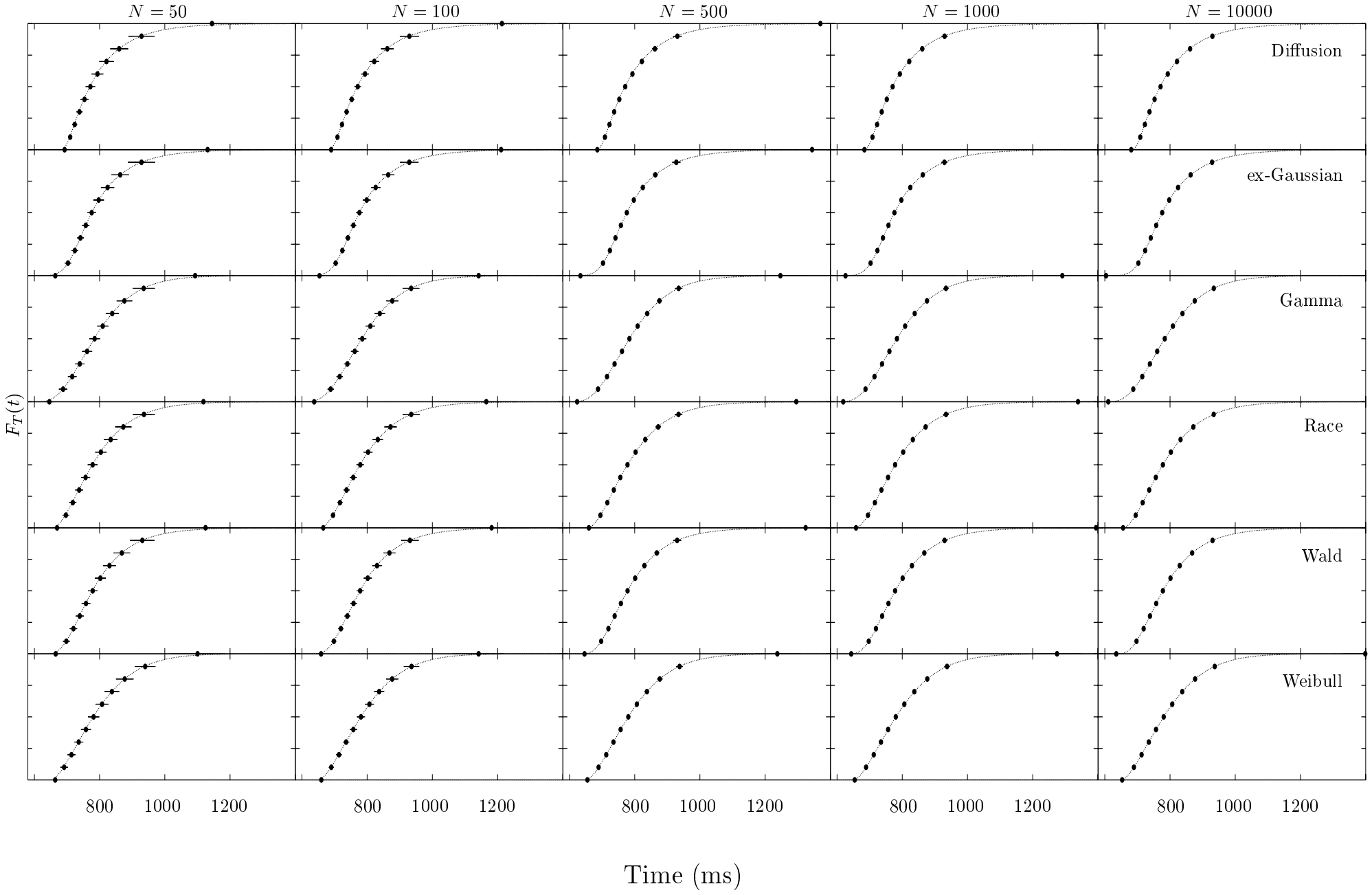

Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
Bias x 10^4
Time (ms)
600
800
1000
1200
N=50
N=100
N=500
N=1000
N=10000

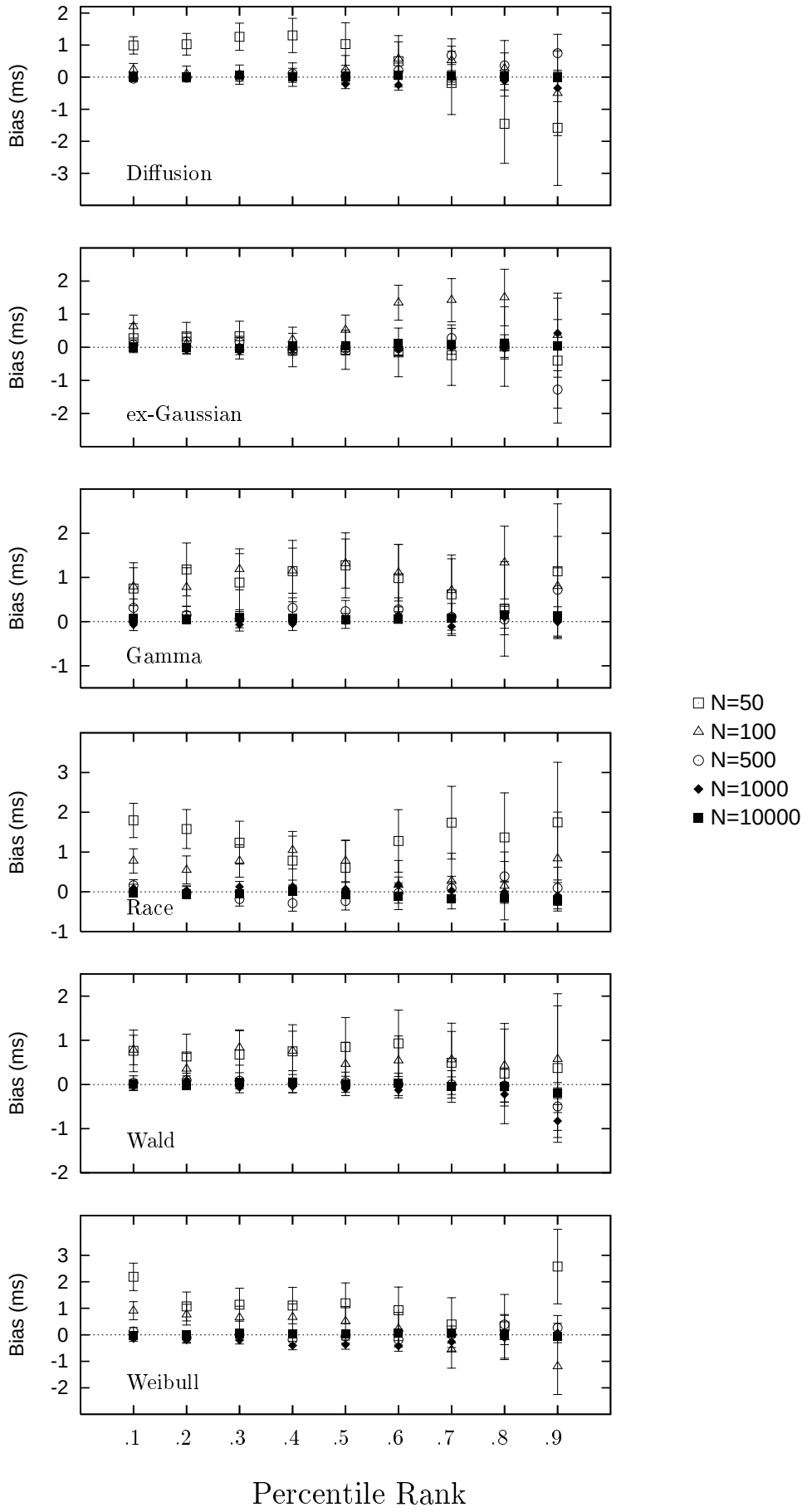

$N = 50$
$N = 100$
$N = 500$
$N = 1000$
$N = 10000$
Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
$f_T(t)$
800
1000
1200
Time (ms)

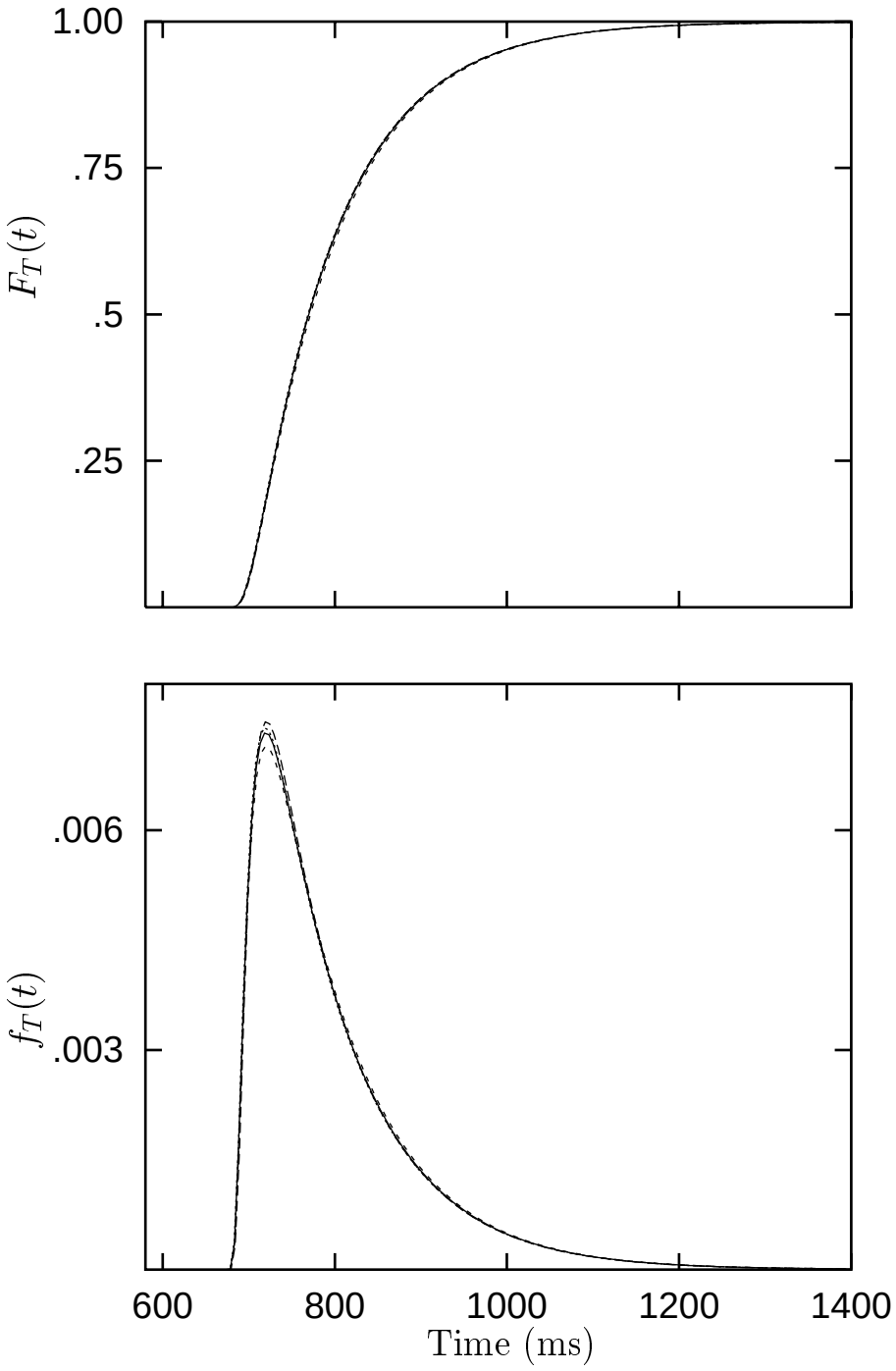

Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
Bias x 10^4
Time (ms)
600
800
1000
1200
N=50
N=100
N=500
N=1000
N=10000

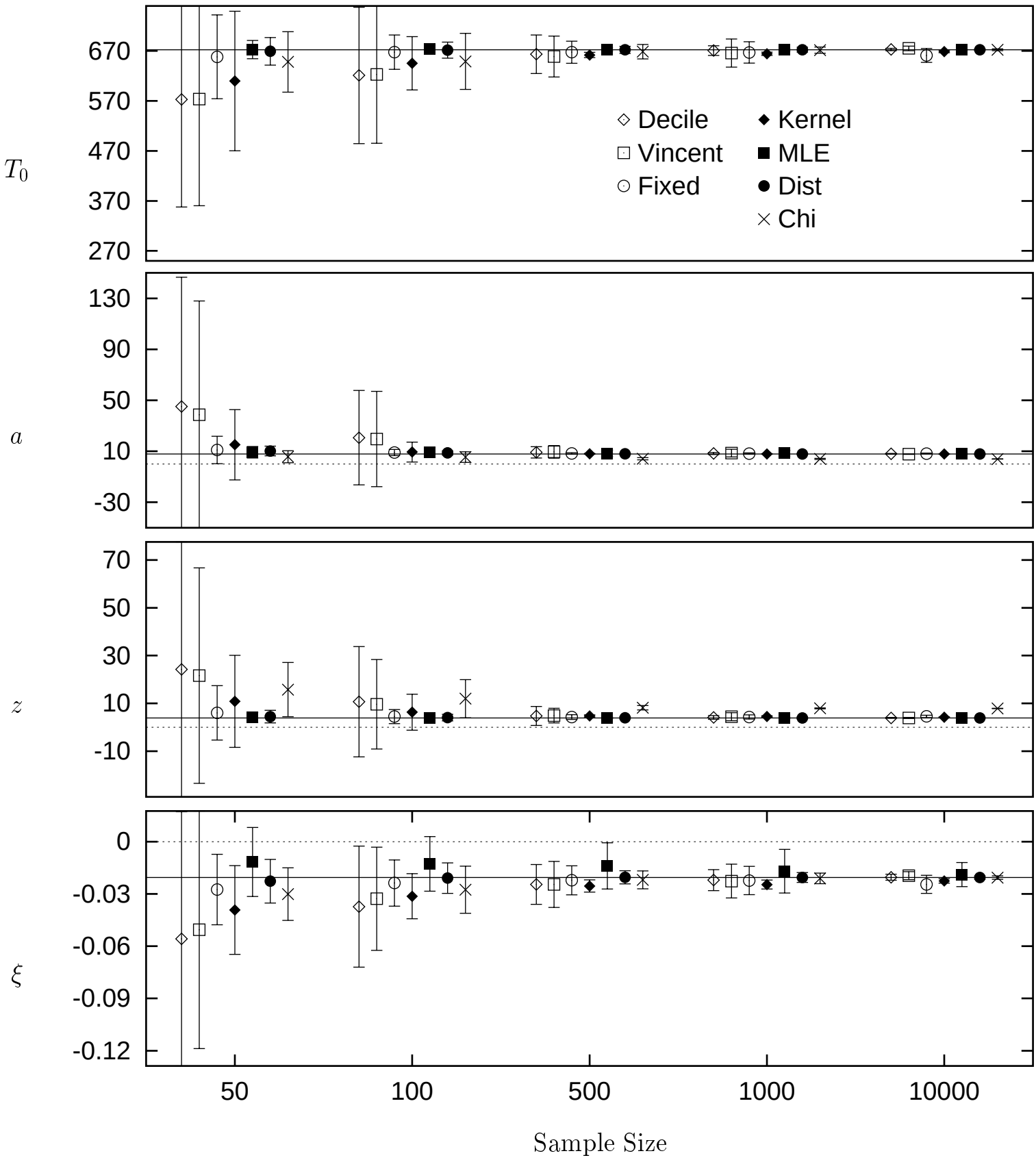

$N = 50$
$N = 100$
$N = 500$
$N = 1000$
$N = 10000$
Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
$f_T(t)$
800
1000
1200
Time (ms)

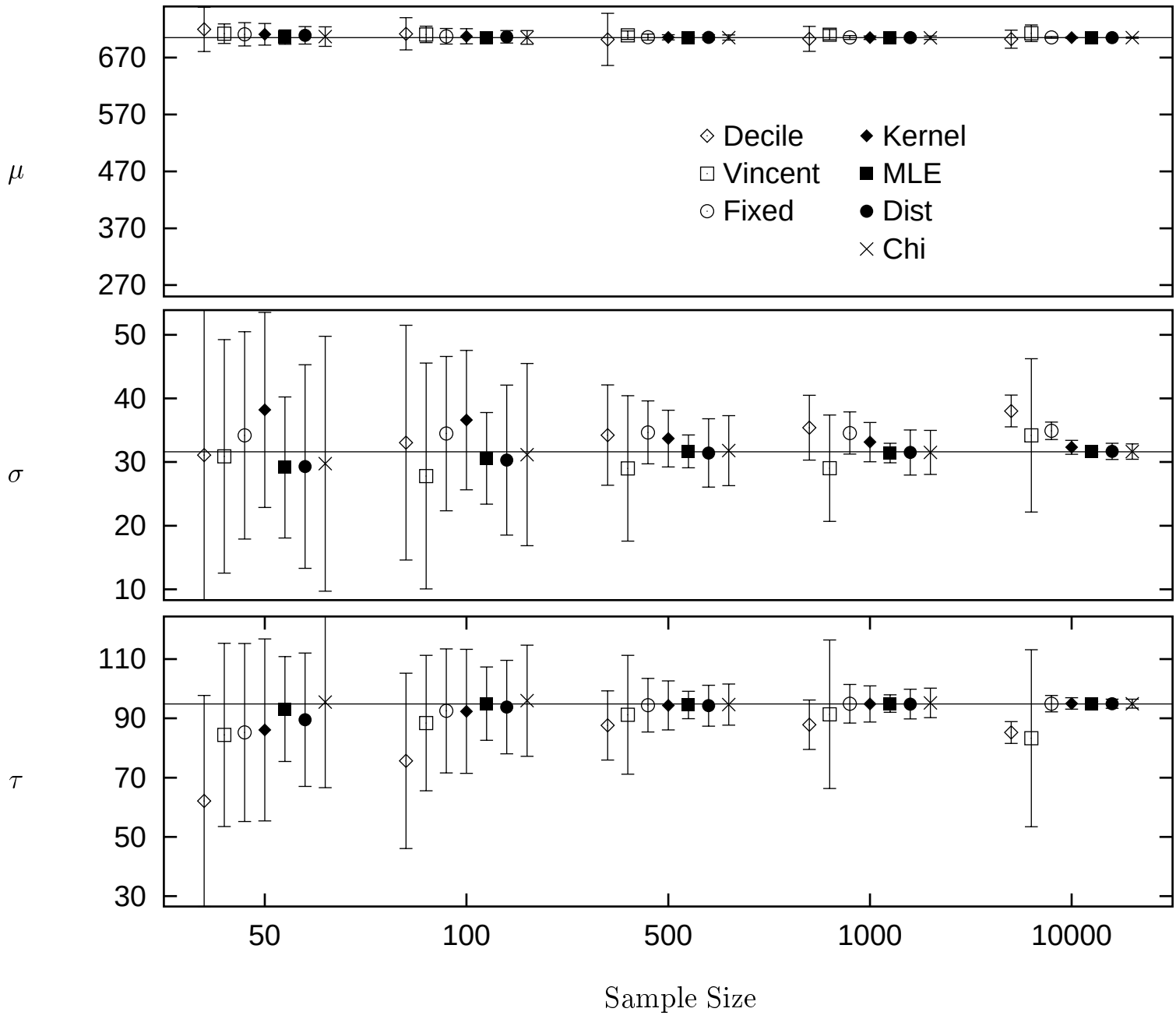

Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
Bias x 10^4
Time (ms)
600
800
1000
1200
N=50
N=100
N=500
N=1000
N=10000

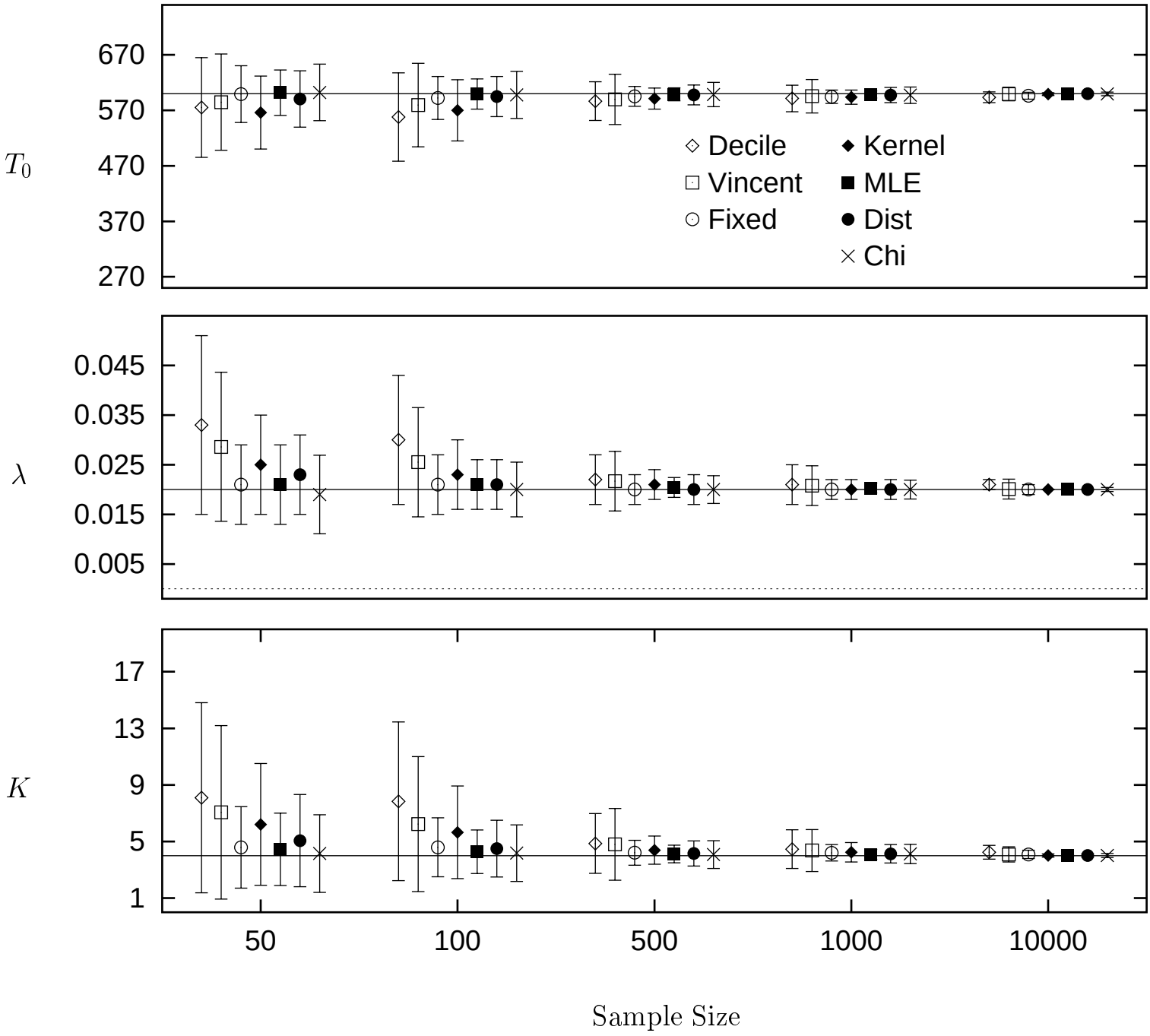

$N = 50$
$N = 100$
$N = 500$
$N = 1000$
$N = 10000$
Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
$F_T(t)$
800
1000
1200
Time (ms)

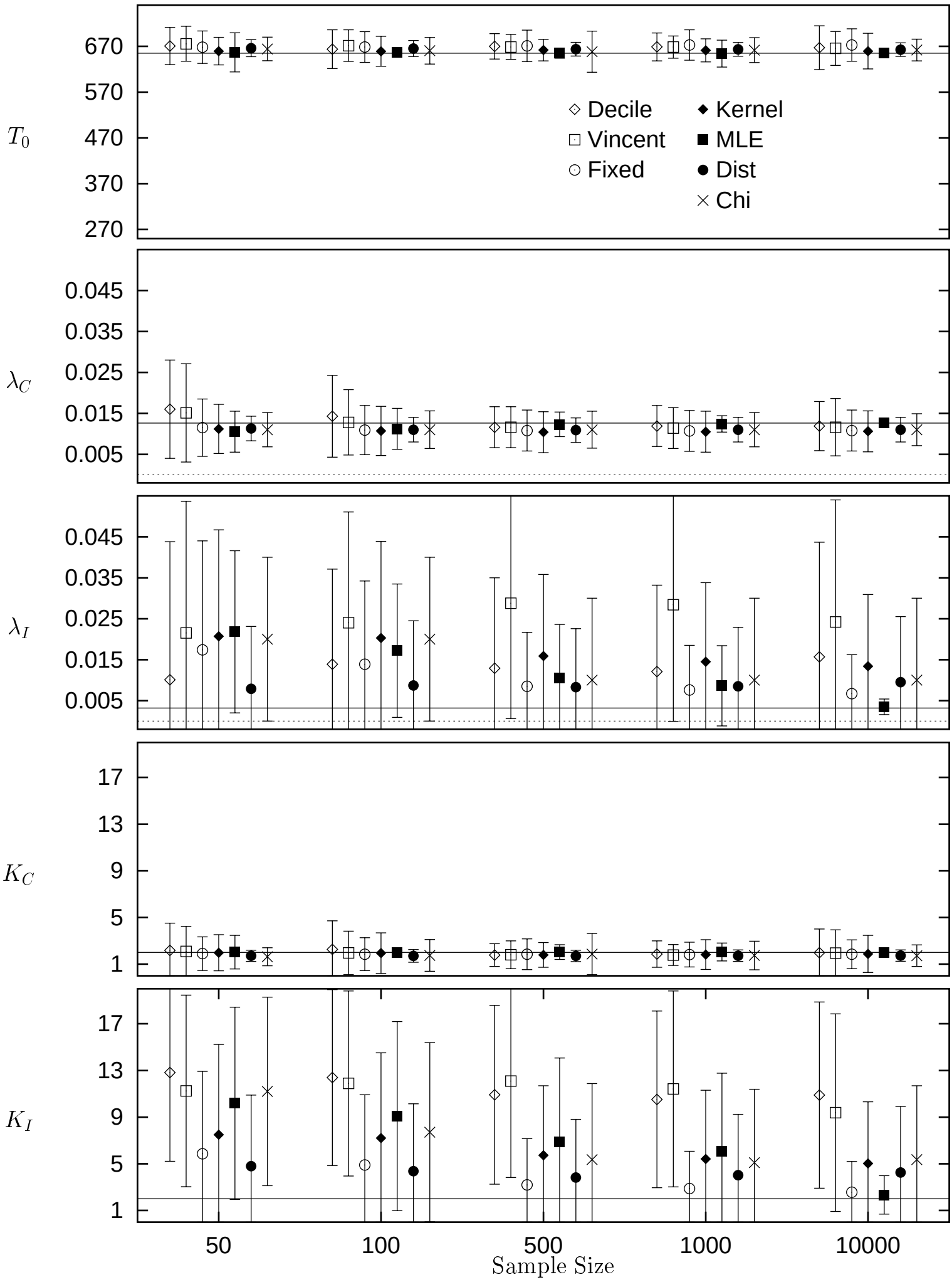

Bias (ms)
Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
.1
.2
.3
.4
.5
.6
.7
.8
.9
Percentile Rank
N=50
N=100
N=500
N=1000
N=10000

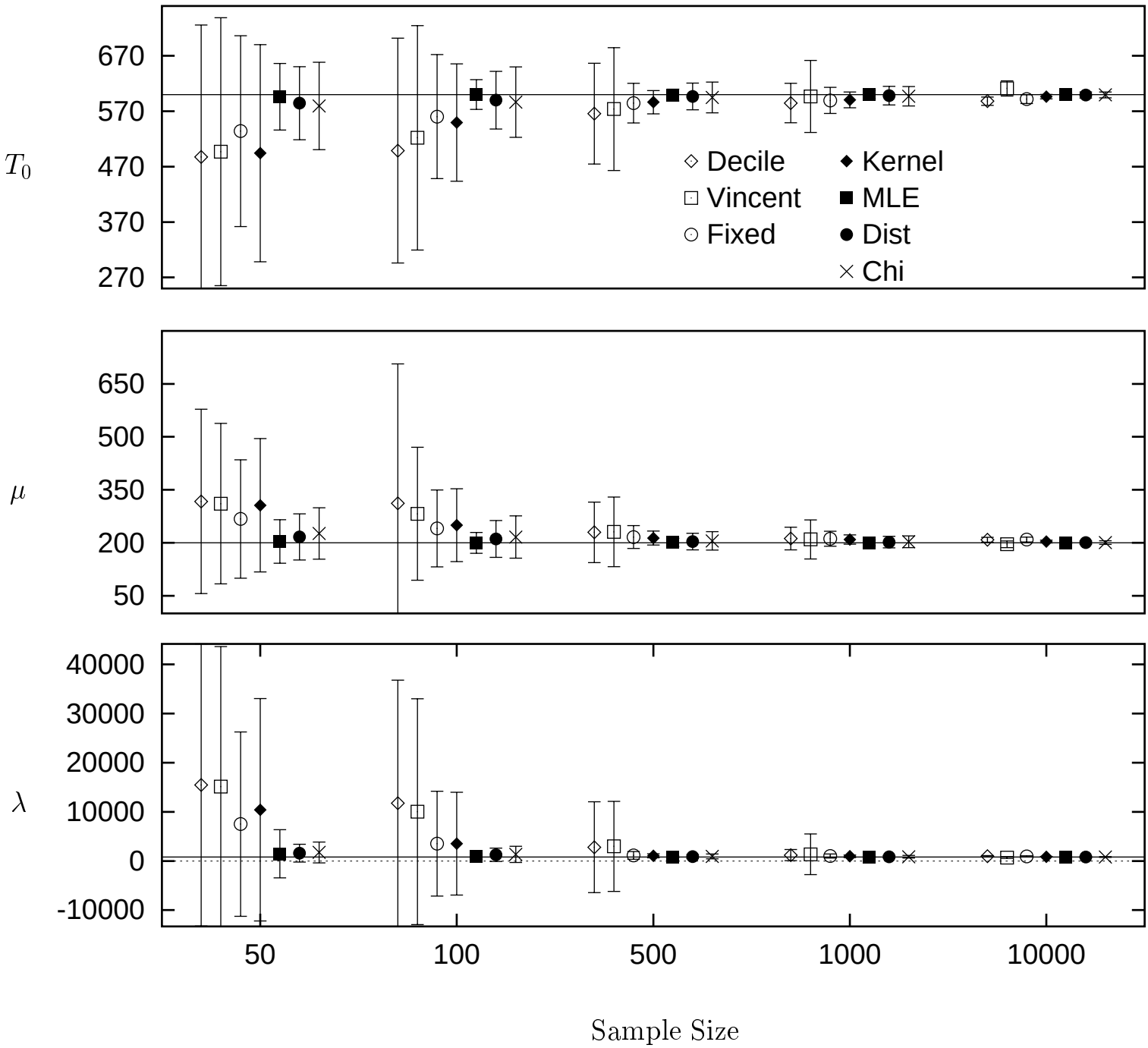

1.00
.75
.5
.25
$F_T(t)$
.006
.003
$f_T(t)$
600
800
1000
1200
1400
Time (ms)

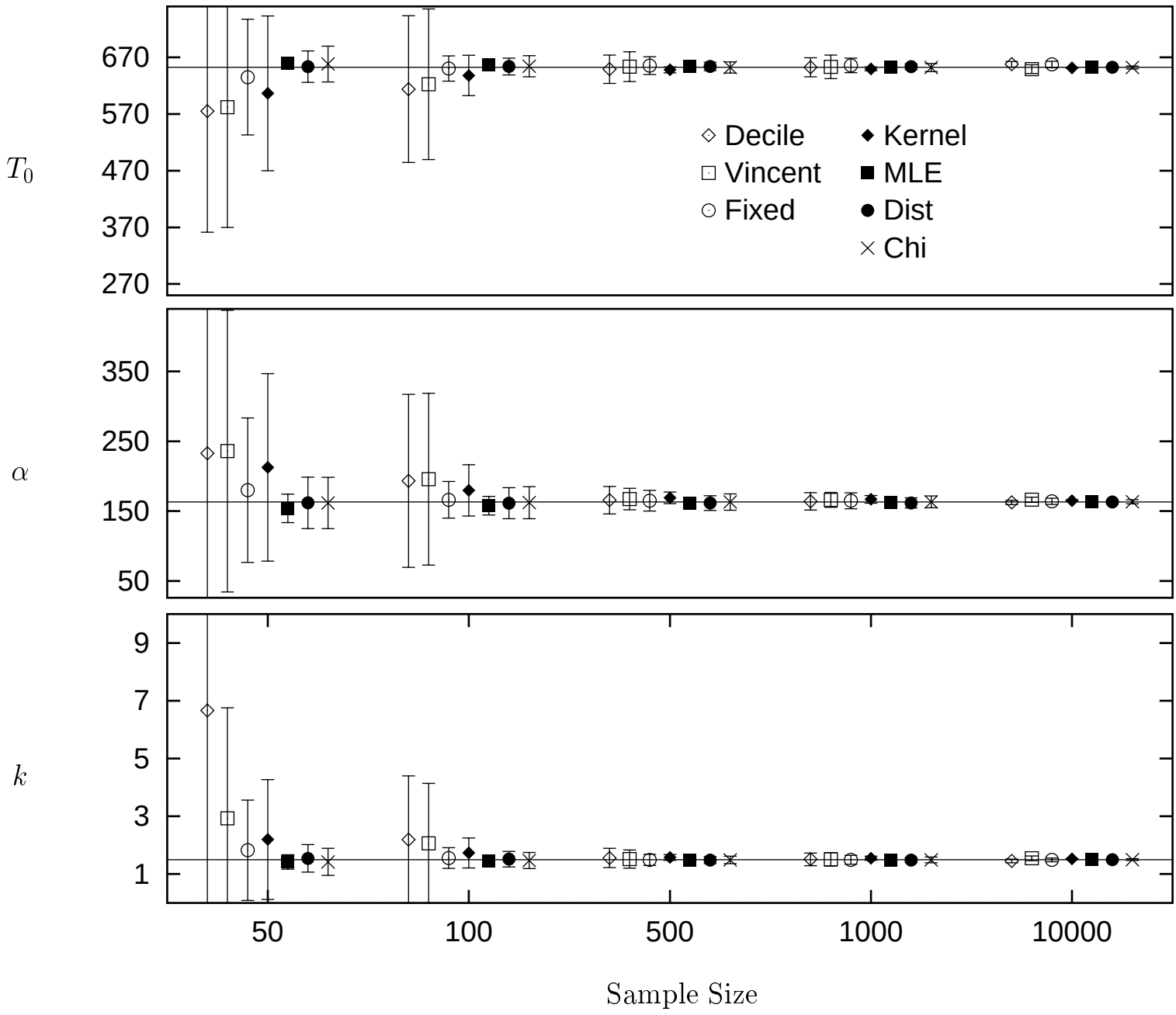

T_0
670
570
470
370
270
Decile
Vincent
Fixed
Kernel
MLE
Dist
Chi
a
130
90
50
10
-30
z
70
50
30
10
-10
ξ
0
-0.03
-0.06
-0.09
-0.12
50
100
500
1000
10000
Sample Size

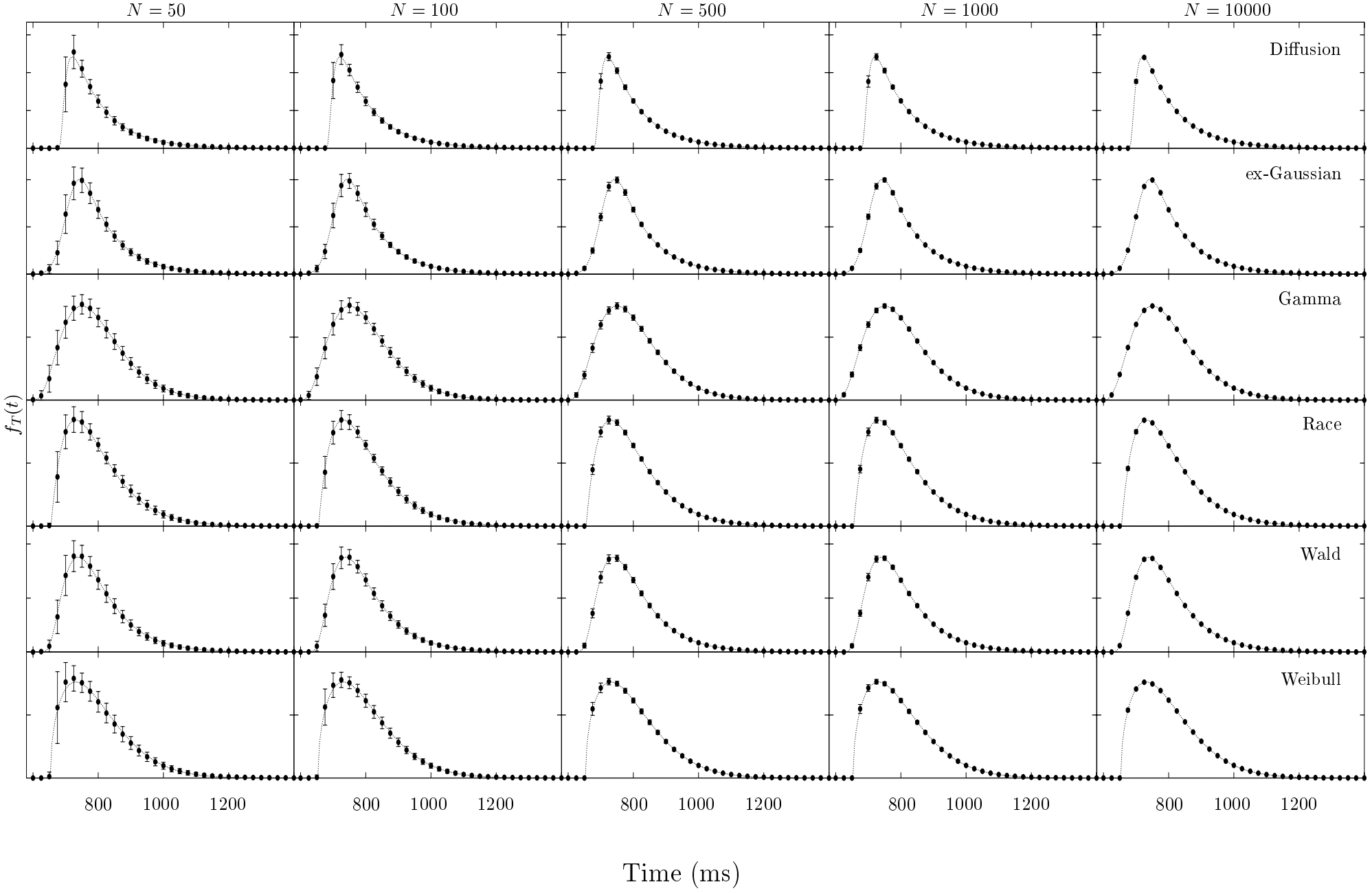

μ
σ
τ
670
570
470
370
270
50
40
30
20
10
110
90
70
50
30
Decile
Vincent
Fixed
Kernel
MLE
Dist
Chi
50
100
500
1000
10000
Sample Size

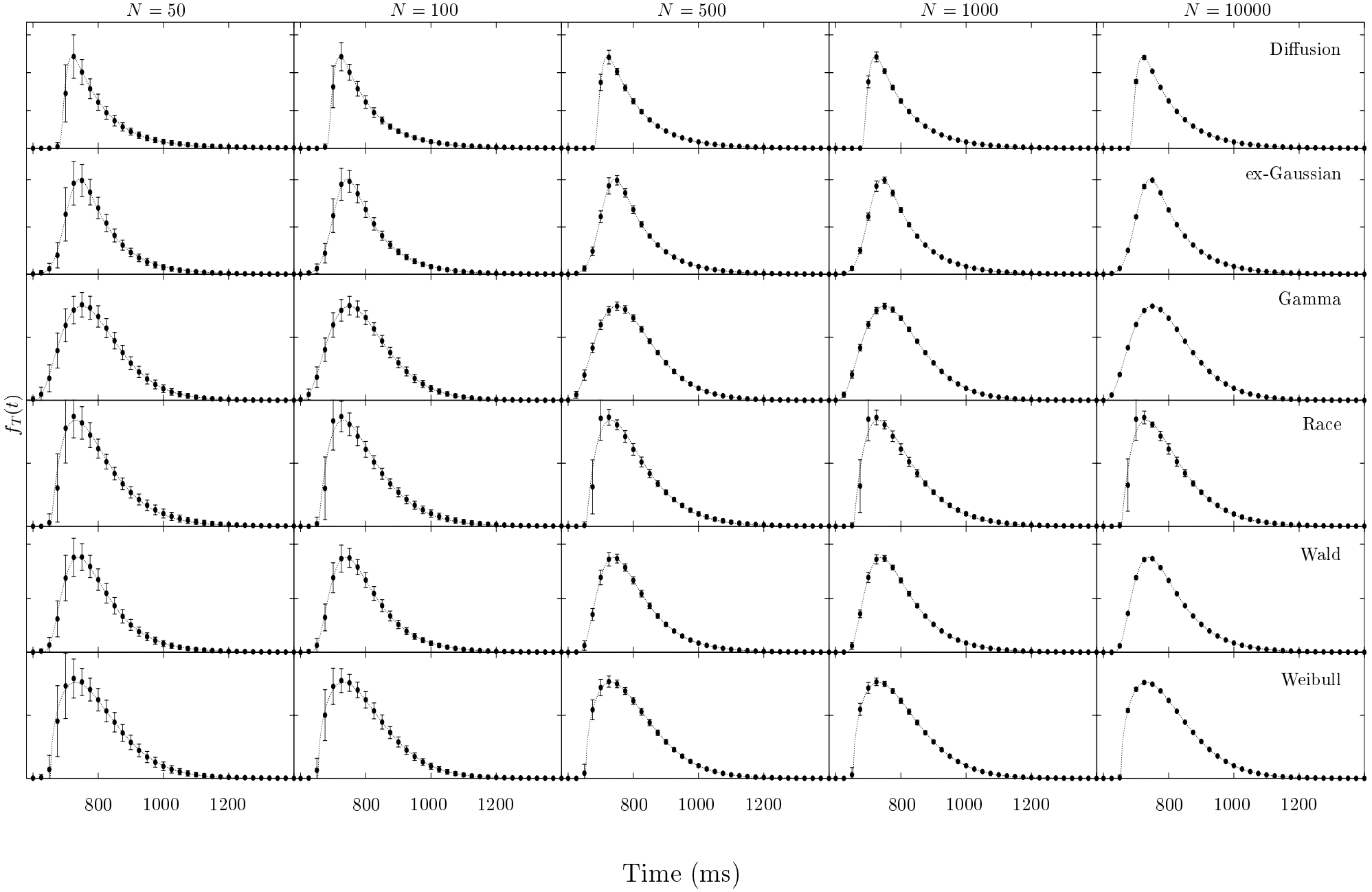

T_0
670
570
470
370
270
◇ Decile
◆ Kernel
□ Vincent
■ MLE
○ Fixed
● Dist
× Chi
λ
0.045
0.035
0.025
0.015
0.005
K
17
13
9
5
1
50
100
500
1000
10000
Sample Size

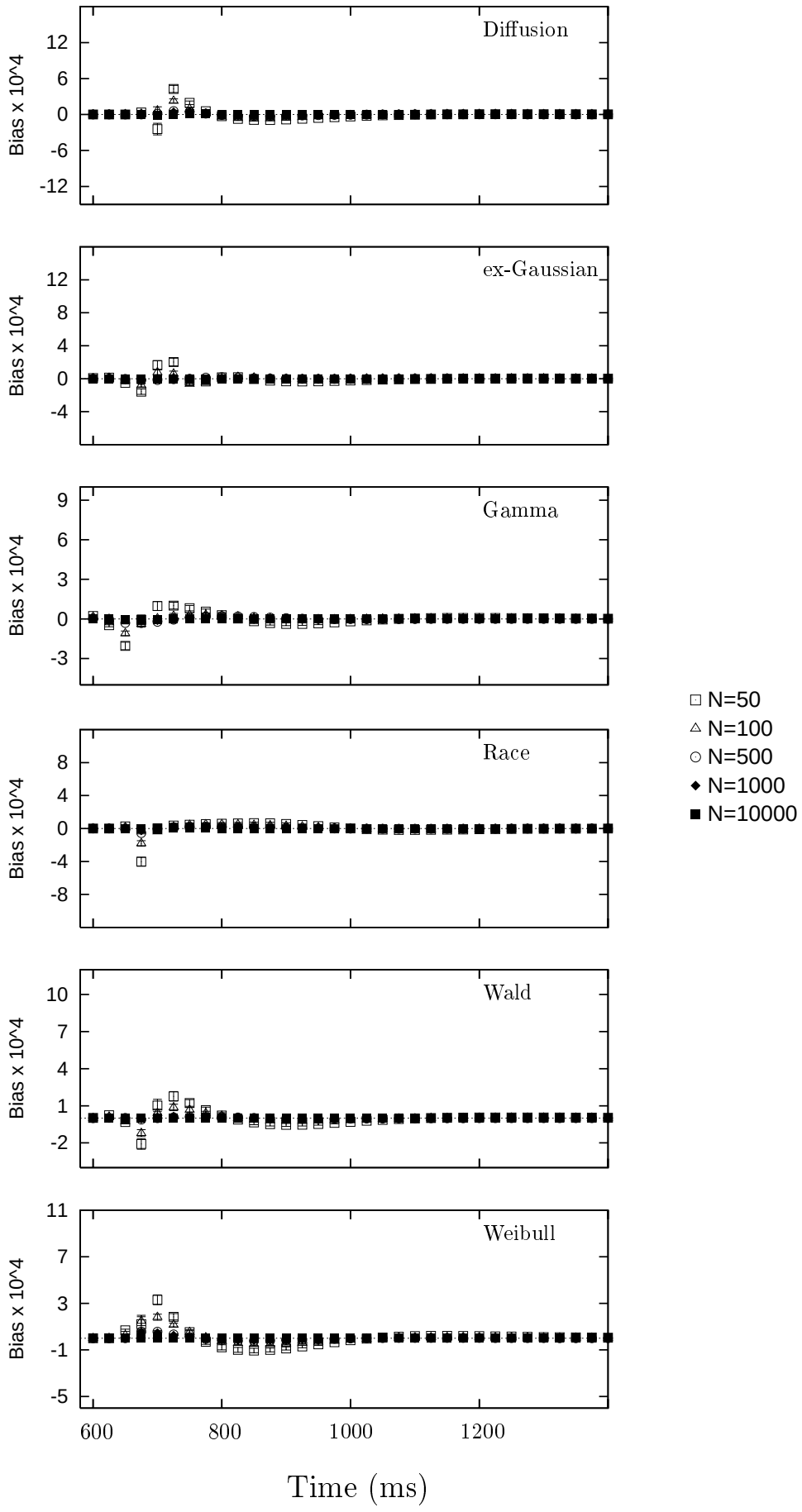

T_0
670
570
470
370
270
Decile
Vincent
Fixed
Kernel
MLE
Dist
Chi
λ_C
0.045
0.035
0.025
0.015
0.005
λ_I
0.045
0.035
0.025
0.015
0.005
K_C
17
13
9
5
1
K_I
17
13
9
5
1
50
100
500
1000
10000
Sample Size

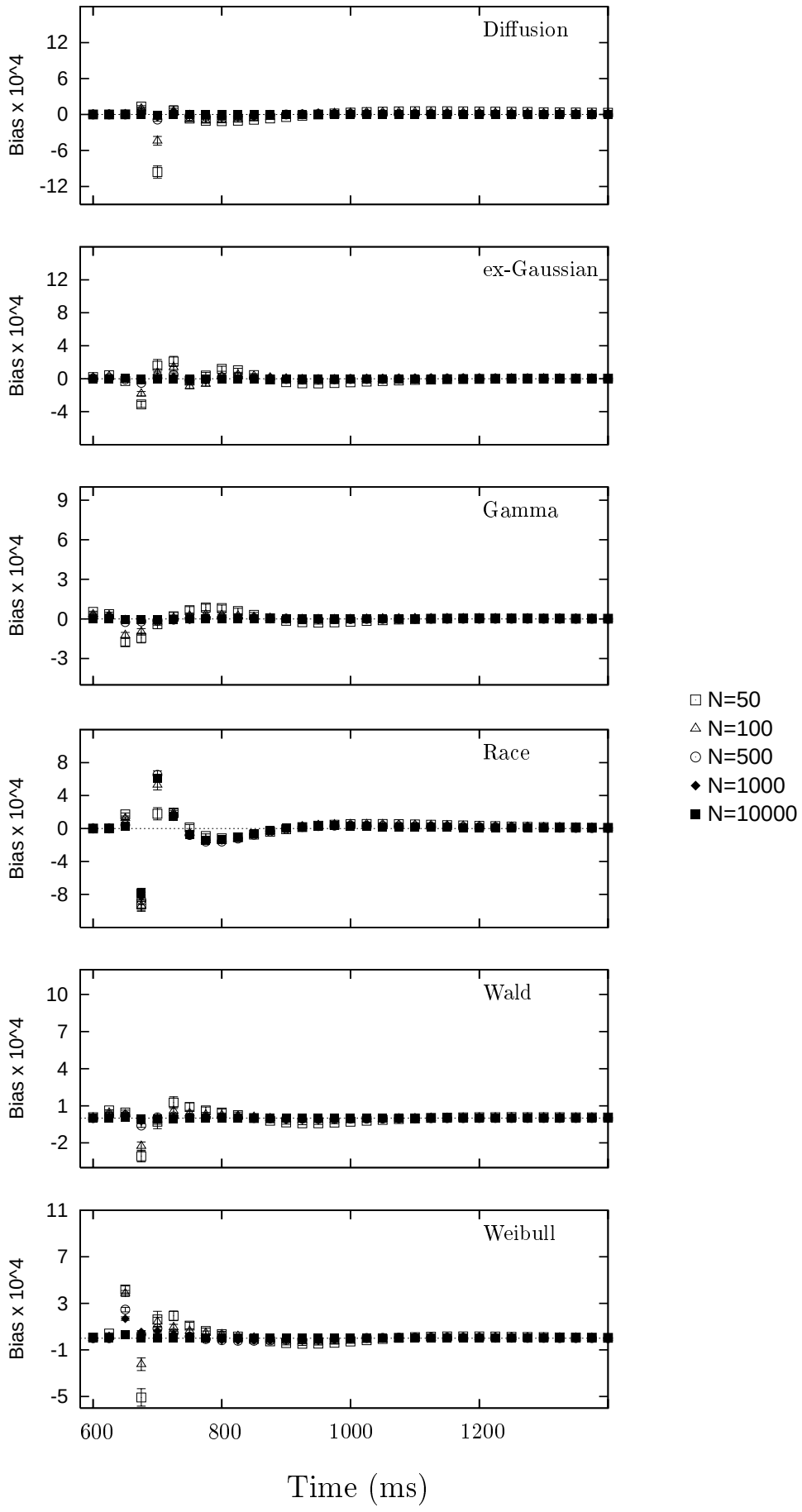

T_0
670
570
470
370
270
Decile
Vincent
Fixed
Kernel
MLE
Dist
Chi
μ
650
500
350
200
50
λ
40000
30000
20000
10000
0
-10000
50
100
500
1000
10000
Sample Size

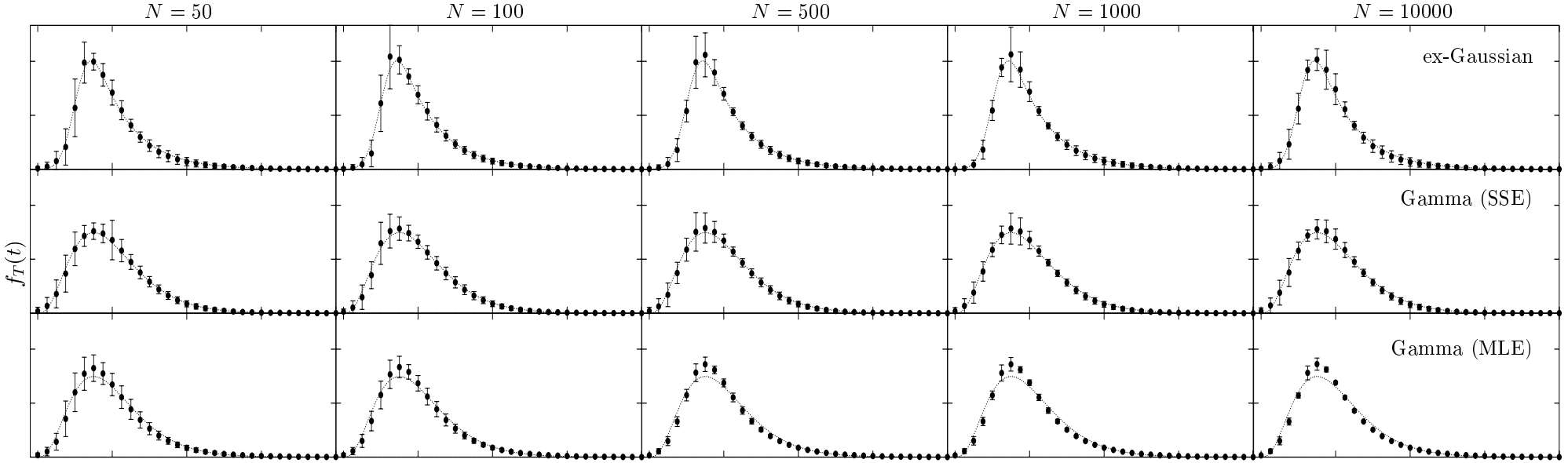

T_0
670
570
470
370
270
Decile
Vincent
Fixed
Kernel
MLE
Dist
Chi
α
350
250
150
50
k
9
7
5
3
1
50
100
500
1000
10000
Sample Size

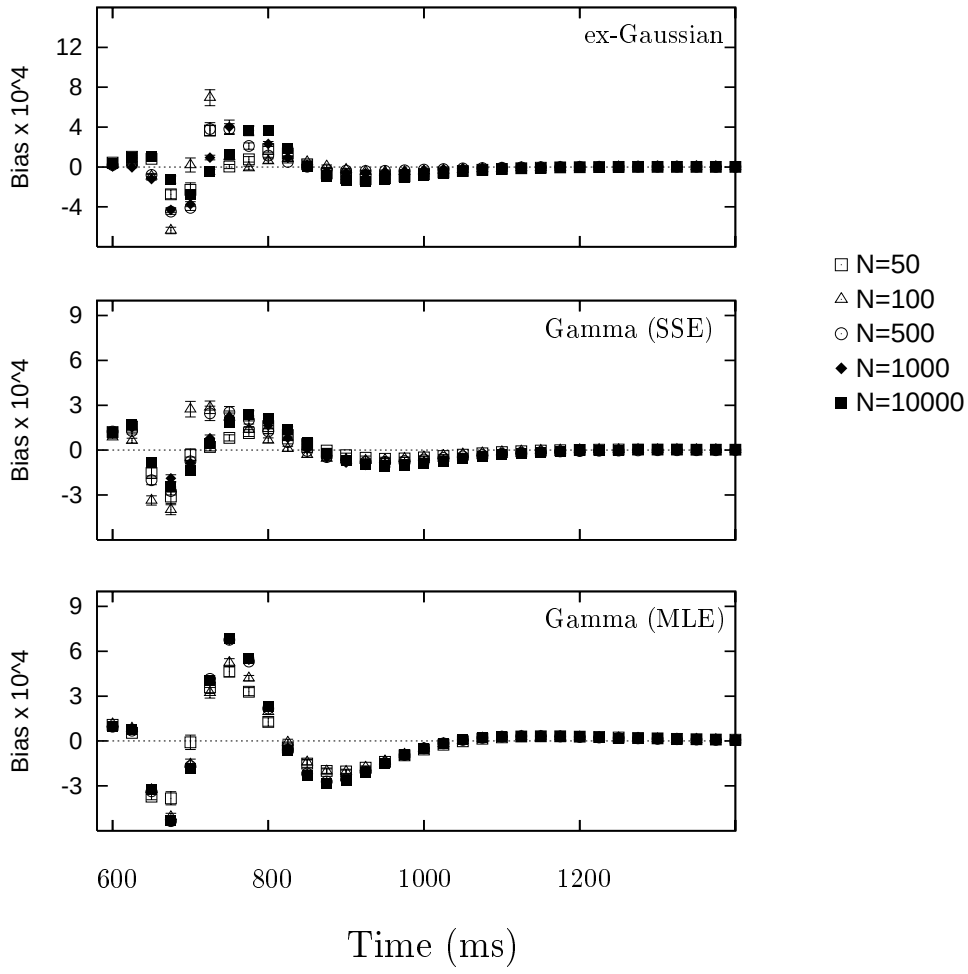

$N = 50$
$N = 100$
$N = 500$
$N = 1000$
$N = 10000$
Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
$f_T(t)$
800
1000
1200
Time (ms)

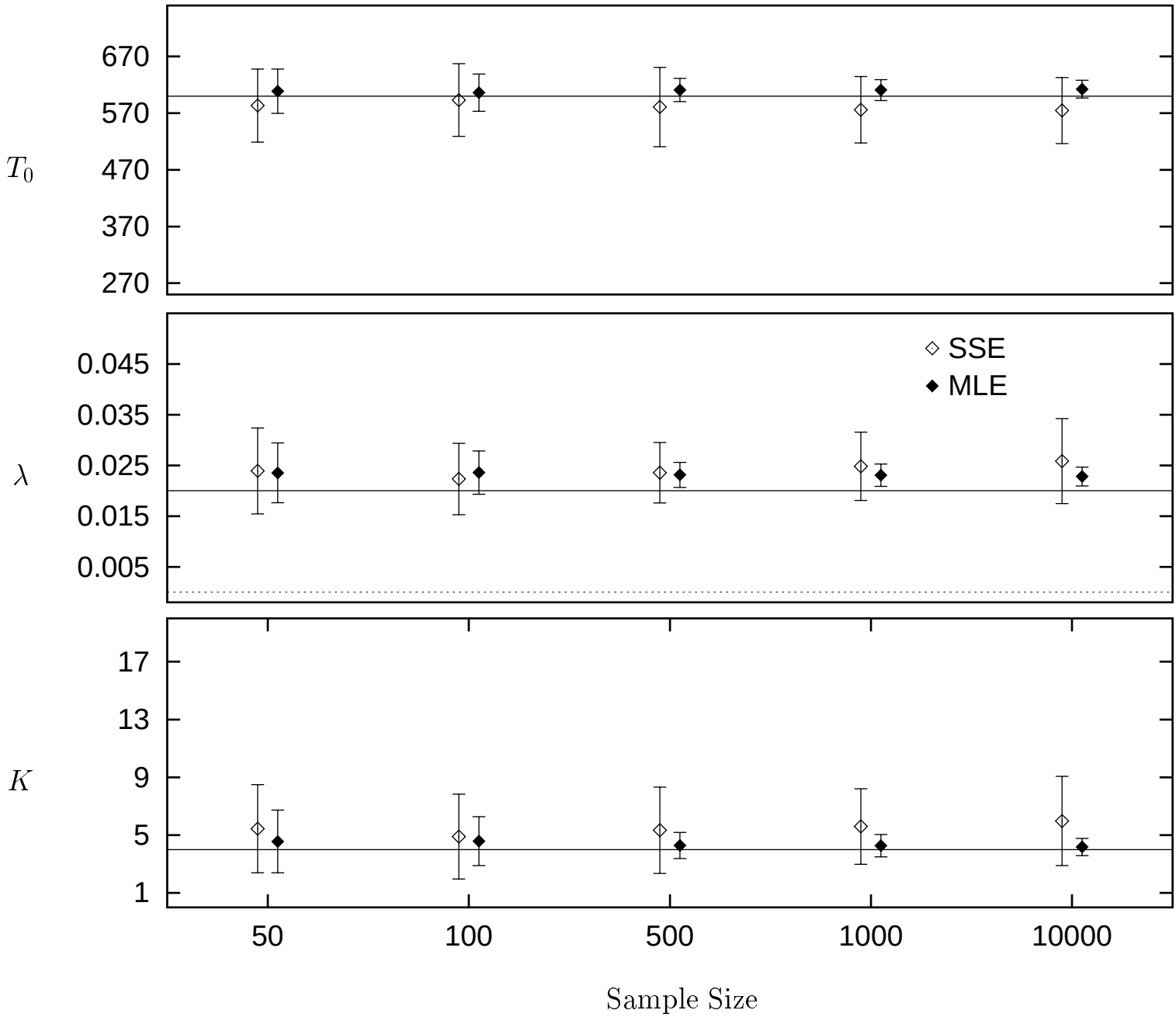

$N = 50$
$N = 100$
$N = 500$
$N = 1000$
$N = 10000$
Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
$f_T(t)$
800
1000
1200
Time (ms)


Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
Bias x 10^4
Time (ms)
600
800
1000
1200
N=50
N=100
N=500
N=1000
N=10000


Diffusion
ex-Gaussian
Gamma
Race
Wald
Weibull
Bias x 10^4
Time (ms)
600
800
1000
1200
N=50
N=100
N=500
N=1000
N=10000


$N = 50$
$N = 100$
$N = 500$
$N = 1000$
$N = 10000$
ex-Gaussian
Gamma (SSE)
Gamma (MLE)
$f_T(t)$


ex-Gaussian
Gamma (SSE)
Gamma (MLE)
Bias x 10^4
12
8
4
0
-4
9
6
3
0
-3
600
800
1000
1200
Time (ms)
N=50
N=100
N=500
N=1000
N=10000


T_0
670
570
470
370
270
λ
0.045
0.035
0.025
0.015
0.005
SSE
MLE
K
17
13
9
5
1
50
100
500
1000
10000
Sample Size