QMPE: Estimating Lognormal, Wald and Weibull RT distributions with a parameter dependent lower bound.

Andrew Heathcote

School of Behavioural Sciences, University of Newcastle, Aviation Building, University Avenue, Callaghan, 2308 NSW, Australia

Ph: 61-2-49215952, andrew.heathcote @newcastle.edu.au

Scott Brown

Department of Cognitive Sciences, University of California at Irvine, Irvine, CA 92697-5100, USA

Ph: +1 949 824 2051 scottb@uci.edu

Denis Cousineau

Département de psychologie Université de Montréal C. P. 6128, succ. Centre-ville Montréal, Québec, H3C 3J7, CANADA

Ph: +1 514 343 7981 denis.cousineau@umontreal.ca

We introduce QMPE, an open source ANSI Fortran90 software for response time distribution estimation. QMPE enables users to estimate parameters for the ex-Gaussian and Gumbel distributions, along with three “shifted” distributions (i.e., distributions with a parameter dependent lower bound), the Lognormal, Wald and Weibull. Estimation can be performed using either the standard maximum likelihood (CML) method, or quantile maximum probability (QMP). We review the properties of each distribution and theoretical evidence showing that CML estimates fail for some cases with shifted distributions, whereas QMP estimates do not. In cases where CML does not fail, a Monte Carlo investigation showed that QMP estimates were usually as good, and in some cases were better, than CML estimates. However, the Monte-Carlo study also uncovered problems that can occur with both CML and QMP estimates, particularly when samples are small and skew is low, highlighting the difficulties of estimating distributions with parameter dependent lower bounds.

This paper describes and tests QMPE (quantile maximum probability estimator), an open source ANSI standard Fortran90 program[1] for estimating the parameters (Θ) of continuous density functions $f(\Theta)$ commonly used to model response time (RT) data. QMPE extends Brown and Heathcote's (in press) QMLE software, which fits only the three parameter ex-Gaussian distribution, to four new positively skewed distributions: the two-parameter Gumbel distribution and the three-parameter shifted Lognormal, shifted Wald, and shifted Weibull distributions. The shifted distributions are bounded below, which makes them attractive as models of RT. Like QMLE, QMPE can fit distributions using continuous maximum likelihood (CML) estimation, and quantile maximum probability estimation (QMP[2], Heathcote, Brown & Mewhort, 2002; Heathcote & Brown, in press). In the next section we describe the distribution functions fit by QMPE. We then describe the estimation methods, and demonstrate that in cases where CML estimation fails for shift distributions, QMP estimates remain efficient. Finally, we report the results of a Monte Carlo study that compares CML and QMP estimation in small samples.

QMPE Distribution Functions

Like Brown and Heathcote's (in press) QMLE program, QMPE fits the ex-Gaussian distribution, a positively skewed distribution produced by the convolution of a normal and exponential distribution (see Heathcote, 1996 for details). The ex-Gaussian has three parameters, the mean (μ) and standard deviation ($\sigma > 0$) of the normal component and the mean of the exponential component ($\tau > 0$). The parameters have a simple relationship to the first three cumulants, the mean (κ_1), the variance (κ_2) and the third central moment ($\kappa_3 = \int (x - \kappa_1)^3 f(x)dx$) given by:

$$\kappa_1 = \mu + \tau \qquad \kappa_2 = \sigma^2 + \tau^2 \qquad \kappa_3 = 2\tau^3$$

The third central moment is a measure of skew, and can be estimated by the method of moments formula: $\hat{\kappa}_3 = \sum (x_i - \bar{x})^3 / n$. The "Fisher Skew" measure, $\gamma_1 = \kappa_3 / \kappa_2^{3/2}$, is also often used to quantify distribution asymmetry as it a dimensionless quantity. Figure 1a shows three examples of ex-Gaussian distributions.

Insert Figure 1 here

Woodworth and Schlosberg (1954) suggested the Lognormal as an empirical approximation to RT distribution. As its name implies, the logarithm of a Lognormally distributed random variable is distributed normally (equivalently, an exponentiated normal random variable has a Lognormal distribution). The Lognormal distribution is the asymptotic distribution of the product of random variables (McClelland's, 1979, cascade model is an example of a multiplicative model, see Ulrich & Miller, 1993 for more details). Hence, the Lognomal distribution can be motivated as an approximation to the finishing time of a series of stages with randomly varying rates (West & Shlesinger, 1990). Breukelen (1995) shows that two well know parallel models are also compatible with the Lognormal distribution.

The Lognormal distribution is bounded below by zero ($x > 0$) and has parameters corresponding to the normal distribution mean (μ) and standard deviation ($\sigma > 0$). In order to allow for a lower bound greater than zero, we added a shift parameter, $\theta > 0$. The corresponding density, which is only defined for $\theta > x$, is:

$$f(x) = \frac{1}{(x-\theta)\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x-\theta)-\mu}{\sigma}\right)^2}$$

Ratcliff and Murdock (1976) reported that the shifted Lognormal provided as good a fit as the ex-Gaussian to RT distribution data from recognition memory experiments.

Figure 1b shows three examples of the Lognormal density. In order to specify the first three cumulants it is convenient to define $\omega = \exp[\sigma^2]$:

$$\kappa_1 = \theta + e^{\mu + \frac{1}{2}\sigma^2} \qquad \kappa_2 = e^{2\mu}\omega(\omega - 1) \qquad \kappa_3 = e^{3\mu}\omega^{\frac{3}{2}}(\omega - 1)^2(\omega + 2)$$

The Wald distribution (Wald, 1947) can be motivated as a model of RT by a continuous approximation to the sequential acquisition of information. Suppose that at each time step identical, normally distributed observations with mean $\mu > 0$ are cumulated. A decision to respond is made when the sum exceeds some criterion, $a > 0$. Given $\mu << a$, the number of steps to exceed the criterion is approximately Wald distributed. As described by McGill (1963), in the limit the discrete steps can be replaced by a continuous time variable, resulting in a diffusion process with an exactly Wald distributed stopping time.

The Wald distribution is bounded below by zero ($x > 0$). We added a shift parameter, $\theta > 0$, to allow for a lower bound greater than zero. The corresponding density, which is only defined for $\theta > x$, is:

$$f(x) = \frac{a}{\sqrt{2\pi(x-\theta)^3}} \exp\left[-\frac{(a - \mu(x-\theta))^2}{2(x-\theta)}\right]$$

Figure 1c shows three examples of the Wald density. The first three cumulants are:

$$\kappa_1 = \theta + a/\mu \qquad \kappa_2 = a/\mu^3 \qquad \kappa_3 = 3a/\mu^5$$

The Wald distribution is also used with a different parameterisation in terms of its mean, a/μ, and its “dispersion”, $\lambda = a^2$. In this form the distribution is often called the “Inverse Gaussian”[3]. When we tested this parameterisation with QMPE, we found that parameter estimates, particularly for λ, were very biased and inefficient. Hence, QMPE uses the “diffusion” parameterisation, which also has the advantage of being interpretable in terms of the information accumulation decision model.

The final two distributions implemented in QMPE, the Weibull and Gumbel, are related in that they both occur as the asymptotic distribution of the minima of samples from sets of random variables (see Weibull, 1951 and Cousineau, Goodman & Shiffrin, 2002, for details). For example, the Weibull arises from the minimum value of samples from random variables that are bounded below by zero, the Gumbel from random variables that are unbounded. Both distributions have two parameters but differ in that the Gumbel distribution is unbounded, whereas the Weibull distribution is bounded below by zero. As for the other distributions that are bounded below, QMPE adds a shift parameter, θ > 0, to the Weibull.

The Weibull distribution is a power transformation (with exponent $c > 0$) of an exponential random variable (with mean τ > 0) as is evident from its density:

$$f(x) = c\tau^{c}\left(x-\theta\right)^{c-1} e^{-\tau^{c}\left(x-\theta\right)^{c}}$$

As for the other shift distributions, the density is only defined for θ > *x*. Figure 1d shows three examples of the Weibull density. The cumulants of the Weibull are expressed in terms of the incomplete Gamma function ($\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$, $x > 0$)

$$\kappa_1 = \theta + \tau\Gamma\left(c^{-1}+1\right)$$
$$\kappa_2 = \tau^2\left[\Gamma\left(2c^{-1}+1\right) - \Gamma^2\left(c^{-1}+1\right)\right]$$
$$\kappa_3 = \tau^3\left[\Gamma\left(3c^{-1}+1\right) - 3\Gamma\left(c^{-1}+1\right)\Gamma\left(2c^{-1}+1\right) + 2\left[\Gamma\left(c^{-1}+1\right)\right]^3\right]$$

The Gumbel density is also named the Type I extreme value, the double-exponential or the Fisher-Tippett distribution. It has a location (μ) and scale (σ > 0) parameter:

$$f(x) = \frac{1}{\sigma}\exp\left[\frac{-(x-\mu)}{\sigma} - \exp\left[\frac{-(x-\mu)}{\sigma}\right]\right]$$

Figure 1e shows three examples of the Gumbel density. Its first three cumulants are:

$$\kappa_1 = \mu + 0.57722\sigma,\ \kappa_2 = (\sigma\pi)^2/6,\text{ and } \kappa_3 = 2.40412\sigma^3$$

The Fisher skew of the Gumbel is fixed (γ_1 = 1.13955) so does not have the flexibility to model changes in RT distribution skew.

Likelihood and Shifted Distributions

The likelihood of a sample given a model is the joint probability of the data assuming the model. When the observations making up the data (x_i, i = 1 … n) are sampled independently, the joint probability is given by the product of the probabilities for each observation. Maximum likelihood methods choose model parameter estimates that maximize the likelihood. However, when applied to continuous distributions, the conventional approach to maximum likelihood estimation maximizes the product of the densities for each observation, $f(x_i,\Theta)$, rather than the product of their probabilities. This procedure is justified using an approximation to the probability of each observation:

$$\Pr(x_i - h_i/2 \le X \le x_i + h_i/2) = \int_{x_i-h_i/2}^{x_i+h_i/2} f(x,\Theta)dx \approx f(x,\Theta)h_i$$

Two assumptions must hold to ensure that maximising the product of the densities ($\prod_{i=1}^{n} f(x_i,\mathbf{\Theta})$) is equivalent to maximising the joint probability. First, the h_i (> 0) must be assumed to be small for the rectangular approximation to be accurate, and the approximation can be made to be exact as the h_i tend to zero. Second, the h_i must be assumed to be independent of Θ, so that the h_i can be ignored in the maximization. We will refer to this approximation as CML. For computational reasons estimates are usually found by maximising the log-likelihood, that is the sum of the logarithms of the densities, $\sum_{i=1}^{n} \ln f(x_i, \Theta)$, which is equivalent to maximising their product.

Unfortunately the assumption underlying the CML approximation can fail in some cases, such as when the distribution's range depends on its parameters, as is the case for shift distributions. When the shift parameter (θ) equals the smallest observation, CML log-likelihood is infinite as the density for the smallest observation x_1 (without loss of generality we assume the observations are ordered, $x_1 \leq x_2 \leq \ldots \leq x_n$) is zero. In this case, CML estimates of the other parameters become inconsistent, in the sense that they do not tend to their true values as sample size increases. The singularity associated with $\theta = x_1$ is not a problem for estimation as long as the log-likelihood always decreases as θ approaches x_1 from below, as is the case, for example, for the shifted Wald distribution (Cheng & Amin, 1981).

For the shifted Lognormal and Weibull distributions, however, log-likelihood can increase as θ approaches x_1 and so inconsistent estimates can be obtained when maximising the log-likelihood by iterative methods. This problem, which we will call the "unbounded likelihood problem", occurs particularly for parameter values where the distribution is highly skewed and has a sharply increasing leading edge. For the Lognormal distribution the unbounded likelihood problem is rarely a practical concern, as the increasing region is usually very small (cf. Giesbrecht & Kempthorne, 1976, Table 1), and a local minimum exists outside this region that produces consistent parameter estimates. Given good starting point estimates, iterative methods will converge on this local minimum and provide consistent parameter estimates. This estimation method is sometimes called local maximum likelihood.

For the Weibull distribution, however, the unbounded likelihood problem can cause more severe difficulties when estimating the highly skewed distributions produced by small values of the Weibull shape parameter, c. For $c = 1$, the Weibull distribution is equivalent to the exponential distribution, which has a sharp and discontinuous leading edge. For $c < 1$ even more skewed distributions with a shape similar to the exponential are obtained. For $c > 1$ the distribution becomes less skewed and the increase of the leading edge more gradual. Symmetry occurs for $c \approx 3.6$, and skew then becomes negative as c increases, approaching a lower bound Fisher skew ≈ –1.14. When $c > 2$ for the shifted Weibull distribution, its likelihood has a consistent global minimum given $\theta < x_1$. For $1 < c < 2$, a local maximum exists which produces consistent estimates (Cheng & Amin, 1983). Even in this range, however, CML likelihood can become unbounded due to sampling error. Similar conditions apply to the shifted Gamma, another distribution commonly used to model RT, for exactly the same values of its shape parameter (Cheng & Amin, 1983).

Cheng and Amin (1983), and independently Ranneby (1984), suggested a solution to the unbounded likelihood problem, called the maximum product of spacings (MPS) method. MPS, like CML, obtains parameter estimates by maximising an objective function. The MPS objective function is exactly proportional to a special case of the QMP objective function, and so produces identical estimates. As defined by Heathcote et al. (2002), QMP estimates are obtained by maximising the multinomial log-likelihood:

$$\sum_{j=1}^{m} N_j \ln\left(D_j\right), \text{ where } D_j = \int_{\hat{q}_{j-1}}^{\hat{q}_j} f\left(x_i, \Theta\right) dt \qquad (1)$$

The $\hat{q}_j$, j = 1 … m-1 are quantile estimates, ($\hat{q}_o, \hat{q}_m$) equals the domain of the distribution (which might depend on Θ), and each inter-quantile range ($\hat{q}_{j-1}, \hat{q}_j$) contains N_j observations (in general N_j may not be an integer).

The MPS estimator is a special case of QMP[4] where order statistics (i.e., x_i) are used to estimate quantiles, $N_j = 1$ and $m = n$. Titterington (1985) suggested a modified version of the MPS objective function that is exactly proportional to the QMP1 objective function examined by Heathcote et al. (2002). By QMP1, we mean estimates obtained by maximising (1) and based on $\hat{q}_j = (x_j + x_{j+1})/2$, $j = 1 \ldots n-1$, and $N_j = 1$. Heathcote et al. showed that QMP1 produced more efficient and less biased estimates than CML for the ex-Gaussian distribution. They also examined QMP4 estimates, where the data set is reduced to a set of $n/4$ equally spaced quantiles (for $n = 4m$, $\hat{q}_j = (x_{4j} + x_{4j+1})/2$ and $N_j = 4$), and found similar estimation performance to CML, despite the fact that QMP4 is clearly not a sufficient estimator (i.e., the values of some observations can varied without affecting its value).

Although Titterington (1985) suggested that MPS can be viewed as maximum likelihood estimation for grouped data, it is important to acknowledge that the equivalence is only approximate in finite samples, because (1) does not take into account the error associated with quantiles estimates. However, the approximation is asymptotically exact in some cases, as shown by Cheng and Amin (1983) for their MPS method. Where the range of the distribution is not parameter dependent MPS and CML are asymptotically equal. Hence, MPS has all of the asymptotic sufficiency, consistency and efficiency properties of CML. Where the range of the distribution is parameter dependent CML and MPS can behave quite differently. Importantly, both the original version of MPS and Titterington's variation (i.e., QMP1) differ from

CML in that they are not subject to the unbounded likelihood problem (Cheng & Iles, 1987). Hence, they continue to give consistent and efficient estimates even when CML completely fails.

In summary, it is clear that unbounded likelihood problem can cause CML to completely fail in cases where QMP continues to work well. It might be argued that such cases are of little interest for RT distribution fitting, as RT distributions rarely have a sufficiently sharp leading edge or degree of skew. We are not aware of any systematic investigation on this point, and caution that sampling error may cause the problem to occur in small samples even if the true distribution comes from a parameter region where CML does not fail. In any case, there is little point comparing the estimation performance of CML and QMP in such cases, as QMP will necessarily be superior. Hence, parameters for the Monte Carlo study were chosen to avoid the "J-shaped" distributions associated with CML failure.

Monte Carlo Study

The Monte Carlo study was modelled after the study reported by Heathcote et al. (2002). It had three aims: 1) to extensively test the QMPE code, 2) to compare the estimation performance of CML and QMP, and 3) to compare estimation performance among the five distributions fit by QMPE. Relatively small sample sizes were used (n = 40, 80 or 160) in order to investigate performance under demanding and realistic conditions. QMP estimation was performed both using QMP1 and QMP4.

For each distribution three sets of parameter values, given in Table 1, were used, (Figure 2 illustrates the corresponding densities). The parameters for the ex-Gaussian distribution were the three sets with medium levels of skew used by Heathcote et al. (2002). The choice of parameters for the other distributions was guided by fitting them to large samples from the three ex-Gaussian distributions, so that results are approximately comparable across distributions. As Table 1 shows, this procedure resulted in a fairly good match on means and standard deviations. Fisher skew varied more between distribution types but covered approximately the same range, except for the Weibull where skew was generally lower, and the Gumbel distribution, where skew is fixed. The smallest value of the Weibull shape parameter investigated (c = 1.5) was large enough to avoid the sort of problematic behaviour for CML estimation illustrated in Figure 1b.

Insert Table 1 here

Examination of Figure 2 indicates that shift estimation in the least skewed cases (labelled 1 in the figure) of the Lognormal and Wald distributions will be challenging, because they have long thin left tails. Such tails make estimation of the shift parameters difficult because samples near the lower bound are rare, and the sampled values of the first order statistic (x_1) are highly variable. As a result, shift estimates are likely to be biased upward and to be more variable for these cases in the Monte Carlo study, particularly in small samples.

Methods

For each distribution type and the nine combinations of sample size and parameter set, 10000 replicates were fit, with the same samples fit by CML, QMP1 and QMP4. The simulated samples were obtained using random number generators provided by the S-plus statistical package[5], and were rounded to the nearest integer.

QMPE uses the same numerical methods as QMLE (see Brown & Heathcote, in press, for more details). CML and QMP estimates are obtained by a conjugate gradient optimisation algorithm. This algorithm requires analytic expressions for the gradient of the objective function. However, QMPE requires only analytic gradients for the density; gradients for CML and QMP are automatically computed from the density gradients. Once search is complete, analytic expressions for the Hessian (second derivative matrix) of the density are used to estimate approximate parameter standard errors and correlations (see Brown & Heathcote, in press, for a proof that these estimates are asymptotically correct for QMP). Although derivative free optimisation methods are available, we have found that analytic gradients greatly speed estimation and that analytic Hessians result in better standard error and correlation estimates.

QMPE automatically obtains starting points for optimisation by substituting method of moments' estimates of cumulants into the equations relating cumulants and parameters. However, this approach fails when sample estimates of skew are negative. In such cases heuristics are used to estimate starting points. For the three distributions with a shift parameter, the heuristic estimates the shift as slightly smaller than the minimum value in the sample (e.g., $\hat{\theta} = p \times x_1$, where p is an appropriately chosen constant), then solves for the other parameters using the first two moments calculated on $\mathbf{x} - \hat{\theta}$. The heuristic is always used for the Lognormal, which we found rarely works with the full method of moments approach. As it has only two parameters, Gumbel start points are obtained from only the first two cumulants.

Good automatic starting point estimates are essential when large numbers of conditions must be fit, and particularly for the shifted distributions when only the local CML solution is useful. QMPE's start point heuristics were fine tuned throughout the course of the Monte Carlo study. We have also found them to work well in real RT data.

The stopping criteria for optimisation were set at a proportional objective function exit tolerance of 10^{-9}, a proportional parameter change tolerance of 10^{-5}, and the maximum number of search iterations was fixed at 250 (see Brown & Heathcote, in press, for details on these settings), resulting in parameter estimates accurate to more than four significant figures. For all parameters bounded below by zero, QMPE sets the objective function to a low value when the estimate is less than 10^{-9}, which ensures both that the bound is respected and that numerical errors do not occur. For distributions with shift parameters, these parameters were restricted to less than the sample minimum for CML fits and less than the minimum quantile for QMP fits.

Results

Overall, 99.98% of fits produced usable parameter estimates, the properties of which are examined below. As shown in Table 2, estimation of parameter standard errors and correlations failed more often, due to non-invertible Hessian estimates. Table 2 averages over parameter sets and sample sizes for brevity. Generally, better performance was obtained with larger samples and for more skewed distributions. In most cases CML estimation resulted in invertible Hessian estimates, although QMP1 performed almost as well for all except the Wald distribution. These results indicate that grouping should be minimised in QMP estimation if parameter standard error and correlation estimates are required. For the Wald distribution, only CML produced adequate performance.

Insert Table 2 here

<u>**Bias and Efficiency**</u>

Bias was estimated as the difference between the mean of the Monte Carlo parameter estimates and the true value. Efficiency was estimated by the standard deviation (SD) of the parameter estimates. Bias and efficiency estimates are described as "consistent" if bias decreases and efficiency increases as sample size increases. Figures 2-4 show bias and efficiency estimates for the shift distributions parameters. Each panel is divided into three areas for the three sets of parameters, and within each area sample size increases from left to right as indicated on the ordinate. Results for CML, QMP1 and QMP4 are presented side by side in groups of three lines, so they can be easily compared.

In order to compare estimation performance across distributions, it is useful to recognise that the shift parameters (θ), and the Weibull scale parameter (τ), have the same units as the data. For these parameters relative estimation performance can be judged on the same scale. Estimation performance can also be judged for all parameters as a proportion of their true values, which are given in Table 1 and reiterated in each caption of the figures displaying results.

The results for the ex-Gaussian distribution replicated Heathcote et al. (2002) with only a slightly different methodology (i.e., rounded samples from a different random number generator), and are omitted for brevity, as is a detailed discussion of the results for the Gumbel distribution, which were uniformly good in all cases and for all estimation methods[6]. In contrast to the Gumbel and ex-Gaussian estimates, estimates for the shift distributions were very poorly behaved in some cases. By "poorly behaved" we mean that estimates were very biased, not always consistent (i.e., bias could increase and efficiency decrease with sample size) and the parameter estimate distributions were not even approximately normal.

The Lognormal estimates were best behaved amongst the shift distributions in terms of consistency and distributions. Parameter estimate distributions were uni-modal with the exception of CML estimates for the least skewed distribution and n =

160, which had second modes overestimating θ and σ and underestimating μ. The uni-modal parameter estimate distributions were slightly skewed, with longer tails for smaller values of θ and μ and larger values of σ. As shown in Figure 2, bias was in the opposite direction to skew, being upward for θ and μ and downward for σ.

Insert Figure 2 here

Figure 2 shows that overestimation of the shift parameter was particularly large for the least skewed Lognormal distribution, which had a small shift parameter (475) and a long left tail. As is evident from Figure 1, samples less than 750 are rare, resulting in a strong upward bias in shift estimates even for larger sample sizes. All parameter estimates were much better behaved for the more skewed distributions, with relatively low bias even for the smallest samples. Efficiency increased with skew for θ but decreased with skew for μ. The σ estimates had similar bias for all three Lognormal distributions.

Overall, QMP estimates were less biased but also less efficient than CML estimates for the Lognormal distribution. Reduced bias was most marked for QMP1, but also occurred for QMP4, with the smallest samples for the two most skewed distributions being exceptions. CML and QMP1 estimates had similar efficiency, with QMP4 estimates being clearly less efficient in most cases, particularly for n = 40.

As shown in Figure 3, the shift parameter of the least skewed Wald distribution was also overestimated, although overestimation was not as pronounced as for the least skewed Lognormal distribution. As for the Lognormal distribution, this can be attributed to a long left tail; Figure 1 shows that samples less than 750 are rare, even though the shift parameter is 625. For μ and *a* there was an underestimation bias for the least skewed distribution. The bias was most pronounced for QMP4 and n = 40, where it was almost the same magnitude as the true parameter values. A large degree of QMP4 underestimation bias for μ and *a* persisted for small samples from the more skewed Wald distributions, but disappeared for larger samples. For the two least skewed distributions and n = 40, QMP4 estimates of μ and *a* were distributed bi-modally, with a mode near zero in both cases. The corresponding shift estimates were uni-modal but some large underestimates did occur.

Insert Figure 3 here

For the more skewed Wald distributions, CML and QMP1 estimation bias was more acceptable for the μ and *a* parameters, even at the smallest sample size. However, relatively large biases in the shift parameter persisted for the more skewed distributions, with the medium skew case producing overestimates and the most skewed case producing underestimates. For all methods, the most skewed distribution produced inconsistent bias estimates that increased with sample size, particularly for the shift parameter. Overall, CML estimates were the least biased and most efficient among estimation methods. Efficiency was consistent for CML, but often inconsistent for the QMP estimates.

For the Weibull, parameter estimate distributions were almost always bimodal to some degree, particularly for the least skewed case, although the second mode tended to disappear as sample size increased. The second mode always underestimated shift (θ) and overestimated the scale (τ) and shape (c) parameters. The problem was particularly marked for CML shape estimates from the least skewed distribution, occurring even for the largest sample size, with outlying estimates many orders of magnitude greater than the true value. Large *c* estimates occurred mainly for samples with negative Fisher skew. The Weibull distribution can have negative Fisher skew, which slowly approaches a bound of approximately –1.14 for large values of *c*. For example, Fisher skew values are -0.08, -0.6, -1, -1.1 and –1.13 for *c* = 4, 10, 100, 1000 and 10000 respectively.

Insert Figure 4 & 5 here

In order to obtain more acceptable behaviour we bounded estimates of *c* above by 10 and repeated the fits for all methods. Although this solution is ad hoc, and the value of the upper bound rather arbitrary, it did improve estimation performance, particularly for CML. For the refit results bias and efficiency were consistent for all cases, except QMP4 bias for the least skewed distribution. For the least skewed Weibull distribution, bias followed the second mode, being downward for θ and upward for τ and *c*, as illustrated in Figure 5 for CML estimates and n = 40. As shown in Figures 4 and 5, underestimation of the shift parameter and overestimation of the τ and *c* parameters was substantial for the least skewed Weibull distribution. CML estimates were less biased than QMP estimates, and for n = 160 CML bias was quite acceptable, whereas QMP bias remained quite large, particularly for QMP4.

For all methods bias was substantially less for the more skewed Weibull distributions. For these cases, QMP estimates, particularly QMP1 estimates, were less biased than CML estimates, although QMP4 bias was larger than CML bias for n = 40 in the medium skew distribution. For all distributions CML estimates were the most efficient, although the advantage over QMP1 was relatively small for the more skewed distributions, particularly at larger sample sizes. QMP4 was clearly the least efficient method, particularly for smaller sample sizes.

Figure 5 also shows that estimates in the deviant mode usually also have ill conditioned Hessians, indicating that the neighbourhood of the solution is not locally quadratic. When estimates with ill-conditioned Hessians were censored, bias was virtually eliminated for CML and greatly reduced for QMP, so that for all estimation methods and the largest sample size it was less than 10 for θ and τ and less than 0.1 for *c*, even for the least skewed distribution. Hence, it appears that censoring estimates with ill conditioned Hessians can greatly improve overall estimation performance. We consider this issue in more detail in the discussion.

Estimating Fisher Skew

Ratcliff (1978) used CML to fit the ex-Gaussian distribution in order to estimate RT distribution skew. He pointed out that that estimates of skew based on the method of moments are both inefficient and non-robust. Hence, unrealistically large sample sizes are required for precise estimates and estimates can be greatly distorted by even small levels of outlier contamination. In this section we compare the indirect method of calculating Fisher skew from CML and QMP1 parameter estimates with direct estimates obtained from the method of moments. Skew estimates for all three-parameter distributions (i.e., those with variable skew) are shown in Figure 6. Results for the method of moments, CML and QMP1 are presented side by side in groups of three lines, so they can be easily compared.

Insert Figure 6 here

For the ex-Gaussian distribution both CML and QMP1 estimates were less biased and more efficient than method of moments estimates, with QMP1 generally having the greatest advantage. For the Lognormal, the method of moments estimates were less biased than CML and QMP and for the two least skewed distributions but more biased for the most skewed distribution. QMP was less biased than CML in all cases, with efficiency being similar. Generally, the methods of moments estimates are less efficient than CML and QMP.

For the Weibull distribution the method of moments fared even better, with less bias than either CML or QMP in all cases. Efficiency was also generally better for method of moments estimates, except for more skewed distribution and larger sample sizes. Generally, QMP1 estimates were less biased but also slightly less efficient than CML estimates. Finally, for the Wald distribution the method of moments estimates were generally the least biased but also the least efficient. CML tends to be a little more biased than QMP1 but were consistently more efficient.

The results presented in Figure 6 provide a basis for comparing estimation performance among distributions with variable skew. Clearly CML and QMP fits of the ex-Gaussian distribution provide better skew estimates than the method of moments. However, for the Wald and Weibull, and to a lesser degree the Lognormal, CML and QMP estimates actually provide poorer skew estimates than the method of moments in most cases. In the next section we discuss possible causes of this poor estimation performance, particularly for the Weibull distribution.

Discussion

The results of the Monte Carlo study confirmed that QMP is generally superior to CML for the two distributions with an unbounded range, the ex-Gaussian and Gumbel. CML and QMP are on a more equal footing for the shift distributions when

Fisher skew is greater than one, although it should be remembered that for more extreme skew CML can fail entirely for the Lognormal and Weibull distributions, due to the unbounded likelihood problem.

No method worked very well for the least skewed Lognormal, Wald and Weibull distributions, particularly for smaller sample sizes. Overestimation of shift for the Lognormal and Wald distributions results from their long thin left tails in the least skewed cases. Overestimation in these cases may be difficult to avoid because information about the shift value is very variable. Hence, QMPE parameter estimates for these distributions should be interpreted with caution when Fisher skew is less than one.

Underestimation of shift for the Weibull appears to be related to a non-quadratic maximum for both CML and QMP, as indicated by ill-conditioned Hessian estimates. Removal of cases with ill-conditioned Hessians greatly improved performance. Heathcote (in press) noted similar behaviour for CML estimates of Wald distributions in small samples (n = 40), and suggested that estimates with ill-conditioned Hessians should be censored when using his software. However, censoring QMPE Wald estimates with ill-conditioned Hessians did not reduce bias appreciably. In contrast, when Heathcote's optimisation methods (the Splus *nlminb* algorithm using analytic first and second derivatives) were applied to the Wald data from the present Monte Carlo study, bias was almost eliminated by censoring. Unfortunately reduced bias was bought at the cost of reduced efficiency relative to the QMPE estimates.

Cheng and Iles (1990) provide a possible explanation for these difficulties, called the "embedded models" problem. They showed that for each of the shift distributions fit by QMPE has a special case as shift approaches $-\infty$, which they call an "embedded" distribution. The embedded distributions (normal for all but the

Weibull, which has an embedded Gumbel distribution) have only two parameters, a scale parameter and a location parameter. When the embedded model fits as well as the shift model it indicates that the sample does not contain sufficient information to estimate shift or shape; instead only location and scale can be reliably estimated. Because the embedded model occurs at an infinite parameter value, iterative estimation of the shift model is difficult, and estimates for all of its parameter become unreliable.

The bimodality and underestimation of shift seen in the QMPE Weibull parameter estimates appears to be due to the embedded model problem. Similar problems found here for some QMP4 estimates produced by QMPE and by Heathcote (in press) for CML estimates of the shifted Wald appear to be examples of the embedded model problem. Heathcote's Wald estimates were probably more prone to this problem than QMPE Wald estimates because his fitting routine used analytic Hessians, and so was more sensitive to the non-quadratic maximum produced by the embedded model problem.

General Discussion

In this paper we have described and tested QMPE, an open source ANSI standard Fortran 90 program, which can estimate the parameters of five continuous density functions commonly used to model RT. QMPE can fit these distributions using continuous maximum likelihood (CML), perhaps the most widely used and recommended method of estimating RT distribution (Heathcote, 1996; Van Zandt, 2000). It can also fit these distributions using Heathcote et al.'s (2002) quantile maximum probability (QMP) method.

A Monte Carlo study replicated Heathcote et al.'s (2002) finding that QMP produces less biased and more efficient parameter estimates than CML for the ex-

Gaussian distribution. QMP was also found to produce less biased, but also slightly less efficient, parameter estimates than CML for the Gumbel distribution. Overall estimation performance was excellent for both distributions, as might be expected in the idealized situation represented by the Monte Carlo study; fitting the true data generating model to uncontaminated data. Of course, real RT data does not conform to this ideal, but at least the excellent performance under ideal conditions is reassuring for the practical application of QMPE.

The ex-Gaussian and Gumbel distributions have an unbounded range. This might be seen as a disadvantage when they are used to model RT data, because RT data must be bounded below by a positive value. QMPE also fits three “shift” distributions, which have a positive, parameter dependent lower bound, and so seem more promising as models of RT. Unfortunately, estimation performance for the shift distributions was much worse than for the ex-Guassian or Gumbel. One reason for fitting parametric distributions to RT data is to obtain a more reliable estimate of skew than is provided by the methods of moments (Ratcliff, 1978). If QMPE is used for this purpose we suggest that the ex-Gaussian be fit (the Gumbel has fixed skew and so is not useful for this purpose). At least when the ex-Gaussian is an accurate approximation to the data, this approach is more efficient and less biased than the method of moments. The shift distributions, in contrast, did not consistently outperform the method of moments, and in some cases produced substantially worse skew estimates.

One reason why the ex-Gaussian outperforms the shift distributions might be the parameterisations used by QMPE, and most packages aimed at fitting RT distribution[7]. The ex-Gaussian has as special cases its least skewed (Gaussian) form when its exponential parameter approaches zero. The shift distributions have their least skewed form as a special case when the shift parameter approaches $-\infty$. In small samples the least skewed case may often be most appropriate because the data contains mainly information about location and scale. Because this case occurs as the shift parameter diverges, parameter estimates for the shift distributions become unreliable, whereas this does not happen for the ex-Gaussian. Cheng and Iles (1990) suggested solving this problem, which dubbed the “embedded model problem” by, using a parameterisation of the shift distributions where the least skewed case occurs at zero rather than infinity (see their Table 1). We have implemented and are testing Cheng and Iles parameterization in a new version of the QMPE software. We suggest that users of the existing version of QMPE, and similar software, exercise caution in interpreting parameter estimates from samples that have ill-conditioned Hessians or negative skew.

The embedded model problem occurs for data with low skew. Highly skewed data can also cause a problem, called the “unbounded likelihood problem”, for CML estimation of the shifted Weibull and Lognormal distributions. The problem occurs because the likelihood maximum occurs when the shift parameter equals the minimum observed value. Although this might seem like a plausible estimate of shift, estimates of the remaining parameters are inconsistent. QMP does not suffer from the unbounded likelihood problem and so remains useful for highly skewed data. Hence, when skew is high, we recommend QMP fitting over CML fitting. QMP might also be useful in other contexts, such as fitting mixtures (e.g., Dolan et al., 2002), as they can also be subject to the unbounded likelihood problem (Cheng & Traylor, 1995).

In the course of this investigation we discovered that QMP has a special case called the maximum of product spacings (MPS), which was advocated by Cheng and Amin (1983) and Ranneby (1984) as a means of overcoming the unbounded likelihood problem. Their work proves that MPS has all of the desirable asymptotic properties of CML when CML estimates exist, and continues to work well when CML fails. Further consideration of the MPS and its generalizations (e.g., Ekstrom, 2001) is beyond the scope of the present work. However, this literature places QMP on a firm theoretical footing, not just as an approximation to likelihood, also as a goodness of fit measure that can be derived from information theory (see also Speckman & Rouder, in press; Heathcote & Brown, in press).

References

Brown, S., & Heathcote, A. (in press). QMLE: Fast, robust and efficient estimation of distribution functions based on quantiles. Behavior Research Methods, Instruments and Computers.

Breukelen, G. J. P. (1995). Parallel processing models compatible with Lognormally distributed processing times. Journal of Mathematical Psychology, 39, 396-399.

Cheng, R. C. H. & Amin, N. A. K. (1981). Maximum likelihood estimation of parameters in the Inverse Gaussian distribution, with unknown origin. Technometrics, 23, 257-263.

Cheng, R. C. H. & Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. Journal of the Royal Statistical Society, Series B, 45, 394-403.

Cheng, R. C. H. & Iles, T. C. (1987). Corrected maximum likelihood in non-regular problems. Journal of the Royal Statistical Society, Series B, 49, 95-101.

Cheng, R. C. H. & Iles, T. C. (1990). Embedded models in three-parameter distribution estimation. Journal of the Royal Statistical Society, Series B, 52, 135-149.

Cheng, R. C. H. & Traylor, L. (1995). Non-regular maximum likelihood problems. Journal of the Royal Statistical Society, Series B, 57, 3-44.

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

Cousineau, D., Goodman, V. & Shiffrin, R. M. (2002). Extending statistics of extremes to distributions varying on position and scale, and implication for race models. Journal of Mathematical Psychology, 46, 431-454.

Cousineau, D. & Larochelle, S. (1997). PASTIS: A Program for Curve and Distribution Analyses. Behavior Research Methods, Instruments, & Computers, 29: 542-548.

Dolan, C. V., van der Maas, H. L. J. & Molenaar, P. C. M. (2002). A framework for ML estimation of parameters (mixtures of) common reaction time distributions given optional truncation or censoring. Behavior Research Methods, Instruments, & Computers.

Ekstrom, M. (2001). Consistency of generalized maximum spacing estimates. Scandinavian Journal of Statistics, 28, 343-354.

Giesbrecht, F. & Kempthorne, O. (1976). Maximum likelihood estimation in the three-parameter Lognormal distribution, Journal of the Royal Statistical Society, Series B, 38, 257-264.

Gumbel, E. J. (1958). The Statistics of Extremes. New York: Columbia University Press.

Heathcote, A. (in press). Fitting the Wald and Ex-Wald distributions to response time data. Behavioural Research Methods, Instruments, & Computers

Heathcote, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. Behavioural Research Methods, Instruments, & Computers, 28, 427-445.

Heathcote, A., & Brown, S. (in press). Reply to Speckman and Rouder: A theoretical basis for QML. Psychonomic Bulletin and Review.

Heathcote, A., Brown, S. & Mewhort, D.J.K. (2002). Quantile Maximum Likelihood Estimation of Response Time Distributions. Psychonomic Bulletin and Review, 9, 394-401.

Hyndman, R.J. & Fan, Y. (1996). Sample quantiles in statistical packages. The American Statistician, 50(4) 361-365.

McClelland, J. L. (1979). On the time relations of mental processes: A framework for analyzing processes in cascade. Psychological Review, 86, 287-330.

Ranneby, B. (1984). The maximum spacing method: an estimation method related to the maximum likelihood method. Scandinavian Journal of Statistics, 11, 93-112.

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

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

Speckman, P. L. & Rouder, J. N. (in press). A comment on Heathcote, Brown and Mewhort's QMLE estimation method for response time distributions. Psychonomic Bulletin & Review.

Titterington, D. M. (1985). Comment on "Estimating parameters in continuous univariate distributions", Journal of the Royal Statistical Society, Series B, 47, 115-116.

Wald, A. (1947). Sequential analysis. New York: John Wiley and sons.

Weibull, W. (1951). A statistical distribution function of wide applicability. Journal of Applied Mechanic, 18, 292-297.

West, J. & Shlesinger, M. (1990). The noise in natural phenomena. American Scientist, 78, 40-45.

Woodworth, R. S., & Schlosberg, H. (1954). Experimental Psychology, Holt, New York.

Van Zandt, T. (2000). How to fit a response time distribution. Psychonomic Bulletin and Review, 7, 424-465.

Footnotes

[1]The software, example data and a manual are available at http://oz.ss.uci.edu/ and in the Software Repository at http://www.newcastle.edu.au/school/behav-sci/ncl/.

[2]Heathcote et al. 2000 described their method as "quantile maximum likelihood estimation" (QMLE). However, Speckman & Rouder (in press) pointed out that that QMLE is not maximum likelihood. Acknowledging this point, Heathcote and Brown (in press) renamed the method "quantile maximum probability estimation" (QMPE).

[3]This parameterisation is convenient when there is no shift parameter as simple analytic maximum likelihood estimates are available for both the mean (just the usual arithmetic sample mean) and $1/\hat{\lambda} = \sum_{i=1}^{n}\left(x_i^{-1} - \hat{\kappa}_1^{-1}\right)/n$. These estimates can also be used to reduce computation when shift must be estimated, using the method of profile likelihood where only the shift parameter is iteratively optimised with the other parameters estimated analytically.

[4] We reserve the term QMP for the general procedure that maximizes (1) based on a set of quantile estimates obtained in an unspecified manner. Many different quantile estimators are available (see Hyndman & Fan, 1996, for a review). Although these estimators are asymptotically equivalent they differ in finite samples and so can result in differing parameter estimates in practice. In the QMPE software we implemented the quantile estimator specified in Heathcote et al. (2002), which produces QMP1 and QMP4 estimates as special cases and corresponds in general to Hyndman and Fan's definition 5. Although we have found this estimator work well, QMPE can be used with other quantile estimators by providing the estimated quantile values rather than raw data as input to the program.

[5]By default S-plus has functions to generate samples from all distributions fit by QMPE except the Wald. Wald random deviates were obtained using the S-plus function available at http://www.statsci.org/s/invgauss.s. This function uses the Inverse Gaussian parameterisation. Heathcote (in press) provides an S-plus Wald random number generator using the diffusion parameterisation.

[6] All estimation methods were consistent for the Gumbel, with negligible bias and good efficiency (SD < 15) for all sample sizes, estimation methods and parameters, with a slight improvement for the less variable distributions. QMP was less biased than CML, but in contrast to findings with the ex-Gaussian

distribution, QMP4 was less biased than QMP1. CML and QMP1 were almost equal in efficiency with QMP4 being slightly less efficient in some cases. The results indicate that all three methods should be useful in practice with samples as small as 40 observations and perhaps less.

[7]PASTIS (Cousineau & Larochelle, 1997) and DISTFIT (Dolan, van der Maas & Molenaar, 2002) use the same parameterisation as QMPE for the ex-Gaussian and Lognormal. For the Weibull, the other packages differ only in that they use a parameter equivalent to the inverse of the τ parameter used by QMPE. For the Wald, both DISTFIT and PASTIS use the "inverse gaussian" parameterisation (see Footnote 3). PASTIS and DISTFIT use the same parameterisation as QMPE for the Gumbel.

Tables

Table 1. Exact parameters of distributions used in the Monte Carlo study (except Wald, where exact parameters are given by (mean,λ) = (375,5000), (275,2000) and (200,800)), and associated moment statistics.

	Set	μ	σ	τ	Mean	SD	γ_1
Ex-Gaussian	1	929.289	70.711	70.711	1000	100.000	0.7071
	2	910.557	44.721	89.443	1000	100.000	1.4311
	3	905.132	31.623	94.868	1000	100.000	1.7076
		θ	σ	μ	Mean	SD	γ_1
Lognormal	1	470	0.18	6.25	996.47	95.538	0.5504
	2	745	0.36	5.45	993.34	92.379	1.1674
	3	800	0.48	5.15	993.49	98.488	1.6590
		θ	μ	a	Mean	SD	γ_1
Wald	1	625	0.1886	70.711	1000	102.698	0.8216
	2	725	0.1626	44.721	1000	101.973	1.1124
	3	800	0.1414	28.284	1000	100.000	1.5000
		θ	τ	c	Mean	SD	γ_1
Weibull	1	700	315	3.2	982.13	96.772	0.1064
	2	800	220	2.0	994.97	101.915	0.6311
	3	840	170	1.5	993.47	104.200	1.0720
		μ	σ	-	Mean	SD	γ_1
Gumbel	1	955	85	-	1004.06	109.017	1.1396
	2	955	74	-	997.71	94.909	1.1396
	3	955	68	-	994.25	87.213	1.1396

Table 2. Percentages of fits with non-invertible Hessian estimates.

	Ex-Gaussian	Lognormal	Wald	Weibull	Gumbel
CML	1.6	0.1	6.0	2.4	44.4
QMP1	1.2	1.7	43.0	5.3	50.9
QMP4	4.0	31.7	47.5	6.1	62.0

Figure Captions

Figure 1: Distributions used in the Monte Carlo study.

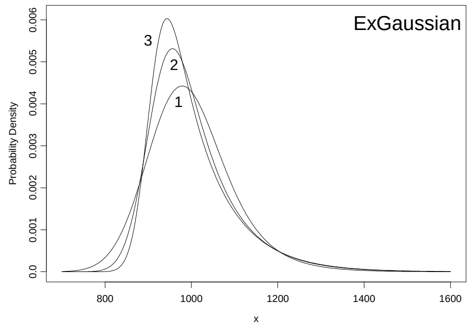
Figure 2. Lognormal mean bias (left column) and standard deviation (SD, right column) for parameters θ (first row, 470, 745 and 800 for panel sections from left to right), μ (second row, .18, .36, .45) and σ (third row, 6.25, 5.45, 5.15). Clusters of three bars represent (from left to right) CML, QMP1 and QMP4 estimates.

Figure 3. Wald mean bias (left column) and standard deviation (SD, right column) for parameters θ (first row, 625, 725 and 800 for panel sections from left to right), μ (second row, .189, .163, .141) and a (third row, 70.7, 44.7, 28.3). Clusters of three bars represent (from left to right) CML, QMP1 and QMP4 estimates.

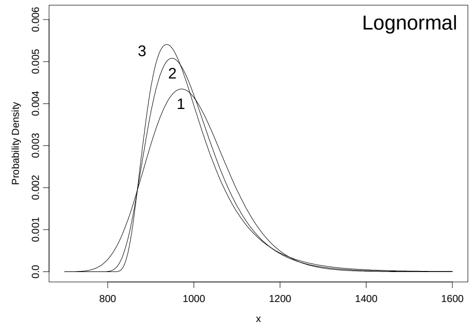
Figure 4. Weibull mean bias (left column) and standard deviation (SD, right column) for parameters θ (first row, 700, 800 and 840 for panel sections from left to right), τ (second row, 315, 220, 170) and σ (third row, 3.2, 2, 1.5). Clusters of three bars represent (from left to right) CML, QMP1 and QMP4 estimates.

Figure 5. CML parameter estimates of the least skewed Weibull distribution and n = 40. A dotted vertical line indicates the true parameter value. Filled bars indicate cases where the Hessian was singular or produced negative parameter variance estimates.

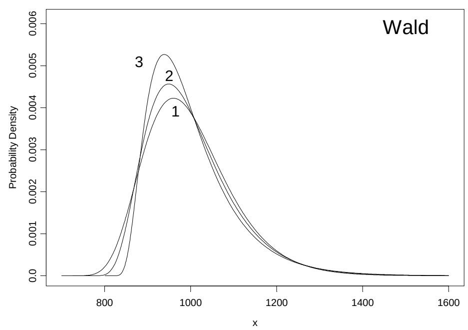
Figure 6. Mean bias (left column) and standard deviation (SD, right column) for Fisher skew estimates for the Ex-Gaussian (first row), Lognormal (second row), Weibull (third row) and Wald (fourth row) data. Clusters of three bars represent (from left to right) estimates based on the method of moments, CML and QMP1.



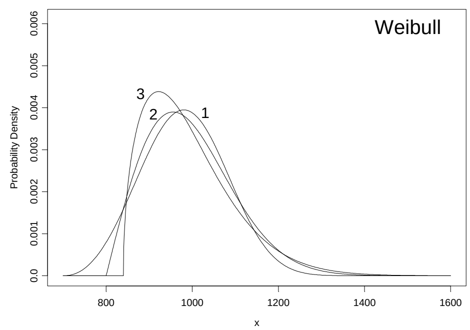
(a)



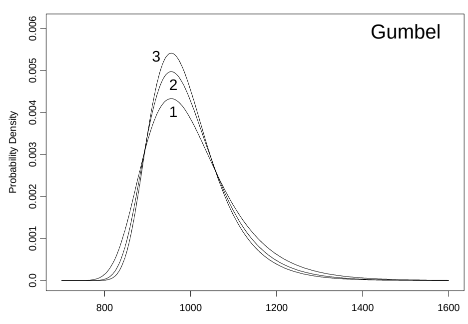
(b)



(c)



(d)



(e)

Figure 1

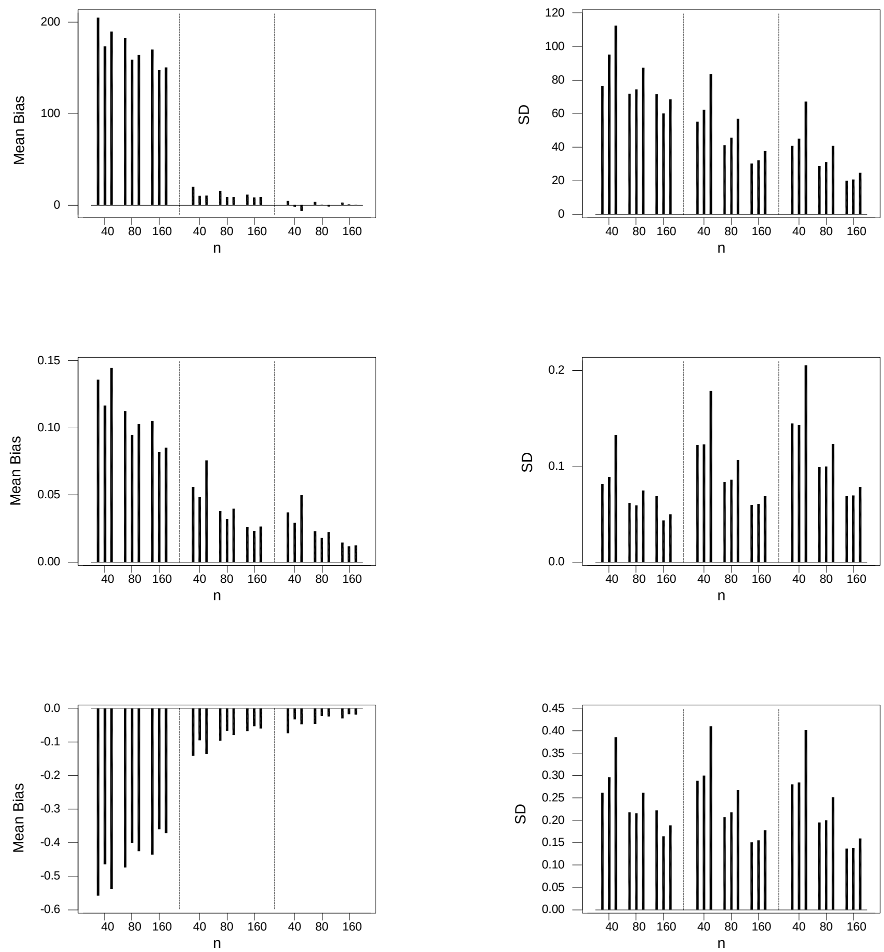


Figure 2.

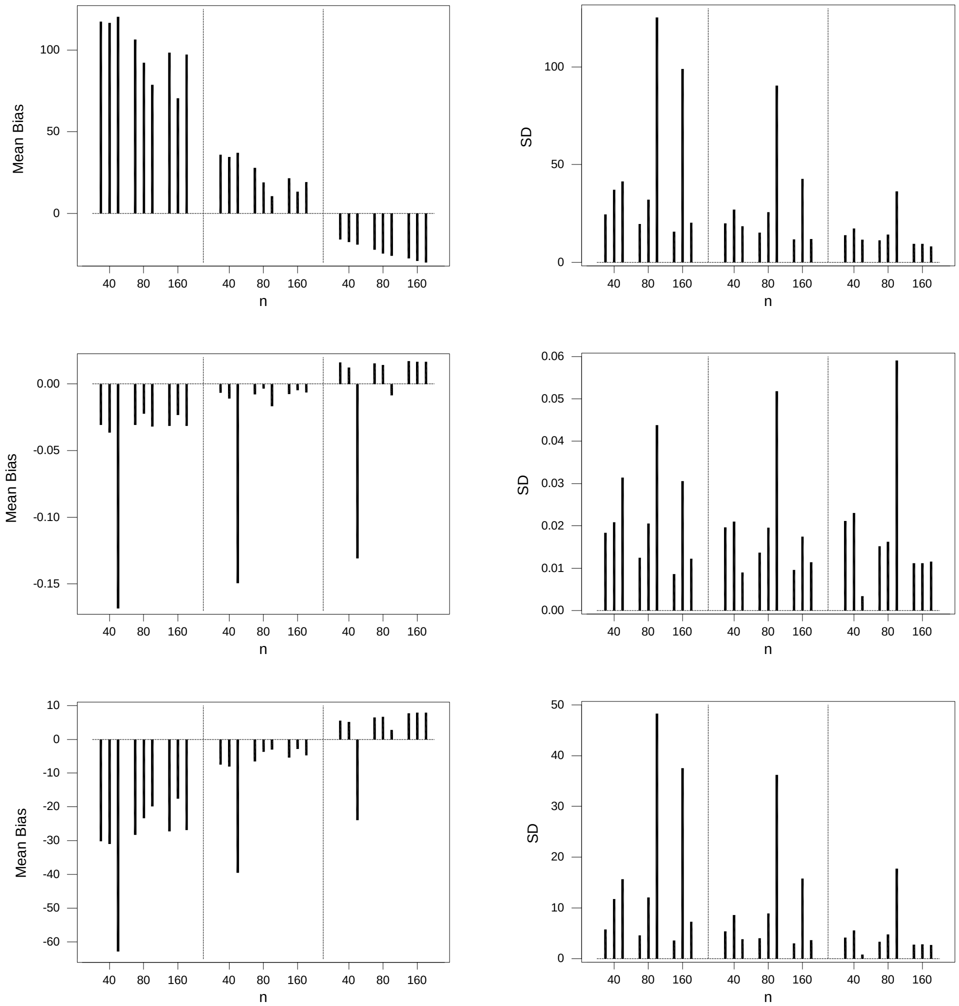


Figure 3.

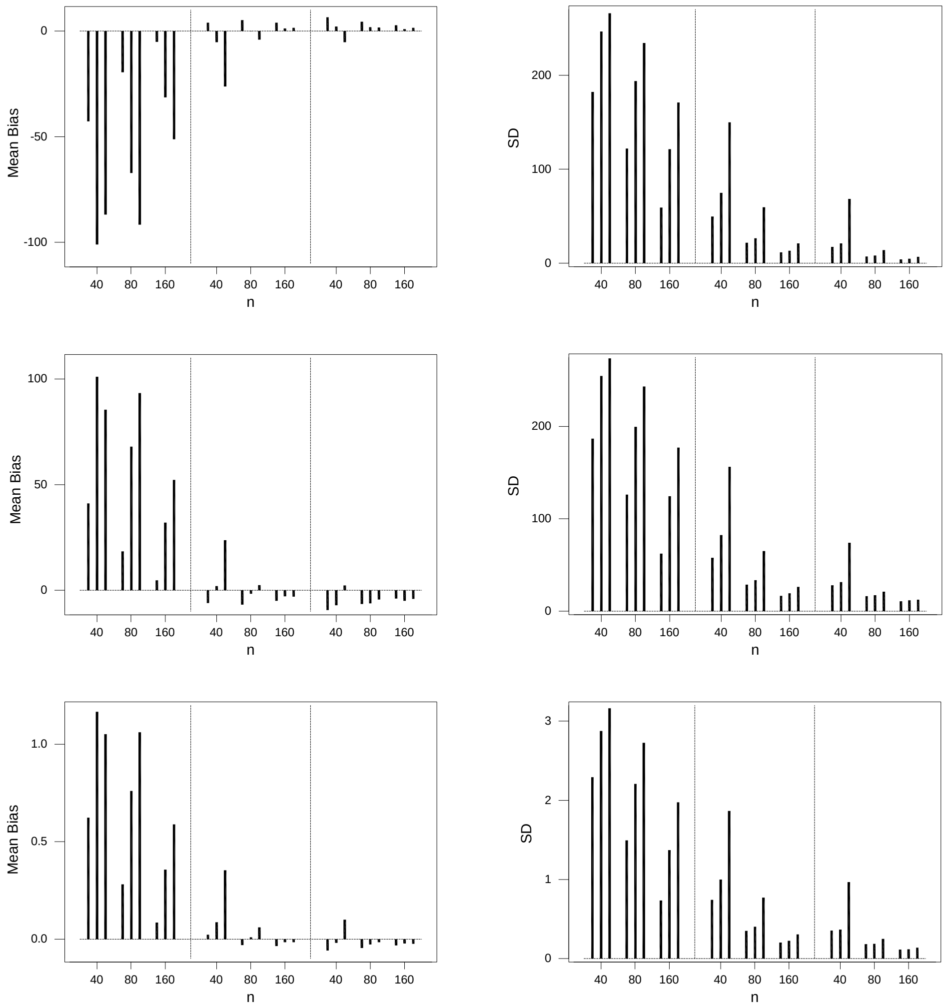


Figure 4.

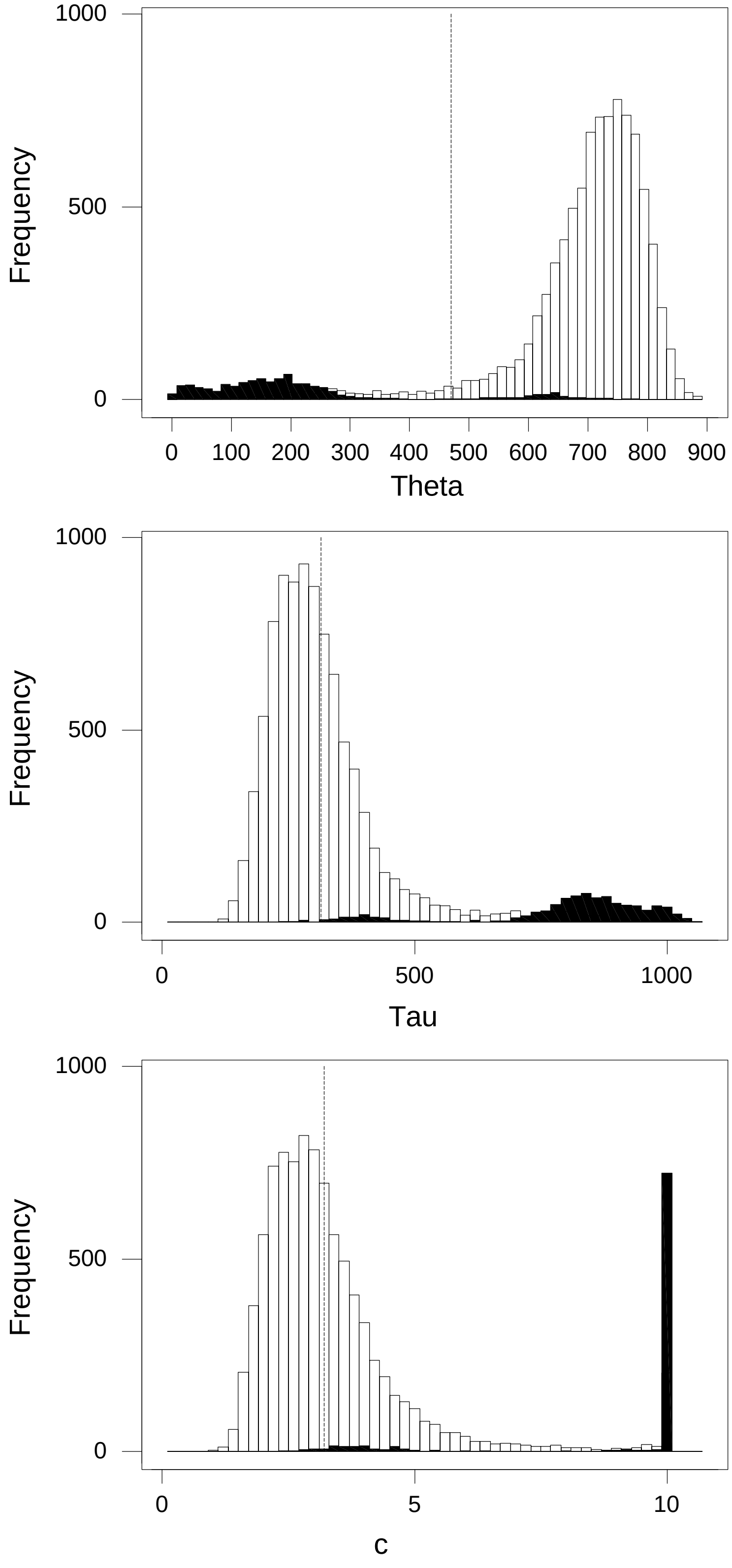

1000
500
0
Frequency
0
100
200
300
400
500
600
700
800
900
Theta
1000
500
0
Frequency
0
500
1000
Tau
1000
500
0
Frequency
0
5
10
c

Figure 5

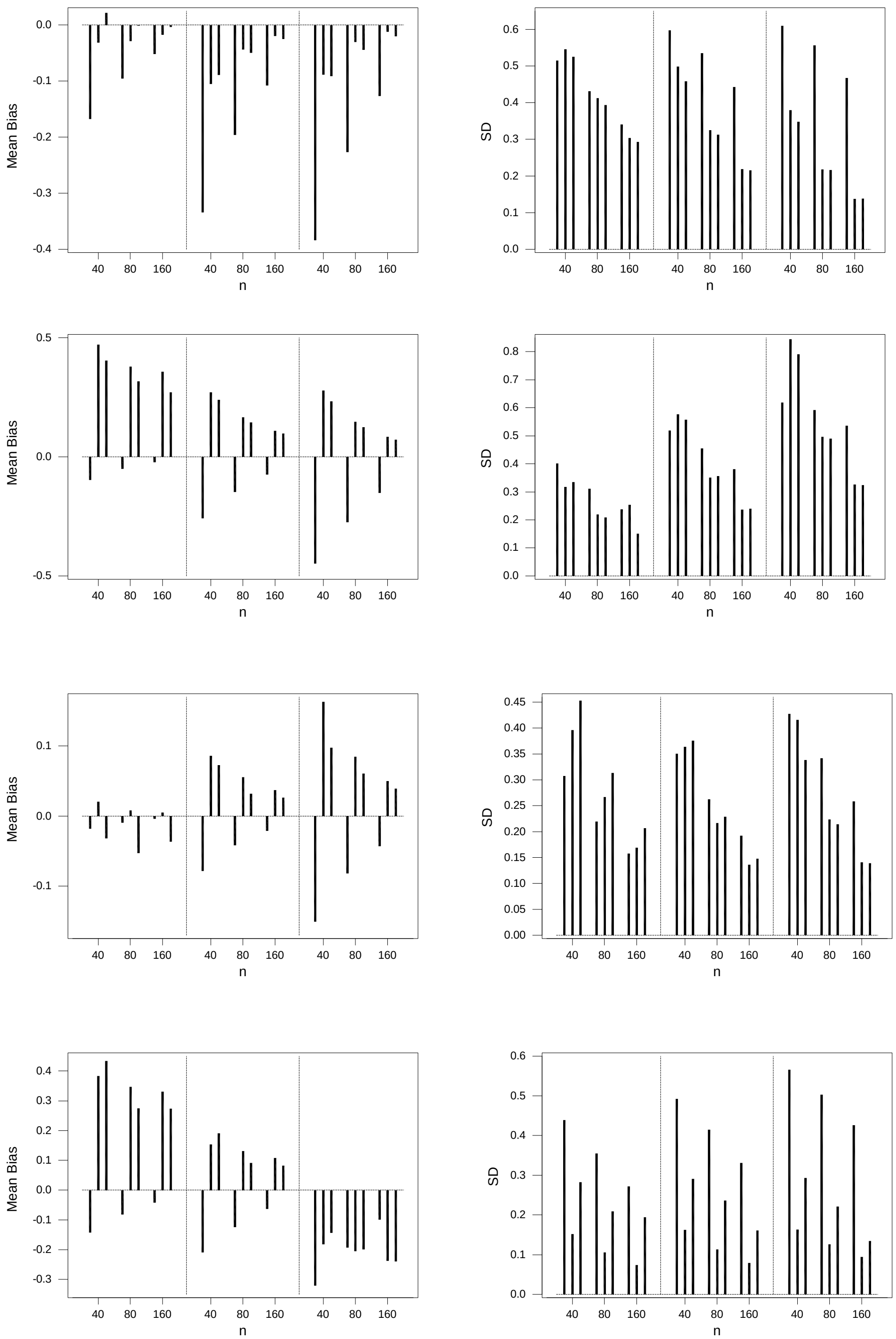


Figure 6.