# Quantile Maximum Likelihood Estimation of Response Time Distributions

Andrew Heathcote[1]

Scott Brown[1]

&

D. J. K. Mewhort[2]

[1]The School of Behavioural Science,
The University of Newcastle, Australia

[2]Department of Psychology
The Queen's University, Canada

Address for Correspondence

Dr. Andrew Heathcote
The School of Behavioural Science,
Building W, The University of Newcastle,
University Avenue, Callaghan, 2308, Australia
Email: heathcote@psychology.newcastle.edu.au

Phone: 61-2-49215952

Suggested running head: Quantile maximum likelihood estimation

**Abstract**

We introduce and evaluate via a Monte Carlo study a robust new estimation technique that fits distribution functions to grouped response time (RT) data, where the grouping is determined by sample quantiles. The new estimator, Quantile Maximum Likelihood (QML), is more efficient and less biased than the best alternative estimation technique when fitting the commonly used ex-Gaussian distribution. Limitations of the Monte Carlo results are discussed and guidance provided for the practical application of the new technique. Because QML estimation can be computationally costly, we make fast open source code for fitting available that can be easily modified to use QML in the estimation of any distribution function.

Recently, psychological researchers have shown increasing interest in characterising the shape of response time (RT) distributions, rather than addressing only a measure of the distribution's central tendency, such as the mean. Ratcliff (1979) demonstrated that moment based estimators, such as skew and kurtosis, are not suitable for characterising the shape of empirical distributions because they suffer from problems with efficiency (i.e., very large sample sizes are required) and robustness (i.e., higher order moments are very sensitive to outliers). He suggested an alternative strategy, characterising shape by fitting an explicit distribution function, most commonly the ex-Gaussian distribution (McGill, 1963), that has been widely adopted (e.g., Andrews & Heathcote, 2001; Balota & Spieler, 1999; Heathcote, Popiel & Mewhort, 1991; Hockley, 1984; Leth-Steenson, Elbaz, & Douglas, 2000; Mewhort, Braun & Heathcote, 1992; Ratcliff & Murdock, 1976; Smith & Mewhort, 1998; Spieler, Balota & Faust, 1996; Wixted & Roher, 1993). In this paper we propose and evaluate a new robust method of fitting distribution functions.

Van Zandt (2000) examined a variety of methods for fitting a distribution function, f, with parameter vector $\boldsymbol{\theta}$, to RT data. She concluded that, for a range of distribution functions commonly used in RT analysis, the generally least variable and biased parameter estimates were obtained by maximum likelihood (ML) estimation. The likelihood of $\boldsymbol{\theta}$ given a data vector **RT**, $L(\boldsymbol{\theta}|\mathbf{RT})$ is proportional to the probability of the data given $\boldsymbol{\theta}$, $p(\mathbf{RT}|\boldsymbol{\theta})$ (likelihood is only defined up to an arbitrary scale factor, Edwards, 1972). For RT data measured with precision 2L (i.e., $RT_i$ falls in the range $RT_i \pm L$, $i = 1 \ldots n$), the probability of observing $RT_i$ is:

$$p(RT_i \mid \boldsymbol{\theta}) = \int_{RT_i - L}^{RT_i + L} f(x, \boldsymbol{\theta}) dx$$

Assuming independent observations, the joint probability of **RT** is the product of the individual probabilities. Using a continuous approximation, $p(RT_i \mid \boldsymbol{\theta}) \approx f(RT_i, \theta)2L$ :

$$L(\boldsymbol{\theta} \mid \mathbf{RT}) \propto \prod_{i=1}^{n} f(RT_i, \boldsymbol{\theta}) \qquad (1)$$

Note that the common factor 2L was absorbed into the arbitrary scale factor (not shown in Equation 1, which is expressed as a proportional relationship) because its value is unrelated to $\boldsymbol{\theta}$. We will call estimates obtained by maximising the right hand side of Equation 1 "Continuous Maximum Likelihood " (CML) estimates. Van Zandt's (2000) results on ML estimation were obtained using the CML method.

While Van Zandt (2000) found CML estimation to be the best method overall, her least-squares CDF estimation method, which minimises the sum of squared deviations between observed and theoretical cumulative probabilities at a set of data quantiles, was almost equally effective. Data quantiles are values below which a given proportion of the observed RT distribution lies, with the median being the most common example. Quantile based methods may actually be superior CML in real data, because appropriately chosen quantiles will not be influenced by outliers. Consistent with superior robustness, Van Zandt, Colonius and Proctor (2000) found that the least-squares CDF method provided more stable estimates of the parameters of the diffusion model (Ratcliff, 1978) than CML fitting for their RT data.

## Quantile Maximum Likelihood Estimation

In this paper we evaluate a new estimation approach, called Quantile Maximum Likelihood (QML) estimation, which combines the robustness of quantiles and the efficiency and consistency of maximum likelihood estimation[1]. While ML estimation based on grouped data is not new (e.g., Kulldorff, 1961), QML differs from earlier approaches in that grouping is determined by sample quantiles. The first step in QML estimation transforms the data vector **RT**, of dimensionality n, into a vector of quantile estimates ($\hat{\mathbf{q}}$) and a vector of counts (**N**) of the number of RTs that occur in each inter-quantile range. We used the following algorithm to calculate quantiles[2].

1. Choose an increasing set of proportions **p**, $0 = p_0 < p_1 < \ldots p_{m-1} < p_m = 1$, $m \leq n$, that correspond to the cumulative probabilities for each quantile.
2. Calculate $N_j = (p_j - p_{j-1})n$ for j = 1 … m, and the quantile estimates $\hat{q}_j = RT_{(I_j^-)} + \left(RT_{(I_j^+)} - RT_{(I_j^-)}\right)\left(I_j - I_j^-\right)$, for j = 1 … (m-1).

$RT_{(k)}$ is the k'th order statistic of **RT** (i.e., the k'th value of **RT** sorted in ascending order), $I_j = p_j n + \frac{1}{2}$, $I_j^-$ is the largest integer less than or equal to $I_j$ and $I_j^+$ is the smallest integer greater than or equal to $I_j$. For example, for the ordered sample (2, 4, 6), and p = (0, 0.3, 0.7, 1):

$$I_1 = 0.3 \text{ x } 3 + 0.5 = 1.4,\ I_j^- = 1,\ I_j^+ = 2.\ \ \hat{q}_1 = 2 + (4\text{-}2)(1.4\text{-}1) = 2.8$$

$$I_2 = 0.7 \text{ x } 3 + 0.5 = 2.6,\ I_j^- = 2,\ I_j^+ = 3.\ \ \hat{q}_2 = 4 + (6\text{-}4)(2.6\text{-}2) = 5.2$$

We set ($\hat{q}_o, \hat{q}_m$) equal to the domain of the distribution function, which is (–∞, +∞) for the commonly used ex-Gaussian.

Maximum likelihood estimation is performed with respect to the transformed data **T** = ($\mathbf{N}, \hat{\mathbf{q}}$). The joint probability of **T** follows a multinomial distribution:

$$p(\mathbf{T} \mid \boldsymbol{\theta}) = \frac{n!}{N_1!N_2!\ldots N_m!}\prod_{j=1}^{m}\left(\int_{\hat{q}_{j-1}}^{\hat{q}_j} f(t,\boldsymbol{\theta})dt\right)^{N_j}$$

Hence the likelihood of the grouped data is**:**

$$L(\boldsymbol{\theta} \mid \mathbf{T}) \propto \prod_{j=1}^{m}\left(\int_{\hat{q}_{j-1}}^{\hat{q}_j} f(t,\boldsymbol{\theta})dt\right)^{N_i} \qquad (2)$$

Note that the multinomial coefficient has been absorbed into the arbitrary scale factor.

We do not claim that QML provides maximum likelihood estimates of $\boldsymbol{\theta}$ conditional on **RT**, because it is not generally true that **T** is a jointly sufficient set of statistics for the estimation of $\boldsymbol{\theta}$. Depending on the choice of **p**, some information relevant to the estimation of $\boldsymbol{\theta}$ may be lost in going from **RT** to **T**. Hence, the QML estimates may differ from ML estimates conditional on **RT**, such as those provided by CML (cf. Example 6.3.1, Edwards, 1972, pp.112-114). However, maximising the right hand side of Equation 2 does provide maximum likelihood estimates of $\boldsymbol{\theta}$ conditional on **T,** so QML estimates have the useful properties of ML estimates, such as consistency. For real data, QML estimates of $\boldsymbol{\theta}$ may be superior to estimates based on CML, because of the robust properties of quantiles.

QML is robust because any RT less than $RT_{\left(l_1^-\right)}$ or greater than $RT_{\left(l_{m-1}^+\right)}$ will have no influence on the quantile likelihood. Selecting **p** involves a trade-off between robustness and a potential loss of information. As the number of quantiles approaches the number of data points, information loss is reduced, as $\hat{\mathbf{q}}$ approaches **RT**, and hence CML and QML estimates converge, but outlying observations can have increasing influence. The next section reports the results of a Monte Carlo study that compared the performance of the CML and QML estimators for the ex-Gaussian distribution. Effects of sample size and numbers of quantiles were also examined. We then discuss the limitations of these results, and the application of QML estimation to alternative RT distribution functions.

## Monte Carlo study

The Monte Carlo study used samples from seven different ex-Gaussian distributions with parameters given in Table 1. The ex-Gaussian distribution is the convolution of a normal distribution (with mean $\mu$ and standard deviation $\sigma$) and an exponential distribution (with mean $\tau$). All distributions had the same mean (1000) and standard deviation (SD = 100) because these values merely fix the measurement units, which are irrelevant to the issue of estimating distribution shape. As in Van Zandt (2000), the value of the standard deviation was chosen to be representative of results in choice RT experiments using milliseconds units, so the results of the simulations can be approximately treated as if they had units of milliseconds.

The shape of a density function may be defined generally as what is left when location and scale are standardized. For the ex-Gaussian distribution, shape can be quantified by the ratio $\underline{K} = \tau/\sigma$. K was varied systematically across the seven distributions. The $\mu$ parameter was varied to maintain a constant overall mean ($\mu + \tau = 1000$) and the magnitudes of $\tau$ and $\sigma$ parameters chosen to maintain a constant standard deviation ($\sqrt{\sigma^2 + \tau^2} = 100$). A useful non-parametric characterization of distribution asymmetry is given by $\underline{A}$ = (Mean – Median)/SD (see Heathcote, 1996 for details). For symmetric distributions, such as the normal, $\underline{A} = 0$, whereas, for choice RT distributions, which are usually positively skewed, $\underline{A} > 0$. For the exponential distribution $\underline{A} = 0.31$, which is also the upper bound for the ex-Gaussian distribution. As shown in Table 1, the simulated ex-Gaussian distribution's shapes varied from almost normal to almost exponential.

-------------------------------
Insert Table 1 about here
-------------------------------

Three sample sizes ($\underline{n}$) were examined, 40, 80, and 160. For each of the 21 combinations of $\underline{n}$ and $\underline{K}$, 35840 separate samples were generated using the normal and exponential random number generators in the Minitab statistical package (Version 12). Both the CML and QML methods were used to estimate ex-Gaussian parameters for each sample. Three different equally spaced sets of quantile estimates were fit using the QML algorithm, so that there were 1 (QML-1, $p_j = j/n$, j = 1 … $n$-1), 2 (QML-2, $p_j = 2j/n$, j = 1 … $n$/2-1), or 4 (QML-4, $p_j = 4j/n$, j = 1 … $n$/4-1) sampled values per inter-quantile range at each sample size.

Fits were obtained by maximising Equation 3 (CML) and Equation 4 (QML) using a conjugate gradients algorithm with a Polak-Ribiere conjugate adjustment to the gradient, and the adaptive Rhomberg method was used to perform numerical integration for the QML objective function (see Press, Teukolsky, Vetterling & Flannery, 1992). Analytic derivatives were used as they greatly reduced the computational cost of QML estimation, which can be expensive for the ex-Gaussian distribution due to the need for numerical integration (see Brown & Heathcote, submitted, for further details of the fitting program, QMLE).

$$\ln\left(L(\boldsymbol{\theta} \mid \mathbf{RT})\right) \propto \sum_{i=1}^{n} \ln f\left(RT_i, \boldsymbol{\theta}\right) \quad (3)$$

$$\ln\left(L(\boldsymbol{\theta} \mid \mathbf{T})\right) \propto \sum_{j=1}^{m} N_j \ln \int_{\hat{q}_{j-1}}^{\hat{q}_j} f(t, \boldsymbol{\theta})\,dt \quad (4)$$

Start points for optimisations were determined by heuristics applied to sample values as described in Heathcote (1996). Generally, this produced faster convergence than using the true values. Less than 1% of fits were removed from further consideration because of failed evaluations of the log-likelihood at convergence. Such results were not due to local maxima and could not be avoided by using alternative start points. Instead, they represented global maxima where either the τ or σ estimates converged to zero because the sampled distribution has no right tail or body respectively. The maxima for the QML method were generally more sharply defined than the maxima for the CML method, as indicated by a lower percentage of fits that terminated due to exceeding the maximum number of iterations allowed by the fitting algorithm (75 iterations, a large value, usually well in excess of the number of iterations required for convergence). Increasing the maximum number of iterations did not result in any improvement in these parameter estimates. The following analyses were carried out both with and without the non-convergent estimates, and the pattern of results was the same, except that variability was reduced slightly when they were excluded. We report the results of analyses with these estimates retained, as that is more representative of the practice with real data, where censoring runs the risk of inducing sampling bias.

## QML Results

Figure 1 presents the results for QML estimation with one observation per inter-quantile range. Performance was excellent for all parameters, especially σ, and especially in the K = 2 … 5 range that is most representative of real choice RT data. Bias, as indicated by the absolute deviation of both the mean and median from the true values, generally decreased with increasing sample size. Hence, QML estimation appears to be consistent (i.e., bias approaches zero with increasing sample size). Generally, bias was positive for μ and negative for τ (as they sum to give the mean such tradeoffs are to be expected) except for K = 1/3, where the reverse held.

--------------------------------
Insert Figure 1 about here
--------------------------------

Sampling variability, indicated by both the range containing 95% of parameter estimates and the inter-quartile range in Figure 1, decreased with sample size. Efficiency improved with increasing K, especially for μ estimates, and was particularly good in the important K = 2 … 5 range. Generally, σ estimates were the most efficient, although μ estimates were equally good for higher values of K (note that the ranges for the panels in Figure 1 differ in order to represent results clearly). The 95% ranges and inter-quartile ranges indicate that the parameter sampling distributions were quite symmetric, which is desirable when parameter estimates are subjected to normal theory analysis, such as ANOVA. Clear exceptions occur for τ estimates for K = 1/2 and 1/3, which produced positively skewed sampling distributions, as τ is bounded below, and its true value is close to zero in these cases.

## QML vs. CML Estimation

Figure 2 compares the bias and efficiency of the CML and QML-1 estimates. The difference in bias is indicated as the absolute deviation from the true value for the CML estimate minus the absolute deviation for the QML-l estimate. The difference in efficiency is indicated by the standard deviation of the CML estimates minus the standard deviation of the QML-1 estimate. Positive values indicate superior performance (less biased, more efficient) for the QML-1 estimates. For almost all cases, except K = 1/3, the QML-1 estimates were less biased than the CML estimates. The difference was most marked of the μ parameter, and for the τ parameter in small samples. A small reversal occurred in τ for larger samples when K = 4. The difference was smaller for σ, but QML-1 clearly did better in the important K = 2 … 5 range.

--------------------------------
Insert Figure 2 about here
--------------------------------

The superior performance for QML-1 was more marked in terms of efficiency, particularly for estimates of μ, where the increase in efficiency for larger K was very large. This reflects the strong decrease in parameter estimate variability with K noted in Figure 1, and shows that the same decrease does not occur for CML estimates. QML-1 σ estimate were also more efficient, especially for larger K, whereas QML-1 τ estimates are only marginally more efficient, although they were still better in almost all cases.

Figure 3 compares the QML-4 estimates (with four observations in each inter-quantile range), with the CML estimates. The pattern of results is essentially identical to the pattern for the QML-1 estimates, except that the advantage over CML estimates is slightly reduced. The results for QML-2 estimates, which are not shown for brevity, fall between those for QML-1 and QML-4. Hence, it appears that robust estimates can be obtained with little cost in terms of bias or efficiency.

--------------------------------
Insert Figure 3 about here
--------------------------------

Results were also obtained for 8 observations per inter-quantile range, in samples of 80 and 160, and for 16 observations per inter-quantile range, for samples of 160. These findings omitted previously to facilitate the display of results. The results for 8 and 16 observations per inter-quantile range followed a similar pattern to those for 1-4 observations per range; both bias and sampling variability increased only slightly as more observations were included in each inter-quantile range. Even with the larger numbers of observations per inter-quantile range, QML retained its marked superiority over CML in the efficiency of μ estimates. Bias differences were also negligible, although QML maintained a slight superiority, on average.

Overall, parameter estimation was markedly more biased and somewhat less efficient for the more symmetric ex-Gaussian distributions. However, for the range K = 2 … 5, bias was negligible even for the smaller sample sizes, particularly for the σ parameter. Bias was stronger for the μ and τ estimates, but was still small even for n = 40. The bias in μ and τ estimates was complementary, with μ being slightly overestimated and τ being slightly underestimated. Tables of bias, both in means and medians, and efficiency, as standard deviations, for the entire simulation study are available via the web[3]. The web site provides examples of how to use the tables to determine the magnitude of potential confounding due to bias and to estimate the sample size required to find an effect of a given magnitude in ex-Gaussian parameter estimates. Due to the large number of replicates used in the Monte Carlo study, the tabled values are precise, and so appropriate for this purpose.

## Discussion

The Monte Carlo study found that QML was generally less biased and much more efficient than CML estimation. These findings support QML as the method of choice for estimating ex-Gaussian distribution parameters. The advantage for QML over CML was largely maintained when the QML estimates were based on up to sixteen times fewer quantiles than the number of data points. This finding indicates that the advantages of QML estimation in terms of robustness against outliers can be exploited with only a small cost in bias and efficiency. Note that no outliers were included in the simulated data. Hence, the QML may enjoy an even greater advantage over the CML in real RT data.

The largest advantage for QML over CML was for estimates of $\mu$, particularly for more asymmetric distributions. The $\mu$ parameter approximately indicates the location of the ex-Gaussian mode. More variable estimates of $\mu$ for more asymmetric distributions occur because the mode is less well defined when the shape of the ex-Gaussian distribution becomes dominated by the exponential component. The improved performance for $\mu$ estimates did not appear to be associated with a cost for $\tau$ estimates, an important finding because studies often focus on both $\mu$ and $\tau$ estimates (e.g., Andrews & Heathcote, in press; Balota & Spieler, 1999; Spieler et al., 1996). The excellent results for estimation of the $\sigma$ parameter suggest that it deserves more attention in characterising RT distribution effects. For example, the relative proportions of observations in the body and right tail of the distribution may be examined through estimates of the $K = \tau/\sigma$ ratio used in the Monte Carlo study.

The simulation results indicate that the ex-Gaussian distribution is less useful for estimating shape when RT distributions are close to symmetric. The A measure provides a practical way of determining asymmetry prior to fitting, as it requires only estimates of the mean, median, and standard deviation. A reasonable heuristic is that estimation of distribution shape using the ex-Gaussian is safe for $A > 0.15$, corresponding to simulation results in the range $K = 2 \ldots 5$.

An important issue in the application of QML estimation is choosing the number of quantiles. The choice represents a trade-off between estimation accuracy and robustness. Using lesser numbers of quantiles provides protection against outliers. However, the simulation results indicate that bias is minimised and efficiency maximised by using the largest number of quantiles compatible with sample size. Fortunately, QML estimates were still superior to CML estimates for up to 16 observations per inter-quantile range, at least when a minimum of ten quantiles was enforced. A strategy that could protect against outliers while minimising bias and efficiency costs, is to use larger numbers of observations in the first and last inter-quantile ranges than are used for the body of the distribution. Larger first and last ranges will reduce information about the tails of the distribution, but where outliers are suspected, parameter estimates may still benefit. The simulation reported here always used equal numbers of observations in each range, so some caution is warranted in extrapolating from the present results.

## Limitations and Extensions

The ex-Gaussian distribution was chosen for the present investigation because it is widely used and usually provides a better characterisation of choice RT distribution than most other simple three-parameter distribution functions (except perhaps the shifted lognormal distribution, see Ratcliff & Murdock, 1976; Wixted & Roher, 1993). Some caution should be exercised in generalising the results of the Monte Carlo study to other distribution functions, as the properties of QML estimates depend on the specific distribution function and quantiles employed. In particular, the Monte Carlo results do not demonstrate that QML estimation will be less biased and more efficient than CML estimates for all distribution functions. Van Zandt (personal communication) found QML-4 parameter estimates for samples of 160 from an exponential distribution with a mean of 100 to be slightly more biased and variable than CML estimates. While the advantage for CML was relatively minor, this case demonstrates the need for further investigation to determine the relative efficiency and bias of CML and QML for other distribution functions. Even in cases where QML is slightly more biased or less efficient than CML, it may still be preferred because any loss of information resulting from reducing the raw data to quantiles may be more than compensated for by increased robustness.

An apriori implausible assumption made by the ex-Gaussian is that RT is not bounded below. While unrealistic, this feature provides robustness against fast anticipatory responses, which can greatly distort parameter estimates for a distribution function that is bounded below. Densities with parameter dependent domains, such as the shifted lognormal or Weibull distributions, are more plausible in this regard, but require the lower bound to be estimated. Where a densities' domain is parameter dependent, the first and last quantiles must be chosen differently than in the Monte

Carlo study, such as by setting the first quantile equal to the minimum observation (i.e., $\hat{q}_o$ = $RT_{(1)}$). The cost is that QML will no longer be robust to anticipatory responses. A more robust approach, but one which is unlikely to be unbiased, or even consistent, is to omit the probability of an observation in the first and/or last inter-quantile range in the calculation of Equation 4. Further investigation is required in order to determine the best method.

The limitations of the Monte Carlo results should not be taken to mean that QML cannot be applied to distribution functions other than the ex-Gaussian. The QML approach is very general, and can even be applied, via Monte Carlo methods, where the distribution function is not known analytically or difficult to evaluate directly due to its complexity. In particular, any RT distribution model, no matter how complex, can be fit by QML if independent samples can be obtained from the model. For any given parameter setting, $\boldsymbol{\theta}$, the model probabilities, $\pi_j = \int_{\hat{q}_{j-1}}^{\hat{q}_j} f(t, \boldsymbol{\theta}) dt$ can be obtained by counting the proportion of samples that fall between each data quantile[4]. For example, for a k choice model (i = 1 … k) specifying RT distributions for both correct and error responses, Monte Carlo estimates of the probability of each type of response in each inter-quantile range (e.g., $\pi_{i,j,Correct}$ and $\pi_{i,j,Error}$) can be used to construct the likelihood : $\sum_{i=1}^{k} \sum_{j=1}^{m_{i,Correct}} N_{i,j,Correct} \ln \pi_{i,j,Correct} + \sum_{i=1}^{k} \sum_{j=1}^{m_{i,Error}} N_{i,j,Error} \ln \pi_{i,j,Error}$ . The number of quantiles for each response type, $\underline{m}_{i,Correct}$ and $\underline{m}_{i,Error}$, can be chosen to suit the sample sizes available, with fewer quantiles employed for rare responses. Even when the model specifies a mixture that is not identifiable in the data (e.g. a small proportion of distracted or anticipatory responses), QML can still be used to fit the mixture distribution with the π values reflecting the effect of the mixture.

QML fitting is also ideal for the estimation of group RT distributions (Ratcliff, 1979) in paradigms that do not allow sufficient observations to be collected to estimate RT distribution for each subject and condition. The estimation method used by Ratcliff (1979) applied CML to quantiles as if they were raw data. This is only exact when the number of quantiles equals the number of RTs, and so is not suitable for use with the larger inter-quantile ranges required for robustness. QML estimation is, therefore, more appropriate, and it was this problem that motivated the initial development of QML. However, unless the components of the group form a scale-location family (Thomas & Ross, 1980) the quantitative form of the group distribution may be distorted. Ratcliff's (1979) results indicate that the distortion is relatively small for the ex-Gaussian distribution in parameter ranges typical of data. The QML approach removes some of the motivation for examining group distribution functions because its robustness, and improved efficiency for the ex-Gaussian, allows it to be applied to smaller samples than CML. However, a group distribution approach will still be necessary for some paradigms. The Monte Carlo results reported here come from a larger study that examined group distribution estimation using CML and QML.

QML estimation is computationally costly because the evaluation of cumulative distribution functions, such as for the ex-Gaussian, often involves numerical integration. Hence, efficient implementations of fitting algorithms, and ideally analytic derivatives for the QML objective function, are required. Brown and Heathcote (submitted) make the Fortran 90 program used in the Monte Carlo study, QMLE, available as open source code[3]. Although approximately an order of magnitude slower than CML, this efficient implementation of QML is fast enough to fit most empirical data sets on a PC. For larger data sets and Monte Carlo studies, QMLE can be compiled for parallel execution, and so can take advantage of multipleprocessor workstations (see Pollard, Mewhort & Weaver, 2000 for discussions of parallel programming issues). The open source approach taken with QMLE allows researchers to augment the code to fit other distribution functions. Brown and Heathcote (submitted) also provide formulae for the derivative and Hessian of a QML objective function that require only knowledge of the derivative and Hessian of the probability density function. Given a specification of the latter, QMLE automatically provides the former, so implementing QML fitting is no harder than implementing CML fitting.

QML estimation is ideally suited to graphical examination of misfit through QQ plots (Cleveland, 1985) of observed versus fitted quantiles. The contribution of each quantile to the overall misfit can be quantified by examining the corresponding terms in the sum in Equation 4. Because QML is a ML method, parameter standard errors and correlations can be estimated through inversion of the Hessian (second partial derivative) matrix of the likelihood function. The Hessian measures the likelihood surface's curvature at the maximum, and so quantifies how sharply the maximum is defined, and consequently how precisely parameters are estimated (see Edwards, 1972 for details and Roher & Wixted, 1994, for an example of this approach)[5]. Brown and Heathcote's (submitted) QMLE program provides parameter standard errors and correlations based on the Hessian, as well as observed and fitted quantiles, and the corresponding terms from Equation 4.

More generally, analytic techniques based on quantiles are now available that rival the scope of classical least-squares methods. Quantile regression (Bassett & Koenker, 1978; Rosseeuw & Leory, 1987) provides robust estimation of covariate models of the median, using the fact that the median minimises the sum of absolute deviations. The latter property can be used to generalise the regression approach to arbitrary quantiles and to provide quantile estimates for large data sets without sorting (Hunter & Lange, 2000). In an approach similar to the Pearson system of distributions, which is estimated by matching the first four moments, Morgenthaller and Tukey (2000) describe a flexible family of distributional shapes that can be easily fit to quantiles. This family can encompass not only variations in symmetry, but also heavy and light tailed cases, so that it is not only more robust, but also more flexible, than the Pearson approach. QML estimation provides one more tool in this growing collection of robust quantile-based methods.

## References

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? Journal of Experimental Psychology: Human Perception and Performance, 27, 514-544.

Bassett, G. & Koenker, R. (1978). Asymptotic theory of least absolute error regression. Journal of the American Statistical Association, 73, 618-622.

Brown, S., & Heathcote, A. (submitted). QMLE: Fast, robust and efficient estimation of distribution functions based on quantiles.

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.

Cleveland, W. S. (1985). The elements of graphing data. Monterey, CA: Wadsworth.

Davison, A. C., & Hinkley, D. V. (1997). Bootstrap methods and their application. New York: Cambridge UP.

Edwards, A. W. F. (1972). Likelihood. London: Cambridge University Press.

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

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (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, 10, 598-615.

Hunter, D. R. & Lange, K. (2000). Quantile regression via an MM algorithm, Journal of Computational and Graphical Statistics, 9, 60-77.

Kulldorff, G. (1961). Estimation from grouped and partially grouped samples. Wiley: New York.

Leth-Steensen, C., Elbaz, Z. K., & Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. Acta Psychologica, 104, 167-190.

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

McGill, W. J. (1963). Stochastic latency mechanisms. In R.D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology, (pp. 193-199). New York: Wiley.

Mewhort, D. J. K., Braun, J. G., & Heathcote, A. (1992). Response time distributions and the Stroop task: A test of the Cohen, Dunbar, and McClelland (1990) model. Journal of Experimental Psychology: Human Perception and Performance, 18, 872-882.

Morgenthaler, S. & Tukey, J. W. (2000). Fitting quantiles: Doubling, HR, HQ, and HHH distributions. Journal of Computational and Graphical Statistics, 9, 180-195.

Pollard, A., Mewhort, D. J. K., &Weaver, D. F. (2000). High Performance Computing Systems and Applications. Boston: Kluwer.

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

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

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

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

Rousseeuw, P. J. & Leroy, A. M. (1987). Robust Regression and Outlier Detection. New Your: Wiley.

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.

Tanner, M. A. (1993). Tools for Statistical Inference, New York: Springer-Verlag.

Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. Journal of Mathematical Psychology, 21, 136-152.

Ulrich, R., & Miller, J. (1994). Effects of outlier exclusion on reaction time analysis. Journal of Experimental Psychology: General, 123, 34-80.

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

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, 208-256.

Wand, M. P. & Jones, M. C. (1995). Kernel Smoothing, London: Chapman & Hall.

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

## Table

Table 1. Parameters and statistics of the simulated ex-Gaussian distributions.

| Mean | SD | $\mu$ | $\sigma$ | $\tau$ | K= $\tau$/ $\sigma$ | A |
|---|---|---|---|---|---|---|
| 1000 | 100 | 968.377 | 94.868 | 31.623 | 1/3 | 0.0098 |
| 1000 | 100 | 955.279 | 89.443 | 44.721 | 1/2 | 0.0245 |
| 1000 | 100 | 929.289 | 70.711 | 70.711 | 1 | 0.0880 |
| 1000 | 100 | 910.557 | 44.721 | 89.443 | 2 | 0.1890 |
| 1000 | 100 | 905.132 | 31.623 | 94.868 | 3 | 0.2420 |
| 1000 | 100 | 902.986 | 24.254 | 97.014 | 4 | 0.2675 |
| 1000 | 100 | 901.942 | 19.611 | 98.058 | 5 | 0.2810 |

## Acknowledgements

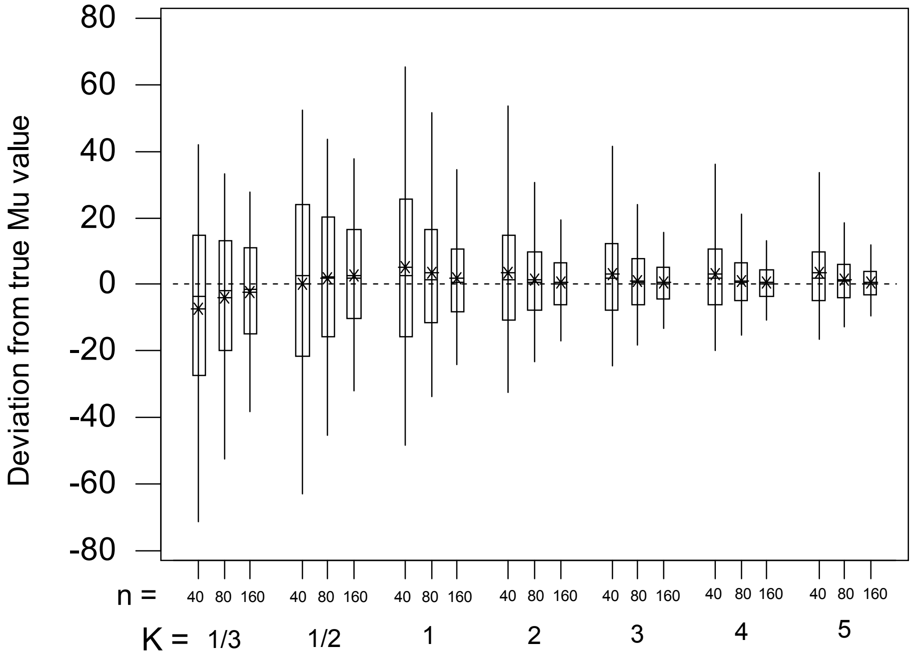
We would like to express our thanks to Trisha Van Zandt and an anonymous reviewer for helpful advice, and to Dr. David Koch of the Faculty of Engineering, University of Newcastle, Australia, for providing the computational support that made this project possible. We acknowledge financial assistance from Australian Research Council Grant A10007174 to S. Andrews and A.H., a grant in support of sabbatical travel provided to A. H. by the Research Management Committee of the University of Newcastle, and an NSERC equipment grant and a Sun AEG equipment grant to D. J. K.M. Correspondence concerning this article should be addressed to A. Heathcote, School of Behavioural Science, University of Newcastle, Callaghan, 2308, NSW, Australia (email: heathcote@psychology.newcastle.edu.au).

## Figure Captions

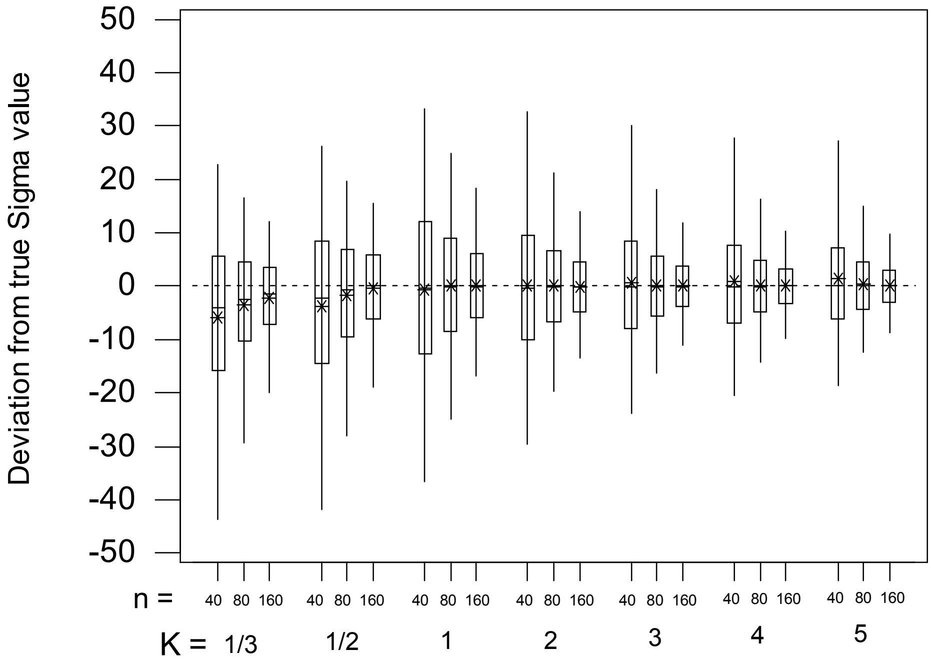
Figure 1. Deviations from the true parameter values for QML-1 estimates of the ex-Gaussian parameters: (a) $\mu$, (b) $\sigma$, and (c) $\tau$. Rectangles indicate the inter-quartile range, the horizontal lines with rectangles indicate the median, stars indicate the mean, and the long vertical lines span 95% of parameter estimates.

Figure 2. CML minus QML-1 Bias (absolute deviation from the true value) and Standard Deviation (SD)

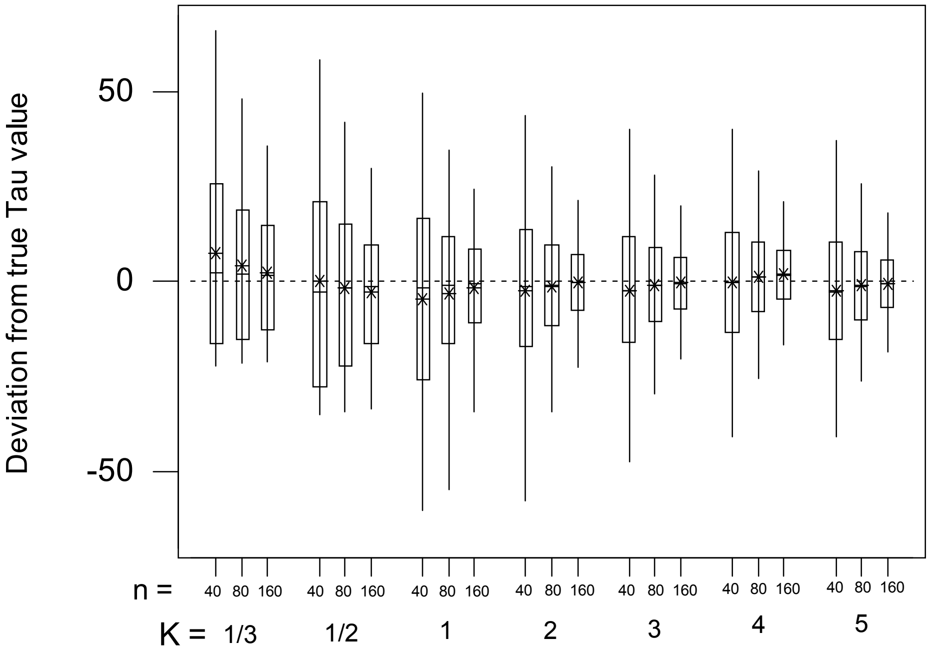
Figure 3. CML minus QML-4 Bias (absolute deviation from the true value) and Standard Deviation (SD)


80
60
40
20
0
-20
-40
-60
-80
Deviation from true Mu value
n =
40 80 160
K = 1/3
1/2
1
2
3
4
5

(a)


Deviation from true Sigma value
50
40
30
20
10
0
-10
-20
-30
-40
-50
n =
40 80 160 40 80 160 40 80 160 40 80 160 40 80 160 40 80 160 40 80 160
K = 1/3
1/2
1
2
3
4
5

(b)



Deviation from true Tau value
50
0
-50
n =
40 80 160 40 80 160 40 80 160 40 80 160 40 80 160 40 80 160 40 80 160
K = 1/3 1/2 1 2 3 4 5


(c)

Figure 1

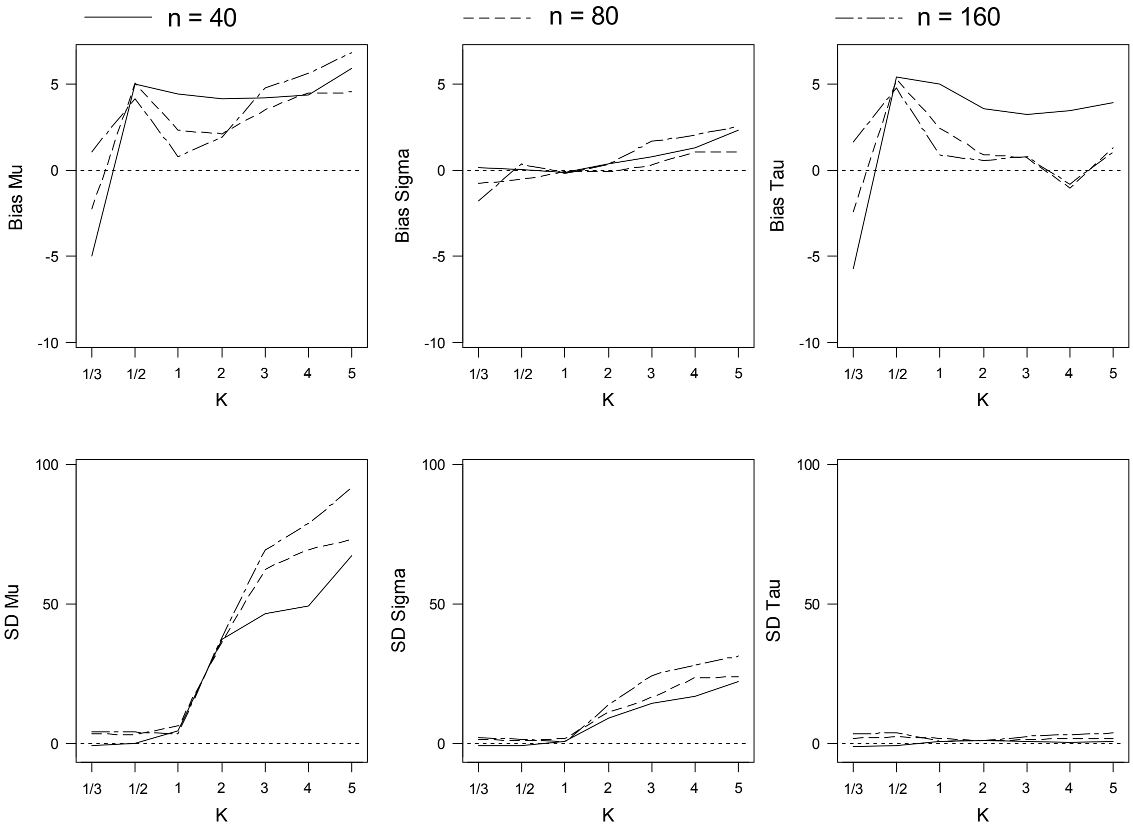

n = 40
n = 80
n = 160
Bias Mu
Bias Sigma
Bias Tau
SD Mu
SD Sigma
SD Tau
K
1/3
1/2
1
2
3
4
5
-10
-5
0
5
50
100

**Figure 2**

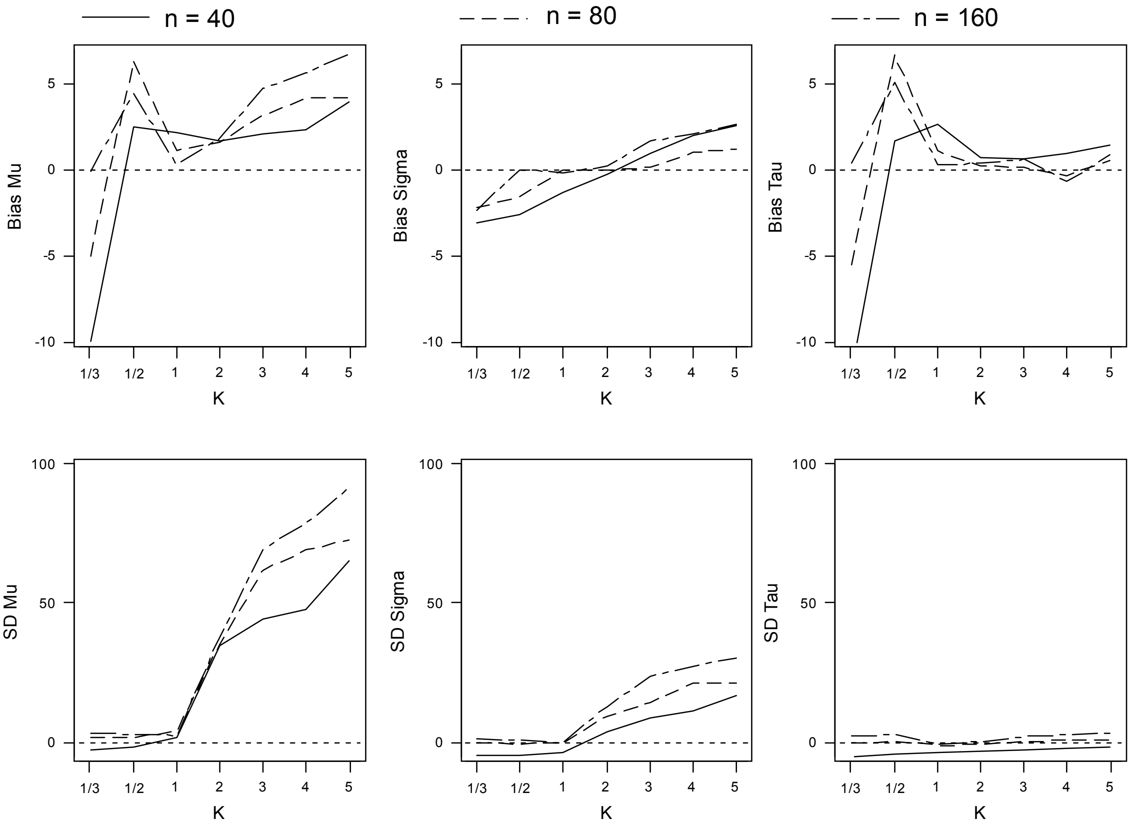

n = 40
n = 80
n = 160
Bias Mu
Bias Sigma
Bias Tau
SD Mu
SD Sigma
SD Tau
K
1/3
1/2
1
2
3
4
5
-10
-5
0
5
50
100

**Figure 3**

## Footnotes

[1]Ulrich and Miller (1994) also suggested a robust likelihood function that combines aspects of Equations 1 and 2. It is used when $r_1$ samples below a lower cut-off, a, and $r_2$ samples above an upper cut-off b have been censored from a sample:

$$L(\boldsymbol{\theta} \mid RT) \propto \left( \int_{-\infty}^{a} f(t,\boldsymbol{\theta})dt \right)^{r_1} \text{ x } \prod_{i=1}^{n-r_1-r_2} f(RT_i,\boldsymbol{\theta}) \text{ x } \left( \int_{b}^{\infty} f(t,\boldsymbol{\theta})dt \right)^{r_2}$$

Ulrich and Miller's approach is implemented in RTSYS (Heathcote, 1996).

[2] The algorithm defines a linearly interpolated quantile estimate. Other estimates are possible, e.g. using $I_j = p_j(n-1)+1$, which converges with the definition used here for large n.

[3]Go to http://psychology.newcastle.edu.au and follow the links to Heathcote's home page.

[4] More sophisticated approaches could use smoothed cumulative distribution function estimates (e.g. Wand & Jones, 1995) and more frugal sampling schemes (e.g., Tanner, 1993).

[5]Note that statistics based on the Hessian are only approximate when estimation is nonlinear. More accurate results can be obtained via bootstrapping (e.g., Davison & Hinkley, 1997).