*Psychonomic Bulletin & Review*
*2004, 11 (3), 574-576*

# Notes and Comment

## A comment on Heathcote, Brown, and Mewhort's QMLE method for response time distributions

PAUL L. SPECKMAN and JEFFREY N. ROUDER
*University of Missouri, Columbia, Missouri*

*Heathcote, Brown, and Mewhort (2002) have introduced a new, robust method of estimating response time distributions. Their method may have practical advantages over conventional maximum likelihood estimation. The basic idea is that the likelihood of parameters is maximized given a few quantiles from the data. We show that Heathcote et al.'s likelihood function is not correct and provide the appropriate correction. However, although our correction stands on firmer theoretical ground than Heathcote et al.'s, it appears to yield worse parameter estimates. This result further indicates that, at least for some distributions and situations, quantile maximum likelihood estimation may have better nonasymptotic properties than a more theoretically justified approach.*

Heathcote, Brown, and Mewhort (2002) have recently introduced a new method of estimating response time distributions that is based on quantiles. The likelihood of the parameters of the distribution is expressed as a function of these quantiles. Parameter values that maximize this likelihood are used as point estimates. The method is called *quantile maximum likelihood estimation* (QMLE). As discussed by Heathcote et al., there are three main advantages of QMLE. First, because QMLE is based on order statistics, it is more robust to outliers than conventional maximum likelihood estimation (MLE). Second, because QMLE uses only a subset of the observations, it is computationally faster than conventional MLE. This additional speed may prove important in large-scale problems with several participants and conditions. Third, the small-sample properties of QMLE may in fact be more desirable than conventional MLE. Given these advantages, QMLE is an attractive tool for research. In this paper, we demonstrate that Heathcote et al.'s framework has a logical flaw and provide a correction. We also confirm Heathcote et al.'s observation that in some situations, QMLE provides superior estimation.

This research is supported by National Science Foundation Grant SES-0095919 to J. N. Rouder, D. Sun, and P. L. Speckman and National Science Foundation Grant DMS-9972598 to D. Sun and P. L. Speckman. We are indebted to Scott Brown and Andrew Heathcote for their support and time. They shared their simulations, evaluated ours, and engaged in several helpful rounds of discussion. Correspondence should be addressed to J. N. Rouder, 210 McAlester Hall, University of Missouri, Columbia, MO 65211 or to jeff@banta. psyc.missouri.edu.

We discuss Heathcote et al.'s (2002) method and the correction within the context of a simple example. Consider the 20 observations, ordered from smallest to largest, in Table 1. Our goal is to fit a parametric distribution, such as the ex-Gaussian distribution (Hohle, 1965), to the data. With Heathcote et al.'s method, one first selects a set of probabilities, $0 < p_1 < \cdots < p_{m-1} < 1$ (we follow their notation here). In our example, we select $p_1 = .125$, $p_2 = .375$, $p_3 = .625$, and $p_4 = .875$. Next, one computes quantile estimates corresponding to these probabilities. There are a number of ways to compute quantile estimates (see Hyndman & Fan, 1996, for a review). In Heathcote et al., quantile estimates are either an order statistic or a weighted average of adjacent order statistics (see Heathcote et al., 2002, for details). Order statistics refer to the case in which data are ranked from smallest to largest. The second order statistic refers to the second smallest observation, the third to the third smallest observation, and so forth. In the example, the selected probabilities, $p_1, \ldots, p_4$, correspond to quantiles estimated by the 3rd, 8th, 13th, and 18th order statistics, respectively. These observations are bold-faced in Table 1 and are denoted as $\vec{q} = (431, 476, 525, 647)$. These quantile estimates divide the observations into cells or bins. A second vector, $\vec{n}$, denotes the number of observations in each cell defined by the quantiles. Heathcote et al. present a formula for calculating $\vec{n}$, and in this example $\vec{n} = (2.5, 5, 5, 5, 2.5)$. The final step is to relate these quantiles to parameters of the distribution. Let's assume the distribution is continuous and denote its density and its cumulative distribution function as $f$ and $F$, respectively. According to Heathcote et al., a multinomial distribution can be used to relate the quantiles to parameters. Their likelihood for the parameters is denoted $L_M$ and is given by

$$L_M\left(\vec{\theta} \mid \vec{q},\vec{n}\right) \propto F\left(q_1;\vec{\theta}\right)^{n_1}\left[F\left(q_2;\vec{\theta}\right) - F\left(q_1;\vec{\theta}\right)\right]^{n_2} \times \cdots \times \left[1 - F\left(q_{m-1};\vec{\theta}\right)\right]^{n_m}, \quad (1)$$

where $m$ is one plus the number of quantiles (5 in this case) and $\vec{\theta}$ is the vector of parameters. In this example,

$$L_M\left(\vec{\theta} \mid \vec{q},\vec{n}\right) \propto F\left(431;\vec{\theta}\right)^{2.5}\left[F\left(476;\vec{\theta}\right) - F\left(431;\vec{\theta}\right)\right]^5 \times\left[F\left(525;\vec{\theta}\right) - F\left(476;\vec{\theta}\right)\right]^5 \times\left[F\left(647;\vec{\theta}\right) - F\left(525;\vec{\theta}\right)\right]^5 \times\left[1 - F\left(647;\vec{\theta}\right)\right]^{2.5}.$$

**Table 1**
**Example With 20 Observations**

| | | | | |
|---|---|---|---|---|
| 413 | 413 | **431** | 435 | 451 |
| 461 | 476 | **476** | 481 | 503 |
| 509 | 511 | **525** | 555 | 565 |
| 613 | 614 | **647** | 697 | 767 |

This is indeed a strange multinomial likelihood. In proper use of the multinomial distribution, the range of the distribution is divided into cells and counts are obtained in these cells. Cell probabilities can be calculated on the basis of parameters, and the multinomial distribution is the basis of the well-known chi-square goodness-of-fit test (Pearson, 1900). For the true multinomial distribution, the cell counts $n_1, \ldots, n_m$ are integers and random, whereas for QMLE, the cell counts are nonrandom and may be fractions. For the true multinomial, the cell boundaries $q_O, \ldots, q_m$ are fixed and nonrandom. Here, these boundaries are random. Hence, the multinomial is not the appropriate likelihood.

A valid likelihood may be derived directly using the well-known formula for a subset of order statistics from a continuous distribution (Hogg & Craig, 1978). Let $\vec{j}$ denote an increasing sequence of ranks, for example, $\vec{j} = (3, 8, 13, 18)$. Let $\vec{q} = q_1, \ldots, q_m$ denote a sequence of increasing order statistics with ranks given by $j$. For the example in Table 1, if $\vec{j} = (3, 8, 13, 18)$, then $\vec{q} = (431, 476, 525, 647)$. The joint density for $\vec{q}$ given the parameters is

$$g\left(\vec{q}\,|\,\vec{\theta}\right) \propto F(q_1)^{j_1-1} f(q_1)\left[F(q_2) - F(q_1)\right]^{j_2-j_1-1} \times f(q_2)\cdots f(q_m)\left[1 - F(q_m)\right]^{N-j_m}.$$

The likelihood of the parameters given the order statistics falls directly from this density and is given by $L_O$:

$$L_O\left(\theta\,|\,\vec{q},\vec{j}\right) \propto F\left(q_1;\vec{\theta}\right)^{j_1-1} f\left(q_1;\vec{\theta}\right) \times\left[F\left(q_2;\vec{\theta}\right) - F\left(q_1;\vec{\theta}\right)\right]^{j_2-j_1-1} \times f\left(q_2;\vec{\theta}\right)\cdots f\left(q_m;\vec{\theta}\right)\left[1 - F\left(q_m;\vec{\theta}\right)\right]^{N-j_m}. \tag{2}$$

For the specific example at hand,

$$L_O\left(\theta\,|\,\vec{q},\vec{j}\right) \propto F\left(431;\vec{\theta}\right)^2 f\left(431;\vec{\theta}\right) \times\left[F\left(476;\vec{\theta}\right) - F\left(431;\vec{\theta}\right)\right]^4 \times f\left(476;\vec{\theta}\right)\left[F\left(525;\vec{\theta}\right) - F\left(476;\vec{\theta}\right)\right]^4 \times f\left(525;\vec{\theta}\right)\left[F\left(647;\vec{\theta}\right) - F\left(525;\vec{\theta}\right)\right]^4 \times f\left(647;\vec{\theta}\right)\left[1 - F\left(647;\vec{\theta}\right)\right]^2.$$

Likelihood $L_O$ is a proper generalization of conventional likelihood in that if one selects every order statistic (i.e., all of the observations), then $L_O$ is equivalent to the conventional likelihood of the parameters. Heathcote et al.'s $L_M$ does not reduce to conventional likelihood. Likelihood $L_O$ is also a generalization of the likelihood for censored data (Dolan, van der Maas, & Molenaar, 2002; Ulrich & Miller, 1994).

Because our alternative is the valid likelihood of parameters given sample statistics, we can make use of standard statistical theory about maximum likelihood estimation. This theory guarantees that under a set of conditions, the regularity conditions, maximizing $L_O$ yields consistent and asymptotically maximally efficient estimates for order statistics (Lehmann, 1991). It may be true that Heathcote et al.'s (2002) method is also asymptotically unbiased and maximally efficient, but we know of no proof of this statement. (Brown & Heathcote, 2003, show that QMLE estimates are asymptotically normal, but they do not show asymptotic maximal efficiency.)

To explore the consequences of the corrected likelihood, we conducted simulation studies with the ex-Gaussian distribution (Hohle, 1965). In the first study, each data set consisted of 20 observations sampled from an ex-Gaussian with parameters $\mu = 400$ msec, $\sigma = 50$ msec, and $\tau = 100$ msec. The 3rd, 8th, 13th, and 18th order statistics were then selected from these observations. Next, parameter estimates were obtained by maximizing[1] $L_M$ and $L_O$. To construct the sampling distribution of the estimators, we repeated this sampling and estimation process 10,000 times. Results are shown in Table 2 under the heading "$N = 20$, 4 order statistics." Surprisingly, maximizing $L_M$ yielded estimates with less bias and smaller root mean squared error (RMSE) than did maximizing $L_O$. To explore this behavior further, we conducted additional studies by increasing the sample size (80 observations per participant) and including all observations. In all cases, maximizing $L_M$ outperformed maximizing $L_O$, although the differences became smaller as sample size was increased.

Heathcote et al. have highlighted the benefits of quantile-based estimation, and this approach is a valuable contribution to response time methodology. Although we take issue with the theoretical grounding of QMLE and function $L_M$, we note that QMLE has more desirable

**Table 2**
**Results of Simulation Studies**

| | Mean | | | RMSE | | |
|---|---|---|---|---|---|---|
| Method | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\tau}$ | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\tau}$ |
| $N = 20$, 4 Order Statistics | | | | | | |
| Maximize $L_M$ | 417.6 | 47.9 | 83.3 | 45.1 | 31.5 | 51.9 |
| Maximize $L_O$ | 420.9 | 41.6 | 78.9 | 47.9 | 36.2 | 54.4 |
| $N = 20$, All Observations | | | | | | |
| Maximize $L_M$ | 410.6 | 47.9 | 90.3 | 37.9 | 25.5 | 42.6 |
| Maximize $L_O$ | 411.6 | 45.0 | 88.4 | 40.7 | 27.9 | 45.2 |
| $N = 80$, 4 Order Statistics | | | | | | |
| Maximize $L_M$ | 405.2 | 49.3 | 94.9 | 23.7 | 18.7 | 28.7 |
| Maximize $L_O$ | 405.7 | 47.1 | 94.0 | 24.9 | 21.6 | 30.1 |
| $N = 80$, All Observations | | | | | | |
| Maximize $L_M$ | 401.7 | 49.4 | 98.7 | 15.8 | 11.7 | 18.5 |
| Maximize $L_O$ | 401.7 | 48.8 | 98.5 | 16.1 | 12.0 | 18.8 |

Note—True values are $\mu = 400$, $\sigma = 50$, $\tau = 100$. RMSE, root mean squared error.

nonasymptotic properties than a valid likelihood approach. The main advantage of maximizing a valid likelihood is that, under regularity conditions, the resulting estimates are guaranteed to be consistent and asymptotically maximally efficient. The ex-Gaussian is regular; therefore, given a set of order statistics, no method is asymptotically better than the likelihood-maximizing method presented here. Asymptotic statistical theory, however, sometimes serves as a poor guide for nonasymptotic behavior. We cannot offer a grounded explanation of why QMLE outperforms maximization of the valid likelihood in small samples. Because there is no known theorem of asymptotic efficiency about QMLE, there may be a downside to using it for other target distributions (or even for other parameter values within the ex-Gaussian). QMLE, at present, is estimation without a safety net.

## REFERENCES

Brown, S., & Heathcote, A. (2003). QMLE: Fast, robust, and efficient estimation of distribution functions based on quantiles. *Behavior Research Methods, Instruments, & Computers*, **35**, 485-492.

Dolan, C. V., van der Maas, H. L. J., & Molenaar, P. C. M. (2002). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods, Instruments, & Computers*, **34**, 304-323.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, **9**, 394-401.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics*. New York: Macmillan.

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

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

Lehmann, E. L. (1991). *Theory of point estimation*. Monterey, CA: Wadsworth.

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

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Magazine*, **50**, 157-175.

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

## NOTE

1. Negative log-likelihood functions were minimized with the simplex algorithm (Nelder & Mead, 1965) in the *R*-statistics package. The ex-Gaussian was reparameterized with parameters $\mu$, $\log(\sigma)$, and $\log(\tau)$ to ensure that $\sigma$ and $\tau$ were positive. The simplex algorithm was started at the true parameter values to speed convergence. Convergence occurred within 500 iterations in all cases.

(Manuscript received April 2, 2003;
revision accepted for publication July 22, 2003.)