Fitting Wald and Ex-Wald Distributions to

Response Time Data: An example using functions for

the S-PLUS package

Andrew Heathcote

School of Behavioural Sciences,

The University of Newcastle, Australia

Address for correspondence:

Andrew Heathcote

School of Behavioural Science, Aviation Building

The University of Newcastle, Callaghan, 2308, NSW, Australia

Email: andrew.heathcote@newcastle.edu.au

Telephone: +61 2 49215952

Abstract

Schwarz (2001, 2002) proposed the Ex-Wald distribution, obtained from the convolution of Wald and Exponential random variables, as a model of simple and go/no-go response time. This paper provides functions for the S-PLUS package that produce maximum likelihood estimates of the parameters of the Ex-Wald, as well as for the shifted Wald and Ex-Gaussian distributions. A Monte Carlo study examined the efficiency and bias of parameter estimates. Results indicated that samples of at least 400 are necessary to obtain adequate estimates of the Ex-Wald, and that for some parameter ranges much larger samples may be required. For shifted Wald estimation smaller samples of around 100 were adequate, at least when fits identified by the software as having an ill-conditioned maximum were excluded. The use of all functions is illustrated using data from Schwarz (2001). The S-PLUS functions, and Schwarz's data, may be downloaded from the Psychonomic Society's web archive, http://www.psychonomic.org/ARCHIVE/.

Schwarz (2001) proposed the Ex-Wald distribution, obtained from the convolution of Wald and Exponential random variables, as a model of Response Time (RT) distribution, particularly in simple RT and go/no-go choice RT tasks. This paper provides functions for the powerful commercially available S-PLUS package that produce maximum likelihood (ML) estimates (Edwards, 1972) of the parameters of the Wald and Ex-Wald distribution, and another distribution based on the Wald, the shifted Wald. Sufficient mathematical detail is also given in this paper to allow implementation in another suitable package. The efficiency and bias of the estimates obtained for the Wald distributions are examined in a Monte Carlo study, and the use of the functions illustrated using data from Schwarz's (2001) participant A.

<u>Ex-Wald Convolution Components</u>

Convolution provides the distribution of the sum of independent random variables, and so is useful in modelling response times that result from the sum of stages with independent and random completion times. Schwarz (2001) modelled decision time with a Wald random variable (W > 0) and non-decision time with an Exponential random variable (X ≥ 0), with RT being the sum of these two times: RT=W $\oplus$ X.

An exponential random variable X has a density $f_E(x \mid t) = \frac{1}{t} e^{-x/t}$ and cumulative density $F_E(x \mid t) = 1 - e^{-x/t}$, where t is the mean (κ_1) of X[1]. The exponential distribution has variance, $\kappa_2 = t^2$ and third central moment (i.e., $\int (x - \kappa_1)^3 f(x) dx$, a measure of skew), $\kappa_3 = 2t^3$. The exponential density (*dexp*) and cumulative density (*pexp*) functions are built-in features of S-PLUS, as is a function that generates exponentially distributed samples (*rexp*). For all of the distributions discussed here density, cumulative density and random

2

functions are provided. Function definitions are given in Table 1, with function names given in italics (e.g., *dwald*) in the body of this paper.

Insert Table 1 about here

The Wald distribution can be derived from a sequential sampling evidence accrual model of the decision process in simple and go/no-go RT, as it is the distribution of first passage times (W) through a level $a > 0$ of a space and time homogenous Wiener diffusion process, with an initial value of zero, drift $m > 0$ and variance $\sigma^2 > 0$ (see Luce, 1986 for a discussion of sequential sampling models). In the RT context, a is a response threshold and m is the mean rate of evidence accrual. Without loss of generality, one of the Wald parameters can be fixed, so we use the conventional normalization $\sigma = 1$ and drop this parameter in the following.

The Wald distribution has a density (*dwald*):

$$f_W(w \mid m, a) = \frac{a}{\sqrt{2\pi w^3}} \exp\left[-\frac{(a - mw)^2}{2w}\right] \qquad w > 0 \qquad (1)$$

The Wald cumulative density function cannot be expressed in closed form, but can be expressed in terms of the cumulative normal distribution function, Φ (built-in function *pnorm* in S-PLUS), which has efficient numerical approximations:

$$F_W(w \mid m, a) = \Phi\left(\frac{mw - a}{\sqrt{w}}\right) + e^{2am}\Phi\left(-\frac{mw + a}{\sqrt{w}}\right) \qquad (2)$$

Schwarz (2001) notes that the second summand in (2) can cause numerical errors when the arguments to the exponential and Φ functions are large. The *pwald* function avoids numerical errors in this case using Derenzo's (1977) approximation to Φ, followed by algebraic simplification of the second summand, as suggested by Schwarz[2]. An S-

3

PLUS function that provides random samples from a different parameterisation of the

Wald distribution is available on the web[3]. The function *rwald* is based on this function,

but uses the parameterisation in Equations 1 and 2.

The Wald distribution has mean a/m, variance, $\kappa_2 = a/m^3$ and third central

moment, $\kappa_3 = 3a/m^5$. It is evident from the moments that a is a scale parameter, and that

m affects all moments. The leading edge of the Wald distribution is fixed at zero (this is

also true of the Ex-Wald), which makes it unrealistic as an exact model of RT, due to the

minimum times required for post-decision processes. In simple RT, for example, Smith

(1995) assumed a substantial shift (relative to overall RT) of around 66ms. This defect

can be corrected by adding a parameter, s, which shifts the leading edge and causes the

mean to become $s + a/m$, but leaves the other central moments unaffected. The Wald

functions in Table 1 include the shift parameter, which is set to zero by default. Equations

1 and 2 can be parameterised to include the shift by substituting $(w - s)$ for w. For the

random function, the shift parameter is added to the output.

The shift parameter provides an alternative way of modelling post-decision time,

as a constant rather than as exponentially distributed random variable in Schwarz's

(2001) Ex-Wald model. A constant for post-decision time is also unrealistic as an exact

model of RT, as post-decision times are variable. Smith (1995) reviewed evidence that

variability in post-decision times had a standard deviation of at most 14ms, and likely

substantially less. For Wolf's (2001) participant A data analysed below this means that

less than 5% of the variability in RT is due to post-decision variability. Hence, a constant

non-decision time may provide a good approximation; it was the approximation adopted

by Smith, and the same assumption is commonly made in fitting sequential sampling

evidence accrual models of choice RT (e.g., Ratcliff & Rouder, 1998).

The Ex-Wald Distribution

Schwarz (2001) derives the density (function *dexw*) of an Ex-Wald random

variable (R) for the case where $m^2 \geq 2/t$. Defining $k = \sqrt{m^2 - 2/t}$:

$$f_{EW}(r \mid m,a,t) = \frac{1}{t}e^{-(r/t - a(m-k))}F_W(r \mid k,a) \tag{3}$$

Schwarz states that this case is of most interest in applications to RT. However, in

practice, it is useful not to have to assume this bound, so the functions given here also

work when $m^2 < 2/t$.

Schwarz (2002) examines the case where $m^2 < 2/t$, which requires complex

function theory. In this case, the Ex-Wald density involves the real part of the complex

function, $w(z) = e^{-z^2}(1 - erf(-iz))$, where $z = x + iy$ is a complex number with real part *x*

and imaginary part *y*, and $erf(z)$ is the complex error function. The real part of $w(z)$

(function *rew*) can be derived from the real and imaginary parts of $erf(z) = u(x,y) + i\,v(x,y)$.

Schwarz favours this method as series approximations exist for $u(x,y)$ and $v(x,y)$ that

converge quickly for all *x*, *y* (Abramowitz & Stegun, 1965, 7.1.29, function *uandv*[4]).

Defining $k' = \sqrt{2/t - m^2}$:

$$f_{EW}(r \mid m,a,t) = \frac{1}{t}e^{-((a-m)^2/(2r))}\mathrm{Re}\left[w\left(k'\sqrt{r/2} + ia/\sqrt{2r}\right)\right]$$

For both cases, the cumulative density (function *pexw*) can be derived from a

general result for convolutions involving exponential random variables (e.g., Ashby &

Townsend, 1980), yielding an expression in terms of the Wald cumulative density and the Ex-Wald density:

$$F_{EW}(r \mid m,a,t) = F_W(r \mid m,a) - t \times f_{EW}(r \mid m,a,t) \qquad (4)$$

Samples from the Ex-Wald distribution are obtained by summing samples from the Wald and Exponential distributions (function *rexw*). The Ex-Wald has mean $t+a/m$, variance, $\kappa_2 = t^2+a/m^3$ and third central moment, $\kappa_3 = 2t^3+a/m^5$

Maximum Likelihood Estimation

The log of the likelihood of a distribution f with parameter vector θ, given a data vector **RT** of length n, is:

$$l(\theta \mid \mathbf{RT}) = \sum_{i=1}^{n} \ln\left(f\left(RT_i \mid \theta\right)\right) \qquad (5)$$

Maximum likelihood parameter estimates are obtained by finding the maximum of Equation 5, or equivalently by minimising minus Equation 5 (functions *negllwald*, *negllswald* and *negllexw* for the Wald, Shifted Wald and Ex-Wald respectively).

For Shifted Wald and Ex-Wald distributions, the minimum cannot be obtained analytically, but instead must be found by a search function. The search function is not guaranteed to find the minimum of the "objective" function (i.e., *negllwald*, *negllswald* or *negllexw*). A number of arguments to the search function must be appropriately set in order to gain good performance. These include bounds on the allowable parameter estimates, a starting point for the search, and ideally functions to calculate the derivatives of the objective function and the relative scales of each parameter.

Fitting Functions

The functions *fitwald* and *fitexw* use the S-PLUS minimization function *nlminb* to obtain maximum likelihood estimates. For example, typing: *fit ← fitwald(rt)* at the command prompt (note ← is typed as a less than sign followed by a minus sign, and should be read as "gets") in S-PLUS creates a list object called *fit* (any name may be used, *fit* will be used generically in the following to indicate the list object created by fitting). The *fit* object contains the estimated parameters for the Wald given a data vector, *rt*, as well as a number of other components. The components can be referenced using the syntax *name$component*. For example, *fit$parameters* is a vector of length 3 containing the estimated parameters in the order (*m,a,s*) for the shifted Wald and (*m,a,t*) for the Ex-Wald.

For both the Wald and Ex-Wald distribution all parameters are bounded below by zero[5]. For the Wald distribution the shift (*s*) parameter must be less than the minimum observed RT (min(**RT**)). A starting point for search can be supplied by the user as the *start* argument to the fitting functions. For example, *fitwald(rt,start= c(0.2,70,100))* sets the starting point to *m*=0.2, *a*=70, and *s*=100. If no *start* argument is supplied the starting point is calculated automatically by the functions *waldstpt* or *exwstpt*. Both functions estimate one parameter using a heuristic then solve for the other two parameters using the first two sample moments. Starting estimates based on the first three sample moments were found to perform poorly due to the high sampling variance of the third moment. For the Wald distribution the shift parameter is estimated by the heuristic $s=p \times \min(\mathbf{RT})$, with *p*=0.9 by default. For the Ex-Wald distribution the exponential parameter is estimated by $t=p \times SD(\mathbf{RT})$, where SD is the sample standard deviation and *p*=0.5 by default. The

default values of *p* may be altered by specifying a single number as the start argument. For example, *fitwald(rt, start=waldstpt(rt=rt,p=0.95))* sets *p* to 0.95.

The *nlminb* search function is faster and converges more reliably if the gradient (first derivatives) vector and Hessian (second derivatives) matrix of the likelihood function are specified. Approximate standard errors and correlations among the parameter estimates can be obtained by inverting the Hessian (Bates & Watts, 1988). The gradient vector and Hessian matrix for minus the log likelihood are:

$$-\frac{\partial l(\mathbf{RT}\mid\boldsymbol{\theta})}{\partial\theta_j}=-\sum_{i=1}^{n}\frac{\partial\ln(f(RT_i\mid\boldsymbol{\theta}))}{\partial\theta_j} \qquad (6)$$

$$-\frac{\partial^2 l(\mathbf{RT}\mid\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}=-\sum_{i=1}^{n}\frac{\partial^2\ln(f(RT_i\mid\boldsymbol{\theta}))}{\partial\theta_j\partial\theta_k} \qquad (7)$$

For the Shifted Wald distribution the first and second derivatives of $\ln(f_W)$ have straightforward analytic solutions, which are given in Table 2. The corresponding results for the Wald without a shift are obtained by setting *s*=0. Functions *negllwald.gh* and *negllswald.gh* calculate the gradient and Hessian for the Wald and Shifted Wald distribution respectively using these equations. Function *fitwald* calls these functions to aid fitting, and they are also used by function *ser.nlminb*, described in more detail below, to calculate approximate parameter standard errors and correlations. For the Ex-Wald distribution, the gradient and Hessian are not so tractable. The *nlminb* function can still fit the Ex-Wald distribution using numerical finite differences, but it tends to be slower.

```
-------------------------------
Insert Table 2 about here
-------------------------------
```

Several conditions are used by *nlminb* to terminate search, as reported in *fit$message*. Normal termination occurs when the objective function decreases by only a

small proportion, by default max(10^{-10}, .Machine\$double.eps$^{2/3}$), where

.Machine\$double.eps is the double precision floating point tolerance of the computer on

which S-PLUS is running. Achieving this criterion is reported as "Relative Function

Convergence" by S-PLUS. Convergence can also occur when parameter estimates change

by only a small amount between iterations of the search (.Machine\$double.eps$^{1/2}$ by

default), called "X Convergence". Abnormal termination, indicating a failure of the

minimization, occurs for a number of reasons, including exceeding a set number of

objective function evaluations (200 by default) or search iterations (150 by default), or

because the size of the finite differences used during search has become very small

(100×.Machine\$double.eps by default), called "False Convergence".

In the Monte Carlo study, False Convergence was a common problem with fits of

the Ex-Wald, even when search was started at the true parameter values. Greatly

improved performance in this regard was obtained by setting the relative scales of the

different parameters to the inverse of the starting point estimate using the *scale* argument

of *nlminb*. For example, for starting point (0.2, 70, 100), *scale=(1/0.2,1/70, 1/100)*. The

fitexw function uses this setting of *scale* by default, but this can be turned off using the

argument *scaleit=F*. Scale does not need to be set for *fitwald* as it is calculated by default

from the Hessian. More rarely, abnormal termination occurs because the function

evaluation and search iteration limits are reached. This problem was avoided in the

Monte Carlo study by doubling their default values in *fitexw*. All default values for

nlminb can be changed using the S-PLUS function *nlminb.control*.

The parameter variance-covariance matrix for the Ex-Wald, and hence

approximate parameter standard errors and correlations, can be obtained by numerical

finite difference methods, using the function *vcov.nlminb* of the MASS library (Venables & Riply, 1999, the library is distributed with S-PLUS but must be loaded using the command *library(MASS)*). The function *vcov.nlminb* can also make use of analytic derivatives if they are supplied to *nlminb*. For convenience, all fitting functions call an augmented version of *vcov.nlminb*, called *ser.nlminb*, which outputs parameter correlations and standard errors and protects against numerical errors. It creates *fit$se*, a vector of length three containing the standard error estimates, in the same order as the *fit$parameters*. Similarly it creates, *fit$cor*, a vector of length three that contains the parameter correlations, in the order r(m,a), r(m,s), r(a,s) for the Wald and r(m,a), r(m,t), r(a,t) for the Ex-Wald. For some fits, *ser.nlminb* sets standard errors and correlations to missing (*fit$se* and *fit$cor* = (NA, NA, NA), where NA is the S-PLUS missing value symbol), either because the Hessian matrix cannot be inverted, because it is singular, or because variance estimates are negative.

Monte Carlo Study

A Monte Carlo study was performed with two aims: to refine the implementation of the S-PLUS functions so that they are numerically robust for small samples, and to determine the bias and efficiency of parameter estimates. A variety of parameters values were selected which span the range usually encountered in applications and also theoretically important cases, particularly for the Ex-Wald. Results for the shifted Wald are described first, followed by results for the Ex-Wald.

Figure 1 shows the three Shifted Wald distributions investigated in the Monte Carlo study, and Table 3 supplies the corresponding parameters. Parameters were chosen so that the distributions were equated on location (mean=1000) and scale (SD=100), but

differed in skew. One thousand samples of sizes 40 to 120 were fit for each distribution.

All shifted Wald fits in the Monte Carlo study used the analytic gradients and Hessian.

Without them, some fits did not converge normally and estimates were sensitive to

starting points, whereas in the Monte Carlo study all fits converged normally and

produced identical parameter estimates for starting points generated by *waldstpt* using *p*

= 0.9, 0.95 and 0.99.

Insert Table 3 about here

As shown in Figure 1, for the Wald distributions examined here, skew was mainly

determined by the left tail, with right tails having very similar shapes for all distributions.

The longer left tail of distribution one caused the greatest problems with parameter

estimates. Table 3 shows that Distribution 1 produced the greatest percentage of irregular

fits (i.e., fits where the Hessian was either singular or produced negative parameter

variance estimates). Irregular fits overestimated the *m* and *a* parameters, and

underestimated the *s* parameter. The underestimates would have been much worse if *s*

were not bounded below by zero.

Insert Figure 1 about here

Figure 2 shows the bias and efficiency of the parameter estimates as a function of

sample size and distribution. Bias was measured by the mean deviation of the estimates

from the true parameter values. Efficiency was measured by the standard deviation (SD)

of the estimates. Bias decreases and efficiency increases with sample size, and as the

distributions became more skewed. Much better estimates were obtained by removing

irregular fits, resulting in bias that was relatively minor even for the smallest sample

sizes. However, parameter estimates were still quite variable for smaller sample sizes,

particularly for the *s* and *a* parameters.

Insert Figure 2 about here

Figure 1 also shows the five Ex-Wald distributions investigated in the Monte

Carlo study. Parameters, which are given in Table 4, were chosen to examine the effect of

varying k^2: Schwarz's (2001) real solution applies for $k^2 \geq 0$, whereas solutions are

complex for $k^2 < 0$. Distribution 2 was chosen to be a reference with $k^2 = 0$. Negative values

of k^2 were produced by decreasing *m* (Distribution 1) or *t* (Distribution 5). Positive values

of k^2 were produced by increasing *m* (Distribution 3) or *t* (Distribution 4). Where k^2 was

manipulated by changing *m* (Distributions 1 and 3), *a* was also altered to maintain a

constant mean. For these parameters, the second and third central moments decreased as

k^2 increased when manipulated via *m*, whereas they increased when k^2 was manipulated

via *t*. Skew, as indicated by Fisher's skewness measure, $\gamma = \mu_3 / \mu_2^{3/2}$, always increased

with k^2.

Insert Table 4 about here

One thousand fits were performed at each sample size in the Monte Carlo study of

the Ex-Wald distribution. These fits produced bimodal and in some cases multi-modal

parameter distributions. Initially parameter sampling distributions were obtained for n =

120 to 400. In order to check if sampling distributions become uni-modal for large

sample sizes, sampling distributions were obtained for n=10000. For Distributions 3 and

4, where the true value of $k^2 > 0$, estimates of *t* were bimodal for n=400, but uni-modal and

relatively symmetric for n=10000. Note that when the fitting algorithm was restricted to

searching parameter combinations where $k^2>0$ the sampling distributions for n=400 were very similar to the unrestricted case, so bimodality was not due to allowing fits to sample the complex case.

For distributions where the true value of $k^2\leq0$, sampling distributions for m and t were not uni-modal even for n=10000. Restricting the fitting algorithm to search for parameter combinations where $k^2<0$ did not help matters for Distributions 1 and 5, where the true value of $k^2<0$, producing similarly shaped sampling distributions to the unrestricted case. Sampling distributions for the three cases were $k^2\leq0$ were obtained for n=50000. For Distribution 2, where the true value of $k^2=0$, unimodality was not obtained even for n=50000. For Distribution 1, sampling distributions were fairly unimodal for n=50000, although a long right tail remained for the m parameter. For Distribution 5 sampling distributions for m and t were trimodal for n=50000, with a sharp third peak appearing at the true parameter value.

Fitting was also less tractable for the Ex-Wald than the Shifted Wald. Table 4 shows that in many cases fits were irregular. In contrast to the Shifted Wald, excluding these fits did not reduce bias or increase efficiency. For Distributions 1 and 2, most fits were well-behaved for n=400. For the other distributions, a substantial proportion of fits had problems for n=400. These problems largely disappeared for Distributions 3 and 4, where the true value of $k^2>0$, when n=10000. However, problems reappeared for Distribution 2 when n=10000, and both Distributions 2 and 5 still had a substantial proportion of irregular fits even for n=50000. Note that all Ex-Wald fits reported here used the true parameter values as the starting point for search. When starting values were

generated using the default heuristic in *exwstpt* similar results were obtained, but with a

slightly increased percentage of irregular fits.

Despite irregularity in fits and multi-modality in parameter estimate distributions,

Figure 3 shows that bias decreased and efficiency increased with sample size for all

distributions. However, overall performance was poor relative to results for the Shifted

Wald, and for the *t* parameter in particular estimates remained variable even for the very

large samples. Distribution 2 showed the greatest bias, and for smaller samples the bias

for the *t* parameter was in the opposite direction to the other distributions. Distributions 1

and 4 produced particularly variable estimates of *t* for smaller samples. Although

variability decreased quickly with sample size for Distribution 4, the decrease was much

slower for Distribution 1.

Insert Figure 3 about here

The results of the Monte Carlo study indicate that maximum likelihood

estimation in finite samples from the shifted Wald and Ex-Wald distributions does not

always produce the uni-modal normally distributed parameter estimate distributions

observed in regular cases, particularly in smaller samples and particularly for the Ex-

Wald. It is important to note that multi-modal parameter estimate distributions were not

due to failures of search because of convergence to a local minimum. The same

parameter estimates were obtained for a range of starting points, and profiles of the

likelihood function (e.g., Figure 6) for particular samples had only one minimum. It was

also not due to failures in the numerical approximations used here. Benchmarking against

numerical integration found these approximations to be accurate, and slightly more

accurate in some cases than Schwarz's (2001) approximation (see Footnote 3).

Multi-modality in likelihood estimates is not unexpected theoretically; it occurs

when distribution functions do not fulfil regularity conditions (Edwards, 1972).

Evidently, likelihood estimation of the shifted Wald, and particularly the Ex-Wald, is

irregular. However, Figures 3 and 4 display consistency, in that both the bias and

standard deviation of parameter estimates decreased with sample size. Hence, estimation

based on likelihood remains useful in practice, as long as sample sizes are sufficiently

large.

Example Application

The application of the fitting functions is illustrated in this section using the data

from participant A shown in Figure 3 of Schwarz (2001), and reproduced in Figure 4

here. Schwarz also examined the fit of the Ex-Gaussian distribution to this data. As

shown in Table 1, S-PLUS functions are supplied to fit the Ex-Gaussian, using similar

notational conventions to the Wald and Ex-Wald functions. Analytic gradients are

supplied by *negllexggrad*. Maximum likelihood estimation is regular for the Ex-Gaussian

and gradients were sufficient for good performance. Heathcote's (1996) Ex-Gaussian

estimation program, RTSYS, can use analytic Hessians in fitting and performed similarly

for this data. The function *exgstpt* estimates starting points using heuristics given in

Heathcote, and usually performs well with the default values.

```
-------------------------------
Insert Figure 4 about here
-------------------------------
```

Schwarz's (2001) data comes from a go/no-go paradigm, so the decision process

might reasonably be modelled by a one barrier sequential sampling (Wiener) process, and

hence a Wald distribution. Table 5 contains the parameter estimates for the Ex-Wald

obtained from *fitexw*, and the estimated distribution functions are plotted as solid lines in

Figure 5. As predicted by the sequential sampling model, manipulation of the probability

of a go trial (p=0.5 or 0.75) mainly effects the *a* (decision criterion) parameter, whereas

manipulation of numerical distance (d=1 or 4) mainly effects the *m* (strength of evidence)

parameter.

Insert Table 5 about here

The exponential component of the Ex-Wald models non-decision times, and

corresponding parameter estimates (*t*) do not vary much across conditions, as expected.

Estimates of the *t* parameter vary slightly from those reported by Schwarz (2001),

especially in the p=.75, d=1 condition, where the fit at the minimum was so ill

conditioned that the Hessian could not be inverted. Some small differences are to be

expected due to variation in the details of the numerical approximations and search

algorithms used. Such numerical differences are magnified when the minimum of the

objective function is ill defined, as appears to be the case for the p=.75, d=1 condition.

Note also that the parameters found by *fitexw* are in the complex case (i.e., $m^2<2/t$)

examined by Schwarz (2002).

Insert Figure 5 about here

In order to explore the p=.75, d=1 case further, Figure 5 plots profiles of the

objective function (*negllexw*) in the region of the solution by holding two parameters

fixed at their estimated values and varying the third. The plots can be interpreted like a

plot of sum-squared error in least squares regression, with smaller values of minus log-

likelihood, which is sometimes called the "deviance", indicating a better fit. Note that the

deviance functions are smooth as changes in *m* and *t* cause a change from the real to

complex solutions. The minimum is sharply defined as a function of *m* and fairly well defined for *a*, and in both cases the minium is approximately quadratic in shape, as assumed by the standard error and correlation estimates based on the Hessian. However, the minimum as a function of *t* is very poorly defined, and clearly non-quadratic as deviance rises dramatically as *t* approaches zero. Similar, although less marked, behaviour for *t* was found in the other conditions, and in these cases the estimates of *t* are within two standard errors of zero, indicating that there is no strong evidence for an exponential component.

Table 6 contains the results of fits of the Shifted Wald using *fitwald*. The deviance values indicate that fit is very slightly better than for the Ex-Wald and the Hessian is well behaved for all conditions. However, standard errors are quite large, especially for the estimated shift (*s*) parameters, which fall within one standard error of zero. Hence, there is no evidence that the shift parameter is non-zero. Figure 8 plots the best fitting Shifted Wald densities as dashed lines; in most cases they are indistinguishable from the Ex-Wald densities, except for a very slight negative shift on the left of the peak.

Insert Table 6 about here

As standard errors for both the Shifted Wald and Ex-Wald models provide little evidence for a reliable non-decision (*s* or *t*) component, the simpler two-parameter Wald distribution, with shift fixed at zero, was fit. The resulting deviance values (Table 7) are only slightly larger than for the three-parameter distributions. As the Wald distribution is "nested" (i.e., is a special case of) the Shifted Wald and Ex-Wald distributions, the reliability of the reduction in deviance can be determined using a likelihood ratio test. Annotating the deviance for the nested (Wald) model as d_n, and the deviance of the full

(Shifted Wald or Ex-Wald) model as d_f, the likelihood ratio statistic is $\lambda=2(d_n-d_f)$. The λ statistic is approximately distributed as χ^2 with degrees of freedom equal to the difference in the number of parameters for the full and nested model: df=3–2=1.

Insert Table 7 about here

Table 7 shows that no reliable differences in deviance were found in any condition[6]. Hence, the two-parameter Wald distribution provides the most parsimonious model of this data. Further, its parameter estimates follow the pattern predicted by the sequential sampling model and have much smaller standard errors than in either of the three-parameter models. The estimated Wald densities are plotted in Figure 4 as solid lines. They are so close to the estimated Ex-Wald densities, which are also plotted as solid lines, that no difference is discernable. It appears that the small estimated vales of t have a negligible effect on the shape of the Ex-Wald.

The results of χ^2 tests based on the histograms in Figure 4 also indicate little difference in fit between the three distributions. Summing across conditions for the Ex-Wald $\chi^2(29) = 45.5$, $p = 0.026$, for the shifted Wald $\chi^2(30) = 46.6$, $p = 0.03$, and for the Wald, $\chi^2(29) = 45.5$, $p = 0.03$. The χ^2 statistic for individual fits was calculated by all fitting functions using a similar algorithm (*chisq*) to Schwarz (2001) and stored in *fit$chisq*. The Ex-Wald results are larger than reported by Schwarz (presumably due to numerical estimation differences and possibly differences in the aggregation of bins with an expected frequency of less than 5) and achieve significance at the 0.05 level. Note that interpretation of these χ^2 statistics as measures of absolute fit is not recommended, both because their sampling distribution is often very poorly approximated by the χ^2

18

distribution, and because they can vary depending on the choice of bin location and width (25ms in this case, following Schwarz).

Table 8 contains the results of fitting the data with the Ex-Gaussian distribution. In agreement with Schwarz's (2001) results, the Ex-Gaussian clearly fits worse than the Wald distributions, both in terms of deviance, and the total histogram based χ^2 over conditions: $\chi^2(31) = 57.9$, $p = .002$. As shown in Figure 4, which plots the estimated Ex-Gaussian densities as dotted lines, they have heavier left tails and sharper peaks than the Wald distributions, but similar right tails. The Ex-Gaussian parameter estimates were, however, quite sharply defined, as indicated by small standard errors.

Insert Table 8 about here

The Ex-Gaussian parameter estimates were less correlated than was the case for any of the Wald distributions. Substantial correlation among parameter estimates can be problematic for fitting, but does not necessarily indicate that the model is inadequate. High correlations may arise because of the parameterisation adopted ("parameter effects curvature", see Bates & Watts, 1988) and alternative parameterisations may reduce correlations. However, limited experimentation with different parameterisations of the Wald distribution, including the mean and dispersion parameterisation adopted by the *invgauss* distribution functions described in Footnote 4, did not result in any improvement, so the sequential sampling parameterisation was retained because of interpretability.

Discussion

This paper provides S-PLUS functions to simulate and fit the Wald, Ex-Wald, shifted Wald and Ex-Gaussian distributions. The Ex-Wald and shifted Wald functions were extensively tested in a Monte Carlo simulation, which allowed automatic starting point estimates and details of the fitting algorithm to be fine-tuned in order to achieve robust numerical performance even with relatively small samples. The Ex-Wald functions cover both the restricted parameter case considered by Schwarz (2001) and the complementary case considered by Schwarz (2002), and so provide a complete estimation solution.

The Monte Carlo study revealed that estimation of the shifted Wald and Ex-Wald distributions can be irregular, in the sense that parameter estimate distributions can be multi-modal, particularly for smaller sample sizes and particularly for the Ex-Wald distribution. However, for both distributions, bias decreased and efficiency increased with sample size, indicating that estimation was consistent, at least for the parameter values examined. Hence, likelihood estimation for these distributions remains useful despite the irregularity. However, caution should be exercised with smaller samples and even with larger samples when the complex case of the Ex-Wald applies. The S-PLUS software can then be used to investigate potentially problematic cases with the same Monte Carlo methods reported here.

The Ex-Wald distribution required larger samples to obtain the same level of bias and efficiency as the shifted Wald. It is difficult to provide global recommendations for the minimum sample size required in applications, as estimation performance varies depending on the true parameter values. However, overall the Monte Carlo results

suggest a minimum sample size of at least 100 for the shifted Wald and 400 or more for the Ex-Wald. For samples of the minimum or greater size, useful Hessian based parameter standard error and correlation estimates are usually obtained for the shifted Wald, and for the Ex-Wald, al least in the real case. However, caution should be exercised in interpreting Hessian based estimates, as they assume estimation is regular. Likelihood ratio testing, as described in the example application section, can be used to provide more reliable inference.

The example application confirmed Schwarz's (2001) finding that Wald distributions provide a clearly superior model of go/no-go RT data compared to the Ex-Gaussian. In particular, the Wald distributions both fit better and provided theoretically meaningful parameter estimates. The shifted Wald fit this data at least as well as the Ex-Wald, indicating that it can also provide a viable approximation to simple and go/no-go RT distribution. The Ex-Wald and shifted Wald represent two extreme approaches to dealing with non-decision time; the shifted Wald assumes non-decision time is always greater than zero but is not variable, whereas the Ex-Wald assumes that non-decision time is usually close to zero, but is variable.

Both approaches are approximations; in reality non-decision time must both have a lower bound greater than zero and be variable. Modelling both aspects of non-decision time would at the very least require one extra parameter, such as the shifted rectangular distribution used by Ratcliff and Tuerlinckx (2002). The greater complexity of this approach increases estimation difficulties, so both the shifted Wald and Ex-Wald remain useful as approximations that at least in some cases are likely sufficiently accurate.

Schwarz (2001) improved the efficiency of estimation by making theoretically motivated assumptions about constancy of parameters across conditions. This represents a superior approach compared to separate estimation of distributions for each condition, and underlines the advantages of using RT distribution models that are theoretically motivated. Likely the greater power of this approach would have led to non-decision time estimates significantly greater than zero, in contrast to the results obtained in the example application here. This application was given mainly to illustrate use of the S-PLUS functions, and in practice Schwarz's approach should be used where possible.

In order to fit models that fix parameters across conditions the functions provided here must be extended. The powerful S-PLUS environment makes this extension relatively straightforward, but requires some knowledge of the S-PLUS language. In particular a new objective function must be created with the appropriate parameterisation and ability to identify data from different conditions. The results of estimation on individual conditions can be used to obtain starting points for the estimation of such models, facilitating search, which must be performed in a higher dimensional parameter space.

References

Ashby, F. G. & Townsend, J. T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influence revisited. Journal of Mathematical Psychology, 21, 93-123.

Bates, D. M., & Watts, D. G. (1988). Nonlinear regression analysis and its applications. New York: Wiley.

Derenzo, S. E. (1977). Approximations for hand calculators using small integer coefficients. Mathematics of Computation, 31, 214-225.

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

Heathcote, A. (1996). RTSYS: A computer program for analysing response time data. Behaviour 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.

Luce, R. D. (1986). Response Times. New York: Oxford University Press.

Ratcliff, R. & Rouder, J. F. (1998). Modeling response times for two-choice decisions. Psychological Science, 9, 347-356.

Ratcliff, R. & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. Psychonomic Bulletin & Review, 9, 438-481.

Schwarz, W. (2001). The Ex-Wald distribution as a descriptive model of response times. Behaviour Research Methods, Instruments & Computers, 33, 457-469.

Schwarz, W. (2002). On the convolution of Inverse Gaussian and exponential random variables, Communications in Statistics: Theory and Methods, 31, 2113-2121.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. Psychological Review, 102, 567-593.

Venables, W. N. & Riply, B. D. (1999). Modern Applied Statistics with S-PLUS, 3rd Edition, Springer-Verlag, New York.

Materials available from the Psychonomic Society web archive

The following materials associated with this article are retrievable from the Psychonomics Society web archive, http://www.psychonomic.org/ARCHIVE/.

FILE: rtfit.scc

DESCRIPTION: This script file contains the S-PLUS functions described in Heathcote (2004), as a 17k ASCII file. Once the script file is saved its functions must be compiled (e.g., source("rtfit.ssc")). These functions work with both S-PLUS 2000 and S-PLUS 6, with the former having been tested more extensively, using both an NT4 and a Windows 2000 system.

FILE: schwarz-A.dat

DESCRIPTION: This file contains the data from Schwarz (2001) participant A as a 24K, three column, tab-delimited, ASCII file. Column names are on the first line, and each following line represents the results for one trial. The first column classifies the go probability for each trial (0.5 or 0.75), the second column classifies numeric distance (1 or 4) and the final column contains the trial RT (ms).

To access the above files or links, please visit http://www.psychonomic.org/ARCHIVE/ and search for this article using the journal (*Behavior Research Methods, Instruments, and Computers*), the first author's name (*Heathcote*) and the publication year (*2004*).

Note that these files, and possible future extensions (e.g., extra distribution functions, updates for new versions of S-PLUS etc.), are also available on the author's web site: http://www.newcastle.edu.au/school/behav-sci/ncl/

Acknowledgements

Tables

Table 1. Function definitions.

Density Functions	Comment
`dwald(w,m,a,s=0)`	`Shifted Wald, w is RT`
`dexw(r,m,a,t)`	`Ex-Wald, r is RT`
`dexg(x,m,s,t)`	`Ex-Gaussian, t is RT, m=μ, s=σ and t=τ`
Cumulative Density Functions	**Comment**
`pwald(w,m,a,s=0)`	`p(W ≤ w) for Shifted Wald`
`pexw(r,m,a,t)`	`p(R ≤ r) for Ex-Wald`
`pexg(x,m,s,t)`	`p(T ≤ t) for Ex-Gaussian`
Random Functions	**Comment (n is number of samples)**
`rwald(n,m,a,s=0)`	`Shifted Wald`
`rexw(n,m,a,t)`	`Ex-Wald`
`rexg(n,m,s,t)`	`Ex-Gaussian`
Objective Functions	**Comment (minus loglikelihood, x is RT)**
`negllswald(p,x)`	`Shifted Wald, p = (m,a,s)`
`negllwald(p,x)`	`Wald with s = 0, p = (m,a)`
`negllexw(p,x)`	`Ex-Wald, p = (m,a,t)`
`negllexg(p,x)`	`Ex-Gaussian, p = (m,s,t)`
Objective Derivatives	**Comment**
`negllswald.gh(p,x)`	`Shifted Wald Gradient and Hessian`
`negllwald.gh(p,x)`	`Gradient and Hessian for Wald with s = 0`
`negllexggrad(p,x)`	`Gradient for Ex-Gaussian`
Heuristic Starting Points	**Comment (p=heuristic parameter)**
`waldstpt(x,shift=T,p=0.9)`	`Wald/Shifted Wald`
`exwstpt(x,p = 0.5)`	`Ex-Wald`
`exgstartpt(rt, p = 0.8)`	`Ex-Gaussian`
Fitting Functions	**Comment (rt is a response time vector)**
`fitwald(rt,shift=T,p=0.9,` ` start=waldstpt(rt,shift,p))`	`Wald (shift=F)and Shifted Wald (shift=T)`
`fitexw(rt,p=.5,start=` `    exwstpt(rt),scaleit=T)`	`Ex-Wald, scaleit=T turns on scaling =` `1/start`
`fitexg(rt,p=0.8,start=` `    exgstartpt(rt,p))`	`Ex-Gaussian`
Auxiliary Functions	**Comment**

`uandv(x,y,firstblock=20,`
`block=0,tol=.Machine$double.`
`eps^(2/3),maxseries=20)`
	`Used by dexw, series approximation to the`
	`real (u) and imaginary (v) parts of`
	`complex error function, erf(x + iy).`

`rew(x,y,...)`	`Used by dexw, real part of` `w(z) = exp(-z^2)[1-erf(-iz)]`

`ser.nlminb(object,tol =`
`1,scale=object$scale,eps=0.0`
`01,eps0=1)`
	`Modified vcov.nlminb, object = fit object`
	`produced by nlminb, called within fitting`
	`functions.`

`chisq(rt,p,bin=25, minn = 5,`
`dist = "wald")`
	`Chis square, bin=bin width, minn=minimum`
	`expected observations in a bin, dist=`
	`"wald", "exw" (Ex-Wald) and "exg" (Ex-`
	`Gaussian), called in fitting functions.`

Table 2. First and second derivatives of $\ln(f_W)$.

$$\partial \ln(f_W)/\partial m = a - m(w - s)$$

$$\partial \ln(f_W)/\partial a = m + 1/a - a/(w - s)$$

$$\partial \ln(f_W)/\partial s = \tfrac{3}{2}(w - s)^{-1} - \tfrac{a^2}{2}(w - s)^{-2} + m^2/2$$

$$\partial^2 \ln(f_w)/\partial m^2 = s - w$$

$$\partial^2 \ln(f_w)/(\partial m \partial a) = 1$$

$$\partial^2 \ln(f_w)/(\partial m \partial s) = m$$

$$\partial^2 \ln(f_w)/\partial a^2 = -1/a^2 - 1/(w - s)$$

$$\partial^2 \ln(f_w)/(\partial a \partial s) = -a/(w - s)^2$$

$$\partial^2 \ln(f_W)/\partial s^2 = \left(\tfrac{3}{2} - a^2(w - s)^{-1}\right)/(w - s)^2$$

Table 3. Shifted Wald parameters and percentages of irregular fits.

Distribution	1	2	3
m	0.18856	0.16262	0.14142
a	70.711	44.721	28.284
s	625	625	625
Mean	1000	1000	1000
SD	102.7	102	100
$\mu_3^{1/3}$	66.6929	73.2623	79.3711
Fisher Skew	0.2739	0.3708	0.5000
Irregular fits, n = 40	12.3	5.0	1.4
Irregular fits, n = 60	9.4	2.4	0.0
Irregular fits, n = 80	6.4	0.6	0.0
Irregular fits, n = 100	4.6	0.4	0.0
Irregular fits, n = 120	2.4	0.1	0.0

Fitting Wald Distributions

Table 4. Ex-Wald parameters and percentages of irregular fits.

Distribution	1	2	3	4	5
m	0.15	0.2	0.25	0.2	0.2
a	52.5	70.0	87.5	70.0	70.0
t	50.0	50.0	50.0	115.0	35.0
k^2	-0.0175	0	0.0225	0.0226	-0.0171
Mean	400	400	400	465	385
SD	134.371	106.066	90	148.24	99.875
$\mu_3^{1/3}$	98.006	77.681	69.768	148.284	67.276
Fisher Skew	0.06292	0.07111	0.08171	0.08216	0.06740
Irregular fits, n = 120	10.4	74.5	26.2	41.1	20.4
Irregular fits, n = 160	42.0	20.5	18.6	17.9	11.7
Irregular fits, n = 200	40.3	1.5	17.5	4.9	10.6
Irregular fits, n = 300	4.9	0.3	17.7	13.0	15.4
Irregular fits, n = 400	0.4	0.2	18.2	11.1	16.2
Irregular fits, n = 10000	0.1	46.7	0.5	0.0	6.1
Irregular fits, n = 50000	4.3	10.1	-	-	15.2

Table 5. Ex-Wald parameter estimates, standard errors (se), correlations (r) and minus loglikelihood (-*l*) for Schwarz's (2001) participant A.

Condition						Ex-Wald				
	m	*a*	*t*	se(*m*)	se(*a*)	se(*t*)	r(*m,a*)	r(*m,t*)	r(*a,t*)	- *l*
p=.5, d=1	.321	108	22.5	.032	6.39	16.3	.94	.91	.73	2206.76
p=.5, d=4	.359	112	19.3	.040	7.57	15.5	.96	.93	.80	2146.22
p=.75, d=1	.316	96	19.3	NA	NA	NA	NA	NA	NA	3279.50
p=.75, d=4	.349	98	20.8	.029	5.08	10.8	.95	.91	.76	3213.06

Table 6. Shifted Wald parameter estimates, standard errors (se), correlations (r) and minus loglikelihood (-*l*) for Schwarz's (2001) participant A.

Condition						Shifted Wald				
	m	*a*	*s*	se(*m*)	se(*a*)	se(*s*)	r(*m,a*)	r(*m,s*)	r(*a,s*)	- *l*
p=.5, d=1	.284	88	49.3	.036	30.2	67.6	.981	-.957	-.99	2206.55
p=.5, d=4	.330	100	25.5	.043	35.7	69.6	.982	-.960	-.99	2146.15
p=.75, d=1	.307	99	0.2	.032	28.5	59.6	.981	-.959	-.99	3279.51
P=.75, d=4	.313	85	30.2	.032	23.5	48.0	.981	-.957	-.99	3212.88

Table 7. Wald parameter estimates, standard errors (se), correlations (r) and minus loglikelihood (-*l*) for Schwarz's (2001) participant A, and likelihood ratio (λ) tests (p = probability under the null hypothesis of no difference) comparing the Wald and Ex-Wald (λ(ew)) and the Wald and Shifted Wald (λ(sw)).

Condition						Wald				
	m	*a*	se(*m*)	se(*a*)	r(*m,a*)	- *l*	λ(ew)	p	λ(sw)	p
p=.5, d=1	.308	111	.011	3.9	.972	2206.74	0	1	0.38	.54
p=.5, d=4	.344	114	.012	4.0	.975	2146.21	0	1	0.11	.74
p=.75, d=1	.307	100	.009	2.9	.969	3279.51	.02	.89	0	1
p=.75, d=4	.332	100	.010	2.9	.971	3213.03	0	1	0.31	.58

Table 8. Ex-Gaussian parameter estimates, standard errors (se), correlations (r) and minus loglikelihood (-*l*) for Schwarz's (2001) participant A.

Condition							Ex-Gaussian			
	m	*s*	*t*	se(*m*)	se(*s*)	se(*t*)	r(*m,s*)	r(*m,s*)	r(*s,t*)	- *l*
p=.5, d=1	318	46.5	41.6	6.62	4.17	6.54	0.76	-0.89	-0.77	2577.66
p=.5, d=4	294	39.3	36.4	5.31	3.39	5.26	0.73	-0.87	-0.74	2515.87
p=.75, d=1	285	44.1	39.2	4.73	2.95	4.66	0.71	-0.87	-0.72	3833.99
p=.75, d=4	266	39.4	35.4	4.12	2.56	4.05	0.70	-0.86	-0.71	3768.00

Figure Captions

Figure 1. (a) Shifted Wald and (b) Ex-Wald distributions used in the Monte Carlo study. For (a), mean and standard deviation are constant and skew increases with the number used to identify the distribution. For (b), distributions 3 and 4 are in the real case studied by Wolfe (2001) and distributions 1 and 5 are in the complex case studied by Wolfe (2002), with distribution 2 on the boundary.

Figure 2. Shifted Wald parameter estimate mean bias (estimated – true) and efficiency (parameter standard deviations, SD) for all fits and for fits with ill-conditioned Hessian estimates censored. The abscissa is sample size (n), and each panel represents results for the *m* (mean rate of evidence accrual), *a* (response threshold) and *s* (shift or minimum) parameters respectively.

Figure 3. Ex-Wald parameter estimate mean bias (estimated – true) and efficiency (parameter standard deviations, SD). The abscissa is sample size (n), on a $\log_{10}$ scale. Each panel represents results for *m* (mean rate of evidence accrual), *a* (response threshold) and *t* (exponential mean) parameters respectively.

Figure 4. Probability histograms of Schwarz's (2001) data for participant A, with fitted probability density functions for the Ex-Wald and Wald (solid lines), Shifted Wald (dashed lines) and the Ex-Gaussian (dotted lines). Each panel provides results for one of four experimental conditions, formed by the factorial combination of the probability of a go response (p_go = .5 or .75) and numerical distance (d = 1 or 4).

Figure 5. Objective function (minus log likelihood) profiles around the fitted Ex-Wald parameter values for Schwarz's (2001) participant A in the p_go=.75, d=1 condition. For each panel the abscissa label indicates the parameter varied (*m*, the mean rate of evidence

accrual, *a*, the response threshold, and *t*, the exponential mean) while holding the other parameters at their fitted values.
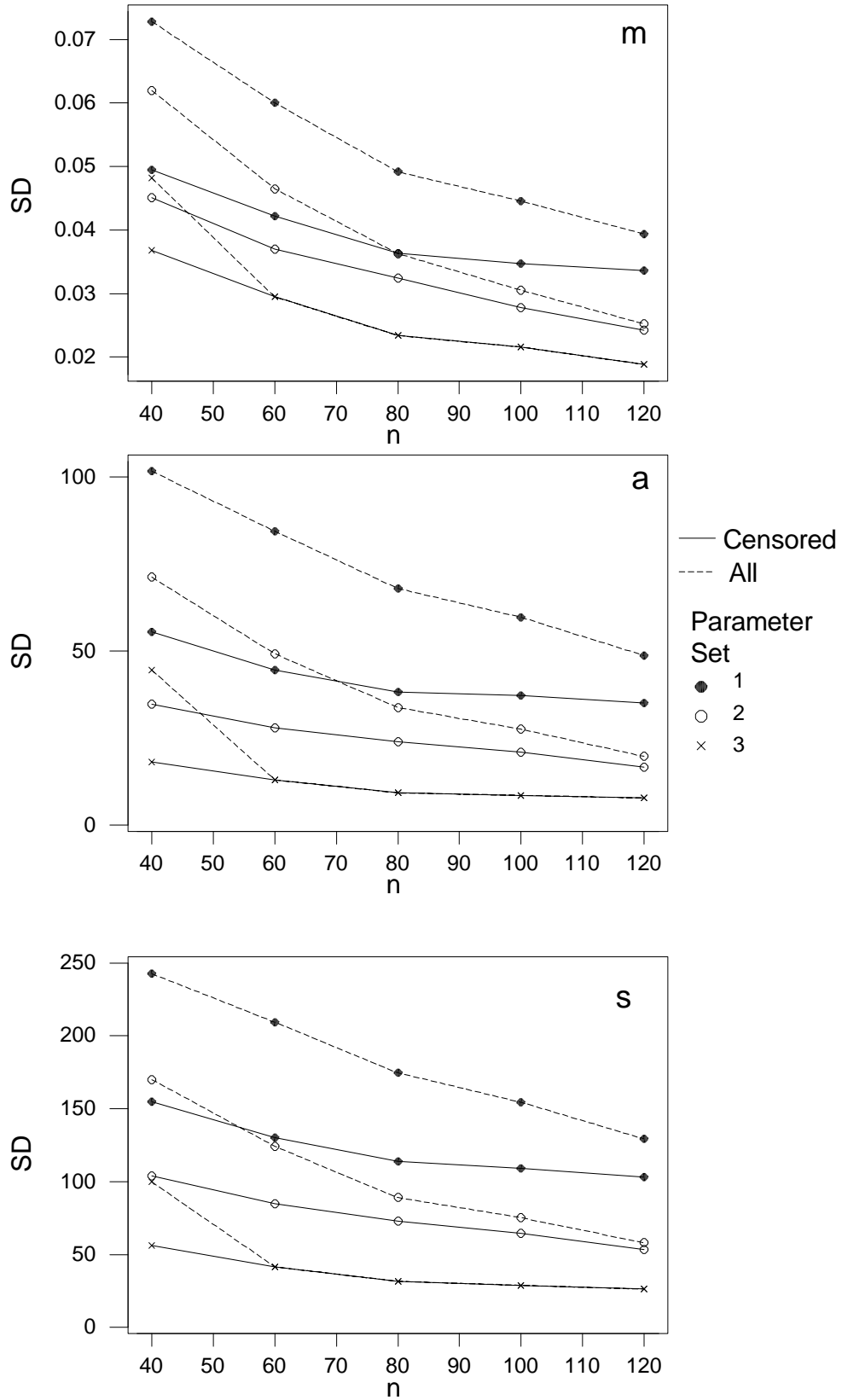
Figures
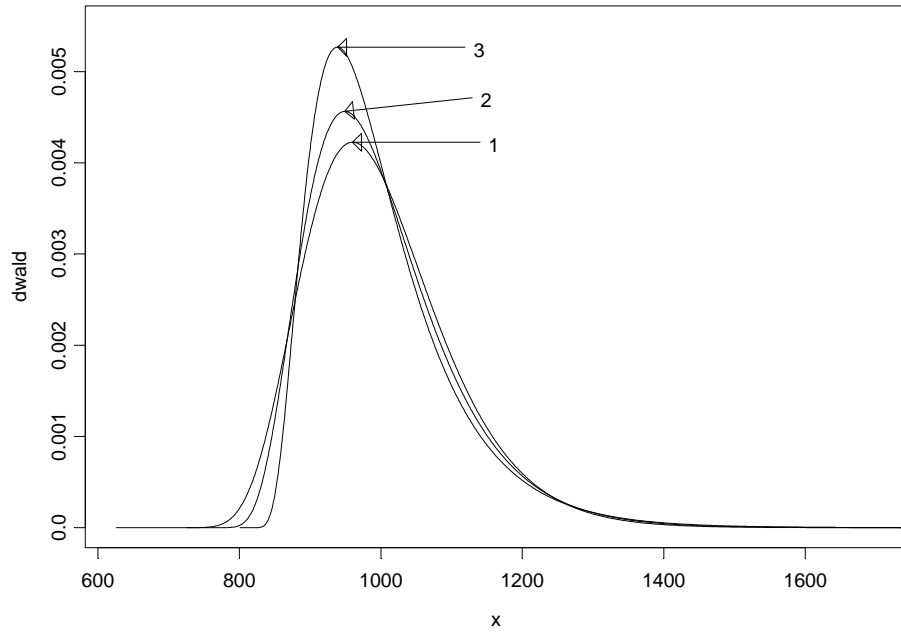
(a)
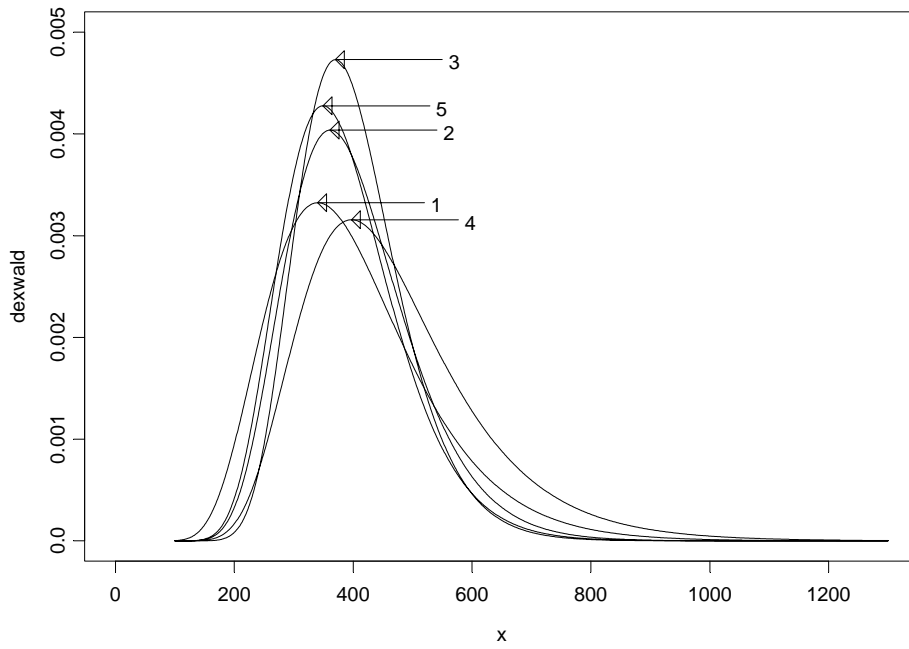
(b)

Figure 1

Figure 2

Figure 3

p_go = 0.5
distance = 1

RT (ms)

p_go = 0.5
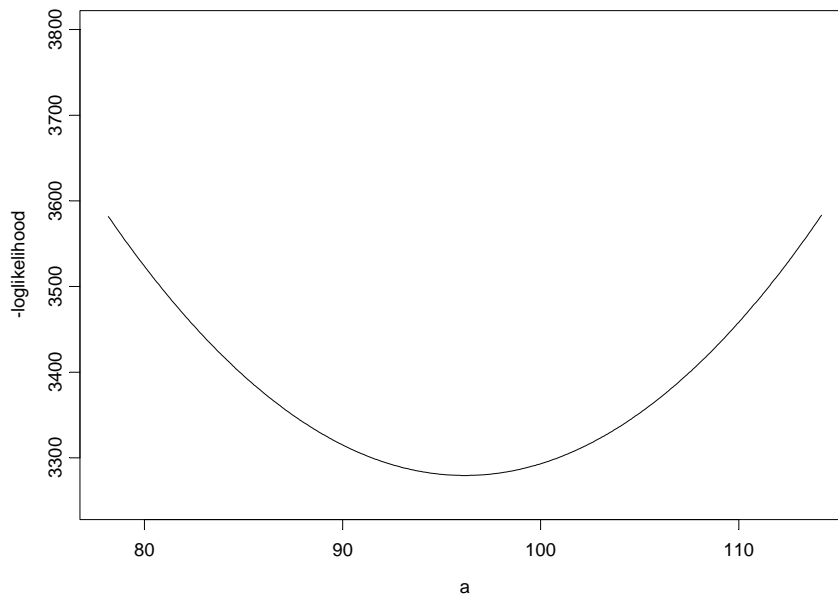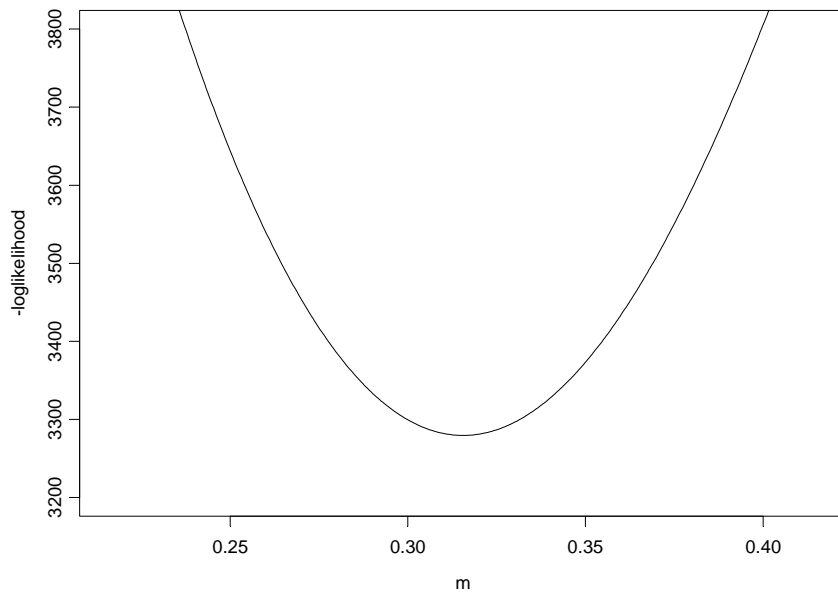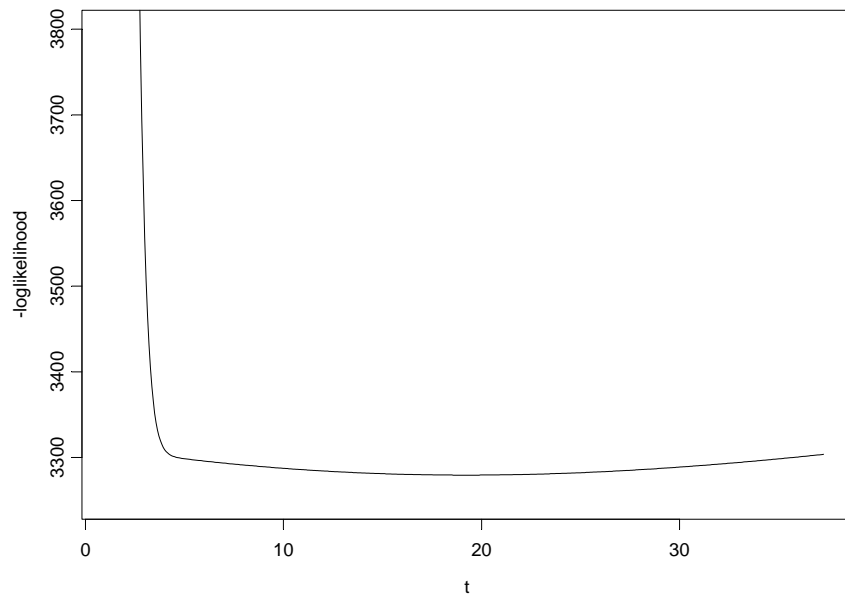distance = 4
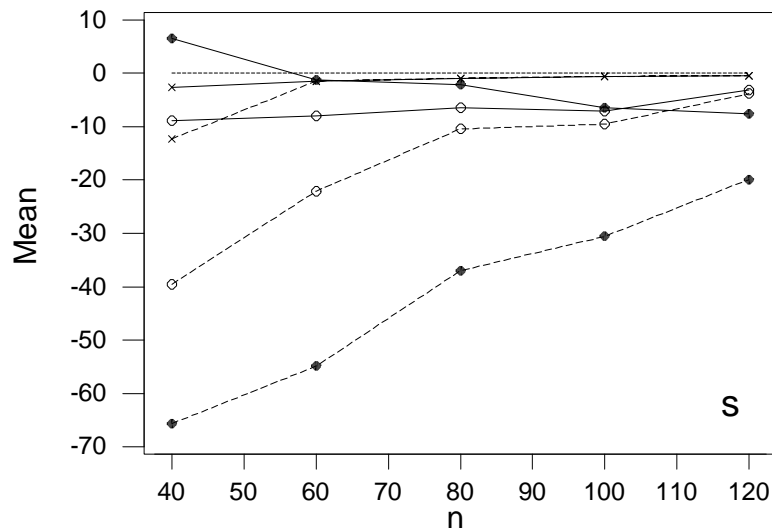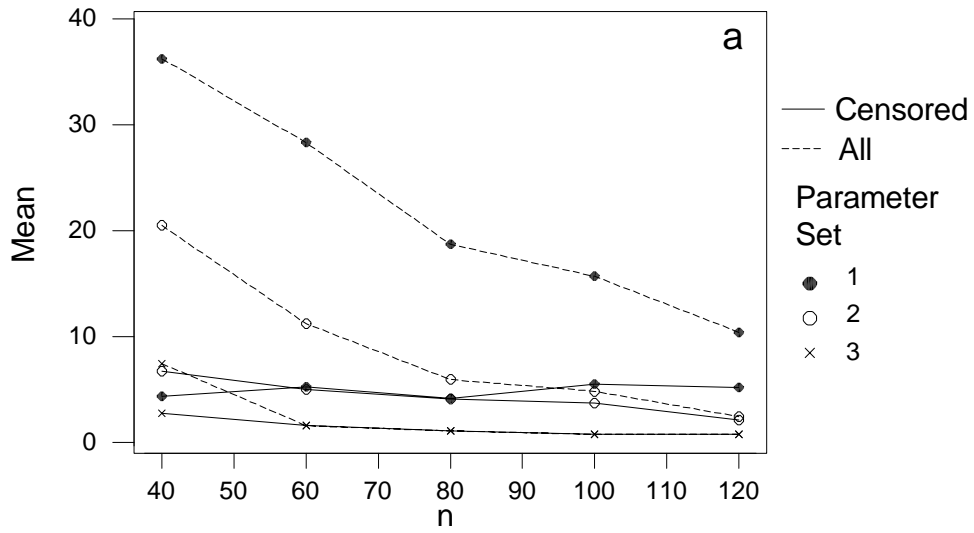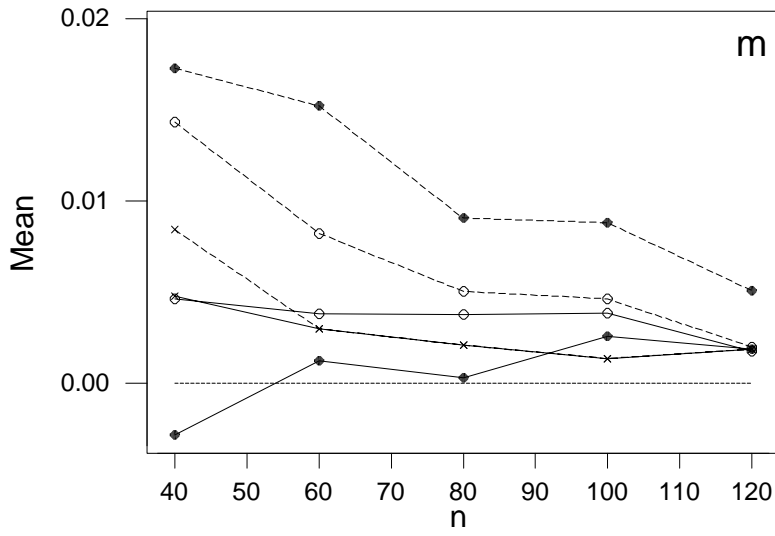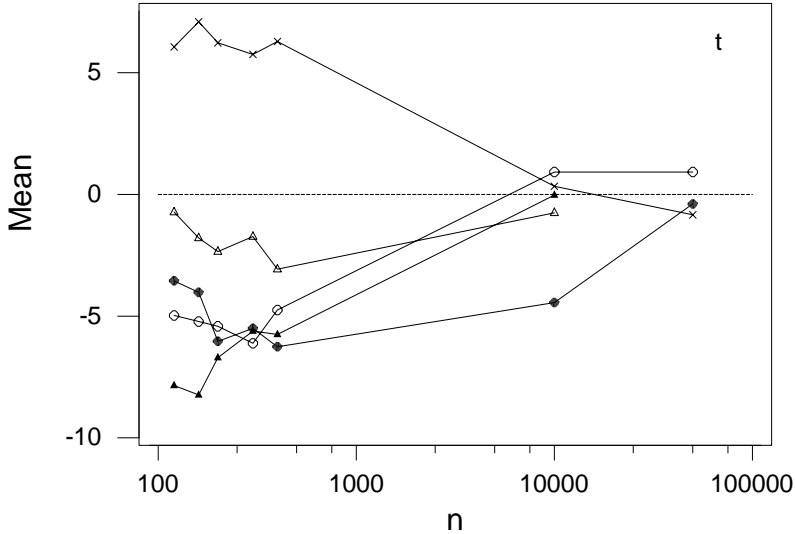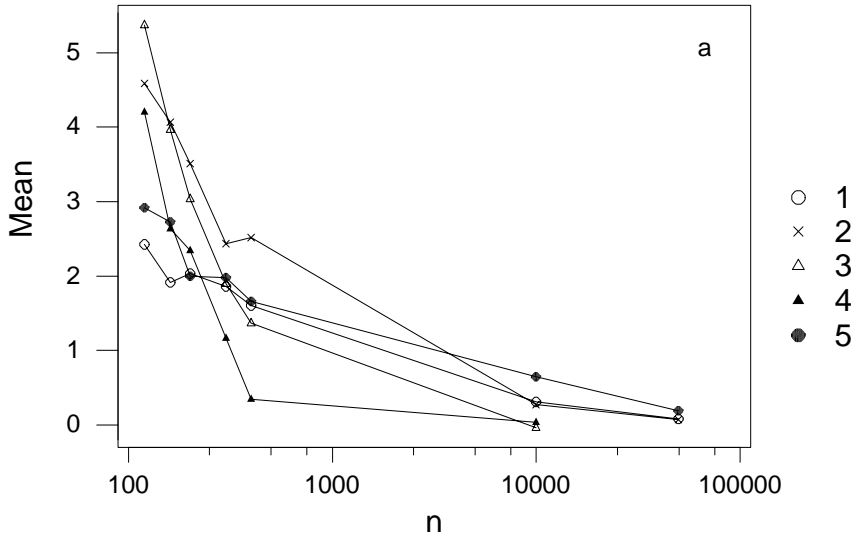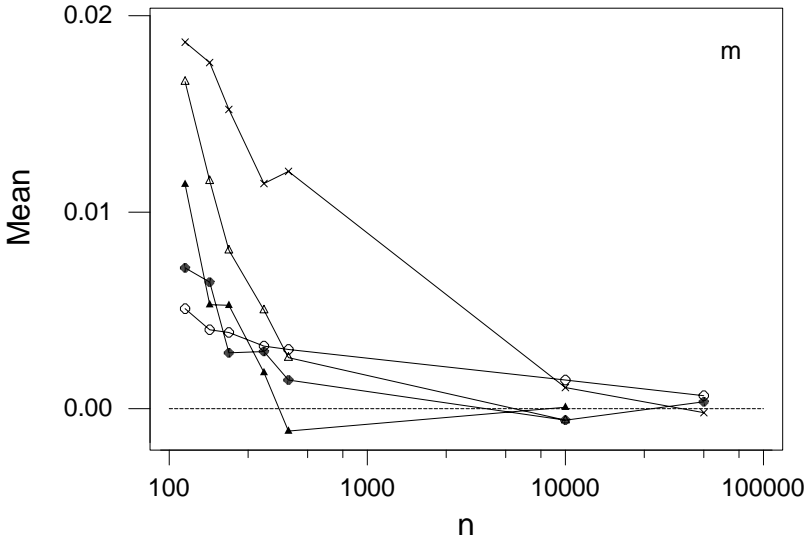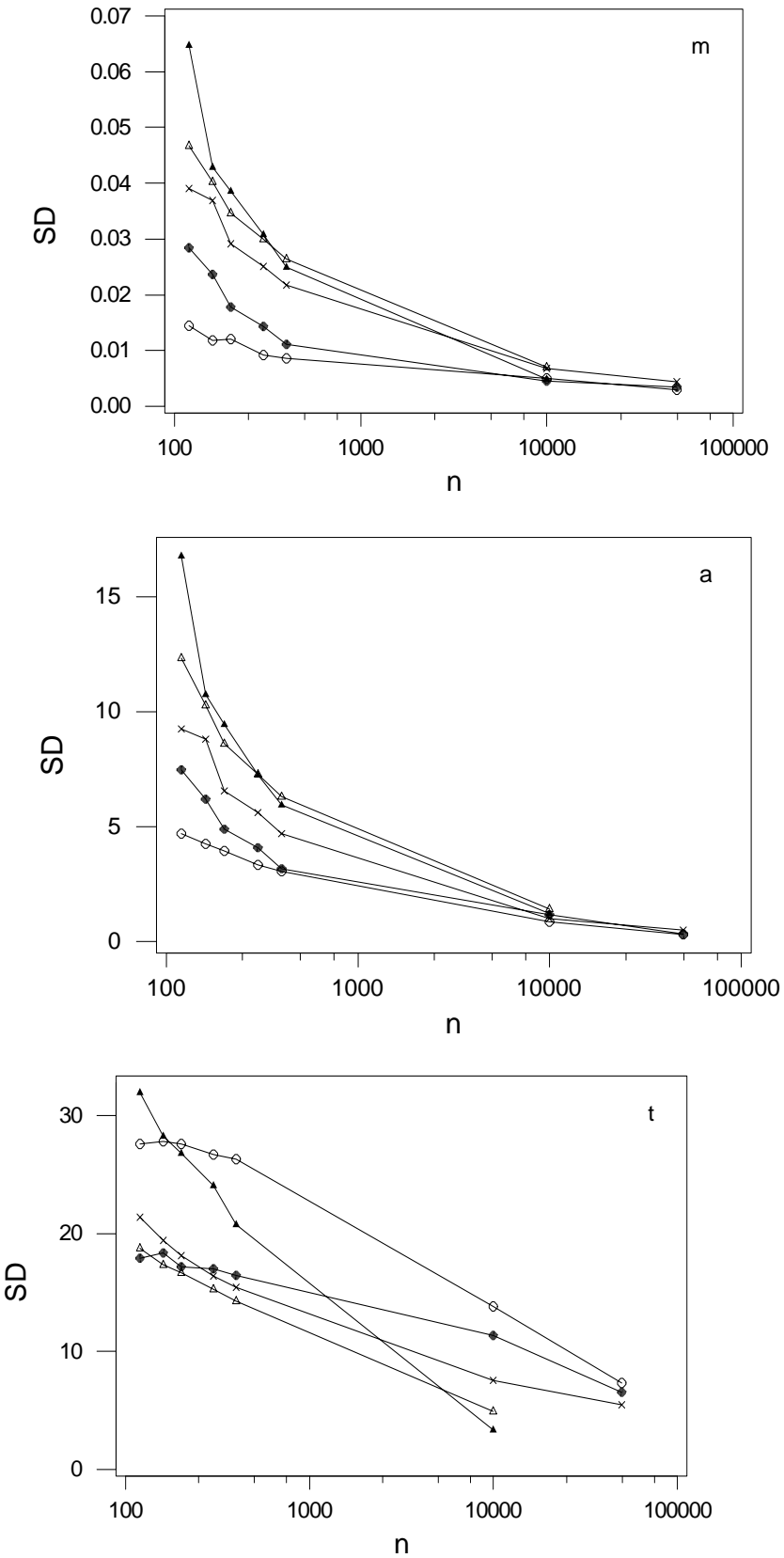
RT (ms)
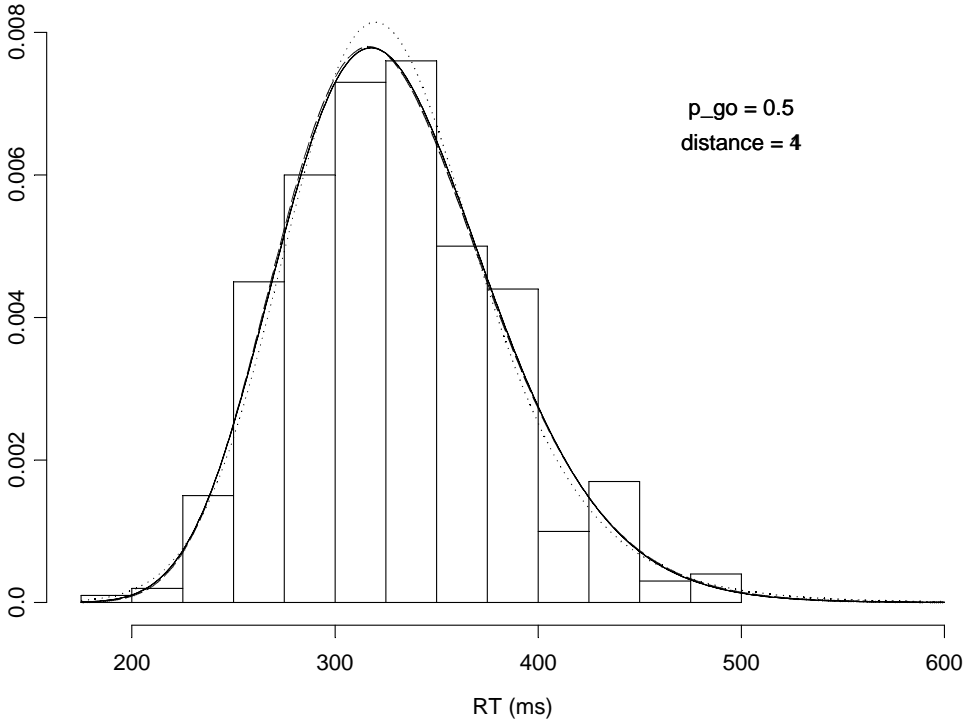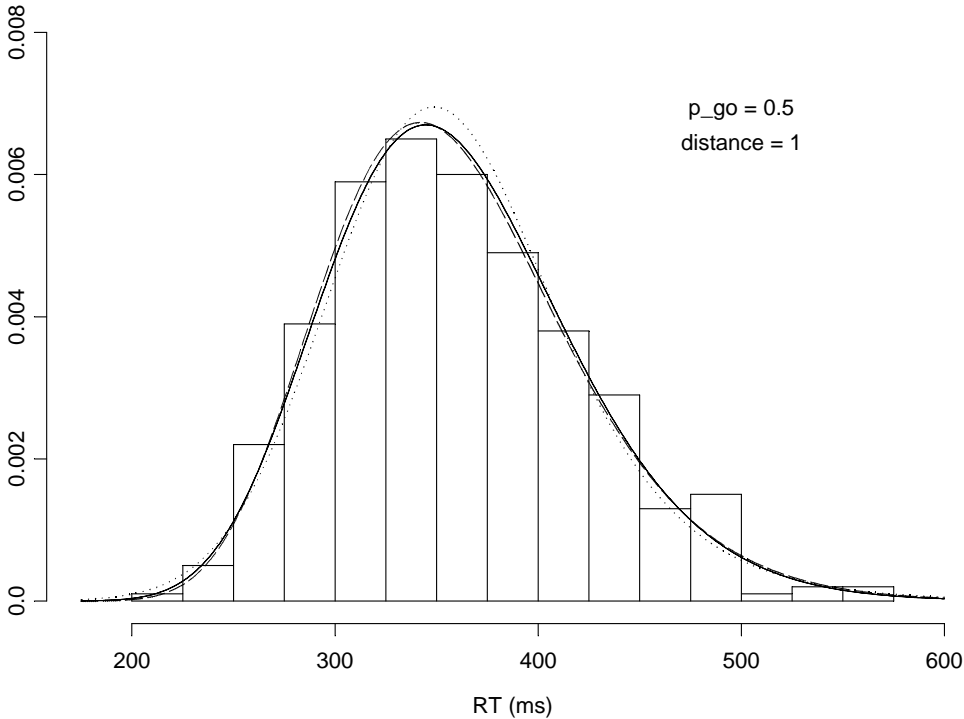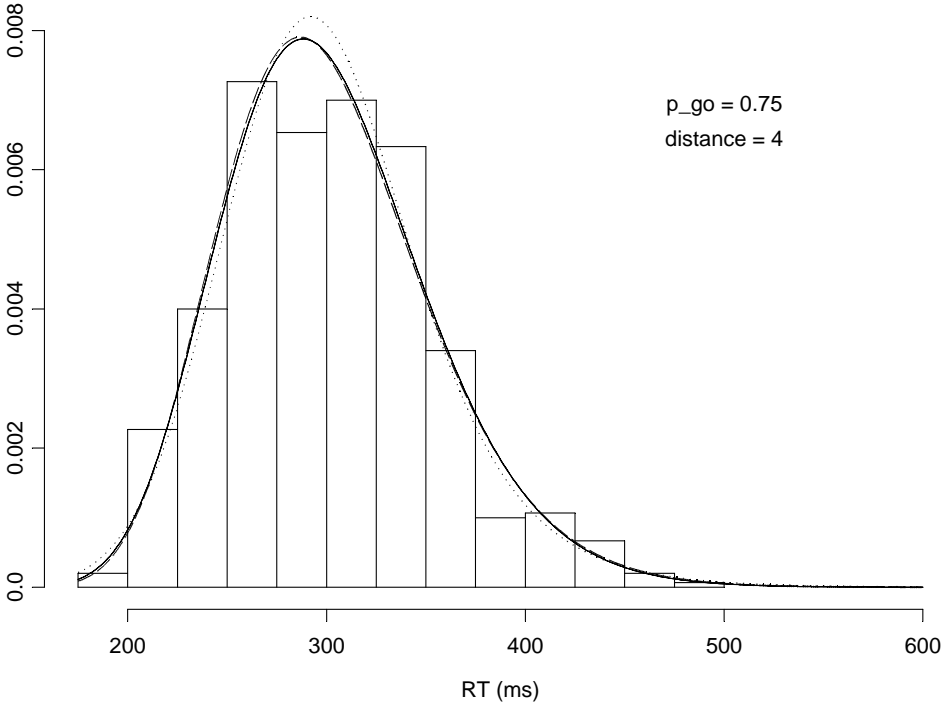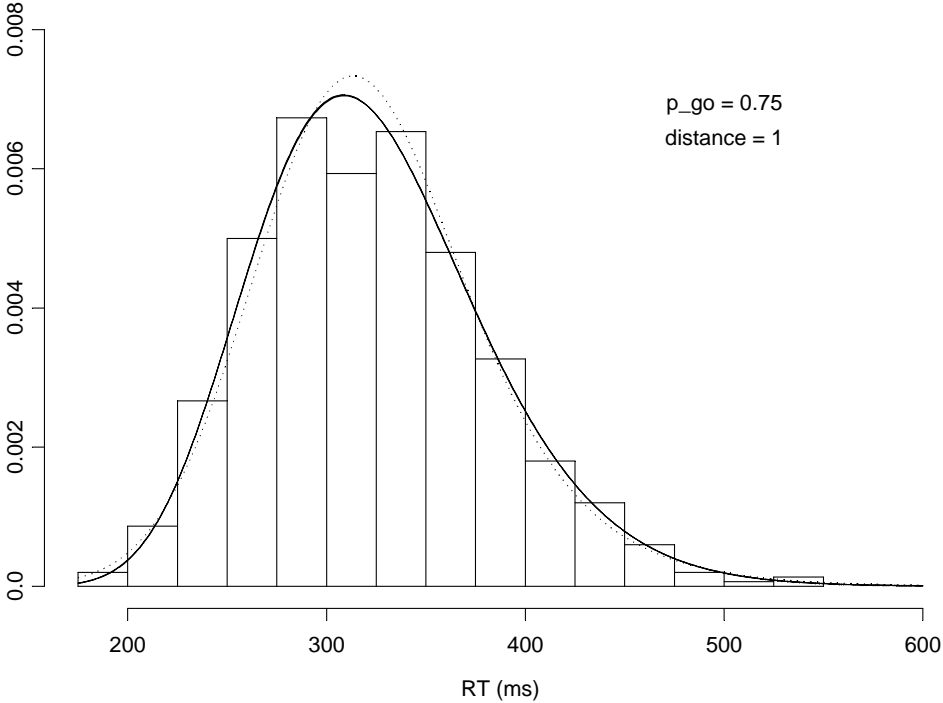
Figure 4

Figure 5

Footnotes

---

[1]All parameters are specified by exactly the same symbols in equations given in the text and the S

functions. Consequently the usual convention of using Greek letters for parameters is not adopted.

[2] Schwarz (2001) suggests the approximation be applied when $(mw + a)/\sqrt{w} \geq 5.5$. However, the *pnorm*

function implemented in S much more accurate than the one used by Schwarz, and benchmarking against

numerical integration results suggests that the second summand as expressed in Equation 2 using *pnorm* is

more accurate than the approximation unless *pnorm* = 0, and/or e^{2am} = Inf, where "Inf" is the value

assigned by S when a number is larger than can be represented in double precision floating point notation.

Hence, *pwald* uses the approximation only when one or both of these conditions is true, in which case it

yields results accurate to approximately 6 significant figures.

[3] http://www.statsci.org/s/invgauss.s, with help at http://www.statsci.org/s/invgauss.html. These functions

are parameterised in terms of the mean of the distribution $\mu = a/m$ and its "dispersion": $\lambda = a^2$. For the Wald

without a shift, these parameters have closed form maximum likelihood estimates, the arithmetic sample

mean for μ and $1/\hat{\lambda} = \sum_{i=1}^{n}\left(x_i^{-1} - \hat{\kappa}_1^{-1}\right)\Big/n$. When this parameterisation is used, the Wald distribution is usually

called the "Inverse Gaussian". Note these functions are not built in to S. To access these functions they

must be saved to a text file, e.g., invgauss.s, and compiled, e.g., source("invgauss.s").

[4] Because of the computational cost of memory allocation costs in S, calculating each term in the series

sequentially with a check for convergence of the approximation at each step is much slower than

calculating a larger number of terms as a vector. The *uandv* function allows vectors of terms to be

calculated and can check for convergence after each set is calculated. The default setting, which calculates

a vector of the first 20 terms and does not check for convergence, was accurate and fast in all cases

examined in the numerical studies reported here. Using more than 20 terms can cause numerical problems

in some cases.

[5] In the Monte Carlo study it was necessary to set the minimum values of (*m*,*a*,*t*) to (1e-8,1e-8,1) for the

Ex-Wald in order to avoid occasional floating point errors.

[6] In theory the full model should never have a greater deviance than the nested model. However, where the full model is not sufficiently constrained by the data, numerical problems can make it difficult for the search algorithm to find the minimum, and so slightly larger deviance values may be found for the full model. This was the case for all but one of the Ex-Wald fits. In these cases λ was set to zero.