Bayesian Estimation of The Ex-Gaussian Distribution

Abir El Haj [1,*], Yousri Slaoui [1], Clara Solier [2,3], Cyril Perret [1,2]

[1] *Laboratoire de Mathématiques et Applications, Université de Poitiers, France*
[2] *CeRCA-CNRS UMR 7295, 5, rue T. Lefebvre, MSHS - Bâtiment A5-TSA 21103, 86073 Poitiers cedex 9, France*
[3] *Maison des Sciences de l'Homme et de la Société, USR 3565, Université de Poitiers, France*

Abstract   Fitting of the exponential modified Gaussian distribution to model reaction times and drawing conclusions from its estimated parameter values is one of the most popular method used in psychology. This paper aims to develop a Bayesian approach to estimate the parameters of the ex-Gaussian distribution. Since the chosen priors yield to posterior densities that are not of known form and that they are not always log-concave, we suggest to use the adaptive rejection Metropolis sampling method. Applications on simulated data and on real data are provided to compare this method to the standard maximum likelihood estimation method as well as the quantile maximum likelihood estimation. Results shows the effectiveness of the proposed Bayesian method by computing the root mean square error of the estimated parameters using the three methods.

Keywords   Adaptive rejection Metropolis sampling, Bayesian estimation approach, Exponential modified Gaussian distribution, Maximum likelihood estimation, Quantile maximum likelihood, Response time

1. Introduction

One of the most commonly used measure in scientific psychology is reaction time (RT) ([25]; [26]), i.e., the time taken by a human being to make a response to a stimulus. The scientific psychology interest on this measure stems from the fact that it allows to explore human cognition thanks to the experimental manipulation of the characteristics of the stimuli used (see [5]). It is generally accepted that reaction times reflects the length of the time taken to realise three activities: perception, cognitive processing and the preparation of motor response. Assume that a psychologist makes the assumption that words of a language are organized in memory according to their frequency of occurrence. The experimental manipulation of this factor should have an impact on cognitive processes. In particular, a human being should be more efficient and so faster to process high frequency words (e.g., maison [house]) than low frequency words (e.g., ornithorynque, [platypus]). In an experimental task, (e.g. handwrite on a graphic tablet a word as soon as it is auditory presented) it should be reflected by shorter reaction times for the first type of item compared to the second type. Durations necessary to perceive the stimulus and to prepare the motor responses being constant, the comparison of reaction times between lexical frequency conditions allows researchers to test the hypothesis of a word organization in memory in function of their frequency in the language (e.g., [17]).

Based on Fisher's proposals ([19]), experimental psychology bases its conclusions on the use of hypothesis testing. Although virtuous, this practice causes several difficulties for studies using reaction times. As early as 1956, [13] drew the attention of researchers to the presence of a large number of very long RTs compared to the number of

---

*Correspondence to: Abir El Haj (Email: abir.el.haj@math.univ-poitiers.fr). Department of Applied Mathematics, University of Poitiers,11 Boulevard Marie et Pierre Curie,Batiment H3 - TSA 61125,86073 POITIERS CEDEX 9, France.

shortest values. The positive skewness of RT distributions is well documented in the literature (e.g., [24]; [10]; [28]; [20]; [16]; [18]; [29]; [21]; [25], [26]). Another well described aspect is the impact of this positive skewness on parameter estimates ([21]) and thus on hypothesis testing (i.e., t-tests, ANOVA, etc.). This issue is mostly managed in scientific psychology using strategies to normalize the RT distribution, such as data transformation using link functions ([6]) or as a procedure for right-censoring the distribution (e.g., [22]; [14]) without, however, ignoring all the problems that these strategies can generate (e.g., [9]). Another approach has been to describe the distribution of RTs from the sum of two stochastically independent random variables where one of these two variables is sampled from an exponentially distribution and the second is Gaussian distributed. This exponential modified Gaussian (ex-Gaussian), (see [24]) allows psychologists to work with a theoretical distribution function closer to those of reaction times. Whereas, as an alternative, [32] used the ex-Wald distribution, while [3], [11], [12] and [16] used the Wald, the Weibull and the log normal distributions.

The aim of this work is to estimate the parameters of the exponential modified Gaussian distribution (ex-Gaussian). The experimental conditions of scientific psychology are a constraint for these estimates. To thread the previous example, a researcher may make the same assumption but this time focusing on a population of young children. With this kind of participants, it is difficult to extract more than 40/50 RT. Most of the solutions proposed to estimate the parameters of the exponential modified Gaussian distribution (ex-Gaussian) are based on frequentist approaches from the maximization of the log-likelihood function. However, it is well known that this approach works effectively with $n > 100$ samples (see [25] and [26]). The aim of this work is to explore the effectiveness of alternative approaches for the estimation of the parameters of the exponential modified Gaussian distribution with $n < 100$ samples. We suggested a Bayesian parametric approach in which these parameters are considered as random variables. Thus, we choose a prior density for each of the variables and we compute the likelihood of the data. Then, we calculate the conditional posterior of each parameter using the chosen priors and the likelihood. Since the computing posteriors are not of known densities form neither log-concave, we propose to use the adaptive rejection Metropolis sampling method to sample from these conditional posteriors.

To prove the validity of the method and to highlight its main features, we set some applications of the proposed approach by using first simulated data and then real data sets. We compare the root mean square error (RMSE) of the three parameters of the ex-Gaussian obtained by applying the proposed Bayesian method, the quantile maximum likelihood (QML) method proposed by [3], [23] and [2] and the frequentist maximum likelihood estimation (MLE) method proposed by [7] on simulated data. Results shows that the RMSE of the three parameters obtained using the proposed Bayesian method is closer to zero than those obtained by using the quantile maximum likelihood and the maximum likelihood estimation methods which prove the effectiveness of our method.

In this paper, we define the ex-Gaussian distribution in section 2. The section 3, which presents a Bayesian parametric approach to estimate the parameters of the ex-Gaussian, is divided into the two subsections 3.1 and 3.2. In subsection 3.1, since the parameters are considered as random variables in the Bayesian framework, a prior density is chosen for each parameter then, the conditional posteriors of these parameters are computed using these priors and the likelihood. The subsection 3.2 presents the algorithm used to estimate the parameters of the ex-Gaussian based on the rejection Metropolis sampling method. In section 4, we recall the quantile maximum likelihood and the maximum likelihood estimation approaches in order to compare them to the proposed Bayesian approach. In section 5, we provide some applications of the proposed method. Indeed, we set some applications of the proposed method by introducing some simulated data in subsection 5.1, then an application on real dataset in subsection 5.2 to prove its effectiveness and highlight its main features. We compare the results obtained by the proposed method to those obtained by using the quantile maximum likelihood and the maximum likelihood estimation method. The final section is a conclusion of the paper.

2. Ex-Gaussian distribution

The ex-Gaussian is a well-known distribution in cognitive psychology (see [20]). It is the sum of independent normal and exponential random variables. An ex-Gaussian random variable Z may be expressed as $Z = X + Y$, where X and Y are independent, X is Gaussian with mean μ and variance σ^2, and Y is exponential of rate λ.

The probability density function of the ex-Gaussian is

$$f(x|\mu,\sigma,\tau) = \frac{1}{\tau}\exp\left(\frac{\mu - x}{\tau} + \frac{\sigma^2}{2\tau^2}\right)\Phi\left(\frac{x-\mu}{\sigma} - \frac{\sigma}{\tau}\right), \tag{1}$$

where $x = (x_1,\ldots,x_n)$ is the data vector of length n and $\Phi(z)$ is the cumulative standard Gaussian distribution, evaluated from $-\infty$ to z, given by

$$\Phi\left(\frac{x-\mu}{\sigma} - \frac{\sigma}{\tau}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\mu}{\sigma} - \frac{\sigma}{\tau}} \exp(-\frac{y^2}{2})\, dy$$

and the parameters μ, σ and τ are the mean, the standard deviation of the Gaussian and the scale of the exponential, respectively. Note that $\sigma > 0$ and $\tau > 0$.

In the sequel, to simplify the computation, we take $\zeta = \sigma^2$ and $\lambda = \frac{1}{\tau}$. Let θ be the set of the model parameters, thus $\theta = (\mu,\zeta,\lambda)$.

3. Bayesian parametric approach

In frequentist statistical inference, the parameters of the model are assumed to be fixed quantities and the data is random. In this case, probability distributions is used to represent the data. However, in Bayesian inference, probability distributions are often used to represent more than just the data. They also represent the uncertainty of the prior in model parameters. Then, these get updated with current data using Bayes' theorem to produce posterior probability distributions.

3.1. Bayesian estimation

In Bayesian inference, we start by choosing a prior probability distribution for the parameters of interest θ. To simplify the computations, we use conjugate prior for the model parameters.. This prior quantifies the existing knowledge about θ:

$$\mu \sim \mathcal{N}(\alpha_1, \alpha_2),$$

$$\zeta \sim \mathcal{IG}(\beta_1, \beta_2),$$

$$\lambda \sim G(\gamma_1, \gamma_2),$$

where $\mathcal{N}, \mathcal{IG}$ and G are the Gaussian, Inverse Gamma and Gamma distribution respectively.

The prior distribution is then updated by the incoming data (i.e., likelihood) to yield a posterior probability distribution under Bayes's rule:

$$f(\mu,\zeta,\lambda|x) = \frac{f(x|\mu,\zeta,\lambda)f(\mu)f(\zeta)f(\lambda)}{f(x)}, \tag{2}$$

where $f(\mu)$, $f(\zeta)$ and $f(\lambda)$ are the prior densities for the parameters μ, ζ and τ respectively and $f(x|\mu,\zeta,\lambda)$ is the joint distribution defined in (1) by considering $\zeta = \sigma^2$ and $\lambda = \frac{1}{\tau}$. The marginal likelihood $f(x)$ is the probability of the observed data x and does not involve the parameters of interest. Hence, the equation (2) can be expressed as

$$f(\mu,\zeta,\tau|x) \propto f(x|\mu,\zeta,\tau)f(\mu)f(\zeta)f(\tau). \tag{3}$$

The symbol $\propto$ denotes proportionality.

The likelihood function is the joint density of the data given the parameters, viewed as a function of the parameters:

$$
\begin{aligned}
L(\mu, \zeta, \lambda) &= \prod_{i=1}^{n} f(x_i | \mu, \zeta, \lambda) \\
&= \prod_{i=1}^{n} \lambda \exp\left((u - x_i)\lambda + \frac{\lambda^2 \zeta}{2} \right) \Phi\left(\frac{(x_i - \mu)}{\sqrt{\zeta}} - \lambda\sqrt{\zeta} \right) \\
&= \lambda^n \exp\left(\sum_{i=1}^{n} (u - x_i)\lambda + \frac{n\lambda^2 \zeta}{2} \right) \prod_{i=1}^{n} \Phi\left(\frac{(x_i - \mu)}{\sqrt{\zeta}} - \lambda\sqrt{\zeta} \right).
\end{aligned}
$$

Now, to find the joint density of x and the parameters, we multiply $f(x|\mu, \zeta, \lambda)$ by the prior distributions

$$
\begin{aligned}
f(x, \mu, \zeta, \lambda) &\propto \lambda^n \exp\left(\sum_{i=1}^{n} (u - x_i)\lambda + \frac{n\lambda^2 \zeta}{2} \right) \prod_{i=1}^{n} \Phi\left(\frac{(x_i - \mu)}{\sqrt{\zeta}} - \lambda\sqrt{\zeta} \right) \\
&\times \frac{1}{\alpha_2} \exp\left(-\frac{(\mu - \alpha_1)^2}{2\alpha_2^2} \right) \frac{\beta_2^{\beta_1} \left(\frac{1}{\zeta} \right)^{\beta_1 + 1} \exp\left(-\frac{\beta_2}{\zeta} \right)}{\Gamma(\beta_1)} \frac{\gamma_2^{\gamma_1} \lambda^{\gamma_1 - 1} \exp(-\gamma_2 \lambda)}{\Gamma(\gamma_1)}.
\end{aligned}
\tag{4}
$$

According to (3), the full conditional posterior distribution is proportional to the joint density (4)

$$
f(\mu, \zeta, \lambda | x) \propto f(x, \mu, \zeta, \lambda).
\tag{5}
$$

Thus, the conditional distribution of μ is given by

$$
\begin{aligned}
f(\mu | x, \zeta, \lambda) &= \frac{f(\mu, \zeta, \lambda | x)}{f(\zeta, \lambda | x)} \\
&\propto f(x, \mu, \zeta, \lambda).
\end{aligned}
\tag{6}
$$

The equality (6) is obtained by using the equation (5) and by noticing that the denominator $f(\zeta, \lambda | x)$ does not involve the parameter μ.

According to (6) and by ignoring all the factor that does not involve μ, we obtain

$$
f(\mu | x, \zeta, \lambda) \propto \exp\left(\sum_{i=1}^{n} (u - x_i)\lambda \right) \prod_{i=1}^{n} \Phi\left(\frac{(x_i - \mu)}{\sqrt{\zeta}} - \lambda\sqrt{\zeta} \right) \exp\left(-\frac{(\mu - \alpha_1)^2}{2\alpha_2^2} \right).
\tag{7}
$$

Now, we are interested in computing the conditional posterior of ζ. By proceeding in the same way, $f(\zeta | x, \mu, \lambda)$ can be expressed as follows

$$
f(\zeta | x, \mu, \lambda) \propto \exp\left(\frac{n\lambda^2 \zeta}{2} \right) \prod_{i=1}^{n} \Phi\left(\frac{(x_i - \mu)}{\sqrt{\zeta}} - \lambda\sqrt{\zeta} \right) \left(\frac{1}{\zeta} \right)^{\beta_1 + 1} \exp\left(-\frac{\beta_2}{\zeta} \right).
\tag{8}
$$

Likewise, the conditional posterior of λ can be expressed as follows

$$
f(\lambda | x, \mu, \lambda) \propto \lambda^n \exp\left(\sum_{i=1}^{n} (u - x_i)\lambda \right) \exp\left(\frac{n\lambda^2 \zeta}{2} \right) \prod_{i=1}^{n} \Phi\left(\frac{(x_i - \mu)}{\sqrt{\zeta}} - \lambda\sqrt{\zeta} \right) \lambda^{\gamma_1 - 1} \exp(-\gamma_2 \lambda).
\tag{9}
$$

Sampling from the conditionals posterior given in (7), (8), and (9) is somewhat complex since they are not functions of known densities. Since the chosen priors do not yield to conditional densities that are always log-concave (Recall that a density is log concave if $\frac{d^2}{dl^2}(\log f(l)) \leq 0$ for all l). We suggest to use the adaptive rejection Metropolis sampling (ARMS) technique proposed by [30] which provides a reasonably efficient method for sampling from the conditional posterior distributions. This algorithm is used for sampling from densities that do not have a closed form. Note that this sampling method is available in the package "HI" (see [8]) of the software R under the name arms.

3.2. *Algorithm of sampling from the posterior functions*

We presented in this section the algorithm of sampling from the conditional posterior distributions (7), (8) and (9) based on the ARMS algorithm (see [30],[31]).

Algorithm 1: Sampling algorithm based on the ARMS algorithm, N is the number of iterations and M is the burn-in value.

REQUIRE: M and N.
1: Random initialization of $\theta^{(0)} = (\mu^{(0)}, \zeta^{(0)}, \lambda^{(0)})$.
2: **for** $j = 1, \ldots, N$ **do**
3:    $\mu^{(j)} \sim f(\mu|x, \zeta^{j-1}, \lambda^{j-1})$, {using `arms` command of the HI package on the software R ([8])} (f is given in (7))
4:    $\zeta^{(j)} \sim f(\zeta|x, \mu^{j}, \lambda^{j-1})$, {using the `arms` command of the HI package} (f is given in (8))
5:    $\lambda^{(j)} \sim f(\lambda|x, \mu^{j}, \zeta^{j})$, {using `arms` command of the HI package} (f is given in (9))
6: **end for**
7: $\widehat{\mu} = (N-M)^{-1} \sum_{j=M}^{N} \mu^{(j)}$, $\widehat{\zeta} = (N-M)^{-1} \sum_{j=M}^{N} \zeta^{(j)}$, $\widehat{\lambda} = (N-M)^{-1} \sum_{j=M}^{N} \lambda^{(j)}$ {average all estimates after the burn-in period}
Ensure: $\widehat{\theta} = (\hat{\mu}, \hat{\zeta}, \hat{\lambda})$.

In the following section, we recall the quantile maximum likelihood (QML) estimation method proposed by [3], [23] and [2] to estimate the ex-Gaussian distribution and the frequentist approach based on the maximum likelihood estimation (MLE) (see [7]) in order to compare them later to the proposed Bayesian approach.

4. Recall on the **QML** and the **MLE** approaches

In this section, we recall the MLE and the QML estimation methods. First, the MLE consists in computing the log-likelihood function of the data vector $x = (x_1, \ldots, x_n)$ sampled from an ex-Gaussian distribution then, in maximizing this function to estimate the parameters μ, σ and τ of the ex-Gaussian. Note that the algorithm is available on the package "emg" of the R software under the name `emg.mle`.

Now, we show how the QML method work to estimate these parameters. This method is a variant of the MLE which combine the robustness of quantiles and the efficiency and consistency of MLE. Following [3], for the QML approach, we transform the data vector $x = (x_1, \ldots, x_n)$ sampled from an ex-Gaussian distribution and of length n into a vector of quantile estimates ($\hat{q}$) and a vector of counts (N) of the length of the portions of the data vector x taking place in the inter-quantile range. Using these vectors, we compute the quantile log-likelihood function, then we maximize it to estimate the parameters of the ex-Gaussian distribution. The full algorithm can be found in [3]. Furthermore this algorithm is available in the package "QMLE.zip" (see [27]) of the Matlab software.

In the following, we present numerical experimental application of the Bayesian, the frequentist and the quantile maximum likelihood applications to compare them.

5. Numerical experimental

This section aims at comparing the proposed Bayesian method to the QML and the MLE methods listed above with application on simulated data first, then on real data set.

5.1. *Simulated data*

We sample $S = 100$ data vectors $x = (x_1, \ldots, x_n)$ of size n, for $n = \{40, 80, 100, 110, 200, 300, 500, 1000\}$, from the ex-Gaussian distribution (sum of an independent normal and an exponential distributions). The parameters used in this simulation are: $\mu = 13, \sigma = 7$ and $\tau = 9$. Note that for these simulations, we use a burn-in of 2500

iterations out of 5000 Gibbs samples, then we estimate posteriors from a further 2500 iterations. To use the proposed Bayesian method (algorithm 1) detailed in section 3.2, We choose the following initialization of the parameters: $\mu^{(0)} = mean(x) - (sd(x) * 0.8)$, $\zeta^{(0)} = (var(x) - (sd(x) * 0.8)^2)$ and $\lambda^{(0)} = 1/(sd(x) * 0.8)$.

We present in Table 1 and Table 2, the root mean square error (RMSE) of the three parameters of the ex-Gaussian for $n = \{40, 80, 100, 110\}$ and $n = \{200, 300, 500, 1000\}$ respectively, defined by

$$RMSE(\mu) = \sqrt{\frac{1}{S}\sum_{s=1}^{S}(\hat{\mu}^{(s)} - \mu)^2}, \quad RMSE(\sigma) = \sqrt{\frac{1}{S}\sum_{s=1}^{S}(\hat{\sigma}^{(s)} - \sigma)^2} \quad \text{and} \quad RMSE(\tau) = \sqrt{\frac{1}{S}\sum_{s=1}^{S}(\hat{\tau}^{(s)} - \tau)^2},$$

where the superscript s labels the estimates obtained in simulation s.

	$n = 40$			$n = 80$			$n = 100$			$n = 110$		
	RMSE(μ)	RMSE(σ)	RMSE(τ)	RMSE(μ)	RMSE(σ)	RMSE(τ)	RMSE(μ)	RMSE(σ)	RMSE(τ)	RMSE(μ)	RMSE(σ)	RMSE(τ)
QML	**2.6**	**1.63**	**2.52**	2.15	**1.24**	**2.11**	2.12	1.32	**1.92**	2.32	1.48	**2.13**
MLE	3.29	2.08	3.29	2.3	1.44	2.33	2.31	1.38	2.2	1.96	1.31	1.9
Bayesian	2.82	1.74	3.5	**1.97**	1.33	2.8	**1.7**	**1.1**	2.4	**1.46**	**1.08**	2.19

Table 1. Root Mean square error of the three parameters μ, σ and τ for the simulated data of sizes $n = 40, 80, 100$ and 110 using the QML, the MLE and the Bayesian approaches.

	$n = 200$			$n = 300$			$n = 500$			$n = 1000$		
	RMSE(μ)	RMSE(σ)	RMSE(τ)	RMSE(μ)	RMSE(σ)	RMSE(τ)	RMSE(μ)	RMSE(σ)	RMSE(τ)	RMSE(μ)	RMSE(σ)	RMSE(τ)
QML	1.57	1.05	1.56	1.21	1.15	1.32	0.93	1.05	0.97	0.7	0.72	0.82
MLE	1.17	0.98	**1.16**	1.06	0.93	1.11	0.73	0.78	0.85	0.59	0.74	0.71
Bayesian	**1.09**	**0.88**	1.23	**0.93**	**0.71**	**1.08**	**0.58**	**0.54**	**0.47**	**0.38**	**0.47**	**0.31**

Table 2. Root Mean square error of the three parameters μ, σ and τ for the simulated data of sizes $n = 200, 300 , 500$ and 1000 using the QML, the MLE and the Bayesian approaches.

Table 1 and table 2 show that the RMSE of the three estimated parameters of the ex-Gaussian obtained using the proposed Bayesian method is closer to zero than those obtained by using the QML and the MLE methods, when n is large, which prove the effectiveness of the proposed method. Furthermore, these tables show that for $n < 100$, the QML is more efficient then the MLE approach since for $n \in \{40, 80, 100\}$, the RMSE is closer to zero, while for $n > 100$, the RMSE of the ex-Gaussian estimated parameters obtained using the MLE method is closer to zero than the one obtained using the QML method. Thus, The QML is more efficient than the MLE approach for sample size less than or equal to 100.

We provide in table 3 the confidence intervals of the three estimators $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\tau}$.

	$n = 200$	$n = 300$	$n = 500$
$\hat{\mu}$	[11.22, 11.55]	[11.45, 11.69]	[11.65, 11.88]
$\hat{\sigma}$	[6.63, 6.81]	[6.71, 6.86]	[6.8, 6.9]
$\hat{\tau}$	[7.25, 8.31]	[7.3, 8.69]	[7.85, 8.9]

Table 3. Confidence intervals of the three estimators $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\tau}$ for the simulated data of sizes $n = 200, 300$ and 500 using the Bayesian approach.

We compute the $\hat{R}$ convergence diagnostic measure ([1]) for each model parameter. Recall that $\hat{R}$ compares the between-chain variability to the within-chain variability. It is available in the package "rstan" of the software R under the name Rhat. To prove that the chains have properly converged, $\hat{R}$ should be lower than 1.1. For the model parameters μ, ζ and λ, $\hat{R}$ is equal to 1, 0.99 and 0.98 respectively. Thus, all the $\hat{R}$ values are lower than 1.1 and

the the chains have properly converged. The coverage probability is equal to 0.942 which show the effectiveness of our proposed method.

Now, we present in Figure 1, the RMSE of the three estimated parameters of the ex-Gaussian distribution using the proposed Bayesian method. Figure 1 shows that the RMSE of the three estimated parameters approaches zero as the sample size n increases.
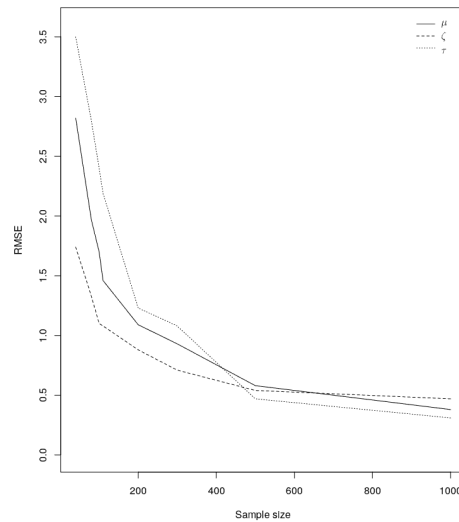

Figure 1. RMSE of the three estimated parameters of the ex-Gaussian distribution.

To check the convergence of the three estimated parameters of the ex-Gaussian toward the true parameters values, we present in Figure 2, the values of these three estimated parameters obtained for different sample size n.
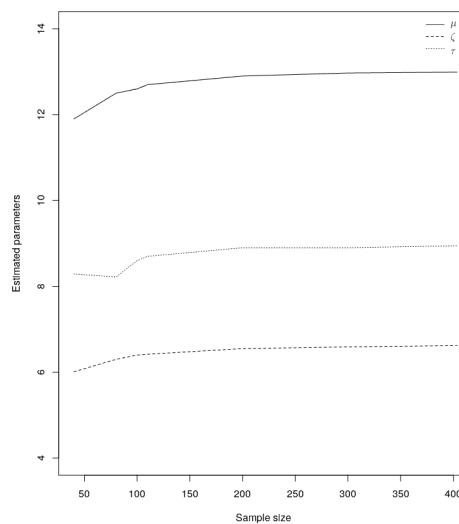

Figure 2. Convergence of the three estimated parameters of the ex-Gaussian distribution.

Figure 2 shows that the three estimated parameters μ, σ and τ converge toward the true values of the parameters which are 13, 7 and 9 respectively.

5.2. Application on real data

We set here two examples of data, one for $n < 100$ and the other for $n > 100$, then we compare the proposed Bayesian approach to the `QML` and the `MLE` methods.

For the first example, we have $n < 100$. More precisely, the number of participants is $n = 52$. Reaction times were recorded in a letter copy task. One undergraduate student had to copy twice in uppercase the 26 letters of the French alphabet on a graphic tablet (Wacom). Letters were presented randomly. A white sheet on the graphic tablet allowed collecting the handwritten productions to check responses. The participants could not see and monitor their production, as a box hided their hand and the graphic tablet.

By applying the Bayesian, the `QML` and the `MLE` methods on the first data example, we obtain the results presented in Table 4.

	$n = 52$		
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\tau}$
`QML`	440.74	60.83	106.26
`MLE`	448.71	53.27	100
`Bayesian`	366.32	14.17	179.92

Table 4. The three estimated parameters $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\tau}$ obtained by applying the `QML`, the `MLE` and the Bayesian approaches on the first data example ($n = 52$).

We compute the $\hat{R}$ for the three Bayesian model parameters μ, σ and τ. We obtain that $\hat{R}$ is equal to 1, 0.99 and 0.99 respectively. Thus, all the $\hat{R}$ values are lower than 1.1 and the chains have properly converged.

Now, we treat the case of $n > 100$. In this example, reaction times were recorded in a picture naming task. Two hundred and thirty-three participants ($n = 233$) had to handwrite the label of a specific drawing (e.g., Table) on a graphic tablet (Wacom). A white sheet on the graphic tablet allowed collecting the handwritten productions to check response.

We apply the three methods on this second data example, results are presented in Table 5

	$n = 233$		
	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\tau}$
`QML`	828.89	162.62	216.71
`MLE`	927.88	198.55	100
`Bayesian`	466.39	180.13	540.46

Table 5. The three estimated parameters $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\tau}$ obtained by applying the `QML`, the `MLE` and the Bayesian approaches on the second data example ($n = 233$).

Now, we compute the $\hat{R}$ for the three Bayesian model parameters μ, σ and τ. We obtain that $\hat{R}$ is equal to 1, 1 and 0.99 respectively. Thus, all the $\hat{R}$ values are lower than 1.1 and the chains have properly converged.

6. Conclusion

Through its experimental approach ([19]), scientific psychology supports these conclusions with hypothesis testing. Using reaction times as a measure of behavior causes difficulties in this context because of a positive skewness in the distribution of this measure. It is generally accepted that the distribution of reaction times fits well with an

exponential modified Gaussian distribution. (ex-Gaussian), (e.g., [24]; [10]; [28]; [20]; [16]; [18]; [29]; [21]; [25]; [26]). As explained in the Introduction, the sample size per experimental condition is a constraint to the use of the three parameters of ex-Gaussian distributions. Since ML methods require larger samples than those found in experimental psychology, the reliability of estimates is questionable ([25]; [26]). The constraint of the sample size is a factor on which the psychologist has little impact regarding the characteristics of the populations with which he works (patients, children, etc.) and the time necessary to collect reliable data (loss attention of participants if the experience is too long).

This work aims at proposing a reliable method for estimating the three parameters of the ex-Gaussian distribution for this type of experimental context. We first replicated the results of [3], ([23]; [4]; [2]) for small samples ($n < 100$): the QML method should be preferred for parameter estimation. Second, for larger samples ($n > 100$), our results suggested that estimates are more reliable with the Bayesian approach. This was particularly the case for very large samples ($n > 300$). Finally, we tested the three approaches on real data (handwritten production of letters ($n = 52$) and of picture names ($n = 233$)). The values proposed by the ML and QML approaches are relatively close whatever the size of the samples. The values proposed for the three parameters by the Bayesian approach were quite far from those of the other two methods. The Bayesian method being the most reliable regarding analyzes on the simulated data (section 5.1.), it seems legitimate to conclude that at least for the largest sample, the Bayesian approach provides the estimates closest to the real values, values not-known in an experimental situation.

As an extension of our work, we can applied for extropy in place of informational energy. For more details, see the definition of extropy in [15].

Acknowledgment

REFERENCES

1. A. Gelman, and D. B. Rubin, *Inference from iterative simulation using multiple sequences*, Statistical Science, vol. 7, pp. 457–472, 1992.
2. A. Heathcote, and S. Brown, *Reply to Speckman and Rouder: A theoretical basis for QML*, Psychonomic bulletin & review, vol. 11, pp. 577–578, 2004.
3. A. Heathcote, S. Brown, and D. J. K. Mewhort, *Quantile maximum likelihood estimation of response time distributions*, Psychonomic bulletin & review, vol. 9, pp. 394–401, (2002).
4. D. Cousineau, S. Brown, and A. Heathcote, *Fitting distributions using maximum likelihood: Methods and packages*, Behavior Research Methods, Instruments, & Computers, vol. 36, pp. 742–756, 2004.
5. F. C. Donders, *On the speed of mental processes: Attention and Performance II. and Trans*, Reprinted from Acta Psychologica, vol. 30, pp. 412–430, 1896.
6. G. E. Box, and D. R. Cox, *An analysis of transformations*, Journal of the Royal Statistical Society: Series B (Methodological), vol. 26, pp. 211–243, 1964.
7. G. Kulldorff, *Estimation from grouped and partially grouped samples.* Wiley: New York, 1961.
8. G. Petris and L. Tardella, *HI: Simulation from distributions supported by nested hyperplanes. URL:https://cran.r-project.org/web/packages/HI/*, 2015.
9. H. Wainer, *Speed vs reaction time as a measure of cognitive performance*, Memory & Cognition, vol. 5, pp. 278–280, 1997.
10. J.H. Dawson, *Probing structure-function relations in heme-containing oxygenases and peroxidases*, Science, vol. 240, pp. 433–439, 1988.
11. J. N. Rouder, D. Sun, P. L. Speckman, J. Lu, and D. Zhou, *A hierarchical Bayesian statistical framework for response time distributions*, Psychometrika, vol 68, pp. 589–606, 2003.
12. J. N. Rouder, *Are unshifted distributional models appropriate for response time?*, Psychometrika, vol 70, pp. 377–381, 2005.
13. L.S. Christie, and R. D. Luce, *Decision structure and time relations in simple choice behavior*, Bulletin of Mathematical Biophysics, vol. 18, pp. 89–112, 1956.
14. M. Van Selst, and P. Jolicoeur, *A solution to the effect of sample size on outlier elimination*, The Quarterly Journal of Experimental Psychology Section A, vol. 47, pp. 631–650, 1994.
15. O. Kamari, and F. Buono, *On extropy of past lifetime distribution*, Ricerche di Matematica, pp. 1–11, 2020.
16. P. A. Luce, *A computational analysis of uniqueness points in auditory word recognition*, Perception & Psychophysics, vol 39, pp. 155–158, 1986.

17. P. Bonin, B. Laroche, C. Perret, *Locus of word frequency effects in spelling to dictation: Still at the orthographic level*, Journal of Experimental Psychology: Learning, Memory, and Cognition, vol. 42, pp. 1814, 2016.
18. P. D. McCormack, and N. M. Wright, *The positive skew observed in reaction time distributions*, Canadian Journal of Psychology /Revue canadienne de psychologie, vol 18, pp. 43, 1964.
19. R. A. Fisher, *The logic of inductive inference.*, Journal of the royal statistical society, vol. 98, pp. 39–82, 1935.
20. R. H. Hohle, *Inferred components of reaction time as a function of foreperiod duration*, Journal of Experimental Psychology, vol 69, pp. 382–386, 1965.
21. R. Ratcliff, *Methods for dealing with reaction time outliers*, Psychological Bulletin, vol 114, pp. 510–532, 1993.
22. R. Ulrich, and J. Miller, *Effects of truncation on reaction time analysis*, Journal of Experimental Psychology: General, vol. 123, pp. 34, 1994.
23. S. Brown, A. Heathcote, *QMLE: Fast, robust, and efficient estimation of distribution functions based on quantiles*, Behavior Research Methods, Instruments, & Computers, vol. 35, pp. 485–492, 2003.
24. S. L. Burbeck, and R. D. Luce, *Evidence from auditory simple reaction times for both change and level detectors*, Perception & psychophysics, vol. 32, pp. 117–133, 1982.
25. T. Van Zandt, *How to fit a response time distribution*, Psychonomic Bulletin & Review, vol. 7, pp. 424–465, 2000.
26. T. Van Zandt, *Analysis of response time distributions*, Stevens' handbook of experimental psychology, 2002.
27. V. Biscione, *QMLE.zip (https://www.mathworks.com/matlabcentral/fileexchange/46330-qmle-zip)*, MATLAB Central File Exchange. Retrieved December 27, 2020.
28. W. E. Hockley, *Analysis of response time distributions in the study of cognitive processes*, Journal of Experimental Psychology: Learning, Memory, and Cognition, vol 10, pp. 598, 1984.
29. W. J. McGill, *Stochastic latency mechanisms*, Handbook of mathematical psychology, vol 1, pp. 309–360, 1963.
30. W. R. Gilks, N. G. Best, and K. K. C. Tan, *Adaptive rejection Metropolis sampling within Gibbs sampling*, Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 44, pp. 455–472, 1995.
31. W. R. Gilks, R. M. Neal, N. G. Best, and K. K. C. Tan, *"Corrigendum: Adaptive Rejection Metropolis Sampling"*, Applied Statistics, vol. 46, pp. 541–542, 1997.
32. W. Schwarz, *he ex-Wald distribution as a descriptive model of response times.*, Behavior Research Methods, Instruments, & Computers, vol. 33, pp. 457–469, 2001.

A. R Source Code

```
# Some required packages
require(invgamma)
require(HI) # Contain the "arms" function to perform Adaptive Rejection Metropolis Sampling

# Bayesian Function (input: data size, data, number of Samples, burn-in | output: mu, sigma and tau)
BayesianExgaussian <-function(n,x,nSamples=5000,Ti=2500){

  # Initialization of the hyper-parameters
  y1 = (0.2)
  y2 =(2.6)^2
  delta1 =  0.0001
  delta2 =0.1
  tau1 = 0.1
  tau2 = 0.1

  # Defining the vectors of the parameters
  mu <- rep(NA, nSamples)
  zeta <- rep(NA, nSamples)
  lambda <- rep(NA, nSamples)

  # Initialization of the parameters
  mu[1]= mean(x)-(sd(x)*0.8)
  zeta[1]=(var(x)-(sd(x)*0.8)^2)
  lambda[1]=1/(sd(x)*0.8)

  # Lower and upper bounds
  minnmu = min(x)
  maxxmu = (max(x)/2)+1
  minnzeta= 0.01
  maxxzeta= ((max(x)-min(x))/4)^2
  minnlambda= 0.001/mean(x)
  maxxlambda= 100/((max(x)-min(x))/2)

  # The posterior functions
  mufunc<-function(mu1){return(sum(((mu1-x))*lambda[i-1]+sum(log(pnorm(((x-mu1)/(sqrt(zeta[i-1])))
```

```
                -(lambda[i-1]*sqrt(zeta[i-1])))))-(((mu1-y1)^2)/(2*y2^2)))}
zetafunc<-function(zeta1){return(((n*lambda[i-1]^2*zeta1)/(2))+sum(log(pnorm(((x-mu[i])/(sqrt(zeta1)))
                -(lambda[i-1]*sqrt(zeta1)))))-(delta1+1)*log(zeta1)-(delta2/zeta1))}
lambdafunc<-function(lambda1){return((n*log(lambda1))+(sum((mu[i]-x))*lambda1)
                +((n*lambda1^2*zeta[i])/(2))+sum(log(pnorm(((x-mu[i])/(sqrt(zeta[i])))
                -(lambda1*sqrt(zeta[i])))))+(tau1-1)*log(lambda1)-(tau2*lambda1))}

# Defining the vectors of the parameters
for(i in 2:nSamples) {
 mu[i]<-arms(mu[i-1],mufunc,function(mu1) (mu1>minnmu)*(mu1<maxxmu),1)
 zeta[i]<-arms(zeta[i-1],zetafunc,function(zeta1) (zeta1>minnzeta)*(zeta1<maxxzeta),1)
 lambda[i]<-arms(lambda[i-1],lambdafunc,function(lambda1) (lambda1>minnlambda)*(lambda1<maxxlambda),1)
}

# The estimation of the parameters
muest=sum(mu[Ti:nSamples])/(nSamples-Ti)
zetaest=sum(zeta[Ti:nSamples])/(nSamples-Ti)
sigmaest=sqrt(zetaest)
lambdaest=sum(lambda[Ti:nSamples])/(nSamples-Ti)
tauest=1/lambdaest
theta=list(mu=muest,sigma=sigmaest,tau=tauest)
return(theta)
}
```