Stochastic Modeling and Applications
Vol.21, No. 2 (December, 2017), 49-66

MUK PUBLICATIONS
Open Access Publisher

# BAYESIAN ANALYSIS USING MCMC METHODS OF RECORD VALUES BASED ON A NEW GENERALISED RAYLEIGH DISTRIBUTION

ROBERT G. AYKROYD, M. A. W. MAHMOUD, AND HASSAN M. ALJOHANI

Abstract. In this paper, we extend the Rayleigh distribution to create a generalised Rayleigh distribution which is more flexible than the standard. The general properties of the new distribution are derived and investigated, with properties of more standard distributions, such as the exponential, standard Rayleigh and the Weibull, appearing as special cases. Further, we consider maximum likelihood estimation and Bayesian inference under the assumptions of gamma prior distributions on model parameters. Point estimates and confidence intervals based on maximum likelihood estimation are computed. The main challenge, however, is that the Bayesian estimators cannot easily be found and hence, Markov chain Monte Carlo (MCMC) techniques are proposed to generate samples from the posterior distributions leading to approximate posterior inference. The approximate Bayes estimators are compared with the maximum likelihood estimators using simulated data showing dramatic superiority of the Bayesian approach.

## 1. Introduction

The standard Rayleigh distribution (`SRay`) is useful in life testing experiments, as its failure rate is a linear function of time. This distribution was originally introduced by Lord Rayleigh [21, 22] in connection with a problem in the field of acoustics. [18] derived the `SRay` distribution as the probability distribution of the distance from the origin to a point $(X_1, X_2, \ldots, X_n)$ in $n$-dimensional Euclidean space, where the $X_i$'s are independent and identically distributed $N(0, \theta)$ random variables. [6] demonstrated the importance of this distribution in communication engineering and [19] noted that some types of electro-vacuum devices have the feature that their rate of ageing changes with time. [12] presented a brief account of the history and properties of this distribution, with other aspects of this distribution discussed in [17]. [14] computed the modified maximum likelihood estimator for the scale parameter of the `SRay` distribution from doubly censored samples and [5] calculated the maximum likelihood estimator for the one parameter standard Rayleigh distribution based on Type-II censoring. [25] wrote the posterior density of the hazard function and also developed Bayesian interval estimates for the one parameter of the standard Rayleigh distribution. [3] obtained the maximum

2000 *Mathematics Subject Classification.* 62N0, 62F15.
*Key words and phrases.* Bayesian inference; maximum likelihood; life testing; Metropolis-Hastings methods; standard Rayleigh distribution; reliability.

likelihood and the modified moment estimators for the two parameter standard Rayleigh distribution.

Record values and associated statistics are of great importance in several real life problems, such as the analysis of patterns in weather, economics and sports as well as to data relating to the usual physical survival times. Record values appear in many statistical applications and are widely used in statistical modelling and inference where the model can be described as random variables in ascending order of magnitude. Considering models of ordered random variables leads to several models of record values. Motivated by extreme weather conditions, record values can be explained as a model for successive extremes in a sequence of independent and identically distributed random variables. Record values had been extensively examined and many useful properties are known, along with many useful applications.

Suppose that $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables with cumulative distribution function $F(x)$. Let $Y_n = \max(\text{or } \min)\{X_1, \ldots, X_n\}$ for $n \geq 1$. We say $X_j$ is an upper (or lower) record value of $\{X_n, n \geq 1\}$, if $Y_j > (\text{or } <) Y_{j-1}, j > 1$. By definition $X_1$ is an upper as well a lower record value. One can transform the upper record by replacing the original sequence of $\{X_j\}$ by $\{-X_j, j \geq 1\}$ or (if $p(X_i > 0) = 1$ for all $i$) by $\{1/X_i, i \geq 1\}$, then the lower record values of this sequence will correspond to the upper record values of the original sequence. The indices at which the upper record values occur are given by the record times $\{U(n)\}, n > 0$, where $U(n) = \min\{j | j > U(n-1), X_j > X_{U(n-1)}, n > 1\}$ and $U(1) = 1$. The record times of the sequence $\{X_n, n \geq 1\}$ are the same as those for the sequence $\{F(X_n), n \geq 1\}$. Since, $F(X)$ has a uniform distribution, it follows that the distribution of $U(n), n \geq 1$ does not depend on $F$.

The rest of the paper is organized as follows. Section 2 introduces our generalized Rayleigh distribution and gives many statistical properties of the distribution. Section 3 presents recurrence relations for single and product moments as well as single and product moment generating functions of upper record values from the generalized Rayleigh distribution. In Section 4 we derive the maximum likelihood estimators and the Fisher information matrix, and consider asymptotic properties. Section 5 considers Bayesian estimation and the construction of credible intervals. This estimation makes use of the Metropolis-Hastings method which is described in Section 6. Section 7 contains a simulation study in order to give assessment of our proposed methods. Some final comments are given in Section 8.

## 2. Generalized Rayleigh distribution (GRay)

**2.1. Distributional results.** Let $X$ be a random variable having distribution with parameters $\gamma$, $\kappa$ and $\lambda$ which we will denote as $\texttt{GRay}(\gamma, \kappa, \lambda)$, then its probability distribution function (PDF) is given by

$$f(x) = \lambda\,\kappa\,x(x^2-\gamma)^{\kappa-1}\exp\left\{-\frac{\lambda(x^2-\gamma)^{\kappa}}{2}\right\}, \qquad x > \sqrt{\gamma}, \tag{2.1}$$

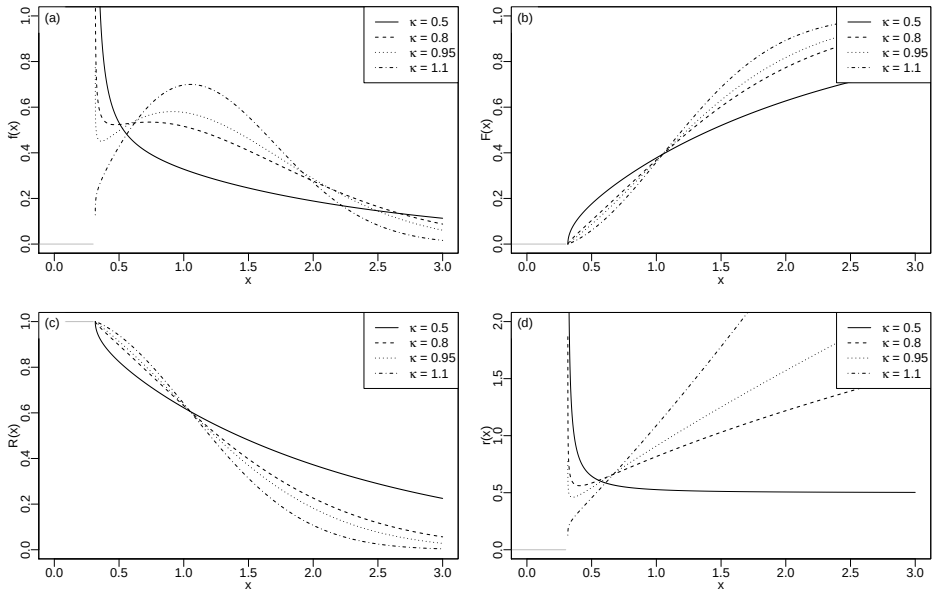


FIGURE 1. Plots of the GRay($\gamma = 0.1, \kappa, \lambda = 1$) for different values of $\kappa$: (a) PDF, (b) CDF, (c) reliability function and (d) failure rate function.

where $\gamma > 0$, $\kappa > 0$ and $\lambda > 0$. From Equation (2.1) it is easy to show that the cumulative distribution function (CDF) is given by

$$F(x) = 1 - \exp\left\{ - \frac{\lambda(x^2 - \gamma)^\kappa}{2} \right\}, \qquad x > \sqrt{\gamma}, \tag{2.2}$$

the reliability, or survivor, function by

$$R(x) = 1 - F(x) = \exp\left\{ - \frac{\lambda(x^2 - \gamma)^\kappa}{2} \right\}, \qquad x > \sqrt{\gamma}, \tag{2.3}$$

and the failure rate function by

$$r(x) = \frac{f(x)}{R(x)} = \lambda\,\kappa\,x(x^2 - \gamma)^{\kappa-1}, \qquad x > \sqrt{\gamma}. \tag{2.4}$$

The behaviour of the density function, cumulative distribution function, failure rate function and the reliability function is shown in Figure (3). These functions shift left and right as $\gamma$ changes, and there is a scale change in the x-direction as $\lambda$ changes, but there are more profound changes with $\kappa$. Here, these figures have fixed $\gamma = 0.1$ and $\lambda = 1.0$, but $\kappa$ is selected to illustrate different shapes.

Note that the derivative of the failure rate function, in Equation (2.4), is

$$\frac{\partial r(x)}{\partial x} = \lambda\kappa(x^2 - \gamma)^{\kappa-2}[(2\kappa - 1)x^2 - \gamma], \quad x > \sqrt{\gamma}. \tag{2.5}$$

This derivative depends on all three parameters $\lambda, \gamma$ and $\kappa$, but the shape only changes as $\kappa$ changes. For $\kappa \geq 1$, one can show that $(2\kappa - 1)x^2 - \gamma > 0$ for all values of $x$, and this shows that $r(x)$ is an increasing function. For $0 < \kappa \leq 0.5$, then $(2\kappa-1)x^2-\gamma < 0$ and this shows that $r(x)$ is a decreasing function. Finally, for $0.5 < \kappa < 1$, then $(2\kappa - 1)x^2 - \gamma < 0$ or $(2\kappa - 1)x^2 - \gamma > 0$ based on the value of $x$, so $r(x)$ has a bathtub shape. The examples in Figure 3 cover these types.

Also, the quantile function, $Q(q)$, can be found by solving $F(Q) = q$, that is

$$1 - \exp\left\{ - \frac{\lambda(Q^2 - \gamma)^\kappa}{2} \right\} = q$$

which leads to the result

$$Q(q) = \sqrt{\left( \frac{2}{\lambda} \ln\left( \frac{1}{1-q} \right) \right)^{1/\kappa} + \gamma}, \qquad 0 \leq q \leq 1. \tag{2.6}$$

Finally, note that values from the `GRay` distribution can be simulated using the probability integral transformation approach, using Equation (2.6), to give

$$x = \sqrt{\left( \frac{2}{\lambda} \ln\left( \frac{1}{1-U} \right) \right)^{1/\kappa} + \gamma}, \qquad \text{where } U \sim \text{Uniform}(0, 1). \tag{2.7}$$

**2.2. Summary measures.** The statistical properties play an important role in the characterization of any distribution. Now some statistical properties of the $\texttt{GRay}(\gamma, \kappa, \lambda)$ are considered.

Mean: To derive the mean of the $\texttt{GRay}(\gamma, \kappa, \lambda)$, consider the definition

$$\mu = \mathrm{E}(X) = [-xR(x)]_{\sqrt{\gamma}}^{\infty} + \int_{\sqrt{\gamma}}^{\infty} R(x)dx. \tag{2.8}$$

After making the substitution $y = (x^2 - \gamma)^k$, then Equation (2.8) becomes

$$\mu = \sqrt{\gamma} + \frac{1}{2\kappa}\Gamma\left(\frac{1}{2\kappa}\right)\left(\frac{2}{\lambda}\right)^{\frac{1}{2\kappa}} + \frac{1}{2\kappa}\sum_{i=1}^{\infty}\binom{-\frac{1}{2}}{i}\gamma^i\eta, \tag{2.9}$$

where $\eta = \Gamma\left(\frac{1}{2\kappa} - \frac{i}{\kappa}\right)\left(\frac{2}{\lambda}\right)^{\frac{1}{2\kappa}-\frac{i}{\kappa}}$.

Variance: Similarly, to find the variance, we have

$$\mathrm{E}(X^2) = \gamma + 2\int_{\sqrt{\gamma}}^{\infty} x \exp\left\{ - \frac{\lambda(x^2 - \gamma)^\kappa}{2} \right\} dx, \tag{2.10}$$

which, using the substitution $y = (x^2 - \gamma)^\kappa$, can be put in the form

$$\mathrm{E}(X^2) = \gamma + \frac{1}{\kappa}\int_0^{\infty} y^{\frac{1}{\kappa}-1} \exp\left\{ - \frac{\lambda}{2}y \right\} dy, = \gamma + \frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\left(\frac{2}{\lambda}\right)^{\frac{1}{\kappa}}. \tag{2.11}$$

From Equations (2.8) and (2.11), the variance of $\texttt{GRay}(\gamma, \kappa, \lambda)$ is given by

$$\mathrm{Var}(X) = \gamma + \frac{1}{\kappa}\Gamma\left(\frac{1}{\kappa}\right)\left(\frac{2}{\lambda}\right)^{\frac{1}{\kappa}} - \left(\sqrt{\gamma} + \frac{1}{2\kappa}\Gamma\left(\frac{1}{2\kappa}\right)\left(\frac{2}{\lambda}\right)^{\frac{1}{2\kappa}} + \frac{1}{2\kappa}\sum_{i=1}^{\infty}\binom{-\frac{1}{2}}{i}\gamma^i\eta\right)^2, \tag{2.12}$$

where, as above, $\eta = \Gamma\left(\frac{1}{2\kappa} - \frac{i}{\kappa}\right)\left(\frac{2}{\lambda}\right)^{\frac{1}{2\kappa}-\frac{i}{\kappa}}$.

Mode: The mode, for $\kappa \geq 1$, is found by equating $\frac{\partial}{\partial x}\ln f(x)$ to zero. This leads to the mode being the solution of the following non-linear equation

$$\frac{1}{x} + \frac{\kappa - 1}{x^2 - \gamma}(2x) - \lambda\kappa x(x^2 - \gamma)^{\kappa-1} = 0. \tag{2.13}$$

This cannot be solved explicitly and hence numerical methods must be used. For other values of $\kappa$, the mode is at the far left, that is at $x = \sqrt{\gamma}$.

Median: The median can simply be found from the quantile function, Equation (2.6), with $q = \frac{1}{2}$ and hence the median is given by

$$\text{Median}(X) = \sqrt{\left(\frac{2}{\lambda}\ln(2)\right)^{1/\kappa} + \gamma}. \tag{2.14}$$

Mean residual life time: Let $m(t)$ denotes the mean residual life time of $\texttt{GRay}(\gamma, \kappa, \lambda)$, then

$$m(t) = \frac{1}{R(t)}\int_t^\infty R(x)dx, \tag{2.15}$$

which can be put in the following form

$$m(t) = \int_{\sqrt{\gamma}}^\infty \exp\left\{\int_t^{t+x} r(y)dy\right\}dx, \tag{2.16}$$

see [26]. Note that for $\kappa \geq 1$, since $r(x)$ is increasing, then $m(t)$ is decreasing, for $0 < \kappa \leq 0.5$, since $r(x)$ is decreasing, then $m(t)$ is increasing. Finally, for $r(x)$ first decreasing and then starting to increase monotonically at some time $x$, means that $m(t)$ is increasing and then decreasing.

## 3. Recurrence equations of upper record value moments

In this section some recurrence relations for the single and product moment and the moment generating function of upper record values from the GRay distribution are stated. Using these recurrence relations, the GRay distribution is characterized.

**Theorem 3.1.** *Single moments of the upper record values: Recurrence relation for single moments of the upper record values from* GRay *distribution is defined in the following theorem.*
*For $n = 1, 2, \ldots$ and $r = 1, 2, \ldots$ we have that*

$$\mu_{n+1}^{r+2} = \left(1 + \frac{r+2}{2\kappa n}\right)\mu_n^{r+2} - \frac{r+2}{2\kappa n}\gamma\mu_n^r, \tag{3.1}$$

*where $\mu_l^s = E(X_{U(l)}^S)$.*

**Theorem 3.2.** *Product moments of the upper record values: Recurrence relation for product moments of the upper record values from GRay distribution* $\texttt{GRay}(\gamma, \kappa, \lambda)$ *are defined in the following theorem.*

(1) *For* $m = 1, 2, \ldots$ *and* $r, s = 1, 2, \ldots$

$$\left(1 + \frac{2\kappa m}{r+2}\right)\mu^{r+2,s}_{m,m+1} = \gamma\mu^{r,s}_{m,m+1} + \frac{2\kappa m}{r+2}\mu^{r+s+2}_{m+1}. \tag{3.2}$$

(2) *For* $1 \leq m \leq n+1$ *and* $r, s = 1, 2, \ldots$

$$\mu^{r+2,s}_{m+1,n} = \left(1 + \frac{r+2}{2\kappa m}\right)\mu^{r+2,s}_{m,n} - \frac{r+2}{2\kappa m}\mu^{r,s}_{m,n}, \tag{3.3}$$

*where* $\mu^{l_1,l_2}_{m,n} = E(X^{l_1}_{U(m)}X^{l_2}_{U(n)})$.

**Theorem 3.3.** *Single moment generating function: Recurrence relation for single moment generating function of the upper record values form* $\texttt{GRay}(\gamma, \kappa, \lambda)$ *is given by*
*For* $n = 1, 2, \ldots$

$$M_{n+1}(t) = -\frac{t}{2\kappa n}\delta'_n(t) + \left(\frac{\gamma t^2}{2\kappa n} + 1\right)M_n(t) - t\delta_n(t) + t\delta_{n+1}(t), \tag{3.4}$$

*where* $M_l(t) = E(e^{tX_{u(l)}})$, $\delta_l(t) = \frac{d}{dt}M_l(t)$, *and* $\delta'_l(t) = \frac{d^2}{dt^2}M_l(t))$.

**Theorem 3.4.** *Joint moment generating functions: Recurrence relation for joint moment generating function of the upper record values form* `GRay` *distribution is stated.*
*For* $n, m = 1, 2, \ldots$

$$\begin{aligned} t_1^2 M''_{n,m}(t_1, t_2) =& (2\kappa m + \gamma t_1^2)M_{n,m}(t_1, t_2) - 2\kappa m t_1 M'_{n,m}(t_1, t_2) \\ &+ 2\kappa m t_1 M'_{n,m+1}(t_1, t_2) - 2\kappa m M_{n,m+1}(t_1, t_2), \end{aligned} \tag{3.5}$$

*where* $M_{n,m}(t_1, t_2)$ *is the joint moment generating function of* $X_{U(n)}$, $X_{U(m)}$ *respectively, then*

$$M_{n,m}(t_1, t_2) = E(e^{t_1 X_{U(n)} + t_2 X_{U(m)}}).$$

## 4. Maximum likelihood analysis

In this section, we estimate $\gamma$, $\kappa$ and $\lambda$, using maximum likelihood and compute the observed Fisher information. Suppose that $\mathbf{x} = \{x_{U_{(1)}}, x_{U_{(2)}}, \ldots, x_{U_{(n)}}\}$ are the first $n$ upper record values from $\texttt{GRay}(\gamma, \kappa, \lambda)$. The general form for a likelihood function for observed record values $\mathbf{x}$, given by [1], is defined as

$$l(\gamma, \kappa, \lambda|\mathbf{x}) = \prod_{i=1}^{n-1} \frac{f(x_{U(i)})}{1 - F(x_{U(i)})} f(x_{U(n)}), \tag{4.1}$$

where $f(\cdot)$ and $F(\cdot)$ are given by Equations (2.1) and (2.2), respectively. Then, substituting from Equations (2.1) and (2.2) into (4.1) gives

$$l(\gamma, \kappa, \lambda|\mathbf{x}) = \begin{cases} \lambda^n \kappa^n \exp\left\{ -\frac{\lambda(x^2_{U(n)} - \gamma)^\kappa}{2} \right\} \times \prod_{i=1}^n x_{U(i)}(x^2_{U(i)} - \gamma)^{\kappa-1}, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } x_{U(1)} < \sqrt{\gamma} \\ 0, \qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{cases} \tag{4.2}$$

The logarithm of the, non-zero part of the, likelihood function in Equation (4.2), that is when $x_{U(1)} < \sqrt{\gamma}$, gives the log-likelihood function

$$L(\gamma, \kappa, \lambda|\mathbf{x}) = n\log\lambda + n\log\kappa - \frac{\lambda(x_{U(n)}^2 - \gamma)^\kappa}{2} + \sum_{i=1}^{n}(\kappa-1)\log(x_{U(i)}^2 - \gamma) + \sum_{i=1}^{n}\log x_{U(i)}. \tag{4.3}$$

Notice that this is monotonic increasing in $\gamma$ and hence the maximum will occur at a boundary of the parameter space for $\gamma$, and in particular $\hat{\gamma} = x_{U(1)}$ that is the minimum of the record value sample. The maximum likelihood estimate of the other two parameters can be found by differentiating Equation (4.3) with respect to $\kappa$ and $\lambda$ and equating the results to zero. This leads to the normal equations for the parameters as

$$\frac{\hat{\lambda}\hat{\kappa}(x_{U(n)}^2 - \hat{\gamma})^{\hat{\kappa}-1}}{2} - \sum_{i=1}^{n}\frac{\hat{\kappa}-1}{x_{U(i)}^2 - \hat{\gamma}} = 0, \tag{4.4}$$

and

$$\frac{n}{\hat{\kappa}} - \frac{\hat{\lambda}}{2}(x_{U(n)}^2 - \hat{\gamma})^{\hat{\kappa}}\log(x_{U(n)}^2 - \hat{\gamma}) + \hat{\kappa}\sum_{i=1}^{n}\log(x_{U(i)}^2 - \hat{\gamma}) = 0. \tag{4.5}$$

The MLEs of $\kappa$ and $\lambda$ can be obtained by simultaneously solving Equations (4.4) and (4.5). However, since Equations (4.4) and (4.5) cannot be solved analytically some numerical method is needed—here the `optim` function in `R` [20] has been used.

In general, the asymptotic variances and covariances of the MLE for parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2)$ say, are given by elements of the inverse of the (expected) Fisher information matrix, $I$, whereas an approximate variance-covariance matrix, $\Sigma_n$, can be obtained from the observed information matrix, $I_n$. That is,

$$\Sigma = \begin{bmatrix} \mathrm{Var}(\hat{\theta}_1) & \mathrm{cov}(\hat{\theta}_1, \hat{\theta}_2) \\ \mathrm{cov}(\hat{\theta}_1, \hat{\theta}_2) & \mathrm{Var}(\hat{\theta}_2) \end{bmatrix} \approx -I_n^{-1}(\hat{\boldsymbol{\theta}}) \tag{4.6}$$

where

$$I_n(\hat{\boldsymbol{\theta}}) = \left[\begin{matrix} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\theta_1^2} & \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2} \\ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\theta_2^2} \end{matrix}\right]\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{4.7}$$

The asymptotic normality of the MLEs, subject to the usual regularity conditions, can be used to compute the approximate confidence intervals for $\theta_1$ and $\theta_2$. Therefore, $(1-\alpha)100\%$ confidence intervals are, respectively,

$$\theta_1 \pm z_{\alpha/2}\ \mathrm{sd}(\hat{\theta}_1), \quad \hat{\theta}_2 \pm z_{\alpha/2}\ \mathrm{sd}(\hat{\theta}_2)$$

where $\mathrm{sd}(\cdot)$ is the standard deviation of the argument and $z_{\alpha/2}$ is the percentile of the standard normal distribution with right-tail probability equal to $\alpha/2$.

As an alternative approach to construction of confidence intervals, consider the asymptotic result, where the parameter vector, $\boldsymbol{\theta}$, is divided into two sets, $\boldsymbol{\theta}^1$ and

$\boldsymbol{\theta}^2$,

$$2\left(L(\hat{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^1, \hat{\boldsymbol{\theta}}^2)\right) \sim \chi^2_d \tag{4.8}$$

where $L(\hat{\boldsymbol{\theta}})$ is the log-likelihood function of the full maximum likelihood estimates, whereas $L(\boldsymbol{\theta}^1, \hat{\boldsymbol{\theta}}^2)$ is the log-likelihood function evaluated with fixed parameters $\boldsymbol{\theta}^1$, and degrees of freedom $d$ is the number of parameters in $\boldsymbol{\theta}^1$. To construct a Wilks confident interval, or region, this can be re-arranged to give

$$\left\{\boldsymbol{\theta}^1 : L(\boldsymbol{\theta}^1, \hat{\boldsymbol{\theta}}^2) \geq L(\hat{\boldsymbol{\theta}}) - \frac{1}{2}\chi^2_d\right\}. \tag{4.9}$$

Particular cases are, for example, the 1D confidence interval for $\boldsymbol{\theta}^1 = \kappa$ giving $d = 1$, with $\boldsymbol{\theta}^2 = (\gamma, \lambda)$, and the 2D confidence region for $\boldsymbol{\theta}^1 = (\kappa, \lambda)$ giving $d = 2$, with $\boldsymbol{\theta}^2 = \gamma$.

Note that although these results can be applied for $\kappa$ and $\lambda$, the non-regularity for $\gamma$ means that an alternative is required. In particular, here the highest density 95% confidence interval for $\hat{\gamma}$, denoted as $(\hat{\gamma}_L, \hat{\gamma}_U)$, can be calculated using the fact that $\hat{\gamma}_U = \hat{\gamma} = x^2_{U(1)}$ and where $\hat{\gamma}_L$ is such that $Pr(\hat{\gamma} \geq \hat{\gamma}_L) = 0.95$. Note that since $x^2_{U(1)} = x^2_1$ with, $x_1 \sim$ `GRay`, the density of $\hat{\gamma}$ is a simple transformation of the `GRay` PDF and hence the value $\hat{\gamma}_L$ can be found easily.

To illustrate maximum likelihood estimation consider Figure 2 which uses a record value dataset of size $n = 5$ from a `GRay`$(\gamma = 0.1, \kappa = 0.5, \lambda = 1)$ distribution. In each of (a)–(c), the profile log-likelihood is shown as a bold curve, showing that $\hat{\gamma}$ is located at the boundary of the parameter space whereas the other parameter estimates occur at turning points. The location of estimated values are shown with a black square and the true values as black triangles. Also shown for each is a confidence interval as a grey bar which has been calculated using the Wilks confidence interval/region result. A bivariate confidence region for $(\kappa, \lambda)$ is shown in (d) superimposed on the 2D log-likelihood surface contours.

## 5. Bayesian estimation

This section describes Bayesian modelling of record values from the `GRay`$(\gamma, \kappa, \lambda)$ distribution with the MCMC algorithm described in the next section. The main idea of the Bayesian approach and computational implementation with MCMC algorithms is to generate samples from the posterior density function and then to compute the Bayes point estimates and also construct the corresponding credible intervals based on the generated posterior samples. By considering the `GRay` model in Equation (2.1), assume the following gamma prior densities for $\gamma$, $\kappa$ and $\lambda$ with parameters $(\alpha_1, \beta_2)$, $(\alpha_2, \beta_2)$ and $(\alpha_3, \beta_3)$ as follows, starting with $\gamma$,

$$\pi(\gamma|\alpha_1, \beta_1) = \frac{\gamma^{\alpha_1 - 1}}{\Gamma(\alpha_1)\beta_1^{\alpha_1}} \exp\left\{-\frac{\gamma}{\beta_1}\right\}, \quad \gamma > 0; \alpha_1, \beta_1 > 0, \tag{5.1}$$

then for $\kappa$,

$$\pi(\kappa|\alpha_2, \beta_2) = \frac{\kappa^{\alpha_2 - 1}}{\Gamma(\alpha_2)\beta_2^{\alpha_2}} \exp\left\{-\frac{\kappa}{\beta_2}\right\}, \quad \kappa > 0; \alpha_2, \beta_2 > 0, \tag{5.2}$$

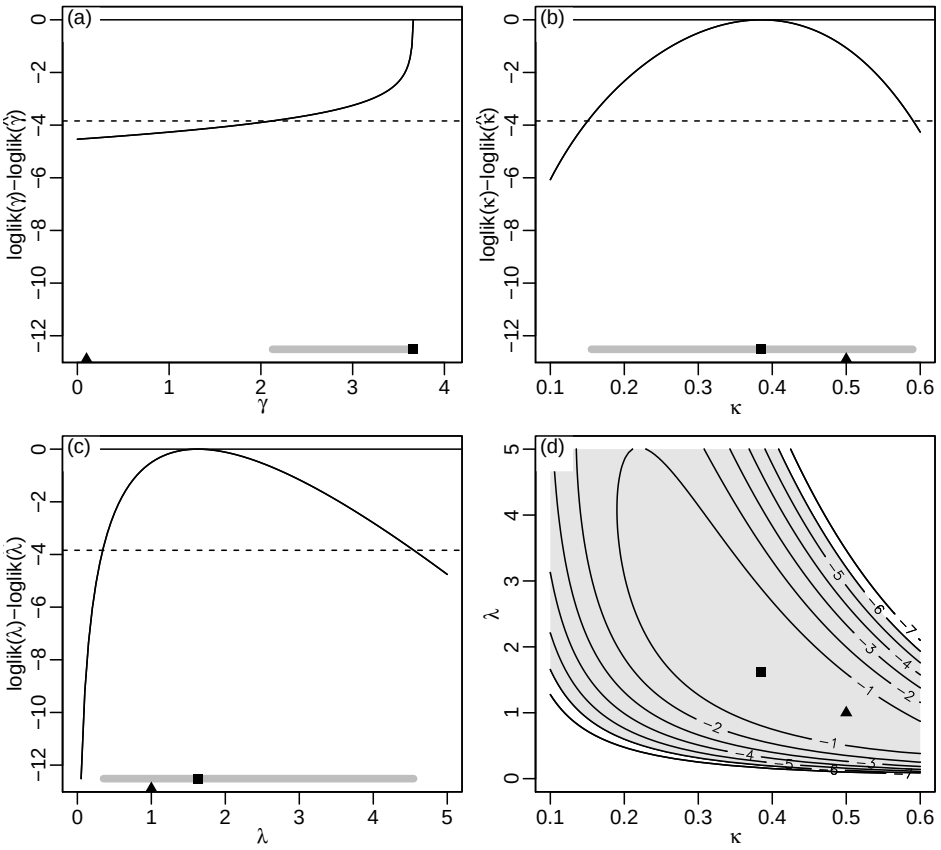


FIGURE 2. Illustration of maximum likelihood estimation showing the profile likelihood functions for: (a) $\gamma$, (b) $\kappa$, and (c) $\lambda$, and in (d) the bivariate profile likelihood for $\kappa$ and $\lambda$. In each, the black square marks the MLE with true value marked as a black triangle, and the grey interval or region shows the approximate 95% confidence interval/region.

and finally for $\lambda$,

$$\pi(\lambda|\alpha_3,\beta_3) = \frac{\gamma^{\alpha_3-1}}{\Gamma(\alpha_3)\beta_3^{\alpha_3}} \exp\left\{-\frac{\lambda}{\beta_3}\right\}, \quad \gamma > 0; \alpha_3, \beta_3 > 0. \tag{5.3}$$

Now, assuming $\gamma$, $\kappa$ and $\lambda$ are independent, then the joint prior density of $\gamma$, $\kappa$ and $\lambda$ can be written as

$$\pi(\gamma,\kappa,\lambda) = \pi(\gamma|\alpha_1,\beta_1)\pi(\kappa|\alpha_2,\beta_2)\pi(\lambda|\alpha_3,\beta_3)$$

$$= \frac{\gamma^{\alpha_1-1}\kappa^{\alpha_2-1}\lambda^{\alpha_3-1}}{\Gamma(\alpha_1)\beta_1^{\alpha_1}\Gamma(\alpha_2)\beta_2^{\alpha_2}\Gamma(\alpha_3)\beta_3^{\alpha_3}} \times \exp\left\{-\left(\frac{\gamma}{\beta_1}+\frac{\kappa}{\beta_2}+\frac{\lambda}{\beta}\right)\right\}. \tag{5.4}$$

This approach follows that of [7], with the assumption that the parameter prior distributions are gamma distribution as suggested by [13]. Note that the hyper-prior parameters, $(\alpha_1, \beta_2)$, $(\alpha_2, \beta_2)$ and $(\alpha_3, \beta_3)$, can be fixed based on expert knowledge or via information from separate calibration experiments. An example of the former is that an expert might be able to provide a mean and variance for a parameter, say $m$ and $v$. Then, as the expectation of the gamma distribution takes form $\mathrm{E}(\theta) = \alpha\beta$, and variance $\mathrm{Var}(\theta) = \alpha\beta^2$, the corresponding hyper-prior parameters can be taken as $\alpha = m^2/v$ and $\beta = v/m$. We have chosen values of $\alpha_1 = \alpha_2 = \alpha_2 = 1$ and $\beta_1 = 1/10$, $\beta_2 = \beta_3 = 1$, as arbitrary values, but particular applications will lead to other choices. In early experimentation, it was found that the only important choice was the value of $\beta_1$ where a prior favouring small values is very worthwhile. Moderate change in the other values has negligible influence on the estimation.

Based on the likelihood function of the observed sample given in Equation (4.2) and the joint prior in Equation (5.4), then the joint posterior density of $\gamma$, $\kappa$ and $\lambda$, given the data, is given by

$$\pi(\gamma, \kappa, \lambda|\mathbf{x}) = \frac{l(\gamma, \kappa, \lambda|\mathbf{x})\pi(\gamma, \kappa, \lambda)}{\int_0^\infty \int_0^\infty \int_0^\infty l(\gamma, \kappa, \lambda|\mathbf{x})\pi(\lambda, \gamma, k)\, d\gamma\, d\kappa\, d\lambda}. \tag{5.5}$$

The Bayes estimator of any function of the parameters $\gamma$, $\kappa$ and $\lambda$, say $g(\gamma, \kappa, \lambda)$, under squared error loss function, is

$$\mathrm{E}_{\gamma,\kappa,\lambda|\mathbf{x}}(g(\gamma, \kappa, \lambda)) = \frac{\int_0^\infty \int_0^\infty \int_0^\infty g(\gamma, \kappa, \lambda) l(\gamma, \kappa, \lambda|\mathbf{x})\pi(\gamma, \kappa, \lambda)\, d\gamma\, d\kappa\, d\lambda}{\int_0^\infty \int_0^\infty \int_0^\infty l(\gamma, \kappa, \lambda|\mathbf{x})\pi(\gamma, \kappa, \lambda)\, d\gamma\, d\kappa\, d\lambda}. \tag{5.6}$$

Evaluating the ratio of the two integrals in Equation (5.6) is too complex and complicated, and hence in this case the MCMC method is proposed to generate samples from the posterior distributions and then compute an approximation to the exact Bayes estimate of $g(\gamma, \kappa, \lambda)$.

The important aspects of the joint posterior are obtained by multiplying the likelihood and the joint prior of $\gamma$, $\kappa$ and $\lambda$, as the normalising constant has no information about the unknown parameters, hence the following statement highlights the key structure,

$$\pi(\gamma, \kappa, \lambda|\mathbf{x}) \propto l(\gamma, \kappa, \lambda|\mathbf{x})\pi(\gamma, \kappa, \lambda)$$

which in our particular case gives

$$\begin{aligned} &= \frac{\gamma^{\alpha_1-1} k^{n+\alpha_2-1} \lambda^{n+\alpha_3-1}}{\Gamma(\alpha_1)\beta_1^{\alpha_1}\Gamma(\alpha_2)\beta_2^{\alpha_2}\Gamma(\alpha_3)\beta_3^{\alpha_3}} \\ &\times \exp\left\{ -\left(\frac{\gamma}{\beta_1} + \frac{k}{\beta_2} + \frac{\lambda}{\beta_3}\right)\right\} \exp\left\{ -\frac{\lambda(x_{U(n)}^2 - \gamma)^k}{2}\right\} \\ &\times \prod_{i=1}^{n} x_{U(i)}(x_{U(i)}^2 - \gamma)^{k-1}. \end{aligned} \tag{5.7}$$

We propose the following MCMC algorithm to draw samples from this posterior density functions, then to compute Bayesian estimates and also to construct the corresponding credible intervals.

## 6. The Metropolis-Hastings method

The estimation of the parameters is based on the approximate posterior distribution computations using a standard Metropolis-Hastings (M-H) algorithm. This is a special case of the Markov chain Monte Carlo (MCMC) approach, whose use has become widespread in the general statistical literature. The M-H algorithm is the first example of a MCMC approach used for parameter estimation and was proposed by [16] and subsequently generalized by [11]. Use of such methods for parameter estimation, and general density exploration, is widespread; a review can be found in [23], and for theoretical details see [8], [15] and [4]. For general practical examples see the collection by [10].

The Markov chain can start at any feasible point in the parameter space, let this arbitrary value be denoted $\boldsymbol{\theta}^0$. From this starting point a discrete time Markov chain is simulated to produce values, $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^K$ say. The algorithm used here is defined as a single-variable random walk MCMC algorithm, see for example [2]. It is one of the simplest schemes, but works well for many applications. It is based on a random walk and uses separate single variable updates. That is, at each step only the value of a single variable is proposed and the proposal is a perturbation of the current value with variance parameter chosen to achieve an acceptable convergence rate. This proposed value is accept with a probability which depends on the posterior distribution. The general structure of the algorithm is given by

(1) Set an initial value for $\boldsymbol{\theta} = \{\gamma, \kappa, \lambda\}$, call this $\boldsymbol{\theta}^0$.
(2) Repeat the following steps for $k = 1, \ldots, K$.

    For $i = 1, 2, 3$, that is for each parameters, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\gamma, \kappa, \lambda)$, in turn.
    (a) Generate a propose new value ${\theta_i}^* = {\theta_i}^k + \epsilon$ where $\epsilon \sim N(0, \tau_i^2)$.
    (b) Evaluate an acceptance probability $\alpha$, as detailed below.
    (c) Generate $u$ from a uniform distribution, $U(0, 1)$.
    (d) If $\alpha > u$ then accept the proposal and set $\theta_i^k = \theta_i^*$, else $\theta_i^k = \theta_i^{k-1}$.
    End

End Repeat

The components of $\boldsymbol{\theta}$ are of the different types, $\gamma$, $\kappa$ and $\lambda$, each allowing different simplifications of the acceptance probability in Step 2(b) above. To explain this, each type will now be considered separately.

Updates of $\gamma$: A proposed new value $\gamma'$ of parameter $\gamma$ is drawn from normal distribution centred on the current parameter value, with variance $\tau_1^2$, chosen to achieve an acceptable convergence rate. Here proposals which are negative or greater than the minimum of the squared data values are rejected, but otherwise the proposal is accepted with probability

$$\alpha(\gamma', \gamma) = \min\left\{1, \frac{l(\gamma', \kappa, \lambda|\mathbf{x})\pi(\gamma')}{l(\gamma, \kappa, \lambda|\mathbf{x})\pi(\gamma)}\right\},$$

otherwise it is rejected and no change is made.

Updates of $\kappa$: A proposed new value $\kappa'$ of parameter $\kappa$ is drawn from normal distribution centred on the current parameter value, with variance $\tau_2^2$, chosen to achieve an acceptable convergence rate. Here negative proposals are rejected, but

if positive the proposal is accepted with probability

$$\min\left\{1, \frac{l(\gamma, \kappa', \lambda|\mathbf{x})\pi(\kappa')}{l(\gamma, \kappa, \lambda|\mathbf{x})\pi(\kappa)}\right\},$$

otherwise it is rejected and no change is made.
Updates of $\lambda$: A proposed new value $\lambda'$ of parameter $\lambda$ is drawn from normal distribution centred on the current parameter value, with variance $\tau_3^2$, chosen to achieve an acceptable convergence rate. Here negative proposals are rejected, but if positive the proposal is accepted with probability

$$\min\left\{1, \frac{l(\gamma, \kappa, \lambda'|\mathbf{x})\pi(\lambda')}{l(\gamma, \kappa, \lambda|\mathbf{x})\pi(\lambda)}\right\},$$

otherwise it is rejected and no change is made.

It is important to realise that both low and high values of $\tau_1^2$, $\tau_2^2$ and $\tau_3^2$, lead to long transient periods and highly correlated samples and hence unreliable estimation [2]. A reasonable proposal variance can usually be chosen adaptively during the early burn-in period, and it has been shown that for a wide variety of problems an acceptance rate of about $20\% - 30\%$ is reasonable [24]. A judgement of when to declare convergence to the equilibrium distribution, and assessment of the efficiency of the algorithm can be made using sample path trace plots and autocorrelation functions. Also, required sample sizes can be calculated (see for example, [2]).

Once the sample has been generated from the posterior distribution, a number of possible estimators are available. After re-labelling, let $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^N$ be the MCMC sample collected after the equilibrium of the Markov chain has been declared, then the posterior mean and variance, for a particular parameter $\theta$, can be estimated by the corresponding sample mean and variance:

$$\widehat{\theta} = \bar{\theta} = \frac{1}{N}\sum_{k=1}^{N}\theta^k, \qquad \widehat{\sigma}^2 = \frac{1}{N-1}\sum_{k=1}^{N}(\theta^k - \bar{\theta})^2.$$

In some cases, for example with very skew distributions or with small sample sizes, it is better to use more robust estimators, such as the posterior median and percentile-based credible intervals. These can easily be obtained by first ordering the sampled values to give $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(N)}$, where $\theta$ represents any of $\gamma$, $\kappa$ and $\lambda$. Then the posterior median estimate is given by

$$\hat{\theta} = \theta_{(N/2)} \tag{6.1}$$

Similarly, the $100(1-\alpha)\%$ credible interval for $\theta$ is given by

$$\left[\theta_{(N\alpha/2)}, \theta_{(N(1-\alpha)/2)}\right].$$

To calculate an approximate MAP estimate, the MCMC algorithm could be converted into a simulated annealing algorithm [9]. In particular a temperature, $T_k$, is included, which decreases as the iterations progress, with $T_k = 2/\log(1+k)$ being one choice of annealing schedule. Hence, the acceptance ratio, $\alpha$, is replaced by $\alpha^{T_k}$. Note that, the MAP estimate is taken as the final iteration, $\widehat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \widehat{\boldsymbol{\theta}}^K$.

The advantage of simulated annealing is that it provides an answer more quickly than a sampling algorithm. On the other hand, the disadvantage is that it does not produce a posterior sample for further investigation [2].

## 7. Experiments

**7.1. Preliminaries.** To evaluate the behaviour of the proposed distribution, different upper record value samples from the `GRay` distribution are simulated and estimates are calculated using the maximum likelihood and Bayesian methods already described. All graphs and calculations have been produced in `R` [20], with code scripts available from the authors.

To generate a dataset of record values, consider the following sequence of values (to 2dp) from the $\texttt{GRay}(0.1, 0.5, 1.0)$ distribution.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **0.87** | 0.64 | **2.26** | 2.17 | **3.81** | 0.48 | 1.83 | 1.69 | 3.22 | 1.76 |
| **7.16** | 0.96 | 0.55 | 1.45 | 4.08 | 4.03 | 2.84 | 0.37 | 0.67 | 3.92 |
| 1.89 | 0.35 | **9.20** | 1.85 | 1.14 | 1.38 | 1.91 | 0.49 | 1.59 | 4.04 |

Those in bold correspond to the record values for a dataset of $n = 5$, that is $U(1) = 1$ with $x_{U(1)} = 0.87$, $U(2) = 3$ with $x_{U(2)} = 2.26$, $U(3) = 4$ with $x_{U(3)} = 3.81$, $U(4) = 11$ with $x_{U(4)} = 7.16$, $U(5) = 23$ with $x_{U(5)} = 9.20$, giving $\mathbf{x} = (0.87, 2.26, 3.81, 7.16, 9.20)$.

Throughout the simulation study $\gamma = 0.1$ and $\lambda = 1$ as these do not have an effect on the distribution shape, but the values $\kappa = 0.5, 0.8, 0.95$ and 1.1, defined as Cases 1–4, as in Figure 2, are considered as representative. Figure 3 shows a summary of $M = 100$ record value datasets, each of size $n = 10$. In each, the grey lines link together the data values within the same dataset. Note that the variability increases with $n$, as well as the mean, and that there is considerable overlap between boxplots. Also, the values are considerably greater in (a), the case with $\kappa = 0.5$, compared to the others.

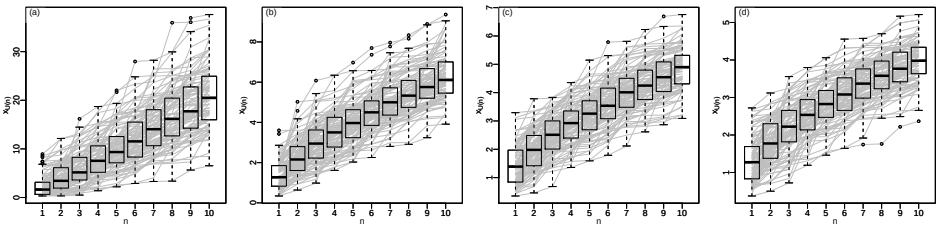


FIGURE 3. Boxplots of record value data, with $n = 10$, from $\text{GRay}(\gamma = 0.1, \kappa, \lambda = 1)$ for different values of $\kappa$ where each boxplot contains $M = 100$ values and values in the same dataset are linked by a grey line: (a) $\kappa = 0.5$, (b) $\kappa = 0.8$, (c) $\kappa = 0.95$, (d) $\kappa = 1.1$.

Each parameter combination will be used with record value sample sizes $n = 5$ and $n = 10$, with the whole simulation process repeated for $M = 100$ replicates to allow a reliable assessment of sampling variability of parameter estimation. For the simulation study based on these upper record values, the maximum likelihood

estimates, asymptotic confidence intervals and Bayesian posterior median with posterior 95% credible interval for $\gamma$, $\kappa$ and $\lambda$ are calculated.

**7.2. Maximum likelihood results.** A collection of $M = 100$ replicates are used to compute different estimates of $\lambda$, $\gamma$ and $\kappa$ with results summarised in Table 1. Recall that in all Cases the true values $\gamma = 0.1$ and $\lambda = 1.0$ are used, but that in: Case 1 $\kappa = 0.5$; Case 2 $\kappa = 0.8$; Case 3 $\kappa = 0.95$; and Case 4 $\kappa = 1.1$.

| | | Case 1 | | Case 2 | | Case 3 | | Case 4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | $n = 5$ | $n = 10$ | $n = 5$ | $n = 10$ | $n = 5$ | $n = 10$ | $n = 5$ | $n = 10$ |
| Maximum likelihood estimation | $\hat{\gamma}$ | 6.36 | 9.88 | 2.65 | 3.18 | 1.94 | 2.46 | 1.73 | 2.08 |
| | Bias | 6.26 | 9.78 | 2.55 | 3.08 | 1.84 | 2.36 | 1.63 | 1.98 |
| | SD | 10.14 | 16.53 | 3.05 | 3.70 | 1.80 | 2.42 | 1.62 | 1.79 |
| | RMSE | 11.87 | 19.14 | 3.96 | 4.80 | 2.57 | 3.38 | 2.29 | 2.66 |
| | Coverage | 0.96 | 0.89 | 0.77 | 0.77 | 0.71 | 0.69 | 0.67 | 0.58 |
| | $\hat{\kappa}$ | 0.34 | 0.40 | 0.41 | 0.52 | 0.44 | 0.56 | 0.46 | 0.60 |
| | Bias | -0.16 | -0.10 | -0.39 | -0.28 | -0.51 | -0.39 | -0.64 | -0.50 |
| | SD | 0.07 | 0.08 | 0.07 | 0.09 | 0.07 | 0.09 | 0.08 | 0.10 |
| | RMSE | 0.18 | 0.13 | 0.40 | 0.29 | 0.51 | 0.40 | 0.65 | 0.51 |
| | Coverage | 0.67 | 0.59 | 0.33 | 0.19 | 0.26 | 0.07 | 0.13 | 0.02 |
| | $\hat{\lambda}$ | 2.38 | 2.12 | 3.63 | 3.17 | 4.27 | 3.79 | 4.89 | 4.43 |
| | Bias | 1.38 | 1.12 | 2.63 | 2.17 | 3.27 | 2.79 | 3.89 | 3.43 |
| | SD | 0.92 | 1.04 | 0.96 | 1.20 | 1.11 | 1.25 | 1.98 | 1.29 |
| | RMSE | 1.66 | 1.52 | 2.80 | 2.48 | 3.45 | 3.06 | 4.36 | 3.66 |
| | Coverage | 0.98 | 0.81 | 0.84 | 0.44 | 0.69 | 0.26 | 0.58 | 0.11 |

TABLE 1. Summary results for $\gamma$, $\kappa$ and $\lambda$ obtained using MLE, averaged over $M = 100$ replicates.

For each replicated data set a triplet of parameter estimates are obtained, producing the complete set $\hat{\gamma}_j$, $\hat{\kappa}_j$, $\hat{\lambda}_j$, for $j = 1, \ldots, M = 100$. The results are then summarised using the following (referring to a general parameter $\theta$): the mean of the estimates, corresponding bias and standard deviation of the estimates

$$\hat{\theta} = \bar{\theta} = \frac{1}{M}\sum_{j=1}^{M}\theta_j, \qquad \text{Bias} = \hat{\theta} - \theta, \qquad \text{SD}^2 = \frac{1}{M-1}\sum_{j=1}^{M}(\theta_j - \bar{\theta})^2,$$

the root mean squared error

$$\text{RMSE}^2 = \frac{1}{M}\sum_{j=1}^{M}(\theta_j - \theta)^2$$

and finally the coverage probability

$$\text{Coverage} = \frac{1}{M}\sum_{j=1}^{M} I_{(\hat{\theta}_L, \hat{\theta}_U)}(\theta)$$

where indicator function $I_{(\hat{\theta}_L, \hat{\theta}_U)}(\theta) = 1$ if $\theta \in (\hat{\theta}_L, \hat{\theta}_U)$ and 0 otherwise.

In all cases the estimation is poor with the key issue regarding the estimation of $\gamma$ which has a knock-on effect on $\kappa$ and $\gamma$ – recalling that $\hat{\gamma} = x^2_{U(1)}$ that is the first record value. As with other non-regular situations, this is a biased estimator and in some cases has produced a very substantial error.

**7.3. Bayesian estimation results.** The same $M = 100$ replicates are used to compute estimates of $\lambda$, $\gamma$ and $\kappa$ using the posterior median and posterior 95% credible intervals, as defined in Section 5, with results summarised in Table 2. Recall, again, that in all Cases the true values $\gamma = 0.1$ and $\lambda = 1.0$ are used, but that in: Case 1 $\kappa = 0.5$; Case 2 $\kappa = 0.8$; Case 3 $\kappa = 0.95$; and Case 4 $\kappa = 1.1$. Typical final proposal standard deviations are 0.2972, 0.4493 and 1.2440 for $\gamma$, $\kappa$ and $\lambda$ respectively. The Markov chain paths, see the examples in Figures 4(a)–(c), show rapid convergence with no discernible trend and good random fluctuations. An initial 100 iterations have been discarded as burn-in, with the remaining $M = 1000$ forming the output sample. Similarly, the autocorrelations function in (d)–(f) show acceptable autocorrelation, although this is more borderline in the case of $\kappa$ in (e). Declaring equilibrium after 100 iterations, the remainder of the sample is used for estimation. The results of the sample size calculations suggested that 63, 587 and 283 iterations are sufficient for the main run, and hence the actual 1000 is more than adequate.

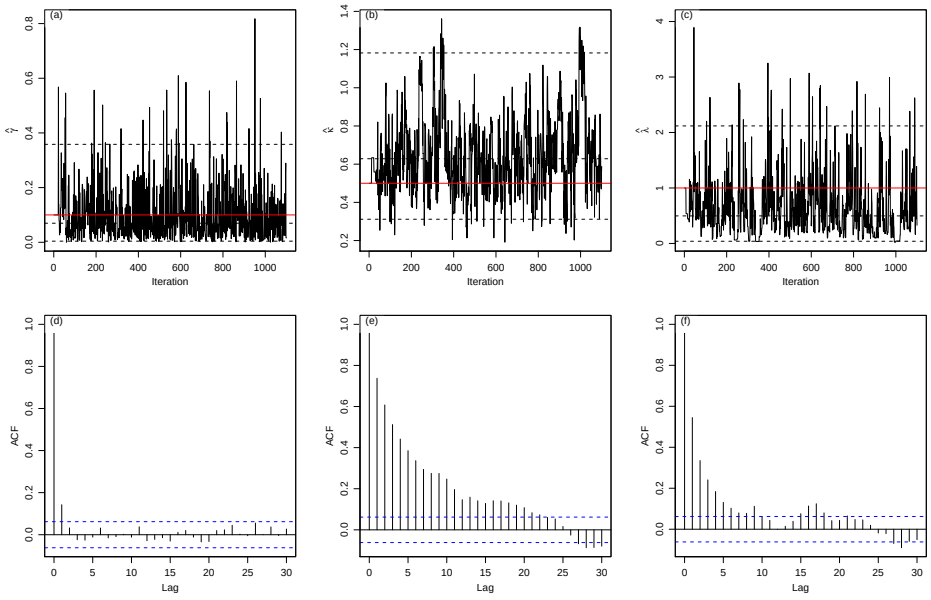


FIGURE 4. Monitoring traces and autocorrelation function generated by the MCMC method.

Figure 5 show example marginal posterior histograms of the model parameters. In (a), the distribution is heavily skew but the majority of values are close to the true value of $\gamma = 0.1$. For $\kappa$, in (b), the distribution is more symmetric with good

concentration around the true value of $\kappa = 0.5$. In (c), again the distribution is skew with many values below the true $\lambda = 1.0$. Overall, these figures indicate that good estimation should be possible, but that all distributions are skew and hence the use of the robust estimators is advisable compared to the more conventional means and variances.

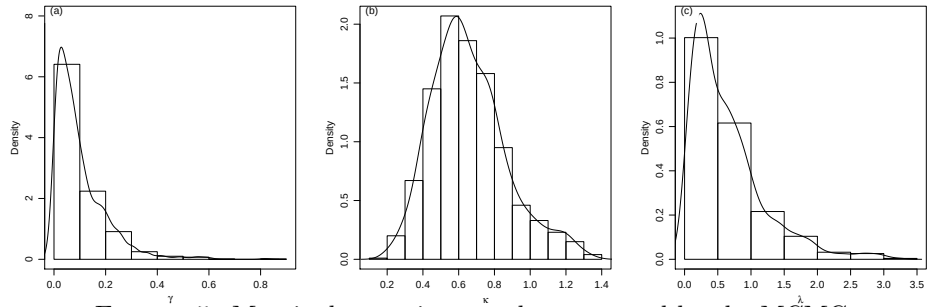


FIGURE 5. Marginal posterior samples generated by the MCMC method summarised using histograms and kernel density curves.

Numerical results are summaries in Table 2 based on posterior median and posterior 95% credible intervals. As with the maximum likelihood estimation, each data produces a triplet of parameter estimates with the results summarised using: the mean estimates, bias, standard deviation, the root mean squared error, and the coverage probability. It is clear that all estimation is now substantial better than with maximum likelihood. In particular biases and RMSE are low with good coverage. There is a slight improvement due to the larger sample size, except for $\gamma$ where there is no change. Recall that in all Cases the true values for $\gamma$ and $\lambda$ are fixed, but that in: Case 1 $\kappa = 0.5$; Case 2 $\kappa = 0.8$; Case 3 $\kappa = 0.95$; and Case 4 $\kappa = 1.1$. In all Cases, there is little change in the estimation properties for $\gamma$ and $\lambda$, but some changes for $\kappa$. Overall, now the estimation is very good and hence the inclusion of prior information has been very successful.

## 8. Summary

In this paper, a new distribution has been proposed with many theoretical properties derived or stated. Methods for parameter estimation from record value data has been developed using the maximum likelihood and Bayesian approaches. In both, care had to be taken because of the non-regular nature of one parameter, and because of non-symmetrical and in particular non-Gaussian parameter sampling distributions. The maximum likelihood approach has been shown to work very badly, but in contrast the proposed Bayesian modelling linked with MCMC estimation has worked very well. Although the Bayesian model included arbitrarily chosen prior parameters the exact values are not overly influential on the results—clearly, a sensitivity analysis would be useful further work.

The application of the standard Rayleigh, and other distributions, to record values and life testing situations is an important branch of statistics and it has previously been found that modifications to standard distributions are often needed

| | | Case 1 | | Case 2 | | Case 3 | | Case 4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | $n=5$ | $n=10$ | $n=5$ | $n=10$ | $n=5$ | $n=10$ | $n=5$ | $n=10$ |
| Posterior median estimates | $\hat{\gamma}$ | 0.08 | 0.08 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 |
| | Bias | -0.02 | -0.02 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| | SD | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| | RMSE | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| | Coverage | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | $\hat{\kappa}$ | 0.59 | 0.56 | 0.92 | 0.86 | 1.12 | 1.02 | 1.23 | 1.17 |
| | Bias | 0.09 | 0.06 | 0.12 | 0.06 | 0.17 | 0.07 | 0.13 | 0.07 |
| | SD | 0.14 | 0.11 | 0.22 | 0.16 | 0.29 | 0.19 | 0.40 | 0.22 |
| | RMSE | 0.16 | 0.12 | 0.24 | 0.17 | 0.33 | 0.20 | 0.42 | 0.23 |
| | Coverage | 1.00 | 0.96 | 1.00 | 0.99 | 1.00 | 0.98 | 0.97 | 0.99 |
| | $\hat{\lambda}$ | 0.75 | 0.82 | 0.86 | 0.88 | 0.88 | 0.90 | 0.96 | 0.92 |
| | Bias | -0.25 | -0.18 | -0.14 | -0.12 | -0.12 | -0.10 | -0.04 | -0.08 |
| | SD | 0.41 | 0.45 | 0.45 | 0.48 | 0.45 | 0.50 | 0.50 | 0.50 |
| | RMSE | 0.47 | 0.48 | 0.47 | 0.50 | 0.46 | 0.51 | 0.50 | 0.50 |
| | Coverage | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |

TABLE 2. Summary results for $\gamma$, $\kappa$ and $\lambda$ obtained using Bayesian analysis, averaged over $M = 100$ replicates.

for different applications. Therefore, the continued development of increasingly more flexible distributions allows the modelling of increasingly more complicated situations. The new generalized Rayleigh distribution proposed in this paper, can now be added to the array available to applied statisticians – and we hope that it will be further studied. Finally, the use of Bayesian modelling has been of great benefit and similar approaches are likely to be helpful in other situations also—this can be another areas of future study.

**Acknowledgment.** The authors thank the Editor and the anonymous referees for the helpful comments.

## References

1. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N.: *Records*, John Wiley & Sons, 1998.
2. Aykroyd, R. G.: Statistical Image Reconstruction, in *Industrial tomography: Systems and Applications*, Woodhead Publishing, pp. 401–424, 2015.
3. Balakrishnan, N. and Cohen, A. C.: *Order Statistics and Inference: Estimation Methods*, Academic Press, 1991.
4. Brooks, S., Gelman, A., Jones, G. and Meng, X-L.: *Handbook of Markov Chain Monte Carlo*, Chapman and Hall/CRC, 2011.
5. Cohen, A. C. and Whitten, B. J.: *Parameter Estimation in Reliability and Life Span Models*, Marcel Dekker, 1988.
6. Dyer, D. D. and Whisenand, C. W.: Best linear unbiased estimator of the parameter of the Rayleigh distribution. Part I: Small sample theory for censored order statistics, *IEEE Transactions on Reliability*, **22**, 27–34, 1973.
7. Fernandez, A. J.: Bayesian inference from Type II doubly censored Rayleigh data, *Statistics & Probability Letters*, **48**, 393–399, 2000.

8. Gamerman, D. and Lopes, H. F.: *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*, 2nd Ed. Chapman and Hall/CRC, 2006.
9. Geman, S. and Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741, 1984.
10. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.: *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, 1996.
11. Hastings, W.K.: Monte Carlo sampling methods using Markov Chains and their applications, *Biometrika* **57**, 97–109, 1970.
12. Hirano, K.: Some properties of the distributions of order k, *Fibonacci Numbers and Their Applications* **42**, 45–61, 1986.
13. Kim, C. and Han, K.: Bayesian estimation of Rayleigh distribution based on generalized order statistics, *Applied Mathematics Sciences* **8**, 7475–7485, 2014.
14. Lee, K.R., Kapadia, C.H. and Brock, D.B.: On estimating the scale parameter of the Rayleigh distribution from doubly censored samples, *Statistische Hefte* **21**, 14–29, 1980.
15. Liu, J.S.: *Monte Carlo strategies in scientific computing*, Springer Science and Business Media, 2001.
16. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Telle, E.: Equation of state calculations by fast computing machines, *The Journal of Chemical Physics* **21**, 1087–1092, 1953.
17. Miller, K.S. and Sackrowitz, H.: Relationships between biased and unbiased Rayleigh distributions, *SIAM Journal on Applied Mathematics* **15**, 1490–1495, 1967.
18. Miller, K.S.: *Multidimensional Gaussian distributions*, John Wiley & Sons Inc, 1964.
19. Polovko, A.M.: *Fundamentals of reliability theory*, Academic Press, 1968.
20. R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016.
21. Lord Rayleigh, On the resultant of a large number of vibrations of the same pitch and of arbitrary phase, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **10**, 73–78, 1880.
22. _____: On the the problem of random vibrations and of random fights in one, two, or three distributions, *Philosophical Magazine* **37**, 321, 1919.
23. Robert, C. and Casella, G.: A short history of Markov Chain Monte Carlo: Subjective recollections from incomplete data, *Statistical Science* **26**, 102–115, 2011.
24. Gelman, A., Gilks, W.R. and Roberts, G.O.: Weak convergence and optimal scaling of random walk Metropolis algorithms, *The Annals of Applied Probability* **7**, 110–120, 1997.
25. Sinha, S.K. and Howlader, H.A.: Credible and hpd intervals of the parameter and reliability of Rayleigh distribution, *IEEE Transactions on Reliability* **32**, 217–220, 1983.
26. Watson G.S. and Wells, W.T.: On the possibility of improving the mean useful life of items by eliminating those with short lives, *Technometrics* **3**, 281–298, 1961.

Robert G. Aykroyd: Department of Statistics, University of Leeds, Leeds, UK
*E-mail address*: r.g.aykroyd@leeds.ac.uk

M. A. W. Mahmoud: Department of Mathematics, Al-Azhar University, Cairo, Egypt
*E-mail address*: mawmahmoud11@yahoo.com

Hassan M. Aljohani: Department of Statistics, University of Leeds, Leeds, UK
*E-mail address*: H.Aljohani09@leeds.ac.uk