



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 12, December 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Bayesian Analysis of the Generalized Half-Cauchy Model: A Comprehensive Study

Dr. Ashwini Kumar Srivastava

Department of Computer Application, Shivharsh Kisan P.G. College, Basti, U.P. India

ABSTRACT: This research paper introduces and explores the Generalized Half-Cauchy (GHC) Model as a novel model for actuarial loss analysis. The GHC Model is a three-parameter extension of the Half-Cauchy Model, offering increased flexibility for modeling positively skewed data with unimodal hazard functions. We present both classical and Bayesian analyses of the GHC Model, including maximum likelihood estimation (MLE), Bayesian model formulation, and posterior analysis. All computational tools are developed in OpenBUGS and R software. A real data set is considered for illustration of the proposed Bayesian approach.

KEYWORDS: Generalized Half-Cauchy Model, actuarial loss modeling, Bayesian analysis, maximum likelihood estimation, survival data

I. INTRODUCTION

Over the past few years, there has been a growing interest in developing new families of probability distributions/models to expand well-established models and enhance their flexibility in practical data modeling. This pursuit has led to the exploration of various techniques to augment existing distributions by introducing additional parameters. A significant focus has been directed towards finding distributions that accurately capture the characteristics of medical survival data, which often exhibit positive skewness and a unimodal hazard function.

In this context, several models such as Log-logistic[1], Lognormal[2], Birnbaum-Saunders[3], Arctangent Survival[4], Generalized Weibull [5], Weibull Extension[6] and Inverse Gaussian[8] have been employed to describe medical survival data. The study also proposes the Generalized Half-Cauchy (GHC) model [9-14], a three-parameter extension of the Half-Cauchy model. The GHC model offers enhanced flexibility for modeling data with a thicker right tail, making it well-suited for positively skewed datasets. The distribution's unimodal hazard shape suggests its applicability in modeling survival data that demonstrates initially increasing and subsequently decreasing death rates.

The remainder of this paper is organized as follows: Section 2 delves into the Model Analysis, providing a detailed discussion of the cumulative distribution function, probability density function, reliability/survival function, hazard rate function, quantile function, and random deviate generation for the GHC model. Section 3 focuses on Maximum Likelihood Estimation (MLE) and the Information Matrix, presenting the log-likelihood function, its differentiation for MLE computation, and the Fisher's information matrix. Section 4 outlines the Bayesian Model Formulation, specifying the probability model, prior distributions, and the data used in the analysis. The Gibbs Sampler implementation for GHC model is detailed, along with the steps for obtaining posterior samples. Section 5 explores the Data Analysis, including the computation of MLE, model validation through Kolmogorov-Smirnov distance, and graphical methods such as Quantile-Quantile (QQ) and Probability-Probability (PP) plots. Bayesian analysis results, convergence diagnostics, and a comparison with MLE are also presented in this section. Additionally, the estimation of hazard and reliability functions, model compatibility assessment through posterior predictive checks, and the conclusion of the study are discussed. Finally, Section 6 concludes the paper by summarizing key findings and suggesting avenues.

II. MODEL DEFINITION AND PROPERTIES

The GHC model is formally defined, and its probability density function (PDF), cumulative distribution function (CDF), reliability/survival function, and hazard rate function are derived.

Cumulative distribution function(cdf):

The distribution function of Generalized Half Cauchy model with three parameters is given by

$$F(x; \alpha, \beta, \theta) = 1 - \left\{ 1 - \frac{2}{\pi} \tan^{-1} \left(\frac{x}{\theta} \right)^\beta \right\}^\alpha$$

where α , β and θ are the parameters. The Generalized Half Cauchy model will be denoted by $GHC(\alpha, \beta, \theta)$.

Probability density function(pdf):

The probability density function is given by

$$f(x; \alpha, \beta, \theta) = \frac{2}{\pi} \frac{\alpha\beta}{\theta} \left(\frac{x}{\theta}\right)^{\beta-1} \left\{ 1 + \left(\frac{x}{\theta}\right)^{2\beta} \right\}^{-1} \left\{ 1 - \frac{2}{\pi} \tan^{-1} \left(\frac{x}{\theta}\right)^\beta \right\}^{\alpha-1}$$

where $(\alpha, \beta, \theta) > 0$ and $x \geq 0$.

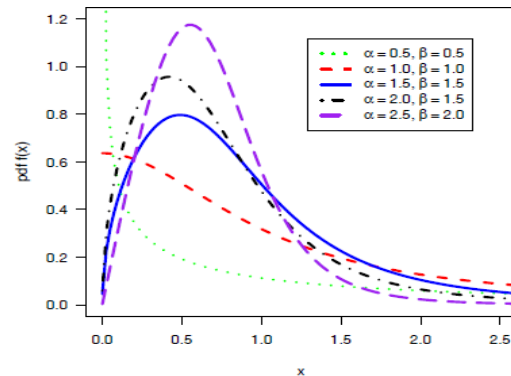


Fig 1 The pdf of GHC model for $\theta = 1$ and different values of α and β .

Some of the typical GHC density functions for different values of α and β for $\theta = 1$ are depicted in Fig 1, which shows that the density function of the GHC model can take different shapes.

The Reliability/Survival function(sf):

The reliability/survival function is

$$R(x; \alpha, \beta, \theta) = \left\{ 1 - \frac{2}{\pi} \tan^{-1} \left(\frac{x}{\theta}\right)^\beta \right\}^\alpha$$

The hazard rate function(hrf):

The hazard rate function is

$$h(x; \alpha, \beta, \theta) = \frac{2}{\pi} \frac{\alpha\beta}{\theta} \left(\frac{x}{\theta}\right)^{\beta-1} \left\{ 1 + \left(\frac{x}{\theta}\right)^{2\beta} \right\}^{-1} \left\{ 1 - \frac{2}{\pi} \tan^{-1} \left(\frac{x}{\theta}\right)^\beta \right\}^{-1}$$

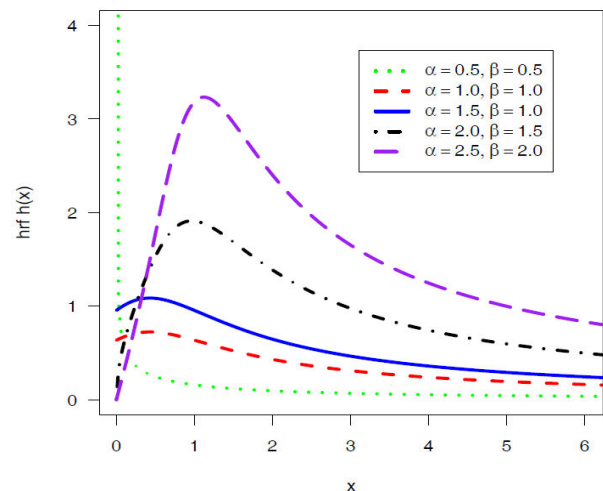


Fig 2 The hrf of GHC model for $\theta = 1$ and different values of α and β .

Some of the typical GHC hazard functions for different values of α and β for $\theta = 1$ are depicted in Fig 2, which shows that the hazard function of the GHC model can take different shapes including the unimodal (upside down bathtub).

The Quantile function:

The quantile function is given by

$$x_p = \theta \left[\tan \left\{ \frac{\pi}{2} \left(1 - (1-p)^{1/\alpha} \right) \right\} \right]^{1/\beta}; 0 < p < 1. \quad (5)$$

The random deviate generation:

The random deviate can be generated from $GHC(\alpha, \beta, \theta)$ by

$$x = \theta \left[\tan \left\{ \frac{\pi}{2} \left(1 - (1-u)^{1/\alpha} \right) \right\} \right]^{1/\beta}; 0 < u < 1.$$

Where u has the $U(0, 1)$ distribution.

We develop some R functions for the computation of pdf, cdf, reliability/survival function, hrf, quantile function and random deviate function respectively.

For model choice based on information criterion, the values of AIC and BIC can be computed. This computation are done by the R function `abic.aks.ghc()` which is developed [15].

III. DATA ANALYSIS

The real data set represents the remission times (in months) of a random sample of 128 bladder cancer patients, [16] is considered for illustration of the proposed methodology. MLE and Bayesian analyses are performed on the data, showcasing the practical application of the GHC model in actuarial loss modelling.

Classical Analysis:

The classical analysis section focuses on the estimation of model parameters using maximum likelihood estimation (MLE). The log-likelihood function is presented, and numerical methods, such as the Newton-Raphson iterative technique, are employed for parameter estimation. Model fit is assessed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)[17].

In this section, we discuss the maximum likelihood estimators (MLE's) of the Generalized Half Cauchy model and their asymptotic properties to obtain approximate confidence intervals based on MLE's[18].

Let $\underline{x} = (x_1, \dots, x_n)$ be a random sample of size n from $GHC(\alpha, \beta, \theta)$, then the log-likelihood function $\ell(\alpha, \beta, \theta | \underline{x})$ can be written as;

$$\ell = n \log\left(\frac{2}{\pi}\right) + n \log\left(\frac{\alpha\beta}{\theta}\right) + (\beta - 1) \sum_{i=1}^n \log\left(\frac{x_i}{\theta}\right) - \sum_{i=1}^n \log\left\{1 + \left(\frac{x_i}{\theta}\right)^{2\beta}\right\} + (\alpha - 1) \sum_{i=1}^n \log\left\{1 - \frac{2}{\pi} \tan^{-1}\left(\frac{x_i}{\theta}\right)^\beta\right\}$$

Differentiating with respect to α , β and θ , we have

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log\left\{1 - (2/\pi) \tan^{-1}(x_i/\theta)^\beta\right\}$$

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^n \log(x_i/\theta) - 2 \sum_{i=1}^n \frac{\log(x_i/\theta)}{\{1 + (x_i/\theta)^{2\beta}\}} - \frac{2(\alpha - 1)}{\pi} \sum_{i=1}^n \frac{(x_i/\theta)^\beta \log(x_i/\theta)}{\{1 - (2/\pi) \{1 + (x_i/\theta)^{2\beta}\} \tan^{-1}(x_i/\theta)^\beta\}}$$

and

$$\frac{\partial \ell}{\partial \theta} = -\frac{n}{\theta} - \frac{n(\beta - 1)}{\theta} + \frac{2\beta}{\theta} \sum_{i=1}^n \frac{(x_i/\theta)^{2\beta}}{\{1 + (x_i/\theta)^{2\beta}\}} + \frac{2\beta(\alpha - 1)}{\pi\theta} \sum_{i=1}^n \frac{(x_i/\theta)^\beta \{1 + (x_i/\theta)^{2\beta}\}^{-1}}{\{1 - (2/\pi) \tan^{-1}(x_i/\theta)^\beta\}}$$

Therefore, to obtain the MLE's of α , β and θ , we can maximize above equation directly with respect to α , β and θ or setting these equations to zero and solving them simultaneously yield the maximum likelihood estimates (MLEs) of the model parameters. Numerical methods can be used to obtain the ML estimates of the parameters. Here, the Newton-Raphson iterative technique could be applied to solve the likelihood equations numerically[19].

Let us denote the parameter vector by $\underline{\delta} = (\alpha, \beta, \theta)$ and the corresponding MLE of $\underline{\delta}$ as $\hat{\underline{\delta}} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$, then the asymptotic normality results in

$$(\hat{\underline{\delta}} - \underline{\delta}) \rightarrow N_3(0, (I(\underline{\delta}))^{-1})$$

where $I(\underline{\delta})$ is the Fisher's information matrix given by

$$I(\underline{\delta}) = - \begin{pmatrix} E\left(\frac{\partial^2 \ell}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \beta}\right) & E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \theta}\right) \\ E\left(\frac{\partial^2 \ell}{\partial \beta \partial \alpha}\right) & E\left(\frac{\partial^2 \ell}{\partial \beta^2}\right) & E\left(\frac{\partial^2 \ell}{\partial \beta \partial \theta}\right) \\ E\left(\frac{\partial^2 \ell}{\partial \theta \partial \alpha}\right) & E\left(\frac{\partial^2 \ell}{\partial \theta \partial \beta}\right) & E\left(\frac{\partial^2 \ell}{\partial \theta^2}\right) \end{pmatrix}$$

In practice, it is useless that the MLE has asymptotic variance $(I(\underline{\delta}))^{-1}$ because we do not know $\underline{\delta}$. Hence, we approximate the asymptotic variance by "plugging in" the estimated value of the parameters. The common procedure is to use observed Fisher information matrix $O(\hat{\underline{\delta}})$ (as an estimate of the information matrix $I(\underline{\delta})$) given by

$$O(\hat{\underline{\delta}}) = - \begin{pmatrix} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \beta} & \frac{\partial^2 \ell}{\partial \alpha \partial \theta} \\ \frac{\partial^2 \ell}{\partial \beta \partial \alpha} & \frac{\partial^2 \ell}{\partial \beta^2} & \frac{\partial^2 \ell}{\partial \beta \partial \theta} \\ \frac{\partial^2 \ell}{\partial \theta \partial \alpha} & \frac{\partial^2 \ell}{\partial \theta \partial \beta} & \frac{\partial^2 \ell}{\partial \theta^2} \end{pmatrix}_{(\hat{\alpha}, \hat{\beta}, \hat{\theta})} = -H(\underline{\delta}) \Big|_{\underline{\delta}=\hat{\underline{\delta}}}$$

where H is the Hessian matrix, $\underline{\delta} = (\alpha, \beta, \theta)$ and $\hat{\underline{\delta}} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$. The Newton-Raphson algorithm to maximize the likelihood produces the observed information matrix. Therefore, the variance-covariance matrix is given by

$$\left(-H(\underline{\delta}) \Big|_{\underline{\delta}=\hat{\underline{\delta}}}\right)^{-1} = \begin{pmatrix} \text{var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\beta}) & \text{cov}(\hat{\alpha}, \hat{\theta}) \\ \text{cov}(\hat{\beta}, \hat{\alpha}) & \text{var}(\hat{\beta}) & \text{cov}(\hat{\beta}, \hat{\theta}) \\ \text{cov}(\hat{\theta}, \hat{\alpha}) & \text{cov}(\hat{\theta}, \hat{\beta}) & \text{var}(\hat{\theta}) \end{pmatrix}$$

Hence, from the asymptotic normality of MLEs, approximate $100(1-\gamma)\%$ confidence intervals for α , β and θ can be constructed as

$$\hat{\alpha} \pm z_{\gamma/2} \sqrt{\text{var}(\hat{\alpha})} \quad \hat{\beta} \pm z_{\gamma/2} \sqrt{\text{var}(\hat{\beta})} \quad \text{and} \quad \hat{\theta} \pm z_{\gamma/2} \sqrt{\text{var}(\hat{\theta})}$$

where $z_{\gamma/2}$ is the upper percentile of standard normal variate.

IV. BAYESIAN MODEL FORMULATION

The Bayesian perspective is introduced, specifying prior distributions for the GHC model parameters. The Gibbs sampling technique is applied to obtain Bayesian estimators of parameters, posterior variances, and credible intervals. We discuss the incorporation of the GHC model into the OpenBUGS software for Bayesian analysis [20].

The Bayesian model is constructed by specifying the prior distributions for the model parameters α , β and θ , and then multiplying with the likelihood function to obtain the posterior distribution function.

- Probability Model : $f(x | \alpha, \beta, \theta)$
- Prior distribution : $p(\alpha, \beta, \theta)$
- Data : $\underline{x} = (x_1, \dots, x_n)$

Given a set of data $\underline{x} = (x_1, \dots, x_n)$, the likelihood function is

$$L(\alpha, \beta, \theta | \underline{x}) = \left(\frac{2}{\pi}\right)^n \left(\frac{\alpha\beta}{\theta}\right)^n \left(\prod_{i=1}^n \left\{1 + \left(\frac{x_i}{\theta}\right)^{2\beta}\right\}^{-1}\right) \left(\prod_{i=1}^n \left(\frac{x_i}{\theta}\right)^{\beta-1}\right) \left(\prod_{i=1}^n \left\{1 - \frac{2}{\pi} \tan^{-1}\left(\frac{x_i}{\theta}\right)^\beta\right\}^{\alpha-1}\right)$$

Denote the prior distribution of α , β and θ as $p(\alpha, \beta, \theta)$. The joint posterior is

$$p(\alpha, \beta, \theta | \underline{x}) = L(\alpha, \beta, \theta | \underline{x}) p(\alpha) p(\beta) p(\theta)$$

Prior distributions:

We assume the independent gamma priors for $\alpha \sim G(a_1, b_1)$ and $\beta \sim G(a_2, b_2)$, and uniform prior for $\theta \sim U(a_3, b_3)$ as

$$p(\alpha) = \frac{b_1^{a_1}}{\Gamma(a_1)} \alpha^{a_1-1} e^{-b_1\alpha} ; \alpha > 0, (a_1, b_1) > 0'$$

$$p(\beta) = \frac{b_2^{a_2}}{\Gamma(a_2)} \beta^{a_2-1} e^{-b_2\beta}; \beta > 0, (a_2, b_2) > 0'$$

and

$$p(\theta) = \frac{1}{b_3 - a_3} ; a_3 < \theta < b_3'$$

Posterior distribution:

Combining the likelihood function with the prior via Bayes' theorem yields the posterior up to proportionality as

$$p(\alpha, \beta, \theta | \underline{x}) \propto \alpha^{a_1+n-1} \beta^{a_2+n-1} \theta^{-n} \exp(-b_1\alpha - b_2\beta) T_1 T_2 T_3$$

where

$$T_1 = \prod_{i=1}^n \left(\frac{x_i}{\theta}\right)^{\beta-1}; \quad T_2 = \prod_{i=1}^n \left\{1 + \left(\frac{x_i}{\theta}\right)^{2\beta}\right\}^{-1}; \quad \text{and} \quad T_3 = \prod_{i=1}^n \left\{1 - \frac{2}{\pi} \tan^{-1}\left(\frac{x_i}{\theta}\right)^\beta\right\}^{\alpha-1}.$$

The posterior is obviously complicated and no close form inferences appear possible. We, therefore, propose to consider MCMC methods to simulate samples from the posterior so that sample-based inferences can be easily drawn.

Markov chain Monte Carlo draws samples by running a cleverly constructed Markov chain that eventually converges to the target distribution (called stationary or equilibrium) which, in our case, is the posterior distribution $p(\alpha, \beta, \theta | \underline{x})$.

Gibbs Sampler : Algorithm

For Gibbs sampler implementation [21, 22], the full conditionals for α , β and θ upto proportionality can be specified as

- (i) Full conditional distribution of the parameter α for given β , θ and \underline{x}

$$p(\alpha | \beta, \theta, \underline{x}) \propto \alpha^{a_1+n-1} \exp(-b_1\alpha) T_3$$

(ii) Full conditional distribution of the parameter β for given α , θ and \underline{x}

$$p(\beta | \alpha, \theta, \underline{x}) \propto \beta^{a_2 + n - 1} \exp(-b_2 \beta) T_2 T_3$$

(iii) Full conditional distribution of the parameter θ for given α , β and \underline{x}

$$p(\theta | \alpha, \beta, \underline{x}) \propto \theta^{-n} T_1 T_2 T_3$$

We shall use OpenBUGS software to obtain posterior samples. A module *aks.dhc(alpha, beta, theta)* is written in Component Pascal for Generalized Half Cauchy to perform full Bayesian analysis in OpenBUGS using the method described in [23, 24]. It is important to note that this module can be used for any set of suitable priors of the model parameters. Almost all aspects of the model in Bayesian framework can be studied using the developed module *aks.dhc(alpha, beta, theta)*

Gibbs Sampler : Implementation

1. Select an initial value $\underline{\delta}^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \theta^{(0)})$ to start the chain.
2. Suppose at the *i*th-step, $\underline{\delta} = (\alpha, \beta, \theta)$ takes the value $\underline{\delta}^{(i)} = (\alpha^{(i)}, \beta^{(i)}, \theta^{(i)})$ then from full conditionals, we generate $\alpha^{(i+1)}$ from $p(\alpha | \beta^{(i)}, \theta^{(i)}, \underline{x})$, $\beta^{(i+1)}$ from $p(\beta | \alpha^{(i+1)}, \theta^{(i)}, \underline{x})$ and $\theta^{(i+1)}$ from $p(\theta | \alpha^{(i+1)}, \beta^{(i+1)}, \underline{x})$
3. This completes a transition from $\underline{\delta}^{(i)}$ to $\underline{\delta}^{(i+1)}$
4. Repeat Step 2, *N* times.

MCMC OUTPUT : POSTERIOR SAMPLE

It is well known that rapid convergence is facilitated by choosing appropriate starting values. In order to guarantee the convergence and to remove the effect of the selection of initial value, the first '*B*' simulate dvariates are discarded. Also to reduce the effect of autocorrelation, select a sampling lag *L* > 1 after which the corresponding autocorrelation are low. Consider $(\underline{\delta}^{(1)}, \dots, \underline{\delta}^{(j)}, \dots, \underline{\delta}^{(M)})$ as the MCMC output (posterior sample) for the posterior analysis

$$\underline{\delta}^{(j)} = (\alpha^{(j)}, \lambda^{(j)}, \theta^{(j)}); j = 1, 2, \dots, M.$$

Thus, MCMC output is referred as the sample after removing the initial iterations (produced during the burn-in period) and considering the appropriate *lag*, which can be used to develop the Bayesian inference.

The Bayes estimates of $\underline{\delta} = (\alpha, \beta, \theta)$, under the square error loss (SEL) function, are given by

$$\hat{\alpha} = \frac{1}{M} \sum_{j=1}^M \alpha^{(j)}; \hat{\beta} = \frac{1}{M} \sum_{j=1}^M \beta^{(j)}; \hat{\theta} = \frac{1}{M} \sum_{j=1}^M \theta^{(j)}$$

The Bayes estimates under absolute and zero-one loss functions are posterior median and mode, respectively.

POSTERIOR ANALYSIS:

The posterior analysis section presents numerical and visual summaries of the Bayesian estimates, including posterior means, modes, standard deviations, quantiles, and credible intervals for model parameters. Convergence diagnostics, such as trace plots and the Brooks-Gelman-Rubin diagnostic, ensure the validity of the Bayesian analysis.

A. Classical Analysis

The estimation of the parameter of proposed model is obtained by the method of maximum likelihood (ML) estimation. To check the validity of the model, we compute the Kolmogorov-Smirnov (KS) distance between the empirical distribution function and the fitted distribution function when the parameters are obtained by method of maximum likelihood. The well known graphical method Probability-Probability (PP) plot is also used for investigating the suitability of the model.

Computation of MLE

The maximum likelihood estimates (MLEs) are obtained by direct maximization of the log-likelihood function $\ell(\alpha, \beta, \theta)$ discussed earlier. The advantage of this procedure is that it runs immediately using existing statistical packages such as R [25]. We consider the software R through the Quasi-Newton algorithm [26] to compute the MLEs.

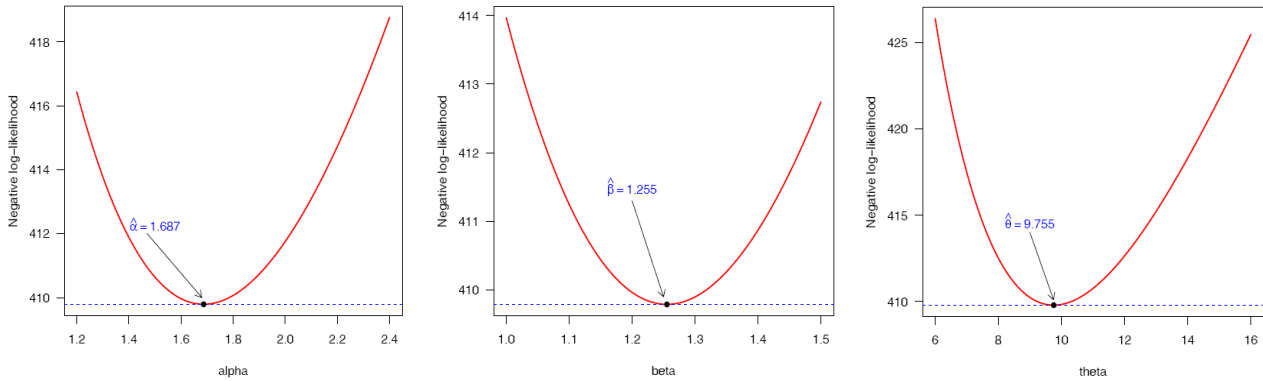


Fig 3 The profile negative log-likelihood plots of α , β and θ

From Fig 3, we show that the likelihood equations have a unique solution. The Table 1 shows the ML estimates, standard error (SE) and 95% Confidence Intervals for parameters α , β and θ . The maximized value of loglikelihood is $\ell(\hat{\alpha}, \hat{\beta}, \hat{\theta}) = -409.7898$.

Table 1 MLE, standard error and 95% confidence interval

Parameter	MLE	Std. Error	95% CI
alpha	1.6869	0.4847	(0.7369, 2.6369)
beta	1.2553	0.1272	(1.0060, 1.5046)
theta	9.7552	2.6928	(4.4773, 15.0331)

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be help us to determine which model is most appropriate for the given data. For the given set, we have AIC=825.5795 and BIC=834.1356. These computation have to made by developing a R function *abic.aks.ghc()*.

Model Validation

To check the validity of the model [27], we compute the Kolmogorov-Smirnov (KS) distance between the empirical distribution function and the fitted distribution function when the parameters are obtained by method of maximum likelihood is 0.0340 and the corresponding p-value is 0.9985. We have developed the R function *ks.aks.ghc()* for computation.

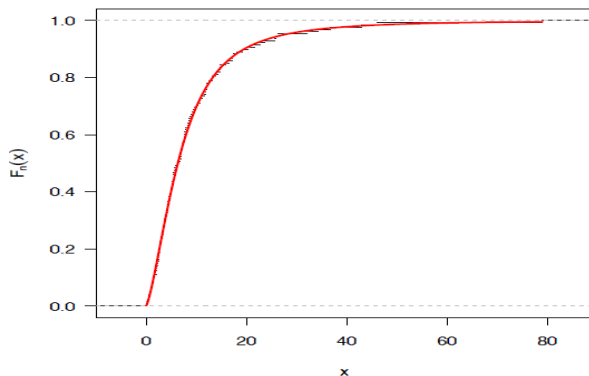


Fig 4 The graph of empirical and fitted distribution function.

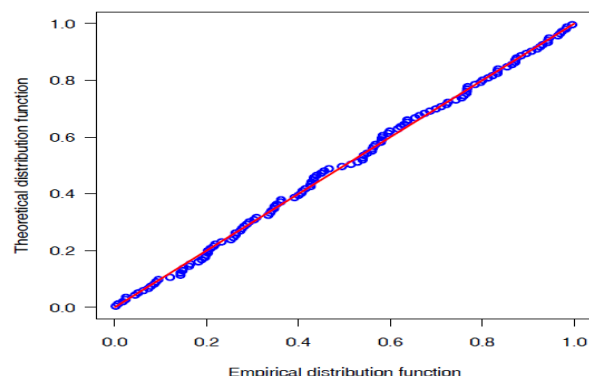


Fig 5 P-P plot using MLEs as estimate.

We have plotted the empirical distribution function and the fitted distribution function in Fig 4 which clearly shows that the fitted Generalized Half Cauchy model provides nice to the given data.

A further validation of this model finding can be obtained by inspecting the probability-probability (P-P) plots (Srivastava and Kumar, 2015)

A P-P plot depicts the points $(F(x_{(i)}), F(x_{(i)}; \hat{\delta}))$; $i = 1, 2, \dots, n$ where $\hat{\delta} = (\hat{\alpha}, \hat{\beta}, \hat{\theta})$ and

$x_{(i)}$ are the order statistics, $F_n(x) = (1/n) \sum_{i=1}^n I(X \leq x)$ is the empirical distribution function and $I(\cdot)$ is the indicator function. The corresponding developed R function `pp.aks.ghc()` for using computation. The P-P plot for the fitted model are shown in Fig 5. It is evident that the fit of the Generalized Half Cauchy model is good.

B. Bayesian Analysis

Using module `aks.dhc(alpha, beta, theta)` for this analysis, we assume the independent uniform prior for $\theta \sim U(a_3, b_3)$ and gamma priors for $\alpha \sim G(a_1, b_1)$ and $\beta \sim G(a_2, b_2)$ with hyper parameter values

$$(a_1 = 0.001, b_1 = 0.001), (a_2 = 0.001, b_2 = 0.001) \text{ and } (a_3 = 0, b_3 = 50.0).$$

We run the model to generate two Markov Chains at the length of 30,000 with different starting points of the parameters. We have chosen initial values $(\alpha = 1.0, \beta = 0.5, \theta = 5.0)$ for the first chain and $(\alpha = 5.0, \lambda = 3.0, \theta = 20.0)$ for the second chain. The convergence is monitored using trace and ergodic mean plots, we find that the Markov Chain converge together after approximately 2000 observations. Therefore, burn-in of 5000 samples is more than enough to erase the effect of starting point(initial values). Finally, samples of size 5000 are formed from the posterior by picking up equally spaced every fifth outcome (to minimize the auto correlation among the generated deviates.), i.e. `thin=5`, starting from 5001.

Therefore, we have the posterior sample $(\alpha_1^{(j)}, \beta_1^{(j)}, \theta_1^{(j)}) ; j = 1, \dots, 5000$ from chain 1 and $(\alpha_2^{(j)}, \beta_2^{(j)}, \theta_2^{(j)}) ; j = 1, \dots, 5000$ from chain 2.

The chain 1 is considered for convergence diagnostics plots. The visual summary is based on posterior sample obtained from chain 1 whereas the numerical summary is presented for both the chains.

Convergence diagnostics

Before examining the parameter estimates or performing other inference, it is a good idea to look at plots of the sequential(dependent) realizations of the parameter estimates and plots thereof. The sequential plot of parameters is the plot that most often exhibits difficulties in the Markov chain. Fig 6 shows the sequential realizations of the parameters of the model.

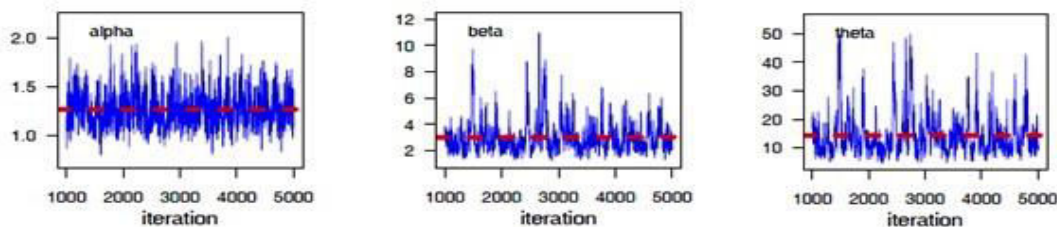


Fig 6 Sequential realization of the parameters α , β and θ .

Visual summary

The visual graphs include the boxplot, density strip plot, histogram, marginal posterior density estimate and rug plots for the parameters. We have also superimposed the 95% HPD intervals.

These graphs provide almost complete picture of the posterior uncertainty about the parameters. We have used the posterior sample $(\alpha_1^{(j)}, \beta_1^{(j)}, \theta_1^{(j)}) ; j = 1, \dots, 5000$ to draw these graphs.

The density strip illustrates a univariate distribution as a shaded rectangular Histograms can provide insights on skewness, behavior in the tails, presence of multi-modal behavior, and data outliers; histograms can be compared to the fundamental shapes associated with standard analytic distributions. Fig 7 represents the histogram, marginal posterior density for α (left panel) and boxplot and density strip plot (right panel). strip, whose darkness at a point is proportional to the probability density [28].

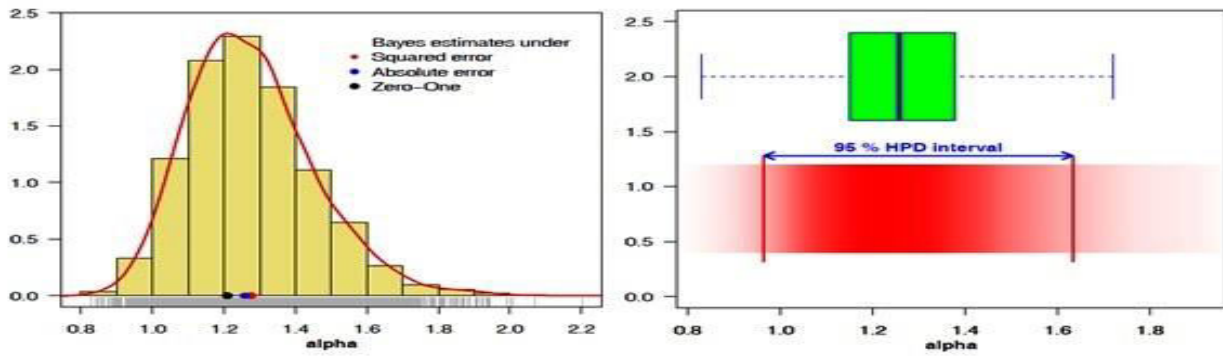


Fig 7 Left panel : Histogram and marginal posterior density, Right panel : boxplot density strip of and 95% HPD interval of α

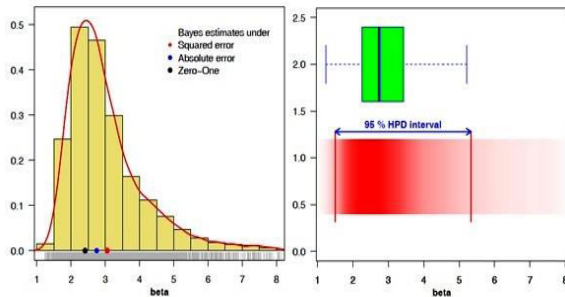


Fig 8 Left panel : Histogram and marginal posterior density, Right panel : boxplot density strip of and 95% HPD interval of β

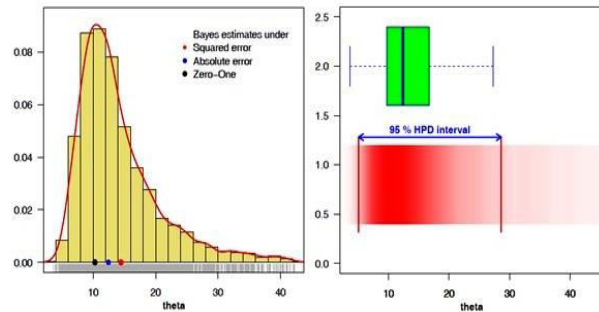


Fig 9 Left panel : Histogram and marginal posterior density, Right panel : boxplot density strip of and 95% HPD interval of θ

The kernel density estimates have been drawn using R with the assumption of Gaussian kernel and properly chosen values of the bandwidths. It can be seen that α is symmetric whereas β and θ show positive skewness. We have plotted the similar graphs for β and θ displayed in Fig 8 and 9.

Numerical Summary

The numerical summary is presented for $(\alpha_1^{(j)}, \beta_1^{(j)}, \theta_1^{(j)}) ; j = 1, \dots, 5000$ from chain 1 and $(\alpha_2^{(j)}, \beta_2^{(j)}, \theta_2^{(j)}) ; j = 1, \dots, 5000$ from chain 2.

We have considered various quantities of interest and their numerical values based on MCMC sample of posterior characteristics for Generalized Half-Cauchy model. The MCMC results of the posterior mean, mode, standard deviation(SD), first quartile, median, third quartile, 2.5th percentile, 97.5th percentile, skewness of the parameters α , β and θ are displayed in Table 2.

Characteristics	Chain 1			Chain 2		
	alpha	lambda	theta	alpha	lambda	theta
Mean	2.445	1.204	14.422	2.539	1.194	14.988
SD	1.316	0.135	7.906	1.340	0.135	8.063
2.5th Percentile	1.034	0.971	6.287	1.076	0.963	6.363
First Quartile (Q ₁)	1.582	1.107	9.277	1.650	1.099	9.619
Median	2.043	1.192	11.930	2.144	1.181	12.500
Third Quartile(Q ₃)	2.839	1.286	16.743	2.981	1.275	17.565
97.5th Percentile	6.331	1.500	37.942	6.435	1.489	39.197
Mode	1.678	1.174	9.803	1.714	1.161	9.880
Skewness	1.915	0.478	1.893	1.870	0.551	1.784
Kurtosis	4.260	0.319	3.858	4.261	0.417	3.436

Table 2 : Numerical summaries based on MCMC sample of posterior characteristics.

The advantage of using the MCMC method over the MLE method is that we can always obtain a reasonable interval estimate of the parameters by constructing the probability intervals based on the empirical posterior distribution. This is often unavailable in maximum likelihood estimation.

The algorithm described by Chen and Shao [29] is used to compute the HPD intervals under the assumption of unimodal marginal posterior distribution. The width of the HPD is another way of measuring uncertainty of beliefs. If

the HPD is wide, then beliefs are uncertain. If the HPD is narrow, then beliefs are fairly certain. Table 3 shows the HPD and Credible intervals for α , β and θ parameters.

Table 3 95% symmetric and HPD credible intervals

Parameter	Symmetric Credible Interval	HPD Credible Interval
alpha	(1.034, 6.331)	(0.852, 5.344)
lambda	(0.971, 1.500)	(0.951, 1.462)
theta	(6.287, 37.942)	(5.089, 32.480)

COMPARISON WITH MLE

We have used graphical method for the comparison of Bayes estimates with ML estimates. In Fig 10, the density functions $f(x; \hat{\alpha}, \hat{\beta}, \hat{\theta})$ using MLEs and Bayesian estimates (the posterior means), computed via MCMC samples, are plotted. It is evident from the Fig 10 that the MLEs and the Bayes estimates are quite close and fit the data very well.

A further support for this finding can be obtained by inspecting the Fig 11. In this figure, we have plotted 2.5th, 50th and 97.5th quantiles of the estimated density, it can be considered as evaluation of model fit, based on posterior sample,

$$(\alpha_1^{(j)}, \beta_1^{(j)}, \theta_1^{(j)}); j = 1, \dots, 5000.$$

We have computed the density function at each observed data point for 5000 posterior samples, using logical function $density()$ in OpenBUGS

$$f(x_i; \alpha_1^{(j)}, \beta_1^{(j)}, \theta_1^{(j)}); j = 1, \dots, 5000; i = 1, \dots, 128$$

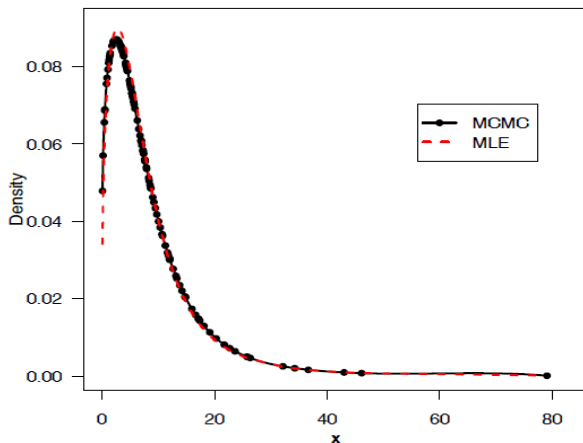


Fig 10 The density functions using ML and Bayesian estimates

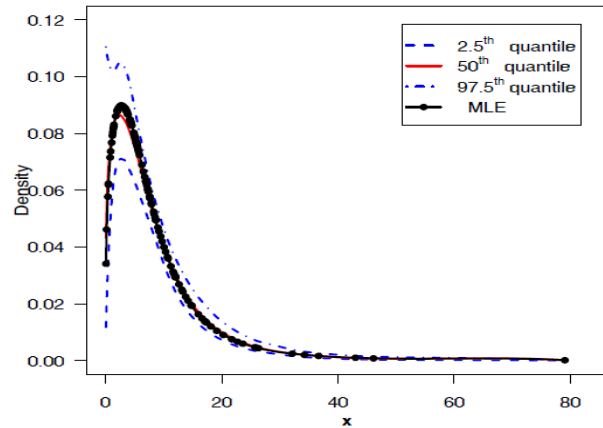


Fig 11 Density estimates

The density corresponding to MLE has been plotted using the “plug-in” estimates of the parameters. It shows that we have a fairly good model for the given data set.

ESTIMATION OF HAZARD AND RELIABILITY FUNCTIONS

In this section, our main aim is to demonstrate the effectiveness of proposed methodology. For this, we have estimated the reliability function using posterior samples. Since we have an effective MCMC technique, we can estimate any function of the parameters. We have used the Kaplan-Meier estimate of the reliability function to make the comparison more meaningful. The Fig 12 using Bayes estimate based on MCMC output and the empirical reliability function. This shows that reliability estimate based on MCMC is very close to the empirical reliability estimates.

The estimated hazard function (dashed blue line & solid red line) using Bayes estimate based on MCMC output has been displayed in the Fig 12 (right panel).

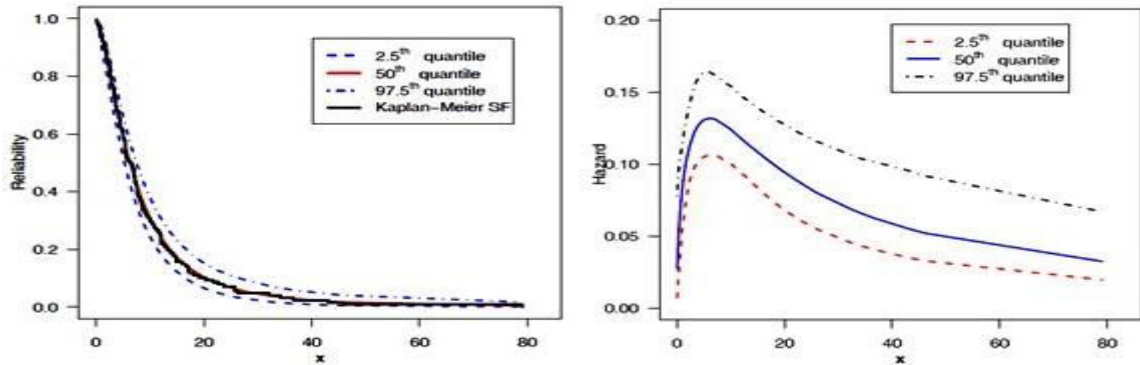


Fig 12 Reliability function(left panel) and hazard function(right panel) estimate using MCMC

ESTIMATION OF HAZARD AND RELIABILITY AT $X_{(20)} : t = 2.26$

Indeed, the MCMC samples may be used to completely summarize the posterior uncertainty about the parameters α , β and θ through a kernel estimate of the posterior distribution. This is also true of any function of the parameters e.g. reliability and hazard functions. Suppose we wish to give point and interval estimates for reliability and hazard functions at the mission time $t=2.26$ (at the 20th observed data point).

We have computed the hazard and reliability functions at mission time $t=2.26$ (at the 20th observed data point) for 5000 posterior samples, using logical function `hrf()` and `reliability()`, (Kumar et al., 2010) in OpenBUGS. It can be computed directly using hazard and reliability functions respectively, which are already discussed in previous section.

$$h(x = 2.26; \alpha_1^{(j)}, \beta_1^{(j)}, \theta_1^{(j)}); j = 1, \dots, 5000 \quad \text{and} \quad R(x = 2.26; \alpha_1^{(j)}, \beta_1^{(j)}, \theta_1^{(j)}); j = 1, \dots, 5000$$

Alternatively, we can develop R functions `hgen.aks.ghc()` and `sgen.aks.ghc()` respectively for computation.

The marginal posterior density estimates of the reliability (left panel) and hazard functions(right panel) and their histograms based on samples of size 5000 are shown in Fig 13 using the Gaussian kernel. The 95% HPD intervals are superimposed.

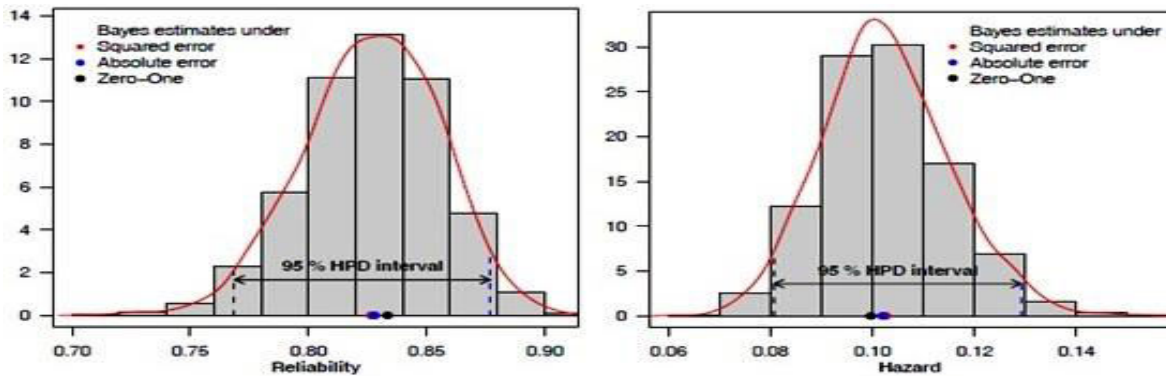


Fig 13 Visual summary of reliability(left panel) and hazard(right panel) at $t=2.26$

The MCMC results of the posterior mean, mode, standard deviation(SD), first quartile, median, third quartile, 2.5th percentile, 97.5th percentile, skewness, kurtosis, 95% symmetric and HPD credible intervals of reliability and hazard functions are displayed in Table 4. The ML estimates of reliability and hazard function at $t=2.26$ are computed using invariance property of the MLE. ML estimates $\hat{h}(t = 2.26) = 0.1032$ and $\hat{R}(t = 2.26) = 0.8361$.

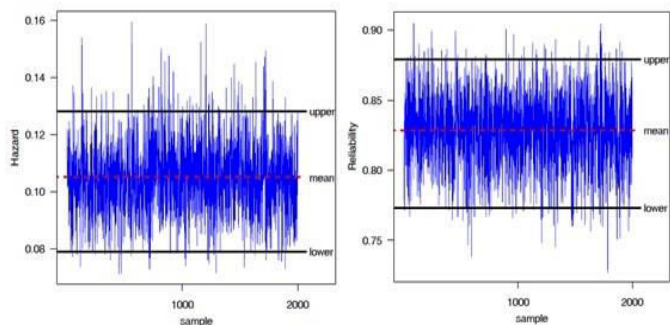


Fig 14 Trace plot for Reliability and Hazard functions at $t=2.26$

A trace plot is a plot of the iteration number against the value of the draw of the parameter at each iteration. Fig 14 display 5000 chain values for the hazard $h(t=2.26)$ and reliability $R(t=2.26)$ functions, with their sample median and 95% credible intervals.

Table 4 Posterior summary for Reliability and Hazard functions at t=2.26

Characteristics	Reliability	Hazard
Mean	0.8268	0.1026
Standard Deviation	0.0288	0.0125
2.5th Percentile	0.7698	0.0796
First Quartile (Q_1)	0.8079	0.0943
Median	0.8280	0.1020
Third Quartile (Q_3)	0.8475	0.1105
97.5th Percentile	0.8790	0.1286
Mode	0.8335	0.0997
Skewness	-0.2634	0.2893
Kurtosis	0.0376	0.2183
95% CI	(0.7698, 0.8791)	(0.0796, 0.1286)
95% HPD CI	(0.7686, 0.8771)	(0.0807, 0.1292)

MODEL COMPATIBILITY

The paper evaluates the compatibility of the GHC model through posterior predictive checks. Graphical representations, including density estimates and quantile-quantile plots, compare observed data with simulated data from the fitted model. The study concludes with a discussion on the adequacy of the GHC model for the given data set.

Posterior Predictive Checks

A natural way to assess the fit of a Bayesian model is to look at how well the predictions from the model agree with the observed data. We do this by comparing the posterior predictive simulations with the data. There are several approaches available for the study of model compatibility in Bayesian framework. Predictive simulation is an easiest and flexible one. The basic idea of studying the model compatibility through predictive simulation is to compare the observed data or some function of it with the data that would have been anticipated from the assumed model called the predictive data. If the two data sets compare favourably, the assumed model can be considered to be an appropriate choice for the data in hand, [30]. Modern Bayesian computational tools however provide straight forward solutions as one can easily simulate predictive samples if MCMC outputs are available from the posterior corresponding to the assumed model. Most of the standard numerical and graphical methods based on predictive distribution can then be easily implemented to study the compatibility of the model.

One of the best ways to assess model adequacy based on posterior predictive distributions is graphically.

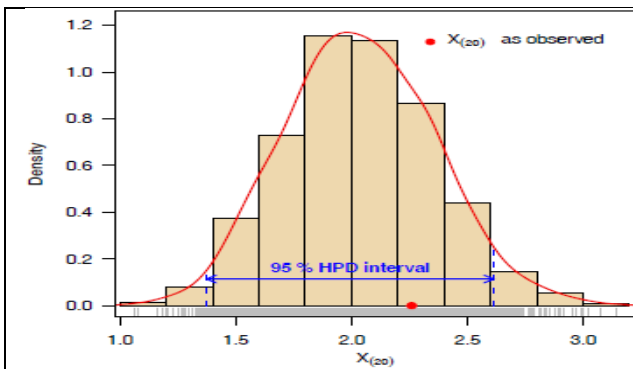


Fig 15 Posterior predictive distribution of $X_{(20)}$

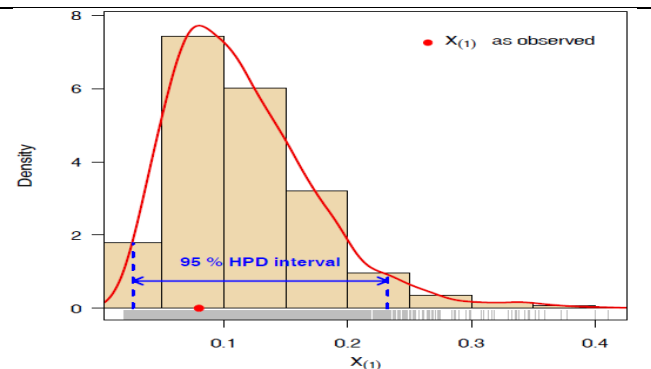


Fig 16 Posterior predictive distribution of $X_{(1)}$

To obtain further clarity on our conclusion for the study of model compatibility, we have considered plotting of density estimates of $(X_{(1)}, X_{(2)}, X_{(20)}, X_{(127)}$ and $X_{(128)})$ replicated future observations from the model with superimposed corresponding observed data. For this purpose, 2000 samples have been drawn from the posterior using MCMC procedure and then obtained predictive samples from the model under consideration using each simulated posterior sample. The size of predictive samples is same as that of observed data.

The posterior predictive distributions based on replicated future data sets are shown in Fig 15 to 19. Fig 16 to 19 represents the estimates corresponding to smallest, second smallest, second largest and largest predictive observations, whereas the same for 20th smallest observations is shown in Fig 15. The corresponding observed values are also shown.

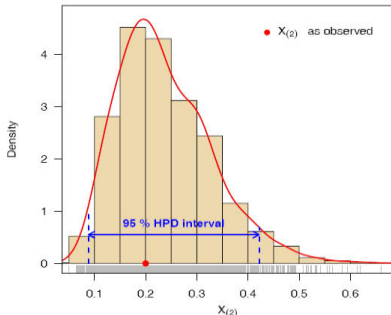


Fig 17 Posterior predictive distribution of $X_{(2)}$

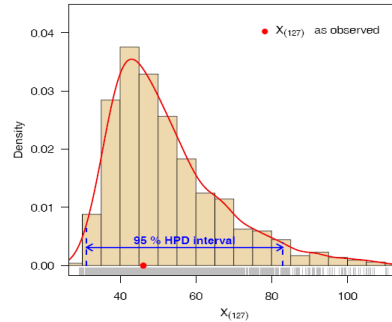


Fig 18 Posterior predictive distribution of $X_{(127)}$

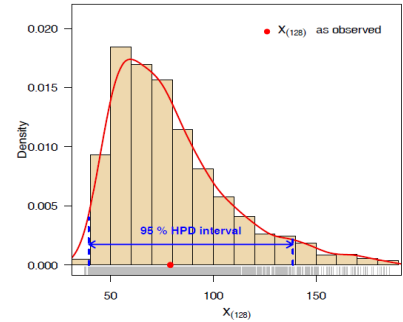


Fig 19 Posterior predictive distribution of $X_{(128)}$

The MCMC results of the posterior mean, median, mode of smallest and largest $(X_{(1)}, X_{(2)}, X_{(127)}$ and $X_{(128)})$ and $X_{(20)}$ are displayed in Table 5.

Table 5: Posterior characteristics

	Observed	Mode	Mean	Median
$X_{(1)}$	0.08	0.07	0.10	0.09
$X_{(2)}$	0.20	0.20	0.24	0.22
$X_{(20)}$	2.26	1.97	2.04	2.03
$X_{(127)}$	46.12	42.61	53.31	48.89
$X_{(128)}$	79.05	61.92	87.81	75.40

As the Figures from 15 to 19 shows, the posterior predictive distributions are centered over the observed values, which indicate good fit. In general, the distribution of replicated data appears to match that of the observed data fairly well. Overall, the results of the posterior predictive simulation indicate that model fits these data particularly well.

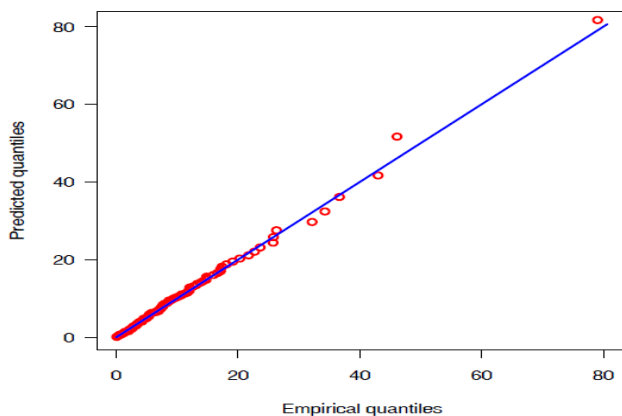


Fig 20 Q-Q plot of predictive quantiles Vs empirical quantiles

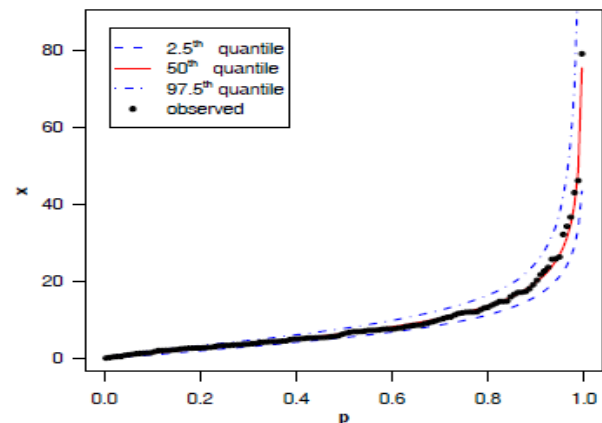


Fig 21 Graphical posterior predictive check of the model adequacy.

The basic idea of graphical model checking is to display the data alongside simulated data from the fitted model, and to look for systematic discrepancies between real and simulated data. In fact, we have predicated the entire data set. Fig 20 represents the Q-Q plot of predicted quantiles vs. observed quantiles. We, therefore, conclude that the generalized half-Cauchy is compatible with the given data set. Fig 21 exhibits graphical posterior predictive check of the model adequacy, solid line represents the posterior median and dashed lines(...) represent lower and upper bounds of 95% probability intervals, observed data is superimposed. The predictive data reflect the expected observations after replicating the experiment in future, having already observed \underline{x} and assuming that the adopted model is true. Overall, the results of the posterior predictive simulation indicate that model fits these data particularly well.

VI. CONCLUSION

The study introduced the Generalized Half Cauchy (GHC) model and conducted a thorough analysis using R. They obtained Maximum Likelihood Estimates (MLE) of parameters, calculated Bayesian estimates using MCMC, and derived probability intervals. The model's compatibility was assessed through a posterior predictive check. The techniques were applied to a real dataset, showcasing the GHC model's practicality. The study highlighted its potential use as an actuarial loss model, providing a valuable tool for Bayesian analysis in actuarial contexts.

ACKNOWLEDGMENT

The authors are thankful to Prof. Vijay Kumar, Professor & Head, Department of Mathematics & Statistics, DDU Gorakhpur University, editor and the referees for their valuable suggestions, which improved the paper to a great extent.

REFERENCES

1. Ali A. Al-Shomrani, A. I. Shawky, Osama H. Arif and Muhammad Aslam (2016), "Log-logistic distribution for survival data analysis using MCMC", SpringerPlus volume 5, Article number: 1774, 1-16.
2. Dey AK, Kundu D. (2009), "Discriminating between the log-normal and log-logistic distributions", *Commun Stat Theory Methods*.39(2):280- 292.
3. Birnbaum, Z. W.; Saunders, S. C. (1969), "A new family of life distributions", *Journal of Applied Probability*, 6 (2): 319–327, doi:10.2307/3212003, JSTOR 3212003, archived from the original on September 23, 2017
4. Glen A. G., Leemis L. M. (1997). "The Arctangent Survival Distribution", William & Mary. URL: <https://www.math.wm.edu/~leemis/1997jqt.pdf>
5. Lai C.D.(2014). "Generalized Weibull Distributions", Springer Briefs in Statistics, Springer, ISBN 978-3-642-39105-7
6. Srivastava, A.K. and Shukla, U.(2021). "Bayesian and Classical Methodologies for Reliability Analysis and Parameter Estimation of Weibull Extension Model under Informative Set of Priors", *International Journal for Modern Trends in Science and Technology*, 7(04): 76-84
7. Srivastava, A.K. (2019). "Bayesian Approach in Estimation of Shape and Scale Parameter of Log-Weibull model". *International Journal of Innovative Science, Engineering & Technology*, 6 (12), 185-196.
8. Seshadri, V. (1999), *The Inverse Gaussian Distribution*, Springer-Verlag, ISBN 978-0-387-98618-0
9. Marshall, A.W. and Olkin, I. (1997). "A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families". *Biometrika* 84, 642-652.
10. Alzaatreh, A., Lee, C., Famoye, F. and Ghosh, I. (2016): The generalized Cauchy family of distributions with applications, *Journal of Statistical Distributions and Applications*, 3-12
11. Jacob, E. and Jayakumar, K. (2012): On half Cauchy distribution and process, *International Journal of Statistika and Matematika*, 3(2), 77–81.
12. Alzaatreh, A., Mansoor, M., Tahir, M. H., Zubair, M. and Ghazali, A. S. (2016): The Gamma-Half-Cauchy distribution : Properties and Applications, *Hacettepe Journal of Mathematics and Statistics*, 45 (4), 1143-1159

13. Cordeiro, G.M and Lemonte, A.J. (2011): The beta-half- Cauchy distribution. *Journal of Probability and Statistics*, Vol. 2011, 1-18, doi: 10.1155/2011/904705.
14. Ghosh, I. (2014): The Kumaraswamy-half Cauchy distribution: Properties and applications, *Journal of Statistical Theory and Applications* 13, 122–134.
15. Srivastava, A.K. and Kumar, V.(2011): Analysis of Software Reliability Data using Exponential Power Model, *International Journal of Advanced Computer Science and Applications*, Vol. 2(2), 38-45, ISSN:2156-5570.
16. Lee, E.T., & Wang, T.W.(2003): *Statistical Methods for Survival Data Analysis* (3rd ed.) John Wiley
17. Gelman, A. (2003): A Bayesian Formulation of Exploratory Data Analysis and Goodness of-fit Testing. *International Statistical Review*, 71(2), 369-382.
18. Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004): *Bayesian Data Analysis*, (2nd ed.). London: Chapman & Hall.
19. Gupta, A., Mukherjee, B., & Upadhyay, S.K. (2008): A Bayes study using Markov Chain Monte Carlo simulation. *Reliability Engineering & System Safety*, 93, 1434-1443.
20. Kumar, V., Ligges, U., & Thomas, A. (2010): *ReliaBUGS User Manual : A subsystem in OpenBUGS for some statistical models. Version 1.0, OpenBUGS 3.2.1*, URL: <http://openbugs.info/w/Downloads/>
21. Geman, S., & Geman, D. (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 6, 721–741.
22. Gelfand, A. E., Hills, S. E., Racine-Poon, A., & Smith, A. F. M. (1990). "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling". *Journal of the American Statistical Association*, 85, 972–985.
23. Thomas, A., O’Hara, B., Ligges, U., & Sturtz, S. (2006): *Making BUGS Open*, *R News*, 6, 12–17, URL <http://mathstat.helsinki.fi/openbugs/>.
24. Kumar, V., Ligges, U., & Thomas, A. (2010). “ReliaBUGS User Manual : A subsystem in OpenBUGS for some statistical models”. Version 1.0, OpenBUGS 3.2.1, Retrieved from <http://openbugs.info/w/Downloads/>
25. R Development Core Team (2016): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL <http://www.R-project.org>.
26. Lange, K. (1999): *Numerical Analysis for Statisticians*. New York, NY: Springer.
27. Srivastava, A.K. (2014): Validation Analysis of Dhillon Model on different Real Data Sets for Reliability Modelling, *International Journal of Advance Foundation and Research in Computer*, Vol.01(09), 18-31, ISSN: 2348 - 4853.
28. Jackson, C.H.(2008). "Displaying uncertainty with shading". *The American Statistician*, 62(4), 340-347.
29. Chen, M. H., & Shao, Q. M. (1999). "Monte Carlo estimation of Bayesian credible intervals and HPD intervals". *Journal of Computational and Graphical Statistics*, 8(1), 69-92.
30. Gupta, A., Mukherjee, B., & Upadhyay, S.K. (2008). "A Bayes study using Markov Chain Monte Carlo simulation". *Reliability Engineering & System Safety*, 93, 1434-1443.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details