"Science Stays True Here"
Journal of Mathematics and Statistical Science, 235-243 | Science Signpost Publishing

Bayesian Estimation and Prediction for Chen Distribution Based on Upper Record Values

M. Faizan[1] and Sana

Department of Statistics and Operations Research
Aligarh Muslim University, Aligarh – 202 002. India

Abstract

Chen (2000) was proposed a two-parameter Chen distribution. This distribution has a bathtub-shape curve or increasing failure rate function which enables it to fit real lifetime data sets. This paper provides maximum likelihood estimates and Bayes estimators of the two unknown parameters of Chen distribution based on upper record value. Finally we have obtained prediction intervals for future observations. Numerical computations are given to illustrate the results using R software.

Keywords: Bayesian inference; Independent gamma prior; Bayesian estimation; Maximum likelihood estimation; Bayesian prediction; Record values; Record times.

1. Introduction: Let X_i, $i \geq 1$ be a sequence of independently and identically distributed (*iid*) continuous random variables with cumulative distribution function *(cdf)* $F(x)$ and probability density function *(pdf)* $f(x)$. An observation X_j will be called an upper record value if its value exceeds that of all previous observations. Thus, X_j is an upper record if $X_j > X_i$, for every $i < j$. Interested readers may refer to the books of Nevzorov (2001) and Ahsanullah (2004) for a review of various developments concerning records and their applications.

For convenience, let us assume that X_j is observed at time j, then the record time sequence $\{T_n, n \geq 1\}$ is defined in the following manner, $T_1 = 1$, with probability 1 and for $n \geq 2$,

$$T_n = min\,\{j: j > T_{n-1},\ X_j > X_{T-1}\}.$$

The upper record value sequence is then defined by

$$R_n = X_{T_n}, n = 1,2,3, \ldots.$$

Record values and the associated statistics are of interest and important in many real life applications. In industry and reliability studies, many products fail under stress. For example, a wooden beam breaks when sufficient perpendicular force is applied to it, an electronic

---

[1] Corresponding author email ID: mdfaizan02@rediffmail.com

component ceases to function in an environment of too high temperature, and a battery dies under the stress of time. But the precise breaking stress or failure point varies even among identical items. Hence, in such experiments, measurements may be made sequentially and only the record values (lower or upper) are observed. Record values arise naturally in many real life applications involving data related to weather, sports, economics and life-tests.

Some work has been done on statistical inference based on record values. See for instance,

Doostparast and Balakrishnan (2010), Raqab *et al.* (2007), Wu *et al.* (2011), Rastogi *et al.* (2012), Ahmed (2014), Huang and Wu (2012), Jozani *et al.* (2006), Soliman (2005), Doostparast (2009).

Statistical prediction is the problem of inferring the values of unknown future variables from current available informative observations. The two books by Aitcheson and Dunsmore (1980) and Geisser (1993) which are primarily concerned with Bayes prediction, give illustrative examples, analysis and possible applications.

Also prediction of future observations is treated as an important problem in clinical, industrial and agricultural experiments. Let $R_n,\ n = 1,2,\ldots.,m$ be the first m upper record values observed from a specific distribution. Then we may be interested in predicting (either point or interval) the value of the next record (R_{m+1}), or more generally, the value of the $s-th$ record (R_s) for some $s > m$. Several researchers have studied Bayesian prediction. Among others are Ebrahimi *et al.* (2010), Geisser (1985), Ahmadi *et al.* (2005), Komaki (2011), Ahmadi *et al.* (2012), Ahmadi and Balakrishnan, (2010) and Alamm *et al.* (2007).

This paper is based on upper record values of Chen distribution. In Subsection 2.1 and 2.2, we have obtained maximum likelihood estimates (*mle's*) and the Bayes estimators for the two unknown parameters of Chen distribution. In Section 3, Bayesian prediction of the $s-th$ upper record is also presented. Numerical results are provided in Section 4.

The *pdf* and the *cdf of* Chen distribution are of the form

$$f\ (x;\ \alpha,\beta)\ =\ \alpha\beta x^{\beta-1}\exp\,[\alpha(1-e^{x^{\beta}})+x^{\beta}],\quad x>0\ ,\ \alpha,\beta>0 \tag{1.1}$$

and

$$F(x;\alpha,\beta)=1-exp[\alpha(1-e^{x^{\beta}})],\qquad x>0\qquad \alpha,\beta>0\ , \tag{1.2}$$

where α and β are unknown parameters.

2. Parameter Estimation

Suppose we observe the first m upper record values $R_1 = r_1, R_2 = r_2, \ldots\ldots, R_m = r_m$, from the *cdf* $F(x;\ \theta)$ and the *pdf* $f(x;\ \theta)$. Then the joint distribution of the first m upper record values is given by [Ahsanullah (2004)]

$$f(\underline{r};\ \theta) = f(r_m;\ \theta) \prod_{i=1}^{m-1} h(r_i; \theta), \qquad -\infty < r_1 < r_2 < \cdots \ldots r_m < \infty \tag{2.1}$$

where $\underline{r} = (r_1, r_2, \ldots, r_m)$, $h(r_i; \theta) = \frac{f(r_i;\ \theta)}{1-F(r_i;\ \theta)}$.

2.1 Maximum Likelihood Estimation

Suppose we observe the first m upper record values $R_1 = r_1, R_2 = r_2, \ldots\ldots, R_m = r_m$, from the $Chen(\alpha, \beta)$ distribution with the *pdf* and the *cdf* given by (1.1) and (1.2). Then from (1.1), (1.2) and (2.1), the joint *pdf* of the first m upper record values is given by

$$L(\alpha, \beta) = f(\underline{r} \mid \alpha, \beta) = (\alpha\beta)^m [\eta(\underline{r})]^{\beta-1} exp[\alpha(1 - e^{r_m^{\beta}})]\, t(\underline{r}),\ 0 \le r_1 < r_2 < \cdots < r_m \tag{2.2}$$

where $\eta(\underline{r}) = \prod_{i=1}^{m} r_i$, $t(\underline{r}) = \prod_{i=1}^{m} e^{r_i^{\beta}}$ and $\underline{r} = (r_1, r_2, \ldots\ldots, r_m)$.

Then, the log-likelihood function is given by,

$$logL(\alpha, \beta) = m\,(log\alpha + log\beta) + (\beta - 1) \sum_{i=1}^{m} log\,(r_i) + \alpha - \alpha e^{r_m^{\beta}} + \sum_{i=1}^{m} r_i^{\beta} \tag{2.3}$$

Taking derivatives with respect to α and β of (2.3) and equating them to zero, we obtain the likelihood equations for α and β be

$$\frac{\partial}{\partial\alpha}[logL(\alpha, \beta)] = \frac{m}{\alpha} + 1 - e^{r_m^{\beta}} = 0 \tag{2.4}$$

and

$$\frac{\partial}{\partial\beta}[logL(\alpha, \beta)] = \frac{m}{\beta} + \sum_{i=1}^{m} log\,(r_i) - \alpha e^{r_m^{\beta}}\, r_m^{\beta}\, log(r_m) + \sum_{i=1}^{m} r_i^{\beta}\, log(r_i) = 0. \tag{2.5}$$

Equation (2.4) yields the MLE of α to be

$$\hat{\alpha} = \frac{m}{(1 - e^{r_m^{\beta}})}. \tag{2.6}$$

Substituting equation (2.6) into equation (2.5), the MLE of β can be obtained by solving the nonlinear equation

$$\frac{m}{\beta} + \sum_{i=1}^{m} log\,(r_i) - \frac{m}{(e^{r_m^{\beta}} - 1)}\, e^{r_m^{\beta}}\, r_m^{\beta}\, log(r_m) + \sum_{i=1}^{m} r_i^{\beta}\, log(r_i) = 0. \qquad (2.7)$$

It should be noted that the equation (2.7) is complicated to solve mathematically because it is a non linear equation, so we may use R software to solve this equation and find the *mle* of the unknown parameters.

2.2 Bayes Estimation

In this subsection, we assume independent gamma prior distribution for α and β which is given by

$$\pi(\alpha,\beta) \propto \alpha^{a-1}\, e^{-b\alpha} \beta^{p-1} e^{-q\beta}, \quad \alpha,\beta,a,b,p,q > 0, \qquad (2.8)$$

where a, b, p and q are chosen to reflect the prior knowledge about the unknown parameters.

By Bayes theorem, the posterior density function of (α,β) is

$$\pi(\alpha,\beta|\, \underline{r}) = C\, f\, (\, r|\, \alpha,\beta)\, \pi(\alpha,\beta) \qquad (2.9)$$

where C is normalizing constant. Using (2.2), (2.8) and applying Bayes theorem, the joint posterior density function of (α,β), for $\alpha,\beta > 0$, becomes

$$\pi(\alpha,\beta|\, \underline{r}) = \frac{\alpha^{m+a-1} \beta^{m+p-1} [\eta(\underline{r})]^{\beta-1} exp\, [\alpha(1- e^{r_m^{\beta}} - b)]\; t(\underline{r}) e^{-q\beta}}{\Gamma(m+a)\; \Phi_1(p,q,a,r_m)} \qquad (2.10)$$

where

$$\Phi_1(p,q,a,r_m) = \int_0^{\infty} \frac{\beta^{m+p-1} [\,\eta(\underline{r})]^{\beta-1}\; e^{-q\beta}\;\; t(\underline{r})}{[b-(1- e^{r_m^{\beta}})]^{m+a}}\, d\beta\,.$$

Assuming, a squared error loss function (SEL), the Bayes estimate is the mean of posterior density. Therefore, from (2.10) the Bayes estimators of α and $\beta$are given by

$$\hat{\alpha} = \frac{\Gamma(m+a+1)}{\Gamma(m+a)} \frac{\Phi_1(p,q,a+1,r_m)}{\Phi_1(p,q,a,r_m)} \qquad (2.11)$$

and

$$\hat{\beta} = \frac{\Phi_1(p+1,q,a,r_m)}{\Phi_1(p,q,a,r_m)}\,, \qquad (2.12)$$

respectively. Generally, the ratio of two integrals given by (2.11) and (2.12) cannot be obtained in a simple closed form, in this case, we may use numerical integration technique in R Software.

3. Bayesian Prediction of Future Records

Assume that we have the first m upper records $R_1 = r_1, R_2 = r_2, \ldots\ldots, R_m = r_m$ from the $Chen(\alpha, \beta)$ distribution. Based on such a sample, prediction, either point or interval, is needed for $s-th$ upper records, $1 \leq m < s$. Now let $Y = R_s$ be the $s-th$ upper record value, $1 \leq m < s$. It is well known that the sequence $\{R_i\,, i \geq 1\}$ is a Markov chain, that is, the conditional pdf of Y=R_s given $\underline{r} = (r_1\,, r_2, \ldots\ldots, r_m)$ is just the conditional *pdf* of $\mathrm{Y}|\mathrm{R_m} = \mathrm{r_m}$. It follows from Arnold *et al.* (1998), that

$$f(y|r_m\,, \theta) = \frac{[H(y)-H(r_m)]^{s-m-1}}{\Gamma(s-m)} \frac{f(y|\,\theta)}{1-F(r_m|\,\theta)}\,, \qquad r_m < y < \infty\,, \tag{3.1}$$

where $H(.) = -In(1 - F(.))$ and *"ln"* is used for the natural logarithm. Also, the Bayes predictive density function of $Y|\,\underline{r}$ is given by

$$h(y|\,\underline{r}) = \int_{\Theta} f(y|\,\underline{r}\,, \theta)\; \pi(\theta|\,\underline{r}\,)\, d\theta, \tag{3.2}$$

from (1.2) ,(2.1) and (3.1) , we have

$$f(y|r_m, \alpha, \beta) = \frac{\alpha^{s-m-1}(e^{y^{\beta}} - e^{r_m^{\beta}})^{s-m-1}\alpha\beta y^{\beta-1}}{\Gamma(s-m)}\, exp[-\alpha(e^{y^{\beta}} - e^{r_m^{\beta}}) + y^{\beta}]\,, \quad y > r_m\,. \tag{3.3}$$

Combining the posterior density given in (2.10), the conditional density given in (3.1) and integrating with respect to α and β, one may get Bayesian predictive density function of $Y = R_s$, (s −*th* future upper record) given the past *m* records, in the form

$$h(y|\,\underline{r}) = \frac{\Phi_2(r_m, y)}{B(m+a, s-m)\;\Phi_1(p,q,a,r_m)}\,, \quad y > r_m, \tag{3.4}$$

where $B(.,.)$ is the complete beta function and

$$\Phi_2(r_m, y) = \int_0^{\infty} \frac{(y^{\beta} - r_m^{\beta})^{s-m-1}\beta^{m+p} y^{\beta-1}[\,\eta(\underline{r})]^{\beta-1}\; e^{(-q\beta+y^{\beta})}\;\; t(\underline{r})}{[b-(1-e^{y^{\beta}})]^{s+a}}\, d\beta\,, \qquad r_m \leq y.$$

The Bayesian predictive bounds of $Y = R_s$ are obtained by evaluating $P(R_s > z|\underline{r})$ for some z. It follows from (3.4) that

$$P\left(R_s > z|\underline{r}\right) = \int_z^\infty \Phi_2(r_m, y)\, dy / \int_{rm}^\infty \Phi_2(r_m, y)\, dy \tag{3.5}$$

It can be shown that, the two-sided $(1-\gamma)$ 100% prediction interval for R_s is given by (L, U), where γ is the level of significance. The lower bound L and the upper U can be obtained by solving the following two equations:

$$1 - \frac{\gamma}{2} = P(R_s > L|\underline{r}) \qquad \text{and} \qquad \frac{\gamma}{2} = P(R_s > U|\underline{r})\,. \tag{3.6}$$

In the special case, $s = m + 1$, it can be shown that

$$P(R_{m+1} > z|\underline{r}) = \Phi_1(p, q, a, z) \,/\, \Phi_1(p, q, a, r_m) \tag{3.7}$$

4. Numerical Computations

In order to illustrate the usefulness of the inference procedures discussed in the previous sections, we generate ten sets of record values of sizes 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 from the Chen distribution with parameters α and β. Taking prior parameters $a = 8, b = 3, p = 6$ and $q = 3$, the Bayes estimators of α and β are shown in the Table 1. The *mle's* are also discussed in the same table.

Moreover, using (3.7), the lower and upper 95% predictive bounds for R_{m+1} (next upper record in the X_n sequence) are shown in the Table 2, respectively.

Table 1: MLE's and Bayes estimates based on generated record values, when the population parameters are $\alpha = 2$ and $\beta = 1$.

Sample size n	Number of Records m	Generated Records	MLE's		Bayes	
			$\hat{\alpha}_M$	$\hat{\beta}_M$	$\hat{\alpha}_B$	$\hat{\beta}_B$
10	2	0.0565541 0.7711888	2.345105	0.734163	2.355048	0.817118
20	2	0.0964869 1.2084080	1.609926	0.969298	2.455991	0.873146
30	5	0.4389127 0.7089264 0.7372351 0.7955152 0.9757005	2.234868	1.296095	2.797496	1.098560

40	4	0.1868247 0.3071563 1.2119232 1.4431430	1.637366	0.936442	2.332438	1.117733
50	3	0.0394928 0.0501206 1.1389907	1.987458	1.007009	2.295220	1.499189
60	7	0.4125921 0.6494116 0.7070127 0.8085855 0.9265034 0.9423748 1.2236942	2.174748	1.086798	2.770517	1.009301
70	2	0.6281383 1.3383497	1.853859	0.981647	2.086564	1.499189
80	6	0.1172860 0.3998376 0.4371462 0.8277019 1.2019725 1.2667020	1.851861	0.931137	2.438576	1.830358
90	5	0.2961271 0.4624653 0.6352003 1.0667923 1.1556717	2.299920	1.152925	2.820634	0.870962
100	2	1.1487200 1.3030270	1.922355	1.010763	2.169718	0.870962

Table 2: Lower and upper 95% prediction bounds for R_{m+1}

Number of Records m	Interval prediction for the next record R_{m+1}		Length
	L	U	
2	0.780001	1.606912	0.826911
3	1.021200	1.416351	0.395151
4	1.009552	1.642618	0.633066
5	0.979122	1.503980	0.524858

Reference:

Ahmed, E. A. (2014): Bayesian estimation based on progressive Type-II censoring from two-parameter bathtub-shaped lifetime model: an Markov chain Monte Carlo approach. *Journal of Applied Statistics*, **41(4)**, 752-768.

Aitchison, J. and Dunsmore, I. R. (1980): *Statistical Prediction Analysis*. CUP Archive.

Ahmadi, J., Doostparast, M. and Parsian, A. (2005): Estimation and prediction in a two-parameter exponential distribution based on k-record values under LINEX loss function. *Communications in Statistics–Theory and Methods*, **34(4)**, 795-805.

Ahmadi, J., Mir Mostafaee, S. M. T. K. and Balakrishnan, N. (2012): Bayesian prediction of k-record values based on progressively censored data from exponential distribution. *Journal of Statistical Computation and Simulation*, **82(1)**, 51-62.

Ahmadi, J. and Balakrishnan, N. (2010): Prediction of order statistics and record values from two independent sequences. *Statistics*, **44(4)**, 417-430.

Alamm, A. A., Raqab, M. Z. and Madi, M. T. (2007): Bayesian prediction intervals for future order statistics from the generalized exponential distribution. *Journal of the Iranian Statistical Society*, **6(1)**, 17-30.

Ahsanullah, M. (2004): *Record Values-Theory and Applications*. University Press of America.

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998): *Record*. John Wiley and Sons.

Chen, Z. (2000): A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters*, **49(2)**, 155-161.

Doostparast, M. (2009): A note on estimation based on record data. *Metrika*, **69(1)**, 69-80.

Doostparast, M. and Balakrishnan, N. (2010): Optimal sample size for record data and associated cost analysis for exponential distribution. *Journal of Statistical Computation and Simulation*, **80(12)**, 1389-1401.

Ebrahimi, N., Soofi, E. S. and Soyer, R. (2010): On the sample information about parameter and prediction. *Statistical Science*, **25(3)**, 348-367.

Geisser, S. (1993): *Predictive Inference*, **55**. CRC press.

Geisser, S. (1985): Interval prediction for Pareto and exponential observables. *Journal of econometrics*, **29(1-2)**, 173-185.

Huang, S. R. and Wu, S. J. (2012): Bayesian estimation and prediction for Weibull model with progressive censoring. *Journal of Statistical Computation and Simulation*, **82(11)**, 1607-1620.

Jozani, M. J., Marchand, É. and Parsian, A. (2006): On estimation with weighted balanced-type loss function. *Statistics and Probability Letters*, **76(8)**, 773-780.

Komaki, F. (2011): Bayesian predictive densities based on latent information priors. *Journal of Statistical Planning and Inference*, **141(12)**, 3705-3715.

Nevzorov, V. B. (2001): Records: *Mathematical Theory. Translation of Mathematical Monographs*, **194**. American Mathematical Society, Providence, RI.

Raqab, M. Z., Ahmadi, J. and Doostparast, M. (2007): Statistical inference based on record data from Pareto model. *Statistics*, **41(2)**, 105-118.

Rastogi, M. K., Tripathi, Y. M. and Wu, S. J. (2012): Estimating the parameters of a bathtub-shaped distribution under progressive type-II censoring. *Journal of Applied Statistics*, **39(11)**, 2389-2411.

Soliman, A. A. (2005): Estimation of parameters of life from progressively censored data using Burr-XII model. *IEEE Transactions on Reliability*, **54(1)**, 34-42.

Wu, S. F., Wu, C. C., Chou, C. H. and Lin, H. M. (2011): Statistical inferences of a two-parameter distribution with the bathtub shape based on progressive censored sample. *Journal of Statistical Computation and Simulation*, **81(3)**, 315-329.

Published: Volume 2018, Issue 6 / June 25, 2018