

Conference Proceedings Paper

Entropy Inference Based on An Objective Bayesian Approach for Upper Record Values Having the Two-Parameter Logistic Distribution

Jung-In Seo

Department of Statistics, Daejeon University, 62 Daehak-ro, Daejeon, Korea; jiseo@dju.kr

Abstract: This paper provides an entropy inference method based on an objective Bayesian approach for upper record values having the two-parameter logistic distribution. We derive the entropy based on i -th upper record value and the joint entropy based on the upper record values, and examine their properties. For objective Bayesian analysis, we provide objective priors such as the Jeffreys and reference priors for unknown parameters of the logistic distribution based on upper record values. Then, an entropy inference method based on the objective priors is developed. In real data analysis, we assess the quality of the proposed models under the objective priors.

Keywords: entropy; logistic distribution; objective Bayesian analysis; upper record value

1. Introduction

Shannon [1] proposed information theory to quantify information loss and introduces statistical entropy. Baratpour et al. [2] provided the entropy of a continuous probability distribution with upper record values and several bounds for this entropy by using the hazard rate function. Abo-Eleneen [3] suggested an efficient computation method for entropy in progressively Type-II censored samples. Kang et al. [4] derived estimators of the entropy of a double-exponential distribution based on multiply Type-II censored samples by using maximum likelihood estimators (MLEs) and approximate MLEs (AMLEs). Seo and Kang [5] developed estimation methods for entropy by using estimators of the shape parameter in the generalized half-logistic distribution based on Type-II censored samples.

This paper provides an entropy inference method based on an objective Bayesian approach for upper record values having the two-parameter logistic distribution. The cumulative distribution function (cdf) and probability density function (pdf) of the random variable X with this distribution are given by

$$F(x) = \frac{1}{1 + e^{-(x-\mu)/\sigma}}$$

and

$$f(x) = \frac{e^{-(x-\mu)/\sigma}}{\sigma [1 + e^{-(x-\mu)/\sigma}]^2}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0, \quad (1)$$

where μ is the location parameter and σ is the scale parameter.

The rest of this paper is organized as follows: Section 2 provides the Jeffreys and reference priors, and derives the entropy inference method based on the provided noninformative priors. Section 3 analyses a real data set to show the validity of the proposed method, and Section 4 concludes this paper.

2. Objective Bayesian Analysis

2.1. Objective Priors

Let $X_{U(i)}, \dots, X_{U(k)}$ be the upper record values X_1, \dots, X_n from the logistic distribution with pdf (1). Then the corresponding likelihood function is given by

$$\begin{aligned} L(\mu, \sigma) &= f(x_{U(k)}) \prod_{i=1}^{k-1} \frac{f(x_{U(i)})}{1 - F(x_{U(i)})} \\ &= \left(\frac{1}{\sigma}\right)^k \frac{\exp\left[-(x_{U(k)} - \mu)/\sigma\right]}{1 + \exp\left[-(x_{U(k)} - \mu)/\sigma\right]} \prod_{i=1}^k \frac{1}{1 + \exp\left[-(x_{U(i)} - \mu)/\sigma\right]}. \end{aligned}$$

The Fisher information matrix for (μ, σ) is given by

$$I(\mu, \sigma) = - \begin{bmatrix} E\left(\frac{\partial^2}{\partial \mu^2} \log L(\mu, \sigma)\right) & E\left(\frac{\partial^2}{\partial \mu \partial \sigma} \log L(\mu, \sigma)\right) \\ E\left(\frac{\partial^2}{\partial \sigma \partial \mu} \log L(\mu, \sigma)\right) & E\left(\frac{\partial^2}{\partial \sigma^2} \log L(\mu, \sigma)\right) \end{bmatrix}, \quad (2)$$

By the result provided in Asgharzadeh et al. [9], all elements of the Fisher information (2) are proportional to $1/\sigma^2$. Therefore, the Jeffreys prior is

$$\pi_J(\mu, \sigma) \propto \frac{1}{\sigma^2} \quad (3)$$

by the definition that it is proportional to the square root of the determinant of the Fisher information. However, the Jeffreys prior has some drawbacks for multi-parameter case such as marginalization paradox, Neyman–Scott problem, and so on. Alternately, Bernardo [10] introduced the reference prior, and Berger and Bernardo [11,12] provided a general algorithm for deriving the reference prior. By using the algorithm, we can obtain the reference prior (μ, σ) as

$$\pi_R(\mu, \sigma) \propto \frac{1}{\sigma}, \quad (4)$$

regardless which parameter is of interest.

Unfortunately, it is impossible to express in closed forms the marginal distribution for μ and σ under the objective priors (3) and (4). To generate the Markov chain Monte Carlo (MCMC) samples from the marginal distributions, we should conduct a MCMC technique.

The full conditional posterior distribution for μ and σ under a joint prior $\pi(\mu, \sigma)$ are given by

$$\pi(\mu|\sigma, \mathbf{x}) \propto \pi(\mu, \sigma) \frac{\exp(\mu/\sigma)}{1 + \exp\left[-(x_{U(k)} - \mu)/\sigma\right]} \prod_{i=1}^k \frac{1}{1 + \exp\left[-(x_{U(i)} - \mu)/\sigma\right]} \quad (5)$$

and

$$\pi(\sigma|\mu, \mathbf{x}) \propto \pi(\mu, \sigma) \left(\frac{1}{\sigma}\right)^k \frac{\exp\left[-(x_{U(k)} - \mu)/\sigma\right]}{1 + \exp\left[-(x_{U(k)} - \mu)/\sigma\right]} \prod_{i=1}^k \frac{1}{1 + \exp\left[-(x_{U(i)} - \mu)/\sigma\right]}, \quad (6)$$

respectively.

Under both objective priors (3) and (4), the full conditional posterior distribution (5) is log-concave. Therefore, we can draw the MCMC samples $\mu_i (i = 1, \dots, N)$ from the conditional

posterior distribution (5) by using the method proposed by Devroye [13]. We also need to note the fact that $\sigma \in \mathbb{R}^+$, but $\mu \in \mathbb{R}$ and $X_{U(i)} \in \mathbb{R}$. In this case, it is not easy to find a suitable proposal distribution for drawing the MCMC samples $\sigma_i (i = 1, \dots, N)$ from the full conditional posterior distribution (6). Therefore, we employ the random-walk Metropolis algorithm based on a normal proposal distribution truncated at zero.

2.2. Entropy

Theorem 1. *The entropy based on i th upper record value $X_{U(i)}$ is*

$$H_{U(i)} = \log \Gamma(i) + \log \sigma + i - (i-1)\psi(i) + \sum_{j=1}^{\infty} \frac{1}{j(j+1)^i}. \quad (7)$$

Remark 1. It is clear that the entropy (7) is an increasing function of σ . Therefore, the larger σ , the less information is provided by the distribution due to increasing entropy.

Remark 2. We can obtain the following relationships between two adjacent entropies:

$$\begin{aligned} \lim_{i \rightarrow \infty} [H_{U(i+1)} - H_{U(i)}] &= \lim_{i \rightarrow \infty} \left[\log(i-1) - \psi(i-1) + \sum_{j=1}^{\infty} \left(\frac{1}{j(j+1)^i} - \frac{1}{j(j+1)^{i-1}} \right) \right] \\ &= \lim_{i \rightarrow \infty} \left\{ \sum_{j=0}^{\infty} \left[\left(\frac{1}{i-1+j} \right) - \log \left(1 + \frac{1}{i-1+j} \right) \right] \right\} \\ &= 0. \end{aligned}$$

Theorem 2. *The joint entropy based on $X_{U(1)}, \dots, X_{U(k)}$ is*

$$H_{U(1), \dots, U(k)} = k(1 + \log \sigma) + \sum_{i=1}^k \sum_{j=1}^{\infty} \frac{1}{j(j+1)^i}, \quad (8)$$

which is an increasing function of σ as Remark 1.

Proof. The joint entropy based on upper record values $X_{U(1)}, \dots, X_{U(k)}$ is defined by Park [14] as

$$\begin{aligned} H_{U(1), \dots, U(k)} &= - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{x_{U(2)}} f_{X_{U(1)}, \dots, X_{U(k)}}(x_{U(1)}, \dots, x_{U(k)}) \\ &\quad \times \log f_{X_{U(1)}, \dots, X_{U(k)}}(x_{U(1)}, \dots, x_{U(k)}) dx_{U(1)}, \dots, dx_{U(k)}, \end{aligned}$$

where $f_{X_{U(1)}, \dots, X_{U(k)}}(x_{U(1)}, \dots, x_{U(k)})$ is the joint density function of $x_{U(1)}, \dots, x_{U(k)}$. In addition, it is simplified to a single-integral by Rad et al. [15] as

$$H_{U(1), \dots, U(k)} = \frac{k(1-k)}{2} - \sum_{i=1}^k \int_{-\infty}^{\infty} \frac{1}{\Gamma(i)} [-\log(1-F(x))]^{i-1} f(x) \log f(x) dx \quad (9)$$

Then the integral term in (9) is given by

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{\Gamma(i)} [-\log(1-F(x))]^{i-1} f(x) \log f(x) dx &= -\log \sigma - 2 \int_0^{\infty} y^i e^{-y} dy \\ &\quad + \int_0^{\infty} y^{i-1} e^{-y} \log(e^y - 1) dy \end{aligned}$$

by using

$$\log\left(\frac{z}{z-1}\right) = \sum_{j=1}^{\infty} \frac{1}{jz^j}, \quad z \leq -1 \text{ or } z > 1.$$

This completes the proof. \square

We present these changes of the entropy (7) and the entropy (8) in Tables 1, 2 and Figure 1.

Table 1. Entropy based on i th upper record value $X_{U(i)}$.

$\sigma \backslash i$	1	2	3	4	5	6	7	8	9	10
0.1	-0.303	-0.370	-0.302	-0.208	-0.115	-0.029	0.048	0.117	0.179	0.234
0.5	1.307	1.239	1.307	1.401	1.494	1.580	1.657	1.727	1.788	1.844
1	2.000	1.932	2.001	2.094	2.187	2.273	2.351	2.420	2.481	2.537
2	2.693	2.625	2.694	2.787	2.880	2.966	3.044	3.113	3.175	3.230
4	3.386	3.319	3.387	3.480	3.574	3.660	3.737	3.806	3.868	3.923
8	4.079	4.012	4.080	4.174	4.267	4.353	4.430	4.499	4.561	4.616

Table 2. Joint entropy based on $X_{U(1)}, \dots, X_{U(k)}$.

$\sigma \backslash k$	1	2	3	4	5	6	7	8	9	10
0.1	-0.303	-1.250	-2.400	-3.632	-4.900	-6.187	-7.481	-8.780	-10.080	-11.382
0.5	1.307	1.969	2.429	2.806	3.147	3.470	3.785	4.096	4.405	4.712
1	2.000	3.355	4.508	5.579	6.612	7.629	8.637	9.641	10.643	11.644
2	2.693	4.741	6.587	8.351	10.078	11.788	13.489	15.186	16.881	18.575
4	3.386	6.128	8.667	11.124	13.544	15.947	18.341	20.731	23.120	25.507
8	4.079	7.514	10.746	13.896	17.010	20.106	23.193	26.276	29.358	32.438

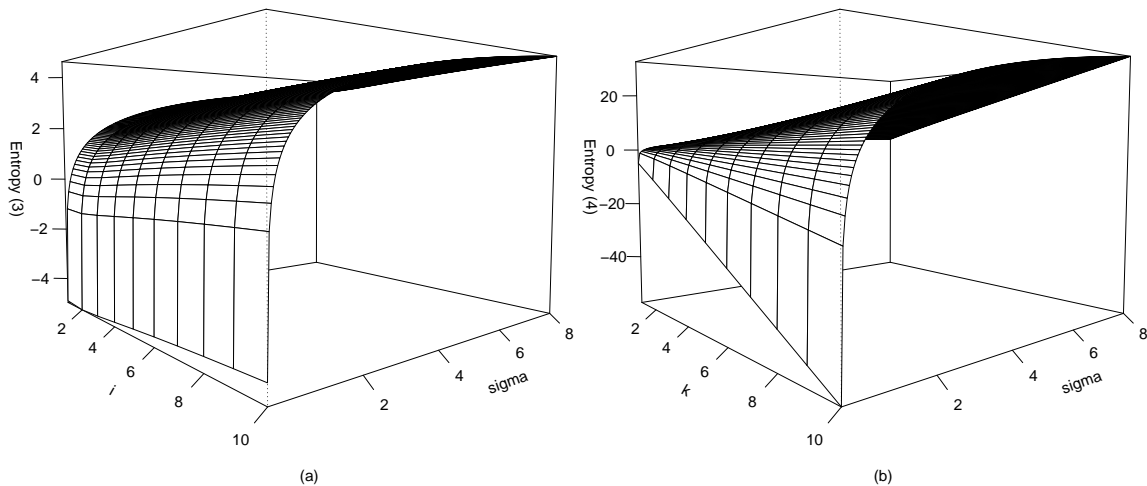


Figure 1. Entropy for upper record values.

Here, we need to note that σ is unknown parameter and should be estimated when real data are observed. The following theorem provides an estimator of the joint entropy (8) in a Bayesian framework.

Theorem 3. *The Bayes estimator of the joint entropy (8) is*

$$\hat{H}_{U(1), \dots, U(k)}^B = k \left[1 + E_{\pi|\mathbf{x}}(\log \sigma) \right] + \sum_{i=1}^k \sum_{j=1}^{\infty} \frac{1}{j(j+1)^i}, \quad (10)$$

where $E_{\pi|\mathbf{x}}(\cdot)$ is the posterior expectation.

Proof. In the Bayesian view, the entropy estimator based on $X_{U(1)}, \dots, X_{U(k)}$ is defined as

$$\hat{H}_{U(1), \dots, U(k)}^B = \int_{\mu} \int_{\sigma} H_{U(1), \dots, U(k)} \pi(\mu, \sigma | \mathbf{x}) d\mu d\sigma. \quad (11)$$

Then, the estimator (11) is given by

$$\hat{H}_{U(1), \dots, U(k)}^B = k + k \int_{\mu} \int_{\sigma} \log \sigma \pi(\mu, \sigma | \mathbf{x}) d\mu d\sigma + \sum_{i=1}^k \sum_{j=1}^{\infty} \frac{1}{j(j+1)^i}.$$

The term in (10), $E_{\pi|\mathbf{x}}(\log \sigma)$, is approximated as

$$E_{\pi|\mathbf{x}}(\log \sigma) \approx \frac{1}{N - M} \sum_{i=M+1}^N \log \sigma_i,$$

where M is the number of burn-in samples. It completes the proof. \square

The following section examines the validity of the provided objective Bayesian method by analyzing a real data set.

3. Application

Asgharzadeh et al. [9] analyzed the upper record values 2.70, 3.78, 4.83, 8.02, 8.37 from the total annual rainfall (in inches) during March recorded at Los Angeles Civic Center from 1973 to 2006. The MCMC samples are generated by using the MCMC algorithm described in Subsection 2.1. To obtain the optimal acceptance rate (see Robert and Rosenthal [7]) under the provided priors (3) and (4), the variances in a truncated normal proposal are assigned to 0.7 and 0.8, respectively. Based on 5,500 MCMC samples with 500 burn-in samples, the Bayes estimates under the square error loss function and corresponding 95% HPD CrIs are computed to compare the MLEs. The results are reported in Table 3. To check the validity of the MCMC samples, we present their autocorrelation functions (ACFs) and trace plots.

Table 3. Estimates and the corresponding 95% HPD CrI for μ and σ .

	μ			σ		
	$\hat{\mu}$	$\hat{\mu}_{JB}$	$\hat{\mu}_{RB}$	$\hat{\sigma}$	$\hat{\sigma}_{JB}$	$\hat{\sigma}_{RB}$
Estimate	2.929	3.196	3.209	0.998	1.010	1.096
HPD CrI		(2.694, 3.904)	(2.392, 4.122)		(0.386, 1.635)	(0.403, 1.935)
Acceptance Rate					0.438	0.439

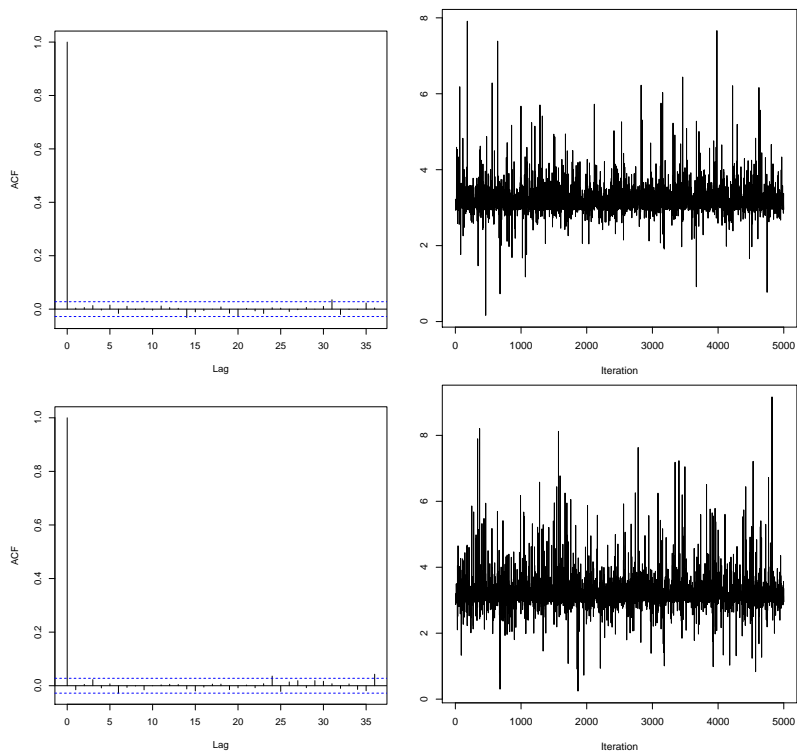


Figure 2. ACFs (left) and trace plots (right) of the MCMC samples μ_i under the priors $\pi_I(\mu, \sigma)$ (top) and $\pi_R(\mu, \sigma)$ (bottom).

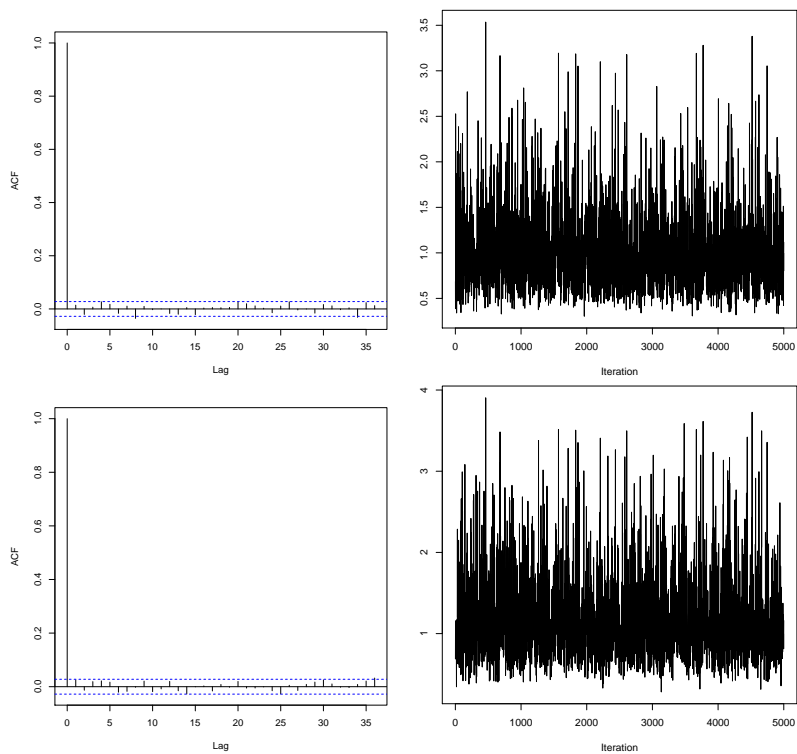


Figure 3. ACFs (left) and trace plots (right) of the MCMC samples σ_i under the priors $\pi_I(\mu, \sigma)$ (top) and $\pi_R(\mu, \sigma)$ (bottom).

From Figures 2 and 3, we can see that the MCMC samples are mixing and converge to the stationary distribution well. Further, we assess the quality of the Bayesian models under the provided priors (3) and (4) based on the replications $X_{U(i)}^{rep}$ ($i = 1, \dots, 5$) of the observed upper record values from the posterior predictive distributions given by

$$f_{X^{rep}}^J(x^{rep}|x) = \int_{\mu} \int_{\sigma} f_{X^{rep}}(x^{rep}|\mu, \sigma) \pi_J(\mu, \sigma|\mathbf{x}) d\mu d\sigma \quad (12)$$

and

$$f_{X^{rep}}^R(x^{rep}|x) = \int_{\mu} \int_{\sigma} f_{X^{rep}}(x^{rep}|\mu, \sigma) \pi_R(\mu, \sigma|\mathbf{x}) d\mu d\sigma, \quad (13)$$

where $f_{X^{rep}}(x^{rep})$ is the marginal density function of X^{rep} . The replications are obtained as

$$X_{U(i)}^{rep} = \frac{1}{N - M} \sum_{j=M+1}^N X_{U(i)}^{rep(j)}, \quad i = 1, \dots, k,$$

where $X_{U(i)}^{rep(j)}$ is a sample from the marginal density function $f_{X^{rep}}(x^{rep})$. The replications and their mean and standard deviation (std) are given in Table 4. The mean and standard deviation (std) of the observed upper record values are 5.54 and 2.541, respectively.

Table 4. Replications, and their Mean and standard deviation (std).

	$X_{U(1)}^{rep}$	$X_{U(2)}^{rep}$	$X_{U(3)}^{rep}$	$X_{U(4)}^{rep}$	$X_{U(5)}^{rep}$	Mean	Std
π_{JB}	3.195	4.846	6.046	7.176	8.204	5.894	2.064
π_{RB}	3.221	5.000	6.288	7.506	8.629	6.129	2.226

The model under the Jeffreys prior (3) shows better performance for the replications $X_{U(i)}^{rep}$ ($i = 1, 2, 3, 5$) and mean, while that under the reference prior (4) shows better performance for the replication $X_{U(4)}^{rep}$ and std. However, there is no significant difference between the replications under the priors (3) and (4). That is, we can not conclude which model is better for this observed upper record values. Therefore, we show the estimation results of the joint entropy (10) under both priors (3) and (4) in Table 5. Further, we present kernel densities of the joint entropies based on the MCMC samples to show the joint entropies graphically.

Table 5. Estimates and the corresponding 95% HPD CrI of the joint entropy $\hat{H}_{U(i), \dots, U(k)}^B$

	$\hat{H}_{U(1), \dots, U(k)}^{JB}$	$\hat{H}_{U(1), \dots, U(k)}^{RB}$
Estimate	6.431	6.803
HPD CrI	(2.648, 9.417)	(3.455, 10.557)

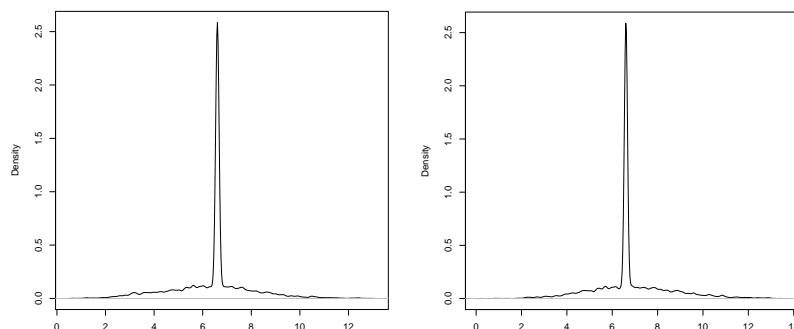


Figure 4. Kernel densities of joint entropy based on the MCMC samples under the priors $\pi_J(\mu, \sigma)$ (left) and $\pi_R(\mu, \sigma)$ (right).

4. Conclusions

This paper proposed an entropy inference method based on an objective Bayesian approach for upper record values having the two-parameter logistic distribution. First, we provided the noninformative priors such as the Jeffreys and reference priors for unknown parameters of the two-parameter logistic distribution, and then derived the entropy based on i -th upper record value and the joint entropy based on the upper record values, and examined their properties. We evaluated the objective Bayesian models under the provided objective priors through the posterior predictive checking that is conducted based on the replications of the observed upper record values. The proposed objective Bayesian approach is useful when there is not enough prior information, and saves a lot of effort and time to obtain prior information.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
2. Baratpour, S.; Ahmadi, J.; Arghami N.R. Entropy properties of record statistics. *Stat. Pap.* **2007**, *48*, 197–213.
3. Abo-Eleneen, Z.A. The entropy of progressively censored samples. *Entropy* **2011**, *13*, 437–449.
4. Kang, S.B.; Cho, Y.S.; Han, J.T.; Kim, J. An estimation of the entropy for a double exponential distribution based on multiply Type-II censored samples. *Entropy* **2012**, *14*, 161–173.
5. Seo, J.I.; Kang, S.B. Entropy estimation of generalized half-logistic distribution (GHLD) based on Type-II censored samples. *Entropy* **2014**, *16*, 443–454.
6. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109.
7. Roberts, G.O.; Rosenthal, J.S. Optimal scaling for various Metropolis-Hastings algorithms. *Stat. Sci.* **2001**, *16*, 351–367.
8. Chen, M.H.; Shao, Q.M. Monte Carlo estimation of Bayesian credible and hpd intervals. *J. Comput. Graph. Stat.* **1998**, *8*, 69–92.
9. Asgharzadeh, A.; Valiollahi, R.; Abdi, M. Point and interval estimation for the logistic distribution based on record data. *Stat. Oper. Res. Trans.* **2016**, *40*, 1–24.
10. Bernardo, J.M. Reference posterior distributions for Bayesian inference (with discussion). *J. Royal Stat. Society. Ser. B* **1979**, *41*, 113–147.
11. Berger, J.O.; Bernardo, J.M. Estimating a product of means: Bayesian analysis with reference priors. *J. Am. Stat. Assoc.* **1989**, *84*, 200–207.
12. Berger, J.O.; Bernardo, J.M. On the development of reference priors. *Bayesian Stat.* **1992**, *4*, 35–60.
13. Devroye, L. A simple algorithm for generating random variates with a log-concave density. *Computing* **1984**, *33*, 247–257.
14. Park, S. Testing exponentiality based on the Kullback-Leibler information with the Type-II censored data. *IEEE Trans. Reliab.* **2005**, *54*, 22–26.

15. Rad, A.H.; Yousefzadeh, F.; Amini, M.; Arghami, N.R. Testing Exponentiality Based on Record Values. *J. Iran. Stat. Soc.* **2007**, *6*, 77–87.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).