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

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

and

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

where \( \mu \) is the location parameter and \( \sigma \) 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 \( X_{U(i)} \), \( i = 1, \ldots, n \), be the upper record values from the logistic distribution with pdf (1). Then the corresponding likelihood function is given by

\[
L(\mu, \sigma) = f \left( x_{U(i)} \right) \prod_{i=1}^{k} \frac{f \left( x_{U(i)} \right)}{1 - F \left( x_{U(i)} \right)}
\]

\[
= \left( \frac{1}{\sigma} \right)^k \exp \left[ - \left( x_{U(k)} - \mu \right) / \sigma \right] \prod_{i=1}^{k} \frac{1}{1 + \exp \left[ - \left( x_{U(i)} - \mu \right) / \sigma \right]}.
\]

The Fisher information matrix for \((\mu, \sigma)\) is given by

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

(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

\[
\pi_J(\mu, \sigma) \propto \frac{1}{\sigma^2}
\]

(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 \((\mu, \sigma)\) as

\[
\pi_R(\mu, \sigma) \propto \frac{1}{\sigma},
\]

(4)

regardless which parameter is of interest.

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

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

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

(5)

and

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

(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 \(\mu_i (i = 1, \ldots, N)\) 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 $\sigma_i(i = 1, \ldots, N)$ 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 $\sigma$. Therefore, the larger $\sigma$, the less information is provided by the distribution due to increasing entropy.

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

$$\lim_{i \to \infty} \left[ H_{U(i+1)} - H_{U(i)} \right] = \lim_{i \to \infty} \left[ \log(i - 1) - \psi(i) + \sum_{j=1}^{\infty} \left( \frac{1}{j(j+1)^i} - \frac{1}{j(j+1)^{i-1}} \right) \right]$$

$$= \lim_{i \to \infty} \left\{ \sum_{j=1}^{\infty} \left[ \left( \frac{1}{i-1+j} \right) - \log \left( 1 + \frac{1}{i-1+j} \right) \right] \right\} = 0.$$

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

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

which is an increasing function of $\sigma$ as Remark 1.

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

$$H_{U(1), \ldots, U(k)} = -\int_{-\infty}^{x_{U(1)}} \cdots \int_{-\infty}^{x_{U(k)}} f_{X_{U(1)}, \ldots, X_{U(k)}}(x_{U(1)}, \ldots, x_{U(k)})$$

$$\times \log f_{X_{U(1)}, \ldots, X_{U(k)}}(x_{U(1)}, \ldots, x_{U(k)}) \, dx_{U(1)}, \ldots, dx_{U(k)},$$

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

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

Then the integral term in (9) is given by

$$\int_{-\infty}^{\infty} \frac{1}{\Gamma(i)} \left[ - \log(1 - F(x)) \right]^{i-1} f(x) \log f(x) \, dx = - \log \sigma - 2 \int_{0}^{\infty} y e^{-y} \, dy + \int_{0}^{\infty} y^{i-1} e^{-y} \log (e^y - 1) \, dy.$$
by using

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

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 \(X_{U(i)}\).

<table>
<thead>
<tr>
<th>(\sigma)</th>
<th>(i)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td></td>
<td>-0.303</td>
<td>-0.370</td>
<td>-0.302</td>
<td>-0.208</td>
<td>-0.115</td>
<td>-0.029</td>
<td>0.048</td>
<td>0.117</td>
<td>0.179</td>
<td>0.234</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td>1.307</td>
<td>1.239</td>
<td>1.307</td>
<td>1.401</td>
<td>1.494</td>
<td>1.580</td>
<td>1.657</td>
<td>1.727</td>
<td>1.788</td>
<td>1.844</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>2.000</td>
<td>1.932</td>
<td>2.001</td>
<td>2.094</td>
<td>2.187</td>
<td>2.273</td>
<td>2.351</td>
<td>2.420</td>
<td>2.481</td>
<td>2.537</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2.693</td>
<td>2.625</td>
<td>2.694</td>
<td>2.787</td>
<td>2.880</td>
<td>2.966</td>
<td>3.044</td>
<td>3.113</td>
<td>3.175</td>
<td>3.230</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>(\sigma)</th>
<th>(k)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td></td>
<td>1.307</td>
<td>1.969</td>
<td>2.429</td>
<td>2.806</td>
<td>3.147</td>
<td>3.470</td>
<td>3.785</td>
<td>4.096</td>
<td>4.405</td>
<td>4.712</td>
</tr>
</tbody>
</table>

Figure 1. Entropy for upper record values.

Here, we need to note that \(\sigma\) 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
\[
\hat{H}^{\text{BU}}_{U(1), \ldots, U(k)} = k \left[ 1 + E_{\pi|x} (\log \sigma) \right] + \sum_{i=1}^{k} \sum_{j=1}^{\infty} \frac{1}{j(j+1)^{2}}, \tag{10}
\]
where \(E_{\pi|x} (\cdot)\) is the posterior expectation.

Proof. In the Bayesian view, the entropy estimator based on \(X_{U(1)}, \ldots, X_{U(k)}\) is defined as
\[
\hat{H}^{\text{BU}}_{U(1), \ldots, U(k)} = \int \int H_{U(1), \ldots, U(k)} \pi(\mu, \sigma | x) d\mu d\sigma. \tag{11}
\]
Then, the estimator (11) is given by
\[
\hat{H}^{\text{BU}}_{U(1), \ldots, U(k)} = k + k \int \int \log \sigma \pi(\mu, \sigma | x) d\mu d\sigma + \sum_{i=1}^{k} \sum_{j=1}^{\infty} \frac{1}{j(j+1)^{2}}.
\]
The term in (10), \(E_{\pi|x} (\log \sigma)\), is approximated as
\[
E_{\pi|x} (\log \sigma) \approx \frac{1}{N-M} \sum_{i=M+1}^{N} \log \sigma_i,
\]
where \(M\) is the number of burn-in samples. It completes the proof. \(\square\)

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 \(\mu\) and \(\sigma\). |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \(\mu\) | \(\hat{\mu}_{\text{JB}}\) | \(\hat{\mu}_{\text{RB}}\) | \(\sigma\) | \(\hat{\sigma}_{\text{JB}}\) | \(\hat{\sigma}_{\text{RB}}\) |
| 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) | (0.438) | (0.439) |
| Acceptance Rate | | | 0.438 | 0.439 | |
Figure 2. ACFs (left) and trace plots (right) of the MCMC samples $\mu_i$ under the priors $\pi_J(\mu, \sigma)$ (top) and $\pi_R(\mu, \sigma)$ (bottom).

Figure 3. ACFs (left) and trace plots (right) of the MCMC samples $\sigma_i$ under the priors $\pi_J(\mu, \sigma)$ (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 $X_{U(i)}^\text{rep}$ ($i = 1, \ldots, 5$) of the observed upper record values from the posterior predictive distributions given by

$$f_{X_{U(i)}^\text{rep}}(x^\text{rep}|x) = \int_\mu \int_\sigma f_{X_{U(i)}^\text{rep}}(x^\text{rep}|\mu,\sigma)\pi_J(\mu,\sigma|x)d\mu d\sigma$$

and

$$f_{X_{U(i)}^\text{rep}}^R(x^\text{rep}|x) = \int_\mu \int_\sigma f_{X_{U(i)}^\text{rep}}(x^\text{rep}|\mu,\sigma)\pi_R(\mu,\sigma|x)d\mu d\sigma,$$

where $f_{X_{U(i)}^\text{rep}}(x^\text{rep})$ is the marginal density function of $X_{U(i)}^\text{rep}$. The replications are obtained as

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

where $X_{U(i)}^{\text{rep}(j)}$ is a sample from the marginal density function $f_{X_{U(i)}^\text{rep}}(x^\text{rep})$. 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>
<thead>
<tr>
<th>Table 4. Replications, and their Mean and standard deviation (std).</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{U(1)}^{\text{rep}}$</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>$\pi_{JB}$</td>
</tr>
<tr>
<td>$\pi_{RB}$</td>
</tr>
</tbody>
</table>

The model under the Jeffreys prior (3) shows better performance for the replications $X_{U(i)}^{\text{rep}}$ ($i = 1, 2, 3, 5$) and mean, while that under the reference prior (4) shows better performance for the replication $X_{U(4)}^{\text{rep}}$ and std. However, there is no significant difference between the replications under the priors (3) and (4). That is, we cannot 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>
<thead>
<tr>
<th>Table 5. Estimates and the corresponding 95% HPD CrI of the joint entropy $\hat{H}_{U(i),…,U(k)}^R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{H}_{U(1),…,U(k)}^J$</td>
</tr>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>6.431</td>
</tr>
<tr>
<td>6.803</td>
</tr>
</tbody>
</table>
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 the $i$-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


© 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/).