British Journal of Mathematics & Computer Science
5(1): 128-132, 2015, Article no.BJMCS.2015.010
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

SCIENCEDOMAIN

Estimation of the Exponential Distribution in the Light of Future Data

Kunio Takezawa[1*]

[1] *Agroinformatics Division, Agricultural Research Center, National Agriculture and Food Research Organization, Kannondai 3-1-1, Tsukuba, Ibaraki 305-8666, Japan.*

Article Information
DOI: 10.9734/BJMCS/2015/13658
Editor(s):
(1) Radko Mesiar, Department of Mathematics, Slovak University of Technology Bratislava, Slovakia.
Reviewers:
(1) D. S. Hooda, Department of mathematics, Jaypee University of Engineering Technology, India.
(2) Anonymous, Payame Noor University, Iran.
Complete Peer review History:
http://www.sciencedomain.org/review-history.php?iid=707&id=6&aid=6656

Short Research Article

Received: 27 August 2014
Accepted: 30 September 2014
Published: 23 October 2014

Abstract

The maximum likelihood estimator does not always maximize the expected log-likelihood when estimating the exponential distribution using a small data set. The maximum likelihood estimator should be multiplied by a positive constant that only depends on the amount of data.

Keywords: Expected log-likelihood; exponential distribution; maximum likelihood estimator; optimization

2010 Mathematics Subject Classification: 60G25; 62F10; 62M20

1 Introduction

The maximum likelihood method produces superior results when there is a large amount of data. However, this is not always true for a small amount of data. Additionally, the maximum likelihood method is not guaranteed to derive the best results in terms of the expected log-likelihood (page 35 in [1]), i.e., the log-likelihood in the light of future data. For example, when we use an estimated normal distribution variance, the third variance outperforms the maximum likelihood variance ([2], page 245 in

**Corresponding author: E-mail: nonpara@gmail.com*

[3]). Nevertheless, there is no general theory on this characteristic, and no method that is applicable to practical applications. Hence, we should accumulate the appropriate techniques and create a rigid theoretical system. In this paper, we improve the maximum likelihood method for estimating the exponential distribution's parameter from a standpoint of the expected log-likelihood.

Section 2 presents theoretical considerations when estimating the parameter of an exponential distribution by maximizing the expected log-likelihood. Using these results, Section 3 presents our derivation of the constant that modifies the maximum likelihood estimator.

2 Log-likelihood of the Exponential Distribution in the Light of Future Data

The probability density function ($f(x)$) of the exponential distribution is

$$f(x) = \begin{cases} \tilde{\lambda}\text{exp}(-\tilde{\lambda}x) & if x \geq 0 \\ 0 & if x < 0. \end{cases} \tag{2.1}$$

Its expectation is

$$\int_0^\infty xf(x)dx = \int_0^\infty \tilde{\lambda}\text{exp}(-\tilde{\lambda}x)xdx = \frac{1}{\tilde{\lambda}}. \tag{2.2}$$

Let us assume that the realizations of the random variable obeying this distribution are $\{x_i\}$ $(1 \leq i \leq n)$. Then, the log-likelihood ($l(\lambda|\{x_i\})$) of λ is

$$\frac{l(\lambda|\{x_i\})}{n} = \text{log}(\lambda) - \frac{\lambda}{n}\sum_{i=1}^{n} x_i. \tag{2.3}$$

To derive the value of λ that maximizes the value above, we differentiate this equation with respect to λ and then set the results equal to 0. Then, we have

$$\frac{1}{\lambda} - \frac{1}{n}\sum_{i=1}^{n} x_i = 0. \tag{2.4}$$

This leads to the maximum likelihood estimator,

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}. \tag{2.5}$$

This $\hat{\lambda}$ is the maximum likelihood estimator in the light of the data at hand ($\{x_i\}$).

Next, we assume that $\{x_i^*\}$ $(1 \leq i \leq m)$ is future data. The log-likelihood ($l(\hat{\lambda}|\{x_i\})$) of $\hat{\lambda}$ in the light of this future data is

$$\frac{l(\hat{\lambda}|\{x_i^*\})}{m} = \text{log}(\hat{\lambda}) - \frac{\hat{\lambda}}{m}\sum_{i=1}^{m} x_i^*. \tag{2.6}$$

Let $\hat{\lambda}^*$ be the value of $\hat{\lambda}$ that maximizes this value. Then,

$$\hat{\lambda}^* = \frac{m}{\sum_{i=1}^{m} x_i^*}. \tag{2.7}$$

Because we have an infinite amount of future data, we let m be infinite. We denote $\hat{\lambda}^*$ in this setting as $\hat{\lambda}^*_\infty$. Then, we have

$$\frac{1}{\hat{\lambda}^*_\infty} = \lim_{m\to\infty} \frac{\sum_{i=1}^{m} x_i^*}{m}. \tag{2.8}$$

The right hand side of the above equation is the expectation of Eq. (2.1). Therefore, we can use Eq. (2.2) to obtain

$$\frac{1}{\hat{\lambda}^*_\infty} = \frac{1}{\tilde{\lambda}}. \tag{2.9}$$

Hence, we have

$$\hat{\lambda}^*_\infty = \tilde{\lambda}. \tag{2.10}$$

That is, the maximum likelihood estimator that is calculated using an infinite amount of future data is the true parameter ($\tilde{\lambda}$) of the population.

When Eq. (2.10) is substituted into Eq. (2.6), we obtain

$$\lim_{m\to\infty} \frac{l(\hat{\lambda}|\{x^*_i\})}{m} = \log(\hat{\lambda}) - \frac{\hat{\lambda}}{\tilde{\lambda}}. \tag{2.11}$$

Using this result and Eq. (2.5), the log-likelihood ($l^*(\hat{\lambda})$) of $\hat{\lambda}$ in the light of an infinite amount of future data is

$$l^*(\hat{\lambda}) = \log\left(\frac{n}{\sum_{i=1}^n x_i}\right) - \frac{n}{\tilde{\lambda}\sum_{i=1}^n x_i}. \tag{2.12}$$

However, $\hat{\lambda}$ given by Eq.(2.5) is not necessarily the best estimator in the light of an infinite amount of future data. We let that best estimator be $\alpha\hat{\lambda}$, where α is a positive constant. Hence, the log-likelihood in the light of an infinite amount of future data is

$$l^*(\alpha\hat{\lambda}) = \log\left(\frac{\alpha n}{\sum_{i=1}^n x_i}\right) - \frac{\alpha n}{\tilde{\lambda}\sum_{i=1}^n x_i}. \tag{2.13}$$

To maximize the above value, we set the derivative of this equation with respect to α equal to 0. Then, we have

$$\alpha = \frac{\tilde{\lambda}\sum_{i=1}^n x_i}{n}. \tag{2.14}$$

We cannot calculate this α because $\tilde{\lambda}$ is unknown.

Therefore, we consider the mean $l^*(\alpha\hat{\lambda})$ from an infinite number of samples of $\{x_i\}$ $(1 \le i \le n)$. That is, we consider the expectation of $l^*(\alpha\hat{\lambda})$. This expectation can be written as

$$E_{\{x_i\}}\left[l^*(\alpha\hat{\lambda})\right] = E_{\{x_i\}}\left[\log\left(\frac{\alpha n}{\sum_{i=1}^n x_i}\right) - \frac{\alpha n}{\tilde{\lambda}\sum_{i=1}^n x_i}\right]. \tag{2.15}$$

Note that the probability density function of the exponential distribution can be represented as in Eq. (2.1). We can replace $\tilde{\lambda}$ with $c\tilde{\lambda}$ (where c is a positive constant). Then, the resultant probability density function gives the same value as Eq. (2.1), in which x takes the value of $\frac{x}{c}$. Let $\{x_i\}$ $(1 \le i \le n)$ be realizations of the random variable obeying Eq. (2.1). Then, we can regard $\{\frac{x_i}{c}\}$ $(1 \le i \le n)$ as realizations of the random variable obeying the exponential distribution with $\tilde{\lambda}$ replaced by $c\tilde{\lambda}$. Hence, when $\tilde{\lambda}$ is replaced with $c\tilde{\lambda}$, we can replace Eq. (2.15) with

$$\begin{aligned} E_{\{\frac{x_i}{c}\}}\left[l^*(\alpha\hat{\lambda})\right] &= E_{\{x_i\}}\left[\log\left(\frac{\alpha n}{\sum_{i=1}^n \frac{x_i}{c}}\right) - \frac{\alpha n}{c\tilde{\lambda}\sum_{i=1}^n \frac{x_i}{c}}\right] \\ &= E_{\{x_i\}}\left[\log\left(\frac{c\alpha n}{\sum_{i=1}^n x_i}\right) - \frac{\alpha n}{\tilde{\lambda}\sum_{i=1}^n x_i}\right] \\ &= E_{\{x_i\}}\left[\log(c) + \log\left(\frac{\alpha n}{\sum_{i=1}^n x_i}\right) - \frac{\alpha n}{\tilde{\lambda}\sum_{i=1}^n x_i}\right]. \end{aligned} \tag{2.16}$$

Thus, we have proved that the α that maximizes Eq. (2.15) is identical to the α that maximizes Eq. (2.16). That is, $\hat{\alpha}$ is independent of $\tilde{\lambda}$ and it only depends on n. Therefore, we can set $\tilde{\lambda}$ to be a specific positive value and derive $\hat{\alpha}$ as a function of n. Then, the resulting $\hat{\alpha}$ is valid, regardless of the value of $\tilde{\lambda}$. It means that we can use the values of $\hat{\alpha}$ even if $\tilde{\lambda}$ is unknown. The third variance has a common feature; the ratio between the estimator that maximizes the expected log-likelihood and the maximum likelihood estimator depends upon only n.

3 Integration Using the Gamma Distribution

The probability density function of the gamma distribution is (page 488 in [4]).

$$g_{p,\tilde{\lambda}}(x) = \frac{\tilde{\lambda}^p x^{p-1}\text{exp}(-\tilde{\lambda}x)}{\Gamma(p)}, \tag{3.1}$$

where $\Gamma(p)$ is Euler's gamma function. It is defined as

$$\Gamma(p) = \int_0^\infty t^{p-1}\text{exp}(-t)dt. \tag{3.2}$$

When $p = 1$ is set in Eq. (3.1), we have

$$g_{1,\tilde{\lambda}}(x) = \frac{\tilde{\lambda}\text{exp}(-\tilde{\lambda}x)}{\Gamma(1)}. \tag{3.3}$$

Becuase $\Gamma(1) = 1$, $g_{1,\tilde{\lambda}}(x)$ is identical to the probability density function of the exponential distribution (Eq. (2.1)).

Next, let us assume that X_1 is the random variable obeying the probability density function of $g_{p,\tilde{\lambda}}(x)$, and that X_2 is that of $g_{q,\tilde{\lambda}}(x)$. Note that X_1 and X_2 are independent. When $Y_1 = X_1 + X_2$, Y_1 obeys the probability density function of $g_{p+q,\tilde{\lambda}}(x)$ (page 489 in [4]). Hence, when we assume that $p = q = 1$, Y_1 obeys the probability density function of $g_{2,\tilde{\lambda}}(x)$. That is, if X_1 and X_2 are from the same exponential distribution, Y_1 obeys the probability density function of $g_{2,\tilde{\lambda}}(x)$. Assume that the random variables of $\{X_1, X_2, \ldots, X_n\}$ are independent of one another and X_i obeys the probability density function of $g_{p(i),\tilde{\lambda}}(x)$. Then, we can generalize this theorem to prove that the probability density function of $\sum_{i=1}^n X_i$ is $g_{\sum_{i=1}^n p(i),\tilde{\lambda}}(x)$ (page 490 in [4]). Therefore, if we assume that $p(1) = p(2) = \ldots = p(n) = 1$, then $\sum_{i=1}^n X_i$ obeys the probability density function of $g_{n,\tilde{\lambda}}(x)$. This indicates that when the random variables of $\{X_1, X_2, \ldots, X_n\}$ obey the same exponential distribution, $\sum_{i=1}^n X_i$ obeys the probability density function of $g_{n,\tilde{\lambda}}(x)$.

Using this theorem, we can replace Eq. (2.15) with Eq. (3.1) to get

$$E_{\{x_i\}}\left[l^*(\alpha\hat{\lambda})\right] = \int_0^\infty \left(\text{log}\left(\frac{\alpha n}{y}\right) - \frac{\alpha n}{\tilde{\lambda}y}\right)\frac{\tilde{\lambda}^n y^{n-1}\text{exp}(-\tilde{\lambda}y)}{\Gamma(n)}dy. \tag{3.4}$$

To obtain $\hat{\alpha}$, we set the derivative of Eq. (3.4) with respect to α equal to 0. Then, we have

$$\int_0^\infty \left(\frac{1}{\alpha} - \frac{n}{\tilde{\lambda}y}\right)y^{n-1}\text{exp}(-\tilde{\lambda}y)dy = 0. \tag{3.5}$$

The integral of Eq. (3.5) using Mathematica 3.0 (Wolfram Research Inc., Champaign, Illinois, U.S.A.) is

$$-\frac{\tilde{\lambda}^n(\alpha n\Gamma(n-1) - \Gamma(n))}{\alpha} = 0. \tag{3.6}$$

$\Gamma(n+1)$ is the factorial of n, so we can conclude that

$$\hat{\alpha} = 1 - \frac{1}{n}. \tag{3.7}$$

This equation is consistent with the characteristic in the previous section: $\hat{\alpha}$ does not depend upon $\tilde{\lambda}$.

4 Conclusions

In this paper, we have proposed an improvement for the maximum likelihood method for estimating the exponential distribution. This is a rudimentary step when developing better estimators than the maximum likelihood estimator. We found that the constant ($\hat{\alpha}$) that is multiplied by the maximum likelihood estimator does not depend upon the true parameter ($\tilde{\lambda}$) of the population. Hence, our improvement to the maximum likelihood estimator for the exponential distribution is easily implemented, even if we consider the favorable condition that the population has only one parameter. Other estimators do not appear to have this advantage.

In the future, we hope to extend our research to improving the maximum likelihood method for a small data set from a number of different perspectives, and to clarify the nature of this new approach. Moreover, this new field of research should play an important role when developing practical algorithms and software based on adding value to an accumulated series of small data sets.

Competing Interests

The author declares that no competing interests exist.

References

[1] Konishi S, Kitagawa G. Information Criteria and Statistical Modeling; 2007. New York: Springer.

[2] Takezawa K. Flexible Model Selection Criterion for Multiple Regression. Open Journal of Statistics. 2012;2(4):401-407.

[3] Takezawa K. Learning Regression Analysis by Simulations; 2013. New York: Springer.

[4] Bickel PJ, Doksum KA. Mathematical Statistics: Basic Ideas and Selected Topics. 2000;I(2nd Edition). New Jersey: Prentice Hall.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=707&id=6&aid=6656