British Journal of Mathematics & Computer Science
11(4): 1-11, 2015, Article no.BJMCS.20522
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Estimation of the Geometric 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/2015/20522
Editor(s):
(1) Metin Basarir, Department of Mathematics, Sakarya University, Turkey.
(2) Dijana Mosic, Department of Mathematics, University of Nis, Serbia.
Reviewers:
(1) Anonymous, SASTRA University, Kumbakonam, India.
(2) Diana Blkova, University of Economics, Prague; Czech Republic.
(3) Manoel F Borges, UNESP-Sao Paulo State University, Brazil.
Complete Peer review History: http://sciencedomain.org/review-history/11298

Original Research Article

Received: 30 July 2015
Accepted: 21 August 2015
Published: 06 September 2015

Abstract

The maximum likelihood method in view of future data (i.e., the maximization of expected log-likelihood) enables estimates of geometric distribution parameter. This estimator is defined as an estimator in which n (number of data) in the maximum likelihood estimator is replaced with $(n + a_0)$; a_0 takes a value such as -1 or -0.5. The value of a_0 reflects knowledge about the range where the parameter is to be found. Therefore, when we know that the true parameter of a population lie in a particular range, this method gives a larger expected log-likelihood than the maximum likelihood estimator. Simple simulations show that this new estimator gives anticipated results. The characteristic of the estimator with $(n + a_0)$ is similar to that for the mean squared error (MSE), that is, the expectation of the sum of the squared difference between the true parameter and its estimate. This new methodology in which estimators are modified using some constants for yielding better estimators in terms of prediction will contribute to various fields where the number of data is not very large.

Keywords: Expected log-likelihood; geometric distribution; maximum likelihood estimator; optimization.

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

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

1 Introduction

The maximum likelihood method does not always maximize the expected log-likelihood (page 35 in [1]). Hence, approaches other than the maximum likelihood method may be preferable if estimations are need for the purpose of prediction. Such estimators are hereinafter referred to as predictive estimators. For example, if the variance of a normal distribution is estimated, the "third variance" gives a larger value of expected log-likelihood than the maximum likelihood variance and unbiased variance ([2], Section 5.5 of [3]). Furthermore, to estimate the parameter of an exponential distribution, the multiplication of the maximum likelihood estimator by $\left(1 - \frac{1}{n}\right)$ (n is the number of data) results in a larger value for the expected log-likelihood than simply using the maximum likelihood estimator ([4]). Similar results are obtained with the binomial distribution ([5]).

This paper therefore presents better estimators than the maximum likelihood estimator for estimating the geometric distribution parameter. Although these new estimators are better than the maximum likelihood estimator in terms of expected log-likelihood, limitation criteria must be imposed. Section 2 shows that an approximate predictive estimator for the geometric distribution can be obtained. Section 3 derives the conditions under which the predictive estimator can be used in numerical simulations.

2 Predictive Estimator of Geometric Distribution

We begin with some basic definitions. For a geometric distribution, the probability density function is

$$f(\xi) = \tilde{p}(1 - \tilde{p})^{\xi - 1}, \tag{2.1}$$

and its expectation is

$$\sum_{r=1}^{\infty} r f(r) = \frac{1}{\tilde{p}}. \tag{2.2}$$

The random variable which obeys the probability density function ($f(\xi)$) is denoted X. The realization (i.e., data) of this random variable is denoted by $\{x_i\}$ ($1 \leq i \leq n$). Given these data, the log-likelihood ($l(p|\{x_i\})$) of p is

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

To derive the value of p which maximizes this value, the above equation is differentiated with respect to p and set equal to 0. The result is

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

Hence, we have the maximum likelihood estimator:

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}, \tag{2.5}$$

where $\hat{p}$ indicates the maximum likelihood estimator given the data ($\{x_i\}$ ($1 \leq i \leq n$)).

Next, future data are denoted $\{x_i^*\}$ ($1 \leq i \leq m$). Given these data, the log-likelihood ($l(\hat{p}|\{x_i^*\})$) of $\hat{p}$ is represented as

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

Let the estimate $\hat{p}$ that maximizes this value be denoted $\hat{p}^*$. Then $\hat{p}^*$ is written as

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

Because the number of future data is infinite, we set m to infinity. Under this condition, $\hat{p}^*$ is denoted $\hat{p}^*_\infty$ and hence Eq. (2.2) results in

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

That is, the maximum likelihood estimator given by an infinite number of future data is the true parameter of the population ($\tilde{p}$). Substitution of Eq. (2.8) into Eq. (2.6) yields

$$\lim_{m\to\infty} \frac{l(\hat{p}|\{x_i^*\})}{m} = \log(\hat{p}) + \log(1-\hat{p})\Big(\frac{1}{\tilde{p}} - 1\Big). \tag{2.9}$$

However, given an infinite number of future data, $\hat{p}$ given by Eq. (2.5) is not guaranteed to be the optimal estimator. This optimal estimator is denoted by $\alpha\hat{p} = \dfrac{\alpha n}{\sum_{i=1}^{n} x_i}$; here α is a constant less than or equal to 1. Therefore, the log-likelihood is

$$\lim_{m\to\infty} \frac{l^*(\alpha\hat{p})}{m} = \log(\alpha\hat{p}) + \log(1-\alpha\hat{p})\Big(\frac{1}{\tilde{p}} - 1\Big). \tag{2.10}$$

Substitution of Eq. (2.5) results in

$$\lim_{m\to\infty} \frac{l^*(\alpha\hat{p})}{m} = \log\Big(\frac{\alpha n}{\sum_{i=1}^{n} x_i}\Big) + \log\Big(1 - \frac{\alpha n}{\sum_{i=1}^{n} x_i}\Big)\Big(\frac{1}{\tilde{p}} - 1\Big). \tag{2.11}$$

If α is larger than 1, the argument of the log in the second term on the right-hand side may be negative. Hence, α must be less than or equal to 1. Moreover, if all of the elements of $\{x_i\}$ are 1, use of the maximum likelihood method (i.e., $\alpha = 1$) causes the argument of the log factor on the right-hand side of Eq. (2.11) to be 0. Therefore, the value of Eq. (2.11) becomes $-\infty$ whatever the value of $\tilde{p}$ may be. Indeed, if all elements of $\{x_i\}$ are 1, $\tilde{p}$ cannot be estimated from the log-likelihood of the maximum likelihood estimator. Thus, if all elements of $\{x_i\}$ are 1, an estimate of the parameter using the method suggested here is not possible.

To maximize Eq. (2.10), its right-hand side is differentiated with respect to α and set equal to 0. We then have

$$\hat{\alpha}\hat{p} = \tilde{p}. \tag{2.12}$$

This equation indicates that $\hat{p}$ multiplied by the optimal value of α obviously gives the true value of p. Because the value of $\tilde{p}$ is unknown in most situations, Eq. (2.12) cannot be calculated.

Thus, let us consider the mean of $\lim_{m\to\infty} \dfrac{l^*(\alpha\hat{p})}{m}$ given by sampling x an infinite number of times; that is, the expectation of $l^*(\alpha\hat{p})$ is required. Note that as data in which all the elements of $\{x_i\}$ are 1 are not treated here, the sampling must be redone. This expectation, calculated using Eq. (2.11), is then

$$E_x\left[\lim_{m\to\infty} \frac{l^*(\alpha\hat{p})}{m}\right] = E_x\left[\log\Big(\frac{\alpha n}{\sum_{i=1}^{n} x_i}\Big) + \log\Big(1 - \frac{\alpha n}{\sum_{i=1}^{n} x_i}\Big)\Big(\frac{1}{\tilde{p}} - 1\Big)\right]. \tag{2.13}$$

Such expectations are used in deriving *AIC*. For example, $E_{G(\mathbf{x}_n)}$ on page 55 in [1] is such an expectation.

If X_i has a geometric distribution, $\displaystyle\sum_{i=1}^{n} X_i$ obeys the negative binomial distribution (e.g., page 87 in

[6]). Hence, Eq. (2.13) is transformed into

$$E_x\left[\lim_{m\to\infty}\frac{l^*(\alpha\hat{p})}{m}\right]=\frac{\sum_{j=n+1}^{\infty}\left(\log\left(\frac{\alpha n}{j}\right)+\left(\frac{1}{\tilde{p}}-1\right)\log\left(1-\frac{\alpha n}{j}\right)\right){}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}}{\sum_{j=n+1}^{\infty}{}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}}. \tag{2.14}$$

Note that the summation on the right-hand side of this equation begins with $j=n+1$ instead of $j=n$. Additionally, a standardization using $\sum_{j=n+1}^{\infty}{}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}$ is performed. Using this strategy, if all the elements of $\{x_i\}$ are 1, the sampling is performed again. To derive an α that maximizes Eq. (2.14), we differentiate this equation with respect to α and set it to 0 to yield

$$\sum_{j=n+1}^{\infty}\left(\frac{1}{\alpha}-\left(\frac{1}{\tilde{p}}-1\right)\frac{n}{j-\alpha n}\right){}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}=0. \tag{2.15}$$

Although it is not easy to obtain analytically, we denote solutions for α by $\hat{\alpha}$.

Next, we replace $\hat{p}$ in Eq. (2.10) with $\hat{p}^+$ expressing it in the form

$$\hat{p}^+=\frac{n+a_0}{\sum_{i=1}^{n}x_i}, \tag{2.16}$$

where a_0 is a constant. Hence, Eq. (2.14) is transformed into

$$\begin{aligned}&E_x\left[\lim_{m\to\infty}\frac{l^*\left(\alpha\frac{n+a_0}{\sum_{i=1}^{n}x_i}\right)}{m}\right]\\&=\frac{\sum_{j=n+1}^{\infty}\left(\log\left(\alpha\frac{n+a_0}{j}\right)+\left(\frac{1}{\tilde{p}}-1\right)\log\left(1-\alpha\frac{n+a_0}{j}\right)\right){}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}}{\sum_{j=n+1}^{\infty}{}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}}.\end{aligned} \tag{2.17}$$

This equation is then differentiated with respect to α and set equal to 0. We have

$$\sum_{j=n+1}^{\infty}\left(\frac{1}{\alpha}-\left(\frac{1}{\tilde{p}}-1\right)\frac{n+a_0}{j-\alpha(n+a_0)}\right){}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}=0. \tag{2.18}$$

If we set $a_0=-1$ and $\alpha=1$, the left-hand side of this equation becomes

$$\sum_{j=n+1}^{\infty}\left(1-\left(\frac{1}{\tilde{p}}-1\right)\frac{n-1}{j-n+1}\right){}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}. \tag{2.19}$$

To simplify this sum, the following sum

$$\sum_{j=n+1}^{\infty}\left(\frac{n-1}{j-n+1}\right){}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n} \tag{2.20}$$

is evaluated using Mathematica 3.0 yielding the result (Fig. 1)

$$\sum_{j=n+1}^{\infty}\left(\frac{n-1}{j-n+1}\right){}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}=-\frac{\tilde{p}-n\tilde{p}^n-\tilde{p}^{n+1}+n\tilde{p}^{n+1}}{1-\tilde{p}}. \tag{2.21}$$

Hence, if $\tilde{p}^n$ is assumed to be 0 and the result of this approximation is substituted into Eq. (2.19), we have

$$\sum_{j=n}^{\infty}\left(1-\left(\frac{1}{\tilde{p}}-1\right)\frac{n-1}{j-n+1}\right){}_{j-1}C_{n-1}\tilde{p}^n(1-\tilde{p})^{j-n}\approx 0. \tag{2.22}$$

Therefore, with $a_0 = -1$ and $\alpha = 1$, Eq. (2.18) holds approximately regardless of the value of $\tilde{p}$. This implies that $\hat{p}^+$ (Eq. (2.16)) with $a_0 = -1$ is an approximate predictive estimator which can be used even if the value of $\tilde{p}$ are unknown.

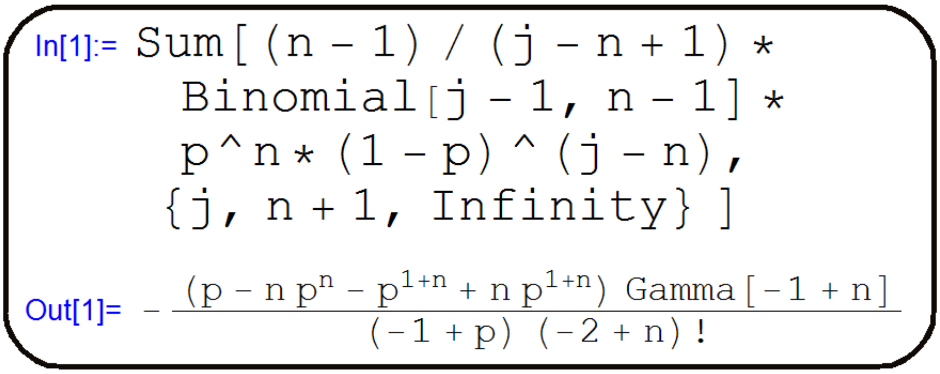

Fig. 1. Result of the calculation of Eq. (2.20) using Mathematica 3.0.

3 Numerical Simulation

Section 2 infers that for $a_0 = -1$ and $\alpha = 1$, Eq. (2.18) is satisfied approximately. This result, however, is based on assuming $\tilde{p}^n$ is 0. To find out under which conditions $\tilde{p}^n$ is close to 0 and whether the left-hand side of Eq. (2.18) is effectively 0, a numerical simulation is required.

Numerical simulations were performed setting $\alpha = 1$ in Eq. (2.17), presuming $a_0 = 0$ (maximum likelihood estimator) or $a_0 = -1.0$ (predictive estimator). The infinite sum $\sum_{j=n+1}^{\infty}$ is then approximated by a finite sum $\sum_{j=n+1}^{2,000}$ in the right-hand side of Eq. (2.17). As the results of this approximation are close to those using the finite sum $\sum_{j=n+1}^{1,000}$, the use of finite sum $\sum_{j=n+1}^{2,000}$ is justified. The value of $\tilde{p}$ is assumed to be one of the set $\{0.05, 0.1, 0.15, \ldots, 0.95\}$. The value of Eq. (2.17) setting $a_0 = 0$ is denoted l_{like} and that setting $a_0 = -1$ is denoted l_{pre}. The values of $(l_{pre} - l_{like})$ are plotted in Figs. 2 (left $n = 5$ and right $n = 10$), and 3(left $n = 20$ and right $n = 30$). The values of $(l_{pre} - l_{like})$ are positive in certain intervals. Specifically if $n = 5$, the interval is $0.05 \le \tilde{p} \le 0.55$; for $n = 10$, $0.05 \le \tilde{p} \le 0.75$ for $n = 20$, $0.05 \le \tilde{p} \le 0.85$; and for $n = 30$, $0.05 \le \tilde{p} \le 0.90$. As $a_0 = -1.0$ is based on the assumption that 0 is a good approximation of $\tilde{p}^n$, we presume that $(l_{pre} - l_{like})$ takes a positive value when the value of $\tilde{p}$ is small and the value of n is large; the positive values of $(l_{pre} - l_{like})$ mean that the predictive estimator ($a_0 = -1.0$) performs better than the maximum likelihood estimator ($a_0 = 0$). Figs. 2 and 3 indicate that this assumption is reasonable.

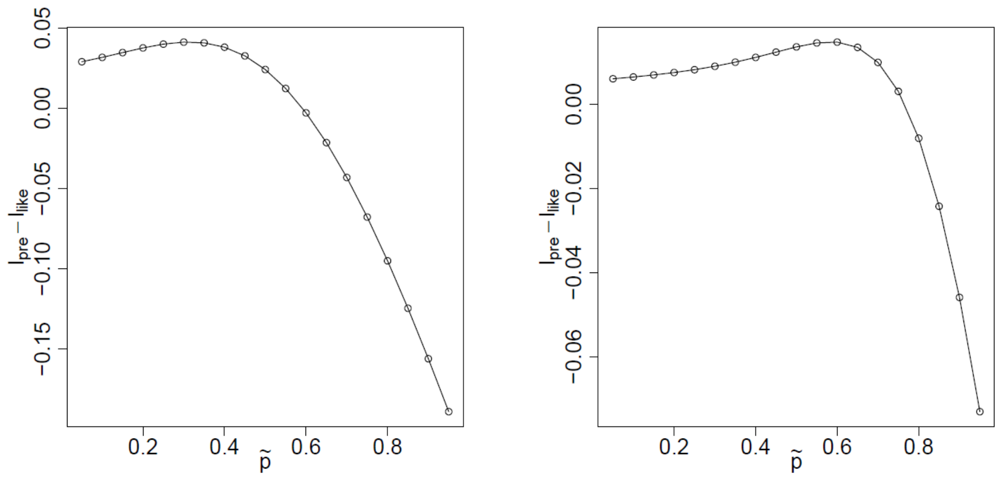

Fig. 2. Values of $(l_{pre} - l_{like})$ setting $a_0 = -1$ in the predictive estimator; $\tilde{p}$ is assumed to be $\{0.05, 0.1, 0.15, \ldots, 0.95\}$. $n = 5$ (left) and $n = 10$ (right).

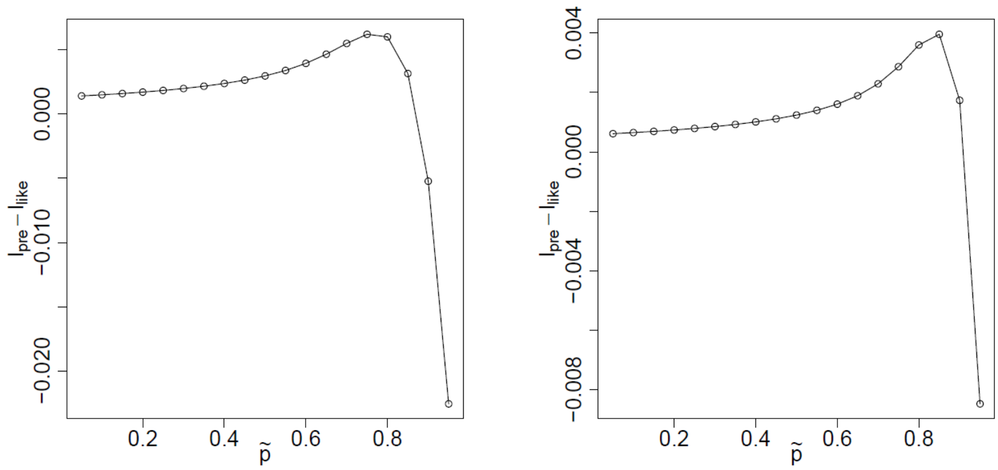

Fig. 3. As for Fig. 3 but with $n = 20$(left). $n = 30$(right).

To extend the interval over which the predictive estimator gives better results than the maximum likelihood estimator, one possible strategy is to make the predictive estimator closer to the maximum likelihood estimator. Assuming $a_0 = -0.5$, we then obtain Figs. 4 and 5. For the various integer values n, the values of $(l_{pre} - l_{like})$ are positive over certain intervals. Specifically, for $n = 5$ the interval is $0.05 \le \tilde{p} \le 0.65$; for $n = 10$, it is $0.05 \le \tilde{p} \le 0.80$; $n = 20$, $0.05 \le \tilde{p} \le 0.90$; and $n = 30$, $0.05 \le \tilde{p} \le 0.90$. Compared with setting $a_0 = -1$, setting $a_0 = -0.5$ widens the interval over which the predictive estimator yields better results that the maximum likelihood estimator. This result shows that the predictive estimator should be adjusted depending on what is a priori known of $\tilde{p}$.

To offer a more realistic analysis of our predictive estimators and the maximum likelihood estimator, we compared the values obtained from Eq. (2.17) given by the two estimators using pseudo-random numbers obeying a geometric distribution. Setting the parameter for the geometric distribution to $\tilde{p} = 0.3$ and $\tilde{p} = 0.8$, 10 data points were sampled from this population using pseudo-random numbers. If all of the 10 data points were 1, the sampling was redone. Using the sampled data, the values of $\hat{p}^{+}$ were calculated setting either $a_0 = -1$ or $a_0 = 0$. Then, the log-likelihood, assuming an infinite number of future data using $\hat{p}^{+}$, is given by either setting of $\hat{p}$ in Eq. (2.9). This calculation was performed $2,000$ times by altering the initial value of the pseudo-random number and the average of the $2,000$ obtained log-likelihood values. This numerical simulation was conducted $1,000$ times. The resultant log-likelihood histograms were generated (Figs. 6 and 7). Setting $\tilde{p} = 0.3$, the predictive estimator performs better than the maximum likelihood estimator. In contrast, with $\tilde{p} = 0.8$, the reverse trend is evident. This tendency coincides with that of Fig. 2(right).

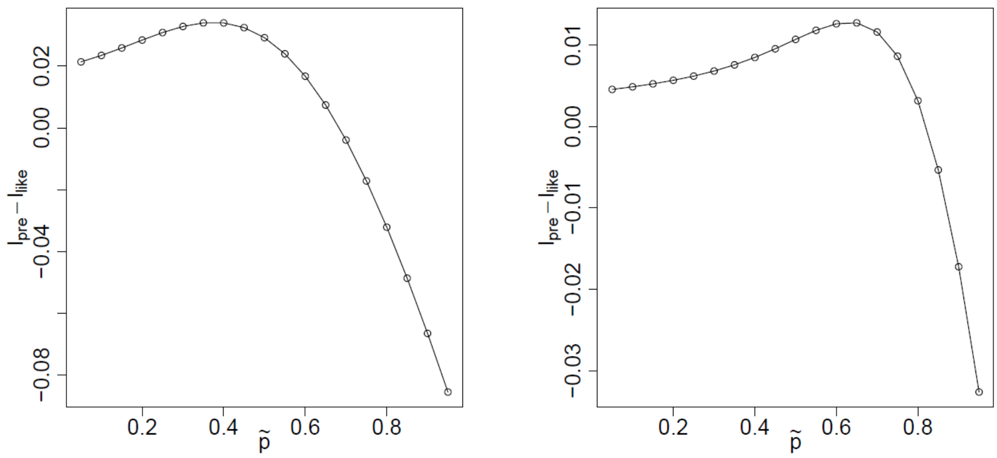

Fig. 4. Values of $(l_{pre} - l_{like})$ setting $a_0 = -0.5$ in the expression for the predictive estimator with $\tilde{p} \in \{0.05, 0.1, 0.15, \ldots, 0.95\}$; left: $n = 5$, right: $n = 10$.

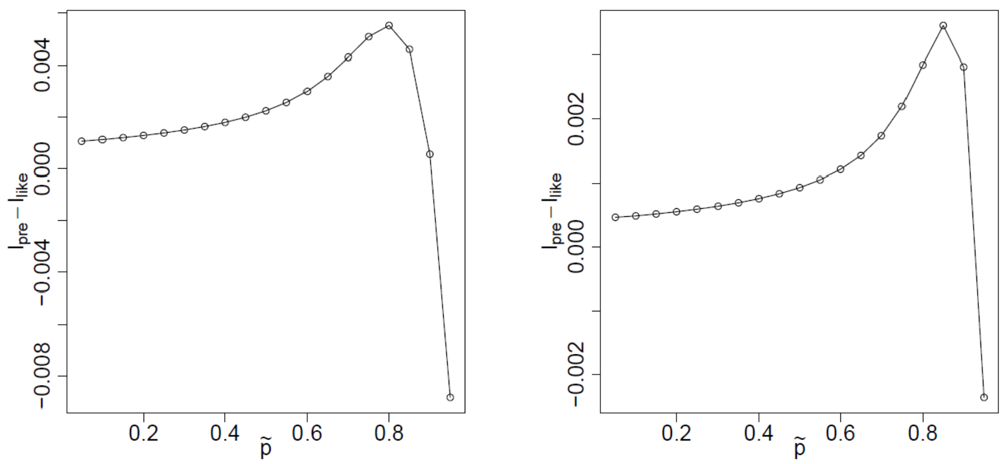

Fig. 5. Values of $(l_{pre} - l_{like})$ when $a_0 = -0.5$ is set in the predictive estimator with $\tilde{p} \in \{0.05, 0.1, 0.15, \ldots, 0.95\}$; left: $n = 20$, right: $n = 30$.

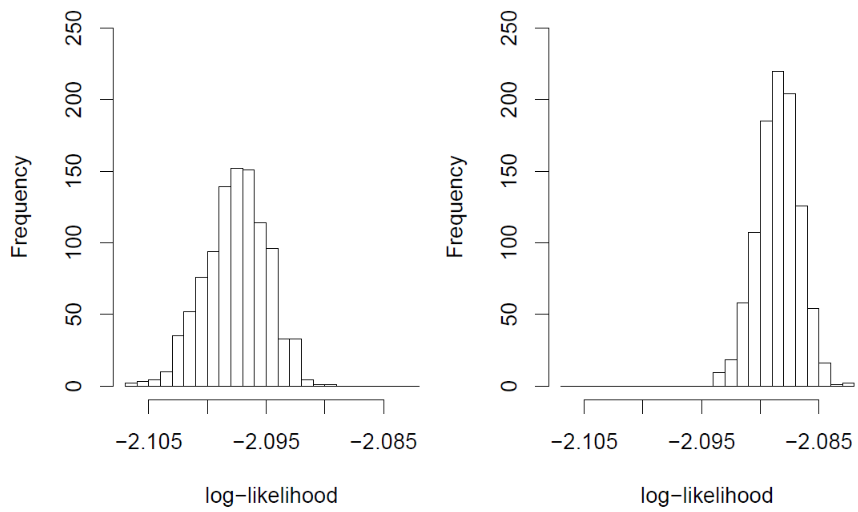

Fig. 6. Comparison between the maximum likelihood estimator ($a_0 = 0$) and the predictive estimator ($a_0 = -1$) in terms of the approximated expected log-likelihood for $n = 10$ and $\tilde{p} = 0.3$. The left histogram gives the distribution of $1,000$ simulated values of the approximated expected log-likelihood given by the maximum likelihood estimator. The average of these values is -2.097548 and the unbiased variance is 6.765919×10^{-6}. The right histogram gives the distribution of $1,000$ simulated values of the approximated expected log-likelihood given by the predictive estimator. The average of these values is -2.088503 and the unbiased variance is 3.116544×10^{-6}.

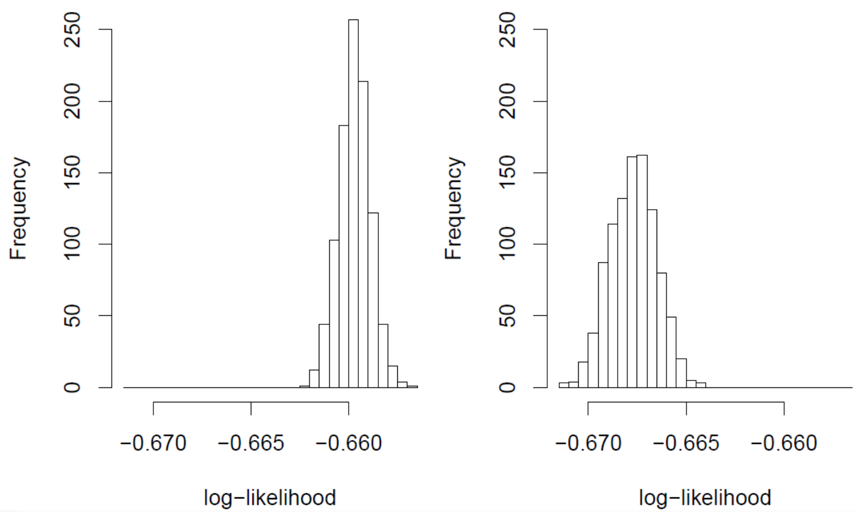

Fig. 7. As for Fig. 6 but with $\tilde{p} = 0.8$. For the left histogram, the average is -0.6596949 and the unbiased variance is 6.558547×10^{-7}. For the right histogram, the average is -0.6677127 and the unbiased variance is 1.412698×10^{-6}.

4 Conclusions

The following passage is taken from page 332 in [7]:

> In general, since MSE is a function of the parameter, there will not be one "best" estimator. Often, the MSEs of two estimators will cross each other, showing that each estimator is better (with respect to the other) in only a portion of the parameter space. However, even this partial information can sometimes provide guidelines for choosing between estimators.

If "MSE" in the above quote is replaced with the "expected log-likelihood", the passage would represent our conclusions accurately:

(1) The value of the expected log-likelihood depends on the true values of the parameters. Hence, these true values are usually needed to obtain predictive estimator values that strictly maximize the expected log-likelihood ([5]). Estimates of the third variance ([2]) and parameter estimation of the exponential distribution ([4]) are exceptional in that true parameter values are not required to obtain the predictive estimator.
(2) A predictive estimator is not unique. For example, when estimating the parameter that specifies the geometric distribution, an estimator of the form $\alpha\hat{p}$ ($\hat{p}$ is the maximum likelihood estimator) is assumed and Eq. (2.17) gives various predictive estimators by setting for example $a_0 = -1$ or $a_0 = -0.5$. Moreover, although Eq. (2.17) is obtained, assuming Eq. (2.22) holds approximately, other predictive estimators are given if an exact solution is found or another approximate expression is generated.
(3) The relative merits of the predictive estimators depend on the interval over which the true parameter exists.
(4) Using the characteristics shown in (3), knowing a little about the parameters helps in the choice of predictive estimator which is more useful than the maximum likelihood estimator.

Definition 7.3.1 on page 330 in [7] contains θ. This indicates that MSE is also one of the criteria derived given an infinite number of future data because such data are required for obtaining the true parameters. Therefore, the estimator given by MSE is a predictive estimator in the broad sense of the term. The difference between MSE and the expected log-likelihood is that the distance between the true parameters given by the infinite number of future data and the estimates of the parameters given by the data is defined in a different manner. However, both estimators are expected to yield beneficial estimates for the purpose of prediction. Nevertheless, the maximum likelihood estimator and the unbiased estimator play main roles in the conventional estimation. This shows that we have not placed any emphasis on finding that the expected log-likelihood is "a formal extension of the classical maximum likelihood" ([8]), except in the context of model selection by AIC or related measures. One of the main reasons for this is that the maximum likelihood and the unbiased estimators lead to mathematically tractable problems on most occasions, whereas predictive estimators inherit the defects raised by (1), (2), and (3) above. Although the maximum likelihood and the unbiased estimators do not always lead to useful results in terms of prediction, we have used them in most situations because the maximum likelihood and the unbiased estimators are treated in a mathematically simple and efficient manner. In this age of sophisticated computers, this strategy requires rethinking. If we construct better estimators than the maximum likelihood and the unbiased estimators by making good use of the abovementioned (4) and by using computers skillfully, our estimations can extract more information from a small number of data.

Although MSE is thought not to be suitable for estimating scale parameters, it works well in estimating location parameters (page 332 in [7]). Therefore, for general estimations of, for example, variance, the expected log-likelihood should be the better choice. Hence, parameter estimation for the purpose of maximizing expected log-likelihood takes advantage of the characteristics of estimation using MSE. That is, both estimations are based on goodness of fit to the infinite number of future data. Both methods are justified qualitatively by considering that results given by maximizing the expected log-likelihood are similar to those by minimizing MSE. Estimation by maximizing the expected log-likelihood is, however, more sophisticated.

This consideration is made clearer by taking the variance estimation as an example. If MSE is used in estimating the variance of the normal distribution, the maximum likelihood estimator proves to be more favorable than the unbiased estimator (page 331 in [7]). Additionally, if we minimize MSE with the condition that the estimator has the same form as the maximum likelihood variance, the number of data n in the maximum likelihood variance is replaced with $(n + 1)$ (page 414 in [9]). That is, the optimal variance in terms of MSE is smaller than the maximum likelihood variance. This results because MSE tends to give a larger penalty for overestimations and a smaller penalty for underestimations (page 332 in [7]). Therefore, a reasonable assertion is that MSE leads to a smaller variance than the maximum likelihood variance not because MSE gives the distance based on an infinite number of future data but because the distance used here tends to result in a smaller variance. Indeed, because the third variance is based on the expected log-likelihood, which is a more effective criterion than the MSE, it results in a larger variance than the maximum likelihood variance. That is, whereas the third variance inherits characteristics from variance estimation using MSE in terms of considering the goodness of fit to an infinite number of future data, it redresses the shortcomings of MSE which leads to underestimation. This is a good example showing that although MSE is a useful criterion, the expected log-likelihood is more useful. For more general discussions on variance estimation, refer to [10 - 14].

Although we are in the age of big data, the number of data is not very large in some domains of science. For example, one experiment ranges over periods of one year to several decades in agricultural science. We should make the best possible use of information contained in data for leading to high predictability in such areas. For this purpose, prior knowledge that the true parameters of a population lie in a particular range should be used efficiently. The methodology suggested here, therefore, will contribute greatly to maximization of the beneficial use of data. To realize this goal, the details of the predictive estimator based on the expected log-likelihood still need to be examined from a number of different perspectives.

Acknowledgement

The author is very grateful to the referees for carefully reading the paper and for their comments and suggestions which have improved the paper.

Competing Interests

The author declares that no competing interests exist.

References

[1] Konishi S, Kitagawa G. Information criteria and statistical modeling. New York: Springer; 2008.

[2] Takezawa K. A revision of AIC for normal error models. Open Journal of Statistics. 2012;2(3): 309-312.

[3] Takezawa K. Learning regression analysis by simulation. New York: Springer; 2013.

[4] Takezawa K. Estimation of the exponential distribution in the light of future data. British Journal of Mathematics & Computer Science. 2015;5(1):128-132.

[5] Takezawa K. Estimation of binomial distribution in the light of future data. British Journal of Mathematics & Computer Science. 2015; 10(2):1-7.

[6] Bickel PJ, Doksum KA. Mathematical statistics: Basic ideas and selected topics, (2nd edition). New Jersey: Prentice Hall. 2000;1.

[7] Casella G, Berger RL. Statistical inference (2nd edition). Pacific Grove (CA): Duxbury; 2001.

[8] Akaike H. Information theory and extension of the maximum likelihood principle. In Second international symposium of information theory, N. B. Petrov and F. Czàki (eds.),. Budapest: Akademiai Kiàdo. 1973;267-281.

[9] DeGroot MH. Probability and statistics (2nd edition). Boston: Addison–Wesley; 1986.

[10] Stein C. Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. Annals of the Institute of Statistical Mathematics. 1964;16(1): 155-160.

[11] Rukhin AL. How much better are better estimators of a normal variance. Journal of the American Statistical Association. 1987;82(399): 925-928.

[12] Goutis C, Casella G. Improved invariant confidence intervals for a normal variance. The Annals of Statistics. 1991;19(4): 2015-2031.

[13] Rukhin AL, Ananda MMA. The risk behavior of estimators of normal variance in linear models. Statistics and Probability Letters. 1992;13:159-166.

[14] Ohtani K. The exact distribution and density functions of the stein-type estimator for normal variance. Communications in Statistics - Theory and Methods. 1993;22(10): 2863-2876.

©2015 Takezawa; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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)
http://sciencedomain.org/review-history/11298