Confidence interval estimation of gamma distribution lifetime data using score and bootstrap methods

Larasati, Rudi Ruswandi, Fitriani*

Universitas Lampung, Indonesia

ARTICLE INFO

Article History
Received : 15-12-2020
Revised : 25-02-2021
Accepted : 23-03-2021
Published : 27-03-2021

Keywords: Confidence Interval Estimation, Gamma Distribution, Maximum Likelihood Estimation, Score Method, Bootstrap Method, Reliability

*Correspondence E-mail: fitriani.1984@fmipa.unila.ac.id

Doi: 10.24042/djm.v4i1.7737

INTRODUCTION

Statistics is an analytical tool that has a significant role in various fields of applied science. One of them is the reliability analysis or the analysis of the system life span’s reliability with continuous random variables. This research is developing previous research conducted by Sangnawakij and Niwitpong in 2019 regarding confidence intervals in gamma distribution parameters (Sangnawakij et al., 2019).

Reliability theory is used to determine the probability of the functioning of a system consisting of several components. In other words, the theory of reliability is increasingly needed in life to improve the product or system performance. Reliability is the chance that a component or system is functioning correctly for a certain period.

Reliability test analysis can be determined by determining reliability function, probability density function, and hazard function. A distribution is needed to perform a reliability test analysis. Some distributions that can describe a system's life span include the Gamma Distribution, the Weibull Distribution, and the Lognormal Distribution. The distribution used in this study is the Gamma
distribution. The Gamma distribution was chosen because it has the most extensive application in analyzing reliability test data. In inferencing statistics, a method is needed to conclude population parameters based on sample data. The method used to estimate the parameters in this research is the Maximum Likelihood method and the Bayes method. The method used to estimate the confidence interval is the Score method, and the data we use is the aircraft engine cooling system's lifetime data.

METHOD

The reliability function (survival) is the probability that an individual can survive until time \( t \) \( (t > 0) \).

\[
R(t) = P(T \geq t) = 1 - F(t) \tag{1}
\]

Furthermore, the Hazard function of an individual's chance of dying in the time interval from \( t \) to \( t + \Delta t \); if known, the individual can still survive up to time \( t \) (Lawless, 1982). The Hazard rate order is an ordering for random variables that compare lifetime to their hazard rate function. It is stronger than the usual stochastic order for random variables yet is weaker than the likelihood ratio ordering (Boland et al., 1994).

The method for estimating the value of distribution parameters from data in the reliability function (survival) uses the maximum likelihood and Bayes methods. The maximum likelihood method uses the parameter space \( \Omega \), which corresponds to the maximum likelihood value of the observed data as an estimate of the unknown parameter. The likelihood function \( L(\theta) \) in its application shows the joint probability density function of a random sample (Bain & Engelhardt, 1992).

Maximum likelihood is widely used to successfully construct statistical estimators for probability distribution parameters (Bertsimas & Nohadani, 2019). If the joint probability density function of \( n \) random variables \( X_1, X_2, ..., X_n \) observed at \( x_1, x_2, ..., x_n \) is denoted by \( f(x_1, x_2, ..., x_n) \), then the likelihood function of the set of observations \( x_1, x_2, ..., x_n \) is expressed as:

\[
L(\theta) = f(x_1; \theta)f(x_2; \theta) ... f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) \tag{2}
\]

With \( \theta \) unknown parameter.

The maximum likelihood estimate of \( \theta \) is obtained by solving the \( \frac{d}{d\theta} \ln L(\theta) = 0 \). Suppose there are \( k \) unknown parameters, then the likelihood parameter estimator of \( \theta_i \) is obtained by solving

\[
\frac{d}{d\theta} \ln L(\theta_1, \theta_2, ..., \theta_k) = 0, \tag{3}
\]

where \( i = 1, 2, 3, ..., k \)

Let \( S \) be the sample space of an experiment and \( A_1, A_2, ..., A_k \) are the events in \( S \) such that \( S \) are independent of each other and \( \bigcup_{i=1}^{k} A_i = S \). It is said that \( A_1, A_2, ..., A_k \) form partitions inside \( S \).

![Figure 1. Bayes' Theorem](image)

If \( k \) events \( A_1, A_2, ..., A_k \) form a partition in \( S \), and then it can be seen in Figure 1 that events \( A_1 \cap B, A_2 \cap B, ..., A_k \cap B \) form partitions in \( B \). So that \( B = (A_1 \cap B) \cup (A_2 \cap B) \cup ... \cup (A_k \cap B) \). Because the events on the right side are independent of each other, then:

\[
P(B) = \sum_{i=1}^{k} P(A_i \cap B) \tag{4}
\]

If \( P(A_i > 0) \) for \( i = 1, 2, ..., k \) then \( P(A_i \cap B) = P(A_i)P(B|A_i) \). So that it is obtained \( P(B) = \sum_{i=1}^{k} P(A_i)P(B|A_i) \). Suppose the events \( A_1, A_2, ..., A_k \) form a partition in the sample space \( S \) such that \( P(A_i > 0) ; i = 1, 2, ..., k \) and let \( B \) be any events such that \( P(B) > 0 \). So for \( i = 1, 2, ..., k \)
\[
P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)} \tag{5}
\]

Bayes’ theorem provides a simple rule for calculating the conditional probability of events \(A_i\) given \(B\) to occur, if each of the unconditional probabilities \(A_i\) and conditional probabilities \(B\) given \(A_i\) occur is known (Soejati & Subanar, 1988). The interval estimator determines the value interval calculated based on statistical measurements and has a certain probability, and will contain the actual parameters.

Let \(\hat{\theta}_L\) and \(\hat{\theta}_U\) be the lower and upper limits for parameter \(\theta\), if
\[
P\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U\right) = 1 - \gamma
\]

In estimating the parameter confidence interval using the Score method, the score or Rao statistics are denoted as:
\[
W_u = U^T (\alpha_0, \hat{\beta}_0) I^{-1}(\alpha_0, \hat{\beta}_0) U(\alpha_0, \hat{\beta}_0) \tag{6}
\]

where \(\hat{\beta}_0\) is the maximum likelihood estimator for \(\beta\) under the null hypothesis \(H_0: \alpha = \alpha_0, \ U(\alpha_0, \hat{\beta}_0)\) is the vector of the score function, and \(I(\alpha_0, \hat{\beta}_0)\) is a matrix from fisher information. \(W_u\) has a Chi-squared distribution. The moment generating function of the Chi-squared distribution with degrees of freedom 1 is \((1 - 2t)^{-1/2}\). If \(Z \sim N(0,1)\), then \(Z^2 \sim \chi^2_1\) (Sangnawakij & Niwitpong, 2017).

Parameterized in terms of a shape parameter \(\alpha\) and an inverse scale parameter \(\beta = 1/\theta\), called a rate parameter (J et al., 2019).

The probability density function of the gamma distribution is as follows:
\[
f(t) = \begin{cases} 
\frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} e^{-\frac{t}{\beta}}, & t > 0 \\
0, & \text{others}
\end{cases}
\]

(Abid & Al-Hassany, 2016). Estimation of parameters using the maximum likelihood method:

The Gamma Distribution Likelihood function is as follows:
\[
L(\alpha, \beta) = \prod_{i=1}^{n} f(t_i; \alpha, \beta) = \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right)^n \prod_{i=1}^{n} t_i^{\alpha-1} e^{-\frac{\sum_{i=1}^{n} t_i}{\beta}}
\]

The natural logarithm is obtained as follows:
\[
\ln L(\alpha, \beta) = \ln\left(\left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right)^n \prod_{i=1}^{n} t_i^{\alpha-1} e^{-\frac{\sum_{i=1}^{n} t_i}{\beta}}\right) = -n \ln \Gamma(\alpha) - n\alpha \ln \beta + n(\alpha - 1) \ln t - \frac{n\bar{t}}{\beta}
\]

(Khamkong, 2018)

The derivative of the natural logarithm of \(\beta\) parameter:
\[
\frac{\partial \ln L(\alpha, \beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \left[-n \ln \Gamma(\alpha) - n\alpha \ln \beta + n(\alpha - 1) \ln t - \frac{n\bar{t}}{\beta}\right] = 0
\]

\(\hat{\beta} = \frac{\bar{t}}{\alpha}\) \tag{8}

The derivative of the natural logarithm of \(\alpha\) parameter:
\[
\frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left[-n \ln \Gamma(\alpha) - n\alpha \ln \beta + n(\alpha - 1) \ln t - \frac{n\bar{t}}{\beta}\right] = 0
\]

\(\hat{\alpha} = \frac{1}{2(\ln \bar{t} - \ln t)}\) \tag{9}

By substituting the estimated value \(\hat{\alpha}\) in Equation (9) to Equation (8), we obtained:
\[
\hat{\beta} = \frac{\bar{t}}{\hat{\alpha}}\]
Likelihood function with the following prior distribution:

\[ \pi(\alpha, \beta | T) = C \cdot f(T | \alpha, \beta) \cdot f(\alpha) \cdot f(\beta) \]

\[ = \left( \frac{1}{(\beta \alpha (\Gamma \alpha)) \Gamma(\alpha)^n} \right)^{\alpha - 1} \exp\left( - \sum_{i=1}^{n} \frac{t_i}{\beta} \right) \int_0^\infty \frac{\lambda^{n-1} \Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\eta \alpha - 1}} d\alpha \]  

(11)

The marginal posterior distribution for \( \alpha \) can be obtained by integrating the posterior distribution of \( \pi(\alpha, \beta | T) \) to \( \beta \).

\[ \pi(\alpha | T) = \int_0^\infty \pi(\alpha, \beta | T) d\beta \]

\[ = \int_0^\infty \frac{\lambda^{n-1} \Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\alpha - 1}} \cdot \frac{\Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\alpha - 1}} d\alpha \]

The marginal posterior distribution for \( \beta \) can be obtained by integrating the posterior distribution of \( \pi(\alpha, \beta | T) \) to \( \alpha \).

\[ \pi(\beta | T) = \int_0^\infty \pi(\alpha, \beta | T) d\alpha \]

\[ = \int_0^\infty \frac{\lambda^{n-1} \Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\alpha - 1}} \cdot \frac{\Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\alpha - 1}} d\alpha \]

Then we get,

\[ \hat{\alpha} = E(\alpha | T) = \int_0^\infty \alpha \pi(\alpha | T) d\alpha \]

\[ \hat{\alpha} = \int_0^\infty \alpha \cdot \left[ \frac{\lambda^{n-1} \Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\alpha - 1}} \cdot \frac{\Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\alpha - 1}} d\alpha \right] d\alpha \]

\[ \hat{\beta} = E(\beta | T) = \int_0^\infty \beta \pi(\beta | T) d\beta \]

\[ \hat{\beta} = \int_0^\infty \beta \pi(\beta | T) d\beta \]

\[ \hat{\beta} = \int_0^\infty \beta \cdot \left[ \frac{\lambda^{n-1} \Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\alpha - 1}} \cdot \frac{\Gamma(\alpha-1)}{(\Gamma(\alpha))^n \left( \sum_{i=1}^{n} t_i \right)^{\alpha - 1}} d\alpha \right] d\beta \]

RESULTS AND DISCUSSION

Before we estimate the confidence intervals using the Score Method, we estimate the system data’s parameter values from the aircraft engine coolant with the maximum likelihood method first.

Table 1. System Life Data

| Machine | Lifetime (hour) | Machine | Lifetime (hour) |
|---------|----------------|---------|----------------|
| 1       | 23             | 16      | 20             |
| 2       | 261            | 17      | 5              |
| 3       | 87             | 18      | 12             |
| 4       | 7              | 19      | 120            |
| 5       | 120            | 20      | 11             |
| 6       | 14             | 21      | 14             |
| 7       | 62             | 22      | 71             |
| 8       | 47             | 23      | 11             |
| 9       | 3              | 24      | 14             |
| 10      | 95             | 25      | 11             |
| 11      | 225            | 26      | 16             |
| 12      | 71             | 27      | 90             |
| 13      | 246            | 28      | 1              |
| 14      | 21             | 29      | 16             |
| 15      | 42             | 30      | 51             |

By substituting the data in Table 1 to Equations (9) and (10), the estimated parameter values in the system data from the aircraft engine coolant with the maximum likelihood method are obtained as follows:
\[ \hat{\alpha} = 0.68525 \]
\[ \hat{\beta} = 86.92590 \]

By substituting the estimated parameters in equations (12) and (13), obtained at the estimation stage through a numerical approach with MATLAB R2013b software. The estimated parameter values in the Gamma distribution use the Bayes method as follows:

\[ \hat{\alpha} = 0.72828 \]
\[ \hat{\beta} = 86.90000 \]

The reliability functions for the Gamma distribution are as follows:

\[
R(t) = 1 - F(t)
= 1 - \frac{\Gamma(a \frac{t}{\beta})}{\Gamma(a)}
\] (14)

The shape of the plot in Figures 2 and 3 shows the aircraft cooling engine’s reliability to carry out the expected function for a long time with a smaller life-span.

The Hazard Rate function for the Gamma distribution is as follows:

\[
h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{\Gamma(a) \beta} a t^{a-1} e^{-\frac{t}{\beta}}}{1 - \frac{\Gamma(a \frac{t}{\beta})}{\Gamma(a)}}
\] (15)

Based on the Hazard plot in Figures 4 and 5, the failure rate decreases with the increasing use of aircraft cooling engines. It is known that the log-likelihood function of the Gamma distribution is as follows:

\[
\ln L(\alpha, \beta) = -n \ln \Gamma(\alpha) - n \ln \beta + (\alpha - 1) \sum_{i=1}^{n} \ln t_i - \frac{1}{\beta} \sum_{i=1}^{n} t_i
\]
Based on Equation (4), the statistical score function is as follows:
\[ W_u = U^T(\alpha_0, \hat{\beta}_0)I^{-1}(\alpha_0, \hat{\beta}_0)U(\alpha_0, \hat{\beta}_0) \]

where \( \hat{\beta}_0 \) is the maximum likelihood estimator for \( \beta \) under the null hypothesis \( H_0: \alpha = \alpha_0, U(\alpha_0, \hat{\beta}_0) \) is the vector of the Score function, and \( I(\alpha_0, \hat{\beta}_0) \) is a matrix from fisher information. We can quickly get the Score function under \( H_0 \) which is:
\[ U(\alpha_0, \hat{\beta}_0) = \left[ \sum_{i=1}^{n} \ln t_i + \frac{n}{2\alpha_0} - n \ln \bar{t} \right] \]

The inverse of fisher’s information is as follows:
\[ I^{-1}(\alpha_0, \hat{\beta}_0) = \left[ \begin{array}{c} 2\alpha_0^2 \\ \frac{2\bar{t}}{n} \\ \frac{2\bar{t}}{n} \\ \frac{\bar{t}^2(2\alpha_0+1)}{n\alpha_0^2} \end{array} \right] \]

So that the score statistics are obtained as follows:
\[ W_u = \frac{2\alpha_0^2}{n} \left( \sum_{i=1}^{n} \ln t_i + \frac{n}{2\alpha_0} - n \ln \bar{t} \right)^2 \]  \hspace{1cm} (18)

If \( Z \sim N(0,1) \) then \( Z^2 \sim \chi_1^2 \), then the \( Z_{\text{score}} \) value can be obtained as follows:
\[ Z_{\text{score}} = \left[ U^T(\alpha_0, \hat{\beta}_0)I^{-1}(\alpha_0, \hat{\beta}_0)U(\alpha_0, \hat{\beta}_0) \right]^{\frac{1}{2}} = \sqrt{\frac{2\alpha_0^2}{n} \left( \sum_{i=1}^{n} \ln t_i + \frac{n}{2\alpha_0} - n \ln \bar{t} \right)} \]  \hspace{1cm} (19)

According to Sangnawakij (2016), the variance of \( \hat{\alpha} \) is \( \frac{2\alpha_0^2}{n} \), to find the confidence interval \( \alpha \) can be obtained by substituting \( \hat{\alpha} \) into the variance as follows:
\[ Z_{\text{score}} = \sqrt{\frac{2\alpha_0^2}{n} \left( \sum_{i=1}^{n} \ln t_i + \frac{n}{2\alpha_0} - n \ln \bar{t} \right)} \]  \hspace{1cm} (20)

Based on the definition of the confidence interval estimation
\[ P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \gamma \]

The confidence interval for \( \alpha \) and \( \beta \) based on the Score method is:
\[ P \left( \frac{3}{n} \left( \frac{1}{\chi^2(2\alpha_0+1)} \right) \right) = (1 - \gamma)100% \]  \hspace{1cm} (21)

\[ P \left( \frac{3}{n} \left( \frac{1}{\chi^2(2\alpha_0+1)} \right) \right) \leq \beta \leq \frac{3}{n} \left( \frac{1}{\chi^2(2\alpha_0+1)} \right) \]

(1 - \gamma)100% \]  \hspace{1cm} (22)

where:
\[ \alpha = \frac{0.5}{\ln \bar{t} - \ln \bar{t}} \text{ and } \gamma = \sum_{i=1}^{n} \ln t_i - n \ln \bar{t} \]

By substituting the estimated parameter values that have been obtained to Equation (21) and Equation (22) with \( \gamma = 0.05 \) for the Maximum Likelihood method, the estimated value of the parameter confidence interval is obtained as follows:
\[ P(0.45438 \leq \alpha \leq 1.39308) = 95\% \]
and
\[ P(42.75885 \leq \beta \leq 131.09290) = 95\% \]

Whereas for the Bayes method, the estimated value of the confidence interval is obtained as follows:
\[ P(0.46406 \leq \alpha \leq 1.30698) = 95\% \]
and
\[ P(32.68960 \leq \beta \leq 141.22559) = 95\% \]

Figure 6. A plot of the reliability function for the lower and upper limits of the confidence interval using the Maximum Likelihood Method.
CONCLUSIONS AND SUGGESTIONS

The reliability test results on the data on the aircraft cooling engine's life span show that the system reliability is decreasing over time. Along with the increasing time in testing, the reliability is lower, as shown by the decreasing graph plot. Based on the Hazard rate graph (failure rate), it is found that the Gamma Distribution has a failure rate that shows a decreasing graph pattern, meaning that the more time a system increases, the failure rate will decrease. In estimating the confidence interval value for parameter $\alpha$, the Maximum Likelihood method is better than the Bayes method because it has a narrower confidence interval. In comparison, in estimating the confidence interval for the $\beta$ parameter, the Bayes method is better than the Maximum Likelihood method.

As a suggestion, this research can be continued to compare the confidence interval between the life span data of systems with a gamma distribution with the Maximum Likelihood and Generalized Least Squares method.

REFERENCE

Abid, S. H., & Al-Hassany, S. A. (2016). On the inverted gamma distribution. *International Journal of Systems Science and Applied Mathematics, 1*(3), 16–22.

Bain, L. J., & Engelhardt, M. (1992). *Introduction to probability and mathematical statistic* (2nd ed.).

Bertsimas, D., & Nohadani, O. (2019). Robust maximum likelihood estimation. *INFORMS Journal on Computing, 31*(3), NFORMS Journal on Computing.

Boland, P. J., El-Neweihi, E., & Proschan, F. (1994). Applications of the Hazard Rate Ordering in Reliability and Order Statistics Reviewed work ( s ): Source: Journal of Applied
J. K., Ngesa, O., & Orwa, G. (2019). On Generalized gamma distribution and its application to survival data. *International Journal of Statistics and Probability, 8*(5), 85–102. https://doi.org/10.5539/ijsp.v8n5p85

Khamkong, M. (2018). Parameter estimation method for the two parameter gamma distribution based on transformation. *International Journal of Applied Engineering Research, 13*(2), 1261–1264.

Lawless, J. F. (1982). *Statistical model and method for life time data.* John Wiley and Sons, Inc.

Sangnawakij, P., Böhning, D., Niwitpong, S. A., Adams, S., Stanton, M., & Holling, H. (2019). Meta-analysis without study-specific variance information: Heterogeneity case. *Statistical Methods in Medical Research, 28*(1), 196–210. https://doi.org/10.1177/0962280217718867

Sangnawakij, P., & Niwitpong, S. A. (2017). Confidence intervals for functions of coefficients of variation with bounded parameter spaces in two gamma distributions. *Songklanakarin Journal of Science and Technology, 39*(1), 27–39. https://doi.org/10.14456/sjst-psu.2017.4

Sharma, V. K., Singh, S. K., & Singh, U. (2017). Classical and bayesian methods of estimation for power lindley distribution with application to waiting time data. *Communications for Statistical Applications and Methods, 24*(3), 193–209. https://doi.org/10.5351/CSAM.2017.24.3.193