Cawangan Kedah
Kampus Sungai Petani

UNIVERSITI TEKNOLOGI MARA

2021 ICMS
INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

Likelihood and Bayesian Intervals in the Stress-Strength Model Using Records From the Pareto Distribution of the Second Kind

Ayman Baklizi

Department of Mathematics, Statistics and Physics

Qatar University

4 - 5 AUGUST 2021

UiTM di hatiku

CONTENTS

Introduction

The probability density function (pdf) and the cumulative distribution function (cdf) of the one parameter Pareto distribution

$$f(x) = \frac{\theta}{(1+x)^{\theta+1}}, x > 0, \theta > 0,$$

$$F(x) = 1 - \frac{1}{(1+x)^{\theta}}, x > 0, \theta > 0.$$

The hazard function is given by

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\theta}{1 + x}, x > 0, \theta > 0.$$

Note that the hazard function is decreasing in x.

Introduction

In some situations, only observations more extreme the current extreme value are recorded. Data of this type are called "Records".

Some examples are:

1 - The hottest day ever,

2 - The longest winning streak in professional basketball,

3 - the lowest stock market figure,

Chandler (1952) introduced and studied some properties of record values. Since then a considerable amount of the literature is devoted to the study of records.

Ahsanullah (2004) and Arnold et al. (1998) provided a detailed treatment for various aspects of records and statistical inference based on records.

Introduction

A component of strength Y subjected to a stress X. The stress and strength are random variables. The system fails as soon as the stress exceeds the strength. The stress strength reliability is the probability that the system functions without failure. This probability is

$$P(X < Y)$$

Let $X \sim \text{Pa}(\theta_1)$ and $Y \sim \text{Pa}(\theta_2)$ be independent Pareto random variables.

Let $R = \Pr(X < Y)$ be the stress strength reliability.

It can be shown that, for the Pareto distribution

$$R = \frac{\theta_1}{\theta_1 + \theta_2}$$

Objectives

Our interest is in interval estimation of R based on record data on both variables. Specifically we will

1- Obtain the Asymptotic likelihood intervals for R

2- Obtain the Bayes intervals for R

3- Obtain the Bootstrap intervals for R.

4- Conduct a simulation study to compare the intervals.

Literature Review

Several authors have considered the problem of constructing confidence intervals for the Stress – Strength Reliability.

Various types of intervals for various distributional assumptions were considered in the literature.

The book by Kotz and Pensky (2003) surveys most of the important developments in this field up to 2003. Later developments and extensions are spread in the literature in many fields like Statistics, Engineering and Medicine.

Methodology

Let $R = (r_0, \dots, r_n)$ **be a set of records from Pa**(θ_1)

Let $S = (s_0, \dots, s_m)$ **be a set of records from Pa**(θ_2)

The likelihood functions are given by (Arnold et al., 1998);

$$L_1(\theta_1|\boldsymbol{R}) = f(r_n)\prod_{i=0}^{n-1} f(r_i)/\big(1 - F(r_i)\big),$$

$$L_2(\theta_2|\boldsymbol{S}) = g(s_m)\prod_{i=0}^{m-1} g(s_i)/\big(1 - G(s_i)\big),$$

where f **and** F **are the pdf and cdf Pa**(θ_1) **of and** g **and** G **are the pdf and cdf of Pa**(θ_2)**.**

Methodology

Substituting for the Pareto distribution we obtain

$$L_1(\theta_1|\boldsymbol{R}) = \frac{\theta_1^n}{(1+r_n)^{\theta_1+1}} \coprod_{i=0}^{n-1}(1+r_i)^{-1},$$

$$L_2(\theta_2|\boldsymbol{S}) = \frac{\theta_2^m}{(1+s_m)^{\theta_2+1}} \coprod_{i=0}^{m-1}(1+s_i)^{-1}.$$

The log likelihood functions are given by;

$$l_1(\theta_1|\boldsymbol{R}) = n\ln(\theta_1) - (\theta_1+1)\ln(1+r_n) + \ln \coprod_{i=0}^{n-1}(1+r_i)^{-1},$$

$$l_2(\theta_2|\boldsymbol{S}) = m\ln(\theta_2) - (\theta_2+1)\ln(1+s_m) + \ln \coprod_{i=0}^{m-1}(1+s_i)^{-1}$$

Methodology

The MLEs are obtained by direct maximization of the corresponding log-likelihood functions. We obtained

$$\hat{\theta}_1 = \frac{n}{ln(1 + r_n)}$$

$$\hat{\theta}_2 = \frac{m}{ln(1 + s_m)}$$

Hence the MLE of R is given by

$$\hat{R} = \frac{\hat{\theta}_1}{\hat{\theta}_1 + \hat{\theta}_2}$$

Methodology

Another confidence interval for R can be obtained from the asymptotic normality of MLEs. Note that

$$\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right)\xrightarrow{D} N\left(0, v_1^2\right),$$

where v_1^2 is the asymptotic variance given by the inverse of the Fisher information.

$$v_1^2 = \left[-E\left(\frac{\partial^2 \ln L(\theta|r_0,\ldots r_n)}{\partial\theta_1^2}\right)\right]^{-1} = \frac{\theta_1^2}{n},$$

Similarly,

$$\sqrt{m}\left(\hat{\theta}_2 - \theta_2\right)\xrightarrow{D}\left(0, v_2^2\right),$$

Where,

$$v_2^2 = \left[-E\left(\frac{\partial^2 \ln L(\theta|s_0,\ldots s_m)}{\partial\theta_2^2}\right)\right]^{-1} = \frac{\theta_2^2}{m}$$

Methodology

Let $n \to \infty, m \to \infty$ **such that** $m/n \to p, 0 < p < 1$**, it follows that;**

$$\sqrt{n}\left(\hat{\theta}_2 - \theta_2\right) \xrightarrow{D} N(0, v_2^2/p).$$

Since $R = \frac{\theta_1}{\theta_1 + \theta_2} = h(\theta_1, \theta_2)$ **say, and** $\hat{R} = \frac{\hat{\theta}_1}{\hat{\theta}_1 + \hat{\theta}_2} = h\left(\hat{\theta}_1, \hat{\theta}_2\right)$

$$\sqrt{n}\left(\hat{R} - R\right) = \sqrt{n}\left[h\left(\hat{\theta}_1, \hat{\theta}_2\right) - h(\theta_1, \theta_2)\right] \xrightarrow{D} N(0, \eta^2)$$

where $\eta^2 = \left(\frac{\partial h(\theta_1,\theta_2)}{\partial \theta_1}\right)^2 v_1^2 + \left(\frac{\partial h(\theta_1,\theta_2)}{\partial \theta_2}\right)^2 v_2^2/p.$

A $(1-\alpha)$**% confidence interval for** R **based on this result is given by**

$$\left\{\hat{R} - z_{1-\frac{\alpha}{2}}\hat{\eta}, \hat{R} + z_{1-\alpha/2}\hat{\eta}\right\}$$

Methodology

The likelihood functions based on the two sets of record values suggests that the conjugate family of prior distributions for θ_1 and θ_2 are the Gamma family of distributions;

$$\psi_1(\theta_1) = \frac{\beta_1^{\delta_1}\theta_1^{\delta_1-1}e^{-\beta\theta_1}}{\Gamma(\delta_1)}, \quad \theta_1 > 0,$$

where β_1 and δ_1 are the parameters of the prior distribution of θ_1 and

$$\psi_2(\theta_2) = \frac{\beta_2^{\delta_2}\theta_2^{\delta_2-1}e^{-\beta_2\theta_2}}{\Gamma(\delta_2)}, \quad \theta_2 > 0,$$

where β_2 and δ_2 are parameters of the prior distribution of θ_2.

Methodology

It can be shown that the posterior distribution of θ_1 given $r_0, \ldots, r_n$ is

$$\pi_1(\theta_1 | r_0, \ldots, r_n) = \frac{\left(\beta_1 + \ln(1+r_n)\right)^{(n+\delta_1)}}{\Gamma(n+\delta_1)} {\theta_1}^{n+\delta_1-1} e^{-\theta_1\left(\beta_1+\ln(1+r_n)\right)}, \theta_1 > 0.$$

Similarly the posterior distribution of θ_2 is given by

$$\pi_1(\theta_2 | s_0, \ldots, s_m) = \frac{\left(\beta_2 + \ln(1+s_m)\right)^{(m+\delta_2)}}{\Gamma(m+\delta_2)} {\theta_2}^{m+\delta_2-1} e^{-\theta_2\left(\beta_2+\ln(1+s_m)\right)}, \theta_2 > 0,$$

The construction of highest posterior density (HPD) regions requires finding the following set;

$$\mathbf{BR}(\pi_\alpha) = \{\theta : \pi(\theta | r_0, \ldots, r_n, s_0, \ldots, s_n) \geq \pi_\alpha\}$$

where π_α is the largest constant such that $\Pr\left(\theta \in \mathbf{BR}(\pi_\alpha)\right) \geq 1 - \alpha$.

Methodology

Bootstrap intervals are computer intensive methods based on resampling with replacement from the original data.

When the parametric form of the distribution from which the data are generated is known except for some unknown parameters, we generate from this distribution after its parameters are replaced by their estimates.

There are several Bootstrap based intervals discussed in the literature (Efron and Tibshirani, 1993). Of the most common ones are the bootstrap-*t* interval and the percentile interval.

Conclusions

The performance of the two bootstrap-t intervals is similar in terms of coverage rates which tend to be anti-conservative with coverage rate less than the nominal. But the situation improves for larger samples. The percentile interval appears to be the best among bootstrap intervals, especially for small sample sizes.

The approximate Bayesian HPD interval appears to perform about as well as the percentile interval for moderate to larger sample sizes. It consistently has the best performance in the case of unequal sample sizes.

The performance of the interval based on the asymptotic normality of the MLE appears to be the good only for large samples. It is highly affected by unequal sample sizes.

The percentile interval appears to be shorter than the approximate HPD intervals for values of R near $1/2$ when the sample sizes are small. For larger sample sizes the percentile is shorter for all values of R under consideration.

In conclusion, we would recommend the use of the percentile interval for small sample sizes. For larger sample sizes the percentile and AHPD have about similar performance. When the sample sizes are unequal the AHPD is the interval of choice.

References

Ahsanullah, M. (2004). *Record values: Theory and applications*. University Press of America Inc., Lanham, Maryland, USA.

Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1998). *Records*. Wiley.

Baklizi, A. (2008b). Estimation of Pr(X<Y) Using Record Values in the One and Two Parameter Exponential Distributions. *Communications in Statistics, Theory and Methods, 37(5), 692 – 698.*

Chandler, K.N. (1952). 'The distribution and frequency of record values.' *Journal of the Royal Statistical Society* B14: 220 – 228

Chen, Ming-Hui and Shao, Qi-Man (1999). 'Monte Carlo estimation of Bayesian credible and HPD intervals.' *Journal of Computational and Graphical Statistics* 8(1): 69-92.

Efron, B., and Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Kotz, S., Lumelskii, Y. and Pensky, M. (2003). *The stress-strength model and its generalizations: Theory and applications*. World Scientific.

Lehmann, E.L. (1998). *Elements of large sample theory*. Springer

THANK YOU