Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.4, 2013

Concepts in Order Statistics and Bayesian Estimation

AL MUTAIRI ANED O.[1,2] , HENG CHIN LOW[1]
[1]School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

[2]Statistics Department, Faculty of Science, Prince Norah bint Abdulrahman University, Riyadh, Saudi Arabia.

* E-mail of the corresponding author: anedomar@hotmail.com

Abstract In this paper, some basic concepts related to order statistics and Bayesian estimation is discussed which includes tests of life, order statistics and concept of reconstruction that has various fields in the application such as reconstruction for industrial units or organisms. This paper also describes significant functions in the reconstruction such as probability of life function or failure probability function, surviving probability function, life probability density function and failure rate function. Furthermore, it illustrated in depth idea concerning Bayesian statistics for the study and evaluation of different Bayesian methods.

Keywords Bayesian statistics, failure probability function, order statistics

1. Introduction

1.1 Concepts in Order Statistics and Bayesian Estimation

1.1.1 Life Tests

In tests of life, it is usually 'n' of matched and independent units under test in particular experience and recorded times of failure for all these units respectively. In such case, the experiment is the complete sample and there is always continuation in the experiment until the failure of all units not be practical especially when the experimental sample is large and units are expensive [1]. Therefore, it is appropriate to stop the experiment after attaining partial information, which distinguishes the field of survival function for other fields in statistics, which is the amputation. It is observed that its use provides accurate estimations not less than the complete sample information of the random variable [2].

1.1.2 Order Statistics

It is observed that order statistics has significant role in statistical inference particularly in the laboratories methods. For instance, let Y= $(Y_1, Y_2,\ldots..Y_n)$ is the random sample selected from the connected population where the possibility of its density function f(y) and its cumulative distribution function is F(y) and by arranging in ascending order $Y_{(1)} < Y_{(2)} <\ldots\ldots..< Y_{(i)} < \ldots\ldots.. < Y_{(n)}$

Where $Y_{(1)}$ is the smallest readings and called the first ordinal statistic, $Y_{(n)}$ is the largest readings and called ordinal statistic n and $Y_{(i)}$ is the ordinal *i* reading and called ordinal statistics *i*. The $Y_{(1)}$,$Y_{(2)}$,……………, $Y_{(n)}$ is the ordinal statistics and provides possibility density function to the ordinal statistics as described Y_i $g_{Y(i)}$

$$g_{Y_{(i)}} = \frac{n!}{(i-1)!(n-i)!}\,[F(y)]^{i-1}\,[1-F(y)]^{n-i} f(y) \quad \ldots\ldots (1)$$

When i=1 the probability density function for the first reading is

$$g_{Y_{(i)}} = n[1-F(y)]^{n-1} f(y) \quad \ldots\ldots (2)$$

When i=n the probability density function for the biggest reading is

$$g_{Y_{(n)}} = n[1 - F(y)]^{n-1} f(y) \ldots\ldots (3)$$

1.1.3 Aging

The concept of Aging in statistics has many applications in different fields including engineering, physical and biological fields. It is observed that reconstruction has three major types including positive aging, fixed aging and negative aging where positive aging is the unit age decrease with the progress of time, may lead over time to the erosion of industrial units, which needs plans for maintenance however, time passing caused agedness, which requires development of plans for remedy. On the other hand, fixed aging that is not influenced by time are those electronic industrial units following exponential model of failure rate fixed in time. Moreover, negative aging with time leads to the enhancement of the industrial unit after the initiation of operation of new units so the proper units remain while manufacturing unit may fail at the beginning of operating immediately from the industrial side [3].

2. Definitions

2.1.1 Survival Function

If random variable is positive Y, then the function of survival after age S(y) described y as follows

$S(y) = \bar{F}(y) = 1 - F(y) = P(Y>y)$ which is the survival function or reliability function or validity function.

The concept of survival probability function analysis after S(y) age of time y in the early research in the science of insurance, life schedules and morality where the contemporary analysis of survival function initiated since an S(y) half century ago in the engineering applications. This paper observed that initially emphasized on the parametric approach for the random variables following the known distribution including normal, exponential, wabil or gamma distribution then researchers emphasize on the assessment and testing of hypotheses for the parameters of these distributions. However, with the growing interest in this field also conducted on non parametric approach which led to the estimation and testing of distribution F(y) hypothesis and utilization of non parametric approach of positive random variable for the life [4].

2.1.2 Life Probability Function

If Y is the positive random variable and F(y) is the life probability function illustrated as

$F(y) = P(Y \leq y)$

2.1.3 Life Probability Density Function

If 'Y' is the positive random variable, then life probability density function 'Y' as f(y) described as follows

$f(y) = \frac{d}{dy}[F(y) = P(Y \leq y)]$

2.1.4 Failure Rate Function

If Y is the positive random variable, h(y) failure rate function (y, y+δy) during the period defined as

$$h(y) = \lim_{\delta \to 0} P(y < Y < y + \delta y \,|Y > y) = \frac{f(y)}{S(y)}; y > 0$$

Where, the failure rate function h(y) takes three significant cases in the application as follows:

$$h(y) = kc\lambda^{c-1} + (1-k) b y^{b-1}\theta \exp(y^k\theta), \theta, b, c, \lambda > 0, 0 \leq k \leq 1, y > 0 \text{------------------------------}(4)$$

The above relation takes a U curve with the axis y of time as demonstrated in the figure below where the curve of failure rate function is described in h(y) the relation (4)

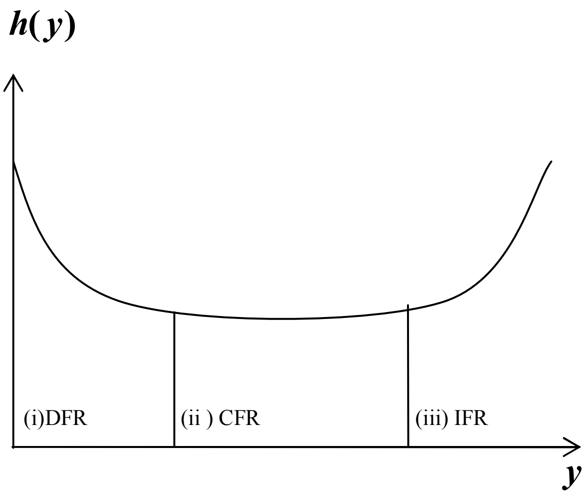

Figure 1: U Curve

In the above figure, part (i) demonstrate figure curve U that failure rate function is decreasing with h(Y) with the time y and symbolized by decreasing failure rate (DFR). However, part (ii) describes that failure rate function is h(y) constant function over time y which is equal to fixed amount C which is not affected by the time as represented as constant failure rate (CFR). The part (iii) of the figure 1 demonstrates that the failure rate function is enhancing h(y) over time 'y' which is represented as increasing failure rate as value of failure rate function is small at the beginning with 'y' and then increases with the time and explained reasons of reconstructions , antiquity and corrosion of industrial machines and biological aspects.

3. Concept of Bayesian Statistic

The Bayesian statistic are significantly utilize for the perspective of decision science through the study of loss associated with the sequence of possible decisions such as patterns of decision of the buyer or investor or institution. The Bayesian statistical inference and mathematical structure are formulated as where group A represents action space and includes all possible procedures so that a ϵ A. While, group H represents the parameters space and includes all levels of possible parameters θ where $\theta \epsilon$ H, which has only one real, level and is unknown to the decision maker who wants to acquire effective estimate for the level. Moreover, it is observed that domain of loss function is the Cartesian product H x A which is all ordered pair for these sequence (θ, a) where a ϵ A and $\theta \epsilon$ A and $\theta \epsilon$ H and its co-domain is the set of real numbers R. Furthermore, group D represents the decisions space and including all possible decisions d, so that d ϵ D and each predictable decision function d has the domain R_Y that is the subset of the set of real numbers R [5].

4. Definitions Related To Bayesian Statistics

4.1.1 Loss Function

The loss function *l(θ,a)* is described as

Upper estimation of procedure

$$l(\theta, a) = \begin{cases} c_1(a-\theta) & ,if \quad a \geq \theta \quad a \\ c_2(\theta - a) & ,if \quad a \leq \theta \quad a \end{cases}$$ Lower predictors of procedure

Where, c_1 and c_2 are consonants and a is the predictable procedure and θ is the parameter, it is observed that loss function depends on the nature of relative under study problem.

If c_1=c_2=1

$l(\theta, a) = |\theta - a|$

In that case, l is defined as absolute error loss function [5]

4.1.2 Squared Error Loss Function

The squared error loss function can be described as

$l(\theta, a) = (\theta - a)^2$ Where a is the any possible procedure and θ is the parameter [6]

4.1.3 LINEX Loss Function

The LINEX linear exponential function is described as follows:

l(Δ) = [$e^{a\Delta}$-aΔ-1]

Where $\Delta = \frac{\hat{\theta}}{\theta}$ - 1, $\hat{\theta}$ is estimator of θ and a is the constant [7]

4.1.4 Risk Function

Risk function R (θ, d) is described as the expected error loss by utilizing the decision d and parameterθ [7]. While the effective way for the development of measuring the quality of decision function comes from searching to enhance the average of error loss function l(θ, d(y)) utilizing risk function R(θ,d) where its domain is H x D and its comparable domain is the real line R [8].

For instance in the situation of continued random variable Y

$R(\theta, d) = \int_{R_Y} l(\theta, d(y)) f(y, \theta) dy$

However in the situation of discrete random variable

$R(\theta, d) = \sum_{y \in R_Y} l(\theta, d(y)) f(y, \theta)$

Where f(y,θ) is the probability distribution density function RY is the random variable space Y. While the risk function $R(\theta, d)$ represents the expected error loss for the probability distribution function f(y, θ) for the decision maker when selecting real level θ and decision d where the risk function can be written in the provided form as R (θ, d) = $E_\theta(l(\theta, d(y))$ [5].

4.1.5 Prior Distribution Function

The prior distribution function $\Pi(\theta)$ is utilized to illustrate the complete information concerning parameters θ before taking the sample $\underline{Y} = (Y_1, Y_2, Y_3, \dots \dots . Y_n)$ from population has the distribution f(y,θ) where prior distribution function $\Pi(\theta)$ is utilized to illustrate the whole information concerning parameter θ before taking the sample Y= (Y_1, Y_2,…..Y_n) from the population has the distribution f(y,θ). The prior distribution function $\Pi(\theta)$ is classified as non-informative prior distribution function and information prior distribution function [7].

4.1.5.1 Non-Informative Prior Distribution Function

Non-distribution function is used in the cases where prior information concerning parameters θ concerning the parameter θ which is comparatively limited and non-informative prior distribution for such cases. If θ is used within limited period m < θ < n on the real axis $(-\infty, \infty)$ where the prior distribution function is illustrated as follows

$$\Pi(\theta) = \frac{1}{n-m}, \qquad m<\theta<n$$

Moreover, if - ∞<θ<∞ where the prior distribution function is illustrated as

$$\Pi(\theta) \propto \text{constant} \; -\infty<\theta<\infty$$

However, if θ is positive that is 0<θ<∞, it is required to take the transformation lnθ= y to acquire regular distribution and $\theta > 0$

$$\Pi(\theta) \propto \frac{1}{\theta^3} \qquad , \theta > 0$$

Hartigan (1964) supposed that prior density function $\Pi(\theta) \propto \frac{1}{\theta^3}$ for distributions with probability density function f (x|θ) which describe the condition [9]

$$\frac{d}{d\theta}\ln \Pi(\theta) = -\frac{E(l_1 l_2}{E(l_2)}$$

Where

$$l_{i=}\frac{d^i}{d\theta^i}[\ln f(x|\theta)]$$

$$E(l_2) \neq 0$$

$$E(l_1^2) + E(l_2) = 0$$

5. Conclusion

It is concluded that order statistics are significant branches of the statistical interference; hence, the ordinal statistics are the effective for finding non-Bayesian and Bayesian predictions where one of the significant application of ordinal statistics is obtain through limits of prediction for the statistics based on the known sample for same distribution.

REFERENCES

[1] A.M. Abouammoh, R. Ahmad, and A. Khalique, "On new renewal better than used classes of life distributions," *Statistics & Probability Letters*, vol. 48, pp. 189 -194, 2000.

[2] R.E. Barlow and F. Proschan, "Statistical Theory of Reliability and Life Testing. ," *To Begin With, Silver*

Spring, vol. 32, no. 1, pp. 3-45, 1981.

[3] M.C. Bryson and M.M. Siddiqui, "Some criteria of ageing. ," *J. Amer. Statist. Assoc.* , vol. 64, no. 1, pp. 1472-1483., 1969.

[4] A.P. Basu and N. Ebrahimi, "Bayesian Approach to Life Testing and Reliability Estimation Using Asymmetric Loss Function ," *Jour.Stat.Plann.Infer*, vol. 29, pp. 21-31, 1991.

[5] Sanku Dey, "COMPARISON OF RELATIVE RISK FUNCTIONS OF THE RAYLEIGH DISTRIBUTION UNDER TYPE-II CENSORED SAMPLES:BAYESIAN APPROACH," *Jordan Journal of Mathematics and Statistics (JJMS)*, vol. 4, no. 1, pp. 61-78, 2011.

[6] A. Zellner, "Bayesian Estimation and Prediction using Asymmetric Loss Functions.," *Journal of American Statistical Association*, vol. 81, pp. 446-451 , 1986.

[7] Hal R Varian, "A Bayesian Approach to Real Estate Assessment in Fienberg," *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, vol. 3, no. 1, pp. 3-23, 1975.

[8] H.A. Howlader and A. Hossain, "On Bayesian Estimation and Prediction from Rayleigh based on Type-II Censored Data," *Communications in Statistics.-Theory and Methods,* vol. 24, no. 9, pp. 2249-2259, 1995.

[9] Pencina MJ, D'Agostino RB Sr, and Song L., "Quantifying discrimination of Framingham risk functions with different survival C statistics.," *PubMed*, vol. 31, no. 15, pp. 1543-53, 2012.

[10] J. A. Hartigan, "Invariant Prior Distribution," *Annals of.Mathematical Statistics*, vol. 34, no. 1, pp. 4-23, 1964.