An Adequate Statistic for the Exponentially Distributed Censoring Data

P. S. Nair and S-C Cheng Department of Mathematics and Computer Science
Creighton University Omaha, Nebraska 68178-2090

Abstract - In the problem of estimating the parameter of the underlying probability distribution, a su¢cient statistic should be one that summarizes and exhausts in itself all the relevant information on the parameter that is contained in the sample. A similar basic problem of statistics is that of predicting a future (that is, not yet observed) random variable on the basis of some existing observable random variables when the parameter of the underlying probability distribution does not concern us directly. Likewise, we are looking for a statistic as one that is exhaustive of all the relevant information on the future random variable that is available in the current observable random variables. The notion of adequate statistics initiated by Fisher [4] and Skibinsky [8] would deal with this concern. Subsequently, it has been extensively investigated in literatures such as [1,2,6,9-11]. In this article, we make a comprehensive study of ...nding an adequate statistic for the total time on test when the data is assumed to be exponentially distributed and censored after the r-th failure.

Key words - Su¢ciency, Factorization criterion, Conditional independence, Adequacy, Order statistics, Censoring data, Total time on test.

I. Introduction

Let $P = fP_\mu : \mu\ 2\ \pounds g$ be the statistical model for $(X; Y)$, that is, a family of probability distributions (measures). In practice, this X could denote a random sample X_1, X_2, ¢¢¢, X_n of size n from a population. A statistic $T = T(X)$ is a measurable map of the observable X. A basic problem of statistics is that of predicting a future (that is, not yet observed) random variable Y on the basis of an observable random variable X, where μ is the parameter that does not concern us directly. The notion of adequate statistics initiated by Fisher [4] and Skibinsky [8] would deal with this concern. Subsequently, it has been extensively investigated in literature such as $[1; 2; 6; 9\ ¡\ 11]$. In this article, we make a comprehensive study of ...nding an adequate statistic for the total time on test when the data is assumed to be exponentially distributed and censored after the r-th failure.

In the sense of Fisher [4], a statistic $T = T(X)$ is said to be a su¢cient statistic for the parameter μ if the conditional probability distribution of X given $T = T(X)$ does not depend on μ. Let P_X denote the family of marginal distribution for the observable X. We write $T\ suff\ (X; P_X)$ (to be read as "T is a su¢cient statistic with respect to P_X") which denotes the fact that $T = T(X)$ is a su¢cient statistic for μ. Two random variables X and Y are said to be independent if the joint probability distribution of X and Y is equal to the product of the marginal probability distributions of X and Y, respectively. We write $X\ ?\ Y j\mu$ to indicate that X and Y are independent with respect to the probability distribution P_μ. Similarly, two random variables X and Y are said to be conditionally independent given a third random variable T if the joint conditional probability distribution of X and Y give T is equal to the product of the marginal conditional probability distributions of X and Y given T, respectively. We also write X ?Y jT; µ to indicate that X and Y are conditionally independent given T with respect to the probability distribution $P_µ$.

II. Sufficient Statistics and Conditional Independence

Let P = ff (x; y; µ) : µ 2 £g be the family of joint probability functions for (X; Y) and P_X = ff (x; µ) : µ 2 £g be the family of marginal joint probability functions for X. Also, let T = T (X) be a measurable function of X. In practice, we let X = (X_1; X_2; ¢¢¢; X_n) denote a random sample of size n and let x = (x_1; x_2; ¢¢¢; x_n) be an observed value of X. The joint probability function of X = (X_1; X_2; ¢¢¢; X_n) is then given by

$$f(x_1; x_2; ¢¢¢; x_n; µ) = \prod_{i=1} f(x_i; µ)$$

The likelihood, L (x_1; x_2; ¢¢¢; x_n; µ), of the sample is de…ned to be the joint probability function evaluate at x_1, x_2, ¢¢¢, x_n, that is,

$$L(x_1; x_2; ¢¢¢; x_n; µ) = \prod_{i=1} f(x_i; µ)$$

The following theorem is the famous factorization criterion for the su¢cient statistics..

Theorem 1 [5] The statistic T is a su¢cient statistic for µ if and only if the likelihood of the sample, L, can be factored into two nonnegative measurable functions,

$$L(x_1; x_2; ¢¢¢; x_n; µ) = g(t; µ)\, h(x_1; x_2; ¢¢¢; x_n);$$

where g (t; µ) is a function only of t and µ and h (x_1; x_2; ¢¢¢; x_n) is not a function involving µ.

We now also present a short list of few theorems concerning the notion of conditional independence. They can be found in the literatures such as [3; 7].

Theorem 2 If X ?Y jT; µ and W = W (X), then .W ?Y jT; µ.

Theorem 3 The following assertions are equivalent.

(1) X ?Y jT; µ.
(2) X ? (Y; T) jT; µ.
(3) (X; T) ? (Y; T) jT; µ.

Theorem 4 If X ?Y jµ and W = W (X), then .X ?Y jW; µ.

Theorem 5 X ?Y jT; µ if and only if , for any W = W (X), E (W jX; Y) = E (W jX).

Theorem 6 Let T = T (X) and W = W (T). If T ?Y jW; µ.and X ?Y jT; µ, then X ?Y jW; µ.

Theorem 7 If X ?Y jW; µ, T = T (X) and W = W (T), then X ?Y jT; µ.

III. Adequate Statistics

In the colorful language of R. A. Fisher [4] we may loosely de…ne the notion of adequacy in the following terms: The statistic T = T (X) is said to be adequate or prediction su¢cient if T summarizes in itself all the relevant information about Y that is contained in X:

In 1967, Skibinsky [8] proposed the following mathematical framework of adequate statistics.

De…nition 8 A statistic T is said to be adequate or prediction su¢cient for X with respect to (Y; P) if

(1) T suff (X; P_X), and
(2) X ?Y jT; µ for all µ 2 £.

Following Skibinsky [8], we write T adq (X; Y; P) [to be read as "T is adequate for X with respect to (Y; P)"] as a mathematical shorthand for the proposition "T adequately summarizes (and exhausts) all information in X about Y with respect to the model P". This is in conformity with our use of the notation $T \text{ suff } (X; P_X)$ when we restrict our attention to the model P_X.

The intuitive content of the above definition appears to be embedded in the following two propositions:

(1) Given $T = T(X)$, no further details about observable X can yields any additional information about θ. In other words, for making an inference on θ, the statistician needs to record only the T-value on X.

(2) If the inferred value of θ is the one that actually obtains, then the conditional distribution of Y given X depends on X only through T. In other words, if the statistician proposes to predict Y by first predicting (or estimating) θ and adequately summarizes (and exhausts) all information in X about Y with respect to the model P".

IV. Main Theorem

We are now ready to present our main result.

Theorem 9 Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ denote the order statistics of a random sample X_1, X_2, $\cdots$, X_n of size n from a population with probability density function from an exponential family:

$$\left\{ f(x;\theta) = B(\theta) H(x) e^{Q(\theta)R(x)} : \theta \in \Theta \right\}:$$

Let $1 \le r \le m \le n$. Consider the problem of predicting a future $X_{(m)}$ after observing $X_{(1)}$, $X_{(2)}$, $\cdots$, $X_{(r)}$. Let P denote the family of joint probability distributions of the observed $X = \left(X_{(1)}; X_{(2)}; \cdots; X_{(r)}\right)$ and not yet observed $X_{(m)}$. Then $T \text{ adq } \left(X; X_{(m)}; P\right)$.

Proof. The likelihood of the observed values of the first r order statistics is given by

$$L\left(x_{(1)}; x_{(2)}; \cdots; x_{(r)}; \theta\right)$$
$$= \prod_{i=1} f\left(x_{(i)};\theta\right)$$
$$= [B(\theta)]^r \left[\prod_{i=1} H\left(x_{(i)}\right)\right] e^{Q(\theta)\sum_{i=1} R\left(x_{(i)}\right)}$$

Define

$$T = \left(\sum_{i=1} R\left(X_{(i)}\right); X_{(r)}\right):$$

Consider

$$g(t;\theta) = [B(\theta)]^r e^{Q(\theta)\sum_{i=1} R\left(x_{(i)}\right)}$$

and

$$h\left(x_{(1)}; x_{(2)}; \cdots; x_{(r)}\right) = \prod_{i=1} H\left(x_{(i)}\right):$$

From the classical factorization criterion (Theorem 1), it follows that $T = \left(\sum_{i=1} R\left(X_{(i)}\right); X_{(r)}\right)$ is a sufficient statistic for θ.

It is well known that the order statistics for a Markov chain. Then it follows that $X \perp X(m) \mid X_{(r)}; \theta$ for all $\theta \in \Theta$. Since $T = \left(\sum_{i=1} R\left(X_{(i)}\right); X_{(r)}\right)$ is a function of $X_{(r)}$, it follows from Theorem 7 that $X \perp X(m) \mid T; \theta$ for all $\theta \in \Theta$. Therefore, $T \text{ adq } \left(X; X_{(m)}; P\right)$. ■

V. An Example

Consider an exponential distribution with mean θ. In this case, $B(\theta) = 1/\theta$, $H(y) = 1$, $Q(\theta) = 1/\theta$, and $R(x) = x$. Now consider a censoring situation in which n units are put on test and observations continues until r units have failed. Suppose that the numbers of millions of revolutions of thirty ball bearings are censored after the twenty-third failure. The ordered data to failure are

17.88	28.92	33.00	41.52	42.12	45.60
48.40	51.84	51.96	54.12	55.56	67.80
68.64	68.64	68.88	84.12	93.12	98.64
105.12	105.84	127.92	128.04	173.40	

We may want to estimate the total time on test $\sum_{i=1}^{30} X_{(i)}$. Assume that the number of millions of revolutions of the ball bearing has an exponential distribution with mean μ. Calculate $\sum_{i=1}^{r} X_{(i)} = 1{,}661.06$. According to Theorem 9, $T = \left(\sum_{i=1}^{23} X_{(i)}, X_{(23)} \right)$ is an adequate statistic for predicting $X_{(m)}$ (for $m = 24, 25, \cdots, 30$) based upon the information of the observed exponentially distributed data $\left(X_{(1)}, X_{(2)}, \cdots, X_{(23)} \right)$. Consequently, an adequate estimator of the total time on test $\sum_{i=1}^{30} X_{(i)}$ will be given by

$$
\begin{aligned}
& \sum_{i=1}^{r} X_{(i)} + (n - r) X_{(r)} \\
= \; & 1{,}661.06 + (7)\,(173.40) \\
= \; & 2{,}874.86.
\end{aligned}
$$

References

[1] Basu, D. and Cheng, S-C (1981). A note on sufficiency in coherent models. Internat. J. Math. & Math. Sci., Vol. 4, No. 3, 571-581.

[2] Cheng, S-C and Mordeson, J. N. (1985). A note on the notion of adequacy. Chinese J. of Math., Vol. 13, No. 1, 13-41.

[3] Chung, K. L. (1968). A Course in Probability Theory (Second Edition). Academic Press, New York.

[4] Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. Phil. Trans. Roy. Soc. London A 222, 309-368.

[5] Halmos, P. R. and Savage, L. J. (1949). Applications of the Radon-Nikodym Theorem to the theory of sufficient statistics. Ann. Statist. 3, 1371-1378.

[6] Lauritzen, S. L.. (1972). Sufficiency and time series analysis. Intst. Math. Statist. , Univ. of Copenhagen 11, 249-269.

[7] Loéve, M. (1968). Probability Theory (Third Edition). Van Nostrand, New York.

[8] Skibinsky, M. (1967). Adequate subfields and sufficiency. Ann. Math. Statist. 38, 155-161.

[9] Sugiura, M. and Moriomoto, H. (1969). Factorization theorem for adequate σ-fields (in Japanese). Sūgaku, 21, 286-289.

[10] Takeuchi, K. and Akahira, M. (1975). Characterizations of prediction sufficiency (adequacy) in terms of risk functions. Ann. Statist. 3, 1018-1024.

[11] Torgenser, E. N. (1977). Prediction sufficiency when the loss function does not depend on the unknown parameter. (adequacy) in terms of risk functions. Ann. Statist. 5, 155-163.