

ON ESTIMATING THE RELIABILITY OF A COMPONENT
SUBJECT TO SEVERAL DIFFERENT STRESSES

by

Satish Chandra

Technical Report No. 109
Department of Statistics ONR Contract

November 22, 1971

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

This document has been approved for public release
and sale; its distribution is unlimited.

DEPARTMENT OF STATISTICS
Southern Methodist University

In the past few years, many authors (e.g., Z. W. Birnbaum, McCarty, Church, Harris, Vandantzig, Govindarajulu, Mann, Whitney, Mazumdar, Owen, Sen, etc.) have attempted to estimate and give the confidence bounds for the reliability of a component using the probability arguments of a certain physical model of failure. According to this model a component fails if at any moment the applied stress (or load), say Y , is greater than the resistance, say X . That is, the problem here is to find an estimate of the probability that X is less than Y , where X and Y are both random variables, having some known or unknown probability structure, and to find the confidence limits on this estimate of the probability. An extensive amount of work has been done on this problem by the above mentioned authors and many others, and some, for example, Mann and Whitney [9], Mazumdar [10], Church and Harris [3] etc. have given practical uses for the results.

The question now arises of what to do if at any moment the applied stress (or load or force) may not be measured in terms of a single random variable, but can be measured in terms of several random variables, say $X_1, X_2, \dots, X_N$, and the resistance is still a single random variable, say Y . These random variables $X_1, X_2, \dots, X_N$, Y may have a known (specified) probability structure to a certain extent or the structure may be completely

unspecified except that in this latter case independence of the X_i 's is

assumed. In some cases, it might be possible to arrive at a suitable

mathematical model which allows one to compute deterministically the

"over-all stress" at different points of time corresponding to a given

set of "initial stresses," and initial conditions, but it may not be pos-

sible in all the cases, especially when all the initial stresses are

random variables. For example, the over-all stress, say X , may be a

suitable linear combination of all the initial stresses $X_1, \dots, X_N$, with

$X_1, \dots, X_N$ being random variables. Thus, writing $X = \sum_{i=1}^N C_i X_i$, C_i 's

being real constants, the problem is to estimate C_i 's, the parameters

involved in order to estimate the over-all stress X . A robust way to

estimate these constants is still under consideration. Some authors have

attempted to estimate these constants using the mixture of probability

distributions techniques in some simple cases, for example when $N = 2$

and X_1, X_2 are independent normal random variables. The problem is even

more complicated if the over-all stress is a nonlinear combination of the

initial stresses.

In many cases, a component does not fail if all the different

stresses $X_1, \dots, X_N$ are simultaneously less than the resistance Y . Then,

a measure of reliability of a component subject to several different stresses

during a given period $[0, T]$ is taken to be the probability that all the

different stresses are simultaneously less than the resistance during the

entire interval, i.e.

$$(1.1) \quad p = \text{Prob} \{X_1 < Y, X_2 < Y, \dots, X_N < Y\}.$$

Church and Harris [3] and Mazumdar [10] considered the example of

missile flights, where the several different stresses are the propulsive force, angles of elevation, changes in atmospheric condition, and so on. They assumed that all these different stresses have known distributions, and a mathematical model is available to compute the over-all stress, and also the distribution of over-all stress is known. However, these assumptions might not be true in a given physical situation.

Another similar problem of interest is that where during a given time period $[0, T]$ the stress or force could be measured in terms of one random variable, say X whereas a component might consist of several different objects subject to different resistances. For example, suppose the electric current is supplied to a component consisting of several different transistors each of which has a different capacity to resist the current. Thus, the component breaks if the current supplied to the component exceeds the capacity of any of the transistors. Hence, denoting the stress or force of the component by X , and the several different resistances by $Y_1, Y_2, \dots, Y_N$, the component does not fail if the applied force or stress X is simultaneously less than the several different resistances.

A measure of reliability of a component subject to single force and various resistances during a given period $[0, T]$ can be taken to be the probability that the applied force is simultaneously less than the several different resistances during the entire interval, i.e.

$$(1.2) \quad q = \text{Prob} \{ X < Y_1, \dots, X < Y_N \}.$$

where $X, Y_1, \dots, Y_N$ are all random variables.

The principal objective of the material presented here is to estimate p and q defined above.

Chapter II consists of parametric estimation of p , assuming that $X_1, \dots, X_N, Y$ have certain probability structures.

Chapter III consists of a non-parametric approach to estimate p for a much more general situation, i.e., $X_1, \dots, X_N, Y$ are all independent random variables having different absolutely continuous distribution functions. A statistic is proposed and shown to estimate p unbiasedly. The variance of this statistic is computed and used to show that this estimate of p is consistent.

Chapter IV is devoted to the estimation of q . Analogous results to Chapter II and III are given and some more results of practical importance are included.

Finally, Chapter V consists of a summary and suggestions for further research.

CHAPTER II

PARAMETRIC ESTIMATION

2.1 Introduction

Owen, Craswell and Hanson [12], Govindarajula [5], Church and Harris [3], etc. have given maximum likelihood estimates and confidence intervals for Prob $\{X < Y\}$ under the assumptions that X and Y have a bivariate normal distribution with some parameters known, and for the case when X and Y are independent, and when they are paired. Mazumdar [10] has given the minimum variance unbiased estimate for Prob $\{X < Y\}$ when X is assumed to be normally distributed with known mean and known variance; also Y is normally distributed with unknown mean, known or unknown variance and X, Y being independent. He also derived interval estimates for these cases. In this chapter, we shall consider the estimation of the probability that N random variables $X_1, \dots, X_N$ are simultaneously less than a random variable Y , where these random variables are mutually independent.

2.2 Useful Lemmas

Let $X_1, \dots, X_N, Y$ be $(N + 1)$ mutually independent random variables with

continuous cumulative distribution functions (c.d.f.'s) $F_1, F_2, \dots, F_N, G$ respectively and let

$$(2.1) \quad T = \max(X_1, \dots, X_N)$$

and have a c.d.f. H . Also, let

$$p = \text{Prob}\{X_1 > Y, \dots, X_N > Y\}.$$

Then, the problem of estimating p is simplified using the following lemmas.

Lemma 2.1

$$P(X_1 > Y, \dots, X_N > Y) = P(T > Y).$$

Lemma 2.2

$$H(t) = P(T \leq t) = \prod_{i=1}^N F_i(t).$$

Proof

$$\begin{aligned} H(t) &= P(T \leq t) = P(\max(X_1, \dots, X_N) \leq t) \\ &= P(X_1 \leq t, \dots, X_N \leq t) \\ &= P(X_1 \leq t) \dots P(X_N \leq t) \\ &= \prod_{i=1}^N F_i(t). \end{aligned}$$

Since $H(t)$ is a function of continuous c.d.f.'s $H(t)$ itself is continuous and so the probability of getting ties from the c.d.f. H is zero. Moreover, the random variable T is independent of the random variable Y .

Now, suppose $F_1 = F_2 = \dots = F_N = F$, i.e., the random variables $X_1, \dots, X_N$

have the same continuous c.d.f. F . Also, F has the density f and Y has the continuous c.d.f. G with the density g , and all these random variables are mutually independent. Then, the density function of the random vari-

able $T = \max(X_1, \dots, X_N)$ is given as

$$(2.2) \quad h(t) = N(F(t))^{N-1} f(t)$$

and thus,

$$(2.3) \quad p = \int_{-\infty}^{\infty} [F(y)]^N dg(y).$$

Proof

$$p = P(T > Y)$$

$$= F_Y P(T > Y | Y)$$

$$= \int_{-\infty}^{\infty} [F(Y)]^N dG(Y) \cdot$$

Lemma 2.4

$$\text{If } F_1 = F_2 = F_N = G, \text{ then}$$

$$p = \frac{1}{N+1} \cdot$$

Proof

In this case, from Equation (2.3)

$$p = \int_{-\infty}^{\infty} F_N^N(Y) dF(Y)$$

$$= \int_1^0 u_N du = \frac{1}{N+1} \cdot$$

2.3 Estimation of p when $X_1, X_2, \dots, X_N, Y$ are all independent normal

Let ϕ and Φ denote the cumulative distribution function and the density of a standard normal random variable respectively, i.e.,

$$(2.4) \quad \phi(y) = \frac{1}{\sqrt{2\pi}} e^{-1/2 y^2} \quad -\infty < y < \infty$$

and

$$(2.5) \quad \Phi(x) = \int_x^{\infty} \phi(y) dy.$$

Further, let the different stresses $X_1, \dots, X_N$ be all independent and

identically distributed normal random variable with mean μ , variance σ^2 . Also, let Y be normal with mean ν , variance τ^2 and independent of all the X_i 's. Then by Lemma 2.3,

$$p = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2} z(x-\mu)^2} \frac{1}{\tau} e^{-\frac{1}{2\tau^2} z(y-\nu)^2} \frac{1}{(2\pi)^{1/2}} dz dy$$

$$= \int_{-\infty}^{\infty} \Phi_N \left(\frac{\nu + \tau z - \mu}{\sigma} \right) \phi(z) dz$$

$$= \int_{-\infty}^{\infty} \Phi_N \left(\frac{\frac{1}{\tau^2} z + \frac{1}{\sigma^2} z + \frac{1}{\tau^2} z}{\frac{1}{\tau^2} z + \frac{1}{\sigma^2} z + \frac{1}{\tau^2} z} \right) \phi(z) dz$$

$$= \int_{-\infty}^{\infty} \Phi_N \left(\frac{\frac{1}{\tau^2} z + \frac{1}{\sigma^2} z}{\frac{1}{\tau^2} z + \frac{1}{\sigma^2} z} \right) \phi(z) dz$$

$$= F_N(H; \rho), \text{ say}$$

$$\text{where } \rho = \frac{\sigma^2}{\tau^2 + \sigma^2}$$

$$H = \frac{(\nu - \mu)}{\tau^2 + \sigma^2} \cdot \frac{1}{2}$$

The function $F_N(H, \rho)$ has been tabulated by Shanti S. Gupta [6] for given values of ρ (≥ 0), N and H . $F_N(H, \rho)$ is the probability that each of N standardized normal variables with equal correlation ρ will not exceed H .

This is the case in the present problem, since

$$P\{X_1 > Y, \dots, X_N > Y\} = P\{X_1 - Y > 0, \dots, X_N - Y > 0\}$$

$$= P \left\{ \frac{(X_1 - Y) - (\mu - \nu)}{\sqrt{(\sigma^2 + \tau^2)/2}} > \frac{(X_2 - Y) - (\mu - \nu)}{\sqrt{(\sigma^2 + \tau^2)/2}}, \dots, \frac{(X_N - Y) - (\mu - \nu)}{\sqrt{(\sigma^2 + \tau^2)/2}} > \frac{(X_N - Y) - (\mu - \nu)}{\sqrt{(\sigma^2 + \tau^2)/2}} \right\}$$

For all $i = 1, \dots, N$, the random variable

$$\frac{(X_i - Y - (u - v))}{(\sigma^2 + \tau^2)^{1/2}}$$

is a standardized normal random variable. Moreover, for $i \neq j$, the correlation coefficient between

$$\frac{(X_i - Y - (u - v))}{(\sigma^2 + \tau^2)^{1/2}} \quad \text{and} \quad \frac{(X_j - Y - (u - v))}{(\sigma^2 + \tau^2)^{1/2}}$$

is $\frac{\tau^2}{\sigma^2 + \tau^2}$.

$F_N^N(H, \rho)$ also gives the probability that the minimum of a set of N equally correlated standardized normal random variables exceeds $-H$.

Thus, if μ, v, σ^2 and τ^2 are known we can evaluate p using Gupta's

table. However if μ, v, σ^2, τ^2 are unknown, or some of them are unknown,

we can use their maximum likelihood estimates. If μ, v, σ^2 and τ^2 are

unknown, and if the samples

$$x_{i1}, \dots, x_{in_i}, \quad i = 1, \dots, N,$$

$$y_1, \dots, y_m$$

are available, let

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad i = 1, \dots, N,$$

$$s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2,$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N \bar{x}_i,$$

$$s^2 = \frac{1}{N} \sum_{i=1}^N \frac{1}{n_i} \sum_{j=1}^{n_i} s_i^2,$$

$$\sigma^2 \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

$\mu_y - \mu_z$ and covariance matrix

Now, $(X - Z, Y - Z)$ is bivariate normally distributed with means $(\mu_x - \mu_y, \mu_y - \mu_z)$

$$= P\{X - Z > 0, Y - Z > 0\}$$

$$p = P\{X > Z, Y > Z\}$$

and equal variance σ^2 . Then we want to estimate

Let X, Y, Z be independent normal with means μ_x, μ_y and μ_z respectively

2.4 Minimum Variance Unbiased Estimation of p in Some Cases

the properties of this estimate are extremely difficult to explore.

This estimate, again, can be evaluated using Gupta's table [6], however

$$\hat{H} = \frac{\bar{Y} - \bar{X}}{s^2 + (s')^2} \cdot \frac{1}{2}$$

where

$$\hat{p} = \frac{s}{s^2 + (s')^2}$$

$$\hat{p} = F_N(H, \hat{p})$$

a one-to-one function of μ, v, σ^2 and t^2 , is

that the maximum likelihood estimate remains invariant although p is not

Then, the maximum likelihood estimate of p , in view of Zehna's [18] result

$$(s')^2 = \frac{1}{m} \sum_{k=1}^m (y_k - \bar{y})^2$$

$$\bar{y} = \frac{1}{m} \sum_{k=1}^m y_k$$

Thus,

$$p = \left\{ \frac{(X - Z) - (\mu_x - \mu_z)}{\mu_z - \mu_x} \frac{(2) 1/2}{\sigma} > \frac{(Y - Z) - (\mu_y - \mu_z)}{\mu_z - \mu_y} \frac{(2) 1/2}{\sigma} \right\} .$$

Now,

$$\left(\frac{(X - Z) - (\mu_x - \mu_z)}{\mu_z - \mu_x} \frac{(2) 1/2}{\sigma} , \frac{(Y - Z) - (\mu_y - \mu_z)}{\mu_z - \mu_y} \frac{(2) 1/2}{\sigma} \right)$$

is bivariate normal with zero means, unit variances and correlation coefficient equal to $1/2$. Thus

$$p = B \left(\frac{\mu_z - \mu_x}{\mu_z - \mu_y} \frac{(2) 1/2}{\sigma} , \frac{\mu_z - \mu_y}{\mu_z - \mu_x} \frac{(2) 1/2}{\sigma} , \frac{1}{2} \right) .$$

where

$$(2.6) \quad B(h, k, \rho) = \frac{1}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi(1-\rho^2)} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right\} dx dy} .$$

D. B. Owen [11] considered evaluation of the function $B(h, k, \rho)$ using the

related function $T(h, a)$ defined by

$$T(h, a) = \frac{1}{2\pi} \int_a^0 \frac{\exp\left\{-\frac{1}{2}h^2(1+x^2)\right\}}{1+x^2} dx .$$

He derived various relations connecting $B(h, k, \rho)$ and $T(h, a)$. The function $T(h, a)$ and its differences are tabulated by Owen in [11]. Using his tables and relations one can obtain the volume under a bivariate surface over any polygon.

Assume that samples $(x_1, \dots, x_n)$, $(y_1, \dots, y_n)$ and $(z_1, \dots, z_n)$ are

available. If μ_x, μ_y, μ_z , and σ are unknown, we substitute their maximum

likelihood estimates to obtain the maximum likelihood estimate of p as

$$\hat{p} = B\left(\frac{\bar{z} - \bar{x}}{\bar{z} - \bar{y}}, \frac{(2)1/2}{(2)1/2}, \frac{1}{2}\right),$$

where

$$\begin{aligned}\bar{x} &= \frac{1}{\lambda} \sum_{i=1}^{\lambda} x_i, \\ \bar{y} &= \frac{1}{m} \sum_{j=1}^m y_j, \\ \bar{z} &= \frac{1}{n} \sum_{k=1}^n z_k,\end{aligned}$$

and

$$s^2 = \frac{1}{\lambda+m+n} \left[\sum_{i=1}^{\lambda} (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2 + \sum_{k=1}^n (z_k - \bar{z})^2 \right].$$

Theorem 2.1

If the means μ_x , μ_y and variance σ^2 are known, but μ_z is unknown and the data $(z_1, \dots, z_n)$ on Z are available, then the minimum variance

unbiased estimate of p is given as

$$\hat{p} = B\left(\frac{\bar{z} - \mu_x}{\bar{z} - \mu_y}, \frac{\sigma^2(2)1/2}{\sigma^2(2)1/2}, \frac{2n-1}{n-1}\right),$$

where

$$\bar{z} = \frac{1}{n} \sum_{k=1}^n z_k.$$

Moreover, this estimate is asymptotically equivalent to the maximum like-

hood estimate.

Proof

Let W and V be independent normal random variables, independent of Z, with means 0 and variance σ^2 , and let

$$I = \begin{cases} 1 & \text{if } -z_1 + W > -\mu^x, \quad -z_1 + V < -\mu^y \\ 0 & \text{otherwise} \end{cases}$$

then

$$\begin{aligned} E(I) &= P(-z_1 + W > -\mu^x, -z_1 + V < -\mu^y) \\ &= P\left(\frac{-z_1 + W + \mu^z}{z} > \frac{\mu^z - \mu^x}{z} \frac{(2)1/2\sigma}{(2)1/2\sigma}, \frac{-z_1 + W + \mu^z}{z} > \frac{\mu^z - \mu^y}{z} \frac{(2)1/2\sigma}{(2)1/2\sigma}\right) \\ &= B\left(\frac{\mu^z - \mu^x}{z} \frac{(2)1/2\sigma}{(2)1/2\sigma}, \frac{\mu^z - \mu^y}{z} \frac{(2)1/2\sigma}{(2)1/2\sigma}, \frac{1}{2}\right) \\ &= p. \end{aligned}$$

Since $\frac{-z_1 + W + \mu^z}{z}$ is bivariate normal with zero

means, unit variances and correlation coefficient equal to $1/2$.

Since $\bar{z}$ is a complete sufficient statistic, it follows from the Rao-Blackwell and Lehmann-Scheffe' theorems that the conditional expectation $E(I|\bar{z})$ is the minimum variance unbiased (m.v.u.) estimate of p .

Let $N^p(\mu, \Sigma)$ denote the p -variate normal distribution with mean vec-

tor $\bar{\mu}$ and variance-covariance matrix Σ ; then,

$$\begin{pmatrix} \bar{z} \\ -z_1 + W \\ -z_1 + V \end{pmatrix} \sim N_3 \left(\begin{pmatrix} \mu^z \\ -\mu^z \\ -\mu^z \end{pmatrix}, \begin{pmatrix} \frac{n}{\sigma^2} & -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2} \\ -\frac{n}{\sigma^2} & \frac{n}{\sigma^2} & \frac{n}{\sigma^2} \\ -\frac{n}{\sigma^2} & \frac{n}{\sigma^2} & \frac{n}{\sigma^2} \end{pmatrix} \right),$$

where $\sim$ denotes "distributed" as.

Now, using 8a.2(v), Page 441, Rao [14], the conditional distribution

of $(-z_1 + W, -z_1 + V)$ given $\bar{z}$ is

$$N_2 \left(\begin{pmatrix} -\bar{z} \\ -\bar{z} \end{pmatrix}, \begin{pmatrix} \sigma^2 - \frac{n}{\sigma^2} & 2\sigma^2 - \frac{n}{\sigma^2} \\ 2\sigma^2 - \frac{n}{\sigma^2} & \sigma^2 - \frac{n}{\sigma^2} \end{pmatrix} \right),$$

and the conditional correlation coefficient is $\frac{n-1}{2n-1}$.

Thus,

$$\hat{p} = E(I|z) = P\{-z_1 + W > -\mu_x, -z_1 + V > -\mu_y | z\}$$

$$P \left\{ \frac{-z_1 + W + z}{\sigma(2 - \frac{1}{n})^{1/2}} > \frac{-z_1 + V + z}{\sigma(2 - \frac{1}{n})^{1/2}}, \frac{-z_1 + W + z}{\sigma(2 - \frac{1}{n})^{1/2}} > \frac{-z_1 + V + z}{\sigma(2 - \frac{1}{n})^{1/2}} \right\}$$

i.e.

$$\hat{p} = B \left(\frac{z - \mu_x}{\sigma(2 - \frac{1}{n})^{1/2}}, \frac{z - \mu_y}{\sigma(2 - \frac{1}{n})^{1/2}}, \frac{2n-1}{n-1} \right)$$

and the maximum likelihood estimate of p in this situation is

$$\tilde{p} = B \left(\frac{z - \mu_x}{\sigma(2 - \frac{1}{n})^{1/2}}, \frac{z - \mu_y}{\sigma(2 - \frac{1}{n})^{1/2}}, \frac{z}{1} \right)$$

Thus, as n becomes large, the minimum variance unbiased estimate approaches in the limit to the maximum likelihood estimate.

Theorem 2.2

If the mean μ_z and variance σ_z^2 are known but the means μ_x and μ_y are unknown, and the data $(x_1, \dots, x_m)$, $(y_1, \dots, y_m)$ on X and Y respectively are available, then the m.v.u. estimate of p is given by

$$\hat{p} = B \left(\frac{\mu_z - x}{\sigma(2 - \frac{1}{m})^{1/2}}, \frac{\mu_z - y}{\sigma(2 - \frac{1}{m})^{1/2}}, \frac{[(2 - \frac{1}{m})(2 - \frac{1}{m})]^{1/2}}{1} \right)$$

where

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i, \quad \bar{y} = \frac{1}{m} \sum_{j=1}^m y_j$$

Moreover, as m and n become large, this estimate approaches in the limit to the maximum likelihood estimate.

Proof

Using the same technique as in the proof of Theorem 2.1, let W be a normal random variable, independent of X and Y , with mean 0 and variance σ^2 , and let

$$I = \begin{cases} 1 & \text{if } x_1 + W < \mu_z \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} E(I) &= P(x_1 + W < \mu_z, y_1 + W < \mu_z^z) \\ &= P\left(\frac{x_1 + W - \mu_x}{\sigma_x} > \frac{\mu_z - \mu_x}{\sigma_x}, \frac{y_1 + W - \mu_y}{\sigma_y} > \frac{\mu_z^z - \mu_y}{\sigma_y}\right) \\ &= P\left(\frac{(2)}{1/2} > \frac{(2)}{1/2}, \frac{(2)}{1/2} > \frac{(2)}{1/2}\right) \\ &= P \cdot \end{aligned}$$

Since $(\bar{x}, \bar{y})$ is a complete sufficient statistic, it follows from the Rao-Blackwell and Lehmann-Scheffe' theorem that the conditional expectation $E(I|\bar{x}, \bar{y})$ is the m.v.u. estimate of p . Now

$$\begin{pmatrix} \bar{x} \\ \bar{y} \\ x_1 + W \\ x_1 + W \end{pmatrix} \sim N_4 \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & 0 & 0 \\ 0 & 0 & \sigma_x^2 & 0 \\ 0 & 0 & 0 & \sigma_y^2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \frac{\sigma_x^2}{m} \\ \frac{\sigma_y^2}{m} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 2\sigma_x^2 \\ 2\sigma_y^2 \end{pmatrix}.$$

Using 8a.2(V) Page 441, Rao [14], the conditional distribution of $(x_1 + W, y_1 + W)$ given $(\bar{x}, \bar{y})$ is

$$N_2 \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}, \begin{pmatrix} 2\sigma_x^2 - \frac{\sigma_x^2}{m} & 0 \\ 0 & 2\sigma_y^2 - \frac{\sigma_y^2}{m} \end{pmatrix}.$$

and conditional correlation coefficient is

$$\frac{1}{[(2 - \frac{1}{\lambda})(2 - \frac{1}{\lambda})]^{1/2}}$$

and hence the m.v.u. estimate of ρ is

$$\hat{\rho} = P\{x_1 + W < \mu_z, y_1 + W < \mu_z | \bar{x}, \bar{y}\}$$

$$= P\left\{ \frac{\mu_z - \bar{x}}{\sigma(2 - \frac{1}{\lambda})^{1/2}} < \frac{\mu_z - \bar{x}}{\sigma(2 - \frac{1}{\lambda})^{1/2}}, \frac{y_1 + W - \bar{y}}{\sigma(2 - \frac{1}{\lambda})^{1/2}} < \frac{\mu_z - \bar{y}}{\sigma(2 - \frac{1}{\lambda})^{1/2}} \right\}$$

$$= P\left\{ \frac{\mu_z - \bar{x}}{\sigma(2 - \frac{1}{\lambda})^{1/2}} < \frac{\mu_z - \bar{y}}{\sigma(2 - \frac{1}{\lambda})^{1/2}}, \frac{y_1 + W - \bar{y}}{\sigma(2 - \frac{1}{\lambda})^{1/2}} < \frac{\mu_z - \bar{y}}{\sigma(2 - \frac{1}{\lambda})^{1/2}} \right\}$$

and this estimate is asymptotically equivalent to the maximum likelihood estimate as λ, m become large.

be samples of X , Y and Z respectively. Furthermore, we arrange the samples

$$(3.2) \quad x_1, \dots, x_n; y_1, \dots, y_m; z_1, \dots, z_n$$

Again let

$$(3.1) \quad p = \text{Prob}\{X < Z, Y < Z\}.$$

and G respectively, and let

be mutually independent random variables with continuous c.d.f.'s F_1, F_2

We shall first start with the simple case when $N = 2$. Let X, Y, Z

Some properties of the proposed estimators are also discussed.

distribution functions (c.d.f.'s) and that all are mutually independent.

variable Y . It is assumed that X 's and Y 's have continuous cumulative

the N random variables $X_1, X_2, \dots, X_N$ are simultaneously less than a random

Here, a statistic is proposed to estimate the probability that all of

been discussed by numerous authors.

is stochastically smaller than Y . Many properties of this estimate have

X 's and Y 's come from the same population versus the alternative that X

samples of X 's and Y 's. It is also used for testing the hypothesis that

ables. The U statistic is based on the ranks of observations in the pooled

the probability that X is less than Y where X and Y are both random vari-

Mann and Whitney [9] proposed a U statistic which estimates unbiasedly

3.1 Introduction

NON-PARAMETRIC ESTIMATION

CHAPTER III

of x 's, m y 's and n z 's in ascending order and let W count the number of times an x and a y precedes a z , i.e.,

W = number of triplets (x_i, y_j, z_k) such that $x_i < z_k$ and $y_j < z_k$

$$i = 1, \dots, l$$

$$j = 1, \dots, m$$

$$k = 1, \dots, n$$

then $\hat{p} = \frac{W}{lmn}$ can be used to estimate the parameter p . The properties of this particular estimator will be explored in the following sections.

3.2 Estimation of p and an Example

Let samples (3.2) be available and let

$$a_{ik} = \begin{cases} 1 & \text{if } x_i < z_k \\ 0 & \text{otherwise,} \end{cases}$$

(3.3)

$$b_{jk} = \begin{cases} 1 & \text{if } y_j < z_k \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, let

$$u_{ijk} = a_{ik} b_{jk},$$

then

$$W = \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^n a_{ik} b_{jk}.$$

Lemma 3.1

$\hat{p} = \frac{W}{lmn}$ is an unbiased estimate of p .

Proof

$$E(\hat{p}) = \frac{1}{lmn} \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^n E(u_{ijk})$$

$$= \frac{1}{lmn} \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^n p = p.$$

Since $U_{ijk} = 1$ if and only if $x_i < z_k$ and $y_j < z_k$ and so

$$E(U_{ijk}) = P(x_i < z_k, y_j < z_k) = p.$$

Example 3.1

Let $\lambda = 4, m = 3, n = 2$ and consider the observations

$$x_2 < y_3 < z_2 < x_1 < x_4 < y_1 < z_1 < x_3 < y_2.$$

There are 6 values of (x, y) which are less than z_1 .
There is 1 value of (x, y) which is less than z_2 .

Thus the experimental value of W is

$$w = 6 + 1 = 7,$$

and

$$\hat{p} = \frac{7}{24}.$$

Also

$$a_{11} = 1, a_{21} = 1, a_{31} = 0, a_{41} = 0, a_{12} = 0, a_{22} = 1, a_{32} = 0, a_{42} = 0,$$

and

$$b_{11} = 1, b_{21} = 0, b_{31} = 1, b_{12} = 0, b_{22} = 0, b_{32} = 1,$$

so that

$$\hat{p} = \frac{1}{\sum_{m=1}^4 \sum_{j=1}^3 \sum_{k=1}^2 a_{ijk} b_{jk}} = \frac{7}{24}.$$

3.3 Properties of $\hat{p}$

In this section we shall compute the variance of $\hat{p}$ and use this variance to show that $\hat{p}$ is a consistent estimate of p . Let

$$T = \max(X, Y),$$

and H denote the c.d.f. of T which is continuous because H is a function of c.d.f.'s of X and Y and the c.d.f.'s of X and Y are continuous by as-

sumption. Then

$$p = \text{Prob}\{X > Z, Y > Z\}$$

$$= \text{Prob}\{T > Z\}$$

$$= E^Z \text{Prob}\{T > Z | Z\}$$

(where E denotes the expectation operator)

$$= \int_{-\infty}^{\infty} dG(z) \int_z^{\infty} dH(t)$$

$$= \int_{-\infty}^{\infty} H(z) dG(z),$$

i.e.,

$$(3.4) \quad p = \int_{-\infty}^{\infty} H(z) dG(z).$$

Also,

$$1-p = P\{Z > T\}$$

$$= E^T \{Z > T | T\}$$

$$= \int_{-\infty}^{\infty} G(t) dH(t).$$

Thus,

$$(3.5) \quad 1-p = \int_{-\infty}^{\infty} G(t) dH(t).$$

Now, let

$$\psi_z = \sigma_z^2 H(z) = E^z [H(z) - E^z H(z)]^2 dG(z)$$

$$= \int_{-\infty}^{\infty} [H(z) - \int_{-\infty}^{\infty} H(z) dG(z)]^2 dG(z)$$

$$= \int_{-\infty}^{\infty} (H(z) - p)^2 dG(z) \quad \text{using (3.4).}$$

Thus, it follows that

$$(3.6) \quad \int_{-\infty}^{\infty} H^2(z) dG(z) = \psi^2 + p^2.$$

Again, let

$$\chi^2 = \sigma^2 G(T) = E_T[G(T) - E_T(G(T))]^2$$

$$= \int_{-\infty}^{\infty} [G(T) - \int_{-\infty}^{\infty} G(T) dH(T)]^2 dH(T)$$

$$= \int_{-\infty}^{\infty} [G(T) - 1 + p]^2 dH(T), \quad \text{using (3.5)}$$

Hence,

$$(3.7) \quad \int_{-\infty}^{\infty} (1 - G(t))^2 dH(t) = \chi^2 + p^2.$$

We shall use the above relations in deriving the variance of $\hat{p}$.

Also used are

$$W = \sum_{m,n} \sum_{j,k} w_{mj} w_{nk}$$

and

$$E(W) = \lambda_{mp}.$$

Theorem 3.1

The variance of W is given by

$$\text{Var}(W) = \lambda_{mm}[(\lambda_{m-1})\psi^2 + (n-1)(\lambda + m - 1)\chi^2 + (\lambda + m - 1)p + \{(n-1)(\lambda + m - 2) - 1\}p^2].$$

Proof

$$\text{Var}(W) = E(W^2) - (E(W))^2$$

$$= E(W^2) - (\lambda_{\text{mp}})^2,$$

so that we need to compute $E(W^2)$.

$$E(W^2) = E\left(\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} U_{njk}^2\right)$$

$$= E\left(\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} U_{njk} U_{i'j'k'}\right).$$

To compute this expected value we shall consider the following cases.

$$1) \quad i = i', j = j', k = k'.$$

$$\text{Here } U_{njk} U_{i'j'k'} = U_{njk}^2 = 1 \quad \text{if and only if (iff)}$$

$$x_i > z_k \quad \text{and} \quad y_j > z_k.$$

Thus,

$$E(U_{njk} U_{i'j'k'}) = P(x_i > z_k, y_j > z_k) = p.$$

$$11) \quad i = i', j = j', k \neq k'.$$

$$\text{Here } U_{njk} U_{i'j'k'} = 1 \quad \text{iff } \max(x_i, y_j) > z_k \text{ and } \max(x_i, y_j) > z_{k'},$$

$$\text{i.e. iff } t > z_k \quad \text{and} \quad t > z_{k'},$$

$$\text{where } t = \max(x_i, y_j),$$

thus

$$E(U_{njk} U_{i'j'k'}) = P(t > z_k \text{ and } t > z_{k'})$$

$$= E_P^t(z_k > t, z_{k'} > t | t)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^t dG(z_k) \int_{-\infty}^t dG(z_{k'})$$

since for $k \neq k'$, $z_k, z_{k'}$ are independent

$$= \int_{-\infty}^{\infty} (1 - G(t))^2 dH(t)$$

$$= \chi^2_2 + p^2 \quad \text{using (3.7)}.$$

$$(iii) \quad i = i', j \neq j', k \neq k'.$$

$$\text{Here } U_{ijk} U_{i'j'k'} = 1 \text{ iff } \max(x_i, y_j) < z_k \text{ and } \max(x_{i'}, y_{j'}) < z_{k'}.$$

$$\text{i.e. iff } t_1 < z_k \text{ and } t_2 < z_{k'},$$

$$\text{where } t_1 = \max(x_i, y_j)$$

$$t_2 = \max(x_{i'}, y_{j'}).$$

Note that t_1 may be equal to t_2 with nonzero probability, i.e. when t_1

and t_2 are the same observations. Thus

$$E(U_{ijk} U_{i'j'k'}) = P(t_1 < z_k, t_2 < z_{k'} | t_1 = t_2 = t)$$

$$+ P(t_1 < z_k, t_2 < z_{k'} | t_1 \neq t_2)$$

$$= \chi^2_2 + p^2 + p^2$$

$$= \chi^2_2 + 2p^2,$$

using case (ii) and noting that when $t_1 \neq t_2$ and $k \neq k'$,

$$P(t_1 < z_k, t_2 < z_{k'}) = P(t_1 < z_k)P(t_2 < z_{k'}) = p^2.$$

$$(iv) \quad i = i', j \neq j', k = k'.$$

$$\text{Here } U_{ijk} U_{i'j'k} = 1 \text{ iff } \max(x_i, y_j) < z_k \text{ and } \max(x_{i'}, y_{j'}) < z_k$$

$$\text{i.e. iff } t_1 < z_k \text{ and } t_2 < z_k,$$

$$\text{where } t_1 = \max(x_i, y_j) \text{ and } t_2 = \max(x_{i'}, y_{j'}).$$

Again, t_1 may be equal to t_2 with nonzero probability. Thus

$$E(U_{1jk} U_{1j',k'}) = P(t_1 < z_k, t_2 < z_k | t_1 = t_2 = t)$$

$$+ P(t_1 < z_k, t_2 < z_k | t_1 \neq t_2)$$

$$= P(t < z_k) + E_{z_k} P(t_1 < z_k, t_2 < z_k | z_k) \\ = p + \int_{-\infty}^{\infty} dG(z_k) \int_{z_k}^{\infty} dH(t_1) \int_{z_k}^{\infty} dH(t_2)$$

$$= p + \int_{-\infty}^{\infty} (H(z_k))^2 dG(z_k)$$

$$= p + \psi_2 + p_2 \quad \text{using (3.6).}$$

$$v) \quad i \neq i', j = j', k = k'.$$

By symmetry as in case (iv) we have

$$E(U_{1jk} U_{1j',k'}) = p + \psi_2 + p_2.$$

$$vi) \quad i \neq i', j = j', k \neq k'.$$

By symmetry as in case (iii), we have

$$E(U_{1jk} U_{1j',k'}) = \chi_2 + 2p_2.$$

$$vii) \quad i \neq i', j \neq j', k = k'$$

$$\text{Here } U_{1jk} U_{1j',k'} = 1 \text{ iff } \max(x_i, y_j) < z_k \text{ and } \max(x_{i'}, y_{j'}) < z_k$$

$$\text{i.e. iff } t_1 < z_k \text{ and } t_2 < z_k,$$

$$\text{where } t_1 = \max(x_i, y_j) \text{ and } t_2 = \max(x_{i'}, y_{j'}).$$

In this case $t_1 \neq t_2$ with probability 1. Thus

$$E(U_{1jk} U_{1j',k'}) = P(t_1 < z_k, t_2 < z_k)$$

$$= \psi_2 + p_2 \quad \text{using the result from case (iv).}$$

viii) $i \neq i', j \neq j', k \neq k'$.

In this case $U_{ijk}, U_{i'j'k'}$ are independent, so that

$$E(U_{ijk} U_{i'j'k'}) = p^2.$$

Hence,

$$E(W^2) = \sum_{i=1}^{\ell} \sum_{j=1}^m \sum_{k=1}^n \sum_{i'=1}^{\ell} \sum_{j'=1}^m \sum_{k'=1}^n E(U_{ijk} U_{i'j'k'})$$

$$= \ell m n p + \ell m n (n-1) (\chi^2 + p^2) + \ell m (m-1) n (n-1) (\chi^2 + 2p^2) + \ell m (m-1) n (p + \psi^2 + p^2) + \ell (\ell-1) m n (n-1) (\chi^2 + 2p^2) + \ell (\ell-1) m (m-1) n (\psi^2 + p^2)$$

$$+ \ell (\ell-1) m (m-1) n (n-1) p^2.$$

Hence denoting by σ^2 , the variance of W , we have

$$\sigma^2 = E(W^2) - (E(W))^2$$

$$= \ell m n p + (n-1) (\chi^2 + p^2) + (m-1) (n-1) (\chi^2 + 2p^2) + (m-1) (p + \psi^2 + p^2) + (\ell-1) (p + \psi^2 + p^2) + (\ell-1) (m-1) (\psi^2 + p^2) + (\ell-1) (m-1) (n-1) p^2 - \ell m n p^2$$

$$= \ell m n [(\ell m-1) \psi^2 + (n-1) (\ell + m - 1) \chi^2 + (\ell + m - 1) p$$

$$+ \{(n-1) (\ell + m - 2) - 1\} p^2].$$

Lemma 3. 2

$$(3.8) \quad \psi^2 \leq p(1-p), \text{ and } \chi^2 \leq p(1-p).$$

Proof

From (3.6) we have

$$0 \leq \psi^2 = \int_{-\infty}^{\infty} (H(z))^2 dg(z) - p^2$$

$$\leq \int_{-\infty}^{\infty} H(z) dG(z) - p^2$$

$$= p - p^2$$

$$= p(1-p) \cdot$$

Similarly, from (3.7) we have

$$0 \leq \chi^2 = \int_{-\infty}^{\infty} (1 - G(t))^2 dH(t) - p^2$$

$$\leq \int_{-\infty}^{\infty} (1 - G(t)) dH(t) - p^2$$

$$= 1 - \int_{-\infty}^{\infty} G(t) dH(t) - p^2$$

$$= p - p^2$$

$$= p(1-p) \cdot$$

Theorem 3.2

$\hat{p} = \frac{\chi_{mn}}{W}$ is consistent estimate of p .

Proof

$$\text{Var}(\hat{p}) = \frac{1}{\chi^2_{mn}} \text{Var}(W)$$

$$= \frac{1}{\chi_{mn}} [(\chi_{m-1}) \psi^2 + (n-1)(\chi + m - 1)\chi^2 + (\chi + m - 1)p$$

$$\leq \frac{1}{\chi_{mn}} [(\chi_{m-1})p(1-p) + (n-1)(\chi + m - 1)p(1-p)$$

$$+ \{(n-1)(\chi + m - 2) - 1\}p^2]$$

$$+ (\chi + m - 1)p + \{(n-1)(\chi + m - 2) - 1\}p^2]$$

$$= \frac{1}{\chi_{mn}} [(\chi_m + \chi n + m - n - 1)p - (\chi_m + n - u + p^2)$$

$$= p \left(\frac{n}{1} + \frac{m}{1} + \frac{\chi}{1} - \frac{\chi_m}{1} - \frac{\chi_m}{1} - p^2 \right) + \frac{n}{1} - \frac{\chi_m}{1} - \frac{\chi_m}{1}$$

as $\chi_{m,n} \rightarrow 0$ as $n \rightarrow \infty$.

$$\hat{p}_1 = \frac{\lambda_n}{n}, \text{ and } \hat{p}_2 = \frac{m}{n}$$

The statistics

$$p_2 = P\{Y > Z\}.$$

$$p_1 = P\{X > Z\}, \text{ and}$$

p_1, p_2 where,

and W can be used to estimate the product of marginal probabilities, viz.,

V to estimate the parameter p unbiasedly. However, the statistics U, V

Unfortunately, we could not find a suitable statistic based on U and

to be bivariate normal.

the null hypothesis H that $G = F_1 = F_2$ and showed the limit distribution nz 's. He also evaluated moments of the joint distribution of U and V under

determine the probability of a given (U, V) in a sample of λx 's, my 's and

give the correct significance level. He obtained recurrence relations to

native $G > F_1 > F_2, U > K_3, V < K_4$ where the constants K_i are chosen to

tive $G > F_1, G > F_2$, he proposed to use $U < K_1, V < K_2$ or with the alter-

$G > F_1 > F_2$. As a critical region for the hypothesis H with the alterna-

sis H that $G = F_1 = F_2$ against the alternative that $G > F_1, G > F_2$ or say

He proposed a test based on the two statistics U and V to test the hypothe-

$$k = 1, \dots, n.$$

$$j = 1, \dots, m;$$

$$i = 1, \dots, \lambda;$$

V = number of pairs (y_j, z_k) such that $y_j < z_k$, where

U = number of pairs (x_i, z_k) such that $x_i < z_k$,

tic. Let

D. R. Whitney [17] considered a bivariate extension of the U statis-

3.4 On a Use of a Bivariate Extension of the U Statistic

satisfy $E(\hat{p}_1) = p_1$, $E(\hat{p}_2) = p_2$. Moreover $\hat{p}_1$, $\hat{p}_2$ have other nice properties by themselves, but

$$E(\hat{p}_1 \hat{p}_2) \neq p_1 p_2$$

since $\hat{p}_1$ and $\hat{p}_2$ are not independent. However, the following theorem gives an unbiased estimate of the product $p_1 p_2$.

Theorem 3.3

$$\hat{p}_3 = \frac{UV - W}{\lambda_{mn}(n-1)}$$

is an unbiased estimate of $p_1 p_2$.

Proof

First, note that

$$U = \sum_{i=1}^n \sum_{k=1}^m a_{ik}, \text{ and } V = \sum_{i=1}^n \sum_{k=1}^m b_{ik},$$

so that

$$\begin{aligned} E(UV) &= E\left(\sum_{i=1}^n \sum_{k=1}^m a_{ik} \sum_{i=1}^n \sum_{k=1}^m b_{ik}\right) \\ &= E\left(\sum_{i=1}^n \sum_{k=1}^m \sum_{i'=1}^n \sum_{k'=1}^m a_{ik} b_{i'k'}\right) \\ &= E\left(\sum_{i=1}^n \sum_{k=1}^m \sum_{i'=1}^n \sum_{k'=1}^m a_{ik} b_{i'k'} + \sum_{i=1}^n \sum_{k=1}^m \sum_{i'=1}^n \sum_{k'=1}^m a_{ik} b_{i'k'}\right) \\ &= E\left(\sum_{i=1}^n \sum_{k=1}^m \sum_{i'=1}^n \sum_{k'=1}^m a_{ik} b_{i'k'}\right) + \sum_{i=1}^n \sum_{k=1}^m \sum_{i'=1}^n \sum_{k'=1}^m a_{ik} b_{i'k'} \\ &= \lambda_{mn} + \lambda_{mn}(n-1)p_1 p_2, \end{aligned}$$

since $E(a_{ik} b_{i'k'}) = p$ and for $k \neq k'$, a_{ik} , $b_{i'k'}$ are independent, so that

$$E(a_{ik} b_{i'k'}) = E(a_{ik}) E(b_{i'k'}) = p_1 p_2.$$

Thus,
$$\hat{p}_3 = \frac{UV - \lambda_{mp}}{UV - W} = \frac{\lambda_{mn}(n-1)}{UV - W}$$
 has expected value p_3 .

Corollary

If $G = F_1 = F_2$, we have $p = 1/3$, $p_1 = 1/2$ and $p_2 = 1/2$

so that

$$E(UV) = \frac{1}{3} \lambda_{mn} + \frac{1}{4} \lambda_{mn}(n-1) = \frac{1}{12} \lambda_{mn} + \frac{1}{4} \lambda_{mn}^2,$$

which is the same result as derived by D. R. Whitney [17] in Section 3

for the case $G = F_1 = F_2$.

Further properties of $\hat{p}_3$ are not explored, because it does not seem

to be of much importance except perhaps this statistic may be used to

test the hypothesis that all observations come from the same distribution

function against the alternative that the departure from the null hypothesis

is in a certain direction.

3.5 A Generalization

Let $X_1, \dots, X_N, Y$ be $(N+1)$ independent random variables with continuous

cumulative distribution functions $F_1, \dots, F_N, G$ respectively, and

$$p = \text{Prob}\{X_1 < Y, \dots, X_N < Y\}.$$

Furthermore, suppose that the samples $(x_{11}, \dots, x_{1m_1}), \dots, (x_{N1}, \dots, x_{Nm_N})$, $(y_1, \dots, y_n)$ of $X_1, \dots, X_N, Y$, respectively, are available. Then arranging

the samples in ascending order and defining

$$a_{ij}^{(1)} = \begin{cases} 1 & \text{if } x_{ij}^{(1)} > z_k \\ 0 & \text{otherwise} \end{cases}$$

$$i=1, \dots, N, \quad k=1, \dots, n$$

$$j=1, \dots, m$$

$$W = \sum_{i=1}^N \sum_{j=1}^m a_{ij}^{(1)}$$

and

$$\hat{p} = \frac{m_1 m_2 \dots m_m \cdot n}{W}$$

Then it can be shown that $\hat{p}$ is unbiased consistent estimate of p .

CHAPTER IV

A DUAL PROBLEM

4.1 Introduction

Let $X, Y_1, \dots, Y_N$ be $(N+1)$ random variables with continuous cumulative

distribution functions $F, G_1, \dots, G_N$ respectively. Then, in certain phys-

cal situations, as was pointed out in Chapter I, it might be of interest

to estimate the probability that the random variable X is simultaneously

less than the N random variables $Y_1, \dots, Y_N$ over a time interval $[0, T]$.

Thus, let

$$(4.1) \quad q = \text{Prob}\{X < Y_1, \dots, Y_N\}.$$

This problem is called a dual problem. We shall use the following nota-

tions and lemmas throughout this chapter. Let

$$T^* = \min(Y_1, \dots, Y_N),$$

and suppose it has a c.d.f. H^* . Then we have the following lemmas.

Lemma 4.1

$$P(X > Y_1, \dots, Y_N) = P(X > T^*).$$

Lemma 4.2

$$H^*(t) = 1 - \prod_{i=1}^N (1 - G_i(t)).$$

Proof

$$H^*(t) = P(T^* \leq t)$$

$$= P(\min(Y_1, \dots, Y_N) \leq t)$$

$$q = P(X > T^*) \\ = E P^X(T^* > X | X)$$

Proof

$$q = \int_{-\infty}^{\infty} [1 - G(x)]^N f(x) dx.$$

Lemma 4.3

and thus

$$h^*(t) = N(1 - G(t))^{N-1} g(t),$$

random variable T^* is given as

variables are mutually independent. Then the density function of the and X has the continuous c.d.f. F with density f and all these random $Y_1, \dots, Y_N$ have the same continuous c.d.f. Also, G has the density g . Next, suppose that $G_1 = G_2 = \dots = G_N \equiv G$, i.e., the random variables zero. Moreover the random variable T^* is independent of X .

tinuous, and so the probability of getting ties from the c.d.f. H^* is Since $H^*(t)$ is a function of continuous c.d.f.'s, $H^*(t)$ itself is con-

$$H^*(t) = G_1(t) + G_2(t) - G_1(t)G_2(t).$$

In the particular case when $N = 2$, $H^*(t)$ reduces to

$$\begin{aligned} &= 1 - \prod_{i=1}^N (1 - G_i(t)). \\ &= 1 - P(Y_1 > t)P(Y_2 > t) - P(Y_1 > t) \\ & \quad (Since \min(Y_1, \dots, Y_N) > t \Leftrightarrow Y_1 > t, Y_2 > t, \dots, Y_N > t) \\ &= 1 - P\{Y_1 > t, Y_2 > t, \dots, Y_N > t\} \\ &= 1 - P\{\min(Y_1, \dots, Y_N) > t\} \end{aligned}$$

$$= \int_{-\infty}^{\infty} (1 - H^*(x)) dF(x)$$

$$= \int_{-\infty}^{\infty} (1 - G(x))^N f(x) dx$$

Lemma 4.4

If $F = G_1 = \dots = G_N = G$, then $q = \frac{1}{N+1}$.

4.2 Estimation of q When $Y_1, \dots, Y_N, X$ are all Independent Normal.

Let $Y_1, \dots, Y_N$ all be independent identically distributed normal

random variables with mean μ , variance σ^2 , and let X be normal with mean ν , variance τ^2 and independent of all the Y 's. Then

$$q = P(X < Y_1, \dots, Y_N) = P(X - Y_1 < 0, \dots, X - Y_N < 0)$$

$$= P\left(\frac{X - Y_1}{\sqrt{\sigma^2 + \tau^2}} < \frac{\nu - \mu}{\sqrt{\sigma^2 + \tau^2}}, \dots, \frac{X - Y_N}{\sqrt{\sigma^2 + \tau^2}} < \frac{\nu - \mu}{\sqrt{\sigma^2 + \tau^2}}\right)$$

$$= P\left(\max \left\{ \frac{X - Y_1}{\sqrt{\sigma^2 + \tau^2}}, \dots, \frac{X - Y_N}{\sqrt{\sigma^2 + \tau^2}} \right\} < \frac{\nu - \mu}{\sqrt{\sigma^2 + \tau^2}}\right)$$

$$= F_N(H, \rho)$$

where

$$\rho = \frac{\tau^2}{\sigma^2 + \tau^2}, \text{ and}$$

$$H = \frac{\nu - \mu}{\sqrt{\sigma^2 + \tau^2}}$$

Since, for all $i = 1, \dots, N$, the random variables

$$\frac{X - Y_i}{\sqrt{\sigma^2 + \tau^2}}$$

known and the data $(x_1, \dots, x_n)$ on X are available, then the m.v.u. estimate of q is given as

$$\hat{q} = \frac{\frac{1}{n} \sum_{i=1}^n x_i - \mu_x}{\frac{\sigma(2 - \frac{1}{n})^{1/2}}{1/2}} \cdot \frac{\frac{1}{n} \sum_{i=1}^n x_i - \mu_x}{\frac{\sigma(2 - \frac{1}{n})^{1/2}}{1/2}} \cdot \frac{1}{2n-1} \cdot \left(\frac{1}{n-1} \right)$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Moreover, this estimate is asymptotically the same as the maximum likelihood estimate for this case.

Theorem 4.2

If the mean μ_x and variance σ^2 are known but the means μ_y and μ_z are unknown, and the data $(y_1, \dots, y_m)$, $(z_1, \dots, z_m)$ on Y and Z , respectively, are available, then the m.v.u. estimate of q is given by

$$\hat{q} = \frac{\frac{1}{m} \sum_{j=1}^m y_j - \mu_y}{\frac{\sigma(2 - \frac{1}{m})^{1/2}}{1/2}} \cdot \frac{\frac{1}{m} \sum_{j=1}^m y_j - \mu_y}{\frac{\sigma(2 - \frac{1}{m})^{1/2}}{1/2}} \cdot \frac{1}{2m-1} \cdot \left(\frac{1}{m-1} \right)$$

where

$$\bar{y} = \frac{1}{m} \sum_{j=1}^m y_j, \text{ and}$$

$$\bar{z} = \frac{1}{m} \sum_{k=1}^m z_k$$

Moreover, as λ and m become large, this estimate approaches in limit to the maximum likelihood estimate.

4.4 Estimation of q When X Has a Normal Distribution with Mean μ , Variance σ^2 and $Y_1, \dots, Y_N$ are All Independent Exponential.

Let $Y_1, \dots, Y_N$ all be independent and identically distributed with the

$$p = \int_{-\infty}^{\infty} [1 - G(x)]^N f(x) dx$$

By Lemma 4.3, we have

Proof

where $\Phi(x)$ is defined by (2.4).

$$(4.4) \quad p = \Phi\left(-\frac{\mu}{\sigma}\right) + \exp(-N\mu + \frac{N^2\sigma^2}{2})(1 - \Phi(N\sigma - \frac{\mu}{\sigma})),$$

(4.2), then

$Y_1, \dots, Y_N$ are independent identically distributed with density function

If X is normally distributed with mean μ and variance σ^2 and

Lemma 4.5

In this situation, we have the following result.

capacity of different transistors, otherwise it fails.

the current supplied to the component is simultaneously less than the capacity of different transistors, otherwise it fails. The component does not fail as long as all exponentially distributed. The component does not fail as long as supplied to a component is normally distributed with certain mean and variance whereas the capacity of the different transistors inside the component σ^2 . This might be a practical situation, for example, the electric current Next, suppose that X is a normal random variable with mean μ and variance

$$(4.3) \quad G(x) = \begin{cases} 1 - e^{-x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

and the distribution function

$$(4.2) \quad g(y) = \begin{cases} e^{-y} & \text{if } y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

density function

$$= \int_0^\infty f(x) dx + \int_0^\infty e^{-Nx} f(x) dx$$

$$= \int_0^\infty \frac{1}{1 - 1/2 \left(\frac{\sigma}{x-\mu} \right)^2} e^{-Nx} dx + \int_0^\infty \frac{1}{1 - 1/2 \left(\frac{\sigma}{x-\mu} \right)^2} e^{-Nx} dx$$

$$= \int_{-\mu/\sigma}^\infty \phi(y) dy + e^{-N\mu + \frac{N^2\sigma^2}{2}} \int_0^\infty \phi(y) dy$$

$$= \phi(-\mu/\sigma) + e^{-N\mu + \frac{N^2\sigma^2}{2}} (1 - \phi(N\sigma)) \cdot \frac{\sigma}{\mu}$$

Corollary: If $\mu = 0$, $\sigma = 1$, then

$$p = \frac{1}{2} + e^{N^2/2} (1 - \phi(N)) \cdot$$

Now, suppose that a random sample distributed as X , say $x_1, \dots, x_m$ is ob-

served. The problem here is to find the minimum variance unbiased estimate

of the q by use of this sample. Letting $\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$, this estimate is

provided by:

Theorem 4.3

If σ is known and $m > 2$, the minimum variance unbiased estimate

of q is

$$(4.5) \quad \hat{q} = \phi\left(\frac{1/2 - \frac{1}{2} \frac{\sigma}{\bar{x}}}{\frac{1}{2} \frac{\sigma}{\bar{x}}}\right) + e^{-N\bar{x} + \frac{N^2(m-1)\sigma^2}{2m}}$$

$$\cdot \left[1 - \phi\left(\frac{(m-1)^{3/2} N \sigma^2 - m(m-1) \frac{1}{2} \frac{\sigma}{\bar{x}}}{1/2 [1 + m(m-2)] \frac{1}{2} \frac{\sigma}{\bar{x}}}\right) \right]$$

And as m becomes large, this estimate approaches in the limit to the maxi-

mum likelihood estimate of q .

Proof

Let $q = q_1 + q_2$,

where $q_1 = \Phi(-\eta/\sigma)$, and

$$q_2 = e^{-\frac{N\mu + N^2\sigma^2}{2}} (1 - \Phi(N\sigma - \frac{\sigma}{\mu})).$$

To find the m.v.u. estimate of q , it would be enough to find an unbiased estimate of q which is a function of a complete-sufficient statistic, in view of the Rao-Blackwell and Lehman-Scheffe' theorems. A complete-sufficient statistic in this case is $\bar{x}$ and if we can find an unbiased estimate $\hat{q}_1$ of q_1 and an unbiased estimate $\hat{q}_2$ of q_2 which are functions of a complete-sufficient statistic, then m.v.u. estimate of q will be

$$\hat{q} = \hat{q}_1 + \hat{q}_2.$$

Note that

$$q_1 = \int_0^\infty \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}(\frac{\sigma}{x-\mu})^2} dx,$$

so that, following Lieberman and Reonikoff [8], the m.v.u. estimate of q_1

is given by

$$\hat{q}_1 = \Phi\left(-\frac{(m)^{1/2}}{x} \frac{1/2}{(m-1)^{1/2}}\right),$$

which is a function of a complete sufficient statistic.

Thus, we only need to find an unbiased estimate of q_2 which is a

function of a complete sufficient statistic.

Let $t_1 = e^{-Nx_1}$.

Then,

$$E(t_1) = \int e^{-Nx_1} \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2} dx_1$$

$$= \int \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2\sigma^2}(x_1 - (\mu - N\sigma^2))^2} dx_1 e^{-\frac{N\mu}{2} + \frac{N^2\sigma^2}{2}} = e$$

Further, let

$$t_2 = \begin{cases} 0 & \text{if } -\infty < x_2 < N\sigma^2 \\ 1 & \text{otherwise} \end{cases}$$

then

$$E(t_2) = 1 - \text{Prob}(-\infty < x_2 < N\sigma^2) = 1 - \int_{N\sigma^2 - \frac{\mu}{\sigma}}^{-\infty} \phi(y) dy = 1 - \Phi(N\sigma - \frac{\mu}{\sigma})$$

Since t_1 is a function of x_1 alone and t_2 is a function of x_2 alone, x_1, x_2 being independent, implies that t_1 and t_2 are independent. Thus,

Let

$$t = t_1 t_2, \text{ and}$$

we obtain

$$E(t) = E(t_1)E(t_2) = e^{-\frac{N\mu}{2} + \frac{N^2\sigma^2}{2}} (1 - \Phi(N\sigma - \frac{\mu}{\sigma})) = q_2.$$

Thus, in view of Rao-Blackwell, Lehmann-Scheffe' theorems, since $\bar{x}$ is a complete-sufficient statistic, $E(t|\bar{x})$ is m.v.u. estimate of q_2 . Note

that

$$t = \begin{cases} e^{-N\bar{x}_1} & \text{if } N\sigma^2 < x_2 < \infty, \text{ and } -\infty < x_1 < \infty \\ 0 & \text{otherwise} \end{cases}$$

$$= e^{-N\bar{x}_1} I_{(N\sigma^2, \infty)}(x_2) \quad -\infty < x_1 < \infty,$$

where

$$I_{(a,b)}(x) = \begin{cases} 1 & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$E(t|x) = E(e^{-Nx_1} I^{(N\sigma^2, \infty)}(x_2) | x)$$

Now,

$$\begin{pmatrix} x \\ x_1 \\ x_2 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} \mu \\ \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \frac{\sigma^2}{m} & \frac{\sigma^2}{m} \\ \frac{\sigma^2}{m} & \sigma^2 & 0 \\ \frac{\sigma^2}{m} & 0 & \sigma^2 \end{pmatrix} \right)$$

Thus, using 8a.2(v) Page 441, Rao [14], the conditional distribution of

(x_1, x_2) given x is

$$\begin{pmatrix} x \\ x \end{pmatrix}^{N_2}, \quad \begin{pmatrix} 1 - \frac{1}{\sigma^2} & -\frac{1}{\sigma^2} \\ -\frac{1}{\sigma^2} & 1 - \frac{1}{\sigma^2} \end{pmatrix}$$

and the conditional correlation coefficient is $-\frac{1}{\sigma^2}$. Therefore, the

conditional density of x_1, x_2 given x , after simplification, reduces to

$$f(x_1, x_2 | x) = \left(\frac{m-2}{m} \right)^{1/2} \frac{1}{2} e^{-\frac{1}{2(m-2)\sigma^2} [(x-x_1)^2 + (x-x_2)^2] - \frac{1}{\sigma^2} (x_1-x_2)^2}$$

Hence,

$$E(t|x) = E(e^{-Nx_1} I^{(N\sigma^2, \infty)}(x_2) | x)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-Nx_1} \left(\frac{m-2}{m} \right)^{1/2} \frac{1}{2} e^{-\frac{1}{2(m-2)\sigma^2} [(x-x_1)^2 + (x-x_2)^2] - \frac{1}{\sigma^2} (x_1-x_2)^2} dx_2 dx_1$$

$\cdot dx_2 dx_1$

$$= \int_{-\infty}^{\infty} e^{-Nx_1} \frac{1}{2} \frac{1}{2} e^{-\frac{1}{2(m-2)\sigma^2} (x-x_1)^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2(m-2)\sigma^2} (x-x_2)^2} dx_2 dx_1$$

The exponent term, except for the constant, of the second integral, is

$$\begin{aligned} & \left(\bar{x} - x_2 \right)^2 - \frac{m}{1} \left(x_1 - x_2 \right)^2 = \bar{x}^2 + x_2^2 - 2x_1x_2 - \frac{m}{1} \left(x_2^2 + x_1^2 - 2x_1x_2 \right) \\ & = \bar{x}^2 - \frac{1}{1} x_2^2 + \frac{m-1}{m} \left[x_2^2 - \frac{m}{m-1} \left(\bar{x} - \frac{1}{m} x_1 \right)^2 \right] \end{aligned}$$

$$- \frac{m}{m-1} \left(\bar{x} - \frac{1}{m} x_1 \right)^2$$

So that,

$$\frac{(m-1)}{1/2} \frac{1/2}{(2\pi)^{1/2}} e^{-\int_{NO}^{\infty} \frac{m}{2(m-2)} \frac{1}{x_2} \left(\bar{x} - \frac{1}{m} x_1 \right)^2 dx_2}$$

$$= \int_{NO}^{\infty} \frac{1}{2} \frac{1}{x_2} \left(\bar{x} - \frac{1}{m} x_1 \right)^2 e^{-\frac{1}{2} \frac{m}{m-1} \left(\bar{x} - \frac{1}{m} x_1 \right)^2} dt$$

$$\text{(where } 1/k = \frac{m-1}{m-2} \frac{1}{2} \text{)}$$

$$= 1 - \phi \left\{ kNO^2 - \frac{m-1}{m} \left(\bar{x} - \frac{1}{m} x_1 \right) \right\}.$$

Thus

$$E(t|\bar{x}) = \int_{-\infty}^{\infty} e^{-Nkx_1 - \frac{m}{2(m-2)} \frac{1}{x_2} \left(\bar{x} - x_1 \right)^2 - \frac{m}{2(m-2)} \frac{1}{x_2} \left(\bar{x} - \frac{1}{m} x_1 \right)^2 - \frac{m}{1} x_2^2 - \frac{m-1}{m} \left(\bar{x} - \frac{1}{m} x_1 \right)^2} dx_1.$$

Simplifying the exponent term again

$$\begin{aligned} & -Nkx_1 - \frac{m}{2(m-2)} \frac{1}{x_2} \left[2(m-1)x_2^2 + (m-1)(1 - \frac{m}{1})x_2^1 - 2(m-1)x_1x_2 - mx_2^2 \right] \\ & = -Nkx_1 - \frac{m}{2(m-2)} \frac{1}{x_2} \left[x_2^2 + x_2^1 - 2x_1x_2 + x_2^1 - \frac{m}{1} x_2^1 - \frac{m-1}{m} \left(x_2^2 + \frac{m}{1} x_2^1 - \frac{m}{2} x_1x_2 \right) \right] \\ & = -Nkx_1 - \frac{m}{2(m-2)} \frac{1}{x_2} \left[\frac{m-1}{1} x_2^2 + \frac{m}{2} x_2^1 + 2x_1x_2 \right] \end{aligned}$$

$$x_1 = \left(\frac{m-1}{m}\right)^{1/2} \sigma t + \bar{x} - \frac{m}{N(m-1)\sigma^2},$$

and the jacobian transformation is

$$dx_1 = \left(\frac{m-1}{m}\right)^{1/2} \sigma dt.$$

Thus the integral becomes

$$\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \phi[Nk\sigma^2 - \frac{m-1}{mk} \{ \bar{x} - \frac{m}{1} \left(\frac{m-1}{m}\right)^{1/2} \sigma t - \frac{m}{N(m-1)\sigma^2} \}] e^{-\frac{1}{2}t^2} dt$$

$$= \int_{-\infty}^{\infty} \phi[Nk\sigma^2 - \frac{m-1}{k} \{ (m-1)\bar{x} - \left(\frac{m-1}{m}\right)^{1/2} \sigma t + \frac{m}{N(m-1)\sigma^2} \}] \phi(t) dt$$

$$= \int_{-\infty}^{\infty} \phi[Nk\sigma^2 - k\bar{x} + \frac{[m(m-1)]^{1/2}}{k} \sigma t - \frac{m}{Nk} \sigma^2] \phi(t) dt$$

$$= \int_{-\infty}^{\infty} \phi(t) \phi \left[\left(1 - \frac{m}{1}\right) Nk\sigma^2 - k\bar{x} + \frac{[m(m-1)]^{1/2}}{k} \sigma t \right] dt$$

$$= \int_{-\infty}^{\infty} \phi(t) \phi(a + bt) dt$$

$$= \phi \left(\frac{a}{(1+b^2)^{1/2}} \right),$$

following 7.6.2, Page 35 of Owen [11] where

$$k = \frac{(m-1)^{1/2}}{(m-2)^{1/2} \sigma},$$

$$a = \left(1 - \frac{m}{1}\right) Nk\sigma^2 - k\bar{x}$$

$$= \frac{m}{m-1} \frac{(m-1)^{1/2}}{(m-2)^{1/2} \sigma} - \left(\frac{m-1}{m-2}\right)^{1/2} \frac{\bar{x}}{\sigma},$$

$$b = \frac{[m(m-1)]^{1/2}}{k\sigma}$$

$$= \frac{[m(m-1)]^{1/2}}{1} = \frac{[m(m-1)]^{1/2}}{1}$$

Therefore,

$$\hat{q}_2 = E(t|x) = e^{-N_x} + \frac{2m}{N^2(m-1)\sigma^2} [1 - \Phi\left(\frac{1+b_2}{a}\right)]$$

$$= e^{-N_x} + \frac{2m}{N^2(m-1)\sigma^2} \left[1 - \Phi\left(\frac{[m(m-2)]^{1/2}}{N\sigma} - \left(\frac{m-1}{3/2}\right)^{1/2} \frac{1}{[1 + \frac{m(m-2)}{1}]^{1/2}}\right) \right]$$

$$= e^{-N_x} + \frac{2m}{N^2(m-1)\sigma^2} \left[1 - \Phi\left(\frac{[m(m-2)]^{1/2}}{N\sigma} - \left(\frac{m-1}{3/2}\right)^{1/2} \frac{1}{[1 + \frac{m(m-2)}{1}]^{1/2}}\right) \right]$$

and the minimum variance unbiased estimate of q is, therefore,

$$\hat{q} = \hat{q}_1 + \hat{q}_2$$

$$= \Phi\left(-\frac{(m)^{1/2}}{1/2} \frac{1}{x} \right) + e^{-N_x} + \frac{2m}{N^2(m-1)\sigma^2} [1 - \Phi\left(\frac{(m-1)^{1/2}}{3/2} \frac{1}{N\sigma} - \left(\frac{m-1}{3/2}\right)^{1/2} \frac{1}{[1 + \frac{m(m-2)}{1}]^{1/2}}\right)]$$

The maximum likelihood estimate of q when σ is known is given as

$$\hat{q} = \Phi\left(-\frac{\bar{x}}{\sigma}\right) + e^{-N_x} + \frac{2m}{N^2\sigma^2} [1 - \Phi(N\sigma)]$$

and thus as m becomes large, the minimum variance unbiased estimate is asymptotically the same as the maximum likelihood estimate.

Corollary 4.1

Let $Y_1, \dots, Y_N$ all be independent exponential random variables with

known parameter α , i.e., the density of Y 's is given as

$$f(y) = \begin{cases} \alpha e^{-\alpha y} & y > 0, \alpha > 0 \\ 0 & \text{otherwise} \end{cases}$$

Also let X be an independent normal random variable with unknown mean μ and known variance σ^2 . Then, the minimum variance unbiased estimate of q based on the random sample $x_1, \dots, x_m$ from X , is

$$\hat{q} = \left[\frac{1}{2} - \frac{(m-1)}{2} \frac{1}{x} \right] + e^{-\alpha N x} + \frac{\alpha^2 N^2 (m-1) \sigma^2}{2m} [1 - \Phi] \left\{ \frac{(m-1)}{3/2} \frac{1}{\alpha N \sigma^2} - \frac{(m-1)}{2} \frac{1}{x} \right\}.$$

Proof

Note that the c.d.f. of Y is

$$G(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\alpha x} & \text{if } x > 0. \end{cases}$$

Thus,

$$q = \int_{-\infty}^{\infty} [1 - G(x)]^N f(x) dx$$

$$= \int_0^{\infty} f(x) dx + \int_{-\infty}^0 e^{-\alpha N x} f(x) dx$$

$$= \Phi\left(-\frac{\mu}{\sigma}\right) + e^{-\alpha N \mu} + \frac{\alpha^2 N^2 \sigma^2}{2} [1 - \Phi(\alpha N \sigma)] , \text{ and therefore}$$

by replacing N by αN in Lemma 4.5 and in theorem 4.3, we get the desired

result.

Note that if μ and σ are both unknown in Lemma 4.5, then the maximum

likelihood estimate of q is

$$\hat{q} = \Phi\left(-\frac{s}{x}\right) + e^{-\frac{N}{s^2} x} + \frac{N}{2} \left(1 - \Phi\left(N \frac{s}{x}\right)\right),$$

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i, \text{ and}$$

$$s^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2.$$

Also, if the parameter α is unknown in Cor. 4.1 and the data

$y_{11}, \dots, y_{1m_1}, \dots, y_{N_1}, \dots, y_{Nm_N}$ are available, then the maximum likelihood

estimate of q is

$$\hat{q} = \phi\left(-\frac{\bar{x}}{s}\right) + e^{-\frac{\bar{y}_N \bar{x}}{2} + \frac{\bar{y}_N^2 s^2}{2}} [1 - \phi\left(\frac{\bar{y}_N s}{\bar{x}}\right)],$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N \bar{y}_i, \text{ where}$$

$$\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}, \quad i=1, \dots, N,$$

$\bar{x}$ and s^2 are defined above.

4.5 Estimation of q when X is uniform and $Y_1, \dots, Y_N$ are all Independent
Exponential

Let $Y_1, \dots, Y_N$ be all independent and identically distributed with

the density function and distribution function given by (4.2) and (4.3) respectively, and X be uniform random variable between 0 and θ , i.e.,

density function of X is given by

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases},$$

then we have the following lemma.

Lemma 4.6

In above case

$$q = \frac{N\theta}{1 - e^{-N\theta}}.$$

Proof

By Lemma 4.3, we have

$$q = \int_{-\infty}^{\infty} [1 - G(x)]^N f(x) dx$$

$$= \int_0^{\infty} \theta \frac{e^{-N\theta}}{1 - e^{-N\theta}} dx$$

$$= \frac{N\theta}{1 - e^{-N\theta}}.$$

Theorem 4.4

In this case the minimum variance unbiased estimate of q is

$$\hat{q} = \frac{m-1}{mNt} + \frac{1}{1} e^{-Nt} (1 - \frac{m-1}{mNt}),$$

where $t = \max(x_1, \dots, x_m)$ and $x_1, \dots, x_m$ is observed random sample distributed as X .

Proof

Let $u = e^{-Nx_1}$, then we obtain

$$E(u) = q.$$

Thus, since t is a complete sufficient statistic, in view of

Rao-Blackwell and Lehmann-Scheffe theorems, $E(u|t)$ is the m.v.u. estimate

of q . The conditional distribution of x_1 given t is of mixed type and its

generalized probability density function with respect to mixture of Lebesgue

and counting measure is, as given by Patil and Wani [13],

$$h(x_1|t) = \begin{cases} (1 - \frac{m}{t})^{\frac{1}{m}} & 0 < x_1 < t \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Thus, } E(u|t) = \int_0^t e^{-Nx_1} (1 - \frac{m}{t})^{\frac{1}{m}} \frac{1}{t} dx_1 + \frac{m}{t} e^{-Nt}$$

$$= \frac{1}{t} e^{-Nt} + \frac{m}{t} \frac{1 - e^{-Nt}}{N}$$

$$= \frac{m-1}{m} e^{-Nt} + \frac{1}{t} (1 - e^{-Nt})$$

4.6 Nonparametric Estimation

Let X, Y, Z be independent random variables with continuous cumulative distribution functions F, G_1, G_2 respectively, and

$$q = \text{Prob}(X < Y \text{ and } X < Z)$$

$$= \text{Prob}(X > T^*),$$

where $T^* = \min(Y, Z)$ and let H^* be c.d.f. of T^* .

Further, assume that

$$x_1, \dots, x_m, y_1, \dots, y_m, z_1, \dots, z_n$$

are samples of X, Y and Z respectively. Then, the problem of estimating q on the basis of samples of X, Y and Z is exactly similar to that of

estimating $p = \text{Prob}(X < Z, Y < Z)$ discussed in Chapter III. However, the main results are stated without proof in the following, since the proof follows exactly in the same way as in Chapter III.

$$\text{Let } a_{ij} = \begin{cases} 1 & \text{if } x_i < y_j \\ 0 & \text{otherwise} \end{cases},$$

$$b_{ik} = \begin{cases} 1 & \text{if } x_i < z_k \\ 0 & \text{otherwise, where} \end{cases}$$

$$i = 1, \dots, \ell; j = 1, \dots, m; k = 1, \dots, n.$$

Further, suppose that

$$u_{ijk} = a_{ij} b_{ik}, \text{ and}$$

$$w_i = \sum_{j=1}^{\ell} \sum_{k=1}^m u_{ijk}$$

$$= \text{number of triplets } (x_i, y_j, z_k) \text{ such that } x_i < y_j, \\ x_i < z_k,$$

i.e., w_i counts the number of times an x precedes a y and a z simultaneously in the arranged sample of x 's, y 's and z 's in ascending order. Then we have the following results.

Lemma 4.7

$$\hat{q} = \frac{w_1}{\ell m} \text{ is an unbiased estimate of } q.$$

Proof

$$u_{ijk} = 1 \text{ if and only if } x_i < y_j \text{ and } x_i < z_k \text{ so that } E(u_{ijk}) = \\ P(x_i < y_j, x_i < z_k) = q, \text{ and the proof follows.}$$

Example 4.1

Let $\ell = 4, m = 3, n = 2$, and consider the observations

$$x_2 < y_3 < z_2 < x_1 < x_4 < y_1 < z_1 < x_3 < y_2.$$

There are 3 values of x which are less than (y_1, z_1) , there is 1 value of x which is less than (y_1, z_2) ,

there are 3 values of x which are less than (y_2, z_1) ,
there is 1 value of x which is less than (y_2, z_2) ,
there is 1 value of x which is less than (y_3, z_1) , and
there is 1 value of x which is less than (y_3, z_2) .

Thus the experimental value of W_1 is

$$W_1 = 3 + 1 + 3 + 1 + 1 + 1 = 10,$$

$$\text{and } \hat{q} = \frac{10}{24}.$$

Also,

$$a_{11} = 1, a_{21} = 1, a_{31} = 0, a_{41} = 1, a_{12} = 1, a_{22} = 1, a_{32} = 1, \\ a_{42} = 1, a_{13} = 0, a_{23} = 1, a_{33} = 1, a_{43} = 0,$$

and

$$b_{11} = 1, b_{21} = 1, b_{31} = 0, b_{41} = 1, b_{12} = 0, b_{22} = 1, b_{32} = 0 \\ b_{42} = 0.$$

Thus,

$$\hat{q} = \frac{1}{10} \sum_{i,j,k} a_{ij} b_{ik} = \frac{10}{24}.$$

Next, suppose that

$$\phi^2 = \sigma^2 F(T^*) = \int_{-\infty}^{\infty} (F(t) - q)^2 dH^*(t), \text{ and} \\ v^2 = \sigma^2 H^*(x) = \int_{-\infty}^{\infty} (H^*(x) - 1 + q)^2 dF(x).$$

Then, we have the following relations.

$$q = \int_{-\infty}^{\infty} F(t) dH^*(t),$$

$$1 - q = \int_{-\infty}^{\infty} H^*(x) dF(x),$$

$$\int_{-\infty}^{\infty} F^2(t) dH^*(t) = q^2 + \phi^2, \text{ and}$$

$$\int_{-\infty}^{\infty} (1 - H^*(x))^2 dF(x) = v^2 + q^2.$$

Thus, we have the following results.

Theorem 4.5

The variance of W_1 is given by

$$\text{Var}(W_1) = \lambda_{mn}[(mn-1)v^2 + (\delta-1)(m+n-1)\phi^2 + (m+n-1)q$$

$$+ \{(\delta-1)(m+n-2) - 1\}q^2].$$

Lemma 4.8

$$\phi^2 \leq q(1-q), \text{ and}$$

$$v^2 \leq q(1-q).$$

Theorem 4.6

$\hat{q}$ is consistent estimate of q .

Proof

$$\frac{1}{\lambda_{mn}} \text{Var}(W_1) = \frac{1}{\lambda_{mn}} \text{Var}(W_1)$$

$$= \frac{1}{\lambda_{mn}} [(mn-1)v^2 + (\delta-1)(m+n-1)\phi^2 + (m+n-1)q$$

$$+ \{(\delta-1)(m+n-2) - 1\}q^2]$$

$$\leq q \left(\frac{m}{1} + \frac{n}{1} - \frac{mn}{1} + \frac{\gamma}{1} - \frac{\gamma_{mn}}{1} \right) + q^2 \left(\frac{\gamma}{1} + \frac{mn}{1} - \frac{\gamma_{mn}}{1} \right)$$

(using Lemma 4.8)

$\rightarrow 0$ as $\lambda, m, n \rightarrow \infty$.

The above results can also be generalized.

The statistic W based on ranks of the pooled sample was suggested to estimate p unbiasedly and was shown to have certain optimum properties. Our knowledge except when σ^2 is known.

then the m.v.u. estimate of $F(x)$ has not yet been found to the best of μ , and variance σ^2 . More generally, if x is a function of μ and σ^2 , and F is cumulative distribution function of a normal variate with mean lies in finding the m.v.u. estimate of $F(c\sigma^2)$, where c is a constant, we could not derive the minimum variance unbiased estimate. The problem in Theorem 4.3, if σ is also unknown, which is the more realistic case, minimum variance unbiased estimates remain an open question. For example, unbiased estimates, whereas in other cases the derivation of the minimum variance unbiased estimates, it was possible to derive the minimum variance

in a very few cases, it was possible to derive the minimum variance estimation were considered. q , as defined in Chapter IV. Parametric and nonparametric methods of estimation were considered. This probability has been called p , and its dual probability, less than a random variable, where all these random variables are mutually the estimate of the probability that N random variables are simultaneously resistances over a time interval $[0, T]$, i.e., the problem is to provide the reliability of a component subject to several different stresses or The major goal of this dissertation is to discuss estimation of

CONCLUSION AND FURTHER RESEARCH

CHAPTER V

ties. It is felt that the statistic W is of much importance, because only the assumption of the continuity of the cumulative distribution function is needed. However, the results could be extended to the discontinuous case with slight modifications. It is not easy to compute W especially when the number of observations available on each random variable is large, even in the case of N (number of random variables) = 2. Though the use of the statistic W has been limited only to estimate p in this dissertation, it could be used to test the hypothesis that the random variables $X_1, \dots, X_N, Y$ have the same continuous cumulative distribution function against the alternative that Y is stochastically larger than the N random variables $X_1, \dots, X_N$ simultaneously, i.e., Y is stochastically larger than the maximum of $X_1, \dots, X_N$. This particular use of the statistic W and possibly some other uses are open for further research. We are presently investigating this particular use of the W statistic for the special case $N = 2$, and as a critical region for the hypothesis, it is proposed to use $W \leq K$, where K is chosen to give the correct significance level. Properties of this test function, the distribution of the statistic W under the null hypothesis, approximation to the normality, a comparison with Whitney's [17] result, etc. are under consideration. Properties of this test function for the general case and other uses of the statistic W are yet to be explored.

1. Birnbaum, Z. W. (1956), "On a Use of the Mann-Whitney Statistic," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, (University of California Press, Berkeley, California), Vol. 1, pp. 13-17.
2. Birnbaum, Z. W. and McCarty, R. C. (1958), "A Distribution Free Upper Bound for $\Pr(Y < X)$ Based on Independent Samples of X and Y," Ann. Math. Stat., Vol. 29, pp. 558-562.
3. Church, J. D. and Bernard Harris (1970), "The Estimation of Reliability from Stress-Strength Relationship," Technometrics, Vol. 12, pp. 49-54.
4. Vandantzig, D. (1951), "On the Consistency and Power of the Wilcoxon's Two-Sample Test," Konink. Neder. Akad. Van Wetensch. Proc. Ser. A, Vol. 54, pp. 1-9.
5. Govindarajulu, Z. (1968), "Distribution-Free Confidence Bounds for $P(X < Y)$," Annals of the Institute of Statistical Mathematics, Vol. 20, pp. 229-238.
6. Gupta, S. S. (1963), "Probability Integrals of Multivariate Normal and Multivariate t," Ann. Math. Stat., Vol. 34, pp. 792-828.
7. Lehmann, E., Notes on the Theory of Estimation, Mimeo notes (University of California Press, Berkeley) (Reprinted 1962).
8. Lieberman, G. J. and Resnikoff, G. J. (1955), "Sampling Plans for Inspection by Variables," Journal of the American Statistical Association, Vol. 50, pp. 457-516.
9. Mann, H. B. and Whitney, D. R. (1947), "On a Test Whether One of Two Random Variables is Stochastically Larger Than the Other," Ann. Math. Stat., Vol. 18, pp. 50-60.
10. Mazumdar, M. (1970), "Some Estimates of Reliability Using Interference Theory," Naval Research Logistics Quarterly, Vol. 17, pp. 159-166.
11. Owen, D. B. (1957), "The Bivariate Normal Probability Distribution," Research Report, SC-3831 (TR) Systems Analysis, Sandia Corporation.

REFERENCES

12. Owen, D. B., K. J. Craswell and D. L. Hanson (1964), "Nonparametric Upper Confidence Bounds for $P(Y < X)$ and Confidence Limits for $P(Y < X)$ when X and Y are Normal," Journal of the American Statistical Association, Vol. 59, pp. 906-924.
13. Patil, G. P. and Want, J. K. (1966), "Minimum Variance Unbiased Estimation of the Distribution Function Admitting a Sufficient Statistic," Ann. Inst. Stat. Math., Vol. 18, pp. 39-47.
14. Rao, C. R. (1965), Linear Statistical Inference and its Applications, John Wiley and Sons, Inc.
15. Sen, P. K. (1960), "On Some Convergence properties of U-Statistic," Calcutta Stat. Bull., Vol. 10, pp. 1-18.
16. Sen, P. K. (1967), "A Note on Asymptotically Distribution-Free Confidence Bounds for $P(X < Y)$ Based on Two Independent Samples," Sankhya, Series A, Vol. 29, pp. 95-102.
17. Whitney, D. R. (1951), "A Bivariate Extension of the U Statistic," Ann. Math. Stat., Vol. 22, pp. 274-282.
18. Zehna, P. W. (1966), "Invariance of Maximum Likelihood Estimation," Ann. Math. Stat., Vol. 37, p. 755.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

UNCLASSIFIED

3. REPORT TITLE

On estimating the reliability of a component subject to several different stresses.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Satish Chandra

6. REPORT DATE

November 22, 1971

8a. CONTRACT OR GRANT NO.

N00014-68-A-0515

b. PROJECT NO.

NR 042-260

c.

d.

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

OFFICE OF NAVAL RESEARCH

13. ABSTRACT

A great deal has been written concerning the estimation of the probability and testing of whether one of two random variables is stochastically larger than the other and its relationship to the estimation of reliability for stress-strength relationships. A more general problem is the estimation and testing of whether one of $N + 1$ random variables is simultaneously stochastically larger (smaller) than the others. An initial paper which deals with this problem for the special case $N = 2$ is that of D. R. Whitney (1951), "A Bivariate Extension of the U Statistic," where he provides a test function and discusses the asymptotic normality of the statistics proposed under the null hypothesis that all the random variables have the same distribution function.

In this dissertation, the problem of estimation of the probability of whether one of $N + 1$ mutually independent random variables, each having a continuous cumulative distribution function, is simultaneously stochastically larger (smaller) than the others has been considered. Parametric and nonparametric methods of estimation are discussed. Applications of the problem involve the estimation of reliability from stress-strength relationships, where a component is subject to several stresses (several strengths) whereas its strength (stress) is a single random variable.

Minimum variance unbiased estimates are provided in certain cases, whereas in some cases the maximum likelihood estimates are given. For a more general situation, where only continuity of the cumulative distribution functions of $N + 1$ random variables is assumed, a statistic W , somewhat similar to the Mann-Whitney U statistic, based on ranks of the pooled sample values from all the random variables, has been suggested and shown to have certain optimum properties; i.e., it's an unbiased & consistent est.