Journal of the KSQC
Vol. 14, No. 1, 1986

A Study on Estimators of $P_r(X_1 < Y < X_2)$

Jae Joo, Kim*
Seong Yeon, Kim**

ABSTRACT

In this paper the minimum variance unbiased, maximum likelihood and empirical estimators of the probability $Pr(X_1 < Y < X_2)$ are obtained, where X_1, X_2 and Y are mutually independent exponential random variables. Comparison of estimators is discussed in the last section for illustraiton.

I. INTRODUCTION

In recent years a fair amount work is available on the problem of estimating $Pr[Y<X]$ in both distribution-free and parametric framworks where X and Y are independent random variables. In most of the work available to date, the estimators of the probability $Pr(Y<X)$ for various distributions are obtained (see Mazumdar (1970), Church and Harris (1970), Downton (1973)). Also some distribution-free estimators and confidence bounds of $Pr(X<Y)$ are obtained (see Birnbaum and MacCarty (1958), Govindarajulu (1968), Owen, Craswell and Hanson (1964) and Enis and Geisser (1971)). However, there are no attempts but Singh's (1980) that have been made to estimate the probability

$$P = Pr(X_1 < Y < X_2)$$

where X_1 and X_2 are independent random stress variables and Y independent of X_1 and X_2 is a random strength variable. The probability P can be interpreted as the probability of the strength of a component falling in between two random tresses to which the component may be expected to be subjected during a specified time interval $(0, t)$. Singh derived the estimators of $Pr(X_1<Y<X_2)$ in the normal case.

In this paper, we obtain the minimum variance unbiased (MVU), maximum likelihood (ML) and empirical estimators of $P_r(X_1<Y<X_2)$ and Compare them in the one parameter exponential case.

In section 2, the estimators are obtained when X_1 and X_2 have known parameters.

In section 3, the MVU and ML estimators are considered when all parameters are unknown.

In section 4, we consider the practical use of results and carried out the comparisons among the estimators in various settings of θ_1, θ_2, θ_Y and n when X_1 and X_2 have known parameters.

* Department of Computer Science & Statistics, Seoul National University.

** This work was partially supported by the Ministry of Education, through Research Institute of Basic Sciences, Seoul National University 1985-86.

II. ESTIMATION WHEN X_1 AND X_2 HAVE KNOWN PARAMETERS

Let X_1 and X_2 be two independent random stress variables distributed as

(2.1) $$f_{X_i}(x_i \mid \theta_i) = \theta_i \exp(-\theta_i x_i), \quad \theta_i > 0, \; x_i > 0, \; i = 1, 2$$

where θ_1 and θ_2 are known and let Y independent of X_1 and X_2 be a random strength variable distributed as

(2.2) $$f_Y(y \mid \theta_Y) = \theta_y \exp(-\theta_Y y), \; \theta_Y > 0, \; y > 0$$

where θ_Y is assumed to be unknown.

Further, let

(2.3) $$P_1 = Pr(X_1 < Y)$$

(2.4) $$P_2 = Pr(X_1 < Y, X_2 < Y)$$

(2.5) $$P = Pr(X_1 < Y < X_2)$$

(2.6) Proposition 2.1. $P = P_1 - P_2$.

Proof :

It can be easily proved as follows.

$$P = Pr(X_1 < Y < X_2) = \int_{-\infty}^{\infty} Pr(X_1 < y, X_2 > y \mid Y = y) dF_y(y)$$

$$= \int_{-\infty}^{\infty} Pr(X_1 < y)\, Pr(X_2 > y)\, dF_y(y) = \int_{-\infty}^{\infty} Pr(X_1 < y)(1 - Pr(X_2 < y))\, dF_y(y)$$

$$= \int_{-\infty}^{\infty} Pr(X_1 < y) dF_Y(y) - \int_{-\infty}^{\infty} Pr(X_1 < y, X_2 < Y) dF_Y(y)$$

$$= Pr(X_1 < Y) - Pr(X_1 < Y, X_2 < Y)$$

where $F_Y(y)$ is cumulative distribution function of Y.

The true value of P which we want to estimate, is given in

Proposition 2.2.

Proposition 2.2 When X_1, X_2 and Y are distributed independently as (2.1) and (2.2) respectively,

(2.7) $$P = \frac{\theta_1 \theta_Y}{(\theta_2 + \theta_Y)(\theta_1 + \theta_2 + \theta_Y)}.$$

Proof :

It can be easily verified.

(2.8) $$P = \frac{\theta_1}{\theta_1 + \theta_Y}$$

and

(2.9) $$P_2 = \frac{\theta_1\theta_2}{\theta_1+\theta_2+\theta_Y}\left(\frac{1}{\theta_2+\theta_Y}+\frac{1}{\theta_1+\theta_Y}\right)$$

(see appendix A). It follows from the relation (2.6) that

$$P = \frac{\theta_1\,\theta_Y}{(\theta_2+\theta_Y)(\theta_1+\theta_2+\theta_Y)}\,.$$

From Proposition 2.2, we can obtain the ML estimator of P.

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample of observation from $f_Y(y\,|\,\theta_Y)$.

If θ_1 and θ_2 are known and θ_Y unknown, then the ML estimator of P, $\hat{P}$ is given by

(2.10) $$\hat{P} = \frac{\theta_1\,\hat{\theta}_Y}{(\theta_2+\hat{\theta}_Y)(\theta_1+\theta_2+\hat{\theta}_Y)}$$

where $\hat{\theta}_Y = n/\sum_{j=1}^{n} Y_j$ this ML estimator is biased (see Appendix C).

Now we state the following well-known result in order to derive MVU estimator of P.

Proposition 2.3. If T and S are MVU estimators of parametric function $\alpha(\theta)$ and $\beta(\theta)$ respectively, then $aT+bS$ is a MVU estimator of $a\alpha(\theta)+b\beta(\theta)$, where a and b are specified constants (see Singh (1980)).

Let $\tilde{P}_1$ and $\tilde{P}_2$ are MVU estimators of P_1 and P_2, respectively.

Then

(2.11) $$\tilde{P}_1 = 1 + \frac{(n-1)!(-1)^n\exp(-\theta_1 t)}{(\theta_1 t)^{n-1}} + (n-1)!\sum_{k=1}^{n-1}\frac{(-1)^k(\theta_1 t)^{-k}}{(n-k-1)!}$$

and

(2.12) $$\tilde{P}_2 = 1+(n-1)!(-1)^{n_t-n+1}\{\exp(-\theta_1 t)/\theta_1^{n-1}$$

$$+\exp(\theta_2 t)/\theta_2^{n-1} - \exp(-(\theta_1+\theta_2)t)/(\theta_1+\theta_2)^{n-1}\}$$

$$+(n-1)!\sum_{k=1}^{n-1}(-t)^{-k}\{\theta_1^{-k}+\theta_2^{-k}-(\theta_1+\theta_2)^{-k}\}/(n-k-1)!$$

where $t=\sum_{j=1}^{n} Y_j$ (see Appendix B). It follows from Proposition 2.1 and Proposition 2.3 that the MVU estimator of P, $\tilde{P}$ can be obtained. That is

(2.13) $$\tilde{P} = (n-1)!(-1)^{n_t-n+1}\{\exp(-(\theta_1+\theta_2)t)/(\theta_1+\theta_2)^{n-1} - \exp(-\theta_2 t)/\theta_2^{n-1}\}$$

$$+(n-1)!\sum_{k=1}^{n-1}(-t)^{-k}\{(\theta_1+\theta_2)^{-k} - \bar{\theta}_2^{k}\}/(n-k-1)!$$

Now let us consider the empirical estimator of P and its variance.

Let X_1 and X_2 be two independent random stree variables with known cumulative distribution functions (cdf) $H_{x_1}(x)$ and $G_{x_2}(x)$ respectively and let Y independent of X_1 and X_2 be a random strength with cdf $F_Y(y)$ which is assumed to be unknown.

We note from relation (2.6) that

$$P=Pr(X_1<Y)-Pr(X_1<Y, X_2<Y) = E_Y(H_{X_1}(Y))$$

$$-E_Y(H_{X_1}(Y)GX_2(Y))=E_Y(H_{X_1}(Y)(1-G_{X_2}(Y))).$$

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample of observations from $F_Y(Y)$ Then an empirical estimator of P is given by

$$\bar{P}=\frac{1}{n}\sum_{j=1}^{n} H_{X_1}(Y_j)(1-G_{X_2}(Y_j)). \tag{2.14}$$

It can easily be verified that $\bar{P}$ is unbiased. Thus $\tilde{P}$ has the smaller variance than $\bar{P}$.

Here, we can consider the relation and bound of the veriance of $\bar{P}$ and $\tilde{P}$.

Proposition 2.4. Let d_L be the Cramér-Rao lower bound of $\tilde{P}$. Then

$$d_L=n^{-1}\left\{\frac{\theta_1\theta_Y(\theta_Y^2-\theta_1\theta_2)}{(\theta_2+\theta_Y)^2(\theta_1+\theta_2+\theta_Y)^2}\right\}^2 = n^{-1}\left\{\frac{p^2(\theta_Y^2-\theta_1\theta_2)}{\theta_1\theta_2}\right\}^2 \tag{2.15}$$

Proof:

Since P satisfied the reqular conditions of Cramer-Rao lower bound, the Fisher information number, $I(\theta_Y)$ is $1/\theta_Y^2$ and $E_{\theta Y}(\tilde{P})$ is given by (2.7), we can obtain the above result.

Proposition 2.5. Let $d_U = 1/4n$ for fixed n.

Then $\text{Var } \bar{P} \le d_U$.

Proof:

$$\text{Var}_{\theta_Y}(\bar{P}) = n^{-1}\text{var}_{\theta_Y}(H_{X_1}(Y)(1-G_{X_2}(Y))) \le n^{-1}(E(H_{X_1}(Y)(1-G_{X_2}(Y)))-P^2)$$

$$=n^{-1}P(1-P) \le 1/4n=d_U.$$

From propositions 2.4, 2.5 and the properties of MVU estimator we have following important results:

$$d_L \le \text{Var}_{\theta_Y}(P) \le \text{Var}_{\theta_Y}(\bar{P}) \le d_U.$$

III. ESTIMATION WHEN ALL PARAMETERS ARE UNKNOWN

Let X_1 and X_2 be two independent random stress variables distributed as

$$f_{X_1}(x_i\,|\,\theta_i)=\theta_i\exp(-\theta_i x_i),\ \theta_i>0,\ x_i>0,\ i=1,2 \tag{3.1}$$

where θ_1 and θ_2 are unknown and let Y be a random strength with same distribution function as in section 2.

And let $X_{11}, X_{12}, \cdots\cdots, X_{1n_1}$ and $X_{21}, X_{22}, \cdots, X_{2n_2}$ and $Y_1, Y_2, \cdots, Y_{n_3}$ be random samples of obserations from

$f_{X_1}(x_1|\theta_1)$, $f_{X_2}(x_2|\theta_2)$ and $f_Y(y|\theta_Y)$, respectively.

From the relation (2.7), the ML estimator of P, $\hat{\hat{P}}$, is obtained, i.e.,

$$(3.2) \qquad \hat{\hat{P}} = \frac{\hat{\theta}_1 \hat{\theta}_Y}{(\hat{\theta}_1 + \hat{\theta}_Y)(\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_Y)}$$

where $\hat{\theta}_i = n_i / \sum_{j=1}^{ni} X_{ij}, \; i=1,2,$

and $\hat{\theta}_Y = n_3 / \sum_{j=1}^{n3} Y_i.$

Finally, we shall obtain the MVU estimator of P.

Let $\tilde{\tilde{P}}_1$ and $\tilde{\tilde{P}}_2$ are MVU estimators of P_1 and P_2 and let $t_i = \sum_{j=1}^{ni} X_{ij}, \; i=1,2.$

$t_3 = \sum_{j=1}^{n3} Y_j.$ Then we can obtain

$$(3.3) \qquad \tilde{\tilde{P}}_1 = \int_0^{t_3/t_1} \int_0^{\min(1, 1/u_1)} g(u_1, u_2)\, du_1\, du_2$$

where $g(u_1, u_2) = (n_1 - 1)(n_3 - 1)(1 - u_1 u_2)^{n_1 - 2}(1 - u_2)^{n_3 - 2} u_2$

$$(3.4) \qquad \tilde{\tilde{P}}_2 = \int_0^{t_3/t_1} \int_0^{t_3/t_2} \int_0^{\min(1, 1/v_1, 1/v_2)} h(v_1, v_2, v_3)\, dv_3 dv_2 dv_1$$

where $h(v_1, v_2, v_3) = (n_1 - 1)(n_2 - 1)(n_3 - 1)(1 - v_1 v_3)^{n_1 - 2} \; x(1 - v_2 v_3)^{n_2 - 2}(1 - v_3)^{n_3 - 2} v_3^2$

(see Appendix D).

From Proposition 2.1 and Proposition 2.3, a MVU estimator of P, $\tilde{\tilde{P}}$, is given by

$$\tilde{\tilde{P}} = \tilde{\tilde{P}}_1 - \tilde{\tilde{P}}_2.$$

IV. COMPARISON IN MODERATE SAMPLES

Unfortunately, the expected values of $\hat{P}$ and $\tilde{P}$ can not be obtained in the closed form because they involve a indefinite integration $\int_a^{\infty} x^{-1} e^{-bx} dx. \; a \geq 0, \; b > 0.$ And so we can not obtain the variances of $\hat{P}$ and $\tilde{P}$ in the closed form (see Appendix C).

If $d\upsilon$ of Proposition 2.5. is not appropriate, how can we use the results in section 2?

Note that $Pr(X_1 < Y < X_2)$ is determined not by the magnitude but by the ratio of θ_1, θ_2 and θ_Y, the extention and application of the following table would yield asymtotic variances and confidence intervales.

In Table 4.1, MSE is mean square error, m, number of samples and n, sample size.

Because of the truncation effects of computer CPU despite the declaration of DOUBLE PRECISION, MVU estimator can not have the better efficiency than ML estimator.

We notice from Table 4.1 that ML estimator has no serious bias and similar efficiency with MVU estimator and that ML and MVU estimators have much better efficiencies than empirical estimator.

Table 1. Estimates of Bias and MSE : m = 500

n	θ_1	θ_2	θ_Y	P	Bias $\hat{P}$	Bias $\tilde{P}$	Bias $\bar{P}$	MSE $\hat{P}$	MSE $\tilde{P}$	MSE $\hat{P}$
6	1.0	3.0	2.0	.0667	−.0011	*	.0006	.000044	.000067	.000192
	1.0	2.0	3.0	.1000	−.0048	*	.0002	.000086	.000069	.000288
	3.0	1.0	1.0	.3000	−.0029	−.0007	.0027	.000922	.001400	.003236
	8.0	1.0	1.0	.4000	.0025	−.0027	.0030	.003566	.004678	.007425
	100.0	1.0	1.0	.4950	. .0077	−.0112	−.0150	.008767	.009653	.013537
10	1.0	3.0	2.0	.667	−.0007	*	.0010	.000023	.000032	.000101
	1.0	2.0	3.0	.1000	−.0029	*	.0002	.000039	.000031	.000194
	3.0	1.0	1.0	.3000	−.0003	.0007	.0001	.000614	.00792	.002134
	8.0	1.0	1.0	.4000	.0010	−.0028	−.0048	.002168	.002535	.004065
	100.0	1.0	1.0	.4950	−.0023	−.0137	−.0136	.005846	.006303	.008360
20	1.0	3.0	2.0	.0667	−.0002	*	.002	.000013	.000015	.000045
	1.0	2.0	3.0	.100	−.0013	.0001	.0005	.000011	.000009	.000082
	3.0	1.0	1.0	.3000	−.0015	−.0012	.0024	.000378	.000428	.001132
	8.0	1.0	1.0	.4000	.0023	.0004	.0005	.001268	.001366	.002372
	100.0	1.0	1.0	.4950	.0026	−.0030	.0014	.002911	.002950	.003940
50	1.0	3.0	2.0	.0667	−.0001	.0001	*	.000006	.000006	.000019
	1.0	2.0	3.0	.1000	−.0005	*	.0002	.000003	.00003	.000038
	3.0	1.0	1.0	.3000	−.0003	−.0002	.0005	.000142	.000150	.000427
	8.0	1.0	1.0	.4000	.0003	−.0005	−.0018	.000517	.000323	.000907
	100.0	1.0	1.0	.4950	−.0027	−.0050	−.0048	.001026	.001034	.001504

* At least the first four digits are zero

Appendix A

Let X_1, X_2 and Y be distributed independently as (2.1) and (2.2), respectively. Then joint probaility density function (pdf) of X_1 and Y is

(A-1) $$f(x_1, y \mid \theta_1, \theta_Y) = \theta_1 \theta_Y \exp(-\theta_1 x_1 - \theta_Y y)$$

Let $T_1 = X_1 - Y$ $T_2 = Y$. Then the joint pdf of T_1 and T_2 is

(A-2) $$g(t_1, t_2 \mid \theta_1, \theta_Y) = \theta_1 \theta_Y \exp\{-\theta_1 t_1 - (\theta_1 + \theta_Y)\, t_2\} \; x I(-\infty < -t_2 < t_1 < \infty).$$

Thus the marginal pdf of T_1 is given by

(A-3) $$g_1(t_1 \mid \theta_1, \theta_Y) = \{\exp(\theta_Y t_1 I(-\infty, 0) + \exp(-\theta_1 t_1)\, I(0, \infty)\} \; x\, \theta_1 \theta_Y / (\theta_1 + \theta_Y).$$

Hence $$P_1 = Pr(X_1 < Y) = Pr(X_1 - Y < 0) = Pr(T_1 < 0)$$

$$= \int_{-\infty}^{0} g_1(t_1 | \theta_1, \theta_Y)\, dt_1 = \theta_1 / (\theta_1 + \theta_Y).$$

Similarly, we can obtain P_2.

Appendix B

The joint pdf of X_1 and $Y_1, Y_2, \cdots, Y_n$ is

(B-1) $$f(x_1, y_1, y_2, \cdots, y_n | \theta_1, \theta_Y) = \theta_1 \theta_Y^n \exp\left(-\theta_1 x_1 - \theta_Y \sum_{i=1}^{n} y_i\right).$$

This joint pdf satisfies the regular conditions of exponential class and has $T = \sum_{j=1}^{n} Y_j$ as complete sufficient statistic. Thus MVU estimator of P_1 is trivial expectation of $I(X_1 < Y_1)$ given $T = t$. Hence

(B-2) $$\tilde{P}_1 = E\{I(X_1 < Y_1) | T = t\} = Pr(X_1 < Y_1 | T = t) = Pr\left(\frac{X_1}{Y_1 / T} < t\right).$$

Let $Y_1 / T = U$. Then $U \sim \beta\varepsilon(1, n-1)$. Thus

(B-3) $$g(x_1, u | \theta_1) = \theta_1 \exp(-\theta_1 x_1)\,(n-1)(I-u)^{n-2}.$$

Let $X_1 / U = V_1$ and $U = V_2$. Then the joint pdf of V_1 and V_2 is given by

(B-4) $$h(v_1, v_2 | \theta_1) = \theta_1 \exp(-\theta_1 v_1 v_2)(n-1)(1-v_2)^{n-2} v_2.$$

Hence we can obtain $\tilde{P}_1$ as (2.11).

Similarly, we can obtain $\tilde{P}_2$.

Appendix C

For example, we shall obtain $E_{\theta_Y}(P)$.

(C-1) $$E_{\theta_Y}(\hat{P}) = E_{\theta_Y}\left(\frac{\theta_1 \hat{\theta}_Y}{(\theta_2 + \theta_Y)(\theta_1 + \theta_2 + \theta_Y)}\right) = E_{\theta_Y}\left(\frac{\hat{\theta}_Y}{\theta_2 + \hat{\theta}_Y}\right) - E_{\theta_Y}\left(\frac{\hat{\theta}_Y}{\theta_1 + \theta_2 + \theta_Y}\right).$$

Here

(C-2) $$E_{\theta_Y}\left(\frac{\hat{\theta}_Y}{\theta_2 + \hat{\theta}_Y}\right) = E\left(\frac{n/\theta_2}{t + n/\theta_2}\right) \quad \text{where} \quad t \sim \Gamma(n, \theta_Y)$$

$$= \int_0^{\infty} \frac{n/\theta_2}{t + n/\theta_2}\, \theta_Y{}^{n_t n-1} \exp(-\theta_Y t) dt$$

$$= a \int_a^{\infty} \theta_Y^n Z^{-1} (Z-a)^{n-1} \exp(-\theta_y (Z-a)) dZ$$

where $a=n/\theta_2$.

Thus

(C-3) $$E_{\theta_Y}\left(\frac{\hat{\theta}_Y}{\theta_2+\hat{\theta}_Y}\right)=-(-a)^n\theta_Y^n \exp(\theta_Y a)\int_a^\infty Z^{-1}\exp(-\theta_Y Z)\,dZ$$

$$+a\theta_Y^n\sum_{k=0}^{n-2}(-a)^k\,\Gamma(k+1,\theta_Y a)/\theta_Y^{k+1}.$$

The right hand side term in the above equation, can be calculated.

Appendix D

A tribial unbiased estimator of P_1 is $I_{(X_{11}<Y_1)}$, and complete sufficient statistics is T_1 and T_2, where $T_1=\sum_{j=1}^{n1}X_{1j}$ and $T_2=\sum_{j=1}^{n3}Y_j$. We know the properties of the joint pdf as follows. The joint pdf of X_{11}/T_1 and Y_1/T_2 does not depend on θ_1 and θ_Y and T_1 and T_2. Thus

(D-1) $$\tilde{\tilde{P}}_1=E(I_{(X_{11}<Y_1)}\,|T_1=t_1,\,T_2=t_2)=Pr\,(X_{11}<Y_1\,|T_1=t_1,\,T_2=t_2)$$

$$=Pr\left(\frac{X_{11}/T_1}{Y_1/T_2}<\frac{t_2}{t_1}\right)$$

where $Z_1=X_{11}/T_1\sim\beta\varepsilon(1,x_1-1)$,

$Z_2=Y_1/T_2\sim\beta\varepsilon(1,n_2-1)$,

and Z_1 and Z_2 are independent.

A little more handling leads us to obtain $\tilde{\tilde{P}}_1$. That is

(D-2) $$P_1=\int_0^{t_3/t_1}\int_0^{\min(1,1/u_1)}g(u_1,u_2)\,du_2du_1$$

where

(D-3) $$g(u_1,u_2)=(n_1-1)(n_3-1)(1-u_1u_2)^{n1-2}\,(1-u_2)^{n3-2}\,u_2.$$

Similarly, we can obtain $\tilde{\tilde{P}}_2$.

REFERENCES

1. Birnbaum, Z.W. and MacCarty, R.C. (1957), "*A distribution-free upper confidence bound for* $Pr(Y<X)$, *based on independent samples of X and Y*," Annals of Math. Statit., 28, 558-562.
2. Church, J.D. and Harris, B. (1970), "*The Estimation of Reliability from Stress-Strength Relationship,*" Tech., 12, 49-54.
3. Downton, F. (1973), "*The estimation of* $Pr(Y<X)$ *in the normal case,*" Tech., 15, 551-558.
4. Enis, P. and Geisser, S. (1971), "*Estimation of the Reliability that* $Y < X$," J. Amer. Statist. Assoc., 66, 162-168.
5. Govindarajulu, Z. (1968), "*Distribution-free Confidence bounds for* $P(X<Y)$," Annals of the Institute of Statist. Math., 20, 229-238.
6. Mazumdar, M. (1970), "*Some estimates of reliability using interference theory,*" Nav. Res. Log. Quart., 17, 159-166.
7. Owen, D.B., Craswell, B.K.J. and Hanson, D.L. (1964), "*Nonparametric confidence bounds for* $Pr(Y<X)$ *and Confidence limits for* $Pr(Y<X)$ *when X and Y are normal,*" J. Amer. Statist. Assoc., 59, 906-924.
8. Singh, N. (1980), "*On the estimation of* $Pr(X_1<Y<X_2)$, Commun. Statist,- Theor. Math., A9 (15), 1551-1561.