ESTIMATION OF JUMPS, RELIABILITY AND HAZARD RATE

By V. K. Murthy

Douglas Aircraft Company

- **0.** Summary. Let F(x) be a probability distribution function. Assuming the singular part to be identically zero, it is well known (see e.g. Cramér [1] pp. 52, 53) that F(x) can be decomposed into $F(x) = F_1(x) + F_2(x)$ where $F_1(x)$ is an everywhere continuous function and $F_2(x)$ is a pure step function with steps of magnitude, say, S_{ν} at the points $x = x_{\nu}$, $\nu = 1, 2, \dots, \infty$ and that finally both $F_1(x)$ and $F_2(x)$ are non-decreasing and are uniquely determined. In this paper the problem of estimating the jump S_i corresponding to the saltus $x = x_i$ is considered. Also considered are the problems of estimation of reliability and hazard rate. Based on a random sample X_1, X_2, \dots, X_n of size n from the distribution F(x), consistent and asymptotically normal classes of estimators are obtained for estimating the jump S_i corresponding to the saltus $x = x_i$. Based on the earlier work of the author [2] on estimation of probability density, consistent and asymptotically normal estimates are obtained for the reliability and hazard rate.
- **1.** Introduction. Let $X_1, X_2, \dots X_n$ be a random sample of size n from the distribution F(x), i.e. $X_1, X_2, \dots X_n$ are independently, identically distributed random variables with the same distribution function F(x). In the particular case when the random variable is time to failure of an item, F(x) is the probability of the event that by time x the item has failed and R(x) = 1 F(x) is the probability of the complementary event that the item survived time instant x and is the so-called reliability of the item. In what follows, for any random variable with distribution function F(x), we call R(x) = 1 F(x) the reliability function. If x is any point of continuity of the distribution F(x) and if the density at x is denoted by f(x), the function Z(x) = f(x)/[1 F(x)] will be referred to as the hazard rate.
- 2. The asymptotic equivalence of an estimate and a class of estimators for the reliability at a point of continuity of F(x). Let

$$F_n(t) = (1/n)[\text{number of observations} \leq t \text{ among } X_1, X_2, \dots X_n]$$

and

(2.1) $R_n(t) = (1/n)$ [number of observations > t among X_1 , X_2 , \cdots X_n]. Clearly $R_n(t)$ is a binomially distributed random variable with

(2.2)
$$E(R_n(t)) = R(t).$$

$$Var(R_n(t)) = (1/n)R(t)(1 - R(t)).$$

Received 12 December 1963; revised 20 December 1964.

Let K(x) be a function satisfying

(2.3)
$$K(x) \ge 0$$
, $K(-x) = K(x)$, $\lim_{|x| \to \infty} xK(x) = 0$, $\int_{-\infty}^{\infty} K(x) dx = 1$.

A function K(x) satisfying (2.3) is called a Window. (Murthy [2]). Let B_n be a sequence of non-negative constants depending on the sample size n such that $B_n \to \infty$ as $n \to \infty$.

Let

$$(2.4) \quad f_n(t) = \int_{-\infty}^{\infty} B_n K(B_n(x-t)) \ dF_n(x) = (B_n/n) \sum_{j=1}^n K(B_n(X_j-t)).$$

It was shown by the author [2] that the class of estimators $\{f_n(x_0)\}$ given by (2.4) consistently and asymptotically normally estimate the density $f(x_0)$ at every point of continuity x_0 of the distribution F(x) and also of f(x) if $\sum_i S_i/|x_i-x_0| < \infty$. We will now propose the class of estimators $\{R_n^*(t)\}$ for estimating the reliability function R(t) where

$$(2.5) \quad R_n^*(t) = \int_t^\infty f_n(x) \, dx = (B_n/n) \sum_{i=1}^n \int_t^\infty K(B_n(X_i - x)) \, dx.$$

We will now prove that at a point of continuity t of the distribution F(t)

$$\lim_{n\to\infty} E(R_n^*(t)) = R(t),$$

and

(2.7)
$$\lim_{n\to\infty} [n \operatorname{Var}(R_n^*(t))] = R(t)(1-R(t)).$$

Let

$$(2.8) G(t) = \int_{-\infty}^{t} K(x) dx.$$

In terms of G(t), $R_n^*(t)$ can be written as

(2.9)
$$R_n^*(t) = (1/n) \sum_{j=1}^n G(B_n(X_j - t)).$$

Taking expectation on both sides of (2.9) we obtain

(2.10)
$$E(R_n^*(t)) = \int_{-\infty}^{\infty} G(B_n(x-t)) dF(x) \\ = 1 - \int_{-\infty}^{\infty} B_n K(B_n(x-t)) F(x) dx.$$

Now

$$(2.11) \quad \int_{-\infty}^{\infty} B_n K(B_n(x-t)) F(x) \ dx = \int_{-\infty}^{\infty} K(\lambda) F(t+\lambda/B_n) \ d\lambda.$$

If t is a point of continuity of the distribution F(x), taking limit on both sides of (2.11) as $n \to \infty$ we have

$$(2.12) \quad \lim_{n\to\infty} \int_{-\infty}^{\infty} B_n K(B_n(x-t)) F(x) \ dx = F(t) \int_{-\infty}^{\infty} K(\lambda) \ d\lambda = F(t).$$

Combining (2.10) and (2.12) we have at a point of continuity t of the distribution F(x) that

(2.13)
$$\lim_{n\to\infty} E(R_n^*(t)) = 1 - F(t) = R(t).$$

Taking the variance of the estimator $R_n^*(t)$ given by (2.9) we obtain

(2.14) Var
$$(R_n^*(t)) = (1/n)$$
 Var $[G(B_n(x-t))]$
= $(1/n)[E(G^2(B_n(x-t))) - E^2(G(B_n(x-t)))].$

Now

(2.15)
$$E(G^{2}(B_{n}(x-t))) = \int_{-\infty}^{\infty} G^{2}(B_{n}(x-t)) dF(x)$$
$$= 1 - 2 \int_{-\infty}^{\infty} G(B_{n}(x-t)) B_{n}K(B_{n}(x-t)) F(x) dx,$$

after integration by parts. Substituting $B_n(x-t) = \lambda$, (2.15) can be written as

$$(2.16) \quad E(G^2(B_n(x-t))) = 1 - 2 \int_{-\infty}^{\infty} G(\lambda)K(\lambda)F(t+\lambda/B_n) d\lambda.$$

Taking limit as $n \to \infty$ on both sides of (2.16) we have at a point of continuity t of the distribution F(x) that

(2.17)
$$\lim_{n\to\infty} E(G^2(B_n(x-t))) = 1 - 2F(t) \int_{-\infty}^{\infty} G(\lambda)K(\lambda) d\lambda$$
$$= 1 - F(t),$$

since $\int_{-\infty}^{\infty} G(\lambda)K(\lambda) d\lambda = \frac{1}{2}$. Combining (2.10), (2.13), (2.14) and (2.17) we discover that

$$(2.18) \quad \lim_{n\to\infty} [n \text{ Var } (R_n^*(t))] = R(t) - R^2(t) = R(t)(1-R(t)),$$

at every point of continuity of the distribution F(x). Also from (2.9) $R_n^*(t)$ can be written as $R_n^*(t) = (1/n) \sum_{j=1}^n V_j$, where $V_j = G(B_n(x_j - t))$ and the V_j 's are independently and identically distributed as a random variable

$$(2.19) y_n = G(B_n(X_j - t)).$$

A sufficient condition for the sequence $\{R_n^*(t)\}$ to be asymptotically normally distributed (see Parzen [3] p. 1069) is that for some $\delta > 0$

$$(2.20) E|y_n - E(y_n)|^{2+\delta}/\{n^{\delta/2}[Var(y_n)]^{1+\delta/2}\} \to 0; as n \to \infty.$$

For y_n given by (2.19) Condition (2.20) is easily verified by noting that both $\lim_{n\to\infty} E|y_n|^{2+\delta} < \infty$, and $\lim_{n\to\infty} \mathrm{Var}(y_n) < \infty$ at every point of continuity t of the distribution F(x). Summing up we have proved the following

THEOREM 1. The estimate $R_n(t)$ given by (2.1) and the class of estimators $\{R_n^*(t)\}$ given by (2.5) are both consistent estimates of R(t) at every point of continuity t of the distribution F(x). Further $R_n(t)$ and $\{R_n^*(t)\}$ are both asymptotically equivalent in the sense that they have the same order of consistency and the same asymptotic variance. The sequence $\{R_n^*(t)\}$ is asymptotically normal.

3. Estimation of the jump S_i at the saltus x_i of the distribution F(x). Assuming the singular part to be identically zero, the distribution F(x) can be decomposed into (see e.g. Cramér [1] pp. 52, 53)

$$(3.1) F(x) = F_1(x) + F_2(x),$$

where $F_1(x)$ is an everywhere continuous function and $F_2(x)$ is a pure step func-

tion with steps of magnitude, say, S_{ν} at the points $x = x_{\nu}$, $\nu = 1, 2, \dots$ and $F_1(x)$ and $F_2(x)$ are non-decreasing and are uniquely determined. Substituting (3.1) in (2.10) we obtain

(3.2)
$$E(R_n^*(t)) = \int_{-\infty}^{\infty} G(B_n(x-t)) dF_1(x) + \int_{-\infty}^{\infty} G(B_n(x-t)) dF_2(x)$$

= $I_1 + I_2$, say.

Following the argument of the previous section, we readily obtain $\lim_{n\to\infty} I_1 = F_1(\infty) - F_1(t)$. Since $F_1(x)$ is continuous at $x = x_i$ we have

(3.3)
$$\lim_{n\to\infty} I_1 = F_1(\infty) - F_1(x_i),$$

at the saltus $x = x_i$ of the distribution F(x).

Now

(3.4)
$$I_{2} = \int_{-\infty}^{\infty} G(B_{n}(x-t)) dF_{2}(x) = \sum_{\nu=1}^{\infty} S_{\nu}G(B_{n}(x_{\nu}-t)).$$

Denoting by $\sum_{x_{\nu}>x_i}$ summation over all ν such that $x_{\nu}>x_i$ and by $\sum_{x_{\nu}<x_i}$ summation over all ν such that $x_{\nu}< x_i$, at the saltus $t=x_i$ of the distribution F(x), I_2 can be written as

$$I_2 = I_{21} + I_{22} + I_{23}$$

where $I_{21} = \sum_{x_{\nu} < x_i} S_{\nu} G(B_n(x_{\nu} - x_i)), I_{22} = S_i G(0) = \frac{1}{2} S_i, \text{ and } I_{23} = \sum_{x_{\nu} > x_i} S_{\nu} G(B_n(x_{\nu} - x_i)).$ Now

$$I_{21} = \Sigma_1 + \Sigma_2$$

where

$$\Sigma_1 = \sum_{x_{\nu} < x_i, |\nu| \leq m} S_{\nu} G(B_n(x_{\nu} - x_i)),$$

and

$$\Sigma_2 = \sum_{x_{\nu} < x_i, |\nu| > m} S_{\nu} G(B_n(x_{\nu} - x_i)).$$

It can be argued as in the proof of the lemma (see Murthy [2]) that Σ_2 can be made arbitrarily small, by choosing m sufficiently large, (no matter what n is) and Σ_1 , for fixed m, can be made arbitrarily small by choosing n sufficiently large, i.e. $\lim_{n\to\infty} I_{21} = 0$. From the fact that

$$I_{23} = \sum_{x_{\nu} > x_i} S_{\nu} - \sum_{x_{\nu} > x_i} S_{\nu} [1 - G(B_n(x_{\nu} - x_i))],$$

we discover $\lim_{n\to\infty} I_{23} = \sum_{x_{\nu}>x_i} S_{\nu}$. Of course, it should be noted (see Murthy [2]) that in proving the above statement it is assumed that $\sum_{\nu\neq i} S_{\nu}/|x_{\nu}-x_{i}| < \infty$. We have therefore proved that at the saltus $t=x_{i}$ of the distribution F(x)

(3.5)
$$\lim_{n\to\infty} I_2 = \frac{1}{2} S_i + \sum_{x_{\nu} > x_i} S_{\nu}.$$

Combining (3.2), (3.3) and (3.5) we obtain

(3.6) $\lim_{n\to\infty} E(R_n^*(x_i)) = F_1(\infty) - F_1(x_i) + \frac{1}{2}S_i + \sum_{x_i>x_i} S_i$, at the saltus x_i of the distribution F(x).

Now

$$F(x_i) = \int_{-\infty}^{x_i} d(F_1(x) + F_2(x)) = F_1(x_i) + \sum_{x_i \le x_i} S_i,$$

and therefore

(3.7)
$$R(x_i) = 1 - F(x_i) = F_1(\infty) + F_2(\infty) - F_1(x_i) - \sum_{x_{\nu} \leq x_i} S_{\nu}$$
$$= F_1(\infty) - F_1(x_i) + \sum_{x_{\nu} > x_i} S_{\nu}.$$

Substituting from (3.7) in (3.6) we discover that

(3.8)
$$\lim_{n\to\infty} E(R_n^*(x_i)) = R(x_i) + \frac{1}{2}S_i.$$

From (2.2) we have that

$$(3.9) E(R_n(x_i)) = R(x_i).$$

Let us write

(3.10)
$$H_n(x_i) = 2[R_n^*(x_i) - R_n(x_i)].$$

In view of (3.8) and (3.9) we obtain

(3.11)
$$\lim_{n\to\infty} E(H_n(x_i)) = S_i,$$

at the saltus x_i of the distribution F(x).

4. Variance of the estimator $H_n(x_i)$. We have

(4.1)
$$\operatorname{Var} (H_n(x_i)) = 4 [\operatorname{Var} (R_n^*(x_i)) + \operatorname{Var} (R_n(x_i)) - 2 \operatorname{cov} (R_n^*(x_i), R_n(x_i))].$$

Since we already know $\operatorname{Var}(R_n(x_i))$ as given by (2.2), we only have to obtain $\operatorname{Var}(R_n^*(x_i))$ and $\operatorname{cov}(R_n^*(x_i), R_n(x_i))$ at the saltus $x = x_i$ of the distribution F(x). We have from (2.14) that

$$(4.2) \quad n \text{ Var } (R_n^*(t)) = E(G^2(B_n(x-t))) - E^2(G(B_n(x-t))).$$

Now

$$E(G^{2}(B_{n}(x-t))) = \int_{-\infty}^{\infty} G^{2}(B_{n}(x-t)) dF_{1}(x) + \int_{-\infty}^{\infty} G^{2}(B_{n}(x-t)) dF_{2}(x)$$

= $J_{1} + J_{2}$, say.

It is easily seen that

(4.3)
$$\lim_{n\to\infty} J_1 = F_1(\infty) - F_1(x_i),$$

$$\lim_{n\to\infty} J_2 = \frac{1}{4} S_i + \sum_{x_p > x_i} S_p,$$

at the saltus $t = x_i$ of F(t). Therefore

(4.4)
$$\lim_{n\to\infty} E[G^2(B_n(x-x_i))] = F_1(\infty) - F_1(x_i) + \frac{1}{4}S_i + \sum_{x_i>x_i} S_i$$

= $R(x_i) + \frac{1}{4}S_i$.

Combining (3.8), (4.2) and (4.4) we obtain

(4.5) $\lim_{n\to\infty} [n \operatorname{Var}(R_n^*(x_i))] = R(x_i) + \frac{1}{4}S_i - (R(x_i) + \frac{1}{2}S_i)^2$, at the saltus $t = x_i$ of the distribution F(t).

To find the co-variance between $R_n(t)$ and $R_n^*(t)$ let us recall

$$(4.6) R_n^*(t) = (1/n) \sum_{j=1}^n G(B_n(X_j - t)),$$

and

(4.7)
$$R_n(t) = (1/n)$$
 [number of observations > t among X_1 , X_2 , \cdots X_n]
$$= (1/n) \sum_{j=1}^n U(X_j - t),$$

where

$$U(x) = 1 for x > 0$$
$$= 0 for x \le 0.$$

Now

(4.8)
$$\operatorname{cov} [R_n^*(t), R_n(t)] = (1/n^2) \sum_{j=1}^n \operatorname{cov} [G(B_n(x_j - t)), U(X_j - t)]$$

= $(1/n) \operatorname{cov} [G(B_n(x - t)), U(x - t)]$

we have

(4.9)
$$\operatorname{cov} \left[G(B_n(x-t)), U(x-t) \right] = M_1 - M_2,$$

where

(4.10)
$$M_1 = \int_{-\infty}^{\infty} U(x-t)G(B_n(x-t)) d(F_1(x) + F_2(x)) = M_{11} + M_{12}$$
, say, and

(4.11)
$$M_2 = E(U(x-t))E(G(B_n(x-t))).$$

It can be easily verified that

(4.12)
$$\lim_{n\to\infty} M_{11} = \lim_{n\to\infty} \int_{-\infty}^{\infty} U(x-t)G(B_n(x-t)) dF_1(x)$$
$$= F_1(\infty) - F_1(x_i)$$

at the saltus $t = x_i$. Also

$$M_{12} = \int_{-\infty}^{\infty} U(x - t) G(B_n(x - t)) dF_2(x) = \sum_{x_{\nu} > x_i} S_{\nu} G(B_n(x_{\nu} - x_i)).$$

Hence

$$\lim_{n\to\infty} M_{12} = \sum_{x_{\nu}>x_i} S_{\nu},$$

at the saltus $t = x_i$. Summing up

$$(4.14) \qquad \lim_{n\to\infty} M_1 = F_1(\infty) - F_1(x_i) + \sum_{x_{\nu}>x_i} S_{\nu} = R(x_i),$$

at the saltus $t = x_i$. We have

(4.15)
$$E(U(x - x_i)) = \int_{-\infty}^{\infty} U(x - x_i) d(F_1(x) + F_2(x))$$

$$= \int_{x_i}^{\infty} dF_1(x) + \sum_{x_r > x_i} S_r$$

$$= F_1(\infty) - F_1(x_i) + \sum_{x_r > x_i} S_r = R(x_i).$$

Combining (3.8), (4.11) and (4.15) we have

(4.16)
$$\lim_{n\to\infty} M_2 = R(x_i)(R(x_i) + \frac{1}{2}S_i),$$

at the saltus $t = x_i$. Combining (4.8), (4.14) and (4.16) we discover that

(4.17)
$$\lim_{n\to\infty} \cos \left[G(B_n(x-t), U(x-t))\right] = R(x_i)(1-R(x_i)-\frac{1}{2}S_i),$$

at the saltus $t = x_i$. Taking (4.1), (2.2), (4.5) and (4.17) we finally obtain

$$\lim_{n\to\infty} [n \operatorname{Var} (H_n(x_i))] = 4[R(x_i) + \frac{1}{4}S_i - R^2(x_i) - \frac{1}{4}S_i^2 - S_i R(x_i) + R(x_i) - R^2(x_i) - 2R(x_i) + 2R^2(x_i) + S_i R(x_i)]$$

$$= S_i (1 - S_i),$$

at the saltus $t = x_i$.

Writing the estimator $H_n(x_i)$ as

$$H_n(x_i) = (1/n) \sum_{j=1}^n \xi_j,$$

where $\xi_j = 2[G(B_n(X_j - x_i)) - U(X_j - x_i)]$, one can easily verify that the sufficient condition for asymptotic normality given by (2.20) is satisfied by the sequence $\{\xi_j\}$ of independently and identically distributed random variables. We have therefore proved

THEOREM 2. The class of estimators $\{H_n(x_i)\}$ are consistent and asymptotically normal for estimating the jump S_i corresponding to the saltus $x = x_i$ of the distribution F(x).

Consider now the estimator $f_n^*(x_i)$ where

$$f_n^*(x_i) = (1/B_n)f_n(x_i)$$

and $f_n(x_i)$ is given by (2.4) at the saltus $t = x_i$. Since

$$f_n^*(x_i) = (1/n) \sum_{j=1}^n K(B_n(X_j - x_i)).$$

a straight forward calculation yields that

(4.20)
$$\lim_{n\to\infty} E[f_n^*(x_i)] = K(0)S_i,$$
$$\lim_{n\to\infty} [n \text{ Var } (f_n^*(x_i))] = K^2(0)S_i(1-S_i)$$

at the saltus $x = x_i$ of the distribution F(x) where the derivative of the absolutely

continuous part f(x) is assumed continuous and finally the estimate $f_n^*(x_i)$ is asymptotically normal. Thus the estimators $H_n(x_i)$ and $[1/K(0)]f_n^*(x_i)$ are asymptotically equivalent for estimating the jump S_i at the saltus $x = x_i$ of the distribution F(x).

5. Estimation of hazard rate. The function Z(t) will be called the hazard rate where

(5.1)
$$Z(t) = f(t)/[1 - F(t)] = f(t)/R(t).$$

Let us now propose $Z_n(t)$ as an estimate of the hazard rate Z(t) where

(5.2)
$$Z_n(t) = f_n(t)/R_n(t),$$

 $f_n(t)$ and $R_n(t)$ being respectively given by (2.4) and (2.1). It was earlier shown by the author (Murthy [2]) that $f_n(t)$ is a consistent estimate of f(t) at every point of continuity t of F(t) and f(t), i.e.

(5.3)
$$\operatorname{Plim}_{n\to\infty} f_n(t) = f(t).$$

It follows from (2.2) that

(5.4)
$$\operatorname{Plim}_{n\to\infty} R_n(t) = R(t).$$

Combining (5.3) and (5.4) and using a well known convergence theorem (see Cramér [1], p. 254) we at once have

(5.5)
$$\operatorname{Plim}_{n\to\infty} Z_n(t) = f(t)/R(t) = Z(t),$$

in other words $Z_n(t)$ is a consistent estimate of the hazard rate Z(t).

It was also shown by the author [2] that at every continuity point x = t of F(x) and f(x)

(5.6)
$$\lim_{n\to\infty} P\{(n/B_n)^{\frac{1}{2}}[(f_n(t) - f(t))/(f(t) \int_{-\infty}^{\infty} K^2(x) dx)^{\frac{1}{2}}] < x\}$$

= $(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy$.

Combining (5.4) and (5.6) and using a well known convergence theorem (see Cramér [1], p. 254) we discover

$$(5.7) \quad \lim_{n \to \infty} P\left\{ \left(\frac{n}{B_n}\right)^{\frac{1}{2}} \frac{\frac{f_n(t)}{R_n(t)} - \frac{f(t)}{R_n(t)}}{\left(\frac{f(t)}{R^2(t)} \int_{-\infty}^{\infty} K^2(x) \ dx\right)^{\frac{1}{2}}} < x \right\} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} \ dy.$$

Consider now

(5.8)
$$y_n = (n/B_n)^{\frac{1}{2}}(R_n(t) - R(t)).$$

We have in view of (2.2) that

(5.9)
$$E(y_n) = 0$$

$$\text{Var } (y_n) = (1/B_n)R(t)(1 - R(t)),$$

and hence

$$(5.10) plim_{n\to\infty} y_n = 0.$$

Combining (5.7) and (5.10) and using the convengence theorem (see Cramér [1], p. 254) again we finally obtain

(5.11)
$$\lim_{n\to\infty} P[(n/B_n)^{\frac{1}{2}} \{ (Z_n(t) - Z(t)) / [(Z(t)/R(t)) \int_{-\infty}^{\infty} K^2(x) dx]^{\frac{1}{2}} \} < x]$$

= $(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy$.

We have therefore proved the following

THEOREM 3. The class of estimators $Z_n(t)$ given by (5.2) for estimating the hazard rate Z(t) is asymptotically normally distributed at every point of continuity x = t of the distribution F(x) and the density f(x).

It may be observed that the estimator $Z_n^*(t)$ for estimating Z(t) where

$$Z_n^*(t) = f_n(t) / R_n^*(t),$$

is consistent and asymptotically normal, the proof being exactly similar to the one given for $Z_n(t)$.

Acknowledgment. The author expresses his gratitude to the referee whose suggestions improved the organization and contents of the paper.

REFERENCES

- [1] CRAMÉR, H. (1946). Mathematical Methods of Statistics, Princeton Univ. Press.
- [2] Murthy, V. K. (1965). Estimation of probability density. Ann. Math. Statist. 36 1027-1031.
- [3] PARZEN, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33 1065-1076.