ESTIMATING FUTURE FROM PAST IN LIFE TESTING

By John E. Walsh

Military Operations Research Division, Lockheed Aircraft Corp., Burbank, Calif.

1. Summary. Let θ_p represent the unique $100p$ per cent point of a continuous statistical population, while x_r is the rth largest value of a sample of size n from this population ($r = 1, \cdots, n$). This paper considers estimation of θ_p on the basis of $x_{r(1)}, \cdots, x_{r(m)}$, where the $r(i)$ differ by $O(\sqrt{n+1})$ and do not necessarily have values near $(n+1)p$. Also considered is estimation of x_R on the basis of $x_{r(1)}, \cdots, x_{r(m)}$, where the $r(i)$ differ by $O(\sqrt{n+1})$ and do not necessarily have values near R. The results are of a nonparametric nature and based on expected value considerations. These estimation procedures may be useful for life-testing situations where time to failure is the variable and some of the items tested have not yet failed when observation is discontinued. Then θ_p and x_R can be estimated for p and R values which extend a moderate way into the region where sample data is not available. Estimation of the x_R value which would be obtained by continuing to observe the experiment represents a prediction of the future from the past. The results of this paper may be of value in the actuarial, population statistics, operations research, and other fields.

2. Introduction. Let us consider a sampling situation where n items are simultaneously life tested to determine their times to failure. Then the time to failure for the first item which fails is the smallest value for this sample of size n. The value for the second item to fail is the next to smallest sample value; etc. Thus life-testing situations have the property that the r smallest order statistics of a sample are determined in advance of the remaining values of the sample. Moreover, the first r items to fail furnish the r smallest values of the sample of size n, even if some or all of the remaining sample values are never determined. Jacobson called attention to these valuable properties of life-testing situations in [1]. A descriptive outline of the life-testing field is given in [2].

The property that the r smallest order statistics of a lifetesting sample can be obtained without the necessity of determining the remaining sample values can be exploited in many ways. The basis for this exploitation is that substantial time and/or cost can often be saved by stopping a life-testing experiment at some convenient time before all the items have failed. The situation of this type considered here is the estimation of θ_p on the basis of $x_1, \cdots, x_r$ when $(n+1)$ $p > r$—that is, estimation of population percentage points in the region not covered by the available data.

The life-testing property that the r smallest order statistics are determined in advance of the remaining sample values furnishes an opportunity for estimating the future from the past. Suppose that r items have failed up to the present time and it is desired to predict the future time at which the Rth item of this set will

Received June 20, 1956; revised September 14, 1956.

fail ($R > r$). That is, $x_1, \cdots, x_r$ are known and estimation of x_R in an expected value sense is desired. This paper derives such an estimate for the case where r and R do not differ too much.

The purpose of this paper is to derive nonparametric expected value estimates which are approximately valid for nearly all continuous statistical populations of practical interest. These estimates are not intended to be competitive with those which can be obtained on the basis of additional information about the population sampled. Instead, the nonparametric estimates presented are for use when more specialized estimation methods are not warranted.

Whether θ_p or x_R is estimated, the arithmetical problem consists in determining the values of $c_1, \cdots, c_m$ for a linear function of the form

$$\sum_{i=1}^{m} c_i x_{r(i)},$$

where m, $r(1), \cdots, r(m)$ are given integers. The procedure for obtaining the values of the c_i consists in solving m specified linear equations in m unknowns. Although the emphasis of the paper is on life-testing situations, the results derived are valid for more general types of situations than those where only x_1, $\cdots$, x_r are available and $r < (n + 1)p, R$. For estimation of θ_p, knowledge of the values of order statistics $x_{r(1)}, \cdots, x_{r(m)}$ such that $r(i) = r(j) + O(\sqrt{n+1})$ and none of the $r(i)$ differ too much from $(n + 1)p$ is sufficient. In estimating x_R, it is sufficient that $x_{r(1)}, \cdots, x_{r(m)}$ are available with $r(i) = r(j) + O(\sqrt{n+1})$ and none of the $r(i)$ differing too much from R. Thus the results of this paper can also be used to estimate the past from the present for life-testing situations where the data for the past was lost or not recorded.

Life-testing situations where population properties are of greater interest than sample properties usually involve inanimate objects such as automobile tires, light bulbs, etc. Often a considerable savings in time and/or expense can be obtained by deliberately stopping a life test of this type when 80 to 90 per cent rather than all of the items have failed. Through use of the method given in this paper, many of the upper population percentage points of interest can be estimated even though the upper 10 per cent to 20 per cent of the data is truncated.

The future mortality occurrences among the now-surviving members of a given set of items can be of interest for some types of life-testing situations. The future mortality among the survivors of a specified group of persons which have already been observed for some time represents a situation of this nature. Estimates of the future mortality among the survivors of such a group of persons can be valuable in actuarial science, population statistics, and other fields. This paper pre sents a rather widely applicable procedure for estimating the first time at which a specified number of additional individuals will have died on the basis of the times to death for the individuals which have already died.

An investigation is made of the variances for the derived estimates. Every estimate considered has an estimate of the form

$$p_t(1 - p_t)/n[f(\theta_{p_t})]^2 + O(n^{-3/2}),$$

where t is a specified number which differs from the $r(i)$ by $O(\sqrt{n+1})$, $p_t = t/(n+1)$, and $f(x)$ is the probability density function (*pdf*) of the statistical population sampled. Thus all the estimates presented are consistent, having standard deviations which are $O(1/\sqrt{n})$.

If all the sample values were available, the corresponding sample percentage point might appear to be the most suitable nonparametric expected value estimate for θ_p. In many cases, however, an estimate for θ_p of the type given in this paper may have a higher efficiency (i.e., smaller variance) than the corresponding sample percentage point. The variance for the sample percentage point corresponding to θ_p is

$$p(1-p)/n[f(\theta_p)]^2 + O(n^{-2}).$$

If n is large and

$$p_t(1-p_t)/f(\theta_{p_t})^2 < p(1-p)/f(\theta_p)^2,$$

the sample percentage point corresponding to θ_p usually has a lower efficiency than an estimate of the type presented here. This inequality is frequently satisfied for unimodal populations where θ_{p_t} is more toward the central part of the probability distribution than θ_p.

In deriving the results, $f(x)$ is assumed to exist, be positive, and of an analytical nature for all x of interest. Such strong restrictions on $f(x)$ are not necessary for the validity of the results presented. However, little generality is gained for practical cases by using weaker restrictions on $f(x)$. There are limitations on the accuracy to which measurements on continuous observations can be made for all applied situations. This data-accuracy limitation indicates that the conditions imposed on $f(x)$ should be acceptable for virtually all practical situations of a continuous type where θ_p is unique for all p.

Section 3 contains a statement of the estimates for θ_p and x_R along with some restrictions on their use. A numerical example of the application of each type of estimate is given in Section 4. Assuming a standard normal population, the approximate properties stated for these estimates are compared with their exact properties. Section 5 contains the derivations and motivation for the material given in Section 3.

3. Statement of estimates. Let us consider an explicit statement of the method for obtaining estimates of θ_p and x_R. The additional notation used is

$$t = \frac{1}{m}\sum_{i=1}^{m} r(i),$$

$$r = \max_{1 \le i \le m} r(i),$$

$$d(i) = t - r(i) = \text{quantity which is } O(\sqrt{n+1}), \qquad (i = 1, \cdots, m);$$

$$d(i) \ne d(k) \text{ if } i \ne k$$

$$p_t = t/(n+1), \qquad q_t = 1 - p_t,$$

$$p_R = R/(n+1), \qquad q_R = 1 - p_R,$$
$$A_j = [\tfrac{1}{2}(p - p_t)]^{j-1}/(j-1)!,$$
$$B_j = \frac{[\tfrac{1}{2}(p_R - p_t)]^{j-1}}{(j-1)!} + \frac{p_R q_R (j-1)(j-2)}{2(n+2)(j-1)!}[\tfrac{1}{2}(p_R - p_t)]^{j-2},$$
$$C_j[d(i), \delta] = \frac{(-\delta)^{j-1}}{(j-1)!} - \frac{(j-1)(-\delta)^{j-2} d(i)}{(n+1)(j-1)!}$$
$$+ \frac{(j-1)(j-2)(-\delta)^{j-3}}{2(n+1)(n+2)(j-1)!}[(n+1)p_t q_t + (p_t - q_t) d(i) + d(i)^2]$$
$$- \frac{(j-1)(j-2)(j-3)(-\delta)^{j-4} d(i)}{2(n+1)^2(n+2)(j-1)!}[(n+1)p_t q_t + \tfrac{1}{3} d(i)^2],$$
$$C(j, i) = C_j[d(i), \tfrac{1}{2}(p - p_t)],$$
$$C'(j, i) = C_j[d(i), \tfrac{1}{2}(p_R - p_t)],$$

where $j = 1, \cdots, m$.

For specified n, m, $r(1), \cdots, r(m)$, p, R, the estimates considered and their principal expected value properties are given by

$$E\left[\sum_{i=1}^{m} a_i x_{r(i)}\right] = \theta_p + O(n^{-3/2}) + O[\,|\tfrac{1}{2}(p - p_t)|^m],$$
$$E\left[\sum_{i=1}^{m} b_i x_{r(i)}\right] = E(x_R) + O(n^{-3/2}) + O[\,|\tfrac{1}{2}(p_R - p_t)|^m].$$

If $m \geqq 4$, $O(n^{-3/2})$ is replaced by $O(n^{-2})$ in these expressions. The sets of linear equations used to determine the values of the a_i and the b_i are

$$\sum_{i=1}^{m} a_i C(j, i) = A_j, \qquad \sum_{i=1}^{m} b_i C'(j, i) = B_j, \qquad j = 1, \cdots, m.$$

The values for the a_i and the b_i can be conveniently expressed in the form of determinants. This form is especially useful for small values of m. Explicitly,

$$a_i = \frac{\text{determinant of the } C(u, v), \text{ with } C(j, i) \text{ replaced by } A_j \text{ for } j = 1, \cdots, m}{\text{determinant of the } C(u, v)},$$

$$b_i = \frac{\text{determinant of the } C'(u, v) \text{ with } C'(j, i) \text{ replaced by } B_j \text{ for } j = 1, \cdots, m}{\text{determinant of the } C'(u, v)}.$$

If the determinant of the $C(u, v)$ is zero or near zero, a change in the values for the $r(i)$ may be required to assure that none of the a_i are of too large a magnitude. Usually a change of one value is enough to eliminate this difficulty. The same will be true if the determinant of the $C'(u, v)$ is zero or near zero.

Let us consider determination of a value for m which seems large enough to assure that the unstated higher order expected value terms can be neglected. Accuracy to terms of order $n^{-3/2}$ implies that $m \geqq 3$. The value used is also required to satisfy the condition

$$m \geqq \begin{cases} \dfrac{\log [\max (n^{-3/2}, 10^{-4})]}{\log | \frac{1}{2}(p - p_t) |}, & \text{if } \theta_p \text{ estimated,} \\ \dfrac{\log [\max (n^{-3/2}, 10^{-4})]}{\log | \frac{1}{2}(p_R - p_t) |}, & \text{if } x_R \text{ estimated.} \end{cases}$$

This second condition for determining m is based on the requirement that

$$\max(n^{-3/2}, 10^{-4}) \geqq \begin{cases} | \frac{1}{2}(p - p_t) |^m, & \text{if } \theta_p \text{ estimated,} \\ | \frac{1}{2}(p_R - p_t) |^m, & \text{if } x_R \text{ estimated.} \end{cases}$$

These two conditions assure that the expected value error of an estimate is $O[\max (n^{-3/2}, 10^{-4})]$. The minimum value of $m = 3$ is acceptable for many of the cases encountered. From a computational viewpoint, the method probably should not be used if the value of m obtained by this procedure exceeds 10.

The convergence rates of the expansions used in obtaining estimates for θ_p depend on $\frac{1}{2}(p - p_t)$, the $d(i)/(n + 1)$, and the properties of the underlying statistical population. In practice, the underlying population properties are usually such that convergence is more rapid for p_t and p near the center of the distribution. On this basis, both $| d(i) |/(n + 1)p_t q_t$ and $\frac{1}{2}| p - p_t |/[\min(pq, p_t q_t)]$ should not be too large. The maximum allowable value for these quantities is taken to be $\frac{2}{3}$ for the type of situations considered. This value is not overly small but should be satisfactory for a large majority of the practical applications. Hence, the method given in this paper for estimating θ_p should not be used if either

$$\max_{1 \leqq i \leqq m} | d(i) | > \tfrac{2}{3} p_t q_t (n + 1),$$

or

$$| p - p_t | > \tfrac{4}{3} \min (pq, p_t q_t).$$

On a similar basis, the method for estimating x_R should not be used if either

$$\max_{1 \leqq i \leqq m} | d(i) | > \tfrac{2}{3} p_t q_t (n + 1),$$

or

$$| p_R - p_t | > \tfrac{4}{3} \min (p_R q_R , p_t q_t).$$

Sometimes the inequality involving the $d(i)$ can be changed from unacceptable to acceptable by using a different value for m which allows a decrease in max $| d(i) |$.

When $x_1 , \cdots , x_r$ are given and $r > (n + 1)p, R$, a recommended selection for the values of the $r(i)$ in both types of estimates is

$$r(i) = r - (m - i)K, \qquad (i = 1, \cdots , m),$$

where

$$K = \max\,[1, \text{largest integer contained in } \frac{1}{m}\sqrt{n+1}].$$

The resulting $r(i)$ differ by $O(\sqrt{n+1})$, are equally spaced, and have desirable properties with respect to the expansions used in deriving the estimates.

Every estimate of the two types considered has approximately the same variance. For all estimates derived, the variances are of the form

$$p_t q_t / n[f(\theta_{p_t})]^2 + O(n^{-3/2}).$$

Thus, each estimate has a standard deviation which is $O(n^{-1/2})$. The order of the standard deviation for an estimate is the reason for neglecting all terms involving n to orders $n^{-3/2}$ and higher in the expected value expressions for these estimates.

4. Numerical Example. To illustrate use of the methods of this paper, let us consider the case where $n = 20$, $x_1, \cdots, x_{15}$ are given, $p = 0.84$, and $R = 17$. The value of m is determined first. This value is the smallest integer which is at least 3 and such that

$$m \geqq \begin{cases} \dfrac{\log\,[\max\,(n^{-3/2}, 10^{-4})]}{\log\,|\,\frac{1}{2}(p - p_t)\,|} \doteq 1.84, & \text{if } \theta_p \text{ estimated,} \\[2ex] \dfrac{\log\,[\max\,(n^{-3/2}, 10^{-4})]}{\log\,|\,\frac{1}{2}(p_R - p_t)\,|} \doteq 1.71, & \text{if } x_R \text{ estimated.} \end{cases}$$

Thus $m = 3$ for both types of estimates.

Next let us evaluate $r(1)$, $r(2)$, and $r(3)$. The value of K is given by

$$K = \max\,[1, \text{largest integer contained in } \frac{1}{m}\sqrt{n+1}] = 1.$$

Hence, for both types of estimates

$$r(1) = 13, \qquad r(2) = 14, \qquad r(3) = 15,$$

since $r(i) = r - (m - i)K$. Thus $t = 14$ and

$$d(1) = 1, \qquad d(2) = 0, \qquad d(3) = -1.$$

Also the relations

$$\max_{1 \leqq i \leqq m} |\,d(i)\,| \leqq \tfrac{2}{3} p_t q_t (n + 1),$$

$$|\,p - p_t\,| \leqq \tfrac{4}{3} \min\,(pq, p_t q_t),$$

$$|\,p_R - p_t\,| \leqq \tfrac{4}{3} \min\,(p_R q_R, p_t q_t)$$

are easily verified so that the methods of the paper are applicable for the case considered.

By direct substitution, the values of the A_j and B_j are found to be

$$A_1 = 1.0000, \quad A_2 = 0.0867, \quad A_3 = 0.0038,$$
$$B_1 = 1.0000, \quad B_2 = 0.0715, \quad B_3 = 0.0028.$$

Thus

$$\begin{array}{lll} C(1, 1) = 1.0000, & C(1, 2) = 1.0000, & C(1, 3) = 1.0000, \\ C(2, 1) = -0.0390 & C(2, 2) = -0.0867 & C(2, 3) = -0.1344, \\ C(3, 1) = 0.0054 & C(3, 2) = 0.0088 & C(3, 3) = 0.0144, \end{array}$$

and

$$\begin{array}{lll} C'(1, 1) = 1.0000, & C'(1, 2) = 1.000, & C'(1, 3) = 1.0000, \\ C'(2, 1) = -0.0238 & C'(2, 2) = -0.0715 & C'(2, 3) = -0.1192, \\ C'(3, 1) = 0.0049 & C'(3, 2) = 0.0076 & C'(3, 3) = 0.0125. \end{array}$$

Consequently,

$$a_1 = \begin{vmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.0867 & -0.0867 & -0.0390 \\ 0.0038 & 0.0088 & 0.0054 \end{vmatrix} \Bigg/ \begin{vmatrix} 1.0000 & 1.0000 & 1.0000 \\ -0.1344 & -0.0867 & -0.0390 \\ 0.0144 & 0.0088 & 0.0054 \end{vmatrix}$$
$$= 3.38.$$

Similarly, $a_2 = -9.41$ and $a_3 = 7.03$. Also

$$b_1 = \begin{vmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.0715 & -0.0715 & -0.0238 \\ 0.0028 & 0.0076 & 0.0049 \end{vmatrix} \Bigg/ \begin{vmatrix} 1.0000 & 1.0000 & 1.0000 \\ -0.1192 & -0.0715 & -0.0238 \\ 0.0125 & 0.0076 & 0.0049 \end{vmatrix}$$
$$= 1.49.$$

In a like fashion, $b_2 = -4.99$ and $b_3 = 4.50$.

Using the values determined for the a_i and b_i, approximate expected value estimates are obtained for $\theta_{0.84}$ and x_{17}. These estimates and their properties are given by

$$\begin{aligned} E(3.38\, x_{13} - 9.41\, x_{14} + 7.03\, x_{15}) &= \theta_{0.84} + O(n^{-3/2}) + O[|\tfrac{1}{2}(p - p_t)|^m] \\ &= \theta_{0.84} + O(0.011) + O(0.00065), \\ E(1.49\, x_{13} - 4.99\, x_{14} + 4.50\, x_{15}) &= E(x_{17}) + O(n^{-3/2}) + O[|\tfrac{1}{2}(p_R - p_t)|^m] \\ &= E(x_{17}) + O(0.011) + O(0.00036). \end{aligned}$$

Here the contribution of order $n^{-3/2}$ seems to be much more important that the contribution of order $|\tfrac{1}{2}(p - p_t)|^m$ or the contribution of order $|\tfrac{1}{2}(p_R - p_t)|^m$.

To check the expected value accuracy of these estimates, let us consider the special case of a normal distribution with zero mean and unit variance. Using a table of the standardized normal distribution and the results of [3],

$$E(3.38\, x_{13} - 9.41\, x_{14} + 7.03\, x_{15}) = 0.995, \qquad \theta_{0.84} = 0.996,$$

$$E(1.49\, x_{13} - 4.99\, x_{14} + 4.50\, x_{15}) = 0.888, \qquad E(x_{17}) = 0.921$$

Thus the expected values of these two estimates are in rather close agreement with the true values for the case of normality. This expected value agreement is much closer than is required on the basis of the standard deviations of these estimates. For the standardized normal case,

$$\text{s.d. of } (3.38\, x_{13} - 9.41\, x_{14} + 7.03\, x_{15}) \doteq 1.03$$

$$\text{s.d. of } (1.49\, x_{13} - 4.99\, x_{14} + 4.50\, x_{15}) \doteq 0.72.$$

Due to the moderately small value of n, these standard deviation values do not agree very closely with the asymptotic value of

$$\sqrt{p_t q_t}/\sqrt{n} f(\theta_{p_t}) \doteq 0.33.$$

The moderately small value of $n = 20$ was selected for the example in order that the results of [3] could be used.

5. Derivations. Here verification is presented for the expected value and variance results stated in Section 3. This verification is based on the material presented by David and Johnson in [4].

Let s be a number such that $1 \leqq s \leqq n$ while $f(x)$ has derivatives of all orders at all points where it is defined and is non-zero at all points considered. Some additional notation is used

$$F(X) = \int_{-\infty}^{X} f(x)\, dx, \qquad F(X_s) = p_s, \qquad X_s^{(u)} = \left.\frac{d^u X}{dF^u}\right|_{X=X_s},$$

$u = 1, 2, \cdots$. Here $X_s^{(0)} = X_s$ while $X_{(n+1)p} = \theta_p$, whether $(n + 1)p$ is an integer or not. On the basis of [4],

$$E[x_{r(i)}]$$

$$= X_{r(i)} + \frac{p_{r(i)} q_{r(i)}}{2(n + 2)} X_{r(i)}^{(2)} + \frac{p_{r(i)} q_{r(i)}}{(n + 2)^2} [\tfrac{1}{3}(q_{r(i)} - p_{r(i)}) X_{r(i)}^{(3)} + \tfrac{1}{8} p_{r(i)} q_{r(i)} X_{r(i)}^{(4)}]$$

to terms of order n^{-3}, where $p_{r(i)} = r(i)/(n + 1)$, and $q_{r(i)} = 1 - p_{r(i)}$. Also

$$E(x_R) = X_R + \frac{p_R q_R}{2(n + 2)} X_R^{(2)} + O(n^{-2}),$$

on the basis of [4].

The first step of the procedure used for developing the estimate of θ_p consists in expanding the $E[x_{r(i)}]$ and θ_p about a probability value which is halfway between p_t and p. The value p_t is considered because of the relation

$$E[x_{r(i)}] = \theta_{t/(n+1)} + O(1/\sqrt{n}), \qquad (i = 1, \cdots, m).$$

Taylor series expansion about the midway probability value of $\frac{1}{2}(p + p_t)$ yields desirable convergence properties for both the $E[x_{r(i)}]$ and θ_p. In particular, these expansions have about the same rate of convergence. Similarly, the first step in developing the estimate for x_R consists in expanding the $E[x_{r(i)}]$ and $E(x_R)$ about the midway probability value of $\frac{1}{2}(p_R + p_t)$. Here the probability p_R is considered because of the relation $E(x_R) = \theta_{R/(n+1)} + O(n^{-1})$.

Next let us consider the expansion of $E[x_{r(i)}]$ about the general probability value of $\frac{1}{2}(Q + p_t)$. By use of Taylor series,

$$X_{r(i)}^{(u)} = \sum_{v=0}^{\infty} \frac{(-1)^v}{v!} \left[\frac{\frac{1}{2}(n+1)(Q - p_t) + d(i)}{n+1}\right]^u X_{\frac{1}{2}[(n+1)Q+t]}^{(u+v)},$$

$(u = 0, 1, \cdots)$, since

$$\frac{d^v X_s^{(u)}}{ds^v} = \frac{X_s^{(u+v)}}{(n+1)^v}, \qquad (v = 1, 2, \cdots).$$

Substitution of these relations into the expression given for $E[x_{r(i)}]$ shows that

$$E[x_{r(i)}] = \sum_{j=1}^{m} C_j[d(i), \tfrac{1}{2}(Q - p_t)] X_{\frac{1}{2}[(n+1)Q+t]}^{(j-1)} + O(n^{-2}) + \sum_{j=m-3}^{m} O\{n^{-(m-j)/2} | \tfrac{1}{2}(Q - p_t)|^j\}.$$

Here the $O(n^{-2})$ and/or the $O[|\frac{1}{2}(Q - p_t)|^m]$ terms are the most important of those which are not explicitly stated.

Now the expansions for θ_p and $E(x_R)$ are considered. The Taylor series expansion of θ_p about the probability value of $\frac{1}{2}(p + p_t)$ is

$$\theta_p = \sum_{u=0}^{\infty} \frac{1}{u!} [\tfrac{1}{2}p - p_t)]^u X_{\frac{1}{2}[(n+1)p+t]}^{(u)}.$$

To obtain the expansion of $E(x_R)$ to $O(n^{-2})$, the Taylor series expansion of the $X_R^{(u)}$ about the probability value of $\frac{1}{2}(p_R + p_t)$ is needed. This is given by

$$X_R^{(u)} = \sum_{v=0}^{\infty} \frac{1}{v!} [\tfrac{1}{2}(p_R - p_t)]^v X_{\frac{1}{2}(R+t)}^{(u+v)}, \qquad (u = 0, 1, \cdots).$$

Substitution of these relations into the expression given for $E(x_R)$ shows that

$$E(x_R) = \sum_{u=0}^{\infty} \left\{ \frac{[\frac{1}{2}(p_R - p_t)]^u}{u!} + \frac{p_R q_R u(u-1)}{2(n+2)u!} [\tfrac{1}{2}(p_R - p_t)]^{u-1} \right\} X_{\frac{1}{2}(R+t)}^{(u)}$$

plus terms of order n^{-2}.

To determine the equations which are used to evaluate $a_1, \cdots, a_m$, first set Q equal to p. Then the coefficient of $X^{(u)}_{\frac{1}{2}[(n+1)p+t]}$ in the expansion for

$$E(\sum a_i x_{r(i)})$$

is required to be the same as the coefficient of this quantity in the expansion of θ_p, $(u = 0, 1, \cdots, m - 1)$, to terms of the prescribed order. Examination of the expansions for $E[x_{r(i)}]$ and θ_p shows that the m linear equations in m unknowns given in Section 3 for evaluating the a_i satisfy this requirement when $m \geqq 3$. If $m \geqq 4$, the terms of order $n^{-3/2}$ are also cancelled out.

To determine the equations used to evaluate $b_1, \cdots, b_m$, set Q equal to p_R. Then the coefficient of $X^{(u)}_{\frac{1}{2}(R+t)}$ in the expansion for $\mathrm{E}(\sum b_i x_{r(i)})$ is required to be the same as the coefficient of this quantity in the expansion for $E(x_R)$, $(u = 0, 1, \cdots, m - 1)$, to terms of the prescribed order. Examination of the expansions for $E[x_{r(i)}]$ and $E(x_R)$ shows that the equations given in Section 3 for evaluating the b_i satisfy this requirement when $m \geqq 3$. If $m \geqq 4$, the terms of order $n^{-3/2}$ also cancel out.

Finally let us consider the variance expressions for the type of estimates considered. Let the $d(i)$ be numbered so that $d(1) > d(2) > \cdots > d(m)$. Then, using the variance results presented in [4] and the general notation c_i to represent either the a_i or the b_i,

$$\begin{aligned}\operatorname{var}\left[\sum_{i=1}^{m} c_i x_{r(i)}\right] &= \sum_{i=1}^{m} \frac{c_i^2}{n}\left[p_t q_t + \frac{(p_t - q_t)\, d(i)}{n+1}\right]\left[X_t^{(1)} - \frac{d(i)}{n+1} X_t^{(2)}\right]^2 \\ &\quad + 2 \sum_{i>j=1}^{m} \frac{c_i c_j}{n}\left[p_t q_t + \frac{p_t\, d(i) - q_t\, d(j)}{n+1}\right]\left[X_t^{(1)} - \frac{d(i)}{n+1} X_t^{(2)}\right] \\ &\quad \cdot \left[X_t^{(1)} - \frac{d(j)}{n+1} X_t^{(2)}\right] + O(n^{-2}).\end{aligned}$$

This follows from the fact that all of the a_i, b_i, and $d(i)/\sqrt{n+1}$ are $O(1)$ with respect to n. Using the condition $\sum_1^m c_i = 1$, which holds in all cases for the estimates derived, it is easily verified that

$$\operatorname{var}\left[\sum_{i=1}^{m} c_i x_{r(i)}\right] = \frac{p_t q_t}{n[f(\theta_{p_t})]^2} + O(n^{-3/2});$$

here the relation $X_t^{(1)} = 1/f(\theta_{p_t})$ is used.

REFERENCES

[1] Paul H. Jacobson, "The relative power of three statistics," *Jour. Amer. Stat. Assoc.*, Vol. 42 (1947), pp. 575–84.

[2] B. Epstein and M. Sobel, "Life testing," *Jour. Amer. Stat. Assoc.*, Vol. 48 (1953), pp. 486–502.

[3] D. Teichroew, "Tables of expected values of order statistics and products of order statistics for samples of size twenty or less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1955), pp. 410–26.

[4] F. N. David and N. L. Johnson, "Statistical treatment of censored data. Part I—fundamental formulae," *Biometrika*, Vol. 41 (1954), pp. 228–40.