REMARK ON THE LINEARIZED MAXIMUM LIKELIHOOD ESTIMATE[1]

By L. K. Chan

University of Western Ontario

1. Introduction. Let

$$y_u < y_{u+1} < \cdots < y_v \tag{1}$$

($u = [np]$ (the greatest integer $\leqq np$), $v = [nq]$, $0 < p < q < 1$) be a Type II doubly censored sample corresponding to an ordered random sample of size n from a continuous distribution whose distribution function is $F((y - \theta_1)/\theta_2)$, where the location parameter θ_1 and scale parameter $\theta_2 > 0$ are the unknown parameters to be estimated. Plackett [6] derived a linearized maximum likelihood estimate of (θ_1, θ_2) which has been referred to in the literature, e.g. [2], [8] and [3]. Unfortunately there is an error in his derivation ([6], page 138) in which he treated all the consecutive order statistics of (1) as sample quantiles. When we study the asymptotic properties of sample quantiles, e.g. [7], Section 2.2, the quantiles are determined by a fixed number, say k, of pre-determined proportions and hence only k order statistics are considered for each n. However for a Type II doubly censored sample, the number $v - u + 1$ of order statistics tends to ∞ as $n \to \infty$. In this remark, a more rigorous derivation under less restricted conditions is given. A consequence of this is that the estimate is asymptotically equivalent to the linear estimate proposed by Weiss [10] and Chernoff, et al. [3].

Most of the notations used will be the same as those in [6]. Since the conditions on the distribution $F((y - \theta_1)/\theta_2)$ were not explicitly given in [6], we shall mention here the conditions required for our argument. We shall let $(1/\theta_2)f((y - \theta_1)/\theta_2)$ be the probability density function corresponding to $F((y - \theta_1)/\theta_2)$, and let $x = (y - \theta_1)/\theta_2$.

C_1. $F((y - \theta_1)/\theta_2)$ is a continuous distribution, and for almost all y

$$\partial^{i+j} \log (1/\theta_2)f((y - \theta_1)/\theta_2)/\partial\theta_r^i\partial\theta_s^j,$$

$$i, j = 0, 1, 2, 3; \qquad 1 \leqq i + j \leqq 3; \qquad r, s = 1, 2,$$

exist for every (θ_1, θ_2) belonging to some two dimensional non-degenerate region $\Theta \subset \{(w, z) \mid -\infty < w < \infty, 0 < z < \infty\}$.

C_2. For every $(\theta_1, \theta_2) \in \Theta$,

$$|\partial^{i+j}(1/\theta_2)f((y - \theta_1)/\theta_2)/\partial\theta_r^i\partial_s^j| < A_{ij}(y),$$

$$|\partial^3 \log (1/\theta_2)f((y - \theta_1)/\theta_2)/\partial\theta_r^2\partial\theta_s| < B_{rs}(y),$$

Received 6 September 1966; revised 5 June 1967.

[1] Work done at the Summer Research Institute of the Canadian Mathematical Congress.

where the $A_{ij}(y)$ are integrable over $(-\infty, \infty)$,

$$\int_{-\infty}^{\infty} B_{rs} \cdot (1/\theta_2) f((y - \theta_1)/\theta_2)\, dy < M_{rs} \quad \text{and} \quad M_{rs}$$

is independent of θ_1 and θ_2.

C_3. The Fisher information matrix $\theta_2^{-2} \|I_{rs}(p, q)\|$ is positive definite, where

$$\begin{aligned} \theta_2^{-2} I_{rs}(p, q) = {} & \int_{\theta_2 F^{-1}(p)+\theta_1}^{\theta_2 F^{-1}(q)+\theta_1} [\partial \log (1/\theta_2) f((y - \theta_1)/\theta_2)/\partial\theta_r \\ & \cdot \partial \log (1/\theta_2) f((y - \theta_1)/\theta_2)/\partial\theta_s] \cdot [(1/\theta_2) f((y - \theta_1)/\theta_2)]\, dy \\ & + 1/p [\int_{-\infty}^{\theta_2 F^{-1}(p)+\theta_1} (\partial(1/\theta_2) f((y - \theta_1)/\theta_2)/\partial\theta_r)\, dy \\ & \cdot \int_{-\infty}^{\theta_2 F^{-1}(p)+\theta_1} (\partial(1/\theta_2) f((y - \theta_1)/\theta_2)/\partial\theta_s)\, dy] \\ & + (1 - q)^{-1} [\int_{\theta_2 F^{-1}(q)+\theta_1}^{\infty} (\partial(1/\theta_2) f((y - \theta_1)/\theta_2)/\partial\theta_r)\, dy \\ & \cdot \int_{\theta_2 F^{-1}(q)+\theta_1}^{\infty} (\partial(1/\theta_2) f((y - \theta_1)/\theta_2)/\partial\theta_s)\, dy], \qquad r, s = 1, 2, \end{aligned}$$

and $F^{-1}(\alpha)$ denotes the αth population quantile of $F(x)$.

C_4. For every $(\theta_1, \theta_2) \varepsilon \Theta$, there exist neighborhoods of $\theta_2 F^{-1}(p) + \theta_1$, $\theta_2 F^{-1}(q) + \theta_1$ in which

$$d[(1/\theta_2) f((y - \theta_1)/\theta_2)]/dy,$$

and the $\partial^{i+j} \log (1/\theta_2) f((y - \theta_1)/\theta_2)/\partial\theta_r^{\,i} \partial\theta_s^{\,j}$ are continuous.

C_5. $\lim_{x \to \infty} |x|^{\epsilon} [1 - F(x) + F(-x)] = 0$ for some $\epsilon > 0$.

C_6. Let $H_1(x) = F(x)$ and $H_2(x) = 1 - F(x)$; then

$$|C_{1r}(x)| = |d^2(f/H_r)/dx^2| \leqq K[F(x)(1 - F(x))]^{-\lambda} \quad \text{for some} \quad \lambda > 0, r = 1, 2,$$

and $|D_1(x)| = |d^2(-d \log f(x)/dx)/dx^2| \leqq K[F(x)(1 - F(x))]^{-\lambda_0}$ for some $\lambda_0 > 0$, where K serves as a generic constant.

Explicit expressions for the elements in $\theta_2^{-2} \|I_{rs}(p, q)\|$ (cf. C_3) can be found in Table 2 of [6] if we replace all t_u and t_v in that table by $F^{-1}(p)$ and $F^{-1}(q)$, respectively.

The existence of C_{1r} and D_1 follows from C_1. From C_5, for sufficiently large x,

$$|x| \leqq K[F(x)(1 - F(x))]^{-1/\epsilon}, \tag{2}$$

and from C_1

$$f(x) > 0 \quad \text{on} \quad \{x \mid 0 < F(x) < 1\}. \tag{3}$$

C_6 and (2) imply that

$$|C_{2r}(x)| = |d^2(xf/H_r)/dx^2| \leqq K[F(1 - F)]^{-\lambda - 1/\epsilon}, \qquad r = 1, 2, \tag{4}$$

and

$$|D_2(x)| = |d^2[-1 - x(d \log f(x)/dx)]/dx^2| \leqq K[F(1 - F)] - \lambda_0 - 1/\epsilon.$$

It is easily seen from (2) that if $C_{1r}(x)/|x|^{\beta}$ and $D_1(x)/|x|^{\beta_0}$ are bounded for some $\beta' > 0$ and $\beta_0 > 0$ when x is sufficiently large, then C_6 and (4) are satisfied. C_6 and

(3) are weaker than the condition that D_1, D_2, C_{1r} and C_{2r} are bounded which was imposed in [6].

C_5 is essential since this condition and (3) are the necessary and sufficient conditions for the existence of the first two moments $E((y_i - \theta_1)/\theta_2)$ and $E((y_i - \theta_1)/\theta_2)^2$, $i = u, u+1, \cdots v$, (cf. Bickel [1], Theorem 2.2a)), which will be needed in our discussion.

2. The linear estimate. Let $x_i = (y_i - \theta_1)/\theta_2$, $f(x) = dF(x)/dx$, $t_i = Ex_i$, $\Sigma = \sum_{i=[np]}^{[nq]}$ and let L_n be the likelihood function of (1). It can be seen that $\partial \log L_n/\partial\theta_r$ is a function of the form

$$g_r((y_u - \theta_1)/\theta_2, \cdots, (y_v - \theta_1)/\theta_2) + h_r(\theta_1, \theta_2).$$

Change the variables to $x_i = (y_i - \theta_1)/\theta_2$ and perform a Taylor's expansion around $x_u = t_u, \cdots, x_v = t_v$. Then (cf. [6], (44), (45))

$$\begin{aligned} n^{-\frac{1}{2}}\partial \log L_n/\partial\theta_r &= n^{-\frac{1}{2}}(\partial \log L_n/\partial\theta_r)_{t_u,\cdots,t_v} + (n^{\frac{1}{2}}\theta_2)^{-1}L_{rn}^0 \\ &\quad - (2n^{\frac{1}{2}}\theta_2)^{-1}(u-1)((y_u - \theta_1)/\theta_2 - t_u)^2 C_{r1}(x_u{}^*) \\ &\quad + (2n^{\frac{1}{2}}\theta_2)^{-1}\sum((y_i - \theta_1)/\theta_2 - t_i)^2 D_r(x_i{}^*) \\ &\quad + (2n^{\frac{1}{2}}\theta_2)^{-1}(n-v)((y_v - \theta_1)/\theta_2 - t_v)^2 C_{r2}(x_v{}^*), \quad r = 1, 2, \end{aligned} \tag{5}$$

where $|x_i{}^* - t_i| < |x_i - t_i|$, $i = u, u+1, \cdots, v$ and

$$\begin{aligned} L_{1n}^0 &= -(u-1)((y_u - \theta_1)/\theta_2 - t_u)(f_u'/p_u - f_u{}^2/p_u{}^2) \\ &\quad - \sum((y_i - \theta_1)/\theta_2 - t_i)\cdot(f_i''/f_i - f_i'^2/f_i^2) \\ &\quad + (n-v)((y_v - \theta_1)/\theta_2 - t_v)(f_v'/q_v + f_v{}^2/q_v{}^2), \\ L_{2n}^0 &= -(u-1)((y_u - \theta_1)/\theta_2 - t_u)(f_u/p_u + t_u f_u'/p_u - t_u f_u{}^2/p_u{}^2) \\ &\quad - \sum((y_i - \theta_i)/\theta_2 - t_i)\cdot(t_i f_i''/f_i - t_i f_i'^2/f_i^2 + f_i'/f_i) \\ &\quad + (n-v)((y_v - \theta_1)/\theta_2 - t_v)(f_v/q_v + t_v f_v'/q_v + t_v f_v{}^2/q_v{}^2), \end{aligned} \tag{6}$$

$f_i = f(t_i)$, $p_u = F(t_u)$, $q_v = 1 - F(t_v)$, $f_i' = f'(t_i)$, $f_i'' = f''(t_i)$, $f'(x) = df(x)/dx$, $f''(x) = d^2f(x)/dx^2$. (Our expansion is different from that of Plackett since we consider the (θ_1, θ_2) in $(y_i - \theta_1)/\theta_2$ of (5) as a constant while he considered it as the maximum likelihood estimate.)

In this remark we prove the following theorem.

THEOREM. *Consider a distribution $F((y - \theta_1)/\theta_2)$ which satisfies the conditions C_1 to C_6. Let $(\theta_1{}^0, \theta_2{}^0)$ be the solution of the linear equations $L_{rn}^0 = 0$, $r = 1, 2$. Then $(\theta_1{}^0, \theta_2{}^0)$ is a strictly unbiased estimate of (θ_1, θ_2) for every n and*

$$n^{\frac{1}{2}}\left\|\begin{matrix}\theta_1{}^0 - \theta_1 \\ \theta_2{}^0 - \theta_2\end{matrix}\right\| = \theta_2 \| J_{rs}^0(n) \|^{-1} \left\|\begin{matrix} n^{-\frac{1}{2}} L_{1n}^0 \\ n^{-\frac{1}{2}} L_{2n}^0\end{matrix}\right\|$$

converges in distribution to the bivariate normal distribution $N((0, 0), \theta_2{}^2 \|I_{rs}(p, q)\|^{-1})$, where $-n\|J_{rs}^0(n)\|$ is the matrix whose elements are the coefficients of θ_1 and θ_2 in the equations $\theta_2 L_{rn}^0 = 0$.

The unbiasedness of (θ_1^0, θ_1^0) follows immediately because from (6) $EL_{rn}^0 = 0$, $r = 1, 2$. For the remainder of this theorem we see that it is sufficient to prove the following Lemmas.

LEMMA 1. *Under* C_1 *to* C_4, $(n^{-\frac{1}{2}}\partial \log L_n/\partial\theta_1, n^{-\frac{1}{2}}\partial \log L_n/\partial\theta_2)$ *converges in distribution to* $N((0, 0), \theta_2^{-2} \|I_{rs}(p, q)\|)$.

LEMMA 2. *Under* C_5, C_6 *and* (3), *each term involving* $((y_i - \theta_1)/\theta_2 - t_i)^2$ *in* (5) *converges in probability to zero.*

LEMMA 3. *Under* C_1 *to* C_6,

$$\lim_{n\to\infty} n^{-\frac{1}{2}}(\partial \log L_n/\partial\theta_r)_{t_u,\cdots t_v} = 0, \qquad r = 1, 2.$$

LEMMA 4. *Under* C_1, C_5 *and* C_6

$$\lim_{n\to\infty} \|J_{rs}^0(n)\| = \|I_{rs}(p, q)\|.$$

PROOF OF LEMMA 1. This lemma is a direct extension to the two-parameter and double censoring case of the expression (3.5.16) given by Halperin [4] (the possibility of such an extension is given in §4 of [4]).

PROOF OF LEMMA 2. It follows from Theorem 2.2b by Bickel [1] that under C_5 and (3) we have for any given α, $0 < \alpha < \frac{1}{2}$,

$$(7) \quad E[x_i - F^{-1}(i/(n+1))]^r = n^{-r/2}[(i/n)(1 - i/n)f^{-2}(F^{-1}(i/n))]^{r/2}\mu_r + o(n^{-r/2}), \quad r = 1, 2,$$

uniformly for $\alpha n \leqq i \leqq (1 - \alpha)n$ when n is sufficiently large, where $\mu_r = \int_{-\infty}^{\infty} x^r(1/(2\pi)^{\frac{1}{2}}) \exp(-x^2/2)\, dx$. Now let us prove for illustration that when $r = 2$ the fourth right-hand term of (5) converges in probability to zero. It is sufficient to prove that

$$(8) \qquad \lim_{n\to\infty} n^{-\frac{1}{2}}\Sigma E[(x_i - t_i)^2 |D_2(x_i^*)|] = 0.$$

Let $m = [\lambda_0 + 1/\epsilon] + 1$. It is easily seen that $[F(1 - F)]^{-m}$ is symmetric about $F = \frac{1}{2}$, decreases in $(0, \frac{1}{2})$ and attains its minimum value $(\frac{1}{2})^{-2m}$ when $F = \frac{1}{2}$. Now let the α in (7) be less than $\min(p, 1 - q)$. Then when n is sufficiently large we see that $(i - 1) > m$, $(n - i) > m$ and from (7) (putting $r = 1$) that $t_i \,\varepsilon\, (F^{-1}(\alpha), F^{-1}(1 - \alpha))$ for $i = u, u + 1, \cdots, v$. Let I_α be the interval $[F^{-1}(\alpha), F^{-1}(1 - \alpha)]$. Then by (4)

$$\begin{aligned} n^{-\frac{1}{2}}\Sigma E[(x_i - t_i)^2 |D_2(x_i^*)|] &\leqq n^{-\frac{1}{2}}\{\Sigma n\tbinom{n-1}{i-1} \textstyle\int_{x\varepsilon I_\alpha}(x - t_i)^2F^{i-1}(1 - F)^{n-i}\cdot K \\ &\cdot(\alpha(1 - \alpha))^{-m}\, dF\} + n^{-\frac{1}{2}} \textstyle\sum \{n(n - 1) \cdots (n - 2m + 1) \\ &\cdot[(i-1)(i-2) \cdots (i-m)]^{-1}[(n - i)(n - i - 1) \cdots (n - i - m + 1)]^{-1} \\ &\cdot(n - 2m)\tbinom{n-2m-1}{i-m-1}\cdot \textstyle\int_{x\notin I_\alpha} 4[(x - F^{-1}((i - m)/(n - 2m + 1)))^2 \\ &+ (F^{-1}((i - m)/(n - 2m + 1)) - F^{-1}(i/(n + 1)))^2 \\ &+ (F^{-1}(i/(n + 1)) - t_i)^2]\cdot K\cdot F^{i-1-m}(1 - F)^{n-i-m}\, dF\}. \end{aligned}$$

If follows from (7) that the above two right-hand terms tend to zero and hence (8) holds.

PROOF OF LEMMA 3. Under C_1 to C_4, we have $E\partial \log L_n/\partial\theta_r = 0, r = 1, 2$ (this is an extension to the two-parameter and double censoring case of the expression (3.3.1) in [4]). Note that $EL_{rn}^0 = 0, r = 1, 2$. From an argument similar to that in the proof of (8) the expectations of all the terms in (5) involving $((y_i - \theta_1)/\theta_2 - t_i)^2$ tend to zero as $n \to \infty$. The lemma follows by taking expectations on both sides of (5) and then letting $n \to \infty$.

PROOF OF LEMMA 4. Let us demonstrate that

$$\lim_{n\to\infty} J_{22}^0(n) = I_{22}(p, q).$$

We see from (6) that

$$\begin{aligned} -J_{22}^0(n) &= n^{-1} \textstyle\sum (t_i)\, d/dx(-1 - x\, d \log f(x)/dx)|_{x=t_i} \\ &\quad - n^{-1}(u-1)(t_u)\, d/dx(xf(x)/F(x))|_{x=t_u} \\ &\quad + n^{-1}(n-v)(t_v)\, d/dx(xf(x)/(1-F(x)))|_{x=t_v}. \end{aligned} \tag{9}$$

Note from (4) that $C_{21}(x)$, $C_{22}(x)$, $D_2(x)$ exist and are bounded when $x \varepsilon I_\alpha$. So we can substitute $F^{-1}(i/(n+1)) + o(n^{-\frac{1}{2}})$ for t_i in (9) (by putting $r = 1$ in (7)), expand $-J_{22}^0(n)$ around $(F^{-1}(u/(n+1)), \cdots, F^{-1}(v/(n+1)))$ and then obtain

$$J_{22}^0(n) = J_{22}^*(n) + o(n^{-\frac{1}{2}}),$$

where $J_{22}^*(n)$ is equal to $J_{22}^0(n)$ with each t_i being replaced by $F^{-1}(i/(n+1))$. By applying the Lemma in [9]

$$\lim_{n\to\infty} J_{22}^*(n) = I_{22}(p, q).$$

In addition to considering the (θ_1, θ_2) in $(y_i - \theta_1)/\theta_2$ of (5) as the maximum likelihood estimate, Plackett also divided the right-hand sides of (5) by n instead of $n^{\frac{1}{2}}$ and showed that the square terms converge in probability to zero. But in such a case the linear terms may also converge in probability to zero. So it is difficult to say that the square terms are asymptotically negligible compared with the linear terms. Rigorous results concerning asymptotic normality of the likelihood function of a censored sample were obtained by LeCam [5] also.

3. A related estimate. Now expand $n^{-\frac{1}{2}}$ ($\partial \log L_n/\partial\theta_r$, $r = 1, 2$, around $x_u = F^{-1}(u/(n+1)), \cdots, x_v = F^{-1}(v/(n+1))$ in the same manner as (5) and let L_{rn}^* be the linear terms corresponding to the L_{rn}^0 in (6). Using the Lemma in [9] it can be proved that

$$\lim_{n\to\infty} n^{-\frac{1}{2}}(\partial \log L_n/\partial\theta_r)_{F^{-1}(u/(n+1)),\cdots,F^{-1}(v/(n+1))} = 0, \qquad r = 1, 2.$$

Then it follows from the same arguments as those of the preceding section that $(n^{-\frac{1}{2}}L_{1n}^*, n^{-\frac{1}{2}}L_{2n}^*)$ converges in distribution to $N((0, 0), \|I_{rs}(p, q)\|)$ and

$$En^{-\frac{1}{2}}L_{rn}^* = o(1), \quad r = 1, 2, \quad \text{as} \quad n \to \infty. \tag{10}$$

Let $-n\,\|J_{rs}^*(n)\|$ be the matrix whose elements are the coefficients of (θ_1, θ_2) in

$$\theta_2 L_{rn}^* = 0, \qquad r = 1, 2. \tag{11}$$

We replace $n\,\|J^*_{rs}(n)\|$ in (11) by $n\,\|I_{rs}(p, q)\|$ and denote the new equations by $\theta_2 L^{\#}_{rn} = 0$, $r = 1, 2$. Then by using the Lemma of [9] again (see also [10], p. 127) it can be shown that

$$\lim_{n\to\infty} n^{\frac{1}{2}}(\|J^*_{rs}(n)\| - \|I_{rs}(p, q)\|) = 0. \tag{12}$$

By (12) we see that $(n^{-\frac{1}{2}}L^{\#}_{1n}, n^{-\frac{1}{2}}L^{\#}_{2n})$ converges in distribution to $N((0, 0), \|I_{rs}(p, q)\|)$. Let $(\theta_1^{\#}, \theta_2^{\#})$ be the solution of $L^{\#}_{rn} = 0$, $r = 1, 2$. Then

$$n^{\frac{1}{2}} \left\| \begin{matrix} \theta_1^{\#} - \theta_1 \\ \theta_2^{\#} - \theta_2 \end{matrix} \right\| = \theta_2 \|I_{rs}(p, q)\|^{-1} \left\| \begin{matrix} n^{-\frac{1}{2}} L^{\#}_{1n} \\ n^{-\frac{1}{2}} L^{\#}_{2n} \end{matrix} \right\| \tag{13}$$

converges in distribution to $N((0, 0), \theta_2^2\|I_{rs}(p, q)\|^{-1})$. If the expectation is taken on both sides of (13), it follows from (10) and (12) that the biasedness of $(\theta_1^{\#}, \theta_2^{\#})$ is $o(n^{-\frac{1}{2}})$. $(\theta_1^{\#}, \theta_2^{\#})$ has the same form as the estimates in [10] and [3].

It can be seen that we can extend (θ_1^0, θ_2^0) and $(\theta_1^{\#}, \theta_2^{\#})$ to the Type II censored sample in which only the order statistics with ranks lying between and including $[np_j]$ and $[nq_j]$ are available, where $0 < p_1 < q_1 < p_2 < q_2 < \cdots < p_k < q_k < 1$, k a fixed integer.

The computation of (θ_1^0, θ_2^0) is more complicated than that of $(\theta_1^{\#}, \theta_2^{\#})$, but (θ_1^0, θ_2^0) has the superiority that it is strictly unbiased, which is important when n is not large.

4. Acknowledgment. The author wishes to thank the referee and Mr. E. R. Mead for their valuable comments.

REFERENCES

[1] Bickel, P. J. (1965). Some contributions to the theory of order statistics. Mimeograph Notes, University of California. See also *Ann. Math. Statist.* **36** 733 (abstract).

[2] Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables.* Wiley, New York.

[3] Chernoff, H., Gastwirth, J. L. and Johns, M. V. Jr. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.* **38** 52–71.

[4] Halperin, M. (1952). Maximum likelihood estimation in truncated samples. *Ann. Math. Statist.* **23** 226–238.

[5] LeCam, L. (1958). Les propriètés asymptotiques des solutions des Bayes. *Publ. Inst. Statist. Univ. Paris* **7** 17–35.

[6] Plackett, R. L. (1958). Linear estimation from censored data. *Ann. Math. Statist* **29** 351–360.

[7] Sarhan, A. E. and Greenberg, B. C. (1962). *Contributions to Order Statistics.* Wiley, New York.

[8] Särndal, C. K. (1962). *Information from Censored Samples.* Almqvist & Wiksell, Uppsala.

[9] Weiss, L. (1963). On the asymptotic distribution of an estimate of a scale parameter. *Naval Res. Logist. Quart.* **10** 1–9.

[10] Weiss, L. (1964). On estimating location and scale parameters from truncated samples. *Naval Res. Logist. Quart.* **11** 125–133.