ESTIMATION OF THE MEAN AND STANDARD DEVIATION BY ORDER STATISTICS. PART III

A. E. Sarhan

University of North Carolina

1. Summary and Introduction. In a previous work [7], the mean and standard deviation were estimated by arranging all the sample elements in ascending order and taking the best linear combination of them. We will use here the same principle to estimate the mean and standard deviation of certain populations from singly and doubly censored samples.

Censored samples may be considered as truncated samples having a known number of unmeasured (missing) observations, i.e., those in which the total number of sample elements is known, but measurements on some of which are lacking.

In life testing, fatigue testing, and in other tests of a destructive nature, we have n items drawn at random from some population which when subjected to a test, fail in order of time. To save time and/or items, it is often required to stop the experiment (to censor the sample) after recording the first r $(<n)$ observations. This is a censored sample from the right.

Again, censored samples are found frequently in biological data where some of the observations in a sample are either below or above a limit in the measure used. The values beyond this limit are believed to form a continuation of the scale of measurement but are unmeasurable in the experiment [6]. For example, in experimental biology, n samples from each animal are tested for antibodies after a certain period of time. Only r of these samples contain measurable amounts while $(n - r)$ of the animals develop the antigen at a level too low for measurement by the prevailing technique. This is a censored sample from the left.

In fact, the estimation of the mean and standard deviation based on the linear combination of all sample elements is a special case, and the general one is considered here.

Censored samples were considered recently in the work of Ipsen [6], Walsh [8], Hald [4], Gupta [3], Cohen [1], Halperin [5], and Epstein et al [2].

2. Rectangular population. The frequency distribution of the rectangular population is

$$f(y) = \frac{1}{\theta_2}, \qquad \theta_1 - \theta_2/2 \leqq y \leqq \theta_1 + \theta_2/2. \tag{2.1}$$

Consider the case where the observations on the smallest r_1 and largest r_2

Received August 23, 1954.

sample elements are missing and those only on the middle $n - r_1 - r_2$ sample elements are known, then we will have

$$(2.2)\quad V^{-1} = (n+1)(n+2)\begin{bmatrix} \frac{r_1+2}{r_1+1} & -1 & 0 \cdots & & 0 \\ & 2 & -1 \cdots & & 0 \\ & & 2 \cdots & & 0 \\ & & \cdots & & \cdots \\ & & & & \frac{n+1}{(n-r_2)(r_2+1)} + \frac{n-r_2-1}{n-r_2} \end{bmatrix}$$

$$(A'V^{-1}A)^{-1} = \frac{1}{(n+2)(n-r_1-r_2-1)}$$

$$(2.3)\quad \begin{bmatrix} (r_1+1)(n-2r_2-1)+(r_2+1)(n-2r_1-1) & -\frac{r_1-r_2}{2} \\ -\frac{r_1-r_2}{2} & r_1+r_2+2 \end{bmatrix},$$

and

$$(A'V^{-1}A)^{-1}A'V^{-1} = \frac{(n+1)}{(n-r_1-r_2-1)}$$

$$(2.4)\quad \begin{bmatrix} \frac{(n-2r_2-1)}{2(n+1)} & 0 \cdots 0 & \frac{n-2r_1-1}{2(n+1)} \\ -1 & 0 \cdots 0 & 1 \end{bmatrix}.$$

Hence,

$$(2.5)\quad \theta_1^* = \frac{1}{2(n-r_1-r_2-1)}\left[(n-2r_2-1)y_{(r_1+1/n)} + (n-2r_1-1)y_{(n-r_2/n)}\right]$$

and

$$(2.6)\quad \theta_2^* = \frac{(n+1)}{(n-r_1-r_2-1)}\left[y_{(n-r_2/n)} - y_{(r_1+1/n)}\right].$$

From (2.3) the variances of the estimates are

$$(2.7)\quad V(\theta_1^*) = \frac{(r_1+1)(n-2r_2-1)+(r_2+1)(n-2r_1-1)}{4(n+1)(n+2)(n-r_1-r_2-1)}\theta_2^2$$

and

$$(2.8)\quad V(\theta_2^*) = \frac{r_1+r_2+2}{(n+2)(n-r_1-r_2-1)}\theta_2^2.$$

The relative efficiencies are the following:

Relative efficiency of θ_1^*

(2.9)
$$= \frac{2(n - r_1 - r_2 - 1)}{(r_1 + 1)(n - 2r_2 - 1) + (r_2 + 1)(n - 2r_1 - 1)};$$

(2.10)
$$\text{relative efficiency of } \theta_2^* = \frac{2(n - r_1 - r_2 - 1)}{(n - 1)(r_1 + r_2 + 2)}.$$

The efficiency is calculated relative to the best linear estimate using all the sample elements.

As a special case, if $r_2 = 0$, i.e., the observations on the smallest r_1 sample elements are missing and we know only the largest $(n - r_1)$ observations, we will get

(2.11)
$$\theta_1^* = \frac{1}{2(n - r_1 - 1)} [(n - 1)y_{(r_1+1/n)} + (n - 2r_1 - 1)y_{(n/n)}]$$

and

(2.12)
$$\theta_2^* = \frac{(n + 1)}{(n - r_1 - 1)} [y_{(n/n)} - y_{(r_1+1/n)}]$$

with variances

(2.13)
$$V(\theta_1^*) = \frac{nr_1 + 2n - 3r_1 - 2}{4(n + 1)(n + 2)(n - r_1 - 1)} \theta_2^2$$

and

(2.14)
$$V(\theta_2^*) = \frac{(r_1 + 2)}{(n + 2)(n - r_1 - 1)} \theta_2^2.$$

Similarly, for the other special case, i.e., where the observations on the largest r_2 sample elements are missing and we have only those on the smallest $(n - r_2)$, we will have

(2.15)
$$\theta_1^* = \frac{1}{2(n - r_2 - 1)} [(n - 2r_2 - 1)y_{(1/n)} + (n - 1)y_{(n-r_2/n)}]$$

and

(2.16)
$$\theta_2^* = \frac{n + 1}{(n - r_2 - 1)} [y_{(n-r_2/n)} - y_{(1/n)}]$$

with variances

(2.17)
$$V(\theta_1^*) = \frac{(nr_2 + 2n - 3r_2 - 2)}{4(n + 1)(n + 2)(n - r_2 - 1)} \theta_2^2$$

and

$$V(\theta_2^*) = \frac{(r_2 + 2)}{(n + 2)(n - r_2 - 1)} \theta_2^2. \tag{2.18}$$

The results in this case and in the previous one are related. If the smallest r_1 observations are missing and $(n - r_1)$ largest are known, then we can use the same estimates (2.15 and 2.16) for estimating θ_1^* and θ_2^* by letting $y_{(1/n)} > y_{(2/n)} > \cdots > y_{(n-r_1/n)}$. Then the coefficients for constructing the best linear estimate of θ_1 will be identical with those given in (2.15). For the case of θ_2^*, the coefficients will be numerically the same but with opposite sign. The variances will be the same if r_2 is replaced by r_1. This, of course, applies for all symmetric distributions.

In particular, if we have

$$f(y) = \frac{1}{\theta_2}, \qquad 0 \leqq y \leqq \theta_2 \tag{2.19}$$

and the smallest r_1 and largest r_2 observations are missing, we will get

$$\theta_2^* = \frac{n+1}{n - r_2} y_{(n-r_2/n)} \tag{2.20}$$

with variance

$$V(\theta_2^*) = \frac{r_2 + 1}{(n - r_2)(n + 2)} \theta_2^2. \tag{2.21}$$

The relative efficiency is

$$\text{relative efficiency of } \theta_2^* = \frac{(n - r_2)}{n(r_2 + 1)}. \tag{2.22}$$

If the largest r_2 observations only are missing, we will get results exactly as above.

Furthermore, if the smallest r_1 observations are missing, we will get results exactly as those obtained by using all the sample elements, i.e.,

$$\theta_2^* = \frac{n+1}{n} y_{(n/n)}, \tag{2.23}$$

$$V(\theta_2^*) = \frac{1}{n(n+2)} \theta_2^2. \tag{2.24}$$

The efficiency of the estimate in this case is 1 and there will be no effect of the missing values on the estimate.

3. Other symmetric distributions. The estimates of the mean and standard deviation of some other symmetric distributions, their variances and their relative efficiencies from singly and doubly censored samples are also worked out. These are only worked for samples up to size 5 and tabulated in tables (I, II, III, IV). These distributions are,

(3.1) u-shaped $$f(y) = \frac{3(y - \theta_1)^2}{2\theta_2^3}, \qquad \theta_1 - \theta_2 \leqq y \leqq \theta_1 + \theta_2$$

(3.2) rectangular $$f(y) = \frac{1}{\theta_2}, \qquad \theta_1 - \tfrac{1}{2}\theta_2 \leqq y \leqq \theta_1 + \tfrac{1}{2}\theta_2$$

(3.3) parabolic $$f(y) = \frac{6(y - \theta_1 + \frac{1}{2}\theta_2)(\theta_1 + \frac{1}{2}\theta_2 - y)}{\theta_2^3}, \qquad \theta_1 - \tfrac{1}{2}\theta_2 \leqq y \leqq \theta_1 + \tfrac{1}{2}\theta_2$$

(3.4) triangular $$f(y) = \frac{4}{\theta_2^2}\left(\tfrac{1}{2}\theta_2 - |y - \theta_1|\right), \qquad |y - \theta_1| \leqq \tfrac{1}{2}\theta_2$$

(3.5) double exponential $$f(y) = \frac{1}{2\sigma} e^{-(|y-\mu|)/\sigma}, \qquad -\infty \leqq y \leqq \infty.$$

The values for the estimates of the normal distribution are also given in order to be used in the discussion.

4. Discussion. Table I is constructed to give the coefficients of the best linear estimate of the mean and standard deviation in singly censored samples (from the right) in different symmetric populations up to samples of size 5. If the sample is censored from the left, i.e., the smallest r_1 observations are missing and the $(n - r_1)$ largest are known, we can use the same table for estimating the mean and standard deviation of the given populations by letting $y_{(1/n)} > y_{(2/n)} > \cdots > y_{(n-r_1/n)}$. Then the coefficients for constructing the best linear estimate of the mean will be identical with these given in table I. For the case of standard deviation the coefficients will be numerically the same, but with opposite signs. The table shows that:

(1) The coefficients of the largest known observation (i.e., $y_{(n-r_2/n)}$) in the estimate of the mean are greater than the corresponding coefficients if the sample was not censored. This is true for all values of n and in different populations. In fact, the coefficient of $y_{(n-r_2/n)}$ is greater than that of $y_{(n)}$ for the same population and sample size except in the case of the u-shaped distribution for $n = 3$, $r_2 = 1$.

(2) For a fixed n, as the number of the missing observations (r_2) increases, the coefficients of ($y_{(n-r_2/n)}$) increase while those of the smallest observation decrease towards zero and then take gradually decreasing negative values. An interesting case is that of the coefficients in the estimate of the mean of the rectangular population. For fixed n, as r_2 increases, the coefficient of the largest known observation increases and that of the smallest one decreases. When $2r_2 = n - 1$, the smallest element will have zero weight. Hence in this case, the mean of the rectangular population is estimated by the largest known observation.

(3) The coefficients of ($y_{(n-r_2/n)}$) in the estimate of the mean for samples of

TABLE I

Coefficients in the best linear estimate of the mean and standard deviation based on the order statistic $y_{(i)}$ from singly censored samples (from the right) in different populations of size n, for the mean $\theta_1^* = \sum_{i=1}^{n-r_2} \beta_{1i} y_{(i)}$, for the standard deviation $\sigma^* = \sum_{i=1}^{n-r_2} \beta_{2i} y_{(i)}$

n	r_2	Population	β_{11}	β_{12}	β_{13}	β_{14}	β_{21}	β_{22}	β_{23}	β_{24}
5	1	U-shaped	.4387230	−.0434338	−.0830626	.6877734	−.5938969	.0330982	−.0488798	.6096784
		Rectangular	.3333333	0	0	.6666667	−.5773502	0	0	.5773502
		Parabolic	.2396929	.0785569	.0889268	.5930233	−.5665168	−.0403093	.0218717	.5849544
		Triangular	.1840550	.1122442	.1777640	.5259371	−.5586542	−.0723442	.0360423	.5949561
		Normal	.12516	.18305	.21472	.47708	−.51173	−.16678	.02740	.65111
		D. Expon.	.0114148	.2163079	.5243191	.2479582	−.4331049	−.4191182	.0037771	.8484460
	2	U-shaped	.0392274	−.0653693	1.0261419		−.9480304	.0136529	.9343775	
		Rectangular	0	0	1.0000000		−.8660253	0	.8660253	
		Parabolic	−.0367471	.0789384	.9578087		−.8390091	−.0399229	.8789321	
		Triangular	−.0449818	.1011039	.9438779		−.8177461	−.0849466	.9026921	
		Normal	−.06377	.14983	.91395		−.76958	−.21212	.98170	
		D. Expon.	−.0664881	.1666260	.8998621		−.6655505	−.6233335	1.2888840	
	3	U-shaped	−1.5000174	2.5000174			−2.3496132	2.3496132		
		Rectangular	−1.0000000	2.0000000			−1.7320506	1.7320506		
		Parabolic	−.8709678	1.8709678			−1.6045191	1.6045191		
		Triangular	−.8013102	1.8013102			−1.5410202	1.5410202		
		Normal	−.74111	1.74111			−1.49713	1.49713		
		D. Expon.	−.5641141	1.5641141			−1.3924746	1.3924746		
4	1	U-shaped	.3332086	−.0923711	.7591625		−.7293108	−.0087866	.7380974	
		Rectangular	.2500000	0	.7500000		−.7216877	0	.7216877	
		Parabolic	.1893850	.1025250	.7080900		−.7223786	−.0190849	.7414635	
		Triangular	.1523683	.1701062	.6775255		−.7205857	−.0423232	.7629089	
		Normal	.11607	.24084	.64310		−.69713	−.12682	.82395	
		D. Expon.	.0662669	.3333287	.6004044		−.7332133	−.2128959	.9461092	
	2	U-shaped	−.6429103	1.6429103			−1.6783682	1.6783682		
		Rectangular	−.5000000	1.5000000			−1.4233755	1.4233755		
		Parabolic	−.4550565	1.4550565			−1.3971840	1.3971840		
		Triangular	−.4291448	1.4291448			−1.3753825	1.3753825		
		Normal	−.40555	1.40555			−1.36544	1.36544		
		D. Expon.	−.3300458	1.3300458			−1.3577955	1.3577955		
3	1	U-shaped	.5400000	.4600000			−.4182840	.4182840		
		Rectangular	.0000000	1.0000000			−1.1547004	1.1547004		
		Parabolic	.0000000	1.0000000			−1.1594426	1.1594426		
		Triangular	.0000000	1.0000000			−1.1664236	1.1664236		
		Normal	.0000000	1.0000000			−1.18164	1.18164		
		D. Expon.	.0000000	1.0000000			−1.2563648	1.2563648		

TABLE II

Variances and efficiencies of the best linear estimates of the mean (θ_1^*) and standard deviation (σ^*) from singly censored samples in different populations ($\sigma = 1$)

n	r_2	Population	Variance of θ_1^*	Efficiency	Variance of σ^*	Efficiency
5	1	U-shaped	.2245995	30.74	.1674434	28.08
		Rectangular	.2380952	59.99	.1428571	50.00
		Parabolic	.2363240	75.75	.1572840	58.84
		Triangular	.2309568	83.74	.1719055	62.80
		Normal	.21772	91.86	.19476	68.45
		D. Expon.	.1586098	99.88	.3097230	73.88
	2	U-shaped	.7209443	9.37	.5574709	8.43
		Rectangular	.4285714	33.33	.2857142	25.00
		Parabolic	.3602716	49.60	.2778808	33.00
		Triangular	.3184536	60.73	.2838755	38.03
		Normal	.28393	62.80	.31809	41.91
		D. Expon.	.1724911	91.85	.4634527	49.37
	3	U-shaped	2.2113788	3.27	1.7935410	2.62
		Rectangular	1.0000000	14.28	.7142855	10.00
		Parabolic	.7920637	22.60	.6414834	14.43
		Triangular	.7002408	27.62	.6330318	17.05
		Normal	.61123	32.72	.69571	19.16
		D. Expon.	1.2743320	2.24	2.8481830	8.03
4	1	U-shaped	.4119197	31.07	.3482922	27.42
		Rectangular	.3500000	57.14	.2500000	44.44
		Parabolic	.3261310	71.00	.2594385	51.49
		Triangular	.3090288	79.07	.2737111	55.32
		Normal	.28701	87.10	.30208	59.60
		D. Expon.	.1860330	98.23	.3339501	89.42
	2	U-shaped	1.4622735	8.75	1.3411707	7.12
		Rectangular	.8000000	25.00	.6666667	16.67
		Parabolic	.6573040	35.23	.6225662	21.46
		Triangular	.5832768	41.89	.6114363	24.37
		Normal	.51299	48.73	.67303	26.75
		D. Expon.	.3335692	62.29	.9457112	31.58
3	1	U-shaped	.5340045	46.84	.3204021	14.69
		Rectangular	.6000000	50.00	.6000000	33.33
		Parabolic	.5324677	60.25	.5887268	37.73
		Triangular	.4928568	66.83	.5986393	40.34
		Normal	.44867	74.29	.63783	43.19
		D. Expon.	.3194197	92.28	.8760051	49.33

Efficiency is calculated relative to the best linear estimate using all sample elements.

TABLE III

Coefficients in the best linear estimate of the mean and standard deviation based on the order statistic $y_{(i)}$ from doubly censored samples in different populations of size n, for the mean $\theta_1^* = \sum_{i=r_1+1}^{n-r_2} \beta_{1i} y_{(i)}$, for the standard deviation $\sigma^* = \sum_{i=r_1+1}^{n-r_2} \beta_{2i} y_{(i)}$

n	r_1	r_2	Population	β_{12}	β_{13}	β_{14}	β_{22}	β_{23}	β_{24}
5	1	1	U-shaped	.5595857	−.1191714	.5595857	−.7832027	0	.7832027
			Rectangular	.5000000	0	.5000000	−.8660253	0	.8660253
			Parabolic	.4509135	.0981730	.4509135	−.9211128	0	.9211128
			Triangular	.4051809	.1896383	.4051809	−.9614817	0	.9614817
			D. Expon.	.2377907	.5244186	.2377907	−1.234223	0	1.234223
		2	U-shaped		1.000000	0		−1.5664142	1.5664142
			Rectangular		1.000000	0		−1.7320506	1.7320506
			Parabolic		1.000000	0		−1.8422256	1.8422256
			Triangular		1.000000	0		−1.9229645	1.9229645
			D. Expon.		1.000000	0		−2.4684456	2.4684456
4	1	1	U-shaped		−1.3053363	1.3053363	.5000000	.5000000	
			Rectangular		−1.4233755	1.4233755	.5000000	.5000000	
			Parabolic		−1.5351864	1.5351864	.5000000	.5000000	
			Triangular		−1.6024692	1.6024692	.5000000	.5000000	
			D. Expon.		−2.0571088	2.0571088	.5000000	.5000000	

size 4 and 5 are the largest in the case of the *u*-shaped and decrease gradually for the other distributions in an order indicated by their arrangement in the table. The coefficients of the smallest observation for the different populations change in the same order.

(4) The coefficients of the largest known element ($y_{n-r_2/n}$) and of the smallest observation in the best linear estimate of the standard deviation for all populations are greater and smaller respectively than the corresponding sample elements if the samples were not censored. Also, for a fixed sample size and for the same distribution as r_2 increases, the coefficient of the largest known element becomes larger and the coefficient of the smallest observation becomes smaller.

Table II is constructed to give the variances and the efficiencies of the estimates of the mean and standard deviation of the different symmetric populations from singly censored samples up to $n = 5$. The efficiencies are calculated relative to the best linear estimate based on all the sample elements.

This table shows that:

(5) For a fixed n, as r_2 increases, the efficiency of both the estimate of the mean and standard deviation decreases.

(6) For every fixed value of n and r_2, the efficiency of both the estimates are low for the *u*-shaped distribution, increases in the rectangular, then the parabolic, the triangular, the normal and greatest in the case of double exponential.

TABLE IV

Variances and efficiencies of the best linear estimate of the mean (θ_1^*) and standard deviation (σ^*) from doubly censored samples in different populations (σ = 1)

n	r_1	r_2	Population	Variance of θ_1^*	Efficiency	Variance of σ^*	Efficiency
5	1	1	U-shaped	.3261912	20.71	.3536064	13.30
			Rectangular	.2857140	49.99	.2857140	25.00
			Parabolic	.2620758	68.30	.3013740	30.71
			Triangular	.2471006	78.27	.3206393	33.67
			D. Expon.	.1587055	99.82	.4386776	52.16
		2	U-shaped	.7217100	9.36	1.1284812	4.17
			Rectangular	.4285714	33.33	.7142857	10.00
			Parabolic	.3611096	49.57	.7146330	12.96
			Triangular	.3198053	60.48	.7300379	14.79
			D. Expon.	.1755904	90.23	.8935550	25.61
4	1	1	U-shaped	.5343228	23.95	.9346768	10.22
			Rectangular	.4000000	50.00	.6666667	16.67
			Parabolic	.354731	65.27	.6755588	19.78
			Triangular	.3278568	74.53	.6948200	21.79
			D. Expon.	.2100694	98.91	.4256344	70.16

Efficiency is calculated relative to the best linear estimate using all sample elements.

In the latter distribution, if the middle sample element is included among the missing observations, the efficiency becomes very low and loses its relatively high standing among the symmetric distributions.

(7) The relative efficiency of the estimate of standard deviation of a given population is less than the corresponding efficiency of the estimate of the mean of the same population. This indicates that the estimate of standard deviation is affected more than the estimate of the mean by missing observations.

Table III is constructed to give the coefficients in the best linear estimate of the mean and standard deviation from doubly censored samples while table IV gives the variances of these estimates and their relative efficiencies. Table III shows that:

(8) The coefficients of the largest known observation vary in the different populations in the order indicated previously. The case of the normal distribution is not indicated, but one would expect that for different values of n, r_1, r_2 the coefficient of the estimates will take values between those of the triangular and those of the double exponential.

(9) When the middle observation and the next to it in order are the only observations known, then only the middle one is used to estimate the mean. Table IV gives the same picture as table II and one would expect that the efficiencies of the estimates of the normal population will lie between those of the triangular and those of the double exponential.

5. Exponential population. The frequency distribution of the exponential population is

$$f(y) = \frac{1}{\sigma} e^{-(y-\mu)/\sigma}, \qquad \mu \leqq y \leqq \infty. \tag{5.1}$$

Consider the case where the observations on the smallest r_1 and largest r_2 elements are missing, then we will have,

(5.2) $V^{-1} =$

$$\begin{bmatrix} \frac{1}{\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2}} + (n-r_1-1)^2 & -(n-r_1-1)^2 & 0 & \cdots & 0 \\ & (n-r_1-1)^2 + (n-r_1-2)^2 & -(n-r_1-2)^2 & \cdots & 0 \\ & & (n-r_1-2)^2 + (n-r_1-3)^2 & \cdots & 0 \\ & & & \cdots & \cdots \\ & & & & (r_2+1) \end{bmatrix}$$

and

(5.3) $(A'V^{-1}A)^{-1} = \dfrac{1}{(n-r_1-r_2-1)}$

$$\begin{bmatrix} \left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right]^2 + (n-r_2-r_1-1) \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2} & -\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)} \\ -\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)} & 1 \end{bmatrix}.$$

Hence, the estimates and their variances are given by

$$\begin{aligned} \mu^* = C\Bigg[&\left\{\frac{1}{C} + (n-r_1) \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right\} y_{(r_1+1/n)} \\ &- r_2 \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)} y_{(n-r_2/n)} - \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)} \sum_{i=r_1+1}^{n-r_2} y_{(i/n)}\Bigg], \end{aligned} \tag{5.4}$$

$$\sigma^* = C\left[\sum_{i=r_1+1}^{n-r_2} y_{(i/n)} - (n-r_1)y_{(r_1+1/n)} + r_2 y_{(n-r_2/n)}\right], \tag{5.5}$$

$$V(\mu^*) = \left\{C\left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right]^2 + \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2}\right\} \sigma^2, \tag{5.6}$$

$$V(\sigma^*) = C\sigma^2. \tag{5.7}$$

The estimate of the mean is

$$\begin{aligned} C\Bigg[&\frac{1}{C} y_{(r_1+1/n)} + \left\{\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)} - 1\right\} \Bigg\{(n-r_1)y_{(r_1+1/n)} \\ &- \sum_{i=r_1+1}^{n-r_2} y_{(i)} - r_2 y_{(n-r_2/n)}\Bigg\}\Bigg] \end{aligned} \tag{5.8}$$

with variance

$$(5.9)\quad \sigma^2 C\left\{\left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right]^2 + \frac{1}{C}\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2} - 2\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)} + 1\right\}$$

where

$$(5.10)\quad C = \frac{1}{(n - r_1 - r_2 - 1)}.$$

The relative efficiencies of the estimates are

relative efficiency of μ^*

$$(5.11)\quad = \frac{1}{n(n-1)}\Big/\left[C\left\{\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right\}^2 + \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2}\right],$$

$$(5.12)\quad \text{rel. eff. of } \sigma^* = \frac{1}{(n-1)}\Big/ C,$$

$$(5.13)\quad \text{rel. eff. of } (\mu^* + \sigma^*) = \frac{1}{n}\Big/ C\left\{\left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right]^2 + \frac{1}{C}\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2} - 2\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)} + 1\right\}.$$

If we have the smallest r_1 observations only missing, we will get

$$(5.14)\quad \mu^* = \frac{1}{(n - r_1 - 1)}\left[\left\{(n - r_1)y_{(r_1+1/n)} - \sum_{i=r_1+1}^{n} y_{(i/n)}\right\} \cdot \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)} + (n - r_1 - 1)y_{(r_1+1/n)}\right]$$

and

$$(5.15)\quad \sigma^* = \frac{1}{(n - r_1 - 1)}\left[\sum_{i=r_1+1}^{n} y_{(i/n)} - (n - r_1)y_{(r_1+1/n)}\right]$$

with variances

$$(5.16)\quad V(\mu^*) = \frac{\sigma^2}{(n - r_1 - 1)} \cdot \left[\left(\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right)^2 + (n - r_1 - 1)\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2}\right]$$

and

$$(5.17)\quad V(\sigma^*) = \frac{\sigma^2}{(n - r_1 - 1)}.$$

The estimate of the mean will be

$$(5.18)\quad \frac{1}{(n-r_1-1)}\left[\left\{1-\sum_{i=1}^{r_1+1}\frac{1}{(n-i+1)}\sum_{i=r_1+1}^{n}y_{(i/n)}\right\} + \left\{(n-r_1)\sum_{i=1}^{r_1+1}\frac{1}{(n-i+1)}-1\right\}y_{(r_1+1/n)}\right]$$

with variance

$$(5.19)\quad \frac{\sigma^2}{(n-r_1-1)}\left[\left\{\sum_{i=1}^{r_1+1}\frac{1}{(n-i+1)}\right\}^2+(n-r_1-1)\cdot\sum_{i=1}^{r_1+1}\frac{1}{(n-i+1)^2}-2\sum_{i=1}^{r_1+1}\frac{1}{(n-i+1)}+1\right].$$

For the case where the observations on the largest r_2 sample elements only are missing, we will get

$$(5.20)\quad \mu^* = \frac{1}{n(n-r_2-1)}\left[n(n-r_2)y_{(1/n)}-\sum_{i=1}^{n-r_2}y_{(i/n)}-r_2y_{(n-r_2/n)}\right],$$

$$(5.21)\quad \sigma^* = \frac{1}{(n-r_2-1)}\left[\sum_{i=1}^{n-r_2}y_{(i/n)}-ny_{(1/n)}+r_2y_{(n-r_2/n)}\right]$$

with variances

$$(5.22)\quad V(\mu^*) = \frac{n-r_2}{n^2(n-r_2-1)}\sigma^2$$

and

$$(5.23)\quad V(\sigma^*) = \frac{\sigma^2}{(n-r_2-1)}.$$

The estimate of the mean is

$$(5.24)\quad \frac{1}{n(n-r_2-1)}\left[(n-1)\sum_{i=1}^{n-r_2}y_{(i/n)}-nr_2y_{(1/n)}+r_2(n-1)y_{(n-r_2/n)}\right]$$

with variance

$$(5.25)\quad \frac{n^2-n-r_2}{n^2(n-r_2-1)}\sigma^2.$$

In particular, if the distribution is

$$(5.26)\quad f(y) = \frac{1}{\sigma}e^{-y/\sigma}, \qquad 0 \leqq y \leqq \infty$$

and the observations in the smallest r_1 and the largest r_2 are missing, we will get

$$(5.27)\quad \sigma^* = \frac{1}{K}\left[\left\{\sum_{i=1}^{r_1+1}\frac{1}{(n-i+1)}\Big/\sum_{i=1}^{r_1+1}\frac{1}{(n-i+1)^2}-(n-r_1)\right\}y_{(r_1+1/n)}+r_2y_{(n-r_2/n)}+\sum_{i=r_1+1}^{n-r_2}y_{(i/n)}\right]$$

with variance

$$V(\sigma^*) = \frac{\sigma^2}{K} \tag{5.28}$$

where

$$K = \left[\left\{\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right\}^2 \Big/ \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2} + (n - r_1 - r_2 - 1)\right] \tag{5.29}$$

and

$$\text{rel. eff.}\ (\sigma^*) = K/n \tag{5.30}$$

where K is given by (5.29).

If only the smallest r_1 sample elements are missing, we will get

$$\begin{aligned}\sigma^* = &\left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2} \Big/ \left\{\left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right]^2 \right.\right.\\ &\left.\left. + (n - r_1 - 1)\left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2}\right]\right\}\right]\left[\left\{\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right.\right.\\ &\left.\left. \Big/ \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2} - (n-r_1)\right\} y_{(r_1+1/n)} + \sum_{i=r_1+1}^{n} y_{(i/n)}\right],\end{aligned} \tag{5.31}$$

$$\begin{aligned}V(\sigma^*) = &\ \sigma^2 \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2}\\ &\Big/ \left\{\left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)}\right]^2 + (n - r_1 - 1) \sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2}\right\}.\end{aligned} \tag{5.32}$$

For the case where the largest r_2 items only are missing, we will get

$$\sigma^* = \frac{\sum_{i=1}^{n-r_2} y_{(i/n)} + r_2 y_{(n-r_2/n)}}{n - r_2} \tag{5.33}$$

and

$$V(\sigma^*) = \frac{\sigma^2}{(n - r_2)}. \tag{5.34}$$

The results (5.33 and 5.34) are exactly the same as the maximum likelihood estimates obtained by Epstein and Sobel [2]. It has also been shown that the estimate obtained has the minimum variance.

Table (VI) is constructed to give the coefficients of the mean and standard deviation of the exponential population (5.1) from singly and doubly censored samples. In singly censored samples, the coefficients of the largest known observations in the estimates of the mean and standard deviation increase as r_2 increases—for fixed n—and that of the smallest observations decrease.

When the sample is censored from the left, as r_1 increases, the reverse situation

TABLE V

Coefficients in the best linear estimate of the mean and standard deviation in singly and doubly censored samples of the skewed distribution (6.1) *for the mean* $\mu^* = \sum_{i=r_1+1}^{n-r_2} \gamma_{1i}\, y_{(i)}$, *for the standard deviation* $\sigma^* = \sum_{i=r_1+1}^{n-r_2} \gamma_{2i}\, y_{(i)}$

n	r_1	r_2	γ_{11}	γ_{12}	γ_{13}	γ_{14}	γ_{15}	γ_{21}	γ_{22}	γ_{23}	γ_{24}	γ_{25}
5	0	1	.2228020	.1041358	.0967091	.5763533		−.5342437	.0063273	−.0605041	.5884204	
	0	2	−.0172304	.1478335	.8693969			−.7793013	.0509398	.7283615		
	0	3	−0.5137645	1.5137645				−1.1952860	1.1952860			
	1	0		.4229770	.1259118	.1064816	.3446296		−.5666606	−.1242249	.0093981	.6814874
	2	0			.7319946	.1199977	.1480077			−.9361914	−.0087097	.9449011
	3	0				2.0497535	−1.0497535				−1.9214554	1.9214554
	1	1		.4899271	.1460117	.3640611			−.9187389	−.1787232	1.0974621	
	1	2		−.5274283	1.5274283				−2.092864	2.092864		
	2	1			1.0938769	−.0938769				−.6520712	.6520712	
4	0	1	.1934078	.1233877	.6832044			−.6565455	−.0987115	.7552570		
	0	2	−.3580357	1.3580357				−1.2661458	1.2661458			
	1	0		.7071773	.1440367	.1487863			−.7609486	−.0361974	.7971460	
	2	0			.5170427	.4829573				−1.5481259	1.5481259	
	1	1		.0858131	.9141869				−1.5228920	1.5228920		
3	0	1	−.6922516	1.6922516				−1.0975877	1.0975877			
	1	0		1.0447763	−.0447763				−1.2450250	1.2450250		

TABLE VI

Coefficients in the best linear estimate of the mean and standard deviation in singly and doubly censored samples of the exponential distribution (5.1) *for the mean* (mean*) $= \sum_{i=r_1+1}^{n-r_2} \delta_{1i}\, y_{(i)}$, *for standard deviation* $\sigma^* = \sum_{i=r_1+1}^{n-r_2} \delta_{2i}\, y_{(i)}$

n	r_1	r_2	δ_{11}	δ_{12}	δ_{13}	δ_{14}	δ_{15}	δ_{21}	δ_{22}	δ_{23}	δ_{24}	δ_{25}
5	0	1	−.0666667	.2666667	.2666667	.5333333		−1.3333333	.3333333	.3333333	.6666667	
	0	2	−.6000000	.4000000	1.2000000			−2.0000000	.5000000	1.5000000		
	0	3	−2.2000000	3.2000000				−4.0000000	4.0000000			
	1	0		.4500000	.1833333	.1833333	.1833333		−1.0000000	.3333333	.3333333	.3333333
	2	0			.7833333	.1083333	.1083333			−1.0000000	.5000000	.5000000
	3	0				1.2833333	−.2833333				−1.0000000	1.0000000
	1	1		.1750000	.2750000	.5500000			−1.5000000	.5000000	1.0000000	
	1	2		−.6500000	1.6500000				−3.0000000	3.0000000		
	2	1			.5666667	.4333333				−2.0000000	2.0000000	
4	0	1	−.1250000	.3750000	.7500000			−1.5000000	.5000000	1.0000000		
	0	2	−1.2500000	2.2500000				−3.0000000	3.0000000			
	1	0		.5833333	.2083333	.2083333			−1.0000000	.5000000	.5000000	
	2	0			1.0833333	−.0833333				−1.0000000	1.0000000	
	1	1		.1666667	.833333				−2.0000000	2.0000000		
3	0	1	−.3333333	1.3333333				−2.0000000	2.0000000			
	1	0		.8333333	.1666667				−1.0000000	1.0000000		

TABLE VII

Variances of the best linear estimate of the mean and standard deviation in singly and doubly censored samples from the skewed distribution (4.1) *and the exponential distribution* (5.1)

n	r_1	r_2	Skewed		Exponential	
			Variance of mean*	Variance of σ*	Variance of mean*	Variance of σ*
5	0	1	.0090718	.1577897	.2533000	.3333333
	0	2	.0124467	.2631055	.3600000	.5000000
	0	3	.0244913	.4627518	.6800000	1.0000000
	1	0	.0100586	.1836198	.2033333	.3333333
	2	0	.0144719	.2911154	.2370830	.5000000
	3	0	.0306681	.6593896	.5438889	1.0000000
	1	1	.0104491	.3564131	.2537500	.5000000
	1	2	.0131271	.9637797	.2605555	1.0000000
	2	1	.0146836	.7282964	.4050000	1.0000000
4	0	1	.0123354	.2678224	.3438000	.5000000
	0	2	.0243523	.6239451	.6250000	1.0000000
	1	0	.0132724	.2739464	.2604166	.5000000
	2	0	.0260043	.6424895	.2152777	1.0000000
	1	1	.0117789	.6891365	.3472222	1.0000000
3	0	1	.0199459	.5958348	.5555555	1.0000000
	1	0	.0214751	.6104800	.3888888	1.0000000

holds true for the coefficients in estimation of the mean. In estimation of the standard deviation in this case, however, the coefficients of the largest known observations increase whereas the smallest known observation is always -1.

In all cases, the coefficient of the largest known observation in the estimate of the mean in samples censored from the right is greater than the coefficients of the smallest known observation in censored samples from the left if the number of missing observations are equal in the two cases.

Table (VII) gives the variances of the estimates of the mean and the standard deviation of the same population. This table shows that the variance of the estimate of the mean varies according to the number of missing observations and to the side from which these are censored, whereas that of the variance of the standard deviation depends only on the total number of missing observations.

6. A skewed distribution. The coefficients in the best linear estimate of the mean and standard deviation of the skewed distribution

$$(6.1)\quad f(y) = \frac{12}{\theta_2}\left(\frac{y-\theta_1}{\theta_2} + \frac{2}{3}\right)^2\left(\frac{1}{3} - \frac{y-\theta_1}{\theta_2}\right), \quad \theta_1 - 2\theta_2/3 \leqq y \leqq \theta_1 + 2\theta_2/3$$

from singly and doubly censored samples up to size 5 are given in table (V).

The same remarks are also to be noted here for the coefficients of the largest or smallest known observations.

Table (VII) gives the variances of the estimates for samples up to size 5 and different values of r_1 and r_2. For singly censored samples, the table shows that the variances of the estimates vary according to the number of the missing observations and to the side from which they are censored.

7. Summary. The best linear estimates of the mean and standard deviation of the rectangular and the exponential populations, their variances and relative efficiencies from singly and doubly censored samples, are given to samples of size n.

The same results are given up to samples of size 5 for some other symmetric distributions. There is a certain pattern in the behavior of the coefficients of the smallest and largest known observations as the number of unknown observations increase in the same population. These coefficients also vary in different populations in a certain order according to their shapes. The effect of the tails of the distribution on the estimates are also considered in the case of the skewed distribution.

It is to be noted that, in some cases where sufficient statistics exist, the best linear estimates can be obtained quickly from them.

Finally, the author wishes to acknowledge the kind help of Dr. B. G. Greenberg, under whose direction this work was done.

REFERENCES

[1] A. C. COHEN, JR., "Estimating mean and variance of normal populations from singly truncated and doubly truncated samples," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 557–569.

[2] B. EPSTEIN AND M. SOBEL, "Life testing," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 486–502.

[3] A. K. GUPTA, "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, Vol. 39 (1952), pp. 260–273.

[4] A. HALD, "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point," *Skan. Aktuarietids.* (1949), pp. 119–134.

[5] M. HALPERIN, "Maximum likelihood estimation in truncated samples," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 226–238.

[6] J. IPSEN, JR., "A practical method of estimating the mean and standard deviation of truncated normal distributions," *Human* Biology, Vol. 21 (1949), pp. 1–16.

[7] A. E. SARHAN, "Estimation of the mean and standard deviation by order statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 317–328.

[8] J. E. WALSH, "Some estimates and tests based on the smallest values in a sample," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 386–397.