It is obvious that $F(p)$ is a continuous function of p with $F(0) < 0$ and $F(1/k(k-1)) > 0$. Hence there exists a p^* with $0 < p^* < 1/k(k-1)$ which is a function of Δ/σ so that $F(p^*) = 0$. Once the Bayes solution relative to $[1 - k(k-1)p, p, \cdots, p]$ has been worked out, it is obvious that to get the Bayes solution relative to $[1 - k(k-1)p^*, p^*, \cdots, p^*]$ it is only necessary to replace q by q_α. If we now substitute $w_i = (\bar{x}_i - \bar{x}_k)/s$ and replace A and B by their values, we find after some simplifications that the Baye's solution relative to $[1 - k(k-1)p^*, p^*, \cdots, p^*]$ reduces to (1) when $\bar{D}_{ij}$ is made to correspond to D_{ij}, $(i, j = 0, 1, 2, \cdots, k; i \neq j \text{ if } i, j > 0)$. Since (1) is an allowable procedure, this proves that it is an optimum one.

REFERENCES

[1] Joyce M. May, "Extended and corrected tables of the upper percentage points of the Studentized range," *Biometrika*, Vol. 39 (1952), pp. 192–193.
[2] Edward Paulson, "An optimum solution to the k sample slippage problem for the normal distribution," *Annals of Math. Stat.*, Vol. 23 (1952), pp. 610–616.
[3] E. S. Pearson and H. O. Hartley, "Biometrika Tables for Statisticians," (Cambridge University Press), Vol. 1 (1954).
[4] K. C. S. Pillai, "On the distribution of Studentized range," *Biometrika*, Vol. 39 (1952), pp. 194–195.
[5] K. V. Ramachandran, "On the Tukey test for the equality of means and the Hartley test for the equality of variances," *Annals of Math. Stat.*, Vol. 27 (1956), pp. 825–831.
[6] S. N. Roy and R. C. Bose, "Simultaneous confidence internal estimation," *Annals of Math. Stat.*, Vol. 24 (1953), pp. 513–536.
[7] Donald R. Truax, "An optimum slippage test for the variances of k normal distributions," *Annals of Math. Stat.*, Vol. 24 (1953), pp. 669–674.
[8] J. W. Tukey, "Allowances for various types of error rates," (unpublished invited address, Blacksburg meeting of the institute of mathematical statistics, March, 1952).
[9] A. Wald, "Statistical decision functions," John Wiley and Sons, 1950.

ESTIMATES OF THE MEAN AND STANDARD DEVIATION OF A NORMAL POPULATION[1]

By W. J. Dixon

University of California, Los Angeles

0. Summary. Several simple estimates of the mean and standard deviation of a normal population are discussed. The efficiencies of these estimates are compared to the sample mean and sample standard deviation and to the best linear unbiased estimates. Little efficiency is lost when simple rather than optimum weights are used.

Received October 29, 1956; revised February 5, 1957.
[1] This research sponsored in part by the Office of Naval Research. It may be reproduced in whole or in part for any purpose of the United States Government.

TABLE I

Several Estimates of Mean of Normal Population with Efficiencies. (Variances to be multiplied by σ^2.)

N	Median		Midrange		$(X_i + X_j)/2$			$\bar{X}_{)1,N(}$		
	Var.	Eff.	Var.	Eff.	i, j	Var.	Eff.	Var.	Eff.	A^*
2	0.500	1.00	0.500	1.00	1, 2	0.500	1.00			
3	0.449	0.743	0.362	0.920	1, 3	0.362	0.920	0.449	0.743	1.000
4	0.298	0.838	0.298	0.838	2, 3	0.298	0.838	0.298	0.838	1.000
5	0.287	0.697	0.261	0.767	2, 4	0.231	0.867	0.227	0.881	0.994
6	0.215	0.776	0.236	0.706	2, 5	0.193	0.865	0.184	0.906	0.992
7	0.210	0.679	0.218	0.654	2, 6	0.168	0.849	0.155	0.922	0.990
8	0.168	0.743	0.205	0.610	3, 6	0.149	0.837	0.134	0.934	0.990
9	0.166	0.669	0.194	0.572	3, 7	0.132	0.843	0.118	0.942	0.990
10	0.138	0.723	0.186	0.539	3, 8	0.119	0.840	0.105	0.949	0.990
11	0.137	0.663	0.178	0.510	3, 9	0.109	0.832	0.0952	0.955	0.991
12	0.118	0.709	0.172	0.484	4, 9	0.100	0.831	0.0869	0.959	0.991
13	0.117	0.659	0.167	0.461	4, 10	0.0924	0.833	0.0799	0.963	0.991
14	0.102	0.699	0.162	0.440	4, 11	0.0860	0.830	0.0739	0.966	0.992
15	0.102	0.656	0.158	0.422	4, 12	0.0808	0.825	0.0688	0.969	0.992
16	0.0904	0.692	0.154	0.392	5, 12	0.0756	0.827	0.0644	0.971	0.993
17	0.0901	0.653	0.151	0.389	5, 13	0.0711	0.827	0.0605	0.973	0.993
18	0.0810	0.686	0.148	0.375	5, 14	0.0673	0.825	0.0570	0.975	0.993
19	0.0808	0.651	0.145	0.363	6, 14	0.0640	0.823	0.0539	0.976	0.993
20	0.0734	0.681	0.143	0.350	6, 15	0.0607	0.824	0.0511	0.978	0.994
∞		0.637		0.000	0.27, 0.73		0.810		1.000	1.000

* A = Var (BLSS)/Var$(\bar{X}_{)1,N(})$.

Since moments of the order statistics are now available for samples of sizes $N \leqq 20$ from normal populations [3] it is a simple matter to find the variances of linear combinations of order statistics. The sample values are denoted $X_1 \leqq X_2 \leqq X_3 \leqq \cdots \leqq X_N$.

1. Estimates of the mean. Table I gives the variance and efficiency of the following estimates of the population mean: (a) median, (b) midrange, (c) mean of best two, and (d) $\bar{X}_{)1,N(} = \sum_{i=2}^{N-1} X_i \,/\, (N - 2)$. The median and midrange are given primarily for comparison purposes, since results are well known. The median is defined as $X_{(N+1)/2}$ for N odd and as $\frac{1}{2}(X_{N/2} + X_{(N+1)/2})$ for N even. The mean of the best two (here "best" is used in the sense of minimum variance) is reported as the small sample equivalent of the estimate commonly used in large samples, the mean of the 27th and 73rd percentiles. It can be seen from Table I that for sample sizes larger than five, the particular ordered observations indicated are not far from the 27th and 73rd percentiles, and efficiencies are close to the asymptotic efficiency (0.810). The efficiency reported for the mean, median, and the mean of best two is the ratio of the variance of the statistic to the variance of the arithmetic mean.

Estimate (d) above is the mean of all observations except the largest and

TABLE II

A linear estimate of the standard deviation. (Variances to be multiplied by σ^2.)

Sample Size	Range			s'			
	k	Var.	Eff.	Estimate	Var.	Eff.	B*
2	0.886	0.571	1.000	$0.8862w$	0.571	1.000	1.000
3	0.591	0.275	0.992	$0.5908w$	0.275	0.992	1.000
4	0.486	0.183	0.975	$0.4857w$	0.183	0.975	0.986
5	0.430	0.138	0.955	$0.4299w$	0.138	0.955	0.966
6	0.395	0.112	0.933	$0.2619(w + w_{(2)})$	0.109	0.957	0.968
7	0.370	0.0949	0.911	$0.2370(w + w_{(2)})$	0.0895	0.967	0.978
8	0.351	0.0829	0.890	$0.2197(w + w_{(2)})$	0.0761	0.970	0.980
9	0.337	0.0740	0.869	$0.2068(w + w_{(2)})$	0.0664	0.968	0.979
10	0.325	0.0671	0.850	$0.1968(w + w_{(2)})$	0.0591	0.964	0.974
11	0.315	0.0616	0.831	$0.1608(w + w_{(2)} + w_{(4)})$	0.0529	0.967	0.977
12	0.307	0.0571	0.814	$0.1524(w + w_{(2)} + w_{(4)})$	0.0478	0.972	0.981
13	0.300	0.0533	0.797	$0.1456(w + w_{(2)} + w_{(4)})$	0.0436	0.975	0.984
14	0.294	0.0502	0.781	$0.1399(w + w_{(2)} + w_{(4)})$	0.0401	0.977	0.985
15	0.288	0.0474	0.766	$0.1352(w + w_{(2)} + w_{(4)})$	0.0372	0.977	0.985
16	0.283	0.0451	0.751	$0.1311(w + w_{(2)} + w_{(4)})$	0.0347	0.975	0.983
17	0.279	0.0430	0.738	$0.1050(w + w_{(2)} + w_{(3)} + w_{(5)})$	0.0325	0.978	0.985
18	0.275	0.0412	0.725	$0.1020(w + w_{(2)} + w_{(3)} + w_{(5)})$	0.0305	0.978	0.986
19	0.271	0.0395	0.712	$0.09939(w + w_{(2)} + w_{(3)} + w_{(5)})$	0.0288	0.979	0.986
20	0.268	0.0381	0.700	$0.10446(w + w_{(2)} + w_{(4)} + w_{(6)})$	0.0272	0.980	0.987

* B = Var (BLSS)/Var(s').

smallest. Interest in this statistic arises when the extreme observations are poorly determined or not available. References [1] and [2] refer to this condition as doubly censored and develop best linear systematic statistics (BLSS) for various amounts of single and double censoring of the sample. The decrease in efficiency of the simple unweighted mean $\bar{X}_{)1,N(}$ compared to the BLSS based on the same observations is not great. In no case is the loss in efficiency more than 1 per cent. This can be noted from the ratio Var(BLSS) / Var($\bar{X}_{)1,N(}$) given in Table I, since this ratio is never less than 0.990. It seems likely that for many applications, one could dispense with the use of unequal weights for the systematic statistics in this case. It can be seen that the efficiency is not greatly affected by the use of coefficients differing greatly from the optimum. The column head "Eff." is efficiency of $\bar{X}_{)1,N(}$ compared with the mean of all observations and is approximately the same for the BLSS.

2. Estimates of standard deviation. The efficiency of the range, w, as an estimate of the standard deviation in small samples, is well known. Similar estimates using additional observations will also give high efficiencies for larger sample sizes. Table II contains the efficiency of the range estimate compared to the unbiased estimate based on the sample standard deviation. The quantity k which satisfies $E(kw) = \sigma$ is tabled for reference. Let us denote the subranges

$X_{N-i+1} - X_i$ by $w_{(i)}$ and $w_{(1)} = w$. The unbiased estimate of the type $s' = k' (\sum w_{(i)})$, where the summation is over the subset of all $w_{(i)}$ which gives minimum variance, is indicated in Table II. The column headed "Eff." refers to the comparison with the unbiased sample standard deviation. The final column gives the ratio of the variance of the best linear systematic statistic as given in [2] to the variance of s'. By examining this ratio we can see that the loss in efficiency due to the use of "zero or one" weights for each range rather than the optimum weights given in [2], is not great.

REFERENCES

[1] 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.

[2] A. E. SARHAN AND B. G. GREENBERG, "Estimation of location and scale parameters by order statistics from singly and doubly censored samples," Part 1, *Ann. Math. Stat.*, Vol. 27 (1956), pp. 427–451.

[3] DAN TEICHROEW, "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 410–426.

THE INDIVIDUAL ERGODIC THEOREM OF INFORMATION THEORY[1]

BY LEO BREIMAN

University of California, Berkeley

1. Introduction. Information theory is largely concerned with stationary stochastic processes $\cdots x_{-1}, x_0, x_1, \cdots$ taking values in a finite "alphabet," $a_1, \cdots, a_s$. In addition, it is usually assumed that the processes are ergodic, that is to say, the shift operator T, defined on the sequence space Ω of the process by shifting each coordinate of a sequence once to the right, is metrically transitive with respect to the probability measure p on Ω.

A question of importance in information theory regarding these processes is the nature and existence, in some sense, of the expression

$$\text{(a)} \qquad \lim_n \left(-\frac{1}{n} \log_2 p(x_0, \cdots, x_{n-1}) \right).$$

In 1948 Shannon [1] showed that for stationary, ergodic Markov chains (a) exists as a limit in probability and is equal to a constant. This limiting constant was termed by Shannon the "entropy" of the process. In 1953 McMillan [2] lifted the restriction to Markov chains and proved that if the process is merely stationary and ergodic, then (a) exists as a limit in L_1 mean and is constant. The purpose of this note is to prove that under the same conditions the limit (a) exists almost surely (a.s.).

Received October 15, 1956.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.