ON SMALL SAMPLES FROM CERTAIN NON-NORMAL UNIVERSES*

By

PAUL R. RIDER
Washington University

INTRODUCTION

The distribution of the ratio

$$z = \frac{\text{mean of sample} - \text{mean of universe}}{\text{standard deviation of sample}}$$

which is of great importance in the theory of small samples, has been derived exactly by theoretical methods for samples of any size from a normal universe.[1] Experimental studies[2] have been

*The writer desires to express his grateful appreciation to the National Research Council, which made possible this study by a grant-in-aid for the assistance of a computer.

[1] See, for example, R. A. Fisher, Applications of "Student's" Distribution, Metron, vol. 5, No. 3 (Dec. 1, 1925), pp. 90-104.

[2] e. g. W. A. Shewhart and F. W. Winters, Small Samples—New Experimental Results, Journal of the American Statistical Association, Vol. 23 (1928), pp. 144-53;

J. Neyman and E. S. Pearson, On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Part I, Biometrika, Vol. 20A (1928), pp. 175-240;

"Sophister," Discussion of Small Samples Drawn from an Infinite Skew Population, Biometrika, Vol. 20A (1928), pp. 389-423;

E. S. Pearson assisted by N. K. Adyanthāya and others, The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations. 2nd paper, Biometrika, Vol. 21 (1929), pp. 259-86.

made of the z -distribution for samples of specific sizes from other types of universe. A theoretical method applicable to samples from a discrete universe was used in a previous paper,[1] in which a rectangular universe was studied in some detail. The rectangular universe was chosen as being the simplest from the standpoint of the method employed, and as a good example of a limited symmetric distribution. It is the purpose of the present paper to apply the method to a triangular population, which is a specimen of a limited skew distribution, and also to a U-shaped universe. The rectangular, triangular and U-shaped universes are shown in Table I in the columns headed R , T , and U , respectively. Their graphs are exhibited in Figure 1.

In addition to the z -distribution, the distributions of means from the triangular and from the U-shaped universe are given.

In the concluding section is discussed the probability corresponding to an interval of three sample standard deviations on each side of the sample mean.

All of the results of the paper are for samples of four.

THE DISTRIBUTION OF z

The distributions of z are shown in Table II,[2], in which the distribution for samples from a normal universe, N , is also given.

The cumulated probability of z for the triangular and for the U-shaped universe are shown in Table III, which may be compared with a similar table for a rectangular and for a normal universe given in Biometrika, Vol. 21 (1929), p. 131.

[1] P. R. Rider, On the Distribution of the Ratio of Mean to Standard Deviation in Small Samples from Non-normal Universes, Biometrika, Vol 21 (1929), pp. 124-143.

[2] For an explanation of the method of deriving these distributions see Rider, loc. cit.

These cumulated probabilities are plotted on probability paper in Figures 2 and 3 and may be compared with similar probabilities for a rectangular universe by reference to Biometrika, Vol. 21 (1929), p. 129, Figure 2.

The principal results to be noted are as follows:

1. The general characteristics of the z -distribution for the U-shaped universe are the same as those for a rectangular universe, viz. a greater number of z 's outside of a certain value of $|z|$, and also a greater clustering of z 's about the origin, than is the case for a normal universe.[1] This is to be expected, since the values of β_2 for U and R are 1.132 and 1.776 respectively, as compared with the value 3 for N .

2. The negative skewness in the triangular universe produces skewness of the opposite type in the distribution of z , as found experimentally by Neyman and E. S. Pearson[2] and by "Sophister."[3] This means (in the case of negative skewness in the universe) that the probability corresponding to an interval from $-\infty$ to z is smaller than when the sampling is from a normal universe.

3. The cumulated probability of $|z|$, or the probability corresponding to an interval from $-z$ to z , is somewhat the same for the triangular universe as for a normal universe;[4] a comparison is made in Table IV.

Results 2 and 3 are apparently due to the fact that in a

---

[1] See Rider, loc. cit., p. 130.

[2] Biometrika, Vol. 20A (1928), p. 198.

[3] Biometrika, Vol. 20A (1928), p. 408.

cf. E. S. Pearson assisted by N. K. Adyanthāya and others. The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations. 2nd paper, Biometrika, Vol. 21 (1929), pp. 259-86.

skew universe the regression of variance on mean[1] is often essentially linear (if parabolic, the vertex of the parabola is well to one side of the scatter diagram). Let us consider the case in which the slope of the regression line is positive. Designating by x the difference between the mean of a sample and the mean of the universe, and by s the standard deviation of the sample, we see that large values of $|x|$ tend to be associated with large values of s^2 (and therefore with large values of s). Thus the values of z tend to be smaller. On the other hand, for large values of $|-x|$, s is smaller and $|-z|$ consequently larger. This means that the frequencies corresponding to the algebraically lower values of z are greater than in the case of a normal universe, or that the use of "Student's" tables would give results too small for the probability that the mean of a sample does not exceed *algebraically* the mean of the universe by more than z times the standard deviation of the sample. The opposite is true in the case studied here, since the universe is negatively skew and the regression line of s^2 on x would have a negative slope.

Since there is a shifting of the whole cumulated z -distribution to the right or left, the effect noted in 3 is readily explained. As a result of this effect we should apparently not be far wrong, when sampling from a skew universe, if we used "Student's" tables to obtain the probability that the mean of a sample does not exceed *numerically* the mean of the universe by more than z times the standard deviation of the sample.[2]

---

[1] For the regression formula see J. Neyman, On the Correlation of the Mean and the Variance in Samples from an "Infinite" Population, Biometrika, Vol. 18 (1926), pp. 401-13.

[2] See E. S. Pearson assisted by N. K. Adyanthāya and others, The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations. 2nd paper, Biometrika, Vol. 21 (1929), pp. 259-86.

THE DISTRIBUTION OF MEANS OF SAMPLES

The distributions of means of samples are shown in Tables V and VI. In these tables x indicates the difference between the mean of the sample and the mean of the universe.

For the difficulties involved in obtaining satisfactory results for the distribution of means of small samples from a U-shaped universe see K. J. Holtzinger and A. E. R. Church, "On the Means of Samples from a U-shaped Population," Biometrika, Vol. 20A (1928), pp. 361-88.

The probability corresponding to an interval of three sample standard deviations on each side of the sample mean.

If M is the mean and σ the standard deviation of a normally distributed variate X, then, as is well known, the probability that an item selected at random will lie within the range $M \pm 3\sigma$ is 0.997. If X and s are the mean and the standard deviation respectively of a sample, the expected or average probability corresponding to the interval $X \pm 3s$ will be different from the probability corresponding to the interval $M \pm 3\sigma$. Shewhart[1] obtained experimentally for the average probability for samples of four associated with the interval $X \pm 3s$ the values 0.90 for a normal universe, 0.91 for a rectangular universe, and 0.91 for a triangular universe.

By analyzing all possible samples of four from the rectangular and triangular universes of Table I it was possible to obtain the probability corresponding to an interval of $3s$ on either side of the sample mean. For example let us consider the sample (1, 1, 2, 2), for which $X = 1.5$, $s = 0.5$. The interval $X \pm 3s$ extends from 0 to 3. This interval includes 0.4 of the rectangular universe R; 0.4 then is the probability that an

[1] W. A. Shewhart, Note on the Probability Associated with the Error of a Single Observation, Journal of Forestry, Vol. 26 (1928) pp. 601-607.

observed value will fall within the interval. Now the particular sample (1, 1, 2, 2) would occur 6 times out of 10,000. If we take all of the samples for which the interval $X \pm 3s$ includes 0.4 of the rectangular universe we find that such samples occur 106 times out of 10,000. Such an analysis leads to Table VII, from which it is ascertained that the average probability corresponding to an interval of $X \pm 3s$ is 0.920. A similar analysis of the triangular universe T gives us Table VIII and yields 0.907 as the average probability associated with $X \pm 3s$. A better understanding of the situation may be obtained from Figure 4.

Paul R. Rider

TABLE I

Rectangular, Triangular and U-Shaped Universes

X	FREQUENCY		
	R	*T*	*U*
0	1		10
1	1	1	5
2	1	2	1
3	1	3	1
4	1	4	1
5	1	5	1
6	1	6	1
7	1	7	1
8	1	8	5
9	1	9	10
10		10	
Total	10	55	36
Mean	4.5	7	4.5
β_1 *	0	0.326	0
β_2 *	$1.\dot{7}\dot{7}\dot{5}$	$2.3\dot{6}$	1.132+

*The values of the β's are uncorrected for grouping. The dots over the digits indicate repeating decimals. The values for a continuous rectangular distribution are $\beta_1 = 0$, $\beta_2 = 1.8$, and for a continuous triangular distribution are $\beta_1 = 0.32$, $\beta_2 = 2.4$.

TABLE II

Probability of z for Samples of 4

z	N	R	T	U
Below −4.25	.0026	.0077	.0015 +	.0384
−4.25 to −3.75	.0011	.0022	.0012	.0004
−3.75 to −3.25	.0018	.0026	.0007	.0009
−3.25 to −2.75	.0032	.0032	.0032	.0077
−2.75 to −2.25	.0062	.0074	.0028	.0016
−2.25 to −1.75	.0131	.0188	.0061	.0106
−1.75 to −1.25	.0314	.0267	.0251	.0147
−1.25 to −0.75	.0829	.0692	.0615	.0256
−0.75 to −0.25	.2047	.2000	.2098	.2299
−0.25 to 0.25	.3058	.3244	.3249	.3405 +
0.25 to 0.75	.2047	.2000	.1741	.2299
0.75 to 1.25	.0829	.0692	.0764	.0256
1.25 to 1.75	.0314	.0267	.0566	.0147
1.75 to 2.25	.0131	.0188	.0118	.0106
2.25 to 2.75	.0062	.0074	.0094	.0016
2.75 to 3.25	.0032	.0032	.0174	.0077
3.25 to 3.75	.0018	.0026	.0000	.0009
3.75 to 4.25	.0011	.0022	.0025 +	.0004
Above 4.25	.0026	.0077	.0150 −	.0383

TABLE III

The cumulated probability of z , or probability that the mean of a random sample of 4 will not exceed (in algebraic sense) the mean of the universe by more than z times the standard deviation of the sample.

z	Cumulated Probability Triangular Universe		Cumulated Probability U-Shaped Universe	
	for - z	for z	for - z	for z
0.0	.51955−	.51955−	.54355+	.54355+
.1	.41649	.54037	.39365−	.60635+
.2	.34497	.61053	.34651	.65349
.3	.28885+	.65136	.30739	.69261
.4	.22719	.70010	.27831	.72193
.5	.18568	.74269	.22081	.77991
.6	.14350−	.76942	.14785+	.85215−
.7	.11580	.79993	.11382	.88618
.8	.09485−	.81086	.09844	.90192
.9	.07784	.83462	.09065+	.90935+
1.0	.06130	.86748	.08285−	.91715+
1.1	.05053	.87456	.07994	.92006
1.2	.04256	.88731	.07471	.92529
1.3	.03716	.88731	.07363	.92637
1.4	.03152	.90787	.07179	.92821
1.5	.02783	.91316	.06614	.93387
1.6	.02334	.91911	.05979	.94021
1.7	.01845−	.93480	.05975−	.94025−
1.8	.01552	.94390	.05941	.94059
1.9	.01410	.94390	.05798	.94202
2.0	.01366	.94810	.05441	.94774
2.1	.01265−	.94810	.04959	.95041
2.2	.01039	.95565−	.04892	.95108
2.3	.00907	.95565−	.04892	.95108
2.4	.00871	.95565−	.04891	.95109
2.5	.00816	.95565−	.04891	.95118
2.6	.00725+	.95565−	.04803	.95197
2.7	.00725+	.95565−	.04732	.95268
2.8	.00661	.96509	.04728	.95272
2.9	.00483	.97910	.04133	.95867
3.0	.00462	.98250−	.03954	.96046
3.5	.00272	.98250−	.03904	.96132
4.0	.00242	.98250−	.03833	.96168

TABLE IV

Cumulated Probability of $|z|$ for Samples of 4.

$\|z\|$ greater than	Probability		$\|z\|$ greater than	Probability	
	Triangular Universe	Normal Universe		Triangular Universe	Normal Universe
0.0	.9219	1.0000	1.6	.1042	.0695−
.1	.8761	.8735+	1.7	.0836	.0603
.2	.7303	.7519	1.8	.0716	.0526
.3	.6375−	.6392	1.9	.0702	.0460
.4	.5271	.5382	2.0	.0652	.0405+
.5	.4423	.4502	2.1	.0646	.0358
.6	.3723	.3751	2.2	.0547	.0318
.7	.3135−	.3121	2.3	.0534	.0283
.8	.2834	.2599	2.4	.0531	.0253
.9	.2432	.2169	2.5	.0525+	.0227
1.0	.1891	.1817	2.6	.0516	.0204
1.1	.1755−	.1528	2.7	.0516	.0185−
1.2	.1552	.1292	2.8	.0415+	.0167
1.3	.1497	.1098	2.9	.0257	.0152
1.4	.1236	.0938	3.0	.0212	.0138
1.5	.1146	.0805+			

TABLE V

Distribution of Means of Samples of 4 from Triangular Universe

x	Probability	x	Probability	x	Probability
−5.25	.00001	−2.25	.01627	0.75	.07202
−5.00	.00004	−2.00	.02200	1.00	.06437
−4.75	.00009	−1.75	.02882	1.25	.05496
−4.50	.00019	−1.50	.03559	1.50	.04462
−4.25	.00038	−1.25	.04501	1.75	.03415+
−4.00	.00070	−1.00	.05362	2.00	.02430
−3.75	.00125	−0.75	.06187	2.25	.01569
−3.50	.00212	−0.50	.06916	2.50	.00881
−3.25	.00344	−0.25	.07484	2.75	.00393
−3.00	.00537	0.00	.07834	3.00	.00109
−2.75	.00805−	0.25	.07918		
−2.50	.01165−	0.50	.07707		

x = (mean of sample) − (mean of universe)

TABLE VI

Distribution of Means of Samples of 4 from U-Shaped Universe

x	Fre-quency	Prob-ability	x	Fre-quency	Prob-ability
-4.50	10000	.0060	0.25	106660	.0635+
-4.25	20000	.0119	0.50	62755	.0374
-4.00	19000	.0113	0.75	51244	.0305+
-3.75	15000	0089	1.00	49270	.0293
-3.50	14225	.0085-	1.25	48376	.0288
-3.25	15300	.0091	1.50	49505	.0295-
-3.00	16690	.0099	1.75	63960	.0381
-2.75	18140	.0108	2.00	89660	.0534
-2.50	35651	.0212	2.25	81224	.0484
-2.25	81224	.0484	2.50	35651	.0212
-2.00	89660	.0534	2.75	18140	.0108
-1.75	63960	.0381	3.00	16690	.0099
-1.50	49505	.0295-	3.25	15300	.0091
-1.25	48376	.0288	3.50	14225	.0085-
-1.00	49270	.0293	3.75	15000	.0089
-0.75	51244	.0305+	4.00	19000	.0113
-0.50	62755	.0374	4.25	20000	.0119
-0.25	106660	.0635+	4.50	10000	.0060
0.00	146296	.0871			
			Total	1679616	1.0001

x = (mean of sample) – (mean of universe)

TABLE VII

Probability Corresponding to the Interval $\bar{X} \pm 3s$
Rectangular Universe

Proportion of universe included in $\bar{X} \pm 3s$*	Number of samples for which this proportion occurs**
0.1	10
0.2	8
0.3	84
0.4	106
0.5	284
0.6	324
0.7	564
0.8	652
0.9	888
1.0	7080
Total	10000

[1] i. e. the probability corresponding to $\bar{X} \pm 3s$.

[2] The probability of the occurrence of this proportion is, of course, obtained by dividing by 10000.

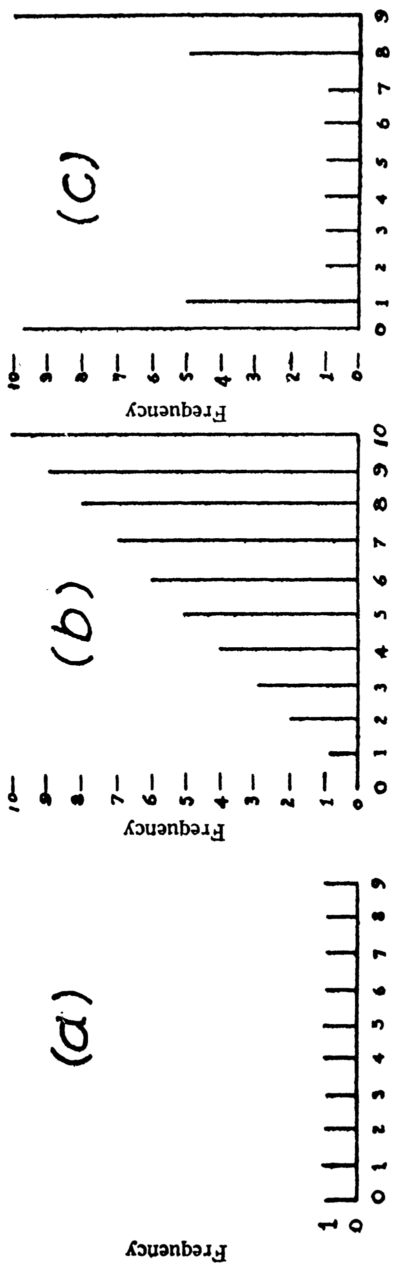
TABLE VIII

Probability Corresponding to the Interval $X \pm 3s$

Triangular Universe

Proportion of universe included in $\bar{X} \pm 3s$	Number of samples for which this proportion occurs	Probability of occurrence of this proportion	Cumulated probability
1/55 = .018	1	—	—
2/55 = .036	16	—	—
3/55 = .055–	89	—	—
4/55 = .073	256	—	—
5/55 = .091	625	.0001	.0001
6/55 = .109	1448	.0002	.0003
7/55 = .127	2401	.0003	.0006
8/55 = .145–	4096	.0004	.0010
9/55 = .164	6993	.0008	.0018
10/55 = .182	11388	.0012	.0030
12/55 = .218	1280	.0001	.0031
13/55 = .236	7776	.0008	.0039
14/55 = .255–	2928	.0003	.0042
15/55 = .273	8762	.0010	.0052
18/55 = .327	12768	.0014	.0066
19/55 = .345+	36000	.0039	.0105
20/55 = .364	8640	.0009	.0114
21/55 = .382	26508	.0029	.0143
22/55 = .400	5400	.0006	.0149
24/55 = .436	32768	.0036	.0185
25/55 = .455–	21600	.0024	.0209
26/55 = .473	10584	.0012	.0221
27/55 = .491	112764	.0123	.0344
28/55 = .509	19698	.0022	.0366
30/55 = .545+	71526	.0078	.0444
33/55 = .600	27116	.0030	.0474
34/55 = .618	296384	.0324	.0798
35/55 = .636	115128	.0126	.0924
36/55 = .655–	37892	.0041	.0965
39/55 = .709	54092	.0059	.1024
40/55 = .727	555924	.0608	.1632
42/55 = .764	57888	.0063	.1695
44/55 = .800	26416	.0029	.1724
45/55 = .818	556520	.0608	.2332
49/55 = .891	774320	.0846	.3178
52/55 = .945+	904676	.0989	.4167
54/55 = .982	879564	.0961	.5128
55/55 = 1.000	4458390	.4872	1.0000
Total	9150625	1.0000	

i. e. the probability corresponding to $\bar{X} \pm 3s$



Rectangular Universe Triangular Universe U-Shaped Universe

FIGURE 1

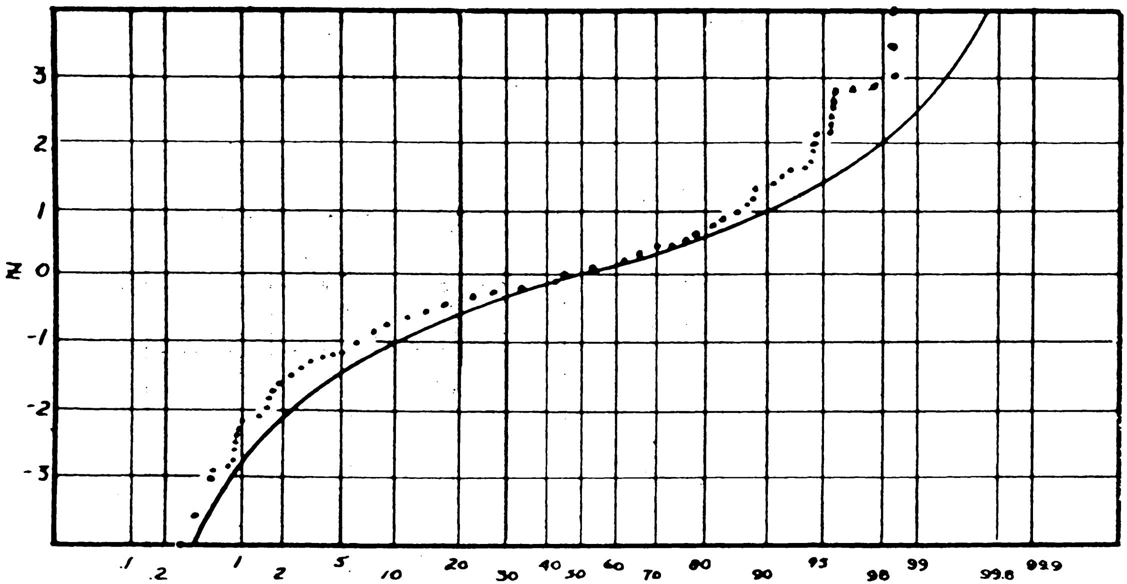


FIGURE 2

Cumulated Probability of z —Triangular Universe
The curve is for samples of 4 from a normal universe.
The dots are for samples of 4 from the universe T .

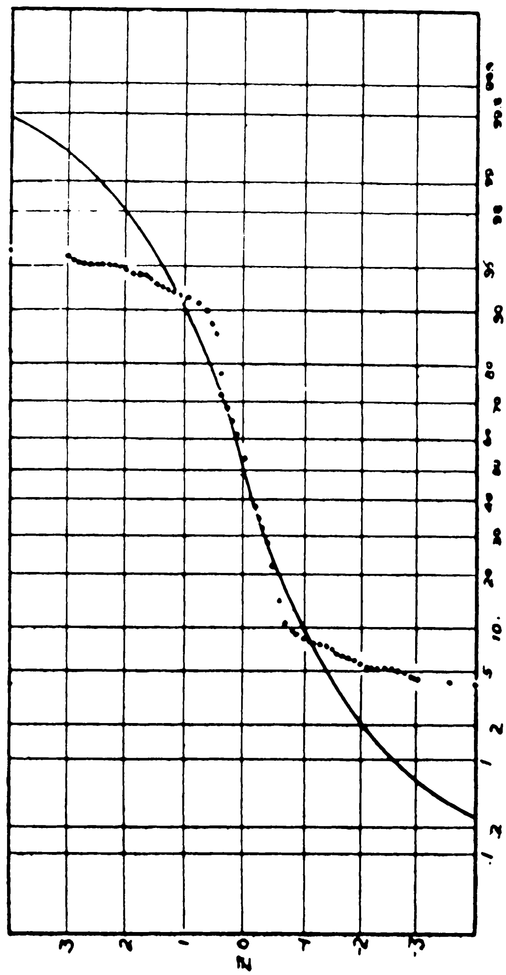


FIGURE 3

Cumulated Probability of z —U-Shaped Universe
The curve is for samples of 4 from a normal universe.
The dots are for samples of 4 from the universe U.

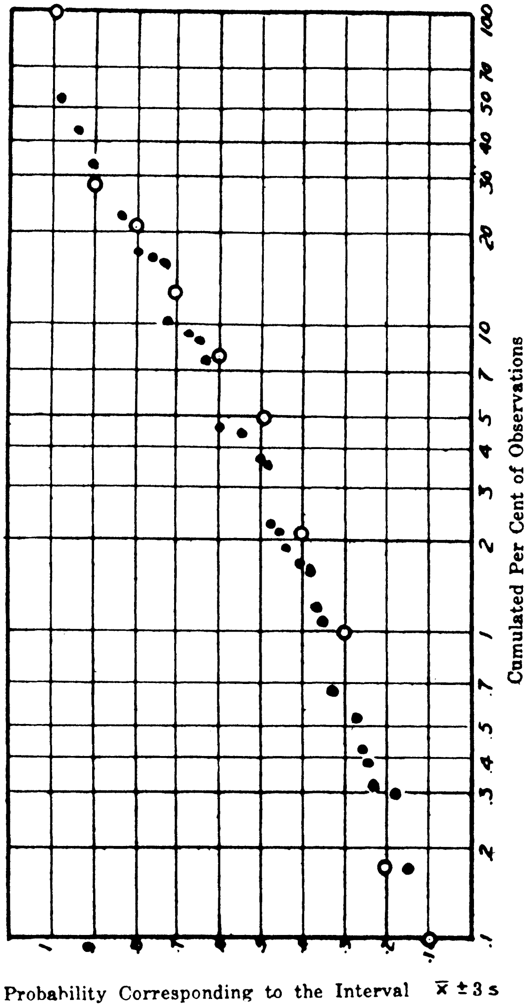


FIGURE 4

Probability Corresponding to the Interval $\bar{X} \pm 3s$
The circles are for samples of 4 from a rectangular universe,
the dots for samples of 4 from a triangular universe.