THE EFFECT OF TRUNCATION ON TESTS OF HYPOTHESES FOR NORMAL POPULATIONS

By Britain J. Williams[1]

The University of Georgia

1. Introduction and summary. When the means of samples drawn from normal populations are employed in performing tests of hypothesis, the experimenter usually assumes that the sample was drawn from a complete normal population. When all members of the population within three standard deviations of the mean admit to sampling, assumption that the sample was drawn from a complete population does not seriously affect the power of the tests. However, there arise situations where the sampling interval is restricted to less than three standard deviations of the mean.

Aggarwal and Guttman [1], [2] have examined the loss of power when using tests based on the assumption that the variable being sampled has a complete normal distribution when, in fact, the distribution is a symmetrically truncated normal distribution. They derived the distribution of means based on samples of size $n \leqq 4$ using convolutions of

$$(1.1) \qquad f(x) = (C/(2\pi)^{\frac{1}{2}}) \exp(-x^2/2), \ |x| < a, \\ = 0, \qquad \text{otherwise,}$$

where C is given by

$$(1.2) \qquad C^{-1} = (2\pi)^{-\frac{1}{2}} \int_{-a}^{a} e^{-\frac{1}{2}t^2}\, dt.$$

Birnbaum and Andrews [2] have pointed out that sums of symmetrically truncated normal random variables have a limiting normal distribution. Thus for large n one can obtain for an approximate cumulative distribution of $\bar{X} = (1/n) \sum_{i=1}^{n} X_i$. For arbitrary n, however, no general formula giving the distribution of means (or sums) of samples of size n drawn from a truncated normal population is available.

In this paper we extend the work of Aggarwal and Guttman [1] to non-symmetric truncation and arbitrary sample size. An asymptotic series for the distribution of sums of samples of size n drawn from a truncated normal population is presented.

2. An asymptotic expansion of the distribution of sums of a truncated variate. We will denote the density function of a truncated normal variate by

$$(2.1) \qquad f(x; \mu, \sigma, a, b) = (C/\sigma(2\pi)^{\frac{1}{2}}) \exp\{-\tfrac{1}{2}(x-\mu)^2/\sigma^2\}, \\ \mu + \sigma a < x < \mu + \sigma b,$$

where C is given by

Received 18 November 1964; revised 20 March 1965.

[1] Currently with RCA Service Company, Missile Test Project, Patrick AFB, Florida.

(2.2) $$C^{-1} = (2\pi)^{-\frac{1}{2}} \int_a^b e^{-\frac{1}{2}t^2}\, dt.$$

Let $X_1, X_2, \cdots, X_n$ be n independent random variables with the density function of X_i; $i = 1, 2, \cdots, n$, given by (2.1) and consider the variable $X = X_1 + X_2 + \cdots + X_n$. If we denote the mean and variance of X_i by μ_1 and σ_1^2 respectively, then the mean and variance of X is given by $\mu = n\mu_1$ and $\sigma^2 = n\sigma_1^2$. For $v = 1, 2, 3, \cdots$, let χ_v denote the cumulants of $X_i - \mu_1$. Then the characteristic function of $X_i - \mu_1$ is given by

$$\psi_1(t) = \exp\{\textstyle\sum_{v=1}^{\infty} (\chi_v/v!)(it)^v\},$$

and the characteristic function of $Y = (X - \mu)/\sigma$ may be expressed as

(2.3) $$\psi(Y) = [\psi_1(t/\sigma_{1n})]^n = \exp\{n \textstyle\sum_{v=1}^{\infty} (\lambda_v/v!)(it/n)^v\}$$

where $\lambda_v = \chi_v/\sigma_1^v$. Expanding the right hand side of (2.3) in powers of $n^{-\frac{1}{2}}$ and collecting terms yields

(2.4) $$\psi(t) = e^{-t^2/2} + \textstyle\sum_{v=1}^{\infty} \{[b_{v,v+2}(it)^{v+2} + \cdots + b_{v,3v}(it)^{3v}]/n^{v/2}\}e^{-t^2/2}$$

where $b_{v,v+2h}$ is a polynomial in $\lambda_3, \lambda_4, \cdots, \lambda_{v-h+3}$ which is independent of n.

Denoting the density function of the standard normal variable by φ, the integral relation

$$\int_{-\infty}^{\infty} e^{itx}\varphi^{[n]}(x)dx = (-it)^n e^{-t^2/2}$$

applied to (2.4) yields the following form of the density function of Y,

(2.5) $$f_n(Y) = \varphi(Y) + \textstyle\sum_{v=1}^{\infty} (-1)^v \cdot [b_{v,v+2}\varphi^{[v+2]}(Y) + \cdots + b_{v,3v}\varphi^{[3v]}(Y)]/n^{v/2}.$$

This series is due to Edgeworth [6]. A corresponding expansion of the distribution function, F_n, of Y can be obtained by replacing φ in (2.5) by the distribution function, Φ, of the standard normal variable. Cramér [3] has shown that this series is asymptotic in powers of $n^{-\frac{1}{2}}$ and that the remainder term is of the same order as the first term neglected.

For computing purposes, the distribution function of Y was expressed in the form

$$F_n(Y) = \textstyle\sum_{h=0}^{N} (1/r_1!r_2!\cdots r_N!)(\lambda_3/3!)^{r_1}\cdots(\lambda_{N+2}/(N+2)!)^{r_N} \cdot \Phi^{[p+2h]}(Y)/n^{p/2} + R_n(N, Y),$$

where the second summand extends over all non-negative integral values of $r_1, r_2, \cdots, r_N, p$, and h satisfying

$$r_1 + r_2 + \cdots + r_N = h$$

and

$$r_1 + 2r_2 + \cdots + Nr_N = p.$$

For $N = 13$, it was found that $R_n(N, Y)$ is of the order 0.0001 when $n = 4$. The accuracy increases with n.

3. Tests of hypotheses. Consider a sample of size n drawn from a normal population with mean μ and variance σ^2. We may assume without loss of generality that $\sigma^2 = 1$.

A uniformly most powerful test of the null hypothesis $H_0:\mu = 0$ against the alternate hypothesis $H_a:\mu > 0$ is given by

$$\text{reject } H_0 \text{ if } \bar{X} > Z_\alpha/n^{\frac{1}{2}}; \text{ accept } H_0 \text{ otherwise,} \tag{3.1}$$

where Z_α is the point exceeded with probability α using the distribution of the standard normal variable, $Z \sim N(0, 1)$. A usual test of size α for a complete population becomes a test of size α' when the population is truncated, where

$$\alpha' = \Pr(Z_t > Z_\alpha/n^{\frac{1}{2}}), \tag{3.2}$$

and Z_t is a truncated normal variable with density function $f(z_t\,; 0, 1, a, b)$.

If $X \sim N(\mu, 1)$, then $\bar{X} \sim N(\mu, 1/n)$ and the usual power function, the probability of accepting H_0 when in fact H_a is true, is given by

$$\begin{aligned} P_u(\mu) &= \Pr(X > Z_\alpha/n^{\frac{1}{2}} \mid X \sim N(\mu, 1/n)) \\ &= \Pr(Z > Z_\alpha - \mu n^{\frac{1}{2}} \mid Z \sim N(0, 1)). \end{aligned} \tag{3.3}$$

If X is a truncated normal variable with density function $f(X;\, \mu, 1, a, b)$ then the actual power of the test is given by

$$\begin{aligned} P_a(\mu) &= \Pr(X > Z_\alpha/n^{\frac{1}{2}} \mid X \sim f_n(X;\, \mu, 1, a, b)) \\ &= \Pr(Z_t > (Z_\alpha - \mu_1 n^{\frac{1}{2}})/\sigma_1 \mid Z_t \sim f_n(Z;\, \mu, 1, a - \mu_1, b - \mu_1)), \end{aligned} \tag{3.4}$$

where μ_1 and σ_1^2 are the mean and variance of $f(X;\, \mu, 1, a, b)$ and $f_n(X;\, \mu, 1, a, b)$ denotes the density function of means based on samples of size n drawn from $f(X;\, \mu, 1, a, b)$.

Let $L(\mu)$ denote the loss of power when using the usual test of the hypothesis (3.1) when in fact the population is truncated. Then $L(\mu) = P_u(\mu) - P_a(\mu)$. The tables express $L(\mu)$ as a percentage of $P_u(\mu)$ for the one sided test (3.1) with $\alpha = 0.05$ for various sample sizes and truncation points. The units used are standard deviations of the parent population.

4. Conclusions. An examination of the tables reveals that there is very little loss of power when the sample is of size 10 or larger and the true value of the mean is more than 0.5 standard deviations away from the value specified in the null hypothesis. Also, as Aggarwal and Guttman [1] have pointed out, there is a change in sign from positive to negative in the loss of power as soon as μ exceeds $Z_\alpha/n^{\frac{1}{2}}$.

When the value of the true mean of the population is equal to the value specified in the null hypothesis, the size of the test and the value of the power function at that point assume the same value. Thus, from column one of the tables, the actual size, α', of the test is easily computed. From Table I we obtain that for $n = 5$ the actual size of the usual test of size $\alpha = 0.05$ is $\alpha' = 0.001$ when the population is symmetrically truncated within one standard deviation of the mean. The actual size increases to 0.014 and 0.031 when the population is symmetrically truncated at 1.5 and 2.0 standard deviations from the mean.

There is no appreciable loss in power or decrease in the size of the test if the population is symmetrically truncated more than 2.5 standard deviations from the mean.

Let $X \sim N(\mu, 1)$ and consider the variable $X_t \sim f(X; \mu, 1, -\infty, b)$ resulting from single truncation on the right at the point $\mu + b$. If we denote the mean and variance of X_t by μ_1 and σ_1^2 respectively, then, for samples of size n, we have the mean and variance of $X \sim f_n(X; \mu, 1, -\infty, b)$ given by μ_1 and σ_1^2/n respectively. Since $\mu_1 < \mu$ we may put $\mu - \mu_1 = \epsilon > 0$ and it follows from the weak law of large numbers that for any constant δ satisfying $0 < \delta < 1$ we have

$$\Pr(|\bar{X}_t - \mu_1| < \epsilon) > 1 - \delta,$$

where $\bar{X}_t$ is based on a sample of size $n > \sigma_1^2/\epsilon^2\delta$.

Thus, if the usual test (3.1) is employed when in fact the distribution is singly truncated on the right, the actual size of the test will tend to zero as n increases. Similarly, if the population is singly truncated on the left and the usual test (3.1) is employed, the actual size of the test tends to one as n increases.

An examination of Table II reveals that when the population is singly truncated at 1.5 times the standard deviation to the right of the mean, the actual size of the usual test of size $\alpha = 0.05$ is $\alpha' = 0.032$ when $n = 5$ and decreases rapidly as n increases. From Table III we see that if the population is singly truncated on the left at 1.5 standard deviations from the mean, the actual size of the test is $\alpha' = 0.075$ for $n = 5$ and increases with n.

Thus, for singly truncated populations, there is not appreciable loss of power involved in using the usual one-sided test when the truncation is beyond 1.5 standard deviations from the mean and the mean is more than 0.5 standard deviations from the value specified in the null hypothesis. However, there is a considerable change in the size of the test for large samples. Hence, when using the usual test on means based on large samples from populations which are singly truncated within 2.5 standard deviations of the population mean, an improvement in the size of the test can be realized by further truncating to obtain a symmetrically truncated population.

5. Acknowledgments. The author acknowledges with pleasure the help and guidance of Dr. A. C. Cohen, Jr. under whose direction this work was carried out. He would also like to express appreciation to Dr. P. V. Rao for his many valuable suggestions.

REFERENCES

[1] Aggarwal, O. P. and Guttman, I. (1960). Truncation and tests of hypothesis. *Ann. Math. Statist.* **31** 230–238. Corrections to "Trucation and tests of hypothesis." *Ann. Math. Statist.* **31** 1231.

[2] Birnbaum, Z. W. and Andrews, F. C. (1949). On sums of symmetrically truncated normal random variables. *Ann. Math. Statist.* **20** 458–461.

[3] Cramér, H. (1937). *Random Variables and Probability Distributions*. Cambridge Univ. Press.

[4] Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[5] Edgeworth, F. Y. (1905). The law of error. *Proc. Cambridge Phil. Soc.* **21** 36.

[6] Kendall, M. G. (1945). *The Advanced Theory of Statistics*. Griffin, London.

TABLE I

Loss in power expressed as percentage of $P_u(\mu)$ for $\alpha = 0.05$ and symmetric truncation at the terminus point a

		μ			
n	a	0.0	0.5	1.0	1.5
5	1.0	99.51	43.90	−18.97	−4.59
	1.5	71.64	19.85	−8.68	−3.37
	2.0	37.03	8.39	−3.69	−1.75
	2.5	13.45	2.96	−1.35	−0.60
6	1.0	99.01	34.79	−17.90	−2.18
	1.5	72.68	15.16	−8.82	−1.88
	2.0	37.15	6.29	−3.80	−1.04
	2.5	13.66	2.35	−1.48	−0.35
7	1.0	98.76	25.99	−15.01	−1.02
	1.5	73.10	10.99	−8.13	−0.97
	2.0	37.39	4.49	−3.60	−0.61
	2.5	13.87	1.69	−1.38	−0.23
8	1.0	98.61	17.94	−11.84	−0.47
	1.5	73.32	7.43	−7.04	−0.47
	2.0	37.61	3.00	−3.22	−0.34
	2.5	14.06	1.12	−1.21	−0.15
9	1.0	98.49	10.76	−9.00	−0.22
	1.5	73.43	4.40	−5.85	−0.22
	2.0	37.78	1.76	−2.79	−0.18
	2.5	14.22	0.65	−1.04	−0.09
10	1.0	98.41	4.49	−6.69	−0.10
	1.5	73.50	1.82	−4.71	−0.10
	2.0	37.92	0.73	−2.34	−0.09
	2.5	14.35	0.27	−0.88	−0.05
15	1.0	98.16	−14.62	−1.31	0.00
	1.5	73.58	−6.09	−1.10	0.00
	2.0	38.29	−2.43	−0.75	0.00
	2.5	14.72	−0.86	−0.32	0.00
20	1.0	98.04	−19.34	−0.24	0.00
	1.5	73.56	−8.78	−0.23	0.00
	2.0	38.42	−3.60	−0.18	0.00
	2.5	14.88	−1.28	−0.09	0.00
25	1.0	97.97	−17.36	−0.04	0.00
	1.5	73.52	−8.83	−0.04	0.00
	2.0	38.47	−3.78	−0.03	0.00
	2.5	14.96	−1.36	−0.02	0.00
50	1.0	97.83	−3.00	0.00	0.00
	1.5	73.41	−2.47	0.00	0.00
	2.0	38.53	−1.40	0.00	0.00
	2.5	15.08	−0.57	0.00	0.00
100	1.0	97.77	−0.04	0.00	0.00
	1.5	73.33	−0.04	0.00	0.00
	2.0	38.53	−0.03	0.00	0.00
	2.5	15.11	−0.02	0.00	0.00

TABLE II

Loss in power expressed as percentage of $P_u(\mu)$ for $\alpha = 0.05$ and single truncation on the right at the terminus point a

n	a	μ = 0.0	0.5	1.0	1.5
5	1.5	35.98	46.29	15.07	1.92
	2.0	28.54	17.99	5.15	0.43
	2.5	9.71	5.62	1.62	0.15
6	1.5	70.43	40.80	9.50	1.01
	2.0	40.90	16.60	3.27	0.61
	2.5	15.05	5.29	0.75	0.30
7	1.5	79.88	40.41	8.08	0.80
	2.0	45.18	16.16	2.57	0.32
	2.5	16.95	5.12	0.56	0.15
8	1.5	83.43	40.19	7.11	0.44
	2.0	47.37	15.74	2.17	0.14
	2.5	17.97	4.91	0.51	0.06
9	1.5	85.26	39.75	6.12	0.17
	2.0	48.87	15.29	1.85	0.05
	2.5	18.69	4.70	0.47	0.02
10	1.5	86.47	39.16	5.16	0.04
	2.0	50.07	14.83	1.55	0.01
	2.5	19.28	4.49	0.42	0.00
15	1.5	90.13	35.56	1.82	0.00
	2.0	54.61	12.65	0.50	0.00
	2.5	21.54	3.67	0.14	0.00
20	1.5	92.44	31.80	0.49	0.00
	2.0	58.09	10.69	0.12	0.00
	2.5	23.31	3.02	0.03	0.00
25	1.5	94.08	28.08	0.11	0.00
	2.0	60.98	8.89	0.02	0.00
	2.5	24.83	2.45	0.00	0.00
50	1.5	97.89	12.52	0.00	0.00
	2.0	70.87	2.80	0.00	0.00
	2.5	30.56	0.67	0.00	0.00
100	1.5	99.60	1.30	0.00	0.00
	2.0	81.27	0.12	0.00	0.00
	2.5	38.13	0.02	0.00	0.00

TABLE III

Loss in power expressed as percentage of $P_u(\mu)$ for $\alpha = 0.05$ and single truncation on the left at the terminus point $-a$. All non-zero entries in the body of the table are negative

		μ			
n	a	0.0	0.5	1.0	1.5
5	1.5	49.31	41.45	21.17	2.36
	2.0	5.81	13.02	7.86	1.36
	2.5	0.65	4.17	2.57	0.41
6	1.5	54.70	37.87	17.09	2.54
	2.0	13.80	11.81	7.06	1.22
	2.5	3.86	3.21	2.52	0.41
7	1.5	59.72	37.50	12.57	1.06
	2.0	17.26	12.46	5.89	0.71
	2.5	4.99	3.40	2.13	0.28
8	1.5	65.61	38.11	9.64	0.45
	2.0	19.64	13.23	4.88	0.36
	2.5	5.64	3.72	1.78	0.17
9	1.5	72.03	38.65	7.49	0.21
	2.0	21.69	13.84	4.00	0.18
	2.5	6.15	4.00	1.48	0.10
10	1.5	78.71	38.80	5.77	0.10
	2.0	23.60	14.25	3.23	0.09
	2.5	6.61	4.21	1.22	0.05
15	1.5	112.99	34.53	1.27	0.00
	2.0	32.53	14.22	0.91	0.00
	2.5	8.75	4.51	0.40	0.00
20	1.5	147.21	27.02	0.24	0.00
	2.0	40.87	12.47	0.20	0.00
	2.5	10.77	4.17	0.10	0.00
25	1.5	181.16	19.85	0.04	0.00
	2.0	48.78	10.22	0.04	0.00
	2.5	12.67	3.60	0.02	0.00
50	1.5	348.27	2.99	0.00	0.00
	2.0	84.36	2.29	0.00	0.00
	2.5	20.82	1.04	0.00	0.00
100	1.5	665.55	0.04	0.00	0.00
	2.0	146.58	0.04	0.00	0.00
	2.5	33.63	0.03	0.00	0.00