U. S. Department of Commerce
National Bureau of Standards

Research Paper RP2007
Volume 43, July 1949

Part of the Journal of Research of the National Bureau of Standards

Formulas for the Percentage Points of the Distribution of the Arithmetic Mean in Random Samples from Certain Symmetrical Universes [1]

By Uttam Chand

Using the method of Fisher and Cornish, the $100\epsilon\%$ point of the distribution of the arithmetic mean in random samples of size N from any universe having finite cumulants, $K_1, K_2, K_3, \ldots$, is expressed to order $1/N^2$ as a function of (*i*), the first five positive integral powers of the $100\epsilon\%$ point of a standardized normal variable, and (*ii*) the quantities K_1, $(K_2)^{1/2}$, and $K_r/(K_2)^{r/2}$ for $r=3$, 4, 5, and 6. The numerical coefficients involved are evaluated for the case of sampling from the normal, rectangular, double-exponential, sech, and sech² distributions, and the accuracy of the resulting formulas illustrated by numerical examples.

Consider $y_1 \ldots y_N$ a random sample of size N from a population having finite cumulants $K_r = K_r(y)$, $(r=1,2, \ldots)$. Let $a_r = a_r(y) = K_r/K_2^{r/2}$ be the relative cumulants, $(r=1,2, \ldots)$, and let λ and σ specify the location and scale parameters of the population distribution, respectively.

Let $\bar{y}$ denote the sample mean and consider the standardized variate

$$t = \frac{\bar{y} - K_1(\bar{y})}{\sqrt{K_2(\bar{y})}}.$$

Applying the formulas of Fisher and Cornish [1, 2] [2] and utilizing the relations

$$K_1(\bar{y}) = K_1(y)$$

$$K_2(\bar{y}) = K_2(y)/N$$

$$a_1(t) = 0$$

$$a_2(t) = 1$$

and

$$a_r(t) = a_r(\bar{y}) = a_r(y)/N^{\frac{r-2}{2}}, (r=3, 4, \ldots),$$

the one-tail ϵ-probability point, t_ϵ, of the standardized variate t can be expressed to order of $1/N^2$ as

$$\begin{aligned} t_\epsilon = x &+ \frac{a_3}{6N^{1/2}}(x^2-1) - \frac{a_3^2}{36N}(2x^3-5x) + \frac{a_4}{24N}(x^3-3x) \\ &+ \frac{a_5}{120N^{3/2}}(x^4-6x^2+3) - \frac{a_3a_4}{24N^{3/2}}(x^4-5x^2+2) \\ &+ \frac{a_3^3}{324N^{3/2}}(12x^4-53x^2+17) \\ &+ \frac{a_6}{720N^2}(x^5-10x^3+15x) \\ &+ \frac{a_3^2a_4}{288N^2}(14x^5-103x^3+107x) \\ &- \frac{a_4^2}{384N^2}(3x^5-24x^3+29x) \\ &- \frac{a_3a_5}{180N^2}(2x^5-17x^3+21x) \\ &- \frac{a_3^4}{7776N^2}(252x^5-1688x^3+1511x) \end{aligned}$$

where $x = x_\epsilon$ is a normal deviate exceeded with probability ϵ.[3]

In particular for a *symmetrical universe* we have

$$a_r(y) = 0 \text{ for } r=3, 5, 7, \ldots,$$

whence

$$t_\epsilon = x + \frac{a_4}{24N}(x^3-3x) + \frac{1}{N^2}\left[\frac{a_6}{720}(x^5-10x^3+15x) - \frac{a_4^2}{384}(3x^5-24x^3+29x)\right].$$

[1] Revision of a paper written during the summer of 1947 when the author was a guest worker at the National Bureau of Standards, and which was read by title at the Eleventh Summer Meeting of the Institute of Mathematical Statistics, held in Madison, Wis., Sept. 7 to 10, 1948.

[2] Figures in brackets indicate the literature references at the end of this paper.

[3] Miriam L. Yevick has drawn to my attention the fact that comparison of the above expression for t_ϵ with the corresponding expression given by Simaika [3], which includes terms involving $\gamma_1(=a_3)$ and $\gamma_2(=a_4)$ only, reveals two errors in his numerical coefficients: the coefficient of γ_2^2 given as "−29/381" should read "−29/384", and the coefficient of $\gamma_1^2\gamma_2$ given as "−107/288" should read "+107/288".

It is interesting to note that for the one-tail 0.04163-probability point, for which $x_\epsilon^2=3$, we have

$$t_\epsilon = x_\epsilon + 0(N^{-2})$$

as $N \rightarrow \infty$.

The density functions and the requisite cumulants of five symmetrical distributions considered explicitly are listed in table 1, and the values of t_ϵ computed for $N=10$ from 1-term (normal approximation), 2-terms, and 3-terms of the immediately preceding formula are compared with the corresponding true values in table 2.

TABLE 1. *Five symmetrical distributions and some of their cumulants*

Distribution	Density function	K_2 (variance)	$a_4(=\gamma_2)$	a_6
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y-\lambda)^2}{2\sigma^2}}$	σ^2	0	0
Rectangular	$\frac{1}{\sigma}$ for $-\frac{1}{2} \leq \frac{y-\lambda}{\sigma} \leq \frac{1}{2}$	$\frac{\sigma^2}{12}$	$-\frac{6}{5}$	$\frac{48}{7}$
Double exponential	$\frac{1}{2} e^{\frac{-\lvert y-\lambda \rvert}{\sigma}}$	$2\sigma^2$	3	30
Sech	$\frac{1}{\pi\sigma} \operatorname{sech}\left(\frac{y-\lambda}{\sigma}\right)$	$\frac{\pi^2\sigma^2}{4}$	2	16
Sech 2	$\frac{1}{2\sigma} \operatorname{sech}^2\left(\frac{y-\lambda}{\sigma}\right)$	$\frac{\pi^2\sigma^2}{12}$	$\frac{6}{5}$	$\frac{16}{7}$

TABLE 2. *Comparison of various approximate and exact values of probability points, t_ϵ, of the standardized variate t $N=10$*

Distribution [a]	One-tail probability levels						
	0.001	0.005	0.010	0.025	0.050	0.100	0.250
Normal:							
1 term (exact)	3. 0902	2. 5758	2. 3263	1. 9600	1. 6449	1. 2816	0. 6745−
Rectangular:							
2 terms	2. 9890	2. 5290	2. 2983	1. 9517	1. 6473	1. 2903	. 6831
3 terms	2. 9837	2. 5270	2. 2973	1. 9515−	1. 6474	1. 2905+	. 6833
exact	2. 9825	2. 5269	2. 2971	1. 9515	1. 6474	1. 2905	. 6833
Double exponential:							
2 terms	3. 3432	2. 6929	2. 3965−	1. 9806	1. 6388	1. 2598	. 6530
3 terms	3. 3038	2. 6839	2. 3941	1. 9822	1. 6409	1. 2612	. 6531
exact	3. 3114	2. 6839	2. 3932	1. 9814	1. 6406	1. 2612	. 6531
Sech:							
2 terms	3. 2589	2. 6539	2. 3731	1. 9737	1. 6408	1. 2671	. 6602
3 terms	3. 2426	2. 6492	2. 3712	1. 9738	1. 6415−	1. 2677	. 6605−
exact	3. 2449	2. 6494	2. 3711	1. 9736	1. 6414	1. 2677	. 6605−
Sech²:							
2 terms	3. 1914	2. 6226	2. 3544	1. 9682	1. 6424	1. 2729	. 6659
3 terms	3. 1840	2. 6219	2. 3548	1. 9691	1. 6430	1. 2730	. 6657
exact	3. 1646	2. 6207	2. 3540	1. 9681	1. 6426	1. 2732	. 6661

[a] The exact values of t_ϵ, for the normal distribution were taken from Kelley [4]; for the rectangular and double-exponential distributions, from an unpublished table prepared in connection with [5]. The exact values of t_ϵ for the sech and sech² distributions were computed under the supervision of Irene A. Stegun of the National Bureau of Standards Computation Laboratory from formulas prepared by Julius Lieblein of the National Bureau of Standards Statistical Engineering Laboratory.

The author expresses appreciation to Churchill Eisenhart for suggesting this problem and to Lola S. Deming and Celia S. Martin of the Statistical Engineering Laboratory, National Bureau of Standards, for the numerical computations.

References

[1] E. A. Cornish and R. A. Fisher, Moments and cumulants in the specification of distributions, Rev. Inst. Intl. Stat. **4,** 1 to 14 (1937).

[2] Maurice G. Kendall, The advanced theory of statistics, **I,** p. 156 to 158 (Charles Griffin & Co., Ltd., London; 1st ed., 1943; 3d ed., 1947).

[3] J. B. Simaika, Interpolation for fresh probability levels between the standard table levels of a function, Biometrika **XXXII,** parts III and IV, p. 263 to 276 (April 1942).

[4] T. L. Kelley, The Kelley statistical tables, revised 1948 (Harvard University Press, Cambridge, Mass.).

[5] Churchill Eisenhart, Lola S. Deming, Celia S. Martin, On the arithmetic mean and median in small samples from the normal and certain nonnormal populations, presented at the Summer Meeting of the Institute of Mathematical Statistics, held in Madison, Wis., Sept. 7 to 10, 1948.

WASHINGTON, March 22, 1949.