NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

THE STANDARD ERRORS OF THE GEOMETRIC AND HARMONIC MEANS AND THEIR APPLICATION TO INDEX NUMBERS[1]

By Nilan Norris

Attempts to derive useful expressions for estimating the standard deviations of the sampling errors of the geometric and harmonic means have not yielded results comparable with those afforded by the modern theory of estimation, including fiducial inference. There are in the literature of probability theory certain theorems which can be applied to obtain these desired results in a straightforward manner. The use of forms for estimating standard errors is subject to certain conditions which are not always fulfilled, particularly in the case of time series. An understanding of these limitations should deter those who may be tempted to judge the significance of phenomena such as price changes solely on the basis of estimated standard errors of indexes.

1. Statement of formulas. The standard error of the geometric mean of a sequence of positive independent chance variables denoted by $x_i = x_1, x_2, \cdots, x_n$, is $\sigma_G = \theta_1 \frac{\sigma_{\log x}}{\sqrt{n}}$, where θ_1 is the population geometric mean of the variates; so that $\sigma_{\log x}$ is the standard deviation of the logarithms in the population as given by $\sigma_{\log x} = [E\{[\log x - E(\log x)]^2\}]^{\frac{1}{2}}$; and n is the number of individuals comprising the sample. The estimate of the standard error of the geometric mean is $s_G = G \frac{s_{\log x_i}}{\sqrt{n-1}}$, where G is the sample geometric mean, that is, the estimate of θ_1; so that $s_{\log x_i}$ is the estimate of $\sigma_{\log x}$; and $n - 1$ is the degree of freedom of the sample.

---

[1] This article summarizes two papers presented at sessions of the Institute of Mathematical Statistics at Detroit, Michigan on December 27, 1938, and at Philadelphia, Pennsylvania on December 27, 1939. The results given herein can be derived by several methods, which vary somewhat as to degree of rigor. The writer wishes to acknowledge his indebtedness to the referee for suggesting a proof based on a probability theorem stated by J. L. Doob, "The limiting distributions of certain statistics," *Annals of Math. Stat.*, Vol. 4 (1935), pp. 160–169. The standard deviation formulas obtained follow as an application of this theorem, as will be seen by reference to it. Obviously the asymptotic variance formulas of many other statistics (estimates of parameters) can be obtained in a similar manner.

The standard error of the harmonic mean of a sequence of positive independent chance variables denoted by $x_i = x_1, x_2, \cdots, x_n$, is $\sigma_H = \theta_2^2 \frac{\sigma_{1/x}}{\sqrt{n}}$, where the population harmonic mean of the variates is $\theta_2 = 1/\alpha = [E(1/x)]^{-1}$; so that the standard deviation of $1/x$ in the population is $\sigma_{1/x} = [E\{[1/x - E(1/x)]^2\}]^{\frac{1}{2}}$; and n is the number of observations comprising the sample. The estimate of the standard error of the harmonic mean is $s_H = \frac{1}{a^2} \frac{s_{1/x_i}}{\sqrt{n-1}}$, where the estimate of α is given by $a = \frac{1}{H} = \frac{1}{n} (\Sigma\, 1/x_i)$; in which s_{1/x_i} is the standard deviation of the reciprocals of the observations comprising the sample; and $n - 1$ is the degree of freedom of the sample.

2. Derivation of formulas. These forms can be obtained by application of the Laplace-Liapounoff theorem[2] as follows: Let $x_i = x_1, x_2, \cdots, x_n$ be a set of positive independent chance variables with the same distribution functions, where the expectations, $E(x_i)$ and $E(x_i^2)$ exist, and where $\sigma_x^2 = E\{[x_i - E(x_i)]^2\} > 0$. The last condition is imposed to eliminate the trivial case in which the x_i are all equal and their distribution is confined to a single point. The geometric mean of the x_i is $G = (x_1 \cdot x_2 \cdot \cdots \cdot x_n)^{1/n}$, and the harmonic mean of the x_i is $H = \left[\frac{1}{n} \sum \frac{1}{x_i}\right]^{-1}$.

It is necessary to assume that both $\sigma_{\log x}^2$ and $\sigma_{1/x}^2$ are finite, and that in the case of both $\log x$ and $1/x$ at least one moment of order higher than any two of the respective variates is also finite. The requirement that the variance and at least one moment higher than the variance be finite can be weakened in various ways, but this is a trivial consideration, since nearly all distributions of any importance have finite third moments.[3] Certain rarely occurring types of distributions, such as the Cauchy distribution, have infinite variance. In such cases, standard error formulas as ordinarily used are not valid.

Let $E(\log x) = \zeta$, and $E(1/x) = \alpha$. By the Laplace-Liapounoff theorem, except for terms of order $1/\sqrt{n}$, the limiting distributions of $\frac{\sqrt{n}(\log G - \zeta)}{\sigma_{\log x}}$ and $\frac{\sqrt{n}(H^{-1} - \alpha)}{\sigma_{1/x}}$ are normal with zero arithmetic means and unit variances. That is, if C represents a set of conditions on chance variables, and $P\{C\}$ is the probability that these conditions are satisfied, then

[2] A. Khintchine, Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, J. Springer, Berlin, 1933, Vol. II, No. 4, pp. 1–8; J. L. Doob, *op. cit.*, pp. 160–169; and S. S. Wilks, *Statistical Inference*, 1936–1937, Edwards Brothers, Inc., Ann Arbor, 1937, pp. 39 *f*.

[3] For a more detailed discussion of this matter see Wilks, *op. cit.*, pp. 39 *f*.

$$\lim_{n\to\infty} P\left\{\frac{\sqrt{n}(\log G - \zeta)}{\sigma_{\log x}} < t\right\} = \lim_{n\to\infty} P\left\{\frac{\sqrt{n}(H^{-1} - \alpha)}{\sigma_{1/x}} < t\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}}\,dx.$$

In order to use these relations in obtaining the limiting distributions of the geometric and harmonic means, it is necessary to suppose that the sequence of random chance variables, V_i, converges in probability (converges stochastically) to ρ, and that the sequence of random chance variables, $\sqrt{n}(V_i - \rho)$, has a normal limiting distribution with zero arithmetic mean and variance σ^2. Also, it is necessary to assume that the real-valued function, $f(x)$, has a Taylor expansion valid in the neighborhood of ρ. If $f'(\rho) \neq 0$, only the first two terms of the series are needed. The required expansion is given by

$$f(x) = f(\rho) + (x - \rho)f'(\rho) + \frac{(x - \rho)^2}{2} f''[\rho + \beta(x - \rho)],$$

where $0 < \beta < 1$, and $f''(x)$ is continuous in the neighborhood of ρ. When these conditions are fulfilled, the limiting distribution of $\sqrt{n}[f(V_i) - f(\rho)]$ is normal with an arithmetic mean of zero and a variance of $\sigma^2[f'(\rho)]^2$.

Let $f(\log G) = e^{\log G}$, and use the expansion given by $e^{\log G} = e^{\zeta} + (\log G - \zeta)e^{\zeta} + \frac{1}{2}(\log G - \zeta)^2 e^{\zeta + \beta(\log G - \zeta)}$. Since $\theta_1 = e^{\zeta}$, it follows that the limiting distribution of $\sqrt{n}(G - \theta_1)$ is normal with an arithmetic mean of zero and a variance of $\theta_1^2\sigma_{\log x}^2$.

Similarly, it can be shown that the limiting distribution of $\sqrt{n}(H - \theta_2)$ is normal with an arithmetic mean of zero and a variance of $\theta_2^4\sigma_{1/x}^2$, where $\theta_2 = \frac{1}{\alpha} = [E(1/x)]^{-1}$.

It is of some interest to observe that the expressions for the standard errors of the geometric and harmonic means correspond with the forms previously given for the standard errors of two efficient ratio-measures of relative variation,[4] namely,

$$\sigma_{G/A} = \frac{\theta_1^2}{\theta^2}\sigma_{A/G}, \quad \text{and} \quad \sigma_{H/G} = \frac{\theta_2^2}{\theta_1^2}\sigma_{G/H},$$

where θ_1/θ is the population geometric-arithmetic ratio, and θ_2/θ_1 is the population harmonic-geometric ratio.

3. Limitations of standard-error estimates. Application of these forms is subject to the usual conditions for drawing sound inferences on the basis of the representative method. Fiducial argument should be employed to avoid certain untenable assumptions of the outmoded method of using standard errors. Estimates of the standard deviations of sampling errors do not constitute an ultimate test of significance which can be applied with a high degree of success to all types of problems. In general, such estimates cannot be relied upon with a

[4] Nilan Norris, "Some efficient measures of relative dispersion," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 214–220.

high degree of confidence when they are used as tests of significance for index numbers, since in nearly all time series there exists an appreciable degree of serial correlation, persistence, or lack of independence among successive items of any sample.

4. Bibliographical note. Certain aspects of the sampling distribution of the geometric mean have been discussed by Burton H. Camp.[5] Attempts to derive forms for estimating the standard errors of index numbers have been made by Truman L. Kelley[6] and Irving Fisher,[7] and an empirical study of the sampling fluctuations of indexes has been made by E. C. Rhodes.[8] Although various special tests of significance for time series have been proposed,[9] at the present time no generally satisfactory procedure has appeared.

HUNTER COLLEGE,
NEW YORK, N. Y.

[5] Burton H. Camp, "Notes on the distribution of the geometric mean," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 221–226.

[6] Truman L. Kelley, "Certain Properties of Index Numbers," *Quarterly Publications of Am. Stat. Assn.*, Vol. 17, New Series 135, Sept., 1921, pp. 826–841.

[7] Irving Fisher, *The Making of Index Numbers*, Houghton Mifflin Company, New York, 1927, 3d ed., pp. 225–229, 342–345, and Appendix I, pp. 407 and 430 *f*.

[8] E. C. Rhodes, "The precision of index numbers," *Roy. Stat. Soc. Jour.*, Vol. 99 (1936), Part I, pp. 142–146, and Part II, pp. 367–369.

[9] Some of the more recent papers dealing with this matter are: G. Tintner, "On tests of significance in time series," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 139–143; "The analysis of economic time series," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 93–100; L. R. Hafstad, "On the Bartels technique for time-series analysis, and its relation to the analysis of variance," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 347–361; and Lila F. Knudsen, "Interdependence in a series," *Am. Stat. Assn. Jour.*, Vol. 35 (1940), pp. 507–514.

A NOTE ON THE USE OF A PEARSON TYPE III FUNCTION IN RENEWAL THEORY

BY A. W. BROWN

One of the methods suggested by A. J. Lotka[1] for the derivation of the renewal function may be briefly summarized as follows.

The method consists of dissecting the total renewal function into "generations". The original installation constitutes the zero generation, the units introduced to replace disused units of the zero generation constitute the first generation, renewal of these the second, and so on. Let $f(x)$ be the "mortality" function, the same for all generations. $f(x)$ is a function satisfying the usual conditions of a distribution function. Adopting Lotka's notation, let N be the number of units in the original collection, $B_1(t)\,dt$ the number of objects intro-

[1] A. J. Lotka, "A Contribution to the Theory of Self Renewing Aggregates, With Special Reference to Industrial Replacement," *Annals of Math. Stat.*, Vol. 10 (1939), p. 1.