THE DIFFERENCE BETWEEN THE INVERSES OF THE HARMONIC AND ARTHIMETIC MEANS

Monroe G. Sirken, National Center for Health Statistics

INTRODUCTION

Given a sequence of N positive integers

$$s = \{s_1, \ldots, s_\alpha, \ldots, s_N\}.$$

The inverse of the harmonic mean is

$$\frac{1}{N}\sum_{\alpha}^{N} 1/s_\alpha$$

and the inverse of the arithmetic mean is

$$\left(\frac{1}{N}\sum_{\alpha}^{N} s_\alpha\right)^{-1} = 1/\bar{s}.$$

Jensen's inequality states that

$$\frac{1}{N}\sum_{\alpha}^{N} 1/s_\alpha \geq 1/\bar{s}.$$

In this paper, we prove the identity

$$\frac{1}{N}\sum_{\alpha}^{N} 1/s_\alpha - 1/\bar{s} \equiv \frac{1}{N}\sum_{\alpha}^{N} s_\alpha \,(1/s_\alpha - 1/\bar{s})^2.$$

PROOF OF THE IDENTITY

$$\frac{1}{N}\sum_{\alpha}^{N} s_\alpha \,(1/s_\alpha - 1/\bar{s})^2$$

$$= \frac{1}{N}\sum_{\alpha}^{N} s_\alpha \left(\frac{1}{s_\alpha^2} - \frac{2}{\bar{s}\, s_\alpha} + \frac{1}{\bar{s}^2}\right)$$

$$= \frac{1}{N}\sum_{\alpha}^{N} \frac{1}{s_\alpha} - \frac{2}{\bar{s}} + \frac{1}{\bar{s}^2} \cdot \frac{1}{N}\sum_{\alpha}^{N} s_\alpha$$

$$= \frac{1}{N}\sum_{\alpha}^{N} \frac{1}{s_\alpha} - \frac{2}{\bar{s}} + \left(\frac{1}{\bar{s}^2}\right)\bar{s}$$

$$= \frac{1}{N}\sum_{\alpha}^{N} 1/s_\alpha - \frac{1}{\bar{s}}.$$

EXAMPLE

In this example, a random variable T is defined such that its variance is equal to the difference between the reciprocals of the harmonic and arithimetic means.

Given a population

$$I = \{I_1, \ldots, I_\alpha, \ldots, I_N\}$$

containing N persons each of whom has $s_\alpha \geq 1$ records in file

$$R = \{R_1, \ldots, R_j, \ldots, R_M\}$$

which contains M records, where

$$M = \sum_{\alpha}^{N} s_\alpha.$$

Let the indicator variable

$$\delta_{\alpha j} = \begin{cases} 1 \text{ if the } R_j^{th}\ (j=1,\ldots,M) \text{ record belongs to } I_\alpha\ (\alpha=1,\ldots,N) \\ 0 \text{ otherwise.} \end{cases}$$

and it follows that

$$M = \sum_{\alpha}^{N}\sum_{j}^{M} \delta_{\alpha j}$$

and

$$N = \sum_{\alpha}^{N}\sum_{j}^{M} \delta_{\alpha j}/s_\alpha.$$

A record R_j (j=1...,M) that belongs to I_α (α=1,...,N) is selected at random from file R. Let the random variable

$$T = \sqrt{\bar{s}}/s_\alpha$$

where

$$\bar{s} = \frac{1}{N}\sum_{\alpha}^{M} s_\alpha = M/N.$$

It follows that

$$E(T) \quad \frac{1}{M}\sum_{j}^{M}\sum_{\alpha}^{N} \delta_{\alpha j}(\sqrt{\bar{s}}/s_\alpha) = \sqrt{\bar{s}}\,\frac{N}{M} = 1/\sqrt{\bar{s}}$$

and

$$E(T^2) = \frac{1}{M}\sum_{j}^{M}\sum_{\alpha}^{N} \delta_{\alpha j}\,(\bar{s}/s_\alpha^2) = \bar{s}\,\frac{1}{M}\sum_{\alpha}^{N} 1/s_\alpha$$

$$= \bar{s}\,\frac{N}{M} \cdot \frac{1}{N}\sum_{\alpha}^{N} 1/s_\alpha = \frac{1}{N}\sum_{\alpha}^{N} 1/s_\alpha.$$

and

$$V(T) = E(T^2) - E^2(T)$$

$$= \frac{1}{N}\sum_{\alpha}^{N} 1/s_\alpha - 1/\bar{s}.$$

Also,

$$V(T) = E[T - E(T)]^2$$

$$= \frac{1}{M}\sum_{\alpha}^{N}\sum_{j}^{M} (\sqrt{\bar{s}}/s_\alpha - 1/\sqrt{\bar{s}})^2 \ \delta_{\alpha j}$$

$$= \frac{1}{M}\sum_{\alpha}^{N} s_\alpha \ (\sqrt{\bar{s}}/s_\alpha - 1/\sqrt{\bar{s}})^2$$

$$= \frac{\bar{s}}{M}\sum_{\alpha}^{N} s_\alpha \ (1/s_\alpha - 1/\bar{s})^2$$

$$= \frac{1}{N}\sum_{\alpha}^{N} s_\alpha \ (1/s_\alpha - 1/\bar{s})^2.$$

Thus,

$$V(T) = \frac{1}{N}\sum_{\alpha}^{N} 1/s_\alpha - 1/\bar{s}$$

$$= \frac{1}{N}\sum_{\alpha}^{N} s_\alpha \ (1/s_\alpha - 1/\bar{s})^2.$$

APPLICATION

There exists a file of $M > N$ records that belong to N persons. M is known and N is unknown. A random sample of m records is selected to estimate N. Assume that every sample record belonging to $I_\alpha(\alpha=1,\dots,N)$ is labeled by s_α, the total number of I_α's records in the file. An unbiased multiplicity estimator [1] of N is

$$\hat{N} = \frac{M}{m}\sum_{\alpha}^{N}\sum_{j}^{M} \delta_{\alpha j}/s_\alpha.$$

Assuming simple random sampling with replacement, the variance of $\hat{N}$ is

$$V(\hat{N}) = V\left(\frac{M}{m}\sum_{\alpha}^{N}\sum_{j}^{M} \delta_{\alpha j}/s_\alpha\right)$$

$$= \frac{M^2}{m\bar{s}} V(T)$$

$$= \frac{M}{m\bar{s}}\left(\frac{1}{N}\sum 1/s_\alpha - 1/\bar{s}\right).$$

$$= \frac{1}{m}\sum_{\alpha}^{N} s_\alpha \ (1/s_\alpha - 1/\bar{s})^2.$$

REFERENCE

[1] Sirken, Monroe G. "Variance Components of Multiplicity Estimators", Biometrics, Vol. 28, No. 3, September 1972, pp. 869-873.