de Carvalho, M. (2016+), "Mean, What do you Mean?" The American Statistician, in press (DOI: 10.1080/00031305.2016.1148632).

Mean, What do you Mean?

Miguel de Carvalho

Abstract

When teaching statistics we often resort to several notions of mean, such as *arithmetic mean*, *geometric mean* and *harmonic mean*, and hence the student is often left with the question: The word mean appears in all such concepts, so what is actually a mean? I revisit Kolmogorov's axiomatic view of the mean, which unifies all these concepts of mean, among others. A population counterpart of the notion of regular mean, along with notions of regular variance and standard deviation will also be discussed here as unifying concepts. Some examples are used to illustrate main ideas.

KEY WORDS: Arithmetic mean; Geometric mean; Harmonic mean; Regular mean.

1 INTRODUCTION

How well have we been teaching arithmetic, harmonic, and geometric means to our students? In a recent paper by C. R. Rao and colleagues (Rao et al., 2014), we read

> "Although the harmonic mean (HM) is mentioned in textbooks along with the arithmetic mean (AM) and the geometric mean (GM) as three possible ways of summarizing the information in a set of observations, its appropriateness in some statistical applications is not mentioned in textbooks."

In the classroom the student is confronted with several notions of mean, and thus a pertinent question that an attentive student could pose is: "The word mean appears in all these concepts, so what is actually a mean?"

Miguel de Carvalho is Associate Professor, Faculty of Mathematics, Pontificia Universidad Católica de Chile, Santiago, Chile (*mmbbcarvalho@gmail.com*). This work was partially supported by the grant Fondecyt 11121186 from the Chilean NSF (Conicyt).

In this short paper I revisit Kolmogorov's axiomatic view of the mean, which unifies all these concepts of mean, among others. While Kolmogorov's axioms of probability are widely known (DeGroot and Schervish, 2011, Section 1.5), it is perhaps less well known that in an often-forgotten note, Kolmogorov also proposed an axiomatic construction for what a unifying concept of mean should be (Kolmogorov, 1930). Let $\mathbf{x} = (x_1, \ldots, x_n)$ and let $\mathbf{1}$ denote a vector of ones. Formally, a '*regular* (type of) *mean*'† is a map $M : \mathbb{R}^n \mapsto \mathbb{R}$ which obeys the following axioms:

A1) $M(\mathbf{x})$ is continuous and increasing in each variable.

A2) $M(\mathbf{x})$ is a symmetric function.

A3) $M(x\mathbf{1}_n) = x$, i.e., the mean of repeated data equals the repeated value.

A4) The mean of the combined sample, $\mathbf{x}$, remains unchanged if a part of the sample is replaced by its corresponding mean, $m = M(x_1, \ldots, x_{n_1})$, i.e.

$$M(\mathbf{x}) = M(\underbrace{m, \ldots, m}_{n_1}, x_{n_1+1}, \ldots, x_n).$$

Kolmogorov (1930) proved that if conditions A1) to A4) hold, then the function $M(\mathbf{x})$ has the form

$$M_g(\mathbf{x}) = g^{-1}\Bigg(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\Bigg), \quad \mathbf{x} = (x_1, \ldots, x_n), \tag{1}$$

where g is a continuous monotone function and g^{-1} is its inverse function. Here are some examples.

Example 1 (Arithmetic Mean)**.** If $g(x) = x$, we obtain $M_x(\mathbf{x}) = 1/n \sum_{i=1}^n x_i$.

Example 2 (Geometric Mean)**.** If $g(x) = \log x$, for $x > 0$, we obtain $M_{\log x}(\mathbf{x}) = (\prod_{i=1}^n x_i)^{1/n}$.

Example 3 (Harmonic Mean)**.** If $g(x) = 1/x$, for $x > 0$, we obtain $M_{1/x}(\mathbf{x}) = n/\sum_{i=1}^n x_i^{-1}$.

Equation (1) also includes as particular cases other examples, such as the so-called power mean: For example, if $g(x) = x^p$, for $p \in (0, \infty)$, we obtain $M_{x^p}(\mathbf{x}) = \{1/n \sum_{i=1}^n x_i^p\}^{1/p}$.

Construction (1) may help the student understand the common link between all concepts of mean, but leaves him with the questions: What is the population counterpart of the notion of regular mean? Can we think of a notion of regular variance?

†I follow the terminology of the English translation in Tikhomirov et al. (1991, p. 144).

2 REGULAR MOMENTS

A population counterpart of the notion of regular mean can be naturally constructed, and originates a broad concept of expected value to which we refer as *Kolmogorov expected value*,

$$E_g(X) = g^{-1}(E_x\{g(X)\}), \tag{2}$$

where $E_x\{g(X)\} = \int g(x)\,\mathrm{d}F(x)$, with $X \sim F$, and g is a continuous monotone function. For ease of notation, below we write 'E' to denote the (usual) expected value 'E_x,' and we apply the same convention for the (usual) standard deviation and variance. Some examples are instructive. For a positive (non-degenerate) random variable X, its geometric and harmonic expected values can be written as

$$E^G(X) := E_{\log x}(X) = \mathrm{e}^{E(\log X)}, \quad E^H(X) := E_{1/x}(X) = \frac{1}{E(1/X)}.$$

Thus, just as the notion of regular mean unifies several concepts of mean, the Kolmogorov expected value unifies the population counterparts. And readily, we can also construct a concept of *Kolmogorov variance*, to be defined as

$$\mathrm{var}_g(X) = g^{-1}(E\{[g(X) - E\{g(X)\}]^2\}) = g^{-1}(\mathrm{var}_x\{g(X)\}). \tag{3}$$

Here g is a continuous monotone function whose inverse function is strictly positive. Examples of Kolmogorov variance include

$$\mathrm{var}^G(X) := \mathrm{var}_{\log x}(X) = \mathrm{e}^{\mathrm{var}(\log X)}, \quad \mathrm{var}^H(X) := \mathrm{var}_{1/x}(X) = \frac{1}{\mathrm{var}\,(1/X)}, \tag{4}$$

with $x > 0$. The interpretation of these 'variances' may challenge one's intuition. For example, if X is degenerate ('constant') then $\mathrm{var}^G(X) = 1$, and not 0 as one could expect. How can we explain this? First, we need to recall that the geometric mean was built to model multiplicative variation, rather than additive variation (Galton, 1879). Geometric variance, like the geometric mean, is a multiplicative concept: Division or multiplication by $1 = \mathrm{var}^G(X)$ produces no changes in the data. The analogy with the standard concept of variance, is that for a degenerate random variable $\mathrm{var}(X) = 0$, which is consistent with the fact that the standard concept of variance is suitable for modeling additive variation: Adding or subtracting $0 = \mathrm{var}(X)$ produces no changes in the data.

Empirical estimators of $\mathrm{var}^G(X)$ and $\mathrm{var}^H(X)$ are

$$\widehat{\mathrm{var}}^G = \exp\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\log(x_i) - \frac{1}{n}\sum_{j=1}^{n}\log(x_j)\right\}^2\right], \quad \widehat{\mathrm{var}}^H = \frac{n}{\sum_{i=1}^{n}\{1/x_i - \frac{1}{n}\sum_{j=1}^{n} 1/x_j\}^2},$$

and by analogy estimators for an arbitrary g can be devised.

By analogy with (2) and (3), we define *Kolmogorov standard deviation* as

$$\mathrm{sd}_g(X) = g^{-1}(\mathrm{sd}_x\{g(X)\}).$$

By defining (2) in this way, $\{\mathrm{var}_g(X)\}^{1/2}$ need not be equal to $\mathrm{sd}_g(X)$, but there is a convenient reason for opting for this definition instead of $\{\mathrm{var}_g(X)\}^{1/2}$. Note first that

$$\mathrm{sd}^H(X) := \mathrm{sd}_{1/x}(X) = \frac{1}{\mathrm{sd}(1/X)},$$

and hence there are cases for which $\{\mathrm{var}_g(X)\}^{1/2}$ equals $\mathrm{sd}_g(X)$, such as $\{\mathrm{var}_{1/x}(X)\}^{1/2} = \mathrm{sd}_{1/x}(X)$, and the obvious case $\{\mathrm{var}_x(X)\}^{1/2} = \mathrm{sd}_x(X)$. Now

$$\mathrm{sd}^G(X) := \mathrm{sd}_{\log x}(X) = \mathrm{e}^{\mathrm{sd}(\log X)}, \tag{5}$$

and hence $\{\mathrm{var}_{\log x}(X)\}^{1/2} \neq \mathrm{sd}_{\log x}(X)$. However, as we shall see in Section 3.2, definition (5) agrees with the concept of geometric standard deviation which was first introduced by Kirkwood (1979), while the alternative definition $\{\mathrm{var}_{\log x}(X)\}^{1/2}$ would not. Empirical estimators of $\mathrm{sd}^G(X)$ and $\mathrm{sd}^H(X)$ are

$$\widehat{\mathrm{sd}}^G = \exp\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\left\{\log(x_i) - \frac{1}{n}\sum_{j=1}^{n}\log(x_j)\right\}\right\}^{1/2}\right], \quad \widehat{\mathrm{sd}}^H = \sqrt{\widehat{\mathrm{var}}^H},$$

and by analogy estimators for an arbitrary g can be devised.

3 EXAMPLES

Casual end-of-class conversations with Julio Avila, and over-the-coffee discussions with Filipe Marques and Vanda Inácio de Carvalho, led me to the following examples.

3.1 Applications in Mathematical Finance and Labor Economics

For an intermediate level course on applied statistics, I recommend the following examples.

Example 4 (Mathematical Finance)**.** It is a basic fact in Finance that an investor may be expected to require a rate of interest which depends on the investment horizon; for example, an investor could be expected to demand a higher rate of interest for an investment over a longer horizon. This basic

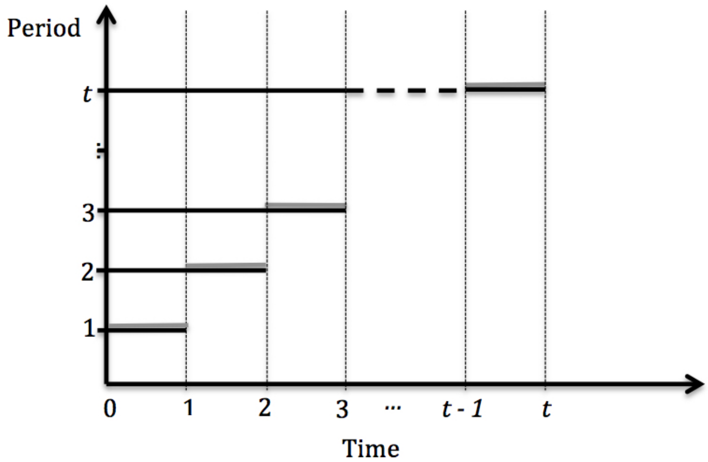

Figure 1: Relevant periods for spot rates of interest (black) and forward rates of interest (gray).

principle motivates the need for concepts of time-varying interest rates, such as those of spot and forward rates of interest. For an investment at time 0, receiving interest over t periods, we suppose that the rate of interest earned depends on the investment horizon. Roughly speaking, the spot rate of interest (i_t^S) consists of the rate from period 0 to period t, whereas the forward rate of interest (i_t^F) is the rate applicable to the period ranging from $t-1$ to t (Chan and Tse, 2011, Ch. 3); see Figure 1 for the relevant periods underlying each of these rates. Formally, the spot and forward rates of interest relate as follows

$$(1 + i_t^S)^t = (1 + i_1^F) \times \cdots \times (1 + i_t^F). \tag{6}$$

See for instance Chan and Tse (2011, Eq. (3.5)). A consequence of (6) is that

$$i_t^S = \underbrace{\{(1 + i_1^F) \times \cdots \times (1 + i_t^F)\}^{1/t}}_{\text{geometric mean}} - 1 = M_{\log x}(1 + i_1^F, \ldots, 1 + i_t^F) - 1, \tag{7}$$

so that the spot rate of interest is given by the geometric mean of the (one plus) forward rates, minus 1. For example, if the forward rates of interest for investments in years 1 and 2 are respectively 3.9% and 4.5%, the spot rate for $t = 2$ is

$$i_2^S = \{(1 + i_1^F)(1 + i_2^F)\}^{1/2} - 1 = \{1.039 \times 1.045\}^{1/2} - 1 \approx 4.2\%.$$

Example 5 (Unemployment Duration Analysis)**.** Here, I discuss the subject of length-bias sampling, in the context of unemployment duration analysis. Suppose that at time t_0 a cross-sectional study is performed by random sampling n subjects who (at that time) have a particular status such as

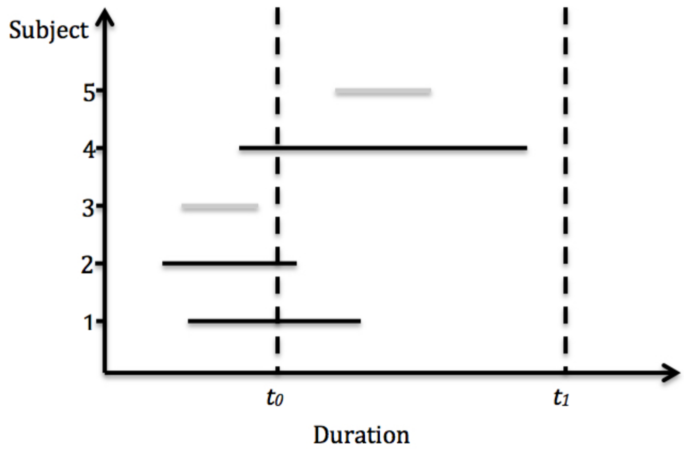

Figure 2: Length-biased sampling. This graphical representation aims to clarify why longer unemployment durations are overrepresented, as they are more likely to be intersected at t_0; subjects in grey would not be in unemployment at t_0, and hence would not be included in the study.

being currently unemployed. These subjects are then followed-up until time t_1, and the aim is on assessing the duration of unemployment in the population; see Figure 2. This sampling scheme is known in econometrics as stock sampling with follow-up (Lancaster, 1990), and in biostatistics as prevalent cohort study (Wang, 1998). (Censoring is also an issue, but I will ignore it here.) With stock sampling with follow-up, instead of random sampling from the true duration T, with survivor function $S(t) := P(T > t)$, we are only able to obtain a random sample from the length-biased duration T_{LB}, with survivor function

$$S_{\mathrm{LB}}(t) = P(T_{\mathrm{LB}} > t) = 1 - \int_0^t \frac{u}{E(T)}\,\mathrm{d}F(u), \quad 0 \leqslant t < \infty, \tag{8}$$

and this leads to an overrepresentation of longer unemployment durations; see Wang (1998) for details. The intuition is that by sampling at a fixed point in time (t_0) we will intersect with a larger probability, subjects who are unemployed for longer periods; see Figure 2. So, how can we compute the expected value of the true unemployment duration (T), from its length-biased counterpart (T_{LB})? It turns out that from (8) we get $f_{\mathrm{LB}}(u) = uf(u)/E(T)$, and thus

$$E\left(\frac{1}{T_{\mathrm{LB}}}\right) = \int_0^\infty \frac{1}{u} f_{\mathrm{LB}}(u)\,\mathrm{d}u = \int_0^\infty \frac{1}{u}\frac{uf(u)}{E(T)}\,\mathrm{d}u = \frac{1}{E(T)},$$

from which we can see an important well-known connection:

$$E(T) = E^H(T_{\mathrm{LB}}) = \frac{1}{E(1/T_{\mathrm{LB}})}. \tag{9}$$

In words: The mean of the 'true' duration, equals the harmonic mean of the length-biased duration. A simple parametric example may be considered to illustrate this further. Suppose T and T_{LB}

respectively denote unemployment duration and length-biased unemployment duration (in weeks). Suppose that the true duration is distributed according to $f(t) = 1/30\,\mathrm{e}^{-t/30}$, for $t > 0$, that is $T \sim \mathrm{Exp}(1/30)$. This implies that, $f_{\mathrm{LB}}(t) = (1/30)^2 t\mathrm{e}^{-t/30}$, for $t > 0$, i.e. $T_{\mathrm{LB}} \sim \mathrm{Gamma}(2, 1/30)$. Thus $E(T) = 30$ weeks, whereas $E(T_{\mathrm{LB}}) = 60$ weeks. Note also that from (9) it follows that $E(1/T_{\mathrm{LB}}) = 1/30$ weeks.

3.2 Geometric Moments of Log Normal

The term *geometric standard deviation* was introduced in a letter to the Editor of *Biometrics* by Kirkwood (1979).† Kirkwood introduced the concept informally—in words—and simply for the log-normal distribution, and this has led to some awkward misinterpretations by some other authors. Let $X \sim \mathrm{LN}(\mu, \sigma^2)$. After some applied motivation justifying the need for the concept, Kirkwood (1979) introduced the concept by claiming that:

> "The geometric mean of X is e^{μ}, and let us define the geometric standard deviation (GSD) to be e^{σ}."

Kirkwood's concept of geometric standard deviation is compatible with the definition of Kolmogorov standard deviation from Section 2. Since $X \sim \mathrm{LN}(\mu, \sigma^2)$, it follows that $X = \mathrm{e}^{\mu+\sigma Z}$ with $Z \sim N(0, 1)$. Thus, if $g(x) = \log x$, we have that

$$\mathrm{var}\{g(X)\} = \mathrm{var}(\mu + \sigma Z) = \sigma^2,$$

and thus $\mathrm{sd}_{\log x} X = g^{-1}(\mathrm{sd}\{g(X)\}) = \mathrm{e}^{\sigma}$. Another way to see this, follows from formula (5),

$$\mathrm{sd}^G(X) := \mathrm{sd}_{\log\, x}(X) = \mathrm{e}^{\mathrm{sd}(\log\, X)} = \mathrm{e}^{\mathrm{sd}(\mu+\sigma Z)} = \mathrm{e}^{\sigma}.$$

4 CONSEQUENCES OF EARLIER DEFINITIONS

4.1 Decision-Theoretic Comment on Regular Means

It is well known that minimization of a quadratic loss leads to the arithmetic mean, i.e.

$$M_x(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i = \arg\min_{\mu} \sum_{i=1}^{n}(x_i - \mu)^2.$$

A natural question from a decision-theoretic viewpoint is thus: "Can we obtain any regular mean, from the first order conditions of an objective function of interest?" It turns out that it is not difficult

†By May 20 2015, Kirkwood's one-page letter to the Editor registered a total of 163 citations in Google Scholar.

to verify that if g is strictly monotone, the first order conditions of $\sum_{i=1}^{n}\{g(x_i)-g(\mu)\}^2$ lead to $M_g(\mathbf{x}) = g^{-1}(1/n\sum_{i=1}^{n} g(x_i))$, so that for example the first order conditions of $\sum_{i=1}^{n}(\log x_i - \log\mu)^2$ yield $M_{\log x}(\mathbf{x}) = (\prod_{i=1}^{n} x_i)^{1/n}$, and that those of $\sum_{i=1}^{n}(1/x_i - 1/\mu)^2$ yield $M_{1/x}(\mathbf{x}) = n/\sum_{i=1}^{n} x_i^{-1}$.

4.2 A Central Limit Theorem for Regular Means

A central limit theorem can also be obtained. Throughout let $\mathbf{X} = (X_1, \dots, X_n)$.

Theorem 1. *Suppose that* $X_1, \dots, X_n \overset{\text{iid}}{\sim} F_X$, *and that* $\operatorname{var}\{g(X)\} < \infty$, *where* $g(x)$ *is a strictly monotone function with derivative* $g'(E_g(X))$ *at* $x = E_g(X)$. *Then,*

$$\sqrt{n}\{M_g(\mathbf{X}) - E_g(X)\} \overset{d}{\to} N\left(0, \frac{\operatorname{var}\{g(X)\}}{\{g'(E_g(X))\}^2}\right), \quad \textit{as } n \to \infty.$$

The proof is immediate but instructive. It follows from a straightforward application of the delta method (Knight, 2000, Theorem 3.4), which in its simplest form implies that if $\sqrt{n}(\widehat{\theta}-\theta) \overset{\text{d}}{\to} N(0,\sigma^2)$, then $\sqrt{n}\{h(\widehat{\theta}) - h(\theta)\} \overset{\text{d}}{\to} N(0, \sigma^2\{h'(\theta)\}^2)$, where $h(x)$ is a function with derivative $h'(\theta)$ at $x = \theta$.

Proof. From the central limit theorem for iid observations (Knight, 2000, Theorem 3.8) it follows

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) - E\{g(X)\}\right) \overset{\text{d}}{\to} N(0, \operatorname{var}\{g(X)\}), \quad \text{as } n \to \infty.$$

From the delta method and the fact that $(g^{-1})'(x) = 1/\{g'(g^{-1}(x))\}$ it follows that as $n \to \infty$

$$\sqrt{n}\Bigg\{\underbrace{g^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right)}_{M_g(\mathbf{X})} - \underbrace{g^{-1}(E\{g(X)\})}_{E_g(X)}\Bigg\} \overset{\text{d}}{\to} N\left(0, \operatorname{var}\{g(X)\}\frac{1}{\{g'(\underbrace{g^{-1}(E\{g(X)\})}_{E_g(X)})\}^2}\right).$$

□

A simple corollary is a central limit theorem for geometric and harmonic means.

Corollary 1. *Suppose that* $X_1, \dots, X_n \overset{\text{iid}}{\sim} F_X$, *where* X *is a positive random variable. Let* $E^G(X) = e^{E(\log X)}$ *and* $E^H(X) = 1/E(1/X)$.

1. *If* $\operatorname{var}(\log X) < \infty$, *then as* $n \to \infty$

$$\sqrt{n}\left\{\left(\prod_{i=1}^{n} X_i\right)^{1/n} - E^G(X)\right\} \overset{d}{\to} N(0, \operatorname{var}(\log X)\{E^G(X)\}^2).$$

2. *If* $\operatorname{var}(1/X) < \infty$, *then as* $n \to \infty$

$$\sqrt{n}\left\{\frac{n}{\sum_{i=1}^{n} 1/X_i} - E^H(X)\right\} \overset{d}{\to} N(0, \operatorname{var}(1/X)\{E^H(X)\}^4).$$

Corollary 1 is tantamount to results discussed by Pakes (1999).

5 DISCUSSION

The idea of thinking in terms of means seems so natural nowadays that it is difficult to realize it has not always been in use. Quoting Plackett (1958):

> "The history of the problem of combining a set of independent observations on the same quantity is traced from antiquity to the appearance in the eighteenth century of the arithmetic mean as a statistical concept."

As a mathematical concept, the history of the notion of the arithmetic mean—as well as that of other concepts of mean—can however be traced much further back in history, and have been used by Greek mathematicians long ago (Heath, 1981, Chapter III).

In many textbooks the student is left with a wealth of notions of mean without a direction on what is the connection between them. Kolmogorov's axiomatic view of the mean offers a rich conceptual framework for teaching the links between pythagorean means, i.e., arithmetic, geometric, and harmonic means. We underscore that M. Nagumo and B. de Finetti obtained results which are tantamount to (1). Indeed, the resemblance between their research is such that Hardy et al. (1934, Theorem 215) and Cifarelli and Regazzini (1996, pp. 270–271) make a reference to Nagumo, de Finetti, and Kolmogorov when presenting their versions of (1). Kolmogorov's expected value as defined in (2) could be constructed directly from a related axiomatic treatment (Hardy et al., 1934, Sections 6.19–6.22), instead of being motivated as a population counterpart of a regular mean.

Geometric and harmonic moments are of interest in a variety of applied settings; see, for instance, Ney and Vidyashankar (2003), Piau (2006), Luati et al. (2012) and Rao et al. (2014), just to name a few recent contexts. Equation (9) can be used for motivating a celebrated estimator proposed by Cox (1969) for the mean duration in the presence of length-bias. Length-biased sampling is still an active field of research; see Ning et al. (2014) and references therein.

The materials and methods in the paper were envisioned for an intermediate level course. Teaching introductory statistics is by all standards challenging in many dimensions (Gelman and Nolan, 2002). Depending on the mathematical background of students, it may be sensible to cover in an introductory course main ideas and examples from Section 1, only, along with a battery of data examples for illustrating the concepts. For an introductory course for mathematicians, it could be instructive to introduce the simple, but enlightening, ideas on the geometry of arithmetic, geometric and harmonic means, discussed by Eves (2003).

REFERENCES

Chan, W.-S., and Tse, Y.-K. (2011), *Financial Mathematics for Actuaries*, Singapore: McGraw-Hill.

Cifarelli, D., M., and Regazzini, E. (1996), "De Finetti's Contribution to Probability and Statistics," *Statistical Science*, 11, 253–282.

Cox, D. R. (1969), "Some Sampling Problems in Technology," in *New Developments in Survey Sampling*, Eds. N. L. Johnson and H. Smith, pp. 506–527, New York: Wiley.

DeGroot, M. H., and Schervish, M. J. (2011), *Probability and Statistics* (4th ed.), Boston: Pearson.

Eves, H. (2003), "Means Appearing in Geometric Figures," *Mathematics Magazine*, 76, 292–294.

Galton, F. (1879), "The Geometric Mean in Vital and Social Statistics," *Proceedings of the Royal Society*, 29, 365–367.

Gelman, A., and Nolan, D. (2002), *Teaching Statistics: A Bag of Tricks*, Oxford, UK: Oxford University Press.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1934), *Inequalities*, Cambridge, UK: Cambridge University Press.

Heath, T. (1981), *A History of Greek Mathematics*, Vol. 1, Oxford, UK: Oxford University Press.

Kirkwood, T. B. L. (1979), "Geometric Means and Measures of Dispersion," *Biometrics*, 35, 908–909.

Knight, K. (2000), *Mathematical Statistics*, Boca Raton, FL: Chapman & Hall/CRC.

Kolmogorov, A. (1930), "Sur la Notion de la Moyenne," Atti della Academia Nazionale dei Lincei, 12, 323–343.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge, UK: Cambridge University Press.

Luati, A., Proietti, T., and Reale, T. (2012), "The Variance Profile," *Journal of the American Statistical Association*, 107, 607–621.

Ney, P. E., and Vidyashankar, A. N. (2003), "Harmonic Moments and Large Deviation Rates for Supercritical Branching Processes," *Annals of Applied Probability*, 13, 475–489.

Ning, J., Qin, J., and Shen, Y. (2014), "Semiparametric Accelerated Failure Time Model for Length-Biased Data with Application to Dementia Study," *Statistica Sinica*, 24, 313–333.

Pakes, A. G. (1999), "On the Convergence of Moments of Geometric and Harmonic Means," *Statistica Neerlandica*, 53, 96–110.

Piau, D. (2006), "Harmonic Moments of Inhomogeneous Branching Processes," *Advances in Applied Probability*, 38, 465–486.

Plackett, R. L. (1958), "Studies in the History of Probability and Statistics: VII. The Principle of the Arithmetic Mean," *Biometrika*, 45, 130–135.

Rao, C. R., Shi, X., and Wu, Y. (2014), "Approximation of the Expected Value of the Harmonic Mean and some Applications," *Proceedings of the National Academy of Sciences*, 111, 15682–15686.

Tikhomirov, V. M. (1991), *Selected Works of A. N. Kolmogorov, Vol. 1: Mathematics and Mechanics*, Dordrecht: Kluwer.

Wang, M.-C. (1998), "Length-Bias," in *Encyclopedia of Biostatistics*, Eds. Peter Armitage and Theodore Colton, Vol. 3 pp. 2223–2226, Wiley: New York.