TRANSACTIONS OF SOCIETY OF ACTUARIES
1990 VOL. 42

A STATISTICAL ANALYSIS OF BANDED DATA WITH APPLICATIONS

ROBERT R. REITANO

ABSTRACT

The goal of this paper is to develop best possible estimates for the higher moments of a distribution of a positive bounded random variable, such as claim amounts, where these estimates are given in terms of the mean. First, upper and lower sharp estimates are developed for the second moment and variance in the case of a discrete random variable. A variety of applications are considered in detail with particular emphasis on claims analysis. Then these upper and lower estimates are generalized to higher-order moments and to continuous random variables, as well as to the associated moment-generating functions.

I. INTRODUCTION

When analyzing the random variable S, defined to equal the aggregate amount of claims in a given time, the sums of in-force policy amounts, sums of squares, and so on must be calculated. Although this is formalized in Section III (Applications), these sums must be calculated for each homogeneous class of policyholders, where homogeneity is defined with respect to the underlying claim probability distribution function (pdf). For sizable portfolios, this summing process can be formidable and is often precluded by the summarization of experience into banded amount classes. The purpose of this paper is to develop sharp estimates for the values of the higher moments of the policy amount distribution, or any pdf of a positive bounded random variable (rv), where these estimates are made using only the first moment or mean.

The motivation for this investigation was the desire to estimate an amount-based standard deviation for life insurance claims. Because experience data were already banded into amount classes, an exact calculation was impossible; however, the means of each of the bands were known, so it was natural to inquire as to what approximations would be possible having only these values. As expected, the error in these approximated standard deviations is greatly influenced by the relative band size. More importantly, given any error tolerance, band size can be predetermined so that the resulting estimates are at least that accurate.

Here, the term "estimate" is used in the sense of "a priori estimates" of mathematical analysis and not in the sense of statistical estimation. That is, for a given pdf defined on a positive bounded rv, the value of all higher moments is constrained, a priori, once the first moment and domain interval are given. For brevity, the upper and lower bounds developed are referred to as estimates rather than a priori estimates. As usual, the qualification of an estimate as "sharp" means that it is the best possible.

In Section II, this definition is formalized, and sharp estimates are developed for the second moment of any finite collection of positive numbers, such as policy or claim amounts. As a corollary, these estimates are translated into estimates of the variance. The relative accuracy of these estimates over an interval $[a,b]$, $a>0$, depends on $r=b/a$ and can be controlled by "banding" amount groups properly. In this section, it also becomes clear which types of discrete distributions maximize the relative error for given values of r and the mean μ, and which discrete distribution maximizes the error for a given r with no restriction on μ.

In Section III, a number of applications are explored in detail. For example, interval and point estimates for the variance of expected claims and modified confidence intervals are developed. Both fixed and variable claim size models are considered. The estimates developed in Section II are shown to be applicable both in the context of current amount bands, as well as in determining, a priori, what banding size is needed for a given maximum error tolerance. In addition, applications are made to the problems of establishing an appropriate level for reinsurance retention limits and of analyzing the variance of decrement estimators.

Section IV generalizes Section II to the case of higher moments of an arbitrary discrete pdf, and in Section V, the higher moments of a continuous pdf are considered. As a corollary, these estimates are used to develop sharp estimates for the associated moment generating functions.

II. SECOND MOMENTS—DISCRETE CASE (SPECIAL)

Let $\{x_i\}_{i=1}^n$ be a collection of numbers from the interval $[a,b]$, $a>0$. In this section, sharp estimates are developed for the second moment, $\mu_2'(x)$, and the ratio $R_2(x)$, where

$$\mu_2'(x) = \frac{1}{n} \sum x_i^2, \tag{2.1}$$

$$R_2(x) = \mu_2'/\mu^2, \tag{2.2}$$

and

$$\mu = \frac{1}{n} \sum x_i.$$

The estimates for $\mu_2'(x)$ are simple functions of μ and r, while those for $R_2(x)$ are functions of r alone.

Considering the distribution $\{\lambda x_i\}_{i=1}^n$ from $[\lambda a, \lambda b]$, $\lambda > 0$, it is clear that:

$$\mu(\lambda x) = \lambda \mu(x), \tag{2.3}$$

$$\mu_2'(\lambda x) = \lambda^2 \mu_2'(x), \tag{2.4}$$

$$R_2(\lambda x) = R_2(x). \tag{2.5}$$

Consequently, letting $\lambda = 1/a$, (2.1) and (2.2) need be estimated only for $\{x_i\}$ in $[1,r]$, since (2.3) through (2.5) can be applied to yield the analogous results for $\{x_i\}$ in $[a,b]$.

An upper estimate $s_2(r,\mu)$ for $\mu_2'(x)$ is defined to be sharp if $\mu_2'(x) \le s_2(r,\mu)$ for all $\{x_i\} \subset [1,r]$, but for any $\epsilon > 0$, $r > 1$, and μ, $1 \le \mu \le r$, there exists a distribution $\{y_i\} \subset [1,r]$ with mean equal to μ and

$$\mu_2'(y) > s_2(r,\mu) - \epsilon. \tag{2.6}$$

An upper estimate $s_2(r)$ for $R_2(x)$ is defined to be sharp in an analogous way. That is, $R_2(x) \le s_2(r)$, but for any $\epsilon > 0$, $r > 1$, there is a distribution $\{y_i\} \subset [1,r]$, so that

$$R_2(y) > s_2(r) - \epsilon. \tag{2.7}$$

Sharp lower estimates are defined analogously.

Since it is certainly true that $R_2(x) \le r^2$, it is clear that $s_2(r)$ will satisfy $s_2(r) \le r^2$ and hence given any $d > 0$, there is an $r > 1$ such that $s_2(r) \le 1 + d$. Consequently, since we also have $R_2(x) \ge 1$ [see (2.36) below],

$$\mu^2 \le \frac{1}{n} \sum x_i^2 \le (1 + d)\, \mu^2, \quad \{x_i\} \subset [1,r]. \tag{2.8}$$

That is, the second moment of $\{x_i\}$ can be approximated with the first moment to any given degree of accuracy by choosing amount bands $[a,b]$ with $r = b/a$ close enough to 1.

As an application, it is shown in Section III.2 that the standard deviation of $\{x_i\}$ can be approximated with the mean to within a 5 percent relative error by choosing $d = 0.22$. Using the observation that $s_2(r) \le r^2$, it is clear that $s_2(r) \leq 1.22$ if $r \leq 1.1045$. If x_i represents a life insurance policy amount, the standard deviation of expected claims, $\sigma(S)$, can therefore be approximated to within 5 percent by "banding" amounts into $[ar^j, ar^{j+1}]$, $j \geq 0$, with $r = 1.1045$. Unfortunately, analysis of experience between $10,000 and $100,000,000 would require more than 90 such bands, and this greatly limits the potential usefulness of this approach. Fortunately, this conclusion is a result of the crudeness of the estimate $s_2(r) \doteq r^2$ and not of the general weakness of this approach. It is shown below, using the actual $s_2(r)$, that $d = 0.22$ can be obtained with the ratio $r = 2.48$, and this reduces the number of bands needed in this example from 93 to 11, an easily workable number.

In order to make the problem more tractable, the analysis of $\mu'_2(x)$ and $R_2(x)$ can be reduced from the collection of all distributions $\{x_i\}_{i=1}^n \subset [1,r]$ to a linearly parametrized collection of distributions, denoted $D(t)$. This parametrization is defined on the interval $[0,n]$ and produces one distribution corresponding to each mean μ, $1 \leq \mu \leq r$. To see this, let $\{x_i\}_{i=1}^n \subset [1,r]$ be given and assume that $1 < x_1 < x_2 < r$. Let δ satisfy $0 < \delta < \min(r - x_2, x_1 - 1)$, and define $\{y_i\}_{i=1}^n$ by:

$$\begin{aligned} y_1 &= x_1 - \delta \\ y_2 &= x_2 + \delta \\ y_i &= x_i, \qquad 3 \leq i \leq n. \end{aligned} \tag{2.9}$$

Then $\mu(y) = \mu(x)$, and

$$\mu'_2(y) = \mu'_2(x) + \frac{2\delta}{n}(\delta + x_2 - x_1). \tag{2.10}$$

Consequently, $\mu'_2(y) > \mu'_2(x)$, and this example illustrates that for a given mean, μ, the distribution over $[1,r]$ that maximizes both $\mu'_2(x)$ and $R_2(x)$ has the property that all but at most one value, x_j, equals 1 or r. Because of this property, such distributions are called *polarized distributions* and can be parametrized over $t \in [0,n]$ by $D(t)$, where $D(t) = \{x_i\}_{i=1}^n$ is defined by:

$$x_i = \begin{cases} 1 & , \quad i \leq n - [\![t]\!] - 1, \\ (t - [\![t]\!])(r - 1) + 1, & \quad i = n - [\![t]\!], \\ r & , \quad i \geq n - [\![t]\!] + 1. \end{cases} \tag{2.11}$$

Here, as usual, $[\![t]\!]$ represents the greatest integer less than or equal to t. If one envisions the distributions $D(t)$ as the bead positions of an abacus with n rods and one bead per rod, the parametrization in (2.11) smoothly moves one bead at a time from one "one-sided" bead position to the other. Alternatively, if $\{x_i\}$ is identified with a point in n-dimensional space, $\mathbf{R}^n$, $D(t)$ can be thought of as a piecewise linear transformation from $[0,n]$ to a one-dimensional edge of the hypercube $[1,r]^n$ extending from $(1, 1, ..., 1)$ to $(r, r, ... r)$.

Since

$$\mu[D(t)] = 1 + \frac{t}{n}(r - 1), \quad 0 \leq t \leq n, \tag{2.12}$$

it is clear that for any $\{x_i\} \subset [1,r]$, with $\mu(x) = \mu$, the associated polarized distribution is defined in (2.11) with

$$t = \frac{n(\mu - 1)}{r - 1}. \tag{2.13}$$

Theorem 1

Let $\{x_i\}_{i=1}^n \subset [1,r]$, with $\mu(x) = \mu$. Then

$$\mu'_2(x) \leq 1 + (r + 1)(\mu - 1), \tag{2.14}$$

$$R_2(x) \leq \frac{(r + 1)^2}{4r}. \tag{2.15}$$

Further, the inequalities in (2.14) and (2.15) are sharp.

Proof

Assume that (2.14) has been established. Then

$$R_2(x) \leq \frac{1 + (r + 1)(\mu - 1)}{\mu^2}. \tag{2.16}$$

As a function of μ in $[1,r]$, the right-hand side of (2.16) is maximized when $\mu = 2r/(r+1)$. Consequently, (2.15) follows by substitution.

In order to establish (2.14), it is sufficient to show that this inequality is satisfied for all polarized distributions $D(t)$. By using (2.11),

$$\mu'_2[D(t)] = \frac{n - [\![t]\!] - 1 + [\![t]\!]r^2 + [(t - [\![t]\!])(r - 1) + 1]^2}{n}, \quad 0 \leq t \leq n. \tag{2.17}$$

Let

$$t = m + s, \quad m = 0, 1, \ldots, n - 1; \quad 0 \leq s < 1. \tag{2.18}$$

Then $[\![t]\!] = m$ and (2.17) becomes

$$\mu'_2[D(t)] = \frac{n - m - 1 + mr^2 + [s(r - 1) + 1]^2}{n},$$
$$m = 0, \ldots, n - 1; 0 \leq s \leq 1. \tag{2.19}$$

The inequality for s was extended to $s=1$, because it is straightforward to verify that the right-hand side of (2.19) achieves the same value for $m=m'$ and $s=0$ as it does for $m=m'-1$ and $s=1$. For a given m, the right-hand side of (2.19) is a quadratic function of s with positive second derivative. Consequently, it is maximized over [0,1] when $s=0$ or 1. Hence, it is sufficient to consider (2.19) only for integral $m=0, 1, \ldots, n$ and $s=0$. For the resulting values of $t=m$,

$$\mu[D(t)] = \frac{n - m + mr}{n}, \quad m = 0, 1, \ldots, n \tag{2.20}$$

$$\mu'_2[D(t)] = \frac{n - m + mr^2}{n}, \quad m = 0, 1, \ldots, n \tag{2.21}$$

and a calculation shows that (2.14) is satisfied with equality at these points. Hence, it follows in general for $0<s<1$.

To see that the inequality in (2.14) is sharp, it is necessary to provide examples of distributions $\{x_i\}$ from $[1,r]$, with a common given mean μ, such that the resultant values of $\mu'_2(x)$ can be chosen arbitrarily close to the upper bound $1+(r+1)(\mu-1)$. Let $r>1$ and μ, $1 \leq \mu \leq r$, be given and define $\rho=(\mu-1)/(r-1)$. Since the inequality in (2.14) clearly provides sharp results when $\mu=1$ or $\mu=r$, only $1<\mu<r$ is considered, and hence $\rho>0$. Let c_j be a sequence of positive rational numbers, $c_j=m_j/n_j$, converging to ρ, so that with $\lambda_j=(c_j-\rho)/c_j$,

$$0 \leq \lambda_j \leq 1, \tag{2.22}$$

$$\lambda_j \to 0 \qquad \text{as } j \to \infty. \tag{2.23}$$

This can be accomplished since $\rho>0$. Now, consider the distribution $\{y_i\}$, $i=1, 2, \ldots, n_j$, defined by:

$$y_i = \begin{cases} 1 & 1 \leq i \leq n_j - m_j, \\ r - \lambda_j(r - 1) & n_j - m_j + 1 \leq i \leq n_j. \end{cases} \tag{2.24}$$

Note that $y_i \in [1,r]$ for all i due to (2.22). Also, a calculation shows that $\mu(y)=\mu$ and

$$\mu'_2(y) = 1 + c_j(r^2 - 1) + c_j\, \lambda_j[\lambda_j(r - 1)^2 - 2r\,(r - 1)]. \quad (2.25)$$

However, since $\lambda_j \to 0$ and $c_j \to \rho$, the right-hand side of (2.25) can be chosen arbitrarily close to its limit value of $1+\rho(r^2-1)$, which equals $1+(\mu-1)(r+1)$. Hence, the inequality in (2.14) is sharp. Letting $\mu=2r/(r+1)$, this example also shows the inequality in (2.15) to be sharp. □

Corollary

Let $\{x_i\}_{i=1}^n \subset [1,r]$, with $\mu(x)=\mu$. Then

$$\sigma^2(x) \le (r - \mu)\,(\mu - 1), \quad (2.26)$$

$$\frac{\sigma^2(x)}{\mu^2(x)} \le \frac{(r - 1)^2}{4r}. \quad (2.27)$$

Further, the inequalities in (2.26) and (2.27) are sharp.

Proof

Since $\sigma^2(x)=\mu'_2(x)-\mu^2$, the results follow directly from Theorem 1. □

Of course, the inequalities in Theorem 1 and the above Corollary can be modified to apply to the interval $[a,b]$ by use of (2.3) through (2.5), and

$$\sigma^2(\lambda x) = \lambda^2\sigma^2(x). \quad (2.28)$$

That is, if $\{x_i\} \subset [a,b]$, choose $\lambda=1/a$. Then, $\{\lambda x_i\} \subset [1,r]$ with $r=b/a$.

From (2.26), since this inequality is sharp, distributions of maximal variance must have a mean equal to $(r+1)/2$, the midpoint of $[1,r]$. Correspondingly, due to (2.13), the associated polarized distribution is given by $t=n/2$. Also, the distributions with maximal ratio of variance to mean squared must have a mean given by:

$$\mu = \frac{2r}{r + 1}. \quad (2.29)$$

Consequently, the associated polarized distribution $D(t)$ is given by:

$$t = \frac{n}{r + 1}. \quad (2.30)$$

That is, the proportion of points at the left endpoint 1, $f(1)$, satisfies

$$f(1) = \frac{r}{r+1} - \frac{\epsilon_t}{n}, \quad 0 < \epsilon_t \leq 1. \tag{2.31}$$

In other words, such distributions are always skewed to the left, with the tendency toward skewness increasing as $r = b/a$ increases,

$$f(1) \to 1, \text{ as } r \to \infty, \tag{2.32}$$

since for r large, ϵ_t is equal to t, which converges to zero as r increases. This is also evidenced by noting that due to (2.29), μ is an increasing function of r with upper bound equal to 2.

A lower bound for $\mu'_2(x)$ is fairly easy to develop by utilizing the well-known Cauchy-Schwarz inequality [2], which states that for given a_i, b_i, $i = 1, \ldots, n$,

$$\sum \left| a_i b_i \right| \leq \left(\sum a_i^2 \right)^{1/2} \left(\sum b_i^2 \right)^{1/2}, \tag{2.33}$$

with equality if and only if there are real numbers α, β, so that

$$\alpha a_i + \beta b_i = 0, \quad i = 1, \ldots, n. \tag{2.34}$$

Letting $a_i = x_i$, $b_i = 1$ yields

$$\sum x_i \leq \left(\sum x_i^2 \right)^{1/2} n^{1/2}, \tag{2.35}$$

or

$$\mu^2 \leq \mu'_2(x), \tag{2.36}$$

with equality if and only if all x_i are equal, due to (2.34). Consequently, μ^2 is a sharp lower bound for $\mu'_2(x)$. Hence, 1 is a sharp lower bound for $R_2(x)$. Finally, although it also follows from above, it is quite obvious by definition that 0 is a sharp lower bound for $\sigma^2(x)$.

Summarizing the above sharp estimates, we have:

$$\mu^2 \leq \mu'_2(x) \leq 1 + (r+1)(\mu - 1), \tag{2.37}$$

$$1 \leq R_2(x) \leq \frac{(r+1)^2}{4r}, \tag{2.38}$$

$$0 \leq \sigma^2(x) \leq (r - \mu)(\mu - 1), \tag{2.39}$$

$$0 \leq \frac{\sigma^2(x)}{\mu^2} \leq \frac{(r-1)^2}{4r}, \tag{2.40}$$

where $\{x_i\} \subset [1,r]$ with $\mu(x) = \mu$.

III. APPLICATIONS

1. *Interval Estimation of the Variance of Expected Claims*

Let A_{ij} equal the exposure amount of the i-th policy in a given class C_j, homogeneous with respect to the claim probability distribution function. Let X_j be the binomial random variable defined on C_j, so that

$$\text{Prob (claim on any policy in } C_j) = \text{Prob } (X_j = 1) = q_j. \tag{3.1}$$

Then S, defined by

$$S = \sum A_{ij} X_j, \tag{3.2}$$

is the random variable that represents the aggregate amount of claims in the time interval during which (3.1) is valid. In the terminology of risk theory [1], this is the individual risk model for aggregate claims. In general, the A_{ij} are also random variables that, for a given class C_j, are assumed to be independent and identically distributed with mean μ_j, and variance σ_j^2. Assuming the X_j to be independent, we have from [1] that

$$\mu(S) = \sum n_j \mu_j q_j, \tag{3.3}$$

$$\sigma^2(S) = \sum n_j \mu_j^2 q_j (1 - q_j) + \sum n_j \sigma_j^2 q_j, \tag{3.4}$$

where n_j is the number of policies in class C_j. Now if $A_{ij} = a_{ij}$ is known and fixed in advance, as is common for life insurance, (3.2) is simply a linear combination of independent binomial variables, and one has:

$$\mu(S) = \sum a_{ij} q_j, \tag{3.5}$$

$$\sigma^2(S) = \sum a_{ij}^2 q_j (1 - q_j). \tag{3.6}$$

In this fixed-claim-amount setting, if in-force policy data are already banded and summarized into intervals, $I_k = [a_k, a_{k+1}]$, $k \geq 1$ where $a_1 > 0$, let

$$r_k = \frac{a_{k+1}}{a_k}, \quad k \geq 1,$$

$$C_{jk} = \text{class of policies in } C_j \text{ with } a_{ij} \in I_k,$$

$$n_{jk} = \text{number of policies in } C_{jk},$$

$$\mu_{jk} = \text{average policy amount in } C_{jk}.$$

Then for each class C_j, the following is true by (2.37):

$$\sum_k n_{jk}\, \mu_{jk}^2 \le \sum_i a_{ij}^2 \le \sum_k n_{jk}\, [a_k^2 + a_k\,(r_k + 1)\,(\mu_{jk} - a_k)]. \tag{3.7}$$

Consequently, by combining (3.6) and (3.7),

$$\sum n_{jk}\, q_j\,(1 - q_j)\, \mu_{jk}^2 \le \sigma^2(S) \le \sum n_{jk}\, q_j\,(1 - q_j)\,[a_k^2 + a_k\,(r_k + 1)\,(\mu_{jk} - a_k)]. \tag{3.8}$$

With policy data already banded, (3.8) provides the resultant interval estimate for $\sigma^2(S)$, which may or may not be acceptable. In addition, the amount of relative error in the point estimate discussed in the next section is also fixed.

However, when it is possible to choose band size, as it is when experience systems are rewritten, the amount of error in the resultant estimates can be controlled. To see this, let $[a,b]$ be an amount interval, $a>0$, such that $a_{ij} \in [a,b]$ for all i, j. For a given value of $r>1$, let N be the solution of:

$$r^x = b/a, \qquad N = [\![x]\!] + 1. \tag{3.9}$$

Define the bands J_k by

$$J_k = [ar^k, ar^{k+1}], \qquad k = 0, 1, \ldots, N - 1. \tag{3.10}$$

Also, let C_{jk}, n_{jk} and μ_{jk} be defined as above only with reference to J_k rather than I_k. Then for each class C_j, the following holds due to (2.38):

$$\sum_k n_{jk}\, \mu_{jk}^2 \le \sum_i a_{ij}^2 \le \frac{(r + 1)^2}{4r} \sum_k n_{jk}\, \mu_{jk}^2. \tag{3.11}$$

Consequently,

$$\sum n_{jk}\, q_j(1 - q_j)\, \mu_{jk}^2 \le \sigma^2(S) \le \frac{(r + 1)^2}{4r} \sum n_{jk}\, q_j(1 - q_j)\, \mu_{jk}^2. \tag{3.12}$$

The problem of choosing r is discussed in detail in the next section. However, from (3.12), any degree of accuracy can be achieved by choosing r close enough to 1.

For the general claim amounts case, assume that the A_{ij} are independent random variables that are identically distributed for each j with mean μ_j and variance σ_j^2. If these distributions are assumed given by mathematical formula, μ_j and σ_j^2 will often be straightforward to compute and $\sigma^2(S)$ will be given directly by (3.4). On the other hand, these distributions may be defined empirically with reference to actual claim data.

If this is the case, Theorem 1 can be utilized in the following way. Let $[a,b]$, $a>0$ be an amount interval that contains all claims, which are assumed to be banded into amount intervals I_k as defined above with associated r_k. Also, let C_{jk} denote the collection of *claims* from class C_j with claim amount in the interval I_k, m_{jk} the number of such claims, μ_{jk} their mean, and σ_{jk}^2 their variance. Since σ_j^2 in (3.4) is defined as the variance of all claims from class C_j and σ_{jk}^2 the variance of such claims conditioned on their being within the amount interval I_k, they are related by a general formula involving conditional expectations [1]. Specifically, letting A_j denote the claims from class C_j,

$$\begin{aligned}\sigma_j^2 &= \text{Var}\,(A_j)\\ &= \text{Var}\,[E\,(A_j|I_k)] + E\,[\text{Var}\,(A_j|I_k)]\\ &= \text{Var}\,[\mu_{jk}] + E\,[\sigma_{jk}^2].\end{aligned}$$

Hence,

$$\sigma_j^2 = \frac{1}{m_j}\sum_k m_{jk}\,(\mu_{jk} - \mu_j)^2 + \frac{1}{m_j}\sum_k m_{jk}\,\sigma_{jk}^2, \tag{3.13}$$

where

$$m_j = \sum_k m_{jk}$$

and μ_j is defined as in (3.4) as the mean claim from class C_j, that is, $\mu_j = E[\mu_{jk}]$. Applying (2.39) to σ_{jk}^2:

$$\begin{aligned}&\frac{1}{m_j}\sum_k m_{jk}\,(\mu_{jk} - \mu_j)^2 \le \sigma_j^2\\ &\qquad\le \frac{1}{m_j}\sum_k m_{jk}\,[(\mu_{jk} - \mu_j)^2 + (r_k\,a_k - \mu_{jk})\,(\mu_{jk} - a_k)],\end{aligned} \tag{3.14}$$

which can be used in (3.4) to provide an estimate of $\sigma^2(S)$ given the current banding classes I_k. As was the case in (3.8), the resulting interval estimate may not be acceptable if the amount bands are too large. However, if it is possible to choose a band size, such as when experience-monitoring systems are revised, the error in the resultant estimates can be controlled.

By applying (2.40) in (3.13) instead of (2.39), as was done above, the interval in (3.14) has length that is a function of $(r_k - 1)^2$. That is, the estimates can be made arbitrarily good by choosing r_k close to 1. Unfortunately, this observation is of little practical value in choosing an amount band size, since in this context, the m_{jk} and μ_{jk} values above are random variables that depend on r_k.

To circumvent this difficulty, we utilize the collective risk model [1] for S. Restricting our attention to a specific risk class C_j, let S represent aggregate claims for this class using a compound Poisson model:

$$S = \sum_{i=1}^{N} X_i = \sum_{l=1}^{n} x_l N_l \tag{3.15}$$

where the number of claims N is assumed to have a Poisson distribution with parameter λ, and the X_i's are independent, identically distributed random variables that are also independent of N. The second representation for S decomposes the original sum into distinct claim sizes of amount x_l, which without loss of generality can be assumed to be finite in number (perhaps quite large), and N_l is the random number of such claims. As it turns out (see [1]), these N_l are independent Poisson random variables with parameter $\lambda_l = \lambda f(x_l)$.

Assume next that amount bands J_k are defined as in (3.10) and let N_k represent the number of claims in J_k. Clearly, N_k equals a sum of N_l values above corresponding to whether x_l is in J_k or not. Also, because the N_l are independent Poisson with parameters given above, the N_k are also independent Poisson with parameters

$$\lambda_k = \lambda\,[F(ar^{k+1}) - F(ar^k)].$$

Hence, S can be decomposed as a sum of independent compound Poisson random variables, S_k, representing the aggregate of all claims in J_k. Reintroducing the subscript j to represent risk class C_j, let μ_{jk} represent the expected value of claims from J_k and σ_{jk}^2 the variance of such claims. Also, let λ_{jk} be the Poisson parameter for N_{jk}. We then have:

$$\begin{aligned} \sigma^2(S) &= \sum \sigma^2(S_{jk}) \\ &= \sum \lambda_{jk}\,(\mu_{jk}^2 + \sigma_{jk}^2). \end{aligned} \tag{3.16}$$

Using (2.40) in (3.16) for σ_{jk}^2, we obtain the following analog to (3.12):

$$\sum \lambda_{jk}\, \mu_{jk}^2 \le \sigma^2(S) \le \frac{(r+1)^2}{4r} \sum \lambda_{jk}\, \mu_{jk}^2 \tag{3.17}$$

The problem of choosing r is discussed next.

2. *Point Estimation of the Standard Deviation of Expected Claims*

Although (3.8), (3.12), (3.14) and (3.17) provide interval estimates for $\sigma^2(S)$, the following simple Lemma provides a point estimate with minimal relative error.

Lemma 1

Let a, c be given positive real numbers and x an unknown real number satisfying:

$$a \le x \le (1 + c)a. \tag{3.18}$$

Let $\hat{x}(\lambda)$ denote the point estimate for x defined by:

$$\hat{x}(\lambda) = (1 + \lambda c)a, \qquad 0 \le \lambda \le 1. \tag{3.19}$$

Then the absolute value of the error of the estimate $\hat{x}(\lambda)$ relative to itself is minimized when $\lambda = 1/2$.

Proof

Utilizing (3.18), (3.19), and the fact that $\hat{x}(\lambda) > 0$, we get

$$\frac{-\lambda c}{1 + \lambda c} \le \frac{x - \hat{x}}{\hat{x}} \le \frac{c(1 - \lambda)}{1 + \lambda c}. \tag{3.20}$$

That is,

$$\left|\frac{x - \hat{x}}{\hat{x}}\right| \le \max\left[\frac{(1 - \lambda)c}{1 + \lambda c}, \frac{\lambda c}{1 + \lambda c}\right]. \tag{3.21}$$

A straightforward analysis yields that $f(\lambda) = c(1-\lambda)/(1+\lambda c)$ is a positive decreasing function over [0,1], which agrees with $g(\lambda) = \lambda c/(1+\lambda c)$, a positive increasing function, at $\lambda = 1/2$. Consequently, the relative error defined in (3.21) is minimized when $\lambda = 1/2$. □

To apply Lemma 1 in the context of (3.12) or (3.17), assume $d > 0$ is given and let r be chosen to satisfy:

$$\frac{(r+1)^2}{4r} = 1 + d. \tag{3.22}$$

Since $h(r) = (r+1)^2/4r$ is an increasing function of r for $r \geq 1$ and $h(1) = 1$, it is clear that the solution of (3.22) must exist, be unique and be greater than 1. Rewriting (3.12) [an identical development holds for (3.17)]:

$$\left(\sum n_{jk}\, q_j\,(1 - q_j)\,\mu_{jk}^2\right)^{1/2} \leq \sigma(S)$$
$$\leq (1 + d)^{1/2} \left(\sum n_{jk}\, q_j\,(1 - q_j)\,\mu_{jk}^2\right)^{1/2}. \tag{3.23}$$

Let $c = (1+d)^{1/2} - 1$ and define

$$\hat{\sigma}(S) = \left(1 + \frac{c}{2}\right)\left[\sum n_{jk}\, q_j\,(1 - q_j)\,\mu_{jk}^2\right]^{1/2}. \tag{3.24}$$

Then, by using (3.21),

$$|\sigma(S) - \sigma(S)| \leq \frac{c}{2 + c}\,\hat{\sigma}(S). \tag{3.25}$$

Hence, in order for the point estimate in (3.24) to have a maximum error relative to itself of $100\epsilon\%$, $0 < \epsilon < 1$, that is,

$$(1 - \epsilon)\,\hat{\sigma}(S) \leq \sigma(S) \leq (1 + \epsilon)\,\hat{\sigma}(S), \tag{3.26}$$

the values c, d, and r must be chosen to satisfy:

$$c = \frac{2\epsilon}{1 - \epsilon} \tag{3.27}$$

$$d = (c + 1)^2 - 1 \tag{3.28}$$

$$r = 1 + 2\,(d + \sqrt{d^2 + d}) \tag{3.29}$$

The value of r in (3.29) is the larger solution of (3.22) expressed as a quadratic equation in r, where this root is given by the quadratic formula. Table 1 provides some numerical results:

TABLE 1

ϵ	0.5	0.1	0.05	0.01	0.005
c	2.0	0.22	0.11	0.02	0.01
d	8.0	0.49	0.22	0.04	0.02
r	33.97	3.70	2.48	1.49	1.33

For example, to achieve a relative error of 5 percent in the point estimate of $\sigma(S)$, the value $r=2.48$ will suffice. Once r is chosen, the number of amount bands N needed to analyze experience in the interval $[a,b]$, $a>0$, is given in (3.9). For example, if $a=10{,}000$ and $b=100{,}000{,}000$, the solution x of (3.9) is approximately 10.14. Consequently, $N=11$ here, as noted in Section II. Limiting the relative error to 1 percent would require 23 bands.

Because such good accuracy can be achieved with only 11 bands, this approach is practical to implement both for industry-wide applications, such as the *TSA Reports of Mortality, Morbidity and Other Experience*, and for individual company experience studies.

As noted above, an identical development holds for the compound Poisson model in (3.17), and Table 1 applies in this context as well. As for the interval estimates of (3.8) and (3.14), where amount bands are already given, again the midpoint estimators are used for minimal relative error. In these cases, the relative error is given in (3.25), where $1+c$ equals the square root of the ratio of the upper to lower bounds of the respective intervals.

3. *Modified Confidence Limits*

Once $\sigma(S)$ is estimated to the required degree of accuracy by $\hat{\sigma}(S)$, it can be used in conjunction with the central limit theorem [4] to produce confidence intervals for S, the aggregate claims during a given period of time.

For example, let Z_α correspond to the positive boundary value of the symmetric $100(1-\alpha)\%$ confidence interval for a normally distributed random variable Z. That is,

$$\text{Prob}\ [|Z| \le Z_\alpha \mid Z \sim N(0,\ 1)] = 1 - \alpha. \tag{3.30}$$

Then according to the central limit theorem, the $100(1-\alpha)\%$ confidence interval for S is approximately:

$$\mu(S) - Z_\alpha\, \sigma(S) \le S \le \mu(S) + Z_\alpha\, \sigma(S). \tag{3.31}$$

If $\hat{\sigma}(S)$ as calculated above has a maximum relative error of $100\epsilon\%$, in the sense of (3.26), the corresponding modified confidence interval for S becomes:

$$\mu(S) - (1 + \epsilon) Z_\alpha \hat{\sigma}(S) \leq S \leq \mu(S) + (1 + \epsilon) Z_\alpha \hat{\sigma}(S). \quad (3.32)$$

4. *Retention Limits*

The appropriate level for retention limits for reinsurance purposes can also be studied with this approach. To see this, let $R>0$ be given and define S^R by:

$$S^R = \sum \min(A_{ij}, R) X_j, \quad (3.33)$$

where X_j and A_{ij} are as defined in Section III.1. Then S^R is the random variable that represents the aggregate amount of claims payable if the retention limit is set at R. If $\mu(S^R)$ is considered an appropriate value for the minimum reserve needed against the contingency insured, an appropriate "surplus" level can be determined by considering the random variable, M^R, defined by:

$$M^R = S^R/\mu(S^R). \quad (3.34)$$

By the central limit theorem, the $100(1-\alpha)\%$ confidence interval for M^R is approximately given by

$$|M^R - 1| \leq Z_\alpha \frac{\sigma(S^R)}{\mu(S^R)}, \quad (3.35)$$

which can be modified as in (3.32) to

$$|M^R - 1| \leq (1 + \epsilon) Z_\alpha \frac{\hat{\sigma}(S^R)}{\mu(S^R)}. \quad (3.36)$$

For the level of confidence desired, α, existing surplus would limit the value of $\hat{\sigma}(S^R)/\mu(S^R)$ that is acceptable. In general, this ratio would be expected to decrease as the value of R decreases. Of course, once $\hat{\sigma}(S)$ has been estimated as in Section III.2, the value of the ratio $\hat{\sigma}(S^R)/\mu(S^R)$ can be readily determined for R equal to any of the amount band boundary points.

For example, assume that $\hat{\sigma}(S)$ has been determined as in (3.24). Then for $R=ar^k$,

$$\frac{\hat{\sigma}(S^R)}{\mu(S^R)} =$$

$$\left(1 + \frac{c}{2}\right) \frac{\left\{\sum_j q_j (1 - q_j) \left[\sum_{l<k} n_{jl}\, \mu_{jl}^2 + R^2 \sum_{l\geq k} n_{jl}\right]\right\}^{1/2}}{\sum_j q_j \left[\sum_{l<k} n_{jl}\, \mu_{jl} + R \sum_{l\geq k} n_{jl}\right]}, \quad (3.37)$$

and these values are straightforward to calculate because the various parameters are assumed known. For $ar^k < R < ar^{k+1}$, a similar formula would be obtained except that $R^2 n_{jk}$ and Rn_{jk} would be replaced by $(\mu_{jk}^R)^2 n_{jk}$ and $\mu_{jk}^R n_{jk}$, respectively, where μ_{jk}^R is defined analogously to μ_{jk} but with all amounts limited to R. For such intermediate values of R, the ratio in (3.37) could be estimated by utilizing an approximation for μ_{jk}^R such as:

$$\mu_{jk}^R = \mu_{jk} - \left(\frac{ar^{k+1} - R}{ar^{k+1} - ar^k}\right)^{\upsilon} (\mu_{jk} - ar^k), \quad R \in J_k, \quad (3.38)$$

where $\upsilon > 0$ is chosen to reflect the magnitude and direction of skewness present in the distribution of amounts in J_k. For example, it is straightforward to check that $\upsilon = 2$ when this distribution is uniform. In general, $\upsilon > 2$ reflects skewness to the left, $\upsilon < 2$ skewness to the right.

5. *Variance of Decrement Estimators*

As a final application, consider the formulas presented in [8] for the moments of $\hat{q}$ defined by (notation changed):

$$\hat{q} = \frac{\sum A_i X_i}{\sum A_i}, \quad (3.39)$$

where A_i is the number of exposure units for individual i, $i = 1, \ldots, n$, and $X_i = X_i(A)$ is a binomial random variable such that:

$$\text{Prob}\,(X_i = 1 \mid A_i = a) = q(a). \quad (3.40)$$

If it is assumed that A_i and X_j are mutually independent for all i and j, and that $q(a) = q$, the variance of $\hat{q}$ as derived in [8] is:

$$\text{Var}\,(\hat{q}) = q(1 - q)\, E\left[\sum A_i^2 / \left(\sum A_i\right)^2\right]. \quad (3.41)$$

Of course, when $A_i = a_i$ is fixed and known in advance, the presence of the expectation E in (3.41) is only notational, and this formula is clearly equivalent to (3.6) restricted to one homogeneous class C_j. In addition, although it is less apparent, if the method utilized to produce (3.41) is applied to S defined in (3.2), the general formula (3.4) is produced.

Let $[a,b]$, $a>0$ be given where this interval contains the range of A_i. For $r>1$, let J_k be a partition of $[a,b]$ as defined in (3.10) and define $\hat{q}_k$ as the restriction of q in (3.39) to J_k. That is,

$$\hat{q}_k = \frac{\sum A_i X_i}{\sum A_i}, \quad A_i \in J_k. \tag{3.42}$$

Since $q(a)=q$, it is clear that $E(\hat{q}_k)=E(\hat{q})=q$, and

$$\text{Var}\,(\hat{q}_k) = q\,(1-q)\,E\left[\sum A_i^2/\left(\sum A_i\right)^2 \,\middle|\, A_i \in J_k\right]. \tag{3.43}$$

To estimate (3.43), let N_k be defined as the random variable representing the number of claims of amount $A_i \in J_k$. Note that N_k is a random variable even when $A_i=a_i$ for all i. Applying (2.38), we have that for $n_k \geq 1$:

$$\frac{1}{n_k} \leq E\left[\sum A_i^2/\left(\sum A_i\right)^2 \,\middle|\, A_i \in J_k, N_k = n_k\right] \leq \frac{(r+1)^2}{4r}\frac{1}{n_k}. \tag{3.44}$$

If expectations are taken in (3.44) with respect to N_k, the following estimate for $\text{Var}(\hat{q}_k)$ is produced:

$$q\,(1-q)\,E\left[\frac{1}{N_k}\right] \leq \text{Var}(\hat{q}_k) \leq \frac{(r+1)^2}{4r}\,q\,(1-q)\,E\left[\frac{1}{N_k}\right]. \tag{3.45}$$

For a given value of r, $E[1/N_k]$ will usually increase as k increases. In particular, both bounds in (3.45) as well as the size of the bounded interval will tend to increase as k increases, implying a general increase in experience volatility as policy amounts increase. As noted before, the estimates in (3.45) are sharp and can usually be utilized to estimate $\text{Var}(\hat{q}_k)$ to any given degree of accuracy by choosing r close enough to 1. Also, $\text{Var}(\hat{q}_k)$ can be approximated by $\hat{\sigma}^2(\hat{q}_k)$, using Lemma 1, where

$$\hat{\sigma}^2(\hat{q}_k) = \left(1+\frac{c}{2}\right)q\,(1-q)\,E\left[\frac{1}{N_k}\right], \quad c = \frac{(r-1)^2}{4r}, \tag{3.46}$$

and the relative error of this approximation is no greater than $\epsilon = c/(2+c)$.

Although (3.45) can also be utilized to estimate $\mathrm{Var}(\hat{q})$ with $r = b/a$ and $N = \Sigma N_k$, the result may be considered quite crude for realistic values of b/a. For example, if $a = 10{,}000$ and $b = 100{,}000{,}000$, (3.45) becomes

$$q\,(1 - q)\,E\left[\frac{1}{N}\right] \leq \mathrm{Var}\,(\hat{q}) \leq 2501\,q\,(1 - q)\,E\left[\frac{1}{N}\right], \quad (3.47)$$

and the resultant value of $\hat{\sigma}^2(\hat{q})$ will have a maximum relative error of almost 100 percent. However, for n large, the absolute error may be quite small and the estimates may have practical value.

If it is assumed that $q(A_i)$ is not constant in general, but is constant over each J_k where $q(A_i) = q_k$, (3.45) and (3.46) can still be utilized but with $q = q_k$.

6. *Practical Considerations*

Throughout this section, q_j has denoted the probability of a claim in class C_j where this probability was defined on an individual policyholder basis. Also, class C_j was assumed homogeneous with respect to the value of this probability and, consequently, would typically be defined in terms of individual risk characteristics and the various underwriting parameters of the insurance product under study. To simplify calculations, however, it is often desirable to combine various risk classes. For example, ages may be quinquennialized or "rated" classes grouped. This is because the parameters q_j must be estimated based on actual experience from each class, and the experience of many classes is too sparse to analyze confidently.

Given some restrictions, the effect of such groupings on the mean and variance of S as given in (3.3) and (3.4) can be analyzed. To this end, let $\{C_j\}$ be a collection of classes to be grouped with respective claim probabilities $\{q_j\}$ and class sizes $\{n_j\}$, $n = \Sigma\, n_j$. As a combined class, the claim probability q is given by

$$q = \sum n_j\, q_j/n. \quad (3.48)$$

For notational convenience, let μ and σ^2 denote that part of the summations in (3.3) and (3.4) that corresponds only to the classes under consideration. Also let $\bar{\mu}$ and $\bar{\sigma}^2$ be analogously defined under the assumption that $\cup C_j$ is a homogeneous class with claim probability q as defined in (3.48):

$$\bar{\mu} = \sum n_j\, \mu_j\, q \quad (3.49)$$

$$\bar{\sigma}^2 = \sum n_j\, \mu_j^2\, q(1 - q) + \sum n_j\, \sigma_j^2\, q \quad (3.50)$$

Lemma 2

Let μ, $\bar{\mu}$, σ^2, $\bar{\sigma}^2$ be defined as above. Further, assume that for the risk classes $\{C_j\}$ combined in (3.48), that $\mu_j = \mu_0$ and $\sigma_j^2 = \sigma_0^2$. Then:

$$\bar{\mu} = \mu, \tag{3.51}$$

$$\bar{\sigma}^2 > \sigma^2. \tag{3.52}$$

Proof

Because $\mu_j = \mu_0$ and $\Sigma n_j\, q = \Sigma n_j\, q_j$, (3.51) follows immediately. For (3.52), first note that by the same line of reasoning,

$$\sum n_j\, \sigma_j^2\, q = \sum n_j\, \sigma_j^2\, q_j. \tag{3.53}$$

Hence,

$$\begin{aligned} \bar{\sigma}^2 - \sigma^2 &= \mu_0^2 \sum n_j\, [q\ (1 - q) - q_j\, (1 - q_j)] \\ &= \mu_0^2 \sum n_j\, q_j\, (q_j - q). \end{aligned} \tag{3.54}$$

By (3.48),

$$q_j - q = \frac{\sum n_l\, (q_j - q_l)}{n}, \tag{3.55}$$

which, when substituted into (3.54), yields:

$$\bar{\sigma}^2 - \sigma^2 = \mu_0^2 \sum_{j,l} \frac{n_j\, n_l}{n}\, q_j\, (q_j - q_l). \tag{3.56}$$

If m represents the number of classes to be combined, it is clear that the summation in (3.56) has $m(m-1)$ terms, because only those terms with $j \neq l$ will be nonzero. By symmetry, these terms can be paired off, yielding

$$\begin{aligned} \bar{\sigma}^2 - \sigma^2 &= \mu_0^2 \sum_{j<l} [n_j\, n_l\, q_j\, (q_j - q_l) + n_j\, n_l\, q_l\, (q_l - q_j)]\, \frac{1}{n} \\ &= \mu_0^2 \sum_{j<l} \frac{n_j\, n_l}{n}\, (q_j - q_l)^2, \end{aligned} \tag{3.57}$$

completing the proof. □

IV. HIGHER MOMENTS—DISCRETE CASE

Let $\{x_i\}_{i=1}^n$ be a collection of numbers from the interval $[a,b]$, $a>0$. Let k be a real number, $k\geq 1$, and define $\mu'_k(x)$ and $R_k(x)$ analogously to the case $k=2$ by:

$$\mu'_k(x) = \frac{1}{n}\sum x_i^k, \tag{4.1}$$

$$R_k(x) = \mu'_k/\mu^k. \tag{4.2}$$

As in Section II, $\mu'_k(x)$ and $R_k(x)$ need only be estimated over the interval $[1,r]$, because

$$\mu'_k(\lambda x) = \lambda^k \mu'_k(x), \tag{4.3}$$

$$R_k(\lambda x) = R_k(x). \tag{4.4}$$

Also, the value of these functions need only be considered on polarized distributions, because if $\{x_i\}$ and $\{y_i\}$ are given as in (2.9),

$$\mu'_k(y) = \mu'_k(x) + \frac{1}{n}\{[(x_2 + \delta)^k - x_2^k] - [(x_1 + \delta)^k - x_1^k]\}, \tag{4.5}$$

which exceeds $\mu'_k(x)$ because $x_2>x_1$ and for $\delta>0$ and $k>1$, $(x+\delta)^k-x^k$ is an increasing function of x.

Theorem 2

Let $\{x_i\}_{i=1}^n \subset [1,r]$ with $\mu(x)=\mu$, and k a real number satisfying $k\geq 1$. Then,

$$\mu'_k(x) \leq 1 + \frac{r^k - 1}{r - 1}(\mu - 1), \tag{4.6}$$

$$R_k(x) \leq \frac{(k - 1)^{(k-1)}(r^k - 1)^k}{k^k (r - 1)(r^k - r)^{(k-1)}}. \tag{4.7}$$

Further, the inequalities in (4.6) and (4.7) are sharp.

Proof

Assuming (4.6), it is clear that

$$R_k(x) \leq \frac{1}{\mu^k}\left[1 + \left(\frac{r^k - 1}{r - 1}\right)(\mu - 1)\right]. \tag{4.8}$$

As a function of μ on $[1,r]$, the right-hand side of (4.8) is maximized when

$$\mu = \frac{k\,(r^k - r)}{(k-1)\,(r^k - 1)}, \tag{4.9}$$

and (4.7) follows by substitution.

To establish (4.6), let $D(t)$ be defined as in (2.11) and t parametrized as in (2.18). Then,

$$\mu'_k[D(t)] = \frac{n - m - 1 + m\,r^k + [s\,(r-1) + 1]^k}{n},$$

$$m = 0, \ldots, n-1;\ 0 \le s \le 1. \tag{4.10}$$

For each m, the right-hand side of (4.10) is a polynomial in s with positive or identically zero second derivative. Consequently, it is maximized over $[0,1]$ when $s=0$ or 1. Hence, it is sufficient to consider (4.10) only for integral $m=0, \ldots, n$ and $s=0$. For such values,

$$\mu[D(t)] = \frac{n - m + m\,r}{n}, \quad m = 0, 1, \ldots, n \tag{4.11}$$

$$\mu'_k[D(t)] = \frac{n - m + m\,r^k}{n}, \quad m = 0, 1, \ldots, n \tag{4.12}$$

and a calculation shows that (4.6) is satisfied with equality at these points. Hence, it follows in general for $0<s<1$.

To see that the inequality in (4.6) is sharp, consider the example given in the proof of Theorem 1. Corresponding to (2.25),

$$\mu'_k(y) = 1 + c_j\,(r^k - 1) + c_j\,\lambda_j\,g(r, \lambda_j), \tag{4.13}$$

where $g(r, \lambda_j)$ is a series or polynomial of order k in r. As j increases, the right-hand side of (4.13) converges to $1+\rho(r^k-1)$, which equals the right-hand side of (4.6), because $\rho=(\mu-1)/(r-1)$. Consequently, the inequality in (4.6) is sharp. Letting μ be defined as in (4.9) shows the inequality in (4.7) to be sharp as well. □

From the above proof, the distribution that maximizes the ratio $R_k(x)$ must have a mean μ given in (4.9). As was the case in (2.29) where $k=2$, this mean is an increasing function of r with upper bound equal to $k/(k-1)$. In addition, the associated polarized distribution $D(t)$ is given by t defined in (2.13), which due to (4.9) equals,

$$t = \frac{n}{k-1}\left[\frac{1}{r-1} - \frac{k}{r^k - 1}\right]. \tag{4.14}$$

Consequently, the proportion of points at the left endpoint 1, $f(1)$, satisfies

$$f(1) = 1 - \frac{1}{k-1}\left[\frac{1}{r-1} - \frac{k}{r^k - 1}\right] - \frac{\epsilon_t}{n}, \quad 0 < \epsilon_t \leq 1. \tag{4.15}$$

Utilizing the fact that the arithmetic mean of any collection of numbers, in particular $\{1, r, \ldots, r^{k-1}\}$, must equal or exceed the geometric mean, it is possible to show that for integral k, t in (4.14) is a decreasing function of r (that is, negative first derivative) and, correspondingly, $f(1)$ is an increasing function of r satisfying

$$f(1) \to 1, \qquad r \to \infty, \qquad k \geq 1. \tag{4.16}$$

This statement holds for nonintegral k as well and can be proved by a more careful analysis of $t'(r)$. Also, for given $r > 1$, t converges to 0 as k increases, therefore

$$f(1) \to 1, \qquad k \to \infty, \qquad r \geq 1. \tag{4.17}$$

Lower bounds for $\mu'_k(x)$ can be developed by utilizing a generalization of (2.33) known as Hölder's inequality [10], [11], which states that for given a_i, b_i, $i = 1, \ldots, n$,

$$\sum |a_i b_i| \leq \left(\sum |a_i|^p\right)^{1/p} \left(\sum |b_i|^q\right)^{1/q}, \tag{4.18}$$

where p, q are real numbers, $1 \leq p, q \leq \infty$, satisfying:

$$1/p + 1/q = 1. \tag{4.19}$$

When $p = 1$, q is taken as equal to ∞, and the corresponding sum is defined as equal to its limiting value as $q \to \infty$,

$$\lim_{q \to \infty} \left(\sum |b_i|^q\right)^{1/q} = \max\{|b_i|\}. \tag{4.20}$$

In addition, (4.18) is satisfied with equality if and only if there are real numbers α, β so that

$$\alpha\, |a_i|^p + \beta\, |b_i|^q = 0, \qquad i = 1, \ldots, n. \tag{4.21}$$

Letting $a_i = x_i$, $b_i = 1$, $p = k$, and $q = k/(k-1)$, (4.18) yields

$$\mu^k \le \mu'_k(x), \tag{4.22}$$

with equality if and only if all x_i are equal due to (4.21).

Consequently, μ^k is a sharp lower bound for $\mu'_k(x)$, and 1 is a sharp lower bound for $R_k(x)$.

As currently stated, Theorem 2 is not applicable to all distributions of a discrete positive bounded rv. This is because it was assumed that the distribution could be realized as a finite collection of points in $[a, b]$, $a > 0$. If $f(x)$ is a probability density function defined on $\{y_j\}_{j=1}^{m} \subset [a, b]$ such that $f(y_j)$ is rational for all j, it can be so realized by defining

$$M = \min \{N \mid N, Nf(y_j) \text{ integral for all } j\},$$

$$n_j = Mf(y_i), j = 1, \ldots, m$$

$$x_i = \begin{cases} y_1 & 1 \le i \le n_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ y_m & n - n_m + 1 \le i \le n \end{cases} \tag{4.23}$$

$$n = \sum n_j.$$

Conversely, every finite collection of points from $[a, b]$, $a > 0$, can be identified with a probability density function $f(x)$ with rational range. However, because every density function can be approximated to any degree of accuracy with density functions of rational range, it should be expected that (4.6), (4.7), and (4.22) are valid in general.

To this end, let $f(x)$ be a probability density function of a discrete rv, $X \in [a, b]$, $a > 0$, and define,

$$\mu'_k = \sum x_i^k f(x_i), \quad k \ge 1, \tag{4.24}$$

$$\mu = \mu'_1,$$

$$R_k = \mu'_k/\mu^k. \tag{4.25}$$

As usual, only the interval $[1, r]$ need be considered.

Theorem 3

Let $f(x)$ be a pdf of a discrete rv defined on $[1,r]$. Then

$$\mu^k \leq \mu'_k \leq 1 + \left(\frac{r^k - 1}{r - 1}\right)(\mu - 1), \tag{4.26}$$

$$1 \leq R_k \leq \frac{(k-1)^{(k-1)}(r^k - 1)^k}{k^k (r-1)(r^k - r)^{(k-1)}}. \tag{4.27}$$

Further, all inequalities are sharp.

Proof

First, assume that $f(x)$ has a finite domain. That is, let $f(x)$ be given and defined on $\{x_i\}_{i=1}^m \subset [1,r]$. Given $\epsilon > 0$, define $g_\epsilon(x_i)$, $i = 1, \ldots, m$, so that $g_\epsilon(x_i)$ is rational and

$$|f(x_i) - g_\epsilon(x_i)| \leq \epsilon/(mr^k), \tag{4.28}$$

$$\sum g_\epsilon(x_i) = 1. \tag{4.29}$$

If $[f(x_i), \ldots, f(x_m)]$ is identified with a point $y \in \mathbf{R}^m$ on the hyperplane defined by $\Sigma\, y_i = 1$, it is clear that (4.28) and (4.29) require the existence of rational points on this hyperplane that are arbitrarily close to y. The existence of such points is a fundamental property of $\mathbf{R}^m$, that is, that rational points are dense in $\mathbf{R}^m$ [3].

Given $g_\epsilon(x)$, it is clear that

$$|\mu'_k(f) - \mu'_k(g_\epsilon)| \leq \epsilon. \tag{4.30}$$

However, the construction in (4.23) shows that Theorem 2 and (4.22) can be applied to $\mu'_k(g_\epsilon)$ and (4.26) is satisfied. Because ϵ can be arbitrarily chosen, (4.26) must also hold for $\mu'_k(f)$.

Now for arbitrary $f(x)$ defined on $\{x_i\}_{i=1}^\infty \subset [1,r]$, if $\mu'_k(f)$ is assumed to exist, it is clear that for every $\epsilon > 0$, k fixed, there is an integer N such that,

$$\sum_{i=N+1} x_i^k f(x_i) < \epsilon, \tag{4.31}$$

$$\sum_{i=N+1} f(x_i) < \epsilon. \tag{4.32}$$

Let $h_N(x)$ be a pdf defined on $\{x_1, \ldots, x_N\}$ so that

$$h_N(x_i) = \frac{f(x_i)}{\sum_1^N f(x_i)}, \quad i = 1, \ldots, N.$$

Applying (4.31) and (4.32),

$$|\mu'_k(f) - \mu'_k(h_N)| \le \frac{\epsilon}{1-\epsilon}[\mu'_k(f) + 1]. \tag{4.33}$$

Hence, because (4.26) is satisfied with $h_N(x)$, it must also hold for $f(x)$ due to (4.33) and the fact that

$$\lim_{N\to\infty} \mu(h_N) = \mu(f). \tag{4.34}$$

The inequalities in (4.27) follow from (4.26) as in Theorem 2. Finally, the inequalities are sharp due to the example in Theorem 2. □

Because (4.3) and (4.4) are valid in general, Theorem 3 can be applied to any pdf $f(x)$ of a discrete rv defined on $[a,b]$, $a>0$.

Corollary

Let $f(x)$ be a pdf of a discrete rv defined on $[1,r]$, and let $M_X(t)$ denote the moment-generating function of x,

$$M_X(t) = \sum_x e^{tx} f(x). \tag{4.35}$$

Then,

$$e^{\mu t} \le M_X(t) \le e^t + \frac{\mu - 1}{r - 1}(e^{rt} - e^t). \tag{4.36}$$

Further, the inequalities in (4.36) are sharp.

Proof

Rewriting (4.35) as

$$M_X(t) = \sum_k \frac{t^k \mu'_k}{k!}, \tag{4.37}$$

(4.36) follows directly from (4.26). For $1 \le \mu \le r$, if the point mass pdf $f_\mu(x)$ is considered, where $f_\mu(\mu) = 1$, $f_\mu(x) = 0$ otherwise, the inequality on the left in (4.36) is seen to be sharp. Also, for $1 \le \mu \le r$, let $g_\mu(x)$ be defined by

$$g_\mu(x) = \begin{cases} (r-\mu)/(r-1) & x = 1 \\ (\mu - 1)/(r-1) & x = r \\ 0 & \text{elsewhere.} \end{cases} \tag{4.38}$$

Clearly,

$$\mu(g_\mu) = \mu, \tag{4.39}$$

$$\mu'_k(g_\mu) = \frac{r - \mu}{r - 1} + \frac{\mu - 1}{r - 1} r^k, \quad k \geq 1,$$

and a calculation shows that

$$\mu'_k(g_\mu) = 1 + \left(\frac{r^k - 1}{r - 1}\right)(\mu - 1), \quad k \geq 1. \tag{4.40}$$

Consequently, the moment-generating function associated with $g_\mu(x)$ is given by the right-hand estimate in (4.36). □

V. HIGHER MOMENTS—CONTINUOUS CASE

Let $f(x)$ be a continuous pdf defined on $[1,r]$ and μ'_k, μ and R_k defined analogously to (4.24) and (4.25), with

$$\mu'_k = \int_1^r x^k f(x)\, dx, \quad k \geq 1. \tag{5.1}$$

Theorem 4

Let $f(x)$ be a continuous pdf defined on $[1, r]$. Then,

$$\mu^k \leq \mu'_k \leq 1 + \left(\frac{r^k - 1}{r - 1}\right)(\mu - 1), \tag{5.2}$$

$$1 \leq R_k \leq \frac{(k - 1)^{(k-1)} (r^k - 1)^k}{k^k (r - 1) (r^k - r)^{(k-1)}}. \tag{5.3}$$

Further, all inequalities are sharp.

Proof

For each n, consider the partition of $[1, r]$ given by:

$$x_i = 1 + i\Delta x, \quad \Delta x = \frac{r - 1}{n}, \quad i = 0, \ldots, n. \tag{5.4}$$

Consider $\sum_{i=0}^{n-1} f(x_i)\,\Delta x$. Because $f(x)$ is continuous and has integral equal to 1 over $[1,r]$, it is clear that

$$\sum_{i=0}^{n-1} f(x_i)\,\Delta x = \upsilon_n, \quad \upsilon_n \to 1 \text{ as } n \to \infty. \tag{5.5}$$

Similarly,

$$\int_1^r x^k f(x)\,dx = \lim_n \sum_{i=0}^{n-1} x_i^k f(x_i)\,\Delta x. \tag{5.6}$$

Let $g_n(x)$ be the pdf defined on the partition $\{x_i\}$ given in (5.4) by

$$g_n(x_i) = \frac{f(x_i)\,\Delta x}{\upsilon_n}. \tag{5.7}$$

Applying Theorem 3 to g_n,

$$\mu^k(g_n) \le \mu'_k(g_n) \le 1 + \left(\frac{r^k - 1}{r - 1}\right)[\mu(g_n) - 1]. \tag{5.8}$$

Taking limits in (5.8) as $n \to \infty$ proves (5.2), because $\mu'_k(g_n) \to \mu'_k(f)$ for all k. As usual, (5.3) follows from (5.2). Finally, the inequalities are sharp because the discrete example given in the proof of Theorem 2 can be approximated to any degree of accuracy by continuous pdf's. □

Corollary

Let $f(x)$ be given as in Theorem 4 and let $M_X(t)$ denote the moment-generating function of x,

$$M_X(t) = \int_1^r e^{tx} f(x)\,dx. \tag{5.9}$$

Then,

$$e^{\mu t} \le M_X(t) \le e^t + \frac{\mu - 1}{r - 1}(e^{rt} - e^t). \tag{5.10}$$

Further the inequalities in (5.10) are sharp.

Proof

The inequalities in (5.10) follow directly from (5.2) and (4.37). Also, the fact that they are sharp follows by considering continuous approximations to the example given in the proof of the Corollary to Theorem 3. □

It was noted in Section IV that the distribution with maximal ratio of μ'_k to μ^k will have μ given as in (4.9). It may be of interest to determine the mean of the distribution for which the interval developed for μ'_k is greatest. A calculation shows that

$$\mu = \left[\frac{r^k - 1}{k(r - 1)}\right]^{1/k}. \tag{5.11}$$

Clearly, μ is an unbounded increasing function of r for each k. Also, for fixed r, μ is an increasing function of k with limit equal to r. The value of this limit can be determined by applying L'Hospital's rule [2] to $\ln\mu$ as a function of k, $k \rightarrow \infty$.

ACKNOWLEDGMENT

An earlier version of this paper was presented at the 17th Actuarial Research Conference in October 1982 and published in [9]. The author gratefully acknowledges references [6] and [7] provided by Piet de Jong and Patrick Brockett at that conference as well as reference [5] provided by an anonymous reviewer.

BIBLIOGRAPHY

1. BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A., AND NESBITT, C.J. *Actuarial Mathematics.* Itasca, Ill.: Society of Actuaries, 1986, Chap. 2, 11.
2. COURANT, R., AND JOHN, F. *Introduction to Calculus and Analysis*, Vol. I. New York: Interscience Publishers, 1965, pp. 15, 464.
3. DUGUNDJI, J. *Topology.* Boston: Allyn and Bacon, Inc. 1970, p. 72.
4. FELLER, W. *An Introduction to Probability Theory and Its Applications*, 3rd ed., Vol. I. New York: John Wiley & Sons, Inc., 1968, Ch. X.
5. GOOVAERTS, M., DE VYLDER, F., AND HAEZENDONCK, J. *Insurance Premiums: Theory and Applications.* New York: Elsevier Publishing (North Holland), 1983.
6. KARLIN, S.J., AND STUDDEN, W.J. *Tchebycheff Systems: with Applications in Analysis and Statistics.* New York: John Wiley & Sons, 1966.
7. KENDALL, M.G., AND STUART, A. *Advanced Theory of Statistics*, Vol. I. *Distribution Theory.* New York: Hafner Press, 1977.

8. KLUGMAN, S.A. "On the Variance and Mean Squared Error of Decrement Estimators,*TSA* XXXIII (1982): 301.
9. REITANO, R.R. "A Statistical Analysis of Banded Data with Applications," *ARCH*, 1983.2, 157–193.
10. RIESZ, F., AND SZ.-NAGY, B. *Functional Analysis*. 2nd ed. New York: Frederick Ungar Publishing Co., 1972, Sec. 21.
11. RUDIN, W. *Real and Complex Analysis*. 2nd ed. *McGraw-Hill Series in Higher Mathematics*. New York: McGraw-Hill, Inc. 1974, Ch. 3.

DISCUSSION OF PRECEDING PAPER

ELIAS S.W. SHIU:

I wish to supplement this paper with an alternative proof for Formulas (4.6), (4.26), (4.36), (5.2), and (5.10). Let ϕ be a continuous function on the interval $[a, b]$. Assume that, over the interval $[a, b]$, the graph of ϕ is below the straight line joining the points $(a, \phi(a))$ and $(b, \phi(b))$; that is,

$$\phi(x) \leq \phi(a) + (x - a)[\phi(b) - \phi(a)]/(b - a), x \in [a, b]. \quad \text{(D.1)}$$

Then, for each random variable X, we have

$$E[\phi(X)] \leq \phi(a) + [E(X) - a][\phi(b) - \phi(a)]/(b - a). \quad \text{(D.2)}$$

Formula (D.1) holds for all convex functions ϕ on $[a, b]$. Now, for $k \geq 1$, the function

$$\phi(x) = x^k$$

is a convex function on each positive interval. With $[a, b] = [1, r]$, formula (D.2) becomes

$$E(X^k) \leq 1 + [E(X) - 1](r^k - 1)/(r - 1). \quad \text{(D.3)}$$

For each fixed t, the function

$$\phi(x) = e^{tx}$$

is convex; hence we have

$$M_X(t) \leq e^t + [E(X) - 1](e^{tr} - e^t)/(r - 1).$$

If ϕ is a convex function, then either (*i*) (D.1) is an equality, that is, ϕ is a linear function, or (*ii*) (D.1) is a strict inequality except for the endpoints a and b. Hence for a nonlinear convex function ϕ, (D.2) is an equality if and only if

$$Pr(X = a) + Pr(X = b) = 1.$$

Inequality (D.3) gives an upper bound for $E(X^k)$ in terms of $E(X)$. Perhaps, lower bounds for $E(X^k)$ are also of interest. Let me now repeat some results given in [4]. It can be proved ([1, section 16], [3, p. 455]) that, for each positive random variable X,

$$[E(X^s)]^{1/s} \leq [E(X^t)]^{1/t}, \qquad -\infty \leq s \leq t \leq \infty.$$

Hence

$$[E(X^s)]^{k/s} \le E(X^k), \qquad 0 < s \le k. \tag{D.4}$$

A special case of (D.4) is

$$[E(X)]^k \le E(X^k), \qquad 1 \le k. \tag{D.5}$$

It is possible to obtain inequalities sharper than (D.5). Given $b>a>0$ and $\rho \ge 1$, we have

$$\begin{aligned} b^\rho &= [a + (b - a)]^\rho \\ &\ge a^\rho + (b - a)^\rho, \end{aligned}$$

which is sharper than $b^\rho \ge a^\rho$. Tong ([5], [6, Lemma 2.3.1]) has applied this observation and (D.4) to obtain a lower bound for $E(X^k)$ in terms of the mean and variance of X. For a non-negative random variable X and a real number $k \ge 2$,

$$\begin{aligned} E(X^k) &\ge [E(X)]^k + E\{[X - E(X)]^k\} \\ &\ge [E(X)]^k + [\mathrm{Var}(X)]^{k/2}. \end{aligned}$$

We can also obtain a lower bound for $E(X^k)$ by means of #5.45 on page 158 of [2]. For a random variable Y, a non-negative Borel function η and a positive number c,

$$E[\eta(Y)] \ge cPr[\eta(Y) \ge c].$$

Consider $\eta(y) = |y|^k$. Then we have

$$E(|Y|^k) \ge cPr(|Y|^k \ge c) = cPr(|Y| \ge c^{1/k}),$$

which is called Markov's inequality or Tchebycheff's inequality.

The inequalities in the paper motivate the following question. Given a function ϕ defined on $[a, b]$, is there a systematic way to construct functions ψ such that

$$\phi(a) = \psi(a), \tag{D.6}$$

$$\phi(b) = \psi(b) \tag{D.7}$$

and either

$$\phi(x) < \psi(x) \qquad \text{for all } x \in (a, b) \tag{D.8}$$

or

$$\phi(x) > \psi(x) \qquad \text{for all } x \in (a, b)? \tag{D.9}$$

Let h be a function defined on $[a, b]$ with $h(b) \neq h(a)$. Then the function

$$\psi(x) = \phi(a) + \{[\phi(b) - \phi(a)]/[h(b) - h(a)]\}[h(x) - h(a)]$$

satisfies (D.6) and (D.7). Consider the difference between the functions ϕ and ψ,

$$d(x) = \phi(x) - \psi(x), x \in [a, b].$$

We seek conditions on ϕ and h such that either

$$d(x) < 0 \qquad \text{for all } x \in (a, b) \tag{D.10}$$

or

$$d(x) > 0 \qquad \text{for all } x \in (a, b). \tag{D.11}$$

Assume that d is differentiable. Since $d(a) = d(b) = 0$, (D.10) holds if d has exactly one minimum in (a, b), and (D.11) holds if d has exactly one maximum in (a, b). The equation

$$d'(y) = 0 \tag{D.12}$$

is equivalent to

$$\phi'(y) = \{[\phi(b) - \phi(a)]/[h(b) - h(a)]\}h'(y).$$

Assume that $\phi'(x)$ is never zero; that is, ϕ is a strictly increasing or decreasing function. Then (D.12) is equivalent to

$$h'(y)/\phi'(y) = [h(b) - h(a)]/[\phi(b) - \phi(a)]. \tag{D.13}$$

Note that y does not appear in the right-hand side of (D.13). If the function $h'(x)/\phi'(x)$ is a strictly increasing or decreasing function, then y is unique.

For an application of the analysis above, consider the functions $\phi(x) = x^k$ and $h(x) = x^j$. The function

$$h'(x)/\phi'(x) = (j/k)x^{j-k}$$

is a strictly decreasing function on $[1, r]$ if $j < k$ (it is strictly increasing on $[1, r]$ if $j > k$). Hence we have, for $j < k$,

$$\frac{x^k - 1}{r^k - 1} < \frac{x^j - 1}{r^j - 1} \text{ for all } x \in (a, b).$$

We conclude this discussion with two applications in life contingencies [2, p. 290, #9.31.c] and [2, p. 532, #18.21.c]. Observe that the functions $\phi'(x)/[\phi(b)-\phi(a)]$ and $h'(x)/[h(b)-h(a)]$ are *weight* functions. It follows from an integration by parts that

$$\int_a^b \left[\frac{\phi'(x)}{\phi(b)-\phi(a)} - \frac{h'(x)}{h(b)-h(a)}\right] g(x)\,dx$$

$$= -\int_a^b \left[\frac{\phi(x)}{\phi(b)-\phi(a)} - \frac{h(x)}{h(b)-h(a)}\right] dg(x). \qquad \text{(D.14)}$$

If g is a monotonic function, then the differential $dg(x)$ is of one sign. For #9.31, consider

$$\phi'(t) = {}_tp_x'^{(j)}$$

and

$$h'(t) = {}_tp_x^{(\tau)}.$$

Then $h'(t)/\phi'(t)$ is a decreasing function. It follows from (D.14) that

$$m_x'^{(j)} - m_x^{(j)} = -\int_0^1 \left(\int_0^t [w^{(j)}(s) - w^{(\tau)}(s)]ds\right) d\mu_{x+t} > 0.$$

For #18.21, consider

$$\phi'(t) = v^t\, {}_tp_x$$

and

$$h'(t) = \phi'(t)e^{-\Delta t},$$

where Δ is a change in the force of interest. The desired result follows from (D.14).

I thank Dr. Reitano for a thought-provoking paper.

REFERENCES

1. Beckenbach, E.F., and Bellman, R. *Inequalities*. New York: Springer-Verlag, 1961.
2. Bowers, N.L., Jr., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. *Actuarial Mathematics*. Itasca, Ill.: Society of Actuaries, 1986.

3. MARSHALL, A.W., AND OLKIN, I. *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press, 1979.
4. SHIU, E.S.W. Discussion of "Some Moment Inequalities and Their Applications," *TSA* XXXVIII (1986): 96–102.
5. TONG, Y.L. "Some Probability Inequalities of Multivariate Normal and Multivariate *t*," *Journal of the American Statistical Association* LXV (1970): 1243–47.
6. TONG, Y.L. *Probability Inequalities in Multivariate Distributions*. New York: Academic Press, 1980.

ERIC S. SEAH:

In the paper, the author provides sharp upper bounds μ'_k, given that μ is available; see Formulas (4.26) and (5.2). In this note, we show that, if we have additional information, namely, the variance σ^2, still better upper an be obtained.

Let $h(x)=x^k$, k being an integer ≥ 2, and $l(x)$ be the linear function joining the points $(1, h(1))$ and $(r, h(r))$. The idea is to find a quadratic function, $f(x)$, which passes through the points $(1, h(1))$ and $(r, h(r))$ and is "sandwiched" between $h(x)$ and $l(x)$ on $[1, r]$. If we require $f'(1)=h'(1)=k$, then

$$f(x) = \frac{1}{r-1}\left[\frac{r^k-1}{r-1} - k\right]x^2 + \frac{1}{r-1}\left[k(r+1) - 2\frac{r^k-1}{r-1}\right]x + \frac{1}{r-1}\left[\frac{r^k-1}{r-1} - (k-1)r - 1\right].$$

We now prove that $f(x)\geq h(x)$ on $[1, r]$. Let $g(x)=f(x)-h(x)$. It follows from the above that $g(1)=g(r)=g'(1)=0$. We note that, for $k=2$, there are at most two sign changes in the coefficients of $g(x)$, while for $k\geq 3$, there are at most three sign changes. By Descartes' Rule of Signs, $g(x)$ has at most two positive real roots when $k=2$ and at most three positive real roots when $k\geq 3$. In the case of $k=2$, $g(x)$ is a quadratic function and has exactly two real roots: 1 and r. For $k\geq 3$, $g(x)$ has exactly three real roots: 1, 1, and r. It is easy to show that $g''(1)\geq 0$. Therefore, $g(x)$ must be non-negative on $[1, r]$.

It is also easy to check that $l(x) \geq f(x)$ on $[1, r]$. Thus for a random variable X bounded between 1 and r, we have $h(X) \leq f(X) \leq l(X)$. Taking expectations yields $E[h(X)] \leq E[f(X)] \leq E[l(X)]$. Hence,

$$\begin{aligned}
\mu'_k \leq & \frac{1}{r-1}\left[\frac{r^k-1}{r-1} - k\right][\sigma^2 + \mu^2] \\
& + \frac{1}{r-1}\left[k(r+1) - 2\frac{r^k-1}{r-1}\right]\mu \\
& + \frac{1}{r-1}\left[\frac{r^k-1}{r-1} - (k-1)r - 1\right] \\
\leq & 1 + \frac{r^k-1}{r-1}(\mu - 1).
\end{aligned}$$

We conclude this discussion with the remark that, for all quadratic functions $m(x)$ passing through $(1, h(1))$ and $(1, h(r))$ such that $m(x) \geq h(x)$ on $[1, r]$, $f(x)$ is the one closest to $h(x)$; that is, $m(x) \geq f(x)$ on $[1, r]$. The proof is not difficult, which we leave to the interested readers.

ESTHER PORTNOY:

The method of moment spaces is an idea that has been thoroughly explored by statisticians (see the additional references at the end of this discussion), and it is good to see it brought into the actuarial literature and applied to a particular situation. Two separate problems should be distinguished. First is the problem of making the most of existing data; second is the question of redesigning data-collection (or data-reporting) methods for the future.

Consider the following numerical example.

Amount Range	Number of Policies	Average Amount
<$2,000	5	$1,380
$2,000–$4,999	15	$3,627
$5,000–$9,999	23	$7,791
$10,000–$24,999	53	$14,708
$25,000–$49,999	25	$32,152
$50,000–$99,999	10	$62,990
$100,000–$500,000	5	$184,220

Note that the "under \$2,000" band cannot be transformed into the form $(1,r)$; fortunately, this is not necessary. Knowing the boundaries (a,b), the average μ and the number n for each band, we can give a sharper estimate than (2.14), namely,

$$\mu_2' \le a^2 + (\mu - a)(a + b) - (b - a)^2 \frac{s(1 - s)}{n}. \qquad (1)$$

Here s is the fractional part of the polarizing parameter

$$t = \frac{n(\mu - a)}{b - a}.$$

Proof of (1): Among all collections $\{x_1, \ldots, x_n\}$ of n points in (a, b) having average μ, the greatest μ_2' is attained by the polarizing distribution $D(t)$:

$$x_i = \begin{cases} a & \text{for } i \le n - m - 1 \\ a + s(b - a) & \text{for } i = n - m \\ b & \text{for } i \ge n - m + 1 \end{cases}$$

where $m = t - s$, the greatest integer in t. But

$$\begin{aligned} n \cdot \mu_2' [D(t)] &= (n - m - 1)a^2 + [a + s(b - a)]^2 + m\, b^2 \\ &= n \cdot a^2 + m(b^2 - a^2) + 2as(b - a) + s^2(b - a)^2 \\ &= n \cdot a^2 + (a + b)[n(\mu - a) - s(b - a)] \\ &\quad + s(b - a)[2a + s(b - a)] \\ &= n \cdot a^2 + n(\mu - a)(a + b) - s(1 - s)(b - a)^2. \end{aligned}$$

Thus for any other collection with the same average μ,

$$\mu_2' \le \mu_2'[D(t)] = a^2 + (\mu - a)(a + b) - (b - a)^2 \frac{s(1 - s)}{n}. \quad QED$$

Since the last term on the right of (1) is positive, its omission gives

$$\mu_2' \le a^2 + (\mu - a)(a + b); \qquad (2)$$

equality will hold only for the polarizing distribution with $s = 0$ or 1, or in the limit as $n \to \infty$. Besides providing a sharper estimate than (2.14), the proof given here for (1) avoids the questionable argument following (2.19), where one must express the right side of (2.19) in terms of μ and then maximize, not vice versa.

For the example given, we have the following estimates from (1) compared to (2.14):

Amount Range	s	R.S. of (1) (000 omitted)	R.S. of (2.14) (000 omitted)
<\$2,000	0.45	2,562	2,760
\$2,000–\$4,999	0.135	15,319	15,389
\$5,000–\$9,999	0.839	66,718	66,865
\$10,000–\$24,999	0.635	263,796	264,780
\$25,000–\$49,999	0.152	1,158,178	1,161,400
\$50,000–\$99,999	0.598	4,388,401	4,448,500
\$100,000–\$500,000	0.053	58,933,042	60,532,000

The two upper bounds thus given for Σx_i^2 are 3.833×10^{11} and 3.920×10^{11}. The corresponding bounds for $s^2 = 1/n\ \Sigma x_i^2 - (1/n\ \Sigma x_i)^2$ are 1.587×10^9 and 2.267×10^9, amounting to about a 20 percent difference in estimates of the standard deviation. Of course, the numbers in this example are quite small; if the numbers in each band are of the order of a few hundred or more, the correction term will be negligible unless the bands are very wide.

A caution about using (1) has to do with round-off error in calculating the fractional part, s, of t. It would be better if Σx_i were reported, rather than $1/n\ \Sigma x_i$; if only the average is given, it might be prudent to experiment a bit to see the possible range of s. For instance, in the \$5,000–\$9,999 band, we have reported an average of \$7,791 on 23 policies; the sum of values could range between \$179,182 and \$179,204 so $12.8364 \leq t \leq 12.8409$. Thus the correction term

$$(b-a)^2 \frac{s(1-s)}{n}$$

is between 145,421 and 148,734. A more conservative upper bound for μ'_2 might be 66,720,000. Other situations might require more substantial modifications.

For the second objective, designing for future data collection, one cannot know n and s in advance, and so (2.14) becomes the best universal bound. However, rather than adjusting bandwidths (and possibly causing problems for those who collect the data and others who use them), it would seem more natural to revise the collection procedure to include and report $1/n\ \Sigma\ x_i^2$ for each band, thus leading to an exact value rather than estimate for the standard deviation.

REFERENCES

This is by no means an exhaustive list. More references can be found in the bibliographies of these papers.

1. BROCKETT, P.L., AND COX, S.H. "Insurance Calculations Using Incomplete Information," *Scandanavian Actuarial Journal* 2 (1985): 94–108.
2. COLLINS, J.R., AND PORTNOY, S.L. "Maximizing the Variance of *M*-Estimators Using the Generalized Method of Moment Spaces," *Annals of Statistics* 9 (1981): 567–77.
3. KAAS, R., AND GOOVAERTS, M.J. "Best Bounds for Positive Distributions with Fixed Moments," *Insurance: Mathematics & Economics* 5 (1986): 87–92.
4. KRIEGER, A. "On Bounding Moments from Grouped Data," *Journal of the American Statistical Association* 70 (1975): 64–66.
5. SKIBINSKI, M. "The Maximum Probability on an Interval When the Mean and Variance Are Known," *Sankhya* (Indian Journal of Statistics) Series A 39 (1977): 144–59.

PATRICK L. BROCKETT* AND SAMUEL H. COX:

We wish to point out an alternative method of bounding higher-order moments and moment generating functions that is capable of extension far beyond the setting of Dr. Reitano's paper. These methods have been discussed and applied to actuarial science problems [2]–[8]. In particular, from these papers it follows that the inequalities (4.26), (4.36), (5.2) and (5.10) are valid for *all* distributions concentrated on $[1, r]$ having a specified mean μ, and in fact, there is no need to assume that the distribution is either discrete or continuous; a single technique applies for all distributions. Moreover, the situations Dr. Reitano develops earlier in his paper are also special cases of the results given in [2]–[8].

The principal result given in the above publications that is applicable to Dr. Reitano's work is given below. We have used this result (for the case that the function h whose expectation is to be bounded is twice differentiable) in [2]. It is proven in the generality given here in [3] and [4] using techniques of Kemperman [8]. Chang [6] also established similar results.

*Dr. Brockett, not a member of the Society, is the Joseph H. Blades Professor of Insurance at the University of Texas at Austin.

Let X denote a random variable concentrated on $[a, b]$ with a known mean μ. If h is a continuous function that is convex over $[a, b]$, then

$$h(\mu) \leq E[h(X)] \leq h(a)p + h(b)(1 - p)$$

where $p = \dfrac{b - \mu}{b - a}$.

The definition of convex used in the above setting is geometric: For each pair of points $P=[x, h(x)]$ and $Q=[y, h(y)]$, the line joining P to Q lies entirely above the portion of the graph of h which joins P to Q. For example, functions such as $h(x)=x^k$, $k\geq 1$ or $h(x)=e^{tx}$ for a fixed real number t are convex functions frequently used in actuarial science. The proof of the above result is also geometrically apparent: To make the expectation of a convex function as large as possible, move as much mass as possible to the extreme end points a and b, while to make the expectation as small as possible, move as much mass as possible towards the center μ. The mean and total mass constraints then uniquely determine the exact extremal distributions as given above.

To obtain (4.26) and (5.2), set $a=1$, $b=r$ and $h(x)=x^k$. Then $h(\mu)=\mu^k$, $E[h(X)]=\mu'_k$, $h(a)=1$, $h(b)=r^k$ and $p=(r-\mu)/(r-1)$. Therefore, we obtain (4.26) and (5.2):

$$\mu^k \leq \mu'_k \leq \frac{r - \mu}{r - 1} + \frac{\mu - 1}{r - 1}r^k = 1 + \frac{\mu - 1}{r - 1}(r^k - 1).$$

To obtain (4.35) and (5.10), set $a=1$, $b=r$ and $h(x)=e^{tx}$. In this situation we have $h(\mu)=e^{t\mu}$, $E[h(X)]=M_X(t)$, $h(a)=e^t$, $h(b)=e^{rt}$ and $p=r-\mu/r-1$, so that we obtain (4.36) and (5.10):

$$e^{t\mu} \leq M_X(t) \leq e^t\frac{r - \mu}{r - 1} + e^{rt}\frac{\mu - 1}{r - 1} = e^t + \frac{\mu - 1}{r - 1}(e^{rt} - e^t).$$

This technique was used in [2] to obtain the same bounds on the moment generating function and was extended much further than the one moment case needed to establish Dr. Reitano's results. It is possible, for example, to obtain both upper and lower "best bounds" for the moments and moment generating function in the situation in which more than just the mean or even just the mean and variance are known. Unimodality constraints and higher-order moment knowledge can also be incorporated into the calculations with equal facility (and with corresponding closed form solutions).

These results were given in [2] and [3]. There is also a large literature on this, and more general related problems involving moments in actuarial science, to be found in the European actuarial literature. These results are summarized in [5] and [7].

As a final comment, we note that the original motivation for Dr. Reitano's investigation was to estimate expectations (in his case the variance) when the data were presented already grouped with only the means of the individual subintervals known. Instead of estimating the expectation directly, he found bounds in terms of the interval limits and the given mean. A general alternative solution to the original problem is available, however, using information theoretic methods. These methods are detailed in [1] and are sufficiently general to include the situation in which means, or medians, or percentiles, or any of a multitude of other information is known about the banded data. Moreover, the estimates are truly estimates in the statistical sense (rather than bounds) with known statistical properties and asymptotic distributions. In addition, unimodality can be incorporated without substantially complicating the problem and the estimates can be readily computed by recourse to an unconstrained convex programming problem. All these results can be established by using the techniques outlined in [1].

REFERENCES

1. BROCKETT, P.L. "Information Theoretic Approach to Actuarial Science: A Unification and Extension of Relevant Theory and Applications," *TSA* XLIII (1991), in press.
2. BROCKETT, P.L., AND COX, S.H. "Optimal Ruin Calculations Using Partial Stochastic Information," *TSA* XXXVI (1984): 49–62.
3. BROCKETT, P.L., AND COX, S.H. "Insurance Calculations Using Incomplete Information," *Scandinavian Actuarial Journal* (1985): 94–108.
4. BROCKETT, P.L., AND COX, S.H. "Best Bounds on Measures of Dispersion for Alpha-Unimodal Random Variables," 1989, working paper.
5. BROCKETT, P.L., GOOVAERTS M., AND TAYLOR, G. "A Review of the Use of Moment Bounds in Actuarial Analysis," 1991, working paper.
6. CHANG, Y.C. "Chance Constrained Programming and Tchebycheff Systems with Applications." Ph.D. diss., University of Texas at Austin, 1990.
7. GOOVAERTS, M., KAAS, R., VAN HEERWAERDEN, A.E., AND BAUWELINCKX, T. *Effective Actuarial Methods*. New York: North Holland Publ., 1990.
8. KEMPERMAN, J.H.B. "Geometry of the Moment Problem," *Proceedings of Symposia in Applied Mathematics* 37 (1987): 16–53.

(AUTHOR'S REVIEW OF DISCUSSION)

ROBERT R. REITANO:

I would like to thank Drs. Shiu, Seah, Portnoy, Brockett, and Cox for their thought-provoking discussions.

Dr. Shiu provides an array of extensions of the inequalities in my paper. Using only the convexity property of the power and exponential functions, he easily and elegantly develops the upper bounds for $\mu'_k(x)$, $k \geq 1$, and for $M_x(t)$. My proof of Theorem 1, being more constructive, simply fills in some of the details to assure sharpness for finite collections of points or distributions of rational range where $Pr(a)+Pr(b) \neq 1$, and perhaps gives more insight via polarized distributions, to the kinds of distributions that need to be considered. However, Dr. Shiu's proof is certainly sufficient for the actual result desired.

By introducing information on the variance, he also provides a better lower bound for $\mu'_k(x)$ than that produced with μ alone. Though not directly usable in my paper's applications, in which variance was assumed unknown, it is clear that in other contexts, the better bound would be preferred. Better yet, however, are the sharp lower bounds of Brockett and Cox [1], [2], if both μ and σ^2 are assumed known.

Dr. Shiu also reminds us of Tchebycheff's inequality, which provides lower bounds for $\mu'_k(x)$ based on information about $F(x)$. Such information, however, is often unknown. In addition, the sharpness of this type of estimate, vis-à-vis the estimate produced with μ, is difficult to predict.

Finally, by developing a general approach to approximating a given function with another which majorizes it, Dr. Shiu produces an interesting generalization of (2.14) of my paper; namely, for $k>j$:

$$\mu'_k(x) \leq 1 + \frac{r^k - 1}{r_j - 1} (\mu'_j(x) - 1),$$

and also utilizes this methodology in a life contingencies context.

Dr. Seah, again assuming information on variance, provides a better upper bound for $\mu'_k(x)$ than that possible with μ alone. His methodology is to find a closest fitting quadratic function that majorizes x^k, $k \geq 2$ on $[1, r]$. His approach is similar to the one used in Reitano [3], where a best-fitting majorant quadratic approximation to $(1+i)^{1-s}$ was sought.

Although it improves my result, which reflects μ alone, this upper bound is not as informative as possible, given the additional information. As referenced above, Brockett and Cox [1], [2] also provide sharp upper bounds to $\mu'_k(x)$, when μ and σ^2 are assumed known.

Dr. Portnoy wonders if I haven't given up too much in Theorem 1 by not using my Equation (2.19) as the upper bound for $\mu'_k(x)$, which she has reduced to her Equation (1). Certainly I have given up something to avoid the extra term involving s, as well as the practical problem of estimating s that she notes. However, it is usually the case that her upper bound will be very close to mine.

First, Dr. Portnoy's example, whereby her estimate of the standard deviation differed by 20 percent from mine, is in error by an order of magnitude. I discovered this as a byproduct of the analysis developed below, then checked her calculations. The mistake occurs in the calculation of s^2, using her Formula (1). The correct answer is 2.202×10^9, not 1.587×10^9, for a relative difference of 2.8 percent compared to my estimate. The corresponding estimates for the standard deviations then differ by about 1.4 percent, not 20 percent.

On the other hand, it is possible to develop examples for which the relative error is greater. Fortunately, the analysis below shows that this will rarely happen in practice.

To this end, let $s^2(s)$ equal Dr. Portnoy's estimate of the variance using her Equation (1), and $s^2(0)$ equal that produced by my Equation (2.26), or equivalently, by her formula with $s=0$. One then obtains:

$$1 - \frac{s^2(s)}{s^2(0)} = \frac{s(1-s)}{n} \times \frac{(r-1)^2}{(\mu-1)\,(r-\mu)}. \tag{D.1}$$

Using my Equation (2.12), (D.1) reduces to:

$$1 - \frac{s^2(s)}{s^2(0)} = \frac{s(1-s)}{n} \times \frac{1}{t/n(1-t/n)}, \tag{D.2}$$

which makes it clear that the relative error is independent of the size of amount band r. Because $0 \le s \le 1$, an upper bound for (D.2) is given by:

$$1 - \frac{s^2(s)}{s^2(0)} \le \frac{1}{4n} \times \frac{1}{p(1-p)}, \tag{D.3}$$

where $p = t/n$ equals the percentile of μ within the range $[1, r]$.

From (D.3), it is clear that for fixed p, the relative discrepancy decreases quickly with n. Using a conditioning argument, it can be shown that the relative discrepancy for the portfolio is less than the weighted average of individual band discrepancies, which by (D.3) could be no more than: 23, 7, 4, 2, 5, 13, and 30 percent, respectively. The weights used reflect the n_j for each band as well as $s_j^2(0)$. Because Dr. Portnoy's portfolio discrepancy for s^2 was 30 percent [1.587 v. 2.267], it was clear that it must have been in error as noted above.

Of course, (D.3) readily provides information on the relative disparity in standard deviation estimates. For example, if $0.1 \leq p \leq 0.9$, the relative discrepancy will be less than 1 percent for $n \geq 140$. Similarly, $n \geq 265$ will suffice for $0.05 \leq p \leq 0.95$.

For amount bands for which p is very close to 0 or 1, (D.3) provides a poor upper bound. For example, if $t<1$, so $s=t$ and $p=s/n<1/n$, we obtain from (D.2):

$$1 - \frac{s^2(s)}{s^2(0)} = 1 - \frac{n-1}{n} \times \frac{s}{1-s/n}. \tag{D.4}$$

Consequently, for fixed n, the relative error can approach 100 percent as $p = s/n$ approaches 0; that is, for μ very close to 1, the lower bound of the interval. In practice, however, this result may be of little concern. A similar analysis holds for p near 1, or μ near the upper bound r. However, for such extreme cases of μ, Dr. Portnoy's Formula (1) is certainly easy to apply.

Dr. Portnoy seems not to follow the logic of my proof of Theorem 1. After showing that $\mu'_k\,[D(t)]$ is a piecewise convex quadratic function of t, which equals the linear function, $1+(r+1)\,\{\mu[D(t)]-1\}$, at the partitioning t values of $0, \ldots, n$, the conclusion is straightforward.

Finally, I agree with Dr. Portnoy's preference for calculating s^2 exactly, rather than estimating it based on banded data, though at the time of my original analysis, this was not an option. However, I disagree that a good banding convention would cause problems for contributors to mortality studies. For example, starting from \$10,000, the following bands produce r values close enough to 2.48 that a 5 percent relative error would be virtually assured:

10–24,999
25–49,999
50–99,999
100–249,999
250–499,999
500–999,999
1,000–2,499,999, etc.

Clearly, this banding convention is generally conservative and fits in well with the amount bands many insurers use as a basis for pricing.

Drs. Brockett and Cox begin their discussion with the observation that "These methods have been discussed and applied to actuarial science problems." Indeed, the cited works follow the 17th Actuarial Research Conference in 1982 where I presented the original version of my research (see the Acknowledgment to the current paper, p. 403), as well the publication of that paper in *ARCH* [4]. This work has subsequently been generalized.

Drs. Brockett and Cox use the convexity property of the power and exponential functions to develop sharp estimates for $\mu'_k(x)$, as did Dr. Shiu, and note that corresponding sharp estimates are also possible when additional information besides μ is known, for example, when σ^2 is known as noted above.

For the applications developed in my paper, however, in which policy data are assumed to be banded, their generalizations may be of theoretical interest only. Indeed, presented with such data, one can hardly imagine having information on the variance, higher moments, medians, or percentiles, let alone knowledge as to whether or not the underlying distribution is unimodal.

In closing, I would like to again thank Drs. Shiu, Seah, Portnoy, Brockett, and Cox for their stimulating discussions and for the additional references to related work.

REFERENCES

1. BROCKETT, P.L., AND COX, S.H. "Optimal Ruin Calculations Using Partial Stochastic Information," *TSA* XXXVI (1984): 49–62.
2. BROCKETT, P.L. AND COX, S.H. "Insurance Calculations Using Incomplete Information," *Scandinavian Actuarial Journal* (1985): 94–108.
3. REITANO, R.R. Discussion of "Asset Share Mathematics," by Peyton J. Huffman, *TSA* XXX (1978): 309–318.
4. REITANO, R.R. "A Statistical Analysis of Banded Data, With Applications," *ARCH*, 1983.2, 157–93.