A Statistical Analysis of Banded Data, with Applications

Robert R. Reitano

ABSTRACT

The purpose of this paper is to develop sharp estimates for the higher moments of a distribution of a positive bounded random variable, where these estimates are given as functions of the first moment. Upper and lower sharp estimates will first be developed for the second moment of a distribution of a discrete random variable. Applications will then be explored in detail concerning interval and point estimates for the variance of expected claims, and modified confidence intervals for this random variable. In addition , applications will be made to the problem of establishing an appropriate level for retention limits as well as analyzing the variance of decrement estimators. These upper and lower estimates will then be generalized to higher order moments. Finally, higher moments of a distribution of a continous random variable will be considered.

I. Introduction

When analyzing the random variable S defined to represent the aggregate amount of expected claims in a given period of time, it is often necessary to calculate the sums of in-force policy amounts, sums of squares, etc. Although this will be formalized in Section III (Applications), these sums must be calculated for each homogeneous class of policyholders, where homogeneity is defined with respect to the expected probability of claim. For sizable portfolios, this summing process can be a formidable task. The purpose of this paper is to develop sharp estimates for the values of higher moments of a policy amount distribution, or any p.d.f. of a positive bounded random variable (r.v.) where these estimates are made using only the first moment.

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 p.d.f. defined on a positive bounded r.v., the value of all higher moments are constrained, a priori, once the first moment and domain interval are given. For brevity, the upper and lower bounds developed will be 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 will be formalized and sharp estimates developed for the second moment of policy amounts or any finite collection of positive numbers. As can be expected, 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 will also become clear which types of discrete distributions maximize the relative error for given values of r and μ where μ is the value of the first moment, and, which discrete distribution maximizes the error for a given r with no restriction on μ. As a corollary, these estimates will be translated into estimates of the variance.

In Section III, a number of applications will be explored in detail. Section IV generalizes II to the case of higher moments of an arbitrary discrete p.d.f., and in Section V, the higher moments of a continuous p.d.f. are considered. As a corollary, these estimates will be 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 will be developed for μ_2' and the ratio $R_2(x)$, where,

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

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

and $\mu = \frac{1}{n}\sum x_i$. The approximation for μ_2' will be a linear function of μ and r, and that for $R_2(x)$ a function of r alone.

If one considers 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 $\lambda = 1/a$, it is possible to estimate (2.1) and (2.2) only for $\{x_i\}$ in $[1,r]$, then apply (2.3) - (2.5) to yield the analogous results for $\{x_i\}$ in $[a,b]$.

An upper estimate $s_2(r,\mu)$ for μ_2' is defined to be sharp if $\mu_2'(x) \leq s_2(r,\mu)$ for all $\{x_i\} \subset [1,r]$, but for any $\varepsilon > 0$, $r > 1$, and μ, $1 \leq \mu \leq r$, there is a distribution $\{y_i\} \subset [1,r]$ such that $\mu(y) = \mu$ and,

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

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

$$R_2(y) > s_2(r) - \mathcal{E} \quad . \tag{2.7}$$

Sharp lower estimates are defined analogously.

Since it is certainly true that $R_2(x) \leq r^2$, it is clear that $s_2(r)$ will satisfy $s_2(r) \leq r^2$ and hence will have the property that for any $c > 0$, there is an $r > 1$ such that $s_2(r) \leq 1 + c$. Consequently,

$$\sum x_i^2 \leq (1 + c)n\mu^2, \qquad \{x_i\} \subset [1,r] \, . \tag{2.8}$$

Hence, 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 will be shown in Section III.2. that the standard deviation of $\{x_i\}$ can be approximated with the mean to within a 5% relative error by choosing $c = .22$. Using the observation that $s_2(r) \leq r^2$, 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% by 'banding' amounts into $[ar^j, ar^{j+1})$, $j \geq 0$, with $r=1.1045$. Unfortunately, to analyze experience between $1,000 and $10,000,000 would require over 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 will be shown later, using the derived $s_2(r)$, that $c = .22$ can be obtained with $r = 2.48$, 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 it is possible to reduce the analysis of $\mu'_2(x)$ and $R_2(x)$ from the collection of all distributions $\{x_i\}_{i=1}^{n} \subset [1,r]$ to a linearly parametrized collection of distributions, D(t), one for each mean μ, $1 \le \mu \le r$. 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:

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

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 from $[1,r]$ 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 will be called <u>polarized distributions</u>, and can be parametrized over $t \in [0,n]$ by D(t), where $D(t) = \{x_i\}_{i=1}^{n}$ is defined by:

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

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 another. Alternatively, if $\{x_i\}$ is identified with a point in n-dimensional space, $\mathbb{R}^n$, D(t) can be thought of as a piecewise linear transformation from $[0,n]$ to a one dimension edge of the hypercube $[1,r]^n$ extending from (1,1,...,1) to (r,r,...r).

Since,

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

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

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

<u>Theorem 1</u> Let $\{x_i\}_{i=1}^{n} \subset [1,r]$. Then,

$$\mu_2'(x) \leq 1 + (r+1)(\mu - 1), \ \mu = \mu(x), \tag{2.14}$$

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

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

<u>proof</u> Assume that (2.14) has been established. Then,

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

As a function of μ on $[1,r]$, 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). Using (2.11),

$$\mu_2'(D(t)) = \frac{n - [t] - 1 + [t]r^2 + ((t - [t])(r-1)+1)^2}{n} \ , \ 0 \leq t \leq n. \tag{2.17}$$

Let,

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

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

$$\mu_2'(D(t)) = \frac{n-m-1 + mr^2 + (s(r-1)+1)^2}{n}, \quad m=o,\ldots, n-1; 0 \leq s \leq 1. \tag{2.19}$$

The inequality for s was extended to s=1 as 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..,n and s=0. For the resulting values of t=m,

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

$$\mu_2'(D(t)) = \frac{n-m+mr^2}{n}, \quad m=0,1..n \tag{2.21}$$

and a calculation proves 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 $1 < \mu < r$ will be considered; hence $\rho > 0$. Let c_j be a sequence of positive rational numbers, $c_j = m_j/n_j$, that converge to ρ and satisfy,

$$0 \leq \lambda_j \leq 1, \quad \lambda_j = \frac{c_j - \rho}{c_j}, \tag{2.22}$$

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

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

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

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

$$\mu_2'(y) = 1 + c_j(r^2-1) + c_j\lambda_j(\lambda_j(r-1)^2 - 2r(r-1)). \tag{2.25}$$

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. ▌

<u>Corollary</u> Let $\{x_i\}_{i=1}^n \subset [1,r]$. Then,

$$\sigma^2(x) \leq (r-\mu)(\mu-1), \quad \mu=\mu(x), \tag{2.26}$$

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

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

<u>proof</u> 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) - (2.5), and,

$$\sigma^2(\lambda x) = \lambda^2\sigma^2(x), \tag{2.28}$$

where $\{x_i\} \subset [a,b]$ and $\lambda = \frac{1}{a}$. Consequently, $\{\lambda x\} \subset [1,r]$ with $r = b/a$.

It is clear from (2.26), since this inequality is sharp, that distributions of maximal variance must have a mean equal to (r+1)/2, the midpoint of $[1, r]$. Also, the distributions with maximal ratio of variance to mean squared must have a mean given by:

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

Consequently, the associated polarized distribution D(t) is given with t defined in (2.13), which due to (2.29) yields,

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

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

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

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

$$f(1) \sim 1, \quad r \to \infty . \qquad (2.32)$$

This is also evidenced by noting that due to (2.29), it is clear that μ is an increasing function of r with upper bound equal to 2.

A lower bound for μ'_2 is fairly easy to develop by utilizing the Cauchy-Schwarz inequality [1], which states that for given a_i, b_i, i = 1,...,n,

$$\sum |a_i b_i| \leq (\sum a_i^2)^{\frac{1}{2}} (\sum b_i^2)^{\frac{1}{2}}, \qquad (2.33)$$

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

$$\alpha a_i + \beta b_i = 0, \ i = 1,..,n. \qquad (2.34)$$

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

$$\sum x_i \leq (\sum x_i^2)^{\frac{1}{2}} n^{\frac{1}{2}}, \text{ or,} \tag{2.35}$$

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

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 $\sigma^2(x)$.

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 probability of becoming a claim. Let X_j be the binomial random variable defined so that,

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

Then S, defined by,

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

is the random variable that represents the total 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 and for a given class C_j, can be 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, (3.2) is simply a linear combination of independent binomial variables, and one has directly,

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

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

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

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

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

n_{jk} = number of policies in C_{jk},

μ_{jk} = average policy amount in C_{jk}.

Then for each class C_j, the following is true according to (2.14) and (2.36):

$$\sum_k n_{jk}\mu_{jk}^2 \leq \sum_i a_{ij}^2 \leq \sum_k n_{jk}\left[a_k^2 + a_k(r_k+1)(\mu_{jk}-a_k)\right]. \qquad (3.7)$$

Consequently,

$$\sum n_{jk}q_j(1-q_j)\mu_{jk}^2 \leq \sigma^2(S) \leq \sum n_{jk}q_j(1-q_j)\left[a_k^2 + a_k(r_k+1)(\mu_{jk}-a_k)\right]. \qquad (3.8)$$

When policy data is already banded, (3.8) provides the resultant interval estimate for $\sigma^2(S)$, which may or may not be acceptable. Consequently, the amount of relative error in the point estimate discussed in the next section will also be fixed. However, if it is possible to choose bands at will, 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:

$$r^x = b/a, \qquad N = [x] + 1. \qquad (3.9)$$

Define the bands J_k by,

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

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 (2.15) and (2.36):

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

Consequently,

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

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

Finally, assume that the A_{ij} are independent random variables which are identically distributed for each j with mean μ_j and variance σ_j^2. If these distributions are assumed given by mathematical formual, μ_j and σ_j^2 will often be straightforward to compute and $\sigma^2(S)$ given directly by (3.4). On the other hand, these distributions may be defined empirically with reference to actual claim data from a recent period of time. 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. For a given value of $r > 1$, let N be the solution of (3.9) and J_k be defined as in (3.10). Also, let C_{jk} denote the collection of <u>claims</u> from class C_j with claim amount in the interval J_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 in amount interval J_k, they are related by a general formula involving conditional expectations [1].

Specifically,

$$\sigma_j^2 = \text{Var}\ (A_{\cdot j})$$

$$= \text{Var}\left[E\ (A_{\cdot j} | J_k)\right] + E\left[\text{Var}\ (A_{\cdot j} | J_k)\right]$$

$$= \text{Var}\left[\mu_{jk}\right] + E\left[\sigma_{jk}^2\right].$$

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, i.e. $\mu_j = E\left[\mu_{jk}\right]$. Combining (3.13) with (2.27) and the remarks following (2.36),

$$\frac{1}{m_j} \sum_k m_{jk}\ (\mu_{jk} - \mu_j)^2 \leq \sigma_j^2 \leq \frac{1}{m_j} \sum_k m_{jk} \left[(\mu_{jk} - \mu_j)^2 + \frac{(r-1)^2}{4r} \mu_{jk}^2 \right] . \tag{3.14}$$

By choosing r close enough to 1, it is clear that the accuracy of (3.14) can be made arbitrarily good. Utilizing (3.14) in (3.4) will then provide the required interval estimate for $\sigma^2(S)$. Also, (3.14) can be used when the Collective Risk Model for S is assumed if (3.4) is appropriately modified [1].

2. Point Estimation of the Standard Deviation of Expected Claims

Although (3.8), (3.12) and (3.14) 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 and unknown real number satisfying:

$$a \leq x \leq (1+c)a. \tag{3.15}$$

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

$$\hat{x}(\lambda) = (1+\lambda c)a \quad , \quad 0 \leq \lambda \leq 1. \tag{3.16}$$

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

<u>proof</u> Utilizing (3.15), (3.16), and the fact that $\hat{x}(\lambda) > 0$,

$$\frac{-\lambda c}{1+\lambda c} \leq \frac{x-\hat{x}}{\hat{x}} \leq \frac{c(1-\lambda)}{1+\lambda c}. \tag{3.17}$$

That is,

$$\left| \frac{x-\hat{x}}{\hat{x}} \right| \leq \max \left[\frac{(1-\lambda)c}{1+\lambda c}, \frac{\lambda c}{1+\lambda c} \right]. \tag{3.18}$$

A straightforward analysis yields that $f(\lambda) = c(1-\lambda) / (1+\lambda c)$ is a positive decreasing function over $[0,1]$ that agrees with $g(\lambda) = \lambda c/ (1+\lambda c)$, a positive increasing function, at $\lambda = \frac{1}{2}$. Consequently, the relative error defined in (3.18) is minimized when $\lambda = \frac{1}{2}$. ∎

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

$$\frac{(r+1)^2}{4r} = 1 + c. \tag{3.19}$$

Since $h(r) = (r+1)^2/4r$ is an increasing function of r for $r \geq 1$ and $g(1) = 1$, it is clear that the solution of (3.19) must exist, be unique and greater than 1. Rewriting (3.12),

$$\left(\sum n_{jk} q_j (1-q_j) \mu_{jk}^2 \right)^{\frac{1}{2}} \leq \sigma(S) \leq (1+c)^{\frac{1}{2}} \left(\sum n_{jk} q_j (1-q_j) \mu_{jk}^2 \right)^{\frac{1}{2}}. \tag{3.20}$$

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

$$\hat{\sigma}(S) = (1+\frac{d}{2}) \quad (\sum n_{jk} q_j \ (1-q_j) \mu_{jk}^2 \quad)^{\frac{1}{2}}. \tag{3.21}$$

Then using (3.18),

$$| \sigma(S) - \hat{\sigma}(S) | \leq \frac{d}{2+d} \ \hat{\sigma}(S). \tag{3.22}$$

Hence, in order for the point estimate in (3.21) to have an error relative to itself of 100ε %, $0 < \varepsilon < 1$, i.e.,

$$(1-\varepsilon)\hat{\sigma}(S) \leq \sigma(S) \leq (1+\varepsilon)\ \hat{\sigma}(S), \tag{3.23}$$

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

$$d = \frac{2\varepsilon}{1 - \varepsilon} \tag{3.24}$$

$$c = (d+1)^2 - 1 \tag{3.25}$$

$$r = 1 + 2 \ (c+\sqrt{c^2+c}) \tag{3.26}$$

The value of r in (3.27) is the larger solution of (3.19) expressed as a quadratic equation in r, where this root is given by the quadratic formula. The following table provides some numerical results:

Table 1

ε	.5	.1	.05	.01	.005
d	.2	.22	.11	.02	.01
c	.8	.49	.22	.04	.02
r	33.97	3.70	2.48	1.49	1.32

For example, to achieve a relative error of 5% 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=1,000, b=10,000,000, the solution x of (3.9) is approximately 10.14. Consequently, N=11 as was stated in section II. To limit the relative error to 1% would require 33 bands. Since $h(r) = (r+1)^2/4r$ has first derivative equal to 0 when r=1, reductions in $\bar{c}$ can only be achieved by relatively large decreases in r and hence, large increases in N.

When estimating $\sigma(S)$ based on (3.4) and (3.14), it is not possible to determine in advance what level of relative error will be produced by a given value of r. In contrast to (3.19), the analogous approach would produce the following equation in r:

$$\frac{\sigma_{max}^2(S)}{\sigma_{min}^2(S)} = 1 + c, \tag{3.27}$$

where $\sigma_{max}^2(S)$ is defined as the right hand side of (3.4) with σ_j^2 set equal to the upper bound in (3.14), and $\sigma_{min}^2(S)$ is analogously defined but with the lower bound in (3.14). When this ratio was taken with the bounds in (3.12), the terms involving n_{jk} and μ_{jk} cancelled out, which was fortunate since these variables are really functions of r. The ratio in (3.27) remains a complicated function of r since it involves m_{jk} and μ_{jk}, both functions of r.

In general, r is chosen equal to some convenient value, the resultant m_{jk} and μ_{jk} evaluated, and c determined by (3.27). The point estimate,

$$\hat{\sigma}(S) = \tfrac{1}{2} \left(\sigma_{max}(S) + \sigma_{min}(S)\right), \tag{3.28}$$

will then have a maximum relative error of 100ε%, where.

$$\varepsilon = \frac{d}{2 + d}, \quad d = (1+c)^{\frac{1}{2}} - 1. \tag{3.29}$$

If the resultant value of ε is unacceptable, the process will need to be repeated with a smaller value of r. One method of obtaining a new trial value of r is as follows. Let $\varepsilon' > 0$ be the targeted value for the desired level of relative error in the point estimate $\hat{\sigma}(S)$. Let d' be defined as in (3.24) but relative to ε', and let c' be the corresponding value from (3.25). The new trial value of r,r', is then taken as the solution of (3.27) with c' used instead of c, and, where m_{jk} and μ_{jk} are taken to be constant and equal to the respective values obtained with the prior value of r. By assumption, $\varepsilon' < \varepsilon$. Consequently, $c' < c$ and $r' < r$. Once r' has been determined, new values of m_{jk} and μ_{jk} can be developed, the <u>actual</u> resultant relative error of the new estimate $\hat{\sigma}(S)$ given in (3.28) evaluated using (3.27) and (3.29), and the process repeated if desired.

The estimate given in (3.28) and resultant error in (3.29) are also valid in the context of (3.8) where $\sigma_{max}(S)$ and $\sigma_{min}(S)$ are redefined with reference to the respective bounds in (3.8), and c chosen according to (3.27).

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,

$$\text{Prob}\left(|z| \le z_\alpha \;\middle|\; z \sim N(0,1) \right) = 1 - \alpha. \tag{3.30}$$

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 $\hat{\sigma}(S)$ is chosen to limit the relative error to 100ε %, in the sense of (3.23), the corresponding modified confidence interval for S becomes:

$$\mu(S) - (1+\varepsilon) z_\alpha \hat{\sigma}(S) \le S \le \mu(S) + (1+\varepsilon) z_\alpha \hat{\sigma}(S). \tag{3.32}$$

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:

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

where X_j and A_{ij} are 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 one considers $\mu(S^R)$ to be 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:

$$M^R = S^R / \mu(S^R). \tag{3.34}$$

By the Central Limit Theorem, the $100(1-\alpha)\%$ confidence interval for M^R is approximately given by:

$$\left| M^R - 1 \right| \le z_\alpha \frac{\sigma(S^R)}{\mu(S^R)}, \tag{3.35}$$

which can be modified as in (3.32) to:

$$\left| M^R - 1 \right| \le (1+\varepsilon) z_\alpha \frac{\hat{\sigma}(S^R)}{\mu(S^R)}. \tag{3.36}$$

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 (3.21). Then for $R=ar^k$.

$$\frac{\hat{\sigma}(S^R)}{\mu(S^R)} = (1+\frac{d}{2})\ \frac{\left(\sum_j q_j\ (1-q_j)\left[\sum_{l<k} n_{jl}\mu_{jl}^2 + R^2\sum_{l\geq k} n_{jl}\ \right)^{\frac{1}{2}}}{\sum_j q_j\left[\sum_{l<k} n_{jl}\mu_{jl} + R\sum_{l\geq k} n_{jl}\right]}, \tag{3.37}$$

and these values are straightforward to calculate since 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:

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

where $\nu > 0$ 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 the $\nu = 2$ when this distribution is uniform. In general, $\nu > 2$ reflects skewness to the left, $\nu < 2$ skewness to the right.

5. <u>Variance of Decrement Estimators</u>

As a last application, consider the formulas presented in [5] for the moments of $\hat{q}$ defined by (notation changed):

$$\hat{q} = \frac{\sum A_i X_i}{\sum A_i}, \tag{3.39}$$

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

$$\text{Prob.} (X_i = 1 \mid A_i = A) = q(A). \tag{3.40}$$

If it is assumed that A_i and X_j are mutually independent for all i and j, and that $q(A_i) = q$, the variance of $\hat{q}$ as derived in [5] is:

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

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

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

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 (2.15) and (2.36), we have that:

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

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

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

For a given value of r, $E\left[\frac{1}{N_k}\right]$ will usually increase as k increases. In particular, both bounds in (3.44) 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 $Var(q_k)$ to any given degree of accuracy by choosing r close enough to 1. Also, $Var(\hat{q}_k)$ can be approximated by $\hat{\sigma}^2(\hat{q}_k)$, using Lemma 1, where:

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

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

Although (3.45) can also be utilized to estimate $Var(\hat{q})$ with r=b/a and $N=\sum N_k$, the result may be considered quite crude for realistic values of b/a. For example, if a=1,000 and b=10,000,000, (3.45) becomes:

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

and the resultant value of $\hat{\sigma}^2(\hat{q})$ will have a maximum relative error of almost 100%. However, for n large, the absolute error will usually be quite small and the estimates of 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 is defined on a policyholder basis. Also, class C_j was assumed homogeneous with respect to the value of this probability and consequently, would be defined in terms of the various underwriting parameters of the insurance product under study. To simplify calculations, it will often be desirable to combine various underwriting 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 many classes are too sparse to confidently analyze.

When $A_{ij}=a_{ij}$ is assumed known and fixed, it is possible to analyze what effect such groupings will have on the variance of S in (3.6). To this end, let $\{C_j\}$ be a collection of classes to be grouped with respective claim probabilities $\{q_j\}$ and class sizes $\{n_j\}$, $n=\sum n_j$. As a combined class, the claim probability q is given by:

$$q = \sum n_j q_j / n. \tag{3.48}$$

For notational convenience, let σ^2 denote that part of the sum in (3.6) which corresponds to the classes under consideration,

$$\sigma^2 = \sum a_{ij}^2 q_j (1-q_j). \tag{3.49}$$

Also, let $\bar{\sigma}^2$ be analogously defined under the assumption that $\bigcup C_j$ is a homogeneous class with claim probability q as defined in (3.48),

$$\bar{\sigma}^2 = \sum a_{ij}^2 q (1-q). \tag{3.50}$$

Finally, let A_j^2 denote the second moment of the amounts in C_j,

$$n_j A_j^2 = \sum_i a_{ij}^2. \tag{3.51}$$

Lemma 2 Let σ^2 and $\bar{\sigma}^2$ be defined as in (3.49) and (3.50) respectively. Further, assume that the amount distributions of the various classes C_j are similar in that $A_j^2 = A^2$ for all j where A_j^2 is defined in (3.51). Then,

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

proof By definition,

$$\bar{\sigma}^2 - \sigma^2 = \sum n_j A_j^2 \left[q(1-q) - q_j(1-q_j) \right]. \tag{3.53}$$

Now since $A_j^2 = A^2$ and $\sum n_j q = \sum n_j q_j$, (3.53) can be rewritten as,

$$\bar{\sigma}^2 - \sigma^2 = A^2 \sum n_j q_j \ (q_j - q). \tag{3.54}$$

By (3.48),

$$q_j - q = \frac{\sum n_1 (q_j - q_1)}{n}, \tag{3.55}$$

which when substituted into (3.54) yields,

$$\bar{\sigma}^2 - \sigma^2 = A^2 \sum_{j,1} \frac{n_j n_1}{n} q_j \ (q_j - q_1). \tag{3.56}$$

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

$$\bar{\sigma}^2 - \sigma^2 = A^2 \sum_{j<1} \left[n_j n_1 q_j \ (q_j - q_1) + n_j n_1 q_1 \ (q_1 - q_j) \right] \frac{1}{n} \tag{3.57}$$

$$= A^2 \sum_{j<1} \frac{n_j n_1}{n} \ (q_j - q_1)^2 .$$

Hence, (3.52) is proved. ∎

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 $[1,r]$, since,

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

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

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

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

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

<u>Theorem 2</u> Let $\{x_i\}_{i=1}^{n} \subset [1,r]$, k a real number satisfying $k \geq 1$. Then,

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

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

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,

$$\mu_k'(D(t)) = \frac{n-m-1+mr^k+(s(r-1)+1)^k}{n}, \qquad \begin{array}{l} m=0,\ldots,n-1 \\ 0 \leq s \leq 1. \end{array} \tag{4.10}$$

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,...,n and s=0. For such values,

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

$$\mu_k'(D(t)) = \frac{n-m+mr^k}{n}, \quad m=0,1,\ldots,n \tag{4.12}$$

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 polynomial of order k. 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) since $\rho = (\mu-1)/(r-1)$. 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, it is clear that 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,

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

Utilizing the well known fact that the arithmetic mean of any collection of numbers, in particualar $\{1, r, \ldots, r^{k-1}\}$, 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 (i.e., negative first derivative), and correspondingly, f(1) an increasing function of r satisfying,

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

This statement holds for non-integral k as well, and can be proved by a more careful analysis of t'(r). Also, it is clear that for given r > 1, t converges to 0 as k increases, therefore,

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

Lower bounds for $\mu_k'(x)$ can be developed by utilizing a generalization of (2.33) known as Hölder's inequality, [6], [7], which states that for given a_i, b_i, i=1, ...,n,

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

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 defined equal to its limiting value as $q \to \infty$,

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

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

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

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

$$\mu^k \leq \mu_k', \tag{4.22}$$

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

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

As it is currently stated, Theorem 2 is not applicable to all distributions of a discrete positive bounded random variable (r.v.). 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,

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

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

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

$$n = \sum n_j .$$

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, since 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 r.v. $X \in [a,b]$, $a>0$, and define,

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

$$\mu = \mu'_1 \tag{4.24}$$

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

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

<u>Theorem 3</u> Let f(x) be a p.d.f. of a discrete r.v. defined on $[1,r]$. Then,

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

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

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 $\varepsilon > 0$, define $g_\varepsilon(x_i)$, $i=1,\ldots,m$, so that $g_\varepsilon(x_i)$ is rational and,

$$\left| f(x_i) - g_\varepsilon(x_i) \right| \leq \varepsilon / (mr^k), \qquad \varepsilon > 0 \tag{4.28}$$

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

If $(f(x_i), \ldots, f(x_m))$ is identified with a point $y \in \mathbb{R}^m$ on the hyperplane defined by $\sum y_i = 1$, 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 $\mathbb{R}^m$, i.e. that rational points are dense in $\mathbb{R}^m$ [3].

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

$$\left| \mu_k'(f) - \mu_k'(g_\varepsilon) \right| \leq \varepsilon \quad , \quad \varepsilon > 0 \,. \tag{4.30}$$

However, the construction in (4.23) shows that Theorem 2 and (4.22) can be applied to $\mu_k'(g_\varepsilon)$ and (4.26) is satisfied. Since 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 $\varepsilon > 0$, there is an integer N such that,

$$\sum_{i=N+1} x_i^k f(x_i) < \varepsilon \quad , \qquad k \text{ fixed}, \tag{4.31}$$

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

Let $h_N(x)$ be a p.d.f. 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),

$$\left| \mu_k'(f) - \mu_k'(h_N) \right| \leq \frac{\varepsilon}{1-\varepsilon} \left[\mu_k'(f) + 1 \right]. \tag{4.33}$$

Hence, since (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. ∎

Since (4.3) and (4.4) are valid in general, Theorem 3 can be applied to any p.d.f. $f(x)$ of a discrete r.v. defined on $[a,b]$, $a > 0$.

Corollary Let $f(x)$ be a p.d.f. of a discrete r.v. 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,

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

Further, the inequalities in (4.36) are sharp.

proof Rewriting (4.35) as:

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

(4.36) follows directly from (4.26). For $1 \leq \mu \leq r$, if the point mass p.d.f. $f_\mu(x)$ 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 $1 \leq \mu \leq r$, 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,

$$\begin{aligned} \mu(g_\mu) &= \mu, \\ \mu'_k(g_\mu) &= \frac{r-\mu}{r-1} + \frac{\mu-1}{r-1}\, r^k, \qquad k \geq 1, \end{aligned} \tag{4.39}$$

and a calculation shows that,

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

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 p.d.f. defined on $[1,r]$ and μ'_k, μ and R_k defined analogously to (4.24) and (4.25), with,

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

Theorem 4 Let $f(x)$ be a continuous p.d.f. 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:

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

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

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

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 p.d.f. defined on the partition $\{x_i\}$ given in (5.4) by,

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

Applying Theorem 3 to g_n,

$$\mu^k(g_n) \leq \mu'_k(g_n) \leq 1 + (\frac{r^k-1}{r-1})\ (\mu(g_n) - 1). \qquad (5.8)$$

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

<u>Corollary</u> Let f(x) be given as in Theorem 4 and let $M_X(t)$ denote the moment generating function of x,

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

Then,

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

Further the inequalities in (5.10) are sharp.

<u>proof</u> The inequalities in (5.10) follow directly from (5.2) and (4.37). Also, the fact that they are sharp follows by considering continous 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,

$$\mu = (\frac{r^k-1}{k(r-1)})^{1/k}. \qquad (5.11)$$

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, $k \to \infty$.

The author gratefully acknowledges the following two references provided by Piet de Jong and Pat Brockett which contain related material:

Carlin and Studden, Tchebychev Systems and Applications, SIAM
Kendall and Stuart, Advanced Theory of Statistics, Volume I.

Bibliography

1. N.L. Bowers, H.U. Gerber, J.C. Hickman, D.A. Jones, and C.J. Nesbitt, Part 5A Study Note: Risk Theory, Society of Actuaries, Chicago, 1982, Ch. 2,3.

2. R. Courant, and F. John, Introduction to Calculus and Analysis, Vol. I, Interscience Publishers, New York, 1965, pp. 15, 464-7.

3. J. Dugundji, Topology, Allyn and Bacon, Inc. Boston, 1970, p. 72.

4. W. Feller, An Introduction to Probability Theory and Its Applications, 3rd Ed., Vol. I, John Wiley & Sons, Inc., New York, 1968, Ch.X.

5. S.A. Klugman, On the Variance and Mean Squared Error of Decrement Estimators, TSA, Vol. XXXIII, Society of Actuaries, Chicago, 1982.

6. F. Riesz, B. Sz-Nagy, Functional Analysis, 2nd Ed., Frederick Ungar Publishing Co., New York, 1972, Sec. 21.

7. W. Rudin, Real and Complex Analysis, 2nd Ed., McGraw-Hill Series in Higher Mathematics, McGraw-Hill, Inc. New York, 1974, Ch. 3.

8. TSA, Reports Numbers, Society of Actuaries, Chicago, 1948-1980.