ABSTRACT

Tchebycheff Type Inequalities

Consider a random variable X, defined on a probability space $(\Omega, \mathcal{A}, P)$, whose distribution is not known completely. With probability one we can say that, for a set $A \in \mathcal{A}$, $0 \leq P(X \in A) \leq 1$. This probability statement does not offer us any significant information in terms of the upper or lower bounds on $P(X \in A)$. Often, a limited amount of information about the distribution of a random variable X is available. This limited amount of information may sometimes enable us to make probability statements on the unknown random variable. When bounds are determined from the available information, the inequalities thus formed, which offer us upper and lower probability bounds for $P(X \in A)$, are known as Tchebycheff type inequalities. This thesis contains a study of Tchebycheff type inequalities. In the thesis we trace the development of a general theorem which will provide us with a method of obtaining sharp Tchebycheff type inequalities for restricted and unrestricted random variables in R_1. We use this general theorem as our background when we discuss the development of a multivariate inequality for different types of regions in R_n. We shall also illustrate various methods of obtaining probability bounds for sums of random variables.

Morty Yalovsky

Department of Mathematics

M.Sc.

TCHEBYCHEFF TYPE INEQUALITIES

Morty Yalovsky

TCHEBYCHEFF TYPE INEQUALITIES

by

Morty Yalovsky

A thesis submitted to the Faculty of
Graduate Studies and Research in partial
fulfilment of the requirements for the
degree of Master of Science.

Department of Mathematics,
McGill University,
Montreal.

August 1968.

I would like to express my gratitude to Dr. V. Seshadri for his encouragement and guidance both as a teacher and as an advisor.

I would like to thank Dr. A. M. Mathai for his suggestion of the topic and for his initial guidance.

I would also like to thank the NRC for their financial support.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION AND NOTATION

Consider a random variable X, defined on a probability space $(\Omega, \mathcal{O}\!\iota, P)$, whose distribution is not known completely. With probability one we can say that, for a set $A \in \mathcal{O}\!\iota$, $0 \leq P(X \in A) \leq 1$. This probability statement does not really offer us any significant information in terms of the upper or lower bounds on $P(X \in A)$. If X is a random variable with mean μ and variance σ^2 we may intuitively expect the variance to have some influence on the distribution of X. If μ and σ^2 are known, to what extent can we make a probability statement about the distribution when the functional form of the distribution function F(x) is not known? Such a probability statement which gives us bounds on the probability of the deviation of the random variable X from its mean value μ in terms of the standard deviation σ was first given by Bienaymé [13] in 1853. Tchebycheff [113] independently obtained the same result in 1867 and Pizetti [95] did likewise in 1892.

Most often, only a limited amount of information about the distribution of a random variable is available. Sometimes this information consists of expected values of functions of the random variable, for example, moments, cumulants, etc. At other times a general information about the shape of the distribution is available, for example, f(x) is monotone, unimodal, etc. This limited amount of information may sometimes enable us to make probability statements on the unknown density function. When bounds are determined from the available information, the inequalities thus formed, which offer us upper and lower probability bounds for the unknown function, are known as Tchebycheff type inequalities.

Historically we note that Tchebycheff's original result was obtained by simple algebraic calculations without using any approximation or any calculus. It was stated as follows and it will be proved in chapter two by means of a general method.

Theorem 1.1.1 [110]. Let a, b, c,... be the mathematical expectations of the quantities X, Y, Z,... and let a_1, b_1, c_1,... be the mathematical expectations of their squares X^2, Y^2, Z^2,... . The probability that the sum X + Y + Z + ... is included within the limits

$$a + b + c + \dots + \alpha(a_1 + b_1 + c_1 + \dots - a^2 - b^2 - c^2 - \dots)^{\frac{1}{2}}$$

and (1.1.1)

$$a + b + c + \dots - \alpha(a_1 + b_1 + c_1 + \dots - a^2 - b^2 - c^2 - \dots)^{\frac{1}{2}}$$

will always be larger than $1 - 1/\alpha^2$, regardless of the size of $\alpha (\alpha \neq 0)$.

Remark: Though α is chosen arbitrarily, no useful information can be obtained for $\alpha \leq 1$.

The first inequality which fell into the category of Tchebycheff type inequalities was that one conjectured by Gauss [31] in 1821 and only later proved by Winckler [124] in 1866. If X is a continuous random variable with a unimodal distribution whose mean is μ and variance is σ^2, then, for any $k > 0$,

$$P(|X - x_0| \geqslant kt) \leqslant \begin{cases} 1 - k/(3)^{\frac{1}{2}}, & k \leqslant 2/(3)^{\frac{1}{2}}, \quad (1.1.2) \\ 4/9k^2, & k \geqslant 2/(3)^{\frac{1}{2}}, \quad (1.1.3) \end{cases}$$

where x_0 is the mode and $t^2 = \sigma^2 + (x_0 - \mu)^2$ is the second moment about the mode. If we now apply the Pearson measure of skewness, $s = (\mu - x_0)/\sigma$, to (1.1.3), then, for all $k > |s|$,

$$P(|X - u| \geqslant k\sigma) \leqslant \frac{4(1 + s^2)}{9(k - |s|)^2} . \quad (1.1.4)$$

Because of the additional information about the distribution, i. e.

unimodality, for small /s/, (1.1.4) offers a sharper bound than the bound obtained in the case of a single random variable from Theorem 1.1.1.

Tchebycheff type inequalities can be used in industrial situations such as quality control, in setting up tests of hypothesis and in setting confidence intervals about a given point. These inequalities are only used when one does not have complete knowledge of the distribution of the random variable. Among its many applications in probability theory it is explicitly used to prove the weak law of large numbers and Bernoulli's Theorem.

There are several expository papers, e.g. Godwin [33], Isii [53], Savage [101], and books, e.g. Godwin [34], Walsh [121], Karlin and Studden [60], Savage [102], concerning these types of inequalities. In this expository work we shall attempt to restrict ourselves to the more recent results without neglecting some of the earlier results. We shall show how some newly formulated general theorems unite some of the previous work done thus enabling us to obtain some of the earlier classical results. We shall review some work, both old and new, which has never been discussed in any of the above expositions. We shall construct examples to illustrate the sharpness of Theorem 2.1.2 and the use of Theorem 2.3.1; we shall give some numerical comparisons for unimodal distributions; we shall partially answer a question raised by Mudholkar and Rao [88] and we shall illustrate an application of Tchebycheff type inequalities to medical problems. Throughout the paper we shall concentrate on developing sharp upper probability bounds.

In chapter two we shall concentrate on the development of a general theorem and we shall show how this theorem enables us to obtain certain univariate inequalities under general conditions.

Chapter three will offer univariate inequalities for distribution which are subject to some restrictions.

In chapter four we shall discuss the development of a multivariate inequality for different types of regions in R_n.

Chapter five deals with inequalities for sums of random variables; some classical limit theorems will be introduced but not emphasized.

In chapter six we offer some applications of these inequalities and in general we discuss their usefulness in practical situations.

The bibliography lists only those papers which have been directly dealt with or referred to in this thesis. For an exhaustive bibliography of Tchebycheff and related Tchebycheff type inequalities, the reader can refer to either Savage's "Bibliography of Nonparametric Statistics" [102] or the bibliography in Savage's paper [101].

Notation

Throughout this paper we shall be dealing with a probability space $(\Omega, \mathfrak{A}, P)$ and random variables or random vectors defined on this probability space.

Notation	Definition
r.v.	random variable
R_{+n}	The Positive Orthant of R_n
μ	$E(X)$
μ'_n	$E(X^n)$
$\mu_n(a)$	$E(X - a)^n$
μ_n	$E(X - \mu)^n$
σ^2	$E(X - \mu)^2$

Notation	Definition
v'_n	$E/X/^n$
$v_n(a)$	$E/X - a/^n$
v_n	$E/X - \mu/^n$
v	$E/X - \mu/$
v_{ni}	$E/X_i - \mu_i/^n$
S_n	$\sum_{i=1}^{n} X_i$
s_n^2	$\sum_{i=1}^{n} \sigma_i^2$
B_n^3	$\sum_{i=1}^{n} v_{3i}$
I	Identity Matrix
i.i.d.	Independent Identically Distributed
ρ_{ij}	$\frac{\sigma_{ij}}{\sigma_i \sigma_j}$
ρ	Correlation Coefficient Between Two Random Variables
When A is a set, A'	The Complement of A
I_A	The Indicator Function of the Set A
When A is a Matrix, A'	The Transpose of A
When A is a Matrix, A	The Determinant of A
When A is a Matrix, trA	The Trace of A
When a is a Vector, a'	The Transpose of a
When $a = (a_1, \ldots, a_n)$	
$D_a = \text{diag}(a_1, \ldots, a_n)$	diag is a Diagonal Matrix
The Vector e	$e = (1, 1, \ldots, 1)$

In this paper we often choose $\mu = 0$. This can be done without any loss of generality since a distribution with mean μ can be transformed to a distribution with mean zero by means of a linear transformation.

In the earlier sections of this paper proofs will be offered for continuous random variables. Analogous proofs for the discrete r.v. can be constructed by the reader if he wishes.

A bound will be called sharp if it is the best possible bound that can be obtained for the situation. In most cases we shall exhibit the sharpness of an inequality through a distribution which attains equality in the inequality.

CHAPTER TWO: UNIVARIATE INEQUALITIES UNDER GENERAL CONDITIONS

2.1 Introduction

If X is a r.v. and if u(x) is an arbitrary real valued function of the real number x, then Y = u(X) is also a r.v. With this in mind we develop Tchebycheff's inequality.

Theorem 2.1.1 (Markov). Let u(X) be a nonnegative function of a r.v. X(discrete or continuous). If $E[u(X)]$ exists, then, for any arbitrary constant $K > 0$,

$$P(u(X) \geqslant K) \leqslant E[u(X)]/K. \qquad (2.1.1)$$

Remark: The above theorem will be referred to throughout this paper as Markov's inequality.

Proof: Let X be a continuous r.v. with distribution function F(x). Let $A = \{x: u(x) \geqslant K\}$.

$$E[u(X)] = \int_0^\infty u(x)dF(x) = \int_A u(x)dF(x) + \int_{A'} u(x)dF(x)$$

$$\geqslant \int_A u(x)dF(x) \geqslant K\int_A dF(x) = KP(u(X) \geqslant K).$$

Therefore $P(u(X) \geqslant K) \leqslant E[u(X)]/K$. A similar proof can be given when X is a discrete valued r.v.

Corollary. If X is a r.v. (discrete or continuous) whose mean μ and variance σ^2 exist, then, for any constant $k > 0$,

$$P(|X - \mu| \geqslant k\sigma) \leqslant 1/k^2. \qquad (2.1.2)$$

Proof: Let $u(X) = (X - \mu)^2 \geqslant 0$ and let $K = k^2\sigma^2$.

$$(2.1.1) \Rightarrow P((X - \mu)^2 \geqslant k^2\sigma^2) \leqslant 1/k^2$$

$$P(|X - \mu| \geqslant k\sigma) \leqslant 1/k^2.$$

Equivalently we can write

$$P(|X - \mu| < k\sigma) \geqslant 1 - 1/k^2. \qquad (2.1.3)$$

Inequality (2.1.2) is known as the Tchebycheff inequality. For a suitably chosen X the inequality is the best possible.

Example. Let X be a discrete valued r.v. such that

$$P(X = k\sigma) = P(X = -k\sigma) = 1/2k^2,\ P(X = 0) = 1 - 1/k^2.$$

$E(X) = 0$, $Var(X) = \sigma^2$, and by (2.1.2)

$$P(/X - \mu/ \geqslant k\sigma) = P\ (/X/ \geqslant k\sigma) \leqslant 1/k^2.$$

Also, $P(/X/ \geqslant k\sigma) = P(X = -k\sigma) + P(X = k\sigma) = 1/k^2$.

Thus equality is attained for this example. This implies that Tchebycheff's inequality is sharp.

In chapter one it was indicated that for a unimodal distribution a sharper inequality could be obtained. A much simpler restriction than unimodality, namely that of boundedness of the r.v., will give us an inequality which is sharper than (2.1.2).

Theorem 2.1.2. Let X be a r.v. with mean μ, variance σ^2 and let $/X - \mu/ \leqslant Z$. Then, for $0 < k < 1$,

$$P(/X - \mu/ \leqslant k\sigma) \leqslant 1 - \frac{\sigma^2(1 - k^2)}{Z^2 - k^2\sigma^2} \qquad (2.1.4)$$

Proof: $Z^2 - k^2\sigma^2 - (X - \mu)^2 + k^2\sigma^2 \geqslant 0$.

By Theorem 2.1.1 we have, for an arbitrary $K > 0$,

$$P(Z^2 - k^2\sigma^2 - (X - \mu)^2 + k^2\sigma^2 \geqslant K) \leqslant$$
$$E(Z^2 - k^2\sigma^2 - (X - \mu)^2 + k^2\sigma^2)/K .$$

Let $K = Z^2 - k^2\sigma^2 > 0$.

$$P(/X - \mu/ \leqslant k\sigma) \leqslant 1 - \frac{\sigma^2(1 - k^2)}{Z^2 - k^2\sigma^2} \ .$$

Example: (2.1.3) is sharp. Consider the r.v. X distributed as follows:

$$P(X = -\sigma/k) = P(X = \sigma/k) = k^2/2,$$
$$P(X = 0) = 1 - k^2.$$

$E(X) = 0$, $Var(X) = \sigma^2$. Let $Z = \sigma^2\ (1/k^2 - 1 + k^2)$. Since $0 < k < 1$,

$$Z^2 \geqslant \sigma^2(1/k^2 - 1) = \sigma^2(1 - k^2)/k^2 \geqslant \sigma^2/k^2.$$

Thus $Z \geqslant / \frac{\sigma}{k} /$. By Theorem 2.1.2 we have

$$P(/X/ \leqslant k\sigma) \leqslant 1 - \frac{\sigma^2(1 - k^2)}{\sigma^2(1/k^2 - 1)} = 1 - k^2.$$

However, $P(/X/ \leqslant k\sigma) = P(X = 0) = 1 - k^2$. Thus equality is attained for this example.

Remark: Had we considered the function $Z^2 - (X - \mu)^2 + k^2\sigma^2 > 0$, we would have obtained Lurquin's result [75], namely,

$$P(/X - \mu/ \leqslant k\sigma) \leqslant 1 - \frac{\sigma^2(1 - k^2)}{Z^2}. \tag{2.1.5}$$

We now consider another restriction on the r.v. X. The following theorem will be used to prove some results in section 5.6 of this paper.

Theorem 2.1.3. If $u(X)$ is a nonnegative function of the r.v. X such that $u(x) \geqslant b$ whenever $x \geqslant a$, then

$$P(X \geqslant a) \leqslant E[u(X)]/b. \tag{2.1.6}$$

Proof: $E[u(X)] = \int_{-\infty}^{\infty} u(x)dF(x) \geqslant \int_a^{\infty} u(x)dF(x)$

$$\geqslant b\int_a^{\infty} dF(x) = bP(X \geqslant a).$$

We now consider the one-sided Tchebycheff inequalities as offered by Cantelli [21].

Theorem 2.1.4. If X is a r.v. with 0 mean and variance σ^2, then

$$P(X \leqslant k) \leqslant \frac{\sigma^2}{\sigma^2 + k^2}, \quad k < 0, \tag{2.1.7}$$

$$P(X \geqslant k) \leqslant \frac{\sigma^2}{\sigma^2 + k^2}, \quad k > 0. \tag{2.1.8}$$

Proof: The proof will be given for the continuous case. $E(X - k) = -k$, $E(X - k)^2 = \sigma^2 + k^2$. Now, for $k < 0$,

$$k^2 \leqslant \left(\int_k^{\infty} (x - k)f(x)dx\right)^2 \leqslant \int_k^{\infty} f(x)dx \int_k^{\infty} (x - k)^2 f(x)dx$$

$$k^2 \leq (1 - P(X \leq k))(k^2 + \sigma^2).$$

Thus (2.1.7) is proven. For a discrete r.v. a similar proof can be given. To prove (2.1.8) we shall use a method based on Theorem 2.1.1. Let $u(X) = (X + c)^2$ be a nonnegative r.v. for $c > 0$.

$$(x + c)^2 \geq (k + c)^2 \ , \ x \geq k > 0.$$

$$P(X \geq k) \leq \frac{E(X + c)^2}{(k + c)^2} . \qquad (2.1.9)$$

Minimizing the right hand side of (2.1.9) with respect to c, we get $c = \sigma^2/k$. Substituting this value of c in (2.1.9) gives us (2.1.8).

Under the given conditions,(2.1.7) and (2.1.8) give the sharpest possible bounds. As an example consider a discrete r.v. taking on two values.

$$P(X = k) = \frac{\sigma^2}{\sigma^2 + k^2} , \qquad P(X = -\sigma^2/k) = \frac{k^2}{\sigma^2 + k^2} .$$

$E(X) = 0$, $Var(X) = \sigma^2$. For $k < 0$, (2.1.7) tells us that

$$P(X \leq k) \leq \frac{\sigma^2}{\sigma^2 + k^2} .$$

However, by our example, $P(X \leq k) = P(X = k) = \frac{\sigma^2}{\sigma^2 + k^2}$.

For $k > 0$, (2.1.8) tells us that

$$P(X \geq k) \leq \frac{\sigma^2}{\sigma^2 + k^2} .$$

However, $P(X \geq k) = P(X = k) = \frac{\sigma^2}{\sigma^2 + k^2}$. Thus (2.1.7) and (2.1.8) are sharp.

2.2 Moments and Absolute Moments of Order r, $r > 2$.

We now extend Markov's inequality to moments of higher order.

Theorem 2.2.1. If X is a r.v. such that $E|X|^r$ exists for $r > 1$, then, for any $k > 0$,

$$P(|X| \geq k) \leq \frac{E|X|^r}{k^r}. \tag{2.2.1}$$

Proof: $$E|X|^r = \int_{-\infty}^{\infty} |x|^r dF(x) = \int_{|x|<k} |x|^r dF(x) + \int_{|x|\geq k} |x|^r dF(x)$$

$$\geq \int_{|x|\geq k} |x|^r dF(x) \geq k^r P(|X| \geq k).$$

This completes proof.

Example: Whenever $k \geq (E|X|^r)^{1/r}$, inequality (2.2.1) is sharp. Let us define the distribution of the r.v. X as follows:

$$P(X = k) = \frac{E|X|^r}{k^r},\ P(X = 0) = 1 - \frac{E|X|^r}{k^r}.$$

By Theorem 2.2.1 $P(|X| \geq k) \leq (E|X|^r)/k^r$. However, $P(|X| \geq k) = P(X = k) = (E|X|^r)/k^r$. Thus (2.2.1) is sharp.

Corollary 1. Consider the r.v. $|X - \mu|$ such that $E|X - \mu|^r$ exists and let $k = t(E|X - \mu|^r)^{1/r}$. Then

$$P(|X - \mu| \geq t(E|X - \mu|^r)^{1/r}) \leq 1/t^r. \tag{2.2.2}$$

This result was obtained by Lurquin [71,72] and by Guldberg [37].

Corollary 2. If $r = 2s$, $s = 1,\ldots,n$, then, for the r.v. $|X - \mu|$ such that $E|X - \mu|^{2s} = \mu_{2s}$ exists, and for $k = t(E|X - \mu|^2)^{\frac{1}{2}}$,

$$P(|X - \mu| \geq t\sigma) \leq \frac{\mu_{2s}}{t^{2s}\mu_2^{\ s}}. \tag{2.2.3}$$

This is Pearson's inequality [93]. Though this result is an improvement over (2.1.2), it still lacks the necessary precision which would make it useful in practical situations requiring statistical analysis.

Corollary 3. For a r.v. $|X - \mu|$ such that $E|X - \mu|^n = \nu_n$ exists and ν_r

exists, and for $k = tv_r^{1/r}$,

$$P(|X - \mu| \geq tv_r^{1/r}) \leq \frac{1}{t^n}\left(\frac{v_n^{1/n}}{v_r^{1/r}}\right)^n . \qquad (2.2.4)$$

This result was obtained by Guldberg [39,40].

Corollary 4. If X is a nonnegative r.v. such that μ'_n exists, then, for $k = t(\mu'_n)^{1/n}$,

$$P(X \geq t(\mu'_n)^{1/n}) \leq 1/t^n . \qquad (2.2.5)$$

This result is due to Guldberg [38].

Let X be a r.v. such that for any arbitrary point $x_o, v_r(x_o) = E|X - x_o|^r$, $r > 1$, exists. By Theorem 2.2.1,

$$P(|X - x_o| \geq k) \leq \frac{v_r(x_o)}{k^r} \qquad (2.2.6)$$

cannot be improved upon for $k \geq (v_r(x_o))^{1/r}$. When $x_o = \mu$, (2.2.4) indicates that an improved bound is possible if two absolute moments about the mean are known; however, (2.2.4) is not sharp. It would appear that if more than two absolute moments about an arbitrary point are known, a sharper inequality than (2.2.4) can be obtained. Wald [119,120] offers a method by which discrete distributions, whose values are all positive and whose absolute moments about the origin are equal to the absolute moments about an arbitrary point of the unknown r.v., can be constructed. Probability bounds are then obtained on these constructed discrete distributions. The restriction to nonnegative random variables is possible for, if $Y = |X - x_o|$, then $EY^r = E|X - x_o|^r$ and $P(Y \geq k) = P(|X - x_o| \geq k)$.

If $v_1(x_o), \ldots, v_n(x_o)$ of the unknown r.v. are available, then for n even the spectrum of the newly constructed distribution will consist of either the point k and n/2 other points or the points 0, k, ∞, and $\frac{1}{2}(n - 2)$ other points. Zero probability is given to all points not in the spectrum of the discrete distribution and an infinitely small probability is given to the point at infinity such that this probability only affects the nth order absolute moment . A discussion of the concept of placing a probability at the point ∞ can be found in Royden [98]. For n odd, the distribution will consist of the points 0, k and $\frac{1}{2}(n - 1)$ other points. Wald shows that the discrete distribution which is constructed is unique in each situation. The sharp lower probability bound of $P(/X - x_o/ < k)$ is obtained by summing all the probabilities corresponding to the points on the left of the point k, and the sharp upper bound of $P(/X - x_o/ < k)$ is equal to the lower bound plus the probability at the point k.

Let us illustrate the use of this method in the case where two absolute moments are given. In order that the numbers v_r, v_s, $s > r$, be realized as absolute moments of a distribution, we must have [120]

$$v_r > 0, \; v_s \geqslant (v_r)^{s/r} \; . \tag{2.2.7}$$

Since two absolute moments are given, we must construct discrete distributions whose spectra consist of either

I) the points k and λ with respective probabilities p and 1 - p

or

II) the points k, 0 and ∞ with respective probabilities p, 1 - p and δ (δ is infinitely small and only affects v_s).

From I), $pk^r + (1 - p)\lambda^r = v_r$, $\lambda^r = (v_r - pk^r)/(1 - p)$

$$pk^s + (1 - p)\lambda^s = v_s \; , \quad \lambda^s = (v_s - pk^s)/(1 - p)$$

Therefore $$\left(\frac{v_r - pk^r}{1 - p}\right)^{s/r} = \frac{(v_s - pk^s)}{1 - p} \quad . \tag{2.2.8}$$

From II), $$pk^r = v_r, \quad k^r = v_r/p \tag{2.2.9}$$

$$pk^s \leq v_s, \quad p(v_r/p)^{s/r} \leq v_s \quad . \tag{2.2.10}$$

Let $s = 2r$. From (2.2.8)

$$1 \geq p = \frac{v_{2r} - v_r^2}{v_{2r} - 2v_r k^r + k^{2r}}$$

$$= \frac{v_{2r} - v_r^2}{v_{2r} - v_r^2 + (k^r - v_r)^2} \geq 0, \tag{2.2.11}$$

by(2.2.7). Also $$\lambda^r = \frac{v_r k^r - v_{2r}}{k^r - v_r} \geq 0. \tag{2.2.12}$$

Thus from (2.2.7) either $k^r \leq v_r$ or $v_r k^r \geq v_{2r}$ and $k^r \geq v_r$. From (2.2.9) we require that $k^r \geq v_r$ such that $p \leq 1$. If $k^r \geq v_{2r}/v_r$ and $k^r \geq v_r$, then, from (2.2.12) and I) we have

$$\lambda^r = (k^r - v_{2r}/v_r)/(k^r/v_r - 1).$$

Thus $\lambda \leq k$ and by Wald's method of obtaining upper and lower bounds

$$1 - p \leq P(/X - \mu/ < k) \leq 1, \tag{2.2.13}$$

where p is defined by (2.2.11). Similarly, if $k^r \leq v_r$, then (2.2.12) and I) tell us that $\lambda \geq k$ and

$$0 \leq P(/X - \mu/ < k) \leq p. \tag{2.2.14}$$

If $v_r \leq k^r \leq v_{2r}/v_r$, then from (2.2.9) and II)

$$1 - v_r/k^r \leq P(/X - \mu/ < k) \leq 1. \tag{2.2.15}$$

The above inequalities were first proven by Cantelli [21]. If we let $k = t\sigma$, $r = 1$ and $\Delta = v_1/\sigma$, then (2.2.13) gives us Peek's inequality [94] as a special case;

$$P(|X - \mu| \geq t\sigma) \leq \frac{1 - \Delta^2}{t^2 - 2t\Delta + 1}, \quad t \geq \Delta. \tag{2.2.16}$$

(2.2.16) is an improvement over (2.1.2) in the case where v_1 is known.

Equation (2.2.7) has stated the conditions which must be satisfied such that v_r and v_s can be absolute moments of a distribution. In general, what conditions must be satisfied in order that a set of numbers can be realized as moments of a distribution? The answer is given by the Hamburger moment problem [106] which states that if

$$\begin{vmatrix} 1 & \mu'_1 & \cdots & \mu'_n \\ \mu'_1 & \mu'_2 & \cdots & \mu'_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu'_n & \mu'_{n+1} & \cdots & \mu'_{2n} \end{vmatrix} \geq 0 , \tag{2.2.17}$$

i.e. nonnegative definite, then the set of moments will determine a probability distribution. If the determinant is strictly positive definite, then there will exist many distributions which are solutions of the moment problem and the set of moments is called non-degenerate. Shohat and Tamarkin [106] have added several other conditions which must be satisfied in order that the set of moments in (2.2.17) will offer a unique probability distribution. In cases when either $\mu'_1, \ldots, \mu'_n$ or $\mu'_1, \ldots, \mu'_{2n}$ are given and satisfy the moment inequality,

Shohat and Tamarkin have given methods of constructing discrete distributions whose moments are equal to those of the unknown distribution. Their method of constructing the discrete distributions and of obtaining the upper and lower bounds is similar to Wald's method. Throughout the rest of this paper we shall assume that for any given set of moments the conditions for the solution of the Hamburger moment problem are satisfied.

If $F(x)$ and $G(x)$ are two distribution functions determined from the same finite set of moments, what can be said about $|F(x) - G(x)|$? Isii [52] and Khamis [63,64] have both offered bounds on the above expression under very general conditions; they have sharpened their results by imposing certain restrictions on the r.v. X and on $f(x)$ and $g(x)$.

2.3 A General Theorem

Theorems 2.1.1 and 2.2.1 have offered us general methods of obtaining some non-trivial Tchebycheff type inequalities which had previously been obtained by various other methods. In 1959 Isii [54] proposed a generalized method of obtaining bounds for $P(X \in A)$ in terms of a given set of moments, where $A \in \mathcal{O}l$ is a closed set in $\Omega = R_1$. This method offered us a unified way of obtaining many of the earlier results which were based on a knowledge of moments. Let I_A be the indicator function of some set $A \in \mathcal{O}l$, let $a = (a_0, \ldots, a_n) \in R_{n+1}$ (a real valued vector) and let f_i, $i = 1, \ldots, n$, be Borel measurable functions from Ω into R_1 such that $f = a_0 + a_1 f_1 + \ldots + a_n f_n$ and $\int f dP$ exists.

When $f \geq I_A$ on Ω, we have

$$P(X \in A) = \int I_A dP \leq \int f dP = a_0 + a_1 M_1 + \ldots + a_n M_n, \tag{2.3.1}$$

for all probability distributions P which realize $M_1, \ldots, M_n$ as moments for $f_1, \ldots, f_n$.

$$P(X \in A) \leq \inf_{a \in C}(a_0 + a_1 M_1 + \dots + a_n M_n), \tag{2.3.2}$$

where $C = \left\{a : a_0 + a_1 f_1 + \dots + a_n f_n \geq I_A \text{ on } \Omega\right\}$.

Similarly, $$P(X \in A) \geq \sup_{a \in D}(a_0 + a_1 M_1 + \dots + a_n M_n), \tag{2.3.3}$$

where $D = \left\{a : a_0 + a_1 f_1 + \dots + a_n f_n \leq I_A \text{ on } \Omega\right\}$. We note that for $f_i = x^i$, $M_i = \mu'_i$. Are inequalities (2.3.2) and (2.3.3) sharp? Isii [54,55,56] has considered the sharpness of these inequalities under the restricted conditions of $\Omega = (-\infty, \infty)$, $\Omega = [0, \infty)$ and $\Omega = [0,1]$, respectively, where the corresponding f_i have respectively been defined by $f_i = x^i$, $f_i = |x|^i$ and f_i = any given function; sharp probability inequalities are obtained through the use of linear functionals on Banach spaces.

In 1963 Isii [57] unified the previous work and offered a general theorem for all abstract spaces Ω. The proof of his theorem is based on the theory of convex sets and the separating hyperplane theorem; a specific case of the theorem was independently obtained by Karlin & Studden [60].

Let $(\Omega, \mathcal{Q})$ be a measurable space and $\mathcal{P}$ be a family of nonnegative measures defined on $(\Omega, \mathcal{Q})$; for $\omega \in \Omega$, let $f(\omega) = (f_0(\omega), \dots, f_n(\omega))$ be a Borel measurable function from Ω to R_{n+1} such that

$$\int |f(\omega)| dP(\omega) < \infty, \text{ for all } P \in \mathcal{P},$$

and let $g(\omega)$ be a real valued measurable function such that

$$\int |g(\omega)| dP(\omega) < \infty, \text{ for all } P \in \mathcal{P}.$$

We define $\mathcal{P}(M) = \left\{P : P \in \mathcal{P}, \int f(\omega) dP(\omega) = M, \text{ for all } M = (M_0, \dots, M_n) \in R_{n+1}\right\}$.

We then have the following theorem.

Theorem 2.3.1 [57]. If $\mathcal{P}$ is convex, $\mathcal{M} = \left\{\int f(\omega) dP(\omega) : P \in \mathcal{P}\right\}$ is a convex set in R_{n+1}. If $M = (M_0, \ldots, M_n)$ is an interior point of $\mathcal{M}$, then

$$U(g;M) = \sup_{P \in \mathcal{P}(M)} \int g(\omega)dP(\omega)$$

$$= \inf_{a_0,\ldots,a_n} \left(\sum_{i=0}^{n} a_i M_i + \sup_{P \in \mathcal{P}} \int (g - \sum_{i=0}^{n} a_i f_i)dP\right). \qquad (2.3.4)$$

Further, if $U(g;M) < \infty$, there exist real constants $a_0, \ldots, a_n$ that attain the inf in the right hand side of (2.3.4).

Proof: Since $\mathcal{P}$ is convex, $\mathcal{M}$ is a convex set in R_{n+1}[67]. For $a \in R_{n+1}$

$$\sup_{P \in \mathcal{P}} \int (g - \sum_{i=0}^{n} a_i f_i)dP \geq \sup_{P \in \mathcal{P}(M)} \int (g - \sum_{i=0}^{n} a_i f_i)dP$$

$$= U(g;M) - \sum_{i=0}^{n} a_i M_i .$$

Therefore $$U(g;M) \leq \inf_{a_0,\ldots,a_n} \left(\sum_{i=0}^{n} a_i M_i + \sup_{P \in \mathcal{P}} \int (g - \sum_{i=0}^{n} a_i f_i)dP\right). \qquad (2.3.5)$$

We must now show that the left hand side of (2.3.4) is greater than the right hand side of (2.3.4). The set $\mathcal{N} = \{(\int f_0 dP, \ldots, \int f_n dP, \int g dP) : P \in \mathcal{P}\}$ is also a convex set in R_{n+2}[67]. Let $\bar{\mathcal{N}}$ denote the closure of $\mathcal{N}$. For $U(g;M) < \infty$, $Z = (M_0, \ldots, M_n, U(g;M))$ is a boundary point of $\bar{\mathcal{N}}$. Since $\bar{\mathcal{N}}$ is a closed convex set, we can construct a supporting hyperplane at Z. Let the equation of the hyperplane be

$$\sum_{i=0}^{n+1} \alpha_i x_i + \beta = 0.$$

Since the projection of $\mathcal{N}$ on R_{n+1} is the point M (an interior point of $\mathcal{M}$), $\alpha_{n+1} \neq 0$ and the equation of the hyperplane becomes

$$x_{n+1} = \sum_{i=0}^{n} a_i x_i + k, \qquad (2.3.6)$$

where $a_i = -\alpha_i/\alpha_{n+1}$, $i = 1, \ldots, n$, and $k = -\beta/\alpha_{n+1}$. For every interior point of $\mathcal{N}$, the definition of $\mathcal{N}$ says that

$$x_{n+1} \leq \sum_{i=0}^{n} a_i x_i + k.$$

Therefore $\int g dP \leq k + \sum_{i=0}^{n} a_i \int f_i dP$, for all $P \in \mathcal{P}$.

$$\therefore \quad k \geq \int g dP - \sum_{i=0}^{n} a_i \int f_i dP, \text{ for all } P \in \mathcal{P}. \tag{2.3.7}$$

However, since Z is a boundary point of $\mathcal{N}$, i.e. lies on the hyperplane,

$$U(g;M) = k + \sum_{i=0}^{n} a_i M_i. \tag{2.3.8}$$

Combining (2.3.7) and (2.3.8), we see that the left hand side of (2.3.4) is greater than the right hand side of (2.3.4) and combining this result with (2.3.5) we see that (2.3.4) is verified. We note that in the proof we have shown that there exist real constants $a_0, \ldots, a_n$ that attain the inf in the right hand side of (2.3.4).

Corollary. $L(g;M) = \inf_{P \in \mathcal{P}(M)} \int g(\omega) dP(\omega)$

$$= \sup_{a_0, \ldots, a_n} \left\{ \left(\sum_{i=0}^{n} a_i M_i + \inf_{P \in \mathcal{P}} \int (g - \sum_{i=0}^{n} a_i f_i) dP \right) \right\}. \tag{2.3.9}$$

The corollary can be proved similarly to the theorem.

Remark 1: If, for any $a = (a_0, \ldots, a_n)$,

$$\sup_{P \in \mathcal{P}} \int (g - \sum_{i=0}^{n} a_i f_i) dP = \begin{cases} 0, & g(\omega) \leq \sum_{i=0}^{n} a_i f_i(\omega) \text{ on } \Omega, \\ \infty, & \text{otherwise,} \end{cases}$$

then
$$\inf_{a_0, \ldots, a_n} \left(\sum_{i=0}^{n} a_i M_i + \sup_{P \in \mathcal{P}} \int (g - \sum_{i=0}^{n} a_i f_i) dP \right) = \inf{}' \sum_{i=0}^{n} a_i M_i, \tag{2.3.10}$$

where inf' is taken over all vectors $(a_0, \ldots, a_n)$ such that

$\sum_{i=0}^{n} a_i f_i(\omega) \geq g(\omega)$ on Ω.

$$U(g;M) = \inf{}' \sum_{i=0}^{n} a_i M_i. \tag{2.3.11}$$

The probability distribution which actually attains the bound in (2.3.11) is called an extremal distribution.

Similarly, $$L(g;M) = \sup{}' \sum_{i=0}^{n} a_i M_i, \tag{2.3.12}$$

where sup' is taken over all vectors $(a_0, \ldots, a_n)$ such that

$\sum_{i=0}^{n} a_i f_i(\omega) \leq g(\omega)$ on Ω.

Isii [58] has also proved the above results by treating Tchebycheff type inequalities as a problem in linear programming. Kingman [65], through the use of convexity arguments offers results which are analogous to those above in the case when $f_0(\omega) = 1$.

Remark 2: If, as a special case, we let $f_0(\omega) = 1$, $M_0 = 1$ and $g(\omega) = I_A(\omega)$, then inequality (2.3.2) is sharp when M is an interior point of $\mathcal{M}$. As mentioned before, this case was treated by Isii [54,55,56]. Using game theoretic arguments, Marshall and Olkin [83] have proved similar results when $f_i = x^i$, $i = 1,\ldots, n$ and Ω is any of the following: $(-\infty,\infty)$, $[0,\infty)$, $[0,1]$.

Remarks 1 and 2 will be used extensively to prove several results which had originally been proved by various methods. When we refer back to Theorem 2.3.1 in this paper, we shall actually be referring to remarks 1 and 2 of Theorem 2.3.1.

Example: Let us consider establishing the Tchebycheff inequality,

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2, \ k > 0,$$

by means of the general theorem. If X is a r.v. such that $E(X) = \mu$, and $Var(X) = \sigma^2$, then $P(|X - \mu| \geq k\sigma) = P(|Y| \geq k)$, where $Y = (X - \mu)/\sigma$, $E(Y) = 0$ and $Var(Y) = 1$. Let $A = (-\infty,-k] \cup [k,\infty)$. Let us consider a polynomial of the form

$$f(y) = a + by + cy^2. \tag{2.3.13}$$

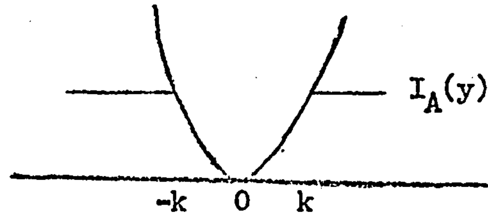
Theorem 2.3.1 tells us that the sharp upper bound of $P(Y \in A)$ is attained by

$$\inf E(f(Y)) = a + c, \tag{2.3.14}$$

where the inf is taken over (a,b,c) satisfying $f(y) \geq I_A(y)$ on Ω.

Since $f(-k) = f(k) = 1$, therefore $b = 0$. Thus $f(y) = a + cy^2$.

Fig. 2.1



We would now like to lower f(y) subject to the condition $f(k) = f(-k) = 1$. By algebra, the lowest point of f(y) occurs at $-b/2c$. Therefore, $a = 0$, $c = 1/k^2$, $f(y) \geq I_A(y)$ and $f(k) = f(-k) = 1$. $E(f(Y)) = 1/k^2$ and $P(Y\in A) \leq 1/k^2$. Thus Tchebycheff's inequality is obtained.

2.4 Bounds in Terms of Mean Deviation

Peek [94] has offered us a probability bound for symmetric interval about the mean in terms of the mean deviation (equation 2.2.16). When the interval about μ is not symmetrical, Glasser [32] offers a probability bound in terms of the mean deviation; his bound is an improvement over Peek's. The reason for choosing v, the mean deviation, rather than σ can be attributed to the fact that extreme values in the distribution will make σ considerably larger than v.

Theorem 2.4.1. Let X be a r.v. with mean μ and finite mean deviation $v > 0$.

$$1 - P(-t_1 v < X - \mu < t_2 v) \leq \begin{cases} 1, & 2t_1t_2 \leq t_1 + t_2, \\ \frac{1}{2}\left(\frac{1}{t_1} + \frac{1}{t_2}\right), & 2t_1t_2 \geq t_1 + t_2. \end{cases} \qquad (2.4.1)$$

Proof: Though Glasser's method of proof is quite simple, we shall prove the result through the use of Theorem 2.3.1 following the development given by Karlin and Studden [60]. Let $Y = (X - \mu)/v$; $E(Y) = 0$, $E|Y| = 1$ and for $A = (-\infty, -t_1] \cup [t_2, \infty)$ we wish to determine $P(Y\in A)$. Let us consider a polynomial of the form

$$f(y) = a_0 + a_1 y + a_2|y|. \qquad (2.4.2)$$

By Theorem 2.3.1 the upper bound on $P(Y\in A)$ is achieved by

$$\inf E(f(Y)) = a_0 + a_2 , \tag{2.4.3}$$

where the inf is taken over all vectors (a_0, a_1, a_2) such that

$$f(y) \geqslant I_A(y) \text{ on } \Omega . \tag{2.4.4}$$

$f(y)$ is linear for $y \in (-\infty, 0]$ and $y \in [0, \infty)$. $f(0) = a_0 \geqslant 0$ and since by (2.4.3) $a_0 + a_2 \leqslant 1$, we can conclude that $a_0 \in [0,1]$. Subject to conditions (2.4.2) and (2.4.4) we can say that $f(-t_1) = f(t_2) = 1$ (i.e. $f(y) < 1$ on $(-t_1, t_2)$).

$$f(-t_1) = a_0 - a_1 t_1 + a_2 t_1 = 1 \tag{2.4.5}$$

$$f(t_2) = a_0 + a_1 t_2 + a_2 t_2 = 1 \tag{2.4.6}$$

Solving the above equations, we obtain

$$a_1 = \frac{1 - a_0}{2}\left(\frac{t_1 - t_2}{t_1 + t_2}\right), \quad a_2 = \frac{1 - a_0}{2}\left(\frac{t_1 + t_2}{t_1 t_2}\right).$$

Accordingly, we must minimize

$$a_0 + \frac{1 - a_0}{2}\left(\frac{1}{t_1} + \frac{1}{t_2}\right) = a_0(1 - p) + p, \tag{2.4.7}$$

where $p = \frac{1}{2}(1/t_1 + 1/t_2)$. Choosing a_0 appropriately, i.e. 0 or 1, the minimum of (2.4.7) is equal to p if $p \leqslant 1$ and is equal to 1 if $p \geqslant 1$. Q.E.D. (2.4.1).

Corollary 1. If $t_1 = t_2 = t$, then

$$1 - P(-tv < X - \mu < tv) \leq \begin{cases} 1, & t \leq 1, \\ 1/t, & t \geq 1. \end{cases} \tag{2.4.8}$$

Corollary 2. If $t_1 \to \infty$ and $t_2 = t$, then

$$1 - P(-\infty < X < \mu + tv) \leq \begin{cases} 1, & t \leqslant \frac{1}{2}, \\ 1/2t, & t \geqslant \frac{1}{2}; \end{cases} \tag{2.4.9}$$

Example: Through a distribution which attains equality in (2.4.1) we show that the result is the best possible. Define the distribution of a r.v. X as follows:

$$P(X = -t_1 v) = \frac{1}{2t_1}, \quad P(X = t_2 v) = \frac{1}{2t_2},$$

$$P(X = 0) = 1 - \tfrac{1}{2}(1/t_1 + 1/t_2).$$

$E(X) = 0$, $E/X/ = v$. By Theorem 2.4.1, $1 - P(-t_1 v < X < t_2 v) \leq \frac{1}{2}(1/t_1 + 1/t_2)$.

However, for our example $1 - P(-t_1 v < X < t_2 v) = P(X = -t_1 v) + P(X = t_2 v)$

$$= \tfrac{1}{2}(1/t_1 + 1/t_2).$$

Therefore, Glasser's inequality is sharp.

2.5 Unsymmetrical Intervals

In this section we shall not restrict ourselves to symmetric intervals about the mean but we shall consider any arbitrary interval containing the mean.

Theorem 2.5.1. Let X be a r.v. with mean μ and finite variance σ^2. Then, for $\beta \geq \alpha > 0$,

$$1-P(-\alpha < X-\mu < \beta) \leq \begin{cases} \dfrac{\sigma^2}{\alpha^2 + \sigma^2}, & \alpha(\beta - \alpha) \geq 2\sigma^2, \quad (2.5.1) \\[2ex] \dfrac{(\beta - \alpha)^2 + 4\sigma^2}{(\alpha + \beta)^2}, & 2\alpha\beta \geq 2\sigma^2 \geq \alpha(\beta - \alpha), \quad (2.5.2) \\[2ex] 1, & \sigma^2 \geq \alpha\beta. \quad (2.5.3) \end{cases}$$

The above, known as Selberg's inequality [103], was first proved through the use of Schwarz's inequality. We will now derive the result through the use of Theorem 2.3.1.

Proof: Let $Y = X - \mu$; thus $E(Y) = 0$ and $Var(Y) = \sigma^2$. We must now find $P(Y \in A)$ where $A = (-\infty, -\alpha] \cup [\beta, \infty)$. To determine $P(Y \in A)$ we must consider all polynomials of the form.

$$f(y) = a_0 + a_1 y + a_2 y^2 \geq I_A(y) \qquad (2.5.4)$$

and we must minimize

$$E(f(y)) = a_0 + a_2\sigma^2 \qquad (2.5.5)$$

with respect to all polynomials of the form (2.5.4). The lowest point of (2.5.4) occurs when $y = -a_1/2a_2$, and (2.5.4) is symmetric about this point. We would like to lower (2.5.4) such that $f(-a_1/2a_2) = 0$.

$$a_0 - a_1^2/2a_2 + a_2 a_1^2/4a_2^2 = 0$$

$$a_0 = a_1^2/4a_2$$

Thus (2.5.4) can be written as

$$f(y) = a_1^2/4a_2 + a_1 y + a_2 y^2 = a_2(y + a_1/2a_2)^2 \;. \qquad (2.5.6)$$

Let $(-a_1/2a_2) = s$, the point about which (2.5.6) is symmetric. Now (2.5.6) must equal 1 at either $y = -\alpha$ or $y = \beta$ depending on whether $-\alpha < s \leq (\beta - \alpha)/2$ or $(\beta - \alpha)/2 \leq s < \beta$, respectively. If $y = -\alpha$, then

$$f(-\alpha) = a_2(-\alpha - s)^2 = 1 \text{ and } a_2 = 1/(\alpha + s)^2.$$

Similarly if $y = \beta$, then $f(\beta) = a_2(\beta - s)^2 = 1$ and $a_2 = 1/(\beta - s)^2$.

Thus (2.5.6) can be written as

$$f(y) = \begin{cases} \dfrac{(y - s)^2}{(\alpha + s)^2}, & -\alpha < s \leq (\beta - \alpha)/2, \qquad (2.5.7) \\[2ex] \dfrac{(y - s)^2}{(\beta - s)^2}, & (\beta - \alpha)/2 \leq s < \beta. \qquad (2.5.8) \end{cases}$$

Corresponding to (2.5.4) and (2.5.5) we must minimize

$$E(f(y)) = \begin{cases} \dfrac{(\sigma^2 + s^2)}{(\alpha + s)^2}, & -\alpha < s \leq (\beta - \alpha)/2, \quad (2.5.9) \\ \dfrac{(\sigma^2 + s^2)}{(\beta - s)^2}, & (\beta - \alpha)/2 \leq s < \beta, \quad (2.5.10) \end{cases}$$

with respect to all polynomials of the forms (2.5.7) and (2.5.8). Minimizing (2.5.9) with respect to s, we see that for $s = \sigma^2/\alpha$ (2.5.9) is a minimum. The minimum value of (2.5.9) is therefore

$$\sigma^2/(\alpha^2 + \sigma^2). \qquad (2.5.11)$$

Similarly minimizing (2.5.10) with respect to s we see that for $s = (\beta - \alpha)/2$, (2.5.10) takes on a minimum value of

$$\frac{(\beta - \alpha)^2 + 4\sigma^2}{(\alpha + \beta)^2}. \qquad (2.5.12)$$

If the condition of (2.5.1) prevails, then the minimum value of $P(Y \in A)$ is given by (2.5.11) and is attained by (2.5.9). If the condition of (2.5.2) prevails, then the minimum of $P(Y \in A)$ is given by (2.5.12) and is attained by (2.5.10). We have thus proved (2.5.1) and (2.5.2).

To obtain (2.5.3) we consider a discrete distribution consisting of two points x_1 and x_2 such that x_1 and x_2 are not in $(-\alpha, \beta)$. Assume $\mu = 0$.

$$E(X) = px_1 + (1 - p)x_2 = 0 \qquad (2.5.13)$$

$$Var(X) = px_1^2 + (1 - p)x_2^2 = \sigma^2 \qquad (2.5.14)$$

Solving (2.5.13) and (2.5.14) we see that $p = x_2^2/(\sigma^2 + x_2^2)$. If $x_1 = (p - 1)x_2/p = -\sigma^2/x_2 < -\alpha$ and $y \in [\beta, \sigma^2/\alpha]$ (i.e. to conform with the condition of (2.5.3)), then, (2.5.3) is satisfied.

Remark 1. [60] There exist extremal distributions which satisfy these bounds. The spectrum of the extremal distribution in the case (2.5.1) consists of the points $\mu - \alpha$ and $\mu + \sigma^2/\alpha$ with respective probabilities

$\sigma^2/(\sigma^2 + \alpha^2)$ and $\alpha^2/(\sigma^2 + \alpha^2)$. In the case of (2.5.2) the spectrum consists of the points $\mu - \alpha$, $\mu + (\beta - \alpha)/2$ and $\mu + \beta$ with respective probabilities of $[\beta(\beta - \alpha) + 2\sigma^2]/(\beta + \alpha)^2$, $4(\beta\alpha - \sigma^2)/(\beta + \alpha)^2$ and $(\alpha^2 - \beta\alpha + 2\sigma^2)/(\beta + \alpha)^2$. The extremal distribution for (2.5.3) was given immediately after (2.5.14). Note that the extremal distributions have been transformed so as to have mean μ.

Let X be a r.v. with mean μ and absolute moments v_r and v_{2r}. We wish to determine an upper bound for the probability statement

$$P(X \in A) = P\Big(X \in (-\infty, \mu - \beta] \cup [\mu - \alpha, \mu + \alpha] \cup [\mu + \beta, \infty)\Big). \qquad (2.5.15)$$

Assume $\mu = 0$. We know that if f(x) and g(x) are two density functions satisfying $g(x) = 0$, $x < 0$, and $g(x) = f(x) + f(-x)$ for $x \geq 0$, then

$$\int_{-\infty}^{\infty} x^{2r} f(x)dx = \int_0^{\infty} x^{2r} g(x)dx, \quad r \geq 1,$$

$$\int_{-\infty}^{\infty} |x|^r f(x)dx = \int_0^{\infty} x^r g(x)dx, \quad r \geq 1,$$

and

$$\int_{-k}^{k} f(x)dx = \int_0^k g(x)dx.$$

The original probability bound of $P(X \in A)$ can be given by the probability bound $P(Y \in A \cup [0, \infty))$, where Y is a nonnegative r.v. such that $E(Y^r) = \mu_r = v_r$ and $E(Y^{2r}) = \mu_{2r} = v_{2r}$. The idea here of considering a nonnegative r.v. is similar to the idea in section 2.2 of this paper. Let $Z = Y^r$; $E(Z) = \mu_r$, $Var(Z) = \mu_{2r} - \mu_r^2$.

$$P(Y \in [0, \alpha] \cup [\beta, \infty)) = P(Z \in [0, \alpha^r] \cup [\beta^r, \infty)) \qquad (2.5.16)$$

From remark 1 of the above theorem we know that the extremal distributions are nonnegative only when $\mu - \alpha \geq 0$. Thus the upper bounds for $P(X \in A)$, when $\mu_r \in (\alpha^r, \beta^r)$, are the same as those for Theorem 2.5.1 provided that $\mu = \mu_r$, $\sigma^2 = \mu_{2r} - \mu_r^2$, $\alpha = \mu_r - \alpha^r$, $\beta = \beta^r - \mu_r$ are substituted in the right hand side of the inequalities. In the case when $\mu_r \notin (\alpha^r, \beta^r)$ the probability bound is 1.

The above development by Karlin and Studden [60] gives a generalization of Guttman's result [41] which states that, for $\lambda > 0$,

$$1 - P(X \in (\mu - k_1\sigma, \mu - k_2\sigma) \cup (\mu + k_2\sigma, \mu + k_1\sigma)) \leq \lambda^{-2}, \tag{2.5.17}$$

where $\mu_4 = (a^2 + 1)\sigma^4$, $k_1 = (1 + \lambda a)^{\frac{1}{2}}$ and $k_2 = (1 - \lambda a)^{\frac{1}{2}}$. To show this connection let $r = 2$ and let us rewrite (2.5.2) in the form of (2.5.15) with the notation used in (2.5.16).

$$1 - P(X \in (\mu - \beta, \mu - \alpha) \cup (\mu + \alpha, \mu + \beta)) \leq \frac{(\beta^2 - 2\mu_2 + \alpha^2)^2 + 4(\mu_4 - \mu_2^2)}{(\beta^2 - \alpha^2)^2}. \tag{2.5.18}$$

If $\beta = k_1\sigma$, $\alpha = k_2\sigma$, the condition of (2.5.2) is satisfied and the right hand side of (2.5.18) equals λ^{-2}.

2.6 Trigonometric Moments

Let us reconsider some of the comments made at the beginning of section 2.3 and apply them to obtain a Tchebycheff type inequality in a situation when trigonometric moments are known.

Theorem 2.6.1 [83]. If X is a random angle in $[0, 2\pi)$ and $E(\sin X) = \alpha$, $E(\cos X) = \beta$, then

$$1 - P(2\theta < X < 2\phi) \leq \frac{1 - \alpha \sin(\theta + \phi) - \beta \cos(\theta + \phi)}{1 - \cos(\phi - \theta)}, \tag{2.6.1}$$

$$P(2\theta \leq X \leq 2\phi) \leq \frac{1 + \alpha \sin(\theta + \phi) + \beta \cos(\theta + \phi)}{1 + \cos(\phi - \theta)}, \tag{2.6.2}$$

where $0 \leq \theta \leq \phi \leq \pi$.

Proof: Let $A' = (2\theta < x < 2\phi)$ where A' denotes the complement of A.

The problem will be to find an upper bound for $P(X \in A)$. Let us consider the following function,

$$f(x) = a_0 + a_1 \sin x + a_2 \cos x \geq I_A(x). \tag{2.6.3}$$

$$E(f(X)) = a_0 + a_1\alpha + a_2\beta \tag{2.6.4}$$

We must now find the minimum of (2.6.4) with respect to all polynomials satisfying (2.6.3).

Subject to the conditions of the hypothesis and (2.6.1) we can let $f(2\theta) = f(2\phi) = 1$ and $f(\theta + \phi) = 0$.

$$\begin{aligned} &a_0 + a_1 \sin(\theta + \phi) + a_2 \cos(\theta + \phi) = 0 \\ &a_0 + a_1 \sin(2\theta) + a_2 \cos(2\theta) = 1 \\ &a_0 + a_1 \sin(2\phi) + a_2 \cos(2\phi) = 1 \end{aligned} \tag{2.6.5}$$

Solving the above equations we can write (2.6.3) as

$$f(x) = \frac{1}{1 - \cos(\phi - \theta)} - \frac{\sin(\theta + \phi) \sin x}{1 - \cos(\phi - \theta)} - \frac{\cos(\theta + \phi) \cos x}{1 - \cos(\phi - \theta)} . \tag{2.6.6}$$

Accordingly (2.6.4) becomes

$$\frac{1 - \alpha \sin(\theta + \phi) - \beta \cos(\theta + \phi)}{1 - \cos(\phi - \theta)} \tag{2.6.7}$$

and we have proved (2.6.1).

If we now let $A = (2\theta \leq x \leq 2\phi)$ and let $f(x)$ be defined as in (2.6.3), then upon setting $f(2\theta) = f(2\phi) = 1$ and $f(\theta + \phi + \pi) = 0$ and carrying out the same procedure as above, we obtain (2.6.2).

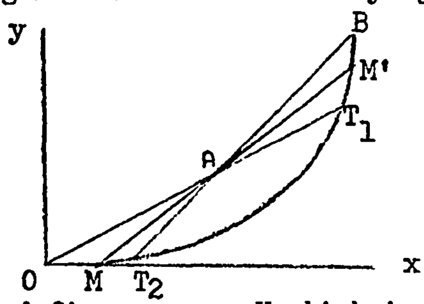
The above results are due to Marshall and Olkin [83].

2.7 Other Inequalities

In this section we shall briefly review two Tchebycheff type inequalities without offering their proofs; a third inequality will also be mentioned.

Von Mises [118] considers a nonnegative r.v. X such that the expected values of any two arbitrary functions of X are known. Let g(X), h(X) be any two monotonic increasing functions of X satisfying the condition $g(0) = h(0) = 0$.

Fig. 2.2



Let $x = g(t)$, $y = h(t)$, and let (x,y) define a curve K which is concave upwards in R_2. Let 0 denote the point (0,0). We define $a = E(g(X))$, $b = E(h(X))$ and let A denote the point (a,b); $p = g(\infty)$, $q = h(\infty)$ and let B denote the point (p,q). We note that if either q or p and q are infinite, we only consider the value for which q/p is finite. Let t_1 and t_2 be defined so as to satisfy the equations

$$\frac{h(t_1)}{g(t_1)} = \frac{b}{a}, \qquad \frac{h(t_2) - b}{g(t_2) - a} = \frac{b - q}{a - p}. \tag{2.7.1}$$

For $t \geq t_1$, or $t \leq t_2$ the equation

$$\frac{a - g(t)}{g(t') - g(t)} = \frac{b - h(t)}{h(t') - h(t)} \tag{2.7.2}$$

admits only one solution t' different from t. Let $x' = g(t')$ and $y' = h(t')$.

Geometrically, we can interpret t_1 and t_2 as being the points satisfying the parametric equations, $x = g(t)$ and $y = h(t)$, when 0A and AB meet K at T_1 and T_2, respectively. For $M = (g(t),h(t))$, $M' = (x',y')$ is the point obtained when MA intersects K.

Von Mises claims that the problem of finding probability bounds for any distribution can be restricted to that of finding probability bounds for a distribution taking on at most three values. Through the laws of statics he obtains the following results.

$$0 \leq P(0 \leq X \leq t) \leq 1 - \frac{a - g(t)}{x' - g(t)}, \quad t_2 \geq t,$$

$$\frac{(a - p)(b - h(t)) - (a - g(t))(b - q)}{ph(t) - qg(t)} \leq P(0 \leq X \leq t)$$

$$\leq 1 - \frac{ah(t) - bg(t)}{ph(t) - qg(t)}, \quad t_2 \leq t \leq t_1,$$

$$\frac{a - g(t)}{x' - g(t)} \leq P(0 \leq X \leq t) \leq 1, \quad t_1 \leq t. \tag{2.7.3}$$

The above inequalities are sharp. The first and third inequalities are obtained by distributions taking on two values and the second inequality is obtained by a distribution taking on three values.

An inequality which may be of some assistance to a statistician dealing with control chart procedures was offered by Winsten[125]. Let us denote the mean range of a sample of size n by w_n,

$$w_n = \int_{-\infty}^{\infty} R_n(F)dx, \quad [62, \text{ page } 339] \tag{2.7.4}$$

where $R_n(F) = 1 - F^n - (1 - F)^n$ and F is a distribution function.

For fixed $t > 0$ we uniquely define an integer m by

$$\sum_{i=1}^{m-1} R_n(i/m) \leq 1/t < \sum_{i=1}^{m} R_n(i/(m + 1)). \tag{2.7.5}$$

Then for $1/(m + 1) < p \leq 1/m$ we define p by the following equation:

$$\sum_{i=1}^{m} R_n(ip) = 1/t. \tag{2.7.6}$$

Subject to the above, Winsten has shown that for any fixed interval whose length is t times w_n,

$$\sup_x \int_x^{x+tw_n} dF(x) \geq p. \tag{2.7.7}$$

The author shows that (2.7.7) is sharp.

Certain ways of representing f(x) for a r.v. X whose distribution is not completely known have been given through Gram-Charlier series and Hermite polynomials. f(x) is represented as an infinite series whose terms are linear functions of the density function of a N(0,1) r.v. and its derivatives. Aoyama [1] uses Hermite polynomials to claim that

$$P(|X - \mu| \geq k\sigma) \leq \frac{1\cdot3\cdot5\cdots(2n-1)}{k^n}\,[1 + g(n,c)], \qquad (2.7.8)$$

where

$$1\cdot3\cdot5\cdots(2n-1)[1 + g(nc)] = \frac{(2n)}{2^n n!}\left\{1 + 1.09c\,\frac{(2^2 2!)^2}{5!}\binom{n}{2} + \frac{(2^3 3!)^2}{7!}\binom{n}{3} + \cdots + \frac{(2^n n!)^2}{(2n+1)!}\binom{n}{n}\right\},$$

$c = \int_{-\infty}^{\infty} \left|\frac{d^2 f}{dx^2}\right| \exp(x^2/4)dx$ and n is a positive integer.

When k = 3 and n = 1, the value in the right hand side of (2.7.8) is 0.1111; this is equal to the value obtained from Tchebycheff's inequality. However, when k = 3 and n = 2, the value in the right hand side of (2.7.8) is 0.0370; this is an improvement over the value obtained from Tchebycheff's inequality.

CHAPTER THREE: INEQUALITIES FOR RESTRICTED UNIVARIATE DISTRIBUTIONS

3.1 Introduction

Most of the inequalities in chapter two were sharp; their sharpness was exhibited by a distribution which attained equality in the inequality. Under a general set of conditions, the results of chapter two cannot be improved upon; however, if additional information about the random variable X, f(x) or F(x) is available, we may be able to improve our inequalities. This additional information may be in the form of an added restriction such as boundedness of the r.v. (Theorem 2.1.2) or it may tell us that f(x) assumes a specific shape, e.g. unimodal.

Throughout chapter two we assumed that the conditions of the Hamburger moment problem were satisfied. In the case of unimodal distributions, Johnson and Rogers [59] have stipulated the moment conditions which must be satisfied so that a set of real numbers can be realized as moments of a unimodal distribution. They have shown that for any set of real numbers, $\mu_1', \ldots, \mu_{n-1}'$, there exists a unimodal distribution function F(x) with mode at zero which satisfies

$$\int_{-\infty}^{\infty} x^r dF(x) = \mu_r' , \quad 1 \leq r < n,$$

iff there exists another distribution function G(x) such that

$$\int_{-\infty}^{\infty} x^r dG(x) = (r + 1)\mu_r', \quad 1 \leq r < n,$$

where n is odd. They have shown that for a r.v. with mean μ, variance σ^2 and mode x_0, their condition simplifies to

$$(\mu - x_0)^2 \leq 3\sigma^2.$$

When dealing with unimodal distributions in this chapter, we shall assume that the conditions for the unimodal moment problem are satisfied.

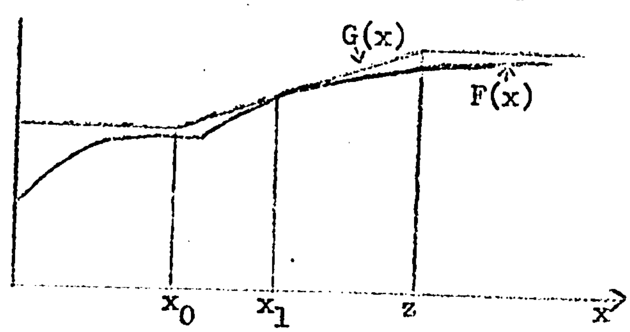
In the next two sections of this chapter we shall obtain probability inequalities through the use of geometrical arguments. In the later sections we shall use remark 1 of Theorem 2.3.1 to obtain upper probability bounds. Our method of proof in the later sections of this chapter will rely heavily on the works of Godwin [34] and Karlin and Studden [60].

It is to be noted that most of the inequalities obtained in this chapter are sharp; when dealing with unimodal distributions, equality is often obtained by distributions which are uniform on some finite interval and which may have some additional mass attached to one or more points.

3.2 Restrictions on F(x) Over Some Given Range

Let X be a nonnegative r.v. such that μ_r' exists for $r \geq 1$ and let us consider three possible values of X, x_0, x_1, and z, where $0 \leq x_0 \leq x_1 \leq z$, such that F(x) is concave downward in (x_0,z). If we join the points $(z,1)$ and $(x_1,F(x_1))$ and extend the line to x_0, F(x), for $x \in (x_0,z)$ will lie entirely below this line. Let G(x) denote the equation of this line.

Fig. 3.1



Then, from geometry,

$$F(x) \leq G(x) = F(x_1) + \frac{1 - F(x_1)}{z - x_1}(x - x_1); \tag{3.2.1}$$

$$z \geq z_0 = x_0 + (x_1 - x_0)/F(x_1). \tag{3.2.2}$$

$$\mu_r' \geq \int_{x_0}^{z} x^r dF(x) \geq \int_{x_0}^{z} x^r dG(x) \geq \int_{x_0}^{z} x^r[(1 - F(x_1))/(z - x_1)]dx$$

$$= \frac{1 - F(x_1)}{z - x_1} \frac{z^{r+1} - x_0^{r+1}}{r + 1} .$$

$$1 - F(x_1) \leq (r + 1)\mu_r'(z - x_1)/(z^{r+1} - x_0^{r+1}). \qquad (3.2.3)$$

The right hand side of (3.2.3) assumes a unique maximum value for some a satisfying

$$a^{r+1} - x_0^{r+1} = (r + 1)a^r(a - x_1) , \quad a > x_0;$$

by rearranging we obtain

$$x_0^{r+1} + ra^{r+1} - (r + 1)x_1 a^r = 0. \qquad (3.2.4)$$

(3.2.3)′ can be written independently of z, i.e.

$$1 - F(x_1) \leq (r + 1)\mu_r'(a - x_1)/(a^{r+1} - x_0^{r+1}) .$$

Combining this with (3.2.4) we obtain

$$1 - F(x_1) \leq \mu_r' /a^r, \qquad (3.2.5)$$

which is the same as

$$P(X \geq x_1) \leq \mu_r' /a^r. \qquad (3.2.6)$$

If a, as obtained in (3.2.4), is greater than z_0 in (3.2.2), (3.2.5) is the best possible. If, however, $a < z_0$, (3.2.5) can be improved upon. Let $z = z_0$; (3.2.3) becomes

$$1 - F(x_1) \leq (r + 1)\mu_r'(z_0 - x_1)/(z_0^{r+1} - x_0^{r+1}). \qquad (3.2.7)$$

By (3.2.2) we have

$$\frac{z_0 - x_1}{z_0 - x_0} \leq \frac{(r + 1)\mu_r'(z_0 - x_1)}{z_0^{r+1} - x_0^{r+1}} ,$$

$$\frac{z_0^{r+1} - x_0^{r+1}}{z_0 - x_0} \leq (r + 1)\mu_r' . \qquad (3.2.8)$$

Substituting (3.2.8) into (3.2.6) we conclude

$$P(X \geq x_1) = 1 - F(x_1) \leq (z_1 - x_1)/(z_1 - x_0), \qquad (3.2.9)$$

where $z_1 \geq z_0$ is the real root obtained from

$$z_1^{r+1} - x_0^{r+1} = (r + 1)\mu_r'(z_1 - x_0).$$

Remark: It can be seen easily that equality can be obtained in (3.2.9); for if X is uniformly distributed over $[x_0, z_1]$,

$$P(X \geq x_1) = (z_1 - x_1)/(z_1 - x_0).$$

The above results were obtained by Von Mises [117]. The idea behind the argument used to obtain this result seem to be quite similar to those arguments used by Narumi [89] to obtain upper probability bounds for a nonnegative r.v. whose density function, f(x), obeyed some constraint. Narumi considers the cases when f(x) is either non-increasing or non-decreasing over some interval beginning at the origin and using geometrical arguments similar to, but more tedious than the ones above, obtained his probability bounds.

An argument somewhat similar to the one used above will be used in a subsequent section to obtain a probability bound for a unimodal distribution.

A much stronger restriction on F(x) than used by Von Mises was imposed by Van Dantzig [116]; in a practical situation it may be very difficult to verify his condition. Let X be a nonnegative r.v. and let $F(x) = 1 - P(x)$ be its distribution function. Let P(x) be differentiable to order h on some interval [a,b] and let $(-1)^j P^{(j)}(x) > 0$ for $0 \leq j \leq h$ and non-increasing for $j = h$. If for some $x_0 \in [a,b]$, h^* is the integral part of

$$(b - x_0)f(x_0)/(P(x_0) - P(b)),$$

where $f(x) = -\dfrac{dP(x)}{dx}$, and $H = \min(h, h^*)$, $r = x_0 + H[P(x_0) - P(b)]/f(x_0)$ and $\alpha_k' = -\int_a^b (x^k - a^k)dP(x)$, then, Van Dantzig proves the following theorem.

Theorem 3.2.1. Subject to the above conditions, ($\rho \neq$ correlation coefficient)

$$P(X \geq x_0) \leq P(b) + \Gamma_{H,k}(a/r)\alpha_k'/x_0^{\,k}, \tag{3.2.10}$$

where $\Gamma_{H,k}(\alpha) = \max_{0 \leq \rho \leq 1} \left\{ \rho^k (1-\rho)^H / k \int_{\alpha\rho}^{1} y^{k-1}(1-y)^H dy \right\}$, $0 \leq \alpha \leq 1$,

can be obtained from the solution of an equation of degree $\leq H + k$.

Proof: Let us consider another distribution function $G(x) = 1 - Q(x)$ satisfying

i) $Q(x) \leq P(x)$, for all $x \geq 0$,

ii) $Q(x_0) = P(x_0)$,

iii) $\beta_k = -\int_0^{\infty} x^k dQ(x) \geq \hat{\beta}_k > 0$.

Here, as before, $G(x) \geq F(x)$ for all x. Using a lemma attributed to Kemperman, Van Dantzig shows that the conditions of the hypothesis permit us to write $Q(x)$ in the following form. For $a \leq x_0 \leq r$,

$$Q(x) = \begin{cases} P(x), & 0 \leq x < a, \\ P(b) + [P(x_0) - P(b)][(r-x)/(r-x_0)]^h, & a \leq x \leq r, \\ P(b), & r \leq x < b, \\ P(x), & x \geq b. \end{cases}$$

We must note that the best $Q(x)$ is obtained for $h = H = \min(h, h^*)$. We shall nevertheless continue the proof using h. The construction of $Q(x)$ also says that $(r - x_0)f(x_0) = -h[P(x_0) - P(b)]$.

$$\beta_k = -\int_0^{\infty} x^k dQ(x) = -\int_0^a x^k dP(x) + \left\{ P(a) - P(b) - (P(x_0) - P(b)) \cdot [(r-a)/(r-x_0)]^h \right\} a^k$$

$$+ \int_a^r \frac{x^k [P(x_0) - P(b)]}{r - x_0}\, h \left(\frac{r-x}{r-x_0} \right)^{h-1} dx - \int_b^{\infty} x^k dP(x) .$$

Let $x_0 = \rho r$, $a = \alpha x_0$, $b = \beta x_0$, and $x = yr$. (3.2.11)

$$\rho^{-1} - 1 = h(P(x_0) - P(b))/x_0 f(x_0).$$

Thus, by a change of variable,

$$\beta_k x_0^{-k} = [-\int_0^a x^k dP(x)]x_0^{-k} - [\int_b^\infty x^k dP(x)]x_0^{-k}$$

$$+ \left\{P(a) - P(b) - [P(x_0) - P(b)]\left(\frac{1-\alpha\rho}{1-\rho}\right)^h\right\}\alpha^k$$

$$+ \frac{h[P(x_0) - P(b)]}{\rho^k(1-\rho)^h}\int_{\alpha\rho}^1 y^k(1-y)^{h-1}dy$$

$$= \left\{-\int_0^a x^k dP(x) - \int_b^\infty x^k dP(x)\right\} x_0^{-k} + [P(a) - P(b)]\,\alpha^k$$

$$+ [P(x_0) - P(b)]\, I(\alpha\rho)/\rho^k(1-\rho)^h, \qquad (3.2.12)$$

where

$$I(\alpha\rho) = h\int_{\alpha\rho}^1 y^k(1-y)^{h-1}dy - (1-\alpha\rho)^h\rho^k\alpha^k$$

$$= k\int_{\alpha\rho}^1 y^{k-1}(1-y)^h dy$$

$$= \frac{h!\,k!}{(h+k)!} - \sum_{j=0}^{h}(-1)^j\frac{h!}{j!(h-j)!}\;\frac{k}{k+j}\,(\alpha\rho)^{k+j} > 0,\quad 0 \le \rho \le 1.$$

$I(\alpha\rho)$ is a polynomial in ρ of degree $h + k$. Let $\hat\rho$ be such that $\rho^k(1-\rho)^h/I(\alpha\rho)$ assumes a maximum value for $0 \le \rho \le 1$. Then, by (3.2.12),

$$-\int_0^\infty x^k dP(x) \ge \beta_k \ge \hat\beta_k = -\int_0^a x^k dP(x) - \int_b^\infty x^k dP(x) + [P(a) - P(b)]a^k$$

$$+ [P(x_0) - P(b)]x_0^k I(\alpha\hat\rho)/[\hat\rho^k(1-\hat\rho)^h]. \qquad (3.2.13)$$

Also,

$$\alpha_k' = -\int_a^b (x^k - a^k)dP(x) = -\int_0^\infty x^k dP(x) + \int_0^a x^k dP(x) + \int_b^k x^k dP(x)$$

$$- [P(a) - P(b)]a^k. \qquad (3.2.14)$$

Combining (3.2.13) and (3.2.14) we see that

$$\alpha_k' \ge [P(x_0) - P(b)]x_0^k/\Gamma_{h,k}(\alpha).$$

Therefore,

$$P(X \ge x_0) \le P(b) + \Gamma_{h,k}(\alpha)\alpha_k'/x_0^k$$

and (3.2.10) follows from (3.2.11).

Van Dantzig also shows that if we wish to avoid the term $\Gamma_{h,k}(\alpha)$ we can obtain a simpler and more general result.

$$P(X \geq x_0) \leq P(b) + \gamma_{h,k} \frac{[\alpha_k + \int_0^a x^k dP(x) + \int_b^\infty x^k dP(x)]}{x_0^k (1 + \phi)}$$

$$+ \frac{[P(a) - P(b)]\psi}{1 + \phi^{-1}} , \qquad (3.2.15)$$

where

$$\gamma_{h,k} = \frac{(h+k)!k^k h^h}{h!k!(h+k)^{h+k}} , \qquad (3.2.16)$$

$$\psi = \frac{I_{k+1,h}(\alpha)}{\alpha^{k+h}} , \quad \alpha_k = -\int_0^\infty x^k dP(x) ,$$

$$I_{L,h}(\alpha) = \sum_{j=0}^{h-1} \binom{L+h-1}{L+j} \alpha^{L+j}(1-\alpha)^{h+1-j}$$

and $\phi = k^k h^h \alpha^{k+h}/(k+h)^{k+h}(1-\alpha)^h$. We note that both Godwin [33] and Isii [53] have made errors in writing out equation (3.2.16). If, as a special case of (3.2.15), we let $h = 1$, $b = \infty$, $k = 2r$, $\int_0^a x^k dP(x) = 0$ and $\psi = 1$, we obtain Camp's inequality [18]; if $a = 0$, $b = \infty$ and $h = 1$, we obtain Meidell's inequalities [84,85].

3.3 Unimodal Distributions: Moments of Higher Order

In subsequent sections we shall obtain sharp inequalities for unimodal distributions in terms of the first two moments. Motivated by Pearson's improvement (2.2.3) over Tchebycheff's inequality through the use of higher even moments, Smith [107] has offered a probability bound for a bounded nonnegative r.v. whose distribution is unimodal. Let $0 \leq X \leq k$ and let $E(X^{2r})$ exist, $r \geq 1$.

$$P(X \geq d) = \int_d^k f(x)dx \leq \frac{1}{d^{2r}} \int_d^k x^{2r} f(x)dx .$$

Also

$$P(X \geq d) \leq \frac{1}{d^{2r}}\left[\mu'_{2r} - \int_0^d x^{2r} f(x)dx\right] = \frac{1}{t^{2r}} - R, \qquad (3.3.1)$$

where $R = (1/d^{2r})\int_0^d x^{2r} f(x)dx$ and $d = t(\mu'_{2r})^{1/2r}$. Let $\phi(x)$ be a function satisfying $(d/dx)[\phi(x)] = f(x)$, $P(x) = \int_x^k f(x)dx$ and $\phi(x) = -P(x)$.

Integrating by parts we get

$$\int_0^d x^{2r} f(x)dx = -d^{2r}P(d) + 2r\int_0^d x^{2r-1}P(x)dx,$$

$$R = -P(d) + (2r/d^{2r})\int_0^d x^{2r-1}P(x)dx. \qquad (3.3.2)$$

Combining (3.3.1) and (3.3.2) we obtain

$$2r\int_0^d x^{2r-1}P(x)dx \leq \mu'_{2r}. \qquad (3.3.3)$$

We now consider $f(x)$ under three different restrictions.

Case I. Let $f(x)$ be a monotone increasing function on $[0,k]$; since the second derivative of $P(x)$ is negative, $P(x)$ is concave downward. Let us draw a line joining the points $(0,P(0))$ and $(d,P(d))$. The equation of this line is

$$y = (x/d)(P(d) - 1) + 1. \qquad (3.3.4)$$

If we substitute y for $P(x)$ in (3.3.3), we obtain

$$(2r/d)(P(d) - 1)\int_0^d x^{2r}dx + 2r\int_0^d x^{2r-1}dx \leq \mu'_{2r}.$$

Upon integration,

$$P(d) \leq (\mu'_{2r}/d^{2r})(1 + 1/2r) - 1/2r. \qquad (3.3.5)$$

If $d = t(\mu'_{2r})^{1/2r}$, then

$$P(d) \leq (1/t^{2r})(1 + 1/2r) - 1/2r. \qquad (3.3.6)$$

In order to apply (3.3.6), $k \leq (1+2r)^{1/2r}(\mu'_{2r})^{1/2r}$. To see this, we consider a uniform distribution $y = f(z)$ on $[0,k]$.

$$E(Z^{2r}) = k^{2r}/(2r+1);\; k = (2r+1)^{1/2r}[E(Z^{2r})]^{1/2r}.$$

Since our $f(x)$ is monotone increasing in $[0,k]$, it follows that

$$(\mu'_{2r})^{1/2r} \geq [E(Z^{2r})]^{1/2r} = k(2r+1)^{-1/2r};$$ hence our limitation on k.

Case II. Let $f(x)$ be a monotone decreasing function on $[0,d]$; $P(x)$ is concave upward and if we draw a tangent to $P(x)$ at $x = \theta d$ $(0 \leq \theta \leq 1)$, $P(x)$ will not cross the tangent at any other place. Let

$$y = -xf(\theta d) + P(\theta d) + \theta df(\theta d) \tag{3.3.7}$$

be the equation of the tangent. As before, if we substitute (3.3.7) into (3.3.3) and integrate, we get

$$P(\theta d) + df(\theta d)[\theta - 2r/(2r+1)] \leq \mu'_{2r}/d^{2r} . \tag{3.3.8}$$

We would like to minimize

$$\mu'_{2r}/d^{2r} - P(\theta d) + df(\theta d)[\theta - 2r/(2r+1)]$$

with respect to θ. Minimizing we obtain $\theta = 2r/(2r+1)$ and substituting this value of θ into (3.3.8) gives

$$P(X \geqslant [2rd/(2r+1)]) \leqslant \mu'_{2r}/d^{2r};$$

$$P(X \geqslant d) \leqslant \mu'_{2r}/d^{2r}(1 + 1/2r)^{2r} . \tag{3.3.9}$$

If $d = t(\mu'_{2r})^{1/2r}$, then,

$$P(X \geqslant d) \leqslant 1/t^{2r}(1 + 1/2r)^{2r}. \tag{3.3.10}$$

Result (3.3.10) was obtained by Meidell [84,85] as well.

Case III. Let $f(x)$ be a monotone increasing function in $[0,c\sigma)$ and a monotone decreasing function in $[c\sigma,k]$. Thus $P(x)$ is concave downward for x $[0,c\sigma)$ and concave upward for $x \geqslant c\sigma$. Let us draw a chord joining the points $(0,P(0))$ and $(c\sigma,P(c\sigma))$ as in case I and let us draw a tangent to some point θd, $c\sigma < \theta d < d$, $0 \leq \theta \leq 1$, as in case II. If, in (3.3.3) $P(x)$ is replaced by the equation of the chord and the equation of the tangent in the respective intervals $[0,c\sigma)$ and $[c\sigma,d]$, then, by carrying out the procedure illustrated in the first two cases, Smith [107] obtains the following result.

$$P(X \geq t\sigma) \leqslant \frac{\frac{\mu'_{2r}}{\sigma^{2r}} - c^{2r}[(2rP(X \geq c\sigma) + 1)/(2r + 1)]}{(t/\theta)^{2r} - c^{2r}}, \qquad (3.3.11)$$

where θ is defined by the equation

$$t = \frac{2r}{2r+1} \frac{t^{2r+1} - (c\theta)^{2r+1}}{\theta[t^{2r} - (c\theta)^{2r}]}.$$

Smith originally showed that $P(X \geq c\sigma)$ in the right hand side of (3.3.11) could be replaced by $P(X \geq t\sigma)$, however, in a subsequent paper [108] he showed that (3.3.11) could be improved upon by replacing $P(X \geq c\sigma)$ by $1 + c[P(X \geq t\sigma) - 1]/t$.

3.4 Gauss' Inequality

Until now we have used different geometrical methods to obtain probability bounds for random variables subject to some restriction. In this and the remaining sections of this chapter we shall show how remark 1 of Theorem 2.3.1 will enable us to obtain some of the classical Tchebycheff type inequalities for unimodally distributed random variables. The methods used will be based on the works of Godwin [34] and Karlin and Studden [60].

In chapter one it was pointed out that Gauss offered an improvement over Tchebycheff's inequality when the r.v. had a unimodal distribution; we shall now prove the result.

Theorem 3.4.1 (Gauss-Winckler). If X is a r.v. such that $E(X) = \mu$, $Var(X) = \sigma^2$ and X has a unique mode at μ, then,

$$P(|X - \mu| \geq k) \leq \begin{cases} 1 - k/(3)^{\frac{1}{2}}\sigma, & 3k^2 \leq 4\sigma^2, \qquad (3.4.1) \\ 4\sigma^2/9k^2, & 3k^2 \geq 4\sigma^2. \qquad (3.4.2) \end{cases}$$

Remark: Through a change of variable we can get the forms (1.1.2) and (1.1.3) given in chapter one.

Proof: Assume $\mu = 0$. By the method discussed in the latter part of section 2.5 of this paper the problem of (3.4.1) and (3.4.2) is reduced to that of determining max $P(X \geq k)$ where max is taken over all unimodal distributions satisfying

$$F(x) = \begin{cases} F(0) + \int_0^x f(t)dt, & 0 \leq x < \infty, \\ 0, & x < 0, \end{cases} \tag{3.4.3}$$

where $f(t)$ is non-increasing. We would also like $f(t)$ to satisfy

$$\int_0^\infty x^2 dF(x) = \sigma^2 = \int_0^\infty t^2 f(t)dt.$$

Assume that $F(0) = 0$. If $F(0) \neq 0$, the distribution function $F(x)$ can be so adjusted that the probability $F(0)$ at 0 becomes uniformly distributed over the interval $[0,\epsilon]$. Using the results for the case when $F(0) = 0$ and letting $\epsilon \to 0$, we can obtain the required solution. $F(0)$ will be defined by

$$F(0) + \int_0^\infty f(t)dt = 1.$$

Let us consider a function $H(t)$ satisfying

$$dH(t) = - df(t). \tag{3.4.4}$$

$$1 = \int_0^\infty f(t)dt = tf(t)\Big|_0^\infty - \int_0^\infty tdf(t) = \int_0^\infty tdH(t). \tag{3.4.5}$$

Similarly, integrating by parts,

$$\sigma^2 = \int_0^\infty t^2 f(t)dt = \int_0^\infty (t^3/3)dH(t) \tag{3.4.6}$$

$$\int_k^\infty f(t)dt = \int_k^\infty (t - k)dH(t). \tag{3.4.7}$$

By (3.4.5), (3.4.6), (3.4.7) and remark 1 of Theorem 2.3.1 we must consider polynomials of the form

$$\phi(t) = at + bt^3 \geq g(t) = \begin{cases} 0, & 0 \leq t \leq k, \\ t - k, & t \geq k, \end{cases} \tag{3.4.8}$$

and we must determine

$$\min(a + 3b\sigma^2), \tag{3.4.9}$$

where the min is taken with respect to all polynomials of the form (3.4.8). To determine a and b we would like to lower $\phi(t)$ such that $\phi(t)$ will touch the non-zero part of $g(t)$ at some point $z \in [k, \infty)$ and $\phi(t)$ will still satisfy (3.4.8).

$$\phi(z) = az + bz^3 = z - k.$$

$$\phi'(z) = a + 3bz^2 = 1.$$

Solving the two simultaneous equations we get

$$a = 1 - 3k/2z, \quad b = k/2z^3.$$

Thus (3.4.9) becomes

$$\min((1 - 3k/2z) + 3k\sigma^2/2z^3). \tag{3.4.10}$$

Minimizing with respect to z, for $a \geq 0$, we obtain $z = (3)^{\frac{1}{2}}\sigma$ whenever $4\sigma^2 \geq 3k^2$ and $z = 3k/2$ whenever $3k^2 \geq 4\sigma^2$. Substituting these values of z into (3.4.10), we obtain (3.4.1) and (3.4.2), respectively.

Remark 1: Equality in (3.4.1) can be obtained by a uniform distribution on $[\mu - (3)^{\frac{1}{2}}\sigma, \mu + (3)^{\frac{1}{2}}\sigma]$. In (3.4.2) equality can be obtained by a distribution which is uniform over $[\mu - (3/2)k, \mu + (3/2)k]$ and which has an extra mass added to the point μ. This extra mass is added at the mode.

Remark 2: Ulin [115] has also obtained probability bounds for unimodal distributions. His bounds, however, are not always sharp.

Under an additional restriction, Karlin and Studden [60] are able to extend Theorem 3.4.1.

Theorem 3.4.2. Let X be a r.v. satisfying the conditions of Theorem 3.4.1 and let the distribution of X also satisfy the additional restriction

$f(\mu - m) = f(\mu + m) = 0$. Then,

$$P(|X - \mu| \geq k) \leq \begin{cases} 1 - k/(3)^{\frac{1}{2}}\sigma, & 3k^2 \leq 4\sigma^2, \quad (3.4.11) \\ 4\sigma^2/9k^2, & 3k^2 \geq 4\sigma^2, \ 3k \leq 2m, \quad (3.4.12) \\ 3\sigma^2(1 - k/m)/m^2, & 3k^2 \geq 4\sigma^2, \ k \leq m \leq 3k/2. \quad (3.4.13) \end{cases}$$

Proof: As in Theorem 3.4.1 we let $\mu = 0$ and our problem is again reduced to that of determining $P(X \geq k)$. Let $t_0 = \min(t:f(t) \leq (1/m))$.

$$\int_0^m [t^2 - t_0^2][f(t) - 1/m]dt \leq 0$$

$$\int_0^m t^2 f(t)dt \leq \int_0^m [t^2/m]dt$$

$$\text{i.e. } \sigma^2 \leq m^2/3. \qquad (3.4.14)$$

If $3k^2 \leq 4\sigma^2$, then by (3.4.14) we see that (3.4.11) = (3.4.1) must hold. Similarly, if $3k^2 \geq 4\sigma^2$, (3.4.2) will also hold whenever $3k \leq 2m$.

Let us consider the situation when $3k^2 \geq 4\sigma^2$ and $k \leq m \leq 3k/2$. We must consider polynomials of the form

$$\phi(t) = at + bt^3 \geq g(t) = \begin{cases} 0, & 0 \leq t \leq k, \\ t - k, & k \leq t \leq m, \end{cases}$$

and we must determine

$$\min(a + 3b\sigma^2)$$

with respect to the above polynomials. If at the point m, for $a > 0$,

$$\phi(m) = am + bm^3 = m - k,$$

then $b = [m(1 - a) - k]/m^3$ and

$$a + 3b\sigma^2 = a(1 - 3\sigma^2/m^2) + 3(1 - k/m)\sigma^2/m^2$$
$$\geq 3(1 - k/m)\sigma^2/m^2, \qquad (3.4.15)$$

since $m^2 \geq 3\sigma^2$. The minimum of $P(X \geq k)$ is given by (3.4.15) since (3.4.15) $\leq$ (3.4.2) whenever $k < m \leq 3k/2$.

Remark: Equality can be obtained in (3.4.13) by a distribution which is uniform over $[\mu - m, \mu + m]$ and which has an extra probability added to the point μ.

3.5 Royden's Inequality [98]

Theorem 3.5.1. Let X be a r.v. such that f(x) has a unique mode at 0 and such that v_1' and v_2' exist. Then

$$P(|X| > k) < \begin{cases} 1 - k/2v_1', \quad 0 \leq k \leq v_1', & (3.5.1) \\ v_1'/2k, \quad v_1' \leq k \leq 3v_2'/4v_1', & (3.5.2) \\ \dfrac{4v_1'^2}{3v_2'} - \dfrac{8v_1'^3 k}{9v_2'^2}, \quad \dfrac{3v_2'}{4v_1'} \leq k \leq \dfrac{v_2'}{v_1'}, & (3.5.3) \\ \dfrac{3v_2' - 4v_1'^2}{3a^2 - 8v_1'a + 3v_2'}, \quad k \geq v_2'/v_1', & (3.5.4) \end{cases}$$

where a is defined as the largest root of

$$2a^3 - (3k + 4v_1')a^2 + 8v_1'ka - 3v_2'k = 0.$$

Proof: As in section 3.4 the problem is reduced to that of determining $P(X \geq k)$ over all distributions (3.4.3) satisfying the moment conditions v_1', v_2'. Again, as in Theorem 3.4.1, we may assume that $F(0) = 0$.

Let $dH(t) = -t\,df(t)$. Integrating by parts as we did in section 3.4,

$$1 = \int_0^\infty f(t)dt = \int_0^\infty dH(t) \tag{3.5.5}$$

$$v_1' = \int_0^\infty tf(t)dt = \int_0^\infty (t/2)dH(t) \tag{3.5.6}$$

$$v_2' = \int_0^\infty t^2 f(t)dt = \int_0^\infty (t^2/3)dH(t) \tag{3.5.7}$$

$$\int_k^\infty f(t)dt = \int_k^\infty (1 - k/t)dH(t) \tag{3.5.8}$$

Remark 1 of Theorem 2.3.1 and conditions (3.5.5) to (3.5.8) tell us that in order to obtain $P(X \geq k)$ we must consider polynomials of the form

$$\phi(t) = a + bt + ct^2 \geq g(t) = \begin{cases} 0, & 0 \leq t \leq k, \\ 1 - k/t, & t \geq k, \end{cases} \tag{3.5.9}$$

and we must find

$$\min(a + 2bv_1' + 3cv_2'), \tag{3.5.10}$$

where the min is taken with respect to all polynomials of the form (3.5.9). Let $\phi(t)$ touch the non-zero part of $g(t)$ at z.

$$\phi(z) = a + bz + cz^2 = 1 - k/z.$$

Differentiating with respect to z we get

$$b + 2cz = k/z^2,$$

and solving the simultaneous equations

$$a = 1 - 2k/z + cz^2 \ , \ b = k/z^2 - 2cz. \tag{3.5.11}$$

(3.5.10) becomes

$$\min[(z^2 - 4v_1'z + 3v_2')c + (z^2 - 2kz + 2v_1'k)/z^2]. \tag{3.5.12}$$

Alternately, the simultaneous equations can be solved so that

$$b = [2 - 3k/z - 2a]/z, \ c = (2k/z - 1 + a)/z^2. \tag{3.5.13}$$

Under these substitutions, (3.5.10) becomes

$$\min[(z^2 - 4v_1'z + 3v_2')a/z^2 + (4v_1'z^2 - (3v_2' + 6kv_1')z + 6v_2'k)/z^3]. \tag{3.5.14}$$

Note: From our representations in (3.5.6) and (3.5.7) and Schwarz's inequality, $3v_2' \geq 4v_1'^2$, i.e. $z^2 - 4v_1'z + 3v_2' \geq 0$.

We shall now determine (3.5.10) when z ranges over the intervals $[2k,\infty)$, $[3k/2,2k]$ and $(k,3k/2]$.

Let us consider the case when $z \geq 2k$. Since $\phi(t) \geq 0$, (3.5.11) tells us that $a \geq 0$ and $b \geq 0$ whenever $c = 0$. Let $c = 0$; to determine (3.5.10) we need only to minimize the right hand term in (3.5.12). Minimizing with respect to z we get $z = 2v_1'$ when $v_1' \geq k$ and $z = 2k$ whenever $v_1' \leq k$. Substituting these values of z into the right hand term of (3.5.12) we get

$$P(X \geq k) \leq \begin{cases} 1 - k/2v_1' \,, & v_1' \geq k \geq 0, \qquad (3.5.1) \\ v_1'/2k \,, & v_1' \leq k. \qquad (3.5.15) \end{cases}$$

Let us consider the interval when $3k/2 \leq z \leq 2k$, i.e. $2k/z \geq 1$, and $3k/z \leq 2$. Since $\phi(t) \geq 0$, we see, from (3.5.13), that for $a = 0$, $b \geq 0$ and $c \geqq 0$. As before we assume that $a = 0$; in order to obtain (3.5.10) we must minimize the right hand term of (3.5.14) with respect to z. Minimizing with respect to z we get $z = 2k$ for $0 \leq k \leq 3v_2'/4v_1'$, $z = 3v_2'/2v_1'$ for $3v_2'/4v_1' \leq k \leq v_2'/v_1'$ and $z = 3k/2$ for $k \geq v_2'/v_1'$. Substituting these values of z into the right hand side of (3.5.14) we get

$$P(X \geq k) \leq \begin{cases} v_1'/2k \,, & 0 \leq k \leq 3v_2'/4v_1' \,, \qquad (3.5.16) \\ \dfrac{4v_1'^2}{3v_2'} - \dfrac{8v_1'^3 k}{9v_2'^2} \,, & \dfrac{3v_2'}{4v_1'} \leq k \leq \dfrac{v_2'}{v_1'} \,, \qquad (3.5.3) \\ 4v_2'/9k^2 \,, & k \geq v_2'/v_1' \,. \qquad (3.5.17) \end{cases}$$

Combining (3.5.15) and (3.5.16) we get (3.5.2).

We must now consider the remaining interval $k < z \leq 3k/2$. One condition which must be satisfied is that $\phi(t) \geq 0$. Let a and b take on the values described in (3.5.11); we thus have a quadratic equation in t which must be ≥ 0. Examining the discriminant D

$$(k/z^2 - 2cz)^2 - 4c(1 - 2k/z + cz^2) \,,$$

we see that $D < 0$ whenever $c \geq k^2/[4z^3(z - k)] = c^* > 0$. Therefore, for $c \geq c^*$, $\phi(t) \geq 0$. Also, $\phi(t) \geq 1 - k/t$ whenever $t \geq k$. Again, by using the values of a and b obtained from (3.5.11) and rearranging the terms,

$$(1 - 2k/z + kt/z^2) + c(z - t)^2 \geq 1 - k/t \,. \qquad (3.5.18)$$

Since the left hand term of (3.5.18) is greater than 0 whenever $c \geq c^*$, it follows that for (3.5.18) to be satisfied, $c \geq c^*$.

Rewriting (3.5.12) with c replaced by c*, we get

$$\min(a+2bv_1'+3cv_2') = \min \frac{4z^4 - 12kz^3 + (8v_1'k + 9k^2)z^2 - 12v_1'k^2z + 3v_2'k^2}{4z^3(z-k)}. \quad (3.5.19)$$

Differentiating with respect to z and equating to 0 we obtain

$$4z^2k(4z - 3k)[2z^3 - (3k + 4v_1')z^2 + 8v_1'kz - 3v_2'k] = 0.$$

We can immediately discard the roots $z = 0$ and $z = 3k/4$ since these values of z do not fall into our interval. We must now examine the third degree polynomial for roots. Differentiating with respect to z and equating to 0 we get

$$(3z - 4v_1')(z - k) = 0.$$

By examining the third degree polynomial for maximum and minimum values we see that the polynomial is negative between the two largest groups and that it assumes its minimum value at $z = k$. Also, by replacing z by $3k/2$ in the third degree polynomial, we see that the largest root will never exceed $3k/2$ iff $k \geq v_2'/v_1'$. Thus the largest root of the third degree polynomial will give a satisfactory minimizing z. This polynomial of degree 3 can alternately be written in the following form:

$$d = \frac{2z^3 - 4v_1'z^2}{3z^2 - 8v_1'z + 3v_2'}.$$

If we substitute this value of d into the right hand side of (3.5.19) and rearrange the terms, we shall get (3.5.4).

Remark: The above proof is due to Karlin and Studden [60]. The inequalities obtained are sharp. When Royden [98] first proved these inequalities, he exhibited distributions which attained equality in each of the four inequalities.

3.6 Bounds for a Nonnegative Random Variable with Unimodal Distribution

Karlin and Studden [60] have offered probability bounds for a non-negative r.v. whose distribution has a unique mode at $x = x_0$.

Theorem 3.6.1. Let X be a nonnegative r.v. such that $E(X) = \mu$ and such that f(x) is unimodal at $x = x_0$, i.e.

$$F(x) = \begin{cases} \int_0^x f(t)dt \ , & 0 \leq x \leq x_0, \ f(t) \text{ is non-decreasing,} \\ F(x_0) + \int_0^x f(t)dt, & x \geq x_0 \ , \ f(t) \text{ is non-increasing,} \end{cases}$$

where $F(x_0) = F(x_0+) - F(x_0-)$.

$$P(X \geq k) \leq \begin{cases} 1 - (k - x_0)/2(\mu - x_0), & k \geq x_0 \ , \ 2\mu - x_0 \geq k + (k^2 - x_0 k)^{\frac{1}{2}} \ , \quad (3.6.1) \\ (2\mu - x_0)/(k^{\frac{1}{2}} + (k - x_0)^{\frac{1}{2}})^2, & k \geq x_0 \ , \ 2\mu - x_0 \leq k + (k^2 - x_0 k)^{\frac{1}{2}} \ , \quad (3.6.2) \\ (2\mu - k)/x_0 \ , & k \leq x_0, \ 2\mu - x_0 \leq k, \quad (3.6.3) \\ 1 \ , & k \leq x_0, \ 2\mu - x_0 \geq k. \quad (3.6.4) \end{cases}$$

Proof: As in the previous two sections, we can, without any loss of generality, let $F(x_0) = 0$. Let us consider a function dH(t) which satisfies $dH(t) = -(t - x_0)df(t)$. Analogous to the last two sections,

$$1 = \int_0^\infty f(t)dt = \int_0^\infty dH(t) \tag{3.6.5}$$

$$\mu = \int_0^\infty tf(t)dt = x_0/2 + \int_0^\infty (t/2)dH(t) \tag{3.6.6}$$

$$\int_k^\infty f(t)dt = \begin{cases} \int_k^\infty [(t - k)/(t - x_0)]dH(t) \ , & k \geq x_0 \\ \int_0^d [(x_0 - k)/(x_0 - t)]dH(t) + \int_k^\infty dH(t) \ , & k < x_0 \ . \end{cases} \tag{3.6.7}$$

Let us consider the case when $k \geq x_0$. Remark 1 of Theorem 2.3.1

and conditions (3.6.5) to (3.6.7) tell us that in order to obtain $P(X \geq k)$ we must consider polynomials of the form

$$\phi(t) = a + bt \geq g(t) = \begin{cases} 0, & 0 \leq t \leq k, \\ (t-k)/(t-x_0), & t \geq k, \end{cases} \tag{3.6.8}$$

and we must find

$$\min(a + b(2\mu - x_0)), \tag{3.6.9}$$

where the min is taken with respect to all polynomials satisfying (3.6.8). Let $z \geq k$ be a point such that

$$a + bz = (z-k)/(z-x_0). \tag{3.6.10}$$

Differentiating (3.6.10) with respect to z we get

$$b = (k - x_0)/(z - x_0)^2 .$$

Substituting this value of b into (3.6.10) gives

$$a = (z^2 - 2kz + kx_0)/(z - x_0)^2 \geq 0 \tag{3.6.11}$$

and solving the quadratic equation in the numerator of (3.6.11) we can see that $z \geq k + (k^2 - x_0 k)^{\frac{1}{2}}$. Substituting the values of a and b into (3.6.9) and rearranging the terms we get

$$\min(1 - 2(k - x_0)(z - \mu)/(z - x_0)^2). \tag{3.6.12}$$

Minimizing this expression with respect to z, we can get (3.6.1) and (3.6.2).

Let us now consider the intervals where $k < x_0$. (3.6.7) tells us that we must consider polynomials of the form

$$\phi(t) = a + bt \geq g(t) = \begin{cases} (x_0 - k)/(x_0 - t), & 0 \leq t \leq k, \\ 1, & t \geq k. \end{cases} \tag{3.6.13}$$

In order to obtain $P(X \geq k)$ we must determine

$$\min(a + b(2\mu - x_0)) , \tag{3.6.14}$$

where the min is taken with respect to all polynomials satisfying

(3.6.13). If $t = 0$, then $a \geq (x_0 - k)/x_0$; therefore $1 \geq a \geq (x_0 - k)/x_0$. Since $(x_0 - k)/(x_0 - t)$ is convex and $b(2\mu - x_0)$ is an increasing function in b, (3.6.14) must be achieved when

$$a + bk = (x_0 - k)/(x_0 - k) = 1,$$

$$\text{i.e. } b = (1 - a)/k.$$

Upon substituting this value of b into (3.6.14) we obtain

$$\min\, (a + (1 - a)(2\mu - x_0)/k)$$

$$= \min\, [(2\mu - x_0)/k - a(2\mu - x_0 - k)/k]. \quad (3.6.15)$$

If in (3.6.15) we let a take on its minimum value $(x_0 - k)/x_0$ whenever $2\mu - x_0 < k$, then we obtain (3.6.3). If in (3.6.15) we let $a = 1$ whenever $2\mu - x_0 > k$, then we get (3.6.4).

Remark: [60] If in the above theorem x_0 was not specified, we could still obtain a bound by maximizing each of the inequalities with respect to x_0.

Karlin and Studden exhibit distributions which attain equality in (3.6.1), (3.6.2) and (3.6.3).

3.7 Numerical Comparisons

In this section we shall offer some numerical results which compare some well known unimodal distributions with the appropriate Tchebycheff type inequalities. We shall consider the following distributions which are symmetrical about the origin, i.e. $\mu = 0$ and the mode lies at 0.

LEGEND	DISTRIBUTION	f(x)	RANGE	VARIANCE	
N	Normal	$(2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$	$-\infty<x<\infty$	1	(3.7.1)
T	Triangular	$\begin{cases}(b+x)/b^2 \\ (b-x)/b^2\end{cases}$	$-b\leq x\leq 0$ $0\leq x\leq b$	$b^2/6$	(3.7.2)
C	Cosine	$(1+\cos x)/2\pi$	$-\pi\leq x\leq\pi$	$(\pi^2/3)-2$	(3.7.3)
L	Logistic	$[\operatorname{sech}^2(x/d)]/2d$	$-\infty<x<\infty$	$(\pi d)^2/12$	(3.7.4)
LA	Laplace	$c\exp(-2c/x/)$, $c>0$	$-\infty<x<\infty$	$1/2c^2$	(3.7.5)

We shall compare the actual probability values attained by each of the above distributions with the best Tchebycheff type inequality which corresponds to the situation. Our legend will be as follows:

G - Gauss (Theorem 3.4.1)

G.T. - Truncated Gauss (Theorem 3.4.2)

Without loss of generality we shall assume that $\sigma^2 = 1$. This can be done since

$$P(/X/ \geq k\sigma) = P(/\frac{X}{\sigma}/ \geq k).$$

However, by assuming that $\sigma^2 = 1$, we must appropriately adjust the truncation points of the triangular and cosine distributions.

The probability values used for distributions (3.7.1) to (3.7.5) have been taken from Chew [22b]; they have been modified to suit our situation.

TABLE I

$P(/X/ \geq k)$

k	G.	N.	L.	LA.
0.2	0.8845	0.8414	0.8206	0.7524
0.4	0.7690	0.6892	0.6524	0.5680

k	G.	N.	L.	LA.
0.6	0.6536	0.5486	0.5040	0.4280
0.8	0.5381	0.4238	0.3796	0.3226
1.0	0.4226	0.3174	0.2804	0.2432
1.2	0.3086	0.2302	0.2038	0.1832
1.4	0.2267	0.1616	0.1462	0.1380
1.6	0.1736	0.1096	0.1040	0.1040
1.8	0.1372	0.0718	0.0736	0.0784
2.0	0.1111	0.0456	0.0508	0.0592
2.2	0.0918	0.0278	0.0364	0.0446
2.4	0.0772	0.0164	0.0254	0.0336
2.6	0.0657	0.0094	0.0178	0.0254
2.8	0.0567	0.0052	0.0124	0.0188
3.0	0.0494	0.0026	0.0086	0.0144

TABLE II

$P(/X/ \geq k)$

k	G.T.	C.	k	G.T.	C.
0.2	0.8845	0.8560	1.6	0.1736	0.1128
0.4	0.7690	0.7156	1.8	0.1372	0.0660
0.6	0.6536	0.5824	2.0	0.1086	0.0336
0.8	0.5381	0.4596	2.2	0.0666	0.0138
1.0	0.4226	0.3496	2.4	0.0519	0.0036
1.2	0.3086	0.2544	2.6	0.0236	0.0004
1.4	0.2267	0.1756	2.7662	0.0000	0.0000

TABLE III

$P(|X| \geq k)$

k	G.T.	T.	k	G.T.	T.
0.2	0.8845	0.8630	1.6	0.1736	0.1202
0.4	0.7690	0.6998	1.8	0.1326	0.0702
0.6	0.6536	0.5698	2.0	0.0918	0.0336
0.8	0.5381	0.4732	2.2	0.0509	0.0104
1.0	0.4226	0.3500	2.4	0.0101	0.0004
1.2	0.3086	0.2602	2.4495	0.0000	0.0000
1.4	0.2267	0.1836			

3.8 Other Inequalities

In this section we briefly mention some other probability bounds which have been obtained for random variables which are restricted in some way. Barlow and Marshall [2,3] obtained a set of inequalities which are of practical use. Let X be a nonnegative r.v. such that $\log(1 - F(x))$ is either concave or convex on $[0,\infty)$; if $\log(1 - F(x))$ is concave, then $F(x)$ is said to have increasing hazard rate while if $\log(1 - F(x))$ is convex $F(x)$ is said to have decreasing hazard rate. Let us consider $q(x) = f(x)/[1 - F(x)]$. $q(x)$ is increasing (decreasing) iff $\log(1 - F(x))$ is concave (convex). $q(x)$ is called the hazard rate and when considering certain life expectancy problems involving either human beings or mechanical components, the authors point out that $q(x)dx$ denotes the conditional probability that A will die in time $(x, x + dx)$ given that A has survived till time x. Using geometrical arguments, Barlow and Marshall are able to obtain sharp probability bounds for distributions having either increasing or decreasing hazard rate.

In chapter one we mentioned that Winsten [125] had obtained probability inequalities in terms of the mean range, w_n, of a sample of size n. Winsten also obtained probability bounds for unimodal symmetric distributions in terms of w_n. The inequalities which he obtained are sharp; Winsten is able to construct distributions which attain equality.

Shohat [105] considers bounded random variables whose distributions are either $\cap$ shaped continuous symmetric, $\cap$ shaped continuous asymmetric, $\cup$ shaped continuous symmetric or $\cup$ shaped continuous asymmetric. He obtained probability bounds for $1 - P(|X| \geq k)$ in terms of μ_{2s}, $s > 1$, k and the upper and lower boundary points of X in each of the above cases.

Mallows [76] has offered a method of obtaining probability bounds through the use of "extremal distributions". He defines a distribution function to be smooth of order k with bound λ, i.e. to satisfy the smoothness condition (k,λ), if

i) the $(k + 1)$th derivative of $F(x)$ exists and is continuous for all x;

ii) there exists $k + 2$ numbers $\beta_0,\beta_1,\ldots,\beta_{k+1}$ such that

$$0 < (-1)^i F^{(k+1)}(x) < \lambda, \quad \beta_i < x < \beta_{i+1}, \quad i = 0,1,\ldots,k.$$

β_0 and β_{k+1} are the two end points of a bounded distribution; when X is unbounded Mallows takes $\beta_0 = -\infty$, $\beta_{k+1} = +\infty$. Mallows describes a method of obtaining two functions $L(x)$ and $U(x)$ such that the distribution function which satisfies the 2m moment conditions of the problem and the smoothness condition (k,λ), i.e. satisfies $(2m,k,\lambda)$ also satisfies the inequality

$$L(x) < F(x) < U(x).$$

To obtain L(x) and U(x) for $\lambda = \infty$, he constructs certain "extremal distributions", E(x) , which satisfy the following conditions. When $x_0, \ldots, x_m$ are the distinct values of X arranged in a non-decreasing order, then,

i) $\int_{-\infty}^{\infty} x^r dE(x) = \mu_r' = \int_{-\infty}^{\infty} x^r dF(x), \quad r = 0,1,\ldots,2m:$

ii) $$E(x) = \begin{cases} 0, & x \leq x_0, \\ 1, & x \geq x_m; \end{cases}$$

iii) in each interval (x_i, x_{i+1}), E(x) is a segment of a polynomial of degree $\leq k$;

iv) at each x_i, $i = 0,1,\ldots,m$, $E^{(k-n_i)}(x)$ has a simple discontinuity: n_i is called the characteristic number of x_i, $i = 0,1,\ldots,m$ and satisfies $n_i \geq 0$, $\sum_{i=0}^{m} n_i = k$. $(n_0,\ldots,n_m)$ is called the character of E(x);

v) E(x) is the limit of a sequence of distribution functions each of which satisfies the smoothness condition (k,∞).

Mallows shows that the distribution function F(x) intersects any extremal distribution function in at most 2m + 1 points and probability bounds can thus be found by examining the E(x)'s over various intervals which contain F(x). Mallows has conjectured the existence of L(x) and U(x) if the extremal distributions satisfy some restricted conditions, but a general solution of the problem $(2m,k,\infty)$ does not exist. As a particular case he shows that when a r.v. X, with zero mean and variance 1, satisfies the smoothness condition $(1,\infty)$, i.e. it is unimodal, then

$$P(X > k) \leq \begin{cases} \dfrac{3 - k^2}{3(1+x^2)}, & 0 \leq k \leq (5/3)^{\frac{1}{2}}, \\ \dfrac{4}{9(1+k^2)}, & k > (5/3)^{\frac{1}{2}}. \end{cases}$$

Mallows, in a later paper [77], extended his definition of a smooth distribution function and offered an alternate method of constructing extremal distributions so as to obtain sharp probability bounds for smooth and bell-shaped distributions. The interested reader may refer to the papers by Mallows [76,77] for a thorough study of this subject or he may refer to Karlin and Studden [60, page 498] for an extension of some of Mallows' results.

CHAPTER FOUR: INEQUALITIES FOR MULTIVARIATE DISTRIBUTIONS

4.1 Introduction

Let $X = (X_1,\ldots,X_n)$ be a random vector defined on $(\Omega,\mathcal{Q},P)$ where $\Omega = R_n$. In this chapter we shall obtain upper probability bounds for $P(X\in A)$, $A\in\mathcal{Q}$, in terms of known moments of the distribution. The inequalities obtained for sets in R_1 cannot be extended to sets in R_n. Using Theorem 2.3.1 we shall concentrate on the development of a general theorem for sets in R_n and we shall show how this theorem enables us to obtain other results, e.g. the inequality in section 4.5 will be obtained from the general theorem by the change of variable technique.

The general theorem will indicate that the solution to the problem of obtaining sharp upper bounds for sets in R_n lies in the solving of a matrix equation; the equation, however, does not always have a simple general solution.

We shall also briefly review some of the earlier work which has been done.

4.2 A General Theorem

Through the use of Theorem 2.3.1 we have been able to obtain many of the inequalities in chapters two and three. In the formulation of the theorem Ω was taken as any abstract space; hence in particular let $\Omega = R_n$ and let $X = (X_1,\ldots,X_n)$ be a random vector defined on $(\Omega,\mathcal{Q},P)$. For $T\in\mathcal{Q}$ let $f(x) = f(x_1,\ldots,x_n)$ be a nonnegative function on R_n such that $f(x) \geq 1$ for all $x\in T$.

$$P(X\in T) \leq \int_{x\in T} f(x)dP + \int_{x\notin T} f(x)dP = Ef(X). \tag{4.2.1}$$

In accordance with Theorem 2.3.1 let f(x) be a nonnegative function of the form

$$f(x) = a_0 + xa' + xAx', \tag{4.2.2}$$

where a_0 is a constant, a is a 1 x n row vector, A is a symmetric n x n matrix and ' denotes transpose. Let T be a symmetric region in R_n such that T' is an open bounded symmetric rectangle and $f(x) \geq 1$ for all $x\epsilon T$. Let $\mathcal{P}$ be a family of probability measures on $(\Omega,\mathcal{A})$ such that

$$\mathcal{P}(M) = \left\{P:P\epsilon\mathcal{P}, E(X_i) = 0, i=1,\ldots,n, E(X_iX_j)=\pi_{ij}, i,j=1,\ldots,n\right\}, \tag{4.2.3}$$

where $\pi_{ii}' = \pi_i^2$, $i = 1,\ldots,n$, $\Pi = \left(\pi_{ij}\right)$, and $M = (1,0,\ldots,0,\pi_1^2,\pi_{12},\ldots,\pi_n^2)$. is an interior point of the moment space

$$\mathcal{M} = (1, E(X_i) = 0, i = 1,\ldots,n, E(X_iX_j) = \pi_{ij}, i,j=1,\ldots,n, P\epsilon\mathcal{P}).$$

Since T is a symmetric set, (4.2.2) becomes

$$a_0 + xa' + xAx' \geq I_T(x), \tag{4.2.4}$$

$$a_0 - xa' + xAx' \geq I_T(x). \tag{4.2.5}$$

From the above equations we see that a is the null vector and we can rewrite f(x) as

$$f(x) = a_0 + xAx' \geq I_T(x). \tag{4.2.6}$$

$Ef(X) = a_0 + trA\Pi$ and by Theorem 2.3.1

$$P(X\epsilon T) = \inf[a_0 + trA\Pi], \tag{4.2.7}$$

where inf is taken with respect to all polynomials satisfying (4.2.6).

Since $P(X\epsilon T) \leq 1$, $0 \leq a_0 \leq 1$. Now

$$a_0 + xAx' \geq I_T(x) \Rightarrow xAx'/(1 - a_0) \geq I_T(x). \tag{4.2.8}$$

By our initial assumptions we note that $g(x) = xAx'/(1 - a_0)$ satisfies all the necessary conditions of the nonnegative function $f(x)$, and that $Eg(X) = [trA\Pi]/(1 - a_0)$. Also, if $P(X\epsilon T) < 1$, then, by the above,

$$[trA\Pi]/(1 - a_0) = trA\Pi + a_0([trA\Pi]/(1 - a_0)) \leq tr\, A\Pi + a_0. \qquad (4.2.9)$$

Thus for $a_0 > 0$, $g(x)$ would provide a lower bound than $f(x)$; therefore, $a_0 = 0$ and

$$f(x) = xAx'. \qquad (4.2.10)$$

$$P(X\epsilon T) \leq \min(tr\, A\Pi), \qquad (4.2.11)$$

where min is taken over all $A\epsilon\mathcal{A}$ and

$$\mathcal{A} = \left\{A: xAx' \geq I_T(x), \text{ for all } x\epsilon R_n\right\}. \qquad (4.2.12)$$

We note that A, as defined above for the set T, is positive definite. $\mathcal{A}$ is closed, convex and bounded from below.

To illustrate the above, let us consider a random vector $Y = (Y_1, Y_2)$ such that $E(Y) = 0$ and $E(Y_iY_j) = \sigma_{ij}, i,j, = 1,2$. Let T' denote the set $\left\{\left|y_i/k_i\sigma_i\right| < 1,\ i = 1,2\right\}$ and consider a nonnegative function

$$f(y_1,y_2) = \frac{1}{1-a^2}\left(\frac{y_1^2}{k_1^2\sigma_1^2} - \frac{2ay_1y_2}{k_1k_2\sigma_1\sigma_2} + \frac{y_2^2}{k_2^2\sigma_2^2}\right) = xAx', \qquad (4.2.13)$$

where $X = (X_1,X_2) = (Y_1/k_1\sigma_1, Y_2/k_2\sigma_2), a^2 < 1$ and $A = \begin{pmatrix}1 & a\\ a & 1\end{pmatrix}^{-1}$.

(4.2.13) ≥ 1 for all $x\epsilon T$. By (4.2.11),

$$P(/X_1/ \geq 1 \text{ or } /X_2/ \geq 1) \leq E(XAX') = \frac{1}{1-a^2}\left(\frac{1}{k_1^2} - \frac{2a\rho}{k_1k_2} + \frac{1}{k_1^2}\right), \qquad (4.2.14)$$

where $\rho = \sigma_{12}/\sigma_1\sigma_2$. Minimizing the right hand side of (4.2.14) with respect to a, we obtain

$$a = \frac{k_1^2 + k_2^2 - [(k_1^2 + k_2^2)^2 - 4\rho^2 k_1^2 k_2^2]^{\frac{1}{2}}}{2\rho k_1k_2}. \qquad (4.2.15)$$

Substituting this value of a into (4.2.14) gives

$$P(/Y_1/ \geq k_1\sigma_1 \text{ or } /Y_2/ \geq k_2\sigma_2) \leq \frac{k_1^2 + k_2^2 + [(k_1^2 + k_2^2)^2 - 4\rho^2 k_1^2 k_2^2]^{\frac{1}{2}}}{2k_1^2 k_2^2}. \quad (4.2.16)$$

The above is Lal's inequality [68].

When $k_1 = k_2 = k$, (4.2.16) becomes

$$P(/Y_1/ \geq k\sigma_1 \text{ or } /Y_2/ \geq k\sigma_2) \leq \frac{1 + (1 - \rho^2)^{\frac{1}{2}}}{k^2}. \quad (4.2.17)$$

The above is Berge's inequality [8]; through an example which attained equality in (4.2.17), Berge has shown that (4.2.17) is sharp.

The method just illustrated can be extended to an n dimensional random vector $Y = (Y_1, \ldots, Y_n)$ by choosing

$$f(y) = \frac{1}{|A|}\left\{\sum_{i=1}^{n} \frac{y_i^2}{k_i^2\sigma_i^2} - 2a\sum_{i \neq j=1}^{n} \frac{y_i y_j}{k_i k_j \sigma_i \sigma_j}\right\}, \quad (4.2.18)$$

where
$$A = \begin{pmatrix} 1 & -a & \ldots & -a \\ -a & 1 & \ldots & -a \\ \cdot\cdot & \cdot\cdot & \ldots & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \ldots & 1 \end{pmatrix}$$

is a n x n positive definite matrix; i.e. $|A| = [1-(n-1)a](1+a)^{n-1} > 0$. For n odd $a < 1/(n-1)$ and for n even $-1 < a < 1/(n-1)$. For an n dimensional rectangle T' defined by $\{/y_i/ < k_i\sigma_i, i=1,\ldots,n\}$, $f(y) \geq 1$ for all $y \in T$. Therefore

$$P(Y \in T) = P(/Y_i/ \geq k_i\sigma_i \text{ for some } i) \leq Ef(Y)$$

$$= \frac{1}{|A|}\left\{\sum_{i=1}^{n} \frac{1}{k_i^2} - 2a\sum_{i=j=1}^{n} \frac{\rho_{ij}}{k_i k_j}\right\}. \quad (4.2.19)$$

Minimizing the right hand side of (4.2.19) with respect to a, we obtain an upper bound.

The drawback in the above method is that a sharp inequality can only be obtained when n = 2. In doing the above examples matrix A was chosen arbitrarily; let us see if can characterize $\mathcal{A}$, the set of all positive definite matrices A satisfying (4.2.12) for some given symmetric region $T \epsilon R_n$ such that T' is an open bounded rectangle in R_n defined by $\{/x_i/ < 1, i=1,\ldots,n\}$. $P(X \epsilon T)$ is equivalent to $P(/X_i/ \geqslant 1$ for some i). The development will be based on the works of Olkin and Pratt [91] and Whittle [123]. Note $E(X'X) = \Pi$.

Since A is positive definite $B = A^{-1}$ is also positive definite.

$$A = \begin{pmatrix} a_{11} & a \\ a' & A_{22} \end{pmatrix}, \; B = \begin{pmatrix} b_{11} & b \\ b' & B_{22} \end{pmatrix} = A^{-1},$$

where a and b are 1 x n-1 vectors. BA = I implies

1) $b_{11}a + bA_{22} = (0,\ldots,0); \; b = -b_{11}aA_{22}^{-1};$ (4.2.20)

2) $b_{11}a_{11} + ba' = 1$; therefore by (4.2.20),

$$b_{11}(a_{11} - aA_{22}^{-1}a') = 1. \tag{4.2.21}$$

Let $z = (x_2,\ldots,x_n)$.

$$\begin{aligned}(x_1,z)\, A(x_1,z)' &= a_{11}x_1^2 + 2az'x_1 + zA_{22}z' \\ &= x_1^2a_{11} - x_1^2aA_{22}a' + (z + aA_{22}^{-1}x_1)A_{22}(z + aA_{22}^{-1}x_1)' \\ &= x_1^2b_{11}^{-1} + (z - b_{11}^{-1}bx_1)A_{22}(z - b_{11}^{-1}bx_1)'. \end{aligned} \tag{4.2.22}$$

Since A_{22} is positive definite

$$xAx' \geqslant x_1^2 b_{11}^{-1}, \tag{4.2.23}$$

and by (4.2.22) the minimum value of xAx' occurs when $b_{11}z = bx_1$. If $x_1 = 1$, then (4.2.10) and (4.2.12) tell us that $b_{11}^{-1} \geqslant 1$. In an equivalent way we obtain the conditions $b_{ii}^{-1} \geqslant 1, i=1,\ldots,n$. We summarize the above results in the following lemma.

Lemma 4.2.1 [91]. $A \epsilon \mathcal{A}$ iff $B = A^{-1}$ is positive definite and $b_{ii} \leqslant 1$, $i=1,\ldots,n$.

Since $E(XAX') = tr\, A\Pi$ is a linear function of A, the minimum value of tr $A\Pi$ is realized at an extreme point of $\mathcal{A}$.

Theorem 4.2.1 [91]. A is extreme in $\mathcal{A}$ iff $B = A^{-1}$ is positive definite and $b_{ii} = 1, i=1,\ldots,n$.

Proof: Sufficiency. Assume $B = A^{-1}$ is positive definite and $b_{ii}= 1$, $i=1,\ldots,n$. If A is not extreme in $\mathcal{A}$, then A can be expressed as a convex combination of A_1 and A_2, where $A_1, A_2 \in \mathcal{A}$. Let

$$A = \tfrac{1}{2}(A_1 + A_2),\ A_1 \neq A_2 \ . \tag{4.2.24}$$

$$1 = \min_{x_i=1} xAx' = b_{ii}^{-1} \geq \tfrac{1}{2}(\min_{x_i=1} xA_1x' + \min_{x_i=1} xA_2x'). \tag{4.2.25}$$

By (4.2.23) $\min_{x_i=1} x\, A_1x' \geq 1$ and $\min_{x_i=1} x\, A_2x' \geq 1$. Thus by (4.2.25) $\min_{x_i=1} xA_1x' = 1 = \min_{x_i=1} xA_2x'$ and by (4.2.22) the min must occur at an identical point and A_1 and A_2 must be identical. Thus we have a contradiction of (4.2.24) and $A \in \mathcal{A}$ is extreme.

Necessity. If $A \in \mathcal{A}$ is extreme in $\mathcal{A}$, then, by Lemma 4.2.1, $B = A^{-1}$ is positive definite. If B is positive definite but $b_{11} < 1$, we would like to show that A is not extreme and thus obtain a contradiction.

$$B(\delta) = B + \begin{pmatrix} \delta & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} b_{11} + \delta & b \\ b' & B_{22} \end{pmatrix} .$$

By Lemma 4.2.1 $(B(\delta))^{-1} = A(\delta) \in \mathcal{A}$ for small δ. Also, since $AB = I$, corresponding to (4.2.20) and (4.2.21), we get

$$a = -a_{11} b B_{22}^{-1} , \tag{4.2.26}$$

$$a_{11}(b_{11} - bB_{22}^{-1}b') = 1. \tag{4.2.27}$$

Using (4.2.26) and (4.2.27) we can write $(B(\delta))^{-1}$ as a function of $a_{11}(\delta)$,

$$A(\delta) = a_{11}(\delta) \begin{pmatrix} 1 & -bB_{22}^{-1} \\ -B_{22}^{-1}b' & B_{22}^{-1}b'bB_{22}^{-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & B_{22}^{-1} \end{pmatrix}. \tag{4.2.28}$$

Since $B(\delta) = (A(\delta))^{-1}$, we can, by matrix algebra and (4.2.27) write

$$a_{11}(\delta) = 1/(b_{11} + \delta - bB_{22}^{-1}b') = a_{11}/(1 + \delta a_{11}). \qquad (4.2.29)$$

For $0 < \theta < 1$, we can choose $\delta_1, \delta_2, \delta_1 \neq \delta_2$, to satisfy

$$\frac{\theta\delta_1}{1 + \delta_1 a_{11}} + \frac{(1 - \theta)\delta_2}{1 + \delta_2 a_{11}} = 0$$

which is equivalent to

$$\theta a_{11}(\delta_1) + (1 - \theta)a_{11}(\delta_2) = a_{11}.$$

Hence, by the above,

$$\theta A(\delta_1) + (1-\theta)A(\delta_2) = A .$$

Thus A is a convex combination of $A(\delta_1)$ and $A(\delta_2)$ and thus not extreme.

Therefore, $b_{11} = 1$ and in general $b_{ii} = 1$, $i=1,\ldots,n$. Q.E.D.

In Theorem 4.2.1 we justified having chosen A as we did in the first example of this section; i.e. ones along the diagonal of A^{-1}. More generally, if from the set $\mathcal{A}$ we consider matrices of the form

$$A = [(1 - a)I + ae'e]^{-1}, \quad -1/(n - 1) < a < 1, \qquad (4.2.30)$$

where I is the identity matrix and $e = (1,\ldots,1)$ is a 1 x n vector, we can see from the above theorem that, providing A is positive definite, A will be extreme in $\mathcal{A}$. It is easy to see that A as defined in (4.2.30) is positive definite. Let $B = A^{-1}$; $|B| = (1+(n-1)a)(1-a)^{n-1} > 0$ provided that $(1+(n-1)a) > 0$ and $(1-a) > 0$; i.e. $1 > a > -1(n-1)$. Therefore, by (4.2.30) B is positive definite and so is A^{-1}.

Theorem 4.2.2 [91]. Let Y be a n dimensional random vector such that $E(Y) = 0$ and $E(Y'Y) = \Sigma$. Let T' be defined by $\left\{ |Y_i|/k_i\sigma_i = |X_i| < 1, i=1,\ldots,n \right\}$. Then,

$$P(X \notin T) = P(|X_i| \geq 1 \text{ for some } i)$$

$$\leq \frac{(n-1)t}{n} - \frac{(n-2)u}{n^2} + \frac{2[u(nt-u)(n-1)]^{\frac{1}{2}}}{n^2}$$

$$= \frac{[u^{\frac{1}{2}} + [(nt-u)(n-1)]^{\frac{1}{2}}]^2}{n^2}, \qquad (4.2.31)$$

where $E(X'X) = \Pi$, $t = \mathrm{tr}\,\Pi$ and $u = e\Pi e'$.

Remark. Birnbaum and Marshall [15] offer an upper bound on $P(X \in T)$ when only certain terms of Π are known. Their inequality is not sharp.

Proof: $\mathrm{tr}[(1-a)I + ae'e]^{-1}\Pi = \mathrm{tr}[(I - \alpha e'e)\Pi]/(1 - a)$

$$= (t - \alpha u)/(1-a), \qquad (4.2.32)$$

where $\alpha = a/(1 + (n-1)a)$. Differentiating (4.2.32) with respect to α and equating to 0 we get

$$a = \frac{t \pm [u(nt - u)/(n - 1)]^{\frac{1}{2}}}{u - (n - 1)t}.$$

By the condition of (4.2.30) only

$$a = \frac{t - [u(nt - u)/(n - 1)]^{\frac{1}{2}}}{u - (n - 1)t} \qquad (4.2.33)$$

is a satisfactory solution. Substituting (4.2.33) into (4.2.32) will give (4.2.31). The value of a in (4.2.33) insures a minimum value since $(4.2.32) \to \infty$ as either $a \to 1$ or $a \to -1/(n-1)$.

If $n = 2$, (4.2.31) is identical to (4.2.15).

Theorem 4.2.2, however, has not yet given us an upper bound which is sharp for all n; i.e. $\min \mathrm{tr}\, A\Pi$, for all A satisfying (4.2.30) does not necessarily give a sharp upper bound.

We restate the problem in an alternate fashion in an attempt to establish a unique sharp upper bound. We must now minimize $\mathrm{tr}\, B^{-1}\Pi$ for all matrices $B \in \beta$, the set of positive definite matrices B such that $b_{ii} = 1, i=1,\ldots,n$.

Lemma 4.2.2 [91]. $\mathrm{tr}\, B^{-1}\Pi$ is a strictly convex function of B for $B \in \beta$, and has a unique minimum which occurs at an interior point B_0 of β.

Proof: $B(t)$ is a linear function of a variable t; dB/dt is a symmetric matrix and $d^2B/dt^2 = 0$.

$$\frac{d}{dt} \operatorname{tr} B^{-1}\Pi = -\operatorname{tr} B^{-1}\left(\frac{dB}{dt}\right)B^{-1}\Pi$$

$$\frac{d^2}{dt^2} \operatorname{tr} B^{-1}\Pi = \frac{d}{dt}\left[-\operatorname{tr} B^{-1}\left(\frac{dB}{dt}\right)B^{-1}\Pi\right]$$

$$= 2 \operatorname{tr} B^{-1}\left(\frac{dB}{dt}\right)B^{-1}\left(\frac{dB}{dt}\right)B^{-1}\Pi > 0,$$

since B and Π are positive definite. $\operatorname{tr} B^{-1}\Pi$ is therefore a strictly convex function of B and it must have a unique minimum. We must show that this unique minimum occurs at an interior point B_o of β. Now,

$$\operatorname{tr} B^{-1}\Pi \geq \operatorname{tr} B^{-1}(\text{smallest characteristic root of } \Pi). \quad (4.2.34)$$

By (4.2.34) $\operatorname{tr} B^{-1}\Pi \to \infty$ as $B \to$ boundary of β; therefore, B_o is an interior point of β.

Lemma 4.2.3 [91]. B_o is the unique point of β such that $B_o^{-1}\Pi B_o^{-1}$ is diagonal.

Proof: In order to determine B_o, the unique minimum, we must take derivatives and equate to zero. Let b_{ij} be a non-diagonal element of $B, i,j=1,\ldots,n, i\neq j$.

$$\frac{d}{db_{ij}} \operatorname{tr} B^{-1}\Pi = -\operatorname{tr} B^{-1}\left(\frac{dB}{db_{ij}}\right)B^{-1}\Pi = -\operatorname{tr} \frac{dB}{db_{ij}} B^{-1}\Pi B^{-1} = -2c_{ij} = 0, \quad (4.2.35)$$

where $C = B^{-1}\Pi B^{-1}$ and dB/db_{ij} is a symmetric matrix whose (i,j)th and (j,i)th elements are one and whose other elements are 0. Thus by (4.2.35) $B_o^{-1}\Pi B_o^{-1}$ must be diagonal.

If we combine Lemmas 4.2.2 and 4.2.3 with equation (4.2.11), we obtain the following sharp upper bound.

Theorem 4.2.3 [91]. $P(|Y_i| \geq k_i\sigma_i$ for some $i) = P(|X_i| \geq 1$ for some $i)$

$$\leq \text{tr } B_o^{-1}\pi = \text{tr } B_o^{-1}\pi B_o^{-1} B_o = \text{tr } B_o^{-1}\pi B_o^{-1}, \qquad (4.2.36)$$

where B_o is the unique positive definite matrix such that $b_{ii} = 1$, $i=1,\ldots,n$ and such that $B_o^{-1}\pi B_o^{-1}$ is diagonal.

Remark: Through an example which attains equality in (4.2.36), Olkin and Pratt illustrate that for $\text{tr } B_o^{-1}\pi \leq 1$, the inequality is sharp.

A problem does arise in that the matrix B_o cannot always be determined from π; i.e. if D is a positive definite diagonal matrix such that

$$D = B_o^{-1}\pi B_o^{-1} ,$$

then

$$B_oDB_o = \pi \qquad (4.2.37)$$

does not have a general solution which can be obtained by ordinary matrix calculus.

Example: Let $X = (X_1,\ldots,X_n)$ be a random vector such that $E(X) = 0$, $E(X_i) = \sigma_i^2$, $i=1,\ldots,n$ and $E(X_iX_j)$ are unknown for $i \neq j$. By Boole's inequality and the univariate Tchebycheff inequality

$$\begin{aligned} P(|X_1| \geq 1 \text{ or } |X_2| \geq 1 \text{ or } \ldots \text{ or } |X_n| \geq 1) &\\ = P(|X_i| \geq 1 \text{ for some } i) &\\ = P(\bigcup_{i=1}^{n} (|X_i| \geq 1)) &\\ \leq \sum_{i=1}^{n} P(|X_i| \geq 1) \leq \sum_{i=1}^{n}\sigma_i^2 . & \qquad (4.2.38) \end{aligned}$$

If $B_o = I$ and the covariance values of π are zero, then $B_o^{-1}\pi B_o^{-1}$ is a diagonal matrix π and by Theorem 4.2.3, inequality (4.2.38) is sharp provided that $\sum_{i=1}^{n} \sigma_i^2 \leq 1$.

4.3 A One-Sided Multivariate Inequality

Corresponding to the theory of section 4.2 of this paper, Marshall and Olkin [80] offer a one-sided multivariate generalization of the Tchebycheff inequality for a random vector $X = (X_1,\dots,X_n)$ with 0 mean and with variance covariance matrix of the form

$$\Pi = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho \\ \vdots & & & \\ \sigma^2\rho & \cdot & \cdots & \sigma^2 \end{pmatrix} = \sigma^2[(1-\rho)I + \rho e'e]; \; -1/(n-1) < \rho < 1. \quad (4.3.1)$$

(4.3.1) is of the form (4.2.30).

Let T' denote a region defined by $\{x_i < 1,\ i=1,\dots,n\}$. We wish to determine a sharp upper bound for $P(X\in T) = P(X_i \geq 1$ for some $i, i=1,\dots,n)$. For a positive definite $n \times n$ matrix A and a $1 \times n$ vector $b = (b_1,\dots,b_n)$, let us, by Theorem 2.3.1, consider a nonnegative function

$$f(x) = a_0 + xb' + xAx', \quad (4.3.2)$$

where $f(x) \geq 1$ for all $x\in T$. Because A is positive definite we can rewrite (4.3.2) as

$$f(x) = (x - a)A(x - a)' + C, \quad (4.3.3)$$

where $a = (a_1,\dots,a_n) = -\frac{1}{2}bA^{-1}$ and $C = a_0 - aAa' \geq 0$. Similar to what was done to equation (4.2.6), we can show that C = 0. (4.3.2) can be rewritten as

$$f(x) = (x - a)A(x - a)', \quad (4.3.4)$$

such that $f(x) \geq I_T(x)$.

$$Ef(X) = \text{tr } A(\Pi + a'a), \quad (4.3.5)$$

and the upper bound for $P(X\in T)$ will be given by min tr $A(\Pi + a'a)$ subject to $f(x) \geq I_T(x)$.

Theorem 4.3.1 [80]. Let $X = (X_1,\dots,X_n)$ be a random vector with $E(X) = 0$ and $E(X'X) = \Pi$ where Π is defined in (4.3.1). If i) $1 - \sigma^2 t > 0$, ii) $n \geq \sigma^2(n - 1)(1 + t)$, where $t = (n - 1)(1 - \rho) - 1$, then,

$$P(X\epsilon T) \leq \frac{n\sigma^2\{[(1+(n-1)\rho)(1+\sigma^2-\sigma^2(n-1)(1-\rho))]^{\frac{1}{2}}+(n-1)(1-\rho)^{\frac{1}{2}}\}^2}{\{n+\sigma^2[1+(n-1)\rho]\}^2}, \quad (4.3.6)$$

otherwise $P(X\epsilon T) \leq 1$.

Remark: Through an example it is illustrated that inequality (4.3.6) is sharp. This fact will justify an intuitive guess and assumption which are used in the proof of the theorem. Karlin and Studden [60, page 520, Theorem 5.1] justify the assumption.

Proof: Let $D_{1-a} = \text{diag}(1-a_1,\ldots,1-a_n)$, $Z = (X - a)(D_{1-a})^{-1}$, $A^* = D_{1-a}AD_{1-a}$ and $B = (A^*)^{-1}$, where D_{1-a} is defined in our notation. By matrix manipulation $A = (D_{1-a})^{-1}A^*(D_{1-a})^{-1}$ and thus

$$\text{tr } A[\Pi + a'a] = \text{tr } B^{-1}(D_{1-a})^{-1}(\Pi + a'a)(D_{1-a})^{-1}. \quad (4.3.7)$$

(4.3.4) becomes

$$f(z) = zD_{1-a}(D_{1-a})^{-1}A^*(D_{1-a})^{-1}D_{1-a}z' = zA^*z'. \quad (4.3.8)$$

$f(z) \geq 0$ for all z and $f(z) \geq 1$ for all $z\epsilon T$. By Theorem 4.2.1, the minimum upper bound will be attained iff $(A^*)^{-1} = B$ is positive definite and $b_{ii} = 1, i = 1,\ldots,n$. If we consider $\mathcal{Q}$, the class of matrices Q of the form (4.2.30).

$$Q = [(1-q)I + qe'e], \quad (4.3.9)$$

we have already shown that Q is positive definite iff $-1/(n-1) < q < 1$.

Now Π/σ^2 is of the form (4.3.9); intuitively, through the symmetry conditions, we expect $B\epsilon\mathcal{Q}$ and $a = \alpha e$, $[\alpha < 1]$.

$$B = [(1-b)I + be'e] \quad (4.3.10)$$

Let P be an n x n orthoganol matrix such that each element of the first row $p_{1j} = 1/\sqrt{n}$, $j = 1,\ldots,n$.

$$PQP' = \text{diag } (1 + (n-1)q, 1-q,\ldots,1-q). \quad (4.3.11)$$

Using (4.3.10) and (4.3.11) we can write (4.3.7) as

$$\text{tr } A[\Pi + a'a] = \frac{\text{tr } (PBP')^{-1}[P\Pi P' + \alpha^2 Pe'eP']}{(1-\alpha)^2}$$

$$= \frac{n[\alpha^2 + \sigma^2 + b(\sigma^2 t - \alpha^2)]}{(1 - \alpha)^2(1 - b)[1 + (n - 1)b]}, \tag{4.3.12}$$

where $t = (n - 1)(1 - \rho) - 1$. We must now minimize (4.3.12) with respect to α and b. Minimizing with respect to the conditions of the hypothesis

$$\alpha = -\sigma^2(1 + bt)/(1 - b),\ b = -\frac{1}{t} + \frac{[(1 + t)(n - 1 - t)]^{\frac{1}{2}}}{t[(n - 1)(1 - \sigma^2 t)]^{\frac{1}{2}}}.$$

The conditions i) and ii) of the theorem are required so as to insure that the roots of the minimizing equation for b are real and that B is positive definite. Substituting the values for α and b, respectively, in (4.3.12) we get (4.3.6).

Example: Let $X = (X_1, \ldots, X_n)$ be a random vector such that $E(X) = 0$, $E(X_i^2) = \sigma_i^2$, $i=1,\ldots,n$, and $E(X_i X_j)$ are unknown, $i \neq j$. By Boole's inequality and the univariate one-sided Tchebycheff inequality,

$$P(X_i \geq 1 \text{ for some } i) = P(\bigcup_{i=1}^{n}(X_i \geq 1))$$

$$= \sum_{i=1}^{n} P(X_i \geq 1) \leq \sum_{i=1}^{n} \sigma_i^2/(1 + \sigma_i^2). \tag{4.3.13}$$

We can obtain this sharp result from Theorem 4.3.1. For if $B = I$, we must minimize (4.3.7).

$$\operatorname{tr}(D_{1-a})^{-1}(\Pi + a'a)(D_{1-a})^{-1} = \sum_{i=1}^{n}(\sigma_i^2 + a_i^2)/(1 - a_i)^2. \tag{4.3.14}$$

Minimizing (4.3.14) with respect to each $a_i [a_i < 1]$,

$$2(1 - a_i)^2 a_i + (\sigma_i^2 + a_i^2)2(1 - a_i) = 0,\ i = 1, \ldots, n.$$

$$a_i = -\sigma_i^2 \text{ and } (4.3.14) = \sum_{i=1}^{n}\sigma_i^2(1 + \sigma_i^2)/(1 + \sigma_i^2)^2 = (4.3.13).$$

4.4. Arbitrary Rectangular Region in R_2

In equation (4.2.16) we obtained $P(X \in T)$ where T was a closed symmetric set whose complement T' was an open bounded symmetric rectangle in R_2. However, in calculating the upper bound we assumed that $E(X) = 0$ and that T' was symmetric about the mean. Let $Y = (Y_1, Y_2)$ be a random vector such that $E(Y) = (\mu_1, \mu_2)$ and $E(Y'Y) = \Sigma$.

Isii [54] removes the above two restrictions and considers determining $P(Y \in R)$ where R' is an open rectangle whose sides L_1, L_2, are parallel to the y_1 and y_2 axis and are in the ratio of $\sigma_1 : \sigma_2$, respectively. The new restriction added is that (μ_1, μ_2) lies within R' on one of the diagonals of R'. By means of a linear transformation $X_i = (Y_i - \mu_i)/\sigma_i$, $i = 1,2$, we obtain a new random vector $X = (X_1, X_2)$ such that $E(X) = 0$ and $E(X'X) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \Pi$. Our problem is now to determine $P(X \in T)$ where T' is an open square, obtained from the transforming of R', with sides of length $(\alpha + \beta)$ and with vertices $(-\alpha, -\alpha)$, $(-\alpha, \beta)$, (β, β) and $(\beta, -\alpha)$, $\beta \geq \alpha > 0$.

Theorem 4.4.1 [54]. Let X and T' be defined as above. The upper bounds of $P(X \in T)$ are as follows:

i) $$P(X \in T) \leq \lambda^2/[\lambda^2 + 1 + \rho], \quad \beta - \alpha \geq 2^{\frac{1}{2}}\lambda, \quad 2\alpha^2 > 1 - \rho \tag{4.4.1}$$

where $\lambda = [2^{\frac{1}{2}}\alpha(1 + \rho) + (2(1 - \rho^2)(\alpha^2 + \rho))^{\frac{1}{2}}]/[2\alpha^2 - (1 - \rho)]$;

ii) $$P(X \in T) \leq \frac{[(\beta - \alpha)^2 + 4 + (16(1 - \rho^2) + 8(1 - \rho)(\beta - \alpha)^2)^{\frac{1}{2}}]}{(\alpha + \beta)^2}, \tag{4.4.2}$$

if the conditions of i) are not satisfied and $\alpha\beta \geq 1$,
$2(\alpha\beta - 1)^2 \geq 2(1 - \rho^2) + (1 - \rho)(\beta - \alpha)^2$;

iii) in all other cases $P(X \in T) \leq 1$.

Proof: Let A be a positive definite matrix. By Theorem 2.3.1 we consider a nonnegative function

$$f(x) = f(x_1, x_2) = a_o + xb' + xAx', \tag{4.4.3}$$

where $b = (b_1,b_2)$ and $f(x) \geqslant I_T(x)$. As shown in section 4.3, (4.4.3) can be reduced to

$$f(x) = (x - a)A(x - a)', \; f(x) \geqslant I_T(x). \tag{4.4.4}$$

Our task is to determine min tr $A[\Pi + a'a]$ subject to the condition of (4.4.4). We can similarly define a function $f_1(x) = f(x_2,x_1)$ to satisfy (4.4.4).

Let us introduce a new function which is a convex combination of $f_1(x)$ and $f(x)$; $g(x) = \frac{1}{2}[f_1(x) + f(x)]$. $g(x) \geqslant 0$ for all x, $g(x) \geqslant 1$ for all $x \in T$ and $Ef(X) = Eg(X)$. By the symmetry conditions involved, we can write

$$g(x) = (x - a)A(x - a)' \geqslant I_T(x), \tag{4.4.5}$$

where $a = (m,m)$ and $A = \begin{pmatrix} c & -ct \\ -ct & c \end{pmatrix}$, $c > 0$, $/t/ < 1$. (4.4.6)

$/t/ < 1$ ascertains that A is positive definite.

$$\begin{aligned} \text{tr } A[\Pi + a'a] &= 2(cm^2(1-t) + c(1-t\rho)) \\ &= \lambda_1 u_1 + \lambda_2 u_2 + \lambda_1 \xi^2 , \end{aligned} \tag{4.4.7}$$

where $\lambda_1 = c(1 - t)$, $\lambda_2 = c(1 + t)$, $u_1 = (1 + \rho)$, $u_2 = (1 - \rho)$ and $\xi = 2^{\frac{1}{2}}m$. Similarly (4.4.5) becomes

$$g(x) = c[(x_1 - m)^2 + (x_2 - m)^2 - 2t(x_1 - m)(x_2 - m)]. \tag{4.4.8}$$

By now considering the extreme points of the square, (4.4.8) can be replaced by

$$\frac{2\lambda_1\lambda_2(\alpha + m)^2}{\lambda_1 + \lambda_2} \geqslant 1, \; \frac{2\lambda_1\lambda_2(\beta - m)^2}{\lambda_1 + \lambda_2} \geqslant 1. \tag{4.4.9}$$

If we now minimize (4.4.7) with respect to the conditions in (4.4.9), we obtain the required results.

4.5 A Symmetric Convex Polygon in R_2

We know that the intersection of a finite number of closed half planes is a convex polygon. Thus, a convex polygon, symmetric about (0,0), can be represented by the intersection of a finite number of strips where the ith strip S_i is derived by rotating the strip $\{/X_1/ < w_i\}$ through an angle α_i, $0 \leq \alpha_i < \pi$, i=1,...,n. If $\alpha_1 = 0$, $\alpha_2 = \pi/2$, the intersection of two strips will give a rectangle and this problem has been solved in section 4.2. In the case when 3 strips intersect we obtain a hexagon.

Let us denote a bounded convex polygon by $T' = \bigcap_{i=1}^{n} S_i$; by DeMorgan's law $T = \bigcup_{i=1}^{n} S_i$. If $X = (X_1, X_2)$ is a random vector such that $E(X) = 0$ and $E(X'X) = \Pi$, then

$$Ef(X) = \text{tr } A\Pi, \tag{4.5.1}$$

where A is a 2 x 2 positive definite matrix and

$$f(x) = xAx' \geq I_T(x). \tag{4.5.2}$$

$$\mathcal{A} = \{ A: f(x) \geq 0;\ f(x) \geq 1 \text{ for } /x_1/ \geq w \} \tag{4.5.3}$$

By Lemma 4.2.1, $A \in \mathcal{A}$ iff $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix} = A^{-1}$ is positive definite and $b_{11} \leq w^2$. It is known [29, page 26] that the transformation matrix used to obtain S_2 by rotating S_1 through an angle α is of the form

$$P = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

Thus, by rotating S_1, A becomes PAP' and $b_{11} \leq w^2$ becomes

$$b_{11}\cos^2\alpha + b_{22}\sin^2\alpha + 2b_{12}\sin\alpha\cos\alpha \leq w^2.$$

If we have m rotations, then, for different w_i, we obtain the following set of conditions which B must satisfy such that $A \in \mathcal{A}$

$$b_{11}\cos^2\alpha_i + b_{22}\sin^2\alpha_i + 2b_{12}\sin\alpha_i\cos\alpha_i \leq w_i^2, \tag{4.5.4}$$

for i = 1,...,n.

As in section 4.2 we must determine min tr $B^{-1}\Pi$, where $B \in \beta$, the closed bounded and convex set of positive definite matrices satisfying (4.5.4). What are the extreme points of β ?

Lemma 4.5.1 [82]. If B is an extreme point of β , then equality in (4.5.4) holds for at least two values of i. Equality in three or more values of i determines B and three equations determine at least one extreme point of B.

Proof: If equality in (4.5.4) holds for i = 1 only, we would like to show that B is not extreme. Let us represent $B = (B_1 + B_2)/2$ as a convex combination of B_1, $B_2 \in \beta$ $(B_1 \neq B_2)$, where

$$B_1 = \begin{pmatrix} b_{11} - \epsilon & b_{12} \\ b_{12} & b_{22} + \delta \end{pmatrix}, \quad B_2 = \begin{pmatrix} b_{11} + \epsilon & b_{12} \\ b_{12} & b_{22} - \delta \end{pmatrix}.$$

If $\epsilon \cos^2 \alpha = \delta \sin^2 \alpha$, then (4.5.4) holds for i = 1. Also we can appropriately choose ϵ and δ small enough such that the inequality sign in (4.5.4) will hold for i = 2,...,m. Thus B is not extreme and by the contradiction we see that equality in (4.5.4) must hold for at least two values of i. If equality holds for three values of i for which the α_i's are distinct, then we have three equations in three unknowns which can be solved by Cramer's rule.

If equality in (4.5.4) holds for i = 1,2, then, b_{11} and b_{22} can be written in terms of b_{12}. (4.5.4) can be expressed as a linear equation in b_{12} and for some value b_{12} and some i = 3,...,m equality can be attained in (4.5.4).

Remark: If we let $u_i = (\cos \alpha_i)/w_i$ and $v_i = (\sin \alpha_i)/w_i$, then, (4.5.4) becomes

$$b_{11}u_i^2 + b_{22}v_i^2 + 2b_{12}u_iv_i \leq 1, \quad i = 1,\ldots,m. \qquad (4.5.5)$$

The boundary line of the strip S_i, defined by $x_1 \cos \alpha_i + x_2 \sin \alpha_i = w_i$, becomes $x_1u_i + x_2v_i = 1$. T' is now defined by

$$T' = \left\{ |x_1u_i + x_2v_i| < 1,\ i = 1,\ldots,m \right\}. \tag{4.5.6}$$

Let $\beta(S) = \left\{ B: xB^{-1}x' \geq 1 \text{ for all } x \in S \right\}$ for some $S \in R_2$. If B_o satisfies

$$\min_{B \in \beta(S)} \operatorname{tr} B^{-1}\Pi = \operatorname{tr} B_o^{-1}\Pi,$$

we call $\left\{ xB_o^{-1}x' < 1 \right\}$ the "best ellipse for S' " and we say that B_o is best for S'. For T' defined as in (4.5.6) Lemma 4.5.1 tells us that for each $S_i \cap S_j$ we must calculate the best B; if for a B the ellipse $\left\{ xB^{-1}x' < 1 \right\}$ lies in T', then this B is used to determine the bound $\operatorname{tr} B^{-1}\Pi$. If none of the $\left\{ xB^{-1}x' < 1 \right\}$ lie in T', we must, for each $S_i \cap S_j \cap S_k$ find a best B. Among these B we determine the one such that the ellipse $\left\{ xB^{-1}x' < 1 \right\}$ lies in T', and we use this matrix B to determine $\operatorname{tr} B^{-1}\Pi$. Lemma 4.5.1 tells us that there exists a B which is best for T'. A computational procedure to determine $\operatorname{tr} B^{-1}\Pi$ is outlined by Marshall and Olkin [82].

Consider the set $T' = \left\{ |x_1u_i + x_2v_i| < 1,\ i = 1,2 \right\}$. If the best ellipse for $S_1 \cap S_2$ is best for T', we obtain the following theorem.

Theorem 4.5.1 [82]. If $X = (X_1, X_2)$ is a random vector such that $E(X) = 0$ and $E(X'X) = \sigma_{ij}$, $i,j = 1,2$, then

$$P(X \in T) \leq \tfrac{1}{2}\left[c_1 + c_2 + \left[(c_1 - c_2)^2 + 4(\sigma_{11}\sigma_{22} - \sigma_{12}^2)(u_1v_2 - u_2v_1)^2\right]^{\frac{1}{2}}\right], \tag{4.5.7}$$

where $c_i = u_i^2\sigma_{11} + v_i^2\sigma_{22} + 2u_iv_i\sigma_{12}$, $i=1,2$.

Proof: Our result will be obtained from (4.2.16) by a direct change of variable. From section 4.2.1 we know that if $Y = (Y_1, Y_2)$ is a random vector such that $E(Y) = 0$ and $E(Y'Y) = \pi_{ij}$, then by (4.2.16),

$$P(X \in S) = P(/Y_1/ \geq 1 \text{ or } /Y_2/ \geq 1) \leq \tfrac{1}{2}[\pi_{11} + \pi_{22} + [(\pi_{11} + \pi_{22})^2 - 4\pi_{12}^2]^{\frac{1}{2}}]. \quad (4.5.8)$$

By means of the transformation matrix

$$M = \begin{pmatrix} u_1 & u_2 \\ v_1 & v_2 \end{pmatrix}^{-1}$$

we can transform S' into T'. Let $X = YM$; the corresponding nonnegative function $yB^{-1}y'$ associated with S becomes $x(M'BM)^{-1}x'$ and (π_{ij}) must be replaced by $(M')^{-1}(\sigma_{ij})M^{-1}$. If we calculate the elements of $(M')^{-1}(\sigma_{ij})M^{-1}$ and respectively substitute them for the elements of (π_{ij}) in (4.5.8), we shall obtain (4.5.7).

4.6 Convex Sets in R_n

Until now the matrix A of our nonnegative function $f(x)$ defined such that $f(x) \geq I_T(x)$, where T' is an open bounded rectangle in R_n, has been positive definite. This, however, is not so when T is a convex set or the union of two convex sets as we shall illustrate. We know from chapter two that when examples are given to illustrate the sharpness of an inequality, probabilities are only given to those points $x \in T$ and $x \notin T$ where $f(x) = 1$ and $f(x) = 0$, respectively. If T is a convex set or the $\cup$ of 2 convex sets, and A is positive definite, then $f(x) = 0$ implies that $x = 0$; also, since $\{xAx' \leq 1\}$ is an ellipsoid, $f(x) = 1$ for all $x \in T$ says that x can have at most two points in T. Since a three point distribution will not generally satisfy $E(X'X) = \Pi$, we conclude that A is not positive definite. Is there a matrix or vector which can replace A?

Marshall and Olkin [81] developed a general theorem for determining sharp upper bounds when T is either a convex set or the union of two convex sets.

Theorem 4.6.1. Let $X = (X_1,\dots,X_n)$ be a random vector with $E(X) = 0$, $E(X'X) = \Pi$. Let $T = T_+ \cup \{x: -x \in T_+\}$, where $T_+ \subseteq R_n$ is a closed convex set. If $\mathcal{A} = \{a = (a_1,\dots,a_n) \in R : ax' \geq 1 \text{ for all } x \in T_+\}$, then,

$$P(X \in T) \leq \inf_{a \in \mathcal{A}} a\Pi a' , \tag{4.6.1}$$

$$P(X \in T_+) \leq \inf_{a \in \mathcal{A}} (a\Pi a')/(1 + a\Pi a'). \tag{4.6.2}$$

Equality in (4.6.1) can be attained whenever $P(X \in T) \leq 1$; equality in (4.6.2) can always be attained.

We note that (4.6.2) is the one-sided analog of (4.6.1).

Remark [88]: If in the conditions of the theorem we only know $\pi_{ii} = \pi_i^2$, $i = 1,\dots,n$, then Π can be written as

$$\Pi = \left(\prod_{i=1}^{n} \pi_i^2\right) C_o ,$$

where $C_o \in \mathcal{C}$ the set of all correlation matrices. (4.6.1) and (4.6.2) are respectively equivalent to

$$P(X \in T) \leq \sup_{C_o \in \mathcal{C}} \inf_{a \in \mathcal{A}} a\Pi a', \tag{4.6.1)'$$

$$P(X \in T_+) \leq \sup_{C_o \in \mathcal{C}} \inf_{a \in \mathcal{A}} (a\Pi a')/(1 + a\Pi a'). \tag{4.6.2)'$$

(4.6.1)' and (4.6.2)' are sharp.

Proof: To prove (4.6.1), let $f(x) = (ax')^2$. Then, $f(x) \geq 0$ for all x, and $f(x) \geq 1$ for all $x \in T$. By Theorem 2.3.1

$$P(X \in T) \leq Ef(X) = \inf_{a \in \mathcal{A}} a\Pi a'.$$

To prove (4.6.2), let $f(x) = (ax' + a\Pi a')^2/(1 + a\Pi a')^2$. $f(x) \geq 0$ for all x and $f(x) \geq 1$ for all $x \in T_+$. By Theorem 2.3.1 we get (4.6.2).

We must now show the sharpness of the inequalities. This will be done through two examples offered by the authors.

Let $q = q(a) = a\Pi a'$, $q^* = q^*(a) = q/(1+q)$ and $w = w(a) = a\Pi/q$.

Lemma 4.6.1 [81]. There exists an $a_0 \in \mathcal{A}$ such that $\inf_{a \in \mathcal{A}} a\Pi a' = a_0\Pi a_0'$; for such an a_0, $w_0 = w(a_0) \in T_+$.

Proof: Since Π is positive definite, a transformation will make $\Pi = I$ and thus there exists a unique value $a_0 \in \mathcal{A}$ such that $\inf a\Pi a' = a_0\Pi a_0'$. If $w_0 \notin T_+$, then, by the geometry of convex sets there exists a separating hyperplane; i.e. there exists a vector $p \in R_n$, $p \neq 0$, and a constant k such that $px' \geq k > pw_0'$ for all $x \in T_+$. We know that $a_0w_0' = a_0(a_0\Pi/a_0\Pi a_0')'$ $= 1$ and thus $[p + (1-k)a_0]w_0' < 1$ and $[p + (1-k)a_0]x' \geq 1$ for all $x \in T_+$. Thus, if we replace p by $[p + (1-k)a_0]$, we can let $k = 1$. Since $pw_0' < 1$, i.e. $p(a_0\Pi/a_0\Pi a_0')' < 1$, we have $a_0\Pi p' < a_0\Pi a_0'$ and for small $\epsilon > 0$,

$$\epsilon(p\Pi p' - 2a_0\Pi p' + a_0\Pi a_0') < 2(a_0\Pi a_0' - a_0\Pi p');$$

$$\text{i.e.} \quad (\epsilon p + (1-\epsilon)a_0)\Pi(\epsilon p + (1-\epsilon)a_0)' < a_0\Pi a_0'. \tag{4.6.3}$$

However, $(\epsilon p + (1-\epsilon)a_0)x' \geq \epsilon + (1-\epsilon) = 1$; therefore $(\epsilon p + (1-\epsilon)a_0) \in \mathcal{A}$ and thus a contradiction by (4.6.3) and the definition of a_0. Therefore $w_0 \in T_+$.

Now to show that (4.6.1) is sharp, we let $q = q(a_0)$, $q^* = q^*(a_0)$ and $w = w(a_0)$. Let $D = \operatorname{diag}(d_1,\ldots,d_r)$ such that $d_i > 0$, $i = 1,\ldots,r$, $r \geq n$, and $\sum_{i=1}^{r} d_i = 1 - q$. Let M be an $r \times n$ matrix such that $M'M = \Pi - qw'w$ and let $C = D^{-\frac{1}{2}}M$ be an $r \times n$ matrix. Let Z be a $1 \times n$ random vector whose distribution is given by

$$\begin{aligned} &P(Z = c_i) = P(Z = -c_i) = d_i/2, \; i = 1,\ldots,r, \\ &P(Z = w) = P(Z = -w) = q/2, \end{aligned} \tag{4.6.4}$$

where c_i is the ith row of C.

$$E(Z) = \sum_{i=1}^{r} (c_i d_i/2 - c_i d_i/2) + wq/2 - wq/2 = 0$$

$$E(Z'Z) = C'DC + qw'w = M'M + qw'w = \Pi.$$

By (4.6.1), $P(Z \in T) \leq q$, however, by Lemma 4.6.1, $w \notin T$ and $P(Z \notin T) \geq q$. Thus $P(Z \notin T) = q$ and any random vector with distribution (4.6.4) can achieve equality in (4.6.1).

Lemma 4.6.2 [81] $\Pi - qw'w$ is positive semi definite and $\Pi - q^*w'w$ is positive definite.

Proof: By Cauchy's inequality [5, page 69],

$$(x\Pi x')(w\Pi^{-1}w') \geq (xw')^2.$$

$$x\Pi x' \geq q(xw')^2 \geq q^*(xw')^2. \qquad (4.6.5)$$

If $x \neq 0$, one of the two inequalities in (4.6.5) must be strict and the result follows.

We shall now show that (4.6.2) is sharp. By the above lemma we know that there exists a non-singular n x n matrix M satisfying $M'M = \Pi - q^*w'w$. Let P be an orthogonal matrix such that $-q^*wM^{-1}P > 0$, let $D = \text{diag}(d_1, \ldots, d_n)$, let $eD^{\frac{1}{2}} = -q^*wM^{-1}P$ and let $C = D^{-\frac{1}{2}}P'M$. Let Z be a random vector whose distribution is given by

$$P(Z = c_i) = d_i, \; i = 1, \ldots, n,$$

$$P(Z = w) = q^*. \qquad (4.6.6)$$

To satisfy a probability distribution $\sum_{i=1}^{n} d_i + q^* = 1$. Now $\sum_{i=1}^{n} d_i = eDe' = (q^*)^2 w(M'M)^{-1}w' = (q^*)^2 w(\Pi - q^*w'w)^{-1}w' = 1/(1 + q)$. Since $q^* = q/(1 + q)$, we have $\sum_{i=1}^{n} d_i + q^* = 1$. Therefore (4.6.6) defines a probability distribution.

$$E(Z) = eDC + wq^* = (-q^*wM^{-1}P)(P'M) + wq^* = 0$$

$$E(Z'Z) = C'DC + q^*w'w = M'M + q^*w'w = \Pi$$

By (4.6.2) $P(Z\epsilon T_+) \leqslant q^*$. However, by Lemma 4.6.1 $P(Z\epsilon T_+) \geqslant q^*$.

Therefore (4.6.2) is sharp.

Corollary. If $X = (X_1,\ldots,X_n)$ and $E(X) = 0$ and $E(X'X) = \Pi$, then,

$$P(X\epsilon T) \equiv P(/\prod_{i=1}^{n} X_i/ > 1 \text{ and } X > 0 \text{ or } X < 0) < \min_{a\epsilon\mathcal{A}} a\Pi a', \qquad (4.6.7)$$

$$P(X\ T_+) = P(/\prod_{i=1}^{n} X_i/ > 1,\ X > 0) < \min_{a\epsilon\mathcal{A}} (a\Pi a')/(1 + a\Pi a'). \qquad (4.6.8)$$

4.7 Inequalities for Concave Functions

For any concave function $f(x)$, Jensen's inequality says that $Ef(X) \leqslant fE(X)$. This property has been utilized by Mudholkar [87] and Mudholkar and Rao [88] to obtain generalizations of the univariate and one-sided univariate Tchebycheff inequalities.

Theorem 4.7.1 [88]. Let $X = (X_1,\ldots,X_n)$ be a nonnegative random vector such that $E(X) = \mu$(1 x n vector). Then, for any nonnegative, concave, homogeneous function ϕ defined on the nonnegative orthant R_{+n} of R_n, and $\epsilon > 0$,

$$P(\phi(X) \geqslant \epsilon) \leqslant \phi(u)/\epsilon\,. \qquad (4.7.1)$$

If $\phi(u)/\epsilon \leqslant 1$, (4.7.1) is sharp.

Proof: $P(\phi(X) \geqslant \epsilon) \leqslant E\phi(X)/\epsilon < \phi[E(X)]/\epsilon = \phi(\mu)/\epsilon$.

We show (4.7.1) to be sharp by constructing a distribution which attains equality in (4.7.1). Let the distribution of the random vector Y be given by

$$P(Y = (\epsilon/\phi(\mu))\mu) = \phi(\mu)/\epsilon\,,$$

$$P(Y = 0) = 1 - \phi(\mu)/\epsilon\,.$$

$$E(Y) = (\epsilon/\phi(\mu))\mu\cdot[\phi(\mu)/\epsilon] = \mu.$$

Since ϕ is a homogeneous function,

$$P(\phi(Y) \geqslant \epsilon) = P(Y = (\epsilon/\phi(\mu))\mu) = \phi(\mu)/\epsilon\,.$$

Thus any random vector X whose distribution is the same as that of Y achieves equality in (4.7.1).

Corollary. Under the above conditions the following inequality is sharp.

$$P(\phi(X) > \epsilon) \leq \phi(\mu)/\epsilon \qquad (4.7.2)$$

Proof: Since $\{X: \phi(X) > \epsilon\} \subseteq \{X: \phi(X) \geq \epsilon\}$, inequality (4.7.2) is true. If (4.7.2) is not sharp, then let there exist $\epsilon_o > \epsilon$, such that for each X such that $E(X) = \mu$, $P(\phi(X) > \epsilon) \leq \phi(\mu)/\epsilon_o$ and $\frac{\phi(\mu)}{\epsilon_o} < \frac{\phi(\mu)}{\epsilon}$. (4.7.3)

Let $\epsilon_o > \epsilon_1 > \epsilon$. By the theorem there exists a random vector Y such that $E(Y) = \mu$ and

$$\phi(\mu)/\epsilon_1 = P(\phi(Y) \geq \epsilon_1) \leq P(\phi(Y) > \epsilon) \leq \phi(\mu)/\epsilon_o .$$

By (4.7.3) a contradiction exists and thus (4.7.2) is sharp.

Various inequalities can be obtained from this theorem by considering various forms of nonnegative random vectors and certain nonnegative, concave, homogeneous functions on R_{+n}.

Example: If $X_1, \ldots, X_n$ are jointly distributed random variables such that $E(X_i^2) = \sigma_i^2$, $i = 1, \ldots, n$, then, for any nonnegative, homogeneous, concave function ϕ on R_{+n},

$$P(\phi(X_1^2, \ldots, X_n^2) \geq \epsilon) \leq \phi(\sigma_1^2, \ldots, \sigma_n^2)/\epsilon. \qquad (4.7.4)$$

If in (4.7.1) we let $X_i = X_i^2$, $i = 1, \ldots, n$, (4.7.4) follows immediately. To show that (4.7.4) is sharp, we consider the joint distribution of the random variables X_i, $i = 1, \ldots, n$, given as follows:

$$P(X_i = \pm \sigma_i/\beta^{\frac{1}{2}}, i = 1, \ldots, n) = \beta/2$$

$$P(X_i = 0, i = 1, \ldots, n) = 1 - \beta.$$

If $\beta = \phi(\sigma_1^2, \ldots, \sigma_n^2)/\epsilon \leq 1$, then $P(\phi(X_1^2, \ldots, X_n^2) \geq \epsilon) = \beta$. Therefore (4.7.4) is sharp.

(4.7.4) can be rewritten as

$$P(\phi(X_1^2,\ldots,X_n^2) \geq \epsilon,\ X > 0 \text{ or } X < 0) \leq \phi(\sigma_1^2,\ldots,\sigma_n^2)/\epsilon. \quad (4.7.5)$$

We now consider a specific example of a nonnegative, homogeneous, concave function on R_{+n} which will enable us to obtain a sharp inequality. If $t = (t_1,\ldots,t_n)$ is a nonnegative real valued vector and $\{\alpha_i\}_{i=1}^n$ are nonnegative real numbers such that $\sum_{i=1}^n \alpha_i = 1$, then,

$$\phi_r(t) = (\sum_{i=1}^n \alpha_i t_i^r)^{1/r},\ r \leq 1, \quad (4.7.6)$$

is a nonnegative, homogeneous, concave function on R_{+n}.

Theorem 4.7.2 [88] Let $X = (X_1,\ldots,X_n)$ be a random vector such that $E(X) = 0$, $E(X'X) = \Pi$ and $E(X_i^2) = \sigma_i^2$, $i = 1,\ldots,n$. Then for any non-negative numbers $\alpha_1,\ldots,\alpha_n$ such that $\sum_{i=1}^n \alpha_i = 1$ and $r \leq \frac{1}{2}$,

$$P(\sum_{i=1}^n \alpha_i |X_i|^{2r} \geq \epsilon,\ X > 0) \leq (\sum_{i=1}^n \alpha_i \sigma_i^{2r})^{1/r} / (\epsilon^{1/r} + (\sum_{i=1}^n \alpha_i \sigma_i^{2r})^{1/r}). \quad (4.7.7)$$

(4.7.7) is sharp.

Proof: If $r \leq \frac{1}{2}$, $(\sum_{i=1}^n \alpha_i |X_i^2|^r)^{1/2r}$ is a concave function (by (4.7.6)).

$$T_+ = \left\{x: (\sum_{i=1}^n |X_i^2|^r)^{1/2r} \geq 1,\ x > 0\right\}$$ is a closed convex set in R_n and by Theorem 4.6.1 equation (4.6.1)', (using the notation defined there),

$$P(X \in T) = P(\sum_{i=1}^n \alpha_i |X_i|^{2r} \geq 1,\ X > 0 \text{ or } X < 0)$$

$$\leq \sup_{C_0 \in \mathcal{C}} \inf_{a \in \mathcal{A}} a \Pi a' . \quad (4.7.8)$$

(4.7.8) is sharp. Since $\phi(x_1^2,\ldots,x_n^2) = (\sum_{i=1}^n \alpha_i (x_i^2)^r)^{1/r}$ is a non-negative, homogeneous, concave function, the above example (4.7.5) tells us that

$$\sup_{C_0 \in \mathcal{C}} \inf_{a \in \mathcal{A}} a \Pi a' = (\sum_{i=1}^n \alpha_i \sigma_i^{2r})^{1/r}. \quad (4.7.9)$$

If we again consult Theorem 4.6.1 equation (4.6.2), then, since $a \Pi a' \geq 0$,

$$P\left(\sum_{i=1}^{n} \alpha_i X_i^{2r} \geq 1,\ X > 0\right) \leq \sup_{C_0 \in \mathcal{C}} \inf_{a \in \mathcal{A}} (a\Pi a'/(1 + a\Pi a'))$$

$$\text{(by (4.7.9))} \qquad \leq \left(\sum_{i=1}^{n} \alpha_i \sigma_i^{2r}\right)^{1/r} \Big/ \left(1 + \left(\sum_{i=1}^{n} \alpha_i \sigma_i^{2r}\right)^{1/r}\right). \quad (4.7.10)$$

(4.7.7) follows from (4.7.10).

To show that (4.7.7) is sharp, we consider the random variables X_i with joint distribution given by

$$P(X_i = \sigma_i/\beta^{\frac{1}{2}},\ i = 1,\dots,n) = \beta/(1 + \beta),$$
$$P(X_i = -\sigma_i/\beta^{\frac{1}{2}},\ i = 1,\dots,n) = 1/(1 + \beta).$$

If $\beta = (\sum_{i=1}^{n} \sigma_i^{2r})^{1/r}/\epsilon^{1/r}$, we can see that equality is attained in (4.7.10).

The same arguments that were used in Theorem 4.7.1 can be extended to concave functions of symmetric random matrices.

Theorem 4.7.3 [88]. Let Z be a n x n symmetric positive semi-definite matrix such that $E(Z) = \Pi$ and $g(z) \geq 0$ is a homogeneous, concave function of Z. Then, for $\epsilon > 0$,

$$P(g(Z) \geq \epsilon) \leq g(\Pi)/\epsilon\ . \qquad (4.7.11)$$

If $g(\Pi)/\epsilon \leq 1$, (4.7.11) is sharp.

Proof: (4.7.11) follows immediately from Jensen's inequality. If $g(\Pi)/\epsilon \leq 1$, let us consider a random matrix Z_0 whose distribution is given by

$$P(Z_0 = (\epsilon/g(\Pi))\cdot\Pi) = g(\Pi)/\epsilon\ ,$$
$$P(Z_0 = 0) = 1 - g(\Pi)/\epsilon\ .$$

$E(Z_0) = \epsilon\Pi\, g(\Pi)/g(\Pi)\epsilon = \Pi$ and $P(g(Z_0) \geq \epsilon) = g(\Pi)/\epsilon$. Thus (4.7.11) is sharp.

Our notation is slightly changed for the remainder of this section.

Mudholkar [87] has offered an inequality for matrix valued random variables. Let $X_1,\dots,X_n$ be jointly distributed p x 1 random vectors such that $E(X_i) = 0$, $E(X_i X_i') = \Pi_i$, $i = 1,\dots,n$. $\Pi = \Pi_1 + \dots + \Pi_n$ and

$X = (X_1,\dots,X_n)$. If $c_1,\dots,c_p$ are the characteristic roots of XX' and $\pi_1,\dots,\pi_p$ are the characteristic roots of Π, then, for any nonnegative symmetric concave function f on R_{+p}, Mudholkar has proven the following result.

Theorem 4.7.3. Subject to the above conditions

$$P(f(c_1,\dots,c_p) \geq \epsilon) \leq f(\pi_1,\dots,\pi_p)/\epsilon, \qquad (4.7.12)$$

where $\epsilon > 0$.

Proof: Marcus [78] has shown that if $c_1,\dots,c_p$ are characteristic roots of a positive semi definite symmetric matrix A, then

$$\min f(y_1'Ay_1, y_2'Ay_2,\dots,y_p'Ay_p) = f(c_1,\dots,c_p),$$

where f is defined as above and the min is taken over all orthonormal sets of p x 1 vectors $y_1,\dots,y_p$.

If $XX' = A$, then, by Marcus' result

$$P[f(c_1,\dots,c_p) \geq \epsilon] \leq P[f(y_1'Ay_1,\dots,y_p'Ay_p) \geq \epsilon]$$

$$\leq Ef(y_1'Ay_1,\dots,y_p'Ay_p)/\epsilon$$

$$\leq f(y_1'\Pi y_1,\dots,y_p'\Pi y_p)/\epsilon$$

$$\leq f(\pi_1,\dots,\pi_p)/\epsilon \,. \qquad (4.7.12)$$

Mudholkar and Rao [88] wish to know whether (4.7.12) can be derived from (4.7.11). We answer that question for a specific case.

Let the p x n matrix X be defined as above. If $Z = XX'$, $E(Z) = \Pi$, and Z and Π commute, then (4.7.12) will follow from (4.7.11). To prove this, we note that two symmetric matrices can be diagonalized by the same orthogonal matrix iff the two matrices commute. Let C be an orthogonal matrix such that $C'ZC = \text{diag}(c_1,\dots,c_p)$ and $C'\Pi C = \text{diag}(\pi_1,\dots,\pi_p)$, where c_i, π_i, $i = 1,\dots,p$, are the characteristic roots of Z and Π,

respectively. Let f be a nonnegative symmetric concave function defined on the diagonal elements of any p x p diagonal matrix.
By (4.7.11)

$$P(f(C'ZC) \geq \epsilon) \leq Ef(C'ZC)/\epsilon$$
$$\leq f(C'\pi C)/\epsilon ,$$

i.e. $$P(f(c_1,\ldots,c_p) \geq \epsilon) \leq f(\pi_1,\ldots,\pi_p)/\epsilon .$$

We note that the condition Z and π commute is too severe a restriction to hold in general.

4.8 An Inequality for a Continuous Stochastic Process

Whittle [122] and Birnbaum and Marshall [15] have both offered Tchebycheff type inequalities for stochastic processes. Whittle obtains his inequality by choosing a certain matrix B and minimizing tr $B^{-1}\pi$, where π is the variance covariance matrix of the process. In the case of a stochastic process, Whittle shows that the matrix B as defined in Theorem 4.2.1 is not satisfactory since tr $B^{-1}\pi \to \infty$ as n increases.

The following theorem and its proof are taken from Parzen [92, page 85]; the theorem coincides with that of Whittle.

Theorem 4.8.1. Let $\{X(t),\ a \leq t \leq b\}$ be a stochastic process which is differentiable in mean square. Let

$$C(t) = \{E(|X(t)|^2)\}^{\frac{1}{2}} < \infty,$$
$$C_1(t) = \{E(|X'(t)|^2)\}^{\frac{1}{2}} < \infty .$$

$$P(\sup_{a\leq t\leq b} |X(t)| > k) \leq E[\sup_{a\leq t\leq b} X^2(t)]/k^2$$
$$\leq [\tfrac{1}{2}(C^2(a) + C^2(b)) + \int_a^b C(t)C_1(t)dt]/k^2, \quad (4.8.1)$$

where $k > 0$.

Proof: By Markov's inequality, the first inequality is true. For $t \in [a,b]$,

$$X^2(t) = X^2(a) + 2\int_a^t X'(u)X(u)du$$

$$= X^2(b) - 2\int_t^b X'(u)X(u)du .$$

$$2X^2(t) = X^2(a) + X^2(b) + 2[\int_a^t X'(u)X(u)du - \int_t^b X'(u)X(u)du]$$

$$\leq X^2(a) + X^2(b) + 2\int_a^b |X'(u)X(u)|du$$

$$\sup_{a \leq t \leq b} X^2(t) \leq \tfrac{1}{2}[X^2(a) + X^2(b)] + \int_a^b |X'(u)X(u)|du$$

$$E(\sup_{a \leq t \leq b} X^2(t)) \leq \tfrac{1}{2}[E(X^2(a)) + E(X^2(b))] + \int_a^b E(|X'(u)X(u)|)du$$

$$\leq \tfrac{1}{2}[E(X^2(a)) + E(X^2(b))] + \int_a^b [E(X^2(u))E(X'^2(u))]^{\frac{1}{2}}du . \quad (4.8.2)$$

(4.8.1) follows directly from (4.8.2).

Corollary. If $E(X(t)) = m(t)$, then

$$P(|X(t) - m(t)| > k) \leq \frac{Var(X(a)) + Var(X(b))}{2k^2}$$

$$+ \int_a^b \frac{[Var(X(t))]^{\frac{1}{2}}[Var(X'(t))]^{\frac{1}{2}}dt}{k^2} . \quad (4.8.3)$$

4.9 Other Inequalities

In this section we shall briefly mention certain other multivariate Tchebycheff type inequalities which are proved by special methods or arguments. Historically speaking, we should note that it was K. Pearson [93] who introduced the basic idea behind equation (4.2.11). He considered an ellipse centered about the origin defined by

$$f(x_1,x_2) = Mx_1^2 + Nx_2^2 + Qx_1x_2 ,$$

and a region T such that $T' = \{(x_1,x_2):f(x_1,x_2) < 1\}$. Pearson found the upper bound of $P((X_1,X_2) \in T)$ in terms of the s th order moment of the function $f(x_1,x_2)$. If we simplify his idea to the case when $s = 1$, we see that for a random vector $X = (X_1,X_2)$ with zero mean and $E(X_iX_j) = \sigma_{ij}, i,j = 1,2$, he showed that

$$P(X\epsilon T) \leq M\sigma_1^2 + N\sigma_2^2 + Q\sigma_{12}. \qquad (4.9.1)$$

The generalizations of Berge [8] and Lal [68] are based on Pearson's method.

Let $X = (X_1,\ldots,X_n)$ be a random vector such that $E(X_i) = \mu_i$ and $E(X_iX_j) = \sigma_{ij}$, $i,j = 1,\ldots,n$, and $\sigma_{ii} = \sigma_i^2$. Chapelon [22] considers an ellipsoid defined by

$$\sum_{i,j} \sigma_{ij}\mu_i\mu_j = 1, \ i,j = 1,\ldots,n, \qquad (4.9.2)$$

which he calls a quadrique type. Also, he considers a parallelepiped which circumscribes (4.9.2) in such a way that its sides are parallel to the x_i axes and touch the x_ith axis at $\pm\sigma_i$, $i = 1,\ldots,n$. He calls this parallelepiped a parallèlépipède type. Chapelon proves that for $t \geq n^{\frac{1}{2}}$, the probability that X will fall in a region similar in construction and t times as large as either the quadrique or parallèlépipède type is greater than $1 - n/t^2, t > 0$.

Leser [69] offers a multivariate inequality in a situation when a restriction is imposed on the distribution. Let λ_i, $i = 1,\ldots,n$ be positive integers and let $\lambda_o^2 = n/[\sum_{i=1}^{n}(1/\lambda_i^2)]$ and $\sigma_o^2 = n/[\sum_{i=1}^{n}(1/\sigma_i^2)]$. We define an ellipsoid by

$$R^2 = \lambda_o^2\sigma_o^2\sum_{i=1}^{n}(x_i/\lambda_i\sigma_i)^2, \qquad (4.9.3)$$

and let $A(R_o)$ be the mean value of the density function $f(x_1,\ldots,x_n)$ when R assumes the value R_o. Let $P = P(\sum_{i=1}^{n}(X_i/\lambda_i\sigma_i)^2 \leq n)$; i. e.

$P = P(R/n^{\frac{1}{2}} \leq \lambda_o\sigma_o)$. If $f(x)$ causes $A(R)$ to be a non-increasing function of R for $R \leq k\sigma_o n^{\frac{1}{2}}$, Leser obtains probability bounds for P over different ranges of k. The bounds are expressed in terms of n, λ_o and k.

Camp [19] obtains a sharp multivariate inequality in terms of a new statistic which he calls the "contour moment" and defines as follows. Let $f(t) = f(t_1,\ldots,t_n)$ be a density function defined on a set $T \epsilon R_n$ such that $0 \leq f(t) \leq L$ and $f(t)$ is Lebesgue integrable. Let Q_λ be the set consisting of all those points $t \epsilon T$ for which $f(t) > \lambda$: x_λ is the measure of Q_λ. x is a unique single valued function of λ and is monotone decreasing in λ ; is also a monotone decreasing function of x for $0 < x < x_0 < \infty$.. We define $y(x) = \lambda_x$ as a single valued monotonic decreasing function of x. To insure that y is single valued, we define $y = \min \lambda_x$ if λ takes on several values at a point x. If λ does not take on any values in an interval, we define y to take on the value it had in the beginning of the interval. The rth contour moment is defined by

$$\hat{u}_r = \int_0^{x_0} x^r y dx. \qquad (4.9.4)$$

If $r = 2$, we obtain the contour variance. Camp proves that

$$1 - \int_{Q_\lambda} f dT \leq [\hat{u}_{2r}/(\delta\hat{\sigma})^{2r}][2r/(2r+1)]^{2r}, \qquad (4.9.5)$$

where λ satisfies $x_\lambda = \delta\hat{\sigma}$.

We shall complete this chapter by reviewing an inequality for minimum components which is obtained from Theorem 2.3.1 through an appropriate choice of $f(x)$.

Theorem 4.9.1 [81]. If $X = (X_1,\ldots,X_n)$ is a random vector such that $E(X) = 0$, $E(X'X) = \Pi$, then, for $T = T_+ \cup \{x: -x \epsilon T_+\}$, where $T_+ \subseteq R_n$ is a closed convex set,

$$P(X \in T) \equiv P(\min_i X_i \geq 1 \text{ or } \min_i (-X_i) \geq 1) \leq \min [1/(e\pi_s^{-1}e')], \qquad (4.9.6)$$

$$P(X \in T_+) = P(\min_i X_i \geq 1) \leq \min [1/(1 + e\pi_s^{-1}e')], \qquad (4.9.7)$$

where the min in the right hand side of the inequalities is taken over all principal submatrices π_s of π such that $e\pi_s^{-1} > 0$.

Proof: Let X_o be a subvector of X whose components are such that $E(X_o'X_o) = \pi_s$. If we let $f(x) = (e\pi_s^{-1}x_o')^2/(e\pi_s^{-1}e')^2$, $f(x)$ satisfies the conditions of Theorem 2.3.1 in the region T and taking $E(f(X))$ we get (4.9.6). In a similar fashion, by letting

$$f(x) = (1 + e\pi_s^{-1}x_o')^2/(1 + e\pi_s^{-1}e')^2$$

we can obtain (4.9.7).

CHAPTER FIVE: SUMS OF RANDOM VARIABLES

5.1 Introduction

In this chapter we shall determine upper probability bounds for either a sum of n random variables, S_n , or for some function of S_n, e.g. $\overline{X}$. Unless otherwise stated, we shall assume the random variables to be independent. We shall mainly concentrate on finding probability bounds for a finite number of random variables since some classical central limit theorems tell us that, as $n \to \infty$, the limiting distribution of

$$\frac{S_n - E(S_n)}{[Var(S_n)]^{\frac{1}{2}}}$$

is a normal distribution; these limit distributions have been thoroughly discussed in the books of Gnedenko [36] and Kolmogorov and Gnedenko [66].

It is easy to see that the Markov, Tchebycheff and one-sided Tchebycheff inequalities can be extended to $\overline{X}$. If $X_1, \ldots, X_n$ are independent random variables such that $E(\overline{X}) = \mu$, $Var(\overline{X}) = \sigma^2/n$, then for $k > 0$, (2.1.2) implies

$$P(|\overline{X} - \mu| \geqslant k) \leqslant \sigma^2/nk^2; \qquad (5.1.1)$$

(2.1.8) implies

$$P(\overline{X} - \mu \geqslant k) \leqslant \sigma^2/(\sigma^2 + nk^2). \qquad (5.1.2)$$

If $X_i > 0$, $i = 1, \ldots, n$, then, by Markov's inequality,

$$P(\overline{X} \geqslant \mu + t) \leqslant \mu/(\mu + t).$$

$$P(\overline{X} - \mu \geqslant t) \leqslant \mu/(\mu + t) \qquad (5.1.3)$$

The above inequalities are sharp, however, to attain equality (n - 1) of the random variables must be identically zero and the remaining X_i must be chosen appropriately. To attain equality in (5.1.3) we let $P(X_1 = 0) = t/(\mu + t)$, $P(X_1 = n(\mu + t)) = \mu/(\mu + t)$, and we define $X_2 = X_3 = \dots = X_n \equiv 0$.

It would appear that under general conditions and even more so under restricted conditions one could obtain sharper bounds than the three above. In this chapter we shall give improved bounds over (5.1.1), (5.1.2) and (5.1.3).

5.2 Normal Approximation for a Sum of Independent Random Variables

Let $X_1,\dots,X_n,\dots$ be a sequence of mutually independent random variables satisfying $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2$, $i = 1,\dots,n,\dots$. Also, let $s_n^2 = \sum_{i=1}^{n} \sigma_i^2 < \infty$. When certain moment conditions, to be indicated, are fulfilled, then, as $n \to \infty$, the distribution of some linear function of the sum of the independent random variables tends to the $N(0,1)$ distribution. The conditions under which the above occurs are stated in the classical limit theorems of Lindeberg-Levy, Liapunov and Lindeberg-Feller. The study of limit distributions is in itself a complete subject and the results cannot, in a strict sense, be classified as Tchebycheff type inequalities.

In this section we shall state some theorems and their consequences without proving either; the proofs of the major theorems can be found in text books such as Gnedenko [35] and Tucker [114]; comprehensive studies of this subject can be found in Gnedenko [36] and Kolmogorov and Gnedenko [66].

Theorem 5.2.1. Lindeberg-Levy Central Limit Theorem

Let $X_1,\dots,X_n,\dots$ be a sequence of i.i.d. random variables such that

$E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, $i = 1,\ldots,n,\ldots$. Then, as $n \to \infty$, the distribution of $n^{\frac{1}{2}}(\bar{X} - \mu)/\sigma$ tends to the $N(0,1)$ distribution.

Theorem 5.2.2. Liapunov Theorem

Let $X_1,\ldots,X_n,\ldots$ be a sequence of independent random variable such that $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2 \neq 0$ and $E|X_i - \mu_i|^3 = v_{3i}$ exist for $i = 1,2,\ldots$. Let $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$ and $B_n^3 = \sum_{i=1}^{n} v_{3i}$. If

$$\lim_{n\to\infty} \frac{B_n}{s_n} = 0 , \qquad \text{(Liapunov condition)}$$

then, as $n \to \infty$,

$$P\left(\left[\sum_{i=1}^{n}(X_i - \mu_i) / s_n\right] \leq x\right) \to \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)dx.$$

Theorem 5.2.3. Lindeberg-Feller Theorem

Let $X_1,\ldots,X_n,\ldots$ be a sequence of independent random variables such that $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2 < \infty, i=1,2,\ldots$. Let G_i and F_n be the distribution functions of X_i and $Y_n = \sum_{i=1}^{n}(X_i - \mu_i)/s_n$, respectively, where s_n is defined as in Theorem 5.2.2. A necessary and sufficient condition that

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \frac{\sigma_i}{s_n} = 0, \quad \lim_{n\to\infty} P(Y_n \leq x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)dx$$

is that for every $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|x-\mu_i| > \epsilon s_n} (x - \mu_i)^2 dG_i(x) = 0.$$

As direct consequences of the above theorem, various people have attempted to obtain numerical values for D_n, the error or remainder value resulting from the normal approximation.

$$\sup_{-\infty<x<\infty} \left| P\left(\sum_{i=1}^{n} \frac{(X_i - \mu_i)}{s_n} \leq x \right) - \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)dx \right| \leq D_n. \quad (5.2.1)$$

Let us assume, without loss of generality, that $E(X_i) = 0$, $Var(X_i) = \sigma_i^2$ and $E|X_i|^3 = v_{3i}$, $i = 1,2,\ldots$. Under these conditions Cramér [24,25] has shown that for $n > 1$,

$$D_n = 3 \log nB_n^3/s_n^3. \tag{5.2.2}$$

In (5.2.2) the term 3 log n was changed by Bergström [9] to 4.8; this new value of D_n offers a better bound for $n \geq 40$. Through the use of Fourier transforms, Berry [12] showed that

$$D_n \leq \frac{1.88}{s_n} \max\left(\frac{v_{31}}{\sigma_1^2}, \ldots, \frac{v_{3n}}{\sigma_n^2}\right), \tag{5.2.3}$$

where $\sigma_i^2 \neq 0$, $i = 1,\ldots,n$. As pointed out by Hsu [50], the value 1.88 in (5.2.3) is incorrect. Takano [112] and Stoker [111] have replaced 1.88 by 2.031 and 1.952, respectively. Godwin [34, page 82], through an example, shows that the numerical constant cannot be less than 0.199.. .

If the X_i's are identically distributed, then

$$2.031v_3/n^{\frac{1}{2}}\sigma^3 < 4.8v_3/n^{\frac{1}{2}}\sigma^3;$$

the Berry-Takano result is superior to Bergström's result. For i.i.d. random variables such that $(B_n/\sigma)^2$ is large, i.e. ≥ 5, Ikeda [51] offers a smaller approximation error than the Berry-Takano result. He shows that for $(B_n/\sigma)^2 \geq 5$, $D_n \leq 1.84076v_3/n^{\frac{1}{2}}\sigma^3$: for $(B_n/\sigma)^2 \geq 6$, $D_n \leq 1.77803v_3/n^{\frac{1}{2}}\sigma^3$.

The following theorem was proved by Offord [90].

Theorem 5.2.4. Let $X_1,\ldots,X_n$ be mutually independent random variables such that $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2$, $E|X_i - \mu_i|^3 = v_{3i}$, $i = 1,\ldots,n$, and $\min_i[\sigma_i/(v_{3i})^{1/3}] = 2k^{1/3}$ for some k. Then for $n > 1$,

$$\sup_t P\left(\left|\sum_{i=1}^n X_i - t\right| \leq x\right) \leq \frac{6 \log n}{k^3 n^{\frac{1}{2}}}\left\{\log n + \frac{kx}{\min_i \sigma_i}\right\} .$$

Remark: Since $(v_{3i})^{1/3} \geq \sigma_i$ for all i (Holder's inequality), $k \leq 1/8$.

Though this inequality seems to fulfill the requirements of a Tchebycheff type inequality, it is to be noted that the proof of the theorem is dependent upon the result (5.2.2); i.e. the result is obtained through the use of the error term involved in the normal approximation.

Similar to the problem introduced in (5.2.1) numerical values have also been obtained for the absolute difference between the characteristic function of the random variable $(\sum_{i=1}^{n} X_i/s_n)$ and of a N(0,1) variable. Ikeda [51] has offered numerical results in the cases when the X_i are independent and when they are i.i.d. Kolmogorov and Gnedenko [66, page 202] also offer a remainder term for the difference of the characteristic functions when the X_i are i.i.d. Their result is inferior to that of Ikeda.

5.3 An Inequality for $\overline{X}$

As seen in chapter two, we can obtain certain inequalities by appropriately choosing μ and μ_2 in Tchebycheff's inequality. If, for example, the X_i are i.i.d. for $i = 1,\ldots,n$, $\mu = 0$ and $Var(X_i) = 1/n$, then, by the Tchebycheff inequality

$$P(|\overline{X}| \geq t/n) \leq E(\overline{X})^2/(t/n)^2 = 1/t^2. \tag{5.3.1}$$

A similar result to the above was obtained by Robbins [96]. He showed that for $n > 1$, and $t > n^{\frac{1}{2}}$, there exists a function $\phi_n(t)$ such that $t^2\phi_n(t) \to 1$ as $t \to \infty$ and

$$P(|\overline{X}| \geq t/n) \leq \phi_n(t) < 1/t^2. \tag{5.3.2}$$

A confidence interval for $\overline{X}$ which is a function of both the sample and population variances was obtained by Guttman [42]. Let $X_1,\ldots,X_n$ be a random sample from a population with mean μ and finite variance σ^2. Using the maximum likelihood principle we define $\overline{X} = \sum_{i=1}^{n} X_i/n$ and $s^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n.$

Theorem 5.3.1. Subject to the above notation, for $k > 1$,

$$P\left\{(\overline{X} - \mu)^2 \geq s^2/(n-1) + \sigma^2(2(k^2-1)/n(n-1))^{\frac{1}{2}}\right\} \leq 1/k^2. \quad (5.3.3)$$

Proof: Consider the r.v. $Y = (\overline{X} - \mu)^2 - s^2/(n-1) - c\sigma^2$, where c is a constant. Taking expectations we see that $E(Y) = -c\sigma^2$ and $E(Y^2) = \sigma^4(2/n(n-1) + c^2)$. By Tchebycheff's inequality,

$$P(/Y/ \geq k(E(Y^2))^{\frac{1}{2}}) \leq 1/k^2.$$

In particular,

$$P\left\{(\overline{X} - \mu)^2 \geq s^2/(n-1) + c\sigma^2 + k\sigma^2(2/n(n-1) + c^2)^{\frac{1}{2}}\right\} \leq 1/k^2. (5.3.4)$$

Since c is arbitrary, we choose an appropriate c so as to minimize the right hand side within the brackets. Minimizing we obtain $c^2 = 2/n(n-1)(k^2-1)$ for $c < 0$; substituting this value into (5.3.4) we obtain (5.3.3). (5.3.3) can also be rewritten as

$$P\left\{/\overline{X} - \mu/ \geq [s^2/(n-1) + \sigma^2(2(k^2-1)/n(n-1))^{\frac{1}{2}}]^{\frac{1}{2}}\right\} \leq 1/k^2. \quad (5.3.5)$$

Similar to (5.3.3), Midzuno [86] has obtained a confidence interval for $\overline{X}$ in terms of approximate values of higher moments. He considers the r.v. $Z = (\overline{X} - \mu)^2 - s^2/(n-1)$ and by means of a lengthy multinomial expansion and approximation, Midzuno shows that

$$E(Z^m) = (\sigma^2/n)^m H_m,$$

where

$$H_m = \left(\sum_{h=0}^{m}\left\{(-1)^h\binom{m}{h}[2(m-h)]!/[2^{m-h}(m-h)!]\right\}\right)\cdot(1 + O(1/n)). \quad (5.3.6)$$

By now applying Markov's inequality, he obtains

$$P(/\overline{X} - \mu/ \geq (s^2/(n-1) + k\sigma^2/n)^{\frac{1}{2}}) \leq H_m/k^m. \quad (5.3.7)$$

Midzuno also used another lengthy multinomial expansion to calculate

$$E/\bar{X} - \mu/^{2L} = \frac{1\cdot 3\cdot 5\cdot\cdots\cdot(2L-1)\sigma^{2L}}{n^L}\left\{1 + O(1/n)\right\}. \quad (5.3.8)$$

If we now apply Markov's inequality, we obtain

$$P(/\bar{X} - \mu/ \geq k\sigma/n^{\frac{1}{2}}) < \frac{1\cdot 3\cdot 5\cdot\cdots\cdot(2L-1)}{k^{2L}}\left\{1 + O(1/n)\right\}. \quad (5.3.9)$$

We note that both of Midzuno's inequalities are not sharp.

Remark: Since $E(X^{2L})$ for a $N(0,1)$ variate is equal to $1\cdot 3\cdot 5\cdot\cdots\cdot(2L-1)$, we might suspect that there is a normal approximation involved when some terms of higher order are neglected in the multinomial expansion. This result preceded Aoyama's result [1], previously mentioned in this paper.

5.4 Bounds for the Sum of Independent Nonnegative Random Variables

Let $X_1,\ldots,X_n$ be independent nonnegative random variables satisfying $E(X_i) = \mu_i$, $i = 1,\ldots,n$. Let $\mathcal{S}$ denote the class of random variables S_n, where $S_n = \sum_{i=1}^{n} X_i$. When $n = 1$, Markov's inequality offers a sharp upper probability bound for $P(X \geq t)$. When Markov's inequality is extended to $\bar{X}$ as in (5.1.3), the resulting inequality is sharp only if $(n - 1)$ of the random variables are identically 0; a sharper bound is desirable.

In this section we shall determine a sharp upper bound for

$$\sup_{S_n \in \mathcal{S}} P(S_n \geq t) \quad (5.4.1)$$

when $n = 2$.

The problem in (5.4.1), for $n = 2$, was solved independently by Birnbaum, Raymond and Zuckerman [16] and by Samuels [100]. Samuels conjectures a result for all n, offers some support for the conjecture, and proves the conjecture for $n = 2$ and $n = 3$; the proof for $n = 3$ is based on a theorem which offers support for the conjecture and on some lemmas which we shall use to obtain the result for $n = 2$. Birnbaum, Raymond and Zuckerman, on the other hand, find their method too complex to extend beyond $n = 2$. Our proofs and discussion will be based upon the works of both authors.

The problem will be reduced to that of finding an upper probability bound for random variables taking on two values, the lower value always being zero. In the case of arbitrary i.i.d. random variables, Hoeffding and Shrikhande [49] also reduce the problem to that of finding probability bounds for random variables taking on a finite number of values; their method offers the best result only for $n = 2$. In the case when the X_i are nonnegative and i.i.d., they offer a bound for S_2 which is only sometimes superior to that of Samuels and Birnbaum, Raymond and Zuckerman.

Lemma 5.4.1 [16]. Let $X_1,\ldots,X_n$ be independent nonnegative random variables and let X_1 take on the values $x_{11} \leqslant x_{12} \leqslant \cdots \leqslant x_{1m}$ with respective probabilities $p_1,\ldots,p_m$. Let x_{1j}, x_{1k}, x_{1L} be three possible values of X_1 such that $0 \leqslant x_{1j} \leqslant x_{1k} \leqslant x_{1L}$. Then, for $t > 0$, there exists a random variable X_1' which has the same distribution as X_1 except that the values x_{1j}, x_{1k}, x_{1L} assume the probabilities p_j', p_k', p_L', such that

$$E(X_1') = E(X_1), \text{ one of } p_j', p_k', p_L' \text{ is zero}$$

and
$$P(X_1' + \sum_{i=2}^{n} X_i \geqslant t) \geqslant P(X_1 + \sum_{i=2}^{n} X_i \geqslant t). \qquad (5.4.2)$$

Remark: We need not worry about the case when the independent random variables are continuous for we know [46] that if $F(x)$ is a distribution function of a continuous random variable X, $F(x)$ can be approximated uniformly by a distribution function $G(x)$, where $G(x)$ is a step function having a finite number of jumps and satisfying the moment conditions of $F(x)$.

Proof: For any α,β let us denote the following:

$$p_j' = p_j + \alpha B,\ p_k' = p_k - \beta,\ p_L' = p_L + (1-\alpha)\beta, \tag{5.4.3}$$

i.e. $p_j' + p_k' + p_L' = p_j + p_k + p_L$.

If $\alpha = (x_{1L} - x_{1k})/(x_{1L} - x_{1j})$, then,

$$x_{1j}p_j' + x_{1k}p_k' + x_{1L}p_L' = x_{1j}p_j + x_{1k}p_k + x_{1L}p_L ,$$

and $E(X_1') = E(X_1)$, for all β.

$$P(X_1 + \sum_{i=2}^{n} X_i \geqslant t) = \sum_{r=1}^{m} P(X_1 = x_{1r})\, P(\sum_{i=2}^{n} X_i \geqslant t - x_{1r})$$

$$= \sum_{r=1}^{m} p_r P(\sum_{i=2}^{n} X_i \geqslant t - x_{1r}). \tag{5.4.4}$$

$$P(X_1' + \sum_{i=2}^{n} X_i \geqslant t) = \sum_{r=1}^{m} p_r'\, P(\sum_{i=2}^{n} X_i \geqslant t - x_{1r}), \tag{5.4.5}$$

where $P(X_1' = x_{1i}) = p_i'$, $i = 1,\dots,m$. From (5.4.3), (5.4.4) and (5.4.5) we see that

$$P(X_1' + \sum_{i=2}^{n} X_i \geqslant t) - P(X_1 + \sum_{i=2}^{n} X_i \geqslant t) = \alpha\beta P(\sum_{i=2}^{n} X_i \geqslant t - x_{1j})$$

$$- \beta P(\sum_{i=2}^{n} X_i \geqslant t - x_{1k}) + (1-\alpha)\beta P(\sum_{i=2}^{n} X_i \geqslant t - x_{1L})$$

$$= \beta C. \tag{5.4.6}$$

(5.4.2) will hold only if βC is positive. If C is positive, let $\beta = p_k$ and thus $p_k' = 0$ by (5.4.3); if C is negative, let

$\beta = \max(-p_j/\alpha, -p_L/(1-\alpha))$ and thus either $p_j' = 0$ or $p_L' = 0$. In either case (5.4.2) holds.

By repeated applications of the above lemma, we can see that our problem is reduced to that of determining probability bounds on random variables taking on at most two values. We must thus determine

$$\sup_{S_n \in \mathcal{S}_o} P(S_n \geqslant t),$$

where $\mathcal{S}_o$ is the set of random variables S_n such that each $X_i, i=1,\ldots,n$, takes on at most two values.

Lemma 5.4.2 [16]. Let the nonnegative r.v. X take on two values x_1 and x_2 with probabilities p_1 and p_2 , respectively. For a given t such that $E(X) < t < x_2$, there exists an $\alpha \geqslant 0$ such that the r.v. X' with values $x_1' = x_1 + \alpha$, $x_2' = t$ and respective probabilities p_1, p_2 satisfies

$$0 \leqslant x_1' \leqslant x_2' , \qquad (5.4.7)$$

$$E(X') = E(X). \qquad (5.4.8)$$

Proof: If $\alpha = p_2(x_2 - t)/p_1$, then $x_1' = x_1 + \alpha$, $x_2' = t$ and (5.4.8) is satisfied. Also, $x_1' = (p_1x_1 + p_2x_2 - p_2t)/p_1 = (E(X) - p_2t)/p_1$

$$\leqslant (t - p_2t)/p_1 = t = x_2'.$$

Therefore, $x_2' \geqslant x_1' \geqslant 0$.

Remark: The above lemma tells us that $X_i \leqslant t$, $i = 1,\ldots,n$, for if the larger of the two values of X is greater than t, the larger value could be replaced by t.

Lemma 5.4.3 [100]. Let the r.v. X_i with mean μ_i take on the values a_i and b_i such that $0 \leqslant a_i \leqslant \mu_i \leqslant b_i \leqslant t$, $i = 1,\ldots,n$. Also, let

$$P(X_i = b_i) = \frac{\mu_i - a_i}{b_i - a_i},\ P(X_i = a_i) = \frac{b_i - \mu_i}{b_i - a_i} \quad .$$

If, for some i,

$$P(S_n = t/X_i = b_i) = 0, \tag{5.4.9}$$

where $S_n \in \mathcal{S}_o$ and $a_i < u_i < b_i$, then there exists an $S_n' \in \mathcal{S}_o$ such that

$$P(S_n' \geq t) > P(S_n \geq t). \tag{5.4.10}$$

Proof: Let

$$y = f(b) = \frac{b - \mu_i}{b - a_i} P(S_n \geq t/X_i = a_i) + \frac{\mu_i - a_i}{b - a_i} P(S_n \geq t/X_i = b)$$

such that $f(b_i) = P(S_n \geq t)$. If (5.4.9) is true, then there exists a $\delta > 0$ such that $P(S_n \geq t/X_i = b)$ is constant for $b \in (b_i - \delta, b_i)$. If we differentiate y with respect to b, then, for $b \in (b_i - \delta, b_i)$,

$$y' = f'(b) = [(u_i - a_i)/(b - a_i)^2]\{P(S_n \geq t/X_i = a_i) - P(S_n \geq t/X_i = b_i)\}.$$

If

$$P(S_n \geq t/X_i = a_i) \neq P(S_n \geq t/X_i = b_i), \tag{5.4.11}$$

then $y' < 0$. If (5.4.11) is true, then f(b) is a decreasing function and (5.4.10) holds. If (5.4.11) is false, then let $X_i \equiv \mu_i$; since

$$P(S_n \geq t/X_i = b_i) \geq P(S_n \geq t/X_i = \mu_i) \geq P(S_n \geq t/X_i = a_i),$$

$P(S_n \geq t)$ remains the same for $X_i \equiv \mu_i$.

By the hypothesis we know that $a_i < \mu_i < b_i$; thus

$$P(S_n \geq t/X_i = b_i) - P(S_n \geq t/X_i = a_i) \geq P(S_n = t/X_i = \mu_i) \geq 0. \tag{5.4.12}$$

Thus, since (5.4.11) is false, (5.4.12) implies that $P(S_n \geq t/X_i = \mu_i) = 0$; therefore for $X_i \equiv \mu_i$, $P(S_n = t/X_i = \mu_i, X_j = b_j) = 0$ for all $i \neq j$. Either

$$P(S_n \geq t/X_i = \mu_i, X_j = b_j) = P(S_n \geq t/X_i = \mu_i, X_j = a_j), \tag{5.4.13}$$

for all $i \neq j$, or (5.4.13) is false for some j. If (5.4.13) is true, then let $X_j \equiv \mu_j$ for all j: $P(S_n \geq t)$ remains the same. Also, if $\sum_{i=1}^{n} \mu_i \geq t$, we would have $P(S_n \geq t) = 1$ which is a trivial bound; we therefore assume that $\sum_{i=1}^{n} \mu_i < t$. If this is so, $P(S_n \geq t) = 0$; (5.4.1) would not have a satisfactory upper bound if (5.4.13) were true. If (5.4.13) is false, then, as shown above, we can, by decreasing b_j increase $P(S_n \geq t)$. Therefore, (5.4.10) is true.

Remark: This lemma tells us to restrict our attention to the S_n such that

$$P(S_n = t/X_i = b_i) > 0. \tag{5.4.14}$$

Lemma 5.4.4 [100]. If (5.4.1) is achieved for S_n, where X_i has a_i as its lower value, $i = 1,\ldots,n$, then for $Y_i = X_i - a_i$ and $T = \sum_{i=1}^{n} Y_i$, T_n attains (5.4.1) with t replaced by $t - \sum_{i=1}^{n} a_i$ and $\{\mu_i\}_{i=1}^{n}$ replaced by $\{\mu_i - a_i\}_{i=1}^{n}$.

Proof: Since $E(Y\ ) = \mu_i - a_i$, the replacements are valid. Also,

$$P(S_n \geq t) = P(T_n \geq t - \sum_{i=1}^{n} a_i). \tag{5.4.15}$$

Lemma 5.4.5 [100]. For all $S_n \in \mathcal{S}_o$ which take on only one possible value in $[0,t)$, $P(S_n \geq t)$ is maximized only if S_n is of the form

$$\begin{aligned} &P(X_i = \mu_i) = 1, \quad i = 1,\ldots,k, \\ &P(X_i = 0) = 1 - P(X_i = t - \sum_{j=1}^{k} \mu_j) = 1 - \mu_i/(t - \sum_{j=1}^{k} \mu_j), \\ &\qquad\qquad i = k+1,\ldots,n, \end{aligned} \tag{5.4.16}$$

where $k = 0, 1,\ldots,n-1$.

Proof: Since $P(S_n = t/X_i = b_i) > 0$ for all i, and since S_n takes on only one value in $[0,t)$,

$$b_i - a_i = t - A, \quad \text{for all } i, \tag{5.4.17}$$

where $A = \sum_{i=1}^{n} a_i$. We define the following notation: if $P(X_i = \mu_i) = 1$ for some i, then, $a_i = \mu_i$ and b_i is defined as in (5.4.17); otherwise we let $a_i = 0$ and $b_i = \mu_i$. By the condition of Lemma 5.4.3,

$$P(S_n \geq t) = 1 - \prod_{i=1}^{n} (t - A + a_i - \mu_i)/(t - A) \, . \tag{5.4.18}$$

If, for fixed A, we maximize the right hand side of (5.4.18) subject to $0 \leq a_i \leq u_i$, $i = 1,\ldots,n$, we see that the maximum is achieved when at most one a_i differs from 0 or μ_i and this a_i must take on a value somewhere in $[0,\mu_i]$ by the given condition above. Thus the maximum is attained only if $a_i = 0$ or $a_i = \mu_i$. Q.E.D. (5.4.16)

Lemma 5.4.6 [100]. If for all $S_n \in \mathcal{S}_0$ such that $a_i = 0$, $i = 1,\ldots,n$, $P(S_n \geq t)$ is maximized by an S_n^* which takes on one value in $[0,t)$, and if this result holds for all t and for all μ_i, $i = 1,\ldots,n$, then $P(S_n^* \geq t)$ is maximum among all $S_n \in \mathcal{S}$.

Proof: By Lemma 5.4.4, the hypothesis and our knowledge that when T_n takes on only one value in $[0,t - A)$, S_n takes on only one value in $[0,t)$, we can by Lemma 5.4.5 construct a S_n^* such that $P(S_n^* \geq t)$ is a maximum for all $S_n \in \mathcal{S}$.

Remark: The above lemma tells us that in determining (5.4.1) we need only to consider those random variables X_i for which $a_i = 0$.

Theorem 5.4.1. Let X_1, X_2 be nonnegative independent random variables such that $E(X_i) = \mu_i$, $i = 1,2$, and $\mu_1 \leq \mu_2$. Then, for $t > 0$,

$$\max_{S_2 \in \mathcal{S}} P(S_2 \geq t) = \begin{cases} 1, & \mu_1 + \mu_2 \geq t, \\ \mu_2/(t - \mu_1), & \mu_1 + \mu_2 \leq t \leq t_o, \\ 1 - (1 - \mu_1/t)(1 - \mu_2/t), & t \geq t_o, \end{cases} \qquad (5.4.19)$$

where $t_o = \frac{1}{2}[\mu_1 + 2\mu_2 + (\mu_1^2 + 4\mu_2^2)^{\frac{1}{2}}]$. Equality is attained in the second case if $b_1 = \mu_1$ and $b_2 = t - \mu_1$; it is attained in the third case if $b_1 = b_2 = t$.

Proof: If $\mu_1 + \mu_2 \geq t$, the bound is obviously one and it is attained by the random variables X_i such that $P(X_i = \mu_i) = 1$, $i = 1,2$.

By the previous lemmas we restrict our attention to random variables X_i taking on two values a_i, b_i such that $a_i = 0$, $i = 1,2$. S_2 can take on the four values: 0, b_1, b_2, $b_1 + b_2$.

i) $b_1 + b_2 \geq t > \max(b_1, b_2)$,

ii) $b_1 \geq t > b_2$,

iii) $b_2 \geq t > b_1$,

iv) $\min(b_1, b_2) \geq t$.

By Lemma 5.4.3 we can neglect cases ii) and iii). Case i) will satisfy (5.4.14) only if $b_1 + b_2 = t$.

$$P(S_2 \geq t) = (\mu_1/b_1)(\mu_2/b_2). \qquad (5.4.20)$$

We must thus minimize $b_1 b_2$ subject to $\mu_1 \leq b_1 \leq t - \mu_2$. Since $b_1 + b_2 = t$ and $\mu_1 \leq \mu_2$, (5.4.20) is maximized for $b_1 = \mu_1$ and $b_2 = t - \mu_1$, i.e. the second inequality.

Let us now consider case iv). Since X_i, $i = 1,2$, has support in $[0,t]$ the only possible values for b_1 and b_2 are t and t (Lemma 5.4.3).

By a calculation one can see that

$$\mu_2/(t - \mu_1) > 1 - (1 - \mu_1/t)(1 - \mu_2/t)$$

iff $t < \frac{1}{2}[\mu_1 + 2\mu_2 + (\mu_1^2 + 4\mu_2^2)^{\frac{1}{2}}]$. This completes the proof.

Corollary. If $X_1 = (X - E(X))^2/s^2$, $X_2 = (Y-E(Y))^2/t^2$, $s > 0$, $t = 1$, we obtain the normally referred to Birnbaum, Raymond and Zuckerman inequality.

We now state a result conjectured by Samuels [100]; a proof for the case when n = 3 based on Lemmas 5.6.1 to 5.6.6 and a theorem to support the conjecture is also given by the same author.

Conjecture: If $X_1,\ldots,X_n$ are independent nonnegative random variables such that $E(X_i) = \mu_i$, $i = 1,\ldots,n$, then,

$$\sup_{S_n \in \mathcal{S}} P(S_n \geq t) = \max(P_0,\ldots,P_{n-1}),$$

where $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$ and for $k = 1,\ldots,n-1$,

$$P_0 = 1 - \prod_{i=1}^{n}(1 - \mu_i/t) \; ;$$
$$P_k = 1 - \prod_{i=k+1}^{n}[1 - \mu_i/(t - \sum_{j=1}^{k}\mu_j)] \; . \qquad (5.4.21)$$

The values P_0, P_k are obtained only if S_n is of the form (5.4.16).

5.5 Inequalities of the Kolmogorov Type

Kolmogorov type inequalities can be said to occupy a place in the over-all framework of Tchebycheff type inequalities. These inequalities deal with the task of obtaining upper probability bounds for partial sums, S_k, of independent random variables whose means and variances exist; i.e. upper bounds for

$$P(\max_{1 \leq k \leq n} |S_k - E(S_k)| \geq \epsilon). \qquad (5.5.1)$$

Kolmogorov originally proved that if $X_1,\ldots X_n$ are independent random variables whose variances exist , then,

$$P(\max_{1\leq k\leq n} /S_k - E(S_k)/ \geq \epsilon) \leq \sum_{i=1}^{n} \text{Var } X_i/\epsilon^2. \qquad (5.5.2)$$

If in addition $/X / \leq k$ for some constant k, i = 1,...,n, then,

$$P(\max_{1\leq k\leq n} /S_k - E(S_k)/ \geq \epsilon) \geq 1 - (\epsilon + k)^2/[\sum_{i=1}^{n} \text{Var}(X_i) + (\epsilon + k)^2 - \epsilon^2]. \qquad (5.5.3)$$

(Proofs for (5.5.2) and (5.5.3) can be found in Tucker [114, page 107]). Hajek and Renyi [43] have generalized (5.5.2) and Birnbaum and Marshall [15] have offered a much more general theorem than the latter.

Theorem 5.5.1.(Birnbaum and Marshall [15]). Let $X_1,\ldots,X_n$ be random variables such that $E/X_i/^r < \infty$, i = 1,...,n, and r > 1. Let

$$E(/X_k/ \big| X_1,\ldots,X_{k-1}) \geq \phi_k /X_{k-1}/ \qquad \text{a.e.} , \qquad (5.5.4)$$

where $\phi_k \geq 0$, k = 2,...,n. Let $a_k > 0$, $b_k = \max(a_k, a_{k+1}\phi_{k+1}, a_{k+2}\phi_{k+1}\phi_{k+2},\ldots,a_n \prod_{i=k+1}^{n} \phi_i)$, k=1,...,n, $b_{n+1} = 0$ and $X_0 = 0$. Then

$$P(\max_{1\leq k\leq n} a_k/X_k/ \geq 1) \leq \sum_{k=1}^{n} (b_k^r - \phi_{k+1}^r b_{k+1}^r) E/X_k/^r$$
$$= \sum_{k=1}^{n} b_k^r (E/X_k/^r - \phi_k^r E/X_{k-1}/^r). \qquad (5.5.5)$$

Proof: Without loss of generality we can assume that r = 1 since (5.5.4) implies

$$E(/X_k^r/ \big| X_1^r,\ldots,X_{k-1}^r) \geq \phi_k^r /X_{k-1}^r / \quad \text{a.e.}, \qquad (5.5.6)$$

where $X_j^r = \text{sgn}/X_j^r/$. Let us consider the sequence of sets

$$A_k = \{ a_i/X_i/ < 1, \ i = 1,\ldots,k-1, \ a_k/X_k/ \geq 1 \}_{k=1}^{n} . \qquad (5.5.7)$$

From (5.5.4) and the definition of conditional expectation

$$\int_{A_k} /X_j/dP = E[I_{A_k} E(/X_j/ \big| X_1,\ldots,X_{j-1})] \geq E[I_{A_k} \phi_j /X_{j-1}/]$$
$$= \phi_j \int_{A_k} /X_{j-1}/dP.$$

Continuing this process j - (k + 1) times, we get

$$\int_{A_k} /X_j/ dP \geq \left(\prod_{i=k+1}^{j} \phi_i\right) \int_{A_k} /X_k/ dP. \tag{5.5.8}$$

Also, by the hypothesis,

$$\sum_{j=k}^{n} (b_j - \phi_{j+1} b_{j+1}) \left(\prod_{i=k+1}^{j} \phi_i\right) = b_k \geq a_k \tag{5.5.9}$$

and $b_k \geq \phi_{k+1} b_{k+1}$, $k = 1,\ldots,n$. Using these facts we get

$$\sum_{j=1}^{n} b_j [E/X_j/ - \phi_j E/X_{j-1}/] = \sum_{j=1}^{n} (b_j - \phi_{j+1} b_{j+1}) E/X_j/$$

$$\geq \sum_{k=1}^{n} \sum_{j=1}^{n} (b_j - \phi_{j+1} b_{j+1}) \int_{A_k} /X_j/ dP$$

$$\geq \sum_{k=1}^{n} \sum_{j=k}^{n} (b_j - \phi_{j+1} b_{j+1}) \int_{A_k} /X_j/ dP$$

by (5.5.8) $$\geq \sum_{k=1}^{n} \sum_{j=k}^{n} (b_j - \phi_{j+1} b_{j+1}) \left(\prod_{i=k+1}^{j} \phi_i\right) \int_{A_k} /X_k/ dP$$

$$\geq \sum_{k=1}^{n} \left[\sum_{j=k}^{n} (b_j - \phi_{j+1} b_{j+1}) \left(\prod_{i=k+1}^{j} \phi_i\right)\right] a_k^{-1} P(A_k)$$

by (5.5.9) $$\geq \sum_{k=1}^{n} P(A_k) = P(\max_{1 \leq k \leq n} a_k /X_k/ \geq 1).$$

Remark: By exhibiting a random vector whose distribution attains equality in (5.5.5), the authors show that (5.5.5) is sharp.

The authors also illustrate certain semi-martingale inequalities which are consequences of (5.5.5).

Corollary 1. If $Y_1,\ldots,Y_n$ are random variables such that $E(Y_i) = 0$, $i = 1,\ldots,n$, and $E(Y_k/Y_1,\ldots,Y_{k-1}) = 0$ a.e., $k = 2,\ldots,n$, then, for $a_i = 1/\epsilon$, $i = 1,\ldots,n$, $r = 2$, $\phi_k = 1$ and $X_k = \sum_{i=1}^{k} Y_i$,

$$P(\max_{1 \leq k \leq n} /X_k/ \geq \epsilon) \leq \frac{E/X_n/^2}{\epsilon^2}\ ; \tag{5.5.10}$$

(5.5.10) was proved by both Doob [26, page 315] and Loeve [70, page 386].

Corollary 2. Let X_1 and X_2 be random variables such that $E(X_i) = 0$, $E(X_i^2) = \sigma_i^2 < \infty$, $i = 1, 2$, and $E(X_1 X_2) = \rho\sigma_1\sigma_2$. If $E(X_2/X_1) = \xi x_1$, i.e. the regression is linear, then for the nonnegative constants a_1, a_2,

$$P(a_1|X_1| \geq 1 \text{ or } a_2|X_2| \geq 1) \leq \begin{cases} a_1^2\sigma_1^2 + a_2^2\sigma_2^2(1-\rho^2), & a_1^2\sigma_1^2 \geq a_2^2\sigma_2^2\rho^2, \\ a_2^2\sigma_2^2, & a_1^2\sigma_1^2 \leq a_2^2\sigma_2^2\rho^2. \end{cases} \qquad (5.5.11)$$

Proof: In Theorem 5.5.1 we let $n = r = 2$. Since

$$E(X_2|X_1) = \xi x_1 = \rho\sigma_2 x_1/\sigma_1$$

$$\xi\sigma_1 = \rho\sigma_2.$$

Now $P(\max_i a_i|X_i| \geq 1, i = 1,2) = P(a_1|X_1| \geq 1 \text{ or } a_2|X_2| \geq 1)$; if in (5.5.5) we let $\phi = \xi$, $b_1 = a_1$ and $b_2 = a_2$, we get (5.5.11).

Corollary 3. Let $Y_1,\ldots,Y_n$ be mutually independent random variables each having 0 mean. If $\phi_k = 1$ and $X_k = Y_1 + \ldots + Y_k$, $k = 1,\ldots,n$, then for any non-increasing sequence of positive numbers a_i, $i = 1,2,\ldots n$, and for $r = 2$, we obtain a particular case of the Hájek-Rényi inequality.

As stated and proved in Gnedenko [35] we have the following theorem of Hájek and Rényi.

Theorem 5.5.2. If $X_1,\ldots,X_n$ are mutually independent random variables such that $E(X_i) = 0$ and $Var(X_i) = \sigma_i^2 < \infty$, $i = 1,\ldots,n$, then, for any non-increasing sequence of positive constants a_i, $i = 1,2,\ldots$ and for any positive integers m and n such that $m < n$,

$$P(\max_{m \leq k \leq n} a_k|S_k| \geq \epsilon) \leq \frac{1}{\epsilon^2}\left(a_m^2\sum_{i=1}^{m}\sigma_i^2 + \sum_{i=m+1}^{n} a_i^2\sigma_i^2\right), \qquad (5.5.12)$$

where $\epsilon > 0$.

Remark: If $m = 1$ and $a_i = 1$ for all i we obtain (5.5.2).

Until now the inequalities which we have dealt with in this section have been of the form (5.5.1). Marshall and Olkin [82] give simple inequalities in terms of minimum values of partial sums of independent random variables.

Theorem 5.5.3 [82]. If $X_1,\ldots,X_n$ are mutually independent random variables such that $E(X_i) = 0$ and $\mathrm{Var}(X_i) = \sigma_i^2$, $i = 1,\ldots,n$, then,

$$P(\min_k S_k \geq 1 \text{ or } \min_k(-S_k) \geq 1) \leq \sigma_1^2 , \tag{5.5.13}$$

$$P(\min_k S_k \geq 1) \leq \sigma_1^2/(1 + \sigma_1^2), \tag{5.5.14}$$

$$P(\min_k /S_k/ \geq 1) \leq \sigma_1^2. \tag{5.5.15}$$

Proof: $\left\{\min_k S_k \geq 1 \text{ or } \min_k (-S_k) \geq 1\right\} \subseteq \left\{\min_k /S_k/ \geq 1\right\} \subseteq \left\{/X_1/ \geq 1\right\}$. Thus by the univariate Tchebycheff inequality we obtain (5.5.13) and (5.5.15). Similarly, $\left\{\min_k S_k \geq 1\right\} \subseteq \left\{X_1 \geq 1\right\}$; by the one-sided Tchebycheff inequality we get (5.5.14).

Remark: By exhibiting three different random vectors, each with mutually independent components, whose distributions attain equality in (5.5.13), (5.5.14) and (5.5.15), respectively, Marshall and Olkin are able to show that the above inequalities are sharp.

Similar to the one-sided generalizations of the univariate and multivariate Tchebycheff inequalities, Marshall [79] has proved a one-sided generalization of Kolmogorov's inequality.

Theorem 5.5.4 [79]. Let $X_1,\ldots,X_n$ be random variables such that $E(X_1) = 0$, $E(X_k/X_1,\ldots,X_{k-1}) = 0$ a.e., $k = 2,\ldots,n$, and $E(X_i^2) = \sigma_i^2 < \infty$, $i = 1,\ldots,n$. Then, for $\epsilon > 0$,

$$P(\max_{1\leq k\leq n} S_k \geq \epsilon) \leq s_n/(\epsilon^2 + s_n), \tag{5.5.16}$$

where $s_n = \sum_{i=1}^{n} \sigma_i^2$.

Proof: Let $A_k = \left\{S_i < \epsilon , i = 1,\ldots,k - 1, S_k \geq \epsilon\right\}$, $k = 1,\ldots,n$, and let $f(x) = (\epsilon \sum_{i=1}^{n} x_i + s_n)^2/(\epsilon^2 + s_n)^2$. Now,

$$s_n/(\epsilon^2 + s_n) = \int f(x)dP \geq \sum_{k=1}^{n} \int_{A_k} f(x)dP$$

$$\geq \frac{1}{(\epsilon^2 + s_n)^2} \sum_{k=1}^{n} \int_{A_k} (\epsilon S_k + s_n)^2 dP$$

$$\geq \sum_{k=1}^{n} P(A_k) = P(\max_{1 \leq k \leq n} S_k \geq \epsilon).$$

The inequalities which have been dealt with in this section have been of the form $P(\max(S_1 \geq \epsilon_1, S_2 \geq \epsilon_2, \ldots, S_n \geq \epsilon_n)$ where $\epsilon_1 = \epsilon_2 = \ldots = \epsilon_n$. If the ϵ_i, $i = 1,\ldots,n$ were not all equal, what would be the resulting change of form of the inequality? No general inequality is easily obtained for all n, however, Marshall [79] has offered a bound in the case of $n = 2$.

Theorem 5.5.5. If X_1, X_2 are random variables such that $E(X_1) = 0$, $E(X_2/X_1) = 0$ a.e., and $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$, $i = 1,2$, then, for $\epsilon_1 > 0$ and $\epsilon_2 > 0$,

$$P(X_1 \geq \epsilon_1 \text{ or } X_1 + X_2 \geq \epsilon_2) \leq [\sigma_2^2 + \sigma_1^2(\alpha_2/\alpha_1)^2]/(\sigma_2^2 + \alpha_2^2/\alpha_1), \quad (5.5.17)$$

where $\alpha_i = \sigma_1^2 + t_1 t_i$, $i = 1,2$, and $t_1 = \min(\epsilon_1, \epsilon_2)$, $t_2 = \epsilon_2$.

Proof: Let $f(x_1,x_2) = c_1 f_1^2(x_1) + c_2 f_2^2(x_1 + x_2)$, where

$c_1 = t_1^2/\alpha_1^2 - t_1^2(\alpha_2 + \sigma_2^2)^2/(\alpha_2^2 + \sigma_2^2\alpha_1)^2$, $c_2 = t_1^2\alpha_2^2/(\alpha_2^2 + \sigma_2^2\alpha_1)^2$,

$f_1(x) = (x + \sigma_1^2/t_1)$, $f_2(x) = x + \sigma_1^2/t_1 + \sigma_2^2\alpha_1/t_1\alpha_2$.

Let us consider the sets $A_1 = \{X_1 \geq t_1\}$, $A_2 = \{X_1 < t_1, X_1 + X_2 \geq t_2\}$. Since $\alpha_2 \geq \alpha_1 > 0$, we know that $c_1 \geq 0$. If we proceed the same way as in Theorem 5.5.4, we obtain

$$\int_{A_1} f(x_1,x_2)dP \geq P(A_1), \int_{A_2} f(x_1,x_2)dP \geq \int_{A_2} c_2 f_2^2(x_1 + x_2)dP = P(A_2).$$

By integration we see that the right hand side of (5.5.17) $= \int f(x_1,x_2)dP$.

$$\int f(x_1,x_2)dP \geq P(A_1) + P(A_2) = P(X_1 \geq t_1 \text{ or } X_1 + X_2 \geq t_2)$$

$$\geqslant P(X_1 \geqslant \epsilon_1 \text{ or } X_1 + X_2 \geqslant \epsilon_2).$$

Through an example which attains equality, Marshall is able to prove that the inequality is sharp.

5.6 Inequalities for Bounded Random Variables

In this chapter we have not yet reviewed any work in which restrictions were imposed on the individual random variables. The restrictions which we shall impose on the random variables of this section are

i) boundedness from above of the absolute moments of the random variables,

ii) boundedness of the random variable.

The impetus for introducing these restrictions has been the work of Bernstein [10,11,23] who obtained an inequality for the sum of a finite number of independent random variables each of whose absolute moments are bounded from above.

Theorem 5.6.1. Let $X_1,\ldots,X_n$ be mutually independent random variables and let $E(X_i) = 0$, $\mathrm{Var}(X_i) = \sigma_i^2$, and

$$v_{ri} \leqslant \tfrac{1}{2}\sigma_i^2 W^{r-2}(r!),\ i = 1,\ldots,n,\ r \geqslant 2, \tag{5.6.1}$$

where W is a constant. Then for $S_n = \sum_{i=1}^{n} X_i$, $s_n^2 = \sum_{i=1}^{n}\sigma_i^2$,

$$P(/S_n/ \geqslant ts_n) \leqslant 2\exp(-t^2/2(1 + Wt/s_n)). \tag{5.6.2}$$

Proof:

$$E(\exp(cX_i)) = 1 + cE(X_i) + \sum_{r=2}^{\infty} \frac{c^r E(X_i^r)}{r!} \tag{5.6.3}$$

$$= 1 + \tfrac{1}{2}c^2\sigma_i^2 \sum_{r=2}^{\infty} \frac{c^{r-2}E(X_i^r)}{\tfrac{1}{2}r!\sigma_i^2}. \tag{5.6.4}$$

Let us denote $F_i = \sum_{r=2}^{\infty}\left(c^{r-2}E(X_i^{\,r})/\frac{1}{2}\,\sigma_i^{\,2}r!\right)$.

$$E(\exp(cX_i)) = 1 + \tfrac{1}{2}c^2\sigma_i^{\,2}F_i \leqslant \exp(\tfrac{1}{2}\,c^2\sigma_i^{\,2}F_i). \qquad (5.6.5)$$

$$E(\exp(cS_n)) \leqslant \prod_{i=1}^{n} \exp(\tfrac{1}{2}\,c^2\sigma_i^{\,2}F_i) \leqslant \exp(\tfrac{1}{2}\,c^2 s_n^{\,2}F)\,, \qquad (5.6.6)$$

where $F = \max(F_i)$, $i = 1,\ldots,n$. (5.6.7)

Since $E(X_i^{\,r}) \leqslant E(/X_i/^r)$, then, by (5.6.1)

$$F_i \leqslant \sum_{r=2}^{\infty} \frac{c^{r-2}\sigma_i^{\,2}W^{r-2}r!}{\sigma_i^{\,2}r!} = \sum_{r=2}^{\infty}(cW)^{r-2} = (1 - cW)^{-1}, \qquad (5.6.8)$$

where $cW < 1$. Thus by (5.6.7)

$$F \leqslant (1 - cW)^{-1}. \qquad (5.6.9)$$

However, from Theorem 2.1.3 we know that if $h(y)$ is a nonnegative function of a r.v. Y such that $h(y) \geqslant b$ whenever $y \geqslant a$,

$$P(Y \geqslant a) \leqslant E(h(Y))/b.$$

If $h(y) = \exp(cy)$, then for $c > 0$,

$$P(Y \geqslant a) \leqslant E(\exp(cY))/\exp(ca). \qquad (5.6.10)$$

If $Y = S_n$ and $a = ts_n$, then

$$P(S_n \geqslant ts_n) \leqslant \exp(\tfrac{1}{2}\,c^2 s_n^{\,2}F - cts_n). \qquad (5.6.11)$$

We would like to minimize the right hand side of (5.6.11) for values in c. Minimizing, we get

$$c = t/s_nF; \qquad (5.6.12)$$

$$P(S_n \geqslant ts_n) \leqslant \exp(-t^2/2F) = \exp(-\tfrac{1}{2}\,ct\varpi). \qquad (5.6.13)$$

Combining (5.6.12) and (5.6.9), we obtain

$$F = (t/cs_n) \leqslant (1 - cW)^{-1},$$

$$c \geqslant t/(s_n + tW) \text{ and } cW \geqslant tW/(s_n + tW).$$

If we put the above value of c in (5.6.13) we get

$$P(S_n \geqslant ts_n) \leqslant \exp(-t^2/2(1 + Wt/s_n)). \qquad (5.6.14)$$

$$P(|S_n| \geqslant ts_n) \leqslant 2\exp(-t^2/2(1 + Wt/s_n)).$$

If we are told that $X_i < K$, for all i, then, since we know that for each X_i

$$v_{ri} \leqslant K^{r-2}\sigma_i^2 \leqslant \tfrac{1}{2}\sigma_i^2 r!(K/3)^{r-2} ,$$

we can, by (5.6.1), conclude that $W = K/3$.

$$P(|S_n| \geqslant ts_n) \leqslant 2\exp(-t^2/2(1 + Kt/3s_n)). \qquad (5.6.15)$$

Two authors, Bennett [6,7] and Hoeffding [48] , have used variations of Bernstein's technique described in equations (5.6.3) to (5.6.11) to obtain improvements over Bernstein's bound.

Theorem 5.6.2 [6]. Let us assume that the conditions of Theorem (5.6.1) prevail and that each X_i is bounded from above by K; i.e. $v_{ri} \leqslant K^{r-2}\sigma_i^2$, $i = 1,\dots,n$, $r > 2$. Then,

$$P(S_n \geqslant ts_n) \leqslant \exp\left(-t^2/\left(1 + Wt/s_n + (1 + 2Wt/s_n)^{\frac{1}{2}}\right)\right) , \qquad (5.6.16)$$

and

$$P(S_n \geqslant ts_n) \leqslant \exp(ts_n/K)(1 + tK/s_n)^{-(t(s_n/K) + (s_n/K)^2)} . \qquad (5.6.17)$$

Remark: (5.6.17) was also obtained by Hoeffding [48 , equation (2.9)].

Proof: Through the use of (5.6.9) we can rewrite (5.6.11) as

$$P(S_n \geqslant ts_n) \leqslant \exp(c^2 s_n^2/(2 - 2cW) - cts_n). \qquad (5.6.18)$$

Minimizing the right hand side of (5.6.18) with respect to c, we obtain $cW = 1 - (1 + 2Wt/s_n)^{-\frac{1}{2}}$ and substituting this value into (5.6.18), we get (5.6.16). Again we can make the substitution $W = K/3$ to get a new inequality. For given values of Wt/s_n , (5.6.16) will always offer lower bounds than (5.6.14).

To determine (5.6.17), we note that

$$F_i = \sum_{r=2}^{\infty}\left(c^{r-2}E(X_i^r)/\tfrac{1}{2}\, r!\sigma_i^2\right) \leqslant \sum_{r=2}^{\infty}\left(c^{r-2}K^{r-2}\sigma_i^2/\tfrac{1}{2}\, r!\sigma_i^2\right)$$

$$= \sum_{r=2} \left((cK)^{r-2} / \tfrac{1}{2}\, r! \right) = 2(e^{cK} - 1 - cK)/(cK)^2.$$

Thus $F \leq 2(e^{cK} - 1 - cK)/(cK)^2$, and (5.6.11) becomes

$$P(S_n \geq ts_n) \leq \exp\left[(s_n/K)^2 (e^{cK} - 1 - cK) - cts_n \right] . \tag{5.6.19}$$

Minimizing the right hand side of (5.6.19) with respect to c, we get $cK = \ln\,(1 + Kt/s_n)$,and substituting this value into (5.6.19) will give us (5.6.17).

Bennett [7] has also utilized this approach to obtain a bound for the sum of independent symmetric bounded random variables.

Bernstein's technique will now be used to establish the following two theorems of Hoeffding [48]. The Bernstein technique is presented in the following way similar to the approach used in the proof of Theorem (5.6.1). Let $c > 0$, then

$$P(\overline{X} - E(\overline{X}) \geq t) = P(S_n - E(S_n) \geq nt) \leq E \exp(c(S_n - E(S_n)))\, / e^{cnt}$$
$$= e^{-cnt} \prod_{i=1}^{n} E\left(\exp(c(X_i - E(X_i))) \right) . \tag{5.6.20}$$

Theorem 5.6.3 [48]. If $X_1, \ldots, X_n$ are independent random variables such that $0 \leq X_i \leq 1$, $i = 1, \ldots, n$, and $E(S_n) = M$, then for $0 \leq t < 1 - \mu$,

$$P(\overline{X} - \mu \geq t) \leq \left((\mu/(\mu + t))^{\mu+t} \,((1 - \mu)/(1 - \mu - t))^{1-\mu-t} \right)^n \tag{5.6.21}$$
$$\leq \exp(-nt^2 g(\mu)) \tag{5.6.22}$$
$$\leq \exp(-2nt^2) , \tag{5.6.23}$$

where $\mu = M/n$,

$$g(\mu) = \begin{cases} (1/(1 - 2\mu))\ln((1 - \mu)/\mu), & 0 < \mu < \tfrac{1}{2}, \\ 1/2\mu(1 - \mu), & \tfrac{1}{2} \leq \mu < 1. \end{cases} \tag{5.6.24}$$

Remark: If we were to say $a \leq X_i \leq b$, $i = 1, \ldots, n$, we would have to replace μ and t by $(\mu - a)/(b - a)$ and $t/(b - a)$, respectively, in the inequalities.[48].

Proof: We begin by proving the following lemma.

Lemma 5.6.1. If X is a r.v. such that $a \leqslant X \leqslant b$, then for any real c,

$$E(\exp(cX)) \leqslant \frac{(b - E(X))e^{ca}}{b - a} + \frac{(E(X) - a)e^{cb}}{b - a} . \qquad (5.6.25)$$

Proof: Let us join the points $(a,\exp(ca))$ and $(b,\exp(cb))$. The line joining these two points lies above $\exp(cx)$ since $\exp(cx)$ is convex.

$$\exp(cx) \leqslant \frac{(b - x)e^{ca}}{b - a} + \frac{(x - a)e^{cb}}{b - a} , \quad a \leqslant x \leqslant b.$$

Taking expectations we get (5.6.25).

We return to the proof of the theorem. For $c > 0$, (5.6.20) tells us that

$$P(\overline{X} - \mu \geqslant t) \leqslant \exp(-cnt - cn\mu)\prod_{i=1}^{n} E(\exp(cX_i)). \qquad (5.6.26)$$

If $E(X_i) = \mu_i$, $i = 1,\ldots,n$, then for $a = 0$, $b = 1$, Lemma 5.6.1 tells us that

$$\prod_{i=1}^{n} E(\exp(cX_i)) \leqslant \prod_{i=1}^{n}(1 - \mu_i + \mu_i \exp(c)). \qquad (5.6.27)$$

Since the arithmetic mean is greater or equal to the geometric mean,

$$\left\{\prod_{i=1}^{n}(1 - \mu_i + \mu_i \exp(c))\right\}^{1/n} \leqslant \frac{\sum_{i=1}^{n}(1 - \mu_i + \mu_i \exp(c))}{n} = 1 - \mu + \mu \exp(c) ,$$

we can combine (5.6.26) and (5.6.27) to give

$$P(\overline{X} - \mu \geqslant t) \leqslant \left\{\exp(-ct - c\mu)(1 - \mu + \mu \exp(c)\right\}^{n} . \qquad (5.6.28)$$

Minimizing (5.6.28) with respect to c we get

$$c = \ln[(1 - \mu)(\mu + t)/(1 - \mu - t)\mu] ,$$

and substituting this value into (5.6.28) we get (5.6.21).

Let the right hand side of (5.6.21) be of the form

$$\exp(-nt^2 G(t,\mu)),$$

where $t^2 G(t,\mu) = (\mu + t)\ln[(\mu + t)/\mu] + (1 - \mu - t)\ln[(1 - \mu - t)/(1 - \mu)]$.

Now $t^2 \frac{\partial}{\partial t} G(t,\mu) = f(t/(1 - \mu)) - f(t/(\mu + t))$, $\qquad (5.6.29)$

where $f(b) = (1 - 2/b)\ln(1 - b)$. For $|b| < 1$ we can expand $f(b)$;

$$f(b) = 2 + \left(\frac{2}{3} - \frac{1}{2}\right)b^2 + \left(\frac{2}{4} - \frac{1}{3}\right)b^3 + \left(\frac{2}{5} - \frac{1}{4}\right)b^4 + \ldots$$

Thus, for $0 < b < 1$, $f'(b) > 0$ and by calculus we know that $f(b)$ is an increasing function of b. By (5.6.29) $\frac{\partial}{\partial t}G(t,\mu) > 0$ iff $t/(1 - \mu) > t/(\mu + t)$, i.e. $t > 1 - 2\mu$. If $1 - 2\mu > 0$, then $G(t,\mu)$ assumes a minimum value at $t = 1 - 2\mu$, i.e. $\min G(t,\mu) = g(\mu)$ as defined in the upper value in (5.6.24); if $1 - 2\mu < 0$, then $G(t,\mu)$ is minimized at $t = 0$ and $\min G(t,\mu) = 1/[2\mu(1 - \mu)] = g(\mu)$ as defined in the lower value in (5.6.24). Q.E.D. (5.6.22). (5.6.23) follows immediately since $g(\mu) \geq g(\frac{1}{2}) = 2$.

Hoeffding has extended the above result to the case when the X_i are not bounded by the same constants.

Theorem 5.6.4 [48]. Let $X_1,\ldots,X_n$ be independent random variables such that $a_i \leq X_i \leq b_i$, $i = 1,\ldots,n$. Then for $t > 0$,

$$P(\overline{X} - \mu \geq t) \leq \exp[-2n^2t^2/\sum_{i=1}^{n}(b_i - a_i)^2], \tag{5.6.30}$$

where $\mu = E(\overline{X})$.

Proof: Let $\mu_i = E(X_i)$. By Lemma 5.6.1 we get

$$E(\exp(c[X_i - \mu_i])) \leq e^{-c\mu_i}\left[\left(\frac{b_i - \mu_i}{b_i - a_i}\right)e^{ca_i} + \left(\frac{\mu_i - a_i}{b_i - a_i}\right)e^{cb_i}\right] = e^{L(h_i)}, \tag{5.6.31}$$

where $L(h_i) = -h_ip_i + \ln(1 - p_i + p_ie^{h_i})$, $h_i = c(b_i - a_i)$, and $p_i = (\mu_i - a_i)/(b_i - a_i)$.

Differentiating twice we obtain

$$L'(h_i) = -p_i + p_i/[(1 - p_i)e^{-h_i} + p_i],$$

$$L''(h_i) = \frac{p_i(1 - p_i)e^{-h_i}}{[(1 - p_i)e^{-h_i} + p_i]^2}. \tag{5.6.32}$$

(5.6.32) is of the form $q(1 - q)$ where $0 < q < 1$ and

$q = [(1 - p_i)e^{-h_i}]/[(1 - p_i)e^{-h_i} + p_i]$. $L''(h_i) \leq \frac{1}{4}$. Using the Taylor expansion we get

$$L(h_i) \leq L(0) + L'(0)h_i + h_i^2/8 = c^2(b_i - a_i)^2/8. \tag{5.6.33}$$

By (5.6.20) we can write

$$P(\overline{X} - \mu \geq t) \leq e^{-cnt} \prod_{i=1}^{n} E[\exp(c(X_i - \mu_i))]. \tag{5.6.34}$$

Combining (5.6.33), (5.6.31) and (5.6.34) we get

$$P(\overline{X} - \mu \geq t) \leq \exp(-cnt + \frac{1}{8} c^2 \sum_{i=1}^{n}(b_i - a_i)^2). \tag{5.6.35}$$

Minimizing (5.6.35) with respect to c, we get $c = 4nt/[\sum_{i=1}^{n}(b_i - a_i)^2]$. Substituting this value of c into (5.6.35), we obtain (5.6.30).

Corollary. If $Y_1,\ldots,Y_m$, $Z_1,\ldots,Z_n$ are independent random variables defined on $[a,b]$, then for $t > 0$,

$$P(\overline{Y} - \overline{Z} - [\, E(\overline{Y}) - E(\overline{Z})\,] \geq t) \leq \exp[-2t^2/(m^{-1} + n^{-1})(b - a)^2]. \tag{5.6.36}$$

Though Bennett [6] and Hoeffding [48] have been able to construct distributions which attain equality in their respective theorems, Kingman [65] points out that the inequalities are in general not the best since the independence of the random variables has not been fully exploited; the inequalities, however, are the sharpest that can be obtained using the Bernstein approach.

5.7 Other Inequalities

In this section we shall briefly mention some inequalities for sums of random variables without offering any proofs. Similar to the ideas in section 5.4, Birnbaum [14] considers a reduction of the problem. He introduces the concept of comparable peakedness as follows. If X_1 and and X_2 are random variables and a_1 and a_2 are any two real constants, then "X_1 is more peaked about a_1 than X_2 about a_2" if, for $t > 0$,

$$P(/X_1 - a_1/ \geq t) \leq P(/X_2 - a_2/ \geq t).$$

Let $X_1,\ldots,X_n$ be a random sample from a continuous symmetrical unimodal (mode at zero) distribution satisfying $P(/X_i/ > a) = 0$, $i = 1,\ldots,n$. Using the concept of comparable peakedness, Birnbaum shows that, for a random sample $Z_1,\ldots,Z_n$ from a uniform distribution on $[-1,1]$, $\overline{X}/a$ is more peaked about zero than $\overline{Z}$. The problem of finding probability bounds for $\overline{X}$ is reduced to that of finding probability bounds for a sum of uniformly distributed random variables.

Authors such as Hoeffding [47] and Samuels [99] have offered inequalities for the number of sucesses occuring in n independent trials. Romanovski [97] has offered an upper bound for the sum of the deviations between the observed and expected frequencies in n independent trials.

In this chapter we have omitted discussion of the class of martingale and semi-martingale inequalities; some results in this field can be found in Doob [26]. Some recent contributors to this field have been Blackwell [17] and Dubins and Savage [27]. They concerned themselves with gambling problems, established certain gambling systems for a sequence of random variables, and they have found probability bounds for the sum of random variables occuring in some intervals. Dubins and Savage [28] have collected the results into a book titled "How to Gamble if You Must". The title is misleading and a subtitle "Inequalities for Stochastic Processes" is given to the book.

Though the field dealing with inequalities for stochastic processes can fall under the heading of Tchebycheff type inequalities, we believe it is well enough established so as to occupy a place of its own; we have thus not dealt with the above type of inequality.

CHAPTER SIX: APPLICATION AND CONCLUSION

Tchebycheff's inequality is a useful tool in probability theory. It can be used to establish the weak law of large numbers and Bernoulli's theorem; it is often used as a background for establishing convergence theorems [4,35]. Many of the inequalities established in chapters two, three and five can be applied in practical situations and in industrial situations such as quality control [125] and polymer research [44,45]; they can also be used in establishing various tests of hypothesis [109].

The inequalities in chapter two can almost always be applied and those of chapter three can be applied if one verifies that the r.v. obeys the restriction imposed upon it. In chapter five most of the inequalities can be applied without too much difficulty whereas the results of chapter four may be quite difficult to apply in general since some probability bounds depend on the solution of a matrix equation which has no general solution.

We now offer an example to illustrate one application of a Tchebycheff type inequality to a medical problem. Type A bacterial meningitis and type B bacterial meningitis are two diseases which have adverse effects on children. Both diseases can be treated by different drugs, however, the treatment must start immediately so as to prevent any permanent brain damage. When a child suffering from either of the diseases is brought to a hospital, it is quite uncommon for the doctor to be able to make an immediate diagnosis; the immediate diagnosis of the exact type of meningitis is often not possible with the available clinical methods. Ideally the patient could be treated with an excess of drugs, i.e. drugs to combat both types of disease. Certain drugs, however, can have negative effects on the patient. Therefore an "educated guess" is often used by the doctors when instituting initial emergency therapy.

A survey of both types of disease has shown that, when plotted against age, the occurence of both diseases is unimodally distributed. Type A disease has a mean value of five years ten months and a mode of three years one month while type B disease has a mean value of eight years seven months and a mode of four years three months. The medical director of a hospital would like to set up certain confidence intervals which would enable his doctors to make an educated guess as to which disease the child is suffering from.

We illustrate the technique of making an educated guess through the use of Theorem 3.6.1. Let us count the children's ages in months. Let X and Y represent the distributions of type A and type B meningitis, respectively. We shall consider the age group greater or equal to sixty months. Applying Theorem 3.6.1 to the r.v. X, we obtain

$$P(X \geqslant 60) \leqslant 0.65.$$

Similarly, if we apply Theorem 3.6.1 to the r.v.Y, we obtain

$$P(Y \geqslant 60) \leqslant 0.91.$$

If we now consider the age group greater or equal to seventy-two months , Theorem 3.6.1 tells us that

$$P(X \geqslant 72) \leqslant 0.50, \ P(Y \geqslant 72) \leqslant 0.81.$$

Combining the above we see that

$$P(60 \leqslant X \leqslant 72) \leqslant 0.15, \qquad P(60 \leqslant Y \leqslant 72) \leqslant 0.10.$$

By repeated applications of Theorem 3.6.1 the doctor could establish upper probability bounds for various age intervals. These probability bounds would enable to make an "educated guess" as to the type of bacterial meningitis. In times of uncertainty when doctors are forced to make "educated guesses", decisions, based on probability bounds derived from Tchebycheff type inequalities, would prove to be very helpful.

We have attempted in this thesis to give results which are of value to the practising statistician without neglecting the basic theory which led to these results. We have traced the development of a general theorem and we have shown how certain inequalities can be established from this general theorem. We have also given other inequalities not obtainable by the general theorem and we have illustrated the various approaches which have led to the establishment of these sharp Tchebycheff type inequalities.

BIBLIOGRAPHY

Papers not seen by the author are denoted by an *.

[1]. Aoyama, H. (1952). On Midzuno's inequality. Ann. Inst. Statist. Math. 3, 65-67.

[2]. Barlow. R. E., and A. W. Marshall (1964). Bounds for distributions with monotone hazard rate I. Ann. Math. Statist. 35, 1234-1257.

[3]. Barlow, R. E., and A. W. Marshall (1964). Bounds for distributions with monotone hazard rate II. Ann. Math Statist. 35, 1258-1274.

[4]. Basman, R. L. (1965). A Tchebycheff inequality for the convergence of a generalized classical linear estimator, sample size being fixed. Econometrika. 33, 608-618.

[5]. Beckenbach, E. F. and R. Bellman (1961). Inequalities. Springer-Verlag, Berlin.

[6]. Bennett, G. (1962). Probability inequalities for the sum of independent random variables. J. Amer. Statist. Assoc. 57, 33-45.

[7]. Bennett, G. (1963). On the probability of large deviations from the expectation for sums of bounded, independent random variables. Biometrika. 55, 528-535.

[8]. Berge, P. O. (1937). A note on a form of Tchebycheff's theorem for two variables. Biometrika. 29, 405-406.

[9]. Bergström, H. (1949). On the central limit theorem in the case of not identically distributed random variables. Skand. Aktuarietidskr. 32, 37-62.

*[10]. Bernstein, S. (1924). Sur une modification de l'inégalité de Tchebichef. Ann. Sc. Instit. Sav. Ukraine, Sect. Math. I. 38-49. (Russian, French summary).

[11]. Bernstein, S. (1937). Sur quelques modifications de l'inégalité de Tchebycheff. Comptes Rendus (Doklady) Acad. Sc. URSS. 17, 279-282.

[12]. Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Amer. Math. Soc. 49, 122-136.

[13]. Bienaymé, J. (1853). Considérations à l'appui de la découverte de Laplace sur la loi des probabilité dans la méthode des moindres carrés. Comptes Rendus (Paris) 37, 309-326.

[14]. Birnbaum, Z. W. (1948). On random variables with comparable peakedness. Ann. Math. Statist. 19, 76-81.

[15]. Birnbaum, Z. W., and Albert W. Marshall (1961). Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. Ann. Math. Statist. 32, 687-703.

[16]. Birnbaum, Z. W., J. Raymond, and H. S. Zucherman (1947). A generalization of Tshebyshev's inequality to two dimensions. Ann. Math. Statist. 18, 70-79.

[17]. Blackwell, D. (1954). On optimal systems. Ann. Math. Statist. 25, 394-397.

[18]. Camp, B. H. (1922). A new generalization of Tchebycheff's statistical inequality. Bull. of the Amer. Math. Soc. 28, 427-432.

[19]. Camp, B. H. (1948). Generalization to N dimensions of inequalities of the Tchebycheff type. Ann. Math. Statist. 19, 568-574.

*[20]. Cantelli, F. P. (1910). Intorno ad un teorema fondamentale della teoria del rischio. Boll dell' Associazione degli Attuari Italiani (Milan). 1-23.

*[21]. Cantelli, F. P. (1928). Sui confini della probabilità. Atti del Congresso Internazional del Matematici (Bologna, 3-10 Settembre 1928). 6, 47-59.

[22]. Chapelon, Jaques M. (1937). Sur l'inégalité fondamentale du calcul des probabilités. Bull. de la Société Mathématique de France (Paris). 65, 100-108.

[22b]. Chew, V. (1968). Some useful alternatives to the normal distribution. American Statistician. June 1968.

[23]. Craig, C.C. (1933). On the Tchebycheff inequality of Bernstein. Ann. Math. Statist. 4, 94-102.

*[24]. Cramér, H. (1923). Das Gesetz von Gauss und die Theorie des Risikos. Skand. Aktuarietidskr. 6, 209-237.

[25]. Cramér, H. (1928). On the composition of elementary errors. Skand. Aktuarietidskr. 11, 13-74, 141-180.

[26]. Doob, J. L. (1953). Stochastic Processes. John Wiley and Sons, N. Y.

[27]. Dubins, L. E., and L. J. Savage (1965). A Tchebycheff-like inequality for stochastic processes. Natl. Acad. Sci. 53, 274-275.

[28]. Dubins, L. E., and L. J. Savage (1965). How to Gamble If You Must. McGraw-Hill, New York.

[29]. Faddeev, D. K. and V. N. Faddeeva (1963). Computational Methods of Linear Algebra. Translation from Russian. W. H. Freeman and Co., San Francisco.

[30]. Feller, W. (1966). An Introduction to Probability Theory and Its Applications. Vol. 2. John Wiley and Sons, N. Y.

*[31]. Gauss, C. F. (1821). Theoria combinationis observationum. Werke. 4, 10-11 (Goettingen).

[32]. Glasser, G. J. (1961). Tchebycheff -type inequalities in terms of the mean deviation. Sankhya. 23A, 397-400.

[33]. Godwin, H. (1955). On generalizations of Tchebychef's inequality. J. of the Amer. Statist. Assoc. 50, 923-945.

[34]. Godwin, H. (1964). Inequalities on Distribution Functions. Griffin's Statistical Monographs and Courses. Hafner, N. Y.

[35]. Gnedenko, B. V. (1962) Theory of Probability. Translated by B. D. Seckler. Chelsea, N.Y.

[36]. Gnedenko B. V. (1944). Limit Theorems for Sums of Independent Random Variables. Uspekhi Mat. Nauk 10, 115-165. English Translation: Translation no. 45, American Mathematical Society, New York.

[37]. Guldberg, A. (1922). Sur le théorème de M. Tchebycheff. Comptes Rendus (Paris). 175, 418-420.

[38]. Guldberg, A. (1922). Sur un théorème de M. Markoff. Comptes Rendus (Paris). 175, 679-680.

[39]. Guldberg, A. (1922). Sur les valeurs moyennes. Comptes Rendus (Paris). 175, 1035-1037.

[40]. Guldberg, A. (1922). Sur quelques inégalités dans le calcul des probabilités. Comptes Rendus (Paris). 175, 1382-1384.

[41]. Guttman, L. (1948). An inequality for kurtosis. Ann. Math. Statist. 19, 277-278.

[42]. Guttman, L. (1948). A distribution-free confidence interval for the mean. Ann. Math. Statist. 19, 410-413.

*[43]. Hájek, J., and A. Rényi. (1955). Generalization of an inequality of Kolmogorov. Acta Math. Sci. Hung. 6, 281-284.

[44]. Herdan G. (1949). Use of statistical inequalities in polymer research. Research. 2, 235-237.

[45]. Herdan, G. (1950). Use of inequalities in polymer research. Research. 3, 35-41.

[46]. Hoeffding, W. (1955). The extrema of the expected value of a function of independent random variables. Ann. Math. Statist. 26, 268-275.

[47]. Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. Ann. Math. Statist. 27, 713-721.

[48]. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13-29.

[49]. Hoeffding, W., amd S. S. Shrinkhande(1955). Bounds for the distribution function of a sum of independent, identically distributed random variables. Ann. Math. Statist. 26, 439-449.

[50]. Hsu, P. L. (1945). The approximate distributions of the mean and variance of a sample of independent variables. Ann. Math. Statist. 16, 1-29.

[51]. Ikeda, S. (1959). A note on the normal approximation to the sum of independent random variables. Ann. Inst. Statist. Math. Tokyo. 11, 121-130.

[52]. Isii, K. (1957). Some investigations of the relation between distribution functions and their moments. Ann. Inst. Statist. Math. 9, 1-11.

[53]. Isii, K. (1959). On Tchebycheff type inequalities. Proc. Inst. Statist. Math. 7, 123-143. (Japanese, English summary).

[54]. Isii, K. (1959). On a method for generalizations of Tchebycheff's inequality. Ann. Inst. Statist. Math. 10, 65-88.

[55]. Isii, K. (1959). Bounds on probability for nonnegative random variables. Ann. Inst. Statist. Math. 11, 89-99.

[56]. Isii, K. (1960). The extrema of probability determined by generalized moments: [I] Bounded random variables. Ann. Inst. Statist. Math. 12, 119-133.

[57]. Isii, K. (1963). On sharpness of Tchebycheff-type inequalities. Ann. Inst. Statist. Math. 14, 185-197.

[58]. Isii, K. (1964). Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming. Ann. Inst. Statist. Math. 16, 277-293.

[59]. Johnson, N. L., and C. A. Rogers (1951). The moment problem for unimodel distributions. Ann.Math.Statist. 22, 433-439.

[60]. Karlin, S. J., and W. J. Studden (1966). Tchebycheff Systems: with Applications in Analysis and Statistics. John Wiley and Sons, New York.

[61]. Kemperman, J. H. B. (1965). On the sharpness of Tchebycheff type inequalities. Indag. Math. 27, 554-601.

[62]. Kendall, M. G., and A. Stuart (1963). The Advances Theory of Statistics. Vol. 1. Charles Griffin and Co. London.

[63]. Khamis, S. H. (1950). A note on the general Tchebycheff inequality. Proc. Intern. Congr. Math. 1, 569-570. Amer. Math. Soc.

[64]. Khamis, S. H. (1954). On the reduced moment problem. Ann. Math. Statist. 25, 113-122.

[65]. Kingman, J. (1963). On inequalities of the Tchebychev type. Proc. Cambridge Phil. Soc. 59, 135-146.

[66]. Kolmogorov, A. N. and B. V. Gnedenko (1954). Limit Distributions for Sums of Independent Random Variables, Addison-Wesley, Reading, Mass.

[67]. Krein, M. G. (1951). The ideas of P. L. Čebyšev and A. A. Markov in the Theory of Limiting Values of Integrals and Their Further Developments. American Mathematical Society Translation, Ser. 2, 12, 1-122.

[68]. Lal, D. N. (1955). A note on a form of Tchebycheff's inequality for two or more variables. Sankhya. 15, 317-320.

[69]. Leser, C. E. V. (1942). Inequalities for multivariate frequency distrubutions. Biometrika. 32, 284-293.

[70]. Loeve, M. (1955). Probability Theory. D. Van Nostrand. N. Y.

[71]. Lurquin, C. (1922). Sur le critérium de Tchebycheff. Comptes Rendus (Paris). 175, 681-683.

[72]. Lurquin, C. (1924). Sur une proposition fondamentale de probabilité. Comptes Rendus (Paris). 178, 306-308.

[73]. Lurquin, C. (1928). Sur une inégalité fondamentale de probabilité. Comptes Rendus (Paris). 187, 868-870.

[74]. Lurquin, C. (1928). Sur une limite de probabilité du sens de Bienaymé-Tchebycheff. Comptes Rendus (Paris).

[75]. Lurquin, C. (1928). Sur une théorème de limite pour les probailitiés au sens de Bienaymé-Tchebycheff. Bull. de l'Acadamie Brux. 14, 641-658.

[76]. Mallows, C. L. (1956). Generalizations of Tchebycheff's inequalities. J. Roy. Statist. Soc. Ser. B. 18, 139-176.

[77]. Mallows, C. L. (1963). A generalization of Chebyshev inequalities. Proc. London Math. Soc. Third Series. 13, 385-412.

[78]. Marcus, M. (1957). Convex functions of quadratic forms. Duke Math. J. 24, 321-325.

[79]. Marshall, A. W. (1960). A one-sided analog of Kolmogoroff's inequality. Ann. Math. Statist. 31, 483-487.

[80]. Marshall, A. W., and I. Olkin (1960). A one-sided inequality of the Chebyshev type. Ann. Math. Statist. 31, 488-491.

[81]. Marshall, A. W. and I. Olkin (1960). Multivariate Chebyshev inequalities. Ann. Math. Statist. 31, 1001-1014.

[82]. Marshall, A. W., and I. Olkin (1960). A bivariate Chebyshev inequality for symmetric convex polygons. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford Univ. Press. 299-308.

[83]. Marshall, A. W., and I. Olkin (1961). Game theoretic proof that Chebyshev inequalities are sharp. Pacific J. of Math. 11, 1421-1429.

[84]. Meidell, B. (1922). Sur une problème du calcul des probabilitiés et les statistiques mathematiques. Comptes Rendus (Paris). 175, 806-808.

[85]. Meidell, B. (1923). Sur la probabilité des erreurs. Comptes Rendus (Paris). 176, 280-282.

[86]. Midzuno, Hiroshi (1950). On certain groups of inequalities. Confidence intervals for the mean. Ann. Inst. Statist. Math. 2, 21-33.

[87]. Mudholkar, G. S. (1965). Some Tchebycheff-type inequalities for matrix-valued random variables. Private manuscript.

[88]. Mudholkar, G. S., and P. S. Rao (1967). Some sharp multivariate Tchebycheff inequalities. Ann. Math. Statist. 38, 393-400.

[89]. Narumi, S. (1923). On further inequalities with possible application to problems in the theory of probability. Biometrika. 15, 245-253.

[90]. Offord, A. C. (1945). An inequality for sums of independent random variables. Proc. London Math. Soc. (2) 48, 467-477.

[91]. Olkin, I., and J. W. Pratt (1958). A multivariate Tchebycheff inequality. Ann. Math. Statist. 29, 226-234.

[92]. Parzen, E. (1962). Stochastic Processes. Holden-Day, San Francisco.

[93]. Pearson, K. (1919). On generalized Tchebycheff theorems in the mathematical theory of statistics. Biometrika. 12, 284-296.

[94]. Peek, R. L. (1933). Some new theorems on limits of variation. Bull. Amer. Math. Soc. 39, 953-959.

*[95]. Pizetti, P. (1892). Fondamenti matematici per la critica dei risultate sperimentali. Ann. Della R. Univ. Di Genova. 184.

* [96]. Robbins, H. (1948). Some remarks on the inequality of Tchebycheff. Courant Anniversary Volume. 345-350. Interscience Publishers, N. Y.

[97]. Romanovski, V. I. (1940). On inductive conclusions in statistics. Comptes Rendus (Doklady) Acad. Sci. URSS. 27, 419-421.

[98]. Royden, H. L. (1953). Bounds on a distribution function when its first n moments are given. Ann. Math. Statist. 24, 361-376.

[99]. Samuels, S. M. (1965). On the number of successes in independent trials. Ann. Math. Statist. 36, 1272-1278.

[100]. Samuels, S. M. (1966). On a Chebyshev-type inequality for sums of independent random variables. Ann. Math. Statist. 37, 248-259.

[101]. Savage, I. R. (1961). Probability inequalities of the Tchebycheff type. J. Res. Natl. Bureau Standards. 65B, 211-222.

[102]. Savage, I. R. (1962). Bibliography of Nonparametric Statistics. Harvard Univ. Press. Cambridge, Mass.

[103]. Selberg, H. L. (1940). Ueber eine Ungleichung der Mathematischen Statistik. Skand. Aktuarietidskr. 23, 114-120.

[104]. Selberg, H. L. (1940). Zwei Ungleichungen zur Ergaenzung des Tchebycheffschen Lemmas. Skand. Aktuarietidskr. 23, 121-125.

[105]. Shohat, J. A. (1929). Inequalities for moments of frequency functions and for various statistical constants. Biometrika. 21, 361-370.

[106]. Shohat, J. A., and J. Tamarkin (1943). The Problem of Moments. Math. Surveys No. 1. Amer. Math. Soc.

[107]. Smith, C. D. (1930). On generalized Tchebycheff inequalities in mathematical statistics. Amer. J. of Math. 52, 109-126.

[108]. Smith, C. D. (1939). On Tchebycheff approximation for decreasing functions. Ann. Math. Statist. 10, 190-192.

[109]. Smith, C. D. (1955). Tchebycheff inequalities as a basis for statistical tests. Math. Mag. 28, 185-195.

[110]. Smith, D. E. (1929). A Source Book in Mathematics. McGraw Hill. New York.

*[111]. Stoker, D. J. (1955). Oor'n Klas van toetsingsgroothede vir die probleem van twee steekproeve. Univ. of Amsterdam (Ph. D. Thesis).

*[112]. Takano, K. (1950). A remark on Berry's paper. Kōkyuroku of Inst. Math. Statist. 5 (Japanese).

[113]. Tchebycheff, P. L. (1867). Des valeurs moyennes. J. Math. (Liouville, 2nd series). 12, 177-184.

[114]. Tucker, H. G. (1967). A Graduate Course in Probability. Academic Press, New York.

[115]. Ulin, B. (1953). An extremal problem in mathematical statistics. Skand. Aktuarietidskr. 36, 158-167.

[116]. Van Dantzig, D. (1951). Une nouvelle généralisation de l'inégalité de Bienayme. Ann. de l'Institut Henri Poincaré. 12, 31-43.

[117]. Von Mises, R. (1938). Sur une inégalité pour les moments d'une distribution quasi-convexe. Bull. Sci. Math. (2). 62, 68-71.

[118]. Von Mises, R. (1939). The limits of a distribution function if two expected values are given. Ann. Math. Statist. 10, 99-104.

[119]. Wald, A. (1938). Generalization of the inequality of Markoff. Ann. Math. Statist. 9, 244-255.

[120]. Wald, A. (1939). Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments. Trans. Amer. Math. Soc. 46, 280-306.

[121]. Walsh, J. E. (1962). Handbook of Nonparametric Statistics. D. Van Nostrand Co. Inc. N. Y.

[122]. Whittle, P. (1958). Continuous generalizations of Chebyshev's inequality. Teor. Veroyatnost. i ee Primenen. 3, 386-394.

[123]. Whittle, P. (1958). A multivariate generalization of Tchebichev's inequality. Quart. J. of Math. (2). 9, 232-240.

*[124]. Winckler, A. (1866). Allgemeine Sätze sur Theorie der unregelmässigen Beobachtungsfehler. Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften-Mathematisch-Naturwissenschaftliche Klasse. 53, 6-41.

[125]. Winsten, C. B. (1946). Inequalities in terms of mean range. Biometrika. 33, 285-295.