Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 104–108

MARKOV'S INEQUALITY FOR RANDOM VARIABLES TAKING VALUES IN A LINEAR TOPOLOGICAL SPACE

BY ALBERT W. MARSHALL
University of British Columbia

Let X be a random variable taking values in the linear topological space $\mathcal{X}$ and let $C \subset \mathcal{X}$ be the closed convex cone which generates the preordering $\precsim$. For an appropriate definition of EX and for $c \in C$, a sharp upper found for $P[X \succsim \varepsilon]$ is obtained in terms of EX. Similarly, a lower bound for $P[X \prec \varepsilon]$ is obtained which is sharp in certain special cases.

1. Introduction. If a random variable X satisfies

(1.1) $$P[X \geq 0] = 1, \quad EX = \mu,$$

and if $\varepsilon > 0$, then according to Markov's inequality,

(1.2) $$P[X \geq \varepsilon] \leq \min\{\mu/\varepsilon, 1\}.$$

Moreover there is a distribution for X satisfying (1.1) for which (1.2) holds with equality. Thus (1.2) is "sharp" in the sense that the bound cannot be improved without information in addition to (1.1) about the distribution of X.

This paper is concerned with inequalities similar to (1.2) which hold for random variables that need not be real-valued, but take values in a real or complex linear topological space $\mathcal{X}$. To obtain such extensions, two preliminaries are required: First, meaning has to be given to inequalities "$a \geq b$" for a,b in $\mathcal{X}$. Second, meaning must be given to the notion of an expectation.

For random variables taking values in the finite dimensional space $\mathcal{R}^n$, the expected value is naturally taken to be the vector of expected values. More generally, the expected value can be defined, e.g., as a Pettis integral: see Perlman (1974) for a similar use of this integral and for the references contained therein. In this paper, it is assumed only that when it exists, $EX = \int X dP \in \mathcal{X}$ and the following properites are satisfied:

(1.3) $$\int(X+Y)dP = \int XdP + \int YdP,$$

(1.4) $$\text{If } A \subset \mathcal{X} \text{ is closed and convex, } P[X \in A] = 1 \text{ implies } \int XdP \in A,$$

(1.5) $$\text{For all events } E \text{ and } c \in \mathcal{X},\ \int_E c\,dP = cP(E).$$

The expression $a \geq b$ can be rewritten as $a - b \in [0,\infty)$ and $a > b$ can be rewritten as $a - b \in (0,\infty)$. Since $[0,\infty)$ is a closed convex cone with interior $(0,\infty)$, it is natural and standard when replacing $(-\infty,\infty)$ by a linear topological space $\mathcal{X}$ to replace $[0,\infty)$ by a closed convex cone $C \subset \mathcal{X}$. For $x,y \in \mathcal{X}$, write

(1.6) $$x \precsim y \quad \text{if } y - x \in C,$$

(1.7) $$x \prec y \quad \text{if } y - x \in C^0,$$

where C^0 is the interior of C. Defined in this way, $\precsim$ is a preordering of $\mathcal{X}$, i.e.,

(1.8) $$x \precsim y \quad \text{for all } x \in \mathcal{X}$$

(1.9) $$x \precsim y \text{ and } y \precsim z \quad \text{implies} \quad x \precsim z,\ x,y,z \in \mathcal{X}.$$

AMS 1980 subject classification. Primary 60E15.
Key words and phrases: Markov's inequality, order preserving functions.

Moreover, $\precsim$ satisfies

(1.10) $x \precsim y$ implies $x+z \precsim y+z$ for all $x,y,z \in \mathcal{X}$,

(1.11) $x \precsim y$ implies $\lambda x \precsim \lambda y$ for all $\lambda \geq 0$, $x,y \in \mathcal{X}$.

Of course (1.2) is equivalent to

(1.2′) $$P[X < \varepsilon] \geq 1 - \min\{\mu/\varepsilon, 1\},$$

but such an equivalence does not hold when $\leq$ is replaced by a partial order $\precsim$. In Section 2 below, upper bounds are obtained for $P[X \succsim \varepsilon]$ and in Section 3, lower bounds for $P[X \prec \varepsilon]$ (upper bounds for $P[X \not\prec \varepsilon]$) are obtained.

For purposes of this paper, certain families $\mathcal{F}$ of real-valued functions defined on $\mathcal{X}$ play a key role. Some conditions that may be imposed on $\mathcal{F}$ are the following:

(1.12) $x \precsim y$ if and only if $f(x) \leq f(y)$ for all $f \in \mathcal{F}$,

(1.12′) $x \prec y$ if and only if $f(x) < f(y)$ for all $f \in \mathcal{F}$,

(1.13) $f \in \mathcal{F}$ implies $f(x) \geq 0$ for all $x \in C$,

(1.14) $f \in \mathcal{F}$ implies $f(ax) \geq af(x)$ for all $a \in [0,1]$, $x \in C$.

In what follows, infima or minima taken over empty sets are to be regarded as ∞.

2. Upper Bounds for P[X$\succsim\varepsilon$].

2.1 Proposition. *Let $C \subset \mathcal{X}$ be a closed convex cone which determines the ordering $\precsim$ via (1.6). Let X be a random variable such that $P[X \in C] = 1$ and $EX = \mu$ exists. Let $\mathcal{F}$ be a set of functions satisfying (1.12), (1.13), (1.14). If $\varepsilon \in C$, then*

(2.1) $$P[X \succsim \varepsilon] \leq \min\{1, \inf_{\{f: f \in \mathcal{F}, f(\varepsilon) > 0\}} f(\mu)/f(\varepsilon)\}$$

Proof. By using (1.3)–(1.6) and (1.10) if follows that

$$\mu = \int X dp = \int_{\{X \succsim \varepsilon\}} X dP + \int_{\{X \not\succsim \varepsilon\}} X dP \succsim \int_{\{X \succsim \varepsilon\}} X dP \succsim \int_{\{X \succsim \varepsilon\}} \varepsilon dP = \varepsilon P[X \succsim \varepsilon].$$

But this implies that

$$f(\mu) \geq f(\varepsilon P[X \succsim \varepsilon]) \geq P[X \succsim \varepsilon] f(\varepsilon) \quad \text{for all } f \in \mathcal{F},$$

i.e.,

$$P[X \succsim \varepsilon] \leq f(\mu)/f(\varepsilon) \quad \text{for all } f \in \mathcal{F} \text{ such that } f(\varepsilon) > 0.$$ □

2.2 Proposition. If (1.14) holds with equality for all $f \in \mathcal{F}$, then for each μ, $\varepsilon \in C$, equality is attainable in (2.1).

Proof. Suppose first that upper bound p of (2.1) is 1 and let Y be a random variable such that $P[Y = \mu] = 1$. By (1.4), $EY = \mu$ so that Y satisfies the conditions of Proposition 2.1. By (1.12) and (1.13) it follows that $\mu \succsim \varepsilon$, that is $P[Y \succsim \varepsilon] = 1$, so equality holds in (2.1).

Next, suppose that $p < 1$ and that

$$P[Y = \varepsilon] = p, \quad P[Y = \alpha] = 1 - p$$

where $\alpha = (\mu - \varepsilon p)/(1 - p)$. Because $p < 1$ it follows from (1.12) and (1.13) that $\mu \not\succsim \varepsilon$ so $\mu - \varepsilon p \not\succsim \varepsilon - \varepsilon p$ or $(1-p)\alpha \not\succsim (1-p)\varepsilon$. Thus $\alpha \not\succsim \varepsilon$, so for this distribution,

$$P[Y \succsim \varepsilon] = P[Y = \varepsilon] = p.$$

To show that $P[Y \in C] = 1$, it is necessary to show only that $\alpha \in C$, since $\varepsilon \in C$ by assumption. Since $f(\mu)/f(\varepsilon) \geq p$ for all $f \in \mathcal{F}$ such that $f(\varepsilon) > 0$, it follows that $f(\mu) \geq pf(\varepsilon) = f(p\varepsilon)$ for all $f \in \mathcal{F}$ hence $\mu \succsim p\varepsilon$, that is, $\alpha \in C$.

From (1.3) and (1.5), it follows that $EY = p\varepsilon + (1-p)(\mu - \varepsilon p)/(1-p) = \mu$. Consequently Y satisfies the condition of Proposition 2.1 and equality is achieved in (2.1). □

2.3 Example. Suppose $\mathcal{X} = \mathcal{R}^n$ and $\mathcal{C} = \{\mathbf{x} = (x_1, \ldots, x_n): x_i \geq 0, i = 1, \ldots, n\} = \mathcal{R}^n_+$ is the nonnegative orthant. Let $\mathcal{F}$ consist of the coordinate functions $f_1, \ldots, f_n$, where $f_i(x) = x_i$. If $\varepsilon \in \mathcal{R}^n_+$ and $\varepsilon \neq 0$, then

$$\inf_{\{f: f\in\mathcal{F}, f(x)>0\}} f(\mu)/f(\varepsilon) = \min_{\{i:\varepsilon_i>0\}} EX_i/\varepsilon_i$$

so that if $\varepsilon_i \geq 0$, $i = 1, \ldots, n$,

$$P[X_i \geq \varepsilon_i, i = 1, \ldots, n] \leq \min_{\{i:\varepsilon_i>0\}} EX_i/\varepsilon_i \tag{2.2}$$

This inequality follows from (4.1) or (7.1) of Marshall and Olkin (1960). It is also equivalent to Corollary 2.1 of Jensen and Foutz (1981).

2.4 Example. Let $\mathcal{X}$ be the linear space of $n \times n$ Hermitian matrices and let $\mathcal{C}$ be the convex cone of positive semi-definite matrices. Take $\mathcal{F}$ to consist of functions of the form $f_\mathbf{a}$ where $\mathbf{a}$ is a unit vector ($\mathbf{aa}^* = 1$) of a complex numbers and $f_\mathbf{a}(\mathbf{A}) = \mathbf{aAa}^*$. Suppose that $\mathbf{C}$ is positive definite. If the random matrix $\mathbf{X}$ is positive definite with probability one, $\inf_{f\in\mathcal{F}} f(E\mathbf{X})/f(\mathbf{C}) = \inf_\mathbf{a} \mathbf{a}E\mathbf{X}\mathbf{a}^*/\mathbf{aCa}^* = \min_{\mathbf{bb}^*=1} \mathbf{bC}^{-1/2}E\mathbf{XC}^{-1/2}\mathbf{b}^* = \lambda_n[\mathbf{C}^{-1/2}(E\mathbf{X})\mathbf{C}^{-1/2}]$, the minimum characteristic root of $\mathbf{C}^{-1/2}(E\mathbf{X})\mathbf{C}^{-1/2}$. Thus

$$P[\mathbf{X} \succsim \mathbf{C}] \leq \lambda_n[\mathbf{C}^{-1/2}(E\mathbf{X})\mathbf{C}^{-1/2}]. \tag{2.3}$$

This result is given in Corollary 3.3 of Jensen and Foutz (1981).

2.5 Example. Let $\mathcal{X} = \mathcal{R}^n$ and suppose that $\precsim_w$ is the ordering of weak submajorization (see Marshall and Olkin, 1979, p. 10). Restricted to $\mathcal{D} = \{\mathbf{x}: x_1 \geq \ldots \geq x_n\}$, this ordering is generated by the convex cone $\mathcal{C} = \{\mathbf{x}: \Sigma_{i=1}^k x_i \geq 0, k = 1, \ldots, n\}$. Replace the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ by $\mathbf{X}_\downarrow = (X_{[1]}, \ldots, X_{[n]})$ where $X_{[1]} \geq \ldots \geq X_{[n]}$ are obtained by ordering $X_1, \ldots, X_n$. Let $\mathcal{F}$ consist of the functions $f_k(\mathbf{x}) = \Sigma_{i=1}^k x_{[i]}$, $k = 1, \ldots, n$. If $\varepsilon \in \mathcal{C}$,

$$\min_{\{f\in\mathcal{F}, f(\varepsilon)>0\}} f(E\mathbf{X})/f(\varepsilon) = \min_{\{k:\Sigma_{i=1}^k \varepsilon_{[i]}>0\}} = \Sigma_{i=1}^k EX_{[i]}/\Sigma_{i=1}^k \varepsilon_{[i]},$$

so that

$$P[\mathbf{X} \succsim_w \varepsilon] = P[\mathbf{X}_\downarrow \succsim_w \varepsilon] \leq \min_{\{k:\Sigma_{i=1}^k \varepsilon_{[i]}>0\}} = \Sigma_{i=1}^k EX_{[i]}/\Sigma_{i=1}^k \varepsilon_{[i]}. \tag{2.4}$$

The bound of this inequality is in terms of $E\mathbf{X}_\downarrow$, not of $E\mathbf{X}$. Because $E\mathbf{X}$ is majorized by $E\mathbf{X}_\downarrow$ (Marshall and Olkin (1979), p. 348), it is not possible to replace $E(X_{[i]})$ by the i-th largest component of $E\mathbf{X}$ in the above bound.

3. Upper Bounds for $P\{X \not\prec \varepsilon\}$. In general, $X \succsim \varepsilon$ implies $X \not\prec \varepsilon$ but not conversely, so it is to be expected that a sharp upper bound for $P[X \not\prec \varepsilon]$ will be larger than the corresponding bound for $P[X \succsim \varepsilon]$ found in Section 2.

The following proposition is less satisfactory than Proposition 2.1 because it is little more than Markov's inequality (1.2) and requires additional steps to yield a bound in terms of EX.

3.1 Proposition. Let $\mathcal{C} \subset \mathcal{X}$ be a closed convex cone and let X be a random variable such that $P[X \in \mathcal{C}] = 1$ and that $EX = \mu$ exists. Let $\mathcal{F}$ be a set of functions satisfying (1.12′) and (1.13). If $\varepsilon \in \mathcal{C}^0$ then

$$P[X \not\prec \varepsilon] \leq \min\{1, E \sup_{f\in\mathcal{F}} f(X)/f(\varepsilon)\}. \tag{3.1}$$

Remark. Because $\varepsilon \succ 0$, it follows from (1.12′) and (1.13) that

$$f(\varepsilon) > f(0) \geq 0 \quad \text{for all } f \in \mathcal{F}.$$

Proof. From (1.12′), (1.13), and Markov's inequality (1.2) it follows that

$$P[X \not\prec \varepsilon] = P[f(X) \geq f(\varepsilon) \text{ for some } f \in \mathcal{F}] \leq P[\sup_{f \in \mathcal{F}} f(X)/f(\varepsilon) \geq 1]$$
$$\leq E \sup_{f \in \mathcal{F}} f(X)/f(\varepsilon). \qquad \square$$

The following examples show that (3.1) sometimes leads to sharp bounds in terms of EX.

3.2 Example. Suppose $\mathcal{X} = \mathcal{R}^n$, $\mathcal{C} = \{\mathbf{x} = (x_1, \ldots, x_n): x_i \geq 0, i = 1, \ldots, n\} = \mathcal{R}^n_+$. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ where each $\varepsilon_i > 0$ and let $\mathcal{F}$ consist of the coordinate functions $f_1, \ldots, f_n$ where $f_i(\mathbf{x}) = x_i$. If $\mathbf{X}$ is an $\mathcal{X}$-valued random variable such that $E\mathbf{X} = \mu$ exists, then

$$P[X_i \geq \varepsilon_i \text{ for some } i = 1, \ldots, n] \leq \min\{1, \Sigma_{i=1}^n \mu_i/\varepsilon_i\}. \tag{3.2}$$

Proof. Since $\sup_{f \in \mathcal{F}} f(x)/f(\varepsilon) \leq \Sigma_{f \in \mathcal{F}} f(x)/f(\varepsilon)$ and since $Ef(X) = f(EX)$ for all $f \in \mathcal{F}$, (3.2) follows from (3.1). $\square$

In spite of its apparent crudeness, inequality (3.2) is sharp. To see this, suppose first that the upper bound is less than one and let $\mathbf{e}_i$ be the vector with i-th coordinate 1 and all other coordinates 0. Let $\mathbf{Y}$ be a random vector such that

$$P[\mathbf{Y} = \varepsilon_i \mathbf{e}_i] = \mu_i/\varepsilon_i, \; i = 1, \ldots, n$$
$$P[\mathbf{Y} = \mathbf{0}] = 1 - \Sigma \mu_i/\varepsilon_i.$$

Then $E\mathbf{Y} = \mu$ and equality is attained in (3.2).

Next, suppose the upper bound of (3.2) is one and let $s = \Sigma_{i=1}^n \mu_i/\varepsilon_i$. Let $\mathbf{Y}$ be a random vector such that

$$P[\mathbf{Y} = s\varepsilon_i \mathbf{e}_i] = \mu_i/s\varepsilon_i.$$

Since $s \geq 1$, $P[Y_i \geq \varepsilon_i$ *for some* $i = 1, \ldots, n] = 1$.

3.3 Example. Suppose $\mathcal{X}$ consists of $n \times n$ Hermitian matrices and $\mathcal{C}$ consists of the positive semi-definite Hermitian matrices. If $P[\mathbf{X} \in \mathcal{C}] = 1$, $E\mathbf{X} = \mu$ exists and $\mathbf{C}$ is positive definite, then

$$P[\mathbf{X} \not\prec \mathbf{C}] \leq \min\{1, \operatorname{tr} \mathbf{C}^{-1/2} \mu \mathbf{C}^{-1/2}\}. \tag{3.3}$$

To obtain (3.3) from (3.1), take $\mathcal{F}$ as in Example 2.4. Denote the largest eigenvalue of an Hermitian matrix $\mathbf{H}$ by $\lambda_1(\mathbf{H})$. Then

$$E \sup_{\mathbf{a}} \mathbf{aXa}^*/\mathbf{aCa}^* = E \sup_{\{\mathbf{a}:\mathbf{aa}^*=1\}} \mathbf{a}\mathbf{C}^{-1/2}\mathbf{X}C^{-1/2}\mathbf{a}^* = E\lambda_1(\mathbf{C}^{-1/2}\mathbf{X}C^{-1/2})$$
$$\leq E \operatorname{tr} \mathbf{C}^{-1/2}\mathbf{X}\mathbf{C}^{-1/2} = \operatorname{tr} \mathbf{C}^{-1/2}(E\mathbf{X})\mathbf{C}^{-1/2}.$$

Thus (3.3) follows from (3.1).

To see that (3.3) is sharp, suppose without loss of generality that $\mathbf{C} = \mathbf{I}$; otherwise replace $\mathbf{X}$ by $\mathbf{C}^{-1/2}\mathbf{X}\mathbf{C}^{-1/2}$. Write μ in the form $\mu = \Gamma\mathbf{D}\Gamma^*$ where $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_n)$ is diagonal and Γ is unitary. Suppose the bound is less than one and let $\mathbf{E}_i = \operatorname{diag} \mathbf{e}_i$ where $\mathbf{e}_i$ is defined in 3.2. If

$$P[\mathbf{Y} = \Gamma\mathbf{E}_i\Gamma^*] = d_i, \quad i = 1, \ldots, n$$
$$P[\mathbf{Y} = \mathbf{0}] = 1 - \Sigma_1^n d_i,$$

then $E\mathbf{Y} = \Sigma\, d_i \Gamma\mathbf{E}_i\Gamma^* = \Gamma(\Sigma d_i \mathbf{E}_i)\Gamma^* = \Gamma\mathbf{D}\Gamma^* = \mu$. Moreover $P[\mathbf{X} \prec \mathbf{I}] = P[\mathbf{X} = \mathbf{0}] = 1 - \operatorname{tr} \mu$ so equality holds in (3.3).

In case the bound of (3.3) is one, the above example can be modified to show that equality is attainable using ideas similar to those used for Example 3.2.

3.4 Example. Let $\mathcal{X} = \mathcal{R}^n$ and supposed that $\precsim_w$ is the ordering of weak submajorization, as in Example 2.5. With $\mathcal{C}$ and $\mathcal{T}$ as in Example 2.5, it follows from (3.1) that

$$P[\mathbf{X} \precsim_w \varepsilon] \leq \Sigma_{k=1}^n [\Sigma_{j=1}^k \mu_{[j]} / \Sigma_{j=1}^n \varepsilon_{[j]}]. \tag{3.4}$$

Acknowledgement. The author would like to thank the editors for their comments on the earlier version of this paper.

REFERENCES

JENSEN, D. R. and FOUTZ, R. V. (1981). Markov Inequalities on partially ordered spaces. *J. Multivariate Anal.* *11* 250–259.

MARSHALL, A. W. and OLKIN, I. (1960). Multivariate Chebyshev Inequalities. *Ann. Math. Statist.* *31* 1001–1014.

MARSHALL, A. W. OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press.

PERLMAN, M. D. (1974). Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *J. Multivariate Anal.* *4* 52–65.