Note on Stochastic Inequalities

Jong-Woo Jeon*

ABSTRACT

In this note, we establish a result which characterizes a partial ordering of probability distributions on a partially ordered space by that of conditional distributions. This result is then reduced to prove the conjecture made by Nevius, Proschan and Sethuraman (1977).

1. Introduction

Let $\boldsymbol{x}$ be a separable complete metric space with closed partial ordering $\leq$. Let $F_1^+(F_1^-)$ be the class of all bounded Borel measurable monotone non-decreasing (non-increasing) functions on $\boldsymbol{x}$. Let $S^+(S^-)$ be the class of all non-decreasing (non-increasing) Borel sets, i.e., sets of which indicator functions are non-decreasing (non-increasing). Let P and Q be probability measures on $\boldsymbol{x}$.

We say that P is stochastically smaller than Q, denote this by $P\alpha Q$, if and only if $\int f\ P(dx) \leq \int f Q(dx)$ for all $f \varepsilon F^+$. It can be readily shown by a simple approximation that this is equivalent to the requirement that $P(A) \leq Q(A)$ for all $A \varepsilon S^+$.

Let X and Y be random elements each taking values in $\boldsymbol{x}$ and have distributions P and Q, respectively. We then say that X is stochastically smaller than Y, denote this by $X\alpha Y$, if and only if $P\alpha Q$.

*University of Maryland

The subject of stochastic comparison has well been developed in the past two decades and its usefulness has been demonstrated in many of probability and statistics. Barlow and Proschan (1975) has made its excellent applications to reliability theory.

2. Main result

For $A \subset x$, write

$$P(A, t) = P_r\{X \varepsilon A \mid h(X) = t\} \text{ and}$$
$$Q(A, t) = P_r\{Y \varepsilon A \mid h(Y) = t\}$$

where $h \varepsilon F^+ \cap F^-$.

Theorem. $X \alpha Y$, or equivalently $P \alpha Q$, if and only if $P(\cdot, t) \alpha Q(\cdot, t)$ for all $t \varepsilon T$ where $P_r\{h(X) \varepsilon T\} = 1$.

Proof. Let $P(\cdot, t) \alpha Q(\cdot, t) \forall t \varepsilon T$, $P_r\{h(X) \varepsilon T\} = 1$.

Then for $f \varepsilon F^+$,

$$\begin{aligned} \int_x f\, P(dx) &= \int_{R_1} \int_x f P(dx,\ t) P_r\{h(X) \varepsilon dt\} \\ &\leq \int_{R_1} \int_x f Q(dx,\ t) P_r\{h(Y) \varepsilon dt\} \\ &= \int_x f\, Q(dx). \end{aligned}$$

Thus, sufficiency is proved.

Now for necessity, suppose $X \alpha Y$. It follows from Strassen (1965) or Kamae, Krengel and O'Brien (1977) that there exist X' and Y' such that

$$P_r\{X' \leq Y'\} = 1$$

and

$$L(X') = L(X), \quad L(Y') = L(Y)$$

where L(·) denotes distribution. Moreover, since $h \varepsilon F^+ \cap F^-$,

$$P_r\{h(X') = h(Y')\} = 1$$

and

$$L(h(X')) = L(h(X)), \quad L(h(Y')) = L(h(Y)).$$

Let $I\{\cdot, C\}$ denote indicator function of C. Denote also

$$P'(A, t) = P_r\{X' \varepsilon A \mid h(X') = t\}$$

$$Q'(A, t)=P_r\{Y'\varepsilon A|h(Y')=t\}$$

and

$$K(A, y)=P_r\{X' \varepsilon A| Y'=y\}.$$

Then for any $A\subset \mathfrak{X}$ and $C\subset R_1$,

$$\int_C P'(A,t)P_r\{h(X') \varepsilon dt\},$$
$$=P_r\{X'\varepsilon A, h(X') \varepsilon C\}$$
$$=\int_{\mathfrak{X}} P_r\{X' \varepsilon A, h(X') \varepsilon C| Y'=y\} P(dy)$$
$$=\int_{\mathfrak{X}} P_r\{X'\varepsilon A| Y'=y\} I\{h(Y'), C\} P(dy)$$
$$=\int_C\int_{\mathfrak{X}} P_r\{X'\varepsilon A| Y'=y\} P_r\{Y'\varepsilon dy|h(Y')=t\} P_r\{h(Y')\varepsilon dt\}$$
$$=\int_C\int_{\mathfrak{X}} K(A,y)Q'(dy, t)P_r\{h(X')\varepsilon dt\}.$$

Thus, $P'(A, t)=\int_{\mathfrak{X}} K(A, y)Q'(dy, t)$, for any $\pm$ in T.

Now, let $f \varepsilon F^+$ and $t \varepsilon T$ be fixed. Since $K(A_y, y)=1$ where $A_y=\{x : x\leq y\}$,

$$\int_{\mathfrak{X}} f(x)P'(dx, t)=\int_{\mathfrak{X}}\int_{\mathfrak{X}} f(x)K(dx, y)Q'(dy, t)$$
$$=\int_{\mathfrak{X}}\int_{A_y} f(x)K(dx, y)Q'(dy, t)$$
$$\leq\int_{\mathfrak{X}} f(y)Q'(dy,t).$$

Therefore $P'(.,t)\alpha Q'(.,t)$, $\forall\, t \varepsilon T$, $P_r\{h(X')\varepsilon t\}=1$ which in turn implies that

$$P(.,t)\alpha Q(.,t),\ \forall\, t \varepsilon T$$

where $P_r\{h(X) \varepsilon T\}=1$. Q.E.D.

Remark. When $\mathfrak{X}=R_1$ and $\leq$ is the usual linear order in R_1, only functions in both F^+ and F^- are constant functions. In this case, the above theorem is trivially true. A non trivial application is shown in Section 4.

3. Majorization and Stochastic majorization

In this section, we briefly give definition of majorization for the sake of completeness.

Given a vector $x=(x_1, \ldots, x_n)$ in R_n, let $x_{(1)}\geq x_{(2)}\geq\ldots\geq x_{(n)}$, denote a non-increasing rearrangement of $x_1, x_2, \ldots, x_n$ · A vector x is said to be majorized by a vector y, denote this by $x \overset{m}{\leq} y$, if

$$\sum_{i=1}^{j} x_{(i)}\leq\sum_{i=1}^{j} y_{(i)},\quad j=1, 2, \ldots, n-1$$

and

$$\sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)}.$$

Note that whenever $(\pi_1, \ldots, \pi_n)$ is a permutation of $(1, \ldots, n)$ and $y = (x_{\pi_1}, \ldots, x_{\pi_n})$, we have $x \overset{m}{\leq} y$ and $y \overset{m}{\leq} x$. Thus majorization constitutes a partial ordering on R_n up to permutation of components. A function f which satisfies the property that $f(x) \leq (\geq) f(y)$ whenever $x \overset{m}{\leq} y$ is called a Schur-convex (Schur-concave) function. Define $f(x) = -\sum_{i=1}^{n} x_i \log x_i$ where $\sum_{i=1}^{n} x_i = 1$, $x_i \geq 0$, $i = 1, 2, \ldots, n$. The function f just defined is known as Shannon entropy function in information theory and is Schur-convex (see Hardy, Littlewood and Polya (1952)). It follows immediately that f takes its maximum when $x_1 = x_2 = \ldots = x_n = \frac{1}{n}$.

Let $X = (X_1, \ldots, X_n)$ be a random vector on R_n. A comparison by majorization among random vector can be defined as before. We say that X is stochastically majorized by Y, denote this by $X \overset{m}{\alpha} Y$, if and only if $\int_{R_n} f(x) P(dx) \leq \int_{R_n} f(y) Q(dy)$ for all bounded Schur-convex function f.

4. Conjecture of Nevius, Proschan and Sethuraman

In their Theorem 2.9, Nevius, Proschan and Sethuraman (1977) proved the following; Let X and Y be random vectors. Then $X \overset{m}{\alpha} Y$ implies that for each bounded Schur-convex function f,

$$\int_{R_n} f(x)\, P_r\{X \varepsilon dx \mid \sum_{i=1}^{n} X_i = t\} \leq \int_{R_n} f(y) P_r\{Y \varepsilon dy \mid \sum_{i=1}^{n} Y_i = t\}$$

for all $t \varepsilon T_f$ where $P_r\{\sum_{i=1}^{n} X_i \varepsilon T_f\} = 1$. The conjecture they made is that the set T_f, chosen above to be dependent on the particular Schur-convex function f, can be free of f. This can be established by identifying $\sum_{i=1}^{n} X_i$ by $h(X)$ in our theorem.

5. Acknowledgments

I wish to thank professor J. Sethuraman for suggesting this problem and helpful comments.

REFERENCES

[1] Barlow, R.E. and Proschan F. (1975). Statistical Theory of Reliability and Life Testing: Probability Models. Holt, Rinehart and Winston, New York.

[2] Hardy, G.H., Littlewood, J.E. and Polya, G. (1952). Inequalities. Cambridge University Press.

[3] Nevius, S.E., Proschan, F. and Sethuraman, J. (1977). Schur functions in statistics. II. Stochastic majorization. Ann. Statist. 5 pp 263-273.

[4] Kamae, T., Krengel, U. and O'Brien, G.L. (1977). Stochastic inequalities on partially ordered space. Ann. Probability 5 pp 899-912.

[5] Strassen, V. (1965). The existence of probability measures with given marginals. Ann. Math. Statist. 36 pp 423-439.