Seite 149 - 155 in:

Horst Rinne, Bernhard Rüger
Heinrich Strecker
(Hrsg.)

# Grundlagen der Statistik und ihre Anwendungen

Festschrift
für
Kurt Weichselberger

Mit 24 Abbildungen

Physica-Verlag
Ein Unternehmen
des Springer-Verlags

# Testing Whether Two Distributions are Stochastically Ordered or Not

Von KARL MOSLER*, Hamburg

**Summary:** The paper presents nonparametric tests on different pairs of hypotheses which involve that two probability distributions are stochastically ordered. Let $F$ and $G$ be two probability distribution functions on the reals. First, we consider the null hypothesis that $F$ is stochastically larger than $G$ holds against the alternative that this is not the case. Second, the null that either $F$ is stochastically larger than $G$ or $G$ is stochastically larger than $F$ holds is tested against the alternative that this is not the case. Third, homogeneity, i.e. $F = G$, is taken as the null hypothesis, and $F$ stochastically larger than $G$ as the alternative. The tests employ Smirnov's supremum statistic in various ways. A new statistic is introduced which is the minimum of the two one-sided Smirnov statistics. Related nonparametric test statistics are shortly surveyed. Applications include location tests with ordinal data, expected utility decisions under risk, comparisons of system reliability and specification tests for duration models.

## 1. Introduction

Let $F$ and $G$ be probability distribution functions on $\mathbb{R}$. $F$ is *stochastically larger* than $G$, $F \geq_{st} G$, if

$$F(x) \leq G(x) \text{ holds for every } x \in \mathbb{R}.$$

$\geq_{st}$ is a partial order on the set of all probability distribution functions. It is called *stochastic order* or, to distinguish it from other orderings of probability distributions, *usual stochastic order*. Another name is *first degree stochastic dominance*. If $F \geq_{st} G$ and two random variables $X$ and $Y$ are distributed according to $F$ and $G$ respectively, we also say that $X$ is stochastically larger than $Y$, $X \geq_{st} Y$. This means that $X$ exceeds any given number with a higher probability than $Y$ does.

In many applications it is necessary to decide whether two given probability distributions are stochastically ordered or not. This paper presents nonparametric statistical procedures to do this on a sample base.

The following properties are immediately seen from the definition of stochastic order: If expectations exist, $X \geq_{st} Y$ implies $EX \geq EY$, but the reverse is not true. $X \geq_{st} Y$ holds if and only if

$$\varphi(X) \geq_{st} \varphi(Y) \text{ for all non-decreasing } \varphi : \mathbb{R} \to \mathbb{R}, \qquad (1)$$

moreover, if and only if

$$E\varphi(X) \geq E\varphi(Y) \text{ for all non-decreasing } \varphi : \mathbb{R} \to \mathbb{R} \qquad (2)$$

holds as far as the expectations exist.

* Prof. Dr. Karl Mosler, Universität der Bundeswehr, Professur für Statistik und Quantitative Ökonomik II, D-22039 Hamburg

The pair $(F, G)$ is called a *shift* if $F(x) = G(x - a)$ holds for all $x$, $a$ being a non-negative number. When $(F, G)$ is a shift, this implies $F \geq_{st} G$. In other words, for a given random variable $X$ and $a \geq 0$ we have $X + a \geq_{st} X$. $(F, G)$ with $F \geq_{st} G$ is also named a *slippage*. Thus, every shift is a slippage.

We give four examples of applications:

1. Two-sample location tests with ordinal data - $X$ and $Y$ may represent the outcomes of two experiments. When the data are measured on an ordinal scale, slippage is the natural notion for a shift in location. Thus testing for stochastic order means testing for a shift in location with ordinal data. See RAHLFS/ZIMMERMANN (1993).
2. Decision under risk - Here $X$ and $Y$ denote the results of two risky prospects and $\varphi$ is a von Neumann-Morgenstern utility function. $X \geq_{st} Y$ implies by (2) that every expected utility maximizer having a non-decreasing utility function will prefer the prospect yielding $X$ over the prospect yielding $Y$.
3. Reliability of a coherent system - Let $X$ and $Y$ be alternative lifelengths of a given component which result in system lifelengths $\varphi(X)$ and $\varphi(Y)$, respectively. If the system is coherent, $\varphi$ is non-decreasing. Then $X \geq_{st} Y$ implies by (1) that $\varphi(X) \geq_{st} \varphi(Y)$ and by (2) that $E[\varphi(X)] \geq E[\varphi(Y)]$, i.e. the expected system lifelength is not less with $X$ than with $Y$. See BARLOW/PROSCHAN (1981).
4. Specification tests for duration models - Let a life distribution depend on an observable and an unobservable covariate and assume a proportional hazards model. If $F$ and $G$ are the conditional distributions of lifelengths at two different values of the observable covariate, then $F$ and $G$ must be stochastically ordered in one or the other direction. This gives rise to a specification test for the proportional hazards model with unobservable covariates. See MOSLER/PHILIPSON (1994).

Many other applications of stochastic order are found in queueing, scheduling, insurance, finance and various other fields. See e.g. STOYAN (1983) and the bibliography by MOSLER/SCARSINI (1993).

In Section 2 of this paper, we present nonparametric statistical procedures for testing different pairs of hypotheses which involve stochastic order as the null and/or as the alternative hypothesis. The tests proposed employ Smirnov's supremum statistic in different ways. A new test statistic is introduced which is the minimum of the two one-sided Smirnov statistics. These tests assume arbitrary probability distributions. The special case of a shift is discussed. The final Section 3 offers remarks on several related problems - other tests for homogeneity against ordered alternatives, tests for stronger orderings, tests with censored data, the case of parametric distributions - and provides references on them.

Our approach differs from most of the operations research and finance literature where so called tests on first and higher degree stochastic dominance have been proposed on a purely deterministic basis; they consist in pointwise comparisons of the empirical distribution functions (or of their integrals), see e.g. BAWA/LINDENBERG/RAFSKY (1979), KROLL/LEVY (1980), ABOUDI/THON (1994). Sample based stochastic comparisons of financial prospects have been considered in BAWA/BROWN/KLEIN (1979) and NELSON/POPE (1991).

## 2. Testing for stochastic order

Let $F$ and $G$ be two given distribution functions, and let $X_1, \ldots X_m, Y_1, \ldots Y_n$ be an independent sample where the $X_i$ are distributed according to $F$, $X_i \sim F$, and the $Y_j$ according to $G$, $Y_j \sim G$. The empirical distribution functions will be denoted by $F_m$ and $G_n$, respectively.

### *1. Stochastic order against non-order*

We start with testing the null hypothesis that $F$ is stochastically larger than $G$ against the alternative that this is not true,

$$H_0: \; F \geq_{st} G \text{ against } H_1: \text{ not } H_0 \,. \tag{A}$$

For this we employ the supremum statistic

$$D^+_{m,n} = \sup_x \left[F_m(x) - G_n(x)\right]$$

and reject $H_0$ when $D^+_{m,n}$ comes out to be too large. $D^+_{m,n}$ is the classic *Smirnov statistic* and the test is the usual Smirnov test, which is also called the one-sided Kolmogorov-Smirnov test. Let $K^+_{m,n}$ denote the distribution function of $D^+_{m,n}$ under $F = G$, and $t_{1-\alpha}$ the $(1-\alpha)$-quantile of $K^+_{m,n}$. $H_0$ is rejected at a level of significance $\alpha$ if $D^+_{m,n} \geq t_{1-\alpha}$. The test is based on the well-known fact that

$$\begin{aligned} &P\left[D^+_{m,n} \geq t | F(x) \leq G(x) \; \forall x\right] \\ &\quad \leq P\left[D^+_{m,n} \geq t | F(x) = G(x) \; \forall x\right] = 1 - K^+_{m,n}(t). \end{aligned} \tag{3}$$

Extensive tables of $K^+_{m,n}$ are found in HARTER/OWEN (1970). Obviously, the Smirnov test is consistent on the whole alternative $H_1$.

When $F$ and $G$ are interchanged in (A) we have a similar problem which is tested by the Smirnov statistic

$$D^-_{m,n} = \sup_x \left[G_n(x) - F_m(x)\right].$$

The distribution function of $D^-_{m,n}$ under $F = G$ is also equal to $K^+_{n,m}$. See DURBIN (1973) for the distribution theory of $D^+_{m,n}$ and $D^-_{m,n}$.

### *2. Stochastic order in either direction*

Motivated by our fourth application – specification tests in duration analysis – we consider the test problem where the hypothesis that either $F$ is stochastically larger than $G$ or $G$ is stochastically larger than $F$ is tested against the alternative that neither is the case,

$$H_0: F \geq_{st} G \text{ or } G \geq_{st} F \text{ against } H_1: \text{ not } H_0. \tag{B}$$

As a test statistic we employ the minimum of the two Smirnov statistics,

$$M_{m,n} = \min\left\{D^+_{m,n}, D^-_{m,n}\right\}.$$

To our knowledge, this test statistic has not appeared in the literature before.

*Proposition 1* There holds
$$\sup_{H_0} P\left[M_{m,n} \geq t\right] \leq \max\left\{1 - K^+_{m,n}(t), 1 - K^+_{n,m}(t)\right\}.$$

*Proof of Proposition 1* We have $H_0 = H_0^+ \cup H_0^-$ where $H_0^+ : F \geq_{st} G$, $H_0^- : G \geq_{st} F$. From (3) follows that under $H_0^+$ the probability of rejection is maximum when $F$ equals $G$,

$$\sup_{H_0^+} P\left[D^+_{m,n} \geq t\right] = P\left[D^+_{m,n} \geq t | F = G\right]. \tag{4}$$

Similarly to (3), we get $P\left[D^-_{m,n} \geq t | G \geq_{st} F\right] \leq P\left[D^-_{m,n} \geq t | F = G\right]$, and therefore

$$\sup_{H_0^-} P\left[D^-_{m,n} \geq t\right] = P\left[D^-_{m,n} \geq t | F = G\right].$$

From $P\left[M_{m,n} \geq t\right] \leq P\left[D^+_{m,n} \geq t\right]$ and $P\left[M_{m,n} \geq t\right] \leq P\left[D^-_{m,n} \geq t\right]$ we conclude

$$\begin{aligned}
\sup_{H_0} P\left[M_{m,n} \geq t\right] &\leq \max\left\{\sup_{H_0^+} P\left[D^+_{m,n} \geq t\right], \sup_{H_0^-} P\left[D^-_{m,n} \geq t\right]\right\} \\
&= \max\left\{P\left[D^+_{m,n} \geq t | F = G\right], P\left[D^-_{m,n} \geq t | F = G\right]\right\} \\
&= \max\left\{1 - K^+_{m,n}(t), 1 - K^+_{n,m}(t)\right\}. \qquad \diamond
\end{aligned}$$

Thus, the critical region $[t_{1-\alpha}, \infty[$ of the classic Smirnov test may be used as a conservative one in the above test. It is also immediately seen that this test is consistent against every alternative contained in $H_1$.

## 3. *Homogeneity against stochastic order*

In this section we envisage the test problem where stochastic order is the alternative which has to be supported by the data, and homogeneity is the null. Consider the following pair of hypotheses,

$$H_0: F = G \text{ against } H_1: G \geq_{st} F \text{ and not } F = G. \tag{C}$$

We reject $H_0$ if

$$\inf\left[F_m(x) - G_n(x)\right] = -D^-_{m,n}$$

is too large, i.e. if $D^-_{m,n}$ is too small. The critical region for $D^-_{m,n}$ is $]-\infty, t_\alpha]$ where $t_\alpha$ is the $\alpha$-quantile of $K^+_{n,m}$.

*4. Stochastic order against the opposite order*

The previous test can be extended to the case where stochastic order in one direction is tested against stochastic order in the opposite direction,

$$H_0: \ F \geq_{st} G \text{ against } H_1: \ G \geq_{st} F \text{ and not } F = G. \qquad \text{(D)}$$

Analogously to (3) we conclude that

$$P[-D^-_{m,n} \geq -t|F \geq_{st} G] \leq P[-D^-_{m,n} \geq -t|F = G],$$

$$\sup_{H_0} P[D^-_{m,n} \leq t] \leq K^+_{n,m}(t). \qquad (5)$$

Therefore the previous test for (C) can be used as well for the pair of hypotheses (D). It is obvious that in both cases the test is consistent on the whole alternative.

*5. Shift alternatives*

It is interesting to see what our tests do when $(F, G)$ or $(G, F)$ is a shift, i.e. $F(x) = G(x - a)$. Then $F \geq_{st} G$ is equivalent to $a \geq 0$, $G \geq_{st} F$ to $a \leq 0$, and $F = G$ to $a = 0$. If $D^+_{m,n}$ is sufficiently large the above test on (A) confirms that $a < 0$. If both $D^+_{m,n}$ and $D^-_{m,n}$ are sufficiently large this is in favour of $a = 0$ by the test on (B). If $D^-_{m,n}$ is small enough the tests on (C) and (D) confirm $a < 0$ over $a \geq 0$.

Thus, for testing the shift alternative,

$$F(x) = G(x - a), \quad H_0: a \geq 0 \text{ against } H_1: a < 0, \qquad \text{(S)}$$

we may employ either $D^+_{m,n}$ or $-D^-_{m,n}$ as a test statistic and reject the null if the statistic is large. Also the sum of these two,

$$D^+_{m,n} - D^-_{m,n} = \sup_x [F_m(x) - G_n(x)] + \inf_x [F_m(x) - G_n(x)],$$

is a proper test statistic for (S). It has been proposed and investigated by ANDREAS WEICHSELBERGER (1993). As WEICHSELBERGER has shown, this test is particularly useful when the kurtosis is not too large.

## 3. Remarks

We conclude the paper with a few remarks on related problems.

**Other tests for homogeneity against stochastic order** – For the pair of hypotheses (C), there exist many other tests which we cannot survey here. We only mention the classic test by Wilcoxon–Mann–Whitney, the Savage–Mantel log-rank test and the tests by LEE/WOLFE (1976) and FRANCK (1984), which are based on likelihood ratio estimators. Comparative surveys are given by CHIKKAGOUDAR/SHUSTER (1974), RAHLFS/ZIMMERMANN (1993) and others.

**Censorized data** – When the data are censorized, the above Smirnov procedures can be applied with modifications. For type II censoring, modified quantiles of the Smirnov statistic can be computed in a way similar to TSAO (1954) and CONOVER (1967). For random right censoring, we refer to FLEMING/HARRINGTON (1981) and SCHUMACHER (1984), where also the log-rank test and the Gehan modification of the Wilcoxon test are dicussed. See also CHAKRABORTY/SEN (1992) and LIU et al. (1993).

**Tests for stronger orderings** – There exist tests on orderings of distributions which are stronger than usual stochastic order. $F$ is called *larger than* $G$ *in failure rate*, $F \geq_{FR} G$, (resp. *in likelihood ratio*, $F \geq_{LR} G$,) if $f(x)/(1-F(x)) \leq g(x)/(1-G(x))$ for all $x$ (resp. $f(x)/f(y) \leq g(x)/g(y)$ for all $x < y$), where $f$ and $g$ denote the densities. There holds $F \geq_{LR} G \Rightarrow F \geq_{FR} G \Rightarrow F \geq_{st} G$. Tests for homogeneity against $F \geq_{FR} G$ are found in BAGAI/KOCHAR (1986), tests for homogeneity against $F \geq_{LR} G$ in DYKSTRA et al. (1991) and in the references quoted there.

**Parametric distributions** – When $F$ and $G$ belong to the same parametric class of distributions, stochastic order can be expressed through their parameters; see MOSLER (1982, p. 122 ff). In particular, if the class is a distribution type then there exist constants $a$ and $b \in \mathbb{R}$, $b > 0$, such that $F(x) = G((x-a)/b)$. Then $b = 1$, $a \geq 0$ implies $F \geq_{st} G$, and the reverse is true if $F$ has unbounded support. In this latter case, testing for stochastic order is equivalent to testing whether $a \geq 0$ and $b = 1$ holds, i.e. whether $(F, G)$ is a shift. For the case of $F$ having bounded support, see WILCOX (1990). Tests for multinomial distributions are given in ROBERTSON et al.(1988).

**Lehmann hypotheses** – A special case of stochastic order is the Lehmann hypothesis, $F = G^k$, $k > 0$. Under this hypothesis, dependent on $k$, the distribution of $D^+_{m,n}$ has been derived by STECK (1969).

## References

ABOUDI, R. / THON, D. (1994): Efficient algorithms for stochastic dominance tests based on financial market data; Management Sci. 40, 508-515.

BAGAI, I. / KOCHAR, S. C. (1986): On tail ordering and comparison of failure rates; Commun. Statist. A 15, 1377-1388.

BARLOW, R. E. / PROSCHAN, F. (1981): Statistical Theory of Reliability and Life Testing; Holt, Rinehart and Winston; New York.

BAWA, V. S. / BROWN, S. J. / KLEIN, R. W. (1979): Estimation Risk and Optimal Portfolio Choice; North Holland; Amsterdam.

BAWA, V. S. / LINDENBERG, E. B. / RAFSKY, L. C. (1979): An efficient algorithm to determine stochastic dominance admissable sets; Management Sci. 25, 609-622.

CHAKRABORTI, S. / SEN, P. K. (1992): Order restricted quantile tests under unequal right-censorship; Sankhya B 54, 150-164.

CHIKKAGOUDAR, M. S. / SHUSTER, J. J. (1974): Comparison of failure rates using rank tests; J. Amer. Statist. Assoc. 69, 411-413.

CONOVER, W. J. (1967): The distribution functions of Tsao's truncated Smirnov statistics; Ann. Math. Statist. 38, 1208-1215.

DURBIN, J. (1973): Distribution Theory for Tests Based on the Sample Distribution Function; SIAM; Philadelphia, PA.

DYKSTRA, R. L. / KOCHAR, S. C. / ROBERTSON, T. (1991): Statistical inference for uniform stochastic ordering in several populations; Ann. Statist. 19, 870-888.

FLEMING, T. R. / HARRINGTON, D. P. (1981): A class of hypothesis tests for one and two sample censored survival data; Commun. Statist. A 10, 763-794.

FRANCK, W. E. (1984): A likelihood ratio test for stochastic ordering; J. Amer. Statist. Assoc. 79, 686-691.

HARTER, H. L. / OWEN, D. B. (1970): Selected Tables in Mathematical Statistics; Vol. 1, Markham; Chicago.

KROLL, Y. / LEVY, H. (1980): Stochastic dominance: A review and some new evidence. In: H. Levy ed.; Research in Finance, Vol. II, JAI Press; Greenwich, CT.

LEE, Y. J. / WOLFE, D. A. (1976): A distribution free test for stochastic ordering; J. Amer. Statist. Assoc. 71, 722-727.

LIU, P. Y. / GREEN, S. / WOLF, M. / CROWLEY, J. (1993): Testing against ordered alternatives for censored survival data; J. Amer. Statist. Assoc. 88, 153-160.

MOSLER, K. (1982): Entscheidungsregeln bei Risiko: Multivariate stochastische Dominanz; Springer; Berlin.

MOSLER, K. / PHILIPSON, T. (1994): Specification tests of the proportional hazard model; Mimeo.

MOSLER, K. / SCARSINI, M. (1993): Stochastic Orders and Applications. A Classified Bibliography; Springer; Berlin.

NELSON, R. D. / POPE, R. D. (1991): Bootstrapped insights into empirical applications of stochastic dominance; Management Sci. 37, 1182-1194.

RAHLFS, V. W. / ZIMMERMANN, H. (1993): Scores: Ordinal data with few categories – How should they be analyzed?; Drug Information J. 27, 1227-1240.

ROBERTSON, T. / WRIGHT, F. T. / DYKSTRA, R. L. (1988): Order Restricted Statistical Inference; J. Wiley; Chichester.

SCHUMACHER, M. (1984): Two–sample tests of Cramér–von–Mises– and Kolmogorov–Smirnov–type for randomly censored data; Internat. Statist. Review 52, 263-281.

STECK, G. P. (1969): The Smirnov two–sample tests as rank tests; Ann. Math. Statist. 40, 1449-1466.

STOYAN, D. (1983): Comparison Methods for Queues and Other Stochastic Models; J. Wiley; Chichester.

TSAO, C. K. (1954): An extension of Massey's distribution of the maximum deviation between two sample cumulative step functions; Ann. Math. Statist. 25, 587-592.

WEICHSELBERGER, A. (1993): Ein neuer nichtparametrischer Anpassungstest zur Beurteilung der Lage von Verteilungen; Vandenhoek & Ruprecht; Göttingen.

WILCOX, R. R. (1990): Determining whether an experimental group is stochastically larger than a control; British J. Math. Statist. Psychol. 43, 327-333.