Negative Dependence and the Simes Inequality

Henry W. Block, Thomas H. Savits and Jie Wang

University of Pittsburgh

May 27, 2008

Outline

0. Introduction

Simes(1986) inequality

- Proven in the case of independence
- Conjectured for multivariate dependence

Sarkar(1998)

- Proved Simes inequality for general positive dependence
- Reversal conjectured for negative dependence

Present paper

- Simes reverses for negative dependence
- Other results for dependence

Impact on testing

- The standard Simes test is conservative under positive dependence
- Standard test is not conservative under negative dependence, so another test required

1. Negative Dependence

Simes inequality for negative dependence

- Samuel-Cahn (1996) - Simes inequality reversed for negatively dependent bivariate normal
- Hochberg and Rom (1995) - more general bivariate result
- Sarkar (1998) - conjecture that Simes reversed for multivariate negative dependence - could not prove result for Karlin and Rinott (1980) concept of $S - MRR_2$ (which is satisfied by most standard negatively dependent distributions)

Negative dependence concept which is stronger than $S - MMR_2$, easier to check, and satisfied by standard negatively dependent distributions called *Condition N*.

Definition 1. The random vector $(X_1, X_2, ..., X_n)$ satisfies *condition N* if there exist $n + 1$ independent random variables $S_0, S_1, ..., S_n$ each having a PF_2 density (or probability function) and a real number s such that

$$(X_1, X_2, ..., X_n) = [(S_1, S_2, ..., S_n)|S_0 + S_1 + ...S_n = s]$$

where PF_2 is defined in Karlin (1968) (see also Block et al, 1982) and the equality means the two quantities have the same distribution.

Block et al (1982) show the following distributions satisfy Condition N:

- 1) the multinomial
- 2) the equicorrelated multivariate normal with nonpositive correlations
- 3) the multivariate hypergeometric
- 4) the Dirichlet
- 5) the Dirichlet compound multinomial

Verification of *condition N* for the above: in general, well known properties that the distributions have structures as in def., e.g., from categorical data analysis - the multinomial distribution is conditional distribution of sums of independent Poisson random variables given that their sum is fixed or from Bayesian analysis - the Dirichlet is the conditional distribution of independent gammas given that its sum is fixed.

Main result:

Theorem 1. Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the ordered components of a distribution satisfying *condition N*. Then
(i) for fixed $a_1 \leq ... \leq a_n$,

$$P\{X_{(1)} \geq a_1, ..., X_{(n)} \geq a_n\} \leq 1 - (1/n)\Sigma_{i=1}^{n} F_i(a_n^-)$$

if $j^{-1}F_i(a_j^-)$ is nonincreasing in $j = 1...., n$ for all $i = 1, ..., n$ where the F_i are the marginal cdfs;
(ii) for fixed $b_1 \leq ... \leq b_n$,

$$P\{X_{(1)} \leq b_1, ..., X_{(n)} \leq b_n\} \leq (1/n)\Sigma_{i=1}^{n} F_i(b_1)$$

if $j^{-1}\overline{F}_i(b_{n-j+1})$ is nonincreasing in $j = 1...., n$ for all $i = 1, ..., n$ where $\overline{F}_i(x) = 1 - F_i(x)$.

The reversed Simes inequality given in the following Corollary is easiest to understand in the following framework:
Consider n hypotheses $H_1, H_2, ..., H_n$ with test statistics $X_1, X_2, ..., X_n$ and random p-values $P_1, P_2, ..., P_n$ (e.g. $P_i = F(X_i)$ for a left-tailed test); to test the overall hypothesis $H_0 = \cap H_i$
Simes used the test: reject H_0 if for some i, $P_{(i)} \leq \frac{i\alpha}{n}$ where $P_{(i)}$ is the ith order statistic.

Corollary. If the X_i in the previous theorem have the same marginal cdf F, then

$$P\{P_{(1)} \geq \alpha/n, P_{(2)} \geq 2\alpha/n, ..., P_{(n)} \geq n\alpha/n\} \leq 1 - \alpha$$

Proof: For a left-tailed test and assuming a continuous F

$$P_{H_0}\{Reject\ H_0\} = P_{H_0}\{\cup\{P_{(i)} \leq i\alpha/n\}$$

$$= P_{H_0}\{\cup\{X_{(i)} \leq a_i\} \geq F(a_n) = \alpha$$

where the last inequality follows from i) of the theorem with $F(a_i) = i\alpha/n$.

Another way of expressing the reversed Simes inequality, which is actually given in the above proof is

$$P_{H_0}\{\cup\{P_{(i)} \leq i\alpha/n\} \geq \alpha.$$

A second negative dependence concept is needed in the following.

Definition 2. A random vector $(X_1, X_2, ..., X_n)$ is said to be *negatively dependent through stochastic ordering* (NDS) if for any nondecreasing function h and any $i = 1, 2, ...n$,

$$E[h(X_1, ...X_{i-1}, X_{i+1}, ...X_n)|X_i]$$

is nonincreasing in X_i.

Related Definition:
$PRDS - I_0$ of Benjamini and Yekutieli(2001) where I_0 is a subset of random variables

Above NDS condition:

- weaker than Condition N (proven in Block et al,1985)
- can be used to prove Theorem 1
- multivariate normal with nonpositive correlations does not satisfy Condition N but does satisfy NDS, so combined with the above remark gives that nonpositively correlated normal satisfies reverse Simes
- NDS condition also observed by Sarkar(1998)
- not easy to show various distributions satisfy this condition; much easier to check Condition N

Example 1. For multivariate normal with nonpositive correlations, easy to show distribution is NDS; then since Theorem 1 applies, as in proof of the corollary, reverse Simes inequality holds.

Negative dependence condition for measures: A bivariate distribution with probability measure μ is said to be RR_2 if the measure μ satisfies a reverse TP_2 inequality (see (2.1) of Block et al (1982)); can be generalized to multivariate RR_2 in pairs (see Remark (vi) of same paper); a distribution satisfying this concept can be shown to satisfy most of the inequalities derived in Karlin et al ((1980) for the stronger concept of $S - MRR_2$; the following relationships hold:

$$\text{Condition } N \Rightarrow S - MRR_2 \Rightarrow \mu RR_2 \text{ in pairs} \Rightarrow \text{various inequalities}$$

$$\text{Condition } N \Rightarrow NDS$$

Not known whether $S - MRR_2$ implies NDS

Positive dependence concept introduced in Block et al (1985)

Definition 3. A random vector $(X_1, X_2, ..., X_n)$ is said to be *positively dependent through stochastic ordering* (PDS) if for any nondecreasing function h and any $i = 1, 2, ...n$

$$E[h(X_1, ...X_{i-1}, X_{i+1}, ...X_n)|X_i]$$

is nondecreasing in X_i.

Notes

1) A slightly more general definition called $PRDS - I_0$ of Def. 3 is given by Benjamini and Yekutieli (2001) (see discussion in Sec. 2.2)

2) These authors observed that the normal distribution with nonnegative correlations was PDS (also observed in Block et al, 1985)

3) Following result is obvious since this dependence implies the main inequality in the proof of Theorem 3.1 of Sarkar(1998)

Theorem 2. Theorem 3.1 of Sarkar(1998) remains valid if MTP_2 is replaced by PDS.

Example 2. The multivariate normal with nonnegative correlations satisfies the Simes inequality. This follows from Theorem 2 and Note 2) above.

Theorem 1 has a counterpart to Theorem 2 where Condition N is replaced by NDS.

- 1. Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc., Ser. B* **57**, 284-300.
- 2. Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165-1188.
- 3. Block, H.W., Savits, T.H. and Shaked, M. (1982). Some concepts of negative dependence. *Annals of Probability* **10**, 765-772.
- 4. Block, H.W., Savits, T.H. and Shaked, M. (1985). A concept of negative dependence using stochastic ordering. *Statistics and Probability Letters* 3, 81-86.
- 5. Hochberg, Y. and Rom, D.M. (1995). Extensions of multiple testing procedures based on Simes test. *J. Statist. Plann. Inference* **48**, 141-152.

- 6. Karlin, S. (1968). *Total Positivity*. Stanford University Press.
- 7. Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities II. Multivariate reverse rules distributions. *J. Multivariate Anal.* **10**. 499-516.
- 8. Samuel-Cahn. E. (1996). Is the Simes improved Bonferroni procedure conservative? *Biometrika* 83, 928-933.
- 9. Sarkar, S.K. (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *Annals of Statistics* **26**, 494-504.
- 10. Simes,R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751-754.