DEPENDENCE ORDERING IN STATISTICAL MODELS AND OTHER NOTIONS

BY TAKEMI YANAGIMOTO

Institute of Statistical Mathematics

The relation of notions of dependence ordering with other notions in statistics such as heaviness of tail and largeness of dispersion are reviewed and developed. The problem of overdispersion in reproductive studies is discussed as a practical, attractive example.

1. Introduction. The notion of positive dependence is closely associated with other various statistical notions. This has been emphasized by many authors including Yanagimoto and Sibuya (1972) and Karlin and Rinott (1980a). Recent theoretical development of the notion of positive dependence permits us better understandings of statistical models. However, as Kimeldorf and Sampson (1987) stressed, the research on dependence orderings does not appear fully developed. We often assume a family of distributions having monotone dependence. To state this clearer, consider a distribution function, $F_\alpha(x)$, of a random variable X on R^n. The suffix α, representing the degree of largeness of dependence, is considered to be favorably parameterized, if $\alpha = 0$ stands for independence and dependence ordering is monotone increasing in α; as a result $\alpha > 0$ means positive dependence of $F_\alpha(x)$.

The situation is largely different according to the value of n. In the bivariate case negative dependence of (X_1, X_2) is reasonably recognized as positive dependence of $(X_1, -X_2)$. The notion of negative dependence is much more complicated in the multivariate case than in the bivariate case. The aim of the present paper is to review and develop dependence orderings with emphasis on the relation with other statistical notions and practical models. In Sections 2 and 4 definitions of dependence orderings are studied in the bivariate case and in the multivariate case. Some relations are discussed in Sections 3 and 5.

As usual conventions, we will employ simple descriptions unless any confusion is anticipated. Therefore, for example, dependence of a distribution, that of a random variable with the distribution and that of a distribution function of the distribution are not distinguished, and an increasing function means a nondecreasing function.

2. Dependence Ordering: Bivariate Case. A systematic definition of notions of positive dependence in the bivariate case was developed by Lehmann

AMS 1980 subject classifications. 62E10.

Key words and phrases. Heaviness of tail, multivariate distribution, overdispersion, spread.

The author expresses his thanks to a referee and the editors for their helpful comments.

(1966), which was followed by many works such as Esary, Proschan, and Walkup (1967), Yanagimoto (1972), and Shaked (1977a). Among these notions, that of association introduced by Esary et al. is inconvenient to be extended to a dependence ordering.

Unified definitions of notions of positive and negative dependence given in Yanagimoto (1972) permit us a straightforward extension to those of dependence orderings. Let $F(x,y)$ and $G(x,y)$ be cumulative distribution functions both having common marginal distribution functions. Following the idea and the notations of Yanagimoto (1972) and Kimeldorf and Sampson (1987), we give the following definition.

DEFINITION 1. Let I_1 and I_2 be real intervals. We say $I_1 < I_2$ if and only if $x_1 \in I_1$ and $x_2 \in I_2$ imply $x_1 < x_2$. Define the four families of products of interval as

$$\begin{aligned} S(1) &= \{(-\infty, x_1] \times (x_1, \infty) \mid -\infty < x_1 < \infty\} \\ S(2') &= \{(x_1, x_2] \times (x_2, \infty) \mid -\infty < x_1 < x_2 < \infty\} \\ S(2'') &= \{(-\infty, x_1] \times (x_1, x_2] \mid -\infty < x_1 < x_2 < \infty\} \\ S(3) &= \{(x_1, x_2] \times (x_2, x_3] \mid -\infty < x_1 < x_2 < x_3 < \infty\}. \end{aligned}$$

Then $G(x,y)$ is said to have larger dependence $F(x,y)$ in the sense of $P(i,j)$, if it holds

$$\begin{aligned} &P_F(I_1,J_1)P_F(I_2,J_2)P_G(I_1,J_2)P_G(I_2,J_1) \\ &\leq P_G(I_1,J_1)P_G(I_2,J_2)P_F(I_1,J_2)P_F(I_2,J_1) \end{aligned}$$

for any $I_1 < I_2$ and $J_1 < J_2$ satisfying $I_1 \times I_2 \in S(i)$ and $J_1 \times J_2 \in S(j)$, where $P_F(I,J)$ represents the probability assigned by F to the rectangle $I \times J$.

The above definition provides the 16 strictly different notions of dependence orderings. The implication scheme among these orderings holds parallel to that of $P(i,j)$'s. Among them two orderings coincide with ones in existing literature. Yanagimoto and Okamoto (1969) called $G(x,y)$ has larger quadrant dependence than $F(x,y)$ when (i,j) takes (1,1). The ordering was also discussed in Tchen (1980). Kimeldorf and Sampson (1987) introduced the ordering called that $G(x,y)$ is more TP_2 than $F(x,y)$. This ordering is strictly stronger than that corresponding $(i,j) = (3,3)$.

Yanagimoto and Okamoto (1969) present another extension of the positive regression dependence ordering, which corresponds to $P(3,1)$ in Yanagimoto (1972).

DEFINITION 2. Suppose that $G(x,y)$ and $F(x,y)$ have both the common marginal distribution functions. $G(x,y)$ is called to have larger regression dependence on x than $F(x,y)$, if it holds

$$F^{-1}(u \mid x') \geq F^{-1}(v \mid x) \text{ implies } G^{-1}(u \mid x') \geq G^{-1}(v \mid x)$$

for any $x' > x$ and $0 < u$, $v < 1$, where $F^{-1}(u \mid x) = \inf\{y \mid F(y \mid x) \geq u\}$.

Schriever (1987) introduced a new dependence ordering, weaker than the above. We assume common marginal distributions to compare dependence orderings of random variables X and Y. Obviously, the assumption is inessential.

Examples of distributions $F_\alpha(x, y)$ whose dependence ordering is monotone in α were presented in Yanagimoto and Okamoto (1969), Yanagimoto (1971) and Kimeldorf and Sampson (1987). Examples cover the normal distribution and families due to Farlie (1960) or Plackett (1965).

3. Some Relations – I. Dependence ordering is expected to possess a close relation with a measure of dependence and the distribution of a test statistic for independence. When $G(x, y)$ has larger dependence than $F(x, y)$, we expect $\mathrm{Cov}_G(Y_1, Y_2) \geq \mathrm{Cov}_F(X_1, X_2)$, if they exist. This is true if $G(x, y)$ has larger quadrant dependence, which was essentially given in Lehmann (1966). Yanagimoto and Okamoto (1969) obtained monotonic properties of the distribution of a test statistic under dependence orderings. The statement (ii) was improved in Schriever (1987).

PROPOSITION 1. *(Yanagimoto and Okamoto, 1969) (i) Suppose $G(x, y)$ has larger quadrant dependence than $F(x, y)$. Let Q be Blomqvist's statistic. Then the distribution of Q under $G(x, y)$ is stochastically larger than that under $F(x, y)$.*

(ii) Suppose $G(x, y)$ has larger regression dependence on x than $F(x, y)$. Let T be a test statistic in a family satisfying regularity conditions, which includes Kendall's and Spearman's statistics. Then the distribution of T under $G(x, y)$ is stochastically larger than that under $F(x, y)$.

Next, we consider the notion of heaviness of tail of a positive distribution function $F(x)$. $F(x)$ is said to have increasing hazard rate, when $-\log\,(1 - F(x))$ is convex. This notion is regarded as describing lighter tails for $F(x)$ than that of the exponential distribution. Let $X_{(1)}$ and $X_{(2)}$ be order statistics in ascending order from independent identically distributed random variables having a distribution function $F(x)$. If $F(x)$ is exponential, $T = (X_{(1)}, X_{(2)} - X_{(1)})$ is independent. Conversely Shanbhag (1970) showed independency of T yields that $F(x)$ is exponential or geometric. Yanagimoto (1972) showed that negative dependence of T can characterize increasing hazard rate.

The notion of heaviness of tail is extended to that of heavier tail. $G(x)$ is said to have heavier HR tail than $F(x)$, if $G^{-1}(F(x))$ is convex.

PROPOSITION 2. *Suppose $G(x)$ and $F(x)$ are continuous. $G(x)$ has heavier HR tail than $F(x)$, if and only if the distribution of T under $G(x)$ has smaller regression dependence on x than that under $F(x)$.*

Results concerning positive dependence and heaviness of tail were explored in Shaked (1977b) and Shaked and Tong (1985). Yanagimoto and Sibuya (1976) showed the close relationship of the notion of heaviness of tail and that of dispersion.

4. Dependence Ordering – Multivariate Case. Only a few number of papers on positive or negative dependence in the multivariate case are found in existing literature. For our purpose, it appears that MTP_2 by Karlin and Rinott (1980a,b), negative association by Joag-Dev and Proschan (1983) and *LOD* and *UOD* by Shaked (1982) are attractive. Suppose $G(x)$ and $F(x)$ have all the common one dimensional marginal functions. $G(x)$ is said to more *LOD* than $F(x)$, if it holds that

$$(1) \qquad P_G(Y_i \le c_i,\ i=1,\ldots,n) \ \ge\ P_F(X_i \le c_i,\ i=1,\ldots,n)$$

for any c_i. The dependence ordering more *UOD* is given by replacing $\le$ by $\ge$ in (1). Karlin and Rinott (1980a,b) defined positive TP_2 dependence by

$$(2) \qquad f(x \vee y) f(x \wedge y) \ \ge\ f(x) f(y)$$

for any $x,\ y \in R^n$, where $x \vee y$ denotes the vector with each component having the greater one of corresponding components of x and y, and $x \vee y + x \wedge y = x + y$. They defined negative TP_2 dependence by reversing the inequality in (2) and adding additional requirements. Since TP_2 dependence of $f(x)$ does not necessarily mean that of a marginal density of $f(x)$, we require additional assumptions. We write a marginal density of $f(x)$ as $\tilde{f}(y)$. Analogous with Definition 1, we present a definition of larger TP_2 dependence ordering.

DEFINITION 3. Let $G(x)$ and $F(x)$ be distribution functions having density functions on a common measure. $G(x)$ is said to have larger TP_2 dependence than $F(x)$, if it holds that

$$\tilde{g}(y_1 \vee y_2)\tilde{g}(y_1 \wedge y_2)\tilde{f}(y_1)\tilde{f}(y_2) \ \ge\ \tilde{g}(y_1)\tilde{g}(y_2)\tilde{f}(y_1 \vee y_2)\tilde{f}(y_1 \wedge y_2)$$

for any marginal density $\tilde{g}(y)$ and $\tilde{f}(y)$ and any $y_1,\ y_2 \in R^m$.

The above definition is not an extension of positive or negative TP_2 dependence in a strict sense, while it is intended to extend negative TP_2 dependence in a simpler way. In the bivariate case the larger TP_2 dependence ordering is weaker than the ordering in Kimeldorf and Sampson (1987), when the density functions exist.

An attractive notion of multivariate negative dependence is negative association (Joag-Dev and Proschan, 1983). The random variable X is called negatively associated, if

$$(3) \qquad \mathrm{Cov}(\varphi(X_1),\ \phi(X_2)) \ \le\ 0,$$

provided it exists, where X_1 and X_2 are any partitions of X, and φ and ϕ are increasing functions. The extension of this notion to positive dependence is straightforward, though the corresponding notion of positive dependence is strictly weaker than well known association (Esary et al. 1967). To avoid confusion of nomenclature, we will call negative association negative weak-association (w-association).

A difficulty arises from the fact that the distributions of $\varphi(X_1)$ and $\phi(Y_1)$ are different in general, even though all the one dimensional marginals of X and Y are common. Note that the condition (3) can be replaced by that $(\varphi(X_1), \phi(X_2))$ has negative quadrant dependence. This fact permits us a definition of w-association ordering.

DEFINITION 4. Let X and Y be random variables having all the common one dimensional marginal distributions. Y is said to have larger w-association than X, if it holds that $(\varphi(Y_1), \phi(Y_2))$ has larger quadrant dependence than $(\varphi(X_1), \phi(X_2))$ for any corresponding partitions X_1, X_2 and Y_1, Y_2 and increasing functions φ and ϕ.

5. Some Relations – II. In this section we will review relations of the notion of multivariate dependence with other notions, with emphasis paid to the problem of overdispersion arising in the multi-generation experiments for reproductive toxicology.

Consider an n dimensional exchangeable random variable $X = (X_1, \ldots, X_n)$. Shaked and Tong (1985) suggested that larger dependence of X is associated with "hanging together" within components of X. They studied the joint distribution of order statistics from components of X, and obtained several relations of larger dependence with larger dispersion and majorization properties. A simple fact of their results is that larger dependence of X is associated with larger dispersion of $T = \Sigma X_i$. Recall that $V(T) = nV(X_1) + n(n-1)\mathrm{Cov}(X_1, X_2)$. Therefore, when we fix all the one dimensional marginal distributions, larger covariance of X means larger variance of T. In other words, positive dependence of X results in overdispersion of T.

The problem of overdispersion is important for a model used for analyzing reproductive experiment data (for example, Krewski, Colin, Hogan, and Yanagimoto, in press). Assume that X represents an underlying (tolerance) distribution of n fetuses in a litter, and that for a critical point c, say $c = 0$, a pathological finding is observed in the ith fetus when $X_i > c$. Define the ith component of $Z(X)$, a statistic, to be 1 for $X_i > c$, and as 0, otherwise. Note $Z(X)$ is expressed as an increasing function of X. It is widely accepted that the random variables X_i, are not necessarily independent and consequently $Z(X)$ is not a vector of independent elements. It is likely to be positively dependent (and possibly negatively dependent). Positive dependence is reasonably interpreted by the fact that fetuses within a litter share common genetic and environmental factors. Potential negative dependence may be interpreted as unequal distribution of toxic substances to fetuses in a pregnant animal.

These two interpretations can be formulated by a mixture distribution. Let $H_\alpha(x)$ be a one dimensional distribution function, which is stochastically monotone in α. Let $K(\alpha)$ be a mixing distribution. The positively dependent distribution function $F(x)$ of X can be expressed by

$$F(x) = \int \Pi_{i=1}^n H_\alpha(x_i) dK(\alpha). \tag{4}$$

When $H_\alpha(x)$ is expressed as $H(x-\alpha)$, we can give another expression,

$$X \sim U + Ve, \tag{5}$$

where $U = (U_1, \ldots, U_n)$, $e = (1, \ldots, 1)$, and V and U_i, $i = 1, \ldots, n$ are random variables having distribution functions $K(x)$ and $H(x)$, respectively. The negatively dependent distribution of X may be expressed by

$$X \sim (U_1, \ldots, U_n) \,|_{\Sigma U_i = u_0} \tag{6}$$

for a constant value u_0.

All the distributions represented in (4)–(6) have been studied as typical multivariate positive or negative distributions. A family of distribution functions given by (4) was pursued in depth by Dykstra, Hewett, and Thompson (1973) and Shaked (1975). A family of distributions given by (6) was discussed by Block, Savits, and Shaked (1982) in a more general manner. They generalized conditioning to $U_0 + \Sigma U_i$ by adding an independent random variable, and showed that such a distribution is *UOD* and also *LOD* if each random variable has a TP_2 density.

In practical applications the distribution function $K(\alpha)$ in (4) contains a parameter representing dispersion. We can expect that larger dispersion of a mixing distribution results in larger dependency of the mixture distribution, though we need marginal adjustments for comparison. The generalization of (6) by Block et al. (1982) looks appealing, and larger dispersion of U_0 is expected to result in smaller dependency of X. Finally, we present examples of the normal case (Ochi and Prentice, 1984) and the beta-binomial case (Skellam, 1948, Williams, 1975).

EXAMPLE 1. As usual the normal case presents us a simple, clear example. Let $U_i \sim N(0, \sigma^2)$, $\sigma^2 \geq 0$, $V \sim N(0, \delta^2)$, $\delta^2 \geq 0$, $U_0 \sim N(0, \tau^2)$, $\tau^2 \geq 0$ and $u_0 = 0$. Then the random variable X in (5) belongs to $N(0, \sigma^2 I + \delta^2 e'e)$ and that in (6) belongs to $N(0, \sigma^2 I - (\sigma^4/(n\sigma^2 + \tau^2))e'e)$. By adjusting all the one dimensional marginal distributions as $N(0,1)$, the former family is written as $N(0, (1-p)I + pe'e)$ for $1 \geq p \geq 0$ and the latter is written as $N(0, (1-p)I + pe'e)$ for $0 > p \geq -1/(n-1)$. Note that both the families superficially share the same form of the normal distribution. It is easily checked that the larger TP_2 dependence ordering of the above families of distributions is monotone increasing in p, that is, monotone increasing in δ^2 and monotone decreasing in τ^2. As a result the more *LOD* or the more *UOD* ordering also is monotone increasing in p.

EXAMPLE 2. The beta-binomial distribution is the most familiar distribution employed in the model for reproductive studies because of its simple form as a mixture distribution of the binomial distribution. The probability function of the beta-binomial distribution is written as

$$(7) \qquad p(x;n,\alpha,\beta) = {}_nC_t \Pi_{r=0}^{x-1}(\mu + r\theta)\Pi_{r=0}^{n-x-1}\{(1-\mu)+r\theta\}/\Pi_{r=0}^{n-1}(1+r\theta).$$

Note that (7) makes sense to some extent, even when θ is less than 0 (Prentice, 1986).

Let W be a random variable having the beta distribution, $Be(\alpha,\beta)$ and $V = \log W/(1-W)$. Let U_i be a random variable having the distribution function $\exp x/(1+\exp x)$. To define $Z(X)$ having the multivariate beta-binomial distribution we set $X = U + Ve$ and $c = 0$. The probability function of $t = \Sigma X_i$ is given by $p(t;n,\theta,\mu)/{}_nC_t$ with $\mu = \alpha/(\alpha+\beta)$ and $\theta = 1/(\alpha+\beta)$. It follows that this distribution is associated if $\theta > 0$. Straightforward calculations yield that the larger TP_2 dependence ordering of the multivariate beta-binomial distribution is monotone increasing in θ for a fixed μ as far as (7) makes sense.

REFERENCES

BLOCK, H.W., SAVITS, T.H. and SHAKED, M. (1982). Some concepts of negative dependence. *Ann. Prob.* **10** 765–772.

DYKSTRA, R.L., HEWETT, J.E. and THOMPSON, W.A. (1973). Events which are almost independent. *Ann. Statist.* **1** 674–681.

ESARY, J., PROSCHAN, F. and WALKUP, D. (1967). Association of random variables with application. *Ann. Math. Statist.* **38** 1466–1474.

FARLIE, D.J.G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47** 307–323.

JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295.

KARLIN, S. and RINOTT, Y. (1980a). Classes of orderings of measures and related correlation inequalities–I. *J. Multivar. Anal.* **10** 467–498.

KARLIN, S. and RINOTT, Y. (1980b). Classes of orderings of measures and related correlation inequalities–II. *J. Multivar. Anal.* **10** 499–516.

KIMELDORF, G. and SAMPSON, A.R. (1987). Positive dependence orderings. *Ann. Inst. Statist. Math.* **39** 113–118.

KREWSKI, D., COLIN, D. and YANAGIMOTO, T. (1989). Statistical methods for developmental toxicity studies. In *Interpretation and Extrapolation of Reproduction Data to Establish Human Safety Standards*, H.C. Grice and K.S. Khera, eds., Springer, New York.

LEHMANN, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137–1153.

OCHI, Y. and PRENTICE, R.L. (1984). Likelihood inference in a correlated probit regression model. *Biometrika* **71** 531–543.

PLACKETT, R.L. (1965). A class of bivariate distributions. *J. Amer. Statist. Assoc.* **60** 516–522.

PRENTICE, R.L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correction induced by covariate measurement errors. *J. Amer. Statist. Assoc.* **81** 321–327.

SCHRIEVER, B.F. (1987). An ordering of positive dependence. *Ann. Statist.* **15** 1208–1214.

SHAKED, M. (1977a). A family of concepts of dependence for bivariate distributions. *J. Amer. Statist. Assoc.* **72** 642–650.

SHAKED, M. (1977b). A concept of positive dependence for exchangeable random variables. *Ann. Statist.* **5** 505–515.

SHAKED, M. (1982). A general theory of some positive dependence. *J. Multivar. Anal.* **12** 219–229.

SHAKED, M. and TONG, Y.L. (1985). Some partial orderings of exchangeable random variables by positive dependence. *J. Multivar. Anal.* **17** 339–349.

SHANBHAG, D.N. (1970). The characterizations for exponential and geometric distributions. *J. Amer. Statist. Assoc.* **65** 1256–1259.

SKELLAM, J.G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *J. Roy. Statist. Soc.* Ser. B **10** 257–261.

TCHEN, A. (1980). Inequalities for distributions with given marginals. *Ann. Prob.* **8** 814–827.

WILLIAMS, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **31** 949–952.

YANAGIMOTO, T. (1971). Examples of families of distributions with positive dependence (in Japanese). Conference Proc. (limited circulation).

YANAGIMOTO, T. (1972). Families of positively dependent random variables. *Ann. Inst. Statist. Math.* **24** 559–573.

YANAGIMOTO, T. and OKAMOTO, M. (1969). Partial orderings of permutations and monotonicity of a rank correlation statistic. *Ann. Inst. Statist. Math.* **21** 489–506.

YANAGIMOTO, T. and SIBUYA, M. (1972). Stochastically larger component of a random vector. *Ann. Inst. Statist. Math.* **24** 259–269.

YANAGIMOTO, T. and SIBUYA, M. (1976). Isotonic test for spread and tail. *Ann. Inst. Statist. Math.* **28** 329–342.

INSTITUTE OF STATISTICAL MATHEMATICS
4-6-7 MINAMI-AZABU
MINATO-KU
TOKYO 106, JAPAN