INFORMATION, CENSORING, AND DEPENDENCE

By Myles Hollander[1], Frank Proschan[1], and James Sconing[1]

Florida State University, Florida State University, and University of Iowa

Hollander, Proschan, and Sconing (1987) used the theory of majorization to develop and study various information measures in the randomly right-censored model where the basic observation is $\underline{Z} = (Z, \delta)$ where $Z = \min(X, Y)$, X is the survival time, Y is the censoring time, Y is assumed to be independent of X, and $\delta = 1$ if $X \leq Y$, $= 0$ otherwise. Here we use coefficients of divergence to derive measures of how dissimilar the joint distribution of $(X, \underline{Z})$ is from the product of its marginals. These measures contain some of the HPS information measures as special cases. We also introduce various concepts of bivariate dependence to measure the degree to which Y inhibits the ability to see X.

1. Introduction and Summary. Consider the randomly censored model where X is the survival time, Y is the censoring time, and where Y is assumed to be independent of X. We observe (Z, δ) where $Z = \min(X, Y)$, $\delta = I(X \leq Y)$, where $I(A)$ denotes the indicator of the event A. Hollander, Proschan, and Sconing (1987) [hereafter referred to as HPS (1987)] used the theory of majorization to develop and study various measures for this model.

One of the measures developed by HPS (1987) for the case where X and Y are discrete is a generalization of Shannon's (1948) information in the uncensored case.

Definition 1.1. For the censored model where X and Y have discrete distributions $p_i = \Pr(X = i), q_i = \Pr(Y = i)$, the information in the experiment (X, Y) is defined to be

$$H(X,Y) = H(\underline{p}, \underline{q}) = -\sum_i q_i \left[\sum_{j \leq i} p_j \log p_j + \bar{P}_{i+1} \log \bar{P}_{i+1} \right] \tag{1}$$

[1]Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR 88-0040. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon.

AMS 1980 subject classifications. Primary 62N05; secondary 62B10.

Key words and phrases. Bivariate dependence, coefficients of divergence, association.

We are grateful to Professor Ian McKeague for helpful conversations about information. We also thank the referee and editors for helpful comments.

where $\underline{p} = (p_1, p_2, \ldots)$, $\underline{q} = (q_1, q_2, \ldots)$, and $\bar{P}_i = \sum_{j \geq i} p_j$. (The choice of the base of the logarithm is unimportant and henceforth will be defined as the base of the natural logarithm.)

When there is no censoring, (1) reduces to Shannon's (1948) measure

$$H(X) = H(\underline{p}) = -\sum_i p_i \log p_i. \tag{2}$$

HPS (1987) showed that (1) is equivalent to Shannon's mutual information $H(X) - H(X|Z, \delta)$. Other properties concerning $H(X, Y)$ established by HPS (1987) include:

I: $H(X) \geq H(X, Y)$.

II: If $Y_1 \overset{st}{\leq} Y_2$, then $H(X, Y_1) \leq H(X, Y_2)$

III: If there exists a k such that for all $i > k$, $p_i > p_{i+1}$ and $\bar{P}_k < e^{-1}$, then $H(X, i+1) - H(X, i)$ is nonincreasing in $i, i > k$. Here $H(X, i)$ is an abbreviation for $H(X, Y)$ where $Y = i$ with probability one.

Property II essentially says that information increases as censoring decreases. However, there will be limits to such an increase and Barlow and Hsiung (1983) state "it would be interesting to see when this (information) gain is marginally decreasing." Property III gives a condition for that effect. HPS (1987) used majorization to prove I and II. Goel (1986) uses (2) and Blackwell's (1951) theory of comparison of experiments to prove I and II.

When X, Y are absolutely continuous with densities p, q, respectively, the analog of (1) is

$$H(X, Y) = H(p, q) = -\int_0^\infty q(y) \left[\int_0^y p(x) \log p(x) dx + \bar{P}(y) \log \bar{P}(y)\right] dy \tag{3}$$

where P is the distribution function of X and $\bar{P} = 1 - P$. This measure of information in the continuous case was introduced and considered in Sconing (1985) and Hollander, Proschan, and Sconing (1985). They noted that unlike (1), (3) is not scale-invariant. Baxter (1989) established analogues of properties I, II, III in the absolutely continuous case using measure $H(X, Y)$ defined by (3). Baxter does not view the lack of scale-invariance to be a serious limitation for the use of (3).

The original motivation for (1) and (3) was intuitive. Suppose in the discrete case the censoring variable assumes the value i. Then the information obtained is the full information $-p_j \log p_j$, if a death occurs prior to the censoring time. Otherwise we receive partial information, $-\bar{P}_{i+1} \log \bar{P}_{i+1}$. (If a death and a censorship occur at the same time we say that a death is observed.) The definition of (expected) information follows by averaging with respect to the censoring variable.

In Section 2 we derive the f-divergence [see (5)] of the Radon-Nikodym derivative of the joint distribution of X and $\underline{Z}$ with respect to the product of marginals.

This measures how dissimilar the joint distribution is from the product of the marginals. The measure derived is seen to contain some of the HPS (1987) information measures as special cases. As the censoring variable increases stochastically (the limiting case of $Y = \infty$ with probability one can be thought of as no censoring), Z and X become more similar and thus the divergence should decrease. Conditions for this to occur are given in Theorem 2.1.

In Section 3 we introduce various notions of bivariate dependence to measure the degree to which the censoring variable Y inhibits the ability to see the survival variable X. Let $Y_1 \overset{st}{\leq} Y_2$ and let $Z_i = \min(X, Y_i), i = 1, 2$. We compare the dependence between X and Z_1 to the dependence between X and Z_2. The notions introduced are "more positive quadrant dependent," "more associated," "more left-tail decreasing," "more right-tail increasing," and "more stochastically increasing." It is then shown that (X, Z_2) is more positive quadrant dependent than (X, Z_1). With the exception of "more associated," similar results are obtained for the other notions of dependence.

2. Coefficients of Divergence. When $X < Y$ we have $Z = X$. Since the variables X, Z are often equal, in some sense their underlying probabilistic structures should be similar. From Kullback (1959), coefficients which increase as two distributions become less similar and are called coefficients of divergence.

We define our information measure in the continuous case to be

$$I_g(p_X p_{\underline{Z}}, p_{X \times \underline{Z}}) = \int_0^{\infty} q(z) \left[\int_0^z p(x) g\{p(x)\} dx + \bar{P}(z) g\{\bar{P}(z)\} \right] dz. \tag{4}$$

This measure is equivalent to a measure of information in the discrete case developed in HPS (1987) and (with $g(x) = -\log x$) advocated in the absolutely continuous case by Baxter (1989).

Note that the information is defined as a relationship between $p_X p_{\underline{Z}}$, the product of the marginal distributions of X and $\underline{Z}$, and $p_{X \times \underline{Z}}$ the joint distribution. Our coefficent I_g is actually a measure of the distance of the joint distribution from the case where X and $\underline{Z}$ are independent. That (4) is actually a coefficient of divergence follows from the results of Csiszàr.

Csiszàr (1963, 1966) generalized the Kullback-Leibler information number in the following fashion. Let $f(x)$ be a convex function on R^+ satisfying $f(0) = \lim_{x \to 0} f(x), 0 \cdot f(0/0) = 0, 0 \cdot f(a/0) = \lim_{x \to \infty} af(x)/x, a > 0$. Let u_1 and u_2 be two probability distributions on some measurable space $(\mathcal{X}, \mathcal{A})$. Let λ be a measure on $(\mathcal{X}, \mathcal{A})$ such that u_i is absolutely continuous with respect to $\lambda, i = 1, 2$. Let p_i be the Radon-Nikodym derivative of u_i with respect to λ. Define

$$I_f(u_1, u_2) = \int p_1(x) f\left[\frac{p_2(x)}{p_1(x)}\right] \lambda(dx). \tag{5}$$

$I_f(u_1, u_2)$ is the $\underline{f\text{-divergence}}$ of u_1 and u_2.

From a completely different point of view, Ali and Silvey (1965a, 1965b, 1966) and independently Ziv and Zakai (1973) obtain an expression similar to (5). Both pairs of authors consider coefficients which quantify the distance between two probability measures. Their coefficient of divergence is defined as

$$d_f(P_1, P_2) = \int_{\phi<\infty} f(\phi)dP_1 + P_2(N) \lim_{\phi\to\infty} f(\phi)/\phi \tag{6}$$

where $f(x)$ is a convex function, $\phi = dP_2/dP_1$, and N is a P_1-null set where P_2 has positive measure. The only difference between (5) and (6) is the dominating measure λ. The two measures will be identical if P_1 and P_2 are mutually absolutely continuous. Note that the measures (5) and (6) are not symmetric in p_1 and p_2. However if $g(x) = xf(1/x)$ then $I_f(p_1,p_2) = I_g(p_2,p_1)$. Further g is convex if and only if f is convex. Define a new function $f^*(x) = f(x) + g(x)$; then the measure $I_{f^*}(p_1,p_2)$ will be symmetric.

Now we can derive the coefficient of (4) using the divergence measures in (5) or (6). Consider X and $\underline{Z}$ as the two variables of interest. We derive the f-divergence of the Radon-Nikodym derivative of the joint distribution of X and $\underline{Z}$ with respect to the product of their marginals. Note that the joint density of X and $\underline{Z}$ puts positive probability on the line where $X = Z$, the 45° line passing through the origin. This line has zero two-dimensional Lebesgue measure. Thus p_1 and p_2 defined as the joint distribution of X and $\underline{Z}$ and the product of the marginals are not mutually absolutely continuous. Hence the measures in (5) and (6) are no longer equivalent. Equation (6) is now useful only if $\lim_{x\to\infty} f(x)/x$ is finite. Equation (5) requires a measure $\lambda(x)$ which dominates both the joint density of X and $\underline{Z}$ and the product of the marginals. Let $\lambda(x)$ be the sum of two-dimensional Lebesgue measure and a measure u, which is Lebesgue measure on the 45° line $\{(x,y) : x = y, x > 0, y > 0\}$. For the joint probability measure of $(X,\underline{Z})$, we write $\Pr\{X = x, \underline{Z} = (z,0)\} = p(x)q(z)$, for $x > z$, 0 otherwise, and $\Pr\{X = x, \underline{Z} = (z,1)\} = p(x)\bar{Q}(x)$, for $x = z$, 0 otherwise. Then (5) becomes

$$\begin{aligned} I_f(p_X p_{\underline{Z}}, p_{X\times\underline{Z}}) &= \int_0^\infty p(x)p(x)\bar{Q}(x) f\left\{\frac{p(x)\bar{Q}(x)}{p(x)p(x)\bar{Q}(x)}\right\} dx + \\ &\quad \int_x \int_{z<x} p(x)q(z)\bar{P}(z) f\left\{\frac{p(x)q(z)}{p(x)q(z)\bar{P}(z)}\right\} dz dx, \end{aligned}$$

which reduces to,

$$\begin{aligned} I_f(p_X p_{\underline{Z}}, p_{X\times\underline{Z}}) &= \int_0^\infty p(x)p(x)\bar{Q}(x) f\{1/p(x)\} dx \\ &\quad + \int_0^\infty q(x)\bar{P}(x)\bar{P}(x) f\{1/\bar{P}(x)\} dx. \end{aligned} \tag{7}$$

Take $g(x) = xf(1/x)$; then (7) becomes

(8) $$I_g(p_X p_{\underline{Z}}, p_{X\times\underline{Z}}) = \int_0^\infty p(x)\bar{Q}(x)g\{p(x)\}dx + \int_0^\infty q(x)\bar{P}(x)g\{\bar{P}(x)\}dx.$$

This can be rewritten to give the coefficient in (4). One would expect that (under reasonable conditions) I_g would decrease as censoring increases stochastically. Such a decrease is equivalent to the term $\psi(z) = \int_0^z p(x)g(p(x))dx + \bar{P}(z)g\bar{P}(z))$ being increasing. Assume g is differentiable; then

$$\psi'(z) = p(z)g\{p(z)\} - p(z)\bar{P}(z)g'\{\bar{P}(z)\} - p(z)g\{\bar{P}(z)\},$$

which is positive if and only if for every z

(9) $$g\{p(z)\} \geq \bar{P}(z)g'\{\bar{P}(z)\} + g\{\bar{P}(z)\}.$$

Unfortunately inequality (9) is not always satisfied. For example, take $g(x) = -\log x$ and $\bar{P}(x) = \exp\{-\lambda x\}$; then the direction of the inequality depends on λ. However some conditions can be found for $g(x)$ and $p(x)$ so that (9) is satisfied. Two such conditions are:

C1: g decreasing on [0,1] and $p(z)\{\bar{P}(z)\}^{-1} \leq 2$
C2: g increasing on [0,1] and $p(z)\{\bar{P}(z)\}^{-1} \geq 2$

The conditions C1 and C2 are introduced to keep the failure rate $p(z)\{\bar{P}(z)\}^{-1}$ from varying too much.

THEOREM 2.1. *If either C1 and C2 hold and $g'(x)$ is continuous on* $[0, \infty]$, *then $I_g(p_X p_{\underline{Z}}, p_{X\times\underline{Z}})$ is decreasing as censoring increases stochastically.*

PROOF. It is enough to show (9). Expand $g(p(z))$ in a Taylor series about $\bar{P}(z)$. Then

$$\begin{aligned} g\{p(z)\} &\geq g\{\bar{P}(z)\} + g'\{\bar{P}(z)\}\{p(z) - \bar{P}(z)\} \\ &\geq g\{\bar{P}(z)\} + \bar{P}(z)g'\{\bar{P}(z)\} + g'\{\bar{P}(z)\}(\{p(z) - 2\bar{P}(z)\} \\ &\geq g\{\bar{P}(z)\} + \bar{P}(z)g'\{\bar{P}(z)\}, \end{aligned}$$

if $g'\{\bar{P}(z)\}\{p(z) - 2\bar{P}(z)\} \geq 0$, which holds if C1 or C2 hold. ||

In terms of the original function $f(x)$, $g(x)$ decreasing is equivalent to $f(x)/x$ increasing, $1 \leq x < \infty$. Most of the functions $f(x)$ which are commonly used in f-divergence satisfy the necessary condition.

EXAMPLE 2.2.

1) $f(x) = x\log x$	$g(x) = -\log x$	Kullback-Leibler Information number
2) $f(x) = (1/2)(x^{1/2} - 1)^2$	$g(x) = (1/2)(x^{1/2} - 1)^2$	Hellinger metric
3) $f(x) = (1/2)\|x - 1\|$	$g(x) = (1/2)\|x - 1\|$	city-block distance
4) $f(x) = (x - 1)^2$	$g(x) = (x - 1)^2/x$	χ^2-distance

It is easy to verify that in the above four cases, $g(x)$ is decreasing. Note that the third function does not satisfy the conditions of Theorem 2.1. However the ordering still holds under slightly more restrictive conditions.

THEOREM 2.3. *If g is decreasing on (0,1) and $p(z)(\bar{P}(z))^{-1} \leq 1$, then $I_g(p_X p_{\underline{Z}}, p_{X\times\underline{Z}})$ is decreasing as censoring increases stochastically.*

PROOF. If g is decreasing and $p(z)/\bar{P}(z) \leq 1, g\{p(z)\} \geq g\{\bar{P}(z)\}$. Equation (9) follows since $g'(x) \leq 0$ on (0,1). ||

These last two theorems use the divergence measure as defined in (5). As was stated previously (6) is not satisfactory unless $\lim_{x\to\infty} f(x)/x < \infty$. Of the four functions cited in Example 2.2 only the second and third functions fit this criterion. In particular the third function, $f(x) = (1/2)|x-1|$ is the one originally proposed by Ali and Silvey (1965a) for measuring dispersion between the joint distribution of two variables and the product of their marginals. In the censored model, the set N corresponds to the set where $X = Z$, or equivalently, where $X \leq Y$. Then (6) becomes

$$(10)\quad d_f(p_X p_{\underline{Z}}, p_{X\times\underline{Z}}) = \int_0^\infty q(x)\bar{P}^2(x) f\{1/\bar{P}(x)\}dx + c\int_0^\infty p(x)\bar{Q}(x)dx$$

where $c = \lim_{x\to\infty} f(x)/x$.

THEOREM 2.4. *If f is such that $\lim_{x\to\infty} f(x)/x = c < \infty$ and $f(x)/x$ is increasing for $1 < x < \infty$, then $d_f(p_X p_{\underline{Z}}, p_{X\times\underline{Z}})$ increases as censoring decreases stochastically.*

PROOF. Consider (10) as an expected loss over the variable Z with loss $\bar{P}(x)f\{1/\bar{P}(x)\}$ when $Z = x$ and $Y < X$, and loss c when $X \leq Y$. So the loss function can be written as $\bar{P}(x)f\{1/\bar{P}(x)\}I(Y < X) + cI(X \leq Y)$. As Y increases stochastically, so does Z. Since $f(x)/x$ increases to c as x increases, the loss function is increasing. Hence the expected loss increases. ||

In Example 2.2 both the Hellinger metric and the city-block distance satisfy the conditions of Theorem 2.4. The conditions in Theorem 2.4 are less restrictive than those of Theorem 2.1 in the sense that there is no condition on the distribution of X. Of course the conditions in Theorem 2.4 are more restrictive in the sense that they allow fewer functions f.

3. Measures of Bivariate Dependence. Dependence measures have typically been developed to test for independence between two variables or to measure the degree to which large values of one variable go with large values of the other. Some general notions of dependence are given in the following definition.

DEFINITION 3.1.

1) Positively quadrant dependent (PQD): U and V are positively quadrant dependent if

$$\Pr(U \le u, V \le v) \ge \Pr(U \le u)\Pr(V \le v) \text{ for all } u, v. \tag{11}$$

2) <u>Associated:</u> U and V are associated if

$$\operatorname{Cov}\{\Gamma(U,V), \Delta(U,V)\} \ge 0 \tag{12}$$

for all Γ, Δ which are componentwise increasing.

3) <u>Left-Tail Decreasing</u> (LTD($V|U$)) : V is left-tail decreasing in U if

$$\Pr(V \le v | U \le u) \text{ is decreasing in } u. \tag{13}$$

4) <u>Right-Tail Increasing</u> (RTI($V|U$)) : V is right-tail increasing in U if

$$\Pr(V > v | U > u) \text{ is increasing in } u. \tag{14}$$

5) <u>Stochastically Increasing</u> (SI($V|U$)) : V is stochastically increasing in U if

$$\Pr(V > v | U = u) \text{ is increasing in } u. \tag{15}$$

These notions are ordered in strength by:

$$\mathrm{SI}(V|U) \Rightarrow \mathrm{RTI}(V|U) \Rightarrow \text{Association} \Rightarrow \text{PQD}. \tag{16}$$

The sequence of implications is the same when RTI($V|U$) is replaced by LTD($V|U$). For verification of the implications and counterexamples to the reverse implications, see Barlow and Proschan (1981). Most of the above definitions were originally given in Lehmann (1966). The notion of association was introduced in Esary, Proschan, and Walkup (1967).

The inequalities in parts 1–5 of Definition 3.1 are notions of positive dependence for a pair of variables. We now generalize these concepts to compare the levels of dependence of two sets of variables.

DEFINITION 3.2. Given four random variables U_1, U_2, V_1, V_2, we say that:

1) <u>U_1 and V_1 are more PQD than U_2 and V_2</u> if for all u, v,

$$\begin{aligned} \Pr(U_1 \le u, V_1 \le v) \;&-\; \Pr(U_1 \le u)\Pr(V_1 \le v) \\ \ge \Pr(U_2 \le u, V_2 \le v) \;&-\; \Pr(U_2 \le u)\Pr(V_2 \le v). \end{aligned} \tag{17}$$

2) <u>U_1 and V_1 are more associated than U_2 and V_2</u> if

$$\begin{gathered} \operatorname{Cov}\{\Gamma(U_1,V_1), \Delta(U_1,V_1)\} - \operatorname{Cov}\{\Gamma(U_2,V_2), \Delta(U_2,V_2)\} \ge 0, \\ \text{for all componentwise increasing functions } \Gamma, \Delta. \end{gathered} \tag{18}$$

3) <u>V_1 is more LTD in U_1 than V_2 is in U_2</u> if for all $v, u' < u$,

$$\begin{aligned}\Pr(V_1 \le v | U_1 \le u') - \Pr(V_1 \le v | U_1 \le u) \;\ge\; & \Pr(V_2 \le v | U_2 \le u') \\ & - \Pr(V_2 \le v | U_2 \le u).\end{aligned} \tag{19}$$

4) <u>V_1 is more RTI in U_1 than V_2 is in U_2</u> if for all $v, u' < u$,

$$\begin{aligned}\Pr(V_1 > v | U_1 > u) - \Pr(V_1 > v | U_1 > u') \;\ge\; & \Pr(V_2 > v | U_2 > u) \\ & - \Pr(V_2 > v | U_2 > u').\end{aligned} \tag{20}$$

5) <u>V_1 is more SI in U_1 than V_2 is in U_2</u> if for all $v, u' < u$,

$$\begin{aligned}\Pr(V_1 > v | U_1 = u) - \Pr(V_1 > v | U_1 = u') \;\ge\; & \Pr(V_2 > v | U_2 = u) \\ & - \Pr(V_2 > v | U_2 = u').\end{aligned} \tag{21}$$

REMARKS. a) With Definition 3.2, comparisons in the censored model are readily made. In our censored data applications we take $U_1 = U_2 = X$ (the survival time random variable), but note that Definition 3.2 does not require that restriction.

b) When U_1 has the same distribution as U_2 and V_1 has the same distribution as V_2, then our notion of "more PQD" given in (17) reduces to Tchen's (1980) notion of the distribution of (U_1, V_1) being "more concordant" than the distribution of (U_2, V_2).

c) Yanagimoto and Okamoto (1969) introduced an ordering which they call monotone regression dependence which is similar but not equivalent to our "more SI" ordering given in (21). They use it to prove monotonicity of some rank correlation statistics with respect to an underlying parameter measuring dependence of the random variables.

d) Schriever (1987) has generalized the ordering of Yanagimoto and Okamato (1969) by introducing an ordering which he terms "more associated." He shows that most well-known rank measures of positive dependence preserve his ordering "more associated" in populations.

e) It is not necessary for the random variables to be positively dependent for any of (17)–(21) to hold.

THEOREM 3.3. *In the censored model the amount of positive quadrant dependence increases as censoring decreases stochastically. That is, if* $Y_1 \overset{st}{\le} Y_2$ *and* $Z_i = \min(X, Y_i)$, $i = 1, 2$, *then* X *and* Z_2 *are more PQD than* X *and* Z_1.

PROOF. Consider $\Pr(X \le x, Z_i \le z) - \Pr(X \le x)\Pr(Z_i \le z)$. There are two cases.

1) If $x \leq z$, then

$$
\begin{aligned}
&\Pr(X \leq x, Z_i \leq z) - \Pr(X \leq x)\Pr(Z_i \leq z) \\
&= \Pr(X \leq x) - \Pr(X \leq x)\Pr(Z_i \leq z) = P(x)\{1 - K_i(z)\} \\
&= P(x)\bar{K}_i(z) = P(x)\bar{P}(z)\bar{Q}_i(z),
\end{aligned}
$$

where $\bar{K}_i(z) = \bar{P}(x)\bar{Q}_i(z)$, the survival function of Z_i.

2) If $x > z$, then

$$
\begin{aligned}
&\Pr(X \leq x, Z_i \leq z) - \Pr(X \leq x)\Pr(Z_i \leq z) \\
&= \Pr\{X \leq x, \min(X, Y_i) \leq z\} - \Pr(X \leq x)\Pr(Z_i \leq z) \\
&= \Pr(X \leq z) + \Pr(z \leq X \leq x, Y_i \leq z) - \Pr(X \leq x)\Pr(Z_i \leq z) \\
&= P(z) + \{P(x) - P(z)\}Q_i(z) - P(x)\{1 - \bar{P}(z)\bar{Q}_i(z)\} \\
&= \bar{Q}_i(z)\{P(z) - P(x) + P(x)\bar{P}(z)\} = \bar{Q}_i(z)P(z)\bar{P}(x). \quad \|
\end{aligned}
$$

The following theorem is an easy consequence of Theorem 3.3.

THEOREM 3.4. *For any increasing function ψ, $\int \psi\{\Pr(X \leq x, Z \leq z) - \Pr(X \leq x)\Pr(Z \leq z)\}dxdz$ will increase as censoring decreases stochastically.*

COROLLARY 3.5. *$Cov(X, Z)$ increases as censoring decreases stochastically.*

PROOF. $\mathrm{Cov}(X, Z) = \int\int\{\Pr(X \leq x, Z \leq z) - \Pr(X \leq x)\Pr(Z \leq z)\}dxdz$ and so the result is immediate from Theorem 3.4. $\|$

Covariance is, of course, a well known measure of positive dependence. Many other such measures can also be shown to increase as censoring decreases stochastically. To show this, we state the following theorem.

THEOREM 3.6. *Let $(U_i, V_i^{(1)})$, $i=1,\ldots,n$, be independent and identically distributed. Let $(U_i, V_i^{(2)})$, $i=1,\ldots,n$, be independent and identically distributed with $(U_i, V_i^{(1)})$ more PQD than $(U_i, V_i^{(2)})$, $i=1,\ldots,n$. Let r, s be concordant functions, that is, both r and s monotonic in the same direction in each argument. Then $\{r(U_1,\ldots,U_n), s(V_1^{(1)},\ldots,V_n^{(1)})\}$ is more PQD than $\{r(U_1,\ldots,U_n), s(V_1^{(2)},\ldots, V_n^{(2)})\}$.*

The proof is by induction along the lines of Theorems 1 and 2 of Lehmann (1966).

COROLLARY 3.7. *Kendall's τ, Spearman's ρ_s, and Blomqvist's q all increase as censoring decreases stochastically.*

PROOF. Kendall's $\tau = \mathrm{Cov}(\mathrm{sign}(X_2 - X_1), \mathrm{sign}(Z_2 - Z_1))$ and hence is increasing by Theorem 3.6 and Corollary 3.5. Spearman's $\rho_s = 3\mathrm{Cov}(\mathrm{sign}(X_2 - X_1), \mathrm{sign}(Z_3 - Z_1))$ and is increasing by Theorem 3.6 and Corollary 3.5. Blomqvist's $q = 2\{\Pr(X > m_x, Z > m_z) + \Pr(X \leq m_x)\Pr(Z \leq m_z)\} - 1$ where m_x and m_z are the medians of X and Z respectively. This reduces to $2\{\Pr(X > m_x, Z > m_z) - \Pr(X > m_x)\Pr(Z > m_z) + \Pr(X \le m_x, Z \le m_z) - \Pr(X \le m_x)\Pr(Z \le m_z)\}$, which (from Theorem 3.3) increases as censoring decreases stochastically. $\|$

In Example 3.8 we show that even though there is less censoring, association may decrease.

EXAMPLE 3.8. Let $\Gamma(X, Z_i) = I(X > x_1, Z_i > z_1), \Delta(X, Z_i) = I(X > x_2, Z_i > z_2)$, $i = 1, 2$, and let $x_1 < x_2 < z_1 < z_2$. Then $\mathrm{Cov}\{\Gamma(X, Z_i), \Delta(X, Z_i)\} = \bar{P}(z_2)\bar{Q}_i(z_2) - \bar{P}(z_1)\bar{Q}_i(z_1)\bar{P}(z_2)\bar{Q}_i(z_2) = \bar{P}(z_2)\bar{Q}_i(z_2)\{1 - \bar{P}(z_1)\bar{Q}_i(z_1)\}$. Choose P, Q_1, Q_2 so that $\bar{P}(z_1) = 1/2$, $\bar{Q}_1(z_1) = 1$, $\bar{Q}_2(z_1) = 1/2$, $\bar{P}(z_2) = 1/4$, $\bar{Q}_1(z_2) = 5/12, \bar{Q}_2(z_2) = 1/3$. Note that $\bar{Q}_1(z_i) \ge \bar{Q}_2(z_i), i = 1, 2$. Then $\mathrm{Cov}\{\Gamma(X, Z_1), \Delta(X, Z_1)\} = 5/96$, and $\mathrm{Cov}\{\Gamma(X, Z_2), \Delta(X, Z_2)\} = 6/96$. Thus here $Y_1 \overset{\mathrm{st}}{\ge} Y_2$ but X and Z_2 are more associated than X and Z_1.

Thus a chain of implications similar to (16) using (17)–(21) is not possible. This result is not that surprising as the ordering defined in (18) does not satisfy the properties for a positive dependence ordering as set down in Kimeldorf and Sampson (1989). In particular they show that a bivariate c.d.f. may not be less associated than its Fréchet upper bound.

This leaves the last three notions: LTD, RTI, and SI.

THEOREM 3.9. *If* $Y_1 \overset{st}{\le} Y_2$ *then*

(i) Z_2 *is more RTI in* X *than* Z_1 *is in* X.

(ii) Z_2 *is more LTD in* X *than* Z_1 *is in* X.

(iii) Z_2 *is more SI in* X *than* Z_1 *is in* X.

PROOF. i) Let $x' < x$. Then

$$\begin{aligned} \Pr(Z > z | X > x) \quad - \quad & \Pr(Z > z | X > x') = \\ & \{\Pr(X > z, Y > z, X > x)/\Pr(X > x)\} - \\ & \{\Pr(X > z, Y > z, X > x')/\Pr(X > x')\}. \end{aligned} \tag{22}$$

There are three cases to consider.

1) Let $x > x' > z$. Then (22) reduces to $\Pr(Y > z) - \Pr(Y > z) = 0$.

2) Let $x > z \ge x'$. Then (22) reduces to $\Pr(Y > z) - \{\Pr(X > z, Y > z)/\Pr(X > x')\} = \bar{Q}(z)[1 - \{\bar{P}(z)/\bar{P}(x')\}]$. This decreases as $\bar{Q}$ decreases.

3) Let $z \ge x > x'$. Then (22) reduces to $\bar{P}(z)\bar{Q}(z)[\{1/\bar{P}(x)\} - \{1/\bar{P}(x')\}] = \bar{P}(z)\bar{Q}(z)\{\bar{P}(x) \cdot \bar{P}(x')\}^{-1}\{\bar{P}(x') - \bar{P}(x)\}$, which decreases as $\bar{Q}$ decreases.

The proofs for LTD and SI follow in an analogous fashion. $\|$

THEOREM 3.10. *Let ψ be an increasing function. Then*

(1) $\int_z \int_{x<x'} \psi\{\Pr(Z \le z|X \le x') - \Pr(Z \le z|X \le x)\}dxdx'dz$ *is increasing as censoring decreases stochastically.*

(2) $\int_z \int_{x<x'} \psi\{\Pr(Z > z|X > x) - \Pr(Z > z|X > x')\}dxdx'dz$ *is increasing as censoring decreases stochastically.*

(3) $\int_z \int_{x<x'} \psi\{\Pr(Z > z|X = x) - \Pr(Z > z|X = x')\}dxdx'dz$ *is increasing as censoring decreases stochastically.*

REFERENCES

ALI, S.M. and SILVEY, S.D. (1965a). Association between random variables and the dispersion of a Radon-Nikodym derivative. *J. Roy. Statist. Soc. Ser. B* **27** 100–107.

ALI, S.M. and SILVEY, S.D. (1965b). A further result on the relevance of the dispersion of a Radon-Nikodym derivative to the problem of measuring association. *J. Roy. Statist. Soc. Ser. B* **27** 108–110.

ALI, S.M. and SILVEY, S.D. (1966). A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser. B* **28** 131–142.

BARLOW, R. and HSIUNG, J. (1983). Information in a life testing experiment. *Statistician* **32** 35–45.

BARLOW, R. and PROSCHAN, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models.* To Begin With, Silver Springs, MD.

BAXTER, L.A. (1989). A note on information and censored absolutely continuous random variables. *Stat. and Dec.* **7** 193–198.

BLACKWELL, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symp. Math. Statist. and Prob.*, University of California Press, 93–102.

CSISZÀR, I. (1963). Eine informations theoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoffschen ketten. *Magyar Tud. Akad. Mat. Kutato Int. Közl* **8** 85–108.

CSISZÀR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientarium Mathmaticarum Hungarica* **2** 299–318.

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

GOEL, P.K. (1986). Comparison of experiments and information in censored data. Technical Report No. 352, Department of Statistics, Ohio State University.

HOLLANDER, M., PROSCHAN, F., and SCONING, J. (1987). Measuring information in right-censored models. *Naval Res. Logist.* **34** 669–681.

HOLLANDER, M., PROSCHAN, F., and SCONING, J. (1985). Measures of dependence for evaluating information in censored models. Technical Report No. M706, Department of Statistics, Florida State University.

KIMELDORF, G. and SAMPSON, A. (1989). A framework for positive dependence. *Ann. Inst. Math. Statist.* **41** 31–45.

KULLBACK, S. (1959). *Information Theory and Statistics.* John Wiley, New York.

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

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

SCONING, J. (1985). Information in censored models. Florida State University Ph.D. Dissertation.

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

YANAGIMOTO, T. and OKAMOTO, M. (1969). Monotonicity of rank correlation based on partial orderings. *Ann. Inst. Statist. Math.* **21** 489–506.

ZIV, J. and ZAKAI, M. (1973). On functionals satisfying a data processing theorem. *IEEE Trans. Info. Theory IT-19* 275–283.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FL 32306-3033

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FL 32306-3033

DEPARTMENT OF STATISTICS
UNIVERSITY OF IOWA
IOWA CITY, IA 52242