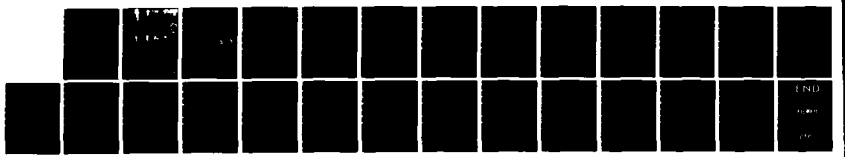
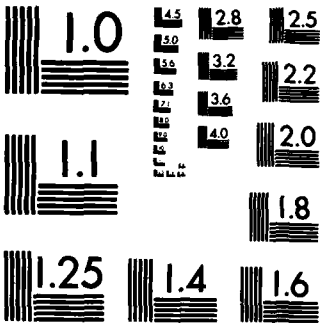
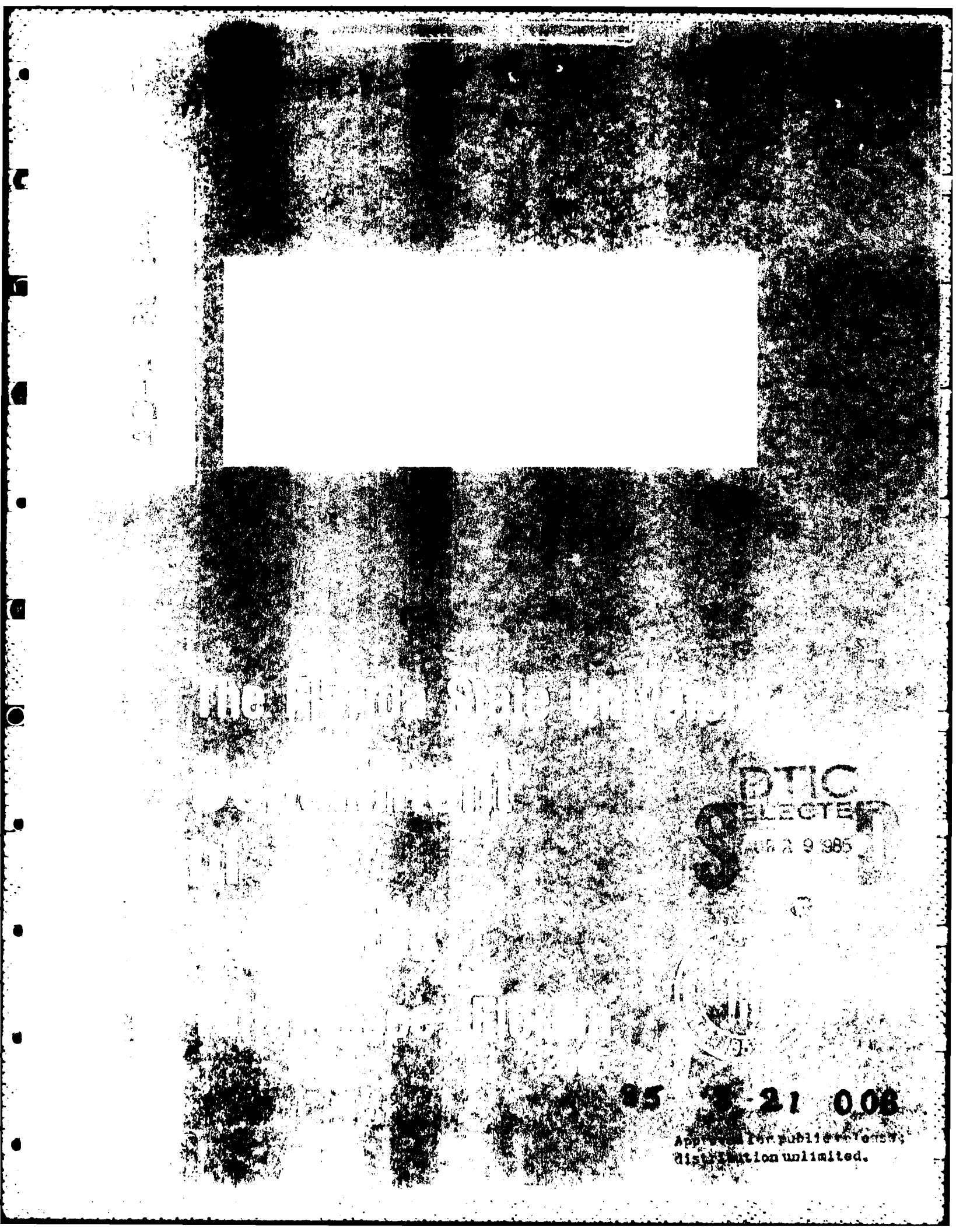
AD-A158 644 INFORMATION IN CENSORED MODELS(U) FLORIDA STATE UNIV 1/1
TALLAHASSEE DEPT OF STATISTICS M HOLLANDER ET AL.
JUN 85 FSU-STATISTICS-M701 AFOSR-TR-85-0617
UNCLASSIFIED F49620-85-C-0007 F/G 12/1 NL







Information in Censored Models

by

Myles Hollander, Frank Proschan and James Sconing

FSU-Statistics Report M701

AFOSR Technical Report No. 85-177

June, 1985

The Florida State University
The Department of Statistics
Tallahassee, Florida 32306

F49620-85-C-0007

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

AMS (1981) Subject Classification: Primary 62B10; secondary 62B15

Key words and Phrases: Information, randomly censored model, majorization.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

AD-A158644

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY ---	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 85-0617
6a. NAME OF PERFORMING ORGANIZATION Florida State University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR
6c. ADDRESS (City, State and ZIP Code) Tallahassee, FL 32306	7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, D.C. 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-85-C-0007
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, D.C. 20332-6448	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 61102F; PROJECT NO.: 2304; TASK NO.: A5; WORK UNIT NO.:
11. TITLE (Include Security Classification) Information in Censored Models	
12. PERSONAL AUTHOR(S) Myles Hollander, Frank Proschan, and James Sconing	
13a. TYPE OF REPORT Technical 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) June 1985 15. PAGE COUNT 20
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD: XXXX; GROUP: XXXXXXXXXX; SUB. GR.: XXXX	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Information, randomly censored model, majorization

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Criteria are developed for measureing information in the randomly right-censored model. Measures which are appropriate include an extension of Shannon's entropy. The measures are seen to satisfy some fundamental theorems including (i) the uncensored case is always at least as invormative as any censored model, (ii) information decreases as censoring increases stochastically, and (iii) the information gain is marginally decreasing.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Brian W. Woodruff, Maj, USAF	22b. TELEPHONE NUMBER (Include Area Code) (202)767-5027 22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

Information in Censored Models

by

Myles Hollander, Frank Proschan, and James Sconing

Florida State University

Abstract

Criteria are developed for measuring information in the randomly right-censored model. Measures which are appropriate include an extension of Shannon's entropy. The measures are seen to satisfy some fundamental theorems including (i) the uncensored case is always at least as informative as any censored model, (ii) information decreases as censoring increases stochastically, and (iii) the information gain is marginally decreasing.

1. Introduction.

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed positive random variables corresponding to the true lifetimes of some items on test. With every X_i there is a corresponding Y_i, independent of X_i. The Y_i's are also independent and identically distributed on the positive real line. Y_i is said to be the censoring variable. The observations consist of the iid pairs (Z_i, δ_i), $i = 1, \ldots, n$, where $Z_i = \min(X_i, Y_i)$, $\delta_i = I(X \leq Y)$, and $I(A)$ denotes the indicator function for the set A. This is the randomly right-censored model.

Typically the goal is to make inferences about some property of the distribution of X. The censoring variable can be thought of as a confounding variable which inhibits the ability to see X. Suppose it is desired to compare experiments where different types of censoring may take place and then decide which experiment is preferred. One approach is to use the "information" in the experiment as a basis for decision.

The term information was first used by Fisher (1925) to describe the efficiency of an estimator of some parametric component of the unknown distribution function.

A more common usage of the term is in the field of communication theory pioneered by Shannon (1948). Shannon's information can be viewed as a measure of uncertainty as to the outcome of a random variable. In our paper Shannon's measure is extended to provide a comparison of experiments in the censored model. In extending Shannon's measure to the censored case and developing other suitable measures of information, we find that the notion that more censoring should yield less information is fundamental. This property should hold for any satisfactory measure of information.

The property of decreasing information as censoring increases has also been

studied by Lindley (1956), Brooks (1982), and Barlow and Hsiung (1983), all in the Bayesian context where information is given in terms of expected risk. The connection between our approach and the Bayesian approach is given by Bernardo (1979).

Some satisfactory notions of information are developed in Sections 2, 3, and 4. Each information measure advanced is shown to satisfy Theorems 1.1 and 1.2 below.

Theorem 1.1. E[Information (X)] $\geq$ E[Information (Z, δ)] for every X and Y.

Theorem 1.2. E[Information (Z_1, δ_1)] $\leq$ E[Information (Z_2, δ_2)] for every X, where (Z_i, δ_i) is the censored variable associated with Y_i, i = 1, 2, and $Y_1 \overset{st}{\leq} Y_2$.

The form of the information measure is utilized in the proof. In the discrete case, information takes the form $E_Y[G]$ where

$$G(i) = \text{Information } (X = j \mid Y = i,\ i > j) + \text{Information } (X > i \mid Y = i).$$

The first term represents the information in observing the X variable directly. The second term is the "partial" information in observing only that X is larger than the observed variable Y. Information is given by taking the expectation over the Y variable. Theorems 1.1 and 1.2 are proved by first showing G(i) < G(i + 1), for every i > 0. This says that information is increased if the experiment is observed for the additional time from i to i + 1. With this preliminary lemma the theorems follow directly.

While information increases as censoring decreases there are limits to this increase. Barlow and Hsiung (]983) state "it would be interesting to see when this (information) gain is marginally decreasing." This leads to the following theorem.

Theorem 1.3. Let $X^{(i)}$ be the lifetime variable which is censored deterministically at time i (Type I censoring). Then for sufficiently large i, E[Information $X^{(i)}$] is a concave increasing function of i.

In Sections 2, 3, and 4 various regularity conditions are imposed to obtain versions of Theorem 1.3 for the particular information measures considered.

In Section 2 Shannon's original measure, entropy, is defined and extended to the censored case and the three fundamental theorems are proved. In Section 3 a more general class of measures is developed based on the theory of majorization. Once again the three basic theorems are proved. In Section 4 Shannon's measure is shown to be inadequate in the continuous case. Several measures are developed based on the variance of the lifetime variable X and the fundamental theorems are proved.

2. Information in the discrete case.

Shannon (1948) axiomatically derived an information measure which satisfies some intuitive requirements. Suppose a variable X takes on only two values with probabilities p and 1 - p. If an information measure is denoted H(p) (or H(X)) then it should satisfy the following requirements:

(i) $H(p) \geq 0$ for all p, $0 \leq p \leq 1$,

(ii) $H(\frac{1}{2}) = 1$, $H(1) = H(0) = 0$,

(iii) $H(X, Y) = H(X|Y) + H(Y)$ for all (X, Y),

where H(X, Y) is the information in the joint experiment (X, Y) and H(X|Y) is the conditional information in the experiment X, given the outcome of experiment Y. Imposing (i) - (iii) leads to the definition of information as

$$H(p) = -\sum_{i=1}^{n} p_i \log_2 p_i, \tag{2.1}$$

where $0\log_2 0 \stackrel{def}{=} 0$ and $\underline{p} = (p_1, p_2, \ldots, p_n)$ with $P(X = i) = p_i$. The choice of the base of the logarithm is unimportant and henceforth will be defined as the base of the natural logarithm. The definition can also be extended to the case where X has a countably infinite number of support points. This measure is termed entropy.

For the censored model where X and Y have discrete distributions, let $p_i = P(X = i)$, $q_i = P(Y = i)$, then extend (2.1) to:

Definition 2.1. The information in the discrete experiment (X, Y) is

$$H(\underline{p}, \underline{q}) = -\sum_i q_i \Big[\sum_{j \le i} p_j \log p_j + \overline{P}_{i+1} \log \overline{P}_{i+1} \Big], \tag{2.2}$$

where $\overline{P}_i = \sum_{j \ge i} p_j$.

Our definition of information in the discrete censored case can be interpreted as follows. Suppose the censoring variable takes the value i. Then the information in our observed variable Z is full information, $-p_j \log p_j$, if a death occurs prior to the censoring time. Otherwise we receive partial information, $-\overline{P}_{i+1} \log \overline{P}_{i+1}$. Note that if a death and a censorship occur at the same time we say that a death is observed. The definition follows by averaging over the censoring times. It is interesting to note that (2.2) is equivalent to Shannon's mutual information, $H(X) - H(X|Z, \delta)$. To see this write $H(X) - H(X|Z, \delta)$ as $-\sum_i p_i \log p_i - \sum_{\underline{j}} r_{\underline{j}} \sum p_{i|\underline{j}} \log p_{i|\underline{j}}$ where $r_{\underline{j}}$ is the probability that $(Z, \delta) = \underline{j} = (j_1, j_2)$ where $j_1 = 1, 2, \ldots$ and $j_2 = 0, 1$. Also $p_{i|\underline{j}}$ is the conditional probability that $X = i$ given that $(Z, \delta) = (j_1, j_2)$. The mutual information can be rewritten as $\sum_{i,\underline{j}} p_{i\underline{j}} \log(p_{i\underline{j}} / p_i r_{\underline{j}})$ where $p_{i\underline{j}}$ is the joint probability that $[X = i, (Z, \delta) = (j_1, j_2)]$. Note that $P[X = i, Z = j, \delta = 0] = p_i q_j$ if $i > j$,

0 otherwise. Also $P[X = i, Z = j, \delta = 1] = p_i\bar{Q}_j$ if $i = j$, 0 otherwise. Finally $P[Z = j, \delta = 0] = q_j\bar{P}_{j+1}$, $P[Z = j, \delta = 1] = p_j\bar{Q}_j$. With these probabilities (2.2) follows from straightforward calculations.

Now with this definition for information in the discrete censored case we show that the three basic theorems stated in Section 1 hold.

Theorem 2.2. $H(\underline{p}) \geq H(\underline{p}, \underline{q})$ for all probability vectors $\underline{p}$ and $\underline{q}$.

The theorem states that any amount of censoring reduces information. In order to prove this we first prove Lemma 2.3. This lemma has appeared in the literature in several different forms. Dobrušin (1963) showed that $H(\underline{p}) \geq H(f(\underline{p}))$ with equality if and only if $f(\underline{p})$ is a one-to-one function. Khinchin (1957) showed that if $\cup A_i = A$ and $A_i \cap A_j = \emptyset$, $i \neq j$, then $H(p(A_i)) \geq H(p(A))$, where A_i represents a set of support points for X and $p(A_i) = P(X \in A_i)$. We prove the lemma directly.

Lemma 2.3. $-\sum_j p_j \log p_j \geq -\sum_{j \leq i} p_j \log p_j - \bar{P}_{i+1} \log \bar{P}_{i+1}$, for every i.

Proof. Since $\log x$ is an increasing function, then $\log \bar{P}_{i+1} \geq \log p_j$, $j \geq i+1$. Hence, $\sum_j p_j \log p_j \leq \sum_{j \leq i} p_j \log p_j + \sum_{j > i} p_j \log \bar{P}_{i+1} = \sum_{j \leq i} p_j \log p_j + \bar{P}_{i+1} \log \bar{P}_{i+1}$. ||

We now proceed with the proof of Theorem 2.2.

Proof of Theorem 2.2.

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

$$\leq \sum q_i \left(-\sum_j p_j \log p_j\right) = H(\underline{p}) \text{ (from Lemma 2.3.)}. \quad ||$$

Next we compare amounts of information available in two models with different censoring distributions, one of which is stochastically larger than the other. First we prove the following lemma.

Lemma 2.4. Let $G_i = \sum_{j \le i} p_j \log p_j + \overline{P}_{i+1} \log \overline{P}_{i+1}$. Then $G_i \ge G_{i+1}$, for $i = 1, 2, \ldots$.

Proof. $G_i - G_{i+1} = -p_{i+1} \log p_{i+1} + \overline{P}_{i+1} \log \overline{P}_{i+1} - \overline{P}_{i+2} \log \overline{P}_{i+2}$

$$= p_{i+1}(\log \overline{P}_{i+1} - \log p_{i+1}) + \overline{P}_{i+2}(\log \overline{P}_{i+1} - \log \overline{P}_{i+2}) \ge 0,$$

since $\overline{P}_{i+1} \ge p_{i+1}$ and $\overline{P}_{i+1} \ge \overline{P}_{i+2}$. ||

We are now ready to prove an analogue of Theorem 1.2.

Theorem 2.5. Let $Y_1 \overset{st}{\le} Y_2$. Let Y_i have outcome probability vector $\underline{q}_i$, $i = 1, 2$. Then $H(\underline{p}, \underline{q}_1) \le H(\underline{p}, \underline{q}_2)$ for every life distribution vector $\underline{p}$.

Proof. From (2.2) we see that $H(\underline{p}, \underline{q}_i) = E_{Y_i}(-G)$, $i = 1, 2$, where G is the function defined by $G(i) = G_i$ as in Lemma 2.4. From Lemma 2.4 G_i is increasing (nondecreasing) in i. Thus $E_{Y_1}(-G) \le E_{Y_2}(-G)$. ||

Thus we see that our intuition has been justified in the simple case where Shannon's entropy is the measure of information. Theorem's 2.2 and 2.5 should represent a sort of "acid test" for the applicability of any measure of information.

The condition of stochastic domination of the censoring variables is also necessary. If stochastic domination does not occur then there exists an interval where $Y_1 \overset{st}{\le} Y_2$ and another interval where $Y_2 \overset{st}{\le} Y_1$. By defining X to have support only on one of these intervals and applying Theorem 2.5, a contradiction arises.

In a similar fashion stochastically decreasing the lifetime variable X also yields more information. Note that (2.1) and (2.2) are scale invariant. By relabeling the axis after stochastically decreasing X and then applying Theorem 2.5 we get analogous results.

We now establish a parallel for Theorem 1.3.

Theorem 2.6. If there exists a k such that for all $i > k$, $p_i > p_{i+1}$, and $\overline{P}_k < e^{-1}$, then $G_i - G_{i+1}$ is decreasing (nonincreasing) in i, $i > k$.

Proof. It is sufficient to show $(G_{i-1} - G_i) - (G_i - G_{i+1}) \geq 0$.

$$(G_{i-1} - G_i) - (G_i - G_{i+1}) = -p_i \log p_i + p_{i+1} \log p_{i+1} + 2\overline{P}_{i+1} \log \overline{P}_{i+1} - \overline{P}_i \log \overline{P}_i - \overline{P}_{i+2} \log \overline{P}_{i+2}$$

$$\geq 2\overline{P}_{i+1} \log \overline{P}_{i+1} - \overline{P}_i \log \overline{P}_i - \overline{P}_{i+2} \log \overline{P}_{i+2} \quad (\text{since } \overline{P}_k < e^{-1})$$

$$\geq 2[\overline{P}_{i+1} \log \overline{P}_{i+1} - (\tfrac{1}{2})(\overline{P}_i + \overline{P}_{i+2}) \log\{(\tfrac{1}{2})(\overline{P}_i + \overline{P}_{i+2})\}] \geq 0. \quad \|$$

The conditions of the theorem assure that the index i is far enough in the right-hand tail for the marginally decreasing property to take hold.

3. Majorization and information.

We wish to consider generalizations of Shannon's entropy measure. In particular requirement (iii) of Section 2 is not universally accepted, and it is this requirement which leads to the specific functional form for Shannon's entropy. By relaxing this assumption we can generalize Shannon's measure. Note that the information in an event is governed solely by the probability of that event. Thus information in the event labeled i is given by $f(p_i)$. What types of measures perform satisfactorily as measures of information? To answer this

question the theory of the majorization ordering is utilized. Majorization is a powerful tool, useful in proving inequalities. The standard reference on majorization is Marshall and Olkin (1979). First we state some definitions and preliminary theorems.

Definition 3.1. Let $\underline{x}, \underline{y} \in R^n$, n-dimensional Euclidean space. We say $\underline{y}$ majorizes $\underline{x}$ ($\underline{y} \overset{m}{\succ} \underline{x}$) if

$$(1) \quad \sum_{i=1}^{k} y_{[i]} \geq \sum_{i=1}^{k} x_{[i]}, \quad k = 1, \ldots, n-1;$$

and,

$$(2) \quad \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i,$$

where $x_{[1]} \geq x_{[2]} \geq \ldots \geq x_{[n]}$, $y_{[1]} \geq y_{[2]} \geq \ldots \geq y_{[n]}$ are the decreasing rearrangements for $\underline{x}$ and $\underline{y}$ respectively.

An equivalent definition of majorization is given by:

Definition 3.2. Let $\underline{x}, \underline{y} \in R^n$. Then $\underline{y} \overset{m}{\succ} \underline{x}$ if and only if there exists a doubly stochastic matrix P, such that $\underline{x} = \underline{y}P$.

A matrix is doubly stochastic if each row sum and each column sum equals one. This definition illuminates why majorization is particularly useful in the study of information under censoring. The $\underline{y}$ vector can be thought of as the probabilities when a censorship has occurred, that is, $(p_1, p_2, \ldots, p_i, \overline{P}_{i+1}, 0, \ldots)$. The $\underline{x}$ vector is the vector of probabilities of the life distribution $(p_1, p_2, \ldots)$. Note that

$$(3.1) \qquad (p_1, p_2, \ldots, p_i, \overline{P}_{i+1}, 0, \ldots) \overset{m}{\succ} (p_1, p_2, \ldots)$$

for every i. The doubly stochastic matrix in this case consists of the

conditional probabilities of surviving to time j given that the item was censored at time i, $i < j$.

Functions which preserve (reverse) the ordering of majorization are called Schur-convex (Schur-concave).

Definition 3.3. A function f: $R^n \to R^1$ is said to be Schur-convex (Schur-concave) if $\underset{\sim}{y} \overset{m}{\succ} \underset{\sim}{x}$ implies $f(\underset{\sim}{y}) \geq (\leq) f(\underset{\sim}{x})$.

Theorem 3.4. (Schur, 1923, Ostrowski, 1952). A permutation invariant function ϕ is Schur-convex (concave) if and only if $(z_i - z_j)(\phi_{(i)}(\underset{\sim}{z}) - \phi_{(j)}(\underset{\sim}{z})) \geq (\leq) 0$, where $\phi_{(i)}(\underset{\sim}{z})$ is the partial derivative of ϕ with respect to z_i.

It is useful to identify specific types of functions which can represent the average information in a random variable.

Theorem 3.5. (Schur, 1923). Let $\phi(\underset{\sim}{x}) = \sum_{i=1}^{n} f(x_i)$ where f: $R^1 \to R^1$. Then ϕ is Schur-convex (concave) if and only if f is a convex (concave) function.

This provides a basis for constructing information measures with a general function f. Let the information in the occurrence of a death at time i be represented by $f(p_i)$. Two possibilities for classes of information measures can be obtained as follows. Define $A = \{f:$ f(x) is decreasing and f(x)/x is concave$\}$ and $B = \{g:$ g(x) is concave and g(x)/x decreasing$\}$. Then there are the following two candidates for general information measures,

$$H_f(\underset{\sim}{p}) = \sum_i p_i f(p_i), \quad f \in A, \tag{3.2}$$

or

$$H_g(\underset{\sim}{p}) = \sum g(p_i), \quad g \in B. \tag{3.3}$$

Solomon (1979) uses a measure similar to (3.3) to measure ecological diversity. The measure given by (3.2) will be adopted here as it represents the average information in an experiment and it facilitates the proof of Theorem 3.12.

Definition 3.6. Let $\underline{p}$ be the vector of probabilities associated with a life variable X. Then the "f-type" information measure in X is given by (3.2).

Note that from the definition of the class A, $H_f(\underline{p})$ is Schur-concave.

Definition 3.7. Let $\underline{p}$ and $\underline{q}$ be the probability vectors associated with X and Y. Then the amount of information in the censored model is defined to be

$$H_f(\underline{p}, \underline{q}) = \sum_i q_i \Big[\sum_{j \le i} p_j f(p_j) + \overline{P}_{i+1} f(\overline{P}_{i+1}) \Big], \tag{3.4}$$

where $f \in A$.

Choosing $f(x) = -\log x$ gives (2.2), however (3.4) cannot be obtained as a measure of Shannon's mutual information. Henceforth, $0f(0) \stackrel{def}{=} 0$ to fix the location.

Lemma 3.8. $\sum_j p_j f(p_j) \ge \sum_{j \le i} p_j f(p_j) + \overline{P}_{i+1} f(\overline{P}_{i+1})$ for every i.

Proof. This follows immediately from (3.1) and the fact that $H_f(\underline{p})$ is Schur-concave. ||

Theorem 3.9. $H_f(\underline{p}) \ge H_f(\underline{p}, \underline{q})$ for every $\underline{p}$ and $\underline{q}$.

Proof. The proof is analogous to that of Theorem 2.2. ||

Lemma 3.10. Let G(f) be a function defined by $G(f(i)) = G_f(i) = \sum_{j \le i} p_j f(p_j) + \overline{P}_{i+1} f(\overline{P}_{i+1})$. Then $G_i(f) \le G_{i+1}(f)$, $i = 1, 2, \ldots$.

Proof. Note that $(p_1, p_2, \ldots, p_{i+1}, \overline{P}_{i+2}, 0, \ldots)$ $\overset{m}{<} (p_1, p_2, \ldots, p_i, \overline{P}_{i+1}, 0, \ldots)$ and that $H_f(\underline{p})$ is Schur-concave. ||

Theorem 3.11. Let Y_1 and Y_2 be censoring variables with probability vectors $\underline{q}_1$ and $\underline{q}_2$ respectively. Let $Y_1 \overset{st}{\leq} Y_2$. Then for every $\underline{p}$, $H_f(\underline{p}, \underline{q}_1) \leq H_f(\underline{p}, \underline{q}_2)$.

Proof. With Lemma 3.10 the proof is analogous to that of Theorem 2.5. ||

Theorem 3.12. If there exists a k such that for all $i > k$, $p_i > p_{i+1}$, then $G_i(f)$ is a concave function of i, $i > k$.

Proof. The proof is similar to the proof in Theorem 2.6. ||

Thus the "f-type" information measures are suitable for the discrete censored model.

4. Information in the continuous case.

Our goal is to extend our definition to include life distributions which are continuous. The obvious analogue of Definition 2.1 would be to define $H(p(x)) = -\int p(x)\log p(x)\,dx$. However, Example 4.1 shows that such a definition is unsatisfactory.

Example 4.1. Let $p(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 < x < \infty, \ \lambda > 0, \\ 0 & \text{otherwise.} \end{cases}$

Then $H(p(x)) = -\int_0^\infty \lambda e^{-\lambda x}[-\lambda x + \log\lambda]\,dx = 1 - \log\lambda$.

From Example 4.1 it is seen that $H(p(x)) \geq 0$ if and only if $\lambda > e$. Thus the base of the logarithm is crucial in determining a key property of information. Furthermore $H(p(x))$ does not have the scale invariant property present in the discrete case. Finally note that if $\lambda < e$, then $H(p(x)) \leq 0$ so that an observation

will decrease our knowledge. All these properties run counter to the properties which measures of information should possess. Thus $H(p(x))$ as defined above is unsatisfactory for defining information.

In order to find a new measure of information, recall the properties Shannon used to define entropy: (i) $H(p) \geq 0$, (ii) $H(\frac{1}{2}, \frac{1}{2}) = 0$ and $H(1, 0) = 1$ and (iii) $H(X, Y) = H(X|Y) + H(Y)$. The first two requirements simply fix the scale and thus are not crucial. It is the third requirement, the so-called additivity criterion, that is crucial in defining entropy. It is desirable to retain this crucial property in the continuous case. Restricting consideration to functions of the form $\partial(X, EX)$, where $\partial(\cdot, \cdot)$ is a metric, leads to $H(X) = E(X - EX)^2 = \sigma_X^2$ (Blyth, 1959). This suggests:

Definition 4.2. Let X be a continuous random variable on the positive real line with p.d.f. $f(x)$ and finite variance. Then the information in X is defined to be $H(X) = H(f) = \int_0^\infty (x - \mu)^2 f(x)\, dx = \sigma_X^2$, where $\mu = \int_0^\infty xf(x)dx$.

Note that information, in any sense, measures the spread of the distribution. From this it seems unreasonable to expect any measure of information to be scale-invariant in the continuous case. Thus when comparing measures of information care must be taken to use the same scale of measurement. Definition 4.2 gives a measure of information in the uncensored case. Recall that in the discrete case there is full information if death occurs prior to censorship, and only partial information, $-\bar{P}_{i+1}\log\bar{P}_{i+1}$, otherwise. In the case of censoring the only constraint is that the remaining probabilities sum to $\bar{P}_{i+1}$. Note that among all discrete probability distributions which have probability $\bar{P}_{i+1}$ remaining, the one that gives the least amount of information is that which puts its entire remaining mass at a single point. This would yield information $-\bar{P}_{i+1}\log\bar{P}_{i+1}$. Thus

$-\overline{P}_{i+1}\log\overline{P}_{i+1}$ can be viewed as a type of "worst case" under the probability constraint. With this "worst case" type of reasoning for the variance measure, information measures can be developed.

In the continuous case, minimizing information is equivalent to minimizing $\vartheta(X, EX)$. Given that the censorship takes place at time c, the constraint is that the remaining probability, $\overline{F}(c)$, must be placed in the set $A = \{x: x > c\}$. It is easy to show that if $c \le EX$, then $\vartheta(X, EX)$ is minimized by placing all the remaining mass at EX. If $c > EX$, then $\vartheta(X, EX)$ is minimized by placing all the mass at c. We now give a definition for information in the continuous censored case.

Definition 4.3. Let X be a lifetime variable with p.d.f. f(x) and finite variance. Let Y be a censoring variable with p.d.f. g(y). Let $Z = \{\min(X, Y), I(X \le Y)\}$ be the observed variable. Then the information in Z is defined to be:

$$H^{(1)}(X, Y) = H^{(1)}(f, g) = \int_0^\infty g(c)\left[\int_0^c (x-\mu)^2 f(x)dx + (c-\mu)^2\overline{F}(c)I(c > \mu)\right]dc;$$

equivalently,

$$H^{(1)}(f, g) = \int_0^\infty g(c)\int_0^c (x-\mu)^2 f(x)dx + \int_\mu^\infty g(c)(c-\mu)^2\overline{F}(c)dc.$$

From this definition results analogous to those of the discrete case are obtained.

Lemma 4.4. Let $k_c^{(1)} = \int_0^c (x-\mu)^2 f(x)dx + (c-\mu)^2\overline{F}(c)I(c > \mu)$. Then for every $c > 0$, $\sigma_X^2 \ge k_c^{(1)}$.

Proof. $\sigma_X^2 - k_c^{(1)} = \int_c^\infty (x-\mu)^2 f(x)dx - (c-\mu)^2\overline{F}(c)I(c > \mu)$.

Case 1. If $c < \mu$, the second term is zero, and $\sigma_X^2 - k_c^{(1)} = \int_c^\infty (x-\mu)^2 f(x)dx \geq 0$.

Case 2. If $c \geq \mu$, then $\sigma_X^2 - k_c^{(1)} \geq (c-\mu)^2 \int_c^\infty f(x)dx - (c-\mu)^2 \bar{F}(c) = 0$. ||

Theorem 4.5. $H^{(1)}(X, Y) \leq H(X)$.

Proof. $H^{(1)}(X) = \int_0^\infty (x-\mu)^2 f(x)dx = \int_0^\infty g(c)[\int_0^\infty (x-\mu)^2 f(x)dx]dc$

$\geq \int_0^\infty g(c)[k_c^{(1)}]dc = H^{(1)}(X, Y)$. ||

Lemma 4.6. $k_c^{(1)}$ is increasing in c.

Proof. $dk_c^{(1)}/dc = (c-\mu)^2 f(c) - (c-\mu)^2 f(c) I(c > \mu) + 2\bar{F}(c)(c-\mu)^2 I(c > \mu)$

$$= \begin{cases} (c-\mu)^2 f(c) & \text{if } c \leq \mu, \\ 2\bar{F}(c)(c-\mu) & \text{if } c > \mu, \end{cases}$$

and each expression is positive. ||

Theorem 4.7. Suppose that Y_1 and Y_2 are censoring variables with d.f.'s G_1 and G_2 respectively. Suppose $Y_1 \overset{st}{\leq} Y_2$. Then $H^{(1)}(X, Y_1) \leq H^{(1)}(X, Y_2)$.

Proof. Define a function $k^{(1)}$ by $k^{(1)}(c) = k_c^{(1)}$ as defined in Lemma 4.4. Then $H^{(1)}(X, Y_i) = E_{Y_i}(k^{(1)})$. From Lemma 4.6 the conclusion follows. ||

Definition 4.8. X is said to have an increasing failure rate (IFR) if $r(t) = f(t)(\bar{F}(t))^{-1}$ is increasing in t.

Theorem 4.9. Let censoring be deterministic at time c and let X be an IFR variable. If there exists a value A such that f(x) is decreasing for $x > A$, then for c sufficiently large, $H^{(1)}(X, c)$ is a concave increasing function of c.

Proof. $H^{(1)}(X, c) = k_c^{(1)}$ which is increasing from Lemma 4.6.

Computing

$$dk_c^{(1)}/dc = \begin{cases} 2(c-\mu)f(c) + (c-\mu)^2 f'(c) & \text{for } c \le \mu, \\ 2\bar{F}(c) - 2(c-\mu)f(c) & \text{for } c > \mu. \end{cases}$$

The first term is negative if f is decreasing; thus we need only consider the second term. We have $2\bar{F}(c) - 2(c-\mu)f(c) \ge 0$ if and only if $(c-\mu)^{-1} \le r(c)$. But $(c-\mu)^{-1} \to 0$ as $c \to \infty$. Thus if X is IFR and c is sufficiently large, then the inequality holds. ||

Theorem 4.9 shows that more censoring yields less information; however, this relationship is not as strong as one would like. Consider two censoring distributions G_1 and G_2, where G_1 is stochastically larger than G_2 up to time μ and equal thereafter. Then the difference in information reduces to $\int_0^\infty (x-\mu)^2 f(x)(\bar{G}_1(x) - \bar{G}_2(x))dx$. This term is positive from Theorem 4.7 but it merely reflects the information in those observations where a death occurred under Model 1 and a censorship occurred under Model 2. The difference for the censored observations is zero even though they are stochastically larger in one case than in the other. This occurs because all censored observations which occur prior to time μ are shifted to μ, regardless of when they actually occur. An alternate measure is sought which will more carefully distinguish among censored observations. This can be achieved by a constraint which was previously ignored, that corresponding to the value of the mean of the distribution, μ. Again, the "worst case" will be used under this new set of restrictions. Given that censorship takes place at time c, consider a new variable, X^0, with p.d.f. $f^0(x)$, which equals f(x) for $x < c$, and minimizes $\int_c^\infty (x-\mu)^2 f^0(x)dx$, under the restrictions that $\int_c^\infty f^0(x) = \bar{F}(c)$ and $\int_c^\infty x f^0(x)dx = \int_c^\infty x f(x)dx$. It can be shown that $f^0(x)$ must put all its mass at the point $\alpha(c) = (\bar{F}(c))^{-1}\int_c^\infty x f(x)dx$. This gives a new definition for information.

Definition 4.10. Let X, Y, Z be defined as in Definition 4.3. Then the information in the random variable Z is defined by $H^{(2)}(X, Y)$
$= \int_0^\infty g(c)[\int_0^c (x-\mu)^2 f(x)dx + (\alpha(c)-\mu)^2\bar{F}(c)]dc$.

Lemma 4.11. Let $k_c^{(2)} = \int_0^c (x-\mu)^2 f(x)dx + (\alpha(c)-\mu)^2\bar{F}(c)$. Then for every $c > 0$, $\sigma_X^2 \geq k_c^{(2)}$.

Proof. $\sigma_X^2 - k_c^{(2)} = \int_c^\infty (x-\mu)^2 f(x)dx - (\alpha(c)-\mu)^2\bar{F}(c)$
$= \int_c^\infty (x-\alpha(c))^2 f(x)dx + 2(\alpha(c)-\mu)\int_c^\infty (x-\alpha(c))f(x)dx - \int_c^\infty (x-\alpha(c))^2 f(x)dx \geq 0$. ||

Theorem 4.12. $H(X) \geq H^{(2)}(X, Y)$, for every X, Y.

Proof. From Lemma 4.11 the proof is the same as that of Theorem 4.5. ||

Lemma 4.13. $k_c^{(2)}$ is increasing in c.

Proof. Direct calculations show that $dk_c^{(2)}/dc = f(c)(c-\mu)^2 \geq 0$. ||

Theorem 4.14. Let X, Y_1, Y_2 be as in Theorem 4.7. Then $H^{(2)}(X, Y_1) \leq H^{(2)}(X, Y_2)$ for every X.

Proof. From Lemma 4.13 the proof follows along the lines of the proof of Theorem 4.7. ||

Definition 4.15. A random variable is said to have increasing (decreasing) mean residual life IMRL(DMRL), if $g(y) = (\bar{F}(y))^{-1}\int_0^\infty \bar{F}(y+t)dt$ is increasing (decreasing) in y.

Theorem 4.16. Suppose censoring is deterministic at time c and X is a DMRL variable. If there exists a number A such that f(x) is decreasing for all $x > A$ then, for sufficiently large c, $H^{(2)}(X, c)$ is a concave increasing function of c.

Proof. $H^{(2)}(X, c) = k_c^{(2)}$ which is increasing by Lemma 4.13. Also $d^2k_c^{(2)}/dc^2$ $= 2f(c)(c - \alpha(c))[1 - (\bar{F}(c))^{-2}(-cf(c)\bar{F}(c) + \alpha(c)f(c)\bar{F}(c))] + (c - \alpha(c))^2 f'(c)$. The second term is negative beyond the point A. The first term is negative if $f(c)(\bar{F}(c))^{-1} \leq (\alpha(c) - c)^{-1}$. We now use the following identity of Meilijson (1971), $g(c)r(c) = 1 + g'(c)$, where $r(c)$ is the failure rate at c, and $g(c) = \alpha(c) - c$ is the mean residual life function at c. If X is a DMRL variable then $g'(c) \leq 0$. Thus $g(c)r(c) \leq 1$ and the conclusion follows. ||

The discrete case can be paralleled in one more fashion. The "worst case" scenario is no longer used. Now the remaining mass $\bar{F}(c)$ is simply moved to the point of censoring. Note that this is not the same as in Definition 4.3. There, mass was sometimes displaced to the right. Here, it is always displaced to the left.

Definition 4.17. Let X, Y, and Z be as in Definition 4.3. Then information in the variable Z is $H^{(3)}(X, Y) = \int_0^\infty g(c)[\sigma_X^2\bar{F}(c) + \int_0^c (x - \mu_c^*)^2 f^*(x)dx]dc$, where μ_c^* $= (\bar{F}(c))^{-1}\int_0^c x f^*(x)dx$ and $f^*(x) = f(x)$ for $x < c$, $= \bar{F}(x)$ for $x = c$.

Lemma 4.18. Let σ_c^2 denote the variance of the truncated density f^*. Then $\sigma_c^2 \leq \sigma_X^2$.

Proof. Let X_1, X_2 be iid copies of X with p.d.f $f(x)$. Then $2\sigma_X^2 = E(X_1 - X_2)^2$ and $2\sigma_c^2 = E(X_1^c - X_2^c)$ where X_i^c is the truncated version of X_i, $i = 1, 2$. Then letting $A = \{X_1 < c, X_2 < c\}$, $B = \{X_1 < c, X_2 \geq c\}$, $C = \{X_1 \geq c, X_2 < c\}$, and $D = \{X_1 \geq c, X_2 \geq c\}$ we have

$$\begin{aligned} 2\sigma^2 &\geq \iint_A (x_1 - x_2)^2 f(x_1)f(x_2)dx_1dx_2 + \iint_B (x_1 - x_2)^2 f(x_1)f(x_2)dx_1dx_2 \\ &\quad + \iint_C (x_1 - x_2)^2 f(x_1)f(x_2)dx_1dx_2 \\ &\geq \iint_A (x_1 - x_2)^2 f(x_1)f(x_2)dx_1dx_2 + \int_0^c (x_1 - c)^2 f(x_1)\bar{F}(c)dx_1 \\ &\quad + \int_0^c (c - x_2)^2 f(x_2)\bar{F}(c)dx_2 = 2\sigma_c^2. \quad \| \end{aligned}$$

Theorem 4.19. $H^{(3)}(X, Y) \le H(X)$.

Proof. From Lemma 4.18 we have

$$H^{(3)}(X, Y) = \int_0^\infty g(c)[F(c)\sigma_X^2 + \bar{F}(c)\sigma_c^2]$$
$$\le \int_0^\infty g(c)[F(c)\sigma_X^2 + \bar{F}(c)\sigma_X^2] = \sigma_X^2 = H(X). \quad \|$$

Lemma 4.20. σ_c^2 is increasing in c.

Proof. Let $c_1 < c_2$. It is enough to show $\sigma_{c_1}^2 \le \sigma_{c_2}^2$. Denote the two random variables as X_{c_1} and X_{c_2}, then X_{c_1} can be obtained from X_{c_2} by truncating X_{c_2} at c_1. The desired result follows from Lemma 4.18. $\|$

Lemma 4.21. Let $L_c = F(c)\sigma_X^2 + \bar{F}(c)\sigma_c^2$. Then L_c is increasing in c.

Proof. $\partial L_c/dc = f(c)(\sigma_X^2 - \sigma_c^2) + \bar{F}(c)(d\sigma_c^2/dc)$. Now, from Lemma 4.18 and Lemma 4.20, $dL_c/dc \ge 0$. $\|$

Theorem 4.22. Let X, Y_1, and Y_2 be defined as in Theorem 4.7. Then $H^{(3)}(X, Y_1) \le H^{(3)}(X, Y_2)$.

Proof. From Lemma 4.21, the conclusion follows as in Theorem 4.7. $\|$

Theorem 4.23. Let X be an IFR variable. Suppose censoring is deterministic at time c. Suppose there exists a value A such that f(x) is decreasing for all $x > A$. Then for c sufficiently large, $H^{(3)}(X, c)$ is a concave, increasing function of c.

Proof. $H^{(3)}(X, c) = L_c$, which is increasing by Lemma 4.21. Also,

$$d^2L_c/dc^2 \le -2f(c)(d\sigma_c^2/dc) + \bar{F}(c)(d^2\sigma_c^2/dc^2).$$

Thus $d^2L_c/dc^2 \le 0$ if

$$f(c)(\bar{F}(c))^{-1} \ge F(c)[3\int_0^c (c - y)^2 f(y)dy]^{-1}.$$

The term on the right decreases to zero. Hence since X is an IFR variable, the result holds. ‖

It is interesting to note that Rao (1983) also suggests variance as a measure of ecological diversity. He considers measures of the form $\int\int k(X, Y)dP_x dP_y$, where $k(\cdot, \cdot)$ is a kernel measuring the distance between X and Y. Taking $k(X, Y) = (X - Y)^2$ gives the variance measure.

We also note that alternate proofs of some of our results can be obtained by using Blackwell's (1951) method for comparing two experiments. For example, to show that the uncensored case is always at least as informative as any censored model, let P denote the distribution of the lifetime variable X, Q the distribution of the independent censoring variable Y. Transform X to (Z, δ) by $(Z = X, \delta = 1)$ if $X \leq Y^*$, $(Z = Y^*, \delta = 0)$ if $X > Y^*$, where Y^* is independent of X and has the distribution Q.

Acknowledgments. We are grateful to Ian McKeague for helpful comments and to Prem Goel for a useful discussion concerning the relationship of our approach to David Blackwell's method of comparing two experiments.

References

Barlow, R. and Hsiung, J. (1983). Information in a Life Test Experiment. The Statistician 32, 35-45.

Bernardo, J. (1979). Expected Information as Expected Utility. Ann. Statist. 7, 686-690.

Blackwell, D. (1951). Comparison of Experiments. Proc. Second Berkeley Symp. Math. Statist. Prob. Univ. California Press. 93-102.

Blyth, C. (1959). Note on Estimating Information. Ann. Math. Stat. 30, 71-79.

Brooks, R. J. (1982). On the Loss of Information through Censoring. Biometrika 69, 137-144.

Dobrušin, R. L. (1963). General Formulation of Shannon's Main Theorem. Amer. Math. Soc. Transl. 33, 323-428.

Fisher, R. A. (1925). Theory of Statistical Estimation. Proc. Cambridge Philos. Soc. 22, 700-725.

Khinchin, A. I. (1957). Mathematical Foundations of Information Theory. Dover, New York.

Lindley, D. L. (1956). On a Measure of Information Provided by an Experiment. Ann. Math. Stat. 27, 986-1005.

Marshall, A. and Olkin, I. (1979). Inequalities: Theory of Majorization and its Applications. Academic Press, New York.

Meilijson, I. (1972). Limiting Properties of the Mean Residual Lifetime Function. Ann. Math. Stat. 43, 354-357.

Ostrowski, A. M. (1952). Sur Quelques Applications des Convexes et Concaves au Sens de I. Schur. J. Pures Appl. 31, 253-292.

Rao, C. R. (1983). Convexity Properties of Entropy Functions and Analysis of Diversity. University of Pittsburgh Technical Report 83-11.

Schur, I. (1923). Über eine Klasse von Mittelbildungen mit Anwendungen die Dieterminaten. Theorie Sitzungsber Berlin Math. Gesellschaft 22, 9-20.

Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal 27, 379-423.

Solomon, D. L. (1979). A Comparative Approach to Species Diversity. Ecological Diversity in Theory and Practice (eds. Grassle, *et al.*). International Cooperative Publishing.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

REPORT NUMBER FSU M701 AFOSR 85-177	2. GOVT ACCESSION NO. AD-A155644	3. RECIPIENT'S CATALOG NUMBER
TITLE (and subtitle) Information in Censored Models		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
AUTHOR(s) Myles Hollander, Frank Proschan, and James Sconing		8. CONTRACT OR GRANT NUMBER(s) AFOSR F49620-85-C-0007
PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Florida State University Tallahassee, FL 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA, & WORK UNIT NUMBERS
CONTROLLING OFFICE NAME AND ADDRESS The U.S. Air Force Air Force Office of Scientific Research Bolling Air Force Base, DC 20332		12. REPORT DATE June, 1985 13. NUMBER OF PAGES 20
MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

DISTRIBUTION STATEMENT (of this report)

distribution unlimited

DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

SUPPLEMENTARY NOTES

KEY WORDS

Information, randomly censored model, majorization

ABSTRACT (Continue on reverse side if necessary and identify by block number)

Criteria are developed for measuring information in the randomly right-censored model. Measures which are appropriate include an extension of Shannon's entropy. The measures are seen to satisfy some fundamental theorems including (i) the uncensored case is always at least as informative as any censored model, (ii) information decreases as censoring increases stochastically, and (iii) the information gain is marginally decreasing.

END

FILMED

10-85

DTIC