Ann. Inst. Statist. Math.
Vol. 44, No. 3, 463–477 (1992)

INFORMATION THEORY AND THE FAILURE TIME OF A SYSTEM*

NADER EBRAHIMI

Division of Statistics, Northern Illinois University, DeKalb, IL 60115-2888, U.S.A.

(Received October 29, 1990; revised August 16, 1991)

Abstract. In this paper we introduce a measure for the rate of generation of information about the failure time of a system using mutual information measure. In the case of a system with multi-components what is really measured is the interaction between one component of the system and the rest. Our definition of measure is only a slight variation of existing classical definitions in the information theory literature. We study properties of our proposed measure and calculate information for several hypothetical systems.

Key words and phrases: Poisson shock model, white noise, entropy, Markov process, first passage time, stress-strength model.

1. Introduction

Reliability of many stochastic systems depends on several characteristics that are time dependent. Consider a system such that its reliability depends on k characteristics. Let $X_i(t)$, $i = 1, \dots, k$, denote the value of i-th characteristic at time t, and let $\bar{A}_i$ be the corresponding permissible set for the i-th characteristic at time t. For example, in engineering-systems, failure might occur when accumulated damages to the system first exceed its known breaking threshold.

Consider the hitting times

$$T_i(A_i) = \mathrm{Inf}\{t > 0 : X_i(t) \in A_i\}, \qquad i = 1, \dots, k, \tag{1.1}$$

where A_i is the complement of $\bar{A}_i$ with respect to the state space of $X_i(t)$.

The random time in (1.1) is clearly the first time when the i-th characteristic of the system is not within permissible limits. One can also reformulate $T_i(A_i)$, $i = 1, \dots, k$, in-terms of a system consisting of k components with $T_i(A_i)$ being the failure time of the i-th component, $i = 1, \dots, k$. Throughout this paper we use the second formulation.

---

* This research was partially supported by the U.S. Air Force Office of Scientific Research Grant AFSOR-89-0402.

One of the most important problems in reliability is to size the amount of information that can be obtained about any $T_i(A_i)$, failure time of the component i, from the knowledge on behaviors of other components, that is, the behaviors of the processes $X_1(t), X_2(t), \ldots, X_{i-1}(t), X_{i+1}(t), \ldots, X_k(t)$. Information about interaction between this component and the rest can help engineers to keep or redesign the component.

Ever since the fundamental work of Shannon (1948) concerning the theory of information, many papers have been devoted to various areas of this theory. In particular, a significant number of works has been done in the following directions: 1) studying the properties of the basic concepts of the theory of information and proving basic theorems; 2) calculating information content of discrete and continuous sources; 3) studying relationship between decision and estimation theory within the framework of the information theory. Among many such works we should call attention to the articles and books by Shannon (1957), Kullback (1968), Blahut (1987) and numerous references therein.

In the present work we introduce a measure for the rate of generation of information about the failure time of any component say the component i, $T_i(A_i)$ by a process $X^{(i)}(t) = (X_1(t), \ldots, X_{i-1}(t), X_{i+1}(t), \ldots, X_k(t))$. The proposed measure is an application of the "mutual information". Mutual information is well-known measure of information and it has many useful properties (see Pinsker (1964)). We give several properties of our measure. Sample calculations of the information rate for a Poisson shock model, a white noise model and bivariate processes are given. Throughout the paper we concentrate on the case when A_i is of the form $A_i = \{x : x > a_i\}$, $i = 1, \ldots, k$. Similar results can be obtained for other cases. It should be emphasized that our goal in this work is *not* to come up with a new measure but try to modify slightly the existing measures in such a way that can be used in the area of "reliability".

2. Preliminaries

In this section we summarize several well-known results that are used in the next section. Most of the results can be found in Pinsker (1964). If (Ω_X, B_X, P_X) and (Ω_Y, B_Y, P_Y) are two probability spaces induced by random variables X and Y respectively, then the product of these spaces $(\Omega_X \times \Omega_Y, B_X \times B_Y)$ is the space of points (c, d), $c \in \Omega_X$, $d \in \Omega_Y$, together with the minimal σ-field over the intervals $C \times D$ which are all pairs (c, d) with $c \in C$ and $d \in D$, and the $C \in B_X$ and $D \in B_Y$. Clearly, the pair of random variables X and Y can be interpreted in a natural way as a single random variable with values in $\Omega_X \times \Omega_Y$. We will write this new random variable as (X, Y). The distribution of the pair $P_{X,Y}(\cdot)$ is a probability measure in $\Omega_X \times \Omega_Y$ which we shall call the joint distribution of X and Y. We also define the product probability measure by $P_X \times P_Y(C \times D) = P_X(C)P_Y(D)$. The information of these random variables is defined by

$$I(X,Y) = \sup \sum_{i,j} P_{X,Y}(C_i \times D_j) \log \frac{P_{X,Y}(C_i \times D_j)}{P_X(C_i)P_Y(D_j)}, \tag{2.1}$$

where the sup is taken over all possible subsets $\{C_i\}$ of the space Ω_X and all possible subsets $\{D_j\}$ of the space Ω_Y. The number $H(X) = I(X,X) = \sup\big[-\sum_i P_X(C_i)\log P_X(C_i)\big]$ is called the entropy of the random variable X.

Let P_{X1} and P_{X2} be the probability measures defined on the Ω_X and let $\{E_i\}$ be partition of Ω_X. We define the entropy $H_{P_{X2}}(P_{X1})$ of P_{X1} with respect to P_{X2} by

$$H_{P_{X2}}(P_{X1}) = \sup\sum_i P_{X1}(E_i)\log\frac{P_{X1}(E_i)}{P_{X2}(E_i)}, \tag{2.2}$$

where the supremum is taken over all partitions of Ω_X. Obviously, $I(X,Y) = H_{P_X\times P_Y}(P_{X,Y})$.

In obtaining the value of the information, the following result due to Gelfand *et al.* (1956) plays an important role.

THEOREM 2.1. (1) *If the distribution $P_{X\times Y}$ is not absolutely continuous with respect to the distribution $P_X \times P_Y(\cdot)$, then $I(X,Y) = \infty$. (2) If the distribution $P_{X\times Y}(\cdot)$ is absolutely continuous with respect to $P_X \times P_Y$, then*

$$\begin{aligned} I(X,Y) &= \int_{\Omega_X\times\Omega_Y} (\log a_{X,Y}(x,y))P_{X,Y}(dx,dy) \\ &= \int_{\Omega_X\times\Omega_Y} i_{X,Y}(x,y)P_{X,Y}(dx,dy), \end{aligned} \tag{2.3}$$

where $a_{X,Y}(x,y)$ is the density function of $P_{X\times Y}(\cdot)$ with respect to $P_X \times P_Y(\cdot)$ and $i_{X,Y}(x,y) = \log a_{X,Y}(x,y)$.

In many cases it is easy to obtain the form of $i_{X,Y}(x,y)$. If spaces Ω_X and Ω_Y contain countably many points $x_1, x_2, \ldots,$ and $y_1, y_2, \ldots,$ then

$$i_{X,Y}(x_i,y_j) = \log\frac{P_{X,Y}(x_i,y_j)}{P_X(x_i)P_Y(y_j)} \quad \text{and}$$

$$I(X,Y) = \sum_{i,j} P_{X,Y}(x_i,y_j)\log\frac{P_{X,Y}(x_i,y_j)}{P_X(x_i)P_Y(y_j)}. \tag{2.4}$$

However, if the distributions $P_X(\cdot)$, $P_Y(\cdot)$ and $P_{X,Y}(\cdot)$ are given in-terms of densities $p_X(\cdot)$, $p_Y(\cdot)$ and $p_{X,Y}(\cdot)$ respectively, then

$$i_{X,Y}(x,y) = \log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \quad \text{and}$$

$$I(X,Y) = \int_{\Omega_X\times\Omega_Y} (p_{X,Y}(x,y))\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\mu_{\Omega_X}(dx)\mu_{\Omega_Y}(dy), \tag{2.5}$$

where $\mu_{\Omega_X}(\cdot)$, $\mu_{\Omega_Y}(\cdot)$ and $\mu_{\Omega_X}\times\mu_{\Omega_Y}(\cdot)$ are measures defined on (Ω_X, B_X), (Ω_Y, B_Y) and $(\Omega_X\times\Omega_Y, B_X\times B_Y)$ respectively. For example if X and Y are finite dimensional spaces then μ_{Ω_X} and μ_{Ω_Y} are Lebesgue measures.

Let X, Y and Z be random variables taking values in the measurable spaces (Ω_X, B_X), (Ω_Y, B_Y) and (Ω_Z, B_Z) respectively. We define

$$\bar{P}_{Y\times Z|X}(E \times F \times N) = \int_{\Omega_X} P_{Y|x}(F \mid x) P_{Z|x}(N \mid x) P_X(dx)$$

and the average conditional information of the pair Y and Z given X as

$$E(I(Y, Z \mid X)) = H_{\bar{P}_{Y\times Z|X}}(P_{X,Y,Z}), \tag{2.6}$$

where $H_{\bar{P}_{Y\times Z|X}}$ is given by (2.2).

In the case where $P_{X,Y,Z}$ are absolutely continuous with respect to the measure $\bar{P}_{Y\times Z|X}$,

$$\begin{aligned} E(I(Y, Z \mid X)) = \int_{\Omega_X} \Bigg[\int_{\Omega_Y} \int_{\Omega_Z} & p_{Y,Z|x}(y, z \mid x) \\ & \times \log \frac{p_{Y,Z|x}(y, z \mid x)}{p_{Y|x}(y \mid x) p_{Z|x}(z \mid x)} \Bigg] p_X(x) dz dy dx, \end{aligned} \tag{2.7}$$

where $p_{Y,Z|x}$ is the conditional density function of Y and Z given $X = x$, $p_{Y|x}$ is the conditional density function of Y given $X = x$, $p_{Z|x}$ is the conditional density function of Z given $X = x$, and p_X is the density function of X. Similarly, the average conditional entropy of P_Z with respect to P_X and the σ-algebra Ω_X is

$$E(H_{P_X}(P_Z \mid \Omega_X)) = H_{\bar{P}_{X(\cdot|\Omega_X)}}(P_Z),$$

where

$$\bar{P}_X(E \times F \mid \Omega_X) = \int_E (P_X(F \mid \Omega_X) P_Z(dw_z \times \Omega_X)). \tag{2.8}$$

3. Definition of the information rate

Let $X_i(U, V]$ be the random variable consisting of the family of random variables $X_i(t)$, $U < t \le V$, $i = 1, 2, \ldots, k$. Then, we have the following definition:

DEFINITION 3.1. The rate of generation of information about the failure time of any component say i $(T_i(a_i))$ by the process $X^{(i)}(t) = (X_1(t), \ldots, X_{i-1}(t), X_{i+1}(t), \ldots, X_k(t))$ is

$$\bar{I}(T_i(a_i), X^{(i)}(t)) = \lim_{V\to\infty} \frac{1}{V} I(T_i(a_i), X^{(i)}(0, V]), \tag{3.1}$$

where $I(X, Y)$ is defined by the equation (2.1), $T_i(a_i) = \text{Inf}\{t : X_i(t) > a_i\}$ and $X^{(i)}(0, V] = (X_1(0, V], \ldots, X_{i-1}(0, V], X_{i+1}(0, V], \ldots, X_k(0, V])$. The proposed measure is an application of the mutual information measure which is defined by the equation (2.1). We say $(T_i(a_i), X^{(i)}(t))$ is *information stable* if $\bar{I}(T_i(a_i), X^{(i)}(t)) = 0$.

Several comments are in order with regard to (3.1). First, in general the equation (3.1) is difficult to compute. But, if $\{X^{(i)}(t)\}$, $t = 1, 2, \ldots$ be discrete parameter process (I believe that this is true in many practical situations or at least this is the way that most processes are observed), then the equation (3.1) reduces to

$$\bar{I}(T_i(a_i), X^{(i)}(t)) = \lim_{n\to\infty} \frac{1}{n} I(T_i(a_i), X^{(i)}(1), \ldots, X^{(i)}(n)). \tag{3.2}$$

Second if $X^{(i)}(t) = X^{(i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$, $X^{(i)}(t)$ is not dynamic, then the equation (3.1) reduces to

$$\bar{I}(T_i(a_i), X^{(i)}(t)) = I(T_i(a_i), X^{(i)}). \tag{3.3}$$

If we have a system consisting of k components such that X_j, $j = 1, \ldots, k$, $j \neq i$, is the life-time of component j, then the equation (3.3) measures the interaction between life-time of component i and life-times of the remaining components.

The number $\bar{H}(T_i(a_i)) = \bar{I}(T_i(a_i), X_i(t))$, $i = 1, \ldots, k$, is called the *relative entropy* of the random variable $T_i(a_i)$.

The following theorem gives several properties of our definition. These properties are similar to properties of "mutual information".

THEOREM 3.1. a) $0 \leq \bar{I}(T_i(a_i), X^{(i)}(t)) < \infty$.

b) *If $X_i(t)$ is independent of $X_{l_1}(t), \ldots, X_{l_m}(t)$, where $A = \{X_{l_i}(t); i = 1, \ldots, m\}$ is a subset of $\{X_1(t), \ldots, X_{i-1}(t), X_{i+1}(t), \ldots, X_k(t)\}$, then*

$$\bar{I}(T_i(a_i), X^{(i)}(t)) = \bar{I}(T_i(a_i), X^{(\bar{A})}(t)),$$

where $\bar{A}$ is the complement of A with respect to $\{X_1(t), X_2(t), \ldots, X_{i-1}(t), X_{i+1}(t), \ldots, X_k(t)\}$.

c) *Suppose $X_l(t)$ is a subordinate to the random process $X_j(t)$, $1 \leq l, j \leq k$. (A random variable X is said to be subordinate to the random variable Y, if Y is everywhere dense in the random variable (X, Y). A random variable X which is a measurable function, $X = f(Y)$, of a random variable Y is obviously subordinate to Y. If $X_l = \{X_l(t); t \geq 0\}$ and $X_j = \{X_j(t); t \geq 0\}$. Then we say that the random process $X_l(t)$ is subordinate to $X_j(t)$ if the random variable X_l is subordinate to the variable X_j.) Then,*

$$I(T_i(a_i), X_j(t)) \geq I(T_i(a_i), X_l(t)).$$

d) *If $I(T_i(a_i), X_l(t)) - I(T_i(a_i), X_j(t)) < \epsilon$, then $\bar{I}(T_i(a_i), X_l(t)) - \bar{I}(T_i(a_i), X_j(t)) < \epsilon$.*

e) *$\bar{I}(T_i(a_i), C(t)X^{(i)}(t)) = \bar{I}(T_i(a_i), X^{(i)}(t))$, where $C(t)$ is some known function of t.*

PROOF. The proof of these properties follows from the corresponding properties of mutual information (see Pinsker (1964) for properties of mutual information).

Suppose $\{X^{(i)}(t)\}$, $t = 1, 2, \ldots$ be discrete parametr process. Then clearly the parameter V, which is given in the equation (3.1), is an integer. In this case we have the following theorem.

THEOREM 3.2. *If* a) *$X^{(i)}(n)$ is a Markov process and* b) *the joint conditional density function of $T_i(a_i)$ and $X^{(i)}(n)$ given $X^{(i)}(j)$, $j = 1, \ldots, n-1$, is equal to the joint conditional density function of $T_i(a_i)$ and $X^{(i)}(n)$ given $X^{(i)}(n-1)$. Then,*

$$\bar{I}(T_i(a_i), X^{(i)}(t)) = \lim_{n\to\infty} \frac{1}{n}[I(T_i(a_i)) - EH(T_i(a_i) \mid X^{(i)}(n))],$$

where $I(T_i(a_i)) = -\int f_{T_i(a_i)}(u) \log f_{T_i(a_i)}(u) du$ and $f_{T_i(a_i)}$ is the density function of $T_i(a_i)$.

PROOF. From the equation (3.2),

$$\bar{I}(T_i(a_i), X^{(i)}(t)) = \lim_{n\to\infty} \frac{1}{n} I(T_i(a_i), X^{(i)}(0, n]).$$

Now,

$$\begin{aligned} I(T_i(a_i), &X^{(i)}(0,n]) \\ &= I(T_i(a_i), (X^{(i)}(1), \ldots, X^{(i)}(n))) \\ &= -\int f_{T_i(a_i)}(u) \log f_{T_i(a_i)}(u) du + E(I(T_i(a_i) \mid X^{(i)}(1))) \\ &\quad + \sum_{j=1}^{n-1} E(I(T_i(a_i), X^{(i)}(j+1) \mid X^{(i)}(1), \ldots, X^{(i)}(j))). \end{aligned} \tag{3.4}$$

Let us consider the l-th term on the right side of (3.4). Using the definition of conditional information, conditional entropy, and assumptions a and b,

$$\begin{aligned} E(I(T_i(a_i), &X^{(i)}(l+1) \mid X^{(i)}(1), \ldots, X^{(i)}(l))) \\ &= -E(H(T_i(a_i) \mid X^{(i)}(l+1))) + E(H(T_i(a_i) \mid X^{(i)}(l))). \end{aligned} \tag{3.5}$$

Combining equations (3.4) and (3.5) we get the result.

THEOREM 3.3. *Let $\{X_i(t)\}$ be a discrete time process, $t = 1, 2, \ldots$, with values in the measurable space (Ω_{X_i}, B_{X_i}) and let $X_i(1), X_i(2), \ldots$ be independent.*

Then,

$$
\begin{aligned}
(3.6)\quad & \bar{H}(T_i(a_i)) \\
&= -\lim_{n\to\infty}\frac{1}{n}\Bigg\{\bar{F}_{i1}(a_i)\log\bar{F}_{i1}(a_i) \\
&\qquad + \left(\sum_{l=2}^{\infty}F_{i1}(a_i)\cdots F_{i,l-1}(a_i)\bar{F}_{i,l}(a_i)\right)\log F_{i1}(a_i) \\
&\qquad + \sum_{j=1}^{n-1}\Bigg[(\bar{F}_{i,j+1}(a_i)F_{ij}(a_i)\cdots F_{i1}(a_i))\log\bar{F}_{i,j+1}(a_i) \\
&\qquad + \sum_{l=j+2}^{\infty}F_{i,j+1}(a_i)\cdots F_{i,l-1}(a_i)\bar{F}_{il}(a_i)\log F_{i,j+1}(a_i)\Bigg]\Bigg\},
\end{aligned}
$$

where $\bar{F}_{ij}(a_i) = P(X_i(j) > a_i)$ *and* $F_{ij}(a_i) = P(X_i(j) \le a_i)$.

PROOF. We will prove this theorem for the case that $X_i(t)$ has a finite state space. Similar arguments can be used for other cases.

We consider the relative entropy of $T_i(a_i)$ defined by

$$\bar{H}(T_i(a_i)) = \lim_{n\to\infty}\frac{1}{n}I(T_i(a_i),(X_i(1),\dots,X_i(n))),$$

where $I(T_i(a_i),(X_i(1),\dots,X_i(n))) = I(T_i(a_i),X_i(1)) + \sum_{j=1}^{n-1}E(I(T_i(a_i),X_i(j+1)\mid X_i(1),\dots,X_i(j)))$. Now,

$$
\begin{aligned}
(3.7)\quad & I(T_i(a_i),X_i(1)) \\
&= \sum_{l_1\in\Omega_{X_i(1)}}\Bigg[P(T_i(a_i)=1,X_i(1)=l_1) \\
&\qquad \times\log\frac{P(T_i(a_i)=1,X_i(1)=l_1)}{P(T_i(a_i)=1)P(X_i(1)=l_1)} \\
&\qquad + \sum_{n=2}^{\infty}P(T_i(a_i)=n,X_i(1)=l_1) \\
&\qquad \times\log\frac{P(T_i(a_i)=n,X_i(1)=l_1)}{P(T_i(a_i)=n)P(X_i(1)=l_1)}\Bigg] \\
&= -\Bigg\{[\bar{F}_{i1}(a_i)\log\bar{F}_{i1}(a_i)] \\
&\qquad + \left[\sum_{n=2}^{\infty}F_{i1}(a_i)F_{i2}(a_i)\cdots F_{i,n-1}(a_i)\bar{F}_{i,n}(a_i)\right]\log F_{i1}(a_i)\Bigg\}.
\end{aligned}
$$

Also, let us consider the l-th term on the right side of the equation (3.4), (second part),

$$\begin{aligned}
&E(I(T_i(a_i), X_i(l+1) \mid X_i(1), \ldots, X_i(l))) \\
&\quad = \sum \Bigg\{ \sum_{n=l+1}^{\infty} P(T_i(a_i) = n, X_i(j) = x(j); j = 1, \ldots, l+1) \\
&\qquad \times [\log(P(T_i(a_i) = n, X_i(l+1) = x(l+1) \mid \\
&\qquad X_i(1) = x(1), \ldots, X_i(l) = x(l))/ \\
&\qquad P(T_i(a_i) = n \mid X_i(1) = x(1), \ldots, X_i(l) = x(l)) \\
&\qquad \times P(X_i(l+1) = x(l+1) \mid X_i(1) = x(1), \ldots, X_i(l) = x(l)))] \Bigg\},
\end{aligned}$$

where the summation is taken over $\Omega_{X_i(l+1)} \times \Omega_{X_i(l)} \times \cdots \times \Omega_{X_i(1)}$,

$$\begin{aligned}
(3.8) \quad = -\Bigg\{ &\bar{F}_{i,l+1}(a_i) F_{il}(a_i) \cdots F_{i1}(a_i) \log \bar{F}_{i,l+1}(a_i) + (\log F_{i,l+1}(a_i)) \\
&\times \left(\sum_{n=l+2}^{\infty} F_{i,l+1}(a_i) \cdots F_{i,n-1}(a_i) \bar{F}_{in}(a_i) \right) \Bigg\}.
\end{aligned}$$

Combining (3.7) and (3.8) we get that

$$\begin{aligned}
(3.9) \quad &I(T_i(a_i), X_i(1), \ldots, X_i(n)) \\
&= -\Bigg\{ \Bigg[\bar{F}_{i1}(a_i) \log \bar{F}_{i1}(a_i) \\
&\quad + \left[\sum_{n=2}^{\infty} F_{i1}(a_i) \cdots F_{i,n-1}(a_i) \bar{F}_{i,n}(a_i) \right] \log F_{i1}(a_i) \Bigg] \\
&\quad + \sum_{j=1}^{n-1} \Bigg\{ [(\bar{F}_{i,j+1}(a_i) F_{ij}(a_i) \cdots F_{i1}(a_i)) \log \bar{F}_{i,j+1}(a_i)] \\
&\quad + \left(\sum_{l=j+2}^{\infty} F_{i,j+1}(a_i) \cdots F_{i,l-1}(a_i) \bar{F}_{il}(a_i) \right) \log F_{i,j+1}(a_i) \Bigg\} \Bigg\}.
\end{aligned}$$

COROLLARY 3.1. *Suppose* $F_{ij}(t) = P(X_i(j) \le t) = F(t)$ *and* $0 < F(a_i) < 1$. *Then,*

$$\bar{H}(T_i(a_i)) = -F(a_i) \log F(a_i).$$

THEOREM 3.4. *Let* $X^{(i)}(t)$ *be a discrete time process,* $t = 1, 2, \ldots$ *with values in the measurable space* $(\Omega_{X^{(i)}}, B_{X^{(i)}})$ *and let* $X^{(i)}(1), \ldots, X^{(i)}(n), \ldots$ *be conditionally independent given* $T_i(a_i)$. *Then,*

$$\bar{I}(T_i(a_i), X^{(i)}(t)) = \lim_{n \to \infty} \frac{1}{n} \left[\sum_{j=1}^{n} I(T_i(a_i), X^{(i)}(j)) + H(X^{(i)}(1), \ldots, X^{(i)}(n)) \right].$$

PROOF. Use the fact that for a sequence of random variables X, $Y_1, \dots, Y_n, \dots$ such that $Y_1, Y_2, \dots, Y_n$ are conditionally independent given X, we have

$$E(I(X,(Y_1,\dots,Y_n))) = \sum_{i=1}^{n} I(X,Y_i) + H(Y_1,\dots,Y_n).$$

4. Examples

Example 1. A device is subjected to shocks occurring randomly in time according to a homogeneous Poisson process (HPP) with intensity λ (for the definition of HPP see Barlow and Proschan (1981)). The i-th shock causes a random amount Y_i of damage, where $Y_1, Y_2, \dots$ are independently distributed with common distribution function F. The device fails when the total accumulated damages exceed a specified threshold, say a, that is the failure time of the device is $T(a) = \mathrm{Inf}\{t : \sum_{i=1}^{N(t)} Y_i > a\}$.

Suppose that the counting process $\{N(t)\}$ was observed at discrete time points, $t = 1, 2, \dots$, $(N(0) = 0)$, and our goal is to obtain the amount of relative information about $T(a)$ by the process $\{N(t); t = 1, 2, \dots\}$.

THEOREM 4.1. *The relative information about* $T(a)$ *by* $\{N(t); t = 1, 2, \dots\}$ *is* $\bar{I}(T(a), N(t)) = b$, *where*

$$b = \lim_{i,n,n-i\to\infty} \left\{ \sum_{h_1=0}^{\infty} \cdots \sum_{h_n=0}^{\infty} \left(\int_0^1 \sum_{k=0}^{h_i} F^{(k+h_1+\cdots+h_{i-1})}(a) \binom{h_i}{k} g(u) \right) \right.$$
$$\left. \times \left(\log \sum_{k=0}^{h_i} F^{(k+h_1+\cdots+h_{i-1})}(a) \binom{h_i}{k} g(u) \right) du \frac{e^{-\lambda n}\lambda^{h_1+\cdots+h_n}}{h_1!\cdots h_n!} \right\},$$

$g(u) = (h_i - k)(u)^{h_i-k-1}(1-u)^k - k(u)^{h_i-k}(1-u)^{k-1}$ *and* $F^{(m)}$ *is the m-fold convolution of F with itself. There are situations that b can not be computed. In those situations we recommend to use appropriate approximations.*

PROOF. By the definition (3.2) we get

$$\bar{I}(T(a), N(t)) = \lim_{n\to\infty} \frac{1}{n} I(T(a), (N(1), N(2), \dots, N(n))),$$

where

$$\begin{aligned} (4.1)\quad & I(T(a),(N(1),\dots,N(n))) \\ &= I(T(a),N(1)) \\ &\quad + \sum_{j=1}^{n-1} E[I(T(a),N(j+1) \mid N(1),\dots,N(j))] \\ &= -\int_0^{\infty} f_{T(a)}(t)\log f_{T(a)}(t)dt - E(H(T(a) \mid N(n),\dots,N(1))). \end{aligned}$$

Now,

$$\begin{aligned}
&P(T(a) > t \mid N(i) = l_i, i = 1, \ldots, n) \\
&= P\left(\sum_{i=1}^{N(t)} Y_i \le a \mid N(i) = l_i, i = 1, \ldots, n\right) \\
&= \sum_{k=0}^{\infty} F^{(k)}(a) P(N(t) = k \mid N(i) = l_i, i = 1, \ldots, n) \\
&= \sum_{k=0}^{h_i} F^{(k+l_{i-1})}(a) \binom{h_i}{k} (t - (i-1))^k (i-t)^{h_i - k} \quad \text{if } i - 1 \le t \le i \\
&(h_i = l_i - l_{i-1}, l_0 = 0, i = 1, \ldots, n) \\
&= \sum_{k=0}^{\infty} (F^{(k+l_n)}(a)(\lambda(t-n))^k \exp(-\lambda(t-n))) \frac{1}{k!} \quad \text{if } t > n.
\end{aligned}$$

(It is assumed that $(0)^0 = 1$.) Consequently the density function of $T(a)$ given $N(i) = l_i$, $i = 1, \ldots, n$, is

$$\text{(4.2)} \quad f(t \mid N(i) = l_i, i = 1, \ldots, n)$$
$$= \begin{cases} \displaystyle\sum_{k=0}^{h_1} F^{(k+l_{i-1})}(a) \binom{h_i}{k} \\ \quad \times [(h_i - k)(i-t)^{h_i-k-1}(t-i+1)^k \\ \quad\quad -k(i-t)^{h_i-k}(t-i+1)^{k-1}], & \text{if } i - 1 \le t \le i \\ \displaystyle\sum_{k=0}^{\infty} F^{(k+l_n)}(a) \frac{1}{k!} \\ \quad \times [\lambda^{k+1}(t-n)^k \exp(-\lambda(t-n)) \\ \quad\quad -\lambda^k k (t-n)^{k-1} \exp(-\lambda(t-n))], & \text{if } t > n. \end{cases}$$

Using the equation (4.2) the second term in the equation (4.1) can be written as

$$\begin{aligned}
\text{(4.3)} \quad &- E(H(T(a) \mid N(i) = l_i, i = 1, \ldots, n)) \\
&= \left\{ \sum_{h_1=0}^{\infty} \sum_{h_2=0}^{\infty} \cdots \sum_{h_n=0}^{\infty} \left[\int_0^{\infty} (f(t \mid N(1) = h_1, N(2) \right.\right. \\
&\quad - N(1) = h_2, \ldots, N(n) - N(n-1) = h_n)) \\
&\quad \left. \times \log(f(t \mid N(1) = h_1, \ldots, N(n) - N(n-1) = h_n)) dt \right] \\
&\quad \left. \times ((\exp(-n\lambda))\lambda^{h_1 + h_2 + \cdots + h_n}) \frac{1}{h_1! h_2! \cdots h_n!} \right\}
\end{aligned}$$

$$
\begin{aligned}
&= \Bigg[\sum_{h_1=0}^{\infty} \cdots \sum_{h_n=0}^{\infty} \Bigg\{ \sum_{i=1}^{n} \Bigg[\int_{i-1}^{i} \Bigg(\sum_{k=0}^{h_i} F^{(k+h_1+\cdots+h_{i-1})}(a) \binom{h_i}{k} \\
&\quad \times ((h_i-k)(i-t)^{h_i-k-1} \times (t-i+1)^k - k(i-t)^{h_i-k}(t-i+1)^{k-1}) \Bigg) \\
&\quad \times \Bigg(\log \sum_{k=0}^{h_i} F^{(k+h_1+\cdots+h_{i-1})}(a) \binom{h_i}{k} ((h_i-k)(i-t)^{h_i-k-1}(t-i+1)^k \\
&\quad - k(i-t)^{h_i-k}(t-i+1)^{k-1}) \Bigg) \Bigg] \Bigg\} \frac{\exp(-n\lambda)}{h_1! \cdots h_n!} \lambda^{h_1+\cdots+h_n} \Bigg] \\
&\quad + \Bigg\{ \sum_{h_1=0}^{\infty} \cdots \sum_{h_n=0}^{\infty} \Bigg\{ \int_n^{\infty} \sum_{k=0}^{\infty} F^{(k+h_1+\cdots+h_n)}(a) \frac{1}{k!} \\
&\quad \times [\lambda^{k+1}(t-n)^k \exp(-\lambda(t-n)) - \lambda^k k(t-n)^{k-1} \exp(-\lambda(t-n))] \\
&\quad \times \Bigg[\log \sum_{k=0}^{\infty} F^{(k+h_1+\cdots+h_n)}(a) \\
&\quad \times \frac{1}{k!} [\lambda^{k+1}(t-n)^k \exp(-\lambda(t-n)) - \lambda^k k(t-n)^{k-1} \\
&\quad \times \exp(-\lambda(t-n))] \Bigg] dt \Bigg\} \frac{e^{-n\lambda} \lambda^{h_1+\cdots+h_n}}{h_1! \cdots h_n!} \Bigg\}.
\end{aligned}
$$

From (4.3) it is clear that

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{-1}{n} E(H(T(a) \mid N(i) = l_i, i = 1, \ldots, n)) \\
&= \lim_{n\to\infty} \frac{1}{n} \Bigg[\sum_{h_1=0}^{\infty} \cdots \sum_{h_n=0}^{\infty} \Bigg\{ \sum_{i=1}^{n} \Bigg[\int_{i-1}^{i} \Bigg(\sum_{k=0}^{h_i} F^{(k+h_1+\cdots+h_{i-1})}(a) \binom{h_i}{k} \\
&\quad \times ((h_i-k)(i-t)^{h_i-k-1}(t-i+1)^k - k(i-t)^{h_i-k}(t-i+1)^{k-1}) \Bigg) \\
&\quad \times \Bigg(\log \sum_{k=0}^{h_i} F^{(k+h_1+\cdots+h_{i-1})}(a) \binom{h_i}{k} ((h_i-k)(i-t)^{h_i-k-1}(t-i+1)^k \\
&\quad - k(i-t)^{h_i-k}(t-i+1)^{k-1}) \Bigg) \Bigg] dt \Bigg\} (\exp(-n\lambda)) \frac{\lambda^{h_1+\cdots+h_n}}{h_1! \cdots h_n!} \Bigg] = b.
\end{aligned}
$$

(Note that the second term in equation (4.3) goes to zero as $n \to \infty$. We use the fact if $a_1, a_2, \ldots$ be a sequence of non-negative numbers such that $a_n = \sum_{i=1}^{n} a_{ni}$, where the a_{ni}, $i = 1, \ldots, n$; $n = 1, 2, \ldots$ are non-negative and bounded such that $\lim_{i,n,n-i\to\infty} a_{ni} = a$, then $\lim_{n\to\infty}(a_n/n) = a$.) The first term in the equation (4.1) can be written as

$$
-\int_0^{\infty} f_{T(a)}(t) \log f_{T(a)}(t) dt
$$

$$= -\left\{\left[\int_0^\infty \sum_{k=0}^\infty \left[F^{(k)}(a)\frac{\lambda e^{-\lambda t}(\lambda t)^k - \lambda^k t^{k-1} k e^{-\lambda t}}{k!}\right.\right.\right.$$
$$\left.\left.\left.\times \log \sum_{k=0}^\infty F^{(k)}(a)\frac{\lambda e^{-\lambda t}(\lambda t)^k - \lambda^k t^{k-1} k e^{-\lambda t}}{k!}\right]\right] dt\right\}.$$

As $n \to \infty$, $-(1/n)\int_0^\infty f_{T(a)}(t)\log f_{T(a)}(t)dt \to 0$. Therefore,

$$\bar{I}(T(a), N(t)) = \lim_{n\to\infty} -\frac{1}{n}E(H(T(a) \mid N(i) = l_i, i = 1, \ldots, n)) = b.$$

COROLLARY 4.1. *Suppose* $F^{(j)}(a) = 1$, $j = 0, 1$ *and* $= 0$ *for* $j > 1$. *Then* $\bar{I}(T(a), N(t)) = 0$, *that is,* $T(a)$ *and* $N(t)$ *are information stable.*

Corollary 4.1 simply tells that for the system which satisfies the assumptions of this corollary, observing only $N(t)$ at discrete time points does not give any information about the failure time of the system.

Example 2. Consider a white noise process $X(t)$, $t = 0, 1, 2, \ldots$ such that $X(t)$ is normally distributed with mean zero and variance 1. (The process $X(t)$ is said to be a white noise process if $X(0), X(1), \ldots$ are independent and have common distribution function.) From Corollary 3.1, $\bar{H}(T(a)) = -\Phi(a)\log\Phi(a)$, where $\Phi(a)$ is the cumulative distribution function of a standard normal distribution. Table 1 gives $\bar{H}(T(a))$ for several values of a. From Table 1 it is clear that we have maximum relative entropy whenever we want to hit 0. That is, given that $X(t)$ is observed at discrete time points we have maximum relative uncertainty about a system failure whose failure time is $T(0)$ and minimum relative uncertainty about a system whose failure time is $T(-3)$.

Table 1.

a	$\bar{H}(T(a))$
−3	.0092
−2	.086
−1	.29
0	.35
1	.15
2	.02
3	.0013

Example 3. Consider a bivariate process $\{(X_n, Y_n); n \geq 1\}$ such that $(X_1, Y_1), (X_2, Y_2), \ldots$ are independent with common joint distribution function $H(x, y) = P(X_i \leq x, Y_i \leq y)$. Let $F(x) = P(X_i \leq x)$, $G(y) = P(Y_i < y)$, $h(x, y)$ be the joint density function of X_i and Y_i, $\bar{W}(x, y) = \int_x^\infty h(z, y)dz$, $W(x, y) = \int_{-\infty}^x h(z, y)dz$, $\bar{U}(x, y) = P(X > x \mid Y = y)$ and $U(x, y) = P(X \leq x \mid Y = y)$. Define $T(x) = \mathrm{Inf}\{n \geq 1 : X_n > x\}$. Then we have the following theorem.

THEOREM 4.2.

$$\bar{I}(T(x), \{Y_n\}) = 0,$$

that is $T(x)$ *and* $\{Y_n\}$ *are stable.*

PROOF. From Definition (3.1) we get that

$$\begin{aligned}
\bar{I}(T(x), \{Y_n\}) &= \lim_{n\to\infty} \frac{1}{n} I(T(x), (Y_1, \dots, Y_n)) \\
&= \lim_{n\to\infty} \frac{1}{n}\Bigg[-\sum_{j=1}^{\infty} P(T(x) = j) \\
&\qquad \times \log(P(T(x) = j)) - E(H(T(x) \mid Y_1, \dots, Y_n))\Bigg].
\end{aligned} \tag{4.4}$$

The first term in the equation (4.4) can be written as

$$\begin{aligned}
&-\sum_{j=1}^{\infty} P(T(x) = j) \log P(T(x) = j) \\
&= -\left[\sum_{j=1}^{\infty} F^{j-1}(x)\bar{F}(x) \log(F^{j-1}(x)\bar{F}(x))\right] \\
&= -\log \bar{F}(x) - \frac{F(x)}{\bar{F}(x)} \log F(x) = -\frac{\bar{F}(x)\log\bar{F}(x) + F(x)\log F(x)}{\bar{F}(x)}.
\end{aligned} \tag{4.5}$$

The second term in the equation (4.4) can be written as

$$\begin{aligned}
&E(H(T(x) \mid Y_1, \dots, Y_n)) \\
&= \int_{(y_1,\dots,y_n)} \sum_{j=1}^{n} P(T(x) = j, Y_1 = y_1, \dots, Y_n = y_n) \\
&\qquad \times \log P(T(x) = j \mid Y_1 = y_1, \dots, y_n = y_n) dy_1 \cdots dy_n \\
&= \sum_{j=1}^{n} \Bigg\{ F^{j-1}(x) \left(\int_{-\infty}^{\infty} \bar{W}(x,y) \log \bar{U}(x,y) dy \right) \\
&\qquad\qquad + (j-1)F^{j-2}(x)\bar{F}(x) \int_{-\infty}^{\infty} W(x,y) \log U(x,y) dy \Bigg\} \\
&\quad + \sum_{j=n+1}^{\infty} \Bigg\{ (n\bar{F}(x)F^{j-1}(x)) \left(\int_{-\infty}^{\infty} W(x,y) \log U(x,y) dy \right) \\
&\qquad\qquad + (j-n-1)F^{j-1}(x)\bar{F}(x) \log F(x) + F^{j-1}(x)\bar{F}(x) \log \bar{F}(x) \Bigg\}.
\end{aligned}$$

Combining (4.4) and (4.5) we get that

$$\begin{aligned}
&\bar{I}(T(x), \{Y_n\}) \\
&= \lim_{n\to\infty} -\frac{1}{n}\left\{ \frac{\bar{F}(x)\log\bar{F}(x) + F(x)\log F(x)}{\bar{F}(x)} \right. \\
&\qquad + \frac{\int_{-\infty}^{\infty} \bar{W}(x,y)\log\bar{U}(x,y)dy}{\bar{F}(x)} \\
&\qquad + \frac{F(x)}{\bar{F}(x)}\int_{-\infty}^{\infty} W(x,y)\log U(n,y)dy \\
&\qquad + nF^n(x)\int_{-\infty}^{\infty} W(x,y)\log U(x,y)dy \\
&\qquad \left. + \frac{F^{n+2}(x)}{\bar{F}(x)}\log F(x) + F^n(x)\log\bar{F}(x) \right\} = 0.
\end{aligned}$$

One application of Theorem 4.2 is that if we have bivariate random vector, where X has the cumulative distribution function F and Y has the cumulative distribution function Y. If we take a sequence of random sample from F and a sequence of random sample from G, then even though X and Y are dependent but still the rate of information about $T(x)$, the first record value larger than x, by $\{Y_n\}$ is 0.

Example 4. Let $X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ be mutually independent random variables, the X_i having the common distribution F and Y_i having the common distribution G. Let $T(0) = \text{Inf}\{n : X_i \le Y_i\}$. Then we have the following result:

THEOREM 4.3. (a) $\bar{I}(T(0), \{Y(n)\}) = -w$, (b) $\bar{I}(T(0), \{X(n)\}) = -a$, *where* $w = \int \bar{F}(y)(\log\bar{F}(y))dG(y)$ *and* $a = \int (G(y)\log G(y))dF(y)$.

PROOF. We will prove part (a). Similar arguments can be used to prove part (b).

Using the definition (3.1) we get that

$$\bar{I}(T(0), \{Y(n)\}) = \lim_{n\to\infty} -\frac{1}{n}\left\{ \sum_{j=1}^{\infty} P(T(0)=j)\log P(T(0)=j) + E[H(T(0) \mid Y_1, \dots, Y_n)] \right\}. \tag{4.6}$$

Given that $R = P(X_i \le Y_i) = \int \bar{G}(x)dF(x)$, $\bar{G} = 1 - G$, $i = 1, 2, \dots,$ the first term in the equation (4.6) can be written as

$$\sum_{j=1}^{\infty} P(T(0)=j)\log P(T(0)=j) \tag{4.7}$$

$$
\begin{aligned}
&= \sum_{j=1}^{\infty} P(X_1 > Y_1, \dots, X_{j-1} > Y_{j-1}, X_j \le Y_j) \\
&\quad \times \log(P(X_1 > Y_1, \dots, X_{j-1} > Y_{j-1}, X_j \le Y_j)) \\
&= \sum_{j=1}^{\infty} (1-R)^{j-1} R \log(1-R)^{j-1} R \\
&= \frac{1}{R}(R \log R + (1-R)\log(1-R)).
\end{aligned}
$$

The second term in the equation (4.6) can be written as

$$
\begin{aligned}
(4.8)\quad & E(H(T(0) \mid Y_1, \dots, Y_n)) \\
&= \frac{1}{R}[w+u] + \sum_{j=n+1}^{\infty} nR(1-R)^{j-n-1} w \\
&\quad + \sum_{j=n+1}^{\infty} R(1-R)^{j-1} \log R \\
&\quad + \sum_{j=n+1}^{\infty} (j-n-1) R(1-R)^{j-1} \log(1-R) \\
&= \frac{1}{R}(w+u) + nw + (1-R)^n \log R + (1-R)^{n+1}(\log(1-R))\frac{1}{R}.
\end{aligned}
$$

Here $u = \int_0^\infty (F(y) \log F(y)) dG(y)$. Combining (4.7) and (4.8) we get that

$$
\bar{I}(T(0), \{Y_n\}) = \lim_{n \to \infty} -\frac{1}{n} I(T(0), (Y_1, \dots, Y_n)) = -w.
$$

An application of Theorem 4.3 in reliability is that one can take X_i to be the strength of a system which is subjected to the stress Y_i at discrete time point i, then $T(0)$ is failure of the system. Theorem 4.3 says that the rate of information about failure time of the system from its strength and the stress are $-a$ and $-w$ respectively.

References

Barlow, R. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models*, To begin with, Silver Spring, MD.

Blahut, R. E. (1987). *Principles and Practice of Information Theory*, Addison-Wesley, Massachusetts.

Gelfand, I. M., Kolmogorov, A. N. and Jaglom, A. M. (1956). On the general definitions of an amount of information, *Dokl. Akad. Nauk SSSR*, **111**, 745–748.

Kullback, S. (1968). *Information Theory and Statistics*, Wiley, New York.

Pinsker, M. S. (1964). *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco.

Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, **27**, 279–423, 623–656.

Shannon, C. E. (1957). Certain results in coding theory for noisy channels, *Inform. and Control*, **1**, 6–25.