NOTE ON EXTREME VALUES, COMPETING RISKS AND SEMI-MARKOV PROCESSES[1]

By Simeon M. Berman

Columbia University

1. Introduction. Let $X_1, X_2, \cdots, X_n$ be independent random variables with continuous distribution functions (d.f.'s); let $Z = \min(X_1, X_2, \cdots, X_n)$, and let N be the (random) index k for which $Z = X_k$. If X_i, $i = 1, 2, \cdots, n$, have a common d.f. $F(x)$, then it is uniquely determined by the d.f. of Z; in fact, Z has the d.f. $P\{Z \leqq x\} = 1 - (1 - F(x))^n$. This note deals with the case where X_i has the d.f. $F_i(x)$, $i = 1, 2, \cdots, n$, where the F_i are not all the same. We show that the joint d.f. of N and Z uniquely determines $F_i(x)$, $i = 1, 2, \cdots, n$, and we give an explicit formula for it. This problem comes directly from a proposed model for the theory of the probability of the death of an individual from one of several competing causes; this model is considered in Section 3. Our main result is also applied to the theory of semi-Markov processes in Section 4.

2. The fundamental equations. The joint d.f. of (N, Z) is specified by the monotonic functions

$$H_k(x) = P\{N = k, Z \leqq x\}, \qquad k = 1, 2, \cdots, n.$$

Theorem. *The set of functions* $\{H_k(x)\}$ *is related to the set* $\{F_j(x)\}$ *by the functional equations*

$$H_k(x) = \int_{-\infty}^{x} \prod_{j \neq k} (1 - F_j(t))\, dF_k(t), \qquad k = 1, 2, \cdots, n. \tag{1}$$

The solution of this set of equations is

$$F_k(x) = 1 - \exp\left\{-\int_{-\infty}^{x} [1 - \sum_{j=1}^{n} H_j(t)]^{-1}\, dH_k(t)\right\}, \qquad k = 1, 2, \cdots, n. \tag{2}$$

Proof. The probability that X_k is the smallest value and that it assumes a value not greater than x is the integral over t in $(-\infty, x]$ of the probability that X_i, $i \neq k$, is greater than $t + dt$, while X_k falls in the infinitesimal interval $[t, t + dt]$. This proves (1).

One can verify that (2) is in fact the solution of (1), but we shall sketch the derivation of the solution. First we notice that

$$1 - \sum_{j=1}^{n} H_j(x) = \prod_{j=1}^{n} [1 - F_j(x)]; \tag{3}$$

Received September 20, 1962; revised April 16, 1963.

[1] Research supported by National Science Foundation Grant NSF-G14146 at Columbia University.

each side, in fact, is the probability that Z exceeds x. We solve (1) in the domain $\{F_j < 1, j = 1, 2, \cdots, n\}$. Inserting (3) in (1), we have

$$H_k(x) = \int_{-\infty}^{x} \prod_{j=1}^{n} [1 - F_j(t)][1 - F_k(t)]^{-1} \, dF_k(t)$$
$$= -\int_{-\infty}^{x} \left[1 - \sum_{j=1}^{n} H_j(t)\right] d\{\log [1 - F_k(t)]\}.$$

Conversion to a differential equation yields $dH_k(x) = -[1 - \sum_{j=1}^{n} H_j(x)] \, d\{\log [1 - F_k(x)]\}$; (2) is the solution of this elementary equation.

3. Competing causes of death. An individual is exposed to several potential causes of death during his lifetime. Stochastic models have long been used to describe the "strength" of the various causes. Recent work on this problem, as well as a brief bibliography, is contained in [1]. Here we propose a general model based on our theorem of Section 2.

Let there be a finite number of causes of death, labelled $1, 2, \cdots, n$. We associate with cause k a nonnegative random variable X_k with a continuous d.f. $F_k(x)$, $k = 1, 2, \cdots, n$. The length of life of an individual and the cause of his death are determined by the following random experiment. The independent random variables X_k, $k = 1, 2, \cdots, n$ are observed; the age at death is $Z = \min (X_1, X_2, \cdots, X_n)$ and the cause of death has the label N given by the index k for which $Z = X_k$. We are assuming that the ages at which different causes strike are stochastically independent. The random variable X_k represents the age at death if cause k were the only cause of death, and measures the "absolute" potency of cause k. For each individual, N and Z are observable but not X_k. The joint distribution of (N, Z), given by $\{H_k(x)\}$, is usually known from actuarial data; the distribution of X_k can be computed from (2).

The quantity

$$\lambda_k(x) = d[\log (1 - F_k(x))]/\sum_{j=1}^{n} d[\log (1 - F_j(x))] \tag{4}$$

is the ratio of the conditional probability that an individual alive at age x will die from cause k in the age interval $[x, x + dx]$ to the conditional probability that he will die in $[x, x + dx]$ from any cause given his survival up to age x. In the classical model, $(d/dx)[-\log (1 - F_k(x))]$ is called the "force of mortality due to cause k" and is assumed to be constant in x, so that $F_k(x)$ is exponential. Chiang [1] supposes only that $\lambda_k(x)$ is constant over intervals of unit length. Our model puts no restriction on the variation of $\lambda_k(x)$.

4. Application to semi-Markov processes. Let $(1, 2, \cdots, n)$ denote the possible states of a physical system and $X(t)$ the state at time t, $t \geqq 0$. Suppose that $X(0) = i$. We associate with state i a set of nonnegative, independent random variables T_{ij}, $j = 1, \cdots, n$, with continuous d.f.'s $F_{ij}(x)$, $j = 1, \cdots, n$. If $Z_i = \min_j T_{ij}$ and N_i is the index k for which $Z_i = T_{ik}$, then the process moves from state i to state k, and this occurs in a time interval of length Z_i, i.e. $X(t) = i$ for $0 \leqq t < Z_i$ and $X(Z_i) = k$. In a similar way, we associate with state k the random variables $\{T_{jk}\}$, and the process moves to state N_k at time $Z_i + Z_k$, etc.

Using our theorem, we can show that $X(t)$ is a semi-Markov process on $(1, 2, \cdots, n)$ with transition functions $H_{ik}(x) = 0$, $x \leqq 0$, $H_{ik}(x) = \int_0^x \prod_{j \neq k}[1 - F_{ij}(t)]\, dF_{ik}(t)$, $x > 0$, $i, k = 1, \cdots, n$, where $H_{ik}(x)$ is the probability that $X(\cdot)$ waits no more than x time units in state i and moves directly to state k. (We refer to [2] for definitions and elementary properties of semi-Markov processes.) It also follows from our theorem that *every* set of continuous transition functions has this representation.

The idea in this section seems to be implicit in recently published work in reliability theory.

REFERENCES

[1] Chiang, Chin Long (1961). On the probability of death from specific causes. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **4** 169–180. Univ. of California Press.

[2] Smith, Walter L. (1958). Renewal theory and its ramifications. *J. Roy. Statist. Soc. Ser. B* **20** 243–302.