RANDOM DISTRIBUTION FUNCTIONS

BY LESTER E. DUBINS AND DAVID A. FREEDMAN

Communicated by J. L. Doob, March 14, 1963

1. Introduction. A random distribution function $F$ is a measurable map from a probability space $(\Omega, \mathcal{F}, Q)$ to the space $\Delta$ of distribution functions on the closed unit interval $I$, where $\Delta$ is endowed with its natural Borel $\sigma$-field, that is, the smallest $\sigma$-field containing the customary weak* topology. It determines a prior probability measure $P = QF^{-1}$ in the space $\Delta$. Of course, $F$ is essentially the same as the stochastic process $\{F_t, 0 \leq t \leq 1\}$ on $(\Omega, \mathcal{F}, Q)$, where $F_t(\omega) = F(\omega)(t)$. Therefore, this note can be thought of as dealing with a certain class of random distribution functions, or a class of stochastic processes, or a class of prior probabilities.

Which class? Practically any base probability $\mu$ on the Borel subsets of the unit square $S$ determines a random distribution function $F$ and so a prior probability $P_\mu$ in $\Delta$, which will be described somewhat informally in §2, by explaining how to select a value of $F$, i.e., a distribution function $F$, at random. §§3, 4 and 5 describe some properties of $P_\mu$. Proofs will be given elsewhere. For ease of exposition, we assume that $\mu$ concentrates on, that is, assigns probability 1 to, the interior of $S$.

2. The construction. To select a value $F$ of $F$ at random, begin by selecting a point $(x, y)$ from the interior of $S$ according to $\mu$. The horizontal and vertical lines through $(x, y)$ divide $S$ into four rectangles; consider the closed lower left rectangle $L$ and the upper right one $R$. The unique (affine) transformation of the form $(u, v) \rightarrow (\alpha u + \beta, \gamma v + \delta)$, $\alpha$ and $\gamma$ positive, which maps $S$ onto $L$ carries $\mu$ into a probability $\mu_L$ concentrated on $L$. The probability $\mu_R$ is defined in a similar way. Now select a point $(x_L, y_L)$ at random from the interior of $L$ according to $\mu_L$, and a point $(x_R, y_R)$ at random from the interior of $R$ according to $\mu_R$. As before, $(x_L, y_L)$ determines four subrectangles of $L$, and $(x_R, y_R)$ determines four subrectangles of $R$. Consider the lower left subrectangle $LL$ in $L$, the upper right subrectangle $RL$ in $L$, and the two analogous subrectangles $LR$ and $RR$ in $R$. The

---

1 Prepared with the partial support of the National Science Foundation, Grant GP-10.

2 We are grateful to Leonard J. Savage for asking us whether there are natural probabilities with interesting properties on the space of all probabilities (on a given space), and to David Blackwell for his interest which persisted through many helpful conversations.
construction may be continued by selecting one point at random from each of these four rectangles, according to the appropriate affine image of $\mu$, and so on. This procedure yields a nested decreasing sequence of closed sets, each being a finite union of closed rectangles: namely, $S, L \cup R, LL \cup RL \cup LR \cup RR$, and so on. The intersection of these closed sets is a nonempty closed set which, with probability 1, is the graph of a distribution function. This function is taken as the random value $F$ of $F$.

With probability 1, $F$ is continuous and strictly monotone, so that $P_\mu$ concentrates on the continuous strictly monotone distribution functions. Unless $\mu$ concentrates on the main diagonal, $F$ is almost certainly purely singular, so that $P_\mu$ concentrates on the purely singular distribution functions.

Three interesting choices for $\mu$ [cited below as examples (1), (2), (3)] are: (1) the uniform distribution on the vertical line segment $x=1/2, 0 \leq y \leq 1$; (2) the uniform distribution on the horizontal line segment $0 \leq x \leq 1, y=1/2$; (3) the uniform distribution on $S$.

3. The average distribution function. A probability $P$ in $\Delta$ determines as usual an average distribution function $F_P$ according to the relation

$$F_P(x) = \int_{G \in \Delta} G(x) dP(G).$$

Consider the mapping $T_\mu$ of $\Delta$ into $\Delta$ defined by

$$(T_\mu F)x = \int_0^1 \int_0^1 \beta F\left(\frac{x}{\alpha}\right)d\mu(\alpha, d\beta)$$

$$+ \int_0^1 \int_0^x \left[ \beta + (1 - \beta)F\left(\frac{x - \alpha}{1 - \alpha}\right) \right] d\mu(\alpha, d\beta).$$

The average $F_{P_\mu}$, or $F_\mu$ for short, satisfies the functional equation $T_\mu F = F$. Since $T_\mu$ is a uniformly strict contraction of the complete metric space $\Delta$ in the sup norm, $T_\mu$ has a unique fixed point, and if $G \in \Delta$, $(T_\mu)^nG \to F_\mu$ as $n \to \infty$.

In example (1) of §2, $F_\mu(x) = x, 0 \leq x \leq 1$; while in examples (2) and (3), $F_\mu(x) = 2\pi^{-1} \sin^{-1} x^{1/2}$. Surprisingly, therefore, the base probabilities $\mu$ of examples (1) and (2) yield different priors $P_\mu$. It follows easily that the base probability of example (3) produces a third distinct prior.

To generalize example (1) slightly, if $\mu$ concentrates on the vertical line segment $x=r, 0 \leq y \leq 1$, and has mean $(r, w)$, the equation $T_\mu F = F$
takes the form

\[ F(x) = wF\left(\frac{x}{r}\right), \quad 0 \leq x \leq r, \]

\[ = w + (1 - w)F\left(\frac{x - r}{1 - r}\right), \quad r \leq x \leq 1 \]

which, as shown in Chapter 6 of [2], has the unique solution

\[ F(x) = Q_w[Q_r^{-1}(x)], \]

where the coin-tossing distribution function \( Q_w \) may be defined as follows. Let \( \{\varepsilon_j, 1 \leq j < \infty\} \) be independent random variables with the common distribution \( P(\varepsilon_j = 0) = w, P(\varepsilon_j = 1) = 1 - w \); then \( Q_w \) is the distribution function of \( \sum_{j=1}^{\infty} \varepsilon_j 2^{-j} \). Since \( Q_r \) is strictly monotone on \( I \), its inverse function \( Q_r^{-1} \) is also a distribution function on \( I \).

The mapping \( T_\mu \) is the usual operator on probabilities associated with a discrete time Markov process having \( I \) for state space and the following transition mechanism: when at \( x \in I \), select \( (\alpha, \beta) \) at random from \( S \) according to \( \mu \) and move to \( x + \alpha(1 - x) \) with probability \( \beta \), or to \( x + \alpha(1 - x) \) with probability \( 1 - \beta \).

4. The uniqueness problem. In examples (1), (2), and (3), distinct base probabilities \( \mu_1 \) and \( \mu_2 \) lead to distinct priors \( P_{\mu_1} \) and \( P_{\mu_2} \). On the other hand, if \( \mu_1 \) and \( \mu_2 \) are distinct but concentrated on the main diagonal of \( S \), then \( P_{\mu_1} \) and \( P_{\mu_2} \) coincide, each assigning probability 1 to the distribution function \( \lambda \), \( \lambda(x) = x, 0 \leq x \leq 1 \). We have found no other exceptions to the conjecture that \( \mu_1 \neq \mu_2 \) implies \( P_{\mu_1} \neq P_{\mu_2} \). This implication does hold when \( \mu_1 \) and \( \mu_2 \) are both concentrated on the same vertical line segment, say, \( x = 1/2, 0 \leq y \leq 1 \). As before, write \( (1/2, w_i) \) for the mean of \( \mu_i \). Then \( F_{\mu_i} = Q_{w_i} \), and for \( w_1 \neq w_2 \), it is well known from the strong law of large numbers that \( Q_{w_1} \) and \( Q_{w_2} \) are mutually singular. It follows easily that \( P_{\mu_1} \) and \( P_{\mu_2} \) are not only different but even mutually singular in the following strong sense. There exist two disjoint Borel subsets \( B_1 \) and \( B_2 \) of \( I \) (e.g., \( B_i \) may be taken as the set of binary irrational numbers whose binary expansion has \( w_i \) for limiting relative frequency of 0's), such that \( P_{\mu_i} \) is concentrated on the collection \( C_i \) of distribution functions, where \( F \in C_i \) if and only if the probability in \( I \) determined by \( F \) concentrates on \( B_i \). Obviously, \( C_1 \) and \( C_2 \) are disjoint Borel subsets of \( \Delta \). If \( w_1 = w_2 \) but \( \mu_1 \neq \mu_2 \), such \( B_i \) do not exist; but \( P_{\mu_1} \) and \( P_{\mu_2} \) are still mutually singular in a fairly strong sense. Namely, there are disjoint Borel subsets \( C_1 \) and \( C_2 \) of \( \Delta \), such that \( P_{\mu_i} \) concentrates on \( C_i \), and having the further property: \( F_i \in C_i \) implies that \( F_1 \) and \( F_2 \) are mutually singular.
5. Consistency. Let $I^\infty$ be the space of sequences $\{x_j\}$, $x_j \in I$, $j = 1, 2, \ldots$, and let $\sigma(I^\infty)$ be its product $\sigma$-field. Let $\xi_n(s)$ be the $n$th coordinate of $s \in I^\infty$. If $\sigma(\Delta)$ denotes the Borel $\sigma$-field in $\Delta$, a probability $P$ on $(\Delta, \sigma(\Delta))$ determines a probability $\bar{P}$ on $(\Delta \times I^\infty, \sigma(\Delta) \times \sigma(I^\infty))$ by the relation

$$
\bar{P}\{A \times [s | \xi_j(s) \in A_j, 1 \leq j \leq n]\} = \int \prod_{j=1}^{n} |F| (A_j) dP(F)
$$

for $A \in \sigma(\Delta)$, $A_j$ Borel in $I$; where $|F|$ denotes the measure in $I$ determined by $F$. Let $P^*$ be a map from all $n$-tuples $\{x_j, 1 \leq j \leq n\}$ of elements of $I$ to probabilities on $(\Delta, \sigma(\Delta))$, so that $P^*(\xi_1(s), \ldots, \xi_n(s))$, as a function of $s$, is a version of the conditional distribution of $F$ under $\bar{P}$, given $\{\xi_j, 1 \leq j \leq n\}$. In other words, $P^*(\xi_1(s), \ldots, \xi_n(s))$ is "the" posterior distribution of $F$ given $\{\xi_j(s), 1 \leq j \leq n\}$.

If $G \in \Delta$, let $|G|^\infty$ denote the unique probability on $(I^\infty, \sigma(I^\infty))$ under which the $\{\xi_n\}$ are independent with common distribution function $G$. Since $\Delta$ is compact metrizable, the space of probabilities on $(\Delta, \sigma(\Delta))$ has a weak* topology, as part of the dual of the space of continuous functions on $\Delta$. Write $\Delta_0$ for the set of all $G \in \Delta$ satisfying the following condition: for $|G|^\infty$-almost all $s \in I^\infty$, $P^*(\xi_1(s), \ldots, \xi_n(s))$ converges to point mass at $G$, in the weak* topology, as $n \to \infty$. Then $\Delta_0 \subset \sigma(\Delta)$, and, as noted by Doob in [1], the forward martingale convergence theorem implies $P(\Delta_0) = 1$. But there is strong evidence that for most $P$, $\Delta_0$ is only of the first category [3, §5]. Here is a result in the other direction. If the base probability $\mu$ concentrates on a vertical line segment $x = r$, $0 \leq y \leq 1$, and assigns positive mass to every nondegenerate subinterval of that segment, then there exists at least one choice of $P^*_\mu$ for which $\Delta_0 = \Delta$; which, in the usual terminology, says that Bayes' estimates constructed from $P^*_\mu$ are consistent.

REFERENCES


2. Lester E. Dubins and Leonard J. Savage, How to gamble if you must, Multilithed notes, University of California, Berkeley, Calif., University of Michigan, Ann Arbor, Mich.


UNIVERSITY OF CALIFORNIA, BERKELEY