RANDOM DISTRIBUTION FUNCTIONS[1]

BY LESTER E. DUBINS[2] AND DAVID A. FREEDMAN[2]

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

1. **Introduction.** A *random distribution function* $\boldsymbol{F}$ is a measurable map from a probability space $(\Omega, \mathfrak{F}, Q)$ to the space Δ of distribution functions on the closed unit interval I, where Δ is endowed with its natural Borel σ-field, that is, the smallest σ-field containing the customary weak* topology. It determines a *prior* probability measure $P = Q\boldsymbol{F}^{-1}$ in the space Δ. Of course, $\boldsymbol{F}$ is essentially the same as the stochastic process $\{F_t, 0 \leqq t \leqq 1\}$ on $(\Omega, \mathfrak{F}, Q)$, where $F_t(\omega) = \boldsymbol{F}(\omega)(t)$. 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* μ on the Borel subsets of the unit square S determines a random distribution function $\boldsymbol{F}$ and so a prior probability P_μ in Δ, which will be described somewhat informally in §2, by explaining how to select a value of $\boldsymbol{F}$, i.e., a distribution function F, at random. §§3, 4 and 5 describe some properties of P_μ. Proofs will be given elsewhere. For ease of exposition, we assume that μ concentrates on, that is, assigns probability 1 to, the interior of S.

2. **The construction.** To select a value F of $\boldsymbol{F}$ at random, begin by selecting a point (x, y) from the interior of S according to μ. 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 $(u, v) \to (\alpha u + \beta, \gamma v + \delta)$, α and γ positive, which maps S onto L carries μ into a probability μ_L concentrated on L. The probability μ_R is defined in a similar way. Now select a point (x_L, y_L) at random from the interior of L according to μ_L, and a point (x_R, y_R) at random from the interior of R according to μ_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.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

construction may be continued by selecting one point at random from each of these four rectangles, according to the appropriate affine image of μ, 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 $\boldsymbol{F}$.

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

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

3. **The average distribution function.** A probability P in Δ 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_μ of Δ into Δ defined by

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

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

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

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

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

takes the form

$$F(x) = wF\left(\frac{x}{r}\right), \qquad 0 \leqq x \leqq r,$$

$$= w + (1 - w)F\left(\frac{x - r}{1 - r}\right), \qquad r \leqq x \leqq 1$$

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 $\{\epsilon_j, 1 \leqq j < \infty\}$ be independent random variables with the common distribution $P(\epsilon_j = 0) = w$, $P(\epsilon_j = 1) = 1 - w$; then Q_w is the distribution function of $\sum_{j=1}^{\infty} \epsilon_j 2^{-j}$. Since Q_r is strictly monotone on I, its inverse function Q_r^{-1} is also a distribution function on I.

The mapping T_μ 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 (α, β) at random from S according to μ and move to αx with probability β, or to $x + \alpha(1 - x)$ with probability $1 - \beta$.

4. **The uniqueness problem.** In examples (1), (2), and (3), distinct base probabilities μ_1 and μ_2 lead to distinct priors P_{μ_1} and P_{μ_2}. On the other hand, if μ_1 and μ_2 are distinct but concentrated on the main diagonal of S, then P_{μ_1} and P_{μ_2} coincide, each assigning probability 1 to the distribution function λ, $\lambda(x) = x$, $0 \leqq x \leqq 1$. 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 μ_1 and μ_2 are both concentrated on the same vertical line segment, say, $x = 1/2$, $0 \leqq y \leqq 1$. As before, write $(1/2, w_i)$ for the mean of μ_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_{μ_1} and P_{μ_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 irrationals whose binary expansion has w_i for limiting relative frequency of 0's), such that P_{μ_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 Δ. If $w_1 = w_2$ but $\mu_1 \neq \mu_2$, such B_i do not exist; but P_{μ_1} and P_{μ_2} are still mutually singular in a fairly strong sense. Namely, there are disjoint Borel subsets C_1 and C_2 of Δ, such that P_{μ_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.

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use

5. **Consistency.** Let I^∞ be the space of sequences $\{x_j\}$, $x_j \in I$, $j=1, 2, \cdots$, and let $\sigma(I^\infty)$ be its product σ-field. Let $\xi_n(s)$ be the nth coordinate of $s \in I^\infty$. If $\sigma(\Delta)$ denotes the Borel σ-field in Δ, a probability P on $(\Delta, \sigma(\Delta))$ determines a probability $\tilde{P}$ on $(\Delta \times I^\infty, \sigma(\Delta) \times \sigma(I^\infty))$ by the relation

$$\tilde{P}\{A \times [s \mid \xi_j(s) \in A_j, 1 \leqq j \leqq n]\} = \int_{F \in A} \prod_{j=1}^{n} |F|(A_j) dP(F)$$

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 $\{x_j, 1 \leqq j \leqq n\}$ of elements of I to probabilities on $(\Delta, \sigma(\Delta))$, so that $P^*(\xi_1(s), \cdots, \xi_n(s))$, as a function of s, is a version of the conditional distribution of F under $\tilde{P}$, given $\{\xi_j, 1 \leqq j \leqq n\}$. In other words, $P^*(\xi_1(s), \cdots, \xi_n(s))$ is "the" posterior distribution of F given $\{\xi_j(s), 1 \leqq j \leqq n\}$.

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 Δ 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 Δ. Write Δ_0 for the set of all $G \in \Delta$ satisfying the following condition: for $|G|^\infty$-almost all $s \in I^\infty$, $P^*(\xi_1(s), \cdots, \xi_n(s))$ converges to point mass at G, in the weak* topology, as $n \to \infty$. Then $\Delta_0 \in \sigma(\Delta)$, 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, Δ_0 is only of the first category [3, §5]. Here is a result in the other direction. If the base probability μ concentrates on a vertical line segment $x=r$, $0 \leqq y \leqq 1$, and assigns positive mass to every nondegenerate subinterval of that segment, then there exists at least one choice of P^*_μ for which $\Delta_0 = \Delta$; which, in the usual terminology, says that Bayes' estimates constructed from P_μ are consistent.

References

1. J. L. Doob, *Application of the theory of martingales*, Colloques Internationaux du Centre National de la Recherche Scientifique Vol., 13, pp. 23–27, Editions du Centre National de la Recherche Scientifique, Paris, 1949.

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.

3. David A. Freedman, *On the asymptotic behavior of Bayes' estimates in the discrete case*, Ann. Math. Statist., (to appear).

UNIVERSITY OF CALIFORNIA, BERKELEY

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use