SIAM REVIEW
Vol. 36, No. 4, pp. 652–654, December 1994

007

THE PROBABILITY INTEGRAL TRANSFORM AND RELATED RESULTS*

JOHN E. ANGUS†

Abstract. A simple proof of the probability integral transform theorem in probability and statistics is given that depends only on probabilistic concepts and elementary properties of continuous functions. This proof yields the theorem in its fullest generality. A similar theorem that forms the basis for the inverse method of random number generation is also discussed and contrasted to the probability integral transform theorem. Typical applications are discussed. Despite their generality and far reaching consequences, these theorems are remarkable in their simplicity and ease of proof.

Key words. random numbers, pivot functions, statistical inference

AMS subject classification. 62E15

1. Introduction. Let X be a real-valued random variable defined on a probability space $(\Omega, \Im, P)$. Let $F(x) = P\{\omega : X(\omega) \leq x\}$, $x \in (-\infty, \infty)$, define the cumulative distribution function (CDF). Let $U(0, 1)$ denote a random variable that is uniformly distributed on $(0, 1)$. The probability integral transform theorem is the following.

THEOREM 1. *If X has CDF $F(\cdot)$ which is continuous, then the random variable $Y = F(X)$ has the distribution of $U(0, 1)$.*

The following theorem, which I will call the quantile function theorem, is related to the probability integral transform theorem but applies to general CDFs.

THEOREM 2. *Let F be a CDF. If $F^{-1} : (0, 1) \to (-\infty, \infty)$ is defined by $F^{-1}(y) = \inf\{x : F(x) \geq y\}$, $0 < y < 1$ and U has the distribution of $U(0, 1)$, then $X = F^{-1}(U)$ has CDF F.*

The usual proof of the probability integral transform theorem given in popular undergraduate textbooks in mathematical statistics or probability assumes that F is absolutely continuous with density $f(x) = dF(x)/dx$ that is strictly positive for all x satisfying $0 < F(x) < 1$. The probability density f_Y of Y is then computed using formal calculus methods:

$$f_Y(y) = f(F^{-1}(y))dF^{-1}(y)/dy = 1 \quad \text{for } 0 < y < 1 \tag{1}$$

which is the probability density of $U(0, 1)$. (A notable exception among contemporary texts at the advanced undergraduate level is Casella and Berger [3, pp. 52-54], which gives a general proof.) Theorem 1 assumes, however, only that F is continuous, not necessarily absolutely continuous, and does not require that F be strictly increasing. Since F is a CDF, it must be nondecreasing. If it is continuous, it may still have intervals of constancy, that is, there may be one or more intervals of the form $[a, b)$ where F is constant throughout $[a, b)$. Of course, if $[a, b)$ is an interval of constancy for F, then $P\{X \in (a, b)\} = \lim_{\varepsilon \searrow 0} F(b - \varepsilon) - F(a) = 0$. Moreover, F is necessarily differentiable almost everywhere with respect to Lebesgue measure, but F need not be recoverable from its derivative via integration. So called "continuous singular" CDFs, CDFs that are continuous but which have derivative 0 almost everywhere with respect to Lebesgue measure, are somewhat pathological but do arise in some applications. For example, in the gambler's ruin problem with bold play and fortunes and wagers normalized to lie in [0, 1], the probability of success with initial fortune x is a continuous singular CDF (see Billingsley [2, pp. 98–101 and 427–429]). More common in applications are continuous CDFs having intervals of constancy. In either of these two cases, Theorem 1 is still true but the formal calculations (1) have no meaning.

*Received by the editors July 26, 1993; accepted for publication February 25, 1994.

†Claremont Graduate School, Department of Mathematics, 143 East Tenth Street, Claremont, California 91711 (Angusj@cgsvax.claremont.edu).

The implications of the integral transform theorem are far reaching in probability and statistics. For example, in statistical inference, if $F \equiv F_\theta$ depends on an unknown parameter θ, then if F_θ is continuous, $F_\theta(X)$ is distributed as $U(0, 1)$, independently of the value of θ, and hence provides a pivotal function which can be used for determining confidence intervals and/or sets. See, for example, [1, p. 366].

In another application, suppose that $X_1, \dots, X_n$ is a random sample from an unknown continuous CDF F, and suppose that one is interested in whether there is evidence to support the hypothesis that $F = F_0$ for some completely known continuous CDF F_0. If $F = F_0$ is true, then by Theorem 1, $F_0(X_{1:n}) < F_0(X_{2:n}) < \cdots < F_0(X_{n:n})$, where $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ are the ordered $X_i s$ (i.e., the "order statistics"), are distributed like the ordered statistics from a random sample of size n from the distribution of $U(0, 1)$. Since the i^{th} smallest ordered value from a sample of size n from the distribution of $U(0, 1)$ has expectation $i/(n + 1)$, a plot of the points $(i/(n + 1), F_0(X_{i:n}))$, $i = 1, \dots, n$ should lie roughly along a straight line of slope 1 if $F = F_0$. Many tests of goodness-of-fit are based on these concepts.

In contrast to the probability integral transform theorem, the quantile function theorem, Theorem 2, is of practical importance because it forms a basis for generating random deviates from an arbitrary (not necessarily continuous) CDF.

Theorems 1 and 2 are proven in the next section. The proof of Theorem 1 given is based on probabilistic concepts, and simple properties of continuous functions, and is actually simpler than the argument of limited scope (1), and both simpler and shorter than that of [3]. The suggested proof of Theorem 2 in [1, problem 5, p. 227] asks the student to consider the special case that F is one-to-one, but the proof presented here is general and suitable for undergraduate mathematical statistics students.

2. Proofs of the theorems. The following lemma is the key to the proof of Theorem 1. Note that it makes no assumptions concerning the CDF F of X.

LEMMA 1. *Let X have CDF F. Then for all real x, $P\{F(X) \le F(x)\} = F(x)$.*

Proof. Decompose the event $\{\omega : F(X(\omega)) \le F(x)\}$ (omitting the "ω" notation) as

$$\{F(X) \le F(x)\} = [\{F(X) \le F(x)\} \cap \{X \le x\}] \cup [\{F(X) \le F(x)\} \cap \{X > x\}].$$

Since $\{X \le x\} \subset \{F(X) \le F(x)\}$ and $\{X > x\} \cap \{F(X) < F(x)\}$ is empty, it follows that

$$\{F(X) \le F(x)\} = \{X \le x\} \cup [\{X > x\} \cap \{F(X) = F(x)\}]. \tag{2}$$

Taking probabilities, the result follows because the last event in brackets in (2) has probability 0 (since it implies that X lies in the interior of an interval of constancy of F). $\square$

Proof of Theorem 1. Let $u \in (0, 1)$. Since F is continuous, there exists a real x such that $F(x) = u$. Then by Lemma 1, $P\{Y \le u\} = P\{F(X) \le F(x)\} = u$. This implies that Y is distributed as $U(0, 1)$.

The proof of Theorem 2 is also very short, but in classroom practice a picture or two showing how particular CDFs F translate into their respective $F^{-1}s$ is helpful. In particular, this will convince the students that F^{-1} is nondecreasing, jumps in F translate into flat regions of F^{-1}, flat regions of F (earlier called intervals of constancy) translate into jumps in F^{-1}, and whereas F is right continuous, F^{-1} is left continuous.

Proof of Theorem 2. First, note that for any x such that $0 < F(x) < 1$ and any $u \in (0, 1)$, $F(x) \ge u$ if, and only if, $x \ge F^{-1}(u)$. For, suppose $x \ge F^{-1}(u) = \inf\{x : F(x) \ge u\}$. Then since F is nondecreasing and right continuous, $\{x : F(x) \ge u\}$ is an (infinite) interval that contains its left-hand endpoint. Hence, x must satisfy $F(x) \ge u$. Conversely, suppose $F(x) \ge u$. Then $x \ge \inf\{x : F(x) \ge u\} = F^{-1}(u)$. It follows now that $P\{F^{-1}(U) \le x\} = P\{U \le F(x)\} = F(x)$, completing the proof.

REFERENCES

[1] L. J. BAIN AND M. ENGELHARDT, *Introduction to Probability and Mathematical Statistics*, 2nd ed., PWS-Kent, Boston, MA, 1992.

[2] P. BILLINGSLEY, *Probability and Measure*, 2nd ed., John Wiley, New York, 1986.

[3] G. CASELLA AND R. L. BERGER, *Statistical Inference*, Brooks/Cole, Pacific Grove, 1990.