The Annals of Probability
1981, Vol. 9, No. 4, 665–670

DISTRIBUTIONAL RESULTS FOR RANDOM FUNCTIONALS OF A DIRICHLET PROCESS

BY ROBERT C. HANNUM,[1,2] MYLES HOLLANDER,[2] AND NAFTALI A. LANGBERG[3]

University of Denver and Florida State University

We obtain an expression for the distribution function of the random variable $\int ZdP$ where P is a random distribution function chosen by Ferguson's (1973) Dirichlet process on (R, B) (R is the real line and B is the σ-field of Borel sets) with parameter α, and Z is a real-valued measurable function defined on (R, B) satisfying $\int |Z| d\alpha < \infty$. As a consequence, we show that when α is symmetric about 0 and Z is an odd function, then the distribution of $\int ZdP$ is symmetric about 0. Our main result is also used to obtain a new result for convergence in distribution of Dirichlet-based random functionals.

1. Introduction and summary. Let P be a random probability measure chosen by Ferguson's (1973) Dirichlet process on (R, B) (R is the real line and B is the σ-field of Borel sets) with parameter α. Let Z be a real-valued measurable function defined on (R, B) satisfying $\int |Z| d\alpha < \infty$. Ferguson's fundamental paper contains many results including, for the purpose of developing Bayesian nonparametric estimators of various parameters, results giving the mean of such random variables (rv's) as $\int ZdP$ and (variance P) and the median of the rv (median P). Yamato (1977) extends Ferguson's results by obtaining the moments of various random estimable parameters. Results pertaining to the distribution functions (df's) of random functionals are less available, an exception being Ferguson's (1973) expression for the df of the random qth percentile of P.

In this note we study the df of $\int ZdP$. Our contributions are:

(i) In Theorem 2.5 we establish the equality

$$\Pr\left\{\int ZdP \le x\right\} = \Pr\{T^x \le 0\}, \tag{1.1}$$

which relates the df of $\int ZdP$ to the distributions of rv's T^x, $x \in R$, defined by (2.3) and Theorem 2.3. In Theorem 2.3 we obtain the characteristic function (cf) of T^x.

(ii) The cf of T^x can be used to obtain results concerning the distribution of $\int ZdP$. For example, in Corollary 2.6 we use the cf of T^x to prove that when Z is odd and α is symmetric about 0, then the distribution of $\int ZdP$ is symmetric about 0. (We also present an alternative proof of Corollary 2.6 using the gamma process representation (see Ferguson, 1973; Ferguson and Klass, 1972) of the Dirichlet process.) Thus, for example, when α is symmetric about 0, the odd moments of the random P, namely $\int \xi^n dP(\xi)$, $n = 1, 3, 5, \cdots$, have distributions that are symmetric about 0.

(iii) In Corollary 2.7 we use the cf of T^x to prove that if $P, P_n, n = 1, 2, \cdots$, are random probability measures chosen by Ferguson's Dirichlet process on (R, B) with parameters

Received April 16, 1980.

[1] Research supported by National Institutes of Environmental Health Sciences Traineeship Grant No. 5 T32 E 50 7011.

[2] Research supported by the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-78-3678.

[3] Research supported by the United States Army Research Office, Durham, under Grant No. DAAG29-79-C-0158.

AMS 1970 *subject classifications.* Primary 60K99; secondary 60E05.

Key words and phrases. Dirichlet process, distribution of random functionals.

α, α_n, respectively, and if α_n converges weakly to α as $n \to \infty$, then (under mild regularity) $\int Z dP_n$ converges in distribution to $\int Z dP$.

2. The distribution of $\int ZdP$. We assume the reader is familiar with definitions and basic results concerning Ferguson's Dirichlet process. If this is not the case, the reader is referred to Ferguson (1973). Throughout this section we take P to be a Dirichlet process on (R, B) with parameter α, and Z to be a real-valued measurable function defined on (R, B) satisfying $\int |Z| d\alpha < \infty$. We first prove two lemmas which will be used to show that the rv's T^x, $x \in R$, defined by (2.3) and Theorem 2.3, have continuous df's.

LEMMA 2.1. *Let r be a positive integer and let $X_1, \cdots, X_r$ be independent and identically distributed* (i.i.d.) rv's *such that* $\sum_{i=1}^r X_i$ *has a continuous* df. *Then X_1 has a continuous* df.

PROOF. Note that for all $y \in R$,

$$(\Pr\{X_1 = y\})^r \leq \Pr\{\textstyle\sum_{i=1}^r X_i = ry\} . \tag{2.1}$$ □

LEMMA 2.2. *Let $x \in R$ and let α be a finite measure on (R, B) such that $\alpha\{\xi\colon\ Z(\xi) \neq x\} > 0$, where Z is a measurable function integrable with respect to α. Then there is a positive integer r such that*

$$\int_0^\infty \exp\left\{-r\int_{-\infty}^\infty \ln[1 + t^2\{Z(\xi) - x\}^2] d\alpha(\xi)\right\} dt < \infty.$$

PROOF. First note that since $\alpha\{\xi\colon\ Z(\xi) \neq x\} > 0$ then there exist two positive integers j and r such that $r \cdot \alpha\{\xi\colon\ |Z(\xi) - x| \geq j^{-1}\} > 1$. Letting $B_j = \{\xi\colon\ |Z(\xi) - x| \geq j^{-1}\}$ we then note that

$$\int_{-\infty}^\infty \ln[1 + t^2\{Z(\xi) - x\}^2] d\alpha(\xi) \geq \int_{B_j} \ln[1 + t^2\{Z(\xi) - x\}^2] d\alpha(\xi)$$
$$\geq \alpha(B_j) \cdot \ln[1 + (t/j)^2],$$

and so we have

$$\int_0^\infty \exp\left\{-r\int_{-\infty}^\infty \ln[1 + t^2\{Z(\xi) - x\}^2] d\alpha(\xi)\right\} dt \leq \int_0^\infty \exp\{-r \cdot \alpha(B_j) \cdot \ln[1 + (t/j)^2]\} dt$$

$$= \int_0^\infty [1 + (t/j)^2]^{-r\cdot\alpha(B_j)} dt \leq \int_0^\infty [1 + (t/j)^2]^{-1} dt < \infty.$$ □

Now we introduce some notation. Let

$$I_{jk} = \begin{cases} \{\xi\colon\ (j-1)/k < Z(\xi) \leq j/k\}, & \text{for } j = -k^2 + 1, \cdots, k^2 \\ \{\xi\colon\ Z(\xi) \leq -k\}, & \text{for } j = -k^2 \\ \{\xi\colon\ Z(\xi) > k\}, & \text{for } j = k^2 + 1. \end{cases}$$

Further let W_{jk}, $j = -k^2, \cdots, k^2 + 1$, be independent gamma rv's with shape parameters $\alpha(I_{jk})$ respectively, and common scale parameter 1. Let $x \in R$ and define, for $k = 1, 2, \cdots$

$$T_k^x = \textstyle\sum_{j=-k^2}^{-1} \{(j/k) - x\} W_{jk} + \sum_{j=0}^{k^2+1} \{[(j-1)/k] - x\} W_{jk}. \tag{2.2}$$

THEOREM 2.3. *Let $x \in R$. Then the sequence T_k^x, $k = 1, 2, \cdots$, converges in law as $k \to \infty$ to a* rv *T^x whose* cf *is given by*

$$\phi_{T^x}(t) = \exp\left\{-\int_{-\infty}^{\infty} \ln[1 - it\{Z(\xi) - x\}]d\alpha(\xi)\right\}. \tag{2.3}$$

PROOF. Let $\delta_\xi(I_{jk}) = 1$ when $\xi \in I_{jk}$ and 0 otherwise, and for each $\xi \in R$ define

$$Z_k(\xi) = \sum_{j=-k^2}^{-1} [j/k]\delta_\xi(I_{jk}) + \sum_{j=0}^{k^2+1} [(j-1)/k]\delta_\xi(I_{jk})$$

for $k = 1, 2, \cdots$. Let $x \in R$ and note from (2.2) that since T_k^x is a linear combination of independent gamma rv's, it has a cf which is infinitely divisible. Hence we may write

$$\begin{aligned}
\ln E(\exp\{itT_k^x\}) &= \textstyle\sum_{j=-k^2}^{-1} \ln E(\exp\{it[(j/k) - x]W_{jk}\}) \\
&\quad + \textstyle\sum_{j=0}^{k^2+1} \ln E(\exp\{it[\{(j-1)/k\} - x]W_{jk}\}) \\
&= \textstyle\sum_{j=-k^2}^{-1} \ln[1 - it\{(j/k) - x\}]^{-\alpha(I_{jk})} \\
&\quad + \textstyle\sum_{j=0}^{k^2+1} \ln[1 - it\{[(j-1)/k] - x\}]^{-\alpha(I_{jk})} \\
&= -\{\textstyle\sum_{j=-k^2}^{-1} \alpha(I_{jk}) \ln[1 - it\{(j/k) - x\}] \\
&\quad + \textstyle\sum_{j=0}^{k^2+1} \alpha(I_{jk}) \ln[1 - it\{[(j-1)/k] - x\}]\} \\
&= -\int_{-\infty}^{\infty} \ln[1 - it\{Z_k(\xi) - x\}]d\alpha(\xi) \\
&= -\int_{-\infty}^{\infty} \ln[(1 + t^2\{Z_k(\xi) - x\}^2)^{1/2}]d\alpha(\xi) \\
&\quad + i\int_{-\infty}^{\infty} \tan^{-1}[t\{Z_k(\xi) - x\}]d\alpha(\xi).
\end{aligned}$$

Now observe that $|Z_k(\xi)| \leq |Z(\xi)|$ for all real ξ and so we have

$$\ln[(1 + t^2\{Z_k(\xi) - x\}^2)^{1/2}] \leq |t| \cdot \{|Z_k(\xi)| + x\} \leq |t| \cdot \{|Z(\xi)| + x\} \tag{2.5}$$

and

$$|\tan^{-1}[t\{Z_k(\xi) - x\}]| \leq |t| \cdot \{|Z_k(\xi)| + x\} \leq |t| \cdot \{|Z(\xi)| + x\}. \tag{2.6}$$

Since the right-hand-sides of (2.5) and (2.6) are integrable with respect to α and since $\lim_{k\to\infty} Z_k(\xi) = Z(\xi)$, then by the dominated convergence theorem we have from (2.4) that

$$\lim_{k\to\infty} E(\exp\{itT_k^x\}) = \exp\left\{-\int_{-\infty}^{\infty} \ln[1 - it\{Z(\xi) - x\}]d\alpha(\xi)\right\}. \tag{2.7}$$

Writing the right-hand-side of (2.7) as $\exp\{-[G(t, x) + iH(t, x)]\}$ where

$$G(t, x) = \frac{1}{2}\int_{-\infty}^{\infty} \ln[1 + t^2\{Z(\xi) - x\}^2]d\alpha(\xi),$$

$$H(t, x) = -\int_{-\infty}^{\infty} \tan^{-1}[t\{Z(\xi) - x\}]d\alpha(\xi),$$

it is straightforward to show that for any real number x, both $|G(t, x) - G(0, x)|$ and $|H(t, x) - H(0, x)|$ are bounded above by $|t| \cdot \int_{-\infty}^{\infty} \{|Z(\xi)| + x\}d\alpha(\xi)$, which goes to zero as $t \to 0$. This shows that the right-hand-side of (2.7) is continuous at $t = 0$ and hence the result of the theorem follows from the continuity theorem for cf's. □

LEMMA 2.4. *Let $x \in R$ and assume that $\alpha\{\xi: \ Z(\xi) \neq x\} > 0$. Then T^x has a continuous* df.

PROOF. By Lemma 2.2 there is a positive integer r such that $\int_{-\infty}^{\infty} | [E(\exp\{itT^x\})]^r | \, dt < \infty$. Now $[E(\exp(itT^x)\})]^r$ is the cf of the convolution of r i.i.d. rv's distributed as T^x. Since the cf of this convolution is absolutely integrable we conclude (cf. Breiman (1968), page 178) that the convolution of r i.i.d. rv's distributed as T^x is continuous. Consequently the result follows from Lemma 2.1 □

We are now ready to obtain an expression for the df of $\int ZdP$.

THEOREM 2.5. *Let P be a Dirichlet process on (R, B) with parameter α and let Z denote a real-valued measurable function defined on (R, B) integrable with respect to α.*

(i) *If there exists a real number x such that $\alpha\{\xi: \ Z(\xi) \neq x\} = 0$, then $\int ZdP$ is degenerate at x.*

(ii) *If $\alpha\{\xi: \ Z(\xi) \neq x\} > 0$ for all $x \in R$, then* (1.1) *holds, where T^x is the* rv *whose* cf *is given by the right-hand-side of* (2.3).

PROOF. (i) Note that $\alpha\{\xi: \ Z(\xi) \neq x\} = 0$ implies that $P\{\xi: \ Z(\xi) \neq x\} = 0$. Thus $\int_{-\infty}^{\infty} Z(\xi) dP(\xi) = x$.

(ii) Let $A_k = \int_{-\infty}^{\infty} Z_k(\xi) dP(\xi)$, for $k = 1, 2, \cdots$, and note that

$$\Pr\{A_k \ lu \ x\} = \Pr\left\{\int_{-\infty}^{\infty} Z_k(\xi) dP(\xi) \le x\right\}$$

$$= \Pr\{\textstyle\sum_{j=-k^2}^{-1} [j/k] P(I_{jk}) + \sum_{j=0}^{k^2+1} [(j-1)/k] P(I_{jk}) \le x\} \,. \tag{2.8}$$

Now for each k the random vector $(P(I_{jk})), j = -k^2, \cdots, k^2 + 1$, has a Dirichlet distribution with parameter $(\alpha(I_{jk})), j = -k^2, \cdots, k^2 + 1$. Letting $W_{jk}, j = -k^2, \cdots, k^2 + 1$, represent independent gamma rv's with shape parameters $\alpha(I_{jk}), j = -k^2, \cdots, k^2 + 1$, respectively and common scale parameter 1, and using the gamma distribution definition of the Dirichlet distribution, (2.8) becomes

$$\begin{aligned} \Pr\{A_k \le x\} &= \Pr\{\textstyle\sum_{j=-k^2}^{-1} [j/k] W_{jk} + \sum_{j=0}^{k^2+1} [(j-1)/k] W_{jk} \le x \sum_{j=-k^2}^{k^2+1} W_{jk}\} \\ &= \Pr\{\textstyle\sum_{j=-k^2}^{-1} \{(j/k) - x\} W_{jk} + \sum_{j=0}^{k^2+1} \{[(j-1)/k] - x\} W_{jk} \le 0\} \\ &= \Pr\{T_k^x \le 0\}. \end{aligned} \tag{2.9}$$

Now note that $\lim_{k\to\infty} A_k = \int ZdP$ almost everywhere with respect to α and thus by (2.9) we have

$$\Pr\left\{\int ZdP \le x\right\} = \lim_{k\to\infty} \Pr\{A_k \le x\} = \lim_{k\to\infty} \Pr\{T_k^x \le 0\} = \Pr\{T^x \le 0\},$$

the last equality following from Theorem 2.3 and Lemma 2.4. □

We illustrate the result of Theorem 2.5 in a simple case. Let $A \in B$, $\alpha(A) \in (0, \infty)$, and $Z(x) = 1, x \in A, = 0, x \notin A$. Then the rv $\int ZdP$ has a Beta distribution with parameters $\alpha(A)$, $\alpha(R) - \alpha(A)$. The cf of T^x, given by (2.3), is $[1 - it(1 - x)]^{-\alpha(A)}[1 + itx]^{-(\alpha(R) - \alpha(A))}$. Thus T^x has the same distribution as $(1 - x)Y_2 - xY_1$, where Y_1, Y_2 are independent and $Y_1 \sim \Gamma(\alpha(R) - \alpha(A), 1)$, $Y_2 \sim \Gamma(\alpha(A), 1)$. Consequently $\Pr\{(1 - x)Y_2 - xY_1 \le 0\} = \Pr\{Y_2(Y_1 + Y_2)^{-1} \le x\} = \Pr\{\int ZdP \le x\}$.

Using Theorem 2.5 in conjunction with the cf of T^x, given in (2.3), it is now possible to gain additional information concerning the distribution of the random mean $\int ZdP$.

COROLLARY 2.6. *Let P be a Dirichlet process on (R, B) with parameter α, where α is a positive finite and symmetric measure about the origin. Let Z denote a real-valued measurable and odd function such that $\int |Z|\, d\alpha < \infty$. Then $\int ZdP$ has a symmetric distribution about the origin.*

PROOF. First assume that there is a real number x such that $\alpha\{\xi\colon\ Z(\xi) \neq x\} = 0$. Then since $Z(\xi) = -Z(-\xi)$ it follows that $\alpha\{\xi\colon\ Z(\xi) \neq 0\} = 0$. Thus by Theorem 2.5 $\int ZdP$ is degenerate at $x = 0$ and hence symmetric.

Next assume that $\alpha\{\xi\colon\ Z(\xi) \neq x\} > 0$ for all $x \in R$. By repeating our arguments it is clear that $\Pr\{\int ZdP \geq -x\} = \Pr\{T^{-x} \geq 0\}$ and thus to show that $\int ZdP$ is symmetric about the origin it suffices to show that, for all $x \in R$,

$$\Pr\{T^x \leq 0\} = \Pr\{-T^{-x} \leq 0\}. \tag{2.10}$$

Since $Z(\xi) = -Z(-\xi)$ and since $d\alpha(\xi) = d\alpha(-\xi)$ we have

$$\begin{aligned}\ln E(\exp\{itT^{-x}\}) &= -\int_{-\infty}^{\infty} \ln[1 - it\{Z(\xi) + x\}]\, d\alpha(\xi)\\ &= \int_{-\infty}^{\infty} \ln[1 + it\{Z(-\xi) - x\}]\, d\alpha(-\xi) = \ln E(\exp\{-itT^x\}).\end{aligned} \tag{2.11}$$

From (2.11) we obtain that T^x has the same *df* as $-T^{-x}$ and hence (2.10) follows directly. □

The proof of Corollary 2.6 given above is based on Ferguson's (1973) original definition of the Dirichlet process. Below we exhibit an alternative proof of Corollary 2.6 which is based on Ferguson's (1973) (also see Ferguson and Klass, 1972) alternative definition of the Dirichlet process as the sum of a countable number of jumps of random height at a countable number of random points. (This alternative representation is often useful in obtaining results for the Dirichlet process, cf. Ferguson (1973), Korwar and Hollander (1973), Yamato (1977).)

ALTERNATIVE PROOF OF COROLLARY 2.6. We have, by Theorem 2 of Ferguson (1973),

$$\int ZdP = \sum_{j=1}^{\infty} Z(V_j)P_j$$

where $V_1, V_2, \cdots$ are i.i.d. according to $Q(\cdot) = \alpha(\cdot)/\alpha(R)$, where the P_j's are independent of the V_j's, and where the P_j's depend on $\alpha(\cdot)$ only through $\alpha(R)$. Thus we have (since Z is odd)

$$\int ZdP = \sum_{j=1}^{\infty} Z(V_j)P_j = -\sum_{j=1}^{\infty} Z(-V_j)P_j. \tag{2.12}$$

Now since α is symmetric about the origin, $-V_j =_d V_j$, for all $j = 1, 2, \cdots$, where "$=_d$" means "has the same distribution as." Hence, since the V_j's are independent of the P_j's we have

$$-\sum_{j=1}^{\infty} Z(-V_j)P_j =_d -\sum_{j=1}^{\infty} Z(V_j)P_j = -\int ZdP. \tag{2.13}$$

From (2.12) and (2.13) we conclude that $\int ZdP =_d -\int ZdP$ and so the df of $\int ZdP$ is symmetric about the origin. □

Finally, we use Theorem 2.5 to obtain

COROLLARY 2.7. *Let α, α_n, be finite measures on (R, B), and P, P_n, be random probability measures chosen by Ferguson's Dirichlet process with parameters α, α_n, respectively, $n = 1, 2, \cdots$. Assume* (i) $\lim_{n\to\infty} \alpha_n\{(-\infty, x]\} = \alpha\{(-\infty, x]\}$ *for all continuity points x of $\alpha\{(-\infty, x]\}$,* (ii) $\lim_{z\to\infty}|Z(z)| = \infty$, (iii) *$Z$ is α-continuous,* (iv) *z is nondegenerate,* (v) *Z is bounded in each interval and* (vi) $\limsup_{n\to\infty} \int |Z|^{1+\delta}\, d\alpha_n < \infty$ *for some $\delta \in (0, \infty)$. Then the sequence of* rv's $\int Z dP_n$, $n = 1, 2, \cdots$, *converges in distribution as $n \to \infty$ to the* rv $\int ZdP$.

PROOF. Let the cf of T_n^x be given by Equation (2.3) when α is replaced by α_n, $n = 1, 2, \cdots$. Clearly

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \ln[1 - it\{Z(z) - x\}]\, d\alpha_n(z) = \int_{-\infty}^{\infty} \ln[1 - it\{(Z(z) - x\}]\, d\alpha(z)$$

[Breiman (1968), page 164]. Thus, $\lim_{n\to\infty} \phi_{T_n^x}(t) = \phi_{T^x}(t)$ for all $t \in (-\infty, \infty)$. By the continuity theorem for cf's [Breiman (1968), page 171] T_n^x converges in distribution as $n \to \infty$ to T^x for all $x \in (-\infty, \infty)$. Consequently by Lemma 2.4 and Theorem 2.5, $\lim_{n\to\infty} P_r\{\int ZdP_n \le x\} = \lim_{n\to\infty} P_r\{T_n^x \le 0\} = Pr\{T^x \le 0\} = P_r\{\int ZdP \le x\}$. □

The result of Corollary 2.7 holds also in the following two cases: (I) Z is bounded, and (II) Z is α-degenerate at x. In case (I) assume that Conditions (i), (iii) hold, and that (vii) $\lim_{n\to\infty} \alpha_n(R) = \alpha(R)$. In case (II) assume that Conditions (i), (v) and (vii) hold.

Acknowledgment. We thank Jayaram Sethuraman for many helpful discussions.

REFERENCES

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, Mass.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

FERGUSON, T. S. and KLASS, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.* **43** 1634–1643.

KORWAR, R. M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probability* **1** 705–711.

YAMATO, H. (1977). Relation between limiting Bayes estimates and the U-statistics for estimable parameters of degrees 2 and 3. *Commun. Statist.—Theor. Math. A6(1)* 55–66.

ROBERT C. HANNUM
DIVISION OF STATISTICS
UNIVERSITY OF DENVER
DENVER, COLORADO 80210

MYLES HOLLANDER
DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306

NAFTALI LANGBERG
DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306