The Annals of Statistics
1982, Vol. 10, No. 4, 1302–1305

THE TAILS OF PROBABILITIES CHOSEN FROM A DIRICHLET PRIOR

BY HANI DOSS AND THOMAS SELLKE

Florida State University and Purdue University

Let α be a finite nonnull measure on $\mathbb{R}$, and let the random distribution function F be distributed according to $\mathscr{D}_\alpha$, where $\mathscr{D}_\alpha$ is the Dirichlet process prior with parameter α; see Ferguson (1973). This note points out that, almost surely, the tails of F are much smaller than the tails of α.

1. Introduction. Let $\mathscr{P}(\mathbb{R})$ denote the set of all probability measures on $\mathbb{R}$. The Dirichlet process priors discussed by Ferguson (1973, 1974) are probability measures on $\mathscr{P}(\mathbb{R})$ which are parameterized by the set of all finite nonnull measures on $\mathbb{R}$. Let α be a finite nonnull measure on $\mathbb{R}$ and write $\alpha = \alpha(\mathbb{R})\,\alpha_0$ so that α_0 is a probability measure. We will use the same symbol to denote both a measure and its distribution function. Ferguson (1973) defines the random distribution function F to have the Dirichlet distribution with parameter α, denoted by $\mathscr{D}_\alpha$, if for every finite measureable partition $\{A_1, \cdots, A_k\}$ of $\mathbb{R}$ the random vector $(F(A_1), \cdots, F(A_k))$ has the Dirichlet distribution with parameter vector $(\alpha(A_1), \cdots, \alpha(A_k))$.

In his 1973 paper, Ferguson demonstrated the existence of the Dirichlet process priors and showed that they could be used to solve certain nonparametric problems.

From the definition of the Dirichlet process, it follows that

$$EF(t) = \alpha_0(t).$$

For this reason, α_0 is often called the prior guess at F. The parameter $\alpha(\mathbb{R})$ indicates the concentration of $\mathscr{D}_\alpha$ around α_0. For example, it is easy to show that if α_0 is fixed and $\alpha(\mathbb{R}) \to \infty$, then $\mathscr{D}_\alpha$ converges to the point mass at α_0 in the weak topology. Ferguson (1973) also showed that if g is a measurable real valued function defined on $\mathbb{R}$ such that $\int |g|\, d\alpha_0 < \infty$, then

$$E\int g\, dF = \int g\, d\,\alpha_0.$$

These considerations might lead one to believe that, in some sense, "on the average, F resembles α_0."

The purpose of this note is to point out that, almost surely, the tails of F are much smaller than the tails of α_0. To simplify the notation, let us temporarily assume that $0 < \alpha_0(t) < 1$ for all real t. Our main result, which follows from Theorem 1 of Fristedt (1967), is that if h is a function that is strictly increasing and convex on $(0, \varepsilon)$ for some sufficiently small $\varepsilon > 0$, then

$$\limsup_{t\to\infty} \frac{1 - F(t)}{h(1 - \alpha_0(t))} = 0 \quad \text{a.s.} \quad [\mathscr{D}_\alpha] \text{ and } \limsup_{t\to-\infty} \frac{F(t)}{h(\alpha_0(t))} = 0 \quad \text{a.s.} \quad [\mathscr{D}_\alpha]$$

if and only if

$$\int_0^\varepsilon \log h(t)\, dt > -\infty. \tag{1}$$

A particular choice for h will yield, for example, that for almost every sample distribution

Received September 1981; revised April 1982.
AMS 1970 *subject classifications.* Primary, 60G17; secondary, 62C10.
Key words and phrases. Dirichlet process, gamma process.

function F, for all sufficiently large t,

$$1 - F(t) \le \exp\left(-\frac{1}{\{1 - \alpha_0(t)\}[\log\{1 - \alpha_0(t)\}]^2} \right).$$

If α_0 is the Cauchy distribution, this means that for all sufficiently large t,

$$1 - F(t) \le \exp\left\{ -\frac{t}{(\log t)^2} \right\},$$

from which it is apparent that with probability one, F has moments of all orders.

Functions h satisfying the integrability condition (1) thus form "upper envelopes" for the tails of F. It will not be possible to provide "sharp" envelopes. Results from Fristedt and Pruitt (1971) will enable us to provide "lower envelopes" for the tails of F. It will then be worth noting that the upper and the lower envelopes are not too far apart.

2. Upper and lower envelopes for the tails of F. Let F be distributed according to $\mathscr{D}_\alpha$, where α is a finite nonnull measure on $\mathbb{R}$. We will first be interested in giving upper bounds for the tail behavior of F that are valid almost surely; i.e., we would like to find functions h and constants c such that

$$\limsup_{\alpha(t)\to 0} \frac{F(t)}{h(\alpha(t))} = c \quad \text{a.s.} \tag{2}$$

Here we define $\limsup_{\alpha(t)\to 0} \dfrac{F(t)}{h(\alpha(t))}$ to mean

$$\limsup_{\alpha(t)\to 0} \frac{F(\alpha^{-1}(\alpha(t)))}{h(\alpha(t))},$$

where $\alpha^{-1}(u) = \inf\{x \mid \alpha(x) \ge u\}$. We use this definition rather than the more natural $\limsup_{t\to-\infty} F(t)/h(\alpha(t))$ because of the possibility that $\alpha(t) = 0$ for some finite t. We will deal only with the left tail of F. It is clear that the corresponding results hold for the right tail.

Let $\mathscr{G}(u, 1)$ denote the gamma distribution with shape parameter u and scale parameter 1. Ferguson (1973, 1974) shows that if $\{\gamma(t);\ t \in [0, \infty)\}$ is a stationary independent increments process with $\gamma(t) \sim \mathscr{G}(t, 1)$ and if

$$F(t) = \frac{\gamma(\alpha(t))}{\gamma(\alpha(\mathbb{R}))} \quad \text{for} \quad t \in \mathbb{R}, \tag{3}$$

then F has the Dirichlet distribution with parameter α.

The process $\{\gamma(t);\ t \in [0, \infty)\}$ has Lévy measure $d\nu(x) = (e^{-x}/x)\,dx$. From Theorem 1 of Fristedt (1967) it follows that if h is a positive convex strictly increasing function on $(0, \infty)$, then

$$\begin{aligned} &\int_0^1 \nu[h(t), \infty)\,dt < \infty \quad \text{implies} \quad P\left\{\limsup_{t\to 0} \frac{\gamma(t)}{h(t)} = 0\right\} = 1,\\ &\int_0^1 \nu[h(t), \infty)\,dt = \infty \quad \text{implies} \quad P\left\{\limsup_{t\to 0} \frac{\gamma(t)}{h(t)} = \infty\right\} = 1. \end{aligned} \tag{4}$$

Now for $u \in (0, 1)$ it is easy to obtain the following bounds:

$$e^{-1} \log \frac{1}{u} \le \nu[u, \infty) \le \log \frac{1}{u} + e^{-1}.$$

Consequently,

$$\int_0^1 \nu[h(t), \infty)\,dt < \infty$$

if and only if

$$\int_0^1 \log \frac{1}{h(t)}\, dt < \infty.$$

This fact, together with the representation for the Dirichlet process mentioned above, allows us to rewrite (4) as

$$\int_0^{\varepsilon} \log h(t)\, dt > -\infty \quad \text{implies} \quad P\left\{\limsup_{t\to 0} \frac{F(t)}{h(\alpha(t))} = 0\right\} = 1$$

while

$$\int_0^{\varepsilon} \log h(t)\, dt = -\infty \quad \text{implies} \quad P\left\{\limsup_{t\to 0} \frac{F(t)}{h(\alpha(t))} = \infty\right\} = 1.$$

We thus see that the only constants that can appear on the right side of (2) are 0 and ∞. In this sense, it is impossible to give sharp upper bounds for the almost sure tail behavior of F (or at least not bounds when the function h above is convex). We will, therefore, exhibit a few functions h such that

$$\limsup_{\alpha(t)\to 0} \frac{F(t)}{h(\alpha(t))} = 0 \quad \text{a.s.} \quad [\mathscr{D}_\alpha].$$

Let $\delta > 0$. Define

$$h_\delta^1(t) = \exp\left\{-\frac{1}{t(|\log t|)^\delta}\right\}, \quad h_\delta^2(t) = \exp\left\{-\frac{1}{t(|\log t|)(\log|\log t|)^\delta}\right\},$$

$$h_\delta^3(t) = \exp\left\{-\frac{1}{t(|\log t|)(\log|\log t|)(\log\log|\log t|)^\delta}\right\},$$

and similarly define $h_\delta^k(t)$ for $k = 4, 5, \cdots,$. Let $k \in \{1, 2, 3, \cdots\}$ and let $\delta > 0$. The function h_δ^k is not well-defined on $(0, \infty)$. However, it will be defined and convex in a sufficiently small neighborhood of 0. Outside this neighborhood, we extend h_δ^k linearly so that it is convex on $(0, \infty)$.

Now for $k = 1, 2, 3, \cdots,$

$$\int_0^1 \log h_\delta^k(t)\, dt \begin{cases} > -\infty & \text{if} \quad \delta > 1, \\ = -\infty & \text{if} \quad \delta \le 1. \end{cases}$$

Thus, for $\delta > 1$, for almost every sample distribution function F,

$$F(t) \le h_\delta^k(\alpha_0(t)) \quad \text{for} \quad \alpha_0(t) \quad \text{sufficiently small,}$$

and

$$1 - F(t) \le h_\delta^k(1 - \alpha_0(t)) \quad \text{for} \quad 1 - \alpha_0(t) \quad \text{sufficiently small.}$$

Let us now find lower envelopes for the tails of F. According to Fristedt and Pruitt (1971, pages 174–175), for any increasing function h on $(0, \infty)$ either

$$P\left\{\liminf_{t\to 0} \frac{\gamma(t)}{h(t)} = 0\right\} = 1 \quad \text{or} \quad P\left\{\liminf_{t\to 0} \frac{\gamma(t)}{h(t)} = \infty\right\} = 1.$$

Thus, we cannot find sharp lower boundaries. It is an easy consequence of Lemma 4 of Fristedt and Pruitt (1971) that for any $\varepsilon > 0$,

$$\liminf_{t\to 0} \gamma(t)\exp\{(1+\varepsilon)\log|\log t|/t\} = \infty \quad \text{a.s.}$$

It is also an easy consequence of Lemma 5 of Fristedt and Pruitt (1971) that for any $\varepsilon > 0$

$$\liminf_{t\to 0} \gamma(t)\exp\{(1-\varepsilon)\log|\log t|/t\} = 0 \quad \text{a.s.}$$

Thus, for almost every sample distribution function,

$$F(t) \geq \exp\{-(1+\varepsilon)\log|\log \alpha_0(t)|/\alpha_0(t)\} \text{ for sufficiently small } \alpha_0(t), \quad \text{and}$$

$$1 - F(t) \geq \exp\{-(1+\varepsilon)\log|\log\{1-\alpha_0(t)\}|/\{1-\alpha_0(t)\}\} \text{ for sufficiently small } 1-\alpha_0(t).$$

EXAMPLE 1. If $\alpha_0(t) = 1 - e^{-t}$ for $t \geq 0$, then for every $\varepsilon > 0$, a.s. for sufficiently large t,

$$\exp\{-(1+\varepsilon)e^t \log t\} \leq 1 - F(t) \leq \exp[-e^t/t^2].$$

EXAMPLE 2. Let

$$f(x) = \begin{cases} \dfrac{c}{x^2(\log x)^2} & x \geq 2, \\ 0 & x < 2, \end{cases}$$

where c is a normalizing constant, and let α_0 be the distribution function with density f. Then

$$\int_{-\infty}^{\infty} x^\delta \, d\alpha_0(x) \begin{cases} = \infty & \text{if } \delta > 1 \\ < \infty & \text{if } 0 < \delta \leq 1. \end{cases}$$

However, it is easy to show that if F is distributed according to $\mathscr{D}_{\alpha_0}$, then F has a finite moment generating function with probability 1.

REMARK 1. In his 1974 paper, Ferguson states (Fact 4, page 617) that if g is a nonnegative measurable function, then $\int g(t)\, d\alpha(t) < \infty$ if and only if $\int g(t)\, dF(t) < \infty$ holds with probability 1. That is not correct. Our examples show that if $\int g(t)\, d\alpha(t) = \infty$, then $\int g(t)\, dF(t)$ may be finite with probability 1, although the expected value of $\int g(t)\, dF(t)$ will be infinite.

REMARK 2. Ferguson (1973) shows that if F is distributed according to $\mathscr{D}_\alpha$, then the posterior distribution of F given a sample $X_1, \cdots, X_n$ is the Dirichlet distribution with parameter $\alpha + \sum_{i=1}^n \delta_{X_i}$ where δ_a denotes the point mass at a. Using the representation of the Dirichlet process in terms of the gamma process (see equation (3)), it is clear that the upper and lower bounds on the prior tails of F also serve as upper and lower bounds on the posterior tails of F. Thus, no set of observations will change the order of magnitude of the tails of F.

Acknowledgment. We thank Professors Persi Diaconis and Thomas Ferguson for some useful comments.

REFERENCES

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

FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629.

FRISTEDT, B. (1967). Sample function behavior of increasing processes with stationary independent increments. *Pac. J. Math.* **21** 21–33.

FRISTEDT, B. E. and PRUITT, W. E. (1971). Lower functions for increasing random walks and subordinators. *Z. Wahrsch. verw. Gebiete* **18** 167–182.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306

MATHEMATICAL SCIENCES RESEARCH INSTITUTE
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720