UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

CONSTRUCTION OF PROCESS-DIFFERENTIABLE REPRESENTATIONS FOR PARAMETRIC FAMILIES OF DISTRIBUTIONS

Peter W. Glynn

Technical Summary Report #2972

February 1987

ABSTRACT

Given a parametric family of distributions $\{F(\theta,\cdot) : \theta \in (a,b)\}$, we consider the problem of constructing a process $\{X(\theta) : \theta \in (a,b)\}$ such that: i.) for all $\theta \in (a,b)$, $X(\theta)$ has distribution $F(\theta,\cdot)$, ii.) $X(\cdot)$ is differentiable in an L^1 sense at θ_0. This problem is motivated by certain computational questions associated with Monte Carlo estimation of the derivative of an expectation which depends on a parameter. A by-product of this work is a partial solution to the problem of constructing, for a given parametric family $\{F(\theta,\cdot)\}$, a function u such that $u(\theta,X(\theta_0))$ has distribution $F(\theta,\cdot)$.

AMS (MOS) Subject Classifications: 62F99, 65C05

Key Words: simulation, parametric statistics, Monte Carlo methods.

Work Unit Number 5 - Optimization and Large Scale Systems

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and by the National Science Foundation under Grant ECS-8404809.

SIGNIFICANCE AND EXPLANATION

Sensitivity analysis for stochastic systems plays an important role in the statistical study of such systems, and in the optimization of stochastic processes. In this paper, we consider questions related to the efficient computation of Monte Carlo estimates for the derivatives corresponding to such systems. These derivatives play a crucial part in the development of numerical schemes for accomplishing the statistical analysis and optimization mentioned above.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

CONSTRUCTION OF PROCESS-DIFFERENTIABLE REPRESENTATIONS FOR PARAMETRIC FAMILIES OF DISTRIBUTIONS

Peter W. Glynn

1. INTRODUCTION

For an index set $\Lambda = (a,b)$ $(-\infty \leq a < b \leq \infty)$, let $\{F(\theta,\cdot) : \theta \in \Lambda\}$ be an associated parametric family of distribution functions. Our goal in this paper is to consider, for $\theta_0 \in \Lambda$, the question of how to construct a probability space $(\Omega,\mathcal{F},P)$ and an associated process $\{X(\theta) : \theta \in \Lambda\}$ such that:

i.) $P\{X(\theta) \leq 0\} = F(\theta,\cdot)$ for all $\theta \in \Lambda$;

ii.) There exists an integrable r.v. $Y(\theta_0)$ such that

$$\delta X(\theta_0;h) \to Y(\theta_0) \tag{1.1}$$

as $h \to 0$ in $L^1(\Omega,\mathcal{F},P)$, where

$\delta X(\theta;h) = h^{-1}(X(\theta+h) - X(\theta))$.

We say that $\{F(\theta,\cdot) : \theta \in \Lambda\}$ admits the <u>process-differentiable representation</u> $\{X(\theta) : \theta \in \Lambda\}$ at $\theta = \theta_0$ if (1.1) holds; the r.v. $Y(\theta_0)$ is called the <u>derivative</u> of $X(\cdot)$ at $\theta = \theta_0$.

This problem is motivated by certain applications arising from Monte Carlo simulation. Specifically, suppose that

$$\alpha(\theta) = \int_{-\infty}^{\infty} r(\theta,x)F(\theta,dx) \tag{1.2}$$

is a parameter which is of interest in a statistical setting in which the "true" value θ^* of θ is unknown. Given an estimator $\{\theta_n : n \geq 1\}$ of θ^*, it frequently can be shown that there exists σ such that

$$n^{1/2}(\theta_n - \theta^*) \Longrightarrow \sigma N(0,1) \tag{1.3}$$

as $n \to \infty$; this typically occurs, for example, when θ_n is a maximum likelihood estimator of θ^* (see Chapter 2 of Ibragimov and Has'minskii (1981)). The natural estimator of the "true" value $\alpha(\theta^*)$ is given by $\alpha(\theta_n)$; $\alpha(\theta_n)$ can also be justified on the basis of

---

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and by the National Science Foundation under Grant ECS-8404809.

asymptotic efficiency principles (see p. 404 of Lehmann (1983)). If $\alpha(\cdot)$ is differentiable at θ^*, a standard "delta method" analysis (see p. 118 of Serfling (1980)) shows that (1.3) implies that

$$n^{1/2}(\alpha(\theta_n) - \alpha(\theta^*)) \Longrightarrow \sigma\alpha'(\theta^*)N(0,1)$$

as $n \to \infty$, where $\alpha'(\theta^*)$ is the derivative of $\alpha(\cdot)$ evaluated at $\theta = \theta^*$.

The asymptotic limit theory for the estimator $\alpha(\theta_n)$ therefore requires evaluation of $\alpha'(\theta^*)$ (or, more precisely, the quantity $\alpha'(\theta_n)$, which acts as an estimator of $\alpha'(\theta^*)$). In many settings, the quantity $\alpha'(\cdot)$ can not be evaluated analytically; numerical evaluation, via the Monte Carlo method, is a practical alternative.

The most obvious Monte Carlo strategy involves approximating $\alpha'(\theta_0)$ by a central difference of the form $\delta\alpha(\theta_0;h) = h^{-1}(\alpha(\theta_0 + h/2) - \alpha(\theta_0 - h/2))$. By generating independent variates $X(\theta_0 \pm h/2)$ with distribution $F(\theta_0 \pm h/2,\cdot)$, one can then estimate $\alpha(\theta_0 \pm h/2)$ via a sample mean of the r.v.'s $r(\theta_0 \pm h/2, X(\theta_0 \pm h/2))$. The key to obtaining a reasonable rate of convergence for this Monte Carlo estimator $\alpha_n'(\cdot)$ of $\alpha'(\cdot)$ is to choose the difference increment h to depend on the sample size n in an optimal fashion. However, as shown in Fox and Glynn (1987), even if $h = h_n$ is chosen optimally, the finite-difference estimator $\alpha_n'(\cdot)$ converges slowly, in the sense that $n^{1/3}(\alpha_n'(\cdot) - \alpha'(\cdot)) \Longrightarrow Z(\cdot)$ as $n \to \infty$, where $Z(\cdot)$ is a non-zero proper r.v. (i.e. roughly speaking, the rate of convergence of $\alpha_n'(\cdot)$ is of order $n^{-1/3}$). Since the rate of convergence typically enjoyed by Monte Carlo estimators is of order $n^{-1/2}$ (due to central limit theorem effects), this suggests that one should look for more suitable estimators.

A more sophisticated approach to the problem is available if $F(\theta,dx) = f(\theta,x)\mu(dx)$, where μ is the distribution of some r.v. W. In such a situation, $\alpha(\theta)$ can be expressed as

$$\alpha(\theta) = E\, r(\theta,W)f(\theta,W) \quad ; \tag{1.4}$$

if the derivative can be interchanged with the expectation, $\alpha'(\theta_0)$ can be estimated via a sample mean of r.v.'s of the form $\frac{\partial}{\partial x_1} r(\theta_0,W)f(\theta_0,W) + r(\theta_0,W)\frac{\partial}{\partial x_1} f(\theta_0,W)$. Under suitable moment hypotheses, this estimator converges to $\alpha'(\theta_0)$ at rate $n^{-1/2}$.

Frequently, $f(\theta,x)$ occurs as a likelihood ratio in a parametric family; as a consequence, this technique is called the likelihood ratio method for derivative estimation. (For further details, see Glynn (1986), Reiman and Weiss (1986), and Rubenstein (1986).)

Our interest here, however, will focus on a second approach. Suppose that $\{F(\theta,\cdot) : \theta \in \Lambda\}$ admits a process-differentiable representation at $\theta = \theta_0$. Then

$$\alpha(\theta) = E\, r(\theta,X(\theta))$$

for $\theta \in \Lambda$. Consequently, if the expectation and differentiation operations can be interchanged (this follows from (1.1) ii if, for some $\varepsilon > 0$, the partial derivatives of r are bounded on the set $\{(x_1,x_2) : |x_1 - \theta_0| < \varepsilon\}$), $\alpha'(\theta_0)$ can be estimated via a sample mean of the form $\frac{\partial}{\partial x_1} r(\theta_0,X(\theta_0)) + \frac{\partial}{\partial x_2} r(\theta_0,X(\theta_0))Y(\theta_0)$, where $Y(\theta_0)$ is the derivative of $X(\cdot)$ at $\theta = \theta_0$. Again, it is clear that under suitable moment hypotheses, this estimator converges to $\alpha'(\theta_0)$ at rate $n^{-1/2}$. Since this procedure involves defining r.v.'s $X(\theta)$ with appropriate marginals on a common probability space, this algorithm is known as a common probability space derivative estimator. As we shall see in Section 2, common probability space estimators can exist in settings where likelihood ratio estimates do not, and vice versa. Further relationships between the two approaches are explored in Glynn (1987).

Before concluding this section, it is worth pointing out that the problem (1.1) also arises naturally in a Monte Carlo systems simulation context. Specifically, the probabilistic dynamics of certain discrete-event systems (e.g. queueing systems) depend on various input distributions (e.g. inter-arrival and service time distributions). If the input distributions are allowed to depend on a parameter θ, then the corresponding steady-state α of the system depends on θ i.e. $\alpha = \alpha(\theta)$. In many design settings, the parameter θ can be viewed as a decision variable (e.g. the service rate in a queue may be subject to optimization); in such a setting, the design goal is to choose θ^* to minimize the steady-state cost α of running the system. Efficient numerical optimization algorithms for minimizing α frequently are based on evaluations of $\alpha'(\cdot)$. One efficient class of Monte Carlo procedures for calculating $\alpha'(\cdot)$ in such a discrete-event setting is the method of perturbation analysis (see Suri (1983)). The idea is to represent the stochastic systems corresponding to different values of θ on a common probability space, and to observe that for certain types of steady-state responses (e.g. waiting times in a queue), the corresponding ergodic sample mean is almost surely differentiable in the parameter θ. The path-wise derivative is then used to estimate $\alpha'(\cdot)$. In any case, the representation problem (1.1) arises naturally in this discrete-event setting (see, for example, the Appendix of Suri (1985).)

In Section 2, we describe the main results of this paper. Unless otherwise stated, the proofs of all results are deferred to Section 3.

2. DESCRIPTION OF MAIN RESULTS

The most obvious approach to constructing a process-differentiable representation for a given parametric family $\{F(\theta,\cdot) : \theta \in \Lambda\}$ is to use a simulation technique based on the method of common random numbers (see Rubinstein (1981), p. 224-229). The idea is based on the observation that if U is a uniform r.v. on $[0,1]$, then $F^{-1}(\theta,U)$ is a r.v. with distribution $F(\theta,\cdot)$, where $F^{-1}(\theta,\cdot)$ is the inverse distribution function defined by $F^{-1}(\theta,x) = \sup\{y : F(\theta,y) < x\}$. If one sets

$$X(\theta) = F^{-1}(\theta,U) \quad , \tag{2.1}$$

the hope is that $X(\theta)$ will be almost surely differentiable in θ, so that $Y(\theta_0)$ may be defined as $Y(\theta_0) = X'(\theta_0)$.

Although this approach is basically sound, it is not universally applicable. As our first result shows, there exist parametric families for which no process-differentiable representation exists.

(2.2) PROPOSITION. Let R be a set which contains no cluster points. Suppose that

$$F(\theta,x) = \sum_{k=1}^{\infty} p_k(\theta) I(x \geq y_k) \tag{2.3}$$

where $y_i \in R$ for all i. If $\{F(\theta,\cdot) : \theta \in \Lambda\}$ admits a process-differentiable representation at $\theta = \theta_0$, then it is necessary that $p_k'(\theta_0)$ exists and equals zero, for all $k \geq 1$.

This proposition basically says that process-differentiable representations do not exist for parametric families of the form (2.3). It is interesting to observe that (2.3) is amenable to likelihood ratio derivative estimation; one choice for $\mu(dx) = P\{W \in dx\}$ is given by

$$\mu((-\infty,x]) = \sum_{k=1}^{\infty} 2^{-k} I(x \geq y_k) \quad .$$

We will now show that there are parametric families $\{F(\theta,\cdot) : \theta \in \Lambda\}$ for which process-differentiable representations exist, and yet the stochastic process $\{X(\theta) : \theta \in \Lambda\}$ defined by the inversion formula (2.1) does not satisfy (1.1). In other words, the inversion representation (2.1) is in some sense, not always the "smoothest" way to stochastically represent a parametric family $\{F(\theta,\cdot) : \theta \in \Lambda\}$.

(2.4) EXAMPLE. For $\theta \in \mathbf{R}$, let

$$F(\theta,x) = \frac{3}{4} I(x \geq \theta) + \frac{1}{4} I(x \geq -\theta) \quad .$$

Set $V = I(U \leq 3/4)$. If we set $X(\theta) = V\theta + (1-V)(-\theta) = \theta(2V-1)$, then it is easily seen that $\{X(\theta) : \theta \in \mathbf{R}\}$ is a process-differentiable representation of $\{F(\theta,\cdot) : \theta \in \mathbf{R}\}$.

On the other hand, observe that for $\theta \geq 0$,

$$F^{-1}(\theta,x) = \begin{cases} -\theta \; ; & 0 \leq x \leq 1/4 \\ \theta \; ; & 1/4 < x \leq 1 \end{cases}$$

whereas

$$F^{-1}(-\theta,x) = \begin{cases} -\theta \; ; & 0 \leq x \leq 3/4 \\ \theta \; ; & 3/4 < x \leq 1 \quad . \end{cases}$$

As a consequence, we find that $F^{-1}(\cdot,x)$ is non-differentiable at $\theta = 0$ for $1/4 < x < 3/4$. Hence, the inversion representation (2.1) fails to satisfy (1.1) (since $Y(\theta_0)$ is assumed to be independent of how h approaches zero) at $\theta_0 = 0$.

Let us expand on this example further. Consider a parametric family of the form

(2.5) $$F(\theta,\cdot) = \sum_{k=1}^{\infty} p_k \, I(\cdot > y_k(\theta)) \quad .$$

Set

(2.6) $$X(\theta) = \sum_{k=1}^{\infty} y_k(\theta) I(P_{k-1} < U < P_k)$$

where $P_0 = 0$, $P_k = p_1 + \ldots + p_k$ for $k > 1$. Then, $\{X(\theta) : \theta \in \Lambda\}$ automatically satisfies (1.1) i; (1.1) ii is then satisfied if the $y_i(\theta)$'s are suitably regular in a neighborhood of θ_0. For example, (2.6) is a process-differentiable representation of $\{F(\theta,\cdot) : \theta \in \Lambda\}$ at θ_0 if the $y_i(\cdot)$'s are twice differentiable in a neighborhood of θ_0 and if there exists $\varepsilon > 0$ such that

$$\sup\{|y_k''(\theta)| : |\theta - \theta_0| < \varepsilon, \, k > 1\} < \infty \quad .$$

(To verify this result, just expand $y_k(\theta_0 + h)$ in a Taylor series about θ_0.) Observe that $X(\cdot) = F^{-1}(\cdot, U)$ a.s. only if $y_1(\theta) < y_2(\theta) < \ldots$ for θ in a neighborhood of θ_0. This confirms the idea that the inversion representation (2.1) breaks down in Example 2.4 because of the fact that the points θ and $-\theta$ reverse order as the parameter θ passes through zero.

Coincidentally, it is interesting to observe that likelihood ratio methods are generally inapplicable to parametric families of the form (2.5). Specifically, (2.5) permits the possibility that $F(\theta_0 + h, \cdot)$ and $F(\theta_0, \cdot)$ are mutually singular for all $h \neq 0$. In such a setting, it is certainly possible to define a "reference" distribution μ which permits likelihood ratio analysis along any sequence $h_n \to 0$; one possibility is

$$\mu(dx) = \sum_{h=1}^{\infty} 2^{-n} F(\theta_0 + h_n, \, dx) \quad .$$

The difficulty is that regardless of how one chooses μ, the densities $f(\theta_0 + h_n, \cdot)$ will be poorly behaved. Specifically, $f(\theta_0 + h_n, \cdot)$ can not converge to $f(\theta_0, \cdot)$, ruling out existence of a derivative. For suppose that $f(\theta_0 + h_n, x) \to f(\theta_0, x)$ μ a.e. as $n \to \infty$. By Scheffe's theorem,

(2.7) $$\int_{\mathbf{R}} |f(\theta_0 + h_n, x) - f(\theta_0, x)| \mu(dx) \to 0$$

as $n \to \infty$. But since $F(\theta_0, \cdot)$ and $F(\theta_0 + h_n, \cdot)$ are mutually singular, there exists a set A_n which supports $f(\theta_0, \cdot)$ and on which $f(\theta_0 + h_n, \cdot)$ vanishes. Clearly,

$$\int_{A_n} |f(\theta_0 + h_n, x) - f(\theta_0,x)|\mu(dx) = \int_{A_n} f(\theta_0,x)\mu(dx) = 1 \quad ,$$

contradicting (2.7).

Our third example shows that the inversion representation may not be the "smoothest" way to stochastically represent a parametric family $\{F(\theta,\cdot) : \theta \in \Lambda\}$, even if the $F(\theta,\cdot)$'s are continuous.

(2.8) EXAMPLE. Let $F(\theta,x) = P\{\theta N(0,1) \leq x\}$; clearly $X(\theta) = \theta N(0,1)$ is a process-differentiable representation of $\{F(\theta,\cdot) : \theta \in \mathbf{R}\}$.

Due to the symmetry of the normal, $F(\theta,\cdot) = F(-\theta,\cdot)$ and consequently $F^{-1}(\theta,\cdot) = F^{-1}(-\theta,\cdot)$. As a result, the inversion representation (2.1) yields a process $X(\theta)$ defined via

$$X(\theta) = |\theta| F^{-1}(1,U) \quad .$$

Clearly, the inversion representation fails to satisfy (1.1) at $\theta_0 = 0$. (Note, however, that if we relaxed (1.1) ii to permit different one-sided derivatives, this example would satisfy the relaxed definition.)

Despite the above examples, it seems fair to say that the inversion representation is generally well-behaved. Our next goal is therefore to show that the derivative $Y(\theta_0)$ of the inversion representation (2.1) can frequently be generated without having to explicitly generate r.v.'s via inverse transform methods. This is an important observation, since inversion can be computationally inefficient; other techniques, like acceptance-rejection (see p. 45 of [8]) are often more suitable for efficient variate generation.

(2.9) THEOREM. Suppose there exists $\varepsilon > 0$ for which:

i.) $F(\theta,dx) = f(\theta,x)I(a(\theta) < x < b(\theta))dx$

for $(\theta,x) \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \times \mathbf{R}$, where

$f(\cdot)$ is strictly positive

ii.) $F(\cdot)$ is continuously differentiable on

$\{(\theta,x) : |\theta-\theta_0| < \varepsilon, a(\theta) < x < b(\theta)\}$.

Then, for $0 < x < 1$, $\frac{\partial}{\partial x_1} F^{-1}(\theta_0,x)$ exists and equals $G(\theta_0,x) = -\frac{\frac{\partial}{\partial x_1} F(\theta_0,F^{-1}(\theta_0,x))}{\frac{\partial}{\partial x_2} F(\theta_0,F^{-1}(\theta_0,x))}$.

If $\{X(\theta) : \theta \in \Lambda\}$ is an inversion representation satisfying (1.1), then $Y(\theta_0)$ can be obtained as an a.s. limit of $\delta X(\theta_0;h_n)$ along some deterministic sub-sequence $h_n \to 0$ (see the proof of Proposition 2.2). It follows that $Y(\theta_0 = \frac{\partial}{\partial x_1} F^{-1}(\theta_0,U)$ a.s., yielding the next corollary.

(2.10) COROLLARY. Let $\{F(\theta,\cdot) : \theta \in \Lambda\}$ be a parametric family satisfying (2.9) i-ii. If $\{X(\theta) : \theta \in \Lambda\}$ is an inversion representation which is process-differentiable at θ_0, then $Y(\theta_0) = G(\theta_0,U)$ a.s.

The above formula for $G(\theta_0,x)$ was previously obtained, using formal methods, by Suri (1985). This theorem supplies an immediate solution to our problem. Since $F^{-1}(\theta_0,U) \overset{\mathcal{D}}{=} X(\theta_0)$ ($\overset{\mathcal{D}}{=}$ denotes equality in distribution),

$$(F^{-1}(\theta_0,U), \frac{d}{d\theta} F^{-1}(\theta_0,U)) \overset{\mathcal{D}}{=} (X(\theta_0), g(\theta_0,X(\theta_0))$$

where $g(\theta_0,x) = -\frac{\partial}{\partial x_1} F(\theta_0,x)/f(\theta_0,x)$. Consequently, to generate both $X(\theta_0)$ and its derivative $Y(\theta_0)$, one may generate $X(\theta_0)$ by one's method of choice, and calculate $Y(\theta_0)$ as $g(\theta_0,X(\theta_0))$. From Section 1, it is clear that the common probability space derivative estimation method can be subsequently applied to the random vector $(X(\theta_0), Y(\theta_0))$ thus obtained.

The above discussion suggests a second approach to obtaining process-differentiable representations for a parametric family $\{F(\theta,\cdot) : \theta \in \Lambda\}$. Specifically, let A be the set of x for which the differential equation

$$\frac{\partial}{\partial x_1} u(\theta,x) = g(\theta,u(\theta,x)) \tag{2.11}$$

subject to $u(\theta_0,x) = x$ has a unique solution $u(\cdot,x)$ on $|\theta-\theta_0| < \varepsilon$.

(2.12) PROPOSITION. Let $X(\theta_0)$ have distribution $F(\theta_0,\cdot)$. Assume the conditions of Theorem 2.9 and that $P\{X(\theta_0) \in A\} = 1$. (If A is non-measurable, replace it by a measurable subset.) Then, $u(\theta,X(\theta_0))$ has distribution $F(\theta,\cdot)$ for $|\theta-\theta_0| < \varepsilon$.

Rather than explicitly invert $F(\theta,\cdot)$ for all θ in the interval $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$, Proposition 2.12 allows us to construct a process satisfying (1.1) i by instead solving the differential equation (2.11). (Of course, in most applications, we expect the process to also satisfy (1.1) ii.) We illustrate this idea with an example.

(2.13) EXAMPLE. For $\theta > 0$, consider the parametric family

$$F(\theta,x) = \begin{cases} 0 & ; \ x < 1 \\ \ell n\, x/\theta & ; \ 1 \leq x \leq e^{\theta} \\ 1 & ; \ x > e^{\theta} \end{cases} .$$

Let $\theta_0 = 1$. Fixing x, we set $u(\cdot) = u(\cdot,x)$ and find that

$$\frac{du}{d\theta} = \frac{1}{\theta}\, u\, \ell n\, u$$

subject to $u(1) = x$. The solution of this differential equation is $u(\theta) = x^{\theta}$. Thus, $u(\theta,X(\theta_0)) = X(1)^{\theta}$. Note that $X(1)$ may be generated according to one's method of choice.

Proposition 2.12 solves, for a large class of parametric families, a certain "inverse" problem. It is well-known that many parametric families are derived from families of r.v.'s of the form $u(\theta,X)$ for some fixed r.v. X (e.g. scale and location families). Proposition 2.12 shows how to solve the inverse problem of constructing a function u and r.v. X which corresponds to a given parametric family.

We conclude this paper with our only uniqueness result for the construction problem (1.1). Before stating our results, it is worth noting that (1.1) determines <u>only</u> the marginal distributions of $\{X(\theta) : \theta \in \Lambda\}$.

(2.14) THEOREM. Let $\{X(\theta) : \theta \in \Lambda\}$ be a process-differentiable representation for $\{F(\theta) : \theta \in \Lambda\}$ at $\theta = \theta_0$. If $Y(\theta_0) \in \mathcal{B}(X(\theta_0))$, the distribution of $(X(\theta_0),Y(\theta_0))$ is uniquely determined.

As argued in the proof of Proposition 2.2, $Y(\theta_0)$ equals the a.s. limit of $\delta X(\theta_0;h_n)$ for some deterministic sequence $h_n \to 0$. Thus,

(2.15) $$Y(\theta_0) \in \bigwedge_{h>0} \mathcal{B}(X(\theta) : |\theta-\theta_0| < h) .$$

In many settings, we would hope that the σ-field in (2.15) is just $\mathcal{B}(X(\theta_0))$. Hence, the hypothesis on $Y(\theta_0)$ given in Theorem 2.14 seems reasonable. Not all derivatives $Y(\theta_0)$ need satisfy this measurability condition, however. For example, for the representation $X(\theta) = \theta N(0,1)$ used in (2.8), $Y(0) = N(0,1) \notin \mathcal{B}(0 \cdot N(0,1))$.

In any case, Theorem 2.14 shows that if $Y(\theta_0) \in \mathcal{B}(X(\theta_0))$, then the joint distribution of $\{X(\theta) : \theta \in \Lambda\}$ is constrained through its infinitesimal behavior at θ_0.

3. PROOFS

Proof of Proposition 2.2: Since $\delta X(\theta_0;h)$ converges to $Y(\theta_0)$ in $L^1(\Omega,\mathcal{F},P)$, it follows that there exists a deterministic subsequence $h_n \to 0$ such that $\delta X(\theta_0;h_n) \to Y(\theta_0)$ a.s. as $n \to \infty$ (see Theorem 4.2.3 of Chung (1974)). Consequently, $X(\theta_0 + h_n) \to X(\theta_0)$ a.s. as $n \to \infty$. Recall that (1.1) i implies that $X(\cdot) \in R$ a.s. Since R contains no cluster points, if therefore follows that for n sufficiently large, $X(\theta_0 + h_n)$ must equal $X(\theta_0)$ a.s. Then, $\delta X(\theta_0;h_n) = 0$ a.s. for n large, from which it is evident that $Y(\theta_0) = 0$ a.s.

Hence, by (1.1) ii, we may conclude that $E|\delta X(\theta_0;h)| \to 0$ as $h \to 0$. Observe that for $\theta_1, \theta_2 \in \Lambda$

$$E|X(\theta_1) - X(\theta_2)| \geq E\{\beta(X(\theta_i))I(X(\theta_1) \neq X(\theta_2))\} \tag{3.1}$$

for $i = 1,2$, where $\beta(x) = \inf\{|y-x| : y \in R\}$. Since R contains no cluster points, $\beta(x) > 0$ for all $x \in R$. From (3.1), it follows that for all $y \in R$,

$$E|X(\theta_0 + h) - X(\theta_0)| \geq \beta(y)P\{X(\theta_0) = y;\ X(\theta_0) \neq X(\theta_0 + h)\} \tag{3.2}$$

$$E|X(\theta_0 + h) - X(\theta_0)| \geq \beta(y)P\{X(\theta_0 + h) = y;\ X(\theta_0) \neq X(\theta_0 + h)\} \quad . \tag{3.3}$$

From (3.2) and the convergence of $E|\delta X(\theta_0;h)| \to 0$ as $h \to 0$, we find that

$$\begin{aligned} P\{X(\theta_0 + h) = y\} &\geq P\{X(\theta_0) = y;\ X(\theta_0) = X(\theta_0 + h)\} \\ &= P\{X(\theta_0) = y\} - P\{X(\theta_0) = y;\ X(\theta_0) \neq X(\theta_0 + h)\} \\ &= P\{X(\theta_0) = y\} - o(h) \quad . \end{aligned}$$

From (3.3), we similarly find that $P\{X(\theta_0) = y\} \geq P\{X(\theta_0 + h) = y\} - o(h)$. Combining, we find that $P\{X(\theta_0 + h) = y\} = P\{X(\theta_0) = y\} + o(h)$, from which it follows that $p_k'(\theta_0) = 0$ for all $k \geq 1$.

Proof of Theorem 2.9: The positivity of f implies that for $0 < x < 1$, $F^{-1}(\theta,x)$ is the unique solution of $F(\theta,F^{-1}(\theta,x)) = x$ for $|\theta-\theta_0| < \epsilon$. Then, for $0 < x < 1$,

$$(3.4)\qquad F(\theta_0 + h, F^{-1}(\theta_0 + h, x)) = F(\theta_0, F^{-1}(\theta_0,x)) \quad .$$

Applying Taylor's theorem to the left-hand side, we find that

$$(3.5)\qquad \frac{\partial}{\partial x_1} F(\xi,\eta)\cdot h + \frac{\partial}{\partial x_2} F(\xi,\eta)\cdot(F^{-1}(\theta_0 + h, x) - F^{-1}(\theta_0,x)) = 0 \quad ,$$

where (ξ,η) lies on the line segment joining $(\theta_0, F^{-1}(\theta_0,x))$ and $(\theta_0 + h, F^{-1}(\theta_0 + h, x))$. By the positivity of f and the continuity of F's partial derivatives, it is evident that $\frac{\partial}{\partial x_2} F(\theta_0,F^{-1}(\theta_0,x)) > 0$ for $0 < x < 1$. We therefore obtain the result by dividing appropriately in (3.5) and letting $h \to 0$.

Proof of Proposition 2.12: Note that $\ell(\theta,x) = F^{-1}(\theta, F(\theta_0,x))$ satisfies (2.11) for $0 < x < 1$. Thus, by definition of A, $\ell(\theta,x) = u(\theta,x)$ for $x \in A$, $|\theta-\theta_0) < \varepsilon$. Since $X(\theta_0) \in A$ a.s., we find that $u(\theta,X(\theta_0)) = \ell(\theta,X(\theta))$ a.s. But since $F(\theta_0,\cdot)$ is a continuous distribution function, $F(\theta_0,X(\theta_0)) \overset{\mathcal{D}}{=} U$, so $\ell(\theta,X(\theta)) \overset{\mathcal{D}}{=} F^{-1}(\theta,u)$, proving our result.

Proof of Theorem 2.14: Since $Y(\theta_0) \in \mathcal{B}(X(\theta_0))$, it follows that $Y(\theta_0) = h(X(\theta_0))$ for some Borel measurable h (depending, in general, on θ_0). For $t \in \mathbf{R}$, set $c(\theta,t) = E \exp(i + X(\theta))$ and observe that $c(\theta,t)$ is uniquely determined by $F(\theta,\cdot)$, which is given. Then,

$$(3.6)\qquad h^{-1}(c(\theta_0 + h,t) - c(\theta_0,t)) = h^{-1}E \exp(it\, X(\theta_0))\{\exp(it(X(\theta_0 + h) - X(\theta_0)) - 1\} \quad .$$

The bracketed term in (3.6) is dominated by $2|t| \cdot |X(\theta_0 + h) - X(\theta_0)|$. Condition (1.1) ii and the boundedness of $\exp(it\, X(\theta_0))$ permit one to pass the limit inside that expectation, showing that $\frac{\partial}{\partial x_1} c(\theta_0,t)$ exists and equals $E \exp(it\, X(\theta_0))Y(\theta_0)$.

Let $Y_i(\theta_0)$ $(i = 1,2)$ be the derivatives associated with two different representations of $\{F(\theta,\cdot) : \theta \in \Lambda\}$. Since $\frac{\partial}{\partial x_1} c(\theta_0,t)$ is uniquely specified, it follows that

$$E \exp(it\, X_1(\theta_0))Y_1(\theta_0) = E \exp(it\, X_2(\theta_0))Y_2(\theta_0) \quad . \tag{3.7}$$

Hence, if $Y_i(\theta_0) = h_i(X_i(\theta_0))$, we find that

$$\int_{-\infty}^{\infty} e^{itx}(h_1(x) - h_2(x))F(\theta_0,dx) = 0 \quad ,$$

i.e.

$$\int_{-\infty}^{\infty} e^{itx}K(dx) = 0$$

for all $t \in R$. Splitting K into its positive and negative parts, this shows that $\int_{-\infty}^{\infty} e^{itx}K_+(dx) = \int_{-\infty}^{\infty} e^{itx}K_-(dx)$. By the inversion theorem for characteristic functions of probability distributions, we may conclude that $K_+ = K_-$. Thus, $h_1 = h_2$ $F(\theta_0,\cdot)$ a.s., proving our result.

REFERENCES

1. Chung, K. L. (1974). A Course in Probability Theory. Academic Press, New York.

2. Fox, B. L. and Glynn, P. W. (1987). Replication schemes for estimating limiting expectations. Submitted for publication.

3. Glynn, P. W. (1986). Optimization of stochastic systems. Proceedings of the 1986 Winter Simulation Conference (Editors: J. R. Wilson, J. O. Henriksen, S. D. Roberts) 52-59.

4. Glynn, P. W. (1987). A relationship between the likelihood ratio and common random number derivative estimates. In preparation.

5. Ibragimov, I. A. and Has'minskii, R. A. (1981). Statistical Estimation - Asymptotic Theory. Springer-Verlag, New York.

6. Lehmann, E. L. (1983). Theory of Point Estimation. John Wiley, New York.

7. Reiman, M. I. and Weiss, A. (1986). Sensitivity analysis via likelihood ratios. Proceedings of the 1986 Winter Simulation Conference (Editors: J. R. Wilson, J. O. Henriksen, S. D. Roberts), 285-289.

8. Rubenstein, R. Y. (1981). Simulation and the Monte Carlo Method. John Wiley, New York.

9. Rubenstein, R. Y. (1986). Sensitivity analysis and performance extrapolation for computer simulation models. Technical Report, Division of Applied Sciences, Harvard University, Cambridge, MA.

10. Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley, New York.

11. Suri, R. (1983). Infinitesimal perturbation analysis of discrete-event dynamic systems: A general theory. Proc. 22nd IEEE Conference on Decision and Control (San Antonio, TX), 1030-1038.

12. Suri, R. (1987). Infinitesimal perturbation analysis for general discrete-event systems. To appear in Journal of the ACM.

PWG/jvs