Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

GENERALIZING THE GENERALIZED LIKELIHOOD RATIO METHOD THROUGH A PUSH-OUT LEIBNIZ INTEGRATION APPROACH

Xingyu Ren[1], and Michael C. Fu[1,2]

[1]Dept. of Electrical and Computer Eng. & Institute for System Research, University of Maryland, College Park, MD, USA
[2]Robert H. Smith School of Business, University of Maryland, College Park, MD, USA

ABSTRACT

We generalize the generalized likelihood ratio (GLR) method through a novel push-out Leibniz integration approach. Extending the conventional push-out likelihood ratio (LR) method, our approach allows the sample space to be parameter-dependent after the change of variables. Specifically, leveraging the Leibniz integral rule enables differentiation of the parameter-dependent sample space, resulting in a surface integral in addition to the usual LR estimator, which may necessitate additional simulation. Furthermore, our approach extends to cases where the change of variables only "locally" exists. Notably, the derived estimator includes existing GLR estimators as special cases and is applicable to a broader class of discontinuous sample performances. Moreover, the derivation is streamlined and more straightforward, and the requisite regularity conditions are easier to understand and verify.

1 INTRODUCTION

Consider an output sample performance parameterized by a real-valued scalar $\theta \in \Theta$:

$$\psi(X,\theta),$$

where Θ is an open interval, $\psi : \mathbb{R} \times \Theta \mapsto \mathbb{R}$ is a real-valued function, and X is the input random variable with density $f(x,\theta)$ and support $\Omega \subset \mathbb{R}$ (independent of θ). Suppose that we are interested in estimating the derivative of the expected sample performance with respect to (w.r.t.) θ:

$$\mathbb{E}(\psi(X,\theta)) = \int_\Omega \psi(x,\theta) f(x,\theta) dx.$$

Typical methods include infinitesimal perturbation analysis (IPA), smoothed perturbation analysis (SPA), the likelihood ratio (LR) method, and weak derivatives (WD) (Fu and Hu 1997; Glasserman 1991; Glynn 1987; Pflug 1996). Assume that ψ and f are differentiable w.r.t. θ, and density f is absolutely continuous w.r.t. a density $f_0 : \Omega \mapsto \mathbb{R}$ independent of θ. Under suitable conditions, we can interchange the order of differentiation and integration:

$$\frac{d}{d\theta}\mathbb{E}(\psi(X,\theta)) = \int_\Omega \frac{d}{d\theta}\left(\psi(x,\theta)\frac{f(x,\theta)}{f_0(x)}\right) f_0(x)dx = \int_\Omega \left(\partial_\theta \psi(x,\theta)h(x,\theta) + \psi(x,\theta)\partial_\theta h(x,\theta)\right) f_0(x)dx,$$

where $h(x,\theta) := f(x,\theta)/f_0(x)$ is the Radon-Nikodym derivative of f w.r.t. f_0. With X sampled from density f_0, $\partial_\theta \psi(X,\theta)h(X,\theta) + \psi(X,\theta)\partial_\theta h(X,\theta)$ is an example of the IPA-LR estimator (L'Ecuyer 1990), where $\partial_\theta \psi(X,\theta)h(X,\theta)$ and $\psi(X,\theta)\partial_\theta h(X,\theta)$ are IPA and LR estimators, respectively.

In some practical scenarios, ψ is not continuous w.r.t. θ (e.g., an indicator function), or not analytically available. Consequently, differentiation cannot be passed through integration, or the partial derivative of ψ may not even exist. Nevertheless, in some cases, through a change of variables, we can "push"

the parameter θ out of the function ψ, to circumvent the need to differentiate a discontinuous function (Rubinstein 1992; Wang et al. 2012). Specifically, assume that there exists a real-valued function $g(x,\theta)$ which is invertible w.r.t. x for each θ and differentiable w.r.t. both arguments, such that we can express $\psi(x,\theta) = \varphi(g(x,\theta))$ for some $\varphi : \mathbb{R} \mapsto \mathbb{R}$. Define a new random variable $Y = g(X,\theta)$, whose density is given by $\tilde{f}(y,\theta) = f(g^{-1}(y,\theta),\theta)\left|\partial_y g^{-1}(y,\theta)\right|$ supported on $\tilde{\Omega} \subset \mathbb{R}$. Make the change of variables:

$$\mathbb{E}(\psi(X,\theta)) = \int_{\tilde{\Omega}} \varphi(y)\tilde{f}(y,\theta)dy = \mathbb{E}(\varphi(Y)),$$

and the LR method applies. Peng et al. (2018), Peng et al. (2020) extend the push-out LR method to scenarios where g is only locally invertible (i.e., its Jacobian matrix J_g is invertible).

Note that the push-out LR method typically requires the support $\tilde{\Omega}$ of Y to be independent of θ. Consider a toy example $\psi(X,\theta) = \mathbf{1}\{X < \theta\}$, where X follows an exponential distribution with parameter $\theta > 0$, having density $f_X(x,\theta) = \theta e^{-\theta x}$ over the support $\Omega = [0,\infty)$. The expected sample performance can be expressed as:

$$\mathbb{E}(\mathbf{1}\{X < \theta\}) = \int_0^\infty \mathbf{1}\{x < \theta\}\theta e^{-\theta x}dx.$$

To apply the push-out LR method, we set $Y = \frac{X}{\theta}$, which follows an exponential distribution with parameter θ^2, with density $f_Y(y,\theta) = \theta^2 e^{-\theta^2 y}$ over the support $\tilde{\Omega} = [0,\infty)$. Make the change of variables:

$$\mathbb{E}(\mathbf{1}\{X < \theta\}) = \int_0^\infty \mathbf{1}\{x < \theta\}\theta e^{-\theta x}dx = \int_0^\infty \mathbf{1}\{y < 1\}\theta^2 e^{-\theta^2 y}dy = \mathbb{E}(\mathbf{1}\{Y < 1\}).$$

Since both the new sample performance $\varphi(y) = \mathbf{1}\{y < 1\}$ and the support of Y are independent of θ, we can apply the LR method w.r.t. Y:

$$\frac{d}{d\theta}\mathbb{E}(\mathbf{1}\{X < \theta\}) = \frac{d}{d\theta}\mathbb{E}(\mathbf{1}\{Y < 1\}) = \mathbb{E}(\mathbf{1}\{Y < 1\}\partial_\theta \log f_Y(Y,\theta)),$$

where $\mathbf{1}\{Y < 1\}\partial_\theta \log f_Y(Y,\theta) = \mathbf{1}\{Y < 1\}(2/\theta - 2\theta Y)$ is an unbiased derivative estimator.

Instead of setting $Y = \frac{X}{\theta}$, an alternative approach to remove θ from the indicator function is to set $Z = X - \theta$. This creates a shifted exponential random variable with density function $f_Z(z,\theta) = \theta e^{-\theta(z+\theta)}$ over the support $[-\theta,\infty)$. Due to the dependence of the support on θ, the LR method cannot be directly applied. However, if we write the integral as

$$\mathbb{E}(\mathbf{1}\{X < \theta\}) = \int_{-\theta}^\infty \mathbf{1}\{z < 0\}\theta e^{-\theta(z+\theta)}dz,$$

we can apply the Leibniz integral rule to differentiate both the lower limit and the integrand simultaneously:

$$\begin{aligned}
&\frac{d}{d\theta}\mathbb{E}(\mathbf{1}\{X < \theta\}) = \frac{d}{d\theta}\int_{-\theta}^\infty \mathbf{1}\{z < 0\}\theta e^{-\theta(z+\theta)}dz \\
&= \int_{-\theta}^\infty \mathbf{1}\{z < 0\}\frac{d}{d\theta}\theta e^{-\theta(z+\theta)}dz - \mathbf{1}\{z < 0\}\theta e^{-\theta(z+\theta)}\big|_{z=-\theta}\frac{d}{d\theta}(-\theta) = \mathbb{E}(\mathbf{1}\{Z < 0\}\partial_\theta \log f_Z(Z,\theta)) + \theta,
\end{aligned}$$

which leads to a standard LR estimator augmented by an extra constant term θ arising from the differentiation w.r.t. the lower limit.

This example suggests that leveraging the Leibniz integral rule extends the applicability of the push-out LR method to broader settings where the support of the newly introduced random variable may depend on the parameter. Furthermore, despite the simplicity of this example, it falls outside the scope of the generalized

LR (GLR) methods proposed by Peng et al. (2018) and Peng et al. (2020), which require either the density function to vanish at the boundary of the support or the input random variables to follow a uniform distribution. In this paper, we will explore the integration of the push-out LR method with the Leibniz integral rule for an output sample performance of the form $\varphi(g(X,\theta))$, where X is a random vector and a change of variables $Y = g(X,\theta)$ removes the parameter from φ. A similar idea is proposed under different regularity conditions by Puchhammer and L'Ecuyer (2022), which focuses on density estimation. The rest of this paper is organized as follows. In Section 2, we formally define the output sample performance and introduce a general form of the Leibniz integral rule for multivariate integrals, subsequently applying it to the sample performance. Specifically, we demonstrate that the new estimator includes the existing GLR estimators as special cases. In Section 3, we extend results in Section 2 to cases where the function $g(x,\theta)$ is only locally invertible w.r.t. x and the sample space is unbounded. Section 4 presents simulation results on the example from Section 1. Section 5 offers conclusions and future research directions.

2 INTEGRATING THE LEIBNIZ INTEGRAL RULE WITH THE PUSH-OUT LR METHOD

Consider an output sample performance $\varphi(g(X,\theta))$, where

- $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ is a bounded measurable function.
- $g(\cdot,\cdot) : \mathbb{R}^n \times \Theta \mapsto \mathbb{R}^n$ is twice continuously differentiable w.r.t. both arguments. For each θ, $g(x,\theta)$ is an invertible function of x. $\Theta \subset \mathbb{R}$ is a bounded open interval.
- X is an $n-$dimensional random vector with bounded support $\Omega \subset \mathbb{R}^n$ (the boundedness condition is relaxed in Section 3).
- X has a density function $f(\cdot,\cdot) : \Omega \times \Theta \mapsto \mathbb{R}$, continuously differentiable w.r.t. both arguments.

Making the change of variables $y = g(x,\theta)$, we can write

$$\mathbb{E}(\varphi(g(X,\theta))) = \int_{g(\Omega,\theta)} \varphi(y) f(g^{-1}(y,\theta),\theta) |\det(J_{g^{-1}}(y,\theta))| dy, \tag{1}$$

where $g(\Omega,\theta)$ is the image of Ω under map g, and $J_{g^{-1}}(y,\theta)$ is the Jacobian matrix of g^{-1} w.r.t. y, i.e., $\{J_{g^{-1}}(y,\theta)\}_{ij} = \partial_{y_j} g_i^{-1}(y,\theta)$. Both the integrand and the domain of integration in Equation (1) involve the parameter θ. The following result introduces the Leibniz integral rule that enables differentiation of the domain w.r.t. θ. Theorem 1 is a special case of the Leibniz integral rule proved in Section 7 and 8 of Flanders (1973), and a more general version is available in Amann et al. (2005).

Theorem 1 Let $D_\theta \subset \mathbb{R}^n$ be a compact set. Suppose that there exists a function $\phi(\cdot,\cdot) : U \times \Theta \mapsto \mathbb{R}^n$, where $U \subset \mathbb{R}^n$ is a fixed domain, such that $D_\theta = \phi(U,\theta)$. Suppose $\phi(\cdot,\cdot) : \mathbb{R}^n \times \Theta \mapsto \mathbb{R}^n$ is twice continuously differentiable in both arguments, and for each θ, $\phi(x,\theta)$ is an invertible function of x. Then, for any function $f(\cdot,\cdot) : \mathbb{R}^n \times \Theta \mapsto \mathbb{R}$ continuously differentiable in both arguments,

$$\frac{d}{d\theta} \int_{D_\theta} f(x,\theta) dx = \int_{D_\theta} (\partial_\theta f(x,\theta) + \operatorname{div}(f(x,\theta)\vec{v}(x))) \, dx,$$

where div is the divergence operator, i.e., $\operatorname{div}(F) = \sum_{i=1}^n \partial_{x_i} F_i$, $F : \mathbb{R}^n \mapsto \mathbb{R}^n$, and $\vec{v}(x) = \partial_\theta \phi(u,\theta)|_{u=\phi^{-1}(x,\theta)}$. In particular, by the divergence theorem (Zorich 2004b), we can write

$$\int_{D_\theta} \operatorname{div}(f(x,\theta)\vec{v}(x)) dx = \int_{\partial D_\theta} f(x,\theta)\vec{v}(x)^T \vec{n}(x) ds,$$

where ∂D_θ is the boundary of D_θ, $\vec{n}(x)$ is the outward normal vector on surface ∂D_θ, and ds is the area element.

The Leibniz integral rule in $\mathbb{R}^n$ is more intricate than in $\mathbb{R}$, as the boundary of the integral domain is an $(n-1)$-dimensional "moving" surface, instead of endpoints of an interval. In fluid mechanics, the Leibniz

Figure 1: The original domain U_θ and the perturbed domain $U_{\theta+\Delta\theta}$.

integral rule is also known as the transport theorem (Frankel 2011). We provide a "physical" interpretation of the Leibniz integral rule in $\mathbb{R}^2$.

Suppose $U_\theta \subset \mathbb{R}^2$ is a domain with a smooth boundary and $F : \mathbb{R}^2 \mapsto \mathbb{R}$ is a smooth function. We are interested in computing $\frac{d}{d\theta}\int_{U_\theta} F(x,y)dxdy$. For small $\Delta\theta$, suppose the domain U_θ moves to $U_{\theta+\Delta\theta}$, as shown in Figure 1. As in Theorem 1, we assume that U_θ is characterized by a smooth function $\phi : \mathbb{R}^2 \times \Theta \mapsto \mathbb{R}^2$ and a fixed domain $U \subset \mathbb{R}^2$, i.e., $U_\theta = \phi(U,\theta)$. Consider the difference $\int_{U_{\theta+\Delta\theta}} F(x,y)dxdy - \int_{U_\theta} F(x,y)dxdy$. The integral over the intersection $U_{\theta+\Delta\theta} \cap U_\theta$ cancels out, leaving only two strips surrounding the boundary ∂U_θ contributing to the difference. We zoom in on a small segment of this strip around a point $x \in \partial U_\theta$, illustrated by the blue region in Figure 1. Here, ds is the arc length element, $\vec{n}$ is the normal vector of the boundary ∂U_θ at x, and $\vec{v}$ is the velocity vector of domain w.r.t. θ, given by $\partial_\theta \phi(u,\theta)|_{u=\phi^{-1}(x,\theta)}$. For sufficiently small $\Delta\theta$ and ds, this region is approximately a rectangle of length ds and width $\vec{v}\cdot\vec{n}\Delta\theta$, the displacement of the domain along the normal vector. Therefore, the area of the blue region is $\vec{v}\cdot\vec{n}\Delta\theta ds$, and

$$\frac{d}{d\theta}\int_{U_\theta} F(x,y)dxdy = \lim_{\Delta\theta\to 0}\frac{1}{\Delta\theta}\left(\int_{U_{\theta+\Delta\theta}} F(x,y)dxdy - \int_{U_\theta} F(x,y)dxdy\right) = \int_{\partial U_\theta} F(x,y)\vec{v}\cdot\vec{n}ds.$$

Notice that Theorem 1 requires that the domain D_θ to be parameterized by a sufficiently smooth function $\phi(u,\theta)$ defined on a fixed set U. In our formulation, these correspond to the function g and the sample space Ω. However, Theorem 1 also requires the integrand to be differentiable, a condition that φ may not satisfy. To address this, we can approximate φ by smooth functions.

Proposition 1 Compactly supported smooth functions are dense in $L^p(\mathbb{R}^n)$, $1 \le p < \infty$ and $C(\mathbb{R}^n)$ (the space of continuous functions on $\mathbb{R}^n$).

See Peng et al. (2018) and Section 8.2 in Folland (1999) for the proof and a method for constructing smooth approximations via convolution with mollifiers. As both Ω and Θ are bounded sets, the set $g(\Omega,\Theta) := \{y \in \mathbb{R}^n \mid y = g(x,\theta),\ (x,\theta) \in \Omega \times \Theta\}$ is also bounded. In our problem formulation, we can restrict φ to this bounded set $g(\Omega,\Theta)$. Since φ is bounded, it is integrable over $g(\Omega,\Theta)$. By Proposition 1, there exists a sequence of smooth functions $\{\varphi_n\}_{n\in\mathbb{N}}$ such that $\varphi_n \to \varphi$ in L^1 as $n \to \infty$. Substituting $\varphi_n,\ n \in \mathbb{N}$ into Equation (1), we can apply Theorem 1:

$$\begin{aligned}\frac{d}{d\theta}\mathbb{E}(\varphi_n(g(X,\theta))) &= \frac{d}{d\theta}\int_{g(\Omega,\theta)} \varphi_n(y) f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|dy \\ &= \int_{g(\Omega,\theta)} \varphi_n(y)\frac{d}{d\theta}(f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|)dy \qquad (2) \\ &+ \int_{g(\Omega,\theta)} \operatorname{div}(\varphi_n(y) f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|\vec{v}(y))dy, \qquad (3)\end{aligned}$$

where $\vec{v}(y) = \partial_\theta g(x,\theta)|_{x=g^{-1}(y,\theta)}$. Notice that for each θ, $g(\cdot,\theta) : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a diffeomorphism, defined as follows (Zorich 2004a).

Definition 1 A mapping $f : U \mapsto V$, where U, V are open subsets of $\mathbb{R}^m$, is a diffeomorphism of order p if f is p-times continuously differentiable, f is a bijection, and $f^{-1} : V \mapsto U$ is p-times continuously differentiable.

The fact that g is a diffeomorphism directly follows from the following inverse function theorem which is very useful for establishing Proposition 2.

Lemma 1 Suppose a mapping $f : G \mapsto \mathbb{R}^m$ of a domain $G \subset \mathbb{R}^m$ is such that f is p-times continuously differentiable, $y_0 = f(x_0)$ at some $x_0 \in G$, and the Jacobian matrix $J_f(x_0)$ invertible. Then there exists a neighborhood $U(x_0) \subset G$ of x_0 and a neighborhood $V(y_0)$ of y_0 such that $f : U(x_0) \mapsto V(y_0)$ is a diffeomorphism of order p. Moreover, if $x \in U(x_0)$ and $y = f(x) \in V(y_0)$, then $J_{f^{-1}}(y) = J_f^{-1}(x)$.

See Section 8.6 in Zorich (2004a) for a proof of Lemma 1. Notice that the image set $g(\Omega, \theta)$ can be complex in high-dimensional spaces, and in some cases, the function g doesn't have a closed-form inverse. Specifically, in Section 3, we study the generalized scenario where g is only locally invertible, meaning there is no global change of variables $y = g(x, \theta)$. Therefore, we would like to reverse the change of variables.

Proposition 2 For $y = g(x, \theta)$, the following equations hold:

$$\frac{d}{d\theta}(f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|) = |\det(J_{g^{-1}}(y,\theta))|(d(x,\theta)+l(x,\theta))f(x,\theta), \tag{4}$$

$$\operatorname{div}(\varphi_n(y)f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|\vec{v}(y)) = |\det(J_{g^{-1}}(y,\theta))|\operatorname{div}(\varphi_n(g(x,\theta))f(x,\theta)s(x,\theta)), \tag{5}$$

where $d(x,\theta) = \operatorname{div}(-f(x,\theta)J_g^{-1}(x,\theta)\partial_\theta g(x,\theta))/f(x,\theta)$ and $l(x,\theta) = \partial_\theta \log f(x,\theta)$ are real-valued functions, and $s(x,\theta) = J_g^{-1}(x,\theta)\partial_\theta g(x,\theta)$ is an $n-$dimensional vector-valued function.

See Appendix A for the proof. To reverse the change of variables, we substitute (4) and (5) into (2) and (3), respectively:

$$\int_{g(\Omega,\theta)} \varphi_n(y)\frac{d}{d\theta}(f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|)dy = \int_\Omega \varphi_n(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx,$$

$$\int_{g(\Omega,\theta)} \operatorname{div}(\varphi_n(y)f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|\vec{v}(y))dy = \int_\Omega \operatorname{div}(\varphi_n(g(x,\theta))f(x,\theta)s(x,\theta))dx.$$

By the divergence theorem, $\int_\Omega \operatorname{div}(\varphi_n(g(x,\theta))f(x,\theta)s(x,\theta))dx = \int_{\partial\Omega} \varphi_n(g(x,\theta))s(x,\theta)^T\vec{n}(x)f(x,\theta)ds$, where $\vec{n}(x)$ is the outward normal vector on surface $\partial\Omega$ (Zorich 2004a). To summarize, we can write

$$\frac{d}{d\theta}\mathbb{E}(\varphi_n(g(X,\theta))) = \int_\Omega \varphi_n(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx + \int_{\partial\Omega} \varphi_n(g(x,\theta))s(x,\theta)^T\vec{n}(x)f(x,\theta)ds. \tag{6}$$

Under suitable conditions, $\frac{d}{d\theta}\mathbb{E}(\varphi_n(g(X,\theta)))$ converges to $\frac{d}{d\theta}\mathbb{E}(\varphi(g(X,\theta)))$ as $n \to \infty$.

Theorem 2 If $\lim_{n\to\infty}\int_{\partial\Omega}\sup_{\theta\in\Theta}|(\varphi(g(x,\theta)) - \varphi_n(g(x,\theta)))s(x,\theta)^T\vec{n}(x)f(x,\theta)|ds = 0$, then

$$\frac{d}{d\theta}\mathbb{E}(\varphi(g(X,\theta))) = \int_\Omega \varphi(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx + \int_{\partial\Omega} \varphi(g(x,\theta))s(x,\theta)^T\vec{n}(x)f(x,\theta)ds, \tag{7}$$

where $d(x,\theta) = \operatorname{div}(-f(x,\theta)s(x,\theta))/f(x,\theta)$, $l(x,\theta) = \partial_\theta \log f(x,\theta)$, and $s(x,\theta) = J_g^{-1}(x,\theta)\partial_\theta g(x,\theta)$.

See Appendix B for the proof. Theorem 2 can be extended to functions $g(\cdot,\cdot) : \mathbb{R}^m \times \Theta \mapsto \mathbb{R}^n$ and $\varphi : \mathbb{R}^m \mapsto \mathbb{R}$, where $m < n$, by replacing $J_g^{-1}(x,\theta)$ with an $m \times m$ invertible submatrix of it (Peng et al. 2018). In Equation (7), the first integral on the right-hand side is derived by differentiating the density of the random variable $Y = g(X,\theta)$ and reversing the change of variables. An unbiased gradient estimator for it is given by

$$\varphi(g(X,\theta))(d(X,\theta)+l(X,\theta)). \tag{8}$$

The second term in Equation (7) is a surface integral that arises from differentiating the domain $g(\Omega,\theta)$ w.r.t. θ. If the domain $g(\Omega,\theta)$ does not depend on θ, the surface integral vanishes. In general, computing the surface integral is challenging unless the surface can be parameterized and the normal vector has a closed-form expression. However, for certain special forms of Ω and φ, the surface integral can be converted into a regular integral that is easier to handle.

2.1 Rectangle Support

Consider the case where $\Omega = [a_1,b_1]\times\cdots\times[a_n,b_n]$, a hyperrectangle in $\mathbb{R}^n$, with boundary given by $\partial\Omega = \cup_{i=1}^n(\Omega_{a_i}\cup\Omega_{b_i})$, a union of surfaces where

$$\Omega_{a_i} := [a_1,b_1]\times\cdots\times\{a_i\}\times\cdots\times[a_n,b_n],\ \Omega_{b_i} := [a_1,b_1]\times\cdots\times\{b_i\}\times\cdots\times[a_n,b_n].$$

For each i, the normal vector $\vec{n}(x)$ for surfaces Ω_{a_i} and Ω_{b_i} are $-e_i$ and e_i, respectively, where $e_i\in\mathbb{R}^n$ is the unit vector with i^{th} component to be one. The surface integral over each Ω_{a_i} and Ω_{b_i} reduces to a standard multivariate integral:

$$\begin{aligned}
&\int_{\partial\Omega}\varphi(g(x,\theta))s(x,\theta)^T\vec{n}(x)f(x,\theta)ds\\
&=\sum_{i=1}^n\int_{x_j\in[a_j,b_j],j=1,\ldots,n,j\neq i}\varphi(g(x,\theta))s(x,\theta)^Te_if(x,\theta)\prod_{j=1,\ldots,n,j\neq i}dx_j\Big|_{x_i=a_i}^{b_i}\\
&=\sum_{i=1}^n(\mathbb{E}(\varphi(g(X,\theta))s(X,\theta)^Te_i|X_i=b_i)f_{X_i}(b_i)-\mathbb{E}(\varphi(g(X,\theta))s(X,\theta)^Te_i|X_i=a_i)f_{X_i}(a_i)),
\end{aligned}$$

where f_{X_i} is the marginal density of X_i. An unbiased gradient estimator for the surface integral is given by

$$\sum_{i=1}^n(\varphi(g(X,\theta))f_{X_i}(b_i)s(X,\theta)^Te_i\big|_{X\sim f_{X|X_i=b_i}}-\varphi(g(X,\theta))f_{X_i}(a_i)s(X,\theta)^Te_i\big|_{X\sim f_{X|X_i=a_i}}),\tag{9}$$

where $f_{X|X_i}$ is the conditional density of X given X_i. In particular, if $\{X_i\}_{i=1,\ldots,n}$ are independent, estimator (9) simplifies to

$$\sum_{i=1}^n(\varphi(g(X,\theta))f_{X_i}(b_i)s(X,\theta)^Te_i\big|_{X_i=b_i}-\varphi(g(X,\theta))f_{X_i}(a_i)s(X,\theta)^Te_i\big|_{X_i=a_i}),$$

which can be simulated by a single sample path, concurrently with estimator (8). Peng et al. (2020) studies a special case where the input consists of an independent sequence of uniform random variables. Another special case occurs when the density function vanishes at the boundary of the support (Peng et al. 2018). For the latter case, the marginal densities $f_{X_i}(a_i)$ and $f_{X_i}(b_i)$ are zero, resulting in the surface integral vanishing, as well.

2.2 Almost Everywhere (a.e.) Differentiable φ

For an a.e. differentiable function $F:\mathbb{R}^n\mapsto\mathbb{R}^n$ with set of discontinuities D_F, the divergence theorem holds under certain conditions (Shapiro 1958). Suppose $\Gamma\subset\mathbb{R}^n$ is a bounded set and its boundary $\partial\Gamma$ is a simple closed curve. Then $\int_\Gamma\operatorname{div}(F(y))dy=\int_{\partial\Gamma}F(y)^T\vec{n}(y)dy$ holds on Γ if the following conditions hold:

- F is continuous on closure$(\Gamma)\setminus D_F$ and is L^2-integrable on Γ.
- $\operatorname{div}F$ exists a.e. and is integrable on Γ.
- div_*F and div^*F are finite on $\Gamma\setminus D_F$, with

$$\operatorname{div}_*F(y):=\liminf_{t\to0}\frac{1}{\operatorname{vol}(B(y,t))}\int_{\partial B(y,t)}F(y)^T\vec{n}(y)dy,$$

where $B(y,t) = \{y' \in \mathbb{R}^n \mid \|y'-y\|_\infty < t\}$ is an open ball centered at y with radius t, and $\text{vol}(B(y,t))$ is its $n-$dimensional volume. $\text{div}^* F$ is defined similarly by replacing lim inf with lim sup.

- The set D_F has logarithmic capacity zero if $n=2$, or Newtonian capacity zero if $n \geq 3$. For a compact set K, the logarithmic capacity is given by $\exp\left(-\min_\mu \int_K \int_K \log(|x-y|^{-1}) d\mu(x) d\mu(y)\right)$, and the Newtonian capacity is given by $\left(\min_\mu \int_K \int_K |x-y|^{-(n-2)} d\mu(x) d\mu(y)\right)^{-1}$, where the minimum is taken over all Borel probability measures on K (Landkof 1972).

Notice that the condition "Newtonian capacity zero" is stronger than the condition "measure zero". For example, in $\mathbb{R}^3$, both a two-dimensional disk and a line segment have Lebesgue measure zero. However, the line segment has zero Newtonian capacity, whereas the two-dimensional disk has a positive capacity (Landkof 1972).

Suppose that φ is bounded and differentiable a.e. except on a set of capacity zero. Since functions g, f and s are continuously differentiable, the divergence theorem holds:

$$\int_\Omega \text{div}(\varphi(g(x,\theta))s(x,\theta)f(x,\theta))dx = \int_{\partial\Omega} \varphi(g(x,\theta))s(x,\theta)^T \vec{n}(x) f(x,\theta)dx.$$

Clearly, $\text{div}(\varphi(g(X,\theta))s(X,\theta)f(X,\theta))/f(X,\theta)$ is an unbiased estimator for the surface integral. Combined with estimator (8), we obtain a single-run unbiased estimator for $\frac{d}{d\theta}\mathbb{E}(\varphi(g(X,\theta)))$:

$$\varphi(g(X,\theta))(d(X,\theta)+l(X,\theta)) + \text{div}(\varphi(g(X,\theta))s(X,\theta)f(X,\theta))/f(X,\theta).$$

3 LOCAL CHANGE OF VARIABLES AND UNBOUNDED SAMPLE SPACE

In this section, we relax the condition for g to be invertible everywhere and instead consider it being locally invertible. Specifically, we only assume that its Jacobian matrix J_g is invertible a.e., which is a necessary but not sufficient condition for global invertibility. By Lemma 1, except on a set of measure zero, for each $x \in \Omega$, there exists a bounded open neighborhood $U(x)$ of x, such that $g(\cdot,\theta)$ is invertible on $U(x)$. Since Ω is bounded, by the Heine-Borel theorem, there exists a finite collection of open neighborhoods $\{U_i\}_{i=1,\cdots,N}$, such that $\text{closure}(\Omega) \subset \cup_{i=1}^N U_i$. For each i, we can derive a "local" version of Equation (6) over $\Omega \cap U_i$:

$$\begin{aligned}
&\frac{d}{d\theta}\mathbb{E}(\varphi_n(g(X,\theta))\mathbf{1}\{X \in (\Omega \cap U_i)\}) \\
=& \int_{\Omega\cap U_i} \varphi_n(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta) + \text{div}(\varphi_n(g(x,\theta))s(x,\theta)f(x,\theta))dx.
\end{aligned}$$

Combining all the open sets $\{U_i\}_{i=1,\cdots,N}$, we can reconstruct Equation (6) over the entire sample space Ω:

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{E}(\varphi_n(g(X,\theta))) &= \sum_{i=1}^N \int_{\Omega\cap U_i} \varphi_n(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta) + \text{div}(\varphi_n(g(x,\theta))s(x,\theta)f(x,\theta))dx \\
&= \int_\Omega \varphi_n(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta) + \text{div}(\varphi_n(g(x,\theta))s(x,\theta)f(x,\theta))dx \\
&= \int_\Omega \varphi_n(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx + \int_{\partial\Omega} \varphi_n(g(x,\theta))s(x,\theta)^T\vec{n}(x)f(x,\theta)ds.
\end{aligned}$$

Therefore, the proof of Theorem 2 still holds for locally invertible function g.

In addition to the local change of variables, Theorem 2 can be extended to the unbounded sample space Ω under appropriate conditions. We provide a brief outline of this extension, leaving the detailed exploration to future research. Consider $\Omega_L := \Omega \cap [-L,L]^n$, the restriction of Ω to the hyperrectangle $[-L,L]^n$. For a fixed $L > 0$, by Proposition 1, there exists a sequence of smooth functions $\{\varphi_{n,L}\}_{n\in\mathbb{N}}$ such

that $\varphi_{n,L} \to \varphi$ in L^1 as $n \to \infty$ over the compact set $g(\Omega_L, \Theta)$. Hence, we can reconstruct Equation (6) over Ω_L:

$$\begin{aligned}&\frac{d}{d\theta}\mathbb{E}(\varphi_{n,L}(g(X,\theta))\mathbf{1}\{X \in \Omega_L\}) \\ =&\int_{\Omega_L} \varphi_{n,L}(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx + \int_{\partial\Omega_L} \varphi_{n,L}(g(x,\theta))s(x,\theta)^T\vec{n}(x)f(x,\theta)ds.\end{aligned}$$

Our goal is to show $\lim_{n\to\infty} \frac{d}{d\theta}\mathbb{E}(\varphi_{n,L}(g(X,\theta))) = \frac{d}{d\theta}\mathbb{E}(\varphi(g(X,\theta))\mathbf{1}\{X \in \Omega_L\})$. By Theorem 2, a sufficient condition is $\lim_{n\to\infty}\int_{\partial\Omega_L}\sup_{\theta\in\Theta}|(\varphi(g(x,\theta))\mathbf{1}\{X \in \Omega_L\} - \varphi_{n,L}(g(x,\theta)))s(x,\theta)^T\vec{n}(x)f(x,\theta)|ds = 0$. Taking $n \to \infty$, we obtain

$$\begin{aligned}&\frac{d}{d\theta}\mathbb{E}(\varphi(g(X,\theta))\mathbf{1}\{X \in \Omega_L\}) \\ =&\int_{\Omega_L} \varphi(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx + \int_{\partial\Omega_L} \varphi(g(x,\theta))s(x,\theta)^T\vec{n}(x)f(x,\theta)ds.\end{aligned} \tag{10}$$

Next, we aim to show that $\lim_{L\to\infty} \frac{d}{d\theta}\mathbb{E}(\varphi(g(X,\theta))\mathbf{1}\{X \in \Omega_L\}) = \frac{d}{d\theta}\mathbb{E}(\varphi(g(X,\theta)))$, for which a sufficient condition is the uniform convergence of both integrals on the right-hand side of Equation (10) over Θ as $L \to \infty$. Specifically, a sufficient condition for the uniform convergence of the first integral is $\int_\Omega \sup_{\theta\in\Theta}|\varphi(g(x,\theta))(d(x,\theta)+l(x,\theta))|f(x,\theta)dx < \infty$. We refer to Theorem 4 in Section 16.3.5 of Zorich (2004b) for the conditions under which the interchange of the order of limit and integral is permissible.

4 SIMULATION EXAMPLE

In this section, we evaluate the generalized GLR method using the toy example introduced in Section 1: $\mathbb{E}(\mathbf{1}\{X < \theta\})$, where X follows an exponential distribution with parameter $\theta > 0$. Notice that its derivative can be computed analytically:

$$\frac{d}{d\theta}\mathbb{E}(\mathbf{1}\{X < \theta\}) = \frac{d}{d\theta}\int_0^\theta \theta e^{-\theta x}dx = (\theta e^{-\theta x})|_{x=\theta}\frac{d}{d\theta}(\theta) + \int_0^\theta \frac{d}{d\theta}(\theta e^{-\theta x})dx = 2\theta e^{-\theta^2}.$$

Using the conventional push-out LR method, we obtain an unbiased estimator as follows:

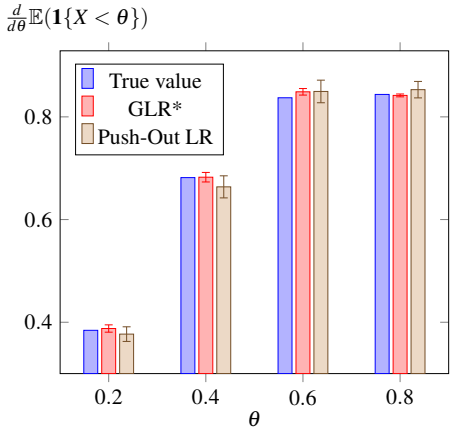
$$\frac{d}{d\theta}\mathbb{E}(\mathbf{1}\{X < \theta\})_{\mathrm{LR}} = \mathbb{E}\left(\mathbf{1}\{Y < 1\}\partial_\theta \log f_Y(Y,\theta)\right) = \mathbb{E}\left(\mathbf{1}\{Y < 1\}(2/\theta - 2\theta Y)\right),$$

where Y follows an exponential distribution with parameter θ^2. The method introduced in Theorem 2 is referred to as the GLR* method, offering another unbiased estimator:

$$\frac{d}{d\theta}\mathbb{E}(\mathbf{1}\{X < \theta\})_{\mathrm{GLR}^*} = \mathbb{E}\left(\mathbf{1}\{X < \theta\}(d(X,\theta)+l(X,\theta)) + \theta\right) = \mathbb{E}\left(\mathbf{1}\{X < \theta\}(1/\theta - \theta - X) + \theta\right),$$

where the constant θ corresponds to estimator (9), the derivative of the parameter-dependent domain.

We simulate both derivative estimators at $\theta = 0.2, 0.4, 0.6, 0.8$ with 2500 independent replications. The simulation results are depicted in Figure 2. Both estimators demonstrate satisfactory accuracy. Notably, the standard errors of the GLR* estimator are half or even less of those of the push-out LR estimator. This observation can be explained as follows. From Equation (7), we observe that the "randomness" of the derivative is split into two components. One component represents a conventional LR estimator (after the change of variables). The other component captures the sensitivity the integration domain w.r.t. θ, and (9) is an unbiased estimator for this component. In this simple example, the latter component is merely the constant term θ, whereas if the input X is a random vector, additional simulations might be required to estimate the value of the surface integral. Thus, the reduction in variance comes at the expense of potentially additional simulation runs. Moreover, for this example, it is more efficient to use the conditional density estimator (L'Ecuyer et al. 2022).



θ	True value	GLR*	Push-Out LR
0.2	0.384	0.388 ± 0.018	0.377 ± 0.038
0.4	0.682	0.683 ± 0.014	0.664 ± 0.032
0.6	0.837	0.849 ± 0.008	0.850 ± 0.026
0.8	0.844	0.842 ± 0.003	0.853 ± 0.019

Figure 2: Simulation results: Point estimates and standard errors for $\frac{d}{d\theta}\mathbb{E}(\mathbf{1}\{X<\theta\})$.

5 CONCLUSION

In this paper, we introduce a novel push-out Leibniz integration approach to generalize the GLR method. The underlying idea of our method is straightforward: "push" the parameter θ out of the performance measure $\varphi(g(X,\theta))$ through a change of variables $Y = g(X,\theta)$, differentiate the transformed density function f_Y and integration domain $g(\Omega,\theta)$ using the Leibniz integral rule, and finally reverse the change of variables $X = g^{-1}(Y,\theta)$. Compared to the push-out LR method, the newly derived estimator can be applied to a wider range of gradient estimation problems where the sample space is parameter-dependent and the function g is only locally invertible. We demonstrate that the newly derived estimator encompasses the existing GLR estimators as special cases. Simulation results suggest that the generalized GLR estimator, compared to the push-out LR method, can reduce variance at the expense of potentially additional simulations. For future research, we aim to extend our results from compact sample spaces to unbounded sample spaces and apply them to more practical scenarios. We also observe that the form of the estimator (9) for the surface integral resembles the form of some SPA estimators (Fu and Hu 1997). Investigating the connection between GLR and SPA estimators is an interesting direction for further research.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant IIS-2123684 and by AFOSR under Grant FA95502010211.

A PROOF OF PROPOSITION 2

We refer to Chapter 2 from Frankel (2011) for the justification of Equations (14) and (17).
Equation (4): By the chain rule,

$$\begin{aligned}\frac{d}{d\theta}(f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|) = (\nabla_x f(g^{-1}(y,\theta),\theta)^T \partial_\theta g^{-1}(y,\theta) + \partial_\theta f(g^{-1}(y,\theta),\theta)) \\ \times |\det(J_{g^{-1}}(y,\theta))| + f(g^{-1}(y,\theta),\theta)\partial_\theta |\det(J_{g^{-1}}(y,\theta))|,\end{aligned} \tag{11}$$

Notice that $g(g^{-1}(y,\theta),\theta) = y$. Therefore, by (implicit) differentiation,

$$0 = \frac{d}{d\theta} g(g^{-1}(y,\theta),\theta) = \partial_\theta g(g^{-1}(y,\theta),\theta) + J_g(g^{-1}(y,\theta),\theta)\partial_\theta g^{-1}(y,\theta),$$

i.e.,

$$\partial_\theta g^{-1}(y,\theta) = -J_g^{-1}(g^{-1}(y,\theta),\theta)\partial_\theta g(g^{-1}(y,\theta),\theta). \tag{12}$$

To compute $\partial_\theta |\det(J_{g^{-1}}(y,\theta))|$, we use the fact that $\partial_\theta \det A(\theta) = \det A(\theta)\text{trace}(\partial_\theta A(\theta)A(\theta)^{-1})$. Since g is twice continuously differentiable, $\partial_\theta \partial_{y_j} g_i^{-1}(y,\theta) = \partial_{y_j}(\partial_\theta g_i^{-1}(y,\theta))$, and

$$\text{trace}\left((\partial_\theta J_{g^{-1}}(y,\theta))J_{g^{-1}}(y,\theta)^{-1}\right) = \sum_{i,j}^{n} \partial_{y_j}(\partial_\theta g_i^{-1}(y,\theta))\partial_{x_i} g_j(x,\theta)|_{x=g^{-1}(y,\theta)} = \sum_{i=1}^{n} \partial_{x_i}(\partial_\theta g_i^{-1}(y,\theta))|_{y=g(x,\theta)}.$$

It follows that

$$\partial_\theta \det(J_{g^{-1}}(y,\theta)) = \det(J_{g^{-1}}(y,\theta)) \left(\sum_{i=1}^{n} \partial_{x_i}(\partial_\theta g_i^{-1}(y,\theta))|_{y=g(x,\theta)} \right). \tag{13}$$

Substituting Equations (12) and (13) into Equation (11), we obtain Equation (4):

$$\begin{aligned}
&\frac{d}{d\theta}(f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|)dy \\
&= (-\nabla_x f(x,\theta)^T J_g^{-1}(x,\theta)\partial_\theta g(x,\theta) + \partial_\theta f(x,\theta))|\det(J_{g^{-1}}(y,\theta))| \\
&+ f(x,\theta)|\det(J_{g^{-1}}(y,\theta))|\text{div}(-J_g^{-1}(x,\theta)\partial_\theta g(x,\theta)) \\
&= |\det(J_{g^{-1}}(y,\theta))|(\text{div}(-f(x,\theta)J_g^{-1}(x,\theta)\partial_\theta g(x,\theta)) + \partial_\theta f(x,\theta)),
\end{aligned}$$

where the last equation follows from the fact that for any real-valued function h and vector-valued function $\vec{v}$,

$$\text{div}(h(x)\vec{v}(x)) = \nabla_x h(x)^T \vec{v}(x) + h(x)\,\text{div}(\vec{v}(x)). \tag{14}$$

Equation (5): By Equation (14), we obtain

$$\begin{aligned}
&\text{div}\left(\varphi_n(y) f(g^{-1}(y,\theta),\theta)|\det(J_{g^{-1}}(y,\theta))|\vec{v}(y)\right) \\
&= \nabla_y(\varphi_n(y) f(g^{-1}(y,\theta),\theta))^T |\det(J_{g^{-1}}(y,\theta))|\vec{v}(y) + \varphi_n(y) f(g^{-1}(y,\theta),\theta)\,\text{div}(|\det(J_{g^{-1}}(y,\theta))|\vec{v}(y)).
\end{aligned} \tag{15}$$

Applying Equation (14) again, we obtain

$$\text{div}(|\det(J_{g^{-1}}(y,\theta))|\vec{v}(y)) = \nabla_y |\det(J_{g^{-1}}(y,\theta))|^T \vec{v}(y) + |\det(J_{g^{-1}}(y,\theta))|\,\text{div}(\vec{v}(y)).$$

For $i = 1, \cdots, n$,

$$\begin{aligned}
\partial_{y_i} \det(J_{g^{-1}}(y,\theta)) &= \det(J_{g^{-1}}(y,\theta))\,\text{trace}(\partial_{y_i} J_{g^{-1}}(y,\theta) J_{g^{-1}}^{-1}(y,\theta)) \\
&= \det(J_{g^{-1}}(y,\theta)) \sum_{j,k}^{n} \partial_{y_i}(J_{g^{-1}}(y,\theta))_{jk} (J_{g^{-1}}^{-1}(y,\theta))_{kj} \\
&= \det(J_{g^{-1}}(y,\theta)) \sum_{j=1}^{n}\sum_{k=1}^{n} (J_{g^{-1}}^{-1}(y,\theta))_{kj} \partial_{y_k}(J_{g^{-1}}(y,\theta))_{ji}.
\end{aligned}$$

Note that for any differentiable function $h : \mathbb{R}^n \mapsto \mathbb{R}$, $\nabla_x h(g(x,\theta)) = J_g(x,\theta)^T \nabla_y h(y)|_{y=g(x,\theta)}$, i.e.,

$$\nabla_y h(y)|_{y=g(x,\theta)} = J_g^{-1}(x,\theta)^T \nabla_x h(g(x,\theta)).$$

Therefore, we can write

$$\nabla_x (J_g^{-1}(x,\theta))_{ji} = J_{g^{-1}}^{-1}(y,\theta)^T \nabla_y (J_g^{-1}(g^{-1}(y,\theta),\theta))_{ji},$$

i.e., $\partial_{x_j}(J_g^{-1}(x,\theta))_{ji} = \sum_{k=1}^n (J_{g^{-1}}^{-1}(y,\theta))_{kj}\partial_{y_k}(J_{g^{-1}}(y,\theta))_{ji}$, and it follows that

$$\partial_{y_i}\det(J_{g^{-1}}(y,\theta)) = \det(J_{g^{-1}}(y,\theta))\sum_{j=1}^n \partial_{x_j}(J_g^{-1}(x,\theta))_{ji}.$$

Hence,

$$\begin{aligned}
&\nabla_y|\det(J_{g^{-1}}(y,\theta))|^T\vec{v}(y)\\
&= |\det(J_{g^{-1}}(y,\theta))|\sum_{i=1}^n(\sum_{j=1}^n \partial_{x_j}(J_g^{-1}(x,\theta))_{ji})\vec{v}_i(y) = \det(J_{g^{-1}}(y,\theta))\operatorname{div}(J_g^{-1}(x,\theta)^T)^T\vec{v}(y),
\end{aligned}$$

where $\operatorname{div}A(x) := (\sum_{j=1}^n \partial_{x_j}A(x)_{1j},\cdots,\sum_{j=1}^n \partial_{x_j}A(x)_{nj})^T$ for any matrix-valued function $A(x)$. Notice that

$$\operatorname{div}(\vec{v}(y)) = \operatorname{trace}(\nabla_y\partial_\theta g(g^{-1}(y,\theta),\theta)) = \operatorname{trace}(J_g^{-1}(x,\theta)^T\nabla_x\partial_\theta g(x,\theta)),$$

where $\nabla_y v(y) := (\nabla_y v_1(y),\cdots,\nabla_y v_n(y))^T$ for any vector-valued function $v(x)$. Therefore,

$$\begin{aligned}
&\operatorname{div}(|\det(J_{g^{-1}}(y,\theta))|\vec{v}(y))\\
&= |\det(J_{g^{-1}}(y,\theta))|\operatorname{div}(J_g^{-1}(x,\theta)^T)^T\vec{v}(y) + |\det(J_{g^{-1}}(y,\theta))|\operatorname{trace}(\nabla_x\partial_\theta g(x,\theta)J_g^{-1}(x,\theta))\\
&= |\det(J_{g^{-1}}(y,\theta))|\operatorname{div}(J_g^{-1}(x,\theta)^T)^T\vec{v}(y) + |\det(J_{g^{-1}}(y,\theta))|\operatorname{trace}(J_g^{-1}(x,\theta)\nabla_x\partial_\theta g(x,\theta))\\
&= |\det(J_{g^{-1}}(y,\theta))|\operatorname{div}(J_g^{-1}(x,\theta)\partial_\theta g(x,\theta)),
\end{aligned} \tag{16}$$

where the last equation follows from the divergence formula for matrix-vector production

$$\operatorname{div}(A(x)v(x)) = \operatorname{div}(A(x)^T)^T v(x) + \operatorname{trace}(A(x)\nabla v(x)). \tag{17}$$

Using the chain rule for gradient $\nabla(pq) = p\nabla q + q\nabla p$, we can write

$$\begin{aligned}
\nabla_y(\varphi_n(y)f(g^{-1}(y,\theta),\theta)) &= \varphi_n(y)\nabla_y f(g^{-1}(y,\theta),\theta) + f(g^{-1}(y,\theta),\theta)\nabla_y\varphi_n(y)\\
&= J_g^{-1}(x,\theta)^T\nabla_x(\varphi_n(g(x,\theta))f(x,\theta)).
\end{aligned} \tag{18}$$

Substituting Equations (16) and (18) into Equation (15), we obtain Equation (5). □

B PROOF OF THEOREM 2

Since $\Omega\times\Theta$ is compact and both f,g are continuously differentiable, $\sup_{(x,\theta)\in\Omega\times\Theta}|f(x,\theta)|$ and $\sup_{(x,\theta)\in\Omega\times\Theta}|\det J_g^{-1}(x,\theta)|$ are bounded. Therefore,

$$\begin{aligned}
\lim_{n\to\infty}|\mathbb{E}(\varphi(g(X,\theta))) - \mathbb{E}(\varphi_n(g(X,\theta)))| &\le \lim_{n\to\infty}\int_\Omega |\varphi(g(x,\theta)) - \varphi_n(g(x,\theta))||f(x,\theta)|dx\\
&\le \lim_{n\to\infty}\int_\Omega |\varphi(g(x,\theta)) - \varphi_n(g(x,\theta))|dx \sup_{(x,\theta)\in\Omega\times\Theta}|f(x,\theta)|\\
&\le \lim_{n\to\infty}\int_{g(\Omega,\theta)} |\varphi(y) - \varphi_n(y)|dy \sup_{(x,\theta)\in\Omega\times\Theta}|f(x,\theta)\det J_g^{-1}(x,\theta)| = 0,
\end{aligned}$$

where the last equation follows from the fact that $\varphi_n\to\varphi$ in L^1. Similarly, $\sup_{(x,\theta)\in\Omega\times\Theta}|d(x,\theta)+l(x,\theta)|$ is bounded, and

$$\begin{aligned}
&\lim_{n\to\infty}\left|\int_\Omega \varphi(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx - \int_\Omega \varphi_n(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx\right|\\
&\le \lim_{n\to\infty}\int_{g(\Omega,\theta)} |\varphi(y) - \varphi_n(y)|dy \sup_{(x,\theta)\in\Omega\times\Theta}|(d(x,\theta)+l(x,\theta))f(x,\theta)\det J_g^{-1}(x,\theta)| = 0.
\end{aligned}$$

By Theorem 4 in Section 16.3.5 of Zorich (2004b), we obtain

$$\begin{aligned}\frac{d}{d\theta}\mathbb{E}(\varphi(g(X,\theta))) &= \lim_{n\to\infty}\frac{d}{d\theta}\mathbb{E}(\varphi_n(g(X,\theta))) \\ &= \int_\Omega \varphi(g(x,\theta))(d(x,\theta)+l(x,\theta))f(x,\theta)dx + \int_{\partial\Omega}\varphi(g(x,\theta))s(x,\theta)^T\vec{n}(x)f(x,\theta)ds. \quad \square\end{aligned}$$

REFERENCES

Amann, H., J. Escher, and G. Brookfield. 2005. *Analysis*. Verlag: Birkhäuser.

Flanders, H. 1973. "Differentiation Under the Integral Sign". *The American Mathematical Monthly* 80(6):615–627.

Folland, G. B. 1999. *Real Analysis: Modern Techniques and Their Applications*. New York: John Wiley & Sons.

Frankel, T. 2011. *The Geometry of Physics: An Introduction*. New York: Cambridge University Press.

Fu, M. C. and J.-Q. Hu. 1997. *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*. Boston: Kluwer Academic.

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis*. Boston: Kluwer Academic.

Glynn, P. W. 1987. "Likelihood Ratio Gradient Estimation: An Overview". In *1987 Winter Simulation Conference (WSC)*, 366–375 https://doi.org/10.1145/318371.318612.

Landkof, N. S. 1972. *Foundations of Modern Potential Theory*, Volume 180. Berlin: Springer.

L'Ecuyer, P. 1990. "A Unified View of the IPA, SF, and LR Gradient Estimation Techniques". *Management Science* 36(11):1364–1383.

L'Ecuyer, P., F. Puchhammer, and A. Ben Abdellah. 2022. "Monte Carlo and Quasi–Monte Carlo Density Estimation via Conditioning". *INFORMS Journal on Computing* 34(3):1729–1748.

Peng, Y., M. C. Fu, J. Hu, P. L'Ecuyer and B. Tuffin. 2020. "Generalized Likelihood Ratio Method for Stochastic Models with Uniform Random Numbers as Inputs". *HAL preprint hal-02652068*.

Peng, Y., M. C. Fu, J.-Q. Hu, and B. Heidergott. 2018. "A New Unbiased Stochastic Derivative Estimator for Discontinuous Sample Performances with Structural Parameters". *Operations Research* 66(2):487–499.

Pflug, G. C. 1996. *Optimization of Stochastic Models: The Interface Between Simulation and Optimization*. Boston: Kluwer Academic.

Puchhammer, F. and P. L'Ecuyer. 2022. "Likelihood Ratio Density Estimation for Simulation Models". In *2022 Winter Simulation Conference (WSC)*, 109–120 https://doi.org/10.5555/3586210.3586220.

Rubinstein, R. Y. 1992. "Sensitivity Analysis of Discrete Event Systems by the "Push Out" Method". *Annals of Operations Research* 39(1):229–250.

Shapiro, V. L. 1958. "The Divergence Theorem for Discontinuous Vector Fields". *Annals of Mathematics* 68(3):604–624.

Wang, Y., M. C. Fu, and S. I. Marcus. 2012. "A New Stochastic Derivative Estimator for Discontinuous Payoff Functions with Application to Financial Derivatives". *Operations Research* 60(2):447–460.

Zorich, V. A. 2004a. *Mathematical Analysis I*. Berlin, Heidelberg: Springer.

Zorich, V. A. 2004b. *Mathematical Analysis II*. Berlin, Heidelberg: Springer.

AUTHOR BIOGRAPHIES

XINGYU REN is a Ph.D. student in the Department of Electrical and Computer Engineering at the University of Maryland, College Park. His research interests include stochastic optimization and Markov decision processes. His e-mail address is renxy@umd.edu.

MICHAEL C. FU holds the Smith Chair of Management Science in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research and an affiliate appointment in the Department of Electrical and Computer Engineering, at the University of Maryland, College Park. His research interests include stochastic gradient estimation, simulation optimization, and applied probability. He served as WSC2011 Program Chair and received the INFORMS Simulation Society's Distinguished Service Award in 2018. He is a Fellow of INFORMS and IEEE. His e-mail address is mfu@umd.edu.