Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Technical Note—Central Limit Theorems for Estimated Functions at Estimated Points

Peter W. Glynn, Lin Fan, Michael C. Fu, Jian-Qiang Hu, Yijie Peng

informs
http://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 68, No. 5, September–October 2020, pp. 1557–1563
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Methods

Technical Note—Central Limit Theorems for Estimated Functions at Estimated Points

Peter W. Glynn,[a] Lin Fan,[a] Michael C. Fu,[b] Jian-Qiang Hu,[c] Yijie Peng[d]

[a] Department of Management Science and Engineering, Stanford University, Stanford, California 94305; [b] Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, College Park, Maryland 20742; [c] Department of Management Science, School of Management, Fudan University, Shanghai 200433, China; [d] Department of Management Science and Information Systems, Guanghua School of Management, Peking University, Beijing 100871, China

Contact: glynn@stanford.edu (PWG); linfan@stanford.edu (LF); mfu@rhsmith.umd.edu (MCF); hujq@fudan.edu.cn (J-QH); pengyijie@pku.edu.cn, http://orcid.org/0000-0003-2584-8131 (YP)

Received: May 26, 2017
Revised: October 24, 2018; May 8, 2019
Accepted: July 22, 2019
Published Online in Articles in Advance: May 21, 2020

Subject Classifications: inventory/production: sensitivity analysis, uncertainty: stochastic; queues: simulation
Area of Review: Simulation

https://doi.org/10.1287/opre.2019.1922

Abstract. We provide a simple proof of the central limit theorem (CLT) for estimated functions at estimated points. Such estimators arise in a number of different simulation-based computational settings. We illustrate the methodology via applications to quantile estimation and related sensitivity analysis, as well as to computation of conditional value-at-risk.

Funding: This work was supported by the National Natural Science Foundation of China [Grants 71901003, 71571048, 71720107003, 71690232, 91846301, and 71790615], the National Science Foundation [Grants CMMI-1434419 and DGE-114747], and by the Air Force Office of Scientific Research (AFOSR) [Grant FA9550-15-10050].

Keywords: simulation • central limit theorem

1. Introduction

In a number of simulation-based numerical schemes, the key computational challenge involves the estimation of a quantity of the form $\alpha = \mathbb{E}[Y(x^*)]$ where $Y = (Y(x) : x \in \mathbb{R})$ is a real-valued, square-integrable stochastic process with sample paths that are right-continuous and have left limits everywhere. When x^* is known to the simulator, the solution is straightforward: generate independent and identically distributed (iid) copies $Y_1, Y_2, \ldots$ of the process, evaluate them at x^*, and estimate α via the sample average $n^{-1}\sum_{i=1}^n Y_i(x^*)$. When x^* must be simultaneously estimated via the simulation, by the estimator $\widehat{X}_n$ say, the natural estimator for α is again obvious; namely, estimate α via the "plug-in" estimator $n^{-1}\sum_{i=1}^n Y_i(\widehat{X}_n)$. Note that the estimator involves the estimated function $n^{-1}\sum_{i=1}^n Y_i(\cdot)$ evaluated at the estimated point $\widehat{X}_n$.

A theoretical challenge in this computational setting is the derivation of a large-sample central limit theorem (CLT) for $n^{-1}\sum_{i=1}^n Y_i(\widehat{X}_n)$ as an estimator for α. Such a CLT is essential to the development of large-sample confidence intervals for α, which are used by simulators to assess the accuracy of their computational procedures.

In this note, we provide simple conditions for such CLTs. There are related results, for example, in the context of sample-average approximation in stochastic programming (Shapiro et al. 2009), which require that the stochastic processes involved have Lipschitz continuous sample paths. However, the literature does not explicitly appear to address the case in which the stochastic process Y is stochastically continuous, but not sample path continuous. (Y is stochastically continuous at x^* if $Y(x) - Y(x^*) \Rightarrow 0$ as $x \to x^*$, where $\Rightarrow$ denotes weak convergence.) The Poisson process is an important example of such a stochastic process, and moreover, many discrete-event processes encountered in operations research are stochastically continuous in an underlying parameter, but not sample path continuous. While empirical process theory together with an infinite-dimensional Delta Theorem (Shapiro et al. 2009) can be used to deal with the case in which Y is not sample path continuous, here we provide an alternative approach from first principles with conditions that are simple to verify. Our results apply to quantiles, conditional value-at-risk, quantile sensitivities, and other computational contexts as well. Section 2 describes our theoretical results, while Section 3 provides applications that illustrate the ease-of-use of our theory.

2. Main Results and Proofs

Set $\bar{Y}_n(\cdot) = n^{-1}\sum_{i=1}^n Y_i(\cdot)$ and $\alpha(\cdot) = \mathbb{E}[Y(\cdot)]$. We start with a heuristic discussion that will later be made rigorous; see Theorems 1 and 2. Suppose the sample path of $Y(\cdot)$ is differentiable, and its derivative $Y'(\cdot)$ is an unbiased estimator of $\alpha'(\cdot)$, the derivative of $\alpha(\cdot)$. One natural approach to developing a CLT for $\bar{Y}_n(\widehat{X}_n)$ would be to approximate $\bar{Y}_n(\widehat{X}_n)$ as

$$\bar{Y}_n\left(\widehat{X}_n\right) \approx \bar{Y}_n(x^*) + \bar{Y}_n'(x^*)\left(\widehat{X}_n - x^*\right),$$

where $\approx$ means "is approximately equal to," and $\bar{Y}_n'(x^*)$ is the derivative of $\bar{Y}_n(\cdot)$ evaluated at x^*. The law of large numbers (LLN) then suggests that $\bar{Y}_n'(x^*)$ converges (either in probability or almost surely) to $\alpha'(x^*)$ as $n \to \infty$, so that we can write

$$\bar{Y}_n\left(\widehat{X}_n\right) \approx \bar{Y}_n(x^*) + \alpha'(x^*)\left(\widehat{X}_n - x^*\right).$$

Hence, if there exist jointly Gaussian random variables (rv's) $\mathcal{N}_1$ and $\mathcal{N}_2$ such that the CLT

$$n^{1/2}\left(\bar{Y}_n(x^*) - \alpha, \widehat{X}_n - x^*\right) \Rightarrow (\mathcal{N}_1, \mathcal{N}_2) \tag{1}$$

as $n \to \infty$ holds, then it should follow that the CLT

$$n^{1/2}\left(\bar{Y}_n\left(\widehat{X}_n\right) - \alpha\right) \Rightarrow \mathcal{N}_1 + \alpha'(x^*)\mathcal{N}_2 \tag{2}$$

holds, as $n \to \infty$. The main theoretical issue with this approach is that in many computational settings, the sample paths of $Y(\cdot)$ may not be differentiable (or even continuous, as discussed in Section 1), even if $\alpha(\cdot)$ is. Therefore, we seek an argument in which only $\alpha(\cdot)$ is assumed differentiable, and the sample paths of $Y(\cdot)$ are not necessarily differentiable or continuous.

Example: Conditional Value-at-Risk. For concreteness, we motivate the theoretical development in this section using the estimation of conditional value-at-risk (CVaR) as a running example. For an rv X and $0 < p < 1$, suppose we wish to estimate the CVaR $\alpha \equiv \mathbb{E}[X|X > q] = \mathbb{E}\left[X\mathbf{1}\{X > q\}\right]/(1-p)$, where q (the x^* in this case) is the quantile defined by $\mathbb{P}(X \le q) = p$, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. We assume that q is unknown and must be estimated from the data, and that q is uniquely defined as the root of $\mathbb{P}(X \le q) = p$. In this case, we set $Y(x) \equiv X\mathbf{1}\{X > x\}/(1-p)$, and given an iid sample $X_1, \ldots, X_n$ with the same distribution as X, we take the estimator $\widehat{X}_n = \inf\{x : n^{-1}\sum_{i=1}^n \mathbf{1}\{X_i \le x\} \ge p\}$, the sample quantile. Then, $\bar{Y}_n(\widehat{X}_n) \equiv n^{-1}\sum_{i=1}^n X_i\mathbf{1}\{X_i > \widehat{X}_n\}/(1-p)$ is the natural "plug-in" estimator of α. Here, the sample paths of Y are not even continuous, but under mild regularity conditions on X, $\alpha(\cdot)$ is differentiable at q.

So, we seek an argument that does not rely on the assumption that the sample paths of $Y(\cdot)$ are differentiable or continuous. The approach we will follow in this paper is to set $\xi_n(x) = \bar{Y}_n(x) - \alpha(x)$ and to write

$$\bar{Y}_n\left(\widehat{X}_n\right) - \alpha = \xi_n\left(\widehat{X}_n\right) + \alpha\left(\widehat{X}_n\right) - \alpha.$$

Suppose we can prove that

$$n^{1/2}\left(\xi_n\left(\widehat{X}_n\right) - \xi_n(x^*)\right) \Rightarrow 0 \tag{3}$$

as $n \to \infty$. Then, if $\alpha(\cdot)$ is differentiable at x^*,

$$\begin{aligned}
&n^{1/2}\left(\bar{Y}_n\left(\widehat{X}_n\right) - \alpha\right)\\
&= n^{1/2}\xi_n(x^*) + n^{1/2}\left(\alpha\left(\widehat{X}_n\right) - \alpha\right) + o_P(1)\\
&= n^{1/2}(\bar{Y}_n(x^*) - \alpha) + (\alpha'(x^*) + o_P(1))n^{1/2}\left(\widehat{X}_n - x^*\right) + o_P(1)\\
&= n^{1/2}(\bar{Y}_n(x^*) - \alpha) + \alpha'(x^*)n^{1/2}\left(\widehat{X}_n - x^*\right) + o_P(1),
\end{aligned} \tag{4}$$

where, for a real sequence a_n, $o_P(a_n)$ denotes a sequence of rv's for which $o_P(a_n)/a_n \Rightarrow 0$ as $n \to \infty$, and we used that (1) implies $\widehat{X}_n - x^* \Rightarrow 0$ as $n \to \infty$. The CLT (2) then follows immediately from (1). Consequently, proving (3) is the key remaining step needed to establish the CLT (2).

Not surprisingly, some sample path regularity of $Y(\cdot)$, short of continuity, is needed to establish (3). Here is what we shall require:

Assumption 1. *In some closed neighborhood $\mathcal{O}$ of x^*, $\alpha(\cdot)$ is continuously differentiable, and there exist nonnegative, nondecreasing functions f and g such that*

i. $\mathbb{E}[(Y(x_1) - Y(x_2))^2] \le f(x_2 - x_1)$, $x_1 < x_2$, $x_1, x_2 \in \mathcal{O}$;

ii. $\mathbb{E}[\min((Y(x_1) - Y(x_2))^2, (Y(x_2) - Y(x_3))^2)] \le g(x_3 - x_1)$, $x_1 < x_2 < x_3$, $x_1, x_2, x_3 \in \mathcal{O}$;

iii. $\int_0^1 f^{1/2}(u)u^{-5/4}du < \infty$;

iv. $\int_0^1 g^{1/2}(u)u^{-3/2}du < \infty$.

We note that (iii) implies that $\lim_{u\downarrow 0} f(u) = 0$, and then (i) and Markov's inequality imply that Y is stochastically continuous at x^*. We also note that because $\alpha(\cdot)$ is continuously differentiable in $\mathcal{O}$, $|\alpha(x) - \alpha(y)| \le c|x-y|$, $x, y \in \mathcal{O}$ for some finite c. It is then easily argued that (i)–(iv) also hold with $Y_c(\cdot) \equiv Y(\cdot) - \alpha(\cdot)$ replacing $Y(\cdot)$.

Theorem 1. *If there exist random variables (rv's) $\mathcal{N}_1$ and $\mathcal{N}_2$ such that*

$$n^{1/2}\left(\bar{Y}_n(x^*) - \alpha, \widehat{X}_n - x^*\right) \Rightarrow (\mathcal{N}_1, \mathcal{N}_2)$$

as $n \to \infty$ and Assumption 1 holds, then

$$n^{1/2}\left(\bar{Y}_n\left(\widehat{X}_n\right) - \alpha\right) \Rightarrow \mathcal{N}_1 + \alpha'(x^*)\mathcal{N}_2$$

as $n \to \infty$.

Proof. It remains only to prove (3). From Bloznelis and Paulauskas (1994), it is known that under Assumption 1, there exists a continuous sample path Gaussian process ξ such that

$$n^{1/2}\xi_n(\cdot) \Rightarrow \xi(\cdot)$$

as $n \to \infty$ in $D[a,b]$ (the space of real-valued functions on $[a,b]$ with sample paths that are right-continuous and have left limits, endowed with the Skorohod topology; see Billingsley 1999 for details), where $a < x^* < b$ and $[a,b] \subset \mathbb{O}$. When the limit $\xi(\cdot)$ has continuous sample paths, this implies that $h(\xi_n) \Rightarrow h(\xi)$ for any mapping $h : D[a,b] \to \mathbb{R}$ for which $h(w_n) \to h(w)$ whenever w_n converges uniformly to w as $n \to \infty$. The mapping $h(w) = \min(\sup_{|x-x^*|\le\delta} |w(x) - w(x^*)|, 1)$ clearly has this property. Furthermore, since this h is bounded, it follows by the portmanteau theorem (Billingsley 1999) that for any $\delta > 0$,

$$\mathbb{E}\left[\min\left(\sup_{|x-x^*|\le\delta} n^{1/2}|\xi_n(x) - \xi_n(x^*)|, 1\right)\right] \to \mathbb{E}\left[\min\left(\sup_{|x-x^*|\le\delta} |\xi(x) - \xi(x^*)|, 1\right)\right] \quad (5)$$

as $n \to \infty$. For $1 > \epsilon, \gamma > 0$, we can now use the sample path continuity of the limit $\xi(\cdot)$ to deduce the existence of $\delta > 0$ for which

$$\mathbb{E}\left[\min\left(\sup_{|x-x^*|\le\delta} |\xi(x) - \xi(x^*)|, 1\right)\right] \le \epsilon\gamma. \quad (6)$$

Hence, for this selection of δ, we find that

$$\begin{aligned}
&\mathbb{P}\left(n^{1/2}|\xi_n\left(\widehat{X}_n\right) - \xi(x^*)| > \epsilon\right) \\
&\le \mathbb{P}\left(|\widehat{X}_n - x^*| > \delta\right) + \mathbb{E}\left[\mathbf{1}\left\{\sup_{|x-x^*|\le\delta} n^{1/2}|\xi_n(x) - \xi_n(x^*)| > \epsilon\right\}\right] \\
&\le \mathbb{P}\left(\left|\widehat{X}_n - x^*\right| > \delta\right) + \mathbb{E}\left[\min\left(\sup_{|x-x^*|\le\delta} n^{1/2}|\xi_n(x) - \xi_n(x^*)|\frac{1}{\epsilon}, 1\right)\right] \\
&\le \mathbb{P}\left(\left|\widehat{X}_n - x^*\right| > \delta\right) + \frac{1}{\epsilon}\mathbb{E}\left[\min\left(\sup_{|x-x^*|\le\delta} n^{1/2}|\xi_n(x) - \xi_n(x^*)|, 1\right)\right],
\end{aligned}$$

where the last inequality holds due to $\epsilon < 1$. We now send $n \to \infty$ and apply (5) and (6) to conclude that

$$\limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}\left|\xi_n\left(\widehat{X}_n\right) - \xi(x^*)\right| > \epsilon\right) \le \gamma.$$

Sending $\gamma \downarrow 0$, we find that

$$\limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}\left|\xi_n\left(\widehat{X}_n\right) - \xi(x^*)\right| > \epsilon\right) = 0,$$

thereby proving (2.3). ◻

We now provide simple sufficient conditions for Assumption 1.

Proposition 1. *The conditions of Assumption 1* (i)–(iv) *are satisfied if either of the following conditions hold:*

a. *There exist $\gamma > 1/2$ and an rv Γ with $\mathbb{E}[\Gamma^2] < \infty$ such that*

$$|Y(x_1) - Y(x_2)| \le \Gamma|x_1 - x_2|^\gamma$$

for $x_1, x_2 \in \mathbb{O}$. (The sample paths of $Y(\cdot)$ are almost surely continuous in this special case, but not necessarily differentiable.)

b. *There exist $b \in \mathbb{R}$, $c > 0$, $\beta > 1/2$, and rv's $\mathcal{W}$ and χ such that $Y(x) = \mathcal{W}\mathbf{1}\{\chi \le x\} - b$, with $\mathbb{E}[\mathcal{W}^2\mathbf{1}\{x_1 < \chi \le x_2\}] \le c|x_2 - x_1|^\beta$ for $x_1, x_2 \in \mathbb{O}$.*

Proof. For (a), choose $f(u) = g(u) = \mathbb{E}[\Gamma^2]u^{2\gamma}$, and note that (i), (iii), and (iv) are easily satisfied. For (ii),

$$\begin{aligned}
&\min((Y(x_2) - Y(x_1))^2, (Y(x_3) - Y(x_2))^2) \\
&\le (Y(x_2) - Y(x_1))^2 + (Y(x_3) - Y(x_2))^2 \\
&\le \Gamma^2(|x_2 - x_1|^{2\gamma} + |x_3 - x_2|^{2\gamma}) \\
&\le \Gamma^2|x_3 - x_1|^{2\gamma}.
\end{aligned}$$

Turning to (b), note that χ lies in either $(x_1, x_2]$ or $(x_2, x_3]$, but not in both, so one of $\mathcal{W}^2\mathbf{1}\{x_1 < \chi \le x_2\}$ or $\mathcal{W}^2\mathbf{1}\{x_2 < \chi \le x_3\}$ is zero. Hence, (ii) and (iv) are trivially satisfied. As for (i) and (iii), they hold for $f(u) = c|u|^\beta$. ◻

Later, case (a) of Proposition 1 will be used in the quantile and quantile sensitivity examples, whereas case (b) will be used in the CVaR example. In the remainder of this section, we investigate a particular setting, in which our goal is to estimate $\mathbb{E}[Y(x^*)]$, where x^* satisfies $\mathbb{E}[Z(x^*)] = 0$. More precisely, let $Z = (Z(x) : x \in \mathbb{R})$ be a square-integrable nondecreasing process with right-continuous sample paths, and suppose $(Y_1, Z_1), (Y_2, Z_2), \ldots$ is an iid sequence of copies of (Y, Z) with Y also square-integrable. (Such a setting arises, for example, when estimating CVaR of an rv X, where we set $Y(x) \equiv X\mathbf{1}\{X > x\}/(1-p)$ and $Z(x) \equiv \mathbf{1}\{X \le x\} - p$.) Set $\phi(x) = \mathbb{E}[Z(x)]$ and assume that $\phi(\cdot)$ is continuously differentiable with positive derivative in some neighborhood of x^*. We suppose that $\widehat{X}_n$ is an estimator for x^* minimizing $|\overline{Z}_n(\cdot)|$ (since $\overline{Z}_n(\cdot)$ may not have a root for a particular sample size n) and satisfying $\widehat{X}_n \Rightarrow x^*$ and

$$\overline{Z}_n\left(\widehat{X}_n\right) = o_P\left(n^{-1/2}\right) \quad (7)$$

as $n \to \infty$, where $o_P(1/a_n)$ is a sequence of rv's for which $a_n o_P(1/a_n) \Rightarrow 0$ as $n \to \infty$. The assumption (7) is, for example, immediate when $\overline{Z}_n(\cdot)$ has continuous sample paths. (Since $\phi'(\cdot)$ is positive and continuous at x^*, there exist $\epsilon > 0$ and x_1, x_2 such that $x_1 < x^* < x_2$, $\phi(x_1) < -\epsilon$ and $\phi(x_2) > \epsilon$. Then the strong LLN yields $\overline{Z}_n(x_1) < -\epsilon/2$ and $\overline{Z}_n(x_2) > \epsilon/2$ almost surely for n sufficiently large, and the desired conclusion follows by the intermediate value theorem.) It is also obvious when $Z(x) = V\mathbf{1}\{\tau \le x\}$, with $\mathbb{E}[V^2] < \infty$, and τ is a continuous rv, since $n^{-1}\sum_{i=1}^n \mathbf{1}\{\tau_i \le \cdot\}$ then jumps by $1/n$, and $|\overline{Z}_n(\widehat{X}_n)| \le n^{-1}\max_{1\le i\le n}|V_i|$. (Then, $n^{-1/2}\max_{1\le i\le n}|V_i| = o_P(1)$ by noticing that for any $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}\left(\max_{1\le i\le n}|V_i| > n^{1/2}\epsilon\right) &\le n\mathbb{P}\left(|V| > n^{1/2}\epsilon\right)\\ &= \mathbb{E}\left[n\mathbf{1}\{V^2 > n\epsilon^2\}\right]\\ &\le \mathbb{E}\left[(V^2/\epsilon^2)\mathbf{1}\{V^2/\epsilon^2 > n\}\right] \downarrow 0\end{aligned}$$

as $n \to \infty$.)

Example: Conditional Value-at-Risk. The above discussion applies to the running example of CVaR estimation, which requires quantile estimation. Suppose that there exists a unique value q for which $\mathbb{P}(X \le q) = p$. We then set $\phi(x) \equiv \mathbb{P}(X \le x) - p$ and $Z(x) \equiv \mathbf{1}\{X \le x\} - p$, so that with x^* taken to be the quantile q, we clearly have $\phi(x^*) = 0$. Taking $\widehat{X}_n$ to be the sample quantile, we then have $|\overline{Z}_n(\widehat{X}_n)| \le n^{-1}$ almost surely from the above discussion, and so (7) holds.

Continuing the theoretical discussion, set $\nu_n(x) = \overline{Z}_n(x) - \phi(x)$ and write

$$\overline{Z}_n\left(\widehat{X}_n\right) = \nu_n\left(\widehat{X}_n\right) + \phi\left(\widehat{X}_n\right) - \phi(x^*).$$

The square-integrability and iid structure of the (Y_i, Z_i)'s guarantee the CLT

$$n^{1/2}\left(\overline{Y}_n(x^*) - \alpha, \overline{Z}_n(x^*)\right) \Rightarrow \left(\mathcal{N}_1', \mathcal{N}_2'\right) \tag{8}$$

as $n \to \infty$, where $\mathrm{Var}(\mathcal{N}_1') = \mathrm{Var}(Y(x^*))$, $\mathrm{Var}(\mathcal{N}_2') = \mathrm{Var}(Z(x^*))$ and $\mathrm{Cov}(\mathcal{N}_1', \mathcal{N}_2') = \mathrm{Cov}(Y(x^*), Z(x^*))$. Then, the proof of Theorem 1 establishes the analogue of (2.3) with $\nu_n(\cdot)$ playing the role of $\xi_n(\cdot)$, so that

$$\begin{aligned}&n^{1/2}\overline{Z}_n\left(\widehat{X}_n\right)\\ &= n^{1/2}\nu_n(x^*) + n^{1/2}\left(\phi\left(\widehat{X}_n\right) - \phi(x^*)\right) + o_P(1)\\ &= n^{1/2}\overline{Z}_n(x^*) + \left(\phi'(x^*) + o_P(1)\right)n^{1/2}\left(\widehat{X}_n - x^*\right) + o_P(1)\\ &= n^{1/2}\overline{Z}_n(x^*) + \phi'(x^*)n^{1/2}\left(\widehat{X}_n - x^*\right) + o_P(1)\end{aligned}$$

as $n \to \infty$. And in view of (7), we see that

$$n^{1/2}\overline{Z}_n(x^*) + \phi'(x^*)n^{1/2}\left(\widehat{X}_n - x^*\right) = o_P(1).$$

as $n \to \infty$. Hence, if Assumption 1 also holds for $Y(\cdot)$ at x^*, then using (4),

$$\begin{aligned}&n^{1/2}\left(\overline{Y}_n\left(\widehat{X}_n\right) - \alpha\right)\\ &\quad = n^{1/2}\left(\overline{Y}_n(x^*) - \alpha\right) + \alpha'(x^*)n^{1/2}\left(\widehat{X}_n - x^*\right) + o_P(1)\\ &\quad = n^{1/2}\left(\overline{Y}_n(x^*) - \alpha\right) - \frac{\alpha'(x^*)}{\phi'(x^*)}n^{1/2}\overline{Z}_n(x^*) + o_P(1)\end{aligned} \tag{9}$$

as $n \to \infty$. In view of (8), we have therefore proved the following theorem.

Theorem 2. *Assume Assumption 1 holds at x^* for Y and also for Z (playing the role of Y). If $\widehat{X}_n$ satisfies (7), then*

$$n^{1/2}\left(\overline{Y}_n\left(\widehat{X}_n\right) - \alpha\right) \Rightarrow \mathcal{N}_1' - \frac{\alpha'(x^*)}{\phi'(x^*)}\mathcal{N}_2' \tag{10}$$

as $n \to \infty$.

3. Illustrative Applications

Example 1 (Quantiles). For $0 < p < 1$, suppose that we wish to compute the quantile q such that $\mathbb{P}(X \le q) = p$. As in the running example from Section 2, we estimate q via any estimator satisfying (7), such as the sample quantile defined previously. We assume that there exists a neighborhood $\mathcal{O}$ of q within which X has a positive and continuous density $f_X(\cdot)$. Such an assumption is standard in the literature on quantile CLTs; see, for example, p. 77 of Serfling 1980.

In this case, we set $Y(x) \equiv x$ (here, $Y(\cdot)$ is actually nonrandom and differentiable), and as we have discussed previously, $Z(x) \equiv \mathbf{1}\{X \le x\} - p$ for $x \in \mathbb{R}$. We need to verify Assumption 1 (i)–(iv) for $Y(\cdot)$ and $Z(\cdot)$ in a neighborhood of q. Application of Proposition 1(a) to Y (with $\gamma = 1$) and 1(b) to Z (with $\chi = X$, $\mathcal{W} = 1$, and $\beta = 1$) then proves that Theorem 2 holds, yielding

$$n^{1/2}\left(\widehat{X}_n - q\right) \Rightarrow \sigma\mathcal{N}(0,1)$$

as $n \to \infty$, where $\sigma^2 = \mathrm{Var}(-Z(q)/f_X(q)) = p(1-p)/f_X^2(q)$.

Example 2 (Conditional Value-at-Risk). Continuing the running example from Section 2, our goal is to compute $\alpha = \mathbb{E}[X \mid X > q] = \mathbb{E}[X\mathbf{1}\{X > q\}]/(1-p)$. Like before, we set $Y(x) \equiv X\mathbf{1}\{X > x\}/(1-p)$ and $Z(x) \equiv \mathbf{1}\{X \le x\} - p$, and let $\widehat{X}_n$ be any estimator of q satisfying (7). It is easily seen using Proposition 1 that both Y and Z satisfy Assumption 1 (i)–(iv) at q whenever $\mathbb{E}[X^2] < \infty$ and X has a positive and continuous density in a neighborhood of q. In that case, Theorem 2 guarantees that

$$n^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \frac{X_i\mathbf{1}\left\{X_i > \widehat{X}_n\right\}}{1-p} - \mathbb{E}[X \mid X > q]\right) \Rightarrow \sigma\mathcal{N}(0,1)$$

as $n \to \infty$, where $\sigma^2 = \mathrm{Var}((X-q)\mathbf{1}\{X > q\}/(1-p))$. It should be noted that our continuous density assumption is weaker than that found in Hong et al. (2014), where X is assumed to have a continuously differentiable density. On the other hand, Pflug and Wozabal (2010) require only that the quantile function (defined by $q(y) = \inf\{x : \mathbb{P}(X \le x) \ge y\}$, for $0 < y < 1$) be continuous at p, but require slightly stronger moment hypotheses on X than we do.

Example 3 (Quantile Sensitivity). Suppose that the rv X depends on a parameter θ, so that $X = X(\theta)$. Then, the quantile q also depends on θ, and $q = q(\theta)$ satisfies $\mathbb{P}(X(\theta) \le q(\theta)) = p$. In many settings, we can write $\mathbb{P}(X(\theta) \le x) = F(\theta; x) = \mathbb{E}[R(\theta; x)]$ for some $R(\cdot;\cdot)$, where R is smooth in θ and x. (For example, such representations can often be obtained by conditioning on some rv S, chosen so that $X(\theta)$ has a computable smooth density; see Fu et al. 2009 for details.) Let θ_0 be the parameter value at which we wish to compute the derivative of $q(\cdot)$. Suppose that F is twice continuously differentiable in a neighborhood $\mathbb{O}$ of $(\theta_0, q(\theta_0))$ with $\partial_x F(\theta_0; q(\theta_0)) > 0$, where $\partial_x F(\theta; x) \doteq \partial F(\theta; x)/\partial x$, and that $R(\cdot;\cdot)$ is almost surely differentiable in that neighborhood and satisfies

$$\partial_\theta F(\theta; x) = \mathbb{E}[\partial_\theta R(\theta; x)],$$
$$\partial_x F(\theta; x) = \mathbb{E}[\partial_x R(\theta; x)],$$

where $\partial_\theta F(\theta; x) \doteq \partial F(\theta; x)/\partial\theta$, $\partial_\theta R(\theta; x) \doteq \partial R(\theta; x)/\partial\theta$, and $\partial_x R(\theta; x) \doteq \partial R(\theta; x)/\partial x$. Finally, assume that $\mathbb{E}[(\partial_\theta R(\theta_0; q(\theta_0)))^2 + (\partial_x R(\theta_0; q(\theta_0)))^2] < \infty$ and that there exist $\gamma > 1/2$ and a square-integrable rv Γ such that

$$|\partial_\theta R(\theta_0; x_1) - \partial_\theta R(\theta_0; x_2)| \le \Gamma |x_1 - x_2|^\gamma,$$
$$|\partial_x R(\theta_0; x_1) - \partial_x R(\theta_0; x_2)| \le \Gamma |x_1 - x_2|^\gamma$$

for $(\theta_0, x_i) \in \mathbb{O}$, $i = 1, 2$.

The implicit function theorem guarantees that the quantile sensitivity is given by $q'(\theta_0) = -\partial_\theta F(\theta_0; q(\theta_0))/\partial_x F(\theta_0; q(\theta_0))$. Furthermore, part (a) of Proposition 1 and (9) then establish that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \partial_\theta R_i\left(\theta_0; \widehat{X}_n\right) \\
&= \partial_\theta F(\theta_0; q(\theta_0)) + \frac{1}{n}\sum_{i=1}^n \{\partial_\theta R_i(\theta_0; q(\theta_0)) - \partial_\theta F(\theta_0; q(\theta_0))\} \\
&\quad - \frac{\partial^2_{\theta,x} F(\theta_0; q(\theta_0))}{\partial_x F(\theta_0; q(\theta_0))} \frac{1}{n}\sum_{i=1}^n \left\{\mathbf{1}\{X_i \le q(\theta_0)\} - p\right\} + o_P\left(n^{-1/2}\right),
\end{aligned}$$

where $\partial^2_{\theta,x} F(\theta; x) \doteq \partial^2 F(\theta; x)/\partial\theta\partial x$, and (via a Taylor series expansion),

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{i=1}^n \partial_x R_i\left(\theta_0; \widehat{X}_n\right)\right)^{-1} \\
&\quad = \left(\partial_x F(\theta_0; q(\theta_0))\right)^{-1} - \left(\partial_x F(\theta_0; q(\theta_0))\right)^{-2} \\
&\qquad \cdot \frac{1}{n}\sum_{i=1}^n \{\partial_x R_i(\theta_0; q(\theta_0)) - \partial_x F(\theta_0; q(\theta_0))\} \\
&\qquad + \partial^2_{x,x} F\left(\theta_0; q(\theta_0)\right)\left(\partial_x F(\theta_0; q(\theta_0))\right)^{-3} \\
&\qquad \cdot \frac{1}{n}\sum_{i=1}^n \left\{\mathbf{1}\{X_i \le q(\theta_0)\} - p\right\} + o_P\left(n^{-1/2}\right),
\end{aligned}$$

where $\widehat{X}_n$ is any estimator of q satisfying (7) and $\partial^2_{x,x} F(\theta; x) \doteq \partial^2 F(\theta; x)/\partial x^2$. It follows that

$$n^{1/2}\left(-\frac{\sum_{i=1}^n \partial_\theta R_i\left(\theta_0; \widehat{X}_n\right)}{\sum_{i=1}^n \partial_x R_i\left(\theta_0; \widehat{X}_n\right)} - q'(\theta_0)\right) \Rightarrow \sigma\mathcal{N}(0,1)$$

as $n \to \infty$, where $\sigma^2 = \mathrm{Var}(M)$, and

$$\begin{aligned}
M = &\frac{\partial_\theta R(\theta_0; q(\theta_0))}{\partial_x F(\theta_0; q(\theta_0))} - \frac{\partial_\theta F(\theta_0; q(\theta_0))}{(\partial_x F(\theta_0; q(\theta_0)))^2}\partial_x R(\theta_0; q(\theta_0)) \\
&+ \left\{\frac{\partial_\theta F(\theta_0; q(\theta_0))\partial^2_{x,x} F(\theta_0; q(\theta_0))}{(\partial_x F(\theta_0; q(\theta_0)))^3}\right. \\
&\left. - \frac{\partial^2_{\theta,x} F(\theta_0; q(\theta_0))}{(\partial_x F(\theta_0; q(\theta_0)))^2}\right\}\mathbf{1}\{X \le q(\theta_0)\},
\end{aligned}$$

providing the CLT for the quantile sensitivity estimator, which in turn, justifies the square-root convergence rate. The convergence rate in Fu et al. (2009) is established by the square-root convergence of the second moment. Without a central limit theorem, there is no theoretical justification for constructing a confidence interval and performing hypothesis testing. An alternative quantile sensitivity estimator can be derived using the generalized likelihood ratio (GLR) method in Peng et al. (2018), which can deal with a large scope of discontinuities in the sample performance. The CLT for the quantile sensitivity estimator using GLR can also be proved by the theory established in this note, and an alternative proof using an empirical process approach can be found in Peng et al. (2017).

4. Confidence Intervals

An important use of results such as (10) in Theorem 2 is in the construction of large-sample confidence intervals, which quantify the estimation accuracy of

Monte Carlo–based computational procedures. To construct a confidence interval for α in (10), a natural approach is to estimate the variance of the limit rv on the right-hand side of (10). However, such an approach will require the estimation of $\alpha'(x^*)$ and $\phi'(x^*)$, which is nontrivial since the respective estimators, $\overline{Y}_n(\cdot)$ and $\overline{Z}_n(\cdot)$, of $\alpha(\cdot)$ and $\phi(\cdot)$, may be nondifferentiable (and even discontinuous) in a neighborhood of the point of interest x^*. One approach to deal with the sample path discontinuity of $\overline{Y}_n(\cdot)$ and $\overline{Z}_n(\cdot)$ is the previously mentioned GLR method (Peng et al. 2018). When $\phi'(\cdot)$ is the density of an rv X (as occurs in the quantile setting), an alternative is to use a density estimator for $\phi'(x^*)$ (see Silverman 1986).

To avoid the implementation complications associated with variance estimation, we now present an easily implemented alternative procedure based on *batching* or *sectioning* (Asmussen and Glynn 2007). These confidence interval procedures bypass the need to estimate the variance of the limit rv's appearing on right-hand side of (10). First, set m to be the (fixed) number of batches (or sections) so that there are $\lfloor n/m \rfloor$ samples per batch, with the samples being iid both within and across batches. We then let $\overline{Y}^{(i)}_{\lfloor n/m \rfloor}(\widehat{X}^{(i)}_{\lfloor n/m \rfloor})$ denote the statistic constructed using the samples in batch i, with $i = 1, \ldots, m$. Since the limit rv on the right-hand side of (10) is normal with mean zero, setting

$$\hat{\mu}_n \equiv \frac{1}{m} \sum_{i=1}^{m} \overline{Y}^{(i)}_{\lfloor n/m \rfloor}\left(\widehat{X}^{(i)}_{\lfloor n/m \rfloor}\right)$$

and

$$\hat{s}^2_n \equiv \frac{1}{m-1} \sum_{i=1}^{m} \left(\overline{Y}^{(i)}_{\lfloor n/m \rfloor}\left(\widehat{X}^{(i)}_{\lfloor n/m \rfloor}\right) - \hat{\mu}_n\right)^2,$$

we have $m^{1/2}(\hat{\mu}_n - \alpha)/\hat{s}_n \Rightarrow t_{m-1}$ as $n \to \infty$, with m fixed, where t_{m-1} is the Student's t-distribution with $m-1$ degrees of freedom. If we now choose z so that $\mathbb{P}(-z \le t_{m-1} \le z) = 1 - \delta$, then it follows that $[\hat{\mu}_n - z\hat{s}_n m^{-1/2}, \hat{\mu}_n + z\hat{s}_n m^{-1/2}]$ is an asymptotic $100(1-\delta)\%$ confidence interval for α. An alternative confidence interval recognizes that $\overline{Y}_n(\widehat{X}_n) = \hat{\mu}_n + o_P(n^{-1/2})$ (by virtue of the linear approximation (9)), so that we may replace $\hat{\mu}_n$ by $\overline{Y}_n(\widehat{X}_n)$ in a confidence interval. In particular, we can alternatively use the large-sample confidence interval $[\overline{Y}_n(\widehat{X}_n) - z\hat{s}_n m^{-1/2}, \overline{Y}_n(\widehat{X}_n) + z\hat{s}_n m^{-1/2}]$. The estimator $\overline{Y}_n(\widehat{X}_n)$ has a large-sample bias that is a factor m^{-1} of the bias associated with $\hat{\mu}_n$, so the latter interval may be preferred on that basis (see Munoz and Glynn 1997).

Acknowledgments

The authors thank the referees and editors for their comments and suggestions that have greatly improved the paper.

References

Asmussen S, Glynn PW (2007) *Stochastic Simulation: Algorithms and Analysis* (Springer, New York).

Billingsley P (1999) *Convergence of Probability Measures* (John Wiley & Sons, New York).

Bloznelis M, Paulauskas V (1994) A note on the central limit theorem for stochastically continuous processes. *Stochastic Processes Appl.* 53(2):351–361.

Fu MC, Hong LJ, Hu JQ (2009) Conditional Monte Carlo estimation of quantile sensitivities. *Management Sci.* 55(12):2019–2027.

Hong LJ, Hu Z, Liu G (2014) Monte Carlo methods for value-at-risk and conditional value-at-risk: A review. *ACM Trans. Model. Comput. Simulation* 24(4):1–37.

Munoz DF, Glynn PW (1997) A batch means methodology for estimation of a nonlinear function of a steady-state mean. *Management Sci.* 43(8):1121–1135.

Peng Y, Fu MC, Glynn PW, Hu JQ (2017) On the asymptotic analysis of quantile sensitivity estimation by Monte Carlo simulation. *Proc. 2017 Winter Simul. Conf.* (IEEE, Piscataway, NJ), 2336–2347.

Peng Y, Fu MC, Hu JQ, Heidergott B (2018) A new unbiased stochastic derivative estimator for discontinuous sample performances with structural parameters. *Oper. Res.* 66(2): 487–499.

Pflug GC, Wozabal N (2010) Asymptotic distribution of law-invariant risk functionals. *Finance Stochastics* 14(3):397–418.

Serfling RJ (1980) *Approximation Theorems of Mathematical Statistics* (John Wiley & Sons, New York).

Shapiro A, Dentcheva D, Ruszczynski A (2009) *Lectures on Stochastic Programming: Modeling and Theory* (Society for Industrial and Applied Mathematics, Philadelphia).

Silverman BW (1986) *Density Estimation for Statistics and Data Analysis* (Chapman and Hall/CRC, Boca Raton, FL).

Peter W. Glynn is the Thomas Ford Professor in the Department of Management Science and Engineering (MS&E) at Stanford University. He is a Fellow of INFORMS and of the Institute of Mathematical Statistics and was the co-winner of the John von Neumann Theory Prize from INFORMS in 2010. In 2012, he was elected to the National Academy of Engineering. His research interests lie in stochastic simulation, queueing theory, and statistical inference for stochastic processes.

Lin Fan is a PhD candidate in the Department of Management Science and Engineering (MS&E) at Stanford University, specializing in operations research. He holds an MS in statistics and an MS in mechanical engineering, both from Stanford University, and a BS in mechanical engineering from Georgia Institute of Technology. His research interests include statistical methods for stochastic processes, stochastic simulation, and online learning and decision-making.

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 affiliate faculty appointment in the Department of Management Science and Engineering, all at the University of Maryland. His research interests include simulation optimization and applied probability, with applications in supply chain management and financial engineering. He is a Fellow of INFORMS and IEEE.

Jian-Qiang Hu is the Hongyi Professor of Management Science in the School of Management, Fudan University. He was an associate professor with the Department of Mechanical Engineering and the Division of Systems Engineering at Boston University before joining Fudan University. His research interests include discrete-event stochastic systems, simulation, stochastic optimization, with applications in supply chain management, financial engineering, and healthcare.

Yijie Peng is an assistant professor in the Department of Industrial Engineering and Management at Peking University. He received his PhD degree from Fudan University in 2014 and was a research fellow at various universities before joining Peking University. His research interests include stochastic modeling and analysis, simulation optimization, machine learning, data analytics, and healthcare.