Submitted to *Operations Research*
manuscript (Please, provide the manuscript number!)

Technical Note: On Estimating Quantile Sensitivities via Infinitesimal Perturbation Analysis

Guangxin Jiang
Department of Mathematics, Tongji University, Shanghai, 200092, China, 1987_gxjiang@tongji.edu.cn

Michael C. Fu
Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, College Park, Maryland 20742, mfu@rhsmith.umd.edu

Hong (2009) introduced a general framework based on probability sensitivities and a conditional expectation relationship for estimating quantile sensitivities by infinitesimal perturbation analysis (IPA). We present an alternative more direct derivation of the IPA estimators that leads to simplified proofs for strong consistency and convergence rate of the unbatched estimator, and strong consistency and a central limit theorem for the batched estimator.

Key words: quantile, gradient estimation, Monte Carlo simulation, sensitivity analysis
History: This note was first submitted on January 2014

1. Introduction

Quantiles provide an alternative performance measure to expected values. In the financial services industry, for example, Value at Risk (VaR) is one of the most widely used standard risk measures (cf. Glasserman et al. 2000). In practice, simulation is often used to estimate quantiles, for which Serfling (1980) provides an overview of quantile estimation for independent and identically distributed data. Hong (2009) was the first to address the important problem of quantile sensitivity estimation in the simulation setting, introducing a general framework based on probability sensitivities. He showed that quantile sensitivities can be written as a conditional expectation, leading to an IPA estimator for which he proved asymptotic unbiasedness. He then introduced a batched IPA estimator, for which weak consistency and a central limit theorem were established. In related work, Hong and Liu (2009) studied the sensitivities of conditional value at risk, and Liu and Hong (2009) proposed a kernel estimator which has a better asymptotic convergence rate than the IPA estimator. It is well known that IPA cannot handle discontinuous performance measures, so Fu, Hong and Hu (2010) applied conditional Monte Carlo to derive a more general estimator that also had an improved convergence rate. Heidergott and Volk-Makarewicz (2012) used measure-valued differentiation method in estimating quantile sensitivities, and applied it to the Variance-Gamma process.

Simplicity of implementation and efficient computational properties make the IPA estimator attractive for quantile sensitivity estimation in practice. The purpose of this note is to provide an alternative more direct way of deriving many of the IPA results in Hong (2009). By simply differentiating the definition of the quantile and comparing it with a well-known IPA expression, the unbatched IPA estimator directly follows, without the need for probabilities sensitivities. As a result the proof for unbiasedness is also simplified. For the batched version of the IPA estimator, strong consistency and a central limit theorem are also established with streamlined proofs.

2. IPA Estimator of Quantile Sensitivities

Let $h(X(\theta);\theta)$ be a performance function, where $X(\theta)=(X_1(\theta),\dots,X_n(\theta))$ is a vector of random variables defined on a given probability space $(\Omega,\mathcal{F},\mathbb{P})$, $\theta\in\Theta\subset\mathbb{R}$ is a parameter that could be in the random variables or directly in the performance function or both, and Θ is an open interval. Let $q_\alpha(\theta)$ denote the α-quantile of $h(X(\theta);\theta)$ for any $0<\alpha<1$, where α is the corresponding probability, i.e., $\Pr\{h(X(\theta);\theta)\le q_\alpha(\theta)\}=\alpha$. The goal is to estimate $q'_\alpha(\theta)=dq_\alpha(\theta)/d\theta$ from samples of $h(X(\theta);\theta)$.

Let $F(\cdot;\theta)$ denote the distribution function of $h(X(\theta);\theta)$, where $F(x;\theta)=\Pr\{h(X(\theta);\theta)\leqslant x\}$. By definition of the quantile

$$\alpha=F(q_\alpha(\theta);\theta). \tag{1}$$

Let ∂_i denote the partial differentiation with respect to (w.r.t.) the ith argument of F. To construct the IPA estimator, we make the following assumption:

A1. In a neighborhood of $x=q_\alpha(\theta)$, $F(x;\theta)$ is continuously differentiable w.r.t. both arguments, and the density $\partial_1F(\cdot;\theta)$ is strictly positive for each $\theta\in\Theta$.

In practice, the distribution function of $h(X(\theta);\theta)$ is usually unavailable, so A1 cannot be verified directly. However, the distribution of X is generally known, so the relationship between X and $h(X(\theta);\theta)$ can be used to establish A1, as in the following proposition, whose proof is provided in Appendix A.

Proposition 1. *Let $y=h(x_1,\dots,x_n;\theta)$ and h_i^{-1} denote the inverse function of h w.r.t. the ith argument. Assume $X(\theta)$ is a continuous random vector with joint density $g(x_1,\dots,x_n;\theta)$. If in a neighborhood of $q_\alpha(\theta)$, $\exists i\in\{1,\dots,n\}$, s.t. $\forall x_j, j=1,\dots n$, the following conditions are satisfied:*
(i) $h_i^{-1}(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n;\theta)$ and $|\partial_y h_i^{-1}(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n;\theta)|$ exist and are continuously differentiable w.r.t. θ, and $|\partial_y h_i^{-1}(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n;\theta)|\neq 0$.

(ii) $g(x_1,\ldots,x_n;\theta)$ *is continuously differentiable w.r.t.* x_i *and* θ*, and* $g(x_1,\ldots,x_n;\theta)>0$ *for all* $\theta\in\Theta$*.*

Then $F(x;\theta)$ *satisfies A1.*

EXAMPLE 1. Let $h(X(\theta);\theta)=\theta e^X$ and $\theta>0$, where $X(\theta)$ follows the standard normal distribution with density function given by $g(x;\theta)=1/\sqrt{2\pi}\exp(-x^2/2)$, which is continuous and differentiable w.r.t. x and θ, and positive everywhere. Since $h^{-1}(y;\theta)=\ln y-\ln\theta$, $|\partial_y h^{-1}(y;\theta)|=|1/y|\neq 0$, $\partial_y\partial_\theta h^{-1}(y;\theta)=0$ exists, and the conditions of Proposition 1 are satisfied. Since the density function of $h(X;\theta)$ is given by

$$f(x;\theta)=\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\ln y-\ln\theta)^2}{2}},\ x>0,$$

it is easy to verify directly that its distribution function satisfies A1.

Under the assumption A1, chain rule differentiation of Equation (1) w.r.t. θ leads to

$$q'_\alpha(\theta)=-\frac{\partial_2F(x;\theta)}{\partial_1F(x;\theta)}\bigg|_{x=q_\alpha(\theta)}. \tag{2}$$

Equation (2) was used to derive quantile sensitivity estimators based on conditional Monte Carlo (smoothed perturbation analysis, cf. Fu and Hu 1997) in Fu, Hong and Hu (2010), and using measure-valued differentiation in Heidergott and Volk-Makarewicz (2012). These applications involved estimating the numerator and denominator of (2) separately.

Instead, we simply note that (2) has the same form as the well-known IPA expression for a single continuous random variable (cf. Suri and Zazanis 1988, Glasserman 1991, Fu 2006), and state the following lemma.

LEMMA 1 (SURI AND ZAZANIS 1988, GLASSERMAN 1991). *Let* $G(\cdot,\theta)$ *be a cumulative distribution function with density function* $\partial_1G(\cdot,\theta)$*, parameterized by* θ *on an open interval* $\theta\in\Theta$*. Suppose that (i) for each* $\theta\in\Theta$*,* $\partial_1G(\cdot,\theta)$ *is strictly positive on an open interval* I_θ *and zero elsewhere; and (ii)* G *is continuously differentiable w.r.t. both arguments on* $\{(x,\theta):\ x\in I_\theta,\theta\in\Theta\}$*. For* $Y(u,\theta)=G^{-1}(u,\theta)$ *defined by the inverse transform representation,*

$$\frac{\partial Y(u,\theta)}{\partial\theta}=-\frac{\partial_2G(Y(u,\theta),\theta)}{\partial_1G(Y(u,\theta),\theta)},\ 0<u<1,\ \theta\in\Theta. \tag{3}$$

Specifically, for the α-quantile of $h(X(\theta);\theta)$, often abbreviated henceforth as simply h, Equation (2) can be connected to Equation (3) by taking $u=\alpha$ and replacing $Y(u;\theta)$ by h and G by F. Therefore, a natural idea to estimate the quantile sensitivity is to first estimate the α-quantile of h for the right-hand side of (3), and then use the corresponding sample derivative (IPA estimator) of h as the estimate of the quantile derivative, representing the left-hand side of (3). This similar idea was used in Heidergott and Volk-Makarewicz (2012) for measure-valued differentiation and can also be used to derive a likelihood ratio method quantile sensitivity estimator.

Remark 2 of Hong (2009) relates Equation (3) to his result on probability sensitivities and a conditional expectation, where the condition is $h=\alpha$. The closed-form probability sensitivity results derived in Hong (2009) are not needed here. As a result, the theoretical proofs that follow are much simplified.

To present the IPA estimator for the quantile sensitivity, the l^{th}-order statistic from a set of i.i.d random variables will be denoted using the subscript (l), i.e., for a sample of size n,

$$h_{(1)} \le h_{(2)} \le \cdots \le h_{(\lceil \alpha n \rceil)} \le \cdots \le h_{(n)}, \tag{4}$$

and let $\hat{q}_\alpha^n(\theta) \triangleq h_{(\lceil \alpha n \rceil)}$ be the standard α-quantile estimator for h. Then (cf. Serfling 1980)

$$\lim_{n\to\infty} \hat{q}_\alpha^n(\theta) = q_\alpha(\theta) \text{ w.p.1.} \tag{5}$$

By A1, we know the conditions in Lemma 1 are satisfied. Then substituting in the α-quantile estimator for h into (3),

$$\left.\frac{\partial h}{\partial \theta}\right|_{h=\hat{q}_\alpha^n} = -\left.\frac{\partial_2 F(h;\theta)}{\partial_1 F(h;\theta)}\right|_{h=\hat{q}_\alpha^n}. \tag{6}$$

To distinguish from the usual IPA estimator, we use $\partial h/\partial\theta$ to denote the estimator given by (6), which is generally unavailable since it requires knowing $F(x;\theta)$. On the other hand, we use $dh/d\theta$ to denote the usual IPA, i.e., sample derivative estimator. These two are related through conditional expectation as follows, which is the main new result.

THEOREM 1. *Assume that $h(X(\theta);\theta)$ is differentiable w.r.t. $\theta\in\Theta$ w.p.1, and A1 is satisfied at the point y. Then,*

$$\mathbb{E}\left[\left.\frac{dh(X(\theta);\theta)}{d\theta}\right| h(X(\theta);\theta)=y\right] = -\frac{\partial_2 F(y;\theta)}{\partial_1 F(y;\theta)}. \tag{7}$$

Proof: Let $Y\sim F$. Then Y can be generated by $Y(\omega;\theta)=F^{-1}(U(\omega);\theta)$, where $U(\omega)$ is a uniform random variable on $[0,1]$. Note that $Y(\omega,\theta)\overset{d}{=}h(X(\theta);\theta)$. Given $U(\omega)=u$, by Lemma 1,

$$\frac{\partial Y(\omega;\theta)}{\partial\theta} = -\frac{\partial_2 F(Y(\omega;\theta);\theta)}{\partial_1 F(Y(\omega;\theta);\theta)}. \tag{8}$$

By the definition of $Y(\omega;\theta)$ and taking the conditional expectation on both sides of Equation (8),

$$\begin{aligned}
\mathbb{E}\left[\left.\frac{dh(X(\theta);\theta)}{d\theta}\right| U(\omega)=u\right] &= \mathbb{E}\left[\left.\frac{dY(\omega;\theta)}{d\theta}\right| U(\omega)=u\right] = \mathbb{E}\left[\left.\frac{\partial Y(\omega;\theta)}{\partial\theta}\right| U(\omega)=u\right] \\
&= \mathbb{E}\left[\left.-\frac{\partial_2 F(Y(\omega;\theta);\theta)}{\partial_1 F(Y(\omega;\theta);\theta)}\right| U(\omega)=u\right] \\
&= \mathbb{E}\left[\left.-\frac{\partial_2 F(F^{-1}(U(\omega);\theta);\theta)}{\partial_1 F(F^{-1}(U(\omega);\theta);\theta)}\right| U(\omega)=u\right] \\
&= -\frac{\partial_2 F(F^{-1}(u;\theta);\theta)}{\partial_1 F(F^{-1}(u;\theta);\theta)}.
\end{aligned} \tag{9}$$

Choose u such that $y = F^{-1}(u;\theta)$, then

$$\mathbb{E}\left[\frac{dh(X(\theta);\theta)}{d\theta}\middle| U(\omega)=u\right] = -\frac{\partial_2 F(y;\theta)}{\partial_1 F(y;\theta)}. \tag{10}$$

The event $\{U(\omega)=u\}$ is equivalent to the event$\{Y(\omega;\theta)=y\}$, i.e. $\{h(X(\theta);\theta)=y\}$, giving (7). □

Example 2. A simple well-known example is the following: $h(X(\theta);\theta) = X\theta$ where X is exponentially distributed with mean 1, i.e., $X \sim Exp(1)$. $dh/d\theta = X = h/\theta$, so $\mathbb{E}[dh/d\theta|h=y] = y/\theta$. On the other hand, $h(X;\theta)$ has the distribution function $F(y;\theta) = 1-\exp(-y/\theta)$, so $-\partial_2 F(y;\theta)/\partial_1 F(y;\theta) = -\partial_\theta(1-\exp(-y/\theta))/\partial_y(1-\exp(-y/\theta)) = y/\theta$.

Remark 1. Hong (2009) used $g(t;\theta) = \mathbb{E}[\partial_\theta h(X;\theta)|h(X;\theta)=t]$ to derive a closed-form expression for the quantile sensitivity. Theorem 1 provides an analytical expression for $g(t;\theta)$, i.e., $g(t;\theta) = -\partial_2 F(t;\theta)/\partial_1 F(t;\theta)$, which is very useful for verifying further assumptions on $g(t;\theta)$.

Example 3. We emphasize the difference between $\partial h/\partial\theta$ and $dh/d\theta$ using the following example. Set $h = \theta X_1 + X_2$, where X_1 and X_2 are independent standard normal random variables. Then, $h \sim N(0,\theta^2+1)$, and $\partial h/\partial\theta = -\partial_2 F(h;\theta)/\partial_1 F(h;\theta) = \theta h/(\theta^2+1)$. By taking partial derivatives directly, we obtain $dh/d\theta = X_1 \neq \partial h/\partial\theta$. However, it is easy to compute the correlation $\rho(X_1,h)$ and joint distribution, and thus show $\mathbb{E}[dh/d\theta|h] = \mathbb{E}[X_1|h] = \theta h/(\theta^2+1) = \partial h/\partial\theta$.

For notational convenience, let

$$\psi(y) \equiv \mathbb{E}\left[\frac{dh(X(\theta);\theta)}{d\theta}\middle| h(X(\theta);\theta)=y\right] = -\frac{\partial_2 F(y;\theta)}{\partial_1 F(y;\theta)}, \tag{11}$$

and $\hat{q'}^n_\alpha$ denote the estimator of quantile sensitivity q'_α. By Equation (7), substituting $y = \hat{q}^n_\alpha$,

$$\hat{q'}^n_\alpha = \psi(\hat{q}^n_\alpha) = -\frac{\partial_2 F(\hat{q}^n_\alpha;\theta)}{\partial_1 F(\hat{q}^n_\alpha;\theta)}. \tag{12}$$

If $dh/d\theta$ satisfies the following assumption, the quantile sensitivity estimator can be easily derived via the quantile estimator.

A2. There exists a function $\phi:\mathbb{R}\to\mathbb{R}$ s.t. for any $X=(X_1,\ldots,X_n)$, $dh(X)/d\theta = \phi(h(X))$.

Example 2 satisfies this condition, whereas Example 3 does not since it also contains another random variable X_2 rather than h only. We have

$$\psi(y) = \mathbb{E}\left[\frac{dh}{d\theta}\middle| h=y\right] = \mathbb{E}[\phi(h)|h=y] = \phi(y). \tag{13}$$

Then,

$$\hat{q'}^n_\alpha = \psi(\hat{q}^n_\alpha) = \phi(\hat{q}^n_\alpha), \tag{14}$$

which means $\partial h/\partial\theta = dh/d\theta$. In the next section, we can prove $\phi(\hat{q}_\alpha^n)$ is a strongly consistent estimator of the quantile sensitivity. However, in general A2 does not hold, so $dh/d\theta$ cannot be written as a function of h only, and an alternative approach is required.

For the sequence (4), we define the corresponding derivatives sequence

$$\frac{dh_{(1)}}{d\theta}, \frac{dh_{(2)}}{d\theta}, \ldots, \frac{dh_{(\lceil \alpha n\rceil)}}{d\theta}, \ldots, \frac{dh_{(n)}}{d\theta}. \tag{15}$$

Although $dh_{(\lceil \alpha n\rceil)}/d\theta \neq \mathbb{E}\left[dh/d\theta \middle| h = \hat{q}_\alpha^n\right]$ in general (cf. Example 3), they have the same expectation:

$$\begin{aligned}\mathbb{E}\left[\frac{dh_{(\lceil \alpha n\rceil)}}{d\theta}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{dh_{(\lceil \alpha n\rceil)}}{d\theta}\middle| h_{(\lceil \alpha n\rceil)}\right]\right] \\ &= \int_{-\infty}^{\infty} \mathbb{E}\left[\frac{dh_{(\lceil \alpha n\rceil)}}{d\theta}\middle| h_{(\lceil \alpha n\rceil)} = y\right] dF_{h_{(\lceil \alpha n\rceil)}}(y) \\ &= \int_{-\infty}^{\infty} \mathbb{E}\left[\frac{dh}{d\theta}\middle| h = y\right] dF_{h_{(\lceil \alpha n\rceil)}}(y) \\ &= \int_{-\infty}^{\infty} \mathbb{E}\left[\frac{dh}{d\theta}\middle| h = y\right] dF_{\hat{q}_\alpha^n}(y) = \mathbb{E}[\psi(\hat{q}_\alpha^n)],\end{aligned} \tag{16}$$

where $F_{h_{(\lceil \alpha n\rceil)}}(y)$ and $F_{\hat{q}_\alpha^n}(y)$ are the distributions of $h_{(\lceil \alpha n\rceil)}$ and $\hat{q}_\alpha^n$ respectively. The third equality holds since $h_{(\lceil \alpha n\rceil)}$ is one of the samples generated in (4). The fourth equality holds since we define $h_{(\lceil \alpha n\rceil)} = \hat{q}_\alpha^n$. Thus the IPA estimator for the quantile is simply given by

$$I_n \triangleq \frac{dh_{(\lceil \alpha n\rceil)}}{d\theta}. \tag{17}$$

In other words, the quantile sensitivity estimator is simply the IPA estimator for h corresponding to the α-quantile estimate. The key insight is to note that the expected value of (12) will converge to q'_α under appropriate conditions simply by applying the relationship (2). By (16), to obtain the expected value of the IPA estimator, we use k independent batches each of size n to form the batched IPA estimator

$$\hat{q'}_\alpha^{n,k} = \frac{1}{k}\sum_{i=1}^{k} I_{n,i}, \tag{18}$$

where $(I_{n,i}, i = 1, 2, \ldots, k)$ are independent realizations of the IPA estimator I_n.

3. Asymptotic Unbiasedness, Consistency and Central Limit Theorem

This section has two parts. In first part, we first consider the unbatched IPA under assumption A2, because of its easy implementation and lower computational cost compared with the batched estimator. We prove that the unbatched estimator (13) is strongly consistent, and provide the convergence rate, which is illustrated by a numerical example given in Appendix B. In the second part, we consider the statistical properties when $dh/d\theta$ cannot be written as a function of h only, i.e., A2 is not satisfied. We prove asymptotic unbiasedness of the unbatched IPA estimator given by (17), and strong consistency and a central limit theorem for the batched IPA estimator given by (18).

3.1. Unbatched Estimator

The IPA estimator of (13) is a strongly consistent estimator with convergence rate $O(n^{-\frac{1}{2}})$. The precise statement is given in the following proposition.

PROPOSITION 2. *If A1 and A2 are satisfied, then $\phi(\hat{q}_\alpha^n) \to q'_\alpha$ w.p.1 as $n \to \infty$. Moreover, if $\phi(x)$ is twice differentiable and bounded by a constant M, then $\phi(\hat{q}_\alpha^n) - q'_\alpha = O(n^{-\frac{1}{2}})$ in distribution as $n \to \infty$.*

Proof: By Equation (5) and the continuity of $\partial_2 F(x;\theta)$ and $\partial_1 F(x;\theta)$ in A1, we have

$$\partial_1 F(\hat{q}_\alpha^n;\theta) \to \partial_1 F(q_\alpha;\theta) \ \text{ w.p.1 as } \ n \to \infty,$$

and

$$\partial_2 F(\hat{q}_\alpha^n;\theta) \to \partial_2 F(q_\alpha;\theta) \ \text{ w.p.1 as } \ n \to \infty.$$

Since $\partial_1 F(\hat{q}_\alpha^n;\theta)$ is strictly positive by A1, and by (2) and (13),

$$\phi(\hat{q}_\alpha^n) = -\frac{\partial_2 F(\hat{q}_\alpha^n;\theta)}{\partial_1 F(\hat{q}_\alpha^n;\theta)} \to -\frac{\partial_2 F(q_\alpha;\theta)}{\partial_1 F(q_\alpha;\theta)} = q'_\alpha \ \text{ w.p.1 as } \ n \to \infty.$$

The following properties of $\hat{q}_\alpha^n$ are well known (cf. Serfling 1980),

$$n^{\frac{1}{2}}(\hat{q}_\alpha^n - q_\alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{(\partial_1 F(q_\alpha;\theta))^2}\right) \ \text{ as } \ n \to \infty, \tag{19}$$

where $\partial_1 F(q_\alpha;\theta)$ is strictly positive by A1, i.e., $\hat{q}_\alpha^n - q_\alpha = O(n^{-\frac{1}{2}})$. By Taylor's theorem,

$$\phi(\hat{q}_\alpha^n) - \phi(q_\alpha) = \phi'(q_\alpha)(\hat{q}_\alpha^n - q_\alpha) + \phi''(\xi)(\hat{q}_\alpha^n - q_\alpha)^2, \tag{20}$$

where ξ is between $\hat{q}_\alpha^n$ and q_α. By condition $\phi''(x) \le M$, $\phi''(\xi) \le M$. Then $\phi(\hat{q}_\alpha^n) - \phi(q_\alpha) = O(n^{-\frac{1}{2}})$ in distribution as $n \to \infty$. □

Under A2, the IPA estimator given by (17) is strongly consistent and there is no need to batch as in Hong (2009), and the quantile sensitivities can be obtained by a single run simulation. We provide an illustrative numerical example in Appendix B.

REMARK 2. For some special cases, $\phi(\hat{q}_\alpha^n)$ converges to a normal random variable with mean q'_α. If $\phi(x) = ax + b$ is a linear function, then $\phi(\hat{q}_\alpha^n) = a\hat{q}_\alpha^n + b$, so by Equation (19),

$$n^{-\frac{1}{2}}\left(\phi(\hat{q}_\alpha^n) - q'_\alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)a^2}{(\partial_1 F(q_\alpha;\theta))^2}\right) \ \text{ as } \ n \to \infty.$$

3.2. Batched Estimator

Generally, $dh/d\theta$ cannot be written as a function of h only, i.e., A2 is not satisfied, and we need to batch the estimator as in (18) and Hong (2009). In this subsection, we establish strong consistency and a central limit theorem for the batched estimator as in Hong (2009). First, asymptotic unbiasedness of the unbatched IPA estimator (17) is given in the following lemma.

LEMMA 2. *Suppose that* $\sup_n \mathbb{E}[I_n^2] < \infty$ *and* h *is differentiable w.r.t.* θ *w.p.1. If A1 is satisfied, then* $\mathbb{E}[I_n] \to q'_\alpha$ *as* $n \to \infty$.

Proof: Similarly as in the proof of Proposition 1, by assumption A1, we can get

$$-\frac{\partial_2 F(\hat{q}_\alpha^n,\theta)}{\partial_1 F(\hat{q}_\alpha^n;\theta)} \to -\frac{\partial_2 F(q_\alpha,\theta)}{\partial_1 F(q_\alpha;\theta)} \quad \text{as} \quad n \to \infty.$$

It suffices to prove $\partial_2 F(\hat{q}_\alpha^n;\theta)/\partial_1 F(\hat{q}_\alpha^n;\theta)$ is uniformly integrable.

$$\mathbb{E}\left[\left(\frac{\partial_2 F(\hat{q}_\alpha^n;\theta)}{\partial_1 F(\hat{q}_\alpha^n;\theta)}\right)^2\right] = \int_{-\infty}^{\infty}\left(\mathbb{E}\left[\frac{dh}{d\theta}\middle| h=y\right]\right)^2 dF_{\hat{q}_\alpha^n}(y) \le \int_{-\infty}^{\infty}\mathbb{E}\left[\left(\frac{dh}{d\theta}\right)^2\middle| h=\hat{q}_\alpha^n\right] dF_{\hat{q}_\alpha^n}(y) = \mathbb{E}[I_n^2], \tag{21}$$

where the inequality follows from Jensen's inequality for conditional expectation, and the second equality holds similarly as Equation (16). Since $\sup_n \mathbb{E}[I_n^2] < \infty$, we know $\partial_2 F(\hat{q}_\alpha^n;\theta)/\partial_1 F(\hat{q}_\alpha^n;\theta)$ is uniformly integrable. Then

$$\mathbb{E}\left[-\frac{\partial_2 F(\hat{q}_\alpha^n;\theta)}{\partial_1 F(\hat{q}_\alpha^n;\theta)}\right] \to \mathbb{E}\left[-\frac{\partial_2 F(q_\alpha;\theta)}{\partial_1 F(q_\alpha;\theta)}\right] \text{ as } n \to \infty.$$

By Equation (16) and Theorem 1,

$$\mathbb{E}[I_n] = \mathbb{E}[\psi(\hat{q}_\alpha^n)] = \mathbb{E}\left[-\frac{\partial_2 F(\hat{q}_\alpha^n;\theta)}{\partial_1 F(\hat{q}_\alpha^n;\theta)}\right] \to \mathbb{E}\left[-\frac{\partial_2 F(q_\alpha;\theta)}{\partial_1 F(q_\alpha;\theta)}\right] = q'_\alpha \text{ as } n \to \infty.$$

□

From asymptotic unbiasedness of the unbatched estimator, we can get strong consistency of the batched estimator.

THEOREM 2. *Under the assumptions in Lemma 2,* $\hat{q'}_\alpha^{n,k} \to q'_\alpha$ *w.p.1 as* $n \to \infty$ *and* $k \to \infty$.

Proof: Since $\hat{q'}_\alpha^{n,k} = 1/k \sum_{i=1}^{k} I_{n,i}$, where $I_{n,i}$ are i.i.d samples of I_n,

$$\mathbb{E}[\hat{q'}_\alpha^{n,k}] = \mathbb{E}[I_n] \quad \text{and} \quad Var[\hat{q'}_\alpha^{n,k}] = \frac{1}{k} Var[I_n]. \tag{22}$$

and $\sup_n \mathbb{E}[I_n^2] < \infty$, which means $\mathbb{E}[I_n]$ exists, we apply Kolmogorov's strong law of large numbers (SLLN),

$$\hat{q'}_\alpha^{n,k} = \frac{1}{k}\sum_{i=1}^{k} I_{n,i} \to \mathbb{E}[I_n], \text{ w.p.1 as } k \to \infty. \tag{23}$$

By Lemma 2, $\lim_{n\to\infty}\lim_{k\to\infty}\hat{q'}_{\alpha}^{n,k}=q'_{\alpha}$ w.p.1.

Similarly, if let $n\to\infty$ first, we have $\lim_{n\to\infty}\mathbb{E}[I_n]=q'_{\alpha}$. Then we apply SLLN, and obtain $\lim_{k\to\infty}\lim_{n\to\infty}\hat{q'}_{\alpha}^{n,k}=q'_{\alpha}$ w.p.1.

□

Next, we provide a central limit theorem for the estimator after stating the following assumptions and lemmas.

A3. $|\partial_2 F(x;\theta)|<M$ for some $M>0$, $\partial_2 F(x;\theta)$ and $\partial_1 F(x;\theta)$ are both twice differentiable w.r.t. the first argument, and bounded by M for all derivatives, i.e.,

$$|\partial_1(\partial_2 F(x;\theta))|\le M,\ |\partial_1(\partial_1 F(x;\theta))|\le M,\ |\partial_1^2(\partial_2 F(x;\theta))|\le M,\ |\partial_1^2(\partial_1 F(x;\theta))|\le M. \tag{24}$$

Remark 3. Similar to Proposition 1, we can verify A3 using the relationship between $X(\theta)$ and $h(X(\theta);\theta)$ when an analytical expression for $F(x;\theta)$ is not available.

Lemma 3 (Hong 2009). *Suppose that the density function $\partial_1 F(x;\theta)$ is continuously differentiable at $x=q_\alpha$ and $\partial_1 F(q_\alpha;\theta)>0$. Then, both $\mathbb{E}[\hat{q}_\alpha^n-q_\alpha]$ and $\mathbb{E}[(\hat{q}_\alpha^n-q_\alpha)^2]$ are of $O(n^{-1})$.*

Note that by A3 we can easily verify the conditions in Lemma 3. Then we have the following lemma, with the proof provided in Appendix C.

Lemma 4. *Under the assumptions in Lemma 2, if A3 is satisfied, then $\mathbb{E}[I_n]-q'_\alpha=O(n^{-1})$.*

Finally, the precise statement of the central limit theorem is as follows.

Theorem 3. *Under the assumptions in Lemma 4, if $\inf_n Var(I_n)>0$, then*

$$\frac{\hat{q'}_{\alpha}^{n,k}-q'_\alpha}{\left(Var[\hat{q'}_{\alpha}^{n,k}]\right)^{\frac{1}{2}}}\xrightarrow{d}\mathcal{N}(0,1)\quad as\quad n\to\infty,\ k\to\infty\ and\ \frac{k^{\frac{1}{2}}}{n}\to 0. \tag{25}$$

Proof: The left-hand side can be written as

$$\frac{\hat{q'}_{\alpha}^{n,k}-q'_\alpha}{\left(Var[\hat{q'}_{\alpha}^{n,k}]\right)^{\frac{1}{2}}}=\frac{\hat{q'}_{\alpha}^{n,k}-\mathbb{E}\left[\hat{q'}_{\alpha}^{n,k}\right]}{\left(Var[\hat{q'}_{\alpha}^{n,k}]\right)^{\frac{1}{2}}}+\frac{\mathbb{E}\left[\hat{q'}_{\alpha}^{n,k}\right]-q'_\alpha}{\left(Var[\hat{q'}_{\alpha}^{n,k}]\right)^{\frac{1}{2}}}. \tag{26}$$

By definition, $\hat{q'}_{\alpha}^{n,k}$ is the sample average of i.i.d replications of I_n. Therefore, by the Lindeberg-Levy central limit theorem, the first term on the right-hand side of (26) satisfies

$$\frac{\hat{q'}_{\alpha}^{n,k}-\mathbb{E}\left[\hat{q'}_{\alpha}^{n,k}\right]}{\left(Var[\hat{q'}_{\alpha}^{n,k}]\right)^{\frac{1}{2}}}\xrightarrow{d}\mathcal{N}(0,1)\quad\text{as}\quad k\to\infty. \tag{27}$$

From Lemma 3, we know $\mathbb{E}[I_n]-q'_\alpha = O(n^{-1})$, i.e., $|\mathbb{E}[I_n]-q'_\alpha|$ can be written as L/n for some constant L. By (22) and $\inf_n Var(I_n)>0$, we can obtain

$$\left|\frac{\mathbb{E}\left[\hat{q'}^{n,k}_\alpha\right]-q'_\alpha}{\left(Var[\hat{q'}^{n,k}_\alpha]\right)^{\frac{1}{2}}}\right| \leq \frac{L}{(Var[I_n])^{\frac{1}{2}}}\frac{k^{\frac{1}{2}}}{n} \to 0 \quad \text{as } \frac{k^{\frac{1}{2}}}{n}\to 0. \tag{28}$$

Therefore, the theorem is proved by Slutsky's theorem (Van der Vaart 2000). □

4. Conclusion

In this note, we first illustrate the relationship between quantile sensitivity estimation and usual gradient estimation, which leads to a simple derivation of the IPA estimator for quantile sensitivity. Under special conditions, we show the unbatched estimator is a strongly consistent estimator, and give the convergence rate. Similarly, we show the batched estimator is also a strongly consistent estimator and follows a limit central theorem. Although the batched estimator is the same as in Hong (2009), the parameter of interest in our setting is slightly more general, but more importantly, our alternative derivation leads to simpler more direct proofs.

Acknowledgments

This work was supported in part by the National Science Foundation (NSF) under Grants CMMI-0856256 and EECS-0901543, and by the Air Force Office of Scientific Research (AFOSR) under Grant FA9550-10-1-0340. The authors thank the Associate Editor and referees for their constructive comments that have led to improvements in the paper, especially Lemma 4, Proposition 2, and Theorem 2.

Appendix.

A. Proof of Proposition 1

We abbreviate $X(\theta)$ as X. Let $Y=(Y_1,\dots,Y_n)$ where $Y_j=X_j, j\neq i$ and $Y_i=h(X;\theta)=h(X_1,\dots,X_n;\theta)$. By (i), $X_i=h_i^{-1}(X_1,\dots,X_{i-1},Y_i,X_{i+1},\dots,X_n;\theta)=h_i^{-1}(Y_1,\dots,Y_n;\theta)$ exists. The joint density function of Y is given by

$$\begin{aligned} f_Y(y_1,\dots,y_n;\theta) &= |J|g(x_1,\dots,x_{i-1},h_i^{-1}(x_1,\dots,x_{i-1},y_i,x_{i+1},\dots,x_n;\theta),x_{i+1},\dots,x_n;\theta) \\ &= |J|g(y_1,\dots,y_{i-1},h_i^{-1}(y_1,\dots,y_{i-1},y_i,y_{i+1},\dots,y_n;\theta),y_{i+1},\dots,y_n;\theta), \end{aligned} \tag{29}$$

where the Jacobian is given by

$$J=\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \frac{\partial}{\partial y_i}h_i^{-1}(y_1,\dots,y_{i-1},y_i,y_{i+1},\dots,y_n;\theta). \tag{30}$$

Let $h_i^{-1}(y_1,\ldots,y_{i-1},y_i,y_{i+1},\ldots,y_n;\theta)$ be abbreviated as h_i^{-1}. The density function of $Y_i = h(X;\theta)$ can be regarded as a marginal distribution and given by

$$\begin{aligned} f(y_i;\theta) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_Y(y_1,\ldots,y_n;\theta)dy_1\cdots dy_{i-1}dy_{i+1}\cdots dy_n \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} |J|g(y_1,\ldots,y_{i-1},h_i^{-1},y_{i+1},\ldots,y_n;\theta)dy_1\cdots dy_{i-1}dy_{i+1}\cdots dy_n. \end{aligned} \tag{31}$$

By condition (i), $|J|$ is continuously differentiable w.r.t. θ. By (i) and (ii), $g(y_1,\ldots,y_n;\theta)$ is continuously differentiable w.r.t y_i and θ and h_i^{-1} is continuously differentiable w.r.t θ, it is easy to check that $g(y_1,\ldots,y_{i-1},h_i^{-1},y_{i+1},\ldots,y_n;\theta)$ is also continuously differentiable w.r.t. θ. Then, $f(t;\theta)$ is continuously differentiable w.r.t. θ, and

$$\frac{\partial F(x;\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\int_{-\infty}^{x} f(t;\theta)dt = \int_{-\infty}^{x}\frac{\partial f(t;\theta)}{\partial\theta}dt \tag{32}$$

exists and is continuous w.r.t. θ. Moreover, $\partial F(x;\theta)/\partial x = f(x;\theta)$ exists and is continuous w.r.t. x.

Finally, since $g(x_1,\ldots,x_n;\theta) > 0$ for all $\theta\in\Theta$ and $J\neq 0$, we obtain $f(t;\theta)>0$, therefore, $F(x;\theta)$ satisfies A1.

B. Numerical example for the unbatched IPA estimator

We consider the α-quantile of the Delta of a European call option. Let h denote the payoff function,

$$h \triangleq h(X;\theta) = (S_0e^X - K)^+, \tag{33}$$

where S_0 is the initial price of an underlying asset, K is strike price of the call option, $X\sim N((r-\frac{1}{2}\sigma^2)T,\sigma\sqrt{T})$, where r is the risk-free interest, T is the maturity and σ is the volatility. Assume that we are interested in the quantile sensitivities of h w.r.t. $\theta = S_0$, i.e. the α-quantile of the Delta. Then, differentiating (33),

$$\frac{dh}{d\theta} = \begin{cases} (h+K)/\theta & \text{if } \theta e^X \geq K, \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

Let $S_0 = 100,\ K = 100,\ r = 0.02,\ \sigma = 0.3,\ T = 1$. Generate $\{h_i, i=1,\ldots,n\}$, the order statistics $\{h_{(i)}, i=1,\ldots,n\}$ to obtain the estimator

$$I_n = \frac{h_{(\lceil\alpha n\rceil)}+K}{\theta}. \tag{35}$$

For different values of n and α, Table 1 shows the quantile sensitivity estimates, where each quantile sensitivity estimate is obtained by one set of sample size n, and the error is estimated by 1000 macroreplications. The true quantile of the payoff function h is given by $q_\alpha = \theta e^{z_\alpha} - K$, where z_α is the quantile of the normal distribution $N((r-\frac{1}{2}\sigma^2)T,\sigma\sqrt{T})$ and the true quantile sensitivities are given by $q'_\alpha = e^{z_\alpha}$.

Table 1 Quantile sensitivity estimates for European Call option

n	$\alpha=0.7$	Error	$\alpha=0.8$	Error	$\alpha=0.9$	Error
1000	1.1249	0.0113	1.2354	0.0137	1.4636	0.0183
4000	1.1429	0.0056	1.2568	0.007	1.4174	0.0093
16000	1.1377	0.0029	1.2567	0.0033	1.4504	0.0049
64000	1.1441	0.0014	1.2573	0.0017	1.4321	0.0023
256000	1.1419	0.0007	1.2547	0.0009	1.4331	0.0012
1024000	1.1409	0.0004	1.2558	0.0004	1.4317	0.0005
true value q'_α	1.1415		1.2554		1.4326	

C. Proof of Lemma 4

For ease of notation, denote $F=F(x;\theta)$ and $\varphi(x)=-\partial_2 F/\partial_1 F$. By A3, in a neighborhood of $x=q_\alpha$, we have $\partial_1 F>0$ and

$$|\partial_2 F|<M,\ |\partial_1\partial_2 F|<M,\ |\partial_1^2 F|<M,\ |\partial_1^2\partial_2 F|<M \text{ and } |\partial_1^3 F|<M. \tag{36}$$

For the first derivative of $\varphi(x)$,

$$|\varphi'(x)|=\left|-\frac{\partial_1\partial_2 F}{\partial_1 F}+\frac{\partial_2 F\partial_1^2 F}{(\partial_1 F)^2}\right|\le\frac{|\partial_1\partial_2 F|}{|\partial_1 F|}+\frac{|\partial_2 F\partial_1^2 F|}{(\partial_1 F)^2}\le M_1<\infty, \tag{37}$$

where $M_1>0$, and for the second derivative of $\varphi(x)$,

$$\begin{aligned}|\varphi''(x)|&=\left|-\frac{\partial_1^2\partial_2 F}{\partial_1 F}+\frac{2\partial_1\partial_2 F\partial_1^2 F+\partial_2 F\partial_1^3 F}{(\partial_1 F)^2}-\frac{2\partial_2 F(\partial_1^2 F)^2}{(\partial_1 F)^4}\right|\\&\le\frac{|\partial_1^2\partial_2 F|}{|\partial_1 F|}+\frac{|2\partial_1\partial_2 F||\partial_1^2 F|+|\partial_2 F||\partial_1^3 F|}{(\partial_1 F)^2}+\frac{2|\partial_2 F|(\partial_1^2 F)^2}{(\partial_1 F)^4}\le M_2<\infty.\end{aligned}$$

where $M_2>0$. By Lemma 3,

$$\mathbb{E}[\hat{q}_\alpha^n-q_\alpha]=O(n^{-1}), \tag{38}$$

$$\mathbb{E}\left[(\hat{q}_\alpha^n-q_\alpha)^2\right]=O(n^{-1}). \tag{39}$$

By Taylor's theorem,

$$\varphi(\hat{q}_\alpha^n)-\varphi(q_\alpha)=\varphi'(q_\alpha)(\hat{q}_\alpha^n-q_\alpha)+\varphi''(\xi)(\hat{q}_\alpha^n-q_\alpha)^2, \tag{40}$$

where ξ is between $\hat{q}_\alpha^n$ and q_α. Since $\lim_{n\to\infty}\hat{q}_\alpha^n=q_\alpha$ w.p.1, when n is sufficiently large, ξ can be in any neighborhood of q_α w.p.1. Then, $|\varphi''(\xi)|\le M_2$ w.p.1. Note that $q'_\alpha=\varphi(q_\alpha)$ and $\mathbb{E}[I_n]=\mathbb{E}\left[\varphi(\hat{q}_\alpha^n)\right]$, so taking the expectation on both sides of Equation (40), we have $\mathbb{E}[I_n]-q'_\alpha=O(n^{-1})$. $\square$

References

Fu, M. C. 2006. Gradient estimation. S. G. Henderson, B. L. Nelson, eds. *Handbooks in Operations Research and Management Science: Simulation*, Chap. 19. Elsevier, Amsterdam, 575-616.

Fu, M. C., Hong, L. J., Hu, J. Q. 2010. Conditional Monte Carlo estimation of quantile sensitivities. *Management Sci.* **55**(12) 2019-2027.

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

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis.* Kluwer Academic Publishers, Norwell, MA.

Glasserman, P., Heidelberger P., and Shahabuddin, P. 2000. Variance reduction techniques for estimating Value-at-Risk. *Management Sci.* **46**(1) 1349-1364.

Heidergott, B., Volk-Makarewicz, W. 2012. Sensitivity analysis of quantiles. *AENORM*, 2012, 26-31. http://www.aenorm.nl/editions/?edt=3&art=18

Hong, L. J. 2009. Estimating quantile sensitivities. *Oper. Res.* **57**(1) 118-130.

Hong, L. J., Liu, G.W. 2009. Simulating sensitivities of conditional value at risk. *Management Sci.* **55**(2) 281-293.

Liu, G. W., Hong, L. J. 2009. Kernel estimation of quantile sensitivities. *Naval Res. Logist* **56**(6) 511-525.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Suri, R., M. A. Zazanis. 1988. Perturbation analysis gives strongly consistent sensitivity estimates for the $M/G/1$ queue. *Management Sci.* **34**(1) 39–64.

Van der Vaart, A.W. 2000. *Asymptotic Statistics.* Cambridge University Press, Cambridge, MA.