Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Bias Reduction in Estimating Quantile Sensitivities ⋆

Guangxin Jiang * **Michael C. Fu** ** **Chenglong Xu** *

* *Department of Mathematics, Tongji University, Shanghai 200092, China (e-mail: 1987_gxjiang@tongji.edu.cn)*
** *Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA (e-mail:mfu@rhsmith.umd.edu)*

Abstract: In this paper, we introduce an Infinitesimal Perturbation Analysis (IPA) estimator with jackknifing to estimate quantile sensitivities, and theoretically prove the two-fold jackknife method reduces bias by eliminating the order $1/n$ bias in the original IPA estimator. Numerical examples in finance on portfolio return and options pricing are presented to illustrate the superiority of the new estimator over the original IPA estimator, especially for high and low quantile levels. Antithetic variates are used to further reduce the variance.

Keywords: Discrete event modeling and simulation, Stochastic hybrid systems, Monte Carlo simulation, Sensitivity analysis, Perturbation analysis.

1. INTRODUCTION

Quantile and quantile-related performance measures play an important role in many fields. For example, the quantile known as Value-at-Risk is used to measure the risk of loss of a portfolio in the finance industry. In the service industry, quantiles are used to measure the service quality level. The most common method for estimating the quantile uses order statistics, introduced by Serfling (1980) and David (1981).

Simulation optimization problems that contain quantiles often require estimates of quantile sensitivities. The classical gradient estimation methods (cf. Fu (2006)) can be applied in quantile sensitivity estimation. Hong (2009) used Infinitesimal Perturbation Analysis (IPA) to provide a batched estimator to estimate the quantile sensitivities. Jiang and Fu (2013) presented an alternative more direct derivation of the IPA estimators for both batched and unbatched estimators. Fu, Hong and Hu (2010) applied conditional Monte Carlo to derive a more general estimator with a wider applicability and improved convergence rate. Heidergott and Volk-Makarewicz (2010) used measure-valued differentiation, i.e., the weak derivatives method, in estimating quantile sensitivities for pricing options.

In this paper, we consider the IPA estimator and apply jackknifing to reduce bias in quantile sensitivity estimation, analogous to what was done for quantile estimation by Seila (1982). A thorough review of the jackknife method can be found in Miller (1974). The rest of the paper is organized as follows. In section 2, we briefly review the IPA estimator of quantile sensitivities. In section 3, the jackknife method is applied to the IPA estimator, and we provide a theoretical proof to show the reduction of the bias from $O(n^{-1})$ to $O(n^{-2})$. In section 4, two numerical examples are given to show the effectiveness of the new jackknife estimator. Section 5 is the conclusion.

⋆ Guangxin Jiang and Chenglong Xu were supported by NSF of China (11171256) and Shanghai Education Committee E-Research Project E03004. Michael C. Fu was supported in part by the National Science Foundation under Grants CMMI-0856256 and EECS-0901543 and by the Air Force Office of Scientific Research under Grant FA9550-10-1-0340.

2. IPA ESTIMATOR

Let $h(X(\theta);\theta)$ be a performance function, where $X(\theta)$ is a vector of random variables defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\theta \in \Theta \subset \mathbb{R}$ is a parameter. Let $F(x;\theta)$ denote the distribution function of $h(X(\theta);\theta)$, and let q_α denote the α-quantile of $h(X(\theta);\theta)$ for any $0 < \alpha < 1$, i.e., $F(q_\alpha;\theta) = \Pr\{h(X(\theta);\theta) \le q_\alpha\} = \alpha$. Then, the quantile sensitivity is given by

$$q'_\alpha = \frac{\partial_2 F(q_\alpha;\theta)}{\partial_1 F(q_\alpha;\theta)}, \tag{1}$$

where ∂_1 and ∂_2 denote the partial differentiation with respect to the first and second argument of a function respectively. This equation is quite similar to the definition of the classical IPA estimator (cf. Suri and Zazanis (1988)). 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, and $h(X(\theta);\theta)$ will often be abbreviated as simply h in following.

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

and let $\hat{q}^n_\alpha \triangleq h_{(\lceil \alpha n \rceil)}$ be the standard α-quantile estimator for h, and $\hat{p}^n_\alpha = \lceil n\alpha \rceil/(n+1)$ be the corresponding probability of $\hat{q}^n_\alpha$. Then (cf. Serfling (1980))

$$\lim_{n\to\infty} \hat{q}^n_\alpha = q_\alpha \text{ w.p.1},$$
$$\lim_{n\to\infty} \hat{p}^n_\alpha = \alpha \text{ w.p.1}. \tag{3}$$

Let $dh/d\theta$ denote the sample derivative obtained by taking partial derivatives directly from the performance function h. For example, $h = \theta Z_1 + Z_2$, $dh/d\theta = d(\theta Z_1 + Z_2)/d\theta = Z_1$. From sequence (2), the corresponding derivative sequence is given by

$$\frac{dh_{(1)}}{d\theta}, \frac{dh_{(2)}}{d\theta}, \ldots, \frac{dh_{(\lceil \alpha n \rceil)}}{d\theta}, \ldots, \frac{dh_{(n)}}{d\theta}. \tag{4}$$

Assume: (i) In a neighborhood of $x = q_\alpha$, $F(x;\theta)$ is continuously differentiable with respect to both the first argument and the second argument, and the density $\partial_1 F(x;\theta)$ is strictly positive for each $\theta \in \Theta$; (ii) $h(X(\theta);\theta)$ is differentiable w.p.1 with respect to $\theta \in \Theta$. Generally, the c.d.f. of h is difficult to obtain, whereas the IPA estimator of h is easy. The relationship between the IPA estimator and the partial derivatives of the c.d.f. is given by (Jiang and Fu , 2013, Theorem 1)

$$\varphi(y;\theta) := \mathbb{E}\left[\frac{dh(X(\theta);\theta)}{d\theta} \middle| h(X(\theta);\theta) = y \right] = -\frac{\partial_2 F(y;\theta)}{\partial_1 F(y;\theta)}, \tag{5}$$

which is slightly different from Suri and Zazanis (1988), since the performance function here is more complicated. Define the IPA estimator $I_n := dh_{(\lceil \alpha n \rceil)}/d\theta$ for sample size n. Since

$$\mathbb{E}[I_n] = \mathbb{E}[\varphi(\hat{q}_\alpha^n;\theta)] \to \mathbb{E}[q_\alpha;\theta] = q'_\alpha \text{ as } n \to \infty, \tag{6}$$

we need to batch the IPA estimator, with the batched IPA estimator corresponding to k independent batches given by

$$\hat{q'}_\alpha^{n,k} = \frac{1}{k} \sum_{i=1}^{k} I_{n,i}, \tag{7}$$

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

3. JACKKNIFE METHOD FOR IPA ESTIMATOR

By Lemma 2 in Jiang and Fu (2013), if $\sup_n \mathbb{E}[I_n^2] < \infty$, the IPA estimator is asymptotically unbiased, i.e.,

$$\mathbb{E}[I_n] \to q'_\alpha \text{ as } n \to \infty. \tag{8}$$

However, in practice, n is finite and the finite estimator is biased, which can lead to simulation coverage probabilities less than the theoretical average value for confidence intervals. Especially in high and low quantile sensitivity estimation, the accuracy of the estimator drops off quickly when the probability level α is close to 1 or 0. The effects of such bias are evident in the numerical results shown in Fig.5 in Hong (2009). In this section, we introduce a simple method, the two-fold jackknife method, which has been used by Seila (1982) to estimate the quantile of the waiting time in an M/M/1 queue, to reduce the bias in estimating the quantile sensitivities.

Suppose there are n samples in the ith batch, and $I_{n,i}$ is the quantile sensitivity estimator. Let $I_{n,i}^1$ and $I_{n,i}^2$ be the quantile sensitivity estimators computed from the first $n/2$ and second $n/2$ samples. The jackknife quantile sensitivity estimator of the ith batch $J_{n,i}$ is given by

$$J_{n,i} = 2I_{n,i} - \frac{1}{2}(I_{n,i}^1 + I_{n,i}^2). \tag{9}$$

We abbreviate $J_{n,i}$ by J_n in the following; similarly for I_n^1, I_n^2 and I_n. If the expected value of the estimator has the form

$$\mathbb{E}[I_n] = q'_\alpha + C/n + O(n^{-2}), \tag{10}$$

for some constant C, then $\mathbb{E}[J_n] - q'_\alpha = O(n^{-2})$, where $A_n = O(B_n)$ as $n \to \infty$ denotes $\lim_{n\to\infty} \sup |A_n/B_n| < \infty$. However in our setting, C also depends on n, so the classical approach cannot be used to prove that the jackknife method reduces bias, and an alternative approach is required. To prove the main theorem, the following two lemmas are needed. Lemma 3.1 provides the bias representation of the quantile estimator, with the proof provided in Appendix A.

Lemma 3.1. Let $F^{-1}(\cdot,\theta)$ be the inverse of the c.d.f of $h(X;\theta)$. Suppose A is an open set which contains q_α, and for $t \in A$, $f(t;\theta)$ is twice continuously differentiable and $f(t;\theta) > 0$. Then,

$$\begin{aligned} \mathbb{E}[\hat{q}_\alpha^n - q_\alpha] &= (F^{-1})'(\alpha;\theta)(\hat{p}_\alpha^n - \alpha) \\ &\quad + \frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{2(n+2)}(F^{-1})''(\hat{p}_\alpha^n;\theta) + O(n^{-2}), \quad (11) \\ \mathrm{Var}[\hat{q}_\alpha^n - q_\alpha] &= \frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{n+1}\left((F^{-1})(\hat{p}_\alpha^n;\theta)\right)^2 + O(n^{-2}). \end{aligned}$$

Remark By (11), it is easy to show that the two-fold jackknife method reduces bias in quantile estimation. Let the quantile estimator $\hat{q}_\alpha^n$ represent the estimator derived from n samples, $\hat{q}_{\alpha,1}^{n/2}$ and $\hat{q}_{\alpha,2}^{n/2}$ represent the estimators derived from the first and second $n/2$ samples. Then it is easy to prove $\mathbb{E}[2\hat{q}_\alpha^n - 1/2(\hat{q}_{\alpha,1}^{n/2} + \hat{q}_{\alpha,2}^{n/2})] - q_\alpha = O(n^{-2})$. Seila (1982) considers the two-fold jackknife estimator in regenerative processes rather than i.i.d random variables, so the bias representation in our setting cannot be written as (11).

By Lemma 3.1, we can derive the bias representation for the quantile sensitivity estimator given in Lemma 3.2. The proof of this lemma is provided in Appendix B.

Lemma 3.2. In addition to the conditions of Lemma 3.1, assume $\varphi(t;\theta)$ is twice differentiable with respect to t and $|\partial_1\varphi(t;\theta)| \le M$ and $|\partial_1^2\varphi(t;\theta)| \le M$ for $M > 0$. Then,

$$\begin{aligned} \mathbb{E}[I_n] - q'_\alpha &= \partial_1\varphi(q_\alpha;\theta)\Bigg((F^{-1})'(\alpha;\theta)(\hat{p}_\alpha^n - \alpha) \\ &\quad + \frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{2(n+2)}(F^{-1})''(\hat{p}_\alpha^n;\theta)\Bigg) + \frac{\partial_1^2\varphi(\tilde{q}_\alpha;\theta)}{2}\mathbb{E}[(\hat{q}_\alpha^n - q_\alpha)^2] \\ &\quad + O(n^{-2}), \end{aligned} \tag{12}$$

where $\tilde{q}_\alpha$ is between q_α and $\mathbb{E}[\hat{q}_\alpha^n]$.

Based on Lemma 3.1 and Lemma 3.2, Theorem 3.3 gives the bias of the jackknife estimator, with the proof provided in Appendix C.

Theorem 3.3. In addition to the conditions of Lemma 3.1, if $\varphi(t;\theta)$ is thrice differentiable with respect to t, and $|\partial_1^{(i)}\varphi(t;\theta)| \le M, i = 1, 2, 3$ for some $M > 0$, then $\mathbb{E}[J_n]$ is $O(n^{-2})$ for $n\alpha/2$ an integer, where J_n is the two-fold jackknife estimator defined by (9).

Remark (i) By Theorem 3.3, even though the coefficient of term $1/n$ in equation (10) is not constant, we can still prove that the jackknife method can eliminate the order $1/n$ term. (ii) The integrality condition can be used to determine the minimum batch size, i.e., $n\alpha/2$ should be an integer. For example, for $\alpha = 0.91$, the minimum n is 200.

(iii) All the conditions in Theorem 3.3 are quite common and almost the same as in Hong (2009) and Jiang and Fu (2013), which means the jackknife method is valid as long as the original estimator works in estimating the quantile sensitivities.

4. NUMERICAL EXPERIMENT

In this section, we use two numerical examples to test our results. The first example is the portfolio return model in Hong (2009), and we compare the bias reduction effect with his results. The second example is the α-quantile of an option.

4.1 Portfolio Return

It is well known that the jackknife method may inflate the variance of the estimator. In this example, we introduce antithetic variates to reduce the variance of the estimator. For details of the antithetic variates method, refer to Glasserman (2004). In this example, when we generate m random vectors by $X = F^{-1}(U)$, correspondingly we can generate another m random vectors by $\tilde{X} = F^{-1}(1-U)$. Let $X_{AV} = X/2 + \tilde{X}/2$ be the new random vector. Note that the antithetic estimator uses approximately twice the effort required to generate original estimator, so when we compare these two estimators, we should compare n replications of the antithetic estimator with $2n$ replications of the original estimator.

Consider a portfolio of three assets. The returns of the assets are denoted by X_1, X_2 and X_3. The portfolio return is

$$h(X;\theta) = \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3,$$

where θ_1, θ_2 and θ_3 are weights of the corresponding assets in this portfolio. Assume that $X = (X_1, X_2, X_3)$ follows a multivariate normal distribution with mean vector $\mu = (0.06, 0.15, 0.25)$ and variance covariance matrix

$$\Sigma = \begin{pmatrix} 0.02 & & \\ & 0.10 & \\ & & 0.22 \end{pmatrix} \begin{pmatrix} 1 & -0.3 & -0.2 \\ -0.3 & 1 & 0.2 \\ -0.2 & 0.2 & 1 \end{pmatrix} \begin{pmatrix} 0.02 & & \\ & 0.10 & \\ & & 0.22 \end{pmatrix}.$$

With $\theta = (\theta_1, \theta_2, \theta_3) = (0.2, 0.3, 0.5)$, we wish to estimate the quantile sensitivity with respect to θ_3. Since the distribution of $h(X;\theta)$ is also a normal distribution, we can obtain the analytical solution of the quantile sensitivity with respect to θ_3, which is used only for evaluating the quality of the simulated estimates. Using the same batch size $n = 200$ and the same number of batches $k = 200$ as in Hong (2009), the results are based on 1000 replications. We compare simulation bias, probability coverage, and mean square error (MSE) for four estimators: the original IPA estimator (nojack-nonantithetic in the figure), jackknife IPA estimator (jack-nonantithetic), antithetic IPA estimator (nojack+antithetic) and jackknife with antithetic IPA estimator (jack+antithetic).

Fig. 1 shows that the jackknife method reduces bias greatly and improves the simulation coverage probabilities significantly for high and low quantile levels, bringing the coverage quite close to the theoretical level of 90%.

Fig. 1. Bias and coverage probabilities (Ex.1)

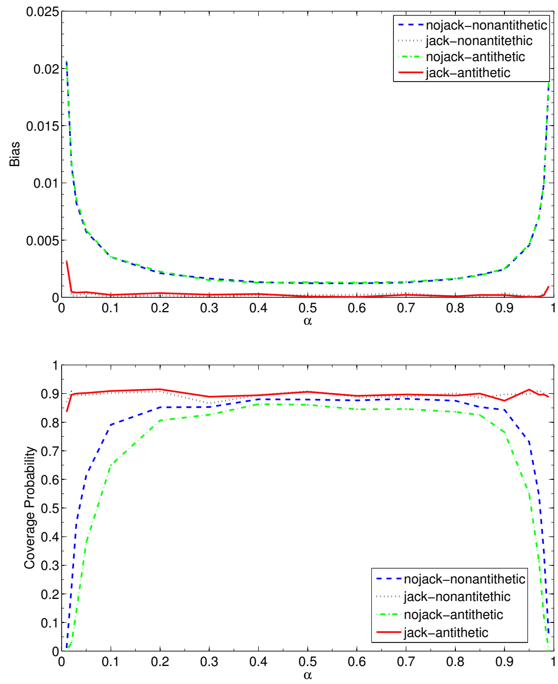
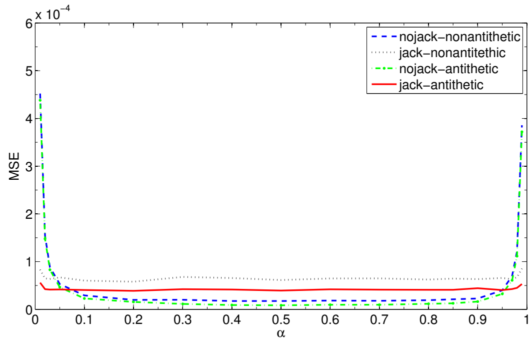


Fig. 2. MSE of estimates (Ex.1)



Although jackknifing reduces bias, it increases the variance of the estimator, so as shown in the figure, overall MSE increases except on the extremes. Antithetic variates helps reducing the gap. Although the jackknifing with antithetic IPA estimator has slightly larger MSE than the original IPA estimator, it reduces the bias at the extremes dramatically and improves the coverage probabilities significantly.

4.2 European Call Options

This example considers quantile sensitivities of European call options. The underlying asset of the option is driven by geometric Brownian motion (GBM)

$$dS_t = rS_t dt + \sigma S_t dW_t, \tag{13}$$

where r is a constant interest rate and σ is a constant volatility. Defining $x^+ = \max(x, 0)$, the payoff function is

$$V = e^{-rT}(S_T - K)^+, \tag{14}$$

where T is the maturity of this option and K is the strike price. For GBM, an explicit expression for the asset price is available, given by

$$S_t = S_0 e^{(r-\frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z}, \tag{15}$$

where Z is a standard normal random variable and S_0 is the initial price of the asset.

The IPA estimator $dV/d\theta$ is found by differentiating (14),

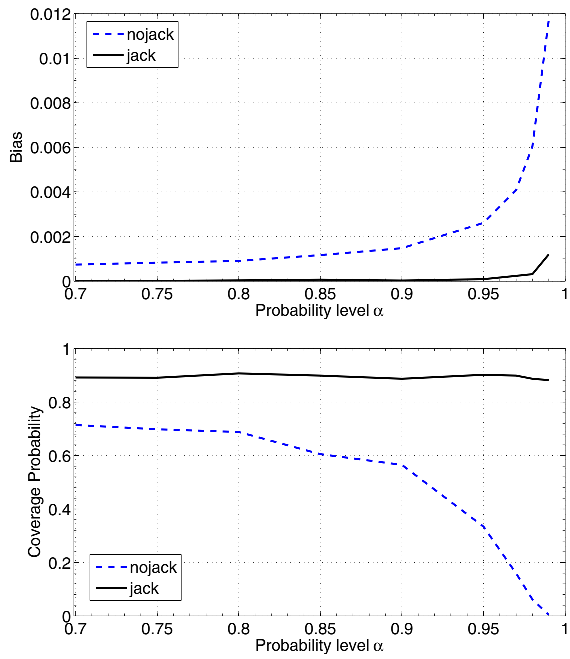
$$\frac{dV}{d\theta} = \left[\frac{d}{d\theta}\left(e^{-rT}\right)\right](S_T - K)^+ \\ +\mathbf{1}\{S_T \geq K\}e^{-rT}\frac{d}{d\theta}(S_T - K). \quad (16)$$

If we are interested in the sensitivity of the α-quantile of V with respect to the initial stock price S_0, the IPA estimator is given by $\exp(-rT)S_T/S_0\mathbf{1}\{S_T \geq K\}$, and the true value of the quantile sensitivity of V is given by

$$e^{-rT}\frac{S_T^\alpha}{S_0}\mathbf{1}\{S_T^\alpha \geq K\}, \quad (17)$$

where $S_T^\alpha = S_0 \exp\left((r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}z_\alpha\right)$, and z_α is the α-quantile of the standard normal distribution. In the numerical experiments, $S_0 = 100$, $K = 100$, $r = 0.02$, $\sigma = 0.1$, $T = 1$, the batch size is $n = 200$, the number of batches is $k = 200$, and the results are based on 1000 independent replications. Since low quantile sensitivities are 0, we are interested in the high quantile sensitivities, and the results are shown in Fig. 3 and Table 1.

Fig. 3. Bias and coverage probabilities (Ex.2)



The top panel in Fig. 3 shows the jackknife method provides significant bias reduction, as the bias is almost reduced to 0 (when $\alpha < 0.95$). In the bottom panel, the jackknifed IPA estimator basically achieves the theoretical coverage level of 90%, whereas the original estimator shows poor coverage probability due to the bias in estimation. For this problem, Table 1 indicates that unlike the last example, the bias dominates the MSE, so the reduction of bias also decreases the overall MSE substantially.

Table 1. MSE of estimates ($\times 10^{-6}$)

α	0.7	0.75	0.8	0.85	0.9
No jackknife	1.07	1.23	1.42	2.15	3.20
jackknife	0.65	0.68	0.76	0.99	1.34
α	0.95	0.96	0.97	0.98	0.99
No jackknife	8.52	14.3	19.0	40.2	144.1
jackknife	2.40	2.61	3.68	5.76	13.5

5. CONCLUDING REMARKS

In this paper, we have presented a new IPA estimator for quantile sensitivity estimation using jackknifing, which provably reduces bias by eliminating the order $O(n^{-1})$ bias in the original IPA estimator. The jackknife method is easy to apply and does not rely on special conditions. Two numerical experiments in finance are presented to illustrate the effectiveness of the jackknife method. The numerical results demonstrate that the jackknife method reduces bias significantly and improves the performance of the IPA estimator, especially for high and low quantile levels. Antithetic variates was shown to reduce the additional variance introduced by using jackknifing.

The framework in this paper cannot be directly applied to steady-state performance measures, since there is no Taylor series expansion for the quantile in dependent sequences such as the steady-state waiting time of a G/G/1 queue. However, this framework can be used for other derivative estimation methods, including the score function/likelihood ratio method and smoothed perturbation analysis, which are fruitful avenues for future research.

REFERENCES

Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37, 577–580.

David, H. (1981). *Order Statistics*, 2nd ed. Wiley, New York.

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., and Hu, J. Q. (2010). Conditional Monte Carlo estimation of quantile sensitivities. *Management Science*, 55, 2019-2027.

Glasserman, P. (2004). *Monte Carlo Simulation in Financial Engineering.* Springer, New York.

Hong, L. J. (2009). Estimating quantile sensitivities. *Operations Research*, 57, 118-130.

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

Jiang, G. and Fu, M. C. (2013). Technical Note: On estimating quantile sensitivities via infinitesimal perturbation analysis. Unpublished, submitted for publication.

Miller, R. G. (1974). The jackknife: A review. *Biometrika*, 61, 1-16.

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

Seila, A. F. (1982). A batching approach to quantile estimation in regenerative simulations. *Management Science*, 28, 573–581.

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

Appendix A. PROOF OF LEMMA 3.1

Since $f(t;\theta)$ is thrice continuously differentiable and $f(t;\theta) > 0$,

$$\begin{aligned}
&(F^{-1})'(\hat{p}^n_\alpha;\theta) = \frac{1}{f(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)},\\
&(F^{-1})''(\hat{p}^n_\alpha;\theta) = -\frac{f'(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}{f^3(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)},\\
&(F^{-1})'''(\hat{p}^n_\alpha;\theta) = \frac{3f'(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}{f^5(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}\\
&-\frac{f''(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)f(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}{f^5(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)},\\
&(F^{-1})''''(\hat{p}^n_\alpha;\theta) = \frac{8f''(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)f(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}{f^7(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}\\
&-\frac{f'''(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)f^2(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}{f^7(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}\\
&-\frac{f''(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)f'(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)f(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}{f^7(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}\\
&-\frac{15f'(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}{f^7(F^{-1}(\hat{p}^n_\alpha;\theta);\theta)}.
\end{aligned} \tag{A.1}$$

Since $f(t;\theta)$ is continuous for $t \in A$, $F^{-1}(y;\theta)$ is continuous in a neighborhood of $y = \alpha$. Furthermore, since $|\hat{p}^n_\alpha - \alpha| \le 1/(n+1)$, we can guarantee $\hat{p}^n_\alpha \in A$ for large n. Then $(F^{-1})^{(i)}(\hat{p}^n_\alpha;\theta);\theta), i = 1,2,3,4$, exist and are bounded for large n. By equation (4.6.3) in David (1981),

$$\begin{aligned}
&\mathbb{E}[\hat{q}^n_\alpha] = F^{-1}(\hat{p}^n_\alpha;\theta) + \frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{2(n+2)}(F^{-1})''(\hat{p}^n_\alpha;\theta)\\
&+\frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{(n+2)^2}\bigg((\frac{1}{2}-\hat{p}^n_\alpha)(F^{-1})'''(\hat{p}^n_\alpha;\theta)\\
&+\frac{1}{8}\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)(F^{-1})''''(\hat{p}^n_\alpha;\theta)\bigg) + o(n^{-2})\\
&= F^{-1}(\hat{p}^n_\alpha;\theta) + \frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{2(n+2)}(F^{-1})''(\hat{p}^n_\alpha;\theta) + O(n^{-2}).
\end{aligned} \tag{A.2}$$

Then,

$$\begin{aligned}
&\mathbb{E}[\hat{q}^n_\alpha - q_\alpha] = F^{-1}(\hat{p}^n_\alpha;\theta) - F^{-1}(\alpha;\theta)\\
&+\frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{2(n+2)}(F^{-1})''(\hat{p}^n_\alpha;\theta) + O(n^{-2})\\
&= (F^{-1})'(\alpha;\theta)(\hat{p}^n_\alpha - \alpha) + \frac{(F^{-1})''(\tilde{\alpha};\theta)}{2}(\hat{p}^n_\alpha-\alpha)^2\\
&+\frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{2(n+2)}(F^{-1})''(\hat{p}^n_\alpha;\theta) + O(n^{-2})\\
&= (F^{-1})'(\alpha;\theta)(\hat{p}^n_\alpha - \alpha) + \frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{2(n+2)}(F^{-1})''(\hat{p}^n_\alpha;\theta)\\
&+O(n^{-2}).
\end{aligned} \tag{A.3}$$

The second equality applies a Taylor series expansion, where $\tilde{\alpha}$ is between α and $\hat{p}^n_\alpha$. The third equality holds since $(\hat{p}^n_\alpha - \alpha)^2 \le 1/(n+1)^2$; hence $(F^{-1})''(\alpha;\theta)(\hat{p}^n_\alpha - \alpha)^2$ is $O(n^{-2})$. Similarly, we can also prove

$$\mathrm{Var}[\hat{q}^n_\alpha - q_\alpha] = \frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{n+1}\left((F^{-1})(\hat{p}^n_\alpha;\theta)\right)^2 + O(n^{-2}). \tag{A.4}$$

Appendix B. PROOF OF LEMMA 3.2

By Theorem 1 in Jiang and Fu (2013), $\mathbb{E}[I_n] = \mathbb{E}[\varphi(\hat{q}^n_\alpha;\theta)]$ and $q'_\alpha = \varphi(q_\alpha;\theta)$. Because $\varphi(t;\theta)$ is twice differentiable with respect to t,

$$\begin{aligned}
&\mathbb{E}[I_n] - q'_\alpha = \mathbb{E}[\varphi(\hat{q}^n_\alpha;\theta) - \varphi(q_\alpha;\theta)]\\
&= \partial_1\varphi(q_\alpha;\theta)\mathbb{E}\left[\hat{q}^n_\alpha - q_\alpha\right] + \frac{\partial_1^2\varphi(\tilde{q}_\alpha,\theta)}{2}\mathbb{E}\left[(\hat{q}^n_\alpha - q_\alpha)^2\right]\\
&= \partial_1\varphi(q_\alpha;\theta)\bigg((F^{-1})'(\alpha;\theta)(\hat{p}^n_\alpha - \alpha)\\
&+\frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{2(n+2)}(F^{-1})''(\hat{p}^n_\alpha;\theta) + O(n^{-2})\bigg)\\
&+\frac{\partial_1^2\varphi(\tilde{q}_\alpha,\theta)}{2}\mathbb{E}\left[(\hat{q}^n_\alpha - q_\alpha)^2\right]\\
&= \partial_1\varphi(q_\alpha;\theta)\bigg((F^{-1})'(\alpha;\theta)(\hat{p}^n_\alpha - \alpha)\\
&+\frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{2(n+2)}(F^{-1})''(\hat{p}^n_\alpha;\theta)\bigg)\\
&+\frac{\partial_1^2\varphi(\tilde{q}_\alpha,\theta)}{2}\mathbb{E}\left[(\hat{q}^n_\alpha - q_\alpha)^2\right] + O(n^{-2}).
\end{aligned} \tag{B.1}$$

The second equality is a Taylor series expansion, where $\tilde{q}_\alpha$ is between q_α and $\mathbb{E}[\hat{q}^n_\alpha]$. The third equality holds by Lemma 3.1, and the fourth equality holds because $\partial_1\varphi(q_\alpha;\theta)$ is bounded.

Appendix C. PROOF OF THEOREM 3.3

Since $J_{n,i} = 2I_{n,i} - 1/2(I^1_{n,i} + I^2_{n,i})$,

$$\begin{aligned}
&\mathbb{E}[J_n] - q'_\alpha = 2\mathbb{E}[I_{n,i} - q'_\alpha] - \frac{1}{2}(\mathbb{E}[I^1_{n,i} - q'_\alpha] + \mathbb{E}[I^2_{n,i} - q'_\alpha])\\
&= 2\bigg\{\partial_1\varphi(q_\alpha;\theta)\bigg((F^{-1})'(\alpha;\theta)(\hat{p}^n_\alpha - \alpha)\\
&+\frac{\hat{p}^n_\alpha(1-\hat{p}^n_\alpha)}{2(n+2)}(F^{-1})''(\hat{p}^n_\alpha;\theta)\bigg)\\
&+\frac{\partial_1^2\varphi(\tilde{q}_\alpha;\theta)}{2}\mathbb{E}[(\hat{q}^n_\alpha - q_\alpha)^2] + O(n^{-2})\bigg\}.\\
&-\bigg\{\partial_1\varphi(q_\alpha;\theta)\bigg((F^{-1})'(\alpha;\theta)(\hat{p}^{n/2}_\alpha - \alpha)\\
&+\frac{\hat{p}^{n/2}_\alpha(1-\hat{p}^{n/2}_\alpha)}{2(n/2+2)}(F^{-1})''(\hat{p}^{n/2}_\alpha;\theta)\bigg)\\
&+\frac{1}{4}\partial_1^2\varphi(\tilde{q}^1_\alpha;\theta)\mathbb{E}[(\hat{q}^{n/2}_{\alpha,1} - q_\alpha)^2]\\
&+\frac{1}{4}\partial_1^2\varphi(\tilde{q}^2_\alpha;\theta)\mathbb{E}[(\hat{q}^{n/2}_{\alpha,2} - q_\alpha)^2]\bigg\} - O(n^{-2})\\
&= \bigg\{\partial_1\varphi(q_\alpha;\theta)(F^{-1})'(\alpha;\theta)\left(2(\hat{p}^n_\alpha - \alpha) - (\hat{p}^{n/2}_\alpha - \alpha)\right)\bigg\}
\end{aligned}$$

$$
\begin{aligned}
&+\Bigg\{\partial_1\varphi(q_\alpha;\theta)\left(\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}(F^{-1})''(\hat{p}_\alpha^n;\theta)\right.\\
&\left.-\frac{\hat{p}_\alpha^{n/2}(1-\hat{p}_\alpha^{n/2})}{2(n/2+2)}(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)\right)\Bigg\}\\
&+\Bigg\{\partial_1^2\varphi(\tilde{q}_\alpha;\theta)\mathbb{E}[(\hat{q}_\alpha^n-q_\alpha)^2]\\
&-\frac{1}{4}\Big(\partial_1^2\varphi(\tilde{q}_\alpha^1;\theta)\mathbb{E}[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2]\\
&+\partial_1^2\varphi(\tilde{q}_\alpha^2;\theta)\mathbb{E}[(\hat{q}_{\alpha,2}^{n/2}-q_\alpha)^2]\Big)\Bigg\}+O(n^{-2}),
\end{aligned}
\tag{C.1}
$$

where $\hat{q}_{\alpha,1}^{n/2}$ and $\hat{q}_{\alpha,1}^{n/2}$ are the quantile sensitivity estimators computed from the first $n/2$ and second $n/2$ samples. If $\lceil n\alpha/2\rceil$ is an integer, $\lceil n\alpha\rceil$ is also an integer, and $\hat{p}_\alpha^{n/2}=n\alpha/(n+2)$ and $\hat{p}_\alpha^n=n\alpha/(n+1)$. Now, consider the first term (in the first braces) on the left hand side of the third equality in equation (C.1):

$$
\begin{aligned}
f_1 :&= \partial_1\varphi(q_\alpha;\theta)(F^{-1})'(\alpha;\theta)\left(2(\hat{p}_\alpha^n-\alpha)-(\hat{p}_\alpha^{n/2}-\alpha)\right)\\
&= \partial_1\varphi(q_\alpha;\theta)(F^{-1})'(\alpha;\theta)\left(\frac{2n\alpha}{n+1}-\frac{n\alpha}{n+2}-\alpha\right)\\
&= \partial_1\varphi(q_\alpha;\theta)(F^{-1})'(\alpha;\theta)\left(\frac{-2\alpha}{(n+1)(n+2)}\right)\\
&= O(n^{-2}).
\end{aligned}
\tag{C.2}
$$

The third equality holds since $\partial_1\varphi(q_\alpha;\theta)$ and $(F^{-1})'(\alpha;\theta)$ are both bounded.

Consider the second term (in the second braces) on the left hand side of the third equality in equation (C.1):

$$
\begin{aligned}
f_2 :=& \partial_1\varphi(q_\alpha;\theta)\left(\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}(F^{-1})''(\hat{p}_\alpha^n;\theta)\right.\\
&\left.-\frac{\hat{p}_\alpha^{n/2}(1-\hat{p}_\alpha^{n/2})}{2(n/2+2)}(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)\right)\\
=& \partial_1\varphi(q_\alpha;\theta)\left(\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}(F^{-1})''(\hat{p}_\alpha^n;\theta)\right.\\
&-\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)+\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)\\
&\left.-\frac{\hat{p}_\alpha^{n/2}(1-\hat{p}_\alpha^{n/2})}{2(n/2+2)}(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)\right)\\
=& \partial_1\varphi(q_\alpha;\theta)\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}\left((F^{-1})''(\hat{p}_\alpha^n;\theta)-(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)\right)\\
&+(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)\left(\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}-\frac{\hat{p}_\alpha^{n/2}(1-\hat{p}_\alpha^{n/2})}{2(n/2+2)}\right)\\
=& \partial_1\varphi(q_\alpha;\theta)\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}\\
&\left((F^{-1})'''(\hat{p}_\alpha^{n/2};\theta)(\hat{p}_\alpha^n-\hat{p}_\alpha^{n/2})+o(n^{-1})\right)\\
&+(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)\left(\frac{n\alpha(n+1-n\alpha)}{(n+1)^2(n+2)}-\frac{n\alpha(n+2-n\alpha)}{(n+2)^2(n+4)}\right)\\
=& \partial_1\varphi(q_\alpha;\theta)\frac{\hat{p}_\alpha^n(1-\hat{p}_\alpha^n)}{(n+2)}\\
&\left((F^{-1})'''(\hat{p}_\alpha^{n/2};\theta)(\frac{n\alpha}{(n+1)(n+2)})+o(n^{-1})\right)\\
&+(F^{-1})''(\hat{p}_\alpha^{n/2};\theta)\left(\frac{3n^3(\alpha-4\alpha^2)+9n^2(\alpha-7\alpha^2)+6n\alpha}{(n+1)^2(n+2)^2(n+4)}\right)\\
=& O(n^{-2}).
\end{aligned}
$$

Consider the third term (in the third braces) on the left-hand side of the third equality in equation (C.1):

$$
\begin{aligned}
f_3 :=& \partial_1^2\varphi(\tilde{q}_\alpha;\theta)\mathbb{E}[(\hat{q}_\alpha^n-q_\alpha)^2]-\frac{1}{2}\partial_1^2\varphi(\tilde{q}_\alpha^1;\theta)\mathbb{E}[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2]\\
=& \left\{\partial_1^2\varphi(\tilde{q}_\alpha;\theta)\mathbb{E}[(\hat{q}_\alpha^n-q_\alpha)^2]-\frac{1}{2}\partial_1^2\varphi(\tilde{q}_\alpha^1;\theta)\mathbb{E}[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2]\right\}\\
&+\frac{1}{2}\left\{\partial_1^2\varphi(\tilde{q}_\alpha;\theta)\mathbb{E}[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2]-\partial_1^2\varphi(\tilde{q}_\alpha^1;\theta)\mathbb{E}[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2]\right\}.
\end{aligned}
\tag{C.3}
$$

For the first term of equation (C.3), since $|\partial_1^2\varphi(\tilde{q}_\alpha;\theta)|\le M$,

$$
\begin{aligned}
g_1 :=& \partial_1^2\varphi(\tilde{q}_\alpha;\theta)\mathbb{E}[(\hat{q}_\alpha^n-q_\alpha)^2]-\frac{1}{2}\partial_1^2\varphi(\tilde{q}_\alpha;\theta)\mathbb{E}[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2]\\
\le& M\left|\mathbb{E}\left[(\hat{q}_\alpha^n-q_\alpha)^2\right]-\frac{1}{2}\mathbb{E}\left[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2\right]\right|\\
=& M\Bigg|\left(\mathbb{E}\left[\hat{q}_\alpha^n-q_\alpha\right]\right)^2+\mathrm{Var}\left[\hat{q}_\alpha^n-q_\alpha\right]\\
&-\frac{1}{2}\left(\mathbb{E}\left[\hat{q}_{\alpha,1}^{n/2}-q_\alpha\right]\right)^2-\frac{1}{2}\mathrm{Var}\left[\hat{q}_{\alpha,1}^{n/2}-q_\alpha\right]\Bigg|.
\end{aligned}
$$

Since $\left(\mathbb{E}\left[\hat{q}_\alpha^n-q_\alpha\right]\right)^2$ and $\left(\mathbb{E}\left[\hat{q}_\alpha^{n/2}-q_\alpha\right]\right)^2$ are both $O(n^{-2})$, by Lemma 3.1,

$$
\begin{aligned}
&\mathrm{Var}\left[\hat{q}_\alpha^n-q_\alpha\right]-\frac{1}{2}\mathrm{Var}\left[\hat{q}_{\alpha,1}^{n/2}-q_\alpha\right]\\
&= \frac{n\alpha(n+1-n\alpha)}{(n+1)^3}-\frac{n\alpha(n+2-n\alpha)}{(n+2)^3}+O(n^{-2})\\
&= O(n^{-2}).
\end{aligned}
$$

Therefore g_1 is $O(n^{-2})$. Now consider the second term of equation (C.3):

$$
\begin{aligned}
g_2 :=& \partial_1^2\varphi(\tilde{q}_\alpha;\theta)\mathbb{E}[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2]-\partial_1^2\varphi(\tilde{q}_\alpha^1;\theta)\mathbb{E}[(\hat{q}_{\alpha,1}^{n/2}-q_\alpha)^2]\\
=& \left(\partial_1^3\varphi(\tilde{q}_\alpha^1;\theta)(\tilde{q}_\alpha-\tilde{q}_\alpha^1)+o(\tilde{q}_\alpha-\tilde{q}_\alpha^1)\right)\mathbb{E}\left[(\hat{q}_\alpha^{n/2}-q_\alpha)^2\right].
\end{aligned}
\tag{C.4}
$$

Since $\tilde{q}_\alpha$ is between q_α and $\mathbb{E}[I_n]$, $|\mathbb{E}[I_n]-\tilde{q}_\alpha|\le|\mathbb{E}[I_n]-q_\alpha|=O(n^{-1})$, so $|\mathbb{E}[I_{n,1}]-\tilde{q}_\alpha^1|$ is $O(n^{-1})$. Then $\left|\mathbb{E}[I_n]-\tilde{q}_\alpha^1\right|=\left|\mathbb{E}[I_n]-q_\alpha+q_\alpha-\mathbb{E}[I_{n,1}]+\mathbb{E}[I_{n,1}]-\tilde{q}_\alpha^1\right|\le|\mathbb{E}[I_n]-q_\alpha|+|\mathbb{E}[I_{n,1}]-q_\alpha|+\left|\mathbb{E}[I_{n,1}]-\tilde{q}_\alpha^1\right|=O(n^{-1})$, so

$$
\left|\tilde{q}_\alpha-\tilde{q}_\alpha^1\right|\le|\mathbb{E}[I_n]-\tilde{q}_\alpha|+|\mathbb{E}[I_n]-\tilde{q}_\alpha^1|=O(n^{-1}).\tag{C.5}
$$

Because $\mathbb{E}\left[(\hat{q}_\alpha^{n/2}-q_\alpha)^2\right]$ is $O(n^{-1})$, by equation (C.5), g_2 is $O(n^{-2})$, then $f_3=g_1+g_2$ is $O(n^{-2})$. Similarly, $f_4:=\left\{\partial_1^2\varphi(\tilde{q}_\alpha;\theta)\mathbb{E}[(\hat{q}_\alpha^n-q_\alpha)^2]-\frac{1}{2}\partial_1^2\varphi(\tilde{q}_\alpha^2;\theta)\mathbb{E}[(\hat{q}_{\alpha,2}^{n/2}-q_\alpha)^2]\right\}$ is also $O(n^{-2})$. Since f_1, f_2, f_3 and f_4 are all $O(n^{-2})$, $\mathbb{E}[J_n]-q'_\alpha=f_1+f_2+f_3/2+f_4/2+O(n^{-2})=O(n^{-2})$.