Variance Comparison between Infinitesimal Perturbation Analysis and Likelihood Ratio Estimators to Stochastic Gradient

Zhenyu Cui[a], Yanchu Liu[b,*], Ruodu Wang[c]

[a] *School of Business, Stevens Institute of Technology, United States.*
[b] *Lingnan College, Sun Yat-sen University, China.*
[c] *Department of Statistics and Actuarial Science, University of Waterloo, Canada.*

Abstract

We theoretically compare variances between the Infinitesimal Perturbation Analysis (IPA) estimator and the Likelihood Ratio (LR) estimator to Monte Carlo gradient for stochastic systems. The results presented in [1] on variance comparison between these two estimators are substantially improved. We also prove a practically interesting result that the IPA estimators to European vanilla and arithmetic Asian options' Delta, respectively, have smaller variance when the underlying asset's return process is independent with the initial price and square integrable.

Keywords: Stochastic Gradient, Infinitesimal Perturbation Analysis, Likelihood Ratio, Variance Comparison, Option Delta

1. Introduction

The need to evaluate system performance in the face of uncertainty often arises in applications of various stochastic systems, including the queuing system, manufacturing and service system, transportation system, and financial system. This calls for stochastic simulation, which refers to the analysis of stochastic processes through the generation of sample paths (realizations) of the processes. Sometimes it is important to evaluate how the stochastic system evolves when there is a small change/perturbation in input parameters,

*Corresponding author. Lingnan College, Sun Yat-sen University, No. 135, Xingangxi Road, Haizhu District, Guangzhou, China. Email: liuych26@mail.sysu.edu.cn.

in order to test the robustness of the system to the input data. This is commonly referred to as stochastic gradient or sensitivity estimation problem.

The central question therein is to estimate the derivative of an expectation of a random performance measure. In particular, in a stochastic simulation setting, assume that we want to estimate the gradient of the expected value of some objective function of the underlying stochastic process with respect to a parameter θ of interest, i.e., $\nabla_\theta \alpha(\theta)$, where $\alpha(\theta) := \mathbb{E}[Y(\theta)]$ for $\theta \in \Theta$, and here Θ is an open subset of $\mathbb{R}$. We are interested in finding a random function $\xi(\theta)$ such that its expected value is equal to the gradient of $\alpha(\cdot)$, i.e., $\mathbb{E}[\xi(\theta)] = \nabla_\theta \alpha(\theta)$.

To estimate this stochastic gradient by simulation, there are two prevailing methods in the literature that can both yield unbiased estimators, i.e., the infinitesimal perturbation analysis (IPA) estimator and the likelihood ratio (LR) estimator. The basic idea behind IPA is to simply propose $\nabla_\theta Y(\theta)$ as the random function or equivalently the estimator. This is a very intuitive construction, and stems from the validity of the interchange $\mathbb{E}[\nabla_\theta Y(\theta)] = \nabla_\theta \mathbb{E}[Y(\theta)]$ under some suitable technical conditions, which has motivated a significant amount of research in the subsequent literature (see, e.g., [2, 3]). The basic idea behind the LR estimator is based on the representation of the expectation as an integral with respect to the probability density function, which depends on the parameter of interest. Thus we can differentiate the probability density function and then construct the LR estimator through weighting the payoff by this derivative of the density function. [4] provide an excellent and comprehensive up-to-date survey for all existing stochastic gradient estimators, including the above two, in various application domains including operations research and machine learning.

Since both estimators are unbiased if they exist, the natural mathematical question to consider is to compare their variances. It has been documented in the numerical examples in the literature (see for example [5] and [6]) that the IPA estimator, whenever it exists, almost always yields smaller variance than an LR estimator. Thus it is empirically recommended to use the IPA estimator as much as possible whenever it exists. However, there was no theoretical foundation justifying the variance comparison until the recent work [1], in which some sufficient conditions are provided that guarantee the variance of the LR estimator is higher than that of the IPA estimator. They also develop several counter examples where the IPA estimator has larger variance than its LR counterpart.

In this paper, we contribute to the literature by substantially extending the theoretical results of variance comparison in [1]. More specifically, we find that the original sufficient conditions in [1] can actually yield a sharper inequality in variance comparison. We further provide a simple and elegant sufficient condition to guarantee the variance of the IPA estimator is smaller by relaxing the conditions in [1]. A surprising result is found that the variance of the LR estimator is at least four times larger than that of the IPA estimator under some structural specifications only imposed on the objective function. Proceeding along this line, we finally obtain important results for financial derivatives practitioners that the IPA estimator to Delta of European vanilla and Asian options has smaller variance than the LR one, only requiring that the log return process of the underlying asset is independent of the initial price and square integrable. This requirement is highly non-restrictive since most of the popular financial models on the underlying asset's dynamics satisfy it, including but not limited to, e.g., the exponential Lévy models and stochastic volatility models with some mild conditions on the parameters to ensure the return process's square-integrability.

The structure of the paper is organized as follows. Section 2 describes the main theoretical comparison result. Section 3 focuses on the results on option's delta. Section 4 concludes the paper. All proofs are collected in the Appendix.

2. General Results

The goal of the stochastic gradient estimation problem is to seek an estimator to the quantity $\nabla_\theta \alpha(\theta) := \nabla_\theta \mathbb{E}[Y(\theta)]$. Here θ is the parameter of interest, $Y(\cdot)$ is a random function or random payoff of the underlying random variable $X(\theta)$, and $f_X(x;\theta)$ is the probability density function (pdf) of $X(\theta)$.

Assume uniform differentiability for both $Y(\theta)$ and $f_X(x;\theta)$ with respect to θ in a neighborhood $\Theta \in \mathbb{R}$ of some value of interest, and finiteness and exchangeability of all integrals that we subsequently encounter. The IPA estimator I_Y and the LR estimator L_Y are respectively given by

$$I_Y := \frac{d}{d\theta} Y(\theta) = Y'(\theta) \quad \text{and}$$

$$L_Y := Y(\theta) \frac{\frac{\partial}{\partial \theta} f_X(X(\theta);\theta)}{f_X(X(\theta);\theta)} = Y(\theta) \left. \frac{\frac{\partial}{\partial \theta} f_X(x;\theta)}{f_X(x;\theta)} \right|_{x=X(\theta)}.$$

It is straightforward to verify that both estimators are unbiased when they exist. If we have both unbiased estimators available, then a natural selection criteria will be based on the variances, or equivalently, the second-order moments, of the estimators. Their second-order moments are denoted by v_1 and v_2 respectively, i.e.,

$$v_1 = \mathbb{E}[(Y'(\theta))^2] \quad \text{and} \quad v_2 = \mathbb{E}\left[\left(Y(\theta)\frac{\frac{\partial}{\partial\theta} f_X(X(\theta);\theta)}{f_X(X(\theta);\theta)}\right)^2\right].$$

We can first show that the sufficient condition to guarantee $v_2 \geqslant v_1$, as stated in Assumption 1 of [1], can in fact yield a much sharper inequality on the variance comparison, i.e., the variance of the LR estimator is at least two times larger than that of the IPA estimator under their conditions.

Proposition 1. *If $Y''(\theta)Y(\theta) \geqslant 0$ and $\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) \leqslant 0$ hold, then $v_2 \geqslant 2v_1$ and $\mathrm{Var}(L_Y) \geqslant 2\mathrm{Var}(I_Y)$.*

On the other hand, if we are only interested in the condition to guarantee $v_2 \geqslant v_1$, as considered in [1], then we can further weaken the current assumption that $f_X(x;\theta)$ is log-concave in θ. We only need to require that the reciprocal of $f_X(x;\theta)$ is convex in θ, which is summarized in the following result.

Proposition 2. *If $Y''(\theta)Y(\theta) \geqslant 0$ holds, and $1/f_X(x;\theta)$ is convex in θ, then $v_2 \geqslant v_1$ and $\mathrm{Var}(L_Y) \geqslant \mathrm{Var}(I_Y)$.*

The above general results for variance comparison depend on some requirements on the pdf of X. It is possible to further relax such assumptions if we have some specific structure on the form of $Y(\theta)$, as shown in the following Proposition 3, which needs a sharp inequality first.

Lemma 1. *Let Z be a positive random variable with differentiable probability density function f_Z, and $n \geqslant 0$ such that $\mathbb{E}[Z^n] < \infty$. Then*

$$\mathbb{E}\left[Z^{n+2}\left(\frac{f_Z'(Z)}{f_Z(Z)}\right)^2\right] > (n+1)^2\mathbb{E}[Z^n], \tag{1}$$

and here the constant $(n+1)^2$ is sharp and can not be further improved.

It turns out that this lemma is new to the literature with a mathematically quite tricky proof which should be of independent interest. A special case of this result is related to the comparison between $\mathbb{E}\left[Z^4\left(\frac{f'_Z(Z)}{f_Z(Z)}\right)^2\right]$ and $\mathbb{E}[Z^2]$. It can provide us a surprising result that the variance of the LR estimator could be at least four times larger than that of the IPA estimator in some specific settings.

Proposition 3. *If $Y(\theta) = g(\theta)Z$ for a differentiable function g with $g(\theta)g'(\theta) \neq 0$ and a positive random variable Z, where Z doesn't depend on θ and has a differentiable pdf satisfying $\mathbb{E}[Z^2] < \infty$, then we have $\mathrm{Var}(L_Y) > 4\mathrm{Var}(I_Y)$.*

3. On Financial Option's Delta Estimators

In financial derivatives markets, practitioners typically need to calculate the so-called Delta of financial options in their daily job. Delta is the first-order partial derivative of the option price with respect to the initial price of the option's underlying asset. We focus on a European call option with a maturity T, while a European put option's Delta can then be obtained by the put-call parity. Let S_t denote the underlying asset's price at time t ($0 \leqslant t \leqslant T$). In the setting of Section 2, we can view the option payoff at T as $Y(\theta)$ with the parameter $\theta = S_0$, i.e., the initial price. Most standard call options popularly traded in the market have the payoff with the form $Y(\theta) := (X(\theta) - K)^+$ where $X(\theta)$ is some functional of the underlying asset's prices at certain time spots, K is a constant denoting the strike price, and $x^+ := \max\{x, 0\}$. For example, at maturity the European vanilla option has the payoff $Y = (S_T - K)^+$ and the arithmetic Asian option has the payoff $Y = (\frac{1}{m}\sum_{i=1}^m S_{t_i} - K)^+$ where $0 < t_1 < \cdots < t_m = T$. We have the following results for the general setting of this practical problem.

Proposition 4. *Suppose that $Y(\theta) = (X(\theta) - K)^+$ with $1 > p := \mathbb{P}(X(\theta) > K) > 0$ for a constant $K > 0$ and $X(\theta) = g(\theta)Z$ for a differentiable function g with $g(\theta)g'(\theta) \neq 0$ and a positive random variable Z with a differentiable pdf satisfying $\mathbb{E}[Z^2] < \infty$. We have*

$$\frac{\mathrm{Var}(L_Y)}{\mathrm{Var}(I_Y)} \geqslant \min\left\{4, \frac{3\delta^2 + 2\delta + (1-p)(1+\delta)^2}{(1-p)(1+\delta)^2}\right\} > 1, \tag{2}$$

where $\delta := \mathbb{E}[(X(\theta) - K)/K | X(\theta) > K]$.

Note that the general results presented in Section 2 depend on some requirements on the probability density function of X. However, the above Proposition 4 reveals that, for a large domain of financial models, as long as the underlying asset's log return process does not depend on its initial price and is square integrable, the IPA estimator to Delta has smaller variance than the LR one. In other words, we actually do not need specific assumption on the shape of the probability function density for the purpose of variance comparison of the two Delta estimator. Such models (subject to parameter restrictions to guarantee the square integrability of the log return) include, e.g., the Black-Scholes model, Merton's jump diffusion model ([7]), Kou's double exponential jump diffusion model ([8]), and variance gamma model ([9]), or generally, the exponential Lévy processes as asset prices models. Our results also hold for the stochastic volatility model, e.g., the Heston model ([10]). This demonstrates the advantage of the IPA estimator in practical task of Delta estimation, and justifies to some extent the wide-spread empirical findings (or, the rule-of-thumb proposed in [6]) that the IPA estimator is usually preferred whenever it is applicable.

Proposition 4 also substantially strengthens the variance comparison results for Delta of Asian options as documented in the Example 2 of Section 4.2 in [1]. They only considered the arithmetic Asian option with Black-Scholes model for the stock price. Directly applying Proposition 4 in the setting of Example 2 in [1], it is straightforward to verify that the IPA estimator has smaller variance than that of the LR estimator for Delta of an arithmetic Asian option case with underlying asset belonging to the models (with parameter restrictions) mentioned above.

4. Conclusion and Future Work

We mainly theoretically advance the comparison problem between variances of IPA and LR estimators in this paper. The sufficient conditions we propose to guarantee that the IPA estimator has smaller variance are easy to verify in concrete applications. For the Delta estimation problem in financial engineering, we assert the smaller variance of IPA estimator for European vanilla and arithmetic Asian options within a large class of underlying asset's dynamics whose log return process does not depend on the initial value and square integrable. Future work may include *sufficient and necessary* conditions for this variance comparison problem and comparing the variance of IPA estimator with other types of unbiased estimators in the literature rather than the LR one, e.g., the Malliavin estimator ([11]).

Acknowledgement

Yanchu Liu has been financially supported by National Natural Science Foundation of China (71991474 and 71721001), and the Fundamental Research Funds for the Central Universities (19wkzd11). Ruodu Wang acknowledges financial support from NSERC grants (RGPIN-2018-03823 and RGPAS-2018-522590).

References

[1] Z. Cui, M. C. Fu, J.-Q. Hu, Y. Liu, Y. Peng, L. Zhu, On the variance of single-run unbiased stochastic derivative estimators, INFORMS Journal on Computing 32 (2) (2020) 390–407.

[2] M. Fu, J.-Q. Hu, Conditional Monte Carlo: Gradient estimation and optimization applications, Vol. 392, Kluwer Academic Publishers, 2012.

[3] J. Hong, Estimating quantile sensitivities, Operations Research 57 (1) (2009) 118–130.

[4] S. Mohamed, M. Rosca, M. Figurnov, A. Mnih, Monte Carlo gradient estimation in machine learning, Journal of Machine Learning Research 21 (132) (2020) 1–62.

[5] M. Broadie, P. Glasserman, Estimating security price derivatives using simulation, Management Science 42 (2) (1996) 269–285.

[6] P. Glasserman, Monte Carlo methods in Financial Engineering, Springer, New York, 2004.

[7] R. Merton, Option pricing when underlying stock returns are discontinuous, Journal of Financial Economics 3 (1-2) (1976) 125–144.

[8] S. Kou, A jump-diffusion model for option pricing, Management Science 48 (8) (2002) 1086–1101.

[9] D. Madan, P. Carr, E. Chang, The variance gamma process and option pricing, Review of Finance 2 (1) (1998) 79–105.

[10] S. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options, The Review of Financial Studies 6 (2) (1993) 327–343.

[11] E. Fournié, J.-M. Lasry, J. Lebuchoux, P.-L. Lions, N. Touzi, Applications of Malliavin calculus to Monte Carlo methods in finance, Finance and Stochastics 3 (4) (1999) 391–412.

Appendix

This appendix section collects the technical proofs of Propositions 1-4 and Lemma 1 in the main text.

Proof of Proposition 1.

Note that we have the following identity

$$\left(y \frac{\frac{\partial}{\partial \theta} f_X(x;\theta)}{f_X(x;\theta)} \right)^2 = y^2 \left(\frac{\frac{\partial}{\partial \theta} f_X(x;\theta)}{f_X(x;\theta)} \right)^2 = y^2 \left(\frac{\frac{\partial^2}{\partial \theta^2} f_X(x;\theta)}{f_X(x;\theta)} - \frac{\partial^2}{\partial \theta^2} \log f_X(x;\theta) \right).$$

Therefore, we can express

$$v_2 = \mathbb{E}\left[\left(Y(\theta) \frac{\frac{\partial}{\partial \theta} f_X(X(\theta);\theta)}{f_X(X(\theta);\theta)} \right)^2 \right] = \mathbb{E}\left[Y^2(\theta) \frac{\frac{\partial^2}{\partial \theta^2} f_X(X(\theta);\theta)}{f_X(X(\theta);\theta)} - Y^2(\theta) \frac{\partial^2}{\partial \theta^2} \log f_X(X(\theta);\theta) \right]. \tag{3}$$

We note that

$$\begin{aligned} \mathbb{E}\left[Y^2(\theta) \frac{\frac{\partial^2}{\partial \theta^2} f_X(X(\theta);\theta)}{f_X(X(\theta);\theta)} \right] &= \frac{\partial^2}{\partial x^2} \mathbb{E}\left[Y^2(\theta) \frac{f_X(X(\theta),x)}{f_X(X(\theta);\theta)} \right] \Big|_{x=\theta} \\ &= \frac{\partial^2}{\partial x^2} \mathbb{E}^{Q(x)} \left[Y^2(x) \right] \Big|_{x=\theta} \\ &= 2\mathbb{E}[Y'(\theta)^2] + 2\mathbb{E}[Y''(\theta)Y(\theta)] \\ &= 2v_1 + 2\mathbb{E}[Y''(\theta)Y(\theta)], \end{aligned} \tag{4}$$

where in the second equality we have used the following measure change argument: By defining a new probability measure $Q(x)$ with the Radon-Nikodym derivative given by $\frac{dQ(x)}{dP} := \frac{f_X(X(\theta),x)}{f_X(X(\theta);\theta)}$ where P is the probability measure underlying the expectation operator $\mathbb{E}$, we then have $\mathbb{E}[\cdot] \equiv \mathbb{E}^{Q(\theta)}[\cdot]$ and (4) holds.

As a consequence of (4), and noting the identity (3), we have

$$v_2 - 2v_1 = 2\mathbb{E}[Y''(\theta)Y(\theta)] - \mathbb{E}\left[Y^2(\theta)\frac{\partial^2}{\partial\theta^2}\log f_X(X(\theta);\theta)\right], \tag{5}$$

and it is clear that with the given (sufficient) assumptions, we have $v_2 \geqslant 2v_1$. Furthermore, we have the following result

$$\mathrm{Var}(L_Y) = v_2 - (\mathbb{E}[Y'(\theta)])^2 \geqslant 2v_1 - (\mathbb{E}[Y'(\theta)])^2 \geqslant 2v_1 - 2(\mathbb{E}[Y'(\theta)])^2 = 2\mathrm{Var}(I_Y).$$

This completes the proof. $\square$

Proof of Proposition 2.
Note that by (5) we have

$$v_2 - v_1 = \mathbb{E}[Y'(\theta)^2] + 2\mathbb{E}[Y''(\theta)Y(\theta)] - \mathbb{E}\left[Y^2(\theta)\frac{\partial^2}{\partial\theta^2}\log f_X(X(\theta);\theta)\right]. \tag{6}$$

Given the assumption that $Y''(\theta)Y(\theta) \geqslant 0$ holds, there is

$$\mathbb{E}[Y'(\theta)^2] + 2\mathbb{E}[Y''(\theta)Y(\theta)] \geqslant \mathbb{E}[Y'(\theta)^2] + \mathbb{E}[Y''(\theta)Y(\theta)] = \frac{1}{2}\mathbb{E}\left[Y^2(\theta)\frac{\frac{\partial^2}{\partial\theta^2}f_X(X(\theta);\theta)}{f_X(X(\theta);\theta)}\right], \tag{7}$$

where the last equality is based on (4) in the proof of Proposition 1.

In order for $v_2 \geqslant v_1$, by (6) and (7), it suffices to have

$$\mathbb{E}\left[Y^2(\theta)\frac{\partial^2}{\partial\theta^2}\log f_X(X(\theta);\theta)\right] \leqslant \frac{1}{2}\mathbb{E}\left[Y^2(\theta)\frac{\frac{\partial^2}{\partial\theta^2}f_X(X(\theta);\theta)}{f_X(X(\theta);\theta)}\right],$$

which can be implied if we have,

$$\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) \leqslant \frac{1}{2}\frac{\frac{\partial^2}{\partial\theta^2}f_X(x;\theta)}{f_X(x;\theta)}. \tag{8}$$

Because

$$\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) = \frac{\frac{\partial^2}{\partial\theta^2}f_X(x;\theta)}{f_X(x;\theta)} - \left(\frac{\frac{\partial}{\partial\theta}f_X(x;\theta)}{f_X(x;\theta)}\right)^2,$$

we know that (8) can be further implied by the following condition

$$\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) \leqslant \left(\frac{\frac{\partial}{\partial\theta}f_X(x;\theta)}{f_X(x;\theta)}\right)^2. \tag{9}$$

On the other hand, we have

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\frac{1}{f_X(x;\theta)} &= \frac{\partial^2}{\partial\theta^2}e^{-\log f_X(x;\theta)} \\
&= \left(-\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) + \left(\frac{\partial}{\partial\theta}\log f_X(x;\theta)\right)^2\right)e^{-\log f_X(x;\theta)} \\
&= \left(-\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) + \left(\frac{\frac{\partial}{\partial\theta}f_X(x;\theta)}{f_X(x;\theta)}\right)^2\right)e^{-\log f_X(x;\theta)} \\
&\geqslant 0,
\end{aligned}$$

if and only if (9) holds. This completes the proof. □

Proof of Lemma 1. Assume $\mathbb{E}\left[Z^n\left(\frac{Zf_Z'(Z)}{f_Z(Z)}\right)^2\right] < \infty$ in the following, since there is nothing to show otherwise. If $n = 0$, then Jensen's inequality yields

$$\mathbb{E}\left[\left(\frac{Zf_Z'(Z)}{f_Z(Z)}\right)^2\right] > \left(\mathbb{E}\left[\frac{Zf_Z'(Z)}{f_Z(Z)}\right]\right)^2 = \left(zf_Z(z)\Big|_0^\infty - 1\right)^2 = 1,$$

where we have used the fact that $zf_Z(z)\big|_0^\infty = 0$, which is guaranteed by $\mathbb{E}\left[\frac{Zf_Z'(Z)}{f_Z(Z)}\right] < \infty$; see the explanation below. Thus in the sequel, we shall assume $n > 0$.

For $\alpha \in \mathbb{R}$ and $\alpha \neq 0$ (note that α can be negative), let $W = Z^{1/\alpha}$, or equivalently, $Z = W^\alpha$. For $z > 0$, we can calculate

$$f_Z(z) = f_W(z^{1/\alpha})\frac{1}{|\alpha|}z^{1/\alpha-1}.$$

Furthermore, we can calculate its derivative

$$f'_Z(z) = \frac{1}{|\alpha|}\left(\frac{1}{\alpha}(z^{1/\alpha-1})^2 f'_W(z^{1/\alpha}) + \left(\frac{1}{\alpha}-1\right) z^{1/\alpha-2} f_W(z^{1/\alpha})\right).$$

Therefore, by noticing that $Z^{1/\alpha} = W$, we have

$$\frac{f'_Z(Z)}{f_Z(Z)} = \frac{\frac{1}{\alpha} W Z^{-1} f'_W(W) + \left(\frac{1}{\alpha}-1\right) Z^{-1} f_W(W)}{f_W(W)} = \frac{W f'_W(W) + (1-\alpha) f_W(W)}{Z\alpha f_W(W)}.$$

It follows that

$$\begin{aligned}
\mathbb{E}\left[Z^n\left(\frac{Z f'_Z(Z)}{f_Z(Z)}\right)^2\right] &= \mathbb{E}\left[Z^n\left(\frac{W f'_W(W) + (1-\alpha) f_W(W)}{\alpha f_W(W)}\right)^2\right] \\
&= \mathbb{E}\left[Z^n\left(\frac{W f'_W(W)}{\alpha f_W(W)}\right)^2\right] + \mathbb{E}\left[\frac{2(1-\alpha) Z^n W f'_W(W)}{\alpha^2 f_W(W)}\right] + \left(\frac{1-\alpha}{\alpha}\right)^2 \mathbb{E}\left[Z^n\right].
\end{aligned}$$

Note that both $\mathbb{E}\left[Z^n\left(\frac{Z f'_Z(Z)}{f_Z(Z)}\right)^2\right]$ and $\mathbb{E}[Z^n]$ are finite, therefore we have that

$$\mathbb{E}\left[Z^n\left(\frac{W f'_W(W)}{\alpha f_W(W)}\right)^2\right] = \mathbb{E}\left[Z^n\left(\frac{Z f'_Z(Z)}{f_Z(Z)} - \left(\frac{1-\alpha}{\alpha}\right)\right)^2\right]$$

is finite. As a consequence, $\mathbb{E}\left[\frac{2(1-\alpha) Z^n W f'_W(W)}{\alpha^2 f_W(W)}\right]$ is also finite.

Let us choose $\alpha = -1/n$, which is actually the *optimal* choice of α as we shall later demonstrate. It follows that

$$\mathbb{E}\left[Z^n\left(\frac{Z f'_Z(Z)}{f_Z(Z)}\right)^2\right] = \mathbb{E}\left[Z^n\left(\frac{W f'_W(W)}{\alpha f_W(W)}\right)^2\right] + \mathbb{E}\left[\frac{2(1-\alpha) f'_W(W)}{\alpha^2 f_W(W)}\right] + (n+1)^2\,\mathbb{E}\left[Z^n\right].$$

Note that

$$\mathbb{E}\left[\frac{f'_W(W)}{f_W(W)}\right] = f_W(w)\Big|_0^\infty = f_Z(z) z^{n+1}\Big|_\infty^0,$$

which is a finite number as analyzed above; that is, $f_Z(z)z^{n+1}$ has a limit both as $z \to 0$ and as $z \to \infty$. The integrability condition $\mathbb{E}[Z^n] < \infty$ guarantees that this limit has to be zero in both cases. Therefore, $\mathbb{E}\left[\frac{f_W'(W)}{f_W(W)}\right] = 0$ and

$$\mathbb{E}\left[Z^n\left(\frac{Zf_Z'(Z)}{f_Z(Z)}\right)^2\right] = \mathbb{E}\left[Z^n\left(\frac{Wf_W'(W)}{\alpha f_W(W)}\right)^2\right] + (n+1)^2\,\mathbb{E}\left[Z^n\right].$$

Moreover, the random variable $M = \frac{Wf_W'(W)}{f_W(W)}$ is not almost surely zero, since $\mathbb{E}[M] = -1$, assuming that it is integrable. As a consequence, we have

$$\mathbb{E}\left[Z^n\left(\frac{Wf_W'(W)}{\alpha f_W(W)}\right)^2\right] > 0,$$

which implies

$$\mathbb{E}\left[Z^n\left(\frac{Zf_Z'(Z)}{f_Z(Z)}\right)^2\right] > (n+1)^2\mathbb{E}\left[Z^n\right],$$

which is the inequality (1).

Next we demonstrate that this inequality obtained is actually sharp, i.e., the constant $(n+1)^2$ can not be further improved upon. Note that the ratio $(n+1)^2$ is achieved asymptotically by choosing $f_Z(z) = c(z+d)^{-\beta}$ for $z > 0$, where $c, d > 0$ are constants and $\beta > n+1$. In this case,

$$\mathbb{E}\left[Z^n\left(\frac{Zf_Z'(Z)}{f_Z(Z)}\right)^2\right] = \mathbb{E}\left[Z^n\left(\frac{Z\beta(Z+d)^{-\beta-1}}{(Z+d)^{-\beta}}\right)^2\right] = \beta^2\mathbb{E}\left[\frac{Z^{n+2}}{(Z+d)^2}\right] \leqslant \beta^2\mathbb{E}\left[Z^n\right].$$

By letting $\beta \downarrow n+1$ and using (1), one obtains

$$\frac{\mathbb{E}\left[Z^n\left(\frac{Zf_Z'(Z)}{f_Z(Z)}\right)^2\right]}{\mathbb{E}[Z^n]} \to (n+1)^2.$$

Thus $(n+1)^2$ is the sharp lower bound for the ratio between $\mathbb{E}\left[Z^n\left(\frac{Zf_Z'(Z)}{f_Z(Z)}\right)^2\right]$ and $\mathbb{E}[Z^n]$. Note that $\beta = n+1$ is the critical point for $\mathbb{E}[Z^n] = \infty$.

Below we explain why the choice in the proof above, e.g., $\alpha = -1/n$, is optimal, and how we have arrived at this value. Note that, via integration by parts, there is

$$\mathbb{E}\left[\frac{W^{n\alpha+1}f_W'(W)}{f_W(W)}\right] = w^{n\alpha+1}f_W(w)\Big|_0^\infty - \int_0^\infty (n\alpha+1)w^{n\alpha}f_W(w)\mathrm{d}w = -(n\alpha+1)\mathbb{E}\left[W^{n\alpha}\right],$$

where we rely on the fact that $w^{n\alpha+1}f_W(w) \to 0$ as $w \to \infty$ and as $w \to 0$, which can be shown using a similar argument for $z^{n+1}f_Z(z) \to 0$ as in the proof of (1) above. It follows that

$$\begin{aligned}\mathbb{E}\left[\frac{2\,(1-\alpha)\,Z^n W f_W'(W)}{\alpha^2 f_W(W)}\right] &= \frac{2\,(1-\alpha)}{\alpha^2}\mathbb{E}\left[\frac{W^{n\alpha+1}f_W'(W)}{f_W(W)}\right] \\ &= \frac{2\,(\alpha-1)}{\alpha^2}(n\alpha+1)\mathbb{E}\left[W^{n\alpha}\right] = \frac{\alpha-1}{\alpha^2}(2n\alpha+2)\mathbb{E}\left[Z^n\right].\end{aligned}$$

Therefore there is

$$\begin{aligned}\mathbb{E}\left[Z^n\left(Z\frac{f_Z'(Z)}{f_Z(Z)}\right)^2\right] &= \mathbb{E}\left[Z^n\left(\frac{Wf_W'(W)}{\alpha f_W(W)}\right)^2\right] + \mathbb{E}\left[Z^n\right]\left(\left(\frac{1-\alpha}{\alpha}\right)^2 + \frac{\alpha-1}{\alpha^2}(2n\alpha+2)\right) \\ &= \mathbb{E}\left[Z^n\left(\frac{Wf_W'(W)}{\alpha f_W(W)}\right)^2\right] + \frac{(\alpha-1)((2n+1)\alpha+1)}{\alpha^2}\mathbb{E}\left[Z^n\right].\end{aligned}$$

Note that α can be arbitrarily chosen. If we write $\beta = 1/\alpha$, then we have

$$\frac{(\alpha-1)((2n+1)\alpha+1)}{\alpha^2} = (1-\beta)(2n+1+\beta) = 2n+1-2n\beta-\beta^2.$$

The function $\beta \mapsto 2n+1-2n\beta-\beta^2$ has a maximum value at $\beta = -n$, or equivalently, when $\alpha = -1/n$, which gives a value of $(n+1)^2$. This establishes the optimality of $\alpha = -1/n$. This completes the proof. □

Proof of Proposition 3. We define $h(\theta) = 1/g(\theta)$, and divide the discussion into the following two sub-cases.

- If $h(\theta) > 0$, then one can calculate $f_Y(y;\theta) = h(\theta)f_Z(h(\theta)y)$ for $y \in \mathbb{R}$. Therefore,

$$\frac{\partial}{\partial\theta}f_Y(y;\theta) = h(\theta)h'(\theta)yf_Z'(h(\theta)y) + h'(\theta)f_Z(h(\theta)y).$$

- If $h(\theta) < 0$, then $f_Y(y;\theta) = -h(\theta) f_Z(h(\theta)y)$ for $y \in \mathbb{R}$, and

$$\frac{\partial}{\partial\theta} f_Y(y;\theta) = -h(\theta)h'(\theta) y f_Z'(h(\theta)y) - h'(\theta) f_Z(h(\theta)y).$$

In both cases, the LR estimator is

$$\begin{aligned} Y(\theta)\frac{\frac{\partial}{\partial\theta} f_Y(Y(\theta);\theta)}{f_Y(Y(\theta);\theta)} &= Y(\theta)\frac{h(\theta)h'(\theta)Y(\theta)f_Z'(h(\theta)Y(\theta)) + h'(\theta)f_Z(h(\theta)Y(\theta))}{h(\theta)f_Z(h(\theta)Y(\theta))} \\ &= Y(\theta)\frac{h'(\theta)}{h(\theta)}\frac{Zf_Z'(Z) + f_Z(Z)}{f_Z(Z)} \\ &= \frac{h'(\theta)}{h^2(\theta)}\frac{Z^2 f_Z'(Z) + Zf_Z(Z)}{f_Z(Z)} = -g'(\theta)\left(\frac{Z^2 f_Z'(Z)}{f_Z(Z)} + Z\right), \end{aligned} \tag{10}$$

with the second moment given by

$$\begin{aligned} v_2 &= (g'(\theta))^2 \mathbb{E}\left[Z^4\left(\frac{f_Z'(Z)}{f_Z(Z)}\right)^2 + 2Z^3\frac{f_Z'(Z)}{f_Z(Z)} + Z^2\right] \\ &= (g'(\theta))^2\left(\mathbb{E}\left[Z^4\left(\frac{f_Z'(Z)}{f_Z(Z)}\right)^2\right] - 5\mathbb{E}[Z^2]\right). \end{aligned}$$

The IPA estimator has the second moment

$$v_1 = (g'(\theta))^2 \mathbb{E}[Z^2].$$

Observe the following equivalence:

$$v_1 \leqslant v_2 \quad \Leftrightarrow \quad 6\mathbb{E}[Z^2] \leqslant \mathbb{E}\left[Z^4\left(\frac{f_Z'(Z)}{f_Z(Z)}\right)^2\right], \tag{11}$$

which is independent of the choice of g. Moreover, because $g'(\theta) \neq 0$, we have

$$\frac{v_2}{v_1} = \frac{\mathbb{E}\left[Z^4\left(\frac{f_Z'(Z)}{f_Z(Z)}\right)^2\right]}{\mathbb{E}[Z^2]} - 5. \tag{12}$$

Using (11), (12), and Lemma 1 for $n = 2$, one can deduce that $v_2 > 4v_1$, and consequently, $\mathrm{Var}(L_Y) > 4\mathrm{Var}(I_Y)$. This completes the proof. $\square$

Proof of Proposition 4. Fix θ. If $p = 0$, then there is nothing to show. Below we assume $p > 0$. First, we note that

$$I_Y = g'(\theta) Z \mathbb{1}_{\{X(\theta) > K\}},$$

and

$$L_Y = (X(\theta) - K) \mathbb{1}_{\{X(\theta) > K\}} \frac{\frac{\partial}{\partial \theta} f_X(X(\theta); \theta)}{f_X(X(\theta); \theta)}.$$

Using (10) in the proof of Proposition 3 with $Y(\theta)$ replaced by $X(\theta)$, we get

$$L_Y = (X(\theta) - K) \mathbb{1}_{\{X(\theta) > K\}} \frac{-g'(\theta)}{g(\theta)} \left(\frac{Z f_Z'(Z)}{f_Z(Z)} + 1 \right).$$

Write $m := K/g(\theta)$. We have

$$\mathbb{E}[I_Y^2] = (g'(\theta))^2 \mathbb{E}[Z^2 \mathbb{1}_{\{Z > m\}}], \tag{13}$$

and

$$\mathbb{E}[L_Y^2] = (g'(\theta))^2 \mathbb{E}\left[(Z - m)^2 \left(\frac{Z f_Z'(Z)}{f_Z(Z)} + 1 \right)^2 \mathbb{1}_{\{Z > m\}} \right]. \tag{14}$$

Next, let W_1 be a random variable distributed as $(Z - m | Z > m)$. Note that W_1 is positive and it has a density and a finite second moment. Noting that $f_{W_1}'(W_1)/f_{W_1}(W_1) = f_Z'(Z)/f_Z(Z)$ conditional on $Z > m$, we can rewrite (13) and (14) as

$$a := \frac{\mathbb{E}[I_Y^2]}{p(g'(\theta))^2} = \mathbb{E}\left[(W_1 + m)^2 \right] = \mathbb{E}[W_1{}^2] + 2m\mathbb{E}[W_1] + m^2,$$

and

$$b := \frac{\mathbb{E}[L_Y^2]}{p(g'(\theta))^2} = \mathbb{E}\left[W_1{}^2 \left(\frac{(W_1 + m) f_{W_1}'(W_1)}{f_{W_1}(W_1)} + 1 \right)^2 \right].$$

We can further derive

$$b = \mathbb{E}\left[(W_1^4 + W_1^2 m^2 + 2W_1^3 m)\left(\frac{f'_{W_1}(W_1)}{f_{W_1}(W_1)}\right)^2 + 2W_1^3 \frac{f'_{W_1}(W_1)}{f_{W_1}(W_1)} + 2W_1^2 m \frac{f'_{W_1}(W_1)}{f_{W_1}(W_1)} + W_1^2\right]$$
$$= \mathbb{E}\left[(W_1^4 + W_1^2 m^2 + 2W_1^3 m)\left(\frac{f'_{W_1}(W_1)}{f_{W_1}(W_1)}\right)^2\right] - 5\mathbb{E}[W_1^2] - 4m\mathbb{E}[W_1].$$

Applying Lemma 1 to the random variable W_1, we get

$$\mathbb{E}\left[(W_1^4 + W_1^2 m^2 + 2W_1^3 m)\left(\frac{f'_{W_1}(W_1)}{f_{W_1}(W_1)}\right)^2\right] > 9\mathbb{E}[W_1^2] + m^2 + 8m\mathbb{E}[W_1].$$

Therefore,

$$b > 9\mathbb{E}[W_1^2] + m^2 + 8m\mathbb{E}[W_1] - 5\mathbb{E}[W_1^2] - 4m\mathbb{E}[W_1] = 4\mathbb{E}[W_1^2] + 4m\mathbb{E}[W_1] + m^2 > a.$$

Then we have $v_2 = \mathbb{E}[L_Y^2] > \mathbb{E}[I_Y^2] = v_1$, and $\mathrm{Var}(L_Y) > \mathrm{Var}(I_Y)$ follows.

Next, we show the last statement. Note that $\mathbb{E}[W_1] = \delta m$, and

$$\mathbb{E}[L_Y] = \mathbb{E}[I_Y] = pg'(\theta)(\mathbb{E}[W_1] + m).$$

We have

$$\begin{aligned}\frac{\mathrm{Var}(I_Y)}{p(g'(\theta))^2} &= \mathbb{E}[W_1^2] + 2m\mathbb{E}[W_1] + m^2 - p((\mathbb{E}[W_1])^2 + 2m\mathbb{E}[W_1] + m^2)\\ &= \mathbb{E}[W_1^2] + m^2(2\delta + 1 - p(1+\delta)^2)\\ &= (\mathbb{E}[W_1^2] - \delta^2 m^2) + m^2((1-p)(1+\delta)^2),\end{aligned}$$

and

$$\begin{aligned}\frac{\mathrm{Var}(L_Y)}{p(g'(\theta))^2} &= 4\mathbb{E}[W_1^2] + 4m\mathbb{E}[W_1] + m^2 - p((\mathbb{E}[W_1])^2 + 2m\mathbb{E}[W_1] + m^2)\\ &= 4(\mathbb{E}[W_1^2] - \delta^2 m^2) + m^2(3\delta^2 + 2\delta + (1-p)(1+\delta)^2).\end{aligned}$$

Noting that $\mathbb{E}[W_1^2] - \delta^2 m^2 = \mathrm{Var}(W_1) \geqslant 0$, the desired inequality (2) follows. This completes the proof. $\square$