Advances in Derivative Estimation:

Volk-Makarewicz, W.M.

2014

Chapter 2

An Investigation into the Derivative Estimation of Ranked-Data Statistics

2.1 Introduction

Let $\mathbf{Z} = (Z_i : 1 \le i \le n)$ be a sequence of independent and identically distributed (i.i.d.) continuous copies of a real-valued random variable Z. The order statistic, $Z_{l:n}$, $l = 1, \ldots, n$, is the l^{th}, smallest random variable value from an ordered sequence of $\mathbf{Z}$. With probability one there is a strict order relation amongst[1] the random variables where

$$Z_{1:n} < Z_{2:n} < \ldots < Z_{l:n} < \ldots < Z_{n:n}.$$

We assume that the distribution of Z depends on some controllable distributional parameter θ, where we assume for the sake of simplicity that $\theta \in \Theta = (a, b) \subset \mathbb{R}$, $a < b$. If we want to stress the dependence of Z on θ, we will write $Z(\theta)$.

Given $l \in \{1, \ldots, n\}$ as one of the ranks, we consider the following derivative estimation problem:

$$\frac{d}{d\theta}\mathbb{E}_\theta[Z_{l:n}], \tag{2.1}$$

referred to as the *order statistic problem* (OSP).

We denote the cumulative distribution function (c.d.f.) of Z by F and additionally the probability density function (p.d.f.) function by f. If we want to emphasize the dependence on θ, we write the distribution function by F_θ and respectively f_θ for the density function. We denote the quantile of Z, respectively of F, at a level $\alpha \in (0, 1)$ by q_α, see Equation (1.39). As Z is continuous, from Equation (1.39), we can write the quantile as the inverse of the distribution

[1] For a discrete distribution however, there is ambiguity in the relation due to the (countable) support.

function, since F is non-decreasing. For convenience, we have repeated the expression:

$$q_\alpha = \sup\{y : F(y) = \alpha\} = F^{-1}(\alpha). \tag{2.2}$$

The quantile is also denoted by $q_\alpha(\theta)$ if we want to emphasise the parameter dependence. A standard statistical estimator for the quantile is the order statistic $Z_{\lceil \alpha n \rceil : n}$, where $\lceil x \rceil$ denotes the smallest integer greater than x.

The relationship between quantiles and order statistics was first determined by Bahadur in [4] for i.i.d. random variables by showing almost sure convergence of $Z_{\lceil \alpha n \rceil : n}$ towards q_α. The requirement of independence in the result has been weakened progressively, for instance, in [92] to m-dependent random variables[2], and in [93] to ϕ-mixing random variables in which the random variables are asymptotically independent. The most recent significant generalization was introduced in [105], where the same convergence result was established for non-linear time series including stochastic recursive sequences satisfying a geometric matching condition. That paper also provided further refinements in the Bahadur-type result for short and long-run dependent linear sequences.

We consider the following derivative estimation problem:

$$\frac{\partial}{\partial \theta} \lim_{n \to \infty} Z_{\lceil \alpha n \rceil : n} = \frac{\partial}{\partial \theta} q_\alpha(\theta) \quad \text{a.s.}, \tag{2.3}$$

which will be referred to as the *quantile problem* (QP). Note that the QP is the limiting version of the OSP in (2.1) and the OSP will serve as a first step towards the analysis of QP. However, the OSP is a problem in its own right as in many applications the basic characteristic is based on a fixed, finite number of observations as we will explain later in the text with examples from finance and quality control.

Firstly, we will derive unbiased estimators for (2.1). For the expression in (2.3) we develop estimators that are strongly consistent, asymptotically unbiased, and that satisfy a central limit theorem, allowing us to construct confidence intervals for $\partial_\theta q_\alpha(\theta)$. To establish these estimators we will work with the derivative estimation methods of *Measure-Valued Differentiation* (MVD), [83], *Score Function Method* (SF), [89], and *Infinitesimal Perturbation Analysis* (IPA) [47] [36]. An overview for the underlying approach and the mechanics of the individual method for these derivative estimation methods has been provided in the Introduction chapter, Section 1.2. Numerical examples will illustrate the performance of the resulting estimators.

[2]A sequence (X_n) of random variables are called m-dependent if X_n and X_{n+k} are independent for any n provided that $k \geq m$.

This chapter provides a unified analysis for all three gradient estimation approaches. The sensitivity analysis of the QP problem is based on what appears to be a generic argument for perturbation analysis of quantiles. Working out the specific details for IPA, SF and MVD will then educe results for the corresponding estimators. In the course of this eduction, we will establish almost sure convergence of the IPA estimator, which improves upon the convergence result in [57], where only convergence in probability is proven.

This chapter is organized as follows. In Section 2.2 we present the ranked-data derivative estimators. The order statistic problem will be discussed in Section 2.3. In Section 2.4 we will discuss the quantile problem. We conclude this chapter with guidelines for derivative estimation for order statistic related performance measures in Section 2.5.

2.2 Derivative Estimation of Ranked-Data

In Section 2.2.1 we introduce the basic ranked-data gradient estimators for Infinitesimal Perturbation Analysis, Score Function, and Measure Valued Differentiation. In Section 2.2.2, we confirm that for distributional approaches, differentiation can be conducted with respect to aggregated random variables.

In this section we denote $X \in \mathbb{R}$ as an input random variable to a mapping h. If we want to stress the dependence due to parameter we denote the random variable by $X(\theta)$.

2.2.1 The Basic Estimators

The starting point of our analysis is a sample of m i.i.d. random variables $\mathbf{Z} = (Z_i : 1 \leq i \leq m)$. For $\alpha \in (0,1)$ and $m \in \mathbb{N}$, we denote the projection of the ordered sample of $\mathbf{Z} \in \mathbb{R}^m$ on element $\lceil \alpha m \rceil$ by $\mathrm{ord}_{\alpha,m}$; i.e.,

$$Z_{\lceil \alpha m \rceil : m} = \mathrm{ord}_{\alpha,m}(Z_1, \ldots, Z_m). \tag{2.4}$$

In the following, let (c_θ, Z^+, Z^-) be a representation for the weak derivative of Z, and let $((Z_j^+, Z_j^-) : 1 \leq j \leq m)$ be an i.i.d. sequence. We firstly introduce the two MVD derivative estimators as follows. The *standard* MVD derivative estimator is attained from successively replacing the elements Z_j in Z by Z_j^+ and Z_j^- respectively due to the product rule representation of the weak derivative, [52]

$$\begin{aligned} D_1^{\mathrm{MVD}}(m) = c_\theta \sum_{j=1}^{m} \Big(&\mathrm{ord}_{\alpha,m}\Big(Z_1, \ldots, Z_{j-1}, Z_j^+, Z_{j+1}, \ldots, Z_m\Big) \\ &- \mathrm{ord}_{\alpha,m}\Big(Z_1, \ldots, Z_{j-1}, Z_j^-, Z_{j+1}, \ldots, Z_m\Big)\Big). \end{aligned} \tag{2.5}$$

The *symmetric* MVD derivative estimator is attained from respectively substituting (w.l.o.g.) the last entry Z_m of Z by Z_m^+ and Z_m^-, and re-scaling the difference by m:

$$D_2^{\mathrm{MVD}}(m) = mc_\theta \left(\mathrm{ord}_{\alpha,m}\left(Z_1,\ldots,Z_{m-1},Z_m^+\right) - \mathrm{ord}_{\alpha,m}\left(Z_1,\ldots,Z_{m-1},Z_m^-\right)\right).$$

Secondly, we introduce the *standard* score function estimator by

$$D_1^{\mathrm{SF}}(m) = \mathrm{ord}_{\alpha,m}(Z_1,\ldots,Z_m)\sum_{j=1}^{m}\mathrm{SF}_\theta(Z_j),$$

and the *symmetric* score function estimator, which is a re-scaled version of the standard estimator applied only to entry Z_m, by

$$D_2^{\mathrm{SF}}(m) = m\,\mathrm{ord}_{\alpha,m}(Z_1,\ldots,Z_m)\,\mathrm{SF}_\theta(Z_m).$$

Finally, we introduce the IPA derivative estimator

$$D^{\mathrm{IPA}}(m) = \frac{\partial}{\partial\theta}Z_k(\theta), \quad \text{where } k = (\lceil\alpha m\rceil : m).$$

The IPA estimator suppresses the dependence of the index $(\lceil\alpha m\rceil : m)$ on θ. Note that in many models this is a reasonable assumption: If Z is monotone as a mapping of θ with probability one, then for a given sample-path the index $(\lceil\alpha m\rceil : m)$ is independent of θ, as it is constant, only depending on the underlying randomness. As we will explain in the proof of Theorem 2.1, this holds under quite general conditions that the index $(\lceil\alpha m\rceil : m)$ is independent of θ for a given realization.

2.2.2 A Lemma

As differentiation w.r.t. the parameter θ is conducted w.r.t. only one of our sample of m random variables, we need to extend on Section 1.2.2 and to verify that differentiation of $X(\theta)$ carries over to aggregated random variables. This is derived in the next Lemma.

Lemma 2.1. *Consider random variables $X(\theta) \in \mathbb{R}$ and $Y \in \mathbb{R}^n$, and let $Z(\theta) = h(Y, X(\theta))$ for a measurable mapping h from $\mathbb{R}^{n+1}$ to $\mathbb{R}$. Suppose that $X(\theta)$ is $\mathfrak{D}$-differentiable for $\mathfrak{D} = \mathfrak{B}_v, \mathfrak{C}_v$, and that Y is not parameterized by θ. In addition, let $\hat{\mathfrak{D}}$ be the induced function space such that for any function $g \in \hat{\mathfrak{D}}$, the conditional expectation $\mathbb{E}[g(h(Y, X(\theta)))|X(\theta) = (\cdot)] \in \mathfrak{D}$. Then the following two statements hold:*

(i) *Suppose that* $X(\theta)$ *has a weak derivative* $(c_\theta, X^+(\theta), X^-(\theta))$*, then* $Z(\theta)$ *is* $\hat{\mathfrak{D}}$*-differentiable with weak derivative* $(c_\theta, Z^+(\theta), Z^-(\theta))$ *where*

$$Z^+(\theta) = h(Y, X^+(\theta)) \quad \text{and} \quad Z^-(\theta) = h(Y, X^-(\theta)).$$

(ii) *If the score function of* $X(\theta)$ *is given by Equation (1.19), then for the random variable* $Z(\theta)$*,* $\mathrm{SF}_\theta(Z(\theta)) = \mathrm{SF}_\theta(X(\theta))$.

Proof: Part (i). As $\mathbb{E}[g(h(Y, X(\theta)))|X(\theta) = (\cdot)] \in \mathfrak{D}$, given $g \in \hat{\mathfrak{D}}$, the results follows from the $\mathfrak{D}$-differentiability of $X(\theta)$

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{E}[g(Z)] &= \frac{d}{d\theta}\int_{\mathbb{R}} \mathbb{E}[g(h(Y, X(\theta)))|X(\theta) = x]\, f_\theta(x)\, dx \\
&= c_\theta \left(\int_{\mathbb{R}} \mathbb{E}[g(h(Y, X(\theta)))|X(\theta) = x]\, f_\theta^+(x)\, dx \right. \\
&\qquad \left. - \int_{\mathbb{R}} \mathbb{E}[g(h(Y, X(\theta)))|X(\theta) = x]\, f_\theta^-(x)\, dx \right) \\
&= c_\theta \big(\mathbb{E}\big[g(Z^+(\theta))\big] - \mathbb{E}\big[g(Z^-(\theta))\big] \big).
\end{aligned}$$

Part (ii). Alternatively, given the above requirements, interchange of derivative of integral yields

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{E}[g(Z)] &= \int_{\mathbb{R}} \mathbb{E}[g(h(Y, X(\theta)))|X(\theta) = x]\, \frac{\partial}{\partial\theta} f_\theta(x)\, dx \\
&= \int_{\mathbb{R}} \mathbb{E}[g(h(Y, X(\theta)))|X_\theta = x]\, \mathrm{SF}_\theta(x)\, f_\theta(x)\, dx \\
&= \mathbb{E}\big[g(Z(\theta))\, \mathrm{SF}_\theta(X(\theta))\big],
\end{aligned}$$

which concludes both proofs. □

Example 2.1. *Let* X *be* $\mathfrak{B}_v$*-differentiable and let* $h(x) = 1\{[a,\infty)\}(x) \in \mathfrak{B}$ *be the indicator mapping of the interval* $[a,\infty)$*. Since* $v(x) \geq 1$*, for all* $x \in \mathbb{R}$*,* $\mathfrak{B} \subset \mathfrak{B}_v$ *and so* $h(X)$ *is also* $\mathfrak{B}_v$*-differentiable. Note that if* X *is a continuous random variable, then* $h(X)$ *becomes a discrete random variable.*

2.3 The Order Statistic Problem

In this section we present our results for the OSP. First, we will provide the technical analysis in Section 2.3.1. Then, in Section 2.3.2 we will present numerical examples.

2.3.1 Technical Analysis

For the OSP, our assumptions are based upon the assumption needed for the three derivative estimation approaches.

(A1) $\mathbf{Z} = (Z_i : 1 \le i \le m)$ is an i.i.d. collection of random variables with a continuous distribution function.

(A2 MVD) The random variable Z is $\mathfrak{D}$-differentiable with $\mathfrak{D} = \mathfrak{B}_v, \mathfrak{C}_v$, and a version of its weak derivative is given by (c_θ, Z^+, Z^-). For estimators D_l^{MVD}, for $l = 1,2$, we assume that each pair (Z_j^+, Z_j^-) is independent of the nominal random variables $\mathbf{Z} \backslash Z_j = (Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_m)$. However, no assumption on the dependence structure of (Z_j, Z_j^+, Z_j^-), $j = 1, \ldots, n$, is imposed.

(A2 SF) There exists an open neighbourhood Θ of θ such that

(1) $f_\theta(x)$ is differentiable w.r.t. θ on Θ for all x,

(2) for any $\theta' \in \Theta$, it holds that $f_\theta(x) = 0$ implies $f_{\theta'}(x) = 0$ for all x,

(3) it holds that

$$\int_{\mathbb{R}} (1 + |z|) \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta} f_\theta(z) \right| dz < \infty.$$

(A2 IPA) The random variable $Z(\theta)$ is almost surely differentiable w.r.t. θ and $Z(\theta)$ is Lipschitz continuous with integrable modulus of continuity; i.e., an integrable random variable $K(Z(\theta))$ exists such that

$$|Z(\theta + \Delta) - Z(\theta)| \le |K(Z(\theta))| \Delta$$

with probability one, and $\mathbb{E}[K(Z(\theta))] < \infty$.

Under the above conditions, the estimators presented in Section 2.2.1 are unbiased. The precise statement is given in the following theorem.

Theorem 2.1. *Suppose condition* **(A1)** *holds. If, in addition,*

(i) *condition* **(A2 MVD)** *is satisfied, then for any m*

$$\frac{d}{d\theta} \mathbb{E}_\theta \big[Z_{\lceil \alpha m \rceil : m} \big] = \mathbb{E}_\theta \big[D_l^{\text{MVD}}(m) \big], \quad l = 1,2,$$

(ii) *condition* **(A2 SF)** *is satisfied, then for any m*

$$\frac{d}{d\theta} \mathbb{E}_\theta \big[Z_{\lceil \alpha m \rceil : m} \big] = \mathbb{E}_\theta \big[D_l^{\text{SF}}(m) \big], \quad l = 1,2,$$

(iii) *condition* **(A2 IPA)** *is satisfied, then for any m*

$$\frac{d}{d\theta}\mathbb{E}\big[Z_{\lceil \alpha m\rceil:m}(\theta)\big] = \mathbb{E}\big[D^{\mathrm{IPA}}(m)\big].$$

Proof: Part (i). Let $\|f\|_v = \sup_x |f(x)|/v(x)$. Then $\mathfrak{D} = \mathfrak{B}_v, \mathfrak{C}_v$, equipped with $\|\cdot\|_v$, is a Banach space. Since $\mathrm{ord}_{\alpha,m}(z_1,\dots,z_{j-1},\cdot,z_{j+1},\dots,z_m)$ is a measurable and continuous mapping in each of its components, Assumption **(A2 MVD)** together with the product rule of $\mathfrak{D}$-differentiation, see Theorem 6.1 [52], implies

$$\frac{d}{d\theta}\mathbb{E}_\theta\big[Z_{\lceil \alpha m\rceil:m}\big] = c_\theta \sum_{j=1}^{m}\Big(\mathbb{E}_\theta\Big[\mathrm{ord}_{\alpha,m}\Big(Z_1,\dots,Z_{j-1},Z_j^+,Z_{j+1},\dots,Z_m\Big)\Big] - \mathbb{E}_\theta\Big[\mathrm{ord}_{\alpha,m}\Big(Z_1,\dots,Z_{j-1},Z_j^-,Z_{j+1},\dots,Z_m\Big)\Big]\Big).$$

Observing that $\mathrm{ord}_{\alpha,m}$ is invariant with respect to permutations of the arguments, implies

$$\frac{d}{d\theta}\mathbb{E}\big[Z_{\lceil \alpha m\rceil:m}\big] = m c_\theta \left(\mathbb{E}_\theta\big[\mathrm{ord}_{\alpha,m}\left(Z_1,\dots,Z_{m-1},Z_m^+\right)\big] - \mathbb{E}_\theta\big[\mathrm{ord}_{\alpha,m}\left(Z_1,\dots,Z_{m-1},Z_m^-\right)\big]\right),$$

which concludes the proof (i).

Part (ii). The form of the score function estimator follows from simple rules of analysis applied to the product of densities. It remains to show that the estimator is unbiased; i.e., that interchanging integration and differentiation is justified. For this to see, note that for $(z_1,\dots,z_m) \in \mathbb{R}^m$ the Mean-Value Theorem implies

$$\begin{aligned}
&\int_{\mathbb{R}^m} \frac{1}{|\Delta|}\mathrm{ord}_{\alpha,m}(z_1,\dots,z_m)\left|\prod_{j=1}^{m} f_{\theta+\Delta}(z_i) - \prod_{j=1}^{m} f_\theta(z_i)\right| dz_1\dots dz_m \\
&\le \int_{\mathbb{R}^m}\left(\sum_{k=1}^{m}|z_k|\right)\sum_{j=1}^{m}\prod_{i=1}^{j-1} f_\theta(z_i)\left(\sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta}f_\theta(z_j)\right|\right)\prod_{i=j+1}^{m} f_\theta(z_i)\; dz_1\dots dz_m \\
&\le m(m-1)\,\mathbb{E}[|Z|]\int_{\mathbb{R}}\sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta}f_\theta(z)\right| dz \;+\; m\int_{\mathbb{R}}(1+|z|)\sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta}f_\theta(z)\right| dz,
\end{aligned}$$

which is finite by **(A2 SF)**. Hence, the first statement in (ii) follows by dominated convergence. The proof of the second part of (ii) follows the same line as the proof of the second part of (i) and is therefore omitted.

Part (iii). Let $k(\theta)$ denote the index such that

$$Z_{k(\theta)}(\theta) = Z_{\lceil \alpha m\rceil:m}(\theta).$$

Let $Z_j(\theta) = F_\theta^{-1}(U_j)$, with $\mathbf{U} = (U_j : 1 \le j \le m)$ an i.i.d. sequence of uniform-$(0,1)$-distributed random variables. The fact that $F_\theta(x)$ is monotone non-decreasing as a mapping of x, implies that $F_\theta^{-1}(u)$ is monotone non-decreasing as a mapping of u, which implies that $k(\theta) = j$ is constant as a mapping of θ for given sequence $\mathbf{U}$. This, together with **(A2 IPA)**, implies that $Z_{\lceil \alpha m\rceil:m}(\theta)$ is almost surely differentiable.

Lipschitz continuity of $Z_j(\theta)$ implies that

$$\begin{aligned}&\mathrm{ord}_{\alpha,m}(\mathbf{Z}_1(\theta+\Delta),\ldots,Z_m(\theta+\Delta)) - \mathrm{ord}_{\alpha,m}(Z_1(\theta),\ldots,Z_m(\theta))|\\ &\quad\le \sum_{k=1}^m |Z_k(\theta+\Delta) - Z_k(\theta)|.\end{aligned}$$

By **(A2 IPA)** it holds that

$$\mathbb{E}\left[\sum_{k=1}^m |Z_k(\theta+\Delta) - Z_k(\theta)|\right] \le m\,\mathbb{E}[K(Z(\theta))]\,\Delta < \infty,$$

and the proof follows from the dominated convergence theorem. □

Remark 2.1. *The symmetric estimators apply differentiation only to $Z_m(\theta)$. The random variables $Z_1(\theta)$ to $Z_{m-1}(\theta)$ are unaffected by the differentiation operation. More precisely, suppose that the derivative is evaluated at θ_0, then*

$$\begin{aligned}\mathbb{E}_\theta\big[D_2^{\mathrm{MVD}}(m)\big] &= \frac{d}{d\theta}\bigg|_{\theta=\theta_0} \mathbb{E}\big[m\,\mathrm{ord}_{\alpha,m}(Z_1(\theta_0),\ldots,Z_{m-1}(\theta_0),Z(\theta))\big]\\ &= c_\theta\,\mathbb{E}\big[m\,\big(\mathrm{ord}_{\alpha,m}\big(Z_1(\theta_0),\ldots,Z_{m-1}(\theta_0),Z_m^+(\theta)\big)\\ &\qquad - \mathrm{ord}_{\alpha,m}\big(Z_1(\theta_0),\ldots,Z_{m-1}(\theta_0),Z_m^-(\theta)\big)\big)\big].\end{aligned}$$

In the same vein, the symmetric SF estimator reads

$$\begin{aligned}\mathbb{E}_\theta\big[D_2^{\mathrm{SF}}(m)\big] &= \frac{d}{d\theta}\bigg|_{\theta=\theta_0} \mathbb{E}\big[m\,\mathrm{ord}_{\alpha,m}(Z_1(\theta_0),\ldots,Z_{m-1}(\theta_0),Z_m(\theta))\big]\\ &= \mathbb{E}\big[m\,\mathrm{ord}_{\alpha,m}(Z_1(\theta_0),\ldots,Z_{m-1}(\theta_0),Z_m(\theta))\,\mathrm{SF}_\theta(Z_m(\theta))\big].\end{aligned}$$

2.3.2 Numerical Examples

In this section we compare the estimators for the OS problem for two examples. For each derivative estimator we compute the mean, root mean square (RMS) error, standard deviation, and the work-normalized variance. The RMS error statistic is the l^2 measure of the distance between the sensitivity estimate and the actual value. This is similar to the standard deviation which is a measure for the distance between the sensitivity estimate and its mean value. The WNV is computed as the product of the mean computation time and the standard deviation [43]. In addition, we will also consider the coverage probability, which is determined as the proportion of experiments for which the actual value lies within the 95% confidence interval attained form the derivative estimator.

2.3.2.1 Quality Control of Steel Production

In the production of steel, one of the impurities that arises during the manufacturing process is sulfur. Sulfur is present in coal that extracts iron from iron ore. The problem with sulfur is that it weakens steel, amongst other things rendering the steel more brittle. The sulfur can be easily removed during manufacturing, and so the amount of sulfur remaining in steel, above a certain percentage, becomes a manufacturing issue.

Suppose a steel producer has a contract to sell m metric tons of steel rods to a construction firm for concrete reinforcing. This contract specifies that at least α percent of the steel bars has a sulfur content less than 0.04%. For testing the amount of sulfur in the steel bars, one allotment will be considered to be one metric ton, and $\mathbf{Z} = (Z_i, 1 \le i \le m)$ is the amount of sulfur (in %) within each allotment. Put differently, the production is considered to be of satisfactory quality if $\mathbb{E}[Z_{\lceil \alpha m \rceil : m}] \le 0.04$. In order to achieve this goal the company may adjust the production process (which in itself is complex) via some control parameter, say, θ. Thus,

$$\frac{d}{d\theta}\mathbb{E}\big[Z_{\lceil \alpha m \rceil : m}\big]$$

denotes the sensitivity of the overall batch quality with respect to θ.

Suppose that $Z_i = a + Y_i(\sigma)$, where $a \in \mathbb{R}$ is constant and $Y_i(\sigma)$ is a standard normal random variable with mean zero and standard deviation $\sigma > 0$. In the following, we let $\theta = \sigma$ and analyze

$$\frac{d}{d\sigma}\mathbb{E}_\sigma\big[Z_{\lceil \alpha m \rceil : m}\big] = \frac{d}{d\sigma}\mathbb{E}\big[Y_{\lceil \alpha m \rceil : m}(\sigma)\big],$$

which gives insight into the impact of the variability of the amount of sulfur within the steel making process on the quality of the steel delivered to the construction firm. For our sensitivity analysis we will call upon Section 1.2.4.2.

In the example, we will assume that we are delivering m metric tons of steel. The sample size and the number of estimates evaluated is $(m, k) = (10, 10)$. We choose the mean level of sulfur content (in percent) $a = 0.025$ and $\sigma = 0.004$. With the supposed maximal acceptable percentage of content of sulfur within the steel of 0.04%, we consider the probability levels $\alpha = 0.50$, and 0.95. The numerical results are presented in Table 2.1. The simulated data are configured in a $m \times k$ array, where m is the number of samples in each estimate of the order statistic derivative, and k is the number of such estimates within the sample average. The RMS error, standard deviation (Std. Dev.), work-normalized variance (WNV), and the coverage probability (Cov. Prob.) are determined over 500 estimates.

$m = 10$, $k = 10$, $\alpha = 0.50$. Actual Value = -0.1227

	IPA	SF1	SF2	MVD1	MVD2
Mean	-0.1211	0.1372	0.3966	-0.1343	-0.1147
RMS Error	0.9952	7.194×10^{1}	2.094×10^{2}	2.8222	6.3466
Std. Dev.	0.1222	8.8299	2.571×10^{1}	0.3464	0.7793
WNV	2.562×10^{-7}	3.054×10^{-3}	4.910×10^{-2}	6.532×10^{-5}	1.512×10^{-4}
Cov. Prob.	0.9240	0.9200	0.8600	0.9220	0.9480

$m = 10$, $k = 10$, $\alpha = 0.95$. Actual Value = 1.5388

	IPA	SF1	SF2	MVD1	MVD2
Mean	1.5359	1.8514	2.6662	1.5454	1.5328
RMS Error	0.1224	7.6047	2.335×10^{1}	0.4553	0.9950
Std. Dev.	0.1884	1.171×10^{1}	3.594×10^{1}	0.7014	1.5325
WNV	3.828×10^{-7}	4.992×10^{-3}	9.241×10^{-2}	2.038×10^{-4}	5.817×10^{-4}
Cov. Prob.	0.9180	0.9060	0.8280	0.9080	0.8280

Table 2.1: The summary statistics for the five derivative estimators in comparison with the actual value for the Quality Control of Steel Production example. The upper part shows results for $\alpha = 0.50$ and the lower part for $\alpha = 0.95$.

Figure 2.1 provides the RMS errors for increasing values of k, where the number of samples for each estimate is $m = 20$. In Figure 2.1, the plot confirms a decrease of the RMS error of order $k^{-1/2}$. This was conducted by least squares regression of the logarithm of the data. For the Score Function method the values for the RMS error are much larger than for IPA and MVD, which is the reason why we omitted the SF estimators in this figure.

2.3.2.2 Performance of a Basket of Stocks

After a pre-screening process, a trader has a basket of m stocks for which she/he wants to undertake an automated stock trading strategy. It is believed, due to a lack of further information, that the stocks' behaviour is statistically indistin-

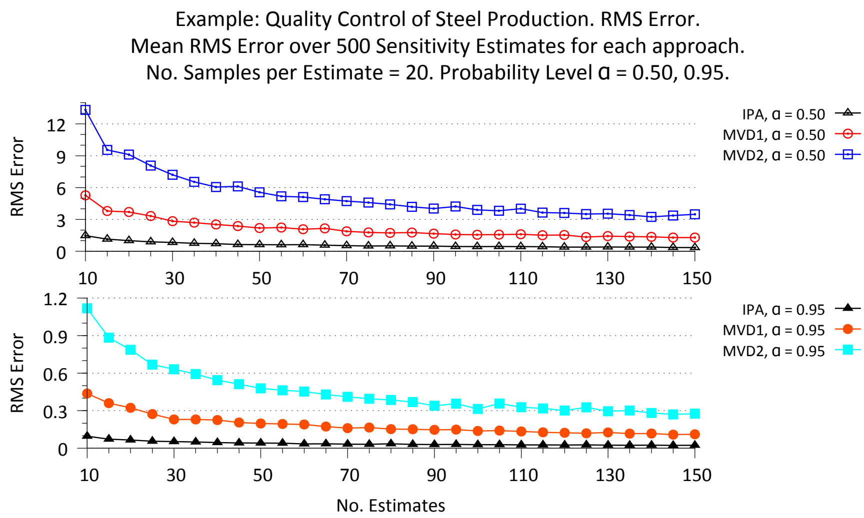

Figure 2.1: The RMS error for IPA and both MVD order statistic derivative estimates in terms of the number of estimates k for the Quality Control of Steel Production example.

guishable. For instance, the stocks may belong to a particular sector such as energy or technology or have a market capitalization within a particular range, e.g., US 1-4 billion. During the trading strategy, the trader monitors the performance of each of the stocks in the basket and is interested in the top (1-α) % returns at some future time t, i.e., the trader is interested to know

$$\mathbb{E}\left[Z_{\lceil \alpha m\rceil:m}(t)\right],$$

where the order statistic is attained from the basket $(Z_i(t) : 1 \le i \le m)$.

Assume that the return of the individual stocks $Z_i(\tau) = S_i(\tau)/S_i(0)$, $0 \le \tau \le t$, where $S_i(\tau)$ is the price of the stock at time τ, follow a Black-Scholes-Merton (BSM) model [81]. The BSM model for $Z_i(\tau)$ is composed of one stock having value $Z_i(\tau)$ at time $\tau \ge 0$ that pays no dividends and a bond which can alternatively be invested in having value $e^{r\tau}$ at time $\tau \ge 0$, where $r \ge 0$. The interest rate is risk free due to default and, for at least short time periods, constant. The evolution of the return of each stock under the BSM model has the form

$$Z_i(\tau) = \exp\left(\left(\beta + \eta - \frac{\sigma^2}{2}\right)\tau + \sigma W_i(\tau)\right).$$

The term β is the historical average market return within the company sector or capitalization range, and η is the additional historical mean out-performance, above the average market return, observed within the screened stocks. The common implied volatility of each stock is given by σ and $W_i(\tau)$, for $1 \le i \le m$, is a Wiener process.

The trader is interested in the influence of the time $\theta = t$ of the top $(1-\alpha)$ % of the returns for the chosen basket of stocks; i.e,

$$\frac{d}{dt}\mathbb{E}\left[Z_{\lceil \alpha m \rceil : m}(t)\right].$$

Note that the sample path of a Wiener process is continuous but almost nowhere differentiable, which implies that $Z_i(t)$ fails to be differentiable w.r.t. t in the pathwise sense, and thus no IPA estimator exists for the desired sensitivity; i.e., condition **(A2 IPA)** does not hold.

The marginal distribution of the Wiener process at time t is given by a normal random variable with mean zero and standard deviation $\sqrt{t}$, this implies that $Z_i(t)$ follows a normal distribution with mean

$$a(t) = \left(\beta + \eta - \frac{\sigma^2}{2}\right)t \tag{2.6}$$

and standard deviation

$$b(t) = \sigma\sqrt{t}. \tag{2.7}$$

Following (1.37), we attain

$$\frac{\partial}{\partial t} f_{a(t),b(t)}(x) = \frac{a(t)}{t}\frac{\partial}{\partial \mu} f_{a(t),b(t)}(x) + \frac{b(t)}{2t}\frac{\partial}{\partial \sigma} f_{a(t),b(t)}(x),$$

for $x \in \mathbb{R}$. From Section 1.2.4.2, the above analytical expression can be represented through random variables in the following way:

$$(c_1 + c_2)\left(\left(p_1 r^+_{a(t),b(t)}(x) + p_2 m_{a(t),b(t)}(x)\right) - \left(p_1 r^-_{a(t),b(t)}(x) + p_2 f_{a(t),b(t)}(x)\right)\right),$$

where

$$c_1 = \frac{a(t)}{\sqrt{2\pi}\, t\, b(t)}, \quad c_2 = \frac{1}{2t}$$

and

$$p_1 = \frac{c_1}{c_1 + c_2} \quad \text{and} \quad p_2 = \frac{c_2}{c_1 + c_2}.$$

A sample of $Z_i^+(t)$ can be attained from sampling with probability p_1 a shifted Rayleigh random variable and with probability p_2 a double-Maxwell random variable. In the same vein, a sample of $Z_i^-(t)$ can be attained from sampling with probability p_1 a shifted Rayleigh random variable and with probability p_2 a standard normal random variable, for details on how to sample the random variables we refer to Section 1.2.4.2.

Following from the same line of argument, the score function is given by

$$\begin{aligned}\mathrm{SF}_t(x) &= \frac{a(t)}{t}\,\frac{\frac{\partial}{\partial\mu} f_{a(t),b(t)}(x)}{f_{a(t),b(t)}(x)} + \frac{b(t)}{2t}\,\frac{\frac{\partial}{\partial\sigma} f_{a(t),b(t)}(x)}{f_{a(t),b(t)}(x)}\\ &= \frac{a(t)}{t}\,\frac{(x-a(t))}{b^2(t)} + \frac{1}{2t}\left(\frac{(x-a(t))^2}{b^2(t)} - 1\right), \quad x \in \mathbb{R}.\end{aligned}$$

Remark 2.2. *While IPA cannot be applied to this problem in a straightforward way, IPA can be applied to the distributional representation of the Wiener process. Indeed, interpreting $Z_i(t)$ as normally distributed with mean $a(t)$ given in (2.6) and standard deviation as in (2.7), IPA can be readily applied to compute $\partial_t Z_i(t)$. It is, however, worth noting that with this change of the model $(Z_i(t) : t \geq 0)$ does not represent the path of the stock price any more, as only the marginal distributions are correct. In the following we will denote IPA for the distributional model by IPA*.*

In the trading strategy we have a total of $n = 400$ securities. The mean historical annual return for all possible stocks in the above economies is set to $\beta = 0.03$, and the selected stocks have an additional mean out-performance of $\eta = 0.04$. The implied volatility $\sigma = 0.15$ is per annum for each stock, and we consider a trading period of one week $t = 1/52 = 0.0192$. These securities are divided into k allotments of m, specifically choosing m and k to equal $(m,k) = (40,10),(10,40)$. The rank of our interest is chosen such that $\alpha = 0.50, 0.90$.

For IPA we present the results for IPA*, i.e., IPA applied to the distributional model as the direct sample path approach fails, see Remark 2.2. The numerical values are presented in Table 2.2. The statistics RMS Error, Std. Dev., WNV, and Cov. Prob. are all determined over 500 estimates.

In Figure 2.2 the coverage probabilities of the 95% confidence interval are presented. The number of samples for each estimate of Z is fixed to $m = 40$ and the number of independent replications k is varied.

The main challenge for IPA is that Lipschitz continuity often fails to hold or is hard to check. When condition **(A2 IPA)** fails, typically the offending random variables are conditioned to smooth out the discontinuities. This approach is called *Smoothed Perturbation Analysis* (SPA), and for details for this method we refer to [31]. A concern for the SPA estimator is that the conditional estimator becomes, in general, more complex than the IPA estimator, and this complexity

$m = 40$, $k = 10$, $\alpha = 0.50$. Actual Value = 4.235×10^{-1}

	IPA*	SF1	SF2	MVD1	MVD2
Mean	4.307×10^{-2}	0.6645	-4.922×10^{2}	4.588×10^{-2}	3.148×10^{-2}
RMS Error	0.7965	1.715×10^{3}	1.070×10^{4}	3.7610	9.7318
Std. Dev.	3.376×10^{-2}	7.269×10^{1}	4.510×10^{2}	0.1594	0.4124
WNV	2.384×10^{-8}	0.4694	1.651×10^{1}	3.408×10^{-5}	5.334×10^{-5}
Cov. Prob.	0.9060	0.9200	0.8240	0.9260	0.9480

$m = 40$, $k = 10$, $\alpha = 0.95$. Actual Value = 0.7291

	IPA*	SF1	SF2	MVD1	MVD2
Mean	0.7286	1.3731	-3.3516	0.7410	0.6585
RMS Error	6.597×10^{-2}	1.023×10^{2}	6.378×10^{2}	0.4247	1.0565
Std. Dev.	4.815×10^{-2}	7.466×10^{1}	4.655×10^{2}	0.3098	0.7678
WNV	3.000×10^{-8}	0.4861	1.697×10^{1}	1.281×10^{-4}	1.822×10^{-4}
Cov. Prob.	0.9040	0.9200	0.8220	0.9020	0.8340

$m = 10$, $k = 40$, $\alpha = 0.50$. Actual Value = -5.885×10^{-2}

	IPA*	SF1	SF2	MVD1	MVD2
Mean	-5.062×10^{-2}	0.1328	-0.5240	-1.030×10^{-2}	-6.380×10^{-2}
RMS Error	5.7327	3.080×10^{3}	9.492×10^{3}	1.580×10^{1}	3.745×10^{2}
Std. Dev.	3.376×10^{-2}	1.814×10^{1}	5.592×10^{1}	9.298×10^{-2}	0.2129
WNV	4.175×10^{-8}	4.978×10^{-2}	4.588×10^{-1}	1.942×10^{-5}	2.890×10^{-5}
Cov. Prob.	0.9440	0.9460	0.9220	0.9380	0.9340

$m = 10$, $k = 40$, $\alpha = 0.95$. Actual Value = 0.6161

	IPA*	SF1	SF2	MVD1	MVD2
Mean	0.6176	0.7773	-0.4683	0.6143	0.5383
RMS Error	7.283×10^{-2}	3.021×10^{2}	9.644×10^{2}	0.2459	0.4983
Std. Dev.	4.489×10^{-2}	1.863×10^{1}	5.947×10^{1}	0.1517	0.2973
WNV	5.101×10^{-8}	5.118×10^{-2}	0.5073	5.149×10^{-5}	5.593×10^{-5}
Cov. Prob.	0.9180	0.9440	0.9180	0.9360	0.8980

Table 2.2: The summary statistics for the five derivative estimators in comparison with the actual value for the Performance Control of a Basket of Stocks example.

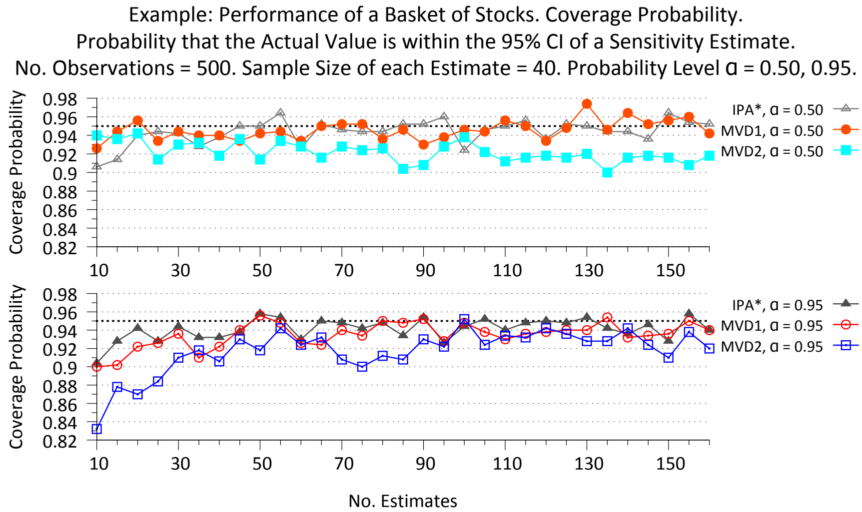

Figure 2.2: The coverage probability when the number of estimates, k, increases in the Performance of a Basket of Stocks example.

may even render the technique impractical. An example of this difficulty is the portfolio credit risk model presented in [32]. In the following we compare the SPA estimator in [32] with both MVD estimators. The example and the numerical settings are as in [32], and for a motivation and detailed discussion of this model we refer to [5]. Let

$$Z = \sum_{i=1}^{r} l_i \, 1\{X_i < x_i\}$$

denote the loss of a portfolio with r obligors, where X_i denotes the default of obligor i and x_i is a threshold indicating when obligor i defaults which will cause a loss of l_i. For the numerical example, we assume that $r = 2$ and that the l_i are independent uniformly distributed on $[0,1]$. We let

$$X_i(\theta) = \frac{\rho V + \sqrt{1-\rho^2}\,\eta_i}{W(\theta)}, \quad 1 \le i \le r,$$

with η_i normally distributed with mean 0 and variance 1 modelling the obligor's idiosyncratic risk, V a standard normally distributed random variable modelling the common factor that affects all obligors, and $W(\theta)$ an exponentially distributed random variable with rate $\theta = 1/0.3$ modelling the shocks that are common to all obligors. In this example the weight factor ρ is set to 0.6. Consequently, the loss of a portfolio is represented by

$$Z(\theta) = \sum_{i=1}^{2} l_i \, 1\left\{\rho V + \sqrt{1-\rho^2}\,\eta_i < x_i W(\theta)\right\}.$$

For the experiments we choose $\alpha = 0.95$ and analyze the sensitivity of the value at risk at this level.

Let $v(x) = 1$, then W is differentiable w.r.t. the space of Borel bounded functions $\mathfrak{B} =: \mathfrak{B}_v$, according to Section 1.2.4.1. Since

$$\mathbb{E}_\theta[Z|W = x] = \mathbb{E}_\theta\left[\sum_{i=1}^{2} l_i \, 1\left\{\rho V + \sqrt{1-\rho^2}\eta_i < x_i W\right\}\middle| W = x\right]$$

as a mapping in x is an element of $\mathfrak{B}$, applying Lemma 2.1 with $\mathfrak{D} = \hat{\mathfrak{D}}_1 = \mathfrak{B}$, elaborating on the presentation of the weak derivative of the exponential distribution as put forward in Section 1.2.4.1, it follows that Z is $\mathfrak{B}$-differentiable. The $\mathfrak{B}$-derivative is given by $Z^+ = Z$,

$$Z^- = \sum_{i=1}^{2} l_i \, 1\left\{\rho V + \sqrt{1-\rho^2}\,\eta_i < x_i(W+Y)\right\},$$

and $c_\theta = 1/\theta$, with Y an exponentially distributed random variable with rate θ, independent of everything else.

For the numerical experiments we have chosen $n = 10^4, 10^6$, and that the array parameters to be equal; i.e., $m = k = 100$ for $n = 10^4$, and $m = k = 1000$ for $n = 10^6$. The results are presented in Table 2.3. We observe that both MVD derivative estimators perform worse than the SPA estimator, with the symmetric estimator (MVD2) having worse RMS error and standard deviation than the standard estimator (MVD1). Most evident in the table are the rates of convergence of $O(n^{-1/2})$ for the sample path estimator and $O(n^{-1/4})$ for the MVD derivative estimators. Between the two MVD derivative approaches, the work-normalized variance is more favourable for the standard estimator.

It is worth noting that the SPA estimator is based on a meticulous sample path analysis, whereas the MVD estimators are attained in a straightforward way. As the results put forward in Table 2.3 illustrate, this flexibility and ease of implementation of the MVD estimators comes at the price of higher computational costs. Note that the MVD estimator is not affected if in place of Z we consider $f(Z)$, where f is, for example, a (discontinuous) utility function, whereas the SPA estimator would have to be computed anew in this case.

Derivative w.r.t. θ. Summary Statistics.

$\alpha = 0.95$ $m = k$	$n = 10^4$. Actual Value $= 6.28 \times 10^{-2}$			$n = 10^6$. Actual Value $= 6.28 \times 10^{-2}$		
	SPA	MVD1	MVD2	SPA	MVD1	MVD2
Mean	6.29×10^{-2}	5.72×10^{-2}	5.92×10^{-2}	6.28×10^{-2}	6.32×10^{-2}	6.48×10^{-2}
RMS Error	1.9×10^{-3}	1.41×10^{-1}	7.49×10^{-1}	1.9×10^{-4}	3.27×10^{-2}	2.81×10^{-1}
Std. Dev.	1.9×10^{-3}	6.81×10^{-3}	4.69×10^{-2}	1.9×10^{-4}	2.02×10^{-3}	1.76×10^{-2}

Table 2.3: Results for both the standard (MVD1) and symmetric (MVD2) MVD derivative estimates for the sensitivity of the $\alpha = 0.95$ quantile of the portfolio loss due to credit risk w.r.t. the common shock parameter θ. The SPA results are cited from [32].

2.4 The Quantile Problem

In this section we present our results for the QP. The technical analysis in provided in Section 2.4.1. Then, in Section 2.4.2, we will present numerical examples.

2.4.1 Technical Analysis

Let $\mathbf{Z}(j) = (Z_i(j) : 1 \le i \le m)$, for $1 \le j \le k$, be a sequence of i.i.d. samples that has the same distribution as $\mathbf{Z} = (Z_i : 1 \le i \le m)$. Furthermore, let $(Z_i^+(j), Z_i^-(j))$, for $1 \le j \le k$ and $1 \le i \le m$, be an i.i.d. sequence of samples that has the same distribution as (Z_i^+, Z_i^-): see **(A2 MVD)**. For $l = 1,2$, we denote the value of $D_l^{\text{MVD}}(m)$ evaluated from sample $\mathbf{Z}(j)$ by $D_{l,j}^{\text{MVD}}$, and we denote the average over k samples by

$$D_l^{\text{MVD}}(m,k) = \frac{1}{k}\sum_{j=1}^{k} D_{l,j}^{\text{MVD}}(m). \quad l = 1,2.$$

For example, by (2.5), $D_1^{\text{MVD}}(m,k)$ reads explicitly

$$D_1^{\text{MVD}}(m,k) = \frac{c_\theta}{k}\sum_{j=1}^{k}\sum_{i=1}^{m}\Big(\text{ord}_{\alpha,m}(Z_1(j),\dots,Z_{i-1}(j),Z_i^+(j),Z_{i+1}(j),\dots,Z_m(j)) \\ - \text{ord}_{\alpha,m}(Z_1(j),\dots,Z_{i-1}(j),Z_i^-(j),Z_{i+1}(j),\dots,Z_m(j))\Big).$$

In the same vein we define

$$D_l^{\text{SF}}(m,k) = \frac{1}{k}\sum_{j=1}^{k} D_{l,j}^{\text{SF}}(m), \quad l = 1,2,$$

and

$$D^{\text{IPA}}(m,k) = \frac{1}{k}\sum_{j=1}^{k} D_j^{\text{IPA}}(m),$$

where

$$D_j^{\text{IPA}}(m) = \frac{\partial}{\partial\theta} Z_{k(j)}(j),$$

and $k(j)$ is the realization of the index $(\lceil \alpha m \rceil : m)$ for the j^{th} sample of $\mathbf{Z}$.

We write $q_\alpha(\theta)$ to express the dependence of the quantile on θ. Since Z is a continuous random variable, F_θ^{-1} is differentiable w.r.t. its argument y, and if F_θ is differentiable w.r.t. θ, $F_\theta^{-1}(y)$ is as well. By definition, see (1.2.4.1),

$$\alpha = F_\theta(q_\alpha(\theta)),$$

and we attain an expression for the quantile sensitivity by implicit differentiation: Section 1.3.1 has the details. The final expression is provided in Equation (1.39) and is again provided:

$$\frac{\partial}{\partial\theta} q_\alpha(\theta) = -\frac{\frac{\partial}{\partial\theta} F_\theta(q_\alpha(\theta))}{f_\theta(q_\alpha(\theta))}. \tag{2.8}$$

The assumptions required for all of the proofs are given below. These mostly focus on the properties of F_θ and f_θ and all of these assumptions are easily verifiable or assessable.

(A3) The random variable Z is square integrable.

(A4) In a neighbourhood of $q_\alpha(\theta)$ it holds that $f_\theta(x) > 0$.

(A5) Let $\Theta_0 \subset \Theta$ be an open neighbourhood of θ_0, where θ_0 denotes the value of θ for which the derivative is evaluated, and let $B(\alpha)$ be an open neighbourhood of $x = q_\alpha(\theta_0)$. Assume that f_θ is continuously differentiable on Θ_0 for all $x \in B(\alpha)$.

(A6) f_θ is differentiable w.r.t. x with bounded derivative in the neighbourhood of $x = q_\alpha(\theta)$.

The first result establishes strong consistency of the estimators presented in Section 2.2.1. The proof requires two ancillary results, Lemmas A.1, and A.2. Lemma A.1 is an almost sure upper bound, given for sufficiently large m, for the distance between a quantile and an order statistic. This allows to justify that the order statistics we have is within a neighbourhood of the quantile. Lemma A.2 is to derive, given our assumptions, that the spacing terms presented in the proofs are uniformly integrable. The SF and MVD derivations employ Taylor series in terms of a spacing function.

Theorem 2.2 (Strong Consistency)**.** *Suppose that Assumptions* **(A1)** *and* **(A3)** *to* **(A6)** *are satisfied. Then,*

(i) *under Assumption* **(A2 MVD)***, for $l = 1, 2$, it holds almost surely that*

$$\lim_{(m,k)\to(\infty,\infty)} D_l^{\mathrm{MVD}}(m,k) = \frac{\partial}{\partial\theta} q_\alpha(\theta).$$

(ii) *under Assumption* **(A2 SF)** *it holds almost surely that*

$$\lim_{(m,k)\to(\infty,\infty)} D_l^{\mathrm{SF}}(m,k) = \frac{\partial}{\partial\theta} q_\alpha(\theta),$$

(iii) *under Assumption* **(A2 IPA)** *it holds almost surely that*

$$\lim_{(m,k)\to(\infty,\infty)} D^{\mathrm{IPA}}(m,k) = \frac{\partial}{\partial\theta} q_\alpha(\theta).$$

Proof: We begin by considering the recursive nature of order statistic $Z_{\lceil\alpha m\rceil:m}$ when an additional random variable, Z_m, is included to the sample of the order statistic formed by $\mathbf{Z}\backslash Z_m = (Z_1,\ldots,Z_{m-1})$. For this discussion and for the proof in general, we require the following two specific order statistics based on $m-1$ elements:

$$Y_1 := Z_{\lceil\alpha m\rceil-1:m-1} = \mathrm{ord}_{\alpha_1,m}(Z_1,Z_2,\ldots,Z_{m-1}),$$

with α_1 such that $\lceil\alpha m\rceil - 1 = \lceil\alpha_1(m-1)\rceil$, and

$$Y_2 := Z_{\lceil\alpha m\rceil:m-1} = \mathrm{ord}_{\alpha_2,m}(Z_1,Z_2,\ldots,Z_{m-1}),$$

where α_2 is defined from $\lceil\alpha m\rceil = \lceil\alpha_2(m-1)\rceil$. The rationale behind the above definition is that adding the random variable Z_m to the ordered collection of size $m-1$, three outcomes are possible: the order statistic $Z_{\lceil\alpha m\rceil:m}$ is equal to Y_1, Y_2, or Z_m itself. If $Z_m \le Y_1$, there are now $\lceil\alpha m\rceil$ random variables with a value less than or equal to Y_1, and so the order statistic $Z_{\lceil\alpha m\rceil:m} = Y_1$ with probability one. If $Z_m > Y_2$, the random variable Z_m is greater in value to at least $\lceil\alpha m\rceil$ random variables in the ordered collection of m, which implies that $Y_2 = Z_{\lceil\alpha m\rceil:m}$ is the value of the order statistic. When $Y_1 < Z_m \le Y_2$, the random variable Z_m

is now the $\lceil \alpha m \rceil^{\text{th}}$ smallest value within the collection, and so with probability one, $Z_{\lceil \alpha m \rceil:m} = Z_m$. Hence, we attain

$$Z_{\lceil \alpha m \rceil:m} \,|\, \mathbf{Z} \backslash Z_m = \begin{cases} \textbf{Random Variable} & \textbf{Event} \\ Y_1 & \{\omega : Z_m \le Y_1\} \\ Z_m & \{\omega : Y_1 < Z_m \le Y_2\} \\ Y_2 & \{\omega : Z_m > Y_2\}. \end{cases} \tag{2.9}$$

We now turn to the analysis of the derivative estimators, and begin with *Part (iii)*, the IPA derivative estimator. For this method, we write $Z_m(\theta)$ to express the fact that Z_m is a function of θ. Let $k(\theta)$ denote the index such that

$$Z_{k(\theta)}(\theta) = Z_{\lceil \alpha m \rceil:m}(\theta).$$

Let $Z_j(\theta) = F_\theta^{-1}(U_j)$ for $(U_j : 1 \le j \le m)$ be an i.i.d. sequence of uniform-$(0,1)$-distributed random variables. The fact that $F_\theta(x)$ is monotone non-decreasing as a mapping of x, implies that $F_\theta^{-1}(u)$ is monotone non-decreasing as a mapping of u, which implies that $k(\theta) = j$ if and only if $U_j = U_{\lceil \alpha m \rceil:m}$ almost surely. Hence, for a given realization of $(U_j : 1 \le j \le m)$, the index $k(\theta)$ does not depend on θ, and $\mathbb{P}(j = k(\theta)) = 1/m$. By computation,

$$\begin{aligned} \mathbb{E}[Z_{\lceil \alpha m \rceil:m}(\theta)] &= \mathbb{E}\big[Z_{k(\theta)}(\theta)\big] \\ &= \sum_{j=1}^{m} \mathbb{E}\big[Z_j(\theta)\, 1\{j = k(\theta)\}\big] \\ &= \sum_{j=1}^{m} \mathbb{E}\big[Z_j(\theta)\big| j = k(\theta)\big]\, \mathbb{P}(j = k(\theta)) \\ &= \sum_{j=1}^{m} \mathbb{E}\big[Z_j(\theta)\big| j = k(\theta)\big] \cdot \frac{1}{m} \\ &= \mathbb{E}[Z_m(\theta) | m = k(\theta)] \\ &= \mathbb{E}[Z_m(\theta) \,|\, Z_m(\theta) \in (Y_1, Y_2]]\,, \end{aligned} \tag{2.10}$$

where for the last but one equality we have used the fact that $(Z_i(\theta), 1 \le i \le m)$ is an i.i.d. sequence.

Since differentiation does not affect the rank of the order statistic,

$$\frac{d}{d\theta}\mathbb{E}[Z_m(\theta) \,|\, Z_m(\theta) \in (Y_1, Y_2]] = \mathbb{E}\left[\frac{\partial}{\partial \theta} Z_m(\theta) \,\middle|\, Z_m(\theta) \in (Y_1, Y_2]\right]. \tag{2.11}$$

Suri and Zazanis, [100], established an alternative representation for the parameter derivative of a random variable via the distribution function F_θ. Specifically, writing $Z_m(\theta) = F_\theta^{-1}(U)$ via its inverse distribution and implicitly differentiating this function w.r.t. θ yields

$$\frac{\partial}{\partial\theta} Z_m(\theta) = -\frac{\frac{\partial}{\partial\theta} F_\theta(Z_m(\theta))}{f_\theta(Z_m(\theta))}. \tag{2.12}$$

By Assumptions **(A4)** and **(A5)**, for m sufficiently large, the RHS of Equation (2.12) is bounded above and continuous at $Z_m(\theta)$ as $Z_m(\theta) \in (Y_1, Y_2]$, and by Lemma A.1, both order statistics are in the neighbourhood of the quantile. Noting that $Z_{l:m}(\theta)$ are almost surely continuous throughout Θ for $l = 1,\dots,m$, applying the Mean-Value Theorem establishes (2.11), and we arrive that the following representation for the IPA estimator

$$\mathbb{E}\left[D^{\text{IPA}}(m)\right] = \mathbb{E}\left[-\frac{\frac{\partial}{\partial\theta} F_\theta(Z_m(\theta))}{f_\theta(Z_m(\theta))} \,\middle|\, Z_m(\theta) \in (Y_1, Y_2] \right].$$

We are nearing the representation of the quantile sensitivity in Equation (2.8), and together with **(A2 IPA)**, to apply the Strong Law of Large Numbers. Since $g := (\partial_\theta F_\theta)/f_\theta$ is continuous and bounded on a neighbourhood given in Lemma A.1, following [66], pp. 280-281, the function g can be written as a difference of two increasing, continuous functions, g_1, g_2, and $g = g_1 - g_2$. Since for every m it holds almost surely that $Y_1 < Z_m(\theta) \le Y_2$, it follows from [96], Property B, p. 183, that

$$\mathbb{E}\left[D^{\text{IPA}}(m)\right] \le \mathbb{E}\left[-\left(g_1(Y_1) - g_2(Y_2)\right)\right].$$

For the lower bound, the order statistics Y_1 and Y_2 are swapped. From [4, 34], $Y_1, Y_2 \overset{a.s.}{\to} q_\alpha(\theta)$, which, by the Continuous Mapping Theorem [101], implies that $g_1(Y_l) \overset{a.s.}{\to} g_1(q_\alpha(\theta))$, and $g_2(Y_l) \overset{a.s.}{\to} g_2(q_\alpha(\theta))$. Then by Slutsky's Theorem, [45],

$$g_1(Y_1) - g_2(Y_2) \overset{a.s.}{\to} g(q_\alpha(\theta)) \;=\; -\frac{\partial}{\partial\theta} q_\alpha(\theta)$$

and

$$g_1(Y_2) - g_2(Y_1) \overset{a.s.}{\to} g(q_\alpha(\theta)) \;=\; -\frac{\partial}{\partial\theta} q_\alpha(\theta).$$

Since g is bounded from above, it follows from the Dominated Convergence Theorem that

$$\lim_{m\to\infty} \mathbb{E}\left[D^{\text{IPA}}(m)\right] = \frac{\partial}{\partial\theta} q_\alpha(\theta).$$

From this position, we apply the Strong Law of Large Numbers, proving the claim for the IPA estimator.

Since SF and MVD are distributional approaches, the θ-dependence is expressed via the distribution function. From Theorem 2.1 it follows that the standard and symmetric estimators are equal in expectation, and in the following we will focus on the symmetric estimators for each approach.

We first turn to the SF estimator, *Part (ii).* For the proof we work with the expected value of $Z_{\lceil \alpha m \rceil : m}$ conditioned on the sample $\mathbf{Z} \backslash Z_m$. As $Z_m \sim F_\theta$, from (2.9) the recursive form of $Z_{\lceil \alpha m \rceil : m}$ conditioned on $\mathbf{Z} \backslash Z_m$ is given by

$$\mathbb{E}_\theta\big[Z_{\lceil \alpha m \rceil : m} \,\big|\, \mathbf{Z} \backslash Z_m\big] = \begin{cases} \textbf{Random Variable} & \textbf{Probability} \\ Y_1 & F_\theta(Y_1) \\ \mathbb{E}_\theta[Z_m \,|\, \mathbf{Z} \backslash Z_m] & F_\theta(Y_2) - F_\theta(Y_1) \\ Y_2 & 1 - F_\theta(Y_2), \end{cases}$$

from which it follows that

$$\begin{aligned} \mathbb{E}_\theta\big[Z_{\lceil \alpha m \rceil : m} \,\big|\, \mathbf{Z} \backslash Z_m\big] = {} & Y_1 F_\theta(Y_1) + Y_2 (1 - F_\theta(Y_2)) \\ & + \mathbb{E}_\theta[Z_m 1\{Z_m \in (Y_1, Y_2]\} \,|\, \mathbf{Z} \backslash Z_m]. \end{aligned} \tag{2.13}$$

Via integration by parts, the third term in Equation (2.13) has the more convenient form

$$\begin{aligned} \mathbb{E}_\theta[Z_m 1\{Z_m \in (Y_1, Y_2]\} \,|\, \mathbf{Z} \backslash Z_m] &= \int_{Y_1}^{Y_2} z f_\theta(z)\, dz \\ &= Y_2 F_\theta(Y_2) - Y_1 F_\theta(Y_1) - \int_{Y_1}^{Y_2} F_\theta(z)\, dz. \end{aligned}$$

This simplifies Equation (2.13) to

$$\mathbb{E}_\theta\big[Z_{\lceil \alpha m \rceil : m} \big| \mathbf{Z} \backslash Z_m\big] = Y_2 - \int_{Y_1}^{Y_2} F_\theta(z)\, dz. \tag{2.14}$$

The order statistic Y_2 has w.r.t. θ a sample path derivative of zero, which stems from the fact that only the distribution depends on θ and not the realizations. The interchange of derivative and integral operators in the above equation is justified by Theorem 2.1. Let p denote an arbitrary Lebesgue density that dominates f_θ, i.e., the support of p includes the support of f_θ for $\theta \in \Theta$, then

$$\begin{aligned} \frac{d}{d\theta} \mathbb{E}_\theta\big[m Z_{\lceil \alpha m \rceil : m} \,\big|\, \mathbf{Z} \backslash Z_m\big] &= -\int_{Y_1}^{Y_2} m \frac{\partial}{\partial \theta} F_\theta(z)\, dz \\ &= -\int_{Y_1}^{Y_2} m \mathbb{E}_p\left[1\{Z_m \le z\} \frac{\frac{\partial}{\partial \theta} f_\theta(Z_m)}{p(Z_m)} \right] dz. \end{aligned} \tag{2.15}$$

We expand Equation (2.15) as a two term Taylor series around $z = Y_1$, where we use the Lagrangian remainder. On the open ball $B_{2r_m}(q_\alpha(\theta))$, with center $y = q_\alpha(\theta)$ and radius r_m, defined in Lemma A.2, it holds

$$\frac{d}{d\theta}\mathbb{E}_\theta\big[mZ_{\lceil\alpha m\rceil:m}\,\big|\,\mathbf{Z}\backslash Z_m\big] \geq -m\frac{\partial}{\partial\theta}F_\theta(Y_1)(Y_2-Y_1) - \frac{1}{2}m\left(\sup_{y\in B_{2r_m}(q_\alpha(\theta))}\left|\frac{\partial^2}{\partial x\partial\theta}F_\theta(y)\right|\right)(Y_2-Y_1)^2. \quad (2.16)$$

For the upper bound, the infimum replaces the supremum. All the following limiting results are as $m\to\infty$. We begin with Bahadur's result [4], [34] in which $Y_1 \overset{a.s.}{\to} q_\alpha(\theta)$ and, together with the Continuous Mapping Theorem [101], $\partial_\theta F_\theta(Y_1) \overset{a.s.}{\to} \partial_\theta F_\theta(q_\alpha(\theta))$. For i.i.d. random variables, the limiting result from Pyke implies $m(Y_2-Y_1) \overset{d}{\to} E/f_\theta(q_\alpha(\theta))$, where E is an exponential random variable with rate one. With the map $a(x) = x^2$, by the Continuous Mapping Theorem, together with Property H, p. 185 from [96], $m(Y_2-Y_1)^2 \overset{a.s.}{\to} 0$. As a consequence, the mixed derivative of F_θ is bounded above by **(A5)**, and we attain uniformly integrability of the expansion in (2.16). Combining all these results with Slutsky's Theorem [45] leads to the following limit

$$\frac{d}{d\theta}\mathbb{E}_\theta\big[mZ_{\lceil\alpha m\rceil:m}\,\big|\,\mathbf{Z}\backslash Z_m\big] \overset{d}{\to} -\frac{\frac{\partial}{\partial\theta}F_\theta(q_\alpha(\theta))}{f_\theta(q_\alpha(\theta))}E, \quad (2.17)$$

and from the definition of the quantile sensitivity in Equation (2.8)

$$\frac{d}{d\theta}\mathbb{E}_\theta\big[mZ_{\lceil\alpha m\rceil:m}\,\big|\,\mathbf{Z}\backslash Z_m\big] \overset{d}{\to} \frac{\partial}{\partial\theta}q_\alpha(\theta)E.$$

By **(A2 SF)**, F_θ is also weakly differentiable, and we attain

$$\left|\frac{\partial}{\partial\theta}F_\theta(Y_1)\right| = c_\theta\left|F_\theta^+(Y_1) - F_\theta^-(Y_1)\right| \leq c_\theta < \infty,$$

and it follows from Lemma A.2 that the upper and lower bounds in (2.16) are uniformly integrable. In addition, as $m\to\infty$, the expected values of both sequences converge to $\partial_\theta q_\alpha(\theta)$. Hence, we attain

$$\lim_{m\to\infty}\mathbb{E}\left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta\big[mZ_{\lceil\alpha m\rceil:m}\,\big|\,\mathbf{Z}\backslash Z_m\big]\right] = \frac{\partial}{\partial\theta}q_\alpha(\theta),$$

using (2.15) together with the fact that c_θ is bounded on a closed neighbourhood of θ, with θ as an interior point, we apply the IPA part of Theorem 2.1, which yields for any m that

$$\frac{d}{d\theta}\mathbb{E}\big[\mathbb{E}_\theta\big[mZ_{\lceil\alpha m\rceil:m}\,\big|\,\mathbf{Z}\backslash Z_m\big]\big] = \mathbb{E}\left[\frac{\partial}{\partial\theta}\mathbb{E}_\theta\big[mZ_{\lceil\alpha m\rceil:m}\,\big|\,\mathbf{Z}\backslash Z_m\big]\right].$$

In the above formula, differentiation is only applied to $Z_m(\theta)$. More specifically, if the derivative is evaluated at θ_0, then the above equation is short hand notation for

$$\left.\frac{d}{d\theta}\right|_{\theta=\theta_0} \mathbb{E}\big[\mathbb{E}\big[m\,\mathrm{ord}_{\alpha,m}(Z_1(\theta_0),\dots,Z_{m-1}(\theta_0),Z_m(\theta))\,\big|\,Z_1(\theta_0),\dots,Z_{m-1}(\theta_0)\big]\big]$$
$$= \mathbb{E}\left[\left.\frac{\partial}{\partial\theta}\right|_{\theta=\theta_0} \mathbb{E}\big[m\,\mathrm{ord}_{\alpha,m}(Z_1(\theta_0),\dots,Z_{m-1}(\theta_0),Z_m(\theta))\,\big|\,Z_1(\theta_0),\dots,Z_{m-1}(\theta_0)\big]\right].$$

By Theorem 2.1, the LHS of the above equation is equal to $\mathbb{E}[D_2^{\mathrm{SF}}(m)]$, see Remark 2.1. To summarize, we have shown that

$$\lim_{m\to\infty} \mathbb{E}_\theta\big[D_2^{\mathrm{SF}}(m)\big] = \frac{\partial}{\partial\theta} q_\alpha(\theta).$$

The proof of the claim now follows from the Strong Law of Large Numbers.

Part (i). For the Strong Consistency result via MVD, this derivation is similar to the SF version. This derivation will provide an integral similar to Equation (2.15). From that position, the proof method is identical. We define the order statistics $Z^+_{\lceil\alpha m\rceil:m}$, $Z^-_{\lceil\alpha m\rceil:m}$, analogously to $Z_{\lceil\alpha m\rceil:m}$ but the random variable Z_m in both cases is replaced in turn by the measure-valued derivative random variables Z^+_m, and Z^-_m; namely,

$$Z^+_{\lceil\alpha m\rceil:m} = \mathrm{ord}_{\alpha,m}\big(Z_1,\dots,Z_{m-1},Z^+_m\big),$$
$$Z^-_{\lceil\alpha m\rceil:m} = \mathrm{ord}_{\alpha,m}\big(Z_1,\dots,Z_{m-1},Z^-_m\big).$$

First note that by (2.15) and the arguments put forward in the analysis of the SF method, we arrive at

$$\mathbb{E}_\theta\Big[Z^+_{\lceil\alpha m\rceil:m}\,\Big|\,\mathbf{Z}\backslash Z_m\Big] = \begin{cases} \textbf{Random Variable} & \textbf{Probability} \\ Y_1 & F^+_\theta(Y_1) \\ \mathbb{E}_\theta\big[Z^+_m\,\big|\,\mathbf{Z}\backslash Z_m\big] & F^+_\theta(Y_2) - F^+_\theta(Y_1) \\ Y_2 & 1 - F^+_\theta(Y_2). \end{cases} \tag{2.18}$$

An equivalent expression holds for $Z^-_{\lceil\alpha m\rceil:m}$. Correspondingly, the above conditional expectation is equivalent to,

$$\mathbb{E}_\theta\Big[Z^+_{\lceil\alpha m\rceil:m}\,\Big|\,\mathbf{Z}\backslash Z_m\Big] = Y_2 - \int_{Y_1}^{Y_2} F^+_\theta(z)\,dz,$$

and so the conditional expectation of the symmetric MVD derivative estimator is an integral of the difference between the MVD distribution functions

$$\begin{aligned}\mathbb{E}_\theta\big[D_2^{\mathrm{MVD}}(m)\,\big|\,\mathbf{Z}\backslash Z_m\big] &= \mathbb{E}_\theta\Big[c_\theta m\left(Z^+_{\lceil\alpha m\rceil:m} - Z^-_{\lceil\alpha m\rceil:m}\right)\Big|\,\mathbf{Z}\backslash Z_m\Big] \\ &= -\int_{Y_1}^{Y_2} c_\theta m\left(F_\theta^+(z) - F_\theta^-(z)\right)dz. \end{aligned} \tag{2.19}$$

Justification for the interchange between integral and derivative is given in Theorem 2.1. For the remainder of the proof, the only difference to the SF derivation is the equivalent derivation leading to Equation (2.17): the distributional convergence of the parameter derivative to the conditional expectation. The θ-derivative of the distribution function is defined via the measure-valued derivative, which can be seen by letting $h(Z) = 1\{Z \le z\}$ in Equation (1.21),

$$\frac{\partial}{\partial\theta}F_\theta(z) = c_\theta\left(F_\theta^+(z) - F_\theta^-(z)\right). \tag{2.20}$$

Equation (2.21) below is the MVD equivalent of Equation (2.17).

$$\begin{aligned}\lim_{m\to\infty}\frac{d}{d\theta}\mathbb{E}_\theta\Big[c_\theta m\left(Z^+_{\lceil\alpha m\rceil:m} - Z^-_{\lceil\alpha m\rceil:m}\right)\Big|\,\mathbf{Z}\backslash Z_m\Big] &= -\frac{c_\theta\left(F_\theta^+(q_\alpha(\theta)) - F_\theta^-(q_\alpha(\theta))\right)}{f_\theta(q_\alpha(\theta))}E \\ &= \frac{\partial}{\partial\theta}q_\alpha(\theta)E. \end{aligned} \tag{2.21}$$

The rest follows as in the Score Function proof using the MVD version of Theorem 2.1. □

The Strong Consistency result in Theorem 2.2 determined that our derivative estimators are asymptotically unbiased w.r.t. the correct value of the quantile sensitivity. In the following we determine the leading order behaviour of the difference between the actual quantile sensitivity and the estimators. Both Lemmas A.1 and A.2 are again applied.

Theorem 2.3 (Rate of Asymptotic Unbiasedness). *Suppose that conditions* **(A1)** *and* **(A3)** *to* **(A6)** *are satisfied, and assume that each MVD random pair together with the nominal random variable; i.e.,* (Z, Z_1^+, Z_1^-)*, possess the same, arbitrary dependence. For* $D(m) \in \{D_1^{\mathrm{MVD}}(m), D_2^{\mathrm{MVD}}(m), D_1^{\mathrm{SF}}(m), D_2^{\mathrm{SF}}(m), D^{\mathrm{IPA}}(m)\}$ *it holds that*

$$\left|\mathbb{E}_\theta\left[D(m) - q_\alpha(\theta)\right]\right| = O\left(\frac{1}{m}\right),$$

provided the appropriate version of condition **(A2)** *holds as well.*

Proof: For IPA the derivation is given in Hong [57] by writing the quantile derivative as a measurable function of the quantile, where the key ingredient of the proof is a Taylor series expansion of the order statistic provided in David and Nagaraja, [18].

For SF and MVD, the proof applies to both derivative estimators since both estimators have the same mean. The proof for the symmetric SF estimator is given. In this proof, we define a relationship operator $\simeq$ as follows: for two families of random variables, (R_m), (S_m), on a common probability space, $(\Omega, \mathscr{F}, \mathbb{P})$, these families possess the relationship $R_m \simeq S_m$ if a measurable set K exists with $\mathbb{P}(K) = 1$ such that for all $\omega \in K$ it holds that $\lim_{m\to\infty} R_m/S_m = 1$. Hence, "$\simeq$" can be considered as the almost sure equivalent to the similarity relationship, denoted by $\sim$, between (deterministic) functions on the real line, which is given as follows: for functions, $r(x)$, $s(x)$, $r(x) \sim s(x)$ if $\lim_{x\to\infty} r(x)/s(x) = 1$.

We begin with the conditional expectation for the symmetric SF estimator in Equation (2.15) in the Strong Consistency result. The random variables $Y_1 = Z_{\lceil \alpha m\rceil -1:m-1}$ and $Y_2 = Z_{\lceil \alpha m\rceil:m-1}$ represent order statistics on the data set $\mathbf{Z}\backslash Z_m = (Z_1, \dots, Z_{m-1})$, and with this notation we write

$$\mathbb{E}_\theta\big[D_2^{\mathrm{SF}}(m)\,\big|\,\mathbf{Z}\backslash Z_m\big] = -\int_{Y_1}^{Y_2} m \frac{\partial}{\partial\theta} F_\theta(z)\, dz.$$

As already mentioned in the proof of Theorem 2.2, the parameter derivative of the distribution function is given via its expectation

$$\frac{\partial}{\partial\theta} F_\theta(z) = \mathbb{E}_q\left[1\{Z_m \le z\}\,\frac{\frac{\partial}{\partial\theta} f_\theta(Z_m)}{p(Z_m)}\right].$$

To find the leading term of the integral, we use a two term Taylor expansion of $\partial_\theta F_\theta$ centred at $z = Y_1$, using the Lagrangian form of the remainder

$$\mathbb{E}_\theta\big[D_2^{\mathrm{SF}}(m)\,\big|\,\mathbf{Z}\backslash Z_m\big] = -\int_{Y_1}^{Y_2} \left(m\frac{\partial}{\partial\theta} F_\theta(Y_1) + m(z - Y_1)\frac{\partial^2}{\partial z\partial\theta} F_\theta\big(\xi_m^{\mathrm{SF}}\big)\right) dz,$$

where $\xi_m^{\mathrm{SF}} \subset (Y_1, Y_2) \subset B_{2r_{m-1}}(q_\alpha(\theta))$, see the proof of Lemma A.2 for the definition of $B_{2r_{m-1}}(q_\alpha(\theta))$,

$$\simeq -m\left(\frac{\partial}{\partial\theta} F_\theta(Y_1)\right)(Y_2 - Y_1) - m c_{1,m}^{\mathrm{SF}} (Y_2 - Y_1)^2. \qquad (2.22)$$

The term $c_{1,m}^{\mathrm{SF}} \ge |\frac{\partial^2}{\partial z\partial\theta} F_\theta(y)|/2$, for $y \in B_{2r_{m-1}}(q_\alpha(\theta))$. This is bounded by Assumption **(A5)**.

The derivation for the MVD symmetric estimator follows from the same line of argument applied to the measure-valued derivative expression for the parameter derivative of the distribution function put forward in Equation (2.20). More specifically, we begin with a similar integral as displayed in Equation (2.19), and we again use a two term Taylor series centred at $z = Y_1$ with Lagrangian remainder. This leads to the following expression

$$\begin{aligned}
&\mathbb{E}_\theta\big[D_2^{\mathrm{MVD}}(m)\,\big|\,\mathbf{Z}\backslash Z_m\big]\\
&= -\int_{Y_1}^{Y_2} c_\theta m\left(F_\theta^+(z) - F_\theta^-(z)\right) dz\\
&= -c_\theta \int_{Y_1}^{Y_2} m\left(F_\theta^+(Y_1) - F_\theta^-(Y_1)\right) + m\left(f_\theta^+\left(\xi_m^{\mathrm{MVD}}\right) - f_\theta^-\left(\xi_m^{\mathrm{MVD}}\right)\right)(z - Y_1)\,dz,
\end{aligned}$$

where $\xi_m^{\mathrm{MVD}} \subset (Y_1, Y_2) \subset B_{2r_{m-1}}(q_\alpha(\theta))$,

$$\simeq -m\left(\frac{\partial}{\partial\theta}F_\theta(Y_1)\right)(Y_2 - Y_1) - mc_\theta c_{1,m}^{\mathrm{MVD}}(Y_2 - Y_1)^2. \tag{2.23}$$

The term $c_{1,m}^{\mathrm{MVD}} \geq |f_\theta^+(y) - f_\theta^-(y)|/2$, for all $y \in B_{2r_{m-1}}(q_\alpha(\theta))$. Note that $c_\theta c_{1,m}^{\mathrm{MVD}} = c_{1,m}^{\mathrm{SF}}$.

In the following, we only consider the asymptotic expansion of the symmetric SF estimator in Equation (2.22). For this estimator, we start by determining a Taylor Series to the necessary number of terms. We define,

$$\begin{aligned}
W_1 &:= \frac{\partial}{\partial\theta}F_\theta(Y_1)\\
W_2 &:= Y_2 - Y_1,
\end{aligned}$$

which gives

$$\mathbb{E}_\theta\big[D_2^{\mathrm{SF}}(m)\,\big|\,\mathbf{Z}\backslash Z_m\big] \simeq -mW_1W_2 - mc_{1,m}^{SF}W_2^2. \tag{2.24}$$

We start by expanding the random variable W_1. We use a two term Taylor polynomial, centred at $z = q_\alpha(\theta)$ with Lagrangian form of the remainder. This Taylor polynomial follows the Taylor series presented in [18], pp. 84, where the difference to the cited case is that the authors expand the expected value of an order statistic, whereas we are expanding the conditional expectation w.r.t. a random variable. The resulting polynomial for W_1 is

$$W_1 \simeq \frac{\partial}{\partial\theta}F_\theta(q_\alpha(\theta)) + 2c_{1,m}^{\mathrm{SF}}(Y_1 - q_\alpha(\theta)). \tag{2.25}$$

We now rewrite $Y_1 - q_\alpha(\theta)$ in terms of the standard Uniform distribution with support $(0,1)$. To this end, we let $T_1 = U_{\lceil \alpha m\rceil -1:m-1}$ and $T_2 = U_{\lceil \alpha m\rceil :m-1}$ denote order statistics of the standard Uniform distribution, equivalent to Y_1, Y_2 for $l = 1,2$, $T_l = F_\theta^{-1}(Y_l)$. We arrive at the following expression for $Y_1 - q_\alpha(\theta)$:

$$Y_1 - q_\alpha(\theta) = F_\theta^{-1}(T_1) - F_\theta^{-1}(\alpha). \tag{2.26}$$

As a next step, we approximate the RHS of the above equation with the help of the Mean-Value Theorem. First, note that

$$\frac{\partial}{\partial y} F_\theta^{-1}(y) = \frac{1}{f_\theta(F_\theta^{-1}(y))}.$$

Let

$$c_{2,m} = \sup_{y \in \hat{B}_{t_{m-1}}(q_\alpha(\theta))} \left| \frac{\partial}{\partial y} F_\theta^{-1}(y) \right|,$$

where the set $\hat{B}_{t_{m-1}}(\alpha)$, via the relation $y = F_\theta^{-1}(t)$, $t \in (0,1)$, is the pre-image to the open ball $B_{r_{m-1}}(q_\alpha(\theta))$, given as follows

$$\hat{B}_{t_m}(q_\alpha(\theta)) = \left\{ t : |t - \alpha| \le t_m = 2\frac{(\ln(m))^{\frac{1}{2}}}{m^{\frac{1}{2}}} \right\}. \tag{2.27}$$

Applying the Mean-Value Theorem to the RHS of (2.26) provides

$$\begin{aligned} Y_1 - q_\alpha(\theta) &= F_\theta^{-1}(T_1) - F_\theta^{-1}(\alpha) \\ &\simeq c_{2,m}(T_1 - \alpha), \end{aligned}$$

and we conclude that $Y_1 - q_\alpha(\theta)$ from Lemma A.1 is absolutely bounded from above, hence uniformly integrable.

For the term W_2, we approximate this expression with a three term Taylor series, centred around $z = \alpha$, This is presented in Equation (2.28), where again we apply the Lagrangian remainder:

$$\begin{aligned} W_2 &\simeq \left(F_\theta^{-1}(\alpha) + \left(\frac{\partial}{\partial z} F_\theta^{-1}(\alpha) \right)(T_2 - \alpha) + c_{3,m}(T_2 - \alpha)^2 \right) \\ &\quad - \left(F_\theta^{-1}(\alpha) + \left(\frac{\partial}{\partial z} F_\theta^{-1}(\alpha) \right)(T_1 - \alpha) + c_{3,m}(T_1 - \alpha)^2 \right) \\ &= \frac{1}{f_\theta(q_\alpha(\theta))}(T_2 - T_1) + c_{3,m}(T_2 - T_1)(T_2 + T_1 - 2\alpha), \end{aligned} \tag{2.28}$$

where $c_{3,m}$ is related to the second derivative of the distribution function; i.e.,

$$c_{3,m} = \sup_{y \in \hat{B}_{t_{m-1}}(\alpha)} \left| \frac{\partial^2}{\partial y^2} F_\theta^{-1}(y) \right| = \sup_{y \in B_{r_{m-1}}(\alpha)} \left| - \frac{\frac{\partial}{\partial y} f_\theta(y)}{f_\theta^3(y)} \right|,$$

which is finite by **(A6)**.

We proceed to combine Equations (2.25) and (2.28) into the symmetric SF integral expression, Equation (2.24). The resulting Taylor expansion for both estimators is given below, noting that the definition of the derivative of the quantile Equation (2.8), is the first term of this expansion

$$\begin{aligned} -mW_1W_2 \simeq{} & m\left(\frac{\partial}{\partial\theta} q_\alpha(\theta)\right)(T_2 - T_1) - mc_{3,m}\left(\frac{\partial}{\partial\theta} F_\theta(q_\alpha(\theta))\right)(T_2 - T_1)(T_2 + T_1 - 2\alpha) \\ & - 2mc_{1,m}^{\mathrm{SF}} c_{2,m}\left(\frac{\partial}{\partial z} F_\theta^{-1}(\alpha)\right)(T_2 - T_1)(T_1 - \alpha) \\ & - 2mc_{1,m}^{\mathrm{SF}} c_{2,m} c_{3,m}(T_2 - T_1)(T_1 - \alpha)(T_2 + T_1 - 2\alpha). \end{aligned}$$

For the second term of Equation (2.24), the square of the expansion of Equation (2.28), contains a number of terms for the sub-leading orders

$$\begin{aligned} -mc_{1,m}^{SF} W_2^2 \simeq{} & -mc_{1,m}^{SF}\left(\frac{\partial}{\partial z} F_\theta^{-1}(\alpha)\right)^2 (T_2 - T_1)^2 \\ & - 2mc_{1,m}^{SF} c_{3,m}\left(\frac{\partial}{\partial z} F_\theta^{-1}(\alpha)\right)(T_2 - T_1)^2(T_2 + T_1 - 2\alpha) \\ & - mc_{1,m}^{SF} c_{3,m}^2 (T_2 - T_1)^2 (T_2 + T_1 - 2\alpha)^2. \end{aligned} \tag{2.29}$$

Here the derivatives are not simplified as there is no need to and typographically not preferred. Equation (2.29) is uniformly integrable due to a combination of both appendix Lemmas A.1 and A.2.

We now have the necessary components of the asymptotic expansion for the symmetric SF estimator, Equation (2.22). For the remainder of this proof, we now determine the unconditional expectation of this estimator. This involves the computation of expectations of functions of Uniform $U(0,1)$ order statistics. These computations were conducted via the Maple 14 programming package and are provided in Appendix A.2, where the needed expectations are Equations (A.8) - (A.10), (A.12) - (A.14), and (A.16).

With these expectations, we attain, after some laborious algebraic computations, our final answer for the rate of asymptotic unbiasedness for each of the SF derivative estimators. Recall that the proof for MVD follows from the same line of argument and will lead to the same result: compare (2.22) and (2.23). Denoting $D(m) \in \{D_1^{\mathrm{SF}}(m), D_2^{\mathrm{SF}}(m), D_1^{\mathrm{MVD}}(m), D_2^{\mathrm{MVD}}(m)\}$ as the catch-all term, the asymptotic expansion is written as the following

$$\begin{aligned}
&\mathbb{E}_\theta[D(m)] - \frac{\partial}{\partial\theta} q_\alpha(\theta) = \mathbb{E}[\mathbb{E}_\theta[D(m)|\mathbf{Z}\setminus Z_m]] - \frac{\partial}{\partial\theta} q_\alpha(\theta) \\
&\approx -\frac{1}{m+1}\left(2(1-\alpha)c_{3,m}\left(\frac{\partial}{\partial\theta}F_\theta(q_\alpha(\theta))\right) + 2c_{1,m}^{SF}\left(\frac{\partial}{\partial z}F_\theta^{-1}(\alpha)\right)^2 + 4\alpha(1-\alpha)c_{1,m}^{SF}c_{2,m}c_{3,m}\right. \\
&\quad \left. - 2\alpha c_{1,m}^{SF}c_{2,m}\left(F_\theta^{-1}(\alpha)\right)\right) - \frac{1}{(m+1)(m+2)}\left(16\alpha(1+\alpha)c_{1,m}^{SF}c_{2,m}c_{3,m}\right. \\
&\quad \left. + \left(4(3-4\alpha)c_{1,m}^{SF}c_{3,m}\left(\frac{\partial}{\partial x}F_\theta^{-1}(\alpha)\right) - 8\alpha(1-\alpha)c_{3,m}\right)\right) - \frac{24(1-\alpha)^2 c_{1,m}^{SF}c_{3,m}^2}{(m+1)(m+2)(m+3)}.
\end{aligned}$$

More simply, the expansion above is written as

$$\mathbb{E}_\theta[D(m)] - \frac{\partial}{\partial\theta} q_\alpha(\theta) = O\left(\frac{1}{m}\right),$$

which proves the claim. □

For a statistical analysis of the simulation output, it is desirable to be able to construct a confidence interval for $\partial_\theta q_\alpha(\theta)$. The following result states that this is indeed possible.

Theorem 2.4 (Central Limit Theorem)**.** *Suppose that Assumptions* **(A1)** *and* **(A3)** *to* **(A6)** *are satisfied and suppose we require the limiting behaviour between k and m to be, $k^{1/2}/m \to 0$ as $(m,k) \to (\infty,\infty)$. For $D(m,k) \in \{D_1^{\mathrm{MVD}}(m,k), D_2^{\mathrm{MVD}}(m,k), D_1^{\mathrm{SF}}(m,k), D_2^{\mathrm{SF}}(m,k), D^{\mathrm{IPA}}(m,k)\}$ it holds that*

$$\frac{D(m,k) - \frac{\partial}{\partial\theta} q_\alpha(\theta)}{(\mathrm{Var}(D(m,k)))^{1/2}} \xrightarrow{d} N(0,1), \tag{2.30}$$

as $(m,k) \to (\infty,\infty)$ provided the appropriate version of Assumption **(A2)** *holds as well.*

Proof: Essentially the proof is the same as in [57]. For $l = 1,2$, the LHS of Equation (2.30) can be rewritten as

$$\frac{D(m,k) - \frac{\partial}{\partial\theta} q_\alpha(\theta)}{(\mathrm{Var}(D(m,k)))^{1/2}} = \frac{D(m,k) - \mathbb{E}[D(m,k)]}{(\mathrm{Var}(D(m,k)))^{1/2}} + \frac{\mathbb{E}[D(m,k)] - \frac{\partial}{\partial\theta} q_\alpha(\theta)}{(\mathrm{Var}(D(m,k)))^{1/2}}. \tag{2.31}$$

The first term on the RHS of Equation (2.31) is equal to the standard normal distribution as $(m,k) \to (\infty,\infty)$ due to the Lévy Central Limit Theorem.

For this term, the conditions apply since the $\mathbf{Z}(j)$, $1 \le j \le k$, in $\mathbf{Z}$ are i.i.d. For the MVD derivative estimators, the measure-valued derivative random variable pairs, $(Z_i^+(j), Z_i^-(j))$, $1 \le i \le m$, are i.i.d. and are independent of $\mathbf{Z}$ with possible exception of the random variables in $\mathbf{Z}$ the MVD random variable pair relates to. Hence, once the second term equals zero in the limit, under the required conditions, the proof is complete.

For the second term in Equation (2.31), from Theorem 2.3, $\mathbb{E}[D(m,k)] - \partial_\theta q_\alpha(\theta) = O(1/m)$. In addition, $\mathrm{Var}(D(m,k)) = O(1/k)$. Finiteness is assured by Assumption **(A3)**, and the variance is computed as a sum of i.i.d. replications of estimators. Combining these behaviours for each component of the second term, as $(m,k) \to (\infty,\infty)$, it is possible to compute (for the sake of brevity we omit the details) an upper bound $L < \infty$ such that

$$\left| \frac{\mathbb{E}[D(m,k)] - \frac{\partial}{\partial\theta} q_\alpha(\theta)}{(\mathrm{Var}(D(m,k)))^{1/2}} \right| \le L \frac{k^{\frac{1}{2}}}{m}. \tag{2.32}$$

Given our requirement on the relation between m and k, the LHS in Equation (2.32) tends to zero as $(m,k) \to (\infty,\infty)$. $\square$.

From this result we can deduce our confidence intervals. Again, let the random variable $D(m,k) \in \{D^{\mathrm{IPA}}(m,k), D_1^{\mathrm{MVD}}(m,k), D_2^{\mathrm{MVD}}(m,k), D_1^{\mathrm{SF}}(m,k), D_2^{\mathrm{SF}}(m,k)\}$ denote any of the presented derivative estimators and write $D(m,k) = (1/k)\sum D_i(m))$ with $D_i(m)$ the i^{th} realization of the estimator represented by $D(m,k)$.

Corollary 2.1 (Confidence Interval at level $\beta \in (0,1)$)**.** *Suppose that Assumptions* **(A1)** *to* **(A6)** *are satisfied. Then the symmetric two-sided confidence interval at level* $\beta \in (0,1)$ *for each of the quantile sensitivity estimators,* $D(m,k)$*, is written as*

$$\frac{\partial}{\partial\theta} q_\alpha(\theta) \in \left(D(m,k) - \frac{q_{1-\beta/2}}{\sqrt{k}} \sqrt{\mathrm{Var}(D_1(m))},\, D(m,k) + \frac{q_{1-\beta/2}}{\sqrt{k}} \sqrt{\mathrm{Var}(D_1(m))} \right)$$

where q_γ *is the quantile function for the standard* t*-distribution,* $T(k-1)$*, with* $k-1$ *degrees of freedom at level* $\gamma \in (0,1)$*. By standard methods, we can also construct one-sided intervals.*

To compute the confidence intervals, we provide in the last part of this section a statistic that almost surely converges to the variance of the derivative estimator for a given column.

Lemma 2.2 (Variance Estimator)**.** *Provided that Assumptions* **(A1)** *to* **(A6)** *are satisfied, it holds for each of the derivative estimators* $D(m,k)$ *as* $k \to \infty$ *that*

$$S(m,k) := \frac{1}{k-1} \left(\sum_{i=1}^{k} (D_i(m))^2 - \frac{1}{k} \left(\sum_{i=1}^{k} D_i(m) \right)^2 \right) \overset{a.s.}{\to} \mathrm{Var}(D_1(m)).$$

Proof: We begin with a rearrangement of $S(m,k)$:

$$S(m,k) = \frac{k}{k-1}\left(\frac{1}{k}\sum_{i=1}^{k}(D_i(m))^2 - \left(\frac{1}{k}\sum_{i=1}^{k}D_i(m)\right)^2\right). \tag{2.33}$$

The estimator is finite, for each derivative estimator, by Assumption **(A3)**. By the Strong Law of Large Numbers [96], we attain as $k \to \infty$

$$\frac{1}{k}\sum_{i=1}^{k}(D_i(m))^2 \overset{a.s.}{\to} \mathbb{E}\left[(D_1(m))^2\right]$$

and

$$\frac{1}{k}\sum_{i=1}^{k}D_i(m) = D(m,k) \overset{a.s.}{\to} \mathbb{E}[D_1(m)]. \tag{2.34}$$

Combining Equation (2.34) with the Continuous Mapping Theorem [101], using the map $a(x) = x^2$ for the second term in Equation (2.33), and a Slutsky Theorem [45], we get as $k \to \infty$,

$$S(m,k) \overset{a.s.}{\to} \mathbb{E}\left[(D_1(m))^2\right] - \mathbb{E}^2[D_1(m)] = \mathrm{Var}(D_1(m)).$$

as required. □.

2.4.2 Examples

In this section we present numerical examples to illustrate the performance of the derivative estimators for the QS problem.

It is worth noting that the computation of the symmetric MVD estimator require sorting each sample $\mathbf{Z}(j)$, $1 \le j \le k$, after the substitution of a realization of Z_m in each sample by a realization of Z_m^+ and Z_m^-, respectively. A naïve implementation of the standard MVD derivative estimators would require to run a sort algorithm $2m$ times, which would cause a large computational burden. To circumvent this problem for the standard estimator, we elaborate on Equation (2.18) for the evaluation of $D_1(m)$. More specifically, the values for Y_1, Y_2 can be attained by ordering $\mathbf{Z}\backslash Z_m$, and inferring the result for both order statistics, each being one of three possible outcomes[3]. These pairs depend on whether the rank $l = 1,\dots,m$ is less than, equal to, or greater to $\lceil \alpha m \rceil$. We then keep the order sample $\mathbf{Z}\backslash Z_m$ and compute Equation (2.18) for m independent realizations of Z_m^+ and Z_m^-, respectively.

[3]There are only two possible outcomes for each order statistic for the extreme ranks $\lceil \alpha m \rceil = 1, m$.

2.4.2.1 A PERT Graph

A Stochastic Activity Network (SAN) is a weighted, directed, acyclic graph in which the following two conditions hold:

1. the activity times (edges), representing times, are fixed or stochastic in nature and are strictly positive, and

2. the incoming activities to each event (node) represent a set of precedence conditions in which each activity must be completed before the event occurs.

In the following we will discuss an application of SAN for the Project Evaluation Review Technique (PERT). Here, the SAN is an abstraction of the stages or intermediate steps (nodes) during the course of the project and the tasks that are required by the different groups of people to get to each stage (edges). The weight of each edge represents the amount of time a group requires to complete the task. This weight can be deterministic or stochastic. The precedence conditions are merely the completion of the tasks set towards each group at a particular stage. The PERT graphs have one source and one sink node, representing the beginning and end of the project. A more detailed discussion on Stochastic Activity Networks and PERT in particular are given in [21], [22], [84].

We use a simple graph used before in [27] and [54]. The graph is presented in Figure 2.3, consisting of five nodes and six edges. Node one is the source; node six in the sink. In this graph, there are three paths, denoted by $\zeta_1 = (1,4,6)$,

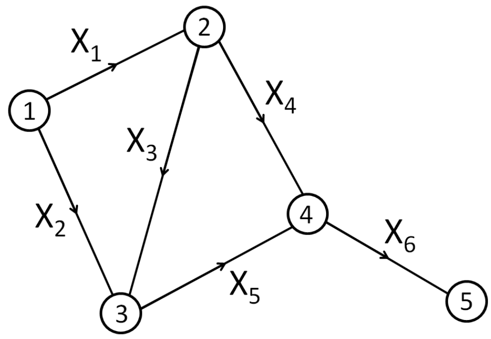

Figure 2.3: Graph of the Stochastic Activity Network (SAN).

$\zeta_2 = (1,3,5,6)$, and $\zeta_3 = (2,5,6)$. The weights of the edges are denoted by X_i, for

$1 \le i \le 6$, where X_i represents the duration of the execution of task i. The duration time for each path, denoted by $Z = h(X, \zeta_i)$, is the sum of the weights of the travelled edges along that path. Due to the precedence conditions, the completion time of the project is given by the path with the maximal weight. Specifically, let ζ^* denote this maximal path, also called the *critical path*, then

$$Z = h(X, \zeta^*) = \begin{cases} X_1 + X_4 + X_6 & \text{if } \zeta^* = \zeta_1, \\ X_1 + X_3 + X_5 + X_6 & \text{if } \zeta^* = \zeta_2, \\ X_2 + X_5 + X_6 & \text{if } \zeta^* = \zeta_3. \end{cases}$$

For each of these activities, the random variables X_i are distributed according to a combination of a fixed component, denoted by t_i, and an exponential random variable with rate λ_i, denoted by Y_i, i.e., $X_i = t_i + Y_i$. In terms of a PERT network, the choice of distribution for the random variables, X_i, $1 \le i \le 6$, represents that the completion of the task requires a certain minimal amount of time as well as an uncertain component, inherent when forecasting the completion time for a set of tasks. The exponential shape comes from a memoryless consideration for the amount of time a set of tasks needs to be completed given the extent of completion of these tasks.

In this example we are interested in determining the derivative of the quantile q_α of the SAN's completion time with regard to the parameters λ_4 and λ_5, related to durations X_4 and X_5. The duration time X_4 is only present on one path, ζ_1, whilst the duration time X_5 is present on paths ζ_2, ζ_3. As the duration times are stochastic, each of the three paths is critical with positive probability and so the distribution of Z is an amalgam of sums of exponential random variables: not easily known and the quantile not analytically derivable. In order to apply our quantile derivative results developed in this chapter, we considered an i.i.d. sequence $(Z_i : 1 \le i \le m)$, where Z_i represents the i^{th} sample of the duration time of the SAN. Since each path is effectively a sum of exponentially distributed random variables, Assumptions **(A1)** to **(A6)** hold.

To derive the sensitivity estimator from either the IPA or SF method, the methods require knowledge on the critical path. Following Section 1.2.4.1, the critical path dependent IPA derivative estimator w.r.t. the parameter λ_4 is

$$\frac{\partial Z}{\partial \lambda_4} = \begin{cases} -\frac{X_4 - t_4}{\lambda_4} & \text{if } \zeta^* = \zeta_1, \\ 0 & \text{if } \zeta^* = \zeta_2 \text{ or } \zeta_3, \end{cases}$$

and w.r.t. the parameter λ_5

$$\frac{\partial Z}{\partial \lambda_5} = \begin{cases} -\frac{X_5 - t_5}{\lambda_5} & \text{if } \zeta^* = \zeta_2 \text{ or } \zeta_3, \\ 0 & \text{if } \zeta^* = \zeta_1. \end{cases}$$

For the score function, if $\zeta^* = \zeta_1$

$$\mathrm{SF}_{\lambda_4}(x) = \frac{1}{\lambda_4}(1 - \lambda_4(x_4 - t_4))$$

and zero otherwise. While for $\zeta^* = \zeta_2, \zeta_3$

$$\mathrm{SF}_{\lambda_5}(x) = \frac{1}{\lambda_5}(1 - \lambda_5(x_5 - t_5)),$$

and zero if $\zeta^* = \zeta_1$.

Conversely, the MVD sensitivity is determined following the evaluation of the completion time due perturbed weights introduced by the measure-valued derivative distributions. The original critical path has no role in evaluating the derivative. Specifically, let $h(x_1, x_2, \ldots, x_6)$ be the weight of the critical path for weights $X_i = x_i$, $1 \leq i \leq 6$; i.e., $Z = h(X_1, X_2, \ldots, X_6)$. Let Y_i be exponentially distributed with rate λ_i, $i = 4, 5$, independent of everything else.

Let $v(x) = 1 + x$, then X_i, for $1 \leq i \leq 6$, is $\mathfrak{B}_v$-differentiable according to Section 1.2.4.1. Since

$$\mathbb{E}[Z|X_i = x] = \mathbb{E}[h(X_1, X_2, \ldots, X_6)|X_i = x], \quad 1 \leq i \leq 6,$$

as a mapping in x is an element of $\mathfrak{B}_v$, and we can apply Lemma 2.1 with $\mathfrak{D} = \hat{\mathfrak{D}} = \mathfrak{B}_v$. Elaborating on the presentation of the weak derivative of the exponential distribution put forward in Section 1.2.4.1, yields

$$\frac{\partial Z}{\partial \lambda_4} = c_{\lambda_4}\left(Z - Z_4^-\right)$$

where

$$Z_4^- = h(X_1, X_2, X_3, X_4 + Y_4, X_5, X_6) \quad \text{and} \quad c_{\lambda_4} = \frac{1}{\lambda_4},$$

and

$$\frac{\partial Z}{\partial \lambda_5} = c_{\lambda_5}\left(Z - Z_5^-\right)$$

where

$$Z_5^- = h(X_1, X_2, X_3, X_4, X_5 + Y_5, X_6) \quad \text{and} \quad c_{\lambda_5} = \frac{1}{\lambda_5}.$$

For the PERT Graph, the probability levels are, $\alpha = 0.50, 0.90, 0.95$. For the completion times of the set of tasks representing each edge, $X_i = t_i + Y_i$, with

Y_i being an exponential random variable with rate λ_i) independent of everything else, the rates for each exponentially distributed random are given by $\lambda = (\lambda_1, \ldots, \lambda_6) = (2/3, 1/2, 1/2, 1/4, 2/3, 1)$, and the minimal completion times are $t = (4.5, 6, 3, 4, 3.5, 6)$. The mean time of completion for each edge is equal to $(\mu_1, \ldots, \mu_6) = (6, 8, 5, 8, 5, 7)$, and the consequent mean time of completion for each path is $(\zeta_1, \zeta_2, \zeta_3) = (21, 23, 20)$. The choice of the parameters is to ensure that each of the paths is critical with positive probability. By extensive simulation, we determined the probability of each path being critical as $(\pi_1, \pi_2, \pi_3) = (0.2455, 0.6476, 0.1069)$. The sensitivity estimates are compared to a Finite Difference (FD) estimate attained from brute-force simulation.

The numerical examples are presented in Table 2.4. RMS, std. dev. , WNV and cov. prob. are determined over 500 replications.

In Figure 2.4 the RMS error is presented for $\alpha = 0.90$. The number of simulated data per estimate is given by m, and the number of estimates is denoted by k. Results are given in terms of the number of simulated data $m \times k$. As the results show, IPA has larger RMS error than the MVD estimators for $m = k^{\frac{1}{2}}$.

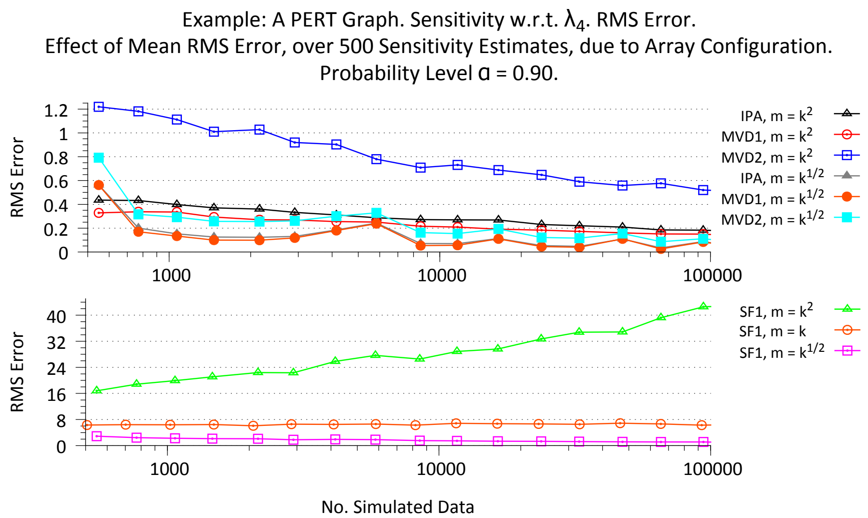

Figure 2.4: The effect of the RMS error for PERT Graph example.

2.4.2.2 A Queueing Network Example

Consider a simple stationary queuing network, with Poisson-λ arrival stream, consisting of three queues with infinite buffer capacity. Jobs arrive from the outside to station 1 and are being served with exponential service rate μ_1. From

Derivative w.r.t. λ_4: $n = 2^{16}$, with $m = 1598$, $k = 41$, $\alpha = 0.90$. FD Value = -1.778×10^1

	IPA	SF1	SF2	MVD1	MVD2
Mean	-1.782×10^1	1.615×10^1	8.990×10^2	-1.774×10^1	-1.770×10^1
RMS Error	0.20250	4.144×10^1	1.631×10^3	0.1556	0.5820
Std. Dev.	3.6041	7.368×10^2	2.901×10^4	2.7701	1.036×10^1
WNV	7.021×10^{-4}	2.475×10^2	9.366×10^4	0.1138	0.1070
Cov. Prob.	0.9460	0.9360	0.9120	0.9320	0.8160

Derivative w.r.t. λ_4: $n = 2^{16}$, with $m = 41$, $k = 1598$, $\alpha = 0.90$. FD Value = -1.778×10^1

	IPA	SF1	SF2	MVD1	MVD2
Mean	-1.737×10^1	-1.646×10^1	-1.339×10^1	-1.739×10^1	-1.743×10^1
RMS Error	4.017×10^{-2}	1.0677	6.7034	3.316×10^{-2}	8.409×10^{-2}
Std. Dev.	0.5873	1.896×10^1	1.192×10^2	0.4434	1.4543
WNV	5.038×10^{-5}	0.1674	4.3635	2.912×10^{-3}	3.1543×10^{-3}
Cov. Prob.	0.8720	0.9420	0.9500	0.8660	0.9400

Derivative w.r.t. λ_5: $n = 2^{16}$, with $m = 1598$, $k = 41$, $\alpha = 0.90$. FD Value = -2.6770

	IPA	SF1	SF2	MVD1	MVD2
Mean	-2.6743	-1.706×10^1	-8.900×10^2	-2.6618	-2.6064
RMS Error	0.1897	1.006×10^2	4.287×10^3	0.1557	0.9323
Std. Dev.	0.5084	2.691×10^2	1.145×10^3	0.4169	2.4973
WNV	1.441×10^{-5}	3.312×10^1	1.458×10^4	2.604×10^{-3}	6.133×10^{-3}
Cov. Prob.	0.9440	0.9400	0.9060	0.9280	0.6800

Derivative w.r.t. λ_5: $n = 2^{16}$, with $m = 41$, $k = 1598$, $\alpha = 0.90$. FD Value = -2.6770

	IPA	SF1	SF2	MVD1	MVD2
Mean	-2.6191	-2.9748	-1.6988	-2.6159	-2.6316
RMS Error	3.916×10^{-2}	2.5766	1.703×10^1	3.439×10^{-2}	0.1172
Std. Dev.	8.747×10^{-2}	6.8982	4.562×10^1	6.890×10^{-2}	0.3107
WNV	1.363×10^{-6}	2.220×10^{-2}	0.6387	7.236×10^{-5}	1.476×10^{-4}
Cov. Prob.	0.9460	0.9440	0.9600	0.9580	0.9420

Table 2.4: The summary statistics for the five derivative estimators for the PERT Graph example.

station 1 they continue with probability p to station 2 where they are served with exponential service rate μ_2 or, alternatively, they continue to station 3 with probability $1-p$ where they are served with exponential service rate μ_3. The jobs leave the system once the service at station 2, respectively, at station 3 is completed. We assume that the queueing network is stable; i.e., $\lambda < \min(\mu_1, \mu_2/p, \mu_3/(1-p))$.

For $1 \leq i \leq 3$, let X_i denote exponential random variables with the following means: X_1 has mean $1/(\mu_1 - \lambda)$, X_2 has mean $1/(\mu_2 - \lambda p)$, whilst the mean of random variable X_3 is $1/(\mu_3 - \lambda(1-p))$. In addition, let $\delta(p)$ be the random variable indicating the Bernoulli routing of the jobs from the first station, i.e., $\delta(p) = 2$ with probability p and $\delta(p) = 3$ with probability $1-p$. By Burke's theorem [11], it is possible to compute the distribution of the departure process at the first station as a Poisson-λ process. Correspondingly, the arrival process at station 2, respectively, station 3 are Poisson-λp, respectively, Poisson-$\lambda(1-p)$ processes. Additionally, since this network satisfies the "overtake-free“ condition[4] , from Theorem 2.2 in [9], the sojourn times at each station are independent with the distribution function specified by the respective random variables. The sojourn time of a job; i.e., the total time elapsed between entering and leaving the system, is then the sum of the sojourn times of each station the job enters into and in distribution is equal to

$$X_1 + X_{\delta(p)}. \tag{2.35}$$

And the expected stationary sojourn time is given by

$$\frac{1}{\mu_1 - \lambda} + \frac{p}{\mu_2 - \lambda p} + \frac{1-p}{\mu_3 - \lambda(1-p)}.$$

Let $q_\alpha(\theta)$ denote the α-quantile of the stationary sojourn time. In the following we apply our estimators to estimate the sensitivity of $q_\alpha(\theta)$ with respect to the first service rate $\theta = \mu_1$. Even though it is possible to attain a closed expression for the distribution of the stationary sojourn time, a closed analytical expression cannot be attained for the α-quantile and hence has to be solved numerically. Also, from the above expression for the mean sojourn times it can be seen that the sensitivity of the mean stationary sojourn time with respect to μ_1 can be easily computed.

To compute each quantile sensitivity, we have our i.i.d. sequence of m random variables, where Z_i is a sample of the stationary sojourn time as provided in (2.35). As in the Stochastic Activity Network, Assumptions **(A1)** to **(A6)** are satisfied. For the derivative estimators, we consider the overall sojourn time to be the sum of observable sojourn times through the stations; i.e, we write

$$Z_i = h(X_1, X_2, X_3, \delta(p))$$

with $h(x_1, x_2, x_3, d) = x_1 + x_2 1\{d = 2\} + x_3 1\{d = 3\}$. Moreover, from Section 1.2.4.1, we attain

$$\frac{\partial}{\partial \mu_1} Z_i = \frac{\partial}{\partial \mu_1} h(X_1, X_2, X_3, \delta(p)) = -\frac{1}{\mu_1 - \lambda} X_1$$

[4]because for any two jobs that both arrive at either station 2 or 3, the job that arrives first at station 1 will arrive first at the subsequent station.

for IPA, and the score function is given by

$$\mathrm{SF}_{\mu_1}(x_1, x_2, x_3, d) = \frac{1}{\mu_1 - \lambda} - x_1.$$

For the MVD estimator, let $v(x) = 1+x$, then X_i, for $1 \le i \le 3$, is $\mathfrak{B}_v$-differentiable according to Section 1.2.4.1. Since, for $1 \le i \le 3$,

$$\mathbb{E}[Z|X_i = x] = \mathbb{E}[h(X_1, X_2, X_3, \delta(p))|X_i = x]$$

as a mapping in x is an element of $\mathfrak{B}_v$, by Lemma 2.1, with $\mathfrak{D} = \hat{\mathfrak{D}} = \mathfrak{B}_v$, elaborating on the presentation of the weak derivative of the exponential distribution put forward in Section 1.2.4.1, yields the measure-valued derivative triple:

$$Z^+ = Z \quad \text{and} \quad Z^- = h(X_1 + Y, X_2, X_3, \delta(p)) = Z + Y,$$

and $c_{\mu_1} = 1/(\mu_1 - \lambda)$. The random variable Y is a independent copy of an exponential random variable X_1 with rate $\mu_1 - \lambda$.

Remark 2.3. *Note that both the IPA and the Score Function estimator require that the individual sojourn times at stations 1, 2 and 3 are observable. This is in contrast to the MVD estimator that only requires that the realized sojourn time is observable.*

In this example, we choose the probability levels to be $\alpha = 0.50, 0.95$. The arrival rate is set to $\lambda = 1$, and after completing service at the first station, jobs are routed to the second station with probability $p = 3/4$ and the third station with probability $1 - p = 1/4$. The service rates are $\mu_1 = 5/4$, $\mu_2 = 8/5$, $\mu_3 = 14/15$. The actual value of the quantile derivative is computed analytically via Equation (2.8) using a root finding algorithm.

The numerical results are presented in Table 2.5. The $n = 2^{12}$, respectively $n = 2^{18}$, observations are arranged in a $m \times k$ array in which the estimate is a sample average of $k = 64$ estimates each containing $m = 64$ observations and $k = m = 256$ observations for the second case. As with the other examples the statistics, RMS error, standard deviation (Std. Dev.), work-normalized variance (WNV), and 95% coverage probability (Cov. Prob.) are determined over 500 independent replications.

Figure 2.5 presents the coverage probabilities. Results are given in terms of the number of simulated data $n = m \times k$. Only the IPA and both MVD derivative estimates are given and the probability level is chosen to be $\alpha = 0.90$.

$n = 2^{12}$, $m = k$, $\alpha = 0.50$. Actual Value $= -1.120 \times 10^1$

	IPA	SF1	SF2	MVD1	MVD2
Mean	-1.102×10^1	-1.042×10^1	-2.097×10^1	-1.115×10^1	-1.120×10^1
RMS Error	5.413×10^{-2}	1.5482	1.200×10^1	0.1237	0.2934
Std. Dev.	0.5809	1.734×10^1	1.341×10^2	1.3855	3.2881
WNV	1.864×10^{-5}	8.877×10^{-3}	1.4820	1.146×10^{-3}	2.937×10^{-3}
Cov. Prob.	0.9360	0.9360	0.9360	0.9320	0.9080

$n = 2^{12}$, $m = k$, $\alpha = 0.95$. Actual Value $= -4.664 \times 10^1$

	IPA	SF1	SF2	MVD1	MVD2
Mean	-4.532×10^1	-4.423×10^1	-3.596×10^1	-4.506×10^1	-4.435×10^1
RMS Error	4.117×10^{-2}	1.1990	9.4165	0.1162	0.3850
Std. Dev.	1.3953	5.593×10^1	4.395×10^2	5.1896	1.783×10^1
WNV	9.590×10^{-5}	8.759×10^{-2}	1.549×10^1	2.617×10^{-2}	8.628×10^{-2}
Cov. Prob.	0.8320	0.9420	0.9260	0.9140	0.8620

$n = 2^{18}$, $m = k$, $\alpha = 0.50$. Actual Value $= -1.120 \times 10^1$

	IPA	SF1	SF2	MVD1	MVD2
Mean	-1.117×10^1	-1.121×10^1	-2.645×10^1	-1.117×10^1	-1.118×10^1
RMS Error	1.686×10^{-2}	1.4829	3.551×10^1	4.468×10^{-2}	9.695×10^{-2}
Std. Dev.	0.1865	1.662×10^1	3.978×10^2	0.5003	1.0865
WNV	4.000×10^{-6}	7.924×10^{-2}	2.869×10^1	9.239×10^{-3}	3.688×10^{-3}
Cov. Prob.	0.9580	0.9460	0.9340	0.9420	0.9600

$n = 2^{18}$, $m = k$, $\alpha = 0.95$. Actual Value $= -4.673 \times 10^1$

	IPA	SF1	SF2	MVD1	MVD2
Mean	-4.673×10^1	-4.664×10^1	1.771×10^1	-4.657×10^1	-4.637×10^1
RMS Error	8.188×10^{-3}	1.1742	2.638×10^1	4.094×10^2	0.1466
Std. Dev.	0.3339	5.482×10^1	1.230×10^3	1.9102	6.8393
WNV	8.567×10^{-6}	0.8038	2.614×10^2	0.3214	0.1515
Cov. Prob.	0.9540	0.9440	0.9400	0.9660	0.9420

Table 2.5: The summary statistics for the five derivative estimators for a Queueing Network Example. Sensitivities are with respect to the service rate of the first queue.

2.5 Guidelines for Derivative Estimation of Order Statistic Related Performance Characteristics

The IPA estimator from [57] is the preferred approach having the least standard deviation of the three techniques, being computationally fast and relatively easy to implement (to program the IPA estimator is almost as simple as to program

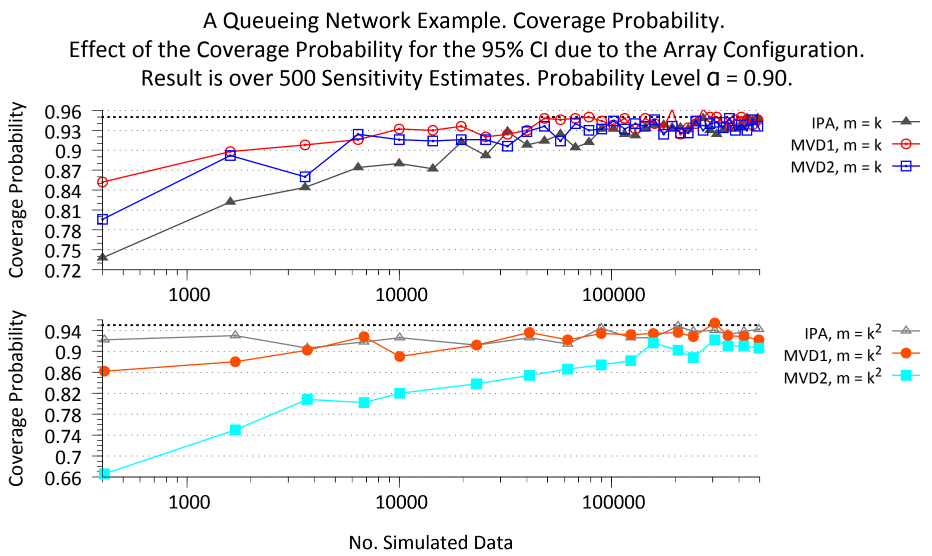

Figure 2.5: The change in coverage probabilities due to a change in the array configuration for a Queueing Network Example.

the derivative $Z_1(\theta)$). Of the distributional methods, the MVD derivative estimators are the only choice. Either the standard or symmetric estimator is a viable choice with the work-normalized variance for both estimators are commensurate to each other. Generally speaking, the standard derivative estimator has the better standard deviation and results relating to the coverage probability, whereas the symmetric estimator is quicker and easier to compute. The SF estimators are not viable choices due to their high standard deviation. The relative speed in computing the SF derivative estimators does not compensate for this defect.

The numerical properties of the estimators can best be understood via the strong consistency proofs. From Equation (2.11), the IPA estimator is determining the parameter derivative of the sample performance falling between two order statistics. These order statistics will not vary too much in the estimate of $D^{\text{IPA}}(m)$, being, from Lemma A.1, away from the quantile of $O(\ln^{1/2}(m)/m^{1/2})$, and from each other by the same extent. For the MVD derivative estimators, particularly for the symmetric estimator, the problem lies in Equation (2.21) where the limiting spacing result for $D_2^{\text{MVD}}(m)$ is an exponential random variable, implying significant variability. The symmetric MVD estimator is at its heart a mean of the difference of two estimators. Though the absolute bound on the difference is a spacing, for each estimate the size of the spacing differs. And for the symmetric derivative estimator, the number of estimates $D_2^{\mathrm{MVD}}(m)$ that have a positive value, a negative value, or no value differs for each realization. The standard MVD derivative estimator is in effect providing an additional sample average within $D_1^{\mathrm{MVD}}(m)$ through the successive single substitutions of the m random variables.

In terms of work-normalized variance, the IPA estimator is the preferred choice when applicable. For example, in the PERT Graph example, Table 2.4, the work-normalized variance is 60-200 times smaller than that of the standard MVD derivative estimator. For the coverage probability, we only consider IPA and both MVD derivative estimators due to the poor performance of the Score Function Method. The coverage probability plots, either for the order statistic sensitivity, Figure 2.2, or for the quantile sensitivity, Figure 2.5, informs us that we need a sufficient number of estimates for the coverage probability to approach a value near 0.95. It is best to choose the sample size to be at least as large as the number of estimates, $m \geq k$. For IPA, the data suggests to choose $m = k^2$. For fixed m, increasing the number of estimates degrades the coverage probability. Though not presented, in the quantile sensitivity case, the coverage probability for $m = k^{1/2}$ degrades towards zero, which is not surprising given the Central Limit Theorem. The symmetric MVD estimator is best in this case by our inference to the greater standard deviation of this approach.

An interesting observation is that the work-normalized variance makes the two MVD derivative estimators competitive. The symmetric derivative estimator, with only one substitution of an MVD random variable for each estimate of $D_1^{\mathrm{MVD}}(m)$ is quicker (and easier) to compute than the standard version and compensates for the deficiency in the standard deviation. In the examples all values for the work-normalized variance for both standard and symmetric estimators are comparable. For the most part the standard estimator bested the symmetric derivative estimator with the exception of the Queueing Network Example, Table 2.5, where the work-normalized variance of the symmetric estimator is a factor of two less. The Score Function derivative estimators are faster and easier to compute than either of the MVD versions though the improvement of time in attaining an estimate does not nearly compensate for their ill performance.

The main challenge for IPA is that Lipschitz continuity often fails to hold or is hard to check. See, for example, the portfolio credit risk model in [32], where a conditioning approach is proposed in order to come up with a weaker set of assumptions with respect to Lipschitz continuity. While this approach allows to deal with non-Lipschitz continuous mappings h, it requires a careful analysis of the model conditioned on each of the discontinuities of h. This is in contrast to the MVD estimator that can be attained in a straightforward manner, see Section 2.3.2.2. Apart from the fact that conditioning via SPA becomes infeasible for complex models, it requires in mathematical terms that h in $Z = h(X)$ is invertible. A simple example were h fails to be invertible is provided in [59]. An extreme example of a model where pathwise differentiability itself is an issue, is the basket of stock example, where the sample path derivative fails to exist with probability one for any t.

To summarize, IPA is the preferred method, when applicable. In case that IPA is not applicable, one of the MVD estimators should be used with a preference for the symmetric estimator due to simplicity of implementation. Moreover, an MVD estimator is preferable when one is interested in a robust quantile sensitivity estimator that can be easily adapted to different choices of performance function h.

Conclusion

In this chapter we have provided a unified analysis of all three main gradient estimation approaches. We have developed the theory for the derivative estimation of ranked-data over a finite data collection and for quantiles. From our numerical comparison of the derivative estimators we conclude that IPA is the preferred method, when applicable, otherwise one of the presented MVD estimators should be used with a preference for the symmetric estimator due to ease of implementation. Moreover, an MVD estimator is preferable when one is interested in a robust quantile sensitivity estimator that can be easily adapted to different choices of performance function. In addition, a fast and efficient implementation of the standard MVD estimator has been provided.