MONTE CARLO COMPUTATION OF CONDITIONAL EXPECTATION QUANTILES

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ENGINEERING-ECONOMIC SYSTEMS AND OPERATIONS RESEARCH

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

OPERATIONS RESEARCH

By

Shing-Hoi Lee

August 1998

UMI Number: 9908809

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Peter Glynn
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Donald L. Iglehart

Prof. Donald Iglehart

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Nicholas Bambos

Approved for the University Committee on Graduate Studies:

Thomas Wasow

Abstract

In this thesis, we examine different ways of numerically computing the distribution function and the quantiles of conditional expectations. Both the conditional expectation and the distribution function itself are computed via Monte Carlo simulation.

Given a limited (and fixed) computer budget, the quality of the estimator is gauged by the inverse of its mean square error. It is a function of the fraction of the budget allocated to estimating the conditional expectation versus the amount of sampling done relative to the "conditioning variable".

We will present the asymptotically optimal rates of convergence for different estimators and resolve the trade-off between the bias and variance of the estimators. Moreover, central limit theorems are established for the estimators proposed. We will also provide algorithms for the practical implementation of the estimators and illustrate how confidence intervals can be formed in each case.

Major areas of application of the computation of the quantiles of conditional expectations include Bayesian estimation and the calculation of Earnings at Risk (EaR) in the field of mathematical finance.

Acknowledgements

I wish to take this opportunity to acknowledge the many people to whom I am indebted for their invaluable assistance during the course of my Ph.D. study here at Stanford.

First, I wish to thank my supervisor, Prof. Peter Glynn, for his expert guidance, highly incredible insights and the large amount of time spent on careful scrutinizing of the draft copies of this thesis and all my writings. Without his inexhaustive support and encouragement I could not have possibly made it through this thesis. He has definitely set forth a perfect example of a responsible and respectful supervisor, to whom I truly owe a lot.

Many thanks to Prof. Donald Iglehart and Prof. Nicholas Bambos for reading my thesis and to Prof. Tze-Leung Lai and Prof. Sam Chiu for rounding out the orals committee. Thanks to all the professors who have taught me during these four years in Stanford University. I am deeply indebted to Prof. Darrell Duffie, Prof. George Papanicolau, Prof. Walter Murray, and Prof. Michael Sauders for their strengthening help and invaluable advice.

I would also like to express my thanks to the staff of the Operations Research Department, especially Ms. Kitt Martinez, Ms. Maggie Barstow-Taylor, Ms. Gretchen Lutz Spalinger, and Ms. Lorrie Papadakis. Sincere thanks are due to my fellow postgraduate friends; they made university life so much more enjoyable and memorable. In addition, financial support in the form of fellowship and assistantship, has been greatly appreciated.

I wish to thank my family for their endurance, support, and never-ending encouragement, to whom I am eternally indebted. Finally, I would like to thank my fiáncé, Lili Kong, for her constant encouragement, patience, and her sense of humor. Lili, you have made my time here in America both colorful and enjoyable.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Summary of Results

1.1 Introduction and Problem Motivation

In this thesis, we examine different ways of numerically computing the distribution function and quantiles of conditional expectations. Recall that for two random variables X and Z, the conditional expectation $\mathbb{E}(X|Z)$, as a measurable function of Z, is also a random variable. For the sake of ease of exposition, we will call the random variable (r.v.) X the *target variable* and Z the *conditioning random element*.

We would like to estimate its distribution function, i.e. $\mathbb{P}(\mathbb{E}(X|Z) \leq x)$. Another closely related problem is to find its quantiles. In particular, for a given $0 < \alpha < 1$, we would also like to find the α-quantile, q_α, such that $\mathbb{P}(\mathbb{E}(X|Z) \leq q_\alpha) = \alpha$.

In a more general setting, the conditioning random element need not be a random variable; it could be a vector of r.v. or even a sample path of a stochastic process. We will delve into the details in Chapter 3.

Typically, the above two problems cannot be solved analytically in "closed form". We can, however, estimate them by numerical methods. One way to do it is by Monte Carlo simulation. In particular, we will use Monte Carlo simulation to estimate both the conditional expectation and its distribution function.

Given a limited (and fixed) computer budget, the goodness of the estimator is gauged by the inverse of its mean square error. It is a function of the fraction of the budget allocated to estimating the conditional expectation versus the amount of sampling done relative to the conditioning random element.

In this thesis, we will present the asymptotically optimal rates of convergence for different estimators and resolve the trade-off between the bias and the variance of the estimators. Moreover, central limit theorems (CLT) are derived for these estimators, which will facilitate the formation of confidence intervals for the estimators and the estimation of asymptotically optimal sample sizes associated with each estimator.

The computation of the quantiles of conditional expectations include the calculation of Earnings at Risk (EaR) in the field of mathematical finance and risk management. Related problems also arise in the setting of Bayesian statistical analysis.

1.1.1 Earnings at Risk

EaR is a measure of the downside risk of a portfolio at a fixed time horizon, τ. Specifically, for a prespecified confidence level α, EaR tells us the α-quantile of the value of the portfolio at time τ. Knowledge of the EaR helps to guide financial decision makers in selecting appropriate investment strategies over the corresponding period.

The setting we have adopted is as follows: We take as given an adapted short-rate process r, with $\int_0^\infty |r_t|\, dt < \infty$ almost surely, and an Itô security price process S in $\mathbb{R}^N$ with

$$dS_t = \mu_t\, dt + \sigma_t\, dB_t.$$

for appropriate μ and σ. Here $(B_t : 0 \le t \le T)$ is a d-dimensional standard Brownian motion.

Aside from technical conditions, the absence of arbitrage is equivalent to the existence of a probability measure Q with special properties, called an equivalent martingale measure, EMM; see [13]. Under Q, all expected rates of return are equivalent to the riskless rate r.

By the definition of an EMM, any random variable that has finite variance with respect to $\mathbb{P}$ has finite expectation with respect to Q. Moreover, for any such random variable Z,

$$\mathbb{E}_t^Q(Z) = \frac{1}{\xi_t}\mathbb{E}_t(\xi_T Z),$$

where $\mathbb{E}_t(\cdot)$ and $\mathbb{E}_t^Q(\cdot)$ are the conditional expectation of $(\cdot)$ given events corresponding to the information available at time t, under the original probability measure $\mathbb{P}$ and under the new measure Q resp. Here,

$$\xi_t = \exp\left(-\int_0^t \eta_s\,dB_s - \frac{1}{2}\int_0^t \eta_s\cdot\eta_s\,ds,\right)$$

and where η is a market price of risk process, that is, an adapted process in $\mathbb{R}^d$ solving the family of linear equations

$$\sigma_t\eta_t = \mu_t - r_t S_t, \quad t\in[0,T].$$

In fact, ξ_T is equal to the Radon Nikodym derivative, $dQ/d\mathbb{P}$.

By Girsanov's theorem, under Q, there is a standard Brownian motion $\hat{B}$ in $\mathbb{R}^d$ such that if the given securities pay no dividends before T, then

$$dS_t = r_t S_t\,dt + \sigma_t\,d\hat{B}_t.$$

More generally, suppose the securities with price process S are claims to a cumulative dividend process D. In this case, we have for any $t<T$,

$$S_t = \mathbb{E}_t^Q\left[\exp\left(\int_t^T -r_s\,ds\right)S_T + \int_t^T \exp\left(\int_t^s -r_u\,du\right)\,dD_s\right]. \tag{1.1}$$

It has to be emphasized that the above equation determines how to price the security in an arbitrage-free market at time t.

Let θ be an adapted process, denoting the trading strategy. Then, at each state ω and time t, the vector $\theta_t(\omega)$ specifies the number of units of the securities to hold. Hence, at time t, the value of the portfolio V_t is equal to $\theta_t\cdot S_t$.

We assume that the securities in the portfolio consist of equities and fixed income derivatives at time τ. This includes non-convertible bonds, swaps, futures, and various kinds of European options, which in turn include Asian options, lookback options etc. As a matter of fact, all these securities can be priced from EMM (see for example, [13, 16].) However, in this thesis, we exclude derivatives that allow holders to exercise prior to their expiration dates, i.e. American derivatives. The reasons are that though approximate numerical methods are available to price "American" derivatives, they typically take longer to price and the analysis of their complexities are comparatively more complicated[1].

Given a prescribed quantile level α and a fixed time horizon τ, EaR is the number q_α such that

$$\begin{aligned}
\alpha &= \mathbb{P}(V_\tau \leq q_\alpha) \\
&= \mathbb{P}\left(\theta_\tau \cdot \mathbb{E}^Q_\tau\left[\exp\left(\int_\tau^T -r_s\,ds\right) S_T + \int_\tau^T \exp\left(\int_\tau^s -r_u\,du\right)\,dD_s\right] \leq q_\alpha\right) \\
&= \mathbb{P}\left(\mathbb{E}_\tau\left[\frac{\theta_\tau \eta_T}{\eta_\tau} \cdot \left[\exp\left(\int_\tau^T -r_s\,ds\right) S_T + \int_\tau^T \exp\left(\int_\tau^s -r_u\,du\right)\,dD_s\right]\right] \leq q_\alpha\right).
\end{aligned}$$

The conditioning random element in this case is the sample path from time 0 to time τ. For Markovian models, the conditioning random element is simply the price at time τ. It is clear that EaR as defined above has exactly the form of our quantile estimation problem. Below, we'll consider two examples illustrating the idea of EaR.

Example 1. Consider a portfolio that consists of a long position of one forward contract on a security, made at time 0. The forward contract is a commitment to pay an amount F (the forward price), which is agreed upon at time 0 and paid at time T, in return for the amount S_T at time T. For the sake of ease of illustration, let's assume that the portfolio position remains unchanged till time τ. Assume that the short interest rate is equal to a constant r and that the security price, S, is a geometric Brownian motion with constant

[1] From empirical results, it was found that the convergence rate of the multi-dimensional binomial method appears to be $O(\text{work}^{-1/(n+1)})$, where n is the dimension of underlying factors and that of the stochastic mesh [9] appears to be $O(\text{work}^{-1/4})$.

parameters μ and σ; i.e., the Black-Scholes assumptions. Then, at any time $0 < t < T$, the arbitrage-free value of the contract is, according to the pricing formula (1.1), equal to

$$V_t = e^{-r(T-t)}\left[\mathbb{E}_t^Q(S_T) - F\right].$$

where $F = S_0 \exp(rT)$ is the forward's price. Now, the probability that the value of the contract at time τ is negative will be given by

$$\begin{aligned}
\mathbb{P}(V_\tau \leq 0) &= \mathbb{P}(\mathbb{E}_\tau^Q(S_T) - F \leq 0) \\
&= \mathbb{P}\left(\frac{\mathbb{E}_\tau^{\mathbb{P}}\left(S_T \frac{dQ}{d\mathbb{P}}\right)}{\mathbb{E}_\tau^{\mathbb{P}}\left(\frac{dQ}{d\mathbb{P}}\right)} \leq F\right) \\
&= \mathbb{P}\left(\mathbb{E}_\tau^{\mathbb{P}}\left((S_T - F)\frac{dQ}{d\mathbb{P}}\right) \leq 0\right)
\end{aligned}$$

which can again be viewed as determining the distribution function of a conditional expectation at 0. In fact, under the current assumptions, the above probability can be computed in "closed form", namely,

$$\begin{aligned}
\mathbb{P}(V_\tau \leq 0) &= \mathbb{P}(S_\tau - Fe^{-r(T-\tau)} \leq 0) \\
&= \mathbb{P}(S_0 \exp((\mu - \sigma^2/2)\tau + \sigma\sqrt{\tau}B(1)) \leq S_0 \exp(r\tau)) \\
&= \Phi\left(\frac{1}{\sigma\sqrt{\tau}}\left[\left(r - \mu + \sigma^2/2\right)\tau\right]\right),
\end{aligned}$$

where $\Phi(\cdot)$ is the normal(0,1) distribution function. We can similarly express the α-quantile of V_τ as the quantile of a conditional expectation, namely, q_α that satisfies

$$\begin{aligned}
\alpha = \mathbb{P}(V_\tau \leq q_\alpha) &= \mathbb{P}\left(e^{-r(T-\tau)}\left(\mathbb{E}_\tau^Q(S_T) - F\right) \leq q_\alpha\right) \\
&= \mathbb{P}\left(\mathbb{E}_\tau^{\mathbb{P}}\left[e^{-r(T-t)}\left(\frac{\xi_T S_T}{\xi_\tau} - F\right)\right] \leq q_\alpha\right),
\end{aligned}$$

where $\xi_t = \mathbb{E}_t^{\mathbb{P}}(dQ/d\mathbb{P})$ as defined earlier.

Example 2. Suppose that the portfolio now consists of one plain European call option on the same underlying security as in Example 1. A European call option gives the holder the right, but not the obligation, to buy the stock at a given exercise price K on a given expiry date $T > \tau$. At time T, the option is rationally exercised if and only if $S_T > K$. Again assume that the call option is not exercised at time τ. Then, at time τ, the value of the call option is, by the Black-Scholes formula [13], equal to

$$V_\tau = S_\tau \Phi\left(\frac{\log\left(\frac{S_\tau}{K}\right) + \left(\mu + \frac{\sigma^2}{2}\right)(T-\tau)}{\sigma\sqrt{T-\tau}}\right) - Ke^{-r(T-\tau)}\Phi\left(\frac{\log\left(\frac{S_\tau}{K}\right) + \left(\mu - \frac{\sigma^2}{2}\right)(T-\tau)}{\sigma\sqrt{T-\tau}}\right).$$

Now, under the assumptions associated with the Black-Scholes' formula, the ratio of the prices S_τ and S_0, under the real-world probability measure $\mathbb{P}$, is log-normally distributed with parameters $(\mu - \sigma^2/2)\tau$ and $\sigma\sqrt{\tau}$. We can thus estimate the quantile of the value of the portfolio by Monte Carlo simulation. Specifically, we can generate samples of r.v. $Z \sim \log \mathrm{N}((\mu - \sigma^2/2)\tau, \sigma\sqrt{\tau})$. For each Z, we can set $S_\tau = S_0 Z$ and compute $\mathbb{P}(V_\tau \leq x)$ by Monte Carlo simulation.

However, in reality, unlike the two simple examples above, the value of the portfolio at time τ (i.e. the conditional expectation evaluation) cannot be solved in closed analytical form. This happens when the portfolio consists of securities with more complicated price processes (e.g. those involving stochastic volatility models.) We may then have to resort to numerical methods, such as solving by finite differences the PDE governing the dynamics of the derivatives' price or by using Monte Carlo simulation under the risk-neutral probability Q. Typically, the number of computations for solving the PDE grows exponentially with the number of state variables, as opposed to the linear growth of computations with the Monte Carlo approach. Moreover, the error convergence rate is independent of the dimension of the problem [7]. For problems with one or two state variables, however, it is typically the case that the finite-difference approach requires fewer computations in total than the Monte Carlo approach in order to obtain similar accuracy; see [7] and Chapter 11 of [13].

1.1.2 Bayesian Estimation

Consider an M/M/1 queue with inter-arrival rate λ and service rate μ that have a prior distribution. Suppose that we can deduce from the observations the posterior (joint) distributions, $\mathbb{P}$, of the two parameters. We restrict our attention to the case where the traffic intensity, $\rho \triangleq \lambda/\mu$ is strictly less than 1 so that steady state of the queueing process $Q(\cdot)$ exists. Let $r(n)$ be the rate of reward when the number of customers in the queue is n. Suppose that the simulator would like to find out, for large t, the probability that the total reward from time 0 to t is less than x, i.e.

$$\mathbb{P}\left(\int_0^t r(Q(s))\,ds \leq x\right).$$

An unbiased estimator is obtained by taking a sample average of

$$I\left(\int_0^t r(Q(s;\lambda,\mu))\,ds \leq x\right)$$

by Monte Carlo simulation. Specifically, we obtain a sample of (λ,μ), according to their posterior distribution. For each such (λ,μ), we simulate the queue until time t, and set the indicator function, $I(\int_0^t r(Q(s))\,ds \leq x)$ accordingly. We then take the average of the samples of indicator functions as an estimate of the desired probability.

One disadvantage of the above approach is that the simulation of the queue will become very computationally expensive if the time horizon $[0,t]$ is large. We know, however, that for each (λ,μ), by the strong law of large numbers (SLLN), the time average $\frac{1}{t}\int_0^t r(Q(s;\lambda,\mu))\,ds \to \mathbb{E}_{\lambda,\mu}r(Q(\infty))$ almost surely, where $\mathbb{E}_{\lambda,\mu}r(Q(\infty))$ is the expectation of the rate of reward of the queue at steady-state conditioned on (λ,μ). Hence, one remedy to the above mentioned problem is to approximate the integral by $t\cdot\mathbb{E}_{\lambda,\mu}r(Q(\infty))$; namely, we approximate the desired probability by

$$\mathbb{P}(t\cdot\mathbb{E}_{\lambda,\mu}r(Q(\infty)) \leq x).$$

This has the same form as the problem of computing the distribution function of conditional expectation.

1.1.3 Literature Review

From a practical point of view, the problem of estimating quantiles is quite important. There exists a considerable literature on estimating quantiles from independent, identically distributed observations (see Cramér [11]). In particular, CLT's have been established for the sample quantiles in the independent and identically distributed (i.i.d.) setting [1, 26]. Iglehart [19] established such CLT's in the context of regenerative stochastic systems. The rate at which such CLT's converge is $O(n^{-1/2})$, where n is the number of samples. Bahadur [1] showed that it was possible to express sample quantiles as sums, via representation as a linear transform of the sample distribution function evaluated at the relevant quantile. From these representations, a number of important insights and properties follow. A precise and illuminating technical discussion of the empirical and quantile processes taken together has been given in the appendix of a paper by Shorack [27].

To the best of our knowledge no literature exists which gives the simulator any guidance for efficiently estimating the distribution functions or the quantiles of conditional expectation and provides their corresponding limit theorems.

Regulatory agencies and investors are increasingly concerned about the risk exposures of financial institutions through their large positions in over-the-counter (OTC) derivatives. This concern is due in part to the losses by both finanical and non-financial corporations on derivative positions during 1994 and 1995. There have been increased calls for the measurement and disclosure of risks associated with derivatives. Recent discussions between regulators and their constituent financial institutions have resulted in a widely applied measure of market risk called "capital at risk" or "value-at-risk" (VaR); see [20]. Two uses of VaR are reporting risk exposures to the SEC for their use in assessing overall systematic risk in the financial system and in internal monitoring of risk exposures.

VaR for derivatives is typically obtained by approximating the value of the derivatives over a short time period by the "delta". "delta-gamma" or "delta-vega" approaches. Specifically, this relies on the linear and parabolic approximations to the exact price function. The drawback of this approach is that the accuracy of these approximations will tend to deteriorate if the time horizon for the VaR is large.

As an alternative, one can "build" an approximate pricing formula for each derivative for which there is no explicit formula, such as Black-Scholes, at hand. For instance, by Monte Carlo or lattice-based calculations, one can estimate the price of a derivative at each of a small number of underlying prices, and from these fit a spline, or some other low-dimensional analytic approximation, for the derivative price. The major restriction is that the dimension of the random factors must be sufficiently low.

To overcome these two problems, we estimate the actual VaR by simulating the market value of the entire portfolio, where the prices of individual instruments are also computed by Monte Carlo. The advantage of this approach is that given a suitable model for the stochastic behaviour of the underlying stochastic factors, it is applicable even for a moderately large time horizon. Morever, the rate of convergence of this approach is independent of the dimension of the underlying factors. We will resolve the problem of how the to allocate a fixed amount of computational budget in order to achieve the fastest rate of convergence.

1.1.4 Thesis Organization

The remainder of the thesis is organized as follows: In the next section, we will provide a rigorous mathematical formulation for our problem. We will also describe the estimators and summarize their convergence properties. In Chapter 2, we will discuss the case in which the conditioning random element, Z, is discrete. The settings for the more general conditioning element Z will be covered in Chapter 3. In these two chapters, we will discuss the problem of estimating the distribution function of the conditional expectation and examine the closely related problem of finding the quantile of a conditional expectation. Limit theorems for the associated quantile estimators will be shown to be closely tied to those derived for probability distributions of conditional expectations. We will finally validate the various theoretical properties with some numerical experimental results.

1.2 Mathematical Overview

In this section, we will define our two problems. In Subsection 1.2.1, we will review the definition of conditional expectation. In Subsection 1.2.2 we will define the quantile, empirical distribution and empirical quantile. We will outline a "recipe" for obtaining the central limit theorems for the estimators of the distribution functions and quantiles in Subsection 1.2.3. A summary of the estimators will be summarized in Subsection 1.2.4. All notation that will be utilized in this thesis will be summarized in SubSection 1.2.5.

1.2.1 Conditional Expectation

Kolmogorov [22] gave the following fundamental definition for conditional expectation.

Definition 1.2.1 *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space (triple), and X a random variable with $\mathbb{E}(|X|) < \infty$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then there exists a random variable Y such that*

1. *Y is $\mathcal{G}$ measurable.*
2. *$\mathbb{E}(|Y|) < \infty$.*
3. *for every set G in $\mathcal{G}$ (equivalently, for every set G in some π-system which contains Ω and generates $\mathcal{G}$), we have*

$$\int_G Y \, d\mathbb{P} = \int_G X \, d\mathbb{P}, \qquad \forall G \in \mathcal{G}.$$

Morever, if $\tilde{Y}$ is another r.v. with these properties then $\tilde{Y} = Y$, a.s., that is, $\mathbb{P}(\tilde{Y} = Y) = 1$. A random variable Y with properties 1-3 is called a version of the conditional expectation $\mathbb{E}(X|\mathcal{G})$ of X given $\mathcal{G}$, and we write $Y = \mathbb{E}(X|\mathcal{G})$.

Let Z be a random variable defined on the above probability space. We write $\mathbb{E}(X|Z)$ for $\mathbb{E}(X|\sigma(Z))$, $\mathbb{E}(X|Z_1, Z_2, \dots)$ for $\mathbb{E}(X|\sigma(Z_1, Z_2, \dots))$, etc. Note that if $\mathcal{G}$ is the trivial σ-algebra $\{\emptyset, \Omega\}$, then $\mathbb{E}(X|\mathcal{G})(\omega) = \mathbb{E}(X)$ for all ω.

This definition agrees with traditional usage. Suppose that X and Z are r.v.'s which have a joint probability density function (pdf) $f_{X,Z}(x,z)$. Then, $f_Z(z) = \int_{\mathbb{R}} f_{X,Z}(x,z)\,dx$ acts as a probability density function for Z. Define the elementary conditional pdf $f_{X|Z}$ of X given Z via

$$f_{X|Z}(x|z) \triangleq \begin{cases} f_{X,Z}(x,z)/f_Z(z) & \text{if } f_Z(z) \neq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Let h be a Borel function on $\mathbb{R}$ such that

$$\mathbb{E}|h(X)| = \int_{\mathbb{R}} |h(x)| f_X(x)\,dx < \infty,$$

where $f_X(x) = \int_{\mathbb{R}} f_{X,Z}(x,z)\,dz$ gives a pdf for X. Set

$$g(z) \triangleq \int_{\mathbb{R}} h(x) f_{X|Z}(x|z)\,dx.$$

Then, $Y \triangleq g(Z)$ is a version of the conditional expectation of $h(X)$ given $\sigma(Z)$.

We can also view conditional expectations as least-squares-best predictors. In particular, if $\mathbb{E}(X^2) < \infty$, then the conditional expectation $Y = \mathbb{E}(X|\mathcal{G})$ is a version of the orthogonal projection of X onto $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. Hence, Y is the least-squares-best $\mathcal{G}$-measurable predictor of X: amongst all $\mathcal{G}$-measurable functions (i.e. amongst all predictors which can be computed from the available information), Y minimizes

$$\mathbb{E}[(Y - X)^2].$$

In the case of the computation of VaR (say at time τ), the random variable X in this case will be the discounted future cash flow. The conditional expectation is taken with respect to an Equivalent Martingale Measure (EMM) (whose definition and proof of existence can be found in, for example, [13]). The "conditioning random element" Z in this case will be the sample paths of the Brownian motion B of the SDE satisfied by the price process in the EMM up to time τ. Intuitively, it contains every event based on the history of the price process up to time τ. For technical reasons, however, one must be able to assign probabilities to the null sets of Ω, the subsets of events of zero probability. For this reason, we will fix the standard filtration $\{\mathcal{F}_t : t \geq 0\}$ of B, with $\mathcal{F}_t$ defined as the tribe generated by the union of $\sigma\{B_s : 0 \leq s \leq t\}$ and the null sets. In this case, the probability measure $\mathbb{P}$ is also extended by letting $\mathbb{P}(A) = 0$ for any null set A.

1.2.2 Quantiles

For any univariate distribution function F, and for $0 < p < 1$, the quantity

$$q_p \triangleq F^{-1}(p) \triangleq \inf\{x : F(x) \geq p\}$$

is called the p-th *quantile* or *fractile* of F. The following proposition, giving useful properties of F and F^{-1}, is easily checked.

Lemma 1.2.1 *Let F be a distribution function. The function $F^{-1}(t), 0 < t < 1$, is nondecreasing and left-continuous, and satisfies*

1. $F^{-1}(F(x)) \leq x,\ -\infty < x < \infty,$
2. $F(F^{-1}(t)) \geq t,\ 0 < t < 1.$
3. $F(x) \geq t$ *if and only if* $x \geq F^{-1}(t)$.

Note that the p-th quantile, $F^{-1}(p)$ satisfies

$$F(q_p-) \leq p \leq F(q_p).$$

We would like to remark that the univariate r.v. in our problem context is the conditional expectation $\mathbb{E}(X|Z)$.

Consider an i.i.d. sequence $\{X_i\}$ with distribution function F. For each sample of size n, $\{X_1, \dots, X_n\}$, a corresponding *sample distribution function* F_n is constructed by placing at each observation X_i a mass of $1/n$. Thus, F_n may be represented as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \qquad -\infty < x < \infty.$$

For each fixed sample, $F_n(\cdot)$ is a distribution function, considered as a function of x. On the other hand, for each fixed value of x, $F_n(x)$ is a random variable, considered as a function of the sample.

Corresponding to a sample $\{X_1, \ldots, X_n\}$ of observations on F, the sample p-th quantile, $q_{p,n}$ is defined as the p-th quantile of the sample distribution function F_n. An equivalent way of defining $q_{p,n}$ is to let $X_{(i)}$ be the "i-th order statistic" of the sample $X_1, \ldots, X_n$, so that

$$X_{(1)} \leq \cdots \leq X_{(n)}.$$

Then, $q_{p,n} = X_{(\lfloor np \rfloor)}$, namely the $\lfloor np \rfloor$-th largest observation in the sample $X_1, \ldots, X_n$. Here, for any $x \in \mathbb{R}^+$, $\lfloor x \rfloor$ is the largest integer not greater than x; i.e., it is the "floor" function.

For the rest of this thesis, unless stated otherwise, we will denote by $F(\cdot)$ the distribution function of $\mathbb{E}(X|Z)$ and by $F_c(\cdot) \triangleq \frac{1}{n(c)} \sum_{i=1}^{n(c)} I(\bar{X}_{m(c)}(Z_i) \leq x)$ its estimate when the computation budget available is equal to c. Here, $\bar{X}_{m(c)}(Z_i) \triangleq m(c)^{-1} \sum_{j=1}^{m(c)} X_j(Z_i)$, where $X_j(Z_i)$ has distribution function $\mathbb{P}(X \in \cdot | Z = Z_i)$. Similarly, we will denote the α-quantile and sample quantile of the conditional expectation by q_α and $q_{\alpha,c}$ resp.

1.2.3 Central Limit Theorem for the Estimators

In this subsection, we will outline a "recipe" for obtaining central limit theorems for the estimators.

Notice that one way to gauge the quality of an estimator, $\hat{\alpha}$, of a performance measure, α, is via the inverse of its mean squared error (MSE); the smaller the MSE the better. Recall that MSE can be decomposed as the sum of the its variance and the square of its bias. Typically, both the bias and the variance terms have certain asymptotic forms in terms of the estimator parameters (In our case, the parameters will be the number of Z and X to be sampled.)

We obtain the fastest rate of convergence to α by minimizing the MSE (as a function of c) with respect to the estimator parameters. Let $g(c)$ be the asymptotic MSE corresponding to the minimizer above. Usually, $g(c)$ is some power function of c. Then, the CLT can typically be expressed as

$$g(c)^{-1/2}(\hat{\alpha} - \alpha) \Rightarrow \text{bias term} + \text{"noise"}.$$

If we choose other (sub-optimal) growth rates for the parameters, the estimator will then converge to either the bias or the noise term (but at the expense of a slower convergence rate of the estimator.)

These CLT's provide hints on how confidence intervals (c.i.'s) for the estimators can be formed. One slight drawback of choosing parameters to achieve the fastest rate of convergence of the estimator is that the formation of c.i.'s with the correct coverage probabilities is non-trivial, since the bias term is usually unobservable. To get around this, one may need to resort to choices of the parameters according to some suboptimal rates of convergence of the estimator, so that the corresponding CLT removes the biasterm. Alternatively, one may need to allocate a small fraction of the computational budget to estimate the size of the bias term. We will investigate these ideas in later chapters.

1.2.4 Summary of Results

The distribution function estimator that we will be using in the cases of the general conditioning random element and the discrete conditionaing random element with infinite state space is

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \leq 0\right).$$

In particular, we sample n Z's; for each Z_i sampled, we sample m number of X's according to the distribution function $\mathbb{P}(X \in \cdot | Z = Z_i)$. The estimator is then set to be the sample average of the indicator functions. Here, the sample sizes m and n are the estimator parameters whose optimal rates of growth corresponding to the fastest rate of convergence of the estimator will be examined in later chapters.

For a given computational budget c, it can be shown that the optimal convergence rate of the estimator is $O(c^{-1/3})$ for the general conditioning random element case and is $O(\sqrt{\log c/c})$ for the discrete infinite case. Further details of this estimator can be found in Chapter 3 and Section 2.2 of Chapter 2.

For the case in which the conditioning random element takes on only a finite number of values, we will consider another estimator that makes heavy use of the assumption that the probabilility mass function of the conditioning random element is known *a priori*. It turns out that in that case the estimator converges to α exponentially fast. The details of this result can be found in the first section of the next chapter.

As for the quantile estimator, we'll be using the empirical quantile estimator of $\{\bar{X}_m(Z_i) : i = 1, \ldots, n\}$ (which are the non-biased estimator of $\mathbb{E}(X|Z = Z_i)$. It will be shown that under certain technical conditions, the convergence properties of the quantile estimator is closed tied to that of the distribution function estimator. We refer the reader to Section 3.2 for further details.

1.2.5 Notation

In order that the reader may have an easy reference, all notation used repeatedly in this dissertation is summarized in this subsection. The terms listed will be explained in more detail when they are used for the first time.

We use the notation that capital letters stand for random variables. Throughout this thesis, we assume a fixed computational budget c. The number n usually refers to the number of conditioning random elements we sample; for each conditioning random element, the number of X's we sample is usually denoted by m. The symbol "$\Rightarrow$" stands for weak convergence or convergence in distribution.

We adopt the notation that $X(z)$ is a r.v. with *regular distribution function* $\mathbb{P}(X \leq x|Z = z)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A regular distribution function is a map

$$\mathbb{P}(\cdot, \cdot) : \Omega \times \mathcal{F} \to [0, 1]$$

such that

1. for $F \in \mathcal{F}$, the function $\omega \mapsto \mathbb{P}(\omega, F)$ is a version of $\mathbb{P}(F|\mathcal{G})$;

2. for almost every ω, the map
$$F \mapsto \mathbb{P}(\omega, F)$$
is a probability measure on $\mathcal{F}$.

Breiman [8] showed that if X takes values in a Borel space $(\mathbb{R}, \mathcal{B})$, then there is a regular conditional distribution for X given Z. Moreover, if Z is a r.v., we can define $\mathbb{P}(X \in \cdot|Z(\omega)) = \varphi(Z(\omega))$, where $\varphi(\cdot)$ is a $\mathcal{B}$ measurable function. (See, for example, [23] for other conditions under which regular conditional probabilities exist.)

We will henceforth use the notation that for any two real valued functions: $f(\cdot)$ and $g(\cdot)$, $(f \cdot g)(x) \triangleq f(x) \cdot g(x)$.

List of Notation

Below is a list of the notations that we will be utilizing in this thesis:

$X(z)$: $X(z)$ is a r.v. with *regular* distribution function $\mathbb{P}(X \leq x|Z = z)$.

$\mu(z)$: $\mathbb{E}(X|Z = z)$

$\sigma^2(z)$: $\mathrm{Var}(X|Z = z)$

Y: $\mu(Z)/\sigma(Z)$;

notice that Y is the inverse of the conditional coefficient of variation of the target variable, X. As we will later see, this Z-measurable variable and its probability density (or mass) function plays a very critical role in the determination of the rate of convergence of our estimators for the discrete and general conditioning random element cases. Henceforth in this thesis, we will denote it by $ICCV$ (short form for the Inverse of Conditional Coefficient of Variation.)

$\rho_k(z)$: $\mathbb{E}((X - \mu(Z))^k|Z = z)/\sigma(z)^k$

$\beta_k(z)$: $\mathbb{E}(|X - \mu(Z)|^k|Z = z)/\sigma(z)^k$

$\bar{\beta}_k(y)$: $\mathbb{E}(\beta_k(Z)|Y = y)$

$\gamma_k(z)$: the k-th cumulant of X given $Z = z$:

in particular, $\gamma_k(z) = \frac{1}{i^k}\left[\frac{d^k}{d\theta^k}\log c(\theta, z)\right]_{\theta=0}$, where $c(\cdot, z)$ is the characteristic function of X given that $Z = z$.

$\nu_z(t)$: $\mathbb{E}\exp(itX(z))$, i.e. the characteristic function of $X(z)$

$\Phi(\cdot)$: $\mathbb{P}(\mathrm{N}(0,1) \le \cdot)$; the d.f. of a $\mathrm{N}(0,1)$ r.v.

$\phi(\cdot)$: $\frac{1}{\sqrt{2\pi}}e^{-(\cdot)^2/2}$; the density function of a $\mathrm{N}(0,1)$ r.v.

$f_Y(y)$: (probability) density function of the r.v. Y (assumed to exist for the moment)

C_b^k: the class of k-times continuously differentiable functions whose derivatives of order less than k are bounded globally.

$\overset{\mathcal{D}}{=}$: means "is distributed as".

$\overset{\mathcal{D}}{\approx}$: means "is approximately distributed as".

Chapter 2

Estimators for Discretely Conditioned Expectations

In this chapter, we assume that X and Z are random variables and that the range of Z is discrete; i.e., Z takes values in $\Gamma = \{z_1, z_2, \dots\}$. We will consider two different cases: (a) the sample size of Z is finite and small and (b) the sample size of Z is large or is infinite. In each case, we will propose an estimator of the quantity $\mathbb{P}(\mathbb{E}(X|Z) \leq 0)$, establish the fastest rate of convergence, and obtain the asymptotically best estimation parameters. It turns out that the fastest rate of convergence in the first case is exponential; whereas in the second case, it's equal to $O(\sqrt{\frac{\log c}{c}})$, where c is the computation budget.

2.1 Finite Conditioning Space

Assume that $|\Gamma| = K < \infty$. Assume further that $\mathbb{P}(Z = z) = p(z)$ for $z \in \Gamma$ are known to us *a priori*. We can now rewrite $\alpha = \mathbb{P}(\mathbb{E}(X|Z) \leq 0) = \mathbb{E}(I(\mathbb{E}(X|Z) \leq 0))$ as

$$\alpha = \sum_{z \in \Gamma} p(z) I(\mathbb{E}(X|Z = z) \leq 0).$$

Consider the following estimator of α:

$$\hat{\alpha} = \sum_{z\in\Gamma} p(z) I\left(\frac{1}{m(z)}\sum_{j=1}^{m(z)} X_j(z) \le 0\right),$$

where $X_j(z)$ has distribution function $\mathbb{P}(X \le \cdot | Z = z)$ for $z \in \Gamma$, $j \ge 1$.

Specifically, for each $z \in \Gamma$, we sample $m(z)$ X's under the d.f. $\mathbb{P}(X \in \cdot | Z = z)$. The sample average $(m(z))^{-1}\sum_{j=1}^{m(z)} X_j(z)$ is then an unbiased estimator of the conditional expectation $\mathbb{E}(X|Z = z)$. We then take a weighted sum of the the indicator functions with the weights being equal to the probability mass $p(z)$'s.

Given a fixed computational budget c, it stands to reason that we should assume that $(m(z) : z \in \Gamma)$ satisfies the constraints that $\sum_{z\in\Gamma} m(z) = \beta\cdot c$ and $m(z) > 0$ for $z \in \Gamma$, where β is the number of X's we can sample for each unit of computational budget. Henceforth, without loss of generality, we assume that $\beta = 1$. For a given choice of $(m(1), \ldots, m(z))$, we will, in this section, determine the bias, variance and hence the MSE of the estimator. We will then derive the optimal choice of $(m(z) : z \in \Gamma)$ that will asymptotically minimize the MSE, and we will derive a limit theorem for such an estimator.

Before we delve into the theoretical results of this estimator, let us take a look at a motivational example for this problem setting. Recall the Bayesian Estimation example discussed in Subsection 1.1.2. Assume that in that example, the M/M/1 queue corresponds to the student services office of an academic department of a university. In particular, imagine that at the beginning of the quarter, the department wants to consider whether it is worthwhile to increase the number of staff members. The criterion used is the 95% quantile of the total waiting time of all the students over the whole academic year. However, though the department has given offers to n prospective students, only $S \le n$ students will eventually accept the offer. Assume that for each student, the time between visits to the student services department is independent and exponentially distributed with parameter λ_1 and that each service request requires a service time that is exponentially distributed with parameter μ. Assume further that all the students are independent so that the students' requests for service arrive at the queue as a Poisson process with parameter $S\cdot\lambda_1$. Suppose that S has a posterior distribution, which is binomial(n, p). Hence, in this case we have that $Z = S$ and $\Gamma = \{0, 1, \ldots, n\}$.

2.1.1 Mean Squared Error of $\hat{\alpha}$

Notice that since the indicator function $I(\cdot)$ is a nonlinear function, though $\bar{X}_{m(z)}(z) \triangleq \frac{1}{m(z)} \sum_{j=1}^{m(z)} X_j(z)$ is an unbiased estimator of $\mathbb{E}(X|Z = z)$, $\hat{\alpha}$ is unbiased only asymptotically. Let's now examine a bit closer the bias of the estimator.

We say that the sequence $\{X_n\}$ are lattice random variables if X_i only takes on values in the set $\{c + kd : k = 0, \pm 1, \pm 2, \ldots\}$, where c is a fixed constant and d is the lattice spacing.

Theorem 2.1.1 *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables such that $\mathbb{E}X_1 = 0$ and $c(\theta) \equiv \mathbb{E}\{\exp[\theta X_1]\} < \infty$ for all $\theta \in \mathbb{R}$. Let $a > 0$ and assume that $\mathbb{P}(X_1 > a) > 0$. Then, if X_1 is not of lattice type,*

$$\lim_{n\to} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq a\right) \exp[nJ(a)]\sqrt{n} = \frac{1}{\sqrt{2\pi}\theta_a\sigma}.$$

where

$$J(a) \triangleq \sup_{-\infty<\theta<\infty} \{\theta a - \log c(\theta)\} = \theta_a a - \log c(\theta_a).$$

$$a = \frac{\mathbb{E}\{X_1 \exp[\theta_a X_1]\}}{c(\theta_a)} = \frac{c'(\theta_a)}{c(\theta_a)}.$$

and

$$\sigma^2 \triangleq \frac{c''(\theta_a)}{c(\theta_a)} - \left[\frac{c'(\theta_a)}{c(\theta_a)}\right]^2 = \frac{c''(\theta_a)}{c(\theta_a)} - a^2.$$

and

The above theorem [10, p.121] shows that for each z, there exists constants $\gamma(z)$ and $\eta(z)$ such that

$$\mathbb{P}(\bar{X}_m(z) \leq 0) \sim \frac{\gamma(z)}{\sqrt{m}} \exp(-m\eta(z)),$$

as $m \to \infty$. For the case where the conditional distribution of X given $Z = z$ is normal, clearly, $\eta(z) = \frac{1}{2}Y(z)^2$, where $Y(z) = \mu(z)/\sigma(z)$, $\mu(z) \triangleq \mathbb{E}(X|Z = z)$, and $\sigma(z)^2 \triangleq$

$\mathrm{Var}(X|Z=z)$.

Now denote by Γ_+ the index set $\{z \in \Gamma : \mu(z) > 0\}$. The bias term is given by:

$$
\begin{aligned}
\mathbb{E}(\hat{\alpha}) - \alpha &= \sum_{z\in\Gamma} p(z)\mathbb{P}(\bar{X}_{m(z)}(z) \le 0) - \alpha \\
&= \sum_{z\in\Gamma_+} p(z)\mathbb{P}(\bar{X}_{m(z)}(z) \le 0) + \sum_{z\notin\Gamma_+} p(z)\left(1 - \mathbb{P}(\bar{X}_{m(z)}(z) > 0)\right) - \alpha \\
&= \sum_{z\in\Gamma_+} p(z)\mathbb{P}(\bar{X}_{m(z)}(z) \le 0) - \sum_{z\notin\Gamma_+} p(z)\mathbb{P}(\bar{X}_{m(z)}(z) > 0).
\end{aligned}
$$

On the other hand, the variance of the estimator is equal to

$$
\mathrm{Var}(\hat{\alpha}) = \sum_{z\in\Gamma} p(z)^2\mathbb{P}(\bar{X}_{m(z)}(z) \le 0)\cdot\mathbb{P}(\bar{X}_{m(z)}(z) > 0)
$$

The mean square error is then given by

$$
\mathrm{MSE}(\hat{\alpha}) = [\mathbb{E}(\hat{\alpha}) - \alpha]^2 + \mathrm{Var}(\hat{\alpha}).
$$

The idea now is to minimize $\mathrm{MSE}(\hat{\alpha})$ with respect to $(m(z) : z \in \Gamma)$. In order to make the minimization problem a bit more tractable, we will determine some bounds on $\mathrm{MSE}(\hat{\alpha})$. Specifically, we use the inequality $(\sum_{i=1}^n x_i)^2 \le n\sum_{i=1}^n x_i^2$, to bound $\mathrm{MSE}(\hat{\alpha})$ by

$$
\begin{aligned}
\mathrm{MSE}(\hat{\alpha}) &\le \mathrm{Var}(\hat{\alpha}) + K\sum_{z\in\Gamma} p(z)^2\frac{\gamma(z)^2}{m(z)}e^{-2m(z)\eta(z)} \\
&\le (K+1)\sum_{z\in\Gamma} p(z)^2\gamma(z)\exp(-m(z)\eta(z)). \qquad (2.1)
\end{aligned}
$$

Denote by V the term on the right hand side (RHS) of the last inequality.

Now, let's minimize (2.1) with respect to $(m(z) : z \in \Gamma)$ subject to $\sum_{z\in\Gamma} m(z) = c$ and $m(z) > 0$ for all $z \in \Gamma$. Let λ be the Lagrange multiplier of the equality constraint. Taking partials with respect to $m(z)$, the optimizer $m^*(z)$ must satisfy

$$
\frac{\partial V}{\partial m(z)} = p(z)^2\gamma(z)e^{-m^*(z)\eta(z)}(-\eta(z)) + \lambda = 0.
$$

In other words,

$$m^*(z) = \frac{-1}{\eta(z)} \log\left(\frac{\lambda}{p(z)^2\gamma(z)\eta(z)}.\right)$$

Substituting the above expression for $m^*(z)$ into the equality constraint, we deduce that λ satisfies

$$-\log\lambda = \frac{c - \sum_{z\in\Gamma} \frac{1}{\eta(z)} \log(p(z)^2\gamma(z)\eta(z))}{\sum_{z\in\Gamma} \eta(z)^{-1}}.$$

Hence, we have derived the expression for the optimal $m^*(z)$ that minimizes the upper bound:

$$m^*(z) = \frac{\eta(z)^{-1}}{\sum_j \eta(j)^{-1}} + \eta(z)^{-1}\log(p(z)^2\gamma(z)\eta(z)) - \frac{\eta(z)^{-1}}{\sum_{w\in\Gamma}\eta(w)^{-1}} \sum_{w\in\Gamma} \frac{\log(p(w)^2\gamma(w)\eta(w))}{\eta(w)}.$$

The minimized upper bound on $\mathrm{MSE}(\hat{\alpha})$ is thus equal to

$$\begin{aligned} V_{\min} &= \sum_{z\in\Gamma} \exp\left(\frac{-c}{\sum_{w\in\Gamma}\eta(w)^{-1}}\right) \exp\left(\sum_{w\in\Gamma} \frac{\eta(w)^{-1}}{\sum \eta(\cdot)^{-1}} \log(p(w)^2\gamma(w)\eta(w))\right) \\ &= K\left(\prod_{z\in\Gamma} (p(z)^2\gamma(z)\eta(z))^{\frac{\eta(z)^{-1}}{\sum_{w\in\Gamma}\eta(w)^{-1}}}\right) \exp\left(\frac{-c}{\sum_z \eta(z)^{-1}}\right). \end{aligned}$$

In other words, the upper bound, V, on $\mathrm{MSE}(\hat{\alpha})$ is converging to zero exponentially fast at the rate $O(e^{-\gamma c})$, where $\gamma \triangleq (\sum_z \eta(z)^{-1})^{-1}$. Notice that γ is largely determined by the term $\eta(\underline{z})^{-1}$, where $\underline{z} \triangleq \arg\min\{|\eta(z)| : z \in \Gamma\}$.

Let's now examine a lower bound on the MSE. Notice that

$$\mathrm{MSE}(\hat{\alpha}) \geq \mathrm{Var}(\hat{\alpha}) \geq \frac{1}{2}\sum_{z\in\Gamma} p(z)^2 \frac{\gamma(z)}{\sqrt{m(z)}} \exp(-m(z)\eta(z))$$

for large $m(z)$, since any one of two terms, $\mathbb{P}(\bar{X}_{m(z)}(z) \leq 0)$ or $\mathbb{P}(\bar{X}_{m(z)}(z) > 0)$ is greater than or equal to 1/2.

Note that in order to drive $\mathrm{MSE}(\hat{\alpha})$ to 0, we must have that the $m_c(\cdot)$ that minimizes $\mathrm{MSE}(\hat{\alpha})$ must satisfy $m_c(z) \nearrow +\infty$ for all $z \in \Gamma$ as $c \nearrow +\infty$. On the other hand, for each $\epsilon > 0$, there exists M_ϵ such that for all $m \geq M_\epsilon$, $m^{-1/2} \geq \exp(-\epsilon m)$. Combining these two remarks, we have that for all $\epsilon > 0$, there exists C_ϵ such that for all $c > C_\epsilon$

$$\begin{aligned}\min \mathrm{MSE}(\hat{\alpha}) &\geq \min \frac{1}{2} \sum_z p(z)^2 \gamma(z) \exp(-(\eta(z) + \epsilon) m(z)) \\ &= \text{constant} \cdot \exp\left(\frac{-c}{\sum_z (\eta(z) + \epsilon)^{-1}} \right).\end{aligned}$$

In other words, for all $c > C_\epsilon$,

$$\frac{-1}{\sum_z \eta(z)^{-1}} \geq \frac{1}{c} \log(\min \mathrm{MSE}(\hat{\alpha})) \geq \frac{-1}{\sum_z (\eta(z) + \epsilon)^{-1}}.$$

Since $\epsilon > 0$ is arbitrary, the above argument leads us to the following proposition.

Proposition 2.1.1 *For any choice of $(m_c(z) : z \in \Gamma)$, the $MSE(\hat{\alpha})$ satisfies*

$$\liminf_{c \to \infty} \frac{1}{c} \log MSE(\hat{\alpha}) \geq -\gamma.$$

The above equality can be attained if we choose

$$m_c(z) \triangleq \gamma \eta(z)^{-1} c + o(c).$$

In this case,

$$\lim_{c \to \infty} \frac{1}{c} \log MSE(\hat{\alpha}) = -\gamma.$$

In summary, we have identified the asymptotically optimal choice of $m(\cdot)$ and shown that the corresponding optimal rate of convergence is exponential with parameter γ.

2.2 Infinite State Space

In this section, we will consider the case in which the probability mass function of Z is not explicitly known *a priori* or the sample space Γ of Z is so large that it is impractical to use the estimator proposed in the previous section. We will assume further that that $|\eta(z)|$ is uniformly bounded away from 0, so that

$$l \triangleq \inf_{z\in\Gamma} \eta(z) = \eta(\underline{z}) > 0,$$

and the conditions of Theorem 2.1.1 hold for all $z \in \Gamma$: (namely, for all $z \in \Gamma$ and $\theta \in \mathbb{R}$, $\mathbb{E}\exp(\theta \cdot X(z)) < \infty$), so that

$$\mathbb{P}(\bar{X}_m(z) \leq 0) \sim \frac{\gamma(z)}{\sqrt{m}} \exp(-m\eta(z)).$$

The estimator we will consider in this section is given by

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \leq 0\right).$$

Specifically, we sample n Z's via Monte Carlo simulation. For each sampled Z_i, we sample m X's according to the conditional distribution $\mathbb{P}(X \in \cdot | Z = Z_i)$ and set the indicator function appropriately. The estimator is then set to be the sample average of the indicator functions.

Again, notice that even though $\bar{X}_m \triangleq m^{-1}\sum_{j=1}^{m} X_j(Z_i)$ is an unbiased estimator of $\mathbb{E}(X|Z = Z_i)$, $\hat{\alpha}$ is in general unbiased only asymptotically.

2.2.1 Mean Square Error

As in the previous section, we will examine how the mean squared error (MSE) of $\hat{\alpha}$ can be minimized. Define as in the previous section the index set $\Gamma_+ \triangleq \{z \in \Gamma : \mu(z) > 0\}$. Observe that the bias of the estimator is given by

$$\begin{aligned}
\mathbb{E}(\hat{\alpha}) - \alpha &= \sum_{z\in\Gamma_+} p(z)\mathbb{P}(\bar{X}_m(z) \leq 0) + \sum_{z\notin\Gamma_+} p(z)\left(1 - \mathbb{P}(\bar{X}_m(z) > 0)\right) - \alpha \\
&= \sum_{z\in\Gamma_+} p(z)\mathbb{P}(\bar{X}_m(z) \leq 0) - \sum_{z\notin\Gamma_+} p(z)\mathbb{P}(\bar{X}_m(z) > 0).
\end{aligned}$$

Unless in some trivial cases in which the above bias is exactly zero (for instance, when the density function of $\mathbb{E}(X|Z)$ is symmetric about the point 0), the magnitude of the above bias is, given any arbitrarily small $\epsilon > 0$, bounded by

$$p(\underline{z})e^{-m(l+\epsilon)} \leq |\mathbb{E}(\hat{\alpha}) - \alpha| \leq e^{-ml} \tag{2.2}$$

for sufficiently large m.

Also, for sufficiently large m, the asymptotic variance of the estimator is given by

$$\begin{aligned} \mathrm{Var}(\hat{\alpha}) &= \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{P}(\bar{X}_m(Z_i) \leq 0) \cdot \mathbb{P}(\bar{X}_m(Z_i) > 0) \\ &\sim \frac{\alpha(1-\alpha)}{n}. \end{aligned}$$

Since $m \cdot n = c$, in order to minimize the MSE, we want to choose the asymptotically optimal m that minimizes

$$V(m) \triangleq e^{-2ml} + \frac{\alpha(1-\alpha)m}{c}.$$

Taking the derivative of $V(m)$ with respect to (w.r.t.) m and setting the derivative to zero, we have that

$$-2le^{-2m^*(c)l} + \frac{\alpha(1-\alpha)}{c} = 0$$

that is, $m^*(c)$ is given by

$$m^*(c) = \log\left(\frac{\alpha(1-\alpha)}{2cl}\right) \cdot \frac{-1}{2l}.$$

Then, the asymptotically minimized MSE is given by

$$\frac{\alpha(1-\alpha)}{2lc} + \frac{\alpha(1-\alpha)}{c} \cdot \frac{1}{2l}\log\left(\frac{2cl}{\alpha(1-\alpha)}\right) = \Theta\left(\frac{\log c}{c}\right).$$

To summarize, if we choose m to grow with c at the rate of $\frac{1}{2l}\log c$ and set $n = \frac{c}{m}$, the asymptotic MSE will be minimized and the root mean squared error $\mathrm{RMSE}(\hat{\alpha}) \to 0$ at the rate of $\Theta(\sqrt{\log c/c})$ [1].

[1] We say that a function of c, $h(c)$, is $\Theta(g(c))$ if there exist positive finite constants a and b such that

We can see here a big degradation in the convergence rate from the case where $m_c^*(z)$ depends on z (as analyzed in the last section) to where we force $m_c(z)$ to be constant in z.

2.2.2 Central Limit Theorem

In fact, one can derive a CLT for the estimator $\hat{\alpha}$. The CLT is summarized by the proposition below. The proof of the proposition utilizes the Lindeberg-Feller Theorem [5, p.359]. Before we prove the proposition, let's consider the following lemma which shows that under some mild conditions, if a triangular array of r.v. is uniformly integrable then it will satisfy the Lindeberg-Feller condition.

Lemma 2.2.1 *Assume that the following conditions hold:*

1. *the random variables $(X_{c,j} : j = 1, 2, \ldots, n(c), c > 0)$ is independent and identically distributed where $n(c) \nearrow +\infty$ as $c \nearrow +\infty$;*
2. $\mathbb{E}X_{c,1} = 0$, $\sigma_c^2 \triangleq \mathbb{E}X_{c,1}^2$;
3. $\lim_{c\to\infty} \sigma_c^2 = \sigma^2 \in (0, \infty)$;
4. *the family $\{X_{c,1}^2 : c > 0\}$ is uniformly integrable.*

Then, $\{X_{c,i}^2 : i = 1, \ldots, n(c), c > 0\}$ satisfies the Lindeberg-Feller condition: namely,

$$\lim_{c\to\infty} \frac{1}{\sigma_c^2} \int_{|X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\sigma_c} X_{c,1}^2 \, d\mathbb{P} = 0$$

for all $\epsilon > 0$.

$a \leq (h(c)/g(c)) \leq b$ as $c \nearrow +\infty$.

Proof.

We need to show that if Conditions 1–3 hold, then for all $\epsilon > 0$ and $\eta > 0$, there exists $C(\epsilon, \eta)$ such that for all $c \geq C(\epsilon, \eta)$,

$$\frac{1}{\sigma_c^2}\mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\sigma_c\right] < \eta.$$

By Condition 3, we know that there exists $\bar{C}$ such that for all $c > \bar{C}$, $\sigma_c^2 \in (\sigma^2/2, 3\sigma^2/2)$. Let $\xi = \sigma/\sqrt{2}$. Then, for all $c \geq \bar{C}$, we have that $\sigma_c^2 \geq \xi^2$.

Now, by Condition 4 and the assumption that $n(c) \nearrow +\infty$ as $c \nearrow +\infty$, there exists $C(\epsilon, \eta) \geq \bar{C}$ such that

$$\mathbb{E}\left[X_{c,1}^2; |X_{c,1}| > \epsilon\sqrt{n(C(\epsilon, \eta))}\,\xi\right] < \eta\xi^2 \quad \forall c > 0.$$

Then, for all $c \geq C(\epsilon, \eta)$,

$$\begin{aligned}
\frac{1}{\sigma_c^2}\mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\sigma_c\right] &\leq \frac{1}{\xi^2}\mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\sigma_c\right] \\
&\leq \frac{1}{\xi^2}\mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(c)}\,\xi\right] \quad \text{since } c \geq \bar{C} \\
&\leq \frac{1}{\xi^2}\mathbb{E}\left[X_{c,1}^2; |X_{c,1}| \geq \epsilon\sqrt{n(C(\epsilon, \eta))}\,\xi\right] \\
&< \frac{1}{\xi^2} \cdot \eta\xi^2 = \eta.
\end{aligned}$$

Since $\epsilon > 0$ and $\eta > 0$ are arbitrary, we have proved the lemma. $\square$

We are now ready to state and prove a CLT for our estimator $\hat{\alpha}$. In particular, we will apply the lemma to the family of centered indicator functions:

$$\left\{I(\bar{X}_{m(c)}(Z_i) \leq 0) - \mathbb{P}(\bar{X}_{m(c)}(Z_1)) : i = 1, 2, \ldots, n(c), c > 0\right\}.$$

Condition 4 is immediate since the family is bounded between 0 and 1. Moreover, as $c \nearrow +\infty$, $\mathrm{Var}(I(\bar{X}_{m(c)}(Z) \leq 0)) = \mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0) \cdot \mathbb{P}(\bar{X}_{m(c)}(Z) > 0) \to \sigma^2$, where

$\sigma^2 \triangleq \alpha(1-\alpha)$.

Proposition 2.2.1 *Suppose that*

1. $\mathbb{P}(\mu(Z) = 0) = 0$;

2. *for all* $z \in \Gamma$ *and all* $\theta \in \mathbb{R}$, $\mathbb{E}\exp(\theta X(z)) < \infty$;

3. $l = \inf_{z\in\Gamma} \eta(z) > 0$.

Then, if $(\log c)^{-1}m(c) \to a \geq (2l)^{-1}$, $n(c) = c/m(c)$, *and* $n(c) \nearrow +\infty$, *then as* $c \nearrow +\infty$,

$$\sqrt{n(c)}(\hat{\alpha}(c) - \alpha) \Rightarrow \sqrt{\alpha(1-\alpha)}\mathrm{N}(0,1).$$

If $(\log c)^{-1}m(c) \to a \in [0, (2l)^{-1})$, $m(c) \nearrow +\infty$, *and* $n(c) = c/m(c)$, *then as* $c \nearrow +\infty$,

$$e^{m(c)l}(\hat{\alpha}(c) - \alpha) \Rightarrow 1.$$

Proof.

Define $\chi_i(m) \triangleq I(\bar{X}_m(Z_i) \leq 0)$. Note that

$$\hat{\alpha} - \alpha = \frac{1}{n}\sum_{i=1}^{n} \hat{\chi}_i(m) + \mathbb{P}(\bar{X}_m(Z) \leq 0) - \alpha,$$

where $\hat{\chi}_i(m) = \chi_i(m) - \mathbb{P}(\bar{X}_m(Z) \leq 0)$ is the centered version of $\hat{X}_i(m)$. Then,

$$\hat{\alpha} - \alpha = n^{-1/2}\left(\sum_{i=1}^{n} \frac{\hat{\chi}_i(m)}{\sqrt{n}}\right) + \mathbb{P}(\bar{X}_m(Z) \leq 0) - \alpha.$$

Observe that for each i,

$$\begin{aligned} \mathbb{E}|\hat{\chi}_i(m)|^2 &= \mathrm{Var}[I(\bar{X}_m(Z) \leq 0)] \\ &= \mathbb{P}(\bar{X}_m(Z) \leq 0) \cdot \mathbb{P}(\bar{X}_m(Z) > 0) < \frac{1}{4}. \end{aligned}$$

from which it follows that $\{\hat{\chi}_i(m(c)) : i = 1, \ldots, n(c), c > 0\}$ is uniformly integrable. By Lemma 2.2.1, the Lindeberg-Feller theorem holds here. That is, as $c \to +\infty$,

$$\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c))}{\sqrt{n(c)}} \Rightarrow \sigma \mathrm{N}(0, 1),$$

where $\sigma = \alpha(1-\alpha)$.

So, if $(\log c)^{-1} m(c) \to a \geq (2l)^{-1}$, $n(c) = c/m(c)$, and $n(c) \nearrow +\infty$, then, as $c \nearrow +\infty$,

$$\sqrt{n(c)} n(c)^{-1/2} \left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c))}{\sqrt{n(c)}} \right) \Rightarrow \sqrt{\alpha(1-\alpha)} \mathrm{N}(0, 1) \tag{2.3}$$

and

$$\sqrt{n(c)} (\mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0) - \alpha) = \sqrt{c/m(c)} e^{-m(c)l} \cdot e^{m(c)l} (\mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0) - \alpha) \to 0 \tag{2.4}$$

since $\sqrt{c/m(c)} e^{-m(c)l} \to 0$ and $e^{m(c)l}(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0) - \alpha) \to 1$ (by (2.2).) By the converging together theorem [5, p.340], we obtain the first result by combining the converging results (2.3) and (2.4).

On the other hand, If $(\log c)^{-1} m(c) \to a \in [0, (2l)^{-1})$, $m(c) \nearrow +\infty$, and $n(c) = c/m(c)$, then, as $c \nearrow +\infty$,

$$e^{m(c)l} n(c)^{-1/2} \left(\sum_{i=1}^{n} \frac{\hat{\chi}_i(m)}{\sqrt{n}} \right) \Rightarrow 0 \quad (\text{since } e^{m(c)l} \sqrt{m(c)/c} \to 0) \tag{2.5}$$

and

$$e^{m(c)l} (\mathbb{P}(\bar{X}_m(Z) \leq 0) - \alpha) \to 1 \quad (\text{by } 2.2.) \tag{2.6}$$

Again, by the converging together theorem, we obtain the second result by combining the converging results (2.5) and (2.6). □

2.3 Methodology

We will provide in this section some methodology for the implementation of the estimators. $\hat{\alpha}_c$ of the distribution function of $\mathbb{E}(X|Z)$.

Let us first consider the finite state space setting as in Section 2.1. To achieve the asymptotically fastest rate of convergence, we choose the sample sizes $\{m_c(z) : z \in \Gamma\}$ according to Proposition 2.1.1; namely, $m_c(z) \triangleq \eta(z)^{-1}/\gamma$, where $\gamma = \sum_{z\in\Gamma} \eta(z)^{-1}$. One difficulty with this choice of $m_c(\cdot)$ is that $\eta(z)$ is not known *a priori* and it is non-trivial to estimate it with a high degree of accuracy.

However, we noted earlier that γ is largely determined by those $\eta(\cdot)^{-1}$'s with small magnitudes. For conditioned r.v. $X(z)$, corresponding to $\eta(\cdot)$ that are sufficiently small, the normal random variable provides a good approximation to the sample mean, $\bar{X}_{m(z)}$, of the X's for large $m(z)$. Since η for a $\mathrm{N}(\mu, \sigma^2)$ r.v. is exactly equal to $(1/2)(\mu/\sigma)^2$, this suggests using $(1/2)(\mu(z)/\sigma(z))^2$ as a first-cut approximation to $\eta(z)$. For target variable with large $\eta(\cdot)$, $(1/2)(\mu(z)/\sigma(z))^2$ typically will not give a good approximation to its η. However, such $\eta(\cdot)$'s have small contribution towards γ. Hence, for each $z \in \Gamma$, we would conjecture spending a small portion of the computational budget to estimate $(1/2)(\mu(z)/\sigma(z))^2$ and using it as an approximate estimate to $\eta(z)$.

The algorithm below gives a practical methodology for the implementation of $\hat{\alpha}$.

Algorithm 2.3.1

Step 0. *Initialization.* Input c, $0 < r < 1$, and $\{p(z) : z \in \Gamma\}$.

Step 1. *Estimate the* $\eta(\cdot)$. Let $m^* = c^r/|\Gamma|$. For each $z \in \Gamma$, we sample m^* X's according to $\mathbb{P}(X \in \cdot|Z = z)$ and set

$$\hat{\eta}(z) \triangleq \frac{1}{2} \frac{\bar{X}_{m^*}(z)^2}{\frac{1}{m^*-1}\sum_{j=1}^{m^*}(X_j(z) - \bar{X}_{m^*}(z))^2}.$$

Step 2. *Estimate the optimal* $m_c(\cdot)$. Let $\gamma \triangleq \sum_{z\in\Gamma} \hat{\eta}(z)^{-1}$. Set $m(z) \triangleq \eta(z)^{-1}/\gamma$.

Step 3. *Determine* $\hat{\alpha}$. Set

$$\hat{\alpha} = \sum_{z\in\Gamma} I\left(\bar{X}_{m(z)}(z) \leq 0\right);$$

i.e., we sample $m(z)$ X's under the d.f. $\mathbb{P}(X \in \cdot|Z = z)$ and take $\hat{\alpha}$ as the weighted sum of the indicator functions with the weights being equal to $p(z)$'s.

Numerical results using this algorithm will be presented in the next section.

Let's now consider the case for an infinite state space setting as in Section 2.2. We appeal to the CLT developed in Proposition 2.2.1 to guide us in setting up the c.i. for the estimator,

$$\hat{\alpha}_c \triangleq \frac{1}{n(c)} \sum_{i=1}^{n(c)} I\left(\frac{1}{m(c)} \sum_{j=1}^{m(c)} X_j(Z_i) \leq 0\right).$$

The absolute fastest rate of convergence to α can be attained by choosing $m(c) = (2l)^{-1}\log c$ and $n(c) = c/m(c)$, according to Proposition 2.2.1. However, the values of the $\eta(\cdot)$ (and the value of l in particular) are not known *a priori*. Also, since the state space of the conditioning random element is assumed to be very large, it's impractical to estimate all the $\eta(\cdot)$'s.

In order to construct a c.i. for the estimator, we can alternatively set $m(c) \sim ac^{\nu}$ for some $\nu \in (0,1)$ and $a > 0$ chosen by the user and set $n(c) = c/m(c)$. In this case, as $c \nearrow +\infty$, the convergence rate of the estimator will be $O(c^{(1-\nu)/2})$, by Proposition 2.2.1. Note that there is a trade-off for the size of ν. For ν close to zero, the convergence rate is fast but we need c large for the CLT to be a good approximations; on the other hand, for ν close to 1, the convergence rate will be slow but we do not require c to be very large to obtain c.i.'s with the correct coverage probability.

The whole algorithm is summarized as follows:

Algorithm 2.3.2

Step 0. *Initialization.* Input c, ν, and a.

Step 1. *Determine the sample sizes.* Set $(m,n) \triangleq (ac^{\nu}, a^{-1}c^{1-\nu})$.

Step 2. *Determine* $\hat{\alpha}$. Set

$$\hat{\alpha} \triangleq \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \leq 0\right).$$

Step 3. *Form the* $(1-\xi)\times 100\%$ *confidence interval for* $\hat{\alpha}$. A consistent estimate of the standard error (s.e.), $s_{\hat{\alpha}}$, of $\hat{\alpha}$ is $\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{n}}$ since $\hat{\alpha} \sim \alpha$. The c.i., according to Proposition 2.2.1, is equal to

$$\left[\hat{\alpha} - z_{\xi/2}s_{\hat{\alpha}}, \hat{\alpha} + z_{\xi/2}s_{\hat{\alpha}}\right],$$

where $z_{\xi/2}$ is the $\xi/2$-quantile of a N(0, 1) r.v.

2.4 Numerical Results

In this section, we will report the numerical results of the algorithms proposed in the last section. The example we have used is as follows: Assume that the conditioning random element $Z \stackrel{\mathcal{D}}{=}$ binomial(10, 0.4) and that conditioned on $Z = z$, $X \stackrel{\mathcal{D}}{=} N(z/2 - 2.3, 1)$. In Algorithm 2.3.1, we set $r = 0.75$. The exact value of $\alpha \triangleq \mathbb{P}(\mathbb{E}(X|Z) \leq 0)$ is given by

$$\sum_{z=0}^{4} \mathbb{P}(Z = z) = \sum_{z=0}^{4}\binom{10}{z}0.4^{z}0.6^{10-z} = 0.6331.$$

We replicate the estimator 200 times. Denote by $\{\hat{\alpha}_i(c) : i = 1, \ldots, 200\}$ the values of the 200 replicates of the estimator $\hat{\alpha}$, given that the computational budget is equal to c. We estimate the mean, standard error, bias, and MSE of the estimator as follows:

mean: set $\bar{\alpha}(c) \triangleq (1/200)\sum_{i=1}^{200}\hat{\alpha}_i(c)$;

standard error: set $s_{\hat{\alpha}}(c) \triangleq \sqrt{(200-1)^{-1}\sum_{i=1}^{200}(\hat{\alpha}_i(c) - \bar{\alpha}(c))^2}$;

bias: set $b_{\hat{\alpha}}(c) \triangleq \bar{\alpha}(c) - \alpha$, where α is the exact theoretical value;

MSE: set $\text{MSE}_{\hat{\alpha}}(c) \triangleq (200)^{-1}\sum_{i=1}^{200}(\hat{\alpha}_i(c) - \alpha)^2$.

c	mean	s.d.	bias	log(MSE)	log(MSE)/c
100	0.6044	0.0930	-0.0287	-4.6634	-0.0466
200	0.6258	0.0535	-0.0073	-5.8416	-0.0292
300	0.6333	0.0121	0.0002	-8.8384	-0.0295
400	0.6309	0.0214	-0.0022	-7.2306	-0.0181
500	0.6331	0.0004	0.0000	-9.7549	-0.0195
600	0.6329	0.0029	-0.0002	-11.7212	-0.0195

Table 2.1: Numerical Results for Algorithm 2.3.1

The table above summarizes the numerical results. Recall from Proposition 2.1.1 that $\liminf_{c\to\infty} \log(\text{MSE}_{\hat{\alpha}}(c))/c \geq -\gamma$. The theoretical value of $-\gamma$ of this example is equal to -0.012 and the table above shows that, empirically, log(MSE)/c is converging to about -0.020 which is quite close to its theoretical value.

Let's now apply Algorithm 2.3.2 to the same problem. We choose $\nu = 0.2$ in this example. The table above summarizes the numerical results.

c	mean	s.d.	bias	log(MSE)	log(MSE)/c
1024	0.6454	0.1226	0.0123	-4.1925	-0.0041
2048	0.6475	0.0963	0.0144	-4.6636	-0.0023
4096	0.6454	0.0784	0.0123	-5.0726	-0.0012
8192	0.6386	0.0589	0.0055	-5.6606	-0.0007
16384	0.6383	0.0468	0.0052	-6.1185	-0.0004
32768	0.6347	0.0352	0.0016	-6.6968	-0.0002
65536	0.6327	0.0251	-0.0004	-7.3742	-0.0001
131072	0.6332	0.0189	0.0001	-7.9468	-0.0001

Table 2.2: Numerical Results for Algorithm 2.3.2

From Table 2.2, we see that the log(MSE)/c is a lot larger than -0.02 and the MSE of the algorithm is larger than that of Algorithm 2.3.1. To deduce the rate of convergence of the estimator, we plot the log(MSE(c)) vs. $\log c$ and the plot (Figure 2.1) turns out to be linear. This suggests that MSE(c) $\sim Vc^{\lambda}$ for some constants V and λ. We can estimate $\log V$ and λ by the y-intercept of the plot and its slope respectively. The theoretical slope and intercept are equal to $-(1-0.2) = -0.8$ and 0.7680 resp.; whereas the empirical slope

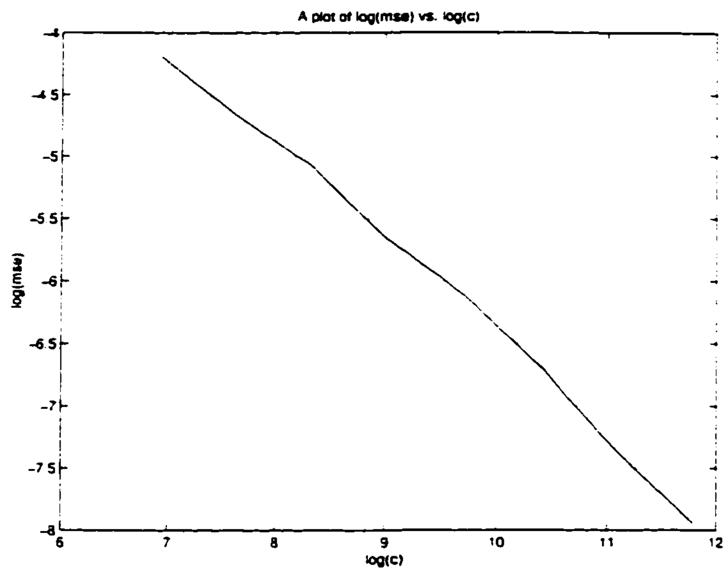


Figure 2.1: Distribution Function Estimator for the Discrete Case Example

and intercept are equal to −0.78 and 1.29 resp., which match quite well the theoretical values.

Out of the 200 experiments, we tested the number of times, N, the confidence intervals covered the true value. The corresponding estimated coverage probability is then set to $\hat{p} \stackrel{\Delta}{=} N/200$. The standard error of the estimated coverage probability is given by $\sqrt{\hat{p}(1-\hat{p})/200}$ and is expressed inside the parenthesis beside the corresponding probability in the Table 2.3. All coverage probabilities converge to the correct values.

c	90% cov.		95% cov.		99% cov.	
1024	0.89	(0.02)	0.91	(0.02)	0.97	(0.01)
2048	0.89	(0.02)	0.93	(0.02)	0.98	(0.01)
4096	0.89	(0.02)	0.91	(0.02)	0.99	(0.01)
8192	0.91	(0.02)	0.95	(0.02)	0.98	(0.01)
16384	0.88	(0.02)	0.94	(0.02)	0.98	(0.01)
32768	0.90	(0.02)	0.95	(0.02)	0.99	(0.01)
65536	0.90	(0.02)	0.97	(0.01)	0.99	(0.01)
131072	0.91	(0.02)	0.97	(0.01)	1.00	(0.00)

Table 2.3: Confidence Interval Coverage Probabilities of the Discrete Case Example

Chapter 3

Estimators for Continuously Conditioned Expectations

In this chapter, we examine the case where the conditioning random element is more general. This includes the case where Z is a (vector-valued) random variable living on a continuous state space or the case where Z is the sample path of a stochastic process. In the latter case. the sample space Γ will be equal to the space of continuous functions; i.e.. $\Gamma = C[0.\infty)$ or that of functions that are right-continuous with left-hand limits; i.e.. $\Gamma = D[0.\infty)$ (see [4] for details.)

We will use the same estimator as that proposed in Chapter 2: namely.

$$\hat{\alpha} \triangleq \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \leq 0\right).$$

Specifically, we sample n Z's via Monte Carlo simulation. For each sampled Z_i. we sample m X's according to the conditional distribution $\mathbb{P}(X \in \cdot|Z = Z_i)$ and set the indicator function appropriately. The estimator is then set to be the sample average of the indicator functions.

For this estimator, it can be shown under certain technical conditions, its bias term can be expressed as a polynomial in powers of $1/m$. Having established these results. we will derive the optimal rate of convergence of the estimator and the corresponding CLTs. As a by-product, we will be able to identify the asymptotically optimal choice of m and n.

3.1 Distribution Function Estimation

3.1.1 Bias of the Estimator

We will examine closely the bias term of the estimator $\hat{\alpha}$. We will consider separately the cases in which the conditional probability distribution is normal or non-normal. In any case, we will first show that, under suitable technical conditions, the bias term is equal to $a/m + o(1/m)$ for some constant a that is identifiable. Then, by assuming strong conditions, we will demonstrate a stronger result, namely, the bias can be expressed as a polynomial of $1/m$.

Normal Conditional Distribution Function

Define $Y(z) \triangleq \mu(z)/\sigma(z)$ for each possible $z \in \Gamma$. Assume that the density function, $f_Y(\cdot)$, of Y exists. We will call Y the *ICCV* (short form for Inverse of Conditional Coefficient of Variation.) This variable and its density function will play a critical role in the results that follow. Denote by $f_Y^{(k)}(\cdot)$ the k-th derivative of $f_Y(\cdot)$. The following proposition states that if the conditional distribution of X given Z is normal, then, except for the remainder term, the bias is inversely proportional to m.

Let C_b^k be the class of k-times continuously differentiable functions whose derivatives of order less than or equal to k are bounded globally.

Proposition 3.1.1 *Assume that*

1. $\mathbb{P}(X \le \cdot | Z) = \Phi(\frac{\cdot - \mu(Z)}{\sigma(Z)})$;
2. $f_Y(\cdot) \in C_b^2$.

Then, the bias $\mathbb{E}\hat{\alpha}(m,n) - \alpha = f_Y^{(1)}(0)/(2m) + o(1/m)$.

Proof. Observe that the bias of the estimator is given by

$$\begin{aligned}
\mathbb{E}\hat{\alpha}(m,n)-\alpha &= \mathbb{E}\mathbb{P}\left(\frac{1}{m}\sum_{j=1}^{m}X_j(Z)\leq 0\right)-\alpha \\
&= \mathbb{E}\mathbb{P}\left(\sqrt{m}\frac{\frac{1}{m}\sum_{j=1}^{m}X_j(Z)-\mu(Z)}{\sigma(Z)}\leq -\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right)-\alpha \\
&= \mathbb{E}\Phi\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right)-\alpha
\end{aligned}$$

By a change of variable,

$$\begin{aligned}
\mathbb{E}\hat{\alpha}(m,n)-\alpha &= \int_{\mathbb{R}}\Phi(-\sqrt{m}y)f_Y(y)\,dy-\alpha \\
&= \int_{-\infty}^{0}(1-\Phi(\sqrt{m}y))f_Y(y)\,dy+\int_0^{\infty}\Phi(-\sqrt{m}y)f_Y(y)\,dy-\alpha \\
&= \int_0^{\infty}\Phi(-\sqrt{m}y)[f_Y(y)-f_Y(-y)]\,dy \\
&= \frac{1}{\sqrt{m}}\int_0^{\infty}\Phi(-y)\left[f_Y(y/\sqrt{m})-f_Y(-y/\sqrt{m})\right]\,dy.
\end{aligned}$$

By the Taylor expansion, we have that

$$\begin{aligned}
f_Y(y/\sqrt{m}) &= f_Y(0)+\frac{y}{\sqrt{m}}f_Y^{(1)}(0)+\frac{1}{2}\frac{y^2}{m}f_Y^{(2)}(\xi^+_{y/\sqrt{m}}) \\
f_Y(-y/\sqrt{m}) &= f_Y(0)-\frac{y}{\sqrt{m}}f_Y^{(1)}(0)+\frac{1}{2}\frac{y^2}{m}f_Y^{(2)}(\xi^-_{y/\sqrt{m}})
\end{aligned}$$

for some $0<\xi^+_{y/\sqrt{m}}<y/\sqrt{m}$ and $-y/\sqrt{m}<\xi^-_{y/\sqrt{m}}<0$. Hence,

$$\left|\mathbb{E}\hat{\alpha}(m,n)-\alpha-\frac{2f_Y^{(1)}(0)}{m}\int_0^{\infty}y\cdot\Phi(-y)\,dy\right|\leq\frac{k_1}{m^{3/2}}\int_0^{\infty}y^2\cdot\Phi(-y)\,dy$$

for some $k_1>0$.

Moreover,

$$\begin{aligned}
&\int_0^\infty y\Phi(-y)\,dy \\
&= \frac{1}{\sqrt{2\pi}}\int_0^\infty \left(\int_{-\infty}^{-y} e^{-\frac{1}{2}x^2}dx\right)\, y\,dy \\
&= \frac{1}{\sqrt{2\pi}}\int_0^\infty \left(\int_0^{x} ydy\right)\, e^{-\frac{1}{2}x^2}\,dx \\
&= \frac{1}{4}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty x^2 e^{-\frac{1}{2}x^2}\,dx = \frac{1}{4}.
\end{aligned}$$

Similarly, one can show that $\int_0^\infty y^2\Phi(-y)\,dy = \frac{1}{3}\sqrt{\frac{2}{\pi}}$. Hence,

$$\mathbb{E}\hat{\alpha}(m,n) - \alpha = \frac{f_Y^{(1)}(0)}{2m} + o\left(\frac{1}{m}\right). \quad \square$$

Remark: Condition 2 can be relaxed somewhat to $\sup_{-x<t<x} |f_Y^{(2)}(t)| = O(e^{\gamma x^2})$ for some $\gamma < \frac{1}{2}$.

Notice that Propositions 3.1.1 specifies that the bias term is of order m^{-1}. We will actually show that under certain technical conditions, the error term can be expressed as a polynomial in powers of m^{-1}.

Again, we consider, in this subsection, the case in which the conditional distribution of X given Z is normal. Let's first consider a result which will be useful for the proofs to follow. This result can be easily proved by induction; similar results can be found in [21, p.183].

Lemma 3.1.1 *For $k \geq 0$,*

$$\begin{aligned}
\int_0^\infty x^{2k} e^{-x^2/2}\,dx &= \frac{(2k)!}{2^k k!}\sqrt{\frac{\pi}{2}} \\
\int_0^\infty x^{2k+1} e^{-x^2/2}\,dx &= 2^k k!.
\end{aligned}$$

Proof.

It is well-known (for example, from [21]) that for $k = 0, 1, 2, \ldots$,

$$\mathbb{E}[\mathrm{N}(0,1)^{2k}] = \frac{(2k)!}{2^k k!}.$$

Since x^{2k} is an even function of x, we have that

$$\begin{aligned}
\int_0^\infty x^{2k} e^{-x^2/2}\, dx &= \frac{1}{2}\int_{-\infty}^\infty x^{2k} e^{-x^2/2}\, dx \\
&= \sqrt{\frac{\pi}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty x^{2k} e^{-x^2/2}\, dx = \frac{(2k)!}{2^k k!}\sqrt{\frac{\pi}{2}}
\end{aligned}$$

which is just the first result.

We will now prove the second result by induction on k. In particular, denote by $G(n)$ the integral $\int_0^\infty x^n e^{-x^2/2}\, dx$. It is clear that $G(1) = 1$. Suppose that $G(2k-1) = 2^{k-1}(k-1)!$ for some integer $k \geq 0$. Let's now consider $G(2k+1)$. By integration by parts, we can show that

$$\begin{aligned}
G(2k+1) &= -\int_0^\infty x^{2k}\, d(e^{-x^2/2}) \\
&= x^{2k} e^{-x^2/2}\Big|_\infty^0 + \int_0^\infty (2k) \cdot e^{-x^2/2} x^{2k-1}\, dx \\
&= 0 + (2k) \cdot G(2k-1) \quad \text{by L'Hopital's Rule} \\
&= 2^k k! \quad \text{by induction hypothesis.}
\end{aligned}$$

Hence, the second result is established. □

The proposition below specifies that the bias term is of order m^{-1} even when the conditional distribution of X given Z is non-normal.

Proposition 3.1.2 *Assume that for some $k \geq 1$,*

1. $\mathbb{P}(X \leq \cdot | Z) = \Phi\left(\frac{\cdot - \mu(Z)}{\sigma(Z)}\right)$;
2. $f_Y(\cdot) \in C_b^{2k}$.

Then, the bias term is given by

$$\mathbb{E}\hat{\alpha}(m,n) - \alpha = \sum_{i=1}^{k} \frac{f_Y^{(2i-1)}(0)}{2^i i!} \cdot \frac{1}{m^i} + o(1/m^k).$$

Proof.

Recall that

$$\mathbb{E}\hat{\alpha}(m,n) - \alpha = \frac{1}{\sqrt{m}} \int_0^\infty \Phi(-y) \left[f_Y\left(\frac{y}{\sqrt{m}}\right) - f_Y\left(\frac{-y}{\sqrt{m}}\right) \right] dy.$$

With the smoothness assumptions on $f_Y(\cdot)$, we have, by Taylor expansion, that

$$\begin{aligned} f_Y\left(\frac{y}{\sqrt{m}}\right) &= \sum_{i=0}^{2k-1} \frac{1}{i!} f_Y^{(i)}(0) \left(\frac{y}{\sqrt{m}}\right)^i + \frac{1}{(2k)!} f_Y^{(2k)}(\xi^+_{y/\sqrt{m}}) \left(\frac{y}{\sqrt{m}}\right)^{2k} \\ f_Y\left(\frac{-y}{\sqrt{m}}\right) &= \sum_{i=0}^{2k-1} \frac{1}{i!} f_Y^{(i)}(0) \left(\frac{-y}{\sqrt{m}}\right)^i + \frac{1}{(2k)!} f_Y^{(2k)}(\xi^-_{y/\sqrt{m}}) \left(\frac{y}{\sqrt{m}}\right)^{2k}. \end{aligned}$$

Hence,

$$\begin{aligned} f_Y\left(\frac{y}{\sqrt{m}}\right) - f_Y\left(\frac{-y}{\sqrt{m}}\right) &= 2\sum_{i=1}^{k} \frac{f_Y^{(2i-1)}(0)}{(2i-1)!} \left(\frac{y}{\sqrt{m}}\right)^{2i-1} \\ &\quad + \frac{1}{(2k)!} \left(f_Y^{(2k)}(\xi^+_{y/\sqrt{m}}) - f_Y^{(2k)}(\xi^-_{y/\sqrt{m}}) \right) \frac{y^{2k}}{m^k}. \end{aligned}$$

Now,

$$\begin{aligned} & \int_0^\infty \Phi(-y) y^{2i-1}\, dy \\ = {} & \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{2i} x^{2i} e^{-x^2/2}\, dx \\ = {} & \frac{1}{2i} \frac{1}{\sqrt{2\pi}} \cdot \frac{(2i)!}{2^i i!} \sqrt{\frac{\pi}{2}} \quad \text{by Lemma 3.1.1} \\ = {} & \frac{(2i-1)!}{2^{i+1} i!}. \end{aligned}$$

So, by the boundedness assumption on $f_Y^{(2k)}(\cdot)$, there exists a $\kappa > 0$ such that

$$\left|\mathbb{E}\hat{\alpha}(m,n) - \alpha - \sum_{i=1}^{k} \frac{f_Y^{(2i-1)}(0)}{2^i i!} \cdot \frac{1}{m^i}\right| \leq \frac{2^k k!}{\sqrt{2\pi}\cdot(2k+1)} \frac{\kappa}{(2k)!} \frac{1}{m^{k+\frac{1}{2}}} = o\left(\frac{1}{m^k}\right). \quad \square$$

Non-normal Conditional Distribution Function

Proposition 3.1.1 covers the case in which given Z, the conditional distribution of X is normal. It turns out that under certain mild regularity conditions, the results in Proposition 3.1.1 also hold for the case in which given Z, the conditional distribution of X is not necessarily normal. For the details of the proof, we need some results about the asymptotic expansions related to the central limit theorem. One of these results is a generalization of the Berry-Esseen inequality.

Theorem 3.1.1 *Let $X_1, \ldots, X_m$ be independent identically distributed random variables with $\mathbb{E}X_1 = \mu$, $\mathrm{Var}X_1 = \sigma^2 > 0$, and $\mathbb{E}|X_1 - \mu|^3 < \infty$. Then, there exists a universal constant a such that*

$$\left|\mathbb{P}\left(\frac{1}{\sigma\sqrt{m}}\sum_{j=1}^{m}(X_j - \mu) < x\right) - \Phi(x)\right| \leq a\sigma^{-3}\mathbb{E}|X_1 - \mu|^3 m^{-1/2}(1+|x|)^{-3}$$

for all x.

The proof of Theorem 3.1.1 can be found in [25, p.168]. The universal constant, a, is given by $a = 2(5 + 2e^{1/2}\pi^{-1/2}\Gamma(3/2))$, so that $a < 16.5$. In particular, the constant a is independent of the distribution of the random variable X_1.

Define $\mathbb{E}(|X - \mu(Z)|^k)/\sigma(Z)^k$ by $\beta_k(Z)$ for each $k = 1, 2, \ldots$. As expected, we apply Theorem 3.1.1 to the conditional distribution of X, given Z. Specifically, we have the following result:

$$\left|\mathbb{P}\left(\bar{X}_m(Z) \leq 0 | Z\right) - \Phi\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right)\right| \leq a\beta_3(Z)m^{-1/2}(1 + |\sqrt{m}\mu(Z)/\sigma(Z)|)^{-3} \quad \text{a.s.}$$

For ease of disposition, denote the right hand side of the above inequality by ξ_m. This implies that the bias term $\mathbb{E}\hat{\alpha}(m,n) - \alpha$ satisfies the inequality

$$\left(\mathbb{E}\Phi\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right) - \alpha\right) - \mathbb{E}(\xi_m) \leq \mathbb{E}\hat{\alpha}(m,n) - \alpha \leq \left(\mathbb{E}\Phi\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right) - \alpha\right) + \mathbb{E}(\xi_m).$$

We have seen in Proposition 3.1.1 that $\mathbb{E}\Phi(-\sqrt{m}\mu(Z)/\sigma(Z)) - \alpha = \Theta(1/m)$ [1]. Let's now take a closer look at the other term, $\mathbb{E}(\xi_m)$.

For notational purposes, for any two real-valued functions $f(\cdot)$ and $g(\cdot)$, denote $(f\cdot g)(x) \triangleq f(x)\cdot g(x)$. Define $\tilde{\beta}_3(Y) \triangleq \mathbb{E}(\beta_3(Z)|Y)$. Assume further the technical condition that $f_Y(\cdot) \in C_b^2$. We have that

$$\begin{aligned}
\sqrt{m}\mathbb{E}(\xi_m)/a &= \mathbb{E}\left[\frac{\beta_3(Z)}{\left(1+\left|\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right|\right)^3}\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}(\beta_3(Z)|Y)}{(1+|\sqrt{m}Y|)^3}\right] \\
&= \int_{\mathbb{R}} (f_Y\cdot\tilde{\beta}_3)(y)\frac{1}{(1+|\sqrt{m}y|)^3}\,dy \\
&= \int_{-\infty}^{0} \frac{(f_Y\cdot\tilde{\beta}_3)(y)}{(1-\sqrt{m}y)^3}\,dy + \int_0^\infty \frac{(f_Y\cdot\tilde{\beta}_3)(y)}{(1+\sqrt{m}y)^3}\,dy \\
&= \int_0^\infty \frac{(f_Y\cdot\tilde{\beta}_3)(-y) + (f_Y\cdot\tilde{\beta}_3)(y)}{(1+\sqrt{m}y)^3}\,dy \\
&= \frac{1}{\sqrt{m}}\int_0^\infty \frac{1}{(1+u)^3}\left[(f_Y\cdot\tilde{\beta}_3)(u/\sqrt{m}) + (f_Y\cdot\tilde{\beta}_3)(-u/\sqrt{m})\right]\,du.
\end{aligned}$$

As in the proof of Proposition 3.1.1, we can write

$$\begin{aligned}
(f_Y\cdot\tilde{\beta}_3)(u/\sqrt{m}) + (f_Y\cdot\tilde{\beta}_3)(-u/\sqrt{m}) &= 2(f_Y\cdot\tilde{\beta}_3)(0) + (u/\sqrt{m})(f_Y\cdot\tilde{\beta}_3)^{(1)}(\xi^+_{u/\sqrt{m}}) \\
&\quad - (u/\sqrt{m})(f_Y\cdot\tilde{\beta}_3)^{(1)}(\xi^-_{u/\sqrt{m}}),
\end{aligned}$$

[1] We say that a function of c, $h(c)$, is $\Theta(g(c))$ if there exist positive finite constants a and b such that $a \leq |h(c)/g(c)| \leq b$ as $c \nearrow +\infty$.

where $0 < \xi^{+}_{u/\sqrt{m}} < u/\sqrt{m}$ and $-u/\sqrt{m} < \xi^{-}_{u/\sqrt{m}} < 0$.

Assume that $\tilde{\beta}_3(\cdot)$ is C_b^1. Then, there exists some k_2 such that

$$\begin{aligned}
\mathbb{E}\left[\frac{\beta_3(Z)}{\left(1+\left|\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right|\right)^3}\right] &\le \frac{(f_Y\cdot\tilde{\beta}_3)(0)}{\sqrt{m}} + \frac{k_2}{m}\int_0^\infty \frac{u\,du}{(1+u)^3} \\
&= \frac{(f_Y\cdot\tilde{\beta}_3)(0)}{\sqrt{m}} + \frac{k_2}{2m}.
\end{aligned}$$

Hence, we have proved that $\mathbb{E}(\xi_m) = \Theta(1/m)$. Since, we have also shown in Proposition 3.1.1 that $\mathbb{E}\Phi\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right) - \alpha = f_Y^{(1)}(0)/(2m)$, we have proved the following proposition:

Proposition 3.1.3 *Assume that*

1. $f_Y(\cdot) \in C_b^2$;

2. $\tilde{\beta}_3(\cdot) \in C_b^1$.

Then, the bias $|\mathbb{E}\hat{\alpha}(m,n) - \alpha| = \Theta(1/m)$.

Remark:

1. Again we can relax Condition 1 to $\sup_{-x<t<x}|f_Y^{(2)}(t)| = O(e^{\gamma x^2})$ for some $\gamma < 0.5$ and Condition 2 to $\sup_{-x<t<x}|(f_Y\cdot\tilde{\beta}_3)^{(1)}(t)| = O(x^p)$ for some $p < 1$.

2. We can, in fact, derive a stronger result than simply saying that $|\mathbb{E}\hat{\alpha}(m,n) - \alpha| = \Theta(1/m)$. We can actually show that bias$(\hat{\alpha}) \sim c/m$ for some $c \neq 0$. What's more, we are going to identify the constant c and even expand the bias term as a polynomial in powers of $1/m$.

In order to obtain a similar (polynomial) expansion of the bias term for the case in which the conditional distribution of X given Z is *not* normal, we need some results about the rate of convergence in the Central Limit Theorem (CLT). In particular, we'll quote some results about an asymptotic expansion in the CLT, namely the Edgeworth expansion. One such version expresses the difference between the distribution function of the normalized sample average and the normal distribution function in terms of a polynomial in powers of $m^{-1/2}$.

Consider a sequence of independent identically distributed r.v. $\{X_n : n = 1, 2, \cdots\}$. Suppose that $\mathbb{E}X_1 = 0$, $\mathrm{Var}X_1 = \sigma^2 > 0$. Set

$$V(x) = \mathbb{P}(X_1 < x), \quad \nu(t) = \mathbb{E}e^{itX_1}, \quad F_m(x) = \mathbb{P}\left(\sigma^{-1}m^{-1/2}\sum_{j=1}^{m} X_j < x\right).$$

for $x, t \in \mathbb{R}$.

Theorem 3.1.2 *If $\mathbb{E}|X_1|^k < \infty$ for some integer $k \geq 3$, then the following inequality holds for all x and m:*

$$\begin{aligned}
&\left|F_m(x) - \Phi(x) - \sum_{\nu=1}^{k-2} Q_\nu(x) m^{\nu/2}\right| \\
\leq\ & c(k)\left\{\sigma^{-k} m^{-(k-2)/2}(1+|x|)^{-k}\int_{|y|\geq\sigma m^{1/2}(1+|x|)} |y|^k\, dV(y)\right. \\
&\quad + \sigma^{-k-1} m^{-(k-1)/2}(1+|x|)^{-k-1}\int_{|y|<\sigma m^{1/2}(1+|x|)} |y|^{k+1}\, dV(y) \\
&\quad \left. + \left(\sup_{|t|\geq\delta}|\nu(t)| + \frac{1}{2m}\right)^m m^{k(k+1)/2}(1+|x|)^{-k-1}\right\}.
\end{aligned}$$

Here $\delta = \sigma^2/(12\mathbb{E}|X_1|^3)$ and $c(k)$ is a positive universal constant depending only on k (and not on the distribution of X_1.) The functions $Q_\nu(x)$ are defined as

$$Q_\nu(x) = -(2\pi)^{-1/2}e^{-x^2/2}\sum H_{\nu+2s-1}(x)\prod_{l=1}^{\nu}\frac{1}{k_l!}\left(\frac{\gamma_{l+2}}{(l+2)!\sigma^{l+2}}\right)^{k_l}. \tag{3.1}$$

where γ_l is the cumulant of order l of the random variable X_1. The summation in (3.1) is extended over all non-negative integer solution $(k_1, k_2, \ldots, k_\nu)$ of the equation $k_1 + 2k_2 + \cdots + \nu k_\nu = \nu$, and $s = k_1 + k_2 + \cdots + k_\nu$. The functions $H_m(x)$ are (Hermite) polynomials of x defined as

$$H_m(x) \triangleq m!\sum_{k=0}^{[m/2]}\frac{(-1)^k x^{m-2k}}{k!(m-2k)!2^k}$$

for every integer $m \geq 0$.

The proof of this theorem of Osipov can be found. for instance. in [24. p.159–168].

Remark:

1. If X is a random variable with the characteristic function $c(\theta) \equiv \mathbb{E}e^{i\theta X}$. then the cumulant (or semi-invariant) of order l of this random variable is defined by the formal equality
$$\gamma_l \triangleq \frac{1}{i^l}\left[\frac{d^l}{d\theta^l}\log c(\theta)\right]_{\theta=0}.$$

2. The polynomial $H_m(x)$ is known as the Chebyshev-Hermite polynomial of degree m and is alternatively defined by the equality
$$H_m(x) \triangleq (-1)^m e^{x^2/2}\frac{d^m}{dx^m}e^{-x^2/2}.$$
In particular. $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$. $H_4(x) = x^4 - 6x^2 + 3$, $H_5(x) = x^5 - 10x^3 + 15x$.

3. The polynomials have an important orthogonal property [21. p.222], namely. that
$$\left.\begin{aligned} \textstyle\int_{-\infty}^{\infty} H_m(x)H_n(x)\phi(x)\,dx &= 0, \quad m \neq n \\ &= n!, \quad m = n. \end{aligned}\right\}$$

4. If the additional condition (Cramér's condition) $\limsup_{|t|\to\infty}|\nu(t)| < 1$ is satisfied. then $\sup_{|t|\geq\delta}|\nu(t)| < 1$ for $\delta > 0$, so that the multiplier $(\sup_{|t|\geq\delta}|\nu(t)| + 1/(2m))^m$ is less than m^{-p} for every $p > 0$ and sufficiently large m. One sufficient condition for Cramér's condition is that the distribution of X is absolutely continuous, as a result of the Riemann-Lebesgue Lemma (see [3. p.43].)

Before we proceed, let's consider the following three lemmas which play an important role in the proof of a later proposition. Henceforth, we adopt the notation that $(f \cdot g)(x) \triangleq f(x) \cdot g(x)$.

Lemma 3.1.2 *Let $\phi(x)$ be the density function of a* N(0, 1) *r.v. and $H_p(x)$ be defined as above. Then,*

$$\int_0^\infty (\phi \cdot H_p)(u)\, du = 0 \quad \textit{for } p = 2, 4, \ldots$$

and

$$\int_0^\infty (\phi \cdot H_p)(u) \cdot u\, du = 0 \quad \textit{for } p = 1, 3, \ldots.$$

Proof. For p even,

$$H_p(x) = p! \sum_{k=0}^{p/2} \frac{(-1)^k x^{p-2k}}{k!(p-2k)!2^k}.$$

So,

$$\begin{aligned}
\int_0^\infty (\phi \cdot H_p)(u)\, du &= \frac{p!}{\sqrt{2\pi}} \sum_{k=0}^{p/2} \frac{(-1)^k}{k!(p-2k)!2^k} \int_0^\infty e^{-x^2/2} \cdot x^{p-2k}\, dx \\
&= \frac{p!}{\sqrt{2\pi}} \sum_{k=0}^{p/2} \frac{(-1)^k}{k!(p-2k)!2^k} \cdot \sqrt{\frac{\pi}{2}} \cdot \frac{(p-2k)!}{2^{p/2-k}(p/2-k)!} \\
&= \frac{p!}{2^{p/2+1}} \sum_{k=0}^{p/2} \frac{(-1)^k}{k!(p/2-k)!} \\
&= \frac{1}{(p/2)!} \frac{p!}{2^{p/2+1}} (1 + (-1))^{p/2} = 0.
\end{aligned}$$

On the other hand, for p odd,

$$H_p(x) = p! \sum_{k=0}^{\frac{p-1}{2}} \frac{(-1)^k x^{p-2k}}{k!(p-2k)!2^k}.$$

So,

$$\begin{aligned}
\int_0^\infty (\phi \cdot H_p)(u) \cdot u\, du &= \frac{p!}{\sqrt{2\pi}} \sum_{k=0}^{\frac{p-1}{2}} \frac{(-1)^k}{k!(p-2k)!2^k} \int_0^\infty e^{-x^2/2} \cdot x^{p-2k+1}\, dx \\
&= \frac{p!}{\sqrt{2\pi}} \sum_{k=0}^{\frac{p-1}{2}} \frac{(-1)^k}{k!(p-2k)!2^k} \cdot \sqrt{\frac{\pi}{2}} \cdot \frac{(p+1-2k)!}{2^{\frac{p+1}{2}-k}(\frac{p+1}{2}-k)!}.
\end{aligned}$$

On simplification, we have that

$$\begin{aligned}\int_0^\infty (\phi \cdot H_p)(u) \cdot u\, du &= \frac{p!}{2^{\frac{p+1}{2}}} \sum_{k=0}^{\frac{p-1}{2}} \frac{(-1)^k}{k!(\frac{p-1}{2}-k)!} \\ &= \frac{p!}{2^{\frac{p+1}{2}} \left(\frac{p-1}{2}\right)!}(1+(-1))^{\frac{p-1}{2}} = 0. \quad \square\end{aligned}$$

The next lemma will make heavy use of the previous one. It helps us to identify the coefficients in the polynomial (in powers of $1/m$) of the bias expansion.

Lemma 3.1.3 *Let $\gamma(\cdot)$ be a real-valued function such that $\mathbb{E}|(\phi \cdot H_p)(-\sqrt{m}\mu(Z)/\sigma(Z)) \cdot \gamma(Z)| < \infty$. Let $Y \triangleq \mu(Z)/\sigma(Z)$, $\tilde{\gamma}(y) \triangleq \mathbb{E}(\gamma(Z)|Y=y)$ for $y \in \mathbb{R}$ and $h(\cdot) \triangleq (f_Y \cdot \tilde{\gamma})(\cdot)$. Finally, let $\phi(\cdot)$, $H_p(\cdot)$ and $(\phi \cdot H_p)(\cdot)$ be defined as in the previous lemma.*

1. *For $p = 1, 3, 5, \ldots$ and $k = 1, 2, 3, \ldots$, if $h(\cdot) \in C_b^{2k}$, then*

$$\mathbb{E}\left[(\phi \cdot H_p)(-\sqrt{m}\mu(Z)/\sigma(Z))\gamma(Z)\right] = \sum_{i=2}^{k} \frac{c_{p,i}}{m^i} + o(1/m^k),$$

where for $2 \le i \le k$,

$$c_{p,i} \triangleq -\frac{2h^{(2i-1)}(0)}{(2i-1)!} \int_0^\infty (\phi \cdot H_p)(u) \cdot u^i\, du$$

are constants depending only on p and $h^{(2i-1)}(0)$.

2. *For $p = 2, 4, 6, \ldots$ and $k = 1, 2, 3, \ldots$, if $h(\cdot) \in C_b^{2k-1}$, then*

$$\mathbb{E}\left[(\phi \cdot H_p)(-\sqrt{m}\mu(Z)/\sigma(Z))\gamma(Z)\right] = \sum_{i=2}^{k} \frac{d_{p,i}}{m^{i-\frac{1}{2}}} + o(1/m^{k-\frac{1}{2}}),$$

where for $2 \le i \le k$,

$$d_{p,i} \triangleq \frac{2h^{(2(i-1))}(0)}{(2(i-1))!} \int_0^\infty (\phi \cdot H_p)(u) \cdot u^{i-1}\, du$$

are constants depending only on p and $h^{(2(i-1))}(0)$.

Proof. We will only give the proof to the first part of the lemma: the second part of the lemma can be proved in an analogous fashion.

By the definition of $\tilde{\gamma}(y)$, we have that

$$\begin{aligned}
& \mathbb{E}\left[(\phi \cdot H_p)(-\sqrt{m}\mu(Z)/\sigma(Z))\gamma(Z)\right] \\
= & \mathbb{E}\left[(\phi \cdot H_p)(-\sqrt{m}Y)\tilde{\gamma}(Y)\right] \\
= & \int_{-\infty}^{\infty}(\phi \cdot H_p)(-\sqrt{m}y)h(y)\,dy \\
= & \int_{-\infty}^{0}(\phi \cdot H_p)(-\sqrt{m}y)h(y)\,dy + \int_{0}^{\infty}(\phi \cdot H_p)(-\sqrt{m}y)h(y)\,dy \\
= & \int_{0}^{\infty}(\phi \cdot H_p)(\sqrt{m}y)h(-y)\,dy - \int_{0}^{\infty}(\phi \cdot H_p)(\sqrt{m}y)h(y)\,dy \\
= & \frac{1}{\sqrt{m}}\int_{0}^{\infty}(\phi \cdot H_p)(u)\left[h\left(\frac{-u}{\sqrt{m}}\right) - h\left(\frac{u}{\sqrt{m}}\right)\right]du.
\end{aligned}$$

Now, by the smoothness assumptions on the function $h(\cdot)$, we have that

$$\begin{aligned}
h\left(\frac{-u}{\sqrt{m}}\right) &= \sum_{i=0}^{2k-1}\frac{1}{i!}h^{(i)}\left(\frac{-u}{\sqrt{m}}\right)^{i} + \frac{1}{(2k)!}h^{(2k)}(\xi^{-}_{u/\sqrt{m}})\left(\frac{-u}{\sqrt{m}}\right)^{2k} \\
h\left(\frac{u}{\sqrt{m}}\right) &= \sum_{i=0}^{2k-1}\frac{1}{i!}h^{(i)}\left(\frac{u}{\sqrt{m}}\right)^{i} + \frac{1}{(2k)!}h^{(2k)}(\xi^{+}_{u/\sqrt{m}})\left(\frac{u}{\sqrt{m}}\right)^{2k}
\end{aligned}$$

for some $-u/\sqrt{m} < \xi^{-}_{u/\sqrt{m}} < 0$ and $0 < \xi^{+}_{u/\sqrt{m}} < u/\sqrt{m}$. So,

$$\begin{aligned}
h\left(\frac{-u}{\sqrt{m}}\right) - h\left(\frac{u}{\sqrt{m}}\right) &= -2\sum_{i=1}^{k}\frac{h^{2i-1}(0)}{(2i-1)!}\left(\frac{-u}{\sqrt{m}}\right)^{2i-1} \\
&+ \frac{1}{(2k)!}\left(h^{(2k)}(\xi^{-}_{\frac{u}{\sqrt{m}}}) - h^{(2k)}(\xi^{+}_{\frac{u}{\sqrt{m}}})\right)\left(\frac{-u}{\sqrt{m}}\right)^{2k}.
\end{aligned}$$

Now since $\int_0^\infty(\phi \cdot H_p)(u)\cdot u\,du = 0$, by the boundedness of the $(2k)$-th derivative of $h(\cdot)$, there exists $c > 0$ such that

$$\left|\mathbb{E}\left[(\phi \cdot H_p)(-\sqrt{m}Y)\tilde{\gamma}(Y)\right] - \sum_{i=2}^{k}\frac{c_{p,i}}{m^{i}}\right| \leq \frac{c}{(2k)!}\int_0^\infty(\phi \cdot H_p)(u)u^{2k}\,du \cdot \frac{1}{m^{k+\frac{1}{2}}} = o(1/m^{k}).$$

where

$$c_{p,i} \triangleq -\frac{2h^{(2i-1)}(0)}{(2i-1)!}\int_0^\infty (\phi \cdot H_p)(u)\cdot u^i\,du.$$

The finiteness of $c_{p,i}$ is guaranteed by Lemma 3.1.1. □

The next simple lemma (and the final one before the main proposition) tells us the rate at which a "tail expectation" converges to zero.

Lemma 3.1.4 *Let X be a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for some $p > 1$, $\mathbb{E}|X|^{p+1} < \infty$. Then,*

$$\mathbb{E}\left[|X|^p; |X| > m\right] \le \frac{\mathbb{E}|X|^{p+1}}{m^{p+1}}$$

for all $m > 0$.

Proof.

Observe that since $|X|/m \ge 1$ on $\{|X| \ge m\}$, we can immediately conclude that

$$\begin{aligned}\mathbb{E}\left(|X|^p I(|X| \ge m)\right) &\le \mathbb{E}\left(\frac{|X|^{p+1}}{m^{p+1}} I(|X| \ge m)\right)\\ &\le m^{-p-1}\mathbb{E}|X|^{p+1}. \quad \square\end{aligned}$$

We are (finally) ready to present a proposition for the expansion of the bias term in the case in which the conditional distribution of X given Z is not normal.

Recall that $\gamma_p(z)$ is the p-th cumulant of X given $Z = z$. Given $\nu \ge 1$, define

- $\Sigma(\nu) \triangleq \{s : \exists(k_1, \ldots, k_\nu) \in \mathbb{Z}^+$ solving $k_1 + 2k_2 + \cdots + \nu k_\nu = \nu,\ s = k_1 + \cdots + k_\nu\}$
- $\kappa(\nu, s) \triangleq \{(k_1, \ldots, k_\nu) \in \mathbb{Z}^+ : k_1 + 2k_2 + \cdots + \nu k_\nu = \nu,\ s = k_1 + \cdots + k_\nu\}$ for each $s \in \Sigma(\nu)$
- $$\chi_{\nu,s}(z) \triangleq \sum_{(k_1,\ldots,k_\nu)\in\kappa(\nu,s)} \prod_{p=1}^{\nu} \frac{1}{k_p!}\left(\frac{\gamma_{p+2}(z)}{(p+2)!\sigma^{p+2}(z)}\right)^{k_p}$$

  for all $z \in \Gamma$ and $s \in \Sigma(\nu)$.

Notice that $\chi_{\nu,s}(z)$ is the coefficient of $H_{\nu+2s-1}$ in the Edgeworth expansion of $\mathbb{P}(\bar{X}_m(z) \leq x)$. Also, define $\tilde{\chi}_{\nu,s}(y) \triangleq \mathbb{E}(\chi_{\nu,s}(Z)|Y = y)$.

Proposition 3.1.4 *Assume that for some $k \geq 1$,*

1. $\mathbb{E}|X|^{2k+2} < \infty$;
2. *there exists a density $g(\cdot)$ and $\epsilon > 0$ such that $\mathbb{P}(X_1 \in \cdot|Z) \geq \epsilon g(\cdot)$ almost surely;*
3. $\tilde{\beta}_{2k+2}(\cdot) \in C_b^1$;
4. $f_Y(\cdot) \in C_b^{2k}$;
5. $(f_Y \cdot \tilde{\chi}_{1,1})(\cdot) \in C_b^{2k-1}$;

Assume further that for all $s = k_1 + \cdots + k_\nu$, where $(k_1, \ldots, k_\nu)$ is any non-negative integer solution of the equation $k_1 + 2k_2 + \cdots + \nu k_\nu = \nu$,

6. $(f_Y \cdot \tilde{\chi}_{2\nu,s})(\cdot) \in C_b^{2(k-\nu)}$ *for $s \in \Sigma(2\nu)$, $1 \leq \nu \leq k-1$,*
7. $(f_Y \cdot \tilde{\chi}_{2\nu+1,s})(\cdot) \in C_b^{2(k-\nu)-1}$ *for $s \in \Sigma(2\nu+1)$, $1 \leq \nu \leq k-1$;*

Then, there exist constants $a_2, \ldots, a_k$ such that

$$\mathbb{E}\hat{\alpha}(m,n) - \alpha = \frac{f_Y^{(1)}(0)}{2m} + \sum_{i=2}^{k} \frac{a_i}{m^i} + o(1/m^k).$$

Proof. Observe that $\mathbb{E}\hat{\alpha}(m,n) = \mathbb{E}\mathbb{P}(\bar{X}_m(Z) \leq 0|Z)$, where

$$\bar{X}_m(Z) \triangleq (1/m) \sum_{j=1}^{m} X_j(Z).$$

In view of Theorem 3.1.2 and Assumption 1, we have that

$$\begin{aligned} &\left| \mathbb{P}(\bar{X}_m(Z) \leq 0|Z) - \Phi\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right) - \sum_{\nu=1}^{2k-1} Q_\nu\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}; Z\right) m^{\nu/2} \right| \\ \leq\ & \frac{\varepsilon\left(m^{1/2}\left(1+\sqrt{m}\frac{|\mu(Z)|}{\sigma(Z)}\right); Z\right)}{m^{\frac{2k-1}{2}}\left(1+\sqrt{m}\frac{|\mu(Z)|}{\sigma(Z)}\right)^{2k+1}} \quad \text{a.s.,} \end{aligned} \tag{3.2}$$

where, by Theorem 4.1.2, $\epsilon(\cdot)$ is given by

$$\begin{aligned}
&\varepsilon(u;z)\\
=\;& c(2k+1)\left\{\sigma(z)^{-(2k+1)}\int_{|y-\mu(z)|>\sigma(z)u}|y-\mu(z)|^{2k+1}\,\mathbb{P}(X(z)\in dy)\right.\\
&+\frac{1}{\sigma(z)^{2k+1}}\int_{|y-\mu(z)|<\sigma(z)u}|y-\mu(z)|^{2k+1}\frac{|y-\mu(z)|/\sigma(z)}{u}\,\mathbb{P}(X(z)\in dy)\\
&\left.+\left(\sup_{|t|\geq\delta(z)}|\nu_z(t)|+\frac{1}{2m}\right)^m m^{2k^2+4k+1}\frac{1}{u}\right\},
\end{aligned}$$

where $\delta(z)=1/(12\sigma(z)\beta_3(z))$ and $\beta_3(z)=\mathbb{E}(|X-\mu(Z)|^3|Z=z)/\sigma(z)^3$.

In particular, denoting $\Gamma(z)\triangleq\left\{y:|y-\mu(z)|>\sigma(z)m^{1/2}\left(1+\sqrt{m}\frac{|\mu(z)|}{\sigma(z)}\right)\right\}$, we have that

$$\begin{aligned}
&\varepsilon\left(m^{1/2}\left(1+\sqrt{m}\frac{|\mu(z)|}{\sigma(z)}\right);z\right)\\
=\;& c(2k+1)\left\{\sigma(z)^{-(2k+1)}\int_{y\in\Gamma(z)}|y-\mu(z)|^{2k+1}\,\mathbb{P}(X(z)\in dy)\right.\\
&+\frac{1}{\sigma(z)^{2k+1}}\int_{y\notin\Gamma(z)}|y-\mu(z)|^{2k+1}\frac{|y-\mu(z)|/\sigma(z)}{m^{1/2}\left(1+\sqrt{m}\frac{|\mu(z)|}{\sigma(z)}\right)}\,\mathbb{P}(X(z)\in dy)\\
&\left.+\left(\sup_{|t|\geq\delta(z)}|\nu_z(t)|+\frac{1}{2m}\right)^m m^{(2k+1)(k+1)+(2k-1)/2}\left(1+\sqrt{m}\frac{|\mu(z)|}{\sigma(z)}\right)^{-1}\right\}.
\end{aligned}$$

Here $c(k)$ is a positive constant depending only on k. The functions $Q_\nu(x;z)$ are defined as

$$Q_\nu(x;z)=-\phi(x)\sum H_{\nu+2s-1}(x)\prod_{l=1}^{\nu}\frac{1}{k_l!}\left(\frac{\gamma_{l+2}(z)}{(l+2)!\sigma(z)^{l+2}}\right)^{k_l}.$$

where $\gamma_l(z)$ is the cumulant of order l of the random variable $X(z)$. The summation above is extended over all non-negative integer solution $(k_1,k_2,\ldots,k_\nu)$ of the equation $k_1+2k_2+\cdots+\nu k_\nu=\nu$, and $s=k_1+k_2+\cdots+k_\nu$.

By Assumption 4, we have seen in Proposition 3.1.2 that

$$\mathbb{E}\Phi\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right) - \alpha = \sum_{i=1}^{k} \frac{f_Y^{(2i-1)}(0)}{2^i i!} \cdot \frac{1}{m^i} + o(1/m^k).$$

Set $R_m(Z)$ to be the right hand side of inequality (3.2). In the following steps, we are going to consider $\mathbb{E}\left[\sum_{\nu=1}^{2k-1} Q_\nu\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}; Z\right)\right]$ and $\mathbb{E}R_m(Z)$.

By the definition of $Q_\nu(\cdot)$, we have that

$$\begin{aligned}
&\mathbb{E}\left[-\sum_{\nu=1}^{2k-1} Q_\nu\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}; Z\right) m^{-\nu/2}\right] \\
= &\sum_{\nu=1}^{2k-1} m^{-\nu/2} \sum_{s\in\Sigma(\nu)} \mathbb{E}\left[(\phi \cdot H_{\nu+2s-1})\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right) \sum \prod_{p=1}^{\nu} \frac{1}{k_p!}\left(\frac{\gamma_{p+2}(Z)}{(p+2)!\sigma^{p+2}(Z)}\right)^{k_p}\right] \\
= &\sum_{\nu=1}^{2k-1} m^{-\nu/2} \sum_{s\in\Sigma(\nu)} \mathbb{E}\left[(\phi \cdot H_{\nu+2s-1})\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}\right) \chi_{\nu,s}(Z)\right] \\
= &\sum_{\nu=1}^{2k-1} m^{-\nu/2} \sum_{s\in\Sigma(\nu)} \mathbb{E}\left[(\phi \cdot H_{\nu+2s-1})(-\sqrt{m}Y)\tilde{\chi}_{\nu,s}(Y)\right] \\
= &\; m^{-1/2}\mathbb{E}\left[(\phi \cdot H_2)(-\sqrt{m}Y)\tilde{\chi}_{1,1}(Y)\right] \\
&+\sum_{\nu=1}^{k-1} m^{-\nu} \sum_{s\in\Sigma(2\nu)} \mathbb{E}\left[(\phi \cdot H_{2\nu+2s-1})(-\sqrt{m}Y)\tilde{\chi}_{2\nu,s}(Y)\right] \\
&+\sum_{\nu=1}^{k-1} m^{-\nu-\frac{1}{2}} \sum_{s\in\Sigma(2\nu+1)} \mathbb{E}\left[(\phi \cdot H_{(2\nu+1)+2s-1})(-\sqrt{m}Y)\tilde{\chi}_{(2\nu+1),s}(Y)\right] \\
= &\; m^{-1/2}\left[\sum_{i=2}^{k} \frac{d_{2,i}}{m^{i-\frac{1}{2}}} + o(1/m^{k-\frac{1}{2}})\right] \quad \text{(by Assumption 5 and Lemma 3.1.3)} \\
&+\sum_{\nu=1}^{k-1} m^{-\nu}\left[\sum_{s\in\Sigma(2\nu)} \sum_{i=2}^{k-\nu} \frac{c_{2\nu+2s-1,i}}{m^i} + o(1/m^{k-\nu})\right] \quad \text{(by Assumption 6)} \\
&+\sum_{\nu=1}^{k-1} m^{-\nu-\frac{1}{2}}\left[\sum_{s\in\Sigma(2\nu+1)} \sum_{i=2}^{k-\nu} \frac{d_{(2\nu+1)+2s-1,i}}{m^{i-\frac{1}{2}}} + o(1/m^{k-\nu-\frac{1}{2}})\right] \quad \text{(by Assumption 7)}
\end{aligned}$$

Hence,

$$
\begin{aligned}
&\mathbb{E}\left[-\sum_{\nu=1}^{2k-1} Q_\nu\left(-\sqrt{m}\frac{\mu(Z)}{\sigma(Z)}; Z\right) m^{-\nu/2}\right] \\
&= \sum_{j=2}^{k} m^{-j}\left[d_{2,j} + \sum_{\nu=1}^{k-1}\left(\sum_{s\in\Sigma(2\nu)} c_{2\nu+2s-1,j-\nu} + \sum_{s\in\Sigma(2\nu+1)} d_{(2\nu+1)+2s-1,j-\nu}\right)\right] + o(1/m^k).
\end{aligned}
$$

where $c_{\cdot,1} \equiv 0$ and $d_{\cdot,1} \equiv 0$.

We will be done if we can show that $\mathbb{E}R_m(Z) = o(1/m^k)$. The proof proceeds as follows: By Assumption 2,

$$
|\nu_z(t)| \leq (1-\epsilon) + \epsilon|\nu_g(t)|,
$$

where $\nu_g(t) \triangleq \mathbb{E}e^{itX}$ and the expectation here is taken w.r.t. the density $g(\cdot)$. By Riemann-Lesbegue lemma,

$$
\limsup_{|t|\nearrow+\infty} |\nu_g(t)| < 1.
$$

By [25], this implies that

$$
\sup_{|t|\geq\delta} |\nu_Z(t)| < (1-\epsilon) + \epsilon \sup_{|t|\geq\delta} |\nu_g(t)| < 1 \quad \text{a.s.}
$$

for every $\delta > 0$. So, for a fixed δ, there exists a m_0 such that for all $m \geq m_0$

$$
\left((1-\epsilon) + \epsilon \sup_{|t|\geq\delta} |\nu_g(t)| + \frac{1}{2m}\right)^m m^{2k^2+4k+1} \leq 1.
$$

Define the upper bound M by

$$
M \triangleq 1 \vee \max_{1\leq m\leq m_0}\left\{\left((1-\epsilon) + \epsilon \sup_{|t|\geq\delta} |\nu_g(t)| + \frac{1}{2m}\right)^m m^{2k^2+4k+1}\right\}.
$$

It is now apparent that, almost surely,

$$\begin{aligned}
&\varepsilon\left(m^{1/2}\left(1+\sqrt{m}\frac{|\mu(Z)|}{\sigma(Z)}\right):Z\right)\\
\leq\ & c(2k+1)\left\{\frac{\beta_{2k+2}(Z)+M}{\sqrt{m}\left(1+\sqrt{m}\frac{|\mu(Z)|}{\sigma(Z)}\right)}\right.\\
&\left.+\mathbb{E}\left[\left|\frac{X-\mu(Z)}{\sigma(Z)}\right|^{2k+1}:\left|\frac{X-\mu(Z)}{\sigma(Z)}\right|>\sqrt{m}\left(1+\sqrt{m}\frac{|\mu(Z)|}{\sigma(Z)}\right)\,\middle|\,Z\right]\right\}\\
\leq\ & c(2k+1)\frac{(M+1)+2\beta_{2k+2}(Z)}{m^{\frac{1}{4(k+1)}}} \quad \text{by Lemma 3.1.4.}
\end{aligned}$$

In other words,

$$R_m(Z)\leq\frac{c(2k+1)(2\beta_{2k+2}(Z)+M+1)/m^{\frac{1}{4k+4}}}{m^{\frac{2k-1}{2}}\left(1+\sqrt{m}\frac{|\mu(Z)|}{\sigma(Z)}\right)^{2k+1}} \quad \text{a.s.}$$

Now, for all $m>0$,

$$\begin{aligned}
&\mathbb{E}\left[m^{k+\frac{1}{4k+4}}\frac{c(2k+1)(2\beta_{2k+2}(Z)+M+1)}{m^{\frac{2k-1}{2}}\left(1+\sqrt{m}\frac{|\mu(Z)|}{\sigma(Z)}\right)^{2k+1}}\right]\\
=\ &\mathbb{E}\left[\sqrt{m}\frac{c(2k+1)(2\tilde{\beta}_{2k+2}(Y)+M+1)}{(1+\sqrt{m}|Y|)^{2k+1}}\right]\\
=\ &\int_{-\infty}^{\infty}\sqrt{m}\frac{c(2k+1)(2\tilde{\beta}_{2k+2}(y)+M+1)f_Y(y)}{(1+\sqrt{m}|y|)^{2k+1}}\,dy.
\end{aligned}$$

By Assumptions 3 and 4, there exists $\tilde{M}>0$ such that $(2\tilde{\beta}_{2k+2}(\cdot)+M+1)f_Y(\cdot)<\tilde{M}$. Thus,

$$\begin{aligned}
&\mathbb{E}\left[m^{k+\frac{1}{4k+4}}\frac{c(2k+1)(2\beta_{2k+2}(Z)+M+1)}{m^{\frac{2k-1}{2}}\left(1+\sqrt{m}\frac{|\mu(Z)|}{\sigma(Z)}\right)^{2k+1}}\right]\\
\leq\ & 2\tilde{M}c(2k+1)\int_0^{\infty}\frac{\sqrt{m}\,dy}{(1+\sqrt{m}y)^{2k+1}}<\frac{\tilde{M}\cdot c(2k+1)}{k}=O(1).
\end{aligned}$$

Hence, $\{m^{k+\frac{1}{4(k+1)}} R_m(Z) : m \geq 1\}$ is bounded in $\mathcal{L}^1$, uniformly in $m > 0$. In other words, $\mathbb{E}R_m(Z)$ is of order $O(m^{-k-\frac{1}{4(k+1)}})$, and $o(m^{-k})$, in particular. $\square$

3.1.2 Mean Squared Error

As a guideline in obtaining the appropriate CLTs, we will now try to minimize the MSE of the estimator, $\hat{\alpha}$ and to derive the optimal rate of convergence.

For sufficiently large m, the asymptotic variance of the estimator is given by

$$\begin{aligned} \mathrm{Var}(\hat{\alpha}) &= \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{P}(\bar{X}_m(Z_i) \leq 0) \cdot \mathbb{P}(\bar{X}_m(Z_i) > 0) \\ &\sim \frac{\alpha(1-\alpha)}{n}. \end{aligned}$$

since $\mathbb{P}(\bar{X}_m(Z) \leq 0) \sim \alpha$, as seen in the previous subsection.

Hence, assuming that the conditions of Proposition 3.1.1 or 3.1.3 hold and that $f_Y^{(1)}(0) \neq 0$, we can deduce that the asymptotic MSE of $\hat{\alpha}$ is given by:

$$\mathrm{MSE}(\hat{\alpha}) = \left(\frac{f_Y^{(1)}(0)}{2m}\right)^2 + \frac{\alpha(1-\alpha)m}{\beta c} \tag{3.3}$$

since $m \cdot n = \beta c$. Differentiating $\mathrm{MSE}(\hat{\alpha})$ w.r.t. m and setting the derivative to zero, we obtain that

$$0 = \frac{d\mathrm{MSE}(\hat{\alpha})}{dm} = -\frac{(f_Y^{(1)}(0))^2}{2m^3} + \frac{\alpha(1-\alpha)}{\beta c}.$$

Hence, the optimal m is given by:

$$m^* = \left(\frac{(f_Y^{(1)}(0))^2\beta}{2\alpha(1-\alpha)}\right)^{1/3} c^{1/3},$$

and the optimal n is given by:

$$n^* = \left(\frac{2\alpha(1-\alpha)\beta^2}{(f_Y^{(1)}(0))^2}\right)^{1/3} c^{2/3}.$$

Substituting m^* and n^* into equation (3.3), we obtain that the minimized MSE, as a function of c, is equal to

$$\mathrm{MSE}_{\min}(\hat{\alpha}) = (3/2)(\alpha(1-\alpha)f_Y^{(1)}(0)/2)^{2/3}c^{-2/3} = O\left(c^{-1/3}\right).$$

3.1.3 Central Limit Theorems

Recall again from Section 3.1 that our estimator for $\alpha = \mathbb{P}(\mathbb{E}(X|Z) \leq 0)$ is given by

$$\hat{\alpha}(m,n) = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \leq 0\right).$$

The determination of the best possible m and n introduces a trade-off between the variance of the estimator and its bias. If n is chosen too small (relative to m), the variance contribution to the mean square error will dominate, whereas if n is chosen too large (relative to m), the bias will govern the convergence rate. Given a fixed computation budget c, it turns out that the optimal sample sizes $m = m_c$ and $n = n_c$ are typically of order $c^{1/3}$ and $c^{2/3}$ respectively for the above estimator.

The following proposition states that if the sample sizes are chosen optimally, then the convergence rate of $\hat{\alpha}(m,n)$ to α is of order $c^{-1/3}$.

Proposition 3.1.5 *Assume that*

1. *the conditions of Proposition 3.1.1 or 3.1.4 hold for some* $k \geq 1$*;*
2. $\mathbb{P}(\mu(Z) = 0) = 0$*;*
3. $f_Y^{(1)}(0) \neq 0$.

Then the following results hold.

If $c^{-1/3}m(c) \to a > 0$ *and* $c^{-2/3}n(c) \to b > 0$, *where* $0 < a, b < \infty$, *then as* $c \nearrow +\infty$,

$$c^{1/3}(\hat{\alpha}(m(c),n(c)) - \alpha) \Rightarrow \sqrt{\frac{\alpha(1-\alpha)}{b}}\mathrm{N}(0,1) + \frac{f_Y^{(1)}(0)}{2a}. \tag{3.4}$$

If for some $\gamma \neq 1/3$, $c^{-\gamma}m(c) \to \bar{a}$ *and* $c^{-(1-\gamma)}n(c) \to \bar{b}$, *where* $0 < \bar{a}, \bar{b} < \infty$, *as* $c \nearrow +\infty$, *then*

$$c^{\gamma}(\hat{\alpha}(m(c), n(c)) - \alpha) \Rightarrow \frac{f_Y^{(1)}(0)}{2\bar{a}} \qquad \text{if } \gamma < \frac{1}{3} \tag{3.5}$$

and

$$c^{\frac{1-\gamma}{2}}(\hat{\alpha}(m(c), n(c)) - \alpha) \Rightarrow \sqrt{\frac{\alpha(1-\alpha)}{\bar{b}}}\mathrm{N}(0,1) \qquad \text{if } \gamma > \frac{1}{3}. \tag{3.6}$$

Proof.

Let $\chi_i(m) \triangleq I(\bar{X}_m(Z_i) \leq 0)$. Note that

$$\hat{\alpha}(m(c), n(c)) - \alpha = \frac{1}{n}\sum_{i=1}^{n} \hat{\chi}_i(m(c)) + \mathbb{P}(\bar{X}_m(Z) \leq 0) - \alpha,$$

where $\hat{\chi}_i(m(c)) = \chi_i(m(c)) - \mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0)$ is the centered version of $\chi_i(m(c))$. Then, for $0 < \gamma < 1$,

$$c^{\gamma}(\hat{\alpha}(m(c), n(c)) - \alpha) = c^{\gamma}n^{-1/2}\left(\sum_{i=1}^{n} \frac{\hat{\chi}_i(m(c))}{\sqrt{n(c)}}\right) + c^{\gamma}(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0) - \alpha).$$

Observe that for each i,

$$\begin{aligned} \mathbb{E}|\hat{\chi}_i(m)|^2 &= \mathrm{Var}[I(\bar{X}_m(Z) \leq 0)] \\ &= \mathbb{P}(\bar{X}_m(Z) \leq 0) \cdot \mathbb{P}(\bar{X}_m(Z) > 0) < \frac{1}{4}, \end{aligned}$$

from which it follows that $\{\hat{\chi}_i(m) : m \geq 1\}$ is uniformly integrable. By Lemma 2.2.1, the Lindeberg-Feller theorem applies here. That is, as $n \to +\infty$,

$$\sum_{i=1}^{n} \frac{\hat{\chi}_i(m)}{\sqrt{n}} \Rightarrow \sigma\mathrm{N}(0,1),$$

where $\sigma = \alpha(1-\alpha)$.

When $\gamma = \frac{1}{3}$, $c^{-1/3}m(c) \to a$ and $c^{-2/3}n(c) \to b$, as $c \nearrow +\infty$,

$$c^{1/3}n(c)^{-1/2}\left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c))}{\sqrt{n(c)}}\right) \Rightarrow \sqrt{\frac{\alpha(1-\alpha)}{b}}\mathrm{N}(0,1).$$

Note also that as $c \nearrow +\infty$,

$$c^{1/3}(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0) - \alpha) = \frac{f_Y^{(1)}(0)}{2a} + o(c^{1/3})$$

as a result of Proposition 3.1.1. By the converging together theorem [5, p.340], we obtain the first result by combining the convergence results above.

Now, assume that $\gamma \neq \frac{1}{3}$ and that $c^{\gamma}m(c) \to \bar{a}$ and $c^{-(1-\gamma)}n(c) \to \bar{b}$. If $\gamma < \frac{1}{3}$,

$$c^{\gamma}(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0) - \alpha) \Rightarrow \frac{f_Y^{(1)}(0)}{2\bar{a}} \quad \text{from Assumption 1,} \tag{3.7}$$

and

$$c^{\gamma}n(c)^{-1/2}\left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c))}{\sqrt{n(c)}}\right) \Rightarrow 0 \quad \text{since } \gamma - (1-\gamma)/2 < 0. \tag{3.8}$$

On the other hand, if $\gamma > \frac{1}{3}$, then,

$$c^{\frac{1-\gamma}{2}}(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq 0) - \alpha) \Rightarrow 0 \quad \text{since } \tfrac{1-\gamma}{2} < \gamma \tag{3.9}$$

and

$$c^{\frac{1-\gamma}{2}}n(c)^{-1/2}\left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c))}{\sqrt{n(c)}}\right) \Rightarrow \sqrt{\frac{\alpha(1-\alpha)}{b}}\mathrm{N}(0,1). \tag{3.10}$$

We then obtain (3.5) by combining the converging results (3.7) and (3.8) and obtain (3.6) by combining (3.9) and (3.10), using the converging together theorem [5, p.340]. □

3.2 Quantile Estimation

In this section, we will provide an estimator for the quantile of a conditional expectation. We will derive a CLT for the estimator. Under certain technical conditions, the CLT for the quantile estimator can actually be derived directly from the CLT for the distribution function evaluated at the corresponding quantile. Towards the end of the section, we will provide sufficient conditions for which this direct derivation is possible.

Let $\alpha(x) \triangleq \mathbb{P}(\mathbb{E}(X|Z) \leq x)$ be the distribution function of the conditional expectation evaluated at x and $q(p) \triangleq \sup\{x : \alpha(x) < p\}$ be it's p-th quantile. Furthermore, define $\hat{\alpha}(x, m, n)$ and $\hat{q}(p, m, n)$ to be the sample distribution function and sample quantile respectively; namely,

$$\hat{\alpha}(x, m, n) \triangleq \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \leq x\right),$$

and

$$\hat{q}(p, m, n) \triangleq \sup\{x : \hat{\alpha}(x, m, n) < p\} = \bar{X}_m(Z)_{(\lfloor np \rfloor)}.$$

where $\bar{X}_m(Z)_{(i)}$ is the i-th order statistics of the sample means $\{\bar{X}_m(Z_1), \ldots, \bar{X}_m(Z_n)\}$. The aim is this section is to derive a CLT for the quantile estimator $\hat{q}(p, m, n)$.

3.2.1 A Useful Lemma

We will in this subsection develop and prove a useful lemma that will allow us to immediately write down the CLT for the quantile estimator, when we are given one for the distribution function, as will be seen in the next subsection.

In the setting of our estimator of the quantile q of $\mathbb{E}(X|Z)$,

- $F(\cdot)$ in the next lemma is the distribution function of the conditional expectation: i.e. $F(x) = \mathbb{P}(\mathbb{E}(X|Z) \leq x)$;
- $F_c(x)$ is our distribution function estimator of $F(x)$, given a computation budget of

c, i.e.

$$F_c(x) = \frac{1}{n(c)} \sum_{i=1}^{n(c)} I\left(\frac{1}{m(c)} \sum_{j=1}^{m(c)} X_j(Z_i) \le x\right);$$

- the constant β in our case is equal to $1/3$;
- for each $x \in (q - \epsilon, q + \epsilon)$,

$$Z_\infty(x) = \frac{f^{(1)}_{Y(x)}(0)}{2a} + \sqrt{F(x)(1 - F(x))a}\mathrm{N}(0, 1),$$

where $Y(x) = (\mu(Z) - x)/\sigma(Z)$ (see later) and $a = m(c)c^{-1/3}$ and $n(c) = c/m(c)$.

Lemma 3.2.1 *Let $F(\cdot)$ be probability distribution function and for a fixed $x \in \mathbb{R}$, let $F_c(x)$ be an estimate of $F(x)$ using a computational budget c. Suppose that, for a given $\beta > 0$,*

1. *The family $\left\{c^\beta\left(F_c(x) - F(x)\right) : c > 0, |x - q| < \epsilon\right\}$ is uniformly integrable for some ϵ;*
2. *For all x such that $|x-q| < \epsilon$, there exists $Z_\infty(x)$ such that $c^\beta(F_c(x) - F(x)) \Rightarrow Z_\infty(x)$; moreover, $\sup_{\substack{|x-q|<\epsilon \\ y \in \mathbb{R}}} \left|\mathbb{P}\left(c^\beta(F_c(x) - F(x)) \le 0\right) - \mathbb{P}\left(Z_\infty(x) \le y\right)\right| \to 0.$*

Then,

$$\mathbb{E}F_c(x) = F(x) + \frac{b(x)}{c^\beta} + o(c^{-\beta}) \qquad \textit{uniformly in } |x - q| < \epsilon,$$

where $b(x) = \mathbb{E}Z_\infty(x)$.

Proof.

By Assumption 1, given any $\eta > 0$, there exists $r(\eta) \in (0, \infty)$ such that

$$\mathbb{E}\left[c^\beta |F_c(x) - F(x)| ; c^\beta |F_c(x) - F(x)| > r(\eta)\right] < \eta$$

for all $c > 0$ and $x \in (q - \epsilon, q + \epsilon)$. From Fatou's lemma, it follows that

$$|\mathbb{E}[Z_\infty(x); |Z_\infty(x)| > r(\eta)]| < \mathbb{E}[|Z_\infty(x)|; |Z_\infty(x)| > r(\eta)] < \eta.$$

Decomposing $\mathbb{E}[c^{\beta}(F_c(x) - F(x))]$ into two terms, we have that

$$\begin{aligned}\mathbb{E}[c^{\beta}(F_c(x) - F(x))] &= \mathbb{E}[c^{\beta}(F_c(x) - F(x)); c^{\beta}|F_c(x) - F(x)| < r(\eta)] \\ &\quad + \mathbb{E}[c^{\beta}(F_c(x) - F(x)); c^{\beta}|F_c(x) - F(x)| > r(\eta)].\end{aligned}$$

We have already shown that the second term on the RHS above is less than η. Let's now consider the first term.

Define $G_{c,x}(y) \triangleq \mathbb{P}(c^{\beta}(F_c(x) - F(x)) \leq y)$ and $G_x(y) \triangleq \mathbb{P}(Z_\infty(x) \leq y)$. By Assumption 2, there exists $C(\eta)$ such that for all $c \geq C(\eta)$,

$$|G_{c,x}(y) - G_x(y)| < \eta \cdot \frac{1}{4r(\eta)} \qquad \text{uniformly in } |x - q| < \epsilon, y \in \mathbb{R}.$$

Notice that for any r.v. X, by a change of variable argument,

$$\begin{aligned}\mathbb{E}[X; |X| < r] &= \int_{-r}^{r} x\, dF(x) \\ &= \int_{-r}^{0} x\, dF(x) + \int_{0}^{r} x\, dF(x) \\ &= \int_{-r}^{0}\int_{0}^{x} dy\, dF(x) + \int_{0}^{r}\int_{0}^{x} dy\, dF(x) \\ &= \int_{-r}^{0} (F(-r) - F(y))\, dy + \int_{0}^{r} (F(r) - F(y))\, dy.\end{aligned}$$

Hence,

$$\mathbb{E}[c^{\beta}(F_c(x) - F(x)); c^{\beta}|F_c(x) - F(x)| < r(\eta)] =$$
$$\int_{-r(\eta)}^{0} G_{c,x}(-r) - G_{c,x}(y)\, dy + \int_{0}^{r(\eta)} G_{c,x}(r) - G_{c,x}(y)\, dy.$$

By construction, we then have that

$$
\begin{aligned}
& \left| \mathbb{E}[c^{\beta}(F_c(x) - F(x)); c^{\beta}|F_c(x) - F(x)| < r(\eta)] - \mathbb{E}[Z_\infty(x); |Z_\infty(x)| < r(\eta)] \right| \\
= & \left| \mathbb{E}[c^{\beta}(F_c(x) - F(x)); c^{\beta}|F_c(x) - F(x)| < r(\eta)] \right. \\
& \left. - \int_{-r(\eta)}^{0} G_x(-r(\eta)) - G_x(y)\, dy - \int_0^{r(\eta)} G_x(r(\eta)) - G_x(y)\, dy \right| \\
< & \ \eta
\end{aligned}
$$

uniformly in $c > C(\eta)$ and $x \in (q - \epsilon, q + \epsilon)$.

Therefore,

$$
\begin{aligned}
& \left| \mathbb{E}[c^{\beta}(F_c(x) - F(x))] - \mathbb{E}Z_\infty(x) \right| \\
= & \left| \mathbb{E}[c^{\beta}(F_c)(x) - F(x); c^{\beta}|F_c(x) - F(x)| > r(\eta)] \right. \\
& \left. + \mathbb{E}[c^{\beta}(F_c)(x) - F(x); c^{\beta}|F_c(x) - F(x)| < r(\eta)] - \mathbb{E}Z_\infty(x) \right| \\
\leq & \ 2\eta + |\, \mathbb{E}[Z_\infty(x); |Z_\infty(x)| < r(\eta)] - \mathbb{E}Z_\infty(x)| \\
= & \ 3\eta
\end{aligned}
$$

uniformly in $|x - q| < \epsilon$. Since $\eta > 0$ is arbitrary, we have shown that as $c \nearrow +\infty$

$$
\mathbb{E}[c^{\beta}(F_c(x) - F(x))] \to \mathbb{E}Z_\infty(x) \quad \text{uniformly in } |x - q| < \epsilon. \quad \square
$$

3.2.2 Central Limit Theorems for the Quantile Estimator

We will invoke Lemma 3.2.1 developed in the last subsection to derive a CLT for the quantile estimator defined earlier, namely,

$$
\hat{q}_c(p) = \bar{X}_{m(c)}(Z)_{(\lfloor n(c)p \rfloor)},
$$

where c is the given computational budget and $\bar{X}_{m(c)}(Z)_{(i)}$ is the i-th order statistic of the sample $\{\bar{X}_{m(c)}(Z_j) : j = 1, \ldots, n(c)\}$. Specifically, we assume that Z is a general conditioning "random element". Just for notational purpose, define

$$Y(x) \triangleq \frac{\mu(Z) - x}{\sigma(Z)}.$$

An interesting feature of the CLT for the quantile estimator is that its RHS is exactly equal to that for the distribution function, divided by $-\alpha'(q(p))$.

We will now outline the ideas of how the CLT for $\hat{q}_c(p)$ is established. First, for an arbitrarily small $\epsilon > 0$, we define for each $x \in (q - \epsilon, q + \epsilon)$, the estimator, $\hat{\alpha}_c(x)$, of $\alpha(x) \triangleq \mathbb{P}(\mathbb{E}(X|Z) \leq x)$ as follows:

$$\hat{\alpha}_c(x) \triangleq \frac{1}{n(c)} \sum_{i=1}^{n(c)} I\left(\frac{1}{m(c)} \sum_{j=1}^{m(c)} X_j(Z_i) \leq x\right).$$

Condition A below gurantees that similar CLT as the one described in Proposition 3.1.5 holds for the estimator $\hat{\alpha}(x)$ for each $|x - q| < \epsilon$.

Condition A. For $|x - q| < \epsilon$, $\mathbb{P}(\mu(Z) = x) = 0$ and either

1. $\mathbb{P}(X \leq \cdot|Z) = \Phi\left((\cdot - \mu(Z))/\sigma(Z)\right)$;

2. $f_{Y(x)}(\cdot) \in C_b^2$

or

3. $\mathbb{E}|X|^4 < \infty$;

4. there exists a density function $g(\cdot)$ such that $\mathbb{P}(X_1 \in \cdot|Z) \geq \epsilon g(\cdot)$ a.s.;

5. $\mathbb{E}[\beta_{2k+2}(Z)|Y(x) = \cdot] \in C_b^1$ for $|x - q| < \epsilon$;

6. the density of $Y(x)$, $f_{Y(x)}(\cdot) \in C_b^2$ for $|x - q| < \epsilon$;

7. for $|x - q| < \epsilon$, $(f_{Y(x)} \cdot \tilde{\chi}_x)(\cdot) \in C_b^1$ where $\tilde{\chi}_x(y) \triangleq \mathbb{E}[\chi_{1,1}(Z)|Y(x) = y]$

hold.

For the ease of exposition, we will henceforth consider only the case where $m(c) \sim ac^{1/3}$ and $n(c) \sim a^{-1}c^{2/3}$ for some $a > 0$, so that the rate of the convergence of the estimator is optimal.

As the second step in establishing the CLT for the quantile estimator, we will prove in the next lemma that Condition A together with Condition B below ensure that the family of r.v.'s $\{c^{1/3}(\hat{\alpha}_c(x) - \alpha(x)) : c > 0, |x - q| < \epsilon\}$ satisfies the conditions of Lemma 3.2.1.

Let's denote by $h(\cdot, \cdot)$ the joint density function of $(\mu(Z), \sigma(Z))$, which is assumed to exist.

Condition B. For a fixed $\epsilon > 0$,

1. $\sup_{|x-q|<2\epsilon} \sup_{\xi \in \mathbb{R}} \int_0^\infty z_2^2 h^2(x + \xi z_2, z_2)\, dz_2 < \infty$;

2. $\sup_{|x-q|<\epsilon} \sup_{\xi \in \mathbb{R}} \int_0^\infty z_2^4 \left(\frac{\partial}{\partial z_1} h(x + \xi z_2, z_2) \right)^2 dz_2 < \infty$.

Lemma 3.2.2 *Suppose that both Conditions A and B hold. Then, for some $\epsilon > 0$, the family of r.v.'s $\{c^{1/3}(\hat{\alpha}_c(x) - \alpha(x)) : c > 0, |x - q| < \epsilon\}$ satisfies*

1. *The family $\left\{c^{1/3}\left(\hat{\alpha}_c(x) - \alpha(x)\right) : c > 0, |x - q| < \epsilon\right\}$ is uniformly integrable;*

2. *Set $Z_\infty(x) = f_{Y(x)}^{(1)}(0)/(2a) + \sqrt{\alpha(x)(1 - \alpha(x))a}\mathrm{N}(0, 1)$. Then,*

$$\sup_{\substack{|x-q|<\epsilon \\ y \in \mathbb{R}}} \left| \mathbb{P}\left(c^{1/3}(\hat{\alpha}_c(x) - \alpha(x)) \leq 0\right) - \mathbb{P}\left(Z_\infty(x) \leq y\right) \right| \to 0.$$

Proof.

To establish the first result, it suffices to show that $\mathbb{E}\left[c^{2/3}(\hat{\alpha}_c(x) - \alpha(x))^2\right]$ is uniformly bounded for $c > 0$ and $|x - q| < \epsilon$ (c.f. Section 13.3 of [28]). Analogous to the proof of Proposition 3.1.5, we can rewrite $c^{1/3}(\hat{\alpha}_c(x) - \alpha(x))$ as

$$c^{1/3}(\hat{\alpha}_c(x) - \alpha(x)) = c^{1/3}n(c)^{-1/2}\left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c), x)}{\sqrt{n(c)}}\right) + c^{1/3}\left(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq x) - \alpha(x)\right).$$

where $\hat{\chi}_i(m(c), x) \triangleq I(\bar{x}_{m(c)}(Z_i) \leq x) - \mathbb{P}(\bar{X}_{m(c)}(Z_i) \leq x)$. Since for any two numbers $a, b \in \mathbb{R}$, $(a+b)^2 \leq 2(a^2+b^2)$, we have that

$$c^{2/3}(\hat{\alpha}_c(x) - \alpha(x))^2 \leq 2\left[\frac{c^{2/3}}{n(c)}\left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c), x)}{\sqrt{n(c)}}\right)^2 + c^{2/3}\left(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq x) - \alpha(x)\right)^2\right].$$

Since $m(c) = ac^{1/3}$ and $n(c) = a^{-1}c^{2/3}$, we have that $c^{2/3}/n(c) = a$. Moreover, since the $\hat{\chi}_\cdot(m(c), x)$'s are i.i.d., we have that

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c), x)}{\sqrt{n(c)}}\right)^2 &= \mathrm{Var}\left(\hat{\chi}_1(m(c), x)\right) \\
&= \mathbb{P}(\bar{X}_{m(c)}(Z) \leq x) \cdot \mathbb{P}(\bar{X}_{m(c)}(Z) > x) \leq \frac{1}{4}.
\end{aligned}$$

On the other hand, recall that $m(c) = ac^{1/3}$. Analogous to the proof of Proposition 3.1.5, we deduce from Condition A that for each $x \in (q-\epsilon, q+\epsilon)$,

$$c^{1/3}(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq x) - \alpha(x)) = \frac{f^{(1)}_{Y(x)}(0)}{a} + o(c^{-1/3}) = \frac{f^{(1)}_{Y(x)}(\xi_{c,x})}{a},$$

where $\xi_{c,x} \to 0$ as $c \nearrow +\infty$. Hence, we will have the first result established if we can show that $f^{(1)}_{Y(x)}(\xi_{c,x})$ is uniformly bounded in $c > 0$ and $|x - q| < \epsilon$.

Now consider the distribution function of $Y(x)$.

$$\begin{aligned}
\mathbb{P}(Y(x) \leq y) &= \mathbb{P}(\mu(Z) \leq \sigma(Z)y + x) \\
&= \int_0^\infty \int_{-\infty}^{x+yz_2} h(z_1, z_2)\, dz_1\, dz_2.
\end{aligned}$$

By Condition B1, we know that $\{\int_0^\infty z_2 h(x + \xi z_2, z_2)\, dz_2 : |x - q| < 2\epsilon, \xi \in \mathbb{R}\}$ is uniformly integrable (w.r.t. the Lesbegue measure.) This allows us to express the density of $Y(x)$ as

$$f_{Y(x)}(y) = \frac{d}{dy}\mathbb{P}(Y(x) \leq y) = \int_0^\infty z_2 h(x + yz_2, z_2)\, dz_2.$$

Similarly, Condition B2 allows us to write

$$f^{(1)}_{Y(x)}(y) = \int_0^\infty z_2^2 \frac{\partial}{\partial z_1} h(x + yz_2, z_2)\, dz_2$$

which is uniformly bounded in $y \in \mathbb{R}$ and $|x - q| < \epsilon$. In particular. we have now established the first result.

Let's now proceed to prove the second result. Recall once again that

$$c^{1/3}(\hat{\alpha}_c(x) - \alpha(x)) = \sqrt{a}\left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c), x)}{\sqrt{n(c)}}\right) + c^{1/3}(\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha(x))$$

and

$$Z_\infty(x) = \frac{f^{(1)}_{Y(x)}(0)}{2a} + \sqrt{\alpha(x)(1 - \alpha(x))a}\mathrm{N}(0, 1).$$

As we have seen earlier, as a result of Condition B, $f^{(1)}_{Y(x)}(y)$ is uniformly bounded in $y \in \mathbb{R}$ and $|x - q| < \epsilon$. In particular,

$$|c^{1/3}(\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha(x)) - f^{(1)}_{Y(x)}(0)/(2a)| = |f^{(1)}_{Y(x)}(\xi_{c,x}) - f^{(1)}_{Y(x)}(0)|/(2a)$$

is also bounded. Furthermore, Condition A implies that for each $|x - q| < \epsilon$,

$$c^{1/3}(\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha(x)) - f^{(1)}_{Y(x)}(0)/(2a) = \frac{\eta(c, x)}{c^{1/3}}.$$

where $\eta(c, x) \to 0$ as $c \nearrow +\infty$. These two results imply that

$$\sup_{|x-q|<\epsilon} \left| c^{1/3}(\mathbb{P}(\bar{X}_{m(c)}(Z) \le x) - \alpha(x)) - \frac{f^{(1)}_{Y(x)}(0)}{2a} \right| \to 0 \tag{3.11}$$

as $c \nearrow +\infty$.

Now, utilizing the Berry-Esseen inequality (c.f. Theorem 3.1.1) on the Bernoulli random variables, $\hat{\chi}_i(m(c), x)$'s, we can show that

$$\sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} \left| \mathbb{P}\left(\sqrt{a}\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c), x)}{\sqrt{n(c)}} \leq y\right) - \mathbb{P}\left(\sqrt{a\alpha(x)(1-\alpha(x))}\mathrm{N}(0,1) \leq y\right)\right| \to 0 \tag{3.12}$$

as $c \nearrow +\infty$.

For ease of exposition, let's define

$$\begin{aligned} X_c(x) &\triangleq \sqrt{a}\left(\sum_{i=1}^{n(c)} \frac{\hat{\chi}_i(m(c), x)}{\sqrt{n(c)}}\right); \\ X(x) &\triangleq \sqrt{a\alpha(x)(1-\alpha(x))}\mathrm{N}(0,1); \\ b_c(x) &\triangleq c^{1/3}(\mathbb{P}(\bar{X}_{m(c)}(Z) \leq x) - \alpha(x)); \\ b(x) &\triangleq \frac{f^{(1)}_{Y(x)}(0)}{2a}. \end{aligned}$$

We have already seen from (3.11) and (3.12) that as $c \nearrow +\infty$,

$$\begin{aligned} \sup_{|x-q|<\epsilon} |b_c(x) - b(x)| &\to 0 \\ \sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} |\mathbb{P}(X_c(x) \leq y) - \mathbb{P}(X(x) \leq y)| &\to 0. \end{aligned}$$

To this end, observe that, by the triangle inequality,

$$\begin{aligned} &\sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} |\mathbb{P}(X_c(x) + b_c(x) \leq y) - \mathbb{P}(X(x) + b(x) \leq y)| \\ \leq\ &\sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} |\mathbb{P}(X_c(x) + b_c(x) \leq y) - \mathbb{P}(X(x) + b_c(x) \leq y)| \\ &+ \sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} |\mathbb{P}(X(x) + b_c(x) \leq y) - \mathbb{P}(X(x) + b(x) \leq y)|. \end{aligned}$$

The first term on the RHS of the above inequality converges to 0 by (3.12). As for the second term, we have that

$$\sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} |\mathbb{P}(X(x)+b_c(x)\le y)-\mathbb{P}(X(x)+b(x)\le y)| \le \sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} |f_{X(x)}(\xi(x))||b_c(x)-b(x)|,$$

where $f_{X(x)}(\cdot)$ is the density function of $\mathrm{N}(0,\alpha(x)(1-\alpha(x)))$ and is bounded above by $(\sqrt{2\pi}\zeta)^{-1}$ and $\zeta^2 \triangleq \min_{|x-q|<\epsilon}\alpha(x)(1-\alpha(x))$. Note from (3.11) that

$$\sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} |b_c(x)-b(x)| \to 0$$

as $c \nearrow +\infty$.

By combining the results, we have proved that

$$\sup_{\substack{|x-q|<\epsilon \\ y\in\mathbb{R}}} |\mathbb{P}(X_c(x)+b_c(x)\le y)-\mathbb{P}(X(x)+b(x)\le y)| \to 0.$$

and hence the second result. □

Equipped with the above lemma, we are now ready to present the theorem for the CLT of our quantile estimator.

Theorem 3.2.1 *Let* $0<p<1$. *Assume further that*

1. *Conditions A and B hold.*
2. $\alpha(\cdot)$ *is continuously differentiable at* $q(p)$ *and* $\alpha'(q(p))>0$.
3. $f^{(1)}_{Y(q(p))}(\cdot)$ *is continuous at the point 0, where* $f_{Y(q(p))}(\cdot)$ *is the density of* $(\mu(Z)-q(p))/\sigma(Z)$.

Then the following results hold.

If $c^{-1/3}m(c) \to a$ and $c^{-2/3}n(c) \to b$, where $0 < a, b < \infty$, then as $c \nearrow +\infty$,

$$c^{1/3}(\hat{q}(p) - q(p)) \Rightarrow \sqrt{\frac{p(1-p)}{b}} \cdot \frac{1}{\alpha'(q(p))} N(0,1) - \frac{1}{2a} f^{(1)}_{Y(q(p))}(0) \frac{1}{\alpha'(q(p))}. \tag{3.13}$$

If for some $\gamma > 1/3$, $c^{-\gamma}m(c) \to a$ and $c^{-(1-\gamma)}n(c) \to b$, where $0 < a, b < \infty$, then, as $c \to +\infty$,

$$c^{\frac{1-\gamma}{2}}(\hat{\alpha}(m(c), n(c)) - \alpha) \Rightarrow \sqrt{\frac{p(1-p)}{b}} \cdot \frac{1}{\alpha'(q(p))} N(0,1). \tag{3.14}$$

If for some $\gamma < 1/3$, $c^{-\gamma}m(c) \to a$ and $c^{-(1-\gamma)}n(c) \to b$, where $0 < a, b < \infty$, and $\sup_{|x-q|<\epsilon} \int_0^\infty z_2^4 \left(\frac{\partial^2}{\partial z_1^2} h(x, z_2)\right)^2 dz_2 < \infty$ then, as $c \to +\infty$,

$$c^{\gamma}(\hat{\alpha}(m(c), n(c)) - \alpha) \Rightarrow -\frac{f^{(1)}_{Y(q(p))}(0)}{2a\alpha'(q(p))}. \tag{3.15}$$

Proof.

For the first result, let

$$\begin{aligned} G(t, m, n) &\triangleq \mathbb{P}\left(\frac{n^{1/2}(\hat{q}(p, m, n) - q(p))}{A} \le t\right) \quad \text{(for some } A \text{ to be determined later)} \\ &= \mathbb{P}(\hat{q}(p, m, n) \le q(p) + Atn^{-1/2}) \\ &= \mathbb{P}(p \le \hat{\alpha}(q(p) + Atn^{-1/2}, m, n)) \\ &= \mathbb{P}\left(np \le \sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \le q(p) + Atn^{-1/2}\right)\right) \\ &= \mathbb{P}(np \le Z_n(F(q(p) + Atn^{-1/2}, m))), \end{aligned}$$

where $Z_n(\triangle)$ is a binomial$(n, \triangle)$ r.v. and $F(x, m) \triangleq \mathbb{E}\mathbb{P}(\bar{X}_m(Z) \le x)$.

In terms of the standardized form of $Z_n(\triangle)$,

$$Z_n^*(\triangle) \stackrel{\triangle}{=} \frac{Z_n(\triangle) - n\triangle}{[n\triangle(1-\triangle)]^{1/2}},$$

we have

$$G(t,m,n) = \mathbb{P}(Z_n^*(\triangle_{m,n,t}) \geq -c_{m,n,t}), \tag{3.16}$$

where

$$\begin{aligned}
\triangle_{m,n,t} &\stackrel{\triangle}{=} F(q(p) + Atn^{-1/2}, m) \\
c_{m,n,t} &\stackrel{\triangle}{=} \frac{\sqrt{n}(\triangle_{m,n,t} - p)}{\sqrt{\triangle_{m,n,t}(1-\triangle_{m,n,t})}}.
\end{aligned}$$

Now utilizing the Berry-Esséen inequality, we write

$$\sup_{-\infty<x<\infty} |\mathbb{P}(Z_n^*(\triangle) < x) - \Phi(x)| \leq C\frac{\rho_\triangle}{\sigma_\triangle^3 n^{1/2}} = C\frac{\gamma(\triangle)}{n^{1/2}}, \tag{3.17}$$

where C is a universal constant, $\sigma_\triangle^2 = \mathrm{Var}\{Z_1(\triangle)\} = \triangle(1-\triangle)$, $\rho_\triangle = \mathbb{E}|Z_1(\triangle) - \triangle|^3 = \triangle(1-\triangle)[(1-\triangle)^2 + \triangle^2]$, and thus

$$\gamma(\triangle) \stackrel{\triangle}{=} \frac{\rho_\triangle}{\sigma_\triangle^3} = \frac{(1-\triangle)^2 + \triangle^2}{\sqrt{\triangle(1-\triangle)}}.$$

Using (3.16) to write

$$\begin{aligned}
\Phi(t) - G(t,m,n) &= \mathbb{P}(Z_n^*(\triangle_{m,n,t}) < -c_{m,n,t}) - [1 - \Phi(t)] \\
&= \mathbb{P}(Z_n^*(\triangle_{m,n,t}) < -c_{m,n,t}) - \Phi(-c_{m,n,t}) + \Phi(t) - \Phi(c_{m,n,t}).
\end{aligned}$$

we have, by (3.17),

$$\begin{aligned}
|\mathbb{P}(Z_n^*(\Delta_{m,n,t}) < -c_{m,n,t}) - \Phi(-c_{m,n,t})| &\le \sup_{-\infty<x<\infty} |\mathbb{P}(Z_n^*(\Delta_{m,n,t}) < x) - \Phi(x)| \\
&\le Cn^{-1/2}\gamma(\Delta_{m,n,t}) \\
&= Cn^{-1/2}\frac{(1-\Delta_{m,n,t})^2 + \Delta_{m,n,t}^2}{\sqrt{\Delta_{m,n,t}(1-\Delta_{m,n,t})}}.
\end{aligned}$$

Now, $F(x,m) = \mathbb{E}\mathbb{P}(\bar{X}_m(Z) \le x) = \alpha(x) + \frac{1}{2m} f_{Y(x)}^{(1)}(0) + o\left(\frac{1}{m}\right)$ *uniformly* in $|x - q(p)| < \epsilon$, by Lemma 3.2.1.

So $\Delta_{m,n,t} = \alpha(q(p) + Atn^{-1/2}) + \frac{1}{2m} f_{Y(q(p)+Atn^{-1/2})}^{(1)}(0) + o(1/m)$. As a consequence, as $c \nearrow +\infty$,

$$\gamma(\Delta_{m,n,t}) \to \frac{(1-p)^2 + p^2}{\sqrt{p(1-p)}} > 0$$

uniformly in c. Thus, $\gamma(\Delta_{m,n,t})n^{-1/2} \to 0$ as $m, n \nearrow +\infty$.

It remains to investigate $c_{m,n,t}$ as $m, n \nearrow +\infty$. We note that

$$\begin{aligned}
c_{m,n,t} &= \frac{\sqrt{n}(F(q(p) + Atn^{-1/2}, m) - p)}{\sqrt{\Delta_{m,n,t}(1-\Delta_{m,n,t})}} \qquad (3.18) \\
&= \frac{\alpha(q(p) + Atn^{-1/2}) - \alpha(q(p)) + \frac{1}{2m} f_{Y(q(p)+Atn^{-1/2})}^{(1)}(0) + o(1/m)}{\sqrt{\Delta_{m,n,t}(1-\Delta_{m,n,t})} \cdot tAn^{-1/2}} tA \\
&\to \frac{\alpha'(q(p))}{\sqrt{p(1-p)}} \cdot tA + \frac{\sqrt{b}}{2a} \frac{1}{\sqrt{p(1-p)}} f_{Y(q(p))}^{(1)}(0)
\end{aligned}$$

by Conditions 3 and 4.

Denote $G(t,c) \triangleq G(t, a\cdot c^{1/3}, b\cdot c^{1/3})$ and let $A = \sqrt{p(1-p)}/\alpha'(q(p))$. Then, as $c \nearrow +\infty$, for any $t \in \mathbb{R}$,

$$\begin{aligned}
\lim_{c\nearrow+\infty} G(t,c) &= \lim_{c\nearrow+\infty} \Phi(c_{ac^{1/3}, bc^{2/3}, t}) \\
&= \Phi\left(t + \frac{\sqrt{b}}{2a} \frac{1}{\sqrt{p(1-p)}} f_{Y(q(p))}^{(1)}(0)\right)
\end{aligned}$$

In other words,

$$\lim_{c\nearrow+\infty} \mathbb{P}\left(\frac{\sqrt{b}c^{1/3}(\hat{q}(p, ac^{1/3}, bc^{1/3}) - q(p))}{\sqrt{p(1-p)}/\alpha'(q(p))} \le t\right) = \Phi\left(t + \frac{\sqrt{b}}{2a}\frac{1}{\sqrt{p(1-p)}} f^{(1)}_{Y(q(p))}(0)\right)$$

or

$$\sqrt{b}c^{1/3}\frac{\hat{q}(p, ac^{1/3}, bc^{2/3}) - q(p)}{\sqrt{p(1-p)}/\alpha'(q(p))} \overset{\mathcal{D}}{\Rightarrow} -\frac{\sqrt{b}}{2a}\frac{1}{\sqrt{p(1-p)}} f^{(1)}_{Y(q(p))}(0) + N(0,1)$$

as $c \nearrow +\infty$. We can rewrite this result as

$$c^{1/3}(\hat{q}(p, m, n) - q(p)) \Rightarrow \sqrt{\frac{p(1-p)}{b}} \cdot \frac{1}{\alpha'(q(p))} N(0,1) - \frac{1}{2a} f^{(1)}_{Y(q(p))}(0) \frac{1}{\alpha'(q(p))}$$

as $c \nearrow +\infty$.

The proof for the second result can be obtained similarly, with the only exception that $c_{m,n,t}$ (as in equation(3.18)) in this case has the following limit:

$$\begin{aligned}
c_{m,n,t} &= \frac{\sqrt{n}(F(q(p) + Atn^{-1/2}, m) - p)}{\sqrt{\Delta_{m,n,t}(1 - \Delta_{m,n,t})}} \\
&= \frac{\alpha(q(p) + Atn^{-1/2}) - \alpha(q(p)) + \frac{1}{2m} f^{(1)}_{Y(q(p)+Atn^{-1/2})}(0) + o(1/m)}{\sqrt{\Delta_{m,n,t}(1 - \Delta_{m,n,t})} \cdot tAn^{-1/2}} tA \\
&\to \frac{\alpha'(q(p))}{\sqrt{p(1-p)}} \cdot tA
\end{aligned}$$

as $c \nearrow +\infty$, since $\sqrt{n}/m = \sqrt{b}/a \cdot c^{(1-\gamma)/2-\gamma}$ and $\gamma > 1/3$. As in the proof for the first result, we let $A = \sqrt{p(1-p)}/\alpha'(q(p))$. As $c \nearrow +\infty$, for any $t \in \mathbb{R}$,

$$\lim_{c\nearrow+\infty} \mathbb{P}\left(\frac{\sqrt{b}c^{1/3}(\hat{q}(p, ac^{1/3}, bc^{1/3}) - q(p))}{\sqrt{p(1-p)}/\alpha'(q(p))} \le t\right) = \Phi(t),$$

which is just our second result.

To prove the third result, we appeal to the Hoeffding's lemma [17] which states that for independent random variables $\{Y_1, \ldots, Y_n\}$ satisfying $\mathbb{P}(a \le Y_i \le b) = 1$ for each i, where

$a < b$,

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \mathbb{E}Y_i \geq nt\right) \leq e^{-2nt^2/(b-a)^2}$$

for all $t > 0$. We would like to show that for all $\eta > 0$,

$$\mathbb{P}\left(\left|c^{\gamma}(\hat{q}_c(p) - q(p)) + \frac{f^{(1)}_{Y(q(p))}(0)}{2a\alpha'(q(p))}\right| > \eta\right) \to 0$$

as $c \nearrow +\infty$. Write

$$\begin{aligned} &\mathbb{P}\left(\left|c^{\gamma}(\hat{q}_c(p) - q(p)) + f^{(1)}_{Y(q(p))}(0)/(2a\alpha'(q(p)))\right| > \eta\right) = \\ &\mathbb{P}\left(\hat{q}_c(p) > q(p) + c^{-\gamma}d_1\right) + \mathbb{P}\left(\hat{q}_c(p) < q(p) + c^{-\gamma}d_2\right). \end{aligned} \tag{3.19}$$

where $d_1 \triangleq -f^{(1)}_{Y(q(p))}(0)/(2a\alpha'(q(p))) + \eta$ and $d_2 \triangleq -f^{(1)}_{Y(q(p))}(0)/(2a\alpha'(q(p))) - \eta$. Note that

$$\begin{aligned} \mathbb{P}(\hat{q}_c(p) > q(p) + c^{-\gamma}d_1) &= \mathbb{P}(p > \hat{\alpha}(q(p) + c^{-\gamma}d_1, m, n)) \\ &= \mathbb{P}\left(np > \sum_{i=1}^{n} I(\bar{X}_m(Z_i) \leq q(p) + c^{-\gamma}d_1)\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{n} I(\bar{X}_m(Z_i) > q(p) + c^{-\gamma}d_1) > n(1-p)\right) \\ &= \mathbb{P}\left(\sum_{i=1}^{n} V_i(m) - \sum_{i=1}^{n} \mathbb{E}V_i(m) > n\delta_1(m)\right). \end{aligned}$$

where

$$V_i(m) \triangleq I(\bar{X}_m(Z_i) > q(p) + c^{-\gamma}d_1)$$

and

$$\delta_1(m) \triangleq \mathbb{P}(\bar{X}_m(Z_i) \leq q(p) + c^{-\gamma}d_1) - p.$$

Likewise, we can deduce that

$$\mathbb{P}(\hat{q}_c(p) < q(p) + c^{\gamma}d_2) \leq \mathbb{P}\left(\sum_{i=1}^{n} W_i - \sum_{i=1}^{n} \mathbb{E}W_i \geq n\delta_2(m)\right).$$

where

$$W_i \triangleq I(\bar{X}_m(Z_i) \leq q(p) + c^{-\gamma} d_2)$$

and

$$\delta_2(m) \triangleq p - \mathbb{P}(\bar{X}_m(Z_i) \leq q(p) + c^{-\gamma} d_2).$$

Let's now examine closely the term $\delta_1(m)$. Let $C(\epsilon, \eta) \triangleq (d_1/\epsilon)^{1/\gamma}$ so that for all $c \geq C(\epsilon, \eta)$, we have $|c^{-\gamma} d_1| < \epsilon$.

$$\begin{aligned}
\delta_1(m) &= \mathbb{P}\left(\bar{X}_m(Z_i) \leq q(p) + c^{-\gamma} d_1\right) - p \\
&= \mathbb{P}\left(\bar{X}_m(Z_i) \leq q(p) + c^{-\gamma} d_1\right) - \alpha(q(p) + c^{-\gamma} d_1) \\
&\quad + \alpha(q(p) + c^{-\gamma} d_1) - \alpha(q(p)) \\
&= \frac{1}{2m} f^{(1)}_{Y(q(p)+c^{-\gamma} d_1)}(0) + o(m^{-1}) + c^{-\gamma} d_1 \alpha'(q(p) + \xi_c)
\end{aligned}$$

uniformly for $c \geq C(\epsilon, \eta)$ where ξ_c is between 0 and $c^{-\gamma} d_1$ and $\xi_c \to 0$ as $c \nearrow +\infty$.

Hence, by the definition of d_1, we have that

$$\delta_1(m) = \alpha'(q(p) + \xi_c) \cdot \eta c^{-\gamma} + \frac{f^{(1)}_{Y(q(p))}(0)}{2m} \left[\frac{f^{(1)}_{Y(q(p)+c^{-\gamma} d_1)}(0)}{f^{(1)}_{Y(q(p))}(0)} - \frac{\alpha'(q(p) + \xi_c)}{\alpha'(q(p))} \right] + o\left(\frac{1}{m}\right)$$

uniformly for $c \geq C(\epsilon, \eta)$. By the assumption of the third result, we have that $f^{(1)}_{Y(x)}(0)$ is continuous on $|x - q(p)| < \epsilon$. So, $d_1(m) = \alpha'(q(p)) \cdot \eta c^{-\gamma}(1 + o(1))$ uniformly for $c \geq C(\epsilon, \eta)$.

Recall that $n(c) = bc^{1-\gamma}$ and notice that $0 \leq V_i(m) \leq 1$ for all $i = 1, \ldots, n$. By utilizing Hoeffding's lemma, we have

$$\begin{aligned}
\mathbb{P}(\hat{q}_c(p) > q(p) + c^{-\gamma} d_1) &\leq \exp(-2bc^{1-\gamma}(\alpha'(q(p))\eta c^{-\gamma}(1 + o(1)))^2) \\
&= \exp(-2b(\eta\alpha'(q(p)))^2 c^{1-3\gamma}(1 + o(1))) \\
&\to 0
\end{aligned}$$

as $c \nearrow +\infty$ since $1 - 3\gamma < 0$.

Similarly, we obtain that $d_2(m) = \alpha'(q(p)) \cdot \eta c^{-\gamma}(1 + o(1))$ uniformly for $c \geq C(\epsilon, \eta)$ and that

$$\begin{aligned}
\mathbb{P}(\hat{q}_c(p) < q(p) + c^{-\gamma} d_2) &\leq \exp(-2bc^{1-\gamma}(\alpha'(q(p))\eta c^{-\gamma}(1 + o(1)))^2) \\
&= \exp(-2b(\eta\alpha'(q(p)))^2 c^{1-3\gamma}(1 + o(1))) \\
&\to 0
\end{aligned}$$

as $c \nearrow +\infty$.

Hence, by combining the two convergence results above, we have completed the proof. $\square$

Remark 3.2.1 *Heuristically, we can also verify that the two limit results (3.4) and (3.13) are consistent. By substituting $x = q(p)$ into the left hand side of results (3.4) and using result (3.13), we get that*

$$\begin{aligned}
& c^{1/3}(\hat{\alpha}(q(p), ac^{1/3}, bc^{2/3}) - \alpha(q(p))) \\
= \; & c^{1/3}(\hat{\alpha}(q(p), ac^{1/3}, bc^{2/3}) - \hat{\alpha}(\hat{q}(p, ac^{1/3}, bc^{2/3}), ac^{1/3}, bc^{2/3})) \\
\sim \; & -c^{1/3}\hat{\alpha}'(q(p), ac^{1/3}, bc^{2/3})(\hat{q}(p, ac^{1/3}, bc^{2/3}) - q(p)) \\
\sim \; & -c^{1/3}\alpha'(q(p))(\hat{q}(p, ac^{1/3}, bc^{2/3}) - q(p)) \\
\overset{\mathcal{D}}{\Rightarrow} \; & -\sqrt{\frac{p(1-p)}{b}}\mathrm{N}(0, 1) + \frac{1}{2a} f^{(1)}_{Y(q(p))}(0)
\end{aligned}$$

which is just the R.H.S. of (3.4).

3.3 Methodology

In this section, we will provide some methodology for the implementation of the estimators of both the distribution function and the quantile of $\mathbb{E}(X|Z)$. We will relate each method to the appropriate theoretical results mentioned in earlier chapters.

Let us first start with the distribution function estimator. We have already seen in the Proposition 3.1.5 that the fastest rate of convergence of the estimator

$$\hat{\alpha} = \frac{1}{n(c)} \sum_{i=1}^{n(c)} I\left(\frac{1}{m(c)} \sum_{j=1}^{m(c)} X_j(Z_i) \leq 0\right)$$

to α is $O(c^{-1/3})$. This can be achieved when $m(c) \sim ac^{\gamma}$ and $n(c) \sim a^{-1}c^{1-\gamma}$, where γ is set to $1/3$. (Without loss of generality, we will assume that $mn = c$.) For other choices of γ, the convergence rate of $\hat{\alpha}$ will be suboptimal. To achieve the fastest rate of convergence, we may directly implement the estimator in the following:

Algorithm 3.3.1

Step 0. *Initialization.* Input c and a.

Step 1. *Determine the sample sizes.* Set $(m, n) \triangleq (ac^{1/3}, a^{-1}c^{2/3})$.

Step 2. *Determine $\hat{\alpha}$.* Set $\hat{\alpha} \triangleq \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \leq 0\right)$. To be repetitive, we sample n Z's; for each Z_i sampled, we sample m X's conditioned on $Z = Z_i$. The estimator $\hat{\alpha}$ which is simply taken as the sample average of the indicator functions is our *point estimate* of α.

Step 3. *Form the $(1-\xi) \times 100\%$ confidence interval for $\hat{\alpha}$.* A consistent estimate of the standard error (s.e.), $s_{\hat{\alpha}}$, of $\hat{\alpha}$ is $\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{n}}$ since $\hat{\alpha} \sim \alpha$. The c.i., according to Proposition 3.1.5, is equal to

$$\left[\hat{\alpha} - z_{\xi/2}s_{\hat{\alpha}}, \hat{\alpha} + z_{\xi/2}s_{\hat{\alpha}}\right],$$

where $z_{\xi/2}$ is the $\xi/2$-quantile of a N(0, 1) r.v.

The biggest disadvantage of Algorithm 3.3.1 is that the (unobservable) bias of the estimator diminishes as fast as its standard deviation. (c.f. Part 1 of Proposition 3.1.5) Mathematically, this implies that the coverage error of the c.i. obtained does not converge to zero as $c \nearrow +\infty$. This fact will be confirmed in the next section when we look at the numerical results.

One way to get around this problem is to choose $\gamma > \frac{1}{3}$. By appealing to Part 2 of Proposition 3.1.5, we know that asymptotically, the size of the bias will become insignificant compared to that of the standard error of the estimator and hence the coverage probability of the confidence interval constructed will be correct. However, this way, we need to sacrifice the optimal rate of convergence.

It is now apparent that in order to obtain an estimator with the correct asymptotic coverage probability and with the optimal convergence rate simultaneouly, we need to somehow estimate the unobservable bias. One potential way to achieve this is via bootstrapping. However, here we consider an alternative way.

Specifically, for any fixed computational budget c, we allocate c^{δ} for the purpose of estimating the bias, where $0 < \delta < 1$. (For example, $\delta = 0.5$ will suffice). For c sufficiently large, this portion of the computational budget is negligible compared with the rest of the budget. Let

$$\alpha^{*} \triangleq \frac{1}{n^{*}} \sum_{i=1}^{n^{*}} I\left(\frac{1}{m^{*}} \sum_{j=1}^{m^{*}} X_j(Z_i) \leq 0\right).$$

where $(m^{*}, n^{*}) \triangleq (ac^{\gamma\delta}, a^{-1}c^{(1-\gamma)\delta}/m)$ for some $a > 0$. Here we choose $\gamma < 1/3$ (e.g. $\gamma = 1/4$) so that by Part 2 of Proposition 3.1.5, $c^{\gamma\delta}(\alpha^{*} - \alpha)$ converges weakly to $f_Y^{(1)}(0)/(2a)$ only.

The latter term can be extracted (approximately) by considering $c^{\gamma\delta}(\alpha^{*} - \hat{\alpha})$. Here, we have used $\hat{\alpha}$ as the best esimate of α. By the converging together theorem, we immediately conclude that since $\gamma\delta < 1/3$, as $c \nearrow +\infty$,

$$\begin{aligned} c^{\gamma\delta}(\alpha^{*} - \hat{\alpha}) &= c^{\gamma\delta}(\alpha^{*} - \alpha) - c^{\gamma\delta - \frac{1}{3}} \cdot c^{\frac{1}{3}}(\hat{\alpha} - \alpha) \\ &\Rightarrow \frac{f_Y^{(1)}(0)}{2a}. \end{aligned}$$

Now, let $\tilde{\alpha} = \hat{\alpha} - c^{\gamma\delta - \frac{1}{3}}(\alpha^{*} - \hat{\alpha})$. Once again, by the converging together theorem, we can show that

$$\begin{aligned} c^{1/3}(\bar{\alpha} - \alpha) &= c^{1/3}(\hat{\alpha} - \alpha) - c^{\gamma\delta}(\alpha^{*} - \hat{\alpha}) \\ &\Rightarrow \sqrt{\alpha(1-\alpha)a}\,\mathrm{N}(0,1). \end{aligned}$$

In other words, the improved estimator $\bar{\alpha}$ is still converging at the optimal rate; furthermore, the confidence interval constructed for it will, asymptotically, have the correct coverage probability for c sufficiently large.

Yet, for a moderate size of c, the s.e. of $\bar{\alpha}$ may well be larger than $\sqrt{\frac{\alpha(1-\alpha)}{n}}$ and we need to "fine-tune" it appropriately to obtain the right coverage probability for the c.i. Specifically, we have that

$$\begin{aligned} c^{1/3}(\bar{\alpha} - \alpha) &= c^{1/3}(\hat{\alpha} - \alpha) - c^{\gamma\delta}(\alpha^{*} - \alpha) + c^{\gamma\delta}(\hat{\alpha} - \alpha) \\ &\overset{\mathcal{D}}{\approx} \sqrt{\alpha(1-\alpha)a}\,\mathcal{N}_1 - \sqrt{\alpha(1-\alpha)a}\,c^{\frac{3\gamma-1}{2}\delta}\mathcal{N}_2 \\ &\quad + \sqrt{\alpha(1-\alpha)a}\,c^{\gamma\delta-\frac{1}{3}}\mathcal{N}_1 + c^{\gamma\delta-\frac{1}{3}}\frac{f_Y^{(1)}(0)}{2a}. \end{aligned}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ are two independent $\mathrm{N}(0,1)$ r.v.'s. Hence,

$$\mathrm{Var}(\bar{\alpha}) \sim \frac{\alpha(1-\alpha)}{n}\left[(1 + c^{\gamma\delta-\frac{1}{3}})^2 + c^{(3\gamma-1)\delta}\right].$$

For a moderate size of c, $c^{(3\gamma-1)\delta}$ may still a significantly contribute towards the variance. We conjecture that it is necessary for us to scale up the standard error of $\bar{\alpha}$ accordingly.

The algorithm below gives a practical methodology for the implementation of $\bar{\alpha}$.

Algorithm 3.3.2

Step 0. *Initialization.* Input c, a, $0 < \gamma < \frac{1}{3}$ and $0 < \delta < 1$.

Step 1. *Determine the sample sizes.* Set $(m^{*}, n^{*}) \triangleq (ac^{\gamma\delta}, a^{-1}c^{(1-\gamma)\delta})$; set $(m, n) \triangleq (ac^{1/3}, a^{-1}c^{2/3})$.

Step 2. *Determine* α^* *and* $\hat{\alpha}$. Set

$$\alpha^* \triangleq \frac{1}{n^*}\sum_{i=1}^{n^*} I\left(\frac{1}{m^*}\sum_{j=1}^{m^*} X_j(Z_i) \leq 0\right)$$

and

$$\hat{\alpha} \triangleq \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{m}\sum_{j=1}^{m} X_j(Z_i) \leq 0\right).$$

Step 3. *Estimate the bias.* Set $b \triangleq c^{\gamma\delta}(\alpha^* - \hat{\alpha})$.

Step 4. *Form a bias-reduced estimator.* Set $\tilde{\alpha} \triangleq \hat{\alpha} - c^{-1/3}b$, which is now our *point estimate* of α.

Step 5. *Form the* $(1-\xi)\times 100\%$ *confidence interval for* $\tilde{\alpha}$. A consistent estimate of the standard error, $s_{\tilde{\alpha}}$, of $\tilde{\alpha}$ is $\sqrt{\frac{\tilde{\alpha}(1-\tilde{\alpha})}{n}}\cdot\sqrt{(1+c^{\gamma\delta-1/3})^2 + c^{(3\gamma-1)\delta}}$. The c.i. is equal to

$$\left[\tilde{\alpha} - z_{\xi/2}s_{\tilde{\alpha}}, \tilde{\alpha} + z_{\xi/2}s_{\tilde{\alpha}}\right].$$

where $z_{\xi/2}$ is the $\xi/2$-quantile of a N(0, 1) r.v.

So far, we have considered methodologies for estimating the distribution function of $\mathbb{E}(X|Z)$. Let's consider now the methodology for estimating the pth-quantile of $\mathbb{E}(X|Z)$. Recall that our quantile estimator $\hat{q}(p)$ is given by

$$\hat{q}_c(p) \triangleq \bar{X}_{m(c)}(Z)_{(\lceil np \rceil)},$$

where $\bar{X}_{m(c)}(Z)_{(i)}$ is the i-th order statistic of the sample means

$$\{\bar{X}_{m(c)}(Z_i) : i = 1, \ldots, n(c)\}.$$

Again, we will appeal to the CLT developed in the last section to guide us in setting up the c.i. for the quantile estimator. There are, however, a number of numerical complications that we need to overcome in order to construct confidence intervals for the quantile estimator with the correct coverage probability. The first one is the bias of the estimator. In particular, let $\bar{q}_m(p)$ be the "population" p-th quantile of $\bar{X}_m(Z) = (1/m)\sum_{j=1}^{m} X_j(Z)$. We have already seen that $\bar{q}_m(p)$ is a biased estimator of $q(p)$. Furthermore, it was argued [12] that the p-th quantile from a sample of size z generated from a population having distribution F_X that is twice continuously differentiable is biased with $\mathbb{E}X_{(\lfloor np \rfloor)} = \alpha + \beta/n + o(1/n)$ where α is the p-th quantile of the population. With this supportive result at hand, we conjecture that

$$\mathbb{E}\hat{q}_c(p) \sim \bar{q}_{m(c)}(p) + \frac{b}{n(c)} + \frac{d}{n(c)^2}$$

for large c; i.e., $\hat{q}_c(p)$ is also a biased estimator of $\bar{q}_{m(c)}(p)$, which constitutes the second complication. Third, recall that, according to Theorem 3.2.1, the s.d. of the "noise" term on the RHS of the CLT is $\sqrt{ap(1-p)}/\alpha'(q(p))$. It involves the density $\alpha'(\cdot)$ of $\mathbb{E}(X|Z)$ at the unknown point q_p. It is important that we can accurately estimate the standard error of $\hat{q}_p$ in order to construct c.i. with the correct coverage probability.

We will tackle the second and third complications via the sectioning-with-jackknife approach. The advantages of this approach are that we can simultaneously correct the bias of the estimator and estimate the estimator's standard error. The first complication will be tackled by the bias-correcting technique we have used in Algorithm 3.3.2.

To illustrate the idea, suppose that we sample a total of $n(c) = sk(c)$ replications of $\bar{X}_{m(c)}(Z)$, where s is the number of sections. In particular, we partition all the $n(c)$ replications into s sections; each of which has $k(c)$ replications. Typically, s is relatively small (say 5 or 10), while the number of observations $k(c)$ per section is large. For $i = 1, \ldots, s$, let $\tilde{Q}_n(i)$ be the sample quantile associated with all the n replications, except for those in the i-th section, so that $\tilde{Q}_n(i)$ is the $\lfloor (s-1)k(c)p \rfloor$-th largest observation in the set

$$\{\bar{X}_m(Z_j) : j \in \{1, \ldots, n(c)\} \setminus \{(i-1)k(c)+1, \ldots, ik(c)\}\}.$$

It is clear that

$$\mathbb{E}\tilde{Q}_n(i) \sim \bar{q}_m(p) + \frac{b}{(s-1)k(c)} + \frac{d}{(s-1)^2 k(c)^2}.$$

Consequently, the "pseudo-value", $\xi_n(i)$ as defined by

$$\xi_n(i) = s\hat{q}(p) - (s-1)\tilde{Q}_n(i)$$

has bias (from $\tilde{q}(p)$) given by

$$\mathbb{E}\xi_n(i) \approx \tilde{q}(p) - \frac{d}{(s-1)sk(c)^2},$$

which is much smaller than does $\hat{q}(p)$. This suggests using the sectioning-with-jackknife estimator defined by

$$\xi_n^J = \frac{1}{s}\sum_{i=1}^{s}\xi_n(i)$$

as our estimator for $\tilde{q}(p)$. Let

$$v^J = \frac{1}{s-1}\sum_{i=1}^{s}(\xi_n(i) - \xi_n^J)^2$$

be our jackknifed variance estimator. Then, it can be shown that

$$\left[\xi_n^J - t\sqrt{\frac{v^J}{s}},\ \xi_n^J + t\sqrt{\frac{v^J}{s}}\right]$$

is an approximate $100(1-\delta)\%$ confidence interval for $\tilde{q}(p)$, where t is selected so that $\mathbb{P}(-t \le t_{s-1} \le t) = 1-\delta$, t_{s-1} being a Student-t r.v. with $s-1$ degrees of freedom. Finally, we use the bias-correcting technique in Algorithm 3.3.2 to correct the bias of $\tilde{q}(p)$ from $q(p)$.

The whole algorithm for estimating the p-th quantile of the conditional expectation $\mathbb{E}(X|Z)$ can be summarized as follows:

Algorithm 3.3.3

Step 0. *Initialization.* Input c, a, $0 < \gamma < \frac{1}{3}$, $0 < \delta < 1$, $s = 5, \ldots, 10$ and $0 < p < 1$.

Step 1. *Determine the sample sizes.* Set $(m^*, n^*) \triangleq (ac^{\gamma\delta}, a^{-1}c^{(1-\gamma)\delta})$; set $(m, n) \triangleq (ac^{1/3}, a^{-1}c^{2/3})$.

Step 2. *Determine q^* and $\hat{q}$.* Set

$$q^* \stackrel{\Delta}{=} \bar{X}_m(Z)_{(\lfloor n^* p \rfloor)}$$

and

$$\hat{q} \stackrel{\Delta}{=} \bar{X}_m(Z)_{(\lfloor np \rfloor)}.$$

Step 3. *Compute the s pseudo-values.* For each $i \in \{1, \ldots, s\}$, compute the sample quantile estimator $\tilde{Q}_n(i)$ based on all the observations except those associated with the i-th section. Then, compute the pseudo-values

$$\xi_n(i) = s\hat{q} - (s-1)\tilde{Q}_n(i).$$

Step 4. *Compute the jackknifed estimators.* Set

$$\xi_n^J = \frac{1}{s}\sum_{i=1}^{s}\xi_n(i)$$

and

$$v^J = \frac{1}{s-1}\sum_{i=1}^{s}(\xi_n(i) - \xi_n^J)^2.$$

Step 5. *Estimate the bias.* Set $b \stackrel{\Delta}{=} c^{\gamma\delta}(q^* - \xi_n^J)$.

Step 6. *Form a bias-reduced estimator.* Set $\hat{q} \stackrel{\Delta}{=} \xi_n^J - c^{-1/3}b$, which is now our point estimate of $q(p)$.

Step 7. *Form the $100(1-\eta)\%$ confidence interval for $\hat{q}$.* A consistent estimate of the standard error, $s_{\hat{q}}$, of $\hat{q}$ is $\sqrt{v^J} \cdot \sqrt{(1 + c^{\gamma\delta - 1/3})^2 + c^{(3\gamma-1)\delta}}$. The c.i. is equal to

$$\left[\hat{q} - t\sqrt{\frac{v^J}{s}},\ \hat{q} + t\sqrt{\frac{v^J}{s}}\right],$$

where t is selected so that $\mathbb{P}(-t \le t_{s-1} \le t) = 1 - \eta$, t_{s-1} being a Student-t r.v. with $s-1$ degrees of freedom.

3.4 Numerical Results

In this section, we will report the results of a Monte Carlo study of the coverage characterics of the confidence intervals of both the distribution function and the quantile estimates of $\mathbb{E}(X|Z)$. We will also verify that the theoretical optimal rates of convergence are actually attained if we choose the estimation parameters appropriately and that the bias-reduction scheme is successful, at least empirically.

First, we will give a description of each of the examples. Both the distribution function and the quantile of the conditional expectation in these examples can be evaluated in closed form. These theoretical results will be used as a benchmark against which the numerical results are compared. We will, in particular, show how the theoretical results can be computed.

Then, we will show the numerical results of each example. Special attentions will be paid to the convergence rates of the estimators and to the coverage probabilility of the confidence interval constructed according to the algorithms outlined in the previous section.

Example 3.1 *Let X and Z be two real-valued random variables. Assume that $Z \stackrel{D}{=} N(1,1)$ and that conditioned on $Z = z$, $X \stackrel{D}{=} N(z,1)$, i.e. $\mathbb{E}(X|Z) = Z$. Let $\alpha = \Phi(-1) = 0.1587$. Clearly, we have that $\mathbb{P}(\mathbb{E}(X|Z) \leq 0) = \alpha$ and that $q(\alpha) = 0.0$. The ICCV, Y, in this example is simply equal to Z and has density function $f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-(y-1)^2/2}$. In particular, $f_Y^{(1)}(0) = e^{-1/2}/\sqrt{2\pi}$. As we will soon see, the value of $f_Y^{(1)}(0)$ helps us determine the size of the theoretical MSE of the estimator.*

The next example to be presented is a very simple example of Value at Risk (VaR).

Example 3.2 *Consider a portfolio that consists of a long position of one forward contract (with expiration date $T = 1$ year) on a security in the same settings as in Example 1 in Chapter 1. Specifically, assume that the short interest rate is equal to a constant $r = 0.06$ and that the security price, S, is a geometric Brownian motion with constant parameters $\mu = 0.08$ and $\sigma = 0.12$. Assume that $S_0 = \$100.00$. The no-arbitrage forward price F at* *tiem 0 is equal to* $S_0 \exp(rT) = \$106.18$. *Let* V_t *be the value of the portfolio at time* t *and set* $\tau = 0.5$.

The probability $\alpha \stackrel{\Delta}{=} \mathbb{P}(V_\tau \leq 0)$ *that the value of the portfolio at time* τ *is negative can be solved in closed form and is equal to* $\Phi((r - \mu + \sigma^2/2)\sqrt{\tau}/\sigma) = 0.1587$. *as shown earlier in Chapter 1. In other words, the* α*-th quantile of* V_τ *is equal to 0.*

Recall that the conditional random element Z *in this example is equal to* S_τ *and the target variable* X *is* $S_T - F$. *Denote by* Q *the EMM. (see chapter 1.) Now the conditional expectation is given by* $\mu(S_\tau) = \mathbb{E}^Q(S_T - F|S_\tau) = S_\tau e^{r(T-\tau)} - F$ *and the conditional variance is given by* $\sigma^2(S_\tau) = \mathbb{E}^Q(S_T^2|S_\tau) - (\mathbb{E}^Q(S_T|S_\tau))^2 = S_\tau^2 e^{2r(T-\tau)}(e^{\sigma^2(T-\tau)} - 1)$. *Hence, the ICCV,* Y, *is equal to*

$$Y(S_\tau) = \frac{1 - \frac{S_0}{S_\tau} e^{r\tau}}{\sqrt{e^{\sigma^2(T-\tau)} - 1}}.$$

Since, under the original probability $\mathbb{P}$, S_τ *has a lognormal distribution with parameters* $(\mu\tau, \sigma^2\tau)$, *we can analytically evaluate* $f_Y^{(1)}(0)$ *as*

$$f_Y^{(1)}(0) = (e^{\sigma^2(T-\tau)} - 1)\frac{\phi(\sigma^{-1}(r - \mu + \sigma^2/2)\sqrt{\tau})}{\sigma\sqrt{\tau}}\left(1 - \frac{r - \mu + \sigma^2/2}{\sigma^2}\right) = 0.064.$$

The propsed algorithms presented in this thesis were programmed in ANSI-C and compiled and run on a Pentium-II machine. For each algorithm, we replicate the estimator 200 times. Denote by $\{\hat{\alpha}_i(c) : i = 1, \ldots, 200\}$ the values of the 200 replicates of the estimator $\hat{\alpha}$, given that the computational budget is equal to c. In the numerical experiments, since most of the computational budget is spent on sampling the X's, we take c to be the number of X's to sample and we set $c = 2^e$ where e ranges from 10 to 20. We estimate the mean, standard error, bias, and MSE of the estimator as follows:

mean: set $\bar{\alpha}(c) \stackrel{\Delta}{=} (1/200)\sum_{i=1}^{200} \hat{\alpha}_i(c)$;

standard error: set $s_{\hat{\alpha}}(c) \stackrel{\Delta}{=} \sqrt{(200-1)^{-1}\sum_{i=1}^{200}(\hat{\alpha}_i(c) - \bar{\alpha}(c))^2}$;

bias: set $b_{\hat{\alpha}}(c) \stackrel{\Delta}{=} \bar{\alpha}(c) - \alpha$, where α is the exact theoretical value;

MSE: set $\mathrm{MSE}_{\hat{\alpha}}(c) \stackrel{\Delta}{=} (200)^{-1}\sum_{i=1}^{200}(\hat{\alpha}_i(c) - \alpha)^2$.

The empirical converence rate of the estimator $\hat{\alpha}(c)$ to α can be deduced as follows. First, we plot log(MSE(c)) vs. log(c). If the plot turns out to be linear, it shows that

$$\text{MSE}(c) \sim Vc^{\lambda}. \tag{3.20}$$

where log(V) is estimated by the y-intercept of the plot and λ is estimated by the slope. In other words, $-\lambda$ is equal to the square of the rate of convergence of the estimator. For example, for the MSE associated with Algorithm 3.3.1, we know, according to equation (3.3) that $V \sim (3/2)(\alpha(1-\alpha)f_Y^{(1)}(0)/2)^{2/3}$ and $\lambda \sim -2/3$; i.e., rate of convergence of Algorithm 3.3.1 is $O(c^{-1/3})$.

Estimator	Theoretical Slope	Regressed Slope	Theoretical Intercept	Regressed Intercept
Example 3.1				
$\hat{\alpha}_c^1$	-0.6667	-0.6217	-2.1141	-2.4744
$\hat{q}_c^1(\alpha)$	-0.6667	-0.6763	0.7238	1.0000
$\hat{\alpha}_c^2$	-0.5000	-0.4857	-2.5195	-2.7037
$\hat{q}_c^2(\alpha)$	-0.5000	-0.5136	0.3184	0.5094
$\hat{\alpha}_c^3$	-0.6667	-0.7174	-2.1141	-1.5493
$\hat{q}_c^3(\alpha)$	-0.6667	-0.6917	0.7238	1.1347
Example 3.2				
$\hat{\alpha}_c^1$	-0.6667	-0.6400	-2.8140	-2.8663
$\hat{q}_c^1(\alpha)$	-0.6667	-0.7040	3.4262	4.2589
$\hat{\alpha}_c^2$	-0.5000	-0.4844	-2.9884	-3.1662
$\hat{q}_c^2(\alpha)$	-0.5000	-0.5090	3.2518	3.4127
$\hat{\alpha}_c^3$	-0.6667	-0.7124	-2.8140	-1.2232
$\hat{q}_c^3(\alpha)$	-0.6667	-0.6851	3.4262	4.6246

Table 3.1: Slopes and Intercepts of the plots in Example 3.1 and 3.2

The plots (Figiure 3.1–3.4) all look nearly linear, which suggests that equation (3.20) is a very good model. The theoretical and empircal values of the rate of convergence are summarized in Table 3.1 for comparison. Here $\hat{\alpha}_c^1$ and $\hat{q}_c^1(\alpha)$ are the estimators of α and $q(\alpha)$ respectively when $m(c)$ and $n(c)$ are chosen to achieve the convergence rate of $O(c^{-1/3})$; similarly, $(\hat{\alpha}_c^2, \hat{q}_c^2(\alpha))$ are the estimators whose convergence rates are $O(c^{-1/4})$.

The estimators $(\hat{\alpha}_c^3, \hat{q}_c^3(\alpha))$ are bias-reduced and have convergence rates of $O(c^{-1/3})$. Table 3.1 shows that the empirical values match pretty well the theoretical ones.

Out of the 200 experiments, we tested the number of times, N, the confidence intervals as constructed according to the proposed algorithms covered the true value. The corresponding estimated coverage probability is then set to $\hat{p} \triangleq N/200$. The standard error of the estimated coverage probability is given by $\sqrt{\hat{p}(1-\hat{p})/200}$ and is expressed inside the parenthesis beside the corresponding probability in the tables.

From the tables following, we see that the bias reduction technique is very effective in reducing the bias of the estimator. The coverage probabilities for the c.i. converge to the correct values if we either use an estimator with a sub-optimal rate of convergence or use the bias reduction technique. Moreover, the tables show that the bias-reduction technique does not slow down the rate of convergence of the estimator. It also shows that "scaling-up" of the s.e. of the estimator in Algorithm 3.3.2 and 3.3.3 are necessary.

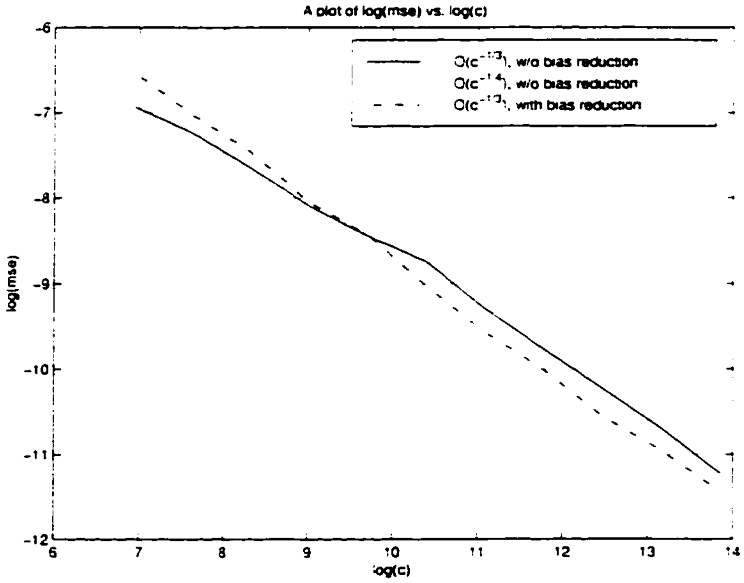


Figure 3.1: Distribution Function Estimator for Example 3.1

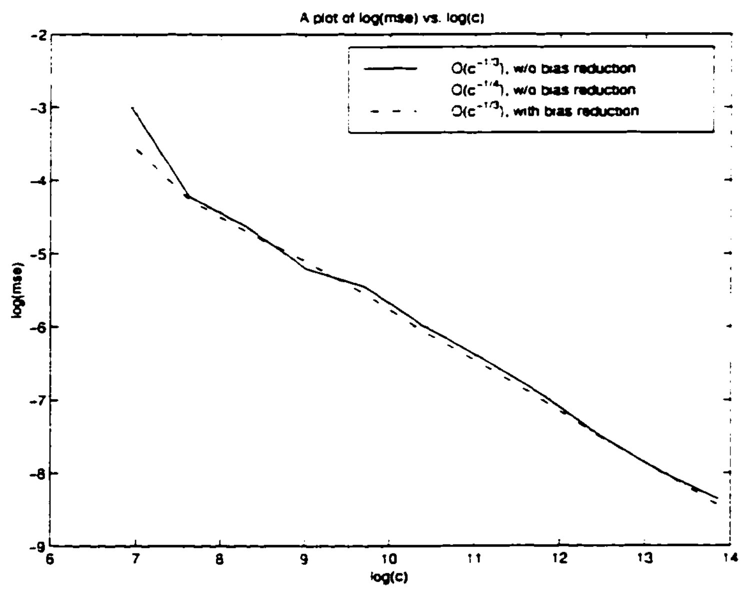


Figure 3.2: Quantile Estimator for Example 3.1

e	mean	s.d.	bias	ln(MSE)	90% cov.		95% cov.		99% cov.	
$O(c^{-1/3})$: without bias-reduction										
10	0.1786	0.0243	0.0199	-6.9288	0.75	(0.03)	0.86	(0.02)	0.95	(0.02)
11	0.1745	0.0218	0.0158	-7.2379	0.86	(0.02)	0.91	(0.02)	0.99	(0.01)
12	0.1711	0.0180	0.0124	-7.6484	0.86	(0.02)	0.94	(0.02)	0.98	(0.01)
13	0.1685	0.0145	0.0098	-8.0925	0.86	(0.02)	0.92	(0.02)	1.00	(0.00)
14	0.1671	0.0120	0.0085	-8.4472	0.81	(0.03)	0.88	(0.02)	0.96	(0.01)
15	0.1671	0.0094	0.0085	-8.7483	0.83	(0.03)	0.91	(0.02)	0.97	(0.01)
16	0.1649	0.0073	0.0063	-9.2972	0.81	(0.03)	0.91	(0.02)	0.96	(0.01)
17	0.1633	0.0060	0.0046	-9.7731	0.79	(0.03)	0.89	(0.02)	0.97	(0.01)
18	0.1625	0.0046	0.0039	-10.2326	0.83	(0.03)	0.88	(0.02)	0.97	(0.01)
19	0.1617	0.0037	0.0030	-10.7020	0.81	(0.03)	0.86	(0.02)	0.98	(0.01)
20	0.1610	0.0028	0.0024	-11.2327	0.82	(0.03)	0.87	(0.02)	0.96	(0.01)
$O(c^{-1/4})$: without bias-reduction										
10	0.1645	0.0481	0.0059	-6.0598	0.93	(0.02)	0.95	(0.02)	0.99	(0.01)
11	0.1607	0.0400	0.0021	-6.4415	0.92	(0.02)	0.97	(0.01)	1.00	(0.00)
12	0.1633	0.0339	0.0046	-6.7528	0.91	(0.02)	0.99	(0.01)	1.00	(0.00)
13	0.1609	0.0278	0.0022	-7.1665	0.94	(0.02)	0.97	(0.01)	1.00	(0.00)
14	0.1605	0.0247	0.0018	-7.4045	0.91	(0.02)	0.95	(0.02)	0.99	(0.01)
15	0.1599	0.0226	0.0012	-7.5849	0.89	(0.02)	0.95	(0.02)	0.99	(0.01)
16	0.1604	0.0172	0.0018	-8.1183	0.93	(0.02)	0.96	(0.01)	1.00	(0.00)
17	0.1598	0.0158	0.0012	-8.2914	0.87	(0.02)	0.93	(0.02)	0.99	(0.01)
18	0.1601	0.0121	0.0014	-8.8186	0.90	(0.02)	0.95	(0.02)	1.00	(0.00)
19	0.1602	0.0099	0.0016	-9.2153	0.92	(0.02)	0.98	(0.01)	1.00	(0.00)
20	0.1592	0.0090	0.0005	-9.4237	0.94	(0.02)	0.95	(0.02)	0.99	(0.01)
$O(c^{-1/3})$: with bias-reduction										
10	0.1726	0.0356	0.0140	-6.5314	0.94	(0.02)	0.97	(0.01)	1.00	(0.00)
11	0.1695	0.0279	0.0108	-7.0238	0.93	(0.02)	0.97	(0.01)	0.99	(0.01)
12	0.1688	0.0218	0.0102	-7.4619	0.91	(0.02)	0.95	(0.02)	1.00	(0.00)
13	0.1651	0.0167	0.0065	-8.0492	0.92	(0.02)	0.97	(0.01)	1.00	(0.00)
14	0.1648	0.0134	0.0062	-8.4333	0.89	(0.02)	0.95	(0.02)	0.99	(0.01)
15	0.1633	0.0099	0.0047	-9.0320	0.91	(0.02)	0.98	(0.01)	1.00	(0.00)
16	0.1625	0.0074	0.0039	-9.5752	0.91	(0.02)	0.95	(0.02)	1.00	(0.00)
17	0.1616	0.0060	0.0029	-10.0162	0.90	(0.02)	0.94	(0.02)	0.99	(0.01)
18	0.1603	0.0048	0.0016	-10.5609	0.91	(0.02)	0.98	(0.01)	1.00	(0.00)
19	0.1605	0.0037	0.0019	-10.9608	0.90	(0.02)	0.95	(0.02)	1.00	(0.00)
20	0.1599	0.0030	0.0012	-11.4471	0.90	(0.02)	0.95	(0.02)	0.99	(0.01)

Table 3.2: Numerical Results of the Distribution Function Estimator in Example 3.1

e	mean	s.d.	bias	ln(mse)	90% cov.		95% cov.		99% cov.	
$O(c^{-1/3})$: without bias-reduction										
10	-0.1851	0.1273	-0.1851	-2.9879	0.77	(0.03)	0.86	(0.02)	0.96	(0.01)
11	-0.0792	0.0921	-0.0792	-4.2194	0.82	(0.03)	0.89	(0.02)	0.95	(0.02)
12	-0.0590	0.0790	-0.0590	-4.6364	0.79	(0.03)	0.85	(0.03)	0.93	(0.02)
13	-0.0494	0.0555	-0.0494	-5.2019	0.75	(0.03)	0.84	(0.03)	0.96	(0.01)
14	-0.0441	0.0487	-0.0441	-5.4483	0.77	(0.03)	0.84	(0.03)	0.96	(0.01)
15	-0.0347	0.0363	-0.0347	-5.9866	0.80	(0.03)	0.90	(0.02)	0.99	(0.01)
16	-0.0280	0.0288	-0.0280	-6.4313	0.78	(0.03)	0.86	(0.02)	0.96	(0.01)
17	-0.0196	0.0246	-0.0196	-6.9221	0.77	(0.03)	0.87	(0.02)	0.96	(0.01)
18	-0.0147	0.0184	-0.0147	-7.4965	0.77	(0.03)	0.86	(0.02)	0.97	(0.01)
19	-0.0120	0.0141	-0.0120	-7.9796	0.77	(0.03)	0.87	(0.02)	0.98	(0.01)
20	-0.0093	0.0121	-0.0093	-8.3664	0.78	(0.03)	0.85	(0.03)	0.97	(0.01)

$O(c^{-1/4})$: without bias-reduction										
10	-0.0833	0.1969	-0.0833	-3.0896	0.77	(0.03)	0.87	(0.02)	0.95	(0.02)
11	-0.0847	0.1626	-0.0847	-3.3970	0.78	(0.03)	0.87	(0.02)	0.98	(0.01)
12	-0.0701	0.1463	-0.0701	-3.6416	0.75	(0.03)	0.87	(0.02)	0.97	(0.01)
13	-0.0407	0.1163	-0.0407	-4.1916	0.76	(0.03)	0.85	(0.03)	0.95	(0.02)
14	-0.0229	0.1009	-0.0229	-4.5421	0.80	(0.03)	0.90	(0.02)	0.97	(0.01)
15	-0.0202	0.0909	-0.0202	-4.7536	0.81	(0.03)	0.91	(0.02)	0.97	(0.01)
16	-0.0163	0.0710	-0.0163	-5.2429	0.83	(0.03)	0.90	(0.02)	0.98	(0.01)
17	-0.0109	0.0660	-0.0109	-5.4151	0.84	(0.03)	0.92	(0.02)	0.98	(0.01)
18	-0.0117	0.0498	-0.0117	-5.9497	0.84	(0.03)	0.90	(0.02)	0.97	(0.01)
19	-0.0092	0.0409	-0.0092	-6.3498	0.89	(0.02)	0.95	(0.02)	0.99	(0.01)
20	-0.0039	0.0376	-0.0039	-6.5553	0.89	(0.02)	0.95	(0.02)	0.99	(0.01)

$O(c^{-1/3})$: with bias-reduction										
10	-0.1193	0.1264	-0.1193	-3.5026	0.89	(0.02)	0.95	(0.02)	0.98	(0.01)
11	-0.0739	0.0931	-0.0739	-4.2622	0.88	(0.02)	0.93	(0.02)	0.99	(0.01)
12	-0.0592	0.0747	-0.0592	-4.7040	0.85	(0.03)	0.92	(0.02)	0.99	(0.01)
13	-0.0498	0.0601	-0.0498	-5.1040	0.88	(0.02)	0.94	(0.02)	0.99	(0.01)
14	-0.0375	0.0500	-0.0375	-5.5483	0.88	(0.02)	0.95	(0.02)	1.00	(0.00)
15	-0.0319	0.0362	-0.0319	-6.0638	0.89	(0.02)	0.95	(0.02)	0.99	(0.01)
16	-0.0266	0.0280	-0.0266	-6.5110	0.89	(0.02)	0.93	(0.02)	0.97	(0.01)
17	-0.0193	0.0234	-0.0193	-6.9930	0.90	(0.02)	0.95	(0.02)	0.99	(0.01)
18	-0.0134	0.0192	-0.0134	-7.5160	0.89	(0.02)	0.95	(0.02)	0.99	(0.01)
19	-0.0116	0.0146	-0.0116	-7.9658	0.90	(0.02)	0.95	(0.02)	0.99	(0.01)
20	-0.0084	0.0120	-0.0084	-8.4516	0.91	(0.02)	0.96	(0.01)	0.99	(0.01)

Table 3.3: Numerical Results of the Quantile Estimator in Example 3.1

e	mean	s.d.	bias	ln(mse)	90% cov.		95% cov.		99% cov.	
$O(c^{-1/3})$: without bias-reduction										
10	0.4833	0.0200	0.0133	-7.4610	0.89	(0.02)	0.93	(0.02)	0.99	(0.01)
11	0.4817	0.0175	0.0117	-7.7264	0.82	(0.03)	0.86	(0.02)	0.98	(0.01)
12	0.4809	0.0133	0.0110	-8.1232	0.83	(0.03)	0.89	(0.02)	0.97	(0.01)
13	0.4767	0.0113	0.0068	-8.6651	0.83	(0.03)	0.89	(0.02)	0.97	(0.01)
14	0.4774	0.0087	0.0074	-8.9459	0.78	(0.03)	0.87	(0.02)	0.98	(0.01)
15	0.4760	0.0063	0.0060	-9.4833	0.82	(0.03)	0.90	(0.02)	0.96	(0.01)
16	0.4744	0.0053	0.0044	-9.9443	0.83	(0.03)	0.88	(0.02)	0.97	(0.01)
17	0.4734	0.0045	0.0035	-10.3355	0.83	(0.03)	0.89	(0.02)	0.96	(0.01)
18	0.4724	0.0034	0.0025	-10.9386	0.85	(0.03)	0.91	(0.02)	0.96	(0.01)
19	0.4718	0.0030	0.0018	-11.2866	0.79	(0.03)	0.87	(0.02)	0.95	(0.02)
20	0.4716	0.0022	0.0017	-11.8154	0.78	(0.03)	0.90	(0.02)	0.98	(0.01)
$O(c^{-1/4})$: without bias-reduction										
10	0.4735	0.0379	0.0036	-6.5397	0.91	(0.02)	0.95	(0.02)	0.99	(0.01)
11	0.4737	0.0321	0.0038	-6.8674	0.89	(0.02)	0.95	(0.02)	1.00	(0.00)
12	0.4772	0.0270	0.0073	-7.1592	0.93	(0.02)	0.96	(0.01)	0.98	(0.01)
13	0.4728	0.0245	0.0029	-7.4086	0.89	(0.02)	0.94	(0.02)	0.98	(0.01)
14	0.4732	0.0189	0.0033	-7.9207	0.93	(0.02)	0.95	(0.02)	0.99	(0.01)
15	0.4726	0.0160	0.0026	-8.2446	0.89	(0.02)	0.95	(0.02)	0.99	(0.01)
16	0.4728	0.0129	0.0029	-8.6616	0.92	(0.02)	0.96	(0.01)	1.00	(0.00)
17	0.4718	0.0116	0.0018	-8.8947	0.90	(0.02)	0.95	(0.02)	0.99	(0.01)
18	0.4706	0.0104	0.0006	-9.1353	0.90	(0.02)	0.94	(0.02)	0.99	(0.01)
19	0.4707	0.0086	0.0008	-9.5055	0.89	(0.02)	0.95	(0.02)	0.99	(0.01)
20	0.4709	0.0071	0.0009	-9.8872	0.90	(0.02)	0.95	(0.02)	1.00	(0.00)
$O(c^{-1/3})$: with bias-reduction										
10	0.4736	0.0428	0.0036	-6.3022	0.97	(0.01)	0.99	(0.01)	1.00	(0.00)
11	0.4760	0.0360	0.0060	-6.6251	0.96	(0.01)	0.97	(0.01)	1.00	(0.00)
12	0.4719	0.0309	0.0019	-6.9527	0.93	(0.02)	0.97	(0.01)	0.99	(0.01)
13	0.4738	0.0215	0.0039	-7.6546	0.92	(0.02)	0.97	(0.01)	1.00	(0.00)
14	0.4720	0.0184	0.0020	-7.9876	0.91	(0.02)	0.96	(0.01)	1.00	(0.00)
15	0.4724	0.0124	0.0024	-8.7438	0.96	(0.01)	0.99	(0.01)	1.00	(0.00)
16	0.4716	0.0094	0.0017	-9.3146	0.94	(0.02)	0.98	(0.01)	1.00	(0.00)
17	0.4713	0.0081	0.0010	-9.6134	0.92	(0.02)	0.97	(0.01)	1.00	(0.00)
18	0.4706	0.0066	0.0007	-10.0427	0.91	(0.02)	0.95	(0.02)	1.00	(0.00)
19	0.4705	0.0048	0.0005	-10.6705	0.93	(0.02)	0.98	(0.01)	1.00	(0.00)
20	0.4701	0.0040	0.0002	-11.0268	0.92	(0.02)	0.98	(0.01)	0.99	(0.01)

Table 3.4: Numerical Results of the Distribution Function Estimator in Example 3.2

e	mean	s.d.	bias	ln(mse)	90% cov.		95% cov.		99% cov.	
$O(c^{-1/3})$: without bias-reduction										
10	-0.4638	0.5570	-0.4638	-0.6465	0.82	(0.03)	0.90	(0.02)	0.95	(0.02)
11	-0.3191	0.4594	-0.3191	-1.1467	0.85	(0.03)	0.90	(0.02)	0.95	(0.02)
12	-0.3267	0.3451	-0.3267	-1.5023	0.77	(0.03)	0.87	(0.02)	0.96	(0.01)
13	-0.1878	0.2899	-0.1878	-2.1298	0.81	(0.03)	0.87	(0.02)	0.96	(0.01)
14	-0.1869	0.2104	-0.1869	-2.5385	0.82	(0.03)	0.90	(0.02)	0.98	(0.01)
15	-0.1522	0.1552	-0.1522	-3.0548	0.84	(0.03)	0.92	(0.02)	0.99	(0.01)
16	-0.1099	0.1299	-0.1099	-3.5452	0.83	(0.03)	0.90	(0.02)	0.99	(0.01)
17	-0.0840	0.1043	-0.0840	-4.0238	0.83	(0.03)	0.93	(0.02)	1.00	(0.00)
18	-0.0595	0.0803	-0.0595	-4.6092	0.86	(0.02)	0.91	(0.02)	0.98	(0.01)
19	-0.0441	0.0703	-0.0441	-4.9825	0.79	(0.03)	0.90	(0.02)	0.97	(0.01)
20	-0.0394	0.0509	-0.0394	-5.4906	0.78	(0.03)	0.88	(0.02)	0.96	(0.01)
$O(c^{-1/4})$: without bias-reduction										
10	-0.2253	0.8806	-0.2253	-0.1956	0.85	(0.03)	0.91	(0.02)	0.96	(0.01)
11	-0.1693	0.7886	-0.1693	-0.4348	0.85	(0.03)	0.93	(0.02)	0.98	(0.01)
12	-0.2705	0.6443	-0.2705	-0.7210	0.79	(0.03)	0.87	(0.02)	0.98	(0.01)
13	-0.1133	0.5727	-0.1133	-1.0812	0.83	(0.03)	0.92	(0.02)	0.97	(0.01)
14	-0.1082	0.4532	-0.1082	-1.5322	0.88	(0.02)	0.94	(0.02)	0.99	(0.01)
15	-0.0854	0.3772	-0.0854	-1.9050	0.88	(0.02)	0.93	(0.02)	0.98	(0.01)
16	-0.0918	0.2872	-0.0918	-2.4022	0.86	(0.02)	0.92	(0.02)	0.95	(0.02)
17	-0.0544	0.2582	-0.0544	-2.6692	0.89	(0.02)	0.94	(0.02)	0.99	(0.01)
18	-0.0315	0.2344	-0.0315	-2.8887	0.89	(0.02)	0.93	(0.02)	0.99	(0.01)
19	-0.0268	0.1952	-0.0268	-3.2535	0.89	(0.02)	0.94	(0.02)	0.98	(0.01)
20	-0.0259	0.1643	-0.0259	-3.5929	0.90	(0.02)	0.94	(0.02)	0.99	(0.01)
$O(c^{-1/3})$; with bias-reduction										
10	-0.3414	0.8438	-0.3414	-0.1924	0.87	(0.02)	0.92	(0.02)	0.99	(0.01)
11	-0.2274	0.7033	-0.2274	-0.6090	0.88	(0.02)	0.93	(0.02)	0.99	(0.01)
12	-0.1249	0.5969	-0.1249	-0.9940	0.86	(0.02)	0.91	(0.02)	0.97	(0.01)
13	-0.1445	0.4459	-0.1445	-1.5199	0.87	(0.02)	0.93	(0.02)	0.99	(0.01)
14	-0.1038	0.3816	-0.1038	-1.8601	0.86	(0.02)	0.94	(0.02)	0.98	(0.01)
15	-0.0904	0.2611	-0.0904	-2.5767	0.91	(0.02)	0.95	(0.02)	0.99	(0.01)
16	-0.0594	0.1969	-0.0594	-3.1679	0.92	(0.02)	0.98	(0.01)	0.99	(0.01)
17	-0.0530	0.1743	-0.0530	-3.4101	0.92	(0.02)	0.95	(0.02)	1.00	(0.00)
18	-0.0257	0.1414	-0.0257	-3.8851	0.88	(0.02)	0.97	(0.01)	1.00	(0.00)
19	-0.0178	0.1066	-0.0178	-4.4547	0.91	(0.02)	0.98	(0.01)	1.00	(0.00)
20	-0.0130	0.0895	-0.0130	-4.8102	0.94	(0.02)	0.96	(0.01)	1.00	(0.00)

Table 3.5: Numerical Results of the Quantile Estimator in Example 3.2

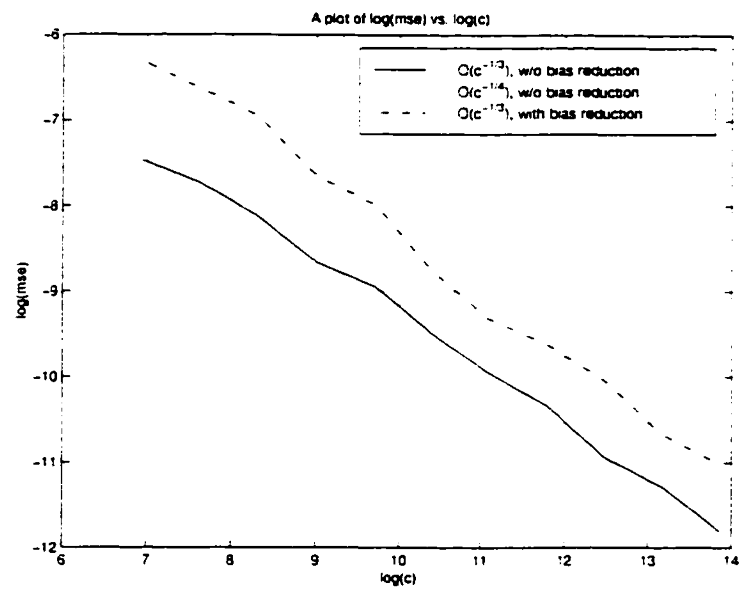


Figure 3.3: Distribution Function Estimator for Example 3.2

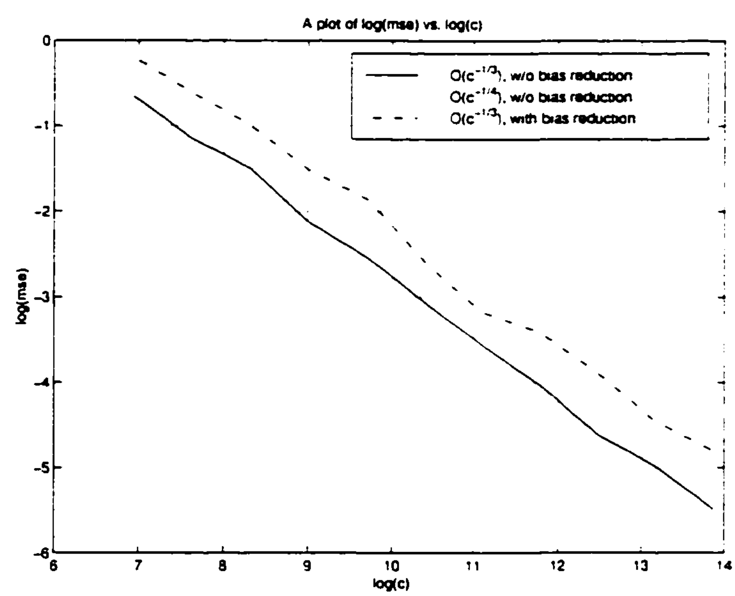


Figure 3.4: Quantile Estimator for Example 3.2

Bibliography

[1] BAHADUR, R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **38** 303–324.

[2] BAHADUR, R. and RAO, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[3] BHATTACHARYA, R. and RAO, R. (1976). *Normal Approximation and Asymptotic Expansions.* John Wiley and Sons. New York.

[4] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* John Wiley and Sons. New York.

[5] BILLINGSLEY, P. (1995). *Probability and Measure. 3rd ed.* John Wiley and Sons. New York.

[6] BOYLE, P. (1977). Options: a Monte Carlo approach. *Journal of Financial Economics* **4** 323–338.

[7] BOYLE, P., BROADIE, M. and GLASSERMAN, P. (1997). Monte Carlo methods for security pricing. *J. Econom. Dynam. Control* **21** 1267–1321.

[8] BREIMAN, L. (1968). *Probability.* Addison-Wesley Publishing Company, Menlo Park. California.

[9] BROADIE, M. and GLASSERMAN, P. (1997). A stochastic mesh method for pricing high-dimensional American options. Research paper. Graduate School of Business. Columbia Univ.

[10] BUCKLEW, J. (1990). *Large Deviation Techniques in Decision, Simulation, and Estimation.* John Wiley and Sons, New York.

[11] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press. Princeton.

[12] DAVID, F. and JOHNSON, N. (54). Statistical treatment of censored data. *Biometrika* **41** 228–240.

[13] DUFFIE, D. (1996). *Dynamic Asset Pricing Theory, 2nd ed.* Princeton Univ. Press. Princeton, New Jersey.

[14] GLYNN, P. (1989). Optimization of stochastic systems via simulation. In *Proceedings of the Winter Simulation Conference.* The Society for Computer Simulation. San Diego. California. pp. 90–105.

[15] GLYNN, P. and WHITT, W. (1992). The asymptotic efficiency of simulation estimators. *Oper. Res.* **40** 505–520.

[16] HARRISON, M. and KREPS, D. (1979). Martingales and arbitrage in multiperiod securities markets. *J. Econom. Theory* **20** 381–408.

[17] HOEFFDING, W. (1963). Probabilities inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc* **58** 13–30.

[18] HULL, J. (1993). *Options, Futures and Other Derivatives, 2nd ed.* Prentice-Hall. Englewood Cliffs, New Jersey.

[19] IGLEHART, D. (1976). Simulating stable stochastic systems, VI: quantile estimation. *J. ACM* **23** 347–360.

[20] Jorion, P. (1997). *Value at Risk, the new benchmark for controlling market risk.* Irwin professional publishing, Chicago.

[21] Kendall, M., Stuart, A. and Ord, J. (1987). *Kendall's Advanced Theory of Statistics, 5th ed.* vol. 1. Charles Griffin & Company Limited, London.

[22] Kolmogorov, A. (1933). *Grundbegriffe der Wahrsheinlichkeitsrechnung* vol. 2 of *Erg. Math.* Springer-Verlag, Berlin.

[23] Parthasarathy, K. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.

[24] Petrov, V. (1975). *Sums of Independent Random Variables.* Springer-Verlag, New York.

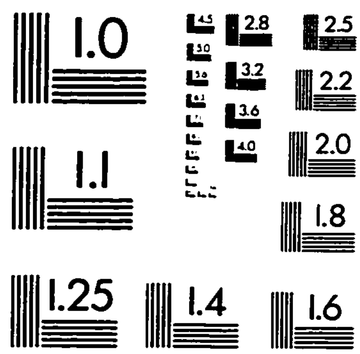
[25] Petrov, V. (1995). *Limit Theorems of Probability Theory, sequences of independent random variables.* Oxford Science Publications, Oxford.

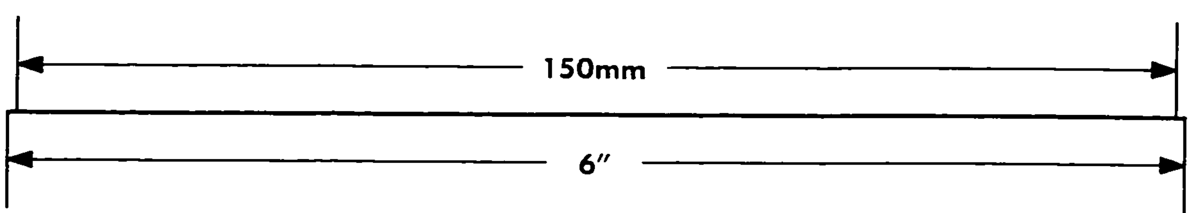
[26] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley and Sons, New York.

[27] Shoract, G. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.

[28] William, D. (1991). *Probability with Martingales.* Cambridge University Press, Cambridge.

IMAGE EVALUATION TEST TARGET (QA-3)





APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989