Asymptotic Representations for Importance-Sampling Estimators of Value-at-Risk and Conditional Value-at-Risk

Lihua Sun, L. Jeff Hong

Department of Industrial Engineering and Logistics Management
The Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, China

Abstract
Value-at-risk (VaR) and conditional value-at-risk (CVaR) are important risk measures. They are often estimated by using importance sampling (IS) techniques. In this paper, we derive the asymptotic representations for IS estimators of VaR and CVaR. Based on these representations, we are able to prove the consistency and asymptotic normality of the estimators and to provide simple conditions under which the IS estimators have smaller asymptotic variances than the ordinary Monte Carlo estimators.

Key words: Value-at-Risk, Conditional Value-at-Risk, Importance Sampling, Asymptotic Representation

1 Introduction
Value-at-risk (VaR) and conditional value-at-risk (CVaR) are two widely used risk measures. They play important roles in investment, risk management, and regulatory control of financial institutions. The Basel Accord II has incorporated the concept of $\alpha$–VaR, which is defined as the $\alpha$–quantile of a portfolio value $L$, and encourages banks to use VaR for daily risk management. The $\alpha$-CVaR, defined as the average of $\beta$–VaR of $L$ for $0 < \beta < \alpha$, has a long history of being used in insurance industry. It provides information on the potential large losses that an investor may suffer.

Risk managers may consider both of VaR and CVaR at the same time to obtain more information about portfolio risk. There are typically three approaches to estimating them: the variance-covariance approach, the historical simulation approach and the Monte Carlo simulation approach. Among the three, the Monte Carlo simulation approach is frequently used, because it is more general and can be applied to a wider range of risk models. However, the Monte Carlo simulation approach is often time-consuming. In risk management, $\alpha$ is typically close to 0. A large number of replications are needed to obtain accurate estimation of the tail behavior. Therefore, variance reduction techniques are often used to increase the efficiency of the estimation. Among these techniques, importance sampling (IS) is a natural choice, because it can allocate more samples to the tail of the distribution that is most relevant to the estimation of VaR and CVaR. In this paper, we study the asymptotic properties of the IS estimators of VaR and CVaR and discuss general conditions for IS to be effective.

Because the IS estimators of both VaR and CVaR are rather complicated compared to the typical sample means, we use the method of asymptotic representations to analyze their asymptotic properties. Bahadur [1] used this method to analyze the asymptotic properties of the ordinary estimator of VaR (quantile) by showing that the estimator can be approximated by a sample mean except for a high-order term. Then, the consistency and asymptotic normality of the estimator can be derived easily. In this paper, we derive the asymptotic representations for the IS estimators of both VaR and CVaR, and use them to prove the consistency and asymptotic normality of both estimators. To the best of our knowledge, we are the first to provide such clear representations.

From the asymptotic normality, we give simple conditions on the IS distributions under which the IS scheme is guaranteed to work asymptotically. A good feature of the conditions is that they are same for both VaR and CVaR. Therefore, one can estimate VaR and CVaR simultaneously using the same IS distribution.

*Corresponding author: Tel: +852 23587096, Email: hong@ust.hk. The authors would like to thank the associate editor and the referees for their helpful comments and suggestions. This research was partially supported by Hong Kong Research Grants Council grant CERG 613907.
This feature will greatly help risk managers who consider both risk measures and use them to complement each other.

The literature on applying IS to estimate VaR is growing rapidly. For example, Glynn [7] considered the use of IS for quantile (VaR) estimators; Glasserman et al. ([3], [4]) used IS to estimate the VaR of a portfolio loss for both light-tail and heavy-tail situations; Glasserman and Li [6] applied IS to estimate the VaR of portfolio credit risk; Glasserman and Juneja [5] used IS to estimate the VaR of a sum of i.i.d random variables. To the best of our knowledge, however, there is no published work on using IS to estimate CVaR.

The rest of the paper is organized as follows. Section 2 reviews the IS estimators of VaR and CVaR. In section 3, we develop asymptotic representations for the IS estimators. From these representations, we can easily prove the consistency and asymptotic normality of the IS estimators. The conclusions are given in section 4. Some lengthy proofs are given in the Appendix.

2 Importance Sampling for VaR and CVaR

Let $L$ be a real-valued random variable with a cumulative distribution function (c.d.f.) $F(\cdot)$, and let $v$ and $c$ denote the $\alpha$-VaR and $\alpha$-CVaR of $L$, respectively, for $0 < \alpha < 1$. Then,

$$v = F^{-1}(\alpha) = \inf\{x: F(x) \geq \alpha\} \quad \text{and} \quad c = v - \frac{1}{\alpha} E[v - L]^+, $$

where $x^+ = \max\{x, 0\}$. Note that $v$ is also the $\alpha$-quantile of $L$, and $c = E[L|L \leq v]$ if $L$ has a positive density at $v$ [8]. Under the definitions, we are interested in the left tail of the distribution of $L$. Therefore, $\alpha$ is often close to 0. Sometimes, $v$ and $c$ are defined for the right tail of $L$ (e.g., [3], [4]). We may convert the right tail to the left tail by adding a negative sign to the random variable.

Ordinary Monte Carlo estimation of $v$ and $c$ involves generating $n$ independent and identically distributed (i.i.d.) random observations of $L$, denoted as $L_1, \ldots, L_n$, and estimating them by

$$\hat{v}_n = \hat{F}_n^{-1}(\alpha) = \inf\{x: \hat{F}_n(x) \geq \alpha\}, \quad (1)$$

$$\hat{c}_n = \hat{v}_n - \frac{1}{n} \sum_{i=1}^{n} (\hat{v}_n - L_i)^+, \quad (2)$$

respectively, where

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{L_i \leq x\} \quad (3)$$

is the empirical distribution of $L$ constructed from $L_1, \ldots, L_n$ and $I\{\cdot\}$ is the indicator function. Note that $\hat{F}_n(x)$ is an unbiased and consistent estimator of $F(x)$. Serfling [9] and Trindade et al. [12] showed that $\hat{v}_n$ and $\hat{c}_n$ are consistent estimators of $v$ and $c$, respectively, as $n \to \infty$.

Now we introduce the IS estimators of $v$ and $c$. Suppose we choose an IS distribution function $G$ for which the probability measure associated with $G$ is absolutely continuous with respect to that associated with $F$, i.e., $F(dx) = 0$ if $G(dx) = 0$ for any $x \in \mathbb{R}$. Let $L(x) = \frac{F(dx)}{G(dx)}$, then $L$ is called the likelihood ratio (LR) function. Then for any $x \in \mathbb{R}$, we may estimate $F(x)$ by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{L_i \leq x\} L(L_i). \quad (4)$$

It is easy to see that $F_n(x)$ is also an unbiased and consistent estimator of $F(x)$ as $\hat{F}_n(x)$ of Equation (3). Let $v_n$ and $c_n$ denote the IS estimators of $v$ and $c$. Similar to Equations (1) and (2), we define

$$v_n = F_n^{-1}(\alpha) = \inf\{x: F_n(x) \geq \alpha\},$$

$$c_n = v_n - \frac{1}{n} \sum_{i=1}^{n} (v_n - L_i)^+ L(L_i).$$

To analyze the asymptotic properties of $v_n$ and $c_n$, we make the following assumptions.

**Assumption 1.** There exists an $\epsilon > 0$ such that $L$ has a positive and differentiable density $f(x)$ for any $x \in (v - \epsilon, v + \epsilon)$. 

2
Assumption 1 requires that \( L \) has a positive and differentiable density in a neighborhood of \( v \). It implies that \( F(v) = \alpha \) and \( c = \text{E}[L|L \leq v] \).

**Assumption 2.** There exist \( \varepsilon > 0 \) and \( C > 0 \) such that \( \mathcal{L}(x) < C \) for any \( x \in (v - \varepsilon, v + \varepsilon) \) and there exists \( p > 2 \) such that \( \text{E}_G[I(L \leq v + \varepsilon)\mathcal{L}(L)] < \infty \), where \( \text{E}_G \) denotes the expectation under the IS measure.

Assumption 2 requires the LR function is bounded above in a neighborhood of \( v \) and it has a finite \( p > 2 \) moment on the left tail of \( L \). An effective IS typically allocates more samples in the set \( \{L < v + \varepsilon\} \) which are most useful in estimating \( v \) and \( c \) (See [3], [4], [5] and [6]). Then, \( f(x) < g(x) \) for \( x < v + \varepsilon \). Therefore, \( \mathcal{L}(x) < 1 \) for \( x < v + \varepsilon \). Thus, Assumption 2 is satisfied for these LR. From Assumption 2 and the positivity of \( \mathcal{L}(x) \), we know that for any \( \varepsilon' \leq \varepsilon \), \( \text{E}_G[I\{L \leq v + \varepsilon'\}|\mathcal{L}(L)] < \infty \). Therefore, \( \text{Var}[I\{L \leq v + \varepsilon'\}|\mathcal{L}(L)] < \infty \) for any \( \varepsilon' \leq \varepsilon \).

### 3 Asymptotic Representations of the IS Estimators

A complicated estimator can often be represented as the sum of several terms whose asymptotic behaviors are clear. This representation is called an asymptotic representation of the estimator. Based on the asymptotic representation of an estimator, many asymptotic properties of the estimator, e.g., consistency and asymptotic normality, can be analyzed easily. A famous example is the asymptotic representation of the VaR estimator \( \hat{v}_n \) (also known as Bahadur representation of the quantile estimator). By Bahadur [1], under Assumption 1,

\[
\hat{v}_n = v + \frac{1}{f(v)} \left( \alpha - \frac{1}{n} \sum_{i=1}^{n} I\{L_i \leq v\} \right) + R_n,
\]

where \( R_n = O_{a.s.} \left( n^{-3/4} \log n \right)^{3/4} \). The statement \( Y_n = O_{a.s.} \left( g(n) \right) \) means that \( Y_n / g(n) \) is bounded by a constant almost surely. Given the representation, many asymptotic properties of \( \hat{v}_n \) can be analyzed easily. For instance, we may use it to prove the strong consistency and asymptotic normality of \( \hat{v}_n \). By the strong law of large numbers [2], \( \frac{1}{n} \sum_{i=1}^{n} I\{L_i \leq v\} \rightarrow F(v) \) w.p.1 as \( n \rightarrow \infty \). Furthermore, because \( F(v) = \alpha \) by Assumption 1, it is clear that \( \hat{v}_n \rightarrow v \) w.p.1 as \( n \rightarrow \infty \). Thus, \( \hat{v}_n \) is a strongly consistent estimator of \( v \). Similarly, by the central limit theorem ([2], p. 64)

\[
\sqrt{n} \left( \alpha - \frac{1}{n} \sum_{i=1}^{n} I\{L_i \leq v\} \right) \Rightarrow \sqrt{\alpha(1-\alpha)} N(0,1) \quad \text{as} \quad n \rightarrow \infty,
\]

where “\( \Rightarrow \)” denotes “converge in distribution” and \( N(0,1) \) denotes a standard normal random variable. Then,

\[
\sqrt{n}(\hat{v}_n - v) \Rightarrow \frac{\sqrt{\alpha(1-\alpha)}}{f(v)} N(0,1) \quad \text{as} \quad n \rightarrow \infty.
\]

Therefore, \( \hat{v}_n \) is asymptotically normally distributed.

In the rest of this section, we develop asymptotic representations of the IS estimators \( v_n \) and \( c_n \), and use them to analyze the consistency and asymptotic normality of \( v_n \) and \( c_n \).

#### 3.1 Asymptotic Representation of \( v_n \)

We first consider the asymptotic representation of \( v_n \). Note that by Taylor expansion ([10], p.110), \( F(v_n) - F(v) = f(v)(v_n - v) - A_{1,n} \), where \( A_{1,n} \) is the remainder term. Then, we have

\[
v_n = v + \frac{F(v_n) - F(v)}{f(v)} + \frac{1}{f(v)} A_{1,n}.
\]

Let \( A_{2,n} = F(v_n) + F_n(v) - F_n(v_n) - F(v) \) and \( A_{3,n} = F_n(v_n) - F(v) \). It is easy to see that

\[ F(v_n) - F(v) = F(v) - F_n(v) + A_{2,n} + A_{3,n}. \]

Therefore, by Equation (7), we have

\[
v_n = v + \frac{F(v) - F_n(v)}{f(v)} + \frac{A_{1,n} + A_{2,n} + A_{3,n}}{f(v)}.
\]

3
In the following lemma, we provide the orders of $A_{1,n}$, $A_{2,n}$ and $A_{3,n}$. The proof of the lemma is included in the Appendix. In the lemma, we use the statement $U_n = o_p(g(n))$ which means $U_n/g(n) \to 0$ in probability as $n \to \infty$.

**Lemma 1.** For a fixed $\alpha \in (0,1)$, suppose that Assumptions 1 and 2 are satisfied. Then, $A_{1,n} = O_{a.s.}((n^{-1+2/p+\delta})$, $A_{2,n} = O_{a.s.}(n^{3/4+1/(2p)+\delta})$, $A_{3,n} = O_{a.s.}(n^{-1})$ for any $\delta > 0$, and $A_{1,n} = o_p(n^{-1/2})$, $A_{2,n} = o_p(n^{-1/2})$, $A_{3,n} = o_p(n^{-1/2})$. Furthermore, if $L(x) < C$ for any $x \in (-\infty, \nu + \varepsilon)$, then $A_{1,n} = O_{a.s.}(n^{-1} \log n)$, $A_{2,n} = O_{a.s.}(n^{-3/4}(\log n)^{3/4})$ and $A_{3,n} = O_{a.s.}(n^{-1})$.

Let $L_i$ denote $L(L_i)$ for all $i = 1, \ldots, n$. By Lemma 1, we can prove the following theorem on the asymptotic representation of $v_n$.

**Theorem 1.** For a fixed $\alpha \in (0,1)$, suppose that Assumptions 1 and 2 are satisfied. Then,

$$v_n = v + \frac{1}{f(v)} \left( \alpha - \frac{1}{n} \sum_{i=1}^{n} I\{L_i \leq v\} \right) L_i + A_n,$$

where $A_n = o_p(n^{-1/2})$ and $A_n = O_{a.s.}(t(n, \delta))$ with $t(n, \delta) = \max\{n^{-1+2/p+\delta}, n^{-3/4+1/(2p)+\delta}\}$ for any $\delta > 0$. Furthermore, if $L(x) < C$ for any $x \in (-\infty, \nu + \varepsilon)$, then $A_n = O_{a.s.}(n^{-3/4}(\log n)^{3/4})$.

**Proof.** By Assumption 1, $F(v) = \alpha$. Then, by Equation (4), $F(v) - F_n(v) = \alpha - \frac{1}{n} \sum_{i=1}^{n} I\{L_i \leq v\}$. Let $A_n = \frac{1}{f(v)}(A_{1,n} + A_{2,n} + A_{3,n})$. Since $f(v) > 0$ by Assumption 1, the conclusions of the theorem follow directly from Equation (8).

Because $\tilde{v}_n$ is a special case of $v_n$ where $L(x) = 1$ for all $x \in \mathbb{R}$, Bahadur representation of Equation (5) may be viewed as a special case of Theorem 1.

Let $\text{Var}_G$ denote the expectation and variance under the IS measure. From Theorem 1, it is also straightforward to prove the following corollary on the strong consistency and asymptotic normality of $v_n$.

**Corollary 1.** For a fixed $\alpha \in (0,1)$, suppose that Assumptions 1 and 2 are satisfied. Then, $v_n \to v$ w.p.1 and

$$\sqrt{n}(v_n - v) \Rightarrow \frac{\sqrt{\text{Var}_G[I\{L \leq v\}L(L)]}}{f(v)} N(0,1) \text{ as } n \to \infty.$$  

**Remark 1.** The conclusions of Corollary 1 have also been proved by Glynn [7] under Assumption 1 and the assumption that $E_G[L^3(L)] < \infty$.

Note that

$$\text{Var}_G[I\{L \leq v\}L(L)] = E_G[I\{L \leq v\}L^2(L)] - E_G^2[I\{L \leq v\}L(L)] = E[I\{L \leq v\}L(L)] - \alpha^2.$$  

(9)

If the IS distribution allocates more samples to the left tail of the distribution of $L$, e.g., $L(x) < 1$ for all $x \leq v$, then by Equation (9), $\text{Var}_G[I\{L \leq v\}L(L)] < \alpha(1 - \alpha)$. Compared to Equation (6), the IS estimator $v_n$ has a smaller asymptotic variance than the ordinary estimator $\tilde{v}_n$. This explains why IS can improve the efficiency of VaR estimation when the IS distribution is selected appropriately.

### 3.2 Asymptotic Representation of $c_n$

We now consider the asymptotic representation of $c_n$. Note that

$$c_n = v_n - \frac{1}{n\alpha} \sum_{i=1}^{n} (v_n - L_i)^+ L_i = v - \frac{1}{n\alpha} \sum_{i=1}^{n} (v - L_i)^+ L_i + (v_n - v) - \frac{1}{n\alpha} \sum_{i=1}^{n} [(v_n - L_i)^+ - (v - L_i)^+] L_i.$$  

Furthermore, note that

$$\frac{1}{n\alpha} \sum_{i=1}^{n} [(v_n - L_i)^+ - (v - L_i)^+] L_i = \frac{1}{n\alpha} \sum_{i=1}^{n} [(v_n - L_i) I\{L_i \leq v_n\} - (v - L_i) I\{L_i \leq v\}] L_i$$

$$= \frac{1}{n\alpha} \sum_{i=1}^{n} [(v_n - v) I\{L_i \leq v_n\}] L_i + \frac{1}{n\alpha} \sum_{i=1}^{n} (v - L_i) [I\{L_i \leq v_n\} - I\{L_i \leq v\}] L_i.$$  

Note that
Therefore,
\[ c_n = v - \frac{1}{n\alpha} \sum_{i=1}^{n} (v - L_i)^+ \mathcal{L}_i + \frac{1}{\alpha} (v_n - v) (\alpha - F_n(v_n)) - \frac{1}{n\alpha} \sum_{i=1}^{n} (v - L_i) [I\{L_i \leq v_n\} - I\{L_i \leq v\}] \mathcal{L}_i. \] (10)

Since
\[ \left| \frac{1}{n\alpha} \sum_{i=1}^{n} (v - L_i) [I\{L_i \leq v_n\} - I\{L_i \leq v\}] \mathcal{L}_i \right| \leq \frac{1}{\alpha} |v_n - v| |F_n(v_n) - F_n(v)|, \] (11)
then by Equations (10) and (11),
\[ c_n = v - \frac{1}{n\alpha} \sum_{i=1}^{n} (v - L_i)^+ \mathcal{L}_i + B_n, \] (12)
where
\[ |B_n| \leq \frac{1}{\alpha} |v_n - v| (2 |F_n(v_n) - F(v)| + |F_n(v) - F(v)|). \]

In the following lemma, we prove the order of \( B_n \). The proof of the lemma is included in the Appendix.

**Lemma 2.** For a fixed \( \alpha \in (0, 1) \), suppose that Assumptions 1 and 2 are satisfied. Then, \( B_n = O_{\text{a.s.}} \left( n^{-1+\epsilon} \right) \) and \( B_n = o_p(n^{-1/2}) \). Furthermore, if \( \mathcal{L}(x) < C \) for any \( x \in (-\infty, v + \epsilon) \), then \( B_n = O_{\text{a.s.}}(n^{-1/2}) \).

Then, we have the following theorem on the asymptotic representation of \( c_n \). Note that the conclusion of the theorem follows directly from Equation (12) and Lemma 2. Therefore, we omit the proof.

**Theorem 2.** For a fixed \( \alpha \in (0, 1) \), suppose that Assumptions 1 and 2 are satisfied. Then,
\[ c_n = c + \left( \frac{1}{n} \sum_{i=1}^{n} \left[ v - \frac{1}{\alpha} (v - L_i)^+ \mathcal{L}_i \right] - c \right) + B_n, \]
where \( B_n = O_{\text{a.s.}} \left( n^{-1+\epsilon/2} \right) \) for any \( \delta > 0 \) and \( B_n = o_p(n^{-1/2}) \). Furthermore, if \( \mathcal{L}(x) < C \) for any \( x \in (-\infty, v + \epsilon) \), then \( B_n = O_{\text{a.s.}}(n^{-1/2}) \).

Note that \( \mathbb{E}_{\tilde{G}} \left[ v - \frac{1}{\alpha} (v - L)^+ \mathcal{L}(L) \right] = c \). Then, by the strong law of large numbers and the central limit theorem, it is straightforward to prove the following corollary on the strong consistency and asymptotic normality of \( c_n \).

**Corollary 2.** For a fixed \( \alpha \in (0, 1) \), suppose that Assumptions 1 and 2 are satisfied and \( \mathbb{E}_{\tilde{G}} \left[ (v-L)^2 \mathcal{L}(L)^2 \right] I\{L < v\} \right) < \infty \). Then, \( c_n \to c \) w.p.1 and
\[ \sqrt{n}(c_n - c) \Rightarrow \frac{\text{Var}_{\tilde{G}} \left[ (v-L)^+ \mathcal{L}(L) \right]}{\alpha} \mathcal{N}(0, 1) \quad \text{as } n \to \infty. \]

Furthermore, we may set \( \mathcal{L}(x) = 1 \) for all \( x \in \mathbb{R} \). Then, the conclusions of Theorem 2 and Corollary 2 apply to \( \tilde{c}_n \), the ordinary Monte Carlo estimator of \( c \). We then have the following corollary.

**Corollary 3.** For a fixed \( \alpha \in (0, 1) \), suppose that Assumption 1 is satisfied. Then,
\[ \tilde{c}_n = c + \left( \frac{1}{n} \sum_{i=1}^{n} \left[ v - \frac{1}{\alpha} (v - L_i)^+ \mathcal{L}_i \right] - c \right) + C_n \]
where \( C_n = O_{\text{a.s.}} \left( n^{-1} \log n \right) \), \( \tilde{c}_n \to c \) w.p.1. Furthermore, if \( \mathbb{E} \left[ (v-L)^2 I\{L < v\} \right] < \infty \), then
\[ \sqrt{n}(\tilde{c}_n - c) \Rightarrow \frac{\text{Var}_{\tilde{G}} [(v-L)^+] \mathcal{L}(L)}{\alpha} \mathcal{N}(0, 1) \quad \text{as } n \to \infty. \]

**Remark 2.** The strong consistency and asymptotic normality of \( \tilde{c}_n \) have also been studied by a number of papers in the literature, including Trindade et al. [12] and Hong and Liu [8], using different methods.

If the IS distribution allocates more samples to the left tail of the distribution of \( L \), e.g., \( \mathcal{L}(x) < 1 \) for all \( x \leq v \), then it is easy to show that \( \text{Var}_{\tilde{G}} [(v-L)^+ \mathcal{L}(L)] < \text{Var} [(v-L)^+] \). Therefore, by Corollaries 2 and 3, the IS estimator \( c_n \) has a smaller asymptotic variance than the ordinary estimator \( \tilde{c}_n \). The likelihood functions in [3], [4], [5], [6] all satisfy the condition \( \mathcal{L}(x) < 1 \) for all \( x < v \). Thus, they can reduce the variance of VaR and CVaR estimators at the same time. This explains why IS can improve the efficiency of VaR and CVaR estimation simultaneously when the IS distribution is selected appropriately.
4 Conclusions

In this paper, we develop asymptotic representations for the IS estimators of both VaR and CVaR, and derive the consistency and asymptotic normality for both estimators. We show that there are simple conditions for choosing IS distributions that guarantee to reduce the asymptotic variances of both estimators. Further research may focus on how to find efficient IS schemes for estimating VaR and CVaR.

A Appendix

A.1 Proof of Lemma 1

A similar result has been proved by Serfling [9] for \( \tilde{v}_n \). In this section, we mainly follow his steps. However, we need to handle the likelihood-ratio term that does not appear in \( \tilde{v}_n \).

A.1.1 Three Propositions

Proposition 1. For a fixed \( \alpha \in (0,1) \), suppose that Assumption 2 is satisfied. Then for any \( \gamma < \varepsilon \),

\[
\Pr \{ |v_n - v| > \gamma \} \leq C_p n^{-p/2} \delta_{n}^{-p}
\]

for sufficiently large \( n \), where \( \delta_{n} = \min \{ F(v + \gamma) - \alpha, \alpha - F(v - \gamma) \} \), and \( C_p \) is a constant related to \( p \). Moreover, if \( \mathcal{L}(L) < C \) for any \( L < v + \varepsilon \), then

\[
\Pr \{ |v_n - v| > \gamma \} \leq 2\varepsilon^{-2n\delta_{n}^{2}/(C+1)^{2}}.
\]

Proof. Note that

\[
\Pr \{ |v_n - v| > \gamma \} = \Pr \{ v_n > v + \gamma \} + \Pr \{ v_n < v - \gamma \}.
\]

Because \( v_n = F_n^{-1}(\alpha) = \inf \{ x : F_n(x) \geq \alpha \} \), \( v_n > v + \gamma \) and \( v_n < v - \gamma \) are equivalent to \( F_n(v + \gamma) < \alpha \) and \( F_n(v - \gamma) \geq \alpha \), respectively. We have

\[
\Pr \{ |v_n - v| > \gamma \} \leq \Pr \{ F_n(v + \gamma) < \alpha \} + \Pr \{ F_n(v - \gamma) \geq \alpha \}\]

\[
\leq \Pr \{ F(v + \gamma) - F_n(v + \gamma) > F(v + \gamma) - \alpha \} + \Pr \{ F_n(v - \gamma) - F(v - \gamma) \geq \alpha - F(v - \gamma) \}. \tag{13}
\]

Moreover,

\[
\Pr \{ F(v + \gamma) - F_n(v + \gamma) > F(v + \gamma) - \alpha \} = \Pr \left\{ \sum_{i=1}^{n} [F(v + \gamma) - I(L_i \leq v + \gamma)\mathcal{L}_i] > n(F(v + \gamma) - \alpha) \right\},
\]

and

\[
\Pr \{ F_n(v - \gamma) - F(v - \gamma) \geq \alpha - F(v - \gamma) \} = \Pr \left\{ \sum_{i=1}^{n} [I(L_i \leq v - \gamma)\mathcal{L}_i - F(v - \gamma)] \geq n(\alpha - F(v - \gamma)) \right\}.
\]

Note that, when Assumption 2 is satisfied, combining with Markov’s inequality ([2], P.14), we have

\[
\Pr \left\{ \sum_{i=1}^{n} F(v + \gamma) - I(L_i \leq v + \gamma)\mathcal{L}_i > n(F(v + \gamma) - \alpha) \right\} \leq \frac{\mathbb{E} \left[ \left| \sum_{i=1}^{n} F(v + \gamma) - I(L_i \leq v + \gamma)\mathcal{L}_i \right|^{p} \right]}{n^{p} (F(v + \gamma) - \alpha)^{p}}.
\]

By Rosenthal’s inequality [11], we know

\[
\mathbb{E} \left[ \left| \sum_{i=1}^{n} F(v + \gamma) - I(L_i \leq v + \gamma)\mathcal{L}_i \right|^{p} \right] \leq \tilde{C}_p \max \left\{ \sum_{i=1}^{n} \mathbb{E} |F(v + \gamma) - I(L_i \leq v + \gamma)\mathcal{L}_i|^{p}, \left( \sum_{i=1}^{n} \mathbb{E} |F(v + \gamma) - I(L_i \leq v + \gamma)\mathcal{L}_i|^2 \right)^{p/2} \right\},
\]

in particular,
where $\hat{C}_p = 2 \max \left\{ p^p, p^{p/2+1}e^p \int_0^\infty x^{p/2-1}(1-x)^{-p}dx \right\}$. When $n$ is sufficiently large,

$$
\sum_{i=1}^n E[F(v+\gamma) - I\{L_i \leq v + \gamma\} L_i]^p \leq \left( \sum_{i=1}^n E[F(v+\gamma) - I\{L_i \leq v + \gamma\} L_i]^2 \right)^{p/2}.
$$

Therefore,

$$
\Pr \left\{ \sum_{i=1}^n [F(v+\gamma) - I\{L_i \leq v + \gamma\} L_i] > n(F(v+\gamma) - \alpha) \right\} \leq \hat{C}_p \frac{\left( E[F(v+\gamma) - I\{L_1 \leq v + \gamma\} L_1]^2 \right)^{p/2}}{n^{p/2}(F(v+\gamma) - \alpha)^p}.
$$

Similarly, we can get

$$
\Pr \left\{ \sum_{i=1}^n [I\{L_i \leq v - \gamma\} L_i - F(v - \gamma)] \geq n(\alpha - F(v - \gamma)) \right\} \leq \hat{C}_p \frac{\left( E[I\{L_1 \leq v - \gamma\} L_1 - F(v - \gamma)]^2 \right)^{p/2}}{n^{p/2}(\alpha - F(v - \gamma))^p}.
$$

Let

$$
C_p = 2\hat{C}_p \max \left\{ \left( E[F(v+\gamma) - I\{L_1 \leq v + \gamma\} L_1]^2 \right)^{p/2}, \left( E[I\{L_1 \leq v - \gamma\} L_1 - F(v - \gamma)]^2 \right)^{p/2} \right\}.
$$

Combing with Equation (13), we have, for sufficiently large $n$,

$$
\Pr \{|v_n - v| > \gamma\} \leq C_p n^{-p/2} \delta^{-p}.
$$

If we further have $\mathcal{L}(L) < C$ for any $L \in (-\infty, v+\epsilon)$, both $F(v+\gamma) - I\{L_i \leq v + \gamma\} L_i$ and $I\{L_i \leq v - \gamma\} L_i - F(v - \gamma)$ are bounded. Thus, we can apply Hoeffding’s Inequality ([9], P.75) to have

$$
\Pr \left\{ \sum_{i=1}^n [F(v+\gamma) - I\{L_i \leq v + \gamma\} L_i] > n(F(v+\gamma) - \alpha) \right\} \leq \exp \left\{ -\frac{2n(F(v+\gamma) - \alpha)^2}{(C + 1)^2} \right\}
$$

$$
\Pr \left\{ \sum_{i=1}^n [I\{L_i \leq v - \gamma\} L_i - F(v - \gamma)] \geq n(\alpha - F(v - \gamma)) \right\} \leq \exp \left\{ -\frac{2n(\alpha - F(v - \gamma))^2}{(C + 1)^2} \right\}.
$$

Then, $\Pr\{|v_n - v| > \gamma\} \leq 2e^{-2n\delta^2/(C+1)^2}$. This completes the proof of the proposition.

**Proposition 2.** Let $\epsilon_{n,\delta} = \frac{2}{f(\epsilon)} n^{-1/2} e^{1/p + \delta}$ with $\delta > 0$. For a fixed $\alpha \in (0, 1)$, suppose Assumptions 1 and 2 are satisfied. Then, $|v_n - v| = O_{a.s.}(\epsilon_{n,\delta})$ for any $\delta > 0$ and $|v_n - v| = o_p \left( n^{-1/2} g(n) \right)$ for any function $g(n) \to \infty$ as $n \to \infty$.

Moreover, if $\mathcal{L}(L) < C$ for any $L \in (-\infty, v+\epsilon)$, then $|v_n - v| = O_{a.s.}(\epsilon_n)$ with $\epsilon_n = \frac{2C}{f(\epsilon)} n^{-1/2} (\log n)^{1/2}$.

**Proof.** By Assumption 1, $v$ is a unique solution to $F(x) = \alpha$. Note that

$$
F(v + \epsilon_{n,\delta}) - \alpha = F(v + \epsilon_{n,\delta} - F(v) = f(v) \epsilon_{n,\delta} + o(\epsilon_{n,\delta}).
$$

When $n$ is sufficiently large and $\delta$ is sufficiently small, $\epsilon_{n,\delta} < \epsilon$, where $\epsilon$ is defined in Assumption 2. Then, $F(v + \epsilon_{n,\delta}) - \alpha \geq f(v) \epsilon_{n,\delta}/2$. Similarly, we can also prove that $\alpha - F(v - \epsilon_{n,\delta}) = \frac{f(v)}{2} \epsilon_{n,\delta} \geq \frac{f(v)}{2} \epsilon_{n,\delta}$.

Hence, by Proposition 1, for $n$ sufficiently large, $\Pr\{|v_n - v| > \epsilon_{n,\delta}\} \leq C_p n^{-1-p\delta}$, where $C_p$ is defined in Proposition 1. Because $\sum_{n=1}^\infty C_p n^{-1-p\delta} < \infty$ for any $\delta > 0$, by Borel-Cantelli Lemma ([2], P.46), we have $|v_n - v| = O_{a.s.}(\epsilon_{n,\delta})$. Similarly, for any $\alpha > 0$, we have

$$
\Pr \left\{ \frac{|v_n - v|}{n^{-1/2} g(n)} > \alpha \right\} = \Pr \left\{ |v_n - v| > \alpha n^{-1/2} g(n) \right\} \leq C_p (\alpha g(n) )^{-p}.
$$

Then, $|v_n - v| = o_p \left( n^{-1/2} g(n) \right)$.

Furthermore, if $\mathcal{L}(L) < C$ for any $L \in (-\infty, v+\epsilon)$, by the same approach, we can prove $|v_n - v| = O_{a.s.}(\epsilon_n)$ with $\epsilon_n = \frac{2C}{f(\epsilon)} n^{-1/2} (\log n)^{1/2}$.

\[7\]
Proposition 3. Let \( T_n = \sup_{|x| \leq \epsilon_n} |F_n(v + x) - F_n(v) - F(v + x) + F(v)|, \) where \( \epsilon_{n, \delta} \) is defined in Proposition 2. Suppose Assumptions 1 and 2 are satisfied. Then, \( T_n = O_{n.a.s.} \left( n^{-3/4 + 1/(2p) + \delta} \right) \) for any \( \delta > 0. \)

Furthermore, let \( K_n = \sup_{|x| \leq \epsilon_n} |F_n(v + x) - F_n(v) - F(v + x) + F(v)|, \) where \( \epsilon_n \) is defined in Proposition 2. If \( \mathcal{L}(x) < C \) for any \( x \in (-\infty, v + \varepsilon) \), then \( K_n = O_{n.a.s.} \left( n^{-3/4} (\log n)^{3/4} \right) \).

Proof. We have \( \epsilon_{n, \delta} < \min\{\varepsilon, \varepsilon\} \) for \( n \) sufficiently large and \( \delta \) sufficiently small, where \( \varepsilon \) and \( \varepsilon \) are defined in Assumptions 1, 2, respectively, and \( \epsilon_{n, \delta} \) is defined in Proposition 2. Let \( b_n = \left\lceil \frac{2}{f(v)} n^{1/4 + 1/p + \delta} \right\rceil \). For any integer \( l \in [-b_n, b_n] \), we let \( \xi_{l,n} = v + \ell \epsilon_{n, \delta}/b_n \). For any \( |x| < \epsilon_{n, \delta} \), we can find \( l \) such that \( v + x \in [\xi_{l,n}, \xi_{l+1,n}) \). Then,

\[
F_n(\xi_{l,n}) - F(\xi_{l+1,n}) \leq F_n(v + x) - F(v + x) \leq F_n(\xi_{l+1,n}) - F(\xi_{l,n}),
\]

which is equivalent to

\[
F_n(\xi_{l,n}) - F_n(v) - F(v) + F(\xi_{l,n}) - F(\xi_{l+1,n}) \\
\leq F_n(v + x) - F_n(v) - F(v + x) + F(v) \\
\leq F_n(\xi_{l+1,n}) - F_n(v) - F(\xi_{l+1,n}) + F(v) + F(\xi_{l+1,n}) - F(\xi_{l,n}).
\]

Then, we have

\[
T_n \leq \sup_{l \in [-b_n, b_n]} \left| F_n(\xi_{l,n}) - F_n(v) - F(\xi_{l,n}) + F(v) \right| + \sup_{l \in [-b_n, b_n]} \left| F(\xi_{l+1,n}) - F(\xi_{l,n}) \right|.
\]

Note that \( |F(\xi_{l+1,n}) - F(\xi_{l,n})| = f(z) \epsilon_{n, \delta}/b_n \) for some \( z \in (\xi_{l,n}, \xi_{l+1,n}) \). Let \( \tilde{f} = \sup_{|x| \leq \epsilon} f(v + x) \). Then,

\[
\sup_{l \in [-b_n, b_n]} \left| F(\xi_{l+1,n}) - F(\xi_{l,n}) \right| \leq \tilde{f} n^{-3/4}.
\]

From Assumption 2, we know \( \mathcal{L}(L) < C \) for any \( L \in (v - \epsilon_{n, \delta}, v + \epsilon_{n, \delta}) \). Let \( \xi_{n, \delta} = n^{-3/4 + 1/(2p) + \delta}/(\log n)^{1/2} \). By Bernstein’s Inequality ([9], P. 95), for any \( c_1 > 0 \)

\[
\Pr \left\{ |F_n(\xi_{l,n}) - F_n(v) - F(\xi_{l,n}) + F(v)| > c_1 \xi_{n, \delta} \right\} \leq 2 \exp \left\{ -\frac{3c_1^2 n \xi_{n, \delta}^2}{6\sigma^2 + 2Cc_1 \xi_{n, \delta}} \right\}, \quad (15)
\]

where \( \sigma^2 = \text{Var}[I, \xi_{l,n}] \mathcal{L}_1 \leq E[I, \xi_{l,n}] \mathcal{L}_2 \leq E[I, \xi_{l,n}] \mathcal{L}_3 \leq C^2 \tilde{f} \epsilon_{n, \delta} \).

Inputting the upper bound of \( \sigma^2 \) into Equation (15), we have

\[
\Pr \left\{ |F_n(\xi_{l,n}) - F_n(v) - F(\xi_{l,n}) + F(v)| > c_1 \xi_{n, \delta} \right\} \leq 2 \exp \left\{ -\frac{3c_1^2 n \xi_{n, \delta}^2}{6C^2 \tilde{f}^2 n^{-1/2} \frac{1}{\log n} + 2Cc_1 n^{-2} \frac{1}{\log n}} \right\}
\]

\[
\leq 2 \exp \left\{ -\frac{3c_1^2 \log n}{6C^2 \tilde{f}^2 n^{-1/2} + 2Cc_1 n^{-2} \frac{1}{\log n}} \right\}.
\]

When \( n \) is sufficiently large, we can choose \( c_1 \) big enough such that \( \frac{3c_1^2}{6C^2 \cdot \tilde{f}^2 n^{-1/2} + 2Cc_1 n^{-2} \frac{1}{\log n}} > 2 \).

Then, \( \Pr \left\{ |F_n(\xi_{l,n}) - F_n(v) - F(\xi_{l,n}) + F(v)| > c_1 \xi_{n, \delta} \right\} \leq 2n^{-2} \). Therefore, when \( \delta < 1/2 \),

\[
\Pr \left\{ \sup_{l \in [-b_n, b_n]} |F_n(\xi_{l,n}) - F_n(v) - F(\xi_{l,n}) + F(v)| > c_1 \xi_{n, \delta} \right\} \leq 2 \left( \frac{2}{\tilde{f}(v)} n^{1/4 + 1/p + \delta} \right) n^{-2}
\]

8
and
\[
\sum_{n=1}^{\infty} \Pr \left\{ \sup_{t \in [-b_n, b_n]} |F_n(\xi_{t,n}) - F_n(v) - F(\xi_{t,n}) + F(v)| > c_n \right\} \leq \sum_{n=1}^{\infty} 2 \left[ \frac{2}{f(v)} n^{1/4 + 1/p + \delta} \right] n^{-2} < \infty. \tag{16}
\]

Note that
\[
\sum_{n=1}^{\infty} \Pr \{ T_n > c_n \} \leq \sum_{n=1}^{\infty} \Pr \left\{ \sup_{t \in [-b_n, b_n]} |F_n(\xi_{t,n}) - F_n(v) - F(\xi_{t,n}) + F(v)| > c_n \right\} + \sum_{n=1}^{\infty} \Pr \left\{ \sup_{t \in [-b_n, b_n]} |F(\xi_{t,n+1}) - F(\xi_{t,n})| > c_n \right\}.
\]
The event \( \sup_{t \in [-b_n, b_n]} |F(\xi_{t+1,n}) - F(\xi_{t,n})| \) is actually a deterministic event. From Equation (14), we know for \( n \) big enough this event will never happen and thus the probability is 0. Combining with Equations (16) and Borel-Cantelli Lemma, we have \( T_n = O_{a.s.}(\xi_{n,\delta}) = O_{a.s.} \left( n^{-\frac{2}{p+\delta} + \frac{1}{p} + \delta} (\log n)^{1/2} \right) \). Note that \( \log n = O(n^\delta) \), then we have \( T_n = O_{a.s.} \left( n^{-\frac{2}{p+\delta} + \frac{1}{p} + \delta} \right) \).

Furthermore, if \( \mathcal{L}(x) < C \), we can similarly prove \( K_n = O_{a.s.}(n^{-3/4}(\log n)^{3/4}) \).

\[\square\]

### A.1.2 Proof of Lemma 1

**Proof.** First, we prove, \( A_{1,n} = O_{a.s.}(n^{-1+2/p+\delta}) \) and \( A_{1,n} = o_p(n^{-1/2}) \). Note that
\[
\Pr \{ A_{1,n} > A \} \leq \Pr \{ A_{1,n} > A, |v_n - v| \leq \epsilon_n, \delta \} + \Pr \{ A_{1,n} > A, |v_n - v| > \epsilon_n, \delta \} \leq \Pr \{ A_{1,n} > A, |v_n - v| \leq \epsilon_n, \delta \} + \Pr \{ |v_n - v| > \epsilon_n, \delta \}.
\]
From Assumption 1 and a second order Taylor expansion, we know that, when \( |v_n - v| < \epsilon, A_{1,n} < M(v_n - v)^2 \) with \( M = \sup_{s \in (-\epsilon, \epsilon)} f'(s) \). When \( n \) is sufficiently large and \( \delta \) is sufficiently small, \( \epsilon_n, \delta < \epsilon \). Let \( A = M \epsilon_n^{2,\delta} \). Then, combining with Proposition 2, we have \( A_{1,n} = O_{a.s.}(\epsilon_n^{2,\delta}) = O_{a.s.}(n^{-1+2/p+\delta}) \) for any \( \delta > 0 \). Let \( A = n^{-1}g(n) \), combining with Proposition 2, we have \( \Pr(\xi_{1,n} > cn^{-1}) \to 0 \) as \( n \to \infty \) for any \( c > 0 \). Thus, \( A_{1,n} = o_p(n^{-1/2}) \).

Second, we prove \( A_{2,n} = O_{a.s.} \left( n^{-\frac{2}{p+\delta} + \frac{1}{p} + \delta} \right) \). With \( \epsilon_n, \delta \) defined in Proposition 2 and \( c_1 \) and \( \xi_{n,\delta} \) defined in the proof of Proposition 3, we have
\[
\Pr \left\{ |A_{2,n}| > c_1 \xi_{n,\delta} \right\} = \Pr \{ |F_n(v_n) - F(v_n) + F(v) - F_n(v)| > c_1 \xi_{n,\delta} \} \leq \Pr \{ |F_n(v_n) - F(v_n) + F(v) - F_n(v)| > c_1 \xi_{n,\delta}, |v_n - v| > \epsilon_n, \delta \} + \Pr \{ |v_n - v| > \epsilon_n, \delta \} + \Pr \{ |v_n - v| > \epsilon_n, \delta \}.
\]
Combining with Propositions 2 and 3, we can easily see \( A_{2,n} = O_{a.s.}(\xi_{n,\delta}) = O_{a.s.} \left( n^{-\frac{2}{p+\delta} + \frac{1}{p} + \delta} (\log n)^{1/2} \right) \) for \( \delta > 0 \). Moreover, we have \( n^{1/2}A_{2,n} = O_{a.s.} \left( n^{-\frac{2}{p+\delta} + \frac{1}{p} + \delta} (\log n)^{1/2} \right) \). For \( \delta \) sufficiently small, we know \( n^{1/2}A_{2,n} \to 0 \) w.p.1 and thus as \( n \to 0 \), \( n^{1/2}A_{2,n} \to 0 \) in probability.

Then, we prove \( A_{3,n} = O_{a.s.}(n^{-1}) \) and \( A_{3,n} = o_p(n^{-1/2}) \). Let \( \mathcal{L}_n = \max \{ L_i, i = 1, \ldots, n : F_n(x) < \alpha \} \). Then, we have \( F_n(u_n) < \alpha \) and \( F_n(u_n) = F_n(u_n) + \mathcal{L}(v_n)/n \).
\[
\Pr \{ A_{3,n} > C/n \} = \Pr \{ F_n(v_n) > \alpha + C/n, |v_n - v| \leq \epsilon_n, \delta \} + \Pr \{ F_n(v_n) > \alpha + C/n, |v_n - v| > \epsilon_n, \delta \} \leq \Pr \{ F_n(v_n) > \alpha + C/n, |v_n - v| \leq \epsilon_n, \delta \} + \Pr \{ |v_n - v| > \epsilon_n, \delta \}.
\]
For \( n \) big enough and \( \delta \) sufficient small, we have \( \epsilon_n, \delta < \epsilon \) with \( \epsilon \) and \( C \) defined in Assumption 2. When \( |v_n - v| < \epsilon, F_n(v_n) = F_n(u_n) + \mathcal{L}(v_n)/n < \alpha + C/n \). Thus the first part of the equation is 0 for \( n \) big enough. Combining with Proposition 2, we have \( A_{3,n} = O_{a.s.}(n^{-1}) \) and therefore \( A_{3,n} = o_p(n^{-1/2}) \).

Furthermore, if \( \mathcal{L}(x) < C \) for any \( x < v + \epsilon \), we can similarly prove that \( A_{1,n} = O_{a.s.}(n^{-1} \log n) \), \( A_{2,n} = O_{a.s.} \left( n^{-3/4}(\log n)^{3/4} \right) \) and \( A_{3,n} = O_{a.s.}(n^{-1}) \). \[\square\]
A.2 Proof of Lemma 2

Proof. By Lemma 1, we know \( F_n(v_n) - F(v) = A_3, n = O_{a.s.}(n^{-1}) \) and \( A_3, n = o_p(n^{-1/2}) \). From Lemma 1, \( v_n - v = O_{a.s.}(n^{-1/2+1/p+\delta}) \) and from Proposition 2, \( v_n - v = o_p(n^{-1/2}g(n)) \) with \( g(n) \) defined in Proposition 2. It suffices to prove that \( F_n(v) - \alpha = O_{a.s.}(n^{-1/2+1/p+\delta}) \) and \( F_n(v) - \alpha = o_p(n^{-1/2}g(n)) \). Similar arguments as in Proposition 1 yield that, for \( n \) sufficiently large

\[
\Pr \{|F_n(v) - \alpha| > \gamma\} = \Pr \left\{ \left| \sum_{i=1}^{n} (I\{L_i \leq v\}L_i - \alpha) \right| > n\gamma \right\} \leq \frac{E[\left| \sum_{i=1}^{n} (I\{L_i \leq v\}L_i - \alpha) \right|^p]}{(n\gamma)^p} < \overline{C}_p \frac{1}{n^{p/2}(\gamma)^p},
\]

where \( \overline{C}_p \) is a constant. For any \( c > 0 \), let \( \gamma = cn^{-1/2}g(n) \). We have

\[
\Pr \left\{ \frac{|F_n(v) - \alpha|}{n^{-1/2}g(n)} > c \right\} = \Pr \left\{ |F_n(v) - \alpha| > \gamma \right\} < \overline{C}_p/[cg(n)]^p.
\]

Thus, \( |F_n(v) - \alpha| = o_p(n^{-1/2}g(n)) \). Similarly, combining with Borel-Cantelli Lemma, we can prove \( |F_n(v) - \alpha| = O_{a.s.}(n^{-1+\delta}) \). When \( \mathcal{L}(L) < C \) for any \( L < v + \varepsilon \), we can similarly prove \( B_n = O_{a.s.}(n^{-1} \log n) \). □

References