IMPORTANCE SAMPLING FOR MONTE CARLO ESTIMATION OF QUANTILES

Peter W. Glynn[1]
Department of Operations Research
Stanford University
Stanford, CA 94305–4022, U.S.A.

Abstract

This paper is concerned with applying importance sampling as a variance reduction tool for computing extreme quantiles. A central limit theorem is derived for each of four proposed importance sampling quantile estimators. Efficiency comparisons are provided in a certain asymptotic setting, using ideas from large deviation theory.
Keywords: quantiles, importance sampling, large deviations.

1 Introduction

Let X be a real-valued random variable with distribution function $F(\cdot)$. For $0 < p < 1$, the quantity

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

is called the p'th quantile of X. The quantile estimation problem is concerned with computing such quantities efficiently.

Quantile estimation is of interest in many applications settings. For example, in computing tables of critical values associated with complicated hypothesis tests in which F cannot be computed analytically in closed form, it may be necessary to resort to Monte Carlo methodology as a means of calculating such critical values; these values are defined as quantiles of an appropriately defined test statistic X. A second setting in which quantile estimation arises naturally is in the manufacturing context, in which a supplier may wish to compute a "promise date" by which the company can guarantee with high probability, delivery of the required product to its customers. Such a computation involves calculating an appropriately defined quantile associated with the (random) time required to process an order, from the instant of order placement to order fulfillment.

This paper is concerned with efficiently computing such quantiles via an application of the variance reduction technique known as importance sampling. Because of the success observed in applying importance sampling to rare-event

[1]This work was supported by the Army Research Office under contract no. DAAH04-94-G-0021.

simulations (see, for example, Chang et al (1992)), the marriage of importance sampling with quantile estimation is natural, particularly when p is either close to 0 or 1.

2 The Basic Idea

Our goal is to compute $\alpha = F^{-1}(p)$. Conventional Monte-Carlo estimation of α involves generating i.i.d. replications $X_1, X_2, \ldots$ having common distribution F and estimating α via the estimator

$$\alpha_n = F_n^{-1}(p) \stackrel{\Delta}{=} \inf\{x : F_n(x) \geq p\},$$

where $F_n(\cdot)$ is the empirical distribution function defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x).$$

To apply importance sampling to this problem, we choose a distribution function $\tilde{F}$, from which variate generation is possible, and for which the probability measure associated with F is absolutely continuous with respect to the probability measure associated with $\tilde{F}$. This implies the existence of a "density" $p(\cdot)$ for which we can write

$$F(dx) = p(x)\tilde{F}(dx) \tag{1}$$

for all $x \in \mathbb{R}$. We assume that $p(\cdot)$ is a known function. Let $P(\cdot)$ ($\tilde{P}(\cdot)$) and $E(\cdot)$ ($\tilde{E}(\cdot)$) correspond to the probability distribution and expectation, respectively, under which the X_i's are i.i.d. with common distribution F ($\tilde{F}$). Relation (1) implies that

$$\begin{aligned} P(X_i \leq x) &= \tilde{E} L_i I(X_i \leq x), \\ P(X_i > x) &= \tilde{E} L_i I(X_i > x), \end{aligned} \tag{2}$$

where $L_i = p(X_i)$ is the "likelihood ratio" associated with X_i. Set $\bar{L}_n = n^{-1} \sum_{i=1}^{n} L_i$. The two above equalities motivate the following estimators for F:

$$\begin{aligned} \tilde{F}_{1n}(x) &= \tfrac{1}{n} \textstyle\sum_{i=1}^{n} L_i I(X_i \leq x), \\ \tilde{F}_{2n}(x) &= 1 - \tfrac{1}{n} \textstyle\sum_{i=1}^{n} L_i I(X_i > x), \\ \tilde{F}_{3n}(x) &= \tilde{F}_{1n}(x)/\bar{L}_n, \\ \tilde{F}_{4n}(x) &= \tilde{F}_{2n}(x)/\bar{L}_n; \end{aligned} \tag{3}$$

it is, of course, understood in (3) that the X_i's appearing there are generated from $\tilde{F}$ rather than F. The corresponding quantile estimators are then defined by

$$\tilde{\alpha}_{in} = \inf\{x : \tilde{F}_{in}(x) \geq p\}$$

for $i = 1, \ldots, 4$.

Our main result in this section is a central limit theorem (CLT) for the four estimators just introduced. In addition to establishing the rate of convergence for those estimators, the CLT below can be used as the basis for large-sample confidence intervals for α.

Theorem 1 *Suppose that $\tilde{E}L_1^3 < \infty$ and assume that F is differentiable at $\alpha = F^{-1}(p)$ with $F'(\alpha) > 0$. Then, for $i = 1, \ldots, 4$,*

$$n^{1/2}(\tilde{\alpha}_{in} - \alpha) \Rightarrow \sigma_i N(0,1)$$

as $n \to \infty$, where

$$\begin{aligned}
\sigma_1^2 &= (\tilde{E}[L_1^2 I(X_1 \le \alpha)] - p^2)/F'(\alpha)^2, \\
\sigma_2^2 &= (\tilde{E}[L_1^2 I(X_1 > \alpha)] - (1-p)^2)/F'(\alpha)^2, \\
\sigma_3^2 &= \tilde{E}[L_1^2 (I(X_1 \le \alpha) - p)^2]/F'(\alpha)^2, \\
\sigma_4^2 &= \tilde{E}[L_1^2 (I(X_1 > \alpha) - (1-p))^2]/F'(\alpha)^2.
\end{aligned}$$

Proof: Since $\tilde{F}_{in}(\cdot)$ is right-continuous and non-decreasing, it follows that for each x

$$\tilde{P}(n^{1/2}(\tilde{\alpha}_{in} - \alpha) \le x) = \tilde{P}(p \le \tilde{F}_{in}(\alpha + xn^{-1/2})). \tag{4}$$

Set

$$\begin{aligned}
s_{1n}(x) &= (\tilde{E}[L_1^2 I(X_1 \le \alpha + xn^{-1/2})] - F(\alpha + xn^{-1/2})^2)^{1/2}, \\
s_{2n}(x) &= (\tilde{E}[L_1^2 I(X_1 > \alpha + xn^{-1/2})] - P(X_1 > \alpha + xn^{-1/2})^2)^{1/2}, \\
s_{3n}(x) &= (\tilde{E}[L_1^2 (I(X_1 \le \alpha + xn^{-1/2}) - F(\alpha + xn^{-1/2}))^2])^{1/2}, \\
s_{4n}(x) &= (\tilde{E}[L_1^2 (I(X_1 > \alpha + xn^{-1/2}) - P(X_1 > \alpha + xn^{-1/2}))^2])^{1/2},
\end{aligned}$$

and $s_i = F'(\alpha)(\sigma_i^2)^{1/2}$. Put

$$\begin{aligned}
\psi_{in}(x) &= n^{1/2}(\tilde{F}_{in}(\alpha + xn^{-1/2}) - F(\alpha + xn^{-1/2}))/s_{in}(x) \text{ and} \\
a_{in}(x) &= n^{1/2}(p - F(\alpha + xn^{-1/2}))/s_{in}(x).
\end{aligned}$$

Then,

$$\begin{aligned}
\tilde{P}(p \le \tilde{F}_{in}(\alpha + xn^{-1/2})) &= \tilde{P}(\psi_{in}(x) \ge a_{in}(x)) - P(N(0,1) \ge a_{in}(x)) \quad (5) \\
&+ P(N(0,1) \ge a_{in}(x)).
\end{aligned}$$

Now, $\tilde{E}L_1^3 < \infty$ guarantees that the appropriate (absolute) third moments of the normalized sum $\psi_{in}(x)$ are uniformly bounded in n, so that the Berry-Esséen theorem guarantees that

$$\sup_y |\tilde{P}(\psi_{in}(x) \ge y) - P(N(0,1) > y)| \to 0 \tag{6}$$

as $n \to \infty$; for $i = 1, 2$, the classical Berry Esséen theorem (see, for example, Feller (1971)) may be applied, whereas for $i = 3, 4$, one may appeal to an extension due to Bhattacharya and Ghosh (1978).

The differentiability of F at α guarantees that $a_{in}(x) \to -xF'(\alpha)/s_i$ as $n \to \infty$. Since $F'(\alpha)$ is assumed to be positive, the theorem then follows from (4) - (6).

□

Note that if $\tilde{F} = F$, then $L_1 = 1$ a.s. and Theorem 1 recovers the CLT for the conventional quantile estimator α_n.

Given a set of simulated values $(X_1, L_1), \ldots, (X_n, L_n)$ generated under $\tilde{P}$, the computation of $\tilde{\alpha}_{in}$ proceeds in much the same way as for α_n. One first orders the X_i's in ascending order, thereby forming the ordered sample $(X_{(1)}, \ldots, X_{(n)})$. Then, $\tilde{\alpha}_{1n}$ is the value $X_{(i_1)}$ associated with the first integer i_1 for which $\sum_{j=1}^{i_1} p(X_{(j)}) \geq pn$, $\tilde{\alpha}_{2n}$ is the value $X_{(i_2)}$ associated with the greatest integer i_2 for which $\sum_{j=i_2}^{n} p(X_{(j)}) \geq (1-p)n$, $\tilde{\alpha}_{3n}$ is the value $X_{(i_3)}$ associated with the first integer i_3 for which $\sum_{j=1}^{i_3} p(X_{(j)}) \geq p\sum_{j=1}^{n} L_j$, and $\tilde{\alpha}_{4n}$ is the value $X_{(i_4)}$ associated with the greatest integer i_4 for which $\sum_{j=i_4}^{n} p(X_{(j)}) \geq n - p\sum_{j=1}^{n} L_j$.

In principle, producing confidence intervals for α based on $\tilde{\alpha}_{in}$ is no more difficult than that for the conventional estimator α_n. The major challenge is finding a good way of estimating $F'(\alpha)$, either explicitly or implicitly.

3 A Candidate Importance Sampling Distribution

Large deviations theory (see, for example, Bucklew (1990)) suggests that, in certain settings, the tail approximation

$$P(X > x) \approx \exp(-x\theta_x + \phi(\theta_x)) \tag{7}$$

is valid for $x \gg EX$, where θ_x is the root of the equation $\phi'(\theta_x) = x$, and ϕ is the cumulant generating function of X. Not surprisingly, the tail approximation (7) suggests a quantile approximation.

In particular, let $\tilde{\theta}_p$ be the root of the equation

$$-\tilde{\theta}_p\phi'(\tilde{\theta}_p) + \phi(\tilde{\theta}_p) = \log(1-p). \tag{8}$$

Then, according to (7), for p close to 1,

$$P(X > \phi'(\tilde{\theta}_p)) \approx 1 - p, \tag{9}$$

suggesting that $\phi'(\tilde{\theta}_p)$ can be used as an approximation to the quantile $\alpha(p) \triangleq F^{-1}(p)$. Of course, typically this approximation will be crude, at best, and Monte Carlo computation is then required to accurately estimate $\alpha(p)$.

Note that if we set $\tilde{F}(dx) = \exp(\tilde{\theta}_p x - \phi(\tilde{\theta}_p))F(dx)$, it is easily verified that this yields an importance sampling distribution having mean $\phi'(\tilde{\theta}_p)$. The relation (9) therefore indicates that, under $\tilde{F}$, sampling from the appropriate tail event associated with the quantile $\alpha(p)$ is no longer a rare event, suggesting the possibility of a variance reduction.

This approach is, of course, only practical when $\phi(\cdot)$ is known and variates from $\tilde{F}$ can be generated relatively easily. Observe also that $x(\theta) = -\theta\phi'(\theta) + \phi(\theta)$ has derivative $-\theta\phi''(\theta)$. Because the cumulant generating function ϕ is known to be convex, it follows that $x(\cdot)$ is typically strictly decreasing, implying the uniqueness of roots to (8).

In order to assess the quality of the estimators $\tilde{\alpha}_{in}$ $(1 \le i \le 4)$, in conjunction with the above candidate importance sampling distribution, we now consider the problem of estimating the p'th quantile of a $N(0,1)$ r.v. for p close to 1. In this case, it is easily verified that $\tilde{\theta}_p = (-2\log(1-p))^{1/2}$ and that $\tilde{F}$ is the distribution of a normal r.v. with mean $y = (-2\log(1-p))^{1/2}$ and unit variance. Each of the quantities $F'(\alpha)^2\sigma_i^2$ $(1 \le i \le 4)$ can be expressed in terms of $\tilde{E}L_1^2$ and $\tilde{E}L_1^2 I(X_1 > \alpha)$. Letting N be a standard normal r.v., note that $p(x) = \exp(-(2xy - y^2)/2)$, so

$$\begin{aligned} \tilde{E}L_1^2 &= E\exp(-(2(N+y)y - y^2)) \\ &= \exp(y^2) \\ &= (1-p)^{-2}. \end{aligned}$$

Similarly, by "completing the square", it is easily shown that

$$\tilde{E}L_1^2 I(X_1 > \alpha) = \exp(y^2)P(N > \alpha + y).$$

But $\alpha \sim y$ as $p \nearrow 1$ (Barlow and Proschan (1975)) and

$$P(N > x) \sim \exp(-x^2/2)/(x\sqrt{2\pi})$$

as $x \to \infty$ (Feller (1968)), so $\log \tilde{E}L_1^2 I(X_1 > \alpha) \sim 2\log(1-p)$ as $p \nearrow 1$. It follows that $F'(\alpha)^2\sigma_1^2 \sim (1-p)^{-2}$, $F'(\alpha)^2\sigma_2^2$ is of order $(1-p)^2$, and $F'(\alpha)^2\sigma_3^2$ and $F'(\alpha)^2\sigma_4^2$ are both of order 1. Thus, the preferred estimator for computing the p'th quantile, when p is close to 1, is $\tilde{\alpha}_{2n}$. (Obviously, when p is close to 0, $\tilde{\alpha}_{1n}$ is to be preferred.)

4 Asymptotic Analysis

The analysis of section 3 suggests that the estimator $\tilde{\alpha}_{2n}$, when used in conjunction with the candidate distribution $\tilde{F}$ described there, can be an effective variance reduction tool. We shall now rigorously establish this conclusion in a certain asymptotic setting. Specifically, suppose that $X = S_m$, with

$S_m = Y_1 + \cdots + Y_m$ and the Y_i's i.i.d. with m large. Assume we are interested in computing the p'th quantile of S_m, where $p = 1 - \exp(-\beta m)$ $(\beta > 0)$.

If $\phi_Y(\theta) = \log E\exp(\theta Y_i)$,then (8) is equivalent to computing a root of

$$-\tilde{\theta}_p \phi'_Y(\tilde{\theta}_p) + \phi_Y(\tilde{\theta}_p) = -\beta. \tag{10}$$

Because $\phi(\theta) = m\phi_Y(\theta)$, it is evident that if $\tilde{F}$ is the candidate distribution proposed in section 3, then

$$\tilde{P}(S_m \in dx) = \exp(\tilde{\theta}_p S_m - m\phi_Y(\theta))P(S_m \in dx).$$

For this particular candidate distribution, we now wish to compare the efficiency of $\tilde{\alpha}_{2n}$ and α_n.

Note that $n^{1/2}(\alpha_n - \alpha) \Rightarrow \sigma N(0,1)$ as $n \to \infty$, where $\sigma^2 = p(1-p)/F'(\alpha)^2$. We shall compare σ_2^2 to σ^2 as $m \to \infty$.

Theorem 2 *Suppose that (10) has a solution and that $\phi_Y(\cdot)$ is differentiable in a neighborhood of $\tilde{\theta}_p$. Then,*

$$\begin{aligned} \lim_{m\to\infty} \frac{1}{m}\log(F'(\alpha)^2\sigma^2) &= -\beta, \\ \limsup_{m\to\infty} \frac{1}{m}\log(F'(\alpha)^2\sigma_2^2) &\leq -2\beta. \end{aligned}$$

Proof: The first assertion is trivial. For the second, note that it is sufficient to prove that $\limsup m^{-1}\log \tilde{E}L^2 I(S_m > \alpha) \leq -2\beta$. Since the root $\tilde{\theta}_p$ to (10) is necessarily positive, it follows that

$$\begin{aligned} \tilde{E}L^2 I(S_m > \alpha) &= \tilde{E}\exp(2(-\tilde{\theta}_p S_m + m\phi_Y(\tilde{\theta}_p)))I(S_m > \alpha) \\ &\leq \tilde{E}\exp(2(-\tilde{\theta}_p \alpha + m\phi_Y(\tilde{\theta}_p)))I(S_m > \alpha) \\ &\leq \exp(2(-\tilde{\theta}_p \alpha + m\phi_Y(\tilde{\theta}_p))). \end{aligned}$$

But by examining the logarithmic limit behavior of $P(S_m > mx)$ for x close to $\phi'_Y(\tilde{\theta}_p)$ (using Cramér's theorem), it is straightforward to verify that $\alpha = m\phi'_Y(\tilde{\theta}_p) + o(m)$. Appealing to (10) completes the proof.

□

Theorem 2 shows that in the context of computing extreme quantiles of random variables that can be represented as sums, substantial variance reductions can be obtained by using the ideas presented in this paper.

References

Barlow, R. E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart, and Winston, New York.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* *6*, 434–451.

Bucklew, J. A. (1990). Large Deviation Techniques in Decision, Simulation, and Estimation. John Wiley, New York.

Chang, C. S., Heidelberger, P., Juneja, S., and Shahabuddin, P. (1992). Effective bandwidth and fast simulation of ATM intree networks. Research Report RC 18586, T. J. Watson Research Center.

Feller, W. (1968). An Introduction to Probability Theory and its Applications, Volume 1. John Wiley, New York.

Feller, W. (1971). An Introduction to Probability Theory and its Applications, Volume 2. John Wiley, New York.