Non-asymptotic confidence intervals for importance sampling estimators of quantiles

Baalu Ketema[1,2,*], Roman Sueur[1], Nicolas Bousquet[1,3,4], Bertrand Iooss[1,2,3], Fabrice Gamboa[2], Francesco Costantino[2]

[1] *EDF R&D, 6 Quai Watier, 78400 Chatou, France*
[2] *Institut de Mathématiques de Toulouse, 31062, Toulouse, France*
[3] *SINCLAIR AI Laboratory, Saclay, France*
[4] *Sorbonne Université, LPSM, 4 place Jussieu, Paris, France*

[*] *Corresponding author - baalu-belay.ketema@edf.fr*

Résumé. La construction d'intervalle de confiance (asymptotiques ou non-asymptotiques) est une étape cruciale pour comprendre la qualité de l'estimation d'une quantité d'intérêt bâtie sur une distribution. Dans cette présentation, nous estimons un quantile q_α d'une variable aléatoire réelle $Y \sim \mu$ dans le cas où seul un échantillon d'une autre distribution μ_0 est disponible et où μ_0 domine μ. La méthode d'estimation utilisée est l'échantillonnage préférentiel. Un TCL est connu pour l'estimateur du quantile mais la variance asymptotique dépend du quantile q_α de μ, l'inconnu, et de sa fonction de répartition F_μ. Nous levons ce verrou en construisant un intervalle de confiance non-asymptotique pour q_α qui peut être utile lorsque l'on ne dispose que d'un échantillon de taille limitée.

Mots-clés. Echantillonnage préférentiel, estimation de quantile, inégalité de concentration, intervalles de confiance non-asymptotique.

Abstract. Building a confidence region (asymptotic or non-asymptotic) is crucial in understanding the quality of point estimators of a distribution. In this presentation, we estimate a quantile q_α of a real random variable $Y \sim \mu$ in the case where only a sample from another distribution μ_0 is available and where μ_0 dominates μ. This estimation procedure is known as importance sampling. A CLT is proved for the quantile estimator but the asymptotic variance depends on the quantile q_α of μ, the unknown, and on its cumulative distribution function F_μ. We lift this barrier by building a non-asymptotic confidence interval for q_α which can be useful when only a limited sample size is available.

Keywords. Importance sampling, quantile estimation, concentration inequality, non-asymptotic confidence intervals.

1 Introduction

In many industrial contexts, quantities of interest (QoI) are defined from real variables $Y \sim \mu$ considered as random with underlying distribution μ, that represent the behavior of a component or system. For instance Y is the output of a code that computes the level of a river [7] or the cladding temperature in a nuclear vessel after an accident [4]. Typical QoIs are the quantile $q_\alpha(Y)$, the superquantile $Q_\alpha(Y)$ [6] or a probability $p_T = \mathbb{P}(Y > T)$ given some threshold T. Usually these QoIs cannot be computed explicitly if μ cannot be easily handled (ie., not in closed form). Hence statistical estimation is required to approximate these quantities. In addition, confidence regions are usually built to understand how far is the estimator to the real value.

The standard estimation method uses a Monte Carlo simulation to approximate the cumulative distribution function (cdf) of Y denoted F_μ: for $Y_1, ..., Y_N$ an iid sample from μ, we have that

$$\widehat{F}(t) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{Y_i \leq t}$$

is a non biased estimator of the cdf of Y for all t in $\mathbb{R}$. It converges a.s. and uniformly in t to F_μ. It also verifies a functional central limit theorem ([9], Chapter 19). We can therefore build plug-in estimators for the quantile $q_\alpha(Y)$, the superquantile $Q_\alpha(Y)$ and the probability threshold p_T:

$$\begin{aligned}
\widehat{q}_\alpha &:= \inf\{t \in \mathbb{R} \mid \widehat{F}(t) \geq \alpha\}, \\
\widehat{Q}_\alpha &:= \frac{1}{N(1-\alpha)} \sum_{i=1}^{N} Y_i \mathbf{1}_{Y_i \geq \widehat{q}_\alpha}, \\
\widehat{p}_T &:= 1 - \widehat{F}(T).
\end{aligned}$$

The associated presentation only focuses on the quantile $q_\alpha(Y)$. From [9], Chapter 21, the estimator $\widehat{q}_\alpha$ satisfies a central limit theorem: assuming that F_μ is differentiable at point $q_\alpha := q_\alpha(Y)$ with $F'_\mu(q_\alpha) > 0$ then we have that

$$\sqrt{N}\big(\widehat{q}_\alpha - q_\alpha\big) \to \mathcal{N}(0, \sigma^2_\infty),$$

where

$$\sigma^2_\infty = \frac{\alpha(1-\alpha)}{F'_\mu(q_\alpha)^2}.$$

The asymptotic variance depends on $q_\alpha = q_\alpha(Y)$, the unknown QoI, as well as on F_μ and therefore this result cannot be used directly to construct asymptotic confidence intervals. In addition, obtaining a large sample of Y, a requirement for asymptotic confidence intervals, can be very time-consuming (for instance Y can be the output of a costly industrial computer code ie., $Y = G(X)$ and a sample of Y is obtained by evaluating G on a sample of X). Consequently, non-asymptotic confidence intervals are more appropriate in this case, as they provide information on the concentration of the estimator around the true value as a function of the sample size N.

The following paper is organized as follows. Section 2 discusses how non-asymptotic confidence intervals can be built for the standard quantile estimator using a uniform concentration inequality on the empirical cdf. Section 3 explains the importance sampling method for the quantile estimator and states a CLT for the latter. Section 4 states the main result of the paper ie., a method for building non-asymptotic confidence intervals for importance sampling estimators of quantiles. And lastly, Section 5 discusses the limitations of this method.

2 Concentration inequality for the quantile estimator

The Dvoretzky–Kiefer–Wolfowitz (DKW) theorem [3, 8] shows that the estimator $\widehat{F}$ of F_μ verifies the following concentration inequality:

$$\mathbb{P}\left(\sup_{t\in\mathbb{R}} |\widehat{F}(t) - F_\mu(t)| > \eta\right) \leq 2e^{-2N\eta^2},$$

for all $\eta > 0$. This concentration inequality can be used to obtain a non-asymptotic confidence interval for the quantile estimator $\widehat{q}_\alpha$: for all $\eta > 0$ small enough and $\alpha \in (0,1)$ fixed

$$\mathbb{P}\Big(\widehat{q}_{\alpha-\eta} \leq q_\alpha(Y) \leq \widehat{q}_{\alpha+\eta}\Big) \geq 1 - 2e^{-2N\eta^2}, \tag{1}$$

since DKW implies that with probability at least $1 - 2e^{-2N\eta^2}$, the functional inequality

$$\widehat{F} - \eta \leq F_\mu \leq \widehat{F} + \eta \tag{2}$$

is verified. And since $\widehat{F}$ and F_μ are non-decreasing functions, taking the generalized inverse in (2) gives for all $\alpha \in (0,1)$

$$(\widehat{F} - \eta)^{\langle -1\rangle}(\alpha) \geq F_\mu^{\langle -1\rangle}(\alpha) = q_\alpha(Y) \geq (\widehat{F} + \eta)^{\langle -1\rangle}(\alpha) \tag{3}$$

with probability at least $1-2e^{-2N\eta^2}$, where $H^{\langle -1\rangle}$ is the generalized inverse of a non-decreasing function H defined as $H^{\langle -1\rangle}(\alpha) := \inf\{t \in \mathbb{R} \mid H(t) \geq \alpha\}$. The non-asymptotic confidence interval's quality heavily depends on the sample size N of Y, see Table 1, as well as on the order of the quantile α ie., for α close to 0 or 1 a very large sample will be required to accurately approximate the quantile.

Sample size N	$N = 10^4$	$N = 10^5$	$N = 2 \times 10^6$
Confidence level ≥ 0.75	$[1.52,\ 1.71]$	$[1.62,\ 1.68]$	$[1.635,\ 1.649]$
Confidence level ≥ 0.95	$[1.50,\ 1.76]$	$[1.61,\ 1.69]$	$[1.632,\ 1.651]$
Confidence level ≥ 0.99	$[1.48,\ 1.79]$	$[1.60,\ 1.70]$	$[1.631,\ 1.654]$

Table 1: Confidence intervals (1) on the 0.95-quantile of $\mathcal{N}(0,1)$ in terms of the sample size and a fixed confidence level. The value of this quantile is approximately 1.6448.

3 Importance sampling estimation procedure

Assume we do not have access to a sample of μ but rather a sample from another distribution μ_0 on $\mathbb{R}$ which dominates μ_0 (ie., $\mu << \mu_0$ meaning that μ admits a density on $\mathbb{R}$ with respect to (wrt) μ_0). Denote $L := \frac{d\mu}{d\mu_0}$ the Radon-Nikodym derivative (also called the likelihood ratio). We would like to build estimators of q_α, a quantile of μ, as well as confidence intervals using an iid sample $Y_1, ..., Y_N$ of μ_0. To do so we can use the importance sampling (IS) method

$$\widehat{F}(t) := \frac{1}{N}\sum_{i=1}^{N} L(Y_i)\mathbf{1}_{Y_i \leq t},$$

which is the standard unbiased Monte Carlo estimator of $F_\mu(t)$. But $\widehat{F}$ is not the cdf of a discrete measure on $\mathbb{R}$ since the weights $\frac{L(Y_i)}{N}$ do not add up to one. Hence we favor instead the following biased estimator

$$\widehat{F}_{\text{is}}(t) := \frac{1}{\sum_{i=1}^N L(Y_i)} \sum_{i=1}^N L(Y_i)\mathbf{1}_{Y_i \le t},$$

which also converges pointwise a.s. to F_μ. It allows us to build an estimator of the quantile of μ by plug-in:

$$\widehat{q}_\alpha^{\ \text{is}} := \inf\{t \in \mathbb{R} \mid \widehat{F}_{\text{is}}(t) \ge \alpha\}.$$

The asymptotic properties of this estimator are already studied in [4, 5], who showed that if (a) L is cube-integrable wrt μ_0; (b) F_μ is differentiable at $q_\alpha := q_\alpha(Y)$ for $Y \sim \mu$; (c) $F'_\mu(q_\alpha) > 0$, then

$$\sqrt{N}\big(\widehat{q}_\alpha^{\ \text{is}} - q_\alpha\big) \underset{N\to\infty}{\longrightarrow} \mathcal{N}(0, \sigma_\infty^2),$$

where

$$\sigma_\infty^2 = \frac{\mathbb{E}_{Y\sim\mu_0}\big[L(Y)^2(\mathbf{1}_{Y\le q_\alpha} - \alpha)^2\big]}{F'_\mu(q_\alpha)^2}.$$

This asymptotic variance depends again on the unknown QoI $q = q_\alpha(Y)$. Hence this result cannot be directly used to build asymptotic confidence intervals.

4 Non-asymptotic confidence intervals for the IS quantile estimator

Now our goal is to build non-asymptotic confidence intervals for the quantile estimator $\widehat{q}_\alpha^{\ \text{is}}$. Assume we have a pointwise confidence interval for F_μ around $\widehat{F}_{\text{is}}$ ie., for each fixed t in $\mathbb{R}$, for all $N \in \mathbb{N}^*$ there exist $\varepsilon_N > 0$ and $\lambda^-_{t,N} < \lambda^+_{t,N}$ such that

$$\mathbb{P}\big(F_\mu(t) \in [\lambda^-_{t,N}, \lambda^+_{t,N}]\big) \ge 1 - \varepsilon_N,$$

where $\varepsilon_N \to 0$ as $N \to \infty$, $t \to \lambda^\pm_{t,N}$ are non-decreasing random functions and $\lambda^-_{t,N}, \lambda^+_{t,N} \underset{N\to\infty}{\longrightarrow} F_\mu(t)$ a.s.. Then we can prove the following result.

Theorem. *Under the previous assumptions, by choosing a decreasing sequence $(a_N)_{N\ge1}$ such that $a_N \le 1$ and $\lim_{N\to\infty} a_N = 0$ we have:*

$$\mathbb{P}\big(q^-_\alpha \le q_\alpha(Y) \le q^+_\alpha\big) \ge 1 - \left\lfloor \frac{1}{a_N} - 1 \right\rfloor \varepsilon_N,$$

where $q^\mp_\alpha$ are the generalized inverse of the functions $t \to \lambda^\pm_{t,N} \pm a_N$ evaluated at $\alpha \in (0,1)$. They are random variables taking values in the original sample $\{Y_1, ..., Y_N\} \cup \{\pm\infty\}$ of μ_0.

Note that this theorem does not guaranty that the quantity $1-\left\lfloor \frac{1}{a_N}-1 \right\rfloor \varepsilon_N$ is positive. This depends on the sequence a_N which has to be cleverly chosen. The proof is inspired from [2] and is based on the following two ingredients:

(i) we can convert the pointwise confidence interval into a uniform one by slightly enlarging it by doing $\lfloor \frac{1}{a_N}-1 \rfloor$ union bounds:

$$\mathbb{P}\left(\cap_{t\in\mathbb{R}}\left\{F_\mu(t)\in[\lambda^-_{t,N}-a_N,\lambda^+_{t,N}+a_N]\right\}\right)\geq 1-\left\lfloor \frac{1}{a_N}-1 \right\rfloor \varepsilon_N,$$

(ii) we then transform the inequality

$$\lambda^-_{t,N}-a_N\leq F_\mu(t)\leq \lambda^+_{t,N}+a_N,\quad \forall t\in\mathbb{R},$$

into an inequality on the quantile $q_\alpha(Y)$ by means of taking the generalized inverse wrt t as in (3).

Now, in order to obtain the $\lambda^\pm_{t,N}$ for the pointwise confidence interval of $F_\mu(t)$, for all t, we can apply Theorem 2 of [1]: for all $s>0$ take $N=e^{\mathcal{D}(\mu|\mu_0)+s}$, then for all $t\in\mathbb{R}$ we have

$$\mathbb{P}\left(\left|\widehat{F}_{\text{is}}(t)-F_\mu(t)\right|\geq \frac{2\varepsilon_s\sqrt{F_\mu(t)}}{1-\varepsilon_s}\right)\leq 2\varepsilon_s, \tag{4}$$

where $\mathcal{D}$ is the Kulback-Leibler divergence and ε_s is given by

$$\varepsilon_s=\left(e^{-s/4}+\sqrt{\mathbb{P}\big(\log L(Y)>\mathcal{D}(\mu|\mu_0)+s/2\big)}\right)^{1/2},$$

where $\log L(Y)$ is the log-likelihood ratio of μ and μ_0 evaluated at $Y\sim\mu_0$.

The concentration inequality (4) can be equivalently rewritten as

$$\mathbb{P}\Big(F_\mu(t)\in[\lambda^-_{t,s},\lambda^+_{t,s}]\Big)\geq 1-2\varepsilon_s,$$

where

$$\lambda^\pm_{t,s}:=\frac{2\widehat{F}_{\text{is}}(t)+\eta_s^2\pm\eta_s\sqrt{4\widehat{F}_{\text{is}}(t)+\eta_s^2}}{2},$$

and $\eta_s:=\frac{2\varepsilon_s}{1-\varepsilon_s}$. Indeed $\lambda^\pm_{t,s}$ verify the necessary assumptions of Theorem 4.

This method can also be used when we have a parametric family $\mathcal{P}:=\{\mu_\theta:\theta\in\Theta\}$ and $\mu,\mu_0\in\mathcal{P}$ and we have an iid sample $Y_1,...,Y_N$ wrt $\mu_{\theta_0}:=\mu_0$, and we want to build an estimator of a quantile on μ_θ and a corresponding confidence interval for all $\theta\in\Theta$.

5 Limitations

The quality of the confidence interval for the quantile $q_\alpha(Y)$ depends on the quality of the initial pointwise confidence interval on F_μ, built from a sample of μ_0. Indeed, the authors of [1] specifically mention that no efforts were made to improve the concentration inequality (4) so the pointwise confidence interval for $F_\mu(t)$ is not necessarily good. In addition, choosing the sequence $(a_N)_N$ is not obvious and requires a compromise between having a small uniform confidence interval around F_μ and a high confidence level. Moreover, the confidence interval obtained for the quantile is actually uniform in $\alpha \in (0,1)$ ie.,

$$\mathbb{P}\Big(\cap_{\alpha\in(0,1)} \left\{q_\alpha^- \leq q_\alpha(Y) \leq q_\alpha^+\right\} \Big) \geq 1 - \left\lfloor \frac{1}{a_N} - 1 \right\rfloor \varepsilon_N.$$

This is because we inverted a uniform bound in t on the cdf. This means that if we want a confidence interval for a specific α, for instance $\alpha = 0.95$, then the confidence level $1 - \left\lfloor \frac{1}{a_N} - 1 \right\rfloor \varepsilon_N$ might be too conservative since

$$\begin{aligned} \mathbb{P}\Big(q_{0.95}^- \leq q_{0.95}(Y) \leq q_{0.95}^+\Big) &\geq \mathbb{P}\Big(\cap_{\alpha\in(0,1)} \left\{q_\alpha^- \leq q_\alpha(Y) \leq q_\alpha^+\right\} \Big) \\ &\geq 1 - \left\lfloor \frac{1}{a_N} - 1 \right\rfloor \varepsilon_N. \end{aligned} \tag{5}$$

Therefore more work is needed to understand how much is lost at (5).

References

[1] S. Chatterjee and P. Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018.

[2] M. Ducoffe, S. Gerchinovitz, and J. S. Gupta. A high probability safety guarantee for shifted neural network surrogates. In *SafeAI@ AAAI*, pages 74–82, 2020.

[3] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.

[4] C. Gauchy, J. Stenger, R. Sueur, and B. Iooss. An information geometry approach to robustness analysis for the uncertainty quantification of computer codes. *Technometrics*, 64(1):80–91, 2022.

[5] P.W. Glynn. Importance sampling for monte carlo estimation of quantiles. In *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*, pages 180–185. Citeseer, 1996.

[6] B. Iooss, V. Vergès, and V. Larget. Bepu robustness analysis via perturbed law-based sensitivity indices. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236(5):855–865, 2022.

[7] P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa, and B. Iooss. Density modification-based reliability sensitivity analysis. *Journal of Statistical Computation and Simulation*, 85(6):1200–1223, 2015.

[8] P. Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The annals of Probability*, pages 1269–1283, 1990.

[9] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.