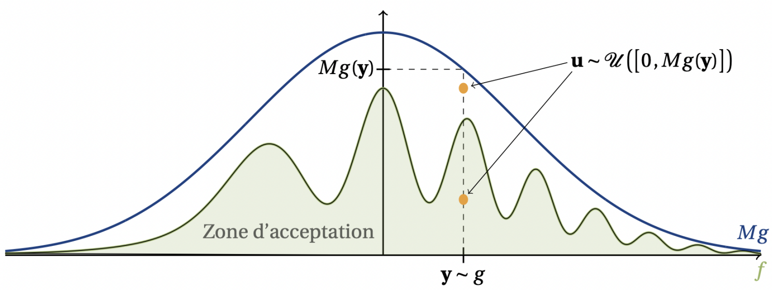
MONTE CARLO PROJECT III :

Samuel Elbaz, Alexandre Zenou,, Sacha Assouly



Dauphine | PSL
UNIVERSITÉ PARIS

Contents

Quantile estimation Naïve Reject algorithm

Question 23: Describe a method using accept-reject method to simulate a random variable X conditional to the event $\{X \in A\}$, $A \subset \mathbb{R}$. Justify theoretically

PREPARED BY Stanislas du Ché

M1: Monte Carlo methods Dauphine Paris University

Question 24: (♠) Using the previous question algorithm, propose a way to simulate $\delta = \mathbb{P}(X \geq q)$. Compute a function `accept_reject_quantile(q,n)` that returns a Monte Carlo estimate $\hat{\delta}_n^{Reject}$ of the target probability using n random variable simulations.
Question 25: (♠) Compute a confidence interval for δ at level 95% for a required precision ϵ.

Importance Sampling

Question 26: Propose a sampling distribution g to realise importance sampling and remind the importance sampling Monte Carlo estimator $\hat{\delta}_n^{IS}$ of δ for $n \in \mathbb{N}$. On which condition, the importance sampling estimator is preferred to the classical Monte Carlo estimator.
From now on, g will be a Cauchy distribution of the parameters (μ_0, γ)
Question 27: Remind the density of a Cauchy distribution. Choose parameters (μ_0, γ), explain your reasoning.
Question 28: (♠) Compute a R function `IS_quantile(q,n)` that gives, for $n = 10000$, the estimator $\hat{\delta}_n^{IS}$ using n simulated random variables. Compute a confidence interval for δ at level 95% at precision ϵ.

Control Variate

Question 29: Remind the definition of the score. Derive the partial derivative of the log-likelihood $\log f(x|\theta_1, \theta_2)$ under μ_1. We note $s_{\mu_1}(x|\theta) = \frac{\partial \log f(x|\theta_1,\theta_2)}{\partial \mu_1}$. (Remind that $\theta_1 = (\mu_1, \sigma_1^2)$ and $\theta_2 = (\mu_2, \sigma_2^2)$)
Question 30: Propose a control variate Monte Carlo Estimator $\hat{\delta}_n^{CV}$ using the control random variable $s_{\mu_1}(X|\theta)$
Question 31: (♠) Compute a R function `CV_quantile(q,n)` that gives for $n = 10000$, the estimator $\hat{\delta}_n^{CV}$ using n simulated random variables. Compute a confidence interval for δ at level 95 % at precision ϵ
Question 32: Compare the 3 methods: Naive, Control Variate and Importance Sampling. You can asses their algorithmic complexity, their computational cost for a required precision, etc.

Figure 1: Questions.

1 Naive Simulation

Question 23:
Let $X \overset{\text{law}}{=} f$. For a set $A \subseteq \mathbb{R}$ such that $\mathbb{P}(X \in A) > 0$, we wish to simulate law of $X|X \in A$ using accept reject method. We have :

$$\begin{aligned} f_{X|X\in A} &:= \frac{f_X \mathbf{1}_A}{\int_A f_X} \\ &\leq f_X \cdot \frac{1}{\int_A f_X} \end{aligned}$$

We recognize the natural M we need to find. As f is a density : $\int_{\mathbb{R}} f_X = 1$ then for our non trivial set A : $\int_A f_X < 1$ so that $M : \frac{1}{\int_A f_X} > 1$ and the instrumental known density is f_X. Hence, we have :

Rejection Method
1. Generate $Y \sim f_X$, $U \sim \mathcal{U}_{[0,1]}$;
2. Accept $Z = Y$ if $U \leq \mathbf{1}_A(Y) \int_A f_X(Y)$;
3. Return to 1 otherwise.

Question 24:

A Monte Carlo estimator is given by

$$\hat{\delta}_N = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{X_i > q\}}.$$

where $(X_k)_{k\leq N} \overset{\text{law}}{=} f$ are i.i.d. It is unbiased and converges towards δ almost surely by strong law of large numbers, since $(\mathbf{1}_{X_k>q})_{k\leq N} \in L^2$ is an iid sequence. Then, the CLT gives that :

$$\sqrt{N}\left(\hat{\delta}_N - \delta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \mathbb{P}(X > q)(1 - \mathbb{P}(X > q))),$$

then by continuity and Slutsky :

$$\sqrt{\frac{N}{\hat{\delta}_N \cdot (1 - \hat{\delta}_N)}}\left(\hat{\delta}_N - \delta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, 1),$$

The 95% confidence interval for δ is given by:

$$\left[\hat{\delta}_N - q_{0.975}\sqrt{\frac{\hat{\delta}_N(1-\hat{\delta}_N)}{N}}, \hat{\delta}_N + q_{0.975}\sqrt{\frac{\hat{\delta}_N(1-\hat{\delta}_N)}{N}}\right],$$

where $q_{0.975}$ is the quantile of the standard normal distribution at level 0.975.
To achieve a precision ε, we impose:

$$2q_{0.975}\sqrt{\frac{\hat{\delta}_N(1-\hat{\delta}_N)}{N}} \leq \varepsilon.$$

Solving for N gives:

$$N_\varepsilon := \left\lceil \frac{4q_{0.975}^2\hat{\delta}_N(1-\hat{\delta}_N)}{\varepsilon^2} \right\rceil.$$

2 Importance Sampling

Question 26: To estimate δ, we introduce an importance sampling distribution g that has larger tails than f to ensure more frequent sampling over $[q, +\infty[$. A basic well-suited choice is the truncated exponential distribution of parameter 1 i.e : $\forall x \in \mathbb{R}$:

$$g(x) := \frac{\exp^{-(x-q)} \mathbf{1}_{x>q}}{\int_q^\infty \exp^{-(s-q)}\, ds} = \exp^{-(x-q)} \mathbf{1}_{x>q}$$

Then :

$$\begin{aligned}\delta &:= \mathbb{P}(X > q) = \mathbb{E}_f[h(X)] \quad \text{where } h(X) := \mathbf{1}_{X>q} \in L^2(f)\\ &= \mathbb{E}_g\left[\frac{fh}{g}(X)\right]\\ &= \mathbb{E}_g\left[f(X)\exp\left(X-q\right)\mathbf{1}_{X>q}\right]\end{aligned}$$

Let us now define an unbiased estimator of δ for a random i.i.d sample $(Z_k)_{(k\leq N)} \overset{\text{law}}{=} \tau(q)\mathcal{E}(1)$ being the "truncated at q" exponential of parameter 1 :

$$\begin{aligned}\hat{\delta}_N^{(IS)} &:= \frac{1}{N}\sum_{k=1}^{N} f(Z_k)e^{Z_k-q}\mathbf{1}_{Z_k>q}\\ &= \frac{e^{-q}}{N(1-a)\sqrt{2\pi}}\sum_{k=1}^{N}\left(e^{Z_k-\frac{Z_k^2}{18}} - ae^{Z_k-\frac{1}{2}(Z_k-1)^2}\right)\mathbf{1}_{Z_k>q}\end{aligned}$$

Since $\{(e^{Z_k-\frac{Z_k^2}{18}} - ae^{Z_k-\frac{1}{2}(Z_k-1)^2})\mathbf{1}_{Z_k>q}\}_{k\leq N} \in L^2(g)$ and are i.i.d, from Strong Law of Large Numbers :

$$\hat{\delta}_N^{(IS)} \xrightarrow[N\to\infty]{a.s} \delta$$

According to theory, **Importance Sampling** is preferred when the estimator has a **smaller variance** than the one using the classical Monte Carlo method [1] It is generally used as in this project, when the tails are so light than estimating it's probability according to a given law, with R, gives 0 almost anytime. The method can be applied only if we found a relevant importance law such that **the weights f/g are well controlled** and that $\text{Var}_g(fh/g) < \infty$.

A way to check wether our g is a good choice or not, comes from the ESS, defined in the course as :

$$ESS(N) = N\frac{\bar{w}_N^2}{w_N^2}$$

where $w_k = \frac{f(Z_k)}{g(Z_k)}$ represents the importance weights associated with the samples $Z_k \sim g$. When the ESS is close to N, it means that our g is a good choice. However, if it is too far from N, then the weights are unstable. This idea of comparing these quantities comes from Cauchy-Schwartz and variance formula of the estimator. One can also use Calculus of Variation to find the optimal map g of Importance Sampling, which is just proportional to $f\cdot|h|$. The optimization problem would be the following :

$$\inf\left\{\int_{\mathbb{R}}\frac{(f(x)h(x))^2}{g(x)}dx \ \text{ with : } g(x)\geq 0 \text{ a.e.}, \int_{\mathbb{R}} g(x)dx = 1,\ g\in C^1(\mathbb{R}),\ \text{supp}(fh)\subseteq\text{supp}(g)\right\}$$

Question 27:

Two main conditions for an importance density to be chosen are :

1. $\text{Var}_g(\frac{f^2h^2}{g^2}) < \infty$
2. The importance weights, $\frac{f}{g}$, must be bounded and should not vary excessively over the set where we estimate h. This ensures that the variance of the weights is controlled enough, giving a sufficient precision for the importance sampling method to be relevant.

The density of a Cauchy distribution of parameters $(\mu_0,\gamma)\in\mathbb{R}\times\mathbb{R}_*^+$ is defined for any $x\in\mathbb{R}-\{\mu_0\}$, by the map :

$$g(x\mid\mu_0,\gamma) := \frac{\gamma}{\pi(\gamma^2+(x-\mu_0)^2)}$$

For some parameters that we are searching for, this density will have larger tails than f, so that it will be a good candidate for the IS method. Let's check for the two conditions given above.
The variance can be written as

$$\text{Var}_g\left(\frac{f(Z)h(Z)}{g(Z\mid\mu_0,\gamma)}\right) = \mathbb{E}_g\left[\left(\frac{f(Z)h(Z)}{g(Z\mid\mu_0,\gamma)}\right)^2\right] - \left(\mathbb{E}_g\left[\frac{f(Z)h(Z)}{g(Z\mid\mu_0,\gamma)}\right]\right)^2.$$

[1]Which is being defined as : $\overline{h}_N = \frac{1}{N}\sum_{k=1}^{N}h(X_k)$, for $(X_k)_{k\leq N}$ i.i.d. of law f. The variance is given by: $\frac{1}{N}\text{Var}_f(h(X))$

The second term is simply δ^2. This implies that the finiteness of the variance depends only on

$$\mathbb{E}_g\left[\left(\frac{f(Z)h(Z)}{g(Z \mid \mu_0,\gamma)}\right)^2\right] = \int_q^{+\infty} \frac{f(z)^2}{g(z \mid \mu_0,\gamma)}dz.$$

But this behaves like $O\left(z^2\exp\left(-\frac{z^2}{18}\right)\right)$ which belongs to $L^1(\mathbb{R})$. So for any (μ_0,γ), the variance of our Importance estimator is finite. One can try to search for (μ_0,γ) minimizing the variance, but it will lead to complex calculations.

Now we are going to search for **sufficient conditions on** (μ_0,γ) **to ensure that** $g(x \mid \mu_0,\gamma) \geq f(x)$ $\forall x \geq q$.
Since $f(x) = \frac{1}{1-a}(f_1(x) - af_2(x))$ and $f_1 \geq f$ everywhere, it is enough to ensure $g(\cdot \mid \mu_0,\gamma) \geq \frac{1}{1-a}f_1(\cdot)$.

Thus, we have :

$$\frac{\gamma}{\pi(\gamma^2 + (x-\mu_0)^2)} \geq \frac{1}{(1-a)\sqrt{2\pi\cdot 9}}\exp\left(-\frac{x^2}{18}\right).$$

Denote $K := \frac{1}{(1-a)\sqrt{2\pi\cdot 9}}$, we invert the previous inequality :

$$\pi\gamma + \frac{\pi(x-\mu_0)^2}{\gamma} \leq \frac{\exp\left(\frac{x^2}{18}\right)}{K}.$$

If we set $\mu_0 = q$, by positivity of our terms and the fact that both of the map are strictly increasing on $[q;+\infty[$, it is enough to have :

$$\pi\gamma \leq \frac{\exp(\frac{q^2}{18})}{K}.$$

Then a sufficient condition to have $g(\cdot \mid \mu_0,\gamma) \geq f(\cdot)$ for any $x \geq q$ is to set $\mu_0 := q$ and

$$\gamma := \frac{1}{\pi}\cdot(1-a)\sqrt{2\pi\cdot 9}\cdot\exp\left(\frac{q^2}{18}\right) > 0.$$

Question 28 :

The following is an importance sampling strongly consistent estimator for our desired $\delta = \mathbb{P}(X \geq q)$:

$$\hat{\delta}_N^{(IS)} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{1}_{Z_k\geq q}\cdot f(Z_k)\cdot\frac{\pi}{\gamma}\cdot\left(\gamma^2 + (Z_k-\mu_0)^2\right)$$

where $(Z_k)_{k\leq N} \overset{\text{law}}{=} g(\cdot \mid \mu_0,\gamma)$ are i.i.d. samples. As the terms inside the sum are i.i.d and square-integrable, by strong law of large numbers :

$$\hat{\delta}_N^{(IS)} \xrightarrow[N\to\infty]{a.s} \delta$$

Then using CLT :

$$\sqrt{N}\left(\hat{\delta}_N^{(IS)} - \delta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0,\sigma^2),$$

where σ^2 is the variance of the estimator given by:

$$\begin{aligned}\sigma^2 &= \mathrm{Var}_g(\hat{\delta}_N^{(IS)})\\ &= \frac{1}{N}\cdot\mathbb{E}_g\left[\left(\frac{f(Z)h(Z)}{g(Z \mid \mu_0,\gamma)} - \mathbb{E}_g\left[\frac{f(Z)h(Z)}{g(Z \mid \mu_0,\gamma)}\right]\right)^2\right]\\ &= \frac{1}{N}\cdot\mathbb{E}_g\left[\left(\frac{f(Z)h(Z)}{g(Z \mid \mu_0,\gamma)} - \delta\right)^2\right]\end{aligned}$$

Thus we define an unbiased estimator for σ^2 :

$$\hat{\sigma}_N^2 = \frac{1}{N-1}\sum_{k=1}^{N}\left(\frac{f(Z_k)}{g(Z_k)}h(Z_k) - \hat{\delta}_N^{(IS)}\right)^2 \xrightarrow[N\to\infty]{a.s} \sigma^2$$

By Slutsky's theorem :

$$\frac{\sqrt{N}}{\hat{\sigma}_N} \cdot \left(\hat{\delta}_N^{(IS)} - \delta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0,1).$$

A 95% confidence interval, where $\alpha = 0.05$ and $q^{95\%}_{\mathcal{N}(0,1)} \approx 1.96$, is given by :

$$\left[\hat{\delta}_N^{(IS)} - \frac{1.96\hat{\sigma}_N}{\sqrt{N}},\ \hat{\delta}_N^{(IS)} + \frac{1.96\hat{\sigma}_N}{\sqrt{N}}\right].$$

If a precision $\varepsilon > 0$ is required, the necessary sample size satisfies:

$$2 \cdot \frac{1.96\hat{\sigma}_N}{\sqrt{N}} \leq \varepsilon.$$

Whence :

$$N_\varepsilon := \left\lceil \left(\frac{3.92\hat{\sigma}_N}{\varepsilon}\right)^2 \right\rceil.$$

3 Control Variate

Question 29 :

We start by recalling the definition of the score:

$$s_{\mu_1}(x|\theta) = \frac{\partial}{\partial\theta} \log f(x|\theta).$$

The log-likelihood of f is :

$$\log f(x) = \log\left(\frac{1}{1-a}\right) + \log\left(f_1(x|\theta_1) - a \cdot f_2(x|\theta_2)\right).$$

Differentiating with respect to μ_1, we have:

$$\frac{\partial}{\partial\mu_1} \log f(x|\theta) = \underbrace{\frac{\partial}{\partial\mu_1} \log\left(\frac{1}{1-a}\right)}_{=0} + \frac{\frac{\partial}{\partial\mu_1}\left(f_1(x|\theta_1) - a \cdot f_2(x|\theta_2)\right)}{f_1(x|\theta_1) - a \cdot f_2(x|\theta_2)}.$$

Since $f_2(\cdot|\theta_2)$ does not depend on μ_1, we compute:

$$\frac{\partial}{\partial\mu_1} f_1(x|\theta_1) = f_1(x|\theta_1) \cdot \frac{x-\mu_1}{\sigma_1^2}, \quad \frac{\partial}{\partial\mu_1} f_2(x|\theta_2) = 0.$$

Then here is a closed form of our score function :

$$\forall x \in \mathbb{R} : s_{\mu_1}(x|(\theta_1;\theta_2)) := \frac{f_1(x|\theta_1)}{f_1(x|\theta_1) - a \cdot f_2(x|\theta_2)} \cdot \frac{x-\mu_1}{\sigma_1^2}.$$

Question 30 :

Since θ depends on μ_1, without loss of generality, we restrict the notation to μ_1 alone, i.e we now work with $s_{\mu_1}(\cdot|\mu_1)$. We compute the expectation of the score function, which is known to be 0. Indeed:

$$\begin{aligned}
\mathbb{E}_f[s_{\mu_1}(x|\mu_1)] &= \int_{\mathbb{R}} s_{\mu_1}(x|\mu_1) f(x|\mu_1)\, dx \\
&= \int_{\mathbb{R}} \frac{\partial}{\partial\mu_1} \log f(x|\mu_1) \cdot f(x|\mu_1)\, dx \\
&= \int_{\mathbb{R}} \frac{\frac{\partial}{\partial\mu_1} f(x|\mu_1)}{f(x|\mu_1)} \cdot f(x|\mu_1)\, dx \\
&= \int_{\mathbb{R}} \frac{\partial}{\partial\mu_1} f(x|\mu_1)\, dx \\
&= \frac{\partial}{\partial\mu_1} \int_{\mathbb{R}} f(x|\mu_1)\, dx \\
&= 0.
\end{aligned}$$

Indeed, by the differentiation theorem under the integral sign, since $f(\cdot|\mu_1)$ is continuously differentiable in μ_1 over $\mathbb{R}$ and $\frac{\partial}{\partial\mu_1} f(\cdot|\mu_1)$ is dominated by an integrable function, we can exchange the order of differentiation and integration.[2]

Thus, a control variate estimator is given, for some $b \in \mathbb{R}$, by:

$$\begin{aligned}\hat{\delta}_N^{(CV)} &= \frac{1}{N}\sum_{k=1}^{N}(\mathbf{1}_{(X_k>q)} - b \cdot s_{\mu_1}(X_k|\mu_1)) \\ &= \frac{1}{N}\sum_{k=1}^{N}\left(1_{(X_k>q)} - b \cdot \frac{f_1(X_k|\theta_1)}{f_1(X_k|\theta_1) - a \cdot f_2(X_k|\theta_2)} \cdot \frac{X_k - \mu_1}{\sigma_1^2}\right)\end{aligned}$$

where $(X_k)_{k\leq N} \overset{\text{law}}{=} f$ is an i.i.d. sample.

But :

$$\begin{aligned}\mathrm{Var}[\hat{\delta}_N^{(CV)}] &= \frac{1}{N}\left(\mathrm{Var}[\mathbf{1}_{(X>q)}] - 2b\,\mathrm{Cov}[\mathbf{1}_{(X>q)}; s_{\mu_1}(X|\mu_1)] + b^2\mathrm{Var}[s_{\mu_1}(X|\mu_1)]\right). \\ &= \mathrm{Var}[\overline{h}_N] + \frac{1}{N}\left(b^2\mathrm{Var}[s_{\mu_1}(X|\mu_1)] - 2b\,\mathrm{Cov}[\mathbf{1}_{(X>q)}, s_{\mu_1}(X|\mu_1)]\right).\end{aligned}$$

Minimizing this variance leads to the optimal choice:

$$b^* = \frac{\mathrm{Cov}[\mathbf{1}_{(X>q)}, s_{\mu_1}(X|\mu_1)]}{\mathrm{Var}[s_{\mu_1}(X|\mu_1)]}.$$

The covariance can be expressed as:

$$\begin{aligned}\mathrm{Cov}[\mathbf{1}_{(X>q)}, s_{\mu_1}(X|\mu_1)] &= \mathbb{E}_f[\mathbf{1}_{(X>q)} s_{\mu_1}(X|\mu_1)] \\ &= \int_q^{+\infty} s_{\mu_1}(x|\mu_1) f(x|\mu_1)dx \\ &= \int_q^{+\infty} \frac{\partial}{\partial\mu_1} f(x|\mu_1)dx \\ &= \frac{\partial}{\partial\mu_1}\int_q^{+\infty} f(x|\mu_1)dx, \text{ Again we switch partial differentiation and integration.} \\ &= \frac{\partial}{\partial\mu_1}\left(1 - F_X(q|\mu_1)\right), \text{ where } F_X \text{ is the cdf of } f(.|(\theta_1;\theta_2) \\ &= -\frac{\partial}{\partial\mu_1} F_X(q|\mu_1) \\ &= -\frac{\partial}{\partial\mu_1}\Psi\left(\frac{q-\mu_1}{\sigma_1}\right) \text{ Where } \Psi \text{ is the cdf of the standard normal law} \\ &= \frac{1}{\sigma_1}\psi\left(\frac{q-\mu_1}{\sigma_1}\right) \text{ Where } \psi \text{ is the density of the standard normal}\end{aligned}$$

Our final theoretical expression for b^* is:

$$b^* := \frac{\psi\left(\frac{q-\mu_1}{\sigma_1}\right)}{\sigma_1 \mathrm{Var}[s_{\mu_1}(X|\mu_1)]}$$

It remains to compute the denominator, where :

$$\begin{aligned}\mathrm{Var}[s_{\mu_1}(X|\mu_1)] &= \mathbb{E}[s_{\mu_1}^2(X|\mu_1)] - (\mathbb{E}[s_{\mu_1}(X|\mu_1)])^2 \\ &= \mathbb{E}[s_{\mu_1}^2(X|\mu_1)] \\ &= \int s_{\mu_1}(x)^2 f(x)\,dx \\ &= \frac{1}{1-a}\int_{-\infty}^{+\infty} \frac{f_1(x)^2}{f_1(x) - af_2(x)}\left(\frac{x-\mu_1}{\sigma_1^2}\right)^2 dx, \text{ A closed form may be difficult to obtain..}\end{aligned}$$

[2]Since $\int_{\mathbb{R}} f(x|\mu_1)\,dx = 1$ for all μ_1, differentiating this constant with respect to μ_1 gives 0.

Therefore, we will provide an estimator :

$$\hat{s}_N(\mu_1) := \frac{1}{N}\sum_{k=1}^{N} s^2(X_k|\mu_1) = \frac{1}{N}\sum_{k=1}^{N}(\frac{f_1(X_k)}{f_1(X_k) - a \cdot f_2(X_k)} \cdot \frac{X_k - \mu_1}{\sigma_1^2})^2 \xrightarrow[N\to\infty]{a.s} \mathbb{E}[s^2_{\mu_1}(X|\mu_1)]$$

where $(X_k)_{k\leq N} \overset{\text{law}}{=} f$ is an i.i.d. sample. Then :

$$\hat{b}^*_N := \frac{\psi\left(\frac{q-\mu_1}{\sigma_1}\right)}{\sigma_1 \hat{s}_N(\mu_1)} \xrightarrow[N\to\infty]{a.s} b^*$$

Using continuity of the inverse map.

We plug this estimator in our Control Variate one :

$$\begin{aligned}
\hat{\delta}_N^{(CV)} &= \frac{1}{N}\sum_{k=1}^{N}(\mathbf{1}_{(X_k>q)} - \hat{b}^*_N \cdot s_{\mu_1}(X_k|\mu_1)) \\
&= \frac{1}{N}\sum_{k=1}^{N}\left(\mathbf{1}_{(X_k>q)} - \frac{\psi\left(\frac{q-\mu_1}{\sigma_1}\right)}{\sigma_1 \hat{s}_N(\mu_1)} \cdot \frac{f_1(X_k|\theta_1)}{f_1(X_k|\theta_1) - a\cdot f_2(X_k|\theta_2)} \cdot \frac{X_k - \mu_1}{\sigma_1^2}\right) \\
&= \frac{1}{N}\sum_{k=1}^{N}\left(\mathbf{1}_{(X_k>q)} - \frac{\psi\left(\frac{q-\mu_1}{\sigma_1}\right)}{\sigma_1^3 \cdot (1-a)} \cdot \frac{f_1(X_k|\theta_1)}{f(X_k|(\theta_1;\theta_2))}\frac{(X_k-\mu_1)}{\hat{s}_N(\mu_1)}\right)
\end{aligned}$$

Question 31 :

Now we aim to use CLT , as $\hat{\delta}_N^{(CV)} \xrightarrow[N\to\infty]{a.s} \mathbb{E}[(\mathbf{1}_{(X_k>q)} - b^* \cdot s_{\mu_1}(X|\mu_1))] = \delta$, by continuity and using strong law of large numbers. Then :

$$\sqrt{N}\cdot\left(\hat{\delta}_N^{(CV)} - \delta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \mathrm{Var}(\mathbf{1}_{(X_k>q)} - b^* \cdot s_{\mu_1}(X|\mu_1))).$$

Yet as b^* is unknown, we provide an estimator of the variance :

$$\hat{V}_N^2 = \frac{1}{N-1}\sum_{k=1}^{N}\left(1_{\{X_k>q\}} - \hat{b}^*_N \cdot s_{\mu_1}(X_k|\mu_1) - \hat{\delta}_N^{(CV)}\right)^2$$

Which converges $a.s$ towards the variance, using continuity again and strong law of large numbers. Then, by continuity and Slutsky :

$$\frac{\sqrt{N}}{\hat{V}_N}\cdot\left(\hat{\delta}_N^{(CV)} - \delta\right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0,1).$$

Therefore we give the following confidence interval :

$$\left[\hat{\delta}_N^{(CV)} - q_{0.975}\cdot\frac{\hat{V}_N}{\sqrt{N}},\ \hat{\delta}_N^{(CV)} + q_{0.975}\cdot\frac{\hat{V}_N}{\sqrt{N}}\right]$$

where $q_{0.975} \approx 1.96$ is the 97.5% quantile of the standard gaussian distribution.

To achieve a precision $\varepsilon > 0$, we solve:

$$2\cdot q_{0.975}\cdot\frac{\hat{V}_N}{\sqrt{N}} \leq \varepsilon$$

which leads to:

$$N_\varepsilon := \left\lceil\left(\frac{2\cdot q_{0.975}\cdot\hat{V}_N}{\varepsilon}\right)^2\right\rceil$$

Question 32 :

For some required precision ε, an estimated variance σ^2 of the estimator derived from the 95% confidence interval :

$$\hat{\delta}_N \pm 1.96 \times \frac{\sigma}{\sqrt{N_\varepsilon}}$$

where N_ε is the sample size induced by the ε. The width of this interval is:

$$W := 2 \times 1.96 \times \frac{\sigma}{\sqrt{N_\varepsilon}}$$

Then :

$$\sigma^2 \approx N_\varepsilon \left(\frac{W}{2 \times 1.96}\right)^2$$

Computational Cost : The computational cost C_i for each method i is calculated as:

$$C_i = T_i \times \sigma_i^2$$

where T_i represents the execution time.

Table 1: Ranking of Monte Carlo Methods by Different Criteria, for values of ε for $q = 2$

Method	Rank (Time)	Rank (Variance)	Rank (Cost)	Rank (Closest to δ)
Naive	2	2	2	1
IS	3	3	3	2
CV	1	1	1	3

In almost any cases, the control variate method is the best one.