Projet de Monte Carlo P3

Samuel Elbaz, Sacha Assouly, Alexandre Zenou

22nd December 2024

Question 23

Let's determine a rejection algorithm which produces a sample that follows the distribution of the random variable X conditioned on the event $X \in A$.

Conditional density Let X be a random variable with a density $f_X(x)$ defined on a measurable space. The conditional density of X given $X \in A$, where A is a measurable subset of the space, is proportional to $f_X(x)$ restricted to A. Specifically:

$$f_{X|X\in A}(x) \propto f_X(x)\mathbf{1}_A(x),$$

This ensures that the conditional density $f_{X|X\in A}$ is obtained by restricting f_X to the set A, then normalizing it so that it integrates to 1.

Rejection algorithm principle The rejection algorithm proceeds as follows:

1. Generate a sequence of variables $X_1, X_2, \ldots$ i.i.d. according to the density $f_X(x)$.
2. Accept a sample X_i if and only if $X_i \in A$.
3. Collect the accepted samples as a new sequence $\{Y_k\}$, where Y_k is the k-th accepted sample.

proof: Each sample X_i is distributed according to $f_X(x)$. The rejection step ensures that only samples within A are retained. Since the density $f_{X|X\in A}$ is proportional to $f_X(x)$ restricted to A, the accepted samples $\{Y_k\}$ are distributed according to:

$$f_{X|X\in A}(x) \propto f_X(x)\mathbf{1}_A(x).$$

The normalization is automatically achieved because the algorithm accepts samples with probability proportional to their density under $f_X(x)$ within the set A.

Independence of accepted samples: The samples $X_1, X_2, \ldots$ are independent and identically distributed. Consequently, the accepted samples Y_k are also independent and identically distributed, following the conditional density $f_{X|X\in A}(x)$.

Conclusion The rejection algorithm is valid for simulating samples from the conditional distribution of X given $X \in A$, as it effectively restricts the samples to the set A and ensures that the resulting distribution is proportional to $f_X(x)$ restricted to A, with proper normalization. This normalization actually gives us $\mathbb{P}(X \in A)$.

Remark: With respect to the question 24, we assume that $\mathbb{P}(X \in A)$ is unknown. If it were known, we could have done a classical reject-algorithm to simulate a sample directly from the density $f_{X|X\in A}(x)$. This method is pretty naive and inefficient, especially for events of small probability. We will look at more effective methods in parts 2 and 3 of the project.

Question 26

Let $g(x)$ be the chosen instrumental density, a Cauchy distribution of parameters q and 1. Our choice is motivated by the fact the density g of a Cauchy is strictly positive for all reel numbers, therefore:

$$\text{supp}(h \cdot f) \subseteq \text{supp}(g)$$

Furthermore, the Cauchy density of median=q is q-symmetric, so we can use an antithetic approach "Z.antithetic = -Z + 2*q" and "1/(Z-q) + q" to reduce the variance of our estimator. The importance sampling estimator for $h(X) = \mathbf{1}_{\{X>q\}}$ is derived as follows:

The expectation of $h(X)$ under f can be written as:

$$\mathbb{E}_f[h(X)] = \int_{\mathbb{R}} h(x)f(x)\,dx = \int_q^\infty f(x)\,dx = \mathbb{P}(X > q).$$

Using importance sampling, we rewrite the integral in terms of the instrumental density $g(x)$:

$$\mathbb{E}_f[h(X)] = \int_q^\infty \frac{f(x)}{g(x)} g(x)\,dx.$$

The corresponding importance sampling estimator, based on n i.i.d. samples $Z_1, Z_2, \ldots, Z_n$ drawn from $g(x)$, is:

$$\hat{\mathbb{E}}_f[h(X)] = \frac{1}{n}\sum_{i=1}^n h(Z_i)\frac{f(Z_i)}{g(Z_i)} = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(Z_i > q)\frac{f(Z_i)}{g(Z_i)}.$$

Properties: The estimator $\hat{\mathbb{E}}_f[h(X)]$ is unbiased and converges to $\mathbb{E}_f[h(X)]$ as $n \to \infty$.

Question 27

The probability density function of a random variable following a Cauchy distribution with parameters x_0 (the location parameter, representing the median and mode) and $\gamma > 0$ (the scale parameter) is given by:

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

As mentioned in question 26, the choice of parameters q and 1 comes from the symmetric properties of the Cauchy distribution and the fact that it gives us bigger tails than the function f.

Question 29

The score is the gradient of the log-likelihood function with respect to the parameter vector: $\log \mathcal{L}(\Theta; x)$. In other worlds the score is :

$$s(\Theta, x) = \nabla_\Theta \log \mathcal{L}(\Theta; x)$$

The likelihood function for the data $x_1, \dots, x_n$ is given by:

$$\mathcal{L}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 \mid x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{1-a}\big(f_1(x_i) - a f_2(x_i)\big)$$

where:

$$f_1(x_i) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right)$$

$$f_2(x_i) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x_i-\mu_2)^2}{2\sigma_2^2}\right)$$

Taking the logarithm of the likelihood:

$$\log \mathcal{L}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 \mid x_1, \dots, x_n) = \sum_{i=1}^{n} \log\left[\frac{1}{1-a}\left(f_1(x_i) - a f_2(x_i)\right)\right]$$

This simplifies to:

$$-n\log(1-a) - n\log(\sqrt{2\pi}) + \sum_{i=1}^{n} \log\left(\frac{1}{\sigma_1}\exp\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right) - \frac{a}{\sigma_2}\exp\left(-\frac{(x_i-\mu_2)^2}{2\sigma_2^2}\right)\right)$$

We compute the derivative of the log-likelihood with respect to μ_1 :
Expanding the derivative:

$$\frac{\partial}{\partial \mu_1} f_1(x_i) = \frac{x_i - \mu_1}{\sigma_1^3} \exp\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right)$$

$$\frac{\partial}{\partial \mu_1} f_2(x_i) = 0 \quad \text{(since } f_2 \text{ depends on } \mu_2\text{, not } \mu_1\text{)}$$

$$\frac{\partial \log \mathcal{L}(x_1, \dots, x_n \mid \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)}{\partial \mu_1} = \sum_{i=1}^{n} \frac{\frac{(x_i-\mu_1)}{\sigma_1^3}\exp\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sigma_1}\exp\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right) - a\frac{1}{\sigma_2}\exp\left(-\frac{(x_i-\mu_2)^2}{2\sigma_2^2}\right)}$$

The score function for μ_1 is defined as:

$$s_{\mu_1}(x \mid \theta_1, \theta_2) = \sum_{i=1}^{n} \frac{\frac{(x_i-\mu_1)}{\sigma_1^3}\exp\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sigma_1}\exp\left(-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}\right) - a\frac{1}{\sigma_2}\exp\left(-\frac{(x_i-\mu_2)^2}{2\sigma_2^2}\right)}.$$

Question 30

Let us take $h_0 : x \mapsto \frac{\partial \log \mathcal{L}(x_1,\dots,x_n \mid \mu_1,\mu_2,\sigma_1^2,\sigma_2^2)}{\partial \mu}$.
Then a control variate Monte Carlo Estimator is :

$$\widehat{\delta}_n^{cv}(b) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \geq q} - b\big[h_0(X_i) - m\big].$$

Since the expectation of the score function is 0, we have $m = 0$.
Thus :
$$\widehat{\delta}_n^{cv}(b) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \geq q} - bh_0(X_i).$$

Question 32

1 Naive method

The computational cost of this method can be evaluated based on the number of samples n. Specifically, the variance of the naive estimator is given by:

$$\hat{\theta}_{\text{naive}} = \frac{1}{n}\sum_{i=1}^{n} f(X_i),$$

where $f(X_i)$ is an indicator function that equals 1 if $X_i \geq q$ and 0 otherwise. The variance of this estimator can be calculated as:
$$\text{Var}(\hat{\theta}_{\text{naive}}) = \frac{1}{n} \cdot \text{Var}(X_i).$$
The computational cost of this method is also evaluated using `microbenchmark`, and by calculating the median execution time for a large number of samples n, we obtain a cost $R_1 = 47758532$ (in time units per sample). This cost is based on the observed data from our experiment.

2 Importance Sampling Method

In our code, the function $f(x)$ is defined as the difference between two densities, and the estimator based on importance sampling is computed using Cauchy weights. This method involves a weighting distribution $g(x) = \text{Cauchy}(x, q, 1)$, and the estimator is given by:

$$\hat{\theta}_{\text{IS}} = \frac{1}{n}\sum_{i=1}^{n} \frac{f(Z_i)}{g(Z_i)} \mathbb{1}(Z_i \geq q),$$

where Z_i are samples generated from a Cauchy distribution. The computational cost of this method is also evaluated using `microbenchmark`, and we obtain a cost $R_2 = 532252.2$.

3 Control Variate Method

The control variate method relies on the use of an auxiliary variable that is correlated with the variable of interest but has a known expectation. This method reduces the variance of the estimator by adjusting the samples based on the control variable.

The computational cost of this method is also evaluated using `microbenchmark`, and we obtain a cost $R_3 = 401456$.

4 Comparison of Methods

We now compare the three methods in terms of computational cost and estimation accuracy. The cost R_1 of the naive method is compared to R_2 for importance sampling and R_3 for the control variate method. The following ratios are calculated to assess the relative efficiency of the methods:

$$\frac{R_1}{R_2} = \frac{47758532}{532252.2} \approx 89.73$$

This ratio shows that the importance sampling estimator is more efficient than the naive estimator in terms of computational cost, as $\frac{R_1}{R_2} > 1$.

$$\frac{R_2}{R_3} = \frac{532252.2}{401456} \approx 1.33$$

This ratio shows that the control variate method outperforms the importance sampling method in terms of computational cost, as $\frac{R_2}{R_3} > 1$.