Computational statistics - Homework 4

Monte Carlo methods

Matteo Silvestri

Professor Maarten Jansen

2024

Table des matières

1 Problem

Find a numerical value of var(X) where X has density $f_X(x) = Kf(x)$ with

$$f(x) = \exp(\cos(2\pi x))$$

on [0,1] (and zero outside the interval), using rejection sampling starting from the density $g_X(x) = Lg(x)$ where

$$g(x) = \frac{1 + 10\left|x - \frac{1}{2}\right|}{2}.$$

(a) Make a plot showing that $f_X(x)$ can be bounded by $M \cdot g_X(x)$ and find a value of M as a function of the constants K,L .

(b) Find the constant L .

(c) Find the CDF $G_X(x) = \int_0^x g_X(u)\,du$ and the corresponding quantile function $G_X^{-1}(p)$.

An alternative approach is to use the symmetry of $g_X(x)$ around $\frac{1}{2}$. So, if $V = \left|X - \frac{1}{2}\right|$, then the distribution and quantile functions of V have easier expressions than those of $G_X^{-1}(p)$. Because of symmetry, given a value of V, a value of X can be generated by $X = V + \frac{1}{2}$ with probability $\frac{1}{2}$ and $X = \frac{1}{2} - V$ with probability $\frac{1}{2}$.

(d) Develop the rejection criterion $U \geq \dfrac{f_X(X)}{Mg_X(X)}$ for $X = G_X^{-1}(V)$ with $V \sim \text{unif}([0,1])$ and $U \sim \text{unif}([0,1])$ in terms of $f(x)$ and $g(x)$.

(e) Implement the rejection sampling, for instance by completing the MATLAB routine randexp-cos.m (using randCauchyplusNormal.m as an example; both MATLAB files are available in the Github Repository). Use the sampler to generate pseudo-observations from which the variance can be estimated.

2 Solution

You can find the MATLAB routine that provides the results in the Github Repository.

(a) We want to find a constant M such that $f_X(x) \leq M g_X(x)$ in $[0,1]$. As we can see from the figure 2.1, it holds that $f(x) < g(x)$ in $[0,1]$. Thus, we obtain

$$f(x) = \frac{f_X(x)}{K} \leq g(x) = \frac{g_X(x)}{L}$$

so we can take $M = \frac{K}{L}$.

(b) The constant L is very easy to compute, since $\int_0^1 g(x)\,dx$ is the sum of the area of a rectangle plus two times the area of a triangle. Thus, we have

$$\int_0^1 g(x)\,dx = \frac{1}{L} = \left[1 \cdot \frac{1}{2}\right] + 2\left[\frac{1}{2} \cdot \left(3 - \frac{1}{2}\right) \cdot \frac{1}{2}\right] = \frac{1}{2} + \frac{5}{4} = \frac{7}{4}$$

so we get $L = \frac{4}{7}$.

(c) Since we have the explicit expression for $g_X(x)$, we can derive the CDF analytically. Indeed, we have

$$G_X(x) = \int_0^x g_X(u)\,du = \int_0^x L g(u)\,du = \int_0^x \frac{4}{7}\left(\frac{1}{2} + 5\left|u - \frac{1}{2}\right|\right)du$$

Let us therefore separate the two cases $x \in [0, \frac{1}{2}]$ and $x \in (\frac{1}{2}, 1]$. If $x \in [0, \frac{1}{2}]$, then we get

$$G_X(x) = \int_0^x \frac{2}{7} + \frac{20}{7}\left(\frac{1}{2} - u\right)du = \frac{12}{7}x - \frac{10}{7}x^2$$

while if $x \in (\frac{1}{2}, 1]$ we obtain

$$G_X(x) = G_X\left(\frac{1}{2}\right) + \int_{\frac{1}{2}}^x \frac{2}{7} + \frac{20}{7}\left(u - \frac{1}{2}\right)du = \frac{10}{7}x^2 - \frac{8}{7}x + \frac{5}{7}.$$

In figure 2.2, we represent the plot of this function. Note that it respects the classical properties of CDF for this particular case, i.e. $G_X(0) = 0\,, G_X\left(\frac{1}{2}\right) = \frac{1}{2}$ and $G_X(1) = 1$.

Now, we calculate the inverse function (i.e. the quantile function) in the classical way. Let's define $p = G_X(x)$; then, for $x \in [0, \frac{1}{2}]$ we have

$$\frac{12}{7}x - \frac{10}{7}x^2 = p \iff 10x^2 - 12x - 7p = 0$$

and the equation is verified by

$$x = \frac{6 \pm \sqrt{36 - 70p}}{10}\,.$$

How do we choose the sign? Just note that if $p = 0$ then the quantile also must be 0 and it should be the same for $p = \frac{1}{2}$. This only happens for the negative sign. If $x \in (\frac{1}{2}, 1]$, then

$$\frac{10}{7}x^2 - \frac{8}{7}x + \frac{5}{7} = p \iff 10x^2 - 8x + (5 - 7p) = 0$$

and the equation is verified by

$$x = \frac{4 \pm \sqrt{16 - 10(5 - 7p)}}{10}\,.$$

As before, we want to impose the conditions $x = \frac{1}{2}$ if $p = \frac{1}{2}$ and $x = 1$ if $p = 1$; then we choose the positive sign. We observe that this reflects the symmetry of the distribution, since we have obtained 1 minus the previous result changing p with (1-p). So to summarise, we have the following quantile function

$$G_X^{-1}(p) = Q(p) = \begin{cases} \frac{1}{10}\left(6 - \sqrt{36 - 70p}\right) & \text{if } x \in [0, \frac{1}{2}] \\ \frac{1}{10}\left(4 + \sqrt{70p - 34}\right) & \text{if } x \in [\frac{1}{2}, 1] \end{cases}$$

and we show its plot in the figure 2.3. Just to be sure, in figure 2.4 we plot the two possible compositions between $G_X(x)$ and $Q(p)$ in order to verify that the calculations are correct. If so, both compositions must correspond to the line $y = x$.

(d) Using the fact that $M = \dfrac{K}{L}$, $f_X(x) = Kf(x)$ and $g_X(x) = Lg(x)$, we obtain that the rejection criterion becomes

$$\text{reject if } U \geq \frac{Kf(X)}{\frac{K}{L}Lg(X)} = \frac{f(X)}{g(X)}\,.$$

(e) So now we can generate samples by rejection sampling method showed in point (d). In particular, we generate n=100000 samples with sample size m = 20 in order to estimate the parameters of the distribution. Notice that, given the symmetry of the distribution, the mean should be $\frac{1}{2}$. Thus, we estimate the variance in two ways : using the sampling variance with known $\mu = \frac{1}{2}$ (variance1 in the code), or using the sampling variance using the sampling mean (variance2 in the code). The following MATLAB code calculates both the mean and the variance of the target distribution.

```
%Parameters estimation
m=20; n=100000;
X = randexpcos(m,n);
mu = mean(mean(X));
%Variance estimation knowing the mean
deviations = X - (1/2);
variance_app1 = zeros(m,1);
for i = 1:m
    deviations_squared = deviations(i,:).^2;
    variance_app1(i,1) = sum(deviations_squared)/(n-1);
end
variance1 = mean(variance_app1);
%Variance estimation without knowning the mean
variance_app2 = var(X,1,2);
variance2 = mean(variance_app2);
```

In the following table, we represent the parameter estimates as the sample size changes.

Parameter	m = 1	m=20	m=200	m=500	m=1000
mean	0.4990	0.4995	0.5000	0.5000	0.5000
variance1	0.1130	0.1130	0.1131	0.1131	0.1131
variance2	0.1130	0.1130	0.1131	0.1131	0.1131

It is evident that the two variance estimates are identical, as the difference between the true mean and the sampling mean is almost zero.

Therefore, we can conclude that the variance is $\sigma^2 = 0.1131$.

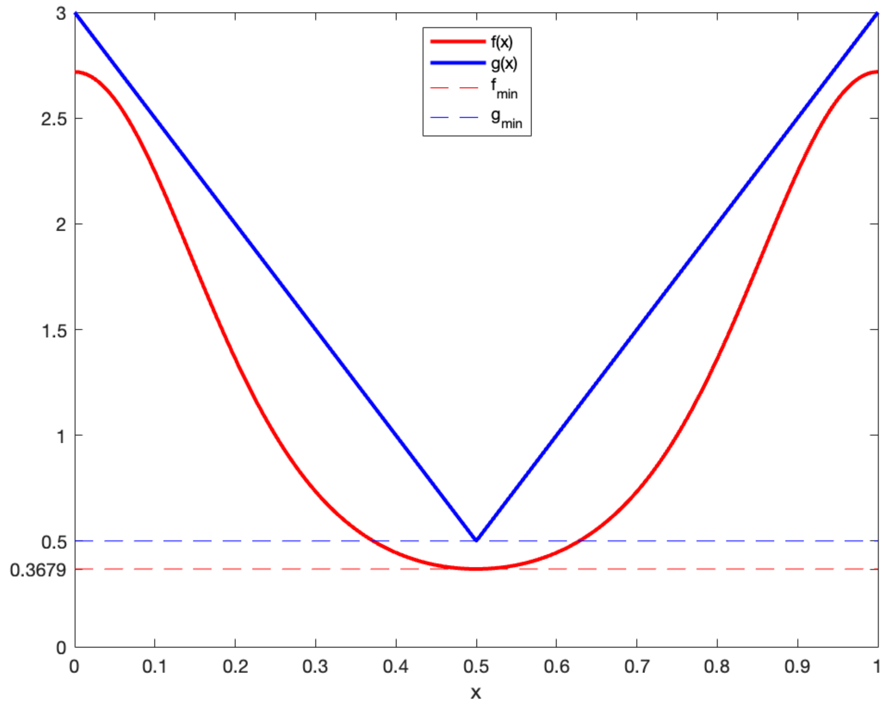

FIGURE 2.1 – Plot of f(x) and g(x) in the interval [0,1]

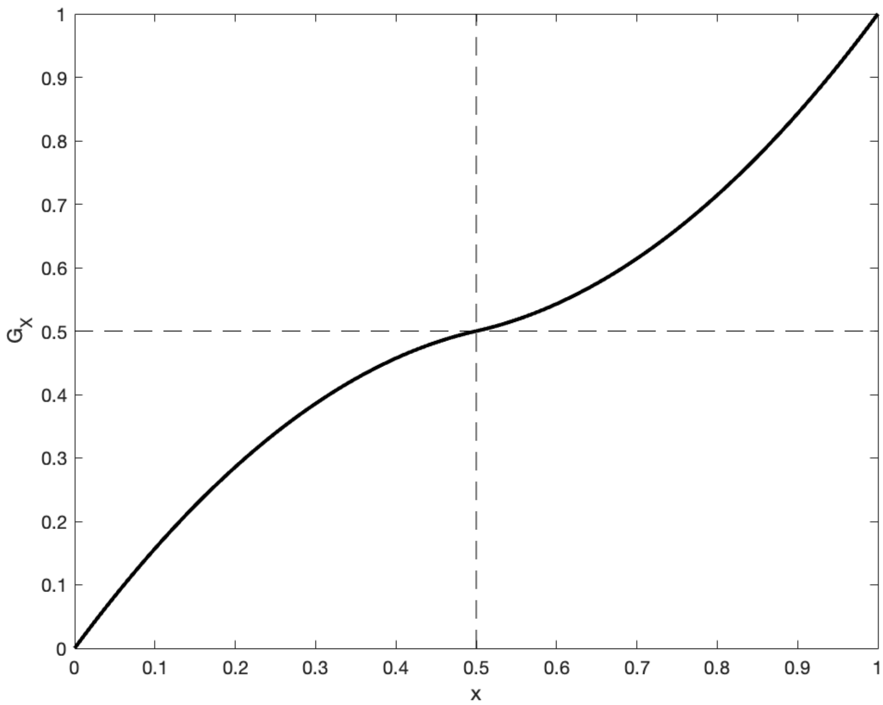

FIGURE 2.2 – Plot of Cumulative Distribution Function $G_X(x)$

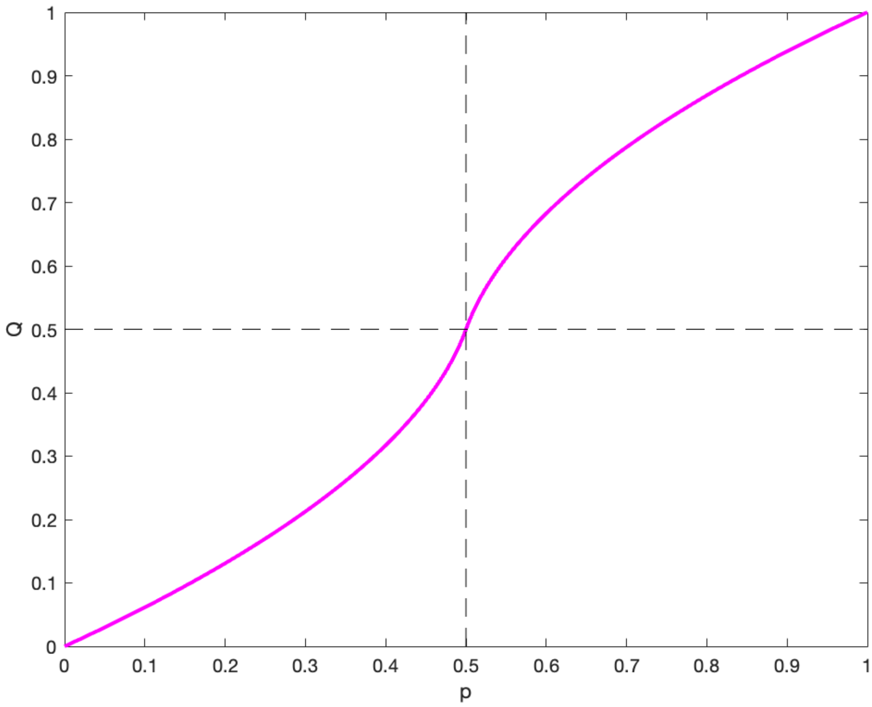

FIGURE 2.3 – Plot of the quantile function $G_X^{-1}(p) = Q(p)$

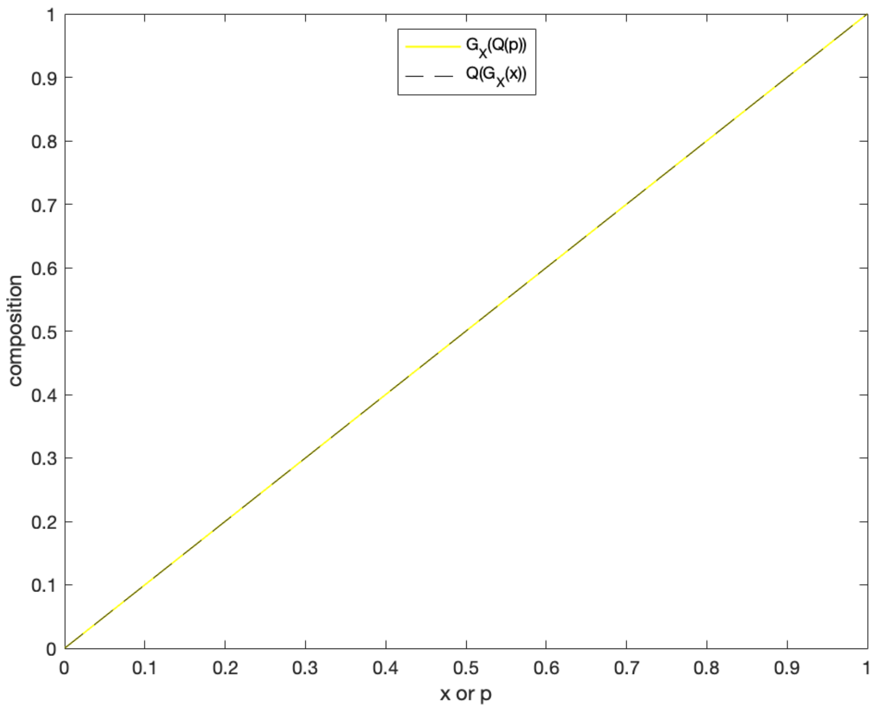

FIGURE 2.4 – Quantile function check