Computational statistics 2021.2

School of Applied Mathematics, Fundação Getulio Vargas
Professor Luiz Max de Carvalho

Problem sheet 1

Lucas Machado Moschen

Exercício 1. *(Inversion and Rejection)*

1. *Let $Y \sim \text{Exp}(\lambda)$ and let $a > 0$. We consider the variable after restricting its support to be $[a, +\infty)$. That is, let $X = Y_{|Y \geq a}$, i.e. X has the law of Y conditionally on being in $[a, +\infty)$. Calculate $F_X(x)$, the cumulative distribution function of X, and $F_X^{-1}(u)$, the quantil function of X. Describe an algorithm to simulate X from $U \sim \text{Unif}[0, 1]$.*

   If $x \geq a$, we have that

$$\begin{aligned} F_X(x) &= \mathbb{P}(Y \leq x \mid Y \geq a) \\ &= \frac{\mathbb{P}(Y \in [a, x])}{\mathbb{P}(Y \geq a)} \\ &= \frac{1 - e^{-\lambda x} - (1 - e^{-\lambda a})}{e^{-\lambda a}} \\ &= 1 - e^{-\lambda(x-a)}, \end{aligned}$$

   otherwise, $F_X(x) = 0$. Let $u = 1 - e^{-\lambda(x-a)}$. Inverting this function, we get that

$$F_X^{-1}(u) = a - \frac{\log(1-u)}{\lambda}.$$

   A simple algorithm is the following

   (i) Let $U \sim \text{Unif}[0, 1]$.

   (ii) Define $X = F_X^{-1}(U)$. Then X has the desired distribution by the inversion method.

2. *Let a and b be given, with $a < b$. Show that we can simulate $X = Y_{|a \leq Y \leq b}$ from $U \sim \text{Unif}[0, 1]$ using*

$$X = F_Y^{-1}(F_Y(a)(1-U) + F_Y(b)U),$$

   i.e. show that if X is given by the formula above, then $\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x \mid a \leq Y \leq b)$. Apply the formula to simulate an exponential random variable conditioned to be greater than a, as in the previous question.

   Using the properties of the (generalized) inverse and some affine transformations, note that

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(F_Y^{-1}(F_Y(a)(1-U) + F_Y(b)U) \leq x) \\ &= \mathbb{P}(F_Y(a)(1-U) + F_Y(b)U \leq F_Y(x)) \\ &= \mathbb{P}(U(F_Y(b) - F_Y(a)) \leq F_Y(x) - F_Y(a)) \\ &= \mathbb{P}\left(U \leq \frac{F_Y(x) - F_Y(a)}{F_Y(b) - F_Y(a)}\right) = \frac{F_Y(x) - F_Y(a)}{F_Y(b) - F_Y(a)}. \end{aligned}$$

However,

$$\frac{F_Y(x) - F_Y(a)}{F_Y(b) - F_Y(a)} = \frac{\mathbb{P}(Y \le x) - \mathbb{P}(Y \le a)}{\mathbb{P}(Y \le b) - \mathbb{P}(Y \le a)} = \mathbb{P}(Y \le x \mid Y \in [a, b]),$$

what concludes that X has the same distribution of $F_Y^{-1}(F_Y(a)(1-U) + F_Y(b)U)$. Taking $b = +\infty$, we can simulate $U \sim \text{Unif}[0,1]$ and use

$$X = F_Y^{-1}(F_Y(a)(1-U) + U).$$

3. *Here is a simple algorithm to simulate $X = Y_{|Y>a}$ for $Y \sim \text{Exp}(\lambda)$:*

   (a) *Let $Y \sim \text{Exp}(\lambda)$. Simulate $Y = y$.*

   (b) *If $Y > a$ then stop and return $X = y$, and otherwise, start again at step (a).*

   Show that this is just a rejection algorithm, by writing the proposal and target densities π and q, as well as the bound $M = \max_x \pi(x)/q(x)$. Calculate the expected number of trials to the first acceptance. Why is inversion to be preferred for $a \gg 1/\lambda$?

   The target density $\pi(x) = \frac{d}{dx}F_X(x) = \lambda e^{-\lambda(x-a)} 1_{\{x \ge a\}}$ is the density of X, while the proposal density is the exponential $q(x) = \lambda e^{-\lambda x} 1_{\{x \ge 0\}}$. Therefore, the bound is

$$M = \sup_{x \ge 0} \frac{\pi(x)}{q(x)} = \sup_{\{x \ge a\}} e^{\lambda a} = e^{\lambda a}.$$

   The probability of accepting $X = y$ is

$$\alpha(y) = \frac{\pi(y)}{Mq(y)} = \begin{cases} 0, & \text{if } y \le a \\ 1, & \text{if } y > a. \end{cases}.$$

   This is only the rejection sampling algorithm. Let N be the number os trials to the first acceptance. We already know that N is geometrically distributed with parameter $M^{-1} = e^{-\lambda a}$. In our case, this is easy to see, because,

$$\mathbb{P}(N > n) = \mathbb{P}(Y \le a)^n = (1 - e^{-\lambda a})^n.$$

   We conclude that $\mathbb{E}[N] = e^{\lambda a}$. When $a \gg 1/\lambda$, we have that $\mathbb{E}[N] \gg e$ and several trials are rejected until a desired sample come. In that case, is much simpler to use the inversion method.

Exercício 2. *(Rejection) Consider the following "squeeze" rejection algorithm for sampling from a distribution with density $\pi(x) = \tilde{\pi}(x)/Z_\pi$ on a state space $\mathbb{X}$ such that*

$$h(x) \le \tilde{\pi}(x) \le M\tilde{q}(x)$$

where h is a non-negative function, $M > 0$ and $q(x) = \tilde{q}(x)/Z_q$ is the density of a distribution that we can easily sample from. The algorithm proceeds as follows.

(a) Draw independently $X \sim q, U \sim \text{Unif}[0,1]$.

(b) Accept X if $U \le h(X)/(M\tilde{q}(X))$.

(c) If X was not accepted, draw an independent $V \sim \mathrm{Unif}[0,1]$ and accept X if

$$V \leq \frac{\tilde{\pi}(X) - h(X)}{M\tilde{q}(X) - h(X)}.$$

1. *Show that the probability of accepting a proposed $X = x$ in either step (b) or (c) is*

$$\frac{\tilde{\pi}(x)}{M\tilde{q}(x)}.$$

$$\begin{aligned}
&\mathbb{P}(\text{Accept } X \mid X = x) = \\
&= \mathbb{P}(U \leq h(x)/(M\tilde{q}(x))) + \mathbb{P}\left(U > h(x)/(M\tilde{q}(x))\right)\mathbb{P}\left(V \leq \frac{\tilde{\pi}(x) - h(x)}{M\tilde{q}(x) - h(x)}\right) \\
&= \int_0^{h(x)/(M\tilde{q}(x))} du + \left(1 - \int_0^{h(x)/(M\tilde{q}(x))} du\right)\int_0^{(\tilde{\pi}(x)-h(x))/(M\tilde{q}(x)-h(x)} q(x)\,du \\
&= \frac{h(x)}{M\tilde{q}(x)} + \left(1 - \frac{h(x)}{M\tilde{q}(x)}\right)\left(\frac{\tilde{\pi}(x) - h(x)}{M\tilde{q}(x) - h(x)}\right) \\
&= \frac{h(x)(M\tilde{q}(x) - h(x)) + M\tilde{q}(x)(\tilde{\pi}(x) - h(x)) - h(x)(\tilde{\pi}(x) - h(x))}{M\tilde{q}(x)(M\tilde{q}(x) - h(x))} \\
&= \frac{\tilde{\pi}(x)(M\tilde{q}(x) - h(x))}{M\tilde{q}(x)(M\tilde{q}(x) - h(x))} \\
&= \frac{\tilde{\pi}(x)}{M\tilde{q}(x)}.
\end{aligned}$$

2. *Deduce from the previous question that the distribution of the samples accepted by the above algorithm is π.*

   We know that given $X = x$, the probability of accepting X is $\dfrac{\pi(x)}{q(x)}\dfrac{Z_\pi}{MZ_q}$. Therefore,

$$\mathbb{P}(\text{Accept } X) = \int_{\mathbb{X}} \frac{\pi(x)}{q(x)}\frac{Z_\pi}{MZ_q} q(x)\,dx = \frac{Z_\pi}{MZ_q}$$

   what implies that, by Bayes Theorem, the density of $X = x$ given that X was accepted is

$$\frac{\pi(x)}{q(x)}\frac{Z_\pi}{MZ_q}\frac{q(x)}{Z_\pi/(MZ_q)} = \pi(x).$$

3. *Show that the probability that step (c) has to be carried out is*

$$1 - \frac{\int_{\mathbb{X}} h(x)\,dx}{MZ_q}$$

   This probability can be written as

$$\mathbb{P}\left(U > h(X)/(M\tilde{q}(X))\right) = 1 - \int_{\mathbb{X}} \frac{h(x)}{M\tilde{q}(x)} q(x)\,dx = 1 - \frac{\int_{\mathbb{X}} h(x)\,dx}{MZ_q}.$$

4. *Let* $\tilde{\pi}(x) = \exp(-x^2/2)$ *and* $\tilde{q}(x) = \exp(-|x|)$*. Using the fact that*

$$\tilde{\pi}(x) \geq 1 - \frac{x^2}{2}$$

for any $x \in \mathbb{R}$*, how could you use the squeeze rejection sampling algorithm to sample from* $\pi(x)$*. What is the probability of not having to evaluate* $\tilde{\pi}(x)$*? Why could it be beneficial to use this algorithm instead of the standard rejection sampling procedure?*

Define $h(x) = \max(1-x^2/2, 0)$. By the fact given above, $\tilde{\pi}(x) \geq h(x)$ for any $x \in \mathbb{R}$. Now, note that,

$$\sup_{x\in\mathbb{R}} \frac{\tilde{\pi}(x)}{\tilde{q}(x)} = \sup_{x\in\mathbb{R}} \exp(-x^2/2 + |x|).$$

In order to maximize the above expression, suppose $x < 0$ is a local extreme, then

$$(-x-1)e^{-x^2/2-x} = 0 \implies x = -1.$$

Suppose now that $x > 0$ is a local extreme, then

$$(-x+1)e^{-x^2/2-x} = 0 \implies x = 1.$$

So the global maximum is attained at $x = -1$, $x = 0$ or $x = 1$. We have that

$$\sqrt{e} = \exp(-(-1)^2/2 + |-1|) = \exp(-1^2/2 + |1|) > \exp(0) = 1.$$

Therefore, $\sup_{x\in\mathbb{R}} \tilde{\pi}(x)/\tilde{q}(x) = \sqrt{e}$. Then, we have that

$$h(x) \leq \tilde{\pi}(x) \leq \sqrt{e}\tilde{q}(x)$$

and we could use the squeeze rejection sampling algorithm.

The probability of not having to evaluate $\tilde{\pi}(x)$ is the probability of accepting X in step (b) that is

$$\int_{\mathbb{R}} \frac{h(x)}{M\tilde{q}(x)} q(x)\,dx = \frac{1}{\sqrt{e}Z_q}\int_{\mathbb{R}} h(x)\,dx = \frac{1}{\sqrt{e}Z_q}\int_{-\sqrt{2}}^{\sqrt{2}} 1 - \frac{x^2}{2}\,dx = \frac{2\sqrt{2}}{3\sqrt{e}} \approx 0.57,$$

since $Z_q = 2$. We have that calculating h is simpler than f and in half operations, we won't need to calculate it.

Exercício 3. *(Transformation) Consider the following algorithm known as Marsaglia's polar method.*

Step (a) *Generate independent* U_1, U_2 *uniformly in* $[-1, 1]$ *until* $Y = U_1^2 + U_2^2 \leq 1$.

Step (b) *Define* $Z = \sqrt{-2\log(Y)}$ *and return* $X_i = ZU_i/\sqrt{Y}$ *for* $i = 1, 2$.

1. *Define* $\vartheta = \arctan(U_2/U_1)$*. Show that the joint distribution of* Y *and* ϑ *has density*

$$f_{Y,\vartheta}(y,\theta) = 1_{[0,1]}(y)\frac{1_{[0,2\pi]}(\theta)}{2\pi}$$

Consider the transformation

$$g(u_1, u_2) = \left(u_1^2 + u_2^2, \arctan(u_2/u_1)\right).$$

The Jacobian of this transformation is

$$\begin{bmatrix} 2u_1 & 2u_2 \\ -u_2/(u_1^2+u_2^2) & u_1/(u_1^2+u_2^2) \end{bmatrix}$$

and its determinant is 2. Therefore, by the Change of Variable formula, since u_2/u_1 has image in $(-\infty, +\infty)$, we know that $\theta \in (-\pi/2, \pi/2)$. Besides that, is clear that $y \in [0, 1]$. Therefore,

$$f_{Y,\vartheta}(y,\theta) = \frac{1}{8\pi/4} 1_{\{\sqrt{y}(\cos(\theta),\sin(\theta)) \in [-1,1]^2\}} 1_{\{y \le 1\}} = \frac{1}{2\pi} 1_{[0,1]}(y) 1_{[0,2\pi]}(\theta).$$

2. *Show that X_1 and X_2 are independent standard normal random variables.*

   Putting $(U_1, U_2) = \sqrt{Y}(\cos(\vartheta), \sin(\vartheta))$, we have that

   $$X_1 = ZU_1/\sqrt{Y} = \sqrt{-2\log(Y)}\cos(\vartheta), \quad X_2 = \sqrt{-2\log(Y)}\sin(\vartheta).$$

   Then, (X_1, X_2) is a transformation of (Y, ϑ) which have uniform distribution over $[0, 1] \times [0, 2\pi]$. The Jacobian of this distribution is

   $$\begin{bmatrix} \frac{-2}{2y\sqrt{-2\log(y)}}\cos(\theta) & -\sqrt{-2\log(y)}\sin(\theta) \\ \frac{-2}{2y\sqrt{-2\log(y)}}\sin(\theta) & \sqrt{-2\log(y)}\cos(\theta), \end{bmatrix}$$

   whose determinant is

   $$\frac{-\cos^2(\theta) - \sin^2(\theta)}{y} = -\frac{1}{y}.$$

   Note that $X_1^2 + X_2^2 = -2\log(Y) \implies Y = \exp\left\{-\frac{1}{2}(X_1^2 + X_2^2)\right\}$ The density of the distribution of (X_1, X_2) is

   $$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2},$$

   which implies that $X_1, X_2 \overset{iid}{\sim} \text{Normal}(0, 1)$.

3. *What are the potential benefits of this approach over the Box-Muller algorithm?*

   The main beneficial part is that is not necessary to calculate any trigonometric functions, which are more expensive that logarithm.

Exercício 4.

Exercício 5 (Rejection and Importance Sampling)**.** *Consider two probability densities π, q on $\mathbb{X}$ such that $\pi(x) > 0 \implies q(x) > 0$ and assume that you can easily draw samples from q. Whenever $\pi(x)/q(x) \le M < \infty$ for any $x \in \mathbb{X}$, it is possible to use rejection sampling to sample from π. When M is unknown or when this condition is not satisfied, we can use importance sampling techniques to approximate expectations with respect to π. However it might be the case that most samples from q have very small importance weights.*

Rejection control is a method combining rejection and importance weighting. It relies on an arbitrary threshold value $c > 0$. We introduce the notation $w(x) = \pi(x)/q(x)$ and

$$Z_c = \int_{\mathbb{X}} \min\{1, w(x)/c\} q(x)\, dx.$$

Rejection control proceeds as follows.

- **Step a.** *Generate independent* $X \sim q$, $U \in \text{Unif}[0,1]$ *until* $U \leq \min\{1, w(X)/c\}$.
- **Step b.** *Return* X.

1. *Give the expression of the probability density* $q^*(x)$ *of the accepted samples.*

   Notice that if $A \subseteq \mathbb{X}$ is measurable,

$$\begin{aligned}\mathbb{P}(X \in A, X \text{ accepted}) &= \int_{\mathbb{X}} \int_0^{\min(1,w(x)/c)} 1_A(x) q(x)\, du\, dx \\ &= \int_{\mathbb{X}} 1_A(x) \min(1, w(x)/c) q(x)\, dx.\end{aligned}$$

   Besides that, $\mathbb{P}(X \text{ accepted}) = Z_c$ using the aboce expression with $A = \mathbb{X}$. Therefore, $q^*(x) = Z_c^{-1} \min(1, w(x)/c) q(x) = Z_c^{-1} \min(q(x), \pi(x)/c)$.

2. Prove that

$$\mathbb{E}_{q^*}\left([w^*(X)]^2\right) = Z_c \mathbb{E}_q(\max\{w(X), c\} w(X)),$$

   where $w^*(x) = \pi(x)/q^*(x)$.

   The left side of the equation is

$$I_1 = \int_{\mathbb{X}} w^*(x)^2 q^*(x)\, dx = Z_c \int_{\mathbb{X}} \frac{\pi(x)^2}{\min(q(x), \pi(x)/c)}\, dx,$$

   while the right side is

$$I_2 = Z_c \int_{\mathbb{X}} \max(w(x), c) w(x) q(x)\, dx = Z_c \int_{\mathbb{X}} \max\left(\frac{\pi^2(x)}{q(x)}, c\pi(x)\right) dx$$

   Define $X_1 = \{x \in \mathbb{X} \mid cq(x) \leq \pi(x)\}$ and $X_2 = \mathbb{X}/X_1$. Then

$$I_1 = Z_c \int_{X_1} \frac{\pi(x)^2}{q(x)}\, dx + Z_c \int_{X_2} c\pi(x)\, dx, ,$$

   and

$$I_2 = Z_c \int_{X_1} \frac{\pi^2(x)}{q(x)}\, dx + Z_c \int_{X_2} c\pi(x)\, dx,$$

   which implies that $I_1 = I_2$ as claimed.

3. Establish that

$$\mathbb{E}_q(\min\{w(X), c\})\mathbb{E}_q(\max\{w(X), c\} w(x)) \leq \mathbb{E}_q(\min\{w(X), c\} \max\{w(X), c\} w(X))$$

   First, let's prove that

$$h(w_1, w_2) = [\min\{w_1, c\} - \min\{w_2, c\}][w_1 \max\{w_1, c\} - w_2 \max\{w_2, c\}] \geq 0$$

   There are three cases:

   (i) $w_1, w_2 \leq c$: In this case, $h(w_1, w_2) = (w_1 - w_2)c(w_1 - w_2) = c(w_1 - w_2) \geq 0$.

   (ii) $w_1, w_2 \geq c$: In this case the first factor is zero and $h(w_1, w_2) = 0$.

(iii) $w_1 < c < w_2$: In this case, $h(w_1, w_2) = (w_1 - c)(cw_1 - w_2^2) = (c - w_1)(w_2^2 - cw_1) > 0$, given that $w_2^2 > c^2 > cw_1$, supposing $w_1 \geq 0$. Notice that $w_2 < c < w_1$ is analogous.

Alongside this result, we see that for every realization of X x_1, x_2, we have that $h(w(x_1), w(x_2)) \geq 0$, which implies that the random variables $\min(w(X), c)$ and $\max(w(X), c)w(X)$ are positively correlated. The claimed result follows.

4. Deduce from the results established in (2) and (3) that

$$\mathrm{Var}_{q^*}(w^*(X)) \leq \mathrm{Var}_q(w(X))$$

First, notice that $\mathbb{E}_{q^*}(w^*(X)) = \int_{\mathbb{X}} \pi(x)\, dx = 1$ and $\mathbb{E}_q(w(X)) = \int_{\mathbb{X}} \pi(x)\, dx = 1$

We have that

$$\begin{aligned}
c\,(1 + \mathrm{Var}_{q^*}(w^*(X))) &= c\mathbb{E}_{q^*}\left([w^*(X)]^2\right) \\
&= cZ_c\mathbb{E}_q(\max\{w(X), c\}w(X))s \quad (2) \\
&= c\mathbb{E}_q(\min(1, w(X)/c))\mathbb{E}_q(\max\{w(X), c\}w(X)) \\
&= \mathbb{E}_q(\min(c, w(X)))\mathbb{E}_q(\max\{w(X), c\}w(X)) \\
&\leq \mathbb{E}_q\left(\min(c, w(X))\max(w(X), c)w(X)\right) \quad (3) \\
&= \int_{\mathbb{X}} \min(c, w(x))\max(c, w(x))w(x)q(x)\, dx \\
&= c\int_{\mathbb{X}} w^2(x)q(x)\, dx \\
&= c\mathbb{E}_q(w(X)^2) \\
&= c\,(1 + \mathrm{Var}_q(w(X)))\,,
\end{aligned}$$

what implies the desired result.

Exercício 6.