Dauphine | PSL UNIVERSITÉ PARIS

MIDO - Master 1 Mathématique & Applications, 2022–2023

Méthodes de Monte Carlo

Final Exam - 10/01/2023

3H00 – DOCUMENTS AND CALCULATOR NOT ALLOWED

Problem 1

...... /**9**

Let f and g be densities with respect to the Lebesgue measure on $\mathbb{R}$ such that

$$\forall x \in \mathbb{R}, \quad f(x) = c\tilde{f}(x) \quad \text{and} \quad g(x) = d\tilde{g}(x),$$

where both positive functions $\tilde{f}$ and $\tilde{g}$ are known and computable, and both constants c and d are unknown. **For questions 1. to 4.**, we consider the special case of an interval $]a,b[\subset \mathbb{R}$ such that

$$\forall x \notin]a,b[, \quad \tilde{f}(x) = 0 \quad \text{and} \quad \sup_{x\in\mathbb{R}} \tilde{f}(x) = M < \infty.$$

1. Considering $U = (U_1, U_2)$ a uniform random point on $\mathscr{R} =]a,b[\times]0,M[$, compute the probability $\mathbb{P}\left[U_2 \leq \tilde{f}(U_1)\right]$. Given an *i.i.d.* sequence $U^1,\dots,U^n$, $n \in \mathbb{N}^*$, of uniform random points on $\mathscr{R}$, deduce a converging estimator of c, $\hat{c}_n$, and justify the convergence of $\hat{c}_n$ in n.

On a

$$\mathbb{P}\left[U_2 \leq \tilde{f}(U_1)\right] = \frac{1}{M(b-a)} \int_a^b \int_0^{\tilde{f}(u_1)} \mathrm{d}u_2 \mathrm{d}u_1 = \frac{1}{M(b-a)} \int_a^b \tilde{f}(u_1)\mathrm{d}u_1 = \frac{1}{cM(b-a)}.$$

2. Is this estimator $\hat{c}_n$ unbiased?

3. Let `df(x)` be a vectorised R function that computes $\tilde{f}(x)$. Write an R code that computes $\hat{c}_n$.

From now on, it is no longer assumed that f is bounded and has bounded support. Nevertheless, we assume from now on that $\{x \in \mathbb{R} \mid f(x) = 0\} \subseteq \{x \in \mathbb{R} \mid g(x) = 0\}$.

4. Let X be a random variable with density f. Show that $\mathbb{E}\left[\tilde{g}(X) / \tilde{f}(X)\right] = c/d$.

5. Given an *i.i.d.* sequence $X_1,\dots,X_n$ of random variables with density f, deduce from question **4.** a converging estimator of c/d.

6. Let $\alpha(\cdot)$ be a positive function on $\mathbb{R}$ such that

$$\int_{\mathbb{R}} \alpha(x)\tilde{f}(x)\tilde{g}(x)\mathrm{d}x < +\infty.$$

Show that if X is a random variable with density f and Y a random variable with density g, then

$$\mathbb{E}\left[\alpha(X)\tilde{g}(X)\right] \Big/ \mathbb{E}\left[\alpha(Y)\tilde{f}(Y)\right] = c \big/ d.$$

7. Deduce from the previous question a converging estimator of c/d based on two sequences $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ of *i.i.d.* random variables with density f and g, respectively. Justify the convergence of this estimator and provide a corresponding R function `ratiof(n)`.

We now assume that d **is known.** For an arbitrary $\omega > 0$, we further consider the **special case** when the auxiliary target density $h(x) \propto \tilde{f}(x) + \omega g(x)$ can be simulated, **even though its normalising constant is unknown**, that is, there exists an R function mixt(N) that returns N i.i.d. realisations with density $h(\cdot)$.

8. Show that a sample from $f(\cdot)$ can be extracted as a random subsample of an existing N-sample from $h(\cdot)$—for instance, produced as mixt(N)—. What is the expected size of this subsample as a function of N?

9. Construct a valid algorithm that partitions an N-sample from $h(\cdot)$ into (i) a sample from $g(\cdot)$ and (ii) a sample from $f(\cdot)$. Deduce a converging estimator of the constant c.

Problem 2

...... /**7.5**

Let $0 < \alpha < 1$ be the shape parameter of the Gamma distribution $\text{Ga}(\alpha, 1)$, with density

$$f(x; \alpha) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-x\} \qquad x > 0$$

The goal is simulate from this distribution using a Generalized Exponential distribution $\text{GE}(\alpha, \lambda)$, with density

$$g(x; \alpha, \lambda) = \frac{\alpha}{\lambda} \left(1 - e^{-x/\lambda}\right)^{\alpha-1} e^{-x/\lambda} \qquad \lambda, x > 0$$

We aim at sampling from f using the accept-reject algorithm with $g(\cdot; \alpha, \lambda)$ as proposal density.

1. Provide the CDF attached to $g(\cdot; \alpha, \lambda)$ and deduce the normalizing constant of $g(\cdot; \alpha, \lambda)$ is correct.

2. Deduce a practical way to generate a random variable with density $g(\cdot; \alpha, \lambda)$.

3. Show that

$$f(x; \alpha) = \frac{1}{\Gamma(\alpha+1)} R(x) g(x; \alpha, 1) \quad x > 0 \tag{1}$$

with

$$R(x) = \left(\frac{x}{1 - e^{-x}}\right)^{\alpha-1} \qquad x > 0$$

and establish that $0 < R(x) \leq 1$ for $x \geq 0$.

4. Construct an accept-reject algorithm to simulate $f(\cdot; \alpha)$ using $g(\cdot; \alpha, 1)$ by providing the acceptance bound on the uniform variate. Indicate the expected number of proposals needed to accept one realisation.

5. Write an executable R code of this algorithm as an R function zeini(N,alpha) with inputs N, the number of simulations, and alpha, the $\text{Ga}(\alpha, 1)$ shape parameter.

6. Since $R(\cdot)$ satisfies *(no proof required!)*

$$\frac{4-(1-\alpha)x}{4+(1-\alpha)x} \le R(x) \le \frac{4+\alpha x}{4-\alpha x}$$

deduce a faster accept-reject algorithm and write a corresponding R function squezze(N,alpha).

7. Since the choice $\lambda = 1$ made above in the proposal is arbitrary, other values of λ could lead to a higher efficiency. Give a precise mathematical meaning to "higher efficiency" and describe how you would run a Monte Carlo experiment to compare the choices $\lambda = 1/2$ and $\lambda = 2$. (**Bonus:** Write the associated R code.)

Problem 3

...... /7

We revisit the simulation of upper truncated Normal $\text{N}^+(a)$ random variables, $a > 0$, with density

$$f(x; a) \propto \exp\{-x^2/2\}\mathbb{I}_{(a,\infty)}(x) \qquad x \in \mathbb{R}$$

where the proportionality symbol is applying to both sides as functions of x.

1. Recall here (i) the exact value of the normalizing constant of $f(\cdot; a)$ and (ii) a standard accept-reject algorithm based on an Exponential proposal translated by a, **as seen in class**.

2. Derive the density of $Z = X^2$ when $X \sim \text{N}^+(a)$ and show that an acceptance-rejection simulation of Z is possible when using as proposal an Exponential $\text{E}(1/2)$ random variable translated by a^2.

3. Deduce an acceptance-rejection algorithm for the simulation of $X \sim \text{N}^+(a)$ and provide an associated R function marsa(N,a) with inputs N, the number of simulations, and a, the $\text{N}^+(a)$ truncation parameter. (**Bonus:** Write a version with no `for`, no `while` and no `repeat` loop.)

4. What is the average acceptance probability $\varrho(a)$ for this algorithm? Given the following asymptotic approximation (when a goes to ∞) of the Normal cdf

$$\Phi(-a) \approx e^{-a^2/2}(a^{-1} - a^{-3})$$

give an asymptotic approximation of $\varrho(a)$.

5. An R experiment on the respective performances of both marsa and truncnorm (the standard algorithm) returns the following execution times:

```
function user system total
truncnorm 0.329 0.011 0.341
function user system total
mars 0.150 0.024 0.174
```

What is the conclusion of this comparison?