STATISTICS 4CI3/6CI3

Winter 2019 Dr. Angelo Canty

TERM TEST Solutions

Note: I decided to mark each question out of 20 instead of 10 which was specified on the exam paper to avoid half-marks. Each question is still equally weighted and the test is still worth the same amount in the calculation of your final grade!

Q. 1 First we need the probability mass function and cumulative distribution function of the random variable X that we wish to generate. From the equation given we have

x	1	2	3	4
$f(x)$	0.30	0.05	0.20	0.45
$F(x)$	0.30	0.35	0.55	1.00

[6 marks]

A straightforward implementation of the inversion method is then

1. Generate $U \sim \text{Uniform}(0,1)$.

2. If $U < 0.30$ Return X=1
Else If $U < 0.35$ Return X=2
Else If $U < 0.55$ Return X=3
Else Return X=4

[10 marks]

An alternative algorithm that is more efficient (not required) is

1. Generate $U \sim \text{Uniform}(0,1)$.

2. If $U < 0.45$ Return X=4
Else If $U < 0.75$ Return X=1
Else If $U < 0.95$ Return X=3
Else Return X=2

Full marks were given for either of these algorithms (or any other equivalent ones).

Note that the two algorithms will not give the same sequence of random numbers. Below I give the results for both algorithms (X_1 comes from the first and X_2 comes from the second) but, of course, you only needed to give the results for your algorithm.

U	0.5197	0.1790	0.9994	0.2873	0.7294	0.5791	0.0361	0.3281	0.2026	0.8213
X_1	3	1	4	1	4	4	1	2	1	4
X_2	1	4	2	4	1	1	4	4	4	3

[4 marks]

Q. 2 This question requires first translating the `R` code into mathematics and then proving the final result.

From the first two lines we get two independent vectors of length n

$$\begin{aligned} U_{11}, \ldots, U_{1n} &\overset{iid}{\sim} \text{Uniform}(0,1) \\ U_{21}, \ldots, U_{2n} &\overset{iid}{\sim} \text{Uniform}(0,1) \end{aligned}$$

[2 marks]

From the next line we have $X_i \;=\; -\log(U_{1i}) \quad i = 1, \ldots, n$.
Since the U_{1i} are *iid* so are the X_i and the cumulative distribution function of X_i is given by

$$\begin{aligned} F_{X_i}(x) &= \mathrm{P}(X_i \leqslant x) \\ &= \mathrm{P}(-\log(U_{1i}) \leqslant x) \\ &= \mathrm{P}\left(U_{1i} \geqslant \mathrm{e}^{-x}\right) \\ &= \begin{cases} 1 - \mathrm{e}^{-x} & x > 0 \\ 0 & x \leqslant 0 \end{cases} \end{aligned}$$

We may recognise this as the cdf for the exponential(1) distribution. **[6 marks]**

The fourth line defines the *iid* vector $Y_1, \ldots, Y_n$ as

$$Y_i \;=\; \begin{cases} X_i & \text{if } U_{2i} < 0.5 \\ -X_i & \text{if } U_{2i} \geqslant 0.5 \end{cases} \qquad i = 1, \ldots, n.$$

Now we get the cdf of Y_i as follows.

$$\begin{aligned} F_{Y_i}(y) \;=\; \mathrm{P}(Y_i \leqslant y) &= \mathrm{P}(Y_i \leqslant y \mid U_{2i} < 0.5)\mathrm{P}(U_{2i} < 0.5) + \mathrm{P}(Y_i \leqslant y \mid U_{2i} \geqslant 0.5)\mathrm{P}(U_{2i} \geqslant 0.5) \\ &= 0.5\mathrm{P}(X_i \leqslant y) + 0.5\mathrm{P}(-X_i \leqslant y) \\ &= 0.5\mathrm{P}(X_i \leqslant y) + 0.5\mathrm{P}(X_i \geqslant -y) \end{aligned}$$

[3 marks]

It is easiest to consider the two cases of $y < 0$ and $y > 0$ separately. First we consider $y < 0$ in which case we have

$$\mathrm{P}(X_i \leqslant y) \;=\; 0 \quad \text{and} \quad \mathrm{P}(X_i \geqslant -y) \;=\; 1 - F_{X_i}(-y_i) \;=\; \mathrm{e}^{y}$$

and so the cdf for Y_i at a point $y < 0$ is

$$F_{Y_i}(y) \;=\; \mathrm{e}^{y} \qquad \text{for } y < 0$$

[3 marks]

Next for $y \geqslant 0$ we have

$$\mathrm{P}(X_i \leqslant y) \;=\; F_{X_i}(y) \;=\; 1 - \mathrm{e}^{-y} \quad \text{and} \quad \mathrm{P}(X_i \geqslant -y) \;=\; 1$$

Hence the cdf for Y at a point $y > 0$ is

$$F_{Y_i}(y) \;=\; 0.5\left(1 - \mathrm{e}^{-y}\right) + 0.5 \;=\; 1 - 0.5\mathrm{e}^{-y} \qquad \text{for } y < 0$$

[3 marks]

Putting these together we have

$$F_{Y_i}(y) \;=\; \begin{cases} \mathrm{e}^{y} & y < 0 \\ 1 - \mathrm{e}^{-y} & Y \geqslant 0 \end{cases}$$

Taking derivatives to get the probability density function for Y_i we have

$$\begin{aligned} g(y) &= \frac{dF_{Y_i}(y)}{dy} \\ &= \begin{cases} 0.5\mathrm{e}^{y} & y < 0 \\ 0.5\mathrm{e}^{-y} & y \geqslant 0 \end{cases} \\ &= 0.5\mathrm{e}^{-|y|} \quad y \in \mathbb{R} \end{aligned}$$

and since the vector of $Y_1, \ldots, Y_n$ is the returned value we see that this function returns an independent sample of size n from the standard Laplace distribution. **[3 marks]**

Q. 3 First we note that

$$x^2 - 2|x| \geqslant -1 \quad \Longleftrightarrow \quad x^2 - 2|x| + 1 \geqslant 0 \quad \Longleftrightarrow \quad (|x| - 1)^2 \geqslant 0$$

and we note that $(|x| - 1)^2 > 0$ unless $|x| = 1$ in which case we have $(|x| - 1)^2 = 0$ so we have proven that $x^2 - 2|x| \geqslant -1$. **[4 marks]**

Now the ratio of densities is

$$\begin{aligned}
\frac{f(x)}{g(x)} &= \frac{\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}}{\frac{1}{2}\mathrm{e}^{-|x|}} \\
&= \sqrt{\frac{2}{\pi}}\exp\left\{-\frac{x^2}{2} + |x|\right\} \\
&= \sqrt{\frac{2}{\pi}}\exp\left\{-\frac{1}{2}\left(x^2 - 2|x|\right)\right\} \\
&\leqslant \sqrt{\frac{2}{\pi}}\exp\left\{\frac{1}{2}\right\} \qquad \text{because } -\frac{1}{2}\left(x^2 - 2|x|\right) \leqslant \frac{1}{2} \text{ from above} \\
&= \sqrt{\frac{2\mathrm{e}}{\pi}}
\end{aligned}$$

[8 marks]

Based on this bound $M = \sqrt{\frac{2\mathrm{e}}{\pi}}$ we see that the acceptance probability for any candidate Y is

$$\mathrm{P(Accept)} = \frac{f(Y)}{Mg(Y} = \frac{\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{1}{2}\left(x^2 - 2|x|\right)\right\}}{\sqrt{\frac{2\mathrm{e}}{\pi}}} = \exp\left\{-\frac{1}{2}\left(|Y| - 1\right)^2\right\}$$

Hence an algorithm to generate standard normals using the standard Laplace candidate density is

1. Generate $Y \sim$ Standard Laplace using code such as given in Question 2.

2. Generate $U \sim$ Uniform$(0, 1)$ independent of Y.

3. If $U \leqslant \exp\left\{-\frac{1}{2}\left(|Y| - 1\right)^2\right\}$ then return Y
  Otherwise discard Y and U and return to Step 1.

[8 marks]

Q. 4 Consider the change of variables $u = e^{-x^2}$ in the integration. Then we have $du = -2xe^{-x^2}$ and $x = \sqrt{-\log(u)}$. **[2 marks]**

We get the limits of integration by noting that

$$x = 0 \;\Rightarrow\; u = 1 \quad \text{and} \quad x \to \infty \;\Rightarrow\; u \to 0$$

[2 marks]

Hence the integral becomes

$$\begin{aligned} I &= \int_0^\infty x^2 e^{-x^2}\, dx = \int_0^\infty \left(-\frac{x}{2}\right) \times \left(-2xe^{x^2}\, dx\right) \\ &= \int_1^0 -\frac{\sqrt{-\log(u)}}{2}\, du \\ &= \int_0^1 \frac{\sqrt{-\log(u)}}{2}\, du \end{aligned}$$

[5 marks]

We recognise this last expression as an expected value for the Uniform(0,1) random variable so we have $I = 0.5\,\mathrm{E}(\sqrt{-\log(U)})$ where $U \sim \text{Uniform}(0,1)$. **[1 mark]**

Hence the Monte Carlo estimator is

$$\hat{I} = \frac{1}{2N}\sum_{i=1}^{N}\sqrt{-\log(U_i)} \qquad \text{where } U_1,\ldots,U_N \overset{iid}{\sim} \text{Uniform}(0,1)$$

[3 marks]

Now we must derive the standard error of the estimator

$$\begin{aligned} \mathrm{Var}\left(\hat{I}\right) &= \frac{1}{4N}\mathrm{Var}\left(\sqrt{-\log(U_1)}\right) \\ &= \frac{1}{4N}\left[\mathrm{E}\left(-\log(U_1)\right) - \left(\mathrm{E}(\sqrt{-\log(U_1)})\right)^2\right] \\ &= \frac{1}{N}\left[\frac{1}{4}\mathrm{E}\left(-\log(U_1)\right) - I^2\right] \\ &= \frac{1}{N}\left(\frac{1}{4} - I^2\right) \end{aligned}$$

The last expression coming because we showed in Question 2 that

$$U \sim \text{Uniform}(0,1) \;\Rightarrow\; -\log(U) \sim \text{exponential}(1) \;\Rightarrow\; \mathrm{E}\left[-\log(U)\right] = 1$$

[5 marks]

Hence the standard error is

$$\mathrm{se}(\hat{I}) = \sqrt{\frac{1-4\hat{I}^2}{4N}}$$

[2 marks]

Q. 5 Here is a reasonable simulation study for this case

1. Decide on d values of $\alpha \geqslant 1$ that will be used. The set **must** include the null case $\alpha = 1$. Let us take $\alpha \in \mathcal{A} = \{1, 1.2, 1.4, 1.5, 1.6, 1.8, 2, 2.5, 3, 4\}$ **[2 marks]**

2. Decide on an appropriate number of simulations such as $N = 10000$. **[2 marks]**

3. For each value of $\alpha_j \in \mathcal{A}$

3.1 Generate N samples each of size $n = 20$ from a gamma distribution with parameters α_j and 1. **[3 marks]**

3.2 Calcualte the N sample medians $\tilde{x}_{j1}, \tilde{x}_{j2}, \ldots, \tilde{x}_j N$. **[3 marks]**

3.3 Calculate to Monte Carlo power estimate

$$\hat{\pi}_j \;=\; \frac{1}{N} \sum_{k=1}^{N} I \left\{ \tilde{x}_{jk} > 1 + \frac{10}{n} = 1.5 \right\}$$

[4 marks]

3.4 Calculate the standard error of the estimate

$$\mathrm{se}\left(\hat{\pi}_j\right) \;=\; \sqrt{\frac{\hat{\pi}_j(1 - \hat{\pi}_j)}{n}}$$

[4 marks]

4. Return the d triples $(\alpha_j, \hat{\pi}_j, \mathrm{se}(\hat{\pi}_j)) \quad j = 1, \ldots, d$ **[2 marks]**