MTH210 2024: Midsem Exam

Instructions:

(a) Write your name and roll number on all supplemental sheets.

(b) Marks will not be given unless all work is shown. Show **all** mathematical details.

Useful densities:

- Gamma(α, β): $f(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, $x > 0$
- Exponential(λ): $f(x) = \lambda e^{-\lambda x}$, $x > 0$.
- Student's t_5-distribution $f(x) = \dfrac{\Gamma(3)}{\sqrt{5\pi}\Gamma(2.5)} \left(1 + \dfrac{x^2}{5}\right)^{-3}, \qquad x \in \mathbb{R}$.

Questions:

1. (20 points) Let $(X, Y) \sim F$, where distribution F is the uniform distribution over a circle of radius 1, centered at the origin. That is, the density of F is

$$f(x, y) = \frac{1}{\pi} \mathbb{I}\left(x^2 + y^2 < 1\right) .$$

   (a) (10 points) Let R and Θ denote the polar coordinates of (X, Y): $X = R\cos\Theta$ and $Y = R\sin\Theta$. Find the joint pdf of R and Θ.

   (b) (10 points) Using the answer to (a) and only using draws from $U(0, 1)$, write an algorithm to obtain a sample from F.

Solution:

The joint density of (X, Y) is

$$f(x, y) = \frac{1}{\pi} \mathbb{I}\left(x^2 + y^2 < 1\right) .$$

With the polar coordinate transformation, $X = R\cos\Theta$ and $Y = R\sin\Theta$. We have to find the density of (R, Θ). Let (r, θ) denote a realization from this distribution. For this, the Jacobian is

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \\ \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \end{vmatrix} = \begin{vmatrix} -r\sin\theta & r\cos\theta \\ \cos\theta & \sin\theta \end{vmatrix} = r .$$

We obtain

$$\begin{aligned} f_{(R,\theta)}(r, \theta) &= |J| f(r, \theta) \\ &= \frac{d}{\pi} \mathbb{I}(d < 1) \mathbb{I}(0 < \theta < 2\pi) \end{aligned}$$

$$= (2d\mathbb{I}(d < 1)) \, \frac{1}{2\pi}\mathbb{I}(0 < \theta < 2\pi)$$
$$= f_R(d) \, f_\Theta(\theta)$$

The above is the joint density of (R, Θ).

Common mistakes:

-4 if distribution of (R^2, Θ) has been found
-2 if Jacobian calculation is not shown

Solution (b):

From the joint pdf above, we know that R and Θ are independent with $\Theta \sim U(0, 2\pi)$ and R has density $f_R(r) = 2r$. Note that the CDF of R is

$$u = F(r) = \int_0^r 2x dx = r^2 \Rightarrow F^{-1}(u) = \sqrt{u} \, .$$

Using the above inverse CDF, we have the following algorithm

(a) Draw $U_1, U_2 \sim U(0,1)$ independently
(b) Set $\Theta = 2\pi U_1$ and $R = \sqrt{U_2}$
(c) Set $X = R\cos\Theta$ and $Y = R\sin\Theta$.

Common mistakes:

-1 if iid not written
-3 if method of drawing from R is not described
-1 for not writing independent

2. (30 points) Consider the distribution F with density

$$f(x) = \frac{(1+x)e^{-x}}{2} \qquad \text{where } 0 < x < \infty \, .$$

(a) (10 points) For an Exponential(λ) proposal distribution, what is the best value of λ for the accept-reject algorithm?
(b) (5 points) Write down the full accept-reject algorithm using the best value of λ. You are only allowed to use draws from $U(0, 1)$.
(c) (10 points) Write down the steps of the Ratio-of-Uniforms algorithm to sample from F.
You are only allowed to use draws from $U(0, 1)$.
(d) (5 points) Which of the two algorithms in (b) and (c) is more efficient and why?

Solution:

(a)Let $g(x) = \lambda e^{-\lambda x}$ denote the density of an Exp(λ) proposal. For the above target, the best value of λ would be the one that would minimize the bound:

$$\sup_{x>0} \frac{f(x)}{g(x)} \leq c(\lambda) \, .$$

First, let's find a form of $c(\lambda)$

$$\begin{aligned}
\log r(x) &:= \log f(x) - \log g(x) \\
&= \log(1+x) - x + \lambda x - \log 2 - \log \lambda \\
\Rightarrow \frac{\partial \log r(x)}{\partial x} &= \frac{1}{1+x} - 1 + \lambda \overset{set}{=} 0 \\
\Rightarrow x &= \frac{\lambda}{1-\lambda} \qquad \text{for } \lambda < 1\,.
\end{aligned}$$

At this value of x, we obtain the optimal $c(\lambda)$ is

$$c(\lambda) = \frac{\frac{1}{1-\lambda} e^{-\lambda/(1-\lambda)}}{2} \frac{1}{\lambda e^{-\lambda^2/(1-\lambda)}} = \frac{1}{2\lambda} \frac{1}{1-\lambda} e^{-\lambda}\,.$$

We now want to minimize this value to obtain the best proposal. Taking log

$$\begin{aligned}
\log c(\lambda) &= -\log 2 - \log \lambda - \log(1-\lambda) - \lambda \\
\Rightarrow \frac{\log c(\lambda)}{d\lambda} &= -\frac{1}{\lambda} + \frac{1}{1-\lambda} - 1 \overset{set}{=} 0
\end{aligned}$$

Solving the equation we get that

$$\lambda^* = \frac{-1 \pm \sqrt{5}}{2}\,.$$

Since λ^* must be between $(0,1)$, we get

$$\lambda^* = \frac{\sqrt{5}-1}{2}\,.$$

Common mistakes:

-1 or -2 for small calculation errors
-2 if not accounted for the fact that $\lambda > 0$.
-5 if $c(\lambda)$ obtained but not attempted to minimize

(b)

Below is the algorithm

(1) Draw U_1, U_2 independently from $U(0,1)$

(2) Set $Y = \log(1-U_1)/\lambda^*$ (obtain this using inverse transform)

(3) If $U_2 \leq \dfrac{f(Y)}{c(\lambda^*)g(Y)}$, then set $X = Y$

(4) Else go to step (1)

Common mistakes:

-1 if iid not written
-1 if inverse transform method not illustrated for $\text{Exp}(\lambda^*)$.
-2 if different U_1 and U_2 are not taken

Solution (c)

To implement RoU, first recall the set

$$D = \left\{ (u, v) : 0 < u < \sqrt{f\left(\frac{v}{u}\right)} \right\} .$$

To implement RoU, we have to draw $(U, V) \sim \text{Unif}(D)$. This we will do using Accept-Reject by sampling from a box that encloses D. Box $[0, a] \times [b, c]$ encloses D where:

$$a = \sup_{x>0} \sqrt{f(x)} = \frac{1}{\sqrt{2}} (\text{ at } x = 0) \qquad \text{(show the steps)}$$

$$b = \inf_{x>0} x\sqrt{f(x)} = 0 \text{ since infimum must occur at } x = 0$$

$$c = \sup_{x>0} x\sqrt{f(x)} = x_c\sqrt{f(x_c)} \text{ where } x_c = \sqrt{3} + 1 \text{ (show this in steps)}$$

The steps of the RoU algorithm are:

(1) Draw $U_1, U_2 \sim U(0, 1)$ independently

(2) Set $U = aU_1$ and $V = cU_2$.

(3) If $U < \sqrt{f(V/U)}$, then set $X = V/U$

(4) Else, go to (1).

Common mistakes:

-1 if independently not written
-1 if drawn directly from $U_1(0, a)$ and $U_2(0, c)$
-3 if major error in RoU algorithm
-1 or -2 if small calculation error
-6 if not able to describe a, b and c.

Solution d:

The algorithm that is better is the one that has the smaller number of expected loops. The expected number of loops of AR is $c(\lambda^*)$.

The expected number of loops for the RoU is $2a(c - b)$. These values are difficult to compare, so whichever one is smaller should be the one chosen

Common mistakes:

-3 if form of expected number of loops is not written
or
-3 if logic is wrong

3. (20 points) Consider a target distribution F with density $f(x)$ with support $\mathcal{X}$. Suppose this density can be written as $f(x) = g_1(x) + g_2(x)$ where $g_1(x) \geq 0$ and $g_2(x) \geq 0$ are functions. Further, let

$$p := \int_{\mathcal{X}} g_2(x) dx .$$

so that $\tilde{g}_2(x) = g_2(x)/p$ is a pdf on $\mathcal{X}$ with distribution $\tilde{G}_2$. Let $h(x)$ be a pdf of a distribution H, so that $g_1(x) \leq h(x)$. With this setup, consider the following algorithm:

Algorithm 1

1: Draw $U \sim U(0,1)$

2: Independently, generate $Y \sim H$

3: If $U \leq \dfrac{g_1(Y)}{h(Y)}$, then accept $X = Y$ and stop

4: Else generate $Z \sim \tilde{G}_2$ and return $X = Z$.

Q: *Prove* that Algorithm 1 returns $X \sim F$.

Solution: We will show that the CDF of the output random variable X, is F. For any $t \in \mathbb{R}$, note that by law of total probability

$$\Pr(X \leq t) = \Pr\left(X \leq t, U \leq \frac{g_1(Y)}{h(Y)}\right) + \Pr\left(X \leq t, U > \frac{g_1(Y)}{h(Y)}\right)$$

We will consider both of these terms:

$$\begin{aligned}
&= \Pr\left(X \leq t, U \leq \frac{g_1(Y)}{h(Y)}\right)\\
&= \Pr\left(Y \leq t, U \leq \frac{g_1(Y)}{h(Y)}\right)\\
&= \mathbb{E}\left(I\left(Y \leq t, U \leq \frac{g_1(Y)}{h(Y)}\right)\right)\\
&= \mathbb{E}\left(I\left(Y \leq t\right)\left(U \leq \frac{g_1(Y)}{h(Y)}\right)\right)\\
&= \mathbb{E}\left[\mathbb{E}\left(I\left(Y \leq t\right) I\left(U \leq \frac{g_1(Y)}{h(Y)}\right) \mid Y\right)\right]\\
&= \mathbb{E}\left[I\left(Y \leq t\right) \mathbb{E}\left(I\left(U \leq \frac{g_1(Y)}{h(Y)}\right) \mid Y\right)\right]\\
&= \mathbb{E}\left[I\left(Y \leq t\right) \frac{g_1(Y)}{h(Y)}\right]\\
&= \int_{-\infty}^{t} \frac{g_1(y)}{h(y)} h(y) dy\\
&= \int_{-\infty}^{t} g_1(y) dy\,.
\end{aligned}$$

For the second term,

$$\begin{aligned}
&\Pr(X \leq t,\ \text{acceptance in Step 3})\\
&= \Pr\left(Z \leq t, U > \frac{g_1(Y)}{h(Y)}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \Pr\left(Z \le t\right) \Pr\left(U > \frac{g_1(Y)}{h(Y)}\right) \qquad \text{due to independence} \\
&= \Pr\left(Z \le t\right) \left[1 - \Pr\left(U \le \frac{g_1(Y)}{h(Y)}\right)\right] \\
&= \Pr\left(Z \le t\right) \left[1 - \int_{\mathcal{X}} g_1(y)dy\right] \\
&= \Pr\left(Z \le t\right) \left[\int_{\mathcal{X}} \left(g_1(y) + g_2(y)\right) dy - \int_{\mathcal{X}} g_1(y)dy\right] \\
&= \int_{-\infty}^{t} \frac{g_2(z)}{p} dz \left[\int_{\mathcal{X}} g_2(y)dy\right] \\
&= \int_{-\infty}^{t} \frac{g_2(z)}{p} dz \cdot p \\
&= \int_{-\infty}^{t} g_2(z)dz\,.
\end{aligned}
$$

Thus,

$$
\Pr(X \le t) = \int_{-\infty}^{t} g_1(y)dy + \int_{-\infty}^{t} g_2(z)dz = F(t)\,.
$$

Common mistakes:

-5 if probabilities not broken up properly in the beginning (but idea is correct)
-5 if second probability you cannot solve
-1 if independence not mentioned at key stage

4. (30 points) Let $Y|(X = x) \sim N(x, 1)$ for $x \in \mathbb{R}$ and let $X \sim$ Student's t_5-distribution.

   (a) (15 points) In the code below, function `denY` implements Simple Monte Carlo estimation. Identify the statistical quantity, θ, being estimated, and mathematically describe the estimator, as we have done in class.

```r
# Function that implements Simple Monte Carlo
denY <- function(y.star, N = 1e3)
{
  x <- rt(N, df = 5)
  hx_at_y.star <- dnorm(y.star, mean = x, sd = 1)
  den_at_y.star <- mean(hx_at_y.star)
  return(den_at_y.star)
}

# The code below may provide more hints on the purpose of denY
# make your choice of grid-values
ys <- seq(-5, 5, length = 5e2)
den <- numeric(length = length(ys))

# Calculate for each value of ys
for(i in 1:length(ys))
```

```
{
  den[i] <- denY(ys[i])
}
plot(ys, den, type = 'l', xlab = "y" = ylab = "Density")
```

Solution: Let $f_{(X,Y)}(x,y)$ denote the joint pdf of the random variable (X,Y) and $f_{Y|X}(y)$ denote the conditional density of $Y|X$ and $f_X(x)$ denote the marginal density of X.

In the function `denY`, we are drawing $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} t_5$, and taking the sample average over $f(y^*|X_t)$. So that my estimator is $\hat{\theta}$:

$$\hat{\theta} = \frac{1}{N}\sum_{t=1}^{N} f_{Y|X}(y^*|X_t) \quad \text{where } X_1, X_2, \ldots, X_N \overset{\text{iid}}{\sim} t_5$$

where $f_{Y|X}(y^*|X_t)$ is the pdf value at y^* of a $N(X_t, 1)$ distribution. By a law of large numbers,

$$\hat{\theta} \xrightarrow{p} \int f_{Y|X}(y^*|x) f_X(x)dx = f_Y(y^*) =: \theta\,.$$

Common mistakes:
-4 if not identified that θ is the marginal density of Y evaluated at a fixed y^*.
-1 if iid not written
-4 if not explained what $h(x) = f(y^*|X_t)$ is
-4 if $X_1, \ldots, X_n$ from t_5 is not identified.

(b) (15 points) Construct a simple importance sampling estimator of θ, as identified in (a). Your estimator should be guaranteed to have finite variance. Provide a mathematical argument for why your estimator has finite variance.

Solution:

For an importance proposal density $g(x)$ and target $f_X(x)$ being the pdf of a t_5 distribution, if we can show that

$$\sup_{x\in\mathbb{R}} \frac{f_X(x)}{g(x)} \le M < \infty$$

then, we know that the simple importance sampling estimator will have finite variance. Consider the standard Cauchy proposal density

$$g(x) = \frac{1}{\pi}\frac{1}{1+x^2} \qquad , x \in \mathbb{R}\,.$$

For this choice of $g(x)$,

$$r(x) = \frac{f_X(x)}{g(x)} = \frac{\Gamma(3)}{\sqrt{5\pi}\,\Gamma(2.5)}\left(1+\frac{x^2}{5}\right)^{-3} \pi(1+x^2)$$

$$\Rightarrow \log r(x) = \log(\text{constant}) - 3\log\left(1+\frac{x^2}{5}\right) + \log(1+x^2)\,.$$

Taking derivative, and setting to 0, we get that $r(x)$ is maximized at $x = \pm 1$, and thus $f_X(x)/g(x) \le r(1)$.

A finite variance simple importance sampling estimator then is the following.

(1) Generation $Z_1, Z_2, \ldots, Z_N \overset{iid}{\sim}$ Cauchy

(2) Estimate θ with

$$\begin{aligned}
\hat{\theta}_{IS} &= \frac{1}{N}\sum_{t=1}^{N} \frac{f_{Y|X}(y^*|Z_t)f_X(Z_t)}{g(Z_t)} \\
&= \frac{1}{N}\sum_{t=1}^{N} \frac{\frac{1}{\sqrt{2\pi}}e^{-(y^*-Z_t)^2/2}\frac{\Gamma(3)}{\sqrt{5\pi}\Gamma(2.5)}\left(1+\frac{Z_t^2}{5}\right)^{-3}}{\frac{1}{\pi}\frac{1}{(1+Z_t^2)}}
\end{aligned}$$

Common mistakes:

-8 if variance is not shown to be finite or if proposal chosen is not one giving finite variance

-1 if iid not written

-5 if sup $f(x)/g(x)$ is not solved and shown to be bounded

-10 if explicit choice of $g(x)$ is not taken