Assignment Sheet: Transformation Methods, Rejection Sampling, and Importance Sampling

Please hand in your answers to at least four of the questions before Monday, March 8th, 11.59pm. You must answer at least one question which involves programming (marked with a "$\star$"). For the programming questions you can use whatever programming language you like (R, Python, etc.). Please hand in your answers to programming questions electronically by emailing your code to Madeleine (`madeleine@aims.ac.za`).

1. The *Polar Marsaglia method* is a modification of the Box-Muller method that avoids having to compute the sine and cosine, however at the price of introducing a rejection sampling step:

   1. Generate $U_1, U_2 \overset{\text{i.i.d.}}{\sim} \mathsf{U}[-1,1]$.
   2. Compute $R^2 = U_1^2 + U_2^2$.
   3. If $R^2 > 1$ reject the pair (U_1, U_2), and go back to step 1.
   4. Set $X_1 = U_1\sqrt{\dfrac{-2\log(R^2)}{R^2}}$ and $X_2 = U_2\sqrt{\dfrac{-2\log(R^2)}{R^2}}$

   (a) Show that X_1 and X_2 are independent and both from the $\mathsf{N}(0,1)$ distribution.

   $\star$ (b) Write a function which draws from the Normal distribution using the Polar Marsaglia method. Use it to draw a sample of size 100 from the $\mathsf{N}(0,1)$ distribution.

2. Let $F(\cdot)$ be a c.d.f., and let $F^-(\cdot)$ be its generalised inverse.

   (a) Let $U_1, U_2 \sim \mathsf{U}(0,1)$. Show that

   $$2 \cdot \text{Cov}\left(F^-(U_1), F^-(1-U_1)\right) = \mathbb{E}\left(\left(F^-(U_1) - F^-(U_2)\right)\left(F^-(1-U_1) - F^-(1-U_2)\right)\right)$$

   (b) Show that for all $u_1, u_2 \in \mathbb{R}$

   $$\left(F^-(u_1) - F^-(u_2)\right)\left(F^-(1-u_1) - F^-(1-u_2)\right) \leq 0.$$

   (c) Deduce from what you have obtained so far that $\text{Cov}\left(F^-(U_1), F^-(1-U_1)\right) \leq 0$.

   (d) Show that

   $$\text{Var}\left(\frac{F^-(U_1) + F^-(1-U_1)}{2}\right) \leq \text{Var}\left(\frac{F^-(U_1) + F^-(U_2)}{2}\right).$$

   Interpret this result.

3. Assume you want to sample from the $\mathsf{N}(0,1)$ distribution using *rejection sampling* with the $\mathsf{N}(1,\sigma^2)$ distribution as instrumental distribution.

   (a) Show that the rejection sampling algorithm yields samples from the target $N(0,1)$ distribution iff $\sigma^2 > 1$.
   Hint: Show that the ratio

   $$\frac{\phi_{(0,1)}(x)}{\phi_{(1,\sigma^2)}(x)} = \sqrt{\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left((\sigma^2-1)x^2 + 2x - 1\right)\right),$$

   and find the values of σ^2 for which it is bounded in x.

   (b) Show that for fixed σ^2 one has to choose

   $$M \geq \sqrt{\sigma^2}\exp\left(\frac{1}{2(\sigma^2-1)}\right).$$

   Hint: Find the maximum of the ratio from part (a).

   (c) Show that the optimal choice of σ^2 is given by $\sigma^2 = \dfrac{\sqrt{5}+3}{2} \approx 2.6180$.
   In the context of rejection sampling, optimality refers to maximising the probability of acceptance.

4. Assume you want to sample from the $\mathsf{N}(0,1)$ distribution using *importance sampling*[1] with the $\mathsf{N}(1,\sigma^2)$ distribution as instrumental distribution.

   (a) Show that the variance of the weights is

$$\mathrm{Var}_g(w(X)) = \frac{\sigma^2}{\sqrt{2\sigma^2-1}} \exp\left(\frac{1}{2\sigma^2-1}\right) - 1,$$

   (b) Show that the variance of the weights is finite iff $\sigma^2 > 1/2$.

   (c) Show that $\sigma^2 \approx 2.2808$ minimises the variance of the weights.

⋆ 5. In this question you will implement the sampling methods studied in questions 3 and 4.

   (a) Write a function which samples from the $\mathsf{N}(0,1)$ distribution using rejection sampling with the $\mathsf{N}(1, 2.6180)$ distribution as instrumental distribution.
   Check your function by drawing a sample of size 1000 and draw a histogram. Compare it to the density $\phi(\cdot)$ of the $\mathsf{N}(0,1)$ distribution.

   (b) Write a function which creates a weighted sample from the $\mathsf{N}(0,1)$ distribution using importance sampling with the $\mathsf{N}(1, 2.2808)$ distribution as instrumental distribution.
   Check your function by drawing a sample of size 1000 and compute both the self-normalised estimate $\hat{\mu}$ and the simple estimate $\tilde{\mu}$ for both $\mu = \mathbb{E}_f(X)$ and $\mu = \mathbb{E}_f(X^2)$. (These should be approximately 0 and 1, respectively).

6. Assume that the instrumental distribution g in importance sampling[1] is chosen such that

$$f(x) < M \cdot g(x)$$

   for all x and a suitable $M \in \mathbb{R}$, where f is the density of the target distribution.

   (a) Show that $\mathrm{Var}_g(w(X)) < M - 1$.

   (b) Show that $\mathrm{Var}_g(w(X) \cdot h(X))$ is finite, if $\mathrm{Var}_f(h(X))$ is finite.

⋆ 7. A random variable is said to be from a left-truncated normal distribution if its density is

$$f_{(\mu,\sigma^2,\tau)}(x) = \begin{cases} 0 & \text{for } x \leq \tau \\ \phi_{(\mu,\sigma^2)}(x)/(1 - \Phi_{(\mu,\sigma^2)}(\tau)) & \text{for } x > \tau, \end{cases}$$

   where $\phi_{(\mu,\sigma^2)}(\cdot)$ is the density of the $\mathsf{N}(\mu,\sigma^2)$ distribution, and $\Phi_{(\mu,\sigma^2)}(\cdot)$ the corresponding c.d.f.

   (a) One way of sampling from the left-truncated normal distribution is rejection sampling with the $N(\mu,\sigma^2)$ distribution as instrumental distribution. Implement this algorithm by writing a function which takes `n` (sample size), `mu` (μ), `sigma` (σ), and `tau` (τ) as parameters and which returns a sample of size n of the left-truncated normal distribution.
   Hint: Note that in this example, the probability of accepting a proposed X is either 0 or 1.

   (b) Use the code from part (a) to draw 10 realisations from a left-truncated normal distribution with parameters $\mu = 0$, $\sigma^2 = 1$, and $\tau = 4$. What is the proportion of rejected values you observed?

   (c) Clearly, the method proposed in part (a) is very inefficient. Propose and implement a more efficient instrumental distribution, i.e. one that yields less rejected values. Attempt to obtain a proportion of rejected values of at most 90%.
   Hint: All the mass of the left-truncated normal distribution is in $(\tau, +\infty)$.

⋆ 8. This question is about estimating $\mathbb{E}(X(1-X))$ for $X \sim \mathsf{Beta}(\alpha,\beta)$ using importance sampling with the uniform distribution $\mathsf{U}[0,1]$ used as instrumental distribution.

   (a) Based on a sample of size 10 compute an importance sampling estimate of $\mu = \mathbb{E}(X(1-X))$ for $\alpha = 2$ and $\beta = 3$, once using the self-normalised estimate $\hat{\mu}$ and once using the "simple" estimate $\tilde{\mu}$.

   (b) Estimate the bias, variance, and the mean-squared error[2] of both methods based on 100,000 replications of your computations in part (a). For computing the bias, use $\mathbb{E}(X(1-X)) = 1/5$. Based on your results, would you prefer the self-normalised estimate?

[1] In this question "importance sampling" refers to importance sampling using $\tilde{\mu}$, i.e. *not* using self-normalised weights.
[2] The bias of $\hat{\mu}$ is $\mathrm{bias}(\hat{\mu}) = \mathbb{E}_g(\hat{\mu}) - \mu$, and the mean square error is $\mathrm{mse}(\hat{\mu}) = \mathbb{E}_g((\hat{\mu}-\mu)^2)$. One can show that $\mathrm{mse}(\hat{\mu}) = (\mathrm{bias}(\hat{\mu}))^2 + \mathrm{Var}_g(\hat{\mu})$.