THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT6011/7611/8305 COMPUTATIONAL STATISTICS
(2021 Fall)

Assignment 1, due on October 5

All numerical computation MUST be conducted in Python, and attach the Python code.

Hints (Useful packages and functions in Python):

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm, expon, t, binom, gamma
from scipy.optimize import minimize
```

1. **Question 1 (Importance Sampling for Rare Event Probability)**

   (a) For $a = 2, 4, 6$, calculate $P(X > a) = \int_a^\infty \phi(x)\mathrm{d}x$, where $X \sim N(0,1)$. With 100000 samples, use (i) the 'Naive' approach (simulate samples from $\phi(x)$, PDF of the standard normal distribution), (ii) the non-ratio version Importance Sampling (simulate samples from the proposal $q(x) = e^{-(x-a)}\mathbf{1}_{[a,\infty)}(x)$), (iii) the ratio version Importance Sampling (simulate samples from the proposal $q(x) = 0.5t_4(x) + 0.5e^{-(x-a)}\mathbf{1}_{[a,\infty)}(x)$).

   (b) Calculate the optimal proposals of both non-ratio or ratio version Importance Sampling. Compare the proposals in (a) with the corresponding optimal proposals by plots.

   (c) For the non-ratio version Importance Sampling, we find that the previous proposal is too flat compared to the optimal proposal. To improve the performance of Importance Sampling, we assume the proposal family to be $q(x|b) = be^{-b(x-a)}\mathbf{1}_{[a,\infty)}(x)$. To find the best parameter b, note that the asymptotic variance can be estimated by

$$
\begin{aligned}
\sigma^2(b) &= \mathbb{E}_b[(w_b(X)f(X) - \mu)^2] \\
&= \int \frac{(\phi(x)\mathbf{1}_{[a,\infty)}(x) - \mu q(x|b))^2}{q(x|b)}\mathrm{d}x = \int \frac{(\phi(x)\mathbf{1}_{[a,\infty)}(x))^2}{q(x|b)}\mathrm{d}x - \mu^2 \\
&\approx \hat{S}(b|b_0) - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} \frac{(\phi(x_i)\mathbf{1}_{[a,\infty)}(x_i))^2}{q(x_i|b)q(x_i|b_0)} - \mu^2,
\end{aligned}
$$

   where $x_i \sim q(x|b_0), i = 1, \dots, n$. So, instead of minimizing $\sigma^2(b)$ with respect to b, we can minimize $\hat{S}(b|b_0)$ for a given b_0. Furthermore, we can iterate the optimization procedure by replacing b_0 by the optimal b in the previous step, which lead to the Adaptive Importance Sampling

(AIS) methodology. When $a = 6$, $b_0 = 1$ and $n = 100$, build the AIS algorithm with 10 iterations to find the optimal b, plot the log value of the asymptotic variance $\sigma^2(b)$ (estimated with 10000 new samples x_i) at the initial b_0 and the optimal b at each iteration step.

(d) Now, we consider the gamma family $q(x|b,c) = b^2(x-c)e^{-b(x-c)}\mathbf{1}_{[c,\infty)}(x)$ as the proposal and obtain the optimal values for b and c. So, consider

$$\hat{S}(b,c|b_0,c_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{(\phi(x_i)\mathbf{1}_{[a^*,\infty)}(x_i))^2}{q(x_i|b,c)q(x_i|b_0,c_0)},$$

where $x_i \sim q(x|b_0,c_0), i = 1,\ldots,n$; $a^* = \min(a, x_{(1-\varepsilon)})$, $x_{(1-\varepsilon)}$ is the $(1-\varepsilon)$-quantile of $\{x_i\}$. When $a = 6$, $b_0 = 1$, $c_0 = 0$, $n = 100$ and $\varepsilon = 0.1$, build an AIS algorithm with $\hat{S}(b,c|b_0,c_0)$ and 20 iterations to find the optimal b and c, plot the log value of the asymptotic variance (estimated with 2000000 new samples) at the initial b_0, c_0 and optimal parameters b, c at each iteration step.

2. **Question 2 (Modified Rejection Sampling)**

To draw samples from the target distribution $\pi(x)$, remember the procedure of Rejection Sampling (RS): 1) Draw initial samples $\{x_1,\ldots,x_m\}$ from an envelope distribution $q(x)$, and calculate the corresponding ratios or weights $\{w_1,\ldots,w_m\}, w_j = w(x_j) = \pi(x_j)/q(x_j)$; 2) Calculate $C = \sup w(x)$, and accept each sample x_j with the probability $p_j = w_j/C$ to obtain the final samples $\{x_1^*,\ldots,x_n^*\}$. A modified version of RS is to replace C by $\max w_j$.

(a) Obviously the first benefit of replacing $\sup w(x)$ by $\max w_j$ is that it saves efforts to calculate the maximum value especially when the shape of the target or proposal are complicated or non-smooth. Let $\pi(x) = 0.5N(x|-2, 0.5^2) + 0.5N(x|1, 1^2)$ be the target and $q(x) = t_1(x)$ be the envelope. Plot $w(x)$ and calculate C.

(b) The acceptance rate for the original RS is $1/C$. As $\max w_j < C$, the modified RS may have higher acceptance rate. Set $m = 50, 100, 200, 500, 1000$. Under each m, implement these two types of Rejection Sampling with 1000 replications to estimate and compare their acceptance rates by plotting the acceptance rate vs $\log(m)$.

(c) The third advantage is that different from $\sup w(x)$, $\max w_j$ can always be obtained even if the tail of $q(x)$ is lighter than that of $\pi(x)$. Let $\pi(x) = N(x|0, 1^2)$, $q(x) = N(x|0, 0.8^2)$. Generate over 3000 final samples using the modified version of RS, and plot the histogram vs the true target $\pi(x)$.

(d) The modified RS procedure is actually a biased method (but asymptotically unbiased), which means that the final samples do not follow the target distribution. Set $m = 5$, $\pi(x) = N(x|0, 1^2)$, $q(x) = N(x|0, 3^2)$. Run $K = 20000$ repetitions for both the RS and the modified RS, for the

kth repetition, obtain a group of final samples $\{x^*_{k,1}, \dots, x^*_{k,n_k}\}$ and assign weights $1/n_k$ for each sample in this group (there would be no sample to set the weights if $n_k = 0$), finally pool the samples in the K groups to obtain the weighted samples $\{(x^*_{1,1}, 1/n_1), \dots, (x^*_{k,i}, 1/n_k), \dots, (x^*_{K,n_K}, 1/n_K)\}$. Plot the two histograms of the pooled weighted samples corresponding to the RS and the modified RS and compare them with $\pi(x)$.