STAT3006 Assignment 2

Cheung Tak Leung 1155157783

April 8, 2023

Contents

1 Question 1

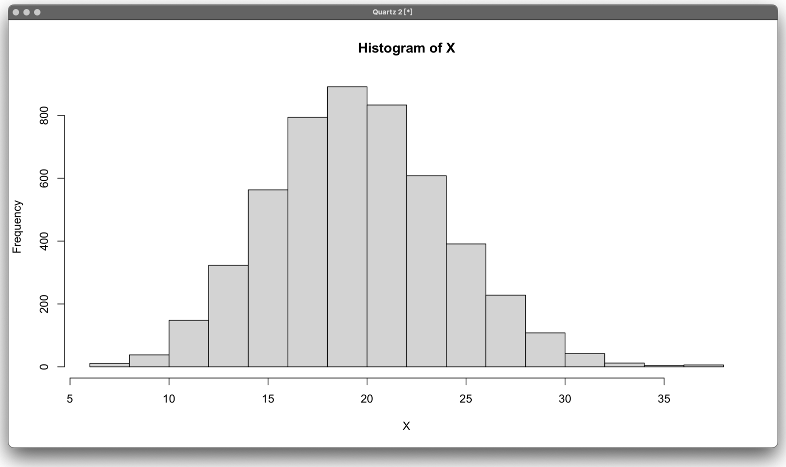
Algorithm 1 Inverse Method

Input: the cdf. of $Poisson(\lambda = 20)$**, total sample number** $N = 5000$

$N \leftarrow 5000$
$i \leftarrow 1$
while $i \leq N$ **do**
    Generate a random number U_i from $U[0, 1]$
    $X_i \leftarrow F^{-}(U_i)$
    $i \leftarrow i + 1$
end while

Output: $\{X_1, ..., X_N\}$ **are** N **samples from** F

The inverse function of cdf. of $Poisson(\lambda = 20)$ in R is the quantile function of itself, in particular `qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)`.



2 Question 2

$$
\begin{aligned}
f(x) &= \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \cdot I(x \le 4)}{\Phi(-4)} \\
g(x) &= 4e^{4(x+4)}I(x \le -4) \\
M &= \frac{e^{-8}}{4\sqrt{2\pi}\Phi(-4)} \\
f(x) \le M \cdot g(x) &\iff M \cdot g(x) - f(x) \ge 0 \\
M \cdot g(x) - f(x) &= \frac{e^{-8}4e^{4(x+4)}I(x \le -4)}{4\sqrt{2\pi}\Phi(-4)} - \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \cdot I(x \le 4)}{\Phi(-4)} \\
&= \frac{I(x \le 4)}{\sqrt{2\pi}\Phi(-4)}\left[e^{-8}e^{4(x+4)} - e^{-\frac{x^2}{2}}\right] \\
&= \frac{I(x \le 4)}{\sqrt{2\pi}\Phi(-4)}\left(e^{4x+8} - e^{-\frac{x^2}{2}}\right)
\end{aligned}
$$

Note that, $\frac{I(x \le 4)}{\sqrt{2\pi}\Phi(-4)} \ge 0 \quad \forall x \in \mathbb{R}$ since Φ is always positive by definition.

$$
\begin{aligned}
\text{Assume } e^{4x+8} - e^{-\frac{x^2}{2}} &< 0 \\
e^{4x+8} &< e^{-\frac{x^2}{2}} \\
4x + 8 + \frac{x^2}{2} &< 0 \\
(x+4)^2 &< 0 \text{ ,which does not make sense.}
\end{aligned}
$$

Therefore, a contradiction arises. Hence, $e^{4x+8} - e^{-\frac{x^2}{2}} \ge 0$

$$\text{Hence, } M \cdot g(x) - f(x) = \frac{I(x \le 4)}{\sqrt{2\pi}\Phi(-4)}\left(e^{4x+8} - e^{-\frac{x^2}{2}}\right) \ge 0$$

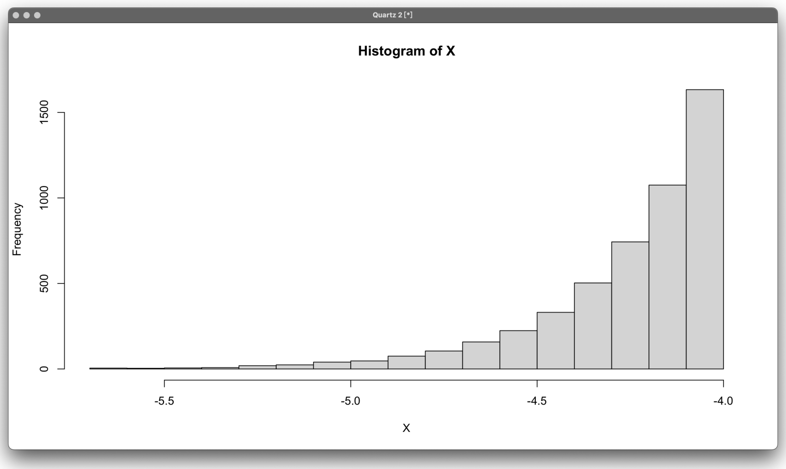
$g(x) = 4e^{4(x+4)}I(x \le -4) = 4e^{-4[-x+(-4)]}I(x \le -4)$ is a shifted exponential distribution which reflected along x-axis.

Algorithm 2 Accept-Reject Method

Input: target pdf f, proposal pdf g, constant M ($f(x) \le Mg(x)$), total sample number $N = 5000$

$N \leftarrow 5000$
$i \leftarrow 1$
while $i \le N$ **do**
  Generate a random number U_i from $U[0,1]$
  Generate a random number Y_i from $g(y)$
  if $U_i \le \frac{f(Y_i)}{M \cdot g(Y_i)}$ **then**
    $X_i \leftarrow Y_i$
    $i \leftarrow i + 1$
  end if
end while

Output: $\{X_1, ..., X_N\}$ **are** N **samples from** f



3 Question 3

3.1 Part (a)

$$\int_{-\infty}^{-4} \cos(2x+3)e^{-x^2/2}dx = \int_{-\infty}^{\infty} \cos(2x+3)\frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \cdot I(x \le 4)}{\Phi(-4)}dx \cdot \sqrt{2\pi}\cdot\Phi(-4) = \mathbf{E}_f(cos(2X+3))\cdot\sqrt{2\pi}\cdot\Phi(-4)$$

where $X \sim f$ and by the strong law of large numbers,

$$\sqrt{2\pi}\cdot\Phi(-4)\cdot\lim_{n\to\infty}\frac{\sum_{i=1}^{n}\cos(2x_i+3)}{n} = \sqrt{2\pi}\cdot\Phi(-4)\cdot\mathbf{E}_f(\cos(2X+3)) = \int_{-\infty}^{-4}\cos(2x+3)e^{-x^2/2}dx$$

Algorithm 3 Classic Monte Carlo Integration using 5000 samples from Question 2

Input: $X_1, ...X_{5000}$**, the 5000 samples from Question 2, sample size** $N = 5000$

$N \leftarrow 5000$
$Y \leftarrow 0$
$i \leftarrow 1$
while $i \le N$ **do**
  $Y \leftarrow Y + \cos(2X_i + 3)$
  $i \leftarrow i + 1$
end while
$Y \leftarrow \sqrt{2\pi}\cdot\Phi(-4)\cdot Y/N$

Output: Mean of $\cos(2X_i + 3)$ **multiplied by** $\sqrt{2\pi}\cdot\Phi(-4)$

In short, I used `mean(cos(2X + 3))*sqrt(2*pi)*pnorm(-4)` where `X` is the vector that contains 5000 samples from question 2 and the result is `4.760799e-05`.

3.2 Part (b)

Denote $f(x) = \dfrac{\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \cdot I(x \le 4)}{\Phi(-4)}$, $g(x) = 4e^{4(x+4)}I(x \le -4)$

$$\int_{-\infty}^{-4} \cos(2x+3)e^{-x^2/2}dx = \int_{-\infty}^{\infty}\cos(2x+3)\frac{f(x)}{g(x)}\cdot g(x)dx\cdot\sqrt{2\pi}\cdot\Phi(-4) = \mathbf{E}_g\left(\cos(2X+3)\frac{f(X)}{g(X)}\right)\cdot\sqrt{2\pi}\cdot\Phi(-4)$$

where $X \sim g$ and by the strong law of large numbers,

$$\sqrt{2\pi}\cdot\Phi(-4)\cdot\lim_{n\to\infty}\frac{\sum_{i=1}^{n}\cos(2x_i+3)\frac{f(x_i)}{g(x_i)}}{n} = \sqrt{2\pi}\cdot\Phi(-4)\cdot\mathbf{E}_g\left(\cos(2X+3)\frac{f(X)}{g(X)}\right) = \int_{-\infty}^{-4}\cos(2x+3)e^{-x^2/2}dx$$

Algorithm 4 Importance Sampling Integration

Input: target pdf f**, proposal pdf** g**, sample size** $N = 5000$

$N \leftarrow 5000$
$Y \leftarrow 0$
$i \leftarrow 1$
while $i \le N$ **do**
  Generate a random number X from g
  $Y \leftarrow Y + cos(2X+3) \times f(X)/g(X)$
  $i \leftarrow i + 1$
end while
$Y \leftarrow \sqrt{2\pi}\cdot\Phi(-4)\cdot Y/N$

Output: Mean of $\cos(2X+3)\frac{f(X)}{g(X)}$ **multiplied by** $\sqrt{2\pi}\cdot\Phi(-4)$

In short, I used `mean(cos(2*Y+3)*f(Y)/g(Y))*sqrt(2*pi)*pnorm(-4)` where `Y` is a vector that contains 5000 samples from g and the result is `4.774457e-05`.

4 Question 4

4.1 Part (a)

By simple random sampling, the standard deviations for each subpopulation are 89.50597, 105.01124 and 201.28166 respectively.

4.2 Part (b)

Obtain the optimal stratified sample sizes for each subpopulation by minimizing $\sum_{i=1}^{3} \frac{\mu^2(S_i)Var(X|X \in S_i)}{n_i}$ subject to $2000 = n_1 + n_2 + n_3$
Let $L = \sum_{i=1}^{3} \frac{\mu^2(S_i)Var(X|X \in S_i)}{n_i} - \lambda(2000 - n_1 - n_2 - n_3)$, for $i = 1, 2, 3$,

$$\begin{aligned}
\frac{\partial L}{\partial n_i} &= -\frac{\mu^2(S_i)Var(X|X \in S_i)}{n_i^2} + \lambda = 0 \\
n_i &= \frac{\mu(S_i)\sqrt{Var(X|X \in S_i)}}{\sqrt{\lambda}} \\
2000 &= \frac{\sum_{k=1}^{3} \mu(S_k)\sqrt{Var(X|X \in S_k)}}{\sqrt{\lambda}} \\
\sqrt{\lambda} &= \frac{\sum_{k=1}^{3} \mu(S_k)\sqrt{Var(X|X \in S_k)}}{2000} \\
n_i &= \frac{2000\mu(S_i)\sqrt{Var(X|X \in S_i)}}{\sum_{k=1}^{3} \mu(S_k)\sqrt{Var(X|X \in S_k)}}
\end{aligned}$$

Using the part (a) answer to estimate $\sqrt{Var(X|X \in S_k)}$. The sample sizes for each subpopulation are 198, 465 and 1337 respectively.

4.3 Part (c)

By randomly drawing n_i samples in strata i for each i, the estimated population mean salary based on these 2000 samples is 4229.49, which is slightly lower than the actual mean salary of 4233.833. The error is -4.342394. The stratified sampling can somehow reflect the accurate mean salary.