102bhw2

Neva Williams

Question 1

a

```
B <- function(x) sin(6*x)^2 + 3*cos(x)^2*sin(4*x)^2 + 1
h <- function(x) exp(-x^2/2) * B(x)
g <- function(x) dnorm(x)
M_acceptable <- sqrt(2*pi) * 5

optB  <- optimize(function(x) B(x), interval = c(0, pi), maximum = TRUE)
B_max_opt <- optB$objective

xx <- seq(0, pi, length.out = 100000)
B_max_grid <- max(B(xx))

B_max <- max(B_max_opt, B_max_grid)
M_opt <- sqrt(2*pi) * B_max

cat(sprintf("M_acceptable = %.6f\n", M_acceptable))
```

```
M_acceptable = 12.533141
```

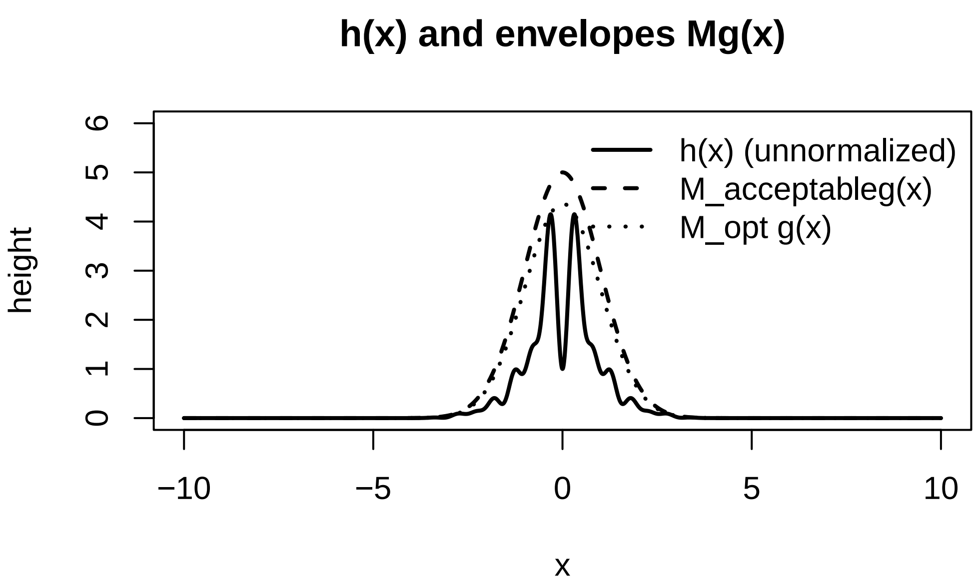
```
cat(sprintf("B_max (optimize) = %.6f, B_max (grid) = %.6f\n", B_max_opt, B_max_grid))
```

```
B_max (optimize) = 2.177928, B_max (grid) = 4.364551
```

```r
cat(sprintf("M_opt = %.6f\n", M_opt))
```

```
M_opt = 10.940306
```

```r
xg <- seq(from = -10, to = 10, length.out = 4000)
y_h    <- h(xg)
y_env1 <- M_acceptable * g(xg)
y_env2 <- M_opt * g(xg)


plot(xg, y_h, type = "l", lwd = 2,ylim=c(0,6),
     xlab = "x", ylab = "height",
     main = "h(x) and envelopes Mg(x)")
lines(xg, y_env1, lty = 2, lwd = 2)
lines(xg, y_env2, lty = 3, lwd = 2)
legend("topright",
       legend = c("h(x) (unnormalized)", "M_acceptableg(x)", "M_opt g(x)"),
       lty = c(1, 2, 3), lwd = 2, bty = "n")
```



b

```r
set.seed(67676767)
n <- 25000
X_acc <- numeric(0)
proposed_counter <- 0

while (length(X_acc) < n) {
  m <- n - length(X_acc)
  X_proposed <- rnorm(m)
  U <- runif(m)

  # With M = sqrt(2*pi)*B_max, accept if U <= B(X)/B_max
  acc <- U <= ( B(X_proposed) / B_max )    # acceptance test

  # store accepted draws
  X_acc <- c(X_acc, X_proposed[acc])

  # Increment counter of proposals
  proposed_counter <- proposed_counter + m
}

X <- X_acc                                  # accepted samples
accept_rate <- length(X) / proposed_counter # ratio of accepts to proposals

list(accepted = length(X), accept_rate = accept_rate)
```

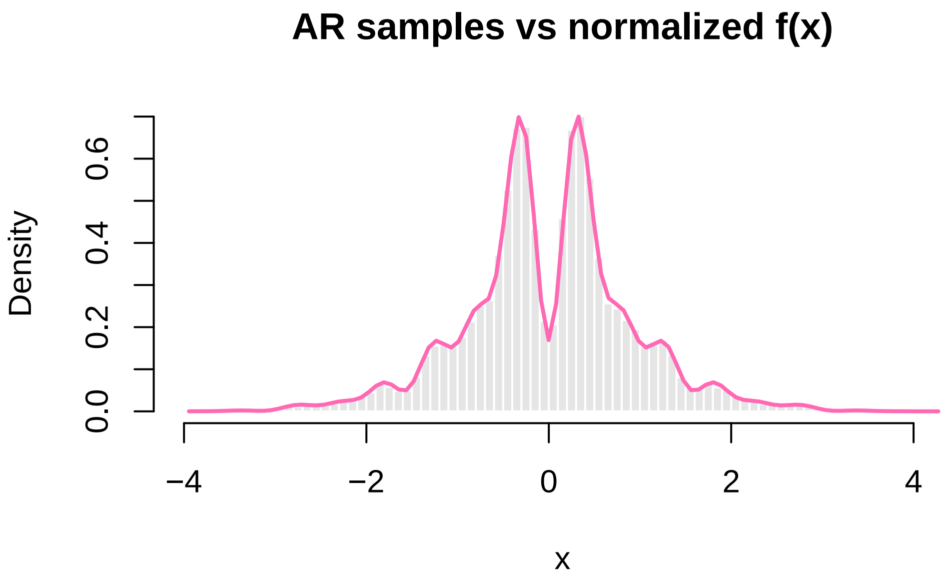
```
$accepted
[1] 25000

$accept_rate
[1] 0.5401085
```

Exactly 25000 samples will be accepted every time since that is specified in the function. The acceptance rate will change if you change the seed.

C

```r
Z_hat <- accept_rate * sqrt(2*pi) * B_max
```

```r
# Normalized target density
f_norm <- function(x) h(x) / Z_hat

# Compare histogram of samples with normalized f
hist(X, breaks = 80, probability = TRUE,
     col = "grey90", border = "white",
     main = "AR samples vs normalized f(x)", xlab = "x")
curve(f_norm, from = min(X), to = max(X), add = TRUE, lwd = 2,col="hotpink")
```



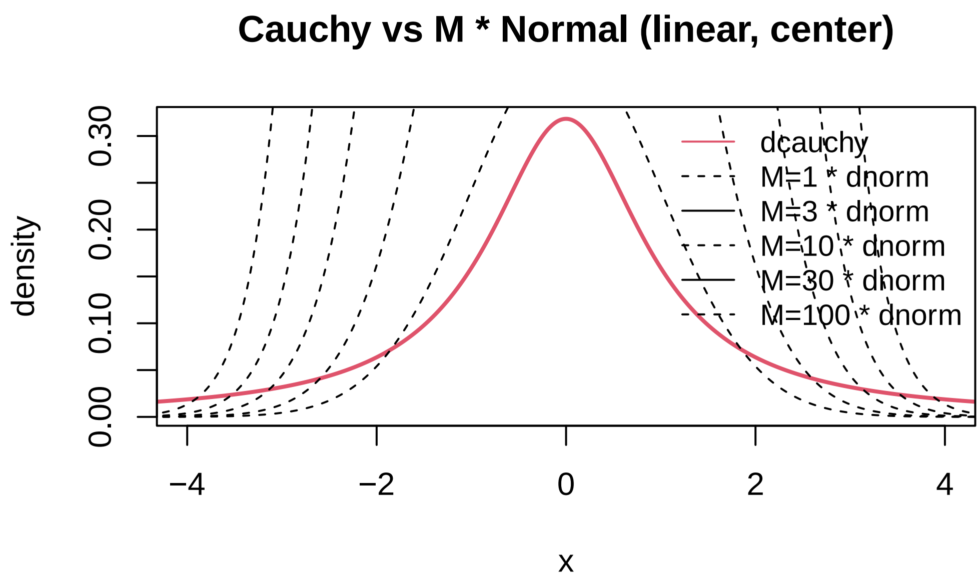

```r
cat(sprintf("Z_hat (from AR acceptance) = %.6f\n", Z_hat))
```

```
Z_hat (from AR acceptance) = 5.908952
```

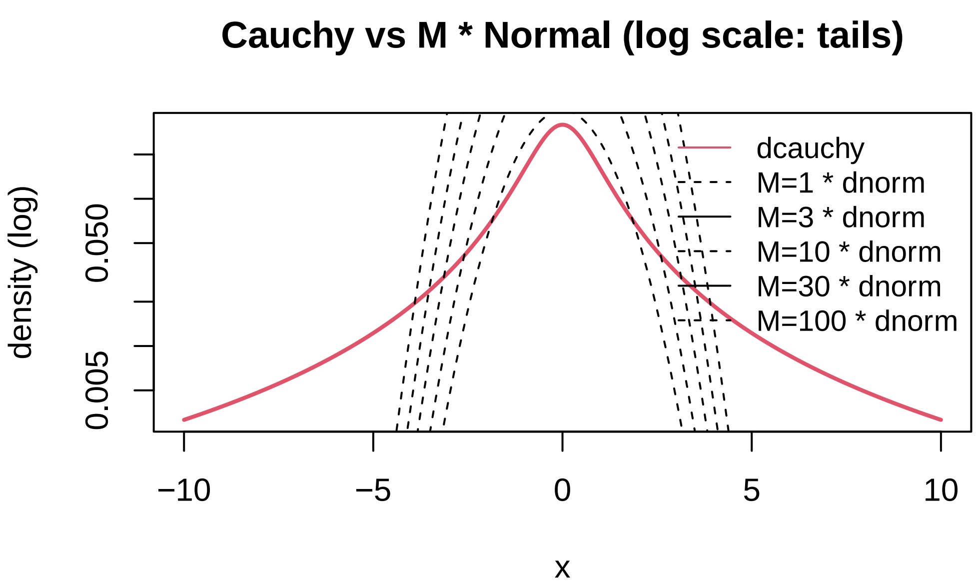
Question 2

```r
x <- seq(-10, 10, length.out = 4000)
f <- dcauchy(x)    # target: Cauchy(0,1)
g <- dnorm(x)      # proposal: N(0,1)
Ms <- c(1, 3, 10, 30, 100)


## (1) Linear-scale overlay near the center
```

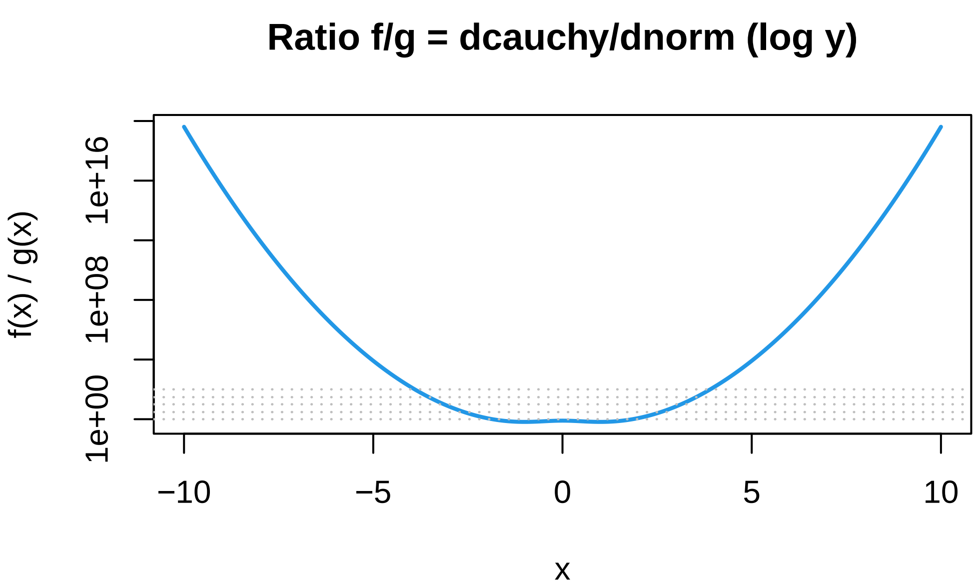
```r
plot(x, f, type = "l", lwd = 2, col = 2,
     main = "Cauchy vs M * Normal (linear, center)",
     xlab = "x", ylab = "density", xlim = c(-4, 4))
for (M in Ms) lines(x, M * g, lty = 2)
legend("topright",
       legend = c("dcauchy", paste0("M=", Ms, " * dnorm")),
       col = c(2, rep(1, length(Ms))), lty = c(1, 2),
       bty = "n", cex = 0.9)
```




```r
plot(x, f, type = "l", lwd = 2, col = 2, log = "y",
     main = "Cauchy vs M * Normal (log scale: tails)",
     xlab = "x", ylab = "density (log)")
for (M in Ms) lines(x, M * g, lty = 2)
legend("topright",
       legend = c("dcauchy", paste0("M=", Ms, " * dnorm")),
       col = c(2, rep(1, length(Ms))), lty = c(1, 2),
       bty = "n", cex = 0.9)
```




```r
ratio <- f / g
plot(x, ratio, type = "l", lwd = 2, col = 4, log = "y",
     main = "Ratio f/g = dcauchy/dnorm (log y)",
     xlab = "x", ylab = "f(x) / g(x)")
abline(h = Ms, lty = 3, col = "gray", lwd = 1.2)
```



You can see in the plots that any multiple of the normal distribution fails to fully envelop the Cauchy distribution (RED) which proves that the normal distribution is not a suitable trial distribution. The Cauchy distribution has very heavy tails that the normal distribution cannot cover. The blue plot shows that no value of M makes f(x) less than Mg(x) across the entire domain of the function.

b (HANDWRITTEN)

2B

$U, V \overset{iid}{\sim} N(0,1) \quad X = U/V$

Joint density

$$f_{U,V}(u,v) = \frac{1}{2\pi} e^{-\left(\frac{u^2+v^2}{2}\right)}$$

$U = XT \quad V = T$

$$\frac{\partial(u,v)}{\partial(x,t)} = \begin{pmatrix} \partial u/\partial x & \partial u/\partial t \\ \partial v/\partial x & \partial v/\partial t \end{pmatrix} = \begin{pmatrix} t & x \\ 0 & 1 \end{pmatrix}, \quad J = \det\begin{pmatrix} t & x \\ 0 & 1 \end{pmatrix} = t \quad |J| = |t|$$

$$f_{X,T}(x,t) = f_{U,V}(xt,t)\,|J| = \frac{1}{2\pi} e^{\left(-\frac{(xt)^2+t^2}{2}\right)} |t| = \frac{1}{2\pi} |t| e^{\left(-\frac{1+x^2}{2}\cdot t^2\right)}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,T}(x,t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |t| e^{\left(-\frac{1+x^2}{2} t^2\right)} dt$$

$$* \text{ even integrand} \rightarrow \frac{1}{\pi}\int_0^{\infty} t e^{\left(-\frac{1+x^2}{2} t^2\right)} dt$$

$$y = \frac{1+x^2}{2} \rightarrow dy = (1+x^2)\,t\,dt, \quad t\,dt = \frac{dy}{1+x^2}$$

$$f_X(x) = \frac{1}{\pi}\int_0^{\infty} e^{-y} \frac{dy}{1+x^2} = \frac{1}{\pi(1+x^2)}\int_0^{\infty} e^{-y} dy$$

$$= \frac{1}{\pi(1+x^2)}$$

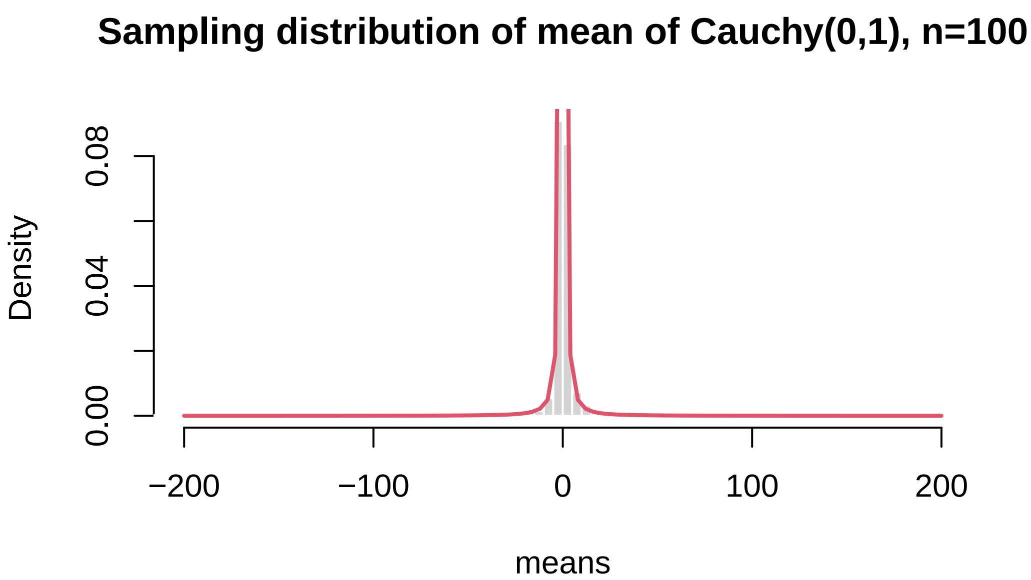
Cauchy PDF : $f(x;\mu,\gamma) = \dfrac{1}{\pi\gamma\left[1+\left(\frac{x+\mu}{\gamma}\right)^2\right]}$

when $\mu = 0$, $\gamma = 1$: $f(x) = \dfrac{1}{\pi(1+x^2)}$

$$X = \frac{U}{V} \sim \text{Cauchy}(0,1)$$

c

```r
set.seed(12)
replications <- 1000
n <- 100
means <- replicate(replications, mean(rcauchy(n)))
bw <- 3.5 * sd(means) * replications^(-1/3)
k <- ceiling((max(means) - min(means)) / bw)
hist(means, breaks = k, probability = TRUE,
     main = "Sampling distribution of mean of Cauchy(0,1), n=100",
     xlab = "means", border = "white", xlim = c(-200, 200))
curve(dcauchy(x), add = TRUE, col = 2, lwd = 2)
```



The Cauchy Distribution doesn't have a defined mean or variance because of its heavy tails so it doesn't follow a Gaussian Distribution.

d

i

The posterior density is

$$\pi(\theta \mid x) \propto L(\theta \mid x)\,\pi(\theta)$$

and the proposal (candidate) density is

$$g(\theta) = \pi(\theta).$$

In the Accept–Reject algorithm we require

$$\pi(\theta \mid x) \le M\, g(\theta) \;\Rightarrow\; L(\theta \mid x) \le M \quad \text{for all } \theta.$$

The smallest valid bound is the supremum of the likelihood:

$$M = \sup_{\theta} L(\theta \mid x) = L(\hat{\theta}_{\text{MLE}}),$$

where $\hat{\theta}_{\text{MLE}}$ is the value of θ that maximizes the likelihood. Thus, the optimal envelope constant is the likelihood evaluated at the MLE.

ii

For the model $X_i \sim \mathcal{N}(\theta, 1)$ and prior $\theta \sim \mathcal{C}(0, 1)$,
the posterior density is

$$\pi(\theta \mid x) \propto \frac{1}{\pi(1+\theta^2)}\, (2\pi)^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right].$$

The unnormalized posterior is

$$h(\theta) = L(\theta)\, \pi(\theta) \propto \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right] (1+\theta^2)^{-1}.$$

Using the Cauchy prior as the candidate distribution $g(\theta) = \pi(\theta)$,
the acceptance ratio simplifies to

$$\frac{h(\theta)}{M\, g(\theta)} = \frac{L(\theta)\, \pi(\theta)}{L(\hat{\theta})\, \pi(\theta)} = \frac{L(\theta)}{L(\hat{\theta})},$$

where $\hat{\theta} = \bar{x}$ is the MLE of θ.
Thus the acceptance probability is

$$\alpha(\theta) = \frac{L(\theta)}{L(\hat{\theta})} = \exp\left[-\tfrac{1}{2} n(\theta - \bar{x})^2\right].$$

Accept–Reject sampler:

```
post_AR <- function(x, N) {
  n <- length(x)
  xbar <- mean(x)
  out <- numeric(0)

  while (length(out) < N) {
    theta <- rcauchy(N)
    acc <- runif(N) < exp(-0.5 * n * (theta - xbar)^2)
    out <- c(out, theta[acc])
  }

  draws <- out[1:N]
  list(
    Interval = quantile(draws, c(0.025, 0.975)),
    mean = mean(draws)
  )
}

theta0 <- 3


n <- 10
x <- rnorm(n, mean = theta0, sd = 1)
post_AR(x, 100)
```

```
$Interval
    2.5%    97.5%
2.320014 3.613213

$mean
[1] 3.011534
```

```
n <- 1e3
x <- rnorm(n, mean = theta0, sd = 1)
post_AR(x, 100)
```

```
$Interval
```

```
    2.5%    97.5%
2.920040 3.046843

$mean
[1] 2.978843
```

```
n <- 1e4
x <- rnorm(n, mean = theta0, sd = 1)
post_AR(x, 100)
```

```
$Interval
    2.5%    97.5%
2.983763 3.022952

$mean
[1] 3.001527
```

As n increases, the posterior credible interval tightens around the true value

$$\theta_o = 3$$

and the posterior mean approaches 3.

Question 3

a (HANDWRITTEN)

3a

R has $f_R(r) = \frac{1}{8} e^{-r/8}, \quad r \geq 0$

with $r = \rho^2$ and $\frac{dr}{d\rho} = 2\rho$

$$f_\rho(\rho) = f_R(\rho^2)\,|2\rho| = \frac{1}{8} e^{-\rho^2/8} \cdot 2\rho = \frac{\rho}{4} e^{-\rho^2/8}$$

$$f_{\rho,\theta}(\rho,\theta) = f_\rho(\rho) f_\theta(\theta) = \left(\frac{\rho}{4} e^{-\rho^2/8}\right) \cdot \frac{1}{2\pi}, \quad \rho \geq 0, \quad \theta \in [0, 2\pi)$$

$X_1 = \rho\cos\theta \quad X_2 = \rho\sin\theta$

$$\left|\frac{\partial(X_1, X_2)}{\partial(\rho,\theta)}\right| = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix} = \rho$$

Joint density of X

w/ $\rho = \sqrt{X_1^2 + X_2^2}$

$$f_X(X_1, X_2) = \frac{f_{\rho,\theta}(\rho,\theta)}{|J|} = \frac{1}{8\pi} e^{\left(-\frac{X_1^2 + X_2^2}{8}\right)}, \quad (X_1, X_2) \in \mathbb{R}^2$$

$$\therefore \boxed{X \sim N_2\left(\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = 4 I_2\right)}$$

b (HANDWRITTEN)

3b

$X_2 \sim N_2(0, \Sigma_X)$, $\Sigma_X = 4I_2$

$cX \sim N(0, c^2\Sigma_X)$

$Z = cX \sim N(0, I_2)$

$c^2 \Sigma_x = I_2$

$\Sigma_X = 4I_2$

$c^2(4I_2) = I_2 \Rightarrow 4c^2 = 1 \Rightarrow c = \frac{1}{2}$

$Z = \frac{1}{2}X \sim N_2(0, I_2)$

$C = \frac{1}{2}$

c (HANDWRITTEN)

$$\mu = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix}, \quad Z \sim N_2(0, I_2)$$

3c

$$\mathbb{E}[Y] = b + A\,\mathbb{E}[Z] = b, \quad \mathrm{cov}(Y) = A\,\mathrm{cov}(Z)A^T = AA^T$$

$$b = \mu \quad AA^T = \Sigma$$

$$A = \begin{bmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{bmatrix} \quad AA^T = \begin{bmatrix} t_{11}^2 & t_{11}t_{21} \\ t_{11}t_{21} & t_{21}^2 + t_{22}^2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix}$$

$$t_{11}^2 = 4 \quad \sqrt{t_{11}^2} = 2 = t_{11}$$

$$t_{11}t_{21} = 2t_{21} = 2 \therefore t_{21} = 1$$

$$t_{21}^2 + t_{22}^2 = 10 = 1 + t_{22}^2 \therefore t_{22} = 3$$

Plug into A

$$A = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$AA^T = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix} = \Sigma$$

So $Y = b + AZ \sim N_2(\mu, \Sigma)$

d

```r
set.seed(102)

# (1) Generate U, V ~ Uniform(0,1)
N <- 5000
U <- runif(N)
V <- runif(N)

# (2)
Theta <- 2 * pi * U
R <- -8 * log(1 - V)

# (3)
X1 <- sqrt(R) * cos(Theta)
X2 <- sqrt(R) * sin(Theta)
X <- cbind(X1, X2)

# (4) Standardize
c <- 1/2
Z <- c * X

# (5)
mu <- c(1, -2)
A <- matrix(c(2, 0,
              1, 3), nrow = 2, byrow = TRUE)
Y <- t(t(Z %*% t(A)) + mu)

# (6
dim(X)
```

```
[1] 5000    2
```

```r
colMeans(X)
```

```
         X1          X2 
-0.02196563  0.05403042 
```

```r
cov(X)
```

```
              X1          X2
X1  3.96964084 -0.03715828
X2 -0.03715828  4.00140038
```

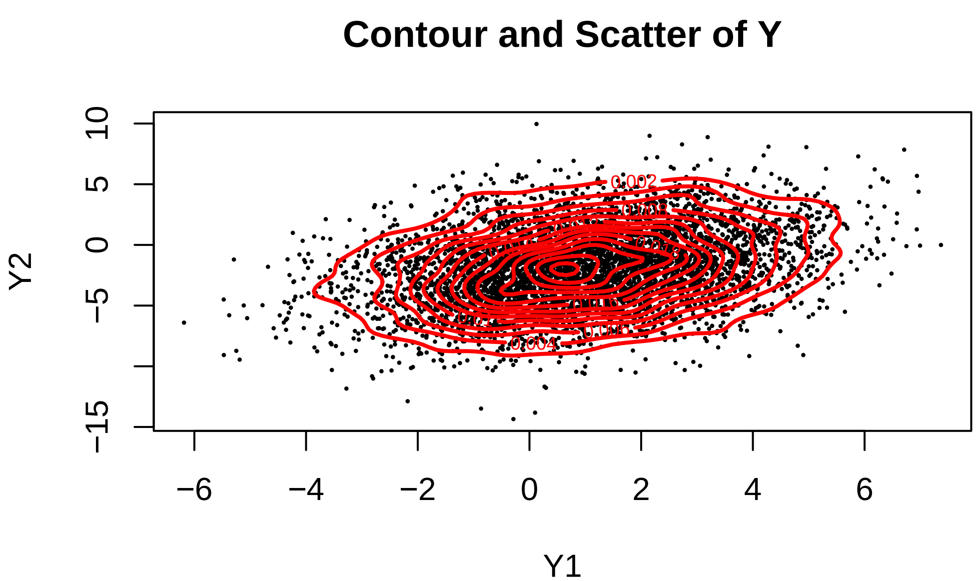
```r
colMeans(Y)
```

```
[1]  0.9780344 -1.9299372
```

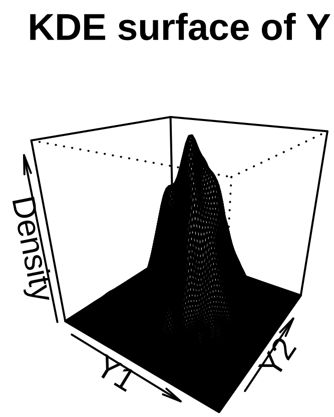
```r
cov(Y)
```

```
         [,1]     [,2]
[1,] 3.969641 1.929083
[2,] 1.929083 9.939824
```

```r
# (7) Visualize 2D kernel density of Y
if (!requireNamespace("MASS", quietly = TRUE)) install.packages("MASS")


# top: scatter with red KDE contours
plot(Y[,1], Y[,2], pch = 16, cex = 0.3,
     xlab = "Y1", ylab = "Y2", main = "Contour and Scatter of Y")
kd <- MASS::kde2d(Y[,1], Y[,2], n = 100)
contour(kd, add = TRUE, col = "red", lwd = 2)
```



```r
# bottom: 3D KDE surface
persp(kd, theta = 35, phi = 25,
      xlab = "Y1", ylab = "Y2", zlab = "Density",
      main = "KDE surface of Y")
```



The simulated values match the analytical results from parts a-c and the KDE plots visually confirm that Y roughly follows the expected elliptical (bivariate normal) shape.

Question 4

$$

$$

a

MGF Derivation

Let $X_i \overset{\text{iid}}{\sim} \text{Exp}(\beta)$ with rate $\beta > 0$, so $f_X(x) = \beta e^{-\beta x}$ for $x \geq 0$.

The MGF of X_i is

$$
\begin{aligned}
M_X(t) = \mathbb{E}[e^{tX}] &= \int_0^\infty e^{tx}\,\beta e^{-\beta x}\,dx \\
&= \beta \int_0^\infty e^{(t-\beta)x}\,dx.
\end{aligned}
$$

For convergence, $t < \beta$. Let $a = t - \beta < 0$, then

$$\int_0^\infty e^{ax}\,dx = \left[\frac{e^{ax}}{a}\right]_0^\infty = -\frac{1}{a}.$$

Substitute back:

$$M_X(t) = \beta\left(-\frac{1}{t-\beta}\right) = \frac{\beta}{\beta - t}, \quad t < \beta.$$

Let $S = \sum_{i=1}^{\alpha} X_i$, with $\alpha \in \mathbb{N}$. Since the X_i are independent,

$$M_S(t) = \prod_{i=1}^{\alpha} M_{X_i}(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha}, \quad t < \beta.$$

This matches the MGF of $\Gamma(\alpha, \beta)$, so

$$S = X_1 + X_2 + \cdots + X_\alpha \sim \Gamma(\alpha, \beta).$$

b

If $X \sim \Gamma(\alpha, \beta)$, then

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0.$$

$Y = \beta X$.

Then $x = y/\beta$ and $\frac{dx}{dy} = \frac{1}{\beta}$.

$$f_Y(y) = f_X\left(\frac{y}{\beta}\right)\left|\frac{dx}{dy}\right| = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{y}{\beta}\right)^{\alpha-1} e^{-\beta(y/\beta)} \cdot \frac{1}{\beta}.$$

$$f_Y(y) = \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}, \quad y > 0.$$

$$Y = \beta X \sim \Gamma(\alpha, 1).$$

In particular, if $X \sim \mathrm{Exp}(\beta) = \Gamma(1, \beta)$, then $Y = \beta X \sim \Gamma(1, 1) = \mathrm{Exp}(1)$.

From part (a), if

$$S = \sum_{i=1}^{\alpha} X_i \sim \Gamma(\alpha, \beta),$$

then multiplying each term by beta gives

$$\sum_{i=1}^{\alpha} Y_i = \sum_{i=1}^{\alpha} \beta X_i = \beta S \sim \Gamma(\alpha, 1).$$

c

$f(x) = \dfrac{x^{n-1}e^{-x}}{\Gamma(n)}$ and $g(x) = \lambda e^{-\lambda x}$.

$f(x) \le M g(x)$ for all $x > 0$. Up to constants, with the ratio:

$$\frac{f(x)}{g(x)} \propto \frac{x^{n-1}e^{-x}}{\lambda e^{-\lambda x}} = \frac{1}{\lambda}\, x^{n-1} e^{-(1-\lambda)x}.$$

$$\ell(x) = \log\left(\frac{f(x)}{g(x)}\right) = (n-1)\log x - (1-\lambda)x + \text{const}.$$

Differentiate and set = to zero

$$\ell'(x) = \frac{n-1}{x} - (1-\lambda) = 0 \quad \Longrightarrow \quad x^* = \frac{n-1}{1-\lambda}.$$

Check $\ell''(x) = -\frac{n-1}{x^2} < 0$ for $n > 1$, confirming a maximum. Since $x^* > 0$, we must have $\lambda < 1$.

Substitute x^* back into the unnormalized ratio:

$$\frac{f(x)}{g(x)} \propto \frac{1}{\lambda}\,(x^*)^{n-1} e^{-(1-\lambda)x^*} = \frac{1}{\lambda}\left(\frac{n-1}{1-\lambda}\right)^{n-1} \exp\big[-(1-\lambda)\tfrac{n-1}{1-\lambda}\big] = \frac{1}{\lambda}\left(\frac{n-1}{1-\lambda}\right)^{n-1} e^{-(n-1)}.$$

Up to constants,

$$M(\lambda) \propto \frac{1}{\lambda}\left(\frac{n-1}{1-\lambda}\right)^{n-1}.$$

To minimize $M(\lambda)$, take logs and differentiate:

$$\log M(\lambda) = -\log\lambda + (n-1)\big[\log(n-1) - \log(1-\lambda)\big] + \text{const},$$

$$\frac{d}{d\lambda}\log M(\lambda) = -\frac{1}{\lambda} + \frac{n-1}{1-\lambda} = 0.$$

Solve for λ:

$$(n-1)\lambda = 1-\lambda \quad \Longrightarrow \quad \lambda^* = \frac{1}{n}.$$

$$\boxed{\lambda^* = \frac{1}{n}.}$$

$$

$$

d

Target and candidate densities (up to constants):

$$f(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, \qquad g(x) = \frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}, \qquad a \leq \alpha.$$

Form the ratio (ignoring constants that do not depend on x):

$$\frac{f(x)}{g(x)} \propto x^{\alpha-a}\, e^{-(1-b)x}.$$

As $x \to 0^+$,

$$\frac{f(x)}{g(x)} \approx x^{\alpha-a}.$$

This is **finite** only if $\alpha - a \geq 0$, i.e.

$$a \leq \alpha.$$

If $a > \alpha$, the ratio $\to \infty$ at zero and no valid envelope exists.

As $x \to \infty$,

$$\frac{f(x)}{g(x)} \approx e^{-(1-b)x}.$$

For this to remain bounded, we require the exponential term to decay, which happens when

$$b < 1.$$

If $b > 1$, the ratio $\to \infty$ as $x \to \infty$.

Choose parameters satisfying

$$a \leq \alpha \quad \text{and} \quad b < 1,$$

then the ratio $\frac{f(x)}{g(x)}$ is bounded over all $x > 0$, so a finite M exists and the Accept–Reject algorithm can be applied.

Therefore, we can use a $\Gamma(a, b)$ proposal to generate $\Gamma(\alpha, 1)$ samples whenever

$$a \leq \alpha \text{ and } b < 1.$$

e

From d we have

$$r(x) := \frac{f(x)}{g(x)} \propto b^{-a} x^{\alpha-a} \exp\{-(1-b)x\}.$$

Let $\ell(x) = \log r(x)$, so

$$\ell(x) = -a \log b + (\alpha - a) \log x - (1-b)x.$$

Differentiate and set = to zero

$$\ell'(x) = \frac{\alpha - a}{x} - (1-b) = 0 \quad \Longrightarrow \quad x^* = \frac{\alpha - a}{1-b}.$$

Since

$$\ell''(x) = -\frac{\alpha - a}{x^2} < 0 \quad (\alpha > a),$$

x^* is a maximizer.

Evaluate the ratio at x^*:

$$r(x^*) \propto b^{-a} \left(\frac{\alpha - a}{1-b}\right)^{\alpha-a} \exp\{-(1-b)\tfrac{\alpha-a}{1-b}\} = b^{-a} \left(\frac{\alpha - a}{(1-b)e}\right)^{\alpha-a}.$$

The bound is

$$\boxed{M = b^{-a} \left(\frac{\alpha - a}{(1-b)e}\right)^{\alpha-a}, \quad b < 1.}$$

f

Let

$$F(b) = b^{-a}(1-b)^{a-\alpha}, \quad b \in (0,1),\ 0 < a \leq \alpha.$$

minimize (F(b)) to minimize the rejection constant (M(b)), which maximia the acceptance rate.

$$\ell(b) = \log F(b) = -a \log b + (a - \alpha) \log(1-b).$$

$$\ell'(b) = -\frac{a}{b} + \frac{a-\alpha}{1-b} = 0.$$

Solve for (b):

$$-a(1-b) + b(a-\alpha) = 0 \implies b = \frac{a}{\alpha}.$$

Convexity

$$\ell''(b) = \frac{a}{b^2} + \frac{a-\alpha}{(1-b)^2} > 0, \quad 0 < b < 1,\ 0 < a \le \alpha.$$

$$\boxed{b^* = \frac{a}{\alpha}}.$$

At this value, the proposal and target have the same mean:

$$\mathbb{E}_g[X] = \frac{a}{b^*} = \frac{a}{a/\alpha} = \alpha, \qquad \mathbb{E}_f[X] = \alpha.$$

g

Among integer values of a, the best choice is

$$\boxed{a = \lfloor \alpha \rfloor.}$$

This is because: The Accept–Reject algorithm requires $a \le \alpha$. The acceptance rate improves as a approaches α from below (since the proposal's shape more closely matches the target's tail behavior). Using $a = \lfloor \alpha \rfloor$ ensures both conditions are met while keeping a an integer when sampling from $\Gamma(a, b)$.

Hence, $a = \lfloor \alpha \rfloor$ yields the most efficient integer-valued proposal.

h

When $\alpha < 1$, direct generation from $\Gamma(\alpha, 1)$ is inefficient.
We use the transformation

$$X = YU^{1/\alpha},$$

where $Y \sim \Gamma(\alpha + 1, 1)$ and $U \sim \mathcal{U}(0, 1)$ are independent.

Let $T = U^{1/\alpha} \in (0, 1)$.
Then

$$F_T(t) = P(T \le t) = P(U^{1/\alpha} \le t) = P(U \le t^\alpha) = t^\alpha, \quad 0 < t < 1.$$

Differentiate to find the PDF:

$$f_T(t) = \alpha t^{\alpha-1}, \quad 0 < t < 1.$$

Now, conditional on $Y = y$, define $X = yT$.
Then the conditional PDF is

$$f_{X|Y}(x|y) = f_T\left(\frac{x}{y}\right)\frac{1}{y} = \alpha y^{-\alpha} x^{\alpha-1}, \quad 0 < x < y.$$

The marginal PDF of X is

$$f_X(x) = \int_x^\infty f_{X|Y}(x|y) f_Y(y)\,dy = \int_x^\infty \alpha x^{\alpha-1} y^{-\alpha} \frac{y^\alpha e^{-y}}{\Gamma(\alpha+1)}\,dy = \frac{\alpha x^{\alpha-1} e^{-x}}{\Gamma(\alpha+1)}.$$

Using $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

$$f_X(x) = \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}, \quad x > 0.$$

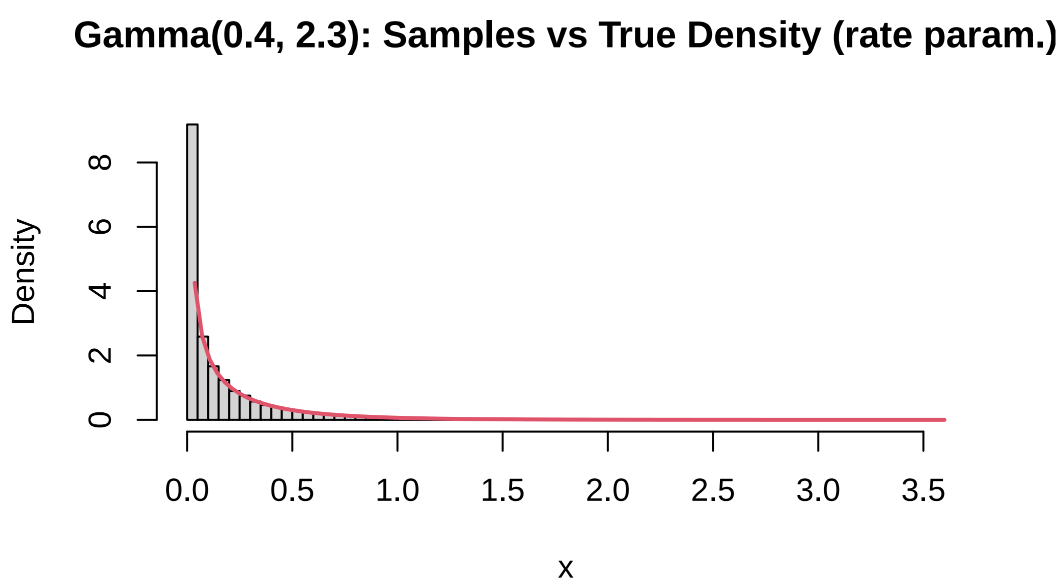
Hence,

$$\boxed{X \sim \Gamma(\alpha, 1)}.$$

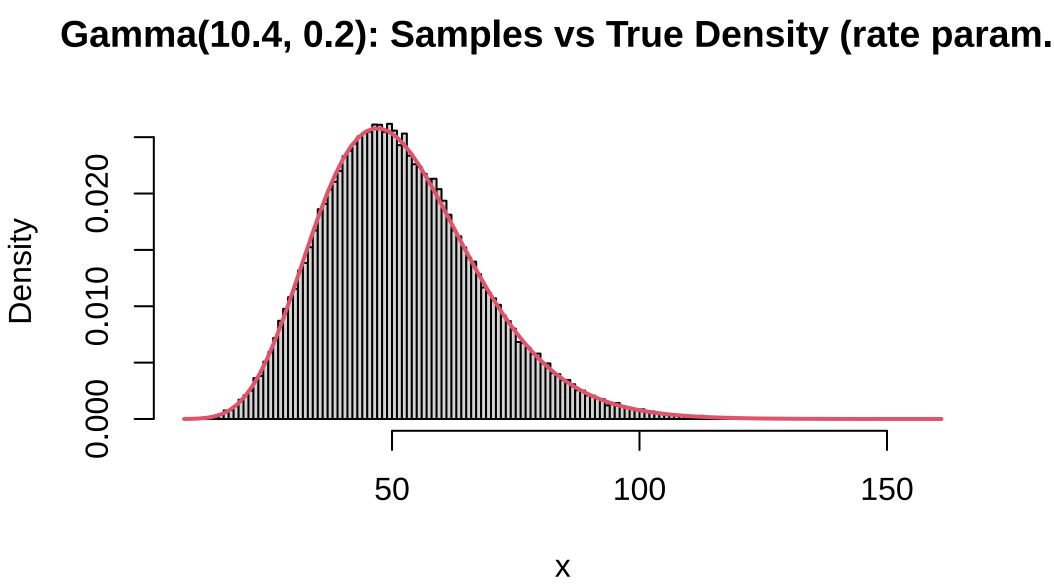
i

```r
set.seed(102)

rgamma_ar_exp <- function(m, alpha, lambda = 1/alpha) {
  out <- numeric(0)
  f_x <- function(x) x^(alpha - 1) * exp(-x) / gamma(alpha)
  g_x <- function(x) lambda * exp(-lambda * x)
  opt_x <- (alpha - 1) / (1 - lambda)
  M <- f_x(opt_x) / g_x(opt_x)
  while (length(out) < m) {
    xprop <- rexp(m, rate = lambda)
    r <- f_x(xprop) / (M * g_x(xprop))
    u <- runif(m)
    keep <- u <= r
    out <- c(out, xprop[keep])
  }
  out[1:m]
```

```r
}

gen_gamma <- function(n, alpha, beta) {
  if (alpha <= 0 || beta <= 0) stop("alpha, beta must be > 0 (rate parameterization).")
  if (alpha > 1) {
    y <- rgamma_ar_exp(n, alpha = alpha, lambda = 1 / alpha)
    return(y / beta)
  }
  y <- rgamma_ar_exp(n, alpha = alpha + 1, lambda = 1 / (alpha + 1))
  u <- runif(n)
  x <- y * u^(1 / alpha)
  return(x / beta)
}

set.seed(80025)
n <- 1e5
alpha <- 0.4
beta <- 2.3
x_a <- gen_gamma(n, alpha, beta)
hist(x_a, breaks = 120, probability = TRUE,
     main = "Gamma(0.4, 2.3): Samples vs True Density (rate param.)",
     xlab = "x")
curve(dgamma(x, shape = alpha, rate = beta), add = TRUE, col = 2, lwd = 2)
```



```r
set.seed(95939)
alpha <- 10.4
beta <- 0.2
x_b <- gen_gamma(n, alpha, beta)
hist(x_b, breaks = 120, probability = TRUE,
     main = "Gamma(10.4, 0.2): Samples vs True Density (rate param.)",
     xlab = "x")
curve(dgamma(x, shape = alpha, rate = beta), add = TRUE, col = 2, lwd = 2)
```



The samples match the true densities nearly perfectly

Question 5

a

Let $Z \sim \mathcal{N}(0,1)$ with pdf

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

We want

$$\mathbb{E}_\phi[h(Z)] = \int_{-\infty}^{\infty} \phi(z) \left(e^{-\frac{(z-3)^2}{2}} + e^{-\frac{(z-6)^2}{2}} \right) dz = I(3) + I(6),$$

where

$$I(a) = \int_{-\infty}^{\infty} \phi(z)\, e^{-(z-a)^2/2}\, dz.$$

Substitute $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$:

$$I(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\Big[-\frac{1}{2}(z^2 + (z-a)^2)\Big]\, dz.$$

Simplify the exponent:

$$z^2 + (z-a)^2 = 2z^2 - 2az + a^2 = 2\left[(z - a/2)^2 + a^2/4\right].$$

Hence

$$I(a) = \frac{1}{\sqrt{2\pi}} e^{-a^2/4} \int_{-\infty}^{\infty} e^{-(z-a/2)^2} dz.$$

Now use the change of variables:

$$t = z - \frac{a}{2} \quad \Rightarrow \quad dz = dt.$$

Then

$$\int_{-\infty}^{\infty} e^{-(z-a/2)^2} dz = \int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}.$$

Thus

$$I(a) = \frac{1}{\sqrt{2\pi}} \sqrt{\pi}\, e^{-a^2/4} = \frac{1}{\sqrt{2}}\, e^{-a^2/4}.$$

Finally,

$$\mathbb{E}_\phi[h(Z)] = I(3) + I(6) = \frac{1}{\sqrt{2}} \left(e^{-9/4} + e^{-9}\right).$$

$$\boxed{\mathbb{E}_\phi[h(Z)] = \frac{1}{\sqrt{2}}(e^{-9/4} + e^{-9}) = 0.0747}$$

b

$$\hat{\mathbb{E}}_f[h(X)] = \frac{1}{N} \sum_{i=1}^{N} h(X_i), \quad X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1).$$

We estimate this using $N_{\text{sim}} = 10^3$ samples and compare it to the true value from part (a).

```r
set.seed(80025)
Nsim <- 1e3


h <- function(x) {
  exp(-(x - 3)^2 / 2) + exp(-(x - 6)^2 / 2)
}


x <- rnorm(Nsim)

mc_hat <- mean(h(x))

true_val <- (1 / sqrt(2)) * (exp(-9/4) + exp(-9))

abs_err <- abs(mc_hat - true_val)

list(mc_hat = mc_hat, true = true_val, abs_err = abs_err)
```

```
$mc_hat
[1] 0.06170153

$true
[1] 0.07461577

$abs_err
[1] 0.01291424
```

$$\mathbb{E}_f[h(X)] = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx$$

where $f = \phi = \mathcal{N}(0, 1)$.

Using importance sampling with $g = \mathcal{U}(-8, -1)$ (density $g(x) = 1/7$ on $(-8, -1)$, 0 otherwise), the estimator is

$$\hat{\mathbb{E}}_f[h(X)] = \frac{1}{N} \sum_{i=1}^{N} h(Y_i)\, \frac{f(Y_i)}{g(Y_i)}, \qquad Y_i \overset{\text{iid}}{\sim} g.$$

Since $g(x) = 0$ for $x \notin (-8, -1)$ but $f(x)h(x) > 0$ for all $x \in \mathbb{R}$, this estimator actually converges to

$$\int_{-8}^{-1} h(x)\, f(x)\, dx$$

(the truncated integral), not the true

$$\int_{-\infty}^{\infty} h(x)\, f(x)\, dx.$$

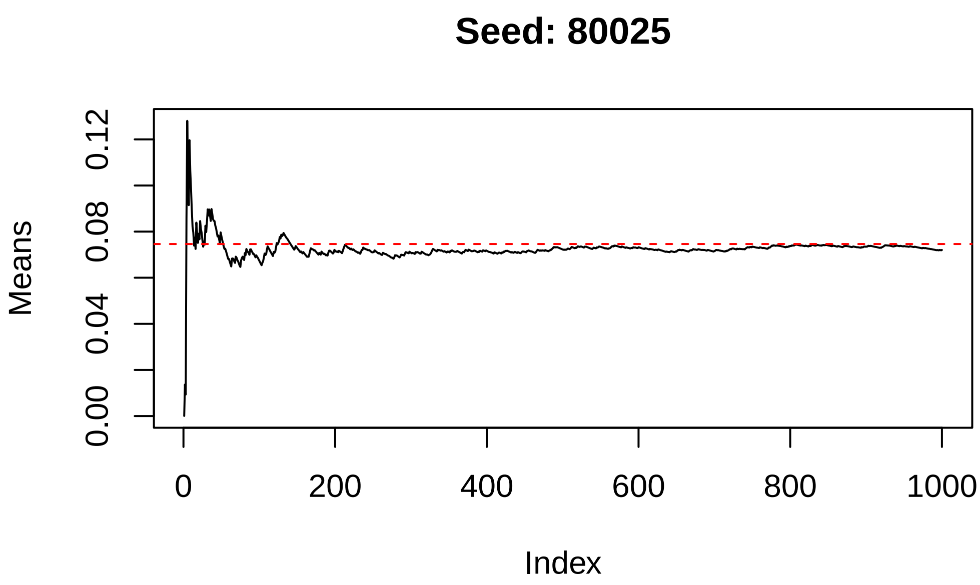
Hence it is biased

```
set.seed(67)

Nsim <- 1e3
h <- function(x) exp(-(x - 3)^2 / 2) + exp(-(x - 6)^2 / 2)
phi <- function(x) dnorm(x)
true_val <- (1 / sqrt(2)) * (exp(-9/4) + exp(-9))

importance_sample_fn <- function(seed) {
  set.seed(seed)
  y <- runif(Nsim, min = -8, max = -1)             # Y ~ U(-8,-1)
  u <- (y + 8) / 7                                 # Map to U ~ U(0,1)
  x_y <- log(u / (1 - u))                          # Logit transform X = log(U/(1-U))
  dx_dy <- -7 / ((y + 8) * (y + 1))                # Jacobian
  f_x_y <- phi(x_y)                                # target density f(x(y))
  g_y <- 1 / 7                                     # uniform g(y)
  w <- h(x_y) * f_x_y / g_y * dx_dy
  is_mean <- mean(w)
  means <- cumsum(w) / seq_along(w)                # cumulative means
  list(means = means, result = is_mean)
}

out1 <- importance_sample_fn(80025)
out2 <- importance_sample_fn(95939)

list(
  IS_hat_seed1 = out1$result,
  IS_hat_seed2 = out2$result,
```

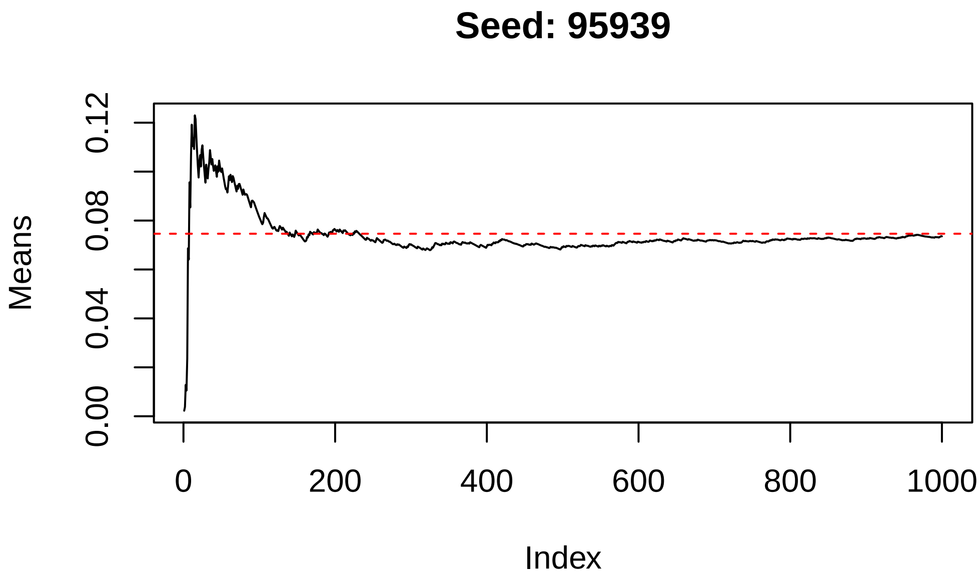
```
  true = true_val
)
```

```
$IS_hat_seed1
[1] 0.07203004

$IS_hat_seed2
[1] 0.07355336

$true
[1] 0.07461577
```

```
plot(out1$means, type = "l", ylab = "Means", main = "Seed: 80025")
abline(h = true_val, lty = 2, col = "red")
```




```
plot(out2$means, type = "l", ylab = "Means", main = "Seed: 95939")
abline(h = true_val, lty = 2, col = "red")
```


Seed: 95939
Means
0.00
0.04
0.08
0.12
0
200
400
600
800
1000
Index