Topics in Statistics: Markov chain Monte Carlo
Problem Set 1

Carsten Jørgensen

September 2025

Exercise 1 - Inverse transformation and rejection sampling

Question 1

Let $Y \sim \text{Exp}(\lambda)$ and $X = Y \mid Y \geq a$ where $a > 0$.
CDF of X:

$$F_X(x) = P(X \leq x) = P(Y \leq x \mid Y \geq a) = \frac{P(a \leq Y \leq x)}{P(Y \geq a)}$$

Since $F_Y(y) = 1 - e^{-\lambda y}$ for $y \geq 0$:

- For $x < a$: $F_X(x) = 0$
- For $x \geq a$:

$$F_X(x) = \frac{F_Y(x) - F_Y(a)}{1 - F_Y(a)} = \frac{(1 - e^{-\lambda x}) - (1 - e^{-\lambda a})}{e^{-\lambda a}} = 1 - e^{-\lambda(x-a)}$$

Therefore:

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ 1 - e^{-\lambda(x-a)} & \text{if } x \geq a \end{cases}$$

Quantile function: Solving $u = F_X(x) = 1 - e^{-\lambda(x-a)}$:

$$F_X^{-1}(u) = a - \frac{1}{\lambda}\ln(1-u)$$

Algorithm:

1. Generate $U \sim \mathcal{U}[0,1]$
2. Return $X = a - \frac{1}{\lambda}\ln(1-U)$

Question 2

We need to show that $X = F_Y^{-1}(F_Y(a)(1-U) + F_Y(b)U)$ has distribution $Y \mid a \leq Y \leq b$.
Let $\alpha = F_Y(a)$ and $\beta = F_Y(b)$. Then:

$$X = F_Y^{-1}(\alpha + U(\beta - \alpha))$$

Since $U \sim \mathcal{U}[0,1]$, we have $V = \alpha + U(\beta - \alpha) \sim \mathcal{U}[\alpha, \beta]$.
For $a \leq x \leq b$:

$$\begin{aligned} P(X \leq x) &= P(F_Y^{-1}(V) \leq x) = P(V \leq F_Y(x)) \\ &= \frac{F_Y(x) - \alpha}{\beta - \alpha} = \frac{F_Y(x) - F_Y(a)}{F_Y(b) - F_Y(a)} \end{aligned}$$

This equals $P(Y \le x \mid a \le Y \le b)$ by definition.
Detailed explanation of why this equals $P(Y \le x \mid a \le Y \le b)$:
By the definition of conditional probability:

$$P(Y \le x \mid a \le Y \le b) = \frac{P(Y \le x \text{ and } a \le Y \le b)}{P(a \le Y \le b)}$$

Since we're considering $a \le x \le b$, the event $\{Y \le x \text{ and } a \le Y \le b\}$ simplifies to $\{a \le Y \le x\}$. Therefore:

$$\begin{aligned}
P(Y \le x \mid a \le Y \le b) &= \frac{P(a \le Y \le x)}{P(a \le Y \le b)} \\
&= \frac{P(Y \le x) - P(Y < a)}{P(Y \le b) - P(Y < a)} \\
&= \frac{P(Y \le x) - P(Y \le a)}{P(Y \le b) - P(Y \le a)} \quad \text{(since } Y \text{ is continuous)} \\
&= \frac{F_Y(x) - F_Y(a)}{F_Y(b) - F_Y(a)}
\end{aligned}$$

This matches exactly what we obtained for $P(X \le x)$, confirming that X has the distribution of Y conditioned on $a \le Y \le b$.
Application to exponential: Taking $b \to \infty$:

$$\begin{aligned}
X &= F_Y^{-1}((1 - e^{-\lambda a})(1 - U) + U) \\
&= -\frac{1}{\lambda} \ln(e^{-\lambda a}(1 - U)) \\
&= a - \frac{1}{\lambda} \ln(1 - U)
\end{aligned}$$

For the exponential distribution, we have:

- $F_Y(y) = 1 - e^{-\lambda y}$ for $y \ge 0$
- $F_Y(a) = 1 - e^{-\lambda a}$
- As $b \to \infty$: $F_Y(b) = 1 - e^{-\lambda b} \to 1$

Substituting into the formula $X = F_Y^{-1}(F_Y(a)(1 - U) + F_Y(b)U)$:

$$\begin{aligned}
X &= F_Y^{-1}((1 - e^{-\lambda a})(1 - U) + 1 \cdot U) \\
&= F_Y^{-1}(1 - e^{-\lambda a} - U + Ue^{-\lambda a} + U) \\
&= F_Y^{-1}(1 - e^{-\lambda a} + Ue^{-\lambda a}) \\
&= F_Y^{-1}(1 - e^{-\lambda a}(1 - U))
\end{aligned}$$

Now, for the exponential distribution, the inverse CDF is:

$$F_Y^{-1}(p) = -\frac{1}{\lambda} \ln(1 - p)$$

Therefore:

$$\begin{aligned}
X &= -\frac{1}{\lambda} \ln(1 - (1 - e^{-\lambda a}(1 - U))) \\
&= -\frac{1}{\lambda} \ln(e^{-\lambda a}(1 - U)) \\
&= -\frac{1}{\lambda} [\ln(e^{-\lambda a}) + \ln(1 - U)] \\
&= -\frac{1}{\lambda} [-\lambda a + \ln(1 - U)] \\
&= a - \frac{1}{\lambda} \ln(1 - U)
\end{aligned}$$

Like in Part 1.

Question 3

The rejection algorithm simulates from:

- **Proposal:** $q(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0,\infty)}(x)$ (standard exponential)
- **Target:** $\pi(x) = \lambda e^{-\lambda(x-a)}\mathbf{1}_{[a,\infty)}(x)$ (truncated exponential)

The ratio for $x \geq a$:

$$\frac{\pi(x)}{q(x)} = \frac{\lambda e^{-\lambda(x-a)}}{\lambda e^{-\lambda x}} = e^{\lambda a}$$

Therefore:

$$M = e^{\lambda a}$$

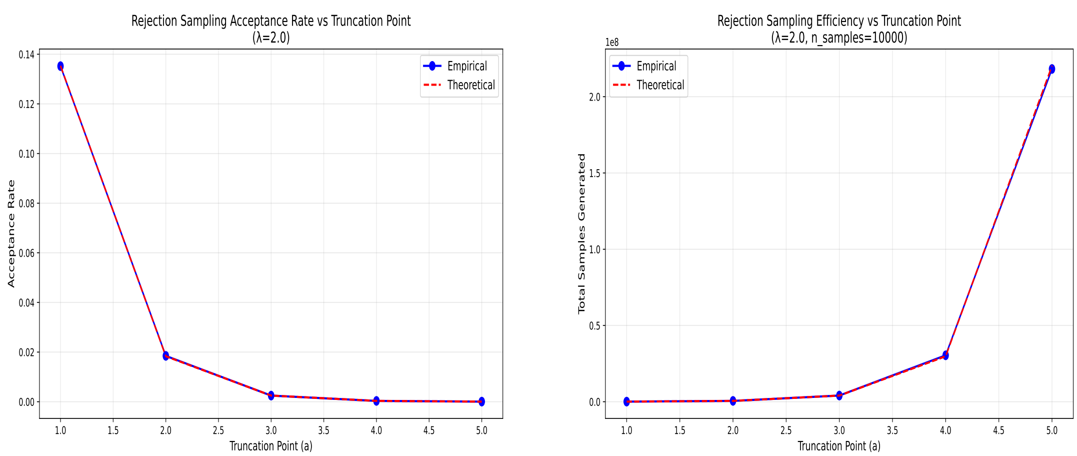
The acceptance probability is $\frac{1}{M} = e^{-\lambda a} = P(Y > a)$.
Expected number of trials:

$$E[\text{trials}] = e^{\lambda a}$$

Why inversion is preferred for $a \gg 1/\lambda$:
When $a \gg 1/\lambda$, we have $\lambda a \gg 1$, so $e^{\lambda a} \gg 1$. This means:

- Rejection algorithm: Expected $e^{\lambda a}$ trials (grows exponentially with a)
- Inversion method: Exactly 1 step regardless of a

Therefore, inversion is much more efficient for large values of a.



(a) Rejection

(b) Rejection Sampling Efficiency

Figure 1: Why inversion is preferred for $a \gg 1/\lambda$

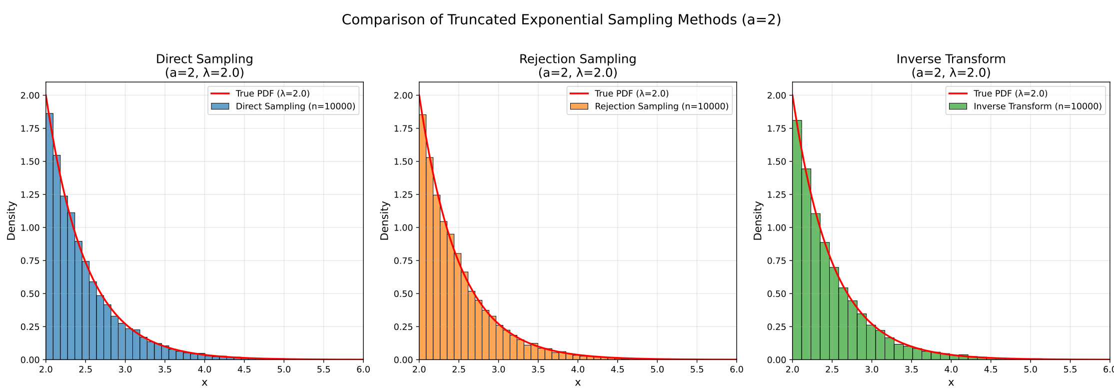


Figure 2: Truncated Exponential Comparison

Exercise 2 - Transformation methods

Question 1: Independence and Distributions of R and S

Given: $X_1 \sim \text{Gamma}(a, 1)$ and $X_2 \sim \text{Gamma}(b, 1)$ are independent.

To prove: $R = \frac{X_1}{X_1+X_2}$ and $S = X_1 + X_2$ are independent, with $R \sim \text{Beta}(a, b)$ and $S \sim \text{Gamma}(a + b, 1)$.
The joint density of (X_1, X_2) is:

$$f_{X_1,X_2}(x_1, x_2) = \frac{x_1^{a-1} x_2^{b-1} e^{-(x_1+x_2)}}{\Gamma(a)\Gamma(b)}$$

Using the transformation $R = \frac{X_1}{X_1+X_2}$, $S = X_1 + X_2$:

$$\begin{aligned} X_1 &= RS \\ X_2 &= S(1 - R) \end{aligned}$$

The Jacobian is:

$$J = \begin{vmatrix} s & r \\ -s & 1-r \end{vmatrix} = s$$

The joint density of (R, S) becomes:

$$\begin{aligned} f_{R,S}(r, s) &= \frac{(rs)^{a-1}[s(1-r)]^{b-1}e^{-s}}{\Gamma(a)\Gamma(b)} \cdot s \\ &= \frac{s^{a+b-1}r^{a-1}(1-r)^{b-1}e^{-s}}{\Gamma(a)\Gamma(b)} \\ &= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} r^{a-1}(1-r)^{b-1}\right] \cdot \left[\frac{s^{a+b-1}e^{-s}}{\Gamma(a+b)}\right] \end{aligned}$$

This factors as $f_R(r) \cdot f_S(s)$, proving independence with $R \sim \text{Beta}(a, b)$ and $S \sim \text{Gamma}(a + b, 1)$.

Question 2: Distribution of $X = U^{1/a}$

Given: $U \sim \text{U}[0, 1]$ and $a > 0$.

To prove: $X = U^{1/a} \sim \text{Beta}(a, 1)$.
For $x \in (0, 1)$:

$$P(X \leq x) = P(U^{1/a} \leq x) = P(U \leq x^a) = x^a$$

The PDF is:

$$f_X(x) = \frac{d}{dx}[x^a] = ax^{a-1}$$

This matches the Beta$(a, 1)$ density: $\frac{\Gamma(a+1)}{\Gamma(a)\Gamma(1)} x^{a-1} = ax^{a-1}$.

Question 3: Conditional Distribution

For $t \in (0, 1)$, the event

$$Y/(Y + Z) \leq t, Y + Z \leq 1$$

is equivalent to

$$(y, z) : y \leq (t/(1-t))z, y \leq 1 - z, y > 0, z > 0$$

.
Splitting at $z = 1 - t$ where $(t/(1 - t))z = 1 - z$:

$$P(Y/(Y + Z) \leq t, Y + Z \leq 1) = \int_{z=0}^{1-t} \int_{y=0}^{(t/(1-t))z} a(1-a)y^{a-1}z^{-a} dydz$$

$$\int_{z=1-t}^{1}\int_{y=0}^{1-z} a(1-a)y^{a-1}z^{-a}dydz = (1-a)t^a(1-t)^{1-a} + (1-a)\int_0^t w^a(1-w)^{-a}dw.$$

In particular,

$$P(Y+Z\le 1) = P(Y/(Y+Z)\le 1, Y+Z\le 1) = a(1-a)\int_0^1 w^{a-1}(1-w)^{-a}dw = a(1-a)B(a,1-a).$$

Differentiate the function of t above: $d/dtP(Y/(Y+Z)\le t, Y+Z\le 1) = a(1-a)t^{a-1}(1-t)^{-a}$.
Therefore, the conditional density of $W = Y/(Y+Z)$ given $Y+Z\le 1$ is

$$f_{W|Y+Z\le 1}(t) = [a(1-a)t^{a-1}(1-t)^{-a}]/[a(1-a)B(a,1-a)] = [1/B(a,1-a)]t^{a-1}(1-t)^{-a}, 0<t<1,$$

which is the Beta(a, 1-a) density. Hence $W|(Y+Z\le 1)\sim$ Beta(a, 1-a).

Question 4

Question 5

This procedure implements **Johnk's algorithm** for generating Gamma$(a,1)$ random variables when $a\in(0,1)$. See page 418 in Non-Uniform Random Variate Generation by L. Devroye, Springer-Verlag, 1986.
Analysis of the algorithm:

1. Steps (a)-(c) generate (Y,Z) conditional on $Y+Z\le 1$ where:
   - $Y = U^{1/a}$ and $Z = V^{1/(1-a)}$ for independent $U,V\sim\mathcal{U}[0,1]$
   - From Question 3, we know that $W = \frac{Y}{Y+Z} \mid Y+Z\le 1 \sim \text{Beta}(a,1-a)$
2. Step (d) generates $T\sim\text{Exp}(1) = \text{Gamma}(1,1)$ since:

$$T = -\log(A)\sim\text{Exp}(1) \text{ when } A\sim\mathcal{U}[0,1]$$

3. Step (e) returns $TW = T\cdot\frac{Y}{Y+Z}$, which from Question 4 has distribution:

$$TW \mid Y+Z\le 1\sim\text{Gamma}(a,1)$$

Relevance for simulations: This algorithm is particularly useful because:

- It generates Gamma$(a,1)$ for $a\in(0,1)$, which cannot be done by summing exponentials
- It only requires uniform random variables
- The acceptance probability is $\frac{1}{B(a,1-a)} = \frac{\Gamma(a)\Gamma(1-a)}{\Gamma(1)} = \Gamma(a)\Gamma(1-a)$
- Combined with the additivity property of Gamma distributions, it enables generation of Gamma$(\alpha,1)$ for any $\alpha>0$

Question 6

To generate Beta(a,b) for any $a>0$ and $b>0$:
Method: Using the results from Question 1 and 5:
If $X_1\sim\text{Gamma}(a,1)$, $X_2\sim\text{Gamma}(b,1)$ independent $\Rightarrow R = \frac{X_1}{X_1+X_2}\sim\text{Beta}(a,b)$

Algorithm:

1. **Generate $X_1\sim$ Gamma$(a,1)$:**

   Decompose $a = \lfloor a\rfloor + \{a\}$ where $\lfloor a\rfloor$ is the integer part and $\{a\}\in[0,1)$ is the fractional part.

- If $\lfloor a \rfloor > 0$: Generate $G_1 = -\sum_{i=1}^{\lfloor a \rfloor} \log(U_i)$ where $U_i \sim \mathcal{U}[0,1]$ - this is Example 3.1 from the lecture notes.
- If $\{a\} > 0$: Use Johnk's algorithm (Question 5) to generate $G_2 \sim \text{Gamma}(\{a\}, 1)$
- Set $X_1 = G_1 + G_2$ (using additivity: $\text{Gamma}(a_1, 1) + \text{Gamma}(a_2, 1) = \text{Gamma}(a_1 + a_2, 1)$)

2. **Generate $X_2 \sim$ Gamma$(b, 1)$:**
   Apply the same decomposition method for $b = \lfloor b \rfloor + \{b\}$.

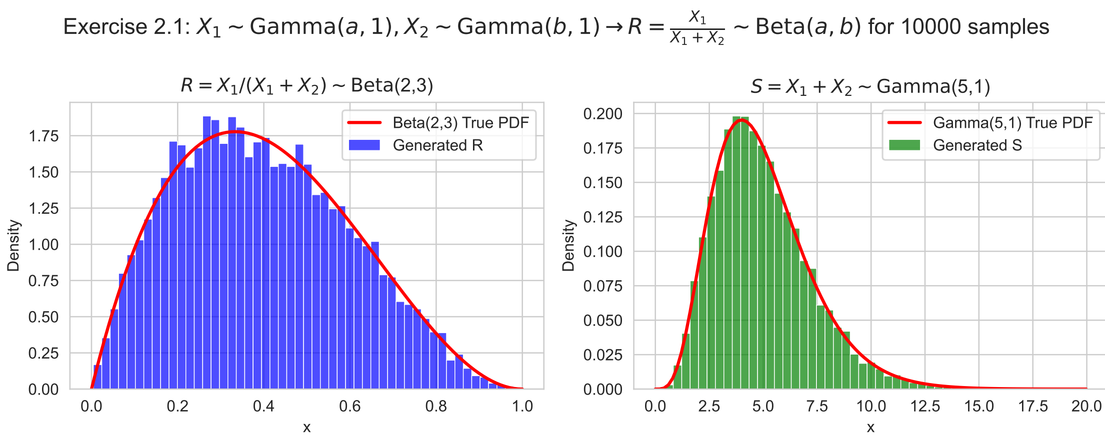
3. **Return:** $R = \frac{X_1}{X_1 + X_2}$

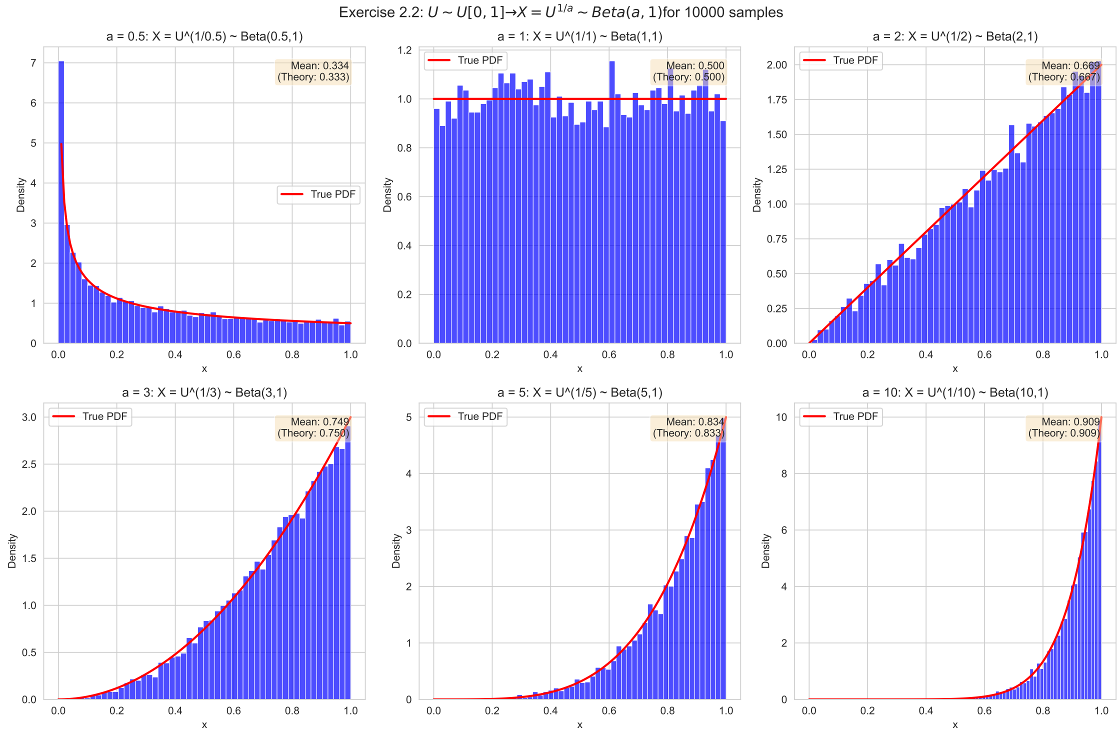
Special cases:

- If $a \in \mathbb{N}$: $X_1 = -\sum_{i=1}^{a} \log(U_i)$ directly
- If $a \in (0, 1)$: Use Johnk's algorithm directly
- For general $a = n + \alpha$ where $n \in \mathbb{N}$ and $\alpha \in (0, 1)$:

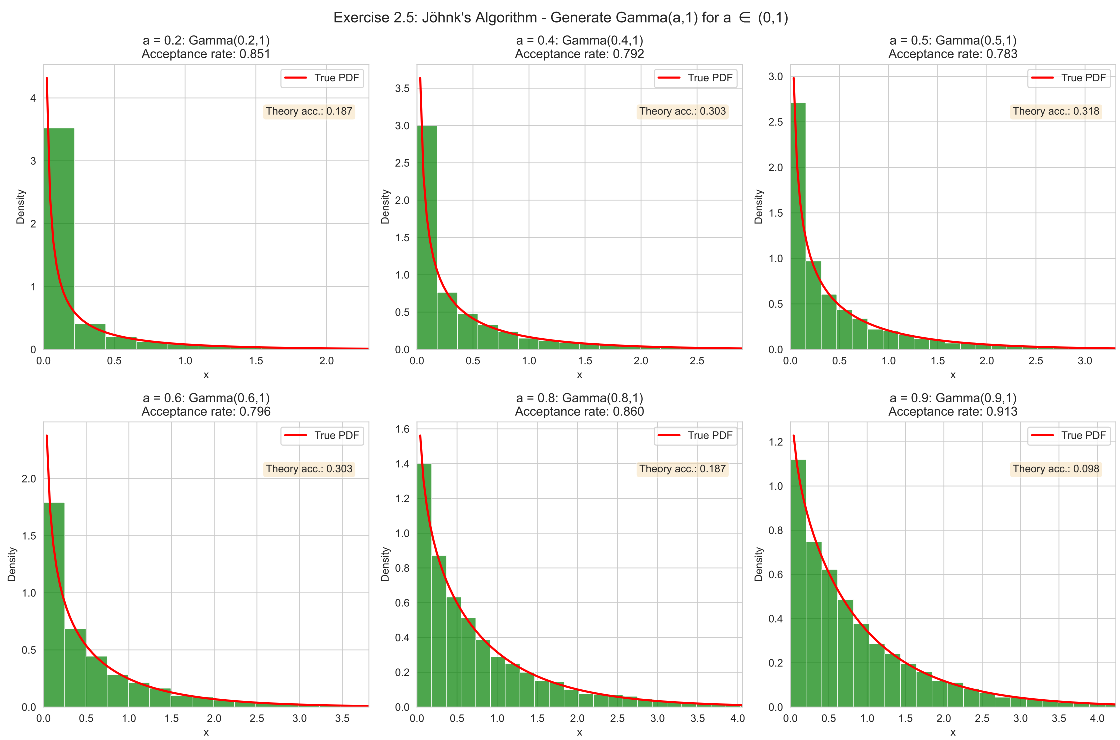
$$X_1 = \underbrace{-\sum_{i=1}^{n} \log(U_i)}_{\text{Gamma}(n,1)} + \underbrace{\text{Johnk}(\alpha)}_{\text{Gamma}(\alpha,1)}$$

This provides a complete method to generate Beta(a, b) random variables for any positive parameters using only uniform random variables.





Exercise 2.2: $U \sim U[0,1] \to X = U^{1/a} \sim Beta(a,1)$for 10000 samples
a = 0.5: X = U^(1/0.5) ~ Beta(0.5,1)
Mean: 0.334
(Theory: 0.333)
True PDF
a = 1: X = U^(1/1) ~ Beta(1,1)
Mean: 0.500
(Theory: 0.500)
a = 2: X = U^(1/2) ~ Beta(2,1)
Mean: 0.669
(Theory: 0.667)
a = 3: X = U^(1/3) ~ Beta(3,1)
Mean: 0.749
(Theory: 0.750)
a = 5: X = U^(1/5) ~ Beta(5,1)
Mean: 0.834
(Theory: 0.833)
a = 10: X = U^(1/10) ~ Beta(10,1)
Mean: 0.909
(Theory: 0.909)
Density
x



Exercise 2.5: Jöhnk's Algorithm - Generate Gamma(a,1) for a ∈ (0,1)
a = 0.2: Gamma(0.2,1)
Acceptance rate: 0.851
Theory acc.: 0.187
a = 0.4: Gamma(0.4,1)
Acceptance rate: 0.792
Theory acc.: 0.303
a = 0.5: Gamma(0.5,1)
Acceptance rate: 0.783
Theory acc.: 0.318
a = 0.6: Gamma(0.6,1)
Acceptance rate: 0.796
Theory acc.: 0.303
a = 0.8: Gamma(0.8,1)
Acceptance rate: 0.860
Theory acc.: 0.187
a = 0.9: Gamma(0.9,1)
Acceptance rate: 0.913
Theory acc.: 0.098
True PDF
Density
x

Exercise 3

Question 1

Show that the probability of accepting a proposed $X = x$ in either step (b) or (c) is $\frac{\tilde{\pi}(x)}{M\tilde{q}(x)}$.
For a proposed value $X = x$, we can accept it in two ways:

- Accept in step (b) with probability: $P(\text{accept in (b)}) = \frac{h(x)}{M\tilde{q}(x)}$
- Reject in step (b) and accept in step (c)

The probability of rejecting in step (b) is: $1 - \frac{h(x)}{M\tilde{q}(x)}$
Given rejection in step (b), the probability of accepting in step (c) is:

$$P\Big(\text{accept in (c)}\Big|\text{reject in (b)}\Big) = \frac{\tilde{\pi}(x) - h(x)}{M\tilde{q}(x) - h(x)}$$

Therefore, the total probability of acceptance is:

$$\begin{aligned}
P(\text{accept}) &= \frac{h(x)}{M\tilde{q}(x)} + \left(1 - \frac{h(x)}{M\tilde{q}(x)}\right) \cdot \frac{\tilde{\pi}(x) - h(x)}{M\tilde{q}(x) - h(x)} \\
&= \frac{h(x)}{M\tilde{q}(x)} + \frac{M\tilde{q}(x) - h(x)}{M\tilde{q}(x)} \cdot \frac{\tilde{\pi}(x) - h(x)}{M\tilde{q}(x) - h(x)} \\
&= \frac{h(x)}{M\tilde{q}(x)} + \frac{\tilde{\pi}(x) - h(x)}{M\tilde{q}(x)} \\
&= \frac{h(x) + \tilde{\pi}(x) - h(x)}{M\tilde{q}(x)} \\
&= \frac{\tilde{\pi}(x)}{M\tilde{q}(x)}
\end{aligned}$$

Question 2

Deduce from the previous question that the distribution of the samples accepted by the above algorithm is π.

Proof. From standard rejection sampling theory, if we propose samples from density $q(x)$ and accept with probability proportional to $\frac{\tilde{\pi}(x)}{M\tilde{q}(x)}$, the accepted samples follow the distribution:

$$p(\text{accepted } X = x) = \frac{q(x) \cdot \frac{\tilde{\pi}(x)}{M\tilde{q}(x)}}{\int_{\mathcal{X}} q(y) \cdot \frac{\tilde{\pi}(y)}{M\tilde{q}(y)} dy}$$

Since $q(x) = \frac{\tilde{q}(x)}{Z_q}$:

$$\begin{aligned}
p(\text{accepted } X = x) &= \frac{\frac{\tilde{q}(x)}{Z_q} \cdot \frac{\tilde{\pi}(x)}{M\tilde{q}(x)}}{\int_{\mathcal{X}} \frac{\tilde{q}(y)}{Z_q} \cdot \frac{\tilde{\pi}(y)}{M\tilde{q}(y)} dy} \\
&= \frac{\frac{\tilde{\pi}(x)}{MZ_q}}{\int_{\mathcal{X}} \frac{\tilde{\pi}(y)}{MZ_q} dy} = \frac{\tilde{\pi}(x)}{\int_{\mathcal{X}} \tilde{\pi}(y) dy} = \frac{\tilde{\pi}(x)}{Z_\pi} = \pi(x)
\end{aligned}$$

$\square$

Question 3

Show that the probability that step (c) has to be carried out is $1 - \frac{\int_{\mathcal{X}} h(x)dx}{MZ_q}$.

Proof. Step (c) is carried out when we reject in step (b). The probability of rejection in step (b) for a sample $X \sim q$ is:

$$
\begin{aligned}
P(\text{reject in (b)}) &= E_q\left[1 - \frac{h(X)}{M\tilde{q}(X)}\right] \\
&= 1 - E_q\left[\frac{h(X)}{M\tilde{q}(X)}\right] \\
&= 1 - \int_{\mathcal{X}} q(x) \cdot \frac{h(x)}{M\tilde{q}(x)} dx \\
&= 1 - \int_{\mathcal{X}} \frac{\tilde{q}(x)}{Z_q} \cdot \frac{h(x)}{M\tilde{q}(x)} dx \\
&= 1 - \frac{1}{MZ_q} \int_{\mathcal{X}} h(x)dx
\end{aligned}
$$

□

Question 4

Let $\tilde{\pi}(x) = \exp(-x^2/2)$ and $\tilde{q}(x) = \exp(-|x|)$. Using the fact that $\tilde{\pi}(x) \geq 1 - x^2/2$ for any $x \in \mathbb{R}$, how could you use the squeeze rejection sampling algorithm to sample from $\pi(x)$. What is the probability of not having to evaluate $\tilde{\pi}(x)$? Why could it be beneficial to use this algorithm instead of the standard rejection sampling procedure?

Solution:

We can set $h(x) = \max\{0, 1 - x^2/2\}$. Since h must be non-negative, we have:

$$
h(x) = \begin{cases} 1 - \frac{x^2}{2} & \text{if } |x| \leq \sqrt{2} \\ 0 & \text{if } |x| > \sqrt{2} \end{cases}
$$

To find M, we need $\tilde{\pi}(x) \leq M\tilde{q}(x)$ for all x:

$$
\exp(-x^2/2) \leq M \exp(-|x|)
$$

$$
M \geq \exp(|x| - x^2/2)
$$

The maximum of $|x| - x^2/2$ occurs at $|x| = 1$, giving $M = \exp(1/2) = \sqrt{e}$.
The probability of not having to evaluate $\tilde{\pi}(x)$ (accepting in step (b)) is:

$$
P(\text{accept in (b)}) = \frac{\int_{\mathcal{X}} h(x)dx}{MZ_q}
$$

where $Z_q = \int_{-\infty}^{\infty} \exp(-|x|)dx = 2$.

$$
\begin{aligned}
\int_{\mathcal{X}} h(x)dx &= \int_{-\sqrt{2}}^{\sqrt{2}} \left(1 - \frac{x^2}{2}\right) dx \\
&= \left[x - \frac{x^3}{6}\right]_{-\sqrt{2}}^{\sqrt{2}} \\
&= 2\sqrt{2} - \frac{2(\sqrt{2})^3}{6} \\
&= 2\sqrt{2} - \frac{2\sqrt{2}}{3} = \frac{4\sqrt{2}}{3}
\end{aligned}
$$

Therefore:

$$P(\text{not evaluating } \tilde{\pi}) = \frac{4\sqrt{2}/3}{\sqrt{e} \cdot 2} = \frac{2\sqrt{2}}{3\sqrt{e}} \approx 0.57$$

Benefits:

- The squeeze algorithm avoids computing the expensive $\exp(-x^2/2)$ function with probability $\frac{2\sqrt{2}}{3\sqrt{e}} \approx 0.57$
- This is particularly beneficial when $\tilde{\pi}(x)$ is computationally expensive to evaluate
- The algorithm maintains the same acceptance rate as standard rejection sampling while reducing computational cost

Exercise 4

Consider Marsaglia's polar method:

- Step a: Generate independent U_1, U_2 according to $U[-1,1]$ until $Y = U_1^2 + U_2^2 \leq 1$.
- Step b: Define $Z = \sqrt{-2\log(Y)}$ and return $X_1 = Z\frac{U_1}{\sqrt{Y}}, X_2 = Z\frac{U_2}{\sqrt{Y}}$.

Question 1: Joint Distribution of Y and ϑ

The joint distribution of $Y = U_1^2 + U_2^2$ and $\vartheta = \arctan 2(U_1, U_2)$ has density:

$$f_{Y,\vartheta}(y,\theta) = \mathbf{1}_{[0,1]}(y)\frac{\mathbf{1}_{[0,2\pi]}(\theta)}{2\pi}$$

Proof:
Since U_1, U_2 are independent and uniformly distributed on $[-1,1]$, their joint density is:

$$f_{U_1,U_2}(u_1,u_2) = \frac{1}{2}\mathbf{1}_{[-1,1]}(u_1)\frac{1}{2}\mathbf{1}_{[-1,1]}(u_2) = \frac{1}{4}\mathbf{1}_{[-1,1]}(u_1)\mathbf{1}_{[-1,1]}(u_2)$$

We condition on the event $\{U_1^2 + U_2^2 \leq 1\}$. The probability of this event is:

$$P(U_1^2 + U_2^2 \leq 1) = \frac{\text{Area of unit disk}}{\text{Area of square}} = \frac{\pi}{4}$$

The conditional density is:

$$f_{U_1,U_2|Y\leq 1}(u_1,u_2) = \frac{f_{U_1,U_2}(u_1,u_2)}{P(U_1^2+U_2^2\leq 1)}\mathbf{1}_{\{u_1^2+u_2^2\leq 1\}} = \frac{1}{\pi}\mathbf{1}_{\{u_1^2+u_2^2\leq 1\}}$$

Now we transform to polar coordinates:

- $Y = U_1^2 + U_2^2 = R^2$ where $R = \sqrt{U_1^2 + U_2^2}$
- $\vartheta = \arctan 2(U_1, U_2)$
- $U_1 = R\cos(\vartheta) = \sqrt{Y}\cos(\vartheta)$
- $U_2 = R\sin(\vartheta) = \sqrt{Y}\sin(\vartheta)$

The Jacobian of the transformation $(U_1, U_2) \to (Y, \vartheta)$ is:

$$J = \begin{vmatrix} \frac{\partial u_1}{\partial y} & \frac{\partial u_1}{\partial \vartheta} \\ \frac{\partial u_2}{\partial y} & \frac{\partial u_2}{\partial \vartheta} \end{vmatrix} = \begin{vmatrix} \frac{\cos(\vartheta)}{2\sqrt{y}} & -\sqrt{y}\sin(\vartheta) \\ \frac{\sin(\vartheta)}{2\sqrt{y}} & \sqrt{y}\cos(\vartheta) \end{vmatrix} = \frac{1}{2}$$

Therefore:

$$f_{Y,\vartheta}(y,\theta) = f_{U_1,U_2|Y\leq 1}(\sqrt{y}\cos(\theta), \sqrt{y}\sin(\theta))|J| = \frac{1}{\pi}\cdot\frac{1}{2} = \frac{1}{2\pi}$$

for $(y,\theta) \in [0,1]\times[0,2\pi]$, which gives us:

$$f_{Y,\vartheta}(y,\theta) = \mathbf{1}_{[0,1]}(y)\frac{\mathbf{1}_{[0,2\pi]}(\theta)}{2\pi}$$

This shows that $Y \sim \text{Uniform}[0,1]$ and $\vartheta \sim \text{Uniform}[0,2\pi]$ are independent.

Question 2: Independence and Normality of X_1 and X_2

To prove: X_1 and X_2 are independent standard normal random variables.
From Question 1, we know that $Y \sim \text{Uniform}[0,1]$ and $\vartheta \sim \text{Uniform}[0,2\pi]$ are independent.
Define:

- $R = \sqrt{Y}$, so $R^2 = Y \sim \text{Uniform}[0,1]$
- $Z = \sqrt{-2\log(Y)}$
- $X_1 = Z\cos(\vartheta) = \sqrt{-2\log(Y)}\cos(\vartheta)$
- $X_2 = Z\sin(\vartheta) = \sqrt{-2\log(Y)}\sin(\vartheta)$

Since $Y \sim \text{Uniform}[0,1]$, we have $-\log(Y) \sim \text{Exponential}(1)$.
Therefore, $-2\log(Y) \sim \text{Exponential}(1/2) = \text{Gamma}(1, 1/2)$.
The transformation $(Y, \vartheta) \to (X_1, X_2)$ is:

- $X_1 = \sqrt{-2\log(Y)}\cos(\vartheta)$
- $X_2 = \sqrt{-2\log(Y)}\sin(\vartheta)$

This is exactly the Box-Muller transformation in disguise!
To find the joint density of (X_1, X_2), we use the inverse transformation:

- $Y = e^{-(x_1^2+x_2^2)/2}$
- $\vartheta = \arctan 2(x_1, x_2)$

The Jacobian is:

$$J = \begin{vmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} \\ \frac{\partial \vartheta}{\partial x_1} & \frac{\partial \vartheta}{\partial x_2} \end{vmatrix} = \begin{vmatrix} -x_1 e^{-(x_1^2+x_2^2)/2} & -x_2 e^{-(x_1^2+x_2^2)/2} \\ \frac{-x_2}{x_1^2+x_2^2} & \frac{x_1}{x_1^2+x_2^2} \end{vmatrix} = \frac{e^{-(x_1^2+x_2^2)/2}}{2\pi}$$

The joint density of (X_1, X_2) is:

$$f_{X_1,X_2}(x_1,x_2) = f_{Y,\vartheta}(e^{-(x_1^2+x_2^2)/2}, \arctan 2(x_1,x_2))|J|^{-1}$$

$$= \frac{1}{2\pi} \cdot \frac{2\pi}{e^{-(x_1^2+x_2^2)/2}} = e^{-(x_1^2+x_2^2)/2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-x_2^2/2}$$

This is the product of two independent standard normal densities, proving that $X_1, X_2 \sim \mathcal{N}(0,1)$ independently.

Question 3: Benefits over Box-Muller Algorithm

Potential benefits of Marsaglia's polar method over Box-Muller:

- No trigonometric functions: Marsaglia's method avoids computing cos and sin functions, which can be computationally expensive.
- Better numerical stability: The method avoids potential numerical issues that can arise with trigonometric function evaluations, especially for arguments close to multiples of π.
- Simpler implementation: The algorithm only requires:
  - Uniform random number generation

  - Basic arithmetic operations (multiplication, addition, square root, logarithm)
  - Simple rejection sampling
- More efficient on average: While the rejection step means some samples are discarded, the acceptance probability is $\pi/4 \approx 0.785$, so on average only about 27

The main trade-off is that Marsaglia's method uses a variable number of uniform random numbers (due to rejection), while Box-Muller always uses exactly two uniform random numbers to produce two normal random numbers.

Exercise 5 Solution

Part 1: Expression of $q^*(x)$

The accepted samples follow a probability density $q^*(x)$. Using the rejection sampling framework, when we accept with probability $\min\{1, w(x)/c\}$, the density of accepted samples is:

$$q^*(x) = \frac{q(x) \cdot \min\{1, w(x)/c\}}{Z_c} \tag{1}$$

where the normalization constant is:

$$Z_c = \int_{\mathcal{X}} \min\{1, w(x)/c\}\, q(x)\, dx \tag{2}$$

This can be rewritten as:

$$q^*(x) = \frac{q(x) \cdot \min\{1, \pi(x)/(cq(x))\}}{Z_c} = \frac{\min\{q(x), \pi(x)/c\}}{Z_c} \tag{3}$$

Part 2: Prove $\mathbb{E}_{q^*}[(w^*(X))^2] = Z_c \mathbb{E}_q(\max\{w(X), c\} w(X))$

First, let's find $w^*(x) = \pi(x)/q^*(x)$:

$$\begin{align} w^*(x) &= \frac{\pi(x)}{q^*(x)} = \frac{\pi(x) \cdot Z_c}{q(x) \cdot \min\{1, w(x)/c\}} \tag{4} \\ &= \frac{w(x) \cdot Z_c}{\min\{1, w(x)/c\}} \tag{5} \end{align}$$

Since $\min\{1, w(x)/c\} = \min\{c, w(x)\}/c$, we have:

$$w^*(x) = \frac{w(x) \cdot c \cdot Z_c}{\min\{c, w(x)\}} = \frac{c \cdot Z_c \cdot w(x)}{\min\{c, w(x)\}} \tag{6}$$

Now, computing $\mathbb{E}_{q^*}[(w^*(X))^2]$:

$$\begin{align} \mathbb{E}_{q^*}[(w^*(X))^2] &= \int_{\mathcal{X}} (w^*(x))^2 q^*(x)\, dx \tag{7} \\ &= \int_{\mathcal{X}} \left(\frac{c \cdot Z_c \cdot w(x)}{\min\{c, w(x)\}} \right)^2 \cdot \frac{q(x) \cdot \min\{1, w(x)/c\}}{Z_c}\, dx \tag{8} \\ &= \frac{c^2 \cdot Z_c^2}{Z_c} \int_{\mathcal{X}} \frac{w(x)^2}{\min\{c, w(x)\}^2} \cdot q(x) \cdot \min\{1, w(x)/c\}\, dx \tag{9} \\ &= c \cdot Z_c \int_{\mathcal{X}} \frac{w(x)^2 \cdot q(x)}{\min\{c, w(x)\}}\, dx \tag{10} \\ &= Z_c \int_{\mathcal{X}} \frac{c \cdot w(x)^2 \cdot q(x)}{\min\{c, w(x)\}}\, dx \tag{11} \end{align}$$

Since $\frac{c \cdot w(x)}{\min\{c, w(x)\}} = \max\{c, w(x)\}$, we get:

$$\mathbb{E}_{q^*}[(w^*(X))^2] = Z_c \cdot \mathbb{E}_q[\max\{w(X), c\} \cdot w(X)] \tag{12}$$

Part 3: Establish the inequality

We need to show:

$$\mathbb{E}_q[\min\{w(X), c\}] \cdot \mathbb{E}_q[\max\{w(X), c\} \cdot w(X)] \leq \mathbb{E}_q[\min\{w(X), c\} \cdot \max\{w(X), c\} \cdot w(X)] \tag{13}$$

Following the hint, define:

$$h(w_1, w_2) = [\min\{w_1, c\} - \min\{w_2, c\}][\max\{w_1, c\} \cdot w_1 - \max\{w_2, c\} \cdot w_2] \tag{14}$$

We need to show $h(w_1, w_2) \geq 0$ for all $w_1, w_2 > 0$.
Consider four cases:

- If $w_1, w_2 \leq c$:
$$h(w_1, w_2) = (w_1 - w_2)(c \cdot w_1 - c \cdot w_2) = c(w_1 - w_2)^2 \geq 0 \tag{15}$$
- If $w_1, w_2 \geq c$:
$$h(w_1, w_2) = (c - c)(w_1^2 - w_2^2) = 0 \tag{16}$$
- If $w_1 \leq c < w_2$:
$$h(w_1, w_2) = (w_1 - c)(c \cdot w_1 - w_2^2) \tag{17}$$
Since $w_1 \leq c$ and $w_2 > c$, we have $w_1 - c \leq 0$ and $c \cdot w_1 \leq c^2 < w_2^2$, so $h \geq 0$
- If $w_2 \leq c < w_1$: By symmetry with the previous case, $h \geq 0$

Since $h(w_1, w_2) \geq 0$, expanding and taking expectations:

$$\mathbb{E}_q[h(w(X), w(Y))] \geq 0 \tag{18}$$

where X and Y are independent with the same distribution. Expanding:

$$\begin{aligned}
0 &\leq \mathbb{E}_q[h(w(X), w(Y))] && (19)\\
&= \mathbb{E}_q[\min\{w(X), c\}] \cdot \mathbb{E}_q[\max\{w(Y), c\} \cdot w(Y)] && (20)\\
&\quad - \mathbb{E}_q[\min\{w(Y), c\}] \cdot \mathbb{E}_q[\max\{w(X), c\} \cdot w(X)] && (21)\\
&\quad + \mathbb{E}_q[\min\{w(Y), c\} \cdot \max\{w(X), c\} \cdot w(X)] && (22)\\
&\quad - \mathbb{E}_q[\min\{w(X), c\} \cdot \max\{w(Y), c\} \cdot w(Y)] && (23)
\end{aligned}$$

Since X and Y have the same distribution, the first two terms cancel, and the last two terms combine to give the desired inequality.

Part 4: Deduce $\mathbb{V}_{q^*}(w^*(X)) \leq \mathbb{V}_q(w(X))$

From part 2:

$$\mathbb{E}_{q^*}[(w^*(X))^2] = Z_c \cdot \mathbb{E}_q[\max\{w(X), c\} \cdot w(X)] \tag{24}$$

Since $\mathbb{E}_{q^*}[w^*(X)] = 1$ (as w^* are importance weights for π with respect to q^*), we have:

$$\mathbb{V}_{q^*}(w^*(X)) = \mathbb{E}_{q^*}[(w^*(X))^2] - 1 = Z_c \cdot \mathbb{E}_q[\max\{w(X), c\} \cdot w(X)] - 1 \tag{25}$$

From part 3's inequality:

$$\begin{aligned}
\mathbb{E}_q[\min\{w(X), c\}] \cdot \mathbb{E}_q[\max\{w(X), c\} \cdot w(X)] &\leq \mathbb{E}_q[\min\{w(X), c\} \cdot \max\{w(X), c\} \cdot w(X)] && (26)\\
&= \mathbb{E}_q[c \cdot w(X)^2] && (27)
\end{aligned}$$

where the last equality uses the fact that $\min\{w(X), c\} \cdot \max\{w(X), c\} = c \cdot w(X)$ for all $w(X) > 0$.
Since $Z_c = \mathbb{E}_q[\min\{w(X), c\}/c] = \mathbb{E}_q[\min\{w(X), c\}]/c$, we have:

$$Z_c \cdot \mathbb{E}_q[\max\{w(X), c\} \cdot w(X)] \leq \mathbb{E}_q[w(X)^2] \tag{28}$$

Therefore:

$$\mathbb{V}_{q^*}(w^*(X)) \leq \mathbb{E}_q[w(X)^2] - 1 = \mathbb{V}_q(w(X)) \tag{29}$$

This proves that rejection control reduces the variance of importance weights, making it a useful technique when standard importance sampling has high variance.

Exercise 6

Question 1

Show that $\pi_X(x) = \pi(x)$
The extended probability density is defined as:

$$\pi_{X,U}(x,u) = \begin{cases} Mq(x) & \text{for } x \in \mathcal{X}, u \in \left[0, \frac{w(x)}{M}\right] \\ 0 & \text{otherwise} \end{cases}$$

To verify that the marginal density $\pi_X(x) = \pi(x)$, we integrate over u:

$$\pi_X(x) = \int \pi_{X,U}(x,u)\, du$$

Since $\pi_{X,U}(x,u) = Mq(x)$ only when $u \in [0, \dfrac{w(x)}{M}]$, we have:

$$\pi_X(x) = \int_0^{w(x)/M} Mq(x)\, du = Mq(x) \cdot \frac{w(x)}{M} = q(x) \cdot w(x) = q(x) \cdot \frac{\pi(x)}{q(x)} = \pi(x)$$

Therefore, $\pi_X(x) = \pi(x)$.

Part 2: Normalized Importance Sampling Estimate

Using the identity:

$$I = \int_0^1 \int_{\mathcal{X}} \phi(x)\pi_{X,U}(x,u)\, dx\, du$$

Under $q_{X,U}(x,u) = q(x) \times \mathbb{I}_{[0,1]}(u)$, we have $X \sim q$, $U \sim \mathcal{U}[0,1]$ independently.
The importance weight function is:

$$w(x,u) = \frac{\pi_{X,U}(x,u)}{q_{X,U}(x,u)}$$

For points where $\pi_{X,U}(x,u) > 0$ (i.e., when $u \in [0, w(x)/M]$):

$$w(x,u) = \frac{Mq(x)}{q(x) \cdot 1} = M$$

However, $w(x,u) = 0$ when $u > w(x)/M$.
More precisely:

$$w(x,u) = M \cdot \mathbb{I}[u \leq w(x)/M]$$

The normalized importance sampling estimate with n samples $(X_i, U_i) \sim q_{X,U}$ is:

$$\hat{I}_n = \frac{\sum_{i=1}^n \phi(X_i) w(X_i, U_i)}{\sum_{i=1}^n w(X_i, U_i)} = \frac{\sum_{i=1}^n \phi(X_i) \cdot M \cdot \mathbb{I}[U_i \leq w(X_i)/M]}{\sum_{i=1}^n M \cdot \mathbb{I}[U_i \leq w(X_i)/M]}$$

Simplifying (canceling M):

$$\hat{I}_n = \frac{\sum_{i: U_i \leq w(X_i)/M} \phi(X_i)}{\sum_{i: U_i \leq w(X_i)/M} 1}$$

This is exactly the rejection sampling estimate! We only include samples where $U_i \leq w(X_i)/M$, which is the acceptance condition in rejection sampling with proposal q and bound M.

Part 3: Show $\mathbb{V}_q(w(X)) \leq \mathbb{V}_{q_{X,U}}(w(X,U))$

First, let's compute the expectations and variances.
Under q:

- $\mathbb{E}_q[w(X)] = \int_{\mathcal{X}} w(x)q(x)\,dx = \int_{\mathcal{X}} \pi(x)\,dx = 1$
- $\mathbb{V}_q(w(X)) = \mathbb{E}_q[w(X)^2] - 1$

Under $q_{X,U}$:

- $w(X,U) = M$ when $U \leq w(X)/M$, and 0 otherwise
- $\mathbb{E}_{q_{X,U}}[w(X,U)] = \mathbb{E}_q[\mathbb{E}_U[w(X,U)|X]]$

Computing the conditional expectation:

$$\begin{aligned}
\mathbb{E}_U[w(X,U)|X] &= M \cdot \mathbb{P}(U \leq w(X)/M) + 0 \cdot \mathbb{P}(U > w(X)/M) \\
&= M \cdot \frac{w(X)}{M} \\
&= w(X)
\end{aligned}$$

Therefore:

$$\mathbb{E}_{q_{X,U}}[w(X,U)] = \mathbb{E}_q[w(X)] = 1$$

For the second moment:

$$\begin{aligned}
\mathbb{E}_{q_{X,U}}[w(X,U)^2] &= \mathbb{E}_q[\mathbb{E}_U[w(X,U)^2|X]] \\
&= \mathbb{E}_q[M^2 \cdot \mathbb{P}(U \leq w(X)/M)] \\
&= \mathbb{E}_q\left[M^2 \cdot \frac{w(X)}{M}\right] \\
&= M \cdot \mathbb{E}_q[w(X)] \\
&= M
\end{aligned}$$

Therefore:

$$\mathbb{V}_{q_{X,U}}(w(X,U)) = \mathbb{E}_{q_{X,U}}[w(X,U)^2] - 1 = M - 1$$

Since $w(x) = \pi(x)/q(x) \leq M$ for all x, we have $w(X)^2 \leq Mw(X)$ almost surely. Thus:

$$\mathbb{E}_q[w(X)^2] \leq M \cdot \mathbb{E}_q[w(X)] = M$$

Thus:

$$\mathbb{V}_q(w(X)) = \mathbb{E}_q[w(X)^2] - 1 \leq M - 1 = \mathbb{V}_{q_{X,U}}(w(X,U))$$

This completes the proof.

Interpretation: This result shows that importance sampling with the original distribution q has lower or equal variance compared to the extended space formulation, which is equivalent to rejection sampling. The key insight is that rejection sampling can be viewed as importance sampling in an extended space, but it generally has higher variance due to the binary accept/reject nature of the weights.

The equality holds only when $w(x) = M$ almost everywhere, which would mean π and q are proportional - a trivial case where rejection sampling would accept every sample.