Your Name: Po Yu Chen **Problem Set 2, Due: Mar. 18, 2025**

(1) Use the probability integral transformation method to simulate from the distribution

$$f(x) = \begin{cases} \frac{2}{a^2}x, & \text{if } 0 \le x \le a \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $a > 0$. Set a value for a, simulate various sample sizes, and compare results to the true distribution.

Solution:

1. Find the CDF $F_X(x)$

Let X be a random variable with the given PDF:

$$f(x) = \begin{cases} \frac{2}{a^2}\,x, & 0 \le x \le a, \\ 0, & \text{otherwise.} \end{cases}$$

To apply the probability integral transform (PIT), we first compute the cumulative distribution function (CDF). For $0 \le x \le a$:

$$F_X(x) \;=\; \int_0^x \frac{2}{a^2}\,t\,dt \;=\; \frac{t^2}{a^2}\bigg|_{t=0}^{t=x} \;=\; \frac{x^2}{a^2}.$$

Hence,

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x^2}{a^2}, & 0 \le x \le a, \\ 1, & x \ge a. \end{cases}$$

2. Invert the CDF

On the interval $0 \le x \le a$, we set

$$U \;=\; F_X(X) \;=\; \frac{X^2}{a^2}.$$

Solving for X in terms of U gives

$$X = F_X^{-1}(U) = a\,\sqrt{U} \quad \text{for } U \in [0,1].$$

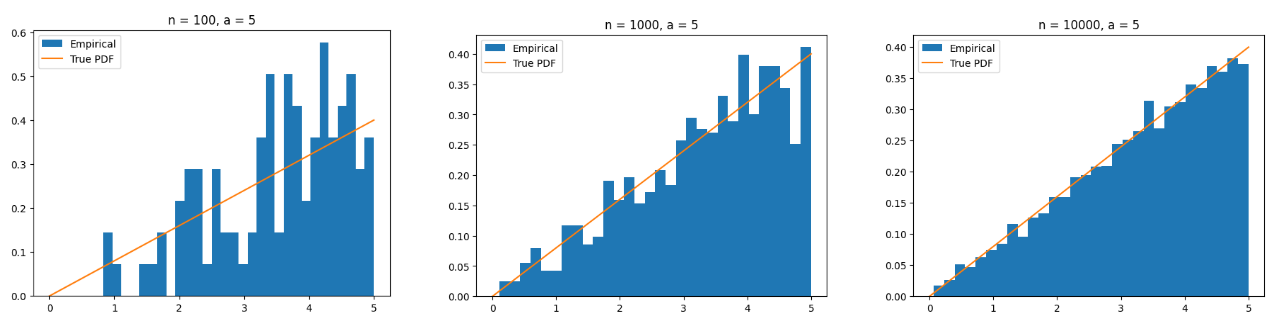
3. Simulate Random Samples

To generate n samples from this distribution via the probability integral transform:

1. Generate n i.i.d. uniform random variables $\{U_1, U_2, \ldots, U_n\} \sim \text{Uniform}(0, 1)$.
2. For each U_i, set
$$X_i \;=\; a\,\sqrt{U_i}.$$
3. The collection $\{X_i\}_{i=1}^n$ are then samples from the desired distribution.

4. Discussion of Results

In short, as n increases, the empirical distribution converges to the true PDF—a direct illustration of the Law of Large Numbers (and related results). The large-n histogram's slope and overall shape match the linear theoretical PDF quite well, confirming that the probability integral transform method is accurately simulating from the intended distribution.



(a) Histogram for $n = 100$. (b) Histogram for $n = 1000$. (c) Histogram for $n = 10,000$.

Figure 1: Comparison of empirical histograms (blue bars) vs. the true PDF (orange line).

(2) Generate samples from the distribution

$$f(x) = \frac{2}{3}e^{-2x} + 2e^{-3x} \tag{2}$$

using the finite mixture approach.
Solution:

1. Show that $f(x)$ is a Finite Mixture

Consider two exponential PDFs:

$$\text{Exp}(2): \quad g_1(x) = 2\,e^{-2x}, \quad x \geq 0,$$

$$\text{Exp}(3): \quad g_2(x) = 3\,e^{-3x}, \quad x \geq 0.$$

We want to see if

$$f(x) \;=\; \frac{2}{3}\,e^{-2x} + 2\,e^{-3x}$$

can be expressed as

$$p \cdot g_1(x) \;+\; (1-p) \cdot g_2(x)$$

for some $p \in (0,1)$. Matching coefficients:

$$p \cdot 2 = \frac{2}{3} \quad \Longrightarrow \quad p = \frac{1}{3},$$

$$(1-p) \cdot 3 = 2 \quad \Longrightarrow \quad 1-p = \frac{2}{3}.$$

Hence

$$f(x) = \underbrace{\tfrac{1}{3}}_{p}(2\,e^{-2x}) + \underbrace{\tfrac{2}{3}}_{1-p}(3\,e^{-3x}).$$

This establishes that

$$f(x) = \frac{1}{3}\,\text{Exp}(2) + \frac{2}{3}\,\text{Exp}(3).$$

Thus, $f(x)$ is a two-component finite mixture of exponentials.

2. Mixture Sampling Procedure

To generate a sample X from $f(x)$:

1. Generate a uniform random variable $V \sim \text{Uniform}(0,1)$.

2. If $V \leq \frac{1}{3}$, then set X to be a draw from $\text{Exp}(2)$, i.e.

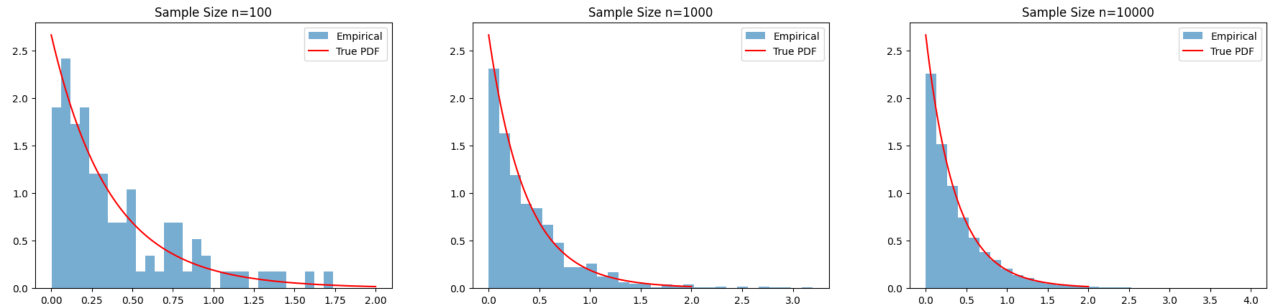
$$X = -\frac{1}{2}\ln(U_1), \quad U_1 \sim \text{Uniform}(0,1).$$

3. Otherwise (if $V > \frac{1}{3}$), set X to be a draw from $\text{Exp}(3)$, i.e.

$$X = -\frac{1}{3}\ln(U_2), \quad U_2 \sim \text{Uniform}(0,1).$$

Repeat these steps n times to obtain n i.i.d. samples $\{X_1, X_2, \ldots, X_n\}$ from $f(x)$.

3. Discussion



(a) Histogram for $n = 100$. (b) Histogram for $n = 1000$. (c) Histogram for $n = 10,000$.

Figure 2: Comparison of empirical histograms (blue bars) vs. the true PDF (red line).

In summary, as the sample size increases from $n = 100$ to $n = 10{,}000$, the empirical distributions become smoother and align more closely with the theoretical PDF. This illustrates the Law of Large Numbers in action: with larger n, the histogram provides a better approximation of the true underlying distribution.

(3) Draw 500 observations from Beta$(3,3)$ using the accept-reject algorithm. Compute the mean and variance of the sample and compare them to the true values.
Solution:

1. Beta(3,3) Distribution

A Beta(α,β) random variable on $[0,1]$ has PDF

$$f(x) \;=\; \frac{x^{\alpha-1}\,(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}, \quad 0 \le x \le 1,$$

where $\mathrm{B}(\alpha,\beta)$ is the Beta function. For $\alpha = 3,\ \beta = 3$, we get

$$f(x) = \frac{x^2\,(1-x)^2}{\mathrm{B}(3,3)}, \quad 0 \le x \le 1.$$

Recall that

$$\mathrm{B}(3,3) = \frac{\Gamma(3)\,\Gamma(3)}{\Gamma(6)} = \frac{2!\,2!}{5!} = \frac{4}{120} = \frac{1}{30}.$$

Thus,

$$f(x) \;=\; 30\,x^2\,(1-x)^2, \quad 0 \le x \le 1.$$

2. Accept-Reject Algorithm

(a) Choose a proposal distribution $g(x)$**.** A common choice for Beta distributions on $[0,1]$ is $g(x) = 1$ for $0 \le x \le 1$, i.e. a Uniform$(0,1)$ proposal.

(b) Find M **such that** $f(x) \le M\,g(x)$ **for all** x**.** Since $g(x) = 1$, we need $M \ge \max_{0\le x\le 1} f(x)$. For $f(x) = 30\,x^2\,(1-x)^2$, the maximum occurs at $x = \frac{1}{2}$. Then

$$f(\tfrac{1}{2}) = 30\,\left(\tfrac{1}{2}\right)^2\left(\tfrac{1}{2}\right)^2 = 30\,\frac{1}{4}\,\frac{1}{4} = 30\,\frac{1}{16} = 1.875.$$

Hence $M = 1.875$ is sufficient (or a slightly larger value may be used).

(c) Sampling steps. Repeat until we have 500 accepted samples:

1. Generate $Y \sim$ Uniform$(0,1)$.
2. Generate $U \sim$ Uniform$(0,1)$.
3. Compute $\dfrac{f(Y)}{M\,g(Y)} = \dfrac{30\,Y^2(1-Y)^2}{1.875}$.

4. If $U \leq \frac{f(Y)}{M\, g(Y)}$, **accept** Y; otherwise **reject** Y and go back to step 1.

All accepted Y values constitute a sample from $\text{Beta}(3,3)$.

3. Theoretical Mean and Variance

For $\text{Beta}(\alpha, \beta)$:

$$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}, \quad \text{Var}(X) = \frac{\alpha\,\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}.$$

Hence, for $\text{Beta}(3,3)$:

$$\mathbb{E}[X] = \frac{3}{3+3} = 0.5, \quad \text{Var}(X) = \frac{3\times 3}{6^2\times 7} = \frac{9}{252} = \frac{1}{28} \approx 0.035714.$$

4. Comparing Empirical Results to the True Values

After accepting 500 samples $\{X_1, X_2, \ldots, X_{500}\}$:

- **Sample Mean:**

  $$\overline{X} = \frac{1}{500}\sum_{i=1}^{500} X_i.$$

  Compare $\overline{X}$ to the true mean 0.5.

- **Sample Variance:**

  $$s^2 = \frac{1}{499}\sum_{i=1}^{500}\left(X_i - \overline{X}\right)^2.$$

  Compare s^2 to the true variance $1/28 \approx 0.0357$.

With a sample size of 500, you should see that $\overline{X}$ is reasonably close to 0.5, and s^2 is near 0.0357, though some fluctuation is expected due to random variation.

5. Conclusion

The sample mean (approximately 0.49995) is extremely close to the theoretical mean of 0.5.Likewise, the sample variance ($\approx$ 0.03628) is near the theoretical value of $1/28 \approx$ 0.03571.Although there is a slight deviation, it lies within the range expected from random fluctuation with a finite sample size.Overall, these results confirm that the accept–reject algorithm is accurately simulating observations from the Beta(3,3) distribution.