The **Probability Integral Transformation (PIT)** is a fundamental concept in probability theory and simulation. It is based on the following principle:

> If X is a continuous random variable with cumulative distribution function (CDF) $F_X(x)$, then the transformed variable $U = F_X(X)$ follows a **uniform distribution** on the interval $[0, 1]$, that is, $U \sim \text{Uniform}(0, 1)$.

Conversely, the **inverse transform method** allows us to generate samples from any continuous distribution using uniform random numbers:

> If $U \sim \text{Uniform}(0, 1)$, and F^{-1} is the inverse of the CDF of a continuous distribution, then the transformed variable $X = F^{-1}(U)$ follows the original distribution with CDF F, i.e., $X \sim F$.

This property provides a general method for generating random numbers from any continuous distribution with a known inverse CDF.

Example: Exponential Distribution

Let X follow an exponential distribution with rate parameter $\lambda > 0$.

- The **CDF** of the exponential distribution is:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

- Solving for x in terms of u gives the **inverse CDF**:

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

- Since $U \sim \text{Uniform}(0, 1)$, and $1 - U \sim \text{Uniform}(0, 1)$, this simplifies to:

$$X = -\frac{1}{\lambda} \ln(U)$$

Using this transformation, we can generate samples from the exponential distribution by:

1. Generating uniform random numbers $U_1, U_2, \ldots, U_n \sim \text{Uniform}(0, 1)$
2. Computing $X_i = -\frac{1}{\lambda} \ln(U_i)$ for each i

This process can be applied to other continuous distributions, as long as the inverse CDF $F^{-1}(u)$ is known or can be approximated.

Theorem: Sampling Distribution of the Sample Mean

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n drawn from a population with mean μ and variance σ^2. Then the sampling distribution of the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ has:

- Mean: $\mathbb{E}[\bar{X}] = \mu$
- Variance: $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$

As n increases, the sampling distribution of $\bar{X}$ tends to a normal distribution due to the Central Limit Theorem (CLT), regardless of the original population distribution.

Special Case: Sampling from a Normal Distribution

If the population is normally distributed, i.e., $X_i \sim \mathcal{N}(\mu, \sigma^2)$, then for any sample size n, the sampling distribution of the sample mean $\bar{X}$ is also normally distributed:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Procedure: Simulating the Sampling Distribution of the Sample Mean

To empirically study the sampling distribution of the sample mean, follow the steps below:

1. **Choose a distribution:** Select a population distribution (e.g., Normal, Binomial, Poisson, Exponential, Gamma) with known parameters (mean μ and standard deviation σ).
2. **Fix parameters:**
   - Sample size: n
   - Number of repetitions (samples): k
3. **Repeat the following process k times:**
   (a) Draw a random sample of size n from the chosen distribution.
   (b) Compute the sample mean $\bar{X}_i$ for the i-th sample.
4. **Collect all sample means:** This results in a set of k sample means: $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k$.
5. **Compute the mean and variance of the sample means:**

$$\text{Empirical Mean of Sample Means} = \frac{1}{k}\sum_{i=1}^{k} \bar{X}_i$$

$$\text{Empirical Variance of Sample Means} = \frac{1}{k-1}\sum_{i=1}^{k} \left(\bar{X}_i - \overline{\overline{X}}\right)^2$$

6. **Compare the empirical values with theoretical expectations:**
   - Theoretical Mean: μ
   - Theoretical Variance: σ^2/n
7. **Visualize the results:** Plot a histogram or density plot of the sample means to observe the shape of the sampling distribution.

Illustrative Example (in R)

Consider a population with a normal distribution, $\mu = 100$, and $\sigma = 15$. We take repeated samples of size $n = 30$ and compute the sample means.

R Code

```
set.seed(123)
n <- 30
num_samples <- 1000
mu <- 100
sigma <- 15

sample_means <- replicate(num_samples, {
  sample <- rnorm(n, mean = mu, sd = sigma)
  mean(sample)
})

mean(sample_means)      # Empirical mean
var(sample_means)       # Empirical variance
```

Expected Results

- Theoretical mean of sampling distribution: $\mu = 100$
- Theoretical variance: $\sigma^2/n = 15^2/30 = 7.5$
- Empirical mean and variance from R simulation should approximate these values.

Practice Questions

Binomial Distribution

1. What is the expected value and variance of the sampling distribution of the mean for a Binomial$(n = 10, p = 0.4)$ population?
2. Simulate the sampling distribution of the mean for Binomial$(n = 10, p = 0.8)$ and observe how skewness changes with p.

Poisson Distribution

1. If $X \sim \text{Poisson}(\lambda = 4)$, simulate the sampling distribution of the mean for sample size $n = 50$. Compare theoretical and empirical mean and variance.

2. How does the distribution of sample means change as λ increases?

Exponential Distribution

1. For $X \sim \text{Exponential}(\lambda = 1.5)$, simulate sample means and check normality for $n = 10$ and $n = 50$.

2. Does the Central Limit Theorem apply here? How does sample size affect the result?

Normal Distribution

1. Use $X \sim \mathcal{N}(70, 10^2)$. Simulate the sampling distribution of the mean for different values of n and confirm that it remains normal.

Gamma Distribution

1. Take $X \sim \text{Gamma}(\text{shape} = 2, \text{rate} = 1)$. Simulate and compare the sampling distribution for $n = 10$ and $n = 100$.

2. How does the shape of the original distribution influence the sample mean distribution?