भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Programming for AI

Sampling Techniques, Inverse Transform Technique, Rejection Sampling (Accept-Reject) Technique

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

07 April 2025

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Contents

- Sampling from a Given Distribution
  - Inverse Transform Technique (ITT)
  - Rejection Sampling (Accept-Reject) Technique

Sampling from a Given Distribution

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

$$F(x) = P(X \leq x)$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Sampling from a Given Distribution

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

Remarks:

- The CDF F may or may not be continuous

The basic building block: Uniform $(0, 1)$ sample(s)

Sampling from a Given Distribution

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

Remarks:

- The CDF F may or may not be continuous
- If F is continuous and differentiable, it admits a probability density function f

The basic building block: Uniform $(0, 1)$ sample(s)

Sampling from a Given Distribution

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

Remarks:

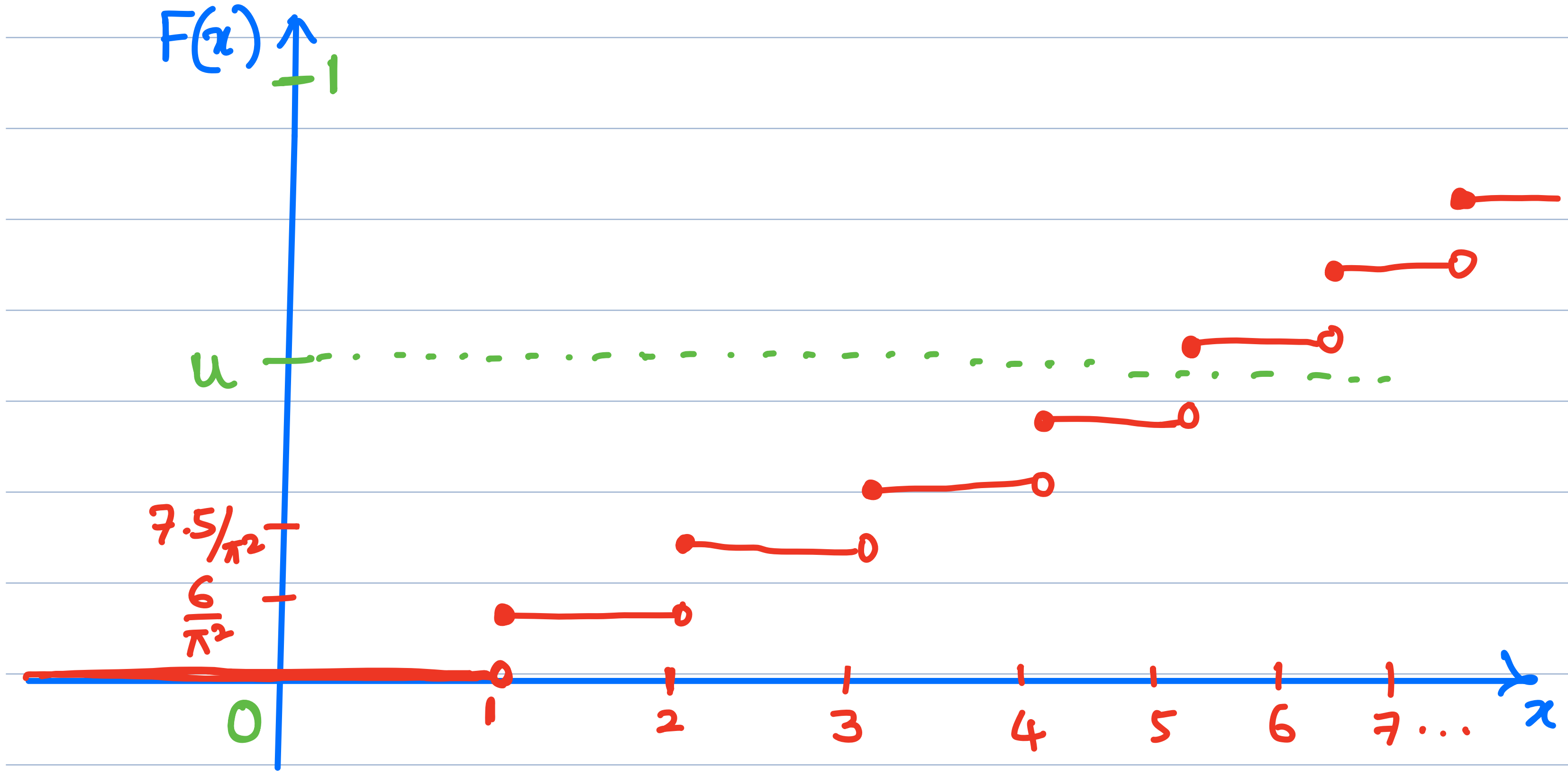
- The CDF F may or may not be continuous
- If F is continuous and differentiable, it admits a probability density function f
- Instead of the CDF F, we may be given a target PMF or PDF from which to sample

The basic building block: Uniform $(0, 1)$ sample(s)

$$\mathbb{P}(X=k) = \frac{6}{\pi^2}\cdot\frac{1}{k^2}, \qquad k\in\mathbb{N}.$$

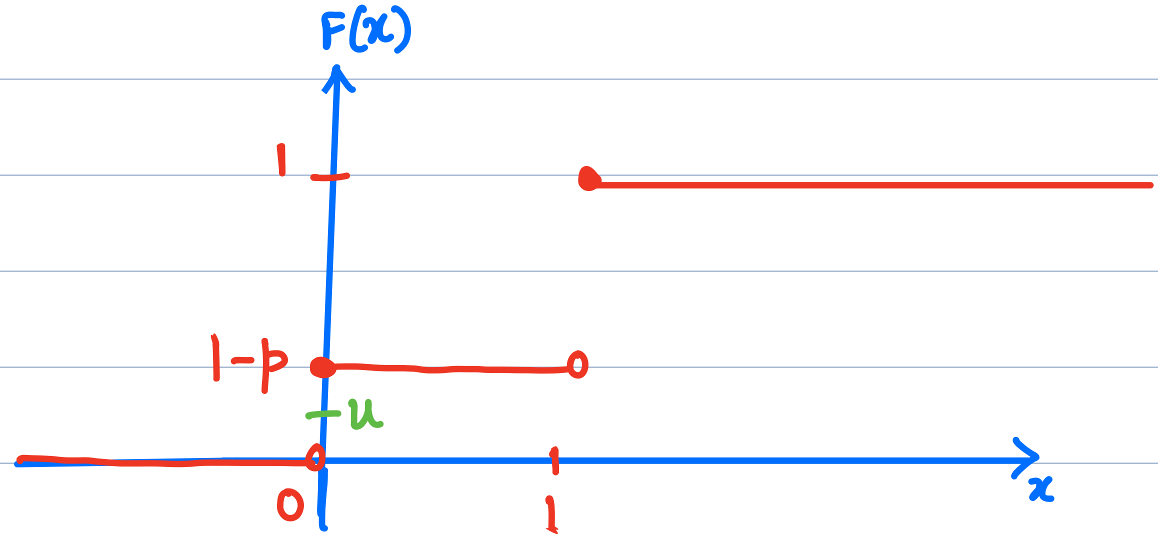
$$\mathbb{P}(X\le x)$$

$$\sum_{k\in\mathbb{N}} \mathbb{P}(X=k) = \frac{6}{\pi^2}\cdot\sum_{k\in\mathbb{N}}\frac{1}{k^2} = 1.$$



$X \sim Ber(p)$

$P(X=1) = p = 1 - \mathbb{P}(X=0)$



$$F^{-1}(u) = \min\{x\in\mathbb{R} : F(x)\ge u\}$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Inverse Transform Technique (ITT)

Inverse Transform Technique

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

Inverse Transform Technique

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

- Sample $U \sim \mathrm{Unif}(0, 1)$

Inverse Transform Technique

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

- Sample $U \sim \text{Unif}(0, 1)$
- Define the function F^{-1} as

$$F^{-1}(u) := \min\{x \in \mathbb{R} : F(x) \geq u\}$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Inverse Transform Technique

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

- Sample $U \sim \mathrm{Unif}(0, 1)$
- Define the function F^{-1} as

$$F^{-1}(u) := \min\{x \in \mathbb{R} : F(x) \geq u\}$$

- F^{-1} is called the quantile function

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Inverse Transform Technique

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

- Sample $U \sim \mathrm{Unif}(0, 1)$
- Define the function F^{-1} as

$$F^{-1}(u) := \min\{x \in \mathbb{R} : F(x) \geq u\}$$

- F^{-1} is called the quantile function
- Set $X = F^{-1}(U)$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Inverse Transform Technique

Objective

Given a cumulative distribution function (CDF) $F : \mathbb{R} \to [0, 1]$, generate a sample $X \sim F$.

- Sample $U \sim \text{Unif}(0, 1)$
- Define the function F^{-1} as

$$F^{-1}(u) := \min\{x \in \mathbb{R} : F(x) \geq u\}$$

- F^{-1} is called the quantile function
- Set $X = F^{-1}(U)$
- **Claim:** The CDF of X is exactly equal to F, i.e., $F_X = F$

Example

- Let X be a discrete random variable with the following PMF:

$\mathbb{P}(X=x)$

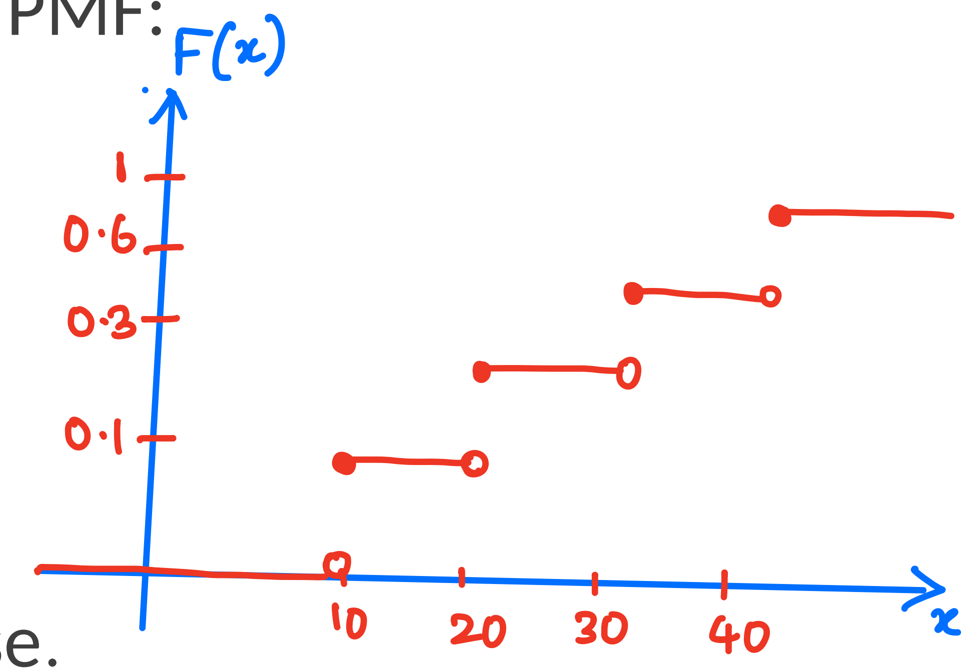
$$p_X(x) = \begin{cases} 0.1, & x = 10, \\ 0.2, & x = 20, \\ 0.3, & x = 30, \\ 0.4, & x = 40, \\ 0, & \text{otherwise.} \end{cases}$$

Use the inverse transform method to generate a sample from the above distribution.

Example

- Let X be a discrete random variable with the following PMF:

$$p_X(x) = \begin{cases} 0.1, & x = 10, \\ 0.2, & x = 20, \\ 0.3, & x = 30, \\ 0.4, & x = 40, \\ 0, & \text{otherwise.} \end{cases}$$



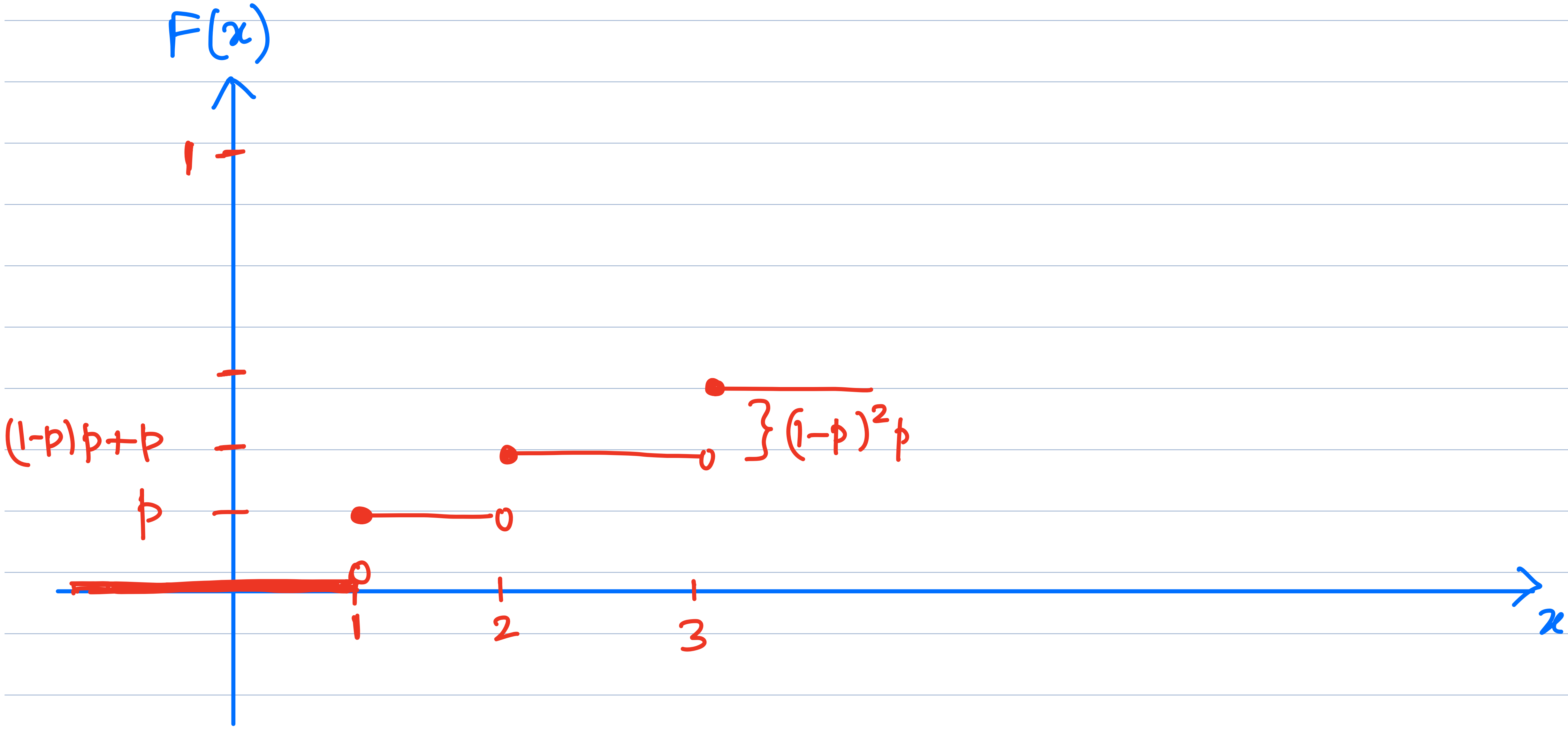
$$X = F^{-1}(U)$$

Use the inverse transform method to generate a sample from the above distribution.

$$F^{-1}(u) = \begin{cases} 10, & 0 < u \leq 0.1, \\ 20, & 0.1 < u \leq 0.3, \\ 30, & 0.3 < u \leq 0.6, \\ 40, & 0.6 < u \leq 1 \end{cases}$$

$X \sim \mathrm{Geo}(p)$

$$\mathbb{P}(X=k) = (1-p)^{k-1}p, \quad k \in \mathbb{N}.$$

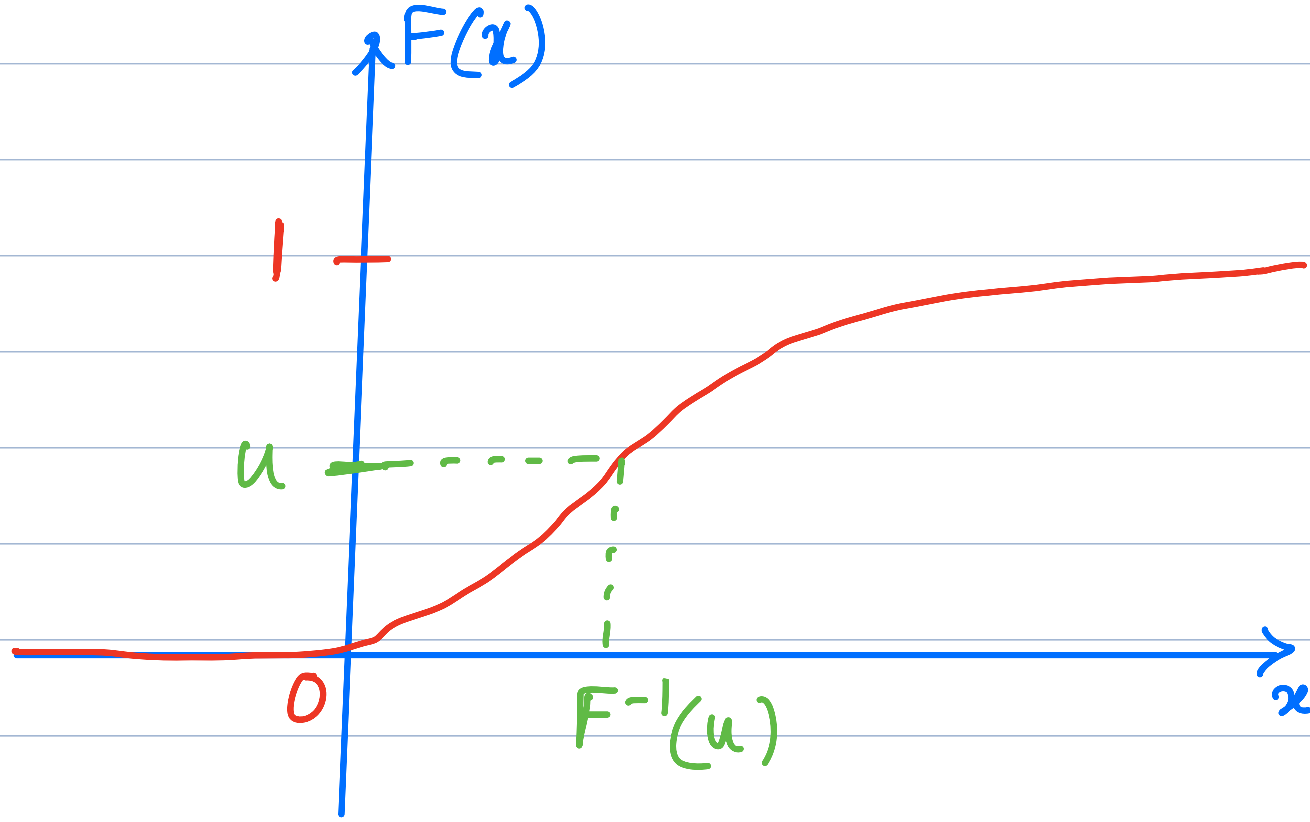


$$F^{-1}(u) = \begin{cases} 1, & 0 < u \le p, \\ 2, & p < u \le p + (1-p)p, \\ 3, & p + (1-p)p < u \le p + (1-p)p + (1-p)^2 p \\ \vdots & \end{cases}$$

$X \sim \text{Exp}(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \geq 0$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases}$$



$$\begin{aligned}
F^{-1}(u) &= \min\{x : F(x) \geq u\} \\
&= \min\{x \geq 0 : 1 - e^{-\lambda x} \geq u\} \\
&= \min\{x \geq 0 : e^{-\lambda x} \leq 1 - u\} \\
&= \min\{x \geq 0 : -\lambda x \leq \ln(1-u)\} \\
&= \min\left\{x \geq 0 : x \geq -\frac{\ln(1-u)}{\lambda}\right\} \\
&= -\frac{\ln(1-u)}{\lambda}
\end{aligned}$$

$$X \sim N(0,1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} , \qquad x \in \mathbb{R}$$

$$F(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Example

- **[Generating a Sample from Rayleigh Distribution]**

  The PDF of the Rayleigh distribution is given by

$$f(r) = r\, e^{-r^2/2}, \quad r > 0.$$

  Use the inverse transform method to generate a sample from the above distribution.

$$f(r) = r\, e^{-r^2/2}, \quad r > 0.$$

$$F(x) = \int_0^x f(t)\,dt = \begin{cases} 1 - e^{-x^2/2}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$

$$F^{-1}(u) = \min\{x \geq 0 : 1 - e^{-x^2/2} \geq u\}$$

$$= \min\{x \geq 0 : x^2 \geq -2\ln(1-u)\}$$

$$= \sqrt{-2\ln(1-u)}.$$

Gaussian Samples on Python via ITT

$X \sim N(\mu, \sigma^2)$

- Python's built-in module

$$\texttt{numpy.random.normal(loc, scale, size)}$$

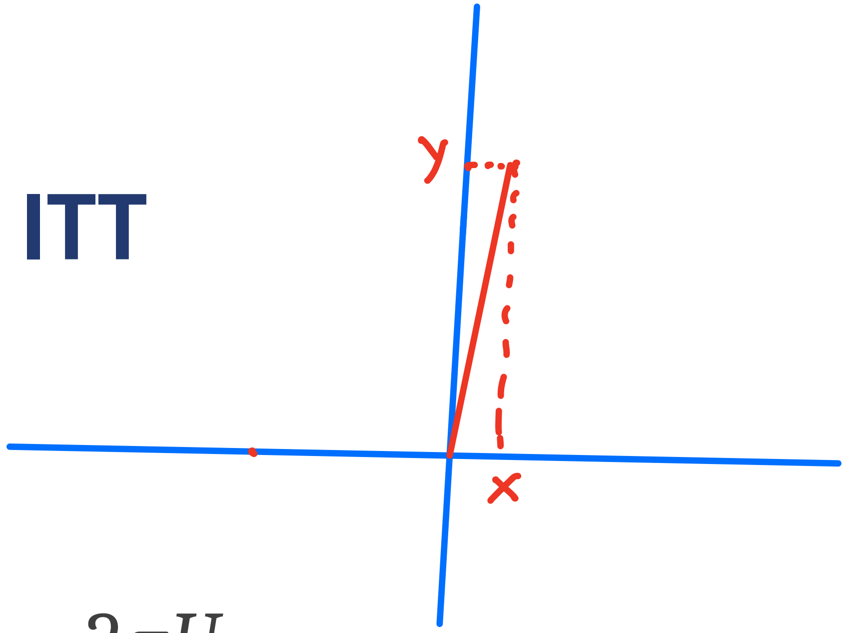
generates n independent samples from $\mathcal{N}(\mu, \sigma^2)$, where

$$n = \texttt{size}, \qquad \mu = \texttt{loc}, \qquad \sigma = \texttt{scale}.$$

- In principle, the above module uses the inverse transform technique

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Gaussian Samples on Python via ITT



1. Let $U_1, U_2 \overset{\text{i.i.d.}}{\sim} \text{Unif}(0, 1)$
2. Let R and Θ be two random variables defined via

$$R = F_1^{-1}(U_1), \qquad\qquad \Theta = 2\pi U_2,$$

   where F_1 is the CDF of the Rayleigh distribution
3. Let X_1 and X_2 be defined as

$$Y_1\,\cancel{X}_1 = R\cos(\Theta), \qquad\qquad Y\,\cancel{X}_2 = R\sin(\Theta).$$

4. Then, $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$
5. To get $X \sim \mathcal{N}(\mu, \sigma^2)$, simply discard Y_2, and

$$X = \sigma\, Y_1 + \mu.$$

6. Repeat steps 1-5 a total of n times to get $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$