COMP 480/580 — Probabilistic Algorithms and Data Structure Sep 25, 2025

Lecture 10

Lecturer: Anshumali Shrivastava *Scribe By:*

Nicky Hennigh (nh73, S01547063)
Leonard Richertz (lr106, S01547909)
Harrison Rey Ghozali (rg157, S01391896)

1 Introduction

In probabilistic algorithms, it is often necessary to generate samples from probability distributions. Two fundamental types of distributions are considered:

- **Discrete distributions**: random variables take values from a countable set.
- **Continuous distributions**: random variables take values from an interval of real numbers.

We denote a random variable by X, and its possible values by lowercase x. The probability that X takes value x is written as $P(X = x)$ in the discrete case, or via a probability density function (PDF) $f(x)$ in the continuous case.

2 Discrete vs Continuous Random Variables

2.1 Discrete Case

For a discrete random variable X, the probability mass function (PMF) $p(x)$ satisfies:

$$p(x) = P(X = x), \quad 0 \leq p(x) \leq 1,$$

and the total probability must sum to 1:

$$\sum_x p(x) = 1.$$

2.2 Continuous Case

For a continuous random variable X, the probability density function (PDF) $f(x)$ is such that:

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Here, $f(x)$ is not itself a probability, but rather a density. The probability of X lying in an interval $[a, b]$ is:

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx.$$

3 The Cumulative Distribution Function (CDF)

Definition 1 *The cumulative distribution function of a random variable X, $F_X(x)$, is the probability that X has a value less than or equal to x:*

$$F_X(x) = P(X \leq x).$$

The CDF must also satisfy the following properties:

- $0 \leq F_X(x) \leq 1$ for all x.
- F_X is non-decreasing: if $a \leq b$ then $F_X(a) \leq F_X(b)$.
- $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$.

3.1 Discrete Case

For a discrete random variable with PMF $p(x)$, $F_X(x) =$ the sum of all PMF results where X $\leq$ x:

$$F_X(x) = \sum_{x_i \leq x} p(x_i).$$

3.2 Continuous Case

For a continuous random variable with PDF $f(x)$, $F_X(x) =$ the integral from $-\infty$ to x of the PDF:

$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt.$$

4 Types of Distributions

When we classify random variables, we can differentiate not just between discrete or continuous variables, but also the complexity of their distribution form.

- Some distributions (e.g., Uniform, Exponential) have closed-form inverse CDFs and can be sampled using the inverse transform method.
- Many important distributions (e.g., Normal, Gamma, Student's t) do not have a simple inverse CDF, motivating the use of rejection sampling or more advanced methods.

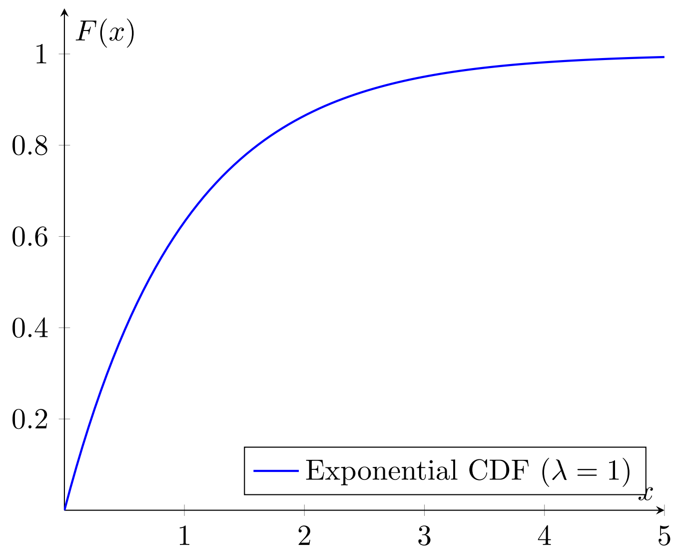
4.1 Exponential Distribution

Let $X \sim \text{Exp}(\lambda)$ with rate $\lambda > 0$.

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

The CDF is obtained by integration:

$$F(x) = \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x}.$$



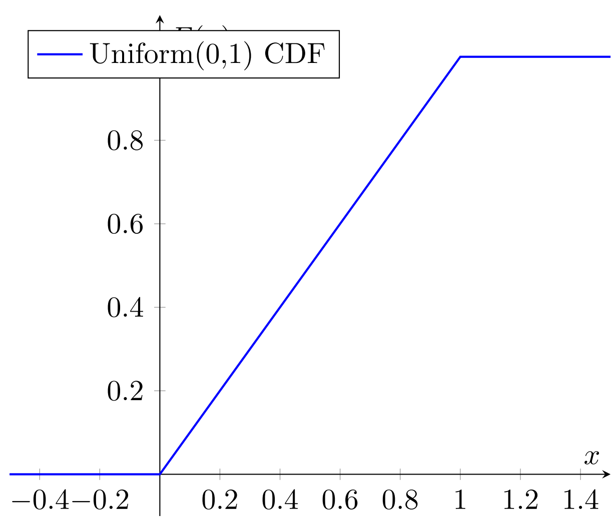
4.2 Uniform Distribution

Let $X \sim U(0, 1)$. Then:

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The CDF is:

$$F(x) = \begin{cases} 0 & x < 0, \\ x & 0 \le x \le 1, \\ 1 & x > 1. \end{cases}$$

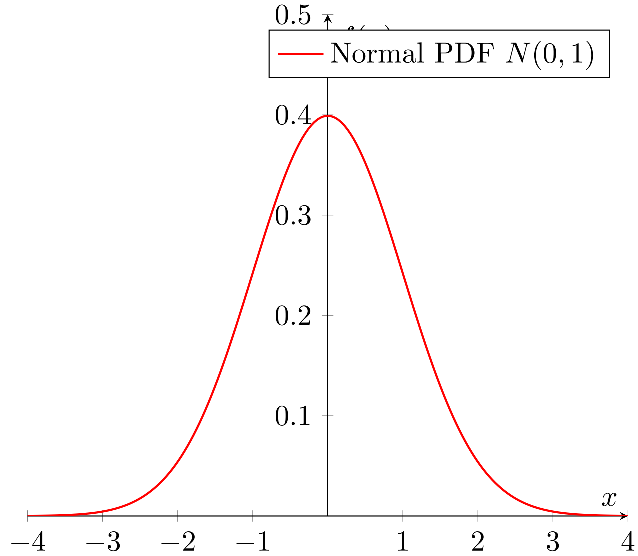


4.3 Normal Distribution

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

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

The CDF has no closed form, but the PDF is shown below:



5 Inverse Transform Sampling

The *inverse transform method* allows us to sample from arbitrary distributions using uniform random numbers. For this to work, we have to know the underlying CDF F(x).

5.1 Algorithm

Suppose F is the CDF of the target distribution.

1. Sample $U \sim U(0, 1)$.
2. Compute $X = F^{-1}(U)$.
3. Return X as a sample from the target distribution.

5.2 Proof of Correctness

To prove why the inverse transform sampling method works, we need to show that the random variable $X = F^{-1}(U)$, where $U \sim \text{Unif}(0, 1)$, indeed follows the distribution with cumulative distribution function (CDF) $F(x)$.

By definition, the CDF of X is

$$P(X \leq x).$$

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

$$P(X \leq x) = P(F^{-1}(U) \leq x).$$

Since F is a monotone increasing function (as all CDFs are), the inequality $F^{-1}(U) \leq x$ is equivalent to $U \leq F(x)$. Therefore,

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)).$$

Now, because $U \sim \text{Unif}(0,1)$, we know that

$$P(U \leq y) = y, \quad \text{for } 0 \leq y \leq 1.$$

Applying this property with $y = F(x)$, we obtain

$$P(U \leq F(x)) = F(x).$$

Thus,

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

which shows that X has distribution function F, i.e., $X \sim F$.

6 Rejection Sampling (aka Accept-Reject Algorithm)

When F^{-1} is not available in closed form, the rejection sampling method is an alternative option.

6.1 Setup

- Let $q(x)$ be the target density.
- Choose a proposal distribution $p(x)$ from which we can easily sample.
- Find a constant $M > 0$ such that:

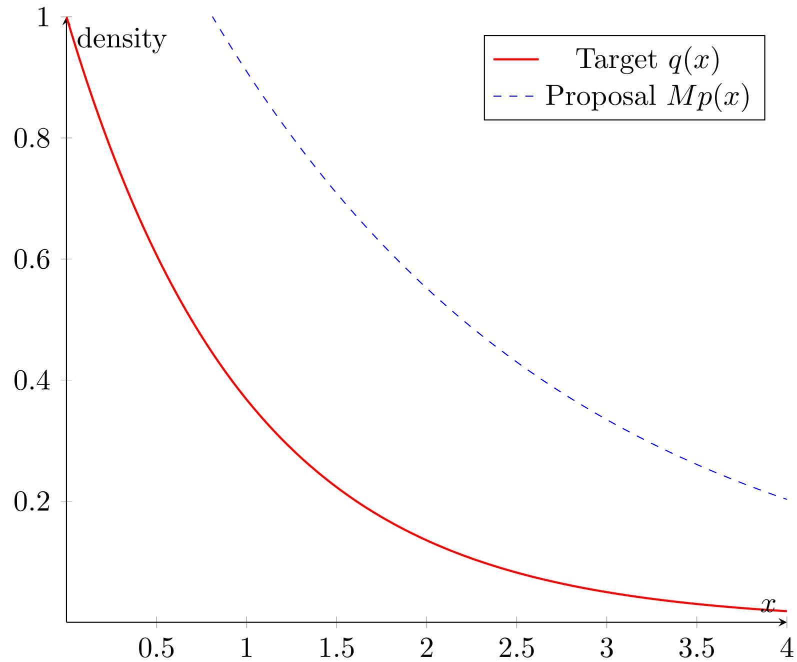
$$q(x) \leq Mp(x), \quad \forall x.$$

6.2 Algorithm

1. Sample $X \sim p(x)$.
2. Sample $U \sim U(0,1)$.
3. If $U \leq \dfrac{q(X)}{Mp(X)}$, accept X as a sample from q.
4. Otherwise, reject X and repeat from step 1.

6.3 Illustration

In the figure below, the red curve is the target density $q(x)$, the blue curve is the proposal density scaled by M, and accepted points fall below $q(x)$.



6.4 Correctness

The probability of accepting a proposed value x is proportional to $q(x)$. In an equation, the density of the accepted samples is:

$$P(Z \in A) = \int_A \frac{q(x)}{Mp(x)} p(x)\, dx = \frac{1}{M} \int_A q(x)\, dx,$$

which implies that the accepted distribution is exactly $q(x)$ (up to normalization).

7 Summary of Lecture

We have covered:

- The difference between discrete and continuous random variables.
- The definition and properties of CDFs.
- Examples of exponential, uniform, and normal distributions.
- Inverse transform sampling and its limitations.
- Rejection sampling.