Common Probability Distributions

Course: Data Mining

Professor: Dr. Tahaei

Herbod Pourali

December 2024

Some common distributions

This section provides descriptions of several common discrete and continuous distributions that are widely used in the context of machine learning and statistics.

Discrete Random Variables

1. Bernoulli Distribution ($X \sim$ Bernoulli(p))
The Bernoulli distribution represents a single trial of a binary experiment. It models a random variable X that can take only two values: 1 (success) with probability p, and 0 (failure) with probability $1-p$. A classic example is a coin toss, where p is the probability of getting heads (success).

The probability mass function (PMF) is given by:

$$p(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

where $0 \leq p \leq 1$.

2. Binomial Distribution ($X \sim$ Binomial(n, p))
The binomial distribution models the number of successes in n independent trials of a Bernoulli experiment, each with success probability p. It is used when the experiment is repeated several times, and each trial is independent of the others.

The PMF is given by:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where x is the number of successes, n is the total number of trials, and p is the probability of success in a single trial.

3. Geometric Distribution ($X \sim$ Geometric(p))
The geometric distribution models the number of trials required to get the first success in a sequence of independent Bernoulli trials. The probability of success in each trial is p, and the random variable represents how many trials are needed before the first success occurs.

The PMF is given by:

$$p(x) = p(1-p)^{x-1}$$

where $x \geq 1$ and $0 < p \leq 1$.

4. Poisson Distribution ($X \sim$ Poisson(λ))
The Poisson distribution is used to model the number of events occurring in a fixed interval of time or space when these events happen independently and at a constant average rate. It is often used to model rare events, such as the number of accidents at an intersection or the number of emails arriving in an inbox per hour.

The PMF is given by:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where $x \geq 0$ and $\lambda > 0$ is the average rate of occurrence.

Continuous Random Variables

5. Uniform Distribution ($X \sim$ Uniform(a, b))
The uniform distribution models a situation where every outcome within a specified range has an equal probability of occurring. If the random variable X is uniformly distributed between a and b, the probability density function (PDF) is constant between these two values.

The PDF is given by:

$$f(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b$$

and $f(x) = 0$ otherwise. This distribution is useful when all outcomes are equally likely within a defined range.

6. Exponential Distribution ($X \sim$ Exponential(λ))
The exponential distribution models the time between events in a Poisson process, where events happen continuously and independently at a constant rate λ. It is commonly used to model the waiting times between events, such as the time between arrivals of customers at a service center or the lifetime of a product.

The PDF is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda > 0$ is the rate parameter.

7. Normal (Gaussian) Distribution ($X \sim$ Normal(μ, σ^2))
The normal distribution is one of the most common continuous distributions, used to model many natural phenomena. It is characterized by its bell-shaped curve and is fully defined by two parameters: the mean μ (the center of the distribution) and the variance σ^2 (which controls the spread).

The PDF is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where $x \in R$, and μ and σ^2 are the mean and variance of the distribution, respectively.

Examples for Each Distribution

Bernoulli Distribution Example
Consider a biased coin with a probability $p = 0.7$ of landing heads. Let X be a random variable representing the outcome of a single coin flip. Then $X \sim$ Bernoulli(0.7). The probability of heads (1) is 0.7, and the probability of tails (0) is 0.3.

Binomial Distribution Example
Suppose you flip a biased coin 5 times, where the probability of heads is $p = 0.6$. Let X represent the number of heads in 5 flips. Then $X \sim$ Binomial$(5, 0.6)$. The probability of getting exactly 3 heads is:

$$p(3) = \binom{5}{3}(0.6)^3(0.4)^2 = 10 \times 0.216 \times 0.16 = 0.3456$$

Geometric Distribution Example

Let's say we are flipping a coin, where the probability of heads is $p = 0.2$. The random variable X represents the number of flips needed to get the first heads. Then $X \sim \text{Geometric}(0.2)$. The probability of getting the first heads on the third flip is:

$$p(3) = 0.2(1 - 0.2)^{3-1} = 0.2 \times 0.8^2 = 0.128$$

Poisson Distribution Examples

1. Suppose that on average, 3 cars pass a particular intersection every minute. The number of cars passing the intersection in a given minute follows a Poisson distribution with $\lambda = 3$. The probability of exactly 4 cars passing in a minute is:

$$p(4) = \frac{3^4 e^{-3}}{4!} = \frac{81e^{-3}}{24} \approx 0.168$$

2. Suppose that the probability that an item produced by a certain machine will be defective is 0.1. Find the probability that a sample of 10 items will contain at most 1 defective item.

The desired probability is given by:

$$P(X \leq 1) = \binom{10}{0}(0.1)^0(0.9)^{10} + \binom{10}{1}(0.1)^1(0.9)^9 = 0.3487 + 0.3874 = 0.7361$$

Poisson Approximation:

Using the Poisson approximation with parameter $\lambda = 10 \times 0.1 = 1$, we have:

$$P(X \leq 1) \approx e^{-1} + e^{-1}$$

Calculating this:

$$P(X \leq 1) \approx e^{-1} + e^{-1} = 2e^{-1} \approx 0.7358$$

Thus, the Poisson approximation gives a value of approximately 0.7358.

Uniform Distribution Example

Suppose a random variable X represents the height of a person in a specific region, and it is uniformly distributed between 150 cm and 200 cm. Then $X \sim \text{Uniform}(150, 200)$, and the probability density function is constant between these two values. The probability of a height X being between 160 cm and 170 cm is:

$$P(160 \leq X \leq 170) = \frac{170 - 160}{200 - 150} = \frac{10}{50} = 0.2$$

Exponential Distribution Example

Suppose a call center receives an average of 5 calls per hour. The time between consecutive calls follows an exponential distribution with rate $\lambda = 5$. The probability that the time until the next call is more than 10 minutes is:

$$P(X > 10) = e^{-5 \times 10/60} \approx 0.607$$

Normal Distribution Example

Let's assume the heights of adult women in a population are normally distributed with a mean of 160 cm and a standard deviation of 6 cm. Then $X \sim \text{Normal}(160, 6^2)$. The probability that a woman's height is between 154 cm and 166 cm can be calculated using the cumulative distribution function (CDF) of the normal distribution.

The CDF of a normal distribution $X \sim \text{Normal}(\mu, \sigma^2)$ is given by:

$$P(154 \leq X \leq 166) = \Phi\left(\frac{166-160}{6}\right) - \Phi\left(\frac{154-160}{6}\right)$$

where $\Phi(x)$ is the CDF of the standard normal distribution.

Using a standard normal table or calculator:

$$P(154 \leq X \leq 166) = \Phi(1) - \Phi(-1) \approx 0.8413 - 0.1587 = 0.6826$$