Statistics and Probability part 2

Garima Malik

May 13, 2022

Outline

Continuous Probability Distributions

Probability Distributions

Statistical Tests

Continuous Probability Distributions

CDF: Cumulative Distribution Function

- The PMF is one way to describe the distribution of a discrete random variable. As we will see later on, PMF cannot be defined for continuous random variables. The cumulative distribution function (CDF) of a random variable is another method to describe the distribution of random variables. The advantage of the CDF is that it can be defined for any kind of random variable (discrete, continuous, and mixed).

Definition 3.10

The cumulative distribution function (CDF) of random variable X is defined as

$$F_X(x) = P(X \leq x), \text{ for all } x \in \mathbb{R}.$$

PDF: Probability Density Function

- We have the following definition for the PDF of continuous random variables:

Definition 4.2

Consider a continuous random variable X with an absolutely continuous CDF $F_X(x)$. The function $f_X(x)$ defined by

$$f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x), \qquad \text{if } F_X(x) \text{ is differentiable at } x$$

is called the probability density function (PDF) of X.

Georgian

Properties of PDF

- Basic properties of PDF:

Consider a continuous random variable X with PDF $f_X(x)$. We have

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f_X(u)du = 1$.
3. $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(u)du$.

Example

- Let X be a continuous random variable with the following PDF :

$$f_X(x) = \begin{Bmatrix} ce^{-x} & x \geq 0 \\ 0 & otherwise \end{Bmatrix} \tag{1}$$

where c is a positive constant.

  - Find c.
  - Find the CDF of X, $F_X(x)$.
  - Find P(1<X<3).

Example

- I choose a real number uniformly at random in the interval [a,b], and call it X. By uniformly at random, we mean all intervals in [a,b] that have the same length must have the same probability. Find the CDF of X. The PDF of X is given as :

$$P(x) = \begin{Bmatrix} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{Bmatrix}$$

Exponential RV

A continuous random variable X is said to have an *exponential* distribution with parameter $\lambda > 0$, shown as $X \sim Exponential(\lambda)$, if its PDF is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Example

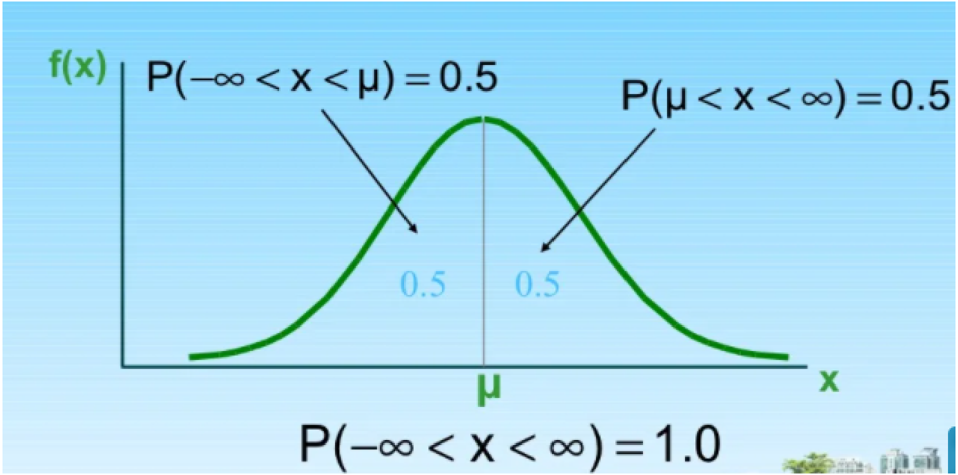
- Let X = amount of time (in minutes) a postal clerk spends with his or her customer. The time is known to have an exponential distribution with the average amount of time equal to four minutes. What will be the probability of when postal clerk takes 5 minutes with the customer?

Standard Normal(Gaussian) RV

A continuous random variable Z is said to be a *standard normal (standard Gaussian)* random variable, shown as $Z \sim N(0,1)$, if its PDF is given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}, \qquad \text{for all } z \in \mathbb{R}.$$

Empirical Properties



Georgian

Example

- If X is normally distributed normally between mean of 100 and standard deviation of 50 find the probability of $P(0 < z < 2.0)$?
  - General way of finding probability in Normal RV
    - Draw the normal curve
    - Translate x values to z values
    - Read the z table

Statistical Tests

- A statistical test provides a mechanism for making quantitative decisions about a process or processes. The intent is to determine whether there is enough evidence to "reject" a conjecture or hypothesis about the process.
- A classic use of a statistical test occurs in process control studies. For example, suppose that we are interested in ensuring that photomasks in a production process have mean linewidths of 500 micrometers. The null hypothesis, in this case, is that the mean linewidth is 500 micrometers. Implicit in this statement is the need to flag photomasks which have mean linewidths that are either much greater or much less than 500 micrometers. This translates into the alternative hypothesis that the mean linewidths are not equal to 500 micrometers.

Statistical Tests

- The null hypothesis is a statement about a belief. We may doubt that the null hypothesis is true, which might be why we are "testing" it. The alternative hypothesis might, in fact, be what we believe to be true. The test procedure is constructed so that the risk of rejecting the null hypothesis, when it is in fact true, is small. This risk, ‘alpha’, is often referred to as the significance level of the test. By having a test with a small value of alpha, we feel that we have actually "proved" something when we reject the null hypothesis.

Common format of hypothesis test

- H_0 : Null hypothesis
- H_a : Alternate hypothesis
- The test statistic is based on the specific hypothesis test.
- α : significance level

Examples

- A researcher thinks that if knee surgery patients go to physical therapy twice a week (instead of 3 times), their recovery period will be longer. Average recovery times for knee surgery patients is 8.2 weeks.
- A principal at a certain school claims that the students in his school are above average intelligence. A random sample of thirty students IQ scores have a mean score of 112.5. Is there sufficient evidence to support the principal's claim? The mean population IQ is 100 with a standard deviation of 15.
- Blood glucose levels for obese patients have a mean of 100 with a standard deviation of 15. A researcher thinks that a diet high in raw cornstarch will have a positive or negative effect on blood glucose levels. A sample of 30 patients who have tried the raw cornstarch diet have a mean glucose level of 140. Test the hypothesis that the raw cornstarch had an effect.