Random Variables

Code authors: Jinhui Li, Viviana Monje.

Corresponding author: vmonje@buffalo.edu

Recommended Matlab: >= 2023a.

Table of Contents

Usage Rights

These materials were created for educational purposes and to be completed as part of material covered in the CE305: Probability, Statistics, and Data Analysis undergraduate course. Students will have a similar notebook to turn in as part of the weekly assignment, which will be graded for completion and not correctness towards the participation category of their final grade.

Learning Objectives

By the end of this notebook, students will be able to:

- **Part I (Lecture 1)**:
- Define a random variable for an experiment;
- Differentiate between discrete and continuous random variables;
- Compute and interpret descriptive statistics of random variables;
- Differentiate between probability mass functions and probability density functions;
- Construct the appropriate probability distribution for a discrete and a continuous random variable;
- Use the definition of cumulative distribution functions (CDFs) to compute event probabilities;
- **Part II (Lecture 2):**
- Understand linear functions and combinations of random variables;
- Interpret the impact of sample size on the variability of statistical measures.

- Determine whether a continuous random variable is appropriate to describe data trends for a discrete random variable.

Part I (Lecture 1)

Random Variables

Basics

A random variable assigns a numerical value to each outcome in a sample space.

Two types of random variables are: discrete and continuous.

Detailed discussion and explanations are found in the Textbook, Ch. 2.4 Random Variables. (Navidi Textbook, 6th Ed)

Example

In a lottery game, there are three possible outcomes:

- Prize A: worth $100;
- Prize B: worth $50;
- No prize: worth $0.

Each time the game is played, exactly one outcome occurs, and all three outcomes are equally likely.

- Question 1: Define the experiment.

- Question 2: Define a random variable Y to represent the outcomes of the experiment.

- Question 3: Is Y a continuous or a discrete random variable?

- Question 4: What values can Y take?

- Question 5: What is the probability of each of the possible outcomes?

Discrete Random Variables (DRVs)

Basics

Definition: A random variable is discrete if it can take distinct (separate), countable values.

Let X be a discrete random variable, then:

- The probability distribution, called specifically probability mass function for DRVs of X is denoted as $p(x) = P(X = x)$. It is the list of associated probabilities for each value X can take.
- The cumulative distribution function of X is denoted as $F(x) = P(X \leq x)$. It is the added probability of all outcomes of X until a value x of interest.
- The following always hold:

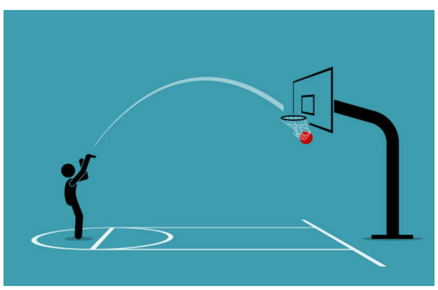
$F(x) = \sum_{t \leq x} p(t) = \sum_{t \leq x} P(X = t)$.

$\sum_{x} p(x) = \sum_{x} P(X = x) = 1$, where the sum is over all the possible values of X.

- The mean of X, or its expected value, is denoted by $E(x) = \mu_X = \sum_x x\mathrm{P}(X = x)$
- The variance of X is denoted by $V(x) = \sigma_X^2 = \sum_x (x - \mu_X)^2 P(X = x) = \sum_x x^2 P(X = x) - \mu_X^2$.
- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\sigma_X^2}$.

More details can be found in Ch. 2.4 Random Variables. (Navidi Textbook, 6th Ed)

Example



Let X be the number of successful free throws made by a basketball player in 2 attempts.

- Question 1: Define the experiment.

- Question 2: Select the definition of the random variable X from the options below:

A) X = the time (in seconds) taken to complete 2 free throws.

B) X = the number of successful free throws made by the basketball player in 2 attempts.

C) X = whether the player makes the first free throw (1 = success, 0 = failure).

- Question 3: List the possible values for X in this experiment and input them in the box below.

- Question 4: The probability distribution of X is the list of its possible values and their associated probabilities. Select and input a probability for each value listed in the previous question and input them below to build the PMF and plot it for this DRV.

- Question 5: Compute the cumulative distribution function F(x) for all values of x∈{0,1,2}.

- Question 6: Compute the expected value E(X).

Continuous Random Variables (CRVs)

Basics

A continuous random variable can take on **any** value over the range it is defined. The probability of a CRV is given by area under its probability distribution, called probability density function. Therefore, the probability of a CRV taking on a single point value is always zero.

Let X be a continuous random variable with probability density function $f(x)$.

- Let a and b be any two numbers, with a < b, then:
  $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = \int_a^b f(x)\text{dx}$; $P(X \leq b) = P(X < b) = \int_{-\infty}^b f(x)\text{dx}$.
- The total probability is: $\int_{-\infty}^{+\infty} f(x)\,\text{dx} = 1$
- The cumulative distribution function of X is: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\text{dt}$.
- The mean of X or expectated value is $E(X) = \mu_X = \int_{-\infty}^{+\infty} \text{xf}(x)\text{dx}$.
- The variance of X is $V(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\text{dx} = \int_{-\infty}^{\infty} x^2 f(x)\text{dx} - \mu_X^2$.
- The standard deviation is the square root of the variance: $\sigma_X = \sqrt{\sigma_X^2}$

More details can be found in Ch. 2.4 Random Variables. (Navidi Textbook, 6th Ed)

Example

A new vending machine can make a drink automatically after payment. Based on records, the preparation time is uniformly distributed between 0 and 2 minutes. Let X be the time (in minutes) it takes to make one drink.

- Question 1: Define the experiment

- Question 2: Select the appropriate definition, in words, for this random variable.

A) X = the number of drinks produced by the vending machine in 2 mins.

B) X = whether the machine successfully makes a drink (1 = success, 0 = failure).

C) X = the time (mins) it takes the vending machine to make a drink.

Suppose that the probability density function (PDF) of X is given by:

$$P(x) = \begin{cases} \frac{1}{2}, & 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

- Question 3: Verify that P(x) above is a valid probability density function. (Tip: $\int_{-\infty}^{\infty} P(x)\,dx = ?$)

- Question 4: What should be the upper limit of the cumulative distribution function for X?

- Questions 5: Find the probabilities for $P(0.5 \le X \le 1.5)$ and $P(X < 1)$.

The Population Median and Percentiles

Basics

Let X be a continuous random variable with probability density function $f(x)$ and cumulative distribution function $F(x)$

- The median of X is the point x_m that splits all possible values in two equal halves. It is found by solving the equation $F(x_m) = P(X \leq x_m) = \int_{-\infty}^{x_m} f(x)\mathrm{dx} = 0.5$.
- If p is any number between 0 and 100, the p-th percentile is the point x_p **below which** the p% of data is distributed. It is found by solving the equation $F(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f(x)\mathrm{dx} = \frac{p}{100}$.
- The median is the 50-th percentile.

More details can be found in Ch. 2.4 Random Variables. (Navidi Textbook, 6th Ed)

Example

A pharmaceutical company is testing a new capsule that slowly releases its active ingredient. Let X be the time (in hours) after swallowing the pill until 50% of the drug is released. Based on lab tests, the release time follows the following probability density function:

$$f(x) = \begin{cases} \frac{1}{4}x, & 0 \leq x \leq 4 \\ 0, & \text{otherwise} \end{cases}$$

- Question 1: Define the experiment.
- Question 2: Define, in words, the random variable for this experiment.
- Question 3: What is the median release time?
- Question 4: Find the 80th percentile of the release time.

Part II (next class)

Operations with Random Variables

Linear Combinations of Random Variables

Independence and Simple Random Samples