Random Variables

Code authors: Jinhui Li, Viviana Monje.

Corresponding author: vmonje@buffalo.edu

Recommended Matlab: >= 2023a.

Table of Contents

Usage Rights

Learning Objectives

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

- **Part I**:
- 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**
- 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.

Recap Part I (prev. lecture)

Random Variables (RVs)

- A random variable assigns a numerical value to each outcome in a sample space. Two types of random variables are: discrete and continuous.
- A DRV can take distinct (separate), countable values, whereas a CRV can take on any value over the range it is defined.

Probability Distributions

- A probability distribution is the list of probabilities associated with the value(s) a RV takes. It is referred to as a PMF for DRV, and as a PDF for CRV.
- The CDF of a random variable, denoted by, gives the probability that a RV takes on all values less than or equal to a value of interest in its range.
- The median of a probability distribution is the value of the RV, x_m, at which the data is divided in half
- The p-th percentile of a distribution is the value of the RV, x_p, below which p% of the distribution is found.
- The definition of the CDF can be used to determine the RV median, x_m, and p-th percentiles of interest, x_p

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

Part II

Linear Functions of Random Variables

Operations with a Constant

If X is a random variable, and a and b are constants, then:

- $\mu_{X+b} = \mu_X + b,\ \sigma^2_{X+b} = \sigma^2_X$
- $\mu_{\mathrm{aX}} = a\mu_X,\ \sigma^2_{\mathrm{aX}} = a^2\sigma^2_X,\ \sigma_{\mathrm{aX}} = |a|\sigma_X$
- $\mu_{\mathrm{aX}+b} = a\mu_X + b,\ \sigma^2_{\mathrm{aX}+b} = a^2\sigma^2_X,\ \sigma_{\mathrm{aX}+b} = |a|\sigma_X$

Linear Combinations of Random Variables

If $X_1,\ldots,X_n$ are random variables and $c_1,\ldots,c_n$ are constants, we can get the linear combination of $X_1,\ldots,X_n$: $c_1X_1 + \ldots + c_nX_n$, then:

- $\mu_{c_1X_1+\ldots+c_nX_n} = c_1\mu_{X_1} + \ldots + c_n\mu_{X_n}$

Linear Combinations of Independent Random Variables

if $X_1,\ldots,X_n$ are independent random variables, $c_1,\ldots,c_n$ are constants, and $S_1,\ldots,S_n$ are sets, then:

- $P(X_1 \in S_1 \text{ and} \ldots \text{and } X_n \in S_n) = P(X_1 \in S_1) \ldots P(X_n \in S_n)$

- $\sigma^2_{c_1X_1+\ldots+c_nX_n} = c_1^2\sigma^2_{X_1} + \ldots + c_n^2\sigma^2_{X_n}$

Independence and Simple Random Samples

Independent and identically distributed (i.i.d.) sample: $X_1, \ldots, X_n$ are independent random variables and all have the same distribution.

If $X_1, \ldots, X_n$ is a simple random sample from a population with mean � and variance σ^2, then:

- The sample mean $\overline{X}$ will be a random variable with $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$
- The standard deviation of $\overline{X}$ is: $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

More details can be found in Ch. 2.5 Linear Functions of Random Variables (Navidi Textbook, 6th Ed)

Example 1

A GPS app tracks a cyclist's average travel time per mile during a ride and provides the following data.

- Mean travel time: $\mu_X = 4$ minutes.
- Standard deviation: $\sigma_X = 0.5$ minutes.
- Question 1: Define the experiment.

- Question 2: Define the random variable in words.

Now, the app wants to convert this to speed in miles per hour. The speed in mph is calculated as: $Y = \frac{60}{X}$. Since this transformation is not linear, the app also considers a linear approximation near $X = 4$, using the function:

- $Y_{\text{approx}} = -3X + 20$

- Question 3: Find the expected speed $\mu_{Y_{\text{approx}}}$.

- Question 4: Find the standard deviation $\sigma_{Y_{\text{approx}}}$ of the approximate speed.

Example 2

At a popular coffee shop, the average waiting time for a customer is known to be $\mu = 6$ minutes, with a standard deviation of $\sigma = 3$ minutes. A manager wants to estimate the average wait time on a given day by randomly sampling the wait times of several customers.

Let $X_1, \ldots, X_n$ be the set of RVs for the waiting time of a random sample of n customers. Define $\overline{X}$ as the sample mean wait time.

- Question 1: Define the experiment

- Question 2: Define, in words, the random variable X_n

- Question 3: If the manager samples n=36 customers, what is the mean and standard deviation of $\overline{X}$?

- Question 4: Explain why the sample mean becomes more stable (less variable) as n increases.

- Question 5: What sample size n is needed to reduce the standard deviation of $\overline{X}$ to 0.2 minute?