Introduction to Statistical Modeling MSDS 598

Probability and Statistics
(Quick Overview)

Michael Ruddy

Probability

- “The likelihood that an event will occur”

- An **experiment** is an activity or procedure that produces distinct, well-defined possibilities called **outcomes.**
- The set of all possible outcomes is called the **sample space.**

- Formally: probability assigns a numerical likelihood to each outcome.
- Follows rules
  - Between 0 and 1
  - Sum (or integral) is equal to 1

Classic Examples

- Experiment: flipping a coin
  - Sample Space: {H, T}
  - P(H) = P(T) = (½)
  - Or maybe not?!

- Experiment: Rolling a 6-sided die
  - Sample Space: {1, 2, 3, 4, 5, 6}
  - P(1) = ... = P(6) = (⅙)
  - P(1 or 2 or 3) = (½)

- Experiment: Rolling two 6-sided dice
  - Sample Space? P((1, 5) or (2, 6))?

If all outcomes are equally likely:

$$P = \frac{\text{\# of outcomes}}{\text{\# of outcomes in sample space}}$$

Other scenarios

- Often the sample space is not discrete, but continuous
  - Counting is replaced with measuring lengths of intervals

- Most of the time we do not know how to exactly assign probabilities to outcomes, even when discrete, i.e. probability of rain in SF tomorrow?

- **Empirical Probability**: Probability estimate obtained by running the experiment many times.
  - Flip a coin many times
  - Historical weather patterns

$$P = \frac{\text{\# of times outcome occurred}}{\text{\# of times experiment performed}}$$

Random Variables

- A **random variable** is a *function* that assigns a numerical value to each outcome in the sample space
  - Rolling two 6-sided dice: Sum of the result
  - Flip a coin 100 times: Number of times heads appears

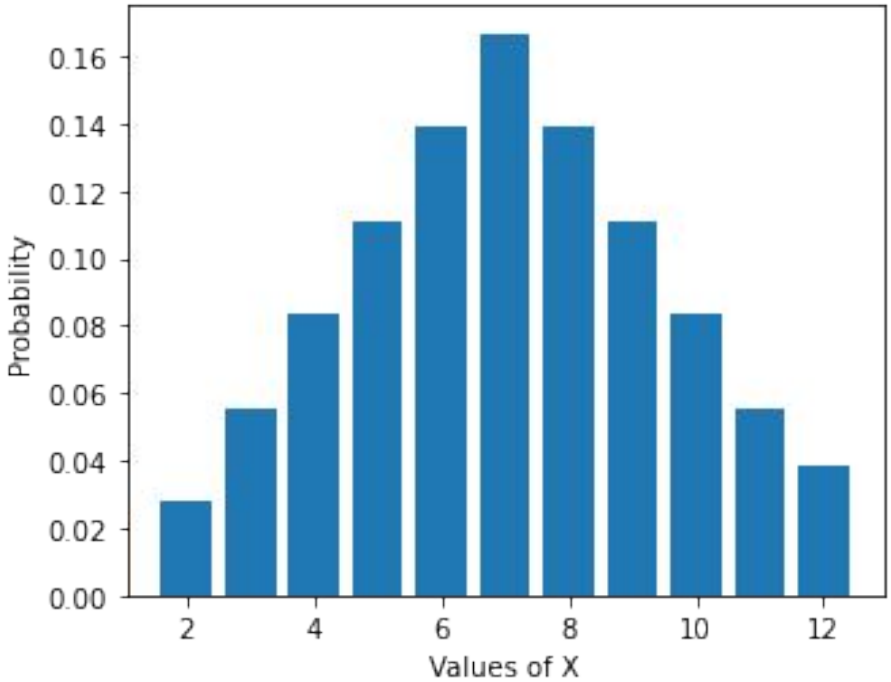
Two 6-sided Dice

- Sample Space: {(1,1), (1,2), (1,3), ... , (6,6)}
- Random Variable X: {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

- Does P(X=6) = P(X=12)? Why?

- What is $\sum_{i=2}^{12} P(X = i)$?

Distribution of X



Probability Density Function P(X = k)