Discrete Random Variables

Ryan Miller

Introduction

- Studying probability is needed to understand *random processes*
  - Key examples include sampling from a population, or assigning into treatment/control group and measuring an outcome

Introduction

- Studying probability is needed to understand *random processes*
  - Key examples include sampling from a population, or assigning into treatment/control group and measuring an outcome
- Statisticians represent the *unknown numeric outcome of a random process* using a **random variable**

Introduction

- Studying probability is needed to understand *random processes*
  - Key examples include sampling from a population, or assigning into treatment/control group and measuring an outcome
- Statisticians represent the *unknown numeric outcome of a random process* using a **random variable**
- Consider flipping a fair coin
  - We can represent this random process with the random variable X
  - Because random variables pertain to numeric outcomes, we'd let $X = 1$ if the outcome "heads" is observed, and $X = 0$ if the outcome "tails" is observed

Example #1

- In the National Football League (NFL), after each touchdown the scoring team gets to choose between a 1-pt and 2-pt attempt to earn additional points (on top of the 6 given for scoring a touchdown)

Example #1

- In the National Football League (NFL), after each touchdown the scoring team gets to choose between a 1-pt and 2-pt attempt to earn additional points (on top of the 6 given for scoring a touchdown)
- We can use the random variable X to denote the number of total points the team earns from a touchdown
  - Notice this is a numeric outcome that is unknowable in advance

Example #1

- In the National Football League (NFL), after each touchdown the scoring team gets to choose between a 1-pt and 2-pt attempt to earn additional points (on top of the 6 given for scoring a touchdown)
- We can use the random variable X to denote the number of total points the team earns from a touchdown
  - Notice this is a numeric outcome that is unknowable in advance
- Since a rule change implemented in 2015, 9.6% of touchdowns were accompanied by zero additional points, 86.5% resulted in one additional point, and 3.9% resulted in two additional points
  - Based upon these data, we can apply the following *probability model* for the points resulting from a touchdown:

X	6	7	8
$P(X = x)$	0.096	0.865	0.039

Probability Models

Probability models are useful because they help us understand a few key aspects of a random process:

1) **Expected Value**, or the "average" numeric outcome
2) **Variance**, or the total amount that the numeric outcomes vary from their *expected value*
3) **Standard Deviation**, or the "average" amount that numeric outcomes vary from their *expected value*

Expected Value

- The **expected value** of a random variable is denoted $E(X)$
- It describes the *expected result*, which is the sum of each possible outcome weighted by its probability

X	6	7	8
P(X = x)	0.096	0.865	0.039

- For a randomly chosen NFL touchdown, $E(X) = 6 * 0.096 + 7 * 0.865 + 8 * 0.039 = 6.94$ points

Variance

To see how much each possible outcome (6, 7, or 8 pts) varies from the expected outcome (6.94 pts) we can calculate their *squared deviations*

Points	6	7	8
Deviation	(6-6.94)^2	(7-6.94)^2	(8-6.94)^2

If we add these squared deviations, weighted by their probabilities, we get **variance**:

$$\text{Var}(X) = 0.096*(6-6.94)^2+0.865*(7-6.94)^2+0.039*(8-6.94)^2 = 0.13$$

Variance

To see how much each possible outcome (6, 7, or 8 pts) varies from the expected outcome (6.94 pts) we can calculate their *squared deviations*

Points	6	7	8
Deviation	(6-6.94)^2	(7-6.94)^2	(8-6.94)^2

If we add these squared deviations, weighted by their probabilities, we get **variance**:

$$\text{Var}(X) = 0.096*(6-6.94)^2+0.865*(7-6.94)^2+0.039*(8-6.94)^2 = 0.13$$

Taking the square-root of the variance, we have the **standard deviation**, or the average deviation of possible outcomes of X from the expected value:

$$\text{SD}(X) = \sqrt{\text{Var}(X)} = \sqrt{0.13} = 0.36$$

So, we expect the average deviation (from the expected value of

Summary

- A **random variable** is used to represent the unknown numeric outcome of a random process
  - We can be creative and use random variables involving 0's and 1's to represent binary outcomes like "heads" or "tails"

Summary

- A **random variable** is used to represent the unknown numeric outcome of a random process
  - We can be creative and use random variables involving 0's and 1's to represent binary outcomes like "heads" or "tails"
- Random variables are connected to probabilities via **probability models**
  - For variables with discrete outcomes, these are often laid out in tabular format

Summary

- A **random variable** is used to represent the unknown numeric outcome of a random process
  - We can be creative and use random variables involving 0's and 1's to represent binary outcomes like "heads" or "tails"
- Random variables are connected to probabilities via **probability models**
  - For variables with discrete outcomes, these are often laid out in tabular format
- Probability models are useful because they allow us to understand the **center** and **spread** of the random variable
  - The center is expressed using the **expected value**, or average outcome
  - The spread is expressed using the **standard deviation**, or the average deviation from the average outcome
  - The standard deviation is calculated as the square-root of the **variance** of the random variable