# CS 112: Modeling Uncertainty in Information Systems

Prof. Jenn Wortman Vaughan

April 16, 2012

Lecture 5

# Today

- Random variables
- Examples of common discrete random variables
  - Bernoulli random variables
  - Binomial random variables
  - Geometric random variables
  - Poisson random variables
- Expectation

# Random Variables

A random variable is a real-valued function of the outcome of an experiment.

# Random Variables

A random variable is a real-valued function of the outcome of an experiment…. though we typically won't write them as functions.

# Random Variables

A random variable is a real-valued function of the outcome of an experiment…. though we typically won't write them as functions.

Example: Suppose we roll two six-sided dice.

- Sample space $\Omega = \{11, 12, 13, ..., 66\}$
- Random variables: the sum of the two rolls, the number of sixes rolled, the product of the two rolls, the value of the first roll, the number of even rolls, etc.

# Random Variables

Example: Suppose we have four disks, each of which fails with probability $p$.

# Random Variables

Example: Suppose we have four disks, each of which fails with probability $p$.

- Sample space $\Omega = \{0000, 0001, 0010, ..., 1111\}$

# Random Variables

Example: Suppose we have four disks, each of which fails with probability $p$.

- Sample space $\Omega = \{0000, 0001, 0010, ..., 1111\}$
- Random variables: the number of disks that failed, a Boolean value indicating if one of the first two disks failed, a Boolean value indicating if at least one disk failed, etc.

# Random Variables

A random variable is called discrete if the set of values that it can take on is finite or countably infinite. For the next few lectures, we will only discuss random variables that are discrete.

# Random Variables

A random variable is called discrete if the set of values that it can take on is finite or countably infinite. For the next few lectures, we will only discuss random variables that are discrete.

In this class, we will use upper case letters for random variables and lower case letters for their values.

# Random Variables & Events

- The event $\{X=k\}$ is the event consisting of all outcomes in $\Omega$ that are mapped to the value $k$ by random variable X

# Random Variables & Events

- The event $\{X=k\}$ is the event consisting of all outcomes in $\Omega$ that are mapped to the value $k$ by random variable X

- Suppose we flip two coins…
  - If X is the number of heads we see, what is $\{X=1\}$?

# Random Variables & Events

- The event $\{X=k\}$ is the event consisting of all outcomes in $\Omega$ that are mapped to the value $k$ by random variable X

- Suppose we flip two coins…
  - If X is the number of heads we see, what is $\{X=1\}$?
  - If Y is a Boolean value that is 1 if the flips are different and 0 otherwise, what is $\{Y=0\}$?

# Random Variables & Events

- The event $\{X=k\}$ is the event consisting of all outcomes in $\Omega$ that are mapped to the value $k$ by random variable X

- Suppose we flip two coins…
  - If X is the number of heads we see, what is $\{X=1\}$?
  - If Y is a Boolean value that is 1 if the flips are different and 0 otherwise, what is $\{Y=0\}$?

- We'll often be lazy and write $P(\{X=k\})$ as $P(X=k)$

# Probability Mass Functions

The probability mass function (PMF) of a random variable X is denoted $p_X$. For any value $k$ that X can take on, $p_X(k)$ is the probability of the event $\{X=k\}$, so

$$p_X(k) = P(\{X=k\}) = P(X=k)$$

# Probability Mass Functions

The probability mass function (PMF) of a random variable X is denoted $p_X$. For any value $k$ that X can take on, $p_X(k)$ is the probability of the event $\{X=k\}$, so

$$p_X(k) = P(\{X=k\}) = P(X=k)$$

Example: Suppose we flip a fair coin twice. Let X be the number of heads. What is $p_X(0)$? What is $p_X(1)$?

# Probability Mass Functions

- Let $\{X \in S\}$ denote the set of outcomes in $\Omega$ that are mapped to values in S by X. What is $P(X \in S)$?

# Probability Mass Functions

- Let $\{X \in S\}$ denote the set of outcomes in $\Omega$ that are mapped to values in S by X. What is $P(X \in S)$?

$$P(X \in S) = \sum_{values\ k \in S} p_X(k)$$

# Probability Mass Functions

- Let $\{X \in S\}$ denote the set of outcomes in $\Omega$ that are mapped to values in S by X. What is $P(X \in S)$?

$$P(X \in S) = \sum_{values\ k \in S} p_X(k)$$

- Example: Suppose we flip a fair coin twice. Let X be the number of heads. What is $P(X \geq 1)$?

# Probability Mass Functions

- Let $\{X \in S\}$ denote the set of outcomes in $\Omega$ that are mapped to values in S by X. What is $P(X \in S)$?

$$P(X \in S) = \sum_{values\ k \in S} p_X(k)$$

- Example: Suppose we flip a fair coin twice. Let X be the number of heads. What is $P(X \geq 1)$?

- What is $\sum_{all\ values\ k} p_X(k)$ ?

# Common Discrete Random Variables

# Bernoulli Random Variables

- Consider a biased coin that lands on heads with probability *p*. Let X be a random variable that takes on the value 1 if the coin is heads and 0 if the coin is tails. Such a variable is called a Bernoulli random variable.

# Bernoulli Random Variables

- Consider a biased coin that lands on heads with probability $p$. Let X be a random variable that takes on the value 1 if the coin is heads and 0 if the coin is tails. Such a variable is called a Bernoulli random variable.

- Examples: A server can be online or not. A new email that you receive can be spam or not. A web user can click or not click on an ad that is displayed.

# Bernoulli Random Variables

- Consider a biased coin that lands on heads with probability $p$. Let X be a random variable that takes on the value 1 if the coin is heads and 0 if the coin is tails. Such a variable is called a Bernoulli random variable.

- Examples: A server can be online or not. A new email that you receive can be spam or not. A web user can click or not click on an ad that is displayed.

- How can we calculate $p_X(k)$ for a particular value of $k$?

# Binomial Random Variables

- Suppose we flip the biased coin $n$ times. Let X be the total number of heads that we see. Such a variable is called a binomial random variable.

# Binomial Random Variables

- Suppose we flip the biased coin *n* times. Let X be the total number of heads that we see. Such a variable is called a binomial random variable.

- Examples: Number of our *n* servers that are online, number of emails that are spam when we receive *n* emails, number of users who click on an ad when *n* users come to a site.

# Binomial Random Variables

- Suppose we flip the biased coin $n$ times. Let X be the total number of heads that we see. Such a variable is called a binomial random variable.

- Examples: Number of our $n$ servers that are online, number of emails that are spam when we receive $n$ emails, number of users who click on an ad when $n$ users come to a site.

- How can we calculate $p_X(k)$ for a particular value of $k$?

# Binomial Random Variables

A computer lab contains 20 computers. Each computer is infected with a virus independently with probability 0.1. What is the probability that exactly two computers are infected?

# Geometric Random Variables

- Suppose we flip the biased coin repeatedly until it lands on heads. Let X be the number of times that we have to flip it. Such a variable is called a geometric random variable.

# Geometric Random Variables

- Suppose we flip the biased coin repeatedly until it lands on heads. Let X be the number of times that we have to flip it. Such a variable is called a geometric random variable.

- Examples: Number of emails we receive until one is spam (including the spam email), number of users who come to a site until one clicks the ad (including the one who clicks).

# Geometric Random Variables

- Suppose we flip the biased coin repeatedly until it lands on heads. Let X be the number of times that we have to flip it. Such a variable is called a geometric random variable.

- Examples: Number of emails we receive until one is spam (including the spam email), number of users who come to a site until one clicks the ad (including the one who clicks).

- How can we calculate $p_X(k)$ for a particular value of $k$?

# Geometric Random Variables

A computer lab contains 20 computers. Each computer is infected with a virus independently with probability 0.1. You examine the computers one by one to see if each has a virus. What is the PMF of the number of computers you need to check before finding one with a virus or determining that there is none with a virus?

# Geometric Random Variables

A computer lab contains 20 computers. Each computer is infected with a virus independently with probability 0.1. You examine the computers one by one to see if each has a virus. What is the PMF of the number of computers you need to check before finding one with a virus or determining that there is none with a virus?

Almost a geometric random variable, except we cap the number of trials.

# Expectation

# Expectation

Suppose you take a trip to Las Vegas and play a slot machine $n$ times. Each time you play, you receive a random payout from a set of possible payouts $\{r_1, r_2, \ldots, r_m\}$, where for each $i$, the probability of receiving $r_i$ is $p_i$.

Let X be a random variable denoting your average payout (that is, total payout divided by $n$).

How much of a payout should you "expect" to get on average each time you play?

# Expectation

- The expected value of a random variable X with probability mass function $p_X$ is

$$E[X] = \sum_k k p_X(k)$$

- We sometimes call this the expectation or mean of X.

# Expectation

- The expected value of a random variable X with probability mass function $p_X$ is

$$E[X] = \sum_{k} k p_X(k)$$

- We sometimes call this the expectation or mean of X.

- This is simply the value we'd "expect" to get for X "on average" if we repeated the same experiment many times.

# Examples of Expectation

- What is the expected value of a Bernoulli random variable with parameter $p$?

# Examples of Expectation

- What is the expected value of a Bernoulli random variable with parameter $p$?

- What is the expected value of a binomial random variable with parameters $p$ and $n$?

# Examples of Expectation

- What is the expected value of a Bernoulli random variable with parameter $p$?

- What is the expected value of a binomial random variable with parameters $p$ and $n$?

  (We'll come back to why this is true…)

# Poisson Random Variables

- Consider a book containing $n$ words. Suppose that the probability of any particular word being misspelled is $p$. Let X be the number of words that are misspelled.

# Poisson Random Variables

- Consider a book containing $n$ words. Suppose that the probability of any particular word being misspelled is $p$. Let X be the number of words that are misspelled.

- We could model X as a binomial random variable, but when $n$ is large and $p$ is small, it is often more convenient to model it as a Poisson random variable.

PMFs are approximately the same for large $n$ and small $p$.

# Poisson Random Variables

- A Poisson random variable with parameter $\lambda$ takes on values in $\{0, 1, 2, ...\}$ and has the PMF

$$p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

- The expected value is simply $\lambda$

# Poisson Random Variables

- A Poisson random variable with parameter λ takes on values in {0, 1, 2, ...} and has the PMF

$$p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

- The expected value is simply λ

- If we want our random variable to represent the number of misspelled words in a book with *n* words, each misspelled with probability *p*, how should we set λ?