Stat 110

Unit 3: Random Variables

Chapter 3 in the Text

Unit 3 Outline

- Random Variables (RVs) and Distributions
- Discrete RVs and Probability Mass Functions (PMFs)
- Bernoulli and Binomial Distributions
- Hypergeometric Distribution
- Discrete Uniform Distribution
- Cumulative Distribution Functions (CDFs)
- Functions of Random Variables
- Independence of Random Variables

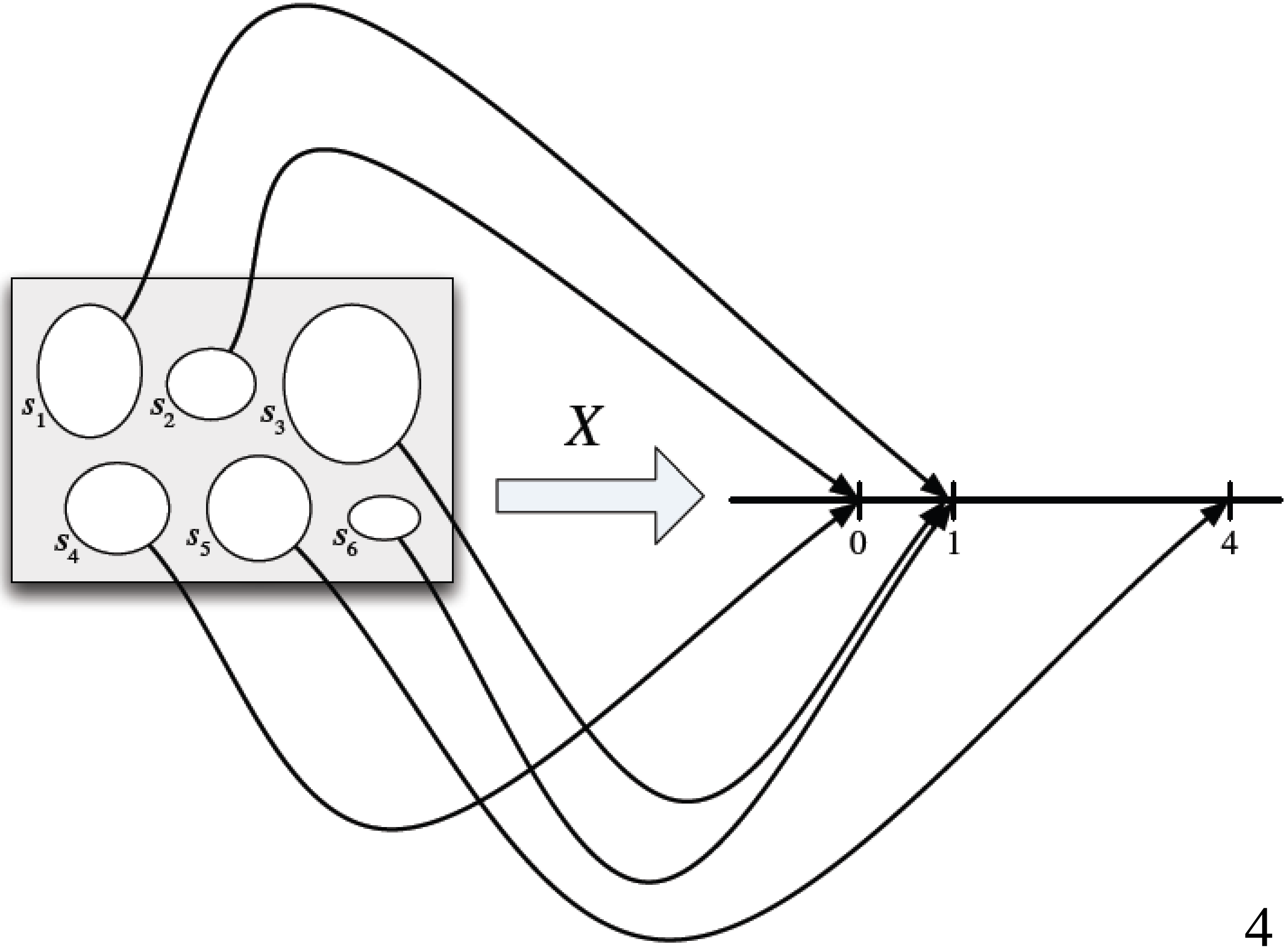
Definition: Random Variable

- Let S be the sample space for an experiment. A function that maps the outcomes in S to the real line is called a ***random variable*** (text, p.92), often written as r.v.
- We usually use capital letters at the end of the alphabet (X, Y, V, etc…) to denote random variables, but this is just convention.
- Example: Let X = # heads in 3 flips of a coin.

S = {TTT,HTT,THT,TTH,THH,HTH,HHT,HHH}

values for X = {0, 1, 2, 3}

Random Variable as a function (mapping)



Definition: Distribution

- Let X be a random variable. The ***distribution*** of X is the collection of all probabilities of the form $P(X \in C)$ for all sets C of real numbers such that $\{X \in C\}$ is an event.
- What the heck does that mean?
- Really, it's just listing/defining what all of the probabilities are for all of the possible values of X (of course, the values of X are on the real line).
- The definition is written in specific language to eventually allow for continuous random variables (that is, random variables that can fall anywhere within a specific range on the real line. Example: time it takes to complete an exam).

Distribution of a Random Variable

- Be careful: you can only find probabilities of events. So be sure to use proper notation.
- $P(X = 2)$ is good notation. $P(X)$ or $P(2)$ is bad.

- They are still probabilities, so they must follow the axioms. Specifically:
  - For each $\{X \in C\}$, $0 \leq P(X \in C) \leq 1$
  - The "sum of all" probabilities equals 1. That for a partition of S: $P(X \in C_1) + \ldots + P(X \in C_n) = 1$.
  - The probability of a union of **disjoint** events add. So if C_1 and C_2 are disjoint, then:
    $P(X \in \{C_1 \cup C_1\}) = P(X \in C_1) + P(X \in C_2)$

Unit 3 Outline

- Random Variables (RVs) and Distributions
- Discrete RVs and Probability Mass Functions (PMFs)
- Bernoulli and Binomial Distributions
- Hypergeometric Distribution
- Discrete Uniform Distribution
- Cumulative Distribution Functions (CDFs)
- Functions of Random Variables
- Independence of Random Variables

Discrete Random Variables

- A random variable X is a ***discrete random variable*** if X can take only a finite number k of different values $x_1, \ldots, x_k$ or at most an infinite sequence of different $x_1, x_2, \ldots$ (aka, countably infinite).
- Examples:
  - X = # heads in 3 flips of a coin.
  - T = sum of two fair 6-sided dice (T for total)
  - W = # games until Harvard football loses its next game.

Probability Mass Function (*PMF*)

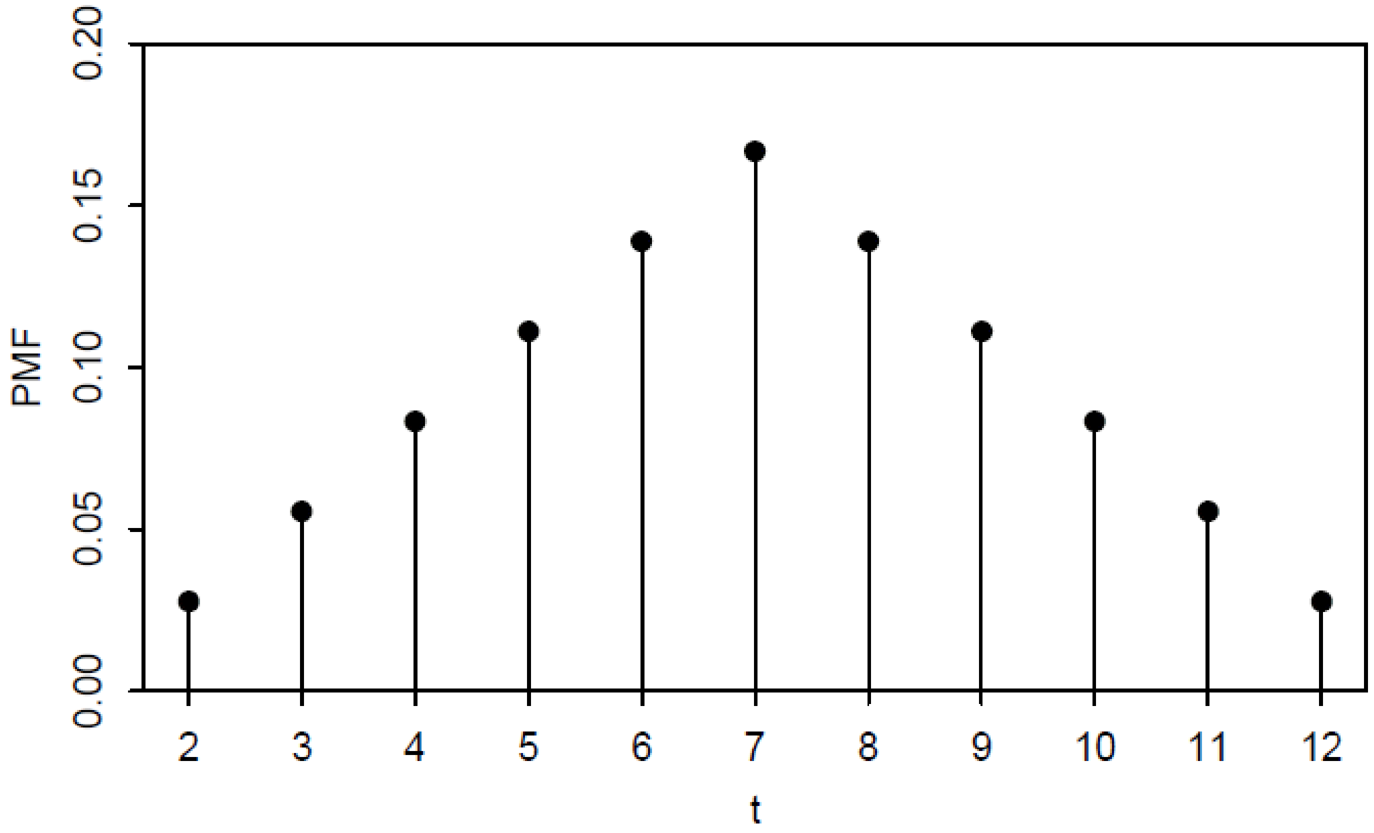
- For a discrete random variable X, the ***probability mass function*** (or simply just *probability function*) of X is defined as the function f such that for every real number x,

$$f(x) = P(X = x)$$

- The closure of the set $\{x: f(x) > 0\}$ is called the ***support*** of the distribution of X.
- Examples:
  - X = # heads in 3 flips of a coin.
  - T = sum of two fair 6-sided dice

Plot of a discrete PMF

- T = sum of two fair 6-sided dice



Unit 3 Outline

- Random Variables (RVs) and Distributions
- Discrete RVs and Probability Mass Functions (PMFs)
- Bernoulli and Binomial Distributions
- Hypergeometric Distribution
- Discrete Uniform Distribution
- Cumulative Distribution Functions (CDFs)
- Functions of Random Variables
- Independence of Random Variables

Story of the Bernoulli Distribution

- Imagine you are going to flip a coin once, with known probability p of showing heads.
- Let X = # heads.
- What is X's distribution? That is, what is the probability mass function for X? (don't forget to mention X's support)
- Extension #1: you are going to perform an experiment once that has probability p of being a success . Let X = # successes.
- Extension #2: you are going to randomly sample a single individual from a population and measure whether or not they have a specific characteristic, and p is the proportion of individuals in the population with that characteristic.
  - Example: did one randomly sampled Massachusetts voter vote for Obama in 2012 ($p = 0.60$)?

Bernoulli Distribution Definition

- An r.v. X is said to have the ***Bernoulli distribution*** with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$, where $0 < p < 1$. We write this as $X \sim \text{Bern}(p)$. The symbol $\sim$ is read "is distributed as".
- Any r.v. whose outcomes are just 0 and 1 has a Bern(p) distribution, where p represents the probability of $X = 1$.
- The number p is called the ***parameter*** of the distribution; it determines which specific Bernoulli distribution that X follows.
- This means that there is not just one distribution, but rather a family of Bernoulli distributions, each defined by this parameter p.

Indicator r.v. and Bernoulli Trials

- The ***indicator random variable*** of an event A is the r.v. which equals 1 if A occurs and 0 otherwise. We will denote the indicator r.v. of A by I_A or $I(A)$. Note that $I_A \sim \text{Bern}(p)$ with $p = P(A)$.
- An experiment that can result in either a “success“ or a “failure" (but not both) is called a ***Bernoulli trial***. A Bernoulli random variable can be thought of as the indicator of success in a Bernoulli trial: it equals 1 if success occurs and 0 if failure occurs in the trial.

Story of the Binomial Distribution

- Imagine you are going to flip a coin n times, with known probability p of showing heads each times. Assume the result of each flip is independent from one another.
- Let X = # heads.
- What is X's distribution? That is, what is the probability mass function for X? (don't forget to mention X's support)
- Imagine instead you are repeatedly performing an experiment n times, where the probability of success, p, for each trial is fixed and each trial is independent from one another.
- Let X = # successes.
- Imagine instead you are sampling n individuals from a huge population (so large that it can be assumed to be infinite), and are measuring each individual for a specific characteristic.

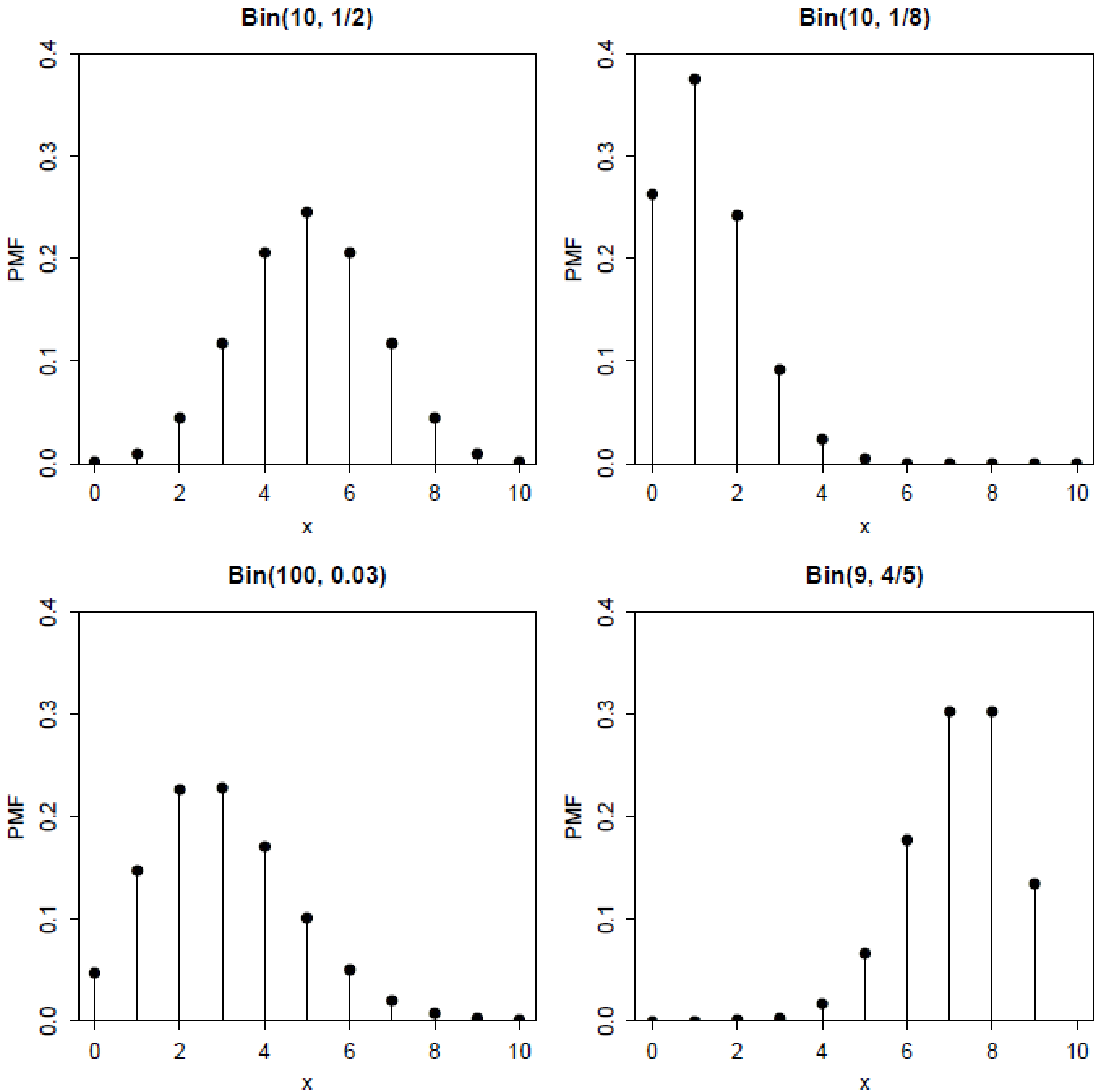
Binomial Distribution Definition

- Suppose that n independent Bernoulli trials are performed, each with the same success probability p. Let X be the number of successes. The distribution of X is called the ***Binomial distribution*** with parameters n and p. We write $X \sim \text{Bin}(n, p)$ to mean that X has the Binomial distribution with parameters n and p, where n is a positive integer and $0 < p < 1$.
- If $X \sim \text{Bin}(n, p)$, then the PMF of X is:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $k = 0, 1, \ldots, n$.

Plot of Binomial PMFs



The Binomial Distribution: Example

- If a couple are both carriers of a certain disease their child has probability 0.25 of being born with the disease. Suppose that a couple has 5 children:
  - What is the probability that none of their children have the disease?
  - What is the probability that at least two children have the disease?
- Let's use the formula to calculate the answers, and check online:
- Binomial Calculator online: http://www.stat.tamu.edu/~west/applets/binomialdemo.html

Binomial Distribution: More Examples

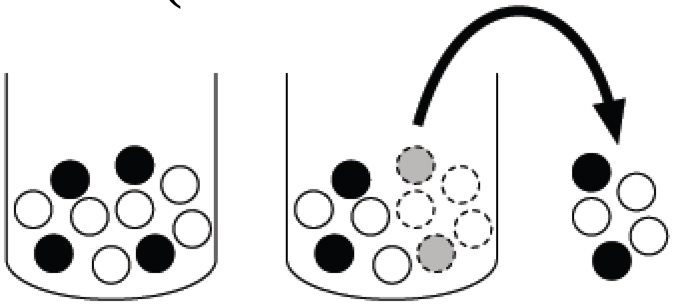
- Let $X \sim \text{Bin}(n = 20, p = 0.3)$.
  - What is $P(X = 10)$?
  - What is $P(X \geq 10)$?
  - What is $P(X > 10)$?
  - What is $P(X < 20)$?
- 20% of Harvard upperclassmen are varsity athletes. A random sample of $n = 64$ students were selected from the Quad houses, and only 8 of them were varsity athletes. [If Harvard housing was assigned randomly], what is the probability of selecting 8 or fewer athletes in this sample of 64 students?
- When is a Binomial distribution symmetric?
- Let $Y = n - X$. What distribution does Y follow?

Unit 3 Outline

- Random Variables (RVs) and Distributions
- Discrete RVs and Probability Mass Functions (PMFs)
- Bernoulli and Binomial Distributions
- Hypergeometric Distribution
- Discrete Uniform Distribution
- Cumulative Distribution Functions (CDFs)
- Functions of Random Variables
- Independence of Random Variables

Story of the Hypergeometric Distribution

- Suppose we have an urn filled with w white and b black balls. Then drawing n balls out of the urn *with replacement* yields a Bin(n, $p = w/(w+b)$) distribution for the number of white balls obtained in n trials, since the draws are independent Bernoulli trials, each with probability $p = w/(w+b)$ of success.
- If we instead sample *without replacement*, as illustrated below, then X = the number of white balls follows a ***Hypergeometric distribution***.
- What is the PMF for X? (be sure to mention the support)



Hypergeometric Distribution Definition

- Consider an urn filled with w white and b black balls. We draw n balls out of the urn at random without replacement, such that all $\binom{w+b}{n}$ samples are equally likely. Let X be the number of white balls in the sample. Then X is said to have the ***Hypergeometric distribution*** with parameters w, b, and n; we denote this by $X \sim \text{HGeom}(w, b, n)$.
- If $X \sim \text{HGeom}(w, b, n)$, then the PMF of X is:

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

for integers k satisfying $0 \leq k \leq w$ and $0 \leq n - k \leq b$.

Hypergeometric Distribution Examples

- In a five-card hand drawn at random from a well-shuffled standard deck, the number of aces in the hand has a HGeom(4, 48, 5) distribution, which can be seen by thinking of the aces as white balls and the non-aces as black balls.
  - What is the probability that the hand has exactly 3 aces?
- A forest has N deer. Today, m of the deer are captured, tagged, and released into the wild. At a later date, n deer will be recaptured at random. Assume that the recaptured deer are equally likely to be any set of n of the deer, e.g., a deer that has been captured does not learn how to avoid being captured again. Let X = # deer in the n recaptured deer that were captured the first time.
  - What distribution does X have?

Hypergeometric vs. Binomial

- The Binomial and Hypergeometric distributions are often confused. Both are discrete distributions taking on integer values between 0 and n for some n, and both can be interpreted as the number of successes in n Bernoulli trials (for the Hypergeometric, each tagged deer in the recaptured sample can be considered a success and each untagged deer a failure).
- However, a crucial part of the Binomial story is that the Bernoulli trials involved are independent. The Bernoulli trials in the Hypergeometric story are dependent, since the sampling is done without replacement: knowing that one deer in our sample is tagged decreases the probability that the second deer will also be tagged.

The HGeome is connected to the Binome

- The Binomial and Hypergeometric distributions are related in a couple very specific ways:

- If $X \sim \text{Bin}(n,p)$ and $Y \sim \text{Bin}(m,p)$ and X, Y are independent, then the conditional distribution of X given $X+Y = r$ is $\text{Hgeom}(n,m,r)$.

- If $X \sim \text{Hgeom}(w,b,n)$ and $N = w + b \to \infty$ such that $w/(w+b) = p$ is fixed, then $X \to \text{Bin}(n,p)$ [X is said to *converge in distribution* to a Binomial]

Unit 3 Outline

- Random Variables (RVs) and Distributions
- Discrete RVs and Probability Mass Functions (PMFs)
- Bernoulli and Binomial Distributions
- Hypergeometric Distribution
- Discrete Uniform Distribution
- Cumulative Distribution Functions (CDFs)
- Functions of Random Variables
- Independence of Random Variables

Discrete Uniform Distribution

- Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X. Then X is said to have the Discrete Uniform distribution with parameter C; we will denote this as $X \sim \text{DUnif}(C)$.
- The PMF of $X \sim \text{DUnif}(C)$ is:

$$P(X = x) = \frac{1}{|C|}$$

for $x \in C$ (and 0 otherwise).

- Common example: random integer from a to b.

Discrete Uniform Dist. Example

- There are 100 slips of paper in a hat, each of which has one of the numbers 1,2,…,100 written on it, with no number appearing more than once. Five of the slips are drawn, one at a time.
- Consider random sampling with replacement (with equal prob's):

a) What is the distribution of how many of the drawn slips have a value of at least 80 written on them?

b) What is the distribution of the value of the j^{th} draw (for $1 \le j \le 5$)?

c) What is the probability that the number 100 is drawn at least once?

- Now consider random sampling without replacement (with all sets of five slips equally likely to be chosen):

d) What is the distribution of how many of the drawn slips have a value of at least 80 written on them?

e) What is the distribution of the value of the jth draw (for $1 \le j \le 5$)?

f) What is the probability that the number 100 is drawn at least once?

Unit 3 Outline

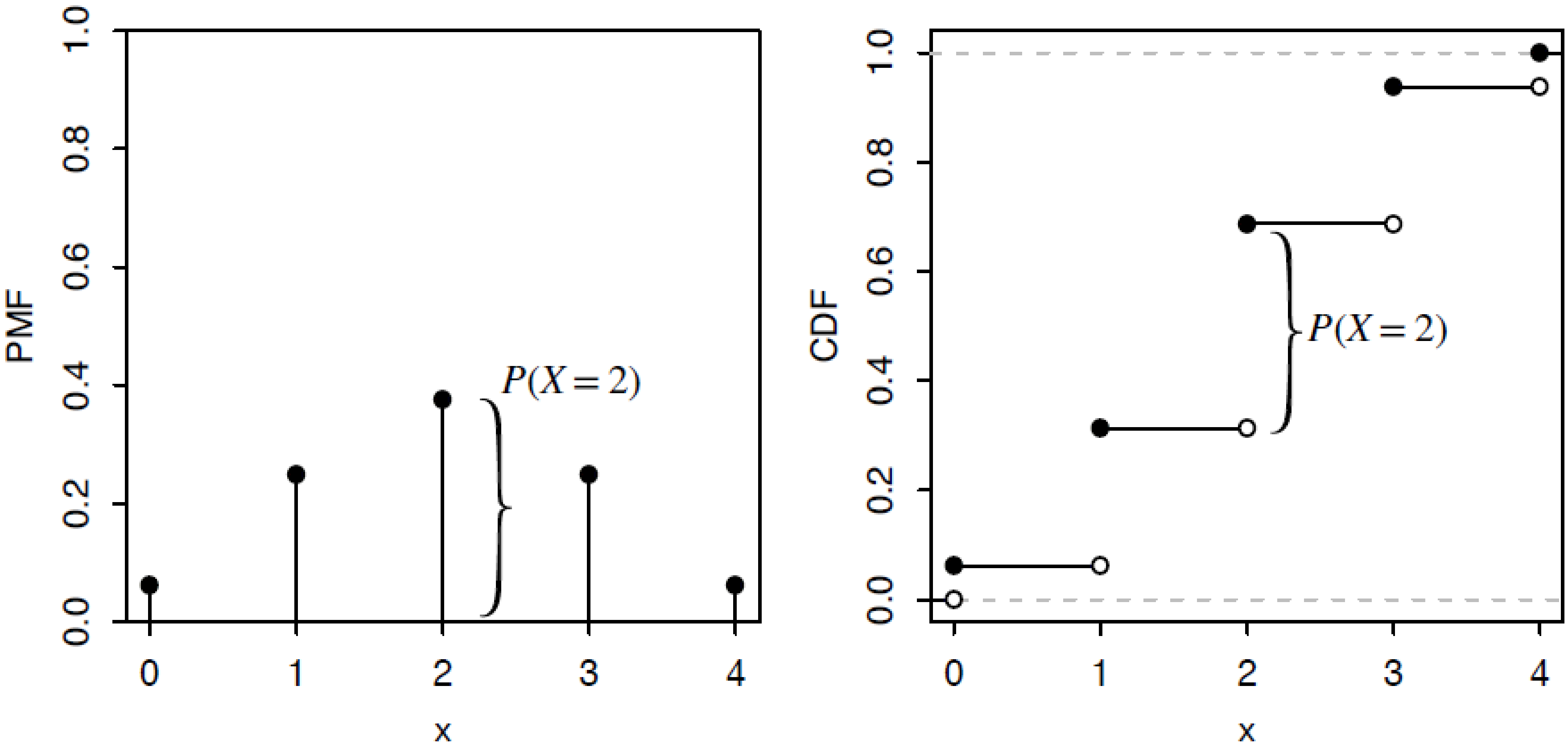
- Random Variables (RVs) and Distributions
- Discrete RVs and Probability Mass Functions (PMFs)
- Bernoulli and Binomial Distributions
- Hypergeometric Distribution
- Discrete Uniform Distribution
- Cumulative Distribution Functions (CDFs)
- Functions of Random Variables
- Independence of Random Variables

Cumulative Distribution Function (CDF)

- The ***cumulative distribution function*** of a r.v. X is the function F_X given by $F_X(x) = P(X \leq x)$. It is often written as just capital F without the subscript, or $F(x)$. (other letters, like G or H, can also be used). (DeGroot p.108)
- Properties of CDFs:
  - CDFs are always non-decreasing
  - $\lim_{x \to -\infty}[F(x)] =$ ____. $\lim_{x \to \infty}[F(x)] =$ ____.
  - CDFs are always continuous from the (right or left?)
- Example: Let $X \sim \text{Bin}(n = 4, p = 0.5)$.
  - Find $F(1.5)$.
  - Find $F(x)$ for all x.

Plot of a CDF

- Let $X \sim \text{Bin}(n = 4, p = 0.5)$.



PMFs vs. CDFs

- We have seen three equivalent ways of expressing the distribution of a random variable. Two of these are the PMF and the CDF: these two functions contain the same information, and we can always figure out the CDF from the PMF and vice versa. Generally the PMF is easier to work with for discrete r.v.s (the CDF requires a summation).
- A third way to describe a distribution is with a story that explains (in a precise way) how the distribution can arise. We used the stories of the Binomial and Hypergeometric distributions to derive the corresponding PMFs. Thus the story and the PMF also contain the same information. We can often achieve more intuitive proofs with the story than with PMF calculations, algebra, etc...

Unit 3 Outline

- Random Variables (RVs) and Distributions
- Discrete RVs and Probability Mass Functions (PMFs)
- Bernoulli and Binomial Distributions
- Hypergeometric Distribution
- Discrete Uniform Distribution
- Cumulative Distribution Functions (CDFs)
- Functions of Random Variables
- Independence of Random Variables

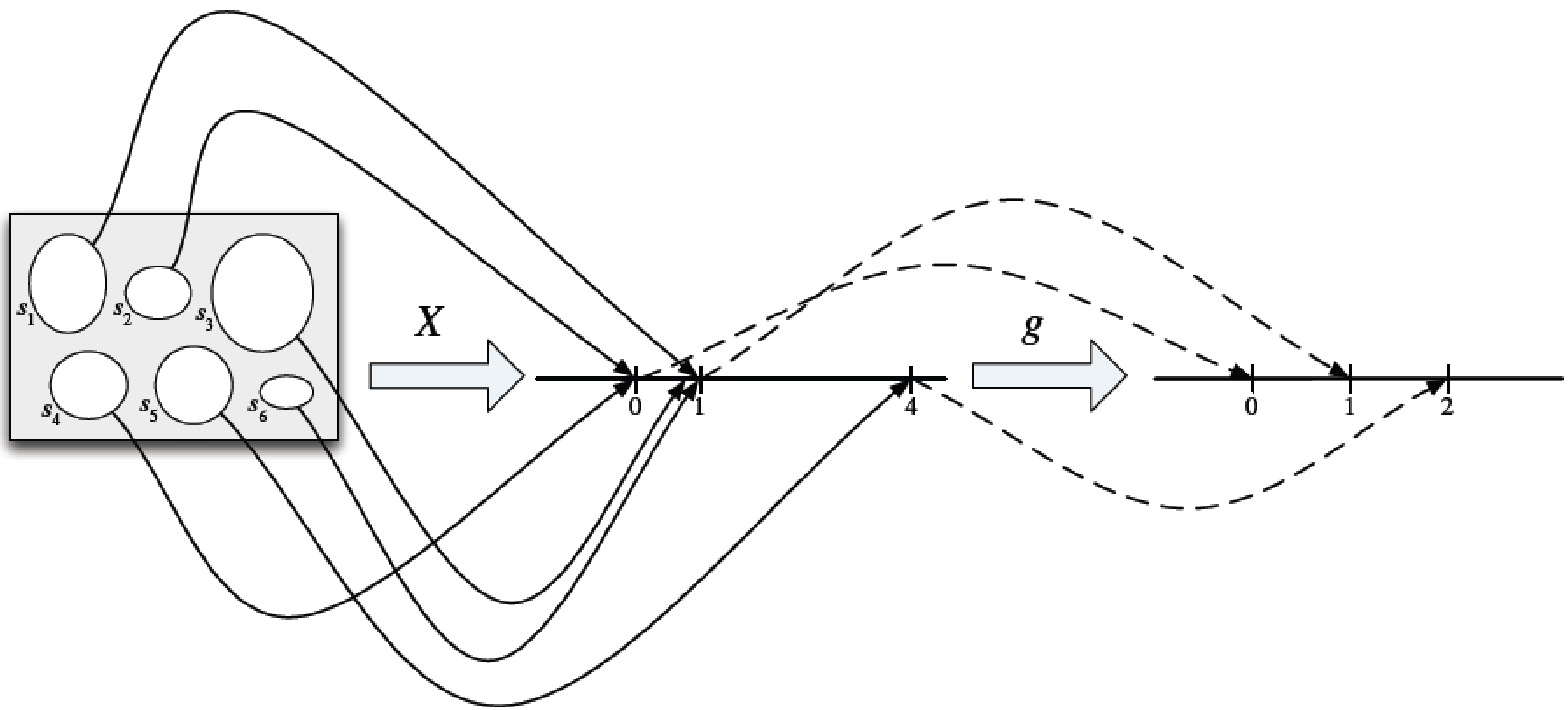
Function of a R.V.

- For an experiment with sample space S, a r.v. X, and a function g: $\mathbb{R} \rightarrow \mathbb{R}$, g(X) is the r.v. that maps s to $g(X(s))$ for all $s \in S$.
- What the heck does that mean? Let's look at an example:
- Let $g(x) = \sqrt{x}$ for concreteness. $g(X)$ is the composition of the functions X and g, saying "first apply X, then apply g".
- So if X crystallizes to 4, then $g(X)$ crystallizes to 2.
- Concrete Example: let X have the following distribution:

x	-2	-1	0	1	2
$P(X = x)$	0.1	0.2	0.3	0.2	0.2

- Find the distribution of $Y = X^2$.

Illustration of a function of a r.v.



Function of a R.V. Example (a Random Walk)

- A particle moves n steps on a number line. The particle starts at 0, and at each step it moves 1 unit to the right or to the left, with equal probabilities. Assume all steps are independent. Let Y be the particle's position after n steps. Find the PMF of Y.
- Hint: how does each step relate to a Bernoulli trial?
- Let $X \sim \text{Bin}(n, p = 1/2)$. Then $Y = 2X - n \rightarrow X = (Y + n)/2$.
- Thus:

$$P(Y = k) = P(2X - n = k) = P(X = (k+n)/2) = \binom{n}{(k+n)/2}\left(\frac{1}{2}\right)^n$$

- For $k = -n, -n+2, -n+4 \ldots, n$.
- Sanity check: does that answer make sense?

PMF of a function of a discrete R.V.

- Let X be a discrete r.v. and g: $\mathbb{R} \rightarrow \mathbb{R}$. Then the support of $g(X)$ is the set of all y such that $g(x) = y$ for at least one x in the support of X, and the PMF of $g(X)$ is:

$$P(g(X) = y) = \sum_{x:g(x)=y} P(X = x)$$

- Concrete Example: continuing on the random walk example, let D be the particle's distance from the origin after n steps. Assume that n is even. Find the PMF of D.
- $D = |Y|$. Thus $P(D = 0) = ???$. For $k = 2, 4, \ldots, n$:

$$P(D = k) = P(Y = k) + P(Y = -k) = 2\binom{n}{(k+n)/2}\left(\frac{1}{2}\right)^n$$

Function of two R.V.s

- For an experiment with sample space S, if X and Y are r.v.s that map $s \in S$ to $X(s)$ and $Y(s)$ respectively, then $g(X,Y)$ is the r.v. that maps s to $g(X(s),Y(s))$.
- What the heck does that mean? Let's look at an example:
- Suppose we roll two fair 6-sided dice, with X the number on the first die and Y the number on the second.
- Let $W = \max(X,Y)$. What is the distribution of W?
- Hint: it's usually easiest to write out the table of s, X, Y, and $W = \max(X,Y)$ for every value of s in S (or at least start to write out the table).

Be careful!

- A common error in probability is to confuse a random variable with its distribution. Two completely different random variables could very well have the same distribution.
- Other common errors:
  - Given a r.v. X, trying to find the PMF of $2X$ by multiplying the PMF of X by 2. Think about what this means for the plot of X and it's distribution.
  - Claiming that because X and Y have the same distribution, X must always equal Y (i.e. $P(X = Y) = 1$). Consider flipping a fair coin once. Let X = indicator it lands heads while $Y = 1–X$ indicator it lands tails. Both X and $Y \sim \text{Bern}(0.5)$, but $X = Y$ is impossible. The PMFs of X and Y are the same function, but X and Y are different mappings from $S \to \mathbb{R}$.
- Always ask yourself: does the result make sense???

Unit 3 Outline

- Random Variables (RVs) and Distributions
- Discrete RVs and Probability Mass Functions (PMFs)
- Bernoulli and Binomial Distributions
- Hypergeometric Distribution
- Discrete Uniform Distribution
- Cumulative Distribution Functions (CDFs)
- Functions of Random Variables
- Independence of Random Variables

Independence of two R.V.s

- Random variables X and Y, are said to be ***independent*** if:

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

for all $x, y \in \mathbb{R}$.

- In the discrete case, this is equivalent to:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all x in the support of X and y in the support of Y.

Independence of many R.V.s

- Random variables $X_1,\ldots,X_n$ are said to be ***independent*** if:

$$P(X_1 \le x_1, \ldots, X_n \le x_n) = P(X_1 \le x_1) \ldots P(X_n \le x_n)$$

for all $x_1,\ldots, x_n \in \mathbb{R}$. For infinitely many r.v.s, we say that they are independent if every finite subset of the r.v.s is independent.

- How does this compare to the independence of events? Why do we only need "one" condition here?

Independence of R.V.s example

- Suppose you roll two dice, with X the number on the first dice and Y the number on the second dice.
- Let $U = X + Y$ and $V = X - Y$.

  - What is the distribution of U?
  - What is the distribution of V?
  - What is $P(U = 12, V = 1)$?
  - Are U and V independent? How do you know?

- What does this mean in layman's terms?

Conditional Distributions and PMFs

- Just like with events, we can equivalently show that two r.v.s are independent based on conditional probability.
- We need to define the conditional distribution:
- For any discrete r.v.s X and Z, the function $\mathrm{P}(X = x | Z = z)$, when considered as a function of x for a fixed z, is called the ***conditional PMF of X given Z = z.***
- Thus, two discrete r.v.s, X and Z, are independent if $P(X = x \mid Z = z) = P(X = x)$ for all possible values of x and z.
- We'll come back to this idea of conditional distributions later.

Independent and Identically Distributed R.V.s (i.i.d.)

- We will often work with random variables that are independent and all came from the same distribution. These r.v.s are called ***independent and identically distributed***, or ***i.i.d.*** for short.
- What is the difference between *independent* and *identically distributed*?
- Give examples of:
  - *independent* and *identically distributed* r.v.s
  - *independent* and **not** *identically distributed* r.v.s
  - ***dependent*** and *identically distributed* r.v.s
  - ***dependent*** and **not** *identically distributed* r.v.s

Binomial is a sum of i.i.d. Bernoullis

- Let $X \sim \text{Bin}(n, p)$. What is it's story?
- We can write $X = X_1 + \ldots + X_n$ where the X_i i.i.d. $\text{Bern}(p)$.

- Further, let $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$ and X is independent of Y.
- What distribution does $T = X + Y$ have?
  - $T \sim \text{Bin}(n + m, p)$

- Why does this make sense?

Last Word: what is random?