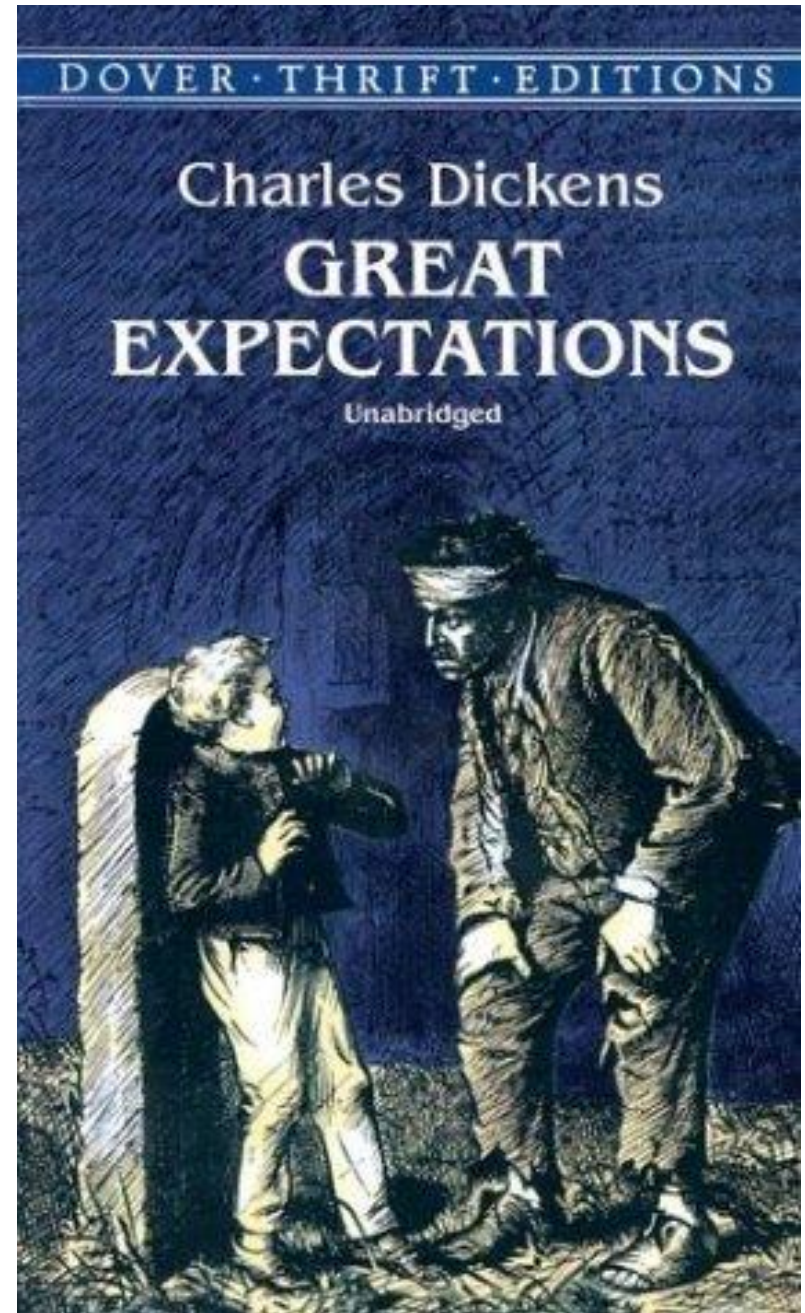# Stat 110

Unit 4: Expectation
## Chapter 4 in the text

Unit 4 Outline

- Definition of Expectation

- Linearity and Monotonicity of Expectation

- LOTUS (expectation of a function)

- Variance and Standard Deviation

- Geometric and Negative Binomial distributions

- Indicator r.v.s and the Fundamental Bridge

- Poisson distribution

Summarizing a r.v.'s Distribution

- Suppose a discrete random variable has the following distribution:

- How would you summarize this distribution?
  - Center?
  - Spread?
  - Shape?
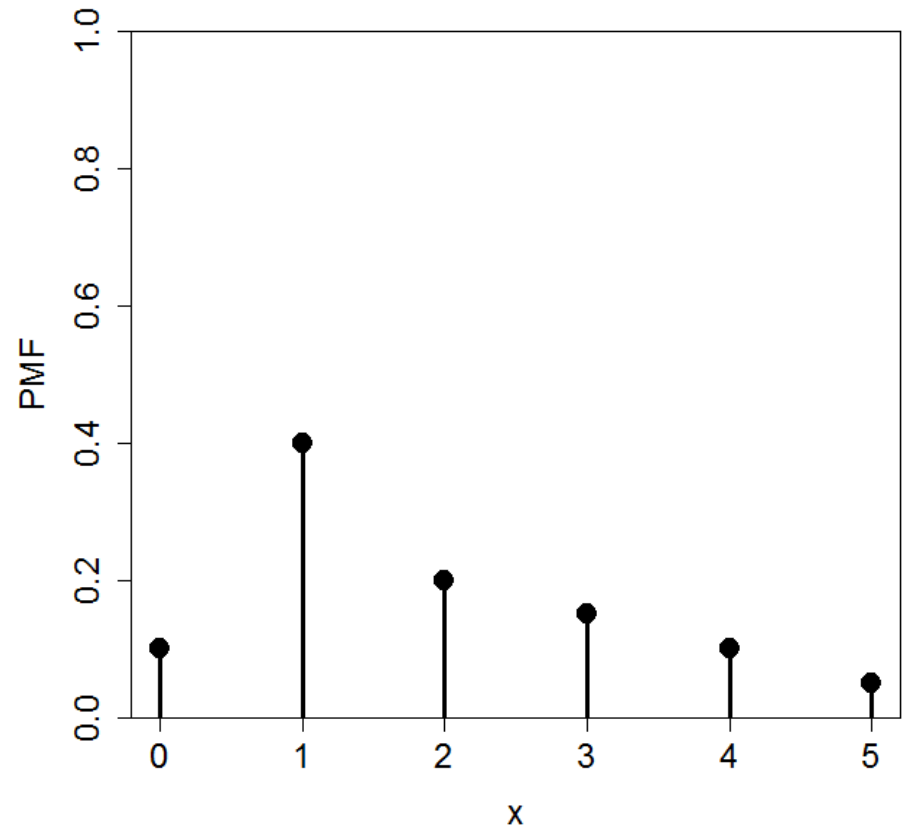
- The concept of *expectation* of a r.v. formalizes these ideas.

Definition: Expectation

- The *expected value* (or *expectation* or *mean*), *E(X)*, of a discrete random variable *X* is defined as (text, p.138-139):

$$E(X) = \sum_{all\ x} xP(X = x)$$

- Sometimes the mean of *X* is written as μ_X.

- Intuitively, what is *E(X)* measuring?

  - It is the theoretical weighted average of *X*, weighted by the probabilities (hence the name, *mean*).

  - It is the measure of the "center" of a distribution, but says nothing about the spread or shape of it.

  - Physically, it is the *balance point* of the PMF.

Expected Value Examples

- Concrete Example: let *X* have the following distribution:

x	0	1	2
$P(X = x)$	0.5	0.4	0.1

- Intuitively, what should be the mean of *X*? Calculate it.

$$E(X) = \sum_{all\ x} xP(X = x) = 0(0.5) + 1(0.4) + 2(0.1) = 0.6$$

- Let $X \sim \text{Bern}(p)$. Intuitively, what should be the mean of *X*? Calculate it.

$$E(X) = \sum_{all\ x} xP(X = x) = 0(1 - p) + 1(p) = p$$

5

Non-uniqueness of Expectation

- Based on it's definition, it can be shown that $E(X)$ only depends on the distribution of X. So two r.v.s with the same distribution (same PMF) will have the same expected value.

- The converse is not true: two r.v.s could have the same expected value, but have completely different distribution.

- Examples: $X \sim \text{Bin}(n=2, p=0.5)$ and $Y = \text{DUnif}(\{0,1,2\})$.

Unit 4 Outline

- Definition of Expectation

- **Linearity and Monotonicity of Expectation**

- LOTUS (expectation of a function)

- Variance and Standard Deviation

- Geometric and Negative Binomial distributions

- Indicator r.v.s and the Fundamental Bridge

- Poisson distribution

Linearity of Expectation

- For any r.v.s X, Y and any constant c, linearity of expectation holds (text p. 140):

$$E(X + Y) = E(X) + E(Y)$$
$$E(aX + b) = aE(X) + b$$

- What do these results mean?

- The second equation says that we can take out constant factors from an expectation; this is both intuitively reasonable and easily verified from the definition.

- The first equation, $E(X + Y) = E(X) + E(Y)$, also seems reasonable when X and Y are independent. What may be surprising is that it holds even if X and Y are dependent!

8

Why $E(X + Y) = E(X) + E(Y)$

- We can equivalently calculate expectations based on the sum of all outcomes in *S* directly. Thus:

$$E(X) = \sum_{all\ s} X(s)P(\{s\})$$

- What the heck does $X(s)$ mean? How about $P(\{s\})$?

- So if *X* and *Y* are based on the outcomes (key: they have to be if they are dependent), then:

$$E(X) + E(Y) = \sum_{all\ s} X(s)P(\{s\}) + \sum_{all\ s} Y(s)P(\{s\})$$

$$= \sum_{all\ s} (X+Y)(s)P(\{s\}) = E(X+Y)$$

- An example is worth a thousand words. ..

Linearity of Expectation is handy!

- Let $X \sim \text{Bin}(n,p)$. Intuitively, what should be the mean of X? Calculate it.
- Two ways to do this:
  1) Brute force:

  2) Applying linearity (much easier):

Linearity of Expectation is handy!

- Let $X \sim \text{HGeom}(w,b,n)$. Intuitively, what should be the mean of X? Calculate it.
  *Hint: define X as a sum of dependent Bernoulli r.v.s.

Monotonicity of Expectation

- Let X and Y be r.v.s such that $X \geq Y$ with probability 1. Then $E(X) \geq E(Y)$, with equality holding if and only if $X = Y$ with probability 1.

- Proof:

- We will prove it only for discrete r.v.s. (but it holds for all r.v.s). The r.v. $Z = X - Y$ is nonnegative (with probability 1), so $E(Z) \geq 0$ since $E(Z)$ is defined as a sum of nonnegative terms. By linearity, $E(X) - E(Y) = E(X - Y) \geq 0$; as desired.

- If $E(X) = E(Y)$, then by linearity we also have $E(Z) = 0$, which implies that $P(X = Y) = P(Z = 0) = 1$ since if even one term in the sum defining $E(Z)$ is positive, then the whole sum is positive.

- Monotonicity is not as useful as linearity.

Unit 4 Outline

- Definition of Expectation

- Linearity and Monotonicity of Expectation

- LOTUS (expectation of a function)

- Variance and Standard Deviation

- Geometric and Negative Binomial distributions

- Indicator r.v.s and the Fundamental Bridge

- Poisson distribution

13

Law of the Unconscious Statistician (LOTUS)

- If *X* is a discrete r.v. and *g* is a function from $\mathbb{R}$ to $\mathbb{R}$, then (text, *p*.156):

$$E\big[g(X)\big] = \sum_{\text{all } x} g(x) P(X = x)$$

- What does this mean?
  This means that we can get the expected value of *g*(*X*) knowing only *P*(*X* = *x*), the PMF of *X*; we don't need to know the PMF of *g*(*X*). The name comes from the fact that in going from E(*X*) to E[*g*(*X*)] it is tempting just to change *x* to *g*(*x*) in the definition, which can be done very easily and mechanically, perhaps in a state of unconsciousness.
- Be careful: E[*g*(*X*)] does not necessarily equal *g*[E(*X*)] !!!

Example to illustrate LOTUS

- 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

- Let $Y = X^2$. Find E(Y) two ways: based on the distribution of Y and using LOTUS.

- Why does it work out to be the same answer?

15

Unit 4 Outline

- Definition of Expectation

- Linearity and Monotonicity of Expectation

- LOTUS (expectation of a function)

- **Variance and Standard Deviation**

- Geometric and Negative Binomial distributions

- Indicator r.v.s and the Fundamental Bridge

- Poisson distribution

16

Variance and Standard Deviation

- The *variance* of a r.v. *X* is defined as (text, *p*.158):

$$Var(X) = E\left[(X - \mu)^2\right]$$

where $\mu = E(X)$. Sometimes variance is written as σ^2_X.

- The square root of the variance is called *standard deviation:*

$$SD(X) = \sqrt{Var(X)}$$

- What is variance measuring? What are its units?
- What is standard deviation measuring? What are its units?
- Which is more interpretable?

Variance: an equivalent formula

- For any r.v. X:

$$Var(X) = E\left(X^2\right) - \mu^2$$

  where $\mu = E(X)$.

- Proof (Expand $(X - \mu)^2$ and use linearity):

$$Var(X) = E\left[(X - \mu)^2\right] = E\left(X^2 - 2\mu X + \mu^2\right)$$
$$= E\left(X^2\right) - 2\mu E(X) + \mu^2 = E\left(X^2\right) - 2\mu^2 + \mu^2$$
$$= E\left(X^2\right) - \mu^2$$

- This result is often useful when calculating variance of a r.v.

Variance Examples

- Concrete Example: let *X* have the following distribution:

x	0	1	2
$P(X = x)$	0.5	0.4	0.1

- What is the variance of *X*? What is its standard deviation?

$$Var(X) = E\left[(X - \mu)^2\right] = \sum_{all\ x}(x - \mu)^2 P(X = x)$$

$$= (0 - 0.6)^2(0.5) + (1 - 0.6)^2(0.4) + (2 - 0.6)^2(0.1) = 0.44$$

$$SD(X) = \sqrt{Var(X)} = \sqrt{0.44} = 0.663$$

- Let *X* ~ Bern(*p*). Calculate Var(*X*).

$$E(X^2) = \sum_{all\ x} x^2 P(X = x) = 0^2(1 - p) + 1^2(p) = p$$

$$Var(X) = E(X^2) - \mu^2 = p^2 - p = p(1 - p)$$

19

Properties of Variance

- $\text{Var}(X + c) = \text{Var}(X)$

- $\text{Var}(cX) = c^2\text{Var}(X)$

- $\text{Var}(X) \geq 0$

- If X and Y are independent, then:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

- Why do these properties make sense, intuitively?

- The 4th property: we will prove later (need to define covariance first). But intuitively, what should $\text{Var}(X+X)$ be?

- When is $\text{Var}(X) = 0$?

Variance of a Binomial r.v.

- Let $X \sim \text{Bin}(n,p)$. What is $\text{Var}(X)$?
- Two ways to do this:
  1) Brute force:

  2) Using properties of variance (much easier):
  Let $X_i \sim \text{Bern}(p)$.

- How should variance of a Hypergeometric r.v. compare?

Unit 4 Outline

- Definition of Expectation

- Linearity and Monotonicity of Expectation

- LOTUS (expectation of a function)

- Variance and Standard Deviation

- **Geometric and Negative Binomial distributions**

- Indicator r.v.s and the Fundamental Bridge

- Poisson distribution

Story of the Geometric Distribution

- Consider a sequence of independent Bernoulli trials, each with the same success probability p, with trial performed until a success occurs. Let X be the number of failures before the first successful trial. Then X has the Geometric distribution with parameter p; we denote this by $X \sim \text{Geom}(p)$.

- What is X's distribution? That is, what is the probability mass function for X? (don't forget to mention X's support)

Geometric Distribution Definition

- If $X \sim \text{Geom}(p)$, then the PMF of X is:

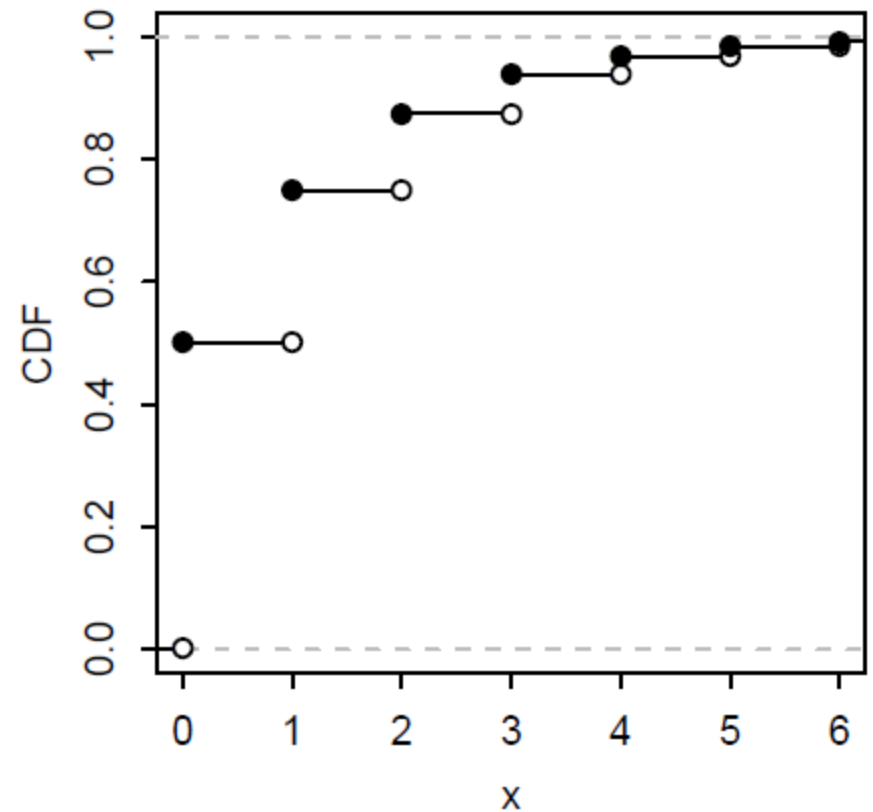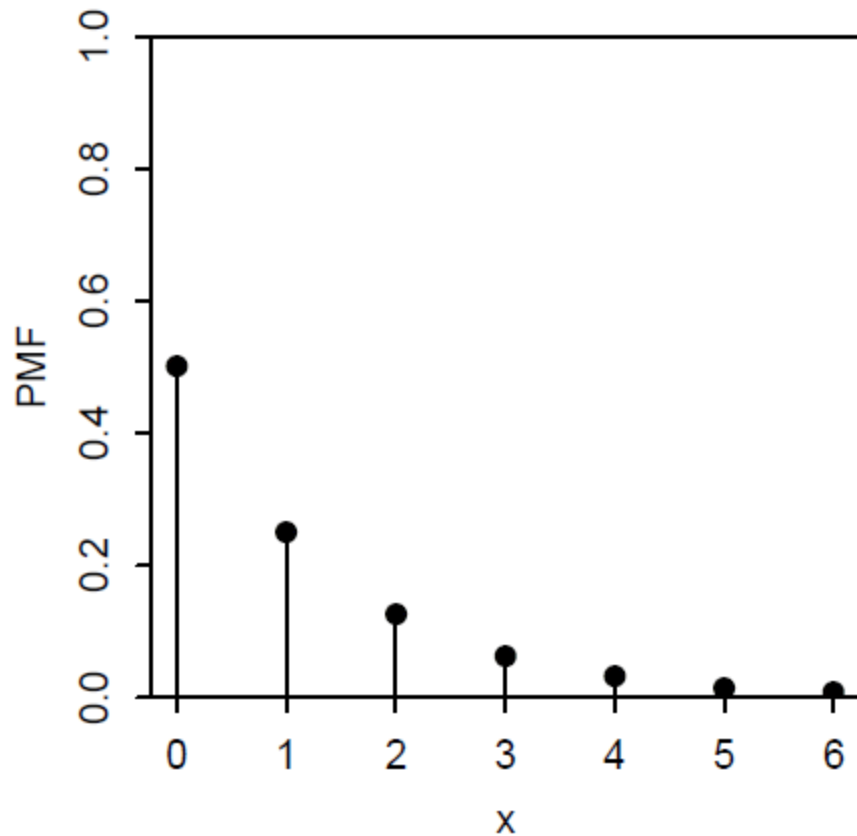$$P(X = k) = (1-p)^k \, p = q^k \, p$$

for $k = 0, 1, \ldots$

- Why is this a valid PMF?
- Key (recall your geometric series):

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

- Note: the First Success distribution, FS(p), counts the 1st success and adds one to the Geom(p) r.v.

24

Plot of a Geometric PMF and CDF
[Geom(0.5)]

The Geometric Distribution: Mean and Variance

- Let $X \sim \text{Geom}(p)$. Find $E(X)$ and $\text{Var}(X)$.

  *Hint: take the derivative of the geometric series w.r.t. q.

$$d\left(\sum_{k=0}^{\infty} q^k\right) / dq = d\left(\frac{1}{1-q}\right) / dq \Rightarrow \sum_{k=0}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$$

$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} kq^k p = p\sum_{k=0}^{\infty} kq^k$$

$$= \frac{pq}{p^2} = \frac{q}{p}$$

The Geometric Distribution: Mean and Variance

- $\text{Var}(X) = \text{E}(X^2) - \mu^2$. *Hint: take the second derivative of q times the geometric series w.r.t. q.

$$\sum_{k=0}^{\infty} kq^k = \frac{q}{(1-q)^2} \Rightarrow \sum_{k=0}^{\infty} k^2 q^{k-1} = \frac{1+q}{(1-q)^3} = \frac{1+q}{p^3}$$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2 q^k p = p \sum_{k=0}^{\infty} k^2 q^k$$

$$= pq\left(\frac{1+q}{p^3}\right) = \frac{q(1+q)}{p^2}$$

$$Var(X) = E(X^2) - \mu^2 = \frac{q(1+q)}{p^2} - \left(\frac{q}{p}\right)^2 = \frac{q}{p^2}$$

27

Story of the Negative Binomial Dist.

- The Negative Binomial distribution is just an extension of the Geometric distribution: instead of waiting for just one success, we can wait for a known pre-determined number, r, of successes.
- So in a sequence of independent Bernoulli trials with success probability p, if X is the number of failures before the rth success, then X is said to have the ***Negative Binomial distribution*** with parameters r and p, denoted $X \sim \text{NBin}(r, p)$
- What is X's distribution? That is, what is the probability mass function for X? (don't forget to mention X's support)

Negative Binomial Dist. Definition

- If $X \sim \text{NBin}(r, p)$, then the PMF of X is:

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

for $n = 0, 1, 2, \ldots$

- How does the Negative Binomial distribution relate to the Geometric Distribution?

- Let i.i.d. $X_i \sim \text{Geom}(p)$, and let $X \sim X_1 + X_2 + \ldots + X_r$

- What distribution does X have?

- So a Negative Binomial r.v. is a sum of i.i.d. Geometric r.v.s

The Negative Binomial Distribution: Mean and Variance

- Let $X \sim \text{NBin}(r,p)$. Find $E(X)$ and $\text{Var}(X)$.

   *Hint: use the fact that Negative Binomial r.v. is a sum of i.i.d. Geometric r.v.s

How everything relates…

- Below is a table of 4 different types of r.v.s. based on sampling where each observation can be considered a result of a Bernoulli trial (not necessarily independent). Here are the 4 types we have seen:

	With replacement	Without replacement
Fixed number of trials	Binomial	Hypergeometric
Fixed number of successes	Negative Binomial	Negative Hypergeometric

- We won't talk about the *Negative Hypergeometric*, but realize there is such a thing (rarely comes into play in practice).

Unit 4 Outline

- Definition of Expectation

- Linearity and Monotonicity of Expectation

- LOTUS (expectation of a function)

- Variance and Standard Deviation

- Geometric and Negative Binomial distributions

- **Indicator r.v.s and the Fundamental Bridge**

- Poisson distribution

Properties of Indicator Variables

- Recall from Unit 3 that an indicator r.v., I_A or $I(A)$, for an event A takes on the value 1 if A occurs and 0 if does not occur.
- Indicator r.v.s have the following properties (for distinct events A and B):
  - $(I_A)^k = I_A$ for any positive integer k.
  - $I_{A^c} = 1 - I_A$
  - $I_{A \cap B} = I_A \cdot I_B$
  - $I_{A \cup B} = I_A + I_B - I_A \cdot I_B$

The Fundamental Bridge

- There is a one-to-one correspondence between events and indicator r.v.s, and the probability of an event A is the expected value of its indicator r.v. I_A:

$$P(A) = E(I_A)$$

- Proof: What distribution does I_A have? What is that distribution's expectation?

The Fundamental Bridge: Example #1

- In a group of *n* people, under the usual assumptions about birthdays, what is the expected number of distinct birthdays among the *n* people, i.e., the expected number of days on which at least one of the people was born?
- Hint: define a set of indicator variables: $I_1,\ldots,I_{365}$ and use the fundamental bridge and linearity.

- What is the expected number of birthday matches, i.e., pairs of people with the same birthday?
- Hint: define a set of indicator variables: $J_1,\ldots,J_{\binom{n}{k}}$ and use the fundamental bridge and linearity.

35

The Fundamental Bridge: Example #2

- Let X be a nonnegative integer-valued r.v. Let F be the CDF of X, and $G(x) = 1 - F(x) = P(X > x)$. The function G is called the *survival function* of X. Then

$$E(X) = \sum_{n=0}^{\infty} G(n)$$

- Let $X \sim \text{Geom}(p)$, and $q = 1 - p$. Using the Geometric story, $\{X > n\}$ is the event that the first $n + 1$ trials are all failures.

- So by expectation via the survival function:

$$E(X) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} q^{n+1} = \frac{q}{1-q} = \frac{q}{p}$$

36

Unit 4 Outline

- Definition of Expectation

- Linearity and Monotonicity of Expectation

- LOTUS (expectation of a function)

- Variance and Standard Deviation

- Geometric and Negative Binomial distributions

- Indicator r.v.s and the Fundamental Bridge

- Poisson distribution

Story of the Poisson Distribution

- Imagine you are trying to determine the number of occurrences ("successes") of a certain rare type of cancer (melanoma) in a large population (like the state of Massachusetts) over a fixed period of time (say a year).
- What would be the Binomial distribution used to model the situation here?
- The Poisson distribution is instead often used in situations like this, where we are counting the number of successes in a particular region or interval of time, and there are a large number of trials, each with a small probability of success.

Poisson Distribution Definition

- A r.v. X has the ***Poisson distribution*** with parameter λ, $X \sim \text{Pois}(\lambda)$, if the PMF of X is:
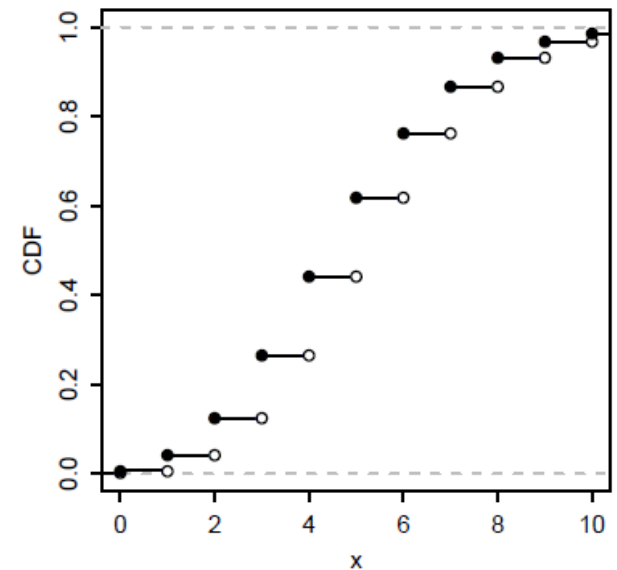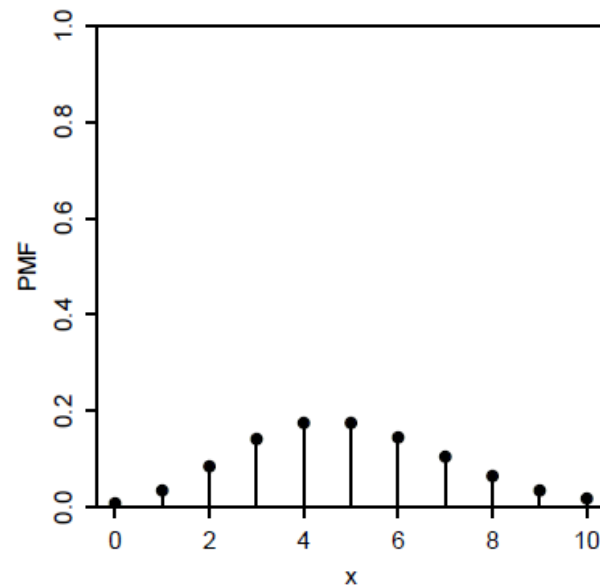
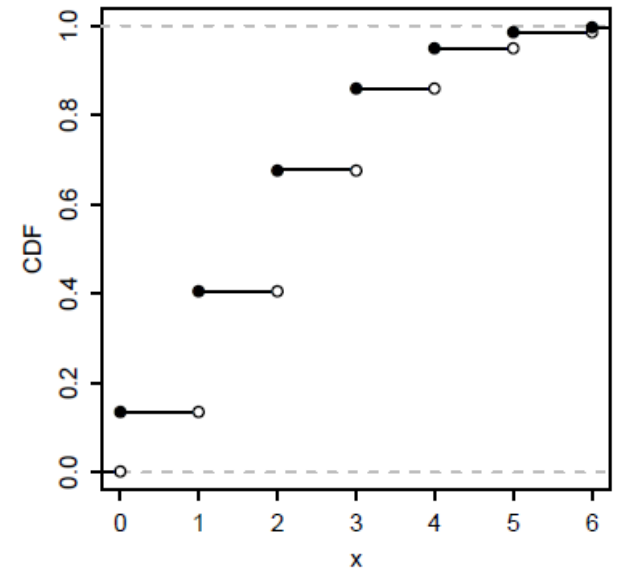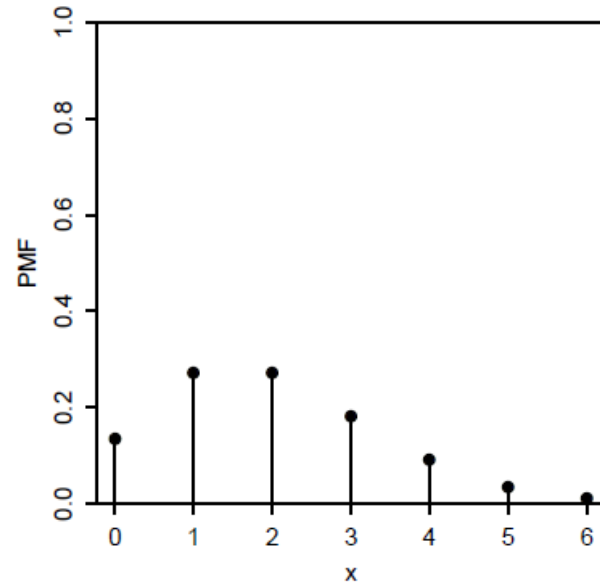$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

for $k = 0, 1, \ldots$

- This is a valid PMF because of the Taylor series:

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$$

Plot of Poisson PMFs and CDFs

[top is Pois(2) and bottom is Pois(5)]

40

The Poisson Distribution Examples

- The following random variables could reasonably follow a distribution that is approximately Poisson:
- The number of emails you receive in an hour. There are a lot of people who could potentially email you in that hour, but it is unlikely that any specific person will actually email you in that hour. Alternatively, imagine subdividing the hour into milliseconds. There are 3.6×10^6 milliseconds in an hour, but in any specific millisecond it is unlikely that you will get an email.
- The number of chips in a chocolate chip cookie. Imagine subdividing the cookie into small cubes; the probability of getting a chocolate chip in a single cube is small, but the number of cubes is large.
- The parameter λ is interpreted as the rate of occurrence of these rare events; in the examples above, could be 20 (emails per hour) or 10 (chips per cookie).

The Poisson Distribution: Mean and Variance

- Let $X \sim \text{Pois}(\lambda)$. Find $E(X)$ and $\text{Var}(X)$.

Poisson Distribution: Concrete Example

- The Boston Bruins score on average 3 goals per 60 minute game. Let X = # goals scored in the next Bruins game.

- Is it reasonable to assume X follows a Poisson distribution?

- What should be the value of λ?

- What is the probability that they get shut out in a game?

- Note: more on the Poisson in the notes for next week…

Last Word: what is a Poisson r.v.

poisson!

A pescetarian meal at The Wooden Table – March 12, 2011