San José State University

Math 161A: Applied Probability & Statistics

Special discrete distributions

Prof. Guangliang Chen

In this lecture, we introduce the following special discrete distributions from **Sections 3.4 – 3.6**:

- Bernoulli
- Binomial (and HyperGeometric)
- Geometric (and Negative Binomial)
- Poisson

Treatment plan for each distribution

- Examples
- Definition (including pmf)
- Expected value
- Variance
- Other useful properties (if any)

Bernoulli distributions

Consider the following experiments and associated random variables. What distributions do they have?

Example 0.1 (Toss a fair coin). $X = 1$ (heads) and 0 (tails).

Example 0.2 (Randomly select a ball from an urn that has 10 red and 20 green balls). Let $Y = 1$ (if the selected ball is red) and 0 (otherwise).

Example 0.3 (Randomly select an individual from a population 40% of which have certain characteristic). Let $Z = 1$ (if the selected individual has the characteristic) and 0 (otherwise).

These experiments all share the following traits:

- There is only **one trial**;
- It has only **two possible outcomes**, "success" or "failure";
- The **probability of having a success is some number** p;
- X **is a indicator variable** for the outcome success:
  $X = 1$ (success) or 0 (failure)

We say that such a random variable has a **Bernoulli distribution with parameter** p, and denote it as $X \sim \text{Bernoulli}(p)$.

Such an experiment is called a **Bernoulli trial**.

Example 0.4 (Toss a fair coin). $X = 1$ (heads) and 0 (tails).

Answer: $X \sim$ Bernoulli$(\frac{1}{2})$

Example 0.5 (Randomly select a ball from an urn that has 10 red and 20 green balls). Let $Y = 1$ (if the selected ball is red) and 0 (otherwise).

Answer: $Y \sim$ Bernoulli$(\frac{1}{3})$

Example 0.6 (Randomly select an individual from a population 40% of which have certain characteristic). Let $Z = 1$ (if the selected individual has the characteristic) and 0 (otherwise).

Answer: $Z \sim$ Bernoulli(0.4)

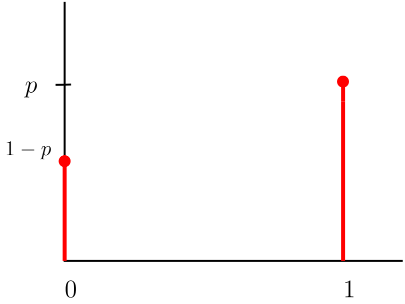
Properties of Bernoulli distributions

Clearly, if a discrete random variable X follows a Bernoulli distribution with parameter p, then its pmf has the following form (and vice versa):

$$f_X(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

(and $f_X(x) = 0$ for all other x)

x	0	1
$P(X = x)$	$1-p$	p



Theorem 0.1. Let $X \sim \text{Bernoulli}(p)$. Then

$$\mathrm{E}(X) = p$$
$$\mathrm{Var}(X) = p(1-p)$$

Proof. We have already obtained these results previously:

$$\mathrm{E}(X) = 0 \cdot (1-p) + 1 \cdot p = p$$
$$\mathrm{E}(X^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

and consequently,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = p - p^2 = p(1-p).$$

$\square$

Binomial

The following experiments (and random variables X) are essentially the same:

- (Toss a fair coin 10 times): $X = \#$heads
- (Answer 10 multiple-choiced questions by random guessing): $X= \#$correctly answered questions
- (Draw with replacement 10 balls at random from an urn containing 30 red balls and 20 blue balls): $X= \#$red balls obtained

We make the following abstraction about the experiments:

- The experiment consists of n **repeated trials**
- **Each trial has only two possible outcomes**: "success", "failure"
- The **probability** p **of getting successes is fixed** throughout the experiment
- The n Bernoulli trials are **independent**
- X counts the **total number of successes**

<u>Distribution of the random variable X:</u>

In short, X counts the total number of successes in n independent Bernoulli trials with fixed probability of success p.

$$\underset{1}{\underline{\qquad}}\quad \underset{2}{\underline{\qquad}}\quad \underline{\qquad}\quad \underline{\qquad}\quad \underline{\qquad}\quad \underline{\qquad}\quad \underset{n}{\underline{\qquad}}$$

In the above scenario, we say that X follows a **binomial distribution with parameters** n, p, and write $X \sim B(n, p)$

Example 0.7. Find the distribution of X in each case below:

- (Toss a fair coin 10 times) $X = \#$heads $B(10, \frac{1}{2})$
- (Answer 10 multiple-choiced questions by random guessing) $X=$ $\#$correctly answered questions $B(10, \frac{1}{4})$
- (Draw with replacement 10 balls from an urn containing 30 red balls and 20 blue balls at random) $X=$ $\#$red balls obtained. $B(10, 0.6)$

Theorem 0.2. The pmf of $X \sim B(n, p)$ is

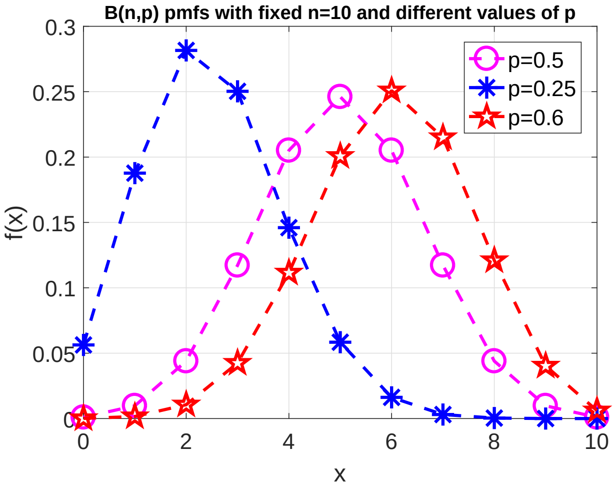
$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

	H	H			H	
___	___	___	___	___	___	___
1	2					n

How to understand this result:

- $\binom{n}{x}$: $\#$ ways of having x successes in n trials
- p^x: probability of having exactly x successes (in each pattern)
- $(1-p)^{n-x}$: probability of having exactly $n-x$ failures

Example 0.8 (Answer 10 multiple-choiced questions by random guessing)**.** Let $X=$ #correctly answered questions. Find $P(X = x)$ for $x = 0, 2, 9$.


B(n,p) pmfs with fixed n=10 and different values of p
0.3
0.25
0.2
0.15
0.1
0.05
0
f(x)
0
2
4
6
8
10
x
p=0.5
p=0.25
p=0.6

Theorem 0.3. Let $X \sim B(n, p)$. Then

$$\mathrm{E}(X) = np, \qquad \mathrm{Var}(X) = np(1-p)$$

Proof. Write

$$X = X_1 + X_2 + \cdots + X_n$$

$$\underset{1}{\underline{\hspace{3em}}}\quad \underset{2}{\underline{\hspace{3em}}}\quad \underline{\hspace{3em}}\quad \underline{\hspace{3em}}\quad \underline{\hspace{3em}}\quad \underline{\hspace{3em}}\quad \underset{n}{\underline{\hspace{3em}}}$$

where the X_i's are *identically and independently distributed* (iid) according to the Bernoulli(p) distribution:

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathrm{Bernoulli}(p)$$

It follows that

$$\mathrm{E}(X_i) = p, \quad \mathrm{Var}(X_i) = p(1-p), \quad \text{for all } i = 1, \ldots, n$$

By linearity and independence,

$$\begin{aligned} \mathrm{E}(X) &= \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = p + \cdots + p = np \\ \mathrm{Var}(X) &= \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = np(1-p). \end{aligned}$$

Hypergeometric

Recall the example in which we draw 10 balls at random from an urn containing 30 red and 20 blue balls, and let $X=$ #red balls obtained.

The experiment was performed with replacement, so we concluded that $X \sim B(n = 10, p = 0.6)$

If the experiment is performed **without replacement** instead, then X is not binomial (why?), but has a so-called **hypergeometric distribution**: $X \sim \text{HyperGeom}(N = 50, r = 30, n = 10)$.

In fact, in the latter case (without replacement), X has the following pmf:

$$f_X(x) = \frac{\binom{30}{x}\binom{20}{10-x}}{\binom{50}{10}}, \quad x = 0, 1, \ldots, 10$$

where

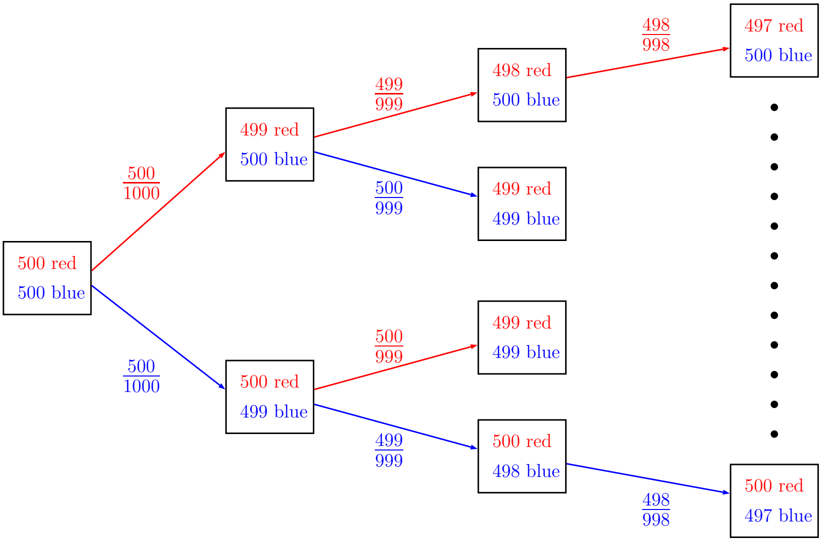
- $\binom{30}{x}$: #ways of choosing x red balls out of 30
- $\binom{20}{10-x}$: #ways of choosing $10 - x$ blue balls out of 20
- $\binom{50}{10}$: #ways of choosing 10 balls out of 50 in total (ignoring color)

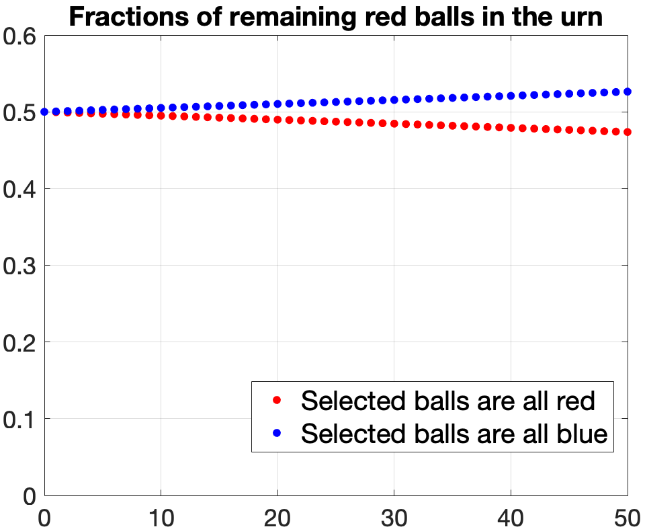
The hypergeometric pmf has a somewhat complicated formula, but in the special setting of large N and large r, it can be well approximated by the binomial pmf.

Theorem 0.4. When N, r are both large (relative to n), then

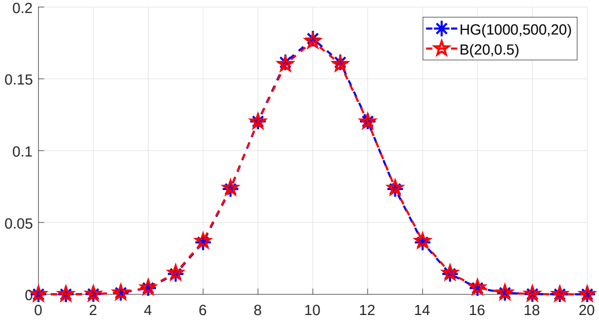
$$\mathrm{HyperGeom}(N, r, n) \approx B(n, p = \frac{r}{N}).$$

Remark. To understand why this approximation seems reasonable, assume an urn containing 500 red balls and 500 blue balls. We select a small number of balls from the urn without replacement, and monitor how the probability of selecting red balls changes gradually (see next slide).


500 red
500 blue
500/1000
500/1000
499 red
500 blue
500 red
499 blue
499/999
500/999
500/999
499/999
498 red
500 blue
499 red
499 blue
499 red
499 blue
500 red
498 blue
498/998
498/998
497 red
500 blue
500 red
497 blue


Fractions of remaining red balls in the urn
0.6
0.5
0.4
0.3
0.2
0.1
0
0
10
20
30
40
50
Selected balls are all red
Selected balls are all blue

Suppose we now draw 20 balls without replacement from an urn containing 500 red and 500 blue balls, and let $X =$ number of red balls in the selection. Below shows the pmf of the exact distribution of X (hypergeometric) and its binomial approximation.



The Hypergeometric distribution has important applications in polling, where it is often suitable to approximate it by the binomial distribution.

Example 0.9. Consider the experiment of drawing, without replacement, n voters at random from the whole pool of N that are registered, r of which support certain presidential candidate. Let $X =$ #supporters of the candidate in the selection. Then $X \sim$ HyperGeom(N, r, n). In reality, both r and N are typically large (e.g., in the order of millions) and n is only around a thousand. Accordingly, we have that $X \overset{\text{approx}}{\sim} B(n, p = \frac{r}{N})$.

Geometric

The following two experiments are identical in nature:

Example 0.10. Consider the experiment of repeatedly and independently flipping a coin until the first head appears. Let $X = \#$flips needed.

$$\underset{1}{\underline{\mathrm{T}}}\quad \underbrace{\underset{2}{\underline{\mathrm{T}}}\quad \underline{\mathrm{T}}\quad \cdots\quad \underline{\mathrm{T}}\quad \underline{\color{red}{\mathrm{H}}}}_{\color{red}{X = \#\text{flips}}}$$

Example 0.11. Consider the experiment of repeatedly drawing balls, with replacement, from an urn containing 5 red balls and 5 blue balls, until a red ball has been selected. Let $X =$ total $\#$ of trials needed.

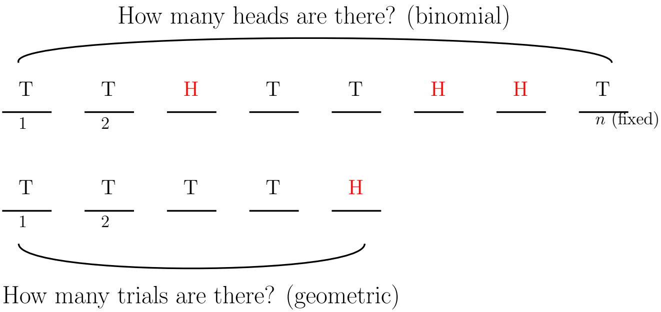
We make the following abstraction:

- The experiment consists of a sequence of **repeated Bernoulli trials** (i.e., each trial has only two outcomes: "success" and "failure");
- The **probability p of getting successes is always fixed**
- The Bernoulli trials are all **independent**;
- The experiment is stopped as soon as **one success has occurred**;
- X denotes the **total number of trials** that have been performed.

In short, X counts the total number of independent trials (with fixed probability of success) that are needed for the first success to occur.

Def 0.1. We say that the previous random variable X has a geometric distribution with parameter p, and write $X \sim \mathrm{Geom}(p)$.

Remark. Binomial (fixed #trials), geometric (fixed #successes):



Example 0.12. The following random variables both have geometric distributions:

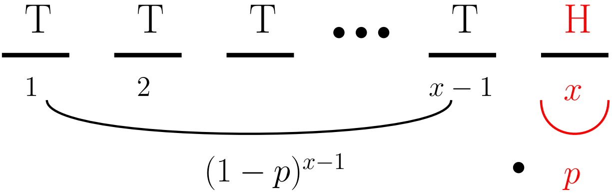
- (Repeatedly flip a fair coin until the first head appears) $X =$ total number of flips needed. Geom($\frac{1}{2}$)

- (Repeatedly draw balls with replacement from an urn containing 5 red balls and 5 blue balls) $X =$ #trials required to obtain a red ball for the first time. Geom($\frac{1}{2}$)

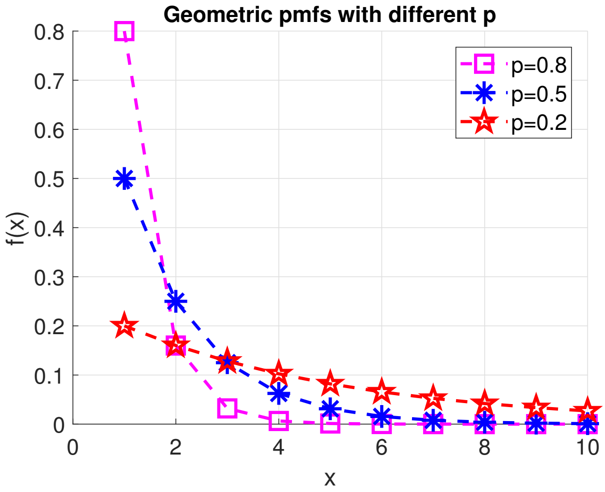
Theorem 0.5. The pmf of of $X \sim \mathrm{Geom}(p)$ is

$$f(x) = (1-p)^{x-1}p, \quad x = 1, 2, \ldots$$

Proof. See the following figure. $\square$



Example 0.13. Suppose X has a geometric distribution with $p = \frac{1}{2}$. Find $P(X = 4)$ and $F(X \geq 4)$.


Geometric pmfs with different p
p=0.8
p=0.5
p=0.2
f(x)
x
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
2
4
6
8
10

Infinitely many mathematicians walk into a bar.

The first says, "I'll have a beer."

The second says, "I'll have half a beer."

The third says, "I'll have a quarter of a beer."

The barman pulls out just two beers.

The mathematicians are all like, "That's all you're giving us? How drunk do you expect us to get on that?"

The bartender says, "Come on guys. Know your limits."

An infinite number of mathematicians walk into a bar.

The first one orders a beer.

The second orders half a beer.

The third orders a third of a beer.

The bartender bellows, “Get the hell out of here, are you trying to ruin me?”

Theorem 0.6. Let $X \sim \text{Geom}(p)$. Then

$$\mathrm{E}(X) = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

Proof. We prove only the formula for expected value. By definition,

$$\mathrm{E}(X) = \sum_{x=1}^{\infty} x \cdot (1-p)^{x-1} p = p \sum_{x=1}^{\infty} x(1-p)^{x-1}$$

This series is of the form $\sum_{x=1}^{\infty} x a^{x-1}$, which is equal to $\frac{1}{(1-a)^2}$. Applying this formula gives that

$$\mathrm{E}(X) = p \cdot \frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

$\square$

Negative Binomial

Briefly speaking, negative binomial distributions are generalizations of geometric distributions by waiting for more than one success.

Example 0.14. Consider the experiment of repeatedly and independently flipping a coin until a total of r heads have been obtained. Let $X = \#$trials needed. Then X follows a *negative binomial* distribution with parameters p (probability of getting heads) and r. We denote it by $X \sim NB(p, r)$.

$$\underbrace{\underset{1}{\underline{\mathrm{T}}}\quad \underset{2}{\underline{\mathrm{T}}}\quad \underline{\mathrm{T}}\quad \underline{\mathrm{H}}\quad \underline{\mathrm{T}}\quad \underline{\mathrm{H}}\quad \underline{\mathrm{T}}\quad \underline{\mathrm{T}}\quad \underline{\mathrm{H}}}_{X = \#\text{flips}}$$

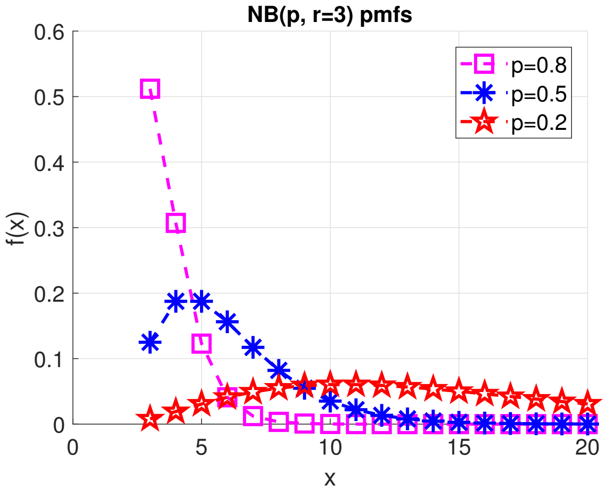
The pmf of X is

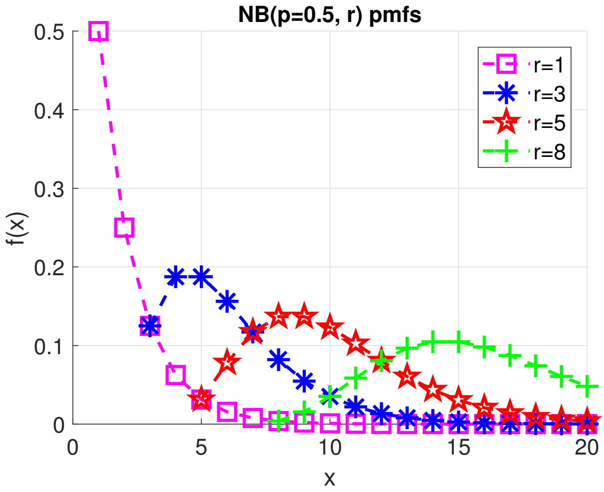
$$f_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = \underbrace{r, r+1, r+2, \ldots}_{\text{range}}$$

in which

- $\binom{x-1}{r-1}$: #ways of getting $r-1$ heads in first $x-1$ trials
- p^r: probability of getting r heads (including the very last head)
- $(1-p)^{x-r}$: probability of getting $x-r$ tails

$$\underbrace{\underset{1}{\underline{\mathrm{T}}}\quad \underset{2}{\underline{\mathrm{T}}}\quad \underline{\mathrm{T}}\quad \underline{\mathrm{H}}\quad \underline{\mathrm{T}}\quad \underline{\mathrm{H}}\quad \underline{\mathrm{T}}\quad \underset{x-1}{\underline{\mathrm{T}}}}_{x-1 \text{ trials, } r-1 \text{ heads}}\quad \underbrace{\underset{x}{\underline{\mathrm{H}}}}_{\text{last head}}$$


NB(p, r=3) pmfs
0.6
0.5
0.4
0.3
0.2
0.1
0
f(x)
0
5
10
15
20
x
p=0.8
p=0.5
p=0.2


NB(p=0.5, r) pmfs
0.5
0.4
0.3
0.2
0.1
0
f(x)
0
5
10
15
20
x
r=1
r=3
r=5
r=8

Theorem 0.7. Let $X \sim \mathrm{NB}(p, r)$. Then

$$\mathrm{E}(X) = \frac{r}{p}, \qquad \mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

Proof. Write X as a sum of iid Geom(p) random variables:

$$X = \underbrace{\underset{1}{\underline{\mathrm{T}}}\quad \underset{2}{\underline{\mathrm{T}}}\quad \underline{\mathrm{T}}\quad \underline{\mathrm{H}}}_{X_1} + \underbrace{\underline{\mathrm{T}}\quad \underline{\mathrm{H}}}_{\cdots} + \underbrace{\underline{\mathrm{T}}\quad \underline{\mathrm{T}}\quad \underline{\mathrm{H}}}_{X_r}$$

Then

$$\mathrm{E}(X_i) = \frac{1}{p}, \quad \mathrm{Var}(X_i) = \frac{1-p}{p^2}, \quad \text{for all } i = 1, \ldots, r$$

By linearity (and independence),

$$\mathrm{E}(X) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_r) = \frac{1}{p} + \cdots + \frac{1}{p} = \frac{r}{p}$$

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_r) = \frac{1-p}{p^2} + \cdots + \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}.$$

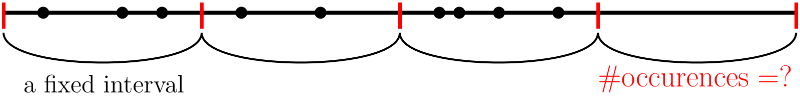
Poisson

Consider the following random variables:

- #hurricanes that hit a region each year
- #earthquakes occurring in certain country in a year
- #car accidents on a certain highway per week
- #phone calls received by a call center per minute
- #customers arriving at a bank counter in every hour
- #typos on each page of certain book

These examples have the following common characteristics:

- X counts the **total number of certain event**
- ... that is **rare** (with a small rate of occurrence λ)
- ... and occurs **independently** of each other
- ... over an **fixed interval of time or space**



Such random variables are modeled by **Poisson distributions**.

Def 0.2 ($X \sim$ Pois(λ)). We say that a discrete random variable has a Poisson distribution with parameter λ, if the pmf has the form

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \ldots$$

Remark. The Poisson pmf is based on the following power series:

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \frac{1}{0!} + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots$$

which implies that

$$1 = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda}$$

Theorem 0.8. If $X \sim \text{Pois}(\lambda)$, then $\mathrm{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$.

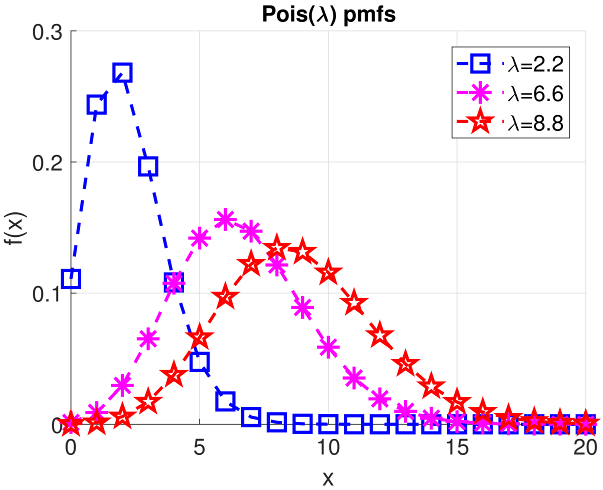
Proof. We only prove the formula for the expected value:

$$\mathrm{E}(X) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x}{x!} e^{-\lambda} = \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} e^{-\lambda} = \lambda \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda} = \lambda \cdot 1 = \lambda.$$

Remark. This theorem indicates that the parameter λ represents the mean value of the random variable.

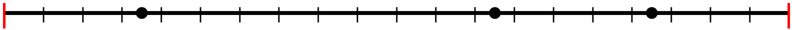
Example 0.15. Suppose, on average, 2.2 hurricanes hit a region each year. Let $X =$ #hurricanes next year. Then $X \sim$ Pois$(\lambda = 2.2)$. It follows that

- $P(X = 0) = e^{-2.2} = 0.1108$
- $P(X = 1) = 2.2e^{-2.2} = 0.2438$
- $P(X = 2) = \frac{2.2^2}{2} e^{-2.2} = 0.2681$
- $P(X \geq 2) = 1 - 3.2e^{-2.2} = 0.6454$


Pois(λ) pmfs
0.3
0.2
0.1
0
f(x)
0
5
10
15
20
x
λ=2.2
λ=6.6
λ=8.8

Theorem 0.9. If n is large and p is small, then $B(n,p) \approx \text{Pois}(\lambda = np)$.

Why this result is true: Consider the hurricane example again (where the number of hurricanes that hit a region in a year has a Poisson distribution with rate λ). We divide a year into n equal subintervals of time (e.g., 12 months, or 365 days).



When n is large, the number of hurricanes that occur in each subinterval is at most 1, thus a Bernoulli random variable with $p = \frac{\lambda}{n}$.

Therefore, the total number of hurricanes during the year is (approximately) binomial, with parameters n, p.

Example 0.16. Verify that for $X \sim \text{Pois}(2.2)$, $P(X = 2) = 0.2681$. In contrast, if $X \sim B(n = 365, p = \frac{2.2}{365})$, then $P(X = 2) = 0.2689$.

Remark. We can also use the binomial approximation to derive the expectation and variance of a Poisson random variable $X \sim \mathrm{Pois}(\lambda)$.

Since $\mathrm{Pois}(\lambda) \approx B(n,p)$ for some large n (and small $p = \frac{\lambda}{n}$), we have

$$\mathrm{E}(X) \approx np = n \cdot \frac{\lambda}{n} = \lambda;$$
$$\mathrm{Var}(X) \approx np(1-p) = n \cdot \frac{\lambda}{n} \cdot \left(1 - \frac{\lambda}{n}\right) \approx \lambda.$$

The above approximations become exact when n goes to infinity.

Example 0.17. The first draft of a probability textbook has 600 pages. Assume that the probability of any given page containing at least one typographical error is 0.015 and the numbers of errors on all the pages are mutually independent. Let T be the total number of pages which have at least one typographical error. Find the probability that $T = 9$.
Answer: .1328 (exact), or .1318 (approx)

Solution. Each page of the textbook corresponds to a Bernoulli trial (whether there is at least a typo on the page). There are 600 repeated trials and they are independent by the assumption. Thus, the exact distribution of T is $T \sim B(n = 600, p = 0.015)$. Using this, we get that

$$P(T = 9) = \binom{600}{9} 0.015^9 (1 - 0.015)^{600-9} = .1328.$$

Since this is a large n, small p setting, the binomial is approximately Poisson:

$$T \overset{\text{approx}}{\sim} \text{Pois}(\lambda = np = 9).$$

Consequently, by the Poisson approximation, we have

$$P(T = 9) \approx \frac{9^9}{9!} e^{-9} = .1318$$

We can see that it is very close to the exact probability given by the binomial distribution.