CSCI 3022-002 Intro to Data Science
Discrete Rvs

Opening **Example**: Suppose the Detroit Lions are secretly a coin with a 25% chance per game (flip) of winning (landing on heads). Define $X =$ the number of wins (heads) the Lions achieve in a 16 game season. What is $P(X = 0)$? What is $P(X = 1)$? What is $P(X = 2)$?

Opening Example Sol'n

bin (n, p)

$X \sim bin(16, 1/4)$ $\quad P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

Suppose the Detroit Lions are secretly a coin with a 25% chance per game (flip) of winning (landing on heads). Define $X =$ the number of wins (heads) the Lions achieve in a 4 game season. What is $P(X = 0)$? What is $P(X = 1)$? What is $P(X = 2)$?

$P(X=2)$: 2 wins:

(# of ways can choose 2 W's & 14 L's)

$\underbrace{P(1 \text{ of those ways})}$

P(14 x L's & 2 W's in that order)

$\binom{16}{2} \quad (p)^2 (1-p)^{14}$

Opening Example Sol'n

Suppose the Detroit Lions are secretly a coin with a 25% chance per game (flip) of winning (landing on heads). Define $X =$ the number of wins (heads) the Lions achieve in a 4 game season. What is $P(X = 0)$? What is $P(X = 1)$? What is $P(X = 2)$?

1. $P(X = 0) = P(\{LLLL\}) = .75^4$.
2. $P(X = 1) = P(\{WLLL\} \cup \{LWLL\} \cup \{LLLW\} \cup \{LLLW\}) = 4 \cdot .25^1 \cdot .75^3$.
3. $P(X = 2) = P(\{WWLL\} \cup \{WLWL\} \cup \{WLLW\} \cup \{LWWL\} \cup \{LWLW\} \cup \{LLWW\}) = 6 \cdot .25^2 \cdot .75^2$.
4. What's the pattern?

Announcements and Reminders

tonight

- Homework due ~~Friday the 27th.~~
- Notebook ~~day on Monday over the discrete rvs.~~

Nb on Friday

Last Time...: Repeated Trials

Counting!

1. Combinations: choose k things out of n; $C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$
2. Permutations: order all n things: $n!$; order r things out of n:

$$P(n,r) = \frac{n!}{(n-r)!}$$

.

The Binomial r.v. counts the total number of successes out of n trials, where X is the number of successes. Important Assumptions:

1. Each trial must be *independent* of the previous experiment.

2. The probability of success must be *identical* for each trial.

Binomials and Bernoullis

A Bernoulli rv is a single trial with success (or "1") with probability p, or:

$$P(X = 1) = p; \quad P(X = 0) = (1-p); \quad \textbf{OR } f(x) = \underbrace{p^x(1-p)^{(1-x)}}_{bin(1,p)}$$

1 2

The last way to write it is the same thing as a *binomial* with $n = 1$:

Now let $X :=$ the number of successes of n trials of a Bern(p). Then:

$$P(X = x) = (\#\text{ of ways that } X = x) \cdot P(\text{of one such outcome})$$

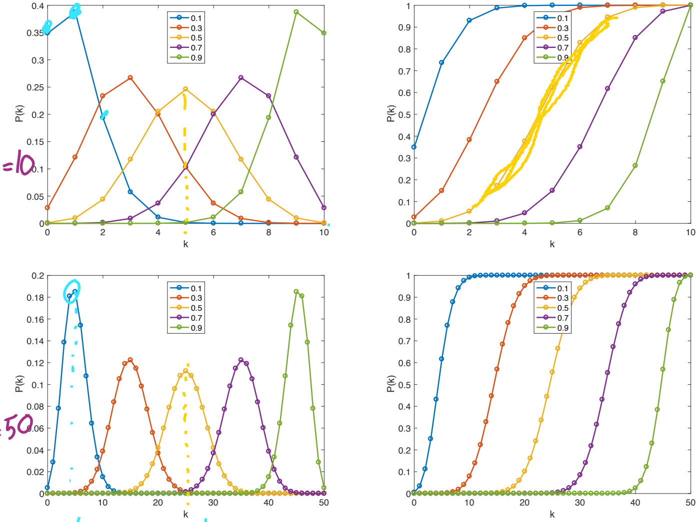
$$P(X = x) = \binom{n}{x} \cdot P(x \text{ successes}) \cdot P(n - x \text{ failures}).$$

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{(n-x)}; \quad x \in \{0, 1, 2, \ldots, n\}$$

Binomial $n = 10, 50$

pmf, cdf

$p = .50$



n=10

n=50

large sample

$F(x) = P(X \le x)$

↓ n: bumps get steeper, smoothe

↑ p→: bumps move with p.

: Heads unlikely $p(H) = .1$

1) pmf shape
2) cdf shape
3) How do parameters matter: n, p

The Geometric

$P(\text{first person matches}) = .1$

$P(\text{it's the second}) =$

$P(\text{not } 1, \text{ yes } 2) = P(\{\text{first} = \text{no}\} \cap \{\text{second} = \text{yes}\})$

$= P(\text{fail}) \cdot P(\text{success}) = (.9)(.1)$

Motivating example: A patient is waiting for a suitable matching kidney donor for a transplant. The probability that a randomly selected donor is a suitable match is 0.1.

What is the probability the first donor tested is the first matching donor? Second? Third?

$P(\text{third}) = P(\text{fail \& then fail \& then succeed}) = (.9)(.9)(.1) = .9^2 \cdot .1$

(The per-donor probability checks are independent and identically distributed!)

Same as binomial

The Geometric pdf

$P(X=x) = P(\text{trial } x \text{ is the first success})$

Continuing in this way, a general formula for the pmf emerges:

$= P(x-1 \text{ failures then a success})$

$= (1-p)^{x-1} \; p$

The parameter p can assume any value between 0 and 1.
Depending on what parameter p is, we get different members of the geometric distribution.

NOTATION: We write $X \sim geom(p)$ to indicate that X is a Geometric rv with success probability p.

The Geometric pdf

OR Count X := # failures until first success

Continuing in this way, a general formula for the pmf emerges:

$$P(X = x) = P(\text{failed x-1 times}) \cdot P(\text{then success!})$$

$(1-p)^x$ p

$$P(X = x) = (1-p)^{x-1} p; \quad x \in \{1, 2, 3, \ldots, \infty\}$$

x=0 is possible

The parameter p can assume any value between 0 and 1.
Depending on what parameter p is, we get different members of the geometric distribution.

NOTATION: We write $X \sim geom(p)$ to indicate that X is a Geometric rv with success probability p.

The Geometric pdf

Continuing in this way, a general formula for the pmf emerges:

$$P(X = x) = P(\text{failed x-1 times}) \cdot P(\text{then success!})$$

$$P(X = x) = (1-p)^{x-1}p; \quad x \in \{1, 2, 3, \ldots, \infty\}$$

The parameter p can assume any value between 0 and 1.
Depending on what parameter p is, we get different members of the geometric distribution.

np.random.geometric.

NOTATION: We write $\underline{X \sim geom(p)}$ to indicate that X is a Geometric rv with success probability p.
Important **note:** sometimes the geometric is counting the number of total *trials*; sometimes it's counting the number of *failures*. Know which one your software is doing!

The Negative Binomial

Motivating example:
A "successful toss" is defined to be the coin landing on heads. Let $X = \#$ of failures/tails before the *second* success/heads.

add 2 geometrics?

$\{THTH\}: X = 2$

$\{HH\}: x = 0$

$\{HTH\}$ or $\{THH\}$ $x = 1$

How is this related to the geometric distribution? The binomial distribution?

The Negative Binomial

Motivating example:
A "successful toss" is defined to be the coin landing on heads. Let $X = \#$ of failures/tails before the *second* success/heads.
Events in $X = 2$: $\{HTH, THH\}$
Events in $X = 3$: $\{HTTH, THTH, TTHH\}$
Events in $X = 4$: $\{HTTTH, THTTH, TTHTH, TTTHH\}$

How is this related to the geometric distribution? The binomial distribution?
It's like adding two geometrics.
The relationship to the binomial is a little harder, but if we know this random variables equals x, what do we know about trial $\#x$? The previous $x - 1$ trials?

trial x: H (success)

previous $x-1$: contained all but 1 H.

The Negative Binomial

& including

In general, let $X = \#$ of trials before the rth success. The pdf/pmf is:

$P(X=x) = P($in the first $x-1$ tries, got $r-1$ successes AND THEN got success $\#r)$.

$= \boxed{P(\text{exactly } r-1 \text{ heads in } x-1 \text{ tries})} \cdot P(\text{success})$.

binomial; $n = x-1$ $P(r-1 \text{ successes})$

$n-x$

$[x-1]-[r-1]$

$= \binom{x-1}{r-1} p^{(r-1)} (1-p) \cdot p$

$= \binom{x-1}{r-1} p^{r} (1-p)^{x-r}$

NOTATION: We write ____________ to indicate that X is a Negative Binomial rv with success probability p and r successes until completion.

The Negative Binomial

In general, let $X = \#$ of trials before the rth success. The pdf/pmf is:

$$P(X = x) = (\#\text{ of ways that } X = x) \cdot P(\text{of one such outcome})$$

NOTATION: We write $\underline{X \sim NB(r,p)}$ to indicate that X is a Negative Binomial rv with success probability p and r successes until completion.

The Negative Binomial

In general, let $X = \#$ of trials before the rth success. The pdf/pmf is:

$$P(X = x) = (\#\text{ of ways that } X = x) \cdot P(\text{of one such outcome})$$

$$(\#\text{ of ways that } x-1 \text{ trials contain exactly } r-1 \text{ successes})$$

$$\cdot P(\text{r successes and } (x-1)-(r-1) \text{ failures}).$$

$$= \binom{x-1}{r-1} p^{r-1}(1-p)^{(x-1)-(r-1)}p$$

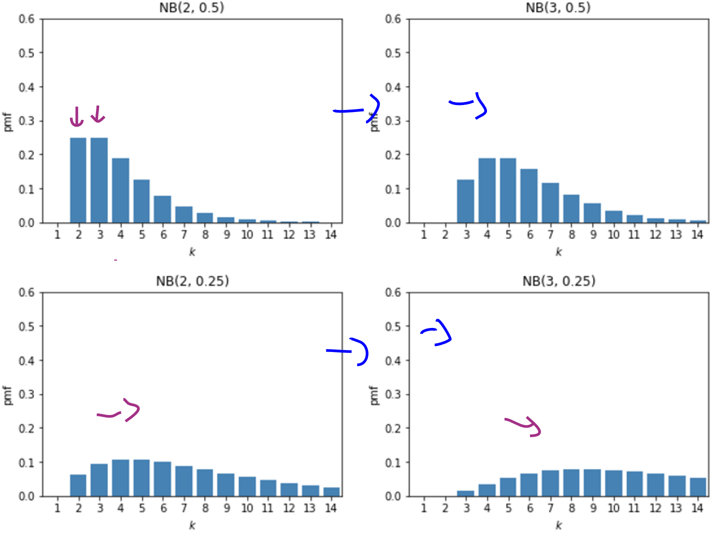
$$P(X = x) = \binom{x-1}{r-1} p^{r}(1-p)^{(x-r)}$$

for $x = \{r, r+1, r+2, \ldots \infty\}$.

NOTATION: We write $\boxed{\underline{X \sim NB(r,p)}}$ to indicate that X is a Negative Binomial rv with success probability p and r successes until completion.

NB pdfs

$r=2$ $NB(r, p)$ $r=3$



$p=0.5$
$=1/2$

$p=0.25$
$=1/4$

as p decreases, it takes longer.

as # of successes needed is higher, it takes longer, and spreads out

The Negative Binomial

Example:
A physician wishes to recruit 5 people to participate in a new health regimen. Let p = .2 be the probability that a randomly selected person agrees to participate. What is the probability that 15 people must be asked before 5 are found who agree to participate?

Solve: Let $X \sim NB(5, 1/5)$ (r, p)

$$P(X = x) = \underbrace{\binom{14}{4}}_{\text{counting part!}} \cdot p^5 \quad (1-p)^{10}$$

$P(X = 15)$

The Negative Binomial

Example:
A physician wishes to recruit 5 people to participate in a new health regimen. Let p = .2 be the probability that a randomly selected person agrees to participate. What is the probability that 15 people must be asked before 5 are found who agree to participate?

For $X \sim NB(5, .2)$, find $P(X = 15)$:

The Negative Binomial

Example:
A physician wishes to recruit 5 people to participate in a new health regimen. Let p = .2 be the probability that a randomly selected person agrees to participate. What is the probability that 15 people must be asked before 5 are found who agree to participate?

For $X \sim NB(5, .2)$, find $P(X = 15)$:

$$P(X = 15) = \binom{15-1}{5-1}.2^5(.8)^{(15-5)}$$

The Poisson Distribution/RV

A Poisson r.v. describes the total number of events that happen in a certain time period.

↳ constant rate (on avg.)

Examples:

$\#$ of vehicles arriving at a parking lot in one week
$\#$ of gamma rays hitting a satellite per hour
$\#$ of cookies sold at a bake sale in 1 hour

The Poisson Distribution/RV

A Poisson r.v. describes the total number of events that happen in a certain time period.

A discrete random variable X is said to have a Poisson distribution with parameter λ $(\lambda > 0)$ if the pdf of X is

NOTATION: We write ____________ to indicate that X is a Poisson r.v. with parameter λ.

The Poisson Distribution/RV

A Poisson r.v. describes the total number of events that happen in a certain time period.

A discrete random variable X is said to have a Poisson distribution with parameter λ $(\lambda > 0)$ if the pdf of X is

aside:

$\sum_{n=0}^{\infty} \frac{2^n}{n!} = e^2$

$$P(X = x) = f(x) = \frac{e^{-\lambda}\lambda^x}{x!}; \quad x \in 0, 1, 2, \infty$$

$= e^{-\lambda} \left(\frac{\lambda^x}{x!} \right)$

λ := rate; units $\frac{\text{events}}{\text{time}}$

NOTATION: We write $X \sim Pois(\lambda)$ to indicate that X is a Poisson r.v. with parameter λ.

The Poisson Distribution/RV

Example:
Let X denote the number of mosquitoes captured in a trap during a given time period. Suppose that X has a Poisson distribution with $\lambda = 4.5$. What is the probability that the trap contains 5 mosquitoes?

$$\lambda = 4.5 \frac{\text{mosq}}{\text{min.}}$$

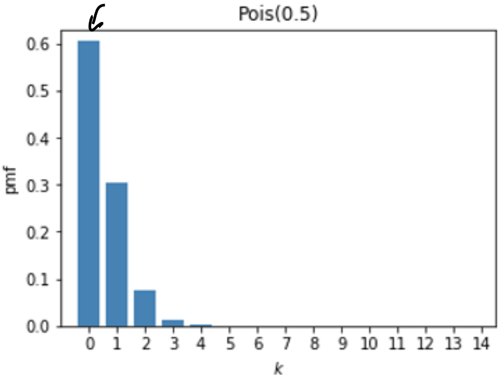
The Poisson Distribution/RV

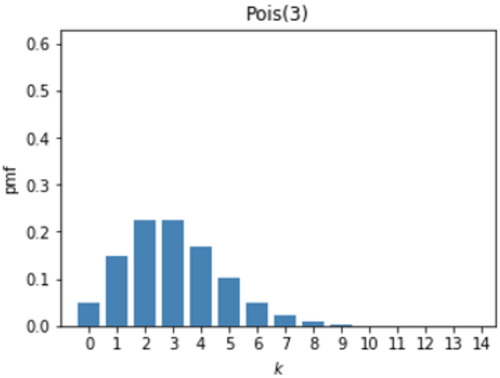
Example:
Let X denote the number of mosquitoes captured in a trap during a given time period. Suppose that X has a Poisson distribution with $\lambda = 4.5$. What is the probability that the trap contains 5 mosquitoes? $P(X = 5) =$

$$e^{-\lambda} \frac{\lambda^x}{x!} : \quad e^{-4.5} \cdot \frac{(4.5)^5}{5!} = \cancel{A}$$

Poisson pdfs

on avg $\frac{1/2 \text{ event}}{1 \text{ time}}$ $\frac{3 \text{ events}}{\text{time}}$





Poisson and... binomial?

One way to generate the Poisson is to take limits of a binomial: suppose you get texts during class ($\frown$) at a rate of 29 texts per hour. What is the probability that you get 29 texts in an hour? 12 texts in an hour? 107 texts in an hour?

λ is the *rate* of the Poisson.

Poisson and... binomial?

One way to generate the Poisson is to take limits of a binomial: suppose you get texts during class $(\frown\!\!\!\!\!\!\cdot\cdot)$ at a rate of 29 texts per hour. What is the probability that you get 29 texts in an hour? 12 texts in an hour? 107 texts in an hour?

λ is the *rate* of the Poisson.

Think about a Bernoulli that represents your friends asking "should I text...?" then flipping a coin with probability p. Then:

Poisson and... binomial?

One way to generate the Poisson is to take limits of a binomial: suppose you get texts during class $\left(\frown\right)$ at a rate of 29 texts per hour. What is the probability that you get 29 texts in an hour? 12 texts in an hour? 107 texts in an hour?

λ is the *rate* of the Poisson.

Think about a Bernoulli that represents your friends asking "should I text...?" then flipping a coin with probability p. Then:

$\lambda = \frac{texts}{hour} \approx \frac{flips}{hour} \cdot \frac{texts}{flip} = np$ for the same n and p as a *binomial*.

Poisson and... binomial?

One way to generate the Poisson is to take limits of a binomial: suppose you get texts during class $\left(\frown\right)$ at a rate of 29 texts per hour. What is the probability that you get 29 texts in an hour? 12 texts in an hour? 107 texts in an hour?

λ is the *rate* of the Poisson.

Think about a Bernoulli that represents your friends asking "should I text...?" then flipping a coin with probability p. Then:

$\lambda = \frac{texts}{hour} \approx \frac{flips}{hour} \cdot \frac{texts}{flip} = np$ for the same n and p as a *binomial*.

...but n might vary a bit from hour to hour, so these are only equivalent *in the limit* (n large, p small)!

Discrete Distributions Example

Example:

A factory makes parts for a medical device company. 6% of those parts are defective. For each one of the problems below:

(i.) Define an appropriate random variable for the experiment.

(ii.) Give the values that the random variable can take on.

(ii.) Find the probability that the random variable equals 2.

(iv.) State any assumptions you need to make.

Problems:

1. Out of 10 parts, X are defective.
2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.
3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

Discrete Distributions Example

6% of those parts are defective.

1. Out of 10 parts, X are defective.

(i.) r.v.:

(ii.) Values of r.v.:

(iii.) $P(X = 2)$:

(iv.) Assumptions:

Discrete Distributions Example

6% of those parts are defective.

1. Out of 10 parts, X are defective.

(i.) r.v.:

$$X \sim bin(10, .06)$$

(ii.) Values of r.v.:

$$X \in \{0, 1, 2, \ldots, 10\}$$

(iii.) $P(X = 2)$:

$$\binom{10}{2} .06^2 .94^8$$

(iv.) Assumptions: Parts are *i.i.d.*

Discrete Distributions Example

6% of those parts are defective.

2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.

(i.) r.v.:

(ii.) Values of r.v.:

(iii.) $P(X = 2)$:

(iv.) Assumptions:

Discrete Distributions Example

6% of those parts are defective.

2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.

(i.) r.v.:

$$X + 1 \sim Geom(.06)$$

(ii.) Values of r.v.:

$$X \in \{0, 1, 2, \ldots, \infty\}$$

(iii.) $P(X = 2)$:

$$.94^2 .06^1$$

(iv.) Assumptions: Parts are *i.i.d.*

Discrete Distributions Example

6% of those parts are defective.

3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

(i.) r.v.:

(ii.) Values of r.v.:

(iii.) $P(X = 2)$:

Discrete Distributions Example

6% of those parts are defective.

3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

(i.) r.v.:

$$X \sim Pois(10)$$

(ii.) Values of r.v.:

$$X \in \{0, 1, 2, \ldots, \infty\}$$

(iii.) $P(X = 2)$:

$$\frac{e^{-10} \cdot 10^2}{2!}$$