Commonly Used Discrete Distributions **Notes 03**

Associated Reading: Wackerly 7, Chapter 3, Sections 4-8

We start with a motivating example. Let's assume a sequence of n coin flips, and let the probability of observing a head be p (not necessarily 0.5!). Let the r.v. Y increase in value by 1 every time we observe a head, and by 0 if not. What we observe is the following:

results of each coin flip.

$\{X_1\ X_2\ X_3 \cdots X_n\}$.

$X_i \in \{0,1\}$

X_i's independent. $P(X_i = 1) = p$.

$Y_n = \sum_{i=1}^{n} X_i$

Chaterize dist of Y_n.

$1 \times 2 \times 3 \cdots \times n$

$\binom{n}{y} = \frac{n!}{y!(n-y)!}$

What is the probability of observing $Y_n = y$ heads, given the probability p?

Y_n : number of heads in n coin flips.

$$\boxed{P(Y_n = y) = \binom{n}{y} p^y (1-p)^{n-y}}$$

number of ways to pick y items from n items.

suppose : $\boxed{\underbrace{HH \cdots H}_{y \text{ flips}}\ \underbrace{T \cdots T}_{n-y \text{ tails}}}$

$p^y \cdot (1-p)^{n-y}$

This probability distribution is the *binomial* distribution.[a] The binomial distribution is used to model binomial experiments, which have the following properties:

- The number of trials, n, is *fixed*.
- Each trial has two possible outcomes: S (success) or F (failure).
- The probability of success remains p throughout the experiment.
- Each trial is *independent* of the others.
- The r.v. of interest is Y, the total number of successes in n trials.

$Y \sim Bin(n, p)$.

∝

It is the case that statisticians will sometimes apply the binomial distribution in a setting where the probability of success changes from trial to trial.

Sampling without replacement from a population of size N.

define n = number of items of interest.

Suppose $\frac{n}{N} \leq 0.1$ people will use bionomial dist & get results that are approximately correct.

(⚠ p changes over time)

[a] Actually, the binomial *family* of distributions; family members are defined by a unique combination of n and p.

Now that this is established, let's look at the interrelationship between the probability distributions that we'll be looking at in Sections 4 to 7 of Chapter 3:

$\{X_1, X_2 \cdots X_n\}$ $X_i \sim$ Bernoulli dist

Bernoulli Process → Sample Without Replacement → Hypergeometric Distribution (number of successes in n trials, finite population N)

Bernoulli Process ↓ Sampling With Replacement ↓ Binomial Distribution (number of successes in n trials)

Sampling With Replacement → Negative Binomial Distribution (Number of trials for s successes) → Geometric Distribution (Number of trials for one success)

Note that each of the named distributions is the subject of its own section. One that isn't listed explicitly is the Bernoulli distribution:

Bernoulli dist. ⇒ just binomial dist with $n=1$. $Bin(1, p)$.

$$p(y) = P(Y=y) = \begin{cases} p & y=1 \\ 1-p & y=0 \end{cases} = \binom{1}{y} p^y (1-p)^{1-y}$$

The Bernoulli distribution, along with one piece of information we haven't explicitly covered yet (from Chapter 5, Section 8), allows us to derive the expected number of successes and the variance on the expected number of successes for a binomial experiment in a much cleaner way than that given in the proof of Theorem 3.7.

$X_i \sim$ Bernoulli dist.

$$E[X_i] = 1\cdot p + 0\cdot(1-p) = p.$$

$$V[X_i] = E[X_i^2] - (E[X_i])^2 = 1^2\cdot p + 0^2\cdot(1-p) - p^2 = p - p^2$$

$$E[Y_n] = E\left[\sum_{i=1}^{n} X_i\right] = n\cdot p. \qquad Y_n = \sum_{i=1}^{n} X_i$$

$$V[Y_n] = \boxed{V\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} V[X_i]} = n(p-p^2) = np(1-p).$$

when $n=2$.

$$V[X_1+X_2] = E(X_1+X_2)^2 - (E(X_1+X_2))^2$$

$E[X_1]=\mu_1$
$E[X_2]=\mu_2$

$$= E[X_1^2] + E[X^2] + 2E[X_1X_2] - (\mu_1+\mu_2)^2$$

$$= E[X_1^2] + E[X_2^2] + 2E[X_1X_2] - \mu_1^2 - \mu_2^2 - 2\mu_1\mu_2.$$

$$= V[X_1] + V[X_2] + 2\left(E(X_1X_2) - \mu_1\mu_2\right)$$

$\boxed{\mathbb{E}[X_1X_2]} = \sum_{x_1 x_2} x_1 x_2\, p(X_1 = x_1, X_2 = x_2) = \sum_{x_1 x_2} x_1 x_2\, p(x_1)\cdot p(x_2).$ ⇒ independent

$= \sum_{x_1} \sum_{x_2} (x_1 p(x_1)) \cdot (x_2 p(x_2))$

$= \sum_{x_1} x_1 p(x_1) \cdot \sum_{x_2} x_2 p(x_2)$

$\boxed{= \mu_1 \cdot \mu_2}$

Let's wrap this up by combining together the previous information:

THE BINOMIAL DISTRIBUTION

NOTATION: $Y \sim \text{Bin}(n, p)$ n: NUMBER OF TRIALS, p: SUCCESS PROBABILITY

PMF: $p(y) = \binom{n}{y} p^y (1-p)^{n-y}$ $0 \le y \le n$

EXPECTED VALUE: $E[Y] = \mu = np$ VARIANCE: $V[Y] = \sigma^2 = np(1-p)$

R FUNCTIONS: `dbinom(y,n,p)` (PMF) / pdf $= P(Y=y)$

`pbinom(y,n,p)` (CDF) $= P(Y \le y)$.

`rbinom(k,n,p)` (SIMULATION OF k BINOMIAL R.V.'S)

→ **EXAMPLE.** Wackerly 7, Exercise 3.51

exactly ... games that A wins among the first ...

3.51 In the 18th century, the Chevalier de Mere asked Blaise Pascal to compare the probabilities of two events. Below, you will compute the probability of the two events that, prior to contrary gambling experience, were thought by de Mere to be equally likely.

a What is the probability of obtaining at least one 6 in four rolls of a fair die?

b If a pair of fair dice is tossed 24 times, what is the probability of at least one double six?

a) $Y =$ "number of 6's in four rolls"

$Y \sim \text{Bin}(n=4, p=\frac{1}{6})$.

$P(Y \ge 1) = \underbrace{P(Y=1) + P(Y=2) \cdots P(Y=4)}$.

$= 1 - P(Y=0)$.

$= 1 - \binom{4}{0} p^0 (1-p)^4$ where $p = \frac{1}{6}$.

$= 0.518$

b) $Y =$ "number of times that double six appears"

$Y \sim \text{Bin}(n=24, p=\frac{1}{36})$

$P(Y \ge 1)$

$= 1 - P(Y=0)$

$= 1 - \binom{24}{0} p^0 (1-p)^{24}$

$= 0.491$

→ **EXAMPLE.** Wackerly 7, Exercise 3.65

***3.65** Refer to Exercise 3.64. The *maximum likelihood estimator* for p is Y/n (note that Y is the binomial random variable, not a particular value of it).

a Derive $E(Y/n)$. In Chapter 9, we will see that this result implies that Y/n is an unbiased estimator for p.

b Derive $V(Y/n)$. What happens to $V(Y/n)$ as n gets large?

$Y \sim \text{Bin}(n, p)$

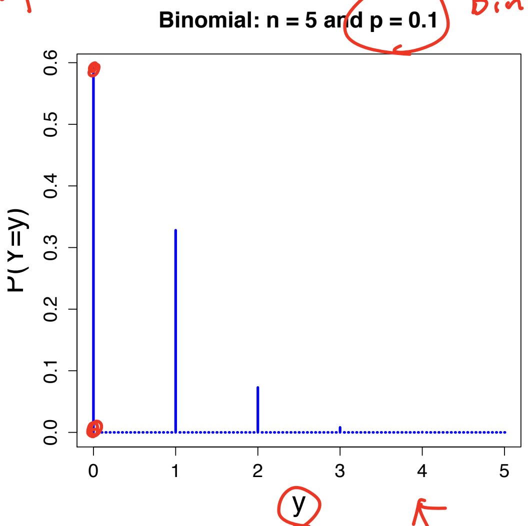
$\mathbb{E}[\frac{Y}{n}] = p = \frac{\mathbb{E}[Y]}{n}$ ↙ $n \cdot p$.

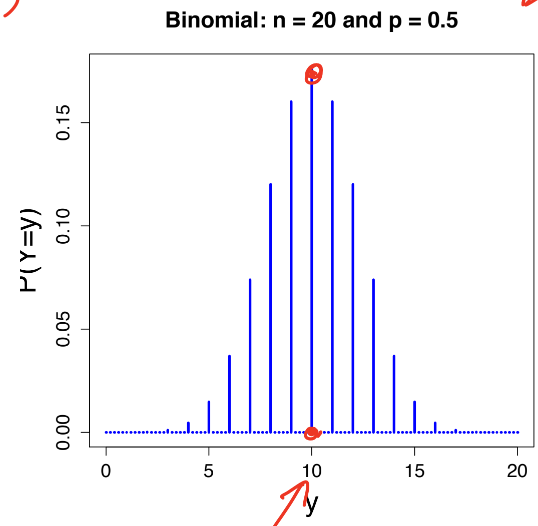
$V[\frac{Y}{n}] = \frac{1}{n^2} \cdot n \cdot p \cdot (1-p) = \frac{p(1-p)}{n} \to 0.$

$Y \sim Bin(n, p)$

pmf

$Bin(5, 0.1)$ $\qquad$ $Bin(20, \frac{1}{2})$





Referring back to the figure on page 2 of these notes, we see that the *geometric distribution* governs the situation where we want to determine the number of trials before observing the first success. Starting from the Bernoulli distribution, it is easy to intuitively define the pmf for the geometric distribution:

Sample space: $\{S, FS, FFS, FFFS, \cdots\}$.

Y: number of total experiments until you see a success.

$$y \geq 1 \Rightarrow \boxed{P(Y=y) = \underbrace{(1-p)}_{\text{prob. failure}}{}^{y-1} \cdot p} \quad \text{pmf.}$$

if each experiment is indep. & success prob $= p$.

$Y \sim Geom(p)$

Less easy is the derivation of $E[Y]$:

$$E[Y] = \sum_{y=1}^{\infty} y \cdot p(y)$$

$$= \sum_{y=1}^{\infty} y \cdot (1-p)^{y-1} \cdot p$$

$$= p \cdot \sum_{y=1}^{\infty} \underbrace{y \cdot (1-p)^{y-1}}_{\boxed{y \cdot q^{y-1}}}$$

let $q = 1-p$

$$\boxed{1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x} \quad (x < 1)}$$

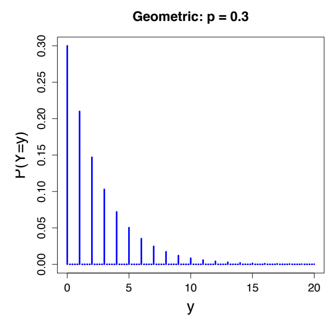
$$= p \cdot \sum_{y=1}^{\infty} \frac{d}{dq} q^y = p \cdot \frac{d}{dq}\left(\sum_{y=1}^{\infty} q^y\right) = p \cdot \frac{d}{dq}\left(\frac{q}{1-q}\right) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

$\frac{1}{(1-q)^2}$

$$q + q^2 + q^3 + \cdots$$
$$= q(1 + q + q^2 + \cdots)$$

$E[Y^2]$.

The Geometric Distribution



Notation: $Y \sim \text{Geom}(p)$ p: success probability

PMF: $p(y) = p(1-p)^{y-1}$ $1 \le y < \infty$

Expected Value: $E[Y] = \mu = 1/p$

Variance: $V[Y] = \sigma^2 = (1-p)/p^2$ $V[Y] = E[Y^2] - (EY)^2$

R Functions: `dgeom(y-1,p)` (PMF)
`pgeom(y-1,p)` (CDF)
`rgeom(k,p)` (Simulation of k geometric r.v.'s)
NOTE: the output will be the number of failures!

`dgeom(9, p)` interested in prob of seeing first success in 10-th trial.

→ **EXAMPLE.** Wackerly 7, Exercise 3.71

3.71 Let Y denote a geometric random variable with probability of success p. $Y \sim \text{Geom}(p)$

a Show that for a positive integer a,
$$P(Y > a) = q^a.$$

b Show that for positive integers a and b,
$$P(Y > a + b | Y > a) = q^b = P(Y > b).$$
This result implies that, for example, $P(Y > 7|Y > 2) = P(Y > 5)$. Why do you think this property is called the *memoryless* property of the geometric distribution?

c In the development of the distribution of the geometric random variable, we assumed that the experiment consisted of conducting identical and independent trials until the first success was observed. In light of these assumptions, why is the result in part (b) "obvious"?

$Y>a$ $Y>a+b$

a) $P(Y>a) = P(Y=a+1) + P(Y=a+2) + P(Y=a+3) \cdots\cdot$

$1 - P(Y \le a)$ $= q^a \cdot p + q^{a+1}\cdot p + q^{a+2}\cdot p + \cdots\cdot$

$= \cancel{p}\cdot q^a \underbrace{(1 + q + q^2 + \cdots)}_{\frac{1}{\cancel{1-q}}} = q^a$

$Y>0$ $Y>b$

$P(Y>a+b \mid Y>a)$
$= P(Y>b)$

b) $P(Y>a+b \mid Y>a)$ $Y>a+b$

$= \dfrac{P(\boxed{Y>a+b \cap Y>a})}{P(Y>a)} = \dfrac{P(Y>a+b)}{P(Y>a)} \overset{\text{part (a)}}{=} \dfrac{q^{a+b}}{q^a} = q^b = P(Y>b)$

"memoryless" property

result in trial $i+1$ does not depend on result on trial i.

→ **EXAMPLE.** Wackerly 7, Exercise 3.77

3.77 If Y has a geometric distribution with success probability p, show that
$$P(Y = \text{an odd integer}) = \frac{p}{1-q^2}.$$
$Y \sim \text{Geom}(p)$

$P(Y \text{ is an odd integer})$

$= P(Y = 2k+1 \text{ for } k=0,1,2\cdots)$

$= P(Y=1) + P(Y=3) + P(Y=5) + \cdots\cdot$

$= p + q^2\cdot p + q^4 \cdot p + q^6 \cdot p + \cdots\cdot$

$= p\underbrace{(1 + q^2 + q^4 + \cdots)}$ $= p \cdot \dfrac{1}{1-q^2} = \dfrac{\cancel{p}}{\cancel{(1-q)}(1+q)} = \dfrac{1}{1+q}.$

outcomes of my experiments. $X_i \in \{S, F\}$.

Now we generalize the geometric distribution to model how many trials are necessary not just for the first success, but for the first r successes. The governing family is the *negative binomial distribution*:

Y: number of trials until you see r successes.

$\{\underbrace{X_1\ X_2\ X_3\ \cdots}_{r-1:\ S}\ \underset{S}{X_y}\}$ → scalar.

$y \geq r$ successes.

$$\boxed{P(Y=y) = \binom{y-1}{r-1} \cdot p^{r-1} \cdot (1-p)^{y-r} \cdot p}$$

① $NB(1,p) = Geom(p)$.

$Y \sim NB(r, p)$

The Negative Binomial Distribution

Notation: $Y \sim \text{NB}(r,p)$ r: number of successes, p: success probability

PMF: $p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$ $r \leq y < \infty$

Expected Value: $E[Y] = \mu = r/p$ Variance: $V[Y] = \sigma^2 = r(1-p)/p^2$

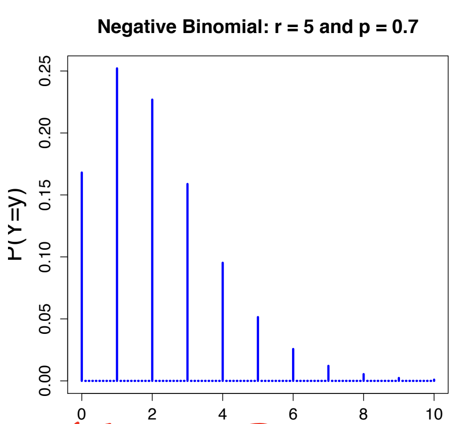
R Functions: `dnbinom(y-r,r,p)` (PMF)

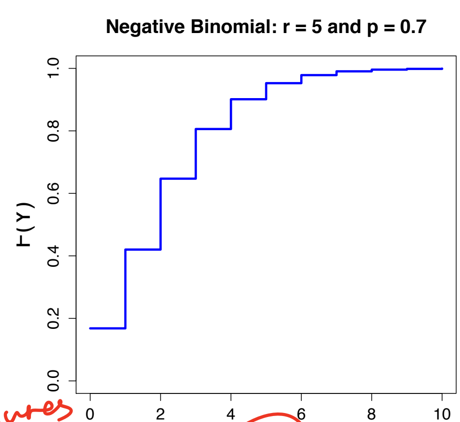
`pnbinom(y-r,r,p)` (CDF)

`rnbinom(k,r,p)` (Simulation of k negative binomial r.v.'s)

NOTE: the output will be the number of failures!

NB(5, 0.7)





number of failures in R.

→ **EXAMPLE.** Wackerly 7, Exercise 3.91 $p = 0.4$

3.91 Refer to Exercise 3.90. If each test costs \$20, find the expected value and variance of the total cost of conducting the tests necessary to locate the three positives.

let Y: number of tests until we see 3 positives.

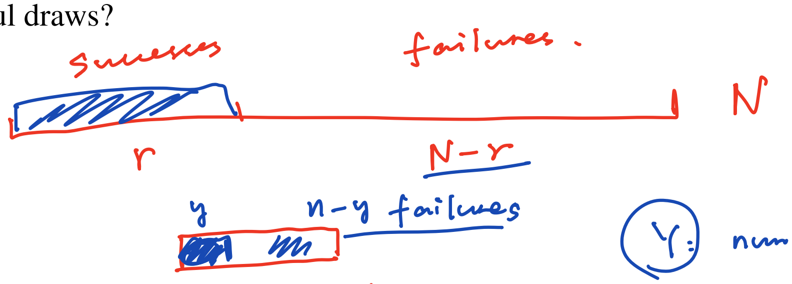
$Y \sim NB(3,\ 0.4)$.

$C = 20 \cdot Y$

$\mathbb{E}[C] = 20\,\mathbb{E}[Y] = 20 \cdot \frac{3}{0.4} = 150.$

$V[C] = \underbrace{400}_{20^2}\ \underbrace{V[Y]}_{r \cdot \frac{(1-p)}{p^2}} = 4500.$

The penultimate distribution that we will look at in this set of notes is the *hypergeometric distribution*. This distribution takes the place of the binomial distribution if we are randomly sampling *without* replacement. Assume that we draw n samples from a population of size N, in which there are r "success" objects and $N - r$ "failure" objects. What is the probability of observing y successful draws?



Y: number of successes in your n samples.

$$P(Y = y) = \frac{\binom{r}{y} \cdot \binom{N-r}{n-y}}{\binom{N}{n}}$$

number of ways to choose successes

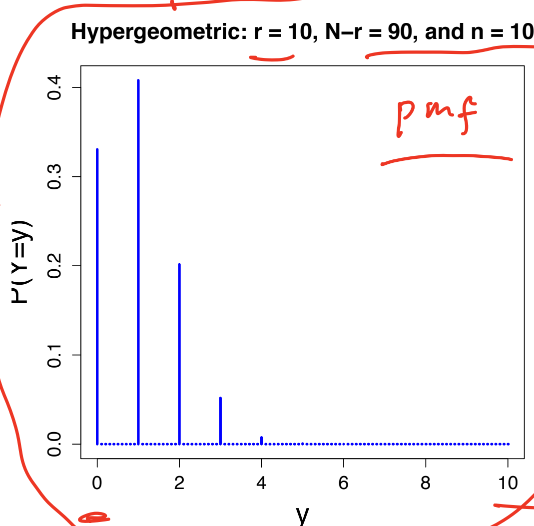
number of ways to select failures.

total number of ways to select n samples.

$Y \sim HG(\boxed{r, N}, \boxed{n})$.

population

number of samples.



$Y \sim HG(10,\ 100,\ 5)$ $\quad P(Y = 3)$

dhyper(3, 10, $\boxed{90}$, 5).

The Hypergeometic Distribution

Notation: $Y \sim HG(r, N, n)$ r: number of successes in population

N: population size

n: number of trials

r N

Sample with replecement. $Y \sim Bin(n, p = \frac{r}{N})$.

$E[Y] = n \cdot \frac{r}{N}$ $V[Y] = n \cdot \frac{r}{N} \cdot \frac{N-r}{N}$

PMF: $p(y) = \dfrac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$ $0 \le y \le n,\ \ y \le r,\ \ n - y \le N - r$

Expected Value: $\boxed{E[Y] = \mu = nr/N}$ Variance: $V[Y] = \sigma^2 = \boxed{n(r/N)[(N-r)/N]}\boxed{[(N-n)/(N-1)]}$

correction factor $\frac{N-n}{N-1} \to 1$ (n small, N large)

R Functions: `dhyper(y,r,N-r,n)` (PMF)

`phyper(y,r,N-r,n)` (CDF)

`rhyper(k,r,N-r,n)` (Simulation of k hypergeometric r.v.'s)

→ **EXAMPLE.** Wackerly 7, Exercise 3.107

3.107 A group of six software packages available to solve a linear programming problem has been ranked from 1 to 6 (best to worst). An engineering firm, unaware of the rankings, randomly selected and then purchased two of the packages. Let Y denote the number of packages purchased by the firm that are ranked 3, 4, 5, or 6. Give the probability distribution for Y.

r = 4 N = 6

n = 2

$Y \sim HG(4, 6, 2)$.

$Y = 0, 1, 2$

$P(Y=0) = \dfrac{\binom{4}{0}\cdot\binom{2}{2}}{\binom{6}{2}} = \frac{1}{15}$

$P(Y=1) = \dfrac{\binom{4}{1}\binom{2}{1}}{\binom{6}{2}} = \frac{8}{15}$ $P(Y=2) = \frac{6}{15}$

→ **EXAMPLE.** Wackerly 7, Exercise 3.113

3.113 A jury of 6 persons was selected from a group of 20 potential jurors, of whom 8 were African American and 12 were white. The jury was supposedly randomly selected, but it contained only 1 African American member. Do you have any reason to doubt the randomness of the selection?

8 12 N = 20

n = 6

Y: number of African American members selected.

unlikely to happen: check the randomness of selection.

$Y \sim HG(8, 20, 6)$.

$\boxed{P(Y \le 1)} = p(0) + p(1) = \dfrac{\binom{8}{0}\binom{12}{6}}{\binom{20}{6}} + \dfrac{\binom{8}{1}\binom{12}{5}}{\binom{20}{6}} = \boxed{0.187}$.

Poisson distribution.

Let's say you, for some reason, want to model the number of deaths per year among Prussian soldiers due to . . . (wait for it) . . . horse kicks. Obviously, the random variable is discrete (zero deaths, one, two, etc.), but because there can be more than two possible outcomes per time period, we are *not* dealing with a Bernoulli process.

Here's data compiled by Ladislaus Bortkiewicz, from 10 Prussian army corps over a 20-year period:

200 corp-year.

$Y \sim$

y	0	1	2	3	4
$N(y)$	109	65	22	3	1

Here, y is the number of deaths observed, and $N(y)$ is the number of corps-years in which y deaths were observed. ($N(y)$ sums to $20 \times 10 = 200$.)

One thing we could do is take the relevant time period τ (e.g., one year in this case) and divide it into n equal subperiods, and treat the number of deaths in each as a Bernoulli random variable (and thus use the binomial distribution to model the total number of deaths). Why is this not an optimal solution?

if n is fixed (e.g. $n=12$) then the number of deaths in sub period can exceed 1. it's therefore not optimal to model Bernoulli.

Let the number of subperiods $n \to \infty$ such that the probability of observing an event $p \to 0$ and such that $np \to \lambda$, a constant. Under these conditions, the binomial distribution transforms into the *Poisson distribution*, or what Bortkiewicz called the *law of small numbers*:[a]

Start with a Bin(n, p).

$$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

$\boxed{\lambda = np} \Rightarrow$

$$= \frac{n!}{y!(n-y)!} \cdot \left(\frac{\lambda}{n}\right)^y \cdot \left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$= \frac{n!}{(n-y)!} \cdot \frac{\lambda^y}{y!\, n^y} \cdot \left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$= \boxed{\frac{n\cdot(n-1)\cdots(n-y+1)}{n^y}} \cdot \frac{\lambda^y}{y!} \left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$\underbrace{\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-y+1}{n}}$$

$\downarrow$ when $n \to \infty$, y fixed.

1

$$= \frac{\lambda^y}{y!} \left(1-\frac{\lambda}{n}\right)^{n-y}$$

$$= \frac{\lambda^y}{y!} \boxed{\left(1-\frac{\lambda}{n}\right)^n} \cdot \boxed{\frac{1}{\left(1-\frac{\lambda}{n}\right)^y}} \to 1$$

when $n \to \infty$ y fixed.

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n$$

$$= \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}\cdot\lambda}$$

let $x = \frac{n}{\lambda}$

$$= \lim_{x\to\infty} \left(1-\frac{1}{x}\right)^{x\cdot\lambda} \boxed{= e^{-\lambda}}$$

$$\lim_{x\to\infty} \left(1-\frac{1}{x}\right)^x = \frac{1}{e}$$

$$\lim_{n\to\infty} \left(1-\frac{1}{n}\right)^n = \frac{1}{e}$$

[a]You would not have to recreate this derivation on Test 1. Just remember that one can derive the Poisson pmf given the binomial pmf.

$$= \frac{\lambda^y}{y!} e^{-\lambda}$$

whenever $\boxed{p(y) = \frac{\lambda^y}{y!} e^{-\lambda}}$ (poisson dist) $Y \sim Pois(\lambda)$.

Not surprisingly, given what you've learned up to now, the Poisson distribution can be used to model the data of the *Poisson process*, which is a continuous-time analog of the discrete-time Bernoulli process. We will not go into the details of Poisson processes in this course; simply realize that a Poisson process is memoryless, like a Bernoulli process.[a]

The Poisson Distribution

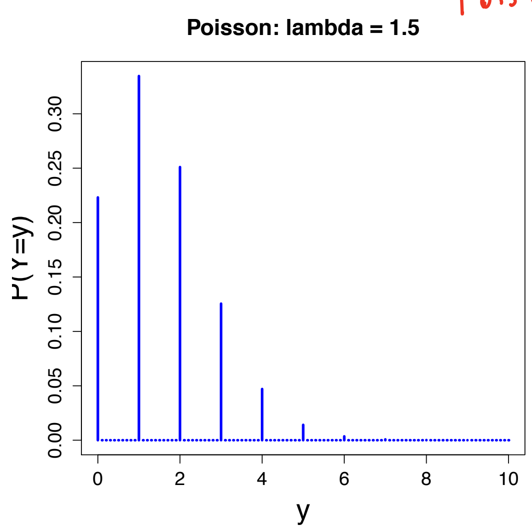
Notation: $Y \sim \text{Pois}(\lambda)$ λ: $\boxed{\text{expected total number of events}}$ (not rate of events!)

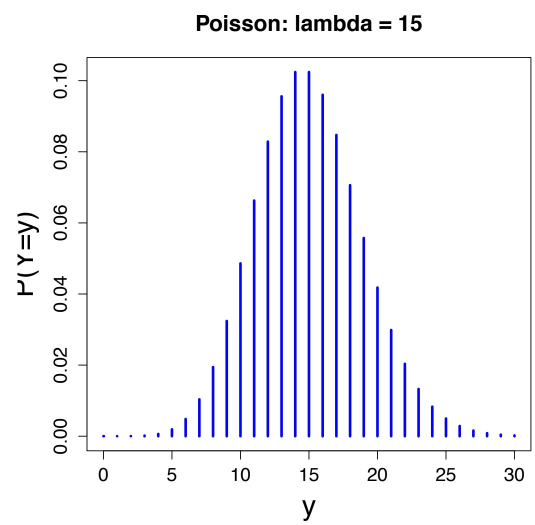
PMF: $p(y) = \frac{\lambda^y}{y!} e^{-\lambda}$ $y \in \mathbb{N}$ (including 0)

Expected Value: $E[Y] = \mu = \lambda$ Variance: $V[Y] = \sigma^2 = \lambda$

R Functions: dpois(y,lambda) (PMF)
ppois(y,lambda) (CDF)
rpois(k,lambda) (Simulation of k Poisson r.v.'s)

$\mathrm{Pois}(\underbrace{\lambda}_{=1.5})$





We derive the expression $E[Y] = \lambda$:

$$E[Y] = \sum_{y=0}^{\infty} y \cdot p(y) = \sum_{y=0}^{\infty} y \cdot \left(\frac{\lambda^y}{y!} e^{-\lambda}\right) = 0 + \sum_{y=1}^{\infty} y \frac{\lambda^y}{y!} e^{-\lambda}$$

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

let $x = y-1$

pmf for Poisson.

$\boxed{V[Y] = \lambda}$ offline.

$1 = \sum_{x=0}^{\infty} p(x)$ $X \sim Pois(\lambda)$.

---

[a] As an aside, the inability of a typical person to appreciate the memoryless-ness of the Bernoulli and Poisson processes leads to such things as believing in a "law of averages" or indulging in the "gambler's fallacy."

→ **EXAMPLE.** Wackerly 7, Exercise 3.123

3.123 The random variable Y has a Poisson distribution and is such that $p(0) = p(1)$. What is $p(2)$?

$Y \sim \text{Pois}(\lambda)$. $p(0) = p(1)$.

$p(0) = e^{-\lambda}$

$p(1) = \lambda \cdot e^{-\lambda}$

$p(0) = p(1)$. $\Downarrow$ $\lambda = 1$

$p(2) = \frac{\lambda^2}{2!} e^{-\lambda} = \frac{e^{-1}}{2}$

→ **EXAMPLE.** Wackerly 7, Exercise 3.129

***3.129** Refer to Exercise 3.128. How long can the attendant's phone call last if the probability is at least .4 that no cars arrive during the call?

ex 3.128

Cars arrive at toll according to poisson process, with average of 80 cars per hour

Y: number of cars arriving at toll within one-hour

$Y \sim \text{Pois}(\lambda = 80)$

$\lambda = \mathbb{E}[Y] = 80$

X: number of cars arriving at toll with t-minute.

$X \sim \text{Pois}\left(\boxed{\lambda = \frac{4}{3} t}\right)$.

$\lambda = \mathbb{E}[X] = \frac{80}{60} \cdot t$

target is to find t: s.t. $P(X=0) \geq 0.4$.

$\downarrow$

$e^{-4/3 t} \geq 0.4$.

$\Rightarrow$ $t \leq -\frac{3}{4} \log(0.4)$

$= 0.687$ minutes

→ **EXAMPLE.** Wackerly 7, Exercise 3.133

3.133 Assume that the tunnel in Exercise 3.132 is observed during ten two-minute intervals, thus giving ten independent observations $Y_1, Y_2, \ldots, Y_{10}$, on the Poisson random variable. Find the probability that $Y > 3$ during at least one of the ten two-minute intervals.

ex 3.132

Cars entering a tunnel with mean per 2-minute is equal to one

Y: number of cars entering tunnel during 2-minute.

$Y \sim \text{Pois}(\lambda = 1)$

$Y_1\ Y_2 \cdots Y_{10}$ independent r.v. $\text{Pois}(1)$.

$P(\text{at least one of } \{Y_1 \cdots Y_{10}\} \text{ that } Y_i > 3)$.

$$P(Y_i > 3) = 1 - P(Y_i = 0) - P(Y_i = 1) - P(Y_i = 2) - P(Y_i = 3)$$

$$= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2!} e^{-\lambda} - \frac{\lambda^3}{3!} e^{-\lambda}$$

$$= 1 - e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!}\right]$$

$\boxed{\text{use } \lambda = 1}$ $= 0.019$.

X: number of r.v. that has value > 3. $X \sim \text{Bin}(10, 0.019)$.

$P(X \geq 1) = 1 - P(X=0) = 1 - \binom{\cdot}{0}(0.019)\cdot(1-0.019) = 0.115$

We will conclude this set of notes by discussing how the Law of Total Probability manifests itself outside the realm of sample spaces (i.e., the realm of Chapter 2).

To remind you, given an event A and a splitting of the sample space $\mathcal{S}$ into n disjoint regions $\{B_1, B_2, \ldots, B_n\}$, we use the LoTP to express $P[A]$ as the sum of n conditional probabilities, each weighted by the "size" of the region B_i:

$$P[A] = \sum_{i=1}^{n} P[A|B_i]P[B_i] \, .$$

$\to P(A \cap B_i)$.

Let's assume here that instead of events, we are dealing with discrete random variables, so that instead of $P[A]$, we are interested in computing $P[Y = y]$. We can write

$$\boxed{P(Y=y)} = \sum_x P(Y=y, X=x) = \sum_x P(X=x)\,P(Y=y \mid X=x)$$

pmf   discret.   pmf

$$\boxed{= \sum_x p(x)\cdot p(y|x).}$$

→ **EXAMPLE.** In your pocket is a random number of coins N, where $N \sim \text{Poisson}(\lambda)$. You toss each coin once, with heads showing with probability p each time. Show that the total number of heads has the Poisson distribution with parameter λp.

H: the number of heads when tossing N coins.

$H \mid N = n \sim \text{Bin}(n, p)$.

$$\boxed{P(H=h)} = \sum_{n \geq h} P(N=n)\,P(H=h \mid N=n).$$

$$= \sum_{n=h}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \cdot \binom{n}{h} p^h (1-p)^{n-h}$$

$$= \sum_{n=h}^{\infty} \frac{\lambda^n}{\cancel{n!}} e^{-\lambda} \cdot \frac{\cancel{n!}}{h!\,(n-h)!} p^h (1-p)^{n-h}.$$

$$= \sum_{n=h} \boxed{\lambda^h \cdot \lambda^{n-h}} \cdot \boxed{e^{-\lambda p} \cdot e^{-\lambda(1-p)}} \cdot \frac{1}{h!\,(n-h)!} p^h (1-p)^{n-h}$$

$$= \frac{\lambda^h p^h}{h!} e^{-\lambda p} \cdot \sum_{n=h}^{\infty} \cdot \frac{\lambda^{n-h}(1-p)^{n-h}}{(n-h)!} e^{-\lambda(1-p)}$$

$\| \; 1$

$Y \sim \text{Pois}(\lambda p)$

$$P(Y=y) = \frac{(\lambda p)^y}{y!} e^{-\lambda p} \qquad = \frac{(\lambda p)^h}{\;} e^{-\lambda p}$$

$h!$

$$\sum_{n=h}^{\infty} \frac{(\lambda(1-p))^{n-h}}{(n-h)!} e^{-\lambda(1-p)} = 1$$

$\sim \text{Pois}(\lambda p)$

let's define $\alpha := \lambda(1-p)$

$x := n-h.$

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

pmf Pois(α).