5 Probability: the mathematics of randomness

The theory of probability is the mathematician's description of random variation. This chapter introduces enough probability to serve as a minimum background for making formal statistical inferences.

Recall this overview:

ch 2. $\boxed{\text{collect Engineering Data}}$ → ch. 3 & 4 $\boxed{\text{summarize data}}$ / ch. 5 $\boxed{\text{draw inference}}$ ⟹ $\boxed{\text{Interpretation}}$

sample $\xrightarrow{\text{"infer"}}$ population

5.1 (Discrete) random variables

The concept of a random variable is introduced in general terms and the special case of discrete data is considered.

Recall:

Discrete data — measurements are separate points (eg. # of pages in a book)

5.1.1 Random variables and distributions

It is helpful to think of data values as subject to chance influences. Chance is commonly introduced into the data collection process through

1. Random sampling technique ⟹
2. Measurement error causes
3. changes in system conditions

POPULATION μ → samples $\bar{x}_1$, $\bar{x}_2$, …, $\bar{x}_n$

All possible $\bar{x}$s are not necessarily equal to eachother or equal to μ

capital letters used to stand for R.V.S (random variables)

Definition 5.1. A *random variable* is a quantity that (prior to observation) can be thought of as dependent on chance phenomena.

X = the value of a coin toss (heads or tails)

Z = the amount of torque required to loosen the next bolt

T = the time you'll have to wait for next bus home

N = the number of defective widgets in a manufacturing process in a day

S = the number of unprovoked shark attacks off coast of Florida next year

Definition 5.2. A *discrete random variable* is one that has isolated or separated possible values (rather than a continuum of available outcomes).

Definition 5.3. A *continuous random variable* is one that can be idealized as having an entire (continuous) interval of numbers as its set of values.

Discrete: X, N, S

continuous: T, Z

Example 5.1 (Roll of a die)**.**

List of possible values:

X = roll of 6-sided fair die - 1, 2, 3, 4, 5, 6

Y = roll of 6-sided unfair die - 1, 2, 3, 4, 5, 6

How to distinguish btwn X and Y?

probability of occurrence!

Definition 5.4. To specify a *probability distribution* for a random variable is to give its set of possible values and (in one way or another) consistently assign numbers between 0 and 1 - called *probabilities* - as measures of the likelihood that the various numerical values will occur

Example 5.2 (Roll of a die, cont'd).

we expect a fair dice to land on the number 3 roughly one out of every 6 tosses.

$P[x=3] = \frac{1}{6}$

suppose the unfair dice is weighted so that the number 3 only lands one out of every 22 tosses.

$P[Y=3] = \frac{1}{22}$

lower case stands for particular value of RV

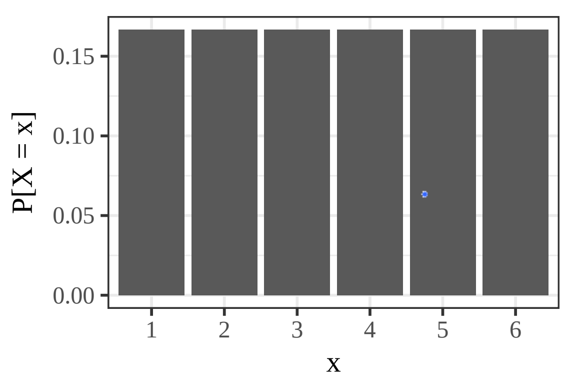
probability that $\underline{\underline{X}}$ R.V. takes value x

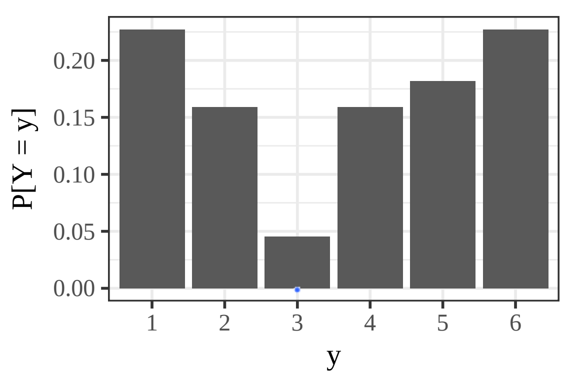
x	1	2	3	4	5	6
$P[X=x]$	1/6	1/6	1/6	1/6	1/6	1/6

Fair dice

y	1	2	3	4	5	6
$P[Y=y]$	5/22	7/44	1/22	7/44	2/11	5/22

unfair dice

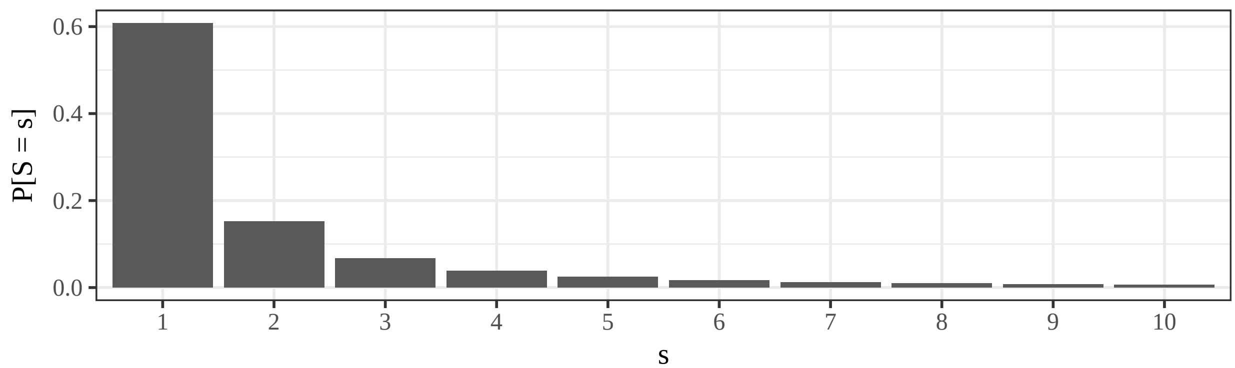




Example 5.3 (Shark attacks)**.** Suppose S is the number of provoked shark attacks off FL next year. This has an infinite number of possible values. Here is one possible (made up) distribution: S can take on values 1, 2, 3, 4

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{1}{1^2} + \frac{1}{2^2} + \cdots = \frac{\pi^2}{6}$$

s	1	2	3	$\cdots$	k	$\cdots$
$P[S=s]$	$\frac{6}{\pi^2}$	$\frac{1}{2^2}\frac{6}{\pi^2}$	$\frac{1}{3^2}\frac{6}{\pi^2}$	$\cdots$	$\frac{1}{k^2}\frac{6}{\pi^2}$	$\cdots$



All the probabilities for a R.V. must add to 1.

5.1.2 Probability mass functions and cumulative distribution functions

The tool most often used to describe a discrete probability distribution is the *probability mass function.*

Definition 5.5. A *probability mass function (pmf)* for a discrete random variable X, having possible values $x_1, x_2, \ldots,$ is a non-negative function $f(x)$ with $f(x_1) = P[X = x_1]$, the probability that X takes the value x_1.

we can also write f_X for the pmF of X and f_S for the pmF of S.

Properties of a mathematically valid probability mass function:

1. $f(x) \geq 0$ for all x (positive probabilities)

2. $\sum_x f(x) = 1$ (sum to 1)

A probability mass function $f(x)$ gives probabilities of occurrence for individual values. Adding the appropriate values gives probabilities associated with the occurrence of multiple values. $P[X \text{ is } x_1 \text{ or } x_2] = P[X = x_1] + P[X = x_2]$

Example 5.4 (Torque)**.** Let Z = the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

z	11	12	13	14	15	16	17	18	19	20
$f(z)$	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

Calculate the following probabilities:

$$\begin{aligned} P(Z \leq 14) &= P[Z=11 \text{ or } Z=12 \text{ or } Z=13 \text{ or } Z=14] \\ &= P[Z=11] + P[Z=12] + P[Z=13] + P[Z=14] \\ &= f(11) + f(12) + f(13) + f(14) \\ &= 0.03 + 0.03 + 0.03 + 0.06 = 0.15 \end{aligned}$$

$$\begin{aligned} P(Z > 16) &= P[Z=17 \text{ or } Z=18 \text{ or } Z=19 \text{ or } Z=20] \\ &= f(17) + f(18) + f(19) + f(20) \\ &= 0.12 + 0.20 + 0.15 + 0.03 = 0.5 \end{aligned}$$

$$\begin{aligned} P(Z \text{ is even}) &= P[Z=12 \text{ or } Z=14 \text{ or } Z=16 \text{ or } Z=18 \text{ or } Z=20] \\ &= f(12) + f(14) + f(16) + f(18) + f(20) \\ &= 0.03 + 0.06 + 0.09 + 0.20 + 0.03 = 0.41 \end{aligned}$$

$$\begin{aligned} P(Z \text{ in } \{15, 16, 18\}) &= P[Z=15 \text{ or } Z=16 \text{ or } Z=18] \\ &= f(15) + f(16) + f(18) \\ &= 0.26 + 0.09 + 0.2 = 0.55 \end{aligned}$$

Another way of specifying a discrete probability distribution is sometimes used.

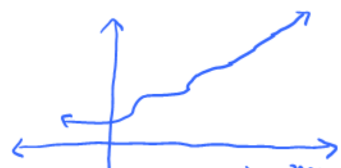
Definition 5.6. The *cumulative probability distribution (cdf)* for a random variable X is a function $F(x)$ that for each number x gives the probability that X takes that value or a smaller one, $F(x) = P[X \leq x]$.

Since (for discrete distributions) probabilities are calculated by summing values of $f(x)$,

$$F(x) = P[X \leq x] = \sum_{y \leq x} f(y)$$

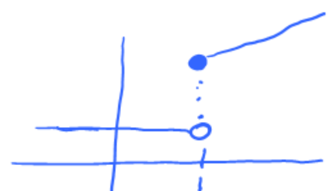
Properties of a mathematically valid cumulative distribution function:

1. $F(x) \geq 0$ for all real numbers x
2. F is monotonically increasing (when x increases, y either stays same or increases)
3. $F(x)$ is right continuous
4. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

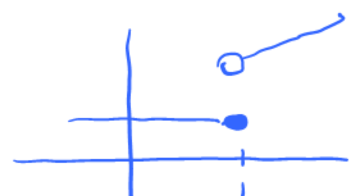


monotonically increasing

NOT monotonically increasing



right-continuous @ $x = 1$



left-continuous @ $x = 1$

In the discrete case, the graph $F(x)$ will be stair-step graph with jumps located at possible values and equal in size to the probabilities associated with these values.

Example 5.5 (Torque, cont'd)**.** Let Z = the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

z	11	12	13	14	15	16	17	18	19	20
$F(z)$	0.03	0.06	0.09	0.15	0.41	0.50	0.62	0.82	0.97	1

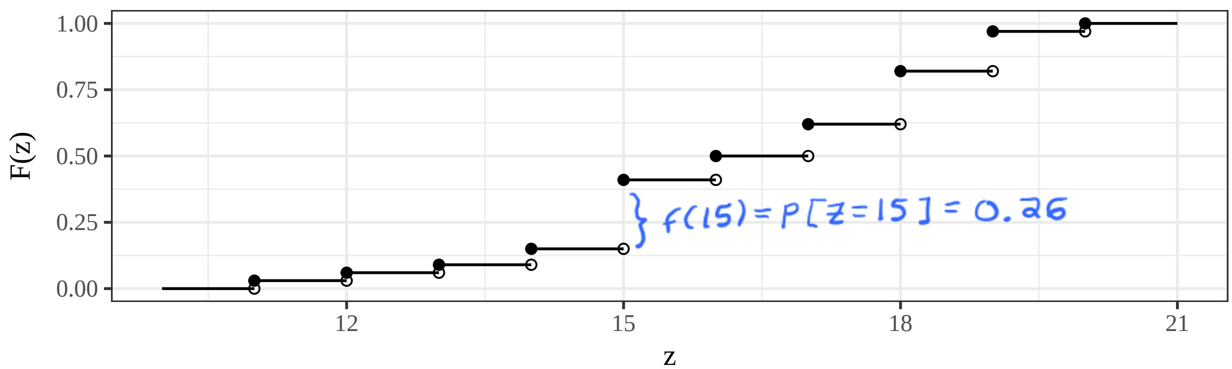


Figure 1: Cdf function for torques.

Calculate the following probabilities using the **cdf only**:

$F(10.7)$ $= P[Z \le 10.7] = P[Z \le 10] = 0$

$P(Z \le 15.5)$ $= P[Z \le 15] = 0.41$

$P(12.1 < Z \le 14)$ $= P[Z = 13 \text{ or } Z = 14]$

$= f(14) + f(13) = \underbrace{[f(14) + f(13) + \cancel{f(12)} + \cancel{f(11)}]}_{P[Z \le 14]} - \underbrace{[\cancel{f(12)} + \cancel{f(11)}]}_{P(Z \le 12)}$

$= P[Z \le 14] - P[Z \le 12]$

$= F[14] - F[12] = 0.15 - 0.06 = 0.09$

$P(15 \le Z < 18)$ $= P[Z = 15, 16, \text{ or } 17]$

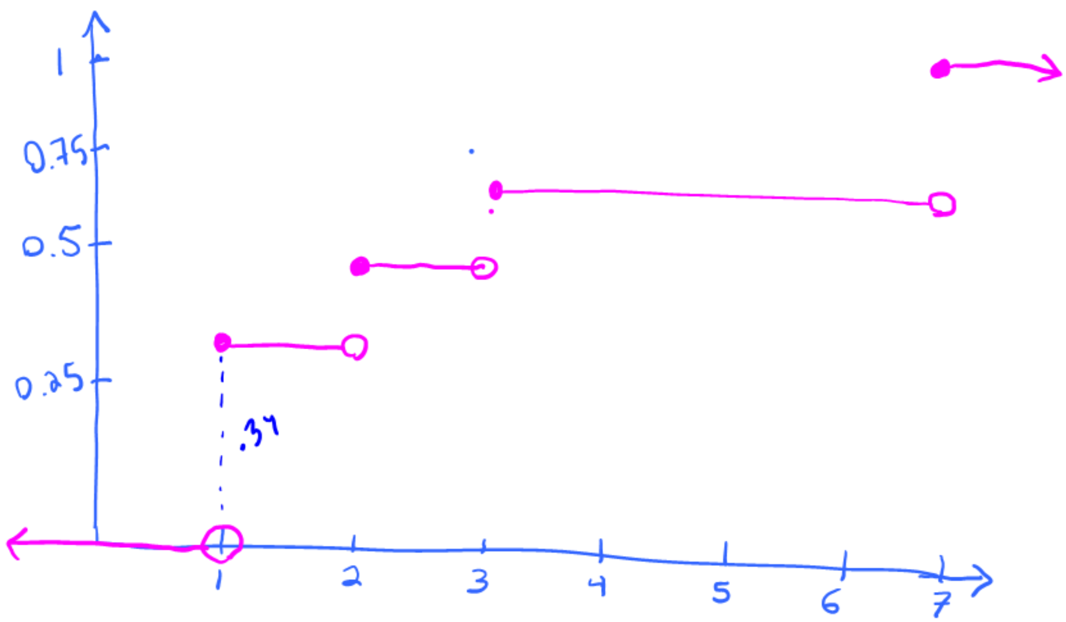
$= P[Z \le 17] - P[Z \le 14]$

$= F(17) - F(14) = 0.62 - 0.15 = 0.47$

Example 5.6. Say we have a random variable Q with pmf:

q	$f(q)$	$F(q)$
1	0.34	0.34
2	0.1	0.44
3	0.22	0.66
7	0.34	1

Draw the cdf.



5.1.3 Summaries

Almost all of the devices for describing relative frequency (empirical) distributions in Ch. 3 have versions that can describe (theoretical) probability distributions.

1. probability histograms (histograms based on theoretical probabilities)
2. mean (measure of location)
3. variance (measure of spread)

Definition 5.7. The *mean* or *expected value* of a discrete random variable X is

$$\mathrm{E}X = \sum_x x f(x)$$

$$\| \quad E(X) \longleftarrow \text{sometimes denoted } \mu$$

EX is the weighted average of all possible values of X, weighted by their probabilities.

EX is the mean of the distribution of X.

Example 5.7 (Roll of a die, cont'd)**.** Calculate the expected value of a toss of a fair and unfair die.

x	1	2	3	4	5	6
$f(x) = P[X = x]$	1/6	1/6	1/6	1/6	1/6	1/6

y	1	2	3	4	5	6
$P[Y = y]$	5/22	7/44	1/22	7/44	2/11	5/22

Fair die:

$$EX = \overset{x\,f(x)}{1\left(\tfrac{1}{6}\right)} + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) + 6\left(\tfrac{1}{6}\right)$$

$$= 3.5$$

Unfair die:

$$EY = 1\left(\tfrac{5}{22}\right) + 2\left(\tfrac{7}{44}\right) + 3\left(\tfrac{1}{22}\right) + 4\left(\tfrac{7}{44}\right) + 5\left(\tfrac{2}{11}\right) + 6\left(\tfrac{5}{22}\right)$$

$$= 3.5909$$

The average roll of the unfair die is 3.5909, higher than the fair die.

Example 5.8 (Torque, cont'd)**.** Let $Z =$ the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

z	11	12	13	14	15	16	17	18	19	20
$f(z)$	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

Calculate the expected torque required to loosen the next bolt.

$$EZ = 11(.03) + 12(.03) + 13(.03) + \dots + 20(.03)$$

$$= 16.35$$

The average torque required to loosen the next bolt is 16.35 units.

Definition 5.8. The *variance* of a discrete random variable X is

$$\mathrm{Var}X = \overbrace{\sum_x \underbrace{(x - \mathrm{E}X)^2}\, \underbrace{f(x)}}^{= E[(X-EX)^2]} = \underbrace{\overbrace{\sum_x x^2 f(x)}^{E(X^2)} - (\mathrm{E}X)^2}_{\text{computationally easier}}. \geq 0 \text{ always!}$$

sometimes $\sigma^2 = \mathrm{Var}(X)$

The *standard deviation* of X is $\sqrt{\mathrm{Var}X}$.

$SD(X) = \sigma$
(notation)

The variance is the average squared deviation of a random variable from its mean.

Example 5.9. Say we have a random variable Q with pmf:

q	$f(q)$
1	0.34
2	0.1
3	0.22
7	0.34

Calculate the variance and the standard deviation.

Long way:

$EQ = 1(.34) + 2(.1) + 3(.22) + 7(.34) = 3.58$

$\to \operatorname{Var} Q = (1-3.58)^2(.34) + (2-3.58)^2(.1) + (3-3.58)^2(.22)$
$+ (7-3.58)^2(.34)$
$= 6.56$

Short way:

$E(Q^2) = \sum_q q^2 f(q)$
$= 1(.34) + 4(.1) + 9(.22) + 49(.34)$
$= 19.38$

$\operatorname{Var} Q = E(Q^2) - (EQ)^2 \leftarrow$
$= 19.38 - (3.58)^2$
$= 6.56$

Example 5.10 (Roll of a die, cont'd)**.** Calculate the variance and standard deviation of a roll of a fair die.

$$EX = 3.5$$

$$EX^2 = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36)$$
$$= 15.17$$

$$\text{Var}\, X = EX^2 - (EX)^2$$
$$= 15.17 - (3.5)^2$$
$$= 2.92$$

$$SD(X) = \sqrt{\text{Var}X} = 1.71$$

5.1.4 Special discrete distributions

Discrete probability distributions are sometimes developed from past experience with a particular physical phenomenon.

On the other hand, sometimes an easily manipulated set of mathematical assumptions having the potential to describe a variety of real situations can be put together.

One set of assumptions is that of independent identical success-failure trials where

1. There is a constant chance of success on each repetition of the scenario (probability p)
2. The repetitions are independent - i.e. knowing the outcome of any one of them does not change assessments of chance related to any others.

Consider a variable

$X =$ the number of successes in n independent identical success-failure trials

Definition 5.9. The *binomial(n, p) distribution* is a discrete probability distribution with pmf

parameters

$$f(x) = \begin{cases} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

name comes from the binomial theorem

for n a positive integer and $0 < p < 1$.

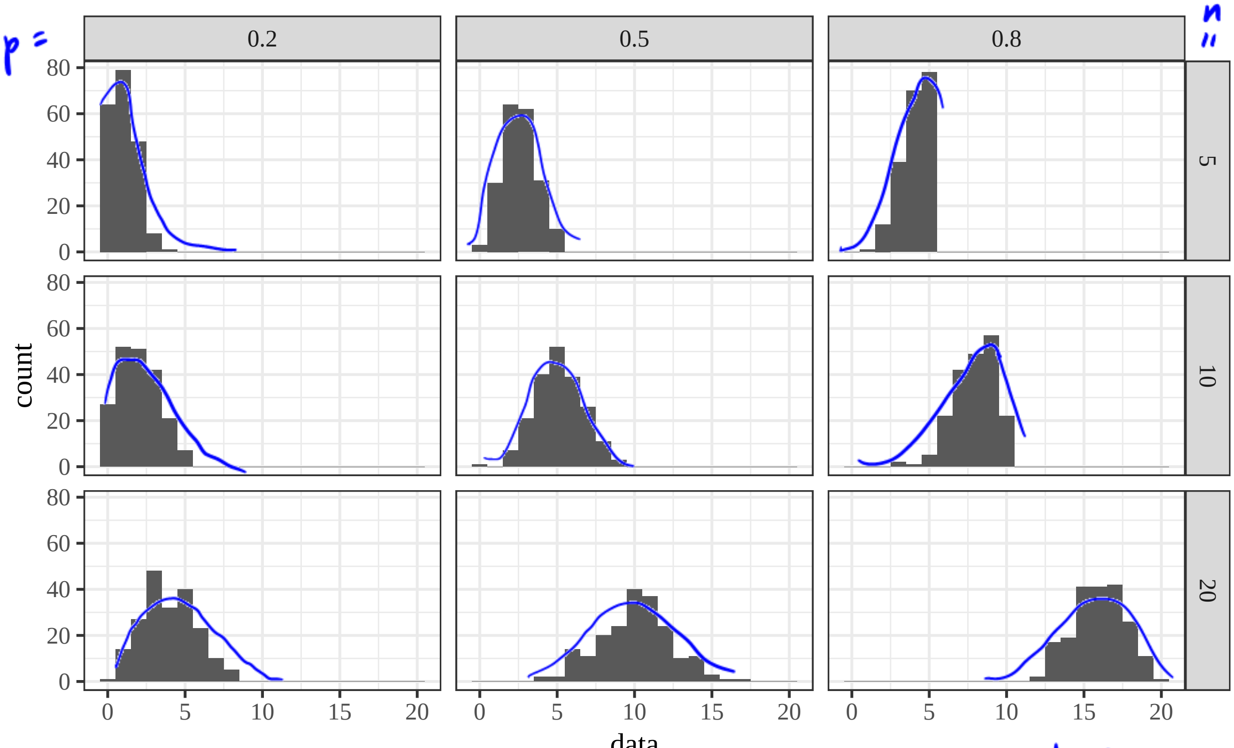
Examples that could follow a binomial(n, p) distribution:

- Number of conforming pellets in a batch of $n=50$ pellets made by a machine.
- Number of runs of the same chemical process w/ % yield above 80%, given you ran the process 1000 times.
- Number of rivets that fail in a boiler of $n=25$ rivets within 3 years of operation.

Note: "success" doesn't have to be good.

p =

n =



skewness decreases as n increases

for different param. values, n & p, the binomial(n,p) has a different shape.

p < .5 ⇒ right skewed

p = .5 ⇒ symmetric

p > .5 ⇒ left skewed

For X a binomial(n, p) random variable,

sometimes written $\binom{n}{x}$ "n choose x"

$$\mu = \mathrm{E}X = \sum_{x=0}^{n} x \underbrace{\left(\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}\right)}_{P(X=x)} = np$$

$$\sigma^2 = \mathrm{Var}X = \sum_{x=0}^{n} (x-np)^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}) = np(1-p)$$

Example 5.11 (10 component machine)**.** Suppose you have a machine with 10 independent components in series. The machine only works if all the components work. Each component succeeds with probability $p = 0.95$ and fails with probability $1 - p = 0.05$.

$p + 1 - p = 1$ (i.e. the probability of succeeding or failing $= 1$)

Let Y be the number of components that succeed in a given run of the machine. Then

$$Y \sim \text{Binomial}(n = 10, p = 0.95)$$

Question: what is the probability of the machine working properly?

$$\boxed{1} \to \boxed{2} \to \boxed{3} \to \cdots \to \boxed{10}$$

$p(Y=x)$

$$p(\text{machine working}) = p(Y = \overset{x}{10})$$

$$\frac{\overset{10}{n!}}{\underbrace{(\overset{10}{n} - \overset{10}{x})}_{0}!\,x!}$$

$$= \frac{10!}{0!\,10!} p^{10} (1-p)^{\cancel{10-10}}$$

$$= p^{10}$$

$$= .95^{10}$$

$$= 0.5987 \text{ (not very reliable)}$$

Example 5.12 (10 component machine, cont'd)**.** What if I arrange these 10 components in parallel? This machine succeeds if at least 1 of the components succeeds.

What is the probability that the new machine succeeds?

$$
\begin{aligned}
P(\text{new machine succeeding}) &= 1 - P(\text{new machine failing}) \\
&= 1 - P(\text{all components failing}) \\
&= 1 - \underbrace{P(Y=0)} \\
&= 1 - \frac{\cancel{10!}}{0!\,\cancel{10!}} p^{\cancel{0}} (1-p)^{10-0} \\
&= 1 - (.05)^{10} \\
&\approx 1
\end{aligned}
$$

Example 5.13 (10 component machine, cont'd)**.** Calculate the expected number of components to succeed and the variance.

$EY = n \cdot p = 10 \cdot .95 = 9.5$ so the number of components to succeed on average per run is 9.5

$\text{Var}\,Y = np(1-p) = 10 \cdot .95(1-.95) = .475$

$SD(Y) = \sqrt{\text{Var}\,Y} = \sqrt{np(1-p)} = .689$

Consider a variable

$$X = \text{ the number of trials required to first obtain a success result}$$

1 param

Definition 5.10. The *geometric(p) distribution* is a discrete probability distribution with pmf

$$f(x) = \begin{cases} p(1-p)^{x-1} & x = 1, \dots \\ 0 & \text{otherwise} \end{cases}$$

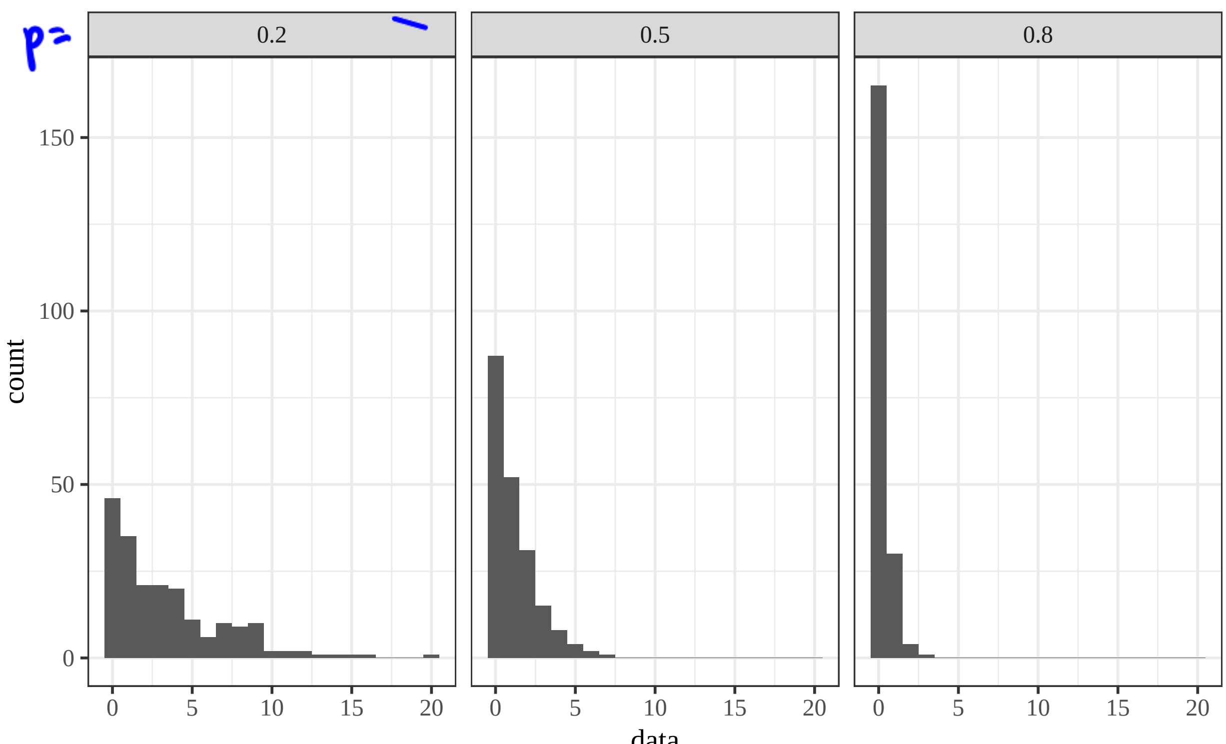
comes from geometric series.

for $0 < p < 1$.

Examples that could follow a ~~binomial(n, p)~~ geometric(p) distribution:

- Number of rolls of a fair die until you land on a 5
- Number of shipments of raw material until you get a defective one
- Number of hexamine pellets you make before you make one that doesn't conform
- Number of buses that come before yours.

p=



probability decay as x increases (at a faster rate as p increases)

For X a geometric(p) random variable,

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

$$\sigma^2 = \mathrm{Var}X = \sum_{x=1}^{\infty} \left(x - \frac{1}{p}\right)^2 p(1-p)^{x-1} = \frac{1-p}{p^2}$$

Cdf derivation:

$f(x) = p(1-p)^{x-1}$ for $x = 1, 2, \ldots$ makes intuitive sense. For X to take the value x, there must be $x-1$ consecutive failure results followed by 1 success.

there are $x-1$ terms $(1-p)$ – failure probability
and 1 term p – success probability.

Another way to think about this:

$$1 - F(x) = 1 - P(X \le x)$$
$$= P(X > x)$$
$$= P(x \text{ failure outcomes in } x \text{ trials})$$

Binomial probability. Binomial (x, p).

$$= \frac{x!}{x!(x-x)!} p^0 (1-p)^x$$
$$= (1-p)^x$$

$$\Rightarrow 1 - F(x) = (1-p)^x$$
$$F(x) = 1 - (1-p)^x \quad x = 1, 2, \ldots$$

Example 5.14 (NiCad batteries). An experimental program was successful in reducing the percentage of manufactured NiCad cells with internal shorts to around 1%

Let T be the test number at which the first short is discovered $\Rightarrow T \sim Geom(p)$

$p = .01$

Calculate

P(1st or 2nd cell tested has the 1st short)

$$= P(T=1 \text{ or } T=2)$$
$$= f(1) + f(2)$$
$$= p(1-p)^0 + p(1-p)$$
$$= .01 + .01(1-.01)$$
$$= .02$$

P(at least 50 cells tested w/o finding a short) $= P(T>50)$

$$= 1 - P(T \leq 50)$$
$$= 1 - F(50)$$
$$= 1 - (1-(1-p)^{50})$$
$$= (1-p)^{50}$$
$$= (1-.01)^{50} = .61$$

Calculate the expected test number at which the first short is discovered and the variance in test numbers at which the first short is discovered.

$$ET = \frac{1}{p} = \frac{1}{.01} = 100$$ tests for the first short to appear, on average.

$$\text{Var}\, T = \frac{1-p}{p^2} = \frac{1-.01}{(.01)^2} = 9900$$

$$SD(T) = \sqrt{\text{Var}\, T} = 99.4987$$

It's often important to keep track of the total number of occurrences of some relatively rare phenomenon.

Consider a variable

$X =$ the count of occurences of a phenomenon across a specified interval of time or space

fixed

1 parameter

Definition 5.11. The *Poisson*(λ) *distribution* is a discrete probability distribution with pmf

$$f(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

for $\lambda > 0$.

These occurrences must:

1. be independent
2. be sequential in time (no two occurences at once)
3. occur at the same constant rate, λ

λ, the rate parameter, is the expected number of occurences in the specified interval of time or space.

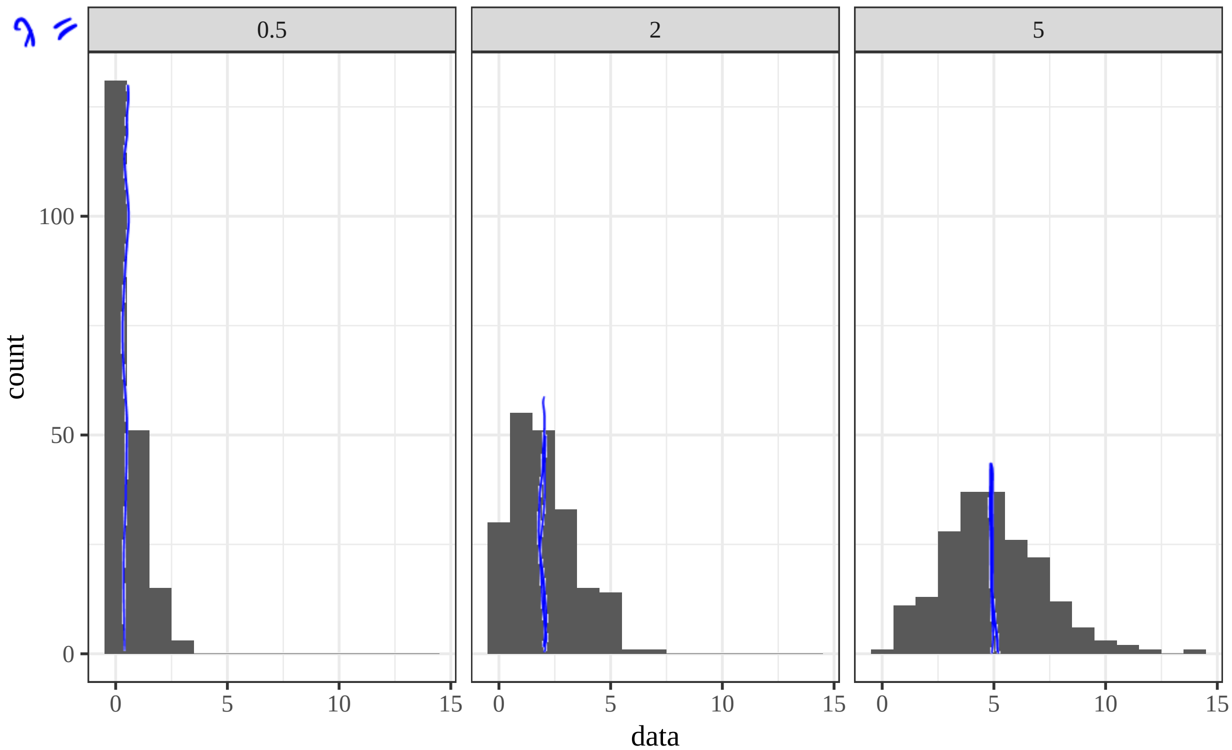
Examples that could follow a Poisson(λ) distribution:

Y is the number of shark attacks off the coast of CA next year, $\lambda = 100$ attacks/year

Z the number of shark attacks off the coast of CA next month
$\lambda = 100/12 = 8.333$ attacks/month

N the number of β particles emitted from a small bar of plutonium, registered by a Geiger counter in a minute, $\lambda = 459.21$ particles/minute.

J is the number of particles per hour, $\lambda = 27552.6$ particles/hour.



right skewed with peak near λ

For X a Poisson(λ) random variable,

$$\mu = \mathrm{E}X = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda$$

$$\sigma^2 = \mathrm{Var}X = \sum_{x=0}^{\infty} (x-\lambda)^2 \frac{e^{-\lambda}\lambda^x}{x!} = \lambda$$

Example 5.15 (Arrivals at the library)**.** Some students' data indicate that between 12:00 and 12:10pm on Monday through Wednesday, an average of around 125 students entered Parks Library at ISU. Consider modeling

125 students per 10 minutes

$M =$ the number of students entering the ISU library between 12:00 and 12:01pm next Tuesday

Model $M \sim \text{Poisson}(\lambda)$. What would a reasonable choice of λ be?

$\lambda = 125/10 = 12.5$ students per minute.

Under this model, the probability that between 10 and 15 students arrive at the library between 12:00 and 12:01 PM is:

$$P(10 \le M \le 15) = f(10) + f(11) + f(12) + f(13) + f(14) + f(15)$$

$$= \frac{e^{-12.5}(12.5)^{10}}{10!} + \frac{e^{-12.5}(12.5)^{11}}{11!} + \cdots + \frac{e^{-12.5}(12.5)^{15}}{15!}$$

$$= 0.6$$

Example 5.16 (Shark attacks)**.** Let X be the number of unprovoked shark attacks that will occur off the coast of Florida next year. Model $X \sim \text{Poisson}(\lambda)$. From the shark data at http://www.flmnh.ufl.edu/fish/sharks/statistics/FLactivity.htm, 246 unprovoked shark attacks occurred from 2000 to 2009. 246 attacks per 10 years

What would a reasonable choice of λ be?

$$\lambda = 246/10 = 24.6 \text{ attacks per year}$$

Under this model, calculate the following:

P[no attacks next year] $= P(X=0)$

$$= f(0) = \frac{e^{-24.6}(24.6)^0}{0!} \approx 2.67 \times 10^{-11}$$

P[at least 5 attacks]

$$= P(X \geq 5) = 1 - P(X<5) = 1 - [f(0)+f(1)+\dots+f(4)]$$

$$= 1 - e^{-24.6}\left[\frac{24.6^0}{0!} + \frac{24.6^1}{1!} + \frac{24.6^2}{2!} + \dots + \frac{24.6^4}{4!}\right) \approx .9999996$$

P[more than 10 attacks] $= P(X>10)$

$$= 1 - P(X \leq 10) = 1 - e^{-24.6}\left[\sum_{x=0}^{10} \frac{24.6^x}{x!}\right]$$

$$= 1 - e^{-24.6}[36266812] = 0.9992$$

Note $\sum_{i=0}^{10} \frac{24.6^i}{i!} = 36266812$