Binomial Random Variables

Counting the number of successes

Download the section 9 .Rmd handout to `STAT240/lecture/sect09-binomial`.

Material in this section is covered by Chapter 11 on the notes website.

We've seen examples of probability distributions.
How do we take a real-life process and figure out its distribution?

Real life is messy and complex. Options:

- Repeat the process infinity times
- Make some **reasonable assumptions**

What is the probability distribution of how many green lights I hit out of three? Let's assume:

- Lights are independent
- Each one has probability p of being green

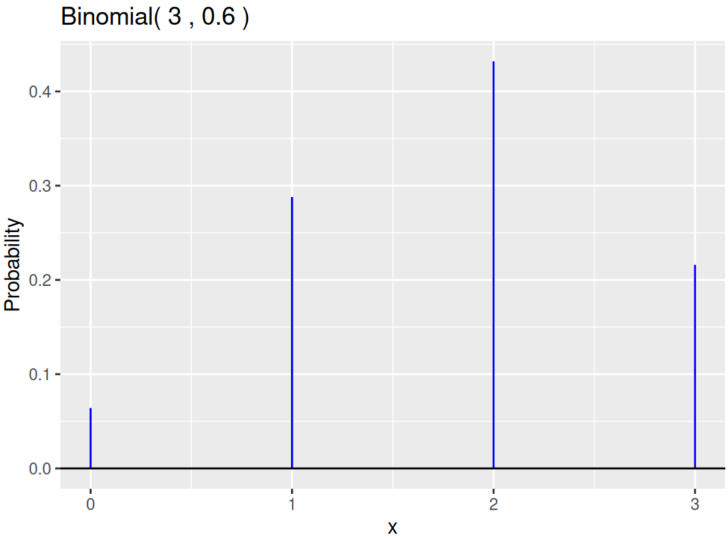
This simplification lets us calculate probabilities. Suppose each light has probability $p = 0.6$ of being green.

X = number of green lights.

Outcomes	x	$P(X = x)$
RRR	0	$(0.4)(0.4)(0.4) = 0.064$
	1	
GGR, GRG, RGG	2	$3(0.6)(0.6)(0.4) = 0.432$
	3	

Complete the pmf for X.

Outcomes	x	$P(X = x)$
RRR	0	$(0.4)(0.4)(0.4) = 0.064$
	1	
GGR, GRG, RGG	2	$3(0.6)(0.6)(0.4) = 0.432$
	3	


Binomial(3 , 0.6)
Probability
0.4
0.3
0.2
0.1
0.0
0
1
2
3
x

X in the traffic lights example is a count of "successes" in 3 tries.

X is a **binomial** RV. A binomial RV counts the number of times a desired outcome occurs in many tries. We say X has a "binomial distribution".

All binomial RVs share specific properties.

The lights (trials) have specific properties:

- **B**: they are binary (success, or failure)
- **I**: they are independent
- **N**: fixed sample size $n = 3$
- **S**: they have the same probability p

A binomial RV is the count of successes in a fixed number (n) of identical trials.

What if we counted how many green lights we hit, before seeing the first red light? Not binomial.

What if we also counted yellow lights? Not binomial.

Write $X \sim Binom(n, p)$.

- n: pre-determined number of trials
- p: individual success probability

So $X \sim Binom(3, 0.6)$ for the traffic lights.

The shape of the binomial distribution depends on n and p.

For the traffic lights,

$$P(X = 2) = 3(0.6)^2(0.4)^1$$

- 2 is the number of successes, 1 is the number of failures
- 0.6 is the sucess probability
- 0.4 is the failure probability
- 3 is the number of ways to have 2/3 successes.

In general,

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

- x is the number of successes, $n - x$ is the number of failures
- p is the sucess probability
- $1 - p$ is the failure probability
- $\binom{n}{x}$ is the number of ways to have x successes out of n.

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

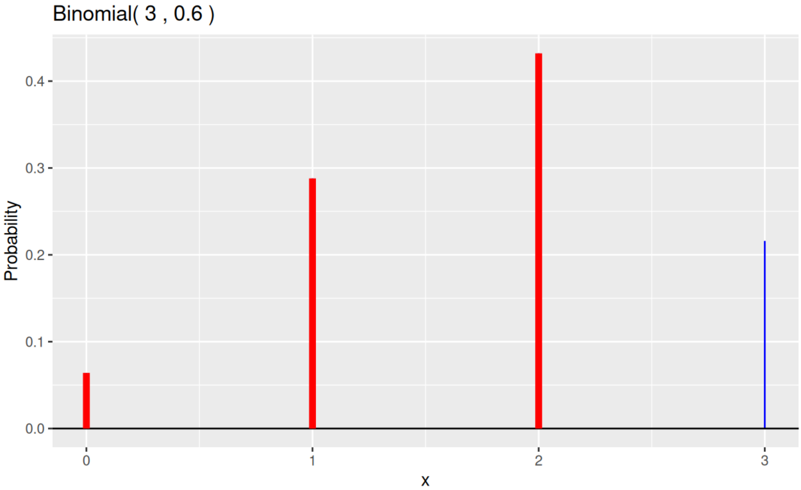
$n!$ is the product of numbers 1 to n.

Can be calculated with R's `choose` or `factorial`.

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

We can use the R command `dbinom` as a shortcut.

We can also use `pbinom` for a "less than or equal" (cumulative) probability.


Binomial(3 , 0.6)
Probability
0.0
0.1
0.2
0.3
0.4
0
1
2
3
x

Remember `pbinom` calculates area to the *left*, including the x value specified.

Let $Y \sim Binom(8, 0.3)$.

- What is $P(Y > 4)$?
- What is $P(Y \geq 4)$?
- What is $P(Y \leq 3 \leq 6)$?

We have shortcut formulas for the mean and variance of $X \sim Binom(n, p)$.

$$\text{mean } \mu = np, \quad \text{var } \sigma^2 = np(1-p)$$

We will hit $3(0.6) = 1.8$ green lights out of 3, on average.

The variance in the number of green lights is $3(0.6)(0.4) = 0.72$.

The mean is the "average" or "typical" values of the binomial process.

Another way to quantify the RV is with a **percentile** or **quantile**.

Choose a probability p. The quantile for p subidvides the probability such that there is p probability to the left and $1 - p$ to the right.

In a population, 35% of individuals are below the 35th percentile, and 65 are above.

Percentiles work a bit differently here since a binomial process is discrete. How large is q before

$$P(Binom \leq q)$$

is at least p?

We can find binomial quantiles with `qbinom`. This is the inverse of `pbinom`.

- `pbinom`: input x value, output cumulative probability
- `qbinom`: input cumulative probability, output x value

Command	In	Out
`dbinom`	A value x	$P(X = x)$
`pbinom`	A value x	$P(X \leq x)$
`qbinom`	A probability p	q for $P(X \leq q) = p$

Let $X \sim Binom(90, 0.7)$.

- Find the mean μ and sd σ of X
- What is $P(X = \mu)$?
- What is $P(\mu - \sigma \leq X \leq \mu + \sigma)$?
- What are the 5th and 95th percentiles of X?