Binomial Random Variables

Counting the number of successes

Download the section 8 .Rmd handout to `STAT240/lecture/08-binomial`.

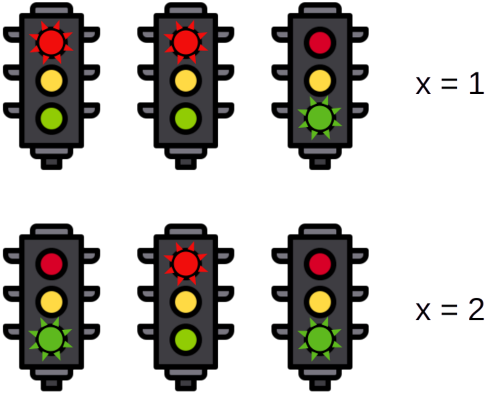
Material in this section is covered by Chapter 10 on the notes website.

How can we find the probability distribution of a real-life process?

Options:

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

Let X be the number of green lights I hit out of 3.



What is the distribution of X? Let's assume:

- Lights are independent
- Each has probability 0.6 of being green

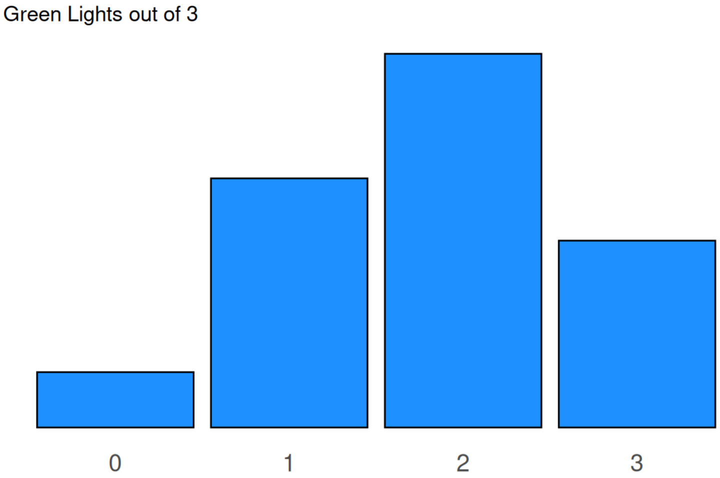
This simplification lets us calculate probabilities.

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 probability distribution of 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	


Green Lights out of 3
0
1
2
3

X is a count of “successes” in 3 tries.

X is a **binomial** RV. A binomial counts the number of times a desired outcome occurs in many tries.

We say X has a “binomial distribution”.

The lights (trials) have specific properties:

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

A binomial is the count of successes in n 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 distribution depends on n and p.

For the traffic lights,

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

- 2 successes, 1 failure
- 0.6 is the sucess probability
- 0.4 is the failure probability
- 3 is the number of orderings

In general,

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

- x successes, $n - x$ failures
- p is the sucess probability
- $1 - p$ is the failure probability
- $\binom{n}{x}$ is the number of orderings

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

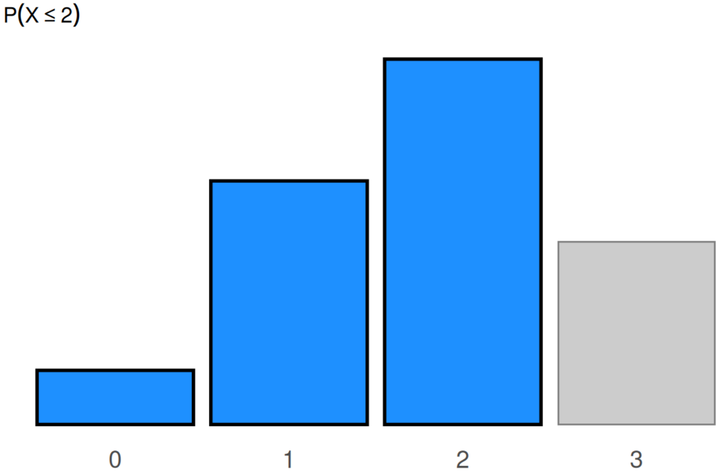
$n!$ is the product of numbers 1 to n. $0! = 1$.

Can be calculated with `choose()` or `factorial()`.

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

R `dbinom()` calculates this value given x, n, p.

`pbinom()` finds a “less than or equal” (cumulative) probability.


P(X ≤ 2)
0
1
2
3

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

Let $Y \sim Binom(8, 0.3)$. Find the following:

- $P(Y \leq 5)$
- $P(Y > 4)$
- $P(Y \geq 4)$
- $P(3 \leq Y \leq 6)$

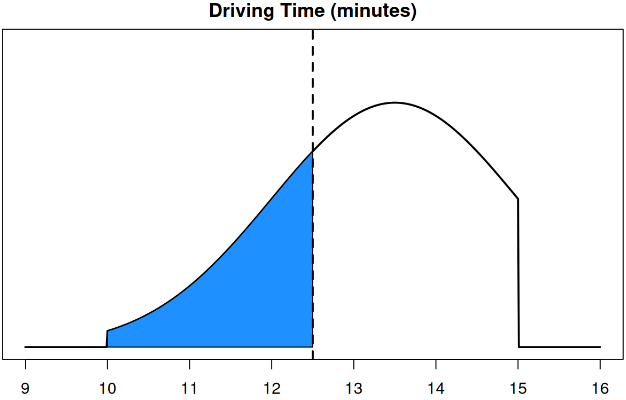
We have shortcuts for the mean/EV and variance of $X \sim Binom(n, p)$.

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

In the lights example, $\mu = 1.8$ and $\sigma^2 = 0.72$.

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

In a population, 30% of individuals are below the 30th percentile, and 70 are above.

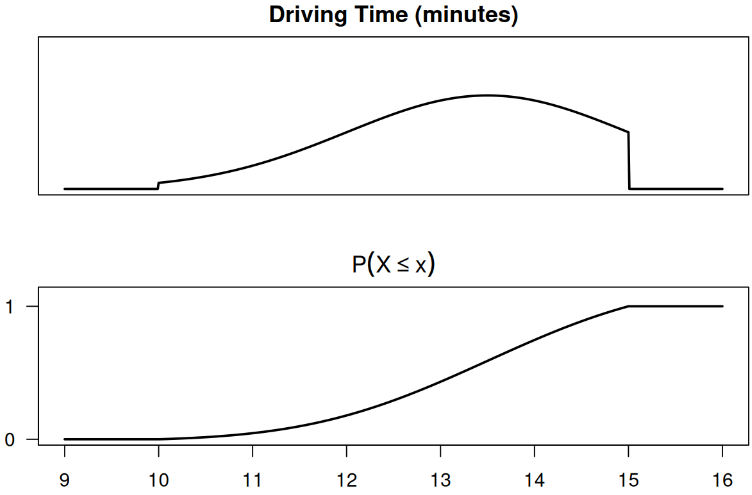


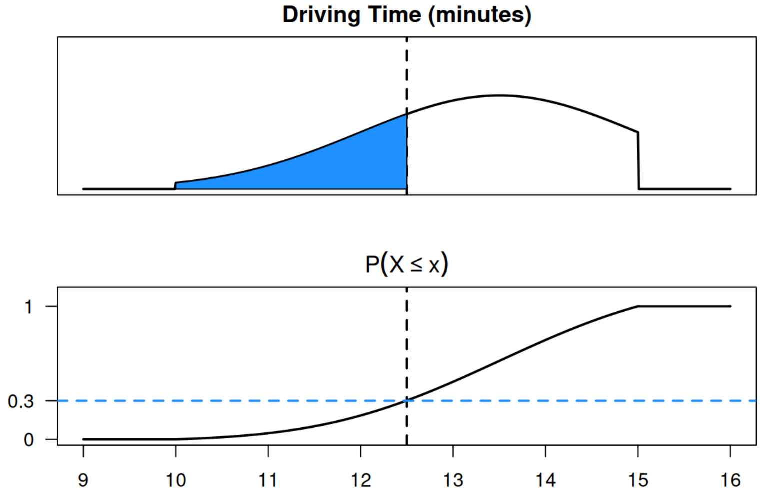
"30% of the time, we get there in less than 12.5 minutes."

In general, quantiles subdivide a specific fraction of the population.

The p percentile is the value such that there is p probability to the left and $1 - p$ to the right.

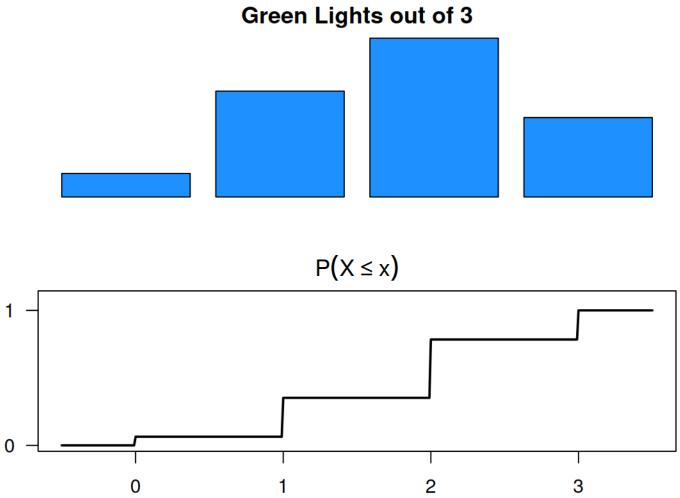
Imagine graphing the *cumulative* probability:

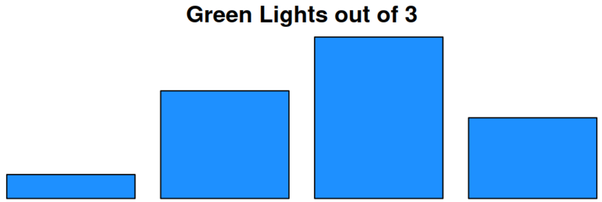


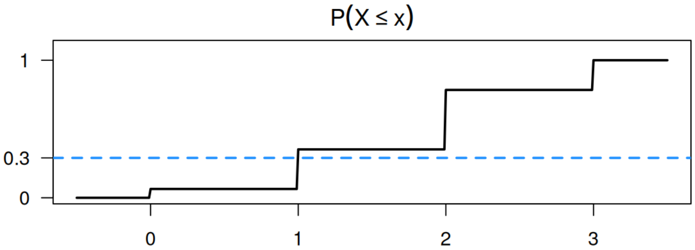


12.5 = 30th percentile

Percentiles work differently for discrete RVs.







The 30th percentile is 1. Verify with ‘qbinom()‘.

In general, the p percentile of a discrete variable is given by q.

How large does q need to be before

$$P(Binom \leq q)$$

is at least p?

For binomial RVs, we have `qbinom()` and `pbinom()`.

- `pbinom()`: input x value, output cumulative probability
- `qbinom()`: input cumulative probability, output x value (i.e. q)

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, var σ^2 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?