PROBABILITY

STEP-BY-STEP EXAMPLES

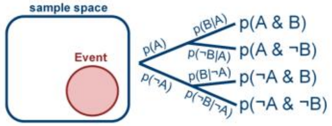
Let's do some examples

p(A) + p(¬A) = 1
p(A & B) = p(A) x p(B)
p(A or B) = p(A) + p(B)

PROBABILITY REVIEW

p(impossible event) = 0
p(certain event) = 1
For any event A, 0 ≤ p(A) ≤ 1
p(A) + p(¬A) = 1
p(A or B) = p(A) + p(B) - p(A & B)
p(A or B) = p(A) + p(B) if A and B are **mutually exclusive**.

p(A & B) = p(A) x p(B|A)
p(A & B) = p(A) x p(B) if A and B are **independent**.



First data set

40 cobras | 60 pythons

Sample space = 100 snakes.
Event, which snake gets chosen.
p(snake)=1 and p(¬snake)=0

$$p(cobra) = \frac{40}{40+60} = \frac{40}{100} = 0.4$$

$$p(python) = \frac{60}{40+60} = \frac{60}{100} = 0.6$$

$$p(cobra\ or\ python) = p(cobra) + p(python) = 0.4 + 0.6 = 1$$

What about choosing two or more snakes?
p(first & second) = p(first) x p(second)

40 cobras | 60 pythons

With replacement: sample space same each time.

An idealistic situation, but mathematically easier.

But when we take a sample of multiple individuals from a population, each one changes the population for the next.

(Accurate if population is big enough and samples small enough)

Without replacement: sample space changes.

Realistic situation, but mathematically harder.

First data set

40 cobras | 60 pythons

Can also use complementation rule for questions.

$$p(cobra\&python) = 1 - p(2\ cobras) - p(2\ pythons) = 1 - 0.16 - 0.36 = 0.48$$

$$p(cobra\&python) = 1 - p(2\ cobras) - p(2\ pythons) = 1 - 0.15758 - 0.35758 = 0.48484$$

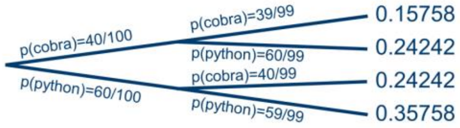
First data set

40 cobras | 60 pythons

Without replacement should use a probability tree.

$$p(cobra\&cobra) = (0.4)(0.39394) = 0.15758$$

$$p(cobra\&python) = 0.24242 + 0.24242 = 0.48484$$



First data set

40 cobras | 60 pythons

Default is usually sampling *with replacement*.
Straightforward addition and multiplication rules work.

$$p(cobra\&cobra) = p(cobra, then\ cobra) = \left(\frac{40}{100}\right)\left(\frac{40}{100}\right) = 0.16$$

$$p(cobra, then\ python) = \left(\frac{40}{100}\right)\left(\frac{60}{100}\right) = 0.24$$

$$p(cobra\&python) = \left(\frac{40}{100}\right)\left(\frac{60}{100}\right) + \left(\frac{60}{100}\right)\left(\frac{40}{100}\right) = 2(0.24) = 0.48$$

What about choosing two or more snakes? e.g., p(two cobras)

With replacement: sample space same each time.

40 cobras / 60 pythons → 0.4; 40 cobras / 60 pythons → 0.4

$$p(two\ cobras) = \left(\frac{40}{100}\right)\left(\frac{40}{100}\right) = (0.4)(0.4) = 0.16$$

Without replacement: sample space changes.

40 cobras / 60 pythons → 0.4; 39 cobras / 60 pythons → 0.3939

$$p(two\ cobras) = \left(\frac{40}{100}\right)\left(\frac{39}{99}\right) = (0.4)(0.3939) = 0.15758$$

First data set

40 cobras | 60 pythons

Sample space = 100 snakes.
Event, which snake gets chosen.
p(snake)=1 and p(¬snake)=0

$$p(cobra) = 0.4$$

$$p(python) = 0.6$$

$$p(cobra\ or\ python) = 0.4 + 0.6 = 1$$

$$p(cobra\&cobra) = 0.16\ or\ 0.15757$$

$$p(python\&python) = 0.36\ or\ 0.35758$$

$$p(cobra\&python) = 0.48\ or\ 0.48484$$

with replacement
without replacement

First data set

The size of the sample space can make a big difference

4 cobras / 6 pythons; 40 cobras / 60 pythons; 400 cobras / 600 pythons

	with replacement	without replacement		
p(cobra & cobra)	*0.16*	*0.13333*	*0.15758*	*0.15976*
p(python & python)	*0.36*	*0.33333*	*0.35758*	*0.35976*
p(cobra & python)	*0.48*	*0.53333*	*0.48484*	*0.48048*

Second data set

20 ducks | 40 eagles | 80 gulls | 60 hawks

Sample space = 200 birds.
Event, which bird gets chosen.
Sampling with replacement.

$$p(duck) = \frac{20}{200} = 0.1$$

$$p(eagle) = \frac{40}{200} = 0.2$$

$$p(duck\ or\ eagle) = (0.1) + (0.2) = 0.3$$

$$p(duck\ \&\ eagle) = (0.1)(0.2) + (0.2)(0.1) = 2(0.02) = 0.04$$

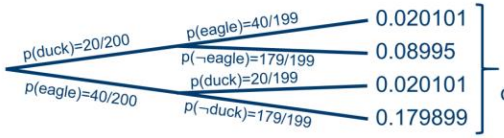
$$p(4gulls) = (0.4)(0.4)(0.4)(0.4) = (0.4)^4 = 0.0256$$

Second data set

20 ducks | 40 eagles | 80 gulls | 60 hawks

Sample space = 200 birds.
Event, which bird gets chosen.
Sampling without replacement.

$$p(duck\ \&\ eagle) = 0.020101 + 0.020101 = 0.0402$$

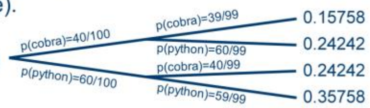


PROBABILITY REVIEW

40 cobras | 60 pythons

20 ducks | 40 eagles | 80 gulls | 60 hawks

Most commonly used:
► p(A or B) = p(A) + p(B)
► p(A & B) = p(A) x p(B)
when A and B independent and we sample with replacement (or population very large).
Otherwise:
p(A & B) = p(A) x p(B|A)



Second data set

The size of the sample space can make a big difference

2 ducks / 4 eagles / 8 gulls / 6 hawks; 20 ducks / 40 eagles / 80 gulls / 60 hawks; 200 ducks / 400 eagles / 800 gulls / 600 hawks

	with replacement	without replacement		
p(duck & eagle)	*0.04*	*0.04211*	*0.04020*	*0.04002*
p(4 gulls)	*0.0256*	*0.01445*	*0.02445*	*0.02548*

Second data set

20 ducks | 40 eagles | 80 gulls | 60 hawks

Sample space = 200 birds.
Event, which bird gets chosen.
Sampling with replacement.

$$p(duck) = \frac{20}{200} = 0.1$$

$$p(eagle) = \frac{40}{200} = 0.2$$

$$p(duck\ or\ eagle) = (0.1) + (0.2) = 0.3$$

$$p(duck\ \&\ eagle) = 0.04\ or\ 0.04020$$

$$p(4\ gulls) = 0.0256\ or\ 0.024451$$

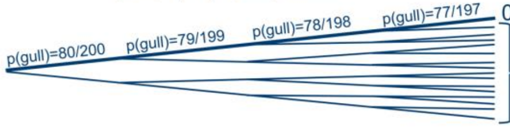
with replacement
without replacement

Second data set

20 ducks | 40 eagles | 80 gulls | 60 hawks

Sample space = 200 birds.
Event, which bird gets chosen.
Sampling without replacement.

$$p(4gulls) = \left(\frac{80}{200}\right)\left(\frac{79}{199}\right)\left(\frac{78}{198}\right)\left(\frac{77}{197}\right) = 0.024451$$



StatsExamples**.com**