Welcome! Probability in R

"When a coincidence seems amazing, that's because the human mind isn't wired to naturally comprehend probability and statistics."

-Neil deGrasse Tyson

Outline

1650's Dice gambling

Chevalier de Méré questioned Pascal which dice game he would be more likely to win:

Game 1: Throw one die, up to 4 times, winning once you roll a six.

Game 2: Throw 2 dice, up to 24 times, winning once you roll a double-six.

Méré's game probabilities

Remember the complement rule and rule for independent events:

Game 1

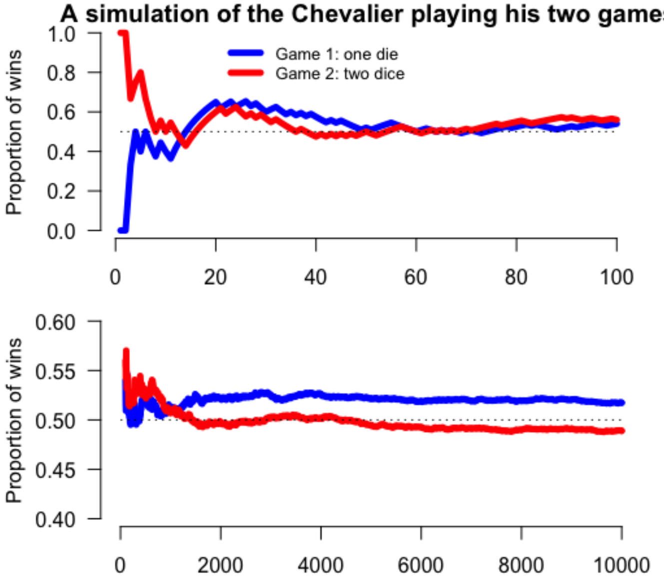
P(rolling at least one 6) =
1 - probability(not rolling a 6) =
$1 - \frac{5}{6}\ x\ \frac{5}{6}\ x\ \frac{5}{6}\ x\ \frac{5}{6} = 1 - \left(\frac{5}{6}\right)^4 = 0.5177$

Game 2

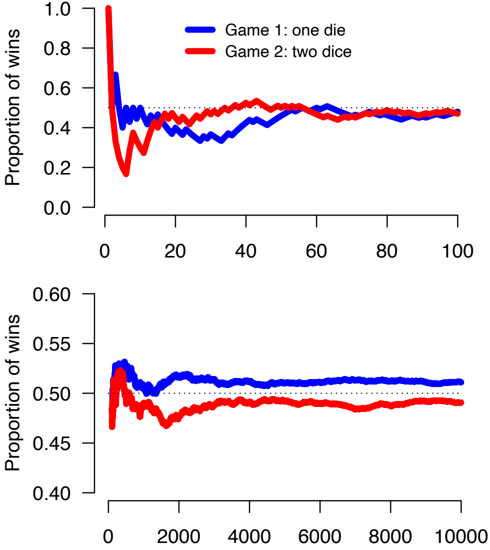
6 outcomes x 6 outcomes = 36 total outcomes;
only 1 is double-six;

P(rolling at least one 6) =
1 - (probability not rolling double-6) =
$1 - \left(\frac{35}{36}\right)^{24} = 0.491$

Simulation: G1 = 0.52 & G2 = 0.49



Simulation 2: G1= 0.52 & G2 = 0.49



Sunburns

We want to determine if on a given day in Miami, two people will get a sunburn. Are the two events independent?

A. Person 1 has a sunburn

and

B. Person 2 has a sunburn

Conditional Independence

No, these are not independent. Event A informs us about Event B.

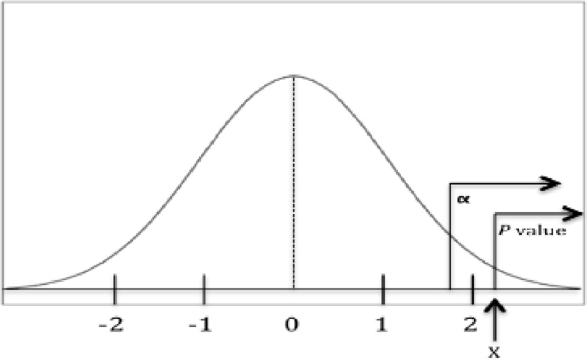
If Person 1 is sunburned this tells us it was a sunny day which increases the probability Person 2 has a sunburn.

If we know the weather on this day was 100°F without an overcast, are Events A and B independent?

Yes, both events are conditionally independent given the information about the weather being sunny.

NHST: under the H_0

A *p*-value is a conditional probability.



The *p*-value is the probability of obtaining a test statistic equal to, or more extreme than, the observed value **under the assumption that the null hypothesis is true.** This is not the same as the probabilility of obtaining your test results.

Frequentist: Conditioned on theory

Frequentist

Parameters are fixed and unknown.

Sample data is random from a population distribution.

Neyman-Pearson's Hypothesis test with Fisher's p-value approach

Probability is interpreted as long-run repeated sampling, under almost identical conditions, from a population distribution.

Bayesian: Conditioned on data

Bayesian

Parameters are random and unknown.

Sample data is fixed.

Probability is interpreted as our beliefs about random variables; a measure of uncertainty.

Distributions are updated with the arrival of new information—use of priors with the likelihood.

Bayes theorem

Bayes Theorem

$$P(B|A) = \frac{P(A \mid B) \ x \ P(B)}{P(A)}$$

$$\mathrm{P(B|A)} = \frac{P(A \mid B) \ x \ P(B)}{P(A \mid B) \ x \ P(B) + P(A \mid B^c) \ x \ P(B^c)}$$

Divide a positive result by all the ways it is possible to get a positive result– true positive plus a false alarm.

Lucy & Ethel

Lucy, Ethel, and a long-term employee are called in because the boss opens a random box of chocolates to discover they are not all wrapped! Furious she pulls out the work schedule....

Each of them works at equal speeds, but Lucy works 30% of the schedule, Ethel works 10%, and the remaining 60% of the hours is an employee that has worked at the company longest.

Lucy & Ethel

The boss is a Frequentist so she calls in the employee that works the most to hand out a pink slip.

Thankfully this employee is a Bayesian and says the boss is forgetting their mistake rates; she makes 0.3% mistakes! Lucy makes 0.7% mistakes, Ethel mistake rate is 1%—fire Ethel!

Ethel over hears and says she works the least hours so isn't her!

Lucy & Ethel

The probability of the mistake was made AND it was the employee:
$P(\text{Employee} \cap \text{unwrapped}) = 0.6(0.003) = 0.0018$

$\text{Lucy} = 0.0021$
$\text{Ethel} = 0.001$

For the denominator, we calculate EVERY possible way this mistake was made:
$0.0018 + 0.0021 + 0.001 = 0.0049$

Lucy & Ethel

Using Bayes' Theorem:

$$\mathrm{P(Employee} \mid \mathrm{unwrapped)} = \frac{0.0018}{0.0049} = 0.37$$

$$\mathrm{P(Lucy} \mid \mathrm{unwrapped\ )} = \frac{0.0021}{0.0049} = 0.43$$

$$\mathrm{P(Ethel} \mid \mathrm{unwrapped\ )} = \frac{0.001}{0.0049} = 0.20$$

Lucy & Ethel

Calculate this in R:

```
prior <- c(0.6, 0.3, 0.1)
likelihood <- c(0.003, 0.007, 0.01)
posterior <- prior * likelihood
Bayesoutcome <- round(posterior/sum(posterior),2)
Bayesoutcome
```

```
[1] 0.37 0.43 0.20
```

Bayes v. Frequentist

I take a coin of my pocket and ask my friends the probability of heads when flipped. Restriction: One of them can only toss the coin 10 times.

Frequentist friend says 50% but wants to flip the coin.
All seven are heads, so he answers:

$$\widehat{\mathrm{H}} = \frac{X}{10} = 0.7 = 70\% \text{ likely to get heads.}$$

Bayes v. Frequentist

Bayesian friend knows my brother is a prankster and one of the coins could be his trick coin.

Bayesian:

$$\text{P(Bias} = \text{H} \mid \text{Heads)} = \frac{P(Heads \mid Bias = H)) \;\; x \;\; (P(Bias = H))}{P(Heads)}$$

Trick coin in R

```
#trick coin####
prob_f <- dbinom(7, 10, .5)
prob_trick <- dbinom(7, 10, .99)

# coin fair? ####
prob_f / (prob_f + prob_trick)

#flipping 10 coins####
fair <- rbinom(20000, 10, .5)
fair
heads <- rbinom(20000, 10, .9)
heads

# 7 heads: fair and trick coin####
fairflip <- sum(fair == 7)
fairflip
trickflip <- sum(heads == 7)
trickflip

# Number of fair coins, showing 7 heads
fairflip / (fairflip + trickflip)
```

Combinations

Out of 34 people, how many 5 person teams can be formed?

Combination: order doesn't matter

$$\binom{n}{k} = \frac{n!}{r!(n-r)!} = n \; C \; r$$

$$\binom{34}{5} = 278{,}256$$

```
> choose(34, 5)
[1] 278256
```

Permutations

On a team of 12, what is the number of ways we can select a first place and second place?

Permutation: order matters

$$\binom{n}{k} = \frac{n!}{(n-r)!} = n\ P\ r$$

$$\frac{10!}{(12-2)!} = \frac{12!}{10!} = 132$$

```
> factorial(12)/factorial(10)
[1] 132
```

Binomial

Consider a coin-flip as a Bernoulli trial, where heads is a success and tails is a failure on a single toss. The binomial distribution can be thought as multiple Bernoulli trials.

Binomial distribution

$$p(x) = \frac{n!}{x!(n-x)!} \textit{success}^{x} \ (\textit{1 - success})^{n-x}$$

Where n is the number of trials, and x is number of successes.

Binomial: example

One migraine medication is estimated to reduce suffering for 80% of patients. When we administer this medication to 10 new people, what is the probabililty 7 of those people get relief?

Binomial distribution

$$p(x) = \frac{n!}{x!(n-x)!} success^{x} \ (1 \text{ - } success)^{n-x}$$

$$\mathrm{p(x)} = \frac{10!}{7!(10-7)!} \ 0.80^{7} \ (1 \text{ - } 0.80)^{10-7}$$

```
> dbinom(7, size=10, prob=0.8)
[1] 0.2013266
```

Poisson

Calculate number of events within an interval.
Invented to determine murder rates, given the average—"Rare events".

Poisson distribution

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

Poisson: example

You're deciding if rush hour traffic will delay your commute home. Between 4pm and 5pm, an average of 600 cars are on your commute per hour. What's the probabilty that 650 or less will be driving in that particular hour?

Poisson distribution

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

$$\text{p(x)} = \frac{600^{650}}{650!} e^{-600}$$

```
> ppois(650, lambda = 600)
[1] 0.9793456
```

Poisson: example

You're deciding if rush hour traffic will delay your commute home. Between 4pm and 5pm, an average of 600 cars are on your commute per hour. What's the probabilty that 650 or **MORE** will be driving in that particular hour?

For the probability of 650 or more use `lower.tail = FALSE`

```
> ppois(650, lambda = 600, lower.tail = FALSE)
[1] 0.02065445
```

Thank you!

If someone asks "Are you a Bayesian or are you a Frequentist?"
the answer should be yes.