Resampling Methods of Solving Problems

Neba Nfonsang
University of Denver

In [1]:
```
require(tidyverse)
```

```
Loading required package: tidyverse
Warning message in library(package, lib.loc = lib.loc, character.only = TRUE, logical.return = TRUE, :
"there is no package called 'tidyverse'"
```

- Resampling methods provide a way of solving a problem using simulations without using an analytic method or formula.

Problem 1

- A construction company owns 50 trucks. Based on past experience, on any given day, the chance that one truck will break down is about 0.1.
- What is the probability that on any particular day, 10 or more trucks owned by the company will break down?

Solve the Problems Using Resampling

Algorithm to be implemented:

- 1) randomly draw a sample of 50 trucks with replacement from a sample of 0 and 1 with a replacement wheere the probability of drawing a 1 (defective truck) is 0.1
- 2) compute the number of of defective trucks in the sample
- 3) if number of defective trucks is equal or greater 10, record "TRUE" or 1 (keep track)
- 4) repeat the process (step 1 to 3) 10000 times
- 5) compute the proportion of samples with number of defective trucks equal to or greater than 10

Note the frequentist definition for probability is that:

- The probability of an event is the proportion of times that the event has taken place in the past, usually based on a long series of trials.
- However, we will be simulating these trials through resampling

In [2]:
```
iterations <- 10000
sim.values <- rep(NA, iterations)

for (i in 1:iterations){
    samp <- sample(c(0, 1), size = 50, replace=T, prob=c(0.9, 0.1))
    sim.values[i] <- sum(samp) >= 10
}

mean(sim.values)
# the mean() function is used to find the proportion of 1's
# in a sequence of 0's and 1's
```

0.022

- Note: we could also sample with replacement from a sample consisting of 9 0's and one 1, where each of the 10 values are equally likely. This is consistent with the chance of one truck breaking down being 0.1.

Solve the Problem Using a Binomial Distribution (Analytically)

- let's check the approximate answer we got using the simulation by solving the problem analytically.

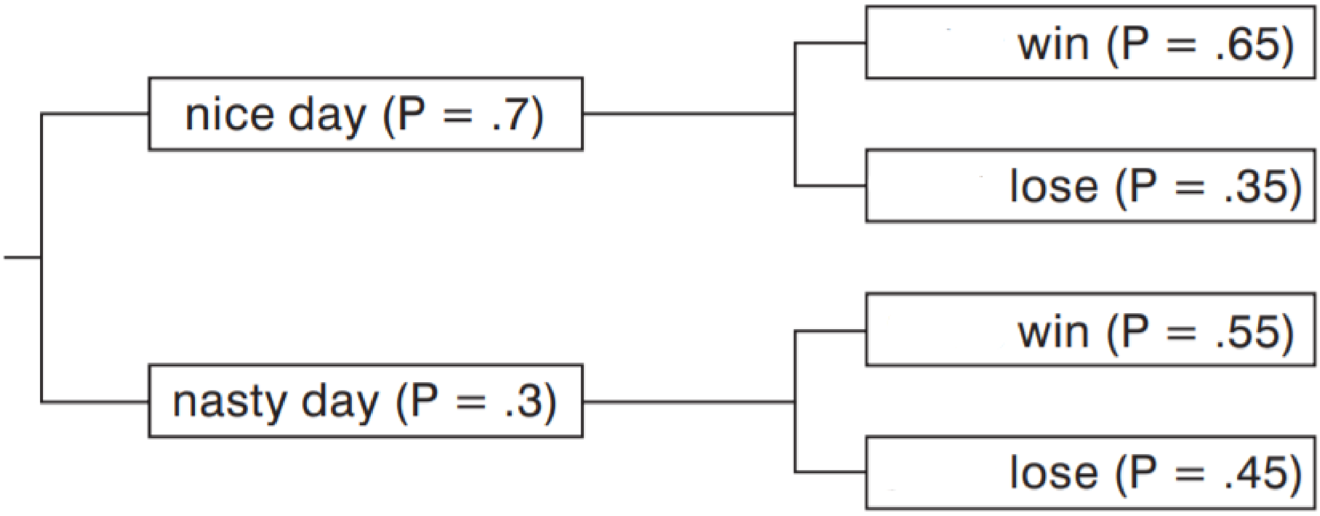
In [3]:
```
pbinom(q=9, size = 50, prob = 0.1, lower.tail = F)
```

0.0245379357045915

Joint Probability Problems and Simulations

Problem 2

- Suppose the probability that the New York Giants will win the football game next Sunday is 0.65 given that the weather on sunday is nice. If the weather on sunday is not nice, the probability that the New York Giants will win is 0.55.
- Based on historic data, the probability that next sunday will be nice is 0.7.
- What is the probability that the weather next sunday will be nice and that the New York Giants will win the football game?
- The tree diagram below is helpful in visualizing the joint probabilities.



set up

- 1) randomly draw a day from {0, 1} where 0=nasty day, 1=nice day with a probability of drawing a nice day being 0.7
- 2) randomly draw a results of the game from {0, 1} where 0=lose, 1=win with the probability of winning being 0.65
- 3) if the draws in step 1 and 2 were a "nice day" and "win", record 1, otherwise, recored 0 (keep track).
- 4) repeat the process (step 1 to 3) 10000 times
- 5) compute the proportion of samples where a "nice day" and "win" was recorded.

In [4]:
```
iterations <- 100000
sim.values <- rep(NA, iterations)

for (i in 1:iterations){
    samp.day <- sample(c(0, 1), size = 1, replace=T, prob=c(0.3, 0.7))
    samp.result <- sample(c(0, 1), size = 1, replace=T, prob=c(0.35, 0.65))

    sim.values[i] <- samp.day*samp.result # a combination of 1*1 give 1, otherwise 0
}

mean(sim.values)
```

0.45525

Analytical Solution to the Joint Probability Problem

- let's check the approximate answer we got using the simulation by solving the problem analytically.
- from the tree diagram, the probability of having a nice day and winning can be computed as P(nice day)*P(win):

In [5]:
```
prob.nice.day <- 0.7
prob.win <- 0.65

prob.nice.day_win <- prob.nice.day*prob.win
prob.nice.day_win
```

0.455

Problem 3

- suppose a sample of students has 5 males and 5 females.
- what is the probability that two students selected with replacement will be both males?

In [6]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

for (i in 1:iterations){
    samp1 <- sample(c(0, 0, 1, 1), size=1, replace=T) # first selection
    samp2 <- sample(c(0, 0, 1, 1), size=1, replace=T) # second selection
    # 0=female, 1=male

    sim.values[i] <- samp1*samp2 # a combination of 1*1 give 1, otherwise 0
}

mean(sim.values)
```

0.2512

In [7]:
```
# analytical solution
prob.drawing.male1 <- 2/4
prob.drawing.male2 <- 2/4
prob.drawing.male12 <- prob.drawing.male1 *prob.drawing.male2
prob.drawing.male12
```

0.25

Problem 4

- What is the probability that two or more people in a room of 25 students will have the same birth day?

set up

- create a list of 1 to 356 representing all possible birth days
- randomly draw a sample of 25 students from the list
- find if there are any duplicates in the sample, if so, record 1, otherwise record 0 (keep track)
- repeat step 1 to 3 several times (10,000 times)
- compute the proportion of samples with duplicates

In [8]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

birth.days <- seq(from=1, to=365, by=1)

for (i in 1:iterations){
    samp <- sample(birth.days, size=25, replace=T)
    duplicates <- sum(duplicated(samp))

    sim.values[i] <- duplicates >= 1
}

mean(sim.values)
```

0.5694

Problem 5

- What is the probability that 3 out of 4 children in a family of 4 will be girls?

set up

- let's assume, for a single birth the probability of getting a boy is 0.5 and the probability of getting a girl is 0.5, just like flipping a coing and getting a head of tail.
- 1) randomly select four kids from {0, 1} with replacement, where 0=boy, 1=girl, and the probability of selecting a girl is 0.5.
- 2) compute the number of girls in the sample
- 3) if the number of girls = 3, record 1 (TRUE), otherwise record 0 (FALSE) (keep track)
- 4) compute the proportion of samples with three girls

In [9]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

for (i in 1:iterations){
    samp <- sample(c(0, 1), size=4, replace=T, prob=c(0.5, 0.5))

    sim.values[i] <- sum(samp)==3
}

mean(sim.values)
```

0.2532

analytical solution

- to solve this problem analytically, a tree diagram could be used.
- this like a problem of tossing 4 coins in a sequence

Problem 5

- Based on past data, the 10%, 60% and 30% of fruits eaten by John are orange, apple, and mango,
- What is the probability that the next three fruits eaten by John will be an orange, apple and a manago (the order in which the fruits are eaten does not matter).

set up

- 1) sample three fruits from {"O", "A", "M"} with replacement with probability .1, .6, and .3 respectively
- 2) if the sample contains "O", "A", and "M", record 1 (TRUE), otherwise record 0 (FALSE)
- 3) repeat step 1 and 2 several times (10,000 times)
- 4) compute the proportion of samples that contained "O", "A", "M"

In [10]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

for (i in 1:iterations){
    samp <- sample(c("O", "A", "M"), size=3, replace=T, prob=c(0.1, 0.6, 0.3))

    sim.values[i] <- sum(sort(samp) == c("A", "M", "O")) == 3
}

mean(sim.values)
```

0.1111

Problem 6

- The distribution of lengths for hammer handles is as follows:
- 20%, 30%, 30%, and 20% of hammer handles are respectively 10.0, 10.1, 10.2, and 10.3 inches long.
- 20%, 20%, 30%, and 20%, and 10% of hammer heads are 2.0, 2.1, 2.2, 2.3, and 2.4 inches long.
- If a hammer handle and a hammer head are randomly selected, what would be the mean of the total length of the hammer?

set up

- 1) randomly draw one hammer handle from the distribution of hammer handles according to their probabilities
- 2) randomly draw one hammer head from the distribution of hammer heads according to their probabilities
- 3) compute the total length and record it (keep track)
- 4) repeat step 1 to 3 several times (10,000 times)
- 5) compute the mean of the total lengths recorded

In [11]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

for (i in 1:iterations){
    handle.length <- sample(c(10, 10.1, 10.2, 10.3), size=1, replace=T,
                           prob=c(0.2, 0.3, 0.3, 0.2))

    head.length <- sample(c(2, 2.1, 2.2, 2.3, 2.4), size=1, replace=T,
                           prob=c(0.2, 0.2, 0.3, 0.2, 0.1))

    sim.values[i] <- handle.length + head.length
}

mean(sim.values)
```

12.33219

Problem 7

What is the probability of selecting four girls and one boy when selecting five students from a class of 25 girls and 25 boys? (hint: select without replacement)

In [12]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

# create a class of 25 girls and 25 boys, 1=girl, 0=boy
class <- rep(c(0, 1), 25)
for (i in 1:iterations){

    # randomly select 5 students without replacement
    samp <- sample(class, size=5, replace=F)

    sim.values[i] <- sum(samp)==4 # check for 4 girl
}

mean(sim.values)
```

0.151

Problem 8

- What is the probability of getting an ordered series of four girls and then one boy, from a class of twenty-five girls and twenty-five boys? (hint: without replacement, order matters)

In [13]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

# create a class of 25 girls and 25 boys, 1=girl, 0=boy
class <- rep(c(0, 1), 25)
for (i in 1:iterations){

    # randomly select 5 students without replacement
    samp <- sample(class, size=5, replace=F)

    # check for the order: 4 girls, 1 boy and record 1 if true, otherwise 0.
    sim.values[i] <- sum(samp == c(1, 1, 1, 1, 0))==5
}

mean(sim.values)
```

0.0295

Problem 9

- 40 couples (20 male, 20 female) came to a party and were randomly paired so they could introduce themselves to each other. The pairing was done by randomly selecting a male from a male group and randomly selecting a female from the female group.
- What is the probability that 4 or more couples are paired with their own partners?

In [14]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

males <- 1:20
females <- 1:20

for (i in 1:iterations){
    male <- sample(males, size=20, replace=F)
    female <- sample(females, size=20, replace=F)

    # check if couples paired with their own partners >= 4
    sim.values[i] <- sum(male == female) >= 4
}

mean(sim.values)
```

0.0167

Problem 10

- Twenty executives are to be assigned to two divisions of a firm. The top managers want to distribute the talents to the divisions evenly. The managers are wondering, what is the probability that the top 10 best are distribution in the ratio 5:5, 4:6, and 3:7?

In [15]:
```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

executives <- rep(c(1, 0), 10) # 1's=best executives

for (i in 1:iterations){
    # select 10 executives without replacement for one division
    samp <- sample(executives, size=10, replace=F)

    # count the number of best executives in the group

    sim.values[i] <- sum(samp)
}

# probability of 5:5
mean(sim.values == 5)

# probability of 4:6
mean(sim.values == 6)

# probability of 3:7
mean(sim.values == 7)
```

0.3394

0.2386

0.0747

Problem 11

- 30 Managers are to be moved to other stations after every 3 years of service but are allow to randomly select their next station, including their present station. There are 30 stations with 30 managers.
- What is the probability that 1, 2, 3... managers will select their current station?

In [16]:

```
# simulation

iterations <- 10000
sim.values <- rep(NA, iterations)

stations <- 1:30

for (i in 1:iterations){
    samp <- sample(stations, size=30, replace=F)


    # count the number of managers who selected their positions

    sim.values[i] <- sum(samp==stations)
}


# probability that 1 managers selected their current station
mean(sim.values==1)

# probability that 2 managers selected their current station
mean(sim.values==2)

# probability that 3 managers selected their current station
mean(sim.values==3)
```

0.3629

0.1855

0.062

In [ ]: