STAT 140: Homework 4

YOUR NAME HERE

10/2/2019

BOOK EXERCISES (OpenIntro Statistics, Fourth Edition)

3.5) **Coin flips.** For (a)-(c), define events (e.g. heads=H) and use appropriate probability notation and find the numerical answer. If you flip a fair coin 10 times, what is the probability of

a) getting all tails?

Answer:

b) getting all heads?

Answer:

c) getting at least one tails?

Answer:

d) What rules did you use to calculate the probabilities in (a)-(c)?

Answer:

3.9) **Disjoint vs. independent.** In parts (a) and (b), identify whether the events are disjoint, independent, or neither (events cannot be both disjoint and independent).

a) You and a randomly selected student from your class both earn A's in this course.

Answer:

b) You and your class study partner both earn A's in this course.

Answer:

c) If two events can occur at the same time, must they be dependent? **Explain in 1 sentence.**

Answer:

3.12) **School absences.** Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness. *Notation:* For (a)-(e), define M as the random variable corresponding to missing school days and use appropriate notation in addition to finding the numerical answer. For example, the probability a random student misses 2 days of school is: $P(M = 2) = 15\%$.

a) What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year?

Answer:

b) What is the probability that a student chosen at random misses no more than one day?

Answer:

c) What is the probability that a student chosen at random misses at least one day?

Answer:

d) If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note in one sentence any assumption you must make to answer this question.

Answer:

e) If a parent has two kids at a DeKalb County elementary school,what is the probability that both kids will miss some school, i.e. at least one day? Note in one sentence any assumption you make.

Answer:

f) If you made an assumption in part (d) or (e), do you think it is reasonable? If you didn't make any assumptions, double check your earlier answers.

Answer:

3.13) **Joint and conditional probabilities.** $P(A) = 0.3$, $P(B) = 0.7$

a) Can you compute $P(A \text{ and } B)$ if you only know $P(A)$ and $P(B)$?

Answer:

b) Assuming that events A and B arise from an independent random process,

---

i) what is $P(A \text{ and } B)$?

Answer:

ii) what is $P(A \text{ or } B)$?

Answer:

iii) what is $P(A|B)$?

Answer:

c) If we are given that $P(A \text{ and } B) = 0.1$, are the random variables giving rise to events A and B independent?

Answer:

d) If we are given that $P(A \text{ and } B) = 0.1$, what is $P(A|B)$?

Answer:

3.29) **College smokers.** At a university, 0.13 of students smoke.

a) Calculate the expected number of smokers in a random sample of 100 students from this university.

Answer:

b) The university gym opens at 9 am on Saturday mornings. One Saturday morning at 8:55 am there are 27 students outside the gym waiting for it to open. Should you use the same approach from part (a) to calculate the expected number of smokers among these 27 students?

Answer:

3.33) **Portfolio return.** A portfolio's value increases by 18% during a financial boom (B) and by 9% during normal times (N). It decreases by 12% during a recession (R). What is the expected return on this scenario if each scenario is equally likely?

Answer:

3.45) **Variance of a mean, Part I.** Suppose we have independent observations X_1 and X_2 from a distribution with mean μ and standard deviation σ. What is the variance of $\frac{X_1+X_2}{2}$?

Answer:

Mountain Day. As we all know, Mountain Day is a glorious Mount Holyoke tradition dating back to 1838, the year after the college was founded. On Mountain Day, classes are cancelled and students and faculty/staff climb to the Summit House atop Mount Holyoke (the college's namesake). Each year, students try to guess when Mountain Day will be (but it still ends up being a surprise). In this problem, you will explore when Mountain Day might fall (month and day of the week) using historical data (1996-2018).

	Monday	Tuesday	Wednesday	Thursday	Friday	**Total**
September	1	6	5	1	0	13
October	0	7	1	1	1	10
Total	1	13	6	2	1	23

When writing your answers, use the following notation:

- M=Monday
- T=Tuesday
- W=Wednesday
- R=Thursday
- F=Friday
- S=September
- O=October

a) What is the probability that Mountain Day falls in September? Use appropriate probability notation and find a numerical answer.

Answer:

b) If Mountain Day does not fall in September, what is the probability that it falls on a Tuesday? Use appropriate probability notation and find a numerical answer.

Answer:

c) What is the probability that Mountain Day falls on an October Wednesday?

Answer:

R EXERCISE

The Law of Large Numbers states that as the sample size grows (as more observations are recorded), the observed probability of an event, $\hat{p}_n$, converges to the probability of that event, p.

In class, we looked at this phenomenon for a fair coin (Slide 9 of Probability lecture). Here we will look at this phenomenon for a fair die.

Sample function:

The sample function allows us to simulate a random process, so we can `flip coins" or`roll dice" on the computer instead of in real life. Looking at the help file (?base::sample), you see that sample() has four arguments:

- **x**: for us this will be a vector of possible outcomes of our random variable
- **size**: n, the sample size (number of observations) - just an integer
- **replace**: a logical; we will set replace=TRUE so that on each die roll we can observe any one of the six outcomes
- **prob**: we do not need to specify this unless some observations are more likely than others. The default is to give equal weight/chance to all possible outcomes.

```
## Demonstrate sample function for fair dice rolls

## Specify set of all outcomes in a vector
roll_outcomes <- c(1,2,3,4,5,6)

## Roll a fair die once
r1 <- sample(x=roll_outcomes, size=1, replace=TRUE)

r1
```

```
## [1] 6
```

```
## Roll a fair die once
r5 <- sample(x=roll_outcomes, size=5, replace=TRUE)

r5
```

```
## [1] 3 4 5 2 5
```

Describe what r1 gives you.

Answer:

If you roll again (rerun the sample line for r1), are you guaranteed get the same result? (Try it if you are not sure!) Why or why not? (**Limit answer to 1 sentence.**)

Answer:

Describe what r5 gives you.

Answer:

Sample function in a loop:

It is tedious to write a separate line of code each time I want to change my sample size. In R, we can write a loop that lets us increase the sample size automatically. I will not ask you to do this independently, but you should be aware that you can.

Here we are only going to keep track of the probability that we roll a 1 and plot the result.

```
## This is the maximum number of observations I am going to make.
N <- 1000

## This creates a place to store my observations
## It is a data frame with two columns, n and p_n
## n is the sample size
## p_n is the observed proportion of times we roll a 1.
LLN_dr <- data.frame(n=1:N, p_n=rep(0,N))
for (n in 1:N){
  LLN_dr$p_n[n] <- sum(sample(x=roll_outcomes,size=n,replace=TRUE)==1)/n
}

## load ggplot2
library(ggplot2)

## ggplot for observed probability versus number of rolls
lln_p <- (ggplot(data=LLN_dr, aes(x=n, y=p_n)) + geom_line()
          + theme_bw()
          + geom_hline(yintercept = 1/6, linetype="dashed", color="red")
          ## Adds a red dashed line at p, the true probability
          + xlab("n (number of rolls)") + ylab(expression(hat(p)[n])))
lln_p
```

Now, you will revisit the example of the weighted die from class (Probability lecture, Slide 63). Simulate the process (use the sample function - you can copy and paste the code from the LLN fair die roll R chunk and then tweak it as necessary for our new problem). Keep track of the probability that we roll a 1 and plot the result as we did before. You will need to add prob=weighted.prob within the parentheses for sample, where weighted.prob is specified below:

Answer

```
## Weighted probability vector (Slide 63)
weighted.prob <- c(3/6,1/6,1/12,1/12,1/12,1/12)

## Copy and paste code from previous R chunk here
```

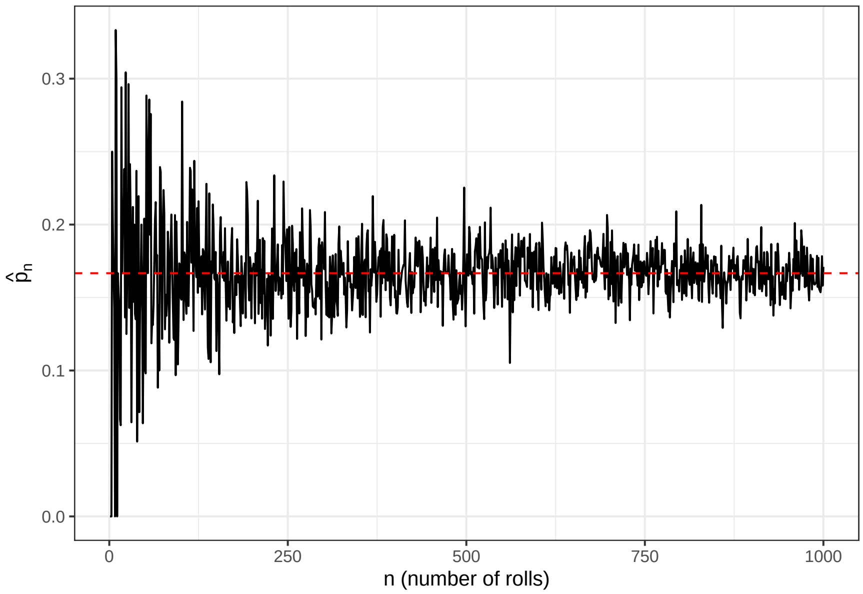

Figure 1: Fair die rolls

```
## Make sure you add the fourth argument, prob=weighted.prob, to the sample function
## Make sure you change the value of yintercept within geom_hline to correspond
## to the expected probability of observing a 1.
## -------------- YOUR ANSWER HERE -------------- ##
```