Ch 3.2 Practice Expected value

Solutions

1. Suppose that a hat contains slips of papers containing the numbers 1, 2, and 3. Two slips of paper are drawn at the same time and the numbers on the slips are multiplied together. Calculate the expected value of the product of the numbers on the slips of paper. Confirm your results using simulation.

The sample space is: $S = \{1 * 2 = 2, 1 * 3 = 3, 2 * 3 = 6\}$.

Set up vectors for the random variable and the vector of probabilities.

```
X <- c(2,3,6)
p_x <- rep(0.33,3)
```

- Theoretical expected value

```
(E_X <- sum(X*p_x))
```

```
## [1] 3.63
```

- Confirm via simulation

```
two.papers <- sample(X, 10000, prob=p_x, replace=TRUE)
mean(two.papers)
```

```
## [1] 3.6555
```

Another approach to simulation that doesn't use the pmf of the product.

```
papers <- replicate(10000,{
  results <- sample(c(1,2,3), 2)
  product <- results[1]*results[2]
})
mean(papers)
```

```
## [1] 3.6442
```

Common Mistakes

Assuming two hats, or drawing with replacement In this case, the sample space for X is expanded.

```
ss <- expand.grid(c(1,2,3), c(1,2,3))
ss$prod <- ss[,1]*ss[,2]
ss
```

```
##   Var1 Var2 prod
## 1    1    1    1
```

```
## 2    2    1    2
## 3    3    1    3
## 4    1    2    2
## 5    2    2    4
## 6    3    2    6
## 7    1    3    3
## 8    2    3    6
## 9    3    3    9
```

```
(x <- unique(ss[,3]))
```

```
## [1] 1 2 3 4 6 9
```

The theoretical pmf then for the product would be:

```
(p.x <- proportions(table(ss[,3])))
```

```
##
##         1         2         3         4         6         9
## 0.1111111 0.2222222 0.2222222 0.1111111 0.2222222 0.1111111
```

which would lead to an Expected value of

```
sum(x*p.x)
```

```
## [1] 4
```

2. In the summer of 2020, the U.S. was considering *pooled testing* of COVID-19. This problem explores the math behind pooled testing. Since the availability of tests is limited, the testing center proposes the following pooled testing technique:

1. Two samples are randomly selected and combined. The combined sample is tested.
2. If the combined sample tests negative, then both people are assumed negative.
3. If the combined sample tests positive, then both people need to be retested for the disease.

Suppose in a certain population, 5 percent of the people being tested for COVID-19 actually have COVID-19, and that individuals' test results are independent. Assume the test is 100% accurate, so if the person is negative they will always test negative. Let X be the total number of tests that are run in order to test two randomly selected people.

a. What is the pmf of X? Explain in words or symbols & equations how you calculated your probabilities.

Let A be the event that person 1 testing positive, and B be the event that person 2 tests positive.

- If the combined sample tests negative, then both people are assumed negative.
  - $P(A^c \cap B^c) = p(A^c) * p(B^c) = .95^2$
  - This results in 1 test, so $X = 1$
- If the combined sample tests positive, then both people need to be retested for the disease.
  - This results in 3 tests, so $X = 3$.
  - Using rules of probability: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .05 + .05 - (.05 * .05) = .0975$
  - Using the combinitorics: $P(A \cap B^c) \cup P(A^c \cap B) \cup P(A \cap B) = (.05 * .95) + (.95 * .05) + (.05 * .05) = .0975$

```
prob_covid <- c(0.05,0.95)
X <- c(1,3)
p_1.test <- 0.95^2
p_3.tests <- .05 + .05 - .05*.05
p_x <- c(p_1.test,p_3.tests)
# cbind = combine vectors as columns in a table
# knitr::kable = display the table in a nice format knitted
knitr::kable(cbind(X, p_x))
```

X	p_x
1	0.9025
3	0.0975

b. What is the theoretical expected value of X? Write your answer out in a sentence.

```
(Exp.tests <- sum(X*p_x))
```

```
## [1] 1.195
```

On average this pooled method of testing uses 1.195 tests to test two people.

c. Now consider this process if three samples are combined. Let Y be the total number of tests that are run in order to test three randomly selected people. If the pooled test is positive, then all three people need to be retested individually. Use simulation to calculate the expected number of tests

> Hint: The function `ifelse()` can be useful here. Try `ifelse(number_positive == 0, 1, 4)`. This returns a 1 if the value of `number_positive` is equal to 0, and 4 otherwise (if number_positive is not 0).

This hint relied on taking the approach of simulating each individual person's sample, and then pooling.

```
test.result <- c("+", "-")
prob.result <- c(.05, .95)

n.tests <- replicate(10000, {
  test.3 <- sample(test.result, prob=prob.result, size=3, replace=TRUE)
  Y <- ifelse(sum(test.3=="+") == 0, 1, 4)
})

mean(n.tests)
```

```
## [1] 1.4443
```

You could also approach this using the pmf:

$P(\text{all three negative}) = (.95^3)$, and that results in $Y = 1$ only a single test being run. Since all pmf's have to add to 1, that means the probability that at least 1 is positive, $(Y = 4)$, is 1-(.95)^3

```
p.all.neg <- .95^3
n.tests <- sample(c(1,4), 10000, replace=TRUE, prob = c(p.all.neg, 1-p.all.neg))
mean(n.tests)
```

```
## [1] 1.4278
```

d. If your only concern is to minimize the expected number of tests given to the population, would you recommending pooling 2 or 3 samples together?

The expected number of tests given was smaller when sampling 2 rather than 3.

3. How costly is it to fly? An airline charges the following baggage fees: $25 for the first bag and $35 for the second. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage and 12% have two pieces. We suppose a negligible portion of people check more than two bags.

a. Define a random variable using a sentence. No R code here.

Let D be the dollar amount a person pays to check their baggage.

a. Build a probability model (pmf) for the revenue per passenger for the airline.

```r
D <- c(0, 25, 25 + 35)
p.d <- c(.54, .34, .12)
knitr::kable(cbind(D, p.d))
```

D	p.d
0	0.54
25	0.34
60	0.12

b. Compute the theoretical average revenue per passenger. Confirm using simulation.

- Theoretical

```r
(mu.d <- sum(D*p.d))
```

```
## [1] 15.7
```

- Simulation

```r
revenue <- sample(D, size=10000, replace=TRUE, prob=p.d)
mean(revenue)
```

```
## [1] 15.576
```

c. Compute the corresponding standard deviation. Confirm using simulation.

- Theoretical

```r
sqrt(sum(p.d*(D - mu.d)^2))
```

```
## [1] 19.95019
```

- Simulation

```r
sd(revenue)
```

```
## [1] 19.79109
```