HW08 - Chance Variability

Stat 20 & 131A, Spring 2017, Prof. Sanchez

Due Mar-16

1) One hundred draws are made at random with replacement from the box `[1, 2]`. $_{0.5pts}$

a. How small can the sum be?

b. How large can the sum be?

c. How many times do you expect the ticket `1` to turn up?

d. How many times do you expect the ticket `2` to turn up?

e. About how much do you expect the sum to be?

To help you understand this chance process, you can use the following R code that simulates repeating the process 1000 times (each time 100 tickets are drawn out of the box). And then plotting a histogram for the sum of the tickets.

```
# box
box = c(rep(1, 50), rep(2, 50))

# 1000 repetitions each of which involves drawing 100 tickets
# (in each repetition the sum of tickets is calculated)
repetitions = 1000
sum_tickets = rep(0, repetitions)

set.seed(12017)
for (i in 1:repetitions) {
  # 100 draws
  draws = sample(box, size = 100, replace = TRUE)
  # sum of tickets
  sum_tickets[i] = sum(draws)
}

# histogram for sum of tickets
hist(sum_tickets, breaks = 30, col = 'gray80', las = 1)
```

2) Fifty draws will be made at random with replacement from one of the two boxes shown below.

`i)  [-1, 2]`

`ii) [-1, -1, 2]`

On each draw, you will be paid in dollars the amount shown on the ticket: if a negative number is drawn, that amount will be taken away from you. To help you get an idea of the possible results,

here's some R code that simulates the previous process (re-run the code in order to draw samples a few more times):

```
# boxes
box1 = c(-1, 2)
box2 = c(-1, -1, 2)

# 50 draws
set.seed(2017)
draws1 = sample(box1, size = 50, replace = TRUE)
draws2 = sample(box2, size = 50, replace = TRUE)

# paid amount
sum(draws1)
sum(draws2)
```

Which box is better? Or are they the same? Explain. $_{0.25pts}$

3) A box contains 10,000 tickets: 4000 `0`'s and 6,000 `1`'s. 10,000 draws will be made at random with replacement from this box. And the number of `1`'s is calculated. Which of the following best describes the situation, and why? $_{0.25pts}$

i. The number of `1`'s will be 6,000 exactly.

ii. The number of `1`'s is very likely to equal 6,000, but there is also some small chance that it will not be equal to 6,000.

iii. The number of `1`'s is likely to be different from 6,000, but the difference is likely to be small compared to 10,000.

To help you understand this chance process, you can use the following R code that simulates repeating the process 1000 times (each time 10,000 tickets are drawn). And then plotting a histogram for the number of 1's.

```
# box
box = c(rep(0, 4000), rep(1, 6000))

# 1000 repetitions each of which involves drawing 10,000 tickets
# (in each repetition the number of 1's is calculated)
repetitions = 1000
ones = rep(0, repetitions)

set.seed(42017)
for (i in 1:repetitions) {
  # 10,000 draws
  draws = sample(box, size = 10000, replace = TRUE)
  # how many 1's
  ones[i] = sum(draws)
}
```

```r
# histogram for number of 1's
hist(ones, breaks = 25, col = 'gray80', las = 1)
```

4) A die will be rolled some number of times, and you win \$1 under the following conditions: $_{1pt}$

a. You win \$1 if it shows an ace more than 20% of the time. Which is better: 60 rolls, or 600 rolls? Explain.

b. You win \$1 if the percentage of aces is more than 15%. Which is better: 60 rolls, or 600 rolls? Explain.

c. You win \$1 if the percentage of aces is between 15% and 20%. Which is better: 60 rolls, or 600 rolls? Explain.

d. You win \$1 if the percentage of aces is exactly $16\frac{2}{3}$. Which is better: 60 rolls, or 600 rolls? Explain.

5) A box contains red and blue marbles; there are more red marbles than blue ones. Marbles are drawn one at a time from the box, at random with replacement. You win a dollar if a red marble is drawn more often than a blue one. There are two choices:

- A. 100 draws are made from the box.
- B. 200 draws are made from the box.

Choose one of the four options below; explain your answer. $_{0.25pts}$

i. A gives a better chance of winning.

ii. B gives a better chance of winning.

iii. A and B give the same chance of winning.

iv. Can't tell without more information.

6) Find the expected value for the sum of 100 draws at random with replacement from each of the following boxes: $_{1pt}$

a. `[0, 1, 1, 6]`

b. `[-2, -1, 0, 2]`

c. `[-2, -1, 3]`

d. `[0, 1, 1]`

7) You gamble 100 times on the toss of a coin. If it lands heads, you win \$1. If it lands tails, you lose \$1. Your net gain will be around _____, give or take _____ or so. Fill in the blanks, using

the options, and show your work: $_{0.5pts}$

$$-\$10 \qquad -\$5 \qquad \$0 \qquad +\$5 \qquad +\$10$$

8) One hundred draws will be mdae at random with replacement from the box: $_{1pt}$

`[1, 1, 2, 2, 2, 4]`

a. The smallest the sum can be is _____

b. The largest the sum can be is _____

c. The sum of the draws will be around _____, give or take _____ or so.

d. The chance that the sum will be bigger than 250 is almost _____%.

To help you understand this chance process, you can use the following R code that simulates repeating the process 1000 times (each time 10,000 tickets are drawn). And then plotting a histogram for the number of 1's.

```r
# box
box = c(1, 1, 2, 2, 2, 4)

# 1000 repetitions each of which involves drawing 10,000 tickets
# (in each repetition the number of 1's is calculated)
repetitions = 1000
results = rep(0, repetitions)

set.seed(52017)
for (i in 1:repetitions) {
  # 100 draws
  draws = sample(box, size = 100, replace = TRUE)
  # how many 1's
  results[i] = sum(draws)
}

# histogram for number of 1's
hist(results, breaks = 25, col = 'gray80', las = 1)
```

9) You can draw either 10 times or 100 times at random with replacement from the box `[-2, 2]`. How many times should you draw: $_{0.75pts}$

a. To win $1 when the sum is 8 or more, and nothing otherwise? Explain your reasoning.

b. To win $1 when the sum is -8 or less, and nothing otherwise? Expain your reasoning.

c. To win $1 when the sum is between -8 and 8, and nothing otherwise? Expain your reasoning.

To help you understand both chance processes, you can use the following R code that simulates repeating each process 10,000 times. And then plotting a histogram for the sum of the tickets.

```r
# box
box = c(-2, 2)

# 10000 repetitions each of which involves drawing 10 tickets
# (in each repetition the sum of tickets is calculated)
repetitions = 10000
sum_tickets10 = rep(0, repetitions)

set.seed(12017)
for (i in 1:repetitions) {
  # 10 draws
  draws10 = sample(box, size = 10, replace = TRUE)
  # sum of tickets
  sum_tickets10[i] = sum(draws10)
}

# freq. histogram for sum of tickets
barplot(table(sum_tickets10), border = NA, las = 1)


# 10000 repetitions each of which involves drawing 100 tickets
# (in each repetition the sum of tickets is calculated)
repetitions = 10000
sum_tickets100 = rep(0, repetitions)

set.seed(12017)
for (i in 1:repetitions) {
  # 10 draws
  draws100 = sample(box, size = 100, replace = TRUE)
  # sum of tickets
  sum_tickets100[i] = sum(draws100)
}

# freq. histogram for sum of tickets
barplot(table(sum_tickets100), border = NA, las = 1)
```

10) A box contains 10 tickets. Each ticket is marked with a whole number between -4 and 4. The numbers are not all the same; their average equals 0. You have two choices:

i. 50 draws are made from the box, and you win \$10 if the sum is between -12 and 12.

ii. 100 draws are made from the box, and you win \$10 if the sum is between -24 and 24.

Choose one of the four options below; explain your reasoning. $_{0.5pts}$

a. (i) and (ii) give the same chance of winning.

b. (i) gives a better chance of winning.

c. (ii) gives a better chance of winning.

d. Can't tell without the SD of the box.

To help you understand this chance process, you can use the following R code that simulates repeating each process 10000 times (in one of them 50 tickets are drawn, in the other 100 tickets are drawn). And then plotting the frequency for the sums of the tickets.

```r
# box
box = rep(c(-4, 4), 5)

# 10000 repetitions each of which involves drawing 50 tickets
# (in each repetition the sum of tickets is calculated)
repetitions = 10000
sum_tickets50 = rep(0, repetitions)

set.seed(12017)
for (i in 1:repetitions) {
  # 50 draws
  draws50 = sample(box, size = 50, replace = TRUE)
  # sum of tickets
  sum_tickets50[i] = sum(draws50)
}

# freq. histogram for sum of tickets
barplot(table(sum_tickets50), border = NA, las = 1)


# 10000 repetitions each of which involves drawing 100 tickets
# (in each repetition the sum of tickets is calculated)
repetitions = 10000
sum_tickets100 = rep(0, repetitions)

set.seed(12017)
for (i in 1:repetitions) {
  # 50 draws
  draws100 = sample(box, size = 100, replace = TRUE)
  # sum of tickets
  sum_tickets100[i] = sum(draws100)
}

# freq. histogram for sum of tickets
barplot(table(sum_tickets100), border = NA, las = 1)
```

11) A coin is tossed 16 times. $_{0.75pts}$

a. The number of heads is like the sum of 16 draws made at random with replacement from one of the following boxes. Which one and why?

```
(i)    [head, tail]

(ii)   [0, 1]

(iii)  [0, 1, 1]
```

b. The number of heads will be around _____, give or take _____ or so. Show your work.

12) A large group of people get together. Each one rolls a die 180 times, and counts the number of 1's. About what percentage of these people should get counts in the range 15 to 45?. Show your work. $_{0.75pts}$

To help you understand this chance processes, you can use the following R code that simulates 10,000 repetitions of rolling a die 180 times. And then plotting a histogram for the number of aces.

```r
# 10000 repetitions each of which involves rolling a die 180 times
# and counting the number of aces
results = rep(0, 10000)

for (i in 1:10000) {
  rolls = sample(1:6, size = 180, replace = TRUE)
  results[i] = sum(rolls == 1)
}

barplot(table(results), col = 'gray80', border = NA, las = 1)
```

13) A coin is tossed 100 times. $_{0.75pts}$

a. The difference *"number of heads - number of tails"* is like the sum of 100 draws from one of the following boxes. Which one, and why?

```
(i)    [heads, tails]
(ii)   [-1, 1]
(iii)  [-1, 0]
(iv)   [0, 1]
(v)    [-1, 0, 1]
```

b. Find the expected value and standard error for the difference. Show your work.

14) A gambler plays roulette 100 times, betting a dollar on a column each time. The bet pays 2 to 1, and there are 12 chances in 38 to win. Fill in the blanks; $_{1.25pts}$

a. In 100 plays, the gambler's net gain will be aorund $_____, give or take $_____ or so.

b. In 100 plays, the gambler should win _____ times, give or take _____ or so.

c. How does the column bet compare with betting on a single number at Keno (example 1 on p. 289)?

15) A letter is drawn 1,000 times, at random, from the word `A R A B I A`. There are two offers.

a. You win a dollar if the number of `A`'s among the draws is 10 or more above the expected number.

b. You win a dollar if the number of `B`'s among the draws is 10 or more above the expected number.

Choose one option and explain. $_{0.5pts}$

i. (a) gives a better chance of winning than (b).

ii. (a) and (b) give the same chance of winning.

iii. (b) gives better chance of winning than (a).

iv. There is not enough information to decide.

To help you understand this chance processes, you can use the following R code that simulates 10,000 repetitions of drawing a letter from the word `ARABIA`. The first set of repetitions involves counting the number of `A`'s. The second set of repetitions involves counting the number of `B`'s. Also, histograms for number of `A`'s and number of `B`'s are produced.

```
word = c('A', 'R', 'A', 'B', 'I', 'A')

# number of A's
set.seed(72017)
number_As = rep(0, 10000)
for (i in 1:10000) {
  draws = sample(word, size = 1000, replace = TRUE)
  number_As[i] = sum(draws == 'A')
}
hist(number_As, breaks = 18, col = 'gray80', las = 1)


# number of B's
set.seed(72017)
number_Bs = rep(0, 10000)
for (i in 1:10000) {
  draws = sample(word, size = 1000, replace = TRUE)
  number_Bs[i] = sum(draws == 'B')
}
hist(number_Bs, breaks = 18, col = 'gray80', las = 1)
```