

MATH 308 Assignment 12: Sequences in Coin Flips

Nakul Joshi

March 26, 2014

1 Expected Number of Occurences 2 Probability of Occurence

For each coin flip i , let X_i be the event that the sequence 01111111 occurs starting at position i . Then:

$$\mathbb{E}(\mathbb{1}(X_i)) = \begin{cases} 2^{-8}, & \text{if } 1 \leq i \leq 93 \\ 0, & \text{if } i > 93 \end{cases}$$

So, the expected number of occurences is:

$$\begin{aligned} N &= \mathbb{E} \left(\sum_{i=1}^{100} \mathbb{1}(X_i) \right) \\ &= \sum_{i=1}^{100} \mathbb{E}(\mathbb{1}(X_i)) \\ &= \sum_{i=1}^{93} \mathbb{E}(\mathbb{1}(X_i)) \\ &= 93/2^8 \approx 0.363 \end{aligned}$$

Using the result above, we can approximate the coin-flipping experiment as a Poisson process with rate parameter $\lambda = 93/2^8$.

The requisite probability can then be calculated as `1-ppois(0,93/2^8)`, which comes out to ≈ 0.30 , which is slightly under the result from Assignment 2.

MATH 308 Assignment 15: Exercises 5.10

Nakul Joshi

March 26, 2014

5

a)

Suppose that cookie orders are represented such that, for example, an order of 1 sugar, 2 chocolate chips, 0 oatmeal, 0 peanut butter, and 2 ginger snaps is given by $|c|cc|cc|$. Then, two of the bars are fixed, giving the string $|??????|$, where the nine ? symbols are to be replaced by five c 's and two $|$'s. If we choose 5 distinct positions in which to place the c 's, the positions of the remaining $|$'s is fixed. Thus, the number of possible cookie orders is $\binom{9}{5}$.

b)

More generally, we can use 'cookie types' to represent occurrences of particular elements in the chosen sets. Then, fixing the position of two $|$'s leaves $n + 1 - 2 = n - 1$ $|$'s and n c 's, for a total of $2n - 1$ symbols. Picking element members requires n position choices, giving the total number of set choices as $\binom{2n-1}{n}$.

c)

For a sample S of size n , each bootstrap sample represents the choice of n elements from S with replacement. Thus, the number of bootstrap samples is as above $\binom{2n-1}{n}$.

6

a)

The number of bootstrap samples with k_1 a_1 's, k_2 a_2 's $\dots k_n$ a_n 's is the same as the number of permutations of the string

$$\underbrace{\overbrace{a_1 a_1 \dots a_1}^{k_1 \text{ times}} \overbrace{a_2 a_2 \dots a_2}^{k_2 \text{ times}} \dots \overbrace{a_n a_n \dots a_n}^{k_n \text{ times}}}_{n \text{ symbols}}$$

Using the formula for string permutations with repeated characters, this yields

$$\frac{n!}{k_1! k_2! \dots k_n!}$$

which, by definition, equals

$$\binom{n}{k_1, k_2 \dots k_n} \quad \square$$

b)

A bootstrap sample with k_i a_i 's requires choosing k_i of n positions, after which the remaining $n - k_i$ positions can be filled by any of the remaining $n - 1$ elements. Thus, the number of such samples is:

$$(n - 1)^{n - k_i} \binom{n}{k_i}$$

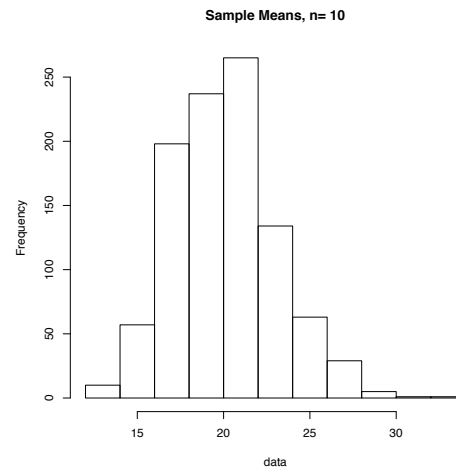
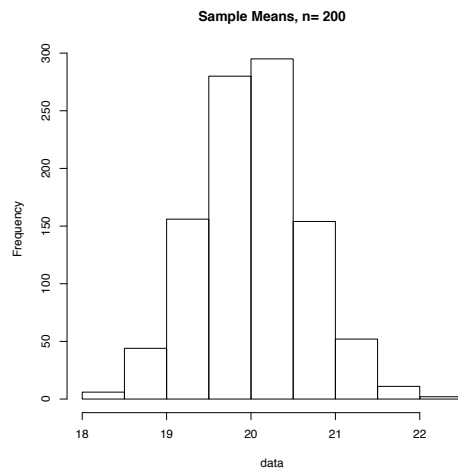
The total number of bootstrap samples is simply n^n . Thus, the required probability is

$$\frac{(n - 1)^{n - k_i}}{n^n} \binom{n}{k_i}$$

8

a)

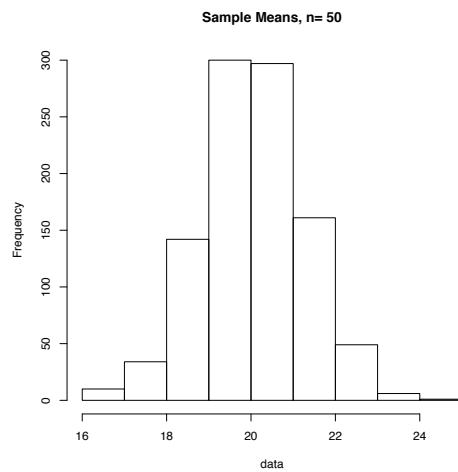
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|-------|
| 16.19 | 19.26 | 20.03 | 20.03 | 20.87 | 24.43 |



| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|-------|
| 18.21 | 19.58 | 20.02 | 20.02 | 20.44 | 22.49 |

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|-------|
| 12.56 | 17.87 | 19.99 | 20.05 | 21.89 | 33.20 |

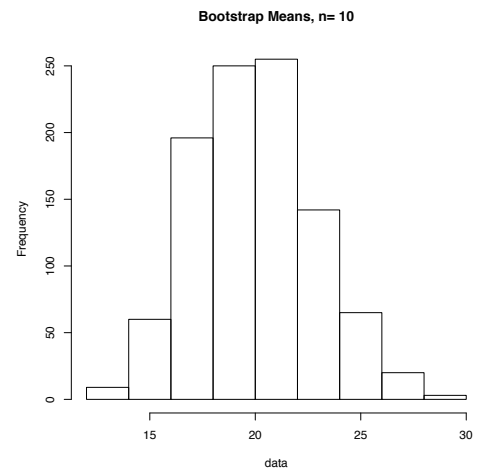
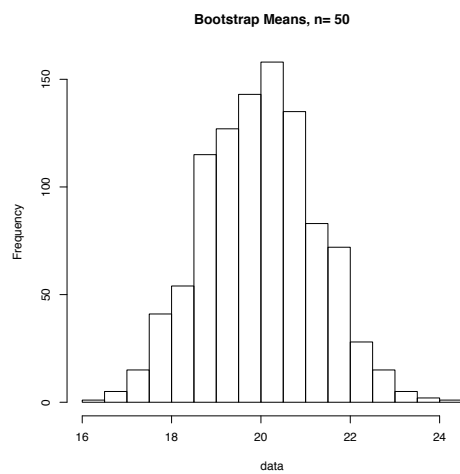
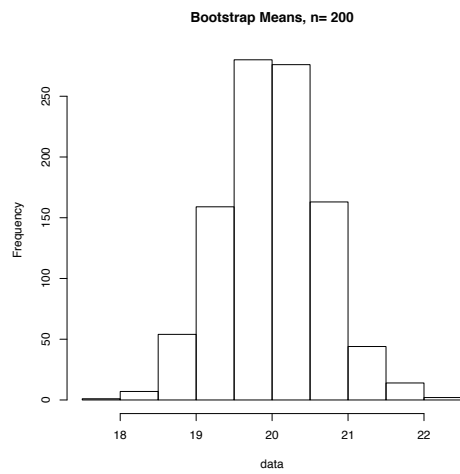
b)



| n | Mean | SD |
|-----|-------|------|
| 200 | 20.00 | 8.94 |
| 50 | 19.98 | 8.98 |
| 10 | 19.98 | 9.00 |

c)

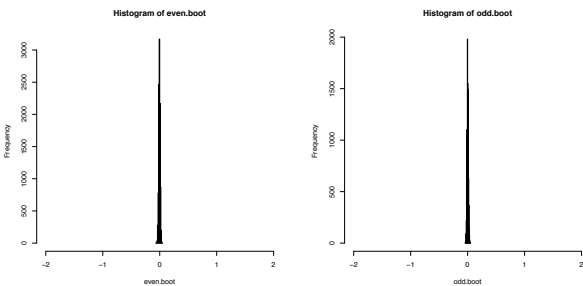
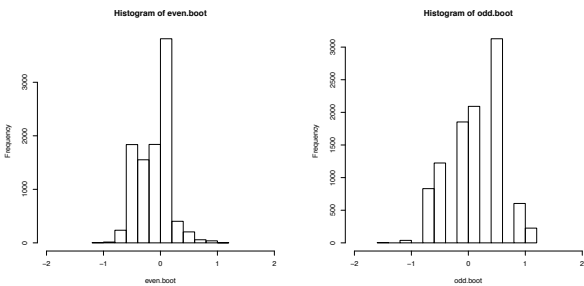
| n | Mean | SE |
|-----|-------|------|
| 200 | 20.00 | 0.65 |
| 50 | 19.98 | 1.25 |
| 10 | 19.98 | 2.82 |



e)

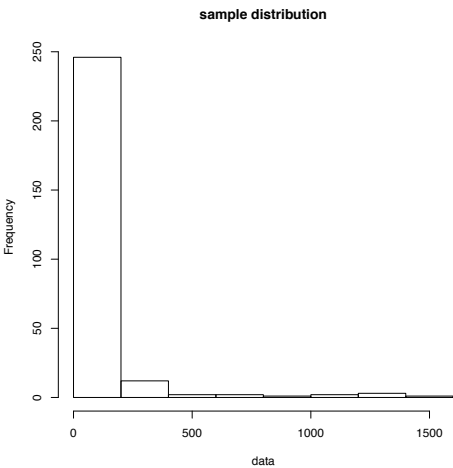
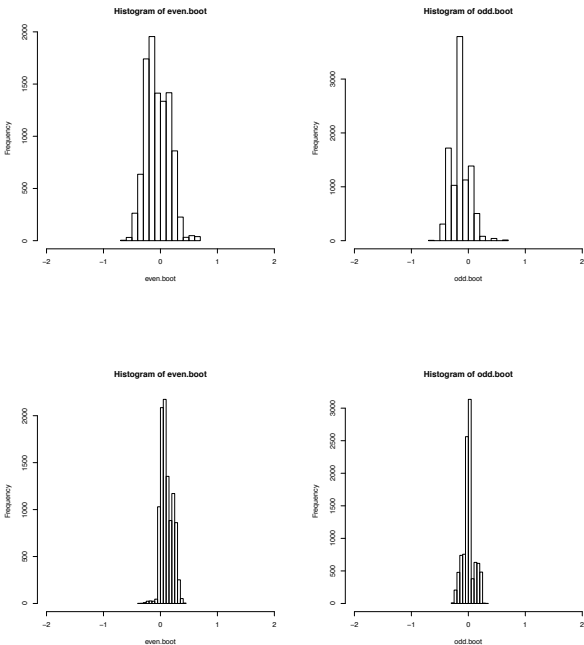
Increasing sample size reduces the bootstrap standard error.

9



10

a)



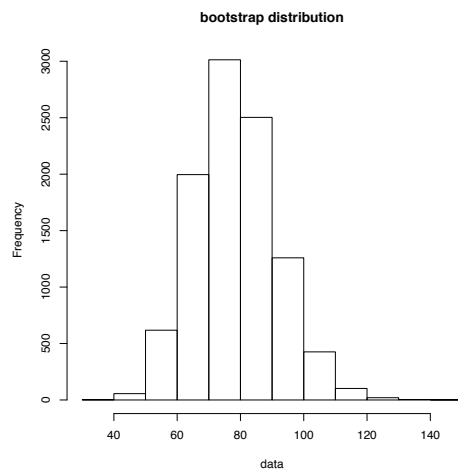
The evenness of N does not seem to affect the histograms, but increasing N reduces the spread.

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|---------|--------|-------|---------|---------|
| 1.00 | 5.00 | 14.20 | 78.08 | 55.50 | 1550.00 |

The data appears to be highly concentrated between 0 and 500, with several outliers that skew the mean.

b)

Mean: 78.03
95% confidence interval: 55.0 – 104.61



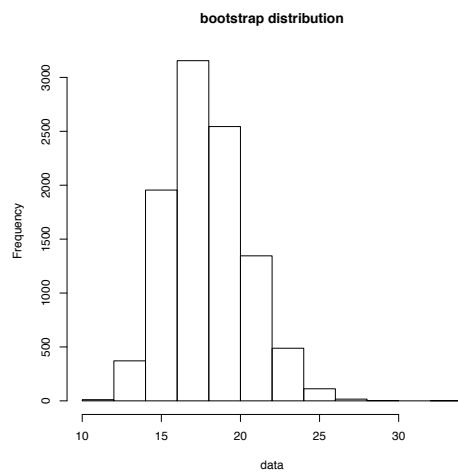
12

14

c)

Bias is 0.37, which is $\approx 3\%$ of the SE.

11



Trimmed Mean: 17.90

95% confidence interval: 13.67 – 23.30

Bias is 0.27, which is $\approx 11\%$ of the SE. The new test statistic has a smaller 95% confidence interval at the cost of higher bias.