Bookmarks

- Module 1: The Basics of R and Introduction to the Course
- Entrance Survey
- Module 2: Fundamentals of Probability, Random Variables, Distributions, and Joint Distributions
- Module 3: Gathering and Collecting Data, Ethics, and Kernel Density Estimates
- Module 4: Joint, Marginal, and Conditional Distributions & Functions of Random Variable

Questions 8 - 15

Bookmark this page

Now we are going to perform some simulations in R. We are going to follow Sara's example in the lecture where we imagine a case where the x_i follow a uniform distribution between 0 and θ $(U[0,\theta])$, and a researcher is trying to figure out the value of θ. (We will set $\theta = 5$). We are going to simulate different random samples from this distribution with a sample size of 100 observations each. These samples will be available to the two researchers, and we are going to plot how $\hat{\theta}$ is distributed for different estimators. There are two types of researchers in this world: researcher A uses as an estimator for θ, $\hat{\theta}_A = 2 * \overline{x}$, where $\overline{x}$ corresponds to the sample mean of the sample he receives from us. Researcher B uses as an estimator $\hat{\theta}_B = 2 * \text{median}(x)$, where $\text{median}(x)$ corresponds to the median of the sample he receives from us. We have provided you with this R code that has some information missing in case you need help for the exercise.

Question 8

1/1 point (graded)

What would be the mean of this distribution of $\overline{x}$?

2.5

✔ **Answer:** 2.5

2.5

Explanation

As it was discussed in the lecture, we have that $\overline{x} = \frac{\sum x_i}{n}$, then:

$$\mathbb{E}[\overline{x}] = \frac{1}{n}\sum \mathbb{E}[x_i] = \frac{n\mu}{n} = \mu$$

In this case since θ is equal to 5, we know that $\mu = \frac{\theta}{2} = 2.5$.

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

Question 9

1/1 point (graded)

What would be the variance of the distribution of $\hat{\theta}_A$?

Please round your answer to the third decimal place (i.e. if you are answer is 0.211111, you would round it to 0.211).

0.083 ✔ **Answer:** 0.083

0.083

Explanation

We know that the variance of $\overline{x} = \frac{\sigma^2}{n}$, where σ^2 corresponds to the variance of x_i and n to the sample size. For a uniform distribution between 0 and θ, we have that the variance is given by $\frac{\theta^2}{12}$. In this case, then we know that $var(\overline{x}) = \frac{25}{12*100} = 0.02083333$. Then, we have that:

$$var(\hat{\theta}_A) = var(2 * \overline{x}) = 4 * var(\overline{x}) = 4 * 0.02083333 = 0.083333$$

Submit You have used 2 of 2 attempts

✔ Correct (1/1 point)

Question 10

1/1 point (graded)

Can we use the CLT discussed in the lecture to infer what would be the distribution of $\hat{\theta}_A$, when $n \to \infty$?

- (•) Yes ✔
- ( ) No

Explanation

From the CLT we know that the distribution of $\overline{x}$ is approximately normal. We have that $\hat{\theta}_A$ is a linear transformation of a normal distributed variable, then we can also infer that it is normal.

Submit You have used 1 of 1 attempts

✔ Correct (1/1 point)

Question 11

1/1 point (graded)

Can we use the CLT discussed in the lecture to infer what would be the distribution of $\hat{\theta}_B$, when $n \to \infty$?

- Yes
- ◉ No ✔

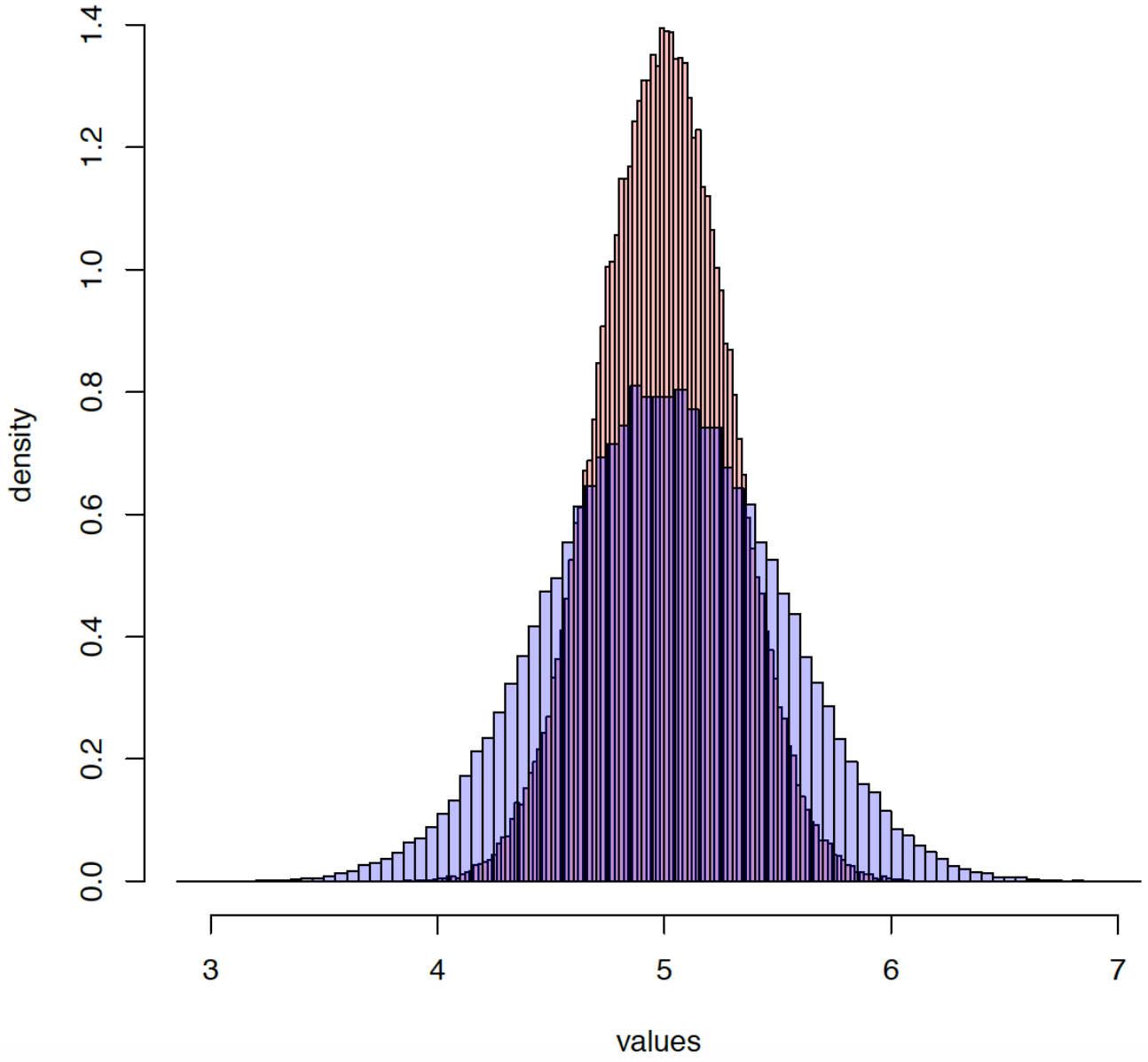
Explanation

No, the CLT we discussed in the lecture was about the distribution of the sample mean. We haven't discussed the distribution of the median of a sample.

Submit You have used 1 of 1 attempts

✔ Correct (1/1 point)

We have run our simulations, simulating 100,000 different samples of size 100. We have provided 200,000 researchers (A and B), each with one of these samples. They have sent us their estimators for $\hat{\theta}$. The following plot shows a histogram of their estimators (Figure 1).


density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
3
4
5
6
7
values

Question 12

1/1 point (graded)

Does the blue histogram correspond to the estimator of researcher A or researcher B?

- a. Researcher A
- b. Researcher B ✔

Explanation

When you fill in the missing parts of the code we sent you and run the simulations, you would realize that the blue distribution corresponds to the distribution of the median of the 100,000 simulations we have run for this exercise.

Submit You have used 1 of 1 attempts

✔ Correct (1/1 point)

Question 13

1/1 point (graded)

As an intuitive answer which one of the two estimators would you prefer to use for the parameter θ, $\hat{\theta}_A$ or $\hat{\theta}_B$?

- a. I would prefer $\hat{\theta}_A$ ✔

b. I would prefer $\hat{\theta}_B$

Explanation
Both estimators seem to be centered around the real value of the parameter θ that is 5. Then, just intuitively you would prefer the one with the lowest variance. This estimator between these two corresponds to $\hat{\theta}_A$.

Submit You have used 1 of 1 attempts

✔ Correct (1/1 point)

Question 14

1/1 point (graded)
Both estimators seem to be centered around the real value of the parameter θ that is 5. Then, just intuitively you would prefer the one with the lowest variance. This estimator between these two corresponds to $\hat{\theta}_A$.

Now, let's increase the sample size to 1000. As an exercise try to use the provided code to code this yourself in R. What would be the new variance of the estimator $\hat{\theta}_A$?

Please round your answer to four decimal points (i.e. if it is 0.41111, please round to 0.4111).

0.0083 ✔ **Answer:** 0.0083

0.0083

Explanation

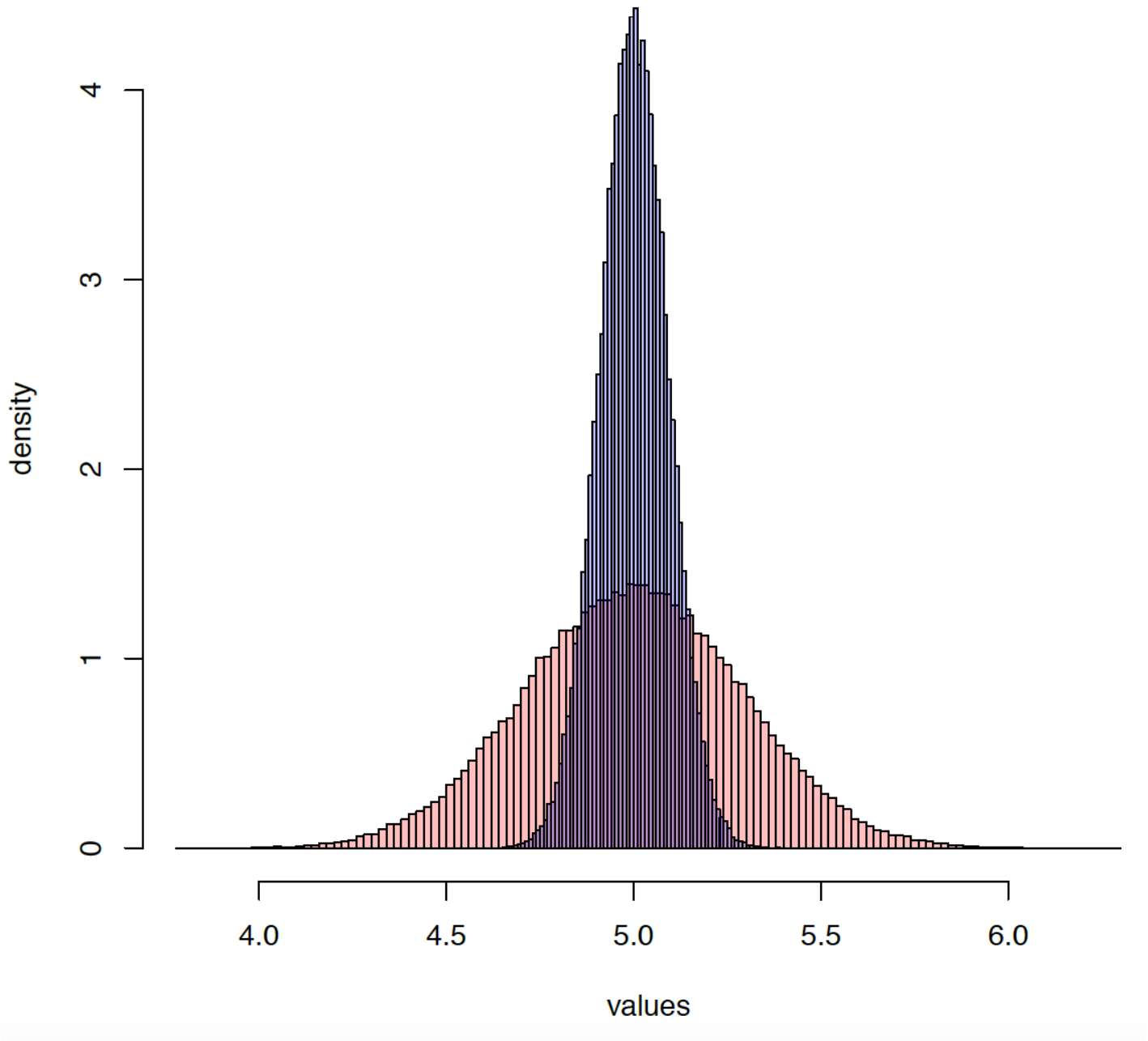
We know that the variance of $\overline{x} = \frac{\sigma^2}{n}$, where σ^2 corresponds to the variance of x_i and n to the sample size. For a uniform distribution between 0 and θ, we have that the variance is given by $\frac{\theta^2}{12}$. In this case, then we know that $var(\overline{x}) = \frac{25}{12*1000} = 0.002083333$. Then, we have that:

$$var(\hat{\theta}_A) = var(2 * \overline{x}) = 4 * var(\overline{x}) = 4 * 0.002083333 = 0.0083333$$

Submit You have used 1 of 2 attempts

✔ Correct (1/1 point)

The following figure shows the distribution for $\hat{\theta}_A$ for $n = 100$, and $n = 1000$


density
0
1
2
3
4
4.0
4.5
5.0
5.5
6.0
values

Question 15

1/1 point (graded)

Does the blue histogram corresponds to a sample size of 100 or of 1000?

- a. 100
- b. 1000 ✔

Explanation

We know that the variance of the estimator when $n = 1000$ is lower. Then, it must correspond to the blue histogram.

Submit You have used 1 of 1 attempts

✔ Correct (1/1 point)