SD - Aug 31

Srijan Kundu

(M.Sc. Data Science - 423)

September 4, 2022

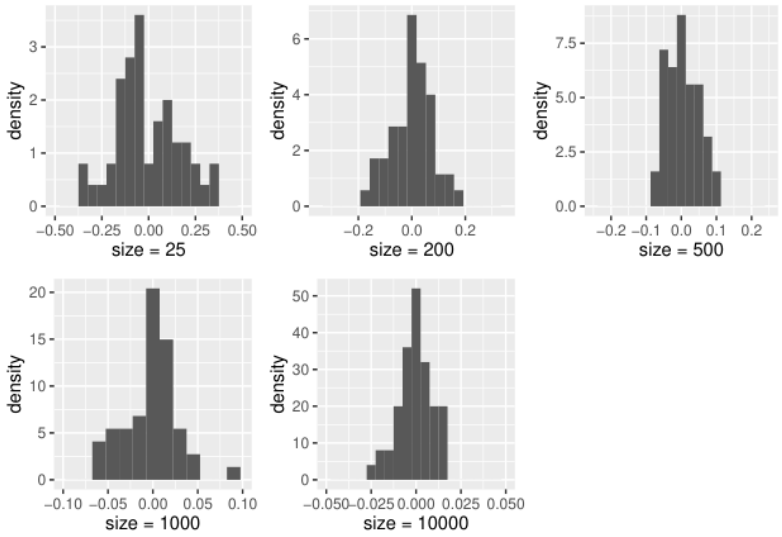
Question

Draw 50 random samples from the standard Normal distribution of sizes:

(i) 25, (ii) 200, (iii) 500, (iv) 1000, (v) 10000;

Obtain the histograms of the frequency distribution of the 'sample mean' for each of these 5 cases. Comment on your findings.

Answer



Explanation:

The 5 histograms of the 5 different cases are shown above.

As we can see in the histograms, the larger the size of the sample, the lesser the 'spread' of the histogram is.

If a random sample $(X_1, X_2, ..., X_n)$ of size n is drawn from a $N(\mu, \sigma^2)$ population, the sample mean $\bar{X}$ will follow a $N\left(\mu, \frac{\sigma^2}{n}\right)$ distribution. Hence, with increase in sample size n, the value of $\frac{\sigma^2}{n}$, which is the variability of the sample mean, eventually decreases.

Here, we have taken samples from the standard normal distribution.

1. For the first case, the sample size is 25, implying 95% of the samples should fall within the range: $(0 \pm 2 * \frac{1}{\sqrt{25}}) = (0 \pm 0.4)$.

2. For the second case, the sample size is 200, implying 95% of the samples should fall within the range: $(0 \pm 2 * \frac{1}{\sqrt{200}}) = (0 \pm 0.14)$.

3. For the third case, the sample size is 500, implying 95% of the samples should fall within the range: $(0 \pm 2 * \frac{1}{\sqrt{500}}) = (0 \pm 0.089)$.

4. For the fourth case, the sample size is 1000, implying 95% of the samples should fall within the range: $(0 \pm 2 * \frac{1}{\sqrt{1000}}) = (0 \pm 0.06)$.

5. For the fifth case, the sample size is 10000, implying 95% of the samples should fall within the range: $(0 \pm 2 * \frac{1}{\sqrt{10000}}) = (0 \pm 0.02)$.

This is exactly the case we can observe in the histograms above.