Spring 2025

Math 408 - Computer Project #1

Instructor: Sergey Lototsky

Due Date: Friday, March 14 (upload to Gradescope as a single PDF file.)

You are welcome to use any programming environment and any help from any source

1. Simulate 100 random normal variables with mean 1.5 and standard deviation 2. Create a histogram which shows the data in the range between -4 and 10, with class intervals of length 0.5

2. Find out how many of your 100 variables are *bigger* than 0, and estimate the probability that a single normal variable with mean 1.5 and standard deviation 2 is above 0. Indicate this estimate on your histogram, and show which area is being used as the estimate. Also, compute and write the theoretical value for this probability.

3. Simulate 100 sample means, each made by taking the average of 20 normal variables with mean 1.5 and standard deviation 2. Make a histogram of these sample means in the range -4 to 10 with class intervals of length 0.5. As in the previous question, estimate the probability that a sample mean is bigger than 0, indicate this estimate on your histogram, and show which area is being used as the estimate. Also, compute and write the theoretical value for this probability.

4. What do you notice about the two histograms and (therefore) about the two estimates of the probability of getting above 0?

5. Repeat items 1–3, but now generating 100 independent random variables of the form $1.5 + (2X/Y)$, where X and Y are independent standard normal random variables. Comment on the results and try to explain the differences compared to the normal case. [Hint: X/Y has the standard Cauchy distribution; $1.5 + (2X/Y)$ has Cauchy distribution with location parameter equal to 1.5 and the scale parameter equal to 2.]

Spring 2025

Math 408 - Computer Project #2

Instructor: Sergey Lototsky

Due Date: Friday, April 25 (upload to Gradescope as a single PDF file.)

You are welcome to use any programming environment and any help from any source

The goal of this project is to investigate various ways of combining information from independent trials.

Part I

Generate 100 independent normal random variables with mean zero and variance 1. Call it vector $V_1 = (x_1, \ldots, x_{100})$. Then generate 100 independent normal random variables with mean zero and variance 1.5. Call it vector $V_{1.5} = (y_1, \ldots, y_{100})$.

Now, how can you tell which vector is which if you only look at the components? Here are four possible ways.

Procedure A. If you are handed only a pair of numbers x_k and y_k without knowing which is which, procedure A is to guess that the number with the smaller absolute value came from vector V_1. Run this procedure on your data and determine how many times you get the correct conclusion. Then compute the theoretical value of the probability that procedure A gives the correct conclusion.

Procedure B. Suppose that instead of a pair of numbers (x_k, y_k), you have the entire collection of numbers, $(x_1, \ldots, x_{100})$ and $(y_1, \ldots, y_{100})$ but without knowing which collection is which. Procedure B says that the collection with the larger *sum of squares* is $V_{1.5}$. Apply procedure B to the data you generated. Does this procedure give the correct answer?

Procedure C. This is a more realistic version of Procedure B, when you pretend that you do not know that the mean in each sample is zero and say that the collection with the larger *sample standard deviation* is $V_{1.5}$. Apply procedure C to the data you generated. Does this procedure give the correct answer?

Procedure D. This is an even more realistic procedure, when you pretend that you do not know the distributions of the sample. Use the result of problem 6, part 4 in Homework number 12 to reduce the setting to a standard shift model, and then use the sign test and at least one other non-parametric test or estimation procedure to answer the question.

Generate five more collections of 100 independent normal random variables with mean zero and variance 1. Then generate five more collections of 100 independent normal random variables with mean zero and variance 1.5. Apply procedures B, C, and D to each of the five pairs. How many times did you get wrong answer?

Compute the theoretical value of the probabilities that procedures B, C, and D give correct answer.

Part II. Repeat Part I when the vector V_1 consists of independent standard Cauchy random variables, and $V_{1.5}$ consists of independent Cauchy random variables with location parameter equal to zero and the scale parameter equal to $\sqrt{1.5}$. Compare and contrast the results with what you got in Part I.[1]

The concluding comment: this project is not about estimation or hypothesis testing, but about *decision making*. The theory of decision making is yet another part of mathematical statistics.

[1]For example, Procedure A works, somewhat, in the Cauchy case: $P(|\text{Cauchy, scale}\sqrt{1.5}| > |\text{Cauchy, scale}\,1|) \approx 0.54$. You can use the note on the Cauchy distribution posted on the class web page under Other material → By other people.