In-Class Exercise 4

For this in-class exercise, work with your group of 2-3 people, to answer the following questions. These questions are not necessarily easy and sometimes they will not have a clear "correct" answer. The goal is to get you thinking about the material we've learned. Some of these questions may require you to discuss and debate with your group members to come up with an answer or can cover topics that we have not yet covered in class.

Be prepared to share your answers with the class and add to the discussion.

After class submit your a do-file with your answers in comments to Moodle for grading. You will be graded as a group on your submission. Only one group member needs to submit the assignment, but make sure add all group member names.

Deriving the Normal Distribution Qualitatively

For this exercise, we will derive the normal distribution "qualitatively." What does that mean? We will not venture to use calculus as it requires Taylor expansions and the intuition doesn't really come through. Instead, we will use reasoning and logic to arrive at the normal distribution.

The normal distribution is also called the Gaussian distribution, named after Carl Friedrich Gauss, who formulated it in the context of errors in astronomical observations. The normal distribution is defined by its probability density function:

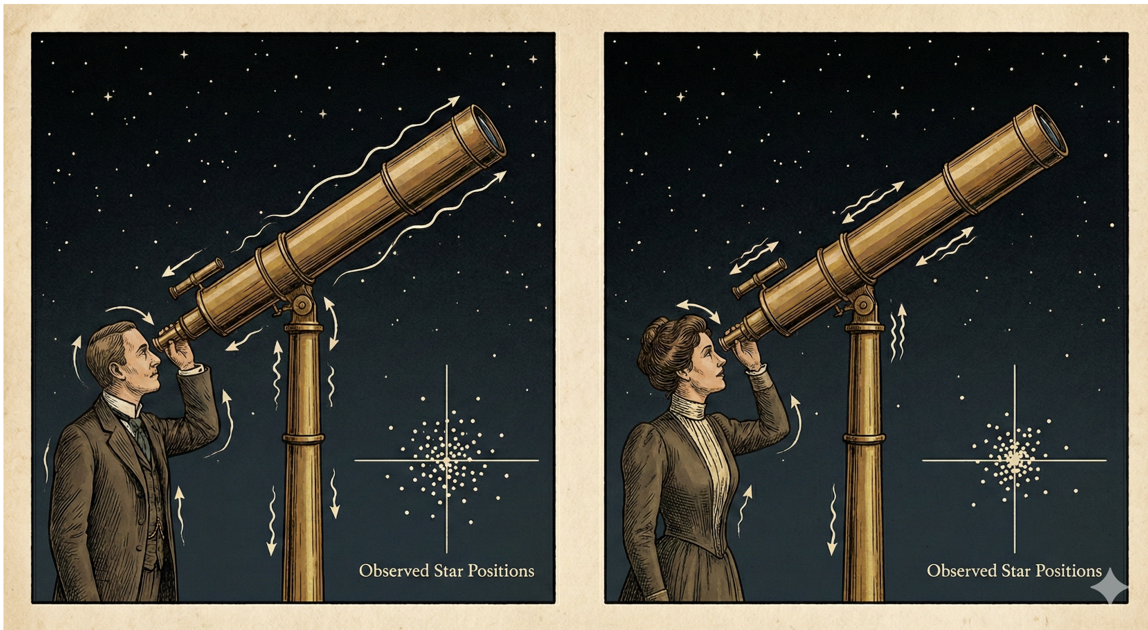
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean and σ is the standard deviation.

So in the spirit of Guass' brilliance, let's do this in the context of astronomy

Introduction

You are a 19th-century astronomer looking through a telescope at a distant star. Your goal is to record its exact position (Right Ascension) on a grid. However, the universe is messy. The star doesn't appear as a single, steady point. It "dances" slightly in the eyepiece due to various tiny disturbances.



Part 1: The Error Curve and Telescope Measurements

Imagine the "True Position" of the star is exactly at 0 on your grid.

But every time you take a measurement, 4 distinct physical factors distort the light slightly to the Left (-1 unit) or Right (+1 unit).

1. Atmosphere: A pocket of warm air passes by. (Left=-1, Right=+1)
2. Wind: A gust shakes the telescope tube. (Left=-1, Right=+1)
3. Gears: The clock-drive gear slips slightly. (Left=-1, Right=+1)

Question 1

If all 4 factors are independent and equally likely to push the measurement Left or Right, what is the probability distribution of your telescope measurements? In other words, for each possible way that the factor can push the telescope, what is the measured position of the telescope?

Use Stata to help you tally up the probabilities. In a do-file, open up a new window and create variables for `atmosphere`, `wind`, and `gears`, each taking on values of −1 or +1. You can then input each combination by using:

```
input
1 1 1
1 1 -1
...
end
```

Then create a new variable `gen measurement = atmosphere + wind + gears` to get the resulting measurement. Finally, use the `tab` command to tally up how many times each measurement occurs. Divide by the total number of combinations to get the probabilities.

> **i Hint**
>
> To make sure you get all combinations, calculate how many there are. Each factor has 2 possible states (Left or Right), and there are 3 factors.
>
> Think about how many possible way there are for the 3 factors to push Left or Right. For each possible combination, what is the resulting measurement? Tally up how many times each measurement occurs.

Question 2

Does the wind shaking the telescope (Factor 2) care about what the atmosphere (Factor 1) is doing? Why?

Question 2

Draw a histogram of the probability distribution you found in Question 1 with the `hist` command. What shape does it resemble? What do you think would happen if instead of 3 factors, we actually measured 1000 factors?

Question 3

1. Where is the highest peak of this curve? Why? Think about how the errors cancel each other out. Relate this to the concept of independence.
2. What can you say about the symmetry of the curve?
3. What about the tails of this curve? How likely is it that a measurement would be at +1,000 or −1,000? As you move away from the center, how quickly does the probability drop?

Part 2: Finding the Functional Form

Let's go through some functional forms and see if they fit the properties we found in Part 1.

Question 4: The Parabola

Consider the function:

$$f(x) = -x^2 + C$$

for some constant C.

Go through each of the properties you found in Part 1 (peak, symmetry, tails) and see if this function satisfies them. If it does not, explain why. You can use Stata to help you plot this function and see its shape. You can use the `twoway function` command to plot it. For example, if $C = 1$, you can use:

```
twoway function y=-x^2+1, range(-3 3)
```

Question 5: The Decay

Consider the function:

$$f(x) = \frac{1}{x^2}$$

Go through each of the properties you found in Part 1 (peak, symmetry, tails) and see if this function satisfies them. If it does not, explain why.

Question 6: The Exponential Decay

Consider the function:

$$f(x) = e^{-x}$$

Go through each of the properties you found in Part 1 (peak, symmetry, tails) and see if this function satisfies them. If it does not, explain why.

Question 7: The Bell

Consider the function:

$$f(x) = e^{-x^2}$$

Go through each of the properties you found in Part 1 (peak, symmetry, tails) and see if this function satisfies them. If it does not, explain why.

Part 3: The Correction Factor

Question 8

We found that the function $f(x) = e^{-x^2}$ satisfies all the properties we found in Part 1. However, let's compare it to the normal distribution function from above.

What does μ and σ have to be equal to in order for $f(x) = e^{-x^2}$ to be equivalent to the normal distribution function (excluding the scaling factor in front)?

Write down the normal distribution function with these values of μ and σ. What is the scaling factor?

Question 9

The reason that scaling factor is there is to ensure that the total area under the curve equals 1 (so that it is a valid probability distribution). To find that scaling factor, we would need to do some calculus (which we won't do here). However, look up the integral of the function e^{-x^2} from $-\infty$ to ∞. What is the value of that integral? For this we can use the `integrate` command in Stata, which needs to be first installed by running `ssc install integrate`.

Look at the help file by typing in `help integrate` to see how to use it. Integrate the function e^{-x^2} from $-\infty$ to ∞. What is the result? Why?