STAT157 HW 8

March 8, 2023

Due Wednesday, March 15 at 11:59pm

Deliberate Practice: Invalidating Considerations

Expected completion time: 120 minutes

This exercise follows the activity in last week's discussion. We will use the following simplified model for invalidating considerations:

- Suppose you want an 80% confidence interval around some quantity
- You have an initial distribution p_0 for what happens in a "normal" world. For example, maybe this is the distribution that you get by looking at a reference class.
- After brainstorming invalidating considerations, you realize that there is a small probability ϵ that you are instead in a "crazy" world, in which case your distribution is instead p_1.
- Your new probability distribution is therefore the mixture $(1 - \epsilon)p_0 + \epsilon p_1$.

For each of the following examples of p_0, p_1, and ϵ, do the following:

- Give an 80% confidence interval when accounting for the ϵ probability of a "crazy" world. Your interval should be centered, i.e. your lower and upper bounds should be the 10th and 90th percentiles of the mixture distribution.
- Confirm your reasoning by simulation: using Python, draw 1000 samples from $(1-\epsilon)p_0+\epsilon p_1$, and use them to estimate the 10th and 90th percentiles of this distribution. Sampling from a mixture happens in two steps:
  - Sample from a Bernoulli(ϵ) to choose between p_0 and p_1.
  - Sample from the chosen distribution (in this exercise the distributions will be uniform).

1. Non-overlapping uniform distributions

    - $p_0 = \text{Uniform}(0, 2)$
    - $p_1 = \text{Uniform}(2.5, 3)$
    - (a) $\epsilon = 0.05$, (b) $\epsilon = 0.1$, (c) $\epsilon = 0.2$

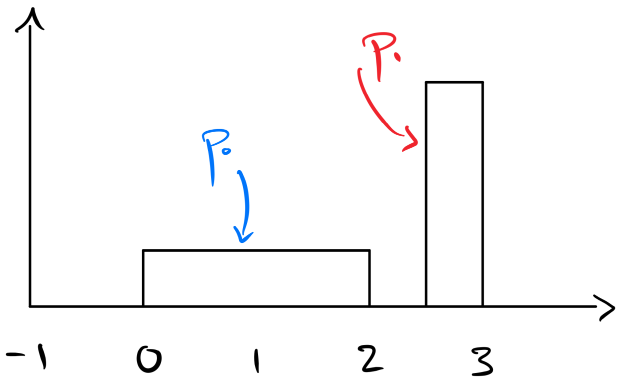

Figure 1: Non-overlapping uniform distributions

2. Overlapping uniform distributions

    - $p_0 = \text{Uniform}(0, 2)$
    - $p_1 = \text{Uniform}(1, 3)$
    - (a) $\epsilon = 0.05$, (b) $\epsilon = 0.1$, (c) $\epsilon = 0.2$

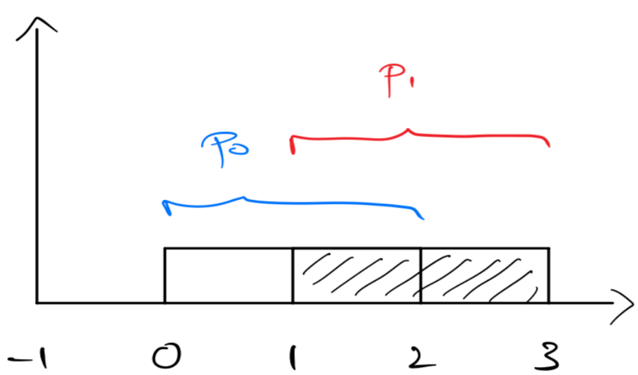

Figure 2: Overlapping uniform distributions

On Gradescope, please also submit the time it took to complete this exercise.

Deliberate Practice: Numerical Sensitivity

Expected completion time: 180 minutes

Consider the formula for the number of days T until the peak of Omicron that we used in the *Turning Considerations into Probabilities* lecture:

$$T = \log_2(N/2N_0) \cdot t + \Delta_0 + \Delta_1,$$

where:

- N is the total number of future UK Omicron cases
- N_0 is the current number of UK Omicron cases
- t is the Omicron doubling time
- Δ_0 is the lag between case peak and hospital peak
- Δ_1 is the lag between single-day hospital peak and 7-day average hospital peak

Using simulations, we will assess the sensitivity of this formula to variations of its five inputs.

1. Suppose the inputs are sampled independently from Normal and LogNormal distributions. For this, sample α, β, γ, and δ independently from Normal(0, 1), and define the inputs N, N_0, t, and Δ as follows:
   - $N = \exp(15.57 + 0.30 \cdot \alpha)$
     - This means that N follows a LogNormal distribution with mean 6.7×10^6 and standard deviation 4×10^6.
   - $N_0 = \exp(12.18 + 0.06 \cdot \beta)$
     - This means that N_0 follows a LogNormal distribution with mean 0.2×10^6 and standard deviation 0.05×10^6.
   - $t = 2.4 + 0.5 \cdot \gamma$
   - $\Delta = \Delta_0 + \Delta_1 = 12 + 3 \cdot \delta$.

   Sample the inputs in this way 1000 times, computing each time the corresponding value of T using the formula above. Plot the histogram of the distribution of T: what are its 10th, 25th, 50th, 75th, and 90th percentiles?

2. Double the standard deviation of N while leaving the other distributions constant: for this, sample α from Normal(0, 4) rather than Normal(0, 1). Do the same but with Δ_0: double its standard deviation, while leaving the other distributions constant. Which of these leads to the greatest increase in the variance of T?

3. Replace the Normal distributions by Student's t distributions: instead of sampling α, β, γ and δ from Normal(0, 1), sample them from Student($\nu = 5$). Plot the histogram of the distribution of T: how did using a t distribution instead of a Normal change the quantiles and variance of T?

4. Repeat your simulation with $\nu = 3$, $\nu = 7$ and $\nu = 9$: how does the degrees of freedom parameter ν affect the distance between the 10th and 90th percentiles of T?

5. Suppose now that instead of being independent, the inputs are correlated. That is, sample $[\alpha, \beta, \gamma, \delta]^\top$ from a multivariate Gaussian distribution:

$$\begin{bmatrix}\alpha \\ \beta \\ \gamma \\ \delta\end{bmatrix} \sim \text{Normal}(\mathbf{0}, \mathbf{\Sigma}), \text{ with } \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}, \text{ and } \rho \in [-1, 1].$$

   Draw 1000 samples, and plot a histogram of the values of T. Do the 10th and 90th quantiles get closer together or farther apart as the correlation parameter ρ increases from 0 to 1? as ρ decreases from 0 to -1?

On Gradescope, please also submit the time it took to complete this exercise.

Predictions

Expected completion time: 120 minutes

Register the following predictions. You can submit them by going to this URL and following the form's instructions. For these predictions, (and all predictions about the future throughout this class), we encourage you to use external sources – by googling things, reading news articles, talking to friends who follow politics or music stats, etc.

1. Will the Montreal REM line between Brossard and Central Station open to the public before May 10th, 2023? This question will resolve based on whether the wikipedia article for Brossard station mentions that the station has opened. The wikipedia article should also cite an article that shows that the station has indeed opened.

2. How many NYT headlines/ledes in the New York section dated between Mar 16 and Mar 23 (included) will include Eric Adams' name?

3. Will California Assembly Bill 1700 pass the California Assembly by May 10th, 2023?

For each question, submit an inclusive 80% confidence interval or your probabilities, as well as an explanation of your reasoning (1-2 paragraphs). **Please include a copy of your google form responses with your Gradescope submission.** On Gradescope, please also submit the time it took to complete this exercise.