GR5224 Final Exam

- You can use all of the course materials, readings, etc. to complete this exam. The only thing that is prohibited is communicating with other people (besides Ben) about the exam before everyone has finished (not all students are taking the exam in the same window). While the use of chatbots is permitted, you have to document your use of them as outlined on the syllabus.
- You have 1 hour and 30 minutes to upload your .qmd or .ipynb file to the Assignments section of Canvas. You can subsequently upload the .pdf file after it finishes rendering.
- Each of the two main problems is weighted equally and each of the subproblems within each problem is weighted equally. Thus, you should try all of the subproblems so that you can at least receive partial credit.

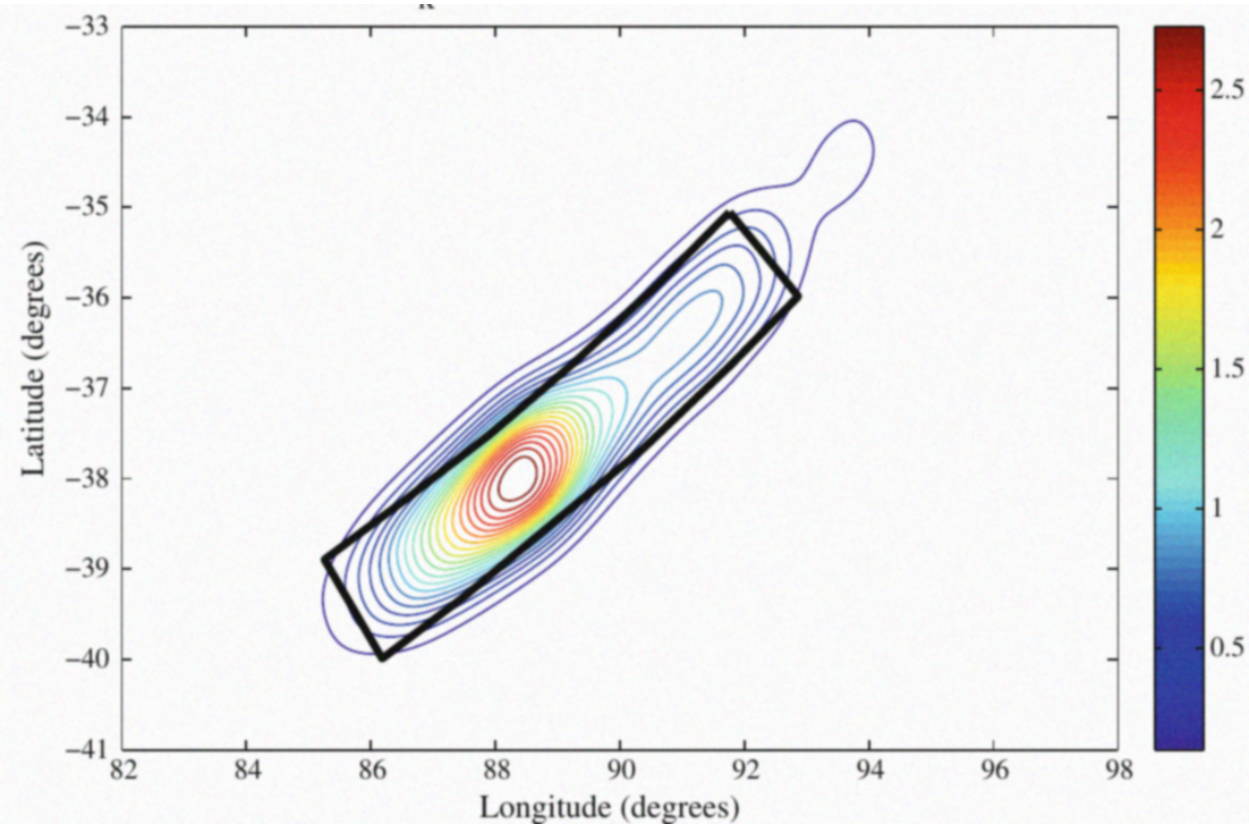
1 Malaysia Airlines Flight 370

On March 8, 2014 a Malaysia Airlines flight from Malaysia to Beijing turned south shortly after takeoff for some reason and has not been found since. If you do not remember this incident, you can glance at

https://en.wikipedia.org/wiki/Malaysia_Airlines_Flight_370

1.1 Prior Distribution

After not immediately finding the plane, Australia asked a bunch of Bayesians where it is likely to be. They said in 2016 that based on the last known trajectory of the plane and projections as to when it would run out of fuel if it mostly followed a straight line, the predictive density looked like this:



What bivariate normal distribution where x is Longitude and y is Latitude is consistent with the above image? The rectangle in black contains about 85% of this distribution, but you would need to utilize a function to evaluate the bivariate normal CDF to match that probability.

1.2 Fuel Update

In 2023, some people who seem knowledgeable about airplanes subsequently said that it was unlikely that the plane could have reached the southern half of the rectangle before running out of fuel. Let's stipulate that anything south of -37.5 degrees latitude is impossible.

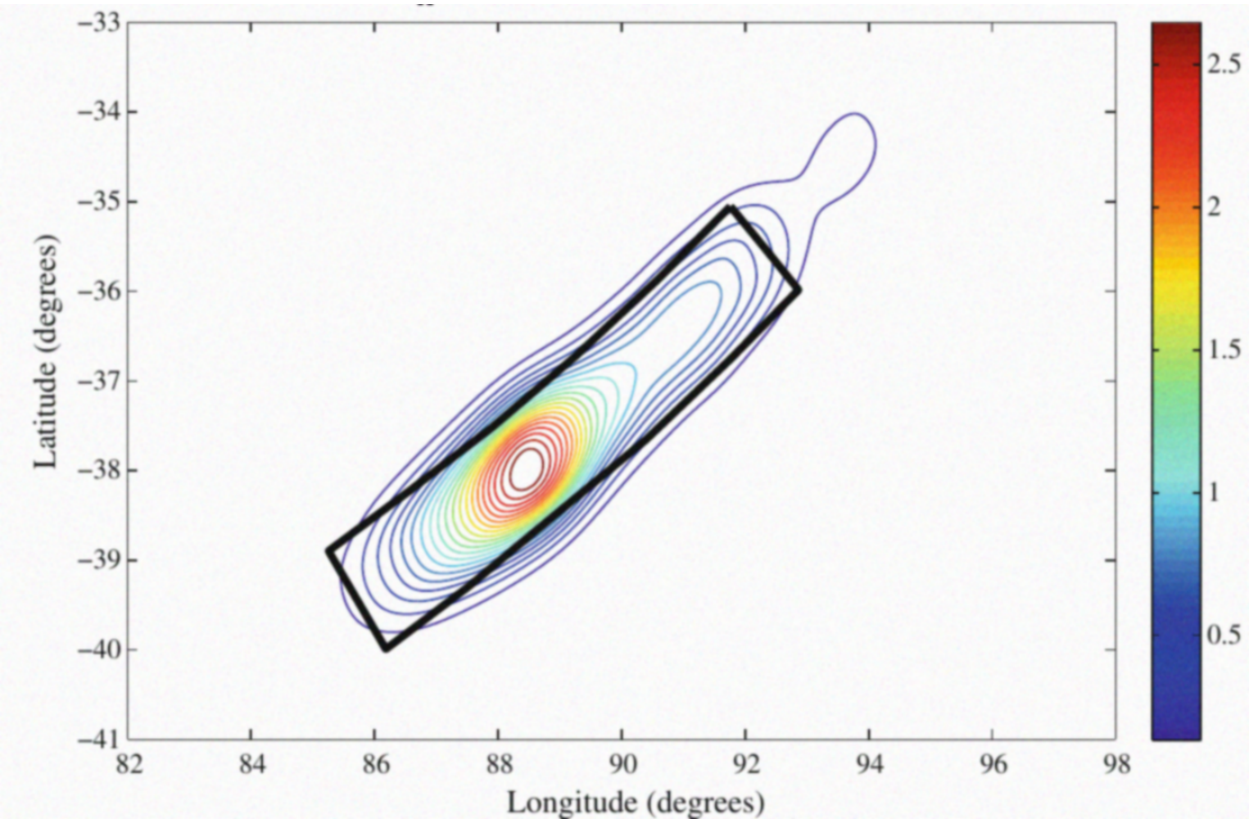
Draw many times from the bivariate normal distribution in the previous subproblem and filter to include only those draws north of -37.5 degrees latitude. How would you describe the new distribution of where the plane is likely to be?

1.3 Satellite Update

In 2017, satellite images that were taken in March 2014 but not analyzed at the time appeared to show multiple objects that could be from the plane. Some analyst said that these satellite images implied with "unprecedented precision and certainty" that the plane crashed near 92.8 degrees longitude and −35.6 degrees latitude. Explain why a point estimate, such as this one, is not very helpful in this context.

1.4 Debris Update

Over a year later, parts of the plane started to wash ashore on the eastern coast of Africa and the islands of Reunion and Madagascar. The ocean currents toward the northeast of the rectangle run west toward Africa. If the plane crashed farther south or east, then the ocean currents would have taken parts of the plane east until they washed ashore on the west coast of Australia but none have been found there. Based upon the debris and the currents, the original Bayesians updated their predictive distribution to



Qualitatively speaking, how informative was the fact that parts of the plane washed ashore in eastern Africa?

1.5 Radio Update

There is a radio protocol called Weak Signal Propagation Reporter (WSPR) that is used by amateur radio enthusiasts to communicate across long distances with low power. Some scientists think that a plane would distort these radio waves enough to detect the end of the flight path. Other scientists disagree that it would yield meaningfully precise estimates.

Based on the radio wave data and evolving discussions of how best to utilize it, an engineer has variously said that

1. In November 2021: near 98 degrees longitude and −33 degrees latitude
2. In mid-2022: near 99.934 degrees longitude and −29.128 degrees latitude, perhaps with a radius of 30 kilometers which is about a quarter of a a degree
3. In January 2023: near 94 degrees longitude and −36 degrees latitude
4. In 2025: Seemingly favoring the mid-2022 estimates again over the 2023 one

In addition, the WSPR data seem to contradict the assumption that the plane flew in a straight path.

How should we think about the posterior distribution of where the plane is likely to have crashed in light of the fact that some of these new estimates are not even in the outer square of the previous distribution based on the debris that was found?

1.6 Ocean Infinity

A private company called Ocean Infinity has searched for the plane in 2018 using robot controlled submarines, starting at 92.8 degrees longitude and −35.6 degrees latitude where the satellite images suggested it would be and moving generally northeast. In total, it searched 112,000 square kilometers.

In 2025, Ocean Infinity is going to try again, but the weather will not be conducive to searching until summer in the southern hemisphere. It will reportedly be paid $70 million by the government of Malaysia if and only if it finds the bulk of the plane by early 2026. Ocean Infinity has to pay its own search costs, and we do not know exactly what those are, in part because the costs would be less if it finds the plane quickly. It has been suggested that crew, supplies, etc. for a six-week search might cost around $16 million. Of course, building robot-controlled submarines is not cheap, but that could be considered a capital investment because the submarines can be used for other purposes in the future.

Do you think it was a good decision for Ocean Infinity to agree to this contract structure with the Malaysian government?

2 September Federal Reserve Meeting

On September 17, 2025, the Federal Reserve Governors will meet to decide whether to change the benchmark interest rate, which is currently between 4.25% and 4.50%. It is possible to buy futures market contracts that pay $1 if your guess for the benchmark interest rate range is correct and pay nothing if your guess is wrong. Thus, the price of a futures market contract can be interpreted as a market-based conditional probability of that interest rate range prevailing. The data on the prices of future contracts was downloaded from the CME FedWatch website and can be loaded with

```
CME_FedWatch <- readr::read_csv("CME_FedWatch.csv", show_col_types = FALSE)
```

The four columns are

1. **Closing_date**: Day of trading
2. **Big_cut**: Probability that the interest rate range will be reduced by 0.5%
3. **Small_cut**: Probability that the interest rate range will be reduced by 0.25%
4. **No_cut**: Probability that the interest rate range will remain between 4.25% and 4.50%.

The last few observations look like

```
tail(CME_FedWatch)
```

```
# A tibble: 6 x 4
  Closing_date Big_cut Small_cut No_cut
  <chr>          <dbl>     <dbl>  <dbl>
1 8/7/2025      0          0.919 0.0814
2 8/8/2025      0          0.889 0.111
3 8/11/2025     0          0.859 0.141
4 8/12/2025     0          0.939 0.0614
5 8/13/2025     0.0572     0.943 0
6 8/14/2025     0          0.921 0.0786
```

In earlier months, the probabilities do not sum to 1 each day because there are many other possible interest rate ranges besides these three that are not shown here.

2.1 Stan Program

Write a Stan program that models the log-odds of a small cut in the interest rate range at the September 17, 2025 meeting (over the span of the above data) as an autoregressive process of order 1, which is to say that

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t$$

where y_t is the log-odds of a small cut on day t, α is an intercept, β is a slope, and ϵ_t is an error term that has a logistic distribution with an expectation of 0 and a scale of λ_t. Stan has functions pertaining to the `logistic` distribution already built in.

Note that the unknown λ_t is not the standard deviation of the logistic distribution, which is actually $\lambda_t \frac{\pi}{\sqrt{3}}$. You should also model λ_t (or perhaps $\gamma_t \equiv \ln \lambda_t$) with some structure that reflects the fact that the uncertainty in y_t is lessened as we get closer to the September 17, 2025 meeting.

2.2 Priors

Why did you choose the prior distributions you used for each unknown parameter in your Stan program?

2.3 Posterior Distribution

Use `cmdstanr` or `cmstanpy` to draw from the posterior distribution of the parameters in your Stan program, making sure to eliminate any warnings that might arise. How would you describe the dependence in the posterior distribution between your beliefs about α and your beliefs about β?

2.4 Influential Observations

Which, if any, dates are inconsistent with the assumption of the ELPD estimator that they could be omitted without having a major effect on the posterior distribution?

2.5 Posterior Predictive Checks

Make a plot (or plots) to substantiate the claim that the model fits the past data reasonably well without substantially overfitting. What about the plot (or plots) leads you to make that conclusion?

2.6 Prediction

There are 22 trading days until the September 17, 2025 (because futures markets are closed on weekends and Labor Day). Draw 4000 times from the posterior predictive distribution of the event that the Federal Reserve makes a small cut to the interest rate range on September 17 that is implied by your model. You

can either do this in the `generated quantities` block of your Stan program or you can do it in Python or R afterward.

How would you describe this predictive distribution and why does it differ from the observe y_t on August 15?