Out-of-OH Q&A

Gustavo Pereira

November 4, 2019

Contents

1 Why

The ideal time to ask questions about the material and problem sets is during office hours. However, many of you can't make it to office hours for valid reasons, so I also try to respond to questions sent by e-mail or other means.

Keeping this document helps you (and yours truly) for a few reasons:

- More clarifications/explanations available
- There is no 'unfair advantage' to those who ask many questions
- I don't have to answer the same thing over and over :)

2 First half - pset 1

2.1 About limits of sets, and Durrett and Billingsley being wrong

"Hi Gustavo,

Thank you for the class today. Can I further clarify Q4 you briefly discussed in class? I initially directly used $\lim(-\infty, x) \to \Omega$ in the steps, but you pointed out it was wrong. I'm actually still a bit confused about why it is wrong. I referred to **Durrett** (who directly used $\lim(-\infty, x) \to \Omega$) and **Billingsley** (who states "clearly"). So, unfortunately, they both would get the proof wrong. Would it be possible for you to give me some hints on which theorems would be useful in the proof? Thank you so much!

[screenshot of Durrett's book, theorem 1.1]

Best,

xx"

(I kept the bold face from the e-mail)

Dear xx,

I wouldn't dare to say that Durrett would get it wrong if he was answering the problem set! But definitely, if he just copied and pasted from his own book, he would be discounted.

Here's the reason: earlier in the book, he states the continuity from above and below of probability measures in terms of collections of sets *indexed by natural numbers*. In order to do that, he defines what it means to say

$$A_i \uparrow A$$

for that class of collections.

Later, in the context of proving the limits of CDFs, he applies a similar statement to the collection $\{(-\infty, x]\}_{x \in \mathbf{R}}$. The problem is that it's not indexed by natural numbers. So if he wanted full credit, he'd need to clarify he meant by

$$(-\infty, x] \uparrow \mathbf{R}$$

and why continuity of $\mathbf{P}$ – defined only for limits of "increasing" sets indexed by natural numbers – also applies for these limits.

The hint is the same one I gave in the recitation. Use the fact (no need to prove it) that

$$\lim_{x\to\infty} F(x) = 1$$

if, and only if, for every *increasing* diverging sequence $x_n \uparrow \infty$,

$$F(x_n) \to 1$$

and try to apply countable additivity.

Sincerely,

Gustavo

3 First half - pset 4

3.1 Clarifying the meaning of posterior in Q3

> In Q3 of PS4, we are asked to compute the posterior mean for β and σ, and I assumed that usually meant the expectation of the conditional distribution of β, conditioned on data, and similarly, the conditional distribution of σ on the data.
>
> In the hint, I'm questioning my understanding because β was conditioned on σ as well. Going forward, does that mean posterior distributions condition on data and all other parameters except the parameter in question? More specifically, do you know of any resources where I could read up on the mechanics behind this?

Computing the distribution of beta given Y and sigma is only supposed to be an intermediate step to make calculations easiser.

The end goal is to find the joint distribution of beta and sigma given data.

3.2 Q4 - all β vs some β

> I was working on question 4 and I realized that I have a bit of a gap in my understanding about admissible decision rules. I know that a Ridge estimator is a Bayesian estimator, and thus is admissible. I know that an admissible decision rule is not dominated, so since the OLS estimator is another decision rule, it is not the case that the risk of the OLS estimator is less than or equal to the risk of the Ridge estimator for all beta and strictly

> less than for some beta. So then, as I understand it, this implies that there is **some** beta for which the OLS estimator has strictly higher risk than the Ridge estimator (but I know nothing about how they compare for any other beta). But, I don't think this tells me anything about how the risks compare for all the other betas, right?
>
> Since in this problem we'd have that the risk for each estimator is the mean squared error, and thus bias + tr(variance), I think I'd want to have an inequality the compares the MSE of the Ridge estimator to the MSE of the OLS estimator, but from the fact that the Ridge estimator is admissible I only see how to write that inequality for **some** beta – how can I extrapolate to being able to write an inequality for **all** beta, in order to be able to make a statement about how tr(var(beta_ols)) compares to tr(var(beta_ridge)) compare?

You're in the right track, but let me point something out first. Admissibility guarantees that

$$\text{MSE}(\hat{\beta}^{\text{Ridge}}, \beta) \leq \text{MSE}(\hat{\beta}^{\text{OLS}}, \beta)$$

holds for at least one β. Rigorously speaking, the inequality is not necessarily strict, because they could have the same risk all over the parameter space and that wouldn't violate admissibility.

Now with that caveat, I would suggest that instead of trying to get a comparison that holds for every possible β, try to use the one β you know exists for which the ranking of MSEs above holds. Write down both sides of the inequality for that particular β, and see if you can compare either side with $V_\beta(\hat{\beta}^{\text{Ridge}})$ – note that this variance does not depend on β.

4 Second half - pset 1

4.1 Approaching Q8

> Could you give us a hint about Q8? We're trying to solve it for a while and we don't even know how start.

It might be useful to write

$$(T_n - \theta) = A_n B_n$$

where $A_n = \frac{1}{\sqrt{n}}$ and $B_n = \sqrt{n}(T_n - \theta)$.

Any deterministic sequence $A_n \to A$ also satisfies $A_n \xrightarrow{d} A$. Can you say something about $A_n B_n$?