Probability and Statistics, From Classical to Bayesian

Week 7

11th June - 18th June 2021

1 Introduction

This week we will be learning about confidence/credible intervals and introduce you guys to another major applications of statistics, i.e. hypothesis testing. We will be seeing both of these in the Bayesian and classical setup. Theoretical portion will come to an end for this project and next week we will talk about sampling and computational aspects of statistics to complete the project.

2 Confidence/Credible intervals

Confidence Intervals(Classical Setup): So far, in estimating of an unknown parameter, we give a single number as our guess for the known parameter. It would be better to give an interval, and say with what confidence we expect the true parameter to lie within it.

Credible Intervals: The idea is the same, we want to predict a range instead of a single number but there is a catch. In the Bayesian setup we think of the parameter to be a random variable so what we are interested in is finding an interval in which the probability of finding the parameter is high.

2.1 Confidence Interval

The general idea is to create a random variable with know distribution, which is dependent on our parameter in some way. Lets deal with normal distributions only for the time being.Let $X_1, .., X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables.
(1)Estimating μ when σ^2 is known. Consider the random variable $\frac{\sqrt{n}(\bar{X_n}-\mu)}{\sigma}$. The variable depends **only** on μ and we know the distribution is the simple N(0,1). This is what we required, now:

$$P(-z_{\alpha/2} < \frac{\sqrt{n}(\bar{X_n}-\mu)}{\sigma} < z_{\alpha/2}) = 1 - \alpha$$

Where $0 < \alpha < 1$ and z_α is the number such that $P(Z > z_\alpha) = \alpha$ (in other words, z_α is the

$(1 - \alpha)$-quantile of the standard normal distribution). On simple rearrangement we can make μ the subject

$$P(\mu \in [\bar{X_n} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X_n} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]) = 1 - \alpha$$

Lets say we wanted an interval with confidence 95 percent, take $\alpha = 0.05$ and we are done. The last equation gives us one interval that fulfils this condition.
The task of finding such a random variable is not easy, students theorem is used to calculate the other cases for Normal Distribution and for other kinds of distribution, we deal with them by approximating there sums as normal random variable using CLT. You will see that in the hand written pdf but I think you got the problems over here.

2.2 Credible Interval

Thankfully things are much more simpler for credible intervals.
Step 1: Find the posterior distribution of the parameter we are interested in.
Step 2: Integrate the posterior over an interval such that the result is more than the required confidence/credibility.
That's it, calculating posteriors is straightforward, no need to rely on some theorem. The integration can also be avoided by computational techniques(In fact even calculation of posteriors can be avoided to a large extent).

Note: The key difference is in the first case parameter is treated as a constant while in the next we treat it as a random variable. In Bayesianism, the probability distributions reflect our degree of belief. So when we computed the credible region above, it's equivalent to saying.
"Given our observed data, there is a 0.95 probability that the true value of μ falls within CR_μ"
In frequentism, on the other hand, μ is considered a fixed value and the data (and all quantities derived from the data, including the bounds of the confidence interval) are random variables. So the frequentist confidence interval is equivalent to saying
"There is a 0.95 probability that when I compute CI_μ from data of this sort, the true mean will fall within CI_μ."

3 Student t- theorem

Recall two basic facts about the normal distribution:

- If X has $N(\mu, \sigma^2)$, then $aX + b$ has $N(a\mu + b, a^2\sigma^2)$.
- If X has $N(\mu_1, \sigma_1^2)$ and Y has $N(\mu_2, \sigma_2^2)$ and they are independent, then $(X+Y)$ has $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

You can prove these results using the MGF technique.

Result : Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$. Then, we have the following:

- $\bar{X_n}$ has $N(\mu, \sigma^2/n)$, where $\bar{X_n} = \frac{1}{n}\Sigma_{k=1}^n X_k$
- $\frac{nW_n}{\sigma^2}$ has χ_n^2, where $W_n = \frac{1}{n}\Sigma_{k=1}^n (X_k - \mu)^2$
- $\frac{(n-1)s_n^2}{\sigma^2}$ has χ_{n-1}^2, where$^2_n \frac{1}{n-1}\Sigma_{k=1}^n (X_k - \bar{X_n})^2$.

- $\bar{X_n}$ and s_n^2 are independent.
- $\frac{\bar{X_n}-\mu}{s_n/\sqrt{n}} = \frac{\sqrt{n}(\bar{X_n}-\mu)}{s_n}$ has t_{n-1} distribution.

proof : *First result* is easy to see.
For *Second Result* $Z_i = (X_i-\mu)/\sigma$ are i.i.d. $N(0,1)$ distributions and Z_i^2 are χ_1^2 distribution(Try to proof that independent Gamma distribution with same scale parameter, when added, their shape parameter are added as such).
Now, The goal is to show that $\bar{Z_n}$ and $\Sigma_{k=1}^n(Z_k\bar{Z_n})^2$ are independent, and the latter has χ_{n-1}^2 distribution. For Latter *three parts*, interested people can ask on Discord Channel for the proof. It is a bit rigorous and we aren't including it here.

4 Hypothesis Testing

Need of Hypothesis Testing: Lets say a psychic claims to have divine visions unavailable to most of us. You are assigned the task of testing her claims. You take a standard deck of cards, shuffle it well and keep it face down on the table. The psychic writes down the list of cards in some order - whatever her vision tells her about how the deck is ordered. Then, you count the number of correct guesses. If the number is 1 or 2, perhaps you can dismiss her claims. If it is 45, perhaps you ought to be take her seriously. Again, where to draw the line? This is a prototype of what are called testing problems.
One hypotheses (we call it the null hypothesis and denote it by H_0) is that the psychic is guessing randomly. The alternate hypothesis (denoted H_1) is that her guesses are better than random guessing (in itself this does not imply existence of psychic powers, it could be that she has managed to see some of the cards, etc.).
Can we decide between the two hypotheses based on the data X? What we need is a rule for deciding which hypothesis is true. A rule for deciding between the hypotheses is called a test. For example, the following are examples of rules (the only condition is that the rule must depend only on the data at hand).
(1) If X is an even number declare that H_0 is false. Else, declare that H_0 is true.
(2) If $X > 5$, then reject H_0, else accept H_0.
(3) If $X > 8$, then reject H_0, else accept H_0.
The first rule does not make much sense as the parity (evenness or oddness) has little to do with either hypothesis. On the other hand, the other two rules make some sense. They rely on the fact that if H_0 is false, then we expect X to be larger than if H_0 is true. But, the question still remains, should we draw the line at 5, 8, or somewhere else?

Technical details: In testing problems there is only one objective, to avoid the following two possible types of mistakes:
Type-I error: H_0 is true, but our rule concludes H_1:
Type-II error: H_1 is true, but our rule concludes H_0:
The probability of Type-I error is called the significance level of the test and usually denoted by σ.
So, $\sigma = P_{H_0}$(the test rejects H_0).
Similarly, one defines the power of the test as, $\beta = P_{H_1}$(the test rejects H_0)
Rejection Region (R): The test will accept the alternate hypothesis is $X \in R$. Here X is the data we have and R is a subset of the possible values of data we can get. We are interested in finding R.
UMP α-test: The test with significance α having the maximum power.

Note that β is the probability of not making Type-II error, and hence we would like it to be close to 1. Given two tests with the same level of significance, the one with higher power is better. Ideally, we would like both errors to be small, but that is not always achievable. We fix the desired level of significance (usually $\sigma = 0.05$ or 0.1) and only consider tests whose probability of Type-I error is at most σ. It may seem surprising that we take σ to be so small. Indeed the two hypotheses are not treated equally. Usually H_0 is the default option (representing traditional belief) and H_1 is a claim that must prove itself. As such, the burden of proof is on H_1.
To use analogy with law, when a person is convicted, there are two hypotheses, one that he is guilty and the other that he is not guilty. According to the maxim "innocent till proved guilty", one is not required to prove his innocence. On the other hand, guilt must be proved. Thus, the null hypothesis is "not guilty" and the alternative hypothesis is "guilty".

Method of determining: The p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. A smaller p-value means that there is stronger evidence in favor of the alternative hypothesis. Standard procedure is to reject null hypothesis if p-value $< \alpha$. It is not always easy to find the p-value, we mostly use the same idea as Confidence intervals. For some specific testing problems, the NP lemma is pretty useful.

Neyman-Pearson lemma: Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where the pdf (or, pmf) corresponding to θ_i is f_{θ_i} for i = 0, 1, using a test with rejection region R that satisfies
i. $x \in R$ if $f_{\theta_1}(x) > kf_{\theta_0}(x)$ and $x \in R^c$ otherwise, for some $k \geq 0$, and
ii. $\alpha = P_{\theta_0}(X \in R)$. Then, any test that satisfies these two conditions is a UMP level α test.

4.1 Bayesian Testing of hypothesis

So, here again say H_0 is some hypothesis(mostly termed as Null Hypothesis) and we have to proof it against some counter hypothesis ,sayH_1. Then for the given data we look up a number(Bayes' factor) and report it mostly. Bayes' factor is $\frac{p}{1-p}$, It can be translated as

$$\frac{\text{Likelihood of given data given H1 is true}}{\text{Likelihood of given data given H0 is true}} = \frac{P(data|H_1)}{P(data|H_0)}.$$

You can play with the term on the right to get different forms of this equation. (Like we switched $P(data|SomeParameter)$ with other similar terms). Normally we decide a certain level and we accept the alternate hypothesis if the Bayes' factor is more than that level.