

CASELLA AND BERGER NOTES AND PROBLEMS - CHAPTER 8

AUSTIN DAVID BROWN

1. PROBLEMS

Chapter 8: 1, 2, 5, 6, 9, 13, 16, 18, 19, 20, 22, 25, 26, 28, 31, 34, 35(a,b), 39, 41

(1) Use CLT

(2) Calculate the density or the distribution function at 15.

(5) Ya got stuck

(6)

(9)

(13)

(16) This follows from the definitions. $\alpha = 1$ for (a), $\alpha = 0$ for (b).

(18) algebra problem

(19) Use NP Lemma

(20) Not sure about this one Cannot figure out how to precisely pick the correct values.

(22)

2. NOTES

8 / 15

THIS IS THE EXAMPLE I SHOULD USE: (See Wasserman)

Comparing algorithms is a great example for statistics. Coin flip

Hypothesis Testing in Stats is a proof by contradiction.

This entire theory is a way around proof by contradiction for some reason.

statistical power is alpha.

Mathematical way to think about this:

Statistical power is just $P_{\theta}(X \in R)$ Suppose $\theta \in \Theta_0$.

Then the "power" is $\alpha = P_{\theta}(X \in R)$

Suppose $\theta \in \Theta_A$.

Then the "power" is $P_\theta(X \in R)$

This is how to write and think of a Confidence Interval:

$$P(\{X : \theta \in C(X)\})$$

LRT is basically comparing the H_0 model to the H_0 model and you choose some constant that you think it should satisfy.

I mean the way Power is defined, maximizing something with respect to the input is going to win....just by definition. I don't really like the way this theory is setup honestly.

DO NOT USE THE THEORY FOR HYP TESTS INSTEAD USE PROOF BY CONTRADICTION and Confidence Intervals!!!!

I am really going to try and avoid this theory and just use proof by contradiction for all of this.

I think point estimation and interval estimation with proof by contradiction and a chosen alpha is the way to go.

The issue with using proof by contradiction is that you have to make choice for alpha. But the way that the power function is defined, it makes maximizing alpha good, which is making the choice that a high alpha is good. But really, the rejection region is another choice.

8 / 16

TODO Hyp test problem set

TODO understand why MLE are asymptotically normal

8 / 16

Likelihood ratio is a form of model comparison.

Neyman-Pearson Lemma: When given α , for a simple hypothesis, the model comparison is the most powerful test.

So these likelihood ratio tests are useful for using the asymptotic chi square distribution.

The important stuff:

Neyman Pearson Lemma

Wald Test for using CLT for example $P(|z| \leq t) = .95$ is the assumption to create the rejection region

Generalized Likelihood Ratio Test for using its asymptotic χ^2 distribution. For example, $P(X^2 \leq t) = .95$ is the assumption to create the rejection region. Note: this relies on CLT in the proof anyway though.

GLRT has the fastest rate of convergence compared to Wald and Score tests.

OK GOT IT. — Wald Test is easy to use and is an application of CLT

Generalized LR Test is harder to use but converges faster than Wald. It is also an application of the CLT.

See Keener's notes on the topic. —

Covariance is an inner product in L_2 . Thus variance is an inner product and we can derive all of the properties like this. For example, if X, Y are independent, then $\text{Cov}(X, Y) = 0$ and they are orthogonal!

Independence is orthogonality!!!! Oh my gosh I just get it now! It makes total sense by the product measure!

Covariance is inner product!! Independence is like orthogonality!!!!

8 / 17

LRT is actually a norm comparison with the sup norm. Note that the density is always positive.

$$\|L\|_{S(\Theta_0)} / \|L\|_{S(\Theta)}$$

Independence really means orthogonal. See product measures.

The real reason that MLE are useful is because they converge in probability to the parameter constant.

The proof that MLE have limiting normal distribution really relies on the fact that they converge in probability.

Moral of the story: We like MLE's because they converge in probability to the constant parameter.

The real insight: We want estimators that converge in probability to the constant.

The real insight to MLE is that we are dealing with the sup norm and there must be uniform convergence somewhere.

This must happen somewhere:

$$\|L - \theta\|_{\infty} \rightarrow 0$$

This must make them converge in probability?

The LLN generalizes to uniform convergence!!

$$\|\bar{X} - \mu\|_{\infty} \xrightarrow{P} 0$$

THIS IS THE INSIGHT TO MLE!!! WE ARE USING UNIFORM CONVERGENCE IN PROBABILITY!!!

The MLE is just a way to use the sup norm actually and then applying LLN to show they are consistent.

REFERENCES

- [1] Casella and Berger. *Statistical Inference*.