Statistical Decision Theory and Bayesian Analysis

Chapter 1: Losses, Risks and Decison Principles

Lili MOU

moull12@sei.pku.edu.cn
http://sei.pku.edu.cn/~moull12

30 April 2015

Reference

§1, James O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer, 1985.

This book covers basic materials of statistical decision theory in an easy-to-understand yet critical manner. The prerequisite is rather low.

- Statistical level: moderately serious statistics
- Mathematical level: easy advanced calculus

This slide mainly picks textual materials in Chapter 1. For detailed math, please refer to other resources.

Introduction

Decision theory, as the name implies, is concerned with the problem of making decisions. Statistical decision theory is concerned with the making of decision in the presence of statistical knowledge which sheds light on some of the uncertainties involved in the decision problem.

Beyond Classical Statistics

Classical statistics is directed towards the use of sample information (the data arising from the statistical investigation) in making inference about θ*.*

Non-sample information

- Loss
- Prior

Loss

Statisticians seem to be pessimistic creatures who think in terms of losses. Decision theorists in economics and business talk instead in terms of gains (utility).

Prior

- A lady, who adds milk to her tea, claims to be able to tell whether the tea or the milk was poured into the cup first. In all of ten trials conducted to test this, she correctly determines which was poured first.
- A music expert claims to be able to distinguish a page of Haydn score from a page of Mozart score. In ten trials conducted to test this, he makes a correct determination each time.
- A drunken friend says he can predict the out come of a flip o f a fair coin. In then trials conducted to test this, he is correct each time.

Frequentist's hypothesis test?

Probability of θ

After all, in most situations there is nothing “random” about θ. A typical example is there θ is an unknown but fixed physical constant (say the speed of light) which is to be concerned. The basic idea is that probability statements concerning θ are then to be interpreted as “personal probabilities” reflecting the degree of personal belief in the likelihood of the given statement.

Abusive Frequentist's Tools

What is statistical inference?

> *In statistical inference the goal is not to make an immediate decision, but is instead to provide as "summary" of the statistical evidence which a wide variety of future "users" of this evidence can easily incorporate into their own decision-making process. . .*
> *Because of this point, many statisticians use "statistical inference" as a shield to ward off consideration of losses and prior information.*

Point Null Hypothesis Test

For a large enough sample size, the classical test will be virtually certain to reject. Likewise a difference that is not significant statistically can nevertheless be very important practically.

In particular, the fact that the **stopping rule** *affects the computation of the p-value means that frequentists often do not terminate experiments early, even when it is obvious what the conclusions are, lest it adversely affect their statistical analysis.*[1]

[1] *Machine Learning: A Probabilistic Perspective*:

Point Null Hypothesis Test

Optional stopping in these various forms is potentially quite common. In a recent survey, 58% of researchers admitted to having collected more data after looking to see whether the results were significant and 22% admitted to stopping an experiment early because they had found the result that they were looking for.[2]

[2] Adam Sanborn, et al., The Frequentists Implications of Optional Stopping on Bayesian Hypothesis Tests.

Point Null Hypothesis Test (2)

Frequentists are hypocrisy!

> *It seems somewhat nonsensical, however, to deliberately formulate a problem wrong, and then in an adhoc fashion explain the final results in more reasonable terms.*

Who cares about nonoccurrence?

Suppose a substance to be analyzed can be sent either to a laboratory in New York or a laboratory in California. The two labs seem equally good, so a fair coin is flipped to choose between them, which "heads" denoting that the lab in New York will be chosen. The coin is flipped and comes up tails, so the California lab is used. After a while, the experimental results come back and a conclusion and report must be developed. Should this conclusion take into account the fact that the coin could have been heads, and hence that the experiment in New York might have been performed instead?

Frequentists v.s. Conditionalists

Common sense (and the conditional viewpoint) cries no, that only the experiment actually performed is relevant, but frequentist reasoning would call for averaging over all possible data, even the possible New York data.

The Likelihood Principle

Definition. For observed data, x, the function $\ell(\theta) = f(x|\theta)$, considered as a function of θ, is called the *likelihood function*.

The Likelihood Principle. In making inferences or decisions about θ after x is observed, all relevant experimental information is contained in the likelihood function for the observed x. Furthermore, two likelihood functions contain the same information about θ if they are proportional to each other (as functions of θ).

Reemphasizing the Conditionalist Perspective

Jeffrey (1961):[3]

> *. . . a hypothesis which may be true may be rejected because it has not predicted observable results which have not occurred.*

"Thus, . . . , the null hypothesis that $\theta = \frac{1}{2}$ certainly would not predict that X would be larger than 9, and indeed such values do not occur. Yet the probabilities of these unpredicted and not occurring observations are included in the classical evidence against the hypothesis."

[3] Jeffreys, Hl, 1961. *Theory of Probability (3rd edn.).* Oxford University Press

Twists and Turns

Pratt (1962):[4]

- An engineer draws a random sample of electron tubes and measures the plate voltages under certain conditions with a very accurate voltmeter, accurate enough so that measurement error is negligible compared with the variability of the tubes.
- A statistician examines the measurements, which look normally distributed and vary from 75 to 99 volts with a mean of 87 and a standard deviation of 4. He makes the ordinary normal analysis, giving a confidence interval for the true mean.

[4] Pratt, J. W., 1962. Discussion of A. Biernbaum's "On the foundations of statistical inference." *J. Amer. Statist. Soc.* (Ser. B) **27**, 169–203. See also Berger, J., Statistical Decision Theory and Bayesian Analysis.

Twists and Turns (2)

- Later he visits the engineer's laboratory, and notices that the voltmeter used reads only as far as 100, so the population appears to be "censored." This necessitates a new analysis, if the statistician is orthodox.
- However, the engineer says he has another meter, equally accurate and reading to 1000 volts, which he would have used if any voltage had been over 100.
- This is a relief to the orthodox statistician, because it means the population was effectively uncensored after all.

Twists and Turns (3)

- But the next day the engineer telephones and says, “I just discovered my high-range voltmeter was not working the day I did the experiment you analyzed for me.”
- The statistician ascertains that the engineer would not have held up the experiment until the meter was fixed, and informs him that a new analysis will be required.
- The engineer is astounded. He says, “But the experiment turned out just the same as if the high-range meter had been working. I obtained the precise voltages of my sample anyway, so I learned exactly what I would have learned if the high-range meter had been available. Next you'll be asking about my oscilloscope.”

The Weak Conditionality Principle.

Suppose one can perform either of two experiments E_1 or E_2, both pertaining to θ, and that the actual experiment conducted is the mixed experiment of first choosing $J = 1$ or 2 with probability $\frac{1}{2}$ each (independent of θ), and then performing experiment E_J. Then the actual information about θ obtained from the overall mixed experiment should depend only on the experiment E_j that is actually performed.

Choosing a Paradigm or Decision Principle

. . . statistics is a collection of useful methodologies, and that one should "keep an open mind as to which method to use in a given application." This is indeed the most common attitude among statisticians.

While we endorse this attitude in a certain practical sense, we do not endorse it fundamentally.

We have argued that this desired fundamental analysis must be compatible with the Likelihood Principle. Furthermore, . . . it is conditional Bayesian analysis that is the only fundamentally correct conditional analysis.

. . . we would strongly argue that conditional (Bayesian) reasoning should be the primary weapon in a statistician's arsenal.