



Foundations of Statistical Inference
Statistical Decision Theory and Bayesian Analysis Report

Author: Massimiliano Pronesti

Instructor: **Motonobu Kanagawa**

November 26, 2021

1 Introduction to Decision Theory

We refer to **statistical decision theory** as the branch of statistics concerned with the problem of making decisions in the presence of statistical knowledge.

Differently from classical statistics - which is mainly directed towards the use of sample information -, decision statistics takes **also** (sample information is still considered) into account other relevant aspects of the problem which usually prove crucial in the decision process. These can be grouped in two categories:

- **knowledge of possible consequences** of the decision, typically quantified from the loss (or the gain) arising from each possible decision;
- **prior information**, typically arising from past experience of similar situations

2 Key elements of decision statistics

2.1 Loss function

As anticipated in section 1, a key element of the decision theory is the loss function, which defines the result of taking an action a given the unknown quantity θ affecting the decision process, commonly referred as **the state of nature**.

For technical convenience, the loss function satisfies the following chain of inequalities

$$L(\theta, a) \geq -K > -\infty$$

assuming, of course, that $L(\theta, a)$ is defined for all $(\theta, a) \in \Theta \times \mathcal{A}$, being Θ the parameter space and \mathcal{A} the set of all possible actions. Nevertheless, a critical **caveat** is that, in many problems, loss function and prior information might not be well defined or even nonunique (at the time of decision making).

2.2 Bayesian expected loss

In light of the above caveat, the natural way of proceeding consists in considering the **expected loss** of making a decision and then picking the "optimal" one according to it.

The **Bayesian expected loss** refers to the uncertainty in θ , treating it as a random quantity associated to a probability distribution $\pi^*(\theta)$ ¹

$$\rho(\pi^*, a) = E^{\pi^*} L(\theta, a) = \int_{\Theta} L(\theta, a) dF^{\pi^*}(\theta)$$

2.3 Frequentist risk and its implications

The so-called "classical statistics" employs a rather different approach towards loss, i.e. the **risk function** given a decision rule $\delta(X)$, trying to minimize that:

$$R(\theta, \delta) = E_{\theta}^X [L(\theta, \delta(x))] = \int_{\mathcal{A}} L(\theta, \delta(x)) dF^X(x|\theta)$$

It is important to highlight that in a no data problem, $R(\theta, \delta) \equiv L(\theta, a)$. Moreover, differently from Bayesian expected loss - which is a number -, the risk is a function on Θ which drives us to the following problem: **being θ unknown, it's not clear what "small" risk means.**

We need to define a (partial) ordering for decision rules to introduce a choice criteria: given two decision rules δ_1, δ_2 , we say δ_1 is R-better than δ_2 if

$$R(\theta, \delta_1) \leq R(\theta, \delta_2) \quad \forall \theta \in \Theta$$

with strict inequality holding for some θ . If no R-better decision rule exists, a decision rule δ is called *admissible* (the definition of *inadmissible* is dual).

Nevertheless, there is usually a large class of admissible decision rules for a large problem, which, typically, have risk functions being better than the others only locally.

¹We use π^* rather than π as the latter one typically refers to the initial prior distribution, while the former one is usually the final posterior distribution of θ (after seeing the data)

3 Decision principles

This far, we introduced how to perform a statistical analysis. We now describe the major methodologies to actually make a decision.

3.1 Conditional Bayes Decision Principle

Let $a \in \mathcal{A}$ be an action minimizing the expected loss $\rho(p_i^*, a)$. Such an action will be called a **Bayes action** and will be denoted a^{π^*} .

3.2 Frequentist Decision Principles

Another possibility to select a decision rule is employing the risk function, although this proves more effortful because of the multiple decision rules admissible. In decision theory, we can identify three main principles that can be used to develop statistical procedures:

- **Bayes risk principle:** as remarked in **section 2.3**, Bayes risk is a number, hence we define a decision rule minimizing it. Notice that, in a no-data problem, the risk is just the loss which implies that this principle gives the same answer of the conditional Bayes decision one.
- **minimax principle:** this principle is often called for consideration of randomized decision rules. According to it, given two randomized decision rules $\delta_1^*, \delta_2^* \in \mathcal{D}^*$, the first is preferred to the second if

$$\sup_{\theta \in \Theta} R(\theta, \delta_1^*) < \sup_{\theta \in \Theta} R(\theta, \delta_2^*)$$

- **invariance principle:** if two problems have same formal structure, then same decision must be taken

4 Foundations

In this last section, we're gonna expand even further the conditional versus frequentist controversy - crucial in statistics - as well as discuss some incorrect usages of classical inference in decision problems.

4.1 Misused inference procedures

4.2 Frequentist Perspective

4.3 Conditional Perspective

4.4 Likelihood Principle