

Lecture 1: Introduction; elements of a decision problem; standard loss functions

Statistical Inference–II ■ Unit I: Decision Theory

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

M.Sc. Statistics, Semester III
Academic Session 2026–2027

Unit I: Decision Theory

Lecture 1 begins Unit I: Decision Theory.

Today

- Introduction
- The Elements of a Decision Problem
 - The State Space (Parameter Space)
 - The Action Space (Decision Space)
 - The Loss Function
 - The Set of Observable Random Variables

Statistical inference is concerned with drawing conclusions about an unknown *state of nature* on the basis of data which are subject to random variation.

In the classical development of the subject, the three principal inference problems — point estimation, interval estimation and testing of hypotheses — are treated as separate topics, each with its own optimality criteria:

- point estimation: unbiasedness and minimum variance;
- interval estimation: coverage probability and expected length;
- testing: size and power.

The three classical inference problems

- **Point estimation:** the task of selecting a numerical value as an estimate of the unknown parameter.
- **Hypothesis testing:** we determine whether the observed evidence is sufficient to reject the null hypothesis; failure to reject it does *not* establish that it is true.
- **Interval estimation:** to select an interval intended to include the true value of the parameter.

In each case, the decision is taken in the light of the observed values of the associated random variables $X_1, X_2, \dots, X_n$.

The decision-theoretic approach

The decision-theoretic approach, developed principally by Wald (1950), provides a single unifying framework within which all these problems — and many others — appear as special cases of one general problem:

choosing an action from a set of available actions when the consequences of the choice depend on an unknown state of nature.

- The essential new ingredient is the explicit quantification of the consequences of a wrong choice through a *loss function*.
- Once the loss function is specified, every inference procedure can be compared with every other by means of its *risk function*, and the search for good procedures becomes a well-posed optimisation problem.

- (i) The decision-theoretic framework is *not* intrinsically Bayesian. It is a framework for comparing procedures. Within it, one may adopt
- a Bayesian attitude (average the risk over a prior),
 - a minimax attitude (guard against the worst case), or
 - a restricted-class attitude (impose unbiasedness — see “Estimation Viewed as a Decision Problem” — and minimise risk within the restricted class).

All three are studied in this unit; restriction by invariance (equivariance) is beyond the scope of this course.

- (ii) The framework does not automatically deliver a unique “best” procedure. As we shall see (Uniformly Best Rules), a uniformly best rule almost never exists. The entire subject may be regarded as a collection of reasonable responses to this inconvenient fact.

The treatment follows, in the main, Ferguson (1967), Berger (1985) and Lehmann and Casella (1998), with the level of detail adjusted to the syllabus.

The four ingredients of a decision problem

A statistical decision problem is completely specified by four ingredients. We describe them in turn:

1. the state space (parameter space) Θ ;
2. the action space (decision space) $\mathcal{A}$;
3. the loss function L ;
4. the set of observable random variables $X_1, X_2, \dots, X_n$.

The state space (parameter space)

The first element of the decision problem is a set of alternative values or forms that will be assumed by the particular characteristics of the part of the real world in which we are interested.

- These characteristics may be assumed to be numerical and are denoted by the parameter θ .
- The set of all possible values of the parameter θ is denoted by the parameter space Θ .
- In the decision-theoretic approach, the parameter space is referred to as the *space of possible states of nature*.

Thus Θ is the set of all values that the unknown “true state of the world” may take. It may be finite (as in $\Theta = \{\theta_1, \theta_2\}$ for a simple-versus-simple testing problem), countable, an interval of the real line, or a subset of $\mathbb{R}^k$.

The action space (decision space)

The second element of decision theory is the set of alternative decisions or actions that may be taken in relation to that part of the real world.

- A particular action is denoted by a (or d).
- The whole set of alternative actions is denoted by $\mathcal{A}$ (or D).
- The set $\mathcal{A}$ is called the *action space* or *decision space* relevant to the problem under consideration.

Example 1.1 (Action spaces in the three classical problems)

- **Point estimation of θ :** $\mathcal{A} = \Theta$ (or the range of the parametric function to be estimated). An action is a proposed numerical value for θ .
- **Testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$:** $\mathcal{A} = \{a_0, a_1\}$, where a_0 denotes “do not reject H_0 ” and a_1 denotes “reject H_0 ”.
- **Interval estimation:** $\mathcal{A}$ is the collection of all intervals $[\theta, \bar{\theta}]$ contained in Θ .

The third element of decision theory is the loss function. When we take an action, say a , and θ is the true state of nature, we may or may not make an error; in either case there is a loss to be incurred. If we take the action a when the true state of nature is θ , this loss is denoted by $L(\theta, a)$.

Definition 1.2 (Loss function)

A *loss function* is a real-valued function

$$L : \Theta \times \mathcal{A} \longrightarrow \mathbb{R},$$

where $L(\theta, a)$ denotes the loss incurred by taking the action a when θ is the true state of nature. It is usually assumed that L is bounded below; most commonly one assumes $L(\theta, a) \geq 0$ for all (θ, a) .

Remark 1.3

The assumption $L \geq 0$ entails no real loss of generality when L is bounded below, since adding a constant to the loss adds the same constant to every risk and therefore does not alter the comparison of any two rules.

- A loss function which is bounded below by $-M$ may be replaced by $L(\theta, a) + M$ throughout.
- Some authors work instead with a *gain* or *utility* function

$$u(\theta, a) = -L(\theta, a)$$

and maximise expected utility; see Savage (1954) and DeGroot (1970).

Let $\gamma(\theta)$ be the parametric function to be estimated and let a denote the estimate.

- (a) **Squared error loss (SELF):** $L(\theta, a) = (a - \gamma(\theta))^2$. Analytically the most tractable; penalises large errors severely.
- (b) **Absolute error loss (ALF):** $L(\theta, a) = |a - \gamma(\theta)|$. More robust to occasional large errors.
- (c) **Weighted squared error loss:** $L(\theta, a) = w(\theta) (a - \gamma(\theta))^2$, with $w(\theta) > 0$. The choice $w(\theta) = 1/\theta$ or $w(\theta) = \{\theta(1 - \theta)\}^{-1}$ arises naturally in Poisson and binomial problems respectively.

(d) **LINEX (linear–exponential) loss:**

$$L(\theta, a) = e^{c\Delta} - c\Delta - 1, \quad \Delta = a - \gamma(\theta), \quad c \neq 0.$$

Asymmetric: over-estimation and under-estimation are penalised differently. Widely used in reliability and life-testing contexts.

(e) **ϵ -neighbourhood zero–one loss:**

$$L(\theta, a) = \begin{cases} 0, & |a - \gamma(\theta)| \leq \epsilon, \\ 1, & \text{otherwise.} \end{cases}$$

For the two-action problem with $\mathcal{A} = \{a_0, a_1\}$ and a partition $\Theta = \Theta_0 \cup \Theta_1$, use the unambiguous error-cost notation

$$L(\theta, a_0) = \begin{cases} 0, & \theta \in \Theta_0, \\ c_{10} (> 0), & \theta \in \Theta_1, \end{cases} \quad L(\theta, a_1) = \begin{cases} c_{01} (> 0), & \theta \in \Theta_0, \\ 0, & \theta \in \Theta_1. \end{cases}$$

- c_{01} is the loss due to a Type I error;
- c_{10} is the loss due to a Type II error.

We refer to this as the *testing loss* (c_{01}, c_{10}) . The special case $c_{01} = c_{10} = 1$ is the 0–1 *loss*.

Note 1.4 (Notation in the prescribed books)

Goon, Gupta and Dasgupta (1980) and several other texts denote the decision space by D , a decision by d and the class of decision rules by Δ ; in these notes they are $\mathcal{A}$ (or D), a (or d) and $\mathcal{D}$.

For testing, those texts write a for the loss of a Type II error and b for the loss of a Type I error. Here these are c_{10} and c_{01} respectively, because the letter a is already used for actions. Thus a result stated there with $a/(a+b)$ reads

$$c_{10}/(c_{01} + c_{10})$$

here (compare Exercise 1.18).

The set of observable random variables

The fourth element is the set of observable random variables $X_1, X_2, \dots, X_n$, whose joint distribution is governed by θ .

- We write $\mathbf{X} = (X_1, \dots, X_n)$, with joint density (or mass function) $f(\mathbf{x} \mid \theta)$ and distribution function F_θ .
- The sample space, that is, the set of possible values of $\mathbf{X}$, will be denoted by $\mathcal{X}$; typically $\mathcal{X} \subseteq \mathbb{R}^n$.

Before reaching any decision we observe $\mathbf{x} = (x_1, \dots, x_n)$, and the action taken is allowed to depend on $\mathbf{x}$.

A decision problem in which no observations are available is called a *no-data problem*; the action is then chosen directly from the loss function and any available prior information.

Note 1.5

A statistical decision problem is thus the quadruple

$$(\Theta, \mathcal{A}, L, \{F_\theta : \theta \in \Theta\}).$$

Wald's fundamental insight was that such a problem is formally a *two-person zero-sum game* between

- the statistician (who chooses a decision rule) and
- Nature (who chooses θ),

with the risk function as the pay-off. See Blackwell and Girshick (1954) for the game-theoretic development.

Practice (Lecture Notes, Chapter 1)

Re-work the examples of this lecture without looking at the solutions.

Next lecture

Lecture 2: Non-randomised decision rules; the risk function; risk of an estimator and of a test.