Introduction to Statistical Decision Theory

Lili Mou
moull12@sei.pku.edu.cn

Outline

Outline

What is a decision?

An example from our research career

- What problem to solve?
- What conference to submit to?
- What track to choose?
- ...

We need to make decisions minutely in our life.

Why statistics?

God does not play dice, but we human do.

The research example

- If I solve problem a, I believe I have 60% chance to achieve a considerable result. The chance dwarfs to 30% for problem b.
- However, provided a state-of-the-art result, problem a has 30% chance to be accepted, and problem b 99%.
- Assume, for convenience, that papers are not to be accepted if the result is not state-of-the-art.

Why statistics?

God does not play dice, but we human do.

The research example

- If I solve problem a, I believe I have 60% chance to achieve a considerable result. The chance dwarfs to 30% for problem b.
- However, provided a state-of-the-art result, problem a has 30% chance to be accepted, and problem b 99%.
- Assume, for convenience, that papers are not to be accepted if the result is not state-of-the-art.
- To solve problem a, the overall acceptance probability is

$$.6 \cdot .3 = .18$$

- To solve problem b, the overall acceptance probability is

$$.3 * .99 \approx .3$$

At this point, I know I had better solve problem b, and that statistics need to be involved when making decisions.

Why statistical decision theory?

Decisions are far from simplicity.

- What if we have some samples (e.g., previous papers) submitted to the conference?
- What if we have consulted some experts in the field?
- What if we know the actual consequences of being accepted/rejected?

Why statistical decision theory?

Decisions are far from simplicity.

- What if we have some samples (e.g., previous papers) submitted to the conference?
- What if we have consulted some experts in the field?
- What if we know the actual consequences of being accepted/rejected?

Decision theory helps people make decisions in a more rational manner (instead of modeling people's decisions).

Humans are irrational!

Basic Elements

θ	:	state of nature / parameter that affects the decision
Θ	:	all possible states
a	:	action
$\mathcal{A}$	:	all possible actions
L	:	$\Theta \times \mathcal{A} \to \mathbb{R}$, loss function
x	:	sample
$\mathcal{X}$	:	sample space
δ	:	$\mathcal{X} \to \mathcal{A}$, a decision rule that depends on x
$\pi(\theta)$	:	subjective prior
$\pi^*(\theta)$	:	subjective posterior, given x

Example I: Estimate the Probability of Acceptance

$$
\begin{array}{rcl}
\theta & : & \text{accpetance probability of a particular problem} \\
\Theta & : & [0,1] \\
a = \hat{\theta} & : & \text{an estimator} \\
\mathcal{A} & : & [0,1] \\
L & : & \|\theta - a\|^2 \\
 & & \text{Loss function is subjective} \\
x & : & \text{None (no-data problem)} \\
\text{or} & & \\
x & : & \text{the probability that an expert suggests with strength } n \\
\mathcal{X} & : & [0,1] \\
\delta(x) = \hat{\theta}(x) & : & \max\{0, x - 0.1\}
\end{array}
$$

$$
\begin{array}{rcl}
\pi(x) & = & \text{Dirichlet}(\alpha_1, \alpha_2) \\
\pi^*(x) & = & \text{Dirichlet}(\alpha_1 + nx, \alpha_2 + n(1 - x))
\end{array}
$$

Example II: Making a Decision

$$
\begin{aligned}
\theta &: \theta_a, \theta_b \text{ accpetance probabilities of problems a and b} \\
\Theta &: [0,1] \times [0,1] \\
a &: \text{either problem a, or problem b} \\
\mathcal{A} &: \{\text{Problem a, Problem b}\} \\
L &: \cdots \\
x &: \text{None (no-data problem)} \\
\text{or} & \\
x &: \text{the probability that an expert suggests with strength } n \\
\mathcal{X} &: [0,1] \\
\delta(x) &= ??? \\
\pi(x) &= \text{Dirichlet}(\alpha_1, \alpha_2) \\
\pi^*(x) &= \text{Dirichlet}(\alpha_1 + nx, \alpha_2 + n(1-x))
\end{aligned}
$$

Philosolphical Remarks

Is θ a random variable? Or, are priors sensible?

- 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 of a fair coin. In ten trials conducted to test this, he is correct each time.

Philosohpical Remarks (II)

Can we quantize losses in all aspects (e.g., time, money) by a single real number?

Yes, both theoretically and operationally, but not practically.

"No heaven or hell": Nothing is infinitely good or bad

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

Philosophical Remarks (III)

Subjectivity of science

“When different reasonable priors yield substantially different answers, can it be right to state that there **is** a single answer?”

“It is indeed rather peculiar that some decision-theorists are ardent anti-Bayesians, precisely because of the subjectivity of the prior, and yet have no qualms about the subjectivity of the loss.”

Outline

All come from the loss

In the beginning HUMAN created the loss.
HUMAN said, let it depend on θ, and it depended on θ.
HUMAN said, let it depend on a, and it depended on a.

$$L(\theta, a)$$

Loss is bad, we want to minimize it. But

- θ is unknown
- x is a random variable

Bayesian Expected Loss

If $\pi^*(\theta)$ is the believed probability distribution of θ, at the time of decision making, the **Bayesian expected loss** of an action a is

$$\rho(\pi^*, a) = \mathbb{E}^{\pi^*}[L(\theta, a)]$$

To minimize, with respect ot a, $\rho(\pi^*, a)$, given data x.

Note that x adjusts your belief about θ, indicated by $\pi^*(\theta)$

Frequentist Risk

A decision rule δ is a function from $\mathcal{X}$ into $\mathcal{A}$. If the data sample is x, and the action is $\delta(x)$

The **frequentist's risk** function of a decision fule $\delta(x)$ is

$$R(\theta, \delta) = \mathbb{E}_\theta^X[L(\theta, \delta(X))] = \int_{\mathcal{X}} L(\theta, \delta(x))\, \mathrm{d}F^X(x|\theta)$$

Choose a δ such that minimizes $R(\theta, \delta)$
But, θ is unknown...

Bayes Risk

The **Bayes risk** of a decision rule δ, with respect to a prior distribution π on Θ is

$$r(\pi, \delta) = \mathbb{E}^{\pi}[R(\theta, \delta)] = \mathbb{E}^{\pi}\,\mathbb{E}_{\theta}^{X}[L(\theta, \delta(X))]$$

Choose a δ such that minimizes $r(\theta, \delta)$

The Big Picture

Which decision rule to choose?

Bayesian

- No-data $\Rightarrow$ Bayes risk
- Data known $\Rightarrow$ Bayes expected loss

Frequentist: The problem is ill-posed

- Admissibility
- Minimax
- Invariance
- Resort to prior $\Rightarrow$ Bayes risk

Example

Loss matrix

	Problem a	Problem c
θ_0 (considerable results in problem a)	0	50
θ_1 (no satisfactory results in problem a)	100	50

No-data problem

Example

Loss matrix

	Problem a	Problem c
θ_0 (considerable results in problem a)	0	50
θ_1 (no satisfactory results in problem a)	100	50

No-data problem

- Frequentist's risk + Admissibility
  $\Rightarrow$ Either *a* or *c*, because each action is admissible
- Frequentist's risk + Minimax
  In the worst scenario, choosing *c* yields 50 loss, whereas *a* yields 100
  $\Rightarrow c$
- Bayesian expected loss, with prior $\pi(\theta_0) = 0.6, \; \pi(\theta_1) = 0.4$

$$r(\pi, a) = .6 \cdot 0 + .4 \cdot 100 = 40$$

$$r(\pi, c) = .6 \cdot 50 + .4 \cdot 50 = 50$$

$\Rightarrow a$

Outline

Take-Home Messages

- Bayesians think everything unknown is a random variable.
- Bayesians condition on data x (known)
- Frequentists condition on θ (unknown)
- Subjectivity versus objectivity
  Good (1973): "The subjectivist states his judgements, whereas the objectivists sweeps them under the carpet by calling assumptions knowledge, and he basks in the glorious objectivity of science."

More philosophical discussions to be continued in May 27 afternoon's seminar.