Chapter 1. Statistics in the path towards decisions

Section 1.1 States of nature, actions, consequences

$a \in A$: action

$\theta \in \Theta$: state of nature

$l(\theta, a)$: consequence of taking action a when state of nature is θ

Example: a patient comes with symptoms that possibly indicate a disease

	Treat	
True disease	a1=yes	a2=no
θ1= present	2	5
θ2=absent	1	0

Issues to consider:

– Does nature try to maximize any loss ?

– Does nature try to update the true value based on our action ?

<u>Data</u> (here, from a diagnostic test).

A test is given, with possible outcomes $x \in \mathcal{X}$: $x = 1$ (test positive), or $x = 2$ (test negative).

The properties of the test are described by the probability distribution of the data across different possible patients, $f(x \mid \theta)$.

True disease	Data: 1(pos)	Data: 2(neg)
θ1= present	.94	.06
θ2=absent	.02	.98

Q1. What information do data provide on patient disease state ?

Q2. What should we believe about the disease state ?

Q3. What should we do (treat or not) ?

Section 1.2 Strategies (decision functions).

We want to have our actions be informed on data.

Definition: A (pure) strategy is a function s from the data space $\mathcal{X}$ to the action space A. Note that $s(X)$ is a random function.

data	s1	s2	s3	s4
X=1	a1	a1	a2	a2
X=2	a1	a2	a1	a2

We will denote by S the set of pure strategies.

Consequences of strategies

In what ways do we evaluate/compare consequences of strategies

One way is by expected loss: $L(\theta, s) = E\{l(\theta, s(X)) \mid \theta)\}$

For example: $L(\theta_1, s_2) = E\{l(\theta_1, s_2(X)) \mid \theta_1\}$=(.94)(2)+(.06)5=2.18
Similarly: $L(\theta_2, s_2) = (.02)1 + (.98)0 = .02$

Notes:
– It matters to be clear as to whom the loss function is specific.
– We could have defined the strategy loss in other ways, e.g., the median
– We will see later that using expectation is actually a *result* of some axioms that try to explain our behaviours.

In the screening example, the losses of the strategies *as functions* of nature are:

disease	s1	s2	s3	s4
θ1	2	2.18	4.82	5
θ2	1	0.02	0.98	0

Notes:

– There is still no direct ordering for some of the strategies, e.g. $s1, s2$.

– Some strategies can be uniformly worse than others, e.g., $s3$ to $s2$.

Admissible strategies

Def. A strategy $s(X)$ is inadmissible if there exists another strategy $s'(X)$ that dominates it for all θ, that is:

$$L(\theta, s') \leq L(\theta, s), \text{for all } \theta,$$

and with at least one inequality strict. Admissible strategies are those that are not inadmissible.

In the screening example, $s3$ is inadmissible (to $s2$).

Randomized strategies

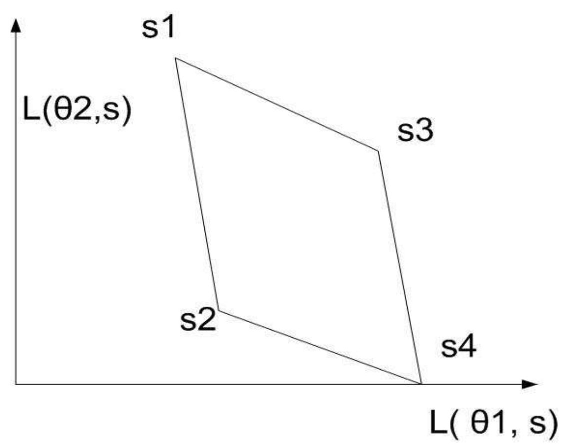
Def. A randomized strategy is one that selects a pure strategy according to a probabiliy distribution on the set of pure strategies.

Example: $s^* = \frac{1}{3}s1 + \frac{2}{3}s4$ means "do $s1$ with probability $1/3$, and do $s4$ with probability $2/3$". (Alternative definition, can place data first, then randomize the strategies).

Extend space S of strategies, to include all randomized strategies.

Geometric representation

For two states of nature, we can represent a strategy by its expected losses

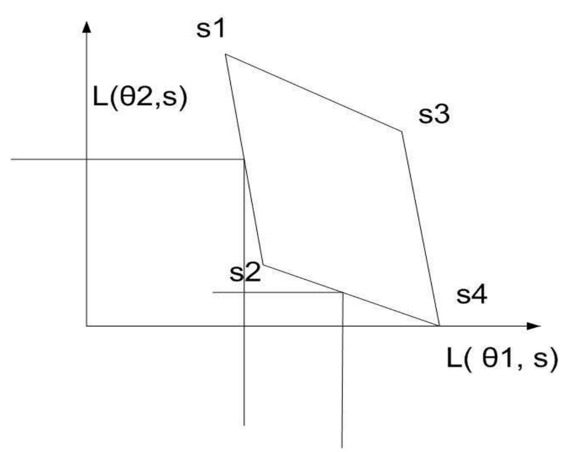


1. By having included randomized strategies, we know that if $s1, s2 \in S$ then $\lambda s1 + (1-\lambda)s2 \in S$, for all $\lambda \in [0,1]$. So, S is the "Convex Hull" of the pure strategies .

2. How are admissible strategies represented geometrically?

<u>Def.</u> The closed "lower quantant" with corner point (x_0, y_0) in the 2-dim space is the set $\{(x, y) : x \le x_0, y \le y_0\}$. (The open lower quantant excludes the boundary).

<u>Result 1.1.</u> A strategy s is admissible iff the closed lower quantant at s shares no other common point with S.



Section 1.3 Ordering the (admissible) strategies; minimax and Bayes rules

Note: among the admissible strategies, there is no uniformly best one

So, ordering strategies requires additional criteria

Three such additional criteria involve the concepts of :
minimax, Bayes, and bias. Here, examine minimax, Bayes; on bias, starting in subsequent sections

Minimax strategies

Def. A minimax strategy s^* is one that minimizes (over strategies) the maximum loss over states of nature:

$$\sup_\theta L(\theta, s^*) = \inf_s \sup_\theta L(\theta, s)$$

Notes:

– a minimax strategy does not necessarily exist; it is not necessarilly unique

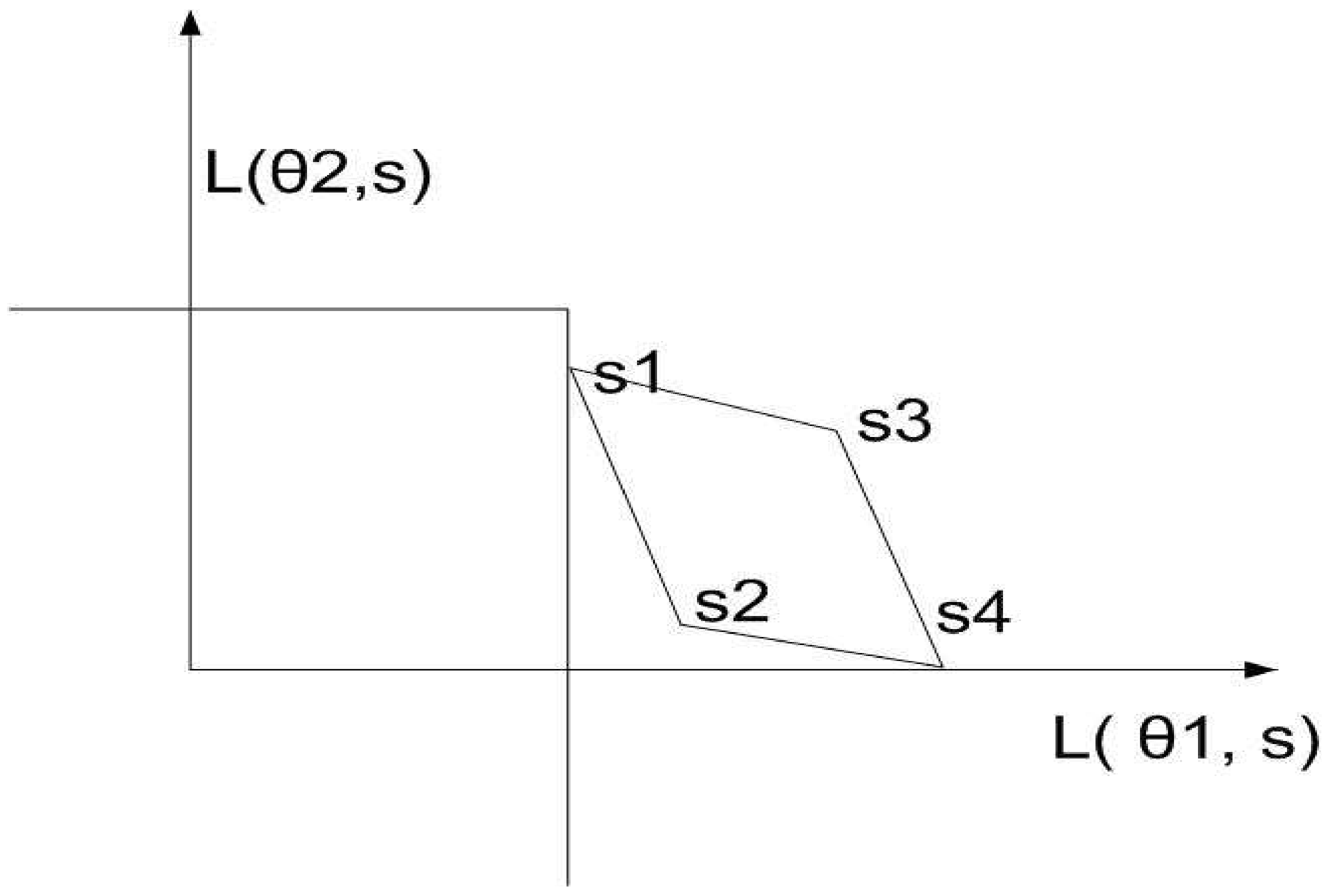
– "conservative" in the sense that it considers the worst case for nature

Geometric representation of minimax

Def. A diagonal lower quantant is a lower quantant with its corner on the $y = x$ line.

If there is a smallest "diagonal closed lower quantant", among those that contact the convex hull S of the losses, then those points of contact are minimax strategies (why ?)

In our screening example, "treat no matter the screening test" is the unique minimax strategy


L(θ2,s)
s1
s3
s2
s4
L(θ1, s)

Sometimes, we can find minimax strategies by finding "equalizers".

Def. An equalizer strategy s is one whose loss is the same for all states of nature

Result 1.2. If an admissible strategy is equalizer, and there is a mimimax, then that equalizer is also minimax (why ?)

Bayes strategies

Def. A Bayes strategy s_π is one that minimizes (over strategies) the expected loss over a prior distribution π on the states of nature:

$$\pi L(\theta_1, s_\pi) + (1-\pi)L(\theta_2, s_\pi) = \inf_s\{\pi L(\theta_1, s) + (1-\pi)L(\theta_2, s)\}$$

Notes:

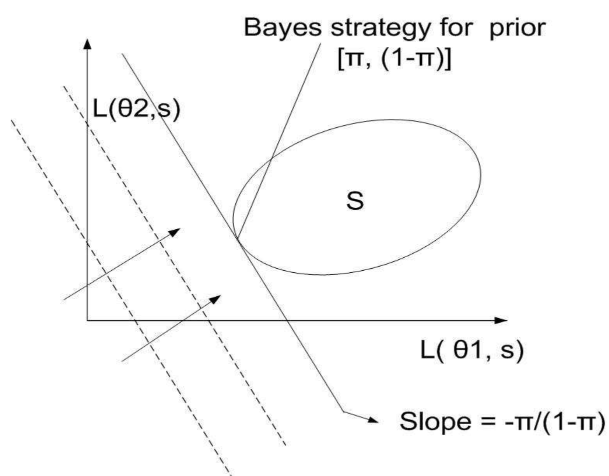
1. More generall, s Bayes for π minimizes $E_\pi\{L(\theta, s(X))\}$
2. Need prior distribution
3. Also other ways to rank strategies over the prior, e.g., with median.

Geometric representation of Bayes strategies

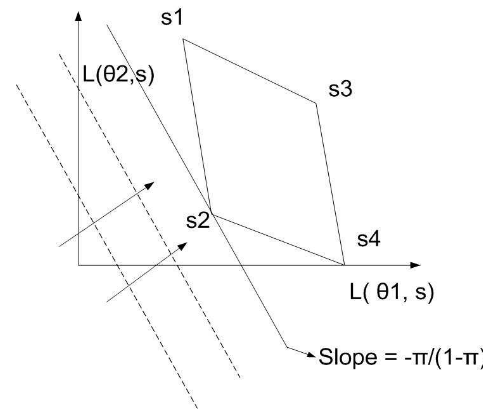
For a prior $(\pi, 1-\pi)$ on the states of nature $(\theta_1, \theta_2,)$, we wish to

minimize $\pi L_1 + (1-\pi)L_2$

subject to $(L_1, L_2) \in S$



In our screening example



1. if $\pi/(1-\pi) > 5.3$ then the Bayes rule is s_1
2. if $\pi/(1-\pi) = 5.3$ then the Bayes rule is any combination of s_1, s_2
3. if $.007 < \pi/(1-\pi) < 5.3$ then the Bayes rule is s_2
4. if $.007 = \pi/(1-\pi)$ then the Bayes rule is any combination of s_2, s_4
5. if $\pi/(1-\pi) < .007$ then the Bayes rule is s_4

Finding Bayes strategies

Suppose we observe data X from the likelihood $f(x \mid \theta)$. Assuming interchangeability of expectations over θ and X,

$$E_\theta[E_{X|\theta}\{l(\theta, s(X)) \mid \theta\}] = E_X[E_{\theta|X}\{l(\theta, s(X)) \mid X\}]$$

So, to minimize the RHS, for observation x, choose $s(x)$ that minimizes the posterior loss

$$E_{\theta|X}\{l(\theta, s(x)) \mid X = x\} = \int_\theta l(\theta, s(x))\text{pr}(\theta \mid X = x)d\theta$$

1. Solve the "no-data problem": assume a "working prior" $\pi(\theta)$ and find s^* that minimizes $\int_\theta l(\theta, s)\pi(\theta)d\theta$.

   Note: that s^* is a function of the working prior, so we write $s^* = s^*_{(\pi)}$

2. Given a prior $\pi(\theta)$, find the posterior distribution

$$\mathsf{pr}(\theta \mid X) = \frac{f(X \mid \theta)\pi(\theta)}{\int f(X \mid \theta)\pi(\theta)d\theta}$$

3. Then, the Bayes strategy for the prior $\pi(\theta)$ and data x from the likelihood $f(x \mid \theta)$ is the Bayes strategy of the "no-data problem", except that we replace the working prior with the posterior distribution. That is, $s^*(X) := s^*_{\mathsf{pr}(\theta|X)}$

Example.

We want to estimate $\theta =$ logit (prevalence) of xerophthelmia in chilren in a developing country. We conduct a study that will give us data X from a likelihood pr$(X \mid \theta)$.

Assuming
– loss function $c_0(s-\theta)^2$, for estimator s
– prior distribution $\pi(\theta)$ for θ
what is the Bayes strategy (estimator) ?

Solving the no-data problem: minimize $\int_\theta (\theta - s)^2 \pi(\theta) d\theta$

$$
\begin{aligned}
0 &= \frac{d}{ds}\int_\theta (\theta - s)^2 \pi(\theta) d\theta \\
&= \frac{d}{ds}\int_\theta (\theta^2 - s\theta + s^2)\pi(\theta) d\theta \\
&= 2s - 2E(\theta) \rightarrow s = E(\theta)
\end{aligned}
$$

So, the Bayes rule is $s^*(X) = E(\theta \mid X)$.

Section 1.4. The relation between Bayesian and admissible estimation

Complete Class Theorem (modified)

(a) A Bayes rule with respect to a non-degenerate prior is admissible.

(b) An admissible rule is Bayes for some prior.

Proof that Bayes rules to non-degenerate prior are admissible.

Suppose s_0 is Bayes for $(\pi, 1-\pi)$. If it is inadmissible, say, to s_1, then

$$L(\theta_i, s_0) \geq L(\theta_i, s_1), \text{ for all } i, \text{ and one strict, so}$$
$$\pi L(\theta_1, s_0) + (1-\pi) L(\theta_2, s_0) > \pi L(\theta_1, s_1) + (1-\pi) L(\theta_2, s_1).$$

But, since s_0 is Bayes, the LHS of the last expression must be at most the RHS (contradiction). So s_0 must be admissible.

Section 1.5 Prospects and their combinations using the likelihood (fixed state of nature)

- why did we assume a *number* describes our loss when acting upon a state of nature ?
- why did we *average* those losses by the likelihood ?
- why did we *average* losses by the prior ? (see 1.6)

We assume here θ fixed, e.g., patients who have disease.

Definitions

1. A pure prospect is a sequence of events that can result after taking a particular action
2. P is the set of all prospects.

   Analogy:

Suppose there are two aspects in life, and they give the following nine prospects

		life length		
		short	med	long
quality	poor	$\tilde{P}_1$	$\tilde{P}_2$	$\tilde{P}_3$
of life	fair	$\tilde{P}_4$	$\tilde{P}_5$	$\tilde{P}_6$
	good	$\tilde{P}_7$	$\tilde{P}_8$	$\tilde{P}_9$

Given those, suppose the prospects in the screening example are

	Treat	
	a1=yes	a2=no
True Disease		
θ=present	$\tilde{P}_2$	$\tilde{P}_4$
θ=absent	$\tilde{P}_6$	$\tilde{P}_9$

3. We need a representation for the prospect that will follow from a *strategy*, i.e., depending on data X. So, we define a mixed prospect p to be a distribution on the pure prospects.

   Analogy:

   Since we are focus on diseased patients, if we use strategy $s_2(X)$, we will have prospect $\tilde{P}_2$ with probability $pr(X = 1 \mid \theta = 1)$, and prospect $\tilde{P}_4$ with probability $pr(X = 0 \mid \theta = 1)$.

4. We denote by P^* the set of all mixed prospects.

Note, then, that a mixed prospect, is a representation of an estimator (or a strategy) through its consequences.

(class notes for axioms for utility)

Section 1.6 Prospects with unknown state of nature

We now recognize that nature (θ) can be many states (values).

Definitions

1. A vector prospect (across states of nature) is the list of the prospects of an action (i.e., list across those states of nature)
2. G is the set of all vector prospects

Analogy (back to screening example):

Suppose there are two aspects in life, and they give the following nine prospects

		life length		
		short	med	long
quality	poor	$\tilde{P}_1$	$\tilde{P}_2$	$\tilde{P}_3$
of life	fair	$\tilde{P}_4$	$\tilde{P}_5$	$\tilde{P}_6$
	good	$\tilde{P}_7$	$\tilde{P}_8$	$\tilde{P}_9$

Given those, suppose the prospects in the screening example are

	Treat	
	a1=yes	a2=no
True Disease		
θ=present	$\tilde{P}_2$	$\tilde{P}_4$
θ=absent	$\tilde{P}_6$	$\tilde{P}_9$

Analogy:

Recall that if we use the strategy $s_2(X)$

(i) for diseased patients, then we will have prospect $\tilde{P}_2$ with probability $pr(X = 1 \mid \theta = 1)$, and prospect $\tilde{P}_4$ with probability $pr(X = 0 \mid \theta = 1)$; call this $P_{\theta=1}(s_2)$

(ii) for healthy patients, then we will have prospect $\tilde{P}_6$ with probability $pr(X = 1 \mid \theta = 2)$, and prospect $\tilde{P}_9$ with probability $pr(X = 0 \mid \theta = 2)$; call this $P_{\theta=2}(s_2)$

3. Thus, without knowing θ, we can represent the prospect that will follow from *strategy* s_2, by

$$[P_{\theta=1}(s_2), P_{\theta=2}(s_2)]\,, \text{ which we call } \underline{P}$$

4. We denote by G* the set of all distributions on prospects $\underline{P}$ (mixed prospects).

(class notes for utility on prospects with unknown state of nature)