Two contestants, Statistician and Nature, put up 1 or 2 fingers simultaneously. The loss to the statistician is $L(\theta, a) = (\theta - a)^2$.

	Nature $\theta = 1$	Nature $\theta = 2$
Statistician $a = 1$	0	1
Statistician $a = 2$	1	0

If it is known $P(\theta = 1) = 1/3, P(\theta = 2) = 2/3$. What is the best action for the statistician? Answer: $a = 2$.
Suppose instead putting up his fingers, the statistician is allowed to put a real number on a piece of paper. What number should he present?
Answer: $a = 5/3$. For squared error loss,
a=$E(\theta) = 1 * 1/3 + 2 * 2/3 = 5/3$.

Add data: X= number of fingers Nature tells Statistician he will put up.

$\theta = 1$: $P(X = 1) = 3/4, P(X = 2) = 1/4$.

$\theta = 2$: $P(X = 2) = 3/4, P(X = 1) = 1/4$.

Assume again the statistician can choose a real number and the loss function is given by $L(\theta, a) = \theta(\theta - a)^2$. What is his best action if $X = 1$ is observed? If $X = 2$ is observed?

$P(\theta = 1|X = 1) = \frac{3}{5}, P(\theta = 2|X = 1) = \frac{2}{5}$. (by Bayes' theorem)

For the given loss function,

$f(a) = E(L) = E[\theta(\theta^2 - 2a\theta + a^2)] = E(\theta)a^2 - 2E(\theta^2)a + E(\theta^3)$,

let $f'(a) = 0$, we can get

$a = \frac{E(\theta^2)}{E(\theta)} = \frac{1^2*3/5+2^2*2/5}{1*3/5+2*2/5} = \frac{11}{7}$.

$P(\theta = 1|X = 2) = \frac{1}{7}, P(\theta = 2|X = 2) = \frac{6}{7}$.

Similarly, we can find $a = \frac{25}{13}$ if $X = 2$ is observed based on $p(\theta|X = 2)$.

Utility function

Can we make decision based on risk (expected loss) only?
Making decisions based on expected utility has a sound mathematical foundation.

Bet or not

1. You receive \$2 if a fair coin fall heads and pay \$1 if it falls tails.
2. You entire fortune is \$1 million. You receive 2 million extra if the coin falls heads and you lose your fortune otherwise.
3. You intend to spend all your cash on beer to drink this evening. You have \$3 which will buy you a good deal of beer. You receive \$ 3 if the coin falls heads and lose your \$3 otherwise.
4. You are desperate to see the "big game". You have \$3 but a ticket costs \$5. You receive \$3 extra if the coin falls heads and lose your \$3 otherwise.

We will use utility to measure the *value* of money and many other things.

Utility function

A function u on a set of prospects to a set of numbers.
Two properties of utility function:
1. $u(P_1) > u(P_2)$ iff the individual prefers P_1 to P_2.
2. If P is a prospect where the individual faces P_1 with prob. λ and faces P_2 with prob. $1 - \lambda$, then
$u(P) = \lambda u(P_1) + (1 - \lambda) u(P_2)$.

Assumptions behind utility

1. Given any two prospects P_1, P_2, the individual can always decide whether he prefers P_1 to P_2, or likes them equally, or he prefers P_2 to P_1.
2. If P_1 is at least as good as P_2, and P_2 is at least as good as P_3, then P_1 is at least as good as P_3.
3. If P_1 is preferred to P_2 which is preferred to P_3, then there is a mixture of P_1 and P_3 which is preferred to P_2, and there is a mixture of P_1 and P_3 over which P_2 is preferred.
4. Suppose P_1 is preferred to P_2 and P_3 is another prospect, then the individual will prefer a mixture of P_1 and P_3 to the same mixture of P_2 and P_3.

Evaluate P if $u(P_0), u(P_1)$ known

P_1 is preferred to P which is preferred to P_0, after sufficient self introspection, he decides he is indifferent between
P and a certain mixture of P_1 and P_0.
$u(P) = \lambda u(P_1) + (1 - \lambda) u(P_0)$
Example: P : ordinary life
P_0: hit by a car and dies
P_1 : cross the street and meets the lady.

Risk set

Two states of nature:
θ_1 : today is a sunny day
θ_2 : today is a rainy day.
Three actions.
a_1 : wear fair-weather clothes
a_2 : wear raincoat
a_3: wear raincoat, boots and rain hat
Loss or utility $l(\theta, a)$

	θ_1 : *sunny*	θ_2 : *rainy*
a_1	0	5
a_2	1	3
a_3	3	2

x_1 : fair weather indicated
x_2 : dubious
x_3 : bad weather indicated.
Probability distribution of x.

	x_1	x_2	x_3
θ_1	0.60	0.25	0.15
θ_2	0.20	0.30	0.50

In total, we have 3^3 possible strategies.

Strategies

	s1	s2	s3	s4	s5	s6	s7	s8	s9
x1	a1	a1	a1	a1	a1	a1	a1	a1	a1
x2	a1	a1	a1	a2	a2	a2	a3	a3	a3
x3	a1	a2	a3	a1	a2	a3	a1	a2	a3
	s10	s11	s12	s13	s14	s15	s16	s17	s18
x1	a2	a2	a2	a2	a2	a2	a2	a2	a2
x2	a1	a1	a1	a2	a2	a2	a3	a3	a3
x3	a1	a2	a3	a1	a2	a3	a1	a2	a3
	s19	s20	s21	s22	s23	s24	s25	s26	s27
x1	a3	a3	a3	a3	a3	a3	a3	a3	a3
x2	a1	a1	a1	a2	a2	a2	a3	a3	a3
x3	a1	a2	a3	a1	a2	a3	a1	a2	a3

Risk table

	s1	s2	s3	s4	s5	s6	s7	s8	s9
θ_1	0.00	0.15	0.45	0.25	0.40	0.70	0.75	0.90	1.20
θ_2	5.00	4.00	3.50	4.40	3.40	2.90	4.10	3.10	2.60

	s10	s11	s12	s13	s14	s15	s16	s17	s18
θ_1	0.60	0.75	1.05	0.85	1.00	1.30	1.35	1.50	1.80
θ_2	4.60	3.60	3.10	4.00	3.00	2.50	3.70	2.70	2.20

	s19	s20	s21	s22	s23	s24	s25	s26	s27
θ_1	1.80	1.95	2.25	2.05	2.20	2.50	2.55	2.70	3.00
θ_2	4.40	3.40	2.90	3.80	2.80	2.30	3.50	2.50	2.00

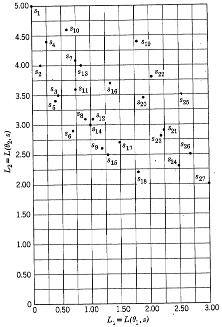
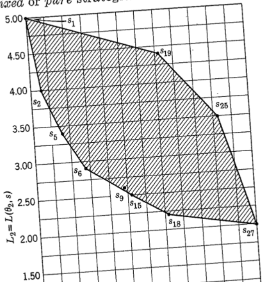


Figure 5.1. Expected loss points $(L_1, L_2) = (L(\theta_1, s), L(\theta_2, s))$ for 27 strategies in Mr. Nelson's rain problem.

each of them is *dominated* by one of these. For example, s_3, which, on reference to Table 5.3 seems to be a reasonable strategy, is is p part of the way from s to s'. Thus every point on the line segment connecting the points representing s and s' represents some strategy. But this means that the set of points representing strategies is *convex*. (A convex set is a set which contains all line segments connecting points of the set.) In fact, the set of points representing strategies is the smallest convex set containing the original *unmixed* or *pure* strategies. See Figure 5.2.



Mixed strategy

Take strategy s_i with probability $p_i, \sum p_i = 1$.
e.g., take s_6 and s_{15} each with prob. $1/2$.
The risk is $\theta_1 : 0.70 * 1/2 + 1.30 * 1/2 = 1$,
$\theta_2 : 2.9 * 1/2 + 2.5 * 1/2 = 2.7$

Pure strategy: it assigns an action to each possible observations.
mixed or randomized strategy: a choice of the set of pure strategies where the choice is made with a random device.
a strategy s **dominates** s^* if $R(\theta_i, s) \leq R(\theta_i, s^*)$ for all i, and $R(\theta_i, s) < R(\theta_i, s^*)$ for at least one i.
A strategy s is **admissible** if it is not dominated by any other strategy (pure or mixed).
The admissible part of the risk set: the lower boundary of the risk set.

Admissibility and completeness

Theorem: The set of admissible risk points is the lower boundary of S if S is closed.
Complete class: a class of decision rules D is a complete class if for any $d \notin D$, there exist $d' \in D$ such that d' is better that d.
(You can do better in D than any place outside).
Essentially complete class: A class of decision rules D is an essentially complete class if for any $d \notin D$, there exist $d' \in D$ that is at leas as good as d.
A class of decision rules D is a **minimal complete class** if it is complete and is a subset of any other complete class.
Roughly: the set of admissible rules is the minimal complete class.

Bayes strategy

A Bayes strategy corresponding to the prior probabilities $(w_i, i = 1, 2, \cdots, k), \sum w_i = 1$ is a strategy which minimizes $R(s) = \sum_i R(\theta_i, s) w_i$.
$R(s)$ is called the **Bayes risk**.

If $P(\theta_1) = 1/3$ and $P(\theta_2) = 2/3$, then s_{18} is the Bayes strategy.

A decision rule may be Bayes vs several prior distributions.
There can be multiple Bayes rules corresponding to the same prior distribution.

Bayesness and admissibility

Every admissible strategy is a Bayes strategy for some prior probability w and $1 - w$. The prior probabilities are determined by the slope of the supporting line of S that passes this point.
supporting line: a supporting line of a set S at the boundary point x if the line passes through x and S is completely on one side of the line.
Any boundary point of a convex set has a supporting line at that point.

Almost all Bayes rules are admissible

If the prior probabilities are positive, the corresponding Bayes strategy is admissible. (Some Bayes strategies are inadmissible in degenerate case.)
proof: Suppose d is not admissible, then there exists d' that is better than d. i.e.,
$R(\theta_i, d') \leq R(\theta_i, d)$ for all i with strict inequality for at least one i.
But this implies
$\sum w_i R(\theta_i, d') < \sum w_i R(\theta_i, d)$ (note we need the Bayes risk is finite there).
This contradicts that d is Bayes.

Bayes rules and pure strategies

At least one Bayes strategy corresponding to prior probabilities (w, 1-w) is a pure strategy $\rightarrow$ we need only to consider pure strategies when we consider Bayes strategies.

Summary

From graphs, it is clear that

1. The set of all randomized strategies is represented by a convex set.
2. All admissible strategies are Bayes strategies for some prior distribution.
3. the Bayes strategies corresponding to positive prior probabilities are admissible.
4. For any prior distribution, there is at least one pure (nonrandomized) Bayes strategy.

exercise

1. In the rain example, find the posterior probability if the prior probability of rain is 0.4 and dubious is observed. What if the prior probability of rain is 0.8? 1.0?

2. Suppose in the contractor problem, the prior probabilities of $\theta_1, \theta_2, \theta_3$ are 0.3, 0.3 and 0.4 respectively.
compute the Bayes strategy and the Bayes risk.

Solutions

1. prior: $P(\theta_1) = 0.6, P(\theta_2) = 0.4$.
x_2=dubious is observed. By Bayes' theorem,
$P(\theta_1|x_2) = \frac{P(x_2|\theta_1)P(\theta_1)}{P(x_2|\theta_1)P(\theta_1)+P(x_2|\theta_2)P(\theta_2)} = \frac{0.25*0.6}{0.25*0.6+0.30*0.4} = 5/9$
and therefore $P(\theta_2|x_2) = 4/9$.
Similarly, if the prior is $P(\theta_1) = 0.2, P(\theta_2) = 0.8$,
$P(\theta_1|x_2) = \frac{0.25*0.20}{0.25*0.20+0.30*0.80} = 5/29$
and therefore $P(\theta_2|x_2) = 24/29$.
If the prior is $P(\theta_1) = 0, P(\theta_2) = 1$, By Bayes' theorem,
we can get $P(\theta_1|x_2) = 0, P(\theta_2|x_2) = 1$.
That is, if the prior distribution puts mass 1 on a certain point, the posterior will also put mass 1 on that point.

Solutions

The prior is given as $P(\theta_1) = P(\theta_2) = 0.3, P(\theta_3) = 0.4$.
First if x_1 is observed, the posterior probabilities are:
$P(\theta_1|x_1) = \frac{P(x_1|\theta_1)P(\theta_1)}{P(x_1|\theta_1)P(\theta_1)+P(x_1|\theta_2)P(\theta_2)+P(x_1|\theta 3)P(\theta_3)} =$
$\frac{1/2*0.3}{1/2*0.3+0*0.3+0*0.4} = 1.$
therefore $P(\theta_2|x_1) = P(\theta_3|x_1) = 0.$
Check the loss table, the expected loss of taking action a_1, a_2, a_3 is 1, 2, 3 respectively, so we take action a_1.
If x_2 is observed, the posterior probabilities are:
$P(\theta_1|x_2) = \frac{1/2*0.3}{1/2*0.3+1/2*0.3+0*0.4} = 0.5,$
and $P(\theta_2|x_2) = \frac{1/2*0.3}{1/2*0.3+1/2*0.3} = 0.5$
and therefore $P(\theta_3|x_2) = 0.$
Check the loss table, the expected loss for taking action a_1, a_2, a_3:
1*0.5+5*0.5=3,
2*0.5+2*0.5=2,
3*0.5+3*0.5=3. So the best action is a_2.

continued

If x_3 is observed, the posterior probabilities are:

$P(\theta_1|x_3) = \frac{0*0.3}{0*0.3+1/2*0.3+1/3*0.4} = 0,$

$P(\theta_2|x_3) = \frac{1/2*0.3}{0*0.3+1/2*0.3+1/3*0.4} = 9/17$

and so $P(\theta_3|x_3) = 8/17$.

Check the loss table, the expected loss for taking the three actions are:

$5 * 9/17 + 7 * 8/17 = 105/17$

$2 * 9/17 + 6 * 8/17 = 66/17$

$3 * 9/17 + 3 * 8/17 = 3 = 51/17$.

the best action is a_3.

If x_4 is observed, the posterior probabilities are

$P(\theta_3|x_4) = \frac{2/3*0.4}{0*0.3+0*0.3+2/3*0.4} = 1$

therefore $P(\theta_1|x_4) = P(\theta_2|x_4) = 0$.

Check the loss table, the best action to take is a_3. which has expected loss 3.

In summary, the Bayes strategy is

$d(x_1) = a_1, d(x_2) = a_2, d(x_3) = a_3, d(x_4) = a_3$.

continued

There are two ways to compute the Bayes risk: average over the distribution of X given θ first, then average over the distribution of θ. Or we can average over the distribution of θ given X first, then average over the distribution of X.

Method 1.

If $\theta = \theta_1$, the action probabilities are

$P(a_1) = 1/2, P(a_2) = 1/2, P(a_3) = P(a_4) = 0$.

check the loss table, the risk is

$1/2 * 1 + 1/2 * 2 = 1.5$.

If $\theta = \theta_2$, the action probabilities are:

$P(a_1) = 0, P(a_2) = 1/2, P(a_3) = 1/2$.

the risk is $1/2 * 2 + 1/2 * 3 = 2.5$.

If $\theta = \theta_3$, the action probabilities are:

$P(a_1) = P(a_2) = 0, P(a_3) = 1$, the risk is 3.

Therefore the Bayes risk is

0.3*1.5+0.3*2.5+0.4*3=2.4.

method 2: compute the marginal distribution of X.
$P(x_1) = P(x_1|\theta_1)P(\theta_1) + P(x_1|\theta_2)P(\theta_2) + P(x_1|\theta_2)P(\theta_3) =$
$1/2 * 0.3 + 0 * 0.3 + 0 * 0.3 = 0.15$,
similarly,
$P(x_2) = 1/2 * 0.3 + 1/2 * 0.3 + 0 * 0.4 = 0.3$,
$P(x_3) = 0 * 0.3 + 1/2 * 0.3 + 1/3 * 0.4 = 0.283$,
and $P(x_4) = 0 * 0.3 + 0 * 0.3 + 2/3 * 0.4 = 0.267$
and the Bayes risk is:
1*0.15+2*0.3+3*0.283+3*0.267=2.4.

More examples about posterior probability

Hemophilia–blood clotting disease
sex linked genetic disease on X chromosome
Males (XY) affected or not.
Females (XX) may have 0 copies of disease gene (not affected) 1 copy (carrier), 2 copies (usually fatal)
Consider a woman, brother is hemophiliac, father is not.
woman's mother must be a carrier.
Woman inherits one X from mother: 50-50 chance of being a carrier.
$\theta = 1$ if woman a carrier, 0 not.
a prior we have $p(\theta = 1) = p(\theta = 0) = 0.5$.
let y_i be the status of the woman's ith son. (1 affected, 0 not).

Example continued

Assume two sons are iid given θ.
$p(y_1 = y_2 = 0|\theta = 1) = 0.5 * 0.5 = 0.25$
$p(y_1 = y_2 = 0|\theta = 0) = 1 * 1 = 1.$
Posterior distribution by Bayes' theorem
$p(\theta = 1|y) = \frac{p(y|\theta=1)p(\theta=1)}{p(y|\theta=1)p(\theta=1)+p(y|\theta=0)p(\theta=0)}$
$= \frac{0.25*0.5}{0.25*0.5+1*0.5} = 0.2$

Updating for new information (suppose a 3rd son is born and unaffected)
note if we observe an affected child then we know $\theta = 1$.
Two approaches to updating analysis
* redo entire analysis (y_1, y_2, y_3 as data).
* update using only new data (y_3) treating previous posterior as new prior.

updating analysis

Redo entire data: now $y = (0, 0, 0)$.
$p(y|\theta = 1) = 0.5 * 0.5 * 0.5 = 0.125$
$p(y|\theta = 0) = 1$.
$p(\theta = 1|y) = \frac{0.125*0.5}{0.125*0.5+1*0.5} = 0.111$.
Updating:
new prior: $p(\theta = 1) = 0.2, p(\theta = 0) = 0.8$.
$p(\theta = 1|y_3) = \frac{0.5*0.2}{0.5*0.2+1*0.8} = 0.111$.
we get the same answer.