My white board writing from week 2

Discussion of the decision theoretic concepts, even in a greater detail than I presented them, is to be found in the copy of the chapter from YS that I posted on moodle. Hence I abstain from reproducing these again here (the minimalist student only needs my discussion from the notes in Lecture 2). However I will discuss in more detail the white board writing related to the

Example 2.5.8:

a) Given that X only can have 3 different values (0,1,2) it is clear that only 8 non-randomized decision rules $(d_1, d_2, \ldots, d_8)$ (as given in the notes) can exists. Then, given that the convex hull is the smallest convex set containing the risk points $\begin{pmatrix} R(\theta_1, d_i) \\ R(\theta_2, d_i) \end{pmatrix}$, $i = 1, 2, \ldots, 8$ we see that it given as shown on the graph. To illustrate the calculation of the (X,Y) coordinates, we note that, for example:

$R(\theta_1, d_1) = L(\theta_1, a_1) * (.81) + L(\theta_1, a_1) * (.18) + L(\theta_1, a_1) * (.01) = 0+0+0=0$

$R(\theta_2, d_1) = L(\theta_2, a_1) * (.25) + L(\theta_2, a_1) * (.5) + L(\theta_2, a_1) * (.25) = 3 * 1 = 3$

hence the risk point that corresponds to d_1 is with coordinates $(0, 3)$.

Similarly, for d_2:

$R(\theta_1, d_2) = L(\theta_1, a_1) * (.81) + L(\theta_1, a_1) * (.18) + L(\theta_1, a_2) * (.01) = 0 + 1 * (.01) = .01$

$R(\theta_2, d_2) = L(\theta_2, a_1) * (.25) + L(\theta_2, a_1) * (.5) + L(\theta_2, \theta_2) * (.25) = 3 * (.25 + .5) =$
$= 3 * (.75) = 2.25$

hence the risk point that corresponds to d_2 is with coordinates $(0.01, 2.25)$.

You can work out the rest in a similar way.

For a prior $(p_1, p_2) = (p_1, 1 - p_1)$ on (θ_1, θ_2), the risk points that have the same value b of their

Bayes risk are on the line with equation

$$p_1 x + p_2 y = b \quad \text{, or equivalently} \quad p_1 x + (1-p_1) y = b.$$

Note that if $x = y$ we get $x = y = b$, hence the x (and equivalently y) coordinate of the intersection of the line $p_1 x + p_2 y = b$ with the line $x = y$ also represents the value of this risk.

b) The minimax rule in the set $\mathcal{D}$ of randomized decision rules is obtained by examining the intersection of the line $y = x$ with the "most south-west" part of the convex risk set. Hence we need to solve the system:

$$\left| \begin{array}{l} y = x \\ y - 0 = \frac{.75 - 0}{.19 - 1}(x - 1) \end{array} \right\}$$

this gives $x = y = \boxed{\frac{25}{52}}$ for the risk point that corresponds to the minimax decision rule δ^* in the set of all randomized decision rules $\mathcal{D}$ that are generated by the set $D = \{d_1, d_2, \dots, d_8\}$ of the non-randomized decision rules; $\boxed{\frac{25}{52}}$ is also the value of the minimax risk.

<u>Important</u>: if we were only looking for the minimax decision rule in the set D (not $\mathcal{D}$(!)) then the answer is different as follows:

Rule	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
$R(\theta_1, d_i)$	0	.01	.18	.19	.81	.82	.99	1
$R(\theta_2, d_i)$	3	2.25	1.5	.75	2.25	1.5	.75	0
max	3	2.25	1.5	.75	2.25	1.5	.99	1

Now the minimal of these maxima is .75 hence d_4 is the minimax decision rule with a minimax risk of .75 (and this risk is $> \frac{25}{52}$ naturally — <u>think</u> about why(!))

If we want to represent the minimax rule δ^* in the set $\mathcal{D}$ as a randomization of the rules d_4 and d_8, we need to find $\alpha \in (0,1)$ to say that
δ^* chooses d_4 with probability α and d_8 with probability $(1-\alpha)$.
But then $\alpha * (.19) + (1-\alpha) * 1 = \frac{25}{52}$ must hold.
Hence we get $\alpha \approx .641$ and can claim that

$$\delta^* = \begin{cases} \text{choose } d_4 \text{ with probability } .641 \\ \text{choose } d_8 \text{ with probability } .359 \end{cases}$$

c) For the least favourable prior, we need to maximize the Bayes risk when we start manipulating the priors $(p, 1-p)$.
Since for any such prior the value of the Bayes risk will be geometrically represented as a X (or equivalently y) coordinate on the line that connects $(0,0)$ with $\left(\frac{25}{52}, \frac{25}{52}\right)$, we obviously maximize when we end up with $\left(\frac{25}{52}, \frac{25}{52}\right)$, i.e. the minimax solution. We are hence looking for a prior in the form $(p, 1-p)$ for which $\left(\frac{25}{52}, \frac{25}{52}\right)$ would be the Bayes solution.

This means $(p, 1-p)$ to be $\perp$ to the line $\overline{d_4 d_8}$. This requirement is the same as to ask that the slope $\frac{.75-0}{.19-1} = -\frac{25}{27}$ of this line to be the same as $-\frac{p}{1-p}$ (since the line $px + (1-p)y = \text{const}$ has a slope $-\frac{p}{1-p}$: indeed $y = -\frac{p}{1-p}x + \text{const}$)

Hence we have to satisfy $-\frac{p}{1-p} = -\frac{25}{27} \Rightarrow$ Hence $p = \frac{25}{52}$ and the least favorable prior is $\underline{\left(\frac{25}{52}, \frac{27}{52}\right)}$.

d) What is the Bayes rule for the prior $(1/3, \frac{2}{3})$ over (θ_1, θ_2)?

The line $\frac{1}{3}X + \frac{2}{3}y$ = const represents points (X, Y) in the risk set with equivalent value of their Bayes risk. The slope of this line is equal to $\frac{-1/3}{2/3} = -\frac{1}{2}$. Hence, to find the Bayes rule w.r. this prior, we need to move lines with a slope of $(-\frac{1}{2})$ "most south-west" while still having an intersection with the risk set. By doing so, you see geometrically that you end up with the rule d_8 (look carefully at the graph). Hence d_8 $(1, 0)$ represents the risk point that corresponds to the Bayesian decision rule w.r. to the prior $(\frac{1}{3}, \frac{2}{3})$ on (θ_1, θ_2). In other words, d_8 is the Bayesian decision rule w.r. to the prior $(\frac{1}{3}, \frac{2}{3})$. Its Bayes risk is equal to:

$$\frac{1}{3} * R(\theta_1, d_8) + \frac{2}{3} R(\theta_2, d_8) = \frac{1}{3} * 1 + \frac{2}{3} * 0 = \boxed{\frac{1}{3}}$$