KTH Computer Science
and Communication

2E1395 - Pattern Recognition
Solutions to Introduction to Pattern Recognition, Chapter 2 a: Conditional Probability and Bayes Rule

Exercise 2A1

We can call X the observation ($X = i$ indicates that the program leader has opened door$_i$), and S the state ($S = j$ indicates that the car is behind door$_j$). We will first reason by intuition and then prove the solution analytically.

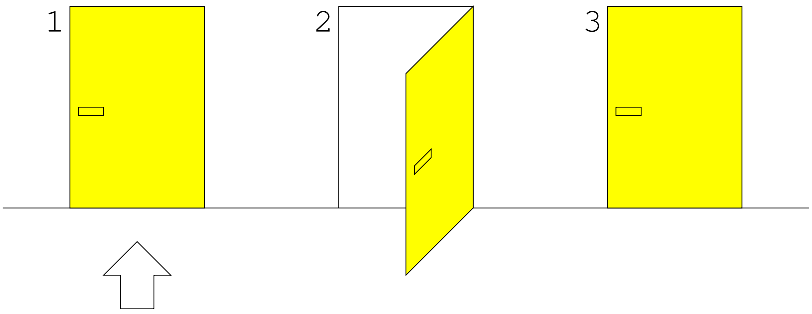

Figure 1. The TV show scenario

Intuition in this case can lead to a wrong conclusion: the *wrong* intuitive assumption can be that the only information brought by opening door$_2$, in the scenario considered in this example (fig. 1), is that the car is not behind it. This assumption would, as a consequence, suggest that, after opening door$_2$ the probability of finding the car behind door$_1$ is the same as the one of finding it behind door$_3$, in formulae: $P_{S|X}(1|2) = P_{S|X}(3|2) = \frac{1}{2}$.

To find what is wrong in the previous argument we consider a new scenario that is just a generalization of the one given by the text. The aim of this new argument will be to show that the choice of door$_1$, made by the player, affects the beaviour of the program leader and, as a consequence the amount of information that he provides opening door$_2$.

We consider this time a large number N of doors (fig. 2). Initially the probability that the car is behind any door is $P_S(i) = \frac{1}{N}$ (equal probabilities). Then the player selects one door (let's say door$_1$).

Every time the program leader opens a door he is constrained to not opening either the door that was selected (door$_1$), nor the door that hides the car. The point of the following argument is that after $N-2$ times the program leader opens a door, we are left with only two doors (let's say door$_1$ and door$_8$, as in fig. 2). Of those, door$_8$ could have been opened every time the program leader made his choice (unless the car was behind it), while door$_1$ could not be opened in any case. This should intuitively convince the reader, that not opening door$_8$ for $N-2$ times brings

Giampiero Salvi (giampi@kth.se)

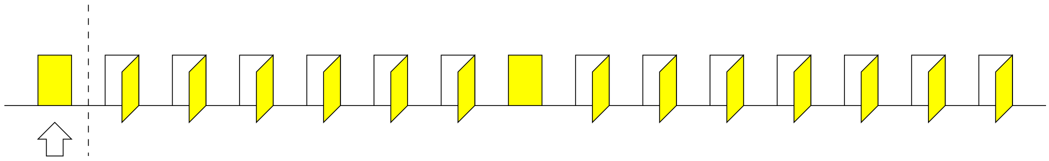

Figure 2. The generalized TV show scenario

more information about the car being behind it, then not opening door_1. In turns, the whole process should convince the reader (and the player) that the probability of the car being behind door_8 is much higher then that for door_1.

We now prove this conclusion analytically.

We go back to figure 1. According to the scenario, the player has chosen door_1. We observe that:

[I] The door that was chosen by the player cannot be opened by the program leader regardless of the position of the car:

$$P_{X|S}(1|i) = 0, \ \forall i$$

[II] If the car is behind door_i then the program leader cannot open that door:

$$P_{X|S}(i|i) = 0, \ \forall \ i$$

[III] The program leader chooses randomly among the doors that are allowed (not chosen by the player, and not hiding the car). The consequence of this is that the probabilities of opening a certain door have to be equally distributed among the available doors.

if we use these observations in the three possible cases we obtain:

1) if the car is behind door_1 $(S = 1)$

$$\begin{array}{ll} P_{X|S}(1|1) = 0 & \text{[I]} \\ P_{X|S}(2|1) = \frac{1}{2} & \text{[III]} \\ P_{X|S}(3|1) = \frac{1}{2} & \text{[III]} \end{array}$$

2) if the car is behind door_2 $(S = 2)$

$$\begin{array}{ll} P_{X|S}(1|2) = 0 & \text{[I]} \\ P_{X|S}(2|2) = 0 & \text{[II]} \\ P_{X|S}(3|2) = 1 & \end{array}$$

3) if the car is behind door_3 $(S = 3)$

$$\begin{array}{ll} P_{X|S}(1|3) = 0 & \text{[I]} \\ P_{X|S}(2|3) = 1 & \\ P_{X|S}(3|3) = 0 & \text{[II]} \end{array}$$

Note that in the first case the program leader can open indifferently the second or third door, while in the other two cases he is forced to open the only door that doesn't hide the car.

Applying Bayes rule, we can find the probability that the car was behind door$_i$ when door$_j$ is opened:

$$P_{S|X}(i|j) = \frac{P_{X|S}(j|i)P_S(i)}{P_X(j)} = \frac{P_{X|S}(j|i)P_S(i)}{\sum_{h=1}^{N} P_{X|S}(j|h)P_S(h)}$$

where all the terms in the formula are now known. Substituting in the cases of interest for this problem:

$$P_{S|X}(1|2) = \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{3}\left(\frac{1}{2}+0+1\right)} = \frac{1}{3}$$

$$P_{S|X}(3|2) = \frac{1\frac{1}{3}}{\frac{1}{3}\left(\frac{1}{2}+0+1\right)} = \frac{2}{3}$$

Showing that after door$_2$ has been opened the player should always decide to change his selection from door$_1$ to door$_3$ if he wants to maximize the probability to win the car.

Exercise 2A3

Since X and Y are statistically independent, their joint distribution is the product of each distributions $f_{XY}(x,y) = f_X(x)f_Y(y)$. Knowing f_{XY} it's always possible to deduct $f_Z(z)$ where Z is a function of X and Y.

a) In this case it's not necessary to use the standard method for deriving $f_Z(z)$. We know that a sum of Gaussian variables is still a gaussian variable. We only need to find μ_Z and σ_Z to characterize the distribution of Z:

$$\mu_Z = E[Z] = E[X+Y] = E[X] + E[Y] = \mu_X + \mu_Y$$

where the third step uses linearity of the expectation function.

$$\begin{aligned}
\sigma_Z^2 &= E[(Z-\mu_Z)^2] = E[(X+Y-\mu_X-\mu_Y)^2] \\
&= E[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)] \\
&= \sigma_X^2 + \sigma_Y^2 + 2E[(X-\mu_X)]E[(Y-\mu_Y)] \\
&= \sigma_X^2 + \sigma_Y^2
\end{aligned}$$

where the fourth step uses independence of X and Y

b) The easiest way to solve this is to note that once the value of X has been determined, this is not a stochastic variable anymore. $Z|X$ is hence a translation of the variable Y ($Z = x_1 + Y$) and its distribution is the same as for Y with an alteration of the mean: $f_{Z|X}(z|x_1)$ is gaussian with $\sigma_{Z|X} = \sigma_Y$ and $\mu_{Z|X} = \mu_Y + x_1$.
c) Same argument as for point b): if Z is fixed then $X|Z$ is a translation of Y ($X = z_1 - Y$), and has gaussian distribution with $\sigma_{X|Z} = \sigma_Y$ and $\mu_{X|Z} = z_1 - \mu_Y$.

Exercise 2A4

By definition the conditional probability density function $f_{X|S}(x|i)$ is the distribution of X when the value of S is known.

a) hence, directly from the text:

$$f_{X|S}(x,0) = N(0,1) \quad \text{and} \quad f_{X|S}(x,1) = N(1,2)$$

b) Since S can assume any of the two values with equal probablility $P_S(0) = P_S(1) = 1/2$. We know that the probability density function $f_X(x)$ is the derivate of the cumulative distribution $F_X(x)$ and that the last corresponds to the probability of the event $X < x$:

$$F_X(x) = P(X < x)$$

This can be written as the sum of the probabilities of the two incompatilbe events: $(X < x) \cap (S = 0)$ and $(X < x) \cap (S = 1)$, in formulae:

$$\begin{aligned} F_X(x) &= F_{XS}(x, 0) + F_{XS}(x, 1) \\ &= P_S(0)F_{X|S}(x|0) + P_S(1)F_{X|S}(x|1) \\ &= \frac{1}{2}F_{X|S}(x|0) + \frac{1}{2}F_{X|S}(x|1) \end{aligned}$$

where for the second passage we have applied Bayes rule. We can now compute the distributions by differentiating both members:

$$f_X(x) = \frac{1}{2}f_{X|S}(x|0) + \frac{1}{2}f_{X|S}(x|1)$$

c) applying Bayes:

$$\begin{aligned} P_{S|X}(0|0.3) &= \frac{P_S(0)P_{X|S}(0.3|0)}{P_X(0.3)} = \frac{P_S(0)P_{X|S}(0.3|0)}{P_S(0)P_{X|S}(0.3|0) + P_S(1)P_{X|S}(0.3|1)} \\ &= \frac{P_{X|S}(0.3|0)}{P_{X|S}(0.3|0) + P_{X|S}(0.3|1)} \end{aligned}$$

where we have simplified $P_S(i) = 1/2,\ \forall i$. Note that, being X a continuous varible, it's always $P_{X|S}(x|i) = 0,\ \forall x, i$. The fraction can be solved as a limit over an interval:

$$P_{S|X}(0|0.3) = \lim_{dx \to 0} \frac{f_{X|S}(0.3|0)dx}{f_{X|S}(0.3|0)dx + f_{X|S}(0.3|1)dx} = \frac{f_{X|S}(0.3|0)}{f_{X|S}(0.3|0) + f_{X|S}(0.3|1)}$$

Exercise 2A5

Two variables X and S are defined with distributions:

$$\begin{aligned} X &\to N(1, 1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}} \\ S &\to P_S(1) = 0.8,\ P_S(2) = 0.2 \end{aligned}$$

a) If $S = 1$ then $Z = X$ and $f_{Z|S}(z|1) = f_X(z) = N(1, 1)$
if $S = 2$ then $Z = 2X$ which is still a gaussian variable with mean and variance given by:

$$\begin{aligned} \mu_Z &= E[Z] = E[2X] = 2\mu_X = 2 \\ \sigma_Z^2 &= E[(Z - \mu_Z)^2] = E[(2X - 2\mu_X)^2] = 4E[(X - \mu_X)^2] = 4\sigma_X^2 = 4 \end{aligned}$$

and hence $f_{Z|S}(z|2) = N(2, 2)$. Summarizing:

$$f_{Z|S}(z|i) = N(i, i) = \frac{1}{\sqrt{2\pi}i}e^{-\frac{(x-i)^2}{2i^2}}$$

b) The unconditioned distribution for the variable Z is given by (look at the previous exercise for a formal proof):

$$f_Z(z) = P_S(1) f_{Z|S}(z|1) + P_S(2) f_{Z|S}(z|2)$$

c) Using Bayes rule,

$$P_{S|Z}(i|z_1) = \frac{P_S(1) P_{Z|S}(z_1|1)}{P_Z(z_1)} = \lim_{dx \to 0} \frac{P_S(1) f_{Z|S}(z_1|1) dx}{f_Z(z_1) dx} = \frac{P_S(1) f_{Z|S}(z_1|1)}{f_Z(z_1)}$$

d) Again using Bayes:

$$P_{X|Z}(x|z_1) = \frac{P_X(x) f_{Z|X}(z_1|x)}{f_Z(z_1)}$$

Exercise 2A6

a) By definition of the transition matrix, a_{ij} is for each t the probability of going to state j at time t when we were at time $t-1$ in state i, that is: $P_{S_t|S_{t-1}}(j|i)$. If we set $t = 13$,

$$P_{S_{13}|S_{12}}(k|2) = a_{2k}$$

That corresponds for each k to the second row in matrix A.

b) Since S_t forms a *first-order* Markow chain, the conditional probability of state S_t to the previous outcomes can be simplified:

$$P_{S_t|S_{t-1}S_{t-2}\cdots S_1} = P_{S_t|S_{t-1}}$$

then in our case:

$$P_{S_{13}|S_{11},S_{12}}(k|1,2) = P_{S_{13}|S_{12}}(k|2) = a_{2k}$$

as in the previous point.

d) This time we know the outcome of S at time t and we want the conditional probability at the previous time step. Applying Bayes:

$$\begin{aligned} P_{S_{11}|S_{12}}(k|2) &= \frac{P_{S_{11}}(k) P_{S_{12}|S_{11}}(2|k)}{P_{S_{12}}(2)} \\ &= \frac{P_{S_{11}}(k) P_{S_{12}|S_{11}}(2|k)}{\sum_k P_{S_{11}}(k) P_{S_{12}|S_{11}}(2|k)} \end{aligned}$$

In principle the unconditioned probabilities $P_{S_{11}}(k)$ that appear in this formula cannot be computed if the prior probability at $t = 0$ is not given. In this case the problem assumes that the prior was such that $P_{S_{11}}(k)$ is the same for each k. We can then simplify those terms, obtaining:

$$P_{S_{11}|S_{12}}(k|2) = \frac{P_{S_{12}|S_{11}}(2|k)}{\sum_k P_{S_{12}|S_{11}}(2|k)} = \frac{a_{k2}}{\sum_k a_{k2}} = \begin{cases} 1/9, & k = 1 \\ 8/9, & k = 2 \\ 0, & k = 3 \end{cases}$$

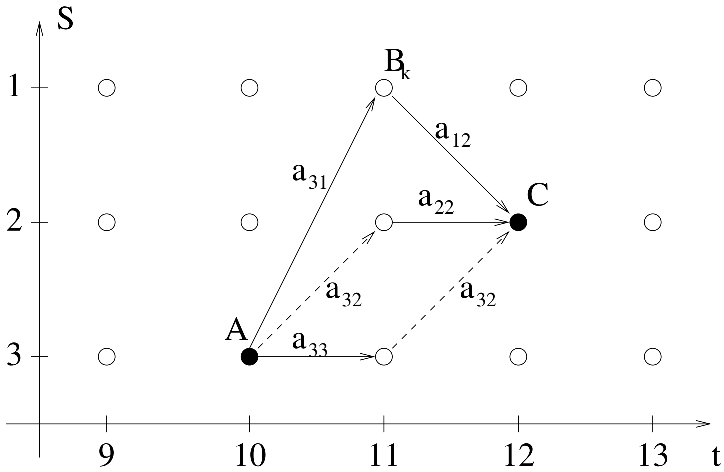

Figure 3. State transitions in the Markov chain

d) The problem is best described looking at figure 3. In the figure all transitions from $S_{10} = 3$ to $S_{12} = 2$ are indicated with an arrow to which the transition probability is attached. Dasched transitions have 0 probability as referring to matrix A. The probability for each path is the product of transition probabilities along the path. Of the three paths in the figure, only the one that goes through $S_{11} = 1$ has non zero probability. This also means that, being only one path possible,

$$P_{S_{11}|S_{10},S_{12}}(k|3,2) = \begin{cases} 1 & k = 1 \\ 0 & \text{otherwise} \end{cases}$$

This intuitive solution can be formalized as follows: for simplicity we call A the event $S_{10} = 3$, B_k the event $S_{11} = k$, and C the event $S_{12} = 2$ (see figure). We can then write:

$$P_{S_{11}|S_{10},S_{12}}(k|3,2) = P(B_k|AC)$$

Applying Bayes rule (with respect to the conditioning event C):

$$P(B_k|AC) = \frac{P(B_k|A)P(C|AB_k)}{P(C|A)}$$

Now we note that $P(C|AB_k) = P(C|B_k)$ because of the assumption of first-order Markov chain (see also point b). We note also that $P(C|A)$ is the probability of $S_{12} = 2$ conditioned to $S_{10} = 3$ for any value of S_{11} and can be hence written as the sum of the product of probabilities along all possible paths:

$$P(C|A) = \sum_k P(B_k|A)P(C|B_k)$$

Substituting we get the general solution to this problem:

$$P_{S_{11}|S_{10},S_{12}}(k|3,2) = \frac{P_{S_{11}|S_{10}}(k|3)P_{S_{12}|S_{11}}(2|k)}{\sum_k P_{S_{11}|S_{10}}(k|3)P_{S_{12}|S_{11}}(2|k)} = \frac{a_{3k}a_{k2}}{\sum_k a_{3k}a_{k2}}$$

That gives the same numerical solution already discussed.