Homework 4

Solutions

1. (Bayesian coin flip)

   (a) The model makes sense because we suspect that the coin that is more prone to be heads than tails, but we are uncertain about the exact bias so we just assume it is uniform between 1/2 (fair coin) and 1 (coin that always lands heads). To compute the probability of heads, we integrate over all the possible values of the Bernouilli parameter. Let $\tilde{r}$ be the result of the coin flip and $\tilde{\theta}$ the parameter of the Bernouilli. Since $\tilde{r}$ is uniform between 1/2 and 1 $f_{\tilde{\theta}}(\theta)$ is equal to 2 for $1/2 \leq \theta \leq 1$ and zero otherwise.

$$P(\text{heads}) = p_{\tilde{r}}(1) = \int_{\theta=1/2}^{1} f_{\tilde{\theta}}(\theta)\, p_{\tilde{r}\,|\,\tilde{\theta}}(1\,|\,\theta)\, \mathrm{d}\theta = \int_{\theta=1/2}^{1} 2\,\theta\, \mathrm{d}\theta = 1 - \frac{1}{4} = \frac{3}{4}, \tag{1}$$

$$P(\text{tails}) = 1 - P(\text{heads}) = \frac{1}{4}. \tag{2}$$

   (b) The conditional pdf on the bias of the coin flip conditioned on tails equals

$$f_{\tilde{\theta}\,|\,\tilde{r}}(\theta\,|\,0) = \frac{f_{\tilde{\theta}}(\theta)\, p_{\tilde{r}\,|\,\tilde{\theta}}(0)}{p_{\tilde{r}}(0)} \tag{3}$$

$$= \frac{f_{\tilde{\theta}}(\theta)\,(1-\theta)}{1/4} = \begin{cases} 8\,(1-\theta) & 1/2 \leq \theta \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

   Note that the conditional pdf is more skewed towards 1/2, as shown on Figure 1. It makes sense because the coin flip is tails. Intuitively the model should be adjusted towards the coin flip being less biased.

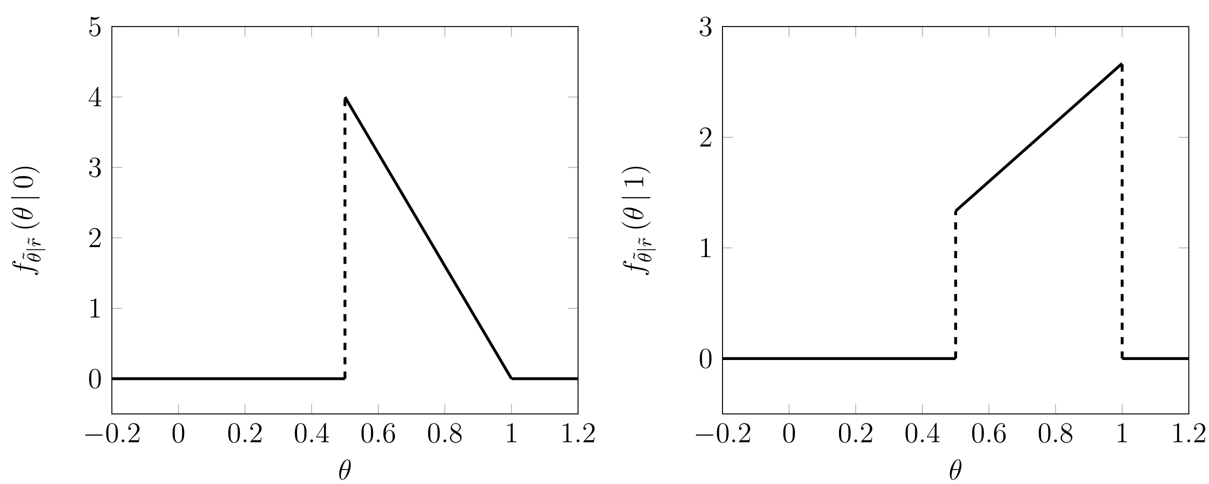


Figure 1: Conditional pdf of the bias of the coin flip given tails (left) and heads (right).

The conditional pdf on the bias of the coin flip conditioned on heads equals

$$f_{\tilde{\theta}|\tilde{r}}(\theta\,|\,1) = \frac{f_{\tilde{\theta}}(\theta)\,p_{\tilde{r}|\tilde{\theta}}(1)}{p_{\tilde{r}}(1)} = \frac{f_{\tilde{\theta}}(\theta)\,\theta}{3/4} = \begin{cases} \frac{8\,\theta}{3} & 1/2 \leq \theta \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

In this case the conditional pdf is more skewed towards 1/2, as shown on Figure 1. Intuitively, you are adjusting your model by incorporating some evidence that supports that the coin might be biased towards heads.

(c) No matter what evidence we observe model we will always assign zero probability density to any value of the bias below 0.5, so we should consider changing the prior.

2. (Halloween parade)

(a)

$$\begin{aligned} \mathrm{P}(\tilde{w} \neq \tilde{r}) &= p_{\tilde{w},\tilde{r}}(0,1) + p_{\tilde{w},\tilde{r}}(1,0) && (6) \\ &= p_{\tilde{r}}(1)\,p_{\tilde{w}\,|\,\tilde{r}}(0\,|\,1) + p_{\tilde{r}}(0)\,p_{\tilde{w}\,|\,\tilde{r}}(1\,|\,0) && (7) \\ &= 0.24. && (8) \end{aligned}$$

(b) It is not reasonable to assume that the forecast and the humidity are independent, even if we know that the forecast does not take the humidity into account. The reason is that both variables are linked through the rain. For example, if $\tilde{w} = 1$ then the humidity is probably high (because it probably will rain) and if $\tilde{w} = 0$ the humidity is probably low (because it probably won't rain). In contrast, conditioned on whether it rains or not, it is quite reasonable to assume independence of $\tilde{w}$ and $\tilde{h}$ because $\tilde{h}$ is not used to produce the forecast.

(c) By the chain rule for discrete and continuous random variables

$$\begin{aligned} p_{\tilde{r}|\tilde{w},\tilde{h}}(r|w,h) &= \frac{f_{\tilde{h}|\tilde{w},\tilde{r}}(h|w,r)p_{\tilde{w}|\tilde{r}}(w|r)p_{\tilde{r}}(r)}{p_{\tilde{w}}(w)f_{\tilde{h}|\tilde{w}}(h|w)} && (9) \\ &= \frac{f_{\tilde{h}|\tilde{r}}(h|r)p_{\tilde{w}|\tilde{r}}(w|r)p_{\tilde{r}}(r)}{\sum_{r=0}^{1} f_{\tilde{h}|\tilde{r}}(h|r)p_{\tilde{w}|\tilde{r}}(w|r)p_{\tilde{r}}(r)}. && (10) \end{aligned}$$

This expression equals to zero unless h is between 0.5 and 0.7. If h is between 0.6 and 0.7 then it is equal to 1. If h is between 0.5 and 0.6 then

$$\begin{aligned} p_{\tilde{r}|\tilde{w},\tilde{h}}(1|1,h) &= \frac{0.2 \cdot 0.8 \cdot 5}{0.2 \cdot 0.8 \cdot 5 + 0.8 \cdot 0.25 \cdot 2} && (11) \\ &= 0.667, && (12) \\ p_{\tilde{r}|\tilde{w},\tilde{h}}(1|0,h) &= \frac{0.2 \cdot 0.2 \cdot 5}{0.2 \cdot 0.2 \cdot 5 + 0.8 \cdot 0.75 \cdot 2} && (13) \\ &= 0.143. && (14) \end{aligned}$$

The conditional pmf is only well defined for $0.1 \le h \le 0.7$,

$$p_{\tilde{r}|\tilde{w},\tilde{h}}(1|w,h) = \begin{cases} 0 & \text{if } 0.1 \le h \le 0.5 \\ 0.667 & \text{if } 0.5 \le h \le 0.6 \quad \text{and } w = 1 \\ 0.143 & \text{if } 0.5 \le h \le 0.6 \quad \text{and } w = 0 \\ 1 & \text{if } 0.6 \le h \le 0.7. \end{cases} \tag{15}$$

We predict no rain (a) if the humidity is between 0.1 and 0.5 or (b) if $0.5 \le h \le 0.6$ and $w = 0$. Otherwise, we predict rain.

(d) According to the model, if the humidity is between 0.1 and 0.5 then we are always right. Similarly, if the humidity is between 0.6 and 0.7 we are also always right. For h between 0.5 and 0.6 we make a mistake under two scenarios: (1) $w = 0$ and it rains and (2) $w = 1$ and it does not rain. The probability of these two events happening equals

$$\begin{aligned} \mathrm{P}(\text{Error}) &= \mathrm{P}(\tilde{r} = 0, w = 1, 0.5 \le \tilde{h} \le 0.6) + \mathrm{P}(\tilde{r} = 1, w = 0, 0.5 \le \tilde{h} \le 0.6) \\ &= \int_{h=0.5}^{0.6} f_{\tilde{h}|\tilde{r}}(h \,|\, 0)\, p_{\tilde{w}|\tilde{r}}(1|0) p_{\tilde{r}}(0)\, \mathrm{d}h + \int_{h=0.5}^{0.6} f_{\tilde{h}|\tilde{r}}(h \,|\, 1)\, p_{\tilde{w}|\tilde{r}}(0|1) p_{\tilde{r}}(1)\, \mathrm{d}h \\ &= \int_{h=0.5}^{0.6} 0.8 \cdot 2 \cdot 0.25 \, \mathrm{d}h + \int_{h=0.5}^{0.6} 0.2 \cdot 5 \cdot 0.2 \, \mathrm{d}h = 0.06. \end{aligned}$$

3. (Markov chain) The transition matrix of the Markov chain is

$$T := \begin{bmatrix} 0 & 0.1 & 0 \\ 1 & 0 & 1 \\ 0 & 0.9 & 0 \end{bmatrix}. \tag{16}$$

T has three eigenvectors

$$q_1 := \begin{bmatrix} 0.0741 \\ 0.741 \\ 0.667 \end{bmatrix}, \qquad q_2 := \begin{bmatrix} 0.0741 \\ -0.741 \\ 0.667 \end{bmatrix}, \qquad q_3 := \begin{bmatrix} 0.707 \\ 0 \\ -0.707 \end{bmatrix}. \tag{17}$$

The corresponding eigenvalues are $\lambda_1 := 1$, $\lambda_2 := -1$ and $\lambda_3 := 0$. The initial state vector for $i = 0$ is

$$p_{\tilde{x}[1]} := \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \tag{18}$$

We use the eigenvector matrix

$$Q := \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \tag{19}$$

to compute

$$Q^{-1} p_{\tilde{x}[1]} = \begin{bmatrix} 0.675 \\ 0.675 \\ 1.27 \end{bmatrix}. \tag{20}$$

So

$$p_{\tilde{x}[1]} = 0.675\, q_1 + 0.675\, q_2 + 1.27\, q_3. \tag{21}$$

We consequently have

$$\begin{aligned} p_{\tilde{x}[i]} &= T^{\,i-1}\, p_{\tilde{x}[1]} && (22) \\ &= T^{\,i} \left(0.675\, q_1 + 0.675\, q_2 + 1.27\, q_3\right) && (23) \\ &= 0.675\, \lambda_1^{i-1} q_1 + 0.675\, \lambda_2^{i-1} q_2 + 1.27\, \lambda_3^{i-1} q_3 && (24) \\ &= 0.675 \left(q_1 + (-1)^{i-1}\, q_2\right) = \begin{cases} \begin{bmatrix} 0.1 \\ 0 \\ 0.9 \end{bmatrix} & \text{if } i \text{ is odd,} \\ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} & \text{if } i \text{ is even.} \end{cases} && (25) \end{aligned}$$

For even i the chain is always at B. For odd i it is either at A or C with probability 0.1 and 0.9.

4. (Heart-disease detection)

   (a) By Bayes rule, the chain rule and the conditional independence assumption

   $$p_{\tilde{h}|\tilde{s},\tilde{c}}(h|s,c) = \frac{p_{\tilde{h}}(h) p_{\tilde{s},\tilde{c}|\tilde{h}}(s,c|h)}{p_{\tilde{s},\tilde{c}}(s,c)} = \frac{p_{\tilde{h}}(h) p_{\tilde{s}|\tilde{h}}(s|h) p_{\tilde{c}|\tilde{h}}(c|h)}{p_{\tilde{s},\tilde{c}}(s,c)}.$$

   Since the denominator does not depend on h, the MAP rule reduces to

   $$\hat{h}(s,c) = \begin{cases} 0 & \text{if } p_{\tilde{h}}(1) p_{\tilde{s}|\tilde{h}}(s|1) p_{\tilde{c}|\tilde{h}}(c|1) < p_{\tilde{h}}(0) p_{\tilde{s}|\tilde{h}}(s|0) p_{\tilde{c}|\tilde{h}}(c|0), \\ 1 & \text{otherwise.} \end{cases}$$

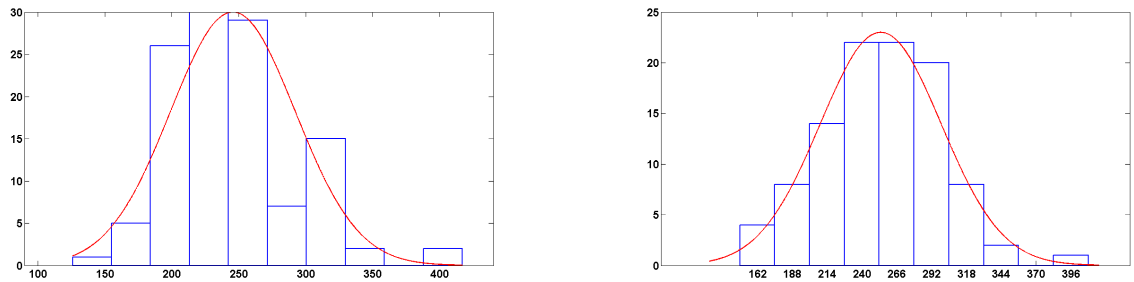
   (b) The error rate is 0.18.

   (c) By Bayes rule for mixed random variables, the chain rule and the conditional independence assumption

   $$p_{\tilde{h}|\tilde{s},\tilde{c},\tilde{x}}(h|s,c,x) = \frac{p_{\tilde{h}}(h) p_{\tilde{s},\tilde{c}|\tilde{h}}(s,c|h) f_{\tilde{x}|\tilde{h},\tilde{s},\tilde{c}}(x|h,s,c)}{f_{\tilde{x}|\tilde{s},\tilde{c}}(x|s,c) p_{\tilde{s},\tilde{c}}(s,c)} = \frac{p_{\tilde{h}}(h) p_{\tilde{s}|\tilde{h}}(s|h) p_{\tilde{c}|\tilde{h}}(c|h) f_{\tilde{x}|\tilde{h}}(x|h)}{f_{\tilde{x}|\tilde{s},\tilde{c}}(x|s,c) p_{\tilde{s},\tilde{c}}(s,c)}.$$

   Since the denominator does not depend on h, the MAP rule reduces to

   $$\hat{h}(s,c) = \begin{cases} 0 & \text{if } p_{\tilde{h}}(1) p_{\tilde{s}|\tilde{h}}(s|1) p_{\tilde{c}|\tilde{h}}(c|1) f_{\tilde{x}|\tilde{h}}(x|1) < p_{\tilde{h}}(0) p_{\tilde{s}|\tilde{h}}(s|0) p_{\tilde{c}|\tilde{h}}(c|0) f_{\tilde{x}|\tilde{h}}(x|0), \\ 1 & \text{otherwise.} \end{cases}$$



$\tilde{h} = 0$ $\tilde{h} = 1$

Figure 2: Histogram (blue) and approximation of the conditional pdf (red) of $\tilde{x}$ given the two possible values of $\tilde{h}$.

(d) Figure 2 shows the histograms and the approximations to the conditional pdfs of $\tilde{x}$ given $\tilde{h}$ (scaled so that they are comparable to the histograms).

(e) The error rate is 0.14. Given that there is not much difference between the cholesterol histograms in the presence and absence of heart disease and that the Gaussian fit is clearly not a very accurate representation of the empirical pdfs, we should not trust the result too much.

(f) If we have enough data to estimate the joint distribution in a stable way, then this approach would probably improve the results. However, we would need more data than just 218 patients. Even without the cholesterol, the joint pmf of $\tilde{c}$, $\tilde{s}$ and $\tilde{h}$ has $2 \cdot 2 \cdot 4 - 1 = 15$ parameters, so we have less than 15 patients per parameter. In general, joint distributions are very difficult to approximate, since the number of parameters that are needed to characterize them are usually very large. Introducing some assumptions that are only approximately true but decrease significantly the number of parameters that we need to estimate often yield better results when there is limited data.