Probability and Statistics

Solution sheet 2

Solution 2.1

(a) By definition of a Laplace Model:

$$\begin{aligned}\mathbb{P}(A_n) &= \frac{|A_n|}{|\Omega|} = \frac{N(N-1)(N-2)...(N-n+1)}{N^n} \\ &= \prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right).\end{aligned}$$

(b) Take $\Omega = \{(\omega_i)_{i=1}^n : \omega_i \in \{1,...,N\}\}$. For the lower bound we will work with $\mathbb{P}(A_n^c)$.

$$\begin{aligned}\mathbb{P}(A_n^c) &= \mathbb{P}(\{\exists j,k \in \{1,..,n\}, k<j\} : \omega_j = \omega_k) \\ &= \mathbb{P}\left(\bigcup_{j=1}^{n}\bigcup_{k=1}^{j-1}\{\omega_k = \omega_j\}\right) \le \sum_{j=1}^{n}\sum_{k=1}^{j-1}\mathbb{P}(\{\omega_k = \omega_j\}) \\ &= \frac{1}{N}\sum_{j=1}^{n}(j-1) = \frac{(n-1)n}{2N},\end{aligned}$$

using that $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n)$ we conclude.

For the upper bound remember that for all $x > -1$, $\ln(1+x) \le x$. Then,

$$\begin{aligned}\mathbb{P}(A_n) &= \prod_{j=1}^{n-1}\left(1-\frac{j}{N}\right) = \prod_{j=1}^{n-1}\exp\left(\ln\left(1-\frac{j}{N}\right)\right) \\ &\le \prod_{j=1}^{n-1}\exp\left(-\frac{j}{N}\right) = \exp\left(-\sum_{j=1}^{n-1}\frac{j}{N}\right) \\ &= \exp\left(-\frac{n(n-1)}{2N}\right).\end{aligned}$$

(c) If $\mathbb{P}(A_n) < \frac{1}{2}$

$$\begin{aligned}&1-\frac{n(n-1)}{2*365} < \frac{1}{2} \\ \Rightarrow & n^2 - n - 365 > 0 \\ \Rightarrow & n > \frac{1+\sqrt{1461}}{2} \sim 19.1.\end{aligned}$$

Given that $\mathbb{P}(n_{\min}) > \frac{1}{2}$, $n_{\min} \ge 20$.

Also, we have that if

$$\begin{aligned}&\exp\left(-\frac{n(n-1)}{2*365}\right) < \frac{1}{2} \\ \Leftrightarrow & n^2 - n - 2*365\ln 2 > 0 \\ \Leftrightarrow & n > \frac{1+\sqrt{2*365\ln 2+1}}{2} \sim 22.9\end{aligned}$$

so $P(A_n) < \frac{1}{2}$ if $n > 23$. Then $n_{\min} \leq 23$. Given that $\mathbb{P}(A_{20}) \sim 0.59$, $\mathbb{P}(A_{21}) \sim 0.55$, $\mathbb{P}(A_{22}) =\sim 0.52$ and $\mathbb{P}(A_{23}) \sim 0.49$, we have that $n_{\min} = 23$.

This problem is similar to the Birthday Problem given that we take the following assumptions:

- The number of people who were born in each day of the year (but February 29th) is the same.
- No one is born on February 29th.

Solution 2.2 Let A_i denote the event: "the ith coin is tossed", and H_n denote "we obtain a head at the nth toss".

(a) The posterior probability that the ith coin is tossed after one toss is $\mathbb{P}(A_i|H_1)$. By the Bayes theorem:

$$\mathbb{P}(A_i|H_1) = \frac{\mathbb{P}(H_1|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{3} \mathbb{P}(H_1|A_j)\mathbb{P}(A_j)}.$$

The prior probability of A_j is $\mathbb{P}(A_j) = 1/3$ since the coin is selected uniformly at random, thus

$$\mathbb{P}(A_i|H_1) = \frac{p_i/3}{(p_1 + p_2 + p_3)/3} = \frac{2p_i}{3} := q_i.$$

That is

$$\mathbb{P}(A_1|H_1) = 1/6,\ \mathbb{P}(A_2|H_1) = 1/3,\ \mathbb{P}(A_3|H_1) = 1/2.$$

(b) We have to compute the probability $\mathbb{P}(H_2|H_1)$.

$$\mathbb{P}(H_2|H_1) = \sum_{i=1}^{3} \mathbb{P}(A_i \cap H_2|H_1) = \sum_{i=1}^{3} p_i \mathbb{P}(A_i|H_1) = 7/12.$$

The second equality is due to the fact that conditioned on A_i, H_1 and H_2 are independent:

$$\begin{aligned}\mathbb{P}(A_i \cap H_2|H_1) &= \frac{\mathbb{P}(A_i \cap H_1 \cap H_2)}{\mathbb{P}(H_1)} = \frac{\mathbb{P}(H_1 \cap H_2|A_i)\mathbb{P}(A_i)}{\mathbb{P}(H_1)} \\ &= \frac{\mathbb{P}(H_1|A_i)\mathbb{P}(H_2|A_i)\mathbb{P}(A_i)}{\mathbb{P}(H_1)} = \frac{p_i \mathbb{P}(H_1|A_i)\mathbb{P}(A_i)}{\mathbb{P}(H_1)} \\ &= p_i \mathbb{P}(A_i|H_1)\end{aligned}$$

(c) Let $\mathbb{Q}(\cdot) = \mathbb{P}(\cdot|B)$, $\mathbb{Q}$ is a probability measure on Ω. For all events X and Y,

$$\begin{aligned}\mathbb{P}(X|B \cap Y) &= \frac{\mathbb{P}(X \cap B \cap Y)}{\mathbb{P}(B \cap Y)} = \frac{\mathbb{P}(X \cap Y|B)\mathbb{P}(B)}{\mathbb{P}(Y|B)\mathbb{P}(B)} \\ &= \frac{\mathbb{Q}(X \cap Y)}{\mathbb{Q}(Y)} = \mathbb{Q}(X|Y).\end{aligned}$$

The formula that we should prove can then be reduced to

$$\mathbb{Q}(A_i|C) = \frac{\mathbb{Q}(A_i)\mathbb{Q}(C|A_i)}{\sum_{j=1}^{k} \mathbb{Q}(A_j)\mathbb{Q}(C|A_j)},$$

which is exactly the Bayes Theorem.

(d) We want to compute $\mathbb{P}(A_i|H_1 \cap H_2)$. Apply the above formula to $B = H_1$, $C = H_2$, A_i as before and $\mathbb{Q}(\cdot) = \mathbb{P}(\cdot|H_1)$. The prior probabilities are $q_i := \mathbb{Q}(A_i)$ calculated in a). Also by the conditional independence of H_1 and H_2 given A_i, we have that under $\mathbb{Q}$, the probability of getting a head from the ith coin is unchanged:

$$\mathbb{Q}(H_2|A_i) = \mathbb{P}(H_2|A_i \cap H_1) = \frac{\mathbb{P}(H_1 \cap H_2 \cap A_i)}{\mathbb{P}(A_i \cap H_1)}$$
$$= \frac{\mathbb{P}(H_1 \cap H_2|A_i)}{\mathbb{P}(H_1|A_i)} = \mathbb{P}(H_2|A_i) = p_i.$$

The posterior probability is analogous to a), by replacing the prior probability by $\mathbb{Q}(A_i)$. Thus

$$\mathbb{P}(A_i|H_1 \cap H_2) = \mathbb{Q}(A_i|H_2) = \frac{p_i q_i}{p_1 q_1 + p_2 q_2 + p_3 q_3} = \frac{p_i q_i}{7/12}.$$

The posterior probabilities are respectively 1/14, 2/7 and 9/14.

(e) The recursive relation is obtained as in the previous question: let $q_{n,i}$ denote the posterior probability after n tosses of "the coin selected at the beginning is the ith coin". Then for every i, $q_{0,i} = 1/3$ and

$$q_{n+1,i} = \frac{p_i q_{n,i}}{p_1 q_{n,1} + p_2 q_{n,2} + p_3 q_{n,3}}.$$

Solution 2.3

(a) Let $k = \sum_{j=1}^{n} x_j$ the amount of red balls taken out. Then

$$\mathbb{P}(X_1 = x_1, ..., X_n = x_n) = \sum_{i=1}^{m} \mathbb{P}(X_1 = x_1, ..., X_n = x_n \mid \text{Urn } i \text{ is chosen})\, \mathbb{P}(\text{Urn } i \text{ is chosen})$$
$$= \sum_{i=1}^{m} \left(\frac{2i-1}{2m}\right)^k \left(\frac{2m-2i+1}{2m}\right)^{n-k} \frac{1}{m}.$$

We have that X_1 and X_2 are not independent because:

$$\mathbb{P}(\{X_1 = 1\} \cap \{X_2 = 1\}) = \sum_{i=1}^{m} \mathbb{P}(\{X_1 = 1\} \cap \{X_2 = 1\} \mid \text{Urn } i \text{ is chosen})\, \mathbb{P}(\text{Urn } i \text{ is chosen})$$
$$= \sum_{i=1}^{m} \left(\frac{2i-1}{2m}\right)^2 \frac{1}{m}$$
$$> \left(\sum_{i=1}^{m} \left(\frac{2i-1}{2m}\right) \frac{1}{m}\right)^2 = \mathbb{P}(X_1 = 1)\, \mathbb{P}(X_2 = 1).$$

Indeed, the first variable gives "information" about the chosen urn so it also gives information about the second variable.

(b) By definition of conditional probabilities

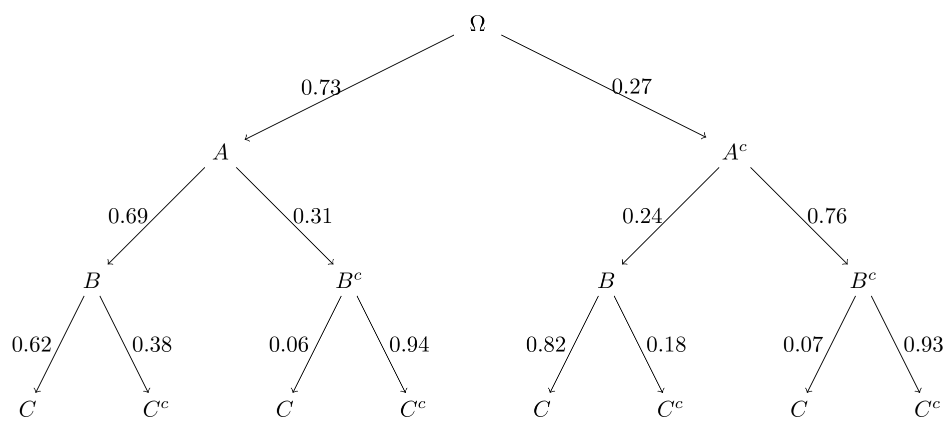
$$\mathbb{P}(\text{The urn chosen is } i \mid X_1 = x_1, ..., X_n = x_n)$$
$$= \frac{\mathbb{P}(\text{The urn chosen is } i, X_1 = x_1, ..., X_n = x_n)}{\mathbb{P}(X_1 = x_1, ..., X_n = x_n)}$$
$$= \frac{\frac{1}{m}\left(\frac{2i-1}{2m}\right)^k \left(\frac{2m-2i+1}{2m}\right)^{n-k}}{\sum_{j=1}^{m} \left(\frac{2j-1}{2m}\right)^k \left(\frac{2m-2j+1}{2m}\right)^{n-k} \frac{1}{m}}$$
$$= \frac{\left(\frac{2i-1}{2m}\right)^k \left(\frac{2m-2i+1}{2m}\right)^{n-k}}{\sum_{j=1}^{m} \left(\frac{2j-1}{2m}\right)^k \left(\frac{2m-2j+1}{2m}\right)^{n-k}}.$$

	$i=1$	$i=2$	$i=3$
$k=0$	0.817	0.176	0.007
$k=1$	0.439	0.474	0.088
$k=2$	0.088	0.474	0.439
$k=3$	0.007	0.176	0.817

(c) Using the previous question, we obtain

Solution 2.4

(a) The tree can be drawn as:



(b) The probability of being accepted, given than you are a woman who postulated at the department I is $P(C \mid A^c \cap B) = 0.82$. That value is bigger than the probability of being accepted, given than you are a man who postulated at the department I, $\mathbb{P}(C \mid A \cap B) = 0.62$. This indicates that in department I females are not disadvantaged.

The probability of being accepted, given than you are a woman who postulated to the department is $P(C \mid A^c \cap B^c) = 0.07$. This value is bigger than the probability of being accepted given than you are a man who postulated at the department II $\mathbb{P}(C \mid A \cap B^c) = 0.06$. This indicates that in department II females are neither disadvantaged.

(c) We compute the following probabilities

$$\begin{aligned}\mathbb{P}[C|A^c] = \frac{\mathbb{P}[C \cap A^c]}{\mathbb{P}[A^c]} &= \frac{\mathbb{P}[C \cap A^c \cap B] + \mathbb{P}[C \cap A^c \cap B^c]}{\mathbb{P}[A^c]} \\ &= \frac{0.82 \cdot 0.27 \cdot 0.24 + 0.07 \cdot 0.27 \cdot 0.76}{0.27} \sim 0.25,\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}[C|A] = \frac{\mathbb{P}[C \cap A]}{\mathbb{P}[A^c]} &= \frac{\mathbb{P}[C \cap A \cap B] + \mathbb{P}[C \cap A \cap B^c]}{\mathbb{P}[A]} \\ &= \frac{0.62 \cdot 0.69 \cdot 0.73 + 0.06 \cdot 0.31 \cdot 0.73}{0.73} \sim 0.45.\end{aligned}$$

This shows that the percentage of women accepted are less than that of the men. This is not explained by the gender, but much more but the fact that women apply to the department with bigger rejection rate.

Solution 2.5

(a) By definition $e = xy \in K_A$ iff $x \in A \wedge y \notin A$ or $x \notin A \wedge y \in A$. Given that

$$|\{x \in A, y \notin A\}| = |2^{V \setminus \{x,y\}}| = 2^{|V|-2},$$

we have that:

$$\begin{aligned}\mathbb{P}(xy \in K_A) &= \mathbb{P}(\{x \in A \wedge y \notin A\} \cup \{x \notin A \wedge y \in A\}) \\ &= \mathbb{P}(\{x \in A \wedge y \notin A\}) + \mathbb{P}(\{x \in A \wedge y \notin A\}) \\ &= \frac{2^{|V|-2}}{2^{|V|}} + \frac{2^{|V|-2}}{2^{|V|}} = \frac{1}{2},\end{aligned}$$

where the last part follows from the symmetry of the problem.

(b) Since $|K_A| = \sum_{k \in K} \mathbb{1}_{k \in K_A}$, we have by linearity, and the previous question

$$\mathbb{E}\left[|K_A|\right] = \mathbb{E}\left[\sum_{k \in K} \mathbb{1}_{k \in K_A}\right] = \frac{|K|}{2}.$$

(c) Given that the expectation of $|K_A|$ is equal to $\frac{|K|}{2}$ there should be a value of A so that $|K_A| \geq \frac{|K|}{2}$. If it is not the case, it holds

$$\begin{aligned}\mathbb{E}\left[|K_A|\right] &= \sum_{k \in \mathbb{N}} k\mathbb{P}(|K_A| = k) \\ &= \sum_{k=1}^{\lceil \frac{|K|}{2} - 1 \rceil} k\mathbb{P}(|K_A| = k) \\ &\leq \left\lceil \frac{|K|}{2} - 1 \right\rceil \sum_{k=1}^{\lceil \frac{|K|}{2} - 1 \rceil} \mathbb{P}(|K_A| = k) \\ &= \left\lceil \frac{|K|}{2} - 1 \right\rceil < \frac{|K|}{2},\end{aligned}$$

where we used $\lceil x \rceil := \inf\{n \in \mathbb{N} : x \leq n\}$, the largest integer smaller than x: it holds $\left\lceil \frac{|K|}{2} - 1 \right\rceil < \frac{|K|}{2}$.

Solution 2.6

(a) Let $\Omega = \{1,2,3\} \times \{2,3\}$. Then, for $i \in \{1,2,3\}$, $B_i = \{i\} \times \{2,3\}$ and $A_2 = \{1,3\} \times \{2\}$ and $A_3 = \{1,2\} \times \{3\}$. We posit the following probabilities:
$\mathbb{P}(B_1) = \mathbb{P}(B_2) = \mathbb{P}(B_3) = \frac{1}{3}$.
$\mathbb{P}(A_2|B_1) = \mathbb{P}(A_3|B_1) = \frac{1}{2}$.
$\mathbb{P}(A_2|B_2) = 0 \quad \mathbb{P}(A_3|B_2) = 1$.
$\mathbb{P}(A_2|B_3) = 1 \quad \mathbb{P}(A_3|B_3) = 0$.

(b) We compute, with Bayes' formula

$$\begin{aligned}\mathbb{P}(B_1|A_2) &= \frac{\mathbb{P}(A_2|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A_2|B_1)\mathbb{P}(B_1) + \mathbb{P}(A_2|B_2)\mathbb{P}(B_2) + \mathbb{P}(A_2|B_3)\mathbb{P}(B_3)} \\ &= \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}\frac{1}{3} + 0\frac{1}{3} + 1\frac{1}{3}} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}.\end{aligned}$$

$$\begin{aligned}\mathbb{P}(B_1|A_3) &= \frac{\mathbb{P}(A_3|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A_3|B_1)\mathbb{P}(B_1)+\mathbb{P}(A_3|B_2)\mathbb{P}(B_2)+\mathbb{P}(A_3|B_3)\mathbb{P}(B_3)}\\ &= \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}\frac{1}{3}+1\frac{1}{3}+0\frac{1}{3}} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}.\end{aligned}$$

It holds that $\mathbb{P}(B_1 \mid A_2) = \mathbb{P}(B_1 \mid A_3) = \mathbb{P}(B_1)$.

(c) You should pick the remaining door, because $\mathbb{P}(B_1 \mid A_j) = \frac{1}{3}$, and $\mathbb{P}(B_l \mid A_j) = \frac{2}{3}$ where $1 \neq l \neq j$. You did not obtain additional information about door No. 1, but you obtained additional information on the last door.