Lecturer: Prof. Dr. Mete SONER
Coordinator: Yilin WANG

ETH Zürich

Solution Series 2

Q1. THE BIRTHDAY PARADOX Take an urn with N balls numerated from $\{1, .., N\}$. Perform the experiment of extracting balls with replacement.

(a) Let $A_n :=$ "The first n balls extracted are different". Calculate $\mathbb{P}(A_n)$ (use a Laplace model).

(b) Prove the following inequalities:

$$1 - \frac{n(n-1)}{2N} \leq \mathbb{P}(A_n) \leq \exp\left(-\frac{n(n-1)}{2N}\right).$$

(c) Calculate $n_{\min} = \inf\{n \in \mathbb{N} : \mathbb{P}(A_n) < \frac{1}{2}\}$ for $N = 365$. Relate this problem with the Birthday Problem: " Find the probability that, in a group of n people, there is at least one pair who have the same birthday".

Solution

(a) By definition of a Laplace Model:

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

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

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

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

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

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

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

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

Given that $P(n_{\min}) > \frac{1}{2}$, $n_{\min} \geq 20$.
Also, we have that if

$$\begin{aligned}&\exp\left(-\frac{n(n-1)}{2*365}\right) < \frac{1}{2} \\ \Leftrightarrow & n^2 - n - 2*365\ln 2 > 0 \\ \Leftrightarrow & n > \frac{1+\sqrt{4*2*365\ln 2+1}}{2} \sim 23.0\end{aligned}$$

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 in February 29th.

Q2. We are interested in studying the probability of success of a student at an entrance exam to two departments of a university. Consider the following events

$$\begin{aligned}A &= \{\text{The student is man}\}, \\ A^c &= \{\text{The student is woman}\}, \\ B &= \{\text{The student applied for department I}\}, \\ B^c &= \{\text{The student applied for department II}\}, \\ C &= \{\text{The student was accepted}\}, \\ C^c &= \{\text{The student wasn't accepted}\}.\end{aligned}$$

We assume that we have the following probabilities (Berkeley 1973):

$\mathbb{P}(A) = 0.73,$
$\mathbb{P}(B \mid A) = 0.69, \quad \mathbb{P}(B \mid A^c) = 0.24,$
$\mathbb{P}(C \mid A \cap B) = 0.62, \quad \mathbb{P}(C \mid A^c \cap B) = 0.82, \quad \mathbb{P}(C \mid A \cap B^c) = 0.06, \quad \mathbb{P}(C \mid A^c \cap B^c) = 0.07.$

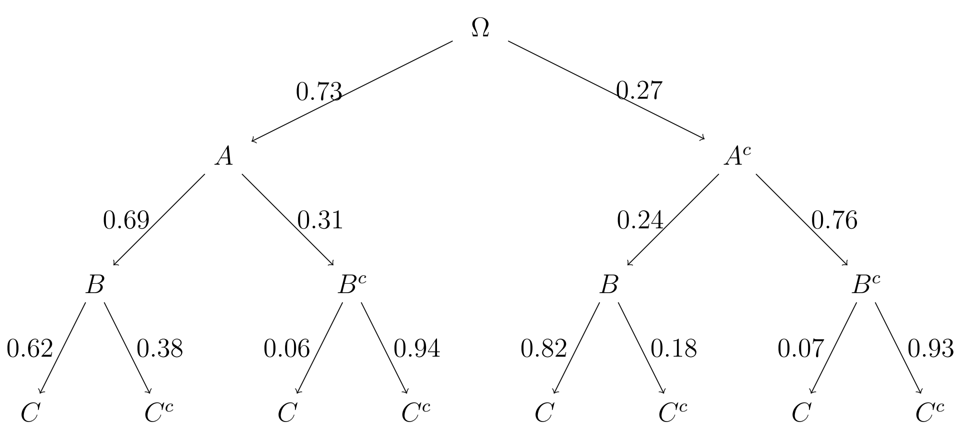
(a) Draw a tree describing the situation with the probabilities associated.

(b) Taking in to consideration the following probabilities $\mathbb{P}[C|A \cap B] = 0.62$, $\mathbb{P}[C|A^c \cap B] = 0.82$, $\mathbb{P}[C|A \cap B^c] = 0.06$, $\mathbb{P}[C|A^c \cap B^c] = 0.07$. With this information, do you think that in this examination women are disadvantaged?

(c) Compute $\mathbb{P}(C \mid A)$ and $\mathbb{P}(C \mid A^c)$. Does this coincide with your answer of b)?

Solution

(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) Just computing

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

and

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

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.

Q3. Posterior probabilies Suppose that a box contains three coins and that for each coin there is a different probability that a head will be obtained when the coin is tossed. Let p_i denote the probability of a head when the ith coin is tossed ($i = 1, \dots, 3$), and suppose that $p_1 = 1/4$, $p_2 = 1/2$, $p_3 = 3/4$.

(a) Suppose that one coin is selected uniformly at random from the box and when it is tossed once, a head is obtained. What is the posterior probability that the ith coin was selected?

(b) If the same coin were tossed again, what would be the probability of obtaining another head?

(c) Prove the Conditional Bayes' Theorem: Let $(A_i)_{i=1\cdots k}$ be a partition of Ω, and B, C are events in Ω,

$$\mathbb{P}(A_i|B \cap C) = \frac{\mathbb{P}(A_i|B)\mathbb{P}(C|A_i \cap B)}{\sum_{j=1}^{k} \mathbb{P}(A_j|B)\mathbb{P}(C|A_j \cap B)}.$$

(d) If the same coin gives another head at the second toss, what is the posterior probability that the ith coin was selected?

(e) Assume that it is always the same coin tossed, and we get always head. What is the recurrence relation of the posterior probability after n tosses that the ith coin was selected?

Solution 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

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

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

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

The formula that we should prove can than be reduced to

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

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 $\mathbf{Q}(\cdot) = \mathbb{P}(\cdot|H_1)$. The prior probabilities are $q_i := \mathbf{Q}(A_i)$ calculated in a). Also by the conditional independence of H_1 and H_2 given A_i, we have that under $\mathbf{Q}$, the probability of getting a head from the ith coin is unchanged:

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

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

$$\mathbb{P}(A_i|H_1 \cap H_2) = \mathbf{Q}(A_i|H_2) = \frac{p_i q_i}{p_1q_1 + p_2q_2 + p_3q_3} = \frac{p_iq_i}{7/12}.$$

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

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

Q4. Introduction to Bayesian Statistics We have m urns with red and white balls inside. The urn $i \in \{1, ..m\}$ has $2i-1$ red balls and $2m-2i+1$ white ones. We randomly select an urn and extract with replacement n times. Define:

$$X_j := \begin{cases} 1 & \text{If the } j\text{-th ball is red,} \\ 0 & \text{If the } j\text{-th ball is white.} \end{cases}$$

We are interested in the following problem " Given that you see $(X_j)_{j=1}^n$, can you say from which urn the balls where taken?"

(a) Compute $\mathbb{P}(X_1 = x_1, ..., X_n = x_n)$ for $x_i \in \{0, 1\}$. Are $X_1, .., X_n$ independent?

(b) Compute the following probability:

$$\mathbb{P}\left(\text{The urn chosen is } i \mid X_1 = x_1, ..., X_n = x_n\right).$$

Show that this only depends on the number or red balls, i.e., $k = \sum_{i=1}^n x_i$.

(c) Compute $\mathbb{P}\left(\text{The urn chosen is } i \mid X_1 = x_1, ..., X_n = x_n\right)$ for $m = 3$ and $n = 3$.

Solution

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

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

We have that X_1 and X_2 are not independent because:

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

This happens because the first variable gives "information" about which urn we have chosen so it also gives information about the second variable.

(b) Just by definition:

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

(c) We just have to compute and get:

	$i=1$	$i=2$	$i=3$
$k=0$	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