Probability

- probability space: $(\Omega, \mathbb{P})$ where Ω countable set of outcomes (sample space), $\mathbb{P}: \mathcal{P}(\Omega) \to [0,1]$ the probability measure satisfying:
  1) $\mathbb{P}(\Omega) = 1$
  2) For $E \subseteq \Omega$, $\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\{\omega\})$
  2*) If $E_1, E_2, \ldots \subseteq \Omega$ w/ $E_i \cap E_j = \emptyset$ for $i \neq j$, then
     $\mathbb{P}(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$
- $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) \quad \mathbb{P}(E \cap F)$
- $E, F \subseteq \Omega$, $\mathbb{P}(F) \neq 0$, $\mathbb{P}(E \mid F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$ Conditional Probability
- E, F independent $\Leftrightarrow \mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$
- E, F mutually exclusive $\Leftrightarrow E \cap F = \emptyset$

- Bayes' Theorem: $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}$

- Law of Total Probability: For partition $B_1, B_2, \ldots, B_n$ of sample space Ω, we have $\mathbb{P}(A) = \mathbb{P}(A \mid B_1)\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\mathbb{P}(B_2) + \ldots + \mathbb{P}(A \mid B_n)\mathbb{P}(B_n)$
  or $\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$

$\hookrightarrow$ Bayes' with Law of Total Probability: $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)}$

- Binomial Distribution: Flip n coins w/ prob p to land heads,
  $\mathbb{P}[k \text{ heads}] = \binom{n}{k} p^k (1-p)^{n-k}$

Problems

1. Email site unrolls new spam-detection tool that correctly identifies a spam email with probability 80%. The probability that it labels a non-spam email as spam is 20%. It is estimated that 40% of all emails are spam. If an email is detected as spam, what is the probability it is actually spam?

Soln:

$\mathbb{P}[\text{spam} \mid \text{detected}]$

$$= \frac{\mathbb{P}[\text{detected} \mid \text{spam}]\,\mathbb{P}[\text{spam}]}{\mathbb{P}[\text{detected}]} \quad \text{by Bayes'}$$

$$= \frac{80\% \cdot 40\%}{\mathbb{P}[\text{detected} \mid \text{spam}]\,\mathbb{P}[\text{spam}] + \mathbb{P}[\text{detected} \mid \text{not spam}]\,\mathbb{P}[\text{not spam}]}$$

$$= \frac{4/5 \cdot 2/5}{4/5 \cdot 2/5 + 1/5 \cdot 3/5}$$

$$= \frac{8}{11}$$

2. You roll a fair 5-sided die 5 times. What is the probability you roll an even number exactly twice? (The sides of the die are numbered 1 through 5).

Soln:

Let $p = \mathbb{P}[\text{even number on one roll}]$
$= 2/5$

Since 2 and 4 are even.

We can treat this as a binomial distribution. For every succesful outcome where we roll an even number exactly twice, we need to choose 2 of 5 rolls to roll an even number. Each such outcome has probability of $(2/5)^2(3/5)^3$ of occuring, since the two even rolls occur with probability $2/5$ and the odd rolls occur with probability $3/5$. Thus, our overall probability is $\binom{5}{2}(2/5)^2(3/5)^3$.

3. You roll a 6-sided die 3 times. What is the probability you roll 3 consecutive numbers in order?

Soln:

$\mathbb{P}[3 \text{ consecutive numbers in order}]$
$= \mathbb{P}[1,2,3] + \mathbb{P}[2,3,4] + \mathbb{P}[3,4,5] + \mathbb{P}[4,5,6]$
$= 4(1/6)^3$

Since we have 4 equally likely outcomes and each one has probability $(1/6)^3$ of occuring.

4. You roll a fair 4-sided die twice. You lose \$1 if your second roll is less than your first. If you lost money, what is the probability you rolled a 2 on your first roll?

Soln:

By Bayes' Theorem:

$$\mathbb{P}[2 \text{ on } 1^{st} \mid \text{lost}] = \frac{\mathbb{P}[\text{lost} \mid 2 \text{ on } 1^{st}]\,\mathbb{P}[2 \text{ on } 1^{st}]}{\mathbb{P}[\text{lost}]}$$

By Law of Total Probability:

$$\begin{aligned}\mathbb{P}[\text{lost}] &= \mathbb{P}[\text{lost} \mid 1 \text{ on } 1^{st}]\,\mathbb{P}[1 \text{ on } 1^{st}] + \mathbb{P}[\text{lost} \mid 2 \text{ on } 1^{st}]\,\mathbb{P}[2 \text{ on } 1^{st}] \\ &\quad + \mathbb{P}[\text{lost} \mid 3 \text{ on } 1^{st}]\,\mathbb{P}[3 \text{ on } 1^{st}] + \mathbb{P}[\text{lost} \mid 4 \text{ on } 1^{st}]\,\mathbb{P}[4 \text{ on } 1^{st}] \\ &= 0 \cdot 1/4 + 1/4 \cdot 1/4 + 2/4 \cdot 1/4 + 3/4 \cdot 1/4 \\ &= 3/8\end{aligned}$$

So, $\mathbb{P}[2 \text{ on } 1^{st} \mid \text{lost}] = \dfrac{1/4 \cdot 1/4}{3/8}$

$= 1/6$