Recall

Independence of events

- E and F are *independent* if $P(EF) = P(E)P(F) \iff P(E|F) = P(E)$ if $P(F) > 0$.
- $\{E_1, \dots, E_n\}$ are *independent* if for **every** $\{i_1, \dots, i_r\} \subset \{1, \dots, n\}$,

$$P(E_{i_1} \cdots E_{i_r}) = P(E_{i_1}) \cdots P(E_{i_r}).$$

- An infinite family of events are *independent* if every finite subset of events from that family are independent.

Example 1 (Pairwise independence $\not\Longrightarrow$ independence)**.** Roll a die twice. Consider the events

$$\begin{aligned} A &= \{ \text{ sum of the two numbers is 7 } \}, \\ B &= \{ \text{ the first number is 3 } \}, \\ C &= \{ \text{ the second number is 4 } \}. \end{aligned}$$

Then

$$P(A) = P(B) = P(C) = \frac{1}{6} \quad \text{and} \quad P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{36}.$$

However,

$$P(A \cap B \cap C) = \frac{1}{36} \quad \text{while} \quad P(A)P(B)P(C) = \frac{1}{216}.$$

Hence the events $\{A, B, C\}$ are pairwise independent but NOT independent.

Remark. It follows from Example 1 that we should check all the 'product' equations in the definition to assure the independence of a finite family of events.

Discrete random variables

A random variable X is a function from the sample space S to the real numbers $\mathbb{R}$. The randomness comes from the probability $P(\cdot)$ on the sample space. If the range of X is a countable set in $\mathbb{R}$, then X is called *discrete* random variable. The *probability mass function* (PMF) of a discrete random variable is defined by

$$p(x) \coloneqq P(X = x).$$

Expectation / expected value / mean.

$$E[X] \coloneqq \sum_{x \colon p(x) > 0} x p(x).$$

Let $g \colon \mathbb{R} \to \mathbb{R}$. Then it is proved that

$$E[g(X)] = \sum_{x \colon p(x) > 0} g(x) p(x).$$

In particular, for any $a, b \in \mathbb{R}$,

$$E[aX + b] = aE[X] + b.$$

Variance.

$$\operatorname{Var}(X) := E[(X - E[X])^2] = E[X^2] - (E[X])^2,$$

And for $a, b \in \mathbb{R}$,

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X).$$

Computing expectations

Example 2. Two balls are randomly chosen from a box containing 8 white balls and 4 black balls. Suppose we win two dollars for each black ball and lose one dollar for each white ball selected. Let random variable X be our winnings. What is X? What are the probabilities associated to each value?

Proof. First, note that the sample space is

$$S = \{(Black, Black), (Black, White), (White, Black), (White, White)\}.$$

We can detect that $X(B, B) = 4$, $X(B, W) = X(W, B) = 2 - 1 = 1$, $X(W, W) = -2$. It means the random variable can only take values from $\{4, 1, -2\}$. Therefore,

$$P(X = k) = \begin{cases} \frac{\binom{4}{2}}{\binom{12}{2}} & \text{if } k = 4, \\ \frac{\binom{4}{1}\binom{8}{1}}{\binom{12}{2}} & \text{if } k = 1, \\ \frac{\binom{8}{2}}{\binom{12}{2}} & \text{if } k = -2. \end{cases}$$

□

Example 3. Randomly choose 3 numbers from $\{1, \ldots, 10\}$. Let X be the smallest number among the 3 chosen numbers. Find $E[X]$.

Solution. First determine the probability mass function

$$p(k) = P(X = k) = \begin{cases} \frac{\binom{10-k}{2}}{\binom{10}{3}} & k = 1, \ldots, 8; \\ 0 & k = 9, 10, \end{cases}$$

where the first case is obtained by choosing the other 2 numbers that are greater than k, and the second case results from the observation that $9, 10$ can never be the smallest numbers among the 3 chosen numbers. Then

$$E[X] = \sum_{k=1}^{10} kp(k) = \sum_{k=1}^{8} k \frac{\binom{10-k}{2}}{\binom{10}{3}} = \frac{11}{4}.$$

□

Example 4. In a game show, there are 3 doors:

$$\boxed{+\$30}, \quad \boxed{-\$10}, \quad \boxed{-\$10}.$$

To start the game, you have to pay \$5 to randomly open a door from these 3 doors which look the same to you. You will get \$30 if opening $\boxed{+\$\ 30}$ and lose \$10 if opening $\boxed{-\$\ 10}$.

Assume you are 'reasonable': if you win \$30, then you quit the game; if you lose \$10, then you flip a coin to decide whether you should continue, i.e., if the coin shows tail then you quit the game; if the coin shows head then you pay another \$5 to randomly open a door from the remaining 2 doors. Assume you are only allowed to play at most 2 rounds.

Let X be the winnings when you leave. Find $E[X]$ and $\mathrm{Var}(X)$.

Solution. First determine the sample space of the door results each of which is represented by a vector

$$S = \{(+30), (-10), (-10, +30), (-10, -10)\}.$$

Let G be the random variable labeling the above outcomes with $1, 2, 3, 4$ from left to right, e.g. the event $\{G = 2\}$ is the outcome (-10). Then the r.v. X is a function of the r.v. G, i.e. $X = f(G)$ for some function $f\colon \mathbb{R} \to \mathbb{R}$ which assigning the winnings to outcomes. Hence

$$\begin{aligned}
E[X] &= E[f(G)] \\
&= \sum_{k=1}^{4} P(G = k) f(k) \\
&= \frac{1}{3} \times (-5 + 30) + \frac{2}{3} \times \frac{1}{2} \times (-5 - 10) + \\
&\quad + \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} \times (-5 - 10 - 5 + 30) + \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} \times (-5 - 10 - 5 - 10) \\
&= 0,
\end{aligned}$$

And since $E[X] = 0$,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = E[(f(G))^2] = \sum_{k=1}^{4} P(G = k)\,(f(k))^2 = 450.$$

□