## 3. Conditional probability and independence

### 3.1. Conditional probability.

**Definition 3.1.** Let $P(F) > 0$. The conditional probability of $E$ given $F$ is defined to be $P(E|F) = P(EF)/P(F)$.

*Remark* 3.1. If $P(s) = 1/|S|$ for all $s \in S$, then $P(E|F) = |EF|/|F|$.

*Example* 3.1. A coin is flipped twice. Suppose all sample points are equally likely to occur. What is the conditional probability that both flips land on heads given (a) the first flip lands on heads? (b) at least one flip lands on heads?

Let $E$ be the event that both flips land on heads, $F$ be the event that the first flip lands on heads and $G$ be the event that at least one flip lands on heads. Then, $E = \{(h,h)\}$, $F = \{(h,h),(h,t)\}$ and $G = \{(h,h),(h,t),(t,h)\}$. This implies

$$P(E|F) = \frac{1/4}{1/2} = \frac{1}{2}, \quad P(E|G) = \frac{1/4}{3/4} = \frac{1}{3}.$$

*Example* 3.2. A bag contains 5 white balls and 6 black balls. In the experiment of randomly withdrawing 3 balls, what are the conditional probability that at least two balls are white given at least one ball is black and the conditional probability that at least one ball is black given at least two balls are white?

Let $E$ be the event that at least two balls are white and $F$ be the event that at least one ball is black. If $W_1, ..., W_5$ denote the white balls and $B_1, ..., B_6$ denote the black balls, then

$$E = \{\{W_i, W_j, B_k\} | i < j\} \cup \{\{W_i, W_j, W_k\} | i < j < k\}, \quad F = S \setminus \{\{W_i, W_j, W_k\} | i < j < k\}.$$

Clearly, $EF = \{\{W_i, W_j, B_k\} | i < j\}$ and, by the combinatorial analysis,

$$|E| = \binom{5}{2}\binom{6}{1} + \binom{5}{3} = 70, \quad |F| = \binom{11}{3} - \binom{5}{3} = 155, \quad |EF| = \binom{5}{2}\binom{6}{1} = 60.$$

This implies $P(E|F) = 60/155 = 12/31$ and $P(F|E) = 60/70 = 6/7$.

**Theorem 3.1** (The multiplication rule). *For events $E_1, ..., E_n$, if $P(E_1E_2\cdots E_{n-1}) > 0$, then*

$$\begin{aligned} P(E_1E_2\cdots E_n) &= P(E_1)\prod_{k=1}^{n-1} P(E_{k+1}|E_1\cdots E_k) \\ &= P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1}). \end{aligned}$$

*Example* 3.3. Recall the match problem that $N$ men randomly select their mixed hats. For $k \le N$, let $E_k$ be the event that there are exactly $k$ men selecting their own hats. Note that $S = E_0 \cup E_1 \cup \cdots \cup E_N$. It has been shown before that $P(E_0) = \sum_{i=0}^{N}(-1)^i/i! \to e^{-1}$ as $N \to \infty$. For $1 \le k \le N$, let $F_k$ be the event that the $k$th man selects his own hat. Then, for $1 \le i_1 < i_2 < \cdots < i_k \le N$,

$$P(F_{i_1}\cdots F_{i_k}) = P(F_{i_1})P(F_{i_2}|F_{i_1})\cdots P(F_{i_k}|F_{i_1}\cdots F_{i_{k-1}})$$

Observe that, for $1 \le l \le N$,

$$P(F_{i_l}|F_{i_1}\cdots F_{i_{l-1}}) = \frac{1}{N-l+1}.$$

By the multiplication rule, this implies

$$P(F_{i_1}\cdots F_{i_k}) = \prod_{l=1}^{k}\frac{1}{N-l+1} = \frac{(N-k)!}{N!}$$

and, hence, for $\{j_1, ..., j_{N-k}\} = \{1, ..., N\} \setminus \{i_1, ..., i_k\}$,

$$P(E_k F_{i_1} \cdots F_{i_k}) = P(F_{i_1} \cdots F_{i_k} F_{j_1}^c \cdots F_{j_{N-k}}^c)$$

$$= P(F_{i_1} \cdots F_{i_k}) P(F_{j_1}^c \cdots F_{j_{N-k}}^c | F_{i_1} \cdots F_{i_k}) = \frac{(N-k)!}{N!} \times \sum_{i=0}^{N-k} \frac{(-1)^i}{i!}.$$

Consequently, we obtain

$$P(E_k) = \sum_{1 \le i_1 < \cdots < i_k \le N} P(E_k F_{i_1} \cdots F_{i_k}) = \frac{1}{k!} \sum_{i=0}^{N-k} \frac{(-1)^i}{i!} \to \frac{e^{-1}}{k!}$$

as $N \to \infty$.

*Example* 3.4. Consider the following molecule dynamics. Suppose that there are two urns, where urn 1 contains $n$ red balls and urn 2 contains $n$ blue balls. Each time, a ball in urn 1 is removed randomly and a ball in urn 2 is removed to urn 1. When urn 2 is empty, balls in urn 1 keep being removed until urn 1 is empty. Find the probability that the last ball to be removed is red. Let $E$ be the desired event. Let $r_1, ..., r_n$ be the red balls and let $F_i$ be the event that ball $r_i$ is the last to be removed. It is clear that $F_i F_j = \emptyset$ if $i \neq j$ and $E = F_1 \cup \cdots \cup F_n$. Note that it does not change the probability of $F_1$ if $r_1$ is re-indexed by $r_i$ and vice versa. This implies $P(F_i) = P(F_1)$ for all $i$ and hence $P(E) = nP(F_1)$. To see the probability of $F_1$, let $G_k$ be the event that ball $r_1$ remains in urn 1 after $k$ withdraws. Clearly, $F_1 = G_1 \cdots G_{2n-1}$ and

$$P(F_1) = P(G_1) P(G_2|G_1) \cdots P(G_{2n-1}|G_1 \cdots G_{2n-2}).$$

Observe that, for $1 \le k \le n$,

$$P(G_{k+1}|G_1 \cdots G_k) = 1 - \frac{1}{n}$$

and, for $n < k < 2n-1$,

$$P(G_{k+1}|G_1 \cdots G_k) = \frac{2n-k-1}{2n-k}.$$

As a result, we achieve

$$P(E) = \left(1 - \frac{1}{n}\right)^n \to e^{-1} \quad \text{as} \quad n \to \infty.$$

3.2. **Bayes's formula.** Let $F$ be an event satisfying $P(F) \in (0,1)$. Then, for any event $E$,

$$P(E) = P(EF) + P(EF^c) = P(F)P(E|F) + P(F^c)P(E|F^c).$$

The following theorem is a generalization of the above computations.

**Theorem 3.2.** *Let $F_1, F_2, ..., F_n$ be events satisfying $P(F_i F_j) = 0$ for $i \neq j$ and $\sum_{i=1}^n P(F_i) = 1$. Suppose $P(F_i) > 0$ for all $1 \le i \le n$. Then, for any event $E$, $P(E) = \sum_{i=1}^n P(F_i) P(E|F_i)$.*

*Remark* 3.2. Note that if $F_1, ..., F_n$ are mutually disjoint, then $P(F_i F_j) = 0$ for $i \neq j$. If $\bigcup_{i=1}^n F_i = S$ is further assumed, then $\sum_{i=1}^n P(F_i) = 1$.

**Corollary 3.3** (Bayes's formula)**.** *Refer to the assumption of Theorem 3.2 and assume that $P(E) > 0$ and $P(F) > 0$. Then,*

$$P(E|F) = \frac{P(E)P(F|E)}{\sum_{i=1}^n P(F_i) P(F|F_i)}.$$

*In particular, for $1 \le j \le n$,*

$$P(F_j|F) = \frac{P(F_j)P(F|F_j)}{\sum_{i=1}^n P(F_i) P(F|F_i)}.$$

*Remark* 3.3. Theorem 3.2 and Corollary 3.3 also hold for any infinite sequence $(F_i)_{i=1}^{\infty}$.

*Example* 3.5. A blood test is 95% effective in detecting a certain disease when it is present. But, the test yields a 1% false positive result of the uninfected. Suppose that $x$% population is infected. What is the probability that a person has the disease given the test result is positive? What is the probability that the test fails?

Let $E$ be the event that a person is infected and $F$ be the event that the test result is positive. In this setting, we have

$$P(F|E) = 0.95, \quad P(F|E^c) = 0.01, \quad P(E) = \frac{x}{100}.$$

Based on this setting, both questions are to find $P(E|F)$ and $P((EF^c) \cup (E^cF))$ respectively. By Bayes's formula, we have

$$P(E|F) = \frac{P(E)P(F|E)}{P(E)P(F|E) + P(E^c)P(F|E^c)} = \frac{0.95x}{0.95x + 0.01(100-x)} = \frac{95}{94 + 100/x}.$$

For the second one, note that $P(E^cF) = P(E^c)P(F|E^c) = (100-x)/10000$ and

$$P(EF^c) = P(E) - P(EF) = P(E) - P(E)P(F|E) = \frac{x}{2000}.$$

This implies $P(\{\text{the test fails}\}) = (25+x)/2500$.

*Example* 3.6. An urn contains two different coins, $A$ and $B$, where $A$ has probability $a$ landing on heads and $B$ has probability $b$ landing on heads. If a cion is randomly selected and flipped, what is the probability that the flipped coin is $A$ given the flip lands on heads.

Let $E$ be the event that the flipping coin is $A$ and $F$ be the event that the flipped coin lands on heads. Following the assumption, we have

$$P(F|E) = a, \quad P(F|E^c) = b, \quad P(E) = 1/2.$$

This implies

$$P(E|F) = \frac{P(E)P(F|E)}{P(E)P(F|E) + P(E^c)P(F|E^c)} = \frac{a}{a+b}.$$

3.3. **Independent events.** Intuitively, two events are said to be independent if whether one event happens won't affect the probability of the appearance of the other. Mathematically, this means that

$$P(E|F) = P(E|F^c) = P(E), \quad P(E^c|F) = P(E^c|F^c) = P(E^c)$$

and

$$P(F|E) = P(F|E^c) = P(F), \quad P(F^c|E) = P(F^c|E^c) = P(F^c).$$

Formally, it is an easy exercise to show that all above equations are equivalent to

$$P(EF) = P(E)P(F).$$

**Definition 3.2.** Two events $E$ and $F$ are said to be **independent** if $P(EF) = P(E)P(F)$. $E$ and $F$ are said to be **dependent** if they are not independent.

**Lemma 3.4.** *Let $E$ be either the empty set or the sample space. Then, for any event $F$, $E$ and $F$ are independent.*

*Proof.* The proof is clear from the fact that $P(F\emptyset) = 0 = P(F)P(\emptyset)$ and $P(FS) = P(F) = P(F)P(S)$. □

*Remark* 3.4. In fact, if $P(E) \in \{0,1\}$, then $E$ is independent of any event.

**Proposition 3.5.** *If $E$ and $F$ are independent, then $E$ and $F^c$ are independent.*

*Proof.* According to the definition of independence of $E$ and $F$, one has $P(EF) = P(E)P(F)$. This implies

$$P(EF^c) = P(E) - P(EF) = P(E) - P(E)P(F) = P(E)[1 - P(F)] = P(E)P(F^c).$$

□

*Remark* 3.5. If $E$ is independent of itself, then $P(E) = P(E)^2$, which implies $P(E) = 0$ or $P(E) = 1$. Equivalently, if $P(E) \in (0, 1)$, then $E$ can not be independent of itself.

*Example* 3.7. Consider an urn with $n$ balls indexed by 1,2,...,n. An experiment is performed by randomly selecting three balls from the urn. Let $E_i$ be the event that one ball has number $i$. Then, $S = \{\{k, l, m\} | k \neq l, l \neq m, k \neq m\}$, $E_i = \{\{i, k, l\} \in S | k \neq l\}$ and, for $i \neq j$, $E_iE_j = \{\{i, j, k\} \in S | k \notin \{i, j\}\}$. This implies $|E_iE_j| = n - 2$ and $|E_i| = \binom{n-1}{2} = (n-1)(n-2)/2$ and $|S| = \binom{n}{3} = n(n-1)(n-2)/6$. Then, for $i \neq j$,

$$P(E_iE_j) = \frac{n-2}{n(n-1)(n-2)/6} = \frac{6}{n(n-1)}, \quad P(E_i) = \frac{(n-1)(n-2)/2}{n(n-1)(n-2)/6} = \frac{3}{n}.$$

This means that $E_i$ and $E_j$ are dependent if $n \geq 4$.

**Definition 3.3.** Events $E_1, ..., E_n$ are independent if

$$P(F_1F_2 \cdots F_n) = P(F_1)P(F_2) \cdots P(F_n),$$

where $F_i \in \{E_i, E_i^c\}$ for $1 \leq i \leq n$. Events $E_1, E_2, ...$ are independent if any finite number of events are independent.

**Lemma 3.6.** *$E_1, E_2, ..., E_n$ are independent if and only if, for any finite set $\{i_1, ..., i_k\} \subset \{1, ..., n\}$,*

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

*Example* 3.8. Consider an infinite sequence of independent trials. At each trial, there are only two outcomes, failure (0) and success (1), and the probability of success at each trial is equal to $p \in (0, 1)$. Fix $n > 0$, let $E_k$ be the event that there are exactly $k$ successes in the first $n$ trial and let $F$ be the event that all trials are successes. Find the probabilities of $E_k$ and $F$.

Note that $E_k = \bigcup_{1 \leq i_1 < \cdots < i_k \leq n} F_{i_1,...,i_k}$, where $F_{i_1,...,i_k}$ denotes the event that, in the first $n$ trials, there are exactly $k$ successes and they are trials $i_1, ..., i_k$. To compute the probability of $F_{i_1,...,i_k}$, let $G_i$ be the event that the $i$th trial is success. By setting $\{j_1, ..., j_{n-k}\} = \{1, ..., n\} \setminus \{i_1, ..., i_k\}$, $F_{i_1,...,i_k}$ can be rewritten as

$$F_{i_1,...,i_k} = G_{i_1}G_{i_2} \cdots G_{i_k}G_{j_1}^cG_{j_2}^c \cdots G_{j_{n-k}}^c.$$

Since all trials are independent, we have

$$P(F_{i_1,...,i_k}) = P(G_{i_1}) \cdots P(G_{i_k})P(G_{j_1}^c) \cdots P(G_{j_{n-k}}^c) = p^k(1-p)^{n-k}.$$

Hence, $P(E_k) = \binom{n}{k}p^k(1-p)^{n-k}$ for $0 \leq k \leq n$. It is worthwhile to remark the binomial theorem since $1 = \sum_{k=0}^n P(E_k) = \sum_{k=0}^n \binom{n}{k}p^k(1-p)^{n-k}$.

To see the probability of $F$, note that $F \subset E_n$. This implies $P(F) \leq p^n$ for all $n > 0$. Letting $n \to \infty$ implies $P(F) = 0$.

*Example* 3.9. Refer to the trials of the previous example. Let $A_{n,m}$ be the event that $n$ successes occur before $m$ failures and $G_1$ be the event that the first trial is success. Set $P_{n,m} = P(A_{n,m})$. By conditioning on the result of the first trial, we have, for $n \geq 1$ and $m \geq 1$,

$$\begin{aligned} P_{n,m} &= P(A_{n,m}|G_1)P(G_1) + P(A_{n,m}|G_1^c)P(G_1^c) \\ &= pP(A_{n-1,m}) + (1-p)P(A_{n,m-1}) = pP_{n-1,m} + (1-p)P_{n,m-1}. \end{aligned}$$

The boundary condition for the above iterative formula contains $P_{0,m} = 1$ for $m \geq 1$ and $P_{n,0} = 0$ for $n \geq 1$.

In fact, Fermat pointed out

$$A_{n,m} = \{\text{there are at least } n \text{ successes in the first } n+m-1 \text{ trials}\},$$

which implies

$$P_{n,m} = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k (1-p)^{n+m-1-k}.$$

*Example* 3.10 (The gambler's ruin problem). Let $A$ and $B$ be two gamblers. They bet on the result of flipping a coin. If the coin lands on heads, then $A$ wins and $B$ gives $A$ one dollar. If the coin lands on tails, then $B$ wins and $A$ gives $B$ one dollar. Suppose that the flips of coins are independent and the probability of heads is always equal to $p$. The game ends if $A$ or $B$ runs out of money. Find the probability that $A$ wins all the money if, in the beginning, $A$ has $m$ dollars and $B$ has $n$ dollars.

Observe that the total money in this game is $N := m+n$ dollars. Let $E$ be the event that $A$ wins all the money if $A$ has $i$ dollars and $B$ has $N-i$ dollars in the beginning. Set $P_i = P(E)$ and let $F$ be the event that the first flip lands on heads. Then, conditioning on the result of the first flip, we have

$$P_i = P(E|F)P(F) + P(E|F^c)P(F^c) = pP_{i+1} + qP_{i-1} \quad \forall 1 \leq i \leq N-1,$$

where $q = 1-p$. The above identity can be rewritten as

$$p(P_{i+1} - P_i) = q(P_i - P_{i-1}) \quad \forall 1 \leq i \leq N-1.$$

This is equivalent to

$$P_{i+1} - P_i = (q/p)^i (P_1 - P_0) \quad \forall 0 \leq i \leq N-1$$

Summing up $i$ yields, for $0 \leq j \leq N$,

$$P_j - P_0 = \sum_{i=0}^{j-1} (P_{i+1} - P_i) = \begin{cases} \frac{(q/p)^j - 1}{q/p - 1}(P_1 - P_0) & \text{if } p \neq q \\ j(P_1 - P_0) & \text{if } p = q \end{cases}.$$

Using the boundary condition $P_0 = 0$, $P_N = 1$, one has

$$1 = P_N - P_0 = \begin{cases} \frac{(q/p)^N - 1}{q/p - 1} P_1 & \text{if } p \neq q \\ NP_1 & \text{if } p = q \end{cases}.$$

As a result, if $p = q$, $P_j = j/N$; if $p \neq q$, $P_j = \frac{(q/p)^j - 1}{(q/p)^N - 1}$. As $A$ starts with $m$ dollars, we obtain

$$P(A \text{ wins all the money}) = \begin{cases} \frac{(q/p)^m - 1}{(q/p)^{n+m} - 1} & \text{if } p \neq q \\ \frac{m}{n+m} & \text{if } p = q \end{cases}.$$

By the symmetry of the game, if $Q_j$ denotes the probability that $B$ wins all the money starting with $N-j$ dollars, then

$$Q_j = \begin{cases} \frac{(p/q)^{N-j} - 1}{(p/q)^N - 1} & \text{if } p \neq q \\ \frac{N-j}{N} & \text{if } p = q \end{cases}$$

Obviously, one may conclude $P_j + Q_j = 1$ for all $0 \leq j \leq N$. This means that the game will end eventually.

3.4. $P(\cdot|F)$ **is a probability.**

**Proposition 3.7.** *Let $F$ be an event with positive probability and set $\overline{P}(E) = P(E|F)$. Then, $\overline{P}$ is a probability.*

*Proof.* Clearly, $\overline{P}(E) = P(E|F) \in [0,1]$ and $\overline{P}(S) = P(S|F) = 1$. For mutually disjoint events $E_1, E_2, ...$, it is obvious that $E_1F, E_2F, ...$ are also mutually disjoint. This implies

$$\overline{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \frac{1}{P(F)} P\left(\bigcup_{n=1}^{\infty} E_nF\right) = \frac{1}{P(F)} \sum_{n=1}^{\infty} P(E_nF) = \sum_{n=1}^{\infty} \overline{P}(E_n).$$

□

**Definition 3.4.** Given $F$, events $E_1$ and $E_2$ are **conditionally independent** if $P(E_1E_2|F) = P(E_1|F)P(E_2|F)$.

*Example* 3.11. Roll a dice twice and assume that all outcomes are equally likely to happen, that is, $S = \{(i,j)|1 \le i,j \le 6\}$ and $P(i,j) = 1/36$ for all $(i,j) \in S$. Let $F$ be the event that the first roll results in 6, $E_1$ be the event that the sum of two rolls is at least 10, and $E_2$ be the event that the sum of two rolls is divided by 3. Precisely, one has

$$F = \{(6,j)|1 \le j \le 6\}, \quad E_1 = \{(i,j)|i+j \ge 10\}, \quad E_2 = \{(i,j)|3|(i+j)\}.$$

Note that

$$E_1F = \{(6,4),(6,5),(6,6)\}, \quad E_2F = \{(6,3),(6,6)\}, \quad E_1E_2F = \{(6,6)\}.$$

This implies $P(E_1E_2|F) = 1/6$ and $P(E_1|F) = 1/2$ and $P(E_2|F) = 1/3$. Thus, given $F$, $E_1$ and $E_2$ are conditionally independent. As $P(E_1) = 1/6$, $P(E_2) = 1/3$ and $P(E_1E_2) = 1/36$, it is clear that $E_1$ and $E_2$ are dependent.

**Proposition 3.8.** *Let $E_1, E_2, F$ be events satisfying $P(E_2F) > 0$. Then, given $F$, $E_1$ and $E_2$ are conditionally independent if and only if $P(E_1|E_2F) = P(E_1|F)$.*

*Proof.* Note that $P(E_2F) > 0$ implies $P(F) > 0$. This proposition is clear from the observation that $E_1$ and $E_2$ are conditionally independent given $F$ if and only if

$$\frac{P(E_1E_2F)}{P(F)} = P(E_1|F)\frac{P(E_2F)}{P(F)},$$

or equivalently,

$$P(E_1|E_2F) = \frac{P(E_1E_2F)}{P(E_2F)} = \frac{P(E_1F)}{P(F)} = P(E_1|F).$$

□

*Remark* 3.6. Refer to the setting in Proposition 3.8 and set $\overline{P}(\cdot) = P(\cdot|F)$. Then, given $F$, $E_1$ and $E_1$ are conditionally independent if and only if $\overline{P}(E_2|E_1) = \overline{P}(E_2)$.