Unions, Intersections, Independence, Conditioning and Bayes' Formula
OPRE 7310 Lecture Notes by Metin Çakanyıldırım
Compiled at 00:50 on Friday 28th August, 2020

1 Unions and Intersections

In a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, interpretation of the events as sets allows us to talk about the intersection and union of the events. Intersection and unions are useful to assess the probability of two events occurring together and the probability of at least one of the two events.

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\} \quad \text{and} \quad A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$$

Since $A \cap B, A \cup B \in \mathcal{F}$ for $A, B \in \mathcal{F}$, we can talk about $\mathrm{P}(A \cap B)$ and $\mathrm{P}(A \cup B)$. The union probability can be related to the intersection probability as

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B).$$

This equality is often used to compute $\mathrm{P}(A \cup B)$ when the remaining three probabilities are known. One way to end up with the equality is to consider an outcome $\omega \in A \cap B$ measured twice in $\mathrm{P}(A) + \mathrm{P}(B)$, to avoid this double counting, we set $\mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cup B)$. Another way to obtain this equality is to consider the partition of $A \cup B$ as $A \cap B^c$, $A \cap B$ and $B \cap A^c$ so that countable additivity of P applies to obtain the first equality below:

$$\mathrm{P}(A \cup B) = \mathrm{P}(A \cap B^c) + \mathrm{P}(A \cap B) + \mathrm{P}(B \cap A^c) = \mathrm{P}(A) - \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B) + \mathrm{P}(B) - \mathrm{P}(A \cap B).$$

The second equality also follows from applying the countable additivity on the partition $A \cap B^c$ and $A \cap B$ of A as well as on the partition $B \cap A^c$ and $A \cap B$ of B.

The equality can also be written as $\mathrm{P}(A) + \mathrm{P}(B) = \mathrm{P}(A \cup B) + \mathrm{P}(A \cap B)$, which can be split as $\mathrm{P}(A) + \mathrm{P}(B) \geq \mathrm{P}(A \cup B) + \mathrm{P}(A \cap B)$ and $\mathrm{P}(A) + \mathrm{P}(B) \leq \mathrm{P}(A \cup B) + \mathrm{P}(A \cap B)$. The former (resp. latter) inequality is the condition for the submodularity (resp. supermodularity) when P is a set function defined over a discrete space Ω. Submodular set functions are used in discrete optimization and cooperative game theory. For our purposes, it is necessary that the probability measure is both submodular and supermodular so it is modular.

The formula for computing the probability of union of finite number sets can be obtained by induction.

$$\mathrm{P}(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{1 \leq i \leq n} \mathrm{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathrm{P}(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} \mathrm{P}(A_i \cap A_j \cap A_k) \dots$$
$$\dots + (-1)^{n+1} \mathrm{P}(A_1 \cap A_2 \cap \dots \cap A_n)$$

Note that $\mathrm{P}(A_i)$s are included, $\mathrm{P}(A_i \cap A_j)$s are excluded, $\mathrm{P}(A_i \cap A_j \cap A_k)$s are included and so on. Hence, the formula is called the inclusion-exclusion identity.

Example: Use the inclusion-exclusion identity to obtain $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$. We can proceed as follows $1 = \mathrm{P}(\Omega) = \mathrm{P}(A \cup A^c) = \mathrm{P}(A) + \mathrm{P}(A^c) + \mathrm{P}(A \cap A^c) = \mathrm{P}(A) + \mathrm{P}(A^c)$ to obtain $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$. $\diamond$

Example: Use the inclusion-exclusion identity to obtain $A \subseteq B$ implies $\mathrm{P}(A) \leq \mathrm{P}(B)$. We can proceed as follows $\mathrm{P}(B) = \mathrm{P}(A \cup (B \setminus A)) = \mathrm{P}(A) + \mathrm{P}(B \setminus A) \geq \mathrm{P}(A)$. $\diamond$

Example: The inclusion-exclusion identity can be proved by induction. For $n = 1$ and $n = 2$ the identity holds. So let us suppose that it holds for n:

$$\begin{aligned} \mathrm{P}(A_1 \cup A_2 \cup \cdots \cup A_n) &= \sum_{1\le i\le n} \mathrm{P}(A_i) - \sum_{1\le i<j\le n} \mathrm{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n} \mathrm{P}(A_i \cap A_j \cap A_k) \ldots \\ &\quad \cdots + (-1)^n \sum_{1\le j\le n} \mathrm{P}(\cap_{i=1\ldots n, i\ne j} A_i) + (-1)^{n+1}\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_n), \end{aligned}$$

where we write the next to last term explicitly. Then

$$\begin{aligned} & \mathrm{P}(A_1 \cup A_2 \cup \cdots \cup A_n \cup A_{n+1}) \\ = & \mathrm{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \\ & + \mathrm{P}(A_{n+1}) - \Big\{ \mathrm{P}((A_1 \cap A_{n+1}) \cup (A_2 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1})) \Big\} \\ = & \sum_{1\le i\le n} \mathrm{P}(A_i) - \sum_{1\le i<j\le n} \mathrm{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n} \mathrm{P}(A_i \cap A_j \cap A_k) \cdots + (-1)^{n+1}\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_n) \\ & + \mathrm{P}(A_{n+1}) - \Big\{ \sum_{1\le i\le n} \mathrm{P}(A_i \cap A_{n+1}) - \sum_{1\le i<j\le n} \mathrm{P}(A_i \cap A_j \cap A_{n+1}) \cdots + (-1)^n \sum_{1\le j\le n} \mathrm{P}(\cap_{i=1\ldots n, i\ne j} A_i \cap A_{n+1}) \\ & \qquad + (-1)^{n+1}\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_n \cap A_{n+1}) \Big\} \\ = & \sum_{1\le i\le n+1} \mathrm{P}(A_i) - \sum_{1\le i<j\le n+1} \mathrm{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n+1} \mathrm{P}(A_i \cap A_j \cap A_k) \cdots + (-1)^{n+1} \sum_{1\le j\le n+1} \mathrm{P}(\cap_{i=1\ldots n+1, i\ne j} A_i) \\ & \qquad + (-1)^{n+2}\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_n \cap A_{n=1}). \quad \diamond \end{aligned}$$

The inclusion-exclusion identity holds not only for a probability measure but also for a counting (cardinality of a set) measure:

$$\begin{aligned} |A_1 \cup A_2 \cup \cdots \cup A_n| &= \sum_{1\le i\le n} |A_i| - \sum_{1\le i<j\le n} |A_i \cap A_j| + \sum_{1\le i<j<k\le n} |A_i \cap A_j \cap A_k| \ldots \\ &\quad \cdots + (-1)^n \sum_{1\le j\le n} |\cap_{i=1\ldots n, i\ne j} A_i| + (-1)^{n+1}|A_1 \cap A_2 \cap \cdots \cap A_n|. \end{aligned}$$

Example: Continuity of probability measure P. i) Let A_n be an increasing sequence of events, i.e., $A_n \subseteq A_{n+1}$, then

$$\mathrm{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathrm{P}(A_n).$$

ii) Above limit also holds when A_n is a decreasing sequence of events, i.e., $A_{n+1} \subseteq A_n$. The proof of i) uses countable additivity of P, which applies to disjoint events. Let $B_1 = A_1$ and $B_n = A_n \cap (\cup_{i=1}^{n-1} A_i^c)$, so B_n consists of outcomes in A_n but not in $A_1, \ldots, A_{n-1}$. Note that B_n are disjoint while A_n do not have to be so. Furthermore, $\cup_{i=1}^n A_i = \cup_{i=1}^n B_i$ as well as $\cup_{i=1}^\infty A_i = \cup_{i=1}^\infty B_i$.

$$\begin{aligned} \mathrm{P}(\lim_{n\to\infty} A_n) &= \mathrm{P}(\cup_{i=1}^\infty A_i) = \mathrm{P}(\cup_{i=1}^\infty B_i) = \sum_{i=1}^\infty \mathrm{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^n \mathrm{P}(B_i) = \lim_{n\to\infty} \mathrm{P}(\cup_{i=1}^n B_i) = \lim_{n\to\infty} \mathrm{P}(\cup_{i=1}^n A_i) \\ &= \lim_{n\to\infty} \mathrm{P}(A_n) \end{aligned}$$

For ii), A_n^c is an increasing sequence so i) yields $\mathrm{P}(\lim_{n\to\infty} A_n^c) = \lim_{n\to\infty} \mathrm{P}(A_n^c)$. This further provides $1 - \mathrm{P}(\lim_{n\to\infty} A_n^c) = 1 - \lim_{n\to\infty} \mathrm{P}(A_n^c) = 1 - \lim_{n\to\infty}(1 - \mathrm{P}(A_n)) = \lim_{n\to\infty} \mathrm{P}(A_n)$. Then $\mathrm{P}(\lim_{n\to\infty} A_n) = \mathrm{P}(\cap_{i=1}^\infty A_i) = 1 - \mathrm{P}(\cup_{i=1}^\infty A_i^c) = 1 - \mathrm{P}(\lim_{n\to\infty} A_n^c) = \lim_{n\to\infty} \mathrm{P}(A_n)$. $\diamond$

Other set operations can be represented in terms of unions and intersections, so above formulas can be helpful. Set operations can be defined without any reference to a probability measure. Next we see a property of probability measures.

2 Independence

The section above related $P(A \cup B)$ to $P(A)$, $P(B)$ and $P(A \cap B)$ through the inclusion-exclusion identity. When $P(A \cup B)$ is unavailable, we search for an equality that would relate $P(A \cap B)$ to only $P(A)$ and $P(B)$. Such an equality is possible only when the events are *independent*: Two events A and B are called independent if and only if $P(A \cap B) = P(A)P(B)$. When events A, B are independent, the probability of both happening can be computed by saying the event A happen first with $P(A)$ and the event B happens afterwards with $P(B)$.

Example: In repeated experiments, each experiment is often independent. Such repeated experiments include dice rolls, coin tosses, picking a number from $\{0, 1, 2, \ldots, 9\}$ with repetition. $\diamond$

Example: If A and B are independent, are A^c and B^c independent? Yes because

$$\begin{aligned} P(A^c \cap B^c) &\overset{1}{=} P((A \cup B)^c) \overset{2}{=} 1 - P(A \cup B) \overset{3}{=} 1 - P(A) - P(B) + P(A \cap B) \\ &\overset{4}{=} 1 - P(A) - P(B) + P(A)P(B) \overset{5}{=} (1 - P(A))(1 - P(B)) \\ &\overset{6}{=} P(A^c)P(B^c) \end{aligned}$$

Equality 1 is from $A^c \cap B^c = (A \cup B)^c$. Equality 2 holds because $(A \cup B)^c$ and $A \cup B$ are complements. Equality 3 is from the inclusion-exclusion identity. Equality 4 is from independence. Equality 5 is an algebraic step. Equality 6 holds because A and A^c are complements as well as B and B^c. $\diamond$

The definition of independence can be extended to multiple events. Events $A_1, A_2, \ldots A_n$ are independent if for each subset $\{i_1, i_2, \ldots, i_k\}$ of $\{1, 2, \ldots, n\}$

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k}).$$

Simply put, the equality must be checked for any combination of events. These combinations can include at least 2 events and at most n events, but they do not include $\emptyset$ or the combinations with 1 event. Hence, the number of equalities to check is $2^n - n - 1$.

Example: An urn contains 3 White and 2 Black balls. 3 balls are drawn without replacement one after another. Let A_i be the event that ball i is White for $i = 1, 2, 3$. Are A_1, A_2, A_3 independent? This experiment creates sequences of ball colors of the form WWW, WWB, etc. Since 3 balls are drawn and each ball can potentially take 2 colors, the sample space has 2^3 elements. Table 1 shows all the outcomes and their probabilities. Note that BBB cannot occur so it has a probability of 0.

Table 1: Each elementary outcome ω and its probability without replacement.

ω with at least 2 WW	$P(\omega)$	ω with at most 1 W	$P(\omega)$
WWW	(3/5)(2/4)(1/3)=0.1	WBB	(3/5)(2/4)(1/3)=0.1
WWB	(3/5)(2/4)(2/3)=0.2	BWB	(2/5)(3/4)(1/3)=0.1
WBW	(3/5)(2/4)(2/3)=0.2	BBW	(2/5)(1/4)(3/3)=0.1
BWW	(2/5)(3/4)(2/3)=0.2	BBB	(2/5)(1/4)(0/3)=0.0

$P(A_1) = P(WWW \text{ or } WWB \text{ or } WBW \text{ or } WBB) = 0.6$, $P(A_2) = P(WWW \text{ or } WWB \text{ or } BWW \text{ or } BWB) = 0.6$ and $P(A_3) = P(WWW \text{ or } WBW \text{ or } BWW \text{ or } BBW) = 0.6$. $P(A_1 \cap A_2) = P(WWW \text{ or } WWB) = 0.3$, $P(A_1 \cap A_3) = P(WWW \text{ or } WBW) = 0.3$, $P(A_2 \cap A_3) = P(WWW \text{ or } BWW) = 0.3$. $P(A_1 \cap A_2 \cap A_3) = P(WWW) = 0.1$. To establish the independence, we check 4=$2^3 - 3 - 1$ equalities: $P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3)$, $P(A_1 \cap A_2) \neq P(A_1)P(A_2)$, $P(A_1 \cap A_3) \neq P(A_1)P(A_3)$, $P(A_2 \cap A_3) \neq P(A_2)P(A_3)$. So A_1, A_2, A_3 are not independent.

Table 2: Each elementary outcome ω and its probability with replacement.

ω with at least 2 WW	$\mathrm{P}(\omega)$	ω with at most 1 W	$\mathrm{P}(\omega)$
WWW	(3/5)(3/5)(3/5)=27/125	WBB	(3/5)(2/5)(2/5)=12/125
WWB	(3/5)(3/5)(2/5)=18/125	BWB	(2/5)(3/5)(2/5)=12/125
WBW	(3/5)(2/5)(3/5)=18/125	BBW	(2/5)(2/5)(3/5)=12/125
BWW	(2/5)(3/5)(3/5)=18/125	BBB	(2/5)(2/5)(2/5)= 8/125

We consider the above experiment with replacement of the drawn balls. Let B_i be the event that ball i is White for $i = 1, 2, 3$. Are B_1, B_2, B_3 independent? Table 2 shows the updated probabilities.

$\mathrm{P}(B_1) = \mathrm{P}(WWW \text{ or } WWB \text{ or } WBW \text{ or } WBB) = (27 + 18 + 18 + 12)/125 = 3/5$, $\mathrm{P}(B_2) = \mathrm{P}(WWW \text{ or } WWB \text{ or } BWW \text{ or } BWB) = 3/5$ and $\mathrm{P}(B_3) = \mathrm{P}(WWW \text{ or } WBW \text{ or } BWW \text{ or } BBW) = 3/5$. $\mathrm{P}(B_1 \cap B_2) = \mathrm{P}(B_1 \cap B_3) = \mathrm{P}(B_2 \cap B_3) = (27 + 18)/125 = 9/25$ and $\mathrm{P}(B_1 \cap B_2 \cap B_3) = 27/125$. To establish the independence, we check $\mathrm{P}(B_1 \cap B_2 \cap B_3) = 27/125 = \mathrm{P}(B_1)\mathrm{P}(B_2)\mathrm{P}(B_3)$, $\mathrm{P}(B_1 \cap B_2) = 9/25 = \mathrm{P}(B_1)\mathrm{P}(B_2)$, $\mathrm{P}(B_1 \cap B_3) = 9/25 = \mathrm{P}(B_1)\mathrm{P}(B_3)$, $\mathrm{P}(B_2 \cap B_3) = 9/25 = \mathrm{P}(A_2)\mathrm{P}(A_3)$. So B_1, B_2, B_3 are independent. $\diamond$

When there are many events $A_1, A_2, \dots, A_N$, pairwise independence of two events A_i, A_j for $1 \le i < j \le N$ is established by checking $\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i)\mathrm{P}(A_j)$. Pairwise independence of every pair of events does not imply the independence of events, the following counterexample from the textbook illustrates this.

Example: An urn contains 4 balls numbered as 1,2,3,4, and a ball is drawn randomly. Let A_2 be the event that the drawn ball is either 1 or 2, so $A_2 = \{1, 2\}$. Similarly define $A_3 = \{1, 3\}$ and $A_4 = \{1, 4\}$. We can check that $\mathrm{P}(A_2 \cap A_3) = 1/4 = \mathrm{P}(A_2)\mathrm{P}(A_3)$, $\mathrm{P}(A_2 \cap A_4) = 1/4 = \mathrm{P}(A_2)\mathrm{P}(A_4)$, and $\mathrm{P}(A_3 \cap A_4) = 1/4 = \mathrm{P}(A_3)\mathrm{P}(A_4)$. However, $\mathrm{P}(A_2 \cap A_3 \cap A_4) = 1/4 \neq 1/8 = \mathrm{P}(A_2)\mathrm{P}(A_3)\mathrm{P}(A_4)$. $\diamond$

Example: When A, B are independent, $|\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)| = 0$. This absolute value cannot be larger than $1/4$ even for dependent events. Let us prove this claim. We first insert $\mathrm{P}(A \cap B) = (\mathrm{P}(A) + \mathrm{P}(A^c))\mathrm{P}(A \cap B)$ and $\mathrm{P}(A)\mathrm{P}(B) = \mathrm{P}(A)(\mathrm{P}(A \cap B) + \mathrm{P}(A^c \cap B))$ into $\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)$ to obtain

$$\begin{aligned}
\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B) &= (\mathrm{P}(A) + \mathrm{P}(A^c))\mathrm{P}(A \cap B) - \mathrm{P}(A)(\mathrm{P}(A \cap B) + \mathrm{P}(A^c \cap B)) \\
&= \mathrm{P}(A^c)\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(A^c \cap B)
\end{aligned}$$

Let $q = \mathrm{P}(A^c)$, then $\mathrm{P}(A \cap B) \le \mathrm{P}(A) \le 1 - q$. $0 \le \mathrm{P}(A^c)\mathrm{P}(A \cap B) \le \max_q\{q(1 - q) : 0 \le q \le 1\} = 1/4$. Similarly, $0 \le \mathrm{P}(A)\mathrm{P}(A^c \cap B) \le 1/4$. Hence, $|\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)| \le 1/4$. The bound is tight for the event A heads on a coin toss and the event B tails on the same coin toss. Then $A = B^C$ and $\mathrm{P}(A \cap B) = 0$ and $\mathrm{P}(A) = \mathrm{P}(B) = 1/2$, and the inequality is tight. $\diamond$

3 Conditioning

The (unconditional) probability of outcome 1 in an experiment of rolling a fair dice is 1/6. Given that an odd number is the outcome, the conditional probability of 1 is 1/3. Conditioning helps us to sharpening our assessment of probabilities.

Many contexts in practice involve decisions made in stages. The decision maker can then have some history of past events to base probabilities of future events on. If the past and future events are independent, the knowledge of the past events does not help to improve the probability assessment of the future events. But dependence of events is fairly common and can be exploited for better assessments. When event A happens before event B, we can ask the probability of event B after event A happens. This probability is conditioned on the fact that event A happens so it is called *conditional probability*. Note that we are not saying whether A has really happened. We are rather investigating the probability of B if A happens. The conditional probability of event B given event A with $\mathrm{P}(A) > 0$ is

$$\mathrm{P}(B|A) := \frac{\mathrm{P}(B \cap A)}{\mathrm{P}(A)}.$$

If events A, B are independent, $\mathrm{P}(B|A) = \mathrm{P}(B)$ whatever happens with event A. At this stage, we do not consider conditioning on an event with probability 0, so we assume $\mathrm{P}(A) > 0$. Appendix shows that the conditional probability defined above is a legitimate probability measure for an appropriately defined probability model.

Example: An instructor gives 5 questions for homeworks but grades only 2 of them. A student wants to solve only the questions that will be graded so he attempts to guess 2 questions correctly. He discovers that the instructor always asks a numerical question and grades it. Subsequently, he confidently guesses 1 question out of 5 correctly. What is the probability that he guesses 2 questions correctly given that he guesses 1 correctly? Let A_i be the event that he guesses i questions correctly for $i = 1, 2$. We are asking for $\mathrm{P}(A_2|A_1)$. Note that $A_2 \subseteq A_1$, so $\mathrm{P}(A_2|A_1) = \mathrm{P}(A_2)/\mathrm{P}(A_1)$. Inserting $\mathrm{P}(A_1) = 2/5$ and $\mathrm{P}(A_2) = 1/C_2^5 = 1/10$, we obtain $\mathrm{P}(A_2|A_1) = 1/4$. $\diamond$

Example: For 3 events A_1, A_2, B, we have $\mathrm{P}(A_1, A_2|B) = \mathrm{P}(A_1 \cap A_2 \cap B)/\mathrm{P}(B) = \mathrm{P}(A_2|A_1 \cap B)\mathrm{P}(A_1 \cap B)/\mathrm{P}(B) = \mathrm{P}(A_2|A_1 \cap B)\mathrm{P}(A_1|B)$. $\diamond$

Example: At a restaurant, there are two waitresses to greet and sit the customers. When you arrive a Young lady waitress sits you and you wonder about the probability of the other waitress to be also a Young lady as opposed to an Experienced lady. For simplicity, let us suppose that the waitresses can be only young or experienced, so the sample space for the waitresses of the restaurant is $\{YY, YE, EY, EE\}$ with equal probabilities. You may answer $1/2$ by thinking that the other waitress is either young or experienced with equal probabilities. You may also answer $1/3$ by computing $\mathrm{P}(YY)/(\mathrm{P}(YY) + \mathrm{P}(YE) + \mathrm{P}(EY)) = (1/4)/(3/4)$.

What is desired here is actually $\mathrm{P}(\text{the other is } Y| \text{ yours is } Y)$, which we can compute by considering $\mathrm{P}(\text{the other is } Y \text{ and yours is } Y)$ and $\mathrm{P}(\text{yours is } Y)$. We can see that $\mathrm{P}(\text{the other is } Y \text{ and yours is } Y) = 1/4$. Can we say that $\mathrm{P}(\text{your waitress is } Y) = 1/2$? No; if the waitresses are YY, this probability is 1; if the waitresses are EE, this probability is 0. This probability needs to be specified only when the waitresses are YE or EY, in which cases let q denote this probability. Then $\mathrm{P}(\text{your waitress is } Y) = 1/4 + q/4 + q/4$. Hence, $\mathrm{P}(\text{the other waitress is } Y| \text{ your waitress is } Y) = (1/4)/(1/4 + q/2) = 1/(1 + 2q)$. If $\mathrm{P}(\text{your waitress is } Y) = 1/2$, $\mathrm{P}(\text{the other waitress is } Y| \text{ your waitress is } Y) = 1/2$. If the restaurant has a policy of $q = 1$, i.e., young waitress sits the customers if available, then $\mathrm{P}(\text{the other waitress is } Y| \text{ your waitress is } Y) = 1/3$. At the other extreme of $q = 0$, $\mathrm{P}(\text{the other waitress is } Y| \text{ your waitress is } Y) = 1$. Note that depending on the restaurant's policy (i.e., q), the correct answer ranges from $1/3$ to $1/2$ then to 1. $\diamond$

The conditional probability equation can be rewritten as $\mathrm{P}(B \cap A) = \mathrm{P}(B|A)\mathrm{P}(A)$. On the other hand, for a partition $\{A_i\}_{i=1}^{\infty}$ of Ω, the *total probability formula* is

$$\mathrm{P}(B) = \mathrm{P}(B \cap \cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathrm{P}(B \cap A_i) = \sum_{i=1}^{\infty} \mathrm{P}(B|A_i)\mathrm{P}(A_i).$$

Since $\{A_i\}_{i=1}^{\infty}$ is a partition of Ω, we have $\cup_{i=1}^{\infty} A_i = \Omega$ and $A_i \cap A_j = \emptyset$ for $i \neq j$. These along with countable additivity and conditional probability equation are used in the above derivation.

Example: Car disc brakes are produced on one of the three machines: 1, 2, 3. Machine 1 produces twice as many discs as Machine 2 and Machine 3. The defective disc percentage on Machine 1 is 2%. The corresponding numbers for Machine 2 and 3 are 4% and 8%. What percentage of discs are defective? In other words, what is the probability that a randomly chosen disc is defective? Let B be the event that the brake is defective. Let A_i be the event that the brake is produced on Machine i for $i = 1, 2, 3$. From the exercise statement $\mathrm{P}(A_1) = 50\%$, $\mathrm{P}(A_2) = 25\%$ and $\mathrm{P}(A_3) = 25\%$. Convince yourself that $\Omega = A_1 \cup A_2 \cup A_3$ and A_1, A_2, A_3 are disjoint. Then we have $\mathrm{P}(B) = \mathrm{P}(B|A_1)\mathrm{P}(A_1) + \mathrm{P}(B|A_2)\mathrm{P}(A_2) + \mathrm{P}(B|A_3)\mathrm{P}(A_3) = 0.02(0.50) + 0.04(0.25) + 0.08(0.25)$. $\diamond$

Example: Does the first bidder have an advantage? Suppose that n suppliers bid for m projects of a buyer for $m \leq n$. The suppliers are to choose their turn to bid and are awarded projects depending on the number of suppliers and projects at the time of their bid. If a supplier is awarded a project, the number of available suppliers and projects both decrease by one. Otherwise, only the number of available suppliers decrease by one. Before the kth bidder, suppose there are n_k available suppliers and m_k projects remaining. Initially, $n_1 = n$ and $m_1 = m$. The buyer accepts the kth bidder randomly, i.e., with probability m_k/n_k. Is the first bidder have a higher chance of getting a project than the second or third?

Let A_k be the event that bidder k gets a project. $\mathrm{P}(A_1) = m/n$. To compute $\mathrm{P}(A_2)$, we use conditioning

$$\mathrm{P}(A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1) + \mathrm{P}(A_1^c)\mathrm{P}(A_2|A_1^c) = \frac{m}{n}\frac{m-1}{n-1} + \frac{n-m}{n}\frac{m}{n-1} = \frac{m(n-1)}{n(n-1)} = \frac{m}{n}.$$

To compute $\mathrm{P}(A_3)$, we use conditioning

$$\begin{aligned}
\mathrm{P}(A_3) &= \mathrm{P}(A_1 \cap A_2)\mathrm{P}(A_3|A_1 \cap A_2) + \mathrm{P}(A_1 \cap A_2^c)\mathrm{P}(A_3|A_1 \cap A_2^c) + \mathrm{P}(A_1^c \cap A_2)\mathrm{P}(A_3|A_1^c \cap A_2) \\
&\quad + \mathrm{P}(A_1^c \cap A_2^c)\mathrm{P}(A_3|A_1^c \cap A_2^c) \\
&= \frac{m}{n}\frac{m-1}{n-1}\frac{m-2}{n-2} + \frac{m}{n}\frac{n-m}{n-1}\frac{m-1}{n-2} + \frac{n-m}{n}\frac{m}{n-1}\frac{m-1}{n-2} + \frac{n-m}{n}\frac{n-1-m}{n-1}\frac{m}{n-2} \\
&= \frac{m}{n}\left\{\frac{(m-1)(m-2) + (n-m)(m-1) + (n-m)(m-1) + (n-m)(n-1-m)}{(n-1)(n-2)}\right\} = \frac{m}{n}
\end{aligned}$$

The first, second an third bidders all have m/n chance of getting a project. Note that we arrive at this conclusion under the unlikely assumption that the buyer randomly awards projects. $\diamond$

Example: Two players toss a fair coin by taking turns. Player A first tosses, then player B, and then player A again. The first player to get head wins. Does the first player, player A, have an advantage? The answer is yes, if the probability of getting a head on an odd toss is higher than the probability of getting a head on an even toss.

$$\begin{aligned}
\mathrm{P}(\text{First Head on odd toss}) &= \sum_{n=1}^{\infty}(1/2)^{2n-1} = (1/2)\sum_{n=0}^{\infty}(1/4)^n = (1/2)(4/3) = 2/3, \\
\mathrm{P}(\text{First Head on even toss}) &= \sum_{n=1}^{\infty}(1/2)^{2n} = (1/4)\sum_{n=0}^{\infty}(1/4)^n = (1/4)(4/3) = 1/3.
\end{aligned}$$

So the first player has an advantage. You should compare this conclusion with that of the previous exercise. $\diamond$

4 Bayes' Formula

Conditional probabilities $\mathrm{P}(A|B)$ and $\mathrm{P}(B|A)$ are not the same. Suppose that we interview a population of breast cancer patients to see if they drank excessive alcohol. The result might very well that they did. If we express being a Breast cancer patient as event B and drinking Alcohol as event A, we observe from the interviews that $P(A|B)$ is high. Perhaps what is more interesting is $P(B|A)$, to check if drinking causes breast cancer. However, we cannot directly find out $P(B|A)$ from $P(A|B)$. This is disappointing because both conditional probabilities have the common term $P(A \cap B)$. It turns out these conditional probabilities can be related to each other.

For a partition $\{A_i\}_{i=1}^{\infty}$ of Ω, we have *Bayes' formula*

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}.$$

Generally, Bayes' formula does not include the middle term provided above, we include the middle term to illustrate the intermediate logical step.

The right-hand side in the Bayes' formula has only $P(B|A_j)$s as the conditional terms while the left-hand side has $P(A_j|B)$, so these two different conditional probabilities are related to each other through the formula. Another way to look at the formula is to check how it expresses the probability of event A_j. Without knowing the outcome of event B, the probability of each event A_j is $P(A_j)$. After learning about the outcome of B, the probability becomes $P(A_j|B)$. Bayes' formula helps us to update the probability $P(A_j)$ with the knowledge of event B to obtain $P(A_j|B)$.

Example: Suppose that a person is accused of an offense Generally, a confession of the offense is treated as a sign that this person has really committed by the offense There are three relevant events in this case: I person is innocent, G the person is guilty, V the person verbally confesses. From the Bayes' formula we have the guilty and innocence probabilities given a confession:

$$P(G|V) = \frac{P(V|G)P(G)}{P(V|G)P(G) + P(V|I)P(I)} \quad \text{and} \quad P(I|V) = \frac{P(V|I)P(I)}{P(V|G)P(G) + P(V|I)P(I)},$$

which can be reorganized as

$$\frac{P(G|V)}{P(I|V)} = \frac{P(V|G)}{P(V|I)} \frac{P(G)}{P(I)}.$$

A verbal confession increases the guilty probability, i.e., $P(G|V)/P(I|V) > P(G)/P(I)$, if $P(V|G) > P(V|I)$. The last inequality says that a guilty person verbally confesses more readily than an innocent person. Is this really true?
ANSWER Suppose that an accused person is believed to be guilty with 70% and makes a verbal confession. A guilty person verbally confesses with 60% chance and an innocent person verbally confesses with 40%. After the verbal confession, what is the likelihood of the person to be guilty? We have $P(G) = 0.7 = 1 - P(I)$, $P(V|G) = 0.6$, $P(V|I) = 0.4$, then $P(G|V)/P(I|V) = (0.6/0.4)(0.7/0.3) = 7/2$ so $P(G|V) = 7/9 = 0.77 > 0.7$. In this case, a verbal confession increases the likelihood of guilt. On the other hand, you can check $P(G|V) < 0.7$ when $P(V|I) > 0.6$. That is, a verbal confession can decrease the likelihood of guilt!

We can also consider the event of a written confession W to obtain

$$\frac{P(G|V \cap W)}{P(I|V \cap W)} = \frac{P(W|G \cap V)}{P(W|I \cap V)} \frac{P(V|G)}{P(V|I)} \frac{P(G)}{P(I)}.$$

A written confession suffers from the same problem as a verbal confession. An innocent person may confess more readily than a guilty person $P(W|I \cap V) > P(W|G \cap V)$, which implies, counter to the intuition, that a written confession can be a sign of innocence. $\diamond$

Example: In a multiple choice exam, the instructor provides 4 alternatives: a), b), c), and d). A student knows the true answer to a particular question with probability 0.6. If the student does not know the answer, he randomly picks one of the alternatives from a) to d). If this question is answered correctly by the student,

what is the probability that it is answered correctly by chance? Let A_T be the event that the student knows the true answer. Let A_C be the event that the student answers the question correctly. We want to find $P(A_T^c|A_C)$.
ANSWER We know $P(A_T) = 1 - P(A_T^c) = 0.6$ and

$$P(A_T^c|A_C) = \frac{P(A_C|A_T^c)P(A_T^c)}{P(A_C|A_T^c)P(A_T^c) + P(A_C|A_T)P(A_T)} = \frac{(0.25)(0.4)}{(0.25)(0.4) + 1(0.6)} = \frac{1}{7} = 0.142.$$

That is, the student guesses the correct answer without knowing what the correct answer should be with 14.2% chance. Said differently, 85.8% of the time the student knowingly marks the correct answer. $\diamond$

Example: Your town has 40,000 potential attendees to a soccer game. Half the time all of these potential attendees go to a game and half of the time only half of these attendees go to a game. Let N be the number of attendees for an arbitrary game, so $P(N = 20,000) = P(N = 40,000) = 1/2$. You are one of the potential attendees and decide to attend the next game and let this event be denoted by A.
a) What is the probability that all 40,000 potential attends along with you when you attend, that is $P(N = 40,000|A)$?
ANSWER We start with $P(A|N = 40,000) = 1$; the number of attendees reaches 40,000 only with you; without you, there would be at most 39,999 attendees. On the other hand, $P(A|N = 20,000) =$ P(Being one of 20,000 attendees out of 40,000 potential attendees) $= 20,000/40,000 = 1/2$. Here we assume that each potential attendee is identical to the others. Then $P(N = 40,000, A) = P(A|N = 40,000)P(N = 40,000) = 1(1/2) = 1/2$ and $P(N = 20,000, A) = P(A|N = 20,000)P(N = 20,000) = (1/2)(1/2) = 1/4$. Note also that $P(N = 40,000, A^c) = 0$ and $P(N = 20,000, A^c) = 1/4$.

$$P(N = 40,000|A) = \frac{P(N = 40,000, A)}{P(N = 40,000, A) + P(N = 20,000, A)} = \frac{1/2}{1/2 + 1/4} = \frac{2}{3}.$$

b) Suppose your town has grown to a population of 60,000 potential attendees while we still have $P(N = 20,000) = P(N = 40,000) = 1/2$. What is $P(N = 40,000|A)$?
ANSWER Following similar steps as above $P(A|N = 40,000) = 2/3$ and $P(A|N = 20,000) = 1/3$. Then $P(N = 40,000, A) = P(A|N = 40,000)P(N = 40,000) = (2/3)(1/2) = 2/6$ and $P(N = 20,000, A) = P(A|N = 20,000)P(N = 20,000) = (1/3)(1/2) = 1/6$. Also $P(N = 40,000, A^c) = 1/6$ and $P(N = 20,000, A^c) = 2/6$.

$$P(N = 40,000|A) = \frac{P(N = 40,000, A)}{P(N = 40,000, A) + P(N = 20,000, A)} = \frac{2/6}{2/6 + 1/6} = \frac{2}{3}.$$

c) When your town grows to an unknown population of $\bar{n} > 40,000$ while $P(N = 20,000) = P(N = 40,000) = 1/2$, what is $P(N = 40,000|A)$?
ANSWER We have $P(A|N = 40,000) = 40,000/\bar{n}$ and $P(A|N = 20,000) = 20,000/\bar{n}$. Then $P(N = 40,000, A) = P(A|N = 40,000)P(N = 40,000) = (40,000/\bar{n})(1/2)$ and $P(N = 20,000, A) = P(A|N = 20,000)P(N = 20,000) = (20,000/\bar{n})(1/2)$.

$$P(N = 40,000|A) = \frac{P(N = 40,000, A)}{P(N = 40,000, A) + P(N = 20,000, A)} = \frac{20,000/\bar{n}}{20,000/\bar{n} + 10,000/\bar{n}} = \frac{2}{3}.$$

We arrive at the same probability of 2/3 regardless of the potential attendees $\bar{n}$ in the town. $\diamond$

Example: In Texas, 60% of the population is Republican and the rest is Democrat. 80% of Democrats are pro-choice, while 30% of Republicans are pro-choice. If Cindy is pro-choice, what is the probability that she is a Democrat?
ANSWER Let D be the event that Cindy is Democrat and C be the event that she is pro-Choice, so we want $P(D|C)$.

$$P(D|C) = \frac{P(C|D)P(D)}{P(C|D)P(D) + P(C|D^c)P(D^c)} = \frac{0.8(0.4)}{0.8(0.4) + 0.3(0.6)} = 0.64. \diamond$$

Example: Car disc brakes are produced on one of the three machines: 1, 2, 3. Machine 2 produces twice as many discs as Machine 1 and Machine 3. The defective disc percentage on Machine 1 is 2%. The corresponding numbers for Machine 2 and 3 are 4% and 8%. If a disc is defective, what is the probability that it is produced on Machine 1?

ANSWER Let D be the event that the disc brake is Defective. Let A_i be the event that the disc brake is produced on Machine i for $i = 1, 2, 3$. We want $\mathrm{P}(A_1|D)$. From the exercise statement $\mathrm{P}(A_1) = 25\%$, $\mathrm{P}(A_2) = 50\%$ and $\mathrm{P}(A_3) = 25\%$, from Bayes' formula

$$\mathrm{P}(A_1|D) = \frac{\mathrm{P}(D|A_1)\mathrm{P}(A_1)}{\mathrm{P}(D|A_1)\mathrm{P}(A_1) + \mathrm{P}(D|A_2)\mathrm{P}(A_2) + \mathrm{P}(D|A_3)\mathrm{P}(A_3)} = \frac{0.02(0.25)}{0.02(0.25) + 0.04(0.50) + 0.08(0.25)} = \frac{1}{9}. \diamond$$

5 Interpretation of Probability

Sticking to the axiomatic probability framework, it is technically possible to say probability is a measure defined for a collection of sets and obeys a few axioms. This however hardly satisfies a practitioner who needs to interpret, explain and decide by using probabilities.

Classical Interpretation: This interpretation dates back to the 19th century and argues that the probability of an event is the fraction of the outcomes leading to the event out of all outcomes in the sample space. The underlying assumption here is that each outcome in the sample space is equally likely, which is called *principle of indifference*. This holds in many examples of finite sample spaces generated by considering equally likely outcomes from experiments such as tossing a coin, rolling a die. But it does not hold when the outcomes do not have the same likelihood; some probability puzzles/parodoxes exploit the humane tendency to assume *principle of indifference* even when it fails.

Example: The experiment of rolling a die can yield two sample spaces $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$ and $\Omega_2 = \{1, 3, 5, \text{Even numbers: } \{2, 4, 6\}\}$. The outcomes in Ω_2 are not equally likely, so we cannot say that the outcome of 1 happens with probability of 1/4 by basing it on Ω_2. $\diamond$

Logical Interpretation and Propensity Interpretation: These two interpretations were put forward in the 1950s respectively by Carnap and Popper. Carnap arguably was inspired by the deductive logic such as the implication $E \implies H$ from E(vidence) to Hypothesis. Implications in deductive logic either hold (logical value of 1) or not (logical value of 0); Carnap attempts to extend this by arguing that probability is a measure of the support of E for H and the measure can take values in $[0, 1]$. "Probability, is the degree of confirmation of a hypothesis H with respect to an evidence statement E, e.g., an observational report" (p.19 Carnap 1963). That is to Carnap, the probability of a hypothesis with respect to some given evidence is the truth of a sequence of logical relations that bridge the hypothesis to the evidence. Either such a bridge and evidence exist or they do not, hence forcing the probability to be either 0 or 1.

Example: We may be interested in the probability of the hypothesis that artificial intelligence (AI) will automate and expedite accounting processes of filing reports. To asses this probability through logical interpretations, we can present the evidence that AI is deployed in accounting and initial AI applications to accounting have automated and expedited reporting. We can also present evidence from other applications of AI, for example call center operations have been automated and expedited by AI. If the evidence is convincing and the logical deduction from the evidence to the hypothesis is correct, we assign probability of 1 to the hypothesis. $\Diamond$

Popper's propensity interpretation has two forms single-case propensity (for an event that happens once) and long-run propensity (for an event that happens many times). Single-case propensity is designed to talk about events that are very rare such as the failure of a nuclear power plant. But making single-case propensity operational is very hard: before the event's occurrence we have no (or limited) data to assess propensity and afterwards we have no interest as the event will not happen again. Long-run propensity interpretation seems to be more or less the same as the frequentist interpretation below. Since computations with logical and propensity interpretations are either inconvenient or simply analogous to frequentist interpretation, these interpretations are not popular among management scientists.

Frequentist Interpretation: The probability of an event is said to be its frequency of occurrence in many repeated experiments. This interpretation does not necessarily require the repeat of the experiment over a long run as experiments can be run in parallel by different agents. Hence, it is readily adopted by statisticians, management scientists and other scientists dealing with a population of agents.

Example: Ten identical dies are rolled by 10 people at once. Let o_i be the outcome of the ith die for $o_i \in \{1, 2, \ldots, 6\}$. Then the probability of 4 on a roll is approximated by

$$\frac{1}{10} \sum_{i=1}^{10} \mathbb{1}_{o_i=4}.$$

The same die can also be rolled by a single person 10 times or 1,000 times. Note that it is easy to repeat the rolling experiment. $\diamond$

Some outcomes such as the failure of a nuclear power plant or an earthquake are so rare that interpretations of their probabilities cannot be done through frequencies. What is meant by the probability of 10^{-5} for the failure of a plant or the crash of a space shuttle? Surely, we do not repeat failures 10^5 times to get to these probabilities. Reliability theory actually offers methods of obtaining probabilities for the failure of complex systems by starting with the analyses of components But this bottom-up approach is a computational one and does not help with the interpretation of the probability 10^{-5}. To fend off the criticism of lack of interpretation for rare events, Popper suggested single-case propensity, which has its own difficulties.

An addressable difficulty of the frequentist approach is interpreting the irrational probabilities such as those involving number π. When the probability is an irrational number, it can not be represented by a ratio of two integers. But the frequentist interpretation is based on the ratio of two integers: the number of occurrences of the desired outcome divided by the number of experiments.

Example: Consider a square dartboard whose each side is 2 units. There is a square of radius 1 unit drawn at the center of the board. Assuming that all the darts you throw hit the square, what is the probability that they are in the circle? This requires comparing the area $\pi 1^2$ of the circle with the area 2^2 of the square, so the probability is $\pi/4$, which is an irrational number. $\diamond$

The particular difficulty of interpreting an irrational probability is overcome by considering this probability as the limit of rational probabilities that can be interpreted via frequentist approach. Reichenbach and von Mises separately worked on formalizing this limit-based interpretation, which is commonly accepted now.

Subjective Interpretation: Probabilities can be interpreted relatively depending on the person doing the interpretation. This immediately leads to two different probability assessments for the same event but by two different people. When personalized, probabilities become beliefs (credence or degrees of confidence). It is very natural for two people to have two beliefs about the same phenomenon. This is the case for example in a trade, where a buyer and a seller could attach different values (utilities) for the same item.

Example: When a graduate student needs a used bike, he first checks a webpage listing items for sale. He finds an appropriate bike sold for \$100. The seller must believe that the value of the bike is \$100. The graduate student evaluates that the value is \$100 if he rides the bike 4 days of each week and the value is \$150 if he rides the bike 6 days of each week. He thinks that there will be several Fall and Spring weeks during which he can ride the bike 6 days. Hence, the buyer (graduate student) has a higher belief for the bike's value than the seller. This in general is the condition for the sales transaction to happen. Otherwise the buyer believes the bike to be less of a value than its price, the transaction does not happen. In summary, presence of sales transactions confirm that at least some buyers and sellers have different beliefs for the utility of the sold item. $\diamond$

The example refers to value of an item as a random variable, but it is related to probabilities such as $\mathrm{P}(\text{Value} = 100)$ and $\mathrm{P}(\text{Value} = 150)$. Subjective probabilities are personal assessments, so they must be from the standpoint of an individual or a group of people. Whose standpoint should we be concerned with? For example in a sale transaction, we consider both the buyer's and seller's probabilities regarding the value of an item. When a firm is marketing to segments of consumers, we consider each segment's value. It seems at least two standpoints are necessary to appreciate the relativity in probabilities and make a difference between subjective probabilities and (objective) probabilities. As a note on terminology, subjective vs. objective probabilities may not be the ideal terms as personal probabilities sound better.

Example: Four people become friends at a university to make up a Thursday evening parlor game group. The group contains a classical interpreter of probability Simon (named after Laplace), a logical (propensity) interpreter Rudolf (after Carnap), a frequentist intrepreter Hans (after Reichenbach) and a subjective interpreter Bruno (after de Finetti). While discussing interpretations of probability, they play a version of monopoly game where a fair dice with 6 faces is rolled. The probability of 1 showing on the dice is 1/6. This number can be interpreted differently by each of these players.
– Simon: There are 6 possible outcomes and each is equally likely, i.e., there is no reason to favor one of the outcomes over others.
– Rudolf: We can start with the hypothesis that the dice shows 1. The available evidence including a fair dice, and the logical relation between this evidence and the hypothesis leads me to conclude that the propensity of each outcome is the same as the outcome of 1.
– Hans: If I roll the dice 6 times, I expect to see 1 exactly once. If I roll the dice 12 times, I expect to see 1 exactly twice, and so on.
– Bruno: I heard of your personal beliefs. As far as I am concerned, each of the 6 outcomes are equally likely, then my belief for outcome 1 is that it happens with probability 1/6. $\Diamond$

Personal interpretation of probabilities is attributed mostly to de Finetti (Cifarelli and Regazzini 1996) and Savage. It is discussed among mathematicians (Marschak 1975), especially via comparisons with Kolmogorov's axiomatic framework (Borkar et al. 2004). Personal interpretation of probabilities have arguably led first to behavioral economics and later to behavioral risk and finance (Thaler 1999).

6 Solved Exercises

1) 90% of vehicles passing a gas station are cars and 10% are buses 5% of the buses stop at the station and 10% of the cars do. What is the probability that a randomly chosen vehicle is a bus that will not stop at the station?
ANSWER Let $A = \{\text{Vehicle is bus}\}$, $B = \{\text{Vehicle does not stop}\}$, we want $\mathrm{P}(A \cap B)$. $\mathrm{P}(A) = 0.1$, $\mathrm{P}(B^c|A) = 0.05$, $\mathrm{P}(B|A) = 1 - \mathrm{P}(B^c|A) = 0.95$. $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B|A) = 0.095$. $\diamond$

2) Two diseases are common in a population. 10% of the population contract disease I and 20% contract disease II, while 5% contract both diseases. Find the probability that a randomly chosen person will contract at least one disease. Find the conditional probability that a randomly chosen person will contract both disease given that s/he has contracted at least one disease.
ANSWER We have $\mathrm{P}(I) = 0.1$, $\mathrm{P}(II) = 0.2$, $\mathrm{P}(I \cap II) = 0.05$. We want $\mathrm{P}(I \cup II)$ and $\mathrm{P}(I \cap II|I \cup II)$. $\mathrm{P}(I \cup II) = \mathrm{P}(I) + \mathrm{P}(II) - \mathrm{P}(I \cap II) = 0.25$. $\mathrm{P}(I \cap II|I \cup II) = \mathrm{P}(I \cap II)/\mathrm{P}(I \cup II) = 0.05/0.25 = 0.2$. $\diamond$

3) An advertising company notices that approximately 1 in 50 potential buyers sees a magazine ad for a product and 1 in 5 sees the corresponding ad on TV. 1 in 100 sees both. 1 in 3 purchases the product after seeing the ad and 1 in 10 purchases without seeing any ad. What is the probability that a randomly selected potential buyer will purchase the product?
ANSWER Let use define events $A = \{\text{Buyer sees magazine ad}\}$, $B = \{\text{Buyer sees TV ad}\}$ and $C = \{\text{Buyer purchases the product}\}$, we want $\mathrm{P}(C)$. We have $\mathrm{P}(A) = 0.02$, $\mathrm{P}(B) = 0.20$, $\mathrm{P}(A \cap B) = 0.01$, $\mathrm{P}(C|A \cup B) = 1/3$ and $\mathrm{P}(C|(A \cup B)^c) = 1/10$. $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) = 0.02 + 0.20 - 0.01 = 0.21$. $\mathrm{P}((A \cup B)^c) = 1 - \mathrm{P}(A \cup B) = 0.79$. $\mathrm{P}(C) = \mathrm{P}(C|A \cup B)\mathrm{P}(A \cup B) + \mathrm{P}(C|(A \cup B)^c)\mathrm{P}((A \cup B)^c) = (1/3)0.21 + (1/10)0.79 = 0.149$. $\diamond$

4) Tim's room has two bookshelves, the first has 5 mystery novels and 3 science fiction novels; the second has 4 mystery novels and 6 science fiction novels. Tim randomly chooses a bookshelf first and then takes a book from that shelf. What is the probability that Tim takes a science fiction novel?
ANSWER Let $B = \{\text{Bookshelf 1 is chosen}\}$, $A = \{\text{Science fiction book is taken}\}$, we want $\mathrm{P}(A)$. $\mathrm{P}(A) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^c) = \mathrm{P}(B)\mathrm{P}(A|B) + \mathrm{P}(B^c)\mathrm{P}(A|B^c) = (1/2)(3/8) + (1/2)(6/20)$. $\diamond$

5) Five machines (numbered 1 through 5) are available for use, and machine 2 is worn out. Machines 1 and 2 come from supplier I and machines 3, 4, 5 come from supplier II. Suppose two machines are randomly selected for production on a particular day. Let A be the event that the wornout machine is selected and B be the event that at least one machine come from supplier I. Find $\mathrm{P}(A)$, $\mathrm{P}(B)$, $\mathrm{P}(A \cap B)$.
ANSWER Ω is the pair of machines selected out of 5, so $|\Omega| = C_2^5 = 10$. $A = \{$(machines 2 and 1), (machines 2 and 3), (machines 2 and 4), (machines 2 and 5)$\}$. $\mathrm{P}(A) = |A|/|\Omega| = 4/10$. $B^c = \{\text{Both machines are from supplier II}\}$ and $|B^c| = C_2^3 = 3$. $\mathrm{P}(B^c) = 3/10$ and $\mathrm{P}(B) = 1 - 3/10 = 7/10$. All of the elements in A are also in B, so $A \subseteq B$. Then $\mathrm{P}(A \cap B) = \mathrm{P}(A) = 4/10$. $\diamond$

6) A car's dashboard light is supposed to flash when the oil pressure is too low. Probability of light flashing when oil is low is 99%. 2% of the time light flashes for no reason. If there is 10% chance that the oil pressure really is low. What is the probability that a driver should be concerned if the warning light goes on?
ANSWER Let $A = \{\text{Light goes on}\}$ and $B = \{\text{Oil is too low}\}$. We have $\mathrm{P}(A|B) = 0.99$, $\mathrm{P}(A|B^c) = 0.02$ and $\mathrm{P}(B) = 0.10$ and want $\mathrm{P}(B|A)$.

$$\mathrm{P}(B|A) = \frac{\mathrm{P}(B)\mathrm{P}(A|B)}{p(B)\mathrm{P}(A|B) + p(B^c)\mathrm{P}(A|B^c)} = \frac{(0.10)(0.99)}{(0.10)(0.99) + (0.90)(0.02)}. \quad \diamond$$

7) A fair dice is rolled 7 times, what is the probability that at least one of the six faces of the dice never shows up? What is the same probability after 14 rolls?
ANSWER Let A_i be the event that face i does not show up for $i = 1 \ldots 6$. We need $\mathrm{P}(\cup_{i=1}^6 A_i)$. We have $\mathrm{P}(A_i) = (5/6)^7$, $\mathrm{P}(A_i \cap A_j) = (4/6)^7$, $\mathrm{P}(A_i \cap A_j \cap A_k) = (3/6)^7$, $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_l) = (2/6)^7$, $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_l \cap A_m) = (1/6)^7$ for $1 \leq i < j < k < l < m \leq 6$, these six events cannot occur simultaneously.

$$\begin{aligned}\mathrm{P}(\cup_{i=1}^6 A_i) &= C_1^6(5/6)^7 - C_2^6(4/6)^7 + C_3^6(3/6)^7 - C_4^6(2/6)^7 + C_5^6(1/6)^7 - 0 \\ &= 6(5/6)^7 - 15(4/6)^7 + 20(3/6)^7 - 15(2/6)^7 + 6(1/6)^7 \ = \ 0.96\end{aligned}$$

For 14 rolls,

$$\begin{aligned}\mathrm{P}(\cup_{i=1}^{14} A_i) &= C_1^6(5/6)^{14} - C_2^6(4/6)^{14} + C_3^6(3/6)^{14} - C_4^6(2/6)^{14} + C_5^6(1/6)^{14} - 0 \\ &= 6(5/6)^7 - 15(4/6)^7 + 20(3/6)^7 - 15(2/6)^7 + 6(1/6)^7 \ = \ 0.42\end{aligned}$$

As the number of rolls double, the probability that at least one of the faces does not show up drops by more than half. $\diamond$

8) Birthday problem. What is the probability that two or more people have the same birthday (out of 365 in a year) in a group of N people?

ANSWER Let A be the event that each of N people has a distinct birthday for $N \leq 365$. We want $\mathrm{P}(A^c)$. There are 365^N ways of assigning days to N people as birthdays. If the birthdays are to be distinct, the first person can be assigned 365 days, the second can be assigned 364 days, ..., the Nth can be assigned $365 - N + 1$ days. So $1 - \mathrm{P}(A^c) = \mathrm{P}(A) = P_N^{365}/(365^N)$. $\diamond$

9) Let π be a random permutation of numbers from 1 to 9. For example, 9, 8, 7, 6, 5, 4, 3, 2, 1 is a permutation that is represented by $\pi(1) = 9$, $\pi(2) = 8$, $\pi(3) = 7$, $\pi(4) = 6$, $\pi(5) = 5$, $\pi(6) = 5, \ldots, \pi(9) = 1$. Another permutation is 5, 9, 8, 7, 6, 4, 3, 2, 1 represented by $\pi(1) = 5$, $\pi(2) = 9$, $\pi(3) = 8$, $\pi(4) = 7$, $\pi(5) = 6$, $\pi(6) = 4, \ldots, \pi(9) = 1$. Note that 9, 8, 7, 6, 5, 4, 3, 2, 1 has $\pi(i) = i$ for $i = 5$, while 5, 9, 8, 7, 6, 4, 3, 2, 1 has no i such that $\pi(i) = i$. What is the probability that at least on i satisfies $\pi(i) = i$ in a random permutation of 9 numbers?

ANSWER Let $A_i = \{\pi(i) = i\}$ for $i = 1 \ldots 9$, we want $\mathrm{P}(\cup_{i=1}^9 A_i)$. $\mathrm{P}(A_i) = 8!/9!$, $\mathrm{P}(A_i \cap A_j) = 7!/9!$, $\mathrm{P}(A_i \cap A_j \cap A_k) = 6!/9!$, $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_l) = 5!/9!$, $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_l \cap A_m) = 4!/9!$, $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_l \cap A_m \cap A_n) = 3!/9!$, $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_l \cap A_m \cap A_n \cap A_o) = 2!/9!$, $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_l \cap A_m \cap A_n \cap A_o \cap A_p) = 1!/9!$, $\mathrm{P}(A_i \cap A_j \cap A_k \cap A_l \cap A_m \cap A_n \cap A_o \cap A_p \cap A_q) = 0!/9!$ for $1 \leq i < j < k < l < m < n < o < p < q \leq 9$. By the inclusion-exclusion identity,

$$
\begin{aligned}
\mathrm{P}(\cup_{i=1}^9 A_i) &= C_1^9 \frac{8!}{9!} - C_2^9 \frac{7!}{9!} + C_3^9 \frac{6!}{9!} - C_4^9 \frac{5!}{9!} + C_5^9 \frac{4!}{9!} - C_6^9 \frac{3!}{9!} + C_7^9 \frac{2!}{9!} - C_8^9 \frac{1!}{9!} + C_9^9 \frac{0!}{9!} \\
&= 1 - 1/2! + 1/3! - 1/4! + 1/5! - 1/6! + 1/7! - 1/8! + 1/9! \\
&= 1 - (1/2! - 1/3! + 1/4! - 1/5! + 1/6! - 1/7! + 1/8! - 1/9!) = 0.632121 \\
&\approx 1 - \exp(-1) = 0.632121.
\end{aligned}
$$

10) Consider three independent events A, B, C.

a) Either show that C is independent of $A \cap B$ or provide a counterexample for dependence.

ANSWER From the independence of A, B, C, we have $\mathrm{P}(C)\mathrm{P}(A \cap B) = \mathrm{P}(C)\mathrm{P}(A)\mathrm{P}(B) = \mathrm{P}(C \cap A \cap B)$. $\mathrm{P}(C)\mathrm{P}(A \cap B) = \mathrm{P}(C \cap A \cap B)$ implies the independence of C and $A \cap B$.

b) Either show that C is independent of $A \cup B$ or provide a counterexample for dependence.

ANSWER From the independence of A, B, C, we have $\mathrm{P}(C)\mathrm{P}(A \cup B) = \mathrm{P}(C)(\mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)) = \mathrm{P}(C \cap A) + \mathrm{P}(C \cap B) - \mathrm{P}(C \cap A \cap B) = \mathrm{P}((C \cap A) \cup (C \cap B)) = \mathrm{P}(C \cap (A \cup B))$. $\mathrm{P}(C)\mathrm{P}(A \cup B) = \mathrm{P}(C \cap (A \cup B))$ implies the independence of C and $A \cap B$.

11) Consider N distinct objects and cycles associated with the permutations of these objects.

a) What is the probability that object 1 is in a k-cycle?

ANSWER We note two facts $i)$ The number of possible k-cycles containing 1 is $C_{k-1}^{N-1}(k-1)! = \frac{(N-1)!}{(N-k)!}$.
$ii)$ The number of ways to complete a permutation once a k-cycle is chosen is $(N-k)!$. So there are $(N-1)!$ permutations of N objects in which object 1 is in a k-cycle.

$$
\mathrm{P}(1 \text{ is in a } k\text{-cycle}) = \frac{\text{Number of permutations in which 1 is in a } k\text{-cycle}}{\text{Total number of permutations}} = \frac{(N-1)!}{N!} = \frac{1}{N}.
$$

b) What is the probability that object N is in the same cycle as object 1 given that object 1 is in a k-cycle?

ANSWER There are $(N-1)!$ permutations of N numbers in which 1 is in a k-cycle. Pair number 1 and N and call it $[1\&N]$, this reduces the number of available objects to $N-1$. Then there are $(N-2)!$ permutations of $N-1$ objects in which $[1\&N]$ is in a k-cycle. Take each k-cycle and separate $[1\&N]$: put 1 into the 1st position and put N into one of the remaining $k-1$ positions. Hence, there are $(N-2)!(k-1)$ permutations of N

objects in which objects 1 and N are in the same k-cycle.

$$\begin{aligned}
\mathrm{P}(N \text{ and } 1 \text{ are in the same } k\text{-cycle}|1 \text{ is in a } k\text{-cycle}) &= \frac{\mathrm{P}(N \text{ and } 1 \text{ are in the same } k\text{-cycle})}{\mathrm{P}(1 \text{ is in a } k\text{-cycle})} \\
&= \frac{(N-2)!(k-1)/N!}{(N-1)!/N!} = \frac{k-1}{N-1}.
\end{aligned}$$

Note that $\mathrm{P}(N$ and 1 are in the same 1-cycle$|$1 is in a 1-cycle$)$=0 and $\mathrm{P}(N$ and 1 are in the same N-cycle$|$1 is in a N-cycle$)$=1.
c) What is the probability that object N is in the same cycle as object 1?
ANSWER

$$\begin{aligned}
\mathrm{P}(N \,\&\, 1 \text{ in the same cycle}) &= \sum_{k=1}^{N} \mathrm{P}(N \,\&\, 1 \text{ in the same } k\text{-cycle}) = \sum_{k=1}^{N} \frac{k-1}{N-1}\frac{1}{N} \\
&= \frac{1}{N(N-1)} \sum_{k=1}^{N} (k-1) = \frac{1}{N(N-1)} \frac{N(N-1)}{2} = \frac{1}{2}.
\end{aligned}$$

Interestingly, this probability does not depend on N.

7 Exercises

1. The world consumption of oil leaves less oil on/in Earth. With the current consumption rates, we know that the probability of having some oil resources available (no depletion) in the next ten years is positive. Let A_n be the event that the world runs out of oil by the nth year from today. Running out of oil means having no oil molecule left on/in Earth. Formally, consider all possible events that can materialize from now and think of each sequence of events (a scenario) as an outcome ω. It is customary to talk of scenarios as outcomes when talking about future.
   a) Are A_1 and A_2 independent?

   b) Is $\{A_n\}$ a monotone sequence of events. If yes, is it increasing or decreasing (both are meant in non-strict sense)? If no, find a counterexample with three events to establish lack of monotonicity.

   c) Describe a scenario that you can create by spending less than \$10 and guarantee that the world does not run out of oil in the next twenty years. Said differently, the scenario must make the world run out of oil with probability of zero: $\mathrm{P}(\cup_{n=1}^{20} A_n) = 0$.

2. For a collection of countable events $\{A_j\}$, establish the inequality $\mathrm{P}(\cup_j A_j) \le \sum_j \mathrm{P}(A_j)$. Hint: Let $B_j = A_j \setminus \cup_{i=1}^{j-1} A_i$ so that $\mathrm{P}(B_j) \le \mathrm{P}(A_j)$ and write $\mathrm{P}(\cup_j B_j)$ in terms of the probability of A_js.

3. Let A and B be two independent events and define $C := (A \cup B) \setminus (A \cap B)$. Is C the event of both A and B occurring; or the event of either A or B occurring; or the event of either one of the events occurring without the other? If C is something else, express it in English. Find $\mathrm{P}(C)$ in terms of $\mathrm{P}(A)$ and $\mathrm{P}(B)$.

4. Events A_1 and A_2 are said to be conditionally independent given event B if

   $$\mathrm{P}(A_1 \cap A_2 | B) = \mathrm{P}(A_1|B)\mathrm{P}(A_2|B).$$

   Even when events A_1 and A_2 are independent, they do not have to be conditionally independent. To establish this, we need a counterexample as follows. In an experiment of two independent dice rolls, let A_i be the event that the roll i shows 6 for $i = 1, 2$. Note that events A_1 and A_2 are independent. Choose an event C such that $\mathrm{P}(A_1|C), \mathrm{P}(A_2|C) > 0$ while $\mathrm{P}(A_1 \cap A_2|C) = 0$. Describe such an event C and compute probabilities $\mathrm{P}(A_1|C), \mathrm{P}(A_2|C)$.

5. Suppose that A, B are two independent events. Either prove that A is independent of B^c or provide a counterexample to show that A and B^c can be dependent.

6. A coin is tossed twice. We consider the following events. T_1: Tails on the first toss. T_2: Tails on the second toss. E: The two tosses are equal (the same).
a) Decide if T_1, T_2, E are pairwise independent.
b) Decide if T_1, T_2, E are independent.
c) Decide if E is independent of $T_1 \cap T_2$.

7. 6 couples attend a tango class. Tango instructor shuffles the couples randomly before the first class to test if a man can dance well with a woman he never danced before. What is the probability that at least one man dances with the woman that he came to the class with?

8. You have a slight fever, short coughs and fatigue for the last two days. Suspecting of covid-19 but unsure, you go to a covid testing center, take a swab, insert it into your nose and get a sample for nasal swab covid test. At the test center, it is explained to you that 10% of people in your neighborhood has covid. Moreover, the nasal swab test does not completely identify the covid virus.
   - If a patient has covid, the test is positive with 60% probability and negative with 40% probability.
   - If a patient has no covid, the test is positive with 10% probability and negative with 90% probability.

   a) Your first covid test comes back negative. To be certain, you ask for a second test which also comes back negative. Assuming that the tests are independent, what is the probability of having covid?

   b) You learn that the testing laboratory lost your second swab and reported the result of the first test as the result of the second test. How does this affect the independence of tests? How would you modify the probability of having covid, if you modify it?

   c) The loss of your second test instigates an investigation of the covid testing procedures. The investigation finds out that the sample collecting nurse does not insert the swab all the way into the potential patient's nasal cavity. She erroneously collects samples only from the bottom of the nose where covid virus does not inhabit. What would be the positive and negative results for this erroneous test?

9. Your department at the university invites $\bar{n}$ people to its weekly research seminar. The number N of people that show up in a seminar has been n_l, n_m, n_h in the last year with probabilities $\mathrm{P}(N = n_l) = p_l$, $\mathrm{P}(N = n_m) = p_m$, $\mathrm{P}(N = n_h) = p_h$ for $2 \le n_l < n_m < n_h \le \bar{n}$ and $p_l + p_m + p_h = 1$. We assume that each invitee is identical to others in terms of probability of showing up. You are also invited to seminars and let the event of your attendance to a seminar be denoted by A.
a) What is the probability of n_l people in attendance if you attend, that is $\mathrm{P}(N = n_l|A)$?

   b) Suppose that $p_h + p_l = 1$ and find $\mathrm{P}(N = n_h|A)$ and compare with $\mathrm{P}(N = n_l|A)$. When you are in attendance, do you expect to find more or fewer people attending the seminar?

   c) While going for a seminar, you run into a friend and both of you realize that you are going to the same seminar. We let event A_i for $i \in \{1, 2\}$ be the attendance of you and your friend. What is the probability of n_l people in attendance if you both attend, that is $\mathrm{P}(N = n_l|A_1, A_2)$?

10. Suppose that you have 5 coins. 2 are double headed; 2 are double tailed and 1 is normal. Double-headed coins have heads on both sides. Double-tailed coins have tails on both sides. Let M, O and N respectively be the event that a double-headed, double-tailed and normal coin is picked. Let H_i^u and H_i^l be the event that the ith coin toss has heads respectively on the upper and lower side.

    a) You close your eyes to pick a coin randomly and toss it. What is the probability that the lower face of this coin is head, i.e., $\mathrm{P}(H_1^l)$? What is the probability that the upper face of this coin is head, i.e., $\mathrm{P}(H_1^u)$?

    b) You open your eyes and see heads on the upper side of the coin, what is the probability that the lower side is head as well, i.e., $\mathrm{P}(H_1^l|H_1^u)$?

c) You pick up this coin and toss it again. What is the probability that the lower side is head, i.e. $P(H_2^l|H_1^u)$?

d) The second toss yields head on the upper side. What is the probability that the lower side is head, i.e. $P(H_2^l|H_1^u \cap H_2^u)$?

e) In parts a-d), we have used a single coin, now discard it so that you have only 4 coins left. What is the probability that the discarded coin is double-headed, i.e., $P(M|H_1^u \cap H_2^u)$?

f) After discarding the coin in e), randomly pick a coin from the remaining 4 coins and toss it, what is the probability that this third toss shows head? Hint: First find $P(H_3^u|H_1^u \cap H_2^u$, double-headed coin discarded) and $P(H_3^u|H_1^u \cap H_2^u$, normal coin discarded)?

11. A private company's revenue can either decrease D or increase I in a year. Obtaining the same amount of revenue two years in a row is a very distant possibility that we ignore. We suppose that increase and decrease are equally likely with probability 50% if nothing else is known. Let A_1 be the event that revenue has increased once and decreased once in the last two years. Let A_2 be the event that the revenue decreased at most once in the last two years.
a) Are A_1 and A_2^c disjoint? Are they independent? Can disjoint events be dependent?
b) We are told that the revenue will increase with probability $6/10$ in the next year if it did so in the last two years. It will increase with probability $3/10$ if it decreased in the last two years. The probability of increase in the next year is 50% if the last two years experienced a revenue increase and a decrease. If you are told that the revenue will increase in the next year, what is the probability that it increased in the last two years?

12. Suppose we select a point at random from a set of four points $\{(0,0,0),(1,1,0),(0,1,1),(1,0,1)\}$. Let event A_i happen if the ith coordinate of the selected point is 1. For example, if $(0,0,0)$ is chosen, A_1, A_2 or A_3 does not happen; if $(1,1,0)$ is chosen, A_1 and A_2 happen but A_3 does not happen. Are any pair of events pairwise independent? Are they independent?

13. A numerical problem is asked to a student. Its answer is an integer between 1 and K. The student can either solve the problem to find the correct answer or does not know how to solve the problem and guesses the correct answer. The probability of solving the problem correctly and then answering correctly for the student is q. With the complementary probability $1-q$, the student cannot solve the problem and guesses an answer. Then the student can give as an answer any of the numbers $1,\dots,K$ with equal probabilities. What is the probability that the problem is correctly solved (without any guessing) if the student gives a correct answer? Check to see if your answer makes sense when $K=1$ and as $K \to \infty$.

14. A student wants to make up a schedule for a seven day period to study one of four subjects on each day. Subjects are Probability, Stochastics, Optimization and Analysis. There are are 4^7 possible schedules. The student creates all possible schedules and randomly selects one. What is the probability that the selected schedule devotes at least one day to each subject?

15. An instructor prepares a test bank of 40 questions. These questions are picked up randomly to make up 5-question exams for 4 students. What is the probability that no two students get the same question, i.e., the probability of completely distinct questions? Find the same probability assuming that each exam has only 1 question.

Appendix: Conditioning Yields a Probability Measure

Starting with the probability model $(\Omega, \mathcal{F}, \mathrm{P})$ and $A \in \mathcal{F}$, we seek to obtain a new conditional probability measure P_A defined on sample space A and σ-algebra $\mathcal{F}_A$. We need to specify what $\mathcal{F}_A$ and P_A really are.

Given a measurable space $(\Omega, \mathcal{F})$ and $A \in \mathcal{F}$, we can construct the restriction of $\mathcal{F}$ to A denoted by $\mathcal{F}_A$. $\mathcal{F}_A$ is defined such that $B \in \mathcal{F}_A$ iff $B \in \mathcal{F}$ and $B \subseteq A$. First, we need to check if $\mathcal{F}_A$ is a σ-field on A.

i) $A \in \mathcal{F}_A$ because $A \in \mathcal{F}$ and $A \subseteq A$.

ii) $B \in \mathcal{F}_A$ implies $B^c \in \mathcal{F}_A$. To obtain this, we first note that A is the sample space for $\mathcal{F}_A$ so $B^c = A \setminus B = A \cap B^c$. Since $B \in \mathcal{F}_A$, we have $A, B \in \mathcal{F}$, so $B^c = A \cap B^c \in \mathcal{F}$. Since $B^c = A \cap B^c$, we have $B^c \subseteq A$. Combining the last two sentences, $B^c \in \mathcal{F}_A$.

iii) $B_i \in \mathcal{F}_A$ implies $\cup_{i=1}^{\infty} B_i \in \mathcal{F}_A$. To obtain this, we start by observing that $B_i \in \mathcal{F}$ and $\cup_{i=1}^{\infty} B_i \in \mathcal{F}$. Then we note that $B_i \subseteq A$ leads to $\cup_{i=1}^{\infty} B_i \subseteq A$. The last two sentences imply $\cup_{i=1}^{\infty} B_i \in \mathcal{F}_A$.

Since $\mathcal{F}_A$ is a σ-field, it can be paired with the sample space A to obtain the measurable space $(A, \mathcal{F}_A)$.

For A with $\mathrm{P}(A) > 0$ and each $B \in \mathcal{F}_A$, we define

$$\mathrm{P}_A(B) := \mathrm{P}(A \cap B)/\mathrm{P}(A).$$

Is P_A a probability measure on $(A, \mathcal{F}_A)$?

i) P_A is defined for every $B \in \mathcal{F}_A$ and $\mathrm{P}_A : \mathcal{F}_A \to [0, 1]$.

ii) $\mathrm{P}_A(A) = \mathrm{P}(A \cap A)/\mathrm{P}(A) = 1$.

iii) For disjoint $B_i \in \mathcal{F}_A$, we have $\cup_{i=1}^{\infty} B_i \in \mathcal{F}_A$ and

$$\mathrm{P}_A(\cup_{i=1}^{\infty} B_i) = \frac{\mathrm{P}((\cup_{i=1}^{\infty} B_i) \cap A)}{\mathrm{P}(A)} = \frac{\mathrm{P}(\cup_{i=1}^{\infty} (B_i \cap A))}{\mathrm{P}(A)} = \frac{\sum_{i=1}^{\infty} \mathrm{P}(B_i \cap A)}{\mathrm{P}(A)} = \sum_{i=1}^{\infty} \mathrm{P}_A(B_i),$$

where the third equality follows from $B_i \cap A \in \mathcal{F}$ and countable additivity of P. Equalities above establish that P_A is countably additive on $\mathcal{F}_A$.

Putting the last two paragraphs together, we have obtained that $(A, \mathcal{F}_A, \mathrm{P}_A)$ is a probability model and in particular P_A is a legitimate probability measure.

$(\Omega, \mathcal{F}, \mathrm{P}_A)$ is also a probability model. Although different probability models can be induced by P_A, the established practice is to stick to the original model of $(\Omega, \mathcal{F}, \mathrm{P})$.

References

V.S. Borkar, V.R. Konda and S.K. Mitter. 2004. On de Finetti coherence and Kolmogorov probability. Statistics and Probability Letters, Vol.66: 417-421.

R. Carnap. 1962. Logical Foundations of Probability, 2nd edition published by The University of Chicago Press.

D.M. Cifarelli and E. Regazzini. 1996. De Finetti's contribution to probability and statistics. Statistical Science, Vol.11, No.4: 243-282.

A. Hajek. 2008. Probability - A Philosophical Overview, in B. Gold and R. Simons (ed.), Proof and Other Dilemmas: Mathematics and Philosophy, published by Mathematical Association of America: 323-339.

J. Marschak. 1975. Personal probabilities of probabilities. Theory and Decisions, Vol.6: 121-153.

R.H. Thaler. 1999. The end of behavioral finance. Financial Analysts Journal, Vol.55, No.6: 12-17.