Module 1

Events and probability

Topics: Random experiment, probability, conditional probability, independence of events

Random experiment

We refer to an experiment, whose outcome cannot be predicted with certainty, as a random experiment. Here the word experiment is used in a broad sense. Consider tossing of a coin. We cannot predict its outcome with certainty. Consider the time you take to read this note. Again, we cannot predict it with certainty. Sometimes, uncertainty of an experiment can be due to the lack of knowledge, instead of the inherent uncertainty. Consider my birthday – it's uncertain for you, but not for my family and friends.

Sample space of a random experiment is defined as the set of all outcomes of the experiment. Sample space of the coin tossing experiment is $\{\text{head}, \text{tail}\}$. For the reading experiment, it's $(0, \infty)$. For the birthday experiment, it's the set of all dates in a year. *Consider some random experiments and try to identify their sample spaces*.

Depending on countability, we classify sample spaces into (1) countable and (2) uncountable sample spaces. Countable sample spaces are further classified into (1a) finite and (1b) infinite sample spaces. Similarly, uncountable sample spaces are classified into (2a) bounded and (2b) unbounded sample spaces. Here the word bounded is related to the notion of size, e.g., length in $\mathbb{R}$, area in $\mathbb{R}^2$, etc. If the size of a set is finite, we call it a bounded set, otherwise it's unbounded. Measurement of probability depends on the type of sample space.

Sample space of the coin tossing experiment is countable and finite. Consider tossing of a coin until head appears for the first time. It's sample space $\{\text{H}, \text{TH}, \text{TTH}, \dots\}$ is countable but infinite. Sample space of the reading experiment is uncountable and unbounded. Consider the time you take to write a 3-hour exam. It's sample space $[0,3]$ is uncountable but bounded.

In random experiments, our main interest is in understanding the chances of occurrence of a single or a group of outcomes. For example, in the coin tossing experiment we may be interested in the chances of occurrence of head, which is a single outcome of the experiment. In the reading experiment we may be interested in the chances of occurrence of the reading time being at most an hour, which is a group of several outcomes of the experiment. We refer to a single or a group of outcomes of a random experiment as events. So, events are subsets of the sample space. We say that an event takes place when any of its outcomes realizes.

Since events are sets, relations and operations involving events are same as those involving sets. Some basic relations are: (i) equality, (ii) subset/superset, (iii) proper subset/superset. Some basic operations are: (i) union, (ii) intersection, (iii) complementation (with respect to the sample space). The set difference operation can be expressed using the basic operations as: $\text{A} \setminus \text{B} = \text{A} \cap \text{B}^\text{c}$. Some useful results involving sets are: (i) $(\text{A} \cup \text{B})^\text{c} = \text{A}^\text{c} \cap \text{B}^\text{c}$, (ii)

$(A \cap B)^c = A^c \cup B^c$, (iii) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, (iv) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$. *If you do not recall these, please do some reading.*

Among events associated with a random experiment, disjoint events have a special place in probability theory. We call two events to be disjoint or mutually exclusive if they have no outcome in common. It is evident that two disjoint events cannot take place simultaneously. In general, we call a group of events to be disjoint if every sub-group of events of size two or more have no outcome in common. It can be proved that if every pair of events in a group are disjoint, then the group of events are disjoint. *Try to prove this yourself.* We refer to a group of disjoint events as partition if their union is the sample space itself. A partition is also known as mutually exclusive and collectively exhaustive events.

Consider tossing of a coin twice. Sample space of this random experiment can be written as $\{HH, HT, TH, TT\}$, where 1st element represents outcome of the first toss and the 2nd element represents outcome of the second toss. Consider the events 'same outcome' and 'one head', i.e., $\{HH, TT\}$ and $\{HT, TH\}$. These are disjoint events, and they constitute a partition of the sample space. Consider the events 'different outcomes' and 'two tails', i.e., $\{HT, TH\}$ and $\{TT\}$. These are disjoint, but they do not constitute a partition. Consider 'same outcome' and 'at least a head', i.e., $\{HH, TT\}$ and $\{HH, HT, TH\}$. These are not mutually exclusive, but they are collectively exhaustive. Finally, consider 'different outcomes' and 'at least one tail', i.e., $\{HT, TH\}$ and $\{HT, TH, TT\}$. They are neither mutually exclusive nor collectively exhaustive.

The above example also tells that an event described in a natural language can be represented as a subset of the sample space. If you are interested in obtaining probability of some event of a random experiment, it's a good idea to specify the sample space first and then express the event as a subset of the sample space. *Consider 4 chairs by the side of a circular dining table, and four members of a family, say F, M, S, D, randomly choose a chair each to sit. Represent the sample space of this random experiment and specify the event that F and M sit next to one another*. Next, we figure out an appropriate probability measure for the experiment, described in the next section, and calculate the probability.

Probability

Probability is a measure for the chances of occurrence of events associated with a random experiment. It can be viewed as a function defined on the set of events. Let Ω denote the sample space. Then the set of events (i.e., the set of subsets of Ω) is the power set of Ω, which is denoted by 2^{Ω}. *Try to construct power set of some of the sample spaces we encountered so far*. By convention, we assign a value between 0 to 1 as the probability of an event, with 0 representing impossibility (i.e., the event never occurs) and 1 representing certainty (i.e., the event always occurs). A probability value in $(0,1)$ implies possibility as well as uncertainty (i.e., the event occurs sometimes, and sometimes it does not). So, the probability function is of the form $P: 2^{\Omega} \to [0,1]$. *If these notations look unfamiliar, do some readings*.

Any arbitrary function from 2^Ω to $[0,1]$ does not qualify as probability function, as it may violate some universally accepted statements. $P: 2^\Omega \to [0,1]$ is a probability measure for Ω if it satisfies three axioms: (i) $P(A) \geq 0$ for all $A \subseteq \Omega$, (ii) $P(\Omega) = 1$, and (iii) $P(A \cup B) = P(A) + P(B)$ whenever A and B are disjoint. The first axiom is implicit in the structure of the probability function itself. The third axiom can also be written as: $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ for every collection of disjoint events $A_1, A_2, \dots, A_n$. *Try to think of something universally accepted that would be violated if we relax any of the three axioms*.

An immediate consequence of the probability axioms is: $P(A^c) = 1 - P(A)$ for all $A \subseteq \Omega$. This is because A and A^c are disjoint events and their union is Ω. Then by 2nd and 3rd axioms, $P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) = 1$, implying $P(A^c) = 1 - P(A)$. If we consider $A = \Omega$, then $A^c = \emptyset$, the empty set, and $P(\emptyset) = 1 - P(\Omega) = 0$. Also, $P(A) = 1 - P(A^c) \leq 1$ for all $A \subseteq \Omega$. This, like the first axiom, is implicit in the structure of the probability function. If we redefine the probability function as $P: 2^\Omega \to \mathbb{R}$, even then the axioms will ensure that the range of the probability function is $[0,1]$.

Let us note some more consequences of the probability axioms. If $A \subset B$, then by the third axiom, $P(B) = P\big(A \cup (B \setminus A)\big) = P(A) + P(B \setminus A) \geq P(A)$. Next, for any events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. By the third axiom, $P(A) = P(A \setminus B) + P(A \cap B)$ and $P(B) = P(B \setminus A) + P(A \cap B)$. Then $P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$. Using this result, for any three events A, B, C, one can show that $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$. *Try to prove this yourself by considering $A \cup B$ as a single event*.

Let us try to construct a probability measure for the random experiment – rolling of a die. Here, $\Omega = \{1,2,3,4,5,6\}$. Let $P_1(\{i\}) = 1/6$ for $i = 1,2,\dots,6$. Probabilities of other events can be found by the third axiom. One can verify that P_1 is consistent with the probability axioms. Now consider $P_2(\{i\}) = 2/9$ for $i = 1,3,5$ and $P_2(\{i\}) = 1/9$ for $i = 2,4,6$. Again, one can verify that P_2 is consistent with the probability axioms. We can construct endless number of valid probability measures, but only one of those must be 'appropriate', and the axiomatic definition of probability fails to identify it. Next, we introduce two specific approaches of measuring probability which are appropriate and consistent with the probability axioms. It shall be noted that the results that we obtain in this course are based on axiomatic definition alone. The specific approach that we employ to measure probability, whether it is appropriate or not, does not change anything if it is consistent with the probability axioms.

The first specific approach is called the classical measure. It is applicable when the outcomes of a random experiment are equally likely. Then it is appropriate to measure probability of an event as the ratio of the 'quantum' of outcomes that support the event to the total outcomes. If Ω is countable, then the quantum is same as count, and then $P(A) = |A|/|\Omega|$ for all $A \subseteq \Omega$, where $|\cdot|$ denotes count. If Ω is uncountable, then the quantum is same as size, i.e., length in one-dimension, area in two-dimension, volume in three-dimension, etc., and then $P(A) = S(A)/S(\Omega)$ for all $A \subseteq \Omega$, where $S(\cdot)$ denotes size. Consider the experiment of rolling a die. If it is a fair die, then the classical measure is appropriate, and we shall choose $\mathrm{P}_1(\{\mathrm{i}\}) = 1/6$ for $\mathrm{i} = 1,2, \dots, 6$ as the probability measure. *Consider the game of dart with a D cm diameter board and a $d < D$ cm diameter bullseye. Assume that a throw always lands on the board and all outcomes are equally likely. Find the probability of hitting the bullseye.*

In the definition of classical measure, we implicitly assumed that Ω is finite in the countable case and bounded in the uncountable case. Otherwise, the denominator in the definition of the classical measure, i.e., $|\Omega|$ or $\mathrm{S}(\Omega)$ is infinite. In that case, probability of every event having finite count or finite size has zero probability, and this leads to violation of the second axiom. For example, consider $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$ with equally likely outcomes. Then $\mathrm{P}(\{\omega_\mathrm{i}\}) = |\{\omega_\mathrm{i}\}|/|\Omega| = 1/\infty = 0$ for all i. By the third axiom, $\mathrm{P}(\Omega) = \mathrm{P}(\cup_{\mathrm{i}=1}^{\infty}\{\omega_\mathrm{i}\}) = \sum_{\mathrm{i}=1}^{\infty}\mathrm{P}(\{\omega_\mathrm{i}\}) = \sum_{\mathrm{i}=1}^{\infty} 0 = 0$, which goes against the second axiom. Now, consider $\Omega = [0, \infty)$ with equally likely outcomes. Then Then $\mathrm{P}\big([\mathrm{i}, \mathrm{i}+1)\big) = \mathrm{S}\big([\mathrm{i}, \mathrm{i}+1)\big)/\mathrm{S}(\Omega) = 1/\infty = 0$ for $\mathrm{i} = 0,1,2, \dots$. By the third axiom, $\mathrm{P}(\Omega) = \mathrm{P}(\cup_{\mathrm{i}=0}^{\infty}[\mathrm{i}, \mathrm{i}+1)) = \sum_{\mathrm{i}=0}^{\infty}\mathrm{P}\big([\mathrm{i}, \mathrm{i}+1)\big) = \sum_{\mathrm{i}=0}^{\infty} 0 = 0$, which goes against the second axiom. Therefore, if I ask you to pick a number from $\mathbb{N}$ or $\mathbb{Z}$ or $\mathbb{R}$ or some such infinite/unbounded set, you will never be able to do it such that all numbers are equally likely to be picked and the probability axioms are satisfied.

The above discussion seems to suggest that the classical measure satisfies probability axioms if the sample space is finite in the countable case and bounded in the uncountable case. Let us verify this formally. Since $|A|$ or $S(A)$ is non-negative for all $A \subseteq \Omega$ and $|\Omega|$ or $S(\Omega)$ is a positive quantity, then $P(A) = |\mathrm{A}|/|\Omega|$ or $\mathrm{S(A)}/\mathrm{S}(\Omega)$ is non-negative, as claimed in the first axiom. Next, $P(\Omega) = |\Omega|/|\Omega|$ or $\mathrm{S}(\Omega)/\mathrm{S}(\Omega)$ is 1, as claimed in the second axiom. Consider disjoint events A and B. Then $|A \cup B| = |A| + |B|$ or $S(A \cup B) = S(A) + S(B)$, and thus, $P(A \cup B) = |A \cup B|/|\Omega| = |A|/|\Omega| + |B|/|\Omega| = P(A) + P(B)$, or $P(A \cup B) = \mathrm{S}(A \cup B)/S(\Omega)$ $= \mathrm{S(A)}/S(\Omega) + \mathrm{S(B)}/S(\Omega) = P(A) + P(B)$, as claimed in the third axiom.

In the classical way of measuring probability, one must calculate $|A|$ or $S(A)$. While the later is quite straightforward, the former may be challenging at times. Finite countable sample spaces are generally represented by permutations, i.e., collection of objects where internal arrangement matters, and combinations, i.e., collection of objects where internal arrangement does not matter. In such cases, standard formulas are useful in obtaining $|A|$. We list down four such formulas: (i) number of permutations using r distinct items from n $(\geq r)$ distinct objects – $P(n, r) = n!/(n-r)!$, (ii) number of combinations using r distinct items from n $(\geq r)$ distinct objects – $C(n, r) = n!/\{r!\,(n-r)!\}$, (iii) number of permutations using r items from n distinct objects with repetitions – n^r, and (iv) number of combinations using r items from n distinct objects with repetitions – $C(n+r-1, r) = (n+r-1)!/\{r!\,(n-1)!\}$. These four formulas can be suitable amended to take care of many other situations. *If these looks unfamiliar, do some reading.*

Consider the game of cards, specifically bridge in which a pack of 52 cards is randomly distributed among four players, say A, B, C, D. There are four types of cards, namely clubs, diamonds, hearts, and spades. Each type has 13 cards labelled as Ace, 2, 3, …, 10, J, Q, K.

Let us determine the probability that each player gets one Ace. Here, the sample space can be viewed as partitioning of the pack of 52 cards into four groups of 13 cards (corresponding to the four players). The event of interest is the collection of those partitions where each of the four groups has one Ace. It is evident that all outcomes are equally likely and Ω is countable and finite. So, the classical measure of probability is applicable. Let us obtain $|\Omega|$ first. There are $C(52,13)$ different ways in which the first group of 13 cards can be chosen. Given the 1st group, there are $C(39,13)$ different ways in which the second group of 13 cards can be chosen. Given the 1st and 2nd groups, there are $C(26,13)$ different ways in which the third group of 13 cards can be chosen. Given the 1st, 2nd, and 3rd groups, the remaining cards go to the fourth group. So, $|\Omega| = C(52,13) \times C(39,13) \times C(26,13) = 52!/(13!)^4$. Now we find $|E|$, where E denotes the event of interest. First, let us distribute the 4 Aces into four groups. Following the above argument, this can be done in $C(4,1) \times C(3,1) \times C(2,1) = 4!$ different ways. After this, the remaining 48 cards are to be distributed into four groups, which can be done in $C(48,12) \times C(36,12) \times C(24,12) = 48!/(12!)^4$. Therefore, $|E| = (4!\,48!)/(12!)^4$, and $P(E) = |E|/|\Omega| = \{4!\,48!\,(13!)^4\}/\{(12!)^4 52!\} = (4!\,13^4)/(52 \times 51 \times 50 \times 49)$. *Now try to find the probability that one player gets all 4 Aces*.

Due to the obvious limitations of the classical measure, we need a probability measure that is applicable in every situation, i.e., for all types of random experiments. Such a measure is the frequentists' measure. If we conduct a random experiment n number of times, and the event A occurs in n_A occasions, then n_A/n provides an estimate for $P(A)$. If we make n large, we can expect n_A/n to get closer to $P(A)$ and eventually converge to it as $n \to \infty$. This 'expectation' is a fact, which we will prove later in this course. So, the frequentists' measure of probability is: $P(A) = \lim_{n\to\infty} n_A/n$ for all $A \subseteq \Omega$. *Verify consistency of the frequentists' measure with the axioms of probability*. Frequentists' measure also have its limitations. First, it is impossible to repeat an experiment infinite number of times. Second, some experiments may be 'costly' to repeat even finitely many times. Third, it may be practically impossible to perform some experiments. Then we use subjective measure of probability, which is nothing but expert opinion, consistent with the probability axioms.

Conditional probability

Consider tossing of a fair coin twice. Sample space of this random experiment can be written as: $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$. We are interested in finding probability of the event that both tosses are the same, i.e., $A = \{HH, TT\}$. Since the coin is fair, outcomes are equally likely and the classical measure applies. Then $P_\Omega(A) = |A|/|\Omega| = 2/4$. We are using Ω as subscript for P to indicate that the probability measure is for the sample space Ω. Now imagine that we have information that at least one toss is head. This information changes the sample space because the outcome TT is no more a possibility. The new sample space can be represented as: $B = \{HH, HT, TH\}$. Observe that B is the event of Ω that at least one toss is head. Outcomes of B are equally likely, as the outcomes of the original sample space were equally likely. So, the classical measure continues to apply. Let us calculate the probability of A in the presence of B. Earlier, A could occur through the outcomes HH and TT, but now it can occur through HH alone. In general, an arbitrary A in the presence of an arbitrary B can occur through $A \cap B$. Therefore, the probability of A in the presence of B is: $P_{\mathrm{B}}(A) = |A \cap B|/|\mathrm{B}| = 1/3$. Observe that the probability has changed.

Let us find a generalized method of obtaining probability of A in the presence of B. Let us consider the case of equally likely outcomes first. Then the classical measure applies to Ω and B both. Following the above discussion, $P_{\mathrm{B}}(A) = |A \cap B|/|\mathrm{B}| = (|A \cap B|/|\Omega|)/(|\mathrm{B}|/|\Omega|) = P_{\Omega}(A \cap B)/P_{\Omega}(B)$. So, we can obtain $P_{\mathrm{B}}(\cdot)$ from $P_{\Omega}(\cdot)$ itself. Now let us consider the case of frequentists' measure. Let n denote the number of repetitions of the random experiment Ω, out of which A occurs in n_A occasions, B occurs in n_B occasions, and $A \cap B$ occurs in $n_{A\cap B}$ occasions. In the presence of B, all outcomes in Ω \ B are discarded, and then A occurs $n_{A\cap B}$ times. Then $P_B(A) = \lim_{n_B \to \infty} n_{A\cap B}/n_B = \lim_{n\to\infty} (n_{A\cap B}/n)/(n_B/n) = P_{\Omega}(A \cap B)/P_{\Omega}(B)$. We get the same expression. Note that $n_B \to \infty \Rightarrow n \to \infty$.

With the above observations, we can calculate probability of A in the presence of B for an arbitrary Ω as: $P_{\mathrm{B}}(A) = P_{\Omega}(A \cap B)/P_{\Omega}(B)$. Since there is no need to get into the new sample space B, we can omit the subscripts from probability function. We denote $P_{\mathrm{B}}(A)$ by $P(A|B)$, meaning probability of A given that B has happened. Let us also omit ∩ and write $A \cap B$ as AB. Then $P(A|B) = P(AB)/P(B)$. Consequences of probability axioms that we noted earlier are applicable in the context of conditional probability. *Show that* $P(A^c|B) = 1 - P(A|B)$. Sometimes, we mistakenly equate $P(A|B^c)$ and $1 - P(A|B)$.

Many interesting questions can be formed with conditional probability. *Consider a fair coin is tossed. If it shows head, the experiment ends, but if it shows tail, the coin is tossed once more and the experiment ends. Given that the outcome of this experiment is head, can you find the probability that the coin was tossed twice*? Conditional probability can be used to determine probability of intersection of two or more events. Since $P(A|B) = P(AB)/P(B)$, we have $P(AB) = P(A|B)P(B)$. We can generalize this formula to calculate probability of intersection arbitrary events as: $P(A_1A_2 \dots A_n) = P(A_n|A_1A_2 \dots A_{n-1})P(A_{n-1}|A_1A_2 \dots A_{n-2}) \cdots P(A_2|A_1)P(A_1)$. Consider this class has $n < 365$ students. Let us calculate the probability that the students have different birthdays. Let A_i denote the event that the i-th student have a birthday different from that of the first $i - 1$ students, for $i = 1,2, \dots, n$. Then $A_1A_2 \dots A_n$ is the event that the students have different birthdays, and its probability can be calculated as: $P(A_n|A_1A_2 \dots A_{n-1})P(A_{n-1}|A_1A_2 \dots A_{n-2}) \cdots P(A_2|A_1)P(A_1) = (1 - (n-1)/365) \times (1 - (n-2)/365) \times \cdots \times (1 - 1/365) \times 1$. Conditional probabilities are obtained by applying the classical measure to $A_1A_2 \dots A_i$ for $i = 1,2, \dots, n-1$.

Another utility of conditional probability comes in association with partition of the sample space. Let $B_1, B_2, \dots, B_n$ denote a partition of Ω. Then by the 3rd axiom, $P(A) = P(\cup_{i=1}^{n} AB_i) = \sum_{i=1}^{n} P(AB_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$. This formula is known as the law of total probability. Consider 10% people in a population are covid infected. The covid detection test is not fully reliable. Consider the probability of false positive (i.e., detecting virus when there is none) to be 0.02 and the probability of false negative (i.e., not detecting the virus which is there) is

0.05. Let us determine the probability of positive test result for a randomly selected person from this population. Let A denote the event that the test result is positive and B denote the event that the person is covid infected. Then the given information can be written as: $P(B) = 0.1$, $P(A|B^c) = 0.02$, $P(A^c|B) = 0.05$, and we want to obtain $P(A)$. Since B, B^c partitions the sampler space (*What is the sample space we are talking about?*), $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = (1 - 0.05) \times 0.1 + 0.02 \times (1 - 0.1) = 0.113$.

In the context of the above problem, we may be interested in $P(B|A)$, i.e., the probability of a person having covid infection given that his test result is positive. In general, our interest may be in determining $P(B|A)$ while conditional probabilities of the other form, i.e., $P(A|B)$ and $P(A|B^c)$ are known. Using the previously developed formulas, we obtain the following:

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

It is known as the Bayes' formula. With this formula, the probability of a person having covid infection given that his test result is positive, $P(B|A) = \{(1 - 0.05) \times 0.1\}/0.113 = 0.84$. *Now, determine the probability of a person having covid infection given that his test result is negative*. Bayes' theorem is at the heart of Bayesian analysis, which aims to refine subjective probability using evidence as and when they appear.

Independence of events

Let us consider tossing of a fair coin twice. Its sample space is: $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$ and the classical measure applies. We are interested in the event that both tosses are the same, i.e., $A = \{HH, TT\}$. Clearly, $P(A) = 1/2$. Assume that we have information that at least one toss is head, i.e., $B = \{HH, HT, TH\}$ has occurred. Then $P(A|B) = P(AB)/P(B) = 0.25/0.75 = 1/3$. Note that $P(A)$ and $P(A|B)$ are different. The information that B has occurred changed the probability of A. However, this is not the always. Consider that we have information that the first toss is head, i.e., $B' = \{HH, HT\}$ has occurred. Then $P(A|B') = P(AB')/P(B') = 0.25/0.50 = 1/2$, which is same as $P(A)$. So, the information that first toss is head does not change the probability of the event that both tosses are the same. We capture this invariance of probability with respect to information through independence of events.

We call two events A and B to be independent if occurrence/non-occurrence of one does not change probability of occurrence/non-occurrence of the other, i.e., A and B are independent if (i) $P(A|B) = P(A)$, (ii) $P(A^c|B) = P(A^c)$, (iii) $P(A|B^c) = P(A)$, (iv) $P(A^c|B^c) = P(A^c)$, and four more conditions by interchanging positions of A and B. All these conditions are equivalent and same as: $P(AB) = P(A)P(B)$. Let us check the first one. $P(A|B) = P(A) \Leftrightarrow P(AB)/P(B) = P(A) \Leftrightarrow P(AB) = P(A)P(B)$, as required. *Verify this equivalence for some of the other conditions*. So, A and B are independent if $P(AB) = P(A)P(B)$.

We can extend the idea of independence to multiple events. We call a collection of events $A_1, A_2, \ldots, A_n$ to be independent if the intersections of events in every pair of non-overlapping sub-collections of $A_1, A_2, \dots, A_n$ are independent. So, three events A_1, A_2, A_3 are independent if (i) A_1, A_2 are independent, (ii) A_1, A_3 are independent, (iii) A_2, A_3 are independent, (iv) A_1, A_2A_3 are independent, (v) A_2, A_1A_3 are independent, and (vi) A_3 and A_1A_2 are independent. The first three requirements are equivalent to: (a) $P(A_1A_2) = P(A_1)P(A_2)$, (b) $P(A_1A_3) = P(A_1)P(A_3)$, and (c) $P(A_2A_3) = P(A_2)P(A_3)$. Given (a-c), the last three requirements for independence are same as: (d) $P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$. *Consider tossing of a fair coin thrice. Let A_1, A_2, A_3 respectively denote the events that 1st, 2nd, 3rd toss is head. Are these events independent*? In general, $A_1, A_2, \dots, A_n$ are independent if $P\left(A_{i_1}A_{i_2} \dots A_{i_k}\right) = P\left(A_{i_1}\right)P\left(A_{i_2}\right) \cdots P\left(A_{i_k}\right)$ for all $1 \le i_1 < i_2 < \cdots < i_k \le n$ and $k = 2,3, \dots, n$. This translates into a total of $C(n,2) + C(n,3) + \cdots C(n,n) = 2^n - n - 1$ conditions.

Earlier, we defined disjoint events, first for $n = 2$ and then for general n. There we observed that pairwise disjoint-ness implies collective disjoint-ness. Here, pairwise independence does not imply collective independence. For example, consider tossing of a fair coin twice, and define $A_1 = \{HH, HT\}$, $A_2 = \{HH, TH\}$, and $A_3 = \{HH, TT\}$. Observe that these events are pairwise independent, but $P(A_1A_2A_3) \ne P(A_1)P(A_2)P(A_3)$, implying that A_1, A_2, A_3 are not independent. Another example can be constructed where $P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$, but at least of the pairwise independence conditions does not hold. *Using the experiment of tossing a fair coin thrice, construct an example of this kind*. Hence, for independence of n events, we must check each of the $2^n - n - 1$ conditions.

Before we conclude this module, we shall distinguish between disjoint and independent events. If two events are disjoint and one of them occurs, then the other cannot occur. On the other hand, if two events are independent, one does not influence the other. Hence, two (non-null) events cannot be disjoint and independent simultaneously. One can easily prove this by contradiction. For example, consider two non-null events A, B to be disjoint. Then $P(AB) = P(\emptyset) = 0$. For independence, we need $P(A)P(B) = P(AB)$, which is zero. It is possible only if at least one of $P(A)$, $P(B)$ is zero, which is impossible for non-null events (*Why?*). Thus, two non-null events cannot be disjoint and independent simultaneously.

Practice problems

Book-1: A Modern Introduction to Probability and Statistics by Dekking et al.
Book-2: Introduction to Probability Theory and Statistical Inference by Larson

Events and probability
Book-1, Chapter-2, Exercise No. 2, 4, 7, 9, 14, 16, 18, 19

Counting techniques
Book-2, Chapter-2, Section-5, Exercise No. 6, 8, 10, 13

Conditional probability
Book-1, Chapter-3, Exercise No. 2, 3, 5, 6, 8, 11, 12, 15, 16, 17

Define

A: she identifies at least 6 cards correctly

B: she identifies exactly 6 cards correctly

C: she identifies all 8 cards correctly.

Since she will call 4 cards red and 4 cards black, it is not possible for her to be correct on exactly 7 cards; thus,

$$A = B \cup C$$

and, since

$$B \cap C = \varnothing,$$

$$P(A) = P(B) + P(C).$$

Clearly, $n(C) = 1$ so $P(C) = \frac{1}{70}$. If B is to occur, she must identify exactly 3 of the 4 red cards correctly and exactly 3 of the 4 black cards correctly. The 1 red card that she is wrong on could be any of the 4 and the 1 black card she is wrong on could be any one of the 4. Thus, the number of 8-tuples having 1 B in the first 4 positions and 1 R in the last 4 positions is

$$n(B) = 4 \cdot 4 = 16$$

and we have

$$P(B) = \tfrac{16}{70}.$$

Thus

$$P(A) = \tfrac{1}{70} + \tfrac{16}{70} = \tfrac{17}{70} \doteq .243;$$

if she only guesses, there is slightly less than 1 chance in 4 of her doing as well as she claims.

EXERCISE 2.5.

1. In how many ways could a dozen oranges be chosen from a table holding 30 oranges? (How many distinct collections of 12 oranges could be made?)

2. How many arrangements could be made of 5 red balls and 1 orange ball?

3. In how many of the arrangements counted in problem 2 are the red balls all together?

4. A certain market uses red boxes and green boxes for displays at Christmas time. In how many ways could the market arrange 20 boxes in a row if 15 of them are red and 5 are green? If there are 10 boxes of each color?

5. Ten people in total are nominated for a slate of 3 offices. If every group of 3 people has the same probability of winning, what is the probability that a particular person will be on the winning slate? That a particular pair of people will be on the winning slate?

6. Two people are to be selected at random to be set free from a prison with a population of 100. What is the probability that the oldest prisoner is 1 of the 2 selected? That the oldest and youngest are the pair selected?

7. In a certain national election year, governors were to be elected in 30 states. Assume that in every state there were only 2 candidates (called the Republican and Democratic candidates, respectively). What is the probability that the Republicans carried all 30 states, assuming that each state was equally likely to elect either party? What is the probability that the same party carried all the states?

8. A, B, and C are going to race. What is the probability that A will finish ahead of C, given that all are of equal ability (and no ties can occur)? What is the probability that A will finish ahead of both B and C?

9. Each of 5 people is asked to distinguish between vanilla ice cream and French vanilla custard (each is given a small sample of both and asked to identify which is ice cream). If all 5 people are guessing, what is the probability that all will correctly identify the ice cream? If all 5 are guessing, what is the probability that at least 4 will identify the ice cream correctly?

10. Compute the probability that a group of 5 cards drawn at random from a 52-card deck will contain

(a) exactly 2 pair
(b) a full house (3 of one denomination and 2 of another)
(c) a flush (all 5 from the same suit)
(d) a straight (5 in sequence, beginning with ace or deuce or trey, . . . , or ten).

11. n people are in a room. Compute the probability that at least 2 have the same birth month. Evaluate this probability for $n = 3, 4, 5, 6$.

12. A student is given a true-false examination with 10 questions. If he gets 8 or more correct, he passes. If he is guessing, what is his probability of passing the examination?

13. Eight black and 2 red balls are randomly laid out in a row. What is the probability that the 2 red balls are side by side? That the 2 red balls are occupying the end positions?

14. A person is to be presented with 3 red and 3 white cards in a random sequence. He knows that there will be 3 of each color; thus he will identify 3 cards as being of each color. If he is guessing, what is his probability of correctly identifying all 6 cards? Of identifying exactly 5 correctly? Exactly 4?

2.6. Conditional Probability

In some applications we shall be given the information that an event A occurred and will be asked the probability that another event B also occurred. For example, we might be given that the card we selected from a regular 52-card bridge deck was red and then might want to know the probability that the card selected was the ace of hearts. Or, when running an opinion poll, we might be given the fact that the person we have selected is a Republican and then might also want to know the probability that he favors our