Probability Theory

Axioms

Axioms are mathematical statements that are accepted without proof. The theory of probability is founded on three axioms from which all other rules are derived. The Russian mathematician Kolmogorov formulated the axioms in a monograph in 1933. An English translation is available in the Foundations of Probability Theory, published by Chelsea in New York in 1950. The axioms of probability states that a probability is between zero and unity and that the probability of mutually exclusive events is additive. Specifically, by denoting the probability of an event E as P(E), the first two axioms read:

$$P(E) \geq 0 \tag{1}$$

$$P(S) = 1 \tag{2}$$

where S is the certain event. The third axiom provides the probability of the union of mutually exclusive events:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) \tag{3}$$

Objectivists (Frequentists) Vs. Subjectivists (Bayesians)

There are two schools of thought in probabilistic analysis. One considers the probability of an event as the relative frequency of occurrence of that event in repeated trials:

$$\mathrm{P}(E) = \lim_{n \to \infty} \frac{n_E}{n} \tag{4}$$

where n_E is the number of occurrences of the event E in n trials. According to this school of thought, which is often referred to as classical statistics, the probability is entirely objective and empirical, i.e., based on observations. In contrast, the Bayesian school of thought is more appropriate in engineering applications where repeated trials are impractical, e.g., structural reliability analysis. Bayesian statistics encompass classical statistics but allow the incorporation of subjective information. As a result the probability is interpreted as a "degree of belief." Importantly, this approach means that the probability is subjective and influenced by the knowledge of the engineer who assigns it.

Chance and Odds

There are other ways of presenting a probability than as a number between zero and one. One option is to multiply the probability by 100 and call it a percent "chance." In other words, a probability equal to 0.5 means that there is a 50% chance. Another option is to work with odds, which expresses a probability as the odds "n to m." Here, n is the expected number of outcomes "for" the event and m is the expected number of outcomes "against" the event. For example, when throwing a die there is a "1 to 5" odds of getting, say, a 6. In this notation the relationship between probability and odds is

$$\mathrm{P}(E)=\frac{n}{n+m} \tag{5}$$

Often m is set equal to unity, so that odds are expressed as “n to one.” By solving for n in Eq. (5), the following relationship between odds and probability appears:

$$n=\frac{P(E)}{1-P(E)} \tag{6}$$

For the example of throwing a dice, this means that the odds are “0.2 to one” in favour of rolling a 6. However, fractions like 0.2 are somewhat awkward. Therefore, it is more common to express the odds of the most likely event; the odds are “five to one” against rolling a 6. There is yet another way of expressing odds. When stating that the odds are “*a* in *b*” then *b* is the total number of outcomes. For example, there is a “1 in 6” odds of getting a 6 when rolling a die. In this case, the relationship between probability and odds is simply

$$\mathrm{P}(E)=\frac{a}{b} \tag{7}$$

The expressions “long odds” and “short odds” are sometimes encountered, particularly in the context of gambling. These are imprecise terms; essentially long odds imply that the probability is very low. Conversely, short odds imply high probability. When using odds to express probability one must avoid using the expression “*x* out of *n* times.” That expression introduces the notion of a series of experiments, i.e., a Bernoulli sequence, which is addressed in another document on this website. For example, saying success will occur “9 out of 10 times” does not mean the probability of occurrence is 0.9 in each trial. In fact, it is unclear whether the event is exactly 9 out of 10 times, or also 10 out of 10 times.

Typical Probabilities

In typical civil engineering applications, such as structural safety where failure is locally dramatic but rather harmless in a wider area, the target probability of failure is around 10^{-4} for the lifetime of the facility. The target failure probability for facilities that are associated with more dramatic failure consequences, such as nuclear power plants, is lower. In comparison, the probability of drawing an ace of spade in a randomized deck of cards is $1/52=19 \cdot 10^{-3}$, i.e., much higher. As a rough rule of thumb, society requires immediate action for hazards with a probability of death per person per year greater than 10^{-3}, while probabilities less than 10^{-6} are associated with events that are so unusual that little can reasonably be done. In comparison, the probability of death per hour for one person flying is around $1.2 \cdot 10^{-8}$, while the annual probability per person, accounting for an average exposure time is $24 \cdot 10^{-6}$. In contrast, the probability death per hour for a person travelling by car is less ($0.7 \cdot 10^{-8}$), but given the larger amount of time we spend in cars the average exposure time is higher and the probability of death per person per year is $200 \cdot 10^{-6}$, i.e., about ten times that of air travel.

Probability Rules

The rules of probability are derived from the axioms. As for the third axiom, visualization by Venn diagrams is often helpful for the understanding the validity of the rules. With the suite of probability rules that is summarized below it is possible to analyze a variety of probabilistic problems. In other words, it is possible to determine the probability of a variety of events, given input probabilities.

Probability of the Complement

Consider two complementary events that are mutually exclusive and collectively exhaustive. The probability of the complement is

$$P(\bar{E}) = 1 - P(E) \tag{8}$$

This rule is derived by combining the second and third axioms of probability:

$$P(S) = P(E \cup \bar{E}) = P(E) + P(\bar{E}) = 1 \quad \Rightarrow \quad P(\bar{E}) = 1 - P(E) \tag{9}$$

Union Rule

This rule provides the probability of the union of two events that are NOT known to be mutually exclusive and collectively exhaustive:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 E_2) \tag{10}$$

This rule is justified by visualizing the two events in a Venn diagram. Unless the last term in Eq. (10) is subtracted, i.e., the intersection event, it will be "counted twice" in the sum $P(E_1)+P(E_2)$.

Inclusion-exclusion Rule

This rule is a generalization of the union rule to problems with more than two events:

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) = &P(E_1) + P(E_2) + P(E_3) \\ &- P(E_1 E_2) - P(E_1 E_3) - P(E_2 E_3) \\ &+ P(E_1 E_2 E_3) \end{aligned} \tag{11}$$

Conditional Probability Rule

A conditional probability is written with a vertical bar. The probability $P(E_1|E_2)$ is read "the probability of E_1 given that E_2 has occurred," or "the probability of E_1 given E_2." The conditional probability rule states that

$$P(E_1 | E_2) = \frac{P(E_1 E_2)}{P(E_2)} \tag{12}$$

This rule can be justified in several ways. One approach is to think of n repeated experiments in which the events E_1 and/or E_2 may occur. Suppose n_2 is the number of times that E_2 occurs, while n_{12} is the number of times E_1 and E_2 occur simultaneously. Then, employing the frequency notion of probability and letting n become large, the fraction in Eq. (12) is expanded as follows:

$$\frac{P(E_1E_2)}{P(E_2)} = \frac{\left(\frac{n_{12}}{n}\right)}{\left(\frac{n_2}{n}\right)} = \frac{n_{12}}{n_2} \tag{13}$$

Now, according to the frequency notion of probability the last term in Eq. (13) is a probability in its own right. It is the number of realizations of E_1E_2 over the total number of realizations of E_2. In other words, the sample space for this probability is E_2. This probability is written $P(E_1|E_2)$, which corroborates Eq. (12).

Multiplication Rule

The multiplication rule is a direct consequence of the conditional probability rule:

$$P(E_1E_2) = P(E_1 \mid E_2)P(E_2) \tag{14}$$

Bayes' Rule

This rule is employed to update the probability of an event, given some information. Suppose E_2 is an event that is known to have occurred, Bayes' rule yields the probability of some event E_1 in light of this information:

$$P(E_1 \mid E_2) = \frac{P(E_2 \mid E_1)}{P(E_2)} P(E_1) \tag{15}$$

This rule is derived by starting with the conditional probability rule and adding the multiplication rule:

$$P(E_1 \mid E_2) = \frac{P(E_1E_2)}{P(E_2)} = \frac{P(E_2 \mid E_1) \cdot P(E_1)}{P(E_2)} \tag{16}$$

This rule is the foundation for an entire discipline within statistics, which employs a particular terminology. In particular, $P(E_1)$ is called the "prior" probability because it is the probability for E_1 before the new information becomes available. Conversely, E_2 is called the "posterior" probability. The probability $P(E_2|E_1)$ is called the "likelihood" and serves a particularly central role in Bayesian updating. It is the probability of observing what was observed, which requires some form of initially assumed underlying model. Finally, the probability $P(E_2)$ generally serves a "normalizing" purpose; in practical applications it often requires the use of the rule of total probability.

Rule of Total Probability

This rule is extensively used in many modern engineering applications. It provides the probability of some event when we know only conditional probabilities for:

$$P(A) = \sum_{i=1}^{n} P(A \mid E_i)P(E_i) \tag{17}$$

Importantly, the event A must be conditioned upon events, E_i, that are mutually exclusive and collectively exhaustive. The rule is derived as follows:

$$
\begin{aligned}
P(A) &= P(AS) \\
&= P\left(A(E_1 \cup E_2 \cup \cdots \cup E_n)\right) \\
&= P(AE_1 \cup AE_2 \cup \cdots \cup AE_n) \\
&= \sum_{i=1}^{n} P(AE_i) \\
&= \sum_{i=1}^{n} P(A \mid E_i) P(E_i)
\end{aligned}
\tag{18}
$$

where the second-last equality invokes the inclusion-exclusion rule for mutually exclusive events, and the last equality invokes the multiplication rule.

Statistical Dependence

Two events are said to be statistically independent if $P(E_1 \mid E_2) = P(E_1)$. A consequence of statistical independence is that $P(E_1 E_2) = P(E_1)P(E_2)$.