Probability Theory

A great reference for the material outlined in this document, and many of the other documents on probabilistic analysis and reliability methods, is the textbook by Der Kiureghian (2022). Another reference for this document is the textbook by Ang & Tang, which contains lots of examples and practice problems (Ang & Tang 2007).

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 state that a probability is between zero and unity and that the probability of the union 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 "chance," measured in percent. In other words, a probability equal to 0.1667 means that there is a 16.67% chance of the event occurring. 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 "against" the event and m is the expected number of outcomes "for" the event. In this notation the relationship between probability and odds is

$$1 - P(\text{event}) = \frac{n}{n+m} \tag{5}$$

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

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

This means that the odds of an event, let's say rolling a four, [: :], with a die, are "five to one." The previous equations correctly say that the corresponding probability is 0.1667. The expression "long odds" means the probability is low, i.e., that n is large compared with m. Conversely, "short odds" means the probability is high, i.e., that n is low compared with m.

Avoid "x in y" and "x out of y"

Notice the word "to" when a probability is expressed as "the odds are n to m." That is the proper terminology. One should avoid saying "x out of y times" and "the probability is x in y." The reason is that those expressions imply a series of trials, in which the event in question occurs x out of y times. That is very different from the probability that x occurs in a single trial. Again, take the probability of rolling a four, [: :], with a die as an example. Saying that a four will appear 1 out of 6 times or that the probability is 1 in 6 implies that we can think of an experiment with 6 trials, i.e., 6 rolls with the die. The probability of a four in each roll is 0.1667. However, any series of trials is a "Bernoulli sequence," described in another document on this website. The binomial distribution from that document gives the probability of a four, [: :], occurring "1 out of 6 times" or "1 in 6" as

$$p(1) = \binom{6}{1} \cdot (0.1667)^1 \cdot (1 - 0.1667)^{6-1} \approx 0.40 \tag{7}$$

In words, there is a 40% chance of seeing one four, [: :], if you roll a die six times. In fact, the probability of seeing one, two, or any number of fours is more than 66%. Those values are very different from the 16.67% chance implied by the person who said that a four occurs "1 out of 6 times" or "1 in 6" while trying to say

$$\mathrm{P}([::]) = \frac{1}{6} \approx 0.1667 \tag{8}$$

This also means that the probabilities “8 in 10” and “4 in 5” are very different, although they seem equal at first glance. An additional comment, about the Poisson occurrence model, addressed in yet another document on this website, fits here. That model governs the occurrence of events along the continuous time axis. One might hear a Poisson process specified as “1 in 50.” What does that mean? Does it mean that there likely will be one earthquake in 50 years? No, that is too simplistic. It means the *rate* of occurrence per year is 1/50 and that the “return period” (defined in the aforementioned document) is 50 years. It also means that the probability of an earthquake in any given year is roughly 1/50=0.02, as long as the rate is low. It is still awkward to say “1 in 50” because it leads the thoughts to “1 earthquake in 50 years.” The probability that 1 earthquake will occur in 50 years is, using the Poisson distribution,

$$\mathrm{P}\left(\text{1 earthquake in 50 years}\right)=\frac{\left(\frac{1}{50}\cdot 50\right)^1}{1!}e^{-\frac{1}{50}\cdot 50}\approx 0.37 \tag{9}$$

The probability that any number of earthquakes will occur in 50 years is

$$\mathrm{P}\left(\text{One or more earthquakes}\right)=1-e^{-\frac{1}{50}\cdot 50}\approx 0.63 \tag{10}$$

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.

Rules of Probability

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

$$P(\overline{E}) = 1 - P(E) \tag{11}$$

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

$$P(S) = P(E \cup \overline{E}) = P(E) + P(\overline{E}) = 1 \quad \Rightarrow \quad P(\overline{E}) = 1 - P(E) \tag{12}$$

Union Rule

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

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 E_2) \tag{13}$$

This rule is justified by visualizing the two events in a Venn diagram. Unless the last term in Eq. (13) 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:

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) \\ - P(E_1E_2) - P(E_1E_3) - P(E_2E_3) \\ + P(E_1E_2E_3) \end{aligned} \tag{14}$$

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

$$P(E_1 \mid E_2) = \frac{P(E_1E_2)}{P(E_2)} \tag{15}$$

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. (15) is expanded as follows:

$$\frac{P(E_1E_2)}{P(E_2)} = \frac{\left(\dfrac{n_{12}}{n}\right)}{\left(\dfrac{n_2}{n}\right)} = \frac{n_{12}}{n_2} \tag{16}$$

Now, according to the frequency notion of probability the last term in Eq. (16) 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. (15).

Multiplication Rule

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

$$P(E_1E_2) = P(E_1 | E_2)P(E_2) \tag{17}$$

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:

$$P(E_1 | E_2) = \frac{P(E_2 | E_1)}{P(E_2)} P(E_1) \tag{18}$$

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

$$P(E_1 | E_2) = \frac{P(E_1E_2)}{P(E_2)} = \frac{P(E_2 | E_1) \cdot P(E_1)}{P(E_2)} \tag{19}$$

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 of E_1 before the new information becomes available. Conversely, $P(E_1|E_2)$ is called the "posterior" probability. The probability $P(E_2|E_1)$ is called the "likelihood" and plays a central role in Bayesian updating. It is the probability of observing what was observed, which requires a model that says something about the connection between E_1 and E_2. Finally, the probability $P(E_2)$ generally serves a "normalizing" purpose. This will become clearer when the rules of probability are applied to random variables. For now, it is said that the calculation of $P(E_2)$ often requires the use of the rule of total probability, which is addressed next.

Rule of Total Probability

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

$$P(A) = \sum_{i=1}^{n} P(A | E_i)P(E_i) \tag{20}$$

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

$$\begin{aligned} P(A) &= P(AS) \\ &= P\big(A(E_1 \cup E_2 \cup \cdots \cup E_n)\big) \\ &= P(AE_1 \cup AE_2 \cup \cdots \cup AE_n) \\ &= \sum_{i=1}^{n} P(AE_i) \\ &= \sum_{i=1}^{n} P(A | E_i)P(E_i) \end{aligned} \tag{21}$$

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 $P(E_1 | E_2) = P(E_1)$. A consequence of statistical independence is that $P(E_1 E_2) = P(E_1)P(E_2)$.

References

Ang, A. H.-S., & Tang, W. H. (2007). *Probability concepts in engineering: emphasis on applications in civil & environmental engineering*. Wiley.

Der Kiureghian, A. (2022). *Structural and System Reliability.* Cambridge University Press.