Chapter 2 Mathematics of Reliability & Safety Analysis

2.1 Elements of Probability Theory

The basic nature of failure data

Probability theory plays a central role in risk analysis. Failure data are basically of a *stochastic* nature, involving ever-changing components and environmental conditions. This means that even under perfect measurement conditions, failure data for a single event cannot be determined by a one-and-only test but rather must be evaluated from a great number of independent observations and describe as a statistical distribution rather than a single value.

The task of the risk engineer is to gather and select all the pertinent data - sometimes the most *probable*, sometimes the most *unfavorable* (according to the circumstances) - needed to tackle the problem considered, evaluate their quality and predict on this basis the expected (most probable) outcomes or consequences. Risk analysis thus involves a phase of statistical evaluation (often only in a rudimentary form, for lack of information) and a phase of probabilistic analysis ("what is the probability that a given value has to be taken into account?", "what is the occurrence probability of a given fracture mode?", etc.).

Statistics & probability

In the modern sense of these words, "statistics" is defined as:

> "the set of mathematical interpretation techniques applied to phenomena for which an exhaustive study of all involved factors is impossible because of their great number and/or complexity",

and "probability" as:

> "a measure of the degree of belief that an event (possible but not certain) will occur".

Different events may have different levels of probability, depending whether we think that they are more likely to be true or false (Fig. 2.1).

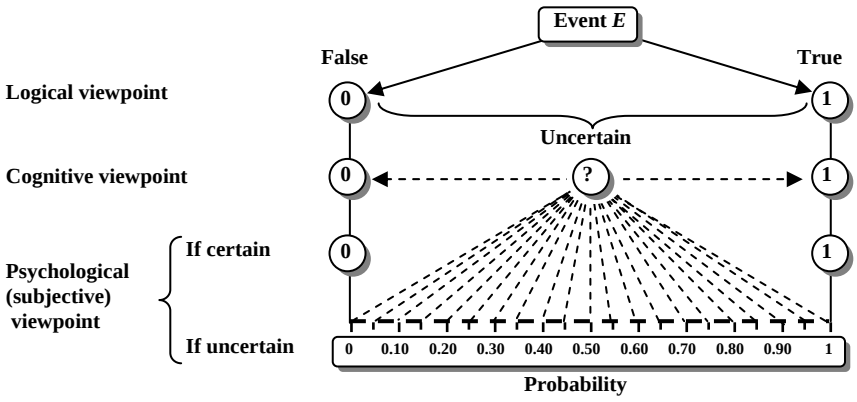


Figure 2.1 ***Probability: from certain to uncertain***

Defining probability as "*degree of belief*" seems "a priori" too vague to be of any use; we need, then, some explanation of its meaning, a tool to evaluate it. There are in fact two basic interpretation of probability: the *relative frequency* approach and the *axiomatic* or *subjective* approach. The relative frequency interpretation requires that a sample space Ω be defined, with event E being a member of Ω. If event E occurs x number of times out of a number N of repeated identical experiments, then the probability $P(E)$ of the outcome of event E is given by:

$$P(E) = \lim_{N \to \infty} \frac{x}{N} \qquad [2.1]$$

For fixed N, the quantity x/N is the *relative frequency of occurrence* of E. Since it is obviously impossible to actually conduct an infinite number of trials, $P(E)$ is in practice approximated by the relative frequency of occurrence calculated for a finite value N. The law of large numbers and the central limit theorem provide a justification that the estimation of $P(E)$ improves with increasing values of N (the larger N, the better the estimation of $P(E)$).

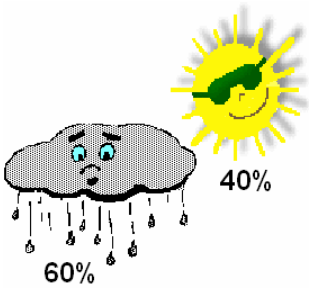
A slightly different formulation is that, "classical", of Laplace, which states that if there are N exhaustive, mutually exclusive and equally likely cases and m of them are favorable to an event E, the probability of E happening is defined as:

$$P(E) = \frac{m}{N} = \frac{\text{number of favorable outcomes}}{\text{total number of possible outcomes}} \qquad [2.2]$$

For example, the probability of randomly drawing a king from a well-shuffled deck of cards is 4/52. Since 4 is the number of favorable outcomes (i.e. 4 kings of diamond, spade, club and heart) and 52 is the number of total outcomes (the number of cards in a deck). This definition of probability is coherent with the concept of probability measuring numerically the degree of certainty or uncertainty of the occurrence of an event.

By definition, the relative frequency interpretation is only applicable when dealing with experiments that can be indefinitely repeated. There are many occasions however in the safety/reliability field when this is not the case, in particular in the common situation when the engineers have to consider rarely occurring events. Then it is necessary to resort to the axiomatic or subjective approach.

The axiomatic interpretation goes back to the literal definition of probability given above, i.e. probability is nothing more than a measure of uncertainty about the likelihood of an event. Stated more precisely, "a probability assignment is a numerical encoding of a state of knowledge when facing uncertainty". To get a better understanding of the subjective definition of probability, let us take the example of odds in betting. It seems reasonable to assume that the amount of money A that someone is willing to pay in order to possibly receive a sum of money B should an event E occurs, is directly proportional to the degree of belief of the better in the actual occurrence of this event. To make a coherent bet, if p is the numerical evaluation of this degree of belief ("probability") then our man should not stake more than $p \cdot B$ (in any case, to be worth betting, A should obviously always be strictly smaller than B) Weather forecasting is another example of subjective approach (though a meteorologist might feel offended to hear that evaluating the probability of rain tomorrow is "not objective"!). Saying that the probability of rain tomorrow is for example 60% doesn't mean of course that it will rain 60%; it will rain (100% occurrence) or it will not rain (0% occurrence). The information is nevertheless useful to decide (or "bet") that it will be wise to take an umbrella when going out. In the same way, evaluating at 5% the probability that a bridge could collapse under the load of a specific vehicle doesn't mean that the bridge in question will collapse 5% (the bridge will either hold on or break), but it will serve deciding if the bridge should be closed to this type of vehicle or not.



Stochastic variables

A *stochastic* (or *random*) *variable* is by definition a *function* X that maps from the sample space Ω to the real numbers $\Re$. Stated differently, a variable is a stochastic variable if the possible values of the variable have different probabilities.

A stochastic variable can be *discrete* or *continuous*. For example, the variable X representing the two possible positions – "open" (value: 1) or "close" (value: 0) – of a dam gate is a discrete variable, whereas the variable representing the elastic limit of the steel used to make a reinforcing bar is a continuous variable that can take any value between 0 and infinity.

As each value of a stochastic variable is associated to an occurrence probability, the description of the whole set of values is given under the form of a *probability distribution function*, $F_X(x)$, also called *cumulative probability function*:

$$F_X(x) \equiv P(X \leq x) , \forall\, x \qquad [2.3]$$

From this definition it follows that the function F_X takes values comprised in the interval [0, 1] when x varies within the range of x_{min} ($x_{min} \geq -\infty$) to x_{max} ($x_{max} > x_{min}$; $x_{max} \leq +\infty$). This function is moreover characterized by the following properties:

· F_x is monotonic and non-decreasing;
. $F_X(x) \geq 0$;
. $F_X(x_{min}) = 0$;
. $F_X(x_{max}) = 1$;

If X is a function that takes only discrete values (x_1, x_2, …x_i, … x_n), then $F_X(x)$ is given by:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} p_X(x_i) \qquad [2.4]$$

where the $p_X(x_i)$ are the probabilities (relative "weights") associated to the different possible values x_i.

In the case of a continuous variable, the equivalent of equation [2.4] takes the form:

$$F_X(x) = P(X \leq x) = \int_{x_{min}}^{x} f_X(u)\, du \qquad [2.5]$$

where $f_X(x)$ is the *probability density*, also simply called *distribution*.

A schematic representation of these two different cases is given in figure 2.2.

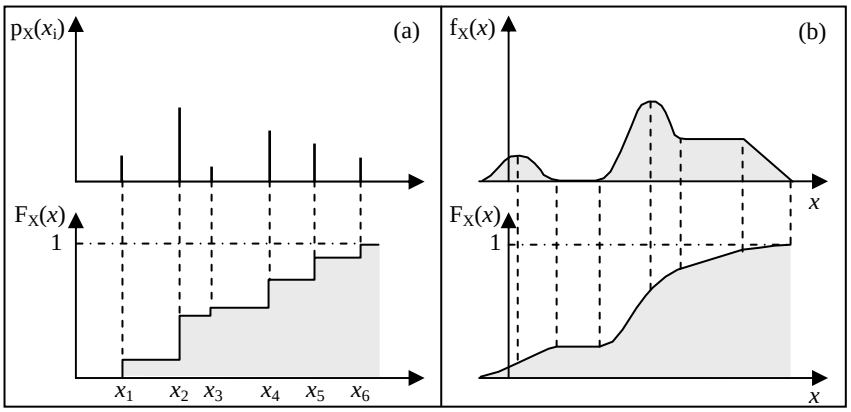


Figure 2.2 ***Variable distributions (a) discrete, (b) continuous***

Note that it is necessary to bring in the *Jacobian* of the transformation when changing the variable of a probability density function, e.g. from $f_X(x)$ to $f_Y(y)$:

$$f_Y(y) = \left|\frac{dx}{dy}\right| \cdot f_X(x)\Big|_{x=y} \qquad [2.6]$$

For example, if we want to express in terms of the variable t, with $t = \exp(\tau)$, a probability density function $f_n(\tau)$, the new probability density function will take the form:

$$f_{ln}(t) = t^{-1} \cdot f_n(\tau)\,|_{\tau = \ln t}$$

Moments

By definition, the moment of order n (with respect to the origin) is, for a given distribution, given by:

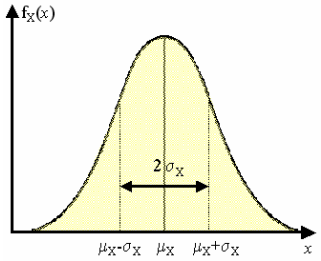
$$\mu_X^n = \sum_i (x_i)^n \cdot p_X(x_i) \qquad [2.7]$$

when the stochastic variable is discrete, and by:

$$\mu_X^n = \int_{x_{min}}^{x_{max}} x^n \cdot f_X(x)\,dx \qquad [2.8]$$

when the stochastic variable is continuous.

The first-order moment (n=1) is the *mean value* or *mathematical expectation* (E[X]) of the variable X. When a sample of limited size is considered, rather than the whole sample population, the mean is noted $\bar{x}$. Finding a *statistic estimator* of μ_X consists in putting $\mu_X \cong \bar{x}$.



The moment of second order with respect to the mean value is the *variance*. For a discrete variable, the variance thus takes the form:

$$V[X] = \sum_i (x_i - \mu_X)^2 \cdot p_X(x_i) \qquad [2.9]$$

and for a continuous variable:

$$V[X] = \left(\sigma_X^2\right) = \int_{x_{min}}^{x_{max}} (x - \mu_X)^2 \cdot f_X(x)\,dx \qquad [2.10]$$

where σ_X is the *standard deviation*. As above for the mean value, for a sample of limited size the standard deviation is noted s_X, with $s_X \cong \sigma_X$.

The ratio between the standard deviation and the mean value is called *coefficient of variation*:

$$COV[X] = \frac{\sigma_X}{\mu_X} \qquad [2.11]$$

This coefficient is of course meaningful only for the distributions having a mean value different from zero.

Some useful distributions in the context of risk/reliability analysis are briefly presented below. A general recapitulative table is given in Appendix 2.1.

Discrete distributions

The *discrete uniform distribution* is one of the simplest probability distributions. In this distribution, all values of the random variable are assigned identical probabilities. There are many situations in which the discrete distribution arises, e.g. the outcome of the throw of a single die (Fig. 2.3).

Throwing a single die						
x_i	1	2	3	4	5	6
$P(X=x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Figure 2.3 ***Example of a discrete uniform distribution***

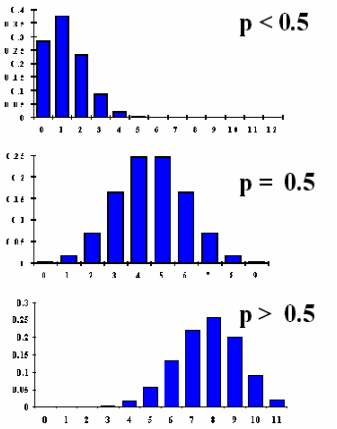
The *binomial distribution* characterizes experiments (*Bernouilli sequences*) that satisfy the following conditions:

1. there are only two possible outcomes on each trial of the experiment; one of the outcome is usually referred to as a *success* and the other as a *failure*;
2. the occurrence probabilities of each of the outcomes (success, probability *p*, or failure, probability *q*=1-*p*) in a trial are constant;
3. the experiment consists of *n* identical, statistically independent, trials.

Tossing a coin 4 times and recording the number of heads is a simple example of such an experiment (Fig. 2.4).

Tossing a coin		
Events	**Number of heads (H)**	**Probability**
TTTT	0	$\frac{1}{16}$
HTTT THTT TTHT TTTH	1	$\frac{4}{16}$
HHTT HTHT HTTH THHT THTH TTHH	2	$\frac{6}{16}$
HHHT HHTH HTHH THHH	3	$\frac{4}{16}$
HHHH	4	$\frac{1}{16}$

Figure 2.4 ***Example of a Bernouilli sequence***



Binomial distribution

The binomial random variable is the count of the number of successes in *n* trials. The probability of obtaining *x* successes in *n* trials is given by the binomial distribution:

$$P(X = x) = C_x^n \, p^x \, (1-p)^{n-x} \qquad x = 0,1,2,\ldots n \qquad [2.12]$$

C_x^n (*binomial coefficient*) represents the number of possible combinations of n objects taken x at the time (without replacement); it is given by:

$$C_x^n = \frac{n!}{x!(n-x)!} \qquad [2.13]$$

For example, the probability that $x = 3$ in the example of Fig. 2.4, given that $n = 4$ et $p = 0.5$, takes the value:

$$P(X = 3) = \frac{4!}{3!(4-3)!} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(1-\frac{1}{2}\right)^{(4-3)} = \frac{4}{16} = 0.25$$

Let us consider as another example, the case of a hospital manager who considers buying three diesel generator sets for power backup. He estimates at 80% the probability that a given make of generator set will remain operational for at least four years (success). Calculate the probability that x ($x = 0, 1, 2, 3$) generators of this type are still operational after four years.

The binomial random variable X to consider is here the number of operational generator sets after four years. The probability of success p is equal to 0.80 and $n = 3$. The answer to the question is thus given by the binomial distribution of figure 2.5.

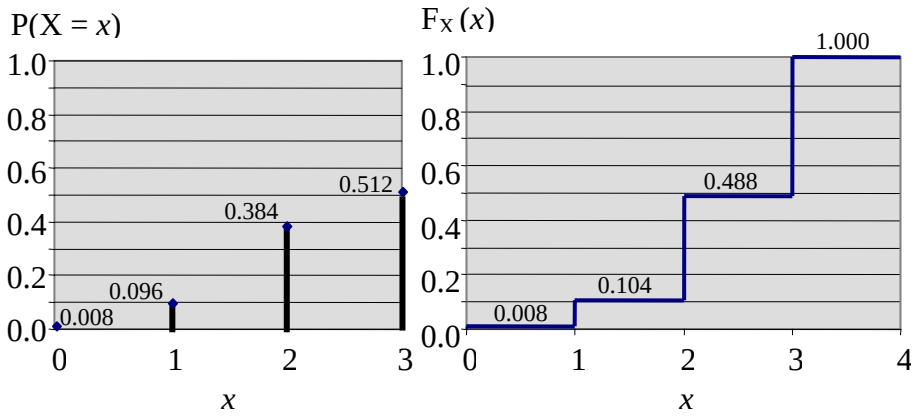


Figure 2.5 ***Binomial distribution of operational generators after four years***

The *expected value* of a binomial random variable can be computed using the expression:

$$E[X] = n \cdot p \qquad [2.14]$$

and its *variance* using the expression:

$$V[X] = n \cdot p \cdot q = n \cdot p \cdot (1-p) \qquad [2.15]$$

In a Bernouilli sequence, the number of trials until the first occurrence of an event is given by the *geometric distribution*. If the first success takes place after the t^{th} trial, this means that only failures were observed in the t-1 preceding trials. Thus, if T is the random variable:

$$P(T = t) = p \cdot (1-p)^{t-1} \qquad [2.16]$$

which is a geometric distribution.

The recurrence time between two events in a Bernouilli sequence is characterized by a geometric distribution.

The mean recurrence time, also known under the name *return period*, is given by:

$$\overline{\mathrm{T}} = \mathrm{E}[\mathrm{T}] = \sum_{t=1}^{\infty} t \cdot p \cdot (1-p)^{n-1} = \frac{1}{p} \qquad [2.17]$$

Moreover (variance):

$$\mathrm{V}[\mathrm{T}] = \frac{1-p}{p^2} \qquad [2.18]$$

Example: a wind turbine is designed to withstand a wind having a return period of 50 years. What is the probability that this wind will be exceeded for the first time 5 years after the wind turbine is brought into service?

The yearly occurrence probability of the design wind is given by:

$$p = \frac{1}{\overline{\mathrm{T}}} = \frac{1}{50} = 0.02$$

The answer to the question is thus:

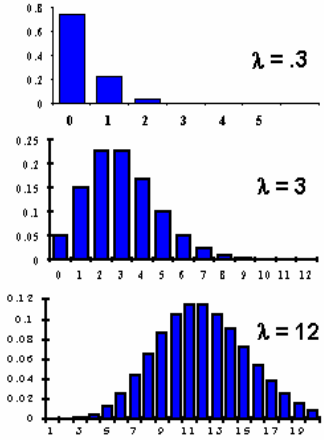
$$\mathrm{P}(\mathrm{T} = 5) = 0.02 \cdot (1 - 0.02)^4 = 0.018$$

The *Poisson distribution* is similar to the binomial in that the random variable represents a count of the total number of "successes". The major difference between these two distributions is that the Poisson does not have a fixed number of trials. Instead, the Poisson distribution uses a fixed interval of time or space in which the number of "successes" is recorded.

Many engineer's problems concern possible occurrences of events distributed in time or space (fatigue rupture anywhere on a cable, pump failures, earthquakes, …). These could be represented by a Bernouilli sequence if time or space were discretized, but in this case the event considered could happen only once (occurrence or non-occurrence) in the given interval. In the general case, a Poisson process is thus a better model for such "experiments".

In order to qualify as a Poisson random variable, an experiment must meet the three following conditions:

1. an event can take place randomly at any point in time or space; "successes" occur one at a time; that is, two or more "successes" cannot occur at exactly the same point in time or exactly at the same point in space;
2. the occurrence of a "success" in any interval is independent of the occurrence of the "successes" in another interval;
3. the occurrence probability of a "success" in a small interval Δt (time or space) is proportional to Δt.



Poisson distribution

The number of occurrences X_t of an event in a (time or space) interval t is in these conditions given by a Poisson distribution:

$$\mathrm{P}(\mathrm{X}_t = x) = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^x}{x!} \qquad [2.19]$$

where λ is the mean number of occurrences of the event in the interval t, which can also be written $\lambda = \nu \cdot t$, with ν the mean occurrence frequency.

The expected value and variance are respectively given by:

$$\mathrm{E}[\mathrm{X}] = \mathrm{V}[\mathrm{X}] = \lambda \qquad [2.20]$$

Example: the number of calls received by the dispatching office of a gas utility on Monday morning between 8:00 a.m. and 9:00 a.m. has a Poisson distribution with λ equal to 4.0. Determine the probability of getting no call between eight and nine in the morning.

With X_t = the number of calls received between 8:00 a.m. and 9:00 a.m.:

$$P[X_t = 0] = \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{e^{-4} \cdot 4^0}{0!} = 0.0183$$

The expected number of calls during the same period is 4.0 and the variance is also 4.0.

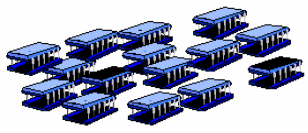
The *hypergeometric distribution* , like the binomial distribution, counts the number of "successes" in *n* trials of an experiment and describes the behavior of a random variable that has only two possible outcomes on each trial of the experiment. It differs however from the binomial distribution in the lack of independence between trials, which implies that the probability of "success" vary between trials. In addition, hypergeometric distributions have finite populations in which the total number of "successes" and "failures" are known.

If *N* is the total population, *S* the total number of "successes" possible and *n* the size of the sample drawn, the hypergeometric probability distribution function takes the form:

$$P(X = x) = \frac{C_x^S \cdot C_{n-x}^{N-S}}{C_n^N} \qquad [2.21]$$

where $0 \le x \le$ minimum of [*S*, *n*].

Example: suppose that a shipment from a semiconductor manufacturer contains 30 memory ships of which two are defective. If a memory board requires 16 ships, what is the probability distribution for the number of defective ships on the memory board?



The random variable to consider here is X = number of defective ships on the memory board. The parameters of the distribution are:

S = 2 ("success" in this case means a defective ship);
N = 30;
n = 16..

The maximum value of *X* in this case is 2.

$$P(X = 0) = \frac{C_0^2 \cdot C_{16\text{-}0}^{30\text{-}2}}{C_{16}^{30}} = 0.209 \; ; \; P(X = 1) = 0.515 \; ; \; P(X = 2) = 0.276$$

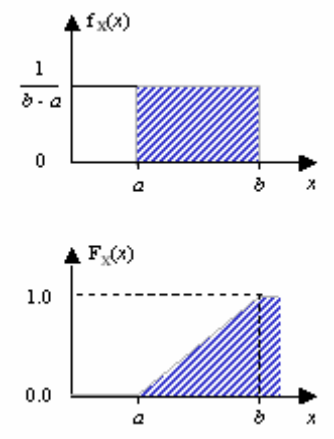
The expected value of a hypergeometric variable is given by:

$$E[X] = n \cdot \frac{S}{N} \qquad [2.22]$$

and its variance by:

$$V[X] = \left\{ n \cdot \frac{S}{N} \cdot \left(1 - \frac{S}{N} \right) \right\} \cdot \frac{N - n}{N - 1} \qquad [2.23]$$

Continuous distributions



Uniform distribution

As in the discrete case, the *uniform distribution* is, from a mathematical viewpoint, the simplest of the continuous distribution functions. In this distribution, the probability density is spread out evenly over some range from *a* to *b* (the two parameters of the distribution) as shown in the figure on the left (upper part).

The mathematical expression of the uniform distribution is therefore given by:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases} \qquad [2.24]$$

The expected value (mean) of the continuous uniform stochastic variable is:

$$E[X] = \frac{a+b}{2} \qquad [2.25]$$

and its variance is:

$$V[X] = \frac{(b-a)^2}{2} \qquad [2.26]$$

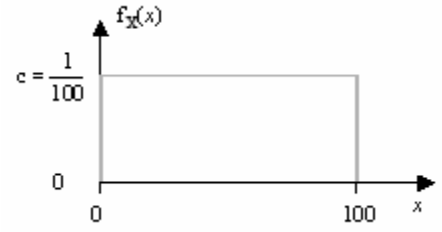
The probability of observing a stochastic variable in some interval is expressed as the area under the density function associated with this interval. Because the density function for the uniform distribution has the shape of a rectangle, calculating the probability for an interval is straightforward.

Example: suppose that a spent nuclear fuel is to be transported along a 100km-long railway. The occurrence probability of an accident is assumed to be uniform along the 100 km of this railway. If X is the stochastic variable giving the distance (from km 0) to the place of the accident, what is the probability to have an accident between km 20 and km 35 (should an accident arise)?

The distribution of the variable X is uniform, therefore:

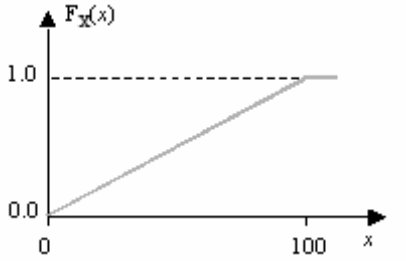
$f_X = c \qquad 0 \le c \le 100$

with $c = \dfrac{1}{100 - 0} = \dfrac{1}{100}$



The probability distribution function is given by:

$$F_X(x) = \int_0^x c\,dx = c \cdot x = \frac{x}{100} \qquad 0 \le x \le 100$$



Thus, the probability to have an accident between km 20 and km 35 can easily be calculated as follows:

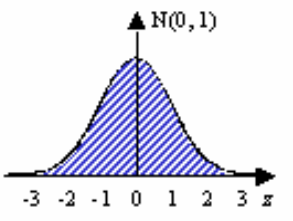
$$P(20 \le x \le 35) = \int_{20}^{35} c\,dx = \frac{35-20}{100} = 0.15$$

The *normal distribution* is incontestably the best known of the continuous distributions. It was originally called the Gaussian distribution, named after Karl Gauss who published a work in 1833 describing the mathematical definition of this distribution, which he developed to describe the error in predicting the orbits of planets.

The general mathematical expression of the normal distribution is:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad [2.27]$$

The two parameters are μ (mean) and σ^2 (variance). For this reason, the normal distribution is often written $N(\mu, \sigma)$ in abbreviated form. The special case of normal distribution with a mean μ of zero and standard deviation σ of one, i.e. N(0, 1), is called the *standard normal*. This is how the normal distribution is usually tabulated; the technique used to translate any normal stochastic variable into a standard one is called a *z-transform*:



Standard normal distribution

$$z = \frac{x - \mu}{\sigma} \qquad [2.28]$$

Example: measures made in view of the construction of a hydroelectric scheme have shown that the annual precipitations in a given catchment area can be represented by a normal distribution N(120 cm, 30 cm). What is the probability that the precipitation will exceed 60 cm?

z-transform of the original *x* value (60 cm) gives in this case:

$$z = \frac{60 - 120}{30} = -2$$

Tables give 0.47725 for the area under the standard normal curve comprised between the abscissas z=0 and z=2. Taking into account the symmetry of the normal distribution, this corresponds also to the area under the curve between z=-2 and z=0. To find the answer to the question (probability that the precipitations will *exceed* 60 cm) it suffices to add 0.5 to the above value to take into account the half distribution to the right of the zero axis, thus $P(X \geq 60 \text{ cm}) = 0.97725$.

One of the most important properties of normal stochastic variables is that within a fixed number of standard deviations from the mean, all normals contain the same fraction of their probabilities. For example, the probabilities of being within one standard deviation ($\pm 1\ \sigma$), two standard deviations ($\pm 2\ \sigma$) and three standard deviations ($\pm 3\ \sigma$) of the mean equal respectively 0.683 (68.3%), 0.954 (95.4%) and 0.997 (99.7%). This explains why in parametric sensitivity studies, the analysis is often limited to values situated within three standard deviations from the mean (see Fig. 2.6).

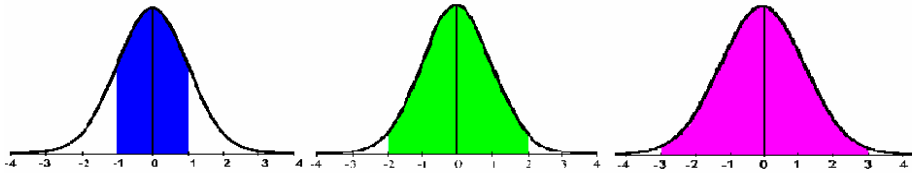
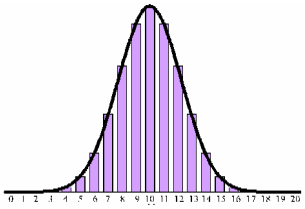


Figure 2.6 ***Probabilities of being within, one, two or three standard deviations from the mean in a normal distribution***

Although the (symmetrical) distribution can range in value from minus infinity to positive infinity, values that are a great distance from the mean rarely occur. Therefore, even if in principle the normal distribution is not well adapted to variables for which negative values are physically ruled out (disintegration time of a radionuclide for example), it constitutes nevertheless a good and suitable approximation in cases where the COV (coefficient of variation) is less than 30%.

The normal distribution can also be used to approximate discrete distributions, specially the binomial and the Poisson. This can prove very useful because as n becomes large, calculating these probabilities can become time consuming. Indeed, the larger the value of n, the more accurate the approximation. In the case of the binomial distribution, the approximation is generally reasonable when the mean $n \cdot p$ is greater or equal to 5 and $n \cdot (1-p)$ is also greater than or equal to 5. The approximation becomes quite good when these values are greater than 10. Approximating a binomial distribution by a normal one requires that both distributions have the same mean and variance:



Approximation of a binomial distribution by a normal distribution

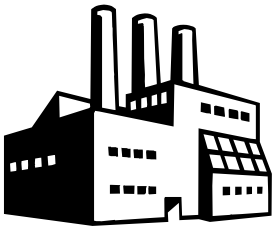
$$\mu = n \cdot p \quad \text{and} \quad \sigma^2 = n \cdot p \cdot (1 - p) \qquad [2.29]$$

Example: to approximate a binomial distribution with n = 20 and p = 0.5 will require a normal distribution with:

$$\mu = (20) \cdot (0.5) = 10 \quad \text{and} \quad \sigma^2 = (20 \cdot 0.5) \cdot (1 - 0.5) = 5$$

In the same way, to approximate a Poisson distribution by a normal one, the mean and variance of the normal should be set to the mean and variance of the Poisson. Since the mean and variance of the Poisson distribution are both λ, the appropriate mean and variance for the normal would be:

$$\mu = \sigma^2 = \lambda \qquad [2.30]$$



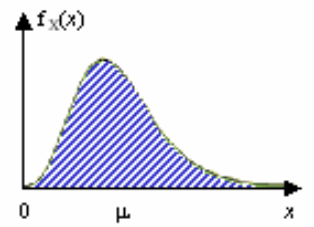
Example: a company manufacturing metal sheets estimates that the number of defects on a 10' by 10' sheet of metal follows a Poisson distribution with an average defect rate of 5 per sheet. Find the probability of observing at least 10 defects per sheet.

Let X be the number of defects on a 10' by 10' metal sheet. X has a Poisson distribution with a mean and variance of 5. Using the normal distribution with the same mean and variance to approximate the Poisson, the answer is given by:

$$P(X \geq 10) = P\left(z \geq \frac{x - \mu}{\sigma} = \frac{10 - 5}{\sqrt{5}} = 2.24\right)$$

Tables give 0.4875 for the area under the standard normal curve comprised between the abscissas z=0 and z=2.24. Therefore, the probability of observing at least 10 defects per sheet is approximately equal to 0.5-0.4875 = 0.0125 (the exact answer would be 0.0317).

A distribution closely related to the normal distribution is the *lognormal distribution*. A variable X is lognormally distributed if Y = ln(X) is normally distributed. The general formula for the probability density function of the lognormal distribution is (taking the relation [2.6] into account):



Lognormal distribution

$$f_X(x) = \frac{e^{-(\ln x - \mu)^2 / 2\sigma^2}}{x \cdot \sigma \cdot \sqrt{2\pi}} \quad x \geq 0 \; ; \; \mu, \sigma > 0 \qquad [2.31]$$

The lognormal distribution is applicable when the quantity of interest must be positive, since ln x exists only for positive values of x.

The *exponential distribution* is a commonly used distribution in reliability engineering. The exponential distribution is a probability density with only one parameter, λ. It is characterized by the fact that its standard deviation is equal to the mean, i.e. its coefficient of variation is 1 (100%). As in the previous case, this distribution is defined only for positive values of the variable.

This distribution can describe a number of physical phenomena, such as the time for a radioactive nucleus to decay, or the time for a component to fail. This distribution describes more generally the random variable T_1 corresponding to the waiting time till the occurrence of the first event in a process governed by a Poisson law. According to [2.19], we have in this case:

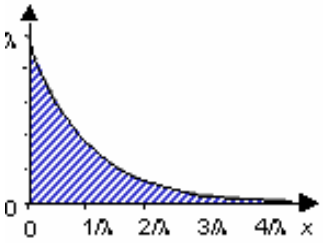
$$P(T_1 > t) = P(X_t = 0) = e^{-\lambda} = e^{-\nu \cdot t} \qquad [2.32]$$

The complement of this expression is precisely the cumulative function of the distribution we are interested in:

$$F_{T1}(t) = P(T_1 \le t) = 1 - e^{-\nu \cdot t} \qquad [2.33]$$

Its derivative (relative to t) gives therefore the mathematical expression of the exponential distribution (substituting here $\nu.t \rightarrow \lambda \cdot x$):

$$f_X(x) = \lambda \cdot e^{-\lambda \cdot x} \qquad [2.34]$$



Exponential distribution

The mean and variance of the exponential distribution are given by :

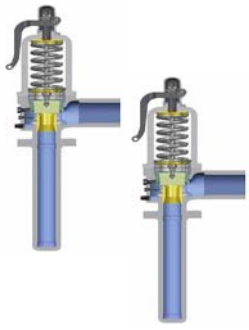
$$E[X] = 1/\lambda, \qquad V[X] = 1/\lambda^2 \qquad [2.35]$$

Example: suppose that the lifetime of a certain electronic component (in hours) is exponentially distributed with rate parameter $\nu = 0.001$.Find the probability that the component lasts at least 2000 hours.

Let T be the stochastic variable denoting the lifetime of the component, then (Eq. [2.32]):

$$P(T > 2000) = e^{-(0.001 \cdot 2000)} = 0.1353$$

Note that the exponential distribution is generally not very useful in modeling data in the real world. The exponential distribution is only useful for items that have a constant failure rate. This means that the population should have no wear-out or infancy problems.



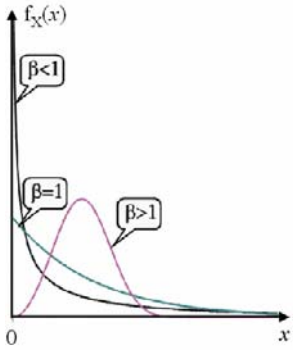
The exponential distribution possesses a special property called the *memoryless* property. Suppose that a device (e.g. safety valve) has a lifetime that can be modeled as an exponential distribution. Then the probability that a *new* device survives t units of time is (Eq. [2.32]):

$$P(T > t) = 1 - P(T \le t) = e^{-\nu \cdot t}$$

Now suppose that an identical device has already survived s units of time. We can think of s as the *age* of the device. What is the probability that this device survives an additional t units of time? The event of interest here is the event that the device survives past time $s+t$. This probability is a *conditional probability* since we are given that lifetime of this device must be greater than its current age s. Therefore:

$$P(X > s+t \mid X > s) = \frac{P(X > s \text{ and } X > s+t)}{P(X > s)} = \frac{e^{-\lambda \cdot (s+t)}}{e^{-\lambda \cdot s}} = e^{-\lambda \cdot t} = P(X > t)$$

That is, the probability that a used device survives an additional t units of time is the same for the used device (with age s) as it is for the new device.



Weibull distributions

The exponential distribution is a particular case of the more general *Weibull distribution*, named in honor of Wallodi Weibull (Swedish materials science engineer, 1887-1979). The Weibull family of distributions is immensely popular in reliability theory, because of the many shapes it attains for various values of its parameters. It can therefore model a great variety of data and life characteristics.

The formula for the probability density function of the general Weibull distribution, defined from zero to positive infinity, is:

$$f_X(x) = \frac{\beta}{\alpha} \cdot \left(\frac{x-\mu}{\alpha} \right)^{(\beta-1)} \cdot \exp\left\{ - \left[(x-\mu)/\alpha \right]^{\beta} \right\} \quad x \geq \mu;\ \beta, \alpha > 0 \qquad [2.36]$$

where β is the *shape parameter*, μ is the *location parameter*, and α is the *scale parameter*. The case where $\mu = 0$ and $\alpha = 1$ is called the *standard Weibull distribution*. The case where only $\mu = 0$ is called the *two-parameter Weibull distribution*.

Note that the general form of probability functions can always be expressed in term of the standard distribution (see normal distribution).

The mean and variance of the general Weibull are respectively given by:

$$E[x] = \mu + \alpha \cdot \Gamma(1 + 1/\beta) \qquad [2.37]$$

$$V[x] = \alpha^2 \cdot \left[\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta) \right] \qquad [2.38]$$

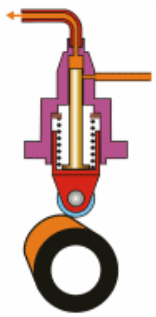
where Γ is the gamma function:

$$\Gamma(a) = \int_0^{\infty} u^{(a-1)} \cdot \exp(-u)\, du \qquad [2.39]$$

When $\beta = 1$, the two-parameter Weibull distribution reduces to the exponential distribution with scale parameter α. The special case $\beta = 2$, is called the Rayleigh distribution with scale parameter α.

The β value gives clues about the failure mechanism:

- $\beta < 1$ implies “infant illnesses” (decreasing failure rate);
- $\beta = 1$ implies random failures, i.e. independent of time, an old device is as good as a new device (see exponential distribution);
- $1 < \beta \leq 4$ implies early wear out (increasing failure rate), due possibly to low cycle fatigue, bearing failures, corrosion, erosion;
- $\beta > 4$ implies old age and rapid wear out; typical failure modes involve some material properties, some corrosion and erosion.



Example: the lifetime T (in hours) of a fuel pump follows a two-parameter Weibull life distribution model with shape parameter $\beta = 1.5$ and scale parameter $\alpha = 8'000$. If a typical fuel pump is used 800 hours a year, what proportion is likely to fail within 5 years?

The cumulative probability function for a two-parameter Weibull distribution is:

$$F_T(t) = 1 - \exp\{-(t/\alpha)^{\beta}\}$$

The probability $P(T \leq 800 \cdot 5)$ is thus 0.2978; this means that about 30% of the pumps will fail in the first 5 years.

The *beta distribution* is another versatile two-parameter family of distributions that has special importance in probability and statistics. The beta distribution presents the characteristic to be defined over a finite range, with end points *a* and *b* (real numbers); it is therefore often used for representing processes with natural lower and upper bounds. Depending on the values of its parameters, the beta distribution generated will have the "U", the "J", the triangle or the general bell shape of the unimodal function..

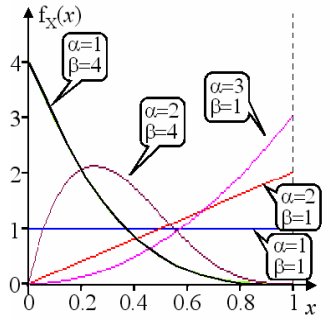
The general formula for the probability density function of the beta distribution is:

$$f_X(x) = \frac{1}{B(\alpha,\beta)} \cdot \frac{(x-a)^{\alpha-1} \cdot (b-x)^{\beta-1}}{(b-a)^{\alpha+\beta-1}} \quad a \le x \le b;\ \alpha, \beta > 0$$

$$f_X(x) = 0 \qquad \text{outside of this range} \qquad [2.40]$$

where α and β are the *shape parameters*, and $B(\alpha, \beta)$ is the *beta function*, given by:

$$B(\alpha,\beta) = \int_0^1 u^{(\alpha-1)} \cdot (1-u)^{(\beta-1)}\, du \qquad [2.41]$$



Standard beta distributions

The beta function can be written in terms of the gamma function (Eq. [2.39]) as follows:

$$B(\alpha, \beta) = \Gamma(\alpha)\cdot\Gamma(\beta)/\Gamma(\alpha+\beta) \qquad [2.42]$$

The case where $a = 0$ and $b = 1$ is called the *standard beta distribution*. The equation for the standard distribution thus is:

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} \quad 0 \le x \le 1;\ \alpha, \beta > 0 \qquad [2.43]$$

The beta distribution is different from the other distributions in that it is defined in terms of its lower and upper bounds rather than in terms of location and scale. However, the location and scale parameters are related to the lower and upper bounds as follows:

$$\text{location} = a \qquad \text{scale} = b - a \qquad [2.44]$$

The mean and variance of the beta distribution are:

$$E[X] = a + \frac{\alpha}{\alpha+\beta} \cdot (b-a) \qquad [2.45]$$

$$V[X] = \frac{\alpha \cdot \beta}{(\alpha+\beta)^2 \cdot (\alpha+\beta+1)} \cdot (b-a)^2 \qquad [2.46]$$

Estimating the α and β parameters is controlled by data availability. Let us consider the case where *a* and *b* are known, and estimates of the mean, $\bar{x}$, and the variance, $\mathbf{s_X}$, are available. In this case, the parameters α and β are given by [Engineering Statistics Handbook, 2003]:

$$\alpha = \tilde{x} \cdot \left(\frac{\tilde{x} \cdot (1-\tilde{x})}{\tilde{s}_X^2} - 1 \right); \quad \beta = (1-\tilde{x}) \cdot \left(\frac{\tilde{x} \cdot (1-\tilde{x})}{\tilde{s}_X^2} - 1 \right) \qquad [2.47]$$

where:

$$\tilde{x} = \frac{\bar{x}-a}{b-a}; \quad \tilde{s}_X^2 = \left(\frac{s_X}{b-a} \right)^2 \qquad [2.48]$$

Multivariate distributions

When outcomes depend on more than one parameter, a *multivariate distribution function* must be defined:

$$P(X_1 \le x_1;\ X_2 \le x_2;\ \ldots\ X_n \le x_n) = F_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots x_n) = \int_{x_{1\min}}^{x_1} \int_{x_{2\min}}^{x_2} \ldots \int_{x_{n,\min}}^{x_n} f_{X_1 X_2 \ldots X_n}(u_1, u_2, \ldots u_n)\, du_1\, du_2 \ldots du_n \qquad [2.49]$$

where: $F_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots x_n)$ is the *multivariate cumulative probability function,*

and: $f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots x_n)$ is *the joint distribution function*.

The case of a bivariate distribution is represented in Fig. 2.7.

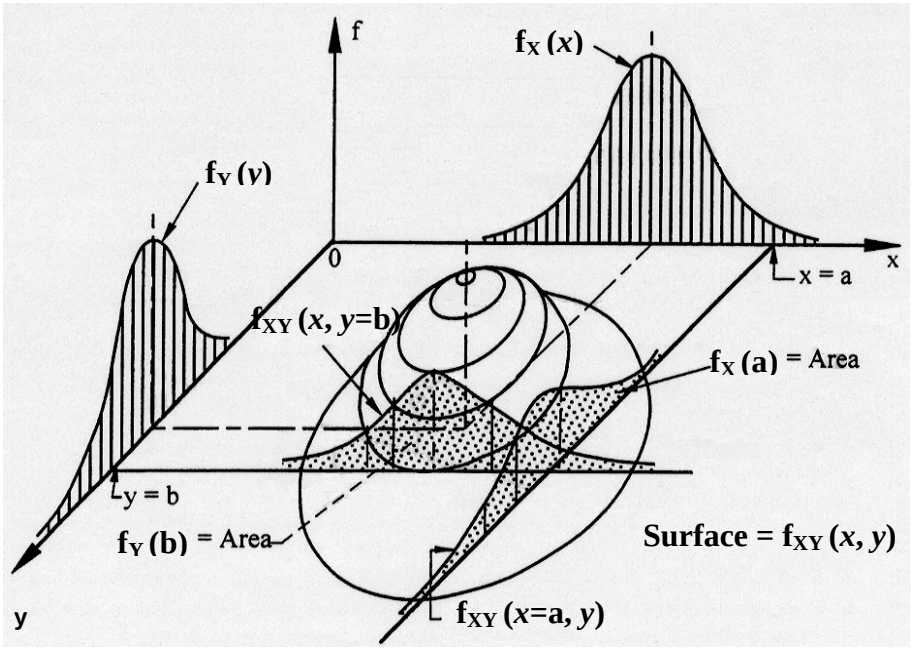


Figure 2.7 ***Bivariate distribution*** **(source: Prof. L. Vulliet)**

The second joint moment relative to the means μ_X and μ_Y is called *covariance* of X and Y:

$$\mathrm{Cov}[X, Y] = E[(X - \mu_x)(Y - \mu_y)] = E[Y, Y] - E[X] \cdot E[Y] \qquad [2.50]$$

The covariance measures the degree of correlation between the variables X and Y; if the variables X and Y are statistically independent, Cov[X, Y] = 0. Instead of the covariance, it is often considered more meaningful to use its normalized expression, the *correlation coefficient* ρ, defined as:

$$\rho = \frac{\mathrm{Cov}[X, Y]}{\sigma_X \cdot \sigma_Y} \qquad [2.51]$$

Example: the bivariate *normal* distribution is given by:

$$f_{XY}(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho^2}} \cdot \exp\left[-\frac{z}{2(1-\rho^2)}\right]$$

where: $$z = \frac{(x-\mu_X)^2}{\sigma_X} - \frac{2\rho(x-\mu_X)\cdot(y-\mu_Y)}{\sigma_X \cdot \sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y}$$

2.2 Elements of Estimation Theory

Relationships between a population and samples drawn from it

The calculation of the occurrence probability of a given event will be based on the use of an appropriate probability distribution. However, neither the shape (type), nor the parameters of this distribution are known *a priori*. These should be inferred from the observation of the physical system of interest. It is the goal of the *estimation theory* to provide a mean of calculating suitable estimates of quantities based on observations of physical systems.

The problem regarding such estimates is that we usually don't have statistics about a whole population but only about some finite-sized sample(s) drawn from the population in question. The task is then, given the observable variables, to determine as accurately as possible the actual distribution and the value of its parameters. This is called *statistical inference and estimation* (Fig. 2.8). In this context, parameters are assumed to be themselves random quantities related statistically to the observation.

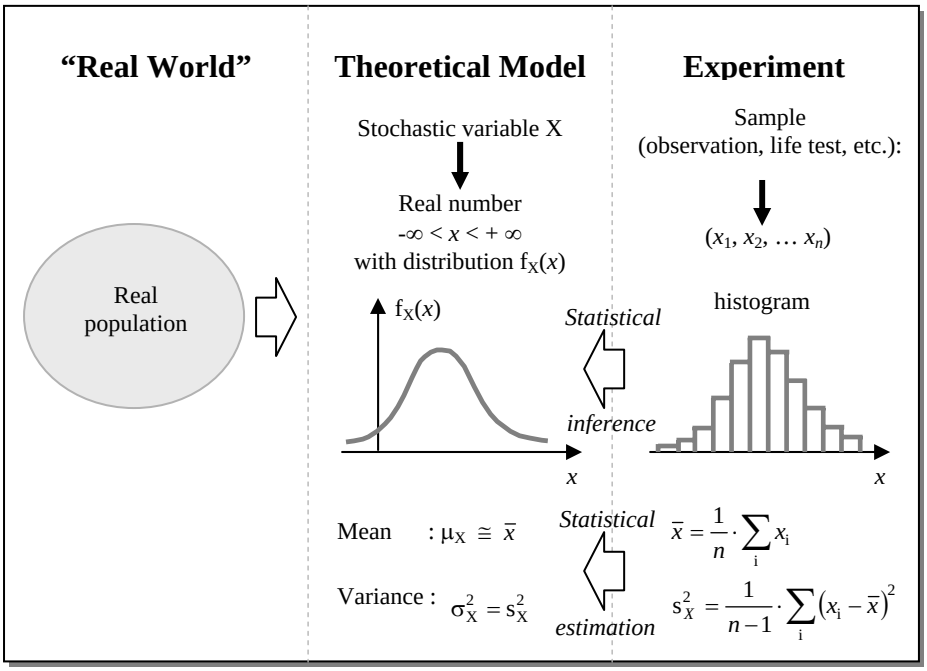


Figure 2.8 ***Statistical inference and evaluation*** **(adapted from: Prof. L. Vulliet)**

The estimation process results in inherent uncertainties affecting for example the moments (mean, variance, etc.) of the distribution, and therefore the numerical results of risk analysis.

Moment method of point estimation

The moments of a distribution allows us to determine the parameters of this distribution. In the case of a normal distribution, the parameters μ and σ^2 are directly the moments of first order relative to the origin and of second order relative to the mean respectively. For the other distributions, the reader is referred to the table given in Appendix 2.1.

It is natural to consider that the moments of a sample constitute a first approximation of the moments corresponding to the whole population. This is the basic principle of the *moment method of point estimation*. Thus, the point estimate of the population mean can be approximated by the mean of the sample, i.e.:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \qquad [2.52]$$

In the same way, a first approximation of the variance relative to the whole population is simply given by :

$$s_X^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad [2.53]$$

The sample mean as calculated in Eq. [2.52] is an *unbiased estimate* $\hat{\mu}_X$ of the population mean μ_X. An estimate is said to be *unbiased* if the expected value of the estimate equals the true value of the parameter. When we have a biased estimate, the bias usually depends on the number of observations *n*. That's the case of the variance estimate given in Eq. [2.53]. To obtain an unbiased estimate of the population variance σ_X^2, it is necessary to use instead the following expression:

$$\hat{\sigma}_X^2 = \frac{n}{n-1} \cdot s_X^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad [2.54]$$

Now suppose all possible samples of size *n* are drawn without replacement from a whole population of size $n_P > n$. If the mean and variance of the sampling distribution of the mean are denoted by $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}^2$ respectively, and the mean and variance of the whole population by μ_X and σ_X^2 as always, then [McCormick, 1981]:

$$\mu_{\bar{X}} = \mu_X \qquad [2.55]$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma_X^2}{n} \cdot \left(\frac{n_P - n}{n_P - 1} \right) \qquad [2.56]$$

In the case where the population is infinite, or the sampling is done with replacement, Eq. [2.56] simplifies to:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_X^2}{n} \qquad [2.57]$$

For large enough *n*, say $n \geq 30$, the sampling distribution of means follows approximately a normal distribution with mean $\mu_{\bar{X}}$ and variance $\sigma_{\bar{X}}^2$, irrespective of the size of n_P (provided $n_P \geq 2n$). The sampling distribution of means is said to be *asymptotically normal*.

Example: ball bearings used in a wind turbine system come from a batch in which the mean mass is12 g with a standard deviation of 0.30 g. If a sample of 100 ball bearings is chosen from a total population of 500, $\mu_{\bar{X}}$ = 12 g and:

$$\sigma_{\bar{X}} = \frac{0.30\,\text{g}}{\sqrt{100}} \cdot \sqrt{\frac{500-100}{500-1}} = 0.027\,\text{g}$$

If the batch population was infinite, or the sampling made with replacement, the standard deviation will be:

$$\sigma_{\bar{X}} = \frac{0.30\,\text{g}}{\sqrt{100}} = 0.030\,\text{g}$$

The equations used to obtain point estimates in reliability analyses depend upon the type of experiment the samples are subjected to during a life test. There are two main possibilities; either the life test is terminated at a given time t_c before all *n* items have failed (*Type I* censoring of the life test), or censoring occurs when a predetermined number of items have failed ($k \leq n$), independently of the time needed to achieve this result (*Type II* censoring).

If the times t_1, t_2, … t_n represent the actual failure times observed for a set of n identical units, then a unbiased estimator of the mean time to failure, defined by:

$$\mu_T = \int_0^\infty t \cdot f(t)\, dt \qquad [2.58]$$

is (from Eq. [2.52]):

$$\hat{\mu}_T = \frac{1}{n} \cdot \sum_{i=1}^{n} t_i \qquad [2.59]$$

For the variance, defined by:

$$\sigma_T^2 = \int^\infty (t - \mu_T)^2 \cdot f(t)\, dt \qquad [2.60]$$

the unbiased estimator is (from Eq. [2.54]):

$$\hat{\sigma}_T^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (t_i - \hat{\mu}_T)^2 \qquad [2.61]$$

Combining equations [2.59] and [2.61] with the appropriate expressions of the means and variances collected from the table in the appendix 2.1 give a couple of equations that allows us to calculate two unknown parameters for the corresponding failure probability distributions.

Example: let us assume that ten identical devices (e.g. valves) are tested with the result of failures occurring at t_1 = 170 hr, t_2= 350 hr, t_3=500 hr, t_4=650 hr, t_5=800 hr, t_6=960 hr, t_7=1100 hr, t_8= 1300 hr, t_9= 1800 hr and t_{10} =2200 hr. Estimate the α and β parameters for a two-parameter Weibull distribution, using moment estimators.

Fig.: Henry Pratt Company

The first step is to calculate $\hat{\mu}_T$ and $\hat{\sigma}_T^2$ using Eqs. [2.59] and [2.61] respectively with the ten recorded time failures.

$$\hat{\mu}_T = \frac{1}{10} \cdot \sum_{i=1}^{n} t_i = 983\,\text{hr}$$

$$\hat{\sigma}_T^2 = \frac{1}{(10-1)} \cdot \sum_{i=1}^{n} (t_i - 983)^2 = 4.114 \cdot 10^5\,\text{hr}^2$$

The two-parameter Weibull distribution takes the form:

$$f_T(t) = \frac{\beta}{\alpha} \cdot \left(\frac{t}{\alpha}\right)^{(\beta-1)} \cdot \exp\left\{-(t/\alpha)^\beta\right\}$$

Taking into account the analytical expressions of the mean and variance of this distribution derived from appendix 2.1 leads to the following system of coupled equations:

$$\hat{\alpha} \cdot \Gamma\left(1 + \hat{\beta}^{-1}\right) = \hat{\mu}_T$$

$$\hat{\alpha}^2 \cdot \left\{\Gamma\left(1 + 2\beta^{-1}\right) - \left[\Gamma\left(1 + \beta^{-1}\right)\right]^2\right\} = \sigma_T^2$$

These two equations must be solved simultaneously by an iterative method to obtain $\hat{\alpha}$ and $\hat{\beta}$. The result is that $\hat{\alpha}$ = 1095 hr and $\hat{\beta}$ = 1.58.

There are other methods for estimating distribution parameters – *least squares*, *maximum likelihood*, *maximum entropy* – that will not be presented here.

Interval estimates

Instead of single numbers for the estimates of the unknown mean and variance, it is often interesting to determine *intervals* of values in which the true values of these parameters are most likely to be found. This approach is called *interval estimates*. As such estimates are used to indicate the precision or accuracy of a point estimate, this approach is also sometimes referred to as *confidence estimates*.

For example, assuming that the sampling distribution of means obeys a normal distribution (see p. 37), the 95% confidence interval for estimation of the population mean μ_X from a sample of large size is given by $\mu_{\overline{X}} \pm 1.96\sigma_{\overline{X}}$, which can also be written:

$$P(\mu_{\overline{X}} - 1.96\sigma_{\overline{X}} \le \mu_X \le \mu_{\overline{X}} + 1.96\sigma_{\overline{X}}) = 0.95 \qquad [2.62]$$

More generally, the confidence limits are given by $\mu_{\overline{X}} \pm t_S \cdot \sigma_{\overline{X}}$, where t_S is obtained from a table of *Student's t-distribution* for two-sided confidence interval estimation. Such a table is given in Table 2.1 for the case of a large sample size *n*.

Table 2.1 ***Confidence levels for the mean of a normal distribution*** **[McCormick, 1981]**

Two-sided confidence level [%]	One-sided confidence level [%]	t_S
99.73	99.86	3.000
99.00	99.50	2.580
98.00	99.00	2.330
96.00	98.00	2.050
95.45	97.72	2.000
95.00	97.50	1.960
90.00	95.00	1.645
80.00	90.00	1.280
68.27	84.14	1.000
50.00	75.00	0.6745

Example: for the sample of 100 ball bearings considered previously (p. 37), drawn without replacement, calculate the limits of the: a) 80%, b) 95% confidence interval.

With $\mu_{\overline{X}} = 12$ g, $\sigma_{\overline{X}} = 0.027$ g and, from Table 2.1, t_s being equal to 1.28 in the first case (a) and 1.96 in the second case (b), we can write:

a) $P(\mu_{\overline{X}} - t_{S_a} \cdot \sigma_{\overline{X}} \le \mu_X \le \mu_{\overline{X}} + t_{S_a} \cdot \sigma_{\overline{X}}) = P(11.965 \le \mu_X \le 12.035) = 0.80$

b) $P(\mu_{\overline{X}} - t_{S_b} \cdot \sigma_{\overline{X}} \le \mu_X \le \mu_{\overline{X}} + t_{S_b} \cdot \sigma_{\overline{X}}) = P(11.947 \le \mu_X \le 12.053) = 0.95$

A one-sided interval estimate may also be calculated. Because the normal distribution is symmetric, the one-sided estimates corresponding to Eq. [2.62] are given by:

$$P(\mu_X \le \mu_{\overline{X}} + 1.96\sigma_{\overline{X}}) = P(\mu_X \ge \mu_{\overline{X}} - 1.96\sigma_{\overline{X}}) = 0.975$$

The one-sided confidence levels from Student's *t*-distribution are also listed in Table 2.1 for a sample of large size.

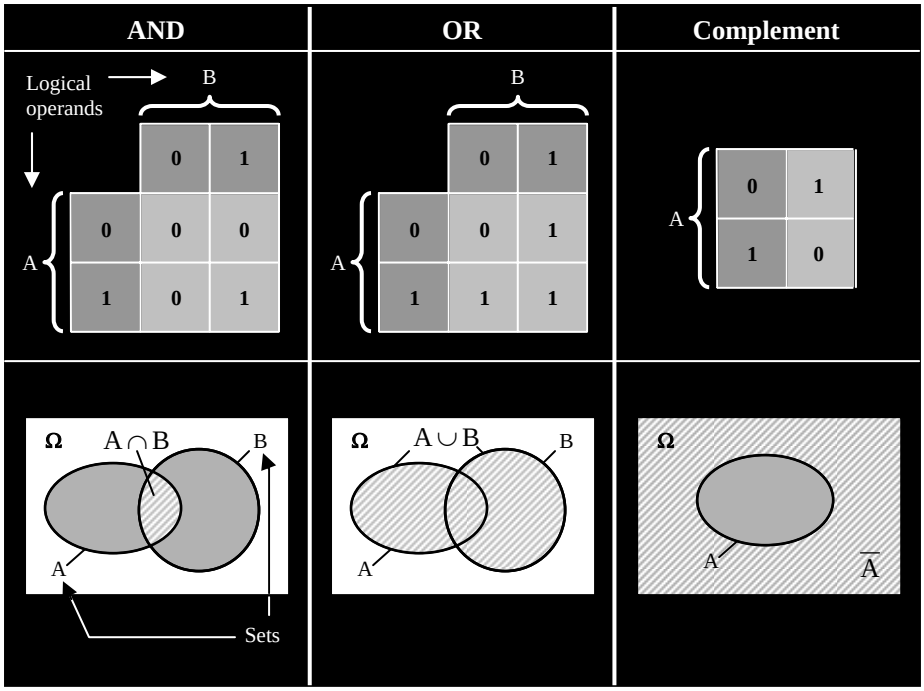
2.3 Elements of Event Algebra (Boolean Algebra)

Risk analysis is about events and their probabilities of occurrence. When different events are involved, we need rules to combine them. In other words, we need an appropriate algebra for events. This mathematical tool is the *Boolean algebra*, from George Boole, English mathematician (1815-1864). George Boole made major contributions to the development of mathematical logic and published a book *The Mathematical Analysis of Logic* in 1847. Boole believed in what he called the 'process of analysis', that is, the process by which combinations of interpretable symbols are obtained. It is the use of these symbols according to well-determined methods of combination that he believed presented 'true calculus'.

Gearge Boole
(1815-1864)

Boolean algebra is thus defined as the study of the manipulation of symbols representing operations according to the rules of logic. The main ingredient in logic analysis is the principles and method used to distinguish between arguments that are valid ("*true*") and those that are not ("*false*"). Logic deals with reasoning and the ability to deduce or come to some appropriate conclusions. Boolean algebra involves the operations of *intersection* ("AND", symbol "$\cap$"), *union* ("OR", symbol "$\cup$") and *complement* ("NOT", symbol "$\bar{}$") on sets. These are defined in Table 2.2, using two types of representation: *truth tables* (with logical operands having the values "1" when" true" and "0" when "false") and *Venn diagrams* (graphical representation of sets by overlapping oval shapes).

Table 2.2 ***Definition of the Boolean operators "AND", "OR" and "complement"***



Ω is the *universal event*; its complement is the *null event*, noted Ø. The universal event corresponds to the union of all possible events. The null event corresponds to the intersection of disjoint events. If the intersection of an event with another one returns the first event itself, this one is said to be *included* in the second (symbol "$\subset$").

The most obvious way to simplify Boolean expressions is to manipulate them in the same way as normal algebraic expressions are manipulated. A set of rules for symbolic manipulation is needed to this end.

Basic Boolean laws

The basic Boolean laws are presented below. Note that every law has two expressions (a) and (b). These are obtained bay changing every "AND" to "OR" and vice-versa. This is known as *duality*.

- **Commutative law**

$A \cup B = B \cup A$ [2.63 a]

$A \cap B = B \cap A$ [2.63 b]

- **Associative law**

$(A \cup B) \cup C = A \cup (B \cup C)$ [2.64 a]

$(A \cap B) \cap C = A \cap (B \cap C)$ [2.64 b]

- **Distributive law**

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ [2.65 a]

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ [2.65 b]

- **Idempotent law**

$A \cup A = A$ [2.66 a]

$A \cap A = A$ [2.66 b]

- **Absorption law**

$A \cup (A \cap B) = A$ [2.67 a]

$A \cap (A \cup B) = A$ [2.67 b]

Additional useful relationships

Some additional frequently used relationships can also be mentioned:

- **Complementation**

$A \cap \overline{A} = \varnothing$ [2.68 a]

$A \cup \overline{A} = \Omega$ [2.68 b]

$\overline{\overline{A}} = A$ [2.68 c]

- **Operations with $\varnothing$ and Ω**

$\varnothing \cap A = \varnothing$ [2.69 a]

$\varnothing \cup A = A$ [2.69 b]

$\Omega \cap A = A$ [2.69 c]

$\Omega \cup A = \Omega$ [2.69 d]

- **Simplification**

$(A \cap B) \cup (A \cap \overline{B}) = A$ [2.70 a]

$(A \cup B) \cap (A \cup \overline{B}) = A$ [2.70 b]

$A \cup (\overline{A} \cap B) = A \cup B$ [2.70 c]

$A \cap (\overline{A} \cup B) = A \cap B$ [2.70 d]

$\overline{A} \cap (A \cup \overline{B}) = \overline{A} \cap \overline{B} = (\overline{A \cup B})$ [2.70 e]

Morgan's theorem

The last equality results from a well-known theorem, often used for simplifying Boolean expressions:

- **Morgan's theorems**

$(\overline{A \cap B}) = \overline{A} \cup \overline{B}$ [2.71 a]

$(\overline{A \cup B}) = \overline{A} \cap \overline{B}$ [2.71 b]

There are several common alternative notations for the Boolean operators, e.g.:

- $A \cap B$ can also be written $A \cdot B$, or more simply AB,
- $A \cup B$ can also be written $A+B$,
- $\overline{A}$ can also be written A', or sometimes $\neg A$.

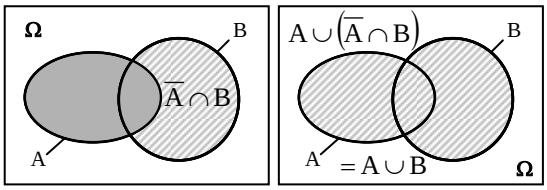
The laws, theorems and relationships listed in page 41 can be proved by using truth tables or Venn diagrams.

Example: prove the truthfulness of the relationship $A \cup (\overline{A} \cap B) = A \cup B$.

a) Using the truth table approach:

A	B	$\overline{A}$	$\overline{A} \cap B$	$A \cup (\overline{A} \cap B)$	$A \cup B$
0	0	1	0	0	0
0	1	1	1	1	1
1	0	0	0	1	1
1	1	0	0	1	1

b) Using the Venn diagram approach:



These laws, theorems and relationships are used to simplify Boolean expressions.

Example: simplify the expression $Z = (A \cup \overline{B} \cup \overline{C}) \cap [A \cup (\overline{B} \cap C)]$:

$$Z = (A \cap A) \cup (A \cap \overline{B} \cap C) \cup (A \cap \overline{B}) \cup (\overline{B} \cap \overline{B} \cap C) \cup (A \cap \overline{C}) \cup (\overline{B} \cap C \cap \overline{C})$$
$$Z = A \cup (A \cap \overline{B} \cap C) \cup (A \cap \overline{B}) \cup (\overline{B} \cap C) \cup (A \cap \overline{C}) \cup (\overline{B} \cap \varnothing)$$
$$Z = \{A \cap [\Omega \cup (\overline{B} \cap C) \cup \overline{B} \cup \overline{C}]\} \cup (\overline{B} \cap C) = (A \cap \Omega) \cup (\overline{B} \cap C)$$
$$Z = A \cup (\overline{B} \cap C)$$

Event algebra and probabilities

An application P that associates a positive real number to any event included in an universal set Ω, with the following properties:

- $0 \le P(A) \le 1$,
- $P(\Omega) = 1$ and $P(\varnothing) = 0$,
- $P(\overline{A}) = 1 - P(A)$,
- $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \varnothing$,
- $P(A) \le P(B)$ if $A \subset B$,

defines a probability relationship (see Fig. 2.9).

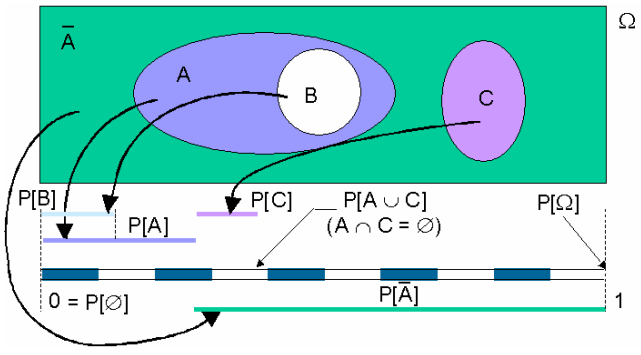


Figure 2.9 ***Definition of a probability relationship on a set of events***

Intersection of events

The *product rule for probabilities* states that:

$$P(A_1 \cap A_2) = P(A_1|\, A_2)\cdot P(A_2) = P(A_2|\, A_1)\cdot P(A_1) \qquad [2.72]$$

$P(A_i|\, A_j)$ represents the *conditional probability* of event A_i given that event A_j has occurred. In the special case that events A_i and A_j are *independent* - i.e. the probability that event A_i occurs is independent of the occurrence of event A_j - then $P(A_i|\, A_j)$ is simply equal to $P(A_i)$.

A second particular case corresponds to events A_i and A_j that are mutually exclusive (i.e. "disjoint"), in which case $P(A_i|\, A_j) = 0$, and therefore $P(A_i \cap A_j) = 0$.

The product rule may be easily generalized to the case of n ($n \geq 2$) events:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1|A_2 \cap A_3 \cap \ldots \cap A_n) \ldots P(A_{n-1}|\, A_n)\cdot P(A_n) \qquad [2.73]$$

If the n events are all independent, then:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\cdot P(A_2) \ldots P(A_n) \qquad [2.74]$$

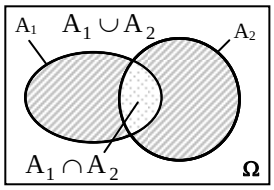
And if they are mutually exclusive:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = 0 \qquad [2.75]$$

Union of events

The general equation expressing the union of two events, $A_1 \cup A_2$, is:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \qquad [2.76]$$



The last term of the right-hand side of the equation appears because of the necessity to avoid possible double counting resulting from the "overlap" caused by the intersection of the two events.

Of course, if the two events are independent, it follows from Eqs. [2.74] and [2.76] that:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1)\cdot P(A_2) \qquad [2.77]$$

On the other hand, if the two events are mutually exclusive, then:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \qquad [2.78]$$

As in the case of the intersection of events, the preceding equations can be generalized to the case of more than two events (*Poincaré's theorem*):

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) =$$

$$\sum_{i=1}^{n} P(A_i) - \sum_{j=2}^{n}\sum_{i=1}^{j-1} P(A_i \cap A_j) + \sum_{j=3}^{n}\sum_{k=2}^{j-1}\sum_{i=1}^{k-1} P(A_i \cap A_j \cap A_k)$$

$$- \ldots + (-1)^n \cdot P(A_1 \cap A_2 \cap \ldots \cap A_n) \qquad [2.79]$$

If the events are all mutually exclusive, then only the first term of the right-hand side of Eq, [2.79] is non vanishing.

If all the events are independent, it is easier to calculate the probability of the union of events from the probability of its complement (application of Morgan's theorem):

$$P(\overline{A_1 \cup A_2 \cup \ldots \cup A_n}) = 1 - P(A_1 \cup A_2 \cup \ldots \cup A_n)$$

$$= P(\overline{A}_1 \cap \overline{A}2 \cap \ldots \cap \overline{A}_n) = \prod_{i=1}^{n}\left[1 - P(A_i)\right] \qquad [2.80]$$

2.4 Bayes' Theorem and Bayesian Inference

Rev. Thomas Bayes
(1702-1761)

The "Bayesian inference method" is based on the Bayesian theorem, which suggests that the degree to which one believes that a proposition is true depends on the a priori belief one has in the truth of the proposition and in the evidence collected to investigate this one.

Bayes' theorem is named after Rev. Thomas Bayes, an 18th century English mathematician who derived a special case of this theorem. Bayes' calculations were published in 1763, two years after his death. Exactly what Bayes intended to do with the calculation, if anything, still remains a mystery today. However, his theorem, as generalized by Laplace, is the basic starting point for inference problems using probability theory as logic.

Bayes' theorem describes the relationships that exist within a set of simple and conditional probabilities. Although its primary application is to situations where "probability" is defined according to the strict relative frequency interpretation of the concept, it perhaps more often applied to situations where "probability" is constructed as an index of subjective confidence (see p. 22).

In the classical statistical approach, model parameters are fixed but unknown constants to be estimated using sample data taken randomly from the population of interest. The Bayesian approach, on the other hands, treats these population model parameters as random, not fixed, quantities. Previous information, or even subjective judgments, are used to construct a *prior distribution model* for these parameters. This model expresses a starting assessment about how likely various values of the unknown parameters are. It is then made use of the current data (via Bayes' formula) to revise this starting assumption and derive what is called the *posterior distribution model* for the population parameters. Parameter estimates, along with confidence intervals, are then calculated from the posterior distribution.

Bayes' fundamental *inverse probability formula* derives from the "product rule for probabilities" (Eq. [2.72]):

$$P(A \cap B) = P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$$

To calculate the probability for event A that incorporates the additional evidence provided by the occurrence of B, the last equality in the above expression is solved to give:

$$P(A \mid B) = P(A) \cdot \left[\frac{P(B \mid A)}{P(B)} \right] \qquad [2.81]$$

The formula expresses that the conditional probability of an event A occurring, given that the event B has occurred, written P(A | B), is equal to the prior unconditional probability of occurrence of A multiplied by a "correction factor" which represents the relative change in the probability of A when B is known to have happened.

A more general formulation of the Bayes' theorem can be developed for a complete set of mutually exclusive events A_i (i = 1, … n); such a set is characterized by the fact that (*theorem of total probabilities*):

$$P\left(\bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} P(A_i) = 1 \qquad [2.82]$$

The above equation remains valid if the probabilities are made conditional to the occurrence of a given event B (this property is general):

$$\sum_{i=1}^{n} P(A_i \mid B) = 1 \qquad [2.83]$$

If this equation is multiplied by P(B), then (with Eq. [2.72]):

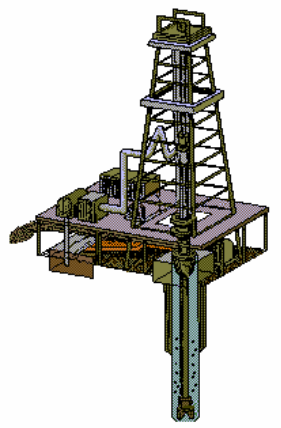
$$P(B) = \sum_{i=1}^{n} P(A_i \mid B) \cdot P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(B \mid A_i) \cdot P(A_i) \qquad [2.84]$$

Equation [2.84] is the *extension rule* of probabilities. It allows P(B) to be expressed in terms of the previously known probabilities $P(A_i)$ and all the conditional probabilities $P(B \mid A_i)$. Introduced in Eq. [2.81], this leads to the final form of the Bayes' formula:

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j) \cdot P(B \mid A_j)} \qquad [2.85]$$

The high degree of symmetry in the last equation shows that once the entire set of probabilities $P(B \mid A_i)$ is known, then the calculation of the posterior $P(A_i \mid B)$ becomes straightforward.

The Bayes' formula is an important tool in reliability analysis when one specifies, by the subjective approach, the possibility of rarely occurring events, because it enables one to "reverse" the order of information gathering about a failure process.



Example of application of Bayes' theorem: consider two oil prospects (drills) A and B, having respectively a chance of success P(A) and P(B). These prospects are on the same basin (i.e. they share a common source rock), therefore a success of A will cause us to revise the chance of success of B and conversely for a failure of A. Calculate the conditional probability of success of prospect B, given that late case (failure of A), assuming that P(A) = 0.3, P(B) = 0.2 and P(B | A) = 0.6.

The probability we are looking for is $P(B \mid \overline{A})$. From the first form of the Bayes' formula (Eq. [2.81]), we have:

$$P(B \mid \overline{A}) = P(B) \cdot \left[\frac{P(\overline{A} \mid B)}{P(\overline{A})} \right]$$

Now, $P(\overline{A}) = 1 - P(A) = 1 - 0.3 = 0.7$,

and $P(A \mid B) + P(\overline{A} \mid B) = 1$ (because $P(A) + P(\overline{A}) = 1$ from Eq. [2.68 b]),

thus: $P(\overline{A} \mid B) = 1 - P(A \mid B) = 1 - P(A) \cdot \left[\frac{P(B \mid A)}{P(B)} \right] = 1 - (0.3 \cdot 0.6 / 0.2)$,

i.e. $P(\overline{A} \mid B) = 0.1$.

Finally: $P(B \mid \overline{A}) = (0.2 \cdot 0.1) / 0.7 \cong 0.03$.

Taking into account the knowledge that A has failed thus lower the probability of having a successful prospect B by a factor of seven compared to the prior unconditional evaluation of the success probability of B. This is probably much less that most people would have guessed.

The Bayes' formula can also be written in terms of probability density functions as follows:

$$g(\lambda \mid x) = \frac{f(x \mid \lambda) \cdot g(\lambda)}{\int_{\lambda_{min}}^{\lambda_{max}} f(x \mid \lambda') \cdot g(\lambda')\, d\lambda'} \qquad [2.86]$$

where $f(x \mid \lambda)$ is the probability model, or *likelihood function*, for the observed data x given the unknown parameter(s) λ, $g(\lambda)$ is the *prior distribution* model for λ and $g(\lambda \mid x)$ is the *posterior distribution* model for λ given that the data *x* have been observed.

A common way to construct a prior distribution is to assume that the prior is a member of a particular parametric family of densities, then choose the parameter of the prior so that the prior represents prior beliefs as closely as possible. When possible, it is very convenient to choose the prior from a parametric family that has the same functional form as the likelihood function; $g(\lambda)$ and $f(x \mid \lambda)$ are in this case called *conjugate distributions* and $g(\lambda)$ is the *conjugate prior* for $f(x \mid \lambda)$.

Example 1: estimating the binomial probability *p*.

Let us consider a binomial process with parameter *p* and assume that we have undertaken *n* trials and obtained *s* success. Estimate the uncertainty about the value of *p*.

Initially, we assume a uniform prior for *p*, that is no prior information about *p*; thus, $g(p) = 1$ $(0 \le p \le 1)$. The likelihood function is the binomial probability:

$$f(s \mid p) = C_s^n\, p^s\, (1-p)^{n-s}$$

and the posterior distribution is therefore given by (Eq. [2.86]):

$$g(p \mid s) = \frac{p^s \cdot (1-p)^{n-s}}{\int_0^1 p'^{\,s} \cdot (1-p')^{n-s}\, dp'}$$

which is the probability density of a beta$(s + 1, n - s + 1)$ distribution in the interval [0, 1]. Note that the beta distribution should only be used to describe the uncertainty about *p* if the sample size *n* is much smaller than the population size, because we assume sampling with replacement.

If, instead of an uniformed prior, we consider now a beta(α, β) prior (conjugate prior), i.e.:

$$g(p) = \frac{p^{(\alpha-1)} \cdot (1-p)^{(\beta-1)}}{\int_0^1 p'^{\,(\alpha-1)} \cdot (1-p')^{(\beta-1)}\, dp'}$$

the posterior distribution becomes:

$$g(p \mid s) = \frac{p^{(\alpha+s-1)} \cdot (1-p)^{(\beta+n-s-1)}}{\int_0^1 p'^{\,(\alpha+s-1)} \cdot (1-p')^{(\beta+n-s-1)}\, dp'}$$

which is a beta$(\alpha + s, \beta + n - s)$ distribution. The posterior belief is in this case just a "rescaling" of the prior belief.

Example 2: estimating the Poisson intensity λ (mean rate at which events occurred).

Let us assume that α events have been observed during a time period t. Estimate the uncertainty associated with λ.

Initially, we assume an uninformed prior for λ, that is:

$$g(\lambda) = \begin{cases} 1/L & 0 \le \lambda \le L \\ 0 & \text{otherwise} \end{cases}$$

where L is some large number representing an upper limit for the possible values of λ.

The likelihood function is here the Poisson probability:

$$f(\alpha \mid \lambda) = e^{-\lambda t} \cdot \frac{(\lambda t)^{\alpha}}{\alpha!}$$

The posterior distribution is thus given by:

$$g(\lambda \mid \alpha) = \frac{e^{-\lambda t} \cdot \lambda^{\alpha}}{\int_0^L e^{-\lambda' t} \cdot \lambda'^{\alpha} \, d\lambda'}$$

Now, because L is large, we can go to the limit $L \to \infty$ and thus replace the denominator in the above equation by (see Appendix 2.1):

$$\int_0^{\infty} e^{-\lambda' t} \cdot \lambda'^{\alpha} \, d\lambda' = \frac{\Gamma(\alpha+1)}{t^{(\alpha+1)}}$$

Therefore, the posterior distribution becomes:

$$g(\lambda \mid \alpha) = \frac{t^{\alpha+1} \cdot e^{-\lambda t} \cdot \lambda^{\alpha}}{\Gamma(\alpha+1)}$$

which is a gamma(α+1, t) distribution.

In short, Bayesian inference consists in:

1. determining a prior estimate of the parameter in the form of a probability density function;
2. finding an appropriate likelihood function for the observed data;
3. calculating the posterior (revised) estimate of the parameter by multiplying the prior distribution and likelihood function and then normalizing.

One issue of concern in Bayesian inference is how strongly the particular selection of a prior distribution influences the results of the process. An uninformative prior is one that provides little or no information. Depending on the situation, uninformative priors may be quite disperse or lead to impossible or quite preposterous values of the parameter. Particularly when results are to be used by people who may question the expert's initial opinion, it is desirable to have enough data available to make the influence of the prior choice slight.

Appendix 2.1 Main Probability Distribution Functions

Distribution	Probability Density Function Cumulative Distribution Function		Mean Variance
Discrete			
Binomial	$p_X(x_i) = \frac{n!}{x_i!(n-x_i)!} p^{x_i} (1-p)^{n-x_i}$	$x_i = 0,1,2,\ldots n$	$E[X] = n \cdot p$
	$P(X \le x) = \sum_{x_i \le x} p_X(x_i)$		$V[X] = n \cdot p \cdot (1-p)$
Geometric	$p_X(x_i) = p \cdot (1-p)^{x_i - 1}$	$x_i = 0,1,2,\ldots n$	$E[X] = 1/p$
	$P(X \le x) = \sum_{x_i \le x} p_X(x_i)$		$V[X] = (1-p)/p^2$
Poisson	$p_X(x_i) = e^{-\lambda} \cdot \frac{\lambda^{x_i}}{x_i!}$	$x_i = 0,1,2,\ldots n$	$E[X] = \lambda$
	$P(X \le x) = \sum_{x_i \le x} p_X(x_i)$		$V[X] = \lambda$
Hypergeometric	$p_X(x_i) = \frac{C_{x_i}^{S} \cdot C_{n-x_i}^{N-S}}{C_n^N}$	$0 \le x_i \le \min[S, n]$	$E[X] = n \cdot \frac{S}{N}$
	$P(X \le x) = \sum_{x_i \le x} p_X(x_i)$		$V[X] = \left\{ n \cdot \frac{S}{N} \cdot \left(1 - \frac{S}{N}\right) \right\} \cdot \frac{N-n}{N-1}$
Continuous			
Uniform	$f_X(x) = \frac{1}{b-a}$	$a \le x \le b$	$E[X] = \frac{b+a}{2}$
	$F_X(x) = \frac{x}{b-a}$		$V[X] = \frac{(b-a)^2}{12}$
Normal	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$-\infty < x < +\infty$	$E[X] = \mu$
	$F_X(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z} e^{-u^2} du = \frac{1}{2}[1 - \mathrm{erf}\lvert z\rvert], \quad x < \mu$		$V[X] = \sigma^2$
	$= \frac{1}{2}[1 + \mathrm{erf}(z)], \quad x > \mu$		
	where erf(z) is the *error function* [McCormick, 1981]:		
	$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-u^2} du$		
	and: $z = (x-\mu)/(\sqrt{2} \cdot \sigma)$		

Distribution	Probability Density Function Cumulative Distribution Function	Mean Variance
Lognormal	$f_X(x) = \frac{e^{-(\ln x - \mu)^2 / 2\sigma^2}}{x \cdot \sigma \cdot \sqrt{2\pi}}$ $x \geq 0$ $F_X(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z} e^{-u^2} du = \frac{1}{2}[1 - \text{erf}\lvert z\rvert],\ \ln x < \mu$ $= \frac{1}{2}[1 + \text{erf}(z)],\ \ln x > \mu$ with: $z = (\ln x - \mu)/(\sqrt{2} \cdot \sigma)$	$E[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$ $V[X] = E^2[X] \cdot \left(e^{\sigma^2} - 1\right)$
Exponential	$f_X(x) = \lambda \cdot e^{-\lambda \cdot x}$ $x \geq 0$ $F_X(x) = 1 - e^{-\lambda \cdot x}$	$E[X] = 1/\lambda$ $V[X] = 1/\lambda^2$
Weibull	$f_X(x) = \frac{\beta}{\alpha} \cdot \left(\frac{x - \mu}{\alpha}\right)^{(\beta - 1)} \cdot \exp\left\{-[(x - \mu)/\alpha]^{\beta}\right\}$ $0 \leq \mu \leq x \leq \infty$ $F_X(x) = 1 - \exp\{-[(x - \mu)/\alpha)^{\beta}\}$	$E[X] = \mu + \alpha \cdot \Gamma(1 + \beta^{-1})$ $V[X] =$ $\alpha^2 \cdot \{\Gamma(1 + 2\beta^{-1}) - [\Gamma(1 + \beta^{-1})]^2\}$ where $\Gamma(x)$ is the *gamma function*: $\Gamma(x) = \int_0^{\infty} y^{x-1} \cdot e^{-y}\, dy$
Beta	$f_X(x) = \frac{1}{B(\alpha, \beta)} \cdot \frac{(x - a)^{\alpha - 1} \cdot (b - x)^{\beta - 1}}{(b - a)^{\alpha + \beta - 1}}$ $a \leq x \leq b$ where : $B(\alpha, \beta) = \int_0^1 u^{(\alpha - 1)} \cdot (1 - u)^{(\beta - 1)}\, du$ $F_X(x) = \frac{1}{B(\alpha, \beta)} \cdot \int_a^x \frac{(x - a)^{\alpha - 1} \cdot (b - x)^{\beta - 1}}{(b - a)^{\alpha + \beta - 1}}\, dx$	$E[X] = a + \frac{\alpha}{\alpha + \beta} \cdot (b - a)$ $V[X] = \frac{\alpha \cdot \beta}{(\alpha + \beta)^2 \cdot (\alpha + \beta + 1)} \cdot (b - a)^2$