Appendix A

Probability theory

A.1 Basic probability notation

In this chapter basic elements of probability theory are reviewed. Readers familiar with probability theory can skip this chapter. Readers which are very unfamiliar with this topic are advised to read an introductory textbook in probability theory.

A.1.1 Event

In order to define probability, we need to work with events. Let as an example A be the event that there is an operator error in a control room. This is written:

$$A = \{\text{operator error}\}$$

An event may occur, or not. We do not know the outcome in advance prior to the experiment or a situation in the "real life". We also use the word event to denote a set of distinct events. For example the event that we get an even number when throwing a die.

A.1.2 Probability

When events are defined, the probability that the event occurs is of interest. Probability is denoted by $\Pr(\cdot)$, i.e.,

$$\Pr(A) = \text{Probability that } A \text{ occur}$$

The numeric value of $\Pr(A)$ may be found by:

- Studying the *sample space*.

- Analysing collected data.
- Look up the value in data hand books.
- "Expert judgement" **?**.

The *sample space* defines all possible events. As an example let A = {It is Sunday}, B = {It is Monday}, .. , G = {It is Saturday}. The sample space is then given by $S = \{A, B, C, D, E, F, G\}$.

So-called Venn diagrams are useful when we want to analyse a subset of the sample space S. A rectangle represents the entire sample space, and closed curves such as a circle are used to represent subsets of the sample space as illustrated in Figure **??**. In the following we will illustrate

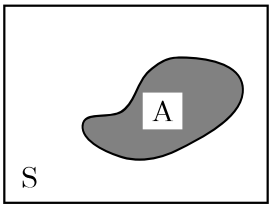
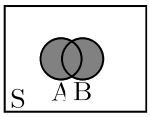
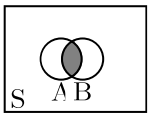


Figure A.1: Venn diagram

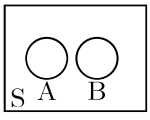
frequently used combinations of events:



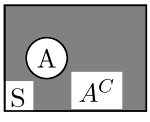
Union. We write $A \cup B$ to denote the union of A and B, i.e., the occurrence of A or B or (A and B). Let A be the event that tossing a die results in a "six", and B be the event that we get an odd number of eyes. We then have $A \cup B = \{1, 3, 5, 6\}$.



Intersection. We write $A \cap B$ to denote the intersection of A and B, i.e. the occurrence of both A and B. As an example, let A be the event that a project is not completed in due time, and let B be the event that the budget limits are exceeded. $A \cap B$ then represent the situation that the project is not completed in due time and the budget limits are exceeded.



Disjoint events. A and B are said to be *disjoint* if they can *not* occur simultaneously, i.e. $A \cap B = \emptyset$ = the empty set. Let A be the event that tossing a die results in a "six", and B be the event that we get an odd number of eyes. A and B are disjoint since they cannot occur simultaneously, and we have $A \cap B = \emptyset$.



Complementary events. The *complement* of an event A is all events in the sample space S except for A. The complement of an event is denoted by A^C. Let A be the event that tossing a die results in an odd number of eyes. A^C is then the event that we get an even number of eyes.

A.1.3 Probability and Kolmogorov's axioms

Probability is a set function Pr() which maps events A_1, $A_{2,}$... in the sample space S to real numbers. The function Pr(·) can only take values in the interval from 0 to 1, i.e. probabilities are greater or equal than 0, and less or equal than 1. Kolmogorov established the following axioms

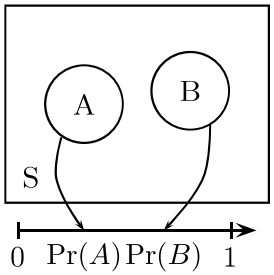


Figure A.2: Mapping of events on the interval [0,1]

which all probability rules could be derived from:

1. $0 \leq \Pr(A)$
2. $\Pr(S) = 1$
3. If A_1, A_2, A_3,... is a sequence of disjoint events we shall then have:
   $\Pr(A_1 \cup A_2 \cup \ldots) = \Pr(A_1) + \Pr(A_2) + \ldots$

The axioms are the basis for establishing calculation rules when dealing with probabilities, but they do not help us in establishing numerical values for the basic probabilities $\Pr(A_1)$, $\Pr(A_2)$, etc. Historically two lines of thoughts have been established, the classical (frequentiest) and the Bayesian approach. In the classical thinking we introduce the concept of a random experiment, where $\Pr(A_i)$ is the relative frequency with which the event A_i occurs. The probability could then be interpreted as a property of the experiment, or a property of the world. By letting nature reveal itself by doing experiments, we could in principle establish all probabilities that are of interest. Within the Bayesian framework probabilities are interpreted as subjective believe about whether A_i will occur or not. Probabilities is then not a property of the world, but rather a measure of the knowledge and understanding we have about a phenomenon.

Before we set up the basic rules for probability theory that we will need, we introduce the concepts of conditional probability and independent events.

Conditional probability. $\Pr(A|B)$ denotes the conditional probability that A will occur given that B has occurred.

Independent events. A and B are said to be *independent* if information about whether B has occurred does *not* influence the probability that A will occur, i.e., $\Pr(A|B) = \Pr(A)$.

Basic rules for probability. The following calculation rules for probability apply:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \tag{A.1}$$

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B) \text{ if A and B are independent} \tag{A.2}$$

$$\Pr(A^C) = \Pr(A \text{ does } \textit{not} \text{ occur}) = 1 - \Pr(A) \tag{A.3}$$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{A.4}$$

Example

Let the two events A and B be defined by A = {It is Sunday} and B = {It is between 6 and 8 pm).

First we note that A and B are independent but not disjoint. We will find $\Pr(A \cap B)$, $\Pr(A \cup B)$ and $\Pr(A|B)$

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B) = \frac{1}{7} \cdot \frac{2}{24} = \frac{1}{84}$$

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = \frac{1}{7} + \frac{2}{24} - \frac{1}{84} = \frac{9}{42}$$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/84}{2/24} = \frac{1}{7}$$

□

A.1.4 The law of total probability

In many situations it is easier to assess the probability of an event B conditionally on some other events, say $A_1, A_2, \ldots, A_r$, than unconditionally. The law of total probability could then be used to assess the unconditional probability. Now, we say that $A_1, A_2, \ldots, A_r$ is a division of the sample space if the union of all A_i's covers the entire sample space, i.e. $\mathrm{A}_1 \cup \mathrm{A}_2 \cup \ldots \cup \mathrm{A}_r = \mathrm{S}$ and the A_i's are pair wise disjoint, i.e. $\mathrm{A}_i \cap \mathrm{A}_j = \emptyset$ for $i \neq j$. An example is shown in Figure **??**.

Let $A_1, A_2, \ldots, A_r$ represent a division of the sample space S, and let B be an arbitrary event in S. The law of total probability now states:

$$\Pr(B) = \sum_{i=1}^{r} \Pr(A_i) \cdot \Pr(B|A_i) \tag{A.5}$$

Example

A special component type is ordered from two suppliers A_1 and A_2. Experience has shown that

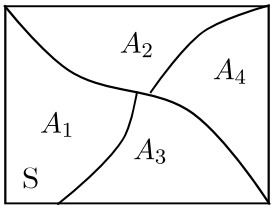


Figure A.3: Divition of the sample space

components from supplier A_1 has a defect probability of 1%, whereas components from supplier A_2 has a defect probability of 2%. In average 70% of the components are provided by supplier A_1. Assume that all components are put on a common stock, and we are not able to trace the supplier for a component in the stock. A component is now fetched from the stock, and we will calculate the defect probability, $\Pr(B)$:

$$\Pr(B) = \sum_{i=1}^{r} \Pr(A_i) \cdot \Pr(B|A_i) = \Pr(A_1) \cdot \Pr(B|A_1) + \Pr(A_2) \cdot \Pr(B|A_2) = \\ 0.7 \cdot 0.01 + 0.3 \cdot 0.02 = 1.3\%$$

□

A.1.5 Bayes theorem

Now consider the example above, and assume that we have got a defect component from the stock (event B). We will derive the probability that the component originates from supplier A_1. We then use Bayes formula that states if $A_1, A_2, \ldots, A_r$ represent a division of the sample space, and B is an arbitrary event then:

$$\Pr(A_j|B) = \frac{\Pr(B|A_j) \cdot \Pr(A_j)}{\sum_{i=1}^{r} \Pr(A_i) \cdot \Pr(B|A_i)} \tag{A.6}$$

Example
We have

$$\Pr(A_1|B) = \frac{\Pr(B|A_1) \cdot \Pr(A_1)}{\sum_{i=1}^{r} \Pr(A_i) \cdot \Pr(B|A_i)} = \frac{0.01 \cdot 0.7}{0.013} = 0.54$$

Thus, the probability of A_1 is reduced from 0.7 to 0.54 when we know that the component is defect. The reason for this is that components from supplier A_1 are the best ones, and hence when we know that the component was defect, it is less likely that it was from supplier A_1. □

A.1.6 Stochastic variables

Stochastic variables are used to describe quantities which can not be predicted exactly. Note that the term *'random quantity'* is often used to denote a stochastic variable.

$$X \text{ is stochastic} \Leftrightarrow \text{Cannot say precisely the value } X \text{ has or will take}$$

To be more precise, a stochastic variable X is a real valued function that assigns a quantitative measure to each event e_i in the sample space S, i.e., $X = X(e_i)$. Often the underlying events, e_i are of little interest. We are only interested in the stochastic variable X measured by some means.

Examples of stochastic variables are given below:

- X = Life time of a component (continuous)
- R = Repair time after a failure (continuous)
- T = Duration of a construction project (continuous)
- C = Total cost of a renewal project (continuous)
- M = Number of delayed trains next month (discrete)
- N = Number of customers arriving today (discrete)
- S = Service time for the first customer arriving today (continuous)
- W = Maintenance and operational cost next year (continuous)

Remark: We distinguish between *continuous* and *discrete* stochastic variables. Continuous stochastic variables can take any value among the real numbers, whereas discrete variables can take only a *finite* (or countable finite) number of values. □

Cumulative distribution function (CDF). A stochastic variable X is characterized by it's *cumulative distribution function*:

$$F_X(x) = \Pr(X \leq x) \tag{A.7}$$

We use subscript X to emphasise the relation to the cumulative distribution function of the quantity X. The argument (lowercase x) states which values the stochastic variable X could take, or is of our interest. From the expression we observe that $F_X(x)$ states the probability that the random quantity X is less or equal than (the numeric value of) x. A typical distribution function is shown in Figure **??**. Note that the distribution function is strictly increasing, and $0 \leq F_X(x) \leq 1$. From $F_X(x)$ we can obtain the probability that X will be within a specified interval, $[a,b)$:

$$\Pr(a < X \leq b) = F_X(b) - F_X(a) \tag{A.8}$$

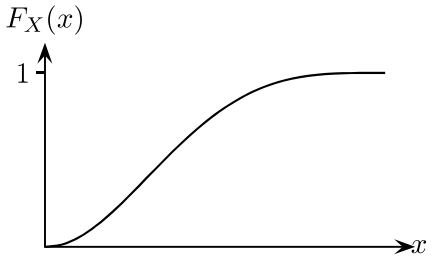


Figure A.4: Cumulative distribution function, $F_X(x)$

Note that the index X representing the stochastic variable X often is dropped if it is obvious which stochastic variable we are working with. Note also the distinction between lowercase and uppercase letters. The uppercase X is used to denote a stochastic variable, for example number of customers arriving next day. The lowercase x is just a representation of possible values X can take. For example $X = 3$.

Example

Assume that the probability distribution function of X is given by $F_X(x) = 1 - e^{-(0.01x)^2}$, and we will find the probability that X is in the interval (100,200]. From Equation (**??**) we have:

$$\Pr(100 < X \leq 200) = F_X(200) - F_X(100) =$$
$$\left[1 - e^{-(0.01 \cdot 200)^2}\right] - \left[1 - e^{-(0.01 \cdot 100)^2}\right] = e^{-1} - e^{-4} \approx 0.35$$

□

Probability density function (PDF). For a continuous stochastic variable, the *probability density function* is given by

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{A.9}$$

The probability density function expresses how likely the various x-values are.

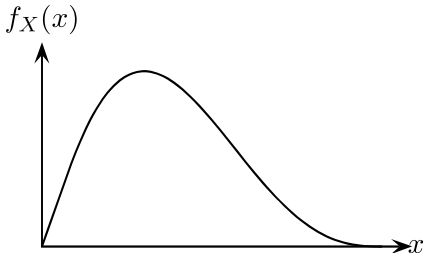


Figure A.5: Probability density function, $f_X(x)$

Note that for continuous random variables the probability that X will take a specific value vanishes. However, the probability that X will fall into a small interval around a specific value is positive. For each x-value given in Figure **??** $f_X(x)$ could be interpreted as the probability that X will fall within a small interval around x divided by the length of this interval. Especially we have:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du \tag{A.10}$$

and

$$\Pr(a < X \leq b) = \int_{a}^{b} f_X(x)\,dx \tag{A.11}$$

The last expression is illustrated in Figure **??**.

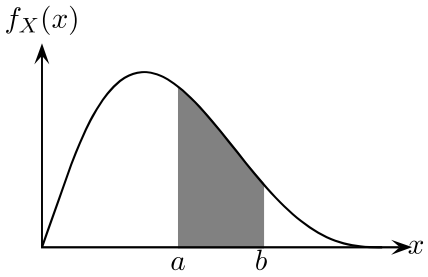


Figure A.6: The shadded area equals $\Pr(a < X \leq b)$

Random quantities that take discrete values are said to be discretely distributed. For such quantities we introduce the point probability for X in the point x_j:

$$p(x_j) = \Pr(X = x_j) \tag{A.12}$$

where $x_1, x_2, \ldots$ are possible values X could take.

Expectation. The expectation (mean) of X is given by

$$\mathrm{E}[X] = \begin{cases} \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx & \text{if } X \text{ is continuous} \\ \sum_j x_j \cdot p(x_j) & \text{if } X \text{ is discrete} \end{cases} \tag{A.13}$$

The expectation can be interpreted as the long time run average of X, if an infinite amount of observations are available.

Median. The median of a distribution is the value m_0 of the stochastic variable X such that $\Pr(X \leq m_0) \geq 1/2$ and $\Pr(X \geq m_0) \geq 1/2$. In other words, the probability at or below m_0 is at least $1/2$, and the probability at or above m_0 is at least $1/2$.

Mode. The mode of a distribution is the value M of the stochastic variable X such that the probability density function, or point probability at M is higher or equal than for any other value of the stochastic variable. We sometimes used the term 'most likely value' rather than *mode*.

Variance. The variance of a random quantity expresses the variation in the value X will take in the long run. We denote the variance of X by:

$$\mathrm{Var}(X) = \begin{cases} \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 \cdot f_X(x)\, dx & \text{if } X \text{ is continuous} \\ \sum_j \big((x_j - \mathrm{E}[X]\big)^2 \cdot p(x_j) & \text{if } X \text{ is discrete} \end{cases} \tag{A.14}$$

Standard deviation. The standard deviation of X is given by

$$\mathrm{SD}(X) = +\sqrt{\mathrm{Var}(X)} \tag{A.15}$$

The standard deviation defines an interval which observations are likely to fall into, i.e., if 100 observations are available, we expect that approximate[1] 67 of these observations fall in the interval $[\mathrm{E}[X] - \mathrm{SD}(X), \mathrm{E}[X] + \mathrm{SD}(X)]$.

Precision. The precision, P, is the reciprocate of the variance, i.e. $P = \frac{1}{\mathrm{Var}(X)}$.

α-percentiles. The upper α-percentile, x_α, in a distribution $F_X(x)$ is the value satisfying $\alpha = \Pr(X > x_\alpha) = 1 - F_X(x_\alpha)$.

[1] This result is valid for the normal distribution. For other distributions there may be deviation from this result.

We end this section by giving some results regarding expectation and variances. These results apply when it is easier to express the expectation and variance of one variable if we condition on the value of another variable.

Result

Double expectation
Let X and Y be stochastic variables. We then have:

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]] \tag{A.16}$$

$$\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathrm{E}[X|Y]) \tag{A.17}$$

□

It follows easily that

$$\mathrm{E}[X] = \mathrm{E}[X|B]\Pr(B) + \mathrm{E}\left[X|B^C\right]\Pr(B^C) \tag{A.18}$$

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(X|B)\Pr(B) + \mathrm{Var}(X|B^C)\Pr(B^C) \\ &+ (\mathrm{E}[X|B] - \mathrm{E}[X])^2\Pr(B) + \left(\mathrm{E}\left[X|B^C\right] - \mathrm{E}[X]\right)^2\Pr(B^C)\end{aligned} \tag{A.19}$$

A.2 Common probability distributions

In this section we will present some common probability distributions. We write $X \sim$ <Name of distribution>(<parameters>) to express that X belongs to <Name of distribution>, and with parameters <parameters>. Sometimes we also use an abbreviation for the distribution, for example we write $X \sim N(3,4)$ to express that X is normally distributed with expectation 3 and variance 4.

A.2.1 The normal distribution

X is said to be normally distributed if the probability density function of X is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{A.20}$$

where μ and σ are parameters that characterise the distribution. The mean and variance are given by:

$$\begin{aligned} \mathrm{E}[X] &= \mu \\ \mathrm{Var}(X) &= \sigma^2 \end{aligned} \tag{A.21}$$

The distribution function for X could not be written on closed from. Numerical methods are required to find $F_X(x)$. It is convenient to introduce a standardised normal distribution for this purpose. We say that U is standard normally distributed if it's probability density function is given by:

$$f_U(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{A.22}$$

We then have

$$F_U(u) = \Phi(u) = \int_{-\infty}^{u} \phi(t)\mathrm{d}t = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t \tag{A.23}$$

and we observe that the distribution function of U does not contain any parameters. We therefore only need one look-up table or function representing $\Phi(u)$. A look-up table is given in Table **??**. To calculate probabilities in the non-standardised normal distribution we use the following result:

Result

If X is normally distributed with parameters μ and σ, then

$$U = \frac{X - \mu}{\sigma} \tag{A.24}$$

is standard normally distributed. □

In many situations we are interested in calculating the "truncated expectation" $\int_{-\infty}^{a} x f(x) dx$. For the normal distribution the following result may be used:

Result

Let X be normally distributed with parameters μ and σ. We then have:

Result

$$\int_{-\infty}^{a} xf(x)dx = \mu\Phi\left(\frac{a-\mu}{\sigma}\right) - \sigma\phi\left(\frac{a-\mu}{\sigma}\right) \tag{A.25}$$

where $\Phi()$ and $\phi()$ are the CDF and PDF for the standard normal distribution respectively. □

To prove Equation (**??**) first introduce $u = (x-\mu)/\sigma$ yielding $\int_{-\infty}^{a} xf(x)dx = \int_{-\infty}^{(a-\mu)/\sigma}(\sigma u - \mu)\phi(u)du$. The $\mu\phi(u)$ part of the integral is directly found by the $\Phi()$ function whereas for the $\sigma u\phi(u)$ part introduce $z = -u^2/2$ yielding $-\sigma/\sqrt{2\pi}\int_{-\infty}^{(a-\mu)^2/2\sigma^2} e^{-z}dz$. The result then follows.

Example **Calculation in the normal distribution**
Let X be normally distributed with parameters $\mu = 5$ and $\sigma = 3$. We will find $\Pr(3 < X \leq 6)$. We have:

$$\Pr(3 < X \leq 6) = \Pr(\frac{3-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{6-\mu}{\sigma}) = \Pr(\frac{3-5}{3} < U \leq \frac{6-5}{3})$$
$$= \Phi\left(\frac{1}{3}\right) - \Phi\left(\frac{-2}{3}\right) = \Phi(0.33) - (1-\Phi(0.67)) = 0.629 - 1 + 0.749 = 0.378$$

□

Problem

Consider the example in Example **??**, and carry out the calculations. □

Problem

Let X be the height of men in a population, and assume X is normally distributed with parameters $\mu = 181$ and $\sigma = 4$. How large percentage of the population is more than 190 cm? □

A.2.2 The exponential distribution

X is said to be exponentially distributed if the probability density function of X is given by:

$$f_X(x) = \lambda e^{-\lambda x} \tag{A.26}$$

The cumulative distribution function is given by:

$$F_X(x) = 1 - e^{-\lambda x} \tag{A.27}$$

and the mean and variance are given by:

$$\begin{aligned} \mathrm{E}[X] &= 1/\lambda \\ \mathrm{Var}(X) &= 1/\lambda^2 \end{aligned} \tag{A.28}$$

Note that for the exponential distribution, X will always be greater than 0. The parameter λ is often denoted the intensity in the distribution Example
We will obtain the probability that X is greater than it's expected value. We then have:

$$\Pr(X > \mathrm{E}[X]) = 1 - \Pr(X \leq \mathrm{E}[X]) = 1 - F_X(\mathrm{E}[X]) = e^{-\lambda \mathrm{E}[X]} = e^{-1} \approx 0.37$$

□

A.2.3 The Weibull distribution

X is said to be Weibull distributed if the probability density function of X is given by:

$$f_X(x) = \alpha\lambda(\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha} \tag{A.29}$$

The cumulative distribution function is given by:

$$F_X(x) = 1 - e^{-(\lambda x)^\alpha} \tag{A.30}$$

and the mean and variance are given by:

$$\begin{aligned} \mathrm{E}[X] &= \frac{1}{\lambda}\Gamma\left(\frac{1}{\alpha}+1\right) \\ \mathrm{Var}(X) &= \frac{1}{\lambda^2}\left[\Gamma\left(\frac{2}{\alpha}+1\right) - \Gamma^2\left(\frac{1}{\alpha}+1\right)\right] \end{aligned} \tag{A.31}$$

where $\Gamma(\cdot)$ is the gamma function. Note that in the Weibull distribution X will always be positive.

The Weibull distribution is often used as a distribution for time to failure of components, partly because it is rather simple, and flexible. For time to failure distributions we introduce the concept of the failure rate function. The failure rate function is given by:

$$z(t) = \frac{f(t)}{1 - F(t)} \tag{A.32}$$

here we use t for running time, and we skip the index for the name of the stochastic variable when it is obvious from the context. It follows that the failure rate function for the Weibull dis-

tribution is given by:

$$z(t) = \alpha\lambda^{\alpha} t^{\alpha-1} \tag{A.33}$$

A.2.4 The gamma distribution

X is said to be gamma distributed if the probability density function of X is given by:

$$f_X(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}(x)^{\alpha-1} e^{-\lambda x} \tag{A.34}$$

α is denoted the intensity parameter whereas λ is denoted the intensity parameter. For integer values of α the gamma distribution is often denoted the *Erlang* distribution. The cumulative distribution function could then be found on closed form:

$$F_X(x) = 1 - \sum_{n=0}^{\alpha-1} \frac{(\lambda x)^{\alpha}}{n!} e^{-(\lambda x)} \tag{A.35}$$

For non-integer values of α numerical methods are required to obtain the cumulative distribution function. The mean and variance are given by:

$$\begin{aligned} \mathrm{E}[X] &= \frac{\alpha}{\lambda} \\ \mathrm{Var}(X) &= \frac{\alpha}{\lambda^2} \end{aligned} \tag{A.36}$$

If we know the expectation E and the variance V in the gamma distribution, we may obtain the parameters α and λ by: $\lambda = E/V$, and $\alpha = \lambda \cdot E$. The gamma distribution is often used as a prior distribution in a Bayesian approach.

For integer values of α the gamma distribution and in particular the Erlang distribution may be seen as a distribution for a sum of exponentially distributed stochastic variables:

Result

Let $Z_1, Z_2, \ldots Z_k$ be independent and exponentially distributed with parameter λ. The variable $X = \sum_{i=1}^{k} Z_i$ is then gamma distributed with shape parameter k and scale parameter λ. □

A.2.5 The inverted gamma distribution

X is said to be inverted gamma distributed if the probability density function of X is given by:

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} e^{-\lambda/x} \tag{A.37}$$

The mean and variance are given by:

$$\begin{aligned} \mathrm{E}[X] &= \lambda/(\alpha-1) \\ \mathrm{Var}(X) &= \lambda^2(\alpha-1)^{-2}(\alpha-2)^{-1} \end{aligned} \tag{A.38}$$

Note that if X is gamma distributed with parameters α and λ, then $Y = X^{-1}$ has an inverted gamma distribution with parameters α and $1/\lambda$. If we know the expectation, E and the variance, V, of an inverted gamma distribution we could obtain α and λ by $\alpha = E^2/V + 2$, and $\lambda = E \cdot (\alpha - 1)$.

A.2.6 The lognormal distribution

X is said to be lognormally distributed if the probability density function of X is given by: eqStream: Lognormal Distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\tau} \frac{1}{x} e^{-\frac{1}{2\tau^2}(\log x - \nu)^2} \tag{A.39}$$

We write $X \sim \mathrm{LN}(\nu,\tau)$. The mean and variance of X is given by

$$\begin{aligned} \mathrm{E}[X] &= e^{\nu + \frac{1}{2}\tau^2} \\ \mathrm{Var}(X) &= e^{2\nu}(e^{2\tau^2} - e^{\tau^2}) \end{aligned} \tag{A.40}$$

The following result could be utilised:

Result

If X is lognormally distributed with parameters ν and τ, then $Y = \ln X$ is normally distributed[2] with expected value ν and variance τ^2. □

[2] $\ln(\cdot)$ is the natural logarithm function

A.2.7 The binomial distribution

Before the binomial distribution is defined, binomial trials are defined. Let A be an event, and assume that the following holds:

i) n trials are performed, and in each trial we record whether A occurrs or not.

ii) The trials are stochastic *independent* of each other.

iii) For each trial $\Pr(A) = p$

When i)-iii) is satisfied, we say that we have binomial trials. Now let X be the number of times event A occurs in such a binomial trial. X is then a stochastic variable with a binomial distribution. This is written $X \sim Bin(n, p)$.

The probability function is given by

$$\Pr(X = x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, .., n \tag{A.41}$$

The cumulative distribution function $\Pr(X \leq x)$ is given in statistical tables. For the binomial distribution, expectation and variance are given by:

$$\begin{aligned} \mathrm{E}[X] &= np \\ \mathrm{Var}(X) &= np(1-p) \end{aligned} \tag{A.42}$$

A.2.8 The Poisson distribution

The Poisson distribution is often appropriate in the situation where the stochastic variable may take the values 0,1,2,..., and where the expected number of occurrences is proportional to an exposure measure such as time or space. For the Poisson distribution we have the following point distribution:

$$p(x) = \Pr(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \tag{A.43}$$

For the poison distribution, expectation and variance are given by:

$$\begin{aligned} \mathrm{E}[X] &= \lambda \\ \mathrm{Var}(X) &= \lambda \end{aligned} \tag{A.44}$$

It can be proved that the Poisson distribution is appropriate if the following situation applies: Consider the occurrence of a certain event (e.g., a component failure) in an interval (a, b), and assume the following:

1. A could occur anywhere in (a,b), and the probability that A occurs in $(t, t+\Delta t)$ is approximately equal to $\lambda\Delta t$, and is independent of t (Δt should be small).

2. The probability that A occurs several times in $(t, t+\Delta t)$ is approximately 0 for small values of Δt.

3. Let I_1og I_2 be disjoint intervals in (a, b). The event A occurs within I_1 is independent of if the event A occurs in I_2.

When the criteria above are fulfilled we say we have a *Poisson point process* with intensity λ. The number of occurrences (X) of A in (a, b) is then Poisson distributed with parameter $\lambda(b-a)$:

$$p(x) = \Pr(X = x) = \frac{[\lambda(b-a)]^x}{x!} e^{-\lambda(b-a)} \tag{A.45}$$

Result: Times between occurrence in the Poisson process

In a Poisson point process with parameter λ the times between the occurrence of the event A are exponentially distributed with parameter λ. □

A.2.9 The inverse-Gauss distribution

The inverse-Gauss distribution is often used when we have an "under laying" deterioration process. If this deterioration process follows a Wiener process with drift η and diffusion constant δ^2, the time[3] T, until the first time the process reaches the value ω will be Inverse-Gauss distributed with parameters $\mu = \omega/\eta$, and $\lambda = \omega^2/\delta^2$.

If the failure progression $\Omega(t)$ follows a Wiener process it could be proven that $\Omega(t)$ - $\Omega(s)$ is normally distributed with expected value $\eta(t-s)$ and variance $\delta^2(t$ - s). That is η is the average growth rate in the process, whereas δ^2 is an expression for the variation of the growth around the average value.

For the inverse-Gauss distribution we have:

$$F_T(t) = \Phi\left(\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}-1\right)\right)) + \Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}+1\right)\right) e^{2\lambda/\mu} \tag{A.46}$$

and

$$\mathrm{E}[T] = \mu \tag{A.47}$$

$$\mathrm{Var}(T) = \mu^3/\lambda \tag{A.48}$$

[3] We use the symbol T rather than the more general symbol X here since this modell is so explicitly linked to the time.

A.3 Distribution of sums, products and maximum values

A.3.1 Distribution of sums

If X_1, X_2,...,X_n are random variables we might obtain the expected value, the variance and the standard deviation of the sum of the x-es:

$$\mathrm{E}[X_1 + X_2 + \ldots + X_n] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i] \tag{A.49}$$

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \tag{A.50}$$

$$\mathrm{SD}\left(\sum_{i=1}^{n} X_i\right) = \sqrt{\sum_{i=1}^{n} [\mathrm{SD}(X_i)]^2} \tag{A.51}$$

Note that Equations (**??**) and (**??**) are only valid if the x-es are stochastically independent. If there is dependency between the x-es we need to include a covariance term, e.g., if we only have two variables X_1 and X_2 we have:

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2) \tag{A.52}$$

where $\mathrm{Cov}(X_1, X_2)$ is the covariance between X_1 and X_2.

The results above help us in determine the expectation and variance of a sum of stochastic variables, but the results could not be used to establish the probability distribution of the sum. In the following we refer some results we could utilise in many situations.

Result: Sum of normally distributed stochastic variables

Let X_1, X_2,...,X_n be independent normally distributed. Let Y be the sum of the x-es, i.e. $Y = \sum_{i=1}^{n} X_i$. Y is then normally distributed with $\mathrm{E}[Y] = \sum_{i=1}^{n} \mathrm{E}[X_i]$ and $\mathrm{Var}(Y) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$. □

Result: Sum of exponentially distributed stochastic variables

Let X_1, X_2,...,X_n independent exponentially distributed with parameter λ. Let Y be the sum of the x-es, i.e. $Y = \sum_{i=1}^{n} X_i$. Y is then gamma distributed with parameters n and λ. □

Result: Sum of gamma distributed stochastic variables

Let X_1, X_2,...,X_n independent gamma distributed with parameters α and λ. Let Y be the sum of the x-es, i.e. $Y = \sum_{i=1}^{n} X_i$. Y is then gamma distributed with parameters $n\alpha$ and λ. □

Result: Central limit theorem

Let X_1, X_2,...,X_n be a sequence of identical independent distributed stochastic variables with expected value μ and standard deviation σ. As n approaches infinity, the average value of the x-es will asymptotically have a normal distribution with expected value μ and standard deviation $\sigma/\sqrt{n}$. Similarly, the sum of the x-es will asymptotically have a normal distribution with expected value $n\mu$ and standard deviation $\sigma\sqrt{n}$. □

Several generalizations for finite variance exist which do not require identical distribution but incorporate some conditions which guarantee that none of the variables exert a much larger influence than the others. Two such conditions are the Lindeberg condition and the Lyapunov condition. Now, as n approaches infinity, the sum of the x-es will asymptotically have a normal distribution with expected value $\sum_{i=1}^{n} \mathrm{E}[X_i]$ and variance $\sum_{i=1}^{n} \mathrm{Var}(X_i)$.

A.3.2 Distribution of a product

If X_1, X_2,...,X_n are *independent* stochastic variables we might obtain the expected value, the variance and the standard deviation of the product of the x-es:

$$\mathrm{E}[X_1 \cdot X_2 \cdot \ldots \cdot X_n] = \mathrm{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathrm{E}[X_i] \tag{A.53}$$

The results for the variance and standard deviation is more complicated, and we only present the results for n=2.

$$\mathrm{Var}(X_1 X_2) = \mathrm{Var}(X_1)\mathrm{Var}(X_2) + \mathrm{Var}(X_1)\left(\mathrm{E}[X_2]\right)^2 + \mathrm{Var}(X_2)\left(\mathrm{E}[X_1]\right)^2 \tag{A.54}$$

$$\mathrm{SD}(X_1 X_2) = \sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2) + \mathrm{Var}(X_1)\left(\mathrm{E}[X_2]\right)^2 + \mathrm{Var}(X_2)\left(\mathrm{E}[X_1]\right)^2} \tag{A.55}$$

Table A.1: The Cumulative Standard Normal Distribution

$$\Phi(z) = \Pr(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \, du$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.500	.504	.508	.512	.516	.520	.524	.528	.532	.536
0.1	.540	.544	.548	.552	.556	.560	.564	.567	.571	.575
0.2	.579	.583	.587	.591	.595	.599	.603	.606	.610	.614
0.3	.618	.622	.626	.629	.633	.637	.641	.644	.648	.652
0.4	.655	.659	.663	.666	.670	.674	.677	.681	.684	.688
0.5	.691	.695	.698	.702	.705	.709	.712	.716	.719	.722
0.6	.726	.729	.732	.732	.739	.742	.745	.749	.752	.755
0.7	.758	.761	.764	.767	.770	.773	.776	.779	.782	.785
0.8	.788	.791	.794	.797	.800	.802	.805	.808	.811	.813
0.9	.816	.819	.821	.824	.826	.829	.831	.834	.836	.839
1.0	.841	.844	.846	.849	.851	.853	.855	.858	.860	.862
1.1	.864	.867	.869	.871	.873	.875	.877	.879	.881	.883
1.2	.885	.887	.889	.891	.893	.894	.896	.898	.900	.901
1.3	.903	.905	.907	.908	.910	.911	.913	.915	.916	.918
1.4	.919	.921	.922	.924	.925	.926	.928	.929	.931	.932
1.5	.933	.934	.936	.937	.938	.939	.941	.942	.943	.944
1.6	.945	.946	.947	.948	.949	.951	.952	.953	.954	.954
1.7	.955	.956	.957	.958	.959	.960	.961	.962	.962	.963
1.8	.964	.965	.966	.966	.967	.968	.969	.969	.970	.971
1.9	.971	.972	.973	.973	.974	.974	.975	.976	.976	.977
2.0	.977	.978	.978	.979	.979	.980	.980	.981	.981	.982
2.1	.982	.983	.983	.983	.984	.984	.985	.985	.985	.986
2.2	.986	.986	.987	.987	.987	.988	.988	.988	.989	.989
2.3	.989	.990	.990	.990	.990	.991	.991	.991	.991	.992
2.4	.992	.992	.992	.992	.993	.993	.993	.993	.993	.994
2.5	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995
2.6	.995	.995	.996	.996	.996	.996	.996	.996	.996	.996
2.7	.997	.997	.997	.997	.997	.997	.997	.997	.997	.997
2.8	.997	.998	.998	.998	.998	.998	.998	.998	.998	.998
2.9	.998	.998	.998	.998	.998	.998	.999	.999	.999	.999
3.0	.999	.999	.999	.999	.999	.999	.999	.999	.999	.999

$\Phi(-z) = 1 - \Phi(z)$