Student Textbook

# Reliability
# Fault Tolerance

00

00

0004

Dan Ștefan Tudose

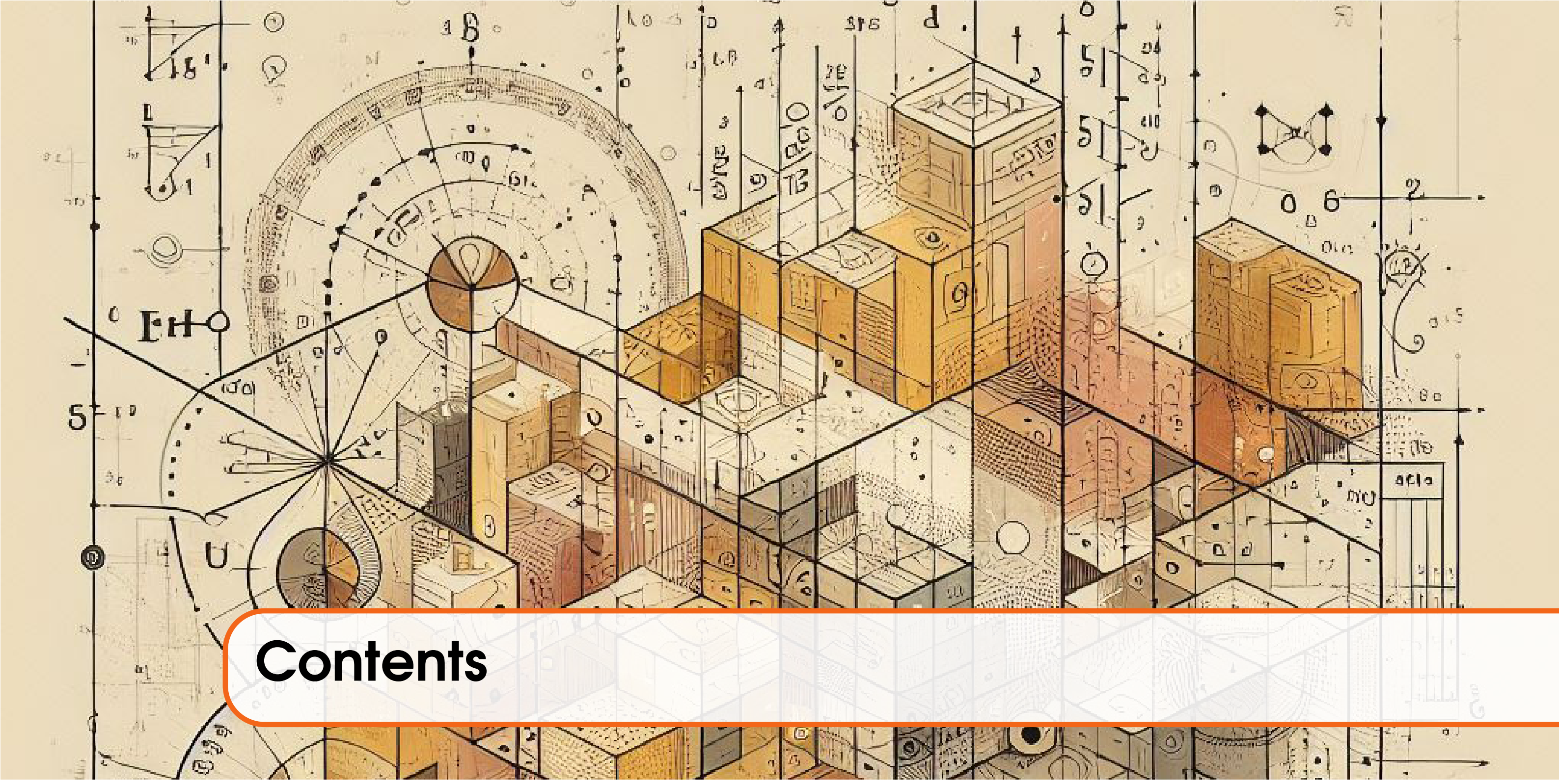
AS OF YET, UNPUBLISHED. PROBABLY NEVER WILL.



*First printing, March 2024*

"Anything that can go wrong will go wrong."

*Murphy's Law*



# Contents


**I — Part One**

**1 Probability Distributions** . . . . . . . . 13

**1.1 Probability Theory Basics** 13

**1.2 Common Probability Distributions** 15

1.2.1 Cumulative Distribution Functions . . . . . . . . 16

1.2.2 Probability Density Functions . . . . . . . . 17

1.2.3 The Expected Value of a Random Variable . . . . . . . . 18

1.2.4 Probability Distributions Commonly Used in Reliability . . . . . . . . 19

**2 Modeling Reliability** . . . . . . . . 25

**2.1 Reliability and Availability** 25

2.1.1 Reliability . . . . . . . . 25

2.1.2 Failure Rate . . . . . . . . 27

2.1.3 Mean Time Between Failures . . . . . . . . 29

2.1.4 Availability . . . . . . . . 30

**2.2 Failure Rate, Reliability, and Mean Time to Failure for an Exponential Fault Distribution** 32

2.2.1 Non-constant Failure Rate . . . . . . . . 34

**II — Part Two**

**3 Reliability Block Diagrams** . . . . . . . . 39

**3.1 Modeling Reliability Through Blocks** 39

3.2 Series Structures 41
3.3 Parallel Structures 43
3.4 Combination of Series and Parallel 47
3.5 k Out of n Systems 47
3.6 Series-Parallel and Parallel-Series Systems 49
3.7 Non-Decomposable Systems 51
3.8 Majority Voted Redundancy 53
3.8.1 Triple Modular Redundancy (TMR) 53
3.8.2 3-out-of-5 Modular Redundancy 55
3.8.3 n-out-of-(2n-1) Modular Redundancy 56
3.9 Standby-Sparing 57
3.9.1 One Spare Reliability 57
3.9.2 Two Spare Reliability 60
3.9.3 N Spare Reliability 61

III Part Three

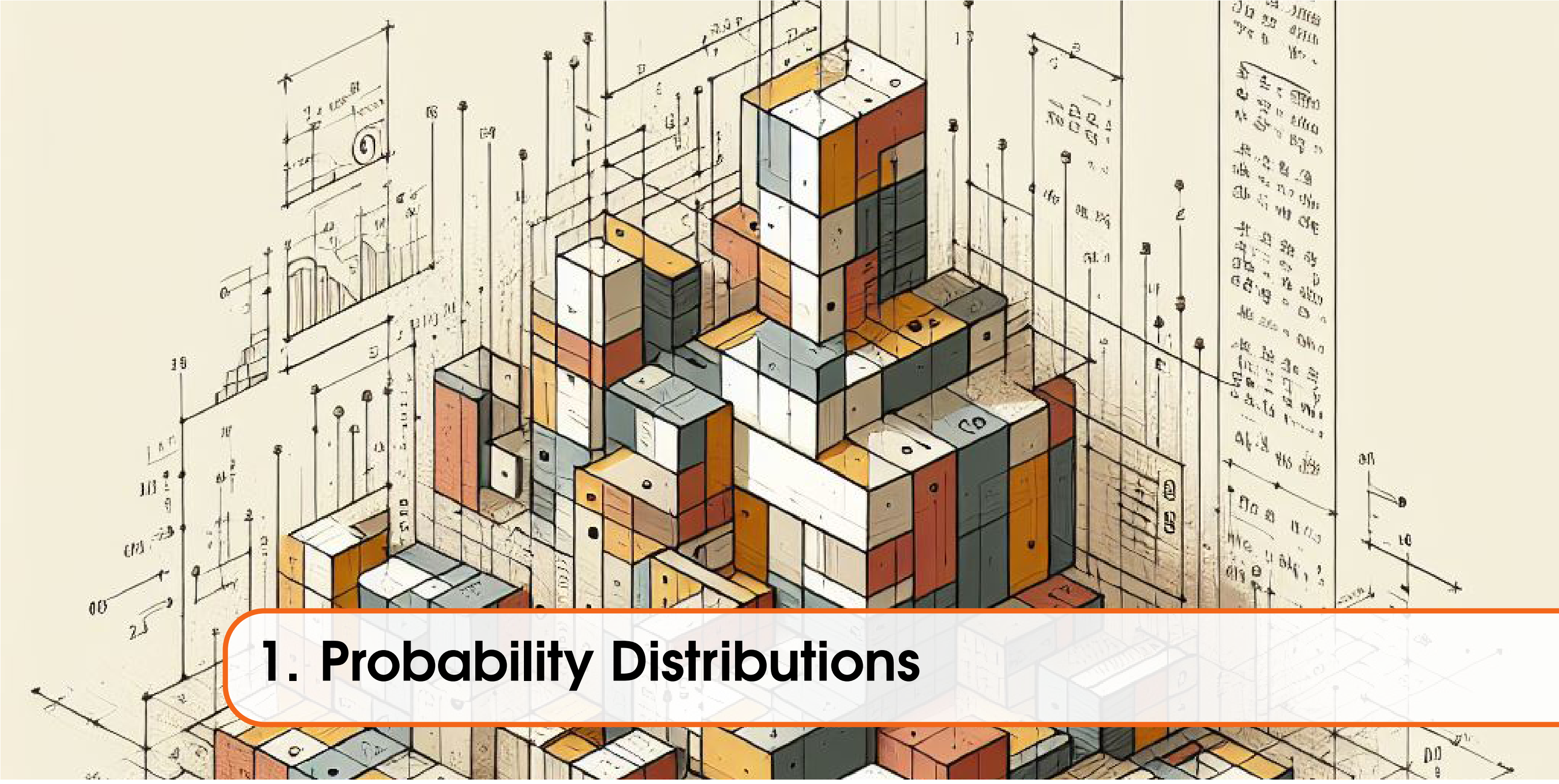
4 Markov Models 65
4.1 Markov Chains 66
4.1.1 Definition and Basic Concepts 66
4.2 Markov Processes 72
4.2.1 Definition and Basic Concepts 72
4.2.2 The Chapman-Kolmogorov Differential Equation System 73
4.2.3 Matricial Form of Chapman-Kolmogorov Equations for a Markov Process 75
4.2.4 Applications of Markov Processes 75
4.3 Single-Component System without Repair 76
4.3.1 State Diagram 77
4.3.2 Transition Rate Matrix 77
4.3.3 System Reliability 77
4.3.4 Mean Time To Failure (MTTF) 78
4.4 Single-Component with Repair 78
4.4.1 State Space 78
4.4.2 Transition Rates 79
4.4.3 Transition Rate Matrix 79
4.4.4 Graph of the Markov Process 79
4.4.5 Solving for State Probabilities (Reliability) 79
4.4.6 System Reliability 80
4.4.7 Mean Time To Failure (MTTF) 80
4.4.8 System Availability for a Single-Component System with Repair 80
4.5 Two-Component System Without Repair 81
4.5.1 State Space 81
4.5.2 Transition Rates 81
4.5.3 Transition Rate Matrix 81
4.5.4 Graph of the Markov Process 82
4.5.5 Solving for State Probabilities (Reliability) 82

**4.6 Two-Component System with Repair** **83**
4.6.1 State Space . . . 83
4.6.2 Transition Rate Matrix . . . 83
4.6.3 Simplified Case: Identical Components . . . 84
**4.7 Two-Component System with a Dependency** **87**
4.7.1 Reliability . . . 88
4.7.2 Mean Time To Failure (MTTF) . . . 89
**4.8 System with one Cold Standby Backup** **89**
4.8.1 State Definitions . . . 89
4.8.2 Transition Rates . . . 90
4.8.3 Kolmogorov Forward Equations . . . 90
4.8.4 Reliability Function of the Cold Standby System . . . 91
**4.9 System with a Warm Standby Backup** **92**
4.9.1 State Definitions . . . 92
4.9.2 Transition Rates . . . 92
4.9.3 State Transition Diagram . . . 92
4.9.4 Kolmogorov Forward Equations . . . 92
4.9.5 Probabilities as a Function of Time . . . 92
4.9.6 Reliability Function of the Warm Standby System . . . 94
**4.10 Cold Standby Backup System with Repair** **94**
4.10.1 State Definitions . . . 94
4.10.2 Kolmogorov Equation System . . . 95
4.10.3 Steady-State Probabilities . . . 95
4.10.4 Steady-State Availability . . . 96
**4.11 Cold Standby System with Two Backups** **96**
4.11.1 State Definitions . . . 97
4.11.2 Kolmogorov Forward Equations . . . 97
4.11.3 Probabilities as a Function of Time . . . 97
4.11.4 Reliability Function . . . 98
**4.12 2-out-of-3 Majority Voting System** **98**
4.12.1 State Definitions . . . 98
4.12.2 Kolmogorov Forward Equations . . . 99
4.12.3 Probabilities as a Function of Time . . . 99
4.12.4 Reliability Function . . . 100
**4.13 2-out-of-3 Majority Voting System with Repair** **100**
4.13.1 State Definitions . . . 100
4.13.2 Kolmogorov Forward Equations . . . 100
4.13.3 State Probabilities . . . 101
4.13.4 Reliability Function . . . 101
4.13.5 Steady-State Reliability of the 2-out-of-3 Majority Voting System with Repair . 101
**4.14 Three-State System with Degraded Mode** **102**
4.14.1 Kolmogorov Equation System . . . 102
4.14.2 Probabilities for Each State . . . 102
4.14.3 Reliability . . . 103
4.14.4 Mean Time To Failure (MTTF) . . . 103
**4.15 Markov Model for a Component with Three States** **103**
4.15.1 Kolmogorov Equation System . . . 103
4.15.2 Probabilities for Each State . . . 104

4.15.3 Reliability . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 104
4.15.4 Mean Time To Failure (MTTF) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105

# Part One


| | | |
|---|---|---|
| **1** | **Probability Distributions** | **13** |
| 1.1 | Probability Theory Basics | |
| 1.2 | Common Probability Distributions | |
| **2** | **Modeling Reliability** | **25** |
| 2.1 | Reliability and Availability | |
| 2.2 | Failure Rate, Reliability, and Mean Time to Failure for an Exponential Fault Distribution | |



# 1. Probability Distributions

## 1.1 Probability Theory Basics

Probability theory is the study of how likely it is that random events will occur. It is a branch of mathematics that is used in many different fields, including statistics, physics, economics, and engineering.

Random events are events that cannot be predicted with certainty, but that can still be described according to their likelihood. For example, the outcome of a coin toss is a random event, because it is impossible to predict whether the coin will land heads or tails before it is tossed. However, we can still describe the likelihood of each outcome based on the fact that there are two equally likely possibilities.

Probability theory provides a mathematical framework for describing the likelihood of random events. It does this by using probability measures, which assign a value between 0 and 1 to each event. A probability of 0 means that the event is impossible, a probability of 1 means that the event is certain, and a probability between 0 and 1 means that the event is possible, but not certain.

There are two main types of probability measures: discrete probability measures and continuous probability measures. Discrete probability measures are used to describe the likelihood of discrete events, such as the outcome of a coin toss or the number of heads that appear when a fair coin is tossed 10 times. Continuous probability measures are used to describe the likelihood of continuous events, such as the height of a randomly selected person or the temperature on a random day in June.

The first axiom of probability theory states that the value of probability of an event $A$ lies between 0 (impossibility) and 1 (certainty):

$$0 \leq P(A) \leq 1 \tag{1.1}$$

$\overline{A}$ denotes the event "*not A*". For example, if $A$ stands for "it rains", $\overline{A}$ stands for "it does not rain". The second axiom of probability theory says that the probability of an event $A$ is equal to 1 minus the probability of the event $A$:

$$P(\overline{A}) = 1 - P(A) \tag{1.2}$$

Suppose that one event, $A$ is dependent on another event, $B$. Then $P(A \mid B)$ denotes the conditional probability of event $A$, given event $B$. The fourth rule of probability theory states that the probability $P(A \cdot B)$ that both $A$ and $B$ will occur is equal to the probability that $B$ occurs times the conditional probability $P(A \mid B)$:

$$P(A \cdot B) = P(A \mid B) \cdot P(B) \tag{1.3}$$

If P(B) is greater than zero, then equation 1.3 can be written as

$$P(A \mid B) = \frac{P(A \cdot B)}{P(B)} \tag{1.4}$$

An important condition that we will often assume is that two events are mutually independent. For events $A$ and $B$ to be independent, the probability *P(A)* does not depend on whether B has already occurred or not, and vice versa.

Thus, $P(A \mid B) = P(A)$. So, for independent events, the rule 1.4 reduces to

$$P(A \cdot B) = P(A) \cdot P(B) \tag{1.5}$$

This is the definition of independence, that the probability of two events both occurring is the product of the probabilities of each event occurring. Situations also arise when the events are mutually exclusive. That is, if $A$ occurs, $B$ cannot, and vice versa. As such, we can write the following

$$P(A \cdot B) = P(B \cdot A) = 0 \tag{1.6}$$

This is the definition of mutually exclusiveness, that the probability of two events both occurring is zero.

Let us now consider the situation when either $A$, or $B$, or both event may occur. The probability $P(A+B)$ is given by

$$P(A+B) = P(A) + P(B) - P(A \cdot B) \tag{1.7}$$

Combining 1.6 with 1.7, we get the following expression for mutually exclusive events

$$P(A+B) = P(A) + P(B) \tag{1.8}$$

## 1.2 Common Probability Distributions

Probability theory operates on a series of fundamental notions:

- Random experiment: any procedure that can be repeated indefinitely and has a well-defined set of possible outcomes.
- Sample space: the set of all possible outcomes of an experiment.
- Event: a subset of the sample space.
- Probability of an event: the number between 0 and 1 assigned to an event by a probability measure.

A simple example of a random experiment, and one that is frequently mentioned in probability theory textbooks is the toss of a coin. It has a defined set of possible outcomes: *heads*, *tails* which constitutes the sample space of the experiment. Each of the two events in the sample space, *heads* or *tails* has an associated probability of occurrence.

Another example of a random experiment is googling something or someone and measuring how fast the search was performed. The sample space is now made from all possible response times the search engine will report $\{t|t > 0\}$ and is no longer discrete, as in the coin toss experiment, but continuous.

A *random variable* is a variable that assigns a numerical value to each outcome in a sample space. In other words, it is a function that takes an outcome as an input and returns a number as an output. Random variables are used to quantify the uncertainty associated with random experiments.

There are two main types of random variables: discrete and continuous.

- *Discrete random variables* take on a finite or countably infinite number of values. A discrete random variable for the coin toss example could map *heads* to 1 and *tails* to 0, for example.
- *Continuous random variables* can take on any value within a certain interval. For the web search example above, the sample space is already composed of numbers, as the possible response times of the search query. This is a case in which the random variable could map the sample space to itself.

Contrary to their definition as functions, random variables are usually denoted with capital letters such as $X, Y, T$ without including their parameter.

Given a random variable $X$ that we use to map search engine response times, we can ask the question: "How likely is that the value of this random variable for a web search is equal to a tenth of a second?". We can write this as a probability $P(X = 0.1)$.

If we record all probabilities of all the outputs of a random variable $X$, we get the *probability distribution of X*.

Then we can ask another question: "What is the probability that a web search will yield a result in less than a tenth of a second?". To answer this, we will need to use the *cumulative distribution function* (CDF).

### 1.2.1 Cumulative Distribution Functions

The CDF of a random variable $X$ is a function that gives the probability that X is less than or equal to a certain value x. We usually note CDF by $F_X(x)$, or, if the random variable is implicit or there's no ambiguity, just by $F(x)$:

$$F_X(x) = P(X \leq x) \tag{1.9}$$

CDF is a non-decreasing function, meaning that as x increases, the CDF(x) also increases.

The CDF is a useful tool for understanding the distribution of a random variable. It can be used to calculate probabilities, such as the probability that a random variable will be between two values.

■ **Example 1.1** For a coin toss experiment, we can define the CDF as:

$$F(x) = \begin{cases} 0 & ,x < 0 \\ q & ,0 \leq x < 1 \\ 1 & ,x \geq 1 \end{cases} \tag{1.10}$$

If the coin is fair, then $q = 0.5$. ■

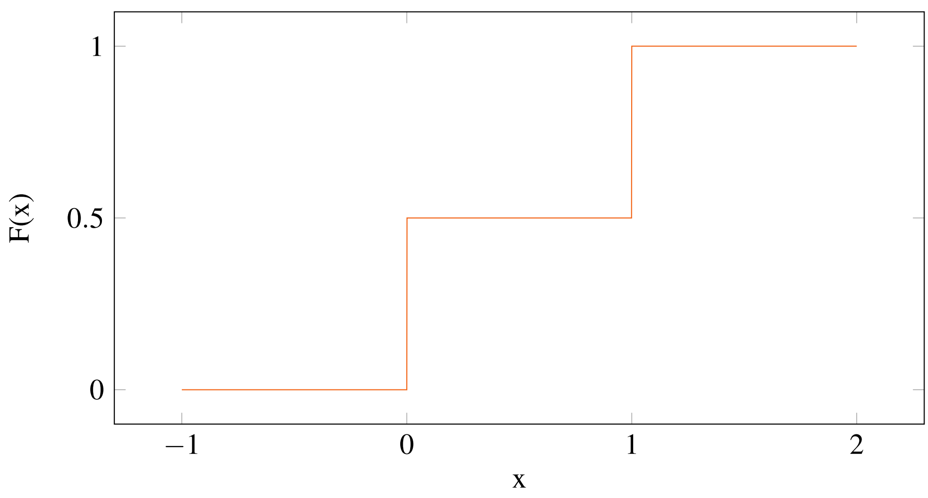



Figure 1.1: CDF of the random coin toss experiment

In computing, we are most often interested in time as a continuous random variable: time of completion of a certain query, system uptime, reliable operation time etc. Therefore, we will be frequently using CDF as $F(t)$ for positive values of time: $0 \leq t < \infty$.

**Definition 1.2.1 — CDF Properties.** The CDF for a continuous random variable that has only positive values has the following properties:

$$0 \leq F(t) \leq 1 \forall t \geq 0 \quad (1.11)$$
$$F(0) = 0 \quad (1.12)$$
$$\lim_{t \to \infty} F(t) = 1 \quad (1.13)$$
$$F(t) \textit{ is a monotone increasing function of time} \quad (1.14)$$

■ **Example 1.2** $F(t) = 1 + 2e^{-3t} - 3e^{-2t}$ is a valid CDF and is plotted in Figure 1.2. ■

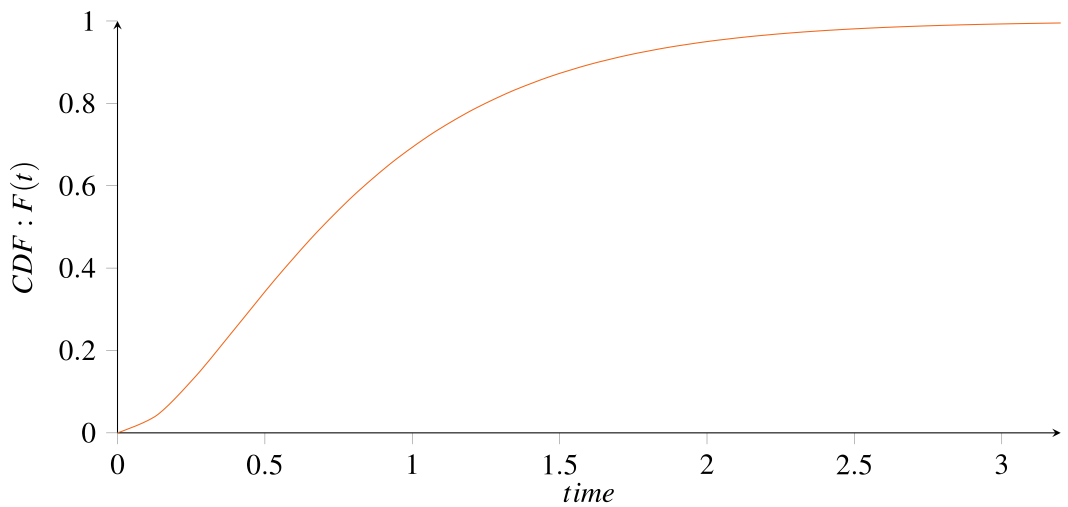



Figure 1.2: CDF of a continuous random variable

### 1.2.2 Probability Density Functions

Another important metric used in reliability is the *probability density function* (PDF) of a continuous random variable $X$. It is a function that describes the relative probability of each value of X and is denoted by $f(x)$.

In other words, the PDF gives the probability that $X$ will take on a value in the infinitesimally small interval from $x$ to $x + dx$.

For a continuous random variable $X$, there is an immediate link between the cumulative distribution function $F(x)$ and the probability density function $f(x)$:

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} \quad (1.15)$$

**Definition 1.2.2 — PDF Properties.** Looking at the properties of the CDF $F(t)$, we can deduce the following properties for the PDF, $f(x)$:

$$\int_0^\infty f(x)dx = 1 \tag{1.16}$$

$$F(t) = \int_0^t f(x)dx \tag{1.17}$$

$$P(X \geq t) = \int_t^\infty f(x)dx \tag{1.18}$$

$$P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a) \tag{1.19}$$

■ **Example 1.3** Taking our previous example of the CDF $F(t) = 1 + 2e^{-3t} - 3e^{-2t}$ we can derive $f(t) = -6e^{-3t} + 6e^{-2t}$, as plotted in Figure 1.3. ■

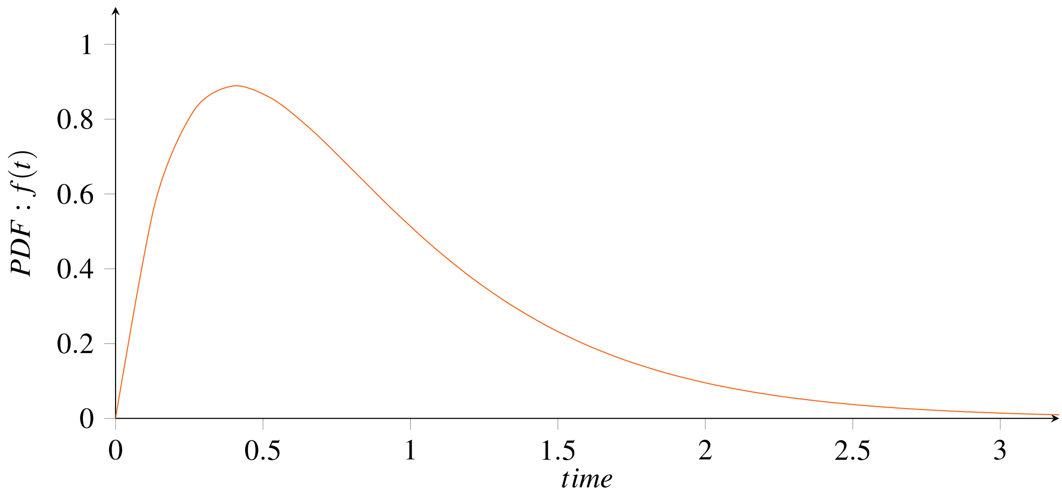



Figure 1.3: PDF of a continuous random variable

### 1.2.3 The Expected Value of a Random Variable

When dealing with random variables, a key objective is to identify the average value that represents the overall outcome of the underlying random experiment.

The *expected value* of a random variable, intuitively, is the long-run average value of repetitions of the experiment it represents.

For example, the expected value in rolling a six-sided die is 3.5 because, roughly speaking, the average of all the numbers that come up in an extremely large number of rolls is very nearly always quite close to three and a half.

**Definition 1.2.3 — Expected value of a random variable.** Suppose random variable $X$ can take value $x_1$ with probability $p_1$, value $x_2$ with probability $p_2$, and so on, up to value $x_k$ with probability $p_k$. Then the expected value of this random variable $X$ is defined as:

$$E[X] = p_1x_1 + p_2x_2 + p_3x_3 + .. + p_kx_k \tag{1.20}$$

Since all probabilities $p_i$ add up to one $(p_1 + p_2 + .. + p_k = 1)$, the expected value can be viewed as the weighted average, with $p_i$ being the weights:

$$E[X] = \frac{p_1x_1 + p_2x_2 + p_3x_3 + .. + p_kx_k}{1} = \frac{p_1x_1 + p_2x_2 + p_3x_3 + .. + p_kx_k}{p_1 + p_2 + .. + p_k} \tag{1.21}$$

■ **Example 1.4** Let X represent the outcome of a roll of a fair six-sided die. More specifically, X will be the number of pips showing on the top face of the die after the toss. The possible values for X are 1, 2, 3, 4, 5, and 6, all equally likely (each having the probability of $\frac{1}{6}$. The expectation of X is

$$E[X] = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6} = 3.5 \tag{1.22}$$

If one rolls the die n times and computes the average (arithmetic mean) of the results, then as n grows, the average will almost surely converge to the expected value, a fact known as the strong law of large numbers. One example sequence of ten rolls of the die is 2, 3, 1, 2, 5, 6, 2, 2, 2, 6, which has the average of 3.1, with the distance of 0.4 from the expected value of 3.5. The convergence is relatively slow: the probability that the average falls within the range 3.5±0.1 is 21.6% for ten rolls, 46.1% for a hundred rolls and 93.7% for a thousand rolls. ■

**Definition 1.2.4 — Expected value for a continuous random variable.** If the probability distribution of X admits a probability density function f(x), then the expected value can be computed as:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \tag{1.23}$$

Since in most reliability calculations the random variable is time, which is quantified from 0 to infinity, we can simplify the previous expression to:

$$E[X] = \int_{0}^{\infty} t f(t) dt, \forall t \geq 0 \tag{1.24}$$

■ **Example 1.5** Taking our previous example of the probability density function $f(t) = -6e^{-3t} + 6e^{-2t}$, we can assess its expected value by plugging it into the previous equation:

$$E[X] = \int_{0}^{\infty} t\left(-6e^{-3t} + 6e^{-2t}\right) dt = \left(\frac{1}{6}e^{-3t}(12t - 9e^{t}(2t+1) + 4) + c\right)\Big|_{0}^{\infty} = \frac{5}{6} \tag{1.25}$$

■

### 1.2.4 Probability Distributions Commonly Used in Reliability

Probability distributions play a crucial role in reliability assessment. For discrete random variables, the binomial and Poisson distributions are particularly useful. When dealing with continuous random variables, the normal, Weibull, and exponential distributions are commonly employed. Additionally, the lognormal, the uniform distribution, Student's t-distribution, and chi-square ($\chi^2$) distribution find applications in specific reliability evaluation scenarios.

#### The Binomial Distribution

The binomial distribution is a discrete probability distribution that calculates the likelihood of obtaining x positive outcomes in n trials, given that the probability of success in each trial is p. It's frequently used to model real-world scenarios involving discrete events, such as the number of heads in multiple coin flips or the number of defective items in a batch.

**Definition 1.2.5 — Probability function of the binomial distribution.** We can attach a probability function to the binomial distribution as follows:

$$p(x) = C_n^x p^x (1-p)^{n-x} \tag{1.26}$$

■ **Example 1.6** Imagine a scenario where you're inspecting a batch of 100 light bulbs to determine the proportion of faulty ones. If you know that the overall defect rate is 5%, the binomial distribution can help you predict the probability of finding a specific number of defective bulbs in your sample. For instance, the probability of finding exactly 2 defective bulbs in your sample can be calculated using the binomial distribution formula:

$$p(x=2) = C_{100}^2 0.05^2 (1-0.05)^{100-2} \approx 0.081(8.1\%) \tag{1.27}$$

■

### The Poisson Distribution

The Poisson distribution is a discrete probability distribution that models the number of events that occur in a fixed interval of time or space, given a known average rate of occurrence. It is named after French mathematician Simeon Denis Poisson, who introduced the distribution in 1837.

The Poisson distribution is better suited than the binomial distribution for events that have a low probability of occurrence.

**Definition 1.2.6 — Probability function of the Poisson distribution.** The function is characterized by the average rate of occurrence $\lambda$ (lambda) that has a value of $\lambda = p \times n$, where n is the number of trials, given that the probability of success in each trial is p:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{1.28}$$

■ **Example 1.7** Using the same light bulb example from the binomial distribution, we can calculate the probability of getting 2 defective light bulbs in a batch of 100, if the defect rate is 5% as:

$$\lambda = 0.05 \times 100 \tag{1.29}$$

$$p(x=2) = \frac{\lambda^2 e^{-\lambda}}{2!} \approx 0.09(9\%) \tag{1.30}$$

■

### The Normal Distribution

The normal distribution, also known as the Gaussian distribution, is one of the most widely used probability distributions in statistics. It is a continuous probability distribution that is bell-shaped, symmetrical, and unimodal. This means that most of the data points in a normal distribution are clustered around the middle of the distribution, and the distribution tails off gradually towards either extreme.

**Definition 1.2.7 — Probability function of the normal distribution.** The normal distribution has the following probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \tag{1.31}$$

The normal distribution is also characterized by its mean $\mu$ and standard deviation $\sigma$. The mean is the average of all the data points in the distribution, and the standard deviation is a measure of how spread out the data is:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}} \tag{1.32}$$

, where $\mu$ is the mean:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1.33}$$

The CDF of the normal distribution is expressed as:

$$F(x) = \int_{-\infty}^{x} f(t),dt = \frac{1}{2}\left[1+\text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{1.34}$$

, where $erf(x)$ is the error function and gives the probability of a random variable with normal distribution falling in the range $[-x,x]$

The normal distribution is often used to model data that is naturally occurring, such as heights of people or test scores. It is also used in many statistical analyses, such as hypothesis testing and confidence intervals.

One particular case is the *standard normal distribution* which is a normal distribution with $\mu = 0$ and $\sigma = 1$ (Figure 1.4)

### The Weibull Distribution

The Weibull distribution is a continuous probability distribution with a wide range of applications in reliability analysis and modeling the lifetime of components. It is characterized by its shape parameter, which determines the shape of the distribution, and its scale parameter, which determines the scale of the distribution.

**Definition 1.2.8 — Probability function of the Weibull distribution.** The Weibull distribution has the following probability density function, where $\lambda$ represents the *scale* parameter and $\beta$ represents the *shape* parameter:

$$f(x) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} e^{-(x/\lambda)^{\beta}} \tag{1.35}$$

The cumulative distribution function (CDF) of the Weibull distribution is given by:

$$F(x) = 1 - e^{-(x/\lambda)^{\beta}} \tag{1.36}$$

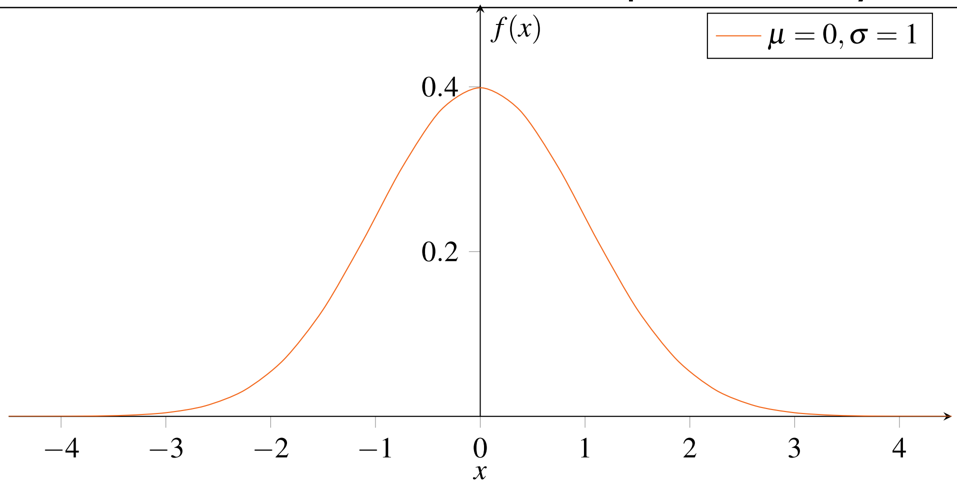



Figure 1.4: Standard normal distribution probability density function

The Weibull distribution is often used to model the failure times of components that experience wear and tear over time. It is also used to model the lifetimes of biological organisms and the times between events in a variety of other applications.

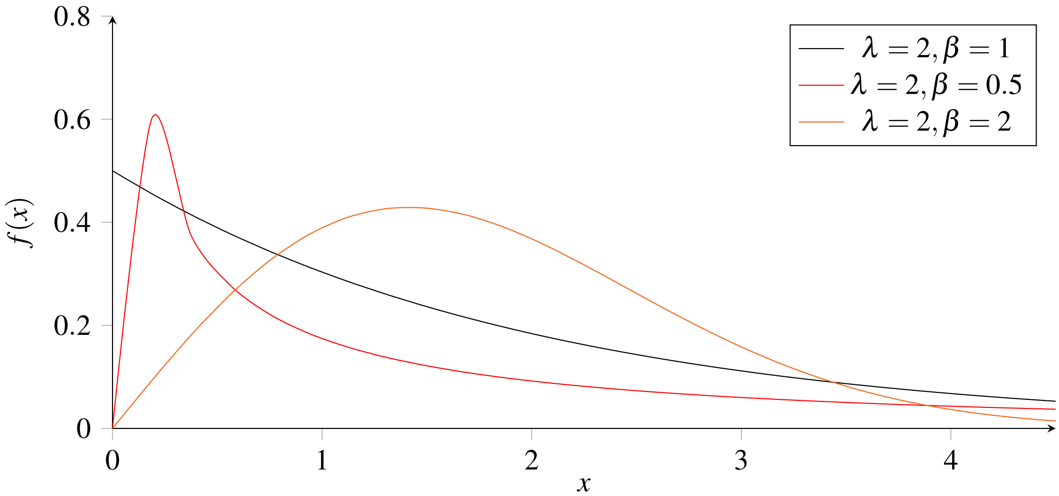



Figure 1.5: Weibull distribution probability density function for different values of $\lambda$ and $\beta$

When the time duration until a failure occurs is represented by the random variable $X$, the Weibull distribution provides a probability distribution where the failure rate is directly proportional to a power of time. The shape parameter, $\beta$, corresponds to that power plus one, allowing for a straightforward interpretation of its value:

- $\beta < 1$: The failure rate decreases over time. This scenario arises when there is a significant amount of "infant mortality," meaning defective items fail early, and the failure rate diminishes as these defective items are gradually eliminated from the population.
- $\beta = 1$: The failure rate remains constant over time. This could indicate that random external events are causing failures or that the underlying failure mechanism is time-independent. In this case, the Weibull distribution simplifies to the exponential distribution.
- $\beta > 1$: The failure rate increases over time. This situation occurs when there is an "aging" process, where components become more prone to failure as time progresses. This could

be due to wear and tear, fatigue, or other cumulative factors that gradually degrade the component's integrity.

In summary, the Weibull distribution offers a flexible framework for modeling failure rates across various scenarios, ranging from decreasing failure rates due to infant mortality to increasing failure rates due to component aging.

#### The Exponential Distribution

The exponential distribution is a continuous probability distribution that describes the time between events in a Poisson process. It is a memoryless distribution, meaning that the probability of an event occurring in a given interval is independent of the time that has elapsed since the last event.

The exponential distribution is a special case of the Weibull distribution in which the shape parameter $\beta = 1$. Therefore, the exponential distribution is characterized only by its rate parameter, $\lambda$ (lambda), which represents the average number of events that occur per unit of time.

**Definition 1.2.9 — Probability function of the exponential distribution.** The probability density function (PDF) of the exponential distribution depends on the *rate* parameter $\lambda$ and can be written as:

$$f(x) = \lambda e^{-\lambda x} \tag{1.37}$$

The cumulative distribution function (CDF) of the exponential distribution is given by:

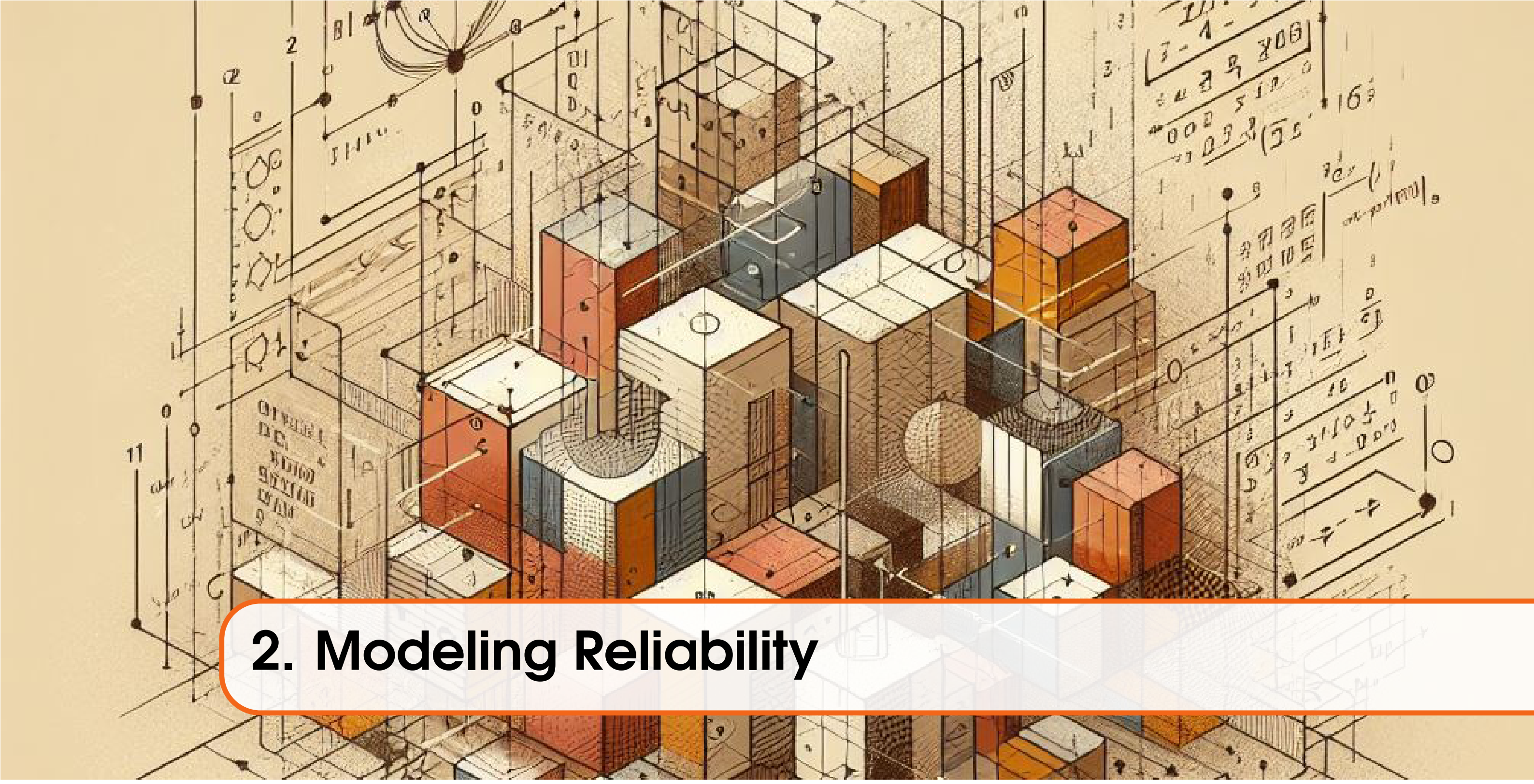
$$F(x) = 1 - e^{-\lambda x} \tag{1.38}$$

The exponential distribution is widely used in reliability analysis to model the time to failure of components. It is also used in other fields, such as queuing theory and survival analysis. Its memoryless property, meaning the likelihood of an event occurring in a specific interval is independent of the time elapsed since the previous event, makes it well-suited for modeling processes with independent inter-arrival times, such as:

- Networking traffic congestion: The exponential distribution is used to model the arrival and departure of data packets in congested networks. This helps in analyzing network performance under varying traffic conditions and identifying bottlenecks.
- Server performance: The exponential distribution is used to analyze the performance of servers in handling requests, such as web servers or file servers. This helps in predicting server response times and ensuring efficient resource utilization.
- Predictive analytics: The exponential distribution is used in predictive analytics models to forecast future events, such as system failures or traffic congestion patterns. This enables businesses to take proactive measures to prevent disruptions and optimize resource allocation.

■ **Example 1.8** We can use the exponential distribution to model the arrival of web requests: suppose that the average number of web requests per minute is 10. The rate parameter can then be calculated as $\lambda = 10$. Using this rate parameter, we can calculate the probability of receiving a web request in any given minute. The probability of receiving a request in the next minute is $1 - e^{-10} \approx 0.995$. ■

■ **Example 1.9** The exponential distribution can also be used to model the waiting time for a web request. The waiting time is the time that it takes for a request to be queued up and processed by the web server. The waiting time can be calculated by using the cumulative distribution function (CDF) of the exponential distribution. For example, the probability of waiting for more than 10 seconds for a web request is $1 - F(10) \approx 0.368$. ■



# 2. Modeling Reliability

## 2.1 Reliability and Availability

Fault tolerance is the ability of a system to continue functioning in spite of malfunctions or faults. As a notion, it is tightly coupled with the concept of reliability, the lack of defects and the availability of a system.

### 2.1.1 Reliability

The reliability of a system is its ability to function correctly over a given time period. Mathematically, the *reliability R(t)* of a system at time t is the probability that the system operates without failure in the interval [0,t), given that the system was performing correctly at time 0. As a probability function, its values lie in the [0, 1] interval.

**Definition 2.1.1 — Reliability function.** We can express the reliability of a system $S$ at time $t$ by:

$$R(t) = P(S \textit{ is fully operational in } [0,t]) \tag{2.1}$$

Notice that we are assuming the system is functioning until it completely stops its normal operation and we are not factoring in the possibility of the system to be repaired. This measure is suitable for applications in which even a momentary disruption can prove costly, for example the autopilot system of a passenger airplane, for which failure would result in catastrophe.

We can consider a random variable $X$ to be the lifetime, or the time until a failure occurs for system $S$. We can also consider $F(t)$ to be the corresponding cumulative distribution function (CDF) for the random variable $X$. We can then write the system reliability as:

$$R(t) = P(X > t) = 1 - F(t) \tag{2.2}$$

It is usually considered that the system in operation at $t = 0$ without any faults, therefore we can write that $R(0) = 1$. Also, as it is deeply ingrained in this Universe that any working system will cease to operate at some future point in time, we can assume $\lim_{t\to\infty} R(t) = 0$.

We can therefore infer that R(t) is a decreasing, continuous, monotone function with values ranging between 0 and 1 in the interval $[0, \infty)$ as in Figure 2.1.

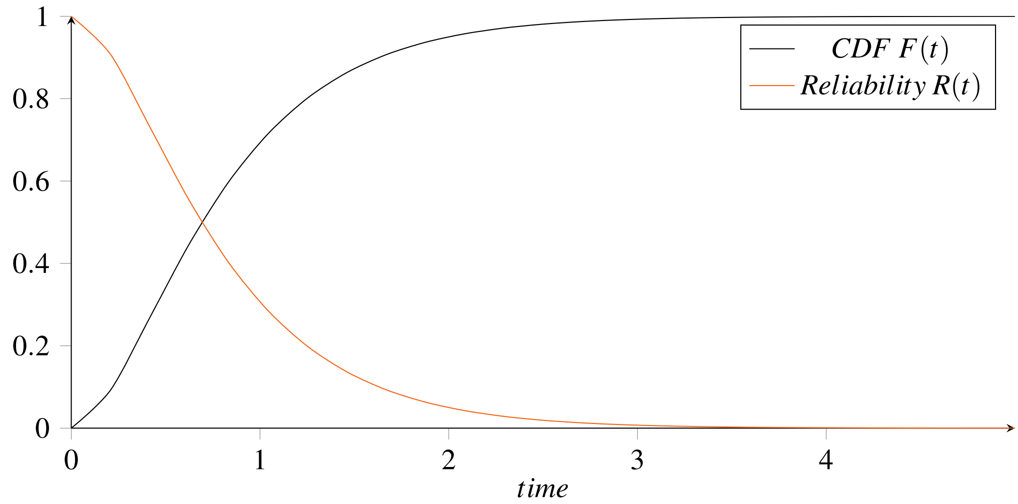



Figure 2.1: Relationship between Reliability and the CDF of a system

Let us consider $f(t)$ the probability density function (PDF) of the system. We have already established its relationship with the CDF of the system to be $F(t) = \int_0^t f(\tau)d\tau$, therefore we can infer that the reliability function, as the inverse of the CDF, can be written as:

$$R(t) = \int_t^\infty f(\tau)d\tau \quad (2.3)$$

Therefore, in a graphical representation, reliability $R(t)$ represents the area under the $f(t)$ curve from t to infinity, as in Figure 2.2.

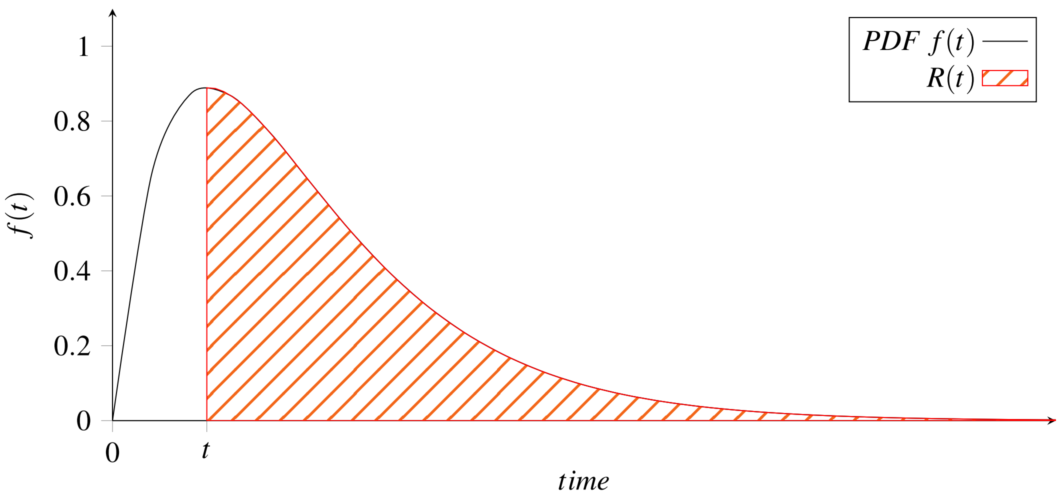



Figure 2.2: Graphical representation of the meaning of Reliability

### 2.1.2 Failure Rate

So far we have quantified the probability of a fault not happening in a given time interval. It is also of interest to determine the probability a fault will happen at a given time, or, in a quantifiable small interval $[t, t+\Delta t]$, given that the system has functioned properly until time $t$. We can write this probability as:

$$P(t < X < t+\Delta t \mid X > t) = \frac{P(t < X < t+\Delta t)}{P(X > t)} = \frac{F(t+\Delta t) - F(t)}{R(t)} \tag{2.4}$$

**Definition 2.1.2 — Failure rate.** The *instantaneous failure rate*, also named *the hazard function* or the age-dependent *failure rate* of the system is defined as:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{R(t)\Delta t} = \lim_{\Delta t \to 0} \frac{R(t) - R(t+\Delta t)}{R(t)\Delta t} = \frac{f(t)}{R(t)} \tag{2.5}$$

Plotting the failure rate as a function of time to time yields a distinctive shape called the "bathtub curve", such as the one in Figure 2.3.

Manufacturing or design defects tend to lead to failure in the initial stage of a product's life. This stage is also known as "infant mortality" and is characterized by a large but decreasing failure rate, as more products are eliminated from the initial batch due to failures. Infant mortality can be eliminated at the manufacturing stage through system testing and accelerated aging of the product before it is sold or released into circulation.

Once this stage is over, the product enters a period in which failure rate is constant. This is usually the stage at which the product experiences its entire useful life. Fault rate is non-zero but low and is typically due to environmental conditions.

In the last stage, failure rate increases due to extensive wear. A failure becomes more likely as more time goes by, until all of the products from the initial population become defective.

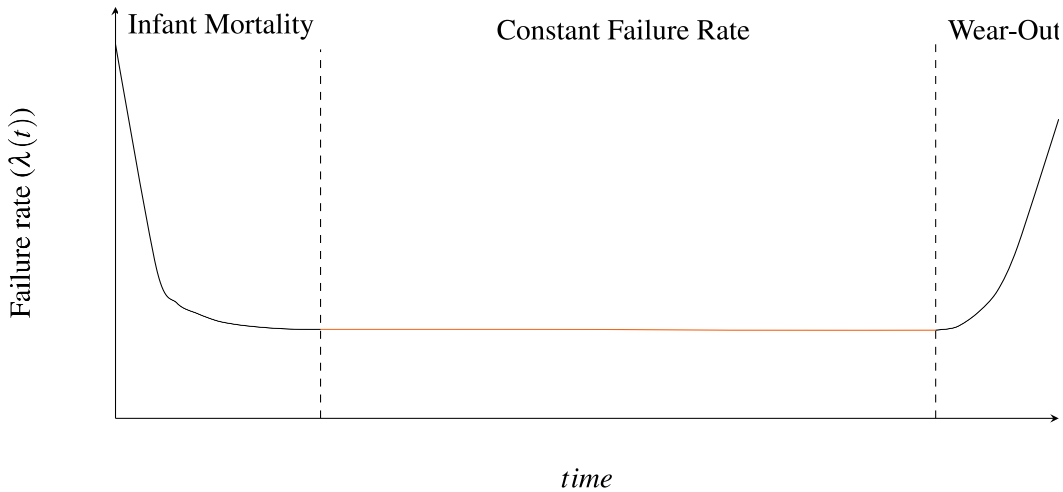



Figure 2.3: Bathtub curve for the failure rate of a system

In engineering and reliability analysis, the failure rate $\lambda$ of a system or component is the frequency with which it fails, expressed in failures per unit of time. It represents the probability of a failure occurring within a specified time interval. The failure rate is a key metric for understanding the reliability and lifespan of a system, and it can be used to make informed decisions about maintenance and replacement schedules.

Failure rates can be calculated in different ways, depending on the type of system and the available data. For example, the failure rate of a component may be calculated based on historical data of failures, or it may be estimated using statistical methods or accelerated life tests.

■ **Example 2.1** Usually, failure rate is expressed as a constant and is measured in failures per unit of time. For example, a light bulb might have a measured failure rate of 0.001 failures per hour, a solid-state drive might have 0.000015 failures per hour and a passenger airplane a typical failure rate of 0.000000001 per hour ■

Failure rates can be affected by a number of factors, including the design of the system, the quality of the components, and the operating environment.

The impact of these factors can be expressed through the following empirical failure rate formula:

$$\lambda = \pi_L \pi_Q (C_1 \pi_T \pi_V + C_2 \pi_E) \tag{2.6}$$

where the notations are as follows:

- $\lambda$ - Failure rate of component.
- $\pi_L$ - Learning factor, associated with how mature the technology is.
- $\pi_Q$ - Quality factor, representing manufacturing process quality control (ranging from 0.25 to 20.00).
- $\pi_T$ - Temperature factor, with values ranging from 0.1 to 1000. It is proportional to $e^{\frac{E_a}{kT}}$, where $E_a$ is the activation energy in electron-volts associated with the technology, k is the Boltzmann constant ($8.6173 \times 10^{-5} eV/K$), and T is the temperature in Kelvin.
- $\pi_V$ - Voltage stress factor for CMOS devices; can range from 1 to 10, depending on the supply voltage and the temperature; does not apply to other technologies (where it is set to 1).
- $\pi_E$ - Environment shock factor; ranges from very low (about 0.4), when the component is in an air-conditioned office environment, to very high (13.0) when it is in a harsh environment.
- $C_1, C_2$ - Complexity factors; functions of the number of gates on the chip and the number of pins in the package.

This formula was taken from MIL-HDBK-217E, MILITARY HANDBOOK: RELIABILITY PREDICTION OF ELECTRONIC EQUIPMENT (27 OCT 1986), written by the U.S. Department of Defense to address reliability modeling of their electronic equipment.

In the harsh environment of space, where charged particles abound and extreme temperature fluctuations occur, electronic devices are more prone to malfunctions compared to their counterparts in the controlled climate of air-conditioned offices. Similarly, computers in automobiles, subjected to intense heat and vibrations, and those in industrial settings, exposed to harsh conditions, face elevated failure rates.

Software failure rate usually decreases as a function of time. A possible curve is shown in Figure 2.4. The three phases of evolution are: test/debug (I), useful life (II) and obsolescence (III).

Software failure rate during useful life depends on the following factors:

1. software process used to develop the design and code
2. complexity of software,

3. size of software,
4. experience of the development team,
5. percentage of code reused from a previous stable project,
6. rigor and depth of testing at test/debug (I) phase.

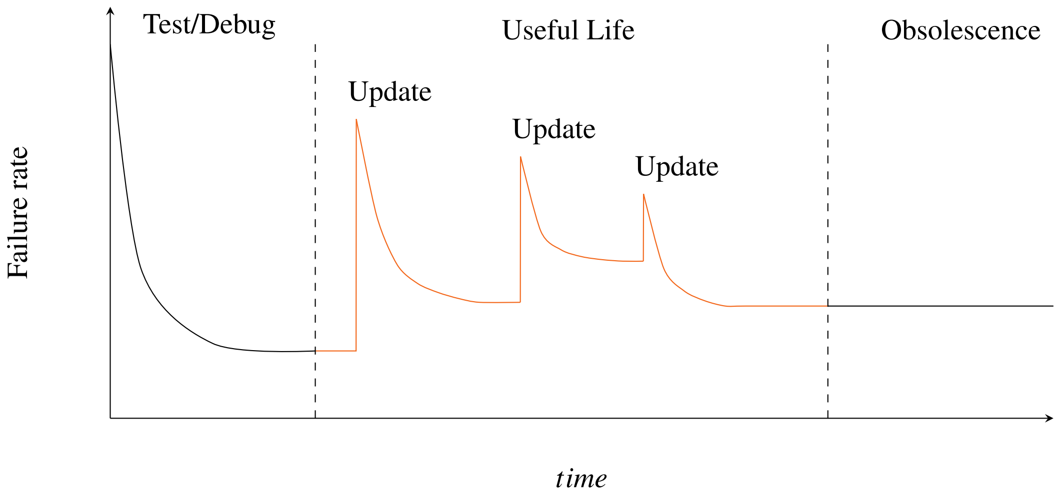



Figure 2.4: The failure rate curve for software versus time

Compared to hardware, software failure curves exhibit two distinct characteristics. Firstly, software's failure rate tends to spike after each feature update during its useful-life phase. This is because upgrades often introduce new functionalities, leading to increased complexity and consequently a higher likelihood of faults. However, following the initial surge in failures, the rate gradually stabilizes, partly due to bug fixes implemented after the upgrades. Secondly, unlike hardware, software does not experience a progressive increase in failure rate during its final phase. In this stage, software approaches obsolescence, and the need for further upgrades or modifications diminishes.

### 2.1.3 Mean Time Between Failures

Another metric to assess a system's fault tolerance is the *Mean Time Between Failures* (MTBF). This parameter is derived from observing the system's behavior during its operational lifespan. The simplest model incorporating fault tolerance assumes the system transitions between two states: fully operational and completely failed. These transitions occur upon failure or after system repair, as depicted in Figure 2.5.

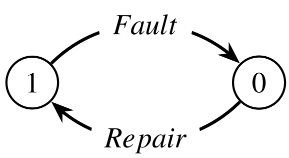



Figure 2.5: Simple state transition graph for a system with failure and repair

This two-state model can be applied to simple systems like light bulbs, which can either illuminate or be burned out, and wires in circuits, which can either be connected or interrupted. It can also be extended to more complex systems like cars and web servers, but the definitions of "operational" and "failed" need to be tailored to the specific context. For instance, an operational web server would be fully responsive to client requests, while a failed web server could be completely unresponsive due to a crash or undergoing maintenance.

Visualizing the system's behavior over time reveals alternating intervals of operational periods and repair downtime. As depicted in Figure 2.6, the system initially operates until it encounters a failure, marking the end of the first Time-To-Fail ($TTF_1$) interval. Subsequently, the system transitions into the first Time-To-Repair ($TTR_1$) interval, representing the time it takes to restore functionality. This pattern of alternating operational and repair intervals persists throughout the system's lifespan.

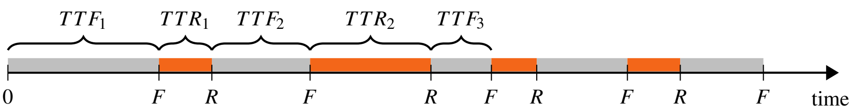



Figure 2.6: The lifetime of a system with consecutive functioning and repair episodes

Measuring these intervals and averaging their values over a long observational period yields two important metrics: the *Mean Time to Failure*, or *MTTF* which is an average of all Time-To-Fail (TTF) intervals, and the *Mean Time to Repair*, or *MTTR*, which is the average of all Time-To-Repair intervals.

$$MTTF = \sum_i \frac{TTF_i}{n} \qquad MTTR = \sum_i \frac{TTR_i}{n} \tag{2.7}$$

**Definition 2.1.3 — Mean Time Between Failures.** Using the above two notions, we can define the *Mean Time Between Failures, MTBF* as the average expected time between two failures for a repairable system:

$$MTBF = MTTF + MTTR \tag{2.8}$$

## 2.1.4 Availability

Few systems are designed to run indefinitely without downtime or maintenance. Typically, we care not only about system reliability but also about failure frequency and recovery time. For example, for web servers, we aim to maximize uptime, the proportion of time the system is operational. This metric is captured by *Availability*.

The system's *Availability A(t)* at time $t$ denotes the likelihood that the system is operating correctly at that specific moment. A(t) is alternatively known as point availability or instantaneous availability. This metric is suitable for scenarios where continuous performance is not crucial, yet prolonged system downtime would incur substantial costs. For instance, an airline reservation system requires high availability to avoid customer dissatisfaction and revenue loss due to downtime. However, occasional very brief failures can be tolerable in such a system.

**Definition 2.1.4 — Interval Availability.** Often it is necessary to determine the *Interval*, or *Mission Availability*. It is defined by:

$$A(T) = \frac{1}{T}\int_0^T A(t)dt \tag{2.9}$$

*A(T)* is the value of the point availability averaged over some interval of time $T$. This interval might be the life-time of a system or the time to accomplish some particular task.

Ultimately, it is frequently observed that following an initial transient impact, point availability stabilizes to a time-independent value. In such instances, we refer to it as *Steady-state availability*, alternatively recognized as Long-term Availability denoted by $A(\infty)$.

**Definition 2.1.5 — Steady-State Availability.**

$$A(\infty) = \lim_{T \to \infty} A(T) = \lim_{T \to \infty} \left( \frac{1}{T} \int_0^T A(t)dt \right) \tag{2.10}$$

The interpretation of $A(\infty)$ lies in its representation as the probability that the system will be operational at a randomly chosen moment, and its relevance is confined to systems incorporating the repair of defective components. In cases where a system is irrepairable, the point availability $A(t)$ aligns with the system's reliability—namely, the probability that the system remains operational from time 0 to t. Consequently, as the time duration $T$ approaches infinity, the steady-state availability of a non-repairable system converges to zero:

$$A(\infty) = 0 \tag{2.11}$$

The long-term availability $A(\infty)$, or more simply written $A$, can be calculated from MTTF, MTBF, and MTTR as follows:

$$A = \frac{MTTF}{MTBF} = \frac{MTTF}{MTTF + MTTR} \tag{2.12}$$

**Example 2.2** A system with low reliability can still exhibit high availability. For example, imagine a communication channel that is down every couple of hours but its takes only 3 seconds to reestablish connection. We can compute the MTBF as 2 hours (7200 seconds) and the MTR as 3 seconds. Even if the reliability of such a communication link is low, its availability is quite high: $A = 7200/7203 = 99.96\%$. ■

Steady-state availability is often specified in terms of downtime per year. Table 2.1 shows examples for some of the values for availability and the corresponding downtime.

| Availability(%) | Downtime per year | Downtime per month | Downtime per week |
|---|---|---|---|
| 90% ("one nine") | 36.5 days | 72 hours | 16.8 hours |
| 99% ("two nines") | 3.65 days | 7.2 hours | 1.68 hours |
| 99.9% ("three nines") | 8.76 hours | 43.2 minutes | 10.1 minutes |
| 99.99% ("four nines") | 52.56 minutes | 4.32 minutes | 1.01 minutes |
| 99.999% ("five nines") | 5.26 minutes | 25.9 seconds | 6.05 seconds |
| 99.9999% ("six nines") | 31.5 seconds | 2.59 seconds | 0.605 seconds |

Table 2.1: Availability and the corresponding downtime per year.

Availability stands as an essential metric, particularly for systems that can endure short interruptions. Networked systems, such as telephone switching and web servers, provide concrete illustrations of this principle. Telephone users anticipate seamless call completion without disruptions, accepting an annual downtime of up to three minutes. Research indicates that web users' tolerance diminishes if websites take more than eight seconds to display results. Consequently, these websites must maintain continuous availability and swift responsiveness, even amid substantial concurrent user traffic.

The electrical power control system serves as another notable example. Consumers expect an uninterrupted power supply 24/7, regardless of weather conditions. Prolonged power outages can pose health risks, disrupting essential services like water pumps, heating, lighting, and medical care. Industries also face substantial financial losses in the event of power disruptions.

## 2.2 Failure Rate, Reliability, and Mean Time to Failure for an Exponential Fault Distribution

In this section we will approach the derivation of reliability and Mean Time Between Failures (MTBF) from the fundamental concept of failure rate. We focus on a component operational at t=0 and sustained in operation until encountering a failure. Our consideration involves the assumption that failures adhere to an exponential probability distribution.

Let's now assume that all failures are permanent and irreparable. Let T represent the random variable "lifetime of the component" (indicating the time until failure). Additionally, let f(t) and F(t) denote the probability density function (PDF) of T and the cumulative distribution function (CDF) of T, respectively. As established in the preceding chapter, we determined that these functions are interrelated as follows:

$$f(t) = \frac{dF(t)}{dt} \qquad F(t) = \int_0^t f(\tau)d\tau \quad \forall t \geq 0 \tag{2.13}$$

The PDF f(t) can be understood as the probability the system will fail at time t. For a tiny $\Delta t, f(t)\Delta t \approx Prob(t \leq T \leq t + \Delta t)$. f(t) is a probability density function, therefore the following will be true:

$$\int_0^\infty f(t)dt = 1 \quad f(t) \geq 0, \forall t \geq 0 \tag{2.14}$$

In the context of the random variable defined above as the lifetime of a component, F(t) can be viewed as probability that the system will exhibit a failure anywhere in (0,t]:

$$F(t) = Prob(t \leq T) \tag{2.15}$$

Conversely, R(t) if the inverse of F(t) and can be defined as the probability the system functions without failure until time t:

$$R(t) = Prob(T > t) = 1 - F(t) \tag{2.16}$$

As we defined it in the previous chapter, the *failure rate* of a system $\lambda(t)$ gives us the probability the system will fail at time t:

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{2.17}$$

Since $f(t) = \frac{dF(t)}{dt} = -\frac{dR(t)}{dt}$, we can rewrite the expression above:

$$\lambda(t) = -\frac{dR(t)}{dt}\frac{1}{R(t)} \tag{2.18}$$

As previously mentioned, for the active life of a system we can consider the failure rate to be constant, $\lambda(t) = \lambda$. This simplifies the previous equation and makes it trivial to solve:

$$\frac{dR(t)}{dt} = -\lambda R(t) \tag{2.19}$$

We can assume R(0) = 1 and we can solve 2.19 for R(t):

$$R(t) = e^{-\lambda t} \tag{2.20}$$

This equation links the reliability of a system to its constant failure rate $\lambda$, if the system is within its normal operational lifetime (the flat constant region of the bathtub curve).

This is the *exponential failure law* and it is plotted in Figure 2.7.

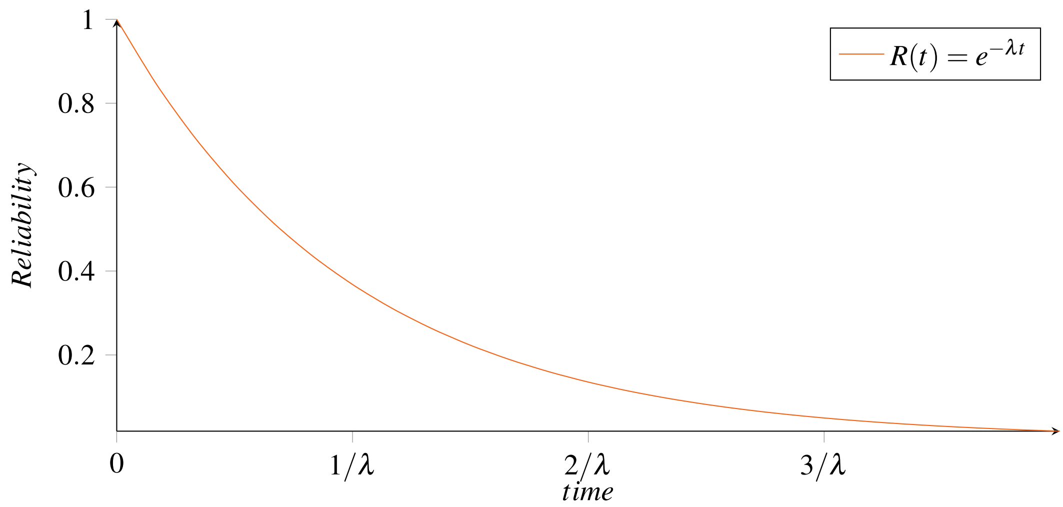



Figure 2.7: Reliability function for an exponential distribution of faults

The exponential failure law is very valuable for the analysis of reliability of components and systems in hardware. However, it can only be used in cases when the assumption that the failure rate is constant is adequate.

To summarize the definitions we have derived above:

$$f(t) = \lambda e^{-\lambda t} \qquad F(t) = 1 - e^{-\lambda t} \qquad R(t) = e^{-\lambda t} \qquad for\, t \geq 0 \tag{2.21}$$

**Definition 2.2.1 — MTBF for an exponential fault distribution.** By definition, the MTBF of an irreparable component is equal to its expected lifetime, E[T]. As the random variable is time, which is always greater than zero, we can rewrite 1.23 as:

$$MTBF = E[T] = \int_0^{\infty} t f(t) dt \tag{2.22}$$

Substituting $f(t) = -\frac{dR(t)}{dt}$ we get,

$$MTBF = -\int_0^\infty t\frac{dR(t)}{dt}dt = -tR(t)\mid_0^\infty + \int_0^\infty R(t)dt \tag{2.23}$$

The value of $-tR(t)$ is equal to 0 at $t = 0$ and also to zero at $t \to \infty$, as the reliability of every system asymptotically drops to zero given a long enough time, $(R(\infty) = 0)$. Thus, we can write:

$$MTBF = \int_0^\infty R(t)dt \tag{2.24}$$

Given an exponential reliability function with a constant failure rate $\lambda$, we can rewrite 2.24 as:

$$MTBF = \int_0^\infty R(t)dt = \int_0^\infty e^{-\lambda t}dt = \frac{1}{\lambda} \tag{2.25}$$

### 2.2.1 Non-constant Failure Rate

Most reliability calculations imply a constant failure rate $\lambda = ct$. partly due to the fact that the analyzed system is thought to be in its operational lifetime, where failure occurrence is random and partly because a failure rate that is dependant of time will further complicate reliability formulas.

If we would like to model the reliability of a system in its "infant mortality" or its "wear-out" phases from Figure 2.3, we will need to employ the Weibull probability distribution, which models appropriately these states.

As presented in a previous section of this work, the Weibull distribution has two parameters, $\lambda$ - the shape parameter and $\beta$ - the scale parameter. We can rewrite equation 1.35 for $t \geq 0$ and get the PDF:

$$f(t) = \lambda\beta t^{\beta-1}e^{-\lambda t^\beta} \tag{2.26}$$

We can derive the failure rate, but in this case it will not be constant:

$$\lambda(t) = \lambda\beta t^{\beta-1} \tag{2.27}$$

By varying the value of $\beta$ we have the following three cases:

- $\beta > 1$: failure rate is a decreasing function of time (used for modeling infant mortality).
- $\beta = 1$: failure rate is constant $\lambda$ and we use the reliability formulas derived in the previous section for the exponential distribution.
- $\beta < 1$: failure rate is an increasing function of time (modeling wear-out).

We can also derive the reliability function when using a Weibull distribution by plugging in the new formula for $\lambda(t)$ in equation 2.18 and solving for R(t):

$$R(t) = e^{-\lambda t^\beta} \tag{2.28}$$

Note that reliability now depends also of $\beta$ and a similar discussion can be made for R(t)'s properties for $\beta < 1$, $beta = 1$ and $beta > 1$.

We can also derive the MTBF from this new reliability formula as:

$$MTBF = \int_0^\infty R(t)dt = \frac{\Gamma(\beta^{-1})}{\beta\lambda^{\beta-1}} \tag{2.29}$$

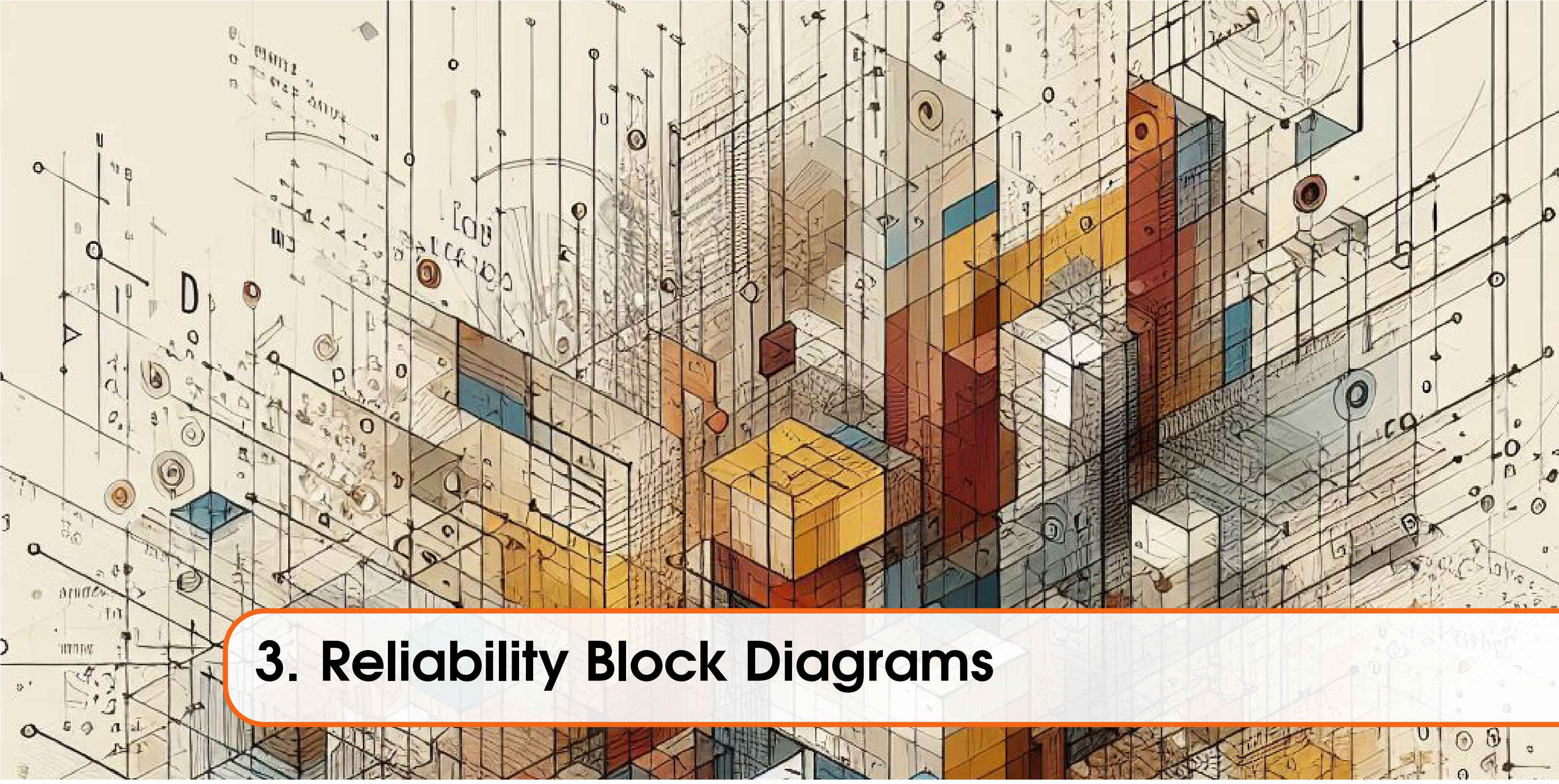
**Definition 2.2.2 — Gamma function.** $\Gamma(x)$ is the gamma function, which is an extension of the factorial function for real number values. It can be computed that $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy$, and, as a factorial function, it will also satisfy the following:

$\Gamma(0) = \Gamma(1) = 1$
$\Gamma(x+1) = (x)\Gamma(x) \ \ \forall x > 1$
If x is a positive integer, then $\Gamma(x) = (x-1)!$

# Part Two


**3 Reliability Block Diagrams . . . . . . . . . . . . . 39**

3.1 Modeling Reliability Through Blocks
3.2 Series Structures
3.3 Parallel Structures
3.4 Combination of Series and Parallel
3.5 k Out of n Systems
3.6 Series-Parallel and Parallel-Series Systems
3.7 Non-Decomposable Systems
3.8 Majority Voted Redundancy
3.9 Standby-Sparing



# 3. Reliability Block Diagrams

## 3.1 Modeling Reliability Through Blocks

Within the realm of combinatorial reliability models, reliability block diagrams (RBDs) have emerged as the most established and prevalent method for analyzing system reliability. These diagrams offer a simplified representation of a system's structure and component reliability, utilizing blocks to denote individual components and interconnections between blocks to depict the operational dependencies among them.

RBDs provide a clear and intuitive graphical representation of system structure and dependencies, facilitating a straightforward understanding of system behavior. They also enable the calculation of various reliability metrics, such as system availability, reliability, and mean time to failure (MTTF), providing valuable insights into system performance.

Also, RBDs can be applied to a wide range of systems, from simple configurations to complex networks, making them a versatile tool for reliability assessment.

Using RBDs, we can represent, for example, components that are tied in series, as in Figure 3.1 a), components that are linked in parallel, as in Figure 3.1(b) or more complex systems that are combinations of series-parallel connections.

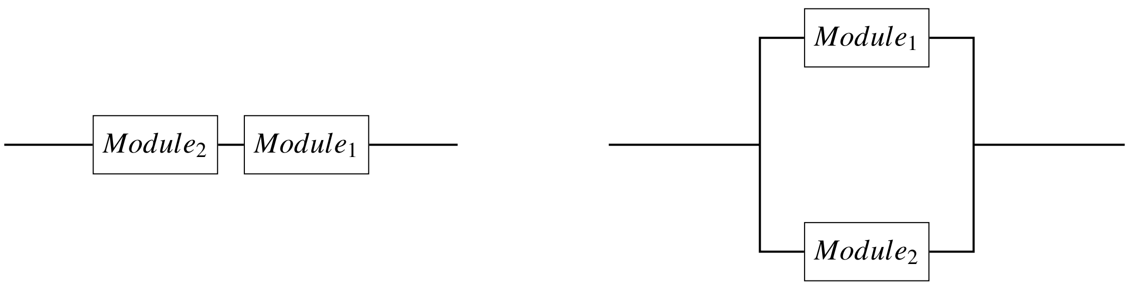



Figure 3.1: Reliability diagrams for a series (a) and a parallel (b) system

For a combination of series, parallel RBD, consider a computing unit that consists of two processor cores that are connected to a shared RAM memory. The reliability block diagram for this system is depicted in Figure 3.2. The processors are arranged in parallel, as only one functioning processor is necessary for system operation. The memory, on the other hand, is connected in series, as its failure would render the entire system inoperable.

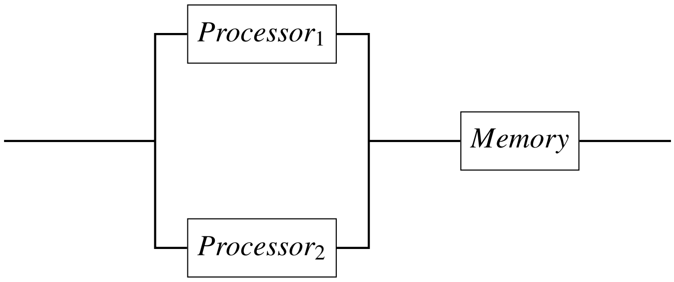



Figure 3.2: Reliability diagram for a three-component system

Despite their widespread use, reliability block diagrams (RBDs) exhibit certain limitations that restrict their applicability in certain situations.

Primarily, RBDs adhere to a simplified assumption that system components can only exist in either an operational or failed state. Additionally, they assume that the system configuration remains constant throughout the mission. These assumptions preclude the modeling of standby components, repair processes, and sophisticated fault detection and recovery mechanisms.

Furthermore, RBDs operate under the assumption of independent component failures. This assumption implies that the sequence in which components fail does not affect the overall system reliability. However, in reality, the order of failures can significantly impact the system's ability to function.

These limitations suggest that RBDs may not be suitable for modeling complex systems where standby components, repair mechanisms, or intricate fault detection and recovery strategies are employed, or where the order of component failures significantly affects system reliability.

In this section, we consider some canonical structures, out of which more complex structures can be constructed.

We start with the basic series and parallel structures, continue with non-series/parallel ones, and then describe some of the many resilient structures that incorporate redundant components (next referred to as modules).

In the next sub-sections, we will use the following notations:

- $R_i = p_i$, the reliability of block $i$, meaning the probability that functional block $i$ is working properly
- $Q_i = q_i = 1 - p_i$, the probability that functional block $i$ is defective
- $R$, the reliability of the whole system (i.e. the probability that the whole system is functioning properly)
- $Q = 1 - R$, the probability that the whole system is defective

## 3.2 Series Structures

A series system consists of N interconnected modules, where the malfunction of any individual module leads to the entire system's failure. It is crucial to note that the diagram in Figure 3.3 represents a reliability diagram, not necessarily an electrical circuit. The output of the first module may not always directly connect to the input of the second module.

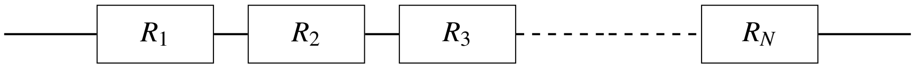



Figure 3.3: Reliability diagram for a series system

For such a system to function properly, all its units must function properly. Assuming that the modules in Figure 3.3 fail independently of each other, the reliability of the entire series system is the product of the reliabilities of its N modules.

Denoting with $R_s(t)$ the reliability of the whole system we can write the following,

$$R_s = P(1 \wedge 2 \wedge 3 \wedge .. \wedge N) = p_1 \cdot p_2 \cdot p_3 \cdot \cdot ... \cdot p_N \tag{3.1}$$

If we denote by $R_i(t)$ the reliability of module i, we can rewrite the equation,

$$R_s(t) = \prod_{i=1}^{N} R_i(t) \tag{3.2}$$

Also,

$$Q_s(t) = 1 - R_s(t) = 1 - \prod_{i=1}^{N}(1 - Q_i(t)) \tag{3.3}$$

$$\prod_{i=1}^{N}(1-Q_i(t)) = 1-(Q_1(t)+Q_2(t)+..+Q_N(t))+(Q_1(t)Q_2(t)+Q_1(t)Q_3(t)+..+Q_{N-1}(t)Q_N(t))-.. \tag{3.4}$$

Usually, in order for the whole system to have a high reliability, each block needs to have a high reliability $R_i \geq 0.9$, which means that $Q_i$ is very small, so we can neglect factors that contain a product of at least two $Q_i$ factors. Therefore, we can rewrite Equation 3.4 as,

$$\prod_{i=1}^{N}(1 - Q_i(t)) \approx 1 - (Q_1(t) + Q_2(t) + .. + Q_N(t)) = 1 - \sum_{i=1}^{N} Q_i(t) \tag{3.5}$$

If we input this into Equation 3.3, we get:

$$Q_s(t) = 1 - (1 - \prod_{i=1}^{N}(1 - Q_i(t))) = \sum_{i=1}^{N} Q_i(t) \tag{3.6}$$

If module $i$ has a constant failure rate, denoted by $\lambda_i$, then, $R_i(t) = e^{-\lambda_i t}$, and consequently:

$$R_s(t) = \prod_{i=1}^{N} e^{\lambda_i t} = e^{-\sum_{i=1}^{N} \lambda_i t} = e^{-\lambda_s t} \quad (3.7)$$

From 3.7 we see that the series system also follows an exponential repartition and has a constant failure rate equal to $\lambda_s$ (the sum of the individual failure rates). Using the relation derived in 2.24, the MTBF of the series system is therefore

$$MTBF_S = \frac{1}{\lambda_s} = \frac{1}{\sum_{i=1}^{N} \lambda_i} = \frac{1}{\sum_{i=1}^{N} \frac{1}{MTBF_i}} \quad (3.8)$$

This means that:

$$MTBF_S < MTBF_i, \forall i = \overline{1,N} \quad (3.9)$$

The failure rate of a series system increases with the number of units that are linked in series.

$$\lambda_S = \sum_{i=1}^{N} \lambda_i \quad (3.10)$$

For identical systems, with the same failure rate, $\lambda_i = \lambda$, we can simplify the equation above to:

$$\lambda_S = N\lambda \quad (3.11)$$

■ **Example 3.1** Consider the series structure in Figure 3.4. The four modules in this diagram represent the instruction decode unit ($R_{ID}$), execution unit ($R_{EU}$), data cache ($R_{DC}$), and instruction cache ($R_{IC}$) in a microprocessor. All four units must be fault-free for the microprocessor to function, although the way they are physically connected does not resemble a series system.

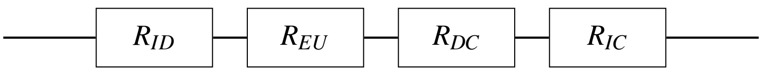



Figure 3.4: Reliability diagram for a series system

Let's assume the modules have the following constant reliabilities: $R_{ID} = 0.9$, $R_{EU} = 0.95$, $R_{DC} = 0.99$, $R_{IC} = 0.89$. Then, the total reliability of the microprocessor is:

$$R_S = R_{ID} \cdot R_{EU} \cdot R_{DC} \cdot R_{IC} = 0.9 \cdot 0.95 \cdot 0.99 \cdot 0.89 \approx 0.75 \quad (3.12)$$

As a general rule, the reliability of a series structure is lower than the reliability of its individual components. This can be explained by the fact there are more states in which two modules can fail when working together than individually. It can be noted that, for the processor to have a 99.9% reliability (which is a common figure for today's PCs), the reliability of each of the four subsystems needs to be at least $R = \sqrt[4]{0.999} \approx 0.9998$ If we increase the number of components that are linked in series even further, the overall reliability will decrease asymptotically towards zero.

For example, if we link together an ever increasing number of systems with reliability $R = 0.9$, we will get the following decrease in overall reliability:

- 2 systems: $R_S = 0.9^2 = 81\%$
- 3 systems: $R_S = 0.9^3 = 72.9\%$
- 4 systems: $R_S = 0.9^4 = 65.61\%$
- 5 systems: $R_S = 0.9^5 = 59.05\%$
- 6 systems: $R_S = 0.9^6 = 53.14\%$

This decrease in reliability is shown in Figure 3.5.

■

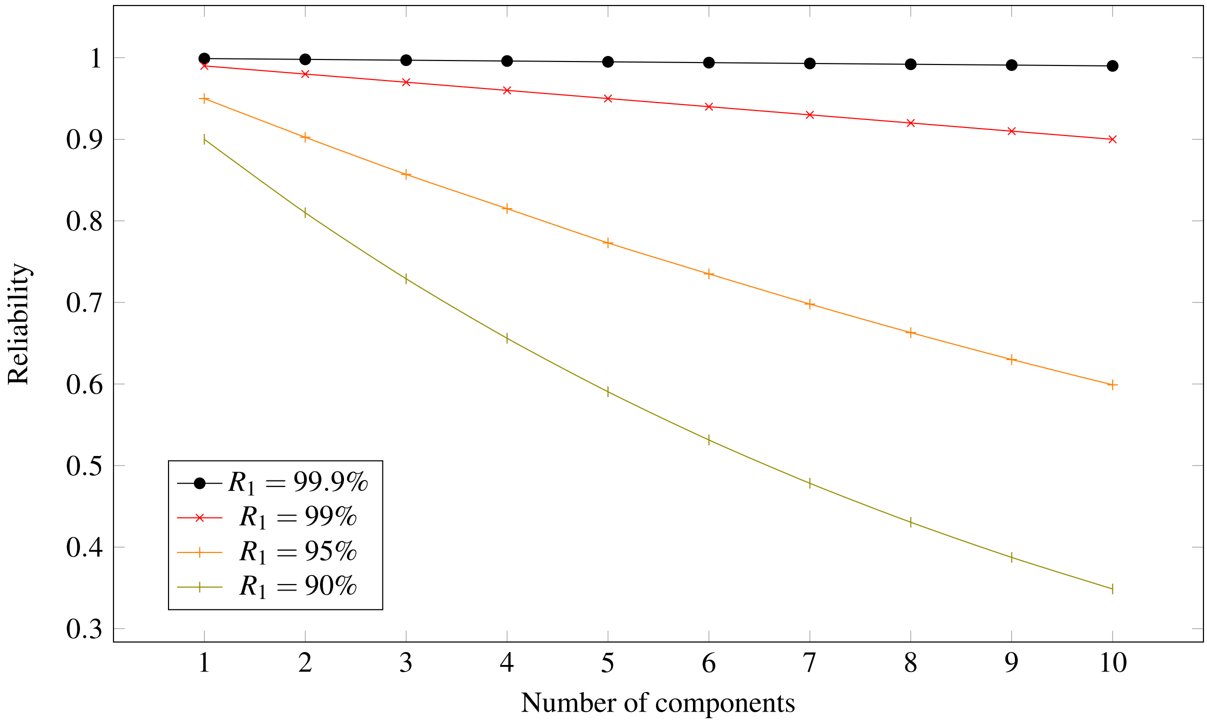



Figure 3.5: Reliability of a series system with increasing number of identical components: single component reliability ($R_1$) from 90% to 99.9%

## 3.3 Parallel Structures

A parallel system is defined as a set of N modules connected together so that it requires the failure of all the modules for the system to fail, as in Figure 3.6.

To get to a reliability formula for the parallel structure, we will have to first consider the probability that the whole system will malfunction ($Q_p(t)$). This will happen when all the blocks malfunction, so block 1, block 2 through block N are all defective. We can express that by:

$$Q_P(t) = P(\overline{1} \wedge \overline{2} \wedge .. \wedge \overline{N}) = \prod_{i=1}^{N} Q_i(t) \tag{3.13}$$

where all blocks are independent and $Q_i(t)$ is the probability that block $i$ is faulty.

We can therefore express the reliability of a parallel structure of N modules by:

$$R_P(t) = 1 - Q_P(t) = 1 - \prod_{i=1}^{N} Q_i(t) = 1 - \prod_{i=1}^{N} (1 - R_i(t)) \tag{3.14}$$

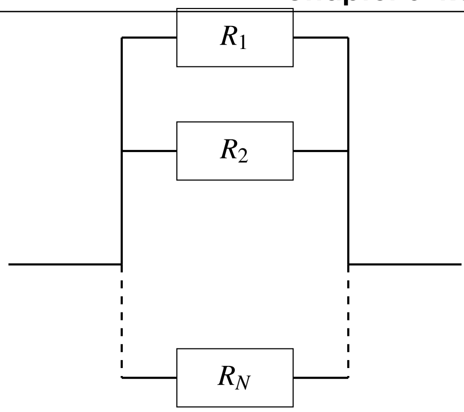



Figure 3.6: Reliability diagram for a parallel system

If every module has a constant failure rate $\lambda_i$, then we can write:

$$R_P(t) = 1 - \prod_{i=1}^{N}(1 - e^{-\lambda_i t}) = \sum_{i=1}^{N} e^{-\lambda_i t} - \sum_{i=1,j=1,i\neq j}^{N} e^{-(\lambda_i+\lambda_j)t} + .. + (-1)^{N+1}\prod_{k=1}^{N} e^{-\lambda_k t} \quad (3.15)$$

To calculate the MTBF, we follow the rule derived in Equation 2.24:

$$MTBF_P = \int_0^\infty R_P(t)dt = \quad (3.16)$$

$$= \sum_{i=1}^{N}\int_0^\infty e^{-\lambda_i t}dt - \sum_{i=1,j=1,i\neq j}^{N}\int_0^\infty e^{-(\lambda_i+\lambda_j)t}dt + .. + (-1)^{N+1}\int_0^\infty \left(\prod_{k=1}^{N} e^{-\lambda_k t}\right)dt = \quad (3.17)$$

$$= \sum_{i=1}^{N}\int_0^\infty e^{-\lambda_i t}dt - \sum_{i=1,j=1,i\neq j}^{N}\int_0^\infty e^{-(\lambda_i+\lambda_j)t}dt + .. + (-1)^{N+1}\int_0^\infty \left(e^{-\Sigma_{k=1}^{N}\lambda_k t}\right)dt \quad (3.18)$$

We can simplify Equation 3.18 by integration:

$$MTBF_P = \sum_{i=1}^{N}\frac{1}{\lambda_i} - \sum_{i=1,j=1,i\neq j}^{N}\frac{1}{\lambda_i+\lambda_j} + .. + (-1)^{N+1}\frac{1}{\Sigma_{k=1}^{N}\lambda_k} \quad (3.19)$$

If all systems have the same failure rate $\lambda_i = \lambda_j = .. = \lambda_N = \lambda$, we can rewrite Equation (78):

$$MTBF_P = \frac{N}{\lambda} - \frac{N}{2\lambda} + .. + (-1)^{N+1}\frac{1}{N\lambda} = \frac{1}{\lambda}\sum_{i=1}^{N}(-1)^{i+1}\frac{C_N^i}{i} \quad (3.20)$$

We can easily substitute the sum in Equation 3.20 with the partial sum of the harmonic series:

$$\sum_{i=1}^{N}(-1)^{i+1}\frac{C_N^i}{i} = \sum_{i=1}^{N}\frac{1}{i} \quad (3.21)$$

Therefore, we can simplify Equation 3.20 and write the MTBF of a parallel system with N identical components as:

$$MTBF_P = \frac{1}{\lambda}\sum_{i=1}^{N}\frac{1}{i} \approx \frac{ln(2N)}{\lambda} \quad (3.22)$$

Note that the harmonic series is divergent, so a parallel system does not have a constant failure rate. The failure rate decreases with the increase of the systems that are linked in parallel. We can derive the global failure rate of a system with N identical modules with failure rate $\lambda$ that are connected in parallel using the result in Equation 3.22:

$$\lambda_P = \frac{\lambda}{\sum_{i=1}^{N}\frac{1}{i}} \quad (3.23)$$

■ **Example 3.2** A system consists of two components in parallel, as in Figure 3.7. What is the total reliability, MTBF and the failure rate of the system?

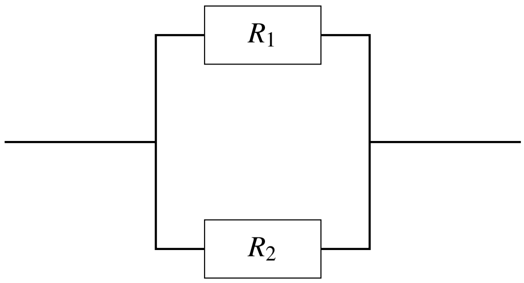



Figure 3.7: Reliability diagrams for a parallel system with two components

We can derive the reliability formula from the general form:

$$R_P(t) = 1 - \prod_{i=1}^{2}(1 - R_i(t)) = 1 - (1 - R_1(t))(1 - R_2(t)) = R_1(t) + R_2(t) - R_1(t)R_2(t) \quad (3.24)$$

Presuming $R_1(t) = e^{-\lambda_1 t}$ and $R_2(t) = e^{-\lambda_2 t}$, the MTBF of the system can be expressed as

$$MTBF_P = \int_0^{\infty} R_P(t)dt = \int_0^{\infty} e^{-\lambda_1 t}dt + \int_0^{\infty} e^{-\lambda_2 t} - \int_0^{\infty} e^{-(\lambda_1+\lambda_2)t}dt \quad (3.25)$$

$$MTBF_P = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2} \quad (3.26)$$

If both modules are identical, meaning $R_1(t) = R_2(t) = R(t)$, then we can simplify the reliability formula to

$$R_P(t) = 2R(t) - R^2(t) \quad (3.27)$$

The MTBF will then be equal to:

$$MTBF_P = \frac{3}{2\lambda} \quad (3.28)$$

and the failure rate of the parallel module will be equal to:

$$\lambda_P = \frac{2}{3}\lambda \tag{3.29}$$

It is worth noting that, as individual reliability functions are $0 \leq R(t) < 1$, $R_P(t)$ will always be greater than $R(t)$, which means that the reliability of the parallel system will always be greater than the reliability of its individual components.

$$R_P(t) = 2R(t) - R^2(t) > R(t), \forall R(t) \in [0,1) \tag{3.30}$$

■

■ **Example 3.3** If the reliability of two individual components is $R_1 = R_2 = 0.9$, then, the total reliability of the parallel system is $R_P = 0.99$. If we increase the number of systems in parallel, as in Figure 3.8, the overall system reliability will also increase:

- 2 systems: $R_P = 1-(1-0.9)^2 = 99\%$
- 3 systems: $R_P = 1-(1-0.9)^3 = 99.9\%$
- 4 systems: $R_P = 1-(1-0.9)^4 = 99.99\%$
- 5 systems: $R_P = 1-(1-0.9)^5 = 99.999\%$
- 6 systems: $R_P = 1-(1-0.9)^6 = 99.9999\%$

■

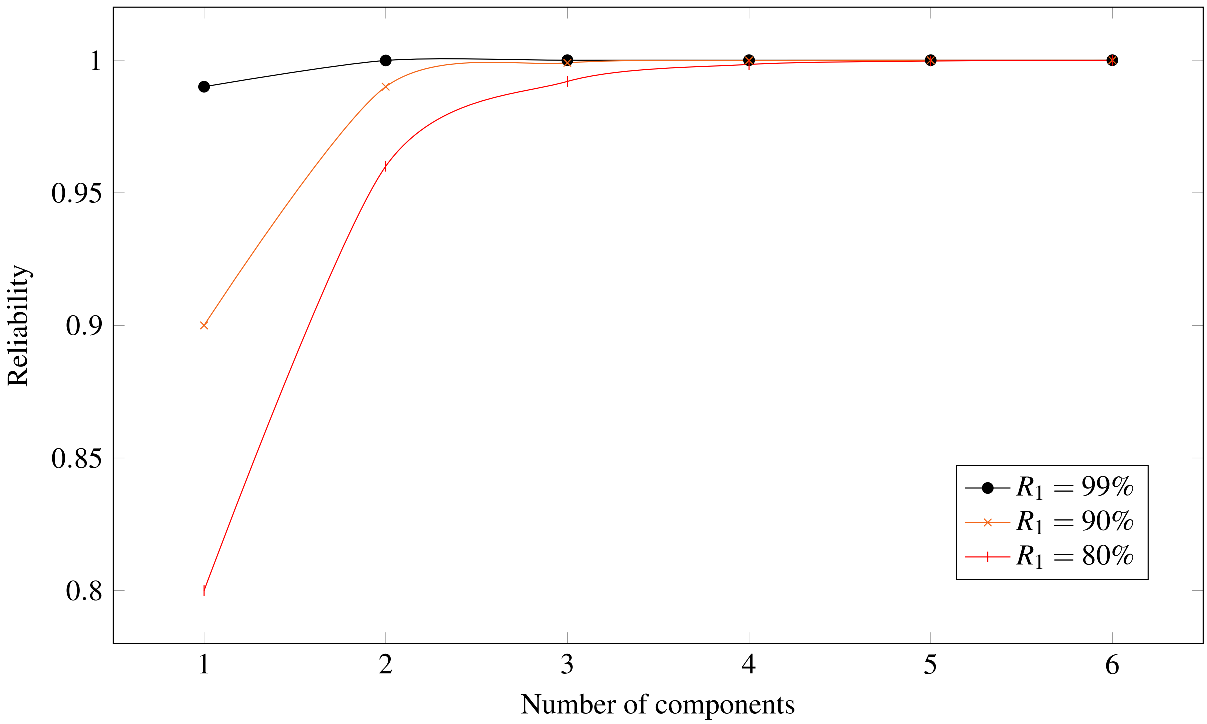



Figure 3.8: Reliability of a parallel system with increasing number of identical components: single component reliability ($R_1$) of 80%, 90% and 99%

## 3.4 Combination of Series and Parallel

While many smaller systems can be accurately represented by either a simple series or parallel configuration, there may be larger systems that involve both series and parallel configurations in the overall system. Such systems can be analysed by calculating the reliabilities for the individual series and parallel sections and then combining them in the appropriate manner. Such a methodology is illustrated in the following example.

■ **Example 3.4** Calculating the Reliability for a Combination of Series and Parallel.

Consider a system with three components. Units 1 and 2 are connected in series and Unit 3 is connected in parallel with the first two, as shown in Figure 3.9. This is meant to illustrate a real-life example of a computing structure in which two processor cores are sharing access to a single RAM memory unit. Each unit has its own reliability function, $R_1$ and $R_2$ for each processor core and $R_3$ for the RAM memory.

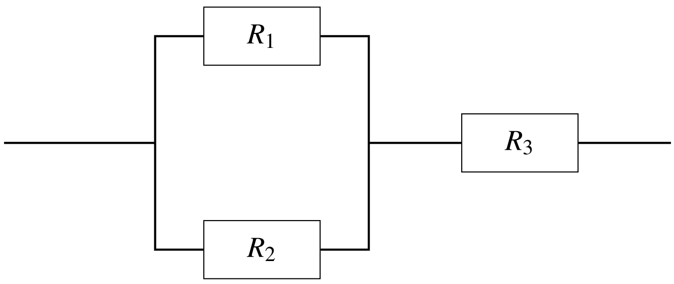



Figure 3.9: Reliability diagram for a simple series-parallel system

What is the reliability of the system if $R_1 = 99.5\%$, $R_2 = 98.7\%$ and $R_3 = 97.3\%$ at 100 hours?

First, the reliability of the parallel segment consisting of Units 1 and 2 is calculated: $R_{12} = 1-(1-R_1)(1-R_2) = 1-(1-0.995)(1-0.987) = 0.999935$

The reliability of the overall system is then calculated by treating Units 1 and 2 as one unit with a reliability of 99.9935% connected in series with Unit 3. Therefore: $R_{123} = R_{12}R_3 = 0.97294$ ■

## 3.5 k Out of n Systems

The k-out-of-n configuration is a special case of parallel redundancy. This type of configuration requires that at least $k$ components succeed out of the total $n$ parallel components for the system to succeed.

■ **Example 3.5** Consider an airplane that has four engines. Furthermore, suppose that the design of the aircraft is such that at least two engines are required to function for the aircraft to remain airborne. This means that the engines are reliability-wise in a k-out-of-n configuration, where k = 2 and n = 4. More specifically, they are in a 2-out-of-4 configuration.

Now, we can derive the overall reliability of such as system if we assume that all four engines have the same reliability function $R(t)$ as a sum of probabilities.

$$R_{2/4}(t) = R^4(t) + 4R^3(t)(1-R(t)) + 6R^2(t)(1-R(t))^2 \tag{3.31}$$

The first term in the probability sum above $R^4(t)$ denotes the probability of all four engines being operational at a certain time. The second term, $R^3(t)(1-R(t))$, denotes the probability of only three engines being operational at a certain time and, as there are four cases in which a single engine could fail, we multiply this by a factor of 4. The last term in the sum gives the probability of a two-engine failure at a certain time. As in the previous case, since there are six instances in which any two engines could fail (equal to $C_4^2$), the total probability of a two-engine failure for our airplane is $6R^2(t)(1-R(t))^2$ ■

Following this example, we can deduce a general formula for k out of n reliability as being:

$$R_{k/n}(t) = C_n^n R^n(t) + C_n^{n-1} R^{n-1}(t)(1-R(t)) + .. + C_n^k R^k(t)(1-R(t))^{n-k}, \forall k < n \tag{3.32}$$

Or, we can express the above sum as:

$$R_{k/n}(t) = \sum_{i=k}^{n} C_n^i R^i(t)(1-R(t))^{n-i}, \forall k < n \tag{3.33}$$

Even though we classified the k-out-of-n configuration as a special case of parallel redundancy, it can also be viewed as a general configuration type. As the number of units required to keep the system functioning approaches the total number of units in the system, the system's behavior tends towards that of a series system. If the number of units required is equal to the number of units in the system, it is a series system. In other words, a series system of statistically independent components is an n-out-of-n system and a parallel system of statistically independent components is a 1-out-of-n system.

This can be easily deduced from the previous equation. if we plug in k = n in the general k-out-of-n reliability formula we get the reliability of a series system:

$$R_{n/n}(t) = \sum_{i=n}^{n} C_n^i R^i(t)(1-R(t))^{n-i} = R^n(t) \tag{3.34}$$

If we plug in k = 1 in the same formula, we get the standard reliability of a parallel system with n identical units:

$$R_{1/n}(t) = \sum_{i=1}^{n} C_n^i R^i(t)(1-R(t))^{n-i} = 1-(1-R(t))^n \tag{3.35}$$

If $R(t) = e^{-\lambda t}$, then k-out-of-n reliability can be written as:

$$R_{k/n}(t) = \sum_{i=k}^{n} C_n^i e^{-i\lambda t}(1-e^{-\lambda t})^{n-i} \tag{3.36}$$

We can also write the MTBF for the k-out-of-n structure:

$$MTBF_{k/n}(t) = \int_0^{\infty} R_{k/n}(t)dt = \frac{1}{\lambda}\sum_{i=k}^{n}\frac{1}{i} \tag{3.37}$$

■ **Example 3.6** For our airplane model that tolerates a two-engine failure, if we assume the fault distribution to be exponential with a constant failure rate $\lambda = 0.00001/hour$ for each engine, we can calculate the MTBF of the airplane as:

$$MTBF_{2/4}(t) = \frac{1}{\lambda}\sum_{i=2}^{4}\frac{1}{i} = \frac{1}{\lambda}\left(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}\right) = \frac{13}{12}\frac{1}{\lambda} \approx 108333 hours \tag{3.38}$$

■

## 3.6 Series-Parallel and Parallel-Series Systems

In this section we consider systems which use multiple identical units of a given reliability that can be linked in groups of series and parallel. We present two ways in which we can link these identical units: series-parallel and parallel-series configurations. Assuming each unit has a reliability function $R(t)$ and using series and parallel reliability formulas, we can quickly derive the overall reliability of the structures presented in Figures 3.10 and 3.11.

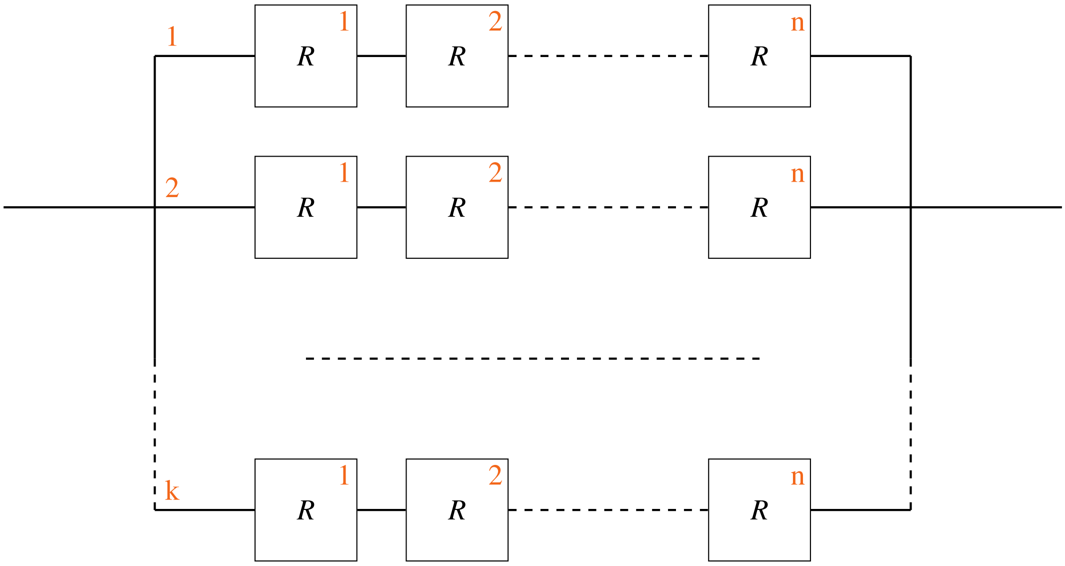



Figure 3.10: Reliability diagram for a $k \times n$ series-parallel system with identical units

For the series-parallel structure, we have n identical units in series on each of the k lines. As such, we can derive the following formula for the series-parallel reliability function $R_{SP}(t)$:

$$R_{SP}(t) = 1-(1-R^n(t))^k \tag{3.39}$$

In a similar fashion, for the parallel-series structure in Figure 3.11, we have a group of k identical units in parallel that are connected in series with another group of k-parallel units and so on repeating n times.

We can derive the following formula for the parallel-series reliability function $R_{PS}(t)$:

$$R_{PS}(t) = \left[1-(1-R(t))^k\right]^n \tag{3.40}$$

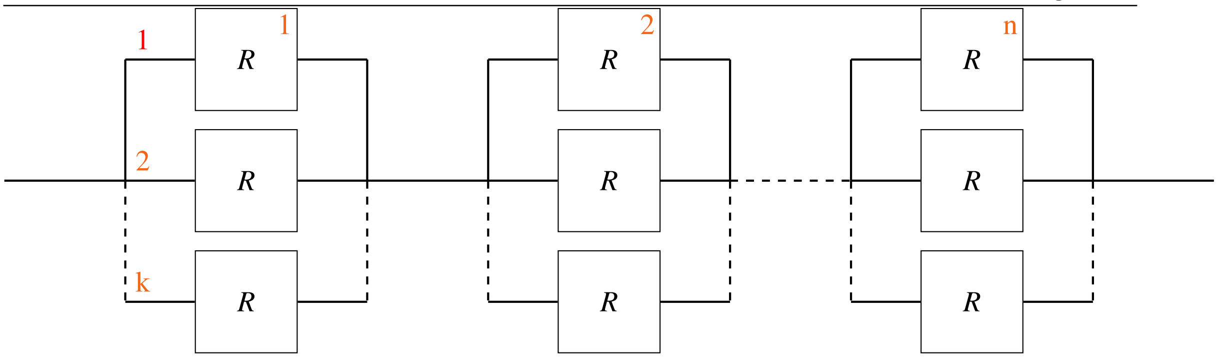



Figure 3.11: Reliability diagram for a $k \times n$ parallel-series system with identical units

Having deduced these formulas, we can plot the two reliability functions as in Figure 3.12. It can be mathematically proven that the reliability of the parallel-series configuration $R_{PS}(t)$ is always greater than the reliability of the series-parallel configuration $R_{SP}(t)$ for any positive integer values of n and k.

What this means is that redundancy at the component level is always more effective than redundancy at the system level in improving system reliability, when using the same number of components.

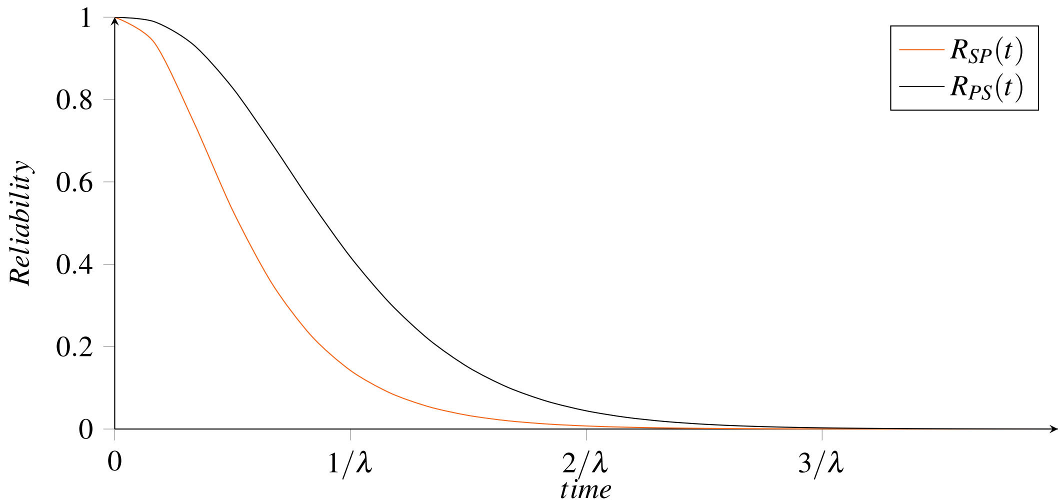



Figure 3.12: $R_{SP}(t)$ versus $R_{PS}(t)$ for n=k=3

■ **Example 3.7** A solar panel is comprised of nine identical solar cells. Each solar cell can have two major failure modes:

- **fail-open**, when the cell or its metallic contacts break due to micro-fissures or bad bonding
- **fail-closed**, when the cell is shorted

In either of these two states the cell's energy production is compromised, but, depending on how the cell is connected to the other cells in the solar panel and the preferred failure mode of the cell, it could have a larger or smaller effect on the overall reliability of the panel.

Given a single solar cell reliability of R=0.9 at a certain time and a fail-open preferred mode of failure, what is the best way to wire the nine cells in the solar panel: series-parallel (Figure 3.13) or parallel-series (Figure 3.14)?

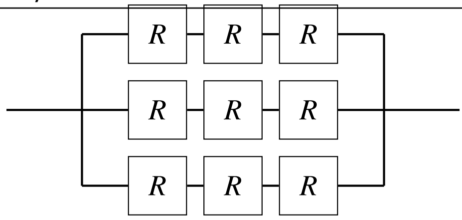



Figure 3.13: Reliability diagram for a $3 \times 3$ series-parallel system with identical units

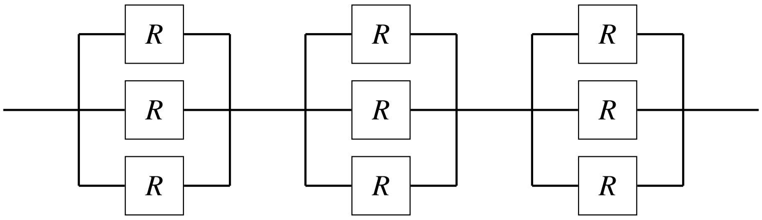



Figure 3.14: Reliability diagram for a $3 \times 3$ parallel-series system with identical units

We can compute the $3 \times 3$ series-parallel and parallel-series reliability functions as:

- $R_{SP} = 1-(1-0.9^3)^3 = 0.98$
- $R_{PS} = (1-(1-0.9)^3)^3 = 0.997.$

It is evident that $R_{SP} < R_{PS}$.

In conclusion, if the preferred failure mode of a solar cell is fail-open, it is more reliable to wire a solar panel as a parallel-series configuration. ■

## 3.7 Non-Decomposable Systems

There are systems that can be described by reliability block diagrams that cannot be decomposed into series or parallel units. An example of such a system is described in Figure 3.15, where the structure does not lead to any type of equivalence to a series or parallel module configuration.

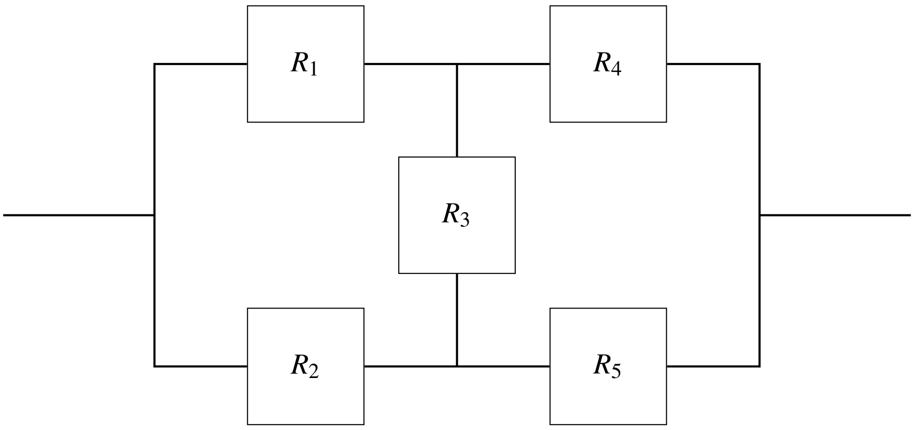



Figure 3.15: Reliability diagram for a non-decomposable system

We can assess the reliability of this type of system by considering the behaviour of one of its constituent modules. Let us pick the module with reliability $R_3$ and estimate the reliability of the system in the following two situations:

**Case 1: module $R_3$ has failed.**
In this case, module $R_3$ can be represented as open in the system reliability diagram (reliability = 0), as in Figure 3.16.

We can estimate the reliability of the system in this case as:

$$R_{C1} = (R_1R_4)||(R_2R_5) = 1-(1-R_1R_4)(1-R_2R_5) = R_2R_5+R_1R_4-R_1R_4R_2R_5 \tag{3.41}$$

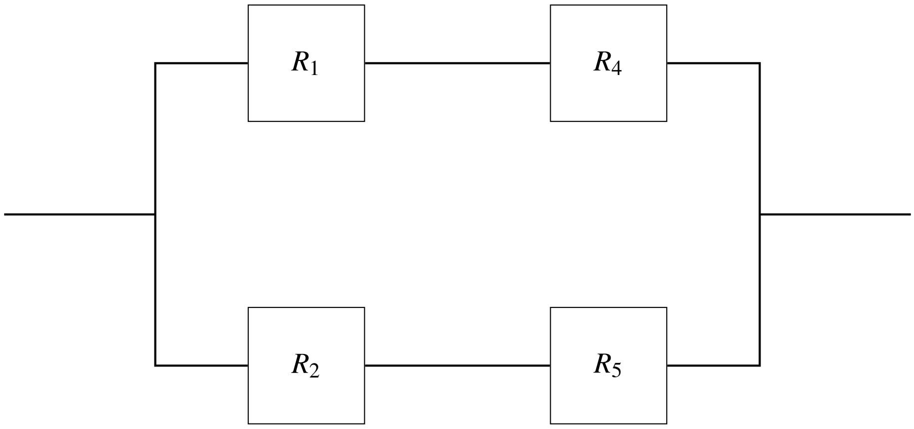



Figure 3.16: Equivalent system reliability diagram for module $R_3$ failed

**Case 2: module $R_3$ is fully operational.**
In this case, module $R_3$ can be represented as connection in the system reliability diagram (reliability = 1), as in Figure 3.17.

We can also estimate the reliability of the system in this case as:

$$\begin{aligned} R_{C2} &= (R_1||R_2)(R_4||R_5) = (R_1+R_2-R_1R_2)(R_4+R_5-R_4R_5) = \\ &= R_1R_4+R_1R_5-R_1R_4R_5+R_2R_4+R_2R_5-R_2R_4R_5-R_1R_2R_4-R_1R_2R_5+R_1R_2R_4R_5 \end{aligned} \tag{3.42}$$

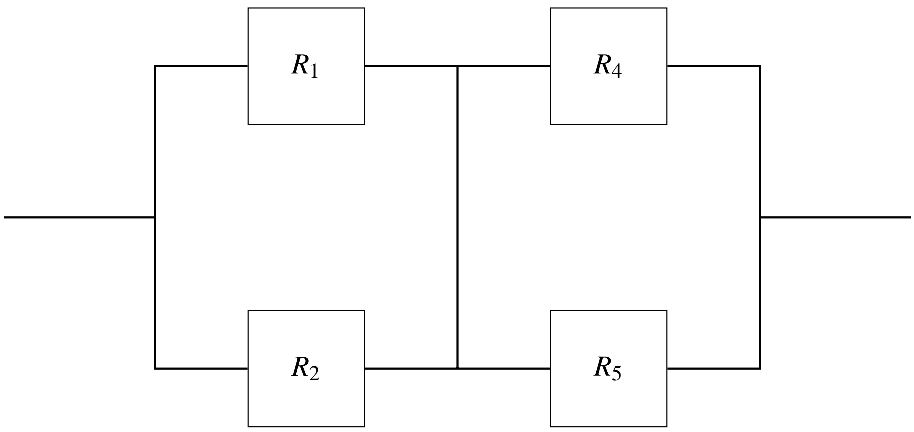



Figure 3.17: Equivalent system reliability diagram for module $R_3$ fully operational

The reliability of the system in Figure 3.15 is going to be a sum of probabilities of the two cases:

$$R_S = R_3 \times P(system\ works|3\ is\ without\ faults) + (1-R_3) \times P(system\ works|3\ is\ faulty) \tag{3.43}$$

$$\begin{aligned}
&= R_3 R_{C2} + (1-R_3) R_{C1} \\
&= R_3(R_1R_4 + R_1R_5 - R_1R_4R_5 + R_2R_4 + R_2R_5 - R_2R_4R_5 - R_1R_2R_4 - R_1R_2R_5 + R_1R_2R_4R_5) \\
&\quad + (1-R_3)(R_2R_5 + R_1R_4 - R_1R_4R_2R_5) \\
&= R_1R_4 + R_2R_5 + R_1R_3R_5 + R_2R_3R_4 - R_1R_3R_4R_5 - R_2R_3R_4R_5 - R_1R_2R_3R_4 - R_1R_2R_3R_5 \\
&\quad - R_1R_2R_4R_5 + 2R_1R_2R_3R_4R_5
\end{aligned}$$

If $R_1 = R_2 = R_3 = R_4 = R_5 = R$, then we can rewrite 3.43 as:

$$R_S = 2R^2 + 2R^3 - 5R^4 + 2R^5 \tag{3.44}$$

## 3.8 Majority Voted Redundancy

Majority voted redundancy is a fault-tolerant technique used in computing systems to enhance reliability and protect against component failures. It employs a redundant approach, utilizing multiple identical copies of a hardware component or software module to execute the same task. The outputs of these redundant components are then fed into a voting mechanism, which determines the correct output based on the majority vote.

### 3.8.1 Triple Modular Redundancy (TMR)

The simplest structure to offer majority voted redundancy is the triple modular scheme in which three identical components, be it hardware or software, are set to simultaneously execute the same function and the results at their outputs are compared by a voter, as in Figure 3.18.

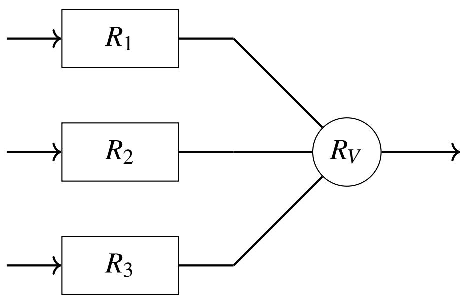



Figure 3.18: Triple modular redundancy majority voting system

The majority voting mechanism determines the correct output based on the majority vote. If two or more components produce the same output, that is considered the correct output. If there is a tie, the system may enter a fail-safe state or attempt to recompute the output.

TMR provides exceptional fault tolerance, as it can withstand one or, in some instances, two component failures while still maintaining system operation. The voting mechanism effectively detects discrepancies among the outputs of the redundant components, enabling error detection and correction.

TMR is widely used in mission-critical systems where reliability is paramount, such as aircraft avionics, medical devices, and industrial control systems. Also, TMR is often employed in data storage systems to protect against data loss due to hardware failures.

The main limitation of TMR comes from its scale, as it introduces additional hardware or software overhead, leading to higher system costs.

The total reliability of a TMR system can be inferred from the fact that it needs at least two of its three modules and its voter to function fault-free in order for the whole system to function properly. As such, we can write:

$$R_{2/3} = [R_1R_2(1-R_3)+R_1(1-R_2)R_3+(1-R_1)R_2R_3+R_1R_2R_3]R_V \tag{3.45}$$

Given the three modules are identical, we can assume that $R_1(t) = R_2(t) = R_3(t) = R(t)$. Also, the voter is far simpler in structure than of any other of the three modules, so we can assume that it can be much more reliable $R_V(t) >> R(t)$, $R_V(t) \approx 1$

Therefore, we can rewrite 3.45 as:

$$R_{2/3}(t) = 3R^2(t) - 2R^3(t) \tag{3.46}$$

Given that $R(t) = e^{-\lambda t}$, we can plug in to 3.46:

$$R_{2/3}(t) = 3e^{-2\lambda t} - 2e^{-3\lambda t} \tag{3.47}$$

One interesting question is whether this TMR structure is more reliable than a single module. We can answer this question by solving this simple inequality:

$$R_{2/3}(t) > R(t) \Rightarrow 3R^2(t)-2R^3(t) > R(t) \Rightarrow 2R^3(t)-3R^2(t)+R(t) < 0 \Rightarrow R(t)(2R^2(t)-3R(t)+1) < 0 \tag{3.48}$$

As $R(t) \geq 0 \;\; \forall t \geq 0$ as a probability function, then:

$$2R^2(t) - 3R(t) + 1 < 0 \Rightarrow \left(R(t) - \frac{1}{2}\right)(R(t)-1) < 0 \tag{3.49}$$

which is true for $R(t) \geq \frac{1}{2}$.

Therefore, it is advisable to use triple modular redundancy only with modules that operate at an individual reliability greater than 50%, as in Figure 3.19. We can calculate the elapsed mission time at which the triple modular redundancy falls below the reliability of a single module $t = ln(2)/\lambda \approx 0.69/\lambda$. In conclusion, TMR can be used for applications that have mission times less than 69% of MTBF.

We can also calculate the MTBF of a TMR structure as:

$$MTBF_{2/3} = \int_0^{\infty} \left(3e^{-2\lambda t} - 2e^{-3\lambda t}\right) dt = \frac{3}{2\lambda} - \frac{2}{3\lambda} = \frac{5}{6\lambda} = \frac{5}{6} MTBF \tag{3.50}$$

We can see that $MTBF_{2/3} < MTBF$ at any time, which says the TMR structure will fail on average more often than one of its constituent modules.

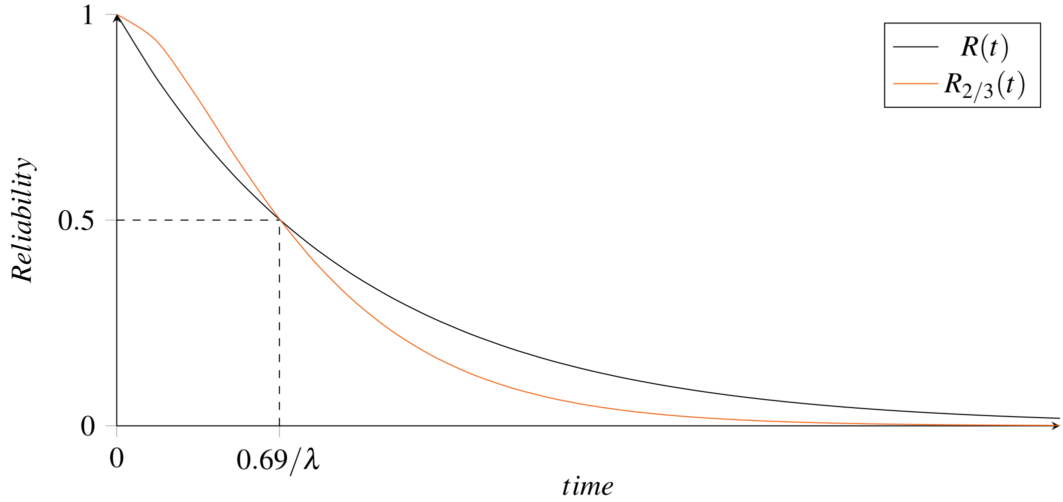



Figure 3.19: $R_{2/3}(t)$ versus $R(t)$ for a TMR system

### 3.8.2 3-out-of-5 Modular Redundancy

We can further replicate the modules in the TMR scheme and build a two out of five majority voting scheme, as in Figure 3.20.

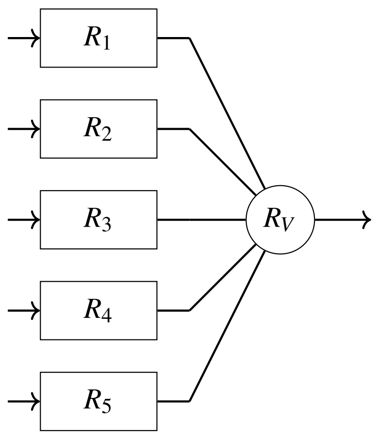



Figure 3.20: 3/5 majority voting system

Operating under the same assumptions that that $R_1(t) = R_2(t) = R_3(t) = R_4(t) = R_5(t) = R(t)$ and $R_V(t) >> R(t)$, $R_V(t) \approx 1$, we can deduce the reliability:

$$R_{3/5}(t) = R^5(t) + 5(1 - R(t))R^4(t) + 10(1 - R(t))^2 R^3(t) = 6R^5(t) - 15R^4(t) + 10R^3(t) \quad (3.51)$$

Given that $R(t) = e^{-\lambda t}$, we can plug in to 3.51:

$$R_{3/5}(t) = 6e^{-5\lambda t} - 15e^{-4\lambda t} + 10e^{-3\lambda t} \quad (3.52)$$

which is also greater than $R(t)$ $if$ $R(t) \in (\frac{1}{2}, 1]$

The MTBF of the structure is:

$$MTBF_{3/5} = \int_0^\infty \left(6e^{-5\lambda t} - 15e^{-4\lambda t} + 10e^{-3\lambda t}\right) dt = \frac{6}{5\lambda} - \frac{15}{4\lambda} + \frac{10}{3\lambda} = \frac{47}{60\lambda} = \frac{47}{60} MTBF \quad (3.53)$$

Therefore, $MTBF_{3/5} < MTBF_{2/3} < MTBF$.

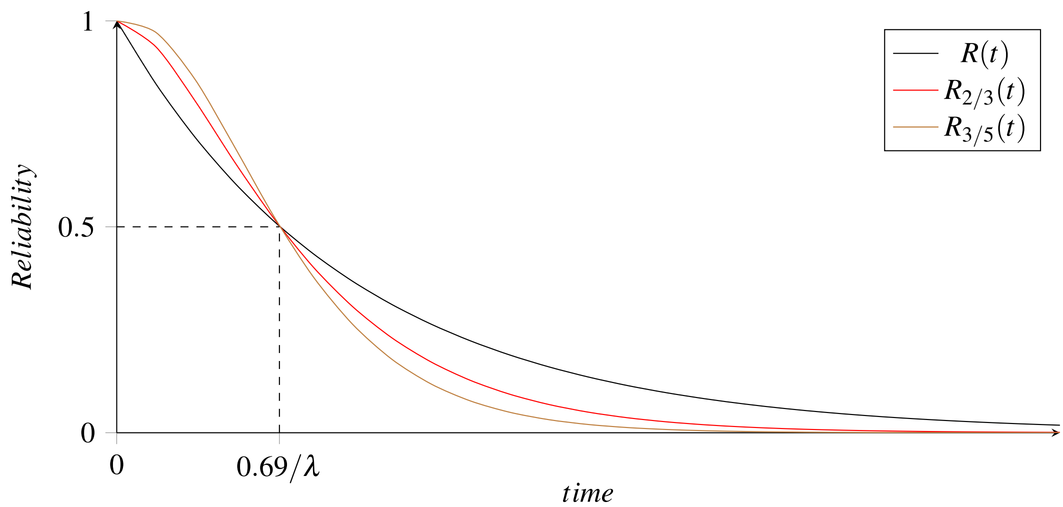



Figure 3.21: $R_{3/5}(t)$ and $R_{2/3}(t)$ versus $R(t)$

### 3.8.3 n-out-of-(2n-1) Modular Redundancy

Given the previous two examples of majority voting, we can expand to a general case of n out of 2n-1 modular redundancy with voting. This is similar to a k-out-of-n structure, with the particularity that n must be an odd number, to allow majority voting.

The general formula for majority voting structures reliability becomes:

$$R_{n/2n-1}(t) = C_{2n-1}^{2n-1} R^{2n-1}(t) + C_{2n-1}^{2n-2}(1-R(t))R^{2n-2}(t) + .. + C_{2n-1}^{n}(1-R(t))^{n-1}R^{n}(t)$$
$$= \sum_{i=n}^{2n-1} C_{2n-1}^{i}(1-R(t))^{2n-1-i}R^{i}(t) \quad (3.54)$$

If $R(t) = e^{-\lambda t}$, we can rewrite 3.54:

$$R_{n/2n-1}(t) = \sum_{i=n}^{2n-1} C_{2n-1}^{i}(1-e^{-\lambda t})^{2n-1-i}e^{-i\lambda t} \quad (3.55)$$

It can be mathematically proven that $R_{n/2n-1}(t) > R(t)$ when $R(t) > 0.5$, so any majority voting system can be used for mission times shorter than 0.69 of a single unit MTBF.

The MTBF for the entire structure is generally decreasing with the increase in number of redundant modules. While this is counter-intuitive, it can be explained by the increasing probability of a number of redundant modules malfunctioning at any given time.

$$MTBF_{n/2n-1} = \int_0^\infty R_{n/2n-1}(t)dt = \frac{1}{\lambda}\sum_{i=n}^{2n-1}\frac{1}{i} \tag{3.56}$$

It can be proven that, when n increases, $MTBF_{n/2n-1}$ decreases asymptotically to:

$$\lim_{n\to\infty} MTBF_{n/2n-1} = \frac{1}{\lambda}\lim_{n\to\infty}\sum_{i=n}^{2n-1}\frac{1}{i} = \frac{ln(2)}{\lambda} \approx 0.69MTBF \tag{3.57}$$

## 3.9 Standby-Sparing

Standby-sparing is a fault-tolerant technique used to improve the reliability of systems. It involves having one or more spare components that are inactive until the primary component fails. When the primary component fails, the spare component is activated and takes over its operation.

There are three types of standby-sparing commonly used in computing:

- Cold Sparing: The spare component is not powered on until the primary component fails. This is the simplest form of standby sparing but also the least efficient.
- Warm Sparing: The spare component is powered on but not actively participating in the system operation. This is more efficient than cold sparing but requires additional resource consumption.
- Hot Sparing: The spare component is fully active and ready to take over operation immediately upon primary component failure. This is the most efficient form of standby sparing but also the most complex and expensive.

There are some immediate benefits of standby sparing over the other types of techniques to improve reliability. It is a structure that offers high reliability, as standby sparing can significantly improve the reliability of systems by providing multiple paths for system operation. Another benefit is reduced downtime. When a primary component fails, the standby component can take over immediately, minimizing downtime and improving system availability.

### 3.9.1 One Spare Reliability

Figure 3.22 describes a one-spare system. There is a primary module with reliability $R_1$ and a spare $R_2$. The spare is coupled into operation only when the failure detection module $FD$ detects a failure of the primary module. Note that the diagram does not implicitly differentiate between cold or warm sparing.

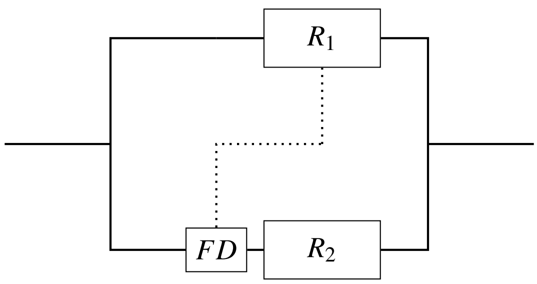



Figure 3.22: One-spare reliability diagram

#### Cold Sparing

We will analyze the system with the assumption that the spare module is not powered until the primary has completely failed.

As previously stated, we note the reliability of the primary module $R_1$ and the reliability of the spare $R_2$. The failure detection module is much simpler in structure than either the primary or the spare, therefore its reliability can be approximated to be $R_{FD} = 1$.

There can be two cases in which this structure successfully completes its mission:

1. The primary module survives for the entire mission duration
2. The primary module shuts down due to a defect at a certain time and the spare switches on for the remainder of the mission duration

If we write $P_1$ to be the probability of case one happening and $P_2$ the probability attached to the second case, then, the reliability of the one-spare structure can be written as $R_{1sp} = P_1 + P_2$

$P_1$ is equal to the reliability of the primary, so $P_1 = R_1$.

To calculate $P_2$ we need to take into account that there are two events that need to happen: the primary breaking down at a certain time and the spare switching on at this time and continuing the mission. Let's write this moment in time as $\tau$.

The probability of the primary breaking down at time $\tau$ is equal to the pdf $f(\tau)$, which, by the relations established in a previous chapter is $f(\tau) = \frac{\mathrm{d}F(\tau)}{\mathrm{d}\tau} = -\frac{\mathrm{d}R_1(\tau)}{\mathrm{d}\tau}$. The probability of the spare switching on at time $\tau$ and then continuing to operate until the mission is completed at a certain time $t$ is $R_2(t-\tau)$. However, the malfunction of the primary can happen anytime between 0 and time $t$, so $P_2 = \int_0^t -\frac{\mathrm{d}R_1(\tau)}{\mathrm{d}\tau} R_2(t-\tau) d\tau$

Therefore, we can write the reliability of a one spare system with cold sparing as:

$$R_{1sp}(t) = P_1 + P_2 = R_1(t) + \int_0^t -\frac{\mathrm{d}R_1(\tau)}{\mathrm{d}\tau} R_2(t-\tau) d\tau \tag{3.58}$$

If $R_1(t) = e^{-\lambda_1 t}$ and $R_2(t) = e^{-\lambda_2 t}$, then we can rewrite 3.58 as:

$$\begin{aligned} R_{1sp}(t) &= e^{-\lambda_1 t} + \int_0^t \lambda_1 e^{-\lambda_1 \tau} e^{-\lambda_2 (t-\tau)} d\tau = e^{-\lambda_1 t} + \lambda_1 e^{-\lambda_2 t} \int_0^t e^{-(\lambda_1 - \lambda_2)\tau} d\tau \\ &= e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_2 t} \left( e^{-(\lambda_1 - \lambda_2)t} - 1 \right) = \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} - \frac{\lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_2 t} \end{aligned} \tag{3.59}$$

We can also calculate the MTBF of this structure as:

$$\begin{aligned} MTBF_{1sp} &= \int_0^\infty R_{1sp}(t) dt = \int_0^\infty \left( \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} - \frac{\lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_2 t} \right) dt \\ &= \frac{\lambda_2}{\lambda_2 - \lambda_1} \int_0^\infty e^{-\lambda_1 t} dt - \frac{\lambda_1}{\lambda_2 - \lambda_1} \int_0^\infty e^{-\lambda_2 t} dt = \frac{\lambda_2}{\lambda_2 - \lambda_1} \frac{1}{\lambda_1} - \frac{\lambda_1}{\lambda_2 - \lambda_1} \frac{1}{\lambda_2} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \end{aligned} \tag{3.60}$$

Therefore, we can conclude that $MTBF_{1sp} = MTBF_1 + MTBF_2$, which means that, on average, a one-spare system will have a longer lifetime than any one of its two components.

Usually, in one-spare systems, the primary and the spare are similar, if not identical modules (e.g. a compute core taking over when an identical compute core breaks down, or a memory drive that backs up a primary drive with the same capacity or specifications). Therefore, we can assume that $R_1(t) = R_2(t) = e^{\lambda t}$.

We can recalculate 3.59 to take into account we are working with identical modules:

$$R_{1sp}(t) = e^{-\lambda t} + \int_0^t \lambda e^{-\lambda \tau} e^{-\lambda(t-\tau)} d\tau = e^{-\lambda t} + \lambda e^{-\lambda t} \int_0^t d\tau = e^{-\lambda t}(1 + \lambda t) \quad (3.61)$$

Similarly, we can infer the MTBF:

$$MTBF_{1sp} = \frac{1}{\lambda} + \frac{1}{\lambda} = 2MTBF \quad (3.62)$$

This shows that the MTBF of a one-spare structure with cold sparing and identical units is twice the MTBF of a single unit.

### Warm Sparing

Now, let us assume that the spare is not completely switched off while the primary is operating. Therefore, it will also be affected by degradation, but at a much lover rate than normal. We will note this as a different reliability function for the spare, $R_{2n}$.

Similar assumptions from the cold sparing case can be applied. The system is considered to be working if:

- The primary operates without interruption for the entire duration of the mission ($P_1 = R_1(t)$).
- The primary breaks down at a certain time $\tau$ and the spare takes over, switching from a standby state into full operation ($P_2$).

The only major difference from cold sparing is for case 2, taking into account that the spare is in a standby state while the primary is fully operational. We can quantify this new reliability as $P_2 = \int_0^t -\frac{\mathrm{d}R_1(\tau)}{\mathrm{d}\tau} R_{2n}(\tau) R_2(t-\tau) d\tau$

Therefore, the reliability of a one-spare system with warm sparing can be written as:

$$R_{1sp'}(t) = P_1 + P_2 = R_1(t) + \int_0^t -\frac{\mathrm{d}R_1(\tau)}{\mathrm{d}\tau} R_{2n}(\tau) R_2(t-\tau) d\tau \quad (3.63)$$

If $R_1(t) = e^{-\lambda_1 t}$, $R_2(t) = e^{-\lambda_2 t}$ and $R_{2n}(t) = e^{-\lambda_{2n} t}$, then we can rewrite 3.63 as:

$$\begin{aligned} R_{1sp'}(t) &= e^{-\lambda_1 t} + \int_0^t \lambda_1 e^{-\lambda_1 \tau} e^{-\lambda_{2n} \tau} e^{-\lambda_2 (t-\tau)} d\tau = e^{-\lambda_1 t} + \lambda_1 e^{-\lambda_2 t} \int_0^t e^{-(\lambda_1 + \lambda_{2n} - \lambda_2)\tau} d\tau \quad (3.64) \\ &= e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_1 + \lambda_{2n} - \lambda_2} \left( e^{-\lambda_2 t} - e^{-(\lambda_1 + \lambda_{2n}) t} \right) \end{aligned}$$

We can also calculate the MTBF of this structure as:

$$MTBF_{1sp'} = \int_0^\infty R_{1sp'}(t)dt = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}\frac{\lambda_1}{\lambda_1 + \lambda_{2n}} = MTBF_1 + MTBF_2\frac{MTBF_{2n}}{MTBF_1 + MTBF_{2n}} \quad (3.65)$$

If the primary and the spare are identical we can assume $R_1(t) = R_2(t) = e^{-\lambda t}$, but the standby reliability of the spare will still need to be factored in. We can write this as $R_{2n} = e^{-\lambda_n t}$

We can rewrite 3.64 as:

$$R_{1sp'}(t) = e^{-\lambda t} + \int_0^t \lambda e^{-\lambda\tau}e^{-\lambda_n\tau}e^{-\lambda(t-\tau)}d\tau = e^{-\lambda t}\left[1 + \frac{\lambda}{\lambda_n}\left(1 - e^{-\lambda_n t}\right)\right] \quad (3.66)$$

MTBF in 3.65 can also be simplified as:

$$MTBF_{1sp'} = \frac{1}{\lambda} + \frac{1}{\lambda + \lambda_n} \quad (3.67)$$

From 3.67 we can infer that the MTBF of the warm spare system is lower than one for the cold spare system, due to the lower MTBF of the spare.

### 3.9.2 Two Spare Reliability

Next, we will focus on the reliability of a two-spare system. There is a primary module of reliability $R_1 = e^{-\lambda_1 t}$ and two spares with reliabilities $R_2 = e^{-\lambda_2 t}$ and $R_3 = e^{-\lambda_3 t}$. We will focus on the cold sparing case, in which both spares are completely switched off when are not used.

We can estimate the reliability of the structure through two iterations: first considering the reliability of the $[R_1, R_2]$ ensemble as a one-spare system $R_{12}$, and then adding $R_3$ as a spare to it.

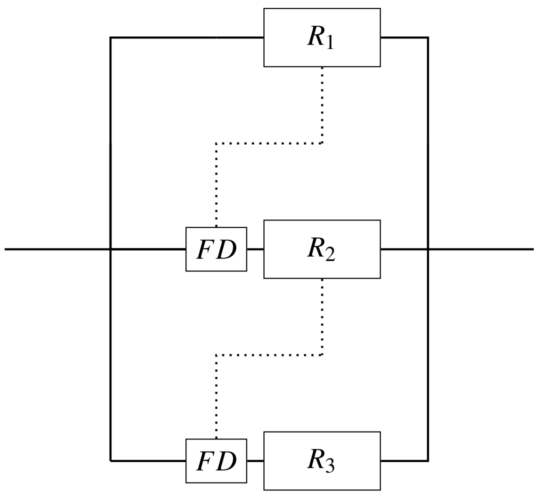



Figure 3.23: Two-spare reliability diagram

We have previously demonstrated in equation 3.59 that:

$$R_{12}(t) = \frac{\lambda_2}{\lambda_2 - \lambda_1}e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_1 - \lambda_2}e^{-\lambda_2 t} \quad (3.68)$$

Factoring the above into the expression for the total system reliability yields the following expression:

$$R_{2sp} = R_{123}(t) = R_{12}(t) + \int_0^t -\frac{\mathrm{d}R_{12}(\tau)}{\mathrm{d}\tau} R_3(t-\tau)d\tau \tag{3.69}$$

After plugging in equation 3.68 in the above expression, we can write:

$$R_{2sp}(t) = \frac{\lambda_2\lambda_3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}e^{-\lambda_1 t} + \frac{\lambda_1\lambda_3}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)}e^{-\lambda_2 t} + \frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}e^{-\lambda_3 t} \tag{3.70}$$

From here it can be easily deduced that MTBF of the two spare structure is:

$$\begin{aligned} MTBF_{2sp} &= \int_0^\infty R_{2sp}(t)dt \\ &= \int_0^\infty \left( \frac{\lambda_2\lambda_3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}e^{-\lambda_1 t} + \frac{\lambda_1\lambda_3}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)}e^{-\lambda_2 t} + \frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}e^{-\lambda_3 t} \right) dt \\ &= \frac{\lambda_2\lambda_3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)}\frac{1}{\lambda_1} + \frac{\lambda_1\lambda_3}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)}\frac{1}{\lambda_2} + \frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}\frac{1}{\lambda_3} \\ &= \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} \end{aligned} \tag{3.71}$$

We can rewrite equation 3.71 as:

$$MTBF_{2sp} = MTBF_1 + MTBF_2 + MTBF_3 \tag{3.72}$$

If all of the three modules are identical, $R_1(t) = R_2(t) = R_3(t) = e^{-\lambda t}$ we can rewrite equation 3.69 as:

$$R_{2sp} = e^{-\lambda t}(1+\lambda t) + \int_0^t -\frac{\mathrm{d}(e^{-\lambda\tau}(1+\lambda\tau))}{\mathrm{d}\tau}e^{-\lambda\tau}d\tau = e^{-\lambda t}\left(1+\lambda t+\frac{\lambda^2 t^2}{2}\right) \tag{3.73}$$

Also,

$$MTBF_{2sp} = \int_0^\infty e^{-\lambda t}\left(1+\lambda t+\frac{\lambda^2 t^2}{2}\right)dt = \frac{3}{\lambda} = 3MTBF \tag{3.74}$$

### 3.9.3 N Spare Reliability

We can generalize the one spare and two spare examples to a system which has any number of spares.

For ease of calculation, we can make the following assumptions:

- Each spare is identical with the primary unit

- The entire standby-sparing system is operating with cold spares

The reliability of a system with $n$ such spares can be deduced iteratively and be written as:

$$R_{nsp} = e^{-\lambda t}\left(1+\lambda t+\frac{\lambda^2 t^2}{2}+\frac{\lambda^3 t^3}{6}+..+\frac{\lambda^n t^n}{n!}\right) = e^{-\lambda t}\sum_0^n \frac{\lambda^n t^n}{n!} \tag{3.75}$$

Similarly, the MTBF of an n-spare system is:

$$MTBF_{nsp} = \int_0^\infty e^{-\lambda t}\sum_0^n \frac{\lambda^n t^n}{n!}dt = n\frac{1}{\lambda} = n\cdot MTBF \tag{3.76}$$

#### Infinite Spares

An interesting, albeit purely theoretical case is when the system has an infinite amount of spares, because the sum $\sum_0^\infty \frac{\lambda^n t^n}{n!}$ is the Taylor series of $e^{\lambda t}$.

Replacing this into 3.75 for $n = \infty$ we get:

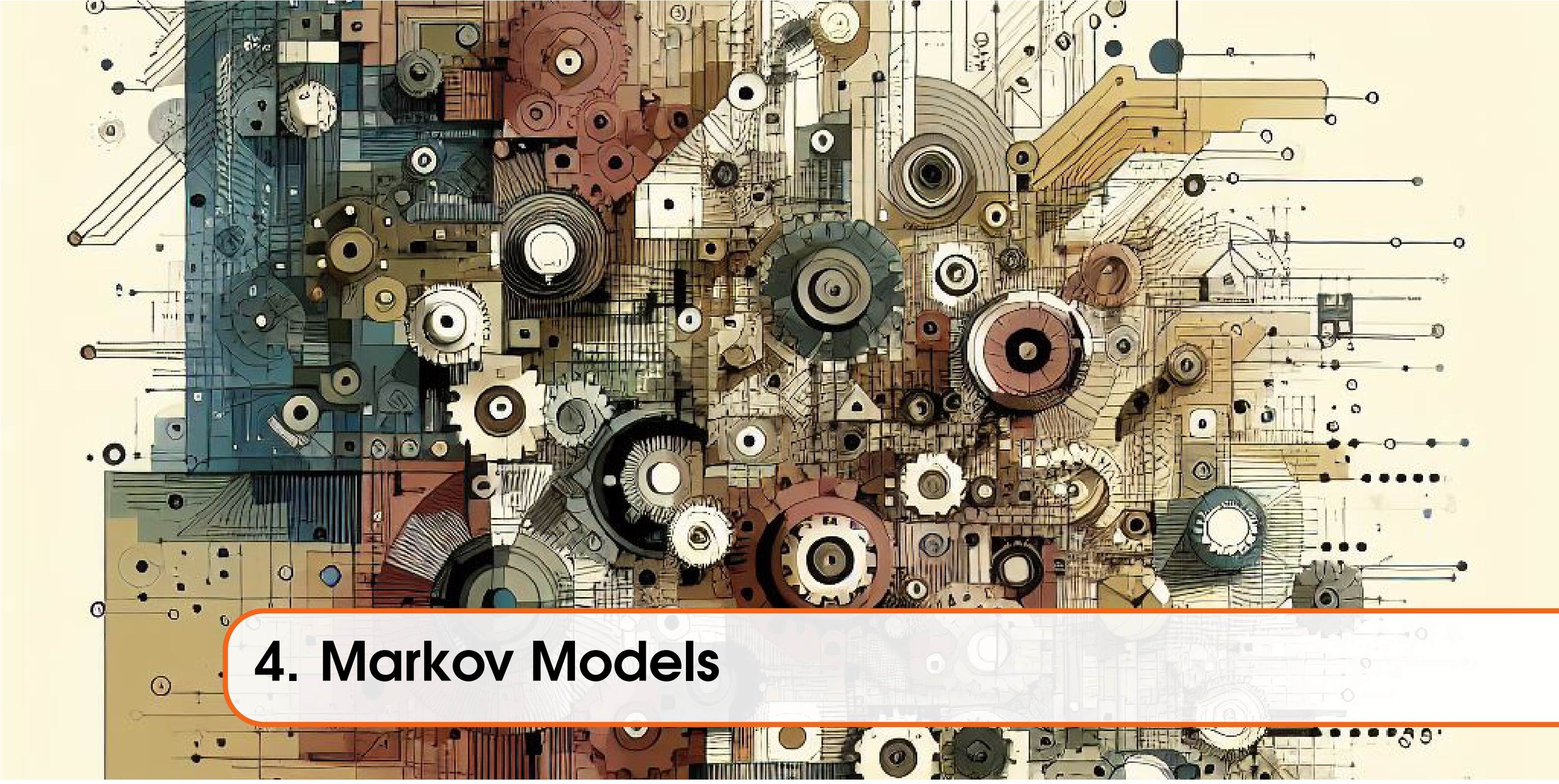
$$R_{\infty sp} = e^{-\lambda t}\sum_0^\infty \frac{\lambda^n t^n}{n!} = e^{-\lambda t}e^{\lambda t} = 1 \tag{3.77}$$

This is the only example in fault tolerance where a system can achieve 100% reliability. Unfortunately, it is also completely unachievable.

# Part Three


**4 Markov Models . . . . . . . . . . . . . . . . . . . . . . . 65**

4.1 Markov Chains
4.2 Markov Processes
4.3 Single-Component System without Repair
4.4 Single-Component with Repair
4.5 Two-Component System Without Repair
4.6 Two-Component System with Repair
4.7 Two-Component System with a Dependency
4.8 System with one Cold Standby Backup
4.9 System with a Warm Standby Backup
4.10 Cold Standby Backup System with Repair
4.11 Cold Standby System with Two Backups
4.12 2-out-of-3 Majority Voting System
4.13 2-out-of-3 Majority Voting System with Repair
4.14 Three-State System with Degraded Mode
4.15 Markov Model for a Component with Three States



# 4. Markov Models

Imagine trying to predict tomorrow's weather. While a perfect forecast is elusive, we often rely on the current conditions. A sunny day is more likely to be followed by another sunny day than by a sudden blizzard. This intuitive idea – that the future state depends only on the present state – is the core principle behind Markov Models.

Building upon this foundation, we can formalize this idea of transitioning between states. A Markov Model is a mathematical system that evolves over time according to specific probabilistic rules. A key characteristic of these models is the Markov property, which states that the future state of the system depends only on its current state, not on the sequence of events that preceded 1 it. In essence, the system has no memory of the past beyond its present condition.

To explore the applications and intricacies of Markov Models, we can broadly categorize them based on how time progresses through the system. This chapter will delve into two primary types:

First, we will examine Markov Chains. These models describe systems where transitions between states occur at discrete points in time. Think of our weather example: we might check the weather once a day, and the state (sunny, rainy, cloudy) changes from one day to the next in distinct steps. Markov Chains are particularly useful for analyzing sequences of events where the timing of transitions is not the primary focus.

Following this, we will explore Markov Processes. In contrast to Markov Chains, Markov Processes model systems where transitions between states can occur at any point in continuous time. Imagine the activity of a radioactive atom that can decay at any moment, or the number of customers in a queue that can change whenever someone arrives or leaves. Markov Processes provide a more nuanced way to model systems where the timing of state changes is crucial.

In the subsequent subchapters, we will explore the mathematical formalism, properties, and applications of both Markov Chains and Markov Processes in detail.

## 4.1 Markov Chains

### 4.1.1 Definition and Basic Concepts

Let's start by defining a Markov Chain and its properties.

**Definition 4.1.1** A Markov Chain is a sequence of random variables $X_0, X_1, X_2, \ldots$ taking values in a countable set $S$, called the state space, such that for any time $n \geq 0$ and any states $i_0, i_1, \ldots, i_n, j \in S$, the following property holds:

$$P(X_{n+1} = j | X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i_n) \tag{4.1}$$

This fundamental property is known as the Markov property or the memoryless property. It implies that the future state of the system depends only on the present state, and is independent of the past history of the system.

The state space $S$ can be finite or countably infinite. The elements of $S$ represent the possible conditions or values that the system can be in at any given time.

#### The Transition Matrix

The probability of transitioning from a state $i \in S$ at time $n$ to a state $j \in S$ at time $n+1$ is called the transition probability and is denoted by $p_{ij}(n)$:

$$p_{ij}(n) = P(X_{n+1} = j | X_n = i) \tag{4.2}$$

If the transition probabilities $p_{ij}(n)$ do not depend on the time $n$, i.e., $p_{ij}(n) = p_{ij}$ for all $n \geq 0$, then the Markov Chain is said to be time-homogeneous. For the remainder of this chapter, we will primarily focus on time-homogeneous Markov Chains.

For a time-homogeneous Markov Chain with a finite state space $S = \{1, 2, \ldots, m\}$, the transition probabilities can be arranged in a matrix called the transition matrix, denoted by $P$:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix} \tag{4.3}$$

Each entry $p_{ij}$ represents the probability of transitioning from state $i$ to state $j$ in one step. The rows of the transition matrix must sum to 1, as from any state $i$, the system must transition to some state $j \in S$ with probability 1:

$$\sum_{j \in S} p_{ij} = 1 \quad \text{for all } i \in S \tag{4.4}$$

A matrix with non-negative entries where each row sums to 1 is called a stochastic matrix. Therefore, the transition matrix of a Markov Chain is a stochastic matrix.

The initial state of the Markov Chain is given by a probability distribution over the state space $S$ at time $n = 0$, denoted by $\pi_0$. If $S$ is finite, $\pi_0$ can be represented as a row vector:

$$\pi_0 = (\pi_{0,1}, \pi_{0,2}, \ldots, \pi_{0,m}) \tag{4.5}$$

where $\pi_{0,i} = P(X_0 = i)$ and $\sum_{i \in S} \pi_{0,i} = 1$.

### n-Step Transition Probabilities and the Chapman-Kolmogorov Equations

The probability of transitioning from state $i$ to state $j$ in $n$ steps is denoted by $p_{ij}^{(n)}$:

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i) \tag{4.6}$$

These $n$-step transition probabilities can be found by raising the transition matrix $P$ to the power of $n$. If $P^n$ denotes the $n$-th power of the transition matrix, then the entry in the $i$-th row and $j$-th column of $P^n$, denoted by $(P^n)_{ij}$, is equal to $p_{ij}^{(n)}$.

A fundamental relationship for calculating $n$-step transition probabilities is given by the Chapman-Kolmogorov equations. For any states $i, j \in S$ and any time $n, m \geq 0$, we have:

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)} \tag{4.7}$$

In matrix notation, this can be written as:

$$P^{n+m} = P^n P^m \tag{4.8}$$

This equation states that to go from state $i$ to state $j$ in $n+m$ steps, the process must be in some state $k$ after $n$ steps, and then transition from state $k$ to state $j$ in the remaining $m$ steps. We sum over all possible intermediate states $k$.

The probability distribution of the Markov Chain at any time $n$, denoted by $\pi_n$, can be obtained by multiplying the initial distribution $\pi_0$ by the $n$-th power of the transition matrix:

$$\pi_n = \pi_0 P^n \tag{4.9}$$

where $\pi_n$ is a row vector whose $j$-th component is $P(X_n = j)$.

### Examples of Markov Chains

#### Random Walk on a Line

Consider a particle moving on the integers. At each time step, the particle moves one step to the right with probability $p$ or one step to the left with probability $q = 1 - p$. The state space is $S = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. The transition probabilities are:

- $P(X_{n+1} = i + 1 | X_n = i) = p$
- $P(X_{n+1} = i - 1 | X_n = i) = q$
- $P(X_{n+1} = j | X_n = i) = 0$ for $|j - i| \neq 1$

This is an example of a Markov Chain with an infinite state space.

#### Weather Prediction

Consider a simplified model of the weather with three states: Sunny (S), Cloudy (C), and Rainy (R). Suppose the transition probabilities are given by the following matrix:

$$P = \begin{pmatrix} P(S|S) & P(S|C) & P(S|R) \\ P(C|S) & P(C|C) & P(C|R) \\ P(R|S) & P(R|C) & P(R|R) \end{pmatrix} = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} \tag{4.10}$$

This matrix tells us, for example, that if it is sunny today, there is a 70% chance it will be sunny tomorrow, a 20% chance it will be cloudy, and a 10% chance it will be rainy.

We can represent this weather prediction Markov Chain with the following graph:

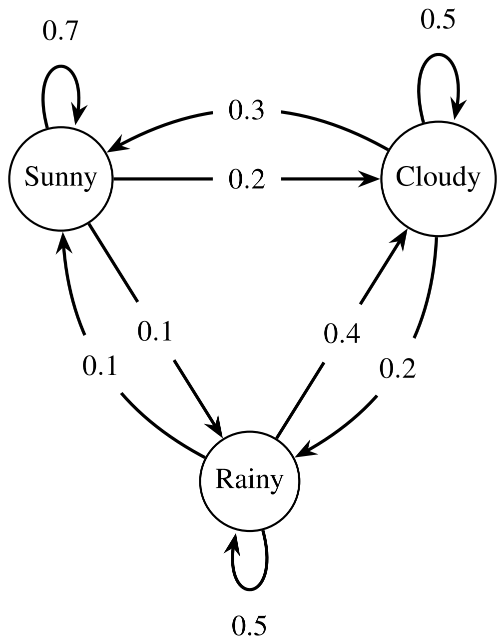



Figure 4.1: Markov Chain in graph form for the weather model example

In this graph:

- The nodes represent the three possible states of the weather: Sunny, Cloudy, and Rainy.
- The arrows represent the transitions between these states.
- The label on each arrow indicates the probability of that transition occurring. For example, the arrow from "Sunny" to "Cloudy" is labeled 0.2, meaning there is a 20% probability of transitioning from a sunny day to a cloudy day.
- A self-loop on a node (e.g., from "Sunny" back to "Sunny") indicates the probability of staying in that state for the next time step.

This graphical representation provides a clear and intuitive way to visualize the transitions and their probabilities in the Markov Chain model.

If today is sunny, the probability of the weather being rainy two days from now can be found by calculating $P^2$:

$$P^2 = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.56 & 0.28 & 0.16 \\ 0.38 & 0.39 & 0.23 \\ 0.24 & 0.42 & 0.34 \end{pmatrix} \tag{4.11}$$

The entry in the first row (Sunny) and third column (Rainy) of $P^2$ is 0.16. Therefore, if today is sunny, there is a 16% chance it will be rainy two days from now.

### Steady-State Vectors

As a Markov Chain evolves over time, its probability distribution over the states may converge to a particular distribution, regardless of the initial state. This limiting distribution, if it exists, is called the steady-state vector or the stationary distribution. It represents the long-term probabilities of the system being in each of the possible states after a sufficiently large number of steps. Understanding the steady-state vector provides valuable insights into the long-term behavior and equilibrium of the system being modeled by the Markov Chain.

The steady-state vector $\pi$ of a time-homogeneous Markov Chain with a transition matrix $P$ is a probability distribution that remains unchanged after one step of the chain. Mathematically, it satisfies the following conditions:

1. $\pi P = \pi$
2. $\sum_{i \in S} \pi_i = 1$, where $S$ is the state space and $\pi_i \geq 0$ for all $i \in S$.

Here, $\pi$ is typically represented as a row vector. Let's consider a Markov Chain with a finite state space $S = \{1, 2, \ldots, m\}$. The steady-state vector is then $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$.

The equation $\pi P = \pi$ can be rewritten as:

$$\pi P - \pi = \begin{pmatrix} 0 & 0 & .. & 0 \end{pmatrix} \tag{4.12}$$

We can factor out $\pi$ to get:

$$\pi(P - I_m) = \begin{pmatrix} 0 & 0 & .. & 0 \end{pmatrix} \tag{4.13}$$

where $I_m$ is the $m \times m$ identity matrix. This matrix equation represents a system of $m$ linear equations for the $m$ unknown components of $\pi$.

Let $P - I_m = Q$. Then the system of equations is $\pi Q = \begin{pmatrix} 0 & 0 & .. & 0 \end{pmatrix}$. Writing this out in component form, we have:

$$\sum_{i=1}^{m} \pi_i q_{ij} = 0 \quad \text{for } j = 1, 2, \ldots, m \tag{4.14}$$

where $q_{ij} = p_{ij} - \delta_{ij}$ ($p_{ij}$ are the elements of $P$, and $\delta_{ij}$ is the Kronecker delta, which is 1 if $i = j$ and 0 otherwise).

However, these $m$ equations are not all linearly independent. Since the rows of the transition matrix $P$ sum to 1, it can be shown that the rows of $(P - I_m)$ are linearly dependent. Therefore, we typically use $m-1$ of these equations along with the normalization condition:

$$\pi_1 + \pi_2 + \cdots + \pi_m = 1 \tag{4.15}$$

to solve for the $m$ unknowns $\pi_1, \pi_2, \ldots, \pi_m$.

In practice, the steps to compute the steady-state vector are as follows:

1. Form the matrix $Q = P - I_m$.
2. Write down the system of linear equations represented by $\pi Q = (0 \quad 0 \quad .. \quad 0)$.
3. Choose $m-1$ linearly independent equations from this system.
4. Include the normalization condition $\sum_{i=1}^{m} \pi_i = 1$ as the $m$-th equation.
5. Solve this system of $m$ linear equations for the $m$ unknowns $\pi_1, \pi_2, \ldots, \pi_m$.

Alternatively, the steady-state vector $\pi$ is the left eigenvector of the transition matrix $P$ corresponding to the eigenvalue 1. That is, if we consider the transpose of $P$, denoted by $P^T$, then $\pi^T$ is the eigenvector of $P^T$ corresponding to the eigenvalue 1:

$$P^T \pi^T = 1 \cdot \pi^T = \pi^T \tag{4.16}$$

The components of this eigenvector must then be normalized to sum to 1 to obtain the steady-state probability distribution.

For irreducible, aperiodic Markov Chains, a unique steady-state distribution exists.

#### Steady-State Vector for the Weather Prediction Example

The transition matrix for the weather prediction Markov Chain was given by:

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} \tag{4.17}$$

The steady-state vector $\pi = (\pi_S, \pi_C, \pi_R)$ represents the long-term probabilities of being in the Sunny (S), Cloudy (C), or Rainy (R) states, respectively. It satisfies the equation $\pi P = \pi$ and the normalization condition $\pi_S + \pi_C + \pi_R = 1$.

The equation $\pi P = \pi$ can be written as:

$$\begin{pmatrix} \pi_S & \pi_C & \pi_R \end{pmatrix} \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix} = \begin{pmatrix} \pi_S & \pi_C & \pi_R \end{pmatrix} \tag{4.18}$$

This leads to the following system of linear equations:

$$0.7\pi_S + 0.3\pi_C + 0.1\pi_R = \pi_S$$
$$0.2\pi_S + 0.5\pi_C + 0.4\pi_R = \pi_C$$
$$0.1\pi_S + 0.2\pi_C + 0.5\pi_R = \pi_R$$

Rearranging these equations, we get:

$$-0.3\pi_S + 0.3\pi_C + 0.1\pi_R = 0$$
$$0.2\pi_S - 0.5\pi_C + 0.4\pi_R = 0$$
$$0.1\pi_S + 0.2\pi_C - 0.5\pi_R = 0$$

We also have the normalization condition:

$$\pi_S + \pi_C + \pi_R = 1 \tag{4.19}$$

Solving the system of equations, the steady-state vector for the weather prediction Markov Chain is:

$$\pi = (0.425, 0.35, 0.225) = \left(\frac{17}{40}, \frac{7}{20}, \frac{9}{40}\right) \tag{4.20}$$

This means that in the long run, the weather will be sunny 42.5% of the time, cloudy 35% of the time, and rainy 22.5% of the time.

### Example of a Markov Chain for System Reliability

Consider a machine that can be in one of two states at the beginning of each day: Operational (O) or Failed (F). We model the transitions between these states using a Markov Chain. Suppose the transition probabilities are as follows:

- If the machine is Operational today, there is a 90% chance it will be Operational tomorrow and a 10% chance it will Fail.
- If the machine is Failed today, there is a 60% chance it will be Repaired and become Operational tomorrow and a 40% chance it will remain Failed.

#### State Space and Transition Matrix

The state space is $S = \{O, F\}$. The transition matrix $P$, where $P_{ij} = P(\text{State tomorrow} = j | \text{State today} = i)$, is given by:

$$P = \begin{pmatrix} P(O|O) & P(F|O) \\ P(O|F) & P(F|F) \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \end{pmatrix} \tag{4.21}$$

#### Reliability Evaluation

Let's assume the machine starts in the Operational state on day 0. The initial state vector is $\pi_0 = \begin{pmatrix} 1 & 0 \end{pmatrix}$.

We can find the probability distribution of the machine's state after $n$ days by calculating $\pi_n = \pi_0 P^n$.

**Reliability after 1 day**

$\pi_1 = \pi_0 P = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 \end{pmatrix}$

The probability that the machine is Operational after 1 day is 0.9. This can be considered the reliability for a 1-day mission.

**Reliability after 2 days**

$\pi_2 = \pi_1 P = \begin{pmatrix} 0.9 & 0.1 \end{pmatrix} \begin{pmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.87 & 0.13 \end{pmatrix}$

The probability that the machine is Operational after 2 days is 0.87.

**Steady-State Reliability (Long-Term Availability)**

To find the steady-state probability vector $\pi = \begin{pmatrix} \pi_O & \pi_F \end{pmatrix}$, we solve the equation $\pi P = \pi$ subject to $\pi_O + \pi_F = 1$.

$\begin{pmatrix} \pi_O & \pi_F \end{pmatrix} \begin{pmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \end{pmatrix} = \begin{pmatrix} \pi_O & \pi_F \end{pmatrix}$

This gives the following system of equations:

$$0.9\pi_O + 0.6\pi_F = \pi_O \tag{4.22}$$

$$0.1\pi_O + 0.4\pi_F = \pi_F \tag{4.23}$$

And the normalization condition:

$$\pi_O + \pi_F = 1 \tag{4.24}$$

From the first equation: $-0.1\pi_O + 0.6\pi_F = 0 \implies \pi_O = 6\pi_F$. Substituting this into the normalization condition: $6\pi_F + \pi_F = 1 \implies 7\pi_F = 1 \implies \pi_F = \frac{1}{7}$. Then, $\pi_O = 6 \times \frac{1}{7} = \frac{6}{7}$.

The steady-state probability of the machine being Operational is $\frac{6}{7} \approx 0.857$. This represents the long-term reliability or availability of the machine.

These examples illustrate how Markov Chains can be used to model various systems that evolve over time based on probabilistic transitions. In the subsequent sections, we will delve deeper into the properties and analysis of Markov Processes, which are more versatile in modeling the reliability of complex systems.

## 4.2 Markov Processes

### 4.2.1 Definition and Basic Concepts

A Markov Process is a type of stochastic process where the future state of the system depends only on its current state, and is independent of its past history. It's the continuous-time equivalent of a Markov Chain.

Key characteristics of a Markov Process:

- Continuous Time: The process evolves over a continuous range of time (as opposed to discrete time steps in Markov Chains).
- State Space: The system can be in one of a countable number of states.
- Markov Property (Memoryless Property): The probability of transitioning to a future state depends solely on the current state, not on the sequence of events that led to that state.

In simpler terms, if you know the current state of a Markov Process, you have all the information you need to predict its future behavior; the past doesn't add any extra predictive power.

**Definition 4.2.1** A Markov Process is a stochastic process $\{X(t), t \geq 0\}$ that takes values in a countable state space $S$, such that for any time $t, s \geq 0$ and any states $i, j \in S$, the conditional probability of being in state $j$ at time $t+s$, given the history of the process up to time $t$ and the state at time $t$, depends only on the state at time $t$. Formally, for any $t \geq 0$, $s > 0$, and states $i, j, i_u \in S$ for $0 \leq u < t$, we have:

$$P(X(t+s) = j | X(t) = i, X(u) = i_u \text{ for all } 0 \leq u < t) = P(X(t+s) = j | X(t) = i) \quad (4.25)$$

This property is the continuous-time analogue of the Markov property for Markov Chains. It signifies that the future evolution of the process depends solely on its present state, and is independent of its past history.

We define the transition probability function $P(i,t; j,s)$ as the probability that the process is in state $j$ at time $s$, given that it was in state $i$ at time $t$, where $t \leq s$:

$$P(i,t; j,s) = P(X(s) = j | X(t) = i) \quad (4.26)$$

If the transition probabilities depend only on the time difference $s-t$, and not on the absolute times $t$ and $s$, the Markov Process is said to be time-homogeneous. In this case, we can write the transition probability as:

$$P(i,t; j,s) = P(i,0; j,s-t) = p_{ij}(s-t) \quad (4.27)$$

For the remainder of this chapter, we will focus on time-homogeneous Markov Processes. The state space $S$ is assumed to be countable.

### 4.2.2 The Chapman-Kolmogorov Differential Equation System

Consider a time-homogeneous Markov Process with transition probabilities $p_{ij}(t)$. For any states $i, j \in S$ and any times $t, s \geq 0$, the Chapman-Kolmogorov equation holds:

$$p_{ij}(t+s) = \sum_{k \in S} p_{ik}(t) p_{kj}(s) \quad (4.28)$$

This equation states that to transition from state $i$ to state $j$ in time $t+s$, the process must be in some intermediate state $k$ at time $t$, and then transition from state $k$ to state $j$ in the remaining time $s$. We sum over all possible intermediate states $k$.

To derive the differential equations governing these transition probabilities, we introduce the concept of transition rates. Let $\lambda_{ij}$ be the rate of transition from state $i$ to state $j$ for $i \neq j$. We define this as:

$$\lambda_{ij} = \lim_{\Delta t \to 0} \frac{p_{ij}(t)}{\Delta t} \quad \text{for } i \neq j \quad (4.29)$$

This represents the instantaneous rate of transitioning from state $i$ to state $j$. We also define $\lambda_{ii}$ as the negative of the rate of leaving state $i$:

$$\lambda_{ii} = \lim_{\Delta t \to 0} \frac{p_{ii}(t) - 1}{\Delta t} = -\sum_{j \neq i} \lambda_{ij} \tag{4.30}$$

Let $\lambda_i = -\lambda_{ii} = \sum_{j \neq i} \lambda_{ij}$ be the total rate of leaving state $i$.

Now, consider the Chapman-Kolmogorov equation and let $s = \Delta t$ be a small time interval. We have:

$$p_{ij}(t + \Delta t) = \sum_{k \in S} p_{ik}(t) p_{kj}(\Delta t) \tag{4.31}$$

We can approximate $p_{kj}(\Delta t)$ for small $\Delta t$ as follows:

- For $k = j$, $p_{jj}(\Delta t) \approx 1 + \lambda_{jj}\Delta t = 1 - \lambda_j \Delta t$.
- For $k \neq j$, $p_{kj}(\Delta t) \approx \lambda_{kj}\Delta t$.

Substituting these approximations into the Chapman-Kolmogorov equation:

$$p_{ij}(t + \Delta t) \approx p_{ij}(t)(1 - \lambda_j \Delta t) + \sum_{k \neq j} p_{ik}(t)(\lambda_{kj}\Delta t) \tag{4.32}$$

Rearranging and dividing by $\Delta t$:

$$\frac{p_{ij}(t + \Delta t) - p_{ij}(t)}{\Delta t} \approx -\lambda_j p_{ij}(t) + \sum_{k \neq j} \lambda_{kj} p_{ik}(t) \tag{4.33}$$

Taking the limit as $\Delta t \to 0$, we obtain the forward Kolmogorov equations (also known as the forward equations or the Kolmogorov forward differential equations):

$$\frac{dp_{ij}(t)}{dt} = -\lambda_j p_{ij}(t) + \sum_{k \in S, k \neq j} \lambda_{kj} p_{ik}(t) = \sum_{k \in S} p_{ik}(t) \lambda_{kj} \tag{4.34}$$

Similarly, by considering $p_{ij}(h+t) = \sum_{k \in S} p_{ik}(h) p_{kj}(t)$ and using the approximations for $p_{ik}(h)$, we can derive the backward Kolmogorov equations (also known as the backward equations or the Kolmogorov backward differential equations):

$$\frac{dp_{ij}(t)}{dt} = -\lambda_i p_{ij}(t) + \sum_{k \in S, k \neq i} \lambda_{ik} p_{kj}(t) = \sum_{k \in S} \lambda_{ik} p_{kj}(t) \tag{4.35}$$

These systems of differential equations describe how the transition probabilities of a time-homogeneous Markov Process change over time, based on the transition rates between the states. Solving these equations (often a challenging task) allows us to determine the probability of being in any state at any time $t$, given the initial state.

### 4.2.3 Matricial Form of Chapman-Kolmogorov Equations for a Markov Process

Let $P(t)$ be the matrix of transition probabilities at time $t$, where the $(i, j)$-th entry is $p_{ij}(t) = P(X(t) = j|X(0) = i)$. Let $Q$ be the transition rate matrix, with entries $\lambda_{ij}$ defined as:

- $\lambda_{ij} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t}$ for $i \neq j$
- $\lambda_{ii} = \lim_{\Delta t \to 0} \frac{p_{ii}(\Delta t)-1}{\Delta t} = -\sum_{j \neq i} \lambda_{ij}$

The forward and backward Kolmogorov differential equation systems can be expressed in matrix form as follows:

#### Forward Kolmogorov Equations (Matrix Form)

The forward Kolmogorov equations are given by:

$$\frac{dp_{ij}(t)}{dt} = \sum_{k \in S} p_{ik}(t)\lambda_{kj} \tag{4.36}$$

In matrix notation, this system of equations can be written as:

$$\frac{dP(t)}{dt} = P(t)Q \tag{4.37}$$

where $\frac{dP(t)}{dt}$ is the matrix whose $(i, j)$-th entry is $\frac{dp_{ij}(t)}{dt}$. This equation describes how the transition probability matrix evolves over time based on the current transition probabilities and the transition rates.

#### Backward Kolmogorov Equations (Matrix Form)

The backward Kolmogorov equations are given by:

$$\frac{dp_{ij}(t)}{dt} = \sum_{k \in S} \lambda_{ik} p_{kj}(t) \tag{4.38}$$

In matrix notation, this system of equations can be written as:

$$\frac{dP(t)}{dt} = QP(t) \tag{4.39}$$

Here, the order of multiplication of the matrices is reversed compared to the forward equations. This formulation considers the state of the process at an earlier time (time 0) and how the initial state influences the probability of being in state $j$ at time $t$.

Both the forward and backward Kolmogorov equations provide fundamental ways to analyze the evolution of a Markov Process in continuous time. The choice of which set of equations to use often depends on the specific problem and the properties of the transition rate matrix $Q$.

### 4.2.4 Applications of Markov Processes

Markov Processes are powerful tools for modeling a wide variety of real-world systems that evolve over time in a probabilistic manner. Their memoryless property makes them tractable while still capturing essential dynamics. Here are some key application areas:

#### Queueing Theory

Markov Processes are fundamental to queueing theory, which studies waiting lines. Models like the M/M/1 queue (Poisson arrivals, exponential service times, single server) are based on Markov Processes. The state of the system is the number of customers in the queue, and transitions occur due to arrivals and departures. Markov Processes allow us to analyze queue lengths, waiting times, and system utilization.

#### Reliability Theory

In reliability theory, Markov Processes can model the failure and repair of systems. The states represent the operational status of the components or the system as a whole (e.g., working, failed, under repair). Transitions occur between these states with certain rates. This allows for the calculation of system availability, mean time to failure, and other reliability metrics.

#### Population Dynamics

Markov Processes can be used to model the growth and decline of populations, including birth and death processes. The state represents the number of individuals in the population, and transitions correspond to births and deaths occurring at certain rates. These models can help predict population sizes and analyze factors affecting population growth.

#### Finance

In finance, Markov Processes are used in various applications, such as modeling stock prices, interest rates, and credit risk. For example, a stock price might be modeled as transitioning between different states (e.g., bull market, bear market) with certain probabilities. These models can be used for option pricing and risk management.

#### Telecommunications

Markov Processes are employed in telecommunications to model network traffic, call arrivals and departures in telephone systems, and the performance of communication channels. The state might represent the number of active connections or the status of a channel. This helps in designing efficient and reliable communication networks.

#### Biology

In biology, Markov Processes find applications in modeling molecular processes, such as protein folding, gene expression, and the movement of molecules within cells. The states represent different configurations or states of the biological system, and transitions occur due to random fluctuations.

Markov Processes provide a versatile and powerful framework for modeling dynamic systems that evolve probabilistically over continuous time, with applications spanning numerous scientific, engineering, and business domains. The Chapman-Kolmogorov differential equations are fundamental for analyzing the behavior of these processes.

## 4.3 Single-Component System without Repair

Consider a single-component system that can be in one of two states: operational or failed. We can model this system as a Markov process with two states:

- State O: Operational
- State F: Failed

The system starts in the operational state at time $t = 0$. The component can transition from the operational state to the failed state with a constant failure rate $\lambda > 0$. Since there is no repair, the system cannot transition back from the failed state to the operational state.

### 4.3.1 State Diagram

The transitions between the states can be represented by the following state diagram:

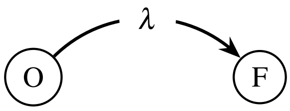



Figure 4.2: State transition graph of the system

### 4.3.2 Transition Rate Matrix

The transition rate matrix $Q$ for this Markov process is a $2 \times 2$ matrix, where $\lambda_{ij}$ represents the rate of transition from state $i$ to state $j$:

$$Q = \begin{pmatrix} \lambda_{00} & \lambda_{01} \\ \lambda_{10} & \lambda_{11} \end{pmatrix}$$

For our system:

- $\lambda_{00} = -\lambda$ (rate of leaving state 0)
- $\lambda_{01} = \lambda$ (rate of transitioning from state 0 to state 1)
- $\lambda_{10} = 0$ (rate of transitioning from state 1 to state 0, as there is no repair)
- $\lambda_{11} = 0$ (rate of leaving state 1, as it is an absorbing state)

Thus, the transition rate matrix is:

$$Q = \begin{pmatrix} -\lambda & \lambda \\ 0 & 0 \end{pmatrix}$$

### 4.3.3 System Reliability

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. The system reliability $R(t)$ is the probability that the system is in the operational state (State 0) at time $t$, given that it started in the operational state at $t = 0$. So, $R(t) = P_0(t)$.

The differential equation for $P_0(t)$ can be written as:

$$\frac{dP_0(t)}{dt} = \lambda_{00} P_0(t) + \lambda_{10} P_1(t)$$

Substituting the values from the transition rate matrix:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + 0 \cdot P_1(t) = -\lambda P_0(t)$$

This is a first-order linear differential equation. With the initial condition $P_0(0) = 1$ (the system starts operational), we can solve it:

$$\int \frac{dP_0}{P_0} = \int -\lambda dt$$

$$\ln |P_0(t)| = -\lambda t + C$$

$$P_0(t) = e^{-\lambda t + C} = e^C e^{-\lambda t}$$

Using the initial condition $P_0(0) = 1$:

$$1 = e^C e^{-\lambda \cdot 0} = e^C \cdot 1 \implies e^C = 1$$

Therefore, the probability of being in the operational state at time $t$ is:

$$P_0(t) = e^{-\lambda t}$$

The system reliability $R(t)$ is thus:

$$R(t) = e^{-\lambda t}$$

### 4.3.4 Mean Time To Failure (MTTF)

The Mean Time To Failure (MTTF) for a non-repairable system is the expected time until the system enters the failed state. It can be calculated as the integral of the reliability function from $t = 0$ to $t = \infty$:

$$MTTF = \int_0^\infty R(t)dt = \int_0^\infty e^{-\lambda t} dt$$

$$MTTF = \left[ -\frac{1}{\lambda} e^{-\lambda t} \right]_0^\infty = \lim_{T \to \infty} \left( -\frac{1}{\lambda} e^{-\lambda T} \right) - \left( -\frac{1}{\lambda} e^{-\lambda \cdot 0} \right)$$

Since $\lambda > 0$, as $T \to \infty$, $e^{-\lambda T} \to 0$. Therefore:

$$MTTF = 0 - \left( -\frac{1}{\lambda} \cdot 1 \right) = \frac{1}{\lambda}$$

Thus, the Mean Time To Failure for a single-component system without repair is the reciprocal of the failure rate.

## 4.4 Single-Component with Repair

Consider a system composed of a single component that can be in one of two states: fully operational or failed. We assume that repair is available for the failed component.

### 4.4.1 State Space

The system can be in one of the following two states:

- State O: Operational
- State F: Failed

### 4.4.2 Transition Rates

We assume the following constant transition rates:

- $\lambda$: Failure rate of the component (from Operational to Failed)
- $\mu$: Repair rate of the component (from Failed to Operational)

### 4.4.3 Transition Rate Matrix

The transition rate matrix $Q$ for this Markov Process is a $2 \times 2$ matrix:

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

### 4.4.4 Graph of the Markov Process

The Markov Process can be represented by the following state transition diagram:

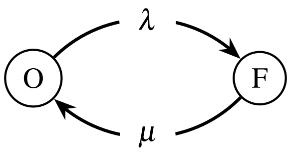



Figure 4.3: State transition graph of the system

### 4.4.5 Solving for State Probabilities (Reliability)

Let $p_1(t)$ be the probability that the system is in the Operational state at time $t$, and $p_2(t)$ be the probability that the system is in the Failed state at time $t$. Assuming the system starts in the Operational state at $t = 0$, the initial conditions are $p_1(0) = 1$ and $p_2(0) = 0$. The forward Kolmogorov differential equations are:

$$\frac{dp_1(t)}{dt} = -\lambda p_1(t) + \mu p_2(t)$$
$$\frac{dp_2(t)}{dt} = \lambda p_1(t) - \mu p_2(t)$$

Solving this system of equations, we obtain:

$$p_1(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}$$

$$p_2(t) = \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} = \frac{\lambda}{\lambda+\mu}(1 - e^{-(\lambda+\mu)t})$$

### 4.4.6 System Reliability

The reliability of the system at time $t$, $R(t)$, is the probability that the system is in the operational state at time $t$, given that it started in the operational state. Therefore,

$$R(t) = p_1(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}$$

### 4.4.7 Mean Time To Failure (MTTF)

The Mean Time To Failure (MTTF) is the expected time until the system enters the failed state for the first time, starting from the operational state. For a simple two-state model where the system transitions from operational to failed with rate $\lambda$, the time to failure follows an exponential distribution with parameter $\lambda$. The mean of this distribution is $1/\lambda$.

Alternatively, MTTF can be calculated by considering the expected sojourn time in the operational state before transitioning to the failed state. For a Markov Process, the sojourn time in a state $i$ is exponentially distributed with rate $-q_{ii}$. In the operational state, the rate of leaving is $\lambda$, so the expected time spent in the operational state before failure is $1/\lambda$.

Thus, the Mean Time To Failure for this single-component system is:

$$MTTF = \frac{1}{\lambda}$$

Note that the repair rate $\mu$ affects the long-term availability and the probability of being in the operational state at any given time, but the mean time to the first failure from an operational state is determined solely by the failure rate $\lambda$.

### 4.4.8 System Availability for a Single-Component System with Repair

The probability of the single-component system being in the operational state at time $t$ was derived as:

$$p_1(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}$$

The availability of the system at time $t$, denoted by $A(t)$, is the probability that the system is operational at that time. Thus,

$$A(t) = p_1(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}$$

In many practical scenarios, we are interested in the steady-state availability, which is the long-term probability that the system is operational. This is obtained by taking the limit of $A(t)$ as $t$ approaches infinity:

$$A_{ss} = \lim_{t\to\infty} A(t) = \lim_{t\to\infty} \left( \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t} \right)$$

Since $\lambda > 0$ and $\mu \geq 0$, we have $\lambda + \mu > 0$, so $\lim_{t\to\infty} e^{-(\lambda+\mu)t} = 0$. Therefore, the steady-state availability is:

$$A_{ss} = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} \times 0 = \frac{\mu}{\lambda+\mu}$$

The steady-state availability represents the proportion of time, on average, that the system is operational in the long run. It is a crucial metric for assessing the performance of repairable systems.

We can also express this in terms of Mean Time To Failure (MTTF) and Mean Time To Repair (MTTR). For a single component, $MTTF = 1/\lambda$ and $MTTR = 1/\mu$. Substituting these into the expression for steady-state availability:

$$A_{ss} = \frac{1/\text{MTTR}}{1/\text{MTTF} + 1/\text{MTTR}} = \frac{1/\text{MTTR}}{(\text{MTTR}+\text{MTTF})/(\text{MTTF}\cdot\text{MTTR})} = \frac{\text{MTTF}}{\text{MTTF}+\text{MTTR}}$$

This alternative form highlights the relationship between system availability, the average time the system operates before failing, and the average time it takes to repair the system.

## 4.5 Two-Component System Without Repair

Consider a system composed of two independent components, C1 and C2. Each component can be in one of two states: fully operational or failed. We assume that once a component fails, it remains in the failed state (no repair).

### 4.5.1 State Space

The system can be in one of the following four states:

- State 1 (OO): C1 Operational, C2 Operational
- State 2 (OF): C1 Operational, C2 Failed
- State 3 (FO): C1 Failed, C2 Operational
- State 4 (FF): C1 Failed, C2 Failed

### 4.5.2 Transition Rates

We assume the following constant failure rates:

- $\lambda_1$: Failure rate of component C1 (from Operational to Failed)
- $\lambda_2$: Failure rate of component C2 (from Operational to Failed)

Since there is no repair, the repair rates $\mu_1 = 0$ and $\mu_2 = 0$.

### 4.5.3 Transition Rate Matrix

The transition rate matrix $Q$ for this Markov Process is a $4 \times 4$ matrix:

$$Q = \begin{pmatrix} -(\lambda_1+\lambda_2) & \lambda_2 & \lambda_1 & 0 \\ 0 & -\lambda_1 & 0 & \lambda_1 \\ 0 & 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

### 4.5.4 Graph of the Markov Process

The Markov Process can be represented by the following state transition diagram:

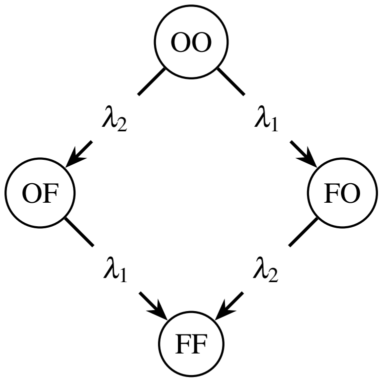



Figure 4.4: State transition graph of the two-component system

### 4.5.5 Solving for State Probabilities (Reliability)

Let $p_i(t)$ be the probability that the system is in state $i$ at time $t$, for $i = 1,2,3,4$. Assuming the system starts in the fully operational state (OO) at $t = 0$, the initial conditions are $p_1(0) = 1$, $p_2(0) = 0$, $p_3(0) = 0$, and $p_4(0) = 0$. The forward Kolmogorov differential equations are:

$$
\begin{aligned}
\frac{dp_1(t)}{dt} &= -(\lambda_1 + \lambda_2)p_1(t) \\
\frac{dp_2(t)}{dt} &= \lambda_2 p_1(t) - \lambda_1 p_2(t) \\
\frac{dp_3(t)}{dt} &= \lambda_1 p_1(t) - \lambda_2 p_3(t) \\
\frac{dp_4(t)}{dt} &= \lambda_1 p_2(t) + \lambda_2 p_3(t)
\end{aligned}
$$

Solving these equations with the given initial conditions:

1. From the first equation, $\frac{dp_1(t)}{dt} = -(\lambda_1 + \lambda_2)p_1(t)$, we get:

$$p_1(t) = e^{-(\lambda_1+\lambda_2)t}$$

2. Substituting $p_1(t)$ into the second equation: $\frac{dp_2(t)}{dt} + \lambda_1 p_2(t) = \lambda_2 e^{-(\lambda_1+\lambda_2)t}$. The solution is:

$$p_2(t) = e^{-\lambda_1 t} - e^{-(\lambda_1+\lambda_2)t}$$

3. Substituting $p_1(t)$ into the third equation: $\frac{dp_3(t)}{dt} + \lambda_2 p_3(t) = \lambda_1 e^{-(\lambda_1+\lambda_2)t}$. The solution is:

$$p_3(t) = e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}$$

4. Since $p_1(t) + p_2(t) + p_3(t) + p_4(t) = 1$, we can find $p_4(t)$:

$$
\begin{aligned}
p_4(t) &= 1 - p_1(t) - p_2(t) - p_3(t) \\
&= 1 - e^{-(\lambda_1+\lambda_2)t} - (e^{-\lambda_1 t} - e^{-(\lambda_1+\lambda_2)t}) - (e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}) \\
&= 1 - e^{-\lambda_1 t} - e^{-\lambda_2 t} + e^{-(\lambda_1+\lambda_2)t}
\end{aligned}
$$

The reliability of each state at time $t$ is given by these probabilities:

- Probability of being in state OO (both operational): $p_1(t) = e^{-(\lambda_1+\lambda_2)t}$
- Probability of being in state OF (C1 operational, C2 failed): $p_2(t) = e^{-\lambda_1 t} - e^{-(\lambda_1+\lambda_2)t}$
- Probability of being in state FO (C1 failed, C2 operational): $p_3(t) = e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}$
- Probability of being in state FF (both failed): $p_4(t) = 1 - e^{-\lambda_1 t} - e^{-\lambda_2 t} + e^{-(\lambda_1+\lambda_2)t}$

These equations describe the reliability of each of the four states of the system over time, given that the system starts in the fully operational state.

If we consider the system tolerates no failures, then its reliability is given by $p_1(t) = e^{-(\lambda_1+\lambda_2)t}$, which is equivalent to the reliability of the two components in series.

If we consider the system tolerates the failure of one out of the two components, its reliability can be expressed as $p_1(t) + p_2(t) + p_3(t) = e^{-\lambda_1 t} + e^{-\lambda_2 t} - e^{-(\lambda_1+\lambda_2)t}$, which is equivalent to the reliability of the two components in parallel.

## 4.6 Two-Component System with Repair

Similarly to the previous example, consider a system made of two independent components, C1 and C2. Each component can be either operational (O) or failed (F). We assume that each component can fail with a rate $\lambda_i$ and be repaired with a rate $\mu_i$, where $i = 1,2$ for components C1 and C2, respectively.

### 4.6.1 State Space

The system has four possible states:

- OO: Both C1 and C2 are operational.
- OF: C1 is operational, C2 is failed.
- FO: C1 is failed, C2 is operational.
- FF: Both C1 and C2 are failed.

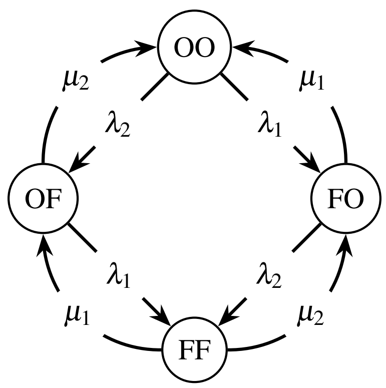



Figure 4.5: State transition graph of the two-component system with repair

### 4.6.2 Transition Rate Matrix

The transition rate matrix $Q$ for this system is given by:

$$Q = \begin{pmatrix} -(\lambda_1+\lambda_2) & \lambda_2 & \lambda_1 & 0 \\ \mu_2 & -(\lambda_1+\mu_2) & 0 & \lambda_1 \\ \mu_1 & 0 & -(\lambda_2+\mu_1) & \lambda_2 \\ 0 & \mu_1 & \mu_2 & -(\mu_1+\mu_2) \end{pmatrix}$$

Solving the associated differential equation system is not trivial and implies a great amount of effort. We can solve an easier case of this system in which the modules have identical failure and repair rates.

### 4.6.3 Simplified Case: Identical Components

Let's consider a simplified case where both components are identical, with failure rate $\lambda$ and repair rate $\mu$ ($\lambda_1 = \lambda_2 = \lambda$, $\mu_1 = \mu_2 = \mu$). Furthermore, due to its symmetry, we can merge the OF and FO states into a single state. The system can be in one of three states based on the number of operational components:

- State 2: Both components are operational.
- State 1: One component is operational, and one is failed.
- State 0: Both components are failed.

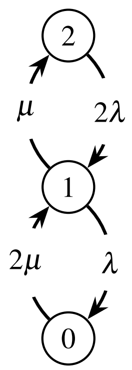



Figure 4.6: State transition graph of the two identical component system with repair

The transition rate matrix can be written as:

$$Q = \begin{pmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -(\mu+\lambda) & \lambda \\ 0 & 2\mu & -2\mu \end{pmatrix}$$

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. Assuming the system starts with both components operational, the initial conditions are $P_2(0) = 1$, $P_1(0) = 0$, and $P_0(0) = 0$.

The Kolmogorov forward differential equations for these probabilities are:

$$\frac{dP_2(t)}{dt} = -2\lambda P_2(t) + \mu P_1(t)$$
$$\frac{dP_1(t)}{dt} = 2\lambda P_2(t) - (\mu+\lambda)P_1(t) + 2\mu P_0(t)$$
$$\frac{dP_0(t)}{dt} = \lambda P_1(t) - 2\mu P_0(t)$$

Solving this system of differential equations is done using Laplace transforms and partial fraction decomposition.

We need to factor in the initial conditions $P_2(0) = 1$, $P_1(0) = 0$, and $P_0(0) = 0$.

**Step 1: Taking the Laplace Transform**

Let $\mathscr{L}\{P_i(t)\} = \tilde{P}_i(s)$. Applying the Laplace transform to each equation, we get:

$$s\tilde{P}_2(s) - P_2(0) = -2\lambda\tilde{P}_2(s) + \mu\tilde{P}_1(s) \tag{4.40}$$

$$s\tilde{P}_1(s) - P_1(0) = 2\lambda\tilde{P}_2(s) - (\mu+\lambda)\tilde{P}_1(s) + 2\mu\tilde{P}_0(s) \tag{4.41}$$

$$s\tilde{P}_0(s) - P_0(0) = \lambda\tilde{P}_1(s) - 2\mu\tilde{P}_0(s) \tag{4.42}$$

Substituting the initial conditions, we obtain the following system of linear algebraic equations:

$$(s+2\lambda)\tilde{P}_2(s) - \mu\tilde{P}_1(s) = 1 \tag{4.43}$$

$$-2\lambda\tilde{P}_2(s) + (s+\mu+\lambda)\tilde{P}_1(s) - 2\mu\tilde{P}_0(s) = 0 \tag{4.44}$$

$$-\lambda\tilde{P}_1(s) + (s+2\mu)\tilde{P}_0(s) = 0 \tag{4.45}$$

**Step 2: Solving the System of Algebraic Equations**

From equation (4.45), we can express $\tilde{P}_0(s)$ in terms of $\tilde{P}_1(s)$:

$$\tilde{P}_0(s) = \frac{\lambda}{s+2\mu}\tilde{P}_1(s)$$

Substitute this into equation (4.44):

$$-2\lambda\tilde{P}_2(s) + (s+\mu+\lambda)\tilde{P}_1(s) - 2\mu\left(\frac{\lambda}{s+2\mu}\tilde{P}_1(s)\right) = 0$$

$$-2\lambda\tilde{P}_2(s) + \left(s+\mu+\lambda - \frac{2\lambda\mu}{s+2\mu}\right)\tilde{P}_1(s) = 0$$

$$-2\lambda\tilde{P}_2(s) + \frac{(s+\mu+\lambda)(s+2\mu) - 2\lambda\mu}{s+2\mu}\tilde{P}_1(s) = 0$$

$$-2\lambda\tilde{P}_2(s) + \frac{s^2+(3\mu+\lambda)s+2\mu^2}{s+2\mu}\tilde{P}_1(s) = 0$$

From equation (4.43), we can express $\tilde{P}_2(s)$ in terms of $\tilde{P}_1(s)$:

$$\tilde{P}_2(s) = \frac{\mu}{s+2\lambda}\tilde{P}_1(s) + \frac{1}{s+2\lambda}$$

Substitute this expression for $\tilde{P}_2(s)$ into the equation above:

$$-2\lambda\left(\frac{\mu}{s+2\lambda}\tilde{P}_1(s) + \frac{1}{s+2\lambda}\right) + \frac{s^2+(3\mu+\lambda)s+2\mu^2}{s+2\mu}\tilde{P}_1(s) = 0$$

Solving for $\tilde{P}_1(s)$, we get:

$$\tilde{P}_1(s) = \frac{2\lambda(s+2\mu)}{s(s+\lambda+\mu)(s+2\lambda+2\mu)}$$

Substituting this back into the expressions for $\tilde{P}_0(s)$ and $\tilde{P}_2(s)$:

$$\tilde{P}_0(s) = \frac{2\lambda^2}{s(s+\lambda+\mu)(s+2\lambda+2\mu)}$$

### Step 3: Partial Fraction Decomposition

For $\tilde{P}_0(s) = \frac{2\lambda^2}{s(s+\lambda+\mu)(s+2(\lambda+\mu))} = \frac{A}{s} + \frac{B}{s+\lambda+\mu} + \frac{C}{s+2(\lambda+\mu)}$

$A = \frac{2\lambda^2}{(\lambda+\mu)(2\lambda+2\mu)} = \frac{\lambda^2}{(\lambda+\mu)^2}$

$B = \frac{2\lambda^2}{(-\lambda-\mu)(\lambda+\mu)} = -\frac{2\lambda^2}{(\lambda+\mu)^2}$

$C = \frac{2\lambda^2}{(-2\lambda-2\mu)(-\lambda-\mu)} = \frac{\lambda^2}{(\lambda+\mu)^2}$

For $\tilde{P}_1(s) = \frac{2\lambda(s+2\mu)}{s(s+\lambda+\mu)(s+2(\lambda+\mu))} = \frac{D}{s} + \frac{E}{s+\lambda+\mu} + \frac{F}{s+2(\lambda+\mu)}$

$D = \frac{4\lambda\mu}{(\lambda+\mu)(2\lambda+2\mu)} = \frac{2\lambda\mu}{(\lambda+\mu)^2}$

$E = \frac{2\lambda(-\lambda-\mu+2\mu)}{(-\lambda-\mu)(\lambda+\mu)} = \frac{2\lambda(\mu-\lambda)}{-(\lambda+\mu)^2} = \frac{2\lambda(\lambda-\mu)}{(\lambda+\mu)^2}$

$F = \frac{2\lambda(-2\lambda-2\mu+2\mu)}{(-2\lambda-2\mu)(-\lambda-\mu)} = \frac{-4\lambda^2}{2(\lambda+\mu)^2} = -\frac{2\lambda^2}{(\lambda+\mu)^2}$

### Step 4: Inverse Laplace Transform

$P_0(t) = \mathscr{L}^{-1}\{\tilde{P}_0(s)\} = \frac{\lambda^2}{(\mu+\lambda)^2} - \frac{2\lambda^2}{(\mu+\lambda)^2}e^{-(\lambda+\mu)t} + \frac{\lambda^2}{(\mu+\lambda)^2}e^{-2(\lambda+\mu)t}$

$P_1(t) = \mathscr{L}^{-1}\{\tilde{P}_1(s)\} = \frac{2\lambda\mu}{(\mu+\lambda)^2} + \frac{2\lambda(\lambda-\mu)}{(\mu+\lambda)^2}e^{-(\lambda+\mu)t} - \frac{2\lambda^2}{(\mu+\lambda)^2}e^{-2(\lambda+\mu)t}$

$P_2(t) = 1 - P_0(t) - P_1(t) = \frac{\mu^2}{(\mu+\lambda)^2} + \frac{2\lambda\mu}{(\mu+\lambda)^2}e^{-(\lambda+\mu)t} + \frac{\lambda^2}{(\mu+\lambda)^2}e^{-2(\lambda+\mu)t}$

We can verify that $P_0(t) + P_1(t) + P_2(t) = 1$ for all $t \geq 0$.

### System Reliability

If we consider the system to be functional only if both components are functional, then the overall reliability is $R(t) = P_2(t) = \frac{\mu^2}{(\mu+\lambda)^2} + \frac{2\lambda\mu}{(\mu+\lambda)^2}e^{-(\lambda+\mu)t} + \frac{\lambda^2}{(\mu+\lambda)^2}e^{-2(\lambda+\mu)t}$

However, if the system tolerates one failed component, then the reliability is that of a parallel system:

$$
\begin{aligned}
R(t) &= P_0(t) + P_1(t) \\
&= \left(\frac{\mu^2}{(\mu+\lambda)^2} + \frac{2\lambda\mu}{(\mu+\lambda)^2}e^{-(\lambda+\mu)t} + \frac{\lambda^2}{(\mu+\lambda)^2}e^{-2(\lambda+\mu)t}\right) \\
&\quad + \left(\frac{2\lambda\mu}{(\mu+\lambda)^2} + \frac{2\lambda(\lambda-\mu)}{(\mu+\lambda)^2}e^{-(\lambda+\mu)t} - \frac{2\lambda^2}{(\mu+\lambda)^2}e^{-2(\lambda+\mu)t}\right) \\
&= \frac{\mu^2+2\lambda\mu}{(\mu+\lambda)^2} + \frac{2\lambda\mu+2\lambda^2-2\lambda\mu}{(\mu+\lambda)^2}e^{-(\lambda+\mu)t} + \frac{\lambda^2-2\lambda^2}{(\mu+\lambda)^2}e^{-2(\lambda+\mu)t} \\
&= \frac{\mu^2+2\lambda\mu}{(\mu+\lambda)^2} + \frac{2\lambda^2}{(\mu+\lambda)^2}e^{-(\lambda+\mu)t} - \frac{\lambda^2}{(\mu+\lambda)^2}e^{-2(\lambda+\mu)t}
\end{aligned}
$$

### Steady-State Probabilities

Let the steady-state probabilities be $\pi_2$, $\pi_1$, and $\pi_0$ for State 2, State 1, and State 0, respectively. These are defined as $\pi_i = \lim_{t\to\infty} P_i(t)$, so we can compute them by taking to the limit the above expressions of $P_0(t)$, $P_1(t)$, and $P_2(t)$, or, we can derive them from the Kolmogorov equation system, by substituting $\frac{dP_i(t)}{dt} = 0$.

The transition rate matrix $Q$ is given by:

$$Q = \begin{pmatrix} -2\lambda & 2\lambda & 0 \\ \mu & -(\mu+\lambda) & \lambda \\ 0 & 2\mu & -2\mu \end{pmatrix}$$

The steady-state probabilities satisfy $\pi Q = \mathbf{0}$ and $\pi_2 + \pi_1 + \pi_0 = 1$, where $\pi = [\pi_2, \pi_1, \pi_0]$. This gives the following system of equations:

$$\begin{aligned} -2\lambda\pi_2 + \mu\pi_1 &= 0 \\ 2\lambda\pi_2 - (\mu+\lambda)\pi_1 + 2\mu\pi_0 &= 0 \\ \lambda\pi_1 - 2\mu\pi_0 &= 0 \\ \pi_2 + \pi_1 + \pi_0 &= 1 \end{aligned}$$

Solving this equation system we derive the steady-state probabilities:

$$\pi_2 = \frac{\mu^2}{(\mu+\lambda)^2} \qquad \pi_1 = \frac{2\lambda\mu}{(\mu+\lambda)^2} \qquad \pi_0 = \frac{\lambda^2}{(\mu+\lambda)^2}$$

Steady-state reliability is $R_{ss} = \pi_2 = \frac{\mu^2}{(\mu+\lambda)^2}$, if the system tolerates no faulty component, or $R_{ss} = \pi_2 + \pi_1 = \frac{\mu^2+2\lambda\mu}{(\mu+\lambda)^2}$, if the system tolerates one faulty component.

## 4.7 Two-Component System with a Dependency

There are situations in which a system is load-sharing and, if a component fails, the other(s) will have to bear a higher load, which will increase their failure rate.

The simplest such system is comprised of two components. Let the four states of the system be:

- State 0: Both components are working.
- State 1: Component 1 has failed, Component 2 is working.
- State 2: Component 2 has failed, Component 1 is working.
- State 3: Both components have failed.

Let $\lambda_1$ be the initial failure rate of Component 1 and $\lambda_2$ be the initial failure rate of Component 2. When Component 2 fails, the failure rate of Component 1 increases to $\lambda_{+1}$. When Component 1 fails, the failure rate of Component 2 increases to $\lambda_{+2}$.

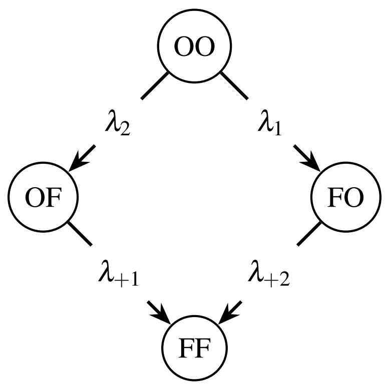



Figure 4.7: State transition graph of a two-component load-sharing system

The probability transition rate matrix **Q** is:

$$\mathbf{Q} = \begin{pmatrix} -(\lambda_1+\lambda_2) & \lambda_1 & \lambda_2 & 0 \\ 0 & -\lambda_{+2} & 0 & \lambda_{+2} \\ 0 & 0 & -\lambda_{+1} & \lambda_{+1} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The Kolmogorov forward equations are:

$$\frac{dP_0(t)}{dt} = -(\lambda_1+\lambda_2)P_0(t)$$
$$\frac{dP_1(t)}{dt} = \lambda_1 P_0(t) - \lambda_{+2}P_1(t)$$
$$\frac{dP_2(t)}{dt} = \lambda_2 P_0(t) - \lambda_{+1}P_2(t)$$
$$\frac{dP_3(t)}{dt} = \lambda_{+2}P_1(t) + \lambda_{+1}P_2(t)$$

With initial conditions $P_0(0) = 1$, $P_1(0) = 0$, $P_2(0) = 0$, and $P_3(0) = 0$.

**Probability of State 0:** $P_0(t)$
The first equation has the solution:

$$P_0(t) = e^{-(\lambda_1+\lambda_2)t}$$

**Probability of State 1:** $P_1(t)$
The second equation is a first-order linear differential equation. The most general solution is when $\lambda_{+2} \neq \lambda_1 + \lambda_2$:

$$P_1(t) = \frac{\lambda_1}{\lambda_{+2} - (\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_{+2}t}\right)$$

**Probability of State 2:** $P_2(t)$
The third equation is also a first-order linear differential equation. The most general solution can be found when $\lambda_{+1} \neq \lambda_1 + \lambda_2$:

$$P_2(t) = \frac{\lambda_2}{\lambda_{+1} - (\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_{+1}t}\right)$$

**Probability of State 3:** $P_3(t)$
The probability of being in State 3 can be found using $P_3(t) = 1 - P_0(t) - P_1(t) - P_2(t)$. We present the general case where $\lambda_{+1} \neq \lambda_1 + \lambda_2$ and $\lambda_{+2} \neq \lambda_1 + \lambda_2$:

$$P_3(t) = 1 - e^{-(\lambda_1+\lambda_2)t} - \frac{\lambda_1}{\lambda_{+2} - (\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_{+2}t}\right) - \frac{\lambda_2}{\lambda_{+1} - (\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_{+1}t}\right)$$

.

## 4.7.1 Reliability

For such a system, the reliability function can be expressed as a sum of probabilities the system is operational, namely $R(t) = P_0(t) + P_1(t) + P_2(t) = e^{-(\lambda_1+\lambda_2)t} + \frac{\lambda_1}{\lambda_{+2}-(\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_{+2}t}\right) + \frac{\lambda_2}{\lambda_{+1}-(\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_{+1}t}\right)$.

We can consider a simplified case, in which the two modules have identical failure rates. Let $\lambda_1 = \lambda_2 = \lambda$ and $\lambda_{+1} = \lambda_{+2} = \lambda_+$. The reliability function $R(t) = P_0(t) + P_1(t) + P_2(t)$ becomes:

$$R(t) = e^{-2\lambda t}\left(1 + \frac{\lambda}{\lambda_+ - 2\lambda} + \frac{\lambda}{\lambda_+ - 2\lambda}\right) - \frac{\lambda}{\lambda_+ - 2\lambda}e^{-\lambda_+ t} - \frac{\lambda}{\lambda_+ - 2\lambda}e^{-\lambda_+ t}$$

$$R(t) = e^{-2\lambda t}\left(1 + \frac{2\lambda}{\lambda_+ - 2\lambda}\right) - \frac{2\lambda}{\lambda_+ - 2\lambda}e^{-\lambda_+ t}$$

$$R(t) = \frac{\lambda_+ - 2\lambda + 2\lambda}{\lambda_+ - 2\lambda}e^{-2\lambda t} - \frac{2\lambda}{\lambda_+ - 2\lambda}e^{-\lambda_+ t}$$

$$R(t) = \frac{\lambda_+}{\lambda_+ - 2\lambda}e^{-2\lambda t} - \frac{2\lambda}{\lambda_+ - 2\lambda}e^{-\lambda_+ t}$$

### 4.7.2 Mean Time To Failure (MTTF)

The MTTF is given by the integral of the reliability function from $t = 0$ to $t = \infty$: $MTTF = \int_0^\infty R(t)dt$. Assuming $\lambda_+ \neq 2\lambda$, we can write:

$$MTTF = \int_0^\infty \left(\frac{\lambda_+}{\lambda_+ - 2\lambda}e^{-2\lambda t} - \frac{2\lambda}{\lambda_+ - 2\lambda}e^{-\lambda_+ t}\right)dt$$

$$MTTF = \frac{\lambda_+}{\lambda_+ - 2\lambda}\int_0^\infty e^{-2\lambda t}dt - \frac{2\lambda}{\lambda_+ - 2\lambda}\int_0^\infty e^{-\lambda_+ t}dt$$

$$MTTF = \frac{\lambda_+}{\lambda_+ - 2\lambda}\left[-\frac{1}{2\lambda}e^{-2\lambda t}\right]_0^\infty - \frac{2\lambda}{\lambda_+ - 2\lambda}\left[-\frac{1}{\lambda_+}e^{-\lambda_+ t}\right]_0^\infty$$

$$MTTF = \frac{\lambda_+}{\lambda_+ - 2\lambda}\left(\frac{1}{2\lambda}\right) - \frac{2\lambda}{\lambda_+ - 2\lambda}\left(\frac{1}{\lambda_+}\right)$$

$$MTTF = \frac{\lambda_+}{2\lambda(\lambda_+ - 2\lambda)} - \frac{2\lambda}{\lambda_+(\lambda_+ - 2\lambda)} = \frac{\lambda_+^2 - 4\lambda^2}{2\lambda\lambda_+(\lambda_+ - 2\lambda)} = \frac{(\lambda_+ - 2\lambda)(\lambda_+ + 2\lambda)}{2\lambda\lambda_+(\lambda_+ - 2\lambda)}$$

$$MTTF = \frac{\lambda_+ + 2\lambda}{2\lambda\lambda_+} = \frac{1}{2\lambda} + \frac{1}{\lambda_+}$$

## 4.8 System with one Cold Standby Backup

We consider a system with a main unit and an identical backup unit. The backup is in a cold standby state, meaning it is not running and cannot fail until the main unit fails. Both units have a failure rate $\lambda$ when they are running.

### 4.8.1 State Definitions

The states of the system are defined as follows:

- **State 0:** The main unit is working, and the backup unit is in a cold standby state.
- **State 1:** The main unit has failed, and the backup unit has started running.
- **State 2:** The backup unit has failed (after taking over from the main unit).

### 4.8.2 Transition Rates

The transitions between these states occur at the following rates:

- From State 0 to State 1: The main unit fails with rate $\lambda$.
- From State 1 to State 2: The backup unit (now running) fails with rate $\lambda$.

State 2 is an absorbing state, representing the failure of the entire system.

#### State Transition Diagram

The Markov process can be represented by the following state transition diagram:

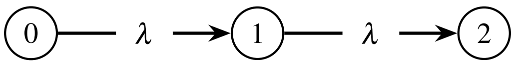



Figure 4.8: State transition graph of a two-component system with cold-sparing back-up

### 4.8.3 Kolmogorov Forward Equations

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. The Kolmogorov forward equations governing the probabilities of being in each state are:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t)$$
$$\frac{dP_1(t)}{dt} = \lambda P_0(t) - \lambda P_1(t)$$
$$\frac{dP_2(t)}{dt} = \lambda P_1(t)$$

The initial conditions are $P_0(0) = 1$, $P_1(0) = 0$, and $P_2(0) = 0$, as the system starts with the main unit working.

We solve these differential equations sequentially.

#### Probability of State 0: $P_0(t)$

The first equation is $\frac{dP_0(t)}{dt} = -\lambda P_0(t)$. This is a simple first-order linear differential equation with the solution:

$$P_0(t) = e^{-\lambda t}$$

#### Probability of State 1: $P_1(t)$

Substituting $P_0(t)$ into the second equation, we get:

$$\frac{dP_1(t)}{dt} + \lambda P_1(t) = \lambda e^{-\lambda t}$$

This is a first-order linear differential equation. The integrating factor is $e^{\int \lambda dt} = e^{\lambda t}$. Multiplying both sides by the integrating factor:

$$e^{\lambda t}\frac{dP_1(t)}{dt} + \lambda e^{\lambda t} P_1(t) = \lambda e^{\lambda t} e^{-\lambda t} = \lambda$$

$$\frac{d}{dt}(e^{\lambda t} P_1(t)) = \lambda$$

Integrating both sides with respect to $t$:

$$e^{\lambda t}P_1(t) = \int \lambda dt = \lambda t + C_1$$

Applying the initial condition $P_1(0) = 0$, we find $C_1 = 0$. Thus,

$$P_1(t) = \lambda t e^{-\lambda t}$$

**Probability of State 2:** $P_2(t)$

Since the sum of the probabilities of all states must be 1, we have $P_0(t) + P_1(t) + P_2(t) = 1$. Therefore,

$$P_2(t) = 1 - P_0(t) - P_1(t)$$

Substituting the expressions for $P_0(t)$ and $P_1(t)$:

$$P_2(t) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}$$

$$P_2(t) = 1 - e^{-\lambda t}(1 + \lambda t)$$

Thus, the probabilities as a function of time for each state of the cold standby system are:

- $P_0(t) = e^{-\lambda t}$
- $P_1(t) = \lambda t e^{-\lambda t}$
- $P_2(t) = 1 - e^{-\lambda t}(1 + \lambda t)$

### 4.8.4 Reliability Function of the Cold Standby System

The reliability function $R(t)$ of the system is the probability that the system is functioning at time $t$. In the context of our cold standby system, the system is functioning if it is in State 0 (main unit working, backup in standby) or State 1 (main unit failed, backup running). The system fails when it enters State 2 (both units failed).

Therefore, the reliability function $R(t)$ is given by the sum of the probabilities of being in State 0 and State 1:

$$R(t) = P_0(t) + P_1(t)$$

From the solutions of the Kolmogorov forward equations, we have:

$$P_0(t) = e^{-\lambda t}$$

$$P_1(t) = \lambda t e^{-\lambda t}$$

Substituting these into the expression for $R(t)$:

$$R(t) = e^{-\lambda t} + \lambda t e^{-\lambda t}$$

$$R(t) = e^{-\lambda t}(1 + \lambda t)$$

Thus, the reliability function of the cold standby system with identical main and backup units (failure rate $\lambda$) is:

$$R(t) = e^{-\lambda t}(1 + \lambda t)$$

## 4.9 System with a Warm Standby Backup

We consider a system with a main unit and an identical backup unit. The main unit has a failure rate $\lambda$. The backup unit, when in warm standby, has a failure rate $\lambda_s$ ($\lambda_s < \lambda$). When the main unit fails, the backup unit starts running with a failure rate $\lambda$.

### 4.9.1 State Definitions

The states of the system are defined as follows:

- **State 0:** The main unit is working, and the backup unit is in warm standby (failure rate $\lambda_s$).
- **State 1:** The main unit has failed, and the backup unit is running (failure rate $\lambda$).
- **State 2:** Both units have failed.

### 4.9.2 Transition Rates

The transitions between these states occur at the following rates:

- From State 0 to State 1: The main unit fails with rate $\lambda$.
- From State 0 to State 2: The backup unit fails in standby with rate $\lambda_s$.
- From State 1 to State 2: The backup unit fails while running with rate $\lambda$.

State 2 is an absorbing state, representing the failure of the entire system.

### 4.9.3 State Transition Diagram

The Markov process can be represented by the following state transition diagram:

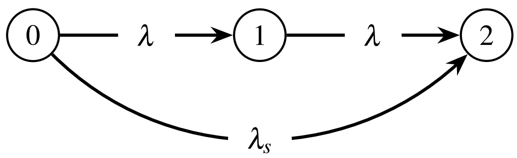



Figure 4.9: State transition graph of a two-component system with warm-sparing back-up

### 4.9.4 Kolmogorov Forward Equations

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. The Kolmogorov forward equations are:

$$\frac{dP_0(t)}{dt} = -(\lambda + \lambda_s)P_0(t)$$
$$\frac{dP_1(t)}{dt} = \lambda P_0(t) - \lambda P_1(t)$$
$$\frac{dP_2(t)}{dt} = \lambda_s P_0(t) + \lambda P_1(t)$$

With initial conditions $P_0(0) = 1$, $P_1(0) = 0$, and $P_2(0) = 0$.

### 4.9.5 Probabilities as a Function of Time

We solve these differential equations sequentially.

**Probability of State 0:** $P_0(t)$

The first equation is $\frac{dP_0(t)}{dt} = -(\lambda + \lambda_s)P_0(t)$, which has the solution:

$$P_0(t) = e^{-(\lambda+\lambda_s)t}$$

**Probability of State 1:** $P_1(t)$

Substitute $P_0(t)$ into the second equation:

$$\frac{dP_1(t)}{dt} + \lambda P_1(t) = \lambda e^{-(\lambda+\lambda_s)t}$$

This is a first-order linear differential equation. The integrating factor is $e^{\int \lambda dt} = e^{\lambda t}$. Multiplying by the integrating factor:

$$e^{\lambda t}\frac{dP_1(t)}{dt} + \lambda e^{\lambda t}P_1(t) = \lambda e^{\lambda t}e^{-(\lambda+\lambda_s)t} = \lambda e^{-\lambda_s t}$$

$$\frac{d}{dt}(e^{\lambda t}P_1(t)) = \lambda e^{-\lambda_s t}$$

Integrating both sides with respect to $t$:

$$e^{\lambda t}P_1(t) = \int \lambda e^{-\lambda_s t} dt = -\frac{\lambda}{\lambda_s}e^{-\lambda_s t} + C_1$$

Using $P_1(0) = 0$, we find $C_1 = \frac{\lambda}{\lambda_s}$. Thus,

$$P_1(t) = e^{-\lambda t}\left(\frac{\lambda}{\lambda_s} - \frac{\lambda}{\lambda_s}e^{-\lambda_s t}\right) = \frac{\lambda}{\lambda_s}(e^{-\lambda t} - e^{-(\lambda+\lambda_s)t})$$

**Probability of State 2:** $P_2(t)$

Since the sum of the probabilities must equal 1: $P_0(t) + P_1(t) + P_2(t) = 1$. Therefore,

$$P_2(t) = 1 - P_0(t) - P_1(t)$$

$$P_2(t) = 1 - e^{-(\lambda+\lambda_s)t} - \frac{\lambda}{\lambda_s}(e^{-\lambda t} - e^{-(\lambda+\lambda_s)t})$$

$$P_2(t) = 1 - e^{-(\lambda+\lambda_s)t}\left(1 - \frac{\lambda}{\lambda_s}\right) - \frac{\lambda}{\lambda_s}e^{-\lambda t}$$

$$P_2(t) = 1 + e^{-(\lambda+\lambda_s)t}\left(\frac{\lambda - \lambda_s}{\lambda_s}\right) - \frac{\lambda}{\lambda_s}e^{-\lambda t}$$

$$P_2(t) = 1 - \frac{\lambda}{\lambda_s}e^{-\lambda t} + \frac{\lambda - \lambda_s}{\lambda_s}e^{-(\lambda+\lambda_s)t}$$

These are the probabilities as a function of time for each state of the system with a warm standby backup.

### 4.9.6 Reliability Function of the Warm Standby System

The reliability function $R(t)$ of the system is the probability that the system is functioning at time $t$. For a warm standby system, the system is functioning if it is in State 0 (main unit working, backup in warm standby) or State 1 (main unit failed, backup running). The system fails when it enters State 2 (both units failed).

Therefore, the reliability function $R(t)$ is given by the sum of the probabilities of being in State 0 and State 1:

$$R(t) = P_0(t) + P_1(t)$$

From the solutions of the Kolmogorov forward equations, we have:

$$P_0(t) = e^{-(\lambda+\lambda_s)t}$$

$$P_1(t) = \frac{\lambda}{\lambda_s}(e^{-\lambda t} - e^{-(\lambda+\lambda_s)t})$$

Substituting these into the expression for $R(t)$:

$$R(t) = e^{-(\lambda+\lambda_s)t} + \frac{\lambda}{\lambda_s}(e^{-\lambda t} - e^{-(\lambda+\lambda_s)t})$$

$$R(t) = e^{-(\lambda+\lambda_s)t} + \frac{\lambda}{\lambda_s}e^{-\lambda t} - \frac{\lambda}{\lambda_s}e^{-(\lambda+\lambda_s)t}$$

$$R(t) = \frac{\lambda}{\lambda_s}e^{-\lambda t} + e^{-(\lambda+\lambda_s)t}\left(1 - \frac{\lambda}{\lambda_s}\right)$$

$$R(t) = \frac{\lambda}{\lambda_s}e^{-\lambda t} + e^{-(\lambda+\lambda_s)t}\left(\frac{\lambda_s - \lambda}{\lambda_s}\right)$$

Thus, the reliability function is:

$$R(t) = \frac{\lambda}{\lambda_s}e^{-\lambda t} + \frac{\lambda_s - \lambda}{\lambda_s}e^{-(\lambda+\lambda_s)t}$$

## 4.10 Cold Standby Backup System with Repair

We consider a system with a main unit and an identical cold standby backup unit. Both units have a failure rate $\lambda$ when operating and a repair rate $\mu$ when failed. There is only one repairman available, so two systems cannot be repaired in parallel.

### 4.10.1 State Definitions

- **State 0:** Main unit is working, backup unit is in cold standby.
- **State 1:** Main unit has failed, backup unit is working.
- **State 2:** Both main and backup units have failed, and one unit is under repair.

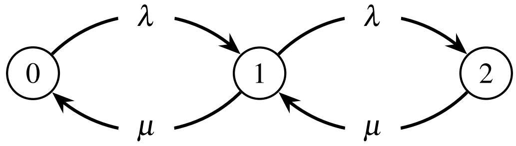



Figure 4.10: State transition graph of a two-component system with cold-sparing back-up with a single repairman

### 4.10.2 Kolmogorov Equation System

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. The Kolmogorov forward equations are:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t)$$
$$\frac{dP_1(t)}{dt} = \lambda P_0(t) - (\lambda + \mu)P_1(t) + \mu P_2(t)$$
$$\frac{dP_2(t)}{dt} = \lambda P_1(t) - \mu P_2(t)$$

With the normalization condition $P_0(t) + P_1(t) + P_2(t) = 1$.

### 4.10.3 Steady-State Probabilities

In steady state, the time derivatives are zero:

$$0 = -\lambda P_0 + \mu P_1$$
$$0 = \lambda P_0 - (\lambda + \mu)P_1 + \mu P_2$$
$$0 = \lambda P_1 - \mu P_2$$

From the first equation, we have:

$$\lambda P_0 = \mu P_1 \implies P_0 = \frac{\mu}{\lambda} P_1$$

From the third equation, we have:

$$\lambda P_1 = \mu P_2 \implies P_2 = \frac{\lambda}{\mu} P_1$$

Now, we use the normalization condition $P_0 + P_1 + P_2 = 1$:

$$\frac{\mu}{\lambda} P_1 + P_1 + \frac{\lambda}{\mu} P_1 = 1$$

$$P_1 \left( \frac{\mu}{\lambda} + 1 + \frac{\lambda}{\mu} \right) = 1$$

$$P_1 \left( \frac{\mu^2 + \lambda\mu + \lambda^2}{\lambda\mu} \right) = 1$$

$$P_1 = \frac{\lambda\mu}{\lambda^2+\lambda\mu+\mu^2}$$

Now we can find $P_0$ and $P_2$:

$$P_0 = \frac{\mu}{\lambda}P_1 = \frac{\mu}{\lambda}\frac{\lambda\mu}{\lambda^2+\lambda\mu+\mu^2} = \frac{\mu^2}{\lambda^2+\lambda\mu+\mu^2}$$

$$P_2 = \frac{\lambda}{\mu}P_1 = \frac{\lambda}{\mu}\frac{\lambda\mu}{\lambda^2+\lambda\mu+\mu^2} = \frac{\lambda^2}{\lambda^2+\lambda\mu+\mu^2}$$

The steady-state probabilities for each state are:

$$P_0 = \frac{\mu^2}{\lambda^2+\lambda\mu+\mu^2}$$

$$P_1 = \frac{\lambda\mu}{\lambda^2+\lambda\mu+\mu^2}$$

$$P_2 = \frac{\lambda^2}{\lambda^2+\lambda\mu+\mu^2}$$

### 4.10.4 Steady-State Availability

The steady-state availability of the system is the probability that the system is in an operational state. In this case, the system is operational in State 0 (Main working, Backup standby) and State 1 (Main failed, Backup working). The steady-state probabilities for these states are:

$$P_0 = \frac{\mu^2}{\lambda^2+\lambda\mu+\mu^2}$$

$$P_1 = \frac{\lambda\mu}{\lambda^2+\lambda\mu+\mu^2}$$

The steady-state availability $A_{ss}$ is the sum of these probabilities:

$$A_{ss} = P_0 + P_1 = \frac{\mu^2}{\lambda^2+\lambda\mu+\mu^2} + \frac{\lambda\mu}{\lambda^2+\lambda\mu+\mu^2}$$

$$A_{ss} = \frac{\mu^2+\lambda\mu}{\lambda^2+\lambda\mu+\mu^2}$$

$$A_{ss} = \frac{\mu(\mu+\lambda)}{\lambda^2+\lambda\mu+\mu^2}$$

## 4.11 Cold Standby System with Two Backups

We consider a system with a main unit and two identical cold standby backup units, all having a failure rate $\lambda$ when running. There is no repair in this system.

### 4.11.1 State Definitions

The states of the system are defined based on the number of failed units:

- **State 0:** 0 failed units (Main working, Backup 1 standby, Backup 2 standby).
- **State 1:** 1 failed unit (Main failed, Backup 1 running, Backup 2 standby).
- **State 2:** 2 failed units (Main failed, Backup 1 failed, Backup 2 running).
- **State 3:** 3 failed units (Main failed, Backup 1 failed, Backup 2 failed).

The transitions between these states occur at a rate $\lambda$ when a running unit fails:

- From State 0 to State 1: The main unit fails (rate $\lambda$).
- From State 1 to State 2: Backup 1 fails (rate $\lambda$).
- From State 2 to State 3: Backup 2 fails (rate $\lambda$).

State 3 is an absorbing state, representing the failure of the entire system.

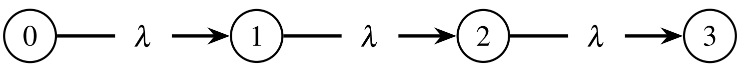



Figure 4.11: State transition graph of a three-component system with cold-sparing back-up

### 4.11.2 Kolmogorov Forward Equations

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. The Kolmogorov forward equations are:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t)$$
$$\frac{dP_1(t)}{dt} = \lambda P_0(t) - \lambda P_1(t)$$
$$\frac{dP_2(t)}{dt} = \lambda P_1(t) - \lambda P_2(t)$$
$$\frac{dP_3(t)}{dt} = \lambda P_2(t)$$

With initial conditions $P_0(0) = 1$, $P_1(0) = 0$, $P_2(0) = 0$, and $P_3(0) = 0$.

### 4.11.3 Probabilities as a Function of Time

Solving these differential equations sequentially:

1. From the first equation:
$$P_0(t) = e^{-\lambda t}$$

2. Substituting $P_0(t)$ into the second equation:
$$\frac{dP_1(t)}{dt} + \lambda P_1(t) = \lambda e^{-\lambda t}$$
Solving this first-order linear ODE yields:
$$P_1(t) = \lambda t e^{-\lambda t}$$

3. Substituting $P_1(t)$ into the third equation:

$$\frac{dP_2(t)}{dt} + \lambda P_2(t) = \lambda(\lambda t e^{-\lambda t}) = \lambda^2 t e^{-\lambda t}$$

Solving this first-order linear ODE yields:

$$P_2(t) = \frac{1}{2}\lambda^2 t^2 e^{-\lambda t}$$

4. From the fact that $P_0(t)+P_1(t)+P_2(t)+P_3(t) = 1$, we have:

$$P_3(t) = 1 - P_0(t) - P_1(t) - P_2(t) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} - \frac{1}{2}\lambda^2 t^2 e^{-\lambda t}$$

$$P_3(t) = 1 - e^{-\lambda t}\left(1 + \lambda t + \frac{1}{2}(\lambda t)^2\right)$$

### 4.11.4 Reliability Function

The reliability function $R(t)$ is the probability that the system is functioning at time $t$, which means the system is in State 0, State 1, or State 2 (less than 3 failed units).

$$R(t) = P_0(t) + P_1(t) + P_2(t)$$

$$R(t) = e^{-\lambda t} + \lambda t e^{-\lambda t} + \frac{1}{2}\lambda^2 t^2 e^{-\lambda t}$$

$$R(t) = e^{-\lambda t}\left(1 + \lambda t + \frac{1}{2}(\lambda t)^2\right)$$

## 4.12 2-out-of-3 Majority Voting System

We consider a system with three identical modules, each having a failure rate $\lambda$. The system employs a 2-out-of-3 majority-voting scheme, meaning the system is functional if at least two of the three modules are working.

### 4.12.1 State Definitions

The states of the system are defined based on the number of failed modules:

- **State 0:** 0 failed modules (3 working).
- **State 1:** 1 failed module (2 working).
- **State 2:** 2 failed modules (1 working).
- **State 3:** 3 failed modules (0 working).

The transitions between these states occur based on the failure of the working modules:

- From State 0 to State 1: Any of the 3 working modules can fail (rate $3\lambda$).
- From State 1 to State 2: Any of the 2 working modules can fail (rate $2\lambda$).
- From State 2 to State 3: The remaining 1 working module can fail (rate $\lambda$).

State 3 is an absorbing state, representing the failure of the entire system (less than 2 working modules).

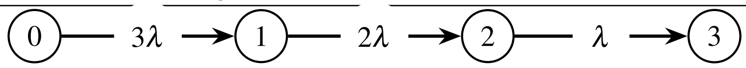



Figure 4.12: State transition graph of a two-out-of-three component system

### 4.12.2 Kolmogorov Forward Equations

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. The Kolmogorov forward equations are:

$$\frac{dP_0(t)}{dt} = -3\lambda P_0(t)$$
$$\frac{dP_1(t)}{dt} = 3\lambda P_0(t) - 2\lambda P_1(t)$$
$$\frac{dP_2(t)}{dt} = 2\lambda P_1(t) - \lambda P_2(t)$$
$$\frac{dP_3(t)}{dt} = \lambda P_2(t)$$

With initial conditions $P_0(0) = 1$, $P_1(0) = 0$, $P_2(0) = 0$, and $P_3(0) = 0$.

### 4.12.3 Probabilities as a Function of Time

Solving these differential equations sequentially:

1. From the first equation:
$$P_0(t) = e^{-3\lambda t}$$

2. Substituting $P_0(t)$ into the second equation:
$$\frac{dP_1(t)}{dt} + 2\lambda P_1(t) = 3\lambda e^{-3\lambda t}$$
Solving this first-order linear ODE yields:
$$P_1(t) = 3e^{-2\lambda t} - 3e^{-3\lambda t}$$

3. Substituting $P_1(t)$ into the third equation:
$$\frac{dP_2(t)}{dt} + \lambda P_2(t) = 2\lambda(3e^{-2\lambda t} - 3e^{-3\lambda t}) = 6\lambda e^{-2\lambda t} - 6\lambda e^{-3\lambda t}$$
Solving this first-order linear ODE yields:
$$P_2(t) = 3e^{-\lambda t} - 6e^{-2\lambda t} + 3e^{-3\lambda t}$$

4. Using $P_0(t) + P_1(t) + P_2(t) + P_3(t) = 1$:
$$P_3(t) = 1 - P_0(t) - P_1(t) - P_2(t) = 1 - e^{-3\lambda t} - (3e^{-2\lambda t} - 3e^{-3\lambda t}) - (3e^{-\lambda t} - 6e^{-2\lambda t} + 3e^{-3\lambda t})$$
$$P_3(t) = 1 - 3e^{-\lambda t} + 3e^{-2\lambda t} - e^{-3\lambda t} = (1 - e^{-\lambda t})^3$$

### 4.12.4 Reliability Function

The system is functional if at least two modules are working, which means 0 or 1 module has failed. This corresponds to State 0 and State 1.

$$R(t) = P_0(t) + P_1(t)$$
$$R(t) = e^{-3\lambda t} + (3e^{-2\lambda t} - 3e^{-3\lambda t})$$
$$R(t) = 3e^{-2\lambda t} - 2e^{-3\lambda t}$$

## 4.13 2-out-of-3 Majority Voting System with Repair

We consider a system with three identical modules, each having a failure rate $\lambda$ when working and a repair rate $\mu$ when failed. The system employs a 2-out-of-3 majority-voting scheme.

### 4.13.1 State Definitions

The states of the system are defined based on the number of failed modules:

- **State 0:** 0 failed modules (3 working).
- **State 1:** 1 failed module (2 working).
- **State 2:** 2 failed modules (1 working).
- **State 3:** 3 failed modules (0 working).

The transitions between these states occur due to failures and repairs:

- From State 0 to State 1: Failure of one of the 3 working modules (rate $3\lambda$).
- From State 1 to State 0: Repair of the 1 failed module (rate $\mu$).
- From State 1 to State 2: Failure of one of the 2 working modules (rate $2\lambda$).
- From State 2 to State 1: Repair of one of the 2 failed modules (rate $2\mu$).
- From State 2 to State 3: Failure of the remaining 1 working module (rate $\lambda$).
- From State 3 to State 2: Repair of one of the 3 failed modules (rate $3\mu$).

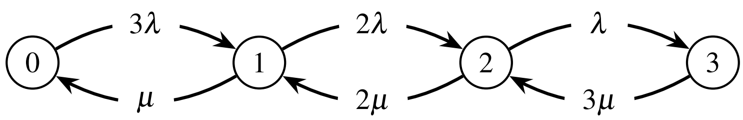



Figure 4.13: State transition graph of a 2-out-of-3 majority vote system with repair

### 4.13.2 Kolmogorov Forward Equations

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. The Kolmogorov forward equations are:

$$\frac{dP_0(t)}{dt} = -3\lambda P_0(t) + \mu P_1(t)$$
$$\frac{dP_1(t)}{dt} = 3\lambda P_0(t) - (2\lambda + \mu)P_1(t) + 2\mu P_2(t)$$
$$\frac{dP_2(t)}{dt} = 2\lambda P_1(t) - (\lambda + 2\mu)P_2(t) + 3\mu P_3(t)$$
$$\frac{dP_3(t)}{dt} = \lambda P_2(t) - 3\mu P_3(t)$$

With initial conditions $P_0(0) = 1$, $P_1(0) = 0$, $P_2(0) = 0$, and $P_3(0) = 0$.

### 4.13.3 State Probabilities

To find the state probabilities $P_i(t)$, this system of linear differential equations with constant coefficients needs to be solved. This typically involves finding the eigenvalues and eigenvectors of the transition rate matrix:

$$Q = \begin{pmatrix} -3\lambda & \mu & 0 & 0 \\ 3\lambda & -(2\lambda+\mu) & 2\mu & 0 \\ 0 & 2\lambda & -(\lambda+2\mu) & 3\mu \\ 0 & 0 & \lambda & -3\mu \end{pmatrix}$$

The solution $P(t) = [P_0(t), P_1(t), P_2(t), P_3(t)]^T$ can be expressed as $P(t) = P(0)e^{Qt}$. Finding the eigenvalues and eigenvectors of this $4 \times 4$ matrix and then computing the matrix exponential is a complex task.

### 4.13.4 Reliability Function

The system is functional if at least two modules are working, which corresponds to State 0 (3 working) and State 1 (2 working). The reliability function $R(t)$ is given by:

$$R(t) = P_0(t) + P_1(t)$$

To obtain an explicit analytical form for $R(t)$, one would need to solve the system of differential equations. This often requires methods like Laplace transforms or matrix diagonalization, which can be quite involved for a system of this size. Numerical methods can be used to find the probabilities at specific times for given values of $\lambda$ and $\mu$.

### 4.13.5 Steady-State Reliability of the 2-out-of-3 Majority Voting System with Repair

In steady state, the probabilities $\pi_i$ are constant, so their time derivatives are zero. The Kolmogorov forward equations become:

$$0 = -3\lambda\pi_0 + \mu\pi_1 \tag{4.46}$$
$$0 = 3\lambda\pi_0 - (2\lambda+\mu)\pi_1 + 2\mu\pi_2 \tag{4.47}$$
$$0 = 2\lambda\pi_1 - (\lambda+2\mu)\pi_2 + 3\mu\pi_3 \tag{4.48}$$
$$0 = \lambda\pi_2 - 3\mu\pi_3 \tag{4.49}$$

With the normalization condition:

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

By solving we can find the steady-state probabilities for the operational states $\pi_0$ and $\pi_1$:

$$\pi_1 = \frac{3\lambda\mu^2}{(\lambda+\mu)^3} \qquad \pi_0 = \frac{\mu^3}{(\lambda+\mu)^3}$$

The steady-state reliability $R_{ss}$ is the probability that the system is in State 0 (3 working modules) or State 1 (2 working modules):

$$R_{ss} = \pi_0 + \pi_1 = \frac{\mu^3}{(\lambda+\mu)^3} + \frac{3\lambda\mu^2}{(\lambda+\mu)^3} = \frac{\mu^2(\mu+3\lambda)}{(\lambda+\mu)^3}$$

## 4.14 Three-State System with Degraded Mode

We consider a system with three states: Fully Operational (State 1), Degraded State (State 2), and Failed State (State 3). There is no repair, so the transition rates are:

- $\lambda_1$: Rate from State 1 to State 2
- $\lambda_2$: Rate from State 1 to State 3
- $\lambda_3$: Rate from State 2 to State 3

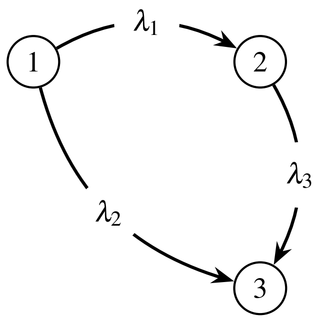



Figure 4.14: State transition graph of a three-state system with degraded mode

### 4.14.1 Kolmogorov Equation System

Let $P_i(t)$ be the probability that the system is in state $i$ at time $t$. The Kolmogorov forward equations are:

$$\frac{dP_1(t)}{dt} = -(\lambda_1 + \lambda_2)P_1(t)$$
$$\frac{dP_2(t)}{dt} = \lambda_1 P_1(t) - \lambda_3 P_2(t)$$
$$\frac{dP_3(t)}{dt} = \lambda_2 P_1(t) + \lambda_3 P_2(t)$$

With initial conditions $P_1(0) = 1$, $P_2(0) = 0$, $P_3(0) = 0$.

### 4.14.2 Probabilities for Each State

Solving the system of equations:

1. From the first equation:
$$P_1(t) = e^{-(\lambda_1+\lambda_2)t}$$

2. Substituting $P_1(t)$ into the second equation:
$$\frac{dP_2(t)}{dt} + \lambda_3 P_2(t) = \lambda_1 e^{-(\lambda_1+\lambda_2)t}$$

   Solving this linear first-order ODE, we get:
$$P_2(t) = \begin{cases} \frac{\lambda_1}{\lambda_3-(\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_3 t}\right) & \text{if } \lambda_3 \neq \lambda_1 + \lambda_2 \\ \lambda_1 t e^{-(\lambda_1+\lambda_2)t} & \text{if } \lambda_3 = \lambda_1 + \lambda_2 \end{cases}$$

3. Using $P_1(t)+P_2(t)+P_3(t)=1$:

$$P_3(t) = 1 - P_1(t) - P_2(t)$$

We assume $\lambda_3 \neq \lambda_1+\lambda_2$:

$$P_3(t) = 1 - e^{-(\lambda_1+\lambda_2)t} - \frac{\lambda_1}{\lambda_3-(\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_3 t}\right)$$

### 4.14.3 Reliability

The reliability function $R(t)$ is the probability that the system is in State 1 (fully operational) or State 2 (degraded). Assuming $\lambda_3 \neq \lambda_1+\lambda_2$ we can derive:

$$R(t) = P_1(t)+P_2(t) = e^{-(\lambda_1+\lambda_2)t} + \frac{\lambda_1}{\lambda_3-(\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_3 t}\right)$$

### 4.14.4 Mean Time To Failure (MTTF)

The MTTF is the expected time until the system enters the failed state (State 3). We can calculate this as the integral of the reliability function from 0 to $\infty$.

$$\begin{aligned} MTTF &= \int_0^\infty \left[e^{-(\lambda_1+\lambda_2)t} + \frac{\lambda_1}{\lambda_3-(\lambda_1+\lambda_2)}\left(e^{-(\lambda_1+\lambda_2)t} - e^{-\lambda_3 t}\right)\right] dt \\ &= \left[\frac{e^{-(\lambda_1+\lambda_2)t}}{-(\lambda_1+\lambda_2)}\right]_0^\infty + \frac{\lambda_1}{\lambda_3-(\lambda_1+\lambda_2)}\left[\frac{e^{-(\lambda_1+\lambda_2)t}}{-(\lambda_1+\lambda_2)} - \frac{e^{-\lambda_3 t}}{-\lambda_3}\right]_0^\infty \\ &= \frac{1}{\lambda_1+\lambda_2} + \frac{\lambda_1}{\lambda_3-(\lambda_1+\lambda_2)}\left(\frac{1}{\lambda_1+\lambda_2} - \frac{1}{\lambda_3}\right) \end{aligned}$$

## 4.15 Markov Model for a Component with Three States

Let us consider a component that can be in one of three states: Operating (State 0), Failed Open (State 1), and Failed Short (State 2). There is no repair operation available.

- **State 0:** Operating
- **State 1:** Failed Open
- **State 2:** Failed Short

The transition rates from the operating state to the failed states are:

- $\lambda_{oo}$: Failure rate from Operating (State 0) to Failed Open (State 1)
- $\lambda_{os}$: Failure rate from Operating (State 0) to Failed Short (State 2)

Since there is no repair, there are no transitions out of the failed states (State 1 and State 2).

### 4.15.1 Kolmogorov Equation System

Let $P_i(t)$ be the probability that the component is in state $i$ at time $t$. The Kolmogorov forward equations are:

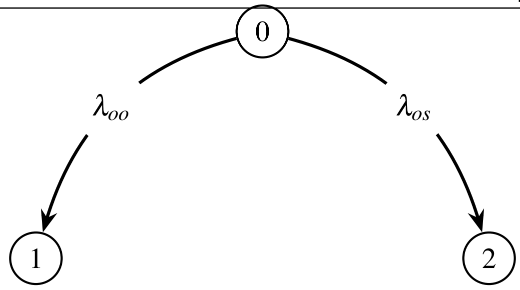



Figure 4.15: State transition graph of a component with three states

$$\frac{dP_0(t)}{dt} = -(\lambda_{oo} + \lambda_{os})P_0(t)$$
$$\frac{dP_1(t)}{dt} = \lambda_{oo}P_0(t)$$
$$\frac{dP_2(t)}{dt} = \lambda_{os}P_0(t)$$

With initial conditions $P_0(0) = 1$, $P_1(0) = 0$, $P_2(0) = 0$.

### 4.15.2 Probabilities for Each State

Solving the system of differential equations:

1. From the first equation:
$$P_0(t) = e^{-(\lambda_{oo}+\lambda_{os})t}$$
2. Substituting $P_0(t)$ into the second equation:
$$\frac{dP_1(t)}{dt} = \lambda_{oo}e^{-(\lambda_{oo}+\lambda_{os})t}$$
Integrating with respect to $t$:
$$P_1(t) = \int_0^t \lambda_{oo}e^{-(\lambda_{oo}+\lambda_{os})\tau}d\tau = \frac{\lambda_{oo}}{\lambda_{oo}+\lambda_{os}}\left[-e^{-(\lambda_{oo}+\lambda_{os})\tau}\right]_0^t = \frac{\lambda_{oo}}{\lambda_{oo}+\lambda_{os}}\left(1-e^{-(\lambda_{oo}+\lambda_{os})t}\right)$$
3. Substituting $P_0(t)$ into the third equation:
$$\frac{dP_2(t)}{dt} = \lambda_{os}e^{-(\lambda_{oo}+\lambda_{os})t}$$
Integrating with respect to $t$:
$$P_2(t) = \int_0^t \lambda_{os}e^{-(\lambda_{oo}+\lambda_{os})\tau}d\tau = \frac{\lambda_{os}}{\lambda_{oo}+\lambda_{os}}\left[-e^{-(\lambda_{oo}+\lambda_{os})\tau}\right]_0^t = \frac{\lambda_{os}}{\lambda_{oo}+\lambda_{os}}\left(1-e^{-(\lambda_{oo}+\lambda_{os})t}\right)$$

### 4.15.3 Reliability

The reliability $R(t)$ of the component is the probability that it is still in the operating state (State 0).

$$R(t) = P_0(t) = e^{-(\lambda_{oo}+\lambda_{os})t}$$

### 4.15.4 Mean Time To Failure (MTTF)

The MTTF is the expected time until the component fails (enters State 1 or State 2). This can be calculated by integrating the reliability function from 0 to $\infty$:

$$MTTF = \int_0^\infty R(t)dt = \int_0^\infty e^{-(\lambda_{oo}+\lambda_{os})t}dt$$

Let $\lambda = \lambda_{oo} + \lambda_{os}$.

$$MTTF = \int_0^\infty e^{-\lambda t}dt = \left[-\frac{1}{\lambda}e^{-\lambda t}\right]_0^\infty = 0 - \left(-\frac{1}{\lambda}e^0\right) = \frac{1}{\lambda} = \frac{1}{\lambda_{oo}+\lambda_{os}}$$