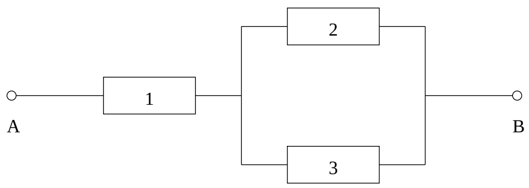
Chapter 7

Analytical Manipulations in Bioinformatics

7.1 Biological Systems

We will study biological systems that in a simple case might look like in Figure 7.1 below:

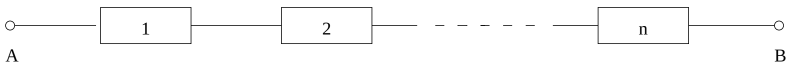
Figure 7.1. *A simple biological system.*



In the biological system in Figure 7.1, the *biological component* number i may be *healthy*, for $i = 1, 2, \ldots$: Otherwise it is *dead* or *unhealthy.* The biological system is healthy if there is a path from point A to point B, which only passes healthy components. Otherwise it is unhealthy.

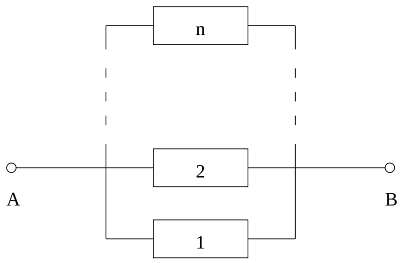
Example 7.1. *The biological* series coupling *in Figure 7.2 is healthy if and only if all its biological components are healthy.*

Figure 7.2. *Biological series coupling.*



Example 7.2. *The biological* parallell coupling *in Figure 7.3 below is healthy if and only if at least one of its biological components are healthy.*

Figure 7.3. *Biological parallell coupling.*



Every biological system can be built be means of a finite number of biological series couplings and biological parallell couplings. See Figures 7.4 and 7.5 below for a simple example of how this works in practical biological applications.

We will study biological systems, the biological components number $i = 1, 2, \ldots$ of which are healthy with certain *health probabilities* $p_1, p_2, \ldots$. Unless otherwise is stated, the biological components of a biological system are assumem to be stochastically independent of each other.

A main characteristic for a biological system is its *health probability.*

Example 7.3. *The health probability for the biological system in Figure 7.4 below is*

$$p_1\,[1-(1-p_2)\,(1-p_3)],$$

because it is a biological series coupling of two biological systems, the health probabilities of which are p_1 and $1-(1-p_2)\,(1-p_3)$, respectively.

Figure 7.4. *A simple biological system with health probabilities.*

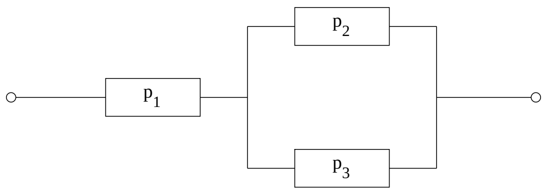
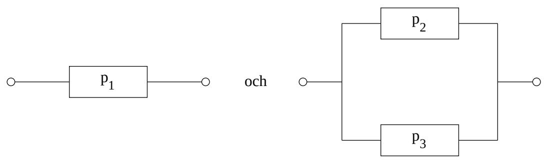


Figure 7.5. *Composition of a simple biological system as a biological series coupling of a single component with a biological parallell coupling.*



The biological components number $i = 1, 2, \ldots$ of a biological system have certain *biological life lengths* $T_1, T_2, \ldots$. The biological life lenghts are modeled as random variables, that are mutually independent, unless otherwise is stated.

The relation between the health probability and the biological life length T_i of biological component number i is the following:

$$p_i = p_i(t) = \mathbf{P}\left\{\text{biological component number } i \text{ is healthy at time }\right\} = \mathbf{P}\{T_i > t\}.$$

The distribution function of the biological life lengths $T_1, T_2, \ldots$ are assume to be continuous, unless otherwise is stated.

Definition 7.1. The *biological survival function* R_T of a bilogical system with biological life length T is given by

$$R_T(t) = \mathbf{P}\{\text{the bilogiocal system is healthy at time } t\} = \mathbf{P}\{T > t\} \quad \text{for } t > 0.$$

Notice that, for a bilogical system with biological life length T with distribution function $F_T(t) = \mathbf{P}\{T \le t\}$, we have

$$R_T(t) = 1 - F_T(t).$$

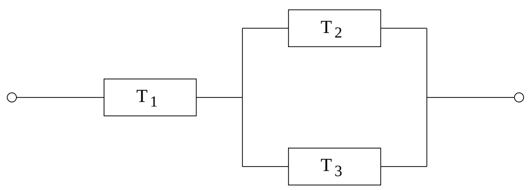
Example 7.4. *The biological life length of the biological system system in Figure 7.6 below is given by*

$$T = \min\{T_1, \max[T_2, T_3]\}.$$

See also Example 7.3. Hence the biological survival function of the the biological system is given by

$$\begin{aligned} R_T(t) = \mathbf{P}\{T > t\} &= \mathbf{P}\{T_1 > t\}\left(1 - (1 - \mathbf{P}\{T_2 > t\})\left(1 - \mathbf{P}\{T_3 > t\}\right)\right) \\ &= R_{T_1}(t)\left(1 - (1 - R_{T_2}(t))(1 - R_{T_3}(t))\right). \end{aligned}$$

Figure 7.6. *A simple biological system with biological life lengths.*



Definition 7.2. The *biological death intensity* r_T of a biological system with biological life length T is given by

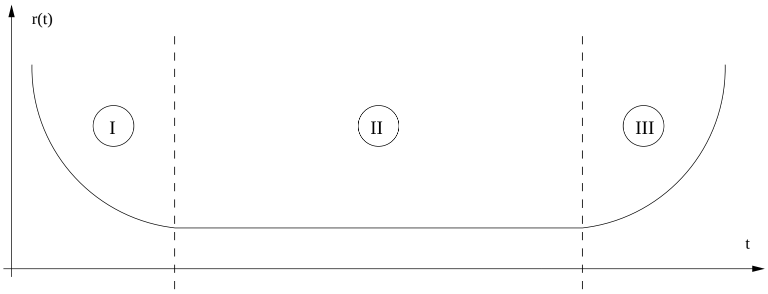
$$r_T(t) = -\frac{d}{dt}\ln\left(R_T(t)\right) \quad \text{for } t > 0.$$

Definition 7.3. A biological system with biological life length T has *Increasing Failure Rate*, IFR, if the biological death intensity is increasing $r'(t) \ge 0$.

A biological system with biological life length T has *Decreasing Failure Rate*, DFR, if the biological death intensity is decreasing $r'(t) \le 0$

Example 7.5. *It is quite common that biological death intensities are neither IFR or DFR, but instead follow a so called bath tub curve, the principal appearance of which is diplayed in Figur 7.7 below. Here region I corresponds to an early phase with a comperatively high probability of unhealth (as foer exampl, for children), the region II corresponds to a component that has survived these early hazards, and has settled at a more normal death intensity, while region III corresponds to an aged biological component, where the death intensity increases with accumlated age, leading to safe eventual death.*

Figure 7.7. *A biological death intensity with a bath tub shape.*



7.2 More on Biological Systems

The following theorem explains that the biological death intensity really is the (infinite-seimal) intensity at which death (sickness, unhealth, ..) occur:

Theorem 7.1. *For a biological life length T with biological death intensity r_T, we have*

$$\mathbf{P}\{T \le t+h | T > t\} = r_T(t)h + o(h) \quad \text{as } h \downarrow 0.$$

Proof. Writing F_T and f_T for the distribution function and probability density function of T, respectively, we have

$$\begin{aligned}\mathbf{P}\{T \le t+h | T > t\} &= \frac{F_T(t+h) - F_T(t)}{1 - F_T(t)} \\ &= \frac{f_T(t)h + o(h)}{R_T(t)} = -\frac{d}{dt}\ln(R_T(t))h + o(h) = r_T(t)h + o(h). \quad \square\end{aligned}$$

Theorem 7.2. *A function $r : (0, \infty) \to [0, \infty)$ is a biological death intensity if and only if*

$$\int_0^\infty r(t)dt = \infty.$$

In that case the corresponding bilogical survival function is given by

$$R(t) = \exp\left\{-\int_0^t r(s)ds\right\}.$$

Proof. *If r is a biological death intensity of a bilogical system with biological survival function R_T, then a differentiation of the function*

$$R(t) = \exp\Big\{-\int_0^t r(s)ds\Big\}$$

gives

$$-\frac{d}{dt}\ln(R(t)) = -\frac{-r(t)\exp\Big\{-\int_0^t r(s)ds\Big\}}{\exp\Big\{-\int_0^t r(s)ds\Big\}} = r(t).$$

As the function $-\ln(R(t))$ has the same derivative as $-\ln(R_T(t))$, namely the biological death intensity $r(t)$, $-\ln(R(t))$ and $-\ln(R_T(t))$ can only differ by an additive constant, so that $R(t)$ and $R_T(t)$ only differ by a multiplicative constant. Since $R(0) = 1 = \mathbf{P}\{T > 0\} = R_T(0)$, it follows that $R(t)$ and $R_T(t)$ are equal. As

$$\lim_{t\to\infty} \exp\Big\{-\int_0^t r(s)ds\Big\} = \lim_{t\to\infty} R_T(t) = \lim_{t\to\infty} \mathbf{P}\{T > t\} = 0$$

we must have $\int_0^\infty r(s)ds = \infty$. Conversely, if we define

$$R(t) = \exp\Big\{-\int_0^t r(s)\,ds\Big\} \quad \textit{where} \quad \int_0^\infty r(s)ds = \infty,$$

then $R(t)$ in decreasing with $R(0) = 1$ and $R(\infty) = 0$, so that $F(t) = 1 - R(t)$ is increasing with $F(0) = 0$ and $F(\infty) = 1$, making makes F a probability distribution function, and thus R a biological survival function. □

Theorem 7.3. *For a biological life length T we have*

$$\mathbf{E}\{T^n\} = \int_0^\infty R_T(t^{1/n})\,dt.$$

Proof. *By integration by parts and a change of variable in the intgeral, we obtain*

$$\mathbf{E}\{T^n\} = \int_0^\infty t\left(-\frac{d}{dt}R_{T^n}(t)\right)dt = [t\,R_{T^n}(t)]_0^\infty + \int_0^\infty R_{T^n}(t)dt = \int_0^\infty R_T(t^{1/n})dt. \quad \square$$

A biological life length T with a constant biological death intensity $r_T(t) = \lambda$ has the *lack of biological memory* property (cf. Theorem 7.1). By Theorem 7.2, a biological life length T lacks biological memory if and only if T is exponentially $\exp(\lambda)$ distributed.

The second simplest form of biological death intensity, after a constant one, is a polynomial one $r_T(t) = ba^bt^{b-1}$. In this case, Theorem 7.2 gives

$$R_T(t) = \exp\left\{-(a\,t)^b\right\},$$

that is, T is Weibull distributed with parameters a and b, Weibull(a, b). And so

$$\mathbf{E}\{T^n\} = \int_0^\infty R_T(t^{1/n})dt = \int_0^\infty \exp\left\{-a^bt^{b/n}\right\}dt = \frac{n}{a^nb}\int_0^\infty t^{n/b-1}\mathrm{e}^{-t}dt$$
$$= \frac{n\Gamma(n/b)}{a^nb} = \frac{\Gamma(n/b+1)}{a^n}$$

by Theorem 7.3, where $\Gamma(x)$ denotes the gamma function.

If the biological life lengths $T_1, \ldots, T_n$ are exponentially $\exp(\lambda)$ distributed, then their sum $T \equiv T_1 + \ldots + T_n$ is *gamma* distributed with parameters n and λ, $\text{gamma}(n, \lambda)$, with probability density function

$$f_T(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} \mathrm{e}^{-\lambda t} \quad \text{for } t > 0.$$

The corresponding biological survival function is

$$R_T(t) = \sum_{k=0}^{n-1} \frac{\lambda^k t^k}{k!} \mathrm{e}^{-\lambda t} \quad \text{for } t > 0.$$

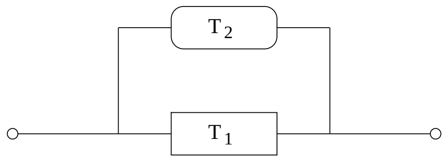
To achieve a high biological health probability of a biological system, the systems may be equipped with more biological components than are actually needed for health, if the components all were healthy. In other words, the biological system is not a pure series coupling, but a biological series coupling of biological subsystems, some of which are biological parallell couplings, to achieve higher biological health probability.

A biological component that is not required for the health of a biological system, when all other biological components of the biological system are healthy, is called a *redundant* biological component.

A *warm redundant* biological components are incorporated with the biological system already from the start of the biological system, while a *cold redundant* biological components is incorporated with the biological system first at the time at which it is required for the health of the system.

Example 7.6. *Figure 7.8 below depicts a biological system where a first biological component with biological life length T_1 is supported by a second redundant biological component with biological life length T_2.*

Figure 7.8. *A biological system where a first biological component is supported by a second redundant biological component.*



For the biological life length T of the biological system we have $T = \max\{T_1, T_2\}$ when the redundant biological component is warm, so that

$$R_T(t) = 1 - (1 - R_{T_1}(t))(1 - R_{T_2}(t)).$$

If the redundant biological component is, we get $T = T_1 + T_2$ instead, so that,

$$R_T(t) = 1 - \int_0^t (1 - R_{T_1}(t - x)) R_{T_2}(x) r_{T_2} dt$$

A quantity of great interest for a biological system, is the probability that biological component number $i = 1, 2, \ldots$ causes the death (unhealth) of the biological system.

That probability, in turn, coincides with the probability that the bilogical life length of biological component number $i = 1, 2, \ldots$ is equal to the biological life length of the whole biological system.

Primarily, biological component that have high probabilities to cause the death (unhealth) of the biological system, are those who should be supported by (warm or cold) redundant biological components.

Example 7.7. *For the biological system in Figure 7.6, we have*

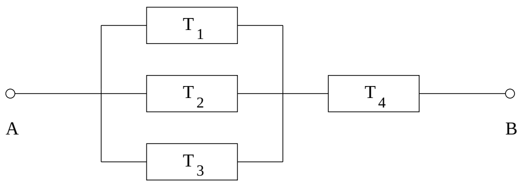
$$\begin{aligned}
&\mathbf{P}\{\textit{biologoical component number } 1 \textit{ causes death}\} \\
&\quad = \mathbf{P}\left\{T_1 = \min[T_1, \max(T_2, T_3)]\right\} \\
&\quad = \mathbf{P}\left\{T_1 \le \max[T_2, T_3]\right\} \\
&\quad = \int_0^\infty \mathbf{P}\{\max(T_2, T_3) \ge t\} f_{T_1}(t)dt \\
&\quad = \int_0^\infty (1 - F_{T_2}(t)F_{T_3}(t)) f_{T_1}(t)dt \\
&\quad = \int_0^\infty (1 - (1 - R_{T_2}(t))(1 - R_{T_3}(t)))\, R_{T_1}(t) r_{T_1}(t)dt.
\end{aligned}$$

Notice that this probability must be $\frac{2}{3}$, by basic combinatorics, when the biological life lengths T_1, T_2 and T_3 are identically distributed.

7.3 Laboration

1. In the biological system in Figure 7.9 below, the first three biological components have biological life lengths T_1, T_2, T_3 that are Weibull$(1, \frac{1}{2})$ distributed, while the fourth biological component have a biological life length T_4 that is $\exp(\frac{1}{2})$ distributed.

Figure 7.9. *A biological system with four biological components.*



a) *Find the expected lifelength $\mathbf{E}\{T\}$ for the biological system. Plot the biological death rate $r_T(t)$, $t \in (0, 10)$, for the biological system: Is the biological system IFR or DFR, or neither IFR nor DFR?*

b) *Find the probability that it is biological component number 4 that causes the desth (unhealth) of the biological system.* Notice that

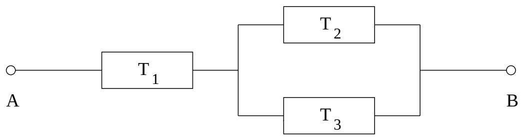
$$\mathbf{P}\{\max[T_1, T_2, T_3] > T_4\} = \int_0^\infty \mathbf{P}\{\max[T_1, T_2, T_3] > t\} f_{T_4}(t)dt.$$

c) Redo task a, first with biological component number 4 doubled with a warm redundant $\exp(\frac{1}{2})$ distributed biological component, and then with biological component number 4 doubled with a cold redundant $\exp(\frac{1}{2})$ distributed biological component. Plot the difference between the biological death rates $r_T(t)$ from tasks a and c, for each of the two ways to incorporate the redundant component.

d) For which values of the parameter ρ does a change of biological component number 4 to an $\exp(\rho)$, $\rho < \frac{1}{2}$, distributed biological component, have the same effect on the expected biological life length $\mathbf{E}\{T\}$ of the biological system, as have the incorporation of the warm and cold redundant $\exp(\frac{1}{2})$ distributed biological component, as described in task d?

2. In the biological system in Figure 7.10 below, the first biological component has a biological life length T_1 that is Weibull$(\mu, \frac{1}{3})$ distributed, while the second and third biological components have biological life length T_2 and T_3 that are Weibull$(\lambda, \frac{1}{3})$. distributed.

Figure 7.10. *A biological system with three biological components.*



The monetary cost of a Weibull$(\gamma, \frac{1}{3})$ distributed biological component is $1/5+1/\gamma$ (in some suitable monetary unit). *Diplay graphically the values of the parameters λ and μ, that maximizes the expected biological life length $\mathbf{E}\{T\}$ of the biological system, at the total costs $1, 2, \ldots, 10$ monitary units, of the biological system. Also plot the expected biological life length $\mathbf{E}\{T\}$ as a function of the costs $1, 2, \ldots, 10$, for the optimal values of the parameters λ and μ.*

In *Mathematica*, it is suitable to define the expected biologica life length as a function of the parameters λ and μ:

```
mean[lambda_,mu_] := ...
```

and then describe how the parameter μ depends on the total costs cost of the biological system, together with thw parameter λ,

```
mu[cost_,lambda_] := ...
```

Then use the command

```
FindMaximum[mean[lambda,mu[cost,lambda]],{lambda,lambda0}]
```

for the total costs $1, \ldots, 10$, for example, using the starting value `lambda0` as the solution to the equation `mu[kost,lambda] = lambda`.

As an alternative to use the `FindMaximum` command, one may use the `NMaximize` command, possibly with suitable constraints in order to avoid some of the analytical labour described above.

Mathematica can be instructed to manufacture a list with the different optimal values of λ in the following manner:

```
Table[lambda/.Last[FindMinimum[-mean[lambda,mu[kost,lambda]],
```

```
{lambda,lambda0/.Last[Solve[mu[kost,lambda0]==lambda0]]}]],
{kost,1,10}]
```

Notice that the command `FindMinimum[.]`returns a list of the form

```
{minimum,{lambda->.}}
```

where `lambda` can be reached with the command

```
Last[FindMinimum[.]]
```

`lambda` thus can be given that value by means of the command

```
lambda/.Last[FindMinimum[.]].
```

In the same way, the command

```
Solve[mu[kost,lambda]==lambda]
```

returns

```
{{lambda->0},{lambda->.}}
```

where the value of `lambda` can be reached with the command

```
Last[Solve[mu[kost,lambda] == lambda]]
```

and `lambda0` is given that value with

```
lambda0/.Last[Solve[mu[kost,lambda]==lambda]].
```