Exercise 1 (2 points)

The time (in days) until the first failure of a vending machine is $T \sim \text{Weibull}(k = 2, \rho = 0.03)$. Thus, the density is $f(t) = k\rho(\rho t)^{k-1} \exp\left(-(\rho t)^k\right) = 0.0018t \exp\left(-(0.03t)^2\right)$.

(a) The probability that the machine works well during 40 and 60 days respectively will be calculated here. First I will show that the survival function of the Weibull distribution is given by $S(t) = \exp\left(-(\rho t)^k\right)$, in general.

$$
\begin{aligned}
S(t) &= \mathrm{P}(T > t) = 1 - \mathrm{P}(T \leq t) = 1 - F(t) = 1 - \int_0^t f(u)\mathrm{d}u \\
&= 1 - \int_0^t k\rho(\rho u)^{k-1} \exp\left(-(\rho u)^k\right)\mathrm{d}u = 1 - \left(-\exp\left(-(\rho u)^k\right)\Big|_0^t\right) \\
&= 1 - \left(1 - \exp\left(-(\rho t)^k\right) = \exp\left(-(\rho t)^k\right).
\end{aligned}
$$

Inserting the parameters in this specific case gives $S(t) = \exp\left(-(0.03t)^2\right)$. The probabilities in question are found by simply calculating $S(40)$ and $S(60)$. Hence, the probability that the machine works well during 40 days is $S(40) = \exp\left(-(0.03 \cdot 40)^2\right) \approx 0.237$. Similarly, the probability that the machine works well during 60 days is $S(60) = \exp\left(-(0.03 \cdot 60)^2\right) \approx 0.0392$.

(b) The 80% quantile of T will be given and interpreted here.

In general, the 80% quantile, $q_{0.8}$, of a Weibull distribution is defined as

$$
0.8 = \mathrm{P}(T \leq q_{0.8}) = F(q_{0.8}) = 1 - S(q_{0.8}) = 1 - \exp\left(-(\rho q_{0.8})^k\right).
$$

Solving the equation for $q_{0.8}$ yields the expression $q_{0.8} = \frac{1}{\rho}(-\ln(1 - 0.8))^{1/k}$. Finally, inserting the given parameters yields $q_{0.8} = \frac{1}{0.03}(-\ln(0.2))^{1/2} \approx 42.288$.

The interpretation of the quantile is that 80% of the machines have failed after approximately 42.3 days, i.e. after approximately 42 days, 7 hours and 12 minutes.

(c) If there is a minimal survival time G, the 3-parameter Weibull distribution can be used to model survival times. The survival function for the 3-parameter Weibull is defined as

$$
S(t) = \begin{cases} 1 & \text{if } t < G, \\ \exp\left(-\rho^k(t - G)^k\right) & \text{if } t \geq G, \end{cases}
$$

Given this survival function, the risk (hazard) function is

$$
\lambda(t) = \frac{f(t)}{S(t)} = -\frac{\mathrm{d}}{\mathrm{d}t}\ln S(t) = \begin{cases} 0 & \text{if } t < G, \\ \rho^k k(t - G)^{k-1} & \text{if } t \geq G, \end{cases}
$$

and the density function is

$$f(t) = -\frac{\mathrm{d}}{\mathrm{d}t}S(t) = S(t)\lambda(t) = \begin{cases} 0 & \text{if } t < G, \\ \rho^k k(t-G)^{k-1} \exp\left(-\rho^k (t-G)^k\right) & \text{if } t \geq G. \end{cases}$$

(d) Given $k = 2, \rho = 0.03$ and assuming a minimum time without any failure of $G = 15$ (days), the mean and median survival times are calculated.

The mean survival time is given by the expression

$$\begin{aligned} \mathrm{E}(T) = \int_G^\infty u f(u) \mathrm{d}u &= \int_0^\infty S(u) \mathrm{d}u \\ &= G + \int_G^\infty \exp\left(-\rho^k (u-G)^k\right) \mathrm{d}u \\ &= G + \frac{1}{\rho} \int_0^\infty \exp\left(-x^k\right) \mathrm{d}x, \end{aligned}$$

where the change of variables $x = \rho(u - G)$ has been used. Inserting $k = 2$ it is apparent that the integral is the Gaussian integral on the positive x-axis, which has the solution $\frac{\sqrt{\pi}}{2}$. Thus, the solution is

$$\mathrm{E}(T) = G + \frac{1}{\rho}\frac{\sqrt{\pi}}{2},$$

which, when inserting $\rho = 0.03$ and $G = 15$ gives the numerical solution $E(T) \approx 44.54$ days.

The median m is the quantile such that

$$0.5 = \mathrm{P}(G \leq T \leq m) = 1 - S(m) = 1 - \exp\left(-\rho^k (m-G)^k\right).$$

Solving the equation gives

$$\begin{aligned} \exp\left(-\rho^k (m-G)^k\right) &= 0.5 \\ \rho(m-G) &= (\ln 2)^{1/k} \\ m &= G + \frac{1}{\rho}(\ln 2)^{1/k}. \end{aligned}$$

Thus, the median survival time with $k = 2, \rho = 0.03$ and $G = 15$ is approximately 42.75 days.

Exercise 2 (2 points)

The hazard function of a piecewise exponential distribution is

$$\lambda(t) = \lambda_j, \ \ \forall t \in [\pi_{j-1}, \pi_j),$$

where $0 < \pi_1 < \ldots < \pi_k < \pi_{k+1} = \infty$ and $j \in \{1, \ldots, k+1\}$.

(a) The expressions of the density and survival functions at a time $t \in [\pi_{j-1}, \pi_j)$ are calculated here.

The expression of the survival function at a time $t \in [\pi_{j-1}, \pi_j)$ is

$$\begin{aligned} S(t) = \exp\left(-\int_0^t \lambda(s)\mathrm{d}s\right) &= \exp\left(-\int_0^{\pi_1} \lambda_1 \mathrm{d}s - \int_{\pi_1}^{\pi_2} \lambda_2 \mathrm{d}s - \ldots - \int_{\pi_{j-1}}^{t} \lambda_j \mathrm{d}s\right) \\ &= \exp\left\{-\left[\lambda_1\pi_1 + \lambda_2(\pi_2 - \pi_1) + \ldots + \lambda_j(t - \pi_{j-1})\right]\right\} \\ &= \exp\left\{-\left[\pi_1(\lambda_1 - \lambda_2) + \pi_2(\lambda_2 - \lambda_3) + \ldots + \pi_{j-1}(\lambda_{j-1} - \lambda_j) + t\lambda_j\right]\right\} \end{aligned}$$

Using the expression $f(t) = S(t)\lambda(t)$, the expression of the density at a time $t \in [\pi_{j-1}, \pi_j)$ is simply

$$f(t) = \lambda_j \exp\left\{-\left[\pi_1(\lambda_1 - \lambda_2) + \pi_2(\lambda_2 - \lambda_3) + \ldots + \pi_{j-1}(\lambda_{j-1} - \lambda_j) + t\lambda_j\right]\right\}$$

(b) Given $k = 2$, the hazard, survival and density functions will be drawn for the values $\lambda_1 = 0.15, \lambda_2 = 0.1, \lambda_3 = 0.2, \pi_1 = 7$ and $\pi_2 = 13$.

These parameters give us the functions

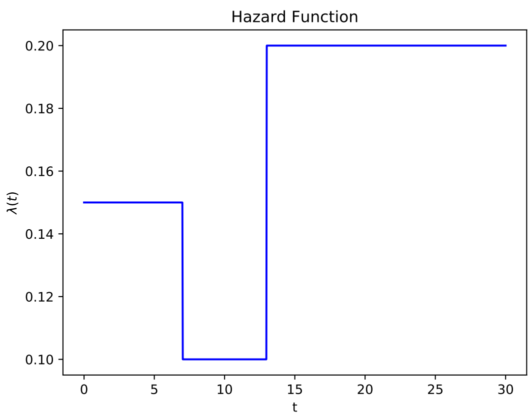
$$\begin{aligned} \lambda(t) &= \begin{cases} \lambda_1 & \text{if } t \in [0, 7), \\ \lambda_2 & \text{if } t \in [7, 13), \\ \lambda_3 & \text{if } t \in [13, \infty) \end{cases} \\ &= \begin{cases} 0.15 & \text{if } t \in [0, 7), \\ 0.1 & \text{if } t \in [7, 13), \\ 0.2 & \text{if } t \in [13, \infty), \end{cases} \end{aligned}$$

$$\begin{aligned} S(t) &= \begin{cases} \exp\left(-t\lambda_1\right) & \text{if } t \in [0, 7), \\ \exp\left(-(\pi_1(\lambda_1 - \lambda_2) + t\lambda_2)\right) & \text{if } t \in [7, 13), \\ \exp\left(-(\pi_1(\lambda_1 - \lambda_2) + \pi_2(\lambda_2 - \lambda_3) + t\lambda_3)\right) & \text{if } t \in [13, \infty), \end{cases} \\ &= \begin{cases} \exp\left(-0.15t\right) & \text{if } t \in [0, 7), \\ \exp\left(-(0.35 + 0.1t)\right) & \text{if } t \in [7, 13), \\ \exp\left(-(0.35 - 1.3 + 0.2t)\right) & \text{if } t \in [13, \infty), \end{cases} \end{aligned}$$

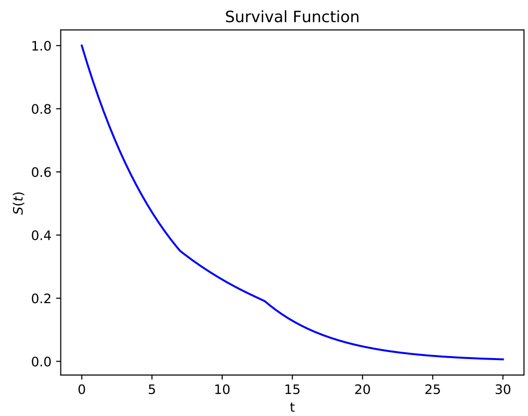
and

$$
\begin{aligned}
f(t) &= \begin{cases} \lambda_1 \exp\left(-t\lambda_1\right) & \text{if } t \in [0,7), \\ \lambda_2 \exp\left(-(\pi_1(\lambda_1-\lambda_2)+t\lambda_2)\right) & \text{if } t \in [7,13), \\ \lambda_3 \exp\left(-(\pi_1(\lambda_1-\lambda_2)+\pi_2(\lambda_2-\lambda_3)+t\lambda_3)\right) & \text{if } t \in [13,\infty), \end{cases} \\
&= \begin{cases} 0.15 \exp\left(-0.15t\right) & \text{if } t \in [0,7), \\ 0.1 \exp\left(-(0.35+0.1t)\right) & \text{if } t \in [7,13), \\ 0.2 \exp\left(-(0.35-1.3+0.2t)\right) & \text{if } t \in [13,\infty), \end{cases}
\end{aligned}
$$

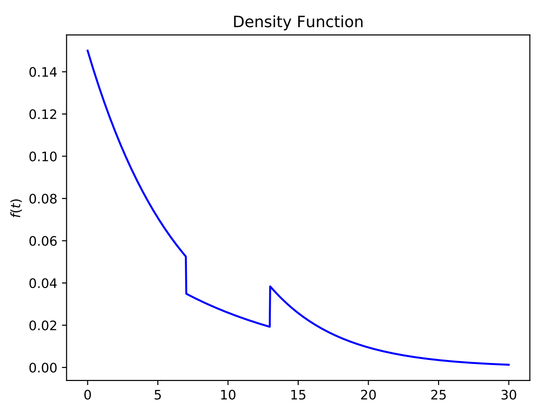
The three functions are plotted in figure 1.



(a) Hazard Function



(b) Survival Function



(c) Density Function

Figure 1: The functions are plotted on the interval $[0,30)$.

Exercise 3 (1.5 points)

First of all, all types of censoring that occur in this experiment are of Type I, since the maximum length of observation is prespecified. I will use the usual definitions, i.e. that $S(t) = \mathrm{P}(T > t)$ is the survival function of T, that $\lambda(t)$ is the hazard function of T and $f(t)$ is the density function of T. Since T is measured from the time of cancer induction, $T = 0$ when the mice are seven weeks old.

(a) Since the first examination shows that the first mouse has got a palpable cancer, there is no censoring in this case, i.e. this is an exact failure time. The time interval that contains T is 42-49 days. Thus, the contribution to the likelihood function is $L_1 = \lambda(t_1)S(t_1)$, where this measurement in time is defined as $t_1 \in [42, 49)$.

(b) For the second mouse, a cancer is detected in week 13, but the cancer was not palpable yet the week before. Since the cancer was not palpable the week before, we know that the time interval that contains T is 91-98 days, since the cancer was detected in week 13. Thus, the contribution to the likelihood function is $L_2 = \lambda(t_2)S(t_2)$, where this measurement in time is defined as $t_2 \in [91, 98)$.

(c) When a mouse dies without any cancer symptoms, on an unknown day before the first examination, the observation is right-censored, since we could not check for cancer symptoms from that point forward. This death might have been due to competing risks, i.e. that the mouse died from something else other than cancer. The contribution to the likelihood function is $L_c = S(C_{R_c})$, where $C_{R_c} \in [0, 42)$, assuming that the first examination takes place on day 42.

(d) A mouse that dies without any symptoms on the 37^{th} day after the cancer induction represents a right-censored data point. This death might also have been due to competing risks. The time interval in this case consists of only the 37^{th} day, since we know that the mouse died then. The contribution to the likelihood function is $L_d = S(37)$.

(e) The mouse that survives the study without any signs of breast cancer is a right-censored observation. The study ended without the event of interest happening to the mouse, which means that this is caused by administrative censoring. Thus, the contribution of this mouse to the likelihood function is $L_e = S(C_{R_e})$, where $C_{R_e} \in [140, 147)$ days, depending on which day in week 20 it is examined the last time.

Exercise 4 (2 points)

Let $T \sim \mathrm{Exp}(\rho)$ be the survival time of interest and $C \sim \mathrm{Exp}(\lambda)$ be the censoring time. Assume that T and C are independent and let

$$Y := \min(T, C) \text{ and } \delta := \begin{cases} 1 & \text{if } T \leq C, \\ 0 & \text{if } T > C. \end{cases}$$

Moreover, I assume that the parameters ρ and λ are rate parameters.

(a) The value of $\mathrm{P}(\delta = 1)$ is calculated as

$$\mathrm{P}(\delta = 1) = \mathrm{P}(T \leq C).$$

In order to find this probability one first needs to know the joint distribution of T and C. Since the two random variables are independent, their joint distribution is given by the product of their distribution functions, i.e. $f_{T,C}(t,c) = f_T(t)f_C(c) = \rho\lambda \exp\left(-(\rho t + \lambda c)\right)$. This means that

$$\begin{aligned}
\mathrm{P}(T \leq C) &= \int_0^\infty \int_0^c f_{T,C}(t,c)\mathrm{d}t\mathrm{d}c = \int_0^\infty \int_0^c \rho\lambda \exp\left(-(\rho t + \lambda c)\right)\mathrm{d}t\mathrm{d}c \\
&= \int_0^\infty \lambda \exp\left(-\lambda c\right)(-\exp\left(-\rho t\right))\big|_0^c \mathrm{d}c = \int_0^\infty \lambda \exp\left(-\lambda c\right)(1 - \exp\left(-\rho c\right))\mathrm{d}c \\
&= \int_0^\infty \lambda \exp\left(-\lambda c\right)\mathrm{d}c - \int_0^\infty \lambda \exp\left(-(\lambda + \rho)c\right)\mathrm{d}c \\
&= (-\exp\left(-\lambda c\right))\big|_0^\infty - \left(-\frac{\lambda}{\lambda + \rho}\exp\left(-(\lambda + \rho)c\right)\right)\Big|_0^\infty \\
&= 1 - \frac{\lambda}{\rho + \lambda} = \frac{\rho}{\rho + \lambda}.
\end{aligned}$$

Thus, the probability of seeing an exact failure time instead of a censored time is $\frac{\rho}{\rho+\lambda}$. From the fraction it becomes apparent that if $\rho \gg \lambda$, then $\mathrm{P}(T \leq C) \to 1$. In the opposite case, when $\lambda \gg \rho$, then $\mathrm{P}(T \leq C) \to 0$. Hence, the interpretation of the result is that when the rate parameter that governs the exact failure times, ρ, grows compared to λ, the probability of seeing exact failure times before censored times increases. On the contrary, if the rate parameter that governs the censored times, λ, grows compared to ρ, the probability of seeing exact failure times before censored times decreases.

(b) The distribution of $Y = \min(T, C)$ is an exponential distribution with rate parameter $\psi = \rho + \lambda$. The proof is

$$\begin{aligned}
\mathrm{P}(Y \leq y) &= 1 - \mathrm{P}(Y > y) = 1 - \mathrm{P}(\min(T, C) > y) = 1 - \mathrm{P}(T > y, C > y) \\
&\overset{Indep.}{=} 1 - \mathrm{P}(T > y)\mathrm{P}(C > y) = 1 - (1 - \mathrm{P}(T \leq y))(1 - \mathrm{P}(C \leq y)) \\
&= 1 - (1 - (1 - \exp\left(-\rho y\right)))(1 - (1 - \exp\left(-\lambda y\right))) \\
&= 1 - \exp\left(-\rho y\right)\exp\left(-\lambda y\right) = 1 - \exp\left(-(\rho + \lambda)y\right) = 1 - \exp\left(-\psi y\right),
\end{aligned}$$

which is the cdf of an exponential distribution with parameter ψ.

Exercise 5 (2.5 points)

We are working with times from bone marrow transplant until relapse and death, respectively, in ten patients with leukemia. By the end of the study, six patients were alive, four of whom had not suffered any relapse neither. Assume that the time until relapse follows an exponential distribution with rate parameter λ, i.e. $T_1 \sim \mathrm{Exp}(\lambda)$. Further assume that the time until death follows a Weibull distribution with parameters ρ and k, i.e. $T_2 \sim \mathrm{Weibull}(k, \rho)$.

(a) The value of the maximum likelihood estimator of λ will be found. Let $U = T_1$ in order to simplify the notation. For the likelihood function this yields

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{10} f_U(u_i)^{\delta_i} S_U(u_i)^{1-\delta_i} = \prod_{i=1}^{10} \lambda^{\delta_i} \exp\left(-\lambda u_i\right) \\ &= \lambda^{\sum_{i=1}^{10} \delta_i} \exp\left(-\lambda \sum_{i=1}^{10} u_i\right) = \lambda^d \exp\left(-\lambda y\right), \end{aligned}$$

where we have defined $d := \sum_{i=1}^{10} \delta_i$ and $y := \sum_{i=1}^{10} u_i$. The log-likelihood function is thus given by

$$l(\lambda) = \log L(\lambda) = d \log \lambda - \lambda y.$$

The score function equated to zero yields

$$\mathcal{S}(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda} l(\lambda) = \frac{d}{\lambda} - y \stackrel{!}{=} 0 \implies \hat{\lambda}_{\mathrm{MLE}} = \frac{d}{y} = \frac{\sum_{i=1}^{10} \delta_i}{\sum_{i=1}^{10} u_i}.$$

Thus, the maximum likelihood estimator is the total number of relapses divided by the total observed time until relapse. In a case where there is no censoring, the reciprocal of this value, $1/\hat{\lambda}_{\mathrm{MLE}}$, is the expected value of the exponential distribution of U, which is the mean observation time until relapse. When there is censoring, the quantity plays a similar role, but the mean observation time until relapse would be overestimated, since some of the δ's in the sum are zero.

(b) The estimated mean residual lifetime of T_1 after 10 months will be calculated here. First of all, the mean residual lifetime for T_1 is defined as

$$\mathrm{mrl}(t) = \mathrm{E}(T_1 - t | T_1 > t) = \frac{\int_t^\infty (u-t) f_{T_1}(u) \mathrm{d}u}{S_{T_1}(t)} = \frac{\int_t^\infty S_{T_1}(u) \mathrm{d}u}{S_{T_1}(t)}$$

Since we know that $T_1 \sim \mathrm{Exp}(\lambda)$, we know that $S_{T_1}(u) = \exp\left(-\lambda u\right)$, such that

$$\mathrm{mrl}(10) = \frac{\int_{10}^{\infty} \exp(-\lambda u)\mathrm{d}u}{\exp(-10\lambda)} = \frac{-\lambda^{-1}\exp(-\lambda u)\big|_{10}^{\infty}}{\exp(-10\lambda)} = \frac{\exp(-10\lambda)}{\lambda\exp(-10\lambda)} = \frac{1}{\lambda},$$

which is the same as the mean lifetime, i.e. the expected value of T_1, which is expected, since the hazard function of the exponential is constant.

(c) The expression of the likelihood function of the parameters ρ and k will be given here. Let $V = T_2$ in order to simplify the notation. Thus the likelihood function is

$$L(\rho, k) = \prod_{i=1}^{10} \lambda_V(v_i)^{\delta_i} S_V(v_i) = \prod_{i=1}^{10} (k\rho(\rho v_i)^{k-1})^{\delta_i} \exp\left(-(\rho v_i)^k\right),$$

since we know that $\lambda_V(v) = k\rho(\rho v)^{k-1}$ and $S_V(v) = \exp\left(-(\rho v)^k\right)$ for the Weibull distribution.

(d) In this part of the problem, the function `fitdistcens` from the R package `fitdistrplus` will be used to estimate the parameters of the Weibull distribution. Since the Weibull distribution describes the time until death, with parameters k and ρ, these two parameters will be estimated using only the data for time until death. The R code has been added below.

```r
library(fitdistrplus)
df <- data.frame(
    left = c(11, 12, 15, 33, 45, 28, 16, 17, 19, 30),
    right = c(11, 12, 15, NA, 45, NA, NA, NA, NA, NA)
)
fitdistcens(censdata = df, distr = "weibull")
```

After running the code, the output shows that the shape parameter has been estimated to $k \approx 1.888$ and the scale parameter has been estimated to $\rho^{-1} \approx 40.019$, which means that $\rho \approx 0.025$.