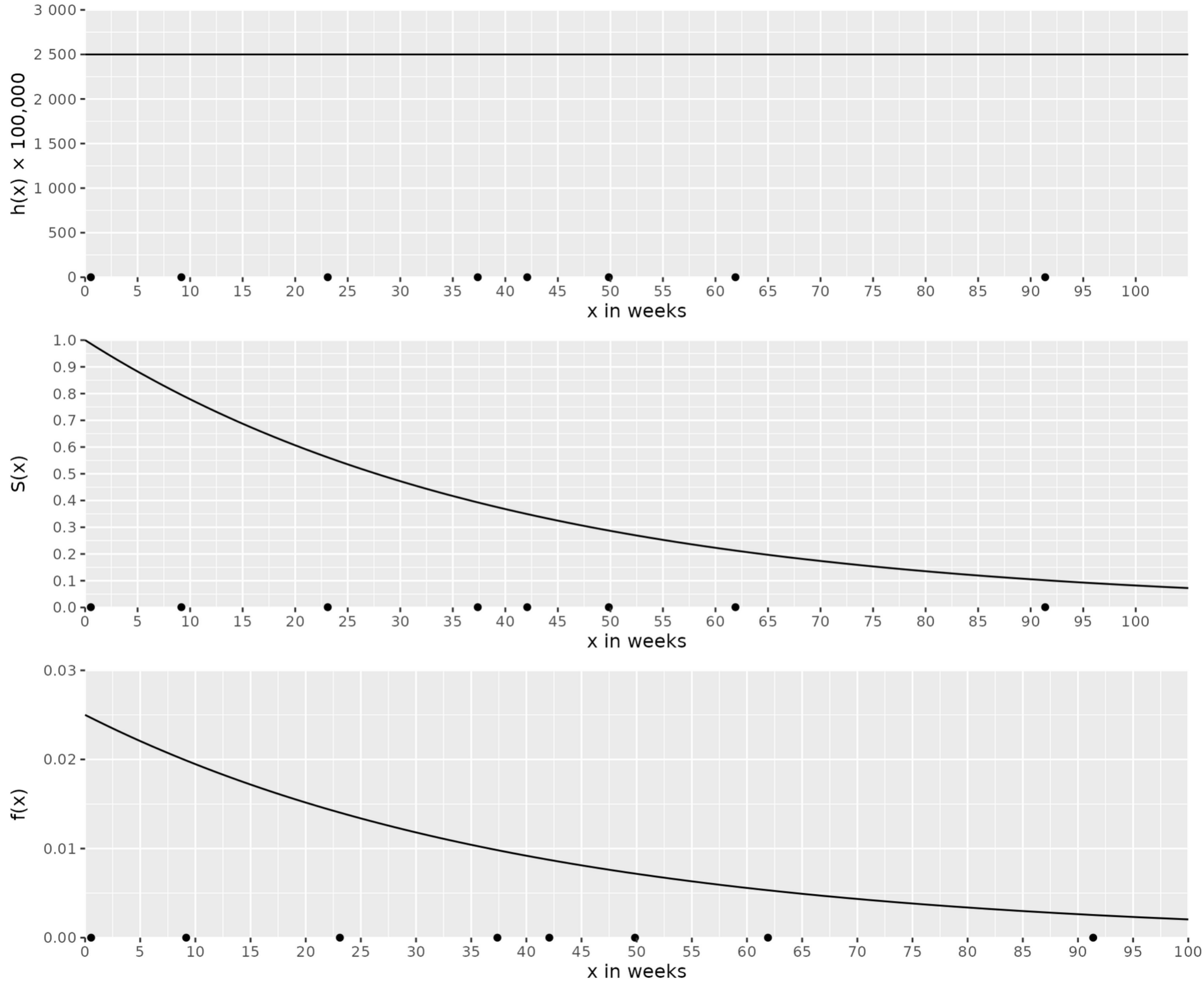
Survival Analysis Preparation Exam
12 May 2022

1. Probabilities of survival

These three graphs describe the distribution of times until COVID-19 infection for children given rates observed in winter 2022 in Rostock.



a) Name these graphs using the terminology of survival analysis.

h(x): Hazard function
S(x): Survival function
f(x): Density function

Assuming the rates of infection stay constant:

b) What share of the population will not get infected within the next 20 weeks?

Given current rates, around 60% of the population are expected not to get infected within the next 20 weeks.

c) What is the probability getting infected within the next 20 weeks?

Given current rates, the probability of getting infected within the next 20 weeks is around 40%.

d) What is the probability of getting infected between week 10 and 20?

Given current rates, the probability of getting infected between week 10 and 20 is around 18%. This can be learned from the S(x) curve by subtracting S(x=20) from S(x=10).

e) What is the name of the distribution corresponding to the graphs shown above?

This is an Exponential distribution.

f) What characteristic makes this distribution special?

The exponential distribution is characterized by a constant hazard.

2. From data to distribution

We have observed the following survival times following cancer diagnosis: ***x*** = (1, 9, 23, 37, 42, 50, 62, 91) weeks. We believe the data is well described by an exponential distribution with $h(x) = \lambda$ and $S(x) = \exp(-\lambda x)$. Calculate the log-likelihood for parameter values λ=0.025 and λ=0.05. Which value for λ is more likely? Remember: $f(x) = h(x)S(x)$.

$L = \prod_i \lambda \exp(-\lambda x_i)$
$\log L = \sum_i \log(\lambda) - \lambda x_i$

under λ=0.025
$\log L_{\lambda.025} = -3.71 + -3.91 + -4.26 + -4.61 + -4.73 + -4.93 + -5.23 + -5.96 =$ **37.38**

under λ=0.05
$\log L_{\lambda.05} = -3.04 + -3.44 + -4.14 + -4.84 + -5.09 + -5.49 + -6.09 + -7.54 =$ **39.71**

The value 0.025 is more likely.

3. Censoring

a) Give two examples of right censored data.

Age at death of birth cohort 1940.

Measuring the temperature of boiling water with a bathtub thermometer.

b) Treat the last two data points in **x** as right-censored. Recalculate the likelihood for λ=0.025 and λ=0.05.

The likelihood contribution from a right-censored observation has to be taken from the corresponding Survival function S(x).

under λ=0.025
$\log L_{\lambda.025}$ = -3.71 + -3.91 + -4.26 + -4.61 + -4.73 + -4.93 + -1.55 + -2.27 = **-30.00**

under λ=0.05
$\log L_{\lambda.05}$ = -3.04 + -3.44 + -4.14 + -4.84 + -5.09 + -5.49 + -3.1 + -4.55 = **-33.72**

4. Research Project

Formulate a research question which can be answered using the methods of Survival Analysis. What is the population of interest? What is the event of interest? When does the observation time start, when does it end? When does censoring occur?