Homework 01

Ihsan Kahveci

2022-04-11

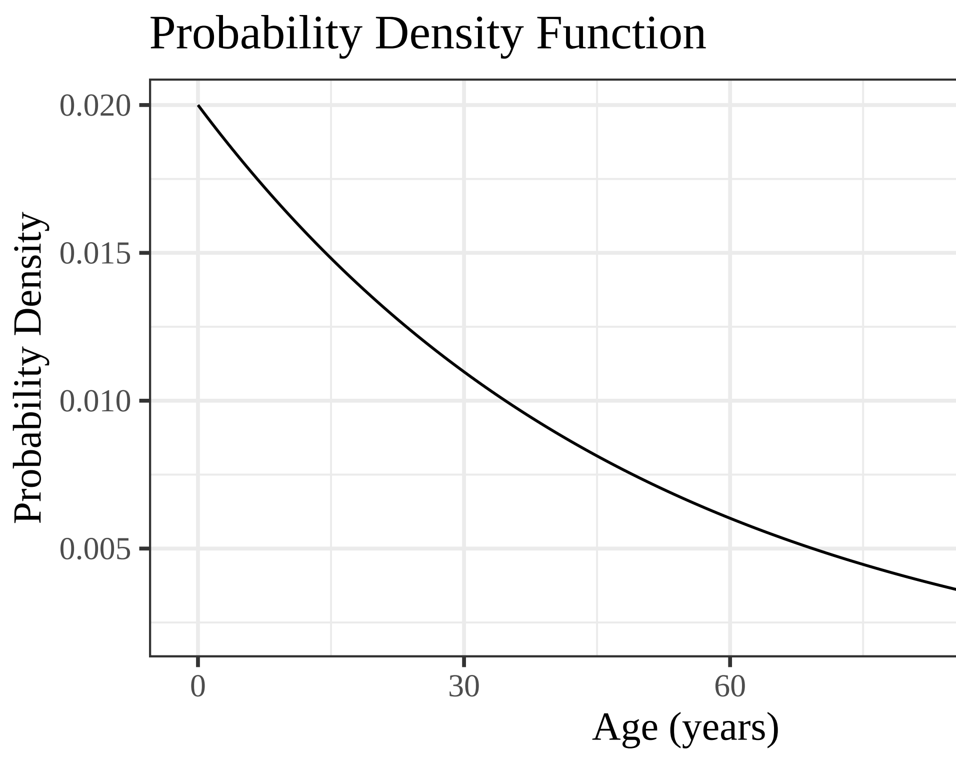
Questions

Q1

If the instantaneous mortality rate is constant (0.02) – independent of age –, the probability distribution follows an exponential distribution. Let x equals to age at death; hence the relevant probability density function $f(x)$, cumulative distribution function $F(x)$, survival function $S(x)$:

$$f(x) = 0.02e^{-0.02x} F(x) = 1 - e^{0.02x} S(x) = e^{-0.02x}$$

(a)



Plotting the probability density function:

(b)

Probability that a member of this population is still alive at age 70: $S(70) = 0.246597$

(c)

Probability that a member of this population dies before age 6: $F(6) = 0.1130796$

(d)

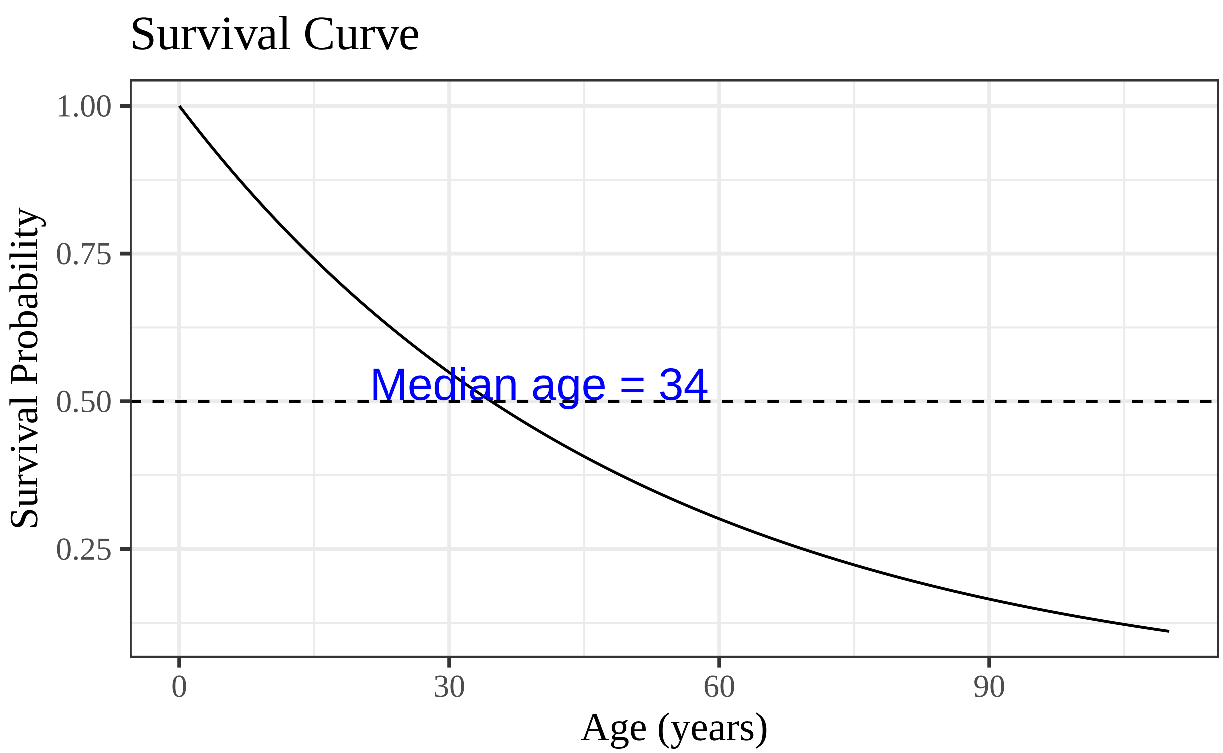
Life expectancy at birth for a member of this population: $e_0 = 50$

(e)

Life expectancy at age 50 for a member of this population: $e_{50} = 50$

(f)

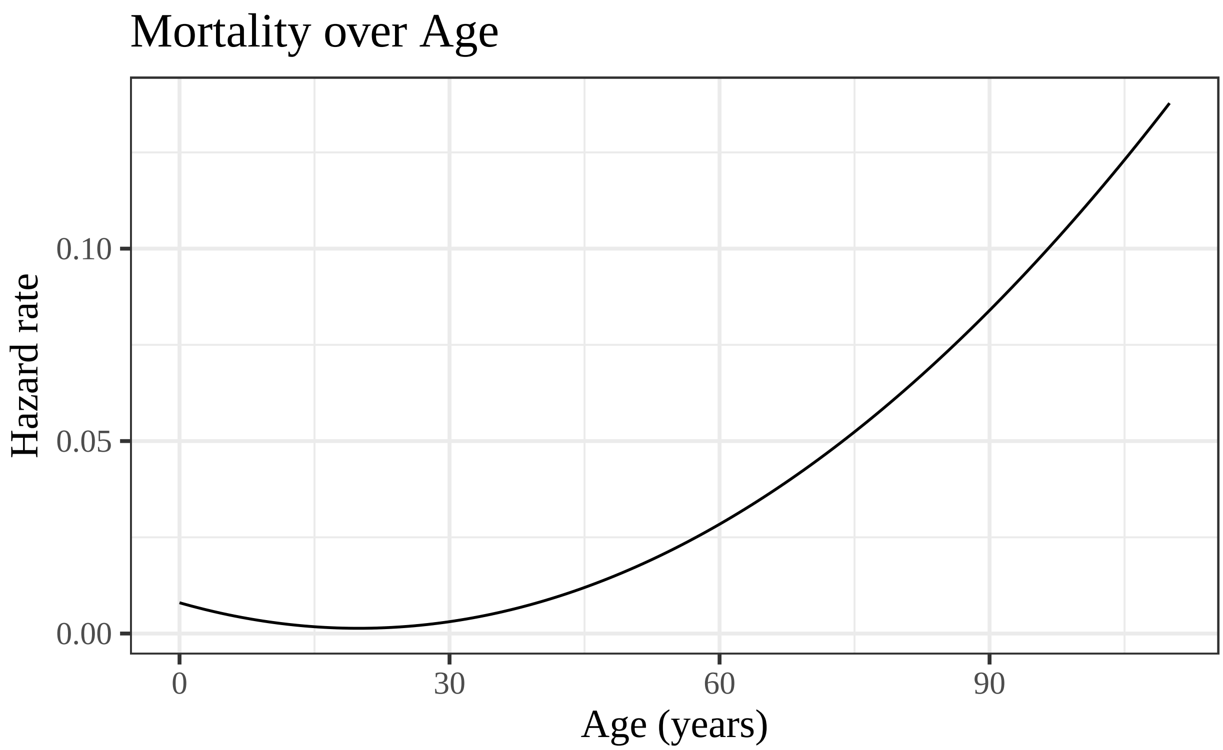
Median age at death for this population: $S(50) = 0.506617$



Q2

(a)

For ages 0 to 110, this mortality rate plot looks like:

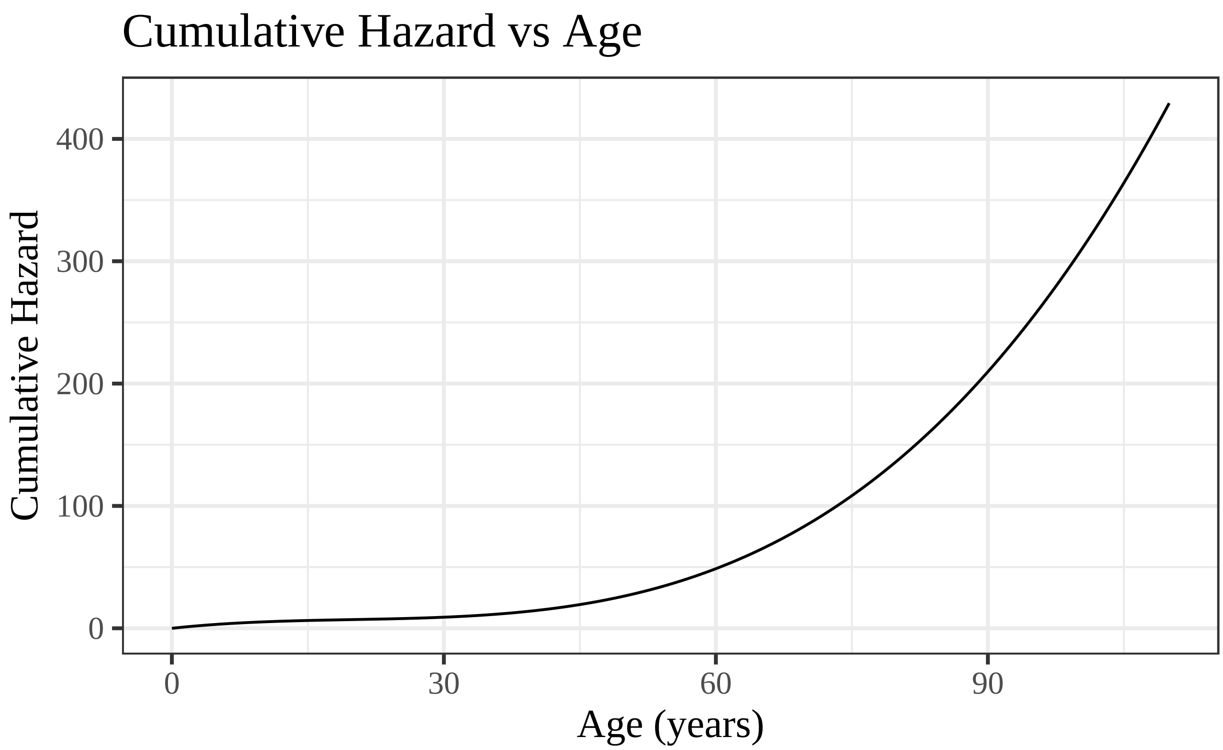


(b)

For a given instantaneous mortality function $\mu(x) = (0.0168x^2 - 0.668x + 8)/1000$, defined as the total area under the curve of $\mu(x)$ bounded on the interval $[0, x]$, or put another way:

$$\begin{aligned}
\Lambda(x) &= \int_0^x \mu(u)du \\
&= \int_0^x \left[(0.0168x^2 - 0.668x + 8)/1000\right] dx \\
&= \int_0^x \frac{21x^2 - 835x + 10000}{1250000} dx \\
&= \frac{21}{125000} \int_0^x x^2 dx - \frac{167}{250000} \int_0^x x dx + \frac{1}{125} \int_0^x 1 dx \\
&= \frac{x \cdot (14x^2 - 835x + 20000)}{25000} + C
\end{aligned}$$

For ages 0 to 110, this cumulative hazard functions looks like:



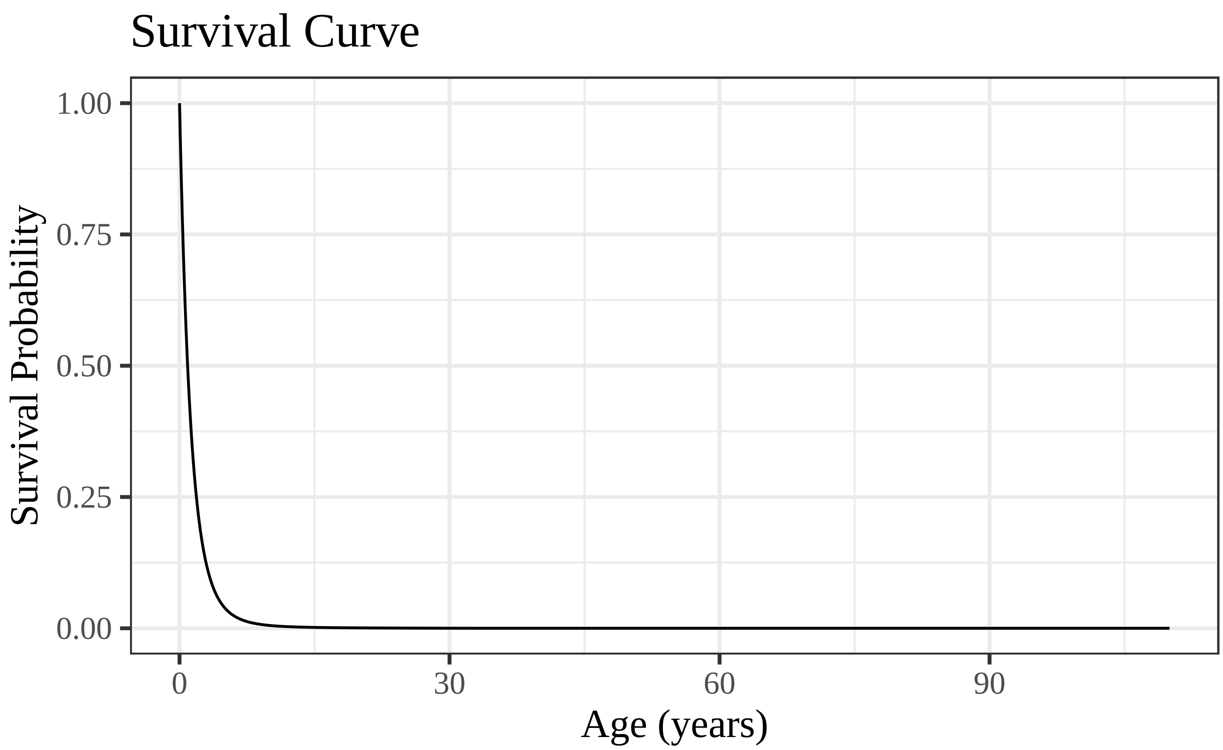
Q2

The survival function, $S(x)$, is defined as the exponentiated negative cumulative hazard function, $e^{-\Lambda(x)}$. Using our calculated cumulative hazard function, the survival function is then:

$$S(x) = \exp\left[\frac{-x \cdot (14x^2 - 835x + 20000)}{25000}\right]$$

(c)

For ages 0 to 110, the survival function then looks like:

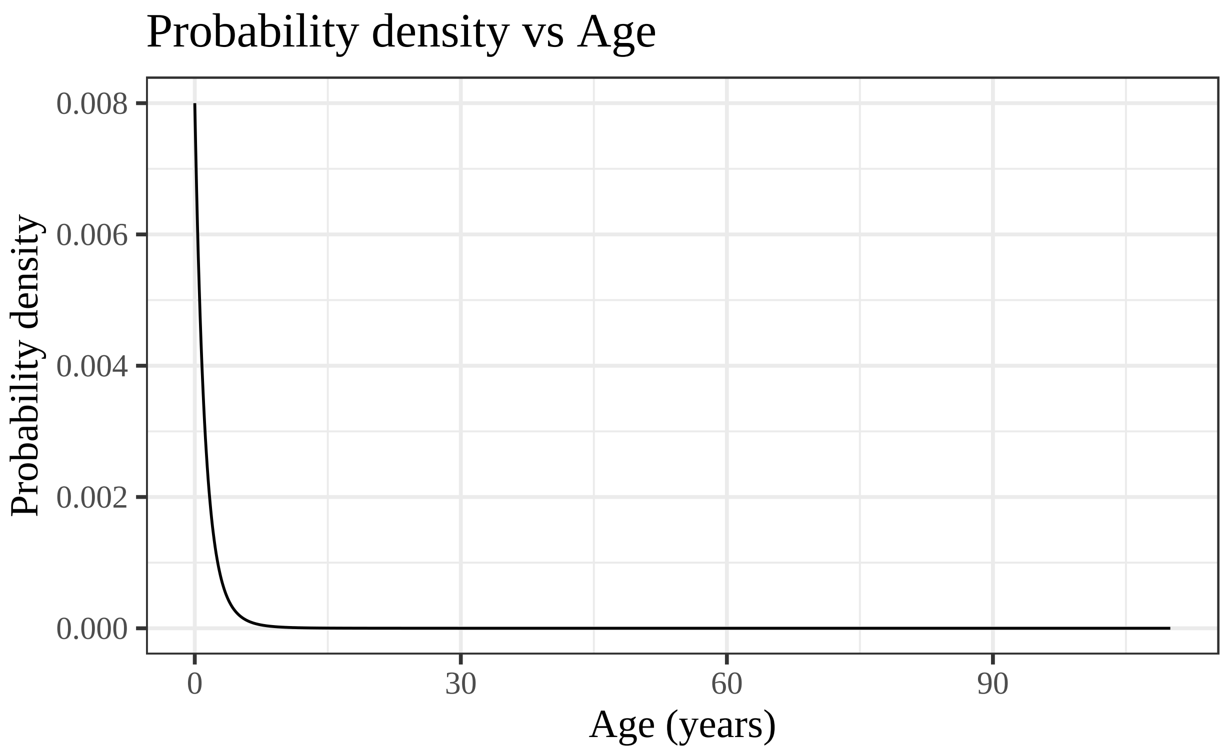


(d)

The probability density function of X, $f(x)$, is the negative derivative of the survival function with respect to x, $f(x) = -\frac{dS(x)}{dx}$. Using our calculated survival function, the probability density function of X is then:

$$f(x) = \frac{-d}{dx} = \mu(x)S(x)$$

For ages 0 to 110, the probability density function looks like:



(e)

Life expectancy at age x, e_x, is defined as:

$$e_x = \frac{\int_x^\infty S(u)du}{S(x)}$$

which simplifies to $\int_0^\infty S(u)du$ for life expectancy at birth, e_0. Using numerical integration, the life expectancy at birth for our cohort is calculated to be **1.442**.

(e)

The life expectancy at age 10 (e_{10}) for a member of this cohort is numerically calculated to be **4.716**.

(f)

The probability that a person aged x dies within the next n years is defined as:

$${}_nq_x = \frac{S(x) - S(x+n)}{S(x)}$$

The ${}_{15}q_{15}$ value for this cohort is then **0.932**.

Appendix

```
# Prep work ---------------------------------------------------------------

# Load libraries
library(tidyverse, quietly = TRUE)


# Make data
age_range <- c(0, 110)
age_data <- tibble(age = seq(age_range[1], age_range[2], .1))

# Question 1
pdf = function(x) 0.02*exp(-0.02*x)
cdf = function(x) 1 - exp(-0.02*x)
survf = function(x) exp(-0.02*x)

pdf_plot =
  age_data %>%
  mutate(PDF = pdf(age)) %>%
  ggplot(aes(x=age, y=PDF)) +
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Probability Density Function",
    x = "Age (years)",
    y = "Probability Density"
    )

median_plot =
  age_data %>%
  mutate(surv = survf(age)) %>%
  ggplot(aes(x=age, y=surv)) +
  geom_line() +
  geom_hline(yintercept = 0.5, linetype="dashed") +
  annotate("text", x=40, y=0.53, label="Median age = 34", color = "blue", size=6) +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Survival Curve",
    x = "Age (years)",
    y = "Survival Probability"
    )
pdf_plot
median_plot
# Question 2
hazard_fun = function(x) (0.0168*x^2 - 0.668*x + 8)/1000

hazard_plot =
  age_data %>%
  mutate(mortality = hazard_fun(age)) %>%
  ggplot(aes(x=age, y=mortality)) +
```

```r
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Mortality over Age",
    x = "Age (years)",
    y = "Hazard rate"
    )

cum_hazard_fun <- function(x) x*(14*x^2 - 835*x + 20000)/25000

chf_plot <-
  ggplot(age_data, aes(x = age, y = cum_hazard_fun(age))) +
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Cumulative Hazard vs Age",
    x = "Age (years)",
    y = "Cumulative Hazard"
  )

survival_fun <- function(x) exp(-1 * cum_hazard_fun(x))

surv_plot =
  age_data %>%
  mutate(surv = survival_fun(age)) %>%
  ggplot(aes(x=age, y=surv)) +
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Survival Curve",
    x = "Age (years)",
    y = "Survival Probability"
    )

pdf_fun = function(x) survival_fun(x) * hazard_fun(x)

pdfun_plot <-
  ggplot(age_data, aes(x = age, y = pdf_fun(age))) +
  geom_line() +
  theme_bw(base_size = plot_size) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Probability density vs Age",
    x = "Age (years)",
    y = "Probability density"
  )


e0 <- integrate(survival_fun, lower = 0, upper = Inf)
e0_val <- round(e0$value, 3)
```

```r
e10 <- integrate(survival_fun, lower = 10, upper = Inf)
e10_val <- round(e10$value / survival_fun(10), digits = 3)

nqx <- function(x, n) (survival_fun(x) - survival_fun(x + n)) / survival_fun(x)
q15_15 <- round(nqx(15, 15), 3)
hazard_plot
chf_plot
surv_plot
pdfun_plot
```