MATH 2310 Reliability

Jakob Balkovec

2024-05-13

```r
plot_function <- function(data_frame,
                          func_name,
                          plot_title,
                          lines = FALSE,
                          line_color = "brown",
                          point_color = "blue") {

  p <- ggplot(data_frame, aes(x = Time, y = Function)) +
    labs(x = "Time", y = func_name, title = plot_title) +
    geom_line(color = line_color) +
    geom_point(color = point_color) +
    ylim(0, max(data_frame$Function + 0.2)) +
    theme_minimal() +
    theme(plot.title = element_text(hjust = 0.5, face = "bold"))

  if (lines) {
    p <- p + geom_vline(xintercept = 5, color = "lightblue", linetype = "dashed") +
             geom_vline(xintercept = 10, color = "lightblue", linetype = "dashed")  +
             geom_vline(xintercept = 15, color = "lightblue", linetype = "dashed")  +

          annotate("text", x = 5, y = max(data_frame$Function), label = "5 yrs", vjust = -0.5, color = "black") +
          annotate("text", x = 10, y = max(data_frame$Function), label = "10 yrs", vjust = -0.5, color = "black")
+
          annotate("text", x = 15, y = max(data_frame$Function), label = "15 yrs", vjust = -0.5, color = "black")
  }
  return(p)
}
```

Activity One

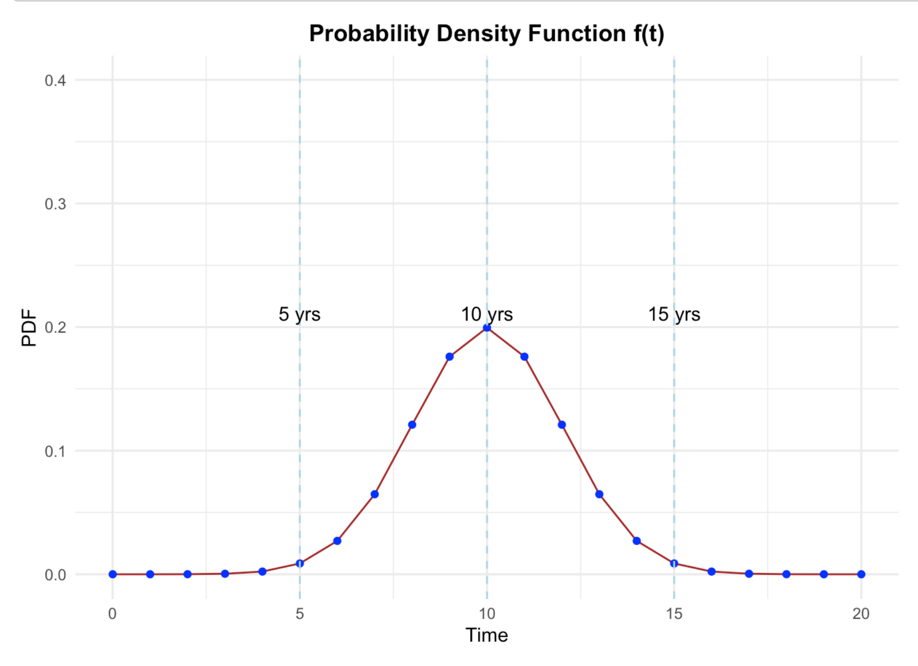
Assume that a certain product can be modeled with a normal failure law having a mean lifetime of 10 years, with a standard deviation of 2 years. In R, define a vector `t <- seq(0,20)` representing the years over which we will examine failure probability.

```r
# Commonly used variables

t <- seq(0,20)

mean_lifetime <- 10
sd_lifetime <- 2
```

[Question A]

- *Generate the probability density function `f(t)` based on the mean and standard deviation lifetimes.*

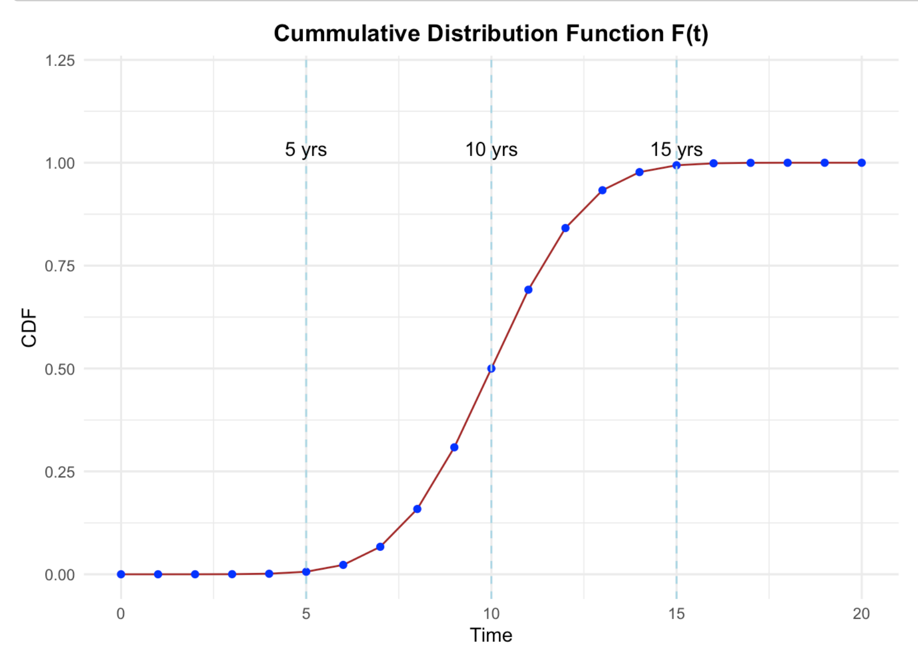
```r
# Normal failure law
#
# Mean lifetime: 10yrs
# SD: 2yrs

# Our probability density function [f(t)]
pdf <- dnorm(t, mean = mean_lifetime, sd = sd_lifetime)

# Combine in a DF for better readability. Using "kable" to prettify the output
pdf_df <- data.frame(Time = t, Function = pdf)
plot_function(data_frame = pdf_df, func_name = "PDF", plot_title = "Probability Density Function f(t)", lines=TRU
E)
```



[Question B]

- *Calculate the cumulative distribution function `F(t)`*

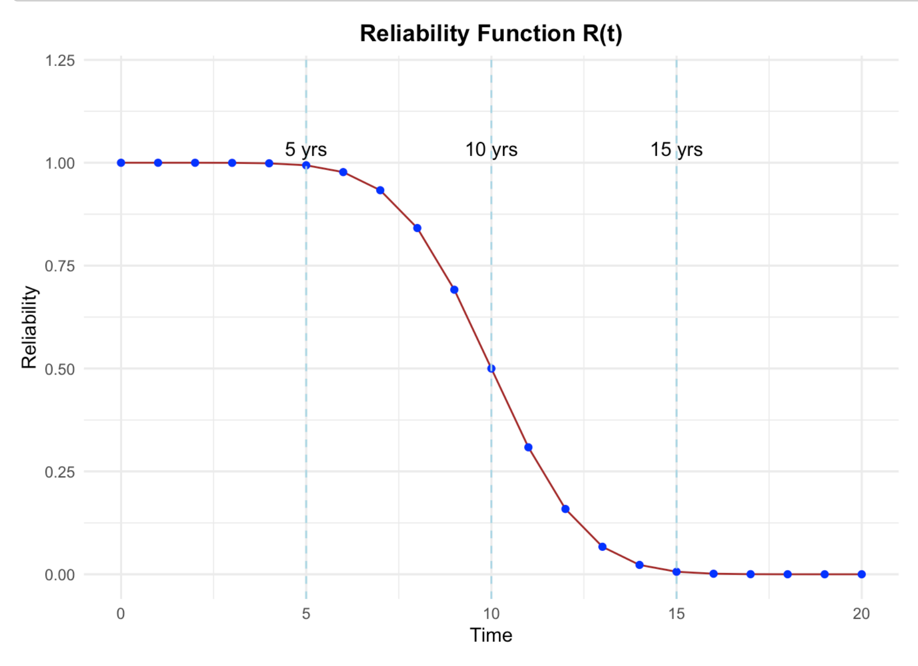
```r
# Normal failure law
#
# Mean lifetime: 10yrs
# SD: 2yrs

# Our cumulative distribution function [F(t)]
cdf <- pnorm(t, mean = mean_lifetime, sd = sd_lifetime)

cdf_df <- data.frame(Time = t, Function = cdf)
plot_function(data_frame = cdf_df, func_name = "CDF", plot_title = "Cummulative Distribution Function F(t)", line
s = TRUE )
```



[Question C]

- *Calculate the reliability function `R(t)`, and plot it as a function of time.*

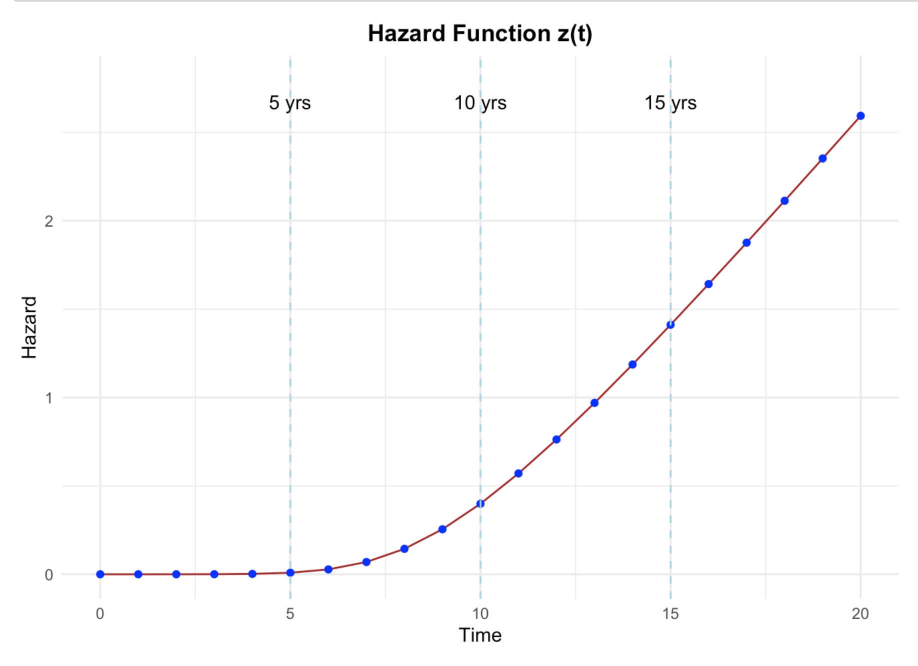
```r
# Normal failure law
#
# Mean lifetime: 10yrs
# SD: 2yrs

# Our reliability function [R(t)]
reliability <- pnorm(t, mean = mean_lifetime, sd = sd_lifetime, lower.tail = FALSE)

# Pack the data into a data-frame
plot_data <- data.frame(Time = t, Function = reliability)
plot_function(data_frame = plot_data, func_name = "Reliability", plot_title = "Reliability Function R(t)", lines
= TRUE)
```



[Question D]

- *Calculate the hazard function `z(t)`, and plot it as a function of time.*

```r
# Normal failure law
#
# Mean lifetime: 10yrs
# SD: 2yrs

# Our hazard function [z(t)]
hazard <- dnorm(t, mean = mean_lifetime, sd = sd_lifetime) / (1 - pnorm(t, mean = mean_lifetime, sd = sd_lifetim
e))

plot_data <- data.frame(Time = t, Function = hazard)
plot_function(data_frame = plot_data, func_name = "Hazard", plot_title = "Hazard Function z(t)", lines = TRUE)
```



[Question E]

- *Based on your result, what type of failure rate do products with normal failure laws have (increasing, decreasing, or constant over time)?*
- Products subject to normal failure laws exhibit a growing likelihood of failure throughout time. When we focus on the first vertical line at 5 years, we notice a very slow but steady increase in the hazard function. Moving into the next segment from 5 years to 10 years, the upward trend becomes more visible, signaling a sharper rise in the function. This trend persists across the remaining sections, highlighting a continuous growth in the probability of failure over time.

Activity Two

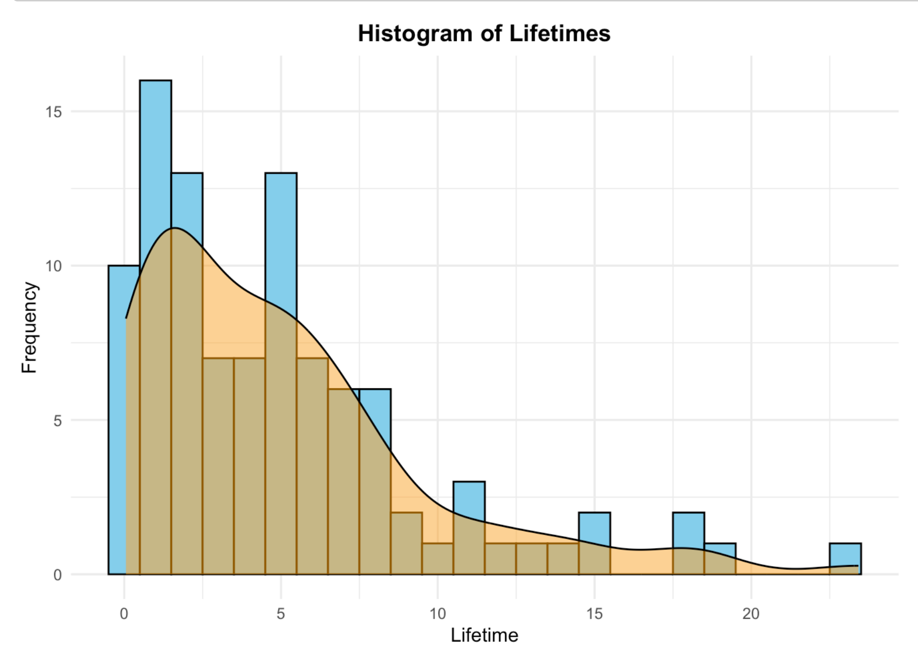
In most real-world applications, we will not know the population distribution for product lifetime. Instead, we will have to estimate it. Suppose that we suspect that the lifetime of a particular product should follow an exponential distribution. We collect data on the lifetimes, in years, for a sample of products in the data set **`lifetimes.txt`**

```r
plot_histogram <- function(data, binwidth = 1, plot_title) {
  hist_plot <-  ggplot(data, aes(x = Lifetime)) +
                geom_histogram(binwidth = binwidth, fill = "skyblue", color = "black") +
                geom_density(aes(y = ..count..), fill = "orange", alpha = 0.5) +
                labs(title = plot_title,
                     x = "Lifetime",
                     y = "Frequency") +
                theme_minimal() +
                theme(plot.title = element_text(hjust = 0.5, face = "bold"))

  return(hist_plot)
}

plot_qq <- function(data) {
  lambda <- 1 / mean(data$Lifetime)
  exp_quantiles <- qexp(ppoints(length(data$Lifetime)), rate = lambda)

  data_sorted <- sort(data$Lifetime)
  empirical_quantiles <- quantile(data_sorted, probs = ppoints(length(data$Lifetime)))

  qq_data <- data.frame(
    Empirical = empirical_quantiles,
    Exponential = exp_quantiles
  )

  qq_plot <- ggplot(qq_data, aes(x = Exponential, y = Empirical)) +
            geom_point(color = "blue1", size = 1.5, shape = 2) +
            geom_abline(slope = 1, intercept = 0, color = "brown4", linetype = "dashed") +
            labs(x = "Exponential Distribution Quantiles", y = "Empirical Quantiles",
                 title = "Q-Q Plot: Lifetime Data vs. Exponential Distribution") +
            theme_minimal() +
            theme(plot.title = element_text(hjust = 0.5, face = "bold"))
  return(qq_plot)
}
```

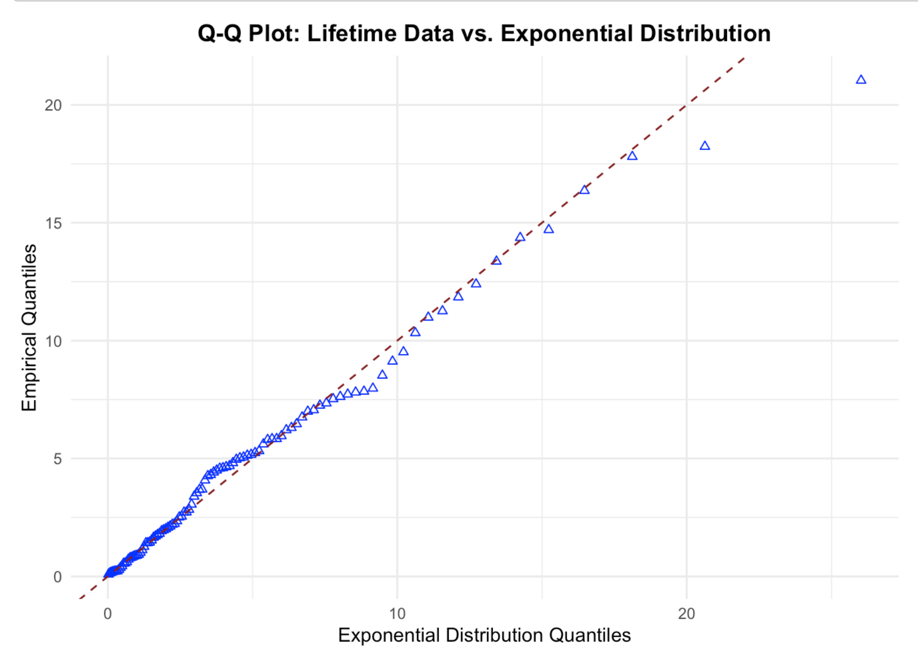
[Question A]

- *Let's examine the distribution of lifetimes. In R,construct a histogram of the lifetimes. Does it look reasonable to assume an exponential distribution for this data?*

```r
lifetimes_df <- read.delim("lifetime.txt") # Done
hist <- plot_histogram(lifetimes_df, plot_title = "Histogram of Lifetimes")
qq_plot <- plot_qq(lifetimes_df)

hist
```




```r
qq_plot
```



- It seems reasonable to assume that our data set follows an exponential distribution. The density plot (Plot A) shows a right-skewed graph, indicating that the distribution decreases as lifetime increases. Plot B confirms this observation. When we compare the quantiles of our dataset (`lifetimes_df`, shown in light blue) with the quantiles of an exponential distribution (the brown dashed line), we notice that the points roughly align along the straight line of the exponential quantiles. This suggests that our data could indeed be exponentially distributed.

[Question B]

- *To use an exponential distribution, we need to know λ. Using the relationship between λ and the mean of an exponential, estimate λ for this population based on the sample mean of the data.*

```r
sample_mean <- mean(lifetimes_df$Lifetime)

lambda_est <- 1/sample_mean

cat("[Estimated λ]: ", round(lambda_est, digits = 4))
```

```
## [Estimated λ]:  0.2036
```

[Question C]

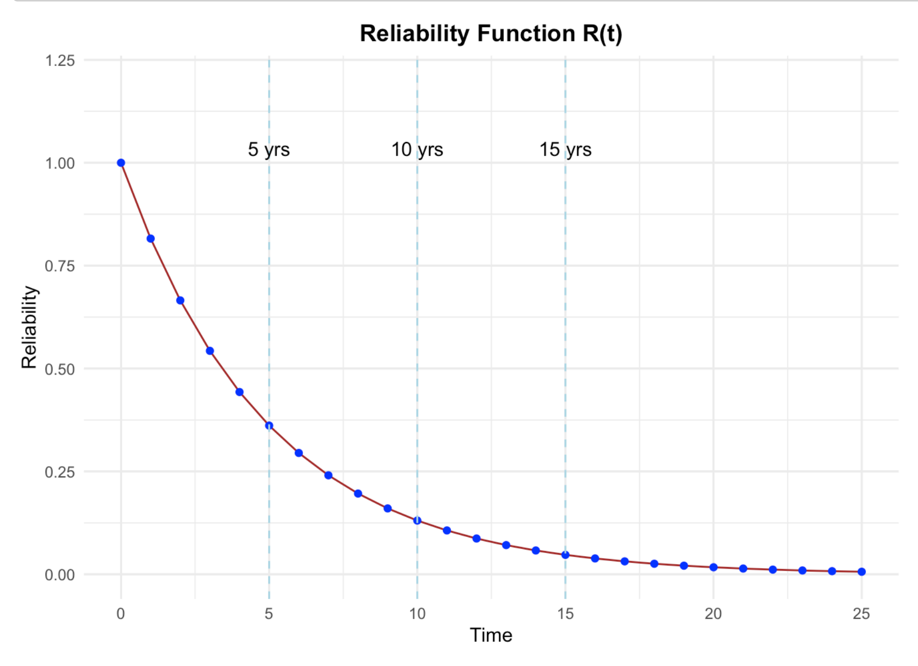
- *Now let's examine the reliability and failure rate for these items. Once again create a vector of time values `t`, this time ranging from 0 to 25, and calculate the reliability function and the failure rate (i.e., the hazard function) based on the exponential distribution, using the estimated value of λ based on your data.*

```r
time <- seq(0, 25)

# Our reliability function [R(t)]
reliability <- pexp(time, rate = lambda_est, lower.tail = FALSE)

# Our hazard function [z(t)]
hazard <- dexp(time, rate = lambda_est) / (1 - pexp(time, rate = lambda_est))
```

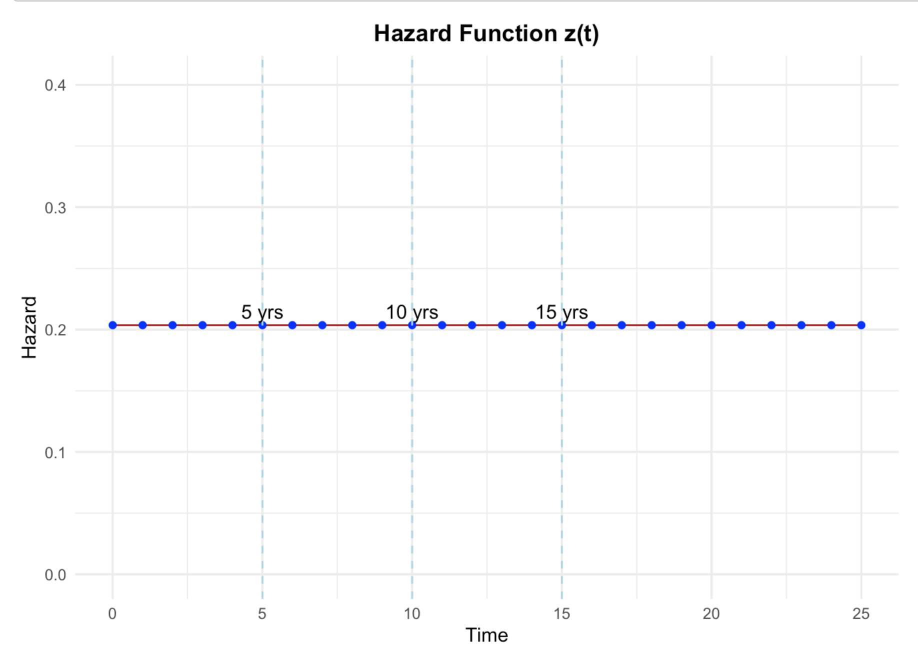
[Question D]

- *Graph the estimated reliability function for this product overtime.*

```r
plot_data <- data.frame(Time = time, Function = reliability)
plot_function(data_frame = plot_data, func_name = "Reliability", plot_title = "Reliability Function R(t)", lines
= TRUE)
```




```r
plot_data <- data.frame(Time = time, Function = hazard)
plot_function(data_frame = plot_data, func_name = "Hazard", plot_title = "Hazard Function z(t)", lines = TRUE)
```



[Question E]

- *Graph the estimated hazard function for this product overtime. Based on your result, what type of failure rate do products with exponential failure laws have (increasing, decreasing, or constant)?*
- Products subject to exponential failure laws exhibit a constant rate of failure throughout time. When we focus on the first vertical line at 5 years, we can't notice a decrease nor can we notice an increase in the hazard function. This trend persists across the remaining sections, highlighting a constant probability of failure over time.

[Question F]

- *For the exponential distribution, rather than using R's built in functionality, we could also use the formulas we've learned previously for `f(t)` and `F(t)` for an exponential distribution to find exact formulas for `R(t)` and `z(t)`. Find `R(t)` and `z(t)` this way. Do your results match the graphs you made?*

```r
# Summary:

# R(t) : e ^(-λt) or exp(-λ*t)
# z(t) : λ
```

Calculations

Probability Density Function (PDF):

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

Cumulative Distribution Function (CDF):

$$\begin{aligned}
\int_t^0 \lambda e^{-\lambda t}\,dt &= \int_0^t \lambda e^{-\lambda t}\,dt \\
&= \lambda \int_0^t e^{-\lambda t}\,dt \\
&= \lambda \left[-\frac{e^{-\lambda t}}{\lambda}\right]_0^t \\
&= \left[-e^{-\lambda t}\right]_0^t \\
&= -e^{-\lambda t} + 1 \\
&= 1 - e^{-\lambda t}
\end{aligned}$$

Reliability Function (Survival Function):

$$\begin{aligned}
R(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-\lambda t}) \\
&= e^{-\lambda t}
\end{aligned}$$

Hazard Function:

$$\begin{aligned}
z(t) &= \frac{f(t)}{R(t)} \\
&= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} \\
&= \lambda
\end{aligned}$$

The obtained results align with those from earlier sections of the problem. The function $R(t) = e^{-\lambda t}$ corresponds to the decreasing exponential curve depicted in part D. Similarly, $z(t) = \lambda$ corresponds to the other graph in part D, representing a constant failure rate. In essence, $z(t) = \lambda$ holds true for all t within the domain.

Jakob Balkovec