Parametric models continued: summary characteristics, fit, and estimation

Shannon Pileggi

STAT 417

OUTLINE

Summary characteristics of T

With a specified probability distribution for T, we can answer:

- What is the average time it takes for a motorist to react aggressively?
- What is the median age that an individual has his/her first drink?
- For patients who have survived 100 days after being diagnosed with lung cancer, what is the average time they have left to live?

Mean survival time

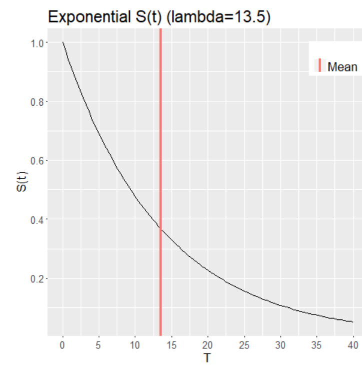
The **mean** (or **expected value**) of the time-to-event random variable T ($T \geq 0$), denoted by $E(T)$ or μ, is given by:

Motorist reaction time: compute $E(T)$

Suppose that the time until a motorist reacts aggressively is exponentially distributed with parameter $\lambda = 5.77$. Find the expected time it takes for a motorist to react aggressively.

Group Exercise

Another measure of the "center" of the distribution for T is the **median survival time**.



The median survival time is _______ the mean survival time of 13.5.

1. greater than
2. less than
3. equal to
4. not enough information to determine

Percentiles of T

- The **median survival time** is a special case of a *percentile* of the distribution of T (the 50^{th} percentile). This is the time at which 50% of the individuals have experienced the event.
- The p^{th} percentile of the distribution of the time-to-event random variable T is the value of T, denoted t_p such that:
- **Interpretation**: The p^{th} percentile, t_p, can be interpreted as:

Graphical representation of percentiles

Age at first drink: median survival time

Suppose that the age at first drink of alcohol follows an exponential distribution with parameter $\lambda = 13.5$. Find the median age at first drink. Recall that $S(t) = \exp(-t/\lambda)$.

Mean residual lifetime

- **Example: Lung Cancer**. Suppose an individual with lung cancer has survived 100 months since being diagnosed. Then how much time (on average) does this individual have left to live? We can set this problem up as a *conditional expectation*:
- This is an example of quantity known as the **mean residual lifetime (mrl)**, denoted mrl(t^*), and is defined to be the average remaining lifetime of all individuals given that the individuals have survived to time t^*, i.e:

Lung cancer example: mean residual lifetime

If the time until death from cancer follows an exponential distribution with parameter $\lambda = 130.18$, then find the average remaining lifetime for those patients who have survived at least 100 days.

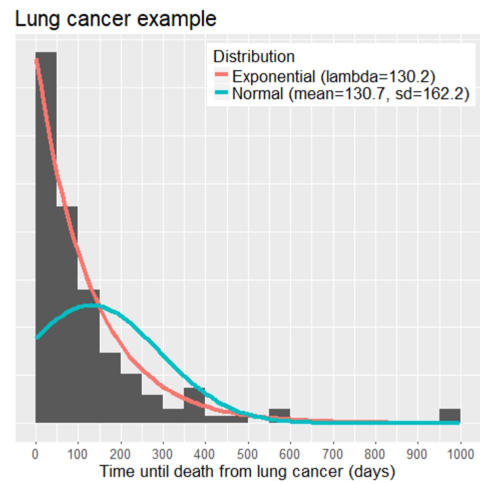
Considerations for parametric models

1.

2.

3.

Veteran's Administration Lung Cancer Study (VALCS)



Lifetimes (since treatment) of 137 lung cancer patients taken from another lung cancer study with two different probability density curves superimposed (exponential and normal).

Strategies for fitting parametric models

1.

2.

3.

Probability plots for any distribution

- Can construct a probability plot for any distribution
- With incomplete data, special techniques are used in the probability plot to account for censoring
- Assess if an exponential distribution fits the data well

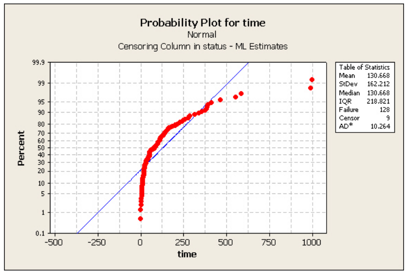
  H_0:

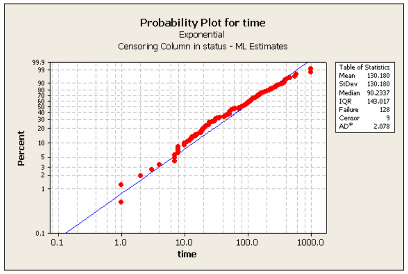
  H_a:

  1.

  2.

VALCS study: probability plots

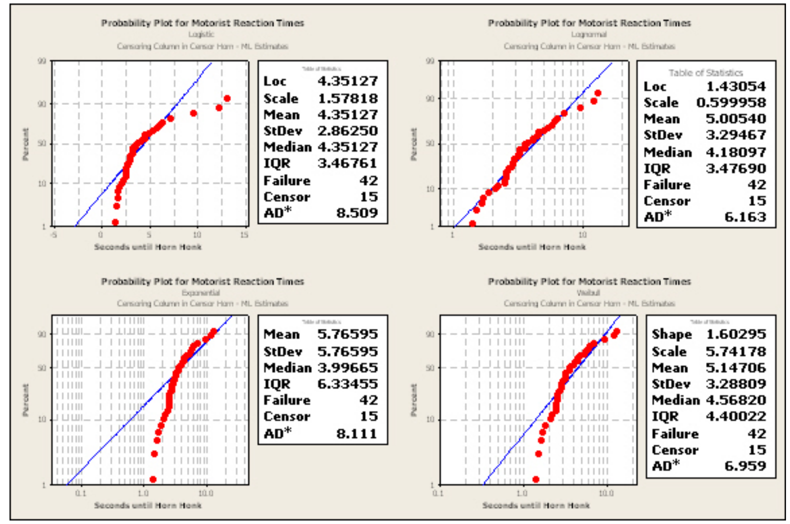




Which distribution appears to fit the data better, and why?

1. normal
2. exponential

Motorist reaction times: probability plots



Parameter estimation

- We have worked through many examples of parametric models for T, and in each example, the value(s) of the parameter(s) were given. For example, the time until death for the lung cancer data was assumed to follow an exponential distribution with parameter $\lambda = 130.18$.
- Also, in the previous probability plots (observe the box on the right side), values of the parameters were provided. How are these values computed?

Maximum likelihood estimation

Suppose we have a random sample of observations $T_1, T_2, \ldots, T_n$ on n individuals, each with the same pdf $f(t_i)$, $i = 1, \ldots, n$ and parameter θ (or possibly more than one parameter).

- Then **likelihood function** is given by:
- The likelihood function tells how likely the observed sample is as a function of the parameter values.
- By *maximizing* the likelihood function, we get the parameter value(s) that most likely would generate our given observed sample.

Maximum likelihood estimation, cont.

Goal: Find an estimate of θ that maximizes the likelihood function, $L(\theta)$.

Process:
1.
2.
3.

MLE example: exponential distribution

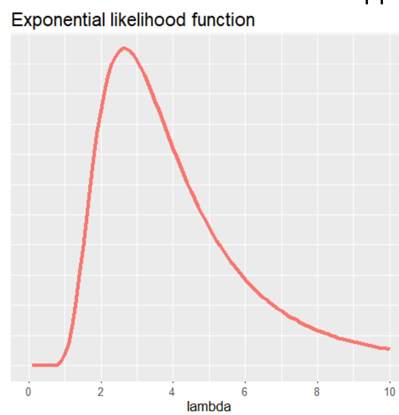
Suppose we obtain a random sample of 5 observations, $T_1, T_2, \ldots, T_5$, from an exponential distribution with parameter λ and pdf:

$$f(t) = \frac{1}{\lambda} e^{-t/\lambda}$$

1. Construct the likelihood function $L(\lambda)$.

MLE example: exponential distribution, cont.

2. Suppose the observed values of $T_1, T_2, \ldots, T_5$ are $5.3, 4.8, 0.4, 2.3$, and 0.4. Plug these into $L(\lambda)$ and examine the graph of $L(\lambda)$ as a function of λ. At which value of λ does the likelihood function appear to be maximized?



MLE example: exponential distribution, cont.

3. Find the value of λ that maximizes $L(\lambda)$.

Likelihood function for right censored data

If there are right censored event times, then can't simply compute the likelihood function L as the product of the pdf's. Contributions to the likelihood function differ:

- If the event time for observation i is complete, then that observation contributes:
- If the event time for observation i is (right) censored, then that observation contributes:
- Putting the contributions of all observation together, the likelihood function is given by:

MLE for right censored data

The following data consist of the times (in months) until relapse of 10 bone marrow transplant patients. Patients 7-10 were free of relapse at the end of the study. Suppose time to relapse follows an exponential distribution with parameter λ.

Patient	1	2	3	4	5	6	7	8	9	10
Relapse Time	5	8	12	24	32	17	16^+	17^+	19^+	30^+

1. Construct the likelihood function for the parameter λ and find the MLE for λ.
2. Assuming that time until relapse follows an exponential probability distribution with λ found in Part (1), what is the mean time until relapse?

MLE for right censored data, cont.

MLE for right censored data, cont.

2. Assuming that time until relapse follows an exponential probability distribution with λ found in Part (1), what is the mean time until relapse?

Parametric distribution analysis in Minitab

- Along with the probability plots, the MLE's, and summary characteristics about the distribution of T are also computed in Minitab when `Parametric Distribution Analysis` is performed.
- The MLE's are provided in the box to the right of the plot, as well as in the output in the Minitab Session Window.
- The mean, median, quartiles, and several other percentiles are displayed in the Session Window.

Motorist reaction times in Minitab (lognormal model)

```
Variable: Seconds until Horn Honk

Censoring Information  Count
Uncensored value          42
Right censored value      15

Censoring value: Censor Horn = 0
```


Parametric distribution analysis in Minitab, cont.

```
Estimation Method: Maximum Likelihood

Distribution:   Lognormal

Parameter Estimates
                      Standard    95.0% Normal CI
Parameter  Estimate      Error      Lower     Upper
Location    1.43054  0.0851775    1.26360   1.59749
Scale      0.599958  0.0665560   0.482718  0.745673

Log-Likelihood = -100.729

Goodness-of-Fit
Anderson-Darling (adjusted) = 6.163
...
```


Parametric distribution analysis in Minitab, cont.

```
Characteristics of Distribution

                                      Standard   95.0% Normal CI
                           Estimate      Error     Lower     Upper
Mean(MTTF)                  5.00540   0.500060   4.11529   6.08804
Standard Deviation          3.29467   0.672961   2.20774   4.91673
Median                      4.18097   0.356124   3.53813   4.94060
First Quartile(Q1)          2.78954   0.249485   2.34102   3.32400
Third Quartile(Q3)          6.26644   0.643405   5.12418   7.66334
Interquartile Range(IQR)    3.47690   0.541312   2.56254   4.71752
```


Parametric distribution analysis in Minitab, cont.

```
Table of Percentiles
                     Standard    95.0% Normal CI
Percent  Percentile     Error     Lower     Upper
      1     1.03545  0.169643  0.751052   1.42753
      2     1.21943  0.181146  0.911401   1.63155
      3     1.35276  0.188314   1.02974   1.77711
...
```

What do you think the first line means?

1. $Pr(T < 1.03545) = 0.01$
2. $Pr(T < 1) = 0.0103545$
3. $Pr(T > 1.03545) = 0.01$
4. $Pr(T > 1) = 0.0103545$

Summary characteristics

Evaluating fit

Parameter estimation

Minitab

Summary

Recap

- Features of time-to-event data and survival studies (e.g. censoring, truncation, etc.)
- Parametric models for the time-to-event random variable T: $f(t)$, $F(t)$, $S(t)$, $h(t)$, $H(t)$
- Descriptive measures for T: $E(T)$, t_p, and, mrl(t^*)
- Strategies for selecting a probability model: probability plots, AD test statistic
- Parameter estimation: maximum likelihood estimation

Future

Nonparametric methods for summarizing and describing samples of time-to-event data:

- Estimators for $S(t)$, $h(t)$, and $H(t)$.
- Estimates of $E(T)$ and t_p
- Inference methods for the estimator of $S(t)$
- Inference procedures for comparing two or more survival experiences.
- Survival analysis in R