TMA4275 LIFETIME ANALYSIS

Slides 2: General concepts for lifetime modeling

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

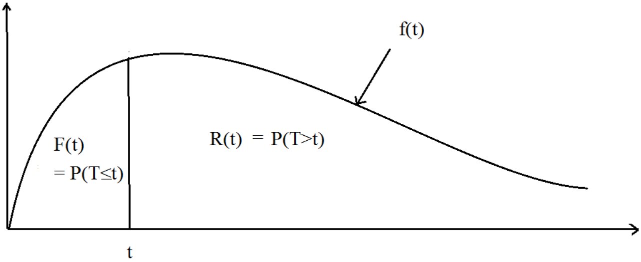
http://www.math.ntnu.no/∼bo/
bo@math.ntnu.no

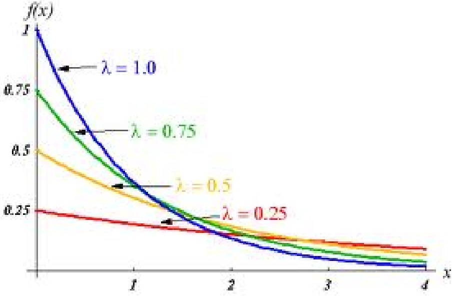
NTNU, Spring 2014

LIFETIME

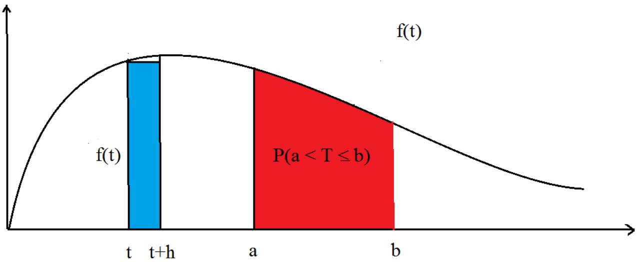
The lifetime T of an individual or unit is a *positive* and *continuously distributed* random variable.

- The probability density function (pdf) is usually called $f(t)$,
- the cumulative distribution function (cdf) $F(t)$ is then given by $F(t) = P(T \leq t) = \int_0^t f(u)du$,
- the reliability (or: survival) function is defined as $R(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(u)du$.





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



$$f(t) = F'(t)$$

$$P(a < T \leq b) = \int_a^b f(u)du = F(b) - F(a)$$

$$P(t < T \leq t + h) = \int_t^{t+h} f(u)du \approx f(t) \cdot h$$

Hence,

$$f(t) \approx \frac{P(t < T \leq t + h)}{h}$$

Suppose we know that unit is alive (functioning) at time t, i.e. $T > t$.

Then it is of interest to consider

$$P(t < T \leq t + h | T > t) = \frac{P(t < T \leq t + h)}{P(T > t)} \approx \frac{f(t)h}{R(t)}$$

(Recall *conditional probability*: $P(A|B) = P(A \cap B)/P(B)$.

From this we define the *hazard function* (also called *hazard rate* or *failure rate*) of T at time t by:

$$z(t) = \lim_{h \to 0} \frac{P(t < T \leq t + h | T > t)}{h} = \frac{f(t)}{R(t)}$$

Example: For the exponential distribution we have
$f(t) = \lambda e^{-\lambda t}$ and $R(t) = e^{-\lambda t}$, so

$$z(t) = \frac{f(t)}{R(t)} = \lambda \text{ (not depending on time!)}.$$

Since $F(t) = 1 - R(t)$ we get, $f(t) = F'(t) = -R'(t)$, and hence

$$z(t) = \frac{f(t)}{R(t)} = -\frac{R'(t)}{R(t)}$$

Thus we can write,

$$\frac{d}{dt}\left(\ln R(t)\right) = -z(t)$$

$$\Rightarrow \ln R(t) = -\int_0^t z(u)du + c$$

$$\Rightarrow R(t) = e^{-\int_0^t z(u)du + c}$$

Since $R(0) = 1$, we have $c = 0$, so

$$R(t) = e^{-\int_0^t z(u)du} \equiv e^{-Z(t)}$$

where $Z(t) = \int_0^t z(u)du$ is called the *cumulative hazard function*.

Recall from last slide:

- $Z(t) = \int_0^t z(u)du$
- $z(t) = Z'(t)$
- $R(t) = e^{-Z(t)}$

Since $f(t) = F'(t) = -R'(t)$, it follows that

$$f(t) = z(t)e^{-\int_0^t z(u)du} = z(t)e^{-Z(t)} \tag{1}$$

For exponential distribution:

$$Z(t) = \int_0^t \lambda du = \lambda t$$

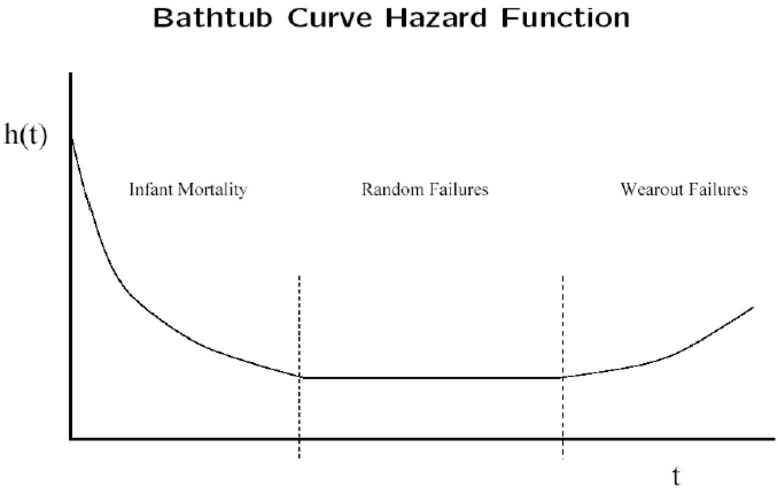
so (1) gives (the well known formula)

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

OVERVIEW OF FUNCTIONS DESCRIBING DISTRIBUTION OF LIFETIME T

Function	Formula	Exponential distr
Density (pdf)	f(t)	$= \lambda e^{-\lambda t}$
Cum. distr. (cdf)	F(t)	$= 1 - e^{-\lambda t}$
Rel/surv function	$R(t) = 1 - F(t)$	$= e^{-\lambda t}$
Hazard function	$z(t) = f(t)/R(t)$	$= \lambda$
Cum hazard function	$Z(t) = \int_0^t z(u)du$	$= \lambda t$
	$R(t) = e^{-Z(t)}$	$= e^{-\lambda t}$
	$f(t) = z(t)e^{-Z(t)}$	$= \lambda e^{-\lambda t}$

1. Suppose the reliability function of T is $R(t) = e^{-t^{1.7}}$.
   Find the functions $F(t), f(t), z(t), Z(t)$.
2. Show that if you get to know only one of the functions $R(t), F(t), f(t), z(t), Z(t)$, then you can still compute all the other!


Bathtub Curve Hazard Function
h(t)
Infant Mortality
Random Failures
Wearout Failures
t

MORE ON THE HAZARD FUNCTION

Recall that $z(t) = \lim_{h\to 0} \frac{P(t<T\leq t+h|T>t)}{h}$.

Thus

$$z(t)h \approx P(t < T \leq t+h|T > t) = P(\text{fail in } (t, t+h)| \text{ alive at } t)$$

Suppose a typical T is large compared to time unit. Then for $h = 1$:

$$z(t) \approx P(t < T \leq t+1|T > t) = P(\text{fail in next time unit } |\text{alive at } t)$$

Thus: Suppose we have n units of age t. How many can we expect to fail in next time unit?

$$e = n \cdot z(t)$$

In practice: Ask an expert: "If you have 100 components (of specific type) of age 1000 hours. How many do you expect to fail in the next hour"? Answer is, say, "2". Looking at $e = n \cdot z(t)$ we estimate;

$$\hat{z}(1000) = \frac{2}{100} = 0.02$$

EXERCISES

Let T be the lifetime of a Norwegian measured in years.

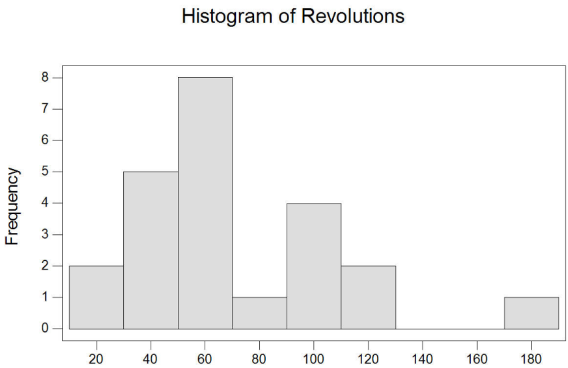
Let $z_M(t)$ be the hazard function for a male person as a function of the age t, while $z_F(t)$ is the corresponding function for a female.

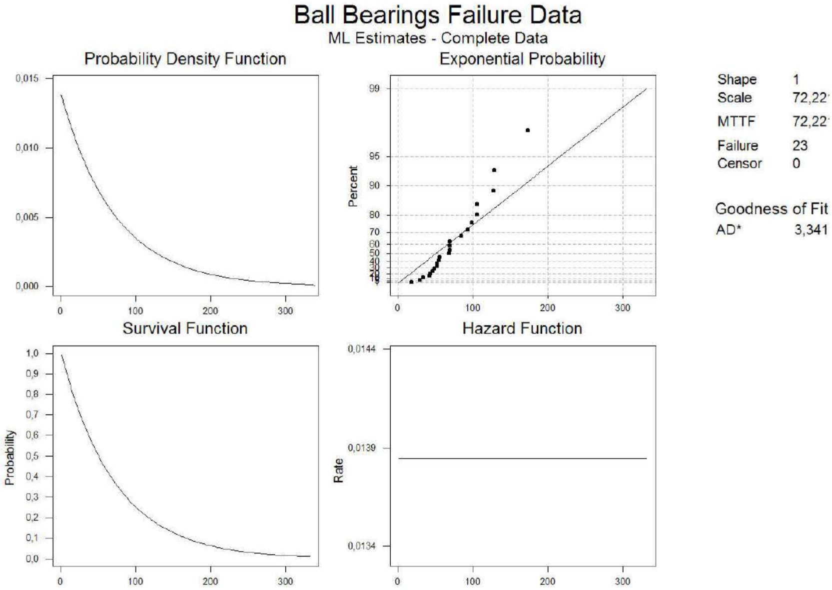
Look at the Mortality tables of Slides 1 and estimate $z_M(21)$ and $z_F(21)$. Compare them and comment.

Do the same at age 72 years.

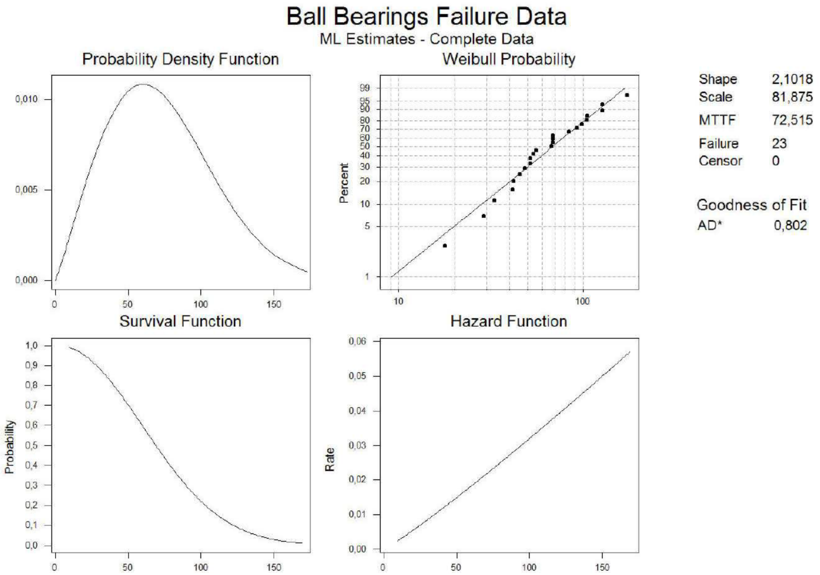
17,88	28,92	33,00	41,52	42,12	45,60	48,40	51,84
51,96	54,12	55,56	67,80	68,64	68,64	68,88	84,12
93,12	98,64	105,12	105,84	127,92	128,04	173,40	

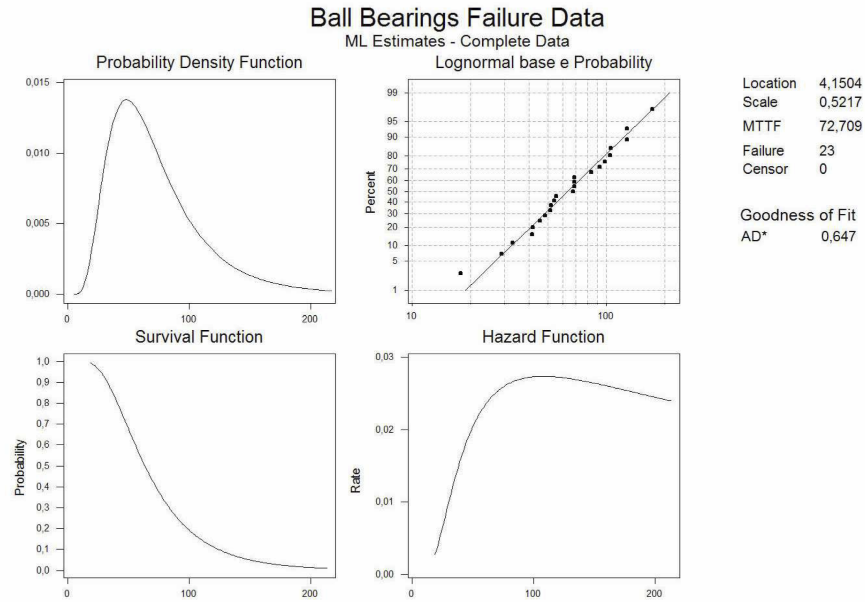
Question: *How can we fit a parametric lifetime model to these data?*



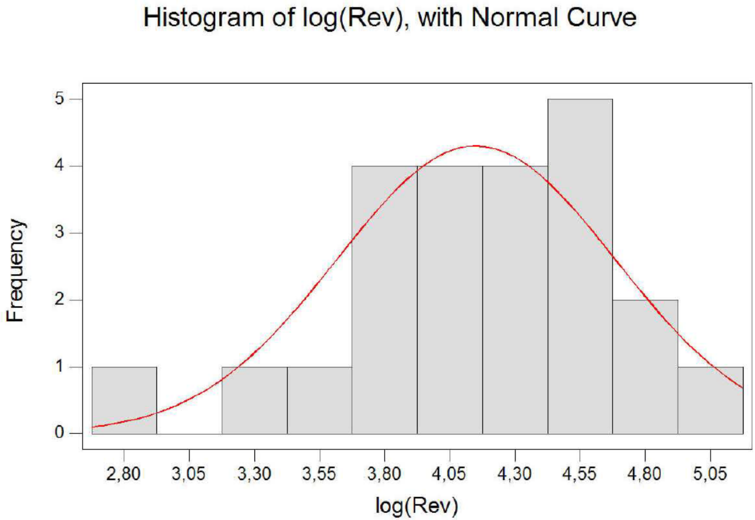

Ball Bearings Failure Data
ML Estimates - Complete Data
Probability Density Function
0,015
0,010
0,005
0,000
0
100
200
300
Exponential Probability
Percent
99
95
90
80
70
60
50
40
0
100
200
300
Survival Function
Probability
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0,0
0
100
200
300
Hazard Function
Rate
0,0144
0,0139
0,0134
0
100
200
300
Shape 1
Scale 72,22
MTTF 72,22
Failure 23
Censor 0
Goodness of Fit
AD* 3,341

BB-DATA: WEIBULL DISTRIBUTION (MINITAB)

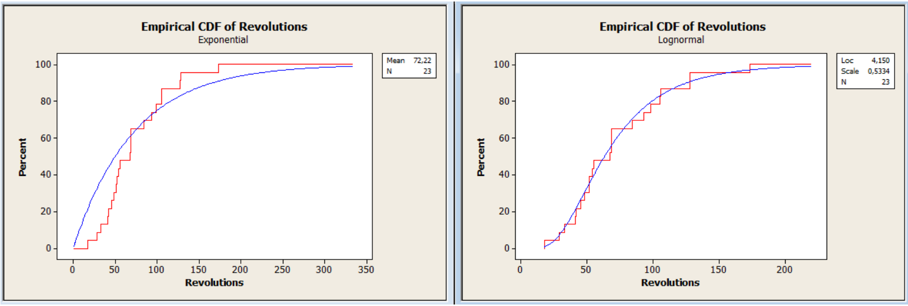



Ball Bearings Failure Data
ML Estimates - Complete Data
Probability Density Function
0,015
0,010
0,005
0,000
0
100
200
Lognormal base e Probability
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
10
100
Location 4,1504
Scale 0,5217
MTTF 72,709
Failure 23
Censor 0
Goodness of Fit
AD* 0,647
Survival Function
Probability
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0,0
0
100
200
Hazard Function
Rate
0,03
0,02
0,01
0,00
0
100
200

BB-DATA: HISTOGRAM OF LOG-LIFETIMES



BB-DATA: EMPIRICAL DISTRIBUTION COMPARED TO PARAMETRIC FITS



MOTIVATION FOR EXPONENTIAL DISTRIBUTION

- Simplest distribution used in the analysis of reliability data.
- Has the important characteristic that its hazard function is constant (does not depend on time t).
- Popular distribution for some kinds of electronic components (e.g., capacitors or robust, high-quality integrated circuits).
- Might be useful to describe failure times for components that exhibit physical wearout only after expected technological life of the system, in which the component would be replaced.

MOTIVATION FOR WEIBULL DISTRIBUTION

- The theory of extreme values shows that the Weibull distribution can be used to model the minimum of a large number of independent positive random variables from a certain class of distributions.
    - Failure of the weakest link in a chain with many links with failure mechanisms (e.g. fatigue) in each link acting approximately independently.
    - Failure of a system with a large number of components in series and with approximately independent failure mechanisms in each component.
- The more common justification for its use is empirical: the Weibull distribution can be used to model failure-time data with a decreasing or an increasing hazard function.

MOTIVATION FOR LOGNORMAL DISTRIBUTION

- The lognormal distribution is a common model for failure times.
- It can be justified for a random variable that arises from the product of a number of identically distributed independent positive random quantities (remember central limit theorem for sum of normals).
- It has been suggested as an appropriate model for failure time caused by a degradation process with combinations of random rates that combine multiplicatively.
- Widely used to describe time to fracture from fatigue crack growth in metals.