STK4080/9080
SURVIVAL AND EVENT HISTORY ANALYSIS
Slides 2: Survival analysis: basic concepts and examples

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

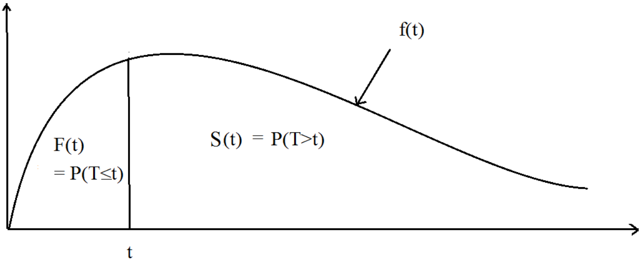
https://www.ntnu.edu/employees/bo.lindqvist
bo.lindqvist@ntnu.no
boli@math.uio.no

University of Oslo, Spring 2021

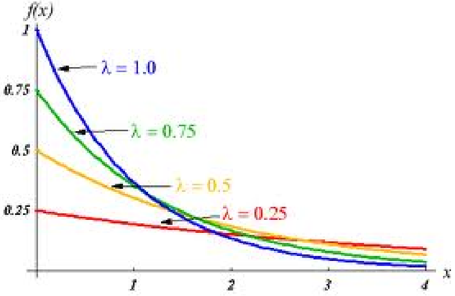
LIFETIME (SURVIVAL TIME)

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 survival) function is defined as $S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(u)du$.

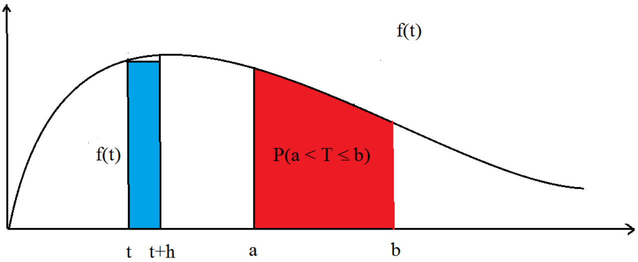


EXAMPLE: EXPONENTIAL DISTRIBUTION



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

INTERPRETATION OF DENSITY FUNCTION



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

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

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

Hence,

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

FROM DENSITY TO HAZARD FUNCTION OF T

From last slide,

$$P(t < T \leq t + h) \approx f(t) \cdot h$$

If we know that the unit is alive (functioning) at time t, i.e. $T > t$, we may be interested in the conditional probability

$$P(t < T \leq t + h | T > t).$$

Using the *conditional probability* formula: $P(A|B) = P(A \cap B)/P(B)$, we get

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

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

$$\alpha(t) = \frac{f(t)}{S(t)}$$

HAZARD FUNCTION OF T

Formal definition of hazard function is

$$\alpha(t) = \lim_{h \to 0} \frac{P(t < T \le t + h | T > t)}{h} = \frac{f(t)}{S(t)}$$

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

$$\alpha(t) = \frac{f(t)}{S(t)} = \lambda \quad \textit{(not depending on time!)}.$$

MORE ON THE HAZARD FUNCTION

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

Thus

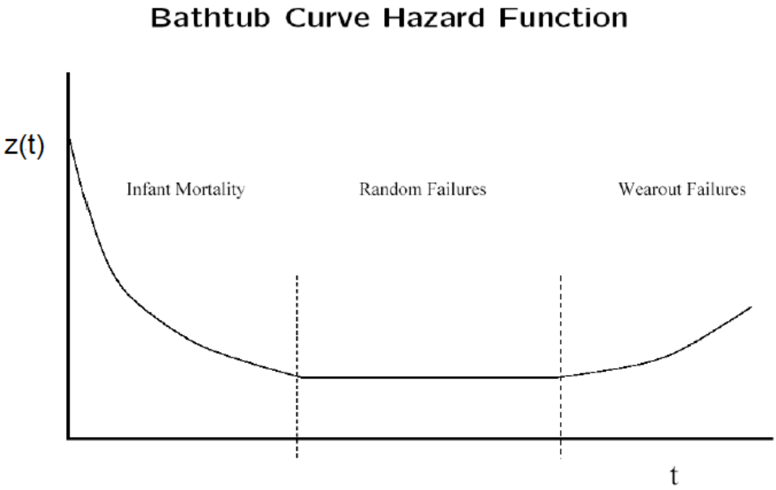
$$\alpha(t)h \approx P(t < T \leq t + h|T > t) = P(\text{fail in } (t, t + h)| \text{ alive at } t)$$

Suppose a typical value of T is large compared to time unit.

Then for $h = 1$:

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

BATHTUB CURVE RELIABILITY ANALYSIS)



EXERCISE

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

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

Look at the Mortality tables of the next slides and estimate $\alpha_M(21)$ and $\alpha_F(21)$. Compare them and comment.

Do the same at age 72 years.

(Hint: Explain why $\alpha_M(t)$ and $\alpha_F(t)$ can be interepreted as the probability of dying within one year for a male and female, respectively, who has reached the age of t years).

MORTALITY TABLE - DEATH HAZARD BY AGE

Statistisk sentralbyrå
Statistics Norway

Tabell

Lukk tabell | Last ned tabellen i Excel | Last ned tabellen som CSV-fil

Dødelighetstabeller

	2012											
	Levende ved alder x, lx			Døde i alder x til x+1, dx			Forventet gjenstående levetid ved alder x, ex			Dødssannsynlighet for alder x (Promille) (Uglattet), qx		
	Begge kjønn	Menn	Kvinner	Begge kjønn	Menn	Kvinner	Begge kjønn	Menn	Kvinner	Begge kjønn	Menn	Kvinner
0 år	100 000	100 000	100 000	248	280	214	81,45	79,42	83,41	2,479	2,798	2,143
1 år	99 752	99 720	99 786	27	25	30	80,65	78,64	82,59	0,274	0,251	0,299
2 år	99 725	99 695	99 756	11	9	13	79,67	77,66	81,61	0,110	0,092	0,129
3 år	99 714	99 686	99 743	11	15	6	78,68	76,67	80,62	0,110	0,153	0,065
4 år	99 703	99 671	99 737	5	0	10	77,69	75,68	79,63	0,048	0,000	0,099
5 år	99 698	99 671	99 727	15	22	7	76,69	74,68	78,64	0,146	0,221	0,066
6 år	99 683	99 649	99 720	2	0	3	75,70	73,70	77,64	0,016	0,000	0,033
7 år	99 682	99 649	99 717	7	6	7	74,70	72,70	76,65	0,066	0,065	0,068
8 år	99 675	99 642	99 710	10	10	10	73,71	71,70	75,65	0,099	0,097	0,102
9 år	99 665	99 633	99 700	8	3	14	72,71	70,71	74,66	0,084	0,033	0,137
10 år	99 657	99 629	99 686	13	13	14	71,72	69,71	73,67	0,134	0,132	0,137
11 år	99 644	99 616	99 673	11	10	13	70,73	68,72	72,68	0,114	0,095	0,133
12 år	99 632	99 607	99 659	8	9	7	69,74	67,73	71,69	0,079	0,093	0,065
13 år	99 624	99 597	99 653	13	6	20	68,74	66,73	70,69	0,127	0,062	0,196
14 år	99 612	99 591	99 633	16	21	10	67,75	65,74	69,71	0,158	0,216	0,097
15 år	99 596	99 570	99 624	15	18	13	66,76	64,75	68,71	0,154	0,180	0,127
16 år	99 581	99 552	99 611	24	27	22	65,77	63,76	67,72	0,245	0,267	0,221
17 år	99 556	99 525	99 589	18	21	16	64,79	62,78	66,74	0,185	0,209	0,159
18 år	99 538	99 504	99 573	34	53	13	63,80	61,79	65,75	0,339	0,537	0,127
19 år	99 504	99 451	99 560	21	29	13	62,82	60,82	64,76	0,214	0,295	0,126
20 år	99 483	99 422	99 548	48	66	28	61,84	59,84	63,76	0,480	0,667	0,279
21 år	99 435	99 355	99 520	46	69	21	60,86	58,88	62,78	0,459	0,694	0,212
22 år	99 389	99 286	99 499	47	72	21	59,89	57,92	61,79	0,473	0,726	0,211
23 år	99 342	99 214	99 478	30	49	9	58,92	56,96	60,81	0,298	0,499	0,091
[illegible]	[illegible]	[illegible]	[illegible]	38	57	[illegible]	[illegible]	55,99	59,81	0,379	0,570	0,184

MORTALITY TABLE - DEATH HAZARD BY AGE

72 år	82 194	78 643	85 922	1 676	1 970	1 387	14,24	12,96	15,27	20,390	25,047	16,139
73 år	80 518	76 673	84 535	1 659	1 953	1 377	13,53	12,28	14,52	20,609	25,468	16,285
74 år	78 859	74 720	83 158	1 765	1 952	1 592	12,80	11,59	13,75	22,386	26,124	19,140
75 år	77 094	72 768	81 567	2 044	2 365	1 746	12,08	10,89	13,01	26,516	32,494	21,409
76 år	75 050	70 404	79 820	2 046	2 343	1 779	11,40	10,24	12,28	27,262	33,274	22,293
77 år	73 004	68 061	78 041	2 381	2 794	2 014	10,70	9,57	11,55	32,611	41,045	25,807
78 år	70 623	65 268	76 027	2 716	3 088	2 385	10,05	8,96	10,84	38,459	47,307	31,370
79 år	67 907	62 180	73 642	2 891	3 405	2 434	9,43	8,38	10,18	42,567	54,768	33,050
80 år	65 016	58 775	71 208	2 997	3 420	2 638	8,83	7,84	9,51	46,098	58,195	37,045
81 år	62 019	55 354	68 570	3 398	3 783	3 082	8,23	7,29	8,86	54,791	68,343	44,952
82 år	58 621	51 571	65 488	3 661	4 101	3 298	7,68	6,79	8,25	62,457	79,529	50,365
83 år	54 960	47 470	62 189	4 013	4 211	3 887	7,15	6,33	7,66	73,026	88,712	62,505
84 år	50 946	43 259	58 302	3 867	4 172	3 667	6,68	5,90	7,14	75,895	96,434	62,888
85 år	47 080	39 087	54 636	3 997	4 117	3 965	6,19	5,48	6,58	84,906	105,321	72,570
86 år	43 082	34 970	50 671	4 236	4 079	4 441	5,71	5,06	6,06	98,324	116,649	87,642
87 år	38 846	30 891	46 230	4 213	4 096	4 399	5,28	4,67	5,59	108,449	132,591	95,146
88 år	34 633	26 795	41 831	4 292	3 894	4 735	4,86	4,30	5,13	123,925	145,341	113,199
89 år	30 341	22 901	37 096	4 114	3 876	4 442	4,48	3,95	4,72	135,592	169,266	119,733
90 år	26 227	19 024	32 655	4 367	3 703	5 047	4,10	3,65	4,29	166,501	194,620	154,562
91 år	21 861	15 322	27 607	3 650	3 069	4 244	3,82	3,41	3,98	166,963	200,273	153,723
92 år	18 211	12 253	23 363	3 732	2 932	4 485	3,49	3,14	3,62	204,937	239,318	191,977
93 år	14 479	9 321	18 878	3 015	2 318	3 665	3,26	2,97	3,36	208,220	248,707	194,145
94 år	11 464	7 003	15 213	2 698	1 823	3 458	2,99	2,79	3,05	235,366	260,306	227,292
95 år	8 766	5 180	11 755	2 301	1 559	2 958	2,75	2,60	2,80	262,529	300,960	251,671
96 år	6 464	3 621	8 797	1 871	1 134	2 496	2,55	2,51	2,57	289,392	313,095	283,759
97 år	4 594	2 487	6 301	1 432	795	1 952	2,39	2,42	2,39	311,787	319,717	309,819
98 år	3 161	1 692	4 349	1 071	646	1 426	2,25	2,32	2,24	338,639	381,843	327,848
99 år	2 091	1 046	2 923	711	316	1 017	2,14	2,45	2,08	340,030	302,252	347,939
100 år	1 380	730	1 906	495	209	711	1,99	2,30	1,93	358,820	286,341	373,123
101 år	885	521	1 195	344	172	479	1,82	2,02	1,78	388,786	329,680	401,004
102 år	541	349	716	234	90	331	1,66	1,76	1,63	432,013	258,075	462,088
103 år	307	259	385	155	148	189	1,54	1,20	1,60	504,285	571,937	491,822
104 år	152	111	196	55	70	61	1,59	1,14	1,67	363,091	632,121	313,445
105 år	97	41	134	28	11	40	1,21	1,24	1,20	290,260	264,859	295,930
106 år	69	30	95	35	17	47	0,50	0,50	0,50	501,251	550,671	494,864

Statistikkbanken kildetabell 07902

USEFUL RELATIONS BETWEEN FUNCTIONS DESCRIBING T

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

$$\alpha(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

Thus we can write,

$$\frac{d}{dt}\left(\ln S(t)\right) = -\alpha(t)$$

$$\Rightarrow \ln S(t) = -\int_0^t \alpha(u)du + c$$

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

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

$$S(t) = e^{-\int_0^t \alpha(u)du} \equiv e^{-A(t)}$$

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

USEFUL RELATIONS (CONT.)

Recall from last slide:

- $A(t) = \int_0^t \alpha(u)du$
- $\alpha(t) = A'(t)$
- $S(t) = e^{-A(t)}$

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

$$f(t) = \alpha(t)e^{-\int_0^t \alpha(u)du} = \alpha(t)e^{-A(t)} \qquad (1)$$

For exponential distribution:

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

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	$S(t) = 1 - F(t)$	$= e^{-\lambda t}$
Hazard function	$\alpha(t) = f(t)/S(t)$	$= \lambda$
Cum hazard function	$A(t) = \int_0^t \alpha(u)du$	$= \lambda t$
	$S(t) = e^{-A(t)}$	$= e^{-\lambda t}$
	$f(t) = \alpha(t)e^{-A(t)}$	$= \lambda e^{-\lambda t}$

EXERCISES

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

EXPECTED VALUE AND VARIANCE OF LIFETIMES

For a lifetime T we have

$$E(T) = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt$$

(The last equality can be proven by partial integration, noting that $R'(t) = -f(t)$. Try to do it! You will need that $\lim_{t\to\infty} t\, R(t) = 0$ which holds if $E(T) < \infty$.)

$$\begin{aligned} Var(T) &= \int_0^\infty (t - E(T))^2 f(t)dt \\ &= \int_0^\infty t^2 f(t)dt - (E(T))^2 \\ &= E(T^2) - (E(T))^2 \end{aligned}$$

$$\mathrm{SD}(T) = (Var(T))^{1/2}$$

EXAMPLE: EXPONENTIAL DISTRIBUTION

Let T be exponentially distributed with density $f(t) = \lambda e^{-\lambda t}$. Then you may check the following computations:

$$\begin{aligned} E(T) &= \int_0^\infty t\lambda e^{-\lambda t} dt = \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda} \\ \text{Var}(T) &= E(T^2) - (E(T))^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \\ \text{SD}(T) &= \frac{1}{\lambda} \end{aligned}$$

Thus: For an exponentially distributed lifetime,

$$E(T) = 1/\text{failure rate}$$

WEIBULL DISTRIBUTION

The lifetime T is Weibull-distributed with *shape* parameter $k > 0$ and *scale* parameter $\theta > 0$, written $T \sim \text{Weib}(k, \theta)$, if

$$S(t) = e^{-(\frac{t}{\theta})^k}$$

From this we can derive:

$$\begin{aligned} A(t) &= \left(\frac{t}{\theta}\right)^k \\ \alpha(t) &= \frac{k}{\theta}\left(\frac{t}{\theta}\right)^{k-1} \\ f(t) &= \alpha(t)e^{-A(t)} = \frac{k}{\theta}\left(\frac{t}{\theta}\right)^{k-1} e^{-(\frac{t}{\theta})^k} \end{aligned}$$

$k = 1$ *corresponds to the exponential distribution;*
$k < 1$ *gives a decreasing failure rate (DFR);*
$k > 1$ *gives an increasing failure rate (IFR).*

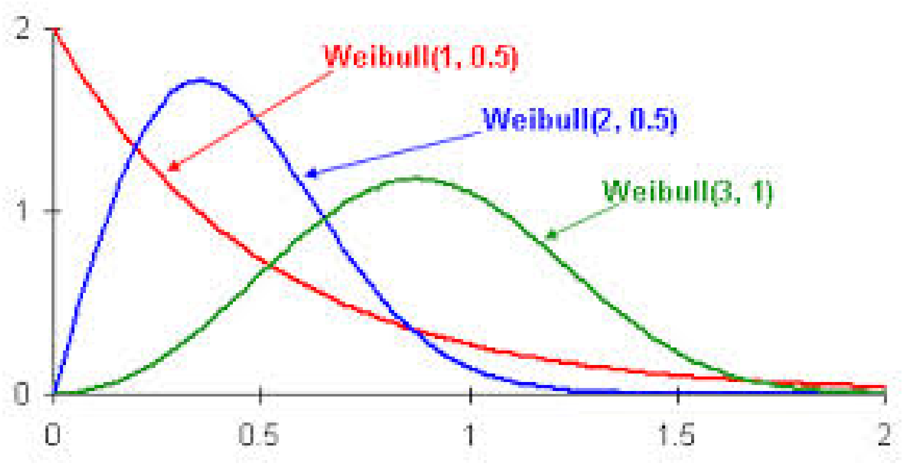
WEIBULL DISTRIBUTION (CONT.)

$$E(T) = \int_0^\infty S(t)dt = \int_0^\infty e^{-(\frac{t}{\theta})^k} dt = \cdots = \theta \cdot \Gamma\left(\frac{1}{k} + 1\right)$$

where $\Gamma(\cdot)$ is the gamma-function defined by $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u} du$.

$$\begin{aligned} \text{Var}(T) &= \theta^2 \left(\Gamma\left(\frac{2}{k} + 1\right) - \Gamma^2\left(\frac{1}{k} + 1\right) \right) \\ SD(T) &= \theta \left(\Gamma\left(\frac{2}{k} + 1\right) - \Gamma^2\left(\frac{1}{k} + 1\right) \right)^{1/2} \end{aligned}$$

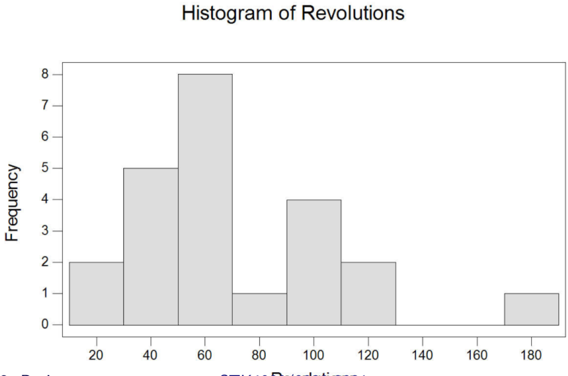
WEIBULL DISTRIBUTION (CONT.)



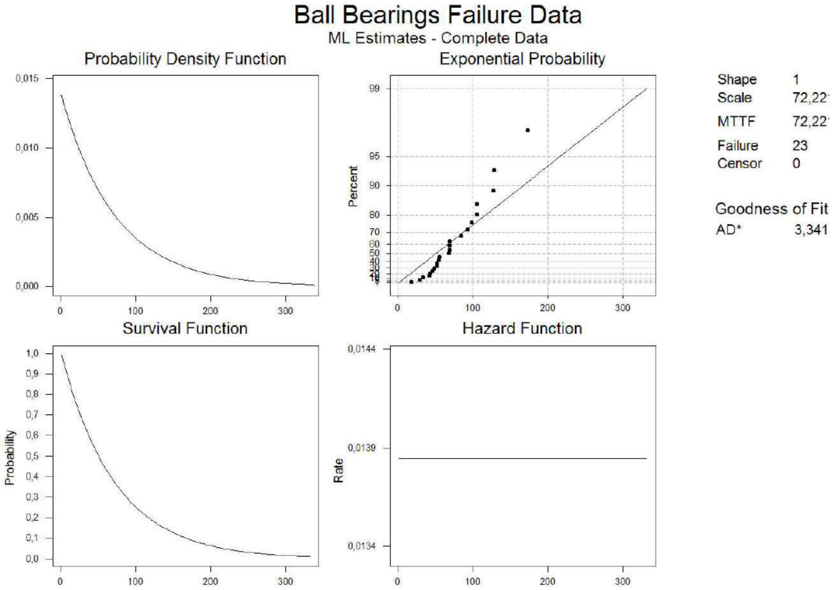
RECALL BALL BEARING FAILURE DATA

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	

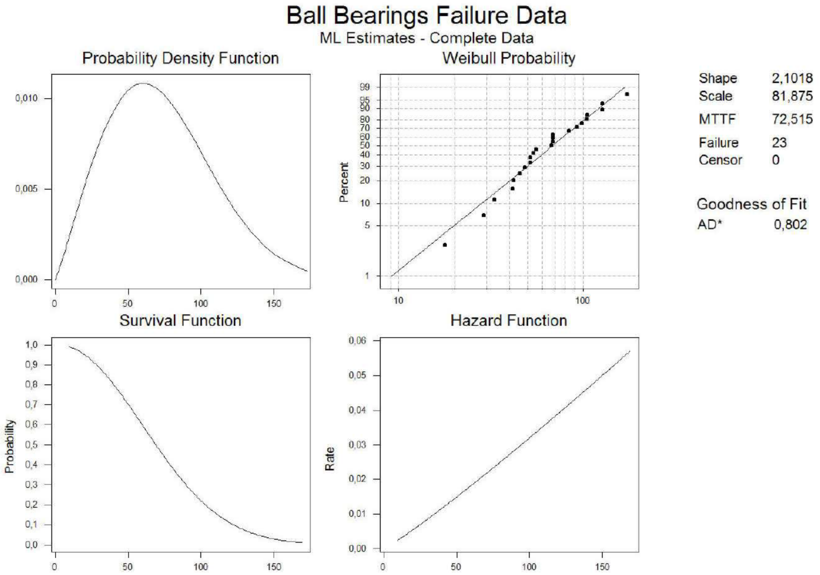
Question: *How can we fit a parametric lifetime model to these data?*



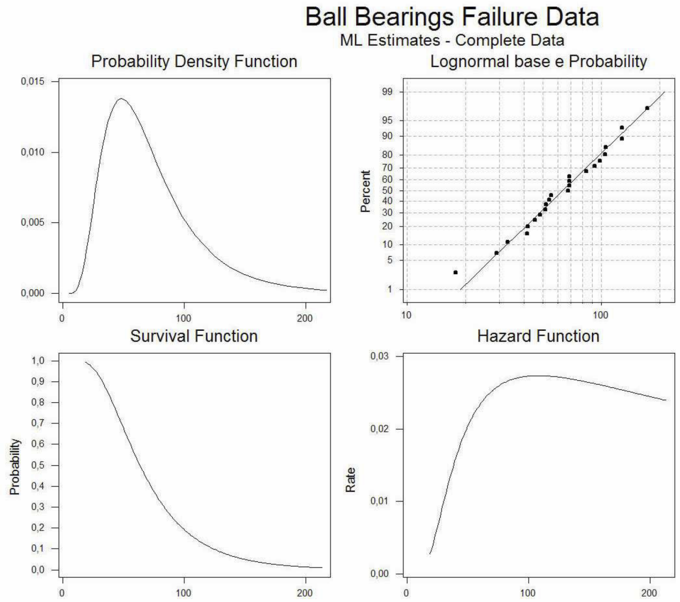
BB-DATA: EXPONENTIAL DISTRIBUTION (MINITAB)



BB-DATA: WEIBULL DISTRIBUTION (MINITAB)



BB-DATA: LOGNORMAL DISTRIBUTION (MINITAB)



Location 4,1504
Scale 0,5217
MTTF 72,709
Failure 23
Censor 0

Goodness of Fit
AD* 0,647

EXAMPLE OF ESTIMATED $S(t)$

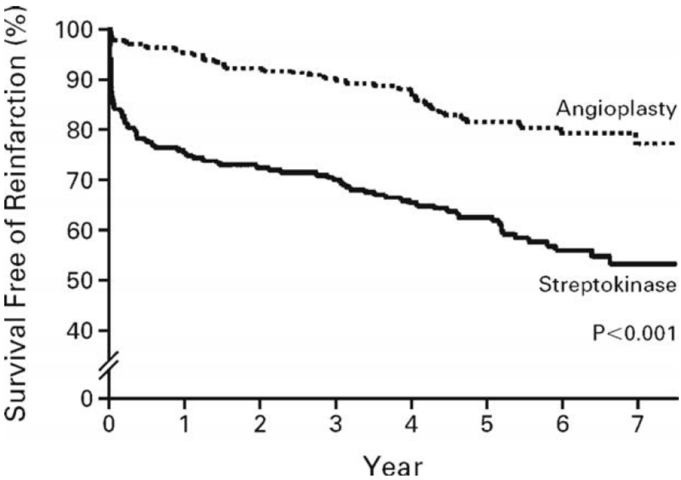


Fig. 1.5 *Survival curves for two treatments of myocardial infarction. Figure reproduced with permission from Zijlstra et al. (1999). Copyright The New England Journal of Medicine.*

EXAMPLE OF ESTIMATED $\alpha(t)$ and $S(t)$

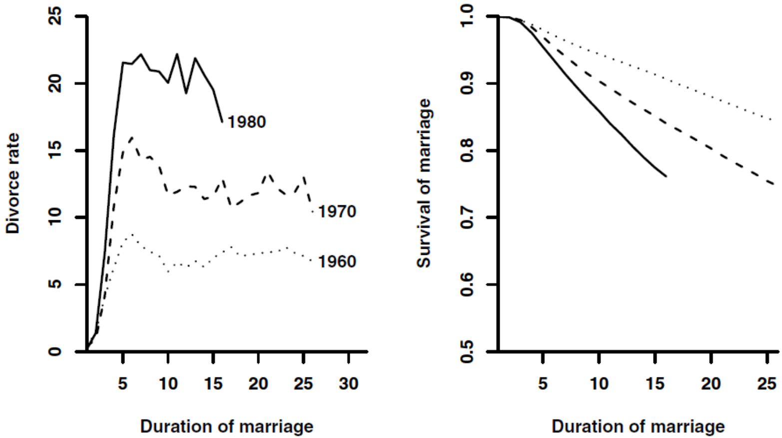


Fig. 1.4 *Rates of divorce per 1000 marriages per year (left panel) and empirical survival curves (right panel) for marriages contracted in 1960, 1970, and 1980. (Based on data from Statistics Norway.)*

Models based on the hazard rate have proven particularly useful:

Cox' regression model.

It is assumed that the *hazard rate* of an individual with *covariates* $x_1, \dots, x_p$ takes the form

$$\alpha(t|x_1, \dots, x_p) = \alpha_0(t) \exp\{\beta_1 x_1 + +\beta_p x_p\}$$

Aalen's additive regression model.

Here it is assumed that

$$\alpha(t|x_1, \dots, x_p) = \beta_0(t) + \beta_1(t)x_1 + \cdots + \beta_p(t)x_p$$

UNOBSERVABLE HETEROGENEITY (ABG 1.5.3 and Ch. 7)

Sometimes one may want to model *unobservable heterogeneity* between individuals.

This may be modeled by assuming that each individual has a *frailty* Z, which varies from individual to individual by some probability distribution (say, with mean value 1).

Then, conditional on the frailty, the hazard rate of an individual is assumed to take the form

$$\alpha(t|Z) = Z \cdot \alpha(t)$$

For derivation of likelihood functions, one will have to unconditon with respect to Z by integration.

Unobservable heterogeneity can also be combined with observed covariates,

$$\alpha(t|x_1, \ldots, x_p, Z) = Z \cdot \alpha_0(t) \exp\{\beta_1 x_1 + + \beta_p x_p\}$$