STK4080/9080
SURVIVAL AND EVENT HISTORY ANALYSIS

Slides 1: Introduction

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

THIS COURSE ...

Course content

The course gives an introduction to the most important concepts and methods in survival and event history analysis. These methods have applications for instance in insurance, medicine and reliability.

Learning outcome

The course gives the background for analyzing a wide specter of models for time to one event as well as models for complex event histories. The course gives a practical introduction to these methods as well as a theoretical understanding of them.

SURVIVAL ANALYSIS (ABG 1.1)

EXAMPLES OF LIFETIMES (SURVIVAL TIMES)

Lifetime

Time to occurrence of some event of interest for individuals in some population.

Medical research:

- Time to death of a patient after start of certain treatment
- Time from entrance to discharge from a hospital
- Times between successive epileptic seizures for patient

Reliability engineering:

- Time to failure of a component or a system
- Number of cycles to failure (fatigue testing)
- Times between successive failures of a machine

WHY COLLECT AND ANALYZE LIFETIME/SURVIVAL/RELIABILITY DATA?

Medical research:

- Compare different treatments with respect to survival or recurrence
- Predict the outcome of an intervention or the life expectancy after the invention
- Identify risk factors for diseases and assess their magnitude

Reliability engineering:

- Assess reliability of a system/component/product
- Compare two or more products with respect to reliability
- Predict product reliability in the design phase
- Predict warranty claims for a product in the market

SPECIAL ASPECTS OF SURVIVAL AND EVENT HISTORY ANALYSIS IN STATISTICS

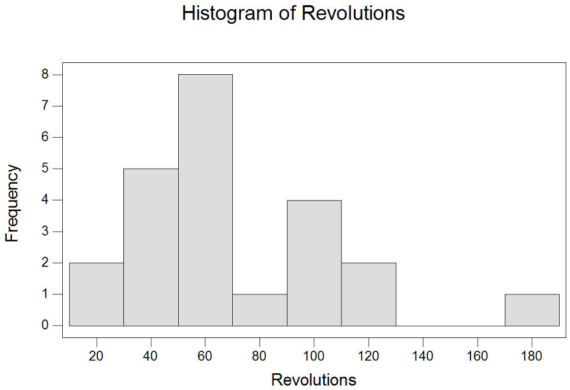
- *Censored data* (how can we use data from individuals or units for which the event of interest has not occurred within the observation period?)
- subjects may not be followed from time 0 (in the study time scale), but only from a later entry time. This is called *delayed entry* or *left-truncation*.
- Definition of *starting time and failure time* may be difficult
- Definition of *time scale* (in reliability: operation time, calendar time or number of cycles?)
- Effect of *covariates* (demographic, medical, environmental)
- What if an individual or unit dies or fails of another cause than the one we would like to study? ("*competing risks*")
- *Recurrent events* – what if the system can fail several times; how to analyze recurring stages of a disease?

COMPLETE SURVIVAL DATA: *BALL BEARING FAILURE DATA*

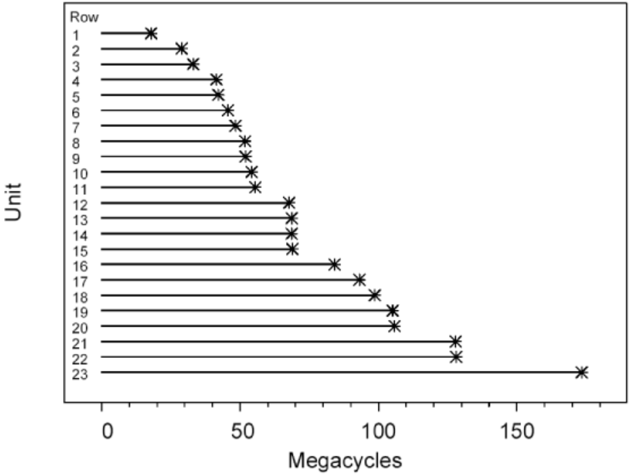
Data: Millions of revolutions to fatigue failure for 23 units

Question: How can we estimate the probability distribution of these 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	



BALL BEARING FAILURE DATA (EVENT PLOT)



HEAVILY CENSORED DATA: IC DATA (MEEKER, 1987)

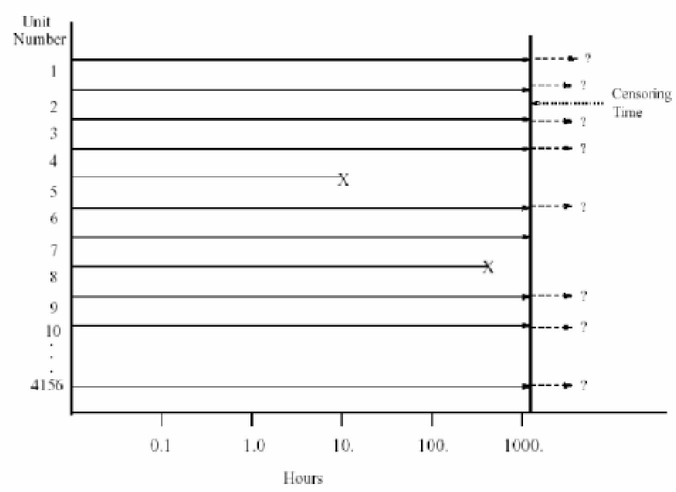
- Integrated circuit failure times in hours
  - n = 4156 ICs tested for 1,370 hours at 80° C and 80% relative humidity
  - There were 28 failures
  - When the test ended at 1,370 hours, 4128 units were still running

.10	.10	.15	.60	.80	.80
1.20	2.5	3.0	4.0	4.0	6.0
10.0	10.0	12.5	20.	20.	43.
43.	48.	48.	54.	74.	84.
94.	168.	263.	593.		

Questions of interest:

- How to estimate the distribution of the failure time when there are censored observations?
- Probability of failure before 100 hours?
- Failure rate by 100 hours?
- Proportion failed after 10^5 hours?

IC DATA (EVENT PLOT)



SURVIVAL OF MULTIPLE MYELOMA PATIENTS

- *Multiple myeloma* is a malignant disease characterised by the accumulation of abnormal plasma cells, a type of white blood cell, in the bone marrow.
- Data (next slide) from Medical Center of the University of West Virginia, USA.
- **Aim:** To examine the association between certain explanatory variables or covariates and the survival time of patients in months from diagnosis until death from multiple myeloma).

MULTIPLE MYELOMA DATA

Table 1.3 *Survival times of patients in a study on multiple myeloma.*

Patient number	Survival time	Status	Age	Sex	*Bun*	*Ca*	*Hb*	*Pcells*	*Protein*
1	13	1	66	1	25	10	14.6	18	1
2	52	0	66	1	13	11	12.0	100	0
3	6	1	53	2	15	13	11.4	33	1
4	40	1	69	1	10	10	10.2	30	1
5	10	1	65	1	20	10	13.2	66	0
6	7	0	57	2	12	8	9.9	45	0
7	66	1	52	1	21	10	12.8	11	1
8	10	0	60	1	41	9	14.0	70	1
9	10	1	70	1	37	12	7.5	47	0
10	14	1	70	1	40	11	10.6	27	0
11	16	1	68	1	39	10	11.2	41	0
12	4	1	50	2	172	9	10.1	46	1
13	65	1	59	1	28	9	6.6	66	0
14	5	1	60	1	13	10	9.7	25	0
15	11	0	66	2	25	9	8.8	23	0
16	10	1	51	2	12	9	9.6	80	0
17	15	0	55	1	14	9	13.0	8	0
18	5	1	67	2	26	8	10.4	49	0
19	76	0	60	1	12	12	14.0	9	0
20	56	0	66	1	18	11	12.5	90	0

EXAM EXERCISE NTNU 2020

Problem 1 *Hospital length of stay of coronavirus (COVID-19) patients*

A group of Chinese medical researchers analyzed data for length of hospital stay for confirmed COVID-19 patients at hospitals in the Sichuan Province. The aim of the study was to obtain knowledge about the new virus that would be important for planning and allocation of medical resources in the COVID-19 pandemic.

The study included 538 patients who were admitted in hospitals after January 16, 2020. 351 out of these (65%) recovered and were discharged before the end of the study, April 4. Only 3 patients died in hospital before April 4.

The data used in this exercise are *simulated* based on the reported results from the study (the full data set was not published in the report).

The data consist of the observed time, `Time` (in days); censoring status C (= 0 or 1); and six binary covariates $\mathtt{x}_1, \ldots, \mathtt{x}_6$, with values 0 and 1 as defined by Table 1.

i	x_i	0	1
1	age (years)	< 45	≥ 45
2	gender	male	female
3	time from onset	< 5	≥ 5
4	hospital grade	non-provincial	provincial
5	density of health workers	< 5.5	≥ 5.5
6	clinical grade	mild	severe

Table 1: The binary covariates

EVENT HISTORY ANALYSIS (ABG 1.2)

- In survival analysis one considers the **time to the occurrence** of a single event for each individual
- Sometimes the event in question may occur *more than once* for an individual: Consider models for **recurrent events**
- In other situations, more than one type of event are of interest: Consider **multistate models**.

MULTISTATE MODELS

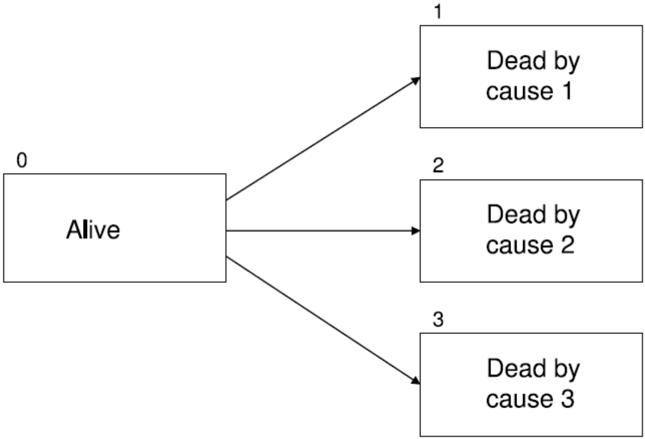


Fig. 1.8 *A model for $k = 3$ competing causes of death.*

MULTISTATE MODELS

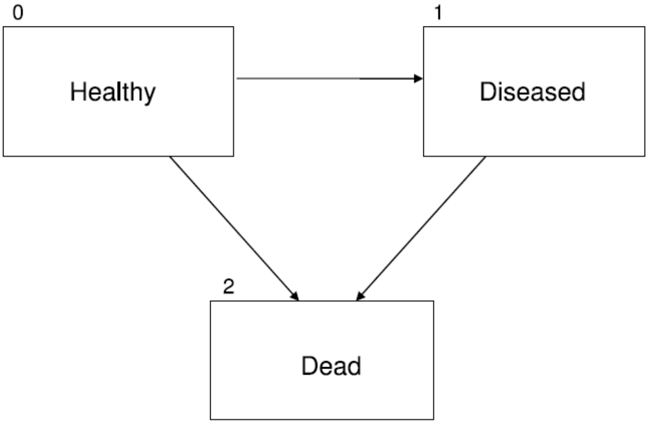


Fig. 1.10 *An illness-death model without recovery.*

RECURRENT EVENTS

Aalen and Husebye (1991): Migratory motor complex (MMC) periods in 19 patients, 1-9 events per individual.

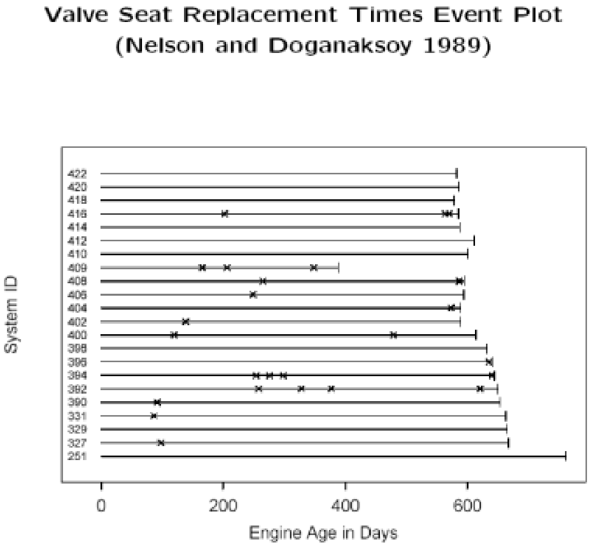
Individual	Observed periods (minutes)					
1	112	145	39	52	21	34
	33	51	(54)			
2	206	147	(30)			
3	284	59	186	(4)		
4	94	98	84	(87)		
5	67	(131)				
6	124	34	87	75	43	38
	58	142	75	(23)		
7	116	71	83	68	125	(111)
8	111	59	47	95	(110)	
9	98	161	154	55	(44)	
10	166	56	(122)			
11	63	90	63	103	51	(85)
12	47	86	68	144	(72)	
⋮			⋮			

AALEN AND HUSEBY: Movements of the small bowel

See Example 7.1 in ABG:

The task is to estimate the average duration of the intervals and the variation within and between individuals

RECURRENT EVENTS/REPAIRABLE SYSTEMS



VALVE SEAT REPLACEMENT DATA

Data on previous slide are collected from valve seats from a fleet of 41 diesel engines. Each engine has 16 valves. (Time unit is days of operation).

Questions of interest:

- Does the replacement rate increase with age?
- How many replacement valves will be needed in the future?
- Can valve life in these systems be modeled as a renewal process?

ESTIMATED NUMBER OF VALVE SEAT REPLACEMENTS

- Middle curve is cumulative estimated number of replacements for one engine, as a function of age.
- Lower and upper curves are 95% confidence limits.