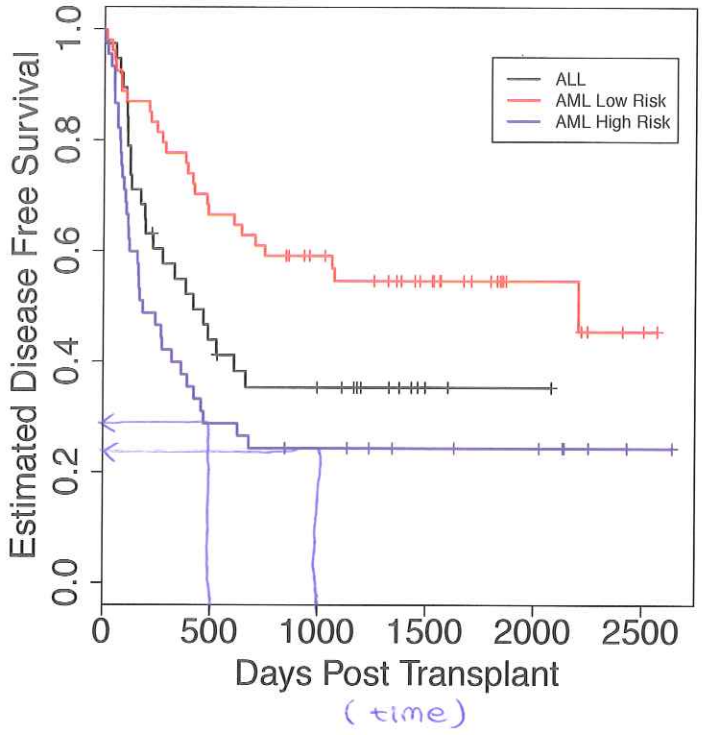
# AMS 276
# Lecture 1: Introduction, Examples and Basic Quantities

Fall 2016

* (KM Section 1.3): Have you ever seen a graph looking like this?




• Prob. of an individual surviving beyond time t

⇔ prob. of experiencing the event after time t.

(time)

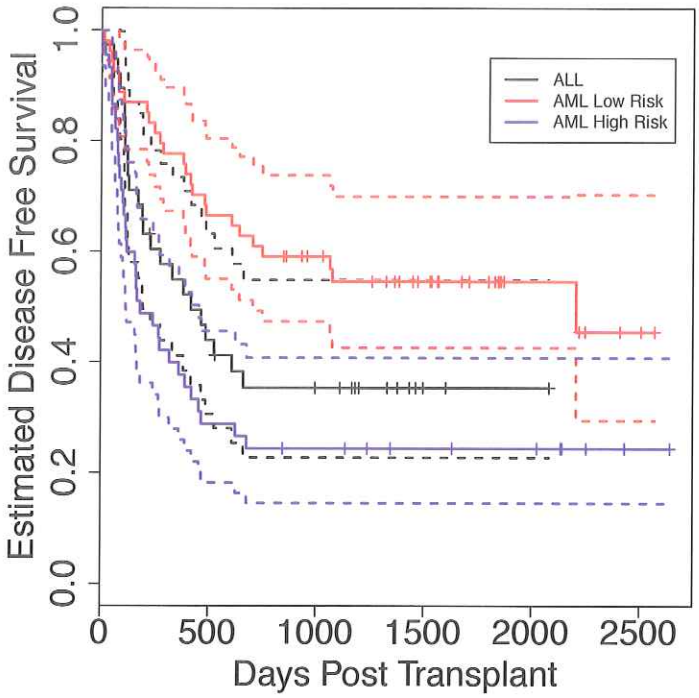
For Blue

• $P(t > 500) > P(t > 1000)$

survival time

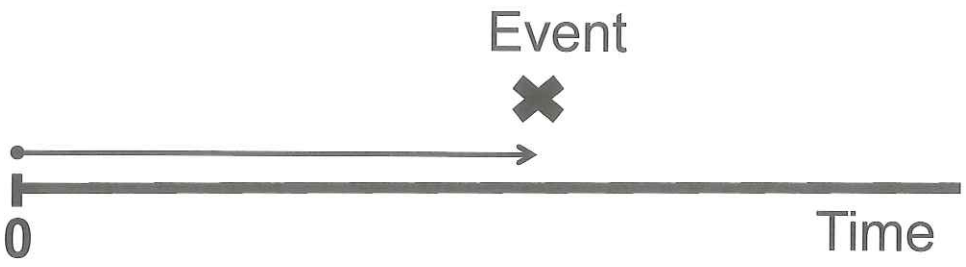
$\{(t > 500)\} \supset \{t \geq 1000\}$

* (KM Section 1.3): Have you ever seen a graph looking like this?
– further express uncertainty!



## † What is Survival Analysis?

1. Outcome variable: time until an event occurs (time to event, failure time, survival time...).




- Examples of failure time
  - ⋆⋆ time from cancer incidence to death
  - ⋆⋆ time from HIV infection to AIDS onset
  - ⋆⋆ time from committee approval to policy implementation
  - ⋆⋆ time from kidney transplantation to organ rejection

- Required that time of origin and end-point be clearly defined

2. Time origin: pre-defined, comparable across subjects

   e.g.: date of randomization in clinical trial, date of enrollment in observational study, date of birth

   Typically not the same calendar date

3. Event:
   - e.g.: death, disease incidence, relapse
   - Usually assume at most one failure per individual.
   - If more than one failure per individual, we need advanced methods.
   - If more than one failure type, we need competing risks methods.

KM chapter 2.7

Each subject may fail due to one of K causes

or

cause - specific mortality

(i) death from heart disease

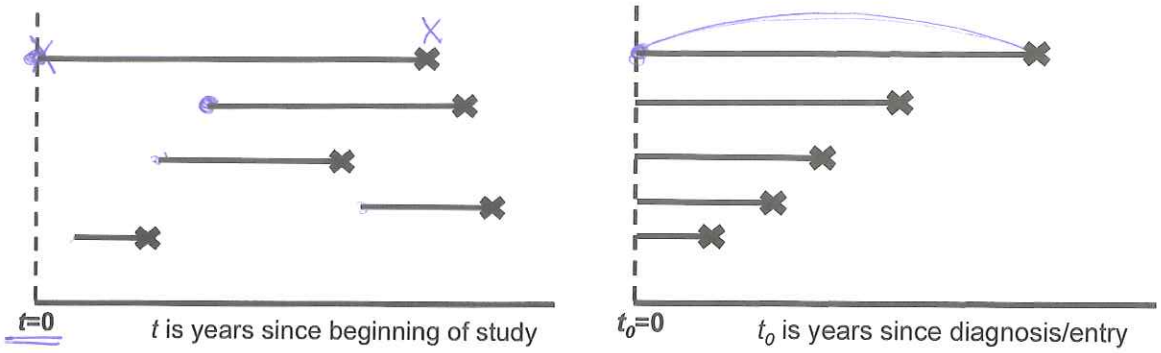
(ii) death from cancer

(iii) death from other causes

Problem: Occurrence of one of the events precludes us from observing the other event(s) on the same patient.

4. Time: measured in years, months, days, etc

- "failure time" or "survival time": positive and continuous

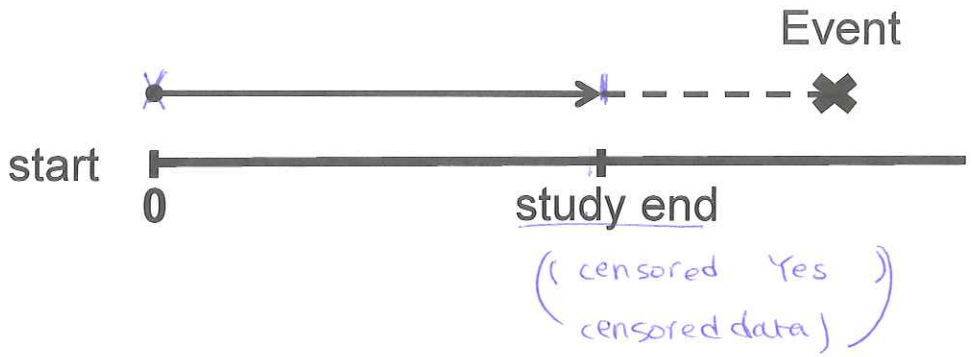



- "failure time" can be discrete, for example, death time is grouped into small intervals (e.g.: days, weeks...)

5. Survival data usually contain censored observations (Incomplete observation of failure time).

   e.g.: right censoring, left censoring, interval censoring. ($\Rightarrow$ will be discussed in detail later)

   - Leukemia patients in remission



- More on **censoring**

  - ⋆⋆ Since any study is of finite duration, event of interest may not be observed for all subjects.
  - ⋆⋆ Generally, standard techniques cannot be applied to survival data since failure times are usually not observed for some individuals under study.
  - ⋆⋆ Subjects whose failure times are unobserved are said to be censored.
  - ⋆⋆ Estimation methods of survival analysis are built around extracting information from all subjects, censored or not.

• Sources of Censoring – subjects may be censored for different reasons:

⋆⋆ Study ends before the event has occurred – known as administrative censoring

often independent of failure process

⋆⋆ Loss to follow-up

↻ subject can no longer be traced; no longer under observation
↻ e.g., cohort study: individual leaves the country and, therefore, can no longer be followed through the national mortality database.

↻ some concern about whether such censoring is related indirectly to the failure time.

- Sources of Censoring– subjects may be censored for different reasons: (contd)

  - ⋆⋆ withdrawal from study

    - ⟳ e.g., patient drops out of clinical trial because he is too sick to participate

    - ⟳ e.g., subject discontinues participation in trial because her symptoms have subsided

  - ⋆⋆ often, dependent censoring ("informative drop-out") will be a concern in such cases.

- Why do we care?

  - ⋆⋆ Most distinct characteristic of survival data: **censoring**
  - ⋆⋆ Suppose we do not treat censoring properly and try several simpler alternatives for estimation in the presence of censoring; e.g.,
    - count censored observations as survival times
      ⇒ under-estimate survival probability
    - delete censored observations
      ⇒ under-estimate survival probability
    - model event indicator as a binary variable
      ⇒ throw away time information

10

$\frac{1+1+1+1+4+5}{6} = 2.17$

$\frac{1+1+1+5}{4}$

$\frac{4}{6} =$

Ex data 1 1 1 1+ 4+ 5 (1+: right censored)

6. Applied fields: medicine, biology, public health, epidemiology, engineering, economics, and demography.

- Our interest is...
  - ⋆⋆ distribution of failure time
  - ⋆⋆ comparison of the failure times in different groups
  - ⋆⋆ effect of explanatory variables

ICS P3

† Example 1.1 Multiple myeloma data

- Multiple myeloma: hematologic cancer characterized by an overproduction of antibodies
- A study undertaken by Eastern Cooperative Oncology Group (ECOG) – E2479
- Patients received a chemotherapy involving alkylating agents
- $n = 479$ patients
- Several covariates measured at diagnosis: blood urea nitrogen, hemoglobin, platelet count, age, white blood cell count, bone fractures, percentage of the plasma cells in bone marrow, proteinuria and serum calcium
- Q: Variable subset selection–which covariates are important predictors of survival?
- Ibrahim and Chen (1998) and Ibrahim et al (1999)

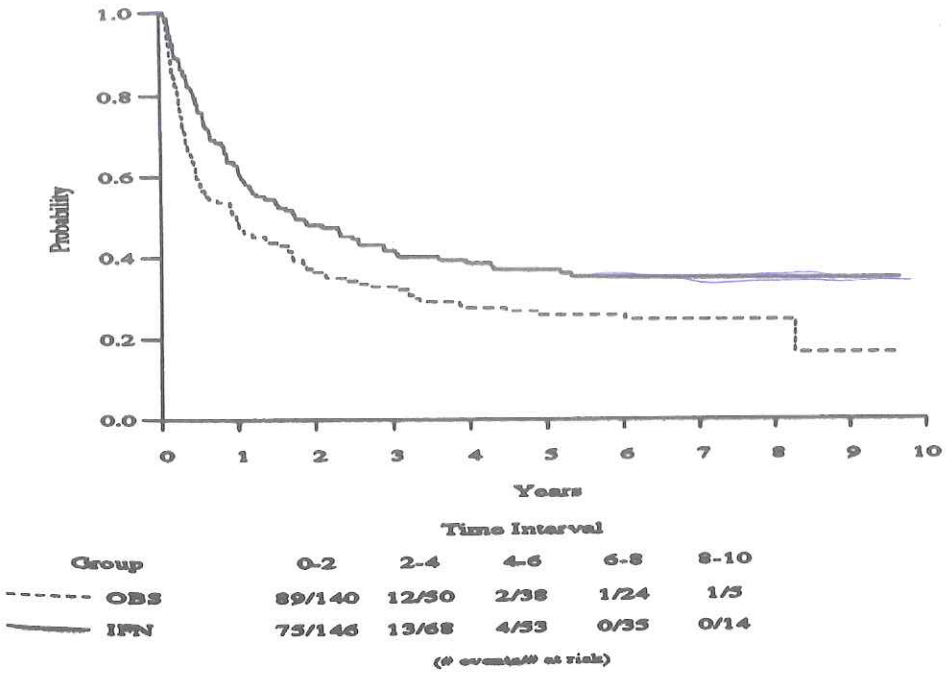



FIGURE 1.1. Kaplan-Meier RFS plots for E1684.

⇒ US FDA approved this regimen as an adjuvant therapy for high-risk melanoma patients.

† Example 1.2 Melanoma data

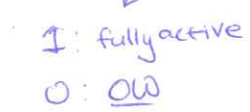
- ECOG phase III clinical trial– E1684: two-arm clinical trial comparing high-dose interferon (IFN) to observations (OBS).
- A total of $n_0 = 286$ patients enrolled in the study, accrued from 1984 to 1990
- Q1: Significant impact of covariates on relapse-free survival (RFS) and survival (OS)
- Q2: Proportion of patients cured after sufficient follow-up

- E1684 Data

study (E1684 and E1690)
failcens (2=relapsed, 1=censored)
survcens (2=dead, 1=alive)
age is in years
trt (1=observation, 2=IFN)
failtime and survtime are in years
nodes1 = nodal category. There are 4 nodal categories. This is a categorical variable.
sex (1=male, 2=female)
breslow is continuous, measured in mm.
perform is categorical.

Breslow and nodes1 have missing values.

| STUDY | TRT | FAILTIME | FAILCENS | SURVTIME | SURVCENS | AGE | NODES1 | SEX | PERFORM | BRESLOW |
|---|---|---|---|---|---|---|---|---|---|---|
| 1684 | 2 | 1.15068 | 2 | 1.57808 | 2 | 35.9945 | 4 | 1 | 0 | 1.27 |
| 1684 | 2 | 0.62466 | 2 | 1.48219 | 2 | 41.9014 | . | 1 | 0 | 0.76 |
| 1684 | 1 | 1.89863 | 1 | 7.33425 | 1 | 70.2164 | 2 | 2 | 0 | 35.00 |
| 1684 | 1 | 0.45479 | 2 | 0.65479 | 2 | 58.1753 | 2 | 2 | 1 | . |
| 1684 | 1 | 2.09041 | 2 | 2.23288 | 2 | 33.7096 | 2 | 1 | 0 | 1.70 |
| 1684 | 2 | 9.38356 | 1 | 9.38356 | 1 | 47.9726 | 2 | 1 | 0 | 1.00 |
| 1684 | 1 | 1.69863 | 2 | 3.27671 | 2 | 31.8219 | 2 | 2 | 0 | 11.00 |
| 1684 | 2 | 0.00000 | 1 | 0.00000 | 1 | 72.3644 | 1 | 2 | 0 | . |
| 1684 | 1 | 0.25753 | 2 | 0.80274 | 2 | 40.7151 | 4 | 2 | 0 | . |
| 1684 | 1 | 9.64384 | 1 | 9.64384 | 1 | 32.9479 | 4 | 1 | 0 | 6.50 |
| 1684 | 2 | 1.61370 | 2 | 1.66575 | 2 | 35.9205 | 4 | 1 | 0 | 3.62 |

† Example 1.2 Melanoma data (contd)

- The treatment effect favoring IFN seen in E1684 was larger than expected and was accompanied by substantial side effects due to the high-dose regimen. $\Rightarrow$ Begin a second trial.
- A second trial was conducted to confirm the results of E1684 and to study the benefit of IFN given at a lower dose.
- ECOG phase III clinical trial– E1690: three-arm clinical trial comparing high-dose interferon (IFN), low-dose IFN and observations (OBS).
- $n = 427$ patients on the high-dose interferon arm and observation arm combined.
- E1690 was initiated right after the completion of E1684 and accrued patients from 1991 until 1995.

† Example 1.2 Melanoma data (contd)

- Designed for exactly the same patient population as E1684 and the high-dose interferon arm in E1690 was identical to that of E1684.
- Q1: Informative prior elicitation in cure rate models using historical data
- Q2: The assessment of the goodness of fit to compare two competing models

† Example 1.2 Melanoma data (contd)

- *response variable*: relapse-free survival (RFS) – possibly right censored
- *covariates*: treatment ($x_1$, IFN or OBS), age ($x_2$), sex ($x_3$), logarithm of Breslow depth ($x_4$), logarithm of size of primary ($x_5$) and type of primary ($x_6$)
- $x_1, x_2, x_3$ are completely observed and $x_4, x_5, x_6$ have missing values (27.4% and 28.6% for E1684 and E1690, respectively)
- Q3: Missing data method in survival analysis (missing covariate data)

† Example 1.3 Breast cancer data

- Finkelstein and Wolfe (1985) : A semi-parametric Model for Regression Analysis of Interval-Censored Failure Time Data
- 46 early breast cancer patients receiving only radiotherapy (covariate value $= 0$) and 48 patients receiving radio-chemotherapy ($x = 1$)
- Monitored for cosmetic change through weekly clinic visits so interval censored data
- Patients missed some of their weekly visits

  e.g.: [7,18] $\Rightarrow$ at the 7th week clinic visit patient had shown no change and then in the next clinic visit at the 18th week the patients tissue showed that the change had already occurred

- Q: The effect of the covariate $x$ on the survival time

## What is coming?

† Let $T$ be the time until some specified event.

† We will define

- ⋆⋆ **Probability density function**: probability of the event's occurring at time $t$
- ⋆⋆ **Survival function**: probability of an individual surviving beyond time $t$
- ⋆⋆ **Hazard rate (or risk function)**: Instantaneous rate of failure at time $t$
- ⋆⋆ **Mean residual life at time** $t$: mean time to the event of interest, given the event has not occurred at $t$

† If we know any one of these four functions, then the other three can be uniquely determined.

## Probability density function

† Let $T$ be the time until some specified event.

⋆⋆ $T \in [0, \infty)$ a nonnegative continuous random variable from a homogeneous population

† Let $f(t)$ (defined over the interval $[0, \infty)$) denote the probability density function of $T$ and let the distribution function be

$$F(t) = \Pr(T < t) = \int_0^t f(u)du.$$

† For simplicity, we assume $F(t)$ is continuous and differentiable.

$f(t) = \frac{\alpha}{\eta}\left(\frac{t}{\eta}\right)^{\alpha-1}\exp\left(-\left(\frac{t}{\eta}\right)^{\alpha}\right)$ : wiki

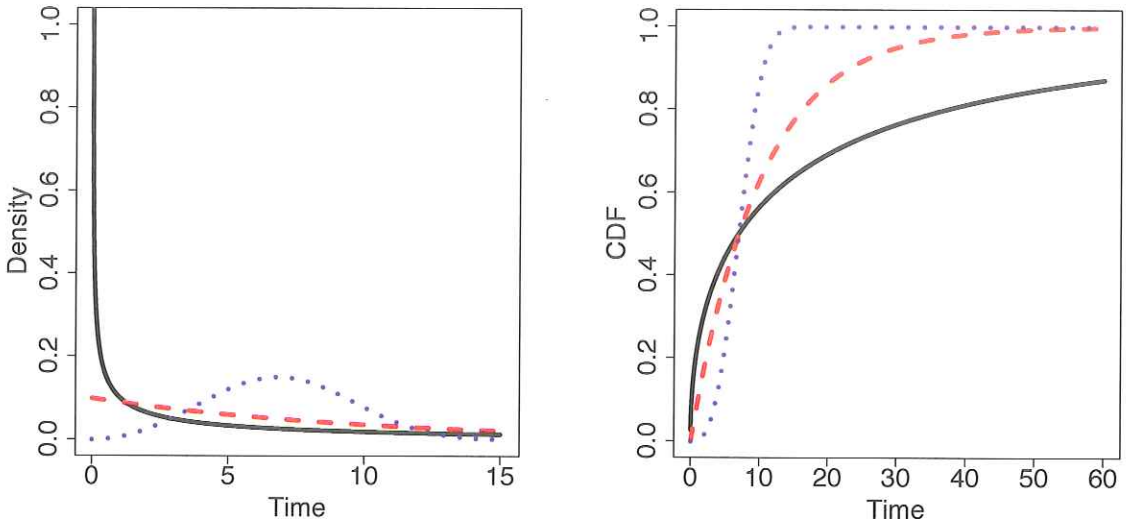
$\eta = \lambda^{-\frac{1}{\alpha}}$

♢ eg.: $T \sim \text{Weibull}(\alpha, \lambda)$ ($\lambda > 0$ and $\alpha > 0$). That is,

$$f(t) = \begin{cases} \alpha\lambda t^{\alpha-1}\exp(-\lambda t^{\alpha}), & \text{for } t > 0, \\ 0 & \text{otherwise .} \end{cases}$$

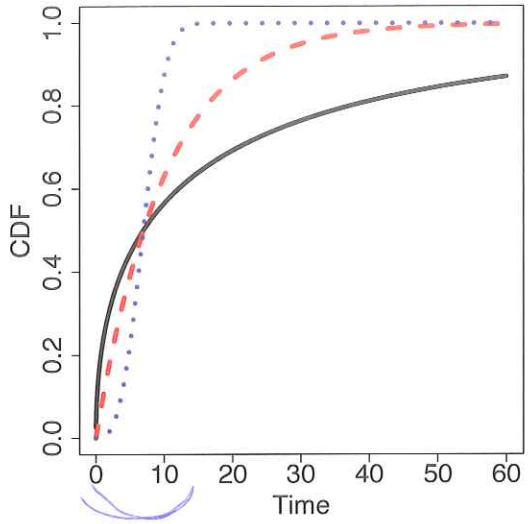
⋆⋆ $(\lambda, \alpha) = (0.26328, 0.5)$ for black solid, $(0.1, \underset{\downarrow}{\overset{\alpha}{1}})$ for red dashed and (0.00208, 3) for blue dotted.

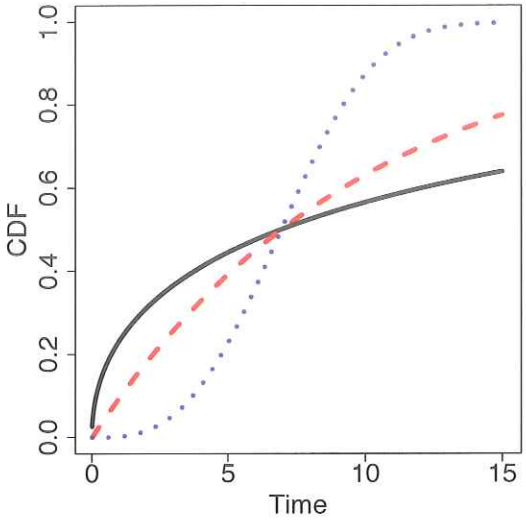
exponential



♢ eg.: $T \sim \text{Weibull}(\alpha, \lambda)$ ($\lambda > 0$ and $\alpha > 0$).

⋆⋆ $(\lambda, \alpha) = (0.26328, 0.5)$ for black solid, (0.1, 1) for red dashed and (0.00208, 3) for blue dotted.

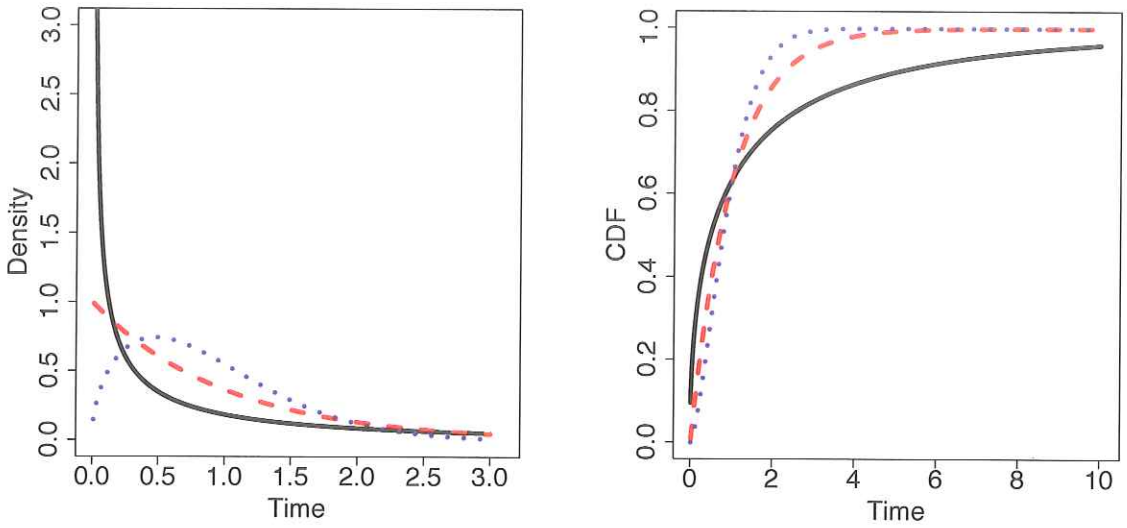


♢ eg.: $T \sim \text{Weibull}(\alpha, \lambda)$ ($\lambda > 0$ and $\alpha > 0$). That is,

$$f(t) = \begin{cases} \alpha\lambda t^{\alpha-1} \exp(-\lambda t^{\alpha}), & \text{for } t > 0, \\ 0 & \text{otherwise .} \end{cases}$$

⋆⋆ $(\lambda, \alpha) = (1, 0.5)$ for black solid, $(1, 1)$ for red dashed and $(1, 1.5)$ for blue dotted.



♢ eg.: $T \sim \text{Weibull}(\alpha, \lambda)$ ($\lambda > 0$ and $\alpha > 0$).

⋆⋆ $(\lambda, \alpha) = (1, 0.5)$ for black solid, (1, 1) for red dashed and (1, 1.5) for blue dotted.