Survival Analysis: Introduction

Survival Analysis typically focuses on **time to event** data. In the most general sense, it consists of techniques for positive-valued random variables, such as

- time to death
- time to onset (or relapse) of a disease
- length of stay in a hospital
- duration of a strike
- money paid by health insurance
- viral load measurements
- time to finishing a doctoral dissertation!

Kinds of survival studies include:

- clinical trials
- prospective cohort studies
- retrospective cohort studies
- retrospective correlative studies

Typically, survival data are not fully observed, but rather are *censored*.

In this course, we will:

- **describe survival data**
- **compare survival of several groups**
- **explain survival with covariates**
- **design studies with survival endpoints**

Some knowledge of discrete data methods will be useful, since analysis of the "time to event" uses information from the discrete (i.e., binary) outcome of whether the event occurred or not.

Some useful references:

- Collett: *Modelling Survival Data in Medical Research*
- Cox and Oakes: *Analysis of Survival Data*
- Kalbfleisch and Prentice: *The Statistical Analysis of Failure Time Data*
- Lee: *Statistical Methods for Survival Data Analysis*
- Fleming & Harrington: *Counting Processes and Survival Analysis*
- Hosmer & Lemeshow: *Applied Survival Analysis*
- Kleinbaum: *Survival Analysis: A self-learning text*

- Klein & Moeschberger: *Survival Analysis: Techniques for censored and truncated data*
- Cantor: *Extending SAS Survival Analysis Techniques for Medical Research*
- Allison: *Survival Analysis Using the SAS System*
- Jennison & Turnbull: *Group Sequential Methods with Applications to Clinical Trials*

Some Definitions and notation

Failure time random variables are always **non-negative**. That is, if we denote the failure time by T, then $T \geq 0$.

T can either be **discrete** (taking a finite set of values, e.g. $a_1, a_2, \ldots, a_n$) or **continuous** (defined on $(0, \infty)$).

A random variable X is called a **censored failure time random variable** if $X = \min(T, U)$, where U is a non-negative censoring variable.

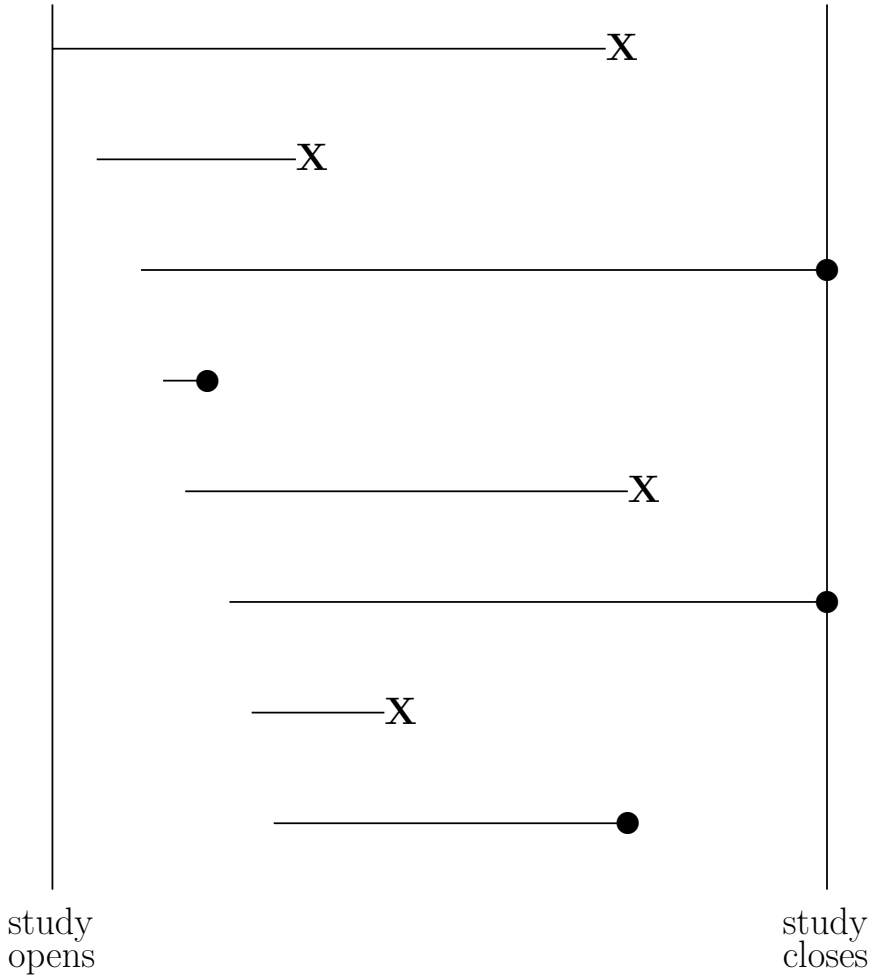
In order to define a failure time random variable, we need:

(1) an unambiguous **time origin**
(e.g. randomization to clinical trial, purchase of car)

(2) a **time scale**
(e.g. real time (days, years), mileage of a car)

(3) definition of the **event**
(e.g. death, need a new car transmission)

Illustration of survival data



● = censored observation
X = event

The illustration of survival data on the previous page shows several features which are typically encountered in analysis of survival data:

- individuals do not all enter the study at the same time
- when the study ends, some individuals still haven't had the event yet
- other individuals drop out or get lost in the middle of the study, and all we know about them is the last time they were still "free" of the event

The first feature is referred to as **"staggered entry"**

The last two features relate to **"censoring"** of the failure time events.

Types of censoring:

- **Right-censoring**:

  only the r.v. $X_i = \min(T_i, U_i)$ is observed due to

  – loss to follow-up
  – drop-out
  – study termination

  We call this right-censoring because the true unobserved event is to the right of our censoring time; i.e., all we know is that the event has not happened at the end of follow-up.

  In addition to observing X_i, we also get to see the **failure indicator**:

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq U_i \\ 0 & \text{if } T_i > U_i \end{cases}$$

  Some software packages instead assume we have a **censoring indicator**:

$$c_i = \begin{cases} 0 & \text{if } T_i \leq U_i \\ 1 & \text{if } T_i > U_i \end{cases}$$

  Right-censoring is the most common type of censoring assumption we will deal with in survival analysis.

- **Left-censoring**

  Can only observe $Y_i = \max(T_i, U_i)$ and the failure indicators:

$$\delta_i = \begin{cases} 1 & \text{if } U_i \leq T_i \\ 0 & \text{if } U_i > T_i \end{cases}$$

  e.g. (Miller) study of age at which African children learn a task. Some already knew (left-censored), some learned during study (exact), some had not yet learned by end of study (right-censored).

- **Interval-censoring**

  Observe (L_i, R_i) where $T_i \in (L_i, R_i)$

  Ex. 1: Time to prostate cancer, observe longitudinal PSA measurements

  Ex. 2: Time to undetectable viral load in AIDS studies, based on measurements of viral load taken at each clinic visit

  Ex. 3: Detect recurrence of colon cancer after surgery. Follow patients every 3 months after resection of primary tumor.

Independent vs informative censoring

- We say censoring is **independent** (non-informative) if U_i is independent of T_i.
  - **Ex. 1** If U_i is the planned end of the study (say, 2 years after the study opens), then it is usually independent of the event times.
  - **Ex. 2** If U_i is the time that a patient drops out of the study because he/she got much sicker and/or had to discontinue taking the study treatment, then U_i and T_i are probably not independent.

  An individual censored at U should be representative of all subjects who survive to U.

  This means that censoring at U *could* depend on prognostic characteristics measured at baseline, but that among all those with the same baseline characteristics, the probability of censoring prior to or at time U should be the same.

- Censoring is considered **informative** if the distribution of U_i contains any information about the parameters characterizing the distribution of T_i.

Suppose we have a sample of observations on n people:

$$(T_1, U_1), (T_2, U_2), ..., (T_n, U_n)$$

There are three main types of (right) censoring times:

- **Type I:** All the U_i's are the same
  e.g. animal studies, all animals sacrificed after 2 years

- **Type II:** $U_i = T_{(r)}$, the time of the rth failure.
  e.g. animal studies, stop when 4/6 have tumors

- **Type III:** the U_i's are random variables, δ_i's are failure indicators:

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq U_i \\ 0 & \text{if } T_i > U_i \end{cases}$$

Type I and **Type II** are called *singly* censored data,
Type III is called *randomly* censored (or sometimes *progressively* censored).

Some example datasets:

Example A. Duration of nursing home stay

(Morris et al., *Case Studies in Biometry*, Ch 12)

The National Center for Health Services Research studied 36 for-profit nursing homes to assess the effects of different financial incentives on length of stay. "Treated" nursing homes received higher per diems for Medicaid patients, and bonuses for improving a patient's health and sending them home.

Study included 1601 patients admitted between May 1, 1981 and April 30, 1982.

Variables include:
LOS - Length of stay of a resident (in days)
AGE - Age of a resident
RX - Nursing home assignment (1:bonuses, 0:no bonuses)
GENDER - Gender (1:male, 0:female)
MARRIED - (1: married, 0:not married)
HEALTH - health status (2:second best, 5:worst)
CENSOR - Censoring indicator (1:censored, 0:discharged)

First few lines of data:
37 86 1 0 0 2 0
61 77 1 0 0 4 0

Example B. Fecundability

Women who had recently given birth were asked to recall how long it took them to become pregnant, and whether or not they smoked during that time. The outcome of interest (summarized below) is time to pregnancy (measured in menstrual cycles).

19 subjects were not able to get pregnant after 12 months.

Cycle	Smokers	Non-smokers
1	29	198
2	16	107
3	17	55
4	4	38
5	3	18
6	9	22
7	4	7
8	5	9
9	1	5
10	1	3
11	1	6
12	3	6
12+	7	12

Example C: MAC Prevention Clinical Trial

ACTG 196 was a randomized clinical trial to study the effects of combination regimens on prevention of MAC (*mycobacterium avium complex*), one of the most common opportunistic infections in AIDS patients.

The **treatment regimens** were:

- clarithromycin (new)
- rifabutin (standard)
- clarithromycin plus rifabutin

Other characteristics of trial:

- Patients enrolled between April 1993 and February 1994
- Follow-up ended August 1995
- In February 1994, rifabutin dosage was reduced from 3 pills/day (450mg) to 2 pills/day (300mg) due to concern over **uveitis**[1]

The main intent-to-treat analysis compared the 3 treatment arms without adjusting for this change in dosage.

[1] *Uveitis* is an adverse experience resulting in inflammation of the uveal tract in the eyes (about 3-4% of patients reported uveitis).

Example D: HMO Study of HIV-related Survival

This is hypothetical data used by Hosmer & Lemeshow (described on pages 2-17) containing 100 observations on HIV+ subjects belonging to an Health Maintenance Organization (HMO). The HMO wants to evaluate the survival time of these subjects. In this hypothetical dataset, subjects were enrolled from January 1, 1989 until December 31, 1991. Study follow up then ended on December 31, 1995.

Variables:

ID	Subject ID (1-100)
TIME	Survival time in months
ENTDATE	Entry date
ENDDATE	Date follow-up ended due to death or censoring
CENSOR	Death Indicator (1=death, 0=censor)
AGE	Age of subject in years
DRUG	History of IV Drug Use (0=no,1=yes)

This dataset is used by Hosmer & Lemeshow to motivate some concepts in survival analysis in Chap. 1 of their book.

Example E: UMARU Impact Study (UIS)

This dataset comes from the University of Massachusetts AIDS Research Unit (UMARU) IMPACT Study, a 5-year collaborative research project comprised of two concurrent randomized trials of residential treatment for drug abuse.

(1) **Program A:** Randomized 444 subjects to a 3- or 6-month program of health education and relapse prevention. Clients were taught to recognize "high-risk" situations that are triggers to relapse, and taught skills to cope with these situations without using drugs.

(2) **Program B:** Randomized 184 participants to a 6- or 12-month program with highly structured life-style in a communal living setting.

Variables:

ID	Subject ID (1-628)
AGE	Age in years
BECKTOTA	Beck Depression Score
HERCOC	Heroin or Cocaine Use prior to entry
IVHX	IV Drug use at Admission
NDRUGTX	Number previous drug treatments
RACE	Subject's Race (0=White, 1=Other)
TREAT	Treatment Assignment (0=short, 1=long)
SITE	Treatment Program (0=A,1=B)
LOT	Length of Treatment (days)
TIME	Time to Return to Drug Use (days)
CENSOR	Indicator of Drug Use Relapse (1=yes,0=censored)

Example F: Atlantic Halibut Survival Times

One conservation measure suggested for trawl fishing is a minimum size limit for halibut (32 inches). However, this size limit would only be effective if captured fish below the limit survived until the time of their release. An experiment was conducted to evaluate the survival rates of halibut caught by trawls or longlines, and to assess other factors which might contribute to survival (duration of trawling, maximum depth fished, size of fish, and handling time).

An article by Smith, Waiwood and Neilson, *Survival Analysis for Size Regulation of Atlantic Halibut* in *Case Studies in Biometry* compares parametric survival models to semi-parametric survival models in evaluating this data.

Obs #	Survival *Time* (min)	*Censor*ing Indicator	*Tow Dur*ation (min.)	Diff in *Depth*	*Length* of Fish (cm)	*Handling* Time (min.)	Total *log(catch)* ln(weight)
100	353.0	1	30	15	39	5	5.685
109	111.0	1	100	5	44	29	8.690
113	64.0	0	100	10	53	4	5.323
116	500.0	1	100	10	44	4	5.323
....							

More Definitions and Notation

There are several equivalent ways to characterize the probability distribution of a survival random variable. Some of these are familiar; others are special to survival analysis. We will focus on the following terms:

- The density function $f(t)$
- The survivor function $S(t)$
- The hazard function $\lambda(t)$
- The cumulative hazard function $\Lambda(t)$

- **Density function (or Probability Mass Function) for discrete r.v.'s**
  Suppose that T takes values in $a_1, a_2, \ldots, a_n$.

$$\begin{aligned} f(t) &= Pr(T = t) \\ &= \begin{cases} f_j & \text{if } \; t = a_j, j = 1, 2, \ldots, n \\ 0 & \text{if } \; t \neq a_j, j = 1, 2, \ldots, n \end{cases} \end{aligned}$$

- **Density Function for continuous r.v.'s**

$$f(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T \leq t + \Delta t)$$

- **Survivorship Function**: $S(t) = P(T \geq t)$.

  In other settings, the cumulative distribution function, $F(t) = P(T \leq t)$, is of interest. In survival analysis, our interest tends to focus on the survival function, $S(t)$.

  For a continuous random variable:

$$S(t) = \int_t^\infty f(u)du$$

  For a discrete random variable:

$$\begin{aligned} S(t) &= \sum_{u \geq t} f(u) \\ &= \sum_{a_j \geq t} f(a_j) \\ &= \sum_{a_j \geq t} f_j \end{aligned}$$

Notes:

- From the definition of $S(t)$ for a continuous variable, $S(t) = 1 - F(t)$ as long as $f(t)$ is absolutely continuous
- For a discrete variable, we have to decide what to do if an event occurs exactly at time t; i.e., does that become part of $F(t)$ or $S(t)$?
- To get around this problem, several books define $S(t) = Pr(T > t)$, or else define $F(t) = Pr(T < t)$ (eg. Collett)

• **Hazard Function** $\lambda(t)$

Sometimes called an *instantaneous failure rate*, the *force of mortality*, or the *age-specific failure rate*.

– **Continuous random variables:**

$$\begin{aligned}
\lambda(t) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T < t + \Delta t | T \geq t) \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{Pr([t \leq T < t + \Delta t] \ \cap \ [T \geq t])}{Pr(T \geq t)} \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{Pr(t \leq T < t + \Delta t)}{Pr(T \geq t)} \\
&= \frac{f(t)}{S(t)}
\end{aligned}$$

– **Discrete random variables:**

$$\begin{aligned}
\lambda(a_j) \equiv \lambda_j &= Pr(T = a_j | T \geq a_j) \\
&= \frac{P(T = a_j)}{P(T \geq a_j)} \\
&= \frac{f(a_j)}{S(a_j)} \\
&= \frac{f(t)}{\Sigma_{k:a_k \geq a_j} f(a_k)}
\end{aligned}$$

• **Cumulative Hazard Function** $\Lambda(t)$

– **Continuous random variables:**

$$\Lambda(t) = \int_0^t \lambda(u) du$$

– **Discrete random variables:**

$$\Lambda(t) = \sum_{k:a_k < t} \lambda_k$$

Relationship between $S(t)$ and $\lambda(t)$

We've already shown that, for a continuous r.v.

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

For a left-continuous survivor function $S(t)$, we can show:

$$f(t) \;=\; -S'(t) \qquad \text{or} \quad S'(t) \;=\; -\,f(t)$$

We can use this relationship to show that:

$$\begin{aligned} -\,\frac{d}{dt}[\log S(t)] \;&=\; -\,\left(\frac{1}{S(t)}\right) S'(t) \\ &=\; -\,\frac{-f(t)}{S(t)} \\ &=\; \frac{f(t)}{S(t)} \end{aligned}$$

So another way to write $\lambda(t)$ is as follows:

$$\lambda(t) \;=\; -\frac{d}{dt}[\log S(t)]$$

Relationship between $S(t)$ and $\Lambda(t)$:

- **Continuous case:**

$$\begin{aligned} \Lambda(t) \;&=\; \int_0^t \lambda(u)du \\ &=\; \int_0^t \frac{f(u)}{S(u)}du \\ &=\; \int_0^t -\frac{d}{du}\log S(u)du \\ &=\; -\log S(t) + \log S(0) \\ &\Rightarrow S(t) = e^{-\Lambda(t)} \end{aligned}$$

- **Discrete case:**

  Suppose that $a_j < t \leq a_{j+1}$. Then

$$\begin{aligned} S(t) \;&=\; P(T \geq a_1, T \geq a_2, \ldots, T \geq a_{j+1}) \\ &=\; P(T \geq a_1)P(T \geq a_2 | T \geq a_1) \cdots P(T \geq a_{j+1} | T \geq a_j) \\ &=\; (1-\lambda_1) \times \cdots \times (1-\lambda_j) \\ &=\; \prod_{k:a_k<t} (1-\lambda_k) \end{aligned}$$

  Cox defines $\Lambda(t) = \Sigma_{k:a_k<t} \log(1-\lambda_k)$ so that $S(t) = e^{-\Lambda(t)}$ in the discrete case, as well.

Measuring Central Tendency in Survival

- **Mean survival** - call this μ

$$\mu = \int_0^\infty u f(u) du \quad \text{for continuous } T$$

$$= \sum_{j=1}^{n} a_j f_j \quad \text{for discrete } T$$

- **Median survival** - call this τ, is defined by

$$S(\tau) = 0.5$$

Similarly, any other percentile could be defined.

In practice, we don't usually hit the median survival at exactly one of the failure times. In this case, the estimated median survival is the *smallest* time τ such that

$$\hat{S}(\tau) \leq 0.5$$

Some hazard shapes seen in applications:

- **increasing**

  e.g. aging after 65

- **decreasing**

  e.g. survival after surgery

- **bathtub**

  e.g. age-specific mortality

- **constant**

  e.g. survival of patients with advanced chronic disease

Estimating the survival or hazard function

We can estimate the survival (or hazard) function in two ways:

- by specifying a parametric model for $\lambda(t)$ based on a particular density function $f(t)$
- by developing an empirical estimate of the survival function (i.e., non-parametric estimation)

If no censoring:
The empirical estimate of the survival function, $\tilde{S}(t)$, is the proportion of individuals with event times greater than t.

With censoring:
If there are censored observations, then $\tilde{S}(t)$ is not a good estimate of the true $S(t)$, so other non-parametric methods must be used to account for censoring (life-table methods, Kaplan-Meier estimator)

Some Parametric Survival Distributions

- The **Exponential** distribution (1 parameter)

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \text{ for } t \geq 0 \\ S(t) &= \int_t^\infty f(u)du \\ &= e^{-\lambda t} \\ \lambda(t) &= \frac{f(t)}{S(t)} \\ &= \lambda \qquad \text{constant hazard!} \\ \Lambda(t) &= \int_0^t \lambda(u)\, du \\ &= \int_0^t \lambda\, du \\ &= \lambda t \end{aligned}$$

Check: Does $S(t) = e^{-\Lambda(t)}$?

median: solve $0.5 = S(\tau) = e^{-\lambda\tau}$:

$$\Rightarrow \tau = \frac{-\log(0.5)}{\lambda}$$

mean:

$$\int_0^\infty u\lambda e^{-\lambda u} du = \frac{1}{\lambda}$$

- The **Weibull** distribution (2 parameters)
  Generalizes exponential:

$$\begin{aligned} S(t) &= e^{-\lambda t^\kappa} \\ f(t) &= \frac{-d}{dt} S(t) = \kappa\lambda t^{\kappa-1} e^{-\lambda t^\kappa} \\ \lambda(t) &= \kappa\lambda t^{\kappa-1} \\ \Lambda(t) &= \int_0^t \lambda(u)du = \lambda t^\kappa \end{aligned}$$

λ - the *scale* parameter
κ - the *shape* parameter

The Weibull distribution is convenient because of its simple form. It includes several hazard shapes:

$\kappa = 1 \rightarrow$ constant hazard

$0 < \kappa < 1 \rightarrow$ decreasing hazard

$\kappa > 1 \rightarrow$ increasing hazard

- **Rayleigh** distribution
  Another 2-parameter generalization of exponential:

$$\lambda(t) = \lambda_0 + \lambda_1 t$$

- **compound exponential**

  $T \sim \exp(\lambda)$, $\lambda \sim g$

$$f(t) = \int_0^\infty \lambda e^{-\lambda t} g(\lambda) d\lambda$$

- **log-normal**, **log-logistic**:
  Possible distributions for T obtained by specifying for $\log T$ any convenient family of distributions, e.g.

  $\log T \sim$ normal (non-monotone hazard)

  $\log T \sim$ logistic

Why use one versus another?

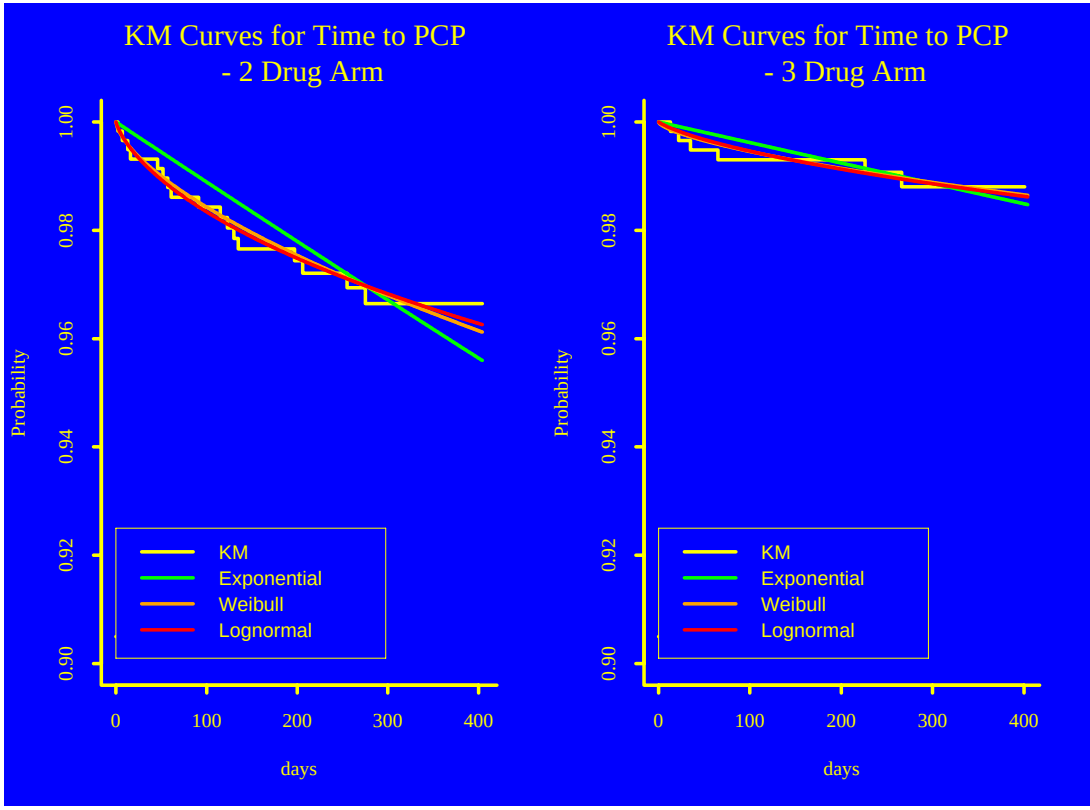
- technical convenience for estimation and inference
- explicit simple forms for $f(t), S(t)$, and $\lambda(t)$.
- qualitative shape of hazard function

One can usually distinguish between a one-parameter model (like the exponential) and two-parameter (like Weibull or log-normal) in terms of the adequacy of fit to a dataset.

Without a lot of data, it may be hard to distinguish between the fits of various 2-parameter models (i.e., Weibull vs log-normal)

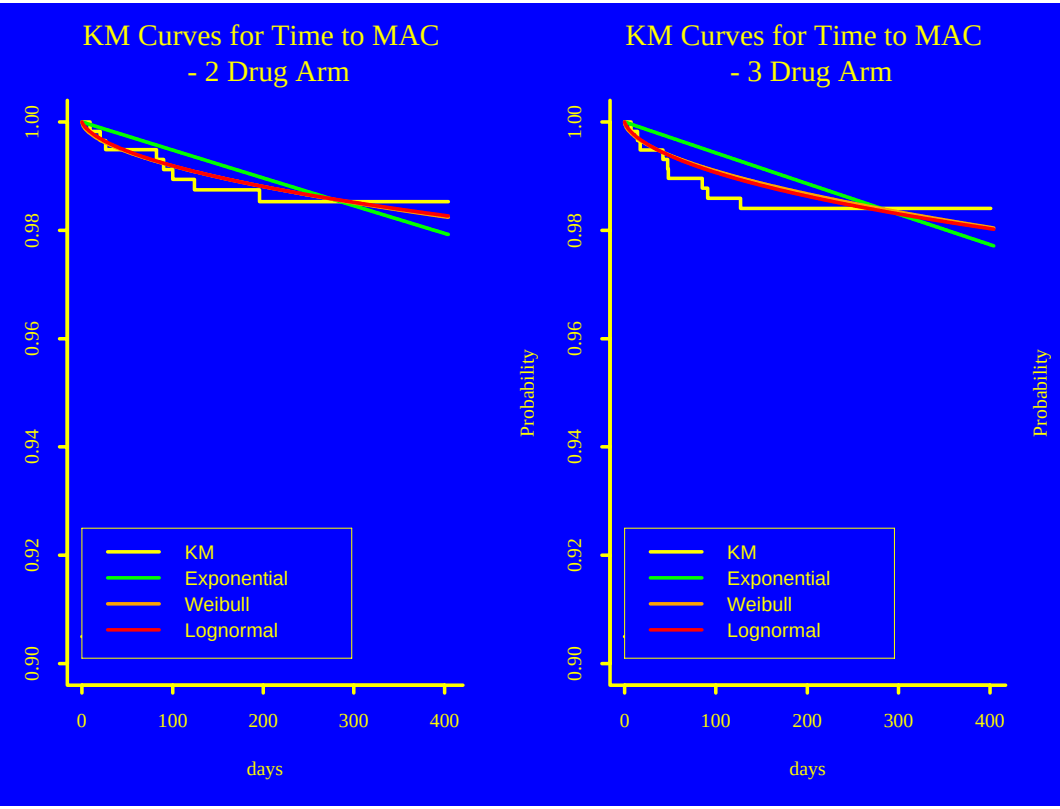
Plots of estimates of $S(t)$

Based on KM, exponential, Weibull, and log-normal
for study of protease inhibitors in AIDS patients
(ACTG 320)



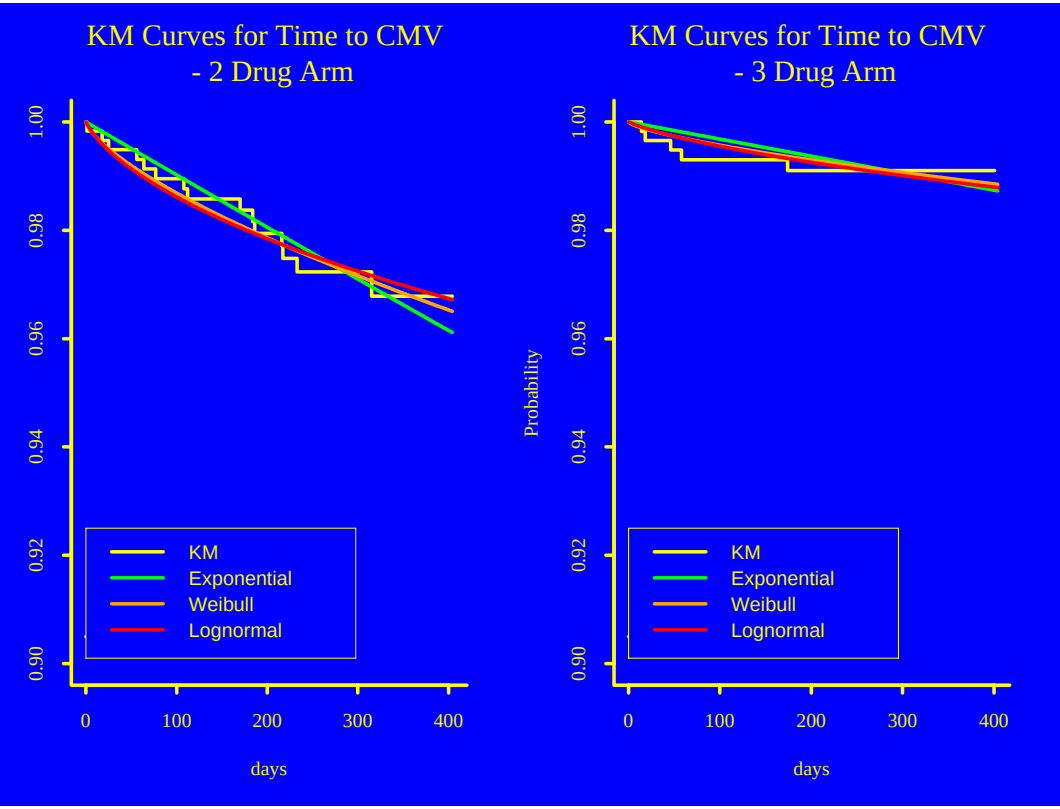
Plots of estimates of $S(t)$

Based on KM, exponential, Weibull, and log-normal
for study of protease inhibitors in AIDS patients
(ACTG 320)



Plots of estimates of $S(t)$

Based on KM, exponential, Weibull, and log-normal
for study of protease inhibitors in AIDS patients
(ACTG 320)



Preview of Coming Attractions

Next we will discuss the most famous non-parametric approach for estimating the survival distribution, called the *Kaplan-Meier estimator*.

To motivate the derivation of this estimator, we will first consider a set of survival times where there is no censoring.

The following are **times to relapse** (weeks) for 21 leukemia patients receiving control treatment (Table 1.1 of Cox & Oakes):

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

How would we estimate S(10), the probability that an individual survives to time 10 or later?

What about $\tilde{S}(8)$? Is it $\frac{12}{21}$ or $\frac{8}{21}$?

Let's construct a table of $\tilde{S}(t)$:

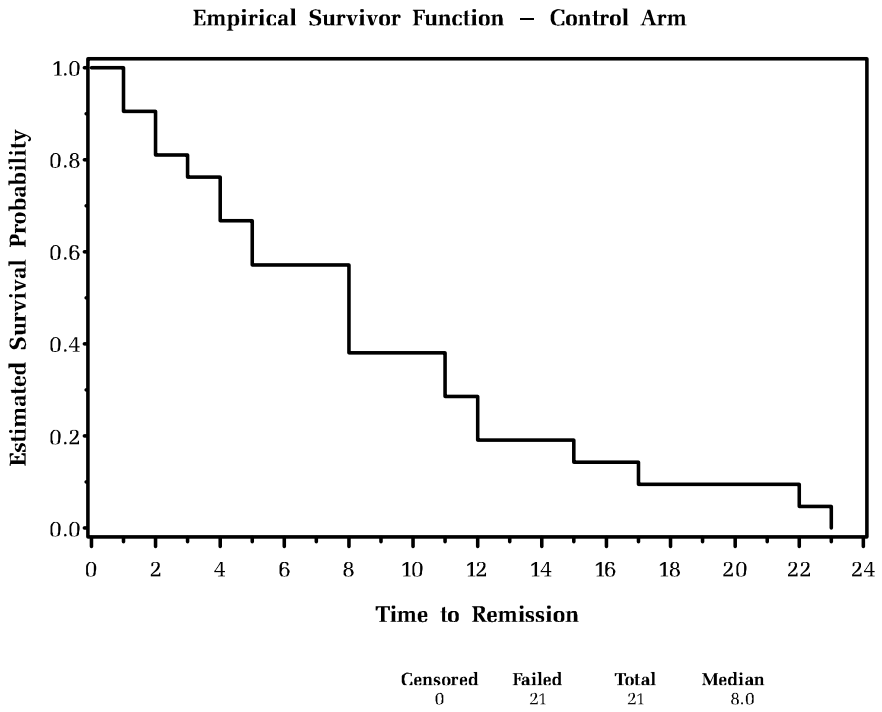
Values of t	$\hat{S}(t)$
$t \leq 1$	21/21=1.000
$1 < t \leq 2$	19/21=0.905
$2 < t \leq 3$	17/21=0.809
$3 < t \leq 4$	
$4 < t \leq 5$	
$5 < t \leq 8$	
$8 < t \leq 11$	
$11 < t \leq 12$	
$12 < t \leq 15$	
$15 < t \leq 17$	
$17 < t \leq 22$	
$22 < t \leq 23$	

Empirical Survival Function:
When there is no censoring, the general formula is:

$$\tilde{S}(t) = \frac{\#\ individuals\ with\ T \geq t}{total\ sample\ size}$$

In most software packages, the survival function is evaluated just after time t, i.e., at t^+. In this case, we only count the individuals with $T > t$.

Example for leukemia data (control arm):



Stata Commands for Survival Estimation

```
.use leukem

.stset remiss status if trt==0       (to keep only untreated patients)
(21 observations deleted)

. sts list
         failure _d:  status
   analysis time _t:  remiss
```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
1	21	2	0	0.9048	0.0641	0.6700	0.9753
2	19	2	0	0.8095	0.0857	0.5689	0.9239
3	17	1	0	0.7619	0.0929	0.5194	0.8933
4	16	2	0	0.6667	0.1029	0.4254	0.8250
5	14	2	0	0.5714	0.1080	0.3380	0.7492
8	12	4	0	0.3810	0.1060	0.1831	0.5778
11	8	2	0	0.2857	0.0986	0.1166	0.4818
12	6	2	0	0.1905	0.0857	0.0595	0.3774
15	4	1	0	0.1429	0.0764	0.0357	0.3212
17	3	1	0	0.0952	0.0641	0.0163	0.2612
22	2	1	0	0.0476	0.0465	0.0033	0.1970
23	1	1	0	0.0000	.	.	.

```
.sts graph
```

SAS Commands for Survival Estimation

```
data leuk;
 input t;
cards;
1
1
2
2
3
4
4
5
5
8
8
8
11
11
12
12
15
17
22
23
;

proc lifetest data=leuk;
 time t;
run;
```


SAS Output for Survival Estimation

The LIFETEST Procedure

Product-Limit Survival Estimates

t	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	21
1.0000	.	.	.	1	20
1.0000	0.9048	0.0952	0.0641	2	19
2.0000	.	.	.	3	18
2.0000	0.8095	0.1905	0.0857	4	17
3.0000	0.7619	0.2381	0.0929	5	16
4.0000	.	.	.	6	15
4.0000	0.6667	0.3333	0.1029	7	14
5.0000	.	.	.	8	13
5.0000	0.5714	0.4286	0.1080	9	12
8.0000	.	.	.	10	11
8.0000	.	.	.	11	10
8.0000	.	.	.	12	9
8.0000	0.3810	0.6190	0.1060	13	8
11.0000	.	.	.	14	7
11.0000	0.2857	0.7143	0.0986	15	6
12.0000	.	.	.	16	5
12.0000	0.1905	0.8095	0.0857	17	4
15.0000	0.1429	0.8571	0.0764	18	3
17.0000	0.0952	0.9048	0.0641	19	2
22.0000	0.0476	0.9524	0.0465	20	1
23.0000	0	1.0000	0	21	0

SAS Output for Survival Estimation (cont'd)

Summary Statistics for Time Variable t

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	12.0000	8.0000	17.0000
50	8.0000	4.0000	11.0000
25	4.0000	2.0000	8.0000

Mean	Standard Error
8.6667	1.4114

Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	Percent Censored
21	21	0	0.00

Does anyone have a guess regarding how to calculate the standard error of the estimated survival?

$$\hat{S}(8^+) \;=\; P(T > 8) = \frac{8}{21} \;=\; 0.381$$

(at $t = 8^+$, we count the 4 events at time=8 as already having failed)

$$se[\hat{S}(8^+)] \;=\; 0.106$$