Survival Analysis

Decision Systems Group
Brigham and Women's Hospital

Harvard-MIT Division of Health Sciences and Technology
HST.951J: Medical Decision Support

Outline

- Basic concepts & distributions
  - Survival, hazard
  - Parametric models
  - Non-parametric models
- Simple models
  - Life-table
  - Product-limit
- Multivariate models
  - Cox proportional hazard
  - Neural nets

What we are trying to do

Predict survival

(or probability of survival)

	Variable 1	Variable 2	days
Case 1	0.7	-0.2	8
Case 2	0.6	0.5	4
	-0.6	0.1	2
	0	-0.9	3
	-0.4	0.4	2
	-0.8	0.6	3
	0.5	-0.7	4

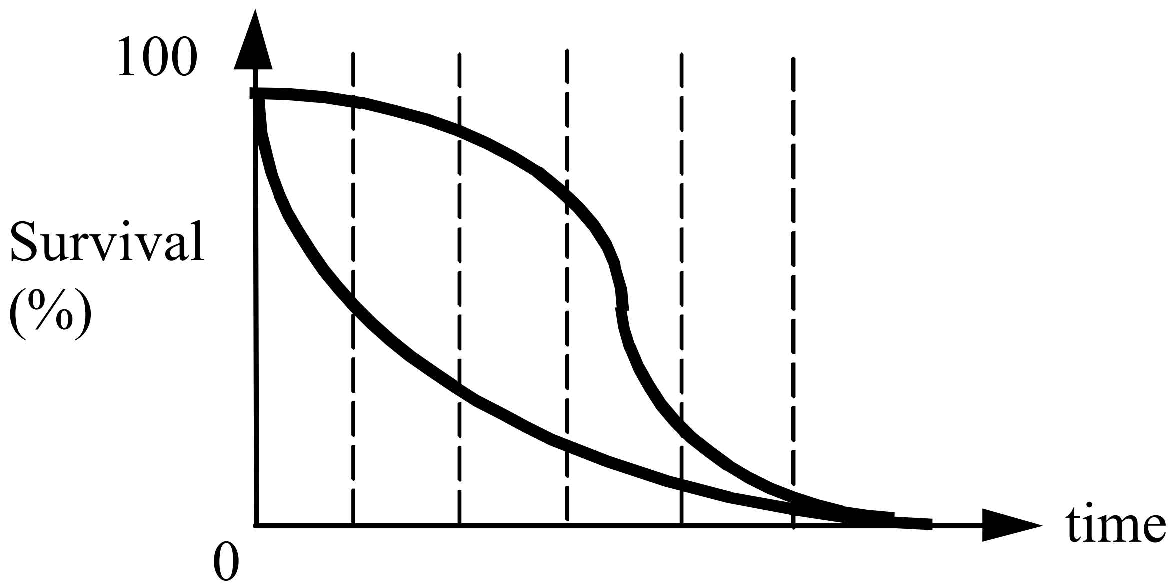
Using these

- and evaluate performance on new cases
- and determine which variables are important

Survival function

Probability that an individual survives at least t

- $S(t) = P(T > t)$
- By definition, $S(0) = 1$ and $S(\infty)=0$
- Estimated by (# survivors at *t* / total patients)

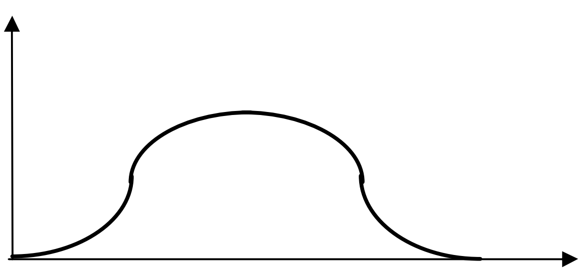


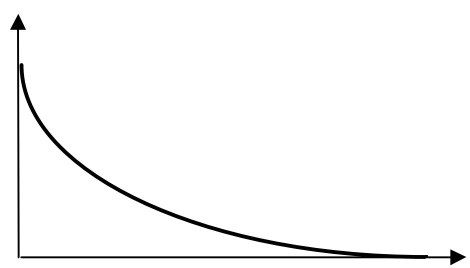
Unconditional failure rate

- Probability density function (of death)
- $f(t) = \lim_{\Delta t \to 0} P(\text{individual dies } (t, t+\Delta t)) / \Delta t$
- f(t) always non-negative
- Area below density is 1
- Estimated by

patients dying in the interval/(total patients*interval_width)

Same as # patients dying per unit interval/total





Some other definitions

- Just like $S(t)$ is "cumulative" survival, $F(t)$ is cumulative death probability
- $S(t) = 1 – F(t)$
- $f(t) = - S'(t)$

Conditional failure rate

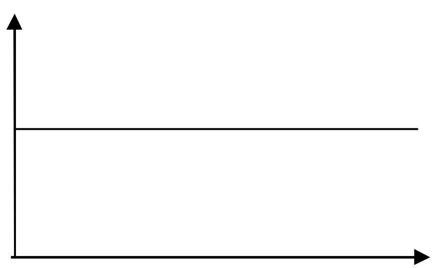
- AKA Hazard function
- $h(t) = \lim_{\Delta t \to 0} P(\text{individual aged t dies } (t, t+\Delta t)) / \Delta t$
- h(t) is instantaneous failure rate
- Estimated by

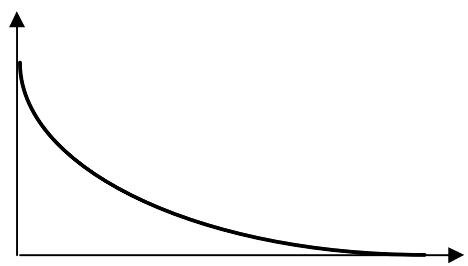
patients dying in the interval/(survivors at t *interval_width)

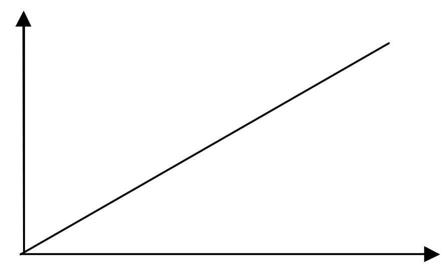
- So can be estimated by

patients dying per unit interval/survivors at t

$h(t) = f(t)/S(t)$ $\quad\quad$ $h(t) = -S'(t)/S(t) = -d \log S(t)/dt$



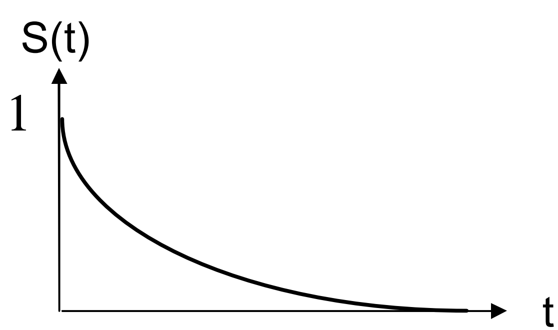


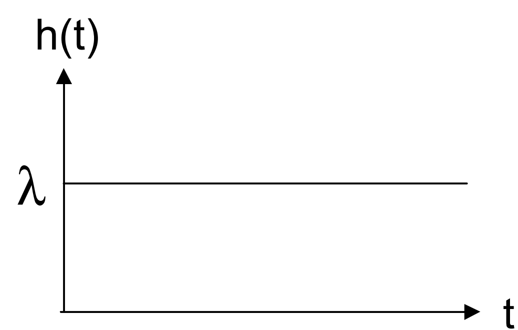


Parametric estimation

Example: Exponential

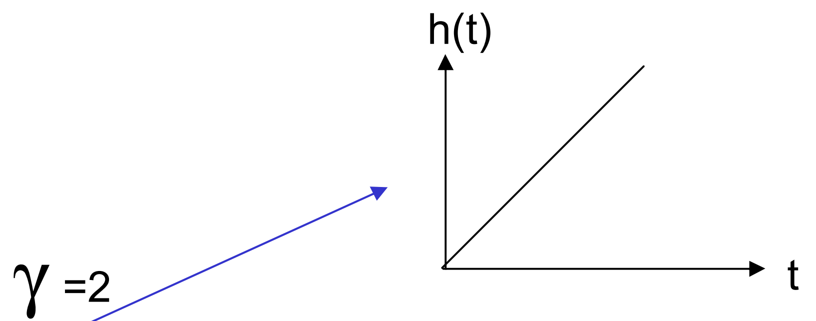
- $f(t) = \lambda e^{-\lambda t}$
- $S(t) = e^{-\lambda t}$
- $h(t) = \lambda$

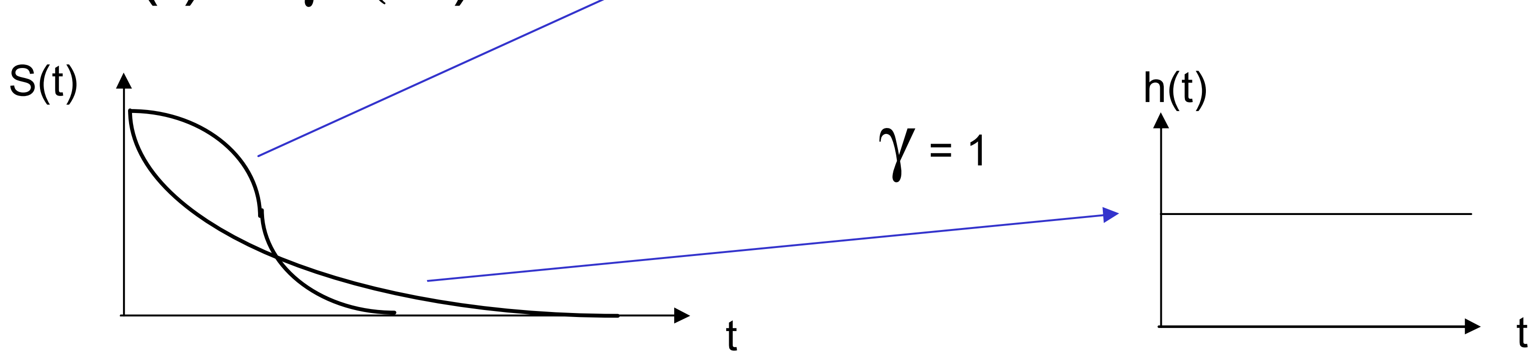




Weibull distribution

- Generalization of the exponential
- For $\lambda,\gamma > 0$
- $f(t) = \gamma\lambda(\lambda t)^{\gamma-1} e^{-\lambda t^{\gamma}}$
- $S(t) = e^{-\lambda t^{\gamma}}$
- $h(t) = \gamma\lambda(\lambda t)^{\gamma-1}$

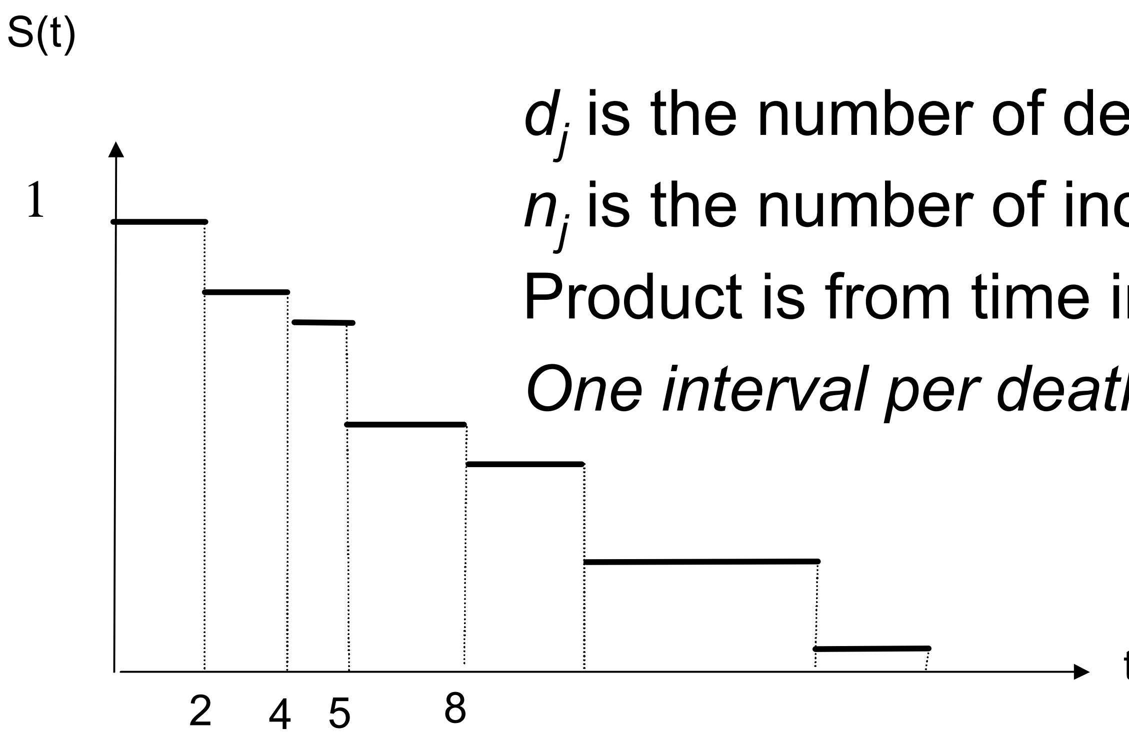




Non-Parametric estimation

Product-Limit (Kaplan-Meier)

$S(t_i) = \Pi\ (n_j - d_j)/ n_j$



d_j is the number of deaths in interval *j*

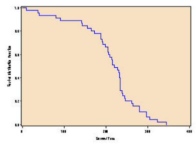
n_j is the number of individuals at risk

Product is from time interval 1 to *j*

One interval per death time

Kaplan-Meier

- Example
- Deaths: 10, 37, 40, 80, 91,143, 164, 188, 188, 190, 192, 206, …



Life-Tables

- AKA actuarial method

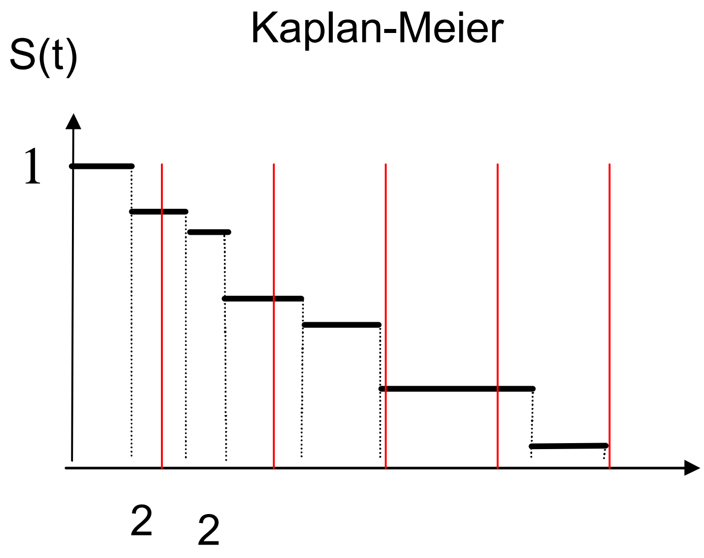
$S(t_i) = \Pi\ (n_j - d_j)/ n_j$

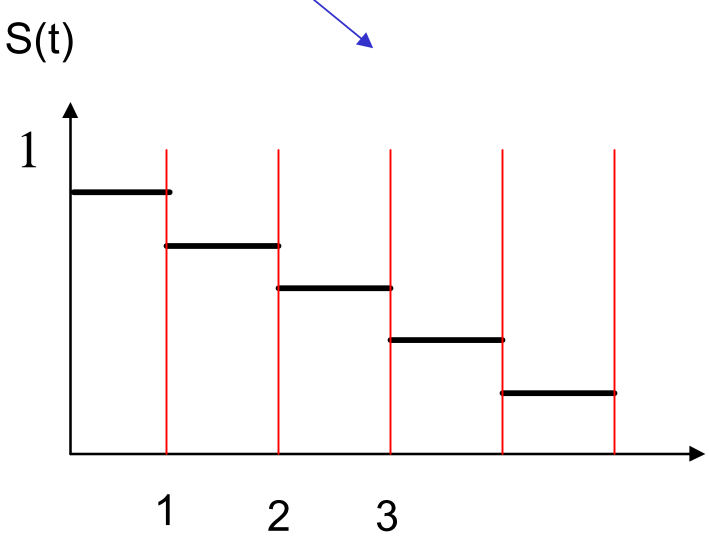
d_j is the number of deaths in interval j

n_j is the number of individuals at risk

Product is from time interval 1 to j

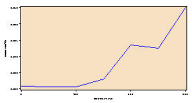
- Pre-defined intervals j are independent of death times



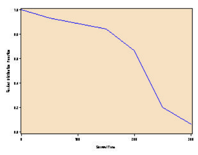


Life-Table

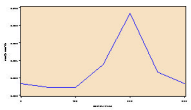
hazard



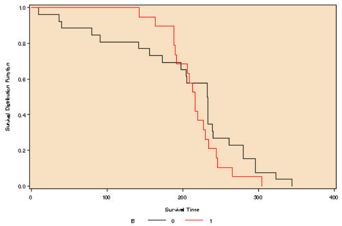
survival



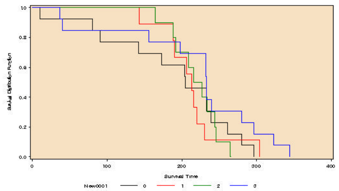
density



Simple models



Multiple strata



Multivariate models

- Several strata, each defined by a set of variable values
- Could potentially go as far as “one stratum per case”?
- Can it do prediction for individuals?

Cox Proportional Hazards

- Regression model
- Can give estimate of hazard for a particular individual relative to baseline hazard at a particular point in time
- Baseline hazard can be estimated by, for example, by using survival from the Kaplan-Meier method

Proportional Hazards

$$\lambda_i = \lambda\, e^{-\beta x_i}$$

where λ is baseline hazard and x_i is covariate for patient

Cox proportional hazards

$$h_i(t) = h_0(t)\, e^{\beta x_i}$$

- Survival

$$S_i(t) = [S_0(t)]^{e^{\beta x_i}}$$

Cox Proportional Hazards

$$h_i(t) = h_0(t)\, e^{\beta x_i}$$

- New likelihood function is how we extimate β
- From the set of individuals at risk at time j (R_j), the probability of picking exactly the one who died is

$$\frac{h_0(t)\, e^{\beta x_i}}{\Sigma_m\, h_0(t)\, e^{\beta x_m}}$$

- Then likelihood function to maximize to all j is
- $L(\beta) = \Pi\, (e^{\beta x_i} / \Sigma_m\, e^{\beta x_m})$

Important details

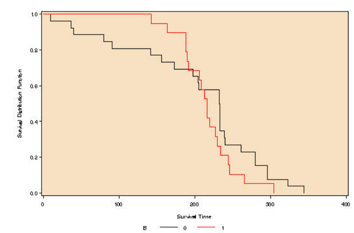
- Survival curves can't cross if hazards are proportional
- There is a common baseline h_0, but we don't need to know it to estimate the coefficients
- We don't need to know the shape of hazard function
- Cox model is commonly used to interpret importance of covariates (amenable to variable selection methods)
- It is the most popular multivariate model for survival
- Testing the proportionality assumption is difficult and hardly ever done

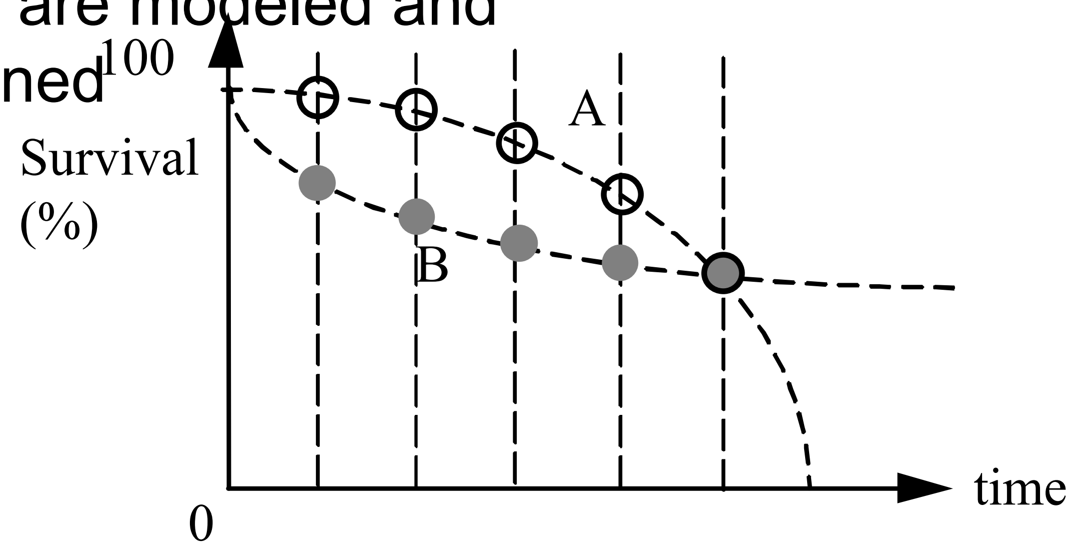
Estimating survival for a patient using the Cox model

- Need to estimate the baseline
- Can use parametric or non-parametric model to estimate the baseline
- Can then create a continuous "survival curve estimate" for a patient
- Baseline survival can be, for example:
  - the survival for a case in which all covariates are set to their means
  - Kaplan-Meier estimate for all cases

What if the proportionality assumption is not OK?

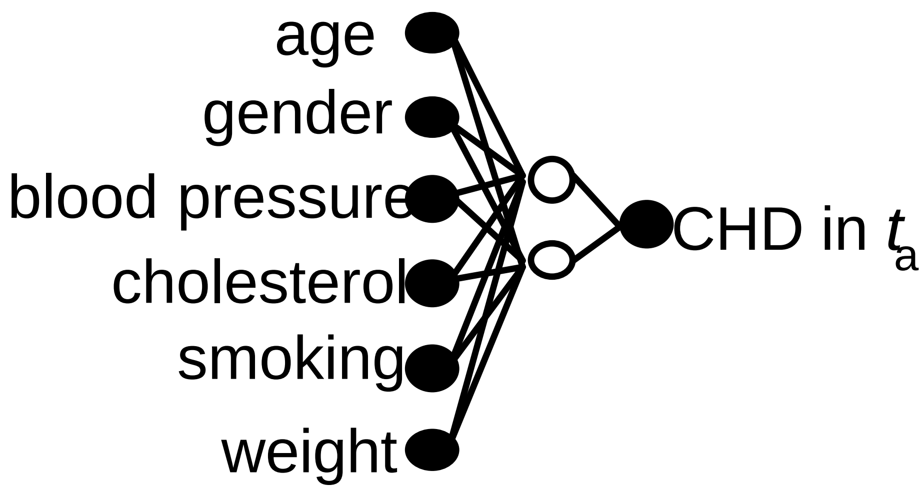
- Survival curves may cross
- Other multivariate models can be built
- Survival at certain time points are modeled and combined





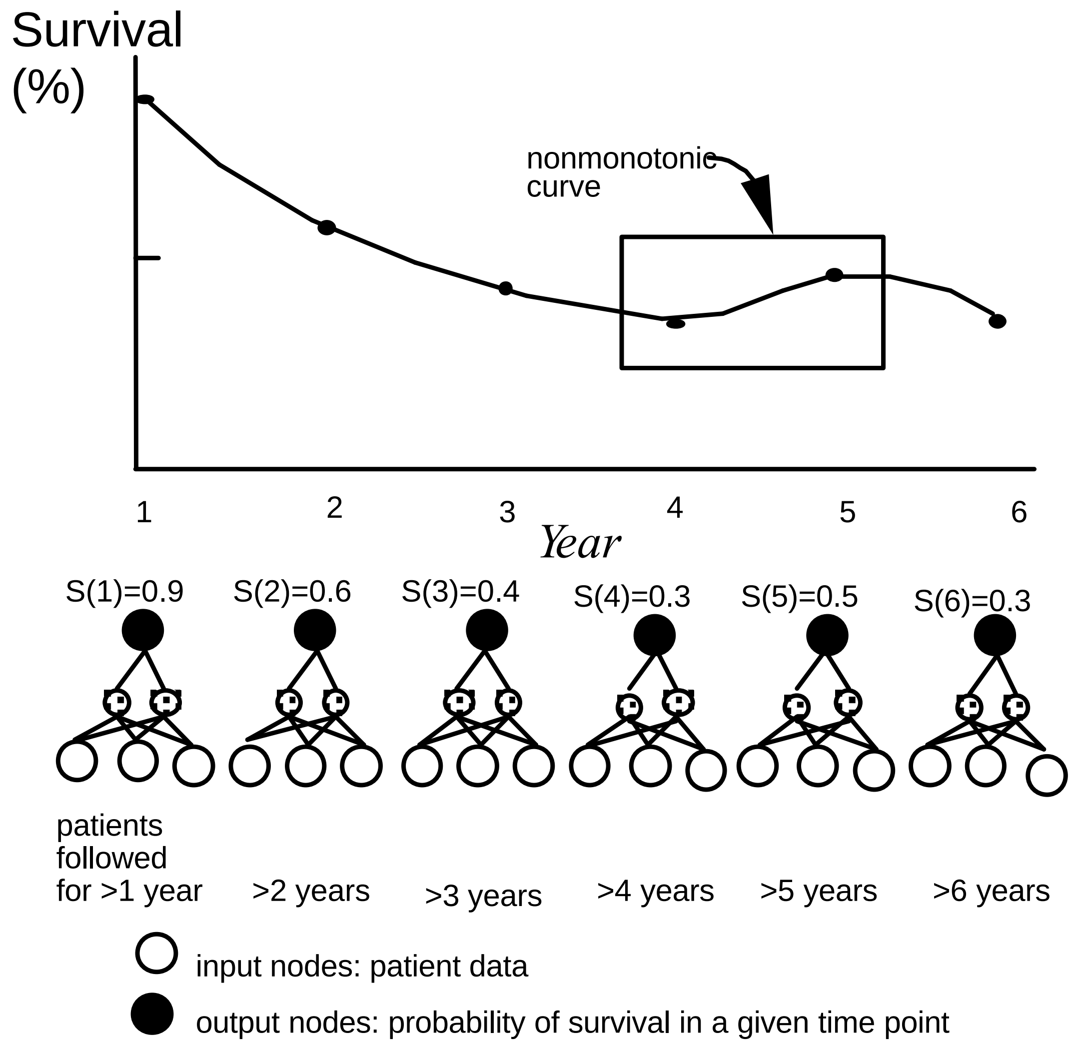
Single-point models

- Logistic regression
- Neural nets



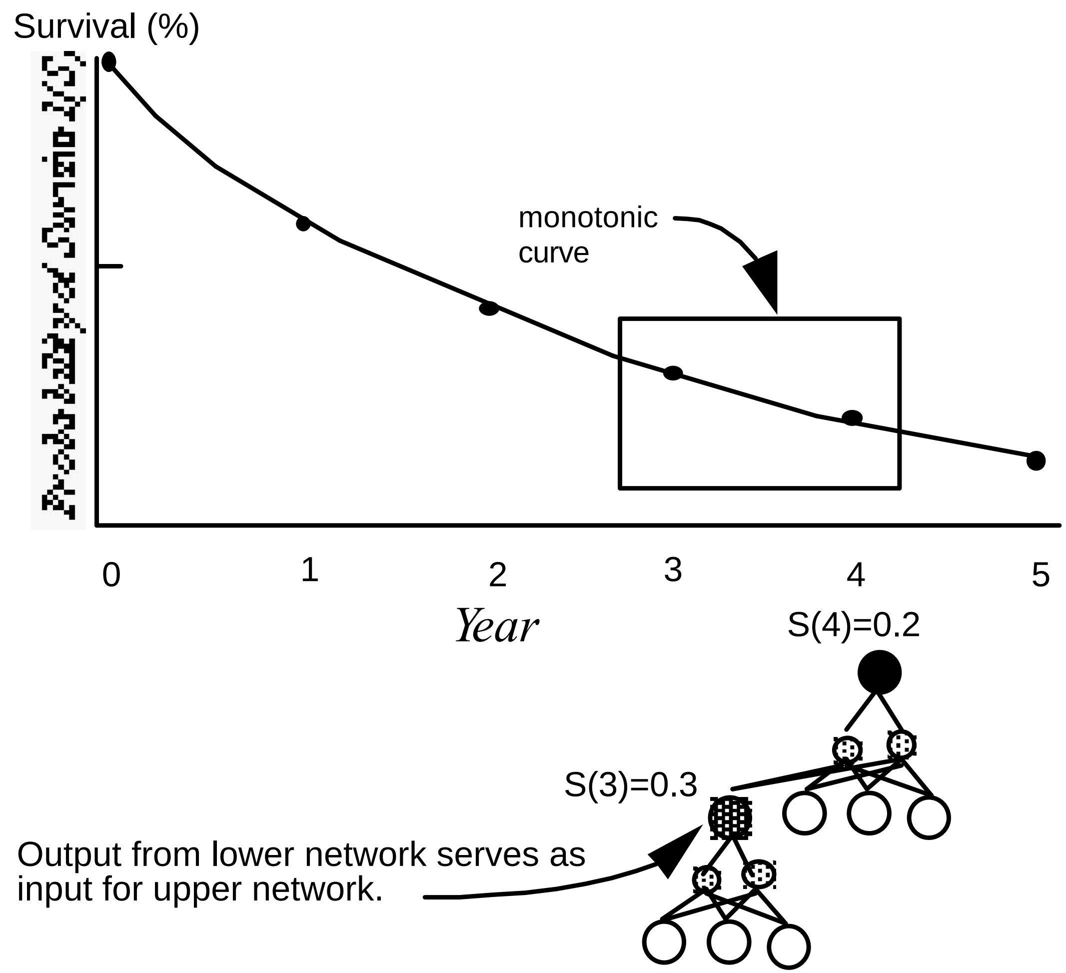
Problems

- Dependency between intervals is not modeled (no links between networks)
- Nonmonotonic curves may appear
- How to evaluate?



Accounting for dependencies

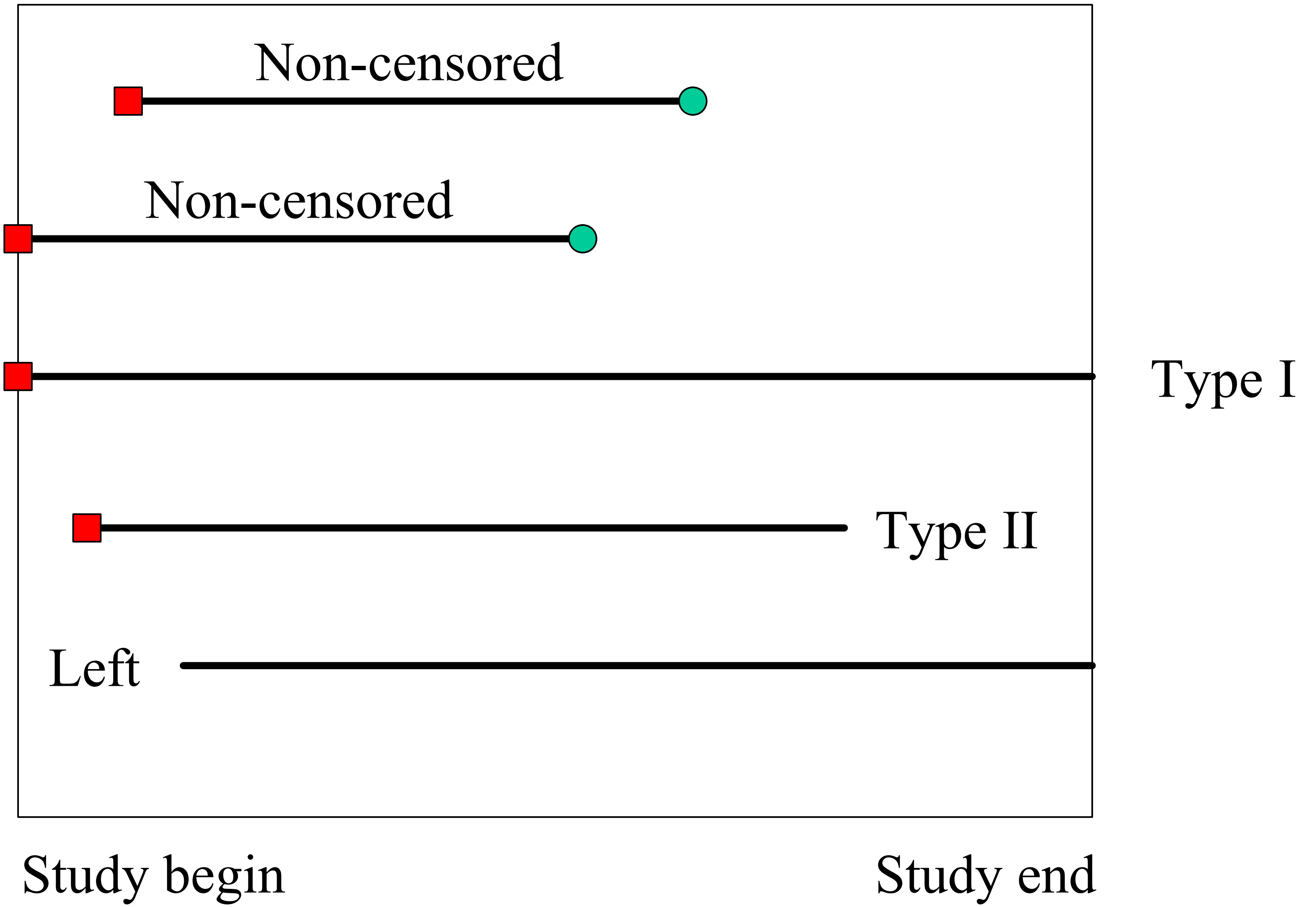
- “Link” networks in some way to account for dependencies



Summary

- Kaplan-Meier for simple descriptive analysis
- Cox Proportional for multivariate prediction if survival curves don’t cross
- Other methods for multivariate survival exist: logistic regression, neural nets, CART, etc.

Censoring


Non-censored
Non-censored
Type I
Type II
Left
Study begin
Study end