Machine Learning for Healthcare

6.871Jx

Survival Modeling

David Sontag

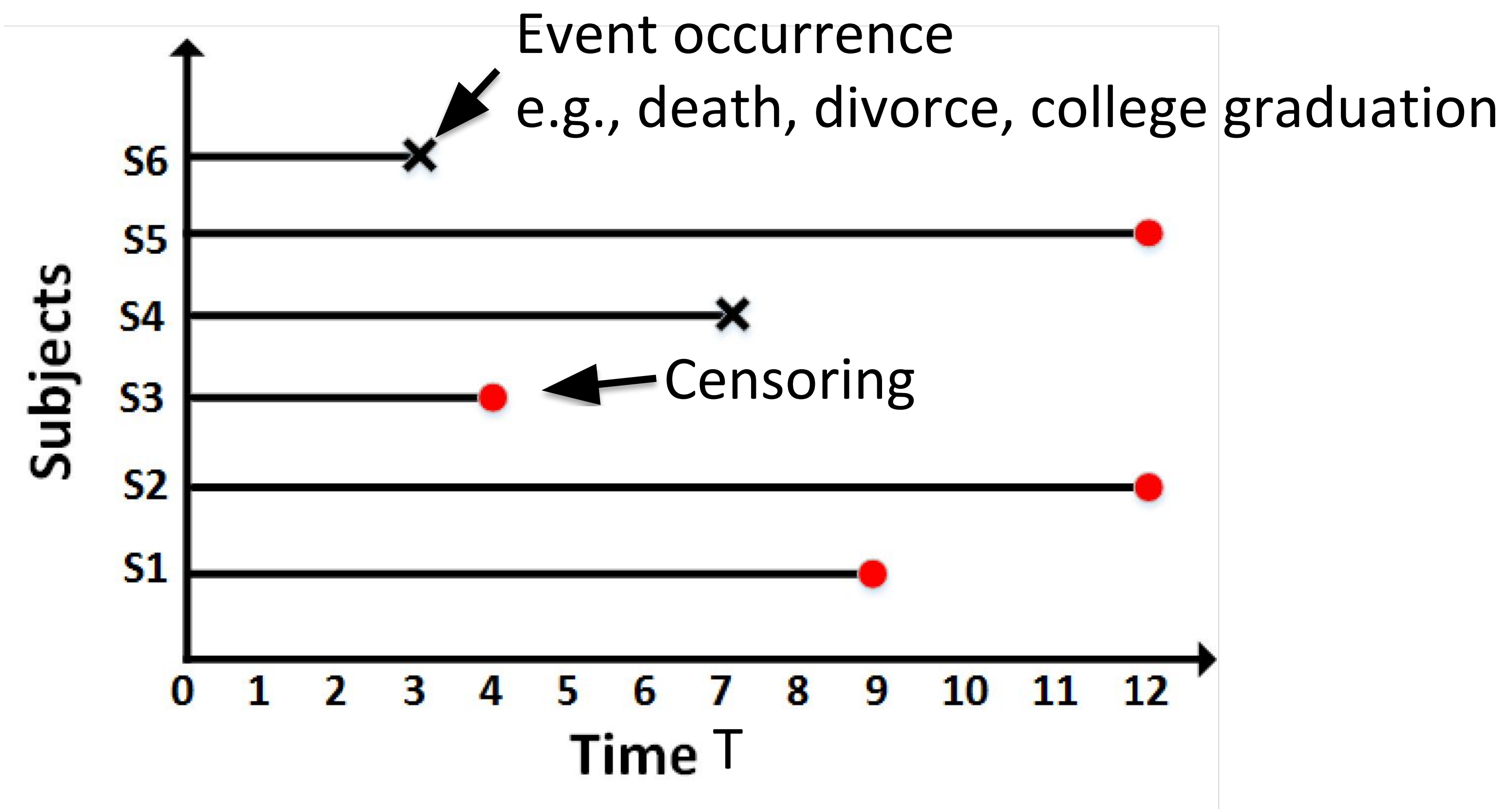
CSAIL

INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

HEALTH SCIENCES & TECHNOLOGY

Survival modeling

- We focus on right-censored data:



[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

Survival modeling

- Why not use classification, as before?
  - Less data for training (due to exclusions)
  - Pessimistic estimates due to choice of window
- What about regression, e.g. minimizing mean-squared error?
  - T is non-negative, may want long tails
  - If we just naively removed censored events, we would be introducing bias

Notation and formalization

- Data are (**x**, T, b)=(features, time, censoring), where *b=0,1* denotes whether time is of censoring or event occurrence
- Let f(t) = P(t) be the probability of death at time t
- Survival function: the probability of an individual surviving beyond time *t*,

$$S(t) = P(T > t) = \int_t^{\infty} f(x)dx$$

[Ha, Jeong, Lee. Statistical Modeling of Survival Data with Random Effects. Springer 2017]

Notation and formalization

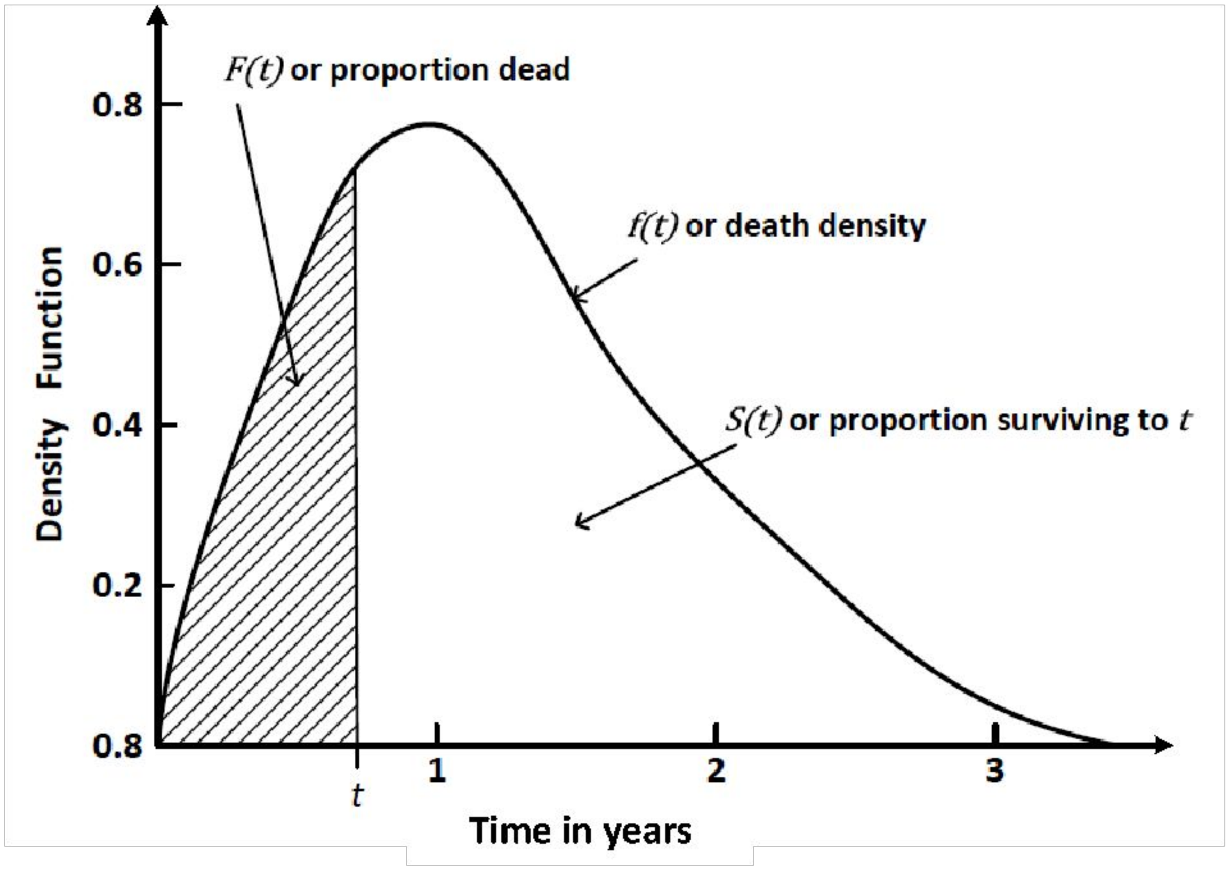
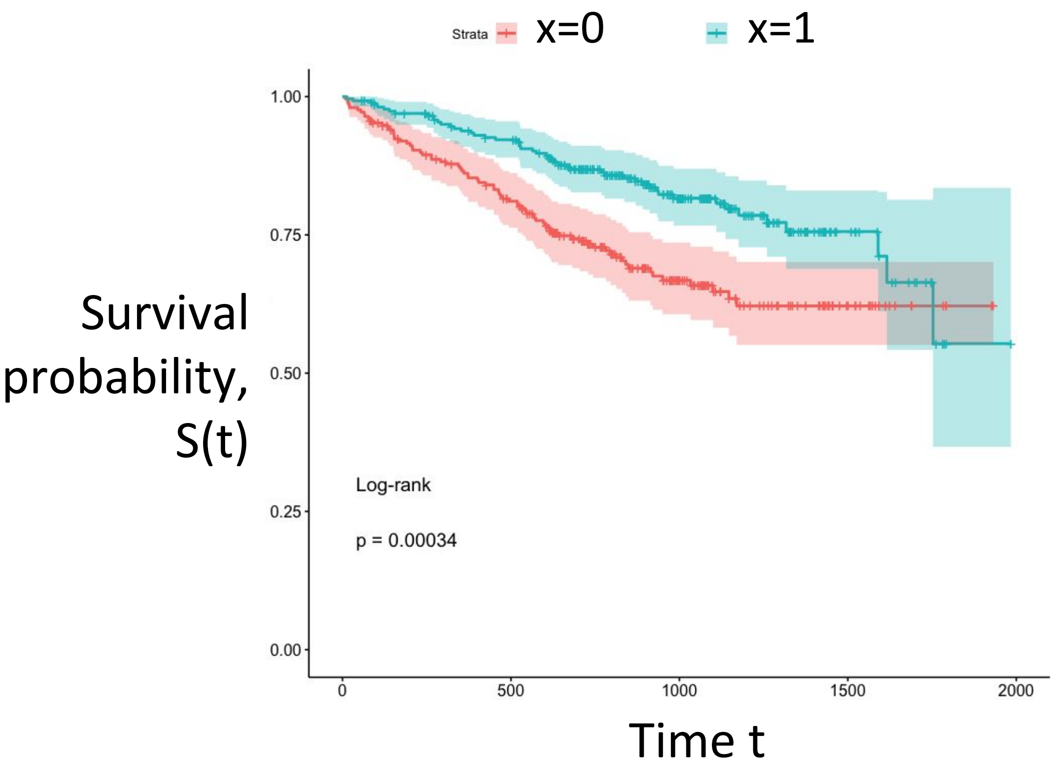


Fig. 2: Relationship among different entities $f(t)$, $F(t)$ and $S(t)$.

[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

Kaplan-Meier estimator

- Example of a non-parametric method; good for unconditional density estimation



Observed event times

$$y_{(1)} < y_{(2)} < \cdots < y_{(D)}$$

$d_{(k)}$ = # events at this time

$n_{(k)}$ = # of individuals alive and uncensored

$$\widehat{S}_{K-M}(t) = \prod_{k:y_{(k)} \leq t} \left\{ 1 - \frac{d_{(k)}}{n_{(k)}} \right\}$$

[Figure credit: Rebecca Peyser]

Maximum likelihood estimation

- Commonly parametric densities for f(t):

Table 2.1 Useful parametric distributions for survival analysis

Distribution	Survival function $S(t)$	Density function $f(t)$
Exponential $(\lambda > 0)$	$\exp(-\lambda t)$	$\lambda \exp(-\lambda t)$
Weibull $(\lambda, \phi > 0)$	$\exp(-\lambda t^{\phi})$	$\lambda \phi t^{\phi-1} \exp(-\lambda t^{\phi})$
Log-normal $(\sigma > 0, \mu \in R)$	$1 - \Phi\{(\ln t - \mu)/\sigma\}$	$\varphi\{(\ln t - \mu)/\sigma\}(\sigma t)^{-1}$
Log-logistic $(\lambda > 0, \phi > 0)$	$1/(1 + \lambda t^{\phi})$	$(\lambda \phi t^{\phi-1})/(1 + \lambda t^{\phi})^2$
Gamma $(\lambda, \phi > 0)$	$1 - I(\lambda t, \phi)$	$\{\lambda^{\phi}/\Gamma(\phi)\}t^{\phi-1} \exp(-\lambda t)$
Gompertz $(\lambda, \phi > 0)$	$\exp\{\frac{\lambda}{\phi}(1 - e^{\phi t})\}$	$\lambda e^{\phi t} \exp\{\frac{\lambda}{\phi}(1 - e^{\phi t})\}$

(parameters can be a function of x)

[Ha, Jeong, Lee. Statistical Modeling of Survival Data with Random Effects. Springer 2017]

Maximum likelihood estimation

- Two kinds of observations: censored and uncensored

Uncensored likelihood

$$p_{\boldsymbol{\theta}}(T = t \mid \mathbf{x}) = f(t)$$

Censored likelihood

$$p_{\boldsymbol{\theta}}^{\text{censored}}(t \mid \mathbf{x}) = p_{\boldsymbol{\theta}}(T > t \mid \mathbf{x}) = S(t)$$

- Putting the two together, we get:

$$\sum_{i=1}^{n} b_i \log p_{\boldsymbol{\theta}}^{\text{censored}}(t \mid \mathbf{x}) + (1 - b_i) \log p_{\boldsymbol{\theta}}(t \mid \mathbf{x})$$

Optimize via gradient or stochastic gradient ascent!

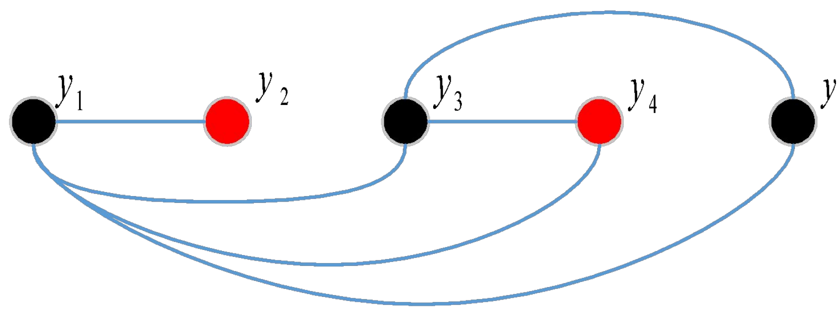
Evaluation for survival modeling

- Concordance-index (also called C-statistic): look at model's ability to predict *relative* survival times:

$$\hat{c} = \frac{1}{num} \sum_{i:b_i=0} \sum_{j:y_i<y_j} I[S(\hat{y}_j|X_j) > S(\hat{y}_i|X_i)]$$

- Illustration – blue lines denote pairwise comparisons:

Black = uncensored
Red = censored



- Equivalent to AUC for binary variables and no censoring

[Wang, Li, Reddy. Machine Learning for Survival Analysis: A Survey. 2017]

Final thoughts on survival modeling

- Could also evaluate:
  - Mean-squared error for uncensored individuals
  - Held-out (censored) likelihood
  - Derive binary classifier from learned model and check calibration

- Partial likelihood estimators (e.g. for cox-proportional hazards models) can be much more data efficient