



Modern methods for analysing survival and time-to-event data

Short course 28 June 2026

Ørnulf Borgan

Clara Bertinelli Salucci

The course contains **classic** and **newer** material:

- Introduction, counting processes, intensity processes and martingales (Ørnulf)
- The Nelson-Aalen and Kaplan-Meier estimators (Ørnulf)
- Regression models for time-to-event data: Cox regression and Aalen's additive regression model (Clara)
- Tutorial
- Unobserved heterogeneity (frailty) in survival analysis, frailty models for clustered data (Ørnulf)
- First hitting time models and threshold regression, including boosting (Clara)
- Cox regression with neural networks (Ørnulf)

Literature

Classic material:

Aalen O.O., Borgan Ø., and Gjessing H.K. (2008). *Survival and Event History Analysis: A Process Point of View*. Springer-Verlag, New York. Selected parts of chapters 1-7 (ebook available at many universities)

First hitting time models and threshold regression

Bertinelli Salucci C., Bakdi A., Glad I.K., Lindqvist B.H., Vanem E., and De Bin R. (2025). Lifetime analysis with monotonic degradation: a boosted first hitting time model based on a homogeneous gamma process. *Lifetime Data Analysis*, **31**: 300-339 (Open access)

Cox regression with neural networks

Kvamme H., Borgan Ø., and Scheel, I. (2019). Time-to-event prediction with neural networks and Cox regression. *Journal of Machine Learning Research*, 20: 1-30. (Open access)

What is survival analysis?

Survival analysis is a set of **statistical concepts, models and methods** for analysing **survival data** and other types of **time-to-event data**

The **subjects under study** may be humans, animals, engines, subscriptions, etc.

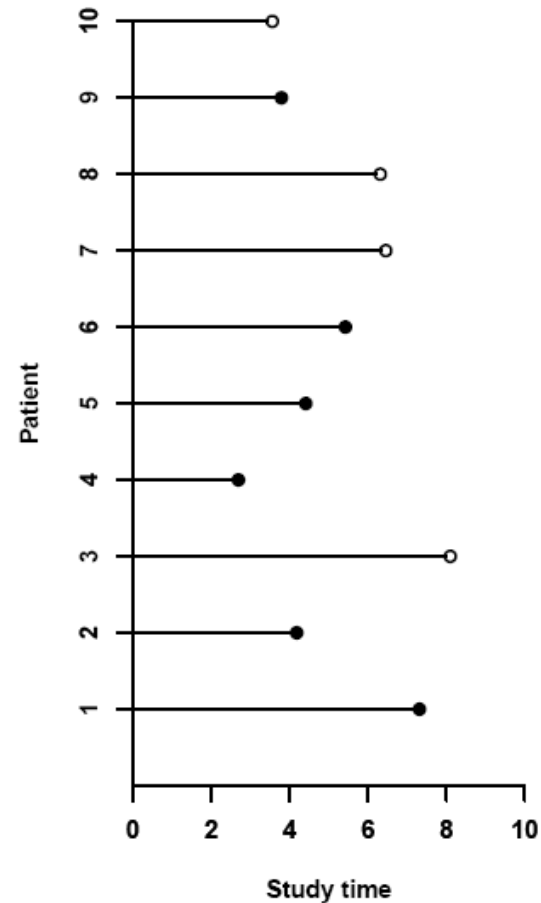
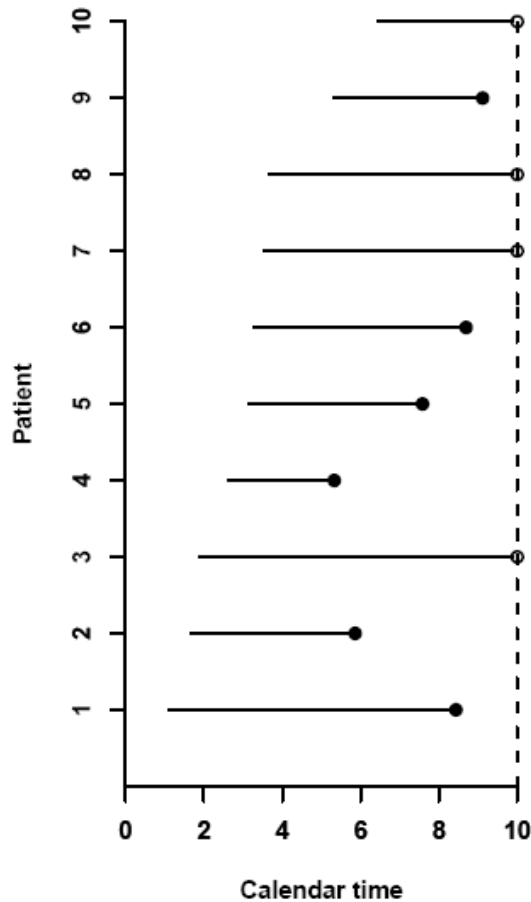
The **event of interest** may be death, cancer diagnosis, childbirth, engine failure, cancellation of a subscription (churn), etc.

The **aim of a study** may be to study the effect of a medical treatment, to establish risk factors for a disease, to assess the reliability of a technical system, etc.

Survival analysis: data

- A **survival time** is the time elapsed from an initial event to a well-defined end-point; e.g.
 - From birth to death (time=age)
 - From birth to breast cancer diagnosis (time=age)
 - From disease onset to death (time=disease duration)
 - From the start of a subscription until cancelation (time=duration of subscription)
- Note that we often use the term **survival time** also when the end-point (the event of interest) is something else than death
- A special feature of survival data is **right-censoring**: we may only know that a true survival time is *larger* than (e.g.) 5 years

One reason for censoring is termination of a study:



Censoring may also be due to other reasons, e.g., withdrawals, lost to follow-up, deaths from another cause than the one under study

It is crucial that censoring does *not* selectively remove subjects at particular high or low risk of experiencing the event. This is the **independent censoring** assumption to be discussed later.

Due to right censoring, traditional statistical method (like t-tests and linear regression) cannot be used to analyse survival data (we cannot even compute a mean).

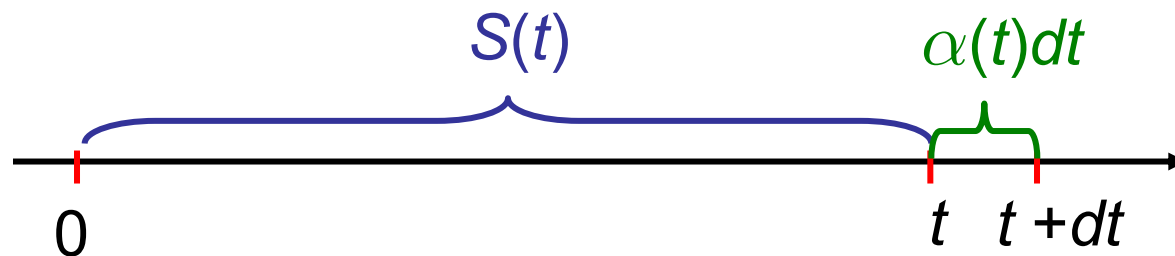
A further complication in some studies is that the subjects are not followed from time 0 (in the study time scale), but only from a later entry time. This is called **delayed entry** or **left-truncation**.

Survival analysis: concepts

In order to analyse survival data, we need the right concepts

The *survival function* $S(t) = P(T^* > t)$ is the probability that the survival time T^* exceeds t (in the study time scale)

The *hazard rate* $\alpha(t)$ is the instantaneous probability that event of interest happens per unit of time, i.e. $\alpha(t)dt$ is the probability that the event will happen between time t and time $t + dt$ given that it has not happened earlier



Formal definitions and relations:

$$S(t) = P(T^* > t)$$

$$\alpha(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} P(t \leq T^* < t + dt | T^* \geq t)$$

$$\alpha(t) = -\frac{S'(t)}{S(t)}$$

$$-\log\{S(t)\} = \int_0^t \alpha(s) ds$$

$$S(t) = \exp \left\{ - \int_0^t \alpha(s) ds \right\}$$

Stochastic processes

A time-continuous **stochastic process** $X = \{X(t)\}$ is a collection of random variables $X(t)$ indexed by time t

A realization of X is a function of t and is called a **sample path**

We will mainly consider stochastic processes that are **adapted**, which means that at time t we know the value of $X(s)$ for all $s \leq t$ (apart, possibly, from unknown parameters)

Formally, we have a *filtration* $\{\mathcal{F}_t\}$, i.e., an increasing family of sigma-algebras, describing how the information available to the researcher increases over time

Then X is adapted if $X(t)$ is $\mathcal{F}_t$ -measurable for all t

Counting processes and martingales

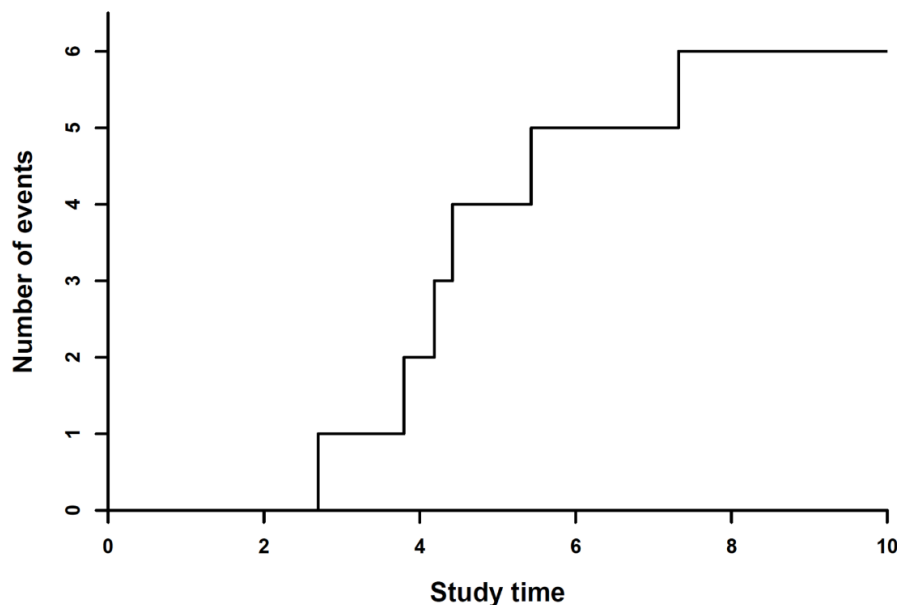
Counting processes and martingales play a key role in formulating models for survival data and in deriving and studying estimators and test statistics

A **counting process** N is a stochastic process whose sample paths are right-continuous step functions with steps of size +1

Typically $N(t)$ counts the number of events that have been observed in the time interval $[0, t]$, assuming no tied event times

Example data from slide 6 (with a * corresponding to censoring)

2.70, 3.50*, 3.80, 4.19, 4.42, 5.43, 6.32*, 6.46*, 7.32, 8.11*



We denote by $\mathcal{F}_t$ all information available to the researcher by time t (in the study time scale)

The **intensity process** λ of the counting process N is (informally) given by

$$\lambda(t)dt = P(dN(t) = 1 \mid \mathcal{F}_{t-})$$

where $dN(t)$ is the number of jumps of the process in $[t, t + dt)$, assumed to be 0 or 1, and $\mathcal{F}_{t-}$ is all information available to the researcher “just before” time t

Since $dN(t)$ is a binary variable, we have

$$\lambda(t)dt = E(dN(t) \mid \mathcal{F}_{t-})$$

It follows that

$$E(dN(t) - \lambda(t)dt \mid \mathcal{F}_{t-}) = 0$$

We now define

$$M(t) = N(t) - \int_0^t \lambda(s) ds$$

From the relation at the bottom of the previous slide, it follows that

$$E(dM(t) \mid \mathcal{F}_{t-}) = 0$$

This shows (informally) that the stochastic process M is a **martingale**

We have the decomposition

$$\underbrace{dN(t)}_{\text{observation}} = \underbrace{\lambda(t)dt}_{\text{signal}} + \underbrace{dM(t)}_{\text{noise}}$$

Statistical models for a counting process are specified by giving the form of the intensity process while the martingale increments represent random noise

Counting processes formulation of survival data

One uncensored survival time

T^* survival time with hazard $\alpha(t)$

Counting process $N^c(t) = I\{T^* \leq t\}$

Then

$$\begin{aligned} P(dN^c(t) = 1 \mid \mathcal{F}_{t-}) &= P(t \leq T^* < t + dt \mid \mathcal{F}_{t-}) \\ &= \begin{cases} \alpha(t)dt & \text{for } T^* \geq t \\ 0 & \text{for } T^* < t \end{cases} \end{aligned}$$

Intensity process:

$$\lambda^c(t) = \alpha(t)I\{T^* \geq t\}$$

Several uncensored survival times

$T_1^*, T_2^*, \dots, T_n^*$ independent survival times (no ties)

Hazard rate for T_i^* is $\alpha_i(t)$

Counting processes $N_i^c(t) = I\{T_i^* \leq t\} \quad i = 1, 2, \dots, n$

Intensity processes (due to independence):

$$\lambda_i^c(t) = \alpha_i(t)I\{T_i^* \geq t\}; \quad i = 1, \dots, n$$

Aggregated process $N^c(t) = \sum_{i=1}^n N_i^c(t)$ has intensity process

$$\lambda^c(t) = \sum_{i=1}^n \lambda_i^c(t) = \sum_{i=1}^n \alpha_i(t)I\{T_i^* \geq t\}$$

Various censoring schemes:

Type I censoring: Observe T_i^* if $T_i^* \leq c_i$, otherwise just observe that $T_i^* > c_i$ for a fixed censoring time c_i

Type II censoring: Observe the r smallest survival times, for the $n - r$ largest survival times we just know that they exceed $T_{(r)}^*$

Random censoring: Similar to Type I censoring, except that the c_i are observed values of random variables C_i that are independent of the survival times T_i^*

We will not assume any of these, but make the weakest possible assumption on the censoring that allows for valid inference. This is **the independent censoring** assumption

Several right-censored survival times

When we have right-censoring, we for each individual observe a (possibly) censored survival time $\tilde{T}_i$ together with an indicator D_i that takes the value 1 when $\tilde{T}_i = T_i^*$ and the value 0 when $\tilde{T}_i < T_i^*$

Then the **independent censoring** assumption is given by

$$P(t \leq \tilde{T}_i < t + dt, D_i = 1 \mid \tilde{T}_i \geq t, \mathcal{F}_{t-}) = P(t \leq T_i^* < t + dt \mid T_i^* \geq t)$$

In words: We have independent censoring if an individual who is at risk at time t has the same probability of experiencing the event in the small time interval $[t, t + dt)$ as would have been the case in the situation without censoring

Note: In the literature, the term «independent censoring» is sometimes used for what we denote «random censoring»

We introduce the counting processes

$$N_i(t) = I\{\tilde{T}_i \leq t, D_i = 1\} \quad i = 1, 2, \dots, n$$

Then, due to independent censoring, the intensity process λ_i of N_i is given by

$$\begin{aligned} \lambda_i(t)dt &= P(dN_i(t) = 1 \mid \mathcal{F}_{t-}) \\ &= P(t \leq \tilde{T}_i < t + dt, D_i = 1 \mid \mathcal{F}_{t-}) \\ &= \begin{cases} 0 & \text{if } \tilde{T}_i < t \\ \alpha_i(t)dt & \text{if } \tilde{T}_i \geq t \end{cases} \end{aligned}$$

Thus

$$\lambda_i(t) = \alpha_i(t)Y_i(t)$$

where $Y_i(t) = I\{\tilde{T}_i \geq t\}$ is an at risk indicator

Several left-truncated and right-censored survival times

Sometimes an individual i is not observed from time zero (in the study time-scale), but only from a later entry time V_i

We then consider the counting processes

$$N_i(t) = I\{V_i < \tilde{T}_i \leq t, D_i = 1\} \quad i = 1, 2, \dots, n$$

Assuming **independent left-truncation and right-censoring** the intensity processes are given by

$$\lambda_i(t) = \alpha_i(t) Y_i(t)$$

where the $Y_i(t) = I\{V_i < t \leq \tilde{T}_i\}$ are at risk indicators

The Nelson-Aalen estimator

We have right-censored and/or left-truncated survival data for a sample of n individuals from a population with hazard rate $\alpha(t)$

Let $N_i(t)$ count the observed number of occurrences (0 or 1) of the event of interest for individual i in the interval $[0, t]$

Provided censoring/truncation is independent, the corresponding intensity process takes the form:

$$\lambda_i(t) = \underbrace{Y_i(t)}_{\text{at risk indicator}} \underbrace{\alpha(t)}_{\text{hazard rate}}$$

The aggregated counting process

$$N(t) = \sum_{i=1}^n N_i(t)$$

has intensity process of the multiplicative form

$$\lambda(t) = \sum_{i=1}^n \lambda_i(t) = Y(t) \alpha(t)$$

where

$$Y(t) = \sum_{i=1}^n Y_i(t)$$

is the **number at risk** "just before" time t

We will estimate the cumulative hazard

$$A(t) = \int_0^t \alpha(u) du$$

We have the decomposition:

$$\begin{aligned} dN(t) &= \lambda(t)dt + dM(t) \\ &= Y(t)\alpha(t)dt + dM(t) \\ &= \underbrace{Y(t)dA(t)}_{\text{signal}} + \underbrace{dM(t)}_{\text{noise}} \end{aligned}$$

Estimating equation

$$dN(t) = Y(t) \cdot d\hat{A}(t)$$

Thus (when $Y(t) > 0$):

$$d\hat{A}(t) = \frac{dN(t)}{Y(t)}$$

The Nelson-Aalen estimator

$$\hat{A}(t) = \int_0^t \frac{dN(s)}{Y(s)} = \sum_{T_j \leq t} \frac{1}{Y(T_j)}$$

is a sum over the jump times $T_1 < T_2 < \dots$ of N

Note that the T_i given here are the $\tilde{T}_i$ with $D_i = 1$ given in increasing order

Example: Melanoma data

In the period 1962-77 a total of 205 patients with malignant melanoma (cancer of the skin) were operated at Odense University hospital in Denmark

A number of covariates were recorded at operation (we will only consider a few of them), and the patients were followed up until death or censoring at the end of the study 31 December 1977

57 patients died of malignant melanoma, 14 died from other causes, and 134 were still alive at the end of the study

Source: Andersen, Borgan, Gill & Keiding: *Statistical Models Based on Counting Processes*, Springer, 1993.

Inspecting the data

You may read the data into R by the commands:

```
mdata="https://www.mn.uio.no/math/english/people/aca/borgan/iwsm2026/melanoma.txt"
melanoma=read.table(mdata,header=T)
```

The data are organized with one line for each of the 205 patients,
and with the following variables in the seven columns:

status: status (1=death from disease, 2=censored, 4=death from other cause)
lifetime: lifetime from operation in years
ulcer: ulceration (1=present, 2=absent)
thickn: tumor thickness in mm
sex: gender (1=female, 2=male)
age: age at operation in years
grthick: grouped tumor thickness
1: 0-1 mm (i.e. below 2 mm)
2: 2-4 mm (i.e. at least 2 mm, but below 5 mm)
3: 5+ mm (i.e. 5 mm or more)]

To see the first five lines of the data file, you may give the command:

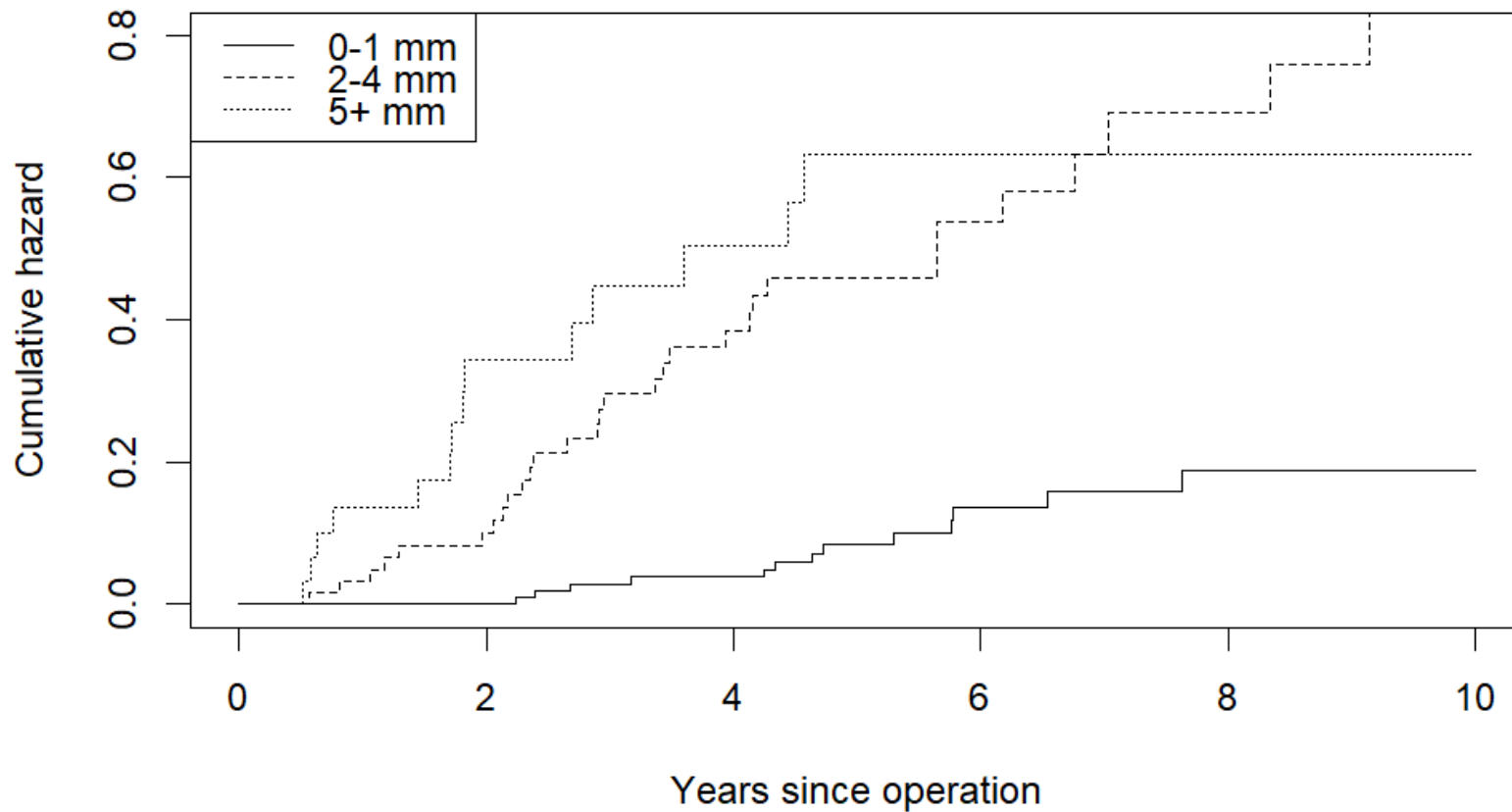
```
head(melanoma)
```

We will study deaths from malignant melanoma, considering deaths from other causes as censorings

Nelson-Aalen plots for gender:



Nelson-Aalen plots for thickness groups



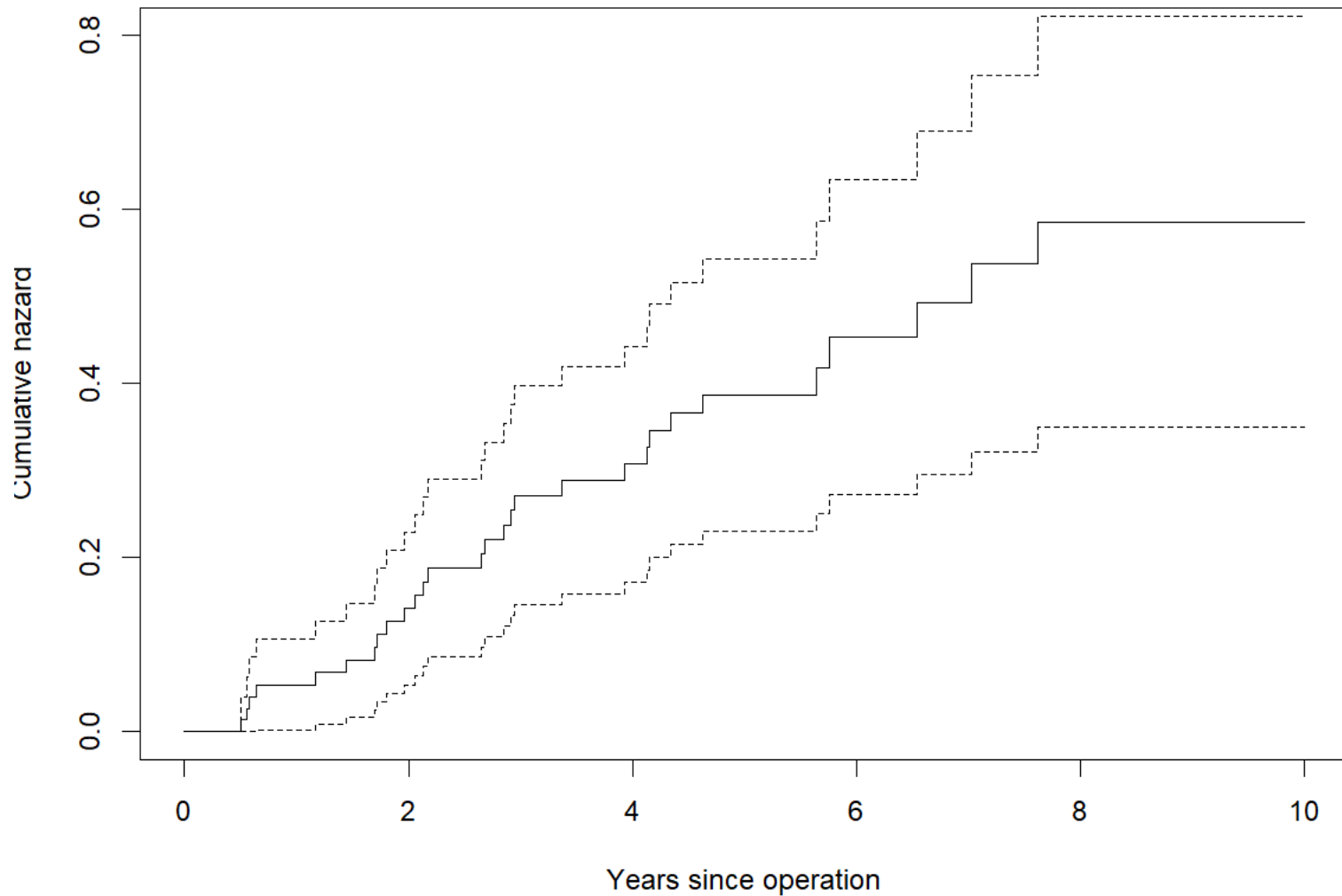
We will show (later) that the Nelson-Aalen $\hat{A}(t)$ estimator is (almost) unbiased and approximately normally distributed around $A(t)$ with a variance that may be estimated by

$$\begin{aligned}\hat{\sigma}^2(t) &= \int_0^t \frac{dN(s)}{Y(s)^2} \\ &= \sum_{T_j \leq t} \frac{1}{Y(T_j)^2}\end{aligned}$$

From this we may obtain approximate 95% pointwise confidence limits for $A(t)$ as

$$\hat{A}(t) \pm 1.96 \cdot \hat{\sigma}(t)$$

Nelson-Aalen plot for males with 95% pointwise confidence limits



R-commands that give plots on the previous slides

We use the survival library, so this must be loaded:

```
library(survival)
```

To plot the Nelson-Aalen estimates for both genders in the same figure,

we may give the commands:

```
surv.s=survfit(Surv(lifetime,status==1) ~ factor(sex), data=melanoma, stype=2)
plot(surv.s, fun="cumhaz", lty=1:2, xlim=c(0,10), ylim=c(0,0.80), main="Gender",
      xlab="Years since operation", ylab="Cumulative hazard")
legend("topleft", c("Females","Males"), lty=1:2)
```

Similar commands give Nelson-Aalen plots for the three thickness groups

Plot of Nelson-Aalen estimate for males with pointwise confidence limits

```
surv.m=survfit(Surv(lifetime,status==1) ~ 1, data=melanoma, subset=(sex==2), stype=2)
plot(surv.m, fun="cumhaz", mark.time=FALSE, xlim=c(0,10), ylim=c(0,0.80),
      main="Males", xlab="Years since operation", ylab="Cumulative hazard")
```

Martingales

To study the properties of the Nelson-Aalen estimator we will use some results for martingales and stochastic integrals

A stochastic process $M = \{M(t)\}$ is a **martingale** relative to the filtration $\{\mathcal{F}_t\}$ if it is adapted to the filtration and satisfies the martingale property:

$$E(M(t) \mid \mathcal{F}_s) = M(s) \text{ for all } t > s$$

Heuristically, M is a martingale provided that

$$E(dM(t) \mid \mathcal{F}_{t-}) = 0$$

where $dM(t)$ is the increment of M over $[t, t + dt)$, and $\mathcal{F}_{t-}$ is all information available to the researcher "just before" time t

We will assume throughout that $M(0) = 0$. Then

$$\begin{aligned} EM(t) &= E\{E(M(t)|\mathcal{F}_0)\} \\ &= E\{M(0)\} = 0 \end{aligned}$$

Thus M is a **mean zero martingale** (denoted just “martingale” in the following)

For $0 \leq s < t < u < v$ we have

$$\begin{aligned} &\text{Cov}\{(M(t) - M(s))(M(v) - M(u))\} \\ &= E\{(M(t) - M(s))(M(v) - M(u))\} \\ &= E\{E[(M(t) - M(s))(M(v) - M(u)) | \mathcal{F}_t]\} \\ &= E\{(M(t) - M(s))E[(M(v) - M(u)) | \mathcal{F}_t]\} \\ &= 0 \end{aligned}$$

Thus a martingale has **uncorrelated increments**

Variation process

The variation of a martingale M may be described by its **predictable variation process** $\langle M \rangle$. This is given informally by

$$d\langle M \rangle(t) = \text{Var}(dM(t) \mid \mathcal{F}_{t-})$$

Note that

$$d\langle M \rangle(t) = \mathbb{E}([dM(t)]^2 \mid \mathcal{F}_{t-}) = \mathbb{E}(dM^2(t) \mid \mathcal{F}_{t-})$$

Hence

$$\mathbb{E}\{d[M^2 - \langle M \rangle](t) \mid \mathcal{F}_{t-}\} = 0$$

Thus $M^2 - \langle M \rangle$ is a martingale

From this we obtain:

$$\text{Var}(M(t)) = \mathbb{E}(M(t)^2) = \mathbb{E}\langle M \rangle(t)$$

(We do not consider optional variation processes in this course)

Now consider a counting process martingale

$$M(t) = N(t) - \int_0^t \lambda(s) ds$$

To determine its predictable variation process, note that

$$\begin{aligned} d\langle M \rangle(t) &= \text{Var}(dM(t) \mid \mathcal{F}_{t-}) \\ &= \text{Var}(dN(t) - \lambda(t)dt \mid \mathcal{F}_{t-}) \\ &= \text{Var}(dN(t) \mid \mathcal{F}_{t-}) \\ &\approx \lambda(t)dt \{1 - \lambda(t)dt\} \\ &\approx \lambda(t)dt \end{aligned}$$

This motivates the important relation

$$\langle M \rangle(t) = \int_0^t \lambda(s) ds$$

Stochastic integrals

Let $M = \{M(t)\}$ be a martingale and $H = \{H(t)\}$ a **predictable process**, which intuitively means that for any time t we know the value of $H(t)$ "just before" time t

A sufficient condition for predictability is that H is adapted and left-continuous

We may then consider the **stochastic integral**

$$I(t) = \int_0^t H(s) dM(s)$$

In general it is quite intricate to define a stochastic integral (cf. the Itô integral)

But we avoid these problems since we only consider stochastic integrals relative to a counting process martingale

$$M(t) = N(t) - \int_0^t \lambda(s) ds$$

Then the stochastic integral may be evaluated as:

$$I(t) = \int_0^t H(s) dM(s) = \underbrace{\int_0^t H(s) dN(s)}_{= \sum_{T_j \leq t} H(T_j)} - \underbrace{\int_0^t H(s) \lambda(s) ds}_{\text{an ordinary integral (pathwise)}}$$

a sum over the
jump times
 $T_1 < T_2 < \dots$

Consider a stochastic integral

$$I(t) = \int_0^t H(s) dM(s)$$

Then

$$\begin{aligned} E\{dI(t) | \mathcal{F}_{t-}\} &= E\{H(t)dM(t) | \mathcal{F}_{t-}\} \\ &= H(t) \underbrace{E\{dM(t) | \mathcal{F}_{t-}\}}_{=0} \quad (\text{H is predictable}) \\ &= 0 \end{aligned}$$

Thus a stochastic integral is a martingale

To obtain the predictable variation process $\langle I \rangle$ we note that

$$\begin{aligned} d\langle I \rangle(t) &= \text{Var}(dI(t) | \mathcal{F}_{t-}) \\ &= \text{Var}\{H(t)dM(t) | \mathcal{F}_{t-}\} \\ &= H^2(t)\text{Var}\{dM(t) | \mathcal{F}_{t-}\} \\ &= H^2(t)d\langle M \rangle(t) \end{aligned}$$

When M is a counting process martingale we have

$$d\langle M \rangle(t) = \lambda(t)dt$$

which shows informally that

$$\langle I \rangle(t) = \int_0^t H^2(s)\lambda(s)ds$$

Statistical properties of Nelson-Aalen estimator

We will study the properties of the Nelson-Aalen estimator

$$\hat{A}(t) = \int_0^t \frac{dN(s)}{Y(s)}$$

for the cumulative hazard $A(t) = \int_0^t \alpha(s)ds$

To handle the possibility that $Y(t) = 0$, we introduce the indicator $J(t) = I\{Y(t) > 0\}$ and interpret $0/0$ as 0

Then we may write the Nelson-Aalen estimator as

$$\hat{A}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s)$$

We have the decomposition

$$dN(s) = \lambda(s)ds + dM(s) = Y(s)\alpha(s)ds + dM(s)$$

Hence we may write

$$\begin{aligned}\hat{A}(t) &= \int_0^t \frac{J(s)}{Y(s)} dN(s) \\ &= \int_0^t \frac{J(s)}{Y(s)} \{Y(s)\alpha(s)ds + dM(s)\} \\ &= \underbrace{\int_0^t J(s)\alpha(s)ds}_{= A^*(t)} + \underbrace{\int_0^t \frac{J(s)}{Y(s)} dM(s)}_{= I(t)}\end{aligned}$$

Thus we have the decomposition:

$$\hat{A}(t) = \underbrace{A^*(t)}_{\text{systematic part}} + \underbrace{I(t)}_{\text{random part}}$$

The random part is the **stochastic integral**:

$$I(t) = \int_0^t \frac{J(s)}{Y(s)} dM(s)$$

where J/Y is a predictable process (its value at time t is known "just before" t)

We may use properties of stochastic integrals to study the Nelson-Aalen estimator

The stochastic integral $I(t)$ is a martingale

In particular

$$E\{I(t)\} = 0$$

Hence

$$\begin{aligned} E\{\hat{A}(t)\} &= E\{A^*(t) + I(t)\} \\ &= E\{A^*(t)\} + 0 \\ &\approx A(t) \end{aligned}$$

So the Nelson-Aalen estimator is approximately unbiased

To obtain the variance of the Nelson-Aalen estimator and an estimator for its variance, we note that

$$\hat{A}(t) - A^*(t) = \int_0^t \frac{J(s)}{Y(s)} dM(s)$$

Thus $\hat{A} - A^*$ is a stochastic integral, and hence a martingale

Its predictable variation process is given by

$$\langle \hat{A} - A^* \rangle(t) = \int_0^t \left(\frac{J(s)}{Y(s)} \right)^2 \lambda(s) ds = \int_0^t \frac{J(s)}{Y(s)} \alpha(s) ds$$

$$\lambda(s) = Y(s)\alpha(s)$$

Hence

$$\text{Var}[\hat{A}(t) - A^*(t)] = E\langle \hat{A} - A^* \rangle(t) = \int_0^t E\left(\frac{J(s)}{Y(s)} \right) \alpha(s) ds$$

Now $A^*(t) \approx A(t)$ and it follows that

$$\text{Var}\hat{A}(t) \approx \int_0^t \mathbb{E} \left(\frac{J(s)}{Y(s)} \right) dA(s)$$

As an estimator for the variance, we may use

$$\hat{\sigma}^2(t) = \int_0^t \frac{J(s)}{Y(s)} d\hat{A}(s) = \int_0^t \frac{J(s)}{Y^2(s)} dN(s)$$

which may be written as the sum

$$\hat{\sigma}^2(t) = \sum_{T_j \leq t} \frac{1}{Y^2(T_j)}$$

over the jump times $T_1 < T_2 < \dots$ of N

We will indicate how one may show that the Nelson-Aalen estimator is approximately normally distributed when n is “large”

To this end we consider the distribution of

$$\sqrt{n}(\hat{A}(t) - A^*(t)) = \int_0^t \sqrt{n} \frac{J(s)}{Y(s)} dM(s)$$

as $n \rightarrow \infty$

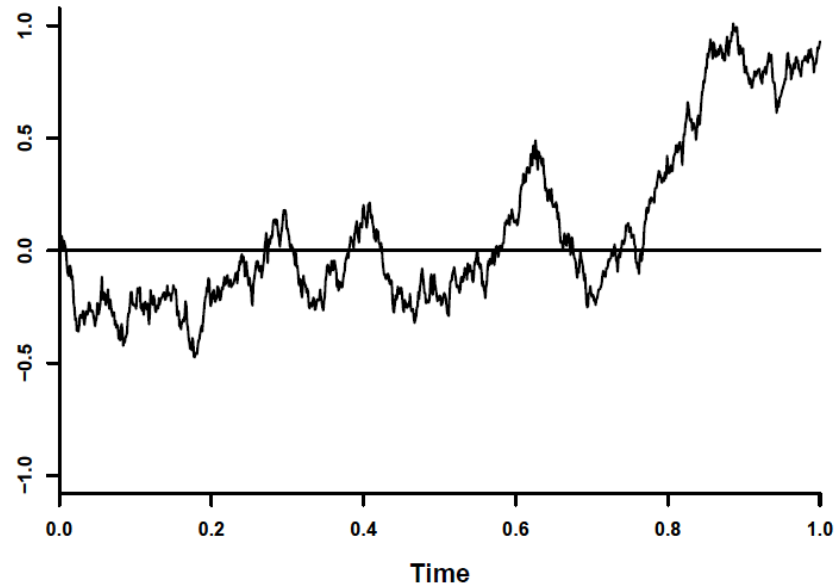
To see that this converges to a normal distributed we will use a martingale central limit theorem

But first we will have a brief look at Wiener processes and Gaussian martingales

Wiener process

A Wiener process (or Brownian motion) $W(t)$ is a stochastic process with the following properties

- $W(0) = 0$
- For $s < t$ the increment $W(t) - W(s)$ is normally distributed with mean 0 and variance $t - s$
- For $s < t < u < v$ the increments $W(t) - W(s)$ and $W(v) - W(u)$ are independent
- The sample paths are continuous



Gaussian martingales

Let $V(t)$ be a strictly increasing continuous function with $V(0) = 0$ and consider the time transformed Wiener process $U(t) = W(V(t))$

The process U inherits the properties of the Wiener process, so U is a stochastic process with the following properties

- $U(0) = 0$
- $U(t) - U(s)$ is normally distributed with mean 0 and variance $V(t) - V(s)$
- increments over disjoint intervals are independent
- the sample paths are continuous

One may show that U is a mean zero martingale with predictable variation $\langle U \rangle(t) = V(t)$

U is called a **Gaussian martingale**

Rebolledo's martingale central limit theorem

Under fairly general assumptions, a sequence of martingales $\{M^{(n)}\}$ will converge in distribution to a Gaussian martingale U as defined on the previous slide

What is needed, is that as $n \rightarrow \infty$

- (i) $\langle M^{(n)} \rangle(t) \rightarrow V(t)$ for all t
- (ii) The sizes of the jumps of $M^{(n)}$ go to zero

(convergence in probability)

We assume that

$$\frac{Y(s)}{n} \rightarrow y(s) > 0 \quad \text{as } n \rightarrow \infty \text{ for all } s$$

Then

$$\langle \sqrt{n}(\hat{A} - A^*) \rangle(t) = \int_0^t n \frac{J(s)}{Y(s)} \alpha(s) ds \rightarrow \int_0^t \frac{\alpha(s)}{y(s)} ds$$

which shows (i) on the previous slide

If an event occurs at time t , the size of the jump of $\sqrt{n}(\hat{A} - A^*)$ at that time has size

$$\frac{\sqrt{n}}{Y(s)} = \frac{1}{\sqrt{n}} \cdot \frac{n}{Y(s)} \rightarrow 0$$

which shows (ii) on the previous slide

Asymptotically the difference between A and A^* vanishes so it follows by Reboledos theorem that

$$\sqrt{n}(\hat{A} - A) \rightarrow U \quad (\text{in distribution})$$

where U is a mean zero Gaussian martingale with variance function

$$\sigma^2(t) = \int_0^t \frac{\alpha(s)}{y(s)} ds$$

In particular for a fixed value of t the Nelson-Aalen estimator $\hat{A}(t)$ is approximately normally distributed around $A(t)$

Survival functions, cumulative hazards, and product integrals: the general case

Uncensored survival time T^*

Survival function: $S(t) = P(T^* > t)$

For the **absolute continuous case**, the hazard is given by:

$$\alpha(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} P(t \leq T^* < t + dt | T^* \geq t)$$

Cumulative hazard: $A(t) = \int_0^t \alpha(u) du$

We have the relations:

$$\alpha(t) = A'(t) = -\frac{S'(t)}{S(t)} \qquad S(t) = \exp\{-A(t)\}$$

For a **general distribution** the hazard rate is not defined, but we may define the cumulative hazard rate as (generalizing the first relation at the bottom of the previous slide):

$$A(t) = -\int_0^t \frac{dS(u)}{S(u-)}$$

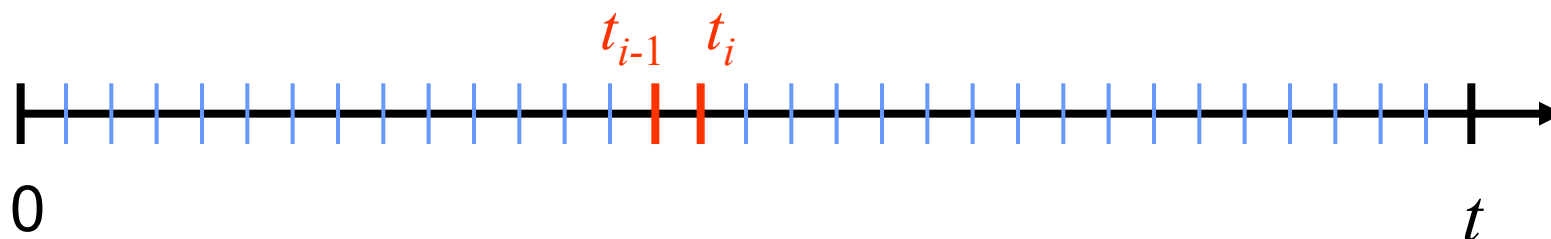
For the **discrete case** $A(t)$ is a step function with increments

$$\begin{aligned}\Delta A(u) &\stackrel{\text{def}}{=} A(u) - A(u-) = -\frac{\Delta S(u)}{S(u-)} \\ &= P(T^* = u | T^* \geq u)\end{aligned}$$

How can we generalize the second relation at the bottom of the previous slide?

Need product-integrals to achieve this generalization

Partition $[0, t]$ into small time intervals:



$$S(t) = \lim_{\max |t_i - t_{i-1}| \rightarrow 0} \prod (1 - \{A(t_i) - A(t_{i-1})\})$$

$$\stackrel{\text{def}}{=} \pi_{0 \leq u \leq t} (1 - dA(u))$$

The limit is a **product-integral**

For the **continuous case** we have:

$$\prod_{0 \leq u \leq t} (1 - dA(u)) = \exp \{-A(t)\}$$

For the **discrete case** we have:

$$\prod_{0 \leq u \leq t} (1 - dA(u)) = \prod_{u \leq t} (1 - \Delta A(u))$$

where $\Delta A(u) = P(T^* = u | T^* \geq u)$ is the increment of the cumulative hazard at time u

For the general case we have a mixture of the two

The Kaplan-Meier estimator

For right censored survival data we observe:

$$\tilde{T}_i = \min\{\text{survival time } T_i^*, \text{ censoring time } C_i\}$$

$$D_i = I\{\tilde{T}_i = T_i^*\}$$

Model: the uncensored survival times T_i^* are independent and identically distributed with hazard rate $\alpha(t)$

Counting and intensity processes:

$$N_i(t) = I\{\tilde{T}_i \leq t, D_i = 1\}$$

$$\lambda_i(t) = I\{\tilde{T}_i \geq t\} \alpha(t) = Y_i(t) \alpha(t)$$

The aggregated counting process

$$N(t) = \sum_{i=1}^n N_i(t)$$

has intensity process

$$\lambda(t) = \sum_{i=1}^n \lambda_i(t) = Y(t)\alpha(t)$$

where

$$Y(t) = \sum_{i=1}^n I\{\tilde{T}_i \geq t\}$$

is the number at risk “just before” time t

Nelson-Aalen estimator:

$$\hat{A}(t) = \int_0^t \frac{dN(u)}{Y(u)} = \sum_{u \leq t} \frac{\Delta N(u)}{Y(u)} = \sum_{T_j \leq t} \frac{1}{Y(T_j)}$$

Plug this into the product-integral expression for the survival function (Nelson-Aalen is a step-function):

$$\begin{aligned} \hat{S}(t) &= \prod_{0 \leq u \leq t} \left(1 - d\hat{A}(u)\right) = \prod_{u \leq t} \left(1 - \Delta \hat{A}(u)\right) \\ &= \prod_{u \leq t} \left(1 - \frac{\Delta N(u)}{Y(u)}\right) = \prod_{T_j \leq t} \left(1 - \frac{1}{Y(T_j)}\right) \end{aligned}$$

This is the **Kaplan-Meier estimator**

An alternative estimator of the survival function is

$$\begin{aligned}\tilde{S}(t) &= \exp\{-\hat{A}(t)\} \\ &= \exp\left\{-\sum_{T_j \leq t} \frac{1}{Y(T_j)}\right\} \\ &= \prod_{T_j \leq t} \exp\left\{-\frac{1}{Y(T_j)}\right\}\end{aligned}$$

For practical purposes there is little difference between the two estimators

But from a theoretical point of view, the Kaplan-Meier estimator is the natural one

Kaplan-Meier estimator: Properties

$$A^*(t) = \int_0^t J(s) \alpha(s) ds \approx A(t)$$

$$S^*(t) = \prod_{0 \leq s \leq t} (1 - dA^*(s)) = \exp\{-A^*(t)\} \approx S(t)$$

May show that (this is Duhamel's equation)

$$\frac{\hat{S}(t)}{S^*(t)} - 1 = - \int_0^t \underbrace{\frac{\hat{S}(s-)}{S^*(s)}}_{\approx 1} d(\hat{A} - A^*)(s) \approx -(\hat{A}(t) - A^*(t))$$

Asymptotically: $\frac{\hat{S}(t)}{S(t)} - 1 \approx -(\hat{A}(t) - A(t))$

Thus:

$$\hat{S}(t) - S(t) \approx -S(t) \cdot (\hat{A}(t) - A(t))$$

The statistical properties for Kaplan-Meier may be derived from those of Nelson-Aalen:

- $\text{Var}\{\hat{S}(t)\} \approx \{S(t)\}^2 \cdot \text{Var}\{\hat{A}(t)\}$
- Variance estimator: $\hat{\tau}^2(t) = \{\hat{S}(t)\}^2 \cdot \hat{\sigma}^2(t)$
with $\hat{\sigma}^2(t) = \int_0^t \{Y(s)\}^{-2} dN(s)$
- $\hat{S}(t)$ is as normally distributed around $S(t)$

Usually the variance is estimated by *Greenwood's formula*:

$$\tilde{\tau}^2(t) = [\hat{S}(t)]^2 \cdot \tilde{\sigma}^2(t)$$

where

$$\tilde{\sigma}^2(t) = \int_0^t [Y(s)\{Y(s) - \Delta N(s)\}]^{-1} dN(s)$$

Only minor difference between the two variance estimators

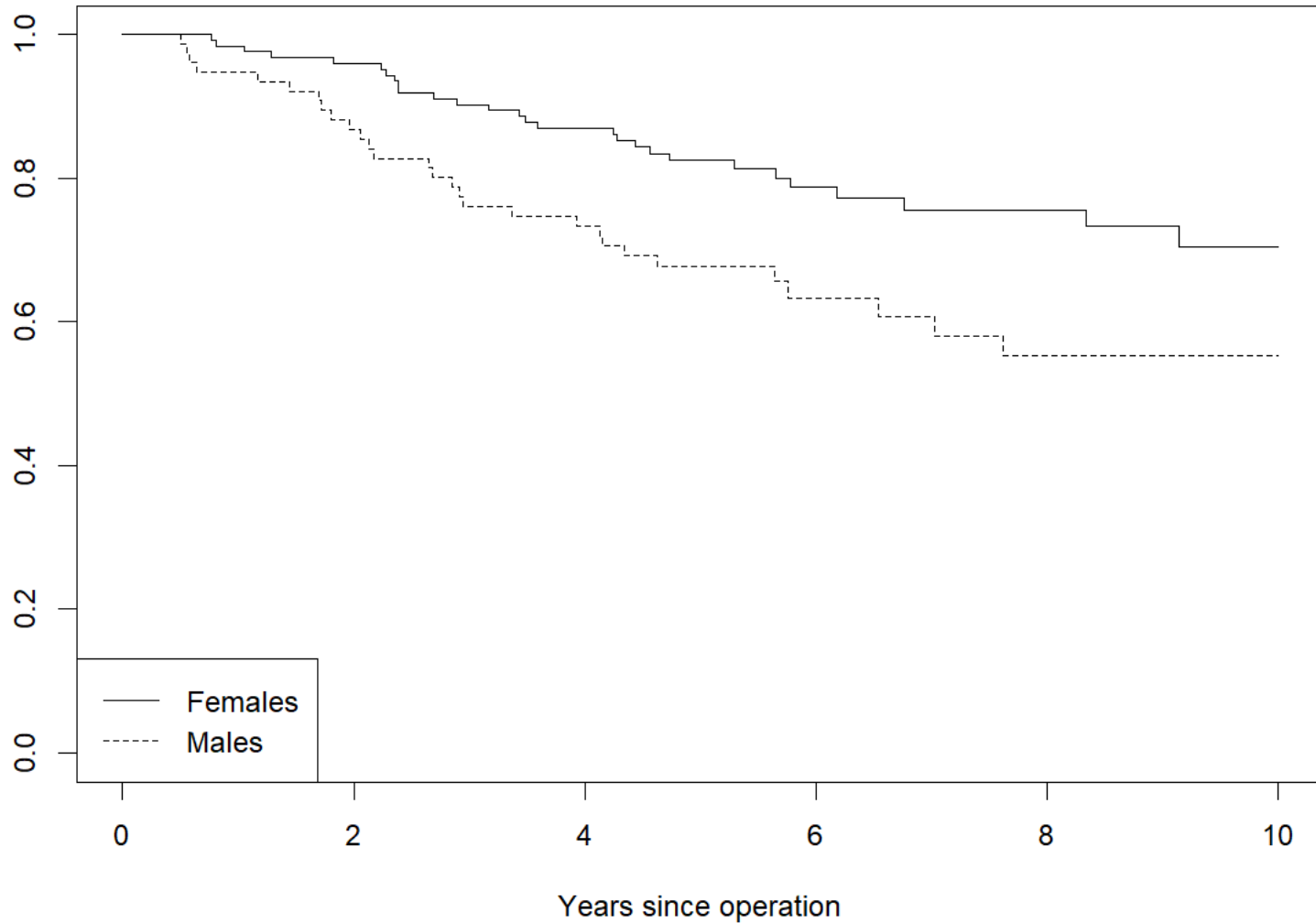
The standard pointwise 95% confidence limits for $S(t)$ takes the form

$$\hat{S}(t) \pm 1.96 \cdot \hat{S}(t) \cdot \hat{\sigma}(t)$$

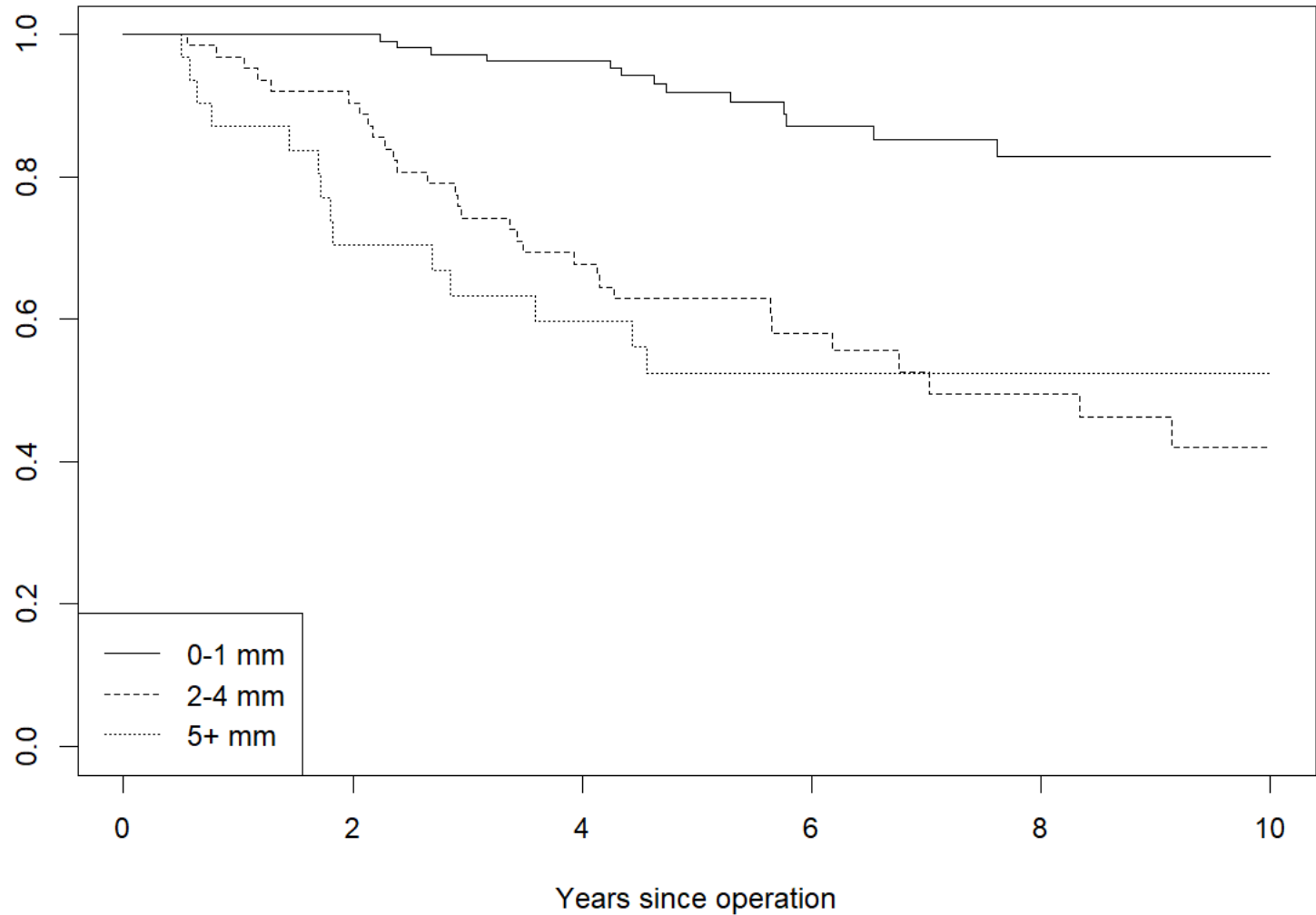
Another option is to use “log-minus-log transformed” confidence limits

Melanoma data

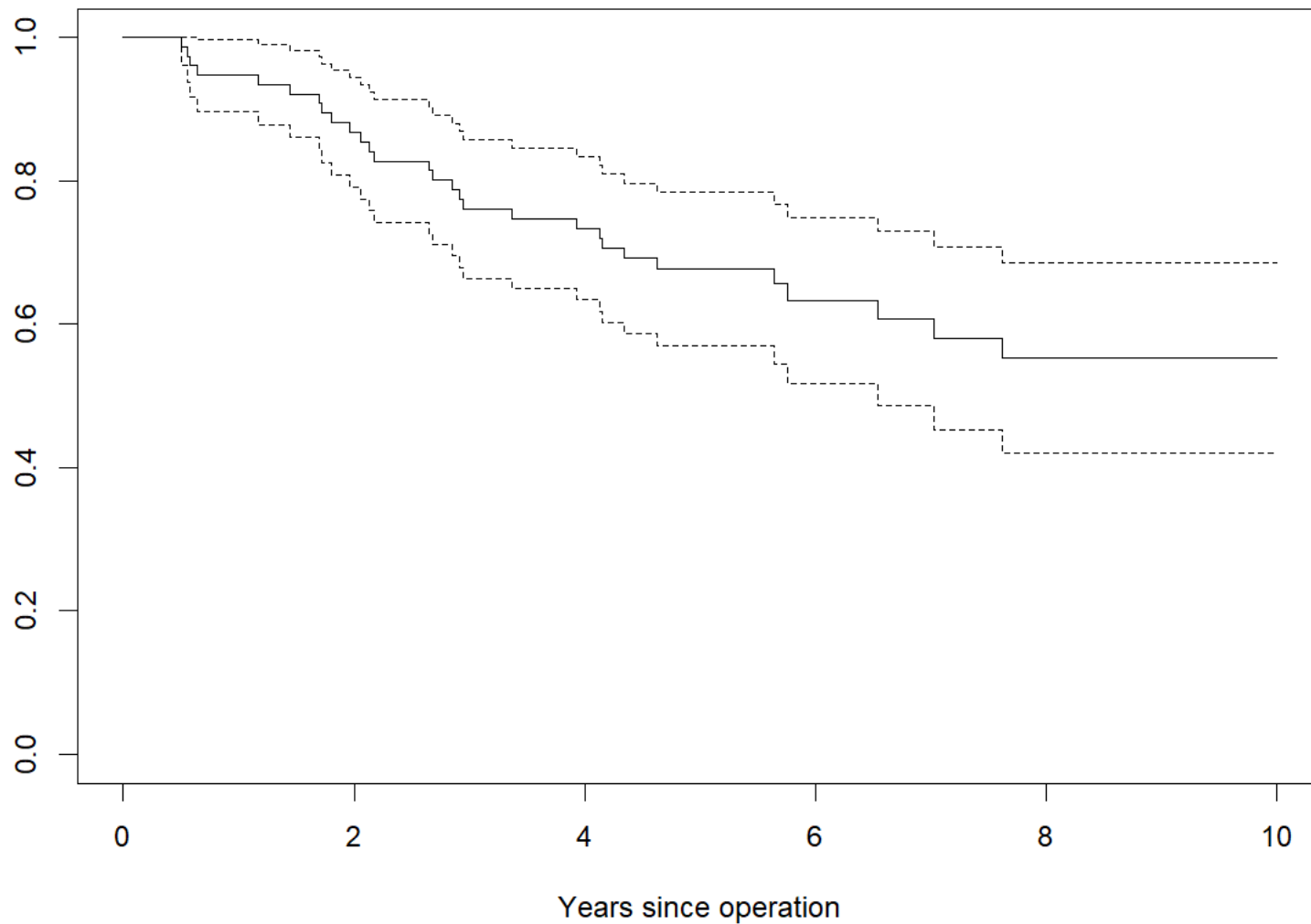
Kaplan-Meier plots for gender



Kaplan-Meier plots plot for thickness groups



Kaplan-Meier plot for males with 95% pointwise standard confidence limits



R-commands that give plots on the previous slides

Read data and load survival library

```
mdata="https://www.mn.uio.no/math/english/people/aca/borgan/iwsm2026/melanoma.txt"
melanoma=read.table(mdata,header=T)
library(survival)
```

To plot the Kaplan-Meier estimates for both genders in the same figure,
we may give the commands:

```
surv.s=survfit(Surv(lifetime,status==1) ~ sex, data=melanoma)
plot(surv.s, lty=1:2, xlim=c(0,10), main="Gender", xlab="Years since operation")
legend("bottomleft", c("Females","Males"), lty=1:2)
```

Similar commands give Kaplan-Meier plots for the three thickness groups

Kaplan-Meier plot for males with (standard) pointwise confidence limits

```
surv.m=survfit(Surv(lifetime,status==1) ~ 1, data=melanoma,
               subset=(sex==2), conf.type= "plain")
plot(surv.m, xlim=c(0,10), main="Males", xlab="Years since operation")
```

The default in R is to use a log-transformed confidence interval, but that is not
a good idea. To obtain confidence intervals for the survival function, you should
use `conf.type="plain"` or `conf.type="log-log"`

Estimation of median survival time and other fractiles of the survival distribution

The p -th fractile ξ_p of the survival distribution is given by

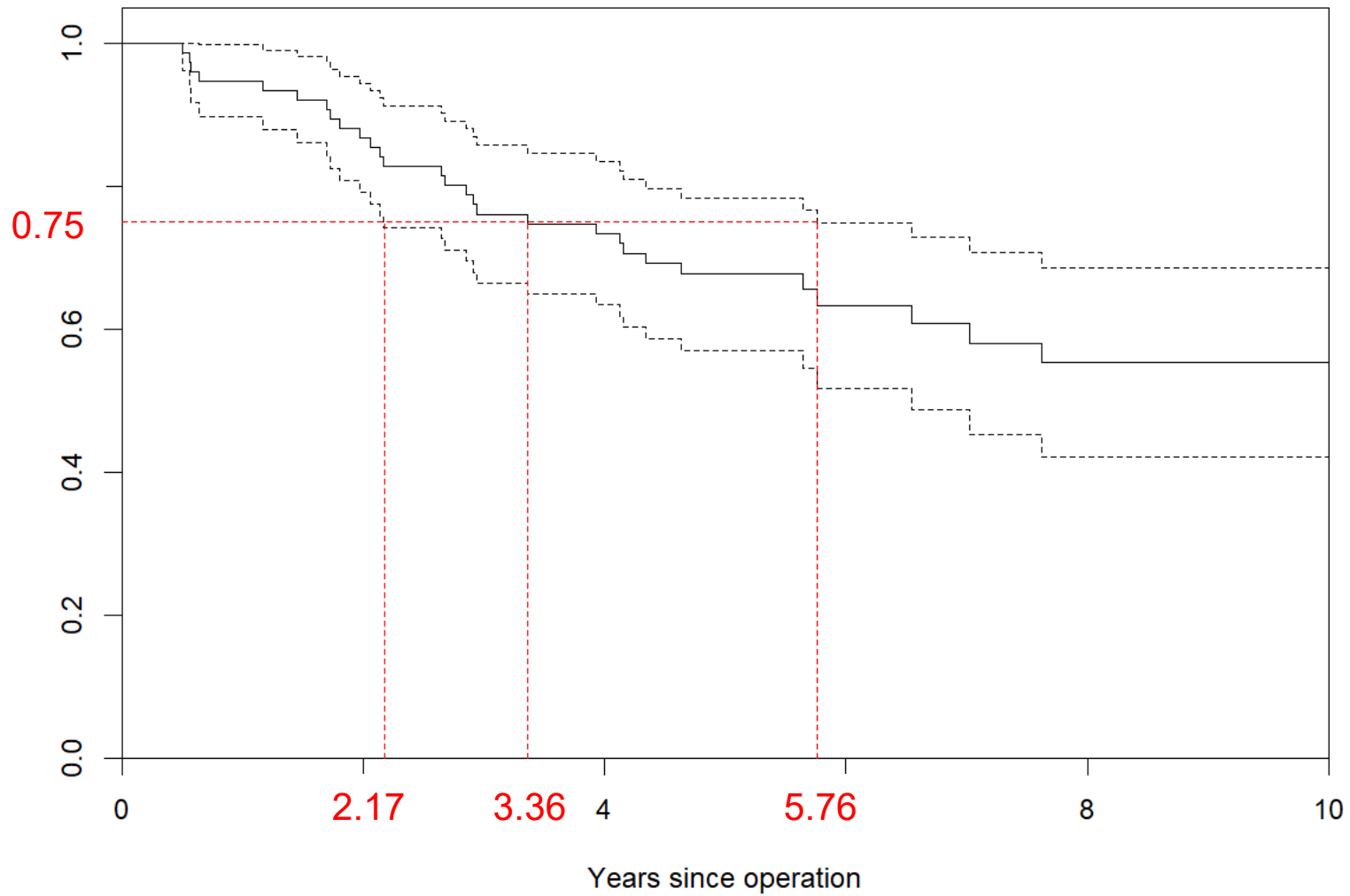
$$F(\xi_p) = p \quad \text{or equivalently} \quad S(\xi_p) = 1 - p$$

It is estimated by

$$\hat{\xi}_p = \inf \left\{ t : \hat{S}(t) \leq 1 - p \right\}$$

Confidence intervals may be found by "inverting" the confidence intervals for the survival function

Lower quartile for males



R-commands for quantiles

You may find the quantiles from by inspecting the output of the summary command
summary(surv.m)

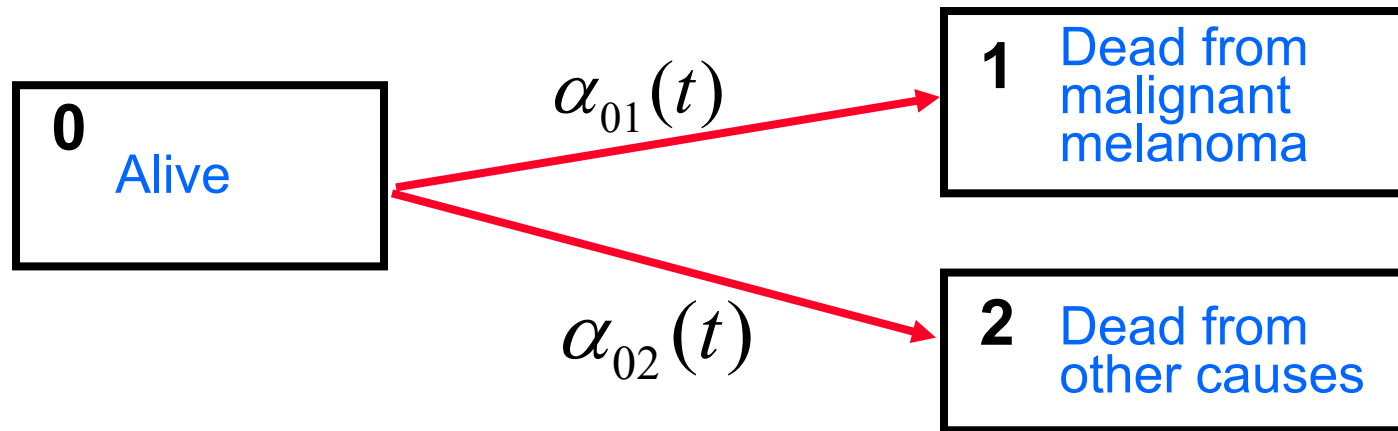
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0.507	76	1	0.987	0.0131	0.961	1.000
0.559	75	1	0.974	0.0184	0.938	1.000
.....						
2.060	65	1	0.854	0.0407	0.774	0.934
2.134	64	1	0.841	0.0422	0.758	0.923
2.173	63	1	0.827	0.0435	0.742	0.913
.....						
2.910	59	1	0.774	0.0482	0.680	0.869
2.945	58	1	0.761	0.0492	0.664	0.857
3.364	57	1	0.747	0.0501	0.649	0.846
3.932	55	1	0.734	0.0510	0.634	0.834
.....						
4.630	47	1	0.677	0.0544	0.570	0.784
5.647	32	1	0.656	0.0567	0.545	0.767
5.762	29	1	0.633	0.0591	0.517	0.749
6.542	25	1	0.608	0.0619	0.487	0.729
.....						

Or you may use the quantile command
quantile(surv.m)

A note on competing risks

For the melanoma data we have treated death from other causes as censorings

It could have been better to use a model for competing risks:



For this model one may estimate the **cumulative incidence function**, which is the probability that an individual has moved from state 0 to state 1 by time t

Kaplan-Meier estimates this probability if we set $\alpha_{02}(t) = 0$