Martingales and properties Nelson-Aalen/Kaplan-Meier

STK4080 September 8 H16

1. Review counting processes, martingales, Nelson-Aalen
2. Kaplan-Meier estimator
3. Gaussian martingale and martingale CLT
4. Application of martingale CLT to Nelson-Aalen and Kaplan-Meier estimators
5. Estimation of expected value $\mu = \mathrm{E}(T)$

Review

- Counting process $N(t)$ and History $\mathcal{F}_t$
- Cumulative intensity process $\Lambda(t) = \int_0^t \lambda(s)ds$:
  $\mathrm{P}(dN(t) = 1|\mathcal{F}_{t-}) = \lambda(t)dt$
- Then $M(t) = N(t) - \Lambda(t)$ is a martingale with resp. to $\mathcal{F}_t$

In particular, with right censored survival data

- Same hazard $\alpha(t)$ for independent individuals
- "Independent censoring": Type I, Random, Type II etc.
- History $\mathcal{F}_t = \{N_i(s), Y_i(s); 0 \leq s \leq t, i = 1, \ldots, n\}$
- Then the cum. int. process for $N(t) = \sum_{i=1}^n N_i(t)$
  becomes $\Lambda(t) = \int_0^t Y(s)\alpha(s)ds$ where $Y(s) = \sum_{i=1}^n Y_i(s)$
- $M(t) = N(t) - \Lambda(t)$ is a martingale with resp. to $\mathcal{F}_t$

Three key results

Assume that M(t) is a (general) martingale with respect to a history $\mathcal{F}_t$ and with $M(0) = 0$. Then $\mathrm{E}M(t) = 0$.

When $N(t)$ has cumulative intensity process $\Lambda(t)$ with respect to (wrt) history $\mathcal{F}_t$ and that $K(t)$ is predictable wrt $\mathcal{F}_t$. Then $M(t) = N(t) - \Lambda(t)$ and $\int_0^t K(s)dM(s)$ are martingales with

- $\mathrm{Var}(M(t)) = \mathrm{E}[\Lambda(t)]$
- $\mathrm{Var}(\int_0^t K(s)dM(s)) = \mathrm{E}[\int_0^t K(s)^2 d\Lambda(s)]$

We used these results to obtain expectation, variance and variance estimators for the Nelson-Aalen estimator of the cumulative hazard $A(t) = \int_0^t \alpha(s)ds$.

We will obtain similar results for the Kaplan-Meier (today) and several other estimators and test statistics (later).

A difference between slides and ABG book

We proved the two variance results on the previous slide in the lecture on Thursday 1st of September.

The text book by Aalen, Borgan & Gjessing use the concept of *predictable variation processes* to arrive at similar results.

We will, however, need the predictable variation process for getting to the asymptotic normality results. Thus we define the concept on the following slides.

Predictable variation process of $M(t) = N(t) - \Lambda(t)$

We argued that

$$\begin{aligned} \text{Var}[(dM(t)|\mathcal{F}_{t-}] &= \text{Var}[(dN(t) - \lambda(t)dt|\mathcal{F}_{t-}] = \text{Var}[(dN(t)|\mathcal{F}_{t-}] \\ &= \lambda(t)dt(1 - \lambda(t)dt) = \lambda(t)dt \end{aligned}$$

We will write $d\langle M\rangle(t) = \lambda(t)dt$ and

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

which is referred to as the *predictable variation process* of $M(t)$. In particular $M^2(t) - \langle M\rangle(t) = M^2(t) - \Lambda(t)$ can be shown to be a martingale, which gives

$$\text{Var}(M(t)) = \text{E}\Lambda(t)$$

The predictable variation process of $\int_0^t K(s)dM(s)$

where $K(t)$ is predictable is then defined as

$$\langle \int_0^t K(s)dM(s) \rangle = \int_0^t \text{Var}[K(s)dM(s)|\mathcal{F}_{s-}]$$

which gives us

$$\langle \int_0^t K(s)dM(s) \rangle = \int_0^t K(s)^2 \text{Var}[dM(s)|\mathcal{F}_{s-}] = \int_0^t K(s)^2 \langle dM(s) \rangle$$

or $\langle \int_0^t K(s)dM(s) \rangle = \int_0^t K(s)^2 \lambda(t)dt = \int_0^t K(s)^2 d\Lambda(t)$.

Again we get a martingale by

$$(\int_0^t K(s)dM(s))^2 - \langle \int_0^t K(s)dM(s) \rangle$$

Nelson-Aalen estimator for $A(t) = \int_0^t \alpha(s)ds$:

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

Let $A^\star(t) = \int_0^t J(s)\alpha(s)ds$ where $J(s) = I(Y(s) > 0)$. Then

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

where $dM(s) = dN(s) - Y(s)\alpha(s)ds$ is the martingale increment and $K(s) = J(s)/Y(s)$ a predictable process, thus itself a martingale.

The first result (on previous page) gives us

$$\mathrm{E}\hat{A}(t) = \mathrm{E}A^\star(t)$$

and thus Nelson-Aalen is practically unbiased.

N-Aa estimator, variance and variance estimator

The second result gives

$$\text{Var}(M(t)) = \text{E}[\int_0^t Y(s)\alpha(s)ds]$$

The third result then leads to

$$\begin{aligned} \text{Var}(\hat{A}(t)) \approx & \quad \text{Var}(\hat{A}(t) - A^\star(t)) = \text{E}[\textstyle\int_0^t \left(\frac{J(s)}{Y(s)}\right)^2 Y(s)\alpha(s)ds] \\ = & \quad \text{E}[\textstyle\int_0^t \frac{J(s)}{Y(s)}\alpha(s)ds] \end{aligned}$$

from which we suggest the variance estimator

$$\hat{\sigma}^2(t) = \widehat{\text{Var}}(\hat{A}(t)) = \int_0^t \frac{dN(s)}{Y(s)^2} = \sum_{\tilde{T}_i \le t} \frac{D_i}{Y(\tilde{T}_i)^2} = \sum_{\tilde{T}_i \le t, D_i = 1} \frac{1}{Y(\tilde{T}_i)^2}$$

Nelson-Aalen approx. normal

In addition we will argue by the martingale central limit theorem that approximately

$$\hat{A}(t) \sim \mathrm{N}(A(t), \hat{\sigma}^2(t))$$

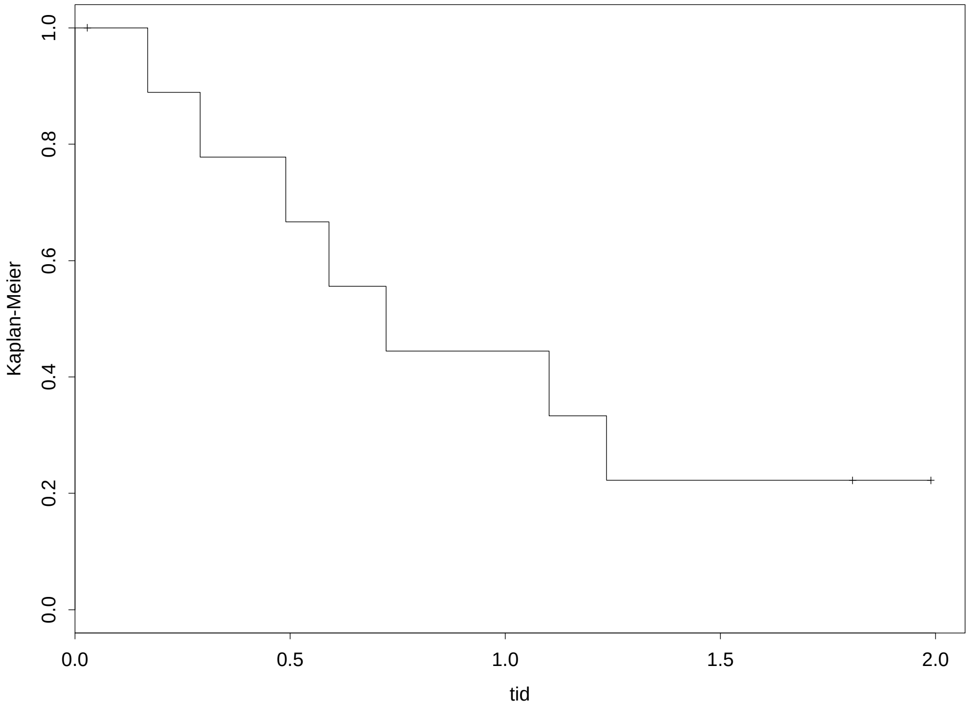
so $\hat{A}(t) \pm 1.96\hat{\sigma}(t)$ is an approx. 95% CI for $A(t)$.

Kaplan-Meier in counting process notation

Product integral: $\prod [1 - K(s) dN(s)] = \prod_i [1 - K(t_i)]$ when $dN(t_i) = 1$ for a finite no. of points t_i and 0 otherwise.

With this notation we may write the Kaplan-Meier estimator by

$$\hat{S}(t) = \prod_{t_i \le t} [1 - \frac{D_i}{Y(\tilde{T}_i)}] = \prod_{s \le t} [1 - \frac{dN(s)}{Y(s)}]$$



Martingal for Kaplan-Meier

It can be shown (exercise), that with $S^\star(t) = \exp(-\int_0^t J(s)\alpha(s)ds)$, we have

$$\frac{\hat{S}(t)}{S^\star(t)} - 1 = -\int_0^t \frac{\hat{S}(s-)}{S^\star(s)} \frac{J(s)}{Y(s)} dM(s)$$

thus as a integral of a predictable function with respect to the martingale $M(t) = N(t) - \Lambda(t)$, thus also a martingale.

Thus, with independent lifetimes with hazard $\alpha(s)$ and ind. censoring,

$$\mathrm{E}\left[\frac{\hat{S}(t)}{S^\star(t)}\right] - 1 = 0$$

and the Kaplan-Meier Estimator is close to unbiased for the survival function $S(t)$.

Variance for the Kaplan-Meier estimator

The representation

$$\frac{\hat{S}(t)}{S^\star(t)} - 1 = -\int_0^t \frac{\hat{S}(s-)}{S^\star(s)} \frac{J(s)}{Y(s)} dM(s) = \int_0^t K(s) dM(s)$$

allows for considering the variance of the Kaplan-Meier using property 3 on slide 3: $\mathrm{Var}(\int_0^t K(s)dM(s)) = \mathrm{E}[\int_0^t K(s)^2 d\Lambda(s)]$

This gives

$$\begin{aligned} \mathrm{Var}(\frac{\hat{S}(t)}{S^\star(t)} - 1) &= \mathrm{E}[\int_0^t K(s)^2 Y(s)\alpha(s)ds] \\ &= \mathrm{E}[\int_0^t \left(\frac{\hat{S}(s-)}{S^\star(s)}\right)^2 \frac{J(s)}{Y(s)}\alpha(s)ds] \end{aligned}$$

Kaplan-Meier variance contd.

But $S^\star(t) \to S(t)$, thus

$$\mathrm{Var}(\hat{S}(t)) \approx S(t)^2 \mathrm{Var}(\frac{\hat{S}(t)}{S^\star(t)}) \approx S(t)^2 \mathrm{E}[\int_0^t \frac{J(s)}{Y(s)} \alpha(s) ds] = S(t)^2 \sigma^2(t)$$

(have used $\hat{S}(s-)/S^\star(s) \approx 1$) where $\sigma^2(t)$ is the same as for the Nelson-Aalen estimator.

An estimator for $\mathrm{Var}(\hat{S}(t))$ is given by the Greenwood's formula (1927)

$$\tilde{\tau}^2(t) = \widehat{\mathrm{Var}}(\hat{S}(t)) = \hat{S}(t)^2 \int_0^t \frac{dN(s)}{Y(s)(Y(s)-1)}$$

where we use a small sample correction $Y(s) - 1$.

Random Walk

Let $X_0 = 0$ and $X_n = X_{n-1} + Y_n$ where Y_n is independent of $X_j, j < n$ and

$$Y_n = \begin{cases} 1 \text{ with prob. } \frac{1}{2} \\ -1 \text{ with prob. } \frac{1}{2} \end{cases}$$

Then $\mathrm{E}(X_n) = 0$ and $\mathrm{Var}(X_n) = n$ and, with $[nt]$ the integer value of nt,

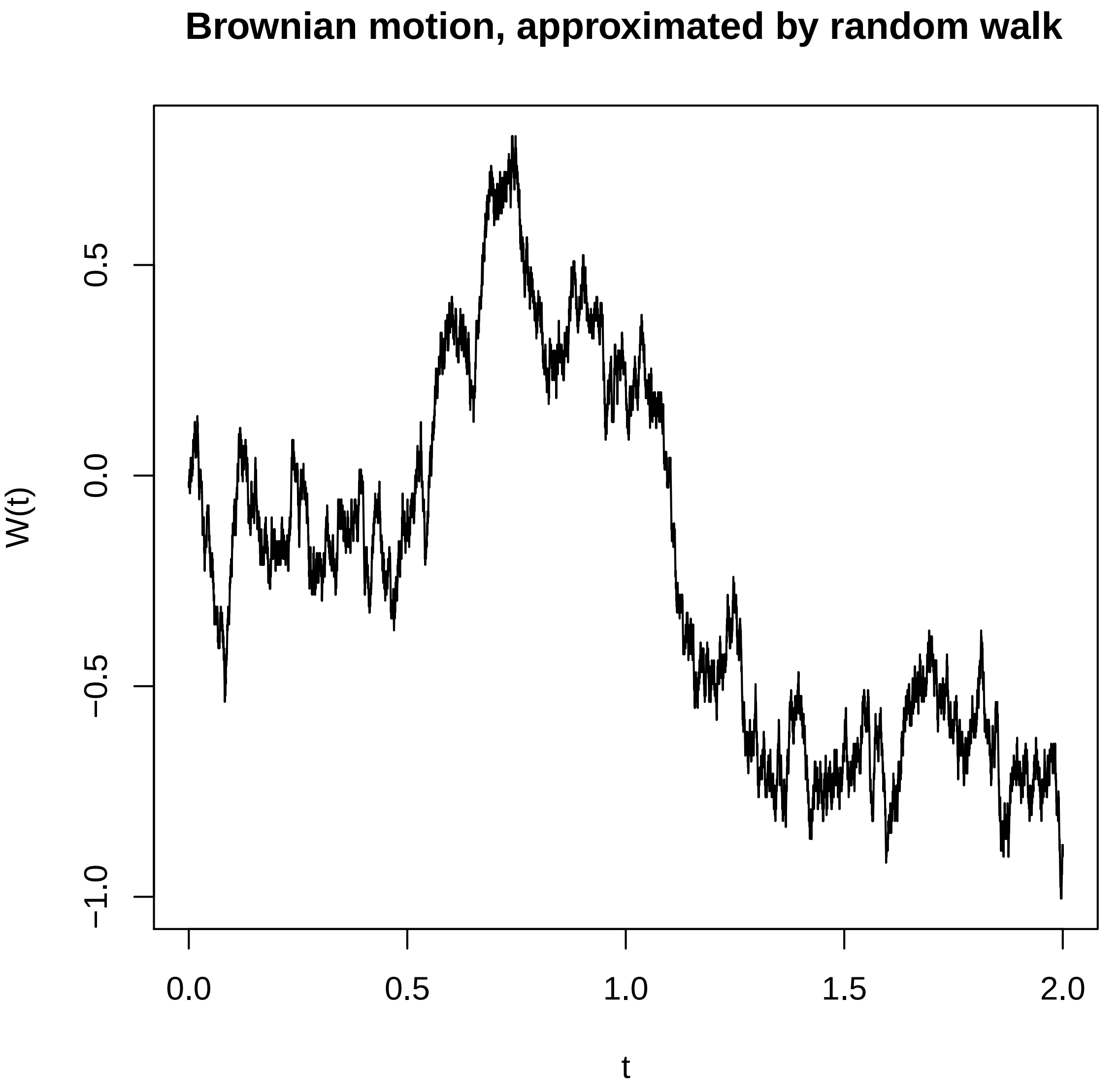
$$\frac{1}{\sqrt{n}} X_{[nt]} \to W(t)$$

where $W(t)$ is Wiener process (Brownian motion) with $\mathrm{E}[W(t)] = 0$, $\mathrm{Var}[W(t)] = t$ and $W(t) \sim \mathrm{N}(0, t)$.

Furthermore $W(t)$ has uncorrelated increments, $v < u < s < t$, $\mathrm{Cov}(W(t) - W(s), W(u) - W(v)) = 0$.

In fact $W(t)$ is a martingale!

Simulated Random Walk



Gaussian martingale

Let $W(t)$ be a Wiener process and $V(t)$ a strictly increasing continuous function with $V(0) = 0$ and let

$$U(t) = W(V(t))$$

Then $U(t)$

- has $\mathrm{E}[U(t)] = 0$
- and $\mathrm{Var}[U(t)] = V(t)$,
- $U(t) \sim \mathrm{N}(0, V(t))$ for a particular argument t
- has independent increments
- is a martingale

and we say $U(t)$ is a gaussian martingale.

Central limit theorem (CLT), non-iid version

Let X_i be random variables with $\mathrm{E}X_i = \mu_i$, $\mathrm{Var}(X_i) = \sigma_i^2$ and X_i-s only "weakly" dependent.

Let $\frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$, $\frac{1}{n}\sum_{i=1}^n \mu_i = \bar{\mu}_n \to \mu$ and $\frac{1}{n}\sum_{i=1}^n \sigma_i^2 \to \sigma^2$. Then

$$\sqrt{n}(\bar{X} - \bar{\mu}_n) \to N(0, \sigma^2)$$

if

- $\max(\frac{\sigma_i^2}{\sum_{i=1}^n \sigma_i^2}) \to 0$
- + some more regularity conditions

Interpret as if all X_i will have little individual effect on $\bar{X}$ for large n.

Martingal CLT

Let $M_n(t)$ be a series of martingales with expectation 0 and predictable variation process $\langle M_n(t)\rangle$.

Then, with $U(t)$ a gaussian martingale with expectation zero and variance function $\sigma^2(t)$,

- $M_n(t) \to U(t)$ as a process
- for a particular argument t this gives $M_n(t) \to N(0, \sigma^2(t))$

if

- $\langle M_n(t)\rangle \to \sigma^2(t)$ for the deterministic function $\sigma^2(t)$
- "Increments" $dM_n(t) \to 0$, become arbitrarily small (uniformly over $s \in [0,t]$) when $n \to \infty$

Martingal CLT applied to Nelson-Aalen

With $M(t) = N(t) - \int_0^t Y(s)\alpha(s)ds$ we have

$$\sqrt{n}[\hat{A}(t) - \int_0^t J(s)\alpha(s)ds] = \sqrt{n}\int_0^t \frac{dM(s)}{Y(s)} \to U(t)$$

for gaussian martingale $U(t)$ with expectation zero and variance function $\sigma^2(t)$ determined by

$$\lim_{n\to\infty}\langle\int_0^t \sqrt{n}\frac{dM(s)}{Y(s)}\rangle = \lim_{n\to\infty}\int_0^t \frac{nJ(s)}{Y(s)}\alpha(s)ds = \sigma^2(t).$$

This limit will exist if $J(s) = 1$ from a certain n and $\frac{n}{Y(s)}$ has a limit ($1/y(s)$).

Under similar conditions

$$\sqrt{n}\frac{dM(s)}{Y(s)} \to 0$$

Martingal CLT for Kaplan-Meier

For the martingale

$$\sqrt{n}\left[\frac{\hat{S}(t)}{S^{\star}(t)} - 1\right] = -\sqrt{n}\int_0^t \frac{\hat{S}(s-)}{S^{\star}(s)}\frac{dM(s)}{Y(s)}$$

we have

- the same limit $\sigma^2(t)$ of pred.var. process as $\langle\sqrt{n}(\hat{A}(t) - A^{\star}(t))\rangle$
- increments similarly become small

Thus, we have convergence to a gaussian martingale, in particular

$$\sqrt{n}\left[\frac{\hat{S}(t)}{S(t)} - 1\right] \to \mathrm{N}(0, \sigma^2(t))$$

Large sample distribution Kaplan-Meier

This leads to

$$\sqrt{n}[\hat{S}(t) - S(t)] \rightarrow \mathrm{N}(0, S(t)^2\sigma^2(t)).$$

and approximately

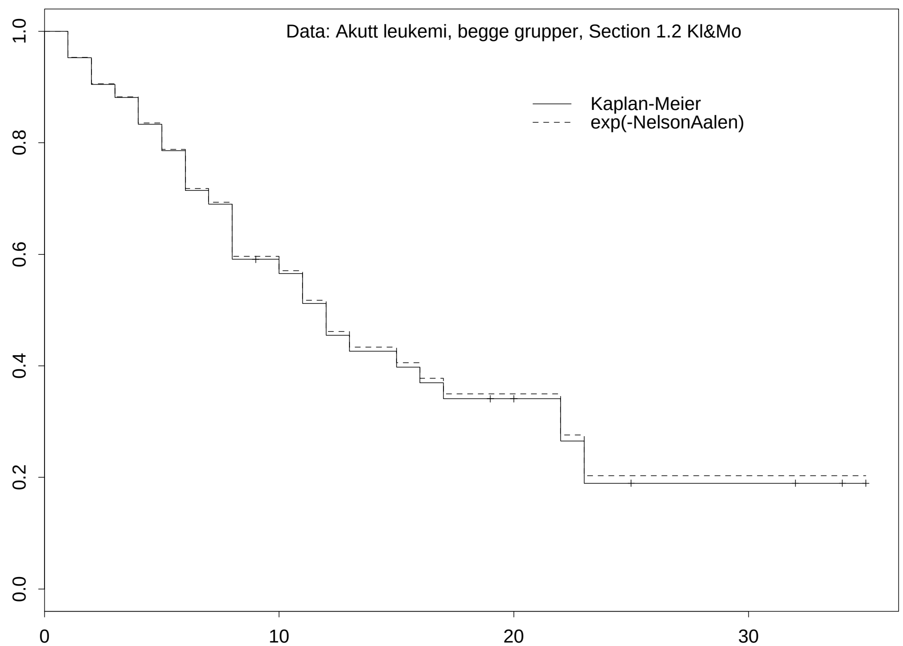
$$\hat{S}(t) \sim \mathrm{N}(S(t), S(t)^2\sigma^2(t)/n)$$

Comparison $-\log$(Kaplan-Meier) and Nelson-Aalen

Both $\hat{A}(t)$ and $-\log(\hat{S}(t))$ are estimators of cumulative hazard $A(t)$.

By a Taylor expansion when $dN(s)/Y(s)$ are small

$$\begin{aligned} -\log(\hat{S}(t)) &= -\int_0^t \log\left(1 - \frac{dN(s)}{Y(s)}\right) \\ &= \hat{A}(t) + \text{ Rest term} \end{aligned}$$



Estimation of expected value

Remember $\mu = \mathrm{E}[X] = \int_0^\infty S(t)dt$. This suggests estimator

$$\hat{\mu} = \int_0^\infty \hat{S}(t)dt$$

for μ.

Problem: can not estimate $S(t)$ longer than $\tau = \max(t : Y(t) > 0)$ and often: $\hat{S}(\tau) > 0$.

Programs will calculate

$$\hat{\mu} = \int_0^\tau \hat{S}(t)dt$$

which is an estimator of $\int_0^\tau S(s)ds$ and need not say much about the real expected value.

Uncertainty of $\hat{\mu}$

Assume $\hat{S}(\tau) = 0$, so that $\hat{\mu}$ is meaningful, and that $S(\tau) = 0$.

Then for $t \leq \tau$

$$\hat{S}(t) = S(t) - S(t) \int_0^t \frac{\hat{S}(s-)}{S(s)} \frac{dM(s)}{Y(s)} \approx S(t) - S(t) \int_0^t \frac{dM(s)}{Y(s)}$$

which gives

$$\begin{aligned} \hat{\mu} &= \int_0^\tau [S(t) - S(t) \int_0^\infty I(s < t) \tfrac{dM(s)}{Y(s)}] dt \\ &= \mu - \int_0^\tau \int_0^\tau I(s < t) S(t) dt \, \tfrac{dM(s)}{Y(s)} \\ &= \mu - \int_0^\tau \int_s^\tau S(t) dt \, \tfrac{dM(s)}{Y(s)} \end{aligned}$$

But $\int_s^\tau S(t)dt$ is a deterministic function, thus known in $\mathcal{F}_{s-}$ and then $\int_0^u \int_s^\tau S(t)dt \, \frac{dM(s)}{Y(s)}$ becomes a martingale.

Uncertainty for $\hat{\mu}$, contd.

But then

$$\text{Var}(\hat{\mu}) = \text{E}[\int_0^\tau (\int_s^\tau S(t)dt)^2 \frac{\alpha(s)}{Y(s)} ds]$$

and

$$\text{E}(\hat{\mu}) = \mu.$$

So we may estimate $\text{Var}(\hat{\mu})$ by

$$\widehat{\text{Var}}(\hat{\mu}) = \int_0^\tau (\int_s^\tau \hat{S}(t)dt)^2 \frac{dN(s)}{Y(s)^2}$$

With $\hat{\mu}_t = \int_0^t \hat{S}(s)ds$ (estimator of $\mu_t = \int_0^t S(s)ds$) we have $\int_s^\tau \hat{S}(t)dt = \hat{\mu}_\tau - \hat{\mu}_s$ which inserted into $\widehat{\text{Var}}(\hat{\mu})$ gives the formula on pg. 97 of ABG.