Outline

- Counting processes
- Martingales
- Applications to survival analysis:
  - Nelson-Aalen estimate
  - Cox partial likelihood (including time-varying covariates)

Primer: Stieltje's integral

For real functions f and g and $a < b$ Stieltje's integral is defined as

$$\int_a^b f(x)g(dx) = \int_a^b f(x)\mathrm{d}g(x) = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)[g(x_i) - g(x_{i-1})]$$

where $a = x_0 < x_1 < \cdots < x_n = b$.

Sufficient condition for existence: f continuous and g of bounded variation (i.e. $g = g_1 - g_2$ where g_1 and g_2 monotone functions).

Example: g continuously differentiable

$$\int_a^b f(x)g(dx) = \int_a^b f(x)g'(x)\mathrm{d}x$$

Example: g right-continuous piecewise constant with jumps $t_1, \dots, t_k$ in $[a, b]$:

$$\int_a^b f(x)g(dx) = \sum_{l=1}^{k} f(t_l)(g(t_l) - g(t_{l-1}))$$

Example: g piecewise continuous differentiable with jumps $t_1, \dots, t_k$ in $[a, b]$ (right-continuous in jumps):

$$\int_a^b f(x)g(dx) = \int_a^b f(x)g'(x)\mathrm{d}x + \sum_{l=1}^{k} f(t_l)(g(t_l) - g(t_l-))$$

Counting process

A continuous time stochastic process $N = \{N(t)\}_{t \geq 0}$ is a counting process if $N(0) = 0$, N is piece-wise constant right-continuous, and with probability one: $N(t) \in \mathbb{N} \cup \{0\}$ with jumps of size 1.

Example: A counting process N is a Poisson process with intensity function λ if for $0 \leq s < t$, $N(t) - N(s) \sim \text{Poisson}(\int_s^t \lambda(u)\mathrm{d}u)$ and if increments on disjoint intervals are independent. $N(t) - N(s)$ is interpreted as the number of "events" in $]s, t]$.

Equivalent definition: $N(t) - N(s) \sim \text{Poisson}(\int_s^t \lambda(u)\mathrm{d}u)$ and conditional on $N(t) - N(s) = n$, the n jump positions in $]s, t]$ are independent with density $f(u) \propto \lambda(u)$, $u \in]s, t]$.

Equivalent definition for constant intensity: the waiting times $W_i = T_i - T_{i-1}$ between jump locations T_i, $i = 1, 2, \ldots$ are independent Exponential(λ) random variables (here $T_0 = 0$ is not a jump location).

The last two definitions show ways to construct a Poisson process N (letting N increase by one at each jump position).

A counting process is also known as a point process - focus is then on the locations of jumps aka the points.

Concept can be generalized to higher dimensions - spatial point processes.

Discrete time martingale

Let $X_1, X_2, \ldots$ be independent with $X_i \in \{-1, 1\}$ and $P(X_i = 1) = p$ (e.g. simple model of changes in stock price).

Define

$$S_n = \sum_{i=1}^{n} X_i$$

Consider expectation given past:

$$\mathbb{E}[S_n | S_{n-1}] = \mathbb{E}[X_n | S_{n-1}] + S_{n-1} = \mathbb{E}X_n + S_{n-1} = 2p - 1 + S_{n-1}$$

Suppose $p = 1/2$. Then $\mathbb{E}[S_n | S_{n-1}] = S_{n-1}$ - best prediction of tomorrow (n) is value today ($n-1$). S_n is a martingale !

Suppose $p > 1/2$. Define compensator $\Lambda_n = n(2p-1)$ and $M_n = S_n - \Lambda_n$.

Then

$$\mathbb{E}[M_n | M_{n-1}] = \mathbb{E}[X_n - [2p-1] | M_{n-1}] + M_{n-1} = M_{n-1}$$

so *compensated* version of S_n is a martingale.

More generally we say that M_n is a martingale with respect to history $\mathcal{F}_n$ if

- M_n is measurable with respect to $\mathcal{F}_n$
- $\mathbb{E}[M_m | \mathcal{F}_n] = M_n$ when $m \geq n$

Same definition in case of continuous time !

For the discrete time cases, increments $S_m - S_n$, $S_p - S_q$ $(q < p < n < m)$ are obviously independent and hence uncorrelated.

This in fact holds in general for any martingale: increments are uncorrelated !

Continuous time martingale

Let for each $t \geq 0$ $\mathcal{F}_t$ denote set of 'information' available up to time t (technically, $\mathcal{F}_t$ is a σ-algebra) such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $0 \leq s \leq t$ (information increasing over time)

For a stochastic process M, $\mathcal{F}_t$ could e.g. represent the history of the process itself up to time t. $\mathcal{F}_t$ could also contain information about other stochastic processes evolving in parallel to M.

Definition: $M = \{M(t)\}_{t \geq 0}$ is a martingale with respect to $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$ if

- $\mathbb{E}[M(t)|\mathcal{F}_s] = M(s)$, $0 \leq s \leq t$.
- $M(t)$ determined by $\mathcal{F}_t$: knowledge of $\mathcal{F}_t$ means we know $M(t)$ (technically speaking, $M(t)$ is $\mathcal{F}_t$ measurable). We say M is adapted to $\mathcal{F}$.

Examples

Suppose N is a Poisson process with intensity $\lambda(\cdot)$. Let $\Lambda(t) = \mathbb{E}N(t) = \int_0^t \lambda(u)\mathrm{d}u$. Then

$$M(t) = N(t) - \Lambda(t)$$

is a martingale with respect to its own past $\mathcal{F}_t = \sigma((N(u))_{0 \le u \le t})$.

A Brownian motion is a martingale with respect to its own past.

Properties:

- If $M(0) = 0$ then $\mathbb{E}M(t) = 0$ for all $t \geq 0$.
- Uncorrelated increments over disjoint intervals: $\mathbb{E}[M(t) - M(s)][M(u) - M(v)] = 0$ for $0 \leq v \leq u \leq s \leq t$.

Martingale central limit theorem: a theorem that says that a sequence of martingales $M_n = \{M_n(t)\}_{t \geq 0}$, $n = 1, 2, \ldots$ converges to a Gaussian process (typically closely related to Brownian motion).

We shall consider survival analysis examples of such sequences.

Definition: a process X is predictable with respect to $\mathcal{F}$ if $X(t)$ is determined by $\mathcal{F}_{t-}$, i.e. information up to but not including t. In other words, $X(t)$ is known given $\mathcal{F}_{t-\mathrm{d}t}$.

Example: a left-continuous process is predictable given its own past: $X(t) = \lim_{h \to 0} X(t - h)$.

Infinitesimal characterization of martingale

Let $\mathrm{d}M(t) = M(\mathrm{d}t) = M((t+\mathrm{d}t)-) - M(t-)$ be increment over infinitesimal interval $[t, t+\mathrm{d}t[$ from t to $t+\mathrm{d}t$.

Then M is a martingale if

$$\mathbb{E}[\mathrm{d}M(t)|\mathcal{F}_{t-}] = 0$$

Heuristically, for $s < t$:

$$\begin{aligned}
\mathbb{E}[M(t)|\mathcal{F}_s] &= M(s) + \mathbb{E}\left[\int_{]s,t]} \mathrm{d}M(u)|\mathcal{F}_s\right] \\
&= M(s) + \int_s^t \mathbb{E}[\mathrm{d}M(u)|\mathcal{F}_s] \\
&= M(s) + \int_s^t \mathbb{E}\big[\mathbb{E}[\mathrm{d}M(u)|\mathcal{F}_{u-}]|\mathcal{F}_s] = M(s)
\end{aligned}$$

(here we used $\mathcal{F}_s \subseteq \mathcal{F}_{u-}$, $s < u$, for the third equality)

Why not define $\mathrm{d}M(t) = M(t + \mathrm{d}t) - M(t)$?

Usually our M is right continuous where left limits exist.

Then, with current definition of $\mathrm{d}M(t)$, $\mathrm{d}M(t)$ is non-zero if M has a jump at t.

For example, for a counting process N, $\mathrm{d}N(t)$ is equal to one if N jumps at t and zero otherwise.

In contrast, $N(t + \mathrm{d}t) - N(t)$ is always zero for infinitesimal $\mathrm{d}t$.

Application in survival analysis

Procedure:

1. express data as counting process N
2. construct martingale $M(t) = N(t) - \Lambda(t)$, $t \geq 0$.
3. Express Nelson-Aalen/Kaplan-Meier/Cox partial likelihood as a stochastic integral

$$\tilde{M}(t) = \int_0^t K(u)\mathrm{d}M(u)$$

   for some predictable process K. Note $\tilde{M}(u)$ is also a martingale (exercise).
4. Apply martingale central limit theorem to $\frac{1}{\sqrt{n}}\tilde{M}_n(t)$ (introducing n, number of subjects, in the notation) to get asymptotic normality.

Independent and identically distributed survival times

Given survival data (T_i, Δ_i), $i = 1, \ldots, n$ define zero or one-step counting processes

$$N_i(t) = 1[T_i \leq t, \Delta_i = 1] = 1[X_i \leq t, X_i \leq C_i]$$

and accumulated process,

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

Note: X_i independent continuous random variables implies N has jumps of size 1. $N(t)$ is number of deaths that happened before or at t

Define $Y_i(t) = 1[T_i \geq t]$. I.e. Y_i is one if ith individual at risk at time t and zero otherwise. Y_i is left-continuous and hence predictable. $Y(t) = \sum_{i=1}^{n} Y_i(t)$ is the number at risk at time t.

$\mathcal{F}_t$: history of N_i and Y_i, $i = 1, \ldots, n$ up to time t.

Compensator

Define

$$\Lambda_i(t) = \int_0^t Y_i(u)h(u)\mathrm{d}u$$

where h is the hazard rate of the X_i.

Then $\Lambda_i(t)$ is a continuous and hence predictable stochastic process.

Moreover, $M_i = N_i - \Lambda_i$ is a martingale: we argue next slide that

$$\mathbb{E}[\mathrm{d}N_i(t)|\mathcal{F}_{t-}] = \mathbb{E}[\mathrm{d}\Lambda_i(t)|\mathcal{F}_{t-}] \Leftrightarrow \mathbb{E}[\mathrm{d}M_i(t)|\mathcal{F}_{t-}] = 0$$

Note: regarding $\mathbb{E}[\mathrm{d}N_i(t)|\mathcal{F}_{t-}]$ two cases: $T_i < t$ (death or censoring already occurred) or $T_i \geq t$ (still at risk)

Case $T_i \geq t$:

$$\begin{aligned}\mathbb{E}[\mathrm{d}N_i(t)|\mathcal{F}_{t-}] &= \mathbb{E}\left[1[T_i \in [t, t+\mathrm{d}t[, C_i \geq X_i] | T_i \geq t\right] \\ `=' & \; P[X_i \in [t, t+\mathrm{d}t[, C_i \geq t | X_i \geq t, C_i \geq t] \\ &= P[X_i \in [t, t+\mathrm{d}t[| X_i \geq t, C_i \geq t]\end{aligned}$$

Under independent censoring, the last probability is $h(t)\mathrm{d}t = Y_i(t)h(t)\mathrm{d}t$ ('=' is because we replace $C_i \geq X_i$ by $C_i \geq t$).

Case $T_i < t$:

$$\mathbb{E}[\mathrm{d}N_i(t)|\mathcal{F}_{t-}] = \mathbb{E}[\mathrm{d}N_i(t)| T_i < t] = 0 = Y_i(t)h(t)\mathrm{d}t$$

(the only possible jump occurred prior to t).

Regarding $\mathrm{d}\Lambda_i(t)$:

$$\mathbb{E}[\mathrm{d}\Lambda_i(t)|\mathcal{F}_{t-}] = \mathbb{E}[Y_i(t)h(t)\mathrm{d}t|\mathcal{F}_{t-}] = Y_i(t)h(t)\mathrm{d}t$$

(where we used $Y_i(t)h(t)\mathrm{d}t$ predictable process, hence given $\mathcal{F}_{t-}$ we know $Y_i(t)$).

Conclusion:

$$\mathbb{E}[\mathrm{d}N_i(t)|\mathcal{F}_{t-}] = \mathbb{E}[\mathrm{d}\Lambda_i(t)|\mathcal{F}_{t-}] \Leftrightarrow \mathbb{E}[\mathrm{d}M_i(t)|\mathcal{F}_{t-}] = 0$$

It follows that

$$M(t) = N(t) - \Lambda(t)$$

is a martingale too where

$$\Lambda(t) = \sum_{i=1}^{n} \Lambda_i(t) = Y(t)h(t)$$

$M(0) = N(0) - \Lambda(0) = 0$ so $\mathbb{E}M(t) = 0$ for all $t \geq 0$.

Nelson-Aalen estimator

Define $0/0 = 0$. Then

$$\mathrm{d}N(u) = \mathrm{d}\Lambda(u) + \mathrm{d}M(u) \Leftrightarrow \frac{\mathrm{d}N(u)}{Y(u)} = 1[Y(u) > 0]h(u)\mathrm{d}u + \frac{\mathrm{d}M(u)}{Y(u)}$$

Integrating we obtain

$$\int_0^t \frac{\mathrm{d}N(u)}{Y(u)} = \int_0^t 1[Y(u) > 0]h(u)\mathrm{d}u + \int_0^t \frac{\mathrm{d}M(u)}{Y(u)}$$

Here:

- $H^*(t) = \int_0^t 1[Y(u) > 0]h(u)\mathrm{d}u$ is equal to $H(t)$ for $t \leq \max\{T_1, \ldots, T_n\}$.
- $W(t) = \int_0^t \frac{\mathrm{d}M(u)}{Y(u)}\mathrm{d}u$ is a zero-mean martingale 'noise' process
- $\hat{H}(t) = \int_0^t \frac{\mathrm{d}N(u)}{Y(u)}$ is an unbiased estimator of $H^*(t)$

Observe:

$$\hat{H}(t) = \sum_{t^* \in D : t^* \leq t} \frac{1}{Y(t^*)}$$

is precisely the Nelson-Aalen estimator.

Martingale central limit theorem for $\frac{1}{\sqrt{n}} W$ can be used to show asymptotic normality of $\hat{H}$.

Score process for Cox regression

We still assume that the counting processes N_i are independent but now with different hazard rates

$$h_i(t) = h_0(t)\exp[\beta^\mathsf{T} Z_i(t)]$$

Note: we immediately seize the opportunity to generalize the Cox regression model by allowing covariates $Z_i(t) = (Z_{i1}(t), \ldots, Z_{ip}(t))$ to be a time-varying predictable random process.

Compensators

$$\Lambda_i(t) = \int_0^t \lambda_i(u)\mathrm{d}u \quad \lambda_i(u) = Y_i(u)h_i(u) \quad \Lambda(t) = \sum_{i=1}^n \Lambda_i(t)$$

Partial log likelihood process:

$$l(\beta, t) = \sum_{i \in D: t_i \leq t} \left(\beta^\mathsf{T} Z_i(t_i) - \log\left[\sum_{l=1}^n Y_l(t_i)\exp(\beta^\mathsf{T} Z_l(t_i))\right]\right)$$

Note: partial log likelihood $l(\beta) = l(\beta, \infty)$. We here used risk process $Y_l(t_i)$ notation instead of risk set $R(t_i)$.

Score process

$$u(\beta, t) = \sum_{i \in D : t_i \le t} \left(Z_i(t_i) - \frac{\sum_{l=1}^{n} Y_l(t_i) Z_l(t_i) \exp(\beta^{\mathsf{T}} Z_l(t_i))}{\sum_{l=1}^{n} Y_l(t_i) \exp(\beta^{\mathsf{T}} Z_l(t_i))} \right)$$

$$= \sum_{i \in D : t_i \le t} (Z_i(t_i) - E(t_i))$$

where $\{E(t)\}_{t \ge 0}$ predictable process.

KM uses notation $(\bar{Z}_1(t), \dots, \bar{Z}_p(t))^{\mathsf{T}})$ for $E(t)$.

We can rewrite score-process to conclude that it is a martingale:

$$u(\beta, t) = \sum_{i=1}^{n} \int_0^t (Z_i(u) - E(u)) \mathrm{d}N_i(u) = \sum_{i=1}^{n} \int_0^t (Z_i(u) - E(u)) \mathrm{d}M_i(u)$$

(stochastic integral of predictable process with respect to a martingale is itself a martingale)

Last equality because

$$\sum_{i=1}^{n}\int_0^t (Z_i(u)-E(u))\mathrm{d}\Lambda_i(u) = \int_0^t \sum_{i=1}^{n}(Z_i(u)-E(u))\mathrm{d}\Lambda_i(u)$$

$$= \int_0^t \Big[\sum_{i=1}^{n} Z_i(u)Y_i(u)\exp(\beta^\mathsf{T} Z_i(u))$$

$$-E(u)\sum_{i=1}^{n} Y_i(u)\exp(\beta^\mathsf{T} Z_i(u))\Big] h_0(u)\mathrm{d}u = \int_0^t 0\mathrm{d}u = 0$$

We can again apply martingale central limit theorem to $\frac{1}{\sqrt{n}}u(\beta,t)$!

Residuals

Score process residuals: simply the p components of score process with β replaced by $\hat{\beta}$ and $\mathrm{d}M_i(u)$ replaced by

$$\mathrm{d}\hat{M}_i(u) = \mathrm{d}N_i(u) - Y_i(u)\exp(\hat{\beta}^\mathsf{T} Z_i(u))\mathrm{d}\hat{H}_0(u) = \mathrm{d}N_i(u) - \mathrm{d}\hat{\Lambda}_i(u)$$

where

$$\mathrm{d}\hat{H}_0(u) = \hat{H}_0(u) - \hat{H}_0(u-) = \begin{cases} \frac{1}{\sum_{l=1}^n Y_l(u)\exp(\hat{\beta}^\mathsf{T} Z_l(u))} & u \text{ death time} \\ 0 & \text{otherwise} \end{cases}$$

Martingale residuals:

$$r_{\mathrm{mart,i}}(t) = N_i(t) - \hat{\Lambda}_i(t)$$

Typically evaluated at $t = \infty$

$$r_{\mathrm{mart,i}}(\infty) = \delta_i - \hat{\Lambda}_i(\infty)$$

Martingale residuals sum to zero

$$\sum_{i=1}^{n} N_i(\infty) - \hat{\Lambda}_i(\infty) = \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} \int_0^{\infty} Y_i(u) \exp(\hat{\beta}^{\mathsf{T}} Z_i(u)) \mathrm{d}\hat{H}_0(u).$$

Last term is

$$\sum_{i=1}^{n} \sum_{k \in D} \frac{Y_i(t_k)) \exp(\hat{\beta}^{\mathsf{T}} Z_i(t_k))}{\sum_{l=1}^{n} Y_l(t_k) \exp(\hat{\beta}^{\mathsf{T}} Z_l(t_k))} = \sum_{k \in D} \frac{\sum_{i=1}^{n} Y_i(t_k)) \exp(\hat{\beta}^{\mathsf{T}} Z_i(t_k))}{\sum_{l=1}^{n} Y_l(t_k) \exp(\hat{\beta}^{\mathsf{T}} Z_l(t_k))}$$

which is equal to $\sum_{j=1}^{n} \delta_j$

Variance of martingale

$$
\begin{aligned}
\mathbb{V}\mathrm{ar} M(t) = \mathbb{V}\mathrm{ar} \int_0^t \mathrm{d}M(u) = \int_0^t \mathbb{V}\mathrm{ar}\, \mathrm{d}M(u) \\
= \int_0^t \mathbb{V}\mathrm{ar}\mathbb{E}[\mathrm{d}M(u)|\mathcal{F}_{u-}] + \mathbb{E}\mathbb{V}\mathrm{ar}[\mathrm{d}M(u)|\mathcal{F}_{u-}] \\
= 0 + \mathbb{E} \int_0^t \mathbb{V}\mathrm{ar}[\mathrm{d}M(u)|\mathcal{F}_{u-}] = \mathbb{E} \int_0^t \mathbb{V}\mathrm{ar}[\mathrm{d}M(u)|\mathcal{F}_{u-}]
\end{aligned}
$$

(note: we used uncorrelated increments for second equality)

Application to variance of Nelson-Aalen

In this case $M(t) = N(t) - \Lambda(t)$ and

$\mathbb{V}\mathrm{ar}[\mathrm{d}M(t)|\mathcal{F}_{t-}] = \mathbb{V}\mathrm{ar}[\mathrm{d}N(t)|\mathcal{F}_{t-}] = \lambda(t)\mathrm{d}t(1-\lambda(t)\mathrm{d}t) \approx \lambda(t)\mathrm{d}t$

where $\lambda(t)\mathrm{d}t = \mathrm{d}\Lambda(t) = Y(t)h(t)\mathrm{d}t$.

Nelson-Aalen estimator has "noise term"

$$\int_0^t \frac{1}{Y(u)}\mathrm{d}M(u)$$

which by exercise 2.1 is a martingale.

Hence variance is

$$\mathbb{V}\mathrm{ar}\hat{H}(t) = \mathbb{E}\int_0^t \mathbb{V}\mathrm{ar}[\frac{1}{Y(u)}\mathrm{d}M(u)|\mathcal{F}_{u-}]$$

$$=\mathbb{E}\int_0^t \frac{1}{Y(u)^2}\mathbb{V}\mathrm{ar}[\mathrm{d}M(u)|\mathcal{F}_{u-}] = \mathbb{E}\int_0^t \frac{1[Y(u)>0]}{Y(u)}h(u)\mathrm{d}u.$$

We estimate this by

$$\int_0^t \frac{1[Y(u) > 0]}{Y(u)} \mathrm{d}\hat{H}(u) = \sum_{\substack{t^* \in D: \\ t^* \leq t}} \frac{1}{Y(t^*)^2}$$

which coincides with (4.2.4) in KM ($Y(t^*) > 0$ for $t^* \in D$).

Predictable variation process

Let M denote a $\mathcal{F}$-martingale.

Conditional variance of martingale increment:

$$\begin{aligned}&\mathbb{V}\mathrm{ar}[\mathrm{d}M(t)|\mathcal{F}_{t-}] = \mathbb{E}[(\mathrm{d}M(t))^2|\mathcal{F}_{t-}) - (\mathbb{E}[\mathrm{d}M(t)|\mathcal{F}_{t-})])^2\\ =&\mathbb{E}[M((t+\mathrm{d}t)-)^2 + M(t-)^2 - 2M((t+\mathrm{d}t)-)M(t-)|\mathcal{F}_{t-}] - 0\\ =&\mathbb{E}[M((t+\mathrm{d}t)-)^2 - M(t-)^2|\mathcal{F}_{t-}] = \mathbb{E}[\mathrm{d}(M(t)^2)|\mathcal{F}_{t-}].\end{aligned}$$

We define the predictable variation process $<M>$ as

$$\mathrm{d}<M>(t) = \mathbb{E}[\mathrm{d}(M(t)^2)|\mathcal{F}_{t-}]$$

Note: $\{M(s)^2- <M>(s)\}_{s\geq 0}$ is yet another martingale.

By previous slide we have
$\mathbb{V}\mathrm{ar}M(t) = \mathbb{E}\int_0^t \mathbb{V}\mathrm{ar}[\mathrm{d}M(u)|\mathcal{F}_{u-}] = \mathbb{E}\int_0^t \mathrm{d}<M>(u)$

Variance of score process

(for ease of notation assume Z_i one-dimensional, use $\mathrm{d} < M_i > (u) = \lambda_i(u)\mathrm{d}u$)

We use Exercise 2.2 for the second equality.

$$\mathbb{V}\mathrm{ar} u(\beta, t) = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar} \int_0^t (Z_i(u) - E(u))\mathrm{d}M_i(u)$$

$$= \mathbb{E} \int_0^t \sum_{i=1}^{n} (Z_i(u) - E(u))^2 \lambda_i(u)\mathrm{d}u$$

$$= \mathbb{E} \int_0^t \Big[\sum_{i=1}^{n} Z_i(u)^2 Y_i(u) \exp(\beta^\mathsf{T} Z_i(u)) + E(u)^2 \sum_{i=1}^{n} Y_i(u) \exp(\beta^\mathsf{T} Z_i(u))$$

$$- 2E(u) \sum_{i=1}^{n} Z_i(u) Y_i(u) \exp(\beta^\mathsf{T} Z_i(u)) \Big] h_0(u)\mathrm{d}u$$

Continues on next slide

$$
\begin{aligned}
=&\mathbb{E}\int_0^t \Big[\sum_{i=1}^n Z_i(u)^2 Y_i(u)\exp(\beta^{\mathsf{T}} Z_i(u)) \\
&- E(u)^2 \sum_{i=1}^n Y_i(u)\exp(\beta^{\mathsf{T}} Z_i(u))\Big] h_0(u)\mathrm{d}u \\
=&\mathbb{E}\int_0^t \sum_{i=1}^n (Z_i(u)^2 - E(u)^2)\lambda_i(u)\mathrm{d}u
\end{aligned}
$$

- is equal to information

Let

$$V(u) = \frac{\sum_{i=1}^n Z_i(u)^2 Y_i(u)\exp(\beta^\mathsf{T} Z_i(u))}{\sum_{i=1}^n Y_i(u)\exp(\beta^\mathsf{T} Z_i(u))} - E(u)^2$$

Then

$$i(\beta,t) = \mathbb{E} j(\beta,t) = \mathbb{E}\int_0^t \sum_{i=1}^n V(u)\mathrm{d}N_i(u)$$

$$= \mathbb{E}\int_0^t \sum_{i=1}^n V(u)\mathbb{E}[\mathrm{d}N_i(u)|\mathcal{F}_{u-}] = \mathbb{E}\int_0^t \sum_{i=1}^n V(u)\lambda_i(u)\mathrm{d}u$$

$$= \mathbb{E}\int_0^t V(u)\Big[\sum_{i=1}^n Y_i(u)\exp(\beta^\mathsf{T} Z_i(u))\Big] h_0(u)\mathrm{d}u =$$

$$\mathbb{E}\int_0^t \Big(\sum_{i=1}^n Z_i(u)^2 Y_i(u)\exp(\beta^\mathsf{T} Z_i(u)) - E(u)^2 Y_i(u)\exp(\beta^\mathsf{T} Z_i(u))\Big) h_0(u)\mathrm{d}u$$

$$= \mathbb{E}\int_0^t \sum_{i=1}^n (Z_i(u)^2 - E(u)^2)\lambda_i(u)\mathrm{d}u$$

A new paradigm for modeling: view data as generated from a counting process. Specify model for compensator.

This set-up allows for

- multiple events for each subject
- subjects being on-off risk (e.g. Vemmetofte data)
- time-varying stochastic covariate processes
- we do not need $\lim_{u\to\infty} H_i(u) = \infty$ (versus the usual survival set-up where we require $P(X_i < \infty) = 1 \Leftrightarrow S_i(\infty) = \exp(-H_i(\infty)) = 0$)
- use of powerful martingale theory for establishing asymptotic results

Exercises

1. A Brownian motion $\{B(s)\}_{s \geq 0}$ is a continuous-time zero-mean Gaussian process[1] with $B(0) = 0$ and $\mathbb{C}\mathrm{ov}(B(s), B(t)) = \min(t, s)$ for $s, t \geq 0$.
   - Show that a Brownian motion has uncorrelated and hence independent increments over disjoint intervals
   - show that a Brownian motion is a martingale with respect to its own history:

$$\mathbb{E}[B(t)|B(u), 0 \leq u \leq s] = B(s)$$

2. Show heuristically that if M is a martingale and K is a predictable process (both with respect to $(\mathcal{F}_t)_{t \geq 0}$) then
   2.1 $\tilde{M}(t) = \int_0^t K(u)\mathrm{d}M(u)$ is a martingale
   2.2 $\tilde{M}$ has predictable variation process $<\tilde{M}>(t) = \int_0^t K(u)^2 \mathrm{d} <M>(u)$.
3. Show that a martingale has uncorrelated increments (cf. slide 11).

[1] I.e. all finite-dimensional distributions are Gaussian