- Let h_j denote the increment in H_0 in interval j, that is,

$$h_j = H_0(s_j) - H_0(s_{j-1}), \quad j = 1, \ldots, J.$$

Assuming $H_0 \sim \mathcal{GP}(c_0 H^\star, c_0)$, we know h_j are **independent** and

$$h_j \overset{indep}{\sim} \text{Gamma}(\alpha_{0,j} - \alpha_{0,j-1}, c_0),$$

where $\alpha_{0,j} = c_0 H^\star(s_j)$.

- Consider the piecewise constant baseline hazard with $h_j = (s_j - s_{j-1})\lambda_j \Rightarrow$ Observe a great similarity.

3.23 (ICS) Relationship to Partial likelihood

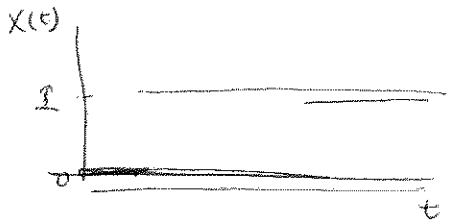
- Cox model under a GP for $H_0(t)$, $GP(c_0 H^*, c_0)$

$$\lim_{c_0 \to 0} \pi(\beta \mid \text{data}) \approx \underbrace{\prod_{i=1}^{n} \left[\frac{\exp(x_i'\beta)}{\sum_{l \in R_i} \exp(x_l'\beta)} \right]^{\nu_i}}_{= \text{Cox partial likelihood}}$$

$c_0 \to \infty$, $H_0 \approx H^*$

$$H^* = \eta_0 t^{\kappa_0}$$

3.24 Gamma process on baseline hazard $h_0(t)$



♣ More on Covariates

- **(fixed-time covariates)** Covariates may be constant (or fixed) values known at time 0.

  e.g. initial disease status.

  X(t)

- **(time-dependent covariates)** Covariates may be time dependent, i.e., their value changes over time.

  e.g. A covariate takes on the value 0 until some intermediate event occurs. It becomes 1 when the event occurs.

  ** Observe time to the intermediate event is random.

  more e.g. current disease status, serial blood pressure measurements.

↻ See KM Chapter 9 and ICS 10.1.3.

- More on Time-Dependent Covariates (KM p307):

Kalbfleisch and Prentice (1980) distinguish between two types of time-dependent covariates.

⋆⋆ **Type 1:** An *external* covariate is one that is not directly related to the failure mechanism.

↺ Case 1: Their value is completely under the control of the investigator.

e.g. a planned schedule of treatments under the control of the investigator

↺ Case 2: Ancillary time-dependent covariates are the output of a stochastic process external to the failure process

e.g. daily temperature as a predictor of survival from a heart attack

$\star\star$ **Type 2:** An *internal* covariate is a value over time generated by the individual under study. These covariats are measured only as long as the individual is still under observation.

e.g. blood pressure, and CD4 counts measured over the course of the study

$\Rightarrow$ The distribution of theses covariates carries information about the failure process.

- What does the difference between the two types imply?

  ⋆⋆ **Type 1:** Consider the survival probability

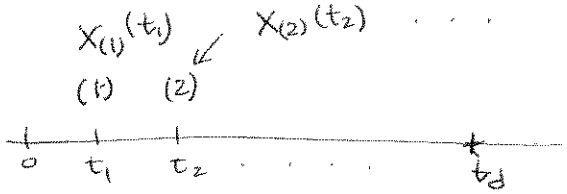
$$\begin{aligned} S(t \mid X(t)) &= P(T > t \mid X(t)) \\ &= \exp\left(-\int_0^t h(u \mid X(t))du\right) \\ &= \exp\left(-\int_0^t h_0(u)e^{\beta X(t)}du\right). \end{aligned}$$

  ⋆⋆ **Type 2:** $S(t \mid X(t)) = 1$ provided that $X(t)$ does not indicate that the individual has died (must be alive and at risk of failure).

  $\Rightarrow$ For internal covariates, the partial likelihood construction is still valid, but it is not possible to estimate the conditional survival function.

- Set-up for time-dependent covariates
  - T: time to some event
  - $\mathbf{X}(t) = (X_1(t), \ldots, X_p(t))'$: a set of covariates at time t which may affect the survival distribution of T.

- Assume:
  - Given $\mathbf{X}_i(t)$, the event and censoring time for subject i is independent.
  - The value of $\mathbf{X}_i(t)$ is known for any time at which the subject is under observation.
  - The value of the covariate is known at an instant just prior to time t (that is, their values are predictable)



- Consider an extension of the proportional hazards model
  - ⋆⋆ The event times are distinct and $t_1 < \ldots < t_d$ denote the ordered event times. $j=1,\ldots,d$
  - ⋆⋆ $\mathbf{X}_{(j)}(t_j)$: the covariate associated with the individual whose failure time is t_j.
  - ⋆⋆ R_j: the risk set at time t_j

- The partial likelihood is

$$\mathrm{PL}(\beta) = \prod_{j=1}^{d} \frac{\exp\left(\sum_{k=1}^{p} \beta_k X_{(j)k}(t_j)\right)}{\sum_{\ell \in R_j} \exp\left(\sum_{k=1}^{p} \beta_k X_{\ell k}(t_j)\right)}.$$

- The partial likelihood is

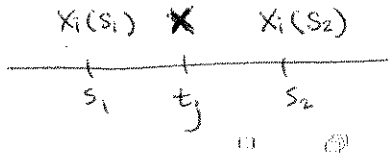
$$\mathrm{PL}(\beta) = \prod_{j=1}^{d} \frac{\exp\left(\sum_{k=1}^{p} \beta_k X_{(j)k}(t_j)\right)}{\sum_{\ell \in R_j} \exp\left(\sum_{k=1}^{p} \beta_k X_{\ell k}(t_j)\right)}.$$

$\star\star$ It is assumed that at each uncensored time t_j, the values of the covariates are observed for all the subjects who are at risk at time t_j.

$\star\star$ If X varies its value continuously over time and is measured only at certain time intervals.

$\Rightarrow X_i(t_j)$ may not be available.

$\star\star$ To overcome the issue, a possible approach is some interpolation between repeated measurements.



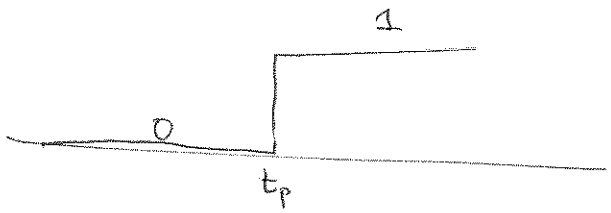
* Revisit Acute Leukemia Example (Example 9.1 of KM)

- Bone marrow transplants are a standard treatment for acute leukemia.
- A total of 137 patients are treated.
- Several risk factors were measured at the time of transplantation.
- One of them is the disease group (fixed covariate); patients were grouped into three risk categories based on their status at the time of transplantation;
  - ALL (38 patients)
  - AML low-risk first remission (54 patients)
  - AML high-risk second remission or untreated first relapse (15 patients) or second or greater relapse or never in remission (30 patients).

* Example 9.1 of KM (contd):

  - An individual is said to be disease-free at a given time after transplant if that individual is alive without the recurrence of leukemia.
  - The event indicator for disease-free survival is $\nu = 1$ if the individual has died or has relapsed.
  - The days on study for a patient is the smaller of their relapse or death time.
  - There is an intermediate event that occurs during the transplant recovery process
    - Return of the patient's platelet count to a self-sustaining level (platelet recovery):

      $X_P(t) = 1$ if platelet recovery occurs prior to time t



* Example 9.1 of KM (contd):

  - The time of the intermediate event, if it occurs, is random.
  - Goal:
    - Examine the relationship of covariates (risk categories, platelet recovery $\mathbb{X}_p$) to the disease-free survival time
    - See how the effects of the fixed covariates change when their intermediate events occur.

* **X**: fixed factors & Platelets

```
> mySurv <- Surv(t1, t2, d)
> myCPH  <- coxph(mySurv ~ as.factor(g) + i)
> myCPH
Call:
coxph(formula = mySurv ~ as.factor(g) + i)

                    coef exp(coef) se(coef)     z       p
as.factor(g)2 -0.494     0.610    0.289 -1.71 0.08773
as.factor(g)3  0.372     1.450    0.269  1.38 0.16743
i             -1.130     0.323    0.329 -3.44 0.00059

Likelihood ratio test=22.6  on 3 df, p=4.82e-05
n= 256, number of events= 82
```

Baseline group: ALL

AML-Low

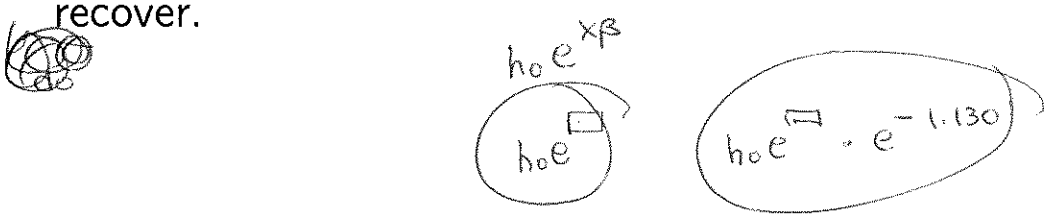
AML-High

having the platelet recovery

* Interpretation of the time-dependent covariate coefficient.

- The estimate of coefficient of platelet recovery is -1.130, which suggests that a patient whose platelets has recovered at a given time has a better chance of survival than a patient who, at that time, has yet to have platelets recover.

- The relative risk of $\exp(-1.130) = 0.323$ suggests that the rate at which patients are relapsing or dying after their platelets recover is about one-third the rate prior to the time at which their platelets recover.



- How about Bayesian approaches to accommodate time-dependent covariates? (ICS 10.1.3)

$$h(t \mid x(t)) = h_0(t)\psi(x(t)).$$

$$\Rightarrow \quad S(t \mid x(t)) = \exp\left\{-\int_0^t h_0(u)\psi(x(u))du\right\}.$$

$$\Rightarrow \quad f(t \mid x(t)) = h_0(t)\psi(x(t)) \exp\left\{-\int_0^t h_0(u)\psi(x(u))du\right\}.$$

$\star\star$ H_0 and ψ no longer separate into a product.

$\star\star$ Furthermore, H_0 does not even appear!

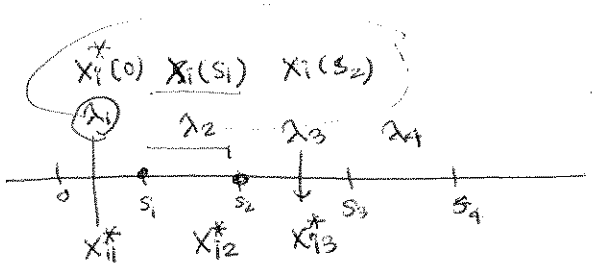
- *What does this imply?*

$**$ We need to compute

$$A(t \mid x(t)) = \exp\left\{-\int_0^t h_0(u)\psi(x(u))du\right\}.$$

$\star\star$ In other words, each likelihood evaluation requires n numerical integrations.

$\star\star$ Sampling from the resulting posterior distribution is very challenging.



• Ex: Consider a Cox piecewise exponential hazard model (Bradshaw et al, 2010)

⋆⋆ Construct a finite partition of the time axis, $0 < s_1 < s_2 < \ldots < s_J$ with $s_J > \max(y_i)$.

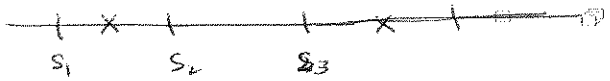
$\Rightarrow$ we have the J intervals, $(0, s_1]$, $(s_1, s_2]$,..., $(s_{J-1}, s_J]$.

⋆⋆ The measurement times for the covariate vector are assumed to fall at the boundaries of the intervals.

⋆⋆ Each subject provides a series of longitudinal measurements for p covariates by $\mathbf{X}_{ik}$ measured at time v_{ik} for $k = 1, \ldots, K_i$, $K_i \geq 1$.

⋆⋆ Thus, it is possible for a measurement to span multiple intervals.

e.g. if measurements on $\mathbf{X}$ are taken every 2 years but the intervals $(s_{j-1}, s_j]$ correspond to 1 year each. $\Rightarrow K_i \leq J$.



- Consider a Cox piecewise exponential hazard model (Bradshaw et al, 2010)

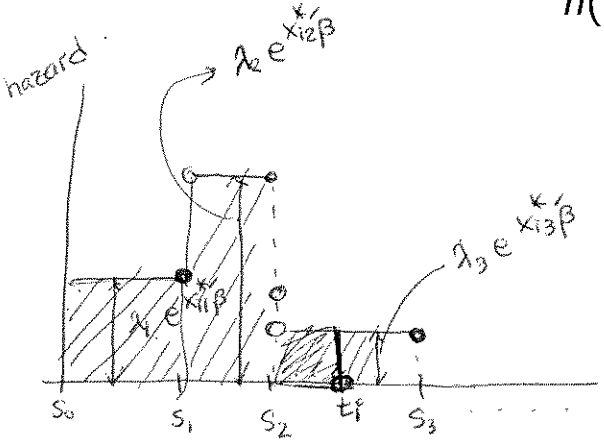
  ⋆⋆ For subject i within interval j, we define $\mathbf{X}^{\star}_{ij}$, which may represent

  * the previous observation of the variable carried forward into the current interval, or,

  * for continuous variables, an interpolated value between two observations.

  ⋆⋆ The hazard function under the piecewise constant hazards model, for $t_i \in (s_{j-1}, s_j]$,

$$h(t_i \mid \mathbf{X}^{\star}_{ij}) = \lambda_j \exp\left(\mathbf{X}^{\star\prime}_{ij}\beta\right).$$



$t_i \in (s_2, s_3]$

$H(t_i \mid x^*, \beta, \lambda)$

$= (s_1 - s_0) \times \lambda_1 e^{x^{*\prime}_{i1}\beta}$

$+ (s_2 - s_1) \times \lambda_2 e^{x^{*\prime}_{i2}\beta}$

$+ (t_i - s_2) \times \lambda_3 e^{x^{*\prime}_{i3}\beta}$

$= \int_0^t h(u \mid x^*, \beta, \lambda)\, du$

- Shall we write down the likelihood?

  ⋆⋆ Define censoring indicator ν_i as

$$\nu_i = \begin{cases} 1 & \text{if subject } i \text{ failed (observed survival time),} \\ 0 & \text{otherwise.} \end{cases}$$

  ⋆⋆ Define δ_{ij} as

$$\delta_{ij} = \begin{cases} 1 & \text{if subject } i \text{ failed or censored in interval } j \\ 0 & \text{otherwise} \end{cases}$$

  ⋆⋆ The likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\lambda})$ is

$$\mathcal{L}_i = \prod_{j=1}^{J} \left(\lambda_j e^{x_{ij}^{*\prime}\beta} \right)^{\nu_i \delta_{ij}} \cdot \exp\left\{ -\sum_{j=1}^{J} \delta_{ij} \left(\sum_{g=1}^{j-1} (s_g - s_{g-1}) \lambda_g \cdot e^{x_{ig}^{*\prime}\beta} + (y_i - s_{j-1}) \lambda_j e^{x_{ij}^{*\prime}\beta} \right) \right\}$$

$$\mathcal{L} = \prod_{i=1}^{n} \mathcal{L}_i$$

$$P(\beta, \lambda \mid \text{data}) \propto \prod_{i=1}^{n} \mathcal{L}_i \quad \pi(\beta) \quad \pi(\lambda)$$

- Possible Error?

of coefficients on the vector of covariates $\mathbf{z}_{ij}^*$. The density for the observed failure time y_i within interval j is then:

$$p_{y,j}(y_i|\mathbf{x}_{ij}^*, \mathbf{z}_{ij}^*, \beta_1, \beta_2, \lambda_j) = (\lambda_j \exp(\mathbf{x}_{ij}^{*\prime}\beta_1 + \mathbf{z}_{ij}^{*\prime}\beta_2))^{\delta_i} (\exp(-\Lambda_j(y_i)))^{\exp(\mathbf{x}_{ij}^{*\prime}\beta_1 + \mathbf{z}_{ij}^{*\prime}\beta_2)}$$

for $y_i \in (s_{j-1}, s_j]$ with cumulative hazard function:

$$\Lambda_j(y_i) = \left((y_i - s_{j-1})\lambda_j \exp(\mathbf{x}_{ij}^{*\prime}\beta_1 + \mathbf{z}_{ij}^{*\prime}\beta_2) + \sum_{g=1}^{j-1} (s_g - s_{g-1})\lambda_g \exp(\mathbf{x}_{ig}^{*\prime}\beta_1 + \mathbf{z}_{ig}^{*\prime}\beta_2) \right).$$

We further let $\lambda = (\lambda_1, \ldots, \lambda_J)'$ denote the $J \times 1$ vector of baseline hazards λ_j and let Δ_{ij} be an indicator of if subject i died or was censored in interval j (i.e. $y_i \in (s_{j-1}, s_j]$). The ith contribution to the complete data likelihood for the piecewise exponential model is then:

$$p_y(y_i|\mathbf{x}_i, \mathbf{z}_i, \beta_1, \beta_2, \lambda) = \prod_{j=1}^{J} (\lambda_j \exp(\mathbf{x}_{ij}^{*\prime}\beta_1 + \mathbf{z}_{ij}^{*\prime}\beta_2))^{\Delta_{ij}\delta_i} \exp\{-\Delta_{ij}[\Lambda_j(y_i)]\exp(\mathbf{x}_{ij}^{*\prime}\beta_1 + \mathbf{z}_{ij}^{*\prime}\beta_2)\} \quad (3)$$

where $\mathbf{x}_{ij}^* = \mathbf{x}_{ik}$ and $\mathbf{z}_{ij}^* = \mathbf{z}_{ik}$ with k and j such that $v_{ik} \leqslant s_{j-1} < v_{i,k+1}$. If we define $t_{ij} = \min(y_i, s_{j+1}) - s_j$ if $y_i \geqslant s_j$ and 0

* Revisit Acute Leukemia Example (Example 9.1 of KM)

* **X**: fixed factors & Platelets

risk groups × (β) time-varying covariates Z (γ)

$\beta \sim N(\bar{\beta}, \Sigma)$

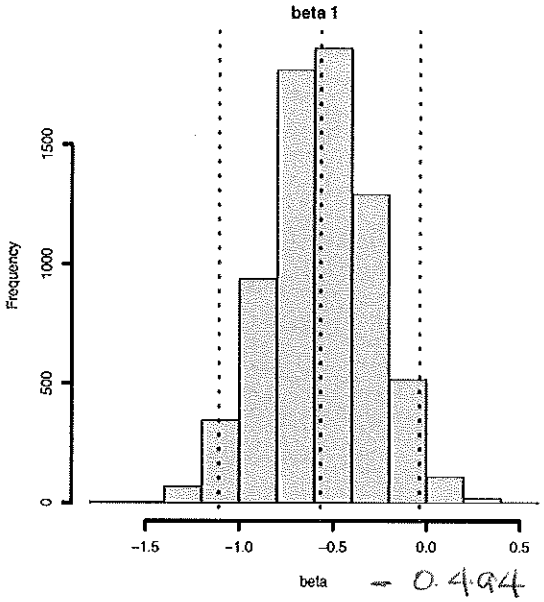
```
> ###  set up hyperparameters
> hyper <- NULL
>
> ### Be \sim N_p(Beta_bar, Sig)
> ## fit the frequentist Cox to set hyperparamters
> hyper$Beta_bar <- c(-0.494, 0.372) ## estimate from frequentist cox
> hyper$Sig <- diag(2.0, p)
> hyper$Inv_Sig <- solve(hyper$Sig)
>
> hyper$gam_bar <- -1.130
> hyper$nu2 <- 2
>
> ##  lambda_j \iid \Ga(a0, lam0)
> hyper$a0 <- 0.1
> hyper$lam0 <- 0.1
```

$\lambda_j \overset{iid}{\sim} Gea(a_0, \lambda_0)$

$\gamma \sim N(\bar{\gamma}, \nu^2)$

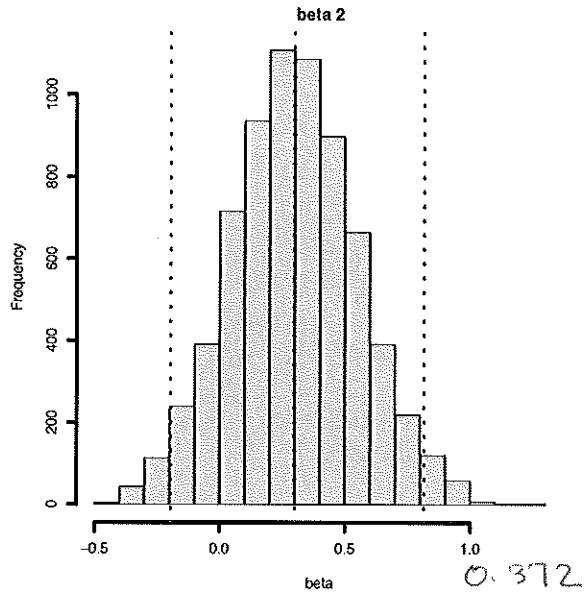
X:

- Piecewise Constant Hazard Model with a Time-Varying Covariate



$\beta_1(\hat{\beta}_1 = -0.565)$

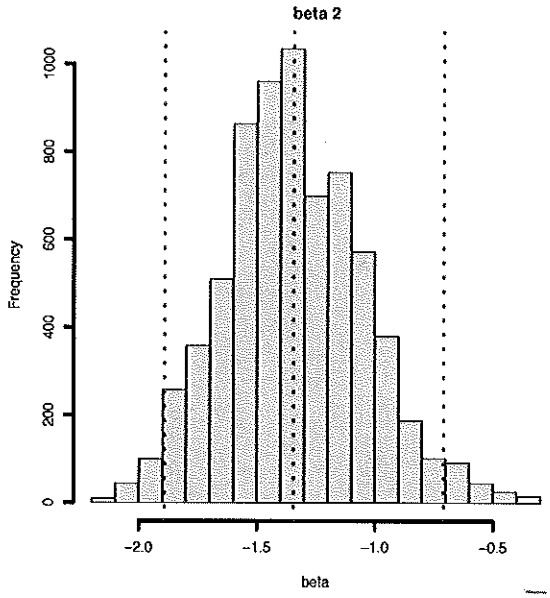
AML low-risk



$\beta_2(\hat{\beta}_2 = 0.299)$

AML high risk

- Piecewise Constant Hazard Model with a Time-Varying Covariate



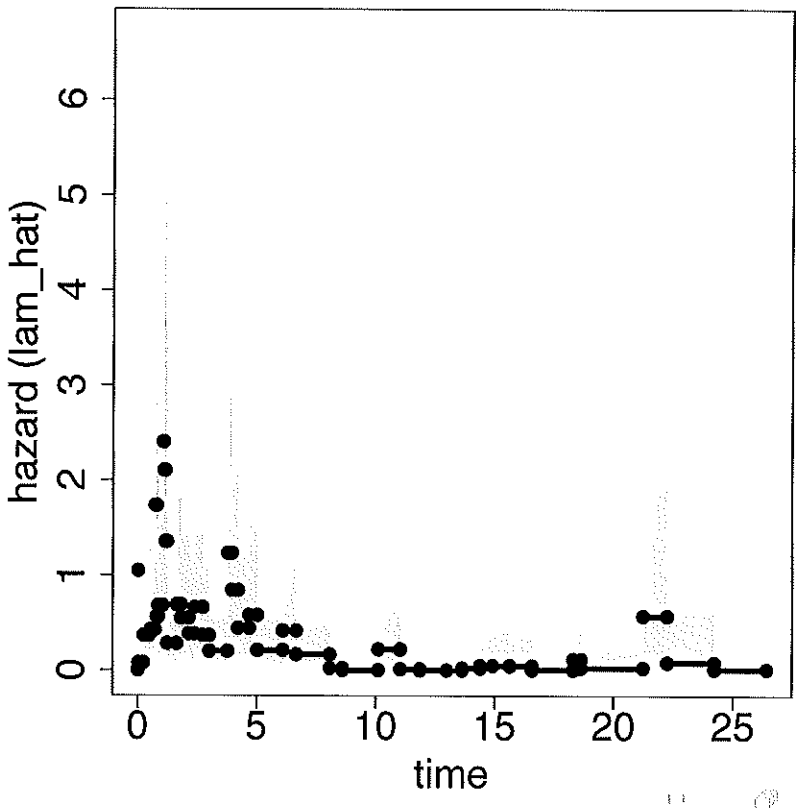
-1.130

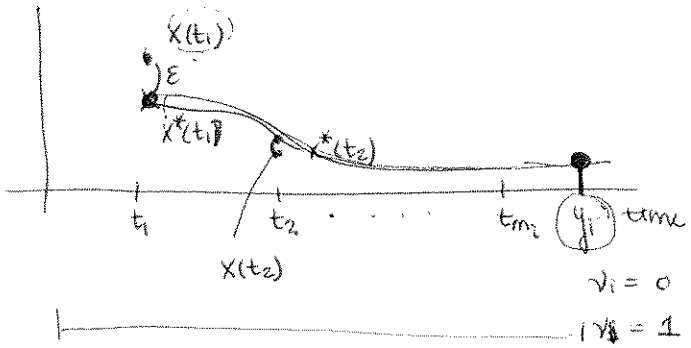
$\beta_3(\hat{\beta}_3 = -1.345)$

platelet recovery

- Piecewise Constant Hazard Model with a Time-Varying Covariate

$\star$ Posterior mean of λ_j with their 95% credible intervals





t_{i1} x_{i1} y_i, ν_i

t_{i2} x_{i2} z_i

t_{im_i} x_{im_i}

$$x_{ij} = \underbrace{X^*(t_{ij})}_{\text{a function of time and } z_i} + \varepsilon_{ij}$$

$f(y_i \mid x_{ij}, j=1,\dots,m_i, z_i, \text{——})$

♣ Alternative Approach (ICS Chapter 7)

- (so far) We assume $\mathbf{X}_i(t)$ is known for any time at which the subject is under observation and treat $\mathbf{X}_i(t)$ as fixed variables.

- We treat $\mathbf{X}_i(t)$ as observations, representing some function of the true covariate, $\mathbf{X}^*(t)$ (referred as the *trajectory function*)

  ↻ Often $\mathbf{X}_i(t)$ are incomplete or may be prone to measurement error.

  ↻ Simply including raw measurements in the analysis leads to bias.

  ⇒ Joint modeling of longitudinal and survival data!

$f(y_i \mid x^*(y_i), z_i, \text{——})$

- Health-related quality of life studies (Chapter 7.1.3)

  $\star\star$ A quality of life (QOL) survey instrument is typically administered to study participants at a number of prespecified time points during treatment and follow-up.

  $\star\star$ *Why is it important to study?* For a patient, quality of life is at times an even more important factor in treatment decisions than any modest survival benefit.

  $\Rightarrow$ Provides more useful information for the decision-making process of both patient and physician.

- Health-related quality of life studies (Chapter 7.1.3 – contd)

  ⋆⋆ Complete QOL data for patients at all of the specified collection times is frequently unavailable due to adverse events such as treatment toxicities or disease progression.

  ↺ Patients who are very ill when they report to the clinic may be less likely to complete the QOL instrument.

  ↺ Clinic personnel may feel that it is unethical to ask a patient to complete such a form when the patient feels so poorly.

  ⇒ The missingness of QOL data is related to the patient's QOL at the assessment time.

  ⇒ Such nonignorable missingness often leads to serious biases and must be taken into account at the time of analysis.

- Health-related quality of life studies (Chapter 7.1.3 – contd)
  - ⋆⋆ Develop a joint model for longitudinal and survival data
    - ↺ The longitudinal measure is QOL.
    - ↺ The survival component of the model acts as a type of non-ignorable missing data mechanism.