The Basic Joint Model

Day 5

- Introduction to shared random effects model

Brief Review

- Longitudinal outcome: Model a continuous repeated measures outcome using a linear mixed effects model

$$y_i(t) = X_i'\beta + Z_i'(t)b_i + \epsilon_i(t), \qquad \epsilon_i(t) \sim N(0, \sigma^2)$$

- Survival outcome: Model a time-to-event outcome using a proportional hazards model

$$h_i(t) = h_0(t)\exp\{\gamma' w_i\}$$

Brief Review

- **New research questions:** What can we do if the longitudinal and survival outcomes are related?
- **Two perspectives:**
  1. Longitudinal studies are often affected by informative drop-out (e.g., due to death)
     - If patients with higher serum bilirubin are more likely to die, will this affect our estimators of the trajectory of serum bilirubin over time?
  2. How to assess if a time-varying biomarker that is measured with error is associated with the event of interest?
     - What if the trajectory of serum bilirubin impacts the risk of death?

Brief Review

- Using the time-varying covariate approach (extended Cox)

$$h_i(t) = h_0(t) \exp\{\gamma' w_i + \alpha y_i(t)\}$$

- If the longitudinal marker is given by

$$\begin{aligned} y_i(t) &= m_i(t) + \epsilon_i(t) \\ &= X_i'\beta + Z_i'(t)b_i + \epsilon_i(t), \qquad \epsilon_i(t) \sim N(0, \sigma^2) \end{aligned}$$

- Using the two-stage approach

$$h_i(t) = h_0(t) \exp\{\gamma' w_i + \alpha \hat{m}_i(t)\}$$

Brief Review

- Issues with these approaches
- The time-varying covariate approach:
  - Assuming the biomarker doesn't change value between observations is a very strong, often implausible assumption
- The two-stage approach:
  - The uncertainty in our estimates form the first stage are not carried through to the second stage

Brief Review

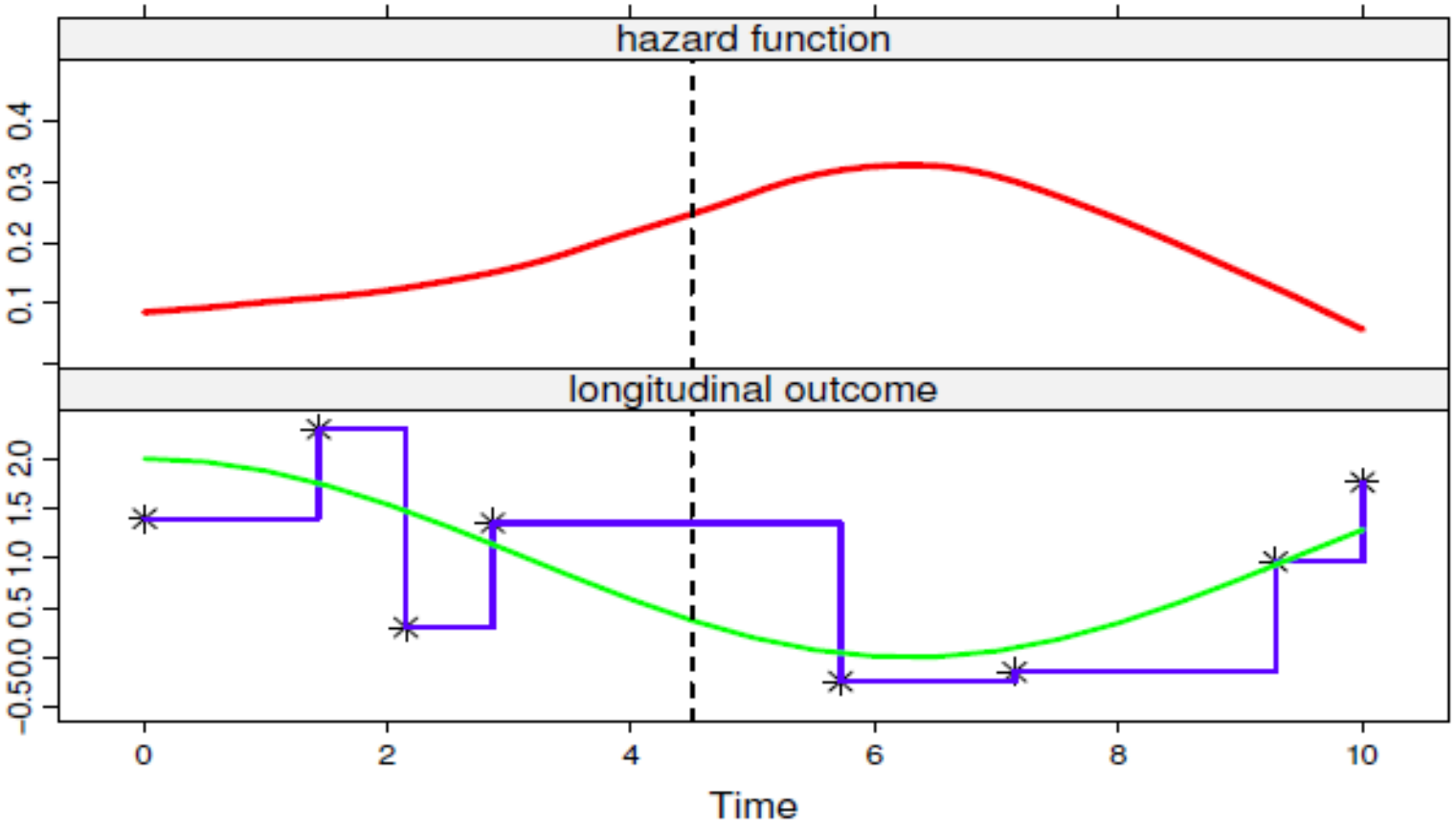
- Inherent features of biomarkers
  - Measured with error
  - Measurements taken on the same individual are correlated
  - Value of the biomarker may be related to prognosis
- To account for the special features of endogenous covariates a new class of models has been developed

  Joint Models for Longitudinal and Time-to-Event Data

Joint Modeling

- Model the longitudinal and survival processes using a single model
  - Account for measurement error
  - Utilizes all available repeated measures
  - Marker levels are not assumed to be constant between visits
  - Reduces bias and maximize efficiency

Joint Modeling Framework



Joint Modeling Framework

- Intuitive idea: Think of it as two component models
  1. Longitudinal part - to describe the evolution of the marker over time for each patient
  2. Survival part - the estimated evolutions are then used in a Cox model
- The component parts share some parameter dependence through shared random effects

Joint Modeling Framework

Notation:

- T_i^*: True event time for patient i
- T_i: Observed event time for patient i
- δ_i: Event indicator, i.e., equals 1 for true events
- y_i(t): Longitudinal responses
- $y_{ij} = \{y_i(t_{ij}), j = 1, \dots, n_i\}$: Observed longitudinal measurements

- We will formulate the joint model in 3 steps…

Longitudinal Submodel

- **Step 1:** From the observed longitudinal response $y_i(t)$ reconstruct the covariate history for each subject
- Mixed effects model

$$\begin{aligned} y_i(t) &= m_i(t) + \epsilon_i(t) \\ &= x_i'(t)\beta + z_i'(t)b_i + \epsilon_i(t), \qquad \epsilon_i(t) \sim N(0, \sigma^2) \end{aligned}$$

- $x_i(t)$ and β: Fixed-effects part
- $z_i(t)$ and b_i: Random-effects part, $b_i \sim N(0, D)$
- $m_i(t)$ is the true unobserved value of the biomarker for the ith patient at time t

$$m_i(t) = x_i'(t)\beta + z_i'(t)b_i$$

Survival Submodel

- **Step 2:** Let's assume that we know $m_i(t)$, i.e., the *true and unobserved* value of the marker at time t
- Then, we can define a standard relative risk model

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma' w_i + \alpha m_i(t)\}$$

- $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ longitudinal history
- α quantifies the strength of the association between the marker and the risk of an event
- w_i baseline covariates

Linking the Submodels

- So how are the changes in the biomarker trajectory associated with survival?
- The true value of the longitudinal response is in the linear predictor of the survival submodel

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma' w_i + \alpha m_i(t)\}$$

- For example,

$$m_i(t) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t$$

- $\alpha m_i(t)$ is termed the current value parameterization

Joint Modeling Framework

- **Step 3:** The two processes are associated -> Define a model for their joint distribution
- Joint models belong to the class of **Shared Parameter Models**

$$p(y_i, T_i^*) = \int p(y_i|b_i)p(T_i^*|b_i)p(b_i)db_i$$

- The association between the longitudinal and survival processes is explained by the *shared* random effects b_i

Joint Modeling Framework

- Key assumption: **Full Conditional Independence** -> random effects explain all interdependencies
- Conditional on the random effects:
- The longitudinal outcome is independent of the time-to-event outcome
- The repeated measurements in the longitudinal outcome are independent of each other

$$p(y_i, T_i, \delta_i | b_i) = p(y_i | b_i) p(T_i, \delta_i | b_i)$$

$$p(y_i | b_i) = \prod_j p(y_{ij} | b_i)$$

- Caveat: Conditional Independence is difficult to test

Joint Modeling Framework

- The standard joint model

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma' w_i + \alpha m_i(t)\}$$
$$y_i(t) = m_i(t) + \epsilon_i(t)$$
$$= x_i'(t)\beta + z_i'(t)b_i + \epsilon_i(t)$$

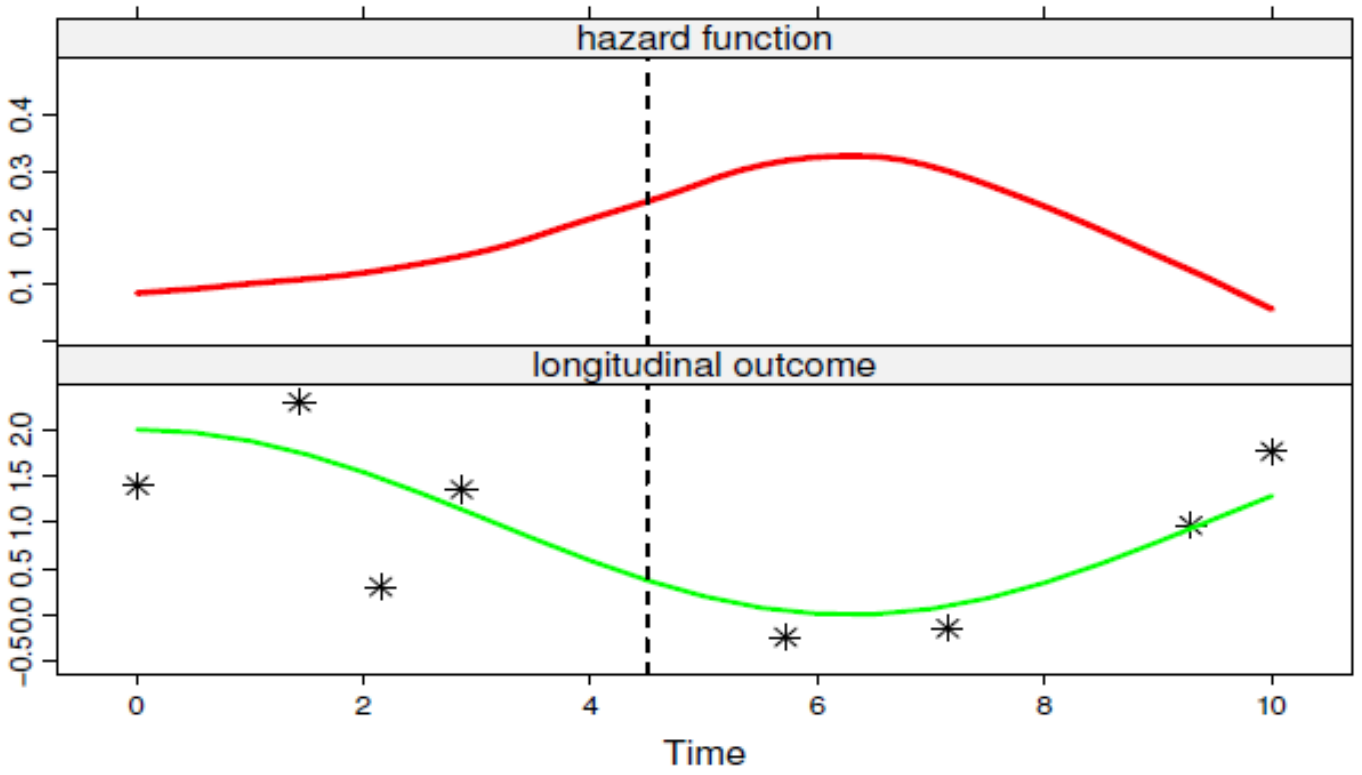
where $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$

- The joint distribution

$$p(y_i, T_i, \delta_i) = \int p(y_i|b_i)\{h(T_i|b_i)^{\delta_i} S(T_i|b_i)\}p(b_i)db_i$$

- $p(\cdot)$ density function; $S(\cdot)$ survival function

Parameterizations



Joint Modeling Framework

- The censoring and measurement processes are assumed non-informative
- Decision to withdraw from the study or appear at the next visit
  - May depend on observed past history (baseline covariates + observed longitudinal responses)
  - No additional dependence on underlying, latent subject characteristics associated with prognosis

Joint Modeling Framework

- The survival function, which is a part of the likelihood of the model, depends on the whole longitudinal history

$$S_i(t|b_i) = \exp\left(-\int_0^t h_0(s)\exp\{\gamma' w_i + \alpha m_i(s)\}ds\right)$$

- Therefore, care should be taken in the definition of the design matrices of the mixed model
- When subjects have nonlinear profiles -> Use splines or polynomials to model them flexibly

Joint Modeling Framework

- Random-effects distribution $p(b_i)$
- In mixed models, it is customary to assume normality
- However, in joint models this distribution plays a more prominent role because the random effects explain all associations
- Nevertheless, we have robustness, especially as the number of unique individuals increase *(Rizopoulos, Biometrika, 2008)*

Joint Modeling Framework

- Assumptions for the baseline hazard function $h_0(t)$
  - Parametric -> possibly restrictive
  - Unspecified -> within JM framework underestimates standard errors
- It is advisable to use parametric but flexible models for $h_0(t)$, e.g., splines

$$\log h_0(t) = \gamma_{h_0,0} + \sum_{q=1}^{Q} \gamma_{h_0,q} B_q(t, v)$$

- $B_q(t, v)$ denotes the q-th basis function of a B-spline with knots $v_1, .., v_Q$
- γ_{h_0} is a vector of spline coefficients

Joint Modeling Framework

- Step-functions for $h_0(t)$ often also work satisfactorily (piecewise-constant baseline hazard)

$$h_0(t) = \sum_{q=1}^{Q} \xi_q I(v_{q-1} < t \leq v_q)$$

- where $0 = v_0 < v_1 < \cdots < v_Q$ denotes a split of the time scale
- Balance bias and variance, and avoid overfitting!
- **Rules of thumb:**
  - Keep the number of parameters between 1/10-1/20 of the total number of events in your sample *(Harrell, 2001)*
  - Knots can be chosen based on percentiles of observed times or true event times

Estimation

- Mainly maximum likelihood but also Bayesian approaches
- Recall our assumptions:
  - The random effects account for the association between the longitudinal and event outcomes (conditional independence)
  - The random effects account for the correlation between repeated measurements of the longitudinal process (conditional independence)
  - Given the observed history, the censoring mechanism and the visiting process are noninformative

Estimation

- The log-likelihood contribution for subject i:

$$
\begin{aligned}
l_i(\theta) &= \log \int p(T_i, \delta_i, y_i, b_i; \theta) db_i \\
&= \log \int p(y_i|b_i;\theta) p(T_i, \delta_i|b_i;\theta) p(b_i;\theta) db_i \\
&= \log \int \left\{ \prod_{j=1}^{n_i} p(y_{ij}|b_i;\theta) \right\} \left\{ h(T_i|b_i;\theta)^{\delta_i} S_i(T_i|b_i;\theta) \right\} p(b_i;\theta) db_i
\end{aligned}
$$

Estimation

- The log-likelihood contribution for subject i:

$$l_i(\theta) = \log \int \left\{ \prod_{j=1}^{n_i} p(y_{ij}|b_i;\theta) \right\} \left\{ h(T_i|b_i;\theta)^{\delta_i} S_i(T_i|b_i;\theta) \right\} p(b_i;\theta) db_i$$

- Where we have our continuous longitudinal outcome

$$p(y_{ij}|b_i;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{[y_{ij} - m_i(t_{ij})]^2}{2\sigma^2} \right\}$$

Estimation

- The log-likelihood contribution for subject i:

$$l_i(\theta) = \log \int \left\{ \prod_{j=1}^{n_i} p(y_{ij}|b_i;\theta) \right\} \left\{ h(T_i|b_i;\theta)^{\delta_i} S_i(T_i|b_i;\theta) \right\} p(b_i;\theta) db_i$$

- Where we have our survival outcome

$$p(T_i, \delta_i|b_i;\theta) = [h_0(T_i)\exp(\alpha m_i(T_i) + \gamma' w_i)]^{\delta_i} \times \exp\left\{ -\int_0^{T_i} h_0(u)\exp(\alpha m_i(u) + \gamma' w_i) du \right\}$$

Estimation

- The log-likelihood contribution for subject i:

$$l_i(\theta) = \log \int \left\{ \prod_{j=1}^{n_i} p(y_{ij}|b_i;\theta) \right\} \left\{ h(T_i|b_i;\theta)^{\delta_i} S_i(T_i|b_i;\theta) \right\} p(b_i;\theta) db_i$$

- Where we have our multivariate normally distributed random effects

$$p(b_i;\theta) = (2\pi|D|)^{-q/2} \exp\left\{ -\frac{b_i' D^{-1} b_i}{2} \right\}$$

where q denotes the dimensionality of the random-effects vector

Estimation

$$
\begin{aligned}
l_i(\theta) &= \log \int \left\{ \prod_{j=1}^{n_i} p(y_{ij}|b_i;\theta) \right\} \left\{ h(T_i|b_i;\theta)^{\delta_i} S_i(T_i|b_i;\theta) \right\} p(b_i;\theta) db_i \\
&= \log \int \left\{ (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{[y_{ij} - m_i(t_{ij})]^2}{2\sigma^2} \right\} \right\} \\
&\times [h_0(T_i) \exp(\alpha m_i(T_i) + \gamma' w_i)]^{\delta_i} \\
&\times \exp\left\{ -\int_0^{T_i} h_0(u) \exp(\alpha m_i(u) + \gamma' w_i) du \right\} \\
&\times (2\pi|D|)^{-q/2} \exp\left\{ -\frac{b_i' D^{-1} b_i}{2} \right\} db_i
\end{aligned}
$$

Estimation

- The log-likelihood, in general, does not have a closed-form solution
- Integrals need to be approximated numerically
- Standard numerical integration algorithms:
  - Gaussian quadrature
  - Monte Carlo
- More difficult is the integral with respect to b_i because it can be of high dimension
  - Laplace approximations
  - Pseudo-adaptive Gaussian quadrature rules

Gauss-Hermite quadrature

- Numerical method to approximate analytically intractable integrals

$$\int_{-\infty}^{\infty} e^{-x^2} f(x) dx \approx \sum_{q=1}^{m} w_q f(x_q)$$

- where m is the number of sample points used
- Can be extended to multivariate integrals (i.e., multiple random effects)

Estimation

- To maximize the approximated log-likelihood

$$l(\theta) = \sum_{i=1}^{n} \log \int p(y_i|b_i;\theta)\{h(T_i|b_i;\theta)^{\delta_i} S_i(T_i|b_i;\theta)\}p(b_i;\theta)db_i$$

- We need to employ an optimization algorithm
- Standard choices:
  - EM (treating b_i as missing data)
  - Newton-type
  - Hybrids (start with EM and continue with quasi-Newton)

EM Algorithm

- Expectation-Maximization (EM): iterative algorithm for MLE when you have incomplete data
- Can sometimes frame problems to look like "missing data" and use this for estimation
- **Intuition:** The complete data log-likelihood (if b_i were observed) often has a closed form and is simpler to maximize
- Two steps: (E) Expectation, (M) Maximization
- **E-step:** We fill in the missing data and replace the log-likelihood of the observed data with a "surrogate function"
- **M-step:** We maximize this "surrogate function"
- Iterate because the E-step might be slightly wrong if the parameters are not already at the MLEs

EM Algorithm

- Our complete data vector $Y = (Y^0, Y^m)$
- Goal: Estimate the parameter θ in the complete data log-likelihood, using only the observed information
- E-step: Compute the expected value of the complete data log-likelihood

$$Q(\theta|\theta^{(k)}) = E\{\log p(y;\theta)|y^o;\theta^{(k)}\}$$
$$= \int \log p(y^m, y^o;\theta) p(y^m|y^o;\theta^{(k)} dy^m) dy^m$$

- M-step: Update the parameters

$$\theta^{(k+1)} = \arg\max_\theta Q(\theta|\theta^{(k)})$$

EM Algorithm

- We can use the EM algorithm to derive the MLEs of the standard joint model

$$h_i(t) = h_0(t)\exp[\gamma' w_i + \alpha\{x_i(t)'\beta + z_i(t)'b_i\}]$$
$$y_i(t) = x_i'(t)\beta + z_i'(t)b_i + \epsilon_i(t)$$
$$b_i \sim N(0, D), \epsilon_i(t) \sim N(0, \sigma^2)$$

- The parameters that we are then interested in estimating are $\theta = (\theta_t', \theta_y', \theta_b')'$ where

$$\theta_y = (\beta', \sigma^2)'$$
$$\theta_t = (\gamma', \alpha, \theta_{h_0}')'$$
$$\theta_b = \text{vech}(D)$$

E-Step

- We treat the random effects as “missing data”
- Goal: Find $\hat{\theta}$ that maximizes the observed data log-likelihood

$$l(\theta) = \sum_i \log p(T_i, \delta_i, y_i; \theta)$$

- Instead, maximize the expected value of the complete data log-likelihood

$$\begin{aligned} Q(\theta|\theta^{(k)}) &= \sum_i \int \log p(T_i, \delta_i, y_i, b_i; \theta) p(b_i|T_i, \delta_i, y_i; \theta^{(k)}) db_i \\ &= \sum_i \int \{\log p(T_i, \delta_i|b_i; \theta_t, \beta) + \log p(y_i|b_i; \theta_y) \\ &\qquad + \log p(b_i; \theta_b)\} p(b_i|T_i, \delta_i, y_i; \theta^{(k)}) db_i \end{aligned}$$

- Need to use numerical integration procedures

M-Step

- The complete data log-likelihood is split into three parts

$$\log p(T_i, \delta_i, y_i, b_i; \theta) = \log p(T_i, \delta_i | b_i; \theta_t, \beta) + \log p(y_i | b_i; \theta_y) + \log p(b_i; \theta_b)$$

- Maximization of $Q(\theta|\theta^{(k)})$ with respect to θ only involves the parts in which the respective parameter appears

M-Step

- For the measurement error variance in the longitudinal measurement model and the covariance matrix of the random effects we have closed-form expressions

$$\begin{aligned}
\hat{\sigma}^2 &= N^{-1} \sum_i \int (y_i - X_i\beta - Z_i b_i)^\top (y_i - X_i\beta - Z_i b_i) p(b_i \mid T_i, \delta_i, y_i; \theta)\, db_i \\
&= N^{-1} \sum_i (y_i - X_i\beta)^\top (y_i - X_i\beta - 2Z_i\tilde{b}_i) + \mathrm{tr}(Z_i^\top Z_i \widetilde{vb}_i) + \tilde{b}_i^\top Z_i^\top Z_i \tilde{b}_i, \\
\widehat{D} &= n^{-1} \sum_i \widetilde{vb}_i + \tilde{b}_i \tilde{b}_i^\top,
\end{aligned}$$

where $N = \sum_i n_i$, $\tilde{b}_i = E(b_i \mid T_i, \delta_i, y_i; \theta^{(it)}) = \int b_i p(b_i \mid T_i, \delta_i, y_i; \theta^{(it)}) db_i$,

and $\widetilde{vb}_i = \mathrm{var}(b_i \mid T_i, \delta_i, y_i; \theta^{(it)}) = \int (b_i - \tilde{b}_i)^2 p(b_i \mid T_i, \delta_i, y_i; \theta^{(it)}) db_i$.

M-Step

- We don't have closed-form solutions for the score equations for the fixed effect β and the parameters of the survival submodel θ_t
- So we use a one-step Newton-Raphson update:

$$\hat{\beta}^{(k+1)} = \hat{\beta} - \{\partial S(\hat{b}^{(k)})/\partial\beta\}^{-1} S(\hat{\beta}^{(k)})$$

$$\hat{\theta}_t^{(k+1)} = \hat{\theta}_t^{(k)} - \{\partial S(\hat{\theta}_t^{(k)})/\partial\theta_t\}^{-1} S(\hat{\theta}_t^{(k)})$$

Score functions

$$
\begin{aligned}
S(\beta) &= \sum_i X_i^\top \{y_i - X_i\beta - Z_i\tilde{b}_i\}/\sigma^2 + \alpha\delta_i x_i(T_i) \\
&\quad - \exp(\gamma^\top w_i) \int \int_0^{T_i} h_0(s)\alpha x_i(s) \exp\big[\alpha\{x_i^\top(s)\beta + z_i^\top(s)b_i\}\big] \\
&\qquad \times p(b_i \mid T_i, \delta_i, y_i; \theta)\, ds\, db_i, \\
S(\gamma) &= \sum_i w_i \Bigg[\delta_i - \exp(\gamma^\top w_i) \int \int_0^{T_i} h_0(s) \exp\big[\alpha\{x_i^\top(s)\beta + z_i^\top(s)b_i\}\big] \\
&\qquad \times p(b_i \mid T_i, \delta_i, y_i; \theta)\, ds\, db_i\Bigg],
\end{aligned}
$$

Score functions

$$
\begin{aligned}
S(\alpha) &= \sum_i \delta_i \{x_i^\top(T_i)\beta + z_i^\top(T_i)\tilde{b}_i\} \\
&\quad - \exp(\gamma^\top w_i) \int \int_0^{T_i} h_0(s) \exp\big[\alpha\{x_i^\top(s)\beta + z_i^\top(s) b_i\}\big] \\
&\quad \times p(b_i \mid T_i, \delta_i, y_i; \theta)\, ds\, db_i,
\end{aligned}
$$

$$
\begin{aligned}
S(\theta_{h_0}) &= \sum_i \delta_i \frac{\partial h_0(T_i; \theta_{h_0})}{\partial \theta_{h_0}^\top} \\
&\quad - \exp(\gamma^\top w_i) \int \int_0^{T_i} \frac{\partial h_0(s; \theta_{h_0})}{\partial \theta_{h_0}^\top} \exp\big[\alpha\{x_i^\top(s)\beta + z_i^\top(s) b_i\}\big] \\
&\quad \times p(b_i \mid T_i, \delta_i, y_i; \theta)\, ds\, db_i.
\end{aligned}
$$

Estimation

- Standard errors: Standard asymptotic MLE

$$\hat{\text{var}}(\hat{\theta}) = \left\{ -\sum_{i=1}^{n} \frac{\partial^2 \log p(y_i, T_i, \delta_i; \theta)}{\partial \theta' \partial \theta}|_{\theta=\hat{\theta}} \right\}^{-1}$$

- That is, standard errors for the parameter estimates can be based on the estimated observed information matrix

$$\hat{\text{var}}(\hat{\theta}) = \{\mathcal{I}(\hat{\theta})\}^{-1}, \text{ with } \mathcal{I}(\hat{\theta}) = -\sum_{i=1}^{n} \frac{\partial S_i(\theta)}{\partial \theta}|_{\theta=\hat{\theta}}$$

Estimation

- Standard errors: Standard asymptotic MLE

$$\hat{\text{var}}(\hat{\theta}) = \left\{ -\sum_{i=1}^{n} \frac{\partial^2 \log p(y_i, T_i, \delta_i; \theta)}{\partial \theta' \partial \theta}|_{\theta=\hat{\theta}} \right\}^{-1}$$

- Standard asymptotic tests + information criteria
  - Likelihood ratio test
  - Score test
  - Wald test
  - AIC, BIC, …

Estimation

- Based on a fitted joint model, estimates for the random effects are based on the posterior distribution

$$p(b_i|T_i,\delta_i,y_i;\theta) = \frac{p(T_i,\delta_i|b_i,\theta)p(y_i|b_i;\theta)p(b_i;\theta)}{p(T_i,\delta_i,y_i;\theta)}$$
$$\propto p(T_i,\delta_i|b_i;\theta)p(y_i|b_i;\theta)p(b_i;\theta)$$

in which θ is replaced by its MLE $\hat{\theta}$

Recap: Joint Modeling Framework

- The standard joint model

$$h_i(t|\mathcal{M}_i(t)) = h_0(t)\exp\{\gamma' w_i + \alpha m_i(t)\}$$
$$y_i(t) = m_i(t) + \epsilon_i(t)$$
$$= x_i'(t)\beta + z_i'(t)b_i + \epsilon_i(t)$$

where $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$

- The joint distribution

$$p(y_i, T_i, \delta_i) = \int p(y_i|b_i)\{h(T_i|b_i)^{\delta_i} S(T_i|b_i)\}p(b_i)db_i$$

- $p(\cdot)$ density function; $S(\cdot)$ survival function

Recap: Joint Modeling Framework

- Joint modeling provides us with a method of linking a longitudinal outcome measured with error (endogenous) with a time to event outcome
- It has been shown to reduce bias and maximize efficiency compared to naïve approaches
- Failing to account for the longitudinal process causes bias in covariate effects on survival when there is a true association between outcomes
- To come:
  - Extensions to joint models
  - Joint models and missing data
  - Dynamic risk prediction with joint models