1 Notation and Assumptions

Suppose there are n_s individuals. For individual i, let C_i be the true class, $\mathbf{S}_i$ be the subject-level covariates predicting class membership, $\mathbf{T}_i$ the observation times, and $\mathbf{X}_{ij}$ and $\mathbf{Y}_{ij}$ be the observed covariates and response at time j. Let m_i be the number of measurements. Let

$$u_{ic} = 1\{C_i = c\}$$

.

We assume that C_i is categorical with possible values $1, ..., n_c$, and category probabilities proportional to $\exp(\mathbf{s}_i\gamma_c)$. One of the classes, R, is designated as the reference class so $\gamma_R = 0$ for identifiability. We assume that $\mathbf{y}_i | (\mathbf{T}_i = \mathbf{t}_i, \mathbf{X}_i = \mathbf{x}_i, C_i = c)$ is multivariate normal with mean $\boldsymbol{\mu}_{ic} = \mathbf{x}_i\boldsymbol{\theta}_c$ and $m_i \times m_i$ covariance matrix $\mathbf{V}_{i|c} = \mathbf{V}(\mathbf{t}_i, \rho, \sigma^2_{s|c}, \sigma^2_{e|c})$.

Suppose the observation times for all subjects are scaled to lie in an interval between $t_{\min}$ and $t_{\max}$. We are assuming $\mathbf{V}(\mathbf{t}_i, \rho, \sigma^2_s, \sigma^2_e)$ is a $m_i \times m_i$ matrix such that

$$\begin{aligned}
\left[\mathbf{V}_{i|c}\right]_{jk} =& \mathrm{Cov}\left(Y_{ij}, Y_{ik} | \left(\mathbf{T}_i = \mathbf{t}_i, \mathbf{X}_i = \mathbf{x}_i, Z_i = z\right)\right) \\
=& \sigma^2_e\delta_{ij} + \sigma^2_s \exp\left(\frac{|t_{ij} - t_{ik}|}{|t_{\max} - t_{\min}|}\log\rho\right).
\end{aligned}$$

where $\delta_{ij} = 1\{i = j\}$, and where ρ and the "proportion nugget" $\sigma^2_{e|c} / \left(\sigma^2_{e|c} + \sigma^2_{s|c}\right)$ are constant across classes but the total variance $\sigma^2_{e|c} + \sigma^2_{s|c}$ may differ across classes.

2 Model Log-Likelihood

Let $\delta(c_i, k) = 1\{c_i = c\}$. Then the log-likelihood contribution for individual i is

$$\begin{aligned}
\mathcal{L}_i =& \log\left(\Pr\left(c_i\right)\right) + \log\left(\Pr\left(y_i | c_i\right)\right). \\
=& \sum_{c=1}^{n_c} u_{ic}\log\left(\pi_{ic}\right) - \frac{1}{2}m_i\log\left(2\pi\right) - \frac{1}{2}\sum_{c=1}^{n_c} u_{ic}\log\det\mathbf{V}(\mathbf{t}_i) \\
& - \frac{1}{2}\sum_{c=1}^{n_c} u_{ic}R_{ic},
\end{aligned}$$

where

$$\begin{aligned}
\pi_{ic} =& \Pr\left(c_i\right) \\
=& \frac{\exp\left(\mathbf{s}_i\gamma_c\right)}{\sum_{k=1}^{n_c}\exp\left(\mathbf{s}_i\gamma_k\right)} \\
=& \frac{\exp\left(\mathbf{s}_i\gamma_c\right)}{1 + \sum_{k\neq R}\exp\left(\mathbf{s}_i\gamma_k\right)}
\end{aligned}$$

and where

$$R_{ic} = (\mathbf{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_c)^T \mathbf{V}_{i|c}^{-1} (\mathbf{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_c).$$

During the EM algorithm, we replace the random variables u_{ic} with the estimated posterior probabilities η_{ic}.

3 Plan for Covariance Estimation

The most straightforward covariance estimate would be

$$\hat{Cov}_1(\theta) = I_1^{-1}$$

where I_1 is the usual information matrix based on the negative Hessian of the log-likelihood. But this ignores uncertainty about latent class membership. From Louis (1982) and Turner (2000), we have

$$\hat{Cov}_2(\theta) = (I_1 - I_2)^{-1}$$

where I_2 is a measure of information loss based on cross-products of first derivatives. But this ignores clustering. The Huber-White, robust, or sandwich formula, which ignores latent class uncertainty but handles clustering, is

$$\hat{Cov}_3(\theta) = I_1^{-1} EC(\theta) I_1^{-1}$$

where $EC(\theta)$ is the empirical covariance matrix, i.e., sum of cross-products of individual score functions. So we propose the ad hoc combination

$$\hat{Cov}_4(\theta) = (I_1 - I_2)^{-1} EC(\theta) (I_1 - I_2)^{-1}$$

which should be greater than or equal to $\hat{Cov}_3(\theta)$.

Based loosely on the order of indexing in the R mixreg package, we will follow this convention: one class at a time, linear part before logistic part. That is, if we have two classes, the parameter vector is $\boldsymbol{\beta}_1$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_2$.

4 Constructing the I_1 Matrix

4.1 First derivatives in linear model coefficients

First of all,

$$\frac{\delta \mathcal{L}_i}{\delta \boldsymbol{\beta}_c} = u_{ic} \boldsymbol{X}_i^T \mathbf{V}_{i|c}^{-1} (\mathbf{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_c).$$

For convenience let $\mathbf{b}_{ic} = \boldsymbol{X}_i^T \mathbf{V}_{i|c}^{-1} (\mathbf{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_c)$ analogously to Turner, although Turner did not have the weights in there.

4.2 Second derivatives in linear model coefficients

Next,

$$\frac{\delta^2 \mathcal{L}_i}{\delta \boldsymbol{\beta}_c \delta \boldsymbol{\beta}_c^T} = u_{ic} \frac{\delta^2 R_{ic}}{\delta \boldsymbol{\beta}_c \delta \boldsymbol{\beta}_c^T}$$

where

$$\frac{\delta^2 R_{ic}}{\delta \boldsymbol{\beta}_c \delta \boldsymbol{\beta}_c^T} = \boldsymbol{X}_i^T \mathbf{V}_{i|c}^{-1} \boldsymbol{X}_i.$$

The cross-derivatives across classes are zero because the $\boldsymbol{\beta}$ vectors for each class are not constrained by each other and are treated as independent. Thus,

$$\frac{\delta^2 \mathcal{L}_i}{\delta \boldsymbol{\beta}_c \delta \boldsymbol{\beta}_k^T} = u_{ic} \delta_{ck} \frac{\delta^2 R_{ic}}{\delta \boldsymbol{\beta}_c \delta \boldsymbol{\beta}_c^T}$$

where $\delta_{ck} = 1\{c = k\}$.

In practice, replacing the unknown random u by the estimated posterior probability ω, we use

$$\mathcal{E} \frac{\delta^2 \mathcal{L}_i}{\delta \boldsymbol{\beta}_c \delta \boldsymbol{\beta}_k^T} = \omega_{ic} \delta_{ck} \frac{\delta^2 R_{ic}}{\delta \boldsymbol{\beta}_c \delta \boldsymbol{\beta}_c^T}$$

4.3 First derivatives in logistic model coefficients

For $c \neq R$,

$$\begin{aligned} \frac{\delta \mathcal{L}_i}{\delta \gamma_c} &= \frac{\delta}{\delta \gamma_c} \sum_{k=1}^{n_c} u_{ik} \log(\pi_{ik}) \\ &= \sum_{k=1}^{n_c} \frac{u_{ik}}{\pi_{ik}} \left(\frac{\delta \pi_{ik}}{\delta \gamma_c} \right) \end{aligned}$$

where

$$\begin{aligned} \frac{\delta \pi_{ik}}{\delta \gamma_c} &= \begin{cases} \pi_{ik}(1 - \pi_{ik})\mathbf{s}_i, & k = c \\ -\pi_{ic}\pi_{ik}\mathbf{s}_i, & k \neq c \end{cases} \\ &= \pi_{ik}(\delta_{ck} - \pi_{ic})\mathbf{s}_i \end{aligned}$$

where $\delta_{ck} = 1\{c = k\}$ (as in, e.g., Linzer & Lewis, 2011). Thus,

$$\frac{\delta \mathcal{L}_i}{\delta \gamma_c} = \sum_{k=1}^{n_c} u_{ik}(\delta_{ck} - \pi_{ic})\mathbf{s}_i.$$

After some algebra, using the fact that the u_{ik} sum to 1 over k, this becomes simply

$$\frac{\delta \mathcal{L}_i}{\delta \gamma_c} = (u_{ic} - \pi_{ic})\mathbf{s}_i.$$

4.4 Second derivatives in logistic model coefficients

By the above, for $c \neq R$ and $c' \neq R$, regardless of whether $c = c'$, we get

$$\frac{\delta^2 \mathcal{L}_i}{\delta \gamma_c \delta \gamma_{c'}^T} = -\frac{\partial \pi_{ic}}{\partial \gamma_{c'}} \mathbf{s}_i^T.$$

So

$$\mathcal{E} \frac{\delta^2 \mathcal{L}_i}{\delta \gamma_c \delta \gamma_{c'}^T} = -\pi_{ic} \left(\delta_{cc'} - \pi_{ic'} \right) \mathbf{s}_i \mathbf{s}_i^T.$$

4.5 Cross-derivatives in linear and logistic model coefficients

We will treat these as zero because $\boldsymbol{\beta}$ does not appear in the expression for $\delta \mathcal{L}_i / \delta \gamma_c$.

5 Cross-products of first derivatives (the I_2 matrix)

Turner (2000) defines the $n_s \times n_s$ matrix G_{ck} for each combination of $c = 1, ..., n_c$ and $k = 1, ..., n_c$ as the matrix whose (i, j)th entry is $\mathcal{E}\left(u_{ic} u_{jk}\right)$ where $\mathcal{E}$ denotes "the operation of taking the expected value with respect to the missing data, given the observed data" (Turner 2000, p. 376). The missing data, here, are the true class memberships where the observed data are the responses and covariates. Thus,

$$\left(G_{ck}\right)_{ij} = \mathcal{E}\left(u_{ic} u_{jk}\right) = \begin{cases} \omega_{ic} \omega_{jk} & i \neq j \\ \omega_{ik} & i = j, c = k \\ 0 & i = j, c \neq k \end{cases}$$

where ω_{ic} is the posterior probability that $C_i = c$ given the data.

The first holds by independence of different observations (so it will require modification in the clusters case). The second holds because the indicator functions $\underline{u}$ can only take on the idempotent values 0 or 1. The third holds because an observation can only belong to one class at a time.

5.1 Cross-products of derivatives in $\boldsymbol{\beta}$ and $\boldsymbol{\beta}$

Regardless of whether c equals k,

$$
\begin{aligned}
\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}_c}\right)\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}_k}\right)^T &= \left(\sum_{i=1}^{n_s} u_{ic}\mathbf{b}_{ic}\right)\left(\sum_{j=1}^{n_s} u_{jk}\mathbf{b}_{jk}\right)^T \\
&= \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} u_{ic}u_{jk}\mathbf{b}_{ic}\mathbf{b}_{jk}^T
\end{aligned}
$$

So

$$
\mathcal{E}\left(\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}_c}\right)\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}_k}\right)^T\right) = \sum_{i=1}^{n_s}\sum_{j=1}^{n_s} \left(G_{ck}\right)_{ij} \mathbf{b}_{ic}\mathbf{b}_{jk}^T
$$

5.2 Cross-products of derivatives in γ and γ

Regardless of whether c equals k,

$$
\begin{aligned}
\left(\frac{\delta \mathcal{L}}{\delta \gamma_c}\right)\left(\frac{\delta \mathcal{L}}{\delta \gamma_k}\right)^T &= \left(\sum_{i=1}^{n_s}(u_{ic}-\pi_{ic})\mathbf{s}_i\right)\left(\sum_{j=1}^{n_s}(u_{jk}-\pi_{jk})\mathbf{s}_j\right) \\
&= \sum_{i=1}^{n_s}\sum_{j=1}^{n_s}\left(u_{ic}-\pi_{ic}\right)\left(u_{jk}-\pi_{jk}\right)\mathbf{s}_i\mathbf{s}_j^T
\end{aligned}
$$

So

$$
\begin{aligned}
\mathcal{E}\left(\frac{\delta \mathcal{L}}{\delta \gamma_c}\right)\left(\frac{\delta \mathcal{L}}{\delta \gamma_k}\right)^T &= \sum_{i=1}^{n_s}\sum_{j=1}^{n_s}\mathcal{E}\left(\left(u_{ic}-\pi_{ic}\right)\left(u_{jk}-\pi_{jk}\right)\right)\mathbf{s}_i\mathbf{s}_j^T \\
&= \sum_{i=1}^{n_s}\sum_{j=1}^{n_s}\left(\left(G_{ck}\right)_{ij}-\omega_{ic}\pi_{jk}-\omega_{jk}\pi_{ic}+\pi_{ic}\pi_{jk}\right)\mathbf{s}_i\mathbf{s}_j^T
\end{aligned}
$$

5.3 Cross-products of derivatives in $\boldsymbol{\beta}$ and γ

Regardless of whether c equals k,

$$
\begin{aligned}
\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}_c}\right)\left(\frac{\partial \mathcal{L}}{\partial \gamma_k}\right)^T &= \left(\sum_{i=1}^{n_s} u_{ic}\mathbf{b}_{ic}\right)\left(\sum_{j=1}^{n_s}\sum_{k'=1}^{n_c} u_{jk'}(\delta_{kk'}-\pi_{jk})\mathbf{s}_j\right)^T \\
&= \sum_{i=1}^{n_s}\sum_{j=1}^{n_s}\sum_{k'=1}^{n_c} u_{ic}u_{jk'}(\delta_{kk'}-\pi_{jk})\mathbf{b}_{ic}\mathbf{s}_j^T
\end{aligned}
$$

So

$$\mathcal{E}\left(\frac{\delta\mathcal{L}}{\delta\gamma_c}\right)\left(\frac{\delta\mathcal{L}}{\delta\gamma_k}\right)^T = \sum_{i=1}^{n_s}\sum_{j=1}^{n_s}\sum_{k'=1}^{n_c}\left(G_{ck'}\right)_{ij}\left(\delta_{kk'} - \pi_{jk}\right)\mathbf{b}_{ic}\mathbf{s}_j^T.$$