Jiaqi's Thesis Progress Report (Updated Jan. 28)

Jiaqi Bi[a]

[a] *Western University, Schulich School of Medicine & Dentistry, Department of Epidemiology and Biostatistics*

1. To Do List

1. Correlated frailty - NR-algorithm
2. Gibb's Sampling in MCEM & MI, partition the **X**
3. Multiple imputation - similar imputation step as MCEM

2. Weibull Parametric Approach and MCEM Method

From the beginning of the discussion, I have obtained the model, i.e., the hazard function is

$$h_{ij}(t_{ij}|z_j) = h_0(t_{ij})\exp(\beta_1 x_{1,ij} + \beta_2 x_{2,ij})z_j \tag{1}$$

There are total n_j individuals in family j, where $i = 1, ..., n_j$, and total J families that $j = 1, ..., J$. $x_{1,ij}$ is the genotype, or say mutation gene status for individual i in family j. $x_{2,ij}$ is the PRS for individual i in family j. The frailty term z_j, has a pdf of $f(z)$, which can be Gamma, log-normal, or other common frailty distributions. The support of $f(z)$ is always non-negative. The Weibull baseline hazard function is defined as

$$h_0(t_{ij}) = \alpha\lambda t_{ij}^{\lambda-1} \tag{2}$$

where λ is the shape parameter and α is the scale parameter. Let $\xi_{ij} = \exp(\beta_1 x_{1,ij} + \beta_2 x_{2,ij})$, the hazard function is

$$h_{ij}(t_{ij}|x_{ij}, g_{ij}, z_j) = \alpha\lambda t_{ij}^{\lambda-1}\xi_{ij}z_j \tag{3}$$

The survival function $S(t)$ can be obtained through cumulative hazard function $H(t)$

$$H(t_{ij}|x_{ij}, g_{ij}, z_j) = \int_0^t h_{ij}(u|x_{ij}, g_{ij}, z_j)du \tag{4}$$

$$= \alpha\xi_{ij}z_j\lambda\int_0^t u^{\lambda-1}du \tag{5}$$

$$= \alpha\xi_{ij}z_j\lambda \cdot \frac{1}{\lambda}t_{ij}^{\lambda} = \alpha\xi_{ij}z_j t_{ij}^{\lambda} \tag{6}$$

Email address: `jbi23@uwo.ca` (Jiaqi Bi)

and the survival function

$$S(t_{ij}|x_{ij}, g_{ij}, z_j) = \exp(-H(t_{ij}|x_{ij}, g_{ij}, z_j)) = \exp(-\alpha\xi_{ij}z_j t_{ij}^{\lambda}) \tag{7}$$

Let $\boldsymbol{\theta} = \{\beta_1, \beta_2, \alpha, \lambda, \boldsymbol{\phi}\}$, where $\boldsymbol{\phi}$ is the parameter vector for the frailty distribution of the choice. Therefore, the likelihood can be written as

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{J} \int_0^{\infty} \prod_{i=1}^{n_j} (\alpha\lambda t_{ij}^{\lambda-1}\xi_{ij}z_j)^{\delta_{ij}} \exp(-\alpha\xi_{ij}z_j t_{ij}^{\lambda}) f(z) dz \tag{8}$$

$$= \prod_{j=1}^{J} \int_0^{\infty} \prod_{i=1}^{n_j} h(t_{ij}|\mathbf{x}_{ij}, z_j)^{\delta_{ij}} \exp(-H(t_{ij}|\mathbf{x}_{ij}, z_j)) f(z) dz \tag{9}$$

So the log-likelihood is

$$\ell(\boldsymbol{\theta}) = \sum_{j=1}^{J} \log\left[\int_0^{\infty} \prod_{i=1}^{n_j} h(t_{ij}|\mathbf{x}_{ij}, z_j)^{\delta_{ij}} \exp(-H(t_{ij}|\mathbf{x}_{ij}, z_j)) f(z) dz\right] \tag{10}$$

3. Gamma Frailty

The Laplace transform of the frailty $z \sim \text{Gamma}(k, k)$, for the simplicity of the mathematical expression, the following Laplace transform will ignore the subscript, denote $\mathscr{L}(f(z)) = \phi(s)$ where $s = \sum_{i=1}^{n_j} H(t_{ij}|\mathbf{x}_{ij})$:

$$\phi(s) = \int_0^{\infty} e^{-sz} f(z) dz \tag{11}$$

$$= \int_0^{\infty} e^{-sz} \frac{k^k}{\Gamma(v)} z^{k-1} e^{-kz} dz \tag{12}$$

Using the Gamma property: $\int_0^{\infty} z^{n-1}e^{-az}dz = \frac{\Gamma(n)}{a^n}$, $\phi(s)$ can be further written as

$$\phi(s) = \frac{k^k}{\Gamma(k)} \int_0^{\infty} e^{-(s+k)z} z^{k-1} dz = \frac{k^k}{\Gamma(k)} \cdot \frac{\Gamma(k)}{(s+k)^k} = (1 + \frac{s}{k})^{-k} \tag{13}$$

The second derivative is $\frac{d^2\phi(s)}{ds^2} = \int_0^{\infty} (-z)^2 e^{-sz} f(z) dz$.
The third derivative is $\frac{d^3\phi(s)}{ds^3} = \int_0^{\infty} (-z)^3 e^{-sz} f(z) dz$, ... Therefore, its d-th derivative, denote $\phi(s)^{(d)}$:

$$\phi(s)^{(d)} = (-1)^d \int_0^{\infty} z^d e^{-sz} f(z) dz \tag{14}$$

$$= (-1)^d \frac{(k+d-1)!}{(k-1)!(s+k)^d} (1 + \frac{s}{k})^{-k} \tag{15}$$

Let $\boldsymbol{\theta} = (\beta_1, \beta_2, \alpha, \lambda, k)$ for Gamma frailty model, the log-likelihood is then written as

$$
\begin{aligned}
\ell(\boldsymbol{\theta}) &= \sum_{j=1}^{k} \log\left[\int_0^\infty \prod_{i=1}^{n_j} (h(t_{ij}|\mathbf{x}_{ij}, z_j))^{\delta_{ij}} \exp(-H(t_{ij}|\mathbf{x}_{ij}, z_j)) f(z_j) dz_j\right] && (16)\\
&= \sum_{j=1}^{J} \log\left[\int_0^\infty \prod_{i=1}^{n_j} (z_j h(t_{ij}|\mathbf{x}_{ij}))^{\delta_{ij}} \exp(-z_j H(t_{ij}|\mathbf{x}_{ij})) f(z_j) dz_j\right] && (17)\\
&= \sum_{j=1}^{J} \log\left[\prod_{i=1}^{n_j} (h(t_{ij}|\mathbf{x}_{ij}))^{\delta_{ij}} \int_0^\infty z_j^{d_j} \exp(-z_j \sum_{i=1}^{n_j} H(t_{ij}|\mathbf{x}_{ij})) f(z_j) dz_j\right] && (18)\\
&= \sum_{j=1}^{J} \log\left[\prod_{i=1}^{n_j} (h(t_{ij}|\mathbf{x}_{ij}))^{\delta_{ij}} \frac{(k+d_j-1)!}{(k-1)!(\sum_{i=1}^{n_j} H(t_{ij}|\mathbf{x}_{ij}) + k)^{d_j}} \Big(1 + \frac{\sum_{i=1}^{n_j} H(t_{ij}|\mathbf{x}_{ij})}{k}\Big)^{-k}\right] && (19)\\
&= \sum_{j=1}^{J} \log\left[\prod_{i=1}^{n_j} ((h(t_{ij}|\mathbf{x}_{ij}))^{\delta_{ij}}) \frac{(k+d_j-1)!}{k! k^{d_j-1}} (1 + \frac{\sum_{i=1}^{n_j} (H(t_{ij}|\mathbf{x}_{ij}))}{k})^{-k-d_j}\right] && (20)\\
&= \sum_{j=1}^{J} \log\left[h(t_{ij}|\mathbf{x}_{ij})^{\delta_{ij}} \frac{(k+d_j-1)!}{k! k^{d_j-1}} (1 + \frac{\sum_{i=1}^{n_j} (H(t_{ij}|\mathbf{x}_{ij}))}{k})^{-k-d_j}\right] && (21)\\
&= \sum_{j=1}^{J} \left[\sum_{i=1}^{n_j} (\delta_{ij} \log h(t_{ij}|\mathbf{x}_{ij})) + \log\Big(\frac{(k+d_j-1)!}{k! k^{d_j-1}} (1 + \frac{\sum_{i=1}^{n_j} (H(t_{ij}|\mathbf{x}_{ij}))}{k})^{-k-d_j}\Big)\right] && (22)
\end{aligned}
$$

For each family j, the ascertainment A_j is defined to be the probability of the proband p being ascertained by the age a_{j_p} at examination. Applying the ascertainment correction for the log-likelihood in family j:

$$\tilde{\ell}_j(\boldsymbol{\theta}) = \ell_j(\boldsymbol{\theta}) - \log A_j(\boldsymbol{\theta}) \tag{23}$$

where $\tilde{\ell}$ is the log-likelihood with ascertainment correction, and ℓ is the crude log-likelihood. Define $\mathbf{x}_{j_p}$ the covariate matrix for proband in family j. Note we can still apply Laplace transform here, such that

$$
\begin{aligned}
A_j(\boldsymbol{\theta}) &= 1 - S_{j_p}(a_{j_p}|\mathbf{x}_{j_p}) && (24)\\
&= 1 - \int_0^\infty S_{j_p}(a_{j_p}|\mathbf{x}_{j_p}, z_j) f(z_j) dz_j && (25)\\
&= 1 - \int_0^\infty \exp(-z_j \cdot H_{j_p}(a_{j_p}|\mathbf{x}_{j_p})) f(z_j) dz_j && (26)\\
&= 1 - (1 + \frac{H_{j_p}(a_{j_p}|\mathbf{x}_{j_p})}{k})^{-k} && (27)
\end{aligned}
$$

4. Log-Normal Frailty

The log-normal frailty is not the power-variance-function (PVF) family, so there is no closed form for Laplace transform or expressions for survivors. But we are able to estimate the Laplace transform using Gauss Hermite Quadrature. We typically standardize the log-normal frailty Z as

$$E(\log Z) = 0 \tag{28}$$

$$\mathrm{Var}(\log Z) = \sigma^2 \tag{29}$$

That is, $z \sim$ log-Normal$(0, \sigma^2)$. The probability density function $f(z)$ is then

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} z^{-1} \exp(-\frac{\log(z)^2}{2\sigma^2}) \tag{30}$$

The Laplace transform is then

$$\phi(s) = \mathscr{L}(f_Z)(s) = \int_0^\infty \exp(-sz) \cdot f(z) dz \tag{31}$$

Using variable transformation, let $y = \frac{\log(z)}{\sqrt{2}\sigma}$, then $z = \exp(\sqrt{2}\sigma y)$, and $dz = \sqrt{2}\sigma \exp(\sqrt{2}\sigma y) dy$. Therefore, for d-th derivative:

$$\phi(s)^d = \int_{-\infty}^\infty z^d \exp(-sz) \cdot \frac{1}{\exp(\sqrt{2}\sigma y)\sigma\sqrt{2\pi}} \cdot \exp(-y^2) \cdot \sqrt{2}\sigma \exp(\sqrt{2}\sigma y) dy \tag{32}$$

$$= \int_{-\infty}^\infty \exp(\sqrt{2}\sigma y)^d \exp(-s \exp(\sqrt{2}\sigma y)) \cdot \frac{1}{\sqrt{\pi}} \exp(-y^2) dy \tag{33}$$

Definition 1 (Gauss-Hermite Quadrature)**.** *The integrand part can be solved using Gauss-Hermite Quadrature. In numerical analysis, the method can be applied in the following form:*

$$\int_{-\infty}^\infty \exp(-x^2) f(x) dx \approx \sum_{i=1}^n \omega_i f(x_i) \tag{34}$$

where n is number of sample points used, and x_i is the roots of Hermite polnomial $H_n(x)$ such that $i = 1, ..., n$, and the weights ω_i is

$$\omega_i = \frac{2^{n-1} n! \sqrt{n}}{n^2 [H_{n-1}(x_i)]^2} \tag{35}$$

Applying Definition 1, the integral of the Laplace transform is then

$$\phi(s)^d = \frac{1}{\sqrt{\pi}} \sum_{\tilde{p}=1}^{N_{\tilde{p}}} \omega_{\tilde{p}} \exp(-s \exp(\sqrt{2}\sigma y_{\tilde{p}})) \exp(\sqrt{2}\sigma y_{\tilde{p}})^d \tag{36}$$

where $\tilde{p}$ denotes the $\tilde{p}$-th element of Gauss Hermite Quadrature, i.e., $\omega_{\tilde{p}}$ denotes the $\tilde{p}$-th weight, $y_{\tilde{p}}$ denotes the $\tilde{p}$-th node, and $N_{\tilde{p}}$ denotes the total number of quadratures. Thus, substituting into the log-likelihood:

$$\ell_j(\boldsymbol{\theta}) = \sum_{i=1}^{n_j} \delta_{ij} \log(h(t_{ij}|\mathbf{x}_{ij})) + \log\Big(\frac{1}{\sqrt{\pi}} \sum_{p=1}^{N_p} \Big[\omega_p \exp(\sqrt{2}\sigma y_p)^{d_j} \exp\Big(-\sum_{i=1}^{n_j} H(t_{ij}|\mathbf{x}_{ij}) \exp(\sqrt{2}\sigma y_p)\Big)\Big]\Big) \tag{37}$$

Similarly, the ascertainment correction in the log-normal frailty can be written as

$$A_j(\boldsymbol{\theta}) = 1 - \int_{-\infty}^{\infty} \exp(-zH(a_{j_p}|\mathbf{x}_{j_p})) f(z) dz \tag{38}$$

$$= 1 - \sum_{\tilde{p}=1}^{N_{\tilde{p}}} \omega_{\tilde{p}} \exp\left(-(\sum_{i=1}^{n_j} H(a_{j_p}|\mathbf{x}_{j_p})) \exp(\sqrt{2}\sigma y_{\tilde{p}_p})\right) \tag{39}$$

5. Missing PRS

5.1. Not Considering the Family Correlations

Given that family j has some subjects containing the missing PRS due to the sampling cost (maybe), that not all subjects are being sampled for the PRS calculation. We propose a Monte Carlo sampling method within the MCEM framework in terms of estimating the distribution of the PRS. The PRS was calculated to infer the relationship between a phenotype and multiple genetic loci, while these information were not gained if one was not involved in the original GWAS. Thus, we propose to sample the PRS using the information that we have already known. Denote $\mathbf{X}_{j,1}$ as the PRS scores vector in family j, and $\mathbf{X}_{j,2}$ the mutation status vector in family j. Take $\mathbf{p}_j$ as the proband indicator in family j, $\mathbf{c}_j$ is the current age for patients in family j. We then define $\mathbf{W} = (\log(\mathbf{t}_j), \boldsymbol{\delta}_j, \log(\mathbf{t}_j) \odot \boldsymbol{\delta}_j, \mathbf{p}_j, \mathbf{c}_j, \mathbf{X}_{j,2})$. We can make the assumption on the conditional distribution of the PRS, take $\mathbf{X}_{j,1}|\mathbf{W}, \boldsymbol{\psi} \sim MVN(\mathbf{W}\boldsymbol{\psi} + \mathbf{u}, \sigma^2\mathbf{I})$. We are interested in modelling the PRS while accounting for the between family variance, so $\mathbf{u} \sim MVN(0, \sigma_u^2\mathbf{I})$. Thus, the E-step for Gamma frailty model with ascertainment correction is then

$$E_{\mathbf{X}_{j,1,mis}}(\ell(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})) = \sum_{j=1}^{J} \Bigg[\sum_{i=1}^{n_j} \int_{\mathbf{X}_{j,1,mis}} (\delta_{ij} \log h(t_{ij}|\mathbf{x}_{ij})) - \tag{40}$$

$$+ \log\Big(\frac{(k+d_j-1)!}{k!k^{d_j-1}}(1 + \frac{\sum_{i=1}^{n_j}(H(t_{ij}|\mathbf{x}_{ij}))}{k})^{-k-d_j}\Big) + \tag{41}$$

$$- \log(A_j(\boldsymbol{\theta})) + \log f(x_{ij,1,mis}|w_{ij}, \boldsymbol{\psi}) dx_{ij,1,mis} \Bigg] \tag{42}$$

Taking a sample of size M when we sample $f(x_{ij,1,mis}|w_{ij},\boldsymbol{\psi})$ for each subject i in family j, $(x^{(1)}_{ij,i,mis},...,x^{(M)}_{ij,i,mis})$. This leads to an E-step:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = \sum_{j=1}^{J}\Bigg[\sum_{i=1}^{n_j}\sum_{m=1}^{M}(\delta_{ij}\log h(t_{ij}|\mathbf{x}^{(m)}_{ij}))- \tag{43}$$

$$+\log\Big(\frac{(k+d_j-1)!}{k!k^{d_j-1}}(1+\frac{\sum_{i=1}^{n_j}(H(t_{ij}|\mathbf{x}^{(m)}_{ij}))}{k})^{-k-d_j}\Big)+ \tag{44}$$

$$-\log(A_j(\boldsymbol{\theta}))+\log f(x^{(m)}_{ij,1,mis}|w_{ij},\boldsymbol{\psi})\Bigg] \tag{45}$$

5.2. *Considering the Family Correlations*

Given that family j has some subjects containing the missing PRS due to the sampling cost (maybe), that not all subjects are being sampled for the PRS calculation. Since subjects within one family are correlated in some genetic associations, we intend to sample the missing PRS using a multivariate normal distribution. Denote $\mathbf{X}_j \sim MVN(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ such that $\mathbf{X}_j$ is the vector of the PRS among all subjects in family j, $\mathbf{X}_j = (x_{1,j}, x_{2,j}, ..., x_{n_j,j})^\top$. When the missing PRS exists in family j, $\mathbf{X}_j$ can be decomposed into $\mathbf{X}_{obs,j}$ and $\mathbf{X}_{mis,j}$. Now suppose $\mathbf{X}_{obs,j} = (x_{1,j},...,x_{\mathring{n}_j,j})^\top$ and $\mathbf{X}_{mis,j} = (x_{\mathring{n}_j+1,j},...,x_{n_j,j})^\top$. We can also partition $\boldsymbol{\mu}_j$ such that

$$\mathbf{X}_j = \begin{bmatrix}\mathbf{X}_{obs,j}\\ \mathbf{X}_{mis,j}\end{bmatrix} \tag{46}$$

and

$$\boldsymbol{\mu}_j = \begin{bmatrix}\boldsymbol{\mu}_{obs,j}\\ \boldsymbol{\mu}_{mis,j}\end{bmatrix} \tag{47}$$

Similarly, the covariace matrix is then decomposed

$$\boldsymbol{\Sigma}_j = \begin{bmatrix}\boldsymbol{\Sigma}_{obs,j} & \boldsymbol{\Sigma}_{obs,mis,j}\\ \boldsymbol{\Sigma}_{obs,mis,j} & \boldsymbol{\Sigma}_{mis,j}\end{bmatrix} \tag{48}$$

6. Monte Carlo EM

The complete data log-likelihood for family j is $\ell_j(\boldsymbol{\theta}; h_{ij})$ where $\boldsymbol{\theta}$ consists all baseline parameters, and model coefficients β's, as well as the frailty parameter ϕ. The E-step for complete data is:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)}) = \int \ell(\boldsymbol{\theta}; h_{ij})\cdot f(x_{mis,i}|x_{obs,i}, z, \boldsymbol{\theta}^{(r)}, t_{ij}, \delta_{ij}, p_j)dx_{mis,ij} \tag{49}$$

We sample the size m_i for each i-th observation, $x^*_{i1},...,x^*_{im_i}$ from the distribution $f(x_{mis,ij}|\cdot)$, and take $M = 1,...,m_i$, such that each X^*_{iM} depends on the iteration number for $r+1$ iterations. In general:

$$\hat{Q}(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}) = \frac{1}{m_i}\sum_{M=1}^{m_i} \ell(x^*_{iM}, x_{obs,ij}, t_{ij}, \boldsymbol{\theta}, z_j) \tag{50}$$

More specifically,

1. We first initialize $m, \theta^{(0)}$, and start the burn-in.
2. Also, we set importance weights $w_t = 1$ for all $t = 1, ..., m$.
3. At the burn-in iteration s, we generate $x_{miss,1}, ..., x_{miss,m} \sim N(\mu_X|X_{obs}, \theta^{(s)}, z)$ using MCMC sample.
4. In the E-step, we estimate $Q(\theta|\theta^{(s)})$ by using the importance weights:

$$Q_m(\theta|\hat{\theta}^{(s)}) = \frac{\sum_{t=1}^m w_t \log f(X_{obs}, X_{miss,t}|\theta)}{\sum_{r=1}^m w_t} \tag{51}$$

5. Note the numerator is actually a weighted log-likelihood. In the M-step, we maximize $Q_m(\theta|\hat{\theta}^{(s)})$ to obtain $\hat{\theta}^{(s+1)}$.
6. Repeat (3.) - (5.) for s burn-in iterations.
7. Then re-initialize $\hat{\theta}^{(0)} = \hat{\theta}^{(s)}$
8. We generate $x_{miss,1}, ..., x_{miss,m} \sim N(\mu_X|X_{obs}, \hat{\theta}^{(0)}, z)$ using MCMC sampler. At iteration $r+1$
9. Compute the importance weights from the ratio of likelihood

$$w_t = \frac{L(\hat{\theta}^{(r)}|X_{miss,t}, X_{obs})}{L(\hat{\theta}^{(0)}|X_{miss,t}, X_{obs})} \tag{52}$$

10. Thus, the E-step can be written as

$$Q_m(\theta|\hat{\theta}^{(r)}) = \frac{\sum_{t=1}^m w_t \log f(X_{miss,t}, X_{obs}|\theta)}{\sum_{t=1}^m w_t} \tag{53}$$

11. Then M-step: we maximize $Q_m(\theta|\hat{\theta}^{(r)})$ to obtain $\hat{\theta}^{(r+1)}$.

This automated MCEM firstly optimizes the importance weights at burn-ins, then performs the actual EM to find $\hat{\theta}$. This importance weight ensures the imputation step of the missing data actually yields to the real distribution.

7. Correlated Frailty using Kinship Matrix

Family members are correlated within one family, that we denote K as the kinship correlation matrix among all observations. This matrix ensures those individuals not from the same family automatically have a correlation of 0. The likelihood construction needs multivariate form. For $\mathbf{Z} \sim \text{MVN}(0, \sigma^2 K)$, that K has the diagonal of 1. The likelihood is

$$
\begin{aligned}
L(\cdot) &= \int_{\mathbb{R}^n} \prod_{i=1}^{n} (h(t|\mathbf{x}_i, \mathbf{z}_i))^{\delta_i} \exp(-H(t|\mathbf{x}_i, \mathbf{z}_i)) f(\mathbf{z}) d\mathbf{z} && (54)\\
&= \int_{\mathbb{R}^n} \prod_{i=1}^{n} (h(t|\mathbf{x}_i))^{\delta_i} \exp(\mathbf{z}_i)^{\delta_i} \exp(-H(t|\mathbf{x}_i) \exp(\mathbf{z}_i)) f(\mathbf{z}) d\mathbf{z} && (55)\\
&= \prod_{i=1}^{n} (h(t|\mathbf{x}_i))^{\delta_i} \int_{\mathbb{R}^n} \exp(\delta_i \mathbf{z}_i - H(t|\mathbf{x}_i) \exp(\mathbf{z}_i)) f(\mathbf{z}) d\mathbf{z} && (56)
\end{aligned}
$$

Applying the Laplace approximation, and taking the log for the likelihood, we obtain

$$
\ell(\cdot) = \sum_{i=1}^{n} \Big[\delta_i \log h(t|\mathbf{x}_i)\Big] + \sum_{i=1}^{n} \Big[\delta_i \hat{\mathbf{z}} - H(t_i|\mathbf{x}_i) \exp(\hat{\mathbf{z}})\Big] - \frac{1}{2} \hat{\mathbf{z}}^\top \Sigma^{-1} \hat{\mathbf{z}} \tag{57}
$$

such that $\Sigma = \sigma^2 K$. Also, we treat the random effect $\mathbf{z}$ as a vector of parameters, and use outer-loop to search for the σ, and use inner-loop to search for other parameters (baseline parameters, and β) including $\mathbf{z}$. The process can be achieved via Newton-Raphson algorithm. For computational efficiency, we can set $\Sigma^{-1} = L^\top L$ through Cholesky Decomposition. In this way, $\mathbf{z}L \sim MVN(0, \sigma^2 I)$. In order to apply NR-algorithm, the gradient and the hessian are required. The gradient for parameters is:

$$
\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \delta_i \mathbf{x}_i + \sum_{i=1}^{n} -H(t_i|\mathbf{x}_i) \mathbf{x}_i \exp(\mathbf{z}) \tag{58}
$$

$$
\frac{\partial \ell}{\partial \mathbf{z}} = \sum_{i=1}^{n} \delta_i - (t_i|\mathbf{x}_i) \exp(\hat{\mathbf{z}}) - \Sigma^{-1} \hat{\mathbf{z}} \tag{59}
$$

$$
\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n} \frac{\delta_i}{\alpha} + \sum_{i=1}^{n} - \frac{H(t_i|\mathbf{x}_i) \exp(\hat{\mathbf{z}})}{\alpha} \tag{60}
$$

$$
\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^{n} \delta_i (\frac{1}{\lambda} + \log(t_i)) + \sum_{i=1}^{n} -H(t_i|\mathbf{x}_i) \exp(\hat{\mathbf{z}}) \log(t_i) \tag{61}
$$

The hessian matrix element, i.e., second partial derivative is

$$
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}^\top \boldsymbol{\beta}} = \sum_{i=1}^{n} -H(t_i|\mathbf{x}_i) \exp(\hat{\mathbf{z}}) x_{ij} x_{ik} \tag{62}
$$

$$
\frac{\partial^2 \ell}{\partial \mathbf{z}^\top \mathbf{z}} = \sum_{i=1}^{n} -H(t_i|\mathbf{x}_i) \exp(\hat{\mathbf{z}}) - \Sigma^{-1} \tag{63}
$$

$$
\frac{\partial^2 \ell}{\partial \alpha^2} = \sum_{i=1}^{n} - \frac{\delta_i}{\alpha^2} \tag{64}
$$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = \sum_{i=1}^{n} -\frac{\delta_i}{\lambda^2} - H(t_i|\mathbf{x}_i) \exp(\hat{\mathbf{z}}) \log(t_i)^2 \tag{65}$$

7.1. Proof of $\Sigma = LL^\top$

Every symmetric positive definite matrix Σ can be decomposed into $\Sigma = LL^\top$, where L is a lower triangular matrix with real and positive diagonal entries.

Proof. Set-ups:

1. Covariance matrix $\mathbf{\Sigma}$ is by definition symmetric and positive definite, e.g.

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2_{X_1} & Cov(X_1, X_2) \\ Cov(X_1, X_2) & \sigma^2_{X_2} \end{pmatrix} \tag{66}$$

   such that $\mathbf{X\Sigma X}^\top > 0$ always, and this matrix is symmetric.

2. Suppose $\mathbf{X}$ has n observations, then Σ is $n \times n$, the first element is $\sigma_{11} > 0$ by definition (For simplicity, we use σ_{11} rather than it's square to denote the variance). Define $l_{11} = \sqrt{\sigma_{11}}$, to be the first element of L. For the first column of L, let $l_{j1} = \frac{\sigma_{j1}}{l_{11}}$ for $j = 2, ...$

Induction step: Assume we have first $k-1$ columns of L, consider k-th column

- For the diagonal element $l_{kk} = \sqrt{\sigma_{kk} - \sum_{j=1}^{k-1} l^2_{kj}}$
- For off-diagonals,

$$l_{ik} = \frac{\sigma_{ik} - \sum_{j=1}^{k-1} l_{ij} l_{kj}}{l_{kk}} \tag{67}$$

  for $i = k+1, ..., n$.

with the repetition for each column $k = 2, ..., n$, the top-left $k \times k$ submatrix of $LL^\top$ matches that of $\mathbf{\Sigma}$. For example, when $k = 3$,

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & & \\ & \sigma_{22} & \\ & & \sigma_{33} \end{pmatrix} \tag{68}$$

and

$$L = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \tag{69}$$

then

$$LL^\top = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{pmatrix} = \begin{pmatrix} l^2_{11} & l_{11}l_{21} & l_{11}l_{31} \\ l_{21}l_{11} & l^2_{21} + l^2_{22} & l_{21}l_{31} + l_{22}l_{32} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l^2_{31} + l^2_{32} + l^2_{33} \end{pmatrix} \tag{70}$$

Take

$$\Sigma = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix} \tag{71}$$

Then by definition of Cholesky Decomposition, we can calculate $l_{11}^2 = \sigma_{11} \implies l_{11} = \sqrt{4} = 2$, and $l_{21} = \frac{\sigma_{21}}{l_{11}} = 2/2 = 1$, and $l_{31} = 1$. Similarly for l_{22}, l_{32}, l_{33}. Therefore,

$$L = \begin{pmatrix} 2 & 0 & 0 \\ 1 & \sqrt{2} & 0 \\ 1 & 0 & \sqrt{2} \end{pmatrix} \tag{72}$$

which implies

$$LL^\top = \begin{pmatrix} 2 & 0 & 0 \\ 1 & \sqrt{2} & 0 \\ 1 & 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix} = \Sigma \tag{73}$$

$\square$

Essentially, the Cholesky Decomposition transforms the multivariate normal to a standard multivariate normal. When $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{\Sigma})$, let $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^\top$, then $\mathbf{Y} = \mathbf{L}^{-1}\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I})$ that $\mathbf{I}$ is the identity matrix, since $\mathbf{L}^{-1}\mathbf{\Sigma}(\mathbf{L}^{-1})^\top = \mathbf{L}^{-1}\mathbf{L}\mathbf{L}^\top(\mathbf{L}^{-1})^\top = \mathbf{I}$. This will simplify the computational process.

8. Multiple Imputation Method