Note on Variational Approximations for Generalized Survival Models

Benjamin Christoffersen

October 23, 2019

In this note, we make a comparison between using variational approximation (VA) in generalized survival models (GSM) compared to generalized linear mixed models (GLMM). Some non-GSMs will also be used. The notation used in this note will be very similar to the notation in Ormerod and Wand (2012) who consider VAs for GLMMs. Particularly, they consider using VAs for GLMMs with grouped data with Gaussian random effects. We will start by doing the same and then consider GSMs afterwards. We will make some attempts to introduce survival analysis although this note is not intended to be completely self-contained.

1 GLMM

We start with the GLMMs as in Ormerod and Wand (2012). The results in this section are essentially those shown by Ormerod and Wand (2012). Let $\boldsymbol{Y}_i$ denote the outcomes of the ith cluster. Each cluster $i = 1, \dots, m$ has n_i members and two associated design matrices denoted by $\mathbf{X}_i$ and $\mathbf{Z}_i$ where the former is for the fixed effects and the latter is for the random effects. Cluster and group will be used interchangeably to denote a group of observations which are assumed, in general, only to be independent conditional on a random effect. We assume that a canonical link function is used in the conditional distribution such that

$$
\begin{aligned}
\boldsymbol{Y}_i \mid \boldsymbol{U}_i &\sim \exp\left(\boldsymbol{Y}_i^\top(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{U}_i) - \mathbf{1}^\top b(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{U}_i) + \mathbf{1}^\top c(\boldsymbol{Y}_i)\right) && (1)\\
&= g_i(\boldsymbol{Y}_i \mid \boldsymbol{U}_i) && (2)\\
\boldsymbol{U}_i &\sim N(\mathbf{0}, \boldsymbol{\Sigma})
\end{aligned}
$$

where the functions b and c are known, the unknown parameters are $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, the outcomes, $\boldsymbol{Y}_i$, and the design matrices, $\mathbf{X}_i$ and $\mathbf{Z}_i$, are observed while the K dimensional random effects, $\boldsymbol{U}_i$, are unknown. Functions are applied elementswise as in Ormerod and Wand (2012) and will be throughout the rest of this note. That is, if $\boldsymbol{s} = (s_1, s_2, s_3)^\top$ then

$$
h(\boldsymbol{s}) = (h(s_1), h(s_2), h(s_3))^\top
$$

We only focus on a single group, i, in this note as deriving the log-likelihood contribution for one group is all we need.

The (often) intractable marginal log-likelihood for each cluster i is

$$\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= \log\left(\int g_i(\boldsymbol{y}_i \mid \boldsymbol{u})\phi(\boldsymbol{u}; \boldsymbol{0}, \boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\right) \\
&= \boldsymbol{y}_i^\top \mathbf{X}_i\boldsymbol{\beta} + \mathbf{1}^\top c(\boldsymbol{y}_i) - \frac{1}{2}\log\lvert\boldsymbol{\Sigma}\rvert - \frac{K}{2}\log 2\pi \\
&\qquad + \log\int \exp\left(\boldsymbol{y}_i^\top \mathbf{Z}_i\boldsymbol{u} - \mathbf{1}^\top b(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}) - \frac{1}{2}\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\right)\mathrm{d}\boldsymbol{u} \qquad (3)
\end{aligned}$$

where ϕ is the density function of the multivariate normal distribution with mean and covariance given by the second and third argument respectively and $\mathbf{1}$ is vector of ones which dimension is given by the context.

Instead of working with the marginal log-likelihood given in Equation (3), we turn to a lower bound. Particularly, we consider the Kullback-Leibler divergence between a multivariate normal distribution with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Lambda}_i$ given by

$$\begin{aligned}
\tilde{l}_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) &= \int \phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\log\frac{g_i(\boldsymbol{y}_i \mid \boldsymbol{u})\phi(\boldsymbol{u}; \boldsymbol{0}, \boldsymbol{\Sigma})}{\phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)}\,\mathrm{d}\boldsymbol{u} \\
&= \boldsymbol{y}_i^\top \mathbf{X}_i\boldsymbol{\beta} + \mathbf{1}^\top c(\boldsymbol{y}_i) - \frac{1}{2}\log\lvert\boldsymbol{\Sigma}\rvert - \frac{K}{2}\log 2\pi \\
&\qquad + \int\Bigg(\boldsymbol{y}_i^\top \mathbf{Z}_i\boldsymbol{u} - \mathbf{1}^\top b(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}) - \frac{1}{2}\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \\
&\qquad\qquad - \log\phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\Bigg)\phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\,\mathrm{d}\boldsymbol{u} \qquad (4)
\end{aligned}$$

Comparing the lower-bound in Equation (4) with the marginal log-likelihood in (3) we observe that

- the $\boldsymbol{y}_i^\top \mathbf{Z}_i\boldsymbol{u}\phi(\cdot)$ terms in the integrand yields a dot product with the mean which has a closed form solution.
- the $\mathbf{1}^\top b(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u})\phi(\cdot)$ term in the integrand yields a sum of univariate integrals which we may have a closed form solution as in the Poisson case or which we can approximate as mentioned in Ormerod and Wand (2012).
- the $-\frac{1}{2}\boldsymbol{u}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\phi(\cdot)$ term in the integrand yields a closed form solution which is $-\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}_i/2 - \boldsymbol{\mu}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i/2$.
- the last term in the integrand yields a closed form solution which is $\frac{K}{2}\log 2\pi + \frac{1}{2}\log\lvert\boldsymbol{\Lambda}_i\rvert + \frac{K}{2}$.

Moreover, as emphasized in Ormerod and Wand (2012), the fact that we only get univariate integrals that we need to approximate rather than multivariate is a great advantage.

The reason we end up with this result is somewhat due to the choice of a conditional distribution for the observed outcomes which is from the exponential family, that the random effect distribution and the distribution we use in the variational approximation are the multivariate normal distribution, and that we use a canonical link function (otherwise we would we end up with an additional set of univariate integral terms which we potentially need to approximate).

2 GSM

In the survival analysis setting, we denote the outcomes as $\widetilde{\boldsymbol{T}}_i$ which is the times to the event for cluster i. We will only consider right censoring here for simplicity. Thus, we only observe some of the event times while other are censored if the independent censoring time denoted by C_{ik} is less than the event time $\tilde{T}_{ik}$. We let $S(t) = \text{P}(\tilde{T} > t)$ denote the survival function and use a class of GSMs of the form

$$g\left(S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})\right) = g\left(\text{P}(\tilde{T} > t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})\right) = g(S_0(t)) + \boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u}$$

where g is a specific link function and S_0 is a baseline survival function.

Let $\delta_{ik} = 1_{\{\tilde{T}_{ik} < C_{ik}\}}$ be an indicator of whether we observe the event time for the kth observation in the ith cluster, let $T_{ik} = \min(\tilde{T}_{ik}, C_{ik})$ denote the potentially censored observed time for the kth observation in the ith cluster and let λ denote the conditional hazard function such that

$$\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) = \lim_{h \rightarrow 0^+} \frac{\text{P}\left(t \leq \tilde{T} < t + h \mid \tilde{T} \geq t, \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}\right)}{h}$$

where standard useful identities are

$$\begin{aligned}
\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \frac{f(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})}{S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})} = -\frac{\partial}{\partial t}\log S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) \\
S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \int_t^\infty f(s \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})\,\text{d}s = \exp\left(-\int_0^t \lambda(s \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})\,\text{d}s\right)
\end{aligned}$$

and f is the conditional density function of $\tilde{T}$. Thus, the observed conditional likelihood of a given observation is

$$\begin{aligned}
p(t, \delta \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= f(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})^\delta\, \text{P}\left(\tilde{T} > t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}\right)^{1-\delta} \\
&= \lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})^\delta S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})
\end{aligned}$$

Using the above, the marginal log-likelihood from the ith cluster is

$$\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= \log\Bigg(\int \exp\big(\mathbf{1}^\top(\boldsymbol{\delta}_i \cdot \log\lambda(\boldsymbol{t}_i \mid \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u})) \\
&\qquad\qquad + \mathbf{1}^\top \log S(\boldsymbol{t}_i \mid \mathbf{X}_i, \boldsymbol{Z}_i, \boldsymbol{u})\big)\,\phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg) \\
&= \log\Bigg(\int \exp\Bigg(\boldsymbol{\delta}_i^\top \log\left(-\frac{S'(\boldsymbol{t}_i \mid \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u})}{S(\boldsymbol{t}_i \mid \mathbf{X}_i, \boldsymbol{Z}_i, \boldsymbol{u})}\right) \\
&\qquad\qquad + \mathbf{1}^\top \log S(\boldsymbol{t}_i \mid \mathbf{X}_i, \boldsymbol{Z}_i, \boldsymbol{u})\Bigg)\,\phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg) \qquad (5)
\end{aligned}$$

where the '$\cdot$' are element-wise. It is not clear from Equation (5) whether we get as neat of an expression by optimizing over a lower bound as in Equation (4). What we will do instead is to consider specific cases.

2.1 Log-log Link

The log-log link function is given by $g(x) = \log(-\log(x))$ such that

$$\begin{aligned}
& \log\left(-\log\left(S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})\right)\right) = \log\left(-\log\left(S_0(t)\right)\right) + \boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u} \\
\Leftrightarrow \qquad & S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) = \exp\left(-\exp\left(\log\left(-\log\left(S_0(t)\right)\right) + \boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u}\right)\right) \\
& \qquad = S_0(t)^{\exp(\boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u})} \qquad (6)
\end{aligned}$$

In particular, setting $\log\left(-\log\left(S_0(t)\right)\right) = \log\lambda + \gamma\log t$ yields

$$\begin{aligned}
\log S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= -\lambda t^\gamma \exp\left(\boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u}\right) \\
\log\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \log\gamma\lambda + (\gamma - 1)\log t + \boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u}
\end{aligned}$$

which is seen to be the Weibull proportional hazard model. Inserting this into Equation (5) yields the following marginal log-likelihood

$$\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = &\left(\sum_k \delta_{ik}\right)\log\gamma\lambda + (\gamma - 1)\boldsymbol{\delta}_i^\top \log\boldsymbol{t}_i + \boldsymbol{\delta}_i^\top\mathbf{X}_i\boldsymbol{\beta} \\
&+ \log\Bigg(\int \exp\left(\boldsymbol{\delta}_i^\top\mathbf{Z}_i\boldsymbol{u} - \mathbf{1}^\top\left(\lambda\boldsymbol{t}_i^\gamma \cdot \exp\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}\right)\right)\right) \\
&\qquad \cdot \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg) \qquad (7)
\end{aligned}$$

Applying a lower bound as in Equation (4) yields

$$
\begin{aligned}
\tilde{l}_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) &= \\
&\left(\sum_k \delta_{ik}\right) \log \gamma\lambda + (\gamma - 1)\boldsymbol{\delta}_i^\top \log \boldsymbol{t}_i + \boldsymbol{\delta}_i^\top \mathbf{X}_i\boldsymbol{\beta} - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{K}{2}\log 2\pi \\
&+ \int \Bigg(\boldsymbol{\delta}_i^\top \mathbf{Z}_i\boldsymbol{u} - \mathbf{1}^\top \left(\lambda \boldsymbol{t}_i^\gamma \cdot \exp\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}\right)\right) \\
&\qquad -\frac{1}{2}\boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{u} - \log\phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\Bigg) \phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\,\mathrm{d}\boldsymbol{u}
\end{aligned} \tag{8}
$$

which is tractable since the only different type of terms in the integrand compared to Equation (4) is the $-\mathbf{1}^\top \left(\lambda \boldsymbol{t}_i^\gamma \cdot \exp\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}\right)\right)\phi(\cdot)$ terms which all have a closed form solutions given by the moment generating function of the multivariate normal distribution.

Next, we return to a general baseline hazard function S_0 in Equation (6). In this case, the hazard function is given by

$$
\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) = -\frac{S'(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})}{S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})} = -\frac{S_0'(t)}{S_0(t)} \exp\left(\boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u}\right)
$$

such that

$$
\begin{aligned}
\log S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \exp(\boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u}) \log S_0(t) \\
\log \lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= L_0(t) + \boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u} \qquad\qquad L_0(t) = \log -\frac{S_0'(t)}{S_0(t)}
\end{aligned}
$$

Inserting this into Equation (5) yields the following marginal log-likelihood

$$
\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \boldsymbol{\delta}_i^\top L_0(\boldsymbol{t}_i)) + \boldsymbol{\delta}_i^\top \mathbf{X}_i\boldsymbol{\beta} + \log \Bigg(\int \exp\big(\boldsymbol{\delta}_i^\top \mathbf{Z}_i\boldsymbol{u} \\
+\mathbf{1}^\top \left(\log S_0(\boldsymbol{t}_i) \cdot \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u})\right)\big)\, \phi(\boldsymbol{u}; 0, \boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg)
\end{aligned} \tag{9}
$$

where we see that applying a lower bound as in Equation (8) yields a similar tractable integral. This again is somewhat to related to Poisson model with the log link function. To see this, we add known offset terms denoted by $\boldsymbol{o}_i$ to the GLMM in Equation (1)

$$
\boldsymbol{Y}_i \mid \boldsymbol{U}_i \sim \exp\left(\boldsymbol{Y}_i^\top(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{U}_i + \boldsymbol{o}_i) - \mathbf{1}^\top b(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{U}_i + \boldsymbol{o}_i) + \mathbf{1}^\top c(\boldsymbol{y}_i)\right)
$$

Next, $b(x) = \exp(x)$ in the Poisson model with the log link function. Thus, the marginal log-likelihood as in Equation (3) is given by

$$
\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \mathbf{1}^\top \left(c(\boldsymbol{y}_i) + \boldsymbol{o}_i\right) + \boldsymbol{y}_i^\top \mathbf{X}_i\boldsymbol{\beta} + \log \Bigg(\int \exp\big(\boldsymbol{y}_i^\top \mathbf{Z}_i\boldsymbol{u} \\
-\mathbf{1}^\top \exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u} + \boldsymbol{o}_i)\big) \cdot \phi(\boldsymbol{u}; 0, \boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg)
\end{aligned} \tag{10}
$$

Observe that the marginal log-likelihoods in Equation (9) and (10) are equal if

$$\begin{aligned}
\boldsymbol{y}_i &= \boldsymbol{\delta}_i \\
\boldsymbol{o}_i &= \log(-\log S_0(\boldsymbol{t}_i)) \\
\boldsymbol{c}(\boldsymbol{y}_i) &= \boldsymbol{\delta}_i \cdot L_0(\boldsymbol{t}_i)) - \boldsymbol{o}_i = \boldsymbol{\delta}_i \cdot \log - \frac{S_0'(\boldsymbol{t}_i)}{S_0(\boldsymbol{t}_i)} - \log(-\log S_0(\boldsymbol{t}_i))
\end{aligned}$$

2.2 Negative Logit Link

Another link function is the minus logit function $g(x) = \log((1-x)/x)$ such that

$$\begin{aligned}
\log\left(\frac{1 - S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})}{S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})}\right) &= \log\left(\frac{1 - S_0(t)}{S_0(t)}\right) + \boldsymbol{x}^\top \boldsymbol{\beta} + \boldsymbol{z}^\top \boldsymbol{u} \\
\Leftrightarrow \qquad S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \left(1 + \exp\left(\log\left(\frac{1 - S_0(t)}{S_0(t)}\right) \right.\right. \\
&\qquad \left.\left. + \boldsymbol{x}^\top \boldsymbol{\beta} + \boldsymbol{z}^\top \boldsymbol{u}\right)\right)^{-1} \\
&= \left(1 + \frac{1 - S_0(t)}{S_0(t)} \exp\left(\boldsymbol{x}^\top \boldsymbol{\beta} + \boldsymbol{z}^\top \boldsymbol{u}\right)\right)^{-1} \\
\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \frac{\exp\left(\boldsymbol{x}^\top \boldsymbol{\beta} + \boldsymbol{z}^\top \boldsymbol{u}\right)(-S_0'(t))}{S_0(t)\left(\exp\left(\boldsymbol{x}^\top \boldsymbol{\beta} + \boldsymbol{z}^\top \boldsymbol{u}\right)(1 - S_0(t)) + S_0(t)\right)}
\end{aligned}$$

Taking the logarithm of the conditional survival function and hazard function yields

$$\begin{aligned}
\log S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= -\log\left(1 + \frac{1 - S_0(t)}{S_0(t)} \exp\left(\boldsymbol{x}^\top \boldsymbol{\beta} + \boldsymbol{z}^\top \boldsymbol{u}\right)\right) \\
\log \lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \boldsymbol{x}^\top \boldsymbol{\beta} + \boldsymbol{z}^\top \boldsymbol{u} + \log(-S_0'(t)) - \log S_0(t) \\
&\qquad - \log\left(\exp\left(\boldsymbol{x}^\top \boldsymbol{\beta} + \boldsymbol{z}^\top \boldsymbol{u}\right)(1 - S_0(t)) + S_0(t)\right)
\end{aligned}$$

The marginal log-likelihood is

$$\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \boldsymbol{\delta}_i^\top \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\delta}_i^\top &\left(\log(-S_0'(\boldsymbol{t}_i)) - \log S_0(\boldsymbol{t}_i)\right) \\
\log\Bigg(\int \exp\Bigg(&\boldsymbol{\delta}_i^\top \mathbf{Z}_i \boldsymbol{u} \\
&+\boldsymbol{\delta}_i^\top \left(-\log\left(\exp\left(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{u}\right)(1 - S_0(\boldsymbol{t}_i)) + S_0(\boldsymbol{t}_i)\right)\right) \\
&+\mathbf{1}^\top \left(-\log\left(1 + \frac{1 - S_0(\boldsymbol{t}_i)}{S_0(\boldsymbol{t}_i)} \exp\left(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{u}\right)\right)\right)\Bigg) \\
&\cdot \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma})\, \mathrm{d}\boldsymbol{u}\Bigg)
\end{aligned}$$

Using the above, we can find that if we apply a lower bound as in Equation (4) we get two new types of terms of the form

$$\int -\log\left(c(t) + f(t)\exp\left(\boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u}\right)\right)\phi(\boldsymbol{u};\boldsymbol{\mu}_i,\boldsymbol{\Lambda}_i)\,\mathrm{d}\boldsymbol{u}$$

for given functions $c : (0,\infty) \rightarrow (0,1]$ and $f : (0,\infty] \rightarrow (0,\infty)$ and fixed t. These need to be approximated (as far as I see). Alternatively, we can consider other types of VAs which may yield a tractable integral. Though, the tangent transform approach suggested by Jaakkola and Jordan (2000) cannot be applied due to the c and f factors. Regardlessly, the comments in Ormerod and Wand (2012) about the tangent transform approach should likely be considered before taking this any further.

2.3 Probit Link

An alternative link function is the probit link function given by $g(x) = -\Phi(x)$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Let $\eta_0(t) = \Phi^{-1}(S_0(t))$. Then the conditional survival function and hazard function are

$$\begin{aligned}
& & -\Phi^{-1}\left(S(t\mid\boldsymbol{x},\boldsymbol{z},\boldsymbol{u})\right) &= -\Phi^{-1}\left(S_0(t)\right) + \boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u} \\
&\Leftrightarrow & S(t\mid\boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) &= \Phi\left(\eta_0(t) - \boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{u}\right) \\
& & \lambda(t\mid\boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) &= (-\eta_0'(t))\frac{\phi\left(\eta_0(t) - \boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{u}\right)}{\Phi\left(\eta_0(t) - \boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{u}\right)}
\end{aligned}$$

where $\phi(\cdot)$ is the standard normal density function unless two additional arguments are applied. Thus, the marginal log-likelihood is

$$\begin{aligned}
l_i(\boldsymbol{\beta},\boldsymbol{\Sigma}) = \boldsymbol{\delta}_i^\top &\log(-\eta_0'(\boldsymbol{t}_i)) \\
&+ \log\Bigg(\int \exp\big(\boldsymbol{\delta}_i^\top\log\phi\left(\eta_0(\boldsymbol{t}_i) - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u}\right) \\
&\qquad\qquad - \boldsymbol{\delta}_i^\top\log\Phi\left(\eta_0(\boldsymbol{t}_i) - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u}\right) \\
&\qquad\qquad + \mathbf{1}^\top\log\Phi\left(\eta_0(\boldsymbol{t}_i) - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u}\right)\big)\phi(\boldsymbol{u};\mathbf{0},\boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg) \\
= \boldsymbol{\delta}_i^\top &\log(-\eta_0'(\boldsymbol{t}_i)) \\
&+ \log\Bigg(\int \exp\big(\boldsymbol{\delta}_i^\top\log\phi\left(\eta_0(\boldsymbol{t}_i) - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u}\right) \\
&\qquad\qquad + (\mathbf{1} - \boldsymbol{\delta}_i^\top)\log\Phi\left(\eta_0(\boldsymbol{t}_i) - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u}\right)\big) \\
&\qquad\qquad \cdot\phi(\boldsymbol{u};\mathbf{0},\boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg) \qquad (11)
\end{aligned}$$

which yields an intractable lower bound due to the $\log\Phi$ terms.

We can compare this to a GLMM with binary outcomes, $y_{ik} \in \{0,1\}$, and the probit link function. In this case, we do not have a canonical link function.

The conditional distribution of the observed outcomes as in Equation (1) is given by

$$\begin{aligned}
\boldsymbol{Y}_i \mid \boldsymbol{U}_i &\sim \exp\left(\boldsymbol{Y}_i^\top h(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{U}_i) - \mathbf{1}^\top b(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{U}_i) + \mathbf{1}^\top c(\boldsymbol{y}_i)\right) \\
h(\eta) &= \log\Phi(\eta) - \log(1 - \Phi(\eta)) = \log\Phi(\eta) - \log\Phi(-\eta) \\
b(\eta) &= -\log(1 - \Phi(\eta)) = -\log\Phi(-\eta) \\
c(y) &= 0
\end{aligned}$$

Thus, the marginal log-likelihood is

$$\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \log\int \exp\big(\boldsymbol{y}_i^\top &\left(\log\Phi(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}) - \log\Phi(-\mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u})\right) \\
&+ \mathbf{1}^\top \log\Phi(-\mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u})\big)\, \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}
\end{aligned}$$

and we end with similar terms and hence similar issues with the lower bound.

One idea for the probit regression is to introduce additional auxiliary variables (Consonni and Marin, 2007; Ormerod and Wand, 2010, section 2.2.4) which is suggested by Albert and Chib (1993) for Gibbs sampling. Sticking to the GLMM example, the auxiliary variables are

$$\boldsymbol{A}_i \mid \boldsymbol{U}_i \sim N(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{U}_i, \mathbf{I})$$

where $\mathbf{I}$ is the identity matrix which dimension is given by the context. We then assume that the conditional distribution of $\boldsymbol{Y}_i \mid \boldsymbol{A}_i$ is degenerate and given by

$$\boldsymbol{Y}_i \mid \boldsymbol{A}_i \sim \exp\left(\boldsymbol{Y}_i^\top \log\mathbf{1}_{\{\boldsymbol{A}_i \geq 0\}} + (\mathbf{1} - \boldsymbol{Y}_i)^\top \log\mathbf{1}_{\{\boldsymbol{A}_i < \mathbf{0}\}}\right)$$

where $\mathbf{1}_{\{\cdot\}}$ is a vector of zeros and ones where each element is one if the corresponding element in the subscript is true. Thus, the joint density is

$$\begin{aligned}
f(\boldsymbol{y}_i, \boldsymbol{a}_i, \boldsymbol{u}_i) = \exp\big(&\boldsymbol{y}_i^\top \mathbf{1}_{\{\boldsymbol{a}_i \geq 0\}} + (\mathbf{1} - \boldsymbol{y}_i)^\top \mathbf{1}_{\{\boldsymbol{a}_i < 0\}} \\
&+ \log\phi(\boldsymbol{a}_i; \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}_i, \mathbf{I}) + \log\phi(\boldsymbol{u}_i; \mathbf{0}, \boldsymbol{\Sigma})\big)
\end{aligned} \tag{12}$$

Following Ormerod and Wand (2010) (and omitting details on how to update distributions in VAs when the used distribution in the approximation factorizes), we use the following factorization in our variational approximation

$$q(\boldsymbol{a}_i, \boldsymbol{u}_i) = q_1(\boldsymbol{a}_i) q_2(\boldsymbol{u}_i)$$

and recursively update $q_1(\boldsymbol{A}_i)$ given $q_2(\boldsymbol{U}_i)$ and vice versa. The first update yields

$$\begin{aligned}
\log q_1^*(\boldsymbol{A}_i) \propto \int \big(&\boldsymbol{y}_i^\top \log\mathbf{1}_{\{\boldsymbol{A}_i \geq 0\}} + (\mathbf{1} - \boldsymbol{y}_i)^\top \log\mathbf{1}_{\{\boldsymbol{A}_i < \mathbf{0}\}} \\
&+ \log\phi(\boldsymbol{A}_i; \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}, \mathbf{I})\big)\, \phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\,\mathrm{d}\boldsymbol{u}
\end{aligned}$$

which is proportional the density of left or right truncated normally distributed variables with mean vector $\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\boldsymbol{\mu}_i$ (not accounting for truncation) and an identity matrix as the covariance matrix. Similarly, the second update yields

$$\log q_2^*(\boldsymbol{U}_i) \propto \int \left(\log \phi\left(\boldsymbol{a}; \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{U}_i, \mathbf{I}\right) + \log\phi(\boldsymbol{U}_i; \mathbf{0}, \boldsymbol{\Sigma})\right) q_1^*(\boldsymbol{a})\, \mathrm{d}\boldsymbol{a}$$

which yields a multivariate normal distribution with covariance matrix and mean

$$\boldsymbol{\Lambda}_i = (\mathbf{Z}_i^\top\mathbf{Z}_i + \boldsymbol{\Sigma}^{-1})^{-1}, \qquad \boldsymbol{\mu}_i = \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top \operatorname{E}_{q_1^*}(\boldsymbol{A}_i) \tag{13}$$

where $\operatorname{E}_{q_1^*}(\boldsymbol{A}_i)$ is the mean of $\boldsymbol{A}_i$ in the q_1^* distribution. We can repeat updating q_1 and q_2 till a convergence criteria is satisfied.

The estimated density at convergence is

$$\begin{aligned}\log q^*(\boldsymbol{a}_i, \boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) = \boldsymbol{y}_i^\top \log\left(\frac{\mathbf{1}_{\{\boldsymbol{a}_i\geq 0\}}}{\Phi\left(\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\boldsymbol{\mu}_i\right)}\right) \\ + (\mathbf{1}-\boldsymbol{y}_i)^\top \log\left(\frac{\mathbf{1}_{\{\boldsymbol{a}_i< 0\}}}{\Phi\left(-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i\right)}\right) \\ + \log\phi\left(\boldsymbol{a}_i; \mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\boldsymbol{\mu}_i, \mathbf{I}\right) + \log\phi\left(\boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i\right)\end{aligned}$$

Next, we compute the Kullback-Leibler divergence with the above and the augmented density in Equation (12) which gives us the following lower bound

$$\begin{aligned}\tilde{l}_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) &= \int\int q^*(\boldsymbol{a},\boldsymbol{u}) \log\frac{f(\boldsymbol{y}_i, \boldsymbol{a}, \boldsymbol{u})}{q^*(\boldsymbol{a},\boldsymbol{u})}\,\mathrm{d}\boldsymbol{a}\,\mathrm{d}\boldsymbol{u} \\ &= \boldsymbol{y}_i^\top \log\Phi\left(\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\boldsymbol{\mu}_i\right) + (\mathbf{1}-\boldsymbol{y}_i)^\top \log\Phi\left(-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i\right) \\ &\quad + \frac{1}{2}\boldsymbol{\mu}_i^\top\mathbf{Z}_i^\top\mathbf{Z}_i\boldsymbol{\mu}_i \\ &\quad + \int\left(-\frac{1}{2}\boldsymbol{u}\mathbf{Z}_i^\top\mathbf{Z}_i\boldsymbol{u} + \log\phi(\boldsymbol{u};\mathbf{0},\boldsymbol{\Sigma}) - \log\phi\left(\boldsymbol{u};\boldsymbol{\mu}_i,\boldsymbol{\Lambda}_i\right)\right) \\ &\qquad\qquad \cdot q_2^*(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \\ &= \boldsymbol{y}_i^\top \log\Phi\left(\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\boldsymbol{\mu}_i\right) + (\mathbf{1}-\boldsymbol{y}_i)^\top \log\Phi\left(-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i\right) \\ &\quad + \frac{1}{2}\boldsymbol{\mu}_i^\top\mathbf{Z}_i^\top\mathbf{Z}_i\boldsymbol{\mu}_i - \frac{1}{2}|\boldsymbol{\Sigma}| + \frac{1}{2}|\boldsymbol{\Lambda}_i| \\ &\quad + \int\left(-\frac{1}{2}\boldsymbol{u}\mathbf{Z}_i^\top\mathbf{Z}_i\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}\boldsymbol{\Sigma}^{-1}\boldsymbol{u} + \frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu}_i)\boldsymbol{\Lambda}^{-1}(\boldsymbol{u}-\boldsymbol{\mu}_i)\right) \\ &\qquad\qquad \cdot q_2^*(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \\ &= \boldsymbol{y}_i^\top \log\Phi\left(\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\boldsymbol{\mu}_i\right) + (\mathbf{1}-\boldsymbol{y}_i)^\top \log\Phi\left(-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i\right) \\ &\quad - \frac{1}{2}|\boldsymbol{\Sigma}\mathbf{Z}_i^\top\mathbf{Z}_i + \mathbf{I}| - \frac{1}{2}\boldsymbol{\mu}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i\end{aligned}$$

as in Ormerod and Wand (2010).

Now, we can make a similar introduction of auxiliary variables in the probit GSM in Equation (11). Keeping with

$$\boldsymbol{A}_i \mid \boldsymbol{U}_i \sim N\left(\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i\boldsymbol{U}_i, \mathbf{I}\right)$$

Let

$$\boldsymbol{T}_i \mid \boldsymbol{A}_i \sim \exp\left(\boldsymbol{\delta}_i^\top \log \mathbf{1}_{\{\boldsymbol{A}_i = \eta_0(\boldsymbol{T}_i)\}} + (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \mathbf{1}_{\{\boldsymbol{A}_i < \eta_0(\boldsymbol{T}_i)\}}\right)$$

Then the joint density is

$$\begin{aligned} f(\boldsymbol{t}_i, \boldsymbol{a}_i, \boldsymbol{u}_i) &= \exp\Big(\boldsymbol{\delta}_i^\top \log \mathbf{1}_{\{\boldsymbol{a}_i = \eta_0(\boldsymbol{t}_i)\}} \\ &\qquad + (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \mathbf{1}_{\{\boldsymbol{a}_i < \eta_0(\boldsymbol{t}_i)\}} \\ &\qquad + \log \phi\left(\boldsymbol{a}_i; \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}_i, \mathbf{I}\right) + \log \phi(\boldsymbol{u}_i; \mathbf{0}, \boldsymbol{\Sigma})\Big) \\ &= \exp\left(\boldsymbol{\delta}_i^\top \log \mathbf{1}_{\{\boldsymbol{a}_i = 0\}} + (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \mathbf{1}_{\{\boldsymbol{a}_i < 0\}}\right. \\ &\qquad \left. + \log \phi\left(\boldsymbol{a}_i; -\eta_0(\boldsymbol{t}_i) + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}_i, \mathbf{I}\right) + \log \phi(\boldsymbol{u}_i; \mathbf{0}, \boldsymbol{\Sigma})\right) \end{aligned}$$

where we shift the $\boldsymbol{A}_i$ elements by $-\eta_0(\boldsymbol{t}_i)$. Proceeding like before gives us the following update formulas

$$\begin{aligned} \log q_1^*(\boldsymbol{A}_i) \propto \int &\left(\boldsymbol{\delta}_i^\top \log \mathbf{1}_{\{\boldsymbol{A}_i = 0\}} + (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \mathbf{1}_{\{\boldsymbol{A}_i < 0\}}\right. \\ &\qquad \left. + \log \phi\left(\boldsymbol{a}_i; -\eta_0(\boldsymbol{t}_i) + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}, \mathbf{I}\right)\right) \phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\, \mathrm{d}\boldsymbol{u} \end{aligned}$$

where each element either follows a degenerate distribution with point mass at zero or follows a right truncated normal distribution. As before, the second update yields a multivariate normal distribution a mean and a covariance matrix like in Equation (13). The estimated density at convergence is

$$\begin{aligned} \log q^*(\boldsymbol{a}_i, \boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) = \boldsymbol{\delta}_i^\top &\log \mathbf{1}_{\{\boldsymbol{a}_i = \mathbf{0}\}} \\ &+ (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log\left(\frac{\mathbf{1}_{\{\boldsymbol{a}_i < \mathbf{0}\}}}{\Phi\left(\eta_0(\boldsymbol{t}_i) - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)}\right) \\ &+ \log \phi\left(\mathbf{K}_i\boldsymbol{a}_i; \mathbf{K}_i(-\eta_0(\boldsymbol{t}_i) + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}_i), \mathbf{I}\right) \\ &+ \log \phi\left(\boldsymbol{u}_i; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i\right) \end{aligned}$$

where $\mathbf{K}_i \in \{0,1\}^{(n_i - \sum_k \delta_{ik}) \times n_i}$ is a matrix with a subset of the n_i unit vectors where unit vector k is included if $\delta_{ik} = 0$. Let $\boldsymbol{k}_i = \eta_0(\boldsymbol{t}_i) - \mathbf{X}_i\boldsymbol{\beta}$ and $\mathbf{L}_i$ be similar to $\mathbf{K}_i$ but for $\delta_{ik} = 1$. Then the lower bound becomes

$$
\begin{aligned}
\tilde{l}_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i) &= (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \Phi\left(\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i\right) \\
&\quad + \int \int \big(\log \phi\left(\boldsymbol{a}; -\boldsymbol{k}_i + \mathbf{Z}_i \boldsymbol{u}, \mathbf{I}\right) + \log \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma}) \\
&\qquad - \log \phi\left(\mathbf{K}_i \boldsymbol{a}; \mathbf{K}_i(-\boldsymbol{k}_i + \mathbf{Z}_i \boldsymbol{\mu}_i), \mathbf{I}\right) \\
&\qquad - \log \phi\left(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i\right)\big)\, q^*(\boldsymbol{a}, \boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\, \mathrm{d}\boldsymbol{a}\, \mathrm{d}\boldsymbol{u} \\
&= (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \Phi\left(\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i\right) - \frac{\sum_k \delta_{ik}}{2} \log 2\pi \\
&\quad + \int \int -\frac{1}{2}(\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{u})^\top \mathbf{L}_i^\top \mathbf{L}_i (\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{u}) \\
&\qquad - \frac{1}{2}(\boldsymbol{a} + \boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{u})^\top \mathbf{K}_i^\top \mathbf{K}_i (\boldsymbol{a} + \boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{u}) \\
&\qquad + \frac{1}{2}(\boldsymbol{a} + \boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i)^\top \mathbf{K}_i^\top \mathbf{K}_i (\boldsymbol{a} + \boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i) \\
&\qquad + \log \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma}) - \log \phi\left(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i\right)\big) \\
&\qquad \cdot q^*(\boldsymbol{a}, \boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\, \mathrm{d}\boldsymbol{a}\, \mathrm{d}\boldsymbol{u} \\
&= (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \Phi\left(\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i\right) - \frac{\sum_k \delta_{ik}}{2} \log 2\pi \\
&\quad + \frac{1}{2} \boldsymbol{\mu}_i \mathbf{Z}_i^\top \mathbf{K}_i^\top \mathbf{K}_i \mathbf{Z}_i \boldsymbol{\mu}_i \\
&\quad + \int \bigg(-\frac{1}{2}(-\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{u})^\top \mathbf{L}_i^\top \mathbf{L}_i (-\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{u}) \\
&\qquad - \frac{1}{2} \boldsymbol{u} \mathbf{Z}_i^\top \mathbf{K}_i^\top \mathbf{K}_i \mathbf{Z}_i \boldsymbol{u} \\
&\qquad + \log \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma}) - \log \phi\left(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i\right)\big)\, q_2^*(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} \\
&= (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \Phi\left(\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i\right) + \boldsymbol{\delta}_i^\top \log \phi\left(\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i\right) \\
&\quad - \frac{1}{2} \operatorname{tr}(\mathbf{Z}_i^\top \mathbf{L}_i^\top \mathbf{L}_i \mathbf{Z}_i \boldsymbol{\Lambda}_i) - \frac{1}{2} \operatorname{tr}(\mathbf{Z}_i^\top \mathbf{K}_i^\top \mathbf{K}_i \mathbf{Z}_i \boldsymbol{\Lambda}_i) \\
&\quad - \frac{1}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda}_i - \frac{1}{2} \boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i \\
&\quad + \frac{1}{2} \log|\boldsymbol{\Lambda}_i| + \frac{1}{2} \operatorname{tr} \mathbf{I} \\
&= (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log \Phi\left(\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i\right) + \boldsymbol{\delta}_i^\top \log \phi\left(\boldsymbol{k}_i - \mathbf{Z}_i \boldsymbol{\mu}_i\right) \\
&\quad - \frac{1}{2} |\boldsymbol{\Sigma} \mathbf{Z}_i^\top \mathbf{Z}_i + \mathbf{I}| - \frac{1}{2} \boldsymbol{\mu}_i^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i
\end{aligned}
$$

It may be worth considering the results by Consonni and Marin (2007) regarding the poor performance of the auxiliary variables approximation in the probit regression model in a Bayesian analysis setting.

3 Accelerated Failure Time Models

In this section, we will consider accelerated failure time (AFT) models with shared random random effects. We define the conditional survival function and the corresponding conditional hazard function as

$$\begin{aligned}
S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= S_0(\exp(-\boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{u})t) \\
\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \lambda_0(\exp(-\boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{u})t)\exp(-\boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{u})
\end{aligned}$$

where S_0 and λ_0 are a given survival function and the corresponding hazard function. Inserting the above into the marginal log-likelihood in Equation (5) yields

$$\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = &-\boldsymbol{\delta}_i^\top \mathbf{X}_i\boldsymbol{\beta} \\
&+ \log\Bigg(\int \exp\big(-\boldsymbol{\delta}_i^\top\mathbf{Z}_i\boldsymbol{u} + \boldsymbol{\delta}_i^\top \log\lambda_0(\exp(-\mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u})\cdot \boldsymbol{t}_i) \\
&\qquad + \mathbf{1}^\top \log S_0(\exp(-\mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{u})\cdot\boldsymbol{t}_i)\big) \\
&\qquad \cdot \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg)
\end{aligned} \tag{14}$$

Whether or not this yields a tractable lower bound depends on the log survival function, $\log S_0$, and log hazard function, $\log\lambda_0$. For instance, if we use the Weibull distribution

$$\begin{aligned}
S_0(t; \lambda, \gamma) &= \exp(-\lambda t^\gamma) \\
\lambda_0(t; \lambda, \gamma) &= \gamma\lambda t^{\gamma-1}
\end{aligned}$$

then

$$\begin{aligned}
\log S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})) &= -\lambda t^\gamma \exp(-\boldsymbol{x}^\top(\gamma\boldsymbol{\beta}) - \boldsymbol{z}^\top(\gamma\boldsymbol{u})) \\
\log\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \log\gamma\lambda + (\gamma - 1)\log t - \boldsymbol{x}^\top(\gamma\boldsymbol{\beta}) - \boldsymbol{z}^\top(\gamma\boldsymbol{u})
\end{aligned}$$

which is equivalent to the log-log link function in Section 2.1 as expected and thus it gives a tractable lower bound.

Choosing a log-logistic model instead implies that the conditional log survival function and log hazard function are

$$\begin{aligned}
\log S_0(t; \lambda, \gamma) &= -\log(1 + \lambda t^\gamma) \\
\log\lambda_0(t; \lambda, \gamma) &= \log\lambda\gamma + (\gamma - 1)\log t - \log(1 + \lambda t^\gamma)
\end{aligned}$$

Consequently,

$$\begin{aligned}
\log S_0(\exp(-\boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{u})t) &= -\log\big(1 + \lambda t^\gamma\exp(-\boldsymbol{x}^\top(\gamma\boldsymbol{\beta}) - \boldsymbol{z}^\top(\gamma\boldsymbol{u}))\big) \\
\log\lambda_0(\exp(-\boldsymbol{x}^\top\boldsymbol{\beta} - \boldsymbol{z}^\top\boldsymbol{u})t) &= \log\lambda\gamma + (\gamma - 1)\log t \\
&\quad - (\gamma - 1)(\boldsymbol{x}^\top(\gamma\boldsymbol{\beta}) + \boldsymbol{z}^\top(\gamma\boldsymbol{u})) \\
&\quad - \log\big(1 + \lambda t^\gamma \exp(-\boldsymbol{x}^\top(\gamma\boldsymbol{\beta}) - \boldsymbol{z}^\top(\gamma\boldsymbol{u}))\big)
\end{aligned}$$

Inserting the above into the marginal log-likelihood in Equation (14) and deriving the lower bound shows that we need to perform approximations of the terms of the following form

$$\int -\log\left(1 + \lambda t^\gamma \exp\left(-\boldsymbol{x}_{ik}^\top(\gamma\boldsymbol{\beta}) - \boldsymbol{z}_{ik}^\top(\gamma\boldsymbol{u})\right)\right)\phi(\boldsymbol{u};\boldsymbol{\mu}_i,\boldsymbol{\Lambda}_i)\,\mathrm{d}\boldsymbol{u}$$

as with the negative logit function in Section 2.2 as expected since setting

$$\frac{1 - S_0(t)}{S_0(t)} = \lambda t^\gamma$$

and re-scaling $\boldsymbol{\beta}$ and $\boldsymbol{u}$ yields the GSM in Section 2.2.

4 GSMs with Time-Varying Effects

The assumption of a constant covariate effect or random effect on the link scale in Section 2 may not be appropriate. Thus, we will consider generalizations to time-varying covariate effects and random effects in this section.

4.1 Log-log Link

Consider again the log-log link function in section 2.1 where

$$\begin{aligned}
&\log S(t \mid \boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) = \exp(\boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u})\log S_0(t) \\
&\log\lambda(t \mid \boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) = L_0(t) + \boldsymbol{x}^\top\boldsymbol{\beta} + \boldsymbol{z}^\top\boldsymbol{u} \qquad L_0(t) = \log -\frac{S_0'(t)}{S_0(t)}
\end{aligned}$$

We generalize this to time-varying fixed effects by letting

$$\begin{aligned}
&\log S(t \mid \boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) = \exp(\boldsymbol{z}^\top\boldsymbol{u})\log S_0(t;\boldsymbol{x}) && S_0(t;\boldsymbol{x}) = S_0(t)^{\exp((\boldsymbol{B}(t)\otimes\boldsymbol{x})^\top\boldsymbol{\beta})} \\
&\log\lambda(t \mid \boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) = L_0(t;\boldsymbol{x}) + \boldsymbol{z}^\top\boldsymbol{u} && L_0(t;\boldsymbol{x}) = \log -\frac{S_0'(t;\boldsymbol{x})}{S_0(t;\boldsymbol{x})}
\end{aligned}$$

where $\otimes$ denotes the Kronecker product and $\boldsymbol{B}(t)$ is a basis expansion. To ease the notation, we assume that all covariates has a time-varying effect using the same basis expansion. The $(\boldsymbol{B}(t)\otimes\boldsymbol{x})^\top\boldsymbol{\beta}$ form is restrictive and can of course be replaced by a general function $h(t;\boldsymbol{x})$. Using similar arguments as before, we find a tractable lower bound as the marginal log-likelihood is

$$\begin{aligned}
l_i(\boldsymbol{\beta},\boldsymbol{\Sigma}) = \boldsymbol{\delta}_i^\top L_0(\boldsymbol{t}_i;\mathbf{X}_i)) + \log\Bigg(\int \exp\big(\boldsymbol{\delta}_i^\top\mathbf{Z}_i\boldsymbol{u} \\
+\mathbf{1}^\top\left(\log S_0(\boldsymbol{t}_i;\mathbf{X}_i)\cdot\exp(\mathbf{Z}_i\boldsymbol{u})\right)\big)\,\phi(\boldsymbol{u};0,\boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg)
\end{aligned}$$

which shows that the lower bound is still tractable as we know the mean and have a closed form solution for the moment generating function. We can further find that

$$
\begin{aligned}
h(t;\boldsymbol{x}) &= (\boldsymbol{B}(t)\otimes\boldsymbol{x})^\top\boldsymbol{\beta} \\
S_0'(t\mid\boldsymbol{x}) &= \exp(g(t;\boldsymbol{x}))S_0(t)^{h(t;\boldsymbol{x})-1}\left(S_0'(t)+S_0(t)\log(S_0(t))h'(t;\boldsymbol{x})\right) \\
\exp L_0(t;\boldsymbol{x}) &= -\exp(h(t;\boldsymbol{x}))S_0(t)^{-1}\left(S_0'(t)+S_0(t)\log(S_0(t))h'(t;\boldsymbol{x})\right) \\
&= \exp\left(\log(-\log S_0(t)+h(t;\boldsymbol{x}))\right) \\
&\qquad\cdot\left((\log(-\log S_0(t)))'+h'(t;\boldsymbol{x})\right) \\
\Leftrightarrow \qquad L_0(t;\boldsymbol{x}) &= \log(-\log S_0(t))+h(t;\boldsymbol{x}) \\
&\qquad+\log\left((\log(-\log S_0(t)))'+h'(t;\boldsymbol{x})\right)
\end{aligned}
$$

which also follows from the more general result

$$
\begin{aligned}
g\left(S(t\mid\boldsymbol{x},\boldsymbol{z},\boldsymbol{u})\right) &= \eta(t;\boldsymbol{x},\boldsymbol{u}) \\
\lambda(t\mid\boldsymbol{x},\boldsymbol{u}) &= -\frac{(g^-)'(\eta(t;\boldsymbol{x},\boldsymbol{u}))}{g^-(\eta(t;\boldsymbol{x},\boldsymbol{u}))}\eta'(t;\boldsymbol{x},\boldsymbol{u}))
\end{aligned}
$$

Using the above results, we can consider extending the model to time-varying random effects as well. Particularity, we consider the model

$$
\log S(t\mid\boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) = \exp((\boldsymbol{M}(t)\otimes\boldsymbol{z})^\top\boldsymbol{u})\log S_0(t;\boldsymbol{x})
$$

where we again use the same basis expansion, $\boldsymbol{M}(t)$, for all random effect covariates. This is similar to the subsurvial model used in Crowther (2014, section 7), Lawrence Gould et al. (2015, section 2.2.2). This leads to the following log hazard function

$$
\begin{aligned}
k(t;\boldsymbol{z},\boldsymbol{u}) &= (\boldsymbol{M}(t)\otimes\boldsymbol{z})^\top\boldsymbol{u} \\
\log\lambda(t\mid\boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) &= \log(-\log S_0(t))+h(t;\boldsymbol{x})+k(t;\boldsymbol{z},\boldsymbol{u}) \\
&\qquad+\log\left((\log(-\log S_0(t)))'+h'(t;\boldsymbol{x})+k'(t;\boldsymbol{z},\boldsymbol{u})\right)
\end{aligned}
$$

and the following marginal log-likelihood

$$
\begin{aligned}
l_i(\boldsymbol{\beta},\boldsymbol{\Sigma}) &= \boldsymbol{\delta}_i^\top\left(\log(-\log S_0(\boldsymbol{t}_i))+h(\boldsymbol{t}_i;\mathbf{X}_i)\right) \\
&\quad+\log\Bigg(\int\exp\Big(\boldsymbol{\delta}_i^\top(\boldsymbol{M}(\boldsymbol{t}_i)^\top\otimes\mathbf{Z}_i)\boldsymbol{u} \\
&\qquad\qquad+\boldsymbol{\delta}_i^\top\log\left((\log(-\log S_0(\boldsymbol{t}_i)))'+h'(\boldsymbol{t}_i;\mathbf{X}_i)+k'(\boldsymbol{t}_i;\mathbf{Z}_i,\boldsymbol{u})\right) \\
&\qquad\qquad+\mathbf{1}^\top\left(\log S_0(\boldsymbol{t}_i;\mathbf{X})\cdot\exp((\boldsymbol{M}(\boldsymbol{t}_i)^\top\otimes\mathbf{Z}_i)\boldsymbol{u})\right)\Big) \\
&\qquad\qquad\cdot\phi(\boldsymbol{u};0,\boldsymbol{\Sigma})\,\mathrm{d}\boldsymbol{u}\Bigg)
\end{aligned}
$$

where the derivative superscripts only implies the diagonal of the Jacobian matrix. This yields additional terms in the lower bound which are potentially intractable integrals. Whether or not the integral is intractable depends on $S_0(t)$, $\boldsymbol{B}(t)$, and $\boldsymbol{M}(t)$.

As an alternative, we can consider a generalization of the proportional hazard model similar to the survival submodel in Lawrence Gould et al. (2015, section 2.2.1), Crowther (2014, section 6), and Crowther et al. (2016). That is, we define the log hazard function as

$$\log \lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) = \log \lambda_0(t) + (\boldsymbol{B}(t) \otimes \boldsymbol{x})^\top \boldsymbol{\beta} + (\boldsymbol{M}(t) \otimes \boldsymbol{z})^\top \boldsymbol{u} \tag{15}$$

such that the log survival function is

$$\begin{aligned}\log S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= -\int_0^t \lambda(s \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})\, \mathrm{d}s \\ &= -\int_0^t \lambda_0(s) \exp\left((\boldsymbol{B}(s) \otimes \boldsymbol{x})^\top \boldsymbol{\beta} + (\boldsymbol{M}(s) \otimes \boldsymbol{z})^\top \boldsymbol{u}\right) \mathrm{d}s \qquad (16)\end{aligned}$$

A related model is used by Yue and Kontar (2019) who use a Gaussian process in a joint model with a related survival submodel as in Equation (15). The terms we derive next are similar to the terms derived by Yue and Kontar (2019). Inserting Equation (15) and (16) into the marginal log-likelihood in Equation (5) yields

$$\begin{aligned}l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= \log\Bigg(\int \exp\big(\boldsymbol{1}^\top(\boldsymbol{\delta}_i \cdot \log \lambda(\boldsymbol{t}_i \mid \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u})) \\ &\qquad\qquad + \boldsymbol{1}^\top \log S(\boldsymbol{t}_i \mid \mathbf{X}_i, \boldsymbol{Z}_i, \boldsymbol{u})\big)\, \phi(\boldsymbol{u}; \boldsymbol{0}, \boldsymbol{\Sigma})\, \mathrm{d}\boldsymbol{u}\Bigg) \\ &= \boldsymbol{\delta}_i^\top \left(\log \lambda_0(\boldsymbol{t}_i) + (\boldsymbol{B}(\boldsymbol{t}_i)^\top \otimes \mathbf{X}_i)\boldsymbol{\beta}\right) \\ &\quad + \log\Bigg(\int \exp\big(\boldsymbol{\delta}_i^\top (\boldsymbol{M}(\boldsymbol{t}_i)^\top \otimes \mathbf{Z}_i)\boldsymbol{u} \\ &\qquad\qquad + \boldsymbol{1}^\top \log S(\boldsymbol{t}_i \mid \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u})\big)\, \phi(\boldsymbol{u}; \boldsymbol{0}, \boldsymbol{\Sigma})\, \mathrm{d}\boldsymbol{u}\Bigg)\end{aligned}$$

Unlike before, the terms from the log hazard are tractable in the lower bound but the terms from log survival function are not in general. Let

$$\lambda_0(t; \boldsymbol{x}) = \lambda_0(t) \exp\left((\boldsymbol{B}(s) \otimes \boldsymbol{x})^\top \boldsymbol{\beta}\right)$$

Then the latter terms in the lower bound from log survival function are of the following form

$$\begin{aligned}&\int \log S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) \phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\, \mathrm{d}\boldsymbol{u} \\ &\qquad = -\int\int_0^t \exp\left(\log \lambda_0(s; \boldsymbol{x}) + (\boldsymbol{M}(s) \otimes \boldsymbol{z})^\top \boldsymbol{u}\right) \phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\, \mathrm{d}s\, \mathrm{d}\boldsymbol{u} \\ &\qquad = -\int_0^t \lambda_0(s; \boldsymbol{x}) \left(\int \exp\left((\boldsymbol{M}(s) \otimes \boldsymbol{z})^\top \boldsymbol{u}\right) \phi(\boldsymbol{u}; \boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i)\, \mathrm{d}\boldsymbol{u}\right) \mathrm{d}s\end{aligned}$$

where we assume that order of integration can be exchanged. The inner integral is the moment generating function of the multivariate normal distribution which has a closed form solution. Hence, the intractable integrals consist of univariate integrals over time. This is a major advantage from the marginal log-likelihood where one, in general, has nested integrals both of which requires numerical approximations.

In general, suppose that we let

$$\begin{aligned}
S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= c\left(\eta(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})\right) \\
\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \frac{-c'\left(\eta(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})\right)}{c\left(\eta(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}\right)}\eta'(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})
\end{aligned}$$

where η is a strictly increasing function (in the first argument) and c is strictly decreasing function which range is $[0, 1]$. Inserting the above survival function and hazard function into the marginal log-likelihood in Equation (5) yields

$$\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \log \Bigg(\int \exp \big(&\boldsymbol{\delta}_i^\top \log(-c'\left(\eta(\boldsymbol{t}_i; \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u})\right)) + \boldsymbol{\delta}_i^\top \log \eta'(\boldsymbol{t}_i; \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u}) \\
&+ (\mathbf{1} - \boldsymbol{\delta}_i)^\top \log c\left(\eta(\boldsymbol{t}_i; \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u})\right)\big) \, \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma}) \, \mathrm{d}\boldsymbol{u} \Bigg)
\end{aligned} \tag{17}$$

This may yield up to three types integrals in the lower bound depending on the c and η functions.

5 AFT Models with Time-Varying Effects

Similar to Section 4, we consider AFT models with time-varying effects in this section. In particular, we consider the model

$$\begin{aligned}
\exp \eta_1(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \int_0^t \exp \eta(s; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) \, \mathrm{d}s \\
S(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= S_0 \left(\int_0^t \exp \eta(s; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) \, \mathrm{d}s \right) \\
&= S_0 \left(\exp \eta_1(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) \right) \\
\lambda(t \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) &= \lambda_0 \left(\exp \eta_1(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) \right) \eta_1'(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u}) \exp \eta_1(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})
\end{aligned}$$

where we specify the $\eta_1(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})$ function and the survival function S_0. This is a special case of the model which marginal log-likelihood is shown in Equation (17). Inserting the above survival function and hazard function into the marginal log-likelihood in Equation (5) yields

$$
\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \log \Bigg(\int \exp \big(& \boldsymbol{\delta}_i^\top \log \lambda_0 \left(\exp \eta_1(\boldsymbol{t}_i; \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u}) \right) \\
& + \boldsymbol{\delta}_i^\top \eta_1(\boldsymbol{t}_i; \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u}) + \boldsymbol{\delta}_i^\top \log \eta_1'(\boldsymbol{t}_i; \mathbf{X}_i, \boldsymbol{Z}_i, \boldsymbol{u}) \\
& + \mathbf{1}^\top \log S_0 \left(\exp \eta_1(\boldsymbol{t}_i; \mathbf{X}_i, \boldsymbol{Z}_i, \boldsymbol{u}) \right) \big) \, \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma}) \, \mathrm{d}\boldsymbol{u} \Bigg)
\end{aligned} \tag{18}
$$

Whether or not this yields a tractable lower bound depends on the log survival function, $\log S_0$, the corresponding log hazard function, $\log \lambda_0$, and the $\eta_1(t; \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u})$ function and the logarithm of its derivative.

As an example, we can choose the survival function and corresponding hazard function as

$$
\begin{aligned}
S_0(t) &= \exp(-\exp(s(\log t))) \\
\lambda_0(t) &= \exp(s(\log t)) \frac{s'(\log t)}{t}
\end{aligned}
$$

where s is a smooth function. Substituting into the marginal log-likelihood in Equation (18) yields

$$
\begin{aligned}
l_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \log \Bigg(\int \exp \big(& \boldsymbol{\delta}_i^\top s(\eta_1(\boldsymbol{t}_i; \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u})) + \boldsymbol{\delta}_i^\top \log s'(\eta_1(\boldsymbol{t}_i; \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{u})) \\
& + \boldsymbol{\delta}_i^\top \log \eta_1'(\boldsymbol{t}_i; \mathbf{X}_i, \boldsymbol{Z}_i, \boldsymbol{u}) \\
& - \mathbf{1}^\top \exp s \left(\eta_1(\boldsymbol{t}_i; \mathbf{X}_i, \boldsymbol{Z}_i, \boldsymbol{u}) \right) \big) \, \phi(\boldsymbol{u}; \mathbf{0}, \boldsymbol{\Sigma}) \, \mathrm{d}\boldsymbol{u} \Bigg)
\end{aligned}
$$

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.

Consonni, G. and Marin, J.-M. (2007). Mean-field variational approximate bayesian inference for latent variable models. *Computational Statistics & Data Analysis*, 52(2):790 – 798.

Crowther, M. J. (2014). *Development and application of methodology for the parametric analysis of complex survival and joint longitudinal-survival data in biomedical research.* PhD dissertation, University of Leicester.

Crowther, M. J., Andersson, T. M.-L., Lambert, P. C., Abrams, K. R., and Humphreys, K. (2016). Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Statistics in Medicine*, 35(7):1193–1209.

Jaakkola, T. S. and Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10(1):25–37.

Lawrence Gould, A., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., and Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. report of the dia bayesian joint modeling working group. *Statistics in Medicine*, 34(14):2181–2195.

Ormerod, J. and Wand, M. (2008). Variational approximations for logistic mixed models. In *Proceedings of the Ninth Iranian Statistics Conference*, pages 450–467. Department of Statistics, University of Isfahan, Isfahan, Iran.

Ormerod, J. T. and Wand, M. P. (2010). Explaining variational approximations. *The American Statistician*, 64(2):140–153.

Ormerod, J. T. and Wand, M. P. (2012). Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 21(1):2–17.

Yue, X. and Kontar, R. (2019). Variational Inference of Joint Models using Multivariate Gaussian Convolution Processes. *arXiv e-prints*, page arXiv:1903.03867.