Application: Gestational age

• Although many epidemiologic studies of preterm birth focus on a 0/1 indicator, gestational age is more accurately modeled as ordered categorical/discrete event time.

• Consider data from a study of pregnancy outcomes \( (n = 1,000) \).

• The outcome variable is the week of delivery & predictors include 0/1 indicators of bleeding, bacterial vaginosis, and intense physical activity.
• We defined a continuation-ratio probit model for the hazard of birth at week $j$ of gestation (standardized to $\{1, \ldots, 18\}$):

$$
\Pr(T_i = j \mid T_i \geq j, x_i) = \Phi(\beta_{1j} + x_i'\beta_2) = \Phi(x_i'\beta),
$$

where $\beta_1 = (\beta_{11}, \ldots, \beta_{1,17})'$ characterize the baseline hazard, $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})'$ are regression coefficients, & $\beta = (\beta_1', \beta_2')'$.

• Chose a multivariate normal prior for $\beta$ with covariance $I_{20 \times 20}$. 

• Ran data augmentation Gibbs sampler 5,000 iterations, discarding first 1,000 iterations as a burn-in.

• Used samples to estimate baseline gestation age distribution among the reference group of women, with $x_i = 0$. 
Baseline gestational age distribution

![Graph showing baseline gestational age distribution. The x-axis represents gestational length, ranging from 30 to 40 weeks. The y-axis represents probability, ranging from 0.0 to 0.25. The graph indicates a peak in probability around the 36th week of gestation.](image-url)
Posterior summaries & mles of parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>MLE</th>
<th>Median</th>
<th>SD</th>
<th>95% credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>week 27</td>
<td>-2.69</td>
<td>-2.74</td>
<td>-2.69</td>
<td>0.16</td>
<td>[-3 -2.38]</td>
</tr>
<tr>
<td>week 28</td>
<td>-3.04</td>
<td>-3.18</td>
<td>-3.03</td>
<td>0.27</td>
<td>[-3.61 -2.54]</td>
</tr>
<tr>
<td>week 29</td>
<td>-2.85</td>
<td>-2.97</td>
<td>-2.84</td>
<td>0.18</td>
<td>[-3.23 -2.52]</td>
</tr>
<tr>
<td>week 30</td>
<td>-2.86</td>
<td>-2.96</td>
<td>-2.86</td>
<td>0.17</td>
<td>[-3.21 -2.55]</td>
</tr>
<tr>
<td>week 31</td>
<td>-2.55</td>
<td>-2.59</td>
<td>-2.54</td>
<td>0.13</td>
<td>[-2.8 -2.3]</td>
</tr>
<tr>
<td>week 32</td>
<td>-2.25</td>
<td>-2.27</td>
<td>-2.25</td>
<td>0.1</td>
<td>[-2.45 -2.05]</td>
</tr>
<tr>
<td>week 33</td>
<td>-2.32</td>
<td>-2.36</td>
<td>-2.31</td>
<td>0.12</td>
<td>[-2.56 -2.09]</td>
</tr>
<tr>
<td>week 34</td>
<td>-2.03</td>
<td>-2.06</td>
<td>-2.03</td>
<td>0.09</td>
<td>[-2.21 -1.86]</td>
</tr>
<tr>
<td>week 35</td>
<td>-1.84</td>
<td>-1.86</td>
<td>-1.84</td>
<td>0.08</td>
<td>[-1.99 -1.68]</td>
</tr>
<tr>
<td>week 36</td>
<td>-1.75</td>
<td>-1.77</td>
<td>-1.75</td>
<td>0.08</td>
<td>[-1.9 -1.6]</td>
</tr>
<tr>
<td>week 37</td>
<td>-1.38</td>
<td>-1.4</td>
<td>-1.38</td>
<td>0.07</td>
<td>[-1.51 -1.25]</td>
</tr>
<tr>
<td>week 38</td>
<td>-0.88</td>
<td>-0.9</td>
<td>-0.88</td>
<td>0.06</td>
<td>[-0.99 -0.78]</td>
</tr>
<tr>
<td>week 39</td>
<td>-0.28</td>
<td>-0.3</td>
<td>-0.28</td>
<td>0.05</td>
<td>[-0.39 -0.18]</td>
</tr>
<tr>
<td>week 40</td>
<td>0.16</td>
<td>0.15</td>
<td>0.16</td>
<td>0.07</td>
<td>[0.03 0.3]</td>
</tr>
<tr>
<td>week 41</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.12</td>
<td>[0.59 1.06]</td>
</tr>
<tr>
<td>week 42</td>
<td>1.32</td>
<td>1.42</td>
<td>1.31</td>
<td>0.31</td>
<td>[0.74 1.99]</td>
</tr>
<tr>
<td>week 43</td>
<td>-0.08</td>
<td>-0.19</td>
<td>-0.07</td>
<td>0.67</td>
<td>[-1.4 1.25]</td>
</tr>
<tr>
<td>bleed</td>
<td>0.15</td>
<td>0.17</td>
<td>0.15</td>
<td>0.05</td>
<td>[0.06 0.24]</td>
</tr>
<tr>
<td>bac vag</td>
<td>0.03</td>
<td>0.05</td>
<td>0.03</td>
<td>0.05</td>
<td>[-0.06 0.12]</td>
</tr>
<tr>
<td>phys act</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>[-0.05 0.15]</td>
</tr>
</tbody>
</table>
Some Conclusions

• The mles and posterior means are quite similar even though the number of parameters is not small.

• Bleeding during pregnancy predicts a shorter gestational age, \( \Pr(\beta_{21} > 0 \mid \text{data}) > 0.99. \)

• Limited evidence of either bacterial vaginosis or physical activity having an impact on gestational age.

• Important to consider possible time-varying effects.
Time-Varying Covariates

Often, in applications one or more of the predictors may vary over time.

For example, suppose that we are interested in assessing the effect of air pollution levels on mortality.

The air pollution levels vary from day to day.
Reasonable model for discrete hazard of death in age group $j$?

$$\Pr(T_i = j \mid T_i = j, x_i, z_{ij}) = h(\alpha_j + x'_i\beta + z_{ij}\psi),$$

where $z_{ij}$ is the level of population for individual $i$ at age $j$

This model accommodates the time-varying covariate

Are we making a restrictive assumption?
In the previous model, we assumed that the effect of air pollution was constant at different ages.

In fact, infants and the elderly are more susceptible to pollution-induced mortality.

How can we generalize the model, to account for this age-dependent susceptibility?
Time-Varying Coefficients

\[ Pr(T_i = j \mid T_i = j, x_i, z_{ij}) = h(\alpha_j + x_i'\beta_j + z_{ij}\psi_j), \]

where we have now added a \( j \) subscript to the parameter \( \psi \) characterizing the air pollution effect.

Potentially, the effect of the time-independent predictors can also vary with time by allowing different \( \beta \)s for the different age intervals.

Dimensionality rapidly becomes problematic - Order Restrictions?
• What about the prior specification?

• We have a potentially high-dimensional vector of time-varying baseline parameters and coefficients.

• Potentially, no individuals with the event in certain intervals.

• What type of information prior is reasonable?
Time-Varying Coefficients

\[ \Pr(T_i = j \mid T_i = j, x_i, z_{ij}) = h(\alpha_j + x'_i \beta + z_{ij} \psi_j), \]

where we have now added a \( j \) subscript to the parameter \( \psi \) characterizing the air pollution effect.

Potentially, the effect of the time-independent predictors can also vary with time by allowing different \( \beta \)s for the different age intervals.

Dimensionality rapidly becomes problematic - Order Restrictions?
• What about the prior specification?

• We have a potentially high-dimensional vector of time-varying baseline parameters and coefficients.

• Potentially, no individuals with the event in certain intervals.

• What type of information prior is reasonable?
• Focusing initially on the model with no time-varying coefficients, we may want to do some smoothing.

• Values of $\alpha_j$ and $\alpha_{j'}$ are likely to be similar if $j$ is close to $j'$. 

• Autoregressive - Gaussian random walk prior:

$$\alpha_j \sim N(\alpha_{j-1}, \tau^{-1}),$$

where $\tau$ is a precision parameter controlling the degree of smoothing.
Penalized Likelihood

• The autoregressive prior essentially penalizes values of $\alpha_j$ that are far from the neighboring values.

• From a frequentist perspective, we can use a similar idea by including a penalty term in the likelihood and then maximizing the resulting penalized likelihood.

• The penalty term can follow many forms, including an autoregressive normal density for the $\alpha$s.
• Often, the degree of smoothing (controlled by the $\tau$ parameter) may be quite subjective

• Potentially, smoothing may obscure real features of the baseline hazard or time-varying coefficient function

• Order-restrictions are a powerful alternative
Some Types of Parameter Restrictions

- **Simple Order**: $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_k$.

- **Umbrella Order**: $\alpha_1 \leq \ldots \leq \alpha_{j^*} \geq \ldots \geq \alpha_k$, where $j^* \in \{1, \ldots, k\}$ is an unknown changepoint.

- **Bathtub Order**: $\alpha_1 \geq \ldots \geq \alpha_{j^*} \leq \ldots \leq \alpha_k$.

Let $\Omega \subset \mathbb{R}^k$ denote the order-restricted space.
Truncated Conjugate Priors for Parameter Constraints

Gelfand, Smith and Lee (1992)

Choose a prior ignoring the parameter constraint

Proceed with the Gibbs sampler as in the unconstrained case

Discard draws inconsistent with the order constraint