1. GLM Basics

(a) Definition of exponential family,

\[ f(y_i; \theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}, \]

where \( \theta_i, \phi \) are location and scale parameters, respectively. Mean and variance are

\[ E(y_i) = b'(\theta_i) \quad \text{and} \quad V(y_i) = b''(\theta_i)a(\phi). \]

(b) Systematic component, link functions:

\[ \eta_i = g(\mu_i) = x_i' \beta, \]

where \( \eta_i \) is the linear predictor, \( g(\cdot) \) is the link function, and \( x_i \) are predictors

(c) Canonical link: \( \theta_i = \eta_i \) - linear, logistic, log for normal, Bernoulli, & Poisson, respectively.
2. Basics of Frequentist Inference

(a) Maximize likelihood using iteratively reweighted least squares (don’t need to memorize details!)

(b) Analysis of deviance:

- Scaled deviance = $2 \times$ difference between log likelihoods for saturated model and current model.

- Exact $\chi^2_{n-p}$ distribution for normal data, otherwise normal approximation - often poor.

- However, the $\chi^2$ approximation typically works well for the difference in scaled deviance between nested models.

(c) Standard errors and confidence intervals can be based on normal approximation at MLE.

(d) For variable selection, stepwise procedures are often used - forward, backward, goodness of fit criteria, disadvantages?
3. Basics of Bayesian Inference in GLMs

(a) Definition of prior, posterior distribution, marginal likelihood, posterior probability, credible intervals, and Bayes factors.

(b) Should be able to derive posterior distributions for regression parameters and error precision in normal linear models with conjugate priors.

(c) Basics of implementing a Markov chain Monte Carlo algorithm with Gibbs and/or Metropolis-Hastings steps.

(d) Commonly used algorithms for Gibbs sampling in GLMs - adaptive rejection sampling, data augmentation - basic details (don’t need to memorize steps in adaptive rejection sampling) and motivation.

(e) How to do inferences based on the posterior in applied problems.
4. Latent Variable Models for Binary Data

(a) How to induce a regression model on a binary or categorical response by defining an underlying continuous variable and a threshold link.

(b) Using an underlying normal specification of probit models, and implementing Albert and Chib (1993) data augmentation algorithm.

(c) Alternative latent variable models - underlying Poisson, etc.
5. Using Hierarchical GLMs for Correlated Data

(a) Definition of a typical generalized linear mixed model (GLMM) that incorporates “random-effects” for a subvector of the regression parameters.

(b) Conjugate prior distributions and Gibbs sampling in normal case.

(c) Problems with improper priors and mixing.

(d) How to implement Gibbs sampling in non-normal cases, probit and otherwise.

(e) Deriving correlations between normal and underlying normal variables as a function of variance components.
6. Survival Analysis

(a) Definitions of different types of censoring, hazard function (discrete and continuous time), survival function, cumulative hazard, and standard relationships.

(b) Proportional hazards model and accelerated failure time model - definitions, and parameter interpretation.

(c) Standard parametric models - constant hazard (exponential), piece-wise constant (piecewise exponential), and Weibull.

(d) Cure rate models - definition and data augmentation trick for posterior computation
7. Missing Data

(a) Standard strategies and associated assumptions.

(b) Pattern mixture and selection model definitions.

(c) MCAR, MAR, informative missingness, non-ignorability, and identifiability issues.

(d) Using data augmentation MCMC to account for missing response data under MAR.

(e) Accounting for missing covariates
8. **Discrete Time Survival Analysis**

(a) Definition and converting a given continuous-time model (e.g., proportional hazards) to discrete time.

(b) Defining a discrete time survival model in terms of a binary response likelihood and GLM.

(c) Incorporating time-varying covariates and time-varying coefficients.

(d) Smoothing the baseline hazard function.

(e) Continuation-ratio probit models - issues in computation and inference & practical justification.

(f) Incorporating parameter restrictions - details in prior specification, posterior implementation, and motivation.
9. Alternatives to Hierarchical Models for Multivariate Binary Data: Bayesian Logistic Regression

(a) Incorporation of random effects induces correlation structure on multiple binary response data.

(b) Alternatively, one can define a multivariate distribution (e.g., multivariate normal, multivariate logistic) for a vector of underlying outcomes.

(c) Issues in parameter interpretation and computation.

(d) Differences between subject-specific and marginal parameter interpretations.
10. **Interval Censored Data**

(a) Timing of examinations differs between subjects and we only know if event occurs between examinations.

(b) How to analyze data of this type using a discrete time Bayesian survival analysis.

(c) In particular, provide details on using data augmentation to simplify the analysis.

(d) How to extend this approach to incorporate surrogate data on the latency time between event occurrence and detection at an examination (e.g., uterus and tumor number/size in uterine fibroid example)?

(e) How to use a 3 state illness-death model to account for informative censoring (e.g., of tumor onset by death) - assumptions?
11. Bayesian Variable Selection and Order Restricted Inference

(a) First suppose we have normal or underlying normal (probit) data and are interested in inferences on effects of ordered categorical predictors.

(b) What are the possibilities for addressing this approach?

(c) Define a mixture prior for addressing this problem and describe posterior computation and inferences.

(d) How does this approach relate to variable selection via SSVS algorithms?
12. **Poisson Log-Linear and Logistic Regression Cases**

(a) Now suppose data are Poisson distributed counts following a log-linear model.

(b) Describe a Poisson-gamma hierarchical model which accommodates over-dispersion relative to the Poisson distribution.

(c) Generalize this model to account for dependency in multiple observations from an individual.

(d) For categorical covariates, define a conditionally-conjugate prior distribution for the regression coefficients.

(e) Modify this prior for variable selection and inferences on ordered categorical covariates.

(f) Use an underlying Poisson modeling strategy to apply this same approach to logistic regression and complementary log-log models for categorical outcomes.
13. **Bayesian Generalized Additive Models with Constraints**

(a) Focusing again on the normal or underlying normal data cases, consider models that allow the regression function to be unknown.

(b) Define a generalized additive model, and place a prior on the unknown regression function(s) with or without monotonicity constraints.

(c) Place a hyperprior on unknown smoothing parameters, show that the prior is conditionally-conjugate and discuss properties including details in posterior computation.

(d) How to apply to real problems - conceptually?
Example Exam Problem Set

Question 1:
Suppose that 2500 pregnant women are enrolled in a study and the outcome is the occurrence of preterm birth. Possible predictors of preterm birth include age of the woman, smoking, socioeconomic status, body mass index, bleeding during pregnancy, serum level of dde, and several dietary factors. Formulate the problem of selecting the important predictors of preterm birth in a generalized linear model (GLM) framework. Show the components of the GLM, including the link function and distribution (in exponential family form). Describe (briefly) how estimation and inference could proceed via a frequentist approach.
Possible Solution:

\( y_i = 1 \) if woman \( i \) has preterm birth and \( y_i = 0 \) otherwise (\( i = 1, \ldots, n \))

\( y_i \sim \text{Bernoulli}(\pi_i) \)

Probability density function:

\[
f(y_i; \pi_i) = \pi^y_i (1 - \pi_i)^{1-y_i}
= \exp \left \{ y_i \log \pi_i + (1 - y_i) \log (1 - \pi_i) \right \}
= \exp \left \{ y_i \log \left( \frac{\pi_i}{1 - \pi_i} \right) + \log (1 - \pi_i) \right \}
= \exp \left \{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right \},
\]

where

\[
\theta_i = \log \left( \frac{\pi_i}{1 - \pi_i} \right), \quad b(\theta_i) = \log (1 + e^{\theta_i}), \quad a(\phi) = \phi = 1,
\]

and \( c(y_i, \phi) = 0 \).

Link function:

Any mapping from \( \mathbb{R} \rightarrow [0, 1] \). A convenient choice is the canonical link,

\[
\eta_i = \theta_i = \log \left( \frac{\pi_i}{1 - \pi_i} \right),
\]

which is the logit. The probit and complementary log-log are alternatives.

Frequentist Estimation:

Maximum likelihood estimates can be obtained for a given model, say

\[
\log \left( \frac{\pi_i}{1 - \pi_i} \right) = x'_i \beta,
\]
(where $\mathbf{x}_i$ is a $p \times 1$ vector of predictors) by iterative weighted least squares

**Frequentist Inference:**

One can select the important predictors to be included in the model by step-wise selection, using the AIC or BIC criterion.

Alternatively, one can just fit the model with all the predictors and then do inferences based on the MLEs and asymptotic standard errors. For example, for continuous predictors included as linear terms in the model, we can do a Wald test. Alternatively, we could do analysis of deviance (see notes for details) to test for significant differences in fit between the nested models with and without a particular predictor.
Question 2:

Women are enrolled in a study when they go off of contraception with the intention of achieving a pregnancy. Suppose there are 350 women in the study who provide information on the number of menstrual cycles required to achieve a pregnancy, whether or not they smoke cigarettes, and their age at beginning the attempt. Describe a statistical model for addressing the question: Is cigarette smoking related to time to pregnancy? Formulate the statistical model within a Bayesian framework and outline the details of model fitting and inference (including the form of the posterior density, an outline of the algorithm for posterior computation, and the approach for addressing the scientific question based on the posterior).
Discrete time survival analysis:

Let $y_{ij} = 1$ if woman $i$ conceives in cycle $j$ and $y_{ij} = 0$ otherwise

Let $r_{ij} = 1$ if woman $i$ is attempting in cycle $j$ and $r_{ij} = 0$ otherwise

Discrete hazard of conception in cycle $j$:

$$
\lambda_{ij} = h(\alpha_j + \mathbf{x}_i'\beta),
$$

where $\alpha_j$ is an intercept parameter and $\mathbf{x}_i = (smk_i, age_i)'$.

Assuming a continuation-ratio probit model, the likelihood is:

$$
\prod_{i=1}^{350} \prod_{j=1}^{T} \left[ \Phi(\alpha_j + \mathbf{x}_i'\beta)^{y_{ij}} \{1 - \Phi(\alpha_j + \mathbf{x}_i'\beta)\}^{1-y_{ij}} \right]^{r_{ij}}
$$
We complete a Bayesian specification of the model with prior densities for
\[ \alpha = (\alpha_1, \ldots, \alpha_T)' \] and \[ \beta, \]
\[ 1(\alpha_1 > \alpha_2 > \ldots > \alpha_T) \prod_{j=1}^{T} N(\alpha_j; \alpha_{0j}, \sigma_{\alpha_j}^2) N(\beta; \beta_0, \Sigma_\beta), \]
where the order constraint models the selection process where more fertile
couples conceive rapidly.

To simplify posterior computation, we augment the observed data with un-
derlying normal variables. In particular, let \[ y_{ij} = 1(z_{ij} > 0), \] where \[ z_{ij} \sim N(\alpha_j + \mathbf{x}_{ij}' \beta, 1) \] are independent. Posterior computation can then proceed via
the following Gibbs sampling algorithm:

1. Choose initial values for \( \alpha \) and \( \beta \).
2. Sample \( z_{ij} \), for all \( i, j : r_{ij} = 1 \), from its full conditional density, which is
\[ N(\alpha_j + \mathbf{x}_{ij}' \beta, 1) \text{ truncated below (above) by 0 for } y_{ij} = 1 (y_{ij} = 0). \]
3. Sample \( \alpha_j \), for \( j = 1, \ldots, T \), from its full conditional density, which is
\[ N(\hat{\alpha}_j, \hat{\sigma}_{\alpha_j}^2) \text{ truncated so that } \alpha_j \in (\alpha_{j+1}, \alpha_{j-1}). \]
4. Sample \( \beta \) from its multivariate normal full conditional density.
5. Repeat steps 2-4 until apparent convergence and calculate posterior sum-
maries based on a large number of additional draws.
Our primary goal is to address the question: Is cigarette smoking related to time to pregnancy.

Based on the Gibbs iterates, we can estimate $\Pr(\beta_1 < 0 \mid data)$. If this posterior probability is high (say, greater than 95%) than we have strong evidence that the hazard of conception is lower for smokers than non-smokers (at least in the population).

This implies that smoking may be associated with an increased time to pregnancy.

To characterize the magnitude of this effect, we could estimate the time to pregnancy distribution for smokers and non-smokers, and obtain credible intervals for these estimates.
Question 3:

A study is conducted examining the impact of alcohol intake during pregnancy on the occurrence of birth defects of 5 different types. Outcome data for a child consist of 5 binary indicators of the presence or absence of the different birth defects. A physician working with you on the study notes that certain children have several birth defects, possibly due to defects in important unmeasured genes, while most children have no defects. Describe a latent variable model for analyzing these data and outline (briefly) the details of a Bayesian analysis (including the form of the posterior density, an outline of the algorithm for posterior computation, and the approach for addressing the scientific question based on the posterior).
Multiple binary outcomes:

\[ y_i = (y_{i1}, \ldots, y_{i5})', \text{ where } y_{ij} = 1 \text{ if child } i \text{ has } j \text{th defect and 0 otherwise} \]

Probit regression model:

\[
\Pr(y_{ij} = 1 \mid \xi_i, x_i) = \Phi(\alpha_j + x_i' \beta_j + \lambda_j \xi_i),
\]

where \( \alpha_j \) is an intercept parameter for defect \( j \),

\( x_i \) is a vector of predictors (level of alcohol, age of mother, etc),

\( \beta_j \) are coefficients specific to the \( j \)th defect,

\( \xi_i \sim N(0, 1) \) is a latent variable measuring genetic susceptibility of child,

\( \lambda_j \) is a factor loading relating overall genetic susceptibility to \( j \)th defect.
Note that we don’t need to assume probit, we could use other link functions.

We complete a Bayesian specification of the model with prior distributions for \( \mathbf{\alpha} = (\alpha_1, \ldots, \alpha_5)' \), \( \mathbf{\beta} = (\beta_1', \ldots, \beta_5')' \) and \( \mathbf{\lambda} = (\lambda_1, \ldots, \lambda_5)' \),

\[
(\mathbf{\alpha}', \mathbf{\beta}')' \sim N(\mathbf{\mu}, \mathbf{\Sigma}) \quad \text{and} \quad \lambda_j \sim N(\lambda_{0j}, \sigma_{\lambda_j}^2) \text{ truncated below by } 0.
\]

Posterior computation can proceed via a Gibbs sampler,

1. Sample underlying variables from their full conditional

\[
\begin{align*}
  z_{ij} & \sim N(\alpha_j + x_i' \beta_j + \lambda_j \xi_i, 1), \\
\end{align*}
\]

truncated below (above) by 0 for \( y_{ij} = 1 \) (\( y_{ij} = 0 \))

2. Sample \( (\mathbf{\alpha}', \mathbf{\beta}')' \) from the multivariate normal full conditional

3. Sample \( \xi_i \) from its normal full conditional

4. Sample each \( \lambda_j \) from their truncated normal full conditionals

To assess the effect of alcohol intake on the different defects, estimate the marginal posterior densities of \( \beta_{j1} \), for \( j = 1, \ldots, 5 \).

For defects having high values of \( \Pr(\beta_{j1} > 0 \mid \text{data}) \), we have evidence of a positive association with alcohol intake.
Question 4:

A toxicology study is conducted in which pregnant mice are exposed to different doses of a chemical. The outcome data consist of an ordinal ranking of the sickness of each pup in each litter, with 1 = healthy, 2 = low birth weight but otherwise healthy, 3 = malformed, and 4 = dead. The goal of the study is to see if dose is associated with health of the pup. Describe a model and analytic strategy. What is the interpretation of the model parameters? What assumptions are being made and can they be relaxed?
Let $y_{ij} \in \{1, 2, 3, 4\}$ be the outcome for the $j$th pup in the $i$th litter, and let $x_i$ be the dose of the test chemical.

A possible model for relating the ordinal ranking of pup health to dose while allowing for within-litter dependency would be

$$
Pr(y_{ij} \leq k \mid x_i, b_i) = \Phi(\alpha_k - \beta x_i - b_i),
$$

where $\alpha_k$ is an intercept parameter or cutpoint on a latent normal density, $\beta$ is a slope parameter characterizing the effect of dose, and $b_i \sim N(0, \psi^{-1})$ is a litter-specific latent variable (random effect).

**Strategy:**

1. Formulate this generalized probit model as an underlying normal regression model.

2. Define conditionally-conjugate priors, restricting $\alpha_1 = 0$ and the $\alpha$’s to be increasing for identifiability and so that the distribution function of $Y$ is a proper distribution function.

The parameter of primary interest is the slope parameter, $\beta$, which is interpretable as the increase in the underlying normal mean attributable to a unit increase in dose.

One assumption being made is that one parameter can be used to characterize the shift in the distribution of $Y$ as dose changes.

Potentially the shape of the distribution may be completely different and one may need to have $\beta$’s specific to $k$.

Another assumption is that we have non-informative cluster size - that is, the number of pups within a litter is not informative. To address this, we could include cluster size as a covariate or even a separate outcome.