Bayesian Hypothesis Testing in GLMs: One-Sided and Ordered Alternatives

- Often interest may focus on comparing a null hypothesis of no difference between groups to an ordered restricted alternative.

- For example, we may have a $k$ level ordered categorical predictor, $w_i$,

$$
y_i = \beta_0 + \sum_{h=1}^{k-1} 1(w_i = h + 1) \beta_h + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$.

- $H_0: \beta_1 = \ldots = \beta_{k-1} = 0$ homogeneity (no association)

- $H_1: \beta_1 \leq \ldots \leq \beta_{k-1}$ simple increasing order

How to assess from Bayesian perspective?
Suppose we assumed the conjugate prior density,
\[ \beta = (\beta_0, \ldots, \beta_{k-1})' \sim N(\mu_0, \Sigma_0) \quad \text{and} \quad \sigma^2 \sim IG(a_0, b_0). \]

Under this prior density, we could easily calculate the posterior density.

Posterior probabilities of \( \beta_h < 0 \) can be calculated.

We can also calculate \( \Pr(\beta_1 \leq \ldots \leq \beta_{k-1} | \text{data}) \)

How can we address \( H_0 \) vs \( H_1 \) using this posterior?

Is there a better way?
The Bayes factor is a standard way of comparing two hypotheses, $H_0$ and $H_1$.

To calculate the Bayes factor, we need to calculate the prior and posterior probabilities of each of the two hypotheses.

What are these probabilities under the conjugate normal prior?

Can we use $\Pr(H_0) = 1 - \Pr(H_1) = 1 - \Pr(\beta_1 \leq \ldots \leq \beta_{k-1})$ as the prior?

Why or why not?
The problem with this approach is that the typical normal conjugate prior assigns zero probability to the null hypothesis.

Thus, the above strategy doesn’t make sense.

Instead, we want to choose a prior density for $\beta$ that allocates probability to $H_0$ and $H_1$, with these probabilities adding to one.

Essentially, we need a prior that has support on the restricted space

$$\Omega = \{\beta : \beta_1 \leq \ldots \leq \beta_{k-1}\},$$

with positive probability assigned to equalities.
We would also like to have a prior is easy to elicit and results in easy computation.

To place order restrictions on parameters in Bayesian models, Gelfand, Smith and Lee (1992) proposed priors of the form

\[ \pi(\beta) \propto 1(\beta \in \Omega) N(\mu_0, \Sigma_0), \]

which is a truncated Gaussian density.

This prior allocates probability one to the restricted space \( \Omega \).

In addition, the full conditional densities of the \( \beta \)'s follow a conditionally conjugate normal form.

Is this approach good for comparing \( H_0 \) and \( H_1 \)?
Actually, we are still assigning zero prior probability to the null hypothesis $H_0$.

By discard draws from the multivariate normal density that are inconsistent with $\beta_1 \leq \ldots \leq \beta_{k-1}$, we ensure that strictly increasing order is satisfied.

However, we never draw a value of $\beta$ such that $\beta_j = \beta_h$.

A generalization is to include point masses to accommodate equalities.
In particular, first reparameterize so that $\gamma_1 = \beta_1$ and $\gamma_j = \beta_j - \beta_{j-1}$ for $j = 2, \ldots, k-1$.

Then choose the following prior density: $\pi(\beta, \gamma) = N(\beta_0; \mu_0, \sigma^2_0) \times$

$$\left\{ \prod_{h=1}^{k-1} \pi_{0h} 1(\gamma_h = 0) + (1 - \pi_{0h}) 1(\gamma_h > 0) \int_0^\infty \frac{N(\gamma_h; \mu_h, \sigma^2_h)}{\int_0^\infty N(z; \mu_h, \sigma^2_h) dz} \right\}$$

The $\gamma_h$ parameters are assigned prior densities consisting of mixtures of point masses at zero (with probability $\pi_{0h}$) and normal densities truncated below by zero.

The prior probability of equivalent means for individuals with $w_i = j$ and $w_i = j+1$ is $\pi_{0j}$, for $j = 1, \ldots, k-1$.

The prior probability of the overall null hypothesis $H_0$ is $\pi_0 = \prod_{j=1}^{k-1} \pi_{0j}$. 
Under this prior, \( \Pr(H_0) = \pi_0 \) and \( \Pr(H_1) = 1 - \pi_0 \).

The prior has support on the restricted space \( \Omega \).

In addition, the prior density is conditionally conjugate with the posterior of \( \gamma_h \) of the form

\[
\tilde{\pi}_h 1(\gamma_h = 0) + (1 - \tilde{\pi}_h) 1(\gamma_h > 0) \frac{N(\gamma_h; \tilde{\mu}_h, \tilde{\sigma}^2_h)}{\int_0^{\infty} N(z; \tilde{\mu}_h, \tilde{\sigma}^2_h) dz},
\]

where \( \tilde{\mu}_h \) and \( \tilde{\sigma}^2_h \) are the posterior mean and variance derived under an unrestricted \( N(\mu_0, \sigma^2_0) \) prior density for \( \gamma_h \).

\( \tilde{\pi}_h \) is the posterior probability of \( \gamma_h = 0 \) given the data and other parameters.
Due to the simplicity of this form, we can simply proceed by a Gibbs sampling algorithm:

1. Specify initial values for $\beta_0$, $\gamma$ and $\sigma^2$.
2. Update $\sigma^2$ by sampling from $\mathcal{IG}$ full conditional
3. Update $\beta_0$ by sampling from normal full conditional
4. Update $\gamma_h$, for $h = 1, \ldots, k - 1$, by sampling from the zero-inflated truncated normal full conditional:
   
   (a) Sample from point mass by using Bernoulli($\tilde{\pi}_h$).
   (b) If not in point mass sample from $N(\tilde{\mu}_h, \tilde{\sigma}_h^2)$ truncated below by 0.
5. Repeat 2-4.
Calculation of Bayes factors for hypothesis testing