A specific example of Gibbs sampling - canonical Markov chain Monte Carlo method for calculations in Bayesian statistical models, as well as a key applied model class.

♠ Simulation of Posteriors

- For any statistical model, simulated values of posterior distributions are nowadays the standard in statistical computation: summarise large samples of parameter values from a posterior distribution to easily understand the information contained in that posterior about the parameters
- MCMC methods, such as Gibbs sampling, generate simulations sequentially using Markov chains
- Gibbs sampling represents an approach in which sets of coupled conditional posteriors derived from the model are used for these simulations
- Here's a key example

♠ Prior and Posterior in the Linear Regression Model: Multiple Shrinkage Example

- Example: Now take $\boldsymbol{\beta} \sim N(0, \mathbf{C}^{-1})$ and we'll take the precision as diagonal, with diagonal elements (multiple shrinkage factors) now to be estimated along with the regression parameter and precision ϕ. Note that this is a bit different to the conditionally conjugate prior of the standard regression earlier – the error precision is no longer included in the prior for $\boldsymbol{\beta}$. This is a modification of the analysis that relaxes that strict adoption of a traditional conjugate prior – we are freed from the need to impose strict conjugacy now that we're using MCMC methods.
- Write $\mathbf{C}^{-1} = \text{diag}(\tau_1, \tau_1, \ldots, \tau_p)$ with prior variances τ_i defining element-wise shrinkage parameters for the individual predictor variables.
- Now allow for learning on the shrinkage factors using priors that are independent gamma priors, $\tau_i^{-1} \sim Ga(k/2, h/2)$ for each j, say, then the relevant conditional posteriors are all available in conjugate forms.

♠ Conditional Posteriors

- $p(\boldsymbol{\beta}|\mathbf{y}, \phi, \mathbf{C})$
  We have a linear regression so that the standard results follow: The posterior (see earlier notes) is multivariate normal
  $$\boldsymbol{\beta}|\mathbf{y}, \phi, \mathbf{C} \sim N(\mathbf{b}, \mathbf{B}^{-1})$$
  with $\mathbf{B} = \mathbf{C} + \mathbf{H}'\mathbf{H}\phi$ and $\mathbf{b} = \mathbf{B}^{-1}\phi\mathbf{H}'\mathbf{y}$ where $\mathbf{H} = \mathbf{X}'$ is the $n \times p$ design matrix.
  For multivariate normals, use programmed functions (multivariate normal simulation is very standard - e.g.. rMNorm.m or similar) or use direct Cholesky decomposition: e.g., by hand in matlab
  $$\mathbf{b} + chol(inv(\mathbf{B})) * randn(p, 1)$$
  for a single draw
- If we know $\mathbf{y}, \boldsymbol{\beta}$ and $\mathbf{C}$, then the conditional posterior for the error precision ϕ depends only on the data and $\boldsymbol{\beta}$ through the residual $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{H}\boldsymbol{\beta}$, and the relevant conditional posterior for sampling values from is just the updated gamma distribution $\phi|\mathbf{y}, \boldsymbol{\beta}, \mathbf{C} \sim Ga((a+n)/2, (b + \boldsymbol{\epsilon}'\boldsymbol{\epsilon})/2)$.
- The relevant conditional posteriors are also independent gammas, namely
  $$\tau_j^{-1} \sim Ga((k+1)/2, (h + \beta_j^2)/2).$$
  This allows for learning on differential shrinkage parameters across variables.
- *Exercise:* Develop the MCMC with these priors.
- Think about choices of prior parameters: One general way to think about ranges of relevant parameter values in binary regression models is to consider how variation in μ_j translates through to the probability scale π_j. Absolute values of μ_j bigger than 2 or so lead to probabilities that are already very extreme.