Other Noninformative Priors

- Other methods for noninformative priors include
  - Bernardo's reference prior, which seeks a prior that will maximize the discrepancy between the prior and the posterior and minimize the discrepancy between the likelihood and the posterior (a "dominant likelihood prior").
  - An improper prior, in which $\int_{\Theta} p(\theta) = \infty$.
  - A highly **diffuse** proper prior, e.g., for normal data with μ unknown, a $N(0, 1000000)$ prior for μ. (This is very close to the improper prior $p(\mu) \propto 1$.)

Informative Prior Forms

- Informative prior information is usually based on expert opinion or previous research about the parameter(s) of interest.

Power Priors

- Suppose we have access to **previous data $\mathbf{x}_0$** that is analogous to the data we will gather.
- Then our "power prior" could be

$$p(\theta|\mathbf{x}_0, a_0) \propto p(\theta)[L(\theta|\mathbf{x}_0)]^{a_0}$$

where $p(\theta)$ is an ordinary prior and $a_0 \in [0, 1]$ is an exponent measuring the influence of the previous data.

Power Priors

- As $a_0 \to 0$, the influence of the previous data is lessened.
- As $a_0 \to 1$, the influence of the previous data is strengthened.
- The posterior, given **our actual** data $\mathbf{x}$, is then

$$\pi(\theta|\mathbf{x}, \mathbf{x}_0, a_0) \propto p(\theta|\mathbf{x}_0, a_0)L(\theta|\mathbf{x})$$

- To avoid specifying a single a_0 value: We could put a, say, beta distribution $p(a_0)$ on a_0 and average over values of a_0 in $[0, 1]$:

$$p(\theta|\mathbf{x}_0) = \int_0^1 p(\theta)[L(\theta|\mathbf{x}_0)]^{a_0} p(a_0)\, \mathrm{d}a_0$$

Prior Elicitation

- A challenge is putting "expert opinion" into a form where it can be used as a prior distribution.
  Strategies:
- Requesting guesses for several quantiles (maybe $\{0.1, 0.25, 0.5, 0.75, 0.9\}$?) from a few experts.
- For a normal prior, note that a quantile $q(\alpha)$ is related to the z-value $\Phi^{-1}(\alpha)$ by:

$$q(\alpha) = \text{mean} + \Phi^{-1}(\alpha) \times (\text{std. dev.})$$

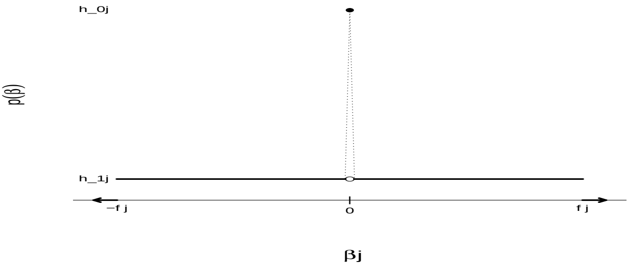
- Via regression on the provided $[q(\alpha), \Phi^{-1}(\alpha)]$ values, we can get estimates for the mean and standard deviation of the normal prior.

Prior Elicitation

- Another strategy asks the expert to provide a "predictive modal value" (most "likely" value) for the parameter.
- Then a rough 67% interval is requested from the expert.
- With a normal prior, the length of this interval is twice the prior standard deviation.
- For a prior on a Bernoulli probability, the "most likely" probability of success is often "clear".

Spike-and-Slab Priors for Linear Models

- In regression, the priors on the regression coefficients are crucial.
- Whether or not $\beta_j = 0$ defines whether X_j is "important" in the regression.
- For any j, a useful prior for β_j is:



Spike-and-Slab Priors for Linear Models

- Here: $P(\beta_j = 0) = h_{0j}$ (= prior probability that X_j is **not** needed in the model)
- $P(\beta_j \neq 0) = 1 - h_{0j} = h_{1j}\big(f_j - (-f_j)\big) = 2f_j h_{1j}$ (where $[-f_j, f_j]$ contains all "reasonable" values for β_j)
- To include X_j in the model with certainty, set $h_{0j} = 0$.
- To reflect **more doubt** that X_j should be in the model, increase the ratio

$$\frac{h_{0j}}{h_{1j}} = \frac{h_{0j}}{(1 - h_{0j})/2f_j} = 2f_j \frac{h_{0j}}{1 - h_{0j}}$$

- Recently, "nonparametric priors" have become popular, typically involving a mixture of Dirichet processes.

CHAPTER 6 SLIDES START HERE

The Monte Carlo Method

- The **Monte Carlo method** involves studying a distribution (e.g., a posterior) and its characteristics by generating many random observations having that distribution.
- If $\theta^{(1)}, \ldots, \theta^{(S)} \overset{\text{iid}}{\sim} \pi(\theta|\mathbf{x})$, then the empirical distribution of $\{\theta^{(1)}, \ldots, \theta^{(S)}\}$ approximates the posterior, when S is large.
- By the **law of large numbers**,

$$\frac{1}{S}\sum_{s=1}^{S} g(\theta^{(s)}) \rightarrow E[g(\theta)|\mathbf{x}]$$

as $S \rightarrow \infty$.

The Monte Carlo Method

So as $S \to \infty$:

$$\bar{\theta} = \frac{1}{S}\sum_{s=1}^{S}\theta^{(s)} \to \text{ posterior mean}$$

$$\frac{1}{S-1}\sum_{s=1}^{S}(\theta^{(s)} - \bar{\theta})^2 \to \text{ posterior variance}$$

$$\frac{\#\{\theta^{(s)} \leq c\}}{S} \to P[\theta \leq c|\mathbf{x}]$$

$$\text{median}\{\theta^{(1)}, \ldots, \theta^{(S)}\} \to \text{ posterior median}$$

(and similarly for **any** posterior quantile).

The Monte Carlo Method

- If the posterior is a "common" distribution, as in many conjugate analyses, we could draw samples from the posterior using R functions.

Example 1: (General Social Survey)

- **Sample 1**: # of children for women age 40+, no bachelor's degree.
- **Sample 2**: # of children for women age 40+, bachelor's degree or higher.
- Assume Poisson(θ_1) and Poisson(θ_2) models for the data.
- We use gamma(2,1) priors for θ_1 and for θ_2.

The Monte Carlo Method

- **Data**: $n_1 = 111$, $\sum_i x_{i1} = 217$
- **Data**: $n_2 = 44$, $\sum_i x_{i2} = 66$
- $\Rightarrow$ Posterior for θ_1 is gamma(219,112).
- $\Rightarrow$ Posterior for θ_2 is gamma(68, 45).
- Find $P[\theta_1 > \theta_2 | \mathbf{x}_1, \mathbf{x}_2]$.
- Find posterior distribution of the ratio $\frac{\theta_1}{\theta_2}$.
- See `R` example using Monte Carlo method on course web page.