# Notes for *Bayesian Data Analysis* by Gelman et al.

Kevin O'Connor

## Ch 2: *Single-Parameter Models*

### 2.4: *Informative Prior Distributions*

Let $\mathcal{F}$ be a class of sampling distributions, $\mathcal{F} = \{p(y|\theta)\}$, and $\mathcal{P}$ be a class of prior distributions. Then $\mathcal{P}$ is **conjugate** for $\mathcal{F}$ if $p(\theta|y) \in \mathcal{P}$, $\forall p(y|\theta) \in \mathcal{F}$, $p(\theta) \in \mathcal{P}$.

To avoid choosing $\mathcal{P}$ as class of all distributions, we refer to **natural conjugate families** taking $\mathcal{P}$ to be a class where all members have same functional form. We like conjugate families because they are easier to work with and tend to be more interpretable.

Exponential families are conjugate to one another,

$$p(y|\theta) \propto g(\theta)^n \exp\left\{\phi(\theta)^T t(y)\right\}$$

$$p(\theta) \propto g(\theta)^\eta \exp\left\{\phi(\theta)^T \nu\right\}$$

$$p(\theta|y) \propto g(\theta)^{n+\eta} \exp\left\{\phi(\theta)^T (\nu + t(y))\right\}$$

### 2.5: *Normal Distribution with Known Variance*

The **posterior predictive distribution** gives distribution of future observation given observed data.

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

For normal sampling distribution with known variance, conjugate prior given by $\theta \sim \mathcal{N}(\mu_0, \tau_0^2)$.

### 2.6: *Other Standard Single-parameter Models*

For normal sampling distribution with unknown variance and known mean, conjugate prior given by $\sigma^2 \sim \text{IG}(\alpha, \beta)$.

$$p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2}$$

For Poisson, conjugate prior given by $\lambda \sim \text{Gamma}(\alpha, \beta)$.

For exponential (special case of Gamma), $\theta \sim \text{Gamma}(\alpha, \beta)$ is conjugate.

### 2.8: *Noninformative Prior Distributions*

A prior is **prior** if it doesn't depend on data and integrates to 1. If the prior integrates to any positive finite value it is called an **unnormalized density**. For some improper priors, one may get a valid posterior distribution.

**Jeffreys' Invariance Principle**: any rule for determining the prior density should yield an equivalent result if applied to the one-to-one transformed parameter. This leads to the definition of the noninformative prior density as $p(\theta) \propto \sqrt{J(\theta)}$ where $J(\theta)$ is Fisher information,

$$J(\theta) = \mathbb{E}\left[\left(\frac{d}{d\theta}\log p(y|\theta)\right)^2 \middle| \theta\right] = -\mathbb{E}\left[\frac{d^2}{d\theta^2}\log p(y|\theta) \middle| \theta\right]$$

Jeffreys' principle doesn't extend that well to multiparameter models. When there are many parameters, tend to abandon noninformative priors in favor of hierarchical models.

A **pivotal quantity** is a function of the data and parameter whose distribution does not depend on the data or the parameter.

Difficulty with noninformative prior distributions is that one must still choose what kind of noninformative prior to use. But also seems inappropriate to establish a single universal method of finding noninformative priors as different situations warrant different approaches.

### 2.9: *Weakly Informative Prior Distributions*

Prior is **weakly informative** if it is proper but set up to provide information intentionally weaker than whatever prior knowledge is available.

We like weakly informative priors because they provide just enough information to ensure that the posterior makes sense.

Two potential methods for constructing weakly informative priors:

1. Start with a noninformative prior then add enough information to make posterior reasonable.
2. Start with strong, highly informative prior and broaden it to account for uncertainty.

## Ch 3: *Introduction to Multiparameter Models*

**Nuisance parameters** are those which are not of immediate interest.

### 3.1: *Averaging over Nuisance Parameters*

Suppose $\theta = (\theta_1, \theta_2)$ and $\theta_2$ is a nuisance parameter. Then we can write

$$p(\theta_1|y) = \int p(\theta_1, \theta_2|y)d\theta_2 = \int p(\theta_1|\theta_2, y)p(\theta_2|y)d\theta_2$$

Noninformative prior for Normal likelihood can be given by

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$$

### 3.3: *Normal Data with a Conjugate Prior*

For $y|\mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$ with both $\mu$ and $\sigma^2$ unknown, need a prior for $(\mu, \sigma^2)$. Conjugate prior given by

$$(\mu, \sigma^2) \sim \text{N-Inv}\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$$

where

$$\sigma^2 \sim \text{Inv}\chi^2(\nu_0, \sigma_0^2)$$

$$\mu|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma_0^2/\kappa_0)$$

Then marginal posteriors for $\mu$ and $\sigma^2$ are $t$ and Inv$\chi^2$ respectively.

### 3.4: *Multinomial Model for Categorical Data*

For multinomial sampling distribution, conjugate prior is Dirichlet,

$$p(\theta|\alpha) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}, \qquad \sum_{j=1}^{k} \theta_j = 1$$

### 3.5: *Multivariate Normal Model*

With known $\Sigma$ and unknown $\mu$, conjugate prior is $\mu \sim \mathcal{N}(\mu_0, \Lambda_0)$.

With both $\mu$ and $\Sigma$ unknown, conjugate prior given by

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1})$$

$$\mu|\Sigma \sim \mathcal{N}(\mu_0, \Sigma/\kappa_0)$$

Noninformative priors for MVN:

1. Setting $\Sigma \sim \text{Inv-Wishart}_{d+1}(I)$ corresponds to each correlation in $\Sigma$ having a marginally uniform distribution.
2. $\Sigma \sim \text{Inv-Wishart}_{d-1}(I)$ gives multivariate Jeffreys prior,

   $$p(\mu, \Sigma) \propto |\Sigma|^{-(d+1)/2}$$

   which corresponds to letting $\kappa_0 \to 0$, $\nu_0 \to -1$, and $|\Lambda_0| \to 0$.