STATS 225: Bayesian Analysis
Supplementary Materials: Convenient Priors

Babak Shahbaba

Department of Statistics, UCI

Winter, 2015

Vague, diffuse, flat, “noninformative”, and reference priors

- The priors we used in the poisson and normal models are called *informative prior*.
- Such priors are intended to provide non-data (at least not the data at hand) information about model parameters and they influence the results to some extent.
- The influence of these priors diminishes very quickly as data increase.
- In some cases, we might not be comfortable with expressing our prior opinion and prefer another class of priors, which are intended to provide the least amount of information.

Vague, diffuse, flat, “noninformative”, and reference priors

- For a normal $N(\theta, \sigma^2)$ model with unknown mean and known variance, we used $P(\theta) = N(\mu_0, \tau_0^2)$ conjugate prior.

- Imagine now that we want to be less specific about our prior opinion. We can do so by increasing τ_0^2.

- In limit, we take τ_0^2 to ∞, which results in the following posterior:

$$\theta | y \sim N(\bar{y}, \sigma^2/n)$$

  which is obtained by $\tau_0^2 \to \infty$ in the $p(\theta|y)$ based on the conjugate informative prior.

Vague, diffuse, flat, “noninformative”, and reference priors

- This is equivalent to using a completely flat (locally uniform) prior, $P(\theta) \propto \text{constant}$, for θ on real line similar to the uniform distribution on $[0, 1]$ that we used for the binomial model.

- Unlike the flat prior on $[0, 1]$, this prior is not a *proper* probability distribution since it does not integrate to 1. We refer to such priors as *improper* priors.

- Although the prior is improper, as we can see, the posterior distribution is still a proper distribution

Vague, diffuse, flat, "noninformative", and reference priors

- We can also use such "noninformative" priors (or as Christensen et al. call it, standard improper reference priors) for σ^2 in normal models with known mean and unknown variance.
- Such priors can be obtained by decreasing the degrees of freedom in the scaled inverse-χ^2 prior.
- In limit, we can set $\nu_0 = 0$ which results in the following posterior distribution:

$$\sigma^2 | y \sim \text{Inv-}\chi^2(n, v)$$

- Setting $\nu_0 = 0$ is equivalent to using $P(\sigma^2) \propto 1/\sigma^2$, which is also an improper prior resulting in a proper posterior distribution.

Vague, diffuse, flat, “noninformative”, and reference priors

- If both θ and σ^2 are unknown, assuming that they are independent *in prior*, we can use a joint impropoer prior as follows:

$$P(\theta, \sigma^2) \propto 1/\sigma^2$$

- The joint posterior distribution is therefore

$$P(\theta, \sigma^2 | y) \propto \sigma^{-n-2} \exp\Big(- \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta)^2 \Big)$$

Vague, diffuse, flat, “noninformative”, and reference priors

- By integrating over θ, the posterior distribution of σ^2 has the following simple form:

$$\sigma^2|y \sim \text{Inv-}\chi^2(n-1, s^2)$$
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- Given σ^2 on the other hand, the conditional posterior distribution of θ becomes

$$\theta|\sigma^2, y \sim N(\bar{y}, \sigma^2/n)$$

- If we integrate the joint posterior distribution over σ^2, the marginal distribution of θ will be a t distribution:

$$\theta|y \sim t_{n-1}(\bar{y}, s^2/n)$$

Jeffreys' prior

- Another systematic approach for setting priors was introduced by Jeffrey.
- The idea is to use a prior $p(\theta)$ that is invariant to transformation such that all parameterizations result in the same prior.
- Recall that for any one-to-one transformation $\phi = h(\theta)$, where h is an invertible function, we have

$$p_\phi(\phi) = p_\theta(\theta)\left|\frac{d\theta}{d\phi}\right|$$

- For example, if $\phi = \theta^2$, then $p_\phi(\phi) = p_\theta(\sqrt{\phi})/(2\sqrt{\phi})$.

Jeffreys' prior

- Recall that the Fisher information for θ is defined as follows:

$$
\begin{aligned}
i(\theta) &= E\left[\left(\frac{d\log(p(y|\theta))^2}{d\theta}\right)|\theta\right] \\
&= -E\left[\frac{d^2\log(p(y|\theta))}{d\theta^2}|\theta\right]
\end{aligned}
$$

- Fisher information in terms of ϕ is

$$
i(\phi) = i(\theta)(\frac{d\theta}{d\phi})^2
$$

Jeffreys' prior

- Therefore,

$$\sqrt{i(\phi)} \quad = \quad \sqrt{i(\theta)}|\frac{d\theta}{d\phi}|$$

- That is, we can achieve the invariance requirement by setting

$$p_\theta \propto \sqrt{i(\theta)}$$

- This is called Jeffreys' prior.

Jeffreys' prior

- For example, for the binomial model, we have:

$$i(\theta) = \frac{n}{\theta(1-\theta)}$$

- That is, $p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$, which is a Beta(1/2, 1/2) distribution.

- Recall the flat uniform prior we use before is equivalent to Beta(1, 1).

Jeffreys' prior

- Jeffrey's prior for the mean and variance of normal $N(\theta, \sigma^2)$ distribution is the same as the improper priors we discussed above:

$$\begin{aligned} p(\theta) &\propto 1 \\ p(\sigma^2) &\propto 1/\sigma^2 \end{aligned}$$

- For the parameter of Poission(θ) model, Jeffrey's prior is $1/\sqrt{\theta}$.