STATS 700 Bayesian Inference and Computation **Fall 2017**

Lecture 4: Sept 25 27

Lecturer: Yang Cheng *Scribes: Anwesha Bhattacharyya, Jingqi Liu*

4.1 Nuisance Parameters

In a given model there might exist certain parameters that help in interpretation on computation, however these are not the parameters of interest, such parameters are called nuisance parameters.

For example if the parameters in a model are (θ_1, θ_2) where we are interested only in $p(\theta_1|y)$, then θ_2 is a nuisance parameter. Here Y is the data available.

It is easy to obtain $p(\theta_1|Y)$ from $p(\theta_1, \theta_2|y)$ by integrating out θ_2.

The necessity of a nuisance parameter may simply arise from the fact that it is easier to sample from $p(\theta_2|y)$ and then conditionally simulate θ_1 from $p(\theta_1|\theta_2, y)$.

4.2 Non-informative priors

To quote Andrew Gelman, "Prior distributions that are uniform, or nearly so, and basically allow the information from the likelihood to be interpreted probabilistically. These are noninformative priors, or maybe, in some cases, weakly informative."

Gaussian example

Let $\{Y_i\} \sim iidGaussian(\mu, \sigma^2)$. We can place a non informative uniform prior on $(\mu, \log(\sigma))$. The reson for inducing a uniform prior on $\log(\sigma)$ is that it can take any value on the real line and is not bounded by the positivity constraint. But, more importantly it induces the conjugate prior inverse gamma or inverse chi-square on the parameter σ^2.

Indeed the prior on (μ, σ) is given as

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

This results in the following posterior distribution.

$$p(\mu, \sigma^2|y) \propto \sigma^{-n-2} \exp\left(\frac{-1}{2\sigma^2}\left[(n-1)s^2 + n(\mu - \bar{y})^2\right]\right)$$

As a consequence, we have the following

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

- $(\sigma^2|y) \sim Inv - \chi^2(n-1, s^2)$
- $(\mu|y) \sim t_{n-1}(\bar{y}, s^2/n)$

It can be noted that as a result of non-informative prior the first conditional posterior of $(\mu|\sigma^2, y)$ is data driven and gives the same inference as a frequentist approach with concentrated variance σ^2/n. HOwever marginalizing out σ^2 we get the conditional posterior to be a t-distribution with heavier tails as compared to the gaussian distribution. This is essentially to account for the uncertainty of the data which is not known anymore and has been integrated out.

4.3 Jeffery's Prior

Jeffery?s Prior is proportional to the square root of the determinant of the Fisher information matrix.

$$\mathbb{P}(\theta) = \sqrt{|det\mathcal{I}(\theta)}$$

Jeffery?s Prior is invariant under reparameterization. Proof: Suppose $\Theta = \Phi(\theta)$

$$\mathcal{I}(\Theta) = \mathbb{E}(\frac{d \log P(y|\Theta)}{d\Theta})^2$$

$$\mathcal{I}(\theta) = \mathbb{E}(...)|\frac{d\Theta}{d\theta}|^2$$

$$\sqrt{\mathcal{I}(\theta)} = \sqrt{\mathcal{I}(\Theta)}|\frac{d\Theta}{d\theta}|$$

4.4 Jeffery's Prior for Binomial Distribution

Suppose $y \sim Binomial(n,p), thenitslog - likelihoodis \mathcal{L} = y \log p + (n-y) \log{(1-p)}$ Then

$$\mathbb{E}(-\frac{\partial L^2}{\partial p^2}) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}$$

So, under Jeffery?s method, the prior should look like the above. The posterior distribution then will be Beta Distribution

$$\mathbb{P}(p|y) = p^{y-\frac{1}{2}}(1-p)^{n-y-\frac{1}{2}}$$

4.5 Jeffery's Prior for Gaussian Distribution

When both μ and σ are known, $\mathbb{P}(\mu)P(\sigma^2)$ are more accepted prior compared with $P(\mu, \sigma^2)$ Suppose $\mu \sim N(0, k_0^2)andy \sim N(\mu, \sigma^2)$ Then the Jeffery?s Prior for μ will be $\mathbb{P}(\mu) \propto exp(-\frac{\mu^2}{2k_0^2})$

$$\mathbb{P}(\mu, \sigma^2) \sim \mathbb{P}(\mu|\sigma^2)\mathbb{P}(\sigma^2) \times Likelihood$$

$$\mathbb{P}(\mu, \sigma^2) = \sigma^2 exp(-\frac{\sum(y_i - y)^2}{2\sigma^2} - \frac{(y-\mu)^2}{2\sigma^2} - \frac{\mu^2}{2k_0^2})$$

If we know σ is 1 and $y \sim N(\mu, 1)$, by intuition, the Jeffery's prior of μ will be flat as the second derivative of log-likelihood is a constant and therefore, it's expectation is a constant.

4.6 Why is Jeffery's Prior useful?

It is useful because of its invariance property. Consider for instance the binomial model with unknown proportion parameter θ and odds parameter $\psi = \frac{\theta}{1-\theta}$. The Jeffrey's posterior on θ reflects as best as possible the information about brought by the data. There is a one-to-one correspondence between θ and ψ. Then, transforming the Jeffrey's posterior on θ into a posterior on ψ (via the usual change-of-variables formula) should yield a distribution reflecting as best as possible the information about ψ. Thus this distribution should be the Jeffrey's posterior about ψ. This is the invariance property. Also, an important point when drawing conclusions of a statistical analysis is scientific communication. Imagine you give the Jeffrey's posterior on θ to a scientific colleague. But he/she is interested in rather than θ. Then this is not a problem with the invariance property: he/she just has to apply the change-of-variables formula.

4.7 Conjugate Priors for Gaussian Distribution

If the posterior distributions $\mathbb{P}(\theta|y)$ are in the same family as the prior probability distribution $\mathbb{P}(\theta)$, the prior is called a conjugate prior. From the example of Binomial distribution, the Jeffery's Prior of p when the number of trials are fixed is beta distribution is also in exponential family. So, the Jeffery's prior for p is a conjugate prior.

prior

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

$$(\sigma^2) \sim Inv - \chi^2(\nu_0, \sigma_0^2)$$

Conditional and marginal posteriors

$$(\mu|\sigma^2, y) \sim N(\mu_n, \frac{\sigma^2}{\kappa_n})$$

$$(\sigma^2|y) \sim Inv - \chi^2(\nu_n, \sigma_n)$$

$$(\mu|y) \sim t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n)$$

Here, we have

$$\kappa_n = n + \kappa_0 \tag{4.1}$$

$$\mu_n = \frac{\sigma^2}{\kappa_n}\Big[\frac{\kappa_0\mu_0 + n\bar{y}}{\sigma^2}\Big] \tag{4.2}$$

$$\nu_n = \nu_0 + n \tag{4.3}$$

$$\sigma_n^2\nu_n = \sigma_0^2\nu_0 + \sum_{i=1}^{n}(y_i - \bar{y})^2 + \frac{(\bar{y} - \mu_0)^2}{1/n + 1/\kappa_0} \tag{4.4}$$

Multivariate Gaussian

If the variance is known we again use a conjugate Gaussian prior on the mean. If the variance is unknown then we can either use a non-informative prior as in the first case or use a conjugate Wishart distribution for the covariance. Once we get the joint posterior distribution little algebraic maneuver will yield the posterior predictive distribution.

4.8 Bioassay Experiment

In clinical trials to determine the dosage of a drug, one needs to carry out bioassay experiments where different dosage levels are implement on different batches of animals/patients and response is analyzed. The setup here includes (n_i, y_i, x_i) where x_i is the dose level administered to i_{th} batch. There are "k" dose levels corresponding to the "k" batches. y_i is the positive response, in this case the number of alive mice out of the n_i mice given the x_i dose. Since y_i is count data with finite options it can be modelled as $Bin(n_i, \theta_i)$ where θ_i is the probability of success for dose x_i. Hence we can use a logit model to explain the dose response relation as

$$logit(\theta_i) = \alpha + \beta x_i$$

Using a uniform prior $p(\alpha, \beta) \propto 1$ we can carry out the posterior inference of the model and use R codes to draw the contour plots. Samling from the posterior we can get a histogram of the parameter $\frac{-\aleph}{\beta}$ which gives us an idea of dosage required for 50% survival rate.

$$\begin{aligned} &\theta_i = .5 \\ \implies &logit(\theta_i) = 0 \\ \implies &\alpha + \beta x_i = 0 \\ \implies &x_i = \frac{-\alpha}{\beta} \end{aligned}$$

References

[1] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*, Volume 2. CRC press Boca Raton, FL, 2014.