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

Lecture 5: Oct 4 Hierarchical Models

Lecturer: CHEN Yang *Scribe: Po-Heng Chen*

5.1 Improper Prior

Setting up an improper prior (or non-informative prior) could be very dangerous to use in practice or in general problem because if you set up such kind of prior and you cannot check this condition. Very likely you will have an improper prior distribution and, whatever your inference is, it will be invalid because you are not making inference from a proper prior.

A general recommendation in practice is to set up a proper prior. For example, it is a positive value such as Gamma prior.

5.2 Exchangeability

Exchangeability means we do not have any information other than the data about how to distinguish the different parameters. In **rat tumor** example, it's we don't have information to distinguish the rats from different labs. We don't have the information to distinguish what is the rat tumor probability from different rats.

Formal mathematical definition: The parameters $(\theta_1, ..., \theta_J)$ are exchangeable if the density $p(\theta_1, ..., \theta_J)$ is invariant to permutations of the indexes $(1, ..., J)$.

Simple form:

$$p(\theta) = \int \left(\prod_{j=1}^{J} p(\theta_j|\phi) \right) p(\phi) d\phi$$

de Finetti's theorem
If you have some exchangeable distribution, you will be able to express it in terms of i.i.d observations and then adding a prior on the parameters.

5.3 Gaussian Example

Observations' distributions: $\theta_1, \theta_2, ..., \theta_j$ where θ are Gaussian means.
Observations: $y_{1j}, y_{2j}, ...y_{nj}$ for θ_j

We also assume that all of the θ values come from a global distribution, which is another Gaussian distribution with (μ, τ^2). Then we can further assume a prior $P_0(\mu, \tau^2)$ for $N(\mu, \tau^2)$.

This is a hierarchical model where the observations at the bottom, parameters at the middle and hyper-parameters at the top.

How to make inference form this model?

Since y_j follows $N(\theta_j, \sigma^2)$, $\bar{y}_j = \frac{\sum_{j=1}^{n_j} y_j}{n_j}$ follows $N(\theta_j, \frac{\sigma^2}{n_j})$.

For each group, we try to calculate what is the sample mean. Because if we forget about the hierarchical model, a very natural estimate of θ_j is the $\bar{y}_j$.

We can simplify the observations using summery statistics:

$$\begin{aligned}
\theta_1 &\rightarrow \bar{y}_1 \sim N(\theta_1, \frac{n_1}{\sigma^2}) \\
\theta_2 &\rightarrow \bar{y}_2 \sim N(\theta_2, \frac{n_2}{\sigma^2}) \\
&\dots \\
\theta_j &\rightarrow \bar{y}_j \sim N(\theta_j, \frac{n_j}{\sigma^2})
\end{aligned}$$

Then we are able to write down the joint distribution of μ, τ^2, all the θ_J, conditioning on the observations:

$$\begin{aligned}
p(\mu, \tau^2, \{\theta_i\}_{i=1}^J | \bar{y}_1, ...\bar{y}_j) &= p(\mu, \tau^2) \prod_{i=1}^J p(\theta_i | \mu, \tau^2) \prod_{i=1}^J p(\bar{y}_i | \theta_i, \sigma_i^2) \\
&= p(\mu, \tau^2) \prod_{i=1}^J \{p(\theta_i | \mu, \tau^2) p(\bar{y}_i | \theta_i, \sigma_i^2)\}
\end{aligned}$$

If we know $p(\mu, \tau^2 | observations)$, posterior of θ_i is $N\big(\frac{\frac{\mu}{\tau^2} + \frac{\bar{y}_i}{\sigma^2}}{\frac{1}{\tau^2} + \frac{1}{\sigma_i^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{1}{\sigma_i^2}}\big)$.

$$\begin{aligned}
p(\mu, \tau^2 | \bar{y}_1, ..., \bar{y}_j) &= \int p(\mu, \tau^2, \{\theta_i\}_{i=1}^J | \bar{y}_1, ...\bar{y}_j) d\theta_1 ... d\theta_J \\
&= p(\mu, \tau^2) \prod_{i=1}^J \int p(\theta_i | \mu, \tau^2) p(\bar{y}_i | \theta_i, \sigma_i^2) d\theta_i
\end{aligned}$$

References

References

[1] Bradley Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, Volume 1. Cambridge University Press, 2012.

[2] 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.