10-801: Advanced Topics in Graphcal Models 10-801, Spring 2007

Lecture 1. Basics of hierarchical Bayesian models

Lecturer: Eric P. Xing *Scribes: Kyung-Ah Sohn*

1 Bayesian Method

Consider a statistical model with parameter θ and observation data X which follow some distributions f and π such that:

$$X \sim f(\cdot \mid \theta), \quad \theta \sim \pi(\theta)$$

Then from Bayes theorem, we have:

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)} = \frac{P(X \mid \theta)P(\theta)}{\int P(X \mid \theta)P(\theta)d\theta}$$

This means that we can compute the *posterior* probability distribution of the parameter θ given the observation X from some *prior* distribution $P(\theta)$ and the likelihood function $P(X \mid \theta)$. Some people criticize this Bayesian framework as too subjective. But is Bayesian really subjective?

Example 1 *Suppose we have a coin and we want to test if the probability of head θ is equal to or greater than 0.5. That is, we do the following hypothesis test:*

$$H_0: \quad \theta = \tfrac{1}{2}$$
$$H_1: \quad \theta \geq \tfrac{1}{2}$$

Now we observed 9 heads and 3 tails in our experiment. Let Z denote the event that the toss is a head. Then our assumption is that

$$Z \sim Ber(\theta)$$

. Here are three different interpretation of this observation.

1. *If we let X the total number of* heads *we observed out of 12 tossing, X would follow Binomial distribution:*

   $$X \sim Binomial(n, \theta)$$

   Then under the null hypothesis, we have

   $$P(X \geq 9) = \sum_{X=9}^{12} P(X \mid n, \theta) = \sum_{X=9}^{12} P(X \mid 12, \frac{1}{2}) = 0.073$$

   and we can use this statistics to decide whether to reject the null hypothesis or not.

2. *Now, we can think of another interpretation of this observation: we tossed the coin until we observed 3 tails and we ended up with 9 heads and 3 tails. Then X would follow the Negative-Binomial distribution:*

$$X \sim NB(3, \theta_t) = \left(\begin{array}{c} 3 + X - 1 \\ X \end{array} \right) \theta_t^3 (1 - \theta_t)^X$$

Then again under the null hypothesis

$$P(X \geq 9) = \sum_{X=9}^{\infty} NB(X \mid 3, \theta_s) = \sum_{X=9}^{\infty} NB(X \mid 3, \frac{1}{2}) = 6.730$$

and this value is different from that estimated in the previous model assumption. This means that parametric approaches are not always objective, and the different interpretation of the experiment can result in an entire different conclusion.

3. *Finally, consider the Bayesian framework where we assume:*

$$X \sim Poisson(\cdot \mid \lambda)$$

$$\lambda \sim Gamma(\lambda \mid r, \frac{1-p}{p})$$

Then,

$$\begin{aligned} P(X \mid r, \frac{1-p}{p}) &= \int P(x, \lambda \mid r, \frac{1-p}{p}) d\lambda \\ &= \int P_0(x \mid \lambda) Gamma(\lambda \mid r, \frac{1-p}{p}) d\lambda \\ &= NB(X \mid r, p) \end{aligned}$$

and this reduces to the same parametric model as in 2.

Example 2 *Suppose*

$$n_i \sim Multinomial(\overrightarrow{\theta})$$

Parametric approach would produce something like:

$$\theta_i = \frac{n_i}{N} \qquad or \qquad \theta_i = \frac{n_i + 1}{N + 1}$$

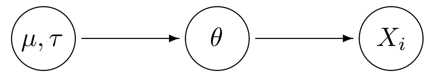
Under the following prior distribution

$$\theta_i \sim Dir(\overrightarrow{\alpha})$$

the Bayes estimator would be derived as:

$$\theta_{BE} = \frac{n_i + \alpha_i}{N + \sum \alpha_i}$$

Normal Bayes estimates



Consider the following model s.t.

$$X_i \sim N(\theta, \sigma^2)$$

$$\theta \sim N(\mu, \tau^2)$$

for some hyper-parameters μ and τ.

Then

$$\begin{aligned} P(\theta, X) &\propto \exp\left(-\frac{(X-\theta)^2}{\sigma^2} - \frac{(\theta-\mu)^2}{\tau^2}\right) \\ &= \exp\left(-\frac{1}{2\rho}\left[\theta - \rho(\frac{X}{\sigma^2} + \frac{\mu}{\tau^2})\right]^2 - \frac{1}{2(\sigma^2+\tau^2)}(X-\mu)^2\right) \\ &= \exp\left(-\frac{1}{2\rho}\left[\theta - \rho(\frac{X}{\sigma^2} + \frac{\mu}{\tau^2})\right]^2\right)\exp\left(-\frac{1}{2(\sigma^2+\tau^2)}(X-\mu)^2\right) \\ &\propto P(\theta \mid X)P(X) \end{aligned}$$

where $\rho = \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}$. Hence, without complicated computation, we can induce:

$$\theta \mid X \sim N\left(\rho\left(\frac{X}{\sigma^2} + \frac{\mu}{\tau^2}\right), \rho\right) \tag{1}$$

$$X \sim N(\mu, \sigma^2 + \tau^2)$$

Given N observations: $X_1, \ldots, X_n$,

$$\overline{X} \sim N(\theta, \sigma^2/n)$$

By replacing X and σ^2 with $\overrightarrow{X}$ and $\frac{\sigma^2}{n}$ in (1), we get the following Bayes estimator:

$$\begin{aligned} \theta_{Bayes} &= \frac{\tau^2}{\sigma^2/n + \tau^2}\overline{X} + \frac{\sigma^2/n}{\sigma^2/n+\tau^2}\mu \\ &= \overline{X} - \frac{\sigma^2/n}{\sigma^2/n+\tau^2}(\overline{X} - \mu) \end{aligned}$$

Lemma 1

$$X \sim h(X)\exp(\theta \cdot X - \phi(\theta))$$

For any prior $\pi(\theta)$, the posterior mean is

$$\delta^\pi(X) = \nabla_X \log m_\pi(X) - \nabla_X \log h(X)$$

where $m_\pi(X) = \int P(X \mid \theta)\pi(\theta)d\theta$.

Consider the model $f(x \mid \theta), \pi(\theta)$ with the conditional assumptions on parameters

$$\pi_1(\theta \mid \theta_1), \pi_2(\theta_1 \mid \theta_2), \ldots, \pi_n(\theta_{n-1} \mid \theta_n), \pi_n(\theta_n)$$

s.t.

$$\pi(\theta) = \int \pi_1(\theta \mid \theta_1)\pi_2(\theta_1 \mid \theta_2) \cdots \pi(\theta_n) d\theta_1 \ldots \theta_n$$

Conditional Decomposition:

$$\pi(\theta \mid X) = \int \pi(\theta \mid X, \theta_1) P(\theta_1 \mid X) d\theta_1$$

where

$$\pi(\theta \mid X, \theta_1) = \frac{f(X \mid \theta)\pi(\theta \mid \theta_1)}{m_1(X \mid \theta_1)}$$

and

$$m_1(X \mid \theta_1) = \int f(X \mid \theta)\pi(\theta \mid \theta_1) d\theta$$

Then

$$\pi(\theta_1 \mid X) = \frac{m_1(X \mid \theta_1)\pi_2(\theta_1)}{m_2}$$

Actually, we have the following decomposition

$$E^{\pi}\left(h(\theta) \mid X\right) = E^{\pi(\theta_1 \mid X)}\left(E^{\pi(\theta \mid \theta_1, X)}[h(\theta)]\right)$$