Corrections for Quiz 1

Yifei Wang

1. (a). $P(Y|X) = \dfrac{P(X|Y)\,P(Y)}{P(X)}$

(b). If $p(y_1, \cdots, y_n) = P(y_{\pi_1}, \cdots, y_{\pi_n})$ for ~~all~~ any finit n and all permutation π, then $y_1, \cdots, y_n$ sequence is finite exchangeble.

(c). For a given mode $P(Y|\theta)$ and its paratmeter θ, and a ~~class~~ ~~of~~ family of distribution $\mathcal{P}$, we say $\mathcal{P}$ is conjugate to $p(Y|\theta)$ and θ, if $P(\theta) \in \mathcal{P}$ we always have $P(\theta|Y) \in \mathcal{P}$

(d) True

(e) True

2. (a). Suppose $y_1 \cdots y_n \overset{iid}{\sim} p(y|\theta)$, we have

$$L(Y) = \log P(y_1 \cdots y_n|\theta) = \sum_{i=1}^{n}\left[-\log 2\theta - \frac{|y_i|}{\theta}\right]$$

$$\frac{dL}{d\theta} = \sum_{i=1}^{n}\left[\frac{-1}{\theta} + \frac{|y_i|}{\theta^2}\right] = \sum_{i=1}^{n}\left[\frac{|y_i| - \theta}{\theta^2}\right]$$

Let the derivitive to be 0, we have $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}|y_i|$

(b). $P(\theta|y_1\cdots y_n) \propto P(y_1\cdots y_n|\theta)\,P(\theta)$

$$\propto P(\theta)\cdot\frac{1}{\theta^n}\cdot\exp\left(-\frac{\Sigma|y_i|}{\theta}\right)$$

Thus, $P(\theta)$ at minimum should have terms like $\frac{1}{\theta^{c_1}}\exp\left(-\frac{c_2}{\theta}\right)$

An Inverse-Gamma prior could satisfy this.

Suppose $P(\theta) \sim \text{InverseGamma}(\alpha, \beta)$

$$P(\theta|y_1\cdots y_n) \propto \theta^{-n}\cdot\exp\left(-\frac{\Sigma|y_i|}{\theta}\right)\cdot\theta^{-(\alpha+1)}\cdot\exp\left(-\frac{\beta}{\theta}\right)$$

$$\propto \theta^{-(\alpha+n+1)}\cdot\exp\left(-\frac{\Sigma|y_i|+\beta}{\theta}\right)$$

$$\propto \text{InverseGamma}\left(\alpha+n,\ \beta+\sum_{i=1}^{n}|y_i|\right)$$

(c) Posterior mean: $E[P(\theta|y_1\cdots y_n)] = \dfrac{\beta+\sum_{i=1}^{n}|y_i|}{\alpha+n-1}$

$$= \underbrace{\frac{\beta}{\alpha-1}}_{\text{prior mean}}\frac{\alpha-1}{\alpha+n-1} + \underbrace{\frac{\sum_{i=1}^{n}|y_i|}{n}}_{\text{MLE}}\frac{n}{\alpha+n-1}$$

$\alpha-1$: prior sample size $\quad\beta$: prior sum of absolute sample values.

n: data sample size

Posterior mean is a weighted sum (weighted by prior/data sample size) of prior mean and the MLE

3. (a). Jeffreys' prior $\propto \sqrt{I(\theta)}$

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log P(y|\theta)\right]$$

$$\log P(y|\theta) = \log\frac{\exp(-\theta)\,\theta^y}{y!} = -\theta + y\log\theta - \log y!$$

$$\frac{\partial^2}{\partial\theta^2}\log P(y|\theta) = \frac{\partial}{\partial\theta}\left(-1+\frac{y}{\theta}\right) = \frac{-y}{\theta^2}$$

Thus, $I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log P(y|\theta)\right] = -E\left[\frac{-y}{\theta^2}\right] = \frac{\theta}{\theta^2} = \frac{1}{\theta}$

Then, we have Jeffrey's prior $\propto \sqrt{I(\theta)} = \frac{1}{\sqrt{\theta}}$

Since $\int_0^{+\infty}\frac{1}{\sqrt{\theta}}d\theta = 2\sqrt{\theta}\Big|_0^{\infty} = \infty$, this prior distribution is not proper.

(b). $P(\theta|y_1,\cdots,y_n) \propto P(\theta)\,P(y_1,\cdots,y_n|\theta)$

$$\propto \frac{1}{\sqrt{\theta}}\cdot\exp(-n\theta)\cdot\theta^{\sum_{i=1}^{n}y_i}$$

$$\propto \theta^{\sum_{i=1}^{n}y_i-\frac{1}{2}}\cdot\exp(-n\theta)$$

$$\propto \text{Gamma}\left(\sum_{i=1}^{n}y_i+\frac{1}{2},\ n\right)$$

This shows that the posterior distribution of this Jeffrey's prior is proper.