Reading: Chp 4, Chp.5

Ex: $X_i \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$, $\mu \mid \sigma^2 \sim N(\mu_0, \sigma^2/\underbrace{k_0}_{\text{psuedo samplesize}})$

Consider log posterior and approximate in as $n \to \infty$

Assume $X_1, \dots, X_n$ iid f_θ , $\theta \in \mathbb{R}^d$

$$\log \mathbb{P}(\theta \mid X) = \pi(\theta) = \log \mathbb{P}(x \mid \theta) + \log \mathbb{P}(\theta) - \log \mathbb{P}(X)$$

Let $\theta^* = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \pi(\theta)$ and do a second order Taylor expansion

$$\pi(\theta) \simeq \pi(\theta^*)^T + (\theta - \theta^*) \underbrace{\frac{\partial \pi}{\partial \theta}(\theta^*)}_{\begin{pmatrix} d\pi/d\theta_1 \\ \vdots \\ d\pi/d\theta_p \end{pmatrix}} + \frac{1}{2}(\theta - \theta^*)^T \underbrace{H_\pi(\theta^*)}_{= \left[\frac{d^2\pi}{d\theta_i\, d\theta_j}\right]_{ij}} (\theta - \theta^*)^T$$

Because θ^* is maximal $d\pi/d\theta\,(\theta^*) = 0$ and

$$\pi(\theta) \simeq \pi(\theta^*) - \frac{1}{2}(\theta - \theta^*)^T \underbrace{\left[-H_\pi(\theta^*)\right]}_{I(\theta^*)} (\theta - \theta^*)$$

$$\Rightarrow \quad \exp(\pi(\theta^*)) \exp\left\{-\frac{1}{2}(\theta - \theta^*)^T I(\theta^*) (\theta - \theta^*)\right\}$$

Thus $\theta | x \approx N\left(\theta^*, I(\theta^*)^{-1}\right)$ "Laplace Approximation"

Rmks: - Easy to summarize posterior

$$- \pi(\theta) = \log \mathbb{P}(\theta) + n\left[\frac{1}{n}\sum_{i=1}^{n} \log \mathbb{P}(x_i | \theta)\right] + C_n$$

So as n grows this is dominated by MLE

$$\Rightarrow \underset{\theta}{\operatorname{argmax}}\ \pi(\theta) \approx \underset{\Theta}{\operatorname{argmax}}\ \mathcal{L}(\theta; x)$$

So... Bernstein von Mises

$$\theta | x \approx N\left(\hat{\theta}_{MLE}, I(\hat{\theta}_{MLE})^{-1}\right)$$

Rmk: - Laplace gives an easy approximate way to summarize the posterior

- Sample size grows, we depend less on prior

- $$I(\theta) = -\frac{d^2 \pi}{d\theta\, d\theta^T} = \frac{-\partial^2 \log \mathbb{P}(\theta)}{\partial\theta\, \partial\theta^T} - \frac{\partial^2 \log \mathbb{P}(x|\theta)}{\partial\theta\,\partial\theta^T}$$

$$= - H_{\log \mathbb{P}(\theta)}(\theta) - n \underbrace{\left[\frac{1}{n}\sum_{i=1}^{n} \frac{-\partial^2 \log \mathbb{P}(x_i|\theta)}{\partial\theta\,\partial\theta^T}\right]}_{SLLN}$$

$$\approx - H_{\log \mathbb{P}(\theta)}(\theta) + n\, \mathbb{E}_{x_i|\theta}\left[\frac{-\partial^2 \log \mathbb{P}(x_i|\theta)}{\partial\theta\,\partial\theta^T}\right]$$

$$= \underbrace{-H_{\log \mathbb{P}(\theta)}(\theta)}_{\text{prior precision}} + \underbrace{nJ(\theta)}_{\text{data precision}}$$

<u>Example:</u> $Y_i \mid \theta \overset{\text{ind}}{\sim} Po(x_i \theta)$, $\theta \sim Gamma(\alpha, \beta)$

$$\theta \mid X \sim Gamma(\alpha + n\bar{y}, \beta + n\bar{x})$$

fact: gamma mode $\frac{\alpha - 1}{\beta}$

So post. mode $\theta^* = \frac{\alpha + n\bar{y} - 1}{\beta + n\bar{x}} = \frac{\bar{y} + \alpha - 1/n}{\bar{x} + \beta/n}$

If $\alpha = 1, \beta = 0$ then $\theta^* = \bar{y}/\bar{x} = \hat{\theta}_{MLE}$ (flat prior gives MLE)

If $n \to \infty$ then $\theta^* \longrightarrow \hat{\theta}_{MLE}$

$$\pi(\theta) = \log \mathbb{P}(\theta \mid x) = (\alpha + n\bar{y} - 1)\log\theta - (\beta + n\bar{x})\theta$$

$$\frac{\partial \pi}{\partial \theta} = \frac{\alpha + n\bar{y} - 1}{\theta} - (\beta + n\bar{x})$$

$$\frac{\partial^2 \pi}{\partial \theta^2} = -\frac{(\alpha + n\bar{y} - 1)}{\theta^2} \Rightarrow I(\theta) = \frac{\alpha + n\bar{y} - 1}{\theta^2}$$

$$\theta \mid x \simeq N\left(\theta^*, \frac{\theta^{*2}}{\alpha + n\bar{y} - 1}\right)$$