206 B Midterm 2 Review

Hierarchical Bayes:

$x \sim f(x|\theta)$, $\theta \sim \pi_1(\theta|\theta_1), \cdots, \theta_n \sim \pi_{n+1}(\theta_n)$

Prior: $\pi(\theta) = \int_{\theta_1 \times \theta_2 \cdots \theta_n} \pi_1(\theta|\theta_1)\pi(\theta_2|\theta_2) \cdots \pi_{n+1}(\theta_n)\, d\theta_1 d\theta_2 \cdots d\theta_n$

$\hookrightarrow$ Marginal prior of the parameter that we are interested in.

Example: Normal Hierachical

$$\begin{tikzcd}
 & \mu, \tau^2 \arrow[dl] \arrow[d] \arrow[dr] & \\
\theta_1 \arrow[d] & \theta_2 \cdots \arrow[d] & \theta_p \arrow[d] \\
x_1 & x_2 \cdots & x_p
\end{tikzcd}$$

$\leftarrow \pi(\mu, \tau^2) = \pi_{21}(\mu|\tau^2) \cdot \pi_{22}(\tau^2) = N(\mu_0, k\tau^2)\, IG(a,b)$

$\leftarrow \theta_i \overset{iid}{\sim} N(\mu, \tau^2)$

$\leftarrow x_i|\theta_i \sim N(\theta_i, \sigma^2)$, σ^2 known

Joint Posterior:

$$\pi(\theta_1, \cdots, \theta_p, \mu, \tau^2|x) \propto \Pi_{i=1}^p f(x_i|\theta_i) \cdot \Pi_{i=1}^p (\theta_i|\mu, \tau^2) \cdot \pi_{21}(\mu|\mu_0, k\tau^2) \cdot \pi_{22}(\tau^2|a,b)$$

$$= \Pi^p (2\pi\sigma^2)^{-\frac{p}{2}} \exp\left(-\frac{\sum_{i=1}^p (x_i-\theta_i)^2}{2\sigma^2}\right)(2\pi\tau^2)^{-\frac{p}{2}} \exp\left(-\frac{\sum_{i=1}^p(\theta_i-\mu)^2}{2\tau^2}\right)(2\pi k\tau^2)^{-\frac{1}{2}} \exp\left(-\frac{(\mu-\mu_0)^2}{2k\tau^2}\right) \cdot (\tau^2)^{-a-1} \exp\left(-\frac{b}{\tau^2}\right) \cdot \frac{b^a}{\Gamma(a)}$$

$$\pi(\mu|\theta_1, \cdots, \theta_p, \tau^2, x) \propto \exp\left(-\frac{\sum(\theta_i-\mu)^2}{2\tau^2} - \frac{(\mu-\mu_0)^2}{2k\tau^2}\right) \propto \exp\left(-\frac{1}{2}\left(\left(\frac{p}{\tau^2}+\frac{1}{k\tau^2}\right)\mu^2 - 2\mu\left(\frac{\sum\theta_i}{\tau^2}+\frac{\mu_0}{k\tau^2}\right)\right)\right)$$

$$\Rightarrow \pi(\mu|\theta_1, \cdots, \theta_p, \tau^2, x) \sim N\left(\frac{\sum\theta_i/\tau^2 + \mu_0/k\tau^2}{p/\tau^2 + 1/k\tau^2}, \left(p/\tau^2 + 1/k\tau^2\right)^{-1}\right)$$

$$\pi(\tau^2|\mu, \theta_1, \cdots, \theta_p, x) \propto (\tau^2)^{-\frac{p}{2}-\frac{1}{2}-a-1} \exp\left(-\frac{1}{\tau^2}\left(\frac{\sum_{i=1}^p(\theta_i-\mu)^2}{2} + \frac{(\mu-\mu_0)^2}{2k}\right)\right) \sim IG$$

$$\pi(\theta_i|\mu, \text{other } \theta, x) \propto \exp\left(-\frac{(x_i-\theta_i)^2}{2\sigma^2} - \frac{(\theta_i-\mu)^2}{2\tau^2}\right) \sim N\left(\frac{x_i/\sigma^2 + \mu/\tau^2}{1/\sigma^2 + 1/\tau^2}, \left(1/\sigma^2 + 1/\tau^2\right)^{-1}\right)$$

PS: Mixture of Conjugate priors are still conjugate.

Non-Informative Priors:

1. Laplace's Priors (Uniform Priors or Flat Priors): Assign the equiprobability to elementary events. When Θ is a finite set, consisting of n elements, the Obvious non-informative prior is to give each element of Θ probability $1/n$.
2. Invariance under reparameterization: Consider $\eta = g(\theta)$, $g(\cdot)$ is monotone over the domain of θ. Find the induced prior for η: $\pi_\eta(\eta) = \pi_\theta(g^{-1}(\eta))\, |dg^{-1}(\eta)/d\theta|$. A more intrinsic and more acceptable notion of noninformative prior should satisfy invariance under reparameterization, i.e. $\pi_\eta(\eta)$ is also a flat prior.

Invariant Priors: for a location parameter θ: $\pi(\theta) = C$, for a scale parameter σ: $\pi(\sigma) = 1/\sigma$.

Fisher Information:

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

$\hookrightarrow$ under regular condition (True for exponential family)

Jeffrey's Prior:

It's non-informative priors in general settings based on fisher information. $\pi^*(\theta) \propto [I(\theta)]^{1/2}$

Jeffrey's is invariant, $\because$ $I(\theta) = I(h(\theta)) \cdot [h'(\theta)]^2$ for $\theta \to h(\theta)$

For multidimensional θ, $\pi^*(\theta) \propto [\det(I(\theta))]^{1/2}$

☆: Jeffrey's Prior doesn't satisfy Likelihood principle.

Reference Prior:

For Multidimensional $\theta=(\theta_1,\theta_2)$, distinguish between θ_1 (Parameters of Primary Interest) and θ_2 (Nuisance Parameters)

Conditional on θ_1, define prior $\pi(\theta_2|\theta_1)$ as the Jeffrey's Prior associated with $f(x|\theta_1,\theta_2)$.

Then, $p(x|\theta_1)=\int_{\theta_2} p(x|\theta_1,\theta_2)\pi(\theta_2|\theta_1)d\theta_2$ $\Rightarrow$ we get a $f(x|\theta_1)$ about only θ_1

Then, find Jeffrey's prior based on $p(x|\theta_1)$ $\Rightarrow$ Then set $\pi(\theta_1,\theta_2)=\pi(\theta_1)\pi(\theta_2|\theta_1)$

Robust Prior Distributions: Parameterized distributions as insensitive as possible to small variation in prior.

Ps: We can robustify the conjugate priors by hierachical modeling.

Exchangeability:

Finite exchangeability: Let $P(y_1,\cdots,y_n)$ be the joint density of $Y_1,\cdots Y_n$, if $P(y_1,\cdots,y_n)=P(y_{\pi_1},\cdots,y_{\pi_n})$ for all permutations π of $\{1,\cdots,n\}$, then $Y_1,\cdots,Y_n$ are exchangeable.

Ps: If $\theta\sim\pi(\theta)$ and $Y_1,\cdots,Y_n$ are conditionally iid, given θ, then marginally (unconditioned on θ) $Y_1,Y_2,\cdots,Y_n$ are exchangeable.

De Finetti's Theorem: Let $y_i\in\mathcal{Y}$. for any n, our belief model for $y_1,\cdots,y_n$ is exchangeable: $P(y_1,\cdots,y_n)=P(y_{\pi_1},\cdots,y_{\pi_n})$ for all permutations π of $(1,2,\cdots n)$. Then our model can be written as: $P(y_1,\cdots,y_n)=\int P(y_1,\cdots,y_n|\theta)\pi(\theta)d\theta$. for some parameter θ, some prior distribution on θ and some sampling model $P(y|\theta)$.

Laplace Approximation: First, we need the posterior distribution.

θ_0: The mode of $\pi(\theta|x)$, i.e. find θ_0, s.t. $\frac{dq(\theta)}{d(\theta)}=0$, $q(\theta)$ is the posterior distribution.

$$A=-\left(\frac{d^2\log(q(\theta))}{d\theta^2}\right)\Big|_{\theta=\theta_0} \Rightarrow \pi(\theta|x)\approx N(\theta_0, 1/A).$$

Ps: If it's MVN case:

$\theta_0=(\theta_{0j}, j=1,\cdots,p)$, such that $\frac{\partial q(\theta)}{\partial\theta_{0j}}=0$

A is Hessian matrix evaluated at θ_0, $A_{ij}=-\frac{\partial^2\log(q(\theta))}{\partial\theta_i\partial\theta_j}\Big|_{\theta=\theta_0}$.

Monte Carlo Method:

Suppose that we we have $\theta^{(1)},\cdots,\theta^{(M)}$ iid samples from $\pi(\theta|x)$, by LLN: as $M\to\infty$

$$\bar\theta=\frac{1}{M}\sum_{i=1}^M\theta^{(m)}\to E(\theta|x).\quad \frac{1}{M-1}\sum_{m=1}^M(\theta^{(m)}-\bar\theta)^2\to Var(\theta|x)$$

$\frac{1}{M}\#(\theta^{(m)}\le c)\to P(\theta\le c|x)$. And empirical distribution of $(\theta^{(1)},\cdots\theta^{(m)})\to\pi(\theta|x)$

Importance Sampling:

$$E(g(\theta)|x)=\frac{\int g(\theta)\pi(\theta|x)d\theta}{}=\frac{\int_\theta g(\theta)f(x|\theta)\pi(\theta)d\theta}{\int_\theta f(x|\theta)\pi(\theta)d\theta}$$

Use $h(x)$ as importance function

$$=\frac{1}{M}\sum_{m=1}^M g(\theta^{m})\cdot\frac{f(x|\theta^{m})\pi(\theta^{(m)})}{h(\theta^{(m)})}$$

A special case when $g(\theta)=1$

Metropolis-Hasting algorithms:

Step 1: Start with an arbitrary initial value $\theta^{(0)}$.

Step 2: Update from $\theta^{(t-1)}$ to $\theta^{(t)}$ by:

2.1: Generate $\xi\sim q(\xi|\theta^{(t-1)})$

2.2: $$\rho(\theta^{(t-1)},\xi)=\min\left\{\frac{\pi(\xi)\,q(\theta^{(t-1)}|\xi)}{\pi(\theta^{(t-1)})\,q(\xi|\theta^{(t-1)})},1\right\}.$$

2.3. $$\theta^{(t)}=\begin{cases}\xi & \text{with } p=\rho\\ \theta^{(t-1)} & \text{o.w.}\end{cases}$$

Gibbs Sampler:

Step 1: Start with an arbitrary value $\eta^{(0)}$.

Step 2: Given $\eta^{(t-1)}$, generate

2.1: $\theta^{(t)}$ from $\pi_1(\theta \mid x, \eta^{(t-1)})$

2.2: $\eta^{(t)}$ from $\pi_2(\eta \mid x, \theta^{(t)})$

Then: $(\theta^{(t)}, \eta^{(t)})$ is from the joint posterior distribution and θ is from marginal, same for η.

Bayes Point Estimation:

Report a point estimation for $h(\theta)$ with associated measure of accuracy.

$\Rightarrow$ Find $\pi(h(\theta) \mid x)$ and use the Bayes rule: i.e. a solution of $\min E^{\pi}(L(\theta,d) \mid x)$ for $d \in D$ and $\theta \in \Theta$.

We can use quadratic loss, absolute error loss, and 0-1 loss.

$\uparrow$ Posterior Mean, $\uparrow$ Posterior Median, $\uparrow$ Posterior Mode.

PS: $MSE(\delta) = Var(\delta) + (E\delta - \theta)^2 = Bias^2 + Var(\delta) \mid x$.

Bayes Hypothesis Testing:

$H_0: \theta \in \Theta_0$. $H_1: \theta \in \Theta_1$. with a loss function: $L(\theta, \varphi) = \begin{cases} 1 & \text{if } \varphi \neq I(\theta \in \Theta_0) \\ 0 & \text{o.w.} \end{cases}$

$\Rightarrow$ The Bayes decision is: $\varphi^{\pi}(x) = \begin{cases} 1 & \text{if } P^{\pi}(\theta \in \Theta_0 \mid x) > P^{\pi}(\theta \in \Theta_0^c \mid x) \\ 0 & \text{o.w.} \end{cases}$

Bayes Factor:

$$B_{01}^{\pi} = \frac{\text{Posterior Odds}}{\text{Prior Odds}} = \frac{\dfrac{P(\theta \in \Theta_0 \mid x)}{P(\theta \in \Theta_1 \mid x)}}{\dfrac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)}}$$

$\log_{10}(B_{01}^{\pi}) \in \begin{cases} (0, 0.5) & \text{okay.} \\ (0.5, 1) & \text{Substantial to reject } H_0. \\ (1, 2) & \text{strong} \\ >2 & \text{Decisive.} \end{cases}$