Feynman Liang
STAT 210A
Lecture #10?? TODO: Check and type up previous
9/22/2016

1 Conjugate priors

Definition 1.1. If the posterior $P(\theta|X)$ is in the same family as prior $P(\theta)$, we say the prior is *conjugate* to the likelihood $P(X|\theta)$.

Example 1.2 (Beta/Binomial)**.** $X \sim \text{Binom}(n, \theta) = \theta^x (1-\theta)^{n-x} \binom{n}{x}$
$\Theta \sim \text{Beta}(\alpha, \beta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}\frac{1}{Z}$

Example 1.3 (Exp. Fam.)**.** For a general s-dimensional exponential family: $X_1, \cdots, X_n \overset{\text{iid}}{\sim} p_\eta(x) = e^{\eta' T(x) - A(\eta)} p_0(x)$
$\lambda_{k,\mu}(\eta) = e^{k\mu'\eta - kA(\eta) - B(k,\mu)} h(\eta)$
Sufficient statistics $\begin{bmatrix} \eta \\ -A(\eta) \end{bmatrix} \in \mathbb{R}^{s+1}$
Natural parameters $\begin{bmatrix} k\mu \\ k \end{bmatrix} \in \mathbb{R}^{s+1}$

$$\lambda(\eta|X_1, \cdots, X_n) \propto_\mu e^{k\mu'\eta - kkA(\eta)} h(\eta) \prod_{i=1}^n e^{\eta' T(x_i) - A(\eta)} \tag{1.1}$$
$$\propto_\mu e^{(k\mu + n\bar{T})'\eta - (k+n)A(\eta)} h(\eta) \tag{1.2}$$
$$= e^{(k\mu + n\bar{T})'\eta - (k+n)A(\eta) - B(k+n, \mu\frac{k}{k+n} + \bar{T}\frac{n}{k+n})} h(\eta) \tag{1.3}$$

where $\bar{T} = \frac{1}{n}\sum_{i=1}^n T(x_i)$.

Pseudo-points interpretation: If we call $\tilde{k} = k+n, \tilde{\mu} = \mu\frac{k}{k+n} + \bar{T}\frac{n}{k+n}$, then starting with $\lambda_{0,0}$ and observing $\tilde{k}$ data points with mean $\tilde{\mu}$ would get us to the same place as observing n points with this prior.

Example 1.4 (Gamma/Poisson)**.** $X_1, \cdots, X_n \sim \text{Pois}(\theta) = \frac{\theta^x e^{-\theta}}{x!}, x \in \mathbb{N}$
$\Theta \sim \text{Gamma}(\nu, \sigma) = \frac{1}{\Gamma(\nu)\sigma^\nu}\theta^{\nu-1}e^{-\theta/\sigma}$, we call ν the *shape parameter* and σ the *scale parameter*.

2 Where does prior come from?

1) Prior knowledge, past experiments. **Eg**: A/B testing, **Eg**: ZIP code classification
2) Subjective beliefs

Issues

- Science demands *objectivity*
- Philosophical conundrum
- Possible to write beliefs exactly?
- Might be overconfident

3) Convenience prior, **Eg**: conjugate priors

4) "Objective" (uninformative) prior, **Eg**: $X \sim \mathcal{N}(\theta, 1)$, $\theta \sim$ flat prior $\lambda(\theta) \propto 1$ ("improper") so $\Theta|X \sim \mathcal{N}(X, 1)$

Definition 2.1. The *Jeffrey's Prior* takes

$$\lambda(\theta) \propto \sqrt{|J(\theta)|} \tag{2.1}$$

where $J(\theta)$ is the Fisher information. The Jeffrey's prior is invariant to reparameterization of θ (Jacobian works to exactly cancel the way the information changes).

Example 2.2 (Jeffrey's prior for binomial)**.** $\text{Beta}(1/2, 1/2) = \left(\sqrt{\theta(1-\theta)}\right)^{-1}$

3 Bayesian Pros and Cons