JASON PACHECO
2/10/11

CS195F - RECITATION WK02

THE EXPONENTIAL DISTRIBUTION

The exponential density is,

$$\text{Exp}(x;\lambda) = \lambda e^{-\lambda x}$$

with rate parameter $\lambda > 0$. We have observations $X = \{x_i\}_{i=1}^n$ iid exponential. The likelihood is,

$$\begin{aligned} L(\lambda; x) = p(x|\lambda) &= \prod_{i=1}^n \lambda \exp(-\lambda x_i) \\ &= \lambda^n \exp\{-\lambda \textstyle\sum_{i=1}^n x_i\} \\ &= \lambda^n \exp\{-\lambda n \bar{x}\} \end{aligned}$$

with sufficient statistic $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$

The conjugate ~~prior~~ for rate parameter λ is the Gamma distribution,

$$\text{Gamma}(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda\beta)$$

$\underline{\text{Def'n}}$: (Conjugate Prior)

A prior $p(\theta) \in \mathcal{F}$ is <u>conjugate</u> to a likelihood $p(D|\theta)$ if the posterior $p(\theta|D)$ is also in $\mathcal{F}$.

$\underline{\text{e.g.}}$ (Gamma-Exponential)

$\lambda \sim \text{Gamma}(\lambda; \alpha, \beta)$

$X|\lambda \sim \text{Exp}(X; \lambda)$, for $X = \{X_i\}_{i=1}^n$ iid

The posterior is,

$$p(\lambda | \bar{x}) \propto \text{Gamma}(\lambda; \alpha, \beta) L(\lambda; \bar{x})$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\lambda\beta) \lambda^n \exp(-\lambda n \bar{x})$$

$$\propto \lambda^{\alpha+n-1} \exp(-\lambda(\beta + n\bar{x}))$$

$$\propto \text{Gamma}(\lambda; n+\alpha, n\bar{x}+\beta)$$

The Normal Distribution

The normal density is,

$$N(x; \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(-\frac{1}{2}(x-\mu)^2/\sigma^2\right)$$

With mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^+$ or standard deviation $\sigma \in \mathbb{R}^+$. Another parameterization is the inverse variance or precision $\tau \in \mathbb{R}^+$

$$N(x; \mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left(-\frac{1}{2}(x-\mu)^2 \tau\right)$$

The likelihood for n i.i.d. observations $X = \{x_i\}_{i=1}^n$ of the precision parameter is,

$$L(\tau; X, \mu) = \prod_{i=1}^n N(x_i; \mu, \tau)$$

$$\propto \prod_{i=1}^n \tau^{+1/2} \exp\left(-\frac{1}{2}(x_i-\mu)^2 \tau\right)$$

$$= \tau^{+n/2} \exp\left(-\frac{1}{2}\tau \sum_{i=1}^n (x_i-\mu)^2\right)$$

The sufficient statistic is,

$$\nu = \frac{1}{n}\sum_{i=1}^n (x_i-\mu)^2$$

So the likelihood is,

$$L(\tau; X, \mu) \propto \tau^{+n/2} \exp\left(-\frac{1}{2}\tau n \nu\right)$$

The conjugate prior for precision τ is,

$\text{Gamma}(\tau; \alpha, \beta)$

e.g. (Normal-Gamma w/ known mean, unknown precision)

$$p(\tau | X) \propto L(\tau; X, \mu)\, \text{Gamma}(\tau; \alpha, \beta)$$

$$\propto \tau^{+n/2} \exp\left(-\tfrac{1}{2} \tau n \nu\right) \tau^{\alpha - 1} \exp(-\tau\beta)$$

$$= \tau^{\alpha + n/2 - 1} \exp\left(-\tau\left(\beta + \tfrac{1}{2} n \nu\right)\right)$$

$$\propto \text{Gamma}\left(\tau; \alpha + \frac{n}{2}, \beta + \frac{n\nu}{2}\right)$$

Bayesian Inference for Gauss-Gamma Model

Suppose now that we would like to predict a new observation $\tilde{X}$. We have 3 Bayesian options:

1) MAP Plug-in

$$\hat{\theta}_{MAP} = \arg\max_{\theta} P(\theta | x_1^n)$$

$$\cancel{P(\tilde{X} | \hat{\theta}_{MAP})}$$

$$\tilde{X} = \arg\max_{X} P(X | \hat{\theta}_{MAP})$$

2) Posterior Mean Plug-in

$\cancel{\bar{\theta} \neq \mathbb{E}}$

$$\bar{\theta} = \mathbb{E}[\theta \mid x_1^n]$$

$$\tilde{X} = \arg\max_x P(x \mid \bar{\theta})$$

3) Posterior Predictive Dist. (Real Bayesian Way)

We integrate out <u>all possible</u> parameters θ

$$P(\tilde{X} \mid x_1^n) = \int P(\theta) P(x_1^n \mid \theta)\, d\theta$$

Under the Normal-Gamma model this is,

$$P(\tilde{X} \mid x_1^n) = \int_0^\infty L(\tau; x_1^n)\, \mathrm{Gamma}(\tau; \alpha, \beta)\, d\tau$$

$$= \mathcal{T}(\tilde{X} \mid \mu, 1, 2\alpha)$$

Which is the Student-T distribution,

$$\mathcal{T}(x \mid \mu, \sigma^2, \nu) \propto \left[1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\left(\frac{\nu+1}{2}\right)}$$