$$P(\theta|\underset{\sim}{y}) = \frac{P(\underset{\sim}{y}|\theta)\,\pi(\theta)}{P(\underset{\sim}{y})} \propto P(\underset{\sim}{y}|\theta)\,\pi(\theta)$$

Single parameter inference

STA 427/527, Fall 2019, Xin Wang

2.4 Choices of priors

2.4.1 Conjugate priors

- **Conjugacy of priors:** If $\mathcal{F}$ is a class of sampling distributions $p(\boldsymbol{y}|\theta)$, and $\mathcal{P}$ is a class of prior distributions $\pi(\theta)$, then the class $\mathcal{P}$ is conjugate for $\mathcal{F}$ if $p(\theta|\boldsymbol{y}) \in \mathcal{P}$.

  prior distributions and posterior distributions are in the same class of distributions.

- **Example 1:** Beta is conjugate for Binomial

  - $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \text{Bin}(m, \theta)$, where m is known, $\theta \sim \text{Beta}(\alpha, \beta)$ $\quad \underset{\sim}{y} = (y_1, y_2, \cdots y_n)$
  - Posterior: $\theta|\boldsymbol{y} \sim \text{Beta}\left(\alpha + \sum_{i=1}^n y_i, \beta + mn - \sum_{i=1}^n y_i\right)$

Data distribution: $P(\underset{\sim}{y}|\theta) = P(y_1, y_2, \cdots y_n|\theta) = \prod_{i=1}^n P(y_i|\theta) = \prod_{i=1}^n \binom{m}{y_i} \theta^{y_i} (1-\theta)^{m-y_i}$

$$= \left(\underbrace{\prod_{i=1}^n \binom{m}{y_i}}\right) \cdot \theta^{\sum_{i=1}^n y_i} (1-\theta)^{nm - \sum_{i=1}^n y_i}$$

prior: $\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1} \rightarrow$ pdf of Beta(α, β)

posterior: $P(\theta|\underset{\sim}{y}) \propto P(\underset{\sim}{y}|\theta) \cdot \pi(\theta)$

$$\propto \theta^{\sum_{i=1}^n y_i} (1-\theta)^{nm - \sum_{i=1}^n y_i} \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$= \theta^{\boxed{\sum_{i=1}^n y_i + \alpha} - 1} (1-\theta)^{\boxed{nm - \sum_{i=1}^n y_i + \beta} - 1}$$

$$\sim \boxed{\text{Beta}\left(\sum_{i=1}^n y_i + \alpha,\ nm - \sum_{i=1}^n y_i + \beta\right)} \rightarrow E(\theta|\underset{\sim}{y}) = \frac{\sum_{i=1}^n y_i + \alpha}{nm + \beta + \alpha}$$

- **Example 2:** Gamma is conjugate for Poisson

  - $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} \text{Poisson}(\lambda)$, prior of λ is $\lambda \sim \text{Gamma}(\alpha, \beta)$

Data distribution: $P(\underset{\sim}{y}|\lambda) = P(y_1, y_2, \cdots y_n|\lambda) = \prod_{i=1}^n P(y_i|\lambda) = \prod_{i=1}^n \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}$

$$= \left(\prod_{i=1}^n \frac{1}{y_i!}\right) \cdot \lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}$$

prior: $\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$

posterior: $P(\lambda|\underset{\sim}{y}) \propto P(\underset{\sim}{y}|\lambda)\,\pi(\lambda) \propto \lambda^{\sum_{i=1}^n y_i} e^{-n\lambda} \cdot \lambda^{\alpha-1} e^{-\beta\lambda}$

$$= \lambda^{\boxed{\sum_{i=1}^n y_i + \alpha} - 1} e^{-\boxed{(n+\beta)}\lambda}$$

$$E(\lambda|\underset{\sim}{y}) = \frac{\sum_{i=1}^n y_i + \alpha}{n+\beta} \qquad \sim \text{Gamma}\left(\sum_{i=1}^n y_i + \alpha,\ n+\beta\right)$$

For. New observation

Poisson example.

– Question: What is prior predictive distribution and posterior predictive distribution?

① prior predictive $P(\tilde{y}) = \int P(\tilde{y}, \lambda)\, d\lambda = \int P(\tilde{y}|\lambda)\, \pi(\lambda)\, d\lambda$

$$= \int \frac{\lambda^{\tilde{y}} e^{-\lambda}}{\tilde{y}!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\, d\lambda$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\tilde{y}!} \int \lambda^{\boxed{\tilde{y}+\alpha}-1} e^{-\boxed{(\beta+1)}\lambda}\, d\lambda$$

$\tilde{y} = 0, 1, 2, \cdots$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\tilde{y}!} \frac{\Gamma(\tilde{y}+\alpha)}{(\beta+1)^{\tilde{y}+\alpha}}$$

If α is an integer >0

$$= \frac{\beta^\alpha}{(\alpha-1)!} \frac{1}{\tilde{y}!} \frac{(\tilde{y}+\alpha-1)!}{(\beta+1)^{\tilde{y}+\alpha}}$$

$$= \frac{(\tilde{y}+\alpha-1)!}{(\alpha-1)!\, \tilde{y}!} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^{\tilde{y}} \rightarrow \text{Negative binomial}$$

Negative binomial: $\left(\alpha, \frac{\beta}{\beta+1}\right)$ $\tilde{y}$: # of failures until α success.

Gamma (α, β)

$$\int \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\, dx = 1$$

$$\Rightarrow \int x^{\alpha-1} e^{-\beta x}\, dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$$

② posterior predictive:

$$p(\tilde{y}|\underline{y}) = \int p(\tilde{y}|\lambda)\, p(\lambda|\underline{y})\, d\lambda = \int \frac{\lambda^{\tilde{y}} e^{-\lambda}}{\tilde{y}!} \boxed{\frac{(n+\beta)^{\sum_{i=1}^n y_i+\alpha}}{\Gamma(\sum_{i=1}^n y_i+\alpha)}} \lambda^{\sum_{i=1}^n y_i+\alpha-1} e^{-(n+\beta)\lambda}\, d\lambda$$

$$= \frac{(n+\beta)^{\sum_{i=1}^n y_i+\alpha}}{\Gamma(\sum_{i=1}^n y_i+\alpha)} \cdot \frac{1}{\tilde{y}!} \int \lambda^{\boxed{\sum_{i=1}^n y_i+\alpha+\tilde{y}}-1} e^{-\boxed{(n+\beta+1)}\lambda}\, d\lambda$$

$$= \frac{1}{\tilde{y}!} \frac{(n+\beta)^{\sum y_i+\alpha}}{\Gamma(\sum y_i+\alpha)} \frac{\Gamma(\sum y_i+\alpha+\tilde{y})}{(n+\beta+1)^{\sum y_i+\alpha+\tilde{y}}} \qquad \tilde{y} = 0, 1, 2 \cdots$$

$\alpha > 0$ integer

$$= \frac{1}{\tilde{y}!} \cdot \frac{(\sum y_i+\alpha+\tilde{y}-1)!}{(\sum y_i+\alpha-1)!} \left(\frac{n+\beta}{n+\beta+1}\right)^{\sum y_i+\alpha} \left(\frac{1}{n+\beta+1}\right)^{\tilde{y}}$$

$\sim$ Negative binomial $\left(\sum y_i+\alpha, \frac{n+\beta}{n+\beta+1}\right)$

$$E(\tilde{y}|\underline{y}) = \frac{\sum y_i+\alpha}{n+\beta}$$

In simulation : (2 steps).

Step 1: Sample $\lambda^{(1)}, \lambda^{(2)}, \cdots, \lambda^{(M)}$ from posterior distribution.
Gamma $(\sum y_i+\alpha,\ n+\beta)$

Step 2: Sample $\tilde{y}^{(m)}$, $m = 1, 2, \cdots M$ from Poisson $(\lambda^{(m)})$

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

$$E(\lambda) = \frac{\alpha}{\beta}$$

$$V(\lambda) = \frac{\alpha}{\beta^2}$$

$$E(\theta \mid \underset{\sim}{y}) = \frac{\Sigma y_i + \alpha}{nm + \alpha + \beta}$$ Binomial – beta

$$= \frac{\Sigma y_i / nm \cdot nm + \frac{\alpha}{\alpha+\beta} \cdot (\alpha+\beta)}{nm + \alpha + \beta}$$

sample size

$$= \boxed{\frac{\Sigma y_i}{nm}} \cdot \frac{nm}{nm+\alpha+\beta} + \boxed{\frac{\alpha}{\alpha+\beta}} \cdot \frac{\alpha+\beta}{nm+\alpha+\beta}$$

$$E(\lambda \mid \underset{\sim}{y}) = \frac{\Sigma y_i + \alpha}{n + \beta}$$ Poisson – Gamma (α, β) $\downarrow$ $\frac{\alpha}{\beta}$

$$= \frac{\Sigma y_i / n \cdot n + \frac{\alpha}{\beta} \cdot \beta}{n + \beta}$$

$$= \boxed{\frac{\Sigma y_i}{n}} \cdot \frac{n}{n+\beta} + \boxed{\frac{\alpha}{\beta}} \cdot \frac{\beta}{n+\beta}$$

- **Example 3: Conjugate priors for exponential family**
  A signle-parameter exponential family: The family consists of any distribution whose pmf/pdf can be written as

  $$f(y|\theta) = \exp\{a(y)\,b(\theta) + c(\theta) + d(y)\}$$

  $a(\cdot),\ b(\cdot),\ c(\cdot),\ d(\cdot)$ are known functions.

  where $a(y)$ and $d(y)$ do not depend on θ, $b(\theta)$ and $c(\theta)$ do not depend on y.

  – If we have n iid observations from the distribution above. The joint distribution of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ is

  Data distribution: $$P(\underset{\sim}{y}|\theta) = \prod_{i=1}^{n} \exp\{a(y_i)\,b(\theta) + c(\theta) + d(y_i)\}$$
  $$= \exp\left\{b(\theta)\sum_{i=1}^{n} a(y_i) + n\,c(\theta) + \sum_{i=1}^{n} d(y_i)\right\}$$

  – A conjugate prior for θ is

  $$\pi(\theta) = g(k, \gamma)\exp\left[k \cdot \gamma \cdot b(\theta) + kc(\theta)\right]$$

  – Posterior distribution

  $$p(\theta|\underset{\sim}{y}) \propto P(\underset{\sim}{y}|\theta)\cdot \pi(\theta)$$
  $$= \exp\left\{b(\theta)\sum_{i=1}^{n} a(y_i) + n\,c(\theta) + \sum_{i=1}^{n} d(y_i)\right\}$$
  $$\cdot\ g(k,\gamma)\exp\left[k\cdot\gamma\, b(\theta) + k\cdot c(\theta)\right]$$
  $$\propto \exp\left\{b(\theta)\left(\sum_{i=1}^{n} a(y_i) + k\gamma\right) + (n+k)\,c(\theta)\right\}$$

- **Comments**
  – Easy to implement
  – Not flexible enough to allow other shapes of priors
  – Still possible to incorporate prior knowledge about θ

- **Some conjugate priors**

Data likelihood	Prior
Bernoulli	Beta
Binomial	Beta
Poisson	Gamma
Normal(σ^2 known)	Normal for μ
Normal(μ known)	Inverse-gamma for σ^2
Uniform$(0, \theta)$	Pareto for θ
Exponential	Gamma
Gamma(β is unknown, α is known)	Gamma

2.5 Noninformative priors

$$p = \frac{\alpha}{\alpha+1} \rightarrow \frac{dp}{d\alpha} = \frac{1}{(\alpha+1)^2}$$

1. Uniform prior

   - **Example:** If we toss a coin, $Y \sim Bernoulli\,(p)$;$p \sim Uniform\,(0,1)$

If considering transformation. $\alpha = \frac{p}{1-p}$, $\alpha > 0$ $\pi(p) = 1$ $0 < p < 1$

? $\pi(\alpha)$ Need to calculate $|J| = \left|\frac{dp}{d\alpha}\right| = \frac{1}{(\alpha+1)^2}$

$$\pi(\alpha) = 1 \cdot \frac{1}{(\alpha+1)^2} = \frac{1}{(\alpha+1)^2}, \quad \alpha > 0$$

↑ prior for α

2. Jeffrey's prior: Fisher information

likelihood function. $\ln L(y|\theta)$ ($\ln L(\theta|y)$)

log likelihood.

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

Then the corresponding Jeffrey's prior is $p(\theta) \propto \sqrt{I(\theta)}$

Consider $\varphi = h(\theta)$, you can derive the prior from $p(\theta)$,

denote as $\pi_1(\varphi)$

It turns out $\pi_1(\varphi) \propto \sqrt{I(\varphi)}$

"invariant" to the parameterization.

   - **Example:** Binomial data $Y|\theta \sim Binomial\,(n,\theta)$ $E(Y) = n\theta$

$$p(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

$$\ln p(y|\theta) = \ln\binom{n}{y} \; y\ln\theta + (n-y)\ln(1-\theta)$$

$$\frac{\partial \ln p(y|\theta)}{\partial\theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta}$$

$$\frac{\partial^2 \ln p(y|\theta)}{\partial\theta^2} = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}$$

$$I(\theta) = -E\left(\frac{\partial^2 \ln p(y|\theta)}{\partial\theta^2}\right) = -E\left[-\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}\right]$$

$$= \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n}{1-\theta} = n\cdot\frac{1}{\theta(1-\theta)}$$

The Jeffreys prior is

$$\pi(\theta) \propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\theta(1-\theta)}} \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$$

$$\sim Beta\left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$