Bayesian Midterm Revision

In Son Zeng

University of Michigan

insonz@umich.edu

November 1, 2018

Overview

Deriving Conditional Distribution

Conditional Distribution Let us define a joint distribution $f_{X,Y}(x,y)$ on $a \leq x \leq b$ and $c \leq y \leq d$. Also, let $f_X(x) > 0, f_Y(y) > 0$, then

1. $$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{\int_a^b f_{X,Y}(x,y)dx}, & \text{if } a \leq x \leq b,\ c \leq y \leq d \\ 0, & \text{otherwise} \end{cases}$$
2. $$f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int_c^d f_{X,Y}(x,y)dy}, & \text{if } a \leq x \leq b,\ c \leq y \leq d \\ 0, & \text{otherwise} \end{cases}$$

Independence

Block 1

If random variables $X_1, X_2, ..., X_n$ are independent, then

- If $X_1, X_2, ..., X_n$ are jointly discrete,
  $p(x_1, x_2, ..., x_n) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot ...... \cdot p_{X_n}(x_n)$
- If $X_1, X_2, ..., X_n$ are jointly continuous,
  $f(x_1, x_2, ..., x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot ...... \cdot f_{X_n}(x_n)$

Block 2

- If X, Y are independent and jointly discrete, and given respectively the marginal of x $p_X(x) > 0$ and marginal of y $p_Y(y) > 0$, then
  $p(x, y) = p_X(x) \cdot p_Y(y) = p_{Y|X}(y|x) \cdot p_X(x) = p_{X|Y}(x|y) \cdot p_Y(y)$
- If X, Y are independent and jointly continuous, and given respectively the marginal of x $f_X(x) > 0$ and marginal of y $f_Y(y) > 0$, then
  $f(x, y) = f_X(x) \cdot f_Y(y) = f_{Y|X}(y|x) \cdot f_X(x) = f_{X|Y}(x|y) \cdot f_Y(y)$

Basics:

- Prior Predictive: $\pi(\tilde{y}) = \int \pi(\tilde{y}|\theta)\pi(\theta)d\theta$
- Posterior Predictive: $\pi(\tilde{y}|y) = \int \pi(\tilde{y}|\theta)\pi(\theta|y)d\theta$
- Bayesian Point Estimator: $\theta = E(\theta|y) = \int \theta\pi(\theta|y)d\theta$
- Posterior Interval $1-\alpha$:
  $P(\theta \in C|y) = \int_C \pi(\theta|y)d\theta = 1-\alpha \rightarrow C = (l, u)$
- Posterior MSE: $MSE = bias^2 + Var(\pi(\theta|y))$

Beta, Binomial Model

- Likelihood and Prior: $\pi(y|\theta) \sim Bin(n, \theta), \theta \sim Beta(\alpha, \beta)$
- Posterior:
  $\theta|Y = y \sim Beta(y + \alpha, n - y + \beta) \rightarrow Var(\theta|y) = \frac{(y+\alpha)(n-y+\beta)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)}$
- Prior predictive: $\pi(\tilde{Y} = \tilde{y}) = \int_0^1 \pi(\tilde{Y} = \tilde{y}|\theta)\pi(\theta|\alpha, \beta)d\theta =$
  $\frac{\Gamma(\tilde{n}+1)}{\Gamma(\tilde{y}+1)\Gamma(\tilde{n}-\tilde{y}+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\tilde{y}+\alpha)\Gamma(\tilde{n}-\tilde{y}+\beta)}{\Gamma(\alpha+\beta+\tilde{n})}$
- Posterior predictive:
  $\pi(\tilde{Y} = \tilde{y}|Y = y) = \int_0^1 \pi(\tilde{Y} = \tilde{y}|\theta)\pi(\theta|y, \alpha, \beta)d\theta =$
  $\frac{\Gamma(\tilde{n}+1)}{\Gamma(\tilde{y}+1)\Gamma(\tilde{n}-\tilde{y}+1)} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(\beta+n-y)} \frac{\Gamma(\alpha+y+\tilde{y})\Gamma(\beta+n-y+\tilde{n}-\tilde{y})}{\Gamma(\alpha+\beta+n+\tilde{n})}$

Normal Mean with natural conjugate prior

If $y|\mu \sim N(\mu, \sigma^2)$, assume that prior: $\mu \sim N(\xi, \tau_0^2)$

1. then posterior:

$$\mu|y \sim N\Big(\frac{\frac{n\bar{y}}{\sigma^2} + \frac{\xi}{\tau_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}\Big) = N(\mu_1, \tau_1^2)$$

2. Bayesian Estimate: $E(\mu|y) = \frac{\frac{n\bar{y}}{\sigma^2} + \frac{\xi}{\tau_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}$ and $Var(\mu|y) = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}$

3. Expression:

$$\pi(\mu|y) \propto \pi(\mu) \cdot \pi(y|\mu) \propto exp\Bigg(-\frac{1}{2}\Big(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\Big)\mu^2 + \Big(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\Big)\mu\Bigg)$$

$$= exp\Big(\tilde{a}\mu^2 + \tilde{b}\mu\Big) \to \tilde{a} = -\frac{1}{2}\Big(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\Big), \tilde{b} = \Big(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\Big)$$

Normal Mean with natural conjugate prior

To continue, $\tilde{\tau}^2 = -\frac{1}{2\tilde{a}} = \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}}$ and $\tilde{\xi} = \tilde{b}\cdot\tilde{\tau}^2 = \frac{\frac{n\bar{y}}{\sigma^2}+\frac{\xi}{\tau_0^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}}$

1. Prior predictive distribution: since $E(\tilde{y}) = E\Big[E(\tilde{y}|\mu)\Big] = E(\mu) = \xi$ and $Var(\tilde{y}) = E\Big[Var(\tilde{y}|\mu)\Big] + Var\Big[E(\tilde{y}|\mu)\Big] = \tau^2+\sigma^2$, we have
$$\pi(\tilde{y}) = \int_{-\infty}^{\infty} \pi(\tilde{y}|\mu)\cdot\pi(\mu)d\mu \sim N(\xi, \tau^2+\sigma^2)$$
2. Posterior predictive distribution: we have
$$\pi(\tilde{y}|y) = \int_{-\infty}^{\infty} \pi(\tilde{y}|\mu)\cdot\pi(\mu|y)d\mu \sim N(\tilde{\xi}, \tilde{\tau}^2+\sigma^2)$$

Normal Variance with natural conjugate prior

1. Likelihood: $\pi(y|\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\Big)$
2. Natural Conjugate Prior:
$\pi(\sigma^2) \propto (\sigma^2)^{-\alpha-1} \cdot exp\Big(-\frac{\beta}{\sigma^2}\Big) \sim Inv - Gamma(\alpha, \beta)$
3. Normalizing Constant for $\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}$
4. Posterior Distribution:
$\pi(\sigma^2|y) \propto \pi(y|\sigma^2) \cdot \pi(\sigma^2) \propto (\sigma^2)^{-\frac{n}{2}-\alpha-1} \cdot exp\Bigg(-\frac{\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}+\beta}{\sigma^2}\Bigg)$

$$\pi(\sigma^2|y) \sim Inv - Gamma\Big(\frac{n}{2}+\alpha, \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}+\beta\Big)$$

$$E(\sigma^2|y) = \frac{\tilde{\beta}}{\tilde{\alpha}-1} = \frac{\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}+\beta}{\frac{n}{2}+\alpha-1}, \frac{n}{2}+\alpha > 1$$

$$Var(\sigma^2|y) = \frac{\tilde{\beta}^2}{(\tilde{\alpha}-1)^2(\tilde{\alpha}-2)} = \frac{\Big(\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}+\beta\Big)^2}{(\frac{n}{2}+\alpha-1)^2(\frac{n}{2}+\alpha-2)}, \frac{n}{2}+\alpha > 2$$

Normal Precision with natural conjugate prior

1. Likelihood: $\pi(y|\sigma^{-2}) \propto (\sigma^{-2})^{\frac{n}{2}} \cdot exp\Big(-\frac{\sigma^{-2}}{2}\sum_{i=1}^{n}(y_i-\mu)^2\Big) \sim G\Big(\frac{n}{2}+1, \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\Big)$
2. Natural Conjugate Prior: $\pi(\sigma^{-2}) \propto (\sigma^{-2})^{\alpha-1} \cdot exp\Big(-\beta\sigma^{-2}\Big) \sim Gamma(\alpha,\beta)$. If $\alpha \to 0, \beta \to 0$, then $\pi(\sigma^{-2}) \propto \sigma^2$, which is improper.
3. Posterior Distribution: $\pi(\sigma^{-2}|y) \propto \pi(y|\sigma^{-2}) \cdot \pi(\sigma^{-2}) \propto (\sigma^{-2})^{\frac{n}{2}+\alpha-1} \cdot exp\Big(-\sigma^{-2} \cdot \Big(\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}+\beta\Big)\Big) \sim G\Big(\frac{n}{2}+\alpha, \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}+\beta\Big)$

$$E(\sigma^{-2}|y) = \frac{\tilde{\alpha}}{\tilde{\beta}} = \frac{\frac{n}{2}+\alpha}{\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}+\beta}$$

$$Var(\sigma^{-2}|y) = \frac{\tilde{\alpha}}{(\tilde{\beta})^2} = \frac{\frac{n}{2}+\alpha}{\Big(\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}+\beta\Big)^2}$$

Normal Precision with natural conjugate prior

1. Prior predictive: Given $\tilde{y}|\sigma^2 \sim N(\mu, \sigma^2)$,
$\pi(\tilde{y}) = \int_0^\infty \pi(\tilde{y}|\sigma^2) \cdot \pi(\sigma^2) d\sigma^2 =$
$\int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{\tilde{(}y-\mu)^2}{2\sigma^2}\right) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} exp\left(-\frac{\beta}{\sigma^2}\right) d\sigma^2 = \int_0^\infty \frac{\beta^\alpha}{\sqrt{2\pi}\Gamma(\alpha)} \cdot$
$\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{1}{2}+1} \cdot exp\left(-\frac{1}{\sigma^2}\left(\frac{(\tilde{y}-\mu)^2}{2}+\beta\right)\right) d\sigma^2 = \frac{\beta^\alpha}{\sqrt{2\pi}\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+\frac{1}{2})}{\left(\frac{(\tilde{y}-\mu)^2}{2}+\beta\right)^{\alpha+\frac{1}{2}}}$

Therefore,

$$\tilde{y} \sim Inv - Gamma(\alpha + \frac{1}{2}, \frac{(\tilde{y} - \mu)^2}{2} + \beta)$$

Normal Precision with natural conjugate prior

1. Posterior predictive:

$\pi(\tilde{y}|y) = \int_0^\infty \pi(\tilde{y}|\sigma^2)\cdot\pi(\sigma^2|y)d\sigma^2 \propto \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} exp\Big(-\frac{(\tilde{y}-\mu)^2}{2\sigma^2}\Big)\cdot$
$(\sigma^2)^{-\frac{n}{2}-\alpha-1} exp\Big(-\frac{1}{\sigma^2}\Big(\frac{\sum_{i=1}^n (y_i-\mu)^2}{2}+\beta\Big)\Big)d\sigma^2 \propto$
$\int_0^\infty \Big(\frac{1}{\sigma^2}\Big)^{\frac{n}{2}+\alpha+\frac{1}{2}+1}\cdot exp\Big(-\frac{1}{\sigma^2}\Big(\frac{(\tilde{y}-\mu)^2}{2}+\frac{\sum_{i=1}^n (y_i-\mu)^2}{2}+\beta\Big)\Big)d\sigma^2 \propto$
$\frac{\Gamma(\frac{n}{2}+\alpha+\frac{1}{2})}{\Big[\frac{\tilde{y}-\mu)^2}{2}+\frac{\sum_{i=1}^n (y_i-\mu)^2}{2}+\beta\Big]^{\frac{n}{2}+\alpha+\frac{1}{2}}} \sim G^{-1}\Big(\frac{n}{2}+\alpha+\frac{1}{2}, \frac{\tilde{(}y-\mu)^2}{2}+\frac{\sum_{i=1}^n (y_i-\mu)^2}{2}+\beta\Big)$

$$\tilde{y}|y \sim Inv - Gamma\Big(\frac{n}{2}+\alpha+\frac{1}{2}, \frac{\tilde{y}-\mu)^2}{2}+\frac{\sum_{i=1}^n (y_i-\mu)^2}{2}+\beta\Big)$$

In fact, to consider all the constants,

$$\pi(\tilde{y}|y) = \frac{\Gamma(\frac{(n}{2}+\alpha+\frac{1}{2})\cdot\Big[\frac{\sum_{i=1}^n (y_i-\mu)^2}{2}+\beta\Big]^{\frac{n}{2}+\alpha}}{\sqrt{2\pi}\Gamma(\frac{n}{2}+\alpha)\Big[\frac{\tilde{y}-\mu)^2}{2}+\frac{\sum_{i=1}^n (y_i-\mu)^2}{2}+\beta\Big]^{\frac{n}{2}+\alpha+\frac{1}{2}}}$$

Poisson model with natural conjugate prior

1. Likelihood: Suppose $y = (y_1, \ldots\ldots, y_n)$ forms
$\pi(y|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{(y_i)!} = \frac{\theta^{\sum_{i=1}^{n} y_i} e^{-n\theta}}{\prod_{i=1}^{n}(y_i)!} \propto exp\Big(-n\theta + \sum_{i=1}^{n} y_i log(\theta)\Big)$
2. Conjugate Prior: Gamma Prior
$\theta \sim Gamma(\alpha, \beta) \rightarrow \pi(\theta) \propto \theta^{\alpha-1} exp(-\beta\theta)$
3. Posterior: $\pi(\theta|y) \propto \Big(\theta^{\sum_{i=1}^{n} y_i} exp(-n\theta)\Big) \cdot \Big(\theta^{\alpha-1} exp(-\beta\theta)\Big) =$
$\theta^{\sum_{i=1}^{n} y_i + \alpha - 1} exp\Big(-(n+\beta)\theta\Big) \sim Gamma(\sum_{i=1}^{n} y_i + \alpha, n+\beta)$
4. Bayesian Estimates: $E(\theta|y) = \frac{\sum_{i=1}^{n} y_i + \alpha}{n+\beta} = \frac{\beta}{n+\beta} \cdot \frac{\alpha}{\beta} + \frac{n}{n+\beta}\bar{y}$
5. Variance Estimate: $Var(\theta|y) = \frac{\sum_{i=1}^{n} y_i + \alpha}{(n+\beta)^2}$
6. Posterior Mode: $Mode(\theta|y) = \frac{\sum_{i=1}^{n} y_i + \alpha - 1}{n+\beta}$
Here the hyperparameters β is the count of prior observations, α is the sum of counts from β prior observations.

Poisson model with natural conjugate prior

1. Limit: When $n \to \infty$, we have $E(\theta|y) \to \frac{n\bar{y}}{n} = \bar{y}$ and $Var(\theta|y) \to \frac{n\bar{y}}{n^2} = \frac{\bar{y}}{n}$
2. Prior Predictive: $\pi(\tilde{y}) = \int_0^\infty \pi(\tilde{y}|\theta)\pi(\theta)d\theta = \int_0^\infty \frac{\theta^{\tilde{y}} exp(-\theta)}{\tilde{y}!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} = \frac{\beta^\alpha}{\tilde{y}!\Gamma(\alpha)} \int_0^\infty \theta^{\tilde{y}+\alpha-1} exp(-(\beta+1)\theta)d\theta = \frac{\beta^\alpha}{\tilde{y}!\Gamma(\alpha)} \cdot \frac{\Gamma(\tilde{y}+\alpha)}{(\beta+1)^{\tilde{y}+\alpha}} = \frac{(\tilde{y}+\alpha-1)!}{\tilde{y}!(\alpha-1)!} \cdot \left(\frac{\beta}{\beta+1}\right)^\alpha \cdot \left(\frac{1}{\beta+1}\right)^{\tilde{y}} = \binom{\tilde{y}+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^\alpha \cdot \left(\frac{1}{\beta+1}\right)^{\tilde{y}}$
   Therefore, the prior predictive distribution $\pi(\tilde{y}) \sim Neg - Binomial(\alpha, \beta)$, where α denotes the number of success until the experiment stops, $\tilde{y}$ denotes the number of failures and $p = \frac{\beta}{\beta+1} \to \beta = \frac{p}{1-p}$ is the odd of success.
3. Expectation: $E(\tilde{y}) = \frac{\alpha}{\beta} = \frac{\alpha(1-p)}{p}$
4. Variance: $Var(\tilde{y}) = \frac{\alpha}{\beta}\left(1 + \frac{1}{\beta}\right) = \frac{(1-p)\alpha}{p^2}$

Poisson model with natural conjugate prior

1. Posterior Predictive: $\pi(\tilde{y}|y) = \frac{\Gamma(\alpha+n\bar{y}+\tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(\alpha+n\bar{y}} \cdot \left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+n\bar{y}} \left(\frac{1}{\beta+n+1}\right)^{\tilde{y}}$
2. Expectation: $E(\tilde{y}|y) = \frac{\alpha+n\bar{y}}{\beta+n} = \frac{(\alpha+n\bar{y})(1-p)}{p}$
3. Variance: $Var(\tilde{y}) = \frac{\alpha+n\bar{y}}{\beta+n}\left(1+\frac{1}{\beta+n}\right) = \frac{(1-p)(\alpha+n\bar{y})}{p^2}$

Exponential Family

1. Likelihood: $\pi(y|\phi) = h(y)c(\phi)exp(\phi t(y)) \propto c(\phi)exp(\phi t(y))$
2. Conjugate Prior:
   $\pi(\phi|n_0, t_0) = g(n_0, t_0)c(\phi)^{n_0}exp(n_0 t_0 \phi) \propto c(\phi)^{n_0}exp(n_0 t_0 \phi)$, where n_0 denotes the sample size (higher n_0 means more informative) and t_0 denotes the prior expectation of $t(y)$. $g(n_0, t_0)$ is independent of ϕ.
3. Posterior: Given $y_i \sim^{iid} \pi(y|\phi)$, we have
   $\pi(\phi|y) \propto \pi(y|\phi) \cdot \pi(\phi|n_0, t_0) \propto c(\phi)^{n+n_0}exp\Big(\phi(n_0 t_0 + \sum_{i=1}^n t(y_i))\Big) = \pi\Big(\phi|n_0 + n, \frac{n_0 t_0 + \sum_{i=1}^n t(y_i)}{n_0+n}\Big) = \pi\Big(\phi|n_0 + n, \frac{n_0 t_0 + n\bar{t}(y)}{n_0+n}\Big)$
   , where $\bar{t}(y) = \frac{1}{n}\sum_{i=1}^n t(y_i)$
4. Binomial, Poisson, Normal, Galenshore, Gamma (Chi-square, Exponential), Beta, Dirichlet, Wishart, Inv-Wishart and Geometric distibutions are examples in exponential family.
5. Uniform, Student's t and most mixture distributions are not in exponential family.

Non-informative Priors

1. Likelihood:
2. Binomial, Poisson, Normal, Galenshore, Gamma (Chi-square, Exponential), Beta, Dirichlet, Wishart, Inv-Wishart and Geometric distibutions are examples in exponential family.
3. Uniform, Student's t and most mixture distributions are not in exponential family.

Jeffrey Priors

1. Objective: Determine $\pi(\theta)$ using $\pi(y|\theta)$ and determine $\pi(\phi)$ using $\pi(y|\phi)$.
2. Transformation: Given $f_X(x)$, let $Y = g(x) \rightarrow x = g^{-1}(y)$, then $f_Y(y) = f_X(g^{-1}(y)) \cdot |\frac{dg^{-1}(y)}{dy}|$
3. Example: $f(x) = 3x^2, 0 < x < 1$ and $Y = X^2 = g(x) \rightarrow x = g^{-1}(y) = \sqrt{y}, 0 < y < 1$, we have $f_Y(y) = f_X(\sqrt{y})|\frac{d\sqrt{y}}{dy}| = 3y\frac{y^{-\frac{1}{2}}}{2} = \frac{3}{2}y^{\frac{1}{2}}, 0 < y < 1.$
4. Jeffrey's general principle: Given the prior density $\pi(\theta)$, and $\phi = h(\theta)$, then the prior on ϕ is: $\pi_\phi(\phi) = \pi_\theta(h^{-1}(\phi))|\frac{dh^{-1}(\phi)}{d\phi}|$. Any rule for determining the prior density $\pi(\theta)$ should obtain the same result if it is applied to any one-to-one transformed parameter $\phi = h(\theta)$

Jeffrey Priors

1. Jeffrey's Prior:

$$\pi(\theta) \propto [I(\theta)]^{\frac{1}{2}} = \left\{ E\left[\left(\frac{d \log \pi(y|\theta)}{d\theta} \right)^2 |\theta \right] \right\}^{\frac{1}{2}} = \left\{ -E\left[\frac{d^2 \log \pi(y|\theta)}{d\theta} |\theta \right] \right\}^{\frac{1}{2}}$$

2. $I(\phi) = E\left[\left\{ \frac{d \log \pi(y|\theta)}{d\theta} \right\}^2 \middle| \theta \right] \left(\frac{d\theta}{d\phi} \right)^2$

$$\rightarrow \pi(\phi) \propto E\left[\left\{ \frac{d \log \pi(y|\theta)}{d\theta} \right\}^2 \middle| \theta \right]^{\frac{1}{2}} \left(\frac{d\theta}{d\phi} \right)$$

Jeffrey Priors for Normal Model:

1. When σ^2 is known, the Jeffrey's prior for μ is

$$y \sim N(\mu, \sigma^2) \rightarrow log(y|\mu) \propto -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 \rightarrow \frac{d^2 log(y|\mu)}{d\mu^2} = -\frac{n}{\sigma^2}$$

$$\pi(\mu) \propto \sqrt{I(\mu)} = \sqrt{-E\left[\frac{-n}{\sigma^2}\right]} \propto 1$$

2. When μ is known, the Jeffrey's prior for σ^2:
$log(y|\sigma^2) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 \rightarrow$
$\frac{d^2 log(y|\sigma^2)}{d(\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(y_i - \mu)^2$
$\pi(\sigma^2) \propto \sqrt{I(\sigma^2)} \propto \sqrt{-E\left[\frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(y_i - \mu)^2\right]} =$
$\sqrt{\frac{n\sigma^2}{(\sigma^2)^3} - \frac{n}{2(\sigma^2)^2}} = \sqrt{\frac{n}{2(\sigma^2)^2}} \propto \frac{1}{\sigma^2}$

Jeffrey Priors for Normal Model:

1. Jeffrey's prior for μ^3: Take $\phi = \mu^3 \rightarrow \mu = \phi^{\frac{1}{3}}, \frac{d\mu}{d\phi} = \frac{1}{3}\phi^{-\frac{2}{3}}$
$\pi(\phi) = \pi(\mu) \cdot |\frac{1}{3}\phi^{-\frac{2}{3}}| \propto 1 \cdot \phi^{-\frac{2}{3}} = \phi^{-\frac{2}{3}} = \mu^{-2}$
2. Jeffrey's prior for $log(\sigma^2)$: Take $\phi = log(\sigma^2) \rightarrow \sigma^2 = e^{\phi} \rightarrow \frac{d\sigma^2}{d\phi} = e^{\phi}$
$\pi(\phi) = \pi(\sigma^2) \cdot |e^{\phi}| \propto \frac{1}{\sigma^2} \cdot \sigma^2 = 1$
3. Jeffrey's prior for μ, σ^{-2} unknown:
$\pi(y|\mu, \sigma^{-2}) \propto (\sigma^{-2})^{\frac{n}{2}} exp\Big(- \frac{\sigma^{-2}}{2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\Big)$
$log\Big(\pi(y|\mu, \sigma^{-2})\Big) \propto \frac{n}{2} log(\sigma^{-2}) - \frac{\sigma^{-2}}{2}(n-1)s^2 - \frac{\sigma^{-2}}{2} n(\bar{y} - \mu)^2$
Taking first and second derivatives, we get: $I(\mu, \sigma^{-2}) =$
$\pi(\mu, \sigma^{-2}) \propto \sqrt{I(\mu, \sigma^{-2})} \propto \frac{1}{\sigma}$

Jeffrey Priors for Binomial Model:

1. Jeffrey's prior for $y \sim Bin(n,\theta) \rightarrow \pi(y|\theta) \propto \theta^y(1-\theta)^{n-y}$
$log(\pi(y|\theta) \propto yln\theta + (n-y)ln(1-\theta) \rightarrow \frac{d^2 log(\pi(y|\theta)}{d\theta^2} = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2} =$
$\frac{-y(1-\theta)^2-(n-y)\theta^2}{\theta^2(1-\theta)^2} = \frac{-y+2\theta y-n\theta^2}{\theta^2(1-\theta)^2} \quad \Big(E(y) = n\theta\Big)$
$I(\theta) = -E\left[\frac{-y+2\theta y-n\theta^2}{\theta^2(1-\theta)^2}\right] = \frac{n}{\theta(1-\theta)}$
$\rightarrow \pi(\theta) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \sim Beta(\frac{1}{2},\frac{1}{2})$
2. Jeffrey's prior for $logit(\theta)$: Take
$\phi = logit(\theta) \rightarrow \theta = \frac{exp(\phi)}{1+exp(\phi)}, \frac{d\theta}{d\phi} = \frac{exp(\phi)}{(1+exp(\phi))^2}$
$\pi(\phi) \propto \pi(\theta) \cdot |\frac{exp(\phi)}{(1+exp(\phi))^2}| \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \cdot |\theta(1-\theta)| = \theta^{\frac{1}{2}}(1-\theta)^{\frac{1}{2}}$

Marginal Posterior Probability

1. Continuous Case: Let $\theta = (\theta_1, \theta_2)$ then $\pi(\theta_1, y) = \int \pi(y|\theta_1, \theta_2)\pi(\theta_1, \theta_2)d\theta_2 \rightarrow \pi(\theta_1|y) = \frac{\pi(\theta_1,y)}{p(y)} = \int \frac{\pi(y|\theta_1,\theta_2)\pi(\theta_1,\theta_2)}{p(y)}d\theta_2$
2. Since $\pi(\theta_1|y)$ is independent of y, $\pi(\theta_1|y) \propto \pi(\theta_1, y) = \int \pi(y|\theta_1, \theta_2)\pi(\theta_1, \theta_2)d\theta_2$
3. Discrete Example: Let $y \in \{0, 1\}, \theta_1 \in \{0, 1\}, \theta_2 \in \{0, 1\}$ If $\pi(y = 1) = 0.3\theta_1 + 0.2\theta_2$, $\pi(\theta_1 = 1) = 0.5$, $\pi(\theta_2 = 1) = 0.1$. If $\{y_1, y_2, y_3\} = \{1, 1, 1\}$, then

$$\pi(\theta_1, \theta_2|y) \propto 0.3\theta_1 + 0.2\theta_2)^3 \cdot 0.5^{\theta}0.5^{1-\theta_1}0.1^{\theta_2}0.9^{1-\theta_2}$$

Hence, $\underbrace{\pi(\theta_1|y)}_{\text{unnormalized}} \propto \sum_{\theta_2=(0,1)} \pi(\theta_1, \theta_2|y) \propto$
$(0.3\theta_1)^3 \cdot 0.5 \cdot 0.9 + (0.3\theta_1 + 0.2)^3 \cdot 0.5 \cdot 0.1$

Joint Posterior Probability for Normal

1. $p(\mu, \sigma^2|y) \propto (\sigma^{-2})(\sigma^{-n}) \cdot exp\Big(- \frac{1}{2\sigma^2} \sum_{i=1}^{n}(y_i - \mu)^2 \Big) =$
   $\sigma^{-n-2} exp\Big(- \frac{1}{2\sigma^2} [\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2] \Big) =$
   $\sigma^{-n-2} exp\Big(- \frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2] \Big)$
2. Keeping σ^2 constant, we have
   $\pi(\mu|\sigma^2, y) \propto exp\Big(- \frac{1}{2\sigma^2} n(\bar{y} - \mu)^2 \Big) \sim N(\bar{y}, \frac{\sigma^2}{n})$
3. We can integrate out μ now:
   $\pi(\sigma^2|y) = \int \sigma^{-n-2} exp\Big(- \frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2] \Big) d\mu =$
   $\sigma^{-n-2} exp\Big(- \frac{1}{2\sigma^2} (n-1)s^2 \Big) \cdot \sqrt{2\pi \frac{\sigma^2}{n}} \propto (\sigma^2)^{-\frac{n+1}{2}} exp\Big(- \frac{1}{2\sigma^2} (n-1)s^2 \Big)$

Joint Posterior Probability for Normal

1. $p(\mu, \sigma^2|y) \propto (\sigma^{-2})(\sigma^{-n}) \cdot exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\Big) =$
   $\sigma^{-n-2}exp\Big(-\frac{1}{2\sigma^2}[\sum_{i=1}^{n}(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2]\Big) =$
   $\sigma^{-n-2}exp\Big(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\Big)$
2. Keeping σ^2 constant, we have
   $\pi(\mu|\sigma^2, y) \propto exp\Big(-\frac{1}{2\sigma^2}n(\bar{y}-\mu)^2\Big) \sim N(\bar{y}, \frac{\sigma^2}{n})$
3. We can integrate out μ now:
   $\pi(\sigma^2|y) = \int \sigma^{-n-2}exp\Big(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\Big)d\mu =$
   $\sigma^{-n-2}exp\Big(-\frac{1}{2\sigma^2}(n-1)s^2\Big) \cdot \sqrt{2\pi\frac{\sigma^2}{n}} \propto (\sigma^2)^{-\frac{n+1}{2}}exp\Big(-\frac{1}{2\sigma^2}(n-1)s^2\Big)$

Bayesian Estimate Interval

1. $\mu|\sigma^2, Y \sim N\Big(E(\mu|Y), Var(\mu|Y)\Big)$
2. If σ is known: $\mu = E(\mu|Y) \pm z_{0.975} SD(\mu|Y)$
3. If σ is unknown: $\mu = E(\mu|Y) \pm t_{0.975, n+2a} SD(\mu|Y)$
4. In Bayesian, we do not have true estimator, so the randomness comes from the prior and what we observed is fixed. That is why there is no Standard Error in Bayesian framework.

Bayesian MAP vs MLE:

1. MAP estimate is the posterior mode $\hat{\theta}_{MAP} = argmax_{\theta}\pi(\theta|Y)$, while MLE maximizes the likelihood: $\hat{\theta}_{MLE} = argmax_{\theta}\pi(Y|\theta) = argmax_{\theta}\pi(\theta|y_1, ...y_n)\pi(\theta) = argmax_{\theta}\left[\sum_{i=1}^{n} log\pi(\theta|y_i) + log(\pi(\theta))\right]$
2. Binomial Example: $y|\theta \sim Bin(n,\theta), \theta \sim Beta(0.5, 0.5)$, then $\pi(\theta|y) \propto \binom{n}{y}\theta^{y}(1-\theta)^{n-y}\frac{\Gamma(1)}{\Gamma(0.5)\Gamma(0.5)}\theta^{0.5-1}(1-\theta)^{0.5-1}$. We take log and set the derivative as 0: $log[\theta^{y-0.5}(1-\theta)^{n-y-0.5}] = (y-0.5)log\theta + (n-y-0.5)log(1-\theta) \rightarrow \frac{\partial log(\pi(\theta|y))}{\partial\theta} = \frac{y-0.5}{\theta} - \frac{n-y-0.5}{1-\theta} = 0 \rightarrow \hat{\theta}_{MAP} = \frac{y-0.5}{n-1}$
3. Generally, $\theta \sim Beta(\alpha_1, \alpha_2)$, so take log we have $log[\theta^{y+\alpha_1-1}(1-\theta)^{n-y+\alpha_2-1}] = (y+\alpha_1-1)log\theta+(n-y+\alpha_2-1)log(1-\theta)\frac{\partial log(\pi(\theta|y))}{\partial\theta} = \frac{y+\alpha_1-1}{\theta} - \frac{n-y+\alpha_2-1}{1-\theta} = 0 \rightarrow \hat{\theta}_{MAP} = \frac{y+\alpha_1-1}{n+\alpha_1+\alpha_2-2}$
4. Dirichlet: $\hat{p}_{i\,MAP} = \frac{x_i+k}{n+mk}$

Bayesian MAP (Normal):

1. Normal Example: $y|\mu \sim N(\mu, \sigma^2), \mu \sim N(\mu_0, \sigma_0^2)$, then $\pi(\mu|y)$ $= \frac{1}{2\pi\sqrt{\sigma_0^2}} exp\Big(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\Big)\cdot\Big(\frac{1}{2\pi\sqrt{\sigma^2}}\Big)^n exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2\Big) =$ $\Big(\frac{1}{\sqrt{2\pi}}\Big)^{n+1}(\sigma_0^2)^{-\frac{1}{2}}(\sigma^2)^{-\frac{n}{2}} exp\Big(-(\frac{1}{2\sigma_0^2}+\frac{n}{2\sigma^2})\mu^2 + (\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2})\mu -$ $(\frac{\mu_0}{2\sigma_0^2}+\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\Big) \sim N\Bigg(\frac{\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\Bigg)$
2. Take log: $log(\pi(\mu|y)) = -(\frac{1}{2\sigma_0^2}+\frac{n}{2\sigma^2})\mu^2 + (\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2})\mu + C$
3. Take derivative as 0: $-(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2})\mu + (\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}) = 0$ $\rightarrow \hat{\mu}_{MAP} = \frac{\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}} = \frac{\mu_0\sigma^2+\sigma_0^2\sum_{i=1}^n y_i}{\sigma^2+\sigma_0^2 n}$

Bayesian CLT:

1. Case 1: For large n, suppose $X_1, X_2, ..., X_n$ are conditionally independent given $\Theta = \theta$
2. Let $c = E\left(\Theta | X_1 = x_1, X_2 = x_2, ..., X_n = x_n\right)$ be the posterior mean and $d^2 = Var\left(\Theta | X_1 = x_1, X_2 = x_2, ..., X_n = x_n\right)$, then if U is the posterior distribution, we have $\frac{U-c}{d} \sim^{CLT} N(0,1)$
3. Bernstein-von Mises Theorem: Under some regularities conditions, as $n \to \infty$, the posterior distribution of θ approaches normality with mean θ_0 and variance $\frac{1}{nI(\theta_0)}$. Now we consider the posterior mode as $\hat{\theta}_{MAP}$ and $\hat{\theta}_{MAP}$ is consistent ($\hat{\theta}_{MAP} \to \theta_0, n \to \infty$)

Bayesian CLT Proof:

1. By Taylor expansion (centered by $\hat{\theta}_{MAP}$, we have $log\pi(\theta|y) \approx log\pi(\hat{\theta}_{MAP}) + (\theta - \hat{\theta}_{MAP})\frac{d}{d\theta}log\pi(\hat{\theta}_{MAP}) + \frac{1}{2}(\theta - \hat{\theta}_{MAP})^2\frac{d^2}{d\theta^2}log\pi(\hat{\theta}_{MAP}) = C - \frac{1}{2}\frac{(\theta-\hat{\theta}_{MAP})^2}{(nI(\hat{\theta}_{MAP}))^{-1}} \sim C - \frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2$
2. It looks like normal if we take $\mu = \hat{\theta}_{MAP}, \sigma^2 = -\left(\frac{d^2}{d\theta^2}log\pi(\theta|y)\right)^{-1}$
3. Take exponential, $\pi(\theta|y) \approx exp\left(C - \frac{1}{2}\frac{(\theta-\hat{\theta}_{MAP})^2}{(nI(\hat{\theta}_{MAP}))^{-1}}\right) \sim N(\hat{\theta}_{MAP}, \left(nI(\hat{\theta}_{MAP}\right)^{-1})$
4. The (j,k) element of the information matrix I is given by: $-\frac{\partial^2}{\partial\theta_j\partial\theta_k}log(\pi(\theta|y))$, evaluated at $\hat{\theta}_{MAP}$

Gibbs Sampling for Gaussian Model

1. Rule: To do Gibbs Sampling, we start from deriving the joint posterior distribution. From that joint posterior distribution, we derive the full conditionals for each parameter FC1, FC2, and so on. The resulting full conditional distributions should have one parameter taking all other parameters as fixed and known.
2. Likelihood: Independent $Y_i|\mu, \sigma^2 \sim N(\mu, \sigma^2)$
3. Priors: $\mu \sim N(\mu_0, \sigma_0^2)$, $\sigma^2 \sim Inv - Gamma(a, b)$
4. Joint Posterior distribution: $\pi(\mu, \sigma^2|Y) \propto \pi(Y|\mu, \sigma^2)\pi(\mu)\pi(\sigma^2) \propto$ $\left(\sqrt{\frac{1}{\sigma^2}}\right)^n exp\left(-\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{\sigma_0^2} exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right)\left(\frac{1}{\sigma^2}\right)^{a+1} exp\left(-\frac{b}{\sigma^2}\right)$

Gibbs Sampling for Gaussian Model

1. FC1 (σ^2 known): $\pi(\mu|\sigma^2, Y) \propto exp\Big(-\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2} - \frac{(\mu-\mu_0)^2}{2\sigma_0^2}\Big) \propto$ $exp\Big(-(\frac{1}{2\sigma_0^2}+\frac{n}{2\sigma^2})\mu^2 + (\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2})\mu\Big) \sim N\Bigg(\frac{\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\Bigg)$
2. FC2 (μ known):
$\pi(\sigma^2|\mu, Y) \propto \Big(\frac{1}{\sigma^2}\Big)^{\frac{n}{2}} \cdot exp\Big(-\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2}\Big) \cdot \Big(\frac{1}{\sigma^2}\Big)^{a+1} \cdot exp\Big(-\frac{b}{\sigma^2}\Big) =$ $\Big(\frac{1}{\sigma^2}\Big)^{\frac{n}{2}+a+1} \cdot exp\Big(-\frac{\frac{1}{2}\sum_{i=1}^n (y_i-\mu)^2+b}{\sigma^2}\Big) \sim$ $Inv-Gamma\Big(\frac{n}{2}+a, \frac{1}{2}\sum_{i=1}^n (y_i-\mu)^2+b\Big)$
3. Full conditional distribution is proportional to the joint density

Gibbs Sampling Algorithm and Convergence

1. Gibbs Sampling is a special case of Metropolis Hasting Algorithm.
2. We first set initial value $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_p^{(0)})$. The for each iteration t, we draw FC1, FC2, ..., FCp. Namely,
3. FC1: $\pi(\theta_1^{(t)}|\theta_2^{(t-1)}, \theta_3^{(t-1)}, ..., \theta_p^{(t-1)})$
4. FC2: $\pi(\theta_2^{(t)}|\theta_2^{(t)}, \theta_3^{(t-1)}, ..., \theta_p^{(t-1)})$,
5. FC3: $\pi(\theta_3^{(t)}|\theta_1^{(t)}, \theta_2^{(t)}, \theta_4^{(t-1)} ......, \theta_p^{(t-1)})$,......,
6. FCp: $\pi(\theta_p^{(t)}|\theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_{p-1}^{(t)})$
7. We run the for loop S times to obtain posterior draws: $\theta^{(1)}, ..., \theta^{(S)}$
8. For convergence, we decide that the time T is sufficient for convergence and we discard the first S-T iterations (the first T samples as "burn-in"). Then the approximate posterior mean of θ_j is $E(\theta_j|Y) \approx \frac{1}{S-T} \sum_{s=S-T+1}^{S} \theta_j^{(s)}$

Practice for Deriving Full Conditional

1. Given $Y|\lambda, b \sim Poisson(\lambda), \lambda|b \sim Gamma(1, b), b \sim Gamma(1, 1)$
2. Probability distributions:
   $\pi(Y|\lambda, b) = \frac{e^{-\lambda}\lambda^{Y}}{Y!}, \pi(\lambda|b) = bexp(-b\lambda), \pi(b) = 1 \cdot exp(-b)$
3. Method: Estimate λ using Y and estimate b using λ.
4. Full Conditionals: $\pi(\lambda|b, Y) \propto \pi(Y|\lambda, b) \cdot \pi(\lambda|b) \propto e^{-\lambda}\lambda^{Y} \cdot exp(-b\lambda) = \lambda^{Y+1-1}exp(-(b+1)\lambda) \sim Gamma(Y+1, b+1)$
5. Full Conditionals (y is constant): $\pi(b|\lambda, Y) \propto \pi(\lambda|b) \cdot \pi(b) \propto bexp(-b\lambda) \cdot exp(-b) = b^{2-1}exp(-b(\lambda+1)) \sim Gamma(2, \lambda+1)$

Metropolis Sampling

1. Situation: Prior is not conjugate ($Y|\mu \sim N(\mu, 1), \mu \sim Beta(a, b)$), not knowing how to make a draw from the FC.
2. $\theta_j^c|\theta_j^* \sim Normal(\theta_j^*, s_j^2) \rightarrow \pi(\theta_j^c|\theta_j^*) = \frac{1}{\sqrt{2\pi s_j^2}} exp\Big(-\frac{1}{2s_j^2}(\theta_j^c - \theta_j^*)^2\Big)$
3. $\theta_j^*|\theta_j^c \sim Normal(\theta_j^c, s_j^2) \rightarrow \pi(\theta_j^*|\theta_j^c) = \frac{1}{\sqrt{2\pi s_j^2}} exp\Big(-\frac{1}{2s_j^2}(\theta_j^* - \theta_j^c)^2\Big)$
4. We propose a random candidate model based on current trials: $\theta_j^c \sim Normal(\theta_j^*, s_j^2)$
5. Metropolis sampling is a version of rejection sampling. The acceptance probability is

$$R = min\Big\{1, \frac{p(\theta_j^c|\theta_{(j)}, Y)}{p(\theta_j^*|\theta_{(j)}, Y)}\Big\}$$

Metropolis Hasting Algorithm

1. Situation: With asymmetric candidate distributions.
2. The acceptance probability is

$$R = min\left\{1, \frac{\frac{p(\theta_j^c|\theta_{(j)}, Y)}{p(\theta_j^*|\theta_{(j)}, Y)}}{\frac{q(\theta_j^c|\theta_{(j)}^*)}{q(\theta_j^*|\theta_{(j)}^c)}}\right\}$$

3. If we take $\theta_j^c \sim p(\theta_j^c|\theta_{(j)}, Y)$, and let $A = \frac{p(\theta_1^c, \theta_2^c, ..., \theta_p^c|Y)}{p(\theta_1^*, \theta_2^*, ..., \theta_p^*|Y}$, we propose for θ_j^c and θ_j^* by updating from θ_j^* and θ_j^c, respectively:
$p(\theta_1^c|\theta_2^*, ..., \theta_p^*, Y) \cdot p(\theta_2^c|\theta_1^c, \theta_3^* ......, \theta_p^*, Y) \cdot ...p(\theta_p^c|\theta_1^c, ..., \theta_{p-1}^c, Y)$
$p(\theta_1^*|\theta_2^c, ..., \theta_p^c, Y) \cdot p(\theta_2^*|\theta_1^*, \theta_3^c ......, \theta_p^c, Y) \cdot ...p(\theta_p^*|\theta_1^*, ..., \theta_{p-1}^*, Y)$
Using the backward substitution for conditional distributions, we obtain

$$B = \frac{p(\theta_1^c|\theta_2^*, ..., \theta_p^*, Y) \cdot ...p(\theta_p^c|\theta_1^c, ..., \theta_{p-1}^c, Y)}{p(\theta_1^*|\theta_2^c, ..., \theta_p^c, Y) \cdot ...p(\theta_p^*|\theta_1^*, ..., \theta_{p-1}^*, Y)} = A \to R = 1$$

Test Reminders

Let us look at the natural conjugate priors for exam 1:

1. Dirichlet is the conjugate prior for multinomial distribution.

The End