Section 2 Bayesian Inference in Gaussian Models

Solutions

Vera Liu

April 6, 2018

0.1

Throughout the section, unless specified, y denotes the observation vector $(y_1...y_n)$

$p(\mu,\omega|y) = \frac{p(y|\mu,\omega)\times p(\mu|\omega)\times p(\omega)}{\int p(y|\mu,\omega)\times p(\mu|\omega)\times p(\omega)}$

$p(\mu,\omega|y) \propto p(y|\mu,\omega) \times p(\mu|\omega) \times p(\omega)$

$\propto \prod_i N(y_i|\mu,\frac{1}{\omega}) \times N(\mu|\mu_0,\frac{1}{\omega k_0}) \times Gamma(\omega|\alpha_0,\beta_0)$

$\propto \omega^{\frac{n}{2}} exp(-\frac{\omega}{2}\sum_i (y_i-\mu)^2)\sqrt{\omega k_0} exp(-\frac{\omega k_0}{2}(\mu-\mu_0)^2)\omega^{\alpha-1}e^{-\beta\omega}$

$\propto \omega^{\alpha+\frac{n}{2}-\frac{1}{2}} exp((-2\beta + k_0(\mu-\mu_0)^2 + \sum_i (y_i-\mu)^2)\frac{\omega}{2})$

$\propto \omega^{\alpha+\frac{n}{2}-\frac{1}{2}} exp(-(k{\mu_0}^2 + \sum {y_i}^2 + 2\beta - \frac{(k_0\mu_0+\sum y_i)^2}{k_0+n})\frac{\omega}{2})$ exp(-$(\mathrm{k}_0+n)(\mu - \frac{k_0\mu_0+\sum y_i}{k_0+n})^2)\frac{\omega}{2})$

$\propto P(\omega|y)P(\mu|\omega,y)$

0.2

$P(\mu|\omega,y) \propto P(y|\mu,\omega)P(\mu|\omega)$

$P(\omega|y) \propto P(y|\mu,\omega)P(\omega)$

From the previous question,

Parameters for posterior of $\mu \sim N(\frac{k_0\mu_0+\sum y_i}{k_0+n}, \frac{1}{\omega(k_0+n)})$

Parameters for posterior of $\omega \sim Gamma(\alpha^*,\beta^*)$, with $\alpha^* = \alpha + \frac{n}{2}$, $\beta^* = \frac{(k{\mu_0}^2+\sum {y_i}^2+2\beta-\frac{(k_0\mu_0+\sum y_i)^2}{k_0+n})}{2}$

0.3

$P(\mu) = \int P(\mu|\omega)P(\omega)d\omega$

$\propto \omega^{\alpha_0-\frac{1}{2}} exp(-(\beta_0 + k_0(\mu-\mu_0)^2/2)\omega)d\omega$

is kernel of $Gamma(\alpha_0+\frac{1}{2}, \beta_0 + k_0(\mu-\mu_0)^2/2)$

Hence previous integral is proportional to the inverse of normalizing constant of the Gamma distribution,

$\propto (\beta_0 + k_0(\mu-\mu_0)^2/2))^{-\alpha_0-\frac{1}{2}}$

$\propto (1 + k_0(\mu-\mu_0)^2/2\beta_0))^{-\alpha_0-\frac{1}{2}}$

$m = \mu_0$

$\nu = 2\alpha_0$

$s = \sqrt{\frac{\beta_0}{\alpha_0 k_0}}$

0.4

Previous exercise shows if $P(\mu|\omega) \sim N(\mu_0,\frac{1}{\omega k_0})$ and $P(\omega) \sim Gamma(\alpha_0,\beta_0)$,

Then, $P(\mu) \sim t(m=\mu_0, s=\sqrt{\frac{\beta_0}{k_0\alpha_0}}, \nu = 2\alpha_0)$

Substitute for the posteriors where $P(\mu|\omega,y) \sim N(\mu_n,\frac{1}{\omega k_n})$ and $P(\omega|y) \sim Gamma(\alpha_n,\beta_n)$

Then $P(\mu) \sim t(m=\mu_n, s=\sqrt{\frac{\beta_n}{k_n\alpha_n}}, \nu = 2\alpha_n)$

0.5

$P(y_1...y_n) = \int P(y_1...y_n|\mu,\omega)P(\mu,\omega)d\mu d\omega$
$\propto \int \omega^{\alpha+\frac{n}{2}-\frac{1}{2}} exp((-2\beta + k_0(\mu-\mu_0)^2 + \sum_i (y_i-\mu)^2)\frac{\omega}{2})d\mu d\omega$
$\propto \int w^{\alpha+\frac{n}{2}-1} exp(-[k_0\mu_0^2 + \sum_{i=1}^n (y_i)^2 + 2\beta - \frac{(k_0\mu_0+\sum_{i=1}^n y_i)^2}{k_0+n}]\frac{\omega}{2})d\omega$
$\propto [k_0\mu_0^2 + \sum_{i=1}^n {y_i}^2 + 2\beta - \frac{(k_0\mu_0+\sum_{i=1}^n y_i)^2}{k_0+n})^{\alpha+\frac{n}{2}}]$

0.6

$P(y_{n+1}...y_{n+m}|y_1...y_n) = \int P(y_{n+1}...y_{n+m}|\mu,\omega)P(\mu,\omega|y_1...y_n)d\mu d\omega$
$\propto \int \omega^{\alpha_n+\frac{m}{2}-\frac{1}{2}} exp((-2\beta_n + k_n(\mu-\mu_n)^2 + \sum_{i=n+1}^{n+m} (y_i-\mu)^2)\frac{\omega}{2})d\mu d\omega$
$\propto \int w^{\alpha_n+\frac{m}{2}-1} exp(-[k_n\mu_n^2 + \sum_{i=n+1}^{n+m} {y_i}^2 + 2\beta_n - \frac{(k_n\mu_n+\sum_{i=n+1}^{n+m} y_i)^2}{k_n+m}]\frac{\omega}{2})d\omega$
$\propto [k_n\mu_n^2 + \sum_{i=n+1}^{n+m} {y_i}^2 + 2\beta_n - \frac{(k_n\mu_n+\sum_{i=1+n}^{n+m} y_i)^2}{k_n+m})^{\alpha_n+\frac{m}{2}}]$
In practice, if can't integrate out in explicit form:
sample $\theta \sim P(\theta)$
then weight by $P(y|\theta)$
$P(y) = \sum P(y|\theta)$

0.7

$P(\Sigma|\nu_0, {\Lambda_0}^{-1}) = \frac{|\Lambda|^{\frac{\nu_0}{2}}}{2^{(\nu_0 d)/2}\Gamma_d(\nu_0/2)}|\Sigma|^{-\frac{\nu_0+d+1}{2}} e^{-\frac{1}{2}tr(\Lambda\Sigma^{-1})}$

When $d = 1$, the univariate case gives: $P(\Sigma|\nu_0, {\Lambda_0}^{-1}) = \frac{|\Lambda|^{\frac{\nu_0}{2}}}{2^{(\nu_0)/2}\Gamma(\nu_0/2)}|\Sigma|^{-\frac{\nu_0+2}{2}} e^{-\frac{1}{2}\Lambda\Sigma^{-1}}$
$= \frac{\frac{\Lambda}{2}^{\frac{\nu_0}{2}}}{\Gamma(\nu_0/2)}|\Sigma|^{-\frac{\nu_0+2}{2}} e^{-\frac{1}{2}\Lambda\Sigma^{-1}}$
Hence $\Sigma^{-1} \sim Gamma(\frac{\nu_0}{2}, \frac{\Lambda_0}{2})$

0.8

$P(\beta|\omega, y_1...y_n) \propto P(\beta, \omega, y_1...y_n)$
$\propto P(y_1...y_n|\beta,\omega) \times P(\beta|\omega) \times P(\omega)$
$= N(X\beta, (\omega\Lambda)^{-1}) \times N(\mu, (\omega K)^{-1}) \times Gamma(a, b)$
$\propto N(X\beta, (\omega\Lambda)^{-1}) \times N(\mu, (\omega K)^{-1})$
$\propto exp(-\frac{1}{2}(Y - X\beta)^T(\omega\Lambda)(Y - X\beta) - \frac{1}{2}(\beta-\mu)^T(\omega K)(\beta-\mu))$
$\propto exp(-\frac{\omega}{2}(\beta^T X^T\Lambda X\beta - 2\beta^T X^T\Lambda y + \beta^T K\beta - 2\beta^T K\mu))$ here eliminating all the terms in exponent without β
$\propto exp(-\frac{\omega}{2}[\beta^T(X^T\Lambda X + K)\beta - 2\beta^T(X^T\Lambda y + K\mu)])$
$\propto exp(-\frac{\omega}{2}[\beta^T(X^T\Lambda X + K)\beta - 2\beta^T(X^T\Lambda y + K\mu)(X^T\Lambda X + K)(X^T\Lambda X + K)^{-1}])$
$\propto exp[-\frac{1}{2}[\beta - (X^T\Lambda X + K)^{-1}(X^T\Lambda Y + K\mu)]^T[\omega(X^T\Lambda X + K)][\beta - (X^T\Lambda X + K)^{-1}(X^T\Lambda Y + K\mu)]]$ – The extra terms from the square are irrelevant, as long as they don't contain β

Hence $\beta|\omega, y_1...y_n \sim N(\mu_n, \Sigma_n)$, where $\mu_n = (X^T\Lambda X + K)^{-1}(X^T\Lambda Y + K\mu)$ and ${\Sigma_n}^{-1} = \omega(X^T\Lambda X + K)$

0.9

$P(\omega|y_1...y_n) \propto P(\omega, y_1...y_n)$
$= \int P(y_1...y_n|\beta,\omega) \times P(\beta|\omega) \times P(\omega)d\beta$
$= \int N(X\beta, (\omega\Lambda)^{-1}) \times N(\mu, (\omega K)^{-1}) \times Gamma(a, b)d\beta$
$= |\frac{\omega\Lambda}{2\pi}|^{\frac{1}{2}}|\frac{\omega K}{2\pi}|^{\frac{1}{2}} exp[-\frac{\omega}{2}(y^T\Lambda Y + \mu^T K\mu - {\mu_n}^T(X_T\Lambda X + K)\mu_n)]\frac{b^a}{\Gamma a}\omega^{a-1}e^{-b\omega}$, all parts involving β integrates (with respect to β) to 1 as a normal density, as shown previously.

$\propto \omega^{\frac{n+d}{2}+a-1} exp[-\omega(b + \frac{1}{2}(y^T\Lambda y + \mu^T K\mu - {\mu_n}^T(X_T\Lambda X + K)\mu_n))]$
Hence $\omega|y_1...y_n \sim Gamma(\frac{n+d}{2} + a, b + \frac{1}{2}(y^T\Lambda y + \mu^T K\mu - {\mu_n}^T(X_T\Lambda X + K)\mu_n))$

0.10

Following is the implementation of the above sampler with a real world data set. In this case, K and λ are fixed as identity matrix.

Set observation and covariate vectors:

```
xmatrix<-as.matrix(dental[,c(3,6,7)])
ymatrix<-as.matrix(dental$distance)
```

Parameters of interest:

```
n=1000
omega<-rep(NA,n)
omega[1]<-1
beta_age<-rep(NA, n)
beta_males<-rep(NA, n)
beta_intercept<-rep(NA, n)
a<-10
b<-10
```

Implementation of Gibbs Sampler:

```
for (i in 1:n){
    beta <- mvrnorm(1, solve(t(xmatrix) %*% xmatrix +diag(dim(xmatrix)[2])) %*% (t(
        xmatrix) %*% ymatrix), 1/omega[i] * solve(t(xmatrix) %*% xmatrix+diag(dim(
        xmatrix)[2])))
    beta_age[i]<-beta[1]
    beta_males[i] <-beta[2]
    beta_intercept[i] <-beta[3]
    omega[i+1]<-rgamma(1,a+(dim(xmatrix)[2]+dim(xmatrix)[1])/2, b+0.5*(t(ymatrix -
        xmatrix %*% (as.matrix(beta))) %*%(ymatrix - xmatrix %*% (as.matrix(beta)))
        + t(as.matrix(beta)) %*% (as.matrix(beta))))}
```

Summary of coefficients:

	Intercept	Age	Males
Mean	12.390	0.913	2.523
Variance	1.210	0.009	0.223
Ridge Estimate	13.944	0.766	2.579 LS Estimate
15.386	0.660	2.321	

The results of ridge estimate and bayesian estimate are very close.

0.11

$P(y, \lambda, \beta, \omega) = N(X\beta, (\omega diag(\lambda_1 ... \lambda_n)) \times N(\mu, (\omega k)^{-1}) \times Gamma(a, b) \times Gamma(\tau, \tau)$

$P(\lambda|y, \beta, \omega) \propto N(X\beta, (\omega diag(\lambda_1 ... \lambda_n)) \times Gamma(\tau, \tau)$

$P(\lambda_i|y_i, \beta, \omega) \propto \sqrt{\omega\lambda_i} exp(-\frac{\omega\lambda_i}{2}(y_i - {x_i}^T\beta)^2){\lambda_i}^{\tau-1}e^{-\tau\lambda_i}$

$\propto {\lambda_i}^{\tau-\frac{1}{2}} exp(-\lambda_i(\tau + \frac{\omega(y_i - {x_i}^T\beta)^2}{2}))$

Hence $\lambda_i \sim Gamma(\tau + \frac{1}{2}, \tau + \frac{\omega(y_i - {x_i}^T\beta)^2}{2}))$

Similarly, $P(\beta|y, \omega, \lambda_i) \propto N(X\beta, \omega\lambda_i) \times N(\mu, (\omega K)^{-1})$

$P(\omega|\lambda_i, y, \beta) \propto N(X\beta, \omega\lambda_i) \times N(\mu, (\omega K)^{-1}) \times Gamma(a, b)$

0.12

Set the priors and parameters of interest. Compared with previous case, there is an extra parameter λ

```
tau<-1
lambda<-rep(1, dim(xmatrix)[1])
omega_mix<-rep(NA,n)
omega_mix[1]<-1
beta_age_mix<-rep(NA, n)
beta_males_mix<-rep(NA, n)
beta_intercept_mix<-rep(NA, n)
```

Implementation of the Gibbs Sampler:

```
for (i in 1:n){
    lambda_pr<-diag(lambda)

    beta <- mvrnorm(1, solve(t(xmatrix) %*%lambda_pr%*% xmatrix +diag(dim(xmatrix)
        [2])) %*% (t(xmatrix) %*% lambda_pr %*% ymatrix), 1/omega_mix[i] * solve(t(
        xmatrix) %*% lambda_pr %*% xmatrix+diag(dim(xmatrix)[2])))

    beta_age_mix[i]<-beta[1]
    beta_males_mix[i] <-beta[2]

    beta_intercept_mix[i] <-beta[3]

    omega_mix[i+1]<-rgamma(1,a+(dim(xmatrix)[2]+dim(xmatrix)[1])/2, b+0.5*(t(
        ymatrix - xmatrix %*% (as.matrix(beta))) %*% lambda_pr %*% (ymatrix -
        xmatrix %*% (as.matrix(beta)))+ t(as.matrix(beta)) %*% (as.matrix(beta))))

    for (j in 1:dim(xmatrix)[1]){
    lambda[j] <- rgamma(1, tau+1/2, shape=1*(tau+0.5*(omega_mix[i+1]*(ymatrix[j] -
        t(xmatrix[j]) %*% beta)^2)))}
}
```

Summary of coefficients:

	Intercept	Age	Males
Mean	14.814	0.754	2.394
Variance	1.412	0.011	0.2449