Statistics 360/601 – Modern Bayesian Theory

Alexander Volfovsky

Lecture ? - Oct 25, 2018

Multivariate normal code

```
> rmvnorm <- function(n,mu,Sigma) {
+ # n = sample size
+ # mu = p dimensional vector of means
+ # Sigma = p x p symmetric positive
+ # definite covariance matrix
+   p<-length(mu)
+   res<-matrix(0,nrow=n,ncol=p)
+   if( n>0 & p>0 ) {
+   E<-matrix(rnorm(n*p),n,p)
+   res<-t(  t(E%*%chol(Sigma)) +c(mu))
+                     }
+   res
+                         }
```

Multivariate normal likelihood code

```
> ldmvnorm<-function(y,mu,Sig){  # log mvn density
+     c(  -(length(mu)/2)*log(2*pi) -.5*log(det(Sig)) -
+           .5*t(y-mu) %*% solve(Sig) %*% (y-mu)   )
```

Wishart and Inverse Wishart code

```
> ### sample from the Wishart distribution
> rwish<-function(n,nu0,S0)
+ {
+   sS0 <- chol(S0)
+   S<-array( dim=c( dim(S0),n ) )
+   for(i in 1:n)
+   {
+      Z <- matrix(rnorm(nu0 * dim(S0)[1]),
+              nrow = nu0, ncol = dim(S0)[1]) %*% sS0
+      S[,,i]<- t(Z)%*%Z
+   }
+   S[,,1:n]
+ }
```

Wishart and Inverse Wishart code

```
> ### sample from the inverse Wishart distribution
> rinvwish<-function(n,nu0,iS0)
+ {
+   sL0 <- chol(iS0)
+   S<-array( dim=c( dim(iS0),n ) )
+   for(i in 1:n)
+   {
+      Z <- matrix(rnorm(nu0 * dim(iS0)[1]),
+           nrow = nu0, ncol = dim(iS0)[1]) %*% sL0
+      S[,,i]<- solve(t(Z)%*%Z)
+   }
+   S[,,1:n]
+ }
```

Data analysis

- Twenty two children take a reading comprehension test before and after a training session.

Data analysis

- Twenty two children take a reading comprehension test before and after a training session.
- The test is graded on a 0 to 100 scale.

Data analysis

- Twenty two children take a reading comprehension test before and after a training session.
- The test is graded on a 0 to 100 scale.
- We think it is jointly normal(θ, Σ).

Data analysis

- Twenty two children take a reading comprehension test before and after a training session.
- The test is graded on a 0 to 100 scale.
- We think it is jointly normal(θ, Σ).
- Need to specify a prior for $\theta \sim N(\mu_0, \Lambda_0)$ and $\Sigma \sim iW(\nu_0, S_0^{-1})$

Data analysis

- Twenty two children take a reading comprehension test before and after a training session.
- The test is graded on a 0 to 100 scale.
- We think it is jointly normal(θ, Σ).
- Need to specify a prior for $\theta \sim N(\mu_0, \Lambda_0)$ and $\Sigma \sim iW(\nu_0, S_0^{-1})$
- Prior for the mean: exam is designed to have a mean of 50, so $\mu_0 = (50, 50)$ seems reasonable as a prior expectation.

Data analysis

- Twenty two children take a reading comprehension test before and after a training session.
- The test is graded on a 0 to 100 scale.
- We think it is jointly normal(θ, Σ).
- Need to specify a prior for $\theta \sim N(\mu_0, \Lambda_0)$ and $\Sigma \sim iW(\nu_0, S_0^{-1})$
- Prior for the mean: exam is designed to have a mean of 50, so $\mu_0 = (50, 50)$ seems reasonable as a prior expectation.
- Want to make sure θ is unlikely to be bigger that 100 or smaller than 0: $\lambda_{01}^1 = \lambda_{02}^2 = (50/2)^2$ puts that probability at 0.05.

Data analysis

- Twenty two children take a reading comprehension test before and after a training session.
- The test is graded on a 0 to 100 scale.
- We think it is jointly normal(θ, Σ).
- Need to specify a prior for $\theta \sim N(\mu_0, \Lambda_0)$ and $\Sigma \sim iW(\nu_0, S_0^{-1})$
- Prior for the mean: exam is designed to have a mean of 50, so $\mu_0 = (50, 50)$ seems reasonable as a prior expectation.
- Want to make sure θ is unlikely to be bigger that 100 or smaller than 0: $\lambda_{01}^1 = \lambda_{02}^2 = (50/2)^2$ puts that probability at 0.05.
- The two exams are measuring the same thing so θ_1 and θ_2 should be close: maybe the prior correlation between then is 0.5

Data analysis

- Twenty two children take a reading comprehension test before and after a training session.
- The test is graded on a 0 to 100 scale.
- We think it is jointly normal(θ, Σ).
- Need to specify a prior for $\theta \sim N(\mu_0, \Lambda_0)$ and $\Sigma \sim iW(\nu_0, S_0^{-1})$
- Prior for the mean: exam is designed to have a mean of 50, so $\mu_0 = (50, 50)$ seems reasonable as a prior expectation.
- Want to make sure θ is unlikely to be bigger that 100 or smaller than 0: $\lambda_{01}^1 = \lambda_{02}^2 = (50/2)^2$ puts that probability at 0.05.
- The two exams are measuring the same thing so θ_1 and θ_2 should be close: maybe the prior correlation between then is 0.5
- Know a little less about Σ but similar rational to the choices for Λ_0 applies. Let $S_0 = \Lambda_0$ and make it loosely centered around that point so let $\nu_0 = p + 2 = 4$.

Gibbs sampler — set up

```
> # read the data
> Y <- dget(readpp.dat)
> # set prior for the mean
> mu0<-c(50,50)
> L0<-matrix( c(625,312.5,312.5,625),nrow=2,ncol=2)
> # set prior for the covariance
> nu0<-4
> S0<-matrix( c(625,312.5,312.5,625),nrow=2,ncol=2)
> # compute statistics for faster updates
> n<-dim(Y)[1] ; ybar<-apply(Y,2,mean)
> Sigma<-cov(Y) ; THETA<-SIGMA<-NULL
> YS<-NULL
> set.seed(1)
```

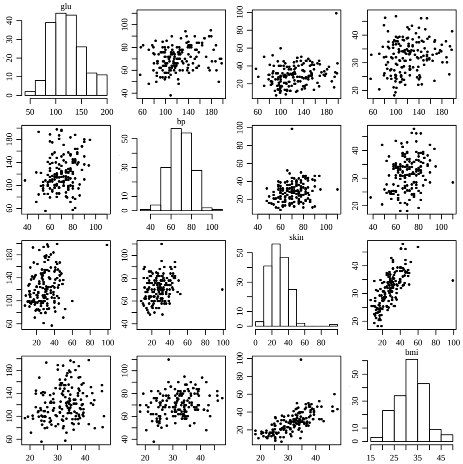
Gibbs sampler — for loop

```
> for(s in 1:5000)
+{
+  ###update theta
+  Ln<-solve( solve(L0) + n*solve(Sigma) )
+  mun<-Ln%*%( solve(L0)%*%mu0 + n*solve(Sigma)%*%ybar )
+  theta<-rmvnorm(1,mun,Ln)
+  ###update Sigma
+  Sn<- S0 + ( t(Y)-c(theta) )%*%t( t(Y)-c(theta) )
+# Sigma<-rinvwish(1,nu0+n,solve(Sn))
+  Sigma<-solve( rwish(1, nu0+n, solve(Sn)) )
+  ### save results
+  THETA<-rbind(THETA,theta); SIGMA<-rbind(SIGMA,c(Sigma))
+  YS<-rbind(YS,rmvnorm(1,theta,Sigma))
+  ###
+  cat(s,round(theta,2),round(c(Sigma),2),"\n")
+}
```

What can we report?

- Might be interested in the probability that the population average increases: $P(\theta_2 > \theta_1 | y_1, \ldots, y_n) = 0.99$
- Might be interested in whether a randomly selected new child do better on the second test: $P(Y_2 > Y_1 | y_1, \ldots, y_n) = 0.71$
- What do these two results mean? Why are they so different?
- Note that $E[\theta_2 - \theta_1 | y_1, \ldots, y_n] = 6.6$ which is pretty small in the grand scheme of things.
- The first probability captures ANY difference in the θs.
- The second probability captures individual uncertainty about testing.

Missing data example



Data look essentially normal. We could model them as

$$Y_i \sim N(\theta, \Sigma)$$

Missing data example

	glu	bp	skin	bmi
1	86	68	28	30.20
2	195	70	33	NA
3	77	82	NA	35.80
4	NA	76	43	47.90
5	107	60	NA	NA
6	97	76	27	NA
7	NA	58	31	34.30
8	193	50	16	25.90
9	142	80	15	NA
10	128	78	NA	43.30

Problem: what is $p(y_2|\theta, \Sigma)$??

Gibbs with missing data

```
> for(s in 1:5000)
+{
+  #update theta...
+  #update Sigma...
+  ###update missing data
+  for(i in 1:n)
+  {
+   b <- ( O[i,]==0 )
+   a <- ( O[i,]==1 )
+   iSa<- solve(Sigma[a,a])
+   beta.j <- Sigma[b,a]%*%iSa
+   s2.j   <- Sigma[b,b] - Sigma[b,a]%*%iSa%*%Sigma[a,b]
+   theta.j<- theta[b]+beta.j%*%(t(Y.full[i,a])-theta[a])
+   Y.full[i,b] <- rmvnorm(1,theta.j,s2.j )
+  }
+}
```