HW 20 Solutions

Eric Slud 12/20/2016

Part (2) first. As we showed in class, if $X_i \sim \mathcal{N}(\mu, 1/\tau)$ are *iid* for $i = 1, \ldots, n$ with prior τ, μ independent following densities $\mu \sim \mathcal{N}(0, 1/\epsilon)$ and $\tau \sim \text{Gamma}(\nu/2, 1/2)$, then the joint density of $\mu, \tau, \mathbf{X}$ is

$$C\, \tau^{(n+\nu)/2-1} \exp\Big(- \frac{\tau}{2} \Big[1 + \sum_{i=1}^n (X_i - \mu)^2\Big]\Big)$$

from which we read off the full (posterior) conditional distributions

$$f_{\tau|\mu,\mathbf{X}}(\tau \,|\, \mu, \mathbf{X}) \sim \text{Gamma}(\frac{n+\nu}{2}, \frac{1}{2}\Big[1 + \sum_{i=1}^n (X_i - \mu)^2\Big])$$

and

$$f_{\mu|\tau,\mathbf{X}}(\mu \,|\, \tau, \mathbf{X}) \sim \mathcal{N}\Big(\frac{n\tau\, \bar{X}}{n\tau + \epsilon}, \frac{1}{n\tau + \epsilon}\Big)$$

where $\bar{X} = \sum_{i=1}^n X_i/n$.

So with $U = \mu$ and $V = n\tau$, we have conditionally given $\mathbf{X}$:

$$U \sim \mathcal{N}(\frac{V\,\bar{X}}{V + \epsilon}, \frac{1}{V+\epsilon}) , \quad V \sim \text{Gamma}(\frac{n+\nu}{2}, \frac{1}{2n}(1 + \sum_{i=1}^n (X_i - \mu)^2))$$

since Gamma variables form a scale family with the second parameter inversely proportional to scale. This says that the conditional distributions given in part (1) of the problem apply, with

$$A = \frac{n+\nu}{2} , \; B_1 = \frac{1}{2n}\Big[1 + \sum_{i=1}^n (X_i - \bar{X})^2\Big] , \; B_2 = \frac{1}{2} , \; b = \bar{X} , \; C = \epsilon , \; D = \bar{X}$$

Next we provide code for the Gibbs Sampler Iteration, mapping 2-vector Uv = (U0,V0) as input to Uout = (U1,V1) as output.

```
GibbsIter = function(Uv, A,B1,B2,b,C,D) {
      var = 1/(Uv[2]+C)
      U1 = rnorm(1,D*Uv[2]*var,sqrt(var))
      U2 = rgamma(1,A,B1+B2*(U1-b)^2)
      c(U1,U2)  }
```

In order to obtain a good representation of the (marginal) density of U, we start with an arbitrary `Uv=c(0,1)`, run the Gibbs Sampler a large number of times (10^5) and show a scaled relative frequency histogram of the last $90,000$ simulated U numbers as our approximate density result, with smoothed `loess` density overplotted along with normal density with the same mean and variance

```
UVarr = array(0 ,c(1e5,2))
uv = c(0,1)
A = 10; B1 = 2; B2 = 0.5; b = 3; C = 0.1; D = 3
for (i in 1:1e5) {
   uv = GibbsIter(uv,A,B1,B2,b,C,D)
   UVarr[i,] = uv }
tmp = hist(UVarr[10001:1e5,1], nclass=50, prob=T,
   xlab="U values", ylab="density",
   main="Histogram of Gibbs-Sampled U's \n after Burn-In")
lines(density(UVarr[10001:1e5,1]), col="blue")
curve(dnorm(x, mean(UVarr[10001:1e5,1]),
   sd(UVarr[10001:1e5,1])), add=T, col="red")

legend(locator(), legend=c("U density","normal"),
   lty=1, col=c("blue","red"))
```

In the figure `HW20Udist.pdf`, you can see that there is a slight difference between the density of Gibbs-sampled U and the corresponding normal density (with same mean and sd as Gibbs-sampled U). The normal density has slightly shorter peak and slightly fatter tails, at least for the given choices of `A, B1, B2, b, C, D` which are meant to conform roughly to the posterior parameters normal data for X_i with $\bar{X} = 3, S^2 = 4, \epsilon = .1, \nu = 1$.

Part (3) – application to female blue crab data.

```
> library(MASS)
> Xv = crabs$FL[crabs$sp=="B" & crabs$sex=="F"]    ### 50 values
> hist(Xv)               ### vaguely normally distributed

> A = (3+50)/2; B1 = 0.01*(1+49*var(Xv)); B2=0.5; b = D = mean(Xv); C=.05
  UVarr = replace(UVarr,T,0)
  uv = c(0,1)
  for (i in 1:1e5) {
        uv = GibbsIter(uv,A,B1,B2,b,C,D)
        UVarr[i,] = uv }
```

We will find posterior expectations of $U = \mu$ from `UVarr[10001:1e5,1]` and of $\sigma^2 = 1/\tau = 50/V$ from `50/UVarr[10001:1e5,2]`.

```
> mean(UVarr[10001:1e5,1])
[1] 13.1792
> mean(50/UVarr[10001:1e5,2])
[1] 6.7888
## Contrast these numbers with mean(Xv) = 13.27 and var(Xv) = 6.9054.
```

A full-credit solution should also contain some checking that the Gibbs Sampler Converged, either by overplotting histograms from different subsets of iterates or goodness of fit tests or something like that. Solutions without such checking had 1 point deducted.