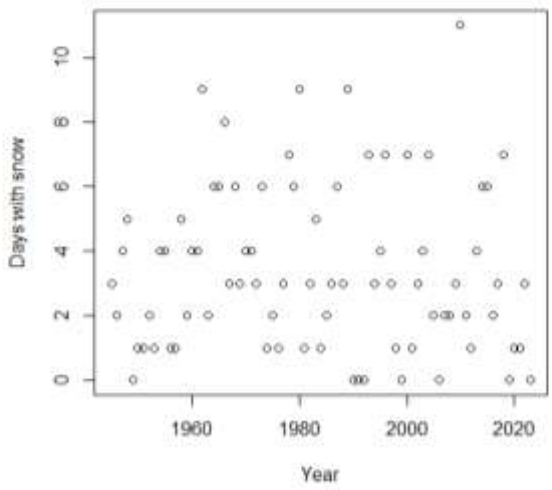
ST440/540 Applied Bayesian Analysis

Lab activity for 2/19/2024

A. QUIZ SOLUTIONS

Let Y_t be the number of days with snow at RDU Airport in year t. The data are below.



(a) Assuming the data are independent and identically distributed across years, specify and justify an appropriate family of distributions for the data (i.e., the likelihood function).

Since Y_t has support {0,1,…,365} I would assume Y_t ~ Binomial(365,θ). Since the counts are nowhere near 365, a Poisson distribution (with gamma prior for the rate) is fine too.

(b) Specify and justify an appropriate family of distributions for the prior distribution of the parameters in the model you defined in (a).

Since θ is a real number between 0 and 1 I would use a beta prior.

B. DISCUSSION QUESTIONS

(1) In this problem we will compute MAP estimates of Gaussian models

(a) Say $Y \sim \mathrm{Normal}(\mu, \sigma^2)$ with prior $\pi(\mu) = 1$ and σ^2 assumed to be known. Compute the MAP estimator of μ.

(b) Say $Y_1, \ldots, Y_n \sim \mathrm{Normal}(\mu, \sigma^2)$ with prior $\pi(\mu) = 1$ and σ^2 assumed to be known. Compute the MAP estimator of μ.

See the solution on the next page

(2) Say $Y_i \sim \mathrm{Normal}(\beta_0 + \beta_1 X_i, \sigma^2)$ for $i = 1, \ldots, n$ with prior $\pi(\beta_0, \beta_1) = 1$ and σ^2 assumed to be known. Show that the MAP estimator of $\beta = (\beta_0, \beta_1)$ is the least squares estimator

$$\hat{\beta} = argmin_{\beta} \sum_{i=1}^{n} (Y_i - \beta_0 - X_i \beta_1)^2$$

(you don’t have to compute the derivation, just get far enough to show equivalence).

See the solution on the next page

(1a) $p(\mu|y) \propto f(y|\mu)\, \overset{1}{\cancel{\pi(\mu)}} \propto f(y|\mu) \propto e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$

$$\log\left[p(\mu|y)\right] = \text{constant} - \frac{1}{2\sigma^2}(y-\mu)^2$$

$$\frac{\partial}{\partial\mu} = \frac{1}{\sigma^2}(y-\mu) = 0 \implies \hat{\mu}_{MAP} = y$$

independence

(1b) $p(\mu|y) \propto f(y_1,\dots,y_n|\mu)\,\pi(\mu) \propto \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \cdot 1$

$$\log\left[f(\mu|y)\right] = \text{constant} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

$$\frac{\partial}{\partial\mu} = \frac{1}{\sigma^2}\sum(y_i-\mu) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} y_i - n\mu\right) = 0$$

$$\implies \hat{\mu}_{MAP} = \frac{\sum y_i}{n} = \bar{y}$$

independence

(2) $p(\beta|y) \propto f(y|\beta)\,\overset{1}{\cancel{\pi(\beta)}} \propto \prod_{i=1}^{n} f(y_i|\beta)$

$$\propto e^{-\frac{1}{2\sigma^2}(y_i-\beta_0-\beta_1 x_i)^2}$$

$$\propto e^{-\frac{1}{2\sigma^2}SSE}$$

So, maximizing $p(\beta|y)$ is equivalent to minimizing SSE.

(3) Consider the multiple regression model

$$Y = \beta_0 + X_1\beta_1 + \cdots + X_p\beta_p + \varepsilon$$

Say our goal is to study the effect of solar radiation (X_1) on ozone (Y), as measured by the slope β_1. Use the output below to approximate the marginal posterior distribution of β_1, $p(\beta_1|Y)$, including (a) a point estimate, (b) 95% credible set, and (c) the posterior probability that solar radiation has a positive effect on ozone. Justify this approximation, including listing your key assumptions.

```
> data(airquality)
> summary(lm(Ozone~Solar.R+Wind+Temp,data=airquality))
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -64.34208   23.05472  -2.791  0.00623 **
Solar.R       0.05982    0.02319   2.580  0.01124 *
Wind         -3.33359    0.65441  -5.094 1.52e-06 ***
Temp          1.65209    0.25353   6.516 2.42e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.18 on 107 degrees of freedom
  (42 observations deleted due to missingness)
Multiple R-squared:  0.6059,    Adjusted R-squared:  0.5948
F-statistic: 54.83 on 3 and 107 DF,  p-value: < 2.2e-16
```

Evoking the Bayesian Central Limit Theorem, we can approximate

beta1|Y ~ Normal(0.05982,0.02319^2).

Therefore approximately (a) the posterior mean is 0.05920, a 95% credible interval is qnorm(c(0.025,0.975), 0.05982,0.02319) and P(beta1>|Y)=1-pnorm(0, 0.05982,0.02319).

(4) Assume the model $Y|\sigma^2, b \sim Normal(0, \sigma^2)$, $\sigma^2|b \sim InvGamma(1, b)$ and $b \sim Gamma(1,1)$.

(a,b) Derive the full conditional distribution of σ^2 and b.

$$p(\sigma^2|Y,b) \propto p(Y|\sigma^2,b)\, p(\sigma^2|b)\, \cancel{p(b)}$$

$$\propto \left[(\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{\sigma^2}\frac{Y^2}{2}}\right]\left[(\sigma^2)^{-(1+1)} e^{-\frac{b}{\sigma^2}}\right]$$

$$\propto (\sigma^2)^{-\left(\frac{3}{2}+1\right)} e^{-\frac{1}{\sigma^2}\left(\frac{Y^2}{2}+b\right)}$$

$$\Rightarrow \sigma^2|b,Y \sim \text{Inv Gamma}\left(3/2+1,\ Y^2/2+b\right)$$

doesn't actually involve b

$$p(b|\sigma^2,Y) \propto \cancel{f(Y|\sigma^2,b)}\, p(\sigma^2|b)\, p(b)$$

$$\propto \left\{b^1 e^{-\frac{b}{\sigma^2}}\right]\left[e^{-b}\right]$$

$$\propto b^{2-1} e^{-b(1/\sigma^2+1)}$$

$$\Rightarrow b|\sigma^2,Y \sim \text{Gamma}\left(2,\ 1/\sigma^2+1\right)$$

(c) Sketch out a Gibbs sampler to draw samples from the joint distribution of $\sigma^2, b|Y$.

Set initial values for σ^2 and b

(1) Draw $\sigma^2|b$ from the inverse gamma distribution in (a)

(2) Draw $b|\sigma^2$ from a gamma distribution (b)

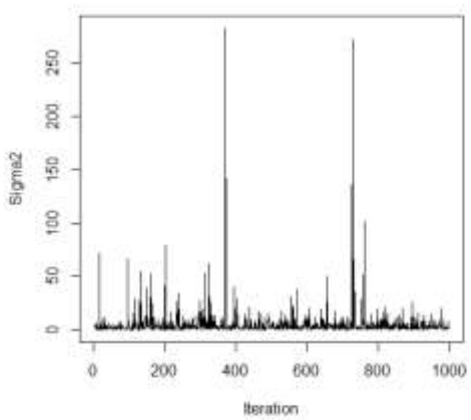
Repeat (1) and (2) S times

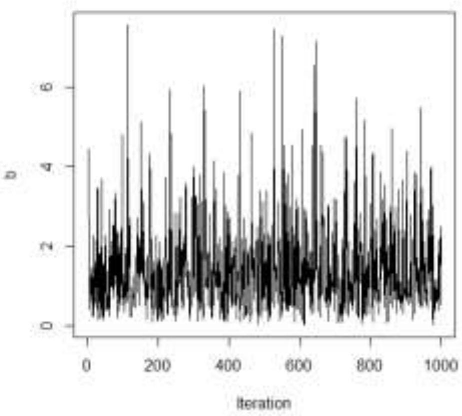
(d) How would you pick initial values?

Since Y has mean zero, Y^2 is an estimate of the variance and could be used as in initial value for σ^2. The model of an InvGamma distribution is b/(a+1), so with a=1 maybe $2\sigma^2$.for the initial value of b

(e) Here are 1000 samples from both parameters (code below) with Y=3, would you say the chain has converged?

I’d say they have converged, but need to run long to give a good approximation to the posterior.





```
S       <- 1000
Y       <- 3
keepers <- matrix(0,S,2)
sigma2  <- 1
b       <- 1
for(iter in 1:S){
   sigma2         <- 1/rgamma(1,1/2+1,Y/2+b)
   b              <- rgamma(1,1+1,1+1/sigma2)
   keepers[iter,] <- c(sigma2,b)
}
plot(keepers[,1],type="l",xlab="Iteration",ylab="Sigma2")
plot(keepers[,2],type="l",xlab="Iteration",ylab="b")
```

(5) Assume the model Y|N,λ ~ Poisson(Nλ). We have been assuming that N is known and λ ~ Gamma(a,b), in which case λ|Y ~ Gamma(Y+a,N+b). How let's say we don't know N and has prior N ~ Gamma(c,d). Below is a Gibbs sampler to approximate the posterior of N,λ|Y.

(a) Would you say the algorithm has converged?

This is borderline but I would run the chains longer. This shows the need for formal diagnostics, which we will cover soon.

(b) How would you approximate the posterior mean and 95% interval of the mean Nλ?

I'd compute N*lambda each iteration and then use sample mean and quantiles.

(c) Why the last plot so strange?

The outcome is Y=100 so we can assume the mean is around 100. The plot is all combinations of N and lambda so that N*lambda=100, e.g., (N,lambda) = (1,100), (10,10) and (100,1)

```
Y <- 100
a <- b <- c <- d <- 0.01

# Initial values
N   <- 10
lam <- 10

# Store output
S <- 10000
samps            <- matrix(10,S,2)
colnames(samps) <- c("N","lambda")

# Go Gibbs!
for(iter in 2:S){
  lam          <- rgamma(1,Y+a,N+b)
  N            <- rgamma(1,Y+c,lam+d)
  samps[iter,] <- c(N,lam)
}

plot(samps[,1],type="l",xlab="Iteration",ylab="N")
plot(samps[,2],type="l",xlab="Iteration",ylab="lambda")
plot(samps,xlab="N",ylab="lambda")
```