STA 360: Assignment 4

Michael Lin

February 2, 2015

1. Posterior predictive derived as follows:
$$X_{n+1} = \theta + Z$$
given $x_{1:n}$. Since $\theta \sim \text{Normal}(3, 1)$ and $Z \sim \text{Normal}(0, \lambda^{-1})$, we can use the linearity of normal distribution to get:
$$X_{n+1} \sim \text{Normal}(3 + 0, 1 + \lambda^{-1}) = \text{Normal}(3, 1 + \lambda^{-1})$$

3. The i.i.d. normal model might not be appropriate because global warming may influence amount of snowfall in a certain direction. For example, if global warming causes higher snowfall, then each city would report increasingly higher snowfall over the years, and the snowfall distribution would be left skewed.

4. Here, we chose a normal-gamma prior with the following parameters:
$$\begin{aligned} m &= 300 \\ c &= 1 \\ a &= 1/2 \\ b &= 80^2 a = 3200 \end{aligned}$$
The method of sampling the random variates from the posterior distribution NormalGamma(M, C, A, B) is simple:
   - The posterior precision τ has distribution Gamma(A, B)
   - The posterior mean μ has distribution Normal$(M, 1/(C\tau))$

   Using Monte Carlo approximation and the above prior hyperparameters, we were able to obtain an empirical value for probability that the mean annual snowfall for Valdez is higher than that of Aomori. With sample size of 10000, we found that:
$$\mathbb{P}(\mu_v > \mu_a | x_{1:n_v}, y_{1:n_a}) \approx 0.9906$$
As a result, we infer that indeed the mean annual snowfall for Valdez is higher than that of Aomori.

5. The following table summarizes the probability (three different trials) that the mean annual snowfall for Valdez is higher than that of Aomori for a few different settings of the prior hyperparameters.

$\mathbb{P}(\mu_v > \mu_a)$	m	c	a	b
0.9906, 0.9903, 0.9916	300	1	1/2	3200
0.9883, 0.9903, 0.9903	500	1	1/2	3200
0.9907, 0.9896, 0.9889	300	1	5	3200
0.9887, 0.9888, 0.9901	300	1	1/2	5000

See below for R code of Monte Carlo approximation.

```
## Load Data
ao = c(188.6, 244.9, 255.9, 329.1, 244.5, 167.7, 298.4, 274.0, 241.3, 288.2, 208.3, 311.4,
       273.2, 395.3, 353.5, 365.7, 420.5, 303.1, 183.9, 229.9, 359.1, 355.5, 294.5, 423.6,
       339.8, 210.2, 318.5, 320.1, 366.5, 305.9, 434.3, 382.3, 497.2, 319.3, 398.0, 183.9,
       201.6, 240.6, 209.4, 174.4, 279.5, 278.7, 301.6, 196.9, 224.0, 406.7, 300.4, 404.3,
       284.3, 312.6, 203.9, 410.6, 233.1, 131.9, 167.7, 174.8, 205.1, 251.6, 299.6, 274.4,
       248.0)

v = c(351.0, 379.3, 196.1, 312.3, 301.4, 240.6, 257.6, 304.5, 296.0, 338.8, 299.9, 384.7,
      353.5, 312.8, 550.7, 327.1, 515.8, 343.4, 341.6, 396.9, 267.3, 230.6, 277.4, 341.0,
      377.0, 391.3, 337.0, 250.4, 353.7, 307.7, 237.5, 275.2, 271.4, 266.5, 318.7, 215.5,
      438.3, 404.6)


## Define prior
m = 300; c = 1; a = 1/2; b = (80)^2*a;
prior.par = c(m,c,a,b);
prior = dnormgam(prior.par, plot=FALSE)


## Define posterior
ao.M = (c*m+sum(ao))/(length(ao)+c);
ao.C = length(ao) + c;
ao.A = a+length(ao)/2;
ao.B = b +  1/2*(c*m^2-ao.C*ao.M^2+sum(ao^2))
ao.par = c(ao.M, ao.C, ao.A, ao.B)
ao.post = dnormgam(ao.par, plot=FALSE)

v.M = (c*m+sum(v))/(length(v)+1);
v.C = length(v) + c;
v.A = a+length(v)/2;
v.B = b +  1/2*(c*m^2-v.C*v.M^2+sum(v^2))
v.par = c(v.M, v.C, v.A, v.B)
v.post = dnormgam(v.par, plot=FALSE)


## Sample Variates for ao
N = 10000
ao.mean.samp = rep(0,N)
ao.pres.samp = rep(0,N)
for (i in 1:N) {
  ao.pres.samp[i] = rgamma(1, ao.A, ao.B)
  ao.mean.samp[i] = rnorm(1, ao.M, 1/sqrt(ao.C*ao.pres.samp[i]));
}


## Sample Variates for v
v.mean.samp = rep(0,N)
v.pres.samp = rep(0,N)
for (i in 1:N) {
  v.pres.samp[i] = rgamma(1, v.A, v.B)
  v.mean.samp[i] = rnorm(1, v.M, 1/sqrt(v.C*v.pres.samp[i]));
```

```
}


## Monte Carlo
index = 0
for (i in 1:N) {
  if(v.mean.samp[i]>ao.mean.samp[i]) {
    index = index + 1
  }
}
prob = index/N

print(prob)
```