CSC 591, HW: Bayesian Parameter Estimation
Student Id: 200261202
Mkhan8@ncsu.edu

Solution 1:

We know that σ^2 is known and μ is unknown. We know that the sample are drawn from $N(\mu, \sigma^2)$.
We also know that $\mu \sim N(\mu_0, \sigma_0^2)$.

1. The posterior distribution of μ is given by:

$$p(\mu|x) \propto (\prod p(x_i\,|\mu, \sigma^2) * p(\mu)$$
$$\propto \left(\exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\ \right\}\right)\left(\exp\left\{-\frac{1}{2\sigma^2}\sum(\mu-\mu_0)^2\ \right\}\right)$$
$$\propto \left(\exp\left\{-\frac{1}{2}\left(\frac{n\sigma_0^2+\sigma^2}{\sigma_0^2\sigma^2}\right)\mu^2 + 2\mu\ \left(\frac{\sigma_0^2\sum x_i + \mu_0\sigma^2}{2\sigma^2\sigma_0^2}\right)\right\}\right)$$
$$\propto \left(\exp\left\{-\frac{1}{2}\left(\frac{n\sigma_0^2+\sigma^2}{\sigma_0^2\sigma^2}\right)\left[\mu^2 - 2\mu\ \left(\frac{\sigma_0^2\sum x_i + \mu_0\sigma^2}{\sigma^2+n\sigma_0^2}\right)\right]\right\}\right)$$

2. We know that the posterior distribution for gaussian is $p(\mu|x)$. And from above calculation we know that

$$p(\mu|x)\ \propto \left(\exp\left\{-\frac{1}{2}\left(\frac{n\sigma_0^2+\sigma^2}{\sigma_0^2\sigma^2}\right)\left[\mu^2 - 2\mu\ \left(\frac{\sigma_0^2\sum x_i + \mu_0\sigma^2}{\sigma^2+n\sigma_0^2}\right)\right]\right\}\right)$$

$$p(\mu|x)\ \propto \left(\exp\left\{-\frac{1}{2}\left(\frac{n\sigma_0^2+\sigma^2}{\sigma_0^2\sigma^2}\right)\left[\mu - \left(\frac{\sigma_0^2\sum x_i + \mu_0\sigma^2}{\sigma^2+n\sigma_0^2}\right)\right]^2\right\}\right)$$

$$p(\mu|x) \sim N\left(\left(\frac{\sigma_0^2\sum x_i + \mu_0\sigma^2}{\sigma^2+n\sigma_0^2}\right), \left(\frac{n\sigma_0^2+\sigma^2}{\sigma_0^2\sigma^2}\right)^{-1}\right)$$
$$p(\mu|x) \sim N(\mu_n, \sigma_n^2)$$

3.
From above question we know,

$$\sigma_n^2 = \left(\frac{n\sigma_0^2+\sigma^2}{\sigma_0^2\sigma^2}\right)^{-1}$$
$$\frac{1}{\sigma_n^2} = \left(\frac{n\sigma_0^2+\sigma^2}{\sigma_0^2\sigma^2}\right)$$
$$\mu_n = \left(\frac{\sigma_0^2\sum x_i + \mu_0\sigma^2}{\sigma^2+n\sigma_0^2}\right)$$
$$\sum x_i = n\overline{x}$$
$$\mu_n = \left(\frac{\sigma_0^2 n\overline{x} + \mu_0\sigma^2}{\sigma^2+n\sigma_0^2}\right)$$
$$\mu_n = \left(\frac{\sigma_0^2 n\overline{x}}{\sigma^2+n\sigma_0^2} + \frac{\mu_0\sigma^2}{\sigma^2+n\sigma_0^2}\right)$$

Or we can write the same as

$$\mu_n = \left(\left(\frac{\sigma_0^2 n}{\sigma^2 + n\sigma_0^2}\right) * \overline{x} + \left(\frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\right) * \mu_0\right)$$

4: from the above equation we can say that the weight average of prior mean $\mu_0 = \left(\frac{\sigma^2}{\sigma^2+n\sigma_0^2}\right)$ and the weighted average of sample mean $\overline{x} = \left(\frac{\sigma_0^2 n}{\sigma^2+n\sigma_0^2}\right)$

5. we can also write the weights as:

$$weighted\ average\ for\ \mu_0 = \left(\frac{1}{1 + \frac{n\sigma_0^2}{\sigma^2}}\right)$$

$$weighted\ average\ for\ \overline{x} = \left(\frac{1}{\frac{\sigma^2}{n\sigma_0^2} + 1}\right)$$

Therefor we can say that the weights are inversely proportional to their variances.

6.

Summing up the weights from question 4:

$$\left(\frac{\sigma_0^2 n}{\sigma^2 + n\sigma_0^2}\right) + \left(\frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\right) = \frac{(\sigma^2 + n\sigma_0^2)}{(\sigma^2 + n\sigma_0^2)} = 1$$

7.

From question 5 equation we can notice that as the value of n increases$n \to \infty$, the value of $\frac{n\sigma_0^2}{\sigma^2} \to \infty\ and\ \frac{1}{\frac{\sigma^2}{n\sigma_0^2}+1} \to 0,$ and the value of $\mu_0 \to 0\ and\ \overline{x} \to 1$ and vice versa.

Therefore, we can say that value of each weight ranges between 0 and 1.

8. we know from question 3 that:

$$\mu_n = \left(\left(\frac{\sigma_0^2 n}{\sigma^2 + n\sigma_0^2}\right) * \overline{x} + \left(\frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\right) * \mu_0\right)$$

Also, we know that the value of $\left(\frac{\sigma_0^2 n}{\sigma^2+n\sigma_0^2}\right)\ and\ \left(\frac{\sigma^2}{\sigma^2+n\sigma_0^2}\right)$ lies between [0,1]. Therefore, we can say that $\mu_n$will also lie between [0,1]

9. If σ2 is known, then for the new instance x^{new}. The posterior predictive is given by:

$$\begin{aligned} p(x^{new}|X) &= \int p(x^{new}|\mu)p(\mu|\, X)\, d\mu \\ &= \int N(x^{new}|\mu,\sigma^2)\, N(\mu|\mu_n,\sigma_n^2)d\mu \\ &= N(x^{new}|\mu_n,\sigma_n^2+\sigma^2) \end{aligned}$$

An alternative proof to this is mentioned in reference:
https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf

The proof says that:

$$x^{new} = (x^{new} - \mu) + \mu$$
$$\rightarrow x^{new} - \mu \sim N(0, \sigma^2)$$
$$and \;\; \mu \sim N(\mu_n, \sigma_n^2)$$

If X, Y are independent, we know $E[X+Y] = E[X] + E[Y]$ and $Var[X+Y] = Var[X] + Var[Y]$.

$$E[x^{new}] = E[(x^{new} - \mu) + \mu] = E[(x^{new} - \mu)] + E[\mu] = 0 + \mu_n = \mu_n$$
$$Var[x^{new}] = Var[(x^{new} - \mu) + \mu] = Var[(x^{new} - \mu)] + Var[\mu] = \sigma^2 + {\sigma_n}^2$$
$$\rightarrow x^{new} \sim N(\mu_n, \sigma^2 + \sigma_n^2)$$

10.

Given: $p(x) \sim N(6, 1.5^2)$ and $p(\mu) \sim N(4, 0.8^2)$

We know: $p(\mu|x) \sim N(\mu_n, \sigma_n^2)$

Where

$$\mu_n = \left(\frac{\sigma_0^2 n\overline{x}}{\sigma^2 + n\sigma_0^2} + \frac{\mu_0 \sigma^2}{\sigma^2 + n\sigma_0^2}\right) = \left(\frac{0.8^2 * 20 * 6 + 4 * 1.5^2}{1.5^2 + 20 * 0.8^2}\right) = \frac{85.8}{15.5} = 5.701$$

$$\sigma_n^2 = \left(\frac{n\sigma_0^2 + \sigma^2}{\sigma_0^2 \sigma^2}\right)^{-1} = \left(\frac{15.5}{1.44}\right)^{-1} = 0.092 = 0.3^2$$

$$p(\mu|x) \sim N(5.7, 0.092)$$

R Code to plot:

```
x <- seq(-10, 10, by = .1)
y <- dnorm(x, mean = 6, sd = 1.5)
y2 <- dnorm(x, mean = 4, sd = 0.8)
y3 <- dnorm(x, mean = 5.7, sd = 0.3)
plot(0,0,xlim = c(0,10),ylim = c(0,1))
lines(x,y,col='red')
lines(x,y2,col='blue')
lines(x,y3,col='green')
```