HW: Bayesian Parameter Estimation

CSC 591: Algorithms for Data-Guided Business Intelligence

Rachit Shah (rshah25@ncsu.edu)

February 17, 2019

Problem

Data: $X = \{x_t\},\ t = 1, \dots, n$ i.i.d with known variance and unknown mean.

Prior Distribution is Gaussian

$$p(\mu) \sim N(\mu_0, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} * e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$$

Likelihood from sample data which is Gaussian.

$$p(X_t|\mu) \sim N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x_t-\mu)^2}{2\sigma^2}}$$

Q1: Derive the formula for the posterior distribution of μ

$$Posterior\ p(\mu|X) \propto p(\mu)\ (Prior) * p(X_t|\mu)\ (Likelihood)$$

$$= p(\mu) * p(x_1|\mu) * \dots * p(x_n|\mu)$$

$$= p(\mu) * \prod_{t=1}^{n} p(x_t|\mu)$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} * e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} * \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x_t-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sqrt{2\pi}^{n+1} * \sqrt{\sigma_0^2 * \sigma^{2n}}} * e^{\frac{-\mu^2+2\mu\mu_0-\mu_0^2}{2\sigma_0^2} - \sum_{t=1}^{n} \frac{x_t^2-2\mu x_t+\mu^2}{2\sigma^2}}$$

$$\propto\ e^{\frac{-\mu^2\sigma^2+2\mu\mu_0\sigma^2-\mu_0^2\sigma^2-\sigma_0^2\sum x_t^2+2\mu\sigma_0^2\sum x_t-n\mu^2\sigma_0^2}{2\sigma_0^2\sigma^2}}$$

$$=\ e^{\frac{-\mu^2(\sigma^2+n\sigma_0^2)+2\mu(\mu_0\sigma^2+\sigma_0^2\sum x_t)-(\mu_0^2\sigma^2+\sigma_0^2\sum x_t^2)}{2\sigma_0^2\sigma^2}}$$

$$p(\mu|X)$$

$$= e^{\frac{-\mu^2+\frac{2\mu\left(\mu_0\sigma^2+\sigma_0^2\sum x_t\right)}{\sigma^2+n\sigma_0^2}-\left(\frac{\mu_0\sigma^2+\sigma_0^2\sum x_t}{\sigma^2+n\sigma_0^2}\right)^2+\left(\frac{\mu_0\sigma^2+\sigma_0^2\sum x_t}{\sigma^2+n\sigma_0^2}\right)^2}{\frac{2\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}}$$

$$* e^{-\frac{\left(\mu_0^2\sigma^2+\sigma_0^2\sum x_t^2\right)}{2\sigma_0^2\sigma^2}}$$

$$= e^{\frac{-\mu^2+\frac{2\mu\left(\mu_0\sigma^2+\sigma_0^2\sum x_t\right)}{\sigma^2+n\sigma_0^2}-\left(\frac{\mu_0\sigma^2+\sigma_0^2\sum x_t}{\sigma^2+n\sigma_0^2}\right)^2}{\frac{2\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}} * e^{\frac{\left(\frac{\mu_0\sigma^2+\sigma_0^2\sum x_t}{\sigma^2+n\sigma_0^2}\right)^2}{\frac{2\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}}$$

$$* e^{-\frac{\left(\mu_0^2\sigma^2+\sigma_0^2\sum x_t^2\right)}{2\sigma_0^2\sigma^2}}$$

$$\propto e^{\frac{-\mu^2+\frac{2\mu\left(\mu_0\sigma^2+\sigma_0^2\sum x_t\right)}{\sigma^2+n\sigma_0^2}-\left(\frac{\mu_0\sigma^2+\sigma_0^2\sum x_t}{\sigma^2+n\sigma_0^2}\right)^2}{\frac{2\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}}$$

$$\therefore \boxed{p(\mu|X) \propto e^{\frac{-\left(\mu-\left(\frac{\mu_0\sigma^2+\sigma_0^2\sum x_t}{\sigma^2+n\sigma_0^2}\right)\right)^2}{\frac{2\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}}} \quad ----(1)$$

Q2: Show that the posterior distribution is the Gaussian, p(μ|X)~N(μn,σ2n)

From equation 1, we can see that the posterior distribution is a Gaussian distribution as it is proportional to a Normal Distribution of the form

$$\boxed{\begin{gathered} e^{-\frac{(\mu-\mu_n)^2}{2\sigma_n^2}} \sim N(\mu_n, \sigma_n^2) \\ Where, \\ \sigma_n = \frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2} \ and \ \mu_n = \frac{\mu_0\sigma^2+\sigma_0^2\sum x_t}{\sigma^2+n\sigma_0^2} \end{gathered}}$$

Q3: Show the derivation and the final estimate for μn and 1/σ2n

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

$$\therefore \boxed{\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} ----(2)$$

$$\mu_n = \frac{\mu_0\sigma^2 + \sigma_0^2 \sum x_t}{\sigma^2 + n\sigma_0^2}$$

$$= \frac{\frac{\mu_0\sigma^2}{\sigma_0^2\sigma^2} + \frac{\sigma_0^2 \sum x_t}{\sigma_0^2\sigma^2}}{\frac{\sigma^2 + n\sigma_0^2}{\sigma_0^2\sigma^2}} (Divide\ whole\ eq\ by\ \sigma_0^2\sigma^2)$$

$$= \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_t}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

$$= \sigma_n^2 * \left(\frac{\mu_0}{\sigma_0^2} + \frac{\frac{\sum x_t}{n}}{\frac{\sigma^2}{n}}\right) (From\ eq.2)$$

$$\therefore \boxed{\mu_n = \sigma_n^2 * \left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\frac{\sigma^2}{n}}\right)} ----(3)$$

Q4: If the mean of the posterior density (which is the MAP estimate), μn is written as the weighted average of the prior mean, μ0, and the sample (likelihood) mean, X¯, then what are the formulas for the weights?

$$\mu_n = \sigma_n^2 * \left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\frac{\sigma^2}{n}}\right) = \left(\frac{\sigma_n^2}{\sigma_0^2} * \mu_0\right) + \left(\frac{n\sigma_n^2}{\sigma^2} * \bar{x}\right) = w_0 * \mu_0 + w_1 * \bar{x}$$

Therefore, the weights are:

$$\boxed{w_0 = \frac{\sigma_n^2}{\sigma_0^2} \; and \; w_1 = \frac{n\sigma_n^2}{\sigma^2} \quad ----(4)}$$

Q5: Are the weights in Question #4 directly or inversely proportional to their variances (justify)?

As you can see from equation 4, weight w_0 corresponds to the weighted average of the prior mean and its equation has the variance σ_0^2 in the denominator. While weight w_1 is the weighted average of the likelihood and its equation also has the corresponding variance σ^2 in the denominator. Hence, both weights are **inversely proportional** to their variances.

Q6: Do the weights in Questions #4 sum up to 1 (justify)?

Yes.

$$w_0 + w_1 = \frac{\sigma_n^2}{\sigma_0^2} + \frac{n\sigma_n^2}{\sigma^2}$$

$$= \frac{1}{\sigma_0^2 * \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)} + \frac{n}{\sigma^2 * \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)} \; (Substitute\; value\; \sigma_n^2\; from\; eq.2)$$

$$= \frac{1}{1 + \frac{n\sigma_0^2}{\sigma^2}} + \frac{n}{\frac{\sigma^2}{\sigma_0^2} + n}$$

$$= \frac{\sigma^2}{\sigma^2 + n\sigma_0^2} + \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2} \quad ----(5)$$

$$= \frac{\sigma^2 + n\sigma_0^2}{\sigma^2 + n\sigma_0^2} = 1$$

$$\boxed{\therefore w_0 + w_1 = 1}$$

Q7: Is each weight between zero and one (justify)?

From equation 5 in the previous equation, we can see that $w_0 = \frac{\sigma^2}{\sigma^2 + n\sigma_0^2} \; and \; w_1 = \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2}$.

Weight w_0 has σ^2 in both numerator and denominator but has an additional $n\sigma_0^2$ in the denominator. Since variance can't be negative and n too can't be negative, the denominator is always greater than or equal to the numerator. When n becomes too large, the denominator will approach infinity and w_0 will get closer to 0. While if n grows closer to 0, w_0 will approach 1. Hence, w_0 always has its value between 0 and 1.

Weight w_1 has $n\sigma_0^2$ in both numerator and denominator but has an additional σ^2 in the denominator. Since variances can't be negative and n too can't be negative, the denominator is always greater than or equal to the numerator. When n is very large such that $n\sigma_0^2 \gg \sigma^2$, the value of w_1 will approach 1. However, for low values of n, if value of σ^2 becomes greater than $n\sigma_0^2$, it will approach 0 but never become negative. Hence, w_1 always has its value between 0 and 1.

Hence, **Yes**, both weights always have value between 0 and 1.

Q8: Given your answers for Questions #4-7, what can you say about the value of μn w.r.t. the values of μ0 and X¯

From question 7, we see that as n (samples) increases, w_0 reaches 0 while w_1 reaches 1. And vice versa. Also, $\mu_n = w_0\mu_0 + w_1\bar{x}$. Since, both weights have their value between 0 and 1, theoretically the maximum value μ_n can reach is $\mu_0 + \bar{x}$ which will happen when both weights are 1 or 0 if both weights are 0. But, since both weights are inversely dependent on n, they will never both be 1 or 0. Hence, the value of μ_n lies anywhere between the values of μ_0 and $\bar{x}$.

Q9: If σ2 is known, then for the new instance xnew, show that p(xnew|X)~N(μn,σ2n+σ2)

$$p(x_{new}|X) = \int p(x_{new}|\mu) * p(\mu|X)\, d\mu$$

$p(x_i|\mu)$ is given to be a normal distribution $N(\mu, \sigma^2)$ and we found the posterior distribution $p(\mu|X)$ to also be a normal distribution $N(\mu_n, \sigma_n^2)$.

Now, $x_{new} = (x_{new} - \mu) + \mu$, where, $x_{new} - \mu$ is the normal distribution $N(0, \sigma^2)$ and μ is the posterior normal distribution $N(\mu_n, \sigma_n^2)$. Hence, we have reframed the equation to be a sum of the 2 normal (gaussian) distributions.

Theorem: If $X_1, X_2, \ldots, X_n$ are mutually independent normal random variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances σ21,σ22,⋯,σ2nσ12,σ22,⋯,σn2, then the linear combination: $Y = \sum_{i=1}^{n} c_i X_i$ follows the normal distribution:

$$N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$$

Since, the problem statement states that the data is i.i.d and follows gaussian distribution, the condition of the theorem is satisfied, and we can say that the sum of the 2 normal distributions is also a normal distribution.

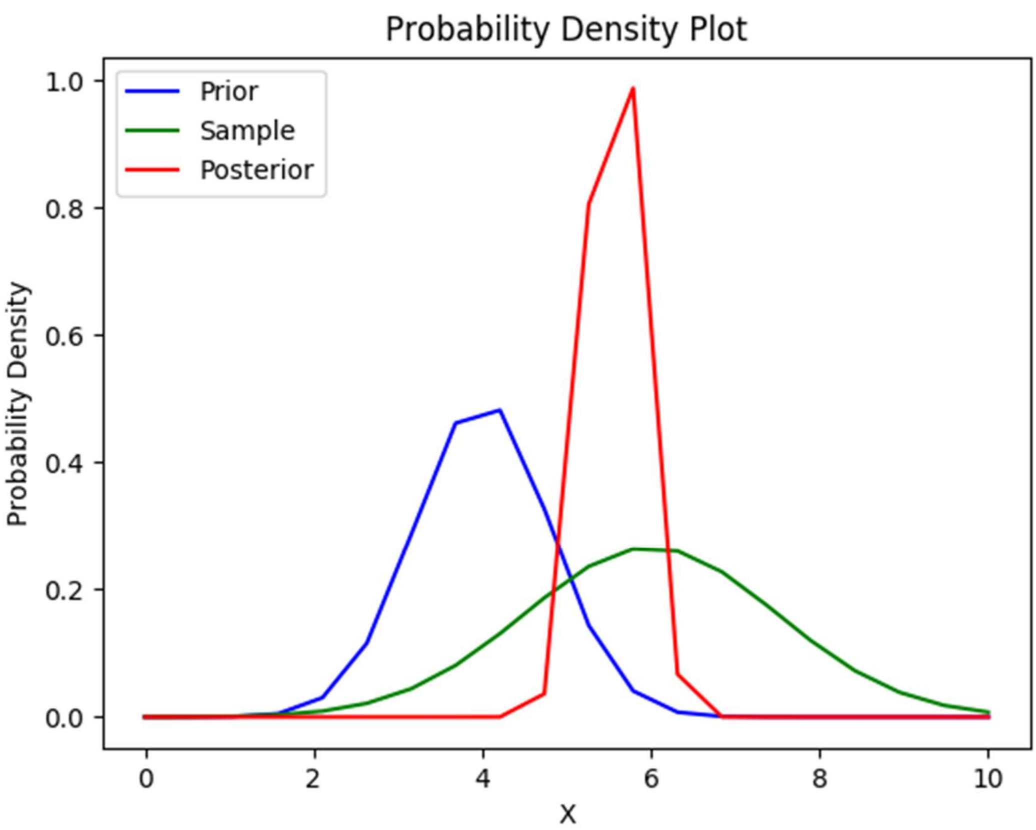
$$\boxed{x_{new} = (x_{new} - \mu) + \mu \sim N(0, \sigma^2) + N(\mu_n, \sigma_n^2) \sim N(\mu_n, \sigma^2 + \sigma_n^2)}$$

Q10: Generate a plot that displays p(x)~N(6,1.52), prior p(μ)~N(4,0.82), and posterior p(μ|X)~N(μn,σ2n) for n=20 sample points. What are the values for μn and σ2n?

Answer:

After running the code below, the output is

('Mean of the posterior distibution is', 5.56336966620394, 'and variance is', 0.09568106312292358)



Code: (Python3)

```
import numpy as np
from matplotlib import pyplot as plt
import scipy.stats as st

samples = 20
x = np.linspace(0, 10, samples)

##Prior
mean_0 = 4
sd_0 = 0.8
prior_dist = st.norm(mean_0, sd_0).pdf(x)

##Sample
```

```
mean_x = 6
sd_x = 1.5
sample_dist = dist = st.norm(mean_x, sd_x).pdf(x)

##Posterior
x_t = st.norm(mean_x,sd_x).rvs(samples)
var_n = 1/((1/sd_0**2)+(samples/sd_x**2))
mean_n = var_n * ((mean_0/sd_0**2)+(np.mean(x_t)*samples/sd_x**2))
print("Mean of the posterior distibution is",mean_n,"and variance is",var_n)
posterior_dist = st.norm(mean_n, np.sqrt(var_n)).pdf(x)

##Plot
plt.plot(x,prior_dist,"b-",label='Prior')
plt.plot(x,sample_dist,"g-",label='Sample')
plt.plot(x,posterior_dist,"r-",label='Posterior')
plt.legend(loc='upper left')
plt.title('Probability Density Plot')
plt.ylabel('Probability Density')
plt.xlabel('X')
plt.show()
```