Name: Alisha Shahane
StudentID: 200311941
UnityID: asshahan

CSC 591 ADBI

HW Solution (Bayesian Parameter Estimation)

1) Derive formula for posterior distribution of μ

Prior distribution of μ is: $p(\mu) \sim N\left(\mu_0, \sigma_0^2\right)$ (given)

$\therefore\ p(\mu) = e^{\left[\frac{1}{-2\sigma_0^2}(\mu - \mu_0)^2\right]}$

Samples are drawn from a normal distribution $N(\mu, \sigma^2)$ whose mean is not known but variance is known.

We can say the likelihood of samples having mean μ is, $p(\mu \mid X) \sim p(X \mid \mu) * p(\mu)$

$p(X \mid \mu) = p(x_1 \mid \mu) * p(x_2 \mid \mu) * \ldots .\, p(x_n \mid \mu)$

$\therefore\ p(X \mid \mu) \sim \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\left(\frac{-1}{2\sigma^2}\right)\sum_i (x_i - \mu)^2}$

Since the variance is known, $\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n$ is a constant. This is why we can ignore it from further derivation.

$\therefore\ p(X \mid \mu) \sim e^{\left(\frac{-1}{2\sigma^2}\right)\sum_i (x_i - \mu)^2}$

Substituting this in $p(\mu \mid X) \sim p(X \mid \mu) * p(\mu)$ we get,

$$p(\mu \mid X) \sim e^{\left[\frac{-1}{2\sigma^2}\sum_i (x_i - \mu)^2\right]} * e^{\left[\frac{-1}{2\sigma_0^2}(\mu - \mu_0)^2\right]}$$

$$p(\mu \mid X) \sim e^{\left[\frac{-1}{2\sigma^2}\sum_i (x_i^2 - 2x_i\mu + \mu^2)\right]} * e^{\left[\frac{-1}{2\sigma_0^2}(\mu^2 - 2\mu\mu_0 + \mu_0^2)\right]}$$

$$p(\mu \mid X) \sim e^{\left[\frac{-1}{2\sigma^2}\sum_i (x_i^2 - 2x_i\mu + \mu^2) + \frac{-1}{2\sigma_0^2}(\mu^2 - 2\mu\mu_0 + \mu_0^2)\right]}$$

$$p(\mu \mid X) \sim \exp\left[-\frac{1}{2}\left(\frac{\sum_i x_i^2}{\sigma^2} - 2\mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_i x_i}{\sigma^2}\right) + \mu^2\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right) + \frac{\mu_0^2}{\sigma_0^2}\right)\right]$$

Ignoring constant terms, we can write $p(\mu \mid X)$ as,

$$\boxed{p(\mu \mid X) \sim \exp\left[-\frac{1}{2}\left(-2\mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_i x_i}{\sigma^2}\right) + \mu^2\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)\right)\right]}$$

2) Show that the posterior distribution is the Gaussian, $p(\mu \mid X) \sim N(\mu_n, \sigma_n^2)$

We can show this by simplifying equation obtained in problem 1.

We need to convert it to a form which is comparable to following form,

$$p(\mu \mid X) \sim e^{\left(\left(\frac{-1}{2\sigma_n^2}\right)(\mu - \mu_n)^2\right)}$$

Comparing coefficients in (a) with this we get –

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

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

After simplifying we get,

$$\boxed{\begin{aligned} \sigma_n^2 &= \left(\frac{\sigma^2\sigma_0^2}{\sigma^2+n\sigma_0^2}\right) \\ \mu_n &= \left(\frac{\mu_0\sigma^2+\sigma_0^2\sum_i x_i}{n\sigma_0^2+\sigma^2}\right) \end{aligned}} \quad \text{and,}$$

3) **Show the derivation and the final estimate for μ_n and $1/\sigma_n^2$**

From problem 2 we have,

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

Therefore,

$$\boxed{\begin{aligned} 1/\sigma_n^2 &= \left(\frac{\sigma^2+n\sigma_0^2\,\sigma^2\sigma_0^2}{\sigma^2\sigma_0^2}\right) \\ \mu_n &= \left(\frac{\mu_0\sigma^2+\sigma_0^2\sum_i x_i}{n\sigma_0^2+\sigma^2}\right) \end{aligned}}$$

4) **If the mean of the posterior density (which is the MAP estimate), μ_n 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?**

From problem 3 we have formula for mean of posterior distribution which is,

$$\mu_n = \left(\frac{\mu_0\sigma^2 + \sigma_0^2\sum_i x_i}{n\sigma_0^2+\sigma^2}\right)$$

In this, $\sum_i x_i$ can be replaced by $n\overline{x}$.

Therefore,

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

From this we can get weights w_0 and w_1 for μ_0 and $\overline{x}$ respectively,

$$\boxed{\begin{aligned} w_0 &= \frac{\sigma^2}{n\sigma_0^2+\sigma^2} \\ w_1 &= \frac{\sigma_0^2 n}{n\sigma_0^2+\sigma^2} \end{aligned}}$$

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

$w_0 \propto \sigma^2$ and $w_0 \propto 1/{\sigma_0}^2$

Weight of prior mean is directly proportional to sample variance, and inversely proportional to prior variance.

$w_1 \propto \sigma_0^2$ and $w_1 \propto 1/\sigma^2$

Weight of sample mean is directly proportional to prior variance, and inversely proportional to sample variance.

Therefore, the weights are inversely proportional to their variances.

6) Do the weights in Question #4sum up to 1 (justify)?

From problem 4,

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

$$\therefore \quad w_0 + w_1 = \frac{n\sigma_0^2 + \sigma^2}{n\sigma_0^2 + \sigma^2} = 1$$

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

The weights sum up to 1.

7) Is each weight between 0 and 1 (justify)?

$$w_0 = \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} = \frac{1}{\frac{n\sigma_0^2}{\sigma^2} + 1}$$

If n = 0, then w_0 = 1 and as n increases w_0 tends to zero 0.

$$w_1 = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} = \frac{1}{1 + \frac{\sigma^2}{n\sigma_0^2}}$$

Similarly, if $\sigma^2 = 0$, then w_1 = 1 and as σ^2 increases w_1 tends to 0.

Hence, both w_1 and w_2 are between 0 and 1.

8) What can you say about μ_n w.r.t values of μ_0 and $\bar{X}$.

Since the weights sum up to 1, both cannot be zero at a time. In corner case, one if the weights in 1 and other is 0. Thus, range μ_n will be,

μ_n lies between $[\min(\mu_0, \bar{X}), \max(\mu_0, \bar{X})]$

9) If σ^2 is known, then for the new instance x^{new}, show that $p(x^{new} \mid X) \sim N(\mu_n, \sigma_n^2+\sigma^2)$

If variance is known,

$$p(x^{new} \mid X) = \int_{-\infty}^{\infty} p(x^{new} \mid \mu) p(\mu, X) d\mu$$

Substituting values from problem 1 and 2,

$$p(x^{new} \mid X) = \frac{1}{\sigma^2 \sigma_n^2 \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(\left(\frac{-1}{2\sigma^2}\right)(x^{new} - \mu)^2\right)} e^{\left(\left(\frac{-1}{2\sigma_n^2}\right)(\mu - \mu_n)^2\right)} d\mu$$

After simplifying the equation,

$$p(x^{new} \mid X) = \frac{1}{(\sigma^2 + \sigma_n^2)\sqrt{2\pi}} e^{\left(\left(\frac{-1}{2}\right)\frac{(x^{new} - \mu_n)^2}{\sigma^2 + \sigma_n^2}\right)}$$

Comparing with equation for Gaussian function,

$\sigma = \sigma^2 + \sigma_n^2$ and $\mu = \mu_n$

Hence,

$$\boxed{p(x^{new} \mid X) \sim N(\mu_n, \sigma^2 + \sigma_n^2)}$$

10) Generate a plot that displays $p(x) \sim N(6, 1.5^2)$, prior $p(\mu) \sim N(4, 0.8^2)$ and posterior $p(\mu \mid X) \sim N(\mu_n, \sigma_n^2)$ for n=20 sample points. What are the values for μ_n and σ_n^2?

Given,

$n = 20$

$\overline{x} = 6$ (likelihood)

$\sigma^2 = 1.5^2$ (likelihood)

$\mu_0 = 4$ (prior)

$\sigma_0^2 = 0.8^2$ (prior)

From question 2 we have,

$$\sigma_n^2 = \left(\frac{\sigma^2\sigma_0^2}{\sigma^2 + n\sigma_0^2}\right) \quad \text{and} \quad \mu_n = \left(\frac{\mu_0\sigma^2 + \sigma_0^2 \sum_i x_i}{n\sigma_0^2 + \sigma^2}\right)$$

Substituting these values,

$$\sigma_n^2 = \left(\frac{\sigma^2\sigma_0^2}{\sigma^2 + n\sigma_0^2}\right) = \left(\frac{2.25 * 0.64}{2.25 + 12.8}\right) = \left(\frac{1.44}{15.05}\right) = 0.0957$$

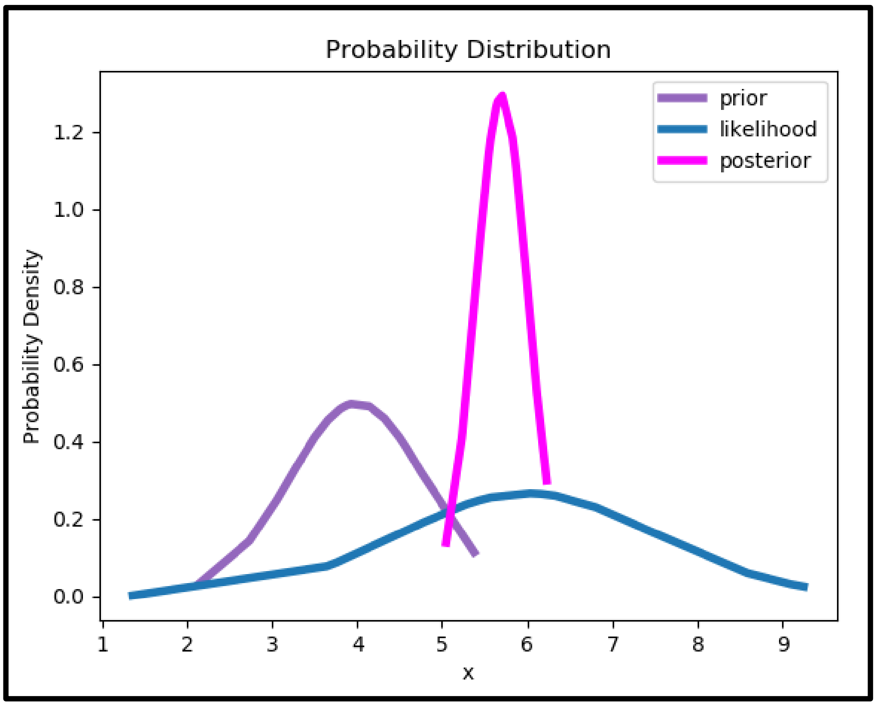
$$\mu_n = \left(\frac{\mu_0\sigma^2 + \sigma_0^2 n\overline{x}}{n\sigma_0^2 + \sigma^2}\right) = \left(\frac{(4 * 2.25) + (0.64 * 120)}{2.25 + 12.8}\right) = \frac{85.8}{15.05} = 5.7$$

Posterior = Prior * likelihood

Therefore, $p(\mu \mid X) \sim N(\mu_n, \sigma_n^2) \sim N(6, 1.5^2) * N(4, 0.8^2)$

$p(\mu \mid X) \sim N(5.7, 0.0957)$

Plot: (Python script attched named 'asshahan_hw_q10.py')



References:
1] “Conjugate Bayesian analysis of the Gaussian distribution” by Kevin P Murphy, https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf
2] “Sum of normally distributed random variables”, https://en.wikipedia.org/wiki/Sum_of_normally_distributed_random_variables
3] “Posterior Predictive Distribution”, http://people.stat.sc.edu/Hitchcock/stat535slidesday18.pdf