Solution: Bayesian Estimation of the Parameters of a Gaussian Distribution

Assumptions:

- **Univariate Case: The data $X = \{x_t\}$,*t*=1,...,*n* is the univariate data, with the i.i.d. samples.**
- **Gaussian (Normal) Distribution: The sample is drawn from the Gaussian (Normal) distribution, $p(x) \sim N(\mu, \sigma^2)$, with parameters μ and σ^2.**
- **Parameters: Unknown mean, known variance**
- **Priors: The conjugate prior for μ is Gaussian, $p(\mu) \sim N(\mu_0, \sigma_0^2)$**

Assignment:

1. **Derive the formula for the posterior distribution of μ**
2. **Show that the posterior distribution is the Gaussian, $p(\mu|X) \sim N(\mu_n, \sigma_n^2)$**
   <u>Answer:</u> Combined answer for Q1 and Q2.

$$p(\mu|X) \propto p(X|\mu)p(\mu)$$

$$p(\mu) \propto exp\left[\frac{1}{-2\sigma_0^2}(\mu - \mu_0)^2\right]$$

$$p(X|\mu) \propto exp\left[\frac{1}{-2\sigma^2}\sum_i (x_i - \mu)^2\right]$$

$$p(\mu|X) \propto \; exp\left[\frac{1}{-2\sigma_0^2}(\mu - \mu_0)^2\right] exp\left[\frac{1}{-2\sigma^2}\sum_i (x_i - \mu)^2\right]$$

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

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

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

In the above equation, we can observe that posterior distribution of μ is quadratic in the exponent. Hence, it comes from some Gaussian distribution with mean μ_n and variance σ_n^2.

3. **Show the derivation and the final estimate for μ_n and $1/\sigma_n^2$**
   <u>Answer:</u>
   A Gaussian distribution with mean μ_n and variance σ_n^2 is given by,

$$p(\mu|\mu_n) \propto exp\left[\frac{1}{-2\sigma_n^2}(\mu - \mu_n)^2\right] = exp\left[\frac{1}{-2\sigma_n^2}(\mu^2 - 2\mu\mu_n + \mu_n^2)\right]$$

Comparing the terms of μ^2 and $-2\mu\mu_n$ from the previous question we get,

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

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

In the above equation, we can rewrite $\sum_i x_i = n\bar{x}$.

4. **If the mean of the posterior density (which is the MAP estimate), $\boldsymbol{\mu_n}$ is written as the weighted average of the prior mean, $\boldsymbol{\mu_0}$, and the sample (likelihood) mean, $\boldsymbol{\bar{x}}$, then what are the formulas for the weights?**
Answer: From the previous question, we can identify the weights for the prior mean, μ_0, w_0, and the sample (likelihood) mean $\bar{x}$, w_1.
$$w_0 = \frac{\sigma^2}{\sigma^2 + n\sigma_0^2}$$
$$w_1 = \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2}$$

5. **Are the weights in Question #4 directly or inversely proportional to their variances (justify)?**
Answer: As shown in the previous answer, the weights are inversely proportional to their variances.

6. **Do the weights in Questions #4 sum up to 1 (justify)?**
Answer: We take a sum of the weights, w_0 and w_1,
$$w_0 + w_1 = \frac{\sigma^2}{\sigma^2 + n\sigma_0^2} + \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2} = \frac{\sigma^2 + n\sigma_0^2}{\sigma^2 + n\sigma_0^2} = 1$$

Yes, they sum to 1.

7. **Is each weight between zero and one (justify)?**
Answer: As per the formulas for the weights,

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

The values of the weights, w_0 and w_1 will always range between 0 and 1. w_0 will approach 1 as the number of samples gets close to 0 and it will approach 0 as the number of samples increases. Similarly, w_1 will approach 1 as the variance σ^2 decreases and it will approach close to 0 as the number of samples increases.

8. **Given your answers for Questions #4-7, what can you say about the value of $\boldsymbol{\mu_n}$ w.r.t. the values of $\boldsymbol{\mu_0}$ and $\boldsymbol{\bar{x}}$.**
Answer: The value of μ_n will always lie between μ_0 and $\bar{x}$ since the weights w_0 and w_1 will range between 0 and 1.

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

Answer:

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

An alternative proof is as follows,

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

Thus, we can show that $p(x^{new}|X) \sim N(\mu_n, \sigma_n^2 + \sigma^2)$.

Reference: Murphy, Kevin P. "Conjugate Bayesian analysis of the Gaussian distribution." *def* 1, no. 2σ2 (2007): 16.

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 $p(\mu|X) \sim N(\mu_n, \sigma_n^2)$ for *n*=20 sample points. What are the values for μ_n and σ_n^2?**

Answer:

R code:

```
n <- 20

x <- seq(0, 10, length.out=n)

mu_x <- 6; var_x <- 1.5^2

mu_0 <- 4; var_0 <- 0.8^2

# Generate samples from p(x) and prior p(mu) distributions

sample_dist <- dnorm(x, mean=mu_x, sd=sqrt(var_x))

prior_dist <- dnorm(x, mean=mu_0, sd=sqrt(var_0))

# Calculate the mu_n and var_n based on the formulas in Q3

x_i <- rnorm(n, mean=mu_x, sd=sqrt(var_x))

var_n <- (var_x*var_0)/(var_x + n*var_0)

t1 <- (var_x*mu_0)/(var_x+n*var_0)

t2 <- (n*mean(x_i)*var_0)/(var_x+n*var_0)

mu_n <- t1 + t2

# Generate samples from the posterior distribution

posterior_dist <- dnorm(x, mean=mu_n, sd=sqrt(var_n))

plot(0, 0, xlim=c(0,10), ylim=c(0,1), main="Probability Density Plot", xlab="X", ylab="Probability Density")
```

```
lines(x, sample_dist, col='red')
lines(x, prior_dist, col='green')
lines(x, posterior_dist, col='blue')
```