MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.s077 Spring 2018
Recitation 3 Friday 2/23

Problem: Let X_1, X_2, W_1, W_2 be random variables such that $X_1 = \theta + W_1$, $X_2 = \theta + W_2$, $W_1 \sim \mathcal{N}(0, 1)$, $W_2 \sim \mathcal{N}(0, 1/4)$ where $\theta \in \mathbb{R}$ is an unknown parameter.

(a) Find the ML estimator $\hat{\Theta}_{ML}$ for θ in terms of X_1 and X_2.

(b) Is $\hat{\Theta}_{ML}$ unbiased?

(c) Calculate the variance of $\hat{\Theta}_{ML}$.

(d) Suppose we have a linear estimator $\hat{\Theta}_L = aX_1 + bX_2$. Find a and b so that $\hat{\Theta}_L$ is unbiased and has minimum variance.

Now suppose θ is drawn from the distribution $\Theta \sim \mathcal{N}(5, 1)$ i.e. $X_1 = \Theta + W_1$, $X_2 = \Theta + W_2$.

(e) Find the posterior distribution $p(\theta|x_1, x_2)$.

(f) Find the peak of the posterior distribution in terms of x_1 and x_2.

(g) Find the variance of the posterior distribution. Is it affected by the values of x_1 and x_2? Compare it to the variance of the prior distribution.

Now suppose we had two additional observations $Y_1 = \Theta + W_1$, $Y_2 = \Theta + W_2$.

(h) Find the posterior distribution $p(\theta|x_1, x_2, y_1, y_2)$.

(i) Find the peak of the posterior distribution in terms of x_1, x_2, y_1, y_2.

(j) Find the variance of the posterior distribution. Compare it to the variance of the previous posterior distribution with two observations.

Solution:

(a) The log-likelihood function (removing additive constants) is

$$l(\theta; x_1, x_2) = -\frac{(x_1 - \theta)^2}{2} - 2(x_2 - \theta)^2$$

We set the derivative to 0 to get the maximum likelihood estimate.

$$\begin{aligned} \frac{dl}{d\theta} &= (x_1 - \theta) + 4(x_2 - \theta) = 0 \\ \Rightarrow \theta_{ML} &= \frac{x_1 + 4x_2}{5} \end{aligned}$$

so the estimator is $\hat{\Theta}_{ML} = \frac{X_1 + 4X_2}{5}$.

(b) Yes.

$$\mathbb{E}[\hat{\Theta}_{ML}] = \frac{\mathbb{E}[X_1] + 4\mathbb{E}[X_2]}{5} = \theta$$

(c)

$$\mathrm{Var}[\hat{\Theta}_{ML}] = \frac{\mathrm{Var}[X_1] + 16\mathrm{Var}[X_2]}{25} = \frac{1}{5}$$

(d) We have $\mathbb{E}[\hat{\Theta}_L] = (a+b)\theta \Rightarrow a + b = 1$ and $\mathrm{Var}[\hat{\Theta}_L] = a^2 + \frac{b^2}{4}$. We therefore want to minimize $a^2 + \frac{(1-a)^2}{4} \Rightarrow 2a - \frac{1-a}{2} = 0 \Rightarrow a = \frac{1}{5}, b = \frac{4}{5}$. Therefore $\hat{\Theta}_L = \hat{\Theta}_{ML}$.

(e) From Bayes' theorem

$$\begin{aligned} p(\theta | x_1, x_2) &= \frac{p(x_1, x_2 | \theta) \cdot p(\theta)}{p(x_1, x_2)} \\ &\propto \exp\left(-\frac{(x_1 - \theta)^2}{2} - 2(x_2 - \theta)^2 - \frac{(\theta - 5)^2}{2}\right) \\ &\propto \exp\left(-3\theta^2 + (x_1 + 4x_2 + 5)\theta\right) \\ &\propto \exp\left(-3\left(\theta - \frac{x_1 + 4x_2 + 5}{6}\right)^2\right) \\ \Rightarrow p(\theta | x_1, x_2) &= \mathcal{N}\left(\frac{x_1 + 4x_2 + 5}{6}, \frac{1}{6}\right) \end{aligned}$$

(f) Since the posterior is a normal distribution, the peak is its mean or $\frac{x_1 + 4x_2 + 5}{6}$. Note that by symmetry of the normal distribution this is also $\mathbb{E}[\Theta | x_1, x_2]$.

(g) The variance can also be easily read off and is $\frac{1}{6}$. We note that this does not depend on x_1 nor x_2 and is smaller than the variance of the prior.

(h) Similar computation yields

$$\begin{aligned}
p(\theta|x_1, x_2, y_1, y_2) &\propto \exp\left(-\frac{11}{2}\theta^2 + (x_1 + 4x_2 + y_1 + 4y_2 + 5)\theta\right) \\
&\propto \exp\left(-\frac{11}{2}\left(\theta - \frac{x_1 + 4x_2 + y_1 + 4y_2 + 5}{11}\right)^2\right) \\
\Rightarrow p(\theta|x_1, x_2, y_1, y_2) &= \mathcal{N}\left(\frac{x_1 + 4x_2 + y_1 + 4y_2 + 5}{11}, \frac{1}{11}\right)
\end{aligned}$$

(i) The peak is $\frac{x_1+4x_2+y_1+4y_2+5}{11}$.

(j) The variance is $\frac{1}{11}$ and is smaller than that for the posterior after two observations.