Department of Industrial Engineering & Operations Research

IEOR 165 (Spring 2016)

Homework 3

Due: Friday, Mar 11

Question 1. Assume we have one observation X drawn from a normal distribution with *unknown* mean μ and *known* σ^2. And μ itself follows a normal distribution with *known* mean θ and *known* variance β^2 (prior). Basically, we have $X|\mu \sim N(\mu, \sigma^2)$ and $\mu \sim N(\theta, \beta^2)$. Show that the posterior distribution is normal and find its mean and variance ($E(\mu|X)$ and $Var(\mu|X)$)
Hint 1: you need to show that

$$f(\mu|x) \propto \exp\{-\frac{1}{2Var(\mu|X)}(\mu - E(\mu|X))^2\}$$

Hint 2: you may want to use the following formula

$$\beta^2(x-\mu)^2 + \sigma^2(\mu-\theta)^2 = (\mu\sqrt{\beta^2+\sigma^2} - \frac{\sigma^2\theta + \beta^2 x}{\sqrt{\beta^2+\sigma^2}})^2 - \frac{(\sigma^2\theta+\beta^2 x)^2}{\beta^2+\sigma^2} + \beta^2 x^2 + \sigma^2\theta^2$$

Solution. By the Bayes' rule, we have

$$\begin{aligned}
f(\mu|X) &\propto f(X|\mu)g(\mu) \\
&\propto \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(X-\mu)^2}{2\sigma^2})\frac{1}{\sqrt{2\pi\beta^2}}\exp(-\frac{(\mu-\theta)^2}{2\beta^2}) \\
&\propto \exp(-\frac{(X-\mu)^2}{2\sigma^2})\exp(-\frac{(\mu-\theta)^2}{2\beta^2}) \\
&\propto \exp(-\frac{\beta^2(X-\mu)^2 + \sigma^2(\mu-\theta)^2}{2\sigma^2\beta^2}) \\
&\propto \exp\Big\{-\frac{(\mu\sqrt{\beta^2+\sigma^2} - \frac{\sigma^2\theta+\beta^2 X}{\sqrt{\beta^2+\sigma^2}})^2 + \text{constant}}{2\sigma^2\beta^2}\Big\} \qquad \text{(use Hint 2)} \\
&\propto \exp\Big\{-\frac{(\beta^2+\sigma^2)(\mu - \frac{\sigma^2\theta+\beta^2 X}{\beta^2+\sigma^2})^2}{2\sigma^2\beta^2}\Big\}
\end{aligned}$$

Then let

$$Var(\mu|X) = \frac{\sigma^2\beta^2}{\sigma^2+\beta^2} \quad E(\mu|X) = \frac{\sigma^2\theta+\beta^2 X}{\beta^2+\sigma^2}$$

From Hint 1, we know that the posterior distribution is normal.

Question 2. Find the Maximum A Posterior estimation of μ in Question 1.

Solution. From Question 1, we know that the posterior distribution is normal. And we also know that the normal distribution achieves its maximum density at its mean (this maximum density is also called mode). So the MAP estimate of μ is $\frac{\sigma^2\theta+\beta^2 X}{\beta^2+\sigma^2}$

Question 3. River floods are often measured by their discharges (in units of feet cubed per second). The following table gives the flood discharges of the Blackstone River in Woonsocket, Rhode Island, in each of the years from 1956 to 1965.

Year	Flood Discharge (ft^3/s)
1956	8710
1957	3850
1958	4970
1959	5398
1960	4780
1961	4020
1962	5790
1963	4510
1964	5520
1965	5300

Take the yearly flood discharge as a random variable X. Use the above data to estimate the cdf $F_X(5000)$ using empirical distribution function.

Solution. From the definition,

$$\hat{F}_X(5000) = \frac{1}{10}\sum_{i=1}^{10} \mathbf{1}(X_i \leq 5000) = 0.5.$$

Question 4. In the context of Question 3, estimate the density at 5000, i.e. $f_X(5000)$, using histogram method with bin edges $\{0, 2000, 3000, 4000, 4500, 5000, 5500, 6000, 10000\}$.

Solution. From the definition,

$$\hat{f}_X^{his}(5000) = \frac{C_j}{10(b_j - b_{j-1})}$$

where $b_{j-1} = 5000$ and $b_j = 5500$. Then $C_j = \sum_{i=1}^{10} \mathbf{1}(5000 \leq X_i < 5500) = 2$ and $\hat{f}_X^{his}(5000) = \frac{2}{10\cdot 500} = 0.0004$.

Question 5. In the context of Question 3, estimate the density at 5000 using kernel density approach. Suppose we choose the uniform kernel and $h = 300$.

Solution. We first compute $K(\frac{u-X_i}{h})$ as follows

$$
\begin{aligned}
&K((5000 - 8710)/300) = 0\\
&K((5000 - 3850)/300) = 0\\
&K((5000 - 4970)/300) = 0.5\\
&K((5000 - 5398)/300) = 0\\
&K((5000 - 4780)/300) = 0.5\\
&K((5000 - 4020)/300) = 0\\
&K((5000 - 5790)/300) = 0\\
&K((5000 - 4510)/300) = 0\\
&K((5000 - 5520)/300) = 0\\
&K((5000 - 5300)/300) = 0.5
\end{aligned}
$$

So the kernel density estimate is

$$\hat{f}_X^k(5000) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{u - X_i}{h}) = \frac{1.5}{10 \cdot 300} = 0.0005.$$