STAT6106, Fall 2017

Mid-term Solution

Problem 1 (10 points)

Uncertainty of any distribution can be quantified by the variance of the distribution, thus we compare the prior variance with the poterior variance:

1. According to $Var(\theta) = Var(E(\theta|y)) + E(Var(\theta|y))$, we know that the prior variance $Var(\theta)$ is no less than the expected posterior variance $E(Var(\theta|y))$.

2. If the prior distribution agrees with the data likelihood (i.e., the two curves have similar shape and center), the posterior variance $Var(\theta|y)$ will be smaller than the prior variance $Var(\theta)$. This denotes the reduced uncertainty after cumulating consistent data/evidence.

3. If the prior distribution is contradictive with the data likelihood (e.g., the two curves have very different center), the posterior variance $Var(\theta|y)$ can be bigger than the prior variance $Var(\theta)$. This denoted that putting the unmatched prior and data together causes more confusion.

Problem 2 (50 points)

(**a**)(**15points**) We can introduce the following conjugate prior distribution of μ to incorporte the knowledge about HKU students' weight and the situation amone the universities in Hong Kong:

$$\mu \sim N(180, 40^2) \tag{1}$$

The likelihood function is:

$$
\begin{aligned}
p(x_1, \cdots, x_{10}|\mu) &= \prod_{i=1}^{10} p(x_i|\mu) \\
&= \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi \cdot 400}} e^{-\frac{(x_i-\mu)^2}{2\cdot 400}} \\
&= (2\pi \cdot 400)^{-10/2} \exp\{-\frac{1}{2\cdot 400}\sum_{i=1}^{10}(x_i-\mu)^2\}
\end{aligned} \tag{2}
$$

The posterior distribution of μ:

$$
\begin{aligned}
p(\mu|x_1, \cdots, x_{10}) &\propto p(\mu) \cdot p(x_1, \cdots, x_{10}|\mu) \\
&\propto \exp\{-\frac{(\mu-180)^2}{2\cdot 40^2}\} \cdot \exp\{-\frac{1}{2\cdot 400}\sum_{i=1}^{10}(x_i-\mu)^2\} \\
&\propto \exp\{-\frac{(\mu-\mu_*)^2}{2\sigma_*^2}\}
\end{aligned} \tag{3}
$$

where $\mu_* = \frac{\frac{180}{40^2}+\frac{10\cdot 150}{400}}{\frac{1}{40^2}+\frac{10}{400}} = 150.7317$, $\sigma_*^2 = \frac{1}{\frac{1}{40^2}+\frac{10}{400}} = 39.02439$.

Thus, the Bayesian inference for the mean weight of CUHK studentd would conclude that:

$$
\mu|x_1, \cdots, x_{10} \sim N(150.7317, 39.02439) \tag{4}
$$

(**b**)(**20points**) This is asking us to calculate the posterior predictive distribution of a new student's weight conditional on the observed sample:

$$
\begin{aligned}
p(x_*|x_1, \cdots, x_{10}) &= \int p(x_*|\mu) \cdot p(\mu|x_1, \cdots, x_{10}) d\mu \\
&\Downarrow \\
x_*|x_1, \cdots, x_{10} &\sim N(150.7317, 39.02439+400) = N(150.7317, 439.02439)
\end{aligned} \tag{5}
$$

Thus, the prediction of this new weight is the normal distribution $x_*|x_1, \cdots, x_{10} \sim N(150.7317, 39.02439+400) = N(150.7317, 439.02439)$.

(**c**)(**15points**)

The 95% posterior interval for the mean weight of CUHK students is (138.4879, 162.9755);
the 95% interval for the new weight is (109.6648, 191.7986).

The new weight has big uncertainty.

Problem 3 (40 points)

(**a**)(**10points**)

The likelihood function is:

$$
\begin{aligned}
p(x_1, \cdots, x_{10}|p) &= \prod_{i=1}^{10} p(1-p)^{x_i-1} = p^{10}(1-p)^{\sum_{i=1}^{10} x_i - 10} \\
&= p^{10}(1-p)^{30}
\end{aligned} \tag{6}
$$

Thus, the conjugate prior distribution for $\boldsymbol{p}$ should have the below form:

$$
p(\boldsymbol{p} = p) \propto p^a(1-p)^b \tag{7}
$$

this is Beta distribution:

$$
\begin{gathered}
X \sim Beta(\alpha, \beta) \\
f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\
0 < x < 1, \alpha > 0, \beta > 0
\end{gathered} \tag{8}
$$

(**b**)(**5points**) Without any prior knowledge, we can set the prior distribution about $\boldsymbol{p}$ to be uniform:

$$
\boldsymbol{p} \sim U(0,1) = Beta(1,1) \tag{9}
$$

(**c**)(**10points**)

The posterior distribution for $\boldsymbol{p}$:

$$
\begin{gathered}
p(\boldsymbol{p} = p|x_1, \cdots, x_{10}) \propto p^{10}(1-p)^{30} \\
\boldsymbol{p} \sim Beta(11, 31)
\end{gathered} \tag{10}
$$

(**d**)(**15points**) this is asking you to calculate the posterior predictive distribution of a new observation conditional on the observed 10 data points:

$$
\begin{aligned}
p(x_*|x_1,\cdots,x_{10}) &= \int p(x_*|p)\cdot p(p|x_1,\cdots,x_{10})dp \\
&= \int_0^1 p(1-p)^{x_*-1}\frac{\Gamma(42)}{\Gamma(11)\Gamma(31)}p^{10}(1-p)^{30}dp \\
&= \frac{\Gamma(42)}{\Gamma(11)\Gamma(31)}\int_0^1 p^{12-1}(1-p)^{30+x_*-1}dp \\
&= \frac{\Gamma(42)}{\Gamma(11)\Gamma(31)}\frac{\Gamma(12)\Gamma(30+x_*)}{\Gamma(42+x_*)}
\end{aligned}
\tag{11}
$$