HW 3

variance = 1/precision

1. a) $y_i \mid \tau \overset{iid}{\sim} N(m, \frac{1}{\tau}) = \frac{1}{\sqrt{2\pi \frac{1}{\tau}}} \cdot e^{\frac{-(y-m)^2}{2 \cdot \frac{1}{\tau}}}$

$$\Rightarrow p(y_1, \dots y_n \mid \tau) = \prod_{i=1}^{n} \frac{1}{\sqrt{\frac{1}{\tau}}\sqrt{2\pi}} \cdot e^{\frac{-1}{2} \cdot (y_i - m)^2 \cdot \tau}$$

$$\boxed{= \left(\frac{\sqrt{\tau}}{\sqrt{2\pi}}\right)^n \cdot e^{-\frac{\tau n}{2} \sum_{i=1}^{n} (y_i - m)^2}}$$

b) $p(\tau \mid y_1, \dots, y_n) \propto p(y_1, \dots y_n \mid \tau) \cdot p(\tau)$

$$= \left(\frac{\sqrt{\tau}}{\sqrt{2\pi}}\right)^n \cdot \frac{b^a}{\Gamma(a)} \tau^{a-1} \cdot e^{-b\tau} \cdot e^{-\frac{\tau n}{2} \sum_{i=1}^{n} (y_i - m)^2}$$

Consider $\sum_{i=1}^{n} (y_i - m)^2 = \sum (y_i - \bar{y} + \bar{y} - m)^2$

$$= \sum [(y_i - \bar{y})^2 + (\bar{y} - m)^2 + 2(y_i - \bar{y})(\bar{y} - m)]$$

$$\propto \sum [(\bar{y} - m)^2 + 2(y_i - \bar{y})(\bar{y} - m)]$$

$$= n(\bar{y} - m)^2 + 2(\bar{y} - m) \sum (y_i - \bar{y})$$

$$= n(\bar{y} - m)^2 + 2(\bar{y} - m)(n\bar{y} - n\bar{y})$$

$$= n(\bar{y} - m)^2$$

$$\Rightarrow e^{-b\tau} \cdot e^{-\frac{\tau n}{2} \sum_{i=1}^{n} (y_i - m)^2}$$

$$= e^{-b\tau + -\frac{\tau n^2}{2} (\bar{y} - m)^2} = e^{-\tau (b + \frac{n^2}{2} (\bar{y} - m)^2)}$$

$$\Rightarrow p(y_1, \dots y_n \mid \tau) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \frac{b^a}{\Gamma(a)} \cdot \tau^{a + \frac{n}{2} - 1} \cdot (\dots)$$

$$\propto \tau^{a+n-1} (\dots)$$

$$\boxed{= \text{Gamma}\left(a + \frac{n}{2}, b + \frac{n^2}{2} (\bar{y} - m)^2\right)}$$

2) $\theta \sim N(m, c^2)$

$$\Rightarrow \theta \mid y \sim N\left(\frac{1/c^2}{1/c^2 + n/s^2} m + \frac{n/s^2}{1/c^2 + n/s^2} \bar{y}, \frac{1}{1/c^2 + n/s^2}\right)$$

$\Rightarrow E[\tilde{Y}] = E[E[\tilde{Y} \mid \theta]] = E[\theta] (= m)$

$\Rightarrow Var[\tilde{Y}] = E[Var[\tilde{Y} \mid \theta]] + Var[E[\tilde{Y} \mid \theta]]$

$$= E[s^2] + Var[\theta] = s^2 + Var[\theta] (= s^2 + c^2)$$

Scanned with CamScanner

$\Rightarrow \tilde{Y} \sim N(m, s^2 + c^2)$

$\Rightarrow \tilde{Y} | y \sim N\left(\frac{1/c^2}{1/c^2 + n/s^2} m + \frac{n/s^2}{1/c^2 + n/s^2} \bar{y},\ s^2 + \frac{1}{1/c^2 + n/s^2} \right)$

4. Essentially, the argument is that since our posterior is basically just our prior updated with data, the data helps lower the variance. Let $\tilde{Y}$ be a prediction variable.

$Var[\theta] = E[Var[\theta | \tilde{Y}]] + Var[E[\theta | \tilde{Y}]]$

$Var[\theta | y] = E[Var[(\theta|y) | \tilde{Y}]] + Var[E[(\theta|y) | \tilde{Y}]]$

$\Rightarrow$ Consider $E[Var[\theta | \tilde{Y}]]$ to be the average variability when predicting the value of $\tilde{Y}$ given only the prior.

The average variability of predicting $\tilde{Y}$ using the prior tuned by the data is less than without the data, because the information lowers the uncertainty associated with our uncertain guessing of a prior $\Rightarrow E[Var[\theta | \tilde{Y}]] \geq E[Var[(\theta|y) | \tilde{Y}]]$

$\Rightarrow Var[E[\theta | \tilde{Y}]] = Var[\tilde{Y}]$

$Var[E[(\theta|y) | \tilde{Y}]] = Var[\tilde{Y}]$

so they're actually the same

Then, considering the 4 terms making up $Var[\theta]$ and $Var[\theta|y]$ and how they compare to each other, $Var[\theta] \geq Var[\theta|y]$.

5. $p(\lambda | y) = Gamma(a + \sum_i y_i, b + n)$

$= \frac{(b+n)^{a + \sum_i y_i}}{\Gamma(a + \sum_i y_i)} \lambda^{(a + \sum_i y_i - 1)} e^{-(b+n)\lambda}$

for prior $\lambda \sim Gamma(a, b)$, which is our prior, too

$\Rightarrow p(\tilde{Y} | y) = \int \frac{(b+n)^{a + \sum_i y_i}}{\Gamma(a + \sum_i y_i)} \cdot \lambda^{(a + \sum_i y_i - 1)} e^{(-b+n)\lambda} \cdot \frac{e^{-\lambda} \lambda^{\tilde{y}}}{\tilde{y}!} d\lambda$

$= \frac{(b+n)^{a + \sum_i y_i}}{\tilde{y}!\, \Gamma(a + \sum_i y_i)} \cdot \int (\lambda)^{\tilde{y} + a + \sum_i y_i - 1} \cdot e^{-\lambda(1 + b + n)} d\lambda$

$\downarrow$ Gamma

$= \frac{(b+n)^{a + \sum_i y_i}}{\tilde{y}!\, \Gamma(a + \sum_i y_i)} \cdot \frac{\Gamma(\tilde{y} + a + \sum_i y_i)}{(b+n+1)^{\tilde{y} + a + \sum_i y_i}}$

$$= \left(\frac{b+n}{b+n+1}\right)^{a+\sum_i y_i} \cdot \frac{1}{(b+n+1)^{\tilde{y}}} \cdot \binom{\tilde{y}+a+\sum_i y_i - 1}{\tilde{y}}$$

$$= \binom{\tilde{y}+a+\sum_i y_i - 1}{\tilde{y}} \cdot \left(\frac{b+n}{b+n+1}\right)^{a+\sum_i y_i} \cdot \left(1-\frac{b+n}{b+n+1}\right)^{\tilde{y}}$$

$$\Rightarrow p(\tilde{y} \mid y) \sim NB\left(a+\sum_i y_i, \frac{b+n}{b+n+1}\right)$$