UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in:	STK4021 — Applied Bayesian statistics - Home exam
Day of examination:	November 30 -2021
Examination hours:	15.00 – 19.00.

This problem set consists of 5 pages.

Appendices:	None
Permitted aids:	Anything available

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

(a) We have

$$r_c(\theta) = E(y-\theta)^2 = \text{Var}[y] = 1.$$

(b) We get directly that (θ, y) is multivariate Gaussian. We have that

$$\begin{aligned}
E[y] =& E[E[y|\theta]] = E[\theta] = 0\\
\text{Var}[y] =& E[\text{Var}[y|\theta]] + \text{Var}[E[y|\theta]] = E[1] + \text{Var}[\sigma^2] = 1 + \sigma^2\\
\text{Cov}[\theta, y] =& E[\text{Cov}[y,\theta|\theta]] + \text{Cov}[E[y|\theta], E[\theta|\theta] = 0 + \text{Cov}[\theta,\theta] = \sigma^2
\end{aligned}$$

(c) We have

$$\begin{aligned}
E[L(\theta,\hat{\theta})|y] =& E[(\theta-\hat{\theta})^2|y]\\
=& E[(\theta - E[\theta|y] + E[\theta|y] - \hat{\theta})^2|y]\\
=& E[(\theta - E[\theta|y])^2|y] + E[(E[\theta|y]-\hat{\theta})^2|y] +\\
& 2E[(\theta - E[\theta|y])(E[\theta|y]-\hat{\theta})|y]\\
\text{Var}[\theta|y] + (E[\theta|y]-\hat{\theta})^2 + 0
\end{aligned}$$

showing that choosing $\hat{\theta} = E[\theta|y]$ is the optimal choise.

Note that $E[\hat{\theta}_B|\theta] = \rho\theta$ and $\text{Var}[\hat{\theta}_B|\theta] = \rho^2$

The risk becomes

$$\begin{aligned}
E[(\theta-\hat{\theta})^2] =& E[(\hat{\theta} - \rho\theta + \rho\theta - \theta)^2]\\
=& E[(\hat{\theta}-\rho\theta)^2] + \theta^2(1-\rho)^2 + 0\\
=& \rho^2 + (1-\rho)^2\theta^2
\end{aligned}$$

(Continued on page 2.)

We then have $\rho^2 + (1-\rho)^2\theta^2 < 1$ when $\theta^2 < (1+\rho)/(1-\rho) = 2\sigma^2 + 1$.

(d) Since $y_i \sim N(0, \sigma^2 + 1)$, we could use $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n y_i^2 - 1$.

Problem 2

(a) We have

$$\begin{aligned} p(\theta|y) \propto & p(\theta)p(y|\theta) \\ \propto & (1-\theta)^{y-2}\theta^2 \propto \text{Beta}(3, y-1) \end{aligned}$$

(b) We have

$$\begin{aligned} p(y) =& p(y|\theta)p(\theta)d\theta \\ =& \int_\theta (y-1)(1-\theta)^{y-2}\theta^2 d\theta \\ =& (y-1)\frac{\Gamma(3)\Gamma(y-1)}{\Gamma(3+y-1)}\int_\theta \frac{\Gamma(3+y-1)}{\Gamma(3)\Gamma(y-1)}\theta^{3-1}(1-\theta)^{y-1-1} \\ =& (y-1)\frac{\Gamma(3)\Gamma(y-1)}{\Gamma(3+y-1)} \end{aligned}$$

(c) We have

$$L(\theta) = \prod_{i=1}^n p(y_i|\theta) = (1-\theta)^{\sum_{i=1}^n y_i - 2n}\theta^{2n}\prod_{i=1}^n (y_i - 1), \tag{1}$$

$$\begin{aligned} l(\theta) =& \sum_{i=1}^n \log p(y_i|\theta) \\ =& (\sum_{i=1}^n y_i - 2n)\log(1-\theta) + 2n\log(\theta) + \sum_{i=1}^n \log(y_i - 1), \\ \frac{\partial}{\partial\theta} l(\theta) =& -\frac{\sum_{i=1}^n y_i - 2n}{1-\theta} + \frac{2n}{\theta} \end{aligned}$$

Putting derivative to zero gives

$$2n(1-\hat{\theta}) = (\sum_{i=1}^n y_i - 2n)\hat{\theta}$$

or $\hat{\theta} = 2n/\sum_{i=1}^n y_i = 2/\bar{y}$.

(d) Normal approximation: We have

$$-\frac{\partial^2}{\partial^2\theta} l(\theta) = \frac{\sum_{i=1}^n y_i - 2n}{(1-\theta)^2} + \frac{2n}{\theta^2} = \hat{\sigma}^{-2}$$

and $p(\theta|y) \approx N(\hat{\theta}, \hat{\sigma}^2)$.

(Continued on page 3.)

(e) We have

$$
\begin{aligned}
p(\theta|y) \propto & p(\theta)p(\boldsymbol{y}|\theta) \\
\propto & \theta^{a-1}(1-\theta)^{b-1}(1-\theta)^{\sum_i y_i - 2n}\theta^{2n} \propto \text{Beta}(a+2n, b+n\bar{y}-2n)
\end{aligned}
$$

which gives

$$
E[\theta|\boldsymbol{y}] = \frac{a+2n}{a+2n+b+n\bar{y}-2n} = \frac{a+2n}{a+b+n\bar{y}}
$$

(f) We have

$$
\begin{aligned}
p(y_{11}|\boldsymbol{y}_{1:10}) = & \int_\theta p(y_{11}|\theta)p(\theta|\boldsymbol{y}_{1:10}d\theta \\
= & \int_\theta p(y_{11}|\theta)\text{Beta}(21,71)d\theta \\
= & \int_\theta (y_{11}-1)(1-\theta)^{y_{11}-2}\theta^2 \frac{\Gamma(92)}{\Gamma(21)\Gamma(71)}\theta^{21-1}(1-\theta)^{71-1}d\theta \\
= & (y_{11}-1)\frac{\Gamma(92)}{\Gamma(21)\Gamma(71)}\frac{\Gamma(23)\Gamma(69+y_{11})}{\Gamma(92+y_{11})} \\
& \int_\theta \frac{\Gamma(92+y_{11})}{\Gamma(23)\Gamma(69+y_{11})}\theta^{23-1}(1-\theta)^{69+y_{11}-1}d\theta \\
= & (y_{11}-1)\frac{\Gamma(92)}{\Gamma(21)\Gamma(71)}\frac{\Gamma(23)\Gamma(69+y_{11})}{\Gamma(92+y_{11})}
\end{aligned}
$$

Problem 3

(a) We have

$$
\begin{aligned}
\Pr(c=1|y) = & \frac{\Pr(c=1)p(y|c=1)}{p(y)} \\
& \propto f_1(y) \\
\Pr(c=2|y) = & \frac{\Pr(c=2)p(y|c=2)}{p(y)} \\
& \propto f_2(y)
\end{aligned}
$$

And due to that $\Pr(c=1|y) + \Pr(c=2|y) = 1$, we obtain

$$
\begin{aligned}
\Pr(c=1|y) = & \frac{f_1(y)}{f_1(y)+f_2(y)} \\
\Pr(c=2|y) = & \frac{f_2(y)}{f_1(y)+f_2(y)}
\end{aligned}
$$

(Continued on page 4.)

(b) We have

$$E[L(c,1)|y] = E[I(\hat{c}=1,c=2)|y] = E[I(c=2|y)] = \Pr(c=2|y) = 1-\Pr(c=1|y)$$

and similarly for $E[L(c,2)]$

(c) We also have

$$E[L(c,D)|y] = 0.1$$

for any y. This means that we will make the decisions

$$\begin{aligned}
\hat{c} =& \begin{cases} 1 & \text{if } 1-\Pr(c=1|y) < \min\{0.1, 1-\Pr(c=2|y)\} \\ 2 & \text{if } 1-\Pr(c=2|y) < \min\{0.1, 1-\Pr(c=1|y)\} \\ D & \text{if } 0.1 < \min\{1-\Pr(c=1|y), 1-\Pr(c=2|y)\} \end{cases} \\
=& \begin{cases} 1 & \text{if } \Pr(c=1|y) > \max\{0.9, \Pr(c=2|y)\} \\ 2 & \text{if } \Pr(c=2|y) > \max\{0.9, \Pr(c=1|y)\} \\ D & \text{if } 0.9 > \max\{\Pr(c=1|y), \Pr(c=2|y)\} \end{cases} \\
=& \begin{cases} 1 & \text{if } \Pr(c=1|y) > 0.9 \\ 2 & \text{if } \Pr(c=2|y) > 0.9 \\ D & \text{if } \max\{\Pr(c=1|y), \Pr(c=2|y) < 0.9\} \end{cases}
\end{aligned}$$

(d) We now have

$$\begin{aligned}
\Pr(c=1|y) =& \frac{(2\pi)^{-1/2}\exp(-0.5(y+1)^2}{(2\pi)^{-1/2}\exp(-0.5(y+1)^2 + (2\pi)^{-1/2}\exp(-0.5(y-1)^2} \\
=& \frac{\exp(-0.5(y+1)^2}{\exp(-0.5(y+1)^2 + \exp(-0.5(y-1)^2} \\
=& \frac{1}{1+\exp(-0.5[(y-1)^2-(y+1)^2]} \\
=& \frac{1}{1+\exp(-0.5[-4y])} = \frac{1}{1+\exp(2y)}
\end{aligned}$$

We then classify to 1 if

$$\frac{1}{1+\exp(2y)} > 0.9$$

which is equivalent to that $y < -0.5\log(9)$.

Similarly we classify to 2 if $y > 0.5\log(9)$.

And then to D if $-0.5\log(9) \le y \le 0.5\log(9)$.

(Continued on page 5.)

Problem 4

(a) We have

$$p(\theta_1|\boldsymbol{y}_1) \propto \prod_{i=1}^{5} \frac{1}{1+(y_{1i}-\theta_1)^2}$$
$$p(\theta_2|\boldsymbol{y}_2) \propto \prod_{i=1}^{5} \frac{1}{1+(y_{2i}-\theta_1)^2}$$

On can compute these on a dense grid on the region $[0,100]$, then normalize and plot these.

(b) A possibility here is to use rejection sampling simulate from the prior and then calculate importance weights. Note that an upper limit will here be 1 so we accept with probabilities

$$\prod_{i=1}^{5} \frac{1}{1+(y_{1i}-\theta_1)^2}$$

for $p(\theta_1|\boldsymbol{y}_1)$ and similarly for θ_2

(c) Given samples $\{\theta_1^s, \theta_2^s\}$ we can obtain samples $\rho^s = \theta_2^s/\theta_1^s$ from which we can make inference through Monte Carlo estimation.