Revision for Michaelmas term

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Exercise 1. Consider a sequence of exchangeable observables $x_{1:n} = (x_1, ..., x_n)$, where it is $x_i \in \mathcal{X}_k$, for $i = 1, ..., n$, where $\mathcal{X}_k = \{x \in \{0,1\}^k | \sum_{j=1}^k x_j = 1\}$. In words, x_i is a k-dimensional vector all of whose elements are equal to 0 except for one which is equal to 1, for $i = 1, ..., n$. Consider the Bayesian model

$$\begin{cases} x_i|\theta & \stackrel{\text{IID}}{\sim} \text{Mu}_k(\theta) \\ \theta & \sim \Pi(\cdot) \end{cases}$$

where $\theta \in \Theta$, with $\Theta = \{\theta \in (0,1)^k | \sum_{j=1}^k \theta_j = 1\}$. Here, Mu_k denotes the Multinomial probability distribution with PMF

$$\text{Mu}_k(x|\theta) = \begin{cases} \prod_{j=1}^k \theta_j^{x_j} & \text{, if } x \in \mathcal{X}_k \\ 0 & \text{, otherwise} \end{cases} \tag{1}$$

1. Show that the parametric model (1) is a member of the $k-1$ exponential family.
2. Compute the likelihood $f(x_{1:n}|\theta)$, and find the sufficient statistic $t_n := t_n(x_{1:n})$.
3. Derive the conjugate prior distribution for θ, and then show that it is a Dirichlet distribution.

   You may use the fact that the Dirichlet distribution $\text{Di}_k(a)$ with parameter $a = (a_1, ..., a_k)$, where $\{a_j > 0;\ j = 1, ..., k\}$ has PDF

   $$\text{Di}_k(\theta|a) = \begin{cases} \frac{\Gamma(\sum_{j=1}^k a_j)}{\prod_{j=1}^k \Gamma(a_j)} \prod_{j=1}^k \theta_j^{a_j-1} & \text{, if } \theta \in \Theta \\ 0 & \text{, otherwise} \end{cases}$$

4. Compute the posterior distribution. State the name of the distribution, and express its parameters with respect to the observations and the hyper-parameters of the prior. Justify your answer.
5. Compute the probability mass function of the predictive distribution for a future observation $y = x_{n+1}$ in closed form.

   Hint $\Gamma(x) = (x-1)\Gamma(x-1)$.

6. Suppose you are interested in checking if a k-sided die is fair or not. You collect n observations $\{x_i\}_{i=1}^n$, where $x_i \in \mathcal{X}_k$, according to the following experiment. You throw the die n times; at the i-th throw, you record the result as $x_{i,j} = 1$ if the result is the j-th side and as $x_{i,j} = 0$ if the result is otherwise for $j = 1, ..., k$.
   (a) Set up the pair of hypotheses, by stating explicitly the pair of hypothesis, and computing the Bayes factor in closed form.
   (b) Suppose that it is a 4-sided die, you throw it 6 times, and it comes up ‘1’, 4 times; ‘2’, 0 times; ‘3’, 1 time; and ‘4’, 1 time. Perform the Bayesian test to check whether the dice is fair or not. State your decision based on Jeffreys’ scale rule.

Solution.

1. It is

$$\mathrm{Mu}_k(x|\theta) = \prod_{j=1}^{k} \theta_j^{x_j} = \prod_{j=1}^{k-1} \theta_j^{x_j} (1 - \sum_{j=1}^{k-1} \theta_j)^{1-\sum_{j=1}^{k-1} x_j}$$
$$= (1 - \sum_{j=1}^{k-1} \theta_j) \exp(\sum_{j=1}^{k-1} x_j \log(\frac{\theta_j}{1 - \sum_{j=1}^{k-1} \theta_j}))$$

   This is the $k-1$ exponential family PDF with

$$u(x) = 1; \qquad g(\theta) = (1 - \sum_{j=1}^{k-1} \theta_j);$$
$$h(x) = (x_1, ...x_{k-1}); \qquad \phi(\theta) = (\log(\frac{\theta_1}{1 - \sum_{j=1}^{k-1} \theta_j}), ..., \log(\frac{\theta_{k-1}}{1 - \sum_{j=1}^{k-1} \theta_j})),$$
$$c = (1, ..., 1)$$

2. The likelihood is

$$f(x_{1:n}|\theta) = \prod_{i=1}^{n} \mathrm{Mu}_k(x_i|\theta) = \prod_{j=1}^{k} \theta_j^{\sum_{i=1}^{n} x_{i,j}} = \prod_{j=1}^{k} \theta_j^{x_{*,j}}$$
$$= (1 - \sum_{j=1}^{k-1} \theta_j)^n \exp(\sum_{j=1}^{k-1} x_{*,j} \log(\frac{\theta_j}{1 - \sum_{j=1}^{k-1} \theta_j}))$$

   and the sufficient statistic is

$$t_n = (n, x_{*,1}, ..., x_{*,k-1})$$

3. Let $\tau = (\tau_0, ..., \tau_{k-1})$. It is

$$\pi(\theta|\tau) \propto (1 - \sum_{j=1}^{k-1} \theta_j)^{\tau_0} \exp(\sum_{j=1}^{k-1} \tau_j \log(\frac{\theta_j}{1 - \sum_{j=1}^{k-1} \theta_j}))$$
$$\propto \prod_{j=1}^{k-1} \theta_j^{\tau_j} (1 - \sum_{j=1}^{k-1} \theta_j)^{\tau_0 - \sum_{j=1}^{k-1} \tau_j} \propto \prod_{j=1}^{k-1} \theta_j^{\tau_j} \, \theta_k^{\tau_0 - \sum_{j=1}^{k-1} \tau_j}$$

   Here, I recognize the Dirichlet distribution with $a_j = \tau_j$ for $j = 1, ..., k-1$ and $a_k = \tau_0 - \sum_{j=1}^{k-1} \tau_j$.

4. Due to conjugacy, it is

$$\pi(\theta|x_{1:n}) = \pi(\theta|\tau + t_n) = \prod_{j=1}^{k-1} \theta_j^{\tau_j + x_{*,j}} \, \theta_k^{\tau_0 + n - \sum_{j=1}^{k-1} (\tau_j + x_{*,j})}$$

   So the posterior is

$$\mathrm{Di}_k(\theta|\tilde{a}) = \frac{\Gamma(\sum_{j=1}^{k} \tilde{a}_j)}{\prod_{j=1}^{k} \Gamma(\tilde{a}_j)} \prod_{j=1}^{k} \theta_j^{\tilde{a}_j - 1} 1_{\Theta}(\theta)$$

   where $\tilde{a} = (\tilde{a}_1, ..., \tilde{a}_k)$, with $\tilde{a}_j = a_j + x_{*,j}$ for $j = 1, ..., k$.

5. It is

$$p(y|x_{1:n}) = \int \text{Mu}_k(y|\theta)\text{Di}_k(\theta|\tilde{a})\text{d}\theta = \int \prod_{j=1}^{k} \theta_j^{y_j} \frac{\Gamma(\sum_{j=1}^{k}\tilde{a}_j)}{\prod_{j=1}^{k}\Gamma(\tilde{a}_j)} \prod_{j=1}^{k}\theta_j^{\tilde{a}_j-1}\text{d}\theta$$

$$= \frac{\Gamma(\sum_{j=1}^{k}\tilde{a}_j)}{\prod_{j=1}^{k}\Gamma(\tilde{a}_j)} \int \prod_{j=1}^{k}\theta_j^{y_j+\tilde{a}_j-1}\text{d}\theta = \frac{\Gamma(\sum_{j=1}^{k}\tilde{a}_j)}{\prod_{j=1}^{k}\Gamma(\tilde{a}_j)} \frac{\prod_{j=1}^{k}\Gamma(y_j+\tilde{a}_j)}{\Gamma(\sum_{j=1}^{k}(y_j+\tilde{a}_j))}$$

$$= \frac{\Gamma(\sum_{j=1}^{k}(a_j+x_{*,j}))}{\prod_{j=1}^{k}\Gamma(a_j+x_{*,j})} \frac{\prod_{j=1}^{k}\Gamma(y_j+a_j+x_{*,j})}{\Gamma(\sum_{j=1}^{k}(y_j+a_j+x_{*,j}))}$$

$$= \frac{\Gamma(a_*+x_{*,*})}{\prod_{j=1}^{k}\Gamma(a_j+x_{*,j})} \frac{\prod_{j=1}^{k}\Gamma(y_j+a_j+x_{*,j})}{\Gamma(\sum_{j=1}^{k}y_j+a_*+x_{*,*})}$$

$$= \frac{\Gamma(a_*+n)}{\prod_{j=1}^{k}\Gamma(a_j+x_{*,j})} \frac{\prod_{j=1}^{k}\Gamma(y_j+a_j+x_{*,j})}{\Gamma(1+a_*+n)}$$

$$= \frac{\Gamma(n+a_*)}{\Gamma(1+a_*+n)} \prod_{j=1}^{k} \frac{\Gamma(y_j+a_j+x_{*,j})}{\Gamma(a_j+x_{*,j})}$$

so $p(y|x_{1:n}) = \frac{1}{n+a_*}(a_{j'}+x_{*,j'})$, where j' such that $y_{j'}=1$.

6.

(a) Obviously the hypothesis test is

$$\begin{cases} \text{H}_0: & \theta = \theta_0, \\ \text{H}_1: & \theta \neq \theta_0, \end{cases}$$

where $\theta_0 = 1/k$. The Bayes factor is

$$\text{B}_{01}(x_{1:n}) = \frac{p_0(x_{1:n})}{p_1(x_{1:n})} = \frac{\prod_{i=1}^{n}\text{Mu}_k(x_i|\theta_0)}{\int_\Theta \prod_{i=1}^{n}\text{Mu}_k(x_i|\theta)\text{Di}_k(\theta|a)\text{d}\theta}$$

So, it is

$$p_0(x_{1:n}) = \prod_{i=1}^{n}\text{Mu}_k(x_i|\theta_0) = \prod_{i=1}^{n}\prod_{j=1}^{k}\theta_0^{x_{i,j}} = (\frac{1}{k})^n$$

and it is

$$p_1(x_{1:n}) = \int_\Theta \prod_{i=1}^{n}\text{Mu}_k(x_i|\theta)\text{Di}_k(\theta|a)\text{d}\theta = \int_\Theta \prod_{i=1}^{n}\prod_{j=1}^{k}\theta_j^{x_{i,j}} \frac{\Gamma(\sum_{j=1}^{k}a_j)}{\prod_{j=1}^{k}\Gamma(a_j)}\prod_{j=1}^{k}\theta_j^{a_j-1}\text{d}\theta$$

$$= \frac{\Gamma(a_*)}{\prod_{j=1}^{k}\Gamma(a_j)} \int_\Theta \prod_{j=1}^{k}\theta_j^{x_{*,j}+a_j-1}\text{d}\theta = \frac{\Gamma(a_*)}{\prod_{j=1}^{k}\Gamma(a_j)} \frac{\prod_{j=1}^{k}\Gamma(a_j+x_{*,j})}{\Gamma(a_*+n)}$$

$$= \frac{\Gamma(a_*)}{\Gamma(a_*+n)} \prod_{j=1}^{k}\frac{\Gamma(a_j+x_{*,j})}{\Gamma(a_j)} = \frac{\prod_{j=1}^{k}\prod_{\ell=0}^{x_{*,j}-1}(a_j+\ell)}{\prod_{\ell=0}^{n-1}(a_*+\ell)}$$

So

$$\text{B}_{01}(x_{1:n}) = (\frac{1}{k})^n \frac{\prod_{\ell=0}^{n-1}(a_*+\ell)}{\prod_{j=1}^{k}\prod_{\ell=0}^{x_{*,j}-1}(a_j+\ell)}$$

(b) I got $k=4$, $n=6$, $x_{*,1:4}=(4,0,1,1)$. Also I consider that I have no a priori information, and hence I can use non-informative prior, e.g. $a_{*,1:4}=(1,1,1,1)$, $a_{*}=4$. So

$$\mathrm{B}_{01}(x_{1:n})=(\frac{1}{4})^{6}\cdot\frac{4\cdot5\cdot6\cdot7\cdot8\cdot9}{1\cdot2\cdot3\cdot4}=0.61523,$$

$\log_{10}(\mathrm{B}_{01}(x_{1:n}))=-0.2109$. Evidence against H_0: not worth more than a bare.

Exercise 2. Consider a Bayesian model

$$\begin{cases} y|\theta & \sim \mathrm{N}(\theta,1) \\ \theta & \sim \mathrm{N}(0,1) \end{cases}$$

1. Compute the Bayes point estimate $\hat{\delta}$ of θ under the loss $\ell(\theta,\delta)=\exp\left(\frac{3}{4}\theta^2\right)(\theta-\delta)^2$
2. Show that $\hat{\delta}$ is inadmissible, and discuss why this happens according to the Theorems in Handout 10.

Solution. The posterior of θ given y is $\theta|y\sim\mathrm{N}\left(\mu=\frac{1}{2}y,\sigma^2=\frac{1}{2}\right)$ –the derivation is easy.

1. According to Proposition in the Handout, because $\mathrm{E}_{\pi}\left(\exp\left(\frac{3}{4}\theta^2\right)|y\right)>0$, it is

$$\hat{\delta}=\frac{\mathrm{E}_{\pi}\left(\theta\exp\left(\frac{3}{4}\theta^2\right)|y\right)}{\mathrm{E}_{\pi}\left(\exp\left(\frac{3}{4}\theta^2\right)|y\right)}$$

I compute $\Delta_j=\mathrm{E}_{\pi}\left(\theta^j\exp\left(\frac{3}{4}\theta^2\right)|y\right)$ up to a multiplicative on j. It is, for $j=0,1$

$$\begin{aligned}\Delta_j=\mathrm{E}(w(\theta)\theta^j|y)&=\int_{\Theta}\exp(\frac{3}{4}\theta^2)\theta^j\mathrm{N}(\theta|\frac{1}{2}y,\frac{1}{2})\mathrm{d}\theta\\
&=\int_{\Theta}\exp(\frac{3}{4}\theta^2)\theta^j\frac{1}{\sqrt{2\pi1/2}}\exp(-\frac{1}{2}\frac{(\frac{1}{2}y-\theta)^2}{1/2})\mathrm{d}\theta\\
&=\int_{\Theta}\theta^j\frac{1}{\sqrt{2\pi1/2}}\exp(-\frac{1}{2}\frac{(\frac{1}{2}y-\theta)^2}{1/2}+\frac{3}{4}\theta^2)\mathrm{d}\theta\\
&=\int_{\Theta}\theta^j\frac{1}{\sqrt{2\pi1/2}}\exp(-\frac{1}{2}\frac{(\theta-2y)^2}{2}+\frac{3}{2}y^2)\mathrm{d}\theta\\
&=\frac{1}{2}\exp(\frac{3}{2}y^2)\int_{\Theta}\theta^j\frac{1}{\sqrt{2\pi2}}\exp(-\frac{1}{2}\frac{(\theta-2y)^2}{2})\mathrm{d}\theta\;=\begin{cases}\frac{1}{2}\exp(\frac{3}{2}y^2)2y & ,c=1\\ \frac{1}{2}\exp(\frac{3}{2}y^2) & ,c=0\end{cases}\end{aligned}$$

So, I get $\hat{\delta}=\frac{\Delta(1)}{\Delta(0)}=2y$.

2. Assume $\delta_c(y)=cy$ where my estimator is a member; i.e. $\hat{\delta}=\delta_2$. The risk for $\delta_c(y)$ is

$$R(\theta,\delta_c)=\mathrm{E}_{\mathrm{N}(\theta,1)}\left(\exp\left(\frac{3}{4}\theta^2\right)(\theta-cy)^2\right)=\exp\left(\frac{3}{4}\theta^2\right)\left(c^2+\theta^2(c-1)^2\right)$$

as

$$
\begin{aligned}
R(\theta,\delta_c) &= \mathrm{E}_F(\ell(\theta,\delta_c(y))|\theta) = \mathrm{E}_{\mathrm{N}(\theta,1)}\left(\exp(\frac{3}{4}\theta^2)(\theta-cy)^2\right) \\
&= \exp\left(\frac{3}{4}\theta^2\right)\mathrm{E}_{\mathrm{N}(\theta,1)}(\theta-cy)^2 \\
&= \exp\left(\frac{3}{4}\theta^2\right)\mathrm{E}_{\mathrm{N}(\theta,1)}(cy-\theta)^2 = \exp\left(\frac{3}{4}\theta^2\right)\mathrm{E}_{\mathrm{N}(\theta,1)}([cy-c\theta]+[c\theta-\theta])^2 \\
&= \exp\left(\frac{3}{4}\theta^2\right)\mathrm{E}_{\mathrm{N}(\theta,1)}([cy-c\theta]^2+[c\theta-\theta]^2-2[cy-c\theta]) \\
&= \exp\left(\frac{3}{4}\theta^2\right)\left(c^2\mathrm{E}_{\mathrm{N}(\theta,1)}[y-\theta]^2+\mathrm{E}_{\mathrm{N}(\theta,1)}[(c-1)\theta]^2-2[c\mathrm{E}_{\mathrm{N}(\theta,1)}(y)-c\theta]\right) \\
&= \exp\left(\frac{3}{4}\theta^2\right)\left(c^2+\theta^2(c-1)^2\right)
\end{aligned}
$$

I observe that $\hat{\delta}=\delta_2$ is dominated by δ_1. It is

$$R(\theta,\delta_2) = \exp\left(\frac{3}{4}\theta^2\right)(4+\theta)^2$$

and

$$R(\theta,\delta_1) = \exp\left(\frac{3}{4}\theta^2\right)$$

where one can see that $R(\theta,\delta_2)=R(\theta,\delta_1)$ for $\theta\in\{0,-4\}$ and $R(\theta,\delta_2)>R(\theta,\delta_1)$ for all the rest θ. Hence $\hat{\delta}=\delta_2$ is inadmissible.

I observe that $\hat{\delta}$ does not produces a finite Bayes risk

$$
\begin{aligned}
r(\pi,\hat{\delta}) &= \int R(\theta,\hat{\delta})\pi(\theta)\mathrm{d}\theta = \int R(\theta,\hat{\delta})\mathrm{N}(\theta|0,1)\mathrm{d}\theta \\
&\propto \int (4+\theta)^2\exp\left(\frac{3}{4}\theta^2\right)\exp\left(-\frac{1}{2}\theta^2\right)\mathrm{d}\theta \\
&= \int (4+\theta)^2\exp\left(\frac{1}{4}\theta^2\right)\mathrm{d}\theta > \int \exp\left(\frac{1}{4}\theta^2\right)\mathrm{d}\theta = \infty
\end{aligned}
$$

and hence Bayesian point estimate $\hat{\delta}$ is not necessarily admissive.

Exercise 3. Consider the Bayesian model

$$
\begin{aligned}
y_i|\theta &\overset{\text{iid}}{\sim} \mathrm{N}\left(\theta,\sigma^2\right),\ i=1,...,n \\
\theta &\sim \mathrm{N}\left(\mu_0,\sigma_0^2\right)
\end{aligned}
$$

where μ_0,σ_0^2 are fixed hyper-parameters, and θ unknown.

1. Derive the $1-\frac{a}{2}$ HPD credible posterior interval for θ.

   Hint-1: It is
   $$\sum_{i=1}^{n}\frac{(x-\mu_i)^2}{\sigma_i^2} = \frac{(x-\hat{\mu})^2}{\hat{\sigma}^2} + \text{const ind of } x$$
   where $\hat{\sigma}^2=\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)^{-1}$ and $\hat{\mu}=\hat{\sigma}^2\left(\sum_{i=1}^{n}\frac{\mu_i}{\sigma_i^2}\right)$.

   Hint-2: The 97.5% quantile of the standard Normal distribution is 1.959964.

2. What size your dataset need to have in order to satisfy a 0.95% HPD credible posterior interval for θ which has length of 1 unit? Consider that $\sigma^2 = 4$ and $\sigma_0^2 = 9$.

Solution. Let $y = (y_1, ..., y_n)$.

1. The posterior pdf of θ is

$$\pi(\theta|y) \propto f(y|\theta) \times \pi(\theta) = \prod_{i=1}^{n} f(y_i|\theta\times)\,\pi(\theta) \propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\frac{(y_i-\theta)^2}{\sigma^2}\right) \times \exp\left(-\frac{1}{2}\frac{(\theta-\mu_0)^2}{\sigma_0^2}\right)$$
$$= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{(\theta-y_i)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\sigma_0^2}\right]\right) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(\theta-\mu_n)^2}{\sigma_n^2} + \text{const...}\right)$$

with

$$\sigma_n^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1} \text{ and } \mu_n = \sigma_n^2\left(\frac{\bar{y}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)$$

Hence

$$\theta|y \sim \mathrm{N}\left(\mu_n, \sigma_n^2\right)$$

To find the 2-sides the $1-\frac{a}{2}$ HPD credible posterior interval for θ, aka $[L, U]$, I consider the theorem in the Handouts. Namely:

$$1-a = \int_L^U \mathrm{N}(\theta|\mu_n, \sigma_n^2)\mathrm{d}\theta = \mathsf{P}_{\mathrm{N}(\mu_n,\sigma_n^2)}(\theta < U) - \mathsf{P}_{\mathrm{N}(\mu_n,\sigma_n^2)}(\theta < L) = \Phi\left(\frac{U-\mu_n}{\sigma_n}\right) - \Phi\left(\frac{L-\mu_n}{\sigma_n^2}\right)$$

where $\Phi(\cdot)$ denotes the CDF of $\mathrm{N}(0, 1)$. Also, it has to be

$$\pi(U|y) = \pi(L|y)$$

and because the PDF of $\mathrm{N}\left(\mu_n, \sigma_n^2\right)$ is symmetric around μ_n

$$L - \mu_n = \mu_n - U \implies L = 2\mu_n - U$$

So

$$1-a = \Phi\left(\frac{U-\mu_n}{\sigma_n}\right) - \Phi\left(\frac{L-\mu_n}{\sigma_n}\right) = \Phi\left(\frac{U-\mu_n}{\sigma_n}\right) - \Phi\left(\frac{2\mu_n-U-\mu_n}{\sigma_n}\right) = 2\Phi\left(\frac{U-\mu_n}{\sigma_n}\right) - 1 \implies$$
$$1-\frac{a}{2} = \Phi\left(\frac{U-\mu_n}{\sigma_n}\right) \implies z_{1-\frac{a}{2}} = \frac{U-\mu_n}{\sigma_n} \implies U = \mu_n + z_{1-\frac{a}{2}}\sigma_n^2$$

and hence $L = \mu_n - z_{1-\frac{a}{2}}\sigma_n^2$.

So the $1-\frac{a}{2}$ HPD credible posterior interval for θ is

$$\left[\mu_n - z_{1-\frac{a}{2}}\sigma_n\ ,\ \mu_n + z_{1-\frac{a}{2}}\sigma_n\right] = \left[\mu_n - z_{1-\frac{a}{2}} \Big/ \sqrt{\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}\ ,\ \mu_n + z_{1-\frac{a}{2}} \Big/ \sqrt{\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}\right]$$

2. The length of the $1-\frac{a}{2}$ HPD credible posterior interval for θ is

$$\ell_n = 2z_{1-\frac{a}{2}} \Big/ \sqrt{\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}$$

So

$$1 = \ell_n = 2z_{1-\frac{a}{2}} \Big/ \sqrt{\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)} = 2 \times 1.959964 \Big/ \sqrt{\left(\frac{n}{4} + \frac{1}{9}\right)}$$

and hence $n \approx 62$.

Example 4. Consider the following 1-way ANOVA problem where the factor has n levels and each level has the same number of repeatations m. Namely, $y_{i,j} = \theta_i + \epsilon_{i,j}$ where θ_i is the effect of the ith level and $\epsilon_{i,j} \sim \mathrm{N}(0, \sigma^2)$ is the error in the j the repeatation, for $j = 1, ..., m$, and $i = 1, ..., n$. Assume $\theta_i \sim \mathrm{N}\left(\mu, \tau^2\right)$ for $i = 1, ..., n$. Assume σ^2 is known, while μ and τ^2 are unknown.

1. Compute the EB estimator of $\{\theta_i\}$ under the square loss where μ and τ^2 are learned via ML-II, and show it can be written in the form $\theta_i = \varpi\mu + (1 - \varpi)\, y_i$.

2. Compute the EB estimator of $\{\theta_i\}$ under the square loss where μ and τ^2 are learned by MoM

3. Compute the EB estimator of $\{\theta_i\}$ by constructing unbiased estimates for ϖ and μ . Hint: if $\xi \sim \chi^2_v$then $\mathrm{E}\left(1/\xi\right) = 1/(v-2)$

Solution. Assume that $y_i = \frac{1}{m}\sum_j y_{i,j}$ and $\sigma_m^2 = \sigma^2/m$. The Bayesian model is

$$\begin{cases} y_i|\theta_i & \sim \mathrm{N}\left(\theta_i, \sigma_m^2\right),\ i = 1, ..., n \\ \theta_i & \sim \mathrm{N}\left(\mu, \tau^2\right),\ i = 1, ..., n \end{cases}$$

By using Bayesian theorem, the posterior distribution of $\theta = (\theta_1, ..., \theta_n)^\top$ is

$$\theta_i|y_i \sim \mathrm{N}\left(\frac{\sigma_m^2}{\sigma_m^2 + \tau^2}\mu + \frac{\tau^2}{\sigma_m^2 + \tau^2}y_i, \frac{\sigma_m^2\tau^2}{\sigma_m^2 + \tau^2}\right) \equiv \mathrm{N}\left(\varpi\mu + (1-\varpi)\, y_i, \sigma_m^2\,(1-\varpi)\right)$$

where $\varpi = \frac{\sigma_m^2}{\sigma_m^2 + \tau^2}$. Hence the Bayesian estimator is $\theta_i = \varpi\mu + (1 - \varpi)\, y_i$.

1. The marginal posterior is $y_i \overset{\text{iid}}{\sim} \mathrm{N}\left(\mu, \sigma_m^2 + \tau^2\right)$ by using standard Normal pdf properties. Essentially this is a regression problem, $y \sim \mathrm{N}\left(1\mu, I_n\left(\sigma_m^2 + \tau^2\right)\right)$ with log likelihood

$$\log\left(g(y)\right) \overset{\text{calcul...}}{=\!=\!=} -\frac{n}{2}\log\left(\sigma_m^2 + \tau^2\right) - \frac{ns^2}{2\left(\sigma_m^2 + \tau^2\right)} - \frac{n\left(\bar{y} - \mu\right)^2}{2\left(\sigma_m^2 + \tau^2\right)} + \text{const}$$

   where $s^2 = \frac{1}{n}\sum_i\left(y_i - \bar{y}\right)^2$. It is maximized for $\hat{\mu}_{\text{ML-II}} = \bar{y}$ regardless τ^2. Then at $\hat{\mu}_{\text{ML-II}} = \bar{y}$, it is

$$0 = \frac{\mathrm{d}}{\mathrm{d}\tau}\log\left(g(y)\right)\Big|_{\tau=\hat{\tau}} = -\frac{n}{2\left(\sigma_m^2 + \hat{\tau}^2\right)} + \frac{ns^2}{2\left(\sigma_m^2 + \hat{\tau}^2\right)}$$

   which implies $\hat{\tau}_{\text{ML-II}} = \max\left(s - \sigma_m^2\right)$. Hence

$$\hat{\theta}_i^{\text{EB}} = \hat{\varpi}_{\text{ML-II}}\bar{y} + \left(1 - \hat{\varpi}_{\text{ML-II}}\right) y_i, \quad \text{where} \quad \hat{\varpi}_{\text{ML-II}} = \frac{\sigma_m^2}{\sigma_m^2 + \max\left(s - \sigma_m^2\right)}$$

2. It is

$$\begin{cases} \mathrm{E}\left(\bar{y}\right) = \bar{y} \\ \mathrm{Var}\left(\bar{y}\right) = \frac{n}{n-1}s^2 \end{cases} \iff \begin{cases} \hat{\mu} = \bar{y} \\ \sigma_m^2 + \hat{\tau}^2 = \frac{n}{n-1}s^2 \end{cases} \iff \begin{cases} \hat{\mu}_{\text{MoM}} = \bar{y} \\ \hat{\tau}^2_{\text{MoM}} = \frac{n}{n-1}s^2 - \sigma_m^2 \end{cases}$$

   Hence

$$\hat{\theta}_i^{\text{EB}} = \hat{\varpi}_{\text{MoM}}\bar{y} + \left(1 - \hat{\varpi}_{\text{MoM}}\right) y_i, \quad \text{where} \quad \hat{\varpi}_{\text{MoM}} = \frac{n-1}{n}\frac{\sigma_m^2}{s^2}$$

3. From Cohran's theorem I know that $\bar{y} \sim \mathrm{N}\left(\mu, \frac{1}{n}(\sigma_m^2+\tau^2)\right)$, $n\frac{s^2}{\sigma_m^2+\tau^2} \sim \chi^2_{n-1}$, and that $\bar{y}$and s^2 are independent. Hence

$$\mathrm{E}\left(\bar{y}\right) = \mu, \text{ and } \qquad \mathrm{E}\left(\frac{(n-3)\sigma_m^2}{ns^2}\right) = \frac{\sigma_m^2}{\sigma_m^2+\tau^2}$$

So

$$\hat{\theta}_i^{\text{EB}} = \hat{\varpi}\bar{y} + \left(1-\hat{\varpi}\right)y_i, \quad \text{where} \quad \hat{\varpi} = \frac{(n-3)\sigma_m^2}{ns^2}$$

Exercise 5. [1-way ANOVA] Consider the following 1-way ANOVA problem where the factor has n levels and each level has the same number of repeatations m. Namely, $y_{i,j} = \theta_i + \epsilon_{i,j}$ where θ_i is the effect of the ith level and $\epsilon_{i,j} \sim \mathrm{N}(0,\sigma^2)$ is the error in the j the repeatation, for $j=1,...,m$, and $i=1,...,n$. Assume $\theta_i \sim \mathrm{N}\left(z_i^\top\beta, \tau^2\right)$ for $i=1,...,n$. Assume σ^2 and $\{z_i\}$ are known, while β and τ^2 are unknown.

1. Compute the EB estimator of $\{\theta_i\}$ under the square loss when β and τ^2 are learned via ML-II.

2. Compute the EB estimator of $\{\theta_i\}$ under the square loss such that β and τ^2 are unbiased estimators.

Hint-1: To fulfill the 'unbiased estimators´ use MoM

Hint-2: Assume known that $\frac{\|y-Z\beta\|^2}{\sigma^2/m+\tau^2} \sim \chi^2_{n-d}$

Hint-3: Assume known that $\mathrm{E}\left(\xi\right) = n-d$ if $\xi \sim \chi^2_{n-d}$

Solution. Assume that $y_i = \frac{1}{m}\sum_j y_{i,j}$ and $\sigma_m^2 = \sigma^2/m$. The Bayesian model is

$$\begin{cases} y_i|\theta_i & \sim \mathrm{N}\left(\theta_i, \sigma_m^2\right), \; i=1,...,n \\ \theta_i & \sim \mathrm{N}\left(z_i^\top\beta, \tau^2\right), \; i=1,...,n \end{cases}$$

By using Bayesian theorem, the posterior distribution of $\theta = (\theta_1,...,\theta_n)^\top$ is

$$\theta_i|y_i \sim \mathrm{N}\left(\frac{\sigma_m^2}{\sigma_m^2+\tau^2}z_i^\top\beta + \frac{\tau^2}{\sigma_m^2+\tau^2}y_i, \frac{\sigma_m^2\tau^2}{\sigma_m^2+\tau^2}\right) \equiv \mathrm{N}\left(\varpi z_i^\top\beta + \left(1-\varpi\right)y_i, \sigma_m^2\left(1-\varpi\right)\right)$$

where $\varpi = \frac{\sigma_m^2}{\sigma_m^2+\tau^2}$. Hence the Bayesian estimator is $\theta_i = \varpi z_i^\top\beta + \left(1-\varpi\right)y_i$.

1. The marginal posterior is $y_i \overset{\text{ind}}{\sim} \mathrm{N}\left(z_i^\top\beta, \sigma_m^2+\tau^2\right)$ by using standard Normal pdf properties. Essentially this is a regression problem, $y \sim \mathrm{N}\left(Z\beta, I_n\left(\sigma_m^2+\tau^2\right)\right)$ with log likelihood

$$\log\left(g(y)\right) = -\frac{n}{2}\log\left(\sigma_m^2+\tau^2\right) - \frac{(y-Z\beta)^\top(y-Z\beta)}{2\left(\sigma_m^2+\tau^2\right)} + \text{const}$$

The likelihood equations are

$$0 = \frac{\mathrm{d}}{\mathrm{d}(\beta,\tau)}\log\left(g(y)\right) = \begin{cases} 0 = \frac{(Z^\top Z)\beta - Z^\top y}{\sigma_m^2+\tau^2} \\ 0 = -\frac{n}{2(\sigma_m^2+\tau^2)} + \frac{S}{2(\sigma_m^2+\tau^2)^2} \end{cases}$$

where $S = \|y-Z\beta\|^2$. They are maximized at $\hat{\beta}_{\text{ML-II}} = \left(Z^\top Z\right)Z^\top Y$ and $\hat{\tau}^2_{\text{ML-II}} = \max\left(\frac{1}{n}\left\|y - Z\hat{\beta}_{\text{ML-II}}\right\|^2 - \sigma_m^2\right)$. Hence the EB of θ is

$$\hat{\theta}_i^{\text{EB}} = \hat{\varpi}_{\text{ML-II}}z_i^\top\beta + \left(1-\hat{\varpi}_{\text{ML-II}}\right)y_i, \quad \text{where} \quad \hat{\varpi}_{\text{ML-II}} = \frac{\sigma_m^2}{\sigma_m^2+\hat{\tau}^2_{\text{ML-II}}}$$

2. By using MoM, I get $\mathrm{E}\left(Y\right) = Z\beta \iff \mathrm{E}\left(\left(Z^\top Z\right) Z^\top Y\right) = \beta$. Hence, $\hat{\beta}_{\text{MoM}} = \left(Z^\top Z\right) Z^\top Y$ with $\mathrm{E}\left(\hat{\beta}_{\text{MoM}}\right) = \beta$. Moreover, because it is $\frac{S}{\sigma_m^2+\tau^2} \sim \chi^2_{n-d}$ I get

$$\mathrm{E}\left(\frac{S}{\sigma_m^2+\tau^2}\right) = n-d \implies \mathrm{E}\left(\frac{S}{n-d} - \sigma_m^2\right) = \tau^2_{\text{MoM}}$$

So $\frac{S}{n-d} - \sigma_m^2$