AMS 221 Homework 4

Diana Gerardo

March 4, 2019

1. (Problem 10.1) A forecaster must announce probabilites $\mathbf{q} = (q_1, ..., q_J)$ for the events $\theta_1, ..., \theta_J$. These events form a partition: that is, one and only one of them will occur. The forecaster will be scored based on the scoring rule

$$s(\theta_j, \mathbf{q}) = \sum_{i=1}^{J} |\mathbf{q}_i - 1_{i=j}|$$

Here $1_{i=j}$ is 1 if $i = j$ and 0 otherwise. Let $\pi = (\pi_1, ..., \pi_J)$ represent the forecaster's own probability for the events $\theta_1, ..., \theta_J$. Show that this scoring rule is not proper. That is, show that there exists a vector $\mathbf{q} \neq \boldsymbol{\pi}$ such that

$$\sum s(\theta_j, \mathbf{q}) \pi_j < \sum s(\theta_j, \boldsymbol{\pi}) \pi_j$$

Because you are looking for a counterexample, it is okay to consider a simplified version of the problem, for example by picking a small J.

Solution:

By definition a scoring rule is "proper" if the expected loss is minimized by the scorer's personal belief:

$$\inf_{q \in Q} S(\mathbf{q}) = S(\boldsymbol{\pi})$$

Thus we want to show that there exists a vector $\mathbf{q} \neq \boldsymbol{\pi}$ such that

$$\sum s(\theta_j, \mathbf{q}) \pi_j > \sum s(\theta_j, \boldsymbol{\pi}) \pi_j$$

Suppose $J = 2$, $q_1 = q$, $q_2 = (1-q)$, $\pi_1 = \pi$, and $\pi_2 = 1-\pi$. Then

$$\begin{aligned}
s(\theta_1, \mathbf{q}) &= \sum_{i=1}^{2} |\mathbf{q}_i - 1_{i=1}| \\
&= |q_1 - 1_{1=1}| + |q_2 - 1_{2=1}| \\
&= |q-1| + |1-q| \\
&= 2|1-q| \\
&= 2(1-q) \text{ since } 0 < q < 1 \\
s(\theta_2, \mathbf{q}) &= \sum_{i=1}^{2} |\mathbf{q}_i - 1_{i=2}| \\
&= |q_1 - 1_{1=2}| + |q_2 - 1_{2=2}| \\
&= |q| + |1-q-1| \\
&= 2|q| \\
&= 2q
\end{aligned}$$

Then

$$\begin{aligned}
\sum_{j=1}^{2} s(\theta_j, \mathbf{q}) \pi_j &= s(\theta_1, \mathbf{q}) \pi_1 + s(\theta_2, \mathbf{q}) \pi_2 \\
&= 2\pi(1-q) + 2q(1-\pi) \\
&= 2\pi - 2\pi q + 2q - 2\pi q \\
&= 2\pi + 2q - 4\pi q
\end{aligned}$$

Now

$$\begin{aligned}
s(\theta_1, \boldsymbol{\pi}) &= \sum_{i=1}^{2} |\boldsymbol{\pi}_i - 1_{i=1}| \\
&= 2(1-\pi) \text{ since } 0 < \pi < 1
\end{aligned}$$

$$\begin{aligned}
s(\theta_2, \boldsymbol{\pi}) &= \sum_{i=1}^{2} |\boldsymbol{\pi}_i - 1_{i=2}| \\
&= 2\pi
\end{aligned}$$

Then

$$\begin{aligned}\sum_{j=1}^{2} s\left(\theta_j, \boldsymbol{\pi}\right)\pi_j &= s\left(\theta_1, \boldsymbol{\pi}\right)\pi_1 + s\left(\theta_2, \boldsymbol{\pi}\right)\pi_2 \\ &= 2\pi\left(1-\pi\right) + 2\pi\left(1-\pi\right) \\ &= 4\pi\left(1-\pi\right)\end{aligned}$$

Let us see when

$$\sum s\left(\theta_j, \mathbf{q}\right)\pi_j < \sum s\left(\theta_j, \boldsymbol{\pi}\right)\pi_j$$

Observe

$$\begin{aligned}2\pi + 2q - 4\pi q &< 4\pi\left(1-\pi\right) \\ q\left(2-4\pi\right) &< 4\pi - 4\pi^2 - 2\pi \\ q &< 2\pi - 4\pi^2 \\ q &< \frac{\pi\left(2-4\pi\right)}{2-4\pi} \\ q &< \pi\end{aligned}$$

Thus in our example

$$\sum s\left(\theta_j, \mathbf{q}\right)\pi_j < \sum s\left(\theta_j, \boldsymbol{\pi}\right)\pi_j$$

holds true when $q < \pi$. $\square$

2. (Problem 10.6) Consider a sequence of independent binary events with probability of success 0.4. Evaluate the two terms in equation (10.8) for the following four forecasters:

Charles: always says 0.4

Mary: randomly chooses between 0.3 and 0.5

Qing: says either 0.2 or 0.3; when he says 0.2 it never rains, when he says 0.3 it always rains

Ana: follows this table

	Rain	No rain
$\pi = 0.3$	0.15	0.35
$\pi = 0.5$	0.25	0.25

Comment on the calibration and refinement of these forecasters.

Solution:

Suppose $x_i \sim Ber\,(0.4)$. Equation 10.8 is

$$BS = \sum_{\pi \in \prod} \nu\,(\pi)\,[\pi - \bar{x}\,(\pi)]^2 + \sum_{\pi \in \prod} \nu\,(\pi)\,[\bar{x}\,(\pi)\,(1 - \bar{x}\,(\pi))]$$

Charles always says $\pi = 0.4$. So we do not know whether $x_i = 0$ or 1 but we can assume $x_i \sim Ber(0.4)$, then $\nu\,(0.4) = \frac{n_k^\pi}{n_k} = \frac{n_k}{n_k} = 1$ and $\bar{x}\,(\pi = 0.4) = \frac{\sum \epsilon_i^\pi x_i}{n_k^\pi} = \frac{\sum x_i}{n_k} = 0.4$

$$\begin{aligned} BS_{\text{Charles}} &= [0.4 - 0.4]^2 + [0.4\,(1 - 0.4)] \\ &= 0 + 0.24 \\ &= 0.24 \end{aligned}$$

Charles's forecast always matches the relative frequency of rainy days which means it is well calibrated. Thus the first term in the BS score is zero. The second term is a measure of refinement of the forcaster. Since $\bar{x}$ is niether close to 0 or 1,Charles can closely be seen as the least refined forecaster. Alternatively we can say Charles exhibits close to zero sharpness.

Qing says $\pi = 0.2$ it never rains ($x_i = 0$) or $\pi = 0.3$ is always rains ($x_i = 1$). Remember that x_i is the indicator of rain on the i-th day. If Qing says $\pi = 0.3$ then $\bar{x}\,(0.3) = \frac{n_k^\pi 1}{n_k^\pi}$=1 and $\nu(0.3) = 0.4$. Thus if Qing says $\pi = 0.2$ then $\bar{x}\,(0.2) = \frac{0}{n_k^\pi} = 0$ and $\nu\,(0.2) = 1 - 0.4 = 0.6$. Then

$$\begin{aligned} BS_{\text{Qing}} &= (0.6)\,[0.2 - 0]^2 + (0.4)\,[0.3 - 1]^2 + (0.6)\,[1\,(1 - 1)] + (0.4)\,[0\,(1 - 0)] \\ &= 0.22 + 0 \\ &= 0.22 \end{aligned}$$

Qing's forcast is not well calibrated since 0.2 is far away from it's relative frequency 0; and 0.3 is far away from the relative frequency of 1. The second term is 0 which implies Qing exhibits perfect sharpness and is thus the best refined among the forcasters in this problem. This can be seen as Qing forcasted all rain events with one π and forcasted all no rain events with another different π.

According to Ana, $\nu\,(0.3) = 0.15 + 0.35 = 0.5$ with $\bar{x}\,(0.3) = \frac{0.15}{0.5} = 0.3$ and $\nu\,(0.5) = 0.25 + 0.25 = 0.5$

with $\bar{x}(0.5) = \frac{0.25}{0.5} = 0.5$. Then

$$\begin{aligned} BS_{\text{Ana}} &\approx (0.5)[0.3-0.3]^2 + (0.5)[0.5-0.5]^2 + (0.5)[0.3(1-0.3)] + (0.5)[0.5(1-0.5)] \\ &= 0 + 0.23 \\ &= 0.23 \end{aligned}$$

The first term is 0 which implies Ana's forcast is well calibrated. However each of her forcasts $\pi = 0.3$ and $\pi = 0.5$ contain both rain and no rain events which implies low sharpness and thus a not so refined forcaster.

Now going back to Mary: Mary randomly chooses between 0.3 and 0.5. In other words Mary chooses her forcast to be either 0.3 or 0.5 at random. Then $\nu(0.3)$ and $\nu(0.5)$ should approximately be about 0.5. We do not know whether $x_i = 0$ or 1 but we can assume $x_i \sim Ber(0.4)$, then $\bar{x}(0.3)$ and $\bar{x}(0.5)$ should approximately be 0.4. Thus

$$\begin{aligned} BS_{\text{Ana}} &= (0.5)[0.3-0.4]^2 + (0.5)[0.5-0.4]^2 + (0.5)[0.4(1-0.4)] + (0.5)[0.4(1-0.4)] \\ &= 0.01 + 0.24 \\ &= 0.25 \end{aligned}$$

Mary forecast is neither well calibrated nor has high sharpness. She is not a refined forcecaster since she is randomly choosing either 0.3 or 0.5 without any reasoning or strategy. Also by randomly choosing 0.3 or 0.5 she will never predict 0.4 precisely to have a well calibrated forcast. □

3. (Problem 13.3) Using simulation, approximate the distribution of $\mathcal{V}_x(\mathcal{E})$ in the example in Section 13.1.4. Suppose each observation costs \$230. Compute the marginal cost-effectiveness ratio for performing the experiment in the example (versus the satus quo of no experimentation). Use the negative of the loss as the measure of effectiveness.

Solution:

We have

$$\begin{aligned} X &\sim N(\theta, \sigma^2) \\ \theta &\sim N(\mu_0, \tau_0^2) \\ \theta | x &\sim N(\mu_x, \tau_x^2) \end{aligned}$$

where

$$\mu_x = \frac{\sigma^2}{\sigma^2 + \tau_0^2}\mu_0 + \frac{\tau_0^2}{\sigma^2 + \tau_0^2}x$$

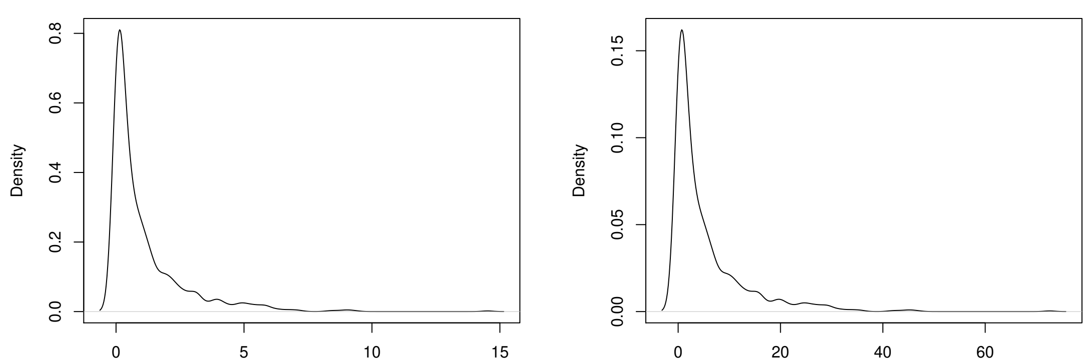
$$\tau_x = \frac{\sigma^2 \tau_0^2}{\sigma^2 + \tau_0^2}$$

And we have that the marginal distribution of x is $N\left(\mu_0, \sigma^2 + \tau_0^2\right)$. Let $\sigma^2 = 4$, $\mu_0 = 0$, and $\tau_0^2 = 1$. First we will simulate 1000 observed values from the marginal distribution of x. Then for each x, we compute the corresponding posterior mean $E\left[\theta|x\right] = \mu_x$. Finally, we can obtain 1000 different $\mathcal{V}_x\left(\mathcal{E}\right)$'s since

$$\begin{aligned} \mathcal{V}_x\left(\mathcal{E}\right) &= \left(E\left[\theta|x\right] - E\left[\theta\right]\right)^2 \\ &= \left(\mu_x - \mu_0\right)^2 \end{aligned}$$

Therefore,

$$\frac{\sigma^2 + \tau_0^2}{\tau_0^4}\mathcal{V}_x\left(\mathcal{E}\right) \sim \chi_1^2$$



(a) Density for $\mathcal{V}_x\left(\mathcal{E}\right)$

(b) $\frac{\sigma^2+\tau_0^2}{\tau_0^4}\mathcal{V}_x\left(\mathcal{E}\right) \sim \chi_1^2$

Figure 1

Suppose each observation costs \$230. We'll use the negative of the loss as the measure of effectiveness.

Then the the marginal cost-effectiveness ratio will be

$$-\frac{230}{(\theta-a)^2}\text{for performing the experiment}$$

$$0\text{for not performing the experiment}$$

Let $a = \mu_x$ where $x = -3.132269$ (the first simulated marginal x value). Here θ is an unknown value so suppose $\theta = 6$. Now suppose we perform the experiment, then $\mu_x = -0.6264538$ and the marginal cost-effectiveness ratio will be $-\frac{230}{(\theta-\mu_x)^2} = -5.238$. $\square$

4. (Problem 13.5) Consider an experiment consisting of a single Bernoulli observation, from a population with success probability θ. Consider a simple versus simple hypothesis testing situation in which $A = \{a_0, a_1\}$, $\Theta = \{0.50, 0.75\}$, and with utilities as shown in Table 13.7. Compute $\mathcal{V}(\mathcal{E}) = E[V(\pi_x)] - V(\pi)$.

Solution:

Our hypothesis is

$$H_0 : \theta = 0.50 \text{ vs } H_1 : \theta = 0.75$$

Set $\pi(H_0) = \pi(H_1) = 1/2$. Let a_i denote the action of accepting hypothesis H_i, $i = 0, 1$, and assume that the utility function is that shown in the following table

Actions	States of nature	
	H_0	H_1
$a_0 =$ accept H_0	0	-1
$a_1 =$ accept H_1	-3	0

The expected utilities for each decision are:

$$\begin{aligned}\mathcal{U}_\pi(a_0) &= u(a_0(H_0))\,\pi(H_0) + u(a_0(H_1))\,\pi(H_1)\\ &= 0(1/2) - 1(1/2)\\ &= -\frac{1}{2}\end{aligned}$$

$$\begin{aligned}\mathcal{U}_\pi(a_1) &= u(a_1(H_0))\,\pi(H_0) + u(a_1(H_1))\,\pi(H_1)\\ &= -3(1/2) + 0(1/2)\\ &= -\frac{3}{2}\end{aligned}$$

with $V(\pi) = \sup_{a \in \mathcal{A}} \mathcal{U}_\pi(a) = -1/2$ and the best decision is action $a^* = a_0$. Assume a conjugate prior $\theta \sim \text{Beta}(\alpha, \beta)$. Then the posterior distribution is

$$\pi(\theta|x) \propto \theta^{x+\alpha-1}(1-\theta)^{\beta+1-x} \sim \text{Beta}(x+\alpha,\ \beta+1-x)$$

Then we can compute the posterior probabilities of H_0 and H_1 as

$$\begin{aligned}
\pi(H_0|x) = \pi(\theta = 0.50|x) &= \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+x)\Gamma(\beta+1-x)}(0.50)^{\alpha+x-1}(0.50)^{\beta+1-x-1} \\
&= \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+x)\Gamma(\beta+1-x)}(0.50)^{\alpha+\beta-1}
\end{aligned}$$

$$\pi(H_1|x) = \pi(\theta = 0.75|x) = \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+x)\Gamma(\beta+1-x)}(0.75)^{\alpha+x-1}(0.25)^{\beta+1-x-1}$$

The expected utilities are

$$\begin{aligned}
\mathcal{U}_{\pi_x}(a_0) &= u(a_0(H_0))\pi_x(H_0) + u(a_0(H_1))\pi_x(H_1) \\
&= 0\pi(H_0|x) - 1\pi(H_1|x) \\
&= -\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+x)\Gamma(\beta+1-x)}(0.75)^{\alpha+x-1}(0.25)^{\beta+1-x-1}
\end{aligned}$$

$$\begin{aligned}
\mathcal{U}_{\pi_x}(a_1) &= u(a_1(H_0))\pi_x(H_0) + u(a_1(H_1))\pi_x(H_1) \\
&= -3\pi(H_0|x) + 0\pi(H_1|x) \\
&= (-3)\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+x)\Gamma(\beta+1-x)}(0.50)^{\alpha+\beta-1}
\end{aligned}$$

If $x = 0$

$$\begin{aligned}
\mathcal{U}_{\pi_x}(a_0) &= -\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha)\Gamma(\beta+1)}(0.75)^{\alpha-1}(0.25)^{\beta} \\
\mathcal{U}_{\pi_x}(a_1) &= (-3)\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha)\Gamma(\beta+1)}(0.50)^{\alpha+\beta-1}
\end{aligned}$$

If $x = 1$

$$\begin{aligned}
\mathcal{U}_{\pi_x}(a_0) &= -\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}(0.75)^{\alpha}(0.25)^{\beta-1} \\
\mathcal{U}_{\pi_x}(a_1) &= (-3)\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}(0.50)^{\alpha+\beta-1}
\end{aligned}$$

Assume $\beta = 1$. Then in R we let α equal a sequence of numbers between 1 and 10 with length 100. If $x = 0$ then $\mathcal{U}_{\pi_x}(a_0) > \mathcal{U}_{\pi_x}(a_1)$ if and only if $\alpha < 5.454545$. If $x = 1$ then $\mathcal{U}_{\pi_x}(a_0) > \mathcal{U}_{\pi_x}(a_1)$ if and only if $\alpha < 2.727273$.

<u>Case 1: $\alpha < 2.727273$, $\beta = 1$</u>

Under case 1,

$$\begin{aligned} V(\pi_x) &= \max\left\{\mathcal{U}_{\pi_x}(a_0), \mathcal{U}_{\pi_x}(a_1)\right\} \\ &= \begin{cases} \mathcal{U}_{\pi_x}(a_0) = -\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.75)^{\alpha-1}(0.25) & \text{if } x = 0 \\ \mathcal{U}_{\pi_x}(a_0) = -\frac{\Gamma(\alpha+2)}{\Gamma(\alpha+1)}(0.75)^{\alpha} & \text{if } x = 1 \end{cases} \end{aligned}$$

Therefore, the observed information is

$$\mathcal{V}_x(\mathcal{E}) = V(\pi_x) - \mathcal{U}_{\pi_x}(a_0) = 0 \text{ for } x = 0, 1$$

Thus

$$\mathcal{V}(\mathcal{E}) = E_x\left[V(\pi_x)\right] - V(\pi) = 0 + \frac{1}{2}$$

<u>Case 2: $2.727273 < \alpha < 5.454545$</u>

Under case 2,

$$\begin{aligned} V(\pi_x) &= \max\left\{\mathcal{U}_{\pi_x}(a_0), \mathcal{U}_{\pi_x}(a_1)\right\} \\ &= \begin{cases} \mathcal{U}_{\pi_x}(a_0) = -\frac{\Gamma(\alpha+2)}{\Gamma(\alpha+1)}(0.75)^{\alpha} & \text{if } x = 1 \\ \mathcal{U}_{\pi_x}(a_1) = (-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.50)^{\alpha} & \text{if } x = 0 \end{cases} \end{aligned}$$

Therefore, the observed information is

$$\begin{aligned} \mathcal{V}_x(\mathcal{E}) &= V(\pi_x) - \mathcal{U}_{\pi_x}(a_0) \\ &= \begin{cases} 0 & \text{if } x = 1 \\ (-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.50)^{\alpha} + \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.75)^{\alpha-1}(0.25) & \text{if } x = 0 \end{cases} \end{aligned}$$

So

$$\mathcal{V}(\mathcal{E}) = E_x\left[V(\pi_x)\right] - V(\pi)$$

where

$$E_x\left[V(\pi_x)\right] = \left[(-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.50)^{\alpha} + \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.75)^{\alpha-1}(0.25)\right] \cdot m(x)$$

The marginal of x is

$$\begin{aligned} m(x) &= \int f(x|\theta)\,\pi(\theta)\,d\theta \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int \theta^{\alpha+x-1}(1-\theta)^{1+\beta-x-1}\,d\theta \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(\alpha+x)\Gamma(1+\beta-x)}{\Gamma(\alpha+\beta+1)} \\ &= \frac{\Gamma(\alpha+x)\Gamma(1+\beta-x)}{\Gamma(a)\Gamma(\beta)(\alpha+\beta)} \\ &= \frac{\Gamma(\alpha+x)\Gamma(2-x)}{\Gamma(\alpha)(\alpha+1)} \text{ for } \beta = 1 \end{aligned}$$

Hence

$$\begin{aligned} \mathcal{V}(\mathcal{E}) &= E_x\left[V(\pi_x)\right] - V(\pi) \\ &= \frac{\Gamma(\alpha)\Gamma(2)}{\Gamma(\alpha)(\alpha+1)} \cdot \left[(-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.50)^{\alpha} + \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.75)^{\alpha-1}(0.25)\right] + \frac{1}{2} \\ &= -3\alpha(0.50)^{\alpha} + \alpha(0.75)^{\alpha-1}(0.25) + \frac{1}{2} \end{aligned}$$

<u>Case 3: $\alpha > 5.454545$</u>

Under case 3,

$$V(\pi_x) = \begin{cases} \mathcal{U}_{\pi_x}(a_1) = -\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.75)^{\alpha-1}(0.25) & \text{if } x = 0 \\ \mathcal{U}_{\pi_x}(a_1) = (-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha+1)}(0.50)^{\alpha} & \text{if } x = 1 \end{cases}$$

Then

$$\begin{aligned} \mathcal{V}_x(\mathcal{E}) &= V(\pi_x) - \mathcal{U}_{\pi_x}(a_0) \\ &= \begin{cases} (-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.50)^{\alpha} + \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.75)^{\alpha-1}(0.25) & \text{if } x = 0 \\ (-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha+1)}(0.50)^{\alpha} + \frac{\Gamma(\alpha+2)}{\Gamma(\alpha+1)}(0.75)^{\alpha} & \text{if } x = 1 \end{cases} \end{aligned}$$

Therefore,

$$
\begin{aligned}
\mathcal{V}(\mathcal{E}) &= E_x\left[V\left(\pi_x\right)\right] - V(\pi) \\
&= \frac{\Gamma(\alpha)\Gamma(2)}{\Gamma(\alpha)(\alpha+1)} \cdot \left[(-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.50)^{\alpha} + \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(2)}(0.75)^{\alpha-1}(0.25)\right] \\
&\quad + \frac{\Gamma(\alpha+1)\Gamma(1)}{\Gamma(\alpha)(\alpha+1)} \cdot \left[(-3)\frac{\Gamma(\alpha+2)}{\Gamma(\alpha+1)}(0.50)^{\alpha} + \frac{\Gamma(\alpha+2)}{\Gamma(\alpha+1)}(0.75)^{\alpha}\right] + \frac{1}{2} \\
&= -3\alpha(0.50)^{\alpha} + \alpha(0.75)^{\alpha-1}(0.25) - 3\alpha(0.50)^{\alpha} + \alpha(0.75)^{\alpha} + \frac{1}{2} \\
&= -6\alpha(0.50)^{\alpha} + (1.33)\alpha(0.75)^{\alpha} + \frac{1}{2} \quad \square
\end{aligned}
$$

5. (Problem 13.11) Consider an experiement $\mathcal{E}$ that consists of observing n conditionally independent random variables $x_1, ..., x_n$ with $x_i \sim \mathcal{N}(\theta, \sigma^2)$, with σ known. Suppose also that a priori $\theta \sim \mathcal{N}(\mu_0, \tau_0^2)$. Show that

$$\mathcal{I}(\mathcal{E}) = \frac{1}{2}\log\left(1 + n\frac{\tau_0^2}{\sigma^2}\right)$$

You can use facts about conjugate priors from Bernardo and Smith (2000) or Berger (1985). However, please rederive $\mathcal{I}$.

Solution:

We have likelihood $x_i \sim \mathcal{N}(\theta, \sigma^2)$ for $i = 1, ..n$ which implies $\bar{x} \sim \mathcal{N}(\theta, \sigma^2/n)$ and we have a prior, $\theta \sim \mathcal{N}(\mu_0, \tau_0^2)$. Thus our marginal distribution of $\bar{x}$ is $\mathcal{N}(\mu_0,\ \sigma^2/n + \tau_0^2)$. By definition of the Lindley information we have

$$\mathcal{I}(\mathcal{E}) = E_{\theta}\left[E_{x|\theta}\left[\log\left(\frac{f(\bar{x}|\theta)}{m(\bar{x})}\right)\right]\right]$$

So the equations for the likelihood and the prior predictive distribution are

$$
\begin{aligned}
f(\bar{x}|\theta) &= \left(\frac{n}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{n}{2\sigma^2}\left(\bar{x}^2 - 2\bar{x}\theta + \theta^2\right)\right\} \\
m(\bar{x}) &= \left(\frac{n}{2\pi\left(\sigma^2 + n\tau^2\right)}\right)^{1/2} \exp\left\{-\frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\bar{x}^2 - 2\bar{x}\mu_0 + \mu_0^2\right)\right\}
\end{aligned}
$$

Then we take the log,

$$
\begin{aligned}
\log f(\bar{x}|\theta) &= \frac{1}{2}\left[\log\left(\frac{n}{2\pi}\right) - \log\left(\sigma^2\right)\right] - \frac{n}{2\sigma^2}\left(\bar{x}^2 - 2\bar{x}\theta + \theta^2\right) \\
\log m(\bar{x}) &= \frac{1}{2}\left[\log\left(\frac{n}{2\pi}\right) - \log\left(\sigma^2 + n\tau^2\right)\right] - \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\bar{x}^2 - 2\bar{x}\mu_0 + \mu_0^2\right)
\end{aligned}
$$

Then the log of the ratio is

$$\begin{aligned}
\log\left(\frac{f\left(\bar{x}|\theta\right)}{m\left(\bar{x}\right)}\right) &= \log f\left(\bar{x}|\theta\right) - \log m\left(\bar{x}\right) \\
&= \frac{1}{2}\left[\log\left(\frac{n}{2\pi}\right) - \log\left(\sigma^2\right)\right] - \frac{n}{2\sigma^2}\left(\bar{x}^2 - 2\bar{x}\theta + \theta^2\right) \\
&\quad - \frac{1}{2}\left[\log\left(\frac{n}{2\pi}\right) - \log\left(\sigma^2 + n\tau^2\right)\right] + \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\bar{x}^2 - 2\bar{x}\mu_0 + \mu_0^2\right) \\
&= \frac{1}{2}\left[\log\left(\sigma^2 + n\tau^2\right) - \log\left(\sigma^2\right)\right] \\
&\quad \underbrace{-\frac{n}{2\sigma^2}\left(\bar{x}^2 - 2\bar{x}\theta + \theta^2\right) + \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\bar{x}^2 - 2\bar{x}\mu_0 + \mu_0^2\right)}_{T}
\end{aligned}$$

So taking the expectation of T with respect $\bar{x}$,

$$E_{x|\theta}\left[T\right] = -\frac{n}{2\sigma^2}\left(E_{x|\theta}\left[\bar{x}^2\right] - 2E_{x|\theta}\left[\bar{x}\right]\theta + \theta^2\right) + \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(E_{x|\theta}\left[\bar{x}^2\right] - 2E_{x|\theta}\left[\bar{x}\right]\mu_0 + \mu_0^2\right)$$

Where

$$\begin{aligned}
E_{x|\theta}\left[\bar{x}\right] &= \theta \\
Var_{x|\theta}\left(\bar{x}\right) &= \frac{\sigma^2}{n} \\
E_{x|\theta}\left[\bar{x}^2\right] &= Var_{x|\theta}\left(\bar{x}\right) + E_{x|\theta}\left[\bar{x}\right]^2 = \frac{\sigma^2}{n} + \theta^2
\end{aligned}$$

So

$$\begin{aligned}
E_{x|\theta}\left[T\right] &= -\frac{n}{2\sigma^2}\left(\frac{\sigma^2}{n} + \theta^2 - 2\theta\theta + \theta^2\right) + \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\frac{\sigma^2}{n} + \theta^2 - 2\theta\mu_0 + \mu_0^2\right) \\
&= -\frac{1}{2} + \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\frac{\sigma^2}{n} + \theta^2 - 2\theta\mu_0 + \mu_0^2\right)
\end{aligned}$$

Then

$$
\begin{aligned}
E_\theta\left[E_{x|\theta}\left[T\right]\right] &= -\frac{1}{2} + \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\frac{\sigma^2}{n} + E_\theta\left[\theta^2\right] - 2E_\theta\left[\theta\right]\mu_0 + \mu_0^2\right) \\
&= -\frac{1}{2} + \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\frac{\sigma^2}{n} + \tau_0^2 + \mu_0^2 - 2\mu_0\mu_0 + \mu_0^2\right) \\
&= -\frac{1}{2} + \frac{n}{2\left(\sigma^2 + n\tau^2\right)}\left(\frac{\sigma^2}{n} + \tau_0^2\right) \\
&= -\frac{1}{2} + \frac{1}{2} \\
&= 0
\end{aligned}
$$

Therefore the Lindley information is

$$
\begin{aligned}
\mathcal{I}\left(\mathcal{E}\right) &= E_\theta\left[E_{x|\theta}\left[\log\left(\frac{f\left(\bar{x}|\theta\right)}{m\left(\bar{x}\right)}\right)\right]\right] \\
&= \frac{1}{2}\left[\log\left(\sigma^2 + n\tau_0^2\right) - \log\left(\sigma^2\right)\right] + E_\theta\left[E_{x|\theta}\left[T\right]\right] \\
&= \frac{1}{2}\left[\log\left(\frac{\sigma^2 + n\tau_0^2}{\sigma^2}\right)\right] \\
&= \frac{1}{2}\left[\log\left(1 + \frac{n\tau_0^2}{\sigma^2}\right)\right] \qquad \square
\end{aligned}
$$