Homework 3

Shao-Ting Chiu (UIN:433002162)

9/27/22

Description

- Course: STAT638, 2022 Fall

  Do the following exercises in Hoff: 3.8, 3.9, 3.14.

  In Exercise 3.9, you should be able to avoid "brute-force" integration by exploiting the fact that the Galenshore distribution is a proper distribution, meaning that the density of the Galenshore(a,b) distribution integrates to one for any $a, b > 0$.

  For 3.14(b), note that $p_U(\theta)$ is proportional to the density of a known distribution.

  Please note that while there are only 3 problems in this assignment, some of them are fairly challenging. So please don't wait too long to get started on this assignment.

- Deadline: `Sept. 27, 12:01pm`

Computational Enviromnent Setup

Third-party libraries

```
%matplotlib inline
import sys # system information
import matplotlib # plotting
import scipy # scientific computing
import random
import pandas as pd # data managing
from scipy.special import comb
```

```python
from scipy import stats as st
from scipy.special import gamma
import numpy as np
import matplotlib.pyplot as plt
# Matplotlib setting
plt.rcParams['text.usetex'] = True
matplotlib.rcParams['figure.dpi']= 300
```

Version

```python
print(sys.version)
print(matplotlib.__version__)
print(scipy.__version__)
print(np.__version__)
print(pd.__version__)
```

```
3.8.12 (default, Oct 22 2021, 18:39:35)
[Clang 13.0.0 (clang-1300.0.29.3)]
3.3.1
1.5.2
1.19.1
1.1.1
```

Problem 3.8

> Coins: Diaconis and Ylvisaker (1985) suggest that coins spun on a flat surface display long-run frequencies of heads that vary from coin to coin. About 20% of the coins behave symmetrically, whereas the remaining coins tend to give frequencies of $\frac{1}{3}$ or $\frac{2}{3}$.

Let θ be the priobability of tossing head.[1]

[1] This solution is referred to the lecutre note about mixture priors. URL: http://www.mas.ncl.ac.uk/~nmf16/teaching/mas3301/week11.pdf

(a)

> Based on the observations of Diaconis and Ylvisaker (1985), use an appropriate mixture of beta distributions as a prior distribution for θ, the long-run frequency of heads for a particular coin. Plot your prior.

Let the prior probability $p_i(\theta)$ be a mixture of $Beta(a_i, b_i)$ with $i = [1, 2, 3]$, and coeifficient $k = [k_1, k_2, k_3]$ with $\sum_{i=1}^{3} k_j = 1$.

Let the prior probabiility be

$$
\begin{aligned}
p(\theta) &= \sum_{i=1}^{3} k_i p_i(\theta) && (1)\\
&= k_1 p_1(\theta) + k_2 p_2(\theta) + k_3 p_3(\theta) && (2)\\
&= 0.2 \times Beta(\theta, a_1, b_1) + 0.4 \times Beta(\theta, a_2, b_2) + 0.4 \times Beta(\theta, a_3, b_3) && (3)\\
&= 0.2 \times Beta(\theta, 3, 3) + 0.4 \times Beta(\theta, 2, 4) + 0.4 \times Beta(\theta, 4, 2) && (4)
\end{aligned}
$$

The distribution is shown in Figure 1.

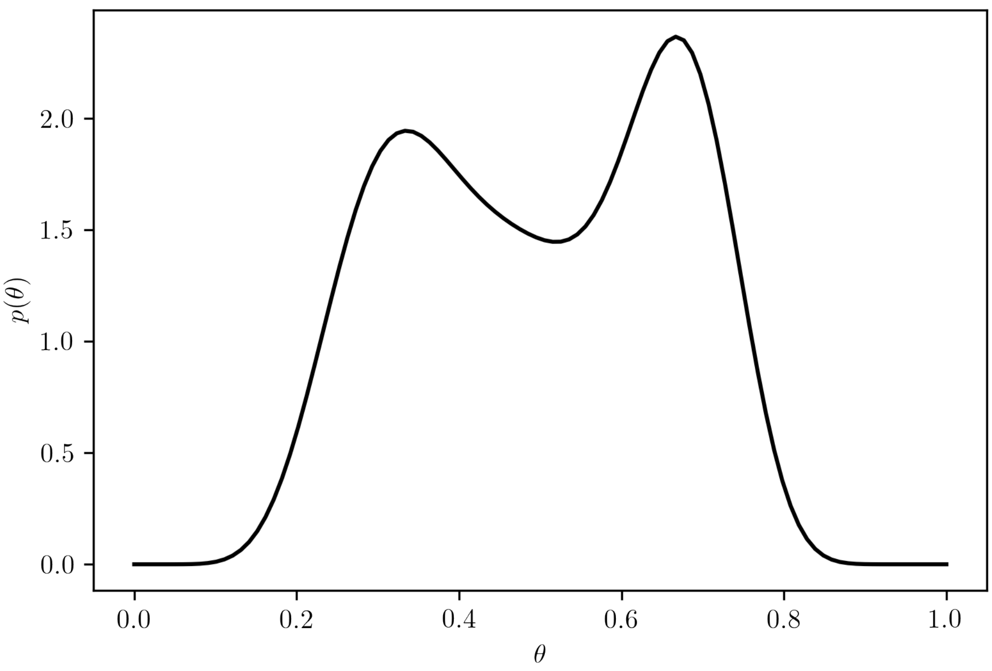

Figure 1: Designed mixture prior.

Table 1: Satistics of the flipping coin experiment

	Properties	Values
0	N	100
1	Number of heads (y=1)	12
2	Number of tails	88

(b)

> Choose a single coin and spin it at least 50 times. Record the number of heads obtained. Report the year and denomination of the coin.

Let $n > 50$ be the number of flips, and x be the number of heads obtained.

```
# A single psudo coin with unknown probability of flipping head
class PseudoCoin:
  def __init__(self, random_state=202209):
    np.random.seed(random_state)
    self.random_state = random_state
    self.ph = np.random.rand()
    self.rv = st.bernoulli(self.ph)

  def flips(self, n):
    return self.rv.rvs(n, random_state=self.random_state)

# parameters setting
n = 100 # number of flips
coin = PseudoCoin()

# Experiment
rs = coin.flips(n)

# Results
print(rs)
```

```
[0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0
 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0]
```

(c)

Compute your posterior for θ, based on the information obtained in (b)

For $i = \{1, 2, 3\}$, the posterior probability of single distribution is

$$\begin{aligned}
p_i(\theta|y) &= \frac{p_i(\theta)p(y|\theta)}{\underbrace{\int_{\theta\in[0,1]} p_i(\theta)p(y|\theta)d\theta}_{=C_j}} && (5)\\
&= \frac{Beta(\theta, a_i, b_i)\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\underbrace{\int_0^1 Beta(\theta, a_i, b_i)\binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta}_{=C_j}} && (6)\\
&= \frac{\frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta} && (7)\\
&= \frac{\frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\theta^{(a_i-1)}(1-\theta)^{b_i-1}\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\theta^{(a_i-1)}(1-\theta)^{b_i-1}\binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta} && (8)\\
&= Beta(\theta, a_i + y, b_i + n - y) && (9)
\end{aligned}$$

The posterior distribution of individual prior is

$$\begin{aligned}
p_1(\theta|y) &= Beta(\theta, a_1 + y, b_1 + n - y) && (10)\\
&= Beta(\theta, 3 + 12, 3 + 100 - 12) = Beta(\theta, 15, 91) && (11)\\
p_2(\theta|y) &= Beta(\theta, a_2 + y, b_2 + n - y) && (12)\\
&= Beta(\theta, 2 + 12, 4 + 100 - 12) = Beta(\theta, 14, 92) && (13)\\
p_3(\theta|y) &= Beta(\theta, a_3 + y, b_3 + n - y) && (14)\\
&= Beta(\theta, 4 + 12, 2 + 100 - 12) = Beta(\theta, 16, 90) && (15)\\
& && (16)
\end{aligned}$$

$$C_j = \int_0^1 \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\theta^{(a_i-1)}(1-\theta)^{b_i-1}\binom{n}{y}\theta^y(1-\theta)^{n-y}d\theta \tag{17}$$

$$= \int_0^1 \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\frac{\Gamma(n)}{\Gamma(y)\Gamma(n-y)}\theta^{(a_i-1)}(1-\theta)^{b_i-1}\theta^y(1-\theta)^{n-y}d\theta \tag{18}$$

$$= \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\frac{\Gamma(n)}{\Gamma(y)\Gamma(n-y)}\int_0^1 \theta^{(a_i-1)}(1-\theta)^{b_i-1}\theta^y(1-\theta)^{n-y}d\theta \tag{19}$$

$$= \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\frac{\Gamma(n)}{\Gamma(y)\Gamma(n-y)}\int_0^1 \theta^{a_i+y-1}(1-\theta)^{b_i+n-y-1}d\theta \tag{20}$$

$$= \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\frac{\Gamma(n)}{\Gamma(y)\Gamma(n-y)}\frac{\Gamma(a_i+y)\Gamma(b_i+n-y)}{\Gamma(a_i+b_i+n)} \tag{21}$$

$$C_1 = \frac{\Gamma(3+3)}{\Gamma(3)\Gamma(3)}\frac{\Gamma(100)}{\Gamma(12)\Gamma(88)}\frac{\Gamma(3+12)\Gamma(3+100-12)}{\Gamma(3+3+100)} \tag{22}$$

$$= \frac{\Gamma(6)}{\Gamma(3)\Gamma(3)}\frac{\Gamma(100)}{\Gamma(12)\Gamma(88)}\frac{\Gamma(15)\Gamma(91)}{\Gamma(106)} \tag{23}$$

$$C_2 = \frac{\Gamma(2+4)}{\Gamma(2)\Gamma(4)}\frac{\Gamma(100)}{\Gamma(12)\Gamma(88)}\frac{\Gamma(2+12)\Gamma(4+100-12)}{\Gamma(2+4+100)} \tag{24}$$

$$= \frac{\Gamma(6)}{\Gamma(2)\Gamma(4)}\frac{\Gamma(100)}{\Gamma(12)\Gamma(88)}\frac{\Gamma(14)\Gamma(92)}{\Gamma(106)} \tag{25}$$

$$C_3 = \frac{\Gamma(4+2)}{\Gamma(4)\Gamma(2)}\frac{\Gamma(100)}{\Gamma(12)\Gamma(100-12)}\frac{\Gamma(4+12)\Gamma(2+100-12)}{\Gamma(4+2+100)} \tag{26}$$

$$= \frac{\Gamma(6)}{\Gamma(4)\Gamma(2)}\frac{\Gamma(100)}{\Gamma(12)\Gamma(88)}\frac{\Gamma(16)\Gamma(90)}{\Gamma(106)} \tag{27}$$

Let

$$C_i^* = \frac{\Gamma(a_i+y)\Gamma(b_i+n-y)}{\Gamma(a_i)\Gamma(b_i)} \tag{28}$$

$$k_j^{(1)} = \frac{k_j^{(0)}C_j}{\sum_{i=1}^J k_i^{(0)}C_i} \tag{29}$$

$$= \frac{k_j^0 C_j^*}{0.2\times\underbrace{\frac{\Gamma(15)\Gamma(61)}{\Gamma(3)\Gamma(3)}}_{C_1^*}+0.4\times\underbrace{\frac{\Gamma(14)\Gamma(92)}{\Gamma(2)\Gamma(4)}}_{C_2^*}+0.4\times\underbrace{\frac{\Gamma(16)\Gamma(90)}{\Gamma(4)\Gamma(2)}}_{C_3^*}} \tag{30}$$

where $j \in \{1, 2, 3\}$.

```
def gammafrac(a,b,c,d,):
    return ((gamma(a)*gamma(b))**-1) * gamma(c) * gamma(d)

c1 = gammafrac(3,3,15,61)
c2 = gammafrac(2,4,14,92)
c3 = gammafrac(4,2,16,90)
C = c1 + c2 + c3

pd.DataFrame({"Variables": ["C1*", "C2*", "C3*", "$k^{1}_{1}$", "$k^{1}_{2}$", "$k^{1}_{3}
```

	Variables	Values
0	C1*	1.813524e+92
1	C2*	1.403157e+149
2	C3*	3.597838e+147
3	k^{1}_{1}	1.260148e-57
4	k^{1}_{2}	9.750000e-01
5	k^{1}_{3}	2.500000e-02

$$
\begin{aligned}
p(\theta|y) &= \sum_{i=1}^{3} k_i^{(1)} p_i(\theta|y) && (31)\\
&= k_1^{(1)} p_1(\theta|y) + k_2^{(1)} p_2(\theta|y) + k_3^{(1)} p_3(\theta|y) && (32)\\
&= 1.260148 \times 10^{-57} \times Beta(\theta, 15, 91) && (33)\\
&+ 9.750000 \times 10^{-01} \times Beta(\theta, 14, 92) && (34)\\
&+ 2.500000e \times 10^{-02} \times Beta(\theta, 16, 90) && (35)
\end{aligned}
$$

(d)

Repeat (b) and (c) for a different coin, but possibly using a prior for θ that includes some information from the first coin. Your choice of a new prior may be informal, but needs to be justified. How the results from the first experiment influence your prior for the θ of the second coin may depend on whether or not the two coins have the same denomination, have a similar year, etc. Report the year and denomination of this coin.

```
# pick another coin
coin2 = PseudoCoin(random_state=202210)

# parameters setting
```

Table 2: Satistics of the flipping coin experiment

	Properties	Values
0	N	100
1	Number of heads (y=1)	51
2	Number of tails	49

```
n2 = 100 # number of flips

# Experiment
rs2 = coin2.flips(n2)

# Results
print(rs2)
```

```
[1 1 0 1 1 0 0 1 1 1 0 1 1 0 1 0 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 1 1 0 1 0 0
 1 1 0 0 0 1 1 0 1 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 1 1 1 0 0 0 1 1 1 0 1 0
 1 0 1 1 0 0 1 0 0 1 0 0 1 1 1 1 1 0 1 0 0 1 1 0 0 1]
```

$$
\begin{aligned}
p^1(\theta) &= 1.260148 \times 10^{-57} \times Beta(\theta, 15, 91) &(36)\\
&+ 9.750000 \times 10^{-01} \times Beta(\theta, 14, 92) &(37)\\
&+ 2.500000e \times 10^{-02} \times Beta(\theta, 16, 90) &(38)
\end{aligned}
$$

Apply Equation 28,

$$
\begin{aligned}
C_1^* &= \frac{\Gamma(15+51)\Gamma(91+100-51)}{\Gamma(15)\Gamma(91)} && (39)\\
&= \frac{\Gamma(66)\Gamma(140)}{\Gamma(15)\Gamma(91)} && (40)\\
C_2^* &= \frac{\Gamma(14+51)\Gamma(92+100-51)}{\Gamma(14)\Gamma(92)} && (41)\\
&= \frac{\Gamma(65)\Gamma(141)}{\Gamma(14)\Gamma(92)} && (42)\\
C_3^* &= \frac{\Gamma(16+51)\Gamma(90+100-51)}{\Gamma(16)\Gamma(90)} && (43)\\
&= \frac{\Gamma(67)\Gamma(139)}{\Gamma(16)\Gamma(90)} && (44)\\
& && (45)
\end{aligned}
$$

	Variables	Values
0	C1*	6.123054e+180
1	C2*	2.028941e+180
2	C3*	1.744410e+181
3	k^{1}_{1}	2.392183e-01
4	k^{1}_{2}	7.926761e-02
5	k^{1}_{3}	6.815141e-01

$$
\begin{aligned}
p^2(\theta|y_2) &= \sum_{i=1}^{3} k_i^{(2)} p_i^{(2)}(\theta|y_2) && (46)\\
&= k_1^{(2)} p_1^{(2)}(\theta|y_2) + k_2^{(2)} p_2^{(2)}(\theta|y_2) + k_3^{(2)} p_3^{(2)}(\theta|y_2) && (47)\\
&= k_1^{(2)} Beta(15+51, 91+49) && (48)\\
&+ k_2^{(2)} Beta(14+51, 92+49) && (49)\\
&+ k_3^{(2)} Beta(16+51, 90+49) && (50)\\
&= 2.32 \times 10^{-1} \times Beta(66, 140) && (51)\\
&+ 7.93 \times 10^{-2} \times Beta(65, 141) && (52)\\
&+ 6.82 \times 10^{-1} \times Beta(67, 139) && (53)\\
& && (54)
\end{aligned}
$$

Problem 3.9

Galenshore distribution: An unknown quantity Y has a Galenshore(α, θ) distribution if its density is given by

$$p(y) = \frac{2}{\Gamma(a)}\theta^{2a}y^{2a-1}e^{-\theta^2 y^2}$$

for $y > 0$, $\theta > 0$ and $a > 0$. Assume for now that a is known. For this density,

$$E[Y] = \frac{\Gamma(a+\frac{1}{2})}{\theta\Gamma(a)}, \quad E[Y^2] = \frac{a}{\theta^2}$$

(a)

Identify a class of conjugate prior densities for θ. Plot a few members of this class of densities.

Identifying Galenshore distribution belongs to exponential family

$$\begin{aligned} p(y|\theta) &= \frac{2}{\Gamma(a)}y^{2a-1}\theta^{2a}e^{-\theta^2 y^2} && (55)\\ &= \left(\frac{2}{\Gamma(a)}y^{2a-1}\right)\left(\theta^2\right)^a\left(e^{-\theta^2 y^2}\right) && (56)\\ & && (57)\end{aligned}$$

- $\phi(\theta) = \theta^2$
- $h(y) = \frac{2}{\Gamma(a)}y^{2a-1}$
- $c(\phi) = \phi^a$
- $t(y) = -y^2$

$$p(y|\phi) = \underbrace{(\frac{2}{\Gamma(a)}y^{2a-1})}_{=h(y)}\underbrace{(\phi^a)}_{=c(\phi)}\exp(\phi\cdot\underbrace{(-1)\cdot y^2}_{=t(y)})$$

Derive the posterior distribution

$$\begin{aligned} p(\phi|n_0,t_0) &= \kappa(n_0,t_0)c(\phi)^{n_0}\exp(n_0 t_0 \phi) && (58)\\ &= \kappa(n_0,t_0)\phi^{an_0}\exp(n_0 t_0\phi) && (59)\\ p(\theta^2|n_0,t_0) &= \kappa(n_0,t_0)\theta^{2an_0}e^{(n_0 t_0 \theta^2)} && (60)\\ & && (61)\end{aligned}$$

Apply change of variables

Let $f(\phi) = \sqrt{\theta} = \theta$ (f is a monotonous function), and $\phi(\theta) = \theta^2$

$$p_\phi(\phi) = p_\theta(f(\phi)) \times |\frac{df}{d\phi}| \tag{62}$$
$$\kappa(n_0, t_0)c(\phi)^{n_0} \exp(n_0 t_0 \phi) = p_\theta(\theta) \times (\frac{1}{2\sqrt{\theta}}) \tag{63}$$
$$\kappa(n_0, t_0)c(\theta^2)^{n_0} \exp(n_0 t_0 \theta^2) = p_\theta(\theta) \times (\frac{1}{2\sqrt{\theta}}) \tag{64}$$
$$p_\theta(\theta) = \kappa(n_0, t_0)\theta^{2a} e^{(n_0 t_0 \theta^2)} \tag{65}$$

$$\begin{aligned} p_\theta(\theta) &= p_\phi(\phi(\theta)) \times |\frac{d\phi(\theta)}{d\theta}| && (66)\\ &\propto \kappa(n_0, t_0)c(\theta^2)^{n_0} \exp(n_0 t_0 \theta^2) \times 2\theta && (67)\\ &\propto \theta^{2an_0+1} e^{(n_0 t_0 \theta^2)} && (68)\\ & && (69)\end{aligned}$$

To avoid alias, let Galenshore pdf function be

$$p_{X\sim Galenshore}(x; a, z) = \frac{2}{\Gamma(a)} z^{2a} x^{2a-1} e^{-z^2 x^2} \tag{70}$$

Combining Equation 70 together,

$$\begin{aligned} p_\theta(\theta|n_0, t_0) &\propto dGalenshore(\theta; an_0 + 1, \sqrt{-n_0 t_0}) && (71)\\ & && (72)\end{aligned}$$

$\because t_0 = -y^2 \therefore - n_0 t_0 = n_0 y^2 \geq 0.$

(b)

> Let $Y_1, \ldots, Y_n \sim$ *i.i.d.* Galenshore(a,θ). Find the posterior distribution of θ given $Y_1, \ldots, Y_n$, using a prior from your conjugate class.

Use the formula described in Hoff (2009, vol. 580, sec. 3.3).

- $n^{(1)} = n_0 + n$
- $t^{(1)} = n_0 t_0 + n\bar{t}(y)$

$$
\begin{aligned}
p(\phi|Y) &\propto Galenshore(an' + 1, \sqrt{-n't'}) && (73)\\
&\propto p(\phi|n_0 + n, n_0 t_0 + n\bar{t}(y)) && (74)\\
p(\theta|Y) &\propto dGalenshore(\theta; a(n_0+n)+1, \sqrt{(n_0+n)(n_0 t_0 + n\bar{t}(y))}) && (75)
\end{aligned}
$$

where $\bar{t}(y) = \frac{\sum t(y_i)}{n}$

(c)

> Write down $\frac{p(\theta_a|Y_1,\ldots,Y_n)}{p(\theta_b|Y_1,\ldots,Y_n)}$ and simplify. Identify a sufficient statistics.

Because $t(y) = -y^2$ is the sufficient statistic of the exponential family, the sufficient statistics for $Y_1, \cdots, Y_n$ is

$$\bar{t}(y) = \frac{\sum t(y_i)}{n} = \frac{\sum y_i^2}{n}$$

(d)

> Determine $E[\theta|y_1, \ldots, y_n]$.

Use the posterior distribution derived in (b).

$$
\begin{aligned}
p(\theta|Y) &\propto dGalenshore(\theta; \underbrace{a(n_0+n)+1}_{a^{(1)}}, \underbrace{\sqrt{(n_0+n)(n_0 t_0 + n\bar{t}(y))}}_{\theta^{(1)}}) && (76)\\
E[\theta|Y] &= \frac{\Gamma(a(n_0+n)+\frac{3}{2})}{\sqrt{(n_0+n)(n_0 t_0 + n\bar{t}(y))}\Gamma(a(n_0+n)+1)} && (77)
\end{aligned}
$$

(e)

Determine the form of the posterior predictive density $p(\tilde{y}|y_1, \dots, y_n)$.

$$p(\tilde{y}|Y) = \int_\theta p(\tilde{y}|\theta)p(\theta|Y)d\theta \tag{78}$$

$$\tag{79}$$

- $p(\tilde{y}|\theta) = \frac{2}{\Gamma(a)}\theta^{2a}\tilde{y}^{2a-1}e^{-\theta^2\tilde{y}^2}$
- $p(\theta|Y) \propto dGalenshore(\theta; \underbrace{a(n_0+n)+1}_{a_{(1)}}, \underbrace{\sqrt{(n_0+n)(n_0t_0+n\bar{t}(y))}}_{b_{(1)}})$
  - $p(\theta|Y) = \frac{2}{\Gamma(a_{(1)})}b_{(1)}^{2a_{(1)}}\theta^{2a_{(1)}-1}e^{-b_{(1)}^2\theta^2}$

$$p(\tilde{y}|Y) = \int_\theta \frac{2}{\Gamma(a)}\theta^{2a}\tilde{y}^{2a-1}e^{-\theta^2\tilde{y}^2} \times \frac{2}{\Gamma(a_{(1)})}b_{(1)}^{2a_{(1)}}\theta^{2a_{(1)}-1}e^{-b_{(1)}^2\theta^2}d\theta \tag{80}$$

$$= \frac{4\tilde{y}^{2a-1}b_{(1)}^{2a_{(1)}}}{\Gamma(a)\Gamma(a_{(1)})}\int_\theta \theta^{2a+2a_{(1)}-1}e^{-\theta^2\tilde{y^2}-b_{(1)}^2\theta^2}d\theta \tag{81}$$

$$= \frac{4\tilde{y}^{2a-1}b_{(1)}^{2a_{(1)}}}{\Gamma(a)\Gamma(a_{(1)})}\int_\theta \theta^{2(a+a_{(1)})-1}e^{-(\tilde{y^2}-b_{(1)}^2)\theta^2}d\theta \tag{82}$$

$$= \tilde{y}^{2a_{(1)}-1}\frac{2\Gamma(an+1)}{\Gamma(a_{(1)})\Gamma(a)}\frac{b_{(1)}^{2an}}{(b_{(1)}+\tilde{y}^2)^{2(an+1)}} \tag{83}$$

Problem 3.14

Unit information prior: Let $Y_1, \dots, Y_n \sim i.i.d. p(y|\theta)$. Having observed the values $Y_1 = y_1, \dots, Y_n = y_n$, the *log likelihood* is given by $l(\theta|y) = \sum \log p(y_i|\theta)$, and the value $\hat{\theta}$ of θ that maximize $l(\theta|y)$ is called the *maximum likelihood estimator*. The negative of the curvature of the log-likelihood, $J(\theta) = -\frac{\partial^2 l}{\partial\theta^2}$, describes the precision of the MLE $\hat{\theta}$ and is called the *observed Fisher information*. For situations in which it is difficult to quantify prior information in terms of a probability distribution, some have suggested that the "prior" distribution be based on the likelihood, for example, by centering the prior distribution around the MLE $\hat{\theta}$. To deal with the fact that the MLE is not really prior information, the curvature of the prior is chosen so that it has only "one nth" as much information as the likelihood, so that $-\frac{\partial^2 \log p(\theta)}{\partial\theta^2} = \frac{J(\theta)}{n}$. Such a prior is called a *unit information prior* (Kass and Wasserman, 1995; Kass and Raftery, 1995), as it has as much information as the average amount of information from a single observation. The unit information prior is not really a prior distribution, as it is computed from the observed data. However, it can be roughly viewed as the prior information of someone with weak but accurate prior information.

(a)

Let $Y_1, \ldots, Y_n \sim i.i.d.$ binary (θ). Obtain the MLE $\hat{\theta}$ and $\frac{J(\hat{\theta})}{n}$.

The Bernoullis distribution can be expressed as[2]

$$p(y_i|\theta) = \theta^{y_i}(1-\theta)^{1-y_i} \quad \text{for } y_i \in \{0, 1\}$$

Because $Y_i, \ldots, Y_n \sim i.i.d.$, $k_1 = \cdots = k_n = k$.

$$\begin{aligned}
l(\theta|y) &= \sum_{i=1}^{n} \log p(y_i|\theta) && (84)\\
&= \sum_{i=1}^{n} \log(\theta^{y_i}(1-\theta)^{1-y_i}) && (85)\\
&= \sum_{i=1}^{n} (y_i \log\theta + (1-y_i)\log(1-\theta)) && (86)\\
&= \log\theta \sum_{i=1}^{n} y_i + \log(1-\theta)(n - \sum_{i=1}^{n} y_i) && (87)
\end{aligned}$$

Thus, $\bar{y} = \sum_{i=1}^{n} y_i$ is the sufficient statistics.

$$l(\theta|y) = \bar{y}\log\theta + (n-\bar{y})\log(1-\theta)$$

$$\begin{aligned}
\frac{\partial l(\theta|y)}{\partial\theta} &= \frac{\bar{y}}{\theta} - \frac{n-\bar{y}}{1-\theta} && (88)\\
\frac{\partial^2 l(\theta|y)}{\partial^2\theta} &= \frac{-\bar{y}}{\theta^2} - \frac{\overbrace{n-\bar{y}}^{\geq 0}}{(1-\theta)^2} \leq 0 && (89)
\end{aligned}$$

[2]General expression of Bernoullis Distribution. Wiki. URL: https://en.wikipedia.org/wiki/Bernoulli_distribution

Thus, the curvature is concave because the second partial derivative is negative. Next, find the maximum $\hat{\theta}$.

$$0 = \frac{\partial l(\hat{\theta}|y)}{\partial \theta} = \frac{\bar{y}}{\hat{\theta}} - \frac{n - \bar{y}}{1 - \hat{\theta}} \tag{90}$$

$$\hat{\theta} = \frac{\bar{y}}{n}_{\#} \tag{91}$$

$$\begin{aligned}
\frac{J(\hat{\theta})}{n} &= -\frac{\partial^2 l}{\partial \theta^2}\frac{1}{n} && (92)\\
&= \left(\frac{\bar{y}}{\hat{\theta}^2} + \frac{n - \bar{y}}{(1 - \hat{\theta})^2}\right)\frac{1}{n} && (93)\\
&= \frac{\hat{\theta}}{\hat{\theta}^2} + \frac{1 - \hat{\theta}}{(1 - \hat{\theta}^2)} && (94)\\
&= \frac{1}{\hat{\theta}} + \frac{1 - \hat{\theta}}{(1 - \hat{\theta}^2)} && (95)\\
& && (96)
\end{aligned}$$

(b)

Find a probability density $p_U(\theta)$ such that $\log p_U(\theta) = \frac{l(\theta|y)}{n} + c$, where c is a constant that does not depend on θ. Compute the information $-\frac{\partial^2 \log p_U(\theta)}{\partial \theta^2}$ of this density.

Part I: Derive $p_U(\theta)$

$$p_U(\theta) = e^{\frac{l(\theta|y)}{n}} e^c \tag{97}$$

$$\int_0^1 p_U(\theta) d\theta = 1 = e^c \int_0^1 e^{\frac{l(\theta|y)}{n}} d\theta \tag{98}$$

$$1 = e^c \int_0^1 \exp(\hat{\theta} \log \theta + (1-\hat{\theta}) \log(1-\theta))) d\theta \tag{99}$$

$$1 = e^c \int_0^1 \theta^{\hat{\theta}} (1-\theta)^{1-\hat{\theta}} d\theta \tag{100}$$

$$1 = e^c \frac{-\pi}{2} (\hat{\theta}-1)\hat{\theta} \csc(\pi\hat{\theta}) \tag{101}$$

$$e^c = \frac{-2}{\pi(\hat{\theta}-1)\hat{\theta}\csc(\pi\hat{\theta})} \tag{102}$$

$$c = \log\left(\frac{-2}{\pi(\hat{\theta}-1)\hat{\theta}\csc(\pi\hat{\theta})}\right) \tag{103}$$

Therefore, we get

$$\log p_U(\theta) = \frac{l(\theta|y)}{n} + \log\left(\frac{-2}{\pi(\hat{\theta}-1)\hat{\theta}\csc(\pi\hat{\theta})}\right)$$

Part II: Fisher inforamtion

$$-\frac{\partial^2 \log p_U(\theta)}{\partial \theta^2} = \frac{-1}{n} \frac{\partial^2 l(\theta|y)}{\partial \theta^2} \tag{104}$$

$$= \frac{\hat{\theta}}{\theta} + \frac{1-\hat{\theta}}{1-\theta^2}{}_{\#} \tag{105}$$

(c)

Obtain a probability density for θ that is proportional to $p_U(\theta) \times p(y_1, \ldots, y_n|\theta)$. Can this be considered a posterior distribution for θ?

$$p_U(\theta) \times p(y_1, \dots, y_n|\theta) \tag{106}$$

$$= \frac{-2}{\pi(\hat{\theta}-1)\hat{\theta}\csc(\pi\hat{\theta})} e^{\frac{l(\theta|y)}{n}} \times \prod_{i=1}^{n} p(y_i|\theta) \tag{107}$$

$$= \frac{-2}{\pi(\hat{\theta}-1)\hat{\theta}\csc(\pi\hat{\theta})} e^{\frac{l(\theta|y)}{n}} \times \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} \tag{108}$$

$$= \frac{-2}{\pi(\hat{\theta}-1)\hat{\theta}\csc(\pi\hat{\theta})} \frac{\theta^{\hat{\theta}}}{(1-\theta)^{\hat{\theta}}} \times \theta^{n\hat{\theta}}(1-\theta)^{n(1-\hat{\theta})} \tag{109}$$

$$= \frac{-2}{\pi(\hat{\theta}-1)\hat{\theta}\csc(\pi\hat{\theta})} \frac{\theta^{\hat{\theta}(n+1)}}{(1-\theta)^{\hat{\theta}-n(1-\hat{\theta})}} \tag{110}$$

- Yes, the resulting is the unit information prior.

(d)

Repeat (a), (b) and (c) but with $p(y|\theta)$ being the Poisson distribution.

From Hoff (2009, vol. 580, sec. 3.2), The PDF of poisson distribution is

$$p(Y=y|\theta) = dpois(y,\theta) = \theta^y \frac{e^{-\theta}}{\Gamma(y)} \quad \text{for } y \in \{0, 1, 2, \dots\}$$

Part I: MLE $\hat{\theta}$

$$l(\theta|y) = \sum_{i=1}^{n} \log p(y_i|\theta) \tag{111}$$

$$= \sum_{i=1}^{n} \log\left(\theta^y \frac{e^{-\theta}}{\Gamma(y)}\right) \tag{112}$$

$$= \log \frac{\theta^{\sum y} e^{-n\theta}}{\sum \Gamma(y)} \tag{113}$$

$$= \log\left(\theta^{\sum y} e^{-n\theta}\right) - \log(\sum \Gamma(y)) \tag{114}$$

$$= \log\left(\frac{\theta^{\sum y} e^{-n\theta}}{\sum \Gamma(y)}\right) \tag{115}$$

Get the MLE $\hat{\theta}$,

$$\frac{\partial l}{\partial \theta} = \frac{\sum y}{\theta} - n \tag{116}$$

$$\hat{\theta} = \frac{\sum_{i=1}^{n} y}{n} \tag{117}$$

Part II: Find Unit information prior

$$\frac{J(\hat{\theta})}{n} = -\frac{\partial^2 l}{\partial \theta^2}\frac{1}{n} \tag{118}$$

$$= \frac{\sum_{i=1}^{n} y_i}{\theta^2}\frac{1}{n} = \frac{1}{\hat{\theta}} \tag{119}$$

Part III: Derive P_U

$$P_U(\theta) = e^c e^{\frac{l(\theta|y)}{n}} \tag{120}$$

$$= e^c \left(\frac{\theta^{\sum y} e^{-n\theta}}{\sum \Gamma(y)}\right)^{\frac{1}{n}} \tag{121}$$

$$\int_0^\infty P_U(\theta) d\theta = 1 = e^c \int_0^\infty \left(\frac{\theta^{\sum y} e^{-n\theta}}{\sum \Gamma(y)}\right)^{\frac{1}{n}} d\theta \tag{122}$$

$$= \frac{e^c}{(\sum \Gamma(y))^{\frac{1}{n}}} \int_0^\infty \theta^{\hat{\theta}} e^{-\theta} d\theta \tag{123}$$

Use the fact that $\int_0^\infty x^a e^{-x} dx = -\Gamma(a+1)$.

$$1 = \frac{e^c}{(\sum \Gamma(y))^{\frac{1}{n}}} \Gamma(\hat{\theta}+1) \tag{124}$$

$$c = \log\left(\frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta}+1)}\right) \tag{125}$$

Therefore,

$$
\begin{aligned}
P_U(\theta) &= e^c e^{\frac{l(\theta|y)}{n}} &(126)\\
&= e^c \left[\frac{\theta^{\sum y} e^{-n\theta}}{\sum \Gamma(y)}\right]^{\frac{1}{n}} &(127)\\
&= \frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta}+1)} \left[\frac{\theta^{\sum y} e^{-n\theta}}{\sum \Gamma(y)}\right]^{\frac{1}{n}} &(128)
\end{aligned}
$$

Part IV: Fisher information of P_U

$$
\begin{aligned}
\log p_U(\theta) &= \log\left(\frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta}+1)}\right) + \frac{1}{n}\log\left[\frac{\theta^{\sum y} e^{-n\theta}}{\sum \Gamma(y)}\right] &(129)\\
&= \log\left(\frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta}+1)}\right) + \frac{1}{n}\log\left[\theta^{\sum y} e^{-n\theta} - \frac{1}{n}\log(\sum \Gamma(y))\right] &(130)
\end{aligned}
$$

Use the fact that $\frac{\partial^2}{\partial x^2}\left[\frac{\log(x^a e^{-nx})}{n}\right] = \frac{-a}{nx^2}$.[3]

$$
-\frac{\partial^2 \log P_U(\theta)}{\partial \theta^2} = \frac{\sum y}{n\theta^2}_{\#} \tag{131}
$$

Part V: Obtain the posterior distribution

$$
\begin{aligned}
p_U(\theta) \times p(y_1, \ldots, y_n|\theta) &= \frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta}+1)} e^{\frac{l(\theta|y)}{n}} \times \prod_{i=1}^{n} p(y_i|\theta) &(132)\\
&= \frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta}+1)} e^{\frac{l(\theta|y)}{n}} \times \frac{\theta^{\sum y} e^{-n\theta}}{\sum \Gamma(y)} &(133)\\
&= \frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta}+1)} \left(\frac{\theta^{n\hat{\theta}} e^{-n\theta}}{\sum \Gamma(y)}\right)^{\frac{1}{n}+1} &(134)
\end{aligned}
$$

References

Diaconis, Persi, and Donald Ylvisaker. 1985. "Quantifying Prior Opinion, Bayesian Statistics. Vol. 2." North Holland Amsterdam:

Hoff, Peter D. 2009. *A First Course in Bayesian Statistical Methods*. Vol. 580. Springer.

[3]https://www.wolframalpha.com/input?i=d%5E2+1%2Fn*log%28x%5Ea*e%5E%28-nx%29%29%2Fdx%5E2