

Homework 3

Shao-Ting Chiu (UIN:433002162)

9/30/22

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 Environment 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
```

```

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

```

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 probability of tossing head.¹

¹This solution is referred to the lecture note about mixture priors.
<http://www.mas.ncl.ac.uk/~nmf16/teaching/mas3301/week11.pdf>

URL:

(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 coefficient $k = [k_1, k_2, k_3]$ with $\sum_{i=1}^3 k_j = 1$.

Let the prior probability be

$$p(\theta) = \sum_{i=1}^3 k_i p_i(\theta) \tag{1}$$

$$= k_1 p_1(\theta) + k_2 p_2(\theta) + k_3 p_3(\theta) \tag{2}$$

$$= 0.2 \times Beta(\theta, a_1, b_1) + 0.4 \times Beta(\theta, a_2, b_2) + 0.4 \times Beta(\theta, a_3, b_3) \tag{3}$$

$$= 0.2 \times Beta(\theta, 3, 3) + 0.4 \times Beta(\theta, 2, 4) + 0.4 \times Beta(\theta, 4, 2) \tag{4}$$

The distribution is shown in Figure 1.

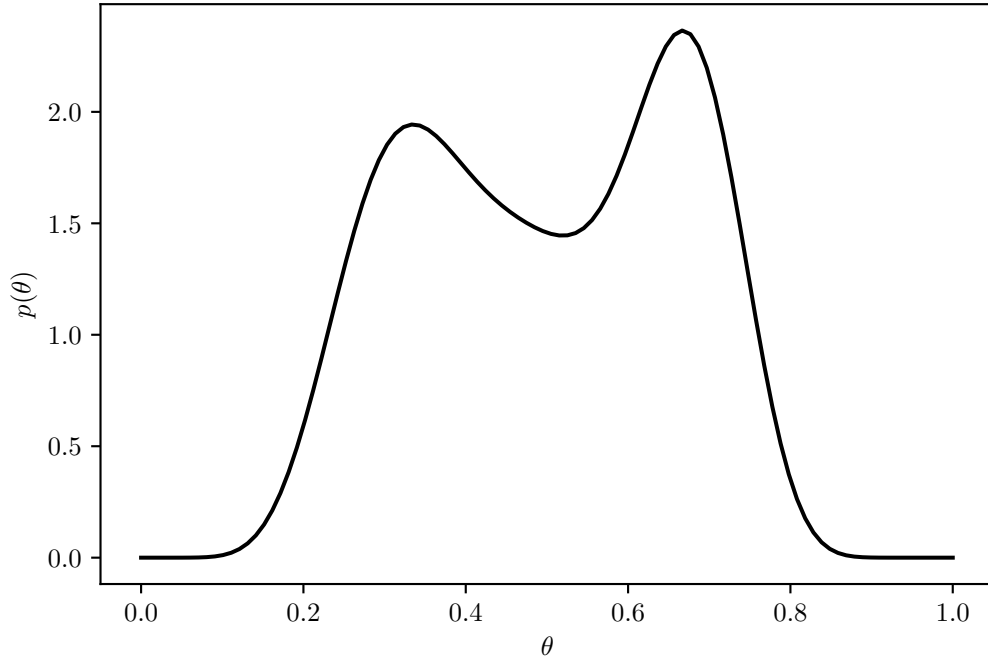


Figure 1: Designed mixture prior.

Table 1: Statistics 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 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 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

$$p_i(\theta|y) = \frac{p_i(\theta)p(y|\theta)}{\underbrace{\int_{\theta \in [0,1]} p_i(\theta)p(y|\theta)d\theta}_{=C_j}} \quad (5)$$

$$= \frac{Beta(\theta, a_i, b_i) \binom{n}{y} \theta^y (1-\theta)^{n-y}}{\underbrace{\int_0^1 Beta(\theta, a_i, b_i) \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta}_{=C_j}} \quad (6)$$

$$= \frac{\frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \binom{n}{y} \theta^y (1-\theta)^{n-y}}{\int_0^1 \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta} \quad (7)$$

$$= \frac{\frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \theta^{(a_i-1)} (1-\theta)^{b_i-1} \binom{n}{y} \theta^y (1-\theta)^{n-y}}{\int_0^1 \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)} \theta^{(a_i-1)} (1-\theta)^{b_i-1} \binom{n}{y} \theta^y (1-\theta)^{n-y} d\theta} \quad (8)$$

$$= Beta(\theta, a_i + y, b_i + n - y) \quad (9)$$

The posterior distribution of individual prior is

$$p_1(\theta|y) = Beta(\theta, a_1 + y, b_1 + n - y) \quad (10)$$

$$= Beta(\theta, 3 + 12, 3 + 100 - 12) = Beta(\theta, 15, 91) \quad (11)$$

$$p_2(\theta|y) = Beta(\theta, a_2 + y, b_2 + n - y) \quad (12)$$

$$= Beta(\theta, 2 + 12, 4 + 100 - 12) = Beta(\theta, 14, 92) \quad (13)$$

$$p_3(\theta|y) = Beta(\theta, a_3 + y, b_3 + n - y) \quad (14)$$

$$= Beta(\theta, 4 + 12, 2 + 100 - 12) = Beta(\theta, 16, 90) \quad (15)$$

$$(16)$$

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

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

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

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

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

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

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

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

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

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

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

Let

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

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

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

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

$$p(\theta|y) = \sum_{i=1}^3 k_i^{(1)} p_i(\theta|y) \quad (31)$$

$$= k_1^{(1)} p_1(\theta|y) + k_2^{(1)} p_2(\theta|y) + k_3^{(1)} p_3(\theta|y) \quad (32)$$

$$= 1.260148 \times 10^{-57} \times Beta(\theta, 15, 91) \quad (33)$$

$$+ 9.750000 \times 10^{-01} \times Beta(\theta, 14, 92) \quad (34)$$

$$+ 2.500000e \times 10^{-02} \times Beta(\theta, 16, 90) \quad (35)$$

(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: Statistics 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 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]
```

$$p^1(\theta) = 1.260148 \times 10^{-57} \times \text{Beta}(\theta, 15, 91) \quad (36)$$

$$+ 9.750000 \times 10^{-01} \times \text{Beta}(\theta, 14, 92) \quad (37)$$

$$+ 2.500000e \times 10^{-02} \times \text{Beta}(\theta, 16, 90) \quad (38)$$

Apply Equation 28,

$$C_1^* = \frac{\Gamma(15+51)\Gamma(91+100-51)}{\Gamma(15)\Gamma(91)} \quad (39)$$

$$= \frac{\Gamma(66)\Gamma(140)}{\Gamma(15)\Gamma(91)} \quad (40)$$

$$C_2^* = \frac{\Gamma(14+51)\Gamma(92+100-51)}{\Gamma(14)\Gamma(92)} \quad (41)$$

$$= \frac{\Gamma(65)\Gamma(141)}{\Gamma(14)\Gamma(92)} \quad (42)$$

$$C_3^* = \frac{\Gamma(16+51)\Gamma(90+100-51)}{\Gamma(16)\Gamma(90)} \quad (43)$$

$$= \frac{\Gamma(67)\Gamma(139)}{\Gamma(16)\Gamma(90)} \quad (44)$$

$$(45)$$

	Variables	Values
0	C1*	6.123054e+180
1	C2*	2.028941e+180
2	C3*	1.744410e+181
3	\hat{k}_{1_1}	2.392183e-01
4	\hat{k}_{1_2}	7.926761e-02
5	\hat{k}_{1_3}	6.815141e-01

$$p^2(\theta|y_2) = \sum_{i=1}^3 k_i^{(2)} p_i^{(2)}(\theta|y_2) \quad (46)$$

$$= k_1^{(2)} p_1^{(2)}(\theta|y_2) + k_2^{(2)} p_2^{(2)}(\theta|y_2) + k_3^{(2)} p_3^{(2)}(\theta|y_2) \quad (47)$$

$$= k_1^{(2)} Beta(15+51, 91+49) \quad (48)$$

$$+ k_2^{(2)} Beta(14+51, 92+49) \quad (49)$$

$$+ k_3^{(2)} Beta(16+51, 90+49) \quad (50)$$

$$= 2.32 \times 10^{-1} \times Beta(66, 140) \quad (51)$$

$$+ 7.93 \times 10^{-2} \times Beta(65, 141) \quad (52)$$

$$+ 6.82 \times 10^{-1} \times Beta(67, 139) \quad (53)$$

$$(54)$$

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

$$p(y|\theta) = \frac{2}{\Gamma(a)} y^{2a-1} \theta^{2a} e^{-\theta^2 y^2} \quad (55)$$

$$= \left(\frac{2}{\Gamma(a)} y^{2a-1} \right) (\theta^2)^a (e^{-\theta^2 y^2}) \quad (56)$$

$$(57)$$

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

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

Derive the posterior distribution

$$p(\phi|n_0, t_0) = \kappa(n_0, t_0) c(\phi)^{n_0} \exp(n_0 t_0 \phi) \quad (58)$$

$$= \kappa(n_0, t_0) \phi^{a n_0} \exp(n_0 t_0 \phi) \quad (59)$$

$$p(\theta^2|n_0, t_0) = \kappa(n_0, t_0) \theta^{2a n_0} e^{(n_0 t_0 \theta^2)} \quad (60)$$

$$(61)$$

Apply change of variables

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

$$p_\phi(\phi) = p_\theta(f(\phi)) \times \left| \frac{df}{d\phi} \right| \quad (62)$$

$$\kappa(n_0, t_0) c(\phi)^{n_0} \exp(n_0 t_0 \phi) = p_\theta(\theta) \times \left(\frac{1}{2\sqrt{\theta}} \right) \quad (63)$$

$$\kappa(n_0, t_0) c(\theta^2)^{n_0} \exp(n_0 t_0 \theta^2) = p_\theta(\theta) \times \left(\frac{1}{2\sqrt{\theta}} \right) \quad (64)$$

$$p_\theta(\theta) = \kappa(n_0, t_0) \theta^{2a} e^{(n_0 t_0 \theta^2)} \quad (65)$$

$$p_\theta(\theta) = p_\phi(\phi(\theta)) \times \left| \frac{d\phi(\theta)}{d\theta} \right| \quad (66)$$

$$\propto \kappa(n_0, t_0) c(\theta^2)^{n_0} \exp(n_0 t_0 \theta^2) \times 2\theta \quad (67)$$

$$\propto \theta^{2an_0+1} e^{(n_0 t_0 \theta^2)} \quad (68)$$

$$(69)$$

To avoid alias, let Galenshore pdf function be

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

Combining Equation 70 together,

$$p_\theta(\theta|n_0, t_0) \propto dGalenshore(\theta; an_0 + 1, \sqrt{-n_0 t_0}) \quad (71)$$

$$(72)$$

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

(b)

Let $Y_1, \dots, Y_n \sim i.i.d. \text{ Galenshore}(a, \theta)$. Find the posterior distribution of θ given Y_1, \dots, Y_n , using a prior from your conjugate class.

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

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

$$p(\phi|Y) \propto \text{Galenshore}(an' + 1, \sqrt{-n't'}) \quad (73)$$

$$\propto p(\phi|n_0 + n, n_0 t_0 + n \bar{t}(y)) \quad (74)$$

$$p(\theta|Y) \propto d\text{Galenshore}(\theta; a(n_0 + n) + 1, \sqrt{(n_0 + n)(n_0 t_0 + n \bar{t}(y))}) \quad (75)$$

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

(c)

Write down $\frac{p(\theta_a|Y_1, \dots, Y_n)}{p(\theta_b|Y_1, \dots, Y_n)}$ 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, \dots, Y_n is

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

(d)

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

Use the posterior distribution derived in (b).

$$p(\theta|Y) \propto d\text{Galenshore}(\theta; \underbrace{a(n_0 + n) + 1}_{a^{(1)}}, \underbrace{\sqrt{(n_0 + n)(n_0 t_0 + n \bar{t}(y))}}_{\theta^{(1)}}) \quad (76)$$

$$E[\theta|Y] = \frac{\Gamma(a(n_0 + n) + \frac{3}{2})}{\sqrt{(n_0 + n)(n_0 t_0 + n \bar{t}(y))} \Gamma(a(n_0 + n) + 1)} \quad (77)$$

(e)

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

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

$$(79)$$

- $p(\tilde{y}|\theta) = \frac{2}{\Gamma(a)} \theta^{2a} \tilde{y}^{2a-1} e^{-\theta^2 \tilde{y}^2}$
- $p(\theta|Y) \propto d\text{Galenshore}(\theta; \underbrace{a(n_0 + n) + 1}_{a_{(1)}}, \underbrace{\sqrt{(n_0 + n)(n_0 t_0 + n \bar{t}(y))}}_{b_{(1)}})$

$$- p(\theta|Y) = \frac{2}{\Gamma(a_{(1)})} b_{(1)}^{2a_{(1)}} \theta^{2a_{(1)}-1} e^{-b_{(1)}^2 \theta^2}$$

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

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

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

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

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

The Bernoulli distribution can be expressed as²

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

Because $Y_1, \dots, Y_n \sim i.i.d.$, $k_1 = \dots = k_n = k$.

$$l(\theta|y) = \sum_{i=1}^n \log p(y_i|\theta) \tag{84}$$

$$= \sum_{i=1}^n \log(\theta^{y_i}(1-\theta)^{1-y_i}) \tag{85}$$

$$= \sum_{i=1}^n (y_i \log \theta + (1-y_i) \log(1-\theta)) \tag{86}$$

$$= \log \theta \sum_{i=1}^n y_i + \log(1-\theta)(n - \sum_{i=1}^n y_i) \tag{87}$$

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

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

$$\frac{\partial l(\theta|y)}{\partial \theta} = \frac{\bar{y}}{\theta} - \frac{n - \bar{y}}{1 - \theta} \tag{88}$$

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

²General expression of Bernoulli 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 $\hat{\theta}$.

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

$$\hat{\theta} = \frac{\bar{y}}{n_{\#}} \quad (91)$$

$$\frac{J(\hat{\theta})}{n} = -\frac{\partial^2 l}{\partial \theta^2} \frac{1}{n} \quad (92)$$

$$= \left(\frac{\bar{y}}{\hat{\theta}^2} + \frac{n - \bar{y}}{(1 - \hat{\theta})^2} \right) \frac{1}{n} \quad (93)$$

$$= \frac{\hat{\theta}}{\hat{\theta}^2} + \frac{1 - \hat{\theta}}{(1 - \hat{\theta})^2} \quad (94)$$

$$= \frac{1}{\hat{\theta}} + \frac{1 - \hat{\theta}}{(1 - \hat{\theta})^2} \quad (95)$$

$$(96)$$

(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 $p_U(\theta)$

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

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

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

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

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

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

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

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 information

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

$$= \frac{\hat{\theta}}{\theta} + \frac{1 - \hat{\theta}}{1 - \theta^2} \quad (105)$$

(c)

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

$$p_U(\theta) \times p(y_1, \dots, y_n | \theta) \quad (106)$$

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

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

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

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

- 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 $\hat{\theta}$

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

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

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

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

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

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

$$\frac{\partial l}{\partial \theta} = \frac{\sum y}{\theta} - n \quad (116)$$

$$\hat{\theta} = \frac{\sum_{i=1}^n y}{n} \quad (117)$$

Part II: Find Unit information prior

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

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

Part III: Derive P_U

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

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

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

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

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

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

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

Therefore,

$$P_U(\theta) = e^c e^{\frac{l(\theta|y)}{n}} \quad (126)$$

$$= e^c \left[\frac{\theta \sum y e^{-n\theta}}{\sum \Gamma(y)} \right]^{\frac{1}{n}} \quad (127)$$

$$= \frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta} + 1)} \left[\frac{\theta \sum y e^{-n\theta}}{\sum \Gamma(y)} \right]^{\frac{1}{n}} \quad (128)$$

Part IV: Fisher information of P_U

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

$$= \log \left(\frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta} + 1)} \right) + \frac{1}{n} \log \left[\theta \sum y e^{-n\theta} - \frac{1}{n} \log(\sum \Gamma(y)) \right] \quad (130)$$

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

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

Part V: Obtain the posterior distribution

$$p_U(\theta) \times p(y_1, \dots, y_n | \theta) = \frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta} + 1)} e^{\frac{l(\theta|y)}{n}} \times \prod_{i=1}^n p(y_i | \theta) \quad (132)$$

$$= \frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta} + 1)} e^{\frac{l(\theta|y)}{n}} \times \frac{\theta \sum y e^{-n\theta}}{\sum \Gamma(y)} \quad (133)$$

$$= \frac{(\sum \Gamma(y))^{\frac{1}{n}}}{\Gamma(\hat{\theta} + 1)} \left(\frac{\theta n \hat{\theta} e^{-n\theta}}{\sum \Gamma(y)} \right)^{\frac{1}{n} + 1} \quad (134)$$

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%2Fnlog%28x%5Eae%5E%28-nx%29%29%2Fdx%5E2>