COGS 118B - Homework1

by Gurpreet, PID: A14220554

1) Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities p(r) = 0.2, p(b) = 0.2, p(g) = 0.6, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then

- What is the probability of selecting an apple?

$$Law\ of\ Alternatives : p(apple) = p(apple|red) * p(red) + p(apple|blue) * p(blue) + p(apple|green) * p(green)$$

$$p(apple|red) = \frac{3}{10} = .30$$

$$p(apple|blue) = \frac{1}{2} = .50$$

$$p(apple|green) = \frac{3}{10} = .30$$

$$p(apple) = (.30)(.20) + (.50)(.20) + (.30)(.60)$$

$$=> p(apple) = 0.34$$

- If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

$$p(green|orange) = \frac{p(orange|green) * p(green)}{p(orange)}$$

$$Bayes\ Rule : p(green|orange)$$

$$= \frac{p(orange|green) * p(green)}{p(orange|green) * p(green) + p(orange|red) * p(red) + p(orange|blue) * p(blue)}$$

$$p(green|orange) = \frac{(.30)(.60)}{(.30)(.60) + (.40)(.20) + (.50)(.20)}$$

$$=> p(green|orange) = .50$$

2) Verify that the Bernoulli distribution satisfies the following properties

- $\sum_{x=0}^{1} p(x|\mu) = 1$

$$\sum_{x=0}^{1} p(x|\mu) = p(x=0|\mu) + p(x=1|\mu)$$
$$p(x=0|\mu) = 1-\mu$$
$$p(x=1|\mu) = \mu$$
$$\sum_{x=0}^{1} p(x|\mu) = 1-\mu+\mu$$
$$=> \sum_{x=0}^{1} p(x|\mu) = 1$$

- $E[x] = \mu$

$$E[x] = p(x=0|\mu)(x=0) + p(x=1|\mu)(x=1)$$
$$E[x] = p(x=1|\mu)(x=1)$$
$$=> E[x] = \mu$$

- $var[x] = \mu(1-\mu)$

$$var[x] = p(x_0)(x_0-\mu)^2 + p(x_1)(x_1-\mu)^2$$
$$var[x] = (1-\mu)(0-\mu)^2 + \mu(1-\mu)^2$$
$$var[x] = (1-\mu)(\mu)^2 + \mu(1-\mu)^2$$
$$var[x] = (\mu^2-\mu^3) + \mu(1-2\mu+\mu^2)$$
$$var[x] = (\mu^2-\mu^3) + \mu 1 - 2\mu^2 + \mu^3$$
$$var[x] = \mu - \mu^2$$
$$=> var[x] = \mu(1-\mu)$$

3) Make use of the result (2.265) to show that the mean, variance, and mode of the beta distribution (2.13) are given respectively by

- $E[\mu] = \frac{a}{a+b}$

$$E[\mu] = \int_0^1 \mu \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1} d\mu$$
$$E[\mu] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \mu^{(a+1)-1}(1-\mu)^{b-1} d\mu$$
$$E[\mu] = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]\left[\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)}\right]$$
$$E[\mu] = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]\left[\frac{a\Gamma(a)\Gamma(b)}{(a+b)\Gamma(a+b)}\right]$$
$$E[\mu] = \frac{a}{(a+b)}$$

- $var[\mu] = \frac{ab}{(a+b)^2(a+b+1)}$

$$var[\mu] = E[\mu^2] - (E[\mu])^2$$
$$(E[\mu])^2 = \left[\frac{a}{(a+b)}\right]^2$$
$$(E[\mu])^2 = \frac{a^2}{(a+b)^2}$$
$$E[\mu^2] = \int_0^1 \mu^2 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1} d\mu$$
$$E[\mu^2] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \mu^{a+2-1}(1-\mu)^{b-1} d\mu$$
$$E[\mu^2] = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]\left[\frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)}\right]$$
$$E[\mu^2] = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]\left[\frac{(a+1)\Gamma(a+1)\Gamma(b)}{(a+b+1)\Gamma(a+b+1)}\right]$$
$$E[\mu^2] = \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]\left[\frac{(a+1)(a)\Gamma(a)\Gamma(b)}{(a+b+1)(a+b)\Gamma(a+b)}\right]$$
$$E[\mu^2] = \left[\frac{(a+1)(a)}{(a+b+1)(a+b)}\right]$$
$$var[\mu] = \frac{(a+1)(a)}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2}$$
$$var[\mu] = \frac{(a+1)(a)(a+b)}{(a+b+1)(a+b)^2} - \frac{(a^2)(a+b+1)}{(a+b+1)(a+b)^2}$$
$$var[\mu] = \frac{(a+1)(a)(a+b)}{(a+b+1)(a+b)^2} - \frac{(a^2)(a+b+1)}{(a+b+1)(a+b)^2}$$
$$var[\mu] = \frac{(a^3 + a^2b + a^2 + ab) - (a^3 + a^2b + a^2)}{(a+b+1)(a+b)^2}$$
$$var[\mu] = \frac{ab}{(a+b+1)(a+b)^2}$$

- $mode[\mu] = \frac{a-1}{a+b-2}$

$$\frac{d}{d\mu}(uv) = u\frac{dv}{d\mu} + v\frac{du}{d\mu}$$
$$u = \mu^{(a-1)}$$
$$v = (1-\mu)^{(b-1)}$$
$$u\frac{dv}{d\mu} = -(b-1)\mu^{(a-1)}(1-\mu)^{(b-2)}$$
$$v\frac{du}{d\mu} = (a-1)\mu^{(a-2)}(1-\mu)^{(b-1)}$$
$$u\frac{dv}{d\mu} + v\frac{du}{d\mu} = (a-1)\mu^{(a-2)}(1-\mu)^{(b-1)} - (b-1)\mu^{(a-1)}(1-\mu)^{(b-2)}$$
$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[(a-1)\mu^{(a-2)}(1-\mu)^{(b-1)} - (b-1)\mu^{(a-1)}(1-\mu)^{(b-2)}\right] = 0$$
$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[(a-1)\mu^{(a-1-1)}(1-\mu)^{(b-1)} - (b-1)\mu^{(a-1)}(1-\mu)^{(b-1-1)}\right] = 0$$
$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[(a-1)\mu\mu^{(a-1)}(1-\mu)^{(b-1)} - (b-1)\mu^{(a-1)}(1-\mu)(1-\mu)^{(b-1)}\right] = 0$$
$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[(a-1)(1-\mu) - (b-1)\mu\right]\mu^{(a-1)}(1-\mu)^{(b-1)} = 0$$
$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[(a-1)(1-\mu) - (b-1)\mu\right]\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = 0$$
$$\left[(a-1)(1-\mu) - (b-1)\mu\right] = 0$$
$$\left[(a-1)(1-\mu) - (b-1)\mu\right] = 0$$
$$\left[(a - a\mu - 1 + \mu) - (b\mu - \mu)\right] = 0$$
$$\left[a - a\mu - 1 + \mu - b\mu + \mu\right] = 0$$
$$\mu = \frac{a-1}{a+b-2}$$

4) We will be following the use of Bayes, theorem as expressed in Bishop equation (LH):

$$posterior \ \propto \ likeihood \ \times prior$$

Assume that we observe a series of NIID coin flips $D = x_i, \ldots x_N$, where each coin flip is modelled as a Bernoulli random variable:

$$p(x_i, \ldots, x_n|\mu) = \mu^{(x_i)}(1 - \mu^{(x_i)}) \ for \ i = 1, \ldots, n$$

The above assumptions lead to the form of the likelihood function seen in Bishop equation (2.5). Make sure you understand this process For the prior on μ, we assume a Beta distribution with a= 1, b=1

$$p(\mu|a = 1, b = 1) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)}\mu^{(1-1)}(1 - \mu^{(1-1)}) = 1$$

a) By plugging the above forms for the likelihood function and the prior distribution over μ into Bishop equation (1,44), show that the posterior distribution has the form seen in Bishop (2.17):

$$p(\mu|D) = p(\mu; m, l, a, b) \propto \mu^{(m+a-1)}(1-\mu)^{(l+b-1)}$$

where $m = \sum_{\mathrm{i}=1}^{N} x_i$ (the number of heads) and $l = \sum_{\mathrm{i}=1}^{N}(1 - x_i)$ (the number of tails). Make particular note of the interpretation of the hyperparameters a and b as "effective observations" mentioned on Bishop page 72.

Suppose we now want to predict the outcome of the next trial (i.e. flip of the coin). The Bayesian way to do this is to evaluate the predictive distribution of the next coin flip x, given the observed data set D. The steps to do this are described on Bishop page 73, from equation (2.19) to (2.20)

$$posterior \ \propto \ likeihood \ \times prior$$

$$BayesTheorem : p(\mu|D) = \frac{p(D|\mu) \times p(\mu)}{p(D)}$$

$$posterior = p(\mu|D)$$

$$likeihood = p(D|\mu)$$

$$prior = p(\mu)$$

$$posterior \propto p(\mu|D) \propto p(D|\mu) \times p(\mu) \propto \prod_{\mathrm{i}=1}^{N} p(x_i|\mu) \times p(\mu|a = 1, b = 1)$$

$$=> (\mu^{(x_i+,\ldots,+x_N)}(1 - \mu^{((1-x_i)+,\ldots,+(1-x_N))})) \times (1) = (\mu^{(m)}(1 - \mu^{(l)}))$$

$$\therefore posterior \propto \mu^{(m+a-1)}(1-\mu)^{(l+b-1)}$$

b) Starting with $p(x = 1|D)$ from the start of Bishop equation (2.19), carefully write out and explain all the steps needed to yield Bishop equation (2.20)

$$p(x = 1|D) = \int_0^1 p(x = 1|\mu)p(\mu|D)du = \int_0^1 \mu p(\mu|D)du$$

$$p(x = 1|D) = \frac{\Gamma(m + a + l + b)}{\Gamma(m + a)\Gamma(l + b)} \int_0^1 \mu\mu^{(m+a-1)}(1-\mu)^{(l+b-1)}$$

$$p(x = 1|D) = \frac{\Gamma(m + a + l + b)}{\Gamma(m + a)\Gamma(l + b)} \int_0^1 \mu^{(m+a+1-1)}(1-\mu)^{(l+b-1)}$$

$$p(x = 1|D) = \frac{\Gamma(m + a + l + b)}{\Gamma(m + a)\Gamma(l + b)} \int_0^1 \mu^{(m+a+1-1)}(1-\mu)^{(l+b-1)}$$

$$p(x = 1|D) = \left[\frac{\Gamma(m + a + l + b)}{\Gamma(m + a)\Gamma(l + b)}\right]\left[\frac{\Gamma(m + a + 1)\Gamma(l + b))}{\Gamma(m + a + l + b + 1)}\right]$$

$$p(x = 1|D) = \left[\frac{\Gamma(m + a + l + b)}{\Gamma(m + a)\Gamma(l + b)}\right]\left[\frac{(m + a)\Gamma(m + a)\Gamma(l + b))}{(m + a + l + b)\Gamma(m + a + l + b)}\right]$$

$$=> p(x=1|D) = \left[\frac{(m+a)}{(m+a+l+b)}\right]$$

5) By setting the derivatives of the log likelihood function (1.54) with respect to μ equal to zero, verify the results (1.55)$Prove: \ \mu = \frac{1}{N}\sum_{n=1}^{N} x_n$

$$ln[\,p(x|\mu,\sigma^2)] = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}ln[\,\sigma^2] - \frac{N}{2}ln[2\pi]$$

$$\frac{\partial}{\partial\mu}\left[ln[\,p(x|\mu,\sigma^2)]\right] = 0$$

$$\frac{\partial}{\partial\mu}\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}ln[\,\sigma^2] - \frac{N}{2}ln[2\pi]\right] = 0$$

$$\frac{\partial}{\partial\mu}\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right] - \frac{\partial}{\partial\mu}\left[\frac{N}{2}ln[\,\sigma^2]\right] - \frac{\partial}{\partial\mu}\left[\frac{N}{2}ln[2\pi]\right] = 0$$

$$\frac{\partial}{\partial\mu}\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right] - [0] - [0] = 0$$

$$\frac{\partial}{\partial\mu}\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right] = 0$$

$$\frac{\partial}{\partial\mu}\left[\sum_{n=1}^{N}(x_n-\mu)^2\right] = 0$$

$$\frac{\partial}{\partial\mu}\left[\sum_{n=1}^{N}(x_n-\mu)^2\right] = 0$$

$$\sum_{n=1}^{N}\frac{\partial}{\partial\mu}(x_n-\mu)^2 = 0$$

$$-2\sum_{n=1}^{N}(x_n-\mu) = 0$$

$$\sum_{n=1}^{N}x_n - \sum_{n=1}^{N}\mu = 0$$

$$\sum_{n=1}^{N}\mu = \sum_{n=1}^{N}x_n$$

$$N\mu = \sum_{n=1}^{N}x_n$$

$$\mu = \frac{1}{N}\sum_{n=1}^{N}x_n$$

6) In the hw1 folder there is a file called hw1p1.mat. Use the Matlab command load to load this data in to Matlab. Once this data is loaded, use the plot command to plot the data. Plot the first column of the data along the x-axis and the second column of the data along the y-axis. Using the text command, write the complete plot command you used at the coordinates x = 5, y = 60. Use title to title the plot, and then label the axes using xlabel and ylabel. Print out this plot to hand in.

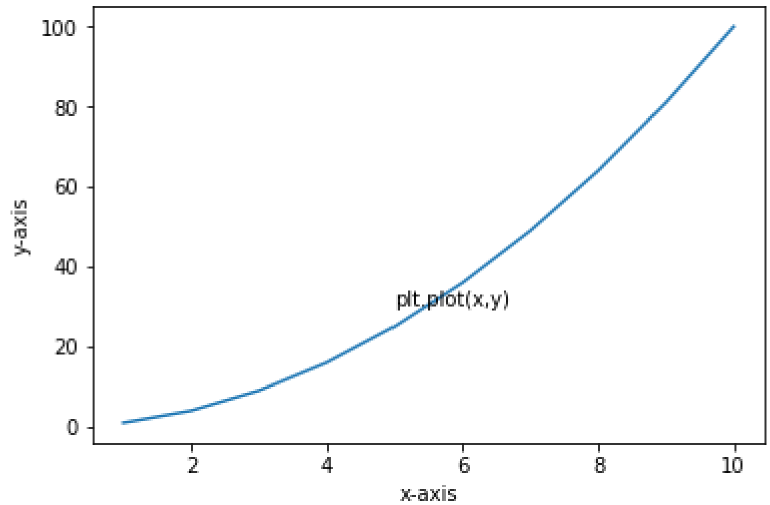
In [20]:
```
import scipy.io as sio
import numpy as np
import matplotlib.pyplot as plt
```

In [36]:
```
hw1p1 = sio.loadmat("hw1p1.mat")
type(hw1p1)
hw1p1
```

Out[36]:
```
{'__header__': b'MATLAB 5.0 MAT-file, Platform: GLNX86, Created on: Fri
Jan  2 18:03:04 2009',
 '__version__': '1.0',
 '__globals__': [],
 'hw1p1data': array([[  1,   1],
        [  2,   4],
        [  3,   9],
        [  4,  16],
        [  5,  25],
        [  6,  36],
        [  7,  49],
        [  8,  64],
        [  9,  81],
        [ 10, 100]], dtype=uint8)}
```

In [53]:
```
x = hw1p1["hw1p1data"][:,0]
y = hw1p1["hw1p1data"][:,1]
fig = plt.plot(x,y)
plt.xlabel('x-axis')
plt.ylabel('y-axis')
plt.annotate("plt.plot(x,y)", (5,30))
```

Out[53]:
```
Text(5,30,'plt.plot(x,y)')
```



7) Load hw1p2.mat into Matlab. Using surf, plot the data with the first column as the x coordinate, the second column as the y coordinate, and the third column as the z coordinate. To do this you will need to use the reshape command to reshape each column from a vector to a 101-by-101 matrix. Print out this plot to hand in.

In [73]:
```
import scipy.io as sio
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
import pandas as pd
```

In [75]:
```
hw1p2 = sio.loadmat("hw1p2.mat")
hw1p2
hw1p2["hw1p2data"]
df = pd.DataFrame(hw1p2["hw1p2data"])
df
```

Out[75]:

	0	1	2
0	-5.0	-5.0	2.210335e-12
1	-5.0	-4.9	3.626051e-12
2	-5.0	-4.8	5.889341e-12
3	-5.0	-4.7	9.470143e-12
4	-5.0	-4.6	1.507660e-11
5	-5.0	-4.5	2.376334e-11
6	-5.0	-4.4	3.708246e-11
7	-5.0	-4.3	5.729103e-11
8	-5.0	-4.2	8.763181e-11
9	-5.0	-4.1	1.327070e-10
10	-5.0	-4.0	1.989680e-10
11	-5.0	-3.9	2.953450e-10
12	-5.0	-3.8	4.340432e-10
13	-5.0	-3.7	6.315291e-10
14	-5.0	-3.6	9.097265e-10
15	-5.0	-3.5	1.297434e-09
16	-5.0	-3.4	1.831964e-09
17	-5.0	-3.3	2.560977e-09
18	-5.0	-3.2	3.544470e-09
19	-5.0	-3.1	4.856844e-09
20	-5.0	-3.0	6.588916e-09
21	-5.0	-2.9	8.849746e-09
22	-5.0	-2.8	1.176806e-08
23	-5.0	-2.7	1.549301e-08
24	-5.0	-2.6	2.019407e-08
25	-5.0	-2.5	2.605967e-08
26	-5.0	-2.4	3.329439e-08
27	-5.0	-2.3	4.211436e-08
28	-5.0	-2.2	5.274077e-08
29	-5.0	-2.1	6.539127e-08
...	...	...	...
10171	5.0	2.1	6.539127e-08
10172	5.0	2.2	5.274077e-08
10173	5.0	2.3	4.211436e-08

	0	1	2
10174	5.0	2.4	3.329439e-08
10175	5.0	2.5	2.605967e-08
10176	5.0	2.6	2.019407e-08
10177	5.0	2.7	1.549301e-08
10178	5.0	2.8	1.176806e-08
10179	5.0	2.9	8.849746e-09
10180	5.0	3.0	6.588916e-09
10181	5.0	3.1	4.856844e-09
10182	5.0	3.2	3.544470e-09
10183	5.0	3.3	2.560977e-09
10184	5.0	3.4	1.831964e-09
10185	5.0	3.5	1.297434e-09
10186	5.0	3.6	9.097265e-10
10187	5.0	3.7	6.315291e-10
10188	5.0	3.8	4.340432e-10
10189	5.0	3.9	2.953450e-10
10190	5.0	4.0	1.989680e-10
10191	5.0	4.1	1.327070e-10
10192	5.0	4.2	8.763181e-11
10193	5.0	4.3	5.729103e-11
10194	5.0	4.4	3.708246e-11
10195	5.0	4.5	2.376334e-11
10196	5.0	4.6	1.507660e-11
10197	5.0	4.7	9.470143e-12
10198	5.0	4.8	5.889341e-12
10199	5.0	4.9	3.626051e-12
10200	5.0	5.0	2.210335e-12

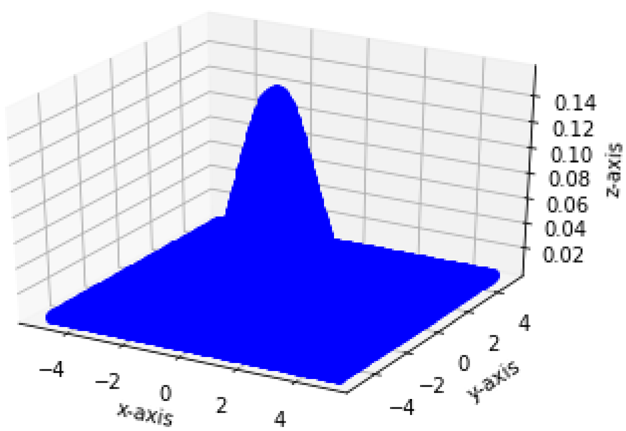
10201 rows × 3 columns

In [82]:
```
x = hw1p2["hw1p2data"][::,0]
y = hw1p2["hw1p2data"][::,1]
z = hw1p2["hw1p2data"][::,2]
print( x.shape )
print( y.shape )
print( z.shape )
```

```
(10201,)
(10201,)
(10201,)
```

In [86]:
```
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.plot(x, y, z, c='b', marker='o')
ax.set_xlabel('x-axis')
ax.set_ylabel('y-axis')
ax.set_zlabel('z-axis')
```

Out[86]:
```
Text(0.5,0,'z-axis')
```



In [ ]: