Bishop 1.3

$$p(\text{apple}) = p(r)\,p(\text{apple}|\text{red}) + p(b)\,p(\text{apple}|b) + p(g)\,p(\text{apple}|g)$$

$$= .2 \times \frac{3}{10} + .2 \times \frac{1}{2} + .6 \times \frac{3}{10}$$

$$= .06 + .1 + .18$$

$$= \underline{\underline{.34}}$$

$$p(g|\text{orange}) = \frac{p(\text{orange}|g)\,p(g)}{p(\text{orange})}$$

$$= \frac{\frac{3}{10} \times .6}{p(r)\,p(\text{orange}|r) + p(b)\,p(\text{orange}|b) + p(g)\,p(\text{orange}|g)}$$

$$= \frac{3/10 \times .6}{.2 \times 4/10 + .2 \times \frac{1}{2} + .6 \times \frac{3}{10}}$$

$$= \frac{.18}{.08 + .1 + .18}$$

$$= 1/2$$

Bishop 2.1

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

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

verify $E[x] = \mu$

$$E[x] = \sum_{x=0}^{1} x\, p(x|\mu) = 1 \times \mu + 0 \times (1-\mu) = \mu \checkmark$$

verify

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

$$\begin{aligned}
\mathrm{Var}[x] &= E[x - E[x]]^2 \\
&= E[x-\mu]^2 \\
&= \sum_{x=0}^{1} (x-\mu)^2 p(x|\mu) \\
&= (1-\mu)^2 \mu + (0-\mu)^2 (1-\mu) \\
&= (1-2\mu+\mu^2)\mu + \mu^2(1-\mu) \\
&= \mu - 2\mu^2 + \cancel{\mu^3} + \mu^2 - \cancel{\mu^3} \\
&= \mu - \mu^2 \\
&= \mu(1-\mu) \checkmark
\end{aligned}$$

2.6 For Beta distribution $\quad \mathrm{Beta}(\mu|a,b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\,\mu^{a-1}(1-\mu)^{b-1}$

show

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

$$E[\mu) = \int_0^1 \mathrm{Beta}(\mu)\,\mu\,d\mu$$

$$= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\,\mu^{a-1}(1-\mu)^{b-1}\mu\,d\mu$$

$$= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\,\mu^{a}(1-\mu)^{b-1}\,d\mu$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\,\cancel{\Gamma(b)}}\,\frac{\Gamma(a+1)\,\cancel{\Gamma(b)}}{\Gamma(a+b+1)} \qquad \text{(using 2.265)}$$

$$= \frac{\cancel{\Gamma(a+b)}\;a\,\cancel{\Gamma(a)}}{\cancel{\Gamma(a)}\,(a+b)\,\cancel{\Gamma(a+b)}} \qquad \text{using } \Gamma(a+1) = \Gamma(a)\times a \text{ (see problem 1.17)}$$

$$= \frac{a}{a+b} \quad \checkmark$$

2.6 cont'd

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

mode $[\mu]$ = peak of pdf — to find peak compute derivative and set it equal to zero

but note this will only be a max as opposed to another extremum when a> 1 and b>1 (look at plots of betapdf to see the different shapes)

$$\mathrm{Beta}(\mu) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1}$$

$$\frac{d\,\mathrm{Beta}(\mu)}{d\mu} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[(a-1)\mu^{a-2}(1-\mu)^{b-1} - (b-1)(1-\mu)^{b-2}\mu^{a-1}\right]$$

setting $=0$ $\Rightarrow$ and restricting to a>1, b>1

$$(a-1)\mu_{mode}^{a-2}(1-\mu_{mode})^{b-1} = (b-1)(1-\mu_{mode})^{b-2}\mu_{mode}^{a-1}$$

$$(a-1)(1-\mu_{mode}) = \mu_{mode}(b-1)$$

$$a-1-a\mu_{mode}+\mu_{mode} = b\mu_{mode}-\mu_{mode}$$

$$\mu_{mode}(2-a-b) = -a+1$$

$$\mu_{mode} = \frac{a-1}{a+b-2}$$ for a>1, and b>1

2.6 cont'd

show $\mathrm{Var}[\mu] = \dfrac{ab}{(a+b)^2(a+b+1)}$

$$\mathrm{Var}[\mu] = \int_0^1 \mathrm{Beta}(\mu)\left(\mu - \frac{a}{a+b}\right)^2 d\mu$$

$$= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1}\left(\mu - \frac{a}{a+b}\right)^2 d\mu$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \mu^{a-1}(1-\mu)^{b-1}\left(\mu^2 - \frac{2\mu a}{a+b} + \frac{a^2}{(a+b)^2}\right) d\mu$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \mu^{a-1}(1-\mu)^{b-1}\left(\mu(\mu-1) + \mu\left(1 - \frac{2a}{a+b}\right) + \frac{a^2}{(a+b)^2}\right) d\mu$$

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

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left[-\frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} + \frac{(a+b-2a)}{(a+b)}\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} + \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\frac{a^2}{(a+b)^2}\right]$$

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

$$= \frac{(a+b)\times(-ab) + (a+b+1)a(b-a) + a^2(a+b+1)}{(a+b)^2(a+b+1)}$$

$$= \frac{\cancel{-a^2b} \cancel{-ab^2} + \cancel{a^2b} + \cancel{ab^2} + ab \cancel{-a^3} \cancel{-a^2b} \cancel{-a^2} + \cancel{a^3} + \cancel{a^2b} + \cancel{a^2}}{(a+b)^2(a+b+1)}$$

$$= \frac{ab}{\quad}$$ ✓

HW 1 Q4

a)

posterior $\propto$ likelihood $\times$ prior

$$\propto \prod_{i=1}^{N} \mu^{x^{(i)}} (1-\mu)^{(1-x^{(i)})} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1}(1-\mu)^{b-1}$$

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

~~for~~ m heads
~~l~~ tail

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

This posterior is related to the prior with m extra μ factors and l extra $(1-\mu)$ factors, so a and b are related to extra head and tail observations respectively

b) $p(x=1|D) = \int_0^1 p(x=1|\mu)\, p(\mu|D)\, d\mu$ ← way done in class using 2.265

$$= \int_0^1 \mu \frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\Gamma(l+b)} \mu^{m+a-1} (1-\mu)^{l+b-1} d\mu$$

$$= \frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\Gamma(l+b)} \int_0^1 \mu^{m+a}(1-\mu)^{l+b-1} d\mu$$

$$= \frac{\Gamma(m+a+l+b)}{\Gamma(m+a)\cancel{\Gamma(l+b)}} \frac{\Gamma(m+a+1)\cancel{\Gamma(l+b)}}{\Gamma(m+a+l+b+1)}$$

$$= \frac{m+a}{m+a+l+b}$$

HW 1 Q4

b) way done in book

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

$$= \int_0^1 \mu\, p(x=1|\mu)\, d\mu$$

$$= E[\mu|D]$$

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

$$= \frac{m+a}{m+a+l+b}$$ (using 2.15 with $a = m+a$, $b = l+b$)

Q5. $\ln P(x \mid \mu, \sigma^2) = \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$

differentiating with respect to μ

$$\frac{\partial \ln P(x \mid \mu, \sigma^2)}{\partial \mu} = \frac{1}{2\sigma^2}\sum_{n=1}^{N}\frac{\partial}{\partial \mu}(x_n-\mu)^2 - 0 - 0$$

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

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

setting the derivative to 0.

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

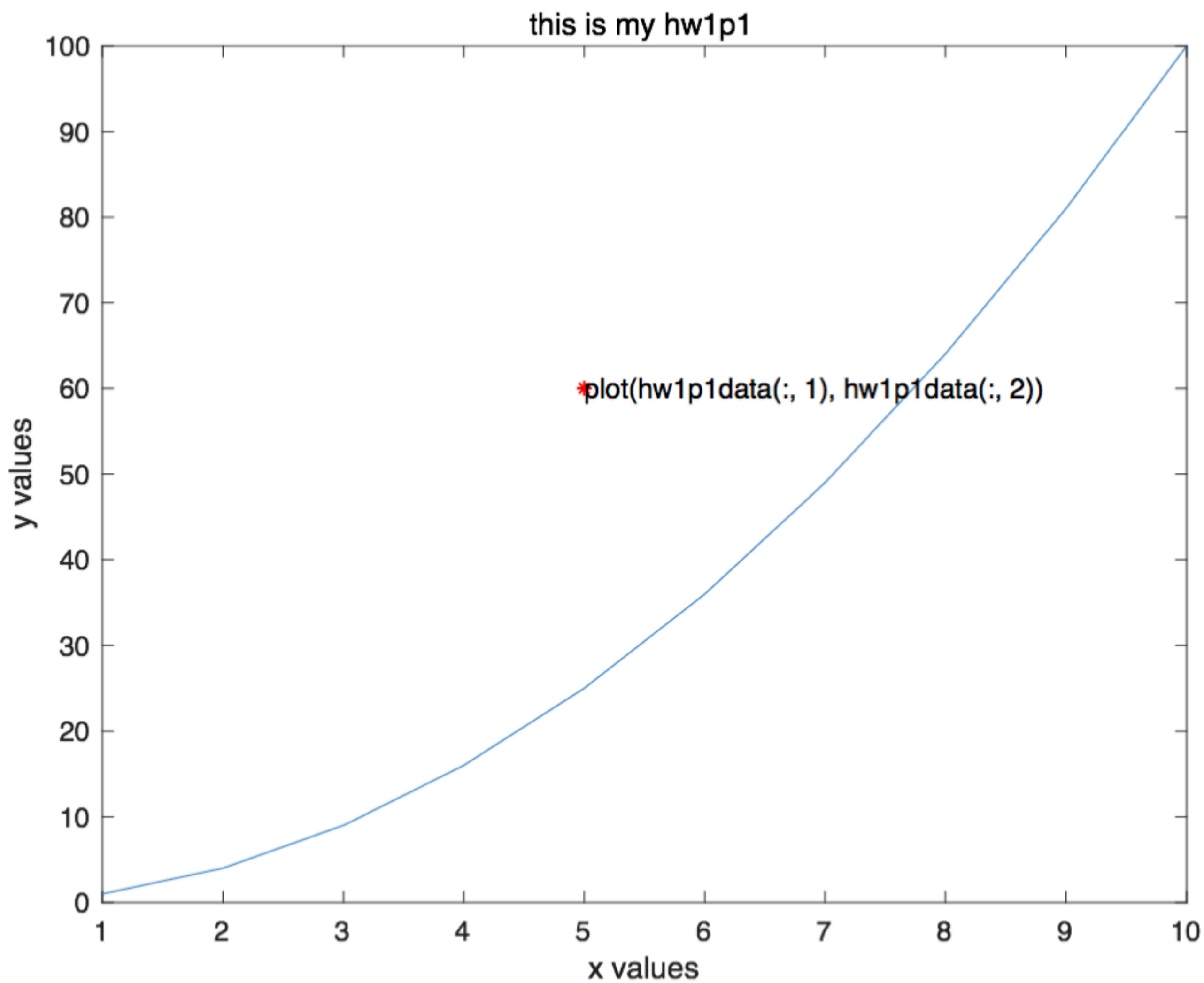
$$\Rightarrow \sum_{n=1}^{N} x_n - \sum_{n=1}^{N}\mu$$

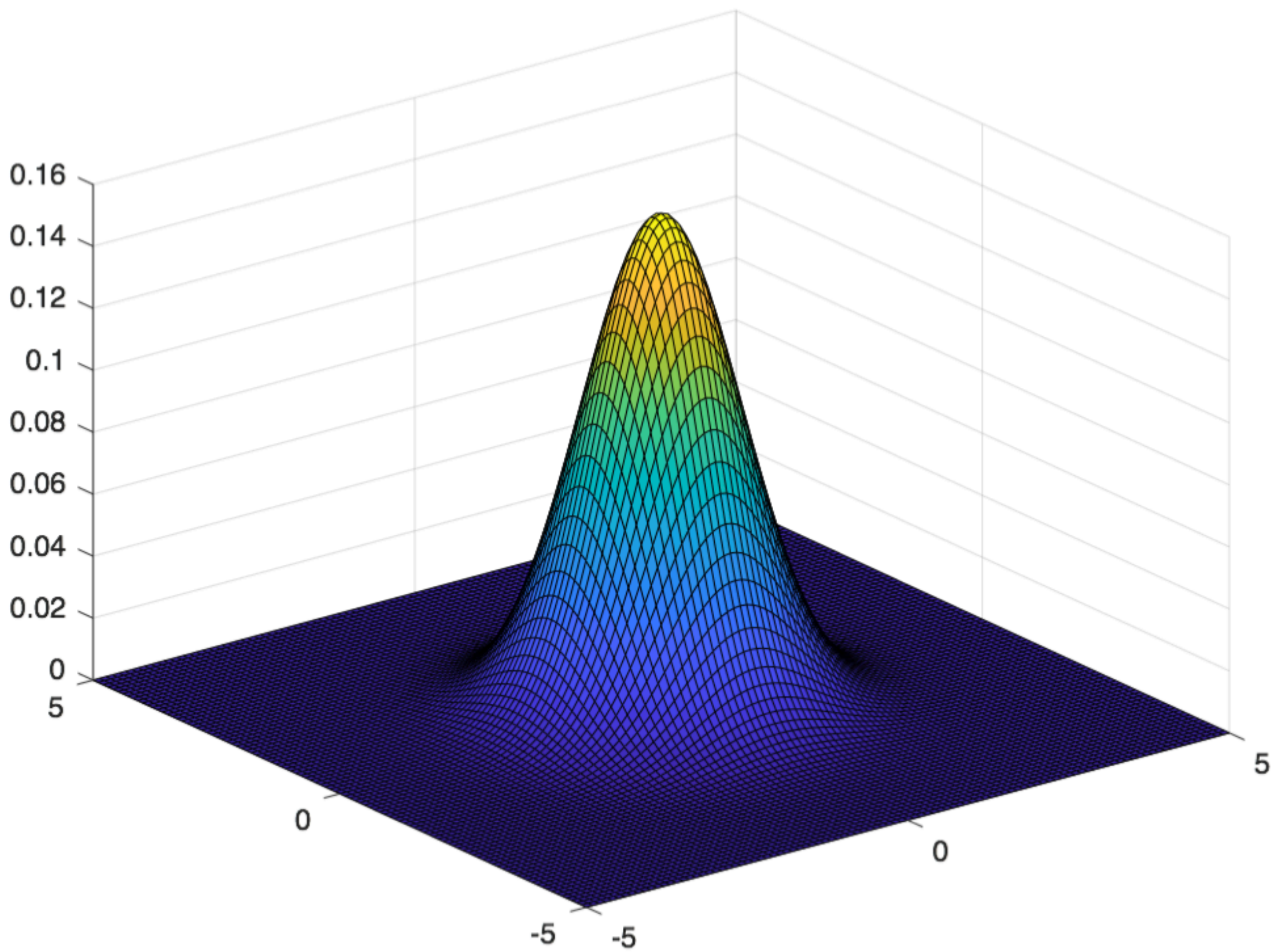
$$\Rightarrow \sum_{n=1}^{N} x_n - N\mu = 0$$

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

```
load hw1p1.mat
plot(hw1p1data(:,1),hw1p1data(:,2))
text(5,60,'plot(hw1p1data(:,1),hw1p1data(:,2))')
title('HW1 Q6')
xlabel('x value')
ylabel('y value')
```

```
load hw1p2
surf(reshape(hw1p2data(:,1),101,101),reshape(hw1p2data(:,2),101,101),reshape(hw1
p2data(:,3),101,101))
```


this is my hw1p1
plot(hw1p1data(:, 1), hw1p1data(:, 2))
y values
x values
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10


0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
5
0
-5
-5
0
5

Solution to HW1 bonus question

We want to find $P(1B, 2G|girlsroom, letter)$ and $P(1G, 2B|girlsroom, letter)$.

Note that all the possibilities for the children are BBB BBG BGG BGB GBB GBG GGB GGG. So $P(1B, 2G) = 3/8 = P(1G, 2B)$

$$\begin{aligned}
P(1B, 2G|girlsroom, letter) &= \frac{P(girlsroom, letter|1B, 2G) * P(1B, 2G)}{P(girlsroom, letter)} \\
&= \frac{P(girlsroom|1B, 2G) * P(letter|1B, 2G) * P(1B, 2G)}{P(girlsroom, letter)} \\
&= \frac{2/3 * a * 3/8}{P(girlsroom, letter)}
\end{aligned}$$

$$\begin{aligned}
P(2B, 1G|girlsroom, letter) &= \frac{P(girlsroom, letter|2B, 1G) * P(2B, 1G)}{P(girlsroom, letter)} \\
&= \frac{P(girlsroom|2B, 1G) * P(letter|2B, 1G) * P(2B, 1G)}{P(girlsroom, letter)} \\
&= \frac{1/3 * a * 3/8}{P(girlsroom, letter)} \\
&= 2 * P(1B, 2G|girlsroom, letter)
\end{aligned}$$

In going from the first lines above to the 2nd lines we have used the fact that the probability of seeing a girls room when you open the door and the probability of seeing the boys letter are conditionally independent given a stated number of boys and girls (they are not independent, but we just need them to be conditionally independent to let $P(girlsroom, letter|1B, 2G) = P(girlsroom|1B, 2B) * P(letter|1B, 2G)$.

(Note that $P(letter|2B, 1G) = P(letter|1B, 2G)$ (as long as there is a boy in the school, there is an equal likely chance of getting the letter and thus of the visitor seeing it whereas $P(girlsroom|1B, 2G)$ is 2/3 (the chance that the randomly opened child's room would be that of a girl)).

Since all other options are not possible we have $P(2B, 1G|girlsroom, letter) + P(1B, 2G|girlsroom, letter) = 1$ which means that

$P(1B, 2G) = 2/3$ and $P(2B, 1G) = 1/3$.