18.05 Exam 2 Solutions

Problem 1. (10 pts: 4,2,2,2) **Concept questions**

(a) Yes and yes. Frequentist statistics don't give the probability an hypothesis is true.

(b) True. Bayesian updating involves multiplying the likelihood and the prior. If the prior is 0 then this product will be 0.

(c) No. The actual experiment that was run would reject the null hypothesis if it were true, more than 5% of the time.

(d) Beta(8,13).

Problem 2. (10 pts)

likelihood $= f(\text{data} \mid \alpha) = \dfrac{\alpha}{5^\alpha} \cdot \dfrac{\alpha}{2^\alpha} \cdot \dfrac{\alpha}{3^\alpha} = \dfrac{\alpha^3}{20^\alpha}$.

Therefore, log likelihood $= \ln(f(\text{data} \mid \alpha) = \ln(\alpha) - \alpha \ln(30)$. We find the maximum likelihood by setting the derivative equal to 0:

$$\frac{d}{d\alpha} \ln(f(\text{data} \mid \alpha) = \frac{3}{\alpha} - \ln(30) = 0.$$

Solving we get $\boxed{\hat{\alpha} = \dfrac{3}{\ln(30)}}$.

Problem 3. (25: 10,5,5,5) **(a)**

hypoth. θ	prior	likelihood $P(x_1 = 5 \mid \theta)$	unormalized posterior	posterior	likelihood $P(x_2 = 7 \mid \theta)$
4-sided	1/2	0	0	0	0
8-sided	1/4	1/8	1/32	$\frac{1}{32T} = \frac{3}{5}$	1/8
12-sided	1/4	1/12	1/48	$\frac{1}{48T} = \frac{2}{5}$	1/22
			$T = \frac{1}{32} + \frac{1}{48} = \frac{5}{96}$		

(b) $P(x_1 = 5) = T = \frac{1}{32} + \frac{1}{48} = \frac{5}{96}$. (Either expression is okay.)

(c) $\text{Odds}(\theta = 12 \mid x_1 = 5) = \dfrac{P(\theta = 12 \mid x_1 = 5)}{P(\theta \neq 12 \mid x_1 = 5)} = \dfrac{2/5}{3/5} = 2/3.$

(d)

$$\begin{aligned} P(x_2 = 7 \mid x_1 = 5) &= \frac{3}{5} \cdot \frac{1}{8} + \frac{2}{5} \cdot \frac{1}{12} = \frac{13}{120} \\ &= \frac{1}{32T} \cdot \frac{1}{8} + \frac{1}{48T} \cdot \frac{1}{12} \\ &= \frac{1}{256T} + \frac{576}{T} \end{aligned}$$

(Any of these answers is okay.)

Problem 4. (15 pts)

This is a normal/normal conjugate prior pair, so we use the normal-normal update formulas.

$\mu =$ Beau's weight.
$n = 3, \quad \overline{x} = 1300$
Prior $\sim \text{N}(1200,\ 200^2)$, so $\quad \mu_{\text{prior}} = 1200, \quad \sigma^2_{\text{prior}} = 200^2$.
Likelihood $\sim \text{N}(\mu \mid 100^2)$, so $\quad \sigma^2 = 100^2$.

$a = \frac{1}{\sigma_{pr}^2} = \frac{1}{200^2}, \qquad b = \frac{n}{\sigma^2} = \frac{3}{100^2}.$

$$\mu_{\text{posterior}} = \frac{a \cdot \mu_{\text{prior}} + b \cdot \overline{x}}{a+b} = \frac{\frac{1}{200^2} \cdot 1200 + \frac{3}{100^2} \cdot 1300}{1/200^2 + 3/100^2}$$

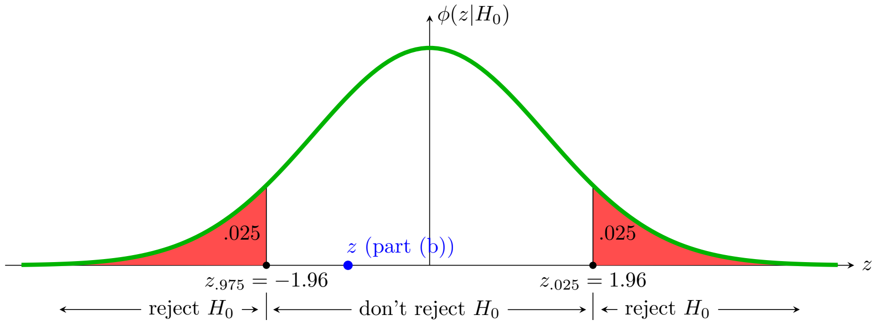
Problem 5. (10 pts: 4,3,3)
(a) Since the H_A is right-sided we use a right-sided rejection region: $\boxed{\text{rejection region } x = 5}$.

x	0	1	2	3	4	5
$\theta = .5$	0.031	0.156	0.313	0.313	0.156	0.031
$\theta = .6$	0.010	0.077	0.230	0.346	0.259	0.078
$\theta = .8$	0.000	0.006	0.051	0.205	0.410	0.328

(b) Power $= P(\text{reject} \mid \theta)$.
$\theta = .6$: power $= .078$
$\theta = .8$: power $= .328$

(c) $p = P(x \geq 4 \mid \theta = .5) = .156 + .031 = .187.$

Problem 6. (15 pts: 5,5,5)
(a) The test statistic is z, so we need a Z-graph



The rejection region is $z < -1.96$ or $z > 1.96$.

(b) We standardize $\overline{x}$ to get z: $z = \dfrac{\overline{x} - 2}{\sigma_{\overline{x}}} = \dfrac{1.5 - 2}{4/\sqrt{16}} = -.5$

(c) $p = 2P(Z \leq -.5) = 2 \cdot (.3085) = .6170$. Since $p > .05$ we do not reject H_0.

Problem 7. (15 pts)
The null hypothesis H_0: For the 4 words counted the long lost book has the same relative frequencies as *Sense and Sensibility*

Total word count of both books combined is 500, so the the maximum likelihood estimate of the relative frequencies assuming H_0 is simply the total count for each word divided by the total word count.

Word	a	an	this	that	Total count
Sense and Sensibility	150	30	30	90	300
Long lost work	90	20	10	80	200
totals	240	50	40	170	500
rel. frequencies under H_0	240/500	50/500	40/500	170/500	500/500

Now the expected counts for each book under H_0 are the total count for that book times the relative frequencies in the above table. The following table gives the counts: (observed, expected) for each book.

Word	a	an	this	that	Totals
Sense and Sensibility	(150, 144)	(30, 30)	(30, 24)	(90, 102)	(300, 300)
Long lost work	(90, 96)	(20, 20)	(10, 16)	(80, 68)	(200, 200)
Totals	(249, 240)	(50, 50)	(40, 40)	(170, 170)	(500, 500)

The chi-square statistic is

$$\begin{aligned} X^2 &= \sum \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{6^2}{144} + \frac{0^2}{30} + \frac{6^2}{24} + \frac{12^2}{102} + \frac{6^2}{96} + \frac{0^2}{20} + \frac{6^2}{16} + \frac{12^2}{68} \\ &\approx 7.9 \end{aligned}$$

There are 8 cells and all the marginal counts are fixed, so we can freely set the values in 3 cells in the table, e.g. the 3 blue cells, then the rest of the cells are determined in order to make the marginal totals correct. Thus $df = 3$.

Looking in the $df = 3$ row of the chi-square table we see that $X^2 = 7.9$ gives p between 0.025 and 0.05. Since this is less than our significance level of 0.1 we reject the null hypothesis that the relative frequencies of the words are the same in both books. Based on the assumption that all her books have similar word frequencies (which is something we could check) we conclude that the book is probably not by Jane Austen.

18.05 Introduction to Probability and Statistics
Spring 2014