STATISTICS 244 **Problem Set 7** **9:30-10:50AM TTh**

Problem Set 7 due Thursday March 7

1. In one study, sentences were classified according to the author's choice for the sentence's first word. Call "This, it, thus, and" Class I words; Class II is "everything else." For each of 215 groups of 5 of James Mill's sentences, the number of Class I words was counted.

# Class I words	0	1	2	3	4	5
# groups	87	11	51	42	20	4

Test whether a Binomial distribution ($n = 5, \theta$) fits these data.

2. The members of a community are classified by Blood type:

O	A	B	AB	Total
121	120	79	33	353

Theory has it that the probabilities of those types depends on gene frequency parameters r, p, q, where $r + p + q = 1$ and P("O") = r^2, P("A") = $p^2 + 2pr$, P("B") = $q^2 + 2qr$, and P("AB") = $2pq$. Using numerical methods (that is, a method such as that described in Chapter 5 of our notes) we can find the MLEs of r, p, q; they are .580, .246, and .173 [you may use these values as the MLEs without verifying that they are]. Test if the community fits the theory.

3. Are fingerprint patterns genetic, or are they developmental? In 1892 Francis Galton compiled the following table on the relationship between the patterns on the same finger of 105 sibling pairs. Test the hypothesis that the patterns are independent – for example, that knowing one sibling (A) has a Whorl on the finger does not help in predicting the pattern of the other (B).

	A children			
B children	Arches	Loops	Whorls	Totals
Arches	5	12	2	19
Loops	4	42	15	61
Whorls	1	14	10	25
Totals	10	68	27	105

4. For the Bortkiewicz Death by Horsekick Data, test the hypothesis that the data follow a Poisson distribution. [You should group the counts for "4 or more" as one category.]

Number of deaths	Frequency count
0	144
1	91
2	32
3	11
4	2
More	0
Total	280

5. In the top tennis matches, a player may challenge an official's line call (e.g. that the serve was "in" or "out"). If so, the official will review the output of an advanced technology camera called Hawk-Eye. Here are the results of 119 challenges made by some top players at the US Open in 2013. Test at the 10% level whether the players all have the same chance of winning a challenge.

Player	Won	Lost	Total
Murray	6	19	25
Ferrer	4	16	20
Azarenka	9	9	18
Nadal	5	12	17
Djokivic	4	12	16
Isner	4	9	13
Federer	1	9	10
Total	33	86	119

6. An American roulette wheel is spun n = 3880 times in order to test if it is fair (i.e. to test if each slot has probability 1/38). Suppose that each of the 36 numbered slots (1, 2, …, 36) comes up exactly 100 times and each of "0" and "00" comes up 140 times.
(a) Test at the 5% level using the Chi-squared test if the wheel is fair.
(b) Now suppose that before you had looked at the data you had suspected that the numbered slots were less likely than the "0" and "00", and you had decided to test the binomial hypothesis H_0: P("0" or "00") = 2/38 vs. H_1: P("0" or "00") > 2/38. We know that the UMP test of these hypotheses rejects H_0 if Z (= total number of "0" and "00"s) is greater than C, where C is chosen for a level 0.05 test. Use the fact that under H_0, Z has approximately a Normal N(n*(2/38), n*(2/38)*(36/38)) distribution [this follows from the Central Limit Theorem, for example] to find C and perform the test.
(c) Compare the result in (b) with that in (a). [This is intended to illustrate that with exactly the same data different conclusions may be reached depending upon how focused the test is upon a narrow hypothesis. The test in (a) tests against <u>all</u> alternatives to a fair wheel and the test of (b) is focused narrowly upon deviations of the probabilities of "0" and "00" slots from the others. As this example should suggest, it would be statistically impermissible to focus the test after seeing the data without compensating for this choice of a narrow hypothesis in some way.]

7. Can under-powered tests be trusted? Consider the following simple situation: A person claims to be able to predict the outcome of a coin toss. To test this claim, a coin will be tossed n=5 times and the hypothesis "the person is just guessing randomly" is tested; if all five tosses are correct, the hypothesis is rejected and the person is said to have extra-sensory powers (ESP). The person takes the test and gets all 5 correct. Since the chance of getting 5 correct if p=1/2 is $(1/2)^5 = 0.03125$, the person then brags about passing a stringent scientific test at the 3% level, an impressive sounding achievement! But is it a reasonable?
(a) Complete the following table (row 2 gives the power of the test for the alternatives in row 1):

Prob correct on one trial p:	.50000	.60000	.70000	.80000	.90000
Prob correct on five trials = p^5	.03125				

(b) The calculations for (a) show the power is low; with only 5 trials it is hard to detect all but extremely large effects. But still, the initial test rejected the "just guessing" hypothesis at about the 3% level, so maybe the bragging is valid?
To examine that bragging claim, let us suppose the *a priori* probability the person has ESP is θ. Now, separately for each of the alternatives p= .6, .7. .8, .9, calculate the posterior probabilities (given the data of "5 correct") that the person has ESP at those p's as a function of θ. Then give these posterior probabilities for θ = .001, .01, .1, .2, .4, .5.

[The intention of this problem is to show that tests with low power may reject, but even then, the credibility of the conclusion is low. In fact, when a low power test rejects, it is indeed true that something unusual has happened, but it is almost as surprising that you reject under an alternative, and unless your *a priori* probability of ESP is high, a rejection will not convince you. Many studies that were later attacked as having given false results used low power tests.]