MATH/STAT 3460, Intermediate Statistical Theory
Winter 2014
Toby Kenney
Sample Final Examination

This Sample Final has more questions than the actual final, in order to cover a wider range of questions.

Critical values for chi-squared distribution:

Degrees of Freedom	10%	5%	1%
1	2.70554345	3.84145882	6.63489660
2	4.60517019	5.99146455	9.21034037
3	6.25138863	7.81472790	11.34486673
4	7.77944034	9.48772904	13.27670414

Some helpful derivatives:

$f(x)$	$f'(x)$	$f''(x)$
$\log(1-e^{-\lambda}(1+\lambda))$	$\frac{\lambda e^{-\lambda}}{1-e^{-\lambda}(1+\lambda)}$	$-\frac{2\lambda e^{-\lambda}+e^{-2\lambda}}{(1-e^{-\lambda}(1+\lambda))^2}$

1. Under a certain model, the number of is assumed to follow a censored Poisson distribution with parameter λ. That is the probabilites of 0, 1 and 2 are $e^{-\lambda}$, $\lambda e^{-\lambda}$ and $1-e^{-\lambda}(1+\lambda)$. The observed frequencies are :

Number	Frequency
0	102
1	131
2	84

   Use Newton's method to find the Maximum Likelihood estimate for λ. [Start with an initial estimate of 1, and perform 1 iteration.]

2. A scientist wants to determine the frequency of a particular version of a gene. The gene has two versions A (probability p) and B (probability $1-p$), and individuals have two copies of the gene, and so can be classified as AA (probability p^2), AB (probability $2p(1-p)$) or BB (probability $(1-p)^2$). The scientist has two tests — one that tests for the state AA and one that tests for the state BB. The scientist plans to test 100 patients with one of the two tests. If the true value of p is 0.3:

   (a) What is the expected information about p for each of the tests?

   (b) Which test should the scientist use?

3. Every year a certain city is flooded with probability p. Based on data from a record of whether or not the city flooded every year for the past 600 years. The maximum likelihood estimate for p is therefore $\hat{p}=\frac{F}{600}$, where F is the number of years that the city is flooded.

An insurance company is interested in the probability that the city floods at any time within the next 2 years. The probability of this is $1-(1-p)^2 = 2p-p^2$. The maximum likelihood estimate of this is therefore $\frac{F}{300} - \frac{F^2}{360000}$. What is the bias of this estimate?

4. Let X_1, X_2, X_3 be samples from a Poisson distribution with parameter λ. Suppose $X_1 = 3$, $X_2 = 0$, and $X_3 = 4$. Find a 95% confidence interval for λ. [Hint: the endpoints of the interval are in the set $\{1.003, 2.001, 3.619, 4.513, 4.925\}$.]

5. Let $X_1, \ldots, X_{20}$ be samples from a Poisson distribution with parameter λ. Suppose that $X_1 + \cdots + X_{20} = 123$. Using a likelihood ratio test, test the hypothesis that $\lambda = 7$ at the 5% significance level.

6. Let $X_1, \ldots, X_{50}$ be samples from a Normal distribution with mean μ and variance σ^2. Suppose that the sample mean is 2.3 and the sample variance is 6.25. Using a likelihood ratio test, test the hypothesis that $\sigma = 3$ at the 5% significance level.

7. In a sample of 100 variables $X_1, \ldots, X_n$, it is believed that the X_i are independent samples from a binomial distribution with $n = 2$ and $p = 0.1$. The following results are obtained:

Value	0	1	2
Frequency	81	16	3

Test the hypothesis using a chi-squared test at the 10% significance level.

8. A scientist is studying the effects of vitamin supplements on intelligence. She takes 1000 subjects, and gives vitamin supplements to 500 of them for a period of 3 months. Then she gives them all a standard test. The results are below:

	Pass	Fail	total
Supplement	284	216	500
No supplement	233	267	500
total	517	483	

(a) Test the hypothesis that results were independent of whether the subjects had taken the vitamin supplement at the 5% significance level.

(b) Does this show that the vitamin supplement causes individuals to get better scores in the tests. If not, give a possible alternative explanation.

9. A scientist is investigating whether global warming is causing hurricanes in more regions. He looks the history of 200 cities, and whether they experienced hurricanes in the period 1990–2000 or in the period 2000–2010. The results are below:

	Hurricanes 2000–2010	No Hurricanes 2000–2010	total
Hurricanes 1990–2000	67	14	81
No Hurricanes 1990–2000	19	100	119
total	86	114	

Test the hypothesis that the probability of a city experiencing a hurricane was the same for these two periods.

10. A coin is tossed 100 times. It comes up heads 39 times. Find a 10% significance region for the probability that the coin comes up heads, based on a chi-squared statistic $\frac{(N-\mathbb{E}(N))^2}{\mathbb{E}(N)} + \frac{(N-\mathbb{E}(N))^2}{100-\mathbb{E}(N)}$.

11. (a) A coin is tossed 10 times. The probability of coming up heads is p. Show that the number of times heads comes up is a sufficient statistic for p.

    (b) What is the probability of the sequence HHTTHHHTHTT given that the number of heads is 5.

12. The number of individuals taking an online survey follows a Poisson distribution with parameter 200. A particular question has two answers. Let the number of people giving each answer be A, B. We want to estimate the probabilities of each answer.

    (a) Show that $A+B$ is an ancilliary statistic.

    (b) In the case, $A = 89$, $B = 98$, use a conditional test, to calculate the significance of this result for the hypothesis that the probability of A is 0.6. [You may use a normal approximation to the Binomial distribution.]

13. In two years, the results in a certain course at a university are:

Grade	Year 1	Year 2
Pass	13	17
Fail	8	4

    Calculate the exact significance level for testing whether the probabilities of passing were the same in the two years. [Hint, the total number of results where the total number of students who pass during the two years is 11058116888.]

14. The number of insurance claims paid on a certain policy is assumed to follow a binomial distribution with $n = 2$ and unknown p. The observed frequencies are :

Number	Frequency
0	87
1	151
2	102

    Calculate the exact significance level, using a conditional test of the observed frequencies under the Binomial hypothesis.