Stat 342 Exam 2
Fall 2014

I have neither given nor received unauthorized assistance on this exam.

Name Signed	**Date**

Name Printed

There are 10 questions on the following 5 pages. Do as many of them as you can in the available time. I will score each question out of 10 points AND TOTAL YOUR BEST 7 SCORES. (That is, this is a 70 point exam.)

1. Here are some facts about geometric distributions that you may use in what follows:

We'll say that $X \sim \text{Geo}(p)$ provided it has pmf

$$f(x \mid p) = \begin{cases} p(1-p)^{(x-1)} & \text{for } x = 1,2,3,4,\ldots \\ 0 & \text{otherwise} \end{cases}$$

For such a variable

$$\text{E}X = \frac{1}{p} \quad \text{and} \quad \text{Var}\, X = \frac{1-p}{p^2}$$

10 pts **a)** What is the Fisher information in a single observation X about the parameter p? (Give an expression involving p for this.)

10 pts **b)** Suppose that $X_1, X_2, \ldots, X_n$ are iid $\text{Geo}(p)$. Identify a 1-dimensional sufficient statistic for this model. Justify your choice (say why you know your statistic is sufficient).

10 pts

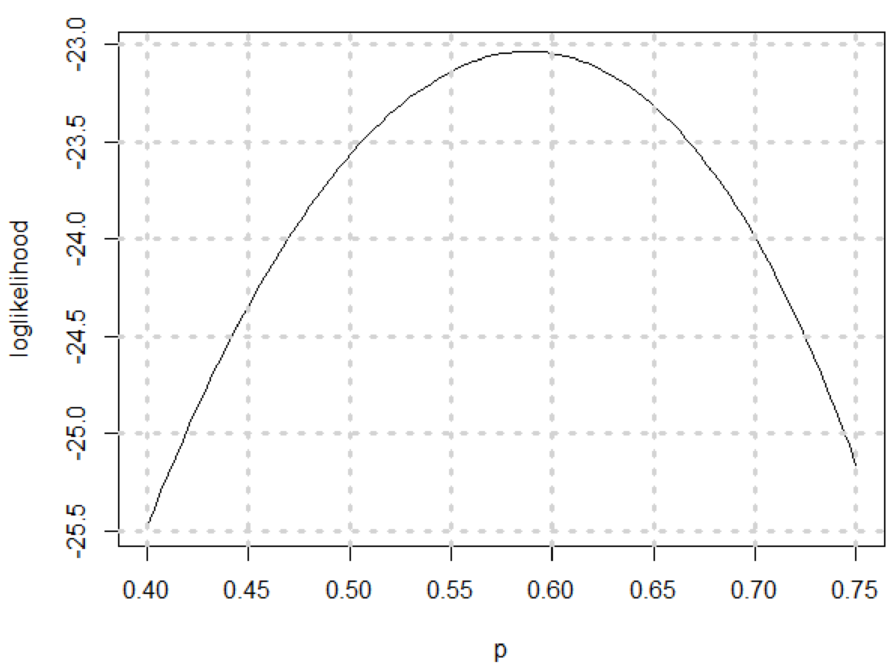
c) Suppose that a sample of $n = 10$ $\text{Geo}(p)$ variables produce the values below.

$$1,1,1,1,3,1,2,1,2,2$$

A maximum likelihood estimate of p based on this sample is $\frac{2}{3}$. Use this fact and find approximate 95% two-sided confidence limits for p .

10 pts

d) Below is a plot of a geometric log-likelihood for a different sample of iid geometric observations. Use it and the fact that the upper 5% point of the χ_1^2 distribution is 3.841 to give approximate 95% confidence limits for this p . (The largest loglikelihood pictured below is about -23.035 .) Indicate how you arrive at your limits.



2. Below is a table specifying pmfs for two possible discrete distributions for a random variable X. Call those $f(x|0)$ and $f(x|1)$. Use them in what follows.

	x									
	1	2	3	4	5	6	7	8	9	10
$f(x\|1)$	.08	.10	.05	.12	.08	.10	.10	.12	.05	.20
$f(x\|0)$	.04	.16	.10	.10	.08	.12	.14	.06	.10	.10

10 pts

a) Identify a minimal sufficient statistic T in the statistical model where observable $X \sim f(x|\theta)$ for $\theta \in \{0,1\}$. Provide values for $T(x)$ in the table below. What guarantees that your answer is minimal sufficient?

x	1	2	3	4	5	6	7	8	9	10
$T(x)$										

Rationale:

10 pts

b) Consider a prior distribution for θ in the statistical model of part **a)** that has $g(0) = .6$ and $g(1) = .4$. Identify any test $a^{\text{opt}}(x)$ that has minimum (0-1 loss) Bayes risk in this context (find a Bayes optimal predictor for θ here). Provide values for $a^{\text{opt}}(x)$ in the table below.

x	1	2	3	4	5	6	7	8	9	10
$a^{\text{opt}}(x)$										

10 pts **c)** For your test in **b)** evaluate Type I and Type II error probabilities.

Type I:

Type II:

10 pts **3.** In a non-parametric bootstrap sample of size n from n distinct values $x_1, x_2, \ldots, x_n$, what is the probability that value x_1 does not occur in the bootstrap sample? This probability has a limit as $n \to \infty$. What is this limit?

What then is the limit of

$$\mathrm{E}\frac{1}{n}\sum_{i=1}^{n} I\left[x_i \text{ is not in the bootstrap sample}\right]$$

(the expected fraction of values not appearing in the bootstrap sample) as $n \to \infty$?

10 pts **4.** Suppose that $X \sim \text{Binomial}(p)$ and that

$$T(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{otherwise} \end{cases}$$

What is the Fisher information in $T(X)$ about p? (Note that $T(X) \sim \text{Binomial}\left(1-(1-p)^2\right)$.)

How do you expect this to compare to the Fisher information in X about p? Why?

10 pts **5.** Suppose that $X \sim \text{Poisson}(\lambda)$ and that what is of interest is (not λ itself, but rather) $\exp(-2\lambda)$. Show that the estimator

$$a(X) = \begin{cases} -1 & \text{if } X \text{ is odd} \\ 1 & \text{if } X \text{ is even} \end{cases}$$

is unbiased for $\exp(-2\lambda)$. (As it turns out, it is the *only* unbiased estimator, and therefore by default the *best* one!)

$a(X)$ is clearly silly. Identify a modification of it that you are sure will have better SEL risk function.