Some Practice Problems

1. Consider a sequence of independent trials in which a success occurs with probability p and a failure occurs with probability $1-p$, and let the random variable X denote the number of failures before the r^{th} success. This problem is concerned with making inferences about the parameter

$$\theta = \frac{1-p}{p},$$

based on a random sample $X_1, \ldots, X_n$.

(a) Show that the probability mass function (pmf) (or frequency function) of X can be written in terms of θ as

$$\mathrm{P}(X=x) = \binom{r+x-1}{x} \frac{\theta^x}{(1+\theta)^{r+x}} \qquad x = 0, 1, \ldots$$

and give $E(X)$ and $Var(X)$ in terms of θ.

(b) Show that $W = \sum_{i=1}^n X_i$ is a sufficient statistic for θ. Give the pmf of W.

(c) Give $E(W)$. Find an unbiased estimate of θ based on W. Give the variance of your estimate. Is your estimate consistent for θ?

(d) Find the Cramer-Rao lower bound for the variance of an unbiased estimate of θ. Is your estimate in (c) an efficient estimate?

(e) Find the maximum likelihood estimate $\hat{\theta}$ of θ. What is the approximate distribution of $\hat{\theta}$ for large n?

(f) Show that the most powerful test of significance level α of $H_0 : \theta = \theta_0$ against $H_A : \theta = \theta_1$ ($\theta_1 < \theta_0$) rejects H_0 if $W \le c$, and explain how you might determine the constant c. Is this test a uniformly most powerful test of H_0 against the composite alternative $H_A^* : \theta < \theta_0$?

2. According to the Hardy-Weinberg formula, the number of flies resulting from certain crossings should be in the proportions $q^2 : 2pq : p^2$, where $p + q = 1$.

(a) Suppose an experiment yielded the frequencies 80, 90, 40. Are these data consistent with this formula with $q = 0.8$?

(b) Suppose q is unspecified. Show that its maximum likelihood estimate is

$$\hat{q} = \frac{2n_1 + n_2}{2(n_1 + n_2 + n_3)},$$

where n_1, n_2, n_3 are the frequencies observed in the three categories.

(c) Are the data in (a) compatible with the formula when q is unspecified?

3. Let $X_1, \ldots, X_n$ be a random sample on X, where X has a gamma distribution with parameters $1/2$ and $1/\theta$, with probability density function

$$f(x|\theta) = \frac{1}{\theta^{\frac{1}{2}}\Gamma(\frac{1}{2})} e^{-\frac{x}{\theta}} x^{-\frac{1}{2}} \qquad x > 0$$

($\Gamma(1/2) = \sqrt{\pi}$), so that $E(X) = \theta/2$ and $Var(X) = \theta^2/2$. Put $Y = \sum_{i=1}^n X_i$.

(a) Show that $W = 2Y/\theta$ has a χ_n^2 distribution.

(b) Find the Cramer-Rao lower bound for the variance of an unbiased estimate of θ.

(c) Show that $2\bar{X}$ is the maximum likelihood estimate of θ. Is it an efficient estimate?

(d) Find a sufficient statistic for θ.

(e) Show that the most powerful test of significance level α of $H_0: \ \theta = \theta_0$ against $H_A: \theta = \theta_1$ $(\theta_1 < \theta_0)$ rejects H_0 if $Y \leq c$. Explain how c can be found using tables of the χ^2 distribution. Is this test a uniformly most powerful test of H_0 against the composite alternative $H_A^*: \ \theta < \theta_0$?

(f) Express the power function $\pi(\theta)$ of the test in (e) as a χ^2 probability. Show that, when $n = 2$, $c = \theta_0 \log(1-\alpha)^{-1}$ and $\pi(\theta) = 1 - (1-\alpha)^{\theta_0/\theta}$, and sketch $\pi(\theta)$.

(g) Describe, as completely as possible, the likelihood ratio test of level α of $H_0: \ \theta = \theta_0$ against the composite alternative $H_A^*: \ \theta < \theta_0$.

4. In an investigation studying the total acidity of coal, three different methods were used on three types of coal and the experiment was replicated twice. The following table shows the values of total acidity (minus 7) that were obtained:

		Coal Type			
		1	2	3	Total
Method	1	1.27 1.17 (2.44)	1.66 1.61 (3.27)	1.44 0.96 (2.40)	8.11
	2	1.03 1.21 (2.24)	1.42 1.58 (3.00)	1.02 0.89 (1.91)	7.15
	3	1.68 1.28 (2.96)	1.69 1.84 (3.53)	1.21 1.51 (2.72)	9.21
	Total	7.64	9.80	7.03	24.47

Subclass totals are in parentheses. The sum of squares of all the observations is 34.6417.

(a) Analyze these data, and state your conclusions clearly.

(b) Give a model for the data and state any assumptions made. Express any hypotheses you tested in your analysis in (a) in terms of the parameters in the model.

(c) Give the fitted values corresponding to each of the observations.

5. The number of cars abandoned in a town of 50,000 people, in the 104 weeks of 1989-90, is given in the following table"

# weeks	40	43	17	4	3	0
# cars	0	1	2	3	4	≥ 5

Let X denote the number of cars abandoned per week. Suggest a probability distribution for X and test, at the 1% level, the null hypothesis that X has this type of distribution.

6. An experimenter knows that the distribution of the lifetime X (in hours) of a certain component has an exponential distribution with mean $1/\lambda$; i.e., the probability density function of X is

$$f(x|\lambda) = \lambda e^{-\lambda x} \quad x \geq 0$$

On the basis of a random sample of n lifetimes $X_1, X_2, \ldots, X_n$, the experimenter is interested in making inferences about the *median* lifetime θ, defined by the equation $\mathrm{P}(X < \theta) = 1/2$.

(a) Show that $\theta = \lambda^{-1} \log 2$.

(b) Find the maximum likelihood estimate $\hat{\theta}$ of θ (based on the sample of n lifetimes). Show that it is an unbiased estimate of θ, and that its variance is θ^2/n.

(c) Find the Cramer-Rao lower bound for the variance of an unbiased estimate of θ. Is the maximum likelihood estimate $\hat{\theta}$ you found in (c) an efficient estimate?

(d) What is the approximate distribution of $\hat{\theta}$ if the sample size n is large?

(e) Put $Y \equiv \sum_{i=1}^{n} X_i$. Using moment generating functions (or otherwise), show that $2\lambda Y$ has a χ^2_{2n} distribution.

(f) Show that the most powerful test with significance level α of the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_A : \theta = \theta_1$, where $\theta_1 < \theta_0$, rejects H_0 if $Y \leq c$, where Y is defined in (e). Explain how the constant c can be found using tables of the χ^2 distribution. Is this test a *uniformly most powerful test* of H_0 against the composite alternative hypothesis $H_A^* : \theta < \theta_0$?

(g) Find, and sketch, the power function of the test you found in (h).

7. Let $X_1, X_2, \ldots, X_n$ be a random sample on X, where X has a $N(\theta, \theta)$ distribution, where both the mean and the variance are equal to θ ($\theta > 0$), which is unknown.

(a) Find a sufficient statistic T for θ.

(b) We know that the sample mean is always an unbiased estimate of the mean, and that the sample variance is always an unbiased estimate of the variance. Here, this means that both

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

are unbiased estimates of θ.

(i) Is either $\bar{X}$ or S^2 a function of the sufficient statistic T you found in (a)?

(ii) Find (or state) the variances of $\bar{X}$ and S^2 and sketch them as functions of θ on the same graph. Does either of these estimates always have a smaller variance than the other? If the unknown value of θ is known to be large, which of these estimates would you prefer?

(c) Without doing the actual calculations, explain how you would find an unbiased estimate of θ with smaller variance than either $\bar{X}$ or S^2.

8. An experiment was designed to investigate the relative merits of four different feeds on the body weight of pigs. Twenty pigs were placed at random into four pens, so that each pen held five pigs. Each pen was given a different feed. The weight gained (in pounds) by each pig after a fixed period of time is given in the following table:

	Feed A	Feed B	Feed C	Feed D
	133	163	210	195
	144	148	233	184
	135	152	220	199
	149	146	226	187
	143	157	229	193
Total	704	766	1118	958

$$133^2 + 144^2 + \cdots + 193^2 = 650{,}848$$

$$704^2 + 766^2 + 1118^2 + 958^2 = 3{,}250{,}060$$

(a) Analyze these data to determine if the feeds have the same effect on weight gain.

(b) If your analysis in (a) leads you to conclude that there are differences between feeds, determine which pairs differ significantly at the 5% level.

(c) Find a 95% confidence interval for $\mu_1 - \mu_2$, the difference between the mean weight gains for feeds A and B.