SAMPLE SOLUTIONS

APPLIED MATHEMATICS 166 | **EXAM I**

March 18, 2003 | Open Book Exam

Please write your name at the top of this page and do your work directly on these sheets.

1. On average, the number of babies born in Cleveland, OH in September is 1472. On January 26, 1977 the city was immobilized by a blizzard. Nine months later, in September 1977, the number of recorded births was 1718, an increase of 246. The following analysis is aimed at determining whether this difference is significant. It is reasonable to assume that the number of births during some month, e.g., September 1977, has a Poisson distribution.

a) Suppose Y is a Poisson distributed rv with parameter λ. Let λ_0 and λ_1 denote two fixed values of λ with $\lambda_1 > \lambda_0$. Suppose we have a single observation of Y. What is the *form* of the rejection region for the most powerful test of

$$H_0 : \lambda = \lambda_0$$

vs.

$$H_a : \lambda = \lambda_1?$$

b) Is the test in part (a) *uniformly most powerful* for testing

$$H_0 : \lambda = \lambda_0$$

vs. the composite alternative

$$H_a : \lambda > \lambda_0?$$

If your answer is affirmative, please state *why*? Otherwise, what rejection region would you use for this hypothesis pair?

c) When λ is large, then Y is approximately normal with mean and variance λ. Use this fact to figure out the *approximate* constants in the specification of the rejection region determined for part (b) above using the Cleveland data and taking $\alpha = 0.05$. Here, $\lambda_0 = 1472$. What is the decision when $Y = 1718$?

d) What is the power of your test if the true value of λ is 1550?

next page

1) a. Simple Null hypothesis vs. Simple alternative. Neymann-Pearson lemma assures that the likelihood ratio test is "most powerful".

$$\text{Ratio} \quad \Lambda = \frac{\lambda_0^Y e^{-\lambda_0}}{Y!} \Big/ \left\{ \frac{\lambda_1^Y e^{-\lambda_1}}{Y!} \right\} = \left(\frac{\lambda_0}{\lambda_1}\right)^Y e^{-(\lambda_0-\lambda_1)}$$

Form of rejection region for likelihood ratio test:

$$\Lambda < k \Leftrightarrow Y\underbrace{[\ln\lambda_0 - \ln\lambda_1]}_{\text{negative}} - \lambda_0 + \lambda_1 < \ln k$$

$$\Leftrightarrow \underline{\underline{Y > c}}$$

b. Test is uniformly most powerful since rejection region is independent of the value of λ for $\lambda > \lambda_0$.

c. $\dfrac{Y-\lambda_0}{\sqrt{\lambda_0}}$ is approximately standard normal if H_0 is true.

$$Y > c \Leftrightarrow \frac{Y-1472}{\sqrt{1472}} > \frac{c-1472}{\sqrt{1472}}$$

$$\alpha = 0.05 = P(Y>c) \cong P\left(Z > \frac{c-1472}{\sqrt{1472}}\right) \Rightarrow \frac{c-1472}{\sqrt{1472}} = z_\alpha = 1.645$$

$$\Rightarrow c = 1472 + \sqrt{1472}\cdot 1.645 = 1538.1$$

Reject when $Y > 1535.1$. When $Y = 1718$, reject H_0.

d. If $\lambda = 1550$, then power is $P(Y > 1535.1) =$

$$P\left(\frac{Y-1550}{\sqrt{1550}} > \frac{1535.1-1550}{\sqrt{1550}}\right) \cong P(Z > -0.38) = 1 - 0.3520$$

$$= \underline{0.6480}$$

2. The first 608 digits in the decimal expansion of π have the following frequencies:

0	1	2	3	4	5	6	7	8	9
60	62	67	68	64	56	62	44	58	67

a) How do the data jibe with the assumption that each digit is equally likely?

b) Test

$$H_0 : P(7) = 0.1$$

vs.

$$H_a : P(7) \neq 0.1?$$

a) Chi-square goodness-of-fit test of

H_0: $p_0 = p_1 = \cdots = p_9 = 0.1$

Compute $\chi^2 = \sum_{i=0}^{9} \frac{(n_i - 608(0.1))^2}{608(0.1)} = \frac{1}{60.8} \sum_{i=0}^{9} (n_i - 60.8)^2$

$\chi^2 \doteq 7.493$

If H_0 is true, this statistic is chi-square with 9 degrees of freedom.

From Table 6, $\chi^2_{9,0.1} = 14.68$, so the p-value exceeds 0.1.

Therefore, DO NOT REJECT H_0.

b) Bernoulli experiment, with $H_0: p = p_7 = 0.1$ vs $H_a: p \neq 0.1$

Z-test, since $n = 608$ is large

$$Z = \frac{X - n \cdot p}{\sqrt{npq}} = \frac{44 - 60.8}{\sqrt{608(0.1)(0.9)}} = 2.27 > z_{0.025} = 1.96$$

$\therefore$ Reject H_0 at level $\alpha = 0.05$.

3. The following table gives the daily rental car rates charged by two companies (Alamo and Avis) in 12 locations (*USA Today,* March 14, 1994). The rate is for a mid-size car rented midweek with three days advance notice. Alamo claims to be a low-budget agency. Do the prices quoted here lend credence to Alamo's claim? Test an appropriate H_0 versus H_a letting $\alpha = 0.05$. Consider different approaches for the test and articulate the assumptions they rely on.

	City:	ATL	ORD	DFW	DEN	LAX	MIA	EWR	PHX	SFO	STL	SEA	DCA
X	Alamo:	48.99	49.99	42.99	34.99	42.99	33.99	59.99	42.89	47.99	47.99	35.99	44.99
Y	Avis:	51.99	55.99	47.00	42.99	44.95	38.99	69.99	50.99	49.99	53.99	42.99	44.99

Both the Sign Test and the Wilcoxon Signed Rank Test are appropriate

a) the Sign Test assumes independent observations from city-to-city. The observations need to be paired. Only at DCA is there a tie. Throw that case out. Out of the remaining $n = 11$ cases, the Difference

$D_i = X_i - Y_i$ is negative in every case. If $p = P(X < Y) = 1/2$,

the probability of this happening is $\binom{11}{0}\left(\frac{1}{2}\right)^{11} = \frac{1}{2048} = 0.000488$

The p-value is .0005 for a one-sided alternative and .001 for a two-sided alternative. Alamo's claim is substantiated.

b) Wilcoxon is even more powerful, so we know what the answer will be already. Wilcoxon assumes that the distributions of X and Y may differ in location, but not in shape. The null hypothesis is that the distributions are the same. Throw out the one tie, so $n = 11$.

$T^+ = 0 \quad T^- = \sum_{i=1}^{11} i = 66$. From Table 9, the p-value is less than 0.005 for a one-sided alternative and less than 0.01 for a two-sided alternative. Alamo's claim is substantiated.

$\mathcal{S}$

4. A study of sterility in the fruit fly ("Hybrid dysgensis in drosophila melanogaster: the biology of female and male sterility," *Genetics*, 1979, pp161–174) reported the following data on the number X of ovaries developed for each female fly in a sample of size 1388. One model for unilateral sterility states that each ovary develops with some probability p independently of every other ovary; and hence X should have a binomial distribution with parameters 2 and p. An opposing theory claims that the development of the ovaries within one individual is *not* independent, and hence X should not be binomial.

X = Number of ovaries developed:	0	1	2	
Observed count:	1212	118	58	$1388 = n$

a) We need to estimate p. Assume that X is binomial. First express $P(X = 0)$, $P(X = 1)$ and $P(X = 2)$ in terms of p and then give the expression for the likelihood function $L(p)$. (Think of $L(p)$ as "the probability of what was observed.")

b) What is the maximum likelihood estimator $\hat{p}$ of p?

c) Carry out the appropriate goodness-of-fit test of the hypothesis that X has the claimed binomial distribution (use $\alpha = 0.01$).

This is just like Problem #2 on the Sample Problems.

a) $P(X=k) = \binom{2}{k} p^k (1-p)^{2-k}$ for $k = 0, 1, 2$.

Let n_k denote the number of times $X_i = k$ in the $n = 1388$ trials.

$$L(p) = \frac{1388!}{n_0!\, n_1!\, n_2!} \left\{(1-p)^2\right\}^{n_0} \left\{2p(1-p)\right\}^{n_1} \left\{p^2\right\}^{n_2}$$ (multinomial)

$$L(p) = \frac{2 \cdot 1388!}{n_0!\, n_1!\, n_2!}\, p^{n_1+2n_2} (1-p)^{n_1+2n_0}$$

b) $\ln L(p) = c + (n_1+2n_2)\ln p + (n_1+2n_0)\ln(1-p)$ differentiate ---

$(n_1+2n_2)/\hat{p} - (n_1+2n_0)/(1-\hat{p}) = 0$ & solve

$\Rightarrow \hat{p} = (n_1+2n_2)/[2(n_0+n_1+n_2)] = 234/2776 = \underline{\underline{0.08429}}$

c) Using $\hat{p} = 0.08429$, $P(X=0) = 0.84 = p_0$; $P(X=1) = 0.15$; $P(X=2) = 0.01$

$0.8385 = p_0$ $\quad$ $0.1544 = p_1$ $\quad$ $0.0071 = p_2$

(next page)

4 continued

k	0	1	2
p_k	0.8385	0.1544	0.0071
np_k	1163.8	214.3	9.9

$$\chi^2 = \sum_{i=0}^{2} \frac{(n_k - n\hat{p}_k)^2}{n\hat{p}_k} = \frac{(1212-1163.8)^2}{1163.8} + \frac{(118-214.3)^2}{214.3} + \frac{(58-9.9)^2}{9.9}$$

$$= 1.99 + 43.27 + 233.70 = \underline{\underline{278.97}}.$$

Under H_0, χ^2 has a chi-square distribution with one degree of freedom. From Table 6, $\chi^2_{1,0.005} = 7.879$

Thus, we reject H_0 with a p-value < 0.005