MATH 1180: Final, May 1, 2012 NAME: Fred Adler

Do all five problems, each worth 40 points. Make sure I can find both the answer and how you got it. You can use 5 pages of notes (each can be written on both front and back). No calculators. Each extra credit problem is worth 2 points.

1. Over 20 years, a dedicated birdwatcher goes looking for Virginia's warblers in Dry Creek Canyon every May 1. He sees 0 warblers in one year, 1 warbler in twelve of the years, 2 warblers in five of the years, and 3 warblers in the remaining two years.

- **a.** Why is a Poisson distribution is a good choice of distribution to describe these data? Why not?
- **b.** Write the likelihood function for the parameter or parameters of the Poisson distribution using these data.
- **c.** Find the maximum likelihood estimate(s).
- **d.** How would you find the expected distribution with these estimates? How many years with 0 warblers would you expect?
- **e.** Explain the steps you would use to test if the observed and expected data are different, illustrating with the years with 0 warblers. What problems could arise with this procedure?

Extra credit: What is one thing you learned about the history of statistics in this class?

number	value
$e^{0.7}$	2.0
$(0.7)^2$	0.5

a. Count data are appropriate for a Poisson distribution, but we don't know that the birds are independent.

b. $\mathcal{L}(\Lambda) = \frac{e^{-\Lambda}\Lambda^0}{0!}\cdot\left(\frac{e^{-\Lambda}\Lambda^1}{1!}\right)^{12}\cdot\left(\frac{e^{-\Lambda}\Lambda^2}{2!}\right)^5\cdot\left(\frac{e^{-\Lambda}\Lambda^3}{3!}\right)^2 = \frac{e^{-20\Lambda}\cdot\Lambda^{28}}{u}$ where u is some number

c. let $S(\Lambda) = \ln(\mathcal{L}(\Lambda)) = \ln\left(\frac{e^{-20\Lambda}\Lambda^{28}}{u}\right) = -20\Lambda + 28\ln(\Lambda) - \ln(u)$

$\frac{dS}{d\Lambda} = -20 + \frac{28}{\Lambda} \Rightarrow \hat{\Lambda} = \frac{28}{20} = 1.4$

d. I'd evaluate the Poisson distribution with $\Lambda = 1.4$, so there would be $20\cdot\frac{e^{-1.4}\cdot 1.4^0}{0!} \approx 20\cdot(e^{-0.7})^2 = \frac{20}{2^2} = 5$

e.

	0	1	2	3
Observed	1	12	5	2
Expected	5	$\frac{20\hat{\Lambda}^1 e^{-\hat{\Lambda}}}{1!}$	etc	

I'd find the χ^2 statistic as $\sum \frac{(O-E)^2}{E}$. The first value is $\frac{(1-5)^2}{5} = 3.2$

The problem would be deciding where to combine data to avoid $E < 5$.

2. A larger survey of Virginia's warblers was conducted in Parched Creek Canyon over the course of 25 years starting in 1930 with an average number seen of 20 birds. The survey was then repeated starting in 1985 for another 25 years with an average of 17 warblers seen per year.

   a. Suppose first that the number of birds found in a given year is known to be normally distributed with a mean of 20.0 and a variance of of 36.0. What is the probability of seeing 17 or fewer warblers in one year? Sketch a graph illustrating this probability

   b. Under the same assumptions as in **a.**, what is the probability of seeing an average of 17 or fewer warblers over the course of 25 years? What is the significance level? Sketch a graph illustrating this probability

   c. Suppose instead that the sample variance was found to be 36.0 in the both the first and second sets of 25 years. Explain the steps you would use to test whether the warbler populations had changed between the two surveys, including a clear statement of the null hypothesis.

   d. What are two problems with this experiment?

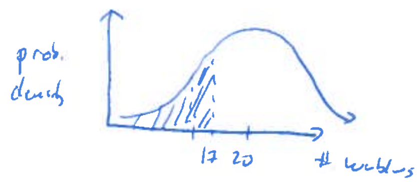
Extra credit: What is one thing you learned about the history of mathematics in this class?

z	$\Phi(z)$
0.0	0.50000
0.5	0.69146
1.0	0.84134
1.5	0.93319
2.0	0.97725
2.5	0.99379
3.0	0.99865
3.5	0.99977
4.0	0.99997

a. $z = \frac{17-20}{\sqrt{36}} = -\frac{3}{6} = -0.5$. $\Phi(-0.5) = 1-\Phi(0.5) = 1-0.69146 = 0.30854$

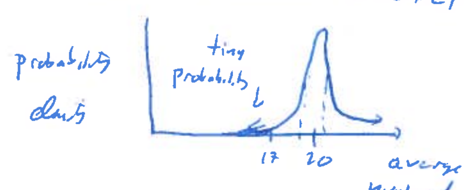
b. The average has a standard deviation of $\sqrt{\frac{\sigma^2}{n}}$

$= \sqrt{\frac{36}{25}} = \frac{6}{5} = 1.2$



Then $z = \frac{17-20}{1.2} = -2.5$. $\Phi(-2.5) = 1-\Phi(2.5) = 1-0.99379 = 0.00621$

This is highly significant.



number of warblers

c. The null hypothesis is that the mean number is the same, or that the difference in the means is 0. I'd find the standard error of this difference by adding those for each set of years, and find the t-statistic by dividing the observed difference of 3 by this value

d. The birds are not necessarily independent in consecutive years, and we don't know whether the survey methods changed.

3. Virginia's warblers arrive in Arid Creek Canyon at rate $\lambda = 3.0$/day during migration for 10 days starting on May 1. They leave at per capita rate $\mu = 0.5$/day starting on May 1 until all of them are gone.

   a. Draw a diagram showing the number of warblers in the canyon from May 1-10.

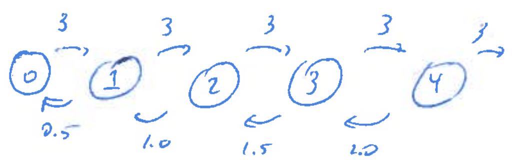
   b. Draw a diagram showing the number of warblers in the canyon after May 10.

   c. Write a differential equation for the expected number of warblers in the canyon from May 1-10 and find the equilibrium.

   d. Suppose that the number of warblers on May 10 is that equilibrium number. How many warblers on average would be left exactly two days later? What is the distribution?

Extra credit: What is one thing you learned about the history of music in this class?

a.



b. Same, but no new birds, a

c. $\frac{dw}{dt} = 3 - 0.5w$, $w^* = 6$

d. If birds leave at rate 0.5, the probability an individual remains after 2 days is $e^{-0.5\cdot 2} = e^{-1}$

On average $6e^{-1}$ would be left, according to a binomial distribution with $n=6$, $p=e^{-1}$

4. Male Virginia's warblers sing to impress females. Some males have songs that are more melodious than others, measured in units of melodiocity. A study in Desiccated Creek Canyon found the following:

Male	Melodiocity	Number of females impressed	$\hat{F}$	ε	Null	ε_{null}
A	5	2	2	0	4	-2
B	8	6	5	1	4	2
C	7	3	4	-1	4	-1
D	6	3	3	0	4	-1
E	9	6	6	0	4	2

a. What is the null hypothesis?

b. Write a linear model for the relationship between melodiocity and the number of females impressed.

c. Find a simple reasonable guess for the parameters of this model and graph it.

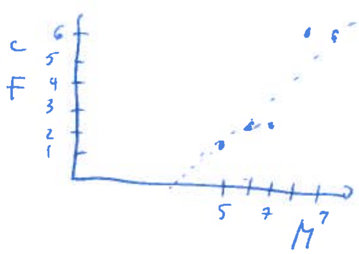
d. Summarize the strength of this relationship in single statistic.

e. What would you have to do to statistically test this relationship?

Extra credit: What is first thing that you learned in this class that you plan to forget?

a. The number of females impressed, F, is not affected by the melodiocity M

b, c



I'd guess a slope of 1 passing through (5, 2) & (9, 6)

$F = 1(M-5)+2 = M-3$

In general, $F = aM + b + \varepsilon$

d. $SS_E = 2$

To find SST, we find $\bar{F} = \dfrac{2+6+3+3+6}{5} = 4$

$SST = (-2)^2 + 2^2 + (-1)^2 + (-1)^2 + (2^2) = 14.$ $\quad r^2 = 1 - \frac{4}{14} = \frac{5}{7} \approx 0.71$

e. Could compare the likelihoods of the maximal likelihood model and the null model, but we need to estimate σ in $\varepsilon \sim N(0, \sigma^2)$ to do this

5. People are worried that surveys are biased because singing birds are easier to locate than silent birds. This is tested by an intensive survey in Dusty Creek Canyon using highly trained Virginia's warbler-tracking flying dogs that find every single bird no matter what it is doing (but without harming or disturbing it in any way). It turns out that human observers locate 90% of singing birds but only 30% of silent birds, and that 1/3 of birds are singing at any one time.

   a. What is the probability that a bird that is located is singing?

   b. Suppose these data had really been collected on 90 birds. Write the table of data that simultaneously describes whether birds were singing and whether they were located by people.

   c. What would the table of data be under the null hypothesis that singing birds are no easier to locate than silent birds?

Extra credit: What is one thing that you learned in this class that you plan to remember?

a. $$\Pr(S|L) = \frac{\overbrace{\Pr(L|S)}^{0.9} \cdot \overbrace{\Pr(S)}^{1/3}}{\underbrace{\Pr(L)}_{0.5}} = \frac{0.3}{0.5} = 0.6$$

$$\Pr(L) = \Pr(L|S)\Pr(S) + \Pr(L|\text{not singing})\Pr(\text{not}) = (0.9)(\tfrac{1}{3}) + (0.3)(\tfrac{2}{3}) = 0.3 + 0.2 = 0.5$$

b. 30 birds singing → 27 located, 3 not

60 birds not singing → 18 located, 42 not

	Singing	not singing
located	27	18
not located	3	42

c. Half are located, $\frac{1}{3}$ are singing, so we would have $\frac{1}{6} \cdot 90 = 15$ located and singing if independent

	Singing	not singing
located	15	30
not located	15	30

6. Extra credit: Before looking, guess the theme of this test: