36-720: Discrete Multivariate Analysis
HW03, Due Wednesday, October 3, 2007

Announcements:

- We are going to take a digression into considering zero counts in tables (Christensen, Chapter 8) before continuing with model selection/criticism (parts of Christensen, Chapter 6; see also Edwards, Ch's 2 & 4 for applications and Ch's 5 & 6 for theory).
- Please remember that, although you are free to talk with one another about HW's, the work you turn in should be your own.
- Best to use a word processor that can handle mathematics (like LaTeX) and can include graphics from R and other programs. Next best is neat handwriting with neatly cut-and-pasted tables and graphs.

Problems:

1. Christensen, p. 209, #5.5.1
2. Christensen, p. 210, #5.5.4. For each of the five examples,
   (a) If the graph is not graphical, or graphical but not decomposable, prove it.
   (b) If the graph is decomposable, give the formula for the MLE of the expected cell counts.
3. The table below shows data (Fienberg, 1980) on an experiment in which the food additive Red Dye No. 2 was fed to two different groups of rats at different dosages. Some rats died during the 131 week experiment; the survivors were sacrificed. All rats biopsied after death to see if they had cancer.

		Age at Death	
Dosage	Cancer	0-131 Weeks	Sacrificed at End
Low	Present	4	0
	Absent	26	14
High	Present	7	7
	Absent	16	14

For brevity, we will use A = Dosage, B = Age at Death, and C = Cancer.

I. Using `glm` and whatever other tools you need in R, compute estimated expected cell counts, the G^2 and X^2 test statistics, and d.f. for each of the following four models:
   (a) C is independent of A and B.
   (b) Given A, C is independent of B.
   (c) Given B, C is independent of A.
   (d) No three-way interaction among A, B and C.

II. Comment on the relationship between X^2 and G^2 for these models. Then say which model fits the best, and why, and interpret that model in terms of the original variables.

4. Continuing with the previous problem...

   (a) Show, using `glm` and appropriate reduced data sets, that the likelihood ratio test for model (a) vs model (c) in Problem 3, yields *exactly the same test statistic* as the G^2 test for independence of B and C in the two-way marginal table for B and C.

   (b) Show, algebraically, that the result in part (a) of this problem *must* hold.

   (c) What feature of model (c) in Problem 3 is this exercise illustrating? Explain.

5. Consider the $2 \times 2 \times 2$ table with cell counts n_{ijk}; $i = 1, 2$; $j = 1, 2$; $k = 1, 2$.

   (a) Derive the likelihood equations for the no three-way interaction model, under Poisson sampling.

   (b) Show that when $n_{ijk} > 0 \ \forall \ i, j, k$, there exists a solution to the likelihood equations and it is unique. [Hint: You need to look at the log-likelihood and show that it is strictly concave.]

6. Consider the same $2 \times 2 \times 2$ table as in the previous problem.

   (a) Suppose that $n_{121} = n_{212} = 0$, and all other cell counts are positive. Explain why the MLE for the model of no three-way interaction, under Poisson sampling, does not exist. [Hint: write out the three 2×2 tables for the minimal sufficient statistics, and try to add $\delta > 0$ to n_{121}; show what happens.]

   (b) Suppose that $n_{111} = n_{122} = 0$, but that all other cell counts are positive. What can you say about the existence of the MLE now? Explain your answer.