36-720: Discrete Multivariate Analysis
HW01, Due September 5, 2007, 5:00pm

Announcements:

- The first part of the course follows Christensen Chapters 2, 3, and 10 (You should review, as needed, Chapters 1 and 9). This homework is mostly drawn from Chapter 2.
- 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. In class we showed that the d.f. for the χ^2 test of independence (either X^2 or G^2) for the $I \times J$ table was $(I-1)(J-1)$ under product-multinomial sampling, by counting parameters. Show the same result under single multinomial sampling.

2. *Partitioning* G^2

   (a) Let $(n_1, \ldots, n_J)$ be a J-cell table of counts following a multinomial distribution with total sample size n_+ and cell probabilities $p_1, \ldots, p_J$. Show that the multinomial probability function can be factored into a product of $J-1$ binomial probability functions.

   (b) Now consider a $2 \times J$ table $(n_{ij} : i = 1, \ldots, 2; j = 1 \ldots, J)$ [so, 2 rows and J columns], where the rows follow independent multinomials (with possibly different row totals n_{1+}, n_{2+}), and consider the likelihood ratio test statistic G^2 for independence of row and column effects (that is, $p_{1j} = p_{2j}\ \forall j$, also known as *homogeneity of proportions*). Use part (a) to show that G^2 can be partitioned into independent 1 d.f. G^2 components for $J-1$ 2×2 tables.

3. Christensen, p. 62, Ex. 2.7.2. You should use all of the tools developed in the first two lectures. Please write a short paragraph summarizing/interpreting your analysis and answering the question about which sampling model is best.

4. Christensen, pp. 64–65, Ex. 2.7.5[1] For part (b), give the name of the probability function you derive for $P(y_1 = r_1 | t = t_0)$, and explain which margins of the 2×2 table can be viewed as fixed, and why, for $P(y_1 = r_1 | t = t_0)$.

5. Consider the following table for differing values of K:

	A	$\overline{A}$	Total
B	$225+K$	$75-K$	300
$\overline{B}$	$75-K$	$25+K$	100
Total	300	100	400

   (a) Find X^2 as a function of K. how large an integer value of K would you need to ensure that the test rejects the independence of rows and columns at the 1% level?

   (b) Repeat part (a) using G^2 in place of X^2.

   (c) What do your answers in parts (a) and (b) suggest about the properties of the two goodness of fit tests?

[1] Agresti (1992 *Statistical Science*) has a nice survey of exact tests for contingency tables. See `hw01/agresti-1992-exact-inference.pdf`.