Lecture last; December 7, 1999

Comments on HW 7,8

In Hw 7, brief comments on testing $H_0 : \theta = \theta_0$ vs. $H_A : \theta = \theta_A \neq \theta_0$. The following solution due to Hong. The likelihood ratio based on $y = (y_1, \ldots, y_n)$ is

$$\mathrm{lr}_{A0}(y) = \begin{cases} \left(\frac{\theta_A}{\theta_0}\right)^{-n} & \text{if } y_{(n)} \leq \theta_0 \\ \infty & \text{if } \theta_0 < y_{(n)} \leq \theta_A \\ \text{undefined} & \text{otherwise} \end{cases}$$

For $\alpha < 1$, any test with level α is most powerful level α if $\{y : y_{(n)} > \theta_0\}$ is in the rejection region, and any such test is UMP level α. We can only get a level α test by randomization: one version that works is (mine)

$$\varphi(y) = \begin{cases} 1 & y_{(n)} > \theta_0 \\ \alpha & y_{(n)} < \theta_0 \end{cases}$$

and another is (Hong's)

$$\varphi(y) = \begin{cases} 1 & \theta_0(1-\alpha) < y_{(n)} < \theta_0 \sqrt[n]{\{\alpha + (1-\alpha)^n\}} \\ 0 & \text{otherwise} \end{cases}$$

In Hw 8, the 2×2 table that I gave you should have been the following:

	Like habits	Unlike habits	Total
Identical twins	44	9	53
Fraternal twins	9	9	18
Total	53	18	71

from which you can compute the p-value to be approximately 0.008. (It was 1 in the table on the homework.) This data is taken from Kalbfleisch, J.G. (1985): Probability and Statistical Inference, Vol.II, Springer-Verlag, who in turn quotes Fisher, R.A. (1959) Smoking and the Cancer Controversy. It refers to the smoking habits of twins, and the evidence is quite strong for a difference in the probabilities of like habits between identical and fraternal twins. This is one piece of evidence used by Fisher to argue that there may be a genetic factor influencing one's smoking habits.

Invariance and hypothesis testing

As we saw last week, all invariant tests are tests determined by the maximal invariant statistic $T = t(Y)$, say. The distribution of the maximal invariant statistic depends only on the maximal invariant $\varphi = \varphi(\theta)$, say, for the induced group on the parameter space. This often means we can reduce a composite H_0 to a simple H_0.
Example: Gamma distribution (CH 5.15)

Let $Y_1, \ldots, Y_n$ be i.i.d from a Gamma(ψ, λ) distribution with density function

$$f(y_i; \psi, \lambda) = \frac{1}{\Gamma(\psi)} y_i^{\psi-1} \exp(-\lambda y_i)$$

and consider testing $H_0 : \psi = \psi_0$ vs $H_1 : \psi > \psi_0$. The testing problem is invariant under multiplication of each observation by the same constant $a > 0$. The group on R^n is thus

$$G = \{g : gY = (aY_1, \ldots, aY_n), a > 0\}$$

where here Y stands for the vector of observations $(Y_1, \ldots, Y_n)$. The induced group on the parameter space $\Theta = \{(\lambda, \psi); \lambda > 0, \psi > 1\}$ is

$$G^* = \{g^* : g\theta = (a\lambda, \psi); a > 0\}.$$

The orbits of G are lines through 0, characterized by the angular tangents, so the maximal invariant is

$$T = (Y_1/Y_n, \ldots, Y_{n-1}/Y_n)$$

with distribution depending only on ψ. (Actually T is ancillary for λ when ψ is fixed, as is obvious from the scale structure of the model.) The joint density of T has a minimal sufficient statistic which is equivalent to $\sum \log Y_i - n \log(\sum Y_i)$, and an expression for the distribution of this statistic can be computed. (It's a bit messy.)

If we first reduce by sufficiency in the full joint distribution, the m.s.s. is $(\sum Y_i, \sum \log Y_i)$, or equivalently $S = (n^{-1} \sum \log Y_i, \log \sum Y_i)$, and under scale transformations this becomes $(\log a + n^{-1} \sum \log Y_i, \log a + \log \sum Y_i)$ with maximal invariant $n^{-1} \sum \log Y_i - \log \sum Y_i$.

In Hw8 you found that the UMP similar test of H_0, based on the conditional distribution of $\sum \log Y_i$, given $\sum Y_i$ was essentially equivalent to this. In this example we have that $\log \sum Y_i - n^{-1} \sum \log Y_i$ is independent of $\sum Y_i$, the former having a distribution free of λ and the latter following a Gamma$(n\psi, \lambda)$ distribution. This is very similar to the normal case, with partitioning based on $(\bar{Y}, \sum(Y_i - \bar{Y})^2)$.

I also gave details on the maximal invariance of $\bar{Y}^T S^{-1} \bar{Y}$ in the multivariate normal problem with $H_0 : \mu = 0$ but I'll skip the details here.

The road not travelled

I didn't get to spend the promised two lectures on estimating functions, generalized linear models and quasi-likelihood. A good reference (imho) is Firth (1991) "Generalized linear models". This is Chapter 3 of *Statistical Theory and Modelling*, edited by Hinkley, Reid and Snell, published by Chapman and Hall.

Suppose that $Y_i, \quad i = 1, \ldots, n$ are independent, and

$$f(y_i; \beta) = \exp\{\frac{y_i\theta_i - c(\theta_i)}{\phi_i} + h(y_i, \phi_i\}$$

where ϕ_i are considered fixed, and θ_i depends on a $p \times 1$ vector β via:

$$\begin{aligned} EY_i &= \mu_i = c'(\theta_i) \\ g(\mu_i) &= x_i^T \beta. \end{aligned}$$

For arcane reasons g is known as a link function. It is then true that

$$\frac{\partial \ell}{\partial \beta_r} = \sum \frac{y_i - \mu_i}{\phi_i V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_r}$$

where $V(\mu_i) = c''(\theta_i)$ is also a function of β. This means that the maximum likelihood estimation equations can be solved by an iterated weighted least squares algorithm, and general methods for doing this are implemented in GLIM and in Splus, for example. (It is possible to relax the assumption on ϕ_i, see Firth for details.)

Note that even if we only assumed $EY_i = \mu_i(\beta)$, $var(Y_i) = \phi_i V(\mu_i)$, and $g(\mu_i) = x_i^T \beta$, we might still use this as an equation for estimating β, by analogy with least squares. A more general approach to estimating functions, also used in studies of robust point estimation, is to consider as an estimate of β the solution to

$$\sum g_r(y_i; \beta) = 0, \quad , r = 1, \ldots, p$$

or in vector form $g(y; \beta) = 0$. To get a consistent estimate we would require $Eg(Y; \beta) = 0$. (This property is satisfied by score functions of course.) The theory of estimating equations starts here and develops a notion of optimal estimating function, etc..

The estimating function for generalized linear models given above has a further property that $E(\partial g / \partial \beta) = -Var\{g(Y; \beta)\}$; another property in common with score functions. A *quasi-likelihood* function is any scalar function of Y and θ whose first derivative has this quasi-score property.

Other major topics that I left out include: decision theory, many aspects of Bayesian inference, approximations to marginal and conditional likelihood, and theoretical aspects of model choice. This last topic includes testing separate families of hypotheses, discussed in Chapter 5 of CH, as well as Bayesian techniques for choosing, for example, the number of regressors in a multiple regression model.

Summary of the lectures

1. Basics: Sept 14 through Oct 5
    - statistical theory and methods
    - probability models
    - parametric models and likelihood
    - sufficiency, minimal sufficiency, factorization theorem

- likelihood map is minimal sufficient
- exponential family, curved exponential family
- ancillarity, maximal ancillarity
- transformation family

2. Likelihood inference: Oct 12 through Nov 9

    - terminology: score, maximum likelihood estimate, Fisher information, likelihood ratio statistic, additivity of log-likelihood
    - limiting distribution of standardized score statistic, standardized m.l.e., likelihood ratio statistic: scalar θ
    - limiting distribution: vector θ and $\theta = (\psi, \lambda)$
    - profile log-likelihood
    - asymptotic normality of posterior

3. Classical theory of testing: Nov 16 through Dec 7

    - terminology: critical region, test function, size, power, p-values, pure significance tests
    - simple H_0: Neyman-Pearson lemma, uniformly most powerful tests, locally most powerful tests, power function, discrete test statistics, conditioning on ancillary statistics
    - composite H_0: reduction by conditioning (UMP similar)
    - composite H_0: reduction by marginalizing (UMP invariant)

The test on **December 14, 9-12 am, SS2101** will cover material from each of these three units. Textbook references are as follows:

1. CH 2.1, 2.2; TPE 1.3, 1.4, 1.5, 1.6; *lecture notes (some on web)*
2. CH 9.1, 9.2(i)(iii)to p.299, 9.3(i)(ii)(iii)to p.324; *BNC 3.1-3.5,3.7*; lecture notes
3. *CH 3.1, 3.2, 4.1-8, 5.1, 5.2, 5.3*; TSH 3.1, 3.2, 3.3, 3.6, 3.7, 3.9, 4.1, 4.2, 4.3, 6.1, 6.2, 6.3; lecture notes

The references in italics are the ones I think are probably best for this topic, but I will try to ensure that questions can be answered based on the information in any one of the references.