STA 3000F (Fall, 2013)

Comments on Homework 2

1. Non-uniqueness of ancillary statistics. Suppose that $(Y_1, Z_1), \dots (Y_n, Z_n)$ are independent and identically distributed and follow a bivariate normal distribution with $\mathrm{E}(Y_i) = \mathrm{E}(Z_i) = 0, \mathrm{var}(Y_i) = \mathrm{var}(Z_i) = 1$, and $\mathrm{core}(Y_i, Z_i) = \theta$, $-1 < \theta < 1$. This is an example of a curved exponential family; it can be written in exponential family form, but the two canonical parameters are constrained to one dimension.

   (a) Show that $\sum Z_i^2$ and $\sum Y_i^2$ are each ancillary for θ, but that $T = \sum(Y_i^2 + Z_i^2)$ is not ancillary.

   (b) Derive the first two moments of $T/\sqrt{n}$, and plot the variance of this as a function of θ.

   This is example 2.30 in Cox & Hinkley, where the conclusion is that the variance of T is "not too strongly dependent on θ", which is why I asked you to plot it. As you can easily see, it varies between 4 and 8, i.e. between 2 and $2\sqrt{2} \doteq 3.4$ on the standard deviation scale, and since the mean is $2\sqrt{n}$ this probably is fairly small for moderate n.

   C&H require ancillary statistics to be functions of the minimal sufficient statistic, which is $(\Sigma Y_i Z_i, \Sigma Y_i^2 + \Sigma Z_i^2)$, so conclude that there is no ancillary statistic for this model. Many other books require only that ancillary statistics be functions of the original data. From that point of view, we could condition on either ΣY_i^2 or ΣZ_i^2, but I haven't worked it out so I don't know if the inference from these two conditional distributions would be different. If they are, then the principle "condition on an ancillary statistic" doesn't help us here.

2. *Logistic regression.* Suppose Y_i are independent Bernoulli random variables, with density

$$f(y_i) = p_i^{y_i}(1-p_i)^{1-y_i}, \quad y = 0, 1,$$

   and that

$$\log \frac{p_i}{1-p_i} = x_i'\beta,$$

   where x_i and β are vectors of length p.

(a) Write the joint density of $(y_1, \ldots, y_n)$ in exponential family form, and give an expression for the minimal sufficient statistic $S = (S_1, \ldots, S_p)$, say.

(b) Show that the conditional distribution of S_j, given $S_{(-j)}$, depends only on β_j.

This result applies to any exponential family when interest is in any linear combination of the canonical parameters. The importance of the result is that inference for β_j can be based on this conditional distribution, which is now a 1-parameter distribution, so no estimation of the nuisance parameters $\beta_{(-j)}$ is needed. The usual 1-parameter pivotal quantities, r_e, r_u, and r are all available for inference, and the limit theorems based on the conditional likelihood function are all valid, even if the dimension of $\beta_{(-j)}$ increases with the sample size.

3. Suppose that Y_i are independent exponential random variables with $\mathrm{E}(Y_i) = \psi\lambda_i$, and Z_i are independent exponential random variables with $\mathrm{E}(Z_i) = \psi/\lambda_i, \quad i = 1, \ldots, n$.

   (a) Find the maximum likelihood estimates of λ_i and ψ.

   (b) Show that $\hat{\psi}$ is not consistent for ψ as $n \to \infty$.

   This class of problems, with one nuisance parameter per pair, or group, of observations, and then with the number of groups going to ∞, but the group size staying fixed, are called "Neyman-Scott" problems, after Neyman & Scott (1948, *Econometrica*). Their two examples were normal theory examples; in the first (Y_i, Z_i) are independent $N(\mu_i, \sigma^2), i = 1, \ldots, n$. Each pair has a different mean, which will be estimated by the sample mean $(Y_i + Z_i)/2$. The maximum likelihood estimator of σ^2 converges to $\sigma^2/2$ as $n \to \infty$. It is not surprising that we'd get poor estimates of the mean, but it would seem on the surface that we should be able to estimate the variance well since all $2n$ observations have the same variance. The other normal theory problem is groups of size n_j with the same mean, but different values of σ^2; in this case the maximum likelihood estimator of μ is consistent, but an estimator with a smaller asymptotic variance can be constructed. See CH Example 5.8 and Exercise 9.3.

4. *Regression-scale models* Suppose $y = (y_1, \ldots, y_n)^T$ have independent components with density

$$\frac{1}{\sigma} f_0(\frac{y_i - x_i^T \beta}{\sigma}),$$

where $f_0(\cdot)$ is a known density on $\mathbb{R}$. In HW 1 you showed that a is ancillary, where $a_i = (y_i - x_i^T \tilde{\beta})/\tilde{\sigma}$, and the estimators $\tilde{\beta}$ and $\tilde{\sigma}$ are given by

$$\tilde{\beta} = (X^T X)^{-1} X^T y, \quad \tilde{\sigma}^2 = (y - X\tilde{\beta})^T (y - X\tilde{\beta})/(n-p).$$

(In HW1 we called these $\hat{\beta}$, $\hat{\sigma}$, but I'll use this notation below for the maximum likelihood estimators.)

(a) Show that under the transformation $y_i \to c y_i + x_i^T b$, where $c > 0$, and $b = (b_1, \ldots, b_p)$ is a vector in $\mathbb{R}^p$, that we have

$$\tilde{\beta} \to c\tilde{\beta} + b, \quad \tilde{\sigma}^2 \to c^2 \tilde{\sigma}^2.$$

Estimators with this property are called equivariant.

(b) Show that the associated ancillary statistic $\tilde{a} = (y - X\tilde{\beta})/\tilde{\sigma}$ is invariant under the transformation in (a).

(c) Show that the maximum likelihood estimators of β and σ are also equivariant, and the associated set of residuals $\hat{a} = (y - X\hat{\beta})/\hat{\sigma}$ is invariant.

(d) Deduce that the distribution of $\hat{a}$ is free of (β, σ), and thus is also ancillary.

Parts (a) - (c) are pretty straightforward. For part (d), one shows that the distribution of $\hat{a}$ is the same for any two different parameter values (β, σ) and (β', σ'), say, using the transformation property of the model, and since these are arbitrarily chosen the distribution of $\hat{a}$ is ancillary. You can also use results on maximal invariants in, for example, TSH Ch. 4(?).

5. *Orthogonal parameters.* In a model $f(y; \theta)$ with $\theta = (\psi, \lambda)$, the component parameters ψ and λ are orthogonal (with respect to expected Fisher information) if $i_{\psi\lambda}(\theta) = 0$.

(a) Assume y_i follows an exponential distribution with mean $\lambda e^{-\psi x_i}$, where x_i is known. Find conditions on the sequence $\{x_i, i = 1, \dots, n\}$ in order that λ and ψ are orthogonal with respect to expected Fisher information. Find an expression for the constrained maximum likelihood estimate $\hat{\lambda}_\psi$ and show the effect of parameter orthogonality on the form of the estimate.

(b) Suppose that $y_1, \dots, y_n$ are independently normally distributed with mean
$$\mathrm{E}(y_i) = \frac{\alpha x_i}{\beta + x_i},$$
where $x_1, \dots, x_n$ are known constants, and variance σ^2. This is called the Michaelis-Menten model, used in chemical kinetics. Show that (α, σ^2, χ) are mutually orthogonal, where
$$\chi = \sum \frac{\alpha^3 x_i^2}{(\beta + x_i)^3}.$$

The last part was trickier than I thought it would be, because it requires keeping careful attention on what is fixed and what varies when one takes partial derivatives. Here is how I did it: Define
$$\ell^*(\alpha, \chi, \sigma^2) = \ell(\alpha, \beta(\chi, \alpha), \sigma^2).$$
Then
$$\frac{\partial \ell^*(\alpha, \chi, \sigma^2)}{\partial \alpha} = \frac{\partial \ell(\alpha, \beta(\chi, \alpha), \sigma^2)}{\partial \alpha} + \frac{\partial \ell(\alpha, \beta(\chi, \alpha), \sigma^2)}{\partial \beta} \frac{\partial \beta(\chi, \alpha)}{\partial \alpha},$$
and
$$\begin{aligned} \frac{\partial^2 \ell^*}{\partial \chi \partial \alpha} &= \frac{\partial}{\partial \chi}\left\{ \frac{\partial \ell(\alpha, \beta(\chi, \alpha), \sigma^2)}{\partial \alpha} + \frac{\partial \ell(\alpha, \beta(\chi, \alpha), \sigma^2)}{\partial \beta} \frac{\partial \beta(\chi, \alpha)}{\partial \alpha} \right\}, \\ &= \frac{\partial^2 \ell}{\partial \alpha \partial \beta} \frac{\partial \beta}{\partial \chi} + \frac{\partial^2 \ell}{\partial \beta^2} \frac{\partial \beta}{\partial \chi} \frac{\partial \beta}{\partial \alpha} + \frac{\partial \ell}{\partial \beta} \frac{\partial^2 \beta}{\partial \chi \partial \alpha}, \end{aligned}$$
and on taking expected values the final term is 0 so
$$\begin{aligned} i^*_{\alpha\chi}(\alpha, \chi, \sigma^2) &= i_{\alpha\beta}(\alpha, \beta, \sigma^2) \frac{\partial \beta}{\partial \chi} + i_{\beta\beta}(\alpha, \beta, \sigma^2) \frac{\partial \beta}{\partial \chi} \frac{\partial \beta}{\partial \alpha} \\ &= \frac{\partial \beta}{\partial \chi}\left(i_{\beta\beta} + i_{\beta\beta} \frac{\partial \beta}{\partial \alpha} \right) \end{aligned}$$

Then

$$\frac{\partial \beta}{\partial \alpha} = \frac{\partial \chi / \partial \alpha}{\partial \chi / \partial \beta},$$

and the proof that the term in brackets above is zero follows from detailed calculation of the derivatives. The other cross-terms are much easier, because, for example

$$\frac{\partial \ell^*}{\partial \chi \partial \sigma^2} = \frac{\partial}{\partial \chi} \left\{ \frac{\partial \ell}{\partial \sigma^2} \right\} = \frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta} \frac{\partial \beta}{\partial \chi},$$

and since the final factor does not depend on y, the orthogonality of χ and σ^2 follows from that of β and σ^2. In any normal theory linear model, the mean parameters are orthogonal to the parameters in the variance.