ST 762, HOMEWORK 3 EXTRA PROBLEMS, FALL 2009

These problems are from previous years and are for you to work on or not as you choose; they are not to be turned in. You should be familiar with the concepts covered by these problems for the midterm test. Solutions will be posted when the homework problems to be turned in are due.

1. *Contaminated normal distribution.* Let $0 \leq \alpha \leq 1$, and let X be a random variable such that

$$\begin{aligned} X &\sim \mathcal{N}(0,1) \text{ with probability } (1-\alpha) \\ &\sim \mathcal{N}(0,b^2) \text{ with probability } \alpha \end{aligned}$$

for $b \geq 1$. Define $\epsilon = X\{(1-\alpha)+\alpha b^2\}^{-1/2}$.

(a) Find $E(\epsilon)$, $E(\epsilon^2)$, $E(\epsilon^3)$, and an expression for κ (the excess kurtosis), where κ satisfies $\text{var}(\epsilon^2) = 2+\kappa$.

(b) Compute κ for $\alpha = 0.01$ and $\alpha = 0.05$ for $b = 2$, 3, and 4. What happens as α and b increase?

Note: The contaminated normal distribution is often used to study "distributional" robustness – sensitivity of an estimator's properties to outliers. The contaminated normal distribution is still symmetric like the normal distribution, but has "heavier tails." Thus, for different choices of α and b, one may study the effect of deviation from normality caused by outliers; this will manifest itself through the implied value of κ. We'll see how important κ is later in the course.

2. (a) Let ϵ be a standard normal random variable. Verify that $E(\epsilon^3) = 0$, and show that the excess kurtosis $\kappa = 0$, where κ is defined by $\text{var}(\epsilon^2) = 2+\kappa$.

(b) Suppose that Y is a gamma random variable with mean μ and variance $\sigma^2\mu^2$. Define $\epsilon = (y-\mu)/(\sigma\mu)$. Derive $E(\epsilon^3)$ and the excess kurtosis for ϵ.

(c) Same as (b), but now Y is a binomial random variable based on k independent trials, each trial having probability of success p.

3. Suppose Y_j, $j = 1,\ldots,n$, are such that

$$E(Y_j|\boldsymbol{x}_j) = f(\boldsymbol{x}_j,\boldsymbol{\beta}); \quad \text{var}(Y_j|\boldsymbol{x}_j) = \sigma^2 g^2(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j),$$

where $\boldsymbol{\beta}$ is $(p \times 1)$, and the distribution of $Y_j|\boldsymbol{x}_j$ has coefficient of skewness ζ_j and excess kurtosis κ_j. Assume that $\boldsymbol{\theta}$ is *known*.

(a) Derive the quadratic estimating equation for $\boldsymbol{\beta}$ and σ jointly of the form given in (5.6) on page 109 of the notes under these conditions. Express your equation in the form

$$\sum_{j=1}^{n} \Big(\boldsymbol{a}_j\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta})\} + \boldsymbol{c}_j[\{Y_j - f(\boldsymbol{x}_j,\boldsymbol{\beta})\}^2 - \sigma^2 g^2(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{x}_j)] \Big) = \boldsymbol{0}$$

for suitable vectors $\boldsymbol{a}_j$ and $\boldsymbol{c}_j$.

(b) Suppose that we are told that, conditional on $\boldsymbol{x}_j$, Y_j is a binomial random variable based on k_j independent trials, each with probability $p(\boldsymbol{x}_j,\boldsymbol{\beta})$, where

$$p(\boldsymbol{x},\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}_j^T\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_j^T\boldsymbol{\beta})}.$$

By substituting appropriate quantities into the first p rows of your answer to (a), give the form of the quadratic estimating equation for $\boldsymbol{\beta}$ in this particular case.

(c) Suppose that we know $\sigma = 1.0$ and $g^2(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) = f(\boldsymbol{x}_j, \boldsymbol{\beta})$, and we are told that $\zeta_j = 1/g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)$ and $\kappa_j = 1/g^2(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)$. From your estimating equations in (a), find the form of the equation corresponding to $\boldsymbol{\beta}$ in this particular case.

(d) Suppose that $\sigma = 0.5$ is known, and suppose that $g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) = f(\boldsymbol{x}_j, \boldsymbol{\beta})$, and we are told that $\zeta_j = 1$ and $\kappa_j = 1.5$. From your estimating equations in (a), find the form of the equation corresponding to $\boldsymbol{\beta}$ in this particular case.

(e) Give a possible explanation for the way your answers to (b), (c), and (d) turned out.

4. Consider an estimating equation for parameters in the usual mean-variance model given in equation (7.1) on page 155 having the general form in (6.13) on page 128 of the notes; i.e.

$$\sum_{j=1}^{n} \boldsymbol{D}_j^T(\boldsymbol{\alpha})\boldsymbol{V}_j^{-1}(\boldsymbol{\alpha})\{\boldsymbol{s}_j(\boldsymbol{\alpha}) - \boldsymbol{m}_j(\boldsymbol{\alpha})\} = \boldsymbol{0}, \quad \boldsymbol{D}_j(\boldsymbol{\alpha}) = \partial/\partial\boldsymbol{\alpha}\, \boldsymbol{m}_j(\boldsymbol{\alpha}), \tag{1}$$

$E\{\boldsymbol{s}_j(\boldsymbol{\alpha})|\boldsymbol{x}_j\} = \boldsymbol{m}_j(\boldsymbol{\alpha})$. Assuming that $\boldsymbol{s}_j(\boldsymbol{\alpha})$ depends on $\boldsymbol{\alpha}$ only through the expression $\{Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})\}^2$ (where $\boldsymbol{\beta}$ is obviously an element of $\boldsymbol{\alpha}$), derive the updating scheme given in (6.16) on page 129 for solving (1), *showing all your steps and giving the explicit rationale for any terms you disregard as negligible in your argument.*

5. *"The Trick" for general transformations.*

Consider the general class of power transformation estimators for η, $\boldsymbol{\theta}$, discussed in Section 6.5 solving equation (6.28) on the bottom of page 137, where $\boldsymbol{\beta}$ is treated as fixed.

(a) In the particular case $\lambda = 1$ (the identity transformation), it turns out that equation (6.28) may in fact be motivated be considering a certain *double exponential* or *Laplace* distribution instead of the normal.

To see this, assume that we have data for which we assume the mean-variance model

$$E(Y_j|\boldsymbol{x}_j) = f(\boldsymbol{x}_j, \boldsymbol{\beta}); \quad \text{var}(Y_j|\boldsymbol{x}_j) = \sigma^2 g^2(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j),$$

and define $\epsilon_j = \{Y_j - f(\boldsymbol{x}_j, \boldsymbol{\beta})\}/\{\sigma g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)\}$. Suppose that we believe that the ϵ_j are i.i.d. and follow a double exponential distribution with density

$$f(\epsilon) = \frac{1}{2\gamma}\exp(-|\epsilon|/\gamma), \quad \gamma > 0,$$

where γ is chosen so that $E(\epsilon^2) = 1$ (it may be verified that $E(\epsilon) = 0$; you will need to find γ).

Under these conditions, write down the joint conditional density of the Y_j given $\boldsymbol{x}_j$.

(b) Under the conditions in (a), define η such that $E(|\epsilon_j|) = e^{\eta}/\sigma$, and rewrite the joint density in (a) in terms of η. Treating this as a likelihood for η and $\boldsymbol{\theta}$ (holding $\boldsymbol{\beta}$ fixed), find the corresponding estimating equations for η and $\boldsymbol{\theta}$ obtained by differentiating the loglikelihood. Verify that these equations are equivalent to (6.28) with $\lambda = 1$.

(c) Now consider equation (6.28) for general $\lambda > 0$. As discussed at the top of page 139, solution of this equation in η and $\boldsymbol{\theta}$ may be implemented using a "trick" similar to that shown explicitly for PL ($\lambda = 2$) pages 130–132 of the class notes.

- Cleverly define a "likelihood" that, when differentiated with respect to η and $\boldsymbol{\theta}$ and set equal to $\mathbf{0}$, yields equation (6.28) (up to a scalar multiple, at least). Thus, you need to give the definition of your "likelihood" for any $\lambda > 0$ and show the equivalence to (6.28) explicitly. *Hint:* Generalize the forms of the likelihood in (b) ($\lambda = 1$) and the normal likelihood ($\lambda = 2$).
- Use your "likelihood" to define a general "trick" for any $\lambda > 0$ for computing the estimator of $\boldsymbol{\theta}$ using nonlinear regression software. Show the derivation of your "trick" explicitly, and demonstrate that the function $F_j(\boldsymbol{\theta})$ analogous to that on page 132 for PL is given by
$$F_j(\boldsymbol{\theta}) = \frac{|Y_j - f(\boldsymbol{\beta}, \boldsymbol{x}_j)|^{\lambda/2} \dot{g}^{\lambda/2}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)}{g^{\lambda/2}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)}.$$
What is this function in the case $g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j) = f^{\theta}(\boldsymbol{x}_j, \boldsymbol{\beta})$?

Note: The hard part is coming up with the "likelihood." This will not necessarily be a true likelihood, but rather is a "fake" likelihood objective function that is of the right form to yield equation (6.28) upon differentiation with respect to η and $\boldsymbol{\theta}$.

6. *The "trick" for REML estimation.* At the bottom of page 172 of the notes, we stated that it is possible to derive a method for estimating $\boldsymbol{\theta}$ in the model
$$E(Y_j|\boldsymbol{x}_j) = f(\boldsymbol{x}_j, \boldsymbol{\beta}), \quad \text{var}(Y_j|\boldsymbol{x}_j) = \sigma^2 g^2(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{x}_j)$$
that requires only nonlinear regression methods.

Treating $\boldsymbol{\beta}$ as fixed, demonstrate that this is possible by showing that maximizing the objective function in (7.11) on page 172 of the notes in $\boldsymbol{\theta}$ and σ^2 may be carried out by minimizing a sum of squares of the form
$$\sum_{j=1}^{n} \{0 - F_j(\boldsymbol{\theta})\}^2. \tag{2}$$
In particular, find the form of the function $F_j(\boldsymbol{\theta})$.