BIOS 7731 Homework 10 Tim Vigers

1.) Let $Y_i = \beta_0 + \beta_1(X_i - \bar{X}) + \varepsilon_i$ with $i = 1, \dots, n$, X_i known, and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Let $\underline{\underline{X}}$ be the design matrix. The least squares estimator is:

$$\hat{\beta} = (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \underline{\underline{X}}^T \underline{Y} = (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \underline{\underline{X}}^T (\underline{\underline{X}} \vec{\beta} + \vec{\varepsilon})$$

We know (from longitudinal and most statistics textbooks) that the limiting distribution of $\hat{\beta}$ is:

$$\hat{\beta} \overset{\cdot}{\sim} N(\beta, (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} \sigma^2 \underline{\underline{I}})$$

Therefore, by the central limit theorem,

$$(\underline{\underline{X}}^T \underline{\underline{X}})^{1/2} (\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, \sigma^2 \underline{\underline{I}})$$

Basically, in the matrix form $(\underline{\underline{X}}^T \underline{\underline{X}})$ is analagous to n in the simpler CLT. Per BD pg. 101 this estimation method is well defined even if $\vec{\varepsilon}$ is not iid normal.

2. BD 6.1.4) Let Y_i satisfy the model:

$$Y_i = \theta + \varepsilon_i \qquad i = 1, \dots, n$$

where $\varepsilon_i = c e_{i-1} + e_i$ for $0 \le c \le 1$ and e_i are iid with mean 0 and variance σ^2. Also, $e_0 = 0$. We can rewrite Y_i in terms of c and e_i, by writing out the first couple of observations:

$$Y_1 = \theta + c e_0 + e_1 = \theta + e_1$$

$$Y_2 = \theta + c e_1 + e_2 = \theta + c(Y_1 - \theta) + e_2$$

$$Y_3 = \theta + c e_2 + e_3 = \theta + c(Y_2 - \theta - c(Y_1 - \theta)) + e_3$$

Rearranging the terms in Y_3, we start to see a pattern:

$$
\begin{aligned}
Y_3 &= \theta + cY_2 - c\theta - c^2 Y_1 + c^2\theta + e_3 \\
&= \theta(1 - c + c^2) + cY_2 - c^2 Y_1 + e_3 \\
&= \theta\left[\frac{1-(-c)^3}{1+c}\right] + cY_2 - c^2 Y_1 + e_3
\end{aligned}
$$

So, we can rewrite the model:

$$
Y_i = \theta\left[\frac{1-(-c)^i}{1+c}\right] + \sum_{j=1}^{i} (-c)^{j-i} Y_i + e_i
$$

Unfortunately here is where I got stuck on the algebra, and I wasn't able to get things in the right form. However, I do understand the important point here, which is that by using the weights a_j, we now have independent errors in the form of e_i. So, with some (complicated) rearranging, we can treat this model like the Gaussian models we are used to.

b. See BD 2.2.34 for details, but for a model with Y_i independent $N(g(\beta, \underline{z}), \underline{w}\sigma^2)$ where $\underline{w}$ is a weight matrix, the least squares estimate is equivalent to maximizing the likelihood.

c. To evaluate bias we simply take the expected value of the estimators:

$$E[\bar{Y}] = \frac{1}{n} E\left[\sum_{i=1}^{n} Y_i\right] = \frac{n\theta}{n} = \theta.$$

Now using the weighted estimator:

$$E[\hat{\theta}] = E\left[\sum_{i=1}^{n} a_j Y_j\right] = \sum_{i=1}^{n} a_j \, E[Y_i] = \theta \sum_{i=1}^{n} a_j$$

In order for the estimator to be unbiased, the sum of the a_j must be 1. We know this is the case because in weighted least squares the weights must sum to 1.

So,

$$E[\hat{\theta}] = \theta \qquad \square.$$

d. Based on all of the above, we know that $\hat{\theta}$ is the UMVUE for θ (BD theorem 6.1.4). Therefore, the variance of $\hat{\theta}$ must be $\leq$ the variance of any other unbiased estimator, including $\bar{Y}$.

e. Following from d. above, if $\hat{\theta}$ is UMVUE then the only way that $Var(\hat{\theta}) = Var(\bar{Y})$ is if $\hat{\theta} = \bar{Y}$. Given

$$Y_i = \theta + ce_{i-1} + e_i$$

The only way that $\bar{Y}$ can also be UMVUE is the case of a standard model with iid errors. So, the middle term (ce_{i-1}) must be

0 for all i. Therefore $Var(\hat{\theta}) < var(\bar{Y})$ for all $0 < c \leq 1$ and $Var(\hat{\theta}) = var(\bar{Y})$ for $c = 0$.

3. BD 6.2.9) Let $Y_1 \dots Y_n$ be iid

$$Y_i = \mu + \sigma \varepsilon_i$$

where ε_i has known density such that $\rho(x) \equiv -\log f(x)$ is strictly convex. So, we have that

$$\varepsilon_i = \frac{Y_i - \mu}{\sigma}$$

which is a simple scale and location shift. Also, because the negative log-likelihood is strictly convex, this implies that the log-likelihood is strictly concave (and that the likelihood is quasiconcave). Because the log-likelihood is strictly concave, there must exist a unique MLE. Also, by definition of ρ, the standard procedure for finding the MLE (setting the derivative of $\ell(\theta)$ w.r.t. θ equal to 0) is simply:

$$\frac{d}{d\mu} \sum_{i=1}^{n} \log f(x) = \sum_{i=1}^{n} \rho'\left(Y_i - \mu / \sigma_0\right) = 0$$

b. This is easiest to solve backwards, given that we know the result:

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \rho'(\theta_1 Y_i - \theta_2^0) = \frac{1}{\theta_1}$$

This can be rearranged with simple algebra.

$$\sum_{i=1}^{n} Y_i\, p'(\theta_1 Y_i - \theta_2^0) = \frac{n}{\theta_1}$$

$$\Rightarrow \sum_{i=1}^{n}\left(Y_i\, p'(\theta_1 Y_i - \theta_2^0) - \frac{1}{\theta_1}\right) = 0$$

This can be re-written as:

$$\sum_{i=1}^{n}\left(Y_i\, p'(\theta_1 Y_i - \theta_2^0) - \partial/\partial\theta_1 \log(\theta_1)\right) = 0$$

Which is again the standard procedure for finding the MLE.

4. Let $X_{i1}, \ldots, X_{in}$ $i = 1, 2$ be iid beta distributed:

$$\theta_i x^{\theta_i - 1} \quad \text{with } \theta_i > 0 \quad \text{and} \quad 0 < x < 1$$

I find the LRT the easiest, so I'm going to start with part b:

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} \theta_1 X_{i1}^{\theta_1 - 1}\, \theta_2 X_{i2}^{\theta_2 - 1} =$$

Take the log:

$$\ell(\theta_1, \theta_2) = n \log(\theta_1) + (\theta_1 - 1)\sum^{n} \log(X_{i1}) + n\log(\theta_2) + (\theta_2 - 1)\sum^{n} \log(X_{i2})$$

Next find the MLE for θ_1 and θ_2:

$$\frac{\partial}{\partial \theta_1} \ell(\theta_1, \theta_2) = \frac{n}{\theta_1} + \sum^n \log(X_{i1}) \stackrel{set}{=} 0$$

$$\boxed{\hat{\theta}_1 = \frac{-n}{\sum \log(X_{i1})}}$$

$$\frac{\partial}{\partial \theta_2} \ell(\theta_1, \theta_2) = \frac{n}{\theta_2} + \sum^n \log(X_{i2}) \stackrel{set}{=} 0$$

$$\boxed{\hat{\theta}_2 = \frac{-n}{\sum^n \log(X_{i2})}}$$

Now, one form of the LRT statistic is:

$$\lambda(X) = \frac{\sup\{p(X, \theta) : \theta \in \Theta_K\}}{\sup\{p(X, \theta) : \theta \in \Theta_H\}}$$

In other words, the maximum likelihood under the alternative over the maximum likelihood under the null. Since we've calculated MLEs for θ_1 and θ_2, we can simply plug them into the above expression for $H: \theta_1 = \theta_2$ and $K: \theta_1 \neq \theta_2$.

a. The score test follows easily from above. We set $S(\theta_0) = \partial / \partial \vec{\theta} \; \ell(\vec{\theta})$, which was already calculated for the LRT. Then the score test is simply:

$$Z_s = \frac{S(\theta_0)}{\sqrt{I_n(\theta_0)}}$$

c. From our class notes, the Wald test statistic is:

$$W_n \xrightarrow{d} \underset{\sim}{Z}^T \underbrace{I(\theta_0)^{-1}}_{\text{many forms.}} \underset{\sim}{Z} \sim \chi^2_s$$

5. BD 6.3.4) a. I simply can't figure this out. It's mind-boggling that something can converge in probability to 0 but converge in law to χ^2. Looking forward to the answer to this!

b. Based on the CLT we know that

$$P(T_n > k) = \alpha \quad \text{means that}$$

$$P\left(\frac{T_n - E[T_n]}{\sqrt{Var(T_n)}} > \frac{k - E[T_n]}{\sqrt{Var(T_n)}}\right) = \alpha$$

Thus $\frac{T_n - E[T_n]}{\sqrt{Var(T_n)}} \mathrel{\dot\sim} N(0,1)$

The result follows from some simple algebra.