Likelihood Theory and Extensions (BIOS:7110)
Breheny

Assignment 9
Due: Tuesday, November 3

1. *Bernstein-von Mises theorem.* Suppose the prior $\pi(\boldsymbol{\theta})$ is proper and continuous with $\pi(\boldsymbol{\theta}) > 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and that regularity conditions (A)-(D) as defined in the "Likelihood: Consistency" lecture are satisfied. Below, $\hat{\boldsymbol{\theta}}$ is the MLE, $\pi(\boldsymbol{\theta} \mid \mathbf{x})$ is the posterior density of $\boldsymbol{\theta}$, and $\boldsymbol{\delta}$ is an arbitrary (fixed) $d \times 1$ vector.

   (a) Consider the log-likelihood ratio

   $$\log\left(\frac{L(\hat{\boldsymbol{\theta}} + \boldsymbol{\delta}/\sqrt{n})}{L(\hat{\boldsymbol{\theta}})}\right);$$

   what does this quantity converge to as $n \to \infty$? Hint: Taylor series.

   (b) Consider the prior ratio

   $$\frac{\pi(\hat{\boldsymbol{\theta}} + \boldsymbol{\delta}/\sqrt{n})}{\pi(\hat{\boldsymbol{\theta}})};$$

   what does this quantity converge to as $n \to \infty$?

   (c) Prove that

   $$\frac{\pi(\hat{\boldsymbol{\theta}} + \boldsymbol{\delta}/\sqrt{n} \mid \mathbf{x})}{\pi(\hat{\boldsymbol{\theta}} \mid \mathbf{x})} \xrightarrow{\text{P}} \exp\{-\tfrac{1}{2}\boldsymbol{\delta}^\top \boldsymbol{\mathscr{I}}(\boldsymbol{\theta}^*)\boldsymbol{\delta}\}.$$

   Note: this follows fairly directly from (a) and (b), but note that "prove" here means that rigorous justifications of each step are required. If you provided such justifications in (a) and (b), you do not have to repeat the arguments, but I do need to see those justifications somewhere.

   (d) In lecture, we stated the theorem in terms of $\pi_n(\boldsymbol{\delta} \mid \mathbf{x})$, the posterior density of $\boldsymbol{\delta} = \sqrt{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$. Show that the ratio in (c) is equal to $\pi_n(\boldsymbol{\delta} \mid \mathbf{x})/\pi_n(\mathbf{0} \mid \mathbf{x})$. In other words, (c) shows that the posterior density of $\boldsymbol{\delta}$ converges to ϕ up to its normalizing constant, where ϕ is the $\text{N}(\mathbf{0}, \boldsymbol{\mathscr{I}}(\boldsymbol{\theta}^*)^{-1})$ density.

   Note: this does not, by itself, prove the Bernstein-von Mises theorem. If we also knew that the normalizing constants converge, then we would have $\pi_n(\boldsymbol{\delta} \mid \mathbf{x}) \xrightarrow{\text{P}} \phi(\boldsymbol{\delta})$ for every $\boldsymbol{\delta}$, and Scheffé's useful convergence theorem (if a sequence of densities converges pointwise to a density, then it converges in total variation) would finish the proof. However, establishing that the normalizing constants converge is not trivial.

2. *Score test with nuisance parameters.* This problem consists of proving the following theorem, which was stated in our lecture on score, Wald, and likelihood ratio approaches:

   Theorem: If (A)-(D) hold, $\boldsymbol{\mathscr{I}}(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}^*$, and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1^*$, then

   $$\mathbf{u}_1(\hat{\boldsymbol{\theta}}_0)^\top \boldsymbol{\mathscr{V}}_{11}^n(\hat{\boldsymbol{\theta}}_0)\mathbf{u}_1(\hat{\boldsymbol{\theta}}_0) \xrightarrow{\text{d}} \chi^2_r,$$

   where $\boldsymbol{\mathscr{V}}^n = \boldsymbol{\mathscr{I}}_n^{-1}$.

In this theorem, you may make use of the following lemma, which you do not need to prove (its proof is essentially identical to the case for the unrestricted MLE $\hat{\boldsymbol{\theta}}$).

Lemma: If (A)-(D) hold and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1^*$, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_0) - \boldsymbol{\theta}_2^*) = \boldsymbol{\mathcal{I}}_{22}^{-1}(\boldsymbol{\theta}^*)\tfrac{1}{\sqrt{n}}\mathbf{u}_2(\boldsymbol{\theta}^*) + o_p(1),$$

where $\mathbf{u} = (\mathbf{u}_1^\top\ \mathbf{u}_2^\top)^\top$ describes the partitioned score vector, $\boldsymbol{\mathcal{I}}_{22}$ is the lower right block of the Fisher information for a single observation, and here the final $o_p(1)$ term is a $(d-r)\times 1$ vector, all of whose elements are converging in probability to zero.

(a) Take a Taylor series expansion of $\frac{1}{\sqrt{n}}\mathbf{u}(\boldsymbol{\theta})$ about $\boldsymbol{\theta}^*$, evaluated at $\hat{\boldsymbol{\theta}}_0$.

(b) Show that, under the null, $\frac{1}{\sqrt{n}}\mathbf{u}_1(\hat{\boldsymbol{\theta}}_0) = \mathbf{A}\frac{1}{\sqrt{n}}\mathbf{u}(\boldsymbol{\theta}^*) + o_p(1)$, where $\frac{1}{\sqrt{n}}\mathbf{u}_1(\hat{\boldsymbol{\theta}}_0)$ is the first component of the expansion in part (a) and $\mathbf{A}$ is a matrix for you to determine.

(c) Using the result from (b), prove the above theorem.