Likelihood Theory and Extensions (BIOS:7110)
Breheny

Assignment 10
Due: Monday, November 8

1. *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)-(C) hold and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1^*$, then

    $$\tfrac{1}{n}\mathbf{u}_1(\hat{\boldsymbol{\theta}}_0)^\top \mathscr{V}_{11}(\hat{\boldsymbol{\theta}}_0)\mathbf{u}_1(\hat{\boldsymbol{\theta}}_0) \xrightarrow{\text{d}} \chi^2_r.$$

    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)-(C) hold and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1^*$, then

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

    where $\mathbf{u} = (\mathbf{u}_1^\top\ \mathbf{u}_2^\top)^\top$ describes the partitioned score vector, $\mathscr{I}_{22}$ 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.

2. *Projection matrices and the χ^2 distribution.* A matrix $\mathbf{P}$ satisfying $\mathbf{P} = \mathbf{PP}$ is known as an *idempotent matrix*, or *projection matrix*. Below, suppose that $\mathbf{P}$ is a symmetric projection matrix.

    (a) Show that every eigenvalue of $\mathbf{P}$ must be either 1 or 0.

    (b) Show that the rank of $\mathbf{P}$ equals the trace of $\mathbf{P}$.

    (c) Show that if $\mathbf{z} \sim \text{N}(\mathbf{0}, \mathbf{I})$, then $\mathbf{z}^\top\mathbf{P}\mathbf{z} \sim \chi^2_r$, where r is the rank of $\mathbf{P}$.

3. *Score, Wald, and likelihood ratio tests for the rate parameter of the Gamma distribution.* This is a continuation of the problem, "Quadratic approximation for the Gamma distribution". In that problem, you derived the score and information, and simulated a random sample from the Gamma distribution. Now, implement score, Wald, and likelihood ratio approaches for carrying out inference regarding the rate parameter β. We discussed these approaches conceptually in class; this problem ensures that you understand all the specifics well enough to program them. Specifically, turn in a separate `.R` file that provides the following answers (please comment your code well enough that I can find each answer easily):

    (a) 99% Wald confidence interval for β.

    (b) Wald test of $H_0 : \beta = 1$.

(c) 99% Score confidence interval for β.

(d) Score test of $H_0 : \beta = 1$.

(e) 99% likelihood ratio confidence interval for β.

(f) Likelihood ratio test of $H_0 : \beta = 1$.

4. *Simulation study comparing likelihood ratio, Wald, and Score tests.* As in the above problem, simulate from the gamma distribution, only this time with a sample size of 20:

```
x <- rgamma(20, shape=2, rate = 1)
```

In this problem, you will carry out simulations in which you generate $N = 10,000$ samples in this manner.

(a) Carry out Wald, likelihood ratio, and score tests of $H_0 : \beta = 1$. Report the observed type I error using $p < 0.05$ as a rejection threshold; note that this is 1 - the coverage of the 95% confidence interval, but the tests are easier to calculate.

(b) Now, suppose we reparameterize the problem in terms of the scale parameter:

$$f(x|\alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}.$$

Re-derive the score and information matrix in terms of this new parameterization.

(c) Carry out Wald, likelihood ratio, and score tests of $H_0 : \theta = 1$. Report the observed type I error rate; note in this particular case, $\beta^* = \theta^* = 1$. Note that you can re-use much of the code you have already written; for example, $\hat{\alpha}$ is the same in both parameterizations.