Likelihood Theory and Extensions (BIOS:7110)
Breheny

Assignment 7
Due: Monday, October 17

1. *Exponential dispersion.* For the exponential dispersion family,

$$\mathbb{E}(\mathbf{s}) = \nabla\psi(\boldsymbol{\theta})$$
$$\mathbb{V}(\mathbf{s}) = \phi\nabla^2\psi(\boldsymbol{\theta}).$$

In this problem, we will derive these results using two different approaches. Note that the first (using properties of the score) is obviously much easier if you've already derived those properties, but it relies on doing quite a bit of work to prove those theorems in the first place.

Both approaches require the ability to change the order of derivation and integration: this is always justified for exponential families (you do not need to show it). Note that in part (a), you are starting with the derivative inside the integral and moving it out, while in part (b), you are starting with the derivative outside the integral and moving it in.

   (a) Derive the above results using the expectation and variance of the score (is very simple, one or two lines).

   (b) Derive the above results using the fact that $exp\{\psi(\boldsymbol{\theta})/\phi\}$ is the normalizing constant, then take first and second derivatives (this approach is longer).

2. *Completing the square.*

   (a) Show that if $\mathbf{A}$ is a symmetric, positive definite matrix, then any expression of the form

$$\mathbf{x}^\top\mathbf{A}\mathbf{x} + \mathbf{b}^\top\mathbf{x} + C$$

   can be rewritten as

$$(\mathbf{x}-\mathbf{m})^\top\mathbf{A}(\mathbf{x}-\mathbf{m}) + D;$$

   provide expressions for $\mathbf{m}$ and D in terms of $\mathbf{A}$, $\mathbf{b}$, and C.

   (b) Using the above result, show that for any likelihood that allows a second-order Taylor series expansion, we can approximate $\ell(\boldsymbol{\theta}|\mathbf{x})$ with the log-likelihood from a multivariate normal distribution:

$$\mathbf{y} \,\dot{\sim}\, \mathrm{N}(\boldsymbol{\theta}, \boldsymbol{\mathcal{I}}^{-1}),$$

   where $\mathbf{y} = \tilde{\boldsymbol{\theta}} - \mathbf{H}^{-1}\mathbf{u} = \tilde{\boldsymbol{\theta}} + \boldsymbol{\mathcal{I}}^{-1}\mathbf{u}$ is the "pseudoresponse" and $\mathbf{u}$ and $\boldsymbol{\mathcal{I}}$ are the score and information evaluated at $\tilde{\boldsymbol{\theta}}$, the point about which we are taking the Taylor series approximation. Note here that $\mathbf{u}$ is the random variable; in reality, the point $\tilde{\boldsymbol{\theta}}$ that we take the approximation about is typically random as well, but this is not accounted for here. In particular, if $\tilde{\boldsymbol{\theta}}$ is the MLE, then $\mathbf{y} = \hat{\boldsymbol{\theta}}$ and we have the typical Wald approximation.

3. *Observed and Fisher information in the exponential family.* Suppose that $x_1, \ldots, x_n \overset{\text{iid}}{\sim} F$, where F is a distribution in the exponential family. Show that for all $\boldsymbol{\theta}$,

$$n\boldsymbol{\mathscr{I}}(\boldsymbol{\theta}) = \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ is the natural parameter.

4. *Quadratic approximation for the Gamma distribution.* The pdf for the Gamma distribution with shape parameter α and rate parameter β is given by

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

Generate a random sample of size 50 from this distribution using the following code:

```
set.seed(1)
x <- rgamma(50, shape=2, rate = 1)
```

To do this problem, please take note of the following information. The derivative of the log of the gamma function is known as the *digamma function*, usually denoted $\psi_0(\alpha)$:

$$\psi_0(\alpha) = \tfrac{d}{d\alpha} \log \Gamma(\alpha).$$

The second derivative of the log of the gamma function is known as the *trigamma function*, and usually denoted $\psi_1(\alpha)$:

$$\psi_1(\alpha) = \tfrac{d^2}{d\alpha^2} \log \Gamma(\alpha).$$

These functions are available in R and called `digamma()` and `trigamma()`, respectively.

(a) Derive the score vector for $\boldsymbol{\theta} = (\alpha, \beta)$.

(b) Solve for the MLE. This is only available in partially-closed form. To find the MLE, solve for $\widehat{\beta}$ in terms of $\hat{\alpha}$ and use this to derive a one-dimensional score function for α, then use the `uniroot()` function to obtain a numerical answer for $\hat{\alpha}$.

(c) Derive the information matrix.

(d) Plot the two-dimensional 95% confidence region based on the quadratic/normal approximation at the MLE. On the plot, label the true value of (α, β). You can do this however you'd like, but my suggestion is to use the `ellipse` package.

(e) Overlay the gamma densities evaluated at three values of $\boldsymbol{\theta}$: the true value $\boldsymbol{\theta}^*$, the MLE $\hat{\boldsymbol{\theta}}$, and $\tilde{\boldsymbol{\theta}} = (3, 1)$. Note that $\tilde{\boldsymbol{\theta}}$ is closer to the MLE than $\boldsymbol{\theta}^*$ is, in the sense that $\|\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\| < \|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}\|$, and yet (if you did the earlier parts correctly) $\boldsymbol{\theta}^*$ is inside the confidence region while $\tilde{\boldsymbol{\theta}}$ is not. What is going on here? Comment briefly on this "paradox".