

Likelihood Theory and Extensions (BIOS:7110)
Breheny

Assignment 7

Due: Tuesday, October 20

1. *Completing the square.* In this problem, we work out a general recipe for approximating any smooth likelihood with positive observed information by a Gaussian likelihood.

- (a) Suppose A is a symmetric, positive definite matrix. Show that any expression of the form

$$-\frac{1}{2}\mathbf{y}^\top \mathbf{A}\mathbf{y} + \mathbf{b}^\top \mathbf{y} + C$$

can be written as

$$-\frac{1}{2}(\mathbf{y} - \mathbf{m})^\top \mathbf{A}(\mathbf{y} - \mathbf{m}) + D.$$

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

- (b) Let $\ell(\boldsymbol{\theta} \mid \mathbf{y})$ be twice continuously differentiable, and consider its second-order Taylor approximation about a fixed point $\tilde{\boldsymbol{\theta}}$. Let

$$\mathbf{u} = \nabla \ell(\tilde{\boldsymbol{\theta}} \mid \mathbf{y})$$

$$\mathbf{H} = \nabla^2 \ell(\tilde{\boldsymbol{\theta}} \mid \mathbf{y})$$

$$\mathcal{I} = -\mathbf{H}$$

and suppose $\mathcal{I}$ is positive definite. Using part (a), show that the quadratic Taylor approximation to $\ell(\boldsymbol{\theta} \mid \mathbf{y})$, viewed as a function of $\boldsymbol{\theta}$, is equivalent to the log likelihood arising from

$$\mathbf{z} \mid \boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \mathcal{I}^{-1}),$$

where $\mathbf{z} = \tilde{\boldsymbol{\theta}} - \mathbf{H}^{-1}\mathbf{u} = \tilde{\boldsymbol{\theta}} + \mathcal{I}^{-1}\mathbf{u}$ is the “pseudoresponse”.

- (c) Consider independent Bernoulli observations $Y_1, \dots, Y_n$, with

$$P(Y_i = 1) = \pi_i, \quad \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \eta_i.$$

Let $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^\top$, and suppose we expand the log likelihood about a fixed vector $\tilde{\boldsymbol{\eta}}$. Find the Gaussian pseudoresponse $\mathbf{z}$ (defined in part b), and provide its approximate multivariate normal distribution. Note that you have already derived the likelihood and taken the relevant derivatives in a much earlier problem, “Logistic regression derivatives”; there is no need to start from scratch here.

2. *Observed and Fisher information in the exponential family.* Suppose that $x \sim F$, where F is a distribution in the exponential family. Show that for all values of x and $\boldsymbol{\theta}$,

$$\mathcal{J}(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})$$

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

3. *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 $\psi_0(\alpha)$:

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

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

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

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

- (a) Derive the score vector for $\theta = (\alpha, \beta)$.
 - (b) Solve for the MLE. This is only available in partially-closed form. To find the MLE, solve for $\hat{\beta}$ 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 θ : the true value θ^* , the MLE $\hat{\theta}$, and $\tilde{\theta} = (3, 1)$. Note that $\tilde{\theta}$ is closer to the MLE than θ^* is, in the sense that $\|\tilde{\theta} - \hat{\theta}\| < \|\theta^* - \hat{\theta}\|$, and yet (if you did the earlier parts correctly) θ^* is inside the confidence region while $\tilde{\theta}$ is not. What is going on here? Comment briefly on this “paradox”.
4. *Neyman-Scott problem.* Suppose $Y_{i1}, Y_{i2} \sim N(\mu_i, \sigma^2)$, for $i = 1, \dots, n$, with all random variables mutually independent.
- (a) Find the maximum likelihood estimate of σ^2 .
 - (b) Show that the MLE of σ^2 is not consistent.
 - (c) Why does our theorem about the consistency of maximum likelihood estimates not hold in this situation? Don't just say “not iid”, because MLEs are consistent in many non-iid scenarios.
 - (d) Find a consistent estimator of σ^2 .