

Empirical Maximum Likelihood: Lecture VIII

Charles B. Moss

5 September 2014

- 1 Empirical Maximum Likelihood
 - Newton-Raphson
 - Example

Empirical Maximum Likelihood

- To demonstrate the estimation of the likelihood functions using maximum likelihood, we formulate the estimation problem for the gamma distribution for the same dataset including a trend line in the mean.
- The basic gamma distribution function is

$$L = \prod_{i=1}^N \frac{1}{\Gamma(\alpha) \beta^\alpha} x_i^{\alpha-1} \exp\left(-\frac{x_i}{\beta}\right)$$
$$\Rightarrow \ln(L) = \sum_{i=1}^N \left[-\ln(\Gamma(\alpha)) - \alpha \ln(\beta) + (\alpha - 1) x_i - \frac{x_i}{\beta} \right] \quad (1)$$

- Next, we add the possibility of a trend line

$$\max_{\alpha_0, \alpha_1, \beta} \ln(L) = \sum_{t=1}^T \left[-\ln(\Gamma(\alpha_0 + \alpha_1 t)) - (\alpha_0 + \alpha_1 t) \ln(\beta) + (\alpha_0 + \alpha_1 t - 1) x_t - \frac{x_t}{\beta} \right] \quad (2)$$

- Given the implicit nonlinearity involved, we will solve for the optimum using the nonlinear optimization techniques.
- Most students have been introduced to the first-order conditions for optimality.
- For our purposes, we will redevelop these conditions within the framework of a second-order Taylor series expansion of a function.

- Taking the second-order Taylor series expansion of a function ($f(x)$) of a vector of variables (x)

$$f(x) = f(x_0) + \nabla_x f(x)(x - x_0) + \frac{1}{2}(x - x_0)' \nabla_{xx}^2 f(x)(x - x_0) + \epsilon(x - x_0) \quad (3)$$

where x_0 is the point of approximation, the gradient vector is defined as

$$\begin{aligned} \nabla_x f(x_0) &= \left[\left. \frac{\partial f(x)}{\partial x_1} \right|_{x=x_0} \quad \dots \quad \left. \frac{\partial f(x)}{\partial x_n} \right|_{x=x_0} \right] \\ \Rightarrow \nabla_x f(x_0)(x - x_0) &= \sum_{i=1}^N \left. \frac{\partial f(x)}{\partial x_i} \right|_{x=x_0} (x_i - x_{0,i}) \end{aligned} \quad (4)$$

- The Hessian matrix is defined as

$$\nabla_{xx}^2 f(x_0) = \begin{bmatrix} \left. \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} \right|_{x=x_0} & \cdots & \left. \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \right|_{x=x_0} \\ \vdots & \ddots & \vdots \\ \left. \frac{\partial^2 f(x)}{\partial x_n \partial x_1} \right|_{x=x_0} & \cdots & \left. \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \right|_{x=x_0} \end{bmatrix} \quad (5)$$

$$\begin{aligned} &\Rightarrow (x - x_0)' \nabla_{xx}^2 f(x_0) (x - x_0) \\ &= \sum_{j=1}^n \sum_{i=1}^n \left. \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right|_{x=x_0} (x_i - x_{0,i}) (x_j - x_{0,j}) \end{aligned}$$

- Given the Taylor series expansion x_0 defines a maximum if

$$f(x) - f(x_0) = \nabla_x f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)' \nabla_{xx}^2 f(x_0)(x - x_0) < 0 \quad (6)$$

- If we restrict our attention to functions whose second derivatives are continuous close to x_0 , this equation implies two sufficient conditions.
 - First, the vector of first derivatives must vanish

$$\|\nabla_x f(x_0)\|_p \rightarrow 0, \|z\|_p \equiv \left| \sum_{i=1}^n z_i^p \right|^{\frac{1}{p}} \quad (7)$$

- Second, the Hessian matrix is negative definite, or

$$x' \nabla_{xx}^2 f(x_0) x < 0 \text{ for all } x \quad (8)$$

Newton-Raphson

- Newton-Raphson is simply a procedure which efficiently finds the zeros of the gradient vector (a vector valued function)

$$\nabla_x f(x) = 0 \Rightarrow \nabla_x f(x_0) - \nabla_{xx}^2 f(x_0)(x - x_0) = 0 \quad (9)$$

- Solving for x based on x_0 we have

$$x = x_0 - [\nabla_{xx}^2 f(x_0)]^{-1} \nabla_x f(x_0) \quad (10)$$

Example

- While the gamma function with a time trend is amenable to solution using these numerical techniques, for demonstration purposes we return to the normal distribution function with a time trend specified as

$$\max_{\alpha_0, \alpha_1, \sigma^2} \ln(L) = -\frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T (\alpha_0 + \alpha_1 t - y_t) \quad (11)$$

- The gradient of the likelihood function becomes

$$\nabla_x \ln(L) = \begin{bmatrix} -\sum_{t=1}^T \frac{\alpha_0 + \alpha_1 t - y_t}{\sigma^2} \\ -\sum_{t=1}^T \frac{t(\alpha_0 + \alpha_1 t - y_t)}{\sigma^2} \\ \sum_{t=1}^T \frac{(\alpha_0 + \alpha_1 t - y_t) - \sigma^2}{2(\sigma^2)^2} \end{bmatrix} \quad (12)$$

- The Hessian matrix becomes

$$\begin{bmatrix} -\sum_{t=1}^T \frac{1}{\sigma^2} & -\sum_{t=1}^T \frac{t}{\sigma^2} & \sum_{t=1}^T \frac{\alpha_0 + \alpha_1 t - y_t}{\sigma^2} \\ -\sum_{t=1}^T \frac{t}{\sigma^2} & -\sum_{t=1}^T \frac{t^2}{\sigma^2} & \sum_{t=1}^T \frac{t(\alpha_0 + \alpha_1 t - y_t)}{(\sigma^2)^2} \\ \sum_{t=1}^T \frac{\alpha_0 + \alpha_1 t - y_t}{\sigma^2} & \sum_{t=1}^T \frac{t(\alpha_0 + \alpha_1 t - y_t)}{(\sigma^2)^2} & -\left(\sum_{t=1}^T \frac{2(\alpha_0 + \alpha_1 t - y_t)^2 - \sigma^2}{2(\sigma^2)^3} \right) \end{bmatrix} \quad (13)$$