Tilastollisen päättelyn jatkokurssi, sl 2010, Exercise 4, week 40

1. Assume that the observations $Y_1, ..., Y_n$ (interpreted as random) are independent with $Y_i \sim P(\mu_i(\theta))$ (Poisson distribution) and the expectation $\mu_i(\theta)$

$$\mu_i(\theta) = \exp\{\alpha + \beta x_i\}$$

depends on the fixed (or nonrandom) variable x_i and the parameter $\theta = (\alpha, \beta) \in \mathbb{R}^2$. Derive the related statistical model and the likelihood function, score function, observed information matrix, and Fisher information matrix of the parameter θ.

Note: The probability mass function of the random variable $Y \sim P(\mu)$ is $f(y;\mu) = \frac{1}{y!}\mu^y e^{-\mu}$, $y = 0, 1, ...,$ $\mu > 0$. Moreover, $(Y) = \mu$ and $\mathsf{Var}(Y) = \mu$.

2. (Continuation for exercise 3.4) (i) Derive the score function and observed information matrix of the parameter $\theta = (\phi, \sigma^2)$.
(ii) Show that the parameters ϕ and σ^2 are orthogonal.

3. Let $\boldsymbol{W} = (W_1, ..., W_n)$ be the observed data (interpreted as random). Partition $W_i = (Y_i, X_i)$ $(i = 1, ..., n)$ and consider the nonlinear regression model

$$Y_i = g(Z_i; \beta) + \varepsilon_i, \quad \beta \in B \subseteq \mathbb{R}^p, \quad \sigma^2 > 0,$$

where Z_i $(p \times 1)$ is a subvector of $(\boldsymbol{W}_{i-1}, X_i)$, $\varepsilon_1, ..., \varepsilon_n \sim \mathsf{N}(0, \sigma^2)$ $\underline{\perp\!\!\!\perp}$ and $(\boldsymbol{W}_{i-1}, X_i)$ $\underline{\perp\!\!\!\perp}$ ε_i for all $i = 1, ..., n$. Moreover, the function $\beta \mapsto g(z; \beta)$ has continuous partial derivatives up to the second order for all $z \in \mathbb{R}^k$ (in the nonlinear case the dimensions of the vector of explanatory variables Z_i and the parameter vector β need not be identical).

(i) Derive the conditional likelihood function $L^{(c)}(\theta; \boldsymbol{w})$ of the parameter $\theta = (\beta, \sigma^2)$.

(ii) Show that $\hat{\beta}$, the ML estimate of β, can be obtained by minimizing the nonlinear sum of squares function

$$S(\beta) = \sum_{i=1}^{n} (y_i - g(z_i; \beta))^2$$

and that $\hat{\sigma}^2$, the ML estimate of σ^2, is obtained by the formula $\hat{\sigma}^2 = n^{-1} S(\hat{\beta})$.

Hint: In part (i) you can use the result of exercise 3.3. Note also that the general assumptions made of the conditional model and marginal model are assumed to hold and that the linear special case is considered on p. 21-22 of the lecture notes. The conditional model and the marginal model are considered on p. 20-21 of the lecture notes. An English reference close to the style of the lecture notes is given in Chapters 5.3 and 5.8 of D. Hendry: Dynamic Econometrics (Oxford Univeristy Press, 1995).

In part (ii) assume that the function $S(\beta)$ is minimized at the point $\beta = b$, consider the conditional log–likelihood function $l^{(c)}(\theta; \boldsymbol{w}) = \log L^{(c)}(\theta; \boldsymbol{w})$, and derive a "suitable" lower bound for the difference $l^{(c)}(b, \sigma^2; \boldsymbol{w}) - l^{(c)}(\beta, \sigma^2; \boldsymbol{w})$. Finally, apply the known inequality $\log x \le x - 1$, $x > 0$, where equality holds if and only if $x = 1$, to the aforementioned lower bound.

4. (Continuation for the preceding one) (i) Derive the score function and observed information matrix of the parameter θ.

(ii) Derive an expression for the Fisher information matrix of the parameter θ (assuming finiteness of the needed moments) and conclude that the parameters β and σ^2 are orthogonal that is, the Fisher information matrix between them is block diagonal.