Problem Set 8: Maximum Likelihood Estimation

Review the Concepts and Proofs

1. What is a likelihood function? What is a conditional maximum likelihood estimator?
2. What are the main assumptions needed to guarantee consistency and asymptotic normality of the CMLE? Discuss possible reasons why these assumptions might be violated.
3. What is the Fisher information matrix?
4. What is the unconditional information matrix equality (UIME)? How does it simplify the asymptotic distribution of the CMLE?
5. Prove the unconditional information matrix equality (UIME).
6. Explain the three testing principles (Wald, LR, LM) graphically.
7. What is the advantage of the LM test when estimation under the alternative is complicated?
8. What is a binary choice model?
9. Compare the advantages and disadvantages of the LPM with those of the logit/probit models.
10. State the log likelihood function of the logit model.
11. Derive score and Hessian of the logit model.
12. Why do the partial effects of the logit/probit model change with $\mathbf{x}\boldsymbol{\theta}$? If you want to state only one partial effect that is representative for the sample, what could you do? Explain.

Exercises

1. Consider a random sample of size N from the Bernoulli distribution

$$f(y) = \theta^y (1-\theta)^{1-y}, \qquad 0 \leq \theta \leq 1, y \in \{0, 1\}.$$

Recall that $\mathrm{E}(y) = \theta$ and $\mathrm{Var}(y) = \theta(1-\theta)$.

(a) Write down the log likelihood function.

(b) Find the ML estimator of θ.

(c) Is the ML estimator biased?

(d) Find the variance of the ML estimator.

(e) Use a CLT to find the asymptotic distribution of the ML estimator.

2. Consider a random sample of size N from the Poisson distribution

$$f(y) = e^{-\lambda} \lambda^y / y!, \qquad \lambda > 0, y \in \{0, 1, 2, \ldots\}.$$

Recall that $\mathrm{E}(y) = \lambda$ and $\mathrm{Var}(y) = \lambda$.

(a) Write down the log likelihood function for observation i and for the full sample.

(b) Derive the score with respect to λ for observation i. Show that it has mean zero and variance $1/\lambda$ and use a CLT to find the asymptotic distribution for the score of the sample.

(c) State the FOC and find the ML estimator of λ.

(d) Is the ML estimator biased?

(e) Find the variance of the ML estimator.

(f) Find the Hessian with respect to λ.

(g) Use a CLT to find the asymptotic distribution of the ML estimator.

3. Consider the poisson regression model used for the nonnegative count variable $y_i \in \{0, 1, 2, \ldots\}$. Assume the conditional mean is $\mathrm{E}(y_i|\mathbf{x}_i) = \mu(\mathbf{x}_i) = \exp(\mathbf{x}_i\boldsymbol{\theta})$ and the distribution of y_i conditional on $\mathbf{x}_i$ is the poisson distribution,

$$f(y|\mathbf{x}_i, \boldsymbol{\theta}) = \exp[-\mu(\mathbf{x}_i)][\mu(\mathbf{x}_i)]^y / y!, \qquad y = 0, 1, 2, \ldots.$$

Note that this distribution has the property that $\mathrm{Var}(y_i|\mathbf{x}_i) = \mathrm{E}(y_i|\mathbf{x}_i)$. (Hint: the solutions to the following questions can all be found in chapter 13 of the textbook.)

(a) Write down the conditional log likelihood function for observation i and for the full sample.

(b) Derive the score with respect to $\boldsymbol{\theta}$ for observation i. Show directly that the score has conditional mean zero.

(c) State the FOC's.

(d) Find the Hessian with respect to $\boldsymbol{\theta}$.

(e) Show directly that the conditional information matrix equality holds.

(f) Find the asymptotic standard errors of $\hat{\boldsymbol{\theta}}$.

4. Consider the linear model $\mathrm{E}(y|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}_o$. Assume the error term $u = y - \mathbf{x}\boldsymbol{\beta}_o$ is, conditional on $\mathbf{x}$, normally distributed with mean zero and variance σ_o^2. Denote the sample size by N.

   (a) Write down the conditional log likelihood function for observation i and for the full sample.

   (b) Derive the score with respect to $\boldsymbol{\theta} = (\boldsymbol{\beta}_o', \sigma^2)'$ for observation i. Show directly that the score has conditional mean zero.

   (c) State the FOC's and solve them for the CMLE's. (Hint: it is sufficient to solve the CMLE of σ_o^2 such that $\hat{\sigma}^2$ is a function of $\hat{\boldsymbol{\beta}}$.)

   (d) Are the CMLE's unbiased?

   (e) Find the Hessian with respect to $\boldsymbol{\theta}$.

   (f) Show directly that the conditional information matrix equality holds.

   (g) Find $\mathbf{A}(x_i, \boldsymbol{\theta})$. Why is advantageous to base the estimator of the variance matric for the CMLE on $\mathbf{A}(x_i, \boldsymbol{\theta})$ instead of using the Hessian directly?

   (h) Find the asymptotic standard errors of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$.

   (i) While an analytical solution is possible, you apply a generalized Gauss-Newton procedure to numerically find the CMLEs for the intercept-only model $\mathbf{x}_i = 1$. Given the sample information $\overline{y} = 1.5$ and $\overline{y^2} = 6.25$, find the CMLEs starting from initial guesses $\beta^{\{0\}} = 0$ and $(\sigma^2)^{\{0\}} = 1$. Report the results for each iteration. How many iterations do you need?

5. Consider again the linear model $\mathrm{E}(y|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}_o$, where the error term $u = y - \mathbf{x}\boldsymbol{\beta}_o$ is, conditional on $\mathbf{x}$, normally distributed with mean zero and variance σ_o^2. Split the regressors into two distinct sets $\mathbf{x}_1$ and $\mathbf{x}_2$ with parameter vectors $\boldsymbol{\beta}_{1o}$ and $\boldsymbol{\beta}_{2o}$ such that $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ and $\boldsymbol{\beta}_o = (\boldsymbol{\beta}_{1o}', \boldsymbol{\beta}_{2o}')'$. Using a sample of size N, you want to test the null hypothesis $H_0 : \boldsymbol{\beta}_{2o} = \mathbf{0}$.

   (a) Show that the likelihood ratio statistic reduces to

   $$LR = N \ \log\left(\tilde{\sigma}^2/\hat{\sigma}^2\right),$$

   where $\tilde{\sigma}^2$ is estimated under H_0 and $\hat{\sigma}^2$ is estimated under H_1. (Hint: substitute the parameter estimators into the log likelihood function both under H_0 and H_1, and simplify each. Then compute the likelihood ratio statistic.)

(b) Show that the LM statistic can be written as

$$LM = \left(\sum_{i=1}^{N} \mathbf{x}_i' \tilde{u}_i \right)' \left(\sum_{i=1}^{N} \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{x}_i' \tilde{u}_i \right) \Big/ \tilde{\sigma}^2,$$

where $\tilde{\sigma}^2$ and $\tilde{u}_i$ are estimated under H_0.

(c) Show that the LM statistic is numerically identical to N times the uncentered R-squared of the auxiliary regression

$$\tilde{u}_i = \mathbf{x}_i \boldsymbol{\gamma} + v_i.$$

Note that the uncentered R-squared is constructed similar to the centered R-squared with the only difference that the original observations instead of the demeaned observations are used. (Hint: write the above LM statistic in matrix form and compare with the uncentered R-squared of the auxiliary regression.)

Empirical Exercises

1. You want to compare the finite-sample properties of Wald, LR and LM tests applied to the restriction analyzed in question 5 in the Exercises section above. To this end, you want to set up a simulation study. The model you simulate and estimate under H_1 is
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$
where $x_1 \sim \text{Normal}(0,1)$, $x_2 \sim \text{Normal}(0,1)$, $u \sim \text{Normal}(0,\sigma^2)$, $\beta_0 = 0$ and $\beta_1 = 1$.

   (a) Write three Matlab functions that compute the Wald, LR and LM test statistics.

   (b) Write a Matlab program that simulates N observations of y for different choices $N = 20, 50, 100$ and $\beta_2 = 0, 0.1, 0.2, \dots, 5$ and invokes the three test functions.

   (c) Extend the program such that the simulation is replicated 1,000 times. For each replication, store the three test statistics. At the end of the program, evaluate the rejection frequencies.

   (d) Extend the program by a graph that shows the power curve: the rejection frequency of each test on the vertical axis against $\beta_2 = 0, 0.1, 0.2, \dots, 5$ on the horizontal axis.

2. You want to analyze the determinants of women's labor force participation. To this end, open the mroz.dta dataset in Stata.

   (a) Re-estimate the baseline specification presented in the textbook and in class by OLS, logit and probit. Compute the APEs and PEAs for the continuous variables.

   (b) Compute the partial effect of age evaluated at the first, second, and third quartile of the distribution of the other regressors.

   (c) Compute the partial effect of experience both analytically (as a general function of $\mathbf{x}$ and $\boldsymbol{\theta}$) and empirically (for the dataset at hand). Take into account that both $exper$ and $expersq$ are included as regressors!

   (d) Add father's years of education, $fatheduc$, and mother's years of education, $motheduc$, as explanatory variables. Test for joint significance of these two regressors using (a) a Wald test and (b) a likelihood ratio test.

   (e) Split the quantitative variable $kidslt6$ into dummy variables $kid0 = 1$ if no young kids and zero else, $kid1 = 1$ if one young kid and zero else, and so on. Which specification is more restrictive? Test the more against the less restrictive specification using (a) a Wald test and (b) a likelihood ratio test.