Advanced Econometrics

Exam Summer Term 2009, September 29, 2009

Prof. Dr. Nikolaus Hautsch
Prof. Dr. Melanie Schienle
Institute for Statistics and Econometrics
Humboldt-Universität zu Berlin

You have to answer 2 out of 3 problems within 90 minutes (plus 10 minutes ”reading time”). If you answer all questions, only the first 2 problems will be taken into account.

You may answer in English or in German. But please stick to one language.

You find necessary underlying results and formulas in the appendix.

Do your best to write legibly. Exams or parts of exams which cannot be read with reasonable effort will not be graded.

Good luck!

Problem 1: Bayesian Inference

a) Explain the major principle of Bayesian inference and discuss (briefly!) the main differences to the frequentist approach. 3 P

b) Compute the kernel and the integrating constant of a $N(\mu, \sigma^2)$ distribution. 2 P

c) Show that the gamma distribution $G(\alpha, \beta)$ is a conjugate prior distribution for the Poisson distribution.

   Hint: The corresponding p.d.f.'s are given in the Appendix. 4 P

d) Write the mean of the posterior distribution in (c) as a weighted average of the mean of the prior distribution and the corresponding maximum likelihood estimator. 4 P

e) Assume a linear regression model $y = X\beta + \varepsilon$ with y being an $(n \times 1)$ vector, X an $(n \times k)$ regressor matrix, β a $(k \times 1)$ parameter vector and ε being an $(n \times 1)$ vector of error terms with $\varepsilon \sim N(0, \sigma^2)$. Assume flat priors for β and $\ln \sigma$ of the form

$$\begin{aligned} \pi(\beta) &\propto c_1, \quad c_1 > 0 \\ \pi(\ln \sigma) &\propto c_2, \quad c_2 > 0. \end{aligned}$$

   Show that

$$\begin{aligned} \pi(\beta|\sigma^2, y) &= N_k(\beta|\hat{\beta}, \sigma^2(X'X)^{-1}), \\ \pi(\sigma^2|y) &= IG(\sigma^2|(n-k)/2, S^2/2), \\ \hat{\beta} &= (X'X)^{-1}X'y, \\ \hat{S}^2 &= (y - X\hat{\beta})'(y - X\hat{\beta}). \end{aligned}$$

5 P

f) Explain briefly the main idea of the Gibbs sampler. 2 P

Problem 2: Nonparametric Estimation

Obtain an estimate $\hat{F}(x)$ of the cumulative distribution function $F(x)$ via integrating the standard (Rosenblatt-Parzen) kernel density estimator of the density $f(x)$. Thus

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(t)\,\mathrm{d}t \qquad \text{for any } x \in \mathbb{R} \, .$$

Assume that f' is continuous at x and the corresponding standard kernel assumptions hold, i.e. $\int K(z)\,\mathrm{d}z = 1$, K symmetric and $\int |z|^2 K(z)\,\mathrm{d}z < \infty$, $\int zK(z)G(z)\,\mathrm{d}z < \infty$ where $G(z) = \int_{-\infty}^{z} K(x)\,\mathrm{d}x$.

a) Derive the asymptotic bias expression

$$\mathbb{E}(\hat{F}(x)) \quad = \quad F(x) + h^2 \frac{f'(x)}{2} \int z^2 K(z)\,\mathrm{d}z + o(h^2) \, .$$

Please be clear and explicit about every step in your calculation.

Hint: Be careful with the boundaries when you interchange the order of integrations. 4 P

b) Assume that you know the asymptotic variance expression

$$\mathbb{V}(\hat{F}(x)) \quad = \quad \frac{1}{n} F(x)(1 - F(x)) - 2\frac{h}{n} f(x) \int zK(z)G(z)\,\mathrm{d}z + o\left(\frac{h}{n}\right) \, .$$

Which conditions on the bandwidth are necessary for $\hat{F}(x)$ being a consistent estimator of $F(x)$? 2 P

c) Use results in a) and b) to show consistency of $\hat{F}$ under the bandwidth assumptions in b). Please provide detailed reasoning. 4 P

d) What is the optimal order h_{opt} for the choice of bandwidth according to MSE or IMSE and what is the resulting asymptotic rate of MSE(h_{opt})? In contrast to density estimation, we can obtain the optimal asymptotic rate of MSE by a wider range of admissible bandwidths including h_{opt}. Specify the admissible interval. If we use the h_{opt} of $\hat{F}$ also for estimating the density function via $\hat{f}$, what can we expect for $\hat{f}_{h_{opt}}$? 5 P

e) Explain why implementing the optimal theoretical choice of bandwidth according to MSE causes problems in practice. Explain the idea of **two** different approaches for the choice of bandwidth in practice. 5 P

Problem 3: Non- and Semiparametric Estimation

Consider a nonparametric regression problem, where you have iid observations $(Y_i, X_i)_{i=1}^n \in \mathbb{R} \times \mathbb{R}$ and you want to estimate m in

$$Y = m(X) + \epsilon \quad \text{with } \mathbb{E}[\epsilon|X] = 0$$

from your data. Assume that the standard kernel assumptions hold, i.e. $\int K(z)\,\mathrm{d}z = 1$, K symmetric and $\int |z|^2 K(z)\,\mathrm{d}z < \infty$, $\int K^2(z)\,\mathrm{d}z < \infty$, m and the density f of the regressor X are smooth $C^2(\mathbb{R})$, and the required moment conditions are fulfilled.

a) True or false? Briefly justify your answer, by stating appropriate formulas and giving the correct interpretation.

   i) When we use a local linear type estimator $\hat{m}^{LL}$, we will underestimate the true function in a small neighborhood around a local maximum, i.e. with high probability for large n it is $\mathbb{E}[\hat{m}^{LL}(x)|X_1, \ldots, X_n] < m(x)$ for x in such an area. 2 P

   ii) The bias of the local constant estimator is generally larger than the bias of the local linear estimator while the variance is the same in both cases 2 P

   iii) When using higher order kernels of order r and corresponding smoothness assumptions, for the optimal choice of bandwidth via MSE or IMSE it is $h_{opt} = O(n^{-1/\rho})$, where ρ increases with r, but ρ can never be an even number. 2 P

   iv) A small number of large outliers in the regressor variable will directly cause local constant type estimators to be inconsistent. 2 P

b) Apart from Kernel smoothing there exist other nonparametric techniques to estimate m. Name two of such methods and explain the basic idea and estimation steps of one of them. Briefly indicate what plays the role of the smoothness parameter in this setting and what restrictions on this parameter apply in order to obtain a consistent estimate. 5 P

c) Now assume the dimension d of regressors is large, $d \gg 1$. Therefore consider now estimating a semiparametric regression model of the form

$$Y = g(X_1'\theta_1) + X_2'\theta_2 + \epsilon.$$

using an iid sample $(Y_i, X_{1i}, X_{2i})_{i=1}^n \in \mathbb{R} \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with $d_1 + d_2 = d$. Here $g : \mathbb{R} \to \mathbb{R}$ is a smooth function, $\theta = (\theta_1, \theta_2)$ is a finite dimensional parameter with $\theta_1 \in \Theta_1 \subset \mathbb{R}^{d_1}$ and $\theta_2 \in \Theta_2 \subset \mathbb{R}^{d_2}$, and $\mathbb{E}[\epsilon|X] = 0$.

   i) In what sense is this a semiparametric model? Briefly explain the main advantages and disadvantages of semiparametric models compared to fully nonparametric and parametric ones. 2 P

ii) Are g and θ identified? If not, can they be identified under additional restrictions? 2 P

iii) Suggest a sensible estimation procedure for (the identfied features of) the unknown parameter vector θ that fits into the general framework of semiparametric, two-step extremum estimators.

Hint: Note that the model is a mixture of a single index and a partial linear one, thus combining standard approaches for these two classes might be a good idea. 3 P

Appendix

(i) The p.d.f. of a Gamma distribution $G(\alpha, \beta)$ is given by

$$f(y; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \quad \alpha, \beta > 0.$$

(ii) The probability mass function of a Poisson distribution is given by

$$p(y; \theta) = \frac{e^{-\theta}\theta^{y}}{y!}, \quad \theta > 0, \quad y = 0, 1, \ldots$$