Aufgabensammlung Mikroökonometrie
2016S

Ulrike Schneider

Aufgaben zu Kapitel 1

1. Sei $\hat{\theta}$ ein Schätzer für den Parameter $\theta \in \mathbb{R}$. Dann ist der *mean square error* MSE definiert als $\mathrm{MSE}_\theta(\hat{\theta}) := \mathrm{E}[(\hat{\theta} - \theta)^2]$, also als die erwartete quadratische Abweichung des Schätzers vom wahren Wert.

   (a) Zeigen Sie, dass $\mathrm{MSE}_\theta(\hat{\theta}) = \mathrm{Bias}_\theta^2(\hat{\theta}) + \mathrm{Var}_\theta(\hat{\theta})$ gilt.

   (b) Es sei $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unbekannt. Geschätzt werden soll der Parameter σ^2. Geben Sie das statistische Modell, den Parameterraum und den Raum des zu schätzenden Parameters an. Berechnen Sie die Werte $\mathrm{MSE}_{\sigma^2}(\hat{\sigma}^2)$ und $\mathrm{MSE}_{\sigma^2}(s^2)$, wobei $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ und $s^2 = \frac{n}{n-1}s^2$. Kommentieren Sie das Ergebnis.

   Hinweis: Sie können die Tatsache verwenden, dass $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$ gilt.

2. Zeigen Sie
$$\mathrm{E}_{\theta_0}\left(\frac{\partial}{\partial\theta}\ln f(Z, \theta_0)\right) = 0$$
   und
$$\mathrm{E}_{\theta_0}\left(\frac{\partial}{\partial\theta}\ln f(Z, \theta_0)\frac{\partial}{\partial\theta'}\ln f(Z, \theta_0)\right) = -\mathrm{E}_{\theta_0}\left(\frac{\partial^2}{\partial\theta\partial\theta'}\ln f(Z, \theta_0)\right).$$
   Dabei dürfen Sie die Reihenfolge von Differentiation und Integration vertauschen.

3. Die Funktion $g : \mathbb{R}^k \to \mathbb{R}$ ist *konkav*, wenn für alle $x, y \in \mathbb{R}^k$ und für alle $t \in (0, 1)$
$$g(tx + (1-t)y) \geq tg(x) + (1-t)g(y)$$
   gilt. Die Funktion ist *strikt konkav*, wenn obige Ungleichung strikt ist.

   (a) Nehmen Sie an, dass g auf $\mathbb{R}^k$ zweimal stetig differenzierbar ist und zeigen Sie, dass die Funktion strikt konkav ist, falls die Hessematrix negativ definit auf ganz $\mathbb{R}^k$ ist.

   (b) Zeigen Sie, dass in dem Fall ein stationärer Punkt $x_0 \in \mathbb{R}^k$ (falls er existiert) ein globales Maximum darstellt.

4. Zeigen Sie, dass im klassischen linearen Regressionsmodell mit normalverteilten Fehlern der F-Test zur Hypothese $R\beta = r$ auch als Wald Test aufgefasst werden kann. Wie passt die asymptotische χ^2-Verteilung ins Bild?

Aufgaben zu Kapitel 2

5. Wir betrachten die logistische (Standard-)Verteilung mit cdf $F(x) = \Lambda(x) = 1/(1 + e^{-x})$.

   (a) Berechnen Sie die Dichtefunktion $\lambda(x)$ und zeigen Sie, dass diese symmetrisch um den Ursprung ist.

   (b) Berechnen Sie Erwartungswert und Varianz.

   Hinweis: Die Berechnung der Varianz ist zB über die mgf $M_X(t)$ möglich. Dazu können folgende Identitäten hilfreich sein:

$$\begin{aligned} B(x,y) &= \int_0^\infty \frac{u^{x-1}}{(1+u)^{x+y}} \quad \text{für } x, y > 0 \\ B(x,y) &= \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \\ \frac{\pi^2}{6} &= \Gamma''(1) - \Gamma'(1)^2 \end{aligned}$$

   (c) Zeigen Sie, dass $\lambda(x) = \Lambda(x)(1 - \Lambda(x))$ gilt.

   (d) Plotten Sie die Dichtefunktion der logistischen Verteilung zusammen mit der Dichtefunktion einer Standardnormalverteilung.

6. Zeigen Sie, dass für unabhängige Gumbel-verteilte Zufallsvariablen u und v die Differenz $u - v$ einer logistischen Verteilung folgt.

   Hinweis: Die cdf einer Gumbel- (oder auch log-Weibull-) Verteilung ist durch $G(x) = \exp(-\exp(-x))$ gegeben.

7. (vorgezogen aus Kapitel 4) Zeigen Sie, dass für eine gestutzte (*truncated*) normalverteilte Zufallsvariable $Y|Y > a$ mit $Y \sim N(\mu, \sigma^2)$ gilt:

   (a) $\mathrm{E}[Y|Y > a] = \mu + \sigma\lambda(\alpha)$, und

   (b) $\mathrm{Var}(Y|Y > a) = \sigma^2[1 + \alpha\lambda(\alpha) - \lambda^2(\alpha)] = \sigma^2[1 - \delta(\alpha)]$,

   wobei $\lambda(\alpha) = \phi(\alpha)/(1 - \Phi(\alpha))$, $\delta(\alpha) = \lambda'(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$ und $\alpha = \frac{a-\mu}{\sigma}$.

8. (vorgezogen aus Kapitel 4) Zeigen Sie, dass der *inverse Mill's ratio* $\lambda(a) = \phi(a)/(1 - \Phi(a))$ die Ungleichung

$$\lambda(a) > a$$

   erfüllt und somit auch $\delta(a) > 0$ gilt.

   Hinweis: Beispiel 7(a).

9. (vorgezogen aus Kapitel 5) Zeigen Sie, dass für eine stetige nicht-negative Zufallsvariable T

$$\mathrm{E}[T] = \int_0^\infty S(t)dt$$

   gilt, wobei $S(t)$ die *survivor function* $S(t) = P(T > t)$ bezeichnet.

10. Leiten Sie aus der allgemeinen Form die Log-likelihood Funktion für das Logit Modell her.

11. Laden Sie den Datensatz *`SwissLabor`* aus dem `R`-package *`AER`*. Details über die Daten und die Verwendung der *`glm`*-Funktion können Sie dem dem entsprechenden Helpfile entnehmen). Regressieren Sie die Variable *`participation`* auf die restlichen Variablen des Datensatzes, sowie auf *`age`*^{2}. Verwenden Sie dabei sowohl Logit als auch Probit Modell.

(a) Kommentieren Sie Ihre Ergebnisse.

(b) Bilden Sie für beide Modelle den *success score* (Trefferquote), der folgendermaßen definiert ist

$$\frac{1}{n}\sum_{i=1}^{n}\left[\, y_i\hat{y}_i + (1-y_i)(1-\hat{y}_i)\right].$$

12. Zeigen Sie, dass der *marginal effect* $\partial P(y_i = 1)/\partial x_{ij}$ im Logit Modell „sinnvoll" durch $0.25\hat{\beta}_{\mathrm{ML},j}$ nach oben abgeschätzt werden kann.

13. Zeigen Sie, dass die Log-Likelihood Funktion $l_n(\beta)$ für das Probit Modell strikt konkav in β ist, falls die Regressormatrix X vollen Spaltenrang k hat.

    Hinweis: Eine Möglichkeit ist es zu zeigen, dass die Funktionen $\ln(\Phi(x))$ und $\ln(1-\Phi(x))$ auf ganz $\mathbb{R}$ strikt konkav sind.

14. Berechnen Sie im Binary-Response Modell den restringierten ML-Schätzer unter der Restriktion $\beta_2 = \cdots = \beta_k = 0$, sowie den Wert der Log-Likelihood Funktion an dieser Stelle. Nehmen Sie dazu an, dass das entsprechende Modell eine Konstante enthält.

15. Berechnen Sie die Hessematrix der Log-Likelihood Funktion im BLM direkt aus dem in der VO hergeleiteten allgemeinen Ausdruck.

16. Berechnen Sie für den Datensatz aus Beispiel 11 jeweils für Logit und Probit Modell das Maß R^2_{RG}.

17. Rechnen Sie für den Datensatz aus Beispiel 11 jeweils für Logit und Probit Modell nach, wie in `R` die Standardfehler und p-Werte für die Koeffizienten gebildet werden. Tun Sie dies nur mithilfe der Daten und dem entsprechenden ML-Schätzer.

Aufgaben zu Kapitel 3

18. Zeigen Sie, dass der Erwartungswert der Gumbelverteilung durch die Euler-Mascheroni Konstante γ gegeben ist und berechnen Sie auch die Varianz.

    Hinweis: $\gamma = -\int_0^\infty \ln x e^{-x} dx$, Beispiel 6.

19. Leiten Sie die *marginal effects* im CLM her.

20. Verwenden Sie den Datensatz ***Fishing*** aus dem `R`-package ***mlogit*** (siehe Datei fishing.pdf für eine Beschreibung). Fitten Sie ein CLM, z.B. mithilfe der Funktion ***logit*** wo Sie die Variable ***mode*** auf die Variablen `price` und ***catch*** regressieren. Das Modell enthält keine Konstante. Beachten Sie, dass die Regressoren "alternative-varying" sind, nicht aber die zu schätzenden Parameter. Schätzen Sie die *marginal effects*. Kommentieren Sie Ihre Ergebnisse.

    Hinweis: In der Funktion ***mlogit*** sollte die Variable ***shape*** auf den Wert ***wide*** gesetzt werden (bezieht sich auf die Struktur des Datensatzes), mit der Variable ***varying*** werden diejenigen Regressoren beschrieben, die "alternative-varying" sind.

21. Überlegen Sie ein Beispiel für ein Multinomial Logit Modell.

22. Leiten Sie die *marginal effects* im MLM her.

23. Fitten Sie für den Datensatz `BankWages` (siehe entsprechendes Helpfile) aus dem R-package `AER` ein MLM indem Sie die Variable `job` auf `education` und `minority` regressieren. Führen Sie die Regression jeweils für ein subset des Datensatzes durch, einmal für `gender = male`, einmal für `gender = female`. Sie können für den fit auch die R-Funktion `mulitnom` verwenden. Schätzen Sie die *marginal effects*.

24. Berechnen Sie den Korrelationskoeffizienten der bivariaten Gumbelverteilung mit cdf $F(x, y) = \exp(-[\exp(-x/\rho) + \exp(-y/\rho)]^{\rho})$ mit $\rho \in (0, 1]$.

    HINWEIS: Achtung schwer.

25. Zeigen Sie, dass das Verhältnis der Wahrscheinlichkeiten der Wahl zweier verschiedener Alternativen innerhalb eines Nests der IIA-Eigenschaft gehorcht, nicht aber wenn die Alternativen in verschiedenen Nestern liegen. Das heißt, zeigen Sie dass

    $$\frac{p_{ijl}}{p_{irs}}$$

    für $j = r$ nicht von den anderen Alternativen abhängt, sehr wohl aber für $j \neq r$.

26. (a) Fitten Sie für den Datensatz `HC` (siehe entsprechendes Helpfile) aus dem R-package `mlogit` ein NLM indem Sie die Variable `depvar` auf die übrigen Variablen (ohne `income`) und alternativabhängige Konstante regressieren. Dabei sollten Sie 2 Nester, jeweils mit den Variablen `gcc, ecc, erc, hpc` (cooling) und `gc, ec, er` (nocooling) bilden. Da die Variable `icca` und `occa` nur für das Nest `cooling` Sinn ergeben, sollten die Werte für die restlichen Variablen auf Null gesetzt werden. (Achtung: das Datenset muss vorher mit der Funktion `mlogit.data` in die richtige Form gebracht werden.) Diskutieren Sie den output.

    (b) Geben Sie das Modell für v_{ijl} an, das dabei verwendet wird. Wieviele Parameter sind enthalten?

    [Alternativ können Sie ein NLM für einen Datensatz Ihrer Wahl fitten!!!]

27. Zeigen Sie, dass im NLM für

    $$v_{ijl} = z'_{ij}\alpha + x'_{ijl}\beta_j,$$

    (wobei $z_{ij} \in \mathbb{R}^p$ und $x_{ijl} \in \mathbb{R}^k$ die erklärenden Variablen und $\alpha \in \mathbb{R}^p$ und $\beta_j \in \mathbb{R}^k$ die unbekannten Parameter) gilt, dass

    $$\left(\sum_{h=1}^{m_j} e^{\frac{v_{ijh}}{\rho_j}}\right)^{\rho_j} = \exp\{z'_{ij}\alpha + \rho_j I_{ij}\},$$

    wobei

    $$I_{ij} := \ln\left(\sum_h e^{x'_{ijh}\frac{\beta_j}{\rho_j}}\right)$$

    die sogenannte *inclusive sum* ist. Zeigen Sie weiters, dass dann

    $$p_{ijl} = p_{il|j}\, p_{ij} = \frac{\exp\{\frac{x'_{ijl}\beta_j}{\rho_j}\}}{\sum_{h=1}^{m_j}\exp\{\frac{x'_{ijh}\beta_j}{\rho_j}\}} \frac{\exp\{z'_{ij}\alpha + \rho_j I_{ij}\}}{\sum_{r=1}^{J}\exp\{z'_{ir}\alpha + \rho_r I_{ir}\}},$$

    gilt.

28. Leiten Sie folgenden Ausdruck für die Log-Likelihood Funktion im NLM her.

    $$\sum_{i=1}^{n}\Big\{-\ln\Big(\sum_{j=1}^{J}\exp\{z'_{ij}\alpha+\rho_j I_{ij}\}\Big)+\sum_{j=1}^{J}\Big[\bar{y}_{ij}\big[z'_{ij}\alpha+\rho_j I_{ij}-\ln\Big(\sum_{l=1}^{m_j}\exp\{\frac{x'_{ijl}\beta_j}{\rho_j}\}\Big)\big]+\sum_{l=1}^{m_j}\frac{y_{ijl}x'_{ijl}\beta_j}{\rho_j}\Big]\Big\},$$

    wobei $\bar{y}_{ij} = \sum_{l=1}^{m_j} y_{ijl}$ Indikator, ob im ersten Schritt Ast j gewählt wurde oder nicht.

Aufgaben zu Kapitel 4

29. Zeigen Sie, dass für eine *left-censored* normalverteilte Zufallsvariable Y, d.h. für

$$Y = \begin{cases} Y^* & \text{für } Y^* > a \\ a & \text{für } Y^* \le a \end{cases}$$

mit $Y^* \sim N(\mu, \sigma^2)$ gilt:

(a) $\mathrm{E}(Y) = \mu + \sigma[\Phi(\alpha)\alpha + \phi(\alpha)]$

(b) (*optional*) $\mathrm{Var}(Y) = \sigma^2(1 - \Phi(\alpha))[1 - \delta(\alpha) + (\lambda(\alpha) - \alpha)^2 \Phi(\alpha)]$,

wobei wieder $\alpha = \frac{a-\mu}{\sigma}$ gilt.

30. Berechnen Sie die *marginal effects* im Standard Tobit Modell

$$\frac{\partial \mathrm{E}(y_i)}{\partial x_i}.$$

31. Fitten Sie ein Standard-Tobit Modell für Tobin's Originaldatensatz. [Z.B. Datensatz `tobin` und Funktion `tobit` im package `AER`.] Vergleichen Sie die Ergebnisse sowohl mit dem Probit-Schätzer $\hat{\alpha}_{\text{PROB}}$ alleine, als auch mit dem Heckit-Schätzer $(\hat{\beta}'_{\text{HECK}}, \hat{\sigma}^2_{\text{HECK}})$.

32. Fitten Sie ein Tobit-II-Modell für den den `Mroz87` Datensatz (*labor supply of married women*) aus dem R-package `sampleSelection`. Die abhängige Variable ist das entsprechende Einkommen (`wage`). Die *participation equation* soll durch

$$\texttt{lfp} = \beta_1^{(1)} + \texttt{age}\,\beta_2^{(1)} + \texttt{age}^2\beta_3^{(1)} + \texttt{faminc}\,\beta_4^{(1)} + \texttt{kids}\,\beta_5^{(1)} + \texttt{educ}\,\beta_6^{(1)} + u_1$$

beschrieben werden, wobei `kids` ein dummy dafür ist, ob es Kinder gibt oder nicht (die Variable muss im Datensatz erst definiert werden). Die Gleichung für das Einkommen soll dann durch

$$\texttt{wage} = \beta_1^{(2)} + \texttt{exper}\,\beta_2^{(2)} + \texttt{exper}^2\beta_3^{(2)} + \texttt{educ}\,\beta_4^{(2)} + \texttt{city}\,\beta_5^{(2)} + u_2$$

gegeben sein. (Für genauere Informationen über die Variablen aus dem Datensatz schauen Sie bitte ins entsprechende helpfile.) Geben Sie sowohl den ML-Schätzer als auch den Heckitschätzer an (beide können mit der R-Funktion `selection` berechnet werden).

33. Berechnen Sie die *marginal effects* im Tobit-II Modell.

34. Es sei $(X, Y)' \sim N(\mu, \Sigma)$ mit $\mu = (\mu_1, \mu_2)'$ und $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$.

(a) $\xi := Y - \mu_2 - \frac{\sigma_{12}}{\sigma_1^2}(X - \mu_1)$ ist unabhängig von X und erfüllt $\xi \sim N(0, \sigma_2^2 - \sigma_{12}^2/\sigma_1^2)$.

(b) Berechnen Sie $\mathrm{E}(Y|X > a)$.

(c) Berechnen Sie $\mathrm{Var}(Y|X > a)$.

Hinweis: (b) und (c): (a) und Beispiel 7.

Aufgaben zu Kapitel 5

35. Zeigen Sie, dass die Likelihoodfunktion für diskrete rechts-zensierte Daten durch

$$l_n(\lambda_1, \ldots, \lambda_p) = \prod_{j=1}^{p} \lambda_j^{d_j} (1 - \lambda_j)^{r_j - d_j}$$

gegeben ist, wobei die Bezeichnungen aus der Vorlesung gelten. Berechnen Sie auch die entsprechenden ML-Schätzer $\hat{\lambda}_{\mathrm{ML},j}$ und argumentieren Sie, warum es keine Korrelation zwischen diesen gibt.

36. Leiten Sie folgenden Varianzschätzer her

$$\widehat{\mathrm{Var}}(\hat{\lambda}_{\mathrm{ML},j}) = \frac{d_j(r_j - d_j)}{r_j^3}.$$

37. Berechnen (plotten) Sie den Kaplan-Meier Schätzer für den Datensatz `aml` aus dem `R`-package `survival`, einmal für Patienten mit erhaltender Chemotherapie und einmal ohne (für Details über den Datensatz siehe entsprechendes helpfile). Geben Sie in Ihren Plots auch Konfidenzbänder mithilfe der Greenwood-Formel an (“`plain`” für Option `conf.type` der Funktion `survfit`).

38. Basierend auf dem Konfidenzintervall für $\ln(-\ln S(t))$, leiten Sie folgendes Konfidenzintervall für $S(t)$ her:

$$\left(\hat{S}_{\mathrm{KM}}(t)^{\exp\{z_{1-\alpha/2}\hat{\sigma}(t)\}}, \hat{S}_{\mathrm{KM}}(t)^{\exp\{-z_{1-\alpha/2}\hat{\sigma}(t)\}}\right),$$

wobei die Bezeichnungen aus der Vorlesung gelten.

39. Plotten Sie für verschiedene Werte der Parameter die Ausfallsrate des *Weibull*, des *generalized Weibull* und des *log-logistic models*.

Aufgaben zu Kapitel 5

40. Berechnen Sie Erwartungswert und Varianz einer Poissonverteilung.

41. Fitten Sie ein Poisson-Regressionsmodell für den Datensatz `DoctorVisits` im `R`-package `AER`. Verwenden Sie zusätzlich die erklärende Variable `age`^{2}. Führen Sie auch einen Test auf *overdispersion* durch.

42. Berechnen Sie die Varianz einer negativen Binomialverteilung.