Aufgabensammlung Mikroökonometrie

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 $\text{MSE}_\theta(\hat{\theta}) := \text{E}[(\hat{\theta} - \theta)^2]$, also als die erwartete quadratische Abweichung des Schätzers vom wahren Wert.

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

   (b) Berechnen Sie für $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ die Werte $\text{MSE}_{\sigma^2}(\hat{\sigma}^2)$ und $\text{MSE}_{\sigma^2}(s^2)$. 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 für den Beweis von Satz 1.9, dass

$$\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) \xrightarrow[n\to\infty]{d} N(0, B_0).$$

   Dazu können Sie das Cramer-Wold device und folgendes CLT benützen:

   Satz. *Seien $Z_1, \ldots, Z_n$ unabhängig mit $E(Z_i) = 0$ und $Var(Z_i) = \sigma_i^2 < \infty$. Für $\sigma^2_{(n)} = \sum_{i=1}^n \sigma_i^2$ gelte $\sigma^2_{(n)}/n \to \sigma^2$ mit $0 < \sigma^2 < \infty$. Dann gilt auch, dass*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i \xrightarrow[n\to\infty]{d} N(0, \sigma^2).$$

Aufgaben zu Kapitel 2

3. 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:

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

(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.

4. 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.

5. 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) Betrachten Sie die Ergebnisse mithilfe der Funktion `summary`. Wie glauben Sie werden die Standardfehler und die z-Werte der Koeffizienten gebildet?

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

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

   wobei *predicted values* $\hat{y}_i$

$$\hat{y}_i := \begin{cases} 1 & x_i\hat{\beta} > 0 \\ 0 & x_i\hat{\beta} \le 0 \end{cases}$$

   sind.

6. Zeigen Sie, dass der *marginal effect* $\partial P(y_i = 1)/\partial x_{ij}$ im logit-Modell „sinnvoll“ durch $0.25\hat{\beta}_j$ nach oben abgeschätzt werden kann.

7. Zeigen Sie, dass die Log-Likelihood Funktion $l_n(\beta)$ für das Logit Modell strikt konkav in β ist, falls die Regressormatrix X vollen Spaltenrank k hat. HINWEIS: Eine Möglichkeit ist zu zeigen, dass die Hessematrix immer negativ definit ist.

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

   HINWEIS: Zeigen Sie, dass die Funktionen $\ln(\Phi(x))$ und $\ln(1 - \Phi(x))$ auf ganz $\mathbb{R}$ strikt konkav sind.

9. Zeigen Sie, dass im Binary-Response Modell mit einer Konstante der restringierte ML-Schätzer unter der Restriktion $\beta_2 = \cdots = \beta_k = 0$ durch $\hat{\beta}^{(0)}_{\mathrm{ML}} = (F^{-1}(\bar{y}), 0, \ldots, 0)'$ gegeben ist. Berechnen Sie auch $l_n(\hat{\beta}^{(0)}_{\mathrm{ML}})$.

10. Berechnen Sie für den Datensatz aus Beispiel 5 jeweils für Logit und Probit Modell das Maß R^2_{RG}.

11. Geben Sie Schätzer für die VC-Matrix von $\hat{\beta}_{\mathrm{ML}}$ in Logit und Probit Modell an und zeigen Sie, dass diese konsistent sind (in dem Sinn dass der entsprechend skalierte Schätzer *unter den Annahmen von Satz 2.6* in Wahrscheinlichkeit gegen die asymptotische VC-Matrix von $\sqrt{n}(\hat{\beta}_{\mathrm{ML}} - \beta_0)$ konvergiert).

    Hinweis: Zeigen Sie, dass $g(z) := \frac{f^2(z)}{F(z)(1-F(z))}$ jeweils global Lipschitz ist.

12. Rechnen Sie für den Datensatz aus Beispiel 5 jeweils für Logit und Probit Modell nach, wie in `R` die Standardfehler und p-Werte für die Koeffzienten gebildet werden. Tun Sie dies nur mithilfe der Daten und des entsprechenden ML-Schätzers.

Aufgaben zu Kapitel 3

13. Verwenden Sie den Datensatz ***Fishing*** aus dem `R`-package ***mlogit*** (dieser Datensatz wurde in der Vorlesung kurz beschrieben). Fitten Sie ein Conditional-Logit Modell, 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. 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.

14. Zeigen Sie, dass die *marginal effects* im Conditional-Logit Modell durch

$$\frac{\partial p_{ij}}{\partial x_{is}} = p_{ij}(\delta_{js} - p_{is})\beta$$

    gegeben sind und schätzen Sie diese für Beispiel 13.

15. Zeigen Sie, dass die Hessematrix der Log-Likelihood Funktion im Conditional-Logit Modell singulär ist, sobald das Modell eine Konstante enthält, das heißt, falls $x_{ij} = (1, \tilde{x}_{ij})$ $j = 1, \ldots, m$ und $i = 1, \ldots, n$ ist.

16. 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.

17. Zeigen Sie, dass der Erwartungswert der Gumbelverteilung durch die Euler-Mascheroni Konstante γ gegeben ist.

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

18. Berechnen Sie die Varianz der Gumbelverteilung.

    Hinweis: Beispiel 4.

19. 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.

20. (a) Fitten Sie für den Datensatz `HC` (wie in der VO besprochen, bzw siehe entsprechendes Helpfile) aus dem R-package `mlogit` ein NLM indem Sie die Variable `depvar` auf die übrigen Variablen (ohne `income`) und alternativ-abhängige 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. Wievielde Parameter werden verwendet?

    [Alternativ können Sie ein NLM für den Datensatz Ihrer Wahl fitten.]

21. 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 $j \neq r$. Verwenden Sie dafür die keine genauere Spezifikation für v_{ijl}.

22. Zeigen Sie, dass ein Nested-Logit Modell mit

$$v_{ijl} \equiv v_{ij}$$

    kein Conditional- oder Multinomial Logit Modell ist.

23. Überprüfen Sie, ob für das Modell aus Beispiel 20 die IIA-Eigenschaft vorliegt oder nicht. Siehe dazu auch Hausman & McFadden (1984) und Train & Croissant (2013).

Aufgaben zu Kapitel 4

24. 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 $\alpha = \frac{a-\mu}{\sigma}$.

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

$$\lambda(a) > a$$

    erfüllt.

    Hinweis Beispiel 24.

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

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

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

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

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

mit den Bezeichnungen aus Beispiel 24 und $\delta(\alpha) = \lambda'(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$.

27. Berechnen Sie die *marginal effects* im Standard-Tobit Modell

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

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

29. 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 24.

30. Berechnen Sie die *marginal effects* im Type-II-Tobit Modell bezüglich $x_i = (x_{i1}, x_{i2}) \in \mathbb{R}^{1\times(k_1+k_2)}$. Schreiben Sie dazu $x_{i1}\beta_1 = x_i\gamma_1$ und $x_{i2}\beta_2 = x_i\gamma_2$ mit $\gamma_1 = (\beta_1', 0')'$ und $\gamma_2 = (0', \beta_2')'$.

31. 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).

32. Beschreiben Sie den Heckit-Schätzer im Type-V-Tobit Modell.

    Hinweis: Erweiteren Sie Beispiel 29, indem Sie auf das Ereignis $X \leq 0$ bedingen. Dazu können Sie verwenden, dass für $Z \sim N(\mu, \sigma^2)$ gilt, dass

$$\begin{aligned} \mathrm{E}(Z|Z \leq 0) &= \mu - \sigma\lambda(\mu/\sigma) \\ \mathrm{Var}(Z|Z \leq 0) &= \sigma^2(1 - \delta(\mu/\sigma)) \end{aligned}$$

Aufgaben zu Kapitel 5

33. Zeigen Sie, dass für eine stetige nicht-negative Zufallsvariable T

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

    gilt, wobei $S(t)$ die *survival function* bezeichnet.

34. 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.

35. Berechnen Sie $\hat{\lambda}_{\mathrm{ML},j}$ und leiten Sie folgenden Varianzschätzer her:

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

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

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

    wobei die Bezeichnungen aus der Vorlesung gelten.

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

39. (a) Geben Sie die Log-Likelihood Funktion für rechts-zensierte Daten (δ_i, T_i) mit $i = 1, \ldots, n$ und zugrundeliegender stetiger Verteilung mit cdf F und pdf f an. Schreiben Sie diese auch in Abhängigkeit von λ und Λ an.

    (b) Geben Sie konkret die Log-Likelihood Funktion für das Weibull Modell an.

40. Zeigen Sie, dass für eine Weibull(α, γ)-verteilte Zufallsvariable T folgende Beziehung gilt:

$$\ln T = -\ln\gamma - \frac{1}{\alpha}W,$$

    wobei W einer Gumbelverteilung folgt.

41. Zeigen Sie, dass auch das *exponential* und das *log-logistic model* ein *AFT-model* darstellen und geben Sie die entsprechenden Fehlerverteilungen an.

42. Hat die stetige Zufallsvariable $W \geq 0$ Hazard Funktion $\lambda(t)$, so hat für $\alpha > 0$ die Zufallsvariable αT Hazard Funktion $\frac{1}{\alpha}\lambda(\frac{t}{\alpha})$.

43. Fitten Sie ein Cox proportional hazard Modell für den Datensatz *ema1996.dta*. Details finden Sie im entsprechenden Kapitel von Cameron & Trivedi.

Aufgaben zu Kapitel 5

44. Berechnen Sie Erwartungswert und Varianz einer Poissionverteilung.

45. Berechnen Sie Erwartungswert und Varianz einer negative Binomialverteilung.