

Fragen:

- Was passiert, wenn wir Likelihood-Inferenz innerhalb von \mathcal{P}_θ betreiben, aber der datengenerierende Prozess $\mathbb{P}_0 \notin \mathcal{P}_\theta$ ist (*Fehlspezifikation*)?
- Was passiert, wenn zwar der Verteilungstyp fehlspezifiziert, jedoch der Erwartungswert korrekt spezifiziert ist (*Quasi-Likelihood*)?
- Kann man auf die Likelihood verzichten und direkt von den Quasi-ML-Schätzgleichungen

$$\mathbb{E} s(\theta) \stackrel{!}{=} 0$$

starten?

Beispiel 3.9 (Lineares Modell). *Wir betrachten wieder die Standard-Annahme*

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$$

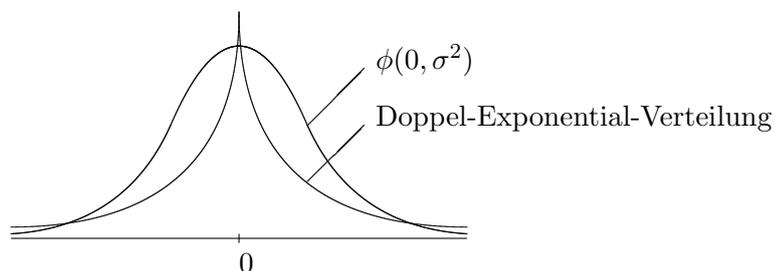
bzw.

$$\mathbf{y} | \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}) \equiv \mathcal{P}_\theta, \quad \theta = (\boldsymbol{\beta}, \sigma^2).$$

Mögliche Fehlspezifikationen:

- (a) Die $N(0, \sigma^2)$ -Annahme für die ε_i ist falsch, zum Beispiel könnte die *wahre* Verteilung die Doppel-Exponential-Verteilung (Laplace-Verteilung) sein:

$$f(\varepsilon_i) \propto \exp(-|\varepsilon_i/\sigma|).$$



Die Doppel-Exponential-Verteilung (oder auch die Cauchy-/ $t(1)$ -Verteilung) ist spitzer im Zentrum und hat breitere Enden (heavy-tails).

⇒ Sie ist ausreißerunempfindlicher.

(b) Die Kovarianzstruktur ist falsch, d.h. $\text{Cov}(\mathbf{y}) \neq \sigma^2 \mathbf{I}$.

Wahre Kovarianzstruktur: $\text{Cov}(\mathbf{y}) = \sigma^2 \mathbf{W}$, zum Beispiel

- $\mathbf{W} = \text{diag}(W_1, \dots, W_n)$ (heteroskedastische Fehler) oder
- \mathbf{W} nichtdiagonal (korrelierte Fehler).

(c) Die Erwartungwertstruktur ist falsch: $\mathbb{E} \mathbf{y} \neq \mathbf{X} \boldsymbol{\beta}$, zum Beispiel wegen

- Fehlspezifikation nichtlinearer Effekte, zum Beispiel $x\beta_1 + x^2\beta_2$ oder $\beta \log x$,
- fehlender Regressoren.

3.4.1 ML-Schätzung bei Fehlspezifikation

Wir beschränken uns auf den i.i.d. Fall: Seien X_1, \dots, X_n i.i.d. wie $X \sim g(x)$ und $g(x)$ die wahre Dichte. Als statistisches Modell betrachten wir die Familie von Dichten

$$\mathcal{P}_\theta = \left\{ f(x|\theta), \theta \in \Theta \right\}.$$

Falls ein $\theta_0 \in \Theta$ existiert mit $g(x) \equiv f(x|\theta_0)$, so ist das Modell korrekt spezifiziert. Falls kein $\theta_0 \in \Theta$ existiert mit $g(x) \equiv f(x|\theta_0)$, ist das Modell fehlspezifiziert.

$$\left(\begin{array}{c} f(x|\theta) \\ \theta \in \Theta \end{array} \right) \bullet g(x) \sim \mathbb{P}_0$$

Definition 3.6 (Kullback-Leibler-Distanz). Die Kullback-Leibler-Distanz von g und f_θ ist definiert durch

$$D(g, f_\theta) = \mathbb{E}_g \left(\log \frac{g(X)}{f(X|\theta)} \right),$$

d.h.

$$D(g, f_\theta) = \int \log \frac{g(x)}{f(x|\theta)} g(x) dx$$

für X stetig. Dabei wird der Erwartungswert bzgl. der „wahren“ Dichte bzw. Wahrscheinlichkeitsfunktion $g(x)$ gebildet.

Es gilt:

$$D(g, f_\theta) \geq 0$$

mit

$$D(g, f_{\theta_0}) = 0 \quad \Leftrightarrow \quad g \equiv f_{\theta_0}.$$

Also:

$$D(g, f_{\theta_0}) = 0 \text{ für ein } \theta_0 \quad \Leftrightarrow \quad \text{Modell korrekt spezifiziert.}$$

Der Beweis erfolgt mit Ungleichung von Jensen (bewiesen im Beweis von Satz 3.4).

Bemerkung. Der (negative) Erwartungswert

$$-\mathbb{E}_g \log g(X) = - \int g(x) \log(g(x)) dx$$

heißt Entropie von g .

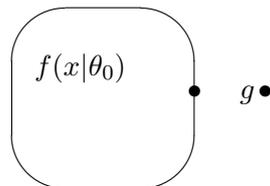
Sei θ_0 „der“ Minimierer der Kullback-Leibler-Distanz:

$$\theta_0 = \operatorname{argmin}_{\theta \in \Theta} \left[\mathbb{E}_g \left\{ \log g(X) \right\} - \mathbb{E}_g \left\{ \log f(X|\theta) \right\} \right].$$

Da $\mathbb{E}_g \left\{ \log g(X) \right\}$ nicht von θ abhängt, gilt auch

$$\theta_0 = \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_g \left\{ \log f(X|\theta) \right\}.$$

Die Dichte $f(x|\theta_0)$ liegt dann im Sinne der Kullback-Leibler-Distanz am „nächsten“ bei g .



Der ML-Schätzer ist

$$\hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \log f(x_i|\theta).$$

Da $\frac{1}{n} \sum_{i=1}^n \log f(x_i|\theta) \xrightarrow{\mathbb{P}} \mathbb{E}_g \log f(X|\theta)$ (Gesetz der großen Zahlen), gilt vermutlich

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0,$$

das heißt der (Quasi-) ML-Schätzer konvergiert gegen jenes θ_0 , dessen Dichte $f(x|\theta_0)$ am nächsten bei g (bezüglich der Kullback-Leibler-Distanz) liegt.

Genauer gilt:

Satz 3.7 (Asymptotische Eigenschaften des ML-Schätzers bei Missspezifikation).

1. *Konsistenz:* Sei θ_0 ein (lokaler) Maximierer von

$$\mathbb{E}_g \log f(X|\theta)$$

(bzw. ein Minimierer von $D(g, f_\theta)$). Unter Regularitätsannahmen (ähnlich wie bei Satz 3.4) existiert eine Folge $\hat{\theta}_n$ von („Quasi-“) ML-Schätzern, das heißt lokalen Maximierern von

$$\frac{1}{n} \sum_{i=1}^n \log f(x_i|\theta)$$

mit

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0.$$

2. *Asymptotische Normalität: Es gilt*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{J}_1^{-1}(\theta_0) \mathbf{I}_1(\theta_0) \mathbf{J}_1^{-1}(\theta_0)\right)$$

mit

$$\mathbf{I}_1(\theta) \equiv \mathbb{E}_g \left(\underbrace{\left(\frac{\partial \log f(X|\theta)}{\partial \theta} \right)}_{\mathbf{s}_1(\theta)} \underbrace{\left(\frac{\partial \log f(X|\theta)}{\partial \theta} \right)^\top}_{\mathbf{s}_1(\theta)^\top} \right)$$

und der (Quasi-) Fisher-Information

$$\mathbf{J}_1(\theta) = \mathbb{E}_g \left(- \frac{\partial^2 \log f(X|\theta)}{\partial \theta \partial \theta^\top} \right).$$

Bemerkung.

- Falls $g(x) \equiv f(x|\theta_0)$, also das Modell korrekt spezifiziert ist, gilt

$$\mathbf{I}_1(\theta) = \mathbf{J}_1(\theta)$$

(vergleiche Satz 2.16), und man erhält die übliche asymptotische Normalverteilung des ML-Schätzers bei korrekter Modellspezifikation.

- Informell gilt

$$\hat{\theta}_n \stackrel{a}{\sim} N \left(\theta_0, \underbrace{\frac{1}{n} \mathbf{J}_1^{-1}(\theta_0) \mathbf{I}_1(\theta_0) \mathbf{J}_1^{-1}(\theta_0)}_{\mathbf{V}(\theta_0)} \right),$$

und $\mathbf{V}(\theta_0)$ wird geschätzt durch

$$\hat{\mathbf{V}}(\hat{\theta}_n) = \mathbf{J}^{-1}(\hat{\theta}_n) \mathbf{I}(\hat{\theta}_n) \mathbf{J}^{-1}(\hat{\theta}_n) \quad (\text{„Sandwich“-Matrix})$$

mit

$$\mathbf{I}(\hat{\theta}_n) = \sum_{i=1}^n \mathbf{s}_i(\hat{\theta}_n) \mathbf{s}_i^\top(\hat{\theta}_n) \quad \text{empirische Fisher-Matrix der Stichprobe,}$$

$$\mathbf{J}(\hat{\theta}_n) = - \sum_{i=1}^n \underbrace{\frac{\partial^2 \log f(x_i|\theta)}{\partial \theta \partial \theta^\top}}_{\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^\top}} \Big|_{\theta=\hat{\theta}_n} \quad \text{empirische beobachtete Informations-Matrix.}$$

Bemerkung.

1. Im i.n.i.d. Fall gilt (informell):

Sei $\ell(\theta, x) = \log f(x|\theta)$ und

$$\theta_0 := \operatorname{argmax}_{\theta} \mathbb{E}_g \ell(\theta, X) = \operatorname{argmax}_{\theta} \mathbb{E}_g \left\{ \sum_{i=1}^n \ell_i(\theta, X_i) \right\},$$

bzw. sei θ_0 die Nullstelle von $\mathbb{E}_g \mathbf{s}(\theta)$, das heißt $\mathbb{E}_g(\mathbf{s}(\theta_0)) = \mathbf{0}$. Außerdem

$$\hat{\theta}_n = \operatorname{argmax}_{\theta} \ell(\theta, x) \quad \text{bzw.} \quad \mathbf{s}(\hat{\theta}_n) = \mathbf{0}.$$

Dann gilt

$$\hat{\theta}_n \stackrel{a}{\sim} N(\theta_0, \hat{\mathbf{V}}(\hat{\theta}_n))$$

wie oben, nur mit $f_i(x_i|\theta)$ an Stelle von $f(x_i|\theta)$.

2. Angenommen, der Modellparameter $\tilde{\theta} = (\theta, \alpha)^\top$ setze sich zusammen aus einem eigentlich interessierenden Parameter θ und einem Nuisance-Parameter α . Die Scorefunktion lautet dann

$$\mathbf{s}(\theta, \alpha) = \begin{pmatrix} s_{\theta}(\theta, \alpha) \\ s_{\alpha}(\theta, \alpha) \end{pmatrix} = \begin{pmatrix} s_{\tilde{\theta}}(\tilde{\theta}) \\ s_{\alpha}(\tilde{\theta}) \end{pmatrix}.$$

Falls trotz fehlspezifizierter Likelihood der eigentlich interessierende Parameter die ML-Gleichung $\mathbb{E}_g(\mathbf{s}_{\theta}(\tilde{\theta}_0)) = 0$ erfüllt, so gilt weiterhin

$$\hat{\theta}_n \stackrel{a}{\sim} N(\theta_0, \hat{\mathbf{V}}(\hat{\theta}_n)) \quad \Rightarrow \quad \text{Quasi-Likelihood.}$$

3.4.2 Quasi-Likelihood und Schätzgleichungen

Frage: Lassen sich Parameter von Interesse wie der Mittelwert μ im i.i.d. Fall oder der Kovariablenvektor β im Regressionsfall noch konsistent und asymptotisch normalverteilt schätzen, wenn das statistische Modell nur teilweise fehlspezifiziert bzw. unvollständig spezifiziert ist?

Beispiel 3.10. Seien Y_1, \dots, Y_n i.i.d. wie $Y \sim f(Y|\mu, \sigma^2)$, f symmetrisch um μ , aber nicht normal, etwa

$$\mathcal{P}_0 = \left\{ f(y|\mu_0) = \frac{1}{2\sigma} e^{-|y-\mu_0|/\sigma} \right\} \quad (\text{Laplace- oder Doppel-Exponential-Verteilung}).$$

Trotzdem wählt man die (Log-) Likelihood

$$\text{ql}(\mu) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 + \text{const}$$

der Normalverteilung als Quasi-(Log-)Likelihood und maximiert diese. So kommt man auf die Quasi-Scorefunktion

$$\text{qs}(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu).$$

Es gilt

$$\mathbb{E}_0 \text{qs}(\mu_0) = \frac{1}{\sigma^2} \sum_{i=1}^n (\underbrace{\mathbb{E}_0(Y_i)}_{=\mu_0} - \mu_0) = 0,$$

also $\hat{\mu}_{QML} = \bar{y}$ wie üblich und wegen $\mathbb{E}_0 \bar{Y} = \mu_0$ erwartungstreu.

Allerdings ist \bar{y} kein (asymptotisch) effizienter Schätzer mehr (die Rao-Cramer-Schranke wird nicht erreicht).

Beispiel 3.11. Seien Y_1, \dots, Y_n unabhängig, $Y_i \sim N(\mu_0, \sigma_i^2)$ und

$$\mathcal{P}_0 = \left\{ \prod_{i=1}^n \phi(y_i | \mu_0, \sigma_i^2) = \frac{1}{(2\pi)^{n/2} \cdot \prod_{i=1}^n \sigma_i} \exp \left(- \sum_{i=1}^n \frac{1}{2} \frac{(y_i - \mu_0)^2}{\sigma_i^2} \right) \right\}.$$

Dann wählt man als Quasi-Log-Likelihood:

$$\text{ql}(\mu) = -\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma} \right)^2,$$

das heißt man ignoriert die Abhängigkeit der Varianz von i und berechnet

$$\begin{aligned} \text{qs}(\mu) &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu), \\ \mathbb{E}_0 \text{qs}(\mu) &= \frac{1}{\sigma^2} \sum_{i=1}^n (\mu_0 - \mu) = 0 \quad \Leftrightarrow \quad \mu_0 = \mu, \\ \hat{\mu}_{QML} &= \bar{y}, \quad \mathbb{E}(\hat{\mu}_{QML}) = \mu_0 \quad \text{erwartungstreu,} \end{aligned}$$

aber

$$\text{Var}_0(\hat{\mu}_{QML}) = \text{Var}_0(\bar{Y}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}_0(Y_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2,$$

das heißt $\hat{\mu}_{QML} = \bar{y}$ ist ineffizient, aber (falls zum Beispiel $\sigma_i^2 \leq c$) konsistent und normalverteilt.

Beispiel 3.12 (Lineares Modell). *Standard-Annahme:*

$$y_i | \mathbf{x}_i \sim N(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2)$$

bzw.

$$\mathbf{y} | \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}).$$

Mögliche Fehlspezifikationen:

- (a) Normalverteilungsannahme falsch,
 (b) Kovarianzstruktur $\text{Cov } \mathbf{y} = \sigma^2 \mathbf{I}$ falsch,
 (c) Erwartungswertstruktur $\mathbb{E} \mathbf{y} = \mathbf{X} \boldsymbol{\beta}$ falsch.

zu (a): Dies ist der Fall, wenn \mathbf{y} nicht normalverteilt ist, aber die Kovarianzstruktur und das Erwartungswertmodell korrekt sind.

Es gilt: $\mathbb{E}_0 \mathbf{y} = \mathbf{X} \boldsymbol{\beta}_0$ ist das wahre Modell.

$$s(\boldsymbol{\beta}) = \frac{1}{\sigma^2} \mathbf{X}^\top (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})$$

$$\mathbb{E}_0 s(\boldsymbol{\beta}_0) = 0$$

Dabei ist $\mathbb{E}_0 s(\boldsymbol{\beta}_0)$ der Erwartungswert im wahren Modell mit wahren Parameter. Es ergibt sich

$$\hat{\boldsymbol{\beta}}_{QML} = \hat{\boldsymbol{\beta}}_{KQ} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

mit

$$\begin{aligned} \mathbb{E}_0(\hat{\boldsymbol{\beta}}_{QML}) &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E} \mathbf{y} = \boldsymbol{\beta}_0 \quad (\text{erwartungstreu}), \\ \text{Cov}_0(\hat{\boldsymbol{\beta}}_{QML}) &= \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}, \end{aligned}$$

also

$$\hat{\boldsymbol{\beta}}_{QML} \stackrel{a}{\sim} N(\boldsymbol{\beta}_0, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1})$$

wie unter NV-Annahme.

zu (b): Die wahre Kovarianzmatrix ist $\sigma^2 \mathbf{W}$ statt $\sigma^2 \mathbf{I}$:

$$\mathbb{P}_0 : \mathbf{y} \sim N(\mathbf{X} \boldsymbol{\beta}_0, \sigma^2 \mathbf{W})$$

$$\mathbb{E}_0 s(\boldsymbol{\beta}_0) = 0$$

$$\hat{\boldsymbol{\beta}}_{QML} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

$$\mathbb{E}_0(\hat{\boldsymbol{\beta}}_{QML}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}_0 = \boldsymbol{\beta}_0$$

$$\begin{aligned} \text{Cov}_0(\hat{\boldsymbol{\beta}}_{QML}) &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \text{Cov}_0(\mathbf{Y}) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \\ &(\neq \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}) \end{aligned}$$

$\hat{\boldsymbol{\beta}}_{QML}$ ist konsistent, aber nicht effizient.

(Ein effizienter Schätzer wäre der gewichtete KQ- bzw. Aitken-Schätzer $\hat{\boldsymbol{\beta}}_{AITKEN} = (\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y}$.)

zu (c): Der wahre Erwartungswert ist ungleich $\mathbf{X} \boldsymbol{\beta}$:

$$\begin{aligned} \text{wahrer Erwartungswert:} & \quad \mathbb{E}_0 \mathbf{y} = \boldsymbol{\mu}_0 = \mathbf{X}_0 \boldsymbol{\beta}_0 \\ \Rightarrow \text{wahres Modell:} & \quad \mathbf{y} \sim N(\mathbf{X}_0 \boldsymbol{\beta}_0, \sigma^2 \mathbf{I}) \end{aligned}$$

(falls N und $\sigma^2 \mathbf{I} = \text{Cov}_0(\mathbf{y})$ richtig). Dann ist

$$\hat{\boldsymbol{\beta}}_{QML} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

$$\mathbb{E}_0(\hat{\boldsymbol{\beta}}_{QML}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}_0 \boldsymbol{\beta}_0 \neq \boldsymbol{\beta}_0 .$$

Somit ist $\hat{\boldsymbol{\beta}}_{QML}$ verzerrter Schätzer, aber liefert das best-approximierende lineare Modell mit Designmatrix \mathbf{X} . Die Kovarianzmatrix ist dann gegeben durch:

$$\text{Cov}_0(\hat{\boldsymbol{\beta}}_{QML}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \underbrace{\text{Cov}_0(\mathbf{y})}_{\sigma^2 \mathbf{I}} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} .$$

Fazit aus den Beispielen:

- Falls die Likelihood oder die Varianzstruktur fehlspezifiziert sind, jedoch die Erwartungswertstruktur

$$\mathbb{E} y_i = \mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$$

korrekt spezifiziert ist, erhält man konsistente Schätzer für $\boldsymbol{\mu}$ bzw. $\boldsymbol{\beta}$.

- Es genügt sogar, die Nullstelle der Quasi-Scorefunktion

$$\text{qs}(\hat{\boldsymbol{\mu}}) \stackrel{!}{=} 0 \quad \text{bzw.} \quad \text{qs}(\hat{\boldsymbol{\beta}}) \stackrel{!}{=} 0$$

zu bestimmen. Falls für das „wahre“ $\boldsymbol{\mu}_0$ bzw. $\boldsymbol{\beta}_0$

$$\mathbb{E}_0 \text{qs}(\boldsymbol{\mu}_0) = 0, \quad \mathbb{E}_0 \text{qs}(\boldsymbol{\beta}_0) = 0$$

gilt, dann ist die Nullstelle $\hat{\boldsymbol{\mu}}$ bzw. $\hat{\boldsymbol{\beta}}$ konsistent und asymptotisch normalverteilt für $\boldsymbol{\mu}$ bzw. $\boldsymbol{\beta}$.

⇒ Idee der „Schätzgleichungen“ (*estimating equations*):

Definiere eine *Schätzfunktion* oder *Quasi-Scorefunktion*

$$\text{qs}(\theta) = \sum_{i=1}^n \psi_i(y_i, \theta)$$

so, dass für den „wahren“ Parameter θ_0

$$\mathbb{E}_0 \text{qs}(\theta_0) = \sum_{i=1}^n \mathbb{E}_0[\psi_i(y_i, \theta_0)] = 0$$

erfüllt ist. Dann ist der *Quasi-ML-Schätzer* oder „*M-Schätzer*“ definiert als Nullstelle

$$\text{qs}(\hat{\theta}_{QML}) \stackrel{!}{=} 0 \quad (\text{Schätzgleichung})$$

der *Schätzfunktion* $\text{qs}(\theta)$.

Beispiel 3.13 (Generalisierte Regression). Sei

$$\begin{aligned} \mathbb{E}_0 y_i &= \mu_i(\boldsymbol{\beta}) && \text{korrekt spezifiziert,} \\ \text{Var}_0 y_i &= \phi v_i(\boldsymbol{\beta}) && \text{(eventuell) fehlspezifiziert.} \end{aligned}$$

Es gilt: $\mathbb{E}_0 s(\boldsymbol{\beta}) = 0$.

Es wird nur eine Annahme hinsichtlich der Schätzgleichung getroffen, jedoch nicht für die Verteilung:

$$\begin{aligned} s(\boldsymbol{\beta}) &= \frac{1}{\phi} \sum_{i=1}^n \left(\frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) v_i(\boldsymbol{\beta})^{-1} \underbrace{(y_i - \mu_i(\boldsymbol{\beta}))}_{\mathbb{E}(y_i - \mu_i(\boldsymbol{\beta}))=0} \\ &\propto \sum_{i=1}^n \left(\frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) v_i(\boldsymbol{\beta})^{-1} (y_i - \mu_i(\boldsymbol{\beta})) \end{aligned}$$

hat Erwartungswert 0 und

$$s(\hat{\boldsymbol{\beta}}) \stackrel{!}{=} 0.$$

$\Rightarrow \hat{\boldsymbol{\beta}}$ ist konsistent und asymptotisch normalverteilt.

Speziell: „generalized estimating equation“ (wie in GLM: $\mu_i(\boldsymbol{\beta}) = h(\mathbf{x}_i^\top \boldsymbol{\beta})$).

Beispiel 3.14 ((Binäre) Longitudinaldaten (repeated measures) oder Clusterdaten). Die Datenpaare $(y_{ij}, \mathbf{x}_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, n_i$, seien je n_i wiederholte Beobachtungen an Individuen oder in „Clustern“, wie zum Beispiel Familien oder Klassen $i = 1, \dots, n$.

n_i : Anzahl der (zeitlich) wiederholten Beobachtungen pro Individuum oder Cluster

y_{ij} : Zielvariable

\mathbf{x}_{ij} : Kovariablenvektor

$y_{ij} | \mathbf{x}_{ij}$ sei aus einer Exponentialfamilie (normal, binomial, Poisson, ...) mit Erwartungswert

$$\mathbb{E}(y_{ij} | \mathbf{x}_{ij}) = h(\mathbf{x}_{ij}^\top \boldsymbol{\beta}) = \mu_{ij}.$$

Die Schätzgleichungen bei Vernachlässigung von (zeitlichen) Korrelationen zwischen den Messwiederholungen lauten

$$\text{qs}(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j=1}^{n_i} \mathbf{x}_{ij} w_{ij}(\boldsymbol{\beta}) (y_{ij} - h(\mathbf{x}_{ij}^\top \boldsymbol{\beta})) \stackrel{!}{=} 0$$

mit

$$\mathbb{E}_{\boldsymbol{\beta}_0} \text{qs}(\boldsymbol{\beta}_0) = 0,$$

wobei die $w_{ij}(\boldsymbol{\beta})$ geeignete Gewichte sind. Somit ist $\hat{\boldsymbol{\beta}}_{\text{QML}}$ konsistent und asymptotisch normal, jedoch unter Effizienzverlust.