Statistische Modellierung III

-Exponentialfamilien-

Dr. Martin Scharpenberg

MSc Medical Biometry/Biostatistics

WiSe 2019/2020

Definition von Exponentialfamilien

Definition von Exponentialfamilien

Definition - Exponentialfamilien

Sei Y eine Zufallsvariable mit Werten in $\mathbb{T} \subseteq \mathbb{R}$, d.h. $\mathbb{T}$ ist der Träger der Verteilungen von Y. Eine Exponentialfamilie ist eine Familie von Dichten bzw. Wahrscheinlichkeitsfunktionen von Y mit der Gestalt

$$f(Y|\theta,\phi,\omega) = \exp\left\{\frac{Y\theta - b(\theta)}{\phi}\omega - c(Y,\phi,\omega)\right\},$$

wobei $\omega \geq 0$ ein (uns bekanntes) **Gewicht**, $\theta \in \Theta$ der (uns unbekannte) **kanonische** (oder natürliche) **Parameter**, $\phi > 0$ der **Dispersionsparameter** (manchmal bekannt und gleich 1) und $b : \Theta \mapsto \mathbb{R}$ eine mindestens dreimal stetig differenzierbare Funktion mit $b''(\theta) > 0$ für alle $\theta \in \Theta$ ist.

Man beachte, dass die erste Ableitung b' stetig und streng wachsend ist

Definition von Exponentialfamilien

Weitere Annahmen:

- b ist streng monoton wachsend
- $c : \mathbb{T} \times \mathbb{R} \mapsto \mathbb{R}$ sei messbar und im zweiten Argument zweimal stetig differenzierbar

Exponentialfamilen - Bemerkungen

- Man kann zeigen, dass $E(Y) = \mu = b'(\theta)$ und $Var(Y) = \phi\, b''(\theta)/\omega$
- Es ist $\log f(Y|\theta,\phi,\omega) = \frac{Y\theta - b(\theta)}{\phi}\omega - c(Y,\phi,\omega)$
- Bei Daten von Einzelbeobachtungen ist typischerweise $\omega = 1$; bei gruppierten Daten ist ω in der Regel gleich der Fallzahl in der Gruppe
- Die Funktion $b : \Theta \longrightarrow \mathbb{R}$ bestimmt (über b') den Erwartungswert und (über b'') die Varianz von Y
- Der Dispersionsparameter bestimmt nur die Varianz von Y. Bei einigen Verteilungen (z.B. Binomialverteilung) ist $\phi = 1$ bekannt

Exponentialfamilen - weitere Bemerkungen

- Beim GLM wird μ und damit θ mit unabhängigen Variablen $\mathbf{x}$ verknüpft, ϕ allerdings nicht
- ϕ ist damit nur von sekundärer Bedeutung und wird als "Störparameter" ("nuissance parameter") bezeichnet. Es wird zur Bestimmung von $Var(Y)$ benötigt
- Da b' monoton wachsend ist, können wir den natürlichen Parameter θ als Funktion des Erwartungswertes μ schreiben:

$$\theta = \theta(\mu) = (b')^{-1}(\mu).$$

- Verwendet man θ als linearen Prädiktor ($\theta = \mathbf{x}_i\beta$), dann ist $g = (b')^{-1}$ die Linkfunktion; sie wird als "natürlicher Link" bezeichnet

Varianzfunktion

Varianzfunktion

- Mit $\theta = \theta(\mu) = (b')^{-1}(\mu)$ erhalten wir

$$f(Y|\mu,\phi,\omega) = \exp\left\{\frac{Y\theta(\mu) - b(\theta(\mu))}{\phi}\omega - c(Y,\phi,\omega)\right\},$$

  also eine Parametrisierung in μ

- Für die Varianz gilt damit

$$Var(Y) = \phi\, v(\mu)/\omega \quad \text{mit} \quad v(\mu) = b''[\theta(\mu)] = b''[(b')^{-1}(\mu)]$$

- Man nennt v die "Varianzfunktion" der Exponentialfamilie. Sie ist bekannt.

Varianzfunktion - Eigenschaften

- Benötigen später (bei der Bestimmung der MLEs) die Ableitung der Funktion $\theta(\mu)$ nach μ

- Es gilt

$$\frac{\partial\,\theta(\mu)}{\partial\mu} = \frac{\partial}{\partial\mu}(b')^{-1}(\mu) = 1/b''[(b')^{-1}(\mu)] = 1/v(\mu)$$

- Zudem folgt aus der vorigen Folie

$$\phi = \frac{Var(Y)}{v(\mu)}\omega = E\left[\omega(Y-\mu)^2/v(\mu)\right]$$

- Man kann zeigen, dass die Exponentialfamilie eindeutig über die Varianzfunktion festgelegt ist

Beispiele für Exponentialfamilien

Bernoulli-Verteilung

- Die Bernoulli-Verteilung hat die Wahrscheinlichkeiten

$$f(y|\pi) = P(Y = y) = \pi^y (1-\pi)^{1-y},$$

wobei $y \in \{0, 1\} = \mathbb{T}$ das Ergebnis des Experiments (1 steht für "Erfolg" und 0 für "Misserfolg") und $0 < \pi < 1$ die Erfolgswahrscheinlichkeit ist

- Damit haben wir

$$\begin{aligned} \log\{f(y|\pi)\} &= y \log \pi + (1-y)\log(1-\pi) = y \log \frac{\pi}{1-\pi} + \log(1-\pi) \\ &= y\,\theta - \log(1+e^\theta), \end{aligned}$$

weil $\log(1-\pi) = \log(\frac{1}{1+e^\theta}) = -\log(1+e^\theta)$

Bernoulli-Verteilung

- Also $f(y|\pi) = \exp\left\{y\,\theta - \log(1+e^\theta)\right\} = \exp\left\{\frac{y\,\theta - \log(1+e^\theta)}{1} \cdot 1 - 0\right\}$
- Die Bernoulli-Verteilung bildet somit eine Exponentialfamilie mit $\theta = \log\frac{\pi}{1-\pi}$ als kanonischen Parameter, $\phi = \omega = 1$, $b(\theta) = \log(1+e^\theta)$ und $c(y,\phi,\omega) = 0$
- Es ergeben sich die bekannten Momente

$$E(Y) = b'(\theta) = \frac{e^\theta}{1+e^\theta} = \pi$$

und

$$Var(Y) = b''(\theta) = \frac{e^\theta(1+e^\theta) - e^{2\theta}}{(1+e^\theta)^2} = \frac{e^\theta}{1+e^\theta}\frac{1}{1+e^\theta} = \pi(1-\pi)$$

- Für die Varianzfunktion gilt $v(\pi) = Var(Y) = \pi(1-\pi)$, weil $Var(Y) = \phi\, v(\mu)/\omega$ und $\phi = \omega = 1$.

Weitere Beispiele

Verteilung	θ	$b(\theta)$	ϕ	ω	$-c(Y, \phi, \omega)$
$B(n, \pi)/n$	$\log(\pi/(1-\pi))$	$\log(1+e^{\theta})$	1	n	$\log\binom{n}{y}$
$NB(n, \pi)/n$	$\log(1-\pi)$	$-\log(1-e^{\theta})$	1	n	$\log\binom{y+n-1}{y}$
$Pois(\lambda)$	$\log(\lambda)$	e^{θ}	1	1	$-\log(y!)$
$N(\mu, \sigma^2)$	μ	$\theta^2/2$	σ^2	1	$-y^2/(2\sigma^2) - 1/2\log(2\pi\sigma^2)$
$Gam(\nu, \lambda)$	$-\lambda/\nu$	$-\log(-\theta)$	$1/\nu$	1	$\nu\log\nu + (\nu-1)\log y - \log\Gamma(\nu)$

Beispiele für Exponentialfamilien