6) Generalizirani linearni modeli (GLM)

Normalni linearni model ("NLM") pretpostavlja

$$Y_i = (x^{(i)})^T \beta + \varepsilon_i, \quad i = 1, \ldots, n$$

neslučajni

za $(\varepsilon_i)_{i=1,\ldots,n} \overset{njd}{\sim} N(0, \sigma^2)$, tj. $Y_1, \ldots, Y_n$ nezavisni t.d.

- $Y_i \sim N(\mu_i, \sigma^2)$, uz
- $\mu_i := \mathbb{E}[Y \mid X = x^{(i)}] = \underbrace{(x^{(i)})^T \cdot \beta}_{=: \eta_i}$, "lin. prediktor"

$\forall i = 1, \ldots, n$.

GLM generalizira gornji model u dva smjera

1° $Y_i \sim$ neka eksponencijalna familija distribucija

($\boxed{B(n,p)}$, $\boxed{Pois(\lambda)}$, Gamma, ...)

2° $\mu_i = u(\eta_i)$ za moguće nelinearnu f-ju u.

6.1 Eksponencijalne familije distribucija

Def. 6.1 Familija distribucija $\mathcal{P} = \{ P_{\theta, \phi} : \theta \in \Theta, \phi \in \Phi \}$ na $\mathbb{R}$ je eksp. familija distr. ako je gustoća (diskretna ili neprekidna) $f_{\theta,\phi}$ od $P_{\theta,\phi}$ oblika

"prirodni" parametar ($\Theta \subseteq \mathbb{R}$), parametar "disperzije" ($\Phi \subseteq \mathbb{R}$)

$$f_{\theta,\phi}(y) = h(y, \phi) \exp\left\{ \frac{1}{\phi}(\theta \cdot y - b(\theta)) \right\}, \quad y \in \mathbb{R} \qquad (6.1)$$

ne ovisi o θ

za neke f-je h i b.

Može se pokazati da za $Y \sim P_{\theta,\phi}$ vrijedi

$$\mathbb{E}[Y] = b'(\theta), \quad Var(Y) = \phi\, b''(\theta), \quad \forall \theta,\phi \qquad (6.2)$$

uz pretp. $Var(Y) > 0, \ \forall \theta, \phi$, imamo da je f-ja

$$\theta \longmapsto \mu(\theta) := \mathbb{E}[Y] \quad (= b'(\theta))$$

<u>invertibilna</u>.

$\hookrightarrow$ umjesto θ možemo koristiti $\underline{\mu \in \mathcal{M}} := \{ \mu(\theta) : \theta \in \Theta \}$

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

$$\hookrightarrow \boxed{Var(Y)} = \phi\, b''(\theta(\mu)) =: \boxed{\phi \cdot \boxed{V(\mu)}}. \qquad (6.3)$$

f-ja "varijance"

<u>Pr. 6.2</u> $(N(\alpha, \sigma^2))$

$$\hookrightarrow f_{N(\alpha,\sigma^2)}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y-\alpha)^2}{2\sigma^2} \right\}$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{ -\frac{y^2}{2\sigma^2} \right\}}_{=: h(y,\sigma^2)} \cdot \exp\left\{ \frac{1}{\sigma^2} \left(\alpha \cdot y - \overbrace{\alpha^2/2}^{=: b(\alpha)} \right) \right\}, \ y \in \mathbb{R}$$

$\Rightarrow \{ N(\alpha, \sigma^2) : \alpha \in \mathbb{R}, \sigma^2 > 0 \}$ je eksponenc. fam. distr.

uz $\boxed{\theta = \alpha}$ i $\boxed{\phi = \sigma^2}$

$\hookrightarrow b'(\theta) = \theta \Rightarrow \boxed{\mu = \theta \ (= \alpha)}$

$b''(\theta) = 1 \Rightarrow \boxed{V(\mu) \equiv 1}$ ⊠

Pr. 6.3 ($Poiss(\lambda)$)

$$\hookrightarrow f_\lambda(y) = \frac{\lambda^y}{y!} e^{-\lambda} \mathbb{1}_{\{y \in \mathbb{N}_0\}}^{(y)}$$

$$= \underbrace{\frac{1}{y!} \mathbb{1}_{\{y\in\mathbb{N}_0\}}^{(y)}}_{=:h(y)} \cdot \exp\{ y \log\lambda - \lambda \}, \quad \lambda = e^{\log\lambda} =: b(\log\lambda)$$

$\Rightarrow$ eksp. fam. distr. uz $\boxed{\theta = \log\lambda \; (\in \mathbb{R})}$ te $\boxed{\phi = 1}$.

$\hookrightarrow$ • $b'(\theta) = e^\theta \Rightarrow \mu = e^\theta \; (=\lambda), \; \boxed{\theta = \log\mu}$

• $b''(\theta) = e^\theta \Rightarrow \boxed{V(\mu) = \mu}$ ∎

Pr. 6.4 ($B(m,p)$)

$$\hookrightarrow f_{m,p}(y) = \binom{m}{y} p^y (1-p)^{m-y} \mathbb{1}_{\{y\in\{0,\dots,m\}\}}$$

$$= \binom{m}{y} \mathbb{1}_{\{y\in\{0,\dots,m\}\}} \exp\left\{ y \log\left(\frac{p}{1-p}\right) - m \log\frac{1}{1-p} \right\}$$

$\Rightarrow Y \sim B(m,p) \Rightarrow Y' := \frac{1}{m} Y \sim B'(m,p)$

$$\hookrightarrow f'_{m,p}(z) = f_{m,p}(m\cdot z)$$

$$= \underbrace{\binom{m}{mz} \mathbb{1}_{\{z\in\{0,\frac{1}{m},\dots,1\}\}}}_{=:h(z,\frac{1}{m})}$$

$$\cdot \exp\left\{ \frac{1}{1/m} \left(z \underbrace{\log\left(\frac{p}{1-p}\right)}_{=:\theta} - \log\frac{1}{1-p} \right) \right\}, \quad \log\frac{1}{1-p} = \log(1+e^\theta) =: b(\theta)$$

$\Rightarrow \{ \frac{1}{m} B(m,p) : m \in \mathbb{N}, p \in \langle 0,1\rangle \}$ je eksp. familija distr.

wz $\boxed{\theta = \operatorname{logit}(p) := \log\left(\frac{p}{1-p}\right)}$; $\boxed{\phi = \frac{1}{m}}$.

- $\mu = b'(\theta) = \frac{e^\theta}{1+e^\theta} =: \operatorname{expit}(\theta) \quad (=p)$
- $V(\mu) = b''(\theta(\mu)) \overset{DZ}{=} \boxed{\mu(1-\mu)}$. ∎