6.2 GLM

Def. 6.5 Za danu eksp. familiju $\{P_{\mu,\phi} : \mu \in M, \phi \in \Phi\}$, GLM pretpostavlja da su $Y_1, \dots, Y_n$ nezavisne t.d.

$$Y_i \sim P_{\mu_i, \phi_i}, \quad i = 1, \dots, n$$

pri čemu $\boxed{\mu_i = g^{-1}(\eta_i)}$

- $g(\mu_i) = (x^{(i)})^T \beta \; (= \eta_i)$ za neki $\beta \in \mathbb{R}^p$, i tzv. **funkciju veze** $g : M \to \mathbb{R}$ (poznata!), te

  npr. (u $\{B(m_i, p)\}$ familiji, $\phi = 1$ te $w_i = m_i$)

- $\phi_i = \dfrac{\phi}{w_i}$, uz "težine" $w_1, \dots, w_n > 0$ (poznate), i $\phi > 0$ potencijalno nepoznat [ϕ katkad zovemo parametar disperzije, kao i ϕ_i]. ◻

Nap. Pretpostavljamo da je $p \leq n$, rang$(X) = p$, g strogo monotona, te da $\exists\, g''$. ◻

Često za g uzimamo tzv. kanonsku f-ju veze def. sa

$$\boxed{g_c(\mu) := \theta(\mu)} \longrightarrow \text{tada} \quad \boxed{\theta_i = \theta(\mu_i) = \theta(g_c^{-1}(\eta_i)) = \eta_i}$$

[g_c ima neka dobra svojstva, ali nije apriori jasno da je to uvijek najbolji izbor za g!]

Pr. 6.6
- $\{N(\mu, \sigma^2)\} \longrightarrow g_c(\mu) = \mu$, tj. $\boxed{\mu_i = \eta_i}$
- $\{\text{Poiss}(\lambda)\} \longrightarrow g_c(\underset{\lambda}{\mu}) = \log \mu$, tj. $\boxed{\mu_i = e^{\eta_i}}$

- $\{\frac{1}{m}B(m,p)\} \longrightarrow g_\tau(\mu) = \underline{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right),$ ($\overset{"}{p}$)

tj. $$\boxed{\mu_i = expit(\eta_i) = \frac{e^{\eta_i}}{1+e^{\eta_i}}}$$

→ "logistički model"

(m = 1 => "logist. regresija") (Y ∈ {0,1})

6.2.1 Procjenitelji za β i ϕ

$$\hat{\beta} := \hat{\beta}_{MLE},$$

gdje $(\hat{\beta}_{MLE}, \hat{\phi}_{MLE})$ maksimizira

$$l(\beta,\phi) := \log \prod_{i=1}^{n} f_{\mu_i,\phi_i}(y_i)$$ (ovisi o β)

$$\overset{(DZ)}{\underset{(6.1)}{=}} \sum_{i=1}^{n} \frac{w_i}{\phi}\{\theta(\underbrace{\mu_i}_{=g^{-1}((x^{(i)})^T\beta)}) \cdot y_i - b(\theta(\mu_i))\}$$

(6.4)

$$+ \sum_{i=1}^{n} \log h(y_i, \phi/w_i)$$

→ $\hat{\beta}_{MLE}$ neovisan o $\hat{\phi}_{MLE}$.

[Interpretacija $\hat{\beta}$-a?]

$$\boxed{\hat{\mu}(x) = \hat{\mathbb{E}}[Y \mid X = x] := g^{-1}(x^T \hat{\beta})}, \quad x \in \mathbb{R}^p$$

Ako je $g = g_\tau$ => $\boxed{\theta(\mu_i) = \eta_i = (x^{(i)})^T\beta}$, pa (DZ) $\hat{\beta}$ zadovoljava

$$\sum_{i=1}^{n} w_i x_{ij}(y_i - \underbrace{\hat{\mu}_i}_{:= \hat{\mu}(x^{(i)})}) = 0, \quad \forall j = 1, \dots p, \qquad (6.5)$$

tj. ako $\underline{\underline{\omega_i = 1}}, \forall i = 1, \ldots, n$, "normalne jdbe"

$$X^T(y - \hat{\mu}) = 0 \qquad (6.6)$$

u NLM $\hat{\mu} = X\hat{\beta} \Rightarrow$ (6.6) postaje

$$X^T y = X^T X \hat{\beta}, \text{ tj. } \hat{\beta} = (X^T X)^{-1} X^T y = \hat{\beta}^{LS}.$$

Inače, osim u NLM, $\hat{\beta}$ aproksimiramo numeričkim (iterativnim) algoritmima (čak i za $g = g_c$).

[Newton-Raphson ili IRLS algoritam]

Ako je ϕ poznat (npr. $\phi = 1$ u $\{B(m,p)\}$ i $\{Poiss(\lambda)\}$ slučaju), $\hat{\phi} := \phi$, a u suprotnom

$$\hat{\phi} := \frac{1}{n - p} \sum_{i=1}^{n} \omega_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \qquad (6.7)$$

$(\neq \hat{\phi}_{MLE})$

zbog procjene $\hat{\beta}_1, \ldots, \hat{\beta}_p$ [generalizirana Pearsonova statistika]

motivirano s $\underbrace{Var(Y_i)}_{= E[(Y_i - \mu_i)^2]} = \frac{\phi}{\omega_i} V(\mu_i)$, $\forall i = 1, \ldots, n$.

Nap. $\hat{\phi} \neq \hat{\phi}_{MLE}$ kada $\boxed{n \to +\infty}$ [generalna teorija MLE]

Nap. Pod nekim uvjetima,

- $\hat{\phi} \xrightarrow{a.s.} \phi$, te
- $\hat{\beta}$ asimptotski normalna $\longrightarrow$ intervali pouzdanosti za β_j

6.2.2 Devijanca i usporedba modela

Ako imamo model ω_0 koji koristi samo $X_1, \dots, X_{p_0}$, i ω_1 koji koristi $X_1, \dots, X_{p_1}$ $(p_0 < p_1 \leq p)$, želimo testirati

$$H_0: \beta_{p_0+1} = \dots = \beta_{p_1} = 0$$

$$H_1: \exists\, i \in \{p_0+1, \dots, p_1\} \text{ t.d. } \beta_i \neq 0 .$$

Ako je H_0 točna, kada $n \to +\infty$,

$$T := 2\big(\underbrace{\ell(\hat{\beta}_1, \phi)}_{\text{iz } \omega_1} - \underbrace{\ell(\hat{\beta}_0, \phi)}_{\text{iz } \omega_0}\big) \sim \chi^2_{p_1 - p_0} \qquad (6.8)$$

$\Rightarrow$ odbacujemo H_0 ako $\boxed{T > (1-\alpha)\text{-kvantila } \chi^2_{p_1-p_0} \text{ razdiobe}}$

Ipak, T moguće izračunati samo ako je ϕ poznat.

Općenito, za svaki (sub)model s p koraka definiramo devijancu modela s

$$D(\hat{\beta}) := 2\big(\ell(\hat{\beta}_{sat}, \phi) - \ell(\hat{\beta}, \phi)\big) \cdot \phi \qquad (6.9)$$

gdje je $\ell(\hat{\beta}_{sat}, \phi)$ dobije tako da u (6.4) stavimo $\hat{\mu}_i := y_i$, $\forall i = 1, \dots, n$ ("saturirani model")

$\Rightarrow$ iz (6.4) slijedi da D ne ovisi o ϕ!

Pod nekim uvjetima, kada $n \to \infty$,

$$T' := \frac{1}{\hat{\phi}}\big(D(\hat{\beta}_0) - D(\hat{\beta}_1)\big) \sim \chi^2_{p_1 - p_0} \qquad (6.9)$$

Nap.) Ako $\hat{\phi} = \phi$, $T' = T$

6.2.3 Odabir modela i generalizacije GLM-a

↳ pretp. da je ϕ poznat.

- Za odabir modela (tj. kovarijata), možemo koristiti CV ili AIC / BIC kriterije gdje je

$$AIC(\hat{\beta}) = -\ell(\hat{\beta}, \phi) + 2 \cdot p$$

p — broj korištenih kovarijata

$$BIC(\hat{\beta}) = -\ell(\hat{\beta}, \phi) + \log(n) \cdot p.$$

- Ridge ili lasso regularizacija:

$$\hat{\beta}^{ridge} := \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} -\ell(\beta, \phi) + \lambda \cdot \|\beta\|_2^2$$

$$\hat{\beta}^{lasso} := \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} -\ell(\beta, \phi) + \lambda \cdot \|\beta\|_1$$

- Generalizirani aditivni model (GAM):

$$g(\mu_i) = \sum_{j=1}^{p} f_j(x_{ij}), \quad i = 1, \ldots, n$$

f_j — procjenjujemo neparametarski