Kapitola 1

VYBRANÉ PARTIE Z TEORIE ODHADU

1.1 ZÁKLADNÍ POJMY MATEMATICKÉ STATISTIKY

V teorii pravděpodobnosti se předpokládá, že je známý pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ a že také známe rozdělení pravděpodobnosti náhodných veličin (resp. náhodných vektorů), které na tomto pravděpodobnostním prostoru uvažujeme.

V matematické statistice však máme k dispozici výsledky n nezávislých pozorování hodnot sledované náhodné veličiny X

$$x_1 = X(\omega_1), \dots, x_n = X(\omega_n),\ \omega_1, \dots, \omega_n \in \Omega$$

a na základě těchto pozorování chceme učinit výpověď o rozdělení zkoumané náhodné veličiny.

Definujme nejprve základní pojmy matematické statistiky.

Definice 1.1.1.
Náhodný vektor

$$\boxed{\mathbf{X}_n = (X_1, \dots, X_n)^{\mathrm{T}}}$$

nazýváme ***náhodným výběrem z rozdělení pravděpodobnosti*** P*, pokud*

(i) $X_1, \dots, X_n$ *jsou nezávislé náhodné veličiny,*

(ii) $X_1, \dots, X_n$ *mají stejné rozdělení pravděpodobnosti* P*.*

Číslo n *nazýváme* ***rozsah náhodného výběru****. Libovolný bod* $\mathbf{x}_n = (x_1, \dots, x_n)^{\mathrm{T}}$*, kde* x_i *je realizace náhodné veličiny* X_i *(*$i = 1, \dots, n$*), budeme nazývat* ***realizací náhodného výběru*** $\mathbf{X}_n = (X_1, \dots, X_n)^{\mathrm{T}}$*. Množinu všech hodnot, kterých může náhodný výběr nabýt, nazýváme* ***výběrový prostor*** *a budeme jej značit* $\mathcal{X}$*.*

Základní dělení matematické statistiky je dané strukturou množiny všech možných rozdělení (označme ji $\mathcal{P}$) náhodného výběru $\mathbf{X}$. Velmi často vybíráme do množiny $\mathcal{P}$ jen rozdělení, která jsou stejného typu a která závisí pouze na nějakém (skalárním či vícerozměrném) parametru. Tento parametr se většinou značí $\boldsymbol{\theta}$ a pravděpodobnostní míry z množiny $\mathcal{P}$ symbolem $P_{\boldsymbol{\theta}}$. Přitom předpokládáme, že parametr $\boldsymbol{\theta}$ nabývá hodnot z nějaké množiny $\boldsymbol{\Theta}$.

Definice 1.1.2.
Množinu $\mathcal{P}$ pravděpodobnostních měr tvaru

$$\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$$

nazýváme ***parametrickou třídou rozdělení****. Vektor $\boldsymbol{\theta}$ nazýváme* ***parametrem rozdělení pravděpodobnosti*** *$P_{\boldsymbol{\theta}}$ a množinu $\boldsymbol{\Theta}$ možných hodnot parametru $\boldsymbol{\theta}$* ***parametrický prostor****.*

Nyní se zmiňme o tzv. rodinách rozdělení.

Definice 1.1.3.
Nechť $g(x)$ je nějaká hustota. Definujme ***rodiny rozdělení***

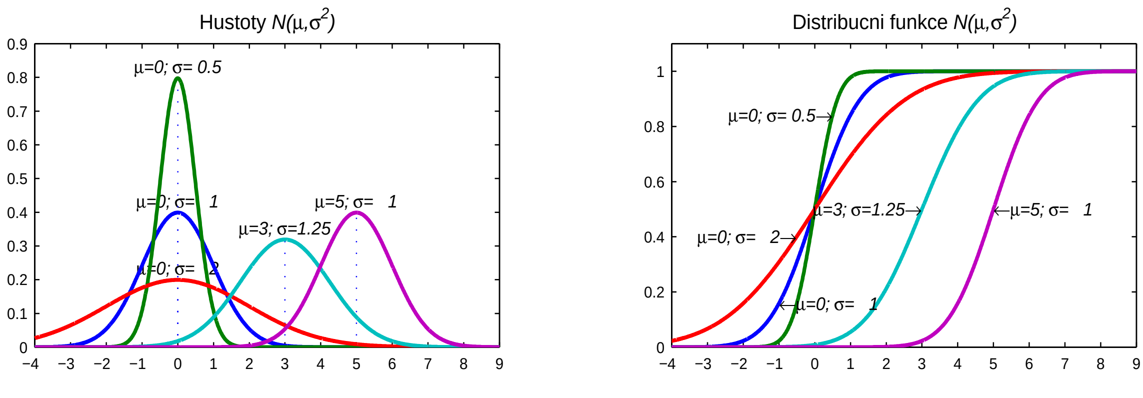
$$\begin{aligned}
\mathcal{F}_1 &= \{f(x;\theta) = g(x-\theta); \theta \in \mathbb{R}\} \\
\mathcal{F}_2 &= \left\{f(x;\delta) = \frac{1}{\delta} g\left(\frac{x}{\delta}\right); \delta > 0\right\} \\
\mathcal{F}_3 &= \left\{f(x;\theta,\delta) = \frac{1}{\delta} g\left(\frac{x-\theta}{\delta}\right); \theta \in \mathbb{R}, \delta > 0\right\}
\end{aligned}$$

Pak říkáme, že $\boxed{\mathcal{F}_1}$ je ***rodina s parametrem polohy*** *(location family), $\boxed{\mathcal{F}_2}$ je* ***rodina s parametrem měřítka*** *(scale family) a $\boxed{\mathcal{F}_3}$ je* ***rodina s parametrem polohy a měřítka*** *(location-scale family).*

Uveďme několik příkladů:
<u>Příklad</u> 1.1.4.
Rodina normálních rozdělení.

$$\begin{aligned}
g(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} & x \in \mathbb{R} \\
f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & x \in \mathbb{R}, ; \mu \in \mathbb{R},\ \sigma > 0
\end{aligned}$$

Hustota $g(x)$ je hustotou tzv. standardizovaného normálního rozdělení.



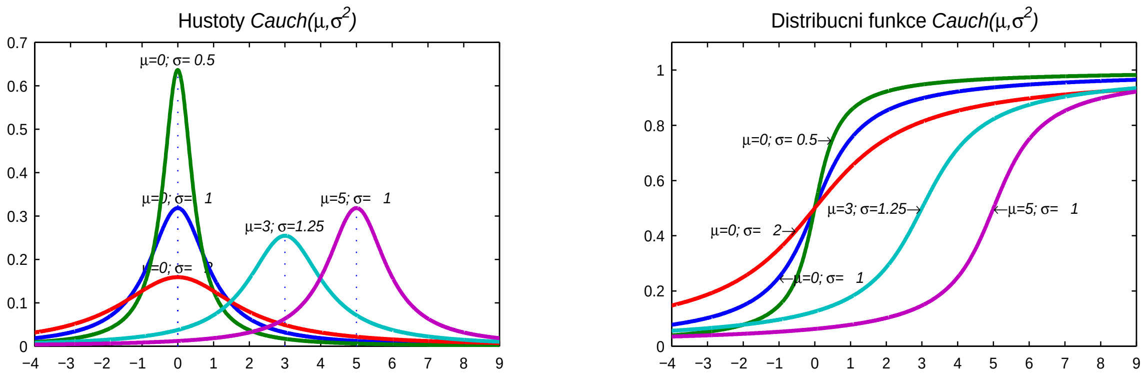
Obrázek 1.1: Ukázky hustot a distribučních funkcí pro $N(\mu, \sigma^2)$.

<u>Příklad</u> 1.1.5.
Rodina Cauchyových rozdělení.

$$g(x) = \frac{1}{\pi}\frac{1}{1+x^2} \qquad\qquad x \in \mathbb{R}$$

$$f(x) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{\sigma}{\pi\left(\sigma^2+(x-\mu)^2\right)} \qquad x \in \mathbb{R}, ; \mu \in \mathbb{R},\ \sigma > 0$$

Hustota $g(x)$ je hustotou Studentova t-rozdělení o 1 stupni volnosti.



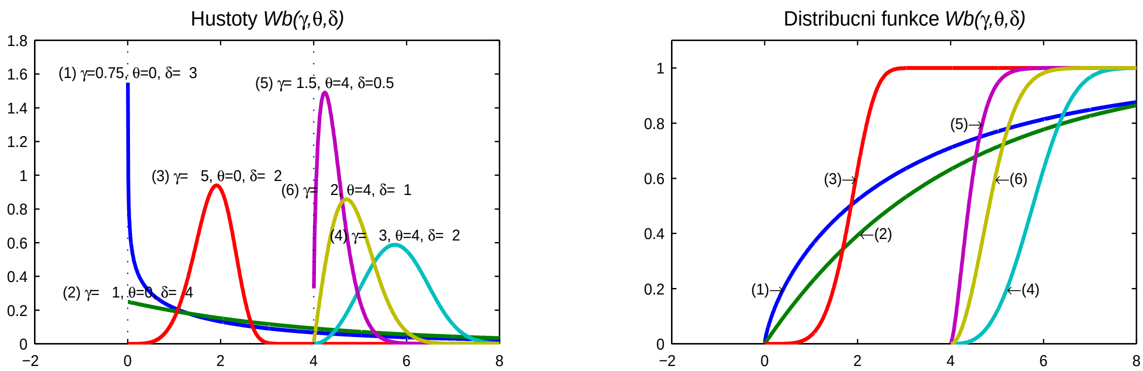
Obrázek 1.2: Ukázky hustot a distribučních funkcí pro $Cauch(\mu, \sigma^2)$.

<u>Příklad</u> 1.1.6.
Rodina Weibullových rozdělení.

$$g(x) = \gamma x^{\gamma-1} e^{-x^{\gamma}} \qquad\qquad x \in \mathbb{R},\ \gamma > 0$$

$$f(x) = \frac{\gamma}{\delta}\left(\frac{x-\theta}{\delta}\right)^{\gamma-1} e^{-\left(\frac{x-\theta}{\delta}\right)^{\gamma}} \qquad x > \theta,\ \theta \in \mathbb{R},\ \gamma > 0,\ \delta > 0$$

Hustota $g(x)$ je hustotou tzv. standardizovaného Weibullova rozdělení.



Obrázek 1.3: Ukázky hustot a distribučních funkcí pro $Wb(\gamma, \theta, \delta)$.

1.2 NESTRANNOST A KONZISTENCE

Cílem teorie odhadu je na základě náhodného výběru odhadnout rozdělení pravděpodobnosti, popřípadě některé parametry tohoto rozdělení, anebo nalézt odhad nějaké funkce parametrů

$\boldsymbol{\theta}$, tj. $\gamma(\boldsymbol{\theta})$. Funkci $\boxed{\gamma(\boldsymbol{\theta})}$ nazýváme **parametrickou funkcí**. V matematické statistice se pro funkce, pomocí kterých budeme odhady provádět, nazývají statistikou. (Tyto funkce jsou navíc měřitelné).

Definice 1.2.1.
*Libovolnou náhodnou veličinu T_n, která vznikne jako funkce náhodného výběru $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$, budeme nazývat **statistikou**, tj.*

$$\boxed{\cdot}$$

Za lepší odhad se považuje ten, jehož rozdělení je více koncentrované okolo neznámé hodnoty parametru. Tento přirozený požadavek koncentrace rozdělení T_n okolo skutečné hodnoty parametru vyjadřujeme pomocí střední hodnoty a rozptylu.

Definice 1.2.2.
*Nechť $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$ je náhodný výběr z rozdělení pravděpodobnosti $P_{\boldsymbol{\theta}}$, kde $\boldsymbol{\theta}$ je vektor neznámých parametrů. Nechť $\gamma(\boldsymbol{\theta})$ je daná parametrická funkce. Řekneme, že statistika je **nestranným odhadem** parametrické funkce $\gamma(\boldsymbol{\theta})$, jestliže pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí*

$$\boxed{ET_n = \gamma(\boldsymbol{\theta}).}$$

(Neboli formálněji zapsáno $E_{\boldsymbol{\theta}}(T(X_1, \ldots, X_n)) = \gamma(\boldsymbol{\theta})$, kde $E_{\boldsymbol{\theta}}$ značí střední hodnotu, která je počítána při hodnotě parametrů $\boldsymbol{\theta}$.)
Pokud

$$\boxed{ET_n \neq \gamma(\boldsymbol{\theta}),}$$

*pak statistiku nazýváme **vychýleným odhadem** parametrické funkce $\gamma(\boldsymbol{\theta})$ a výraz*

$$\boxed{Bias(T_n) = ET_n - \gamma(\boldsymbol{\theta})}$$

*nazýváme **vychýlením.***
*Řekneme, že statistika je **asymptoticky nestranným odhadem** parametrické funkce $\gamma(\boldsymbol{\theta})$, jestliže pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí*

$$\boxed{\lim_{n\to\infty} ET_n = \gamma(\boldsymbol{\theta}).}$$

<u>Poznámka</u> 1.2.3.
Pokud $ET_n > \gamma(\boldsymbol{\theta})$ pro $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, pak je odhad **kladně vychýlený**. Pokud $ET_n < \gamma(\boldsymbol{\theta})$ pro $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, pak je odhad **záporně vychýlený**.

Klasickými příklady nestranných odhadů jsou statistiky

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Velmi snadno lze ukázat, že $\bar{X}$ je nestranným odhadem střední hodnoty $\mu = EX_i$ $(i = 1, \dots, n)$ a S^2 je nestranným odhadem střední hodnoty $\sigma^2 = DX_i$ $(i = 1, \dots, n)$ náhodného výběru $\mathbf{X}_n = (X_1, \dots, X_n)^{\mathrm{T}}$.

Vlastnost nestrannosti však ještě neposkytuje záruku dobrého odhadu, pouze vylučuje systematické chyby.

<u>Příklad</u> 1.2.4.
Nechť náhodná veličina X má geometrické rozdělení, tj. $X \sim Ge(\theta)$ a

$$f_X(x) = P(X = x) = (1-\theta)^x \theta \qquad 0 < \theta < 1 \;\; x = 0, 1, \dots$$

Veličina X udává počet neúspěchů při výběru z alternativního rozdělení před výskytem prvního úspěchu. Hledejme nestranný odhad pro θ.

Je-li $T(X)$ takový nestranný odhad, musí pro něj platit

$$\boxed{ET(X)} = \sum_{x=0}^{\infty} T(x)(1-\theta)^x \theta = \boxed{\theta} \qquad 0 < \theta < 1,$$

Odtud dostáváme

$$\sum_{x=0}^{\infty} T(x)(1-\theta)^x = 1 \qquad 0 < \theta < 1,$$

takže musí platit

$$\begin{aligned} T(0) &= 1 \\ T(x) &= 0 \qquad \text{pro } x \geq 1. \end{aligned}$$

Tento odhad však není pokládán za vhodný, protože jen minimálně přihlíží k počtu neúspěchů před prvním úspěchem. Závisí jen na tom, zda úspěch nastal hned v prvním pokusu či nikoli.

Může se také stát, že nestranný odhad neexistuje.

<u>Příklad</u> 1.2.5.
Nechť náhodná veličina X má binomické rozdělení, tj. $X \sim Bi(n, \theta)$ a

$$f_X(x) = P(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad n \geq 1, \;\; 0 < \theta < 1 \;\; x = 0, 1, \dots, n.$$

Sporem ukážeme, že neexistuje nestranný odhad pro parametrickou funkci

$$\boxed{\gamma(\theta) = \frac{1}{\theta}}.$$

Nechť existuje taková funkce T, že pro každé $\theta \in (0, 1)$ platí

$$\boxed{ET(X)} = \sum_{x=0}^{n} T(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} = \boxed{\frac{1}{\theta}} \qquad 0 < \theta < 1.$$

Na levé straně je však polynom proměnné θ nejvýše stupně n, který samozřejmě nemůže být identicky roven $\frac{1}{\theta}$ na intervalu $(0, 1)$.

Definice 1.2.6.
Řekneme, že statistika $T_n = T(X_1, \ldots, X_n)$ *je* ***konzistentním odhadem*** $\gamma(\boldsymbol{\theta})$*, jestliže pro* $\forall \epsilon > 0$ *a* $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$ *platí:*

$$\boxed{\lim_{n\to\infty} P(|T_n - \gamma(\boldsymbol{\theta})| > \epsilon) = 0.}$$

<u>Poznámka</u> 1.2.7.
Odhad T_n je tedy konzistentním odhadem parametrické funkce $\gamma(\boldsymbol{\theta})$, jestliže T_n konverguje podle pravděpodobnosti k $\gamma(\boldsymbol{\theta})$, což značíme

$$\boxed{T_n \xrightarrow{P} \gamma(\boldsymbol{\theta}).}$$

Někdy se této konvergenci říká slabá a odhadu T_n se také říká, že je **slabě konzistentní**.
Používání konzistentních odhadů zaručuje malou pravděpodobnost velké chyby v odhadu parametru, jakmile rozsah výběru dostatečné roste. Následující věta udává postačující podmínku pro konzistentní odhad.

Věta 1.2.8.
Nechť statistika je nestranný nebo asymptoticky nestranný odhad parametrické funkce $\gamma(\boldsymbol{\theta})$ *a platí*

$$\lim_{n\to\infty} DT_n = 0.$$

Pak je statistika $T_n = T(X_1, \ldots, X_n)$ *konzistentním odhadem parametrické funkce* $\gamma(\boldsymbol{\theta})$.

Důkaz. Nechť $\varepsilon > 0$. Z Čebyševovy nerovnosti plyne:

$$P(|T_n - ET_n| \geq \frac{\varepsilon}{2}) \leq \frac{4DT_n}{\varepsilon^2}.$$

Existuje přirozené číslo n_0 tak, že pro $\forall n > n_0$ platí:

$$-\frac{\varepsilon}{2} < \gamma(\boldsymbol{\theta}) - ET_n < \frac{\varepsilon}{2}.$$

Dále platí

$$\begin{aligned} P(|T_n - \gamma(\boldsymbol{\theta})| \geq \varepsilon) &= 1 - P(|T_n - \gamma(\boldsymbol{\theta})| < \varepsilon) = 1 - P(|T_n - ET_n + ET_n - \gamma(\boldsymbol{\theta})| < \varepsilon) \\ &\leq 1 - P(|T_n - ET_n| < \frac{\varepsilon}{2}) = P(|T_n - ET_n| \geq \frac{\varepsilon}{2}) \leq \frac{4DT_n}{\varepsilon^2} \\ \text{a tedy } \lim_{n\to\infty} P(|T_n - \gamma(\boldsymbol{\theta})| \geq \varepsilon) &\leq \frac{4}{\varepsilon^2} \lim_{n\to\infty} DT_n = 0. \end{aligned}$$

□

Pokud existuje více nestranných odhadů je přirozenou otázkou, který z nich je nejlepší. Za nejlepší můžeme považovat ten, který má nejmenší rozptyl mezi všemi nestrannými odhady. Rozdělení každé statistiky však závisí na parametru $\boldsymbol{\theta}$, z čehož vyplývá, že i rozptyl DT_n nestranné statistiky T_n závisí na parametru $\boldsymbol{\theta}$. Může se stát, že odhad minimalizující rozptyl při určité hodnotě parametru není vhodný pro jinou hodnotu parametru - existuje jiný nevychýlený odhad, který má při této hodnotě parametru menší rozptyl. Pokud taková situace nenastane, mluvíme o rovnoměrně nejlepším nestranném odhadu.

Definice 1.2.9.
Nechť T_n je nestranný odhad parametru $\boldsymbol{\theta}$ a pro všechna $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí

$$DT_n \leq DT_n^*,$$

kde T_n^* je libovolný nestranný odhad parametru $\boldsymbol{\theta}$. Potom odhad T_n nazveme ***rovnoměrně nejlepším nestranným odhadem parametru*** $\boldsymbol{\theta}$.

1.3 POSTAČUJÍCÍ STATISTIKY

Nalezení rovnoměrně nejlepších nestranných odhadů není vždy jednoduché. Abychom nalezli odhad, který má nejmenší rozptyl, je vhodná jistá redukce výběru, tj. nahrazení celého výběru jedinou statistikou, takovou, která bude obsahovat þveškerou informaci o parametru $\boldsymbol{\theta}$ß, která byla obsažená ve výběru. Takováto redukce výběrového prostoru se dosáhne pomocí postačujících statistik.

Definice 1.3.1.
Mějme náhodný výběr $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$ z rozdělení pravděpodobnosti $P_{\boldsymbol{\theta}}$, kde $\boldsymbol{\theta}$ je neznámý parametr. Řekneme, že statistika $\mathbf{S}(\mathbf{X})$ je ***postačující (suficientní) statistikou*** *(*sufficient statistic*), jestliže sdružené rozdělení náhodného výběru $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$ podmíněné jevem $\mathbf{S}(\mathbf{X}) = \mathbf{s}$ je pro každé $\mathbf{s}$ nezávislé na $\boldsymbol{\theta}$.*

<u>Příklad</u> 1.3.2.
Nechť náhodný výběr $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$ pochází z alternativního rozdělení s parametrem $\theta \in (0, 1)$. Nechť

$$S = \sum_{i=1}^{n} X_i.$$

Pak $S \sim Bi(n, \theta)$. Nechť $\mathbf{x}_n = (x_1, \ldots, x_n)^{\mathrm{T}}$ je realizace náhodného výběru. Uvažujme podmíněnou pravděpodobnost

$$P(X_1 = x_1, \ldots, X_n = x_n | S = s).$$

(a) Je-li $\sum_{i=1}^{n} x_i \neq s$, pak je tato podmíněná pravděpodobnost rovna nule.

(b) Nechť $\sum_{i=1}^{n} x_i = s$. Pak

$$\begin{aligned}\boxed{P(X_1 = x_1, \ldots, X_n = x_n | S = s)} &= \frac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(S = s)} \\ &= \frac{\prod_{i=1}^{n} P(X_i = x_i)}{P(S = s)} \\ &= \frac{\theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}}{\binom{n}{s} \theta^s (1-\theta)^{n-s}} = \boxed{\frac{1}{\binom{n}{s}}}.\end{aligned}$$

Výsledek nezávisí na θ, takže statistika $S = \sum_{i=1}^{n} X_i$ je **postačující statistikou**.

Uvedeme větu, která se nazývá také větou o faktorizaci a která zjednodušuje hledání postačujících statistik. Kromě toho umožňuje rychle rozhodnout o tom, či je statistika postačující.

Věta 1.3.3.
Neumanovo faktorizační kritérium. *Mějme náhodný výběr* $\mathbf{X}_n = (X_1, \dots, X_n)^{\mathrm{T}}$ *z rozdělení s pravděpodobnostní funkcí (resp. hustotou)* $f(\mathbf{x}; \boldsymbol{\theta})$*, kde* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*. Potom* $S(\mathbf{X})$ *je postačující statistika pro* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*, právě když existují nezáporné měřitelné funkce* g, h *takové, že sdružené rozdělení náhodného výběru je součinem dvou faktorů:*

$$\boxed{f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) = h(\mathbf{x}) \;\; g\left(\mathbf{S}(\mathbf{x}), \boldsymbol{\theta}\right)}$$

(a říkáme, že hustota f *se dá* ***faktorizovat****).*

DŮKAZ. Tvrzení ukážeme pouze pro diskrétní případ.

$\Rightarrow$ Nechť $\mathbf{S}$ je postačující statistika, pak

$$P(\mathbf{X} = \mathbf{x} | \mathbf{S}(\mathbf{X}) = \mathbf{s}) = h(\mathbf{x})$$

a nezávisí na $\boldsymbol{\theta}$. Dále pro sdruženou pravděpodobnostní funkci platí

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) = P(\mathbf{X} = \mathbf{x}) = \underbrace{P(\mathbf{X} = \mathbf{x} | \mathbf{S}(\mathbf{X} = \mathbf{S}(\mathbf{x})))}_{h(\mathbf{x})} \underbrace{P(\mathbf{S}(\mathbf{X}) = \mathbf{S}(\mathbf{x}))}_{g(\mathbf{S}(\mathbf{x}), \boldsymbol{\theta})}$$

$\Leftarrow$ Předpokládejme, že sdruženou pravděpodobnostní funkci lze vyjádřit ve tvaru

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) = h(\mathbf{x}) \, g\left(\mathbf{S}(\mathbf{x}), \boldsymbol{\theta}\right),$$

tj. že ji lze faktorizovat. Označme

$$B_{\mathbf{s}} = \{\mathbf{x} \in \mathbb{R}^n; \mathbf{S}(\mathbf{x}) = \mathbf{s}\}.$$

Nejprve spočtěme

$$\begin{aligned} P(\mathbf{S}(\mathbf{X}) = \mathbf{s}) &= \sum_{\mathbf{x} \in B_{\mathbf{s}}} P(\mathbf{X} = \mathbf{x}) = \sum_{\mathbf{x} \in B_{\mathbf{s}}} h(\mathbf{x}) \, g\left(\mathbf{S}(\mathbf{x}), \boldsymbol{\theta}\right) \\ &= g\left(\mathbf{S}(\mathbf{x}), \boldsymbol{\theta}\right) \sum_{\mathbf{x} \in B_{\mathbf{s}}} h(\mathbf{x}). \end{aligned}$$

Je-li $P(\mathbf{S}(\mathbf{X}) = \mathbf{s}) > 0$ a $\mathbf{S}(\mathbf{x}) \neq \mathbf{s}$, pak je podmíněná pravděpodobnost

$$P(\mathbf{X} = \mathbf{x} | \mathbf{S}(\mathbf{X}) = \mathbf{s}) = 0.$$

Je-li $P(\mathbf{S}(\mathbf{X}) = \mathbf{s}) > 0$ a $\mathbf{S}(\mathbf{x}) = \mathbf{s}$, pak

$$\begin{aligned} P(\mathbf{X} = \mathbf{x} | \mathbf{S}(\mathbf{X}) = \mathbf{s}) &= \frac{P(\mathbf{X} = \mathbf{x})}{P(\mathbf{S}(\mathbf{X}) = \mathbf{s})} = \frac{h(\mathbf{x}) \, g\left(\mathbf{S}(\mathbf{x}), \boldsymbol{\theta}\right)}{g\left(\mathbf{S}(\mathbf{x}), \boldsymbol{\theta}\right) \sum_{\mathbf{x} \in B_{\mathbf{s}}} h(\mathbf{x})} \\ &= \frac{h(\mathbf{x})}{\sum_{\mathbf{x} \in B_{\mathbf{s}}} h(\mathbf{x})} \end{aligned}$$

a tím je dokázáno, že podmíněné rozdělení vektoru $\mathbf{X}$ při dané hodnotě statistiky $\mathbf{S}$ nezávisí na $\boldsymbol{\theta}$ a $\mathbf{S}$ je postačující statistikou pro prametr $\boldsymbol{\theta}$. $\square$

Příklad 1.3.4.
Nechť náhodný výběr $\mathbf{X}_n = (X_1, \dots, X_n)^{\mathrm{T}}$ pochází z Poissonova rozdělení s parametrem $\theta > 0$ s pravděpodobnostní funkcí

$$f_X(x) = P(X = x) = \frac{e^{-\theta}\theta^x}{x!} \qquad x = 0, 1, 2, \dots.$$

Ukážeme, že statistika

$$T = \sum_{i=1}^n X_i$$

je postačující statistikou pro parametr θ, neboť sdružená hustota náhodného výběru je tvaru

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} = \underbrace{e^{-n\theta}\theta^{\sum_{i=1}^n x_i}}_{g(\mathbf{T}(\mathbf{x}),\boldsymbol{\theta})} \underbrace{\left(\prod_{i=1}^n x_i!\right)^{-1}}_{h(\mathbf{x})}.$$

Než uvedeme větu, která ukazuje praktický význam postačujících statistik pro konstrukci nejlepších nestranných odhadů, všimněme si PODMÍNĚNÝCH STŘEDNÍCH HODNOT.

Nechť $\mathbf{Z} = (X, Y)$ je náhodný vektor, $F(x, y)$ je jeho sdružená distribuční funkce a $F_X(x)$ a $F_Y(y)$ odpovídající marginální distribuční funkce. Nechť vektor středních hodnot $E\mathbf{Z}$ existuje (a je konečný).

(1) Nechť pro každou borelovskou množinu $S \in \mathcal{B}$ a pro každé $x \in \mathbb{R}$ existuje funkce $F(x|y)$ taková, že platí

$$P(X \le x, Y \in S) = \int_S F(x|y) dF_Y(y).$$

Potom tuto funkci $F(x|y)$ nazveme **podmíněnou distribuční funkci náhodné veličiny** X **při daném** $Y = y$ (podmíněnou jevem $Y = y$ nebo také vzhledem k Y).

(2) Nechť $T = T(X, Y)$ je transformovaná náhodná veličina. Potom funkci

$$E(T(X,Y)|Y = y) = \int_{\mathbb{R}} T(x, y) dF(x|y) \qquad y \in \mathbb{R}$$

nazveme **podmíněnou střední hodnotou náhodné veličiny** T **za podmínky** $Y = y$ za předpokladu, že uvedený integrál pro všechna $y \in \mathbb{R}$ existuje (a je konečný).

Položme

$$E(T(X,Y)|Y = y) = h(y)$$

a definujme symbolem

$$\boxed{E(T(X,Y)|Y) = h(Y)}$$

náhodnou veličinu, kterou nazveme **(zobecněnou) podmíněnou střední hodnotou náhodné veličiny** $T(X, Y)$ **při daném** Y.

Důležité vlastnosti podmíněných středních hodnot

(a) Nechť X_1, X_2, Y jsou náhodné veličiny a a_0, a_1, a_2 jsou reálné konstanty, pak pokud střední hodnoty EX_1, EX_2 existují

$$E(a_0 + a_1X_1 + a_2X_2|Y) = a_0 + a_1E(X_1|Y) + a_2E(X_2|Y).$$

(b) Nechť X_1, X_2, Y jsou náhodné veličiny a střední hodnota EX existuje, pak

$$E\left[E(X|Y)\right] = EX.$$

(c) Nechť $T_1 = T_1(X, Y)$ a $T_2 = T_2(Y)$ jsou transformované náhodné veličiny, pak

$$E(T_1T_2|Y) = T_2E(T_1|Y).$$

(3) Nechť $T = T(X, Y)$ je transformovaná náhodná veličina. **Podmíněný rozptyl** při daném $Y = y$ je definován vztahem

$$D(T(X,Y)|Y = y) = E\left\{\left[T - E(t|Y = y)^2|Y = y\right]\right\}$$

a **(zobecněný) podmíněný rozptyl** při daném Y je definován vztahem

$$D(T(X,Y)|Y) = E\left\{\left[T - E(t|Y)^2|Y\right]\right\}.$$

Platí

$$DT = E\left[D(T|Y)\right] + D\left[E(T|Y)\right].$$

Věta 1.3.5.
***Rao-Blackwellova**. Nechť $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$ je náhodný výběr z rozdělení pravděpodobnosti $P_{\boldsymbol{\theta}}$, kde $\boldsymbol{\theta}$ je vektor neznámých parametrů. Nechť existuje postačující statistika $\mathbf{S}(\mathbf{X})$ pro parametr $\boldsymbol{\theta}$. Nechť $\gamma(\boldsymbol{\theta})$ je daná parametrická funkce a statistika $T(\mathbf{X})$ je jejím nestranným odhadem, přičemž $ET(\mathbf{X})^2 < \infty$ pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Pak platí*

(i) Pro parametrickou funkci $\gamma(\boldsymbol{\theta})$ existuje nestranný odhad

$$S^*(\mathbf{X}) = S^*\left(\mathbf{S}(\mathbf{X})\right),$$

který je funkcí postačující statistiky $\mathbf{S}(\mathbf{X})$.

(ii) Pro rozptyl nestranného odhadu $S^*(\mathbf{X})$ platí

$$DS^*(\mathbf{X}) \leq DT(\mathbf{X}) \qquad \text{pro každé } \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{1.1}$$

(iii) V nerovnosti (1.1) platí rovnost právě když

$$S^*(\mathbf{X}) = T(\mathbf{X}) \quad \text{s pravděpodobností } 1 \text{ pro každé } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Důkaz. Nechť $T = T(\mathbf{X})$ je libovolný nestranný odhad parametrické funkce $\gamma(\boldsymbol{\theta})$ a $\mathbf{S} = \mathbf{S}(\mathbf{X})$ je postačující statistika pro parametr $\boldsymbol{\theta}$.

(i) Položme

$$\boxed{S^*(s) = E\left(T(\mathbf{X})|\mathbf{S}(\mathbf{X}) = \mathbf{s}\right)}.$$

Protože $S(\mathbf{X})$ je postačující statistikou, funkce $S^*(s)$ nezávisí na $\boldsymbol{\theta}$, tj.

$$S^* = S^*(\mathbf{S}) = S^*\left(\mathbf{S}(\mathbf{X})\right) = E\left[T(\mathbf{X})|\mathbf{S}(\mathbf{X})\right] = E(T|\mathbf{S})$$

je statistika. Ukážeme, že S^* je nestranný odhad parametrické funkce $\gamma(\boldsymbol{\theta})$. Pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí:

$$ES^* = E\left[E\left(T|\mathbf{S}\right)\right] = ET = \gamma(\boldsymbol{\theta}).$$

(ii) Počítejme a upravujme rozptyl statistiky T

$$\begin{aligned} DT = E\,[T-\gamma(\boldsymbol{\theta})]^2 &= E\,\{[T-S^*]+[S^*-\gamma(\boldsymbol{\theta})]\}^2 \\ &= \underbrace{E\,[T-S^*]^2}_{\geq 0} + 2\,\underbrace{E\,\{[T-S^*]\,[S^*-\gamma(\boldsymbol{\theta})]\}}_{=0} + \underbrace{E\,[S^*-\gamma(\boldsymbol{\theta})]^2}_{DS^*} \end{aligned}$$

tj.

$$DT \geq DS^*,$$

neboť střední hodnotu součinu dvou statistik lze vyjádřit takto

$$\begin{aligned} \underbrace{E\,\{[T-S^*]\,[S^*-\gamma(\boldsymbol{\theta})]\}}_{E(U\cdot V)} &= \underbrace{E\,\{E\,\{[T-S^*]\,[S^*-\gamma(\boldsymbol{\theta})]\,|S\}\}}_{E(E(U\cdot V|\mathbf{S}))} \\ &= E\left\{[S^*-\gamma(\boldsymbol{\theta})]\,\underbrace{E\,\{[T-S^*]\,|\mathbf{S}\}}_{=0}\right\} = 0. \end{aligned}$$ □

(iii) V nerovnosti (1.1) platí rovnost právě když

$$E\,[T-S^*]^2 = 0 \quad \text{pro všechna } \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

tj. když pro všechna $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí

$$S^*(\mathbf{X}) = T(\mathbf{X}) \quad \text{s pravděpodobností 1.}$$

<u>Poznámka</u> 1.3.6.
Z uvedené věty vyplývá, že při hledání nejlepších nestranných odhadů se můžeme omezit na odhady, které jsou funkcemi postačujících statistik. Věta 1.3.5 dává návod, jak určit nestranný odhad, který je funkcí postačující statistiky, jestliže známe libovolný nestranný odhad.
<u>Příklad</u> 1.3.7.
Uvažujme výběr z alternativního rozdělení s parametrem $\theta > 0$ s pravděpodobnostní funkcí

$$f_X(x) = P(X = x) = \theta^x (1-\theta)^{1-x} \qquad x = 0, 1$$

a odhad parametrické funkce $\boxed{\gamma(\theta) = \theta}$ počítejme pomocí podmíněné střední hodnoty $\boxed{S^* = E(T|S)}$, kde T je libovolný nestranný odhad $\gamma(\theta) = \theta$.

Je zřejmé, že nestranným odhadem parametru θ je i statistika

$$T = T(\mathbf{X}) = X_1,$$

tj. první člen výběru, neboť

$$EX_1 = \theta.$$

Jak jsme ukázali v příkladu 1.3.2, postačující statistikou pro parametr θ je statistika

$$S = S(\mathbf{X}) = \sum_{i=1}^{n} X_i.$$

Statistika S je součtem nezávislých náhodných veličin s alternativním rozdělením a tedy má binomické rozdělení s parametry n a θ, tj.

$$S = \sum_{i=1}^{n} X_i \sim Bi(n, \theta).$$

Všimněme si, že pravděpodobnost

$$P\left(X_1 = x, \sum_{i=1}^{n} X_i = s\right) = P\left(X_1 = x, \sum_{i=2}^{n} X_i = s - x\right).$$

Náhodné veličiny $X_1 \sim A(\theta) \equiv Bi(1,\theta)$ a $\sum_{i=2}^{n} X_i \sim Bi(n-1,\theta)$ jsou nezávislé, takže

$$\begin{aligned} P\left(X_1 = x, \sum_{i=1}^{n} X_i = s\right) &= P(X_1 = x)\, P\left(\sum_{i=2}^{n} X_i = s - x\right) \\ &= \theta^x (1-\theta)^{1-x} \binom{n-1}{s-x} \theta^{s-x} (1-\theta)^{n-1-s+x} \\ &= \binom{n-1}{s-x} \theta^s (1-\theta)^{n-s}. \end{aligned}$$

Počítejme podmíněnou střední hodnotu za podmínky, že $S = s$

$$\begin{aligned} S^*(s) = E(T|S=s) &= E\left\{X_1 | \sum_{i=1}^{n} X_i = s\right\} = \sum_{x=0,1} x \frac{P\left(X_1 = x, \sum_{i=1}^{n} X_i = s\right)}{P\left(\sum_{i=1}^{n} X_i = s\right)} \\ &= \frac{\binom{n-1}{s-x}\theta^s(1-\theta)^{n-s}}{\binom{n}{s}\theta^s(1-\theta)^{n-s}} = \frac{(n-1)!s!(n-s)!}{n!(s-1)!(n-s)!} = \frac{s}{n}, \end{aligned}$$

Tedy

$$S^*(S) = E(T|S) = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

což je aritmetický průměr všech pozorování.

Podívejme se, jak to vypadá s rozptyly statistik $T = X_1$ a S^*.

$$\begin{aligned} DT &= DX_1 = \theta(1-\theta) \\ DS^* &= D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{\theta(1-\theta)}{n}, \end{aligned}$$

tedy rozptyl druhého nestranného odhadu se n krát zmenšil.

<u>Příklad</u> 1.3.8.

Uvažujme výběr z Poissonova rozdělení s parametrem $\theta > 0$ s pravděpodobnostní funkcí

$$f_X(x) = P(X = x) = \frac{e^{-\theta}\theta^x}{x!} \qquad x = 0, 1, 2, \ldots$$

a odhad parametrické funkce $\boxed{\gamma(\theta) = \theta}$ počítejme pomocí podmíněné střední hodnoty $\boxed{S^* = E(T|S)}$, kde T je libovolný nestranný odhad $\gamma(\theta) = \theta$.

Je zřejmé, že nestranným odhadem parametru θ je i statistika

$$T = T(\mathbf{X}) = X_1,$$

tj. první člen výběru, neboť

$$EX_1 = \theta.$$

Jak jsme ukázali v příkladu 1.3.4, postačující statistikou pro parametr θ je statistika

$$S = S(\mathbf{X}) = \sum_{i=1}^{n} X_i.$$

Dále je třeba si uvědomit, že statistika S je součtem nezávislých náhodných veličin s Poissonovým rozdělením a má také Poissonovo rozdělení s parametrem $n\theta$, tj.

$$S = \sum_{i=1}^{n} X_i \sim Po(n\theta).$$

Počítejme dále pravděpodobnost

$$P\left(X_1 = x, \sum_{i=1}^{n} X_i = s\right) = P\left(X_1 = x, \sum_{i=2}^{n} X_i = s - x\right).$$

Náhodné veličiny $X_1 \sim Po(\theta)$ a $\sum_{i=2}^{n} X_i \sim Po((n-1)\theta)$ jsou nezávislé, takže

$$\begin{aligned} P\left(X_1 = x, \sum_{i=1}^{n} X_i = s\right) &= P\left(X_1 = x\right) P\left(\sum_{i=2}^{n} X_i = s - x\right) \\ &= \frac{e^{-\theta}\theta^x}{x!} \frac{e^{-(n-1)\theta}\left[(n-1)\theta\right]^{s-x}}{(s-x)!}. \end{aligned}$$

Nyní již počítejme podmíněnou střední hodnotu za podmínky, že $S = s$

$$\begin{aligned} S^*(s) = E(T|S = s) &= E\left\{X_1 | \sum_{i=1}^{n} X_i = s\right\} = \sum_{x=0}^{s} x \frac{P\left(X_1 = x, \sum_{i=1}^{n} X_i = s\right)}{P\left(\sum_{i=1}^{n} X_i = s\right)} \\ &= \sum_{x=0}^{s} x \frac{\frac{e^{-\theta}\theta^x}{x!} \frac{e^{-(n-1)\theta}[(n-1)\theta]^{s-x}}{(s-x)!}}{\frac{e^{-n\theta}(n\theta)^s}{s!}} = \sum_{x=0}^{s} x \binom{s}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{s-x}. \end{aligned}$$

Protože výraz $\sum_{x=0}^{s} x\binom{s}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{s-x}$ je střední hodnotou náhodné veličiny s binomickým rozdělením $Bi(s, \frac{1}{n})$, ihned dostaneme

$$S^*(s) = E(T|S = s) = \frac{s}{n}.$$

Tedy

$$S^*(S) = E(T|S) = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

což je aritmetický průměr všech pozorování.

Stejně jak v předchozím případě, všimněme si rozptylů obou odhadů $T = X_1$ a S^*.

$$\begin{aligned} DT &= DX_1 = \theta \\ DS^* &= D\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} DX_i = \frac{\theta}{n}, \end{aligned}$$

tedy rozptyl druhého nestranného odhadu se n krát zmenšil.

Poznámka 1.3.9.
Nahrazení nestranného odhadu T odhadem $S^* = E(T|\mathbf{S})$ ještě neznamená, že jsme mezi všemi nestrannými odhady našli odhad s nejmenším rozptylem. Úplnost postačující statistiky je pro to dostatečnou podmínkou.

Definice 1.3.10.
*Systém parametrických tříd rozdělení $\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ nazveme **úplným**, pokud pro každou měřitelnou funkci $h(\mathbf{x})$ a náhodnou veličinu X s rozdělením z této třídy platí implikace: jestliže*

$$Eh(X) = 0 \quad \textit{pro každé } \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

pak

$$h(X) = 0 \quad \textit{s pravděpodobností } 1 \textit{ pro každé } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Příklad 1.3.11.
Nechť $\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ je třídou binomických rozdělení

$$X \sim P(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad n \geq 1, \;\; 0 < \theta < 1 \;\; x = 0, 1, \ldots, n.$$

Ukážeme, že tento systém je **úplný**. Uvažujme funkci $h(x)$ na množině $\{0, 1, \ldots, n\}$, pro kterou platí

$$Eh(X) = 0 \quad \text{pro každé } \theta \in (0, 1).$$

Tato funkce musí splňovat podmínku

$$Eh(X) = \sum_{x=0}^{n} h(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} = 0 \quad \text{pro každé } \theta \in (0, 1).$$

Tuto podmínku můžeme napsat takto

$$\begin{aligned} Eh(X) &= \sum_{x=0}^{n} h(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} = \underbrace{(1-\theta)^n}_{(1+z)^{-n}} \sum_{x=0}^{n} h(x) \binom{n}{x} \underbrace{\left(\frac{\theta}{1-\theta}\right)^x}_{z^x} \\ &= (1+z)^{-n} \sum_{x=0}^{n} \binom{n}{x} h(x) z^x = 0 \quad \text{pro } z > 0 \end{aligned}$$

Na jedné straně máme polynom n-tého řadu v proměnné z. Pokud se má identicky rovnat nule, musí se všechny jeho koeficienty rovnat nule, tj.

$$h(x) = 0 \quad \text{pro } x = 0, 1, \ldots, n.$$

Proto

$$\boxed{P(h(X) = 0) = 1 \quad \text{pro každé } \theta \in (0, 1)}.$$

Příklad 1.3.12.
Nechť $\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ je třídou Poissonových rozdělení s pravděpodobnostní funkcí

$$f_X(x) = P(X = x) = \frac{e^{-\theta} \theta^x}{x!} \qquad x = 0, 1, 2, \ldots$$

Tento systém je opět **úplný**. Uvažujme funkci $h(x)$ na množině $\{0, 1, 2, \ldots\}$, pro kterou platí

$$Eh(X) = 0 \quad \text{pro každé } \theta > 0.$$

Tato funkce musí splňovat podmínku

$$Eh(X) = \sum_{x=0}^{\infty} h(x) \frac{e^{-\theta}\theta^x}{x!} = 0 \quad \text{pro každé } \theta > 0.$$

Takže

$$\sum_{x=0}^{\infty} h(x) \frac{\theta^x}{x!} = 0 \quad \text{pro každé } \theta > 0.$$

Tato mocninná řada je rovna nule pro všechna $\theta > 0$, takže všechny její koeficienty musí být rovnu nule, tj.

$$h(x) = 0 \quad \text{pro } x = 0, 1, 2, \ldots.$$

Proto

$$\boxed{P(h(X) = 0) = 1 \quad \text{pro každé } \theta > 0}.$$

<u>Příklad</u> 1.3.13.
Nechť $\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ je třídou normálních rozdělení

$$X \sim \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^2} \quad x \in \mathbb{R}, ; \theta > 0$$

Tento systém není úplný. Definujme

$$h(x) = \begin{cases} -1 & x < 0, \\ 1 & x \geq 0. \end{cases}.$$

Pro libovolné $\theta > 0$ platí

$$\frac{1}{\sqrt{2\pi}\theta} \int_{-\infty}^{\infty} h(x) e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^2} dx = -\underbrace{\frac{1}{\sqrt{2\pi}\theta} \int_{-\infty}^{0} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^2} dx}_{=\frac{1}{2}} + \underbrace{\frac{1}{\sqrt{2\pi}\theta} \int_{0}^{\infty} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^2} dx}_{=\frac{1}{2}} = 0.$$

Tedy z vlastnosti, že $Eh(X) = 0$ neplyne, že $P(h(X) = 0) = 1$.

Definice 1.3.14.
Nechť $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$ *je náhodný výběr z rodělení pravděpodobnosti* $\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. *Statistiku* $T(\mathbf{X})$ *nazveme* ***úplnou vzhledem k*** $\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, *pokud její rozdělení pravděpodobností tvoří úplný systém.*

Nyní vyslovíme větu o jednoznačnosti nestranných odhadů založených na postačujících statistikách.

$\square$

Věta 1.3.15.
První Lehmanova-Sheffého věta. *Nechť* $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$ *je náhodný výběr z rodělení pravděpodobnosti* $\mathcal{P} = \{P_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$*. Předpokládejme, že* $T = T(\mathbf{X})$ *je* ***nestranný odhad parametrické funkce*** $\gamma(\boldsymbol{\theta})$*, přičemž* $ET^2 < \infty$ *pro každé* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*. Nechť* $\mathbf{S} = \mathbf{S}(\mathbf{X})$ *je* ***úplná postačující statistika****. Definujme*

$$S^* = E(T|\mathbf{S}).$$

Pak S^* *je* ***nejlepší nestranný odhad parametrické funkce*** $\gamma(\boldsymbol{\theta})$ *a je* ***jediný****.*

Důkaz. Nechť $T = T(\mathbf{X})$ a $T_2 = T_2(\mathbf{X})$ jsou nestranné odhady parametrické funkce $\gamma(\boldsymbol{\theta})$ s konečnými druhými momenty. Označme $S_2^* = E(T_2|\mathbf{S})$. Pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí

$$\begin{aligned} ES^* &= \gamma(\boldsymbol{\theta}) \qquad DS^* \leq DT \\ ES_2^* &= \gamma(\boldsymbol{\theta}) \qquad DS_2^* \leq DT_2 \end{aligned}$$

Máme tedy

$$E(S^* - S_2^*) = E(E(T|\mathbf{S}) - E(T_2|\mathbf{S})) = 0 \quad \text{pro každé } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Z předpokladu o úplnosti plyne, že

$$P(S^* = S_2^*) = 1 \quad \text{pro každé } \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Z toho plyne závěr, že pro nestranné odhady S^* a T_2 platí

$$DS^* \leq DT_2.$$

Proto S^* je nejlepší. Z Raovy-Blackwellovy věty plyne, že T_2 bude stejně dobrý odhad jako S_2^* právě tehdy, bude-li

$$T_2 = S_2^* \quad \text{skoro jistě při každém } \boldsymbol{\theta}.$$

Jelikož víme, že $S^* = S_2^*$, dostáváme odtud $T_2 = S^*$ skoro jistě. □

Poznámka 1.3.16.
V tomto případě nejmenší možný rozptyl nestranného odhadu parametrické funkce $\gamma(\boldsymbol{\theta})$ je roven DS^*. Přitom jde o skutečné dosažitelné minimum.

Věta 1.3.17.
Druhá Lehmanova-Sheffého věta. *Nechť* $\mathbf{S}$ *je úplná postačující statistika. Nechť*

$$W = g(\mathbf{S})$$

je nestranný odhad parametrické funkce $\gamma(\boldsymbol{\theta})$ *a nechť* $EW^2 < \infty$ *pro každé* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*. Pak* W *je* ***nejlepší nestranný odhad parametrické funkce*** $\gamma(\boldsymbol{\theta})$ *a* ***je jediný****.*

Důkaz. Tvrzení je přímým důsledkem první Lehmannovy-Sheffého věty. □

Příklad 1.3.18.
Nechť $\mathbf{X}_n = (X_1, \ldots, X_n)^{\mathrm{T}}$ je náhodný výběr z alternativního rozdělení s pravděpodobnostní funkcí

$$f(x, \theta) = P(X = x) = \theta^x (1 - \theta)^{n-x} \qquad 0 < \theta < 1 \;\; x = 0, 1$$

s pravděpodobností úspěchu $\theta \in (0, 1)$, kde θ je neznámý parametr. Budeme hledat nejlepší nestranný odhad pro

- $\boxed{\theta}$, což je střední hodnota alternativního rozdělení
- a v případě, že $n \geq 2$ také pro $\boxed{\theta(1-\theta))}$, což je rozptyl alternativního rozdělení

$\boxed{\theta}$: Z příkladů 1.3.2 a 1.3.11 vyplývá, že statistika

$$S = \sum_{i=1}^{n} X_i \sim Bi(n, \theta)$$

je **úplnou postačující statistikou**, takže statistika

$$S^*(S) = E(T|S) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

odvozená pomocí Rao-Blackwellovy věty je podle první Lehmanovy-Sheffého věty **nejlepším nestranným odhadem parametru** θ.

$\boxed{\theta(1-\theta))}$: Pomocí Rao-Blackwellovy věty nejprve hledejme statistiku $S^* = E(T|S)$, kde T je nějaký nestranný odhad parametrické funkce $\gamma(\theta) = \theta(1-\theta)$ a S je postačující statistikou pro parametr θ.

Jako **nestranný odhad parametrické funkce** $\gamma(\theta) = \theta(1-\theta)$ vezměme například

$$T = X_1(1 - X_2),$$

neboť

$$ET = E[X_1(1 - X2)] = \underbrace{EX_1 \cdot E(1 - X_2)}_{\text{nezávislost } X_1, X_2} = \theta(1-\theta).$$

Pro $s = 0, 1, \ldots, n$ počítejme

$$\begin{aligned} S^*(s) = E(T|S = s) &= E\left(X_1(1 - X_2) \,\middle|\, \sum_{i=1}^{n} X_i = s\right) \\ &= \frac{P\left(X_1 = 1, 1 - X_2 = 1, \sum_{i=1}^{n} X_i = s\right)}{P(\sum_{i=1}^{n} X_i = s)} \end{aligned}$$

Je-li $s = 0$, je zřejmé, že

$$E\left(X_1(1 - X_2) \,\middle|\, \sum_{i=1}^{n} X_i = s\right) = 0.$$

Nechť nyní $s > 0$. Pak

$$\begin{aligned} S^*(s) &= \frac{P(X_1 = 1)P(X_2 = 0)P\left(\sum_{i=3}^{n} X_i = s - 1\right)}{P(\sum_{i=1}^{n} X_i = s)} \\ &= \frac{\theta(1-\theta)\binom{n-2}{s-1}\theta^{s-1}(1-\theta)^{n-2-s+1}}{\binom{n}{s}\theta^{s}(1-\theta)^{n-s}} = \frac{(n-2)!s!(n-s)!}{n!(s-1)!(n-s-1)!} \\ &= \frac{s(n-s)}{n(n-1)} = \frac{n}{n-1} \cdot \frac{s}{n} \cdot \left(1 - \frac{s}{n}\right) \end{aligned}$$

a
$$S^*(S) = \frac{n}{n-1}\bar{X}(1-\bar{X}),$$
kde
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
Protože statistika
$$S = \sum_{i=1}^{n} X_i \sim Bi(n,\theta)$$
je **úplnou postačující statistikou**, pak podle první Lehmanovy-Sheffého věty je $S^*(S)$ **nejlepším nestranným odhadem parametrické funkce** $\theta(1-\theta)$.

Veličiny $X_1,\dots,X_n$ můžeme chápat jako výběr z $Bi(1,\theta)$. Toto rozdělení má rozptyl $\theta(1-\theta)$. Všimněme si, že pro $i = 1,\dots,n$ platí
$$X_i^2 = X_i,$$
neboť tyto veličiny nabývají pouze hodnot 0 a 1. Nestranný odhad rozptylu pořízený na základě daného výběru je
$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i - n\bar{X}^2\right) = \frac{1}{n-1}\left(n\bar{X} - n\bar{X}^2\right) \\ &= \frac{n}{n-1}\bar{X}(1-\bar{X}) \end{aligned}$$
a odhad $\boxed{S^2}$ je tedy totožný s **nejlepším nestranným odhadem** parametrické funkce $\theta(1-\theta)$.

<u>Příklad</u> 1.3.19.
Nechť $\mathbf{X}_n = (X_1,\dots,X_n)^{\mathrm{T}}$ je náhodný výběr z Poissonova rozdělení s pravděpodobnostní funkcí
$$f_X(x) = P(X=x) = \frac{e^{-\theta}\theta^x}{x!} \qquad x = 0,1,2,\dots$$
kde θ je neznámý parametr. Budeme hledat nejlepší nestranný odhad pro

- $\boxed{\theta}$, což je střední hodnota Poissonova rozdělení
- $\boxed{e^{-\theta} = P(X=0)}$

$\boxed{\theta}$: Z příkladů 1.3.4 a 1.3.12 vyplývá, že statistika
$$S = \sum_{i=1}^{n} X_i \sim Po(n\theta)$$
je **úplnou postačující statistikou**, takže statistika
$$S^*(S) = E(T|S) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$
odvozená pomocí Rao-Blackwellovy věty je podle první Lehmanovy-Sheffého věty **nejlepším nestranným odhadem parametru** θ.

$\boxed{e^{-\theta}}$: Pomocí Rao-Blackwellovy věty nejprve hledejme statistiku $S^* = E(T|S)$, kde T je nějaký nestranný odhad parametrické funkce $\gamma(\theta) = e^{-\theta}$ a S je postačující statistikou pro parametr θ.

Položme

$$T = I_{\{0\}}(X_1) = I(X_1 = 0) = \begin{cases} 1 & X_1 = 0, \\ 0 & \text{jinak.} \end{cases}$$

Protože

$$ET = 1 \cdot P(T = 1) + 0 \cdot P(T = 0) = P(X_1 = 0) = e^{-\theta},$$

pak statistika T je **nestranným odhadem parametrické funkce** $\gamma(\theta) = e^{-\theta}$.

Je-li $n = 1$, pak statistika T je **nejlepším nestranným odhadem** parametrické funkce $\gamma(\theta) = e^{-\theta}$.

Pro $n > 1$ počítejme

$$\begin{aligned} S^*(s) = E(T|S = s) &= E\left(I(X_1 = 0) \middle| \sum_{i=1}^n X_i = s \right) \\ &= \frac{P\left(T = 1, \sum_{i=1}^n X_i = s\right)}{P(\sum_{i=1}^n X_i = s)} = \frac{P\left(X_1 = 0, \sum_{i=2}^n X_i = s\right)}{P(\sum_{i=1}^n X_i = s)} \\ &= \frac{P(X_1 = 0)P\left(\sum_{i=2}^n X_i = s\right)}{P(\sum_{i=1}^n X_i = s)} = \frac{\frac{e^{-\theta}e^{-(n-1)\theta}[(n-1)\theta]^s}{s!}}{\frac{e^{-n\theta}(n\theta)^s}{s!}} = \left(\frac{n-1}{n}\right)^s \end{aligned}$$

a

$$S^*(S) = \frac{n}{n-1}\bar{X}(1 - \bar{X}),$$

kde

$$\bar{X} = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}.$$

Protože statistika

$$S = \sum_{i=1}^n X_i \sim Po(n\theta)$$

je **úplnou postačující statistikou**, pak podle první Lehmanovy-Sheffého věty je $S^*(S)$ **nejlepším nestranným odhadem parametrické funkce** $e^{-\theta}$.

Spočítejme ještě

$$\begin{aligned} ES^* = ES^*(S) = E\left(\frac{n-1}{n}\right)^S &= \sum_{s=0}^{\infty} \left(\frac{n-1}{n}\right)^s \frac{e^{-n\theta}(n\theta)^s}{s!} \\ &= e^{-n\theta} \underbrace{\sum_{s=0}^{\infty} \frac{[(n-1)\theta]^s}{s!}}_{=e^{(n-1)\theta}} = e^{-\theta} \end{aligned}$$

$$ES^{*2} = \sum_{s=0}^{\infty} \left(\frac{n-1}{n}\right)^{2s} \frac{e^{-n\theta}(n\theta)^s}{s!} = e^{-n\theta} \underbrace{\sum_{s=0}^{\infty} \frac{\left[\frac{(n-1)^2}{n}\theta\right]^s}{s!}}_{=e^{\frac{(n-1)^2}{n}\theta}} = e^{-2\theta + \frac{\theta}{n}}$$

$$DS^* = ES^{*2} - (ES^*)^2 = e^{-2\theta + \frac{\theta}{n}} - e^{-2\theta} = e^{-2\theta}\left(e^{\frac{\theta}{n}} - 1\right).$$

1.4 REGULÁRNÍ SYSTÉM HUSTOT

Definice 1.4.1.
Mějme parametrický prostor $\boldsymbol{\Theta} \subset \mathbb{R}^m$*. Řekneme, že* ***systém*** m***-parametrických hustot***

$$\boxed{\mathcal{F}^m_{reg} = \{f(\mathbf{y};\boldsymbol{\theta}) : \boldsymbol{\theta} = (\theta_1,\dots,\theta_m)^{\mathrm{T}} \in \boldsymbol{\Theta}\}}$$

je ***regulární****, jestliže platí*

(1) $\boldsymbol{\Theta} \subset \mathbb{R}^m$ *je otevřená borelovská množina.*

(2) Množina $M = \{\mathbf{y} \in \mathbb{R}^n : f(\mathbf{y};\boldsymbol{\theta}) > 0\}$ *nezávisí na parametru* $\boldsymbol{\theta}$*.*

(3) Pro každé $\mathbf{y} \in M$ *existuje konečná parciální derivace*

$$f_i'(\mathbf{y};\boldsymbol{\theta}) = \frac{\partial f(\mathbf{y};\boldsymbol{\theta})}{\partial \theta_i} \quad (i = 1,\dots,m).$$

(4) Pro všechny $\boldsymbol{\theta} = (\theta_1,\dots,\theta_m)^{\mathrm{T}} \in \boldsymbol{\Theta}$ *platí*

$$\int_M \frac{f_i'(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta}) = \int_M \frac{\partial \ln f(\mathbf{y};\boldsymbol{\theta})}{\partial \theta_i} dF(\mathbf{y};\boldsymbol{\theta}) = 0 \qquad i = 1,\dots,m,$$

kde $F(\mathbf{y};\boldsymbol{\theta})$ *je odpovídající distribuční funkce.*

(5) Pro všechny $\boldsymbol{\theta} = (\theta_1,\dots,\theta_m)^{\mathrm{T}} \in \boldsymbol{\Theta}$ *je integrál*

$$J_{ij}(\boldsymbol{\theta}) = \int_M \frac{\partial \ln f(\mathbf{y};\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln f(\mathbf{y};\boldsymbol{\theta})}{\partial \theta_j} dF(\mathbf{y};\boldsymbol{\theta}) \qquad i,j = 1,\dots,m$$

konečný a matice $\boxed{\mathbf{J} = \mathbf{J}(\boldsymbol{\theta}) = (J_{ij}(\boldsymbol{\theta}))_{i,j=1}^m}$ *je pozitivně definitní. Matice* $\mathbf{J}$ *se nazývá* ***Fisherova informační matice o parametru*** $\boldsymbol{\theta}$***.***

POZNÁMKA 1.4.2.
Pro jednoduchost někdy hovoříme o **regulárnosti** $f(\mathbf{y};\boldsymbol{\theta})$, ne o **regulárnosti systému hustot**.
POZNÁMKA 1.4.3.
Podmínka (4) souvisí s otázkou, zda při derivování rovnosti

$$1 = \int_M dF(\mathbf{y};\boldsymbol{\theta})$$

lze zaměnit pořadí derivace a integrálu, tj. pro $j = 1,\dots,m$

$$\boxed{0} = \frac{\partial}{\partial \theta_j} 1 = \frac{\partial}{\partial \theta_j} \int_M dF(\mathbf{y};\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} \int_M \frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta}) \stackrel{?}{=} \underbrace{\int_M \frac{\partial}{\partial \theta_j} \frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta}) = \boxed{0}}_{(*)}.$$

Pokud platí $(*)$, pak pořadí lze zaměnit. Uvažujme nejprve spojitý případ

$$\boxed{0} = \int_M \frac{\partial}{\partial\theta_j}\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta}) = \int_M \frac{\partial}{\partial\theta_j}\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}f(\mathbf{y};\boldsymbol{\theta})d\mathbf{y} = \int_M \frac{\partial}{\partial\theta_j}f(\mathbf{y};\boldsymbol{\theta})d\mathbf{y}$$
$$= \boxed{\int_M f_j'(\mathbf{y};\boldsymbol{\theta})d\mathbf{y}} = \int_M f_j'(\mathbf{y};\boldsymbol{\theta})\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}d\mathbf{y} = \boxed{\int_M \frac{f_j'(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta})}.$$

Zcela analogicky pro diskrétní případ máme

$$\boxed{0} = \int_M \frac{\partial}{\partial\theta_j}\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta}) = \sum_{\mathbf{y}\in M} \frac{\partial}{\partial\theta_j}\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}f(\mathbf{y};\boldsymbol{\theta}) = \sum_{\mathbf{y}\in M} \frac{\partial}{\partial\theta_j}f(\mathbf{y};\boldsymbol{\theta})$$
$$= \boxed{\sum_{\mathbf{y}\in M} f_j'(\mathbf{y};\boldsymbol{\theta})} = \sum_{\mathbf{y}\in M} f_j'(\mathbf{y};\boldsymbol{\theta})\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} = \boxed{\int_M \frac{f_j'(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta})}.$$

Definice 1.4.4.
Nechť $f \in \mathcal{F}^m_{reg}$. Pak náhodný vektor

$$\boxed{\mathbf{U} = \mathbf{U}(\boldsymbol{\theta}) = (U_1(\boldsymbol{\theta}), \dots, U_m(\boldsymbol{\theta}))^{\mathrm{T}}} \qquad \text{se složkami} \qquad \boxed{U_i = U_i(\boldsymbol{\theta}) = \frac{\partial \ln f(\mathbf{Y};\boldsymbol{\theta})}{\partial\theta_i}}$$

se nazývá ***skórový vektor*** *příslušný fustotě f.*

<u>POZNÁMKA</u> 1.4.5.
Protože

$$P(\mathbf{Y} \in M) = \int_M dF(\mathbf{y};\boldsymbol{\theta}) = 1,$$

je skórový vektor

$$\mathbf{U}(\boldsymbol{\theta}) = \left(\frac{\partial \ln f(\mathbf{Y};\boldsymbol{\theta})}{\partial\theta_1}, \dots, \frac{\partial \ln f(\mathbf{Y};\boldsymbol{\theta})}{\partial\theta_m}\right) = \left(\frac{f_1'(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})}, \dots, \frac{f_m'(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})}\right)$$

definován s pravděpodobností 1 (s pravděpodobností 1 je $f(\mathbf{y};\boldsymbol{\theta}) > 0$, tj. pro všechna $\mathbf{y} \in M$).

Dále pomocí skórového vektoru lze podmínky (4) a (5) z definice regularity systému hustot vyslovit takto:

(4) existuje $\boxed{E\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}}$
(5) existuje $\boxed{D\mathbf{U}(\boldsymbol{\theta}) = E\left[\mathbf{U}(\boldsymbol{\theta})\mathbf{U}^{\mathrm{T}}(\boldsymbol{\theta})\right] = J(\boldsymbol{\theta})}$

Věta 1.4.6.

(1) Je-li $\boxed{f \in \mathcal{F}^m_{reg}}$ a pro $i, j = 1, \dots, m$ existují

$$f_{ij}''(\mathbf{y};\boldsymbol{\theta}) = \frac{\partial^2 f(\mathbf{y};\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j},$$

pak

$$\boxed{E\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}} \qquad a \qquad \boxed{D\mathbf{U}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta})}.$$

(2) Platí-li navíc pro $i, j = 1, \ldots, m$

$$\boxed{E\left(\frac{f''_{ij}(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})}\right) = 0},$$

pak

$$\boxed{\mathbf{J}(\boldsymbol{\theta}) = -E(\mathbf{U}'(\boldsymbol{\theta})),}$$

kde

$$\mathbf{U}'(\boldsymbol{\theta}) = \left(\frac{\partial U_i(\boldsymbol{\theta})}{\partial \theta_j}\right)_{i,j=1}^m.$$

Důkaz.

(1) z podmínky (4) pro regulární systém hustot a z definice vektoru $\mathbf{U}$ plyne

$$E\mathbf{U} = \mathbf{0}.$$

Pak

$$\boxed{D\mathbf{U}} = \left(E(U_iU_j) - \underbrace{EU_i}_{0}\underbrace{EU_j}_{0}\right)_{i,j=1}^m$$
$$= \left(E\left(\frac{\partial \ln f(\mathbf{Y};\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \ln f(\mathbf{Y};\boldsymbol{\theta})}{\partial \theta_j}\right)\right)_{i,j=1}^m = \boxed{\mathbf{J}(\boldsymbol{\theta})}$$

(2) Protože

$$U_i(\boldsymbol{\theta}) = \frac{\partial \ln f(\mathbf{Y};\boldsymbol{\theta})}{\partial \theta_i} = \frac{f'_i(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})},$$

pak

$$\frac{\partial U_i(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\frac{f'_i(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})} = \frac{f''_{ij}(\mathbf{Y};\boldsymbol{\theta})f(\mathbf{Y};\boldsymbol{\theta}) - f'_i(\mathbf{Y};\boldsymbol{\theta})f'_j(\mathbf{Y};\boldsymbol{\theta})}{[f(\mathbf{Y};\boldsymbol{\theta})]^2}$$
$$= \frac{f''_{ij}(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})} - \underbrace{\frac{f'_i(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})}}_{=U_i}\underbrace{\frac{f'_j(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})}}_{=U_j}$$

a odtud

$$U_iU_j = \frac{f''_{ij}(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})} - \frac{\partial U_i(\boldsymbol{\theta})}{\partial \theta_j}.$$

Protože pro $i, j = 1, \ldots, m$

$$J_{ij}(\boldsymbol{\theta}) = E(U_iU_j),$$

pak

$$\boxed{\mathbf{J}(\boldsymbol{\theta})} = E\left(\frac{f''_{ij}(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})} - \frac{\partial U_i(\boldsymbol{\theta})}{\partial \theta_j}\right)_{i,j=1}^{m}$$

$$= \left(\underbrace{E\frac{f''_{ij}(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})}}_{=0} - E\frac{\partial U_i(\boldsymbol{\theta})}{\partial \theta_j}\right)_{i,j=1}^{m} = \boxed{-E(\mathbf{U}'(\boldsymbol{\theta}))}. \qquad \square$$

<u>Poznámka</u> 1.4.7.
V některých publikacích se hustota splňující podmínky $\boxed{(1)\text{ až }(5)}$ v definici 1.4.1 nazývá <u>**regulární v 1. derivacích**</u>, pokud navíc splňuje podmínku

$$\boxed{E\left(\frac{f''_{ij}(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})}\right) = 0},$$

nazývá se <u>**regulární v 2. derivacích**</u>.

<u>Poznámka</u> 1.4.8.
Podmínka $E\left(\frac{f''_{ij}(\mathbf{Y};\boldsymbol{\theta})}{f(\mathbf{Y};\boldsymbol{\theta})}\right) = 0$ souvisí s otázkou, zda při derivování rovnosti

$$0 = EU_i(\boldsymbol{\theta}) = \int_M \frac{f'_i(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta})$$

lze <u>zaměnit pořadí derivace a integrálu</u>, tj. pro $i, j = 1, \ldots, m$

$$\boxed{0} = \frac{\partial}{\partial \theta_j} EU_i(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j}\int_M \frac{f'_i(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta}) \overset{?}{=} \underbrace{\int_M \frac{\partial}{\partial \theta_j}\frac{f'_i(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta}) = \boxed{0}}_{(*)}.$$

Pokud platí $(*)$, pak pořadí lze zaměnit. Uvažujme nejprve spojitý případ

$$\boxed{0} = \int_M \frac{\partial}{\partial \theta_j}\frac{f'_i(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta}) = \int_M \frac{\partial}{\partial \theta_j}\frac{f'_i(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} f(\mathbf{y};\boldsymbol{\theta}) d\mathbf{y} =$$

$$= \int_M \frac{\partial}{\partial \theta_j} f'_i(\mathbf{y};\boldsymbol{\theta}) d\mathbf{y} = \boxed{\int_M f''_{ij}(\mathbf{y};\boldsymbol{\theta}) d\mathbf{y}} = \int_M f''_{ij}(\mathbf{y};\boldsymbol{\theta}) \frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} d\mathbf{y}$$

$$= \boxed{\int_M \frac{f''_{ij}(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta})}.$$

Zcela analogicky pro diskrétní případ máme

$$\boxed{0} = \int_M \frac{\partial}{\partial \theta_j}\frac{f'_i(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta}) = \sum_{\mathbf{y}\in M} \frac{\partial}{\partial \theta_j}\frac{f'_i(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} f(\mathbf{y};\boldsymbol{\theta})$$

$$= \sum_{\mathbf{y}\in M} \frac{\partial}{\partial \theta_j} f'_i(\mathbf{y};\boldsymbol{\theta}) = \boxed{\sum_{\mathbf{y}\in M} f''_{ij}(\mathbf{y};\boldsymbol{\theta})} = \sum_{\mathbf{y}\in M} f''_{ij}(\mathbf{y};\boldsymbol{\theta}) \frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}$$

$$= \boxed{\int_M \frac{f''_{ij}(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} dF(\mathbf{y};\boldsymbol{\theta})}.$$

Příklad 1.4.9.
NORMÁLNÍ ROZDĚLENÍ. Mějme náhodnou veličinu Y s normálním rozdělením

$$\boxed{Y \sim N(\mu, \sigma^2)} \sim f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2},$$

přičemž:

(a) $\boxed{\sigma^2 \text{ je } \textbf{známé}, \text{ tj. } \theta_1 = \mu.}$

Pak **hustota** $f(y)$ **je regulární v prvních derivacích** (viz body (1) až (5)) a také **je regulární v druhých derivacích** (viz body (1) až (6))

(1) Množina $\mathbf{\Theta}_1 = (-\infty, \infty)$ je neprázdná otevřená množina.

(2) Množina $M = \{y \in \mathbb{R} : f(y) > 0\}$ je $(-\infty, \infty)$ a nezávisí na $\mu \in \mathbf{\Theta}_1$.

(3) Pro každé $\mathbf{y} \in M$ existuje konečná derivace

$$f'_\mu(y) = \frac{d\, f(y)}{d\, \mu} = f(y)\frac{y-\mu}{\sigma^2} \quad \Rightarrow \quad \boxed{U_1 = \frac{Y-\mu}{\sigma^2}}.$$

(4) Pro všechna $\mu \in \mathbf{\Theta}_1$ platí

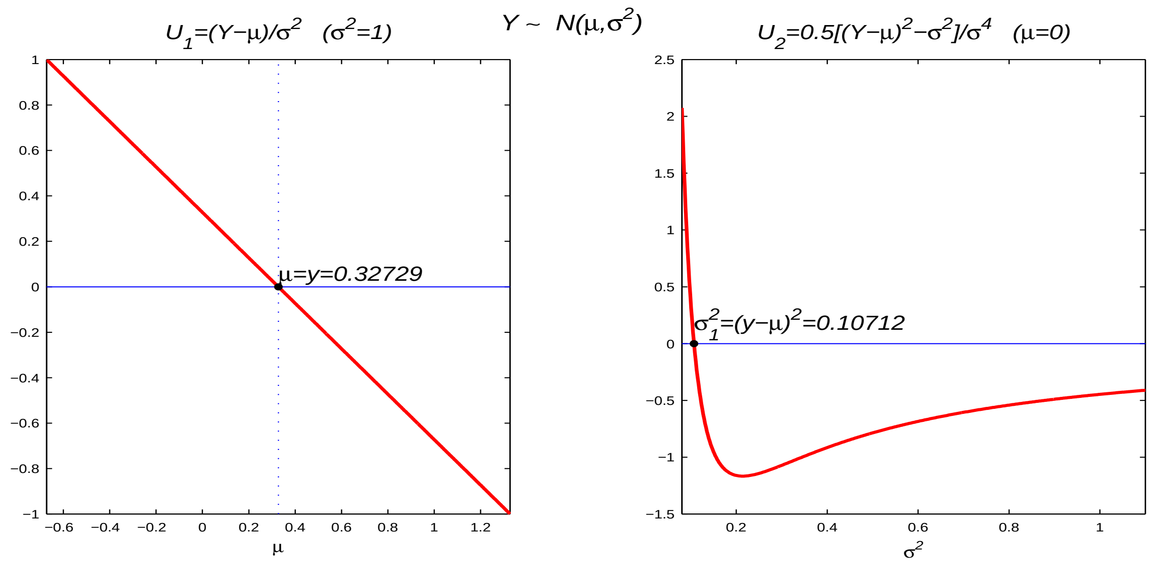
$$\boxed{EU_1} = \int_{-\infty}^{\infty} \frac{f'_\mu(y)}{f(y)} f(y)dy = \int_{-\infty}^{\infty} f'_\mu(y)dy = \frac{1}{\sigma^2} \underbrace{\int_{-\infty}^{\infty} (y-\mu)f(y)dy}_{0} = \boxed{0}.$$

(5) Pro všechna $\mu \in \mathbb{R}$ je integrál J_{11} konečný a kladný

$$\begin{aligned}
\boxed{J_{11}} = EU_1^2 &= \int_{-\infty}^{\infty} \left(\frac{f'_\mu(y)}{f(y)}\right)^2 f(y)dy = \int_{-\infty}^{\infty} \frac{\left(f'_\mu(y)\right)^2}{f(y)} dy \\
&= \frac{1}{\sigma^4} \underbrace{\int_{-\infty}^{\infty} (y-\mu)^2 f(y)dy}_{DY=\sigma^2} \\
&= \boxed{\frac{1}{\sigma^2}} > 0.
\end{aligned}$$

(6) Protože $f''_{\mu\mu}(y) = \frac{d}{d\,\mu}\left(\frac{y-\mu}{\sigma^2} f(y)\right) = \frac{(y-\mu)^2-\sigma^2}{\sigma^4} f(y)$, pak

$$\boxed{E\left(\frac{f''_{\mu\mu}(Y)}{f(Y)}\right)} = E\left(\frac{(Y-\mu)^2-\sigma^2}{\sigma^4}\right) = \frac{1}{\sigma^4} \underbrace{E(Y-\mu)^2}_{DY=\sigma^2} - \frac{1}{\sigma^2} = \boxed{0}.$$



Obrázek 1.4: Ukázky skórových funkcí U_1 (resp. U_2) pro $N(\mu,\sigma^2)$ při známém σ^2 (resp. μ).

(b) $\boxed{\mu \text{ je \textbf{známé}, tj. } \theta_2=\sigma^2.}$

Pak **hustota** $f(y)$ **je regulární v prvních derivacích** (viz body (1) až (5)) a také **je regulární v druhých derivacích** (viz body (1) až (6))

(1) Množina $\mathbf{\Theta}_2=(0,\infty)$ je neprázdná otevřená množina.

(2) Množina $M=\{y\in\mathbb{R}: f(y)>0\}$ je $(-\infty,\infty)$ a nezávisí na $\sigma^2\in\mathbf{\Theta}_2$.

(3) Pro každé $\mathbf{y}\in M$ existuje konečná derivace

$$f'_{\sigma^2}(y)=\frac{d\,f(y)}{d\,\sigma^2}=f(y)\frac{(y-\mu)^2-\sigma^2}{2\sigma^4}\quad\Rightarrow\quad\boxed{U_2=\frac{(Y-\mu)^2-\sigma^2}{2\sigma^4}}.$$

(4) Pro všechna $\sigma^2\in\mathbf{\Theta}_2$ platí

$$\boxed{EU_2}=\int_{-\infty}^{\infty}\frac{f'_{\sigma^2}(y)}{f(y)}f(y)dy=\int_{-\infty}^{\infty}f'_{\sigma^2}(y)dy=\int_{-\infty}^{\infty}f(y)\frac{(y-\mu)^2-\sigma^2}{2\sigma^4}dy=\boxed{0}.$$

(5) Pro všechna $\sigma^2\in\mathbf{\Theta}_2$ je integrál J_{22} konečný a kladný

$$\begin{aligned}\boxed{J_{22}}&=EU_2^2=\int_{-\infty}^{\infty}\left(\frac{f'_{\sigma^2}(y)}{f(y)}\right)^2f(y)dy=\frac{1}{4\sigma^8}\int_{-\infty}^{\infty}\left[(y-\mu)^2-\sigma^2\right]^2f(y)dy\\&=\frac{1}{4\sigma^8}\underbrace{\int_{-\infty}^{\infty}(y-\mu)^4f(y)dy}_{\mu_4=3\sigma^4}-\frac{2\sigma^2}{4\sigma^8}\underbrace{\int_{-\infty}^{\infty}(y-\mu)^2f(y)dy}_{\sigma^2}+\frac{\sigma^4}{4\sigma^8}\underbrace{\int_{-\infty}^{\infty}f(y)dy}_{1}\\&=\boxed{\frac{1}{2\sigma^4}}>0\end{aligned}$$

(6) Protože $f''_{\sigma^2\sigma^2}(y)=\frac{d}{d\,\sigma^2}\left(f(y)\frac{(y-\mu)^2-\sigma^2}{2\sigma^4}\right)=\frac{(y-\mu)^4-6\sigma^2(y-\mu)^2+3\sigma^4}{4\sigma^8}f(y)$, pak

$$\boxed{E\left(\frac{f''_{\sigma^2\sigma^2}(Y)}{f(Y)}\right)}=\frac{1}{4\sigma^8}\left[\underbrace{E(Y-\mu)^4}_{\mu_4=3\sigma^4}-6\sigma^2\underbrace{E(Y-\mu)^2}_{DY=\sigma^2}+3\sigma^4\right]=\boxed{0}.$$

(c) $\boxed{\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}} = (\mu, \sigma^2)^{\mathrm{T}}}$.

Pak **hustota** $f(y)$ **je regulární v prvních derivacích** (viz body (1) až (5)) a také **je regulární v druhých derivacích** (viz body (1) až (6))

(1) Množina $\boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2 = (-\infty, \infty) \times (0, \infty)$ je neprázdná otevřená množina.

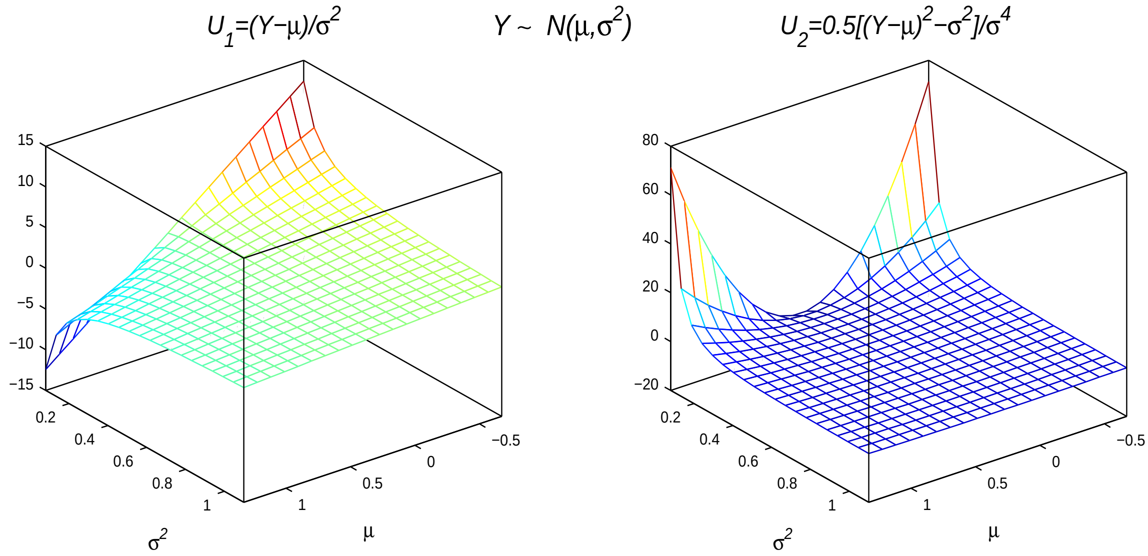
(2) Množina $M = \{y \in \mathbb{R} : f(y) > 0\}$ je $(-\infty, \infty)$ a nezávisí na $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(3) Pro každé $\mathbf{y} \in M$ existují konečné derivace $f'_\mu(y), f'_{\sigma^2}(y)$ (viz předchozí dva příklady).

(4) Pro všechna $(\mu, \sigma^2)^{\mathrm{T}} \in \boldsymbol{\Theta}$ platí $EU_1 = EU_2 = 0$ a

$$\boxed{\mathbf{U} = \left(\frac{Y-\mu}{\sigma^2}, \frac{(Y-\mu)^2-\sigma^2}{2\sigma^4}\right)^{\mathrm{T}}}$$

(viz předchozí dva příklady).



Obrázek 1.5: Ukázky skórových funkcí U_1 a U_2 pro $N(\mu, \sigma^2)$ při neznámém σ^2 a μ.

(5) Pro všechna $(\mu, \sigma^2)^{\mathrm{T}} \in \boldsymbol{\Theta}$ jsou integrály J_{11}, J_{22} a $J_{12} = J_{21}$ konečné, přičemž

$$\begin{aligned}\boxed{J_{12}} &= \int_{-\infty}^{\infty} \frac{f'_\mu(y)}{f(y)} \frac{f'_{\sigma^2}(y)}{f(y)} f(y)dy \\ &= \frac{1}{2\sigma^6} \int_{-\infty}^{\infty} (y-\mu)\left[(y-\mu)^2 - \sigma^2\right] f(y)dy \\ &= \underbrace{\frac{1}{2\sigma^6} \int_{-\infty}^{\infty} (y-\mu)^3 f(y)dy}_{\mu_3=0} - \underbrace{\frac{1}{2\sigma^4} \int_{-\infty}^{\infty} (y-\mu) f(y)dy}_{0} = 0\end{aligned}$$

a Fisherova informační matice

$$\boxed{\mathbf{J}(\mu, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}} > 0$$

je pozitivně definitní.

(6) Protože

$$E\left(\frac{f''_{\mu\mu}(Y)}{f(Y)}\right) = E\left(\frac{f''_{\sigma^2\sigma^2}(Y)}{f(Y)}\right) = 0,$$

(viz předchozí dva příklady), dále

$$f''_{\mu\sigma^2}(y) = f''_{\sigma^2\mu}(y) = f(y)\frac{(y-\mu)^3 - 3\sigma^2(y-\mu)}{2\sigma^6},$$

pak také

$$\boxed{E\left(\frac{f''_{\mu\sigma^2}(Y)}{f(Y)}\right)} = \frac{1}{2\sigma^6}\left[\underbrace{E(Y-\mu)^3}_{0} - 3\sigma^2 \underbrace{E(Y-\mu)}_{0}\right] = 0 = \boxed{E\left(\frac{f''_{\sigma^2\mu}(Y)}{f(Y)}\right)}.$$

<u>Příklad</u> 1.4.10.
Weibullovo 3-parametrické exponenciální rozdělení $Wb(\gamma,\theta,\delta)$. Mějme náhodnou veličinu Y s hustotou

$$f(y;\gamma,\theta,\delta) = \begin{cases} \frac{\gamma}{\delta}\left(\frac{y-\theta}{\delta}\right)^{\gamma-1} e^{-\left(\frac{y-\theta}{\delta}\right)^{\gamma}} & y > \theta,\ \theta\in\mathbb{R},\ \gamma>0,\ \delta>0 \\ 0 & \text{jinak.} \end{cases}$$

Zřejmě nejde o regulární systém hustot, neboť množina M je závislá na parametru θ.

Skórový vektor a informační matice pro náhodné výběry

V dalším budeme uvažovat **náhodný výběr** $\mathbf{Y}_n = (Y_1,\dots,Y_n)^{\mathrm{T}}$ z rozdělení s regulární hustotou $\boxed{f\in\mathcal{F}^m_{\text{reg}}}$. Označme $M = \{y\in\mathbb{R} : f(y;\boldsymbol{\theta}) > 0\}$.

Pak sdružená (simultánní) hustota náhodného vektoru $\mathbf{Y}_n = (Y_1,\dots,Y_n)^{\mathrm{T}}$ je rovna

$$\boxed{f_{\mathbf{Y}_n}(\mathbf{y};\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i,\boldsymbol{\theta})}, \qquad \mathbf{y} = (y_1,\dots,y_n)' \in \mathbb{R}^n$$

neboť náhodný výběr je tvořen systémem nezávislých náhodných veličin.

Zaveďme následující **ZNAČENÍ** pro

<u>funkce</u>:

$$\begin{aligned} l_k &= l(\boldsymbol{\theta};y_k) = \ln f(y_k;\boldsymbol{\theta}) \\ l^*_n &= l^*_n(\boldsymbol{\theta};\mathbf{y}) = \ln f_{\mathbf{Y}_n}(\mathbf{y};\boldsymbol{\theta}) \end{aligned}$$

<u>náhodné vektory</u>:

$$\begin{aligned} \mathbf{U}_k &= \mathbf{U}_k(\boldsymbol{\theta}) = (U_{1,k}(\boldsymbol{\theta}),\dots,U_{m,k}(\boldsymbol{\theta}))^{\mathrm{T}} = \left(\frac{\partial \ln f(Y_k;\boldsymbol{\theta})}{\partial\theta_1},\dots,\frac{\partial \ln f(Y_k;\boldsymbol{\theta})}{\partial\theta_m}\right)^{\mathrm{T}} \\ \mathbf{U}^*_n &= \mathbf{U}^*_n(\boldsymbol{\theta}) = (U^*_1(\boldsymbol{\theta}),\dots,U^*_m(\boldsymbol{\theta}))^{\mathrm{T}} = \left(\frac{\partial \ln f_{\mathbf{Y}_n}(\mathbf{Y};\boldsymbol{\theta})}{\partial\theta_1},\dots,\frac{\partial \ln f_{\mathbf{Y}_n}(\mathbf{Y};\boldsymbol{\theta})}{\partial\theta_m}\right)^{\mathrm{T}} \end{aligned}$$

<u>maticové funkce</u>:

$$\begin{aligned} \mathbf{J} &= \mathbf{J}(\boldsymbol{\theta}) = (J_{ij}(\boldsymbol{\theta}))_{i,j=1}^m = \left(\int_M \frac{\partial \ln f(y;\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial \ln f(y;\boldsymbol{\theta})}{\partial\theta_j} dF(y;\boldsymbol{\theta})\right)_{i,j=1}^m \\ \mathbf{J}_n &= \mathbf{J}_n(\boldsymbol{\theta}) = \left(\int_M\cdots\int_M \frac{\partial \ln f_{\mathbf{Y}_n}(\mathbf{y};\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial \ln f_{\mathbf{Y}_n}(\mathbf{y};\boldsymbol{\theta})}{\partial\theta_j} dF_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta})\right)_{i,j=1}^m \end{aligned}$$

kde $n\in\mathbb{N},\ k=1,\dots,n$.

Lemma 1.4.11.
Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ z rozdělení s hustotou $f \in \mathcal{F}^m_{reg}$. Pak platí

(1) Skórový vektor náhodného výběru $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ je roven

$$\boxed{\mathbf{U}^*_n(\boldsymbol{\theta}) = \sum_{k=1}^{n} \mathbf{U}_k(\boldsymbol{\theta})}$$

(2) Fisherova informační matice náhodného výběru $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ je rovna

$$\boxed{\mathbf{J}_n(\boldsymbol{\theta}) = n\mathbf{J}(\boldsymbol{\theta}).}$$

Důkaz.

(1) Pro j-tý prvek skórového vektoru $\mathbf{U}^*_n(\boldsymbol{\theta})$ náhodného výběru $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ platí

$$\begin{aligned}\boxed{U^*_j(\boldsymbol{\theta})} &= \frac{\partial \ln f_{\mathbf{Y}_n}(\mathbf{Y}; \boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial \ln \prod_{k=1}^{n} f(Y_k, \boldsymbol{\theta})}{\partial \theta_j} \\ &= \sum_{k=1}^{n} \frac{\partial \ln f(Y_k, \boldsymbol{\theta})}{\partial \theta_j} = \boxed{\sum_{k=1}^{n} U_{j,k}(\boldsymbol{\theta}))}.\end{aligned}$$ □

(2) Platí

$$\boxed{D\mathbf{U}^*_n(\boldsymbol{\theta})} = D\left(\sum_{k=1}^{n} \mathbf{U}_k(\boldsymbol{\theta})\right) \overset{nez.}{=} \sum_{k=1}^{n} D\mathbf{U}_k(\boldsymbol{\theta}) = \sum_{k=1}^{n} \mathbf{J}(\boldsymbol{\theta}) = \boxed{n\mathbf{J}(\boldsymbol{\theta})}.$$

Věta 1.4.12.
Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ z rozdělení s hustotou $f \in \mathcal{F}^m_{reg}$.

(1) Pokud pro $i, j = 1, \ldots, m$ existují

$$f''_{ij}(y; \boldsymbol{\theta}) = \frac{\partial^2 f(y; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j},$$

pak

$$\boxed{E\mathbf{U}^*_n(\boldsymbol{\theta}) = \mathbf{0}} \qquad a \qquad \boxed{D\mathbf{U}^*_n(\boldsymbol{\theta}) = n\mathbf{J}(\boldsymbol{\theta})}.$$

(2) Platí-li navíc pro $i, j = 1, \ldots, m$

$$\boxed{E\left(\frac{f''_{ij}(Y; \boldsymbol{\theta})}{f(Y; \boldsymbol{\theta})}\right) = 0},$$

(tj. f je regulární i v 2. derivacích), pak

$$\boxed{E(\mathbf{U}_n^{*'}(\boldsymbol{\theta})) = -n\mathbf{J}(\boldsymbol{\theta}),}$$

kde

$$\mathbf{U}_n^{*'}(\boldsymbol{\theta}) = \left(\frac{\partial U_i^*(\boldsymbol{\theta})}{\partial \theta_j}\right)_{i,j=1}^m .$$

DŮKAZ. Důkaz ihned plyne v věty 1.4.6 a z lemma 1.4.11. □

Následující věta uvádí **asymptotické vlastnosti skórových vektorů náhodných výběrů**.

Věta 1.4.13.
Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ z rozdělení s regulární hustotou $f \in \mathcal{F}_{reg}^m$. Označme $M = \{y \in \mathbb{R} : f(y; \boldsymbol{\theta}) > 0\}$. Nechť pro všechna $y \in M$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ a $i, j = 1, \ldots, m$ existují druhé parciální derivace hustoty $f(y; \boldsymbol{\theta})$.

(1) Pak platí

$$\boxed{\frac{1}{\sqrt{n}}\mathbf{U}_n^*(\boldsymbol{\theta}) \quad \overset{A}{\sim} \quad N_m(\mathbf{0}, \mathbf{J}(\boldsymbol{\theta}))}$$

nebo ekvivalentně

$$\boxed{\mathbf{U}_n^*(\boldsymbol{\theta}) \quad \overset{A}{\sim} \quad N_m(\mathbf{0}, \mathbf{J}_n(\boldsymbol{\theta})).}$$

Dále platí

$$\boxed{\frac{1}{n}\mathbf{U}_n^*(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{J}(\boldsymbol{\theta})^{-1}\mathbf{U}_n^*(\boldsymbol{\theta}) \quad \overset{A}{\sim} \quad \chi^2(m)}$$

nebo

$$\boxed{\mathbf{U}_n^*(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\theta})^{-1}\mathbf{U}_n^*(\boldsymbol{\theta}) \quad \overset{A}{\sim} \quad \chi^2(m).}$$

(2) Platí-li navíc, že f je regulární i v 2.derivacích, tj.

$$E\left(\frac{f_{ij}''(Y; \boldsymbol{\theta})}{f(Y; \boldsymbol{\theta})}\right) = 0,$$

pak matice náhodných veličin

$$\boxed{\frac{1}{n}\mathbf{U}_n^{*'}(\boldsymbol{\theta}) = \frac{1}{n}\left(\frac{\partial U_i^*(\boldsymbol{\theta})}{\partial \theta_j}\right)_{i,j=1}^m = \frac{1}{n}\left(\frac{\partial^2 l_n(\boldsymbol{\theta}; \mathbf{Y})}{\partial \theta_i \partial \theta_j}\right)_{i,j=1}^m \quad \xrightarrow{s.j.} \quad -\mathbf{J}(\boldsymbol{\theta}),}$$

nebo ekvivalentně

$$\boxed{\mathbf{U}_n^{*'}(\boldsymbol{\theta}) \quad \xrightarrow{s.j.} \quad -\mathbf{J}_n(\boldsymbol{\theta}).}$$

Důkaz.

(1) Složky skórového vektoru příslušného k náhodnému výběru lze podle lemma 1.4.11 vyjádřit jako součet nezávislých stejně rozdělených náhodných veličin

$$U_i^*(\boldsymbol{\theta}) = \frac{\partial \ln f_{\mathbf{Y}_n}(\mathbf{Y};\boldsymbol{\theta})}{\partial \theta_i} = \sum_{k=1}^{n} U_{i,k}(\boldsymbol{\theta}), \qquad i = 1,\ldots,m.$$

Podle věty 1.4.12 pak $E\mathbf{U}_n^*(\boldsymbol{\theta}) = \mathbf{0}$ a $D\mathbf{U}_n^*(\boldsymbol{\theta}) = n\mathbf{J}(\boldsymbol{\theta})$. První tvrzení pak plyne z vícerozměrné centrální limitní věty a druhé z vlastností kvadratických forem náhodných vektorů s normálním rozdělením.

(2) Náhodné veličiny

$$\frac{1}{n}U_{ij}^{*'}(\boldsymbol{\theta}) = \frac{\partial\left(\frac{1}{n}U_i^*(\boldsymbol{\theta})\right)}{\partial\theta_j}$$

jsou podle lemma 1.4.11 součtem nezávislých náhodných veličin se stejným rozdělením pravděpodobností, proto

$$\begin{aligned}\frac{1}{n}U_{ij}^{*'}(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{k=1}^{n}\frac{\partial U_{i,k}(\boldsymbol{\theta})}{\partial\theta_j}\\ &= \frac{1}{n}\sum_{k=1}^{n}U'_{ij,k}(\boldsymbol{\theta})\\ &= \frac{1}{n}\sum_{k=1}^{n}\frac{\partial^2 \ln f(Y_k,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} = \frac{1}{n}\frac{\partial^2 l_n(\boldsymbol{\theta};\mathbf{Y})}{\partial\theta_i\partial\theta_j}.\end{aligned}$$ □

Protože pro tyto náhodné veličiny platí silný zákon velkých čísel, tj.

$$\lim_{n\to\infty} P[U_{ij}^{*'}(\boldsymbol{\theta}) = EU_{ij}^{*'}(\boldsymbol{\theta})] = 1,$$

a podle věty 1.4.12

$$E(\mathbf{U}_n^{*'}(\boldsymbol{\theta})) = -n\mathbf{J}(\boldsymbol{\theta}),$$

tak $\frac{1}{n}\mathbf{U}_n^{*'}(\boldsymbol{\theta})$ konverguje skoro jistě k $-\mathbf{J}(\boldsymbol{\theta})$.

Definice 1.4.14.
Mějme náhodný výběr $\mathbf{Y}_n = (Y_1,\ldots,Y_n)^{\mathrm{T}}$ z rozdělení s hustotou $f(\mathbf{y};\boldsymbol{\theta})$. Nechť pro každé $n \in \mathbb{N}$ je definován odhad $T_n(\mathbf{Y})$ skalární parametrické funkce $\tau(\boldsymbol{\theta})$. Řekneme, že $T_n(\mathbf{Y})$ je ***REGULÁRNÍ ODHAD*** *parametrické funkce $\tau(\boldsymbol{\theta})$, pokud jsou splněny následující tři podmínky*

(i) $f \in \mathcal{F}_{reg}^m$,

(ii) $T_n(\mathbf{Y})$ *je* ***nestranným*** *odhadem* $\tau(\boldsymbol{\theta})$,

(iii) pro $j = 1,\ldots,m$ existují: $\tau_j'(\boldsymbol{\theta}) = \frac{\partial\tau(\boldsymbol{\theta})}{\partial\theta_j}$ a platí $\boxed{\tau_j'(\boldsymbol{\theta}) = \int_M T_n(\mathbf{y})\frac{f_j'(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta})}$ *tj.*

$$\boxed{E[\mathbf{U}_n^* T_n(\mathbf{Y})] = C(\mathbf{U}_n^*, T_n(\mathbf{Y})) = (\tau_1'(\boldsymbol{\theta}),\ldots,\tau_m'(\boldsymbol{\theta}))^{\mathrm{T}} = \tau'(\boldsymbol{\theta}).}$$

Poznámka 1.4.15.
Podmínka (iii) souvisí s otázkou, zda při derivování rovnosti

$$\tau(\boldsymbol{\theta}) = ET_n(\mathbf{Y}) = \int_M T_n(\mathbf{y})dF(\mathbf{y};\boldsymbol{\theta})$$

lze zaměnit pořadí derivace a integrálu, tj. pro $j = 1, \dots, m$

$$\boxed{\tau_j'(\boldsymbol{\theta})} = \frac{\partial}{\partial\theta_j}ET_n(\mathbf{Y}) = \frac{\partial}{\partial\theta_j}\int_M T_n(\mathbf{y})dF(\mathbf{y};\boldsymbol{\theta}) \overset{?}{=} \underbrace{\int_M \frac{\partial}{\partial\theta_j}T_n(\mathbf{y})dF(\mathbf{y};\boldsymbol{\theta}) = \boxed{\tau_j'(\boldsymbol{\theta})}}_{(*)}.$$

Pokud platí $(*)$, pak pořadí lze zaměnit. Uvažujme nejprve spojitý případ

$$\begin{aligned}\boxed{\tau_j'(\boldsymbol{\theta})} &= \int_M T_n(\mathbf{y})\frac{\partial}{\partial\theta_j}\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta}) = \int_M T_n(\mathbf{y})\frac{\partial}{\partial\theta_j}\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}f(\mathbf{y};\boldsymbol{\theta})d\mathbf{y}\\ &= \int_M T_n(\mathbf{y})\frac{\partial}{\partial\theta_j}f(\mathbf{y};\boldsymbol{\theta})d\mathbf{y} = \boxed{\int_M T_n(\mathbf{y})f_j'(\mathbf{y};\boldsymbol{\theta})d\mathbf{y}}\\ &= \int_M T_n(\mathbf{y})f_j'(\mathbf{y};\boldsymbol{\theta})\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}d\mathbf{y} = \boxed{\int_M T_n(\mathbf{y})\frac{f_j'(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta})}.\end{aligned}$$

Zcela analogicky pro diskrétní případ máme

$$\begin{aligned}\boxed{\tau_j'(\boldsymbol{\theta})} &= \int_M T_n(\mathbf{y})\frac{\partial}{\partial\theta_j}\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta}) = \sum_{\mathbf{y}\in M} T_n(\mathbf{y})\frac{\partial}{\partial\theta_j}\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}f(\mathbf{y};\boldsymbol{\theta})\\ &= \sum_{\mathbf{y}\in M} T_n(\mathbf{y})\frac{\partial}{\partial\theta_j}f(\mathbf{y};\boldsymbol{\theta}) = \boxed{\sum_{\mathbf{y}\in M} T_n(\mathbf{y})f_j'(\mathbf{y};\boldsymbol{\theta})}\\ &= \sum_{\mathbf{y}\in M} f_j'(\mathbf{y};\boldsymbol{\theta})\frac{f(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})} = \boxed{\int_M T_n(\mathbf{y})\frac{f_j'(\mathbf{y};\boldsymbol{\theta})}{f(\mathbf{y};\boldsymbol{\theta})}dF(\mathbf{y};\boldsymbol{\theta})}.\end{aligned}$$

Věta 1.4.16 (Raova-Cramerova nerovnost).
Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \dots, Y_n)^{\mathrm{T}}$ z rozdělení s hustotou $f \in \mathcal{F}^m_{reg}$, dále pro $i, j = 1, \dots, m$ nechť existují

$$f''_{ij}(\mathbf{y};\boldsymbol{\theta}) = \frac{\partial^2 f(\mathbf{y};\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}$$

a $T_n(\mathbf{Y})$ je regulární odhad $\tau(\boldsymbol{\theta})$. Pak pro každé $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ platí

$$\boxed{DT_n(\mathbf{Y}) \geq [\tau'(\boldsymbol{\theta})]^T[\mathbf{J}_n(\boldsymbol{\theta})]^{-1}\tau'(\boldsymbol{\theta}) = C_n(\boldsymbol{\theta})}$$

a $C_n(\boldsymbol{\theta})$ se nazývá ***DOLNÍ RAOVA-CRAMEROVA HRANICE****.*

Důkaz. Pro jednoduchost předpokládejme, že $\boldsymbol{\theta} = \theta$. Z regularity $T_n(\mathbf{Y})$ plyne

$$|\tau'(\theta)| = |E[\mathbf{U}_n^* T_n(\mathbf{Y})]| = |\underbrace{C(\mathbf{U}_n^*(\theta), T_n(\mathbf{Y}))}_{viz E\mathbf{U}_n^*=0}| \overset{Schwarz.ner.}{\leq} \sqrt{DT_n(\mathbf{Y})}\underbrace{\sqrt{D\mathbf{U}_n^*(\theta)}}_{=\sqrt{\mathbf{J}_n(\theta)}}.$$

tj.

$$(\tau'(\theta))^2 \leq DT_n(\mathbf{Y})\mathbf{J}_n(\theta)$$

a

$$\frac{(\tau'(\theta))^2}{\mathbf{J}_n(\theta)} \leq DT_n(\mathbf{Y}), \tag{1.2}$$

čímž je tvrzení dokázáno. □

<u>Poznámka</u> 1.4.17.

HUSTOTY EXPONENCIÁLNÍHO TYPU

Předpokládejme, že funkce $\tau(\theta)$ není rovna konstantě. Rovnosti v (1.2) bude dosaženo právě tehdy, bude-li platit rovnost

$$(\tau'(\theta))^2 = \underbrace{E\left[T_n(\mathbf{Y}) - \tau(\theta)\right]^2}_{=DT_n(\mathbf{Y})} \underbrace{E\left[\frac{f'(\mathbf{Y},\theta)}{f(\mathbf{Y},\theta)}\right]^2}_{=E[\mathbf{U}_n^*(\theta)]^2 = \mathbf{J}_n(\theta)}$$

Víme, že Schwarzova nerovnost přejde v rovnost právě tehdy, platí-li

(a) buď

$$T_n(\mathbf{y}) - \tau(\theta) = 0$$

skoro všude, což značí, že existuje přesný odhad T_n funkce $\tau(\theta)$, tj. takový odhad, který je roven $\tau(\theta)$ s pravděpodobností 1. Pak bychom měli

$$ET_n(\theta) = \tau(\theta)$$

a z Raovy-Cramérovy nerovnosti by plynulo

$$E\left[T_n(\mathbf{Y}) - \tau(\theta)\right]^2 \geq \frac{(\tau'(\theta))^2}{\mathbf{J}_n(\theta)}.$$

Je-li však $T_n(\theta) = \tau(\theta)$ skoro jistě, platí

$$E\left[T_n(\mathbf{Y}) - \tau(\theta)\right]^2 = 0,$$

což je spor s předchozí nerovností;

(b) nebo když náhodná veličina $U_n^*(\theta)$ je lineární transformací náhodné veličiny $T_n(\mathbf{Y})$, tj.

$$U_n^*(\theta) = \frac{f'(\mathbf{Y},\theta)}{f(\mathbf{Y},\theta)} = K(\theta)T_n(\mathbf{Y}) + B(\theta).$$

Protože platí (neboť f je regulární v prvních derivacích)

$$\boxed{0 = EU_n^*(\theta)} = E\left[K(\theta)T_n(\mathbf{Y}) + B(\theta)\right] = K(\theta)ET_n(\mathbf{Y}) + B(\theta) = K(\theta)\tau(\theta) + B(\theta)$$

$$\Rightarrow \boxed{B(\theta) = -K(\theta)\tau(\theta)} \qquad \Rightarrow \qquad \boxed{U_n^*(\theta) = K(\theta)\left[T_n(\mathbf{Y}) - \tau(\theta)\right]}$$

a existuje tedy taková funkce $K(\theta)$ nezávislá na $\mathbf{y}$, že platí

$$\frac{d\ln f(\mathbf{y},\theta)}{d\theta} = \frac{f'(\mathbf{y},\theta)}{f(\mathbf{y},\theta)} = K(\theta)\left[T_n(\mathbf{y}) - \tau(\theta)\right] \qquad \mathbf{y} \in M. \tag{1.3}$$

Označme $Q(\theta)$ a $R(\theta)$ funkce, pro které platí

$$Q'(\theta) = K(\theta) \qquad \text{a} \qquad R'(\theta) = \tau(\theta)K(\theta) \qquad \Rightarrow \qquad \boxed{\tau(\theta) = \frac{R'(\theta)}{Q'(\theta)}} = \boxed{ET_n(\mathbf{Y})}.$$

Pak integrováním ze vzorce (1.3) dostaneme

$$\ln f(\mathbf{y}, \theta) = Q(\theta)T_n(\mathbf{y}) - R(\theta) + H(\mathbf{y}),$$

kde $H(\mathbf{y})$ nezávisí na θ. Označme nyní

$$h(\mathbf{y}) = e^{H(\mathbf{y})}.$$

Pak máme

$$\boxed{f(\mathbf{y}, \theta) = \exp\{Q(\theta)T_n(\mathbf{y}) - R(\theta) + H(\mathbf{y})\}} = h(\mathbf{y}) \underbrace{\exp\{Q(\theta)T_n(\mathbf{y}) - R(\theta)\}}_{g(T_n(\mathbf{y}),\theta)}$$

a $\boxed{T_n(\mathbf{Y})}$ je **postačující statistikou pro parametr** $\boxed{\theta}$. Dostali jsme tzv. **hustotu exponenciálního typu** s jedním parametrem.

Všimněme si dále, za jakých podmínek je

- splněna podmínka $E\left(\frac{f''}{f}\right) = 0$, která zajišťuje, že hustota f je regulární i v druhých derivacích
- a platí vztah $\tau'(\theta) = E(U_n^*(\theta)T_n(\mathbf{Y}))$ z definice regulárního odhadu.

Počítejme nejprve

$$\begin{aligned}
f'(\mathbf{y}, \theta) &= f(\mathbf{y}, \theta) \left\{K(\theta) \left[T_n(\mathbf{y}) - \tau(\theta)\right]\right\} \\
f''(\mathbf{y}, \theta) &= f'(\mathbf{y}, \theta) \left\{K(\theta) \left[T_n(\mathbf{y}) - \tau(\theta)\right]\right\} + f(\mathbf{y}, \theta) \left\{K'(\theta) \left[T_n(\mathbf{y}) - \tau(\theta)\right] - K(\theta)\tau'(\theta)\right\} \\
\frac{f''(\mathbf{y}, \theta)}{f(\mathbf{y}, \theta)} &= K^2(\theta) \left[T_n(\mathbf{y}) - \tau(\theta)\right]^2 + K'(\theta) \left[T_n(\mathbf{y}) - \tau(\theta)\right] - K(\theta)\tau'(\theta) \\
\tau'(\theta) &= \frac{R''(\theta)Q'(\theta) - R'(\theta)Q''(\theta)}{[Q'(\theta)]^2}
\end{aligned}$$

Pak

- podmínka

$$\begin{aligned}
\boxed{0 = E\left(\frac{f''}{f}\right)} &= E\left\{K^2(\theta) \left[T_n(\mathbf{Y}) - \tau(\theta)\right]^2 + K'(\theta) \left[T_n(\mathbf{Y}) - \tau(\theta)\right] - K(\theta)\tau'(\theta)\right\} \\
&= K^2(\theta) \underbrace{E\left[T_n(\mathbf{Y}) - \tau(\theta)\right]^2}_{DT_n(\mathbf{Y})} + K'(\theta) \underbrace{E\left[T_n(\mathbf{Y}) - \tau(\theta)\right]}_{=0} - K(\theta)\tau'(\theta) \\
&\Rightarrow \quad \boxed{DT_n(\mathbf{Y})} = \frac{\tau'(\theta)}{K(\theta)} = \boxed{\frac{\tau'(\theta)}{Q'(\theta)}}
\end{aligned}$$

- a vztah

$$\boxed{\tau'(\theta) = E[U_n^*(\theta)T_n(\mathbf{Y})]} = E\{U_n^*(\theta)[T_n(\mathbf{Y}) - \tau(\theta)]\} = E\left\{Q'(\theta)\,[T_n(\mathbf{Y}) - \tau(\theta)]^2\right\}$$
$$= Q'(\theta)\underbrace{E\,[T_n(\mathbf{Y}) - \tau(\theta)]^2}_{DT_n(\mathbf{Y})}$$
$$\Rightarrow \quad \boxed{DT_n(\mathbf{Y})} = \frac{\tau'(\theta)}{K(\theta)} = \boxed{\frac{\tau'(\theta)}{Q'(\theta)}}.$$

Vidíme, že dostáváme stejný vztah

$$\boxed{DT_n(\mathbf{Y}) = \frac{\tau'(\theta)}{Q'(\theta)} = \frac{R''(\theta)Q'(\theta) - R'(\theta)Q''(\theta)}{[Q'(\theta)]^3}}.$$

Definice 1.4.18.
Řekneme, že regulární odhad $\boxed{T_n(\mathbf{Y})}$ *je*

(a) ***VYDATNÝM*** *(také EFICIENTNÍM) odhadem* $\tau(\boldsymbol{\theta})$, *pokud*
$$\boxed{\varepsilon_n[T_n(\mathbf{Y})] = \frac{C_n(\boldsymbol{\theta})}{DT_n(\mathbf{Y})} = 1}$$

(b) ***ASYMPTOTICKY VYDATNÝM*** *odhadem* $\tau(\boldsymbol{\theta})$, *pokud* $\boxed{\lim_{n\to\infty} \varepsilon_n[T_n(\mathbf{Y})] = 1}$

a číslo $\boxed{\varepsilon_n[T_n(\mathbf{Y})]}$ *se nazývá* ***vydatnost*** *(eficience)* ***odhadu*** $T_n(\mathbf{Y})$.

Věta 1.4.19.
Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ *z rozdělení s hustotou* $f \in \mathcal{F}^m_{reg}$. *Nechť pro* $i, j = 1, \ldots, m$ *existují*
$$f''_{ij}(\mathbf{y}; \boldsymbol{\theta}) = \frac{\partial^2 f(\mathbf{y}; \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}.$$
Mějme odhad $T_n(\mathbf{Y})$ *parametrické funkce* $\tau(\boldsymbol{\theta})$, *přičemž pro každé* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *a každé* $i = 1, \ldots, m$ *nechť existují*
$$\tau'_i(\boldsymbol{\theta}) = \frac{\partial \tau(\boldsymbol{\theta})}{\partial\theta_i}.$$
Pak odhad $\boxed{T_n(\mathbf{Y})}$ *parametrické funkce* $\tau(\boldsymbol{\theta})$ *je* ***vydatný*** *právě tehdy, když*
$$\boxed{P\left\{T_n(\mathbf{Y}) - \tau(\boldsymbol{\theta}) = [\tau'(\boldsymbol{\theta})]^{\mathrm{T}}[\mathbf{J}_n(\boldsymbol{\theta})]^{-1}\mathbf{U}(\boldsymbol{\theta})\right\} = 1.}$$

Důkaz. viz. např.:
Lamoš, F., Potocký R.: *Pravdepodobnosť a matematická štatistika*, ALFA, Bratislava, 1989.
Anděl, J.: *Matematická statistika*, SNTL Praha 1978 □

1.4.1 MAXIMÁLNĚ VĚROHODNÉ ODHADY

V dalším budeme uvažovat pouze regulární hustoty, tj. $f(\mathbf{y};\boldsymbol{\theta}) \in \mathcal{F}^m_{\text{reg}}$, $\mathbf{y} \in \mathbb{R}^n$.

Definice 1.4.20.

(a) ***VĚROHODNOSTNÍ FUNKCÍ*** *rozumíme funkci vektorového parametru* $\boldsymbol{\theta}$

$$\boxed{L(\boldsymbol{\theta};\mathbf{y}) = f(\mathbf{y};\boldsymbol{\theta})}$$

(b) ***LOGARITMICKOU VĚROHODNOSTNÍ FUNKCÍ*** *nazýváme funkci*

$$\boxed{l(\boldsymbol{\theta};\mathbf{y}) = \ln L(\boldsymbol{\theta};\mathbf{y})}$$

(c) Řekneme, že odhad $\widehat{\boldsymbol{\theta}}_{\text{MLE}} = \widehat{\boldsymbol{\theta}}_{\text{MLE}}(\mathbf{Y})$ *je* ***MAXIMÁLNĚ VĚROHODNÝ ODHAD (MLE)*** *vektorového parametru* $\boldsymbol{\theta}$, *pokud platí*

$$\boxed{L(\widehat{\boldsymbol{\theta}}_{\text{MLE}};\mathbf{Y}) \geq L(\boldsymbol{\theta};\mathbf{Y})}$$

pro všechna $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

<u>Poznámka</u> 1.4.21.
Z <u>regularity systému hustot</u> plyne existence prvních parciálních derivací věrohodnostní funkce podle složek $\boldsymbol{\theta}$. Dále předpokládejme, že existují i její druhé parciální derivace podle složek $\boldsymbol{\theta}$. Maximálně věrohodný odhad $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$ získáme řešením buď systému věrohodnostních rovnic

$$\boxed{L'(\boldsymbol{\theta};\mathbf{Y}) = \mathbf{0}} \quad \text{kde} \quad L'(\boldsymbol{\theta};\mathbf{Y}) = (L'_1(\boldsymbol{\theta}),\dots,L'_m(\boldsymbol{\theta}))' \quad \text{a} \quad L'_i(\boldsymbol{\theta}) = \frac{\partial L(\boldsymbol{\theta};\mathbf{Y})}{\partial \theta_i} \tag{1.4}$$

za podmínky, že matice $L''(\boldsymbol{\theta};\mathbf{Y}) = \left(L''_{ij}(\boldsymbol{\theta})\right)_{i,j=1}^m = \left(\frac{\partial^2 L(\boldsymbol{\theta};\mathbf{Y})}{\partial\theta_i\partial\theta_j}\right)_{i,j=1}^m$ je negativně definitní, nebo

$$\boxed{l'(\boldsymbol{\theta};\mathbf{Y}) = \mathbf{0}} \quad \text{kde} \quad l'(\boldsymbol{\theta};\mathbf{Y}) = (l'_1(\boldsymbol{\theta}),\dots,l'_m(\boldsymbol{\theta}))' \quad \text{a} \quad l'_i(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta};\mathbf{Y})}{\partial \theta_i} \tag{1.5}$$

za podmínky, že matice $l''(\boldsymbol{\theta};\mathbf{Y}) = \left(l''_{ij}(\boldsymbol{\theta})\right)_{i,j=1}^m = \left(\frac{\partial^2 l(\boldsymbol{\theta};\mathbf{Y})}{\partial\theta_i\partial\theta_j}\right)_{i,j=1}^m$ je negativně definitní,

neboť pak (1.4) bude platit právě tehdy, když bude platit i (1.5), protože pro všechna $\mathbf{y} \in M$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ a $i = 1,\dots,m$ platí

$$l'_i = \frac{\partial l(\boldsymbol{\theta};\mathbf{Y})}{\partial\theta_i} = \frac{L'_i(\boldsymbol{\theta};\mathbf{Y})}{L(\boldsymbol{\theta};\mathbf{Y})} \qquad \text{a} \qquad L(\boldsymbol{\theta};\mathbf{Y}) > 0.$$

V následující větě shrňme **základní vlastnosti maximálně věrohodných odhadů**.

Věta 1.4.22.
Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \dots, Y_n)^{\mathrm{T}}$ *z rozdělení s regulární hustotou* $f \in \mathcal{F}^m_{reg}$.

(1) *Nechť* $\boldsymbol{\tau} : \boldsymbol{\Theta} \mapsto \boldsymbol{\Omega}$ *je funkce zobrazující* $\boldsymbol{\Theta}$ *do* $\boldsymbol{\Omega} \subseteq \mathbb{R}^r$ $(1 \leq r \leq m)$. *Pak maximálně věrohodný odhad vektorové parametrické funkce* $\boldsymbol{\tau}(\boldsymbol{\theta})$ *je* $\boldsymbol{\tau}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}\right)$, *tj*

$$\boxed{\widehat{\boldsymbol{\tau}(\boldsymbol{\theta})}_{\mathrm{MLE}} = \boldsymbol{\tau}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}\right).}$$

(2) *Nechť existuje vektorová* ***postačující statistika*** $\mathbf{T}_n(\mathbf{Y}) = (T_1(\mathbf{Y}), \dots, T_m(\mathbf{Y}))^{\mathrm{T}}$ *parametru* $\boldsymbol{\theta}$. *Pak maximálně věrohodný odhad* $\boxed{\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$ *je* ***vektorovou funkcí postačující statistiky****:*

$$\begin{aligned}\boxed{\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}} &= \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta};\mathbf{y}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} f(\mathbf{y};\boldsymbol{\theta}) \\ &= \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \{h(\mathbf{y})\ g\left(\mathbf{T}_n(\mathbf{y},\boldsymbol{\theta})\right)\} = \boxed{h(\mathbf{y})\ \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} g\left(\mathbf{T}_n(\mathbf{y},\boldsymbol{\theta})\right)}.\end{aligned}$$

(3) *Mějme parametrickou funkci* $\tau(\boldsymbol{\theta})$ *a nechť pro každé* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *a každé* $i = 1, \dots, m$ *existují* $\tau_i'(\boldsymbol{\theta}) = \frac{\partial \tau(\boldsymbol{\theta})}{\partial \theta_i}$. *Existuje-li* ***vydatný*** *(eficientní)* ***odhad*** $T_n(\mathbf{Y})$ *parametrické funkce* $\tau(\boldsymbol{\theta})$, *pak je jejím maximálně věrohodným odhadem* $\boxed{\widehat{\boldsymbol{\tau}(\boldsymbol{\theta})}_{\mathrm{MLE}} = \mathbf{T}_n(\mathbf{Y}).}$

(4) *Platí* $\boxed{\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} \xrightarrow{s.j.} \boldsymbol{\theta}}$.

<u>*Poznámka*</u>*: Protože z konvergence skoro jistě plyne konvergence podle pravděpodobnosti, je* $\boxed{\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$ *také* ***konzistentním odhadem*** *parametru* $\boldsymbol{\theta}$.

(5) *Nechť pro* $i, j = 1, \dots, m$ *existují druhé parciální derivace* $\frac{\partial^2 f(y;\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}$, *pak* $\boxed{\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}}}$ *je* ***vydatným odhadem*** *parametru* $\boldsymbol{\theta}$.

DůKAZ. viz. např.:
Lamoš, F., Potocký R.: *Pravdepodobnosť a matematická štatistika*, ALFA, Bratislava, 1989.
Anděl, J.: *Matematická statistika*, SNTL Praha 1978 □

V poslední větě této části uveďme důležité vlastnosti, které se týkají **asymptotického rozdělení maximálně věrohodných odhadů**.

Věta 1.4.23.
Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \dots, Y_n)^{\mathrm{T}}$ *z rozdělení s regulární hustotou* $f \in \mathcal{F}^m_{reg}$. *Označme* $M = \{y \in \mathbb{R} : f(y;\boldsymbol{\theta}) > 0\}$. *Nechť pro všechna* $y \in M$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *a* $i, j = 1, \dots, m$ *existují druhé parciální derivace hustoty* $f(y;\boldsymbol{\theta})$ *a platí* $E\left(\frac{f''_{ij}(Y;\boldsymbol{\theta})}{f(Y;\boldsymbol{\theta})}\right) = 0$. *Pak*

(1) $\boxed{\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} \overset{A}{\sim} N_m(\boldsymbol{\theta}, \mathbf{J}_n(\boldsymbol{\theta})^{-1})}$ *nebo ekvivalentně* $\boxed{\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} - \boldsymbol{\theta}) \overset{A}{\sim} N_m(\mathbf{0}, \mathbf{J}(\boldsymbol{\theta})^{-1})}$

(2) $\boxed{W = (\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta})^{\mathrm{T}} \mathbf{J}_n(\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta}) \quad \overset{A}{\sim} \quad \chi^2(m)}$, *tzv.* ***Waldova statistika***.

DŮKAZ.

(1) Označme $\boldsymbol{\theta}_0$ skutečnou hodnotu neznámého parametru $\boldsymbol{\theta}$ a parciální derivace i-té složky skórového vektoru

$$\mathbf{u}_i^{*'}(\boldsymbol{\theta}) = \Big(U_{i1}^{*'}(\boldsymbol{\theta}), \ldots, U_{im}^{*'}(\boldsymbol{\theta})\Big)^{\mathrm{T}} = \left(\frac{\partial U_i^*(\boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial U_i^*(\boldsymbol{\theta})}{\partial \theta_m}\right)^{\mathrm{T}}, \qquad i = 1, \ldots, m,$$

což je vlastně i-tý řádek (díky symetrii též sloupec) matice druhých parciálních derivací funkce $\ln f_{\mathbf{Y}}(\mathbf{y}, \boldsymbol{\theta})$. Uvažujme pro $i = 1, \ldots, m$, Taylorův rozvoj i-té složky skórového vektoru

$$U_i^*(\boldsymbol{\theta}) = U_i^*(\boldsymbol{\theta}_0) + \mathbf{u}_i^{*'}(\boldsymbol{\theta}^*)^{\mathrm{T}} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \text{ přičemž } \boldsymbol{\theta}^* = t\boldsymbol{\theta}_0 + (1-t)\boldsymbol{\theta}, \qquad t \in (0,1).$$

Levá strana této rovnice je v bodě maximálně věrohodného odhadu $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$ rovna nule. Odtud plyne

$$U_i^*(\boldsymbol{\theta}_0) = -\mathbf{u}_i^{*'}(\boldsymbol{\theta}^*)^{\mathrm{T}} \left(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta}_0\right), \qquad i = 1, \ldots, m,$$

což lze maticově zapsat také takto

$$\mathbf{U}_n^*(\boldsymbol{\theta}_0) = -\mathbf{U}_n^{*'}(\boldsymbol{\theta}^*)(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta}_0),$$

kde nyní

$$\boldsymbol{\theta}^* = t\boldsymbol{\theta}_0 + (1-t)\widehat{\boldsymbol{\theta}}_{\text{MLE}}, \qquad t \in (0,1).$$

Pro matici náhodných veličin, kde sloupce jsou vektory $\mathbf{u}_i^{*'}(\boldsymbol{\theta})$ $(i = 1, \ldots, m)$

$$\mathbf{U}_n^{*'}(\boldsymbol{\theta}) = \left(\frac{\partial U_i^*(\boldsymbol{\theta})}{\partial \theta_j}\right)_{i,j=1}^m = \Big(\mathbf{u}_1^{*'}(\boldsymbol{\theta}), \ldots, \mathbf{u}_m^{*'}(\boldsymbol{\theta})\Big)$$

podle věty 1.4.13 platí

$$\lim_{n\to\infty} P\left\{\mathbf{U}_n^{*'}(\boldsymbol{\theta}_0) = -\mathbf{J}_n(\boldsymbol{\theta}_0)\right\} = 1.$$

Protože $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$ konverguje podle věty 1.4.22 skoro jistě ke skutečnému parametru $\boldsymbol{\theta}_0$, tj.

$$\widehat{\boldsymbol{\theta}}_{\text{MLE}} \quad \xrightarrow{s.j.} \quad \boldsymbol{\theta}_0$$

takže

$$\lim_{n\to\infty} P\left\{\mathbf{U}_n^{*'}(\widehat{\boldsymbol{\theta}}_{\text{MLE}}) = \mathbf{U}_n^{*'}(\boldsymbol{\theta}_0)\right\} = 1.$$

Protože $\boldsymbol{\theta}^*$ leží mezi $\boldsymbol{\theta}_0$ a $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$, pak konverguje i $\boldsymbol{\theta}^*$ skoro jistě k $\boldsymbol{\theta}_0$. tj.

$$\boldsymbol{\theta}^* \quad \xrightarrow{s.j.} \quad \boldsymbol{\theta}_0$$

a odtud ihned dostaneme, že také

$$\lim_{n\to\infty} P\left\{\mathbf{U}_n^{*'}(\boldsymbol{\theta}^*) = -\mathbf{J}_n(\boldsymbol{\theta}_0)\right\} = 1.$$

Tedy uvážíme-li maticový zápis Taylorova rozvoje v bodě $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$

$$\mathbf{U}_n^*(\boldsymbol{\theta}_0) = \underbrace{-\mathbf{U}_n^{*'}(\boldsymbol{\theta}^*)}_{\longrightarrow \mathbf{J}_n(\boldsymbol{\theta}_0)} \left(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta}_0\right)$$

Podle věty 1.4.13 má skórový vektor $\mathbf{U}_n^*(\boldsymbol{\theta}_0)$ asymptoticky normální rozdělení:

$$\tfrac{1}{\sqrt{n}}\mathbf{U}_n^*(\boldsymbol{\theta}_0) \quad \overset{A}{\sim} \quad N_m(\mathbf{0}, \mathbf{J}(\boldsymbol{\theta}_0)) \qquad \text{nebo-li} \qquad \mathbf{U}_n^*(\boldsymbol{\theta}_0) \quad \overset{A}{\sim} \quad N_m(\mathbf{0}, \mathbf{J}_n(\boldsymbol{\theta}_0)),$$

proto také $\boxed{\widehat{\boldsymbol{\theta}}_{\text{MLE}}}$ **má asymptoticky normální rozdělení**.

Při regularitě je střední hodnota rovna

$$E\mathbf{U}_n^*(\boldsymbol{\theta}_0) = 0$$

a pro $n \to \infty$

$$-E\mathbf{U}_n^{*'}(\boldsymbol{\theta}^*) \longrightarrow \mathbf{J}_n(\boldsymbol{\theta}_0) > 0,$$

proto $\boxed{E\widehat{\boldsymbol{\theta}}_{\text{MLE}} = \boldsymbol{\theta}_0}$.

Pro $n \to \infty$ počítejme varianční matici

$$D\mathbf{U}_n^*(\boldsymbol{\theta}_0) = \mathbf{J}_n(\boldsymbol{\theta}_0) = \mathbf{J}_n(\boldsymbol{\theta}_0) D\left(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta}_0\right) \mathbf{J}_n(\boldsymbol{\theta}_0)^{\mathrm{T}}$$

takže pro $n \to \infty$

$$D\left(\widehat{\boldsymbol{\theta}}_{\text{MLE}} - \boldsymbol{\theta}_0\right) = \boxed{D\widehat{\boldsymbol{\theta}}_{\text{MLE}}} = \mathbf{J}_n(\boldsymbol{\theta}_0)^{-1}\mathbf{J}_n(\boldsymbol{\theta}_0)\mathbf{J}_n(\boldsymbol{\theta}_0)^{-1} = \boxed{\mathbf{J}_n(\boldsymbol{\theta}_0)^{-1}}$$

(2) Druhé tvrzení plyne z vlastností kvadratických forem náhodných vektorů s normálním rozdělením. □

Příklad 1.4.24.

NORMÁLNÍ ROZDĚLENÍ.

Mějme náhodnou veličinu Y s normálním rozdělením

$$\boxed{Y \sim N(\mu, \sigma^2)} \sim f(y) = \tfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

a náhodný výběr $\mathbf{Y}_n = (Y_1, \dots, Y_n)^{\mathrm{T}}$ z téhož rozdělení, přičemž:

(a) $\boxed{\sigma^2 \text{ je } \textbf{známé}\text{, tj. } \theta_1 = \mu.}$

(1) Skórová funkce náhodného výběru (viz příklad 1.4.9): $\boxed{U_1^*(\mu) = \sum_{i=1}^{n} \frac{Y_i - \mu}{\sigma^2}}$.

(2) Fisherova informace o parametru μ z náhodného výběru (viz příklad 1.4.9):

$$\boxed{J_n(\mu) = nJ(\mu) = \frac{n}{\sigma^2}}.$$

(3) Uvažujme parametrickou funkci $\boxed{\tau(\mu) = \mu}$ a výběrový průměr, tj. statistiku

$$T_n(\mathbf{Y}) = \boxed{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

(i) Platí $\boxed{E\bar{Y} = \mu}$, tj. $\bar{Y}$ je **nestranným odhadem** parametru μ a $D\bar{Y} = \frac{\sigma^2}{n}$.

(ii) $\bar{Y}$ je **regulárním odhadem** parametrické funkce $\tau(\mu) = \mu$, přičemž $\tau'_\mu(\mu) = 1$, neboť $\bar{Y}$ je nestranným odhadem a platí

$$\begin{aligned}
\boxed{E\big(\bar{Y}U_1^*(\mu)\big)} &= \frac{1}{n\sigma^2} E\left[\left(\sum_{i=1}^{n} Y_i\right)\left(\sum_{i=1}^{n} Y_i - n\mu\right)\right] \\
&= \frac{1}{n\sigma^2}\sum_{i=1}^{n} \underbrace{EY_i^2}_{\sigma^2+\mu^2} + \frac{2}{n\sigma^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \underbrace{E(Y_iY_j)}_{\mu^2 (nez.)} - \frac{n\mu^2}{\sigma^2} \\
&= \frac{\sigma^2+\mu^2}{\sigma^2} + \frac{n(n-1)}{n\sigma^2}\mu^2 - \frac{n\mu^2}{\sigma^2} = \boxed{1 = \tau'_\mu(\mu)}.
\end{aligned}$$

(iii) $\bar{Y}$ je **vydatným odhadem** μ, neboť dolní Raova-Cramerova hranice

$$\boxed{C_n(\mu)} = \frac{\left[\tau'_\mu(\mu)\right]^2}{J_n(\mu)} = \frac{1}{\frac{n}{\sigma^2}} = \boxed{\frac{\sigma^2}{n}} = D\bar{Y}.$$

(b) $\boxed{\mu \text{ je } \textbf{známé}, \text{ tj. } \theta_2 = \sigma^2.}$

(1) Skórová funkce náhodného výběru (viz příklad 1.4.9):

$$U_2^*(\sigma^2) = \sum_{i=1}^{n} \frac{(Y_i-\mu)^2 - \sigma^2}{2\sigma^4} = \frac{1}{2\sigma^4}\sum_{i=1}^{n} Z_i - \frac{1}{2\sigma^2}.$$

(2) Fisherova informace o parametru $\tau(\sigma^2) = \sigma^2$ z náhodného výběru (viz příklad 1.4.9):

$$J_n(\sigma^2) = nJ(\sigma^2) = \frac{n}{2\sigma^4}.$$

(3) Uvažujme parametrickou funkci $\boxed{\tau(\sigma^2) = \sigma^2}$ a výběrový rozptyl, tj. statistiku

$$\begin{aligned}
T_n(\mathbf{Y}) = \boxed{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 &= \frac{1}{n-1}\left[\sum_{i=1}^{n} \underbrace{(Y_i-\mu)^2}_{\text{označme } Z_i} - n(\bar{Y}-\mu)^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} Z_i - n(\bar{Y}-\mu)^2\right].
\end{aligned}$$

Počítejme

$$\begin{aligned}
EZ_i &= DY_i = \sigma^2 \\
DZ_i &= EZ_i^2 - (EZ_i)^2 = \mu_4 - \sigma^4 = 2\sigma^4 \\
C(Z_i, Z_j) &= E(Z_iZ_j) - \underbrace{E(Z_i)E(Z_j)}_{\sigma^4} = 0 \qquad \Rightarrow \qquad E(Z_iZ_j) = \sigma^4 \text{ pro } i \neq j.
\end{aligned}$$

Pak

(i) Snadno lze ukázat, že platí $\boxed{ES^2 = \sigma^2}$, tj. S^2 je **nestranným odhadem** parametru σ^2. Dále obecně pro výběrový rozptyl platí:

$$DS^2 = \frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4$$

a protože v případě normálního rozdělení máme

$$\mu_4 = 3\sigma^2,$$

dostáváme

$$DS^2 = \frac{3\sigma^2}{n} - \frac{n-3}{n(n-1)}\sigma^4 = \frac{\sigma^4\left[3(n-1)-(n-3)\right]}{n(n-1)} = \frac{2\sigma^4}{n-1}.$$

(ii) S^2 je **regulárním odhadem** parametrické funkce $\tau(\sigma^2) = \sigma^2$, přičemž $\tau'_{\sigma^2}(\sigma^2) = 1$, neboť je nestranným odhadem a platí

$$\begin{aligned}
\boxed{E\big(S^2 U_2^*(\sigma^2)\big)} &= \frac{1}{2(n-1)\sigma^4} E\left[\Big(\sum_{i=1}^n Z_i - n(\bar{Y}-\mu)^2\Big)\Big(\sum_{i=1}^n Z_i - n\sigma^2\Big)\right] \\
&= \frac{1}{2(n-1)\sigma^4}\left[\sum_{i=1}^n \underbrace{EZ_i^2}_{\mu^4 = 3\sigma^4} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n \underbrace{E(Z_iZ_j)}_{\sigma^4}\right. \\
&\qquad - n\sum_{i=1}^n \underbrace{E\big(Z_i(\bar{Y}-\mu)^2\big)}_{\frac{(n+2)\sigma^4}{n^2}} \\
&\qquad \left. - n\sigma^2 \sum_{i=1}^n \underbrace{EZ_i}_{\sigma^2} + n^2\sigma^2 \underbrace{E(\bar{Y}-\mu)^2}_{D\bar{Y} = \frac{\sigma^2}{n}}\right] \\
&= \frac{3n\sigma^4 + n(n-1)\sigma^4 - (n+2)\sigma^4 - n^2\sigma^4 + n\sigma^4}{2(n-1)\sigma^4} \\
&= \boxed{1 = \tau'_{\sigma^2}(\sigma^2)},
\end{aligned}$$

přičemž platí

$$\begin{aligned}
E\big(Z_i(\bar{Y}-\mu)^2\big) &= E\left\{Z_i\left[\frac{1}{n}\sum_{i=1}^n (Y_i-\mu)\right]\left[\frac{1}{n}\sum_{i=1}^n (Y_i-\mu)\right]\right\} \\
&= \frac{1}{n^2}\left[\underbrace{EZ_i^2}_{3\sigma^4} + \sum_{i\neq j=1}^n \underbrace{E(Z_iZ_j)}_{\sigma^4}\right. \\
&\qquad \left. + \sum_{i\neq j=1}^n \sum_{i\neq j\neq k=1}^n \underbrace{E\big(Z_i(Y_j-\mu)(Y_k-\mu)\big)}_{0}\right] \\
&= \frac{1}{n^2}\left[3\sigma^4 + (n-1)\sigma^4\right] = \frac{(n+2)\sigma^4}{n^2}.
\end{aligned}$$

(iii) S^2 je **asymptoticky vydatným odhadem** σ^2, neboť dolní Raova-Cramerova hranice je rovna

$$\boxed{C_n(\sigma^2)} = \frac{\left[\tau'_{\sigma^2}(\sigma^2)\right]^2}{J_n(\sigma^2)} = \frac{1}{\frac{n}{2\sigma^4}} = \boxed{\frac{2\sigma^4}{n}} < DS^2 = \frac{2\sigma^4}{n-1} \quad \text{a} \quad \lim_{n\to\infty} \frac{C_n(\sigma^2)}{DS^2} = 1.$$

1.5 EXPONENCIÁLNÍ TŘÍDA ROZDĚLENÍ PRAVDĚPODOBNOSTÍ

Definice 1.5.1.
Řekneme, že pozorování pochází z rozdělení ***exponenciálního typu****, pokud jeho* pravděpodobnostní funkce *(v případě diskrétních rozdělení) či* hustota *(v případě spojitých rozdělení) je tvaru* $\boxed{f(y) = \exp\{a(y)b(\theta) + c(\theta) + d(y)\}}$ *kde*

θ *je (neznámý) tzv.* ***přirozený parametr***
a
$a(y), b(\theta), c(\theta), d(y)$ *jsou známé funkce.*

Pokud

- $\boxed{a(y) = y,}$ *říkáme že pravděpodobnostní funkce, popř. hustota je v* ***kanonické formě****.*
- *v konkrétním rozdělení figurují další neznámé parametry, nazveme je tzv.* ***rušivými parametry****.*

V dalším budeme uvažovat pouze **regulární** a **kanonické** formy spolu s podmínkou $\boxed{b(\theta) = \theta}$ a přitom zavedeme do označení jeden rušivý parametr ϕ :

$$\boxed{f(y) = \exp\left\{\frac{y\theta - \gamma(\theta)}{\psi(\phi)} + d(y, \phi)\right\}},$$

kde θ a ϕ jsou parametry
$\gamma(\theta),\ \psi(\phi) > 0,\ d(y)$ jsou známé funkce,
a pokud
$\psi(\phi) = \frac{\phi}{\omega} > 0,$ $\phi > 0$ je tzv. **faktor měřítka** (scale factor)
$\omega > 0$ je známá **apriorní váha**.

Tato forma se také nazývá **škálovou formou hustoty exponenciálního typu**.
Poznámka 1.5.2.
Jestliže neplatí $b(\theta) = \theta$, stačí provést jednoduchou reparametrizaci a zavést případně nový parametr

$$\boxed{\theta^* = b(\theta)},$$

který se pak nazývá **kanonickým parametrem**.

Lemma 1.5.3.
Mějme náhodnou veličinu Y z rozdělení s regulární hustotou $f \in \mathcal{F}^m_{reg}$ exponenciálního typu:

$$\boxed{f(y) = \exp\left\{\frac{y\theta - \gamma(\theta)}{\psi(\phi)} + d(y, \phi)\right\}}. \tag{1.6}$$

Pak $$\boxed{EY = \gamma'(\theta)}$$
(tento vztah je ekvivalentní s podmínkou (4) v definici regularity 1.4.1.)

Nechť navíc platí (tj. f je regulární i v 2.derivacích)

$$\boxed{E\left(\frac{f''(Y;\theta)}{f(Y;\theta)}\right) = 0}, \tag{1.7}$$

kde $f''(Y;\theta) = \frac{d^2 f(Y;\theta)}{d\theta^2}$, *pak*

$$\boxed{DY = \gamma''(\theta)\psi(\phi)}$$

Funkce $\boxed{\gamma''(\theta) = \dfrac{DY}{\psi(\phi)}}$ *se nazývá* ***rozptylovou funkcí*** *(*variance function*).*

Důkaz.
Označme **logaritmus věrohodnostní funkce** náhodné veličiny Y

$$l(\theta; y) = \ln f(y;\theta)$$

a **skóre**

$$U(\theta) = \frac{dl(\theta;Y)}{d\theta}.$$

Pro rozdělení exponenciálního typu v kanonické formě platí:

$$\frac{df(y;\theta)}{d\theta} = \frac{y - \gamma'(\theta)}{\psi(\phi)} f(y;\theta)$$

$$\frac{d^2 f(y;\theta)}{d\theta^2} = \frac{[y-\gamma'(\theta)]^2 - \gamma''(\theta)\psi(\phi)}{\psi(\phi)^2} f(y;\theta)$$

$$U(\theta) = \frac{dl(\theta;Y)}{d\theta} = \frac{df(Y;\theta)}{d\theta}\frac{1}{f(Y;\theta)} = \frac{Y-\gamma'(\theta)}{\psi(\phi)}$$

$$J(\theta) = J_{11}(\theta) = \int_M \left(\frac{d\ln f(y;\theta)}{d\theta}\right)^2 dF(y;\theta) = \int_M \left(\frac{Y-\gamma'(\theta)}{\psi(\phi)}\right)^2 dF(y;\theta) = EU(\theta)^2 \quad \square$$

Podmínka regularity (4) z definice 1.4.1

$$\int_M \frac{d\ln f(y;\theta)}{d\theta} dF(y;\theta) = EU(\theta) = 0$$

je u rozdělení exponenciálního typu tvaru

$$EU(\theta) = E\left(\frac{Y-\gamma'(\theta)}{\psi(\phi)}\right) = \frac{1}{\psi(\phi)}\left\{EY - \gamma'(\theta)\right\} = 0$$

a je ekvivalentní s podmínkou

$$\boxed{EY = \gamma'(\theta)}.$$

Podmínka z předpokladů tohoto lemma

$$E\left(\frac{f''(Y;\theta)}{f(Y;\theta)}\right) = 0$$

je u rozdělení exponenciálního typu tvaru

$$E\left\{\frac{[Y-\gamma'(\theta)]^2-\gamma''(\theta)\psi(\phi)}{\psi(\phi)^2}\right\}=\frac{1}{\psi(\phi)^2}\left\{\underbrace{E\left[Y-\overbrace{\gamma'(\theta)}^{=EY}\right]^2}_{=DY}-\gamma''(\theta)\psi(\phi)\right\}=0$$

a je ekvivalentní s podmínkou

$$\boxed{DY=\gamma''(\theta)\psi(\phi)}.$$

1.6 PŘÍKLADY ROZDĚLENÍ EXPONENCIÁLNÍHO TYPU

1.6.1 Normální rozdělení

Mějme $\boxed{Y\sim N(\mu,\sigma^2)}$ $\quad \mu\in\mathbb{R},\ \sigma^2>0.$

Pak $f(y)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}=\exp\left\{\frac{y\mu-\overbrace{\frac{1}{2}\mu^2}^{\gamma(\theta)}}{\underbrace{\sigma^2}_{\psi(\phi)}}\underbrace{-\frac{1}{2}\frac{y^2}{\sigma^2}-\frac{1}{2}\ln\left(2\pi\sigma^2\right)}_{d(y,\phi)}\right\}$

a $\quad \gamma(\theta)=\frac{1}{2}\mu^2=\frac{1}{2}\theta^2 \quad\Rightarrow\quad \gamma'(\theta)=\theta=\mu$

$\quad\Rightarrow\quad \gamma''(\theta)=1$

$\psi(\phi)=\sigma^2 \quad\Rightarrow\quad \phi=\sigma^2.$

Skutečně platí

$$EY=\gamma'(\theta)=\mu$$

a

$$DY=\gamma''(\theta)\psi(\phi)=\sigma^2.$$

Tedy	***přirozený parametr***	$\boxed{\theta=\mu}$	***scale factor***	$\phi=\sigma^2$
	rozptylová funkce	$\boxed{V(\mu)=1}$	***váhy***	$\omega=1.$

Ověření regularity pro jednoparametrový systém hustot (σ^2 bereme jako rušivý parametr):

(1) Parametrický prostor pro parametr $\theta=\mu$: $\boldsymbol{\Theta}=(-\infty,\infty)$ je neprázdná otevřená množina.

(2) Množina $M=\{y\in\mathbb{R}:f(y)>0\}=(-\infty,\infty)$ nezávisí na $\theta\in\boldsymbol{\Theta}$.

(3) Existuje $U=\frac{Y-\gamma'(\theta)}{\psi(\phi)}=\frac{Y-\theta}{\sigma^2}$.

(4) Střední hodnota $EU = 0 \Leftrightarrow \gamma'(\theta) = \theta = \mu = EY$.

(5) Fisherova informace $J = EU^2 = E\left(\frac{Y-\theta}{\sigma^2}\right)^2 = \frac{1}{\sigma^2} > 0 \quad \Rightarrow$ hustota je **regulární v 1. derivacích**.

(6) Podmínka $E\left(\frac{f''}{f}\right) = 0 \Leftrightarrow \gamma''(\theta)\psi(\phi) = \sigma^2 = DY \quad \Rightarrow$ hustota je **regulární v 2. derivacích**.

1.6.2 Binomické rozdělení

Mějme $$\boxed{Z = nY \sim Bi(n, \pi)} \qquad n \in \mathbb{N}, \pi \in (0,1)$$

pak $f_Z(z) = \binom{n}{z}\pi^z(1-\pi)^{n-z} = \exp\left\{z\ln\left(\frac{\pi}{1-\pi}\right) + n\ln(1-\pi) + \ln\binom{n}{z}\right\}$
pro $z = 0, \ldots, n,$
přičemž $EZ = \mu = n\pi$ a $DZ = n\pi(1-\pi)$.

Pravděpodobnostní funkce **není** ve **škálové formě**, proveďme **reparametrizaci**

$$\boxed{\theta} = \ln\left(\frac{\pi}{1-\pi}\right) = \ln\left(\frac{n\pi}{n-n\pi}\right) = \boxed{\ln\left(\frac{\mu}{n-\mu}\right)} \quad \Rightarrow \quad \pi = \frac{e^\theta}{1+e^\theta} \text{ a } 1-\pi = \frac{1}{1+e^\theta}.$$

Tedy $$f_Z(z) = \exp\left\{z\theta - \underbrace{n\ln\left(1+e^\theta\right)}_{\gamma(\theta)} + \underbrace{\ln\binom{n}{z}}_{d(y,\phi)}\right\}$$

a
$$\begin{aligned}
\gamma(\theta) = n\ln\left(1+e^\theta\right) \quad &\Rightarrow \quad \gamma'(\theta) = n\frac{e^\theta}{1+e^\theta} = n\pi = \mu \\
&\Rightarrow \quad \gamma''(\theta) = n\frac{e^\theta}{(1+e^\theta)^2} = n\pi(1-\pi) = \mu\left(1-\frac{\mu}{n}\right) \\
\psi(\phi) = \frac{\phi}{\omega} = 1 \quad &\Rightarrow \quad \omega = 1 \quad \phi = 1
\end{aligned}$$

Skutečně platí
$$EZ = \gamma'(\theta) = \mu$$
a
$$DZ = \gamma''(\theta)\psi(\phi) = n\pi(1-\pi).$$

Tedy ***přirozený parametr*** $\boxed{\theta = \ln\left(\frac{\mu}{n-\mu}\right)}$

rozptylová funkce $\boxed{V(\mu) = \mu\left(1-\frac{\mu}{n}\right)}$

scale factor $\phi = 1$

váhy $\omega = 1.$

Poznámka 1.6.1.
Je třeba poznamenat, že ve vztahu

$$\gamma(\theta) = n \ln\left(1 + e^{\theta}\right)$$

se vedle parametru θ vyskytuje i parametr n, který je však vždy znám. Abychom dostáli úvodní definici (1.6), proto ho nebudeme považovat za rušivý parametr, ale za známou konstantu. Tomuto problému se lze vyhnout, když přejdeme od absolutních četností Z k relativním četnostem Y (viz dále).

Ověření regularity pro jednoparametrový systém hustot (n předpokládáme, že známe):

(1) Parametrický prostor pro parametr $\theta = \ln\left(\frac{\pi}{1-\pi}\right)$: $\boldsymbol{\Theta} = (0, \infty)$ je neprázdná otevřená množina.

(2) Množina $M = \{z \in \mathbb{R} : f(z) > 0\} = \{0, 1, \ldots, n\}$ nezávisí na $\theta \in \boldsymbol{\Theta}$.

(3) Existuje $U = \frac{Z-\gamma'(\theta)}{\psi(\phi)} = Z - \frac{ne^{\theta}}{1+e^{\theta}}$.

(4) Střední hodnota $EU = 0 \Leftrightarrow \gamma'(\theta) = n\frac{e^{\theta}}{1+e^{\theta}} = n\pi = EZ$.

(5) Fisherova informace $J = EU^2 = E\left(Z - \frac{ne^{\theta}}{1+e^{\theta}}\right)^2 = n\pi(1-\pi) > 0 \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 1. derivacích**.

(6) Podmínka $E\left(\frac{f''}{f}\right) = 0 \Leftrightarrow \gamma''(\theta)\psi(\phi) = n\frac{e^{\theta}}{(1+e^{\theta})^2} = n\pi(1-\pi) = DZ \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 2. derivacích**.

V případě, že uvažuje místo absolutních **relativní četnosti**: $\boxed{Y = \frac{Z}{n}}$

je pravděpodobnostní funkce nenulová pro $y \in \{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$ a je tvaru

$$f_Y(y) = \binom{n}{ny}\pi^{ny}(1-\pi)^{n-ny} = \exp\left\{\frac{y\ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi)}{1/n} + \ln\binom{n}{ny}\right\}$$

a $\qquad EY = \mu = \pi \quad$ a $\quad DY = \frac{\pi(1-\pi)}{n}$.

Pravděpodobnostní funkce **není** ve **škálové formě**, proveďme **reparametrizaci**

$$\boxed{\theta = \ln\left(\frac{\pi}{1-\pi}\right)} \quad \Rightarrow \quad \pi = \frac{e^{\theta}}{1+e^{\theta}} \text{ a } 1-\pi = \frac{1}{1+e^{\theta}}$$

Tedy $\qquad f_Y(y) = \exp\left\{\frac{y\theta - \overbrace{\ln\left(1+e^{\theta}\right)}^{\gamma(\theta)}}{\underbrace{1/n}_{\psi(\phi)}} + \underbrace{\ln\binom{n}{ny}}_{d(y,\phi)}\right\}$

a

$$\gamma(\theta) = \ln\left(1 + e^{\theta}\right) \quad \Rightarrow \quad \gamma'(\theta) = \frac{e^{\theta}}{1+e^{\theta}} = \pi = \mu$$

$$\Rightarrow \quad \gamma''(\theta) = \frac{e^{\theta}}{(1+e^{\theta})^2} = \pi(1-\pi)$$

$$\psi(\phi) = \frac{\phi}{\omega} = \frac{1}{n} \quad \Rightarrow \quad \omega = n \quad \phi = 1$$

Skutečně platí $EY = \gamma'(\theta) = \mu$

$DY = \gamma''(\theta)\psi(\phi) = \frac{\pi(1-\pi)}{n}$,

Tedy **_přirozený parametr_** $\boxed{\theta = \ln\left(\frac{\mu}{1-\mu}\right)}$

rozptylová funkce $\boxed{V(\mu) = \mu(1-\mu)}$

scale factor $\phi = 1$

váhy $\omega = n$.

Ověření regularity pro jednoparametrový systém hustot (n předpokládáme, že známe):

(1) Parametrický prostor pro parametr $\theta = \ln\left(\frac{\pi}{1-\pi}\right)$: $\mathbf{\Theta} = (0, \infty)$ je neprázdná otevřená množina.

(2) Množina $M = \{y \in \mathbb{R} : f(y) > 0\} = \{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$ nezávisí na $\theta \in \mathbf{\Theta}$.

(3) Existuje $U = \frac{Y-\gamma'(\theta)}{\psi(\phi)} = Y - \frac{e^\theta}{1+e^\theta}$.

(4) Střední hodnota $EU = 0 \Leftrightarrow \gamma'(\theta) = \frac{e^\theta}{1+e^\theta} = \pi = EY$.

(5) Fisherova informace $J = EU^2 = E\left(Y - \frac{e^\theta}{1+e^\theta}\right)^2 = \pi(1-\pi) > 0 \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 1. derivacích**.

(6) Podmínka $E\left(\frac{f''}{f}\right) = 0 \Leftrightarrow \gamma''(\theta)\psi(\phi) = \frac{e^\theta}{(1+e^\theta)^2} = \pi(1-\pi) = DY \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 2. derivacích**.

1.6.3 Poisonovo rozdělení

Mějme $\boxed{Y \sim Po(\lambda)}$ $\quad \lambda > 0$

pak $f(y) = \frac{\lambda^y e^{-\lambda}}{y!} = \exp\{y \ln\lambda - \lambda - \ln y!\} \qquad y = 0, 1, 2, \dots$

přičemž $EY = \mu = \lambda$ a $DY = \lambda$.

Protože pravděpodobnostní funkce **není** ve **škálové formě**, proveďme **reparametrizaci**

$$\boxed{\theta = \ln\lambda} \quad \Rightarrow \quad \lambda = e^\theta$$

a $$f(y) = \exp\left\{y\theta - \underbrace{e^\theta}_{\gamma(\theta)} - \underbrace{\ln y!}_{d(y,\phi)}\right\}$$

$$\gamma(\theta) = e^\theta \qquad \Rightarrow \quad \gamma'(\theta) = e^\theta = \lambda = \mu$$
$$\Rightarrow \quad \gamma''(\theta) = e^\theta = \lambda = \mu$$

$$\psi(\phi) = \frac{\phi}{\omega} = 1 \qquad \Rightarrow \qquad \omega = 1 \quad \phi = 1$$

Skutečně platí

$$EY = \gamma'(\theta) = \lambda = \mu$$

a

$$DY = \gamma''(\theta)\psi(\phi) = e^{\theta} = \lambda = \mu,$$

Tedy

přirozený parametr	$\boxed{\theta = \ln\lambda}$
rozptylová funkce	$\boxed{V(\mu) = \mu}$
scale factor	$\phi = 1$
váhy	$\omega = 1.$

Ověření regularity pro jednoparametrový systém hustot:

(1) Parametrický prostor pro parametr $\theta = \ln\lambda : \mathbf{\Theta} = (0, \infty)$ je neprázdná otevřená množina.

(2) Množina $M = \{y \in \mathbb{R} : f(y) > 0\} = \{0, 1, 2, \ldots\}$ nezávisí na $\theta \in \mathbf{\Theta}$.

(3) Existuje $U = \frac{Y - \gamma'(\theta)}{\psi(\phi)} = Y - e^{\theta} = Y - \lambda$.

(4) Střední hodnota $EU = 0 \Leftrightarrow \gamma'(\theta) = e^{\theta} = \lambda = EY$.

(5) Fisherova informace $J = EU^2 = E\left(Y - e^{\theta}\right)^2 = \lambda > 0 \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 1. derivacích**.

(6) Podmínka $E\left(\frac{f''}{f}\right) = 0 \Leftrightarrow \gamma''(\theta)\psi(\phi) = e^{\theta} = \lambda = DY \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 2. derivacích**.

1.6.4 Gamma rozdělení

Mějme $\boxed{Y \sim G(\alpha, \beta)}$ $\quad \alpha > 0, \beta > 0.$

Pak $\quad f(y) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} e^{-\frac{y}{\beta}} \qquad y > 0$

přičemž $\quad EY = \alpha\beta \quad$ a $\quad DY = \alpha\beta^2.$

Tento tvar hustoty je však pro nás nevhodný (ve střední hodnotě máme rušivý parametr α), proto uvažujme reparametrizaci

$$\boxed{\mu = \alpha\beta} \qquad \Rightarrow \qquad \beta = \frac{\mu}{\alpha},$$

pak $\quad f(y) = \frac{1}{\Gamma(\alpha)}\left(\frac{\alpha}{\mu}\right)^{\alpha} y^{\alpha-1} e^{-\frac{\alpha}{\mu}y} = \exp\left\{\frac{y\left(-\frac{1}{\mu}\right) - \ln\mu}{\frac{1}{\alpha}} + \alpha\ln\alpha + (\alpha - 1)\ln y - \ln\Gamma(\alpha)\right\}$

přičemž $\quad EY = \mu \quad$ a $\quad DY = \frac{\mu^2}{\alpha}.$

Hustota není v kanonickém tvaru, proto parametrizujme $\boxed{\theta = -\dfrac{1}{\mu}} \quad \Rightarrow \quad \mu = -\frac{1}{\theta}$

pak $$f(y) = \exp\left\{ \frac{y\theta - \overbrace{\ln(-\dfrac{1}{\theta})}^{\gamma(\theta)}}{\frac{1}{\alpha}} + \underbrace{\alpha \ln \alpha + (\alpha-1)\ln y - \ln\Gamma(\alpha)}_{d(y,\phi)} \right\}$$

a $$\gamma(\theta) = \ln(\mu) = \ln(-\tfrac{1}{\theta}) \quad \Rightarrow \quad \gamma'(\theta) = -\tfrac{\theta}{\theta^2} = -\tfrac{1}{\theta} = \mu$$
$$\Rightarrow \quad \gamma''(\theta) = \tfrac{1}{\theta^2} = \mu^2$$
$$\psi(\phi) = \tfrac{\phi}{\omega} = \tfrac{1}{\alpha} \quad \Rightarrow \quad \omega = 1 \quad \phi = \tfrac{1}{\alpha}$$

Skutečně platí
$$EY = \gamma'(\theta) = \mu$$
a
$$DY = \gamma''(\theta)\psi(\phi) = \tfrac{\mu^2}{\alpha} = \tfrac{\alpha^2\beta^2}{\alpha} = \alpha\beta^2$$

Tedy ***přirozený parametr*** $\boxed{\theta = -\dfrac{1}{\mu}}$

rozptylová funkce $\boxed{V(\mu) = \mu^2}$

scale factor $\phi = \frac{1}{\alpha}$

váhy $\omega = 1.$

Ověření regularity pro jednoparametrový systém hustot:

(1) Parametrický prostor pro parametr $\theta = -\frac{1}{\mu} : \mathbf{\Theta} = (-\infty, 0)$ je neprázdná otevřená množina.

(2) Množina $M = \{y \in \mathbb{R} : f(y) > 0\} = (0, \infty)$ nezávisí na $\theta \in \mathbf{\Theta}$.

(3) Existuje $U = \frac{Y - \gamma'(\theta)}{\psi(\phi)} = \alpha(Y - \mu)$.

(4) Střední hodnota $EU = 0 \Leftrightarrow \gamma'(\theta) = -\frac{1}{\theta} = \mu = EY$.

(5) Fisherova informace $J = EU^2 = E\left[\alpha(Y-\mu)\right]^2 = \alpha^2 \frac{\mu^2}{\alpha} = \alpha\mu^2 > 0 \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 1. derivacích**.

(6) Podmínka $E\left(\frac{f''}{f}\right) = 0 \Leftrightarrow \gamma''(\theta)\psi(\phi) = \frac{\mu^2}{\alpha} = DY \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 2. derivacích**.

1.6.5 Exponenciální rozdělení

Exponenciální rozdělení je speciálním případem gama rozdělení $\boxed{Y \sim Ex(\lambda) \equiv G(1, \lambda)}$

Pak $f(y) = \frac{1}{\lambda} e^{-\frac{y}{\lambda}} = \exp\{-\frac{1}{\lambda} y - \ln\lambda\} \qquad y > 0$

Přičemž $EY = \mu = \lambda$ a $DY = \lambda^2$.

Hustota není v kanonickém tvaru, proto parametrizujme

$$\boxed{\theta = -\frac{1}{\lambda}} \quad \Rightarrow \quad \lambda = -\tfrac{1}{\theta}$$

Tedy

$$f(y) = \exp\left\{ y\theta - \overbrace{\ln(-\frac{1}{\theta})}^{=\gamma(\theta)} \right\}$$

a

$$\begin{aligned} \gamma(\theta) = \ln(\mu) = \ln(-\tfrac{1}{\theta}) \quad &\Rightarrow \quad \gamma'(\theta) = -\tfrac{1}{\theta} = \lambda = \mu \\ &\Rightarrow \quad \gamma''(\theta) = \tfrac{1}{\theta^2} = \lambda^2 = \mu^2 \\ \psi(\phi) = \tfrac{\phi}{\omega} = 1 \quad &\Rightarrow \quad \omega = 1 \quad \phi = 1 \end{aligned}$$

Skutečně platí

$$EY = \gamma'(\theta) = \mu$$

a

$$DY = \gamma''(\theta)\psi(\phi) = \mu^2 = \lambda^2$$

tj. jde o **regulární systém hustot** a navíc platí podmínka (1.7).

Tedy

Tedy	***přirozený parametr***	$\boxed{\theta = -\frac{1}{\mu}}$
	rozptylová funkce	$\boxed{V(\mu) = \mu^2}$
	scale factor	$\phi = 1$
	váhy	$\omega = 1.$

Ověření regularity pro jednoparametrový systém hustot:

(1) Parametrický prostor pro parametr $\theta = -\frac{1}{\mu}$: $\boldsymbol{\Theta} = (-\infty, 0)$ je neprázdná otevřená množina.

(2) Množina $M = \{y \in \mathbb{R} : f(y) > 0\} = (0, \infty)$ nezávisí na $\theta \in \boldsymbol{\Theta}$.

(3) Existuje $U = \frac{Y-\gamma'(\theta)}{\psi(\phi)} = (Y - \lambda)$.

(4) Střední hodnota $EU = 0 \Leftrightarrow \gamma'(\theta) = -\frac{1}{\theta} = \lambda = EY$.

(5) Fisherova informace $J = EU^2 = E\left(Y - \mu\right)^2 = \lambda^2 > 0 \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 1. derivacích**.

(6) Podmínka $E\left(\frac{f''}{f}\right) = 0 \Leftrightarrow \gamma''(\theta)\psi(\phi) = \lambda^2 = DY \quad \Rightarrow$ pravděpodobnostní funkce je **regulární v 2. derivacích**.

Kapitola 2

GLM MODELY

2.1 KLASICKÝ LINEÁRNÍ REGRESNÍ MODEL

Mějme **regresní model** plné hodnosti:

$$\boxed{\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \ \wedge\ h(\mathbf{X}) = h(\mathbf{X}^{\mathrm{T}}\mathbf{X}) = m = p+1 \ \wedge\ n > m \ \wedge\ \boldsymbol{\varepsilon} \sim \mathcal{L}_n(\mathbf{0}, \sigma^2\mathbf{I}_n)}$$

kde je	vektor závisle proměnných	$\mathbf{Y} = (Y_1, \dots, Y_n)^T$
	matice plánu	$\mathbf{X} = (x_{ij}) \quad i = 1, \dots, n; \quad j = 0, \dots, p$
	vektor chyb	$\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^{\mathrm{T}} \quad E\boldsymbol{\varepsilon} = \mathbf{0}; \quad D\boldsymbol{\varepsilon} = \sigma^2\mathbf{I}_n;$

Odhad neznámých parametrů $\boldsymbol{\beta}$ provedený *metodou nejmenších čtverců* je řešením **normálních rovnic** $\boxed{\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^{\mathrm{T}}\mathbf{Y}}$ a platí: $\boxed{\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\,\mathbf{X}^{\mathrm{T}}\mathbf{Y}}$.

Označme

◇ $\boxed{\widehat{\mathbf{Y}}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \underbrace{\mathbf{X}\,(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\,\mathbf{X}^{\mathrm{T}}}_{\mathbf{H}}\mathbf{Y} = \mathbf{H}\mathbf{Y}$

◇ $\boxed{\hat{\boldsymbol{\varepsilon}}} = \mathbf{Y} - \widehat{\mathbf{Y}} = (\underbrace{\mathbf{I} - \mathbf{H}}_{\mathbf{M}})\mathbf{Y} = \mathbf{M}\mathbf{Y} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})) = \underbrace{\mathbf{M}\mathbf{X}}_{=\mathbf{0}}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}$

◇ $\boxed{s^2} = \frac{SSE}{n-m} = \frac{(\mathbf{Y}-\widehat{\mathbf{Y}})^{\mathrm{T}}(\mathbf{Y}-\widehat{\mathbf{Y}})}{n-m} = \frac{\hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{\boldsymbol{\varepsilon}}}{n-m} = \frac{\mathbf{Y}^{\mathrm{T}}(\mathbf{I}-\mathbf{H})\mathbf{Y}}{n-m} = \frac{\boldsymbol{\varepsilon}^{\mathrm{T}}(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}}{n-m}$

Platí: $*\ E\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$

$*\ Es^2 = \frac{E(SSE)}{n-m} = \sigma^2$, tj. s^2 je nestranným odhadem rozptylu σ^2

$*\ D\boldsymbol{\beta} = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$

Platí-li navíc $\boxed{\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)}$, pak

⋆ $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I_n)$

⋆ $\widehat{\boldsymbol{\varepsilon}} \sim N_n(\mathbf{O}, \sigma^2(\mathbf{I} - \mathbf{H}))$

⋆ $\widehat{\boldsymbol{\beta}} \sim N_m(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$

⋆ $\frac{SSE}{\sigma^2} \sim \chi^2(n-m)$

⋆ $\widehat{\boldsymbol{\beta}}$ a s^2 jsou stochasticky nezávislé

- ⋆ $T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2 v_{jj}}} \sim t(n-m)$, kde $j = 1, \ldots, n$ a $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = (v_{ij})_{i,j=1}^{m}$
- ⋆ $F = \frac{1}{qs^2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)^{\mathrm{T}} W^{-1}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \sim F(q, n-m)$,

  kde $\quad (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{pmatrix} \mathbf{V} & \mathbf{U} \\ \mathbf{U} & \mathbf{W} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}; \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} \quad$ a $\quad h(\mathbf{W}) = q$
- ⋆ $T = \frac{\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \mathbf{c}^{\mathrm{T}}\boldsymbol{\beta}}{\sqrt{s^2 \mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}}} \sim t(n-m)$, kde $\mathbf{c} = (c_0, c_1, \ldots, c_p)^{\mathrm{T}}$ a $E(\mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) = \mathbf{c}^{\mathrm{T}}\boldsymbol{\beta}$

Označme i-tý řádek matice plánu $\mathbf{X}$ jako $\mathbf{x}_i^{\mathrm{T}} = (x_{i0}, \ldots, x_{ip})$, pak

- $\begin{matrix} Y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i \sim N(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \sigma^2), \\ \hat{Y}_i = \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} \sim N(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \sigma^2\mathbf{x}_i^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i) \end{matrix} \quad \Rightarrow \quad Y_i - \hat{Y}_i \sim N(0, \sigma^2(1 + \mathbf{x}_i^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i)).$

Intervaly spolehlivosti		
pro parametry β_j $(j = 1, \ldots, m)$	$\beta_j \mp c_1$	$c_1 = t_{1-\frac{\alpha}{2}}(n-m)s\sqrt{v_{jj}}$
pro střední hodnotu predikce $E\hat{Y}_i = E\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$	$\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} \mp c_2$	$c_2 = t_{1-\frac{\alpha}{2}}(n-m)s\sqrt{\mathbf{x}_i^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i}$
pro predikci $\hat{Y}_i = \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$ $(i = 1, \ldots, n)$	$\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} \mp c_3$	$c_3 = t_{1-\frac{\alpha}{2}}(n-m)s\sqrt{1 + \mathbf{x}_i^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_i}$

kde $t_{1-\frac{\alpha}{2}}(n-m)$ je $1 - \frac{\alpha}{2}$ kvantil Studentova rozdělení o $n-m$ stupních volnosti

2.2 OMEZENÍ KLASICKÉHO LINEÁRNÍHO REGRESNÍHO MODELU

Mějme klasický lineární regresní model plné hodnosti

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \;\wedge\; h(\mathbf{X}) = h(\mathbf{X}^{\mathrm{T}}\mathbf{X}) = m = p + 1 \;\wedge\; n > m \;\wedge\; \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n),$$

kde	$\mathbf{Y} = (Y_1, \ldots, Y_n)^T$	je vektor závisle proměnných,
	$\mathbf{X} = (x_{ij})$	je matice plánu, $(i = 1, \ldots, n;\ j = 0, \ldots, p)$
	$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$	je vektor chyb, přičemž $E\boldsymbol{\varepsilon} = \mathbf{0}$; $D\boldsymbol{\varepsilon} = \sigma^2\mathbf{I}_n$.

Když se podíváme na tento model blíže, zjistíme, že se skládá ze dvou částí:

Systematická (signální) část vyjadřuje lineární vztah pro střední hodnotu a neznámé parametry β_j, tj.

$$\boxed{EY_i = \mu_i = \mathbf{x}_i^T\boldsymbol{\beta}}.$$

Tato část je obvykle cílem zkoumání, snažíme se pomocí ní maximálně možně vysvětlit chování náhodné veličiny Y_i a zjistit skrze parametry β_j velikost a znaménko závislosti na vysvětlujících veličinách $\mathbf{x}_i$.

V reálném světě má mnoho procesů jiný, než lineární vztah závislosti. Např. v ekonomii se ukazuje, že mnoho vztahů má logaritmickou závislost, k vysvětlení procesů v přírodních vědách se užívají reciproké, mocninné i další vztahy. Vysvětlovaná veličina popisující pravděpodobnost přežití člověka, v případě určité nemoci a určitého způsobu léčby, může z definice pravděpodobnosti nabývat hodnot pouze z intervalu $[0, 1]$, což by v případě klasického lineárního modelu bylo možné zajistit jen za přijetí omezení na estimátor $\boldsymbol{\beta}$.

Náhodná část je reprezentovaná náhodnými chybami $\boxed{\varepsilon_i}$, které shrnují v sobě všechny ostatní vlivy, působící na Y_i, kromě již uvedených v systematické části. Rozdělení náhodných veličin ε_i je závislé na rozdělení Y_i a má tvar

$$\boxed{\varepsilon_i \sim N(0, \sigma^2)},$$

kde ε_i jsou nezávislé. Právě **normalita chyb** je často nesplněným předpokladem klasického lineárního regresního modelu. Připomeňme, že normalita se vyznačuje nezávislosti střední hodnoty a rozptylu. Typicky např. u ekonomických veličin s rostoucí střední hodnotou obvykle roste rozptyl náhodné veličiny, přičemž náhodné chyby mají v těchto případech často nesymetrická, kladně sešikmená rozdělení.

Shrneme-li předchozí, můžeme říci, že **klasický lineární regresní model** je sice velmi důležitým stochastickým modelem, avšak má celou řadu omezení:

- Je **omezen pouze na třídu normálních rozdělení**: $\boxed{Y_i \sim N(\mu_i, \sigma^2)}$ $\quad i = 1, \ldots, n$, kde $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ tvoří náhodný výběr.
- Předpokládá **striktní rovnost mezi střední hodnotou náhodné veličiny** Y_i **a lineární kombinací prediktorů**: $\boxed{EY_i = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta}}$, kde

  $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})^T$ je vektor prediktorů a
  $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$ je vektor neznámých parametrů.

Je však možné provést **zobecnění** tohoto klasického lineárního modelu dvěma směry:

1. Zobecnění na nenormální rozdělení, a to na tzv. **třídu exponenciálních rozdělení**

2. Zobecnění na nelineární funkce, které **spojují** neznámé střední hodnoty výchozího rozdělení náhodné veličiny Y_i s prediktivními proměnnými.

2.3 DEFINICE JEDNOROZMĚRNÉHO GLM

Definice 2.3.1 (Zobecněný lineární model).
Mějme náhodný výběr

$$\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$$

a nechť rozdělení Y_i závisí na pevných vektorech

$$\mathbf{x}_i = (x_{i1}, \ldots, x_{im})^{\mathrm{T}} \in \mathbb{R}^m$$

prostřednictím neznámého vektoru parametrů

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^{\mathrm{T}}.$$

Matice

$$\mathbf{X} = (\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_n^{\mathrm{T}})^{\mathrm{T}}$$

má rozměr $n \times m$ a hodnost $m < n$.
*Říkáme, že $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ se řídí **zobecněným lineárním modelem** (generalized linear model, zkratka GLM), jestliže dále platí:*

1. *rozdělení $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ je exponenciálního typu s **regulární** hustotou tvaru*

$$\boxed{f(\mathbf{y}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i, \theta_i) = \exp\left\{\sum_{i=1}^{n} \left[\frac{y_i\theta_i - \gamma(\theta_i)}{\psi_i(\phi)} + d(y_i, \phi)\right]\right\}} \qquad (2.1)$$

2. *parametr θ_i závisí na $\mathbf{x}_i$ a $\boldsymbol{\beta}$ prostřednictvím parametru*

$$\boxed{\eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}, \qquad (2.2)$$

   *který nazveme **lineární prediktor**.*

3. *Existuje známá ryze monotónní diferencovatelná funkce $\boxed{g}$, tzv. **linkovací funkce** (link function), a platí*

$$\boxed{\eta_i = g(\mu_i) \qquad \mu_i = g^{-1}(\eta_i)}, \quad \textit{kde} \quad \mu_i = \mu(\theta_i) = EY_i. \qquad (2.3)$$

   *Řekneme, že linkovací funkce je **kanonická**, pokud $\boxed{\theta_i = \eta_i = g(\mu_i)}$.*

*Matici $\mathbf{X} = (\mathbf{x}_i^{\mathrm{T}})_{i=1}^{n}$ nazýváme **maticí plánu**.*

<u>Poznámka</u> 2.3.2.
Z regularity systému

$$\mathcal{F}_{\mathrm{reg}}^{n} = \{f(\mathbf{y}; \boldsymbol{\theta}) : \boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^{\mathrm{T}} \in \boldsymbol{\Theta}\}$$

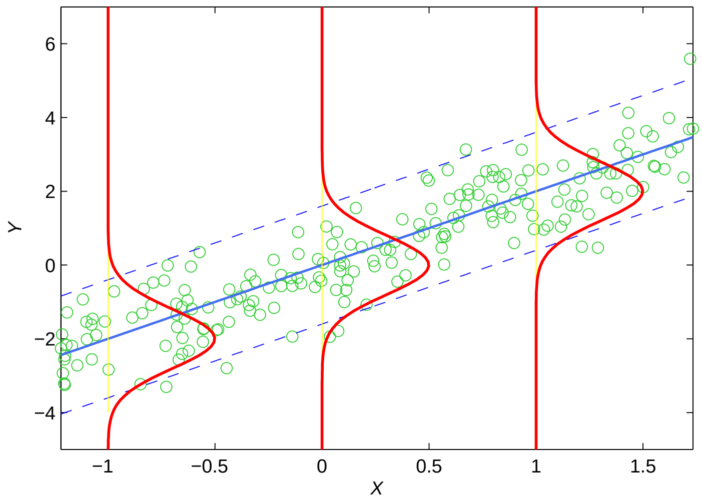
plyne regularita z něj odvozeného systému

$$\mathcal{F}_{\mathrm{reg}}^{m} = \{f(\mathbf{y}, \boldsymbol{\beta}) : \boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^{\mathrm{T}} \in \mathbb{R}^m\}, \quad \text{neboť} \quad EY_i = \mu_i = \mu(\theta_i) = g^{-1}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}).$$

Příklad 2.3.3.
Regresní přímka v klasickém lineárním regresním modelu:

$\boxed{Y_i \sim N(\mu_i, \sigma^2)}$ jsou pro $i = 1, \ldots, n$ nezávislé náhodné veličiny,

$\boxed{g(\mu_i) = \mu_i = \beta_1 + \beta_2 x_i}$ je identická linkovací funkce, β_1, β_2 a σ^2 jsou neznámé parametry (přičemž σ^2 je rušivým parametrem) a x_i jsou známé kovariáty.

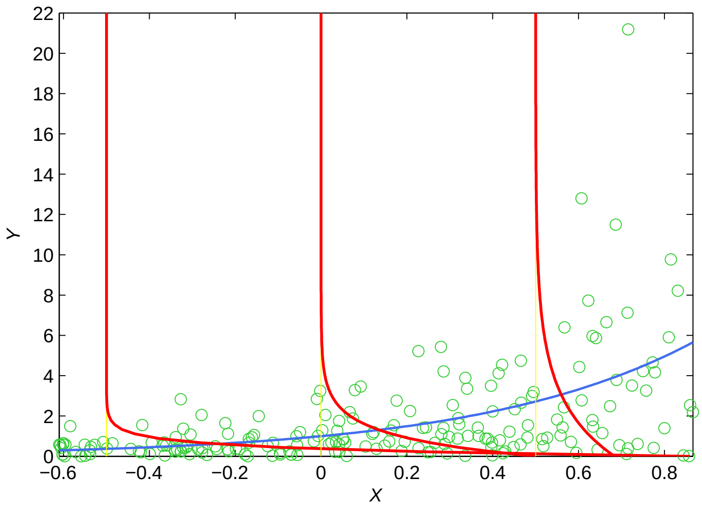


Obrázek 2.1: Ukázka klasického regresního modelu s homogenním rozptylem.

Příklad 2.3.4.
Regresní modely s logaritmickou linkovací funkcí pro exponenciálně a gamma rozdělené závisle proměnné:

$\boxed{Y_i \sim Ex(\lambda_i) \equiv G(1, \lambda_i)}$ jsou pro $i = 1, \ldots, n$ nezávislé náhodné veličiny ($EY_i = \mu_i = \lambda_i$),

$\boxed{g(\mu_i) = \ln \mu_i = \beta_1 + \beta_2 x_i)}$ je logaritmická linkovací funkce, β_1, β_2 jsou neznámé parametry a x_i jsou známé kovariáty.

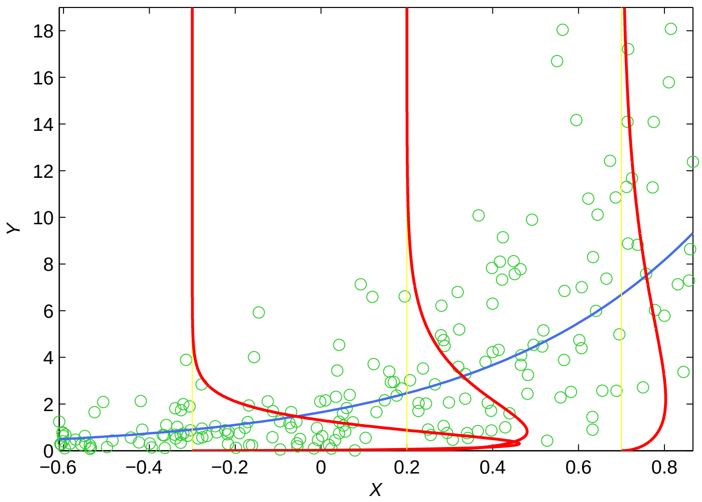


Obrázek 2.2: Ukázka GLM modelu s linkovací funkcí $g(\mu) = \ln \mu$ pro exponenciálně rozdělenou náhodnou veličinu Y.

Jestliže

$\boxed{Y_i \sim G(\alpha, \beta_i = \frac{\mu_i}{\alpha})}$ jsou pro $i = 1, \ldots, n$ nezávislé náhodné veličiny $(EY_i = \mu_i = \alpha\beta_i)$,

$\boxed{g(\mu_i) = \ln \mu_i = \beta_1 + \beta_2 x_i}$ je logaritmická linkovací funkce, β_1, β_2 a $\alpha = \frac{1}{\phi}$ jsou neznámé parametry (α je rušivý parametr) a x_i jsou známé kovariáty.

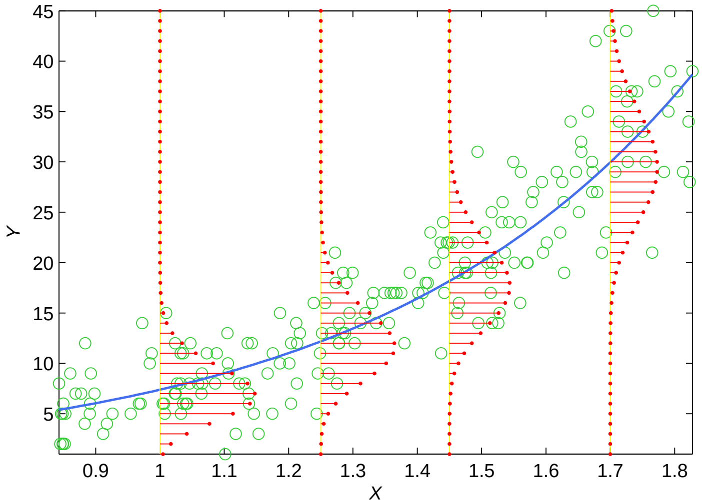


Obrázek 2.3: Ukázka GLM modelu s linkovací funkcí $g(\mu) = \ln \mu$) pro náhodnou veličinu Y s gama rozdělením.

<u>PŘÍKLAD</u> 2.3.5.
Poissonovská regrese:

$\boxed{Y_i \sim Po(\mu_i)}$ jsou pro $i = 1, \ldots, n$ nezávislé náhodné veličiny $(EY_i = \mu_i)$,

$\boxed{g(\mu_i) = \ln \mu_i = \beta_1 + \beta_2 x_i}$ je logaritmická linkovací funkce, β_1, β_2 jsou neznámé parametry a x_i jsou známé kovariáty.



Obrázek 2.4: Ukázka poissonovské regrese s linkovací funkcí $g(\mu) = \ln \mu$.

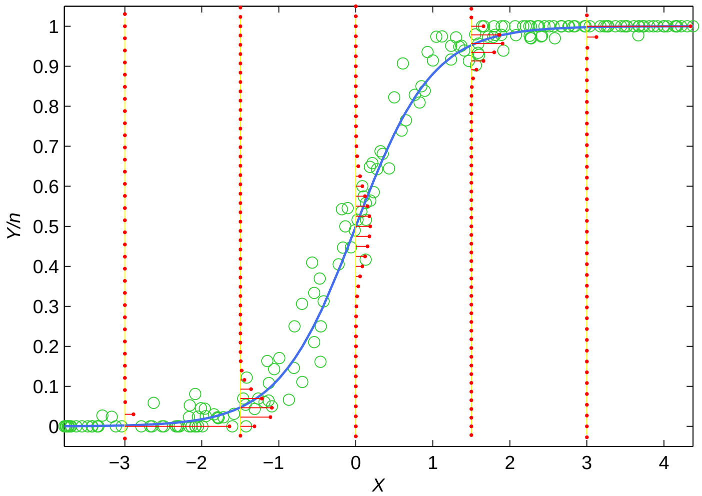
PŘÍKLAD 2.3.6.
Binomická regrese:

$$\boxed{Y_i \sim Bi(n_i, \pi_i)} \text{ jsou pro } i = 1, \ldots, n \text{ nezávislé náhodné veličiny} \quad ,$$

kde $\boxed{g(\pi_i) = \ln\left(\dfrac{\pi_i}{1-\pi_i}\right)}$ je logistická linkovací funkce, β_1, β_2 jsou neznámé parametry a x_i jsou známé kovariáty.

Například ve farmaceutickém experimentu může být n_i počet pacientů, kterým byla podána dávka x_i nového léku a Y_i počet pacientů dávající pozitivní odpověď na danou dávku x_i nového léku.

Jestliže pozorujeme, že $\boxed{\dfrac{Y_i}{n_i}}$ roste spolu s $\boxed{x_i}$, hledáme model, ve kterém π_i je funkcí x_i, ale nabývá hodnot $0 < \pi_i < 1$. Proto model $\pi_i = \beta_1 + \beta_2 x_i$ není vhodný, avšak $\beta_1 + \beta_2 x_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ obvykle pracuje dobře.



Obrázek 2.5: Ukázka binomické regrese s linkovací funkcí $g(\pi) = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$.

PŘÍKLAD 2.3.7.
Kontingenční tabulky: $\boxed{Y_{ij} \sim Mn(n, \pi_{ij})}$ jsou pro $i = 1, \ldots, I$, $j = 1, \ldots, J$, $n = I \cdot J$, $(\sum\limits_{i=1}^{I} \sum\limits_{j=1}^{J} \pi_{ij} = 1)$ nezávislé náhodné veličiny, například počet lidí i-té etnické skupiny, kteří volí politickou stranu j.

Snahou bude testovat hypotézu $\boxed{H_0 : \pi_{ij} = \alpha_i \beta_j}$ pro všechna i, j, kde α_i, β_j jsou neznámé parametry, $\sum_{i=1}^{I} \alpha_i = 1$ a $\sum_{j=1}^{J} \beta_j = 1$, tj. chceme testovat hypotézu, že volba strany a etnická příslušnost jsou nezávislé.

Připomeňme, že	$\mu_{ij} = EY_{ij} = n\pi_{ij}$
takže	$\ln \frac{Y_{ij}}{n} = \ln \pi_{ij}$
a za platnosti hypotézy H_0	$\ln \pi_{ij} = \ln \alpha_i + \ln \beta_j$
ekvivalentně	$\boxed{g(\mu_{ij}) = \ln \mu_{ij} = const_{ij} + a_i + b_j}$ pro nějaké $const_{ij}, a_i, b_j$.

2.4 LOGARITMUS VĚROHODNOSTNÍ FUNKCE V GLM

Všimněme si, že rozdělení náhodných veličin Y_i jsou stejného typu a **logaritmus sdružené věrohodnostní funkce** má tvar

$$l^*(\boldsymbol{\theta};\mathbf{y}) = l^*(\theta_1,\ldots,\theta_n;y_1,\ldots,y_n) = \sum_{i=1}^{n} l_i(\theta_i;y_i) = \sum_{i=1}^{n}\left(\frac{y_i\theta_i-\gamma(\theta_i)}{\psi_i(\phi)}+d(y_i,\phi)\right).$$

Věta 2.4.1.
Mějme náhodný výběr $\mathbf{Y}=(Y_1,\ldots,Y_n)^{\mathrm{T}}$, *který se řídí* ***zobecněným lineárním modelem*** *s linkovací funkcí*

$$\boxed{g(\mu_i)=\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}=\eta_i} \qquad i=1,\ldots,n.$$

Předpokládejme, že pro $i=1,\ldots,n$ *existují příslušné derivace* $\gamma'(\theta_i),\gamma''(\theta_i)$ *a platí*

$$EY_i=\mu_i=\gamma'(\theta_i) \qquad DY_i=\gamma''(\theta_i)\psi_i(\phi).$$

pak

$$\boxed{U_j^*=U_j^*(\boldsymbol{\beta})=\sum_{i=1}^{n}\frac{x_{ij}(Y_i-\mu_i)}{DY_i}\frac{\partial\mu_i}{\partial\eta_i}} \tag{2.4}$$

a

$$\boxed{J_{jk}^*=J_{jk}^*(\boldsymbol{\beta})=\sum_{i=1}^{n}\frac{x_{ij}x_{ik}}{DY_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2}, \tag{2.5}$$

což lze zapsat maticově

$$\boxed{\mathbf{U}_n^*=\mathbf{U}_n^*(\boldsymbol{\beta})=(U_1^*,\ldots,U_m^*)^{\mathrm{T}}=\mathbf{X}^{\mathrm{T}}\mathbf{W}(\boldsymbol{\beta})\mathbf{Q}(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta})=\mathbf{X}^{\mathrm{T}}\mathbf{WQr}} \tag{2.6}$$

a

$$\boxed{\mathbf{J}_n=\mathbf{J}_n(\boldsymbol{\beta})=\left(J_{jk}^*\right)_{j,k=1}^{m}=\mathbf{X}^{\mathrm{T}}\mathbf{W}(\boldsymbol{\beta})\mathbf{X}=\mathbf{X}^{\mathrm{T}}\mathbf{WX}}, \tag{2.7}$$

kde

$$\mathbf{r}=\mathbf{r}(\boldsymbol{\beta})=(r_1(\boldsymbol{\beta}),\ldots,r_n(\boldsymbol{\beta}))^{\mathrm{T}} \qquad r_i=r_i(\boldsymbol{\beta})=Y_i-\mu_i=Y_i-g^{-1}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$$

$$\mathbf{W}=\mathbf{W}(\boldsymbol{\beta})=diag\{w_1(\boldsymbol{\beta}),\ldots,w_n(\boldsymbol{\beta})\} \qquad w_i=w_i(\boldsymbol{\beta})=\frac{1}{DY_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2$$

$$\mathbf{Q}=\mathbf{Q}(\boldsymbol{\beta})=diag\{q_1(\boldsymbol{\beta}),\ldots,q_n(\boldsymbol{\beta})\} \qquad q_i=q_i(\boldsymbol{\beta})=\frac{\partial\eta_i}{\partial\mu_i}.$$

Důkaz.

Podle předpokladů: $EY_i=\mu_i=\gamma'(\theta_i)$ a $DY_i=\gamma''(\theta_i)\psi_i(\phi) \;\Rightarrow\; \gamma''(\theta_i)=\frac{DY_i}{\psi_i(\phi)}$.

Mějme *monotonní diferencovatelnou funkci* g, tzv. **linkovací funkci**, pro kterou platí:

$$\boxed{g(\mu_i)=\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}=\sum_{j=1}^{m}x_{ij}\beta_j=\eta_i}.$$

Pak $\quad EY_i = \mu_i = g^{-1}(\eta_i) = g^{-1}\left(\sum_{j=1}^{m} x_{ij}\beta_j\right) = g^{-1}\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right) = h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) = \mu_i(\boldsymbol{\beta}),$

kde $\quad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^{\mathrm{T}}$ je vektor neznámých parametrů a

$\mathbf{x}_i = (x_{i1}, \ldots, x_{im})^{\mathrm{T}}$ je vektor hodnot regresorů pro i-té pozorování,

Spočítejme $\quad U_j^*(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n}\left(\frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}\right) \quad$ pro $\quad j = 1, \ldots, m.$

Abychom získali $U_j^*(\boldsymbol{\beta})$, vyjděme ze vztahů

$$\frac{\partial l_i}{\partial \theta_i} = \frac{Y_i - \gamma'(\theta_i)}{\psi_i(\phi)} = \frac{Y_i - \mu_i}{\psi_i(\phi)} = \frac{Y_i - h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{\psi_i(\phi)}$$

$$\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \sum_{k=1}^{m} x_{ik}\beta_k}{\partial \beta_j} = x_{ij}$$

$$\frac{\partial \mu_i}{\partial \theta_i} = \gamma''(\theta_i) = \frac{DY_i}{\psi_i(\phi)}$$

a upravujme $\quad \frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{x_{ij}(Y_i - \mu_i)}{DY_i}\frac{\partial \mu_i}{\partial \eta_i}$

a odtud již plyne první tvrzení věty: $\boxed{U_j^* = U_j^*(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{x_{ij}(Y_i - \mu_i)}{DY_i}\frac{\partial \mu_i}{\partial \eta_i}}.$

Označme **informační matici**:

$$\mathbf{J}_n = \mathbf{J}_n(\boldsymbol{\beta}) = \left(J_{jk}^*\right)_{j,k=1}^{m} = \left(E(U_j^* U_k^*)\right)_{j,k=1}^{m} = D\mathbf{U}_n^*(\boldsymbol{\beta}),$$

kde $\quad \mathbf{U}_n^* = \mathbf{U}_n^*(\boldsymbol{\beta}) = (U_1^*, \ldots, U_m^*)^{\mathrm{T}}.$

Protože $\quad E\left(\frac{\partial l_i}{\partial \beta_j} \cdot \frac{\partial l_i}{\partial \beta_k}\right) = \frac{x_{ij}x_{ik}E(Y_i - \mu_i)^2}{(DY_i)^2}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \frac{x_{ij}x_{ik}}{DY_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$

pak $\quad \boxed{J_{jk}^* = J_{jk}^*(\boldsymbol{\beta})} = \sum_{i=1}^{n} E\left(\frac{\partial l_i}{\partial \beta_j} \cdot \frac{\partial l_i}{\partial \beta_k}\right) = \boxed{\sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{DY_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}.$

Zaveďme <u>maticové vyjádření</u> následujícím způsobem:

Označme $\quad w_i = w_i(\boldsymbol{\beta}) = \frac{1}{DY_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \quad$ a $\quad \mathbf{W} = \mathbf{W}(\boldsymbol{\beta}) = diag\{w_1(\boldsymbol{\beta}), \ldots, w_n(\boldsymbol{\beta})\},$

neboť $\quad \eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} \quad$ a $\quad \mu_i = h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \quad$ jsou funkcemi $\boldsymbol{\beta}$.

Pak $\quad \boxed{\mathbf{J}_n = \mathbf{X}^{\mathrm{T}}\mathbf{W}(\boldsymbol{\beta})\mathbf{X} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}},$

kde $\quad \mathbf{X} = (\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_n^{\mathrm{T}})^{\mathrm{T}} = (\mathbf{X}_1, \ldots, \mathbf{X}_m) \quad$ a $\quad \mathbf{X}_j = (x_{1j}, \ldots, x_{nj})^{\mathrm{T}}.$

Označme $\quad r_i = r_i(\boldsymbol{\beta}) = Y_i - \mu_i = Y_i - g^{-1}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \quad$ a $\quad \mathbf{r} = \mathbf{r}(\boldsymbol{\beta}) = (r_1(\boldsymbol{\beta}), \ldots, r_n(\boldsymbol{\beta}))^{\mathrm{T}}$

$q_i = q_i(\boldsymbol{\beta}) = \frac{\partial \eta_i}{\partial \mu_i} \quad$ a $\quad \mathbf{Q} = \mathbf{Q}(\boldsymbol{\beta}) = diag\{q_1(\boldsymbol{\beta}), \ldots, q_n(\boldsymbol{\beta})\},$

pak $\quad \boxed{\mathbf{U}_n^* = \mathbf{U}_n^*(\boldsymbol{\beta}) = (U_1^*, \ldots, U_m^*)^{\mathrm{T}} = \mathbf{X}^{\mathrm{T}}\mathbf{W}(\boldsymbol{\beta})\mathbf{Q}(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta}) = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Q}\mathbf{r}} \quad$ a

$\boxed{U_j^*} = \sum_{i=1}^{n} \frac{x_{ij}(Y_i - \mu_i)}{DY_i}\frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{n} x_{ij}r_i w_i q_i = \boxed{\mathbf{X}_j\mathbf{W}(\boldsymbol{\beta})\mathbf{Q}(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta}) = \mathbf{X}_j^{\mathrm{T}}\mathbf{W}\mathbf{Q}\mathbf{r}}. \qquad \square$

2.5 MAXIMÁLNĚ VĚROHODNÉ ODHADY NEZNÁMÝCH PARAMETRŮ

Odhad neznámých parametrů metodou **maximální věrohodnosti** dostaneme řešením rovnic typu

$$\boxed{\frac{\partial l^*}{\partial \boldsymbol{\beta}} = \mathbf{U}_n^*(\boldsymbol{\beta}) = \mathbf{0}}$$

Aby šlo o maximum, je nutné, aby **matice druhých parciálních derivací logaritmické věrohodnostní funkce podle složek parametru $\boldsymbol{\beta}$ byla negativně definitní**.

Podle věty 1.4.13 konverguje matice druhých parciálních derivací skoro jistě k matici $-\mathbf{J}_n$, která je při regularitě systému hustot negativně definitní.

Aproximujeme-li proto matici $\frac{\partial \mathbf{U}_n^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ maticí $-\mathbf{J}_n$, je řešení systému předešlých rovnic maximálně věrohodným odhadem parametru $\boldsymbol{\beta}$.

Ukažme ještě, že tato aproximace je při **kanonickém linku přesná**, tj. matice

$$\boxed{\mathbf{U}_n^{*'}(\boldsymbol{\beta}) = \frac{\partial \mathbf{U}_n^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\mathbf{J}_n(\boldsymbol{\beta})}.$$

Použijeme-li **kanonický link**, tj.

$$\theta_i = \eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} = \sum_{j=1}^{n} x_{ij}\beta_j$$

a vztahů

$$\frac{\partial \mu_i}{\partial \theta_i} = \gamma''(\theta_i) = V(\mu_i)$$
$$DY_i = \psi_i(\phi)\gamma''(\theta_i) = \psi_i(\phi)V(\mu_i)$$

pak

$$\boxed{U_j^*} = \sum_{i=1}^{n} U_{j,i} = \sum_{i=1}^{n} \frac{x_{ij}(Y_i - \mu_i)}{DY_i}\frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{n} \frac{x_{ij}(Y_i - \mu_i)}{\psi_i(\phi)V(\mu_i)}\underbrace{\frac{\partial \mu_i}{\partial \theta_i}}_{=V(\mu_i)} = \boxed{\sum_{i=1}^{n} \frac{x_{ij}(Y_i - \mu_i)}{\psi_i(\phi)}}$$

$$\boxed{J_{jk}^*} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{DY_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\psi_i(\phi)V(\mu_i)}V^2(\mu_i) = \boxed{\sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\psi_i(\phi)}V(\mu_i)}$$

$$\boxed{\frac{\partial U_j^*}{\partial \beta_k}} = \sum_{i=1}^{n} \frac{\partial U_{j,i}}{\partial \beta_k} = \sum_{i=1}^{n} \frac{\partial U_{j,i}}{\partial \mu_i}\underbrace{\frac{\partial \mu_i}{\partial \eta_i}}_{=V(\mu_i)}\frac{\partial \eta_i}{\partial \beta_k} = \sum_{i=1}^{n} -\frac{x_{ij}}{\psi_i(\phi)}V(\mu_i)x_{ik} = -\sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\psi_i(\phi)}V(\mu_i) = \boxed{-J_{jk}^*}.$$

Nyní se vraťme k <u>řešení věrohodnostních rovnic</u>. Protože obecně rovnice

$$U_j^* = \frac{\partial l^*}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{x_{ij}(Y_i - \mu_i)}{DY_i}\frac{\partial \mu_i}{\partial \eta_i} = 0 \quad j = 1, \ldots, m.$$

nejsou lineární vzhledem k neznámým parametrům, musí se řešit numerickou iterací.

2.5.1 NEWTON-RAPHSONOVA METODA

Chceme-li najít řešení systému nelineárních rovnic $\mathbf{U}_n^*(\boldsymbol{\beta}) = \mathbf{0}$, lze použít následující iterativní postup:

1. Nejprve provedeme **linearizaci pomocí Taylorova rozvoje** v okolí bodu $\boldsymbol{\beta}_0$, kde $\boldsymbol{\beta}_0$ je nějaký počáteční odhad: $\mathbf{U}_n^*(\boldsymbol{\beta}) \approx \mathbf{U}_n^*(\boldsymbol{\beta}_0) + \mathbf{U}_n^{*'}(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. Protože $\mathbf{U}_n^*(\boldsymbol{\beta}) = 0$, pak po jednoduchých úpravách dostaneme $\boxed{\boldsymbol{\beta} \approx \boldsymbol{\beta}_0 - \left[\mathbf{U}_n^{*'}(\boldsymbol{\beta}_0)\right]^{-1} \mathbf{U}_n^*(\boldsymbol{\beta}_0)}$.

2. Odhady parametrů v s-tém kroku jsou získány ze vztahu
$$\widehat{\boldsymbol{\beta}}^{(s)} = \widehat{\boldsymbol{\beta}}^{(s-1)} - \left[\mathbf{U}_n^{*'}\left(\widehat{\boldsymbol{\beta}}^{(s-1)}\right)\right]^{-1} \mathbf{U}_n^*\left(\widehat{\boldsymbol{\beta}}^{(s-1)}\right).$$

3. Iterační proces popsaný v předchozím bodě pokračuje tak dlouho, dokud
$$\widehat{\boldsymbol{\beta}}^{(s+1)} - \widehat{\boldsymbol{\beta}}^{(s)} \approx 0.$$

2.5.2 METODA SKÓROVÁNÍ

Alternativní procedurou k Newton-Raphsonově metodě je tzv. **metoda skórování**, kdy se matice druhých parciálních derivací $\mathbf{U}_n^{*'}(\boldsymbol{\beta})$ nahradí její střední hodnotou, tj. maticí $-\mathbf{J}_n(\boldsymbol{\beta})$, kde $\mathbf{J}_n(\boldsymbol{\beta})$, je **informační matice**.

Druhý iterační krok pak upravíme takto:
$$\widehat{\boldsymbol{\beta}}^{(s)} = \widehat{\boldsymbol{\beta}}^{(s-1)} + \left[\mathbf{J}_n(\widehat{\boldsymbol{\beta}}^{(s-1)})\right]^{-1} \mathbf{U}_n^*(\widehat{\boldsymbol{\beta}}^{(s-1)}).$$

Jednoduchou úpravou dostaneme
$$\mathbf{J}_n(\widehat{\boldsymbol{\beta}}^{(s-1)})\widehat{\boldsymbol{\beta}}^{(s)} = \mathbf{J}_n(\widehat{\boldsymbol{\beta}}^{(s-1)})\widehat{\boldsymbol{\beta}}^{(s-1)} + \mathbf{U}_n^*(\widehat{\boldsymbol{\beta}}^{(s-1)}).$$

Využijme vztahů:
$$\mathbf{U}_n^*(\boldsymbol{\beta}) = \mathbf{X}^{\mathrm{T}}\mathbf{W}(\boldsymbol{\beta})\mathbf{Q}(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta}) \quad \text{a} \quad \mathbf{J}_n(\boldsymbol{\beta}) = \mathbf{X}^{\mathrm{T}}\mathbf{W}(\boldsymbol{\beta})\mathbf{X}$$

a dále upravujme
$$\begin{aligned}\mathbf{X}^{\mathrm{T}}\mathbf{W}(\widehat{\boldsymbol{\beta}}^{(s-1)})\mathbf{X}\widehat{\boldsymbol{\beta}}^{(s)} &= \mathbf{X}^{\mathrm{T}}\mathbf{W}(\widehat{\boldsymbol{\beta}}^{(s-1)})\mathbf{X}\widehat{\boldsymbol{\beta}}^{(s-1)} + \mathbf{X}^{\mathrm{T}}\mathbf{W}(\widehat{\boldsymbol{\beta}}^{(s-1)})\mathbf{Q}(\widehat{\boldsymbol{\beta}}^{(s-1)})\mathbf{r}(\widehat{\boldsymbol{\beta}}^{(s-1)}) \\ &= \mathbf{X}^{\mathrm{T}}\mathbf{W}(\widehat{\boldsymbol{\beta}}^{(s-1)}) \underbrace{\left[\mathbf{X}\widehat{\boldsymbol{\beta}}^{(s-1)} + \mathbf{Q}(\widehat{\boldsymbol{\beta}}^{(s-1)})\mathbf{r}(\widehat{\boldsymbol{\beta}}^{(s-1)})\right]}_{\mathbf{Z}(\widehat{\boldsymbol{\beta}}^{(\mathbf{s-1})})}\end{aligned}$$

přičemž pro $i = 1, \dots, n$
$$Z_i(\widehat{\boldsymbol{\beta}}^{(s-1)}) = \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{(s-1)} + r_i^{(s-1)}q_i^{(s-1)} = \sum_{j=1}^{m} x_{ij}\hat{\beta}_j^{(s-1)} + \left(Y_i - \hat{\mu}_i^{(s-1)}\right)\frac{\partial\hat{\eta}_i^{(s-1)}}{\partial\hat{\mu}_i^{(s-1)}}.$$

Můžeme psát
$$\boxed{\mathbf{X}^{\mathrm{T}}\mathbf{W}(\widehat{\boldsymbol{\beta}}^{(s-1)})\mathbf{X}\widehat{\boldsymbol{\beta}}^{(s)} = \mathbf{X}^{\mathrm{T}}\mathbf{W}(\widehat{\boldsymbol{\beta}}^{(s-1)})\mathbf{Z}(\widehat{\boldsymbol{\beta}}^{(s-1)})} \quad \text{nebo} \quad \boxed{\mathbf{X}^{\mathrm{T}}\mathbf{W}^{(s-1)}\mathbf{X}\widehat{\boldsymbol{\beta}}^{(s)} = \mathbf{X}^{\mathrm{T}}\mathbf{W}^{(s-1)}\mathbf{Z}^{(s-1)}},$$

kde
$$\mathbf{W}^{(s-1)} = diag\{w_1^{(s-1)}, \dots, w_n^{(s-1)}\}$$
a
$$w_i^{(s-1)} = \frac{1}{\widehat{DY}_i^{(s-1)}} \left(\frac{\partial \hat{\mu}_i^{(s-1)}}{\partial \hat{\eta}_i^{(s-1)}}\right)^2 = \frac{\omega_i}{\phi V(\hat{\mu}_i^{(s-1)}) \left(q_i^{(s-1)}\right)^2},$$
kde
$$\widehat{DY}_i^{(s-1)} = \psi(\phi) V\left(\hat{\mu}_i^{(s-1)}\right) = \frac{\phi}{\omega_i} V(\hat{\mu}_i^{(s-1)}).$$

Jde o obdobu vážené metody nejmenších čtverců a v tomto případě mluvíme o **iterační vážené metodě nejmenších čtverců**.

2.6 TESTOVÁNÍ HYPOTÉZ

Statistické modely jsou obvykle konstruovány s cílem rozhodnout o předem definované hypotéze a tuto přijmout či vyvrátit

2.6.1 Testování parametrů

Věta 2.6.1.
*Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \dots, Y_n)^{\mathrm{T}}$, který se řídí **zobecněným lineárním modelem** s maticí vysvětlujících proměnných $\mathbf{X}_{n\times m}$ Předpokládejme, že pro $i = 1, \dots, n$ existují příslušné derivace $\gamma'(\theta_i), \gamma''(\theta_i)$ a platí*
$$EY_i = \mu_i = \gamma'(\theta_i) \qquad DY_i = \gamma''(\theta_i)\psi_i(\phi).$$
*Dále mějme matici $\mathbf{C}_{m\times q}$ s hodností $h(\mathbf{C}) = q < m$. Platí-li hypotéza: $\boxed{H_0 : \mathbf{C}^{\mathrm{T}}\boldsymbol{\beta} = 0}$, pak **Waldova statistika***
$$\boxed{W = \widehat{\boldsymbol{\beta}}^{\mathrm{T}}_{\mathrm{MLE}}\mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta})^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}} \quad \overset{A}{\sim} \quad \chi^2(q)},$$
kde $\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ je maximálně věrohodným odhadem vektorového parametru $\boldsymbol{\beta}$.

Důkaz.
MLE odhad vektorové parametrické funkce $\mathbf{C}^{\mathrm{T}}\boldsymbol{\beta}$ je roven $\mathbf{C}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$. Vektor $\mathbf{C}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ je lineární transformací náhodného vektoru $\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$, který má asymptoticky normální rozdělení:
$$\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}} \quad \overset{A}{\sim} \quad N_m(\boldsymbol{\beta}, \mathbf{J}_n(\boldsymbol{\beta})^{-1}),$$
proto také
$$\mathbf{C}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}} \quad \overset{A}{\sim} \quad N_q(\mathbf{C}^{\mathrm{T}}\boldsymbol{\beta}, \mathbf{C}^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta})^{-1}\mathbf{C}),$$
neboť hodnost matice $h\left(\mathbf{C}^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta})^{-1}\mathbf{C}\right) = q$.

Zbytek důkazu plyne z vlastnosti kvadratických forem normálních vektorů. □

Poznámka 2.6.2.
Hypotézu

$$\boxed{H_0 : \mathbf{C}^{\mathrm{T}}\boldsymbol{\beta} = 0}$$

zamítáme na hladině významnosti α, pokud platí

$$\boxed{W > \chi^2_{1-\alpha}(q)}.$$

Protože odhad $\widehat{\boldsymbol{\beta}}_{\text{MLE}}$ konverguje za předpokladu existence $E(l^*(\boldsymbol{\beta}))$ skoro jistě k $\boldsymbol{\beta}$, aproximujeme při výpočtu Waldovy statistiky W Fisherovou informační matici $\mathbf{J}_n(\boldsymbol{\beta})$ maticí $\mathbf{J}_n(\widehat{\boldsymbol{\beta}}_{\text{MLE}})$.

Poznámka 2.6.3.
Testovat hypotézu

$$\boxed{H_O : \beta_j = 0}$$

pro $j = 1, \ldots, m$ lze více způsoby:

- Pomocí Waldovy statistiky W, a to při speciální volbě

$$\mathbf{C} = \mathbf{c}_{m\times 1} = (0, \ldots, \overset{j}{1}, \ldots, 0)^{\mathrm{T}}.$$

- Pomocí vztahu

$$\boxed{\hat{\beta}_{MLE,j} \quad \overset{A}{\sim} \quad N\left(\beta_j, s^\star_{jj} = \left(\mathbf{J}_n(\boldsymbol{\beta})^{-1}\right)_{jj}\right)},$$

  přičemž hypotézu **zamítáme**, pokud

$$\frac{|\hat{\beta}_{MLE,i}|}{\sqrt{s^\star_{jj}}} > u_{1-\frac{\alpha}{2}},$$

  kde opět Fisherovou informační matici $\mathbf{J}_n(\boldsymbol{\beta})$ aproximujeme maticí $\mathbf{J}_n(\widehat{\boldsymbol{\beta}}_{\text{MLE}})$.

2.7 OVĚŘOVÁNÍ VHODNOSTI MODELU

2.7.1 Minimální, maximální model a submodely

Určení vhodné modelové rovnice je základem všech regresních modelů. Jedním z důležitých principů regresních modelů je **zásada jednoduchosti**, která znamená, že jednodušší model poměrně dobře popisující zkoumaná data dostane přednost před složitějším modelem, který data popisuje téměř dokonale.

Často musíme vzít také v úvahu současně se základním zobecněným lineárním modelem i několik z něj vyplývajících dílčích modelů, kterým se říká submodely.

Definujme nejprve důležité pojmy

Definice 2.7.1.
Maximální GLM*, který označíme GLM_{max}, splňuje následující podmínky*

(1) Maximální model je zobecněný lineární model se stejným typem rozdělení jako zkoumaný GLM model.

(2) Maximální model a zkoumaný mají stejnou linkovací funkci.

(3) Počet parametrů maximálního modelu je roven počtu vysvětlovaných veličin n, maximálně věrohodný odhad parametru $\boldsymbol{\beta}_{max}$ je n-rozměrný vektor $\widehat{\boldsymbol{\beta}}_{max}$.

POZNÁMKA 2.7.2.
Z definice plyne, že vysvětlovaná veličina $\mathbf{Y}$ je maximálním modelem určena s nulovým reziduem, tj. odhadnutá hodnota

$$\widehat{\mathbf{Y}}_{max} = \widehat{\boldsymbol{\mu}}_{max} = (\hat{\mu}_{max,1}, \ldots, \hat{\mu}_{max,n})^{\mathrm{T}} = \mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}.$$

Definice 2.7.3.
***Minimální GLM**, který označíme GLM_{min}, splňuje následující podmínky*

(1) Minimální model je zobecněný lineární model se stejným typem rozdělení jako zkoumaný GLM model.

(2) Minimální model a zkoumaný mají stejnou linkovací funkci.

(3) Počet parametrů minimálního modelu je roven 1, maximálně věrohodný odhad parametru $\boldsymbol{\beta}_{min}$ je skalár $\hat{\beta}_{min}$.

POZNÁMKA 2.7.4.
Pro minimální model, kde $\mathbf{X} = (1, \ldots, 1)^{\mathrm{T}}$, lze snadno ověřit, že

$$\hat{\mu}_{min,i} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad i = 1, \ldots, n.$$

Maximální model tedy slouží jako ukazatel þnejlepšíß regrese a minimální model naopak jako ukazatel þnejhoršíß regrese při daném rozdělení a dané linkovací funkci. Zkoumaný model se bude nacházet někde mezi těmito extrémy a ve srovnání s nimi budeme oceňovat vhodnost modelu.

Definice 2.7.5.
Mějme zobecněný lineární model s maticí plánu $\mathbf{X}_{n\times m}$ a vektorem neznámých parametrů $\boldsymbol{\beta}$.
***Submodel**, který označíme GLM_{sub}, splňuje následující podmínky*

(1) Submodel je zobecněný lineární model se stejným typem rozdělení jako zkoumaný GLM model.

(2) Submodel a zkoumaný model mají stejnou linkovací funkci.

(3) Vektor neznámých parametrů $\boldsymbol{\beta}_{sub} \in \mathbb{R}^q$ a matice plánu $\mathbf{Q}_{n\times q}$, pro kterou platí

$$\mathbf{Q}_{n\times q} = \mathbf{X}_{n\times m}\mathbf{T}_{m\times q}.$$

Aby GLM_{sub} byl submodelem modelu GLM, musí každý sloupec matice $\mathbf{Q}$ patřit do obalu sloupců matice $\mathbf{X}$. To bude splněno právě tehdy, bude-li $\mathbf{Q}$ typu

$$\mathbf{Q}_{n\times q} = \mathbf{X}_{n\times m}\mathbf{T}_{m\times q}.$$

Je třeba si uvědomit, že GLM_{sub} je speciálním případem modelu GLM. Platí-li tudíž pro náhodný výběr $\mathbf{Y}$ model GLM_{sub}, platí pro $\mathbf{Y}$ také model GLM.

Model GLM vybíráme tak bohatý, abychom si mohli být jisti, že popisuje dobře chování $\mathbf{Y}$.

Následně bychom ovšem chtěli vědět, zda lze použít jednodušší model GLM_{sub}. Můžeme usuzovat takto, platí-li GLM_{sub}, pak rozšíření na GLM nepřinese podstatné změny a vektory $\widehat{\boldsymbol{\beta}}$ a $\widehat{\boldsymbol{\beta}}_{sub}$ by se neměly podstatně lišit. Na druhé straně, budou-li $\widehat{\boldsymbol{\beta}}$ a $\widehat{\boldsymbol{\beta}}_{sub}$ příliš odlišné, svědčí to proti možnosti redukce GLM na GLM_{sub}.

2.7.2 DEVIACE

Definice 2.7.6.
Mějme modely GLM a GLM_{max}. Nechť náhodný výběr $\mathbf{Y}$ se řídí modelem GLM_{max}. ***Škálovou deviací modelu* GLM** *(scaled deviance) rozumíme statistiku*

$$D = 2\left[l^*(\widehat{\boldsymbol{\beta}}_{max};\mathbf{Y}) - l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y})\right],$$

kde $\widehat{\boldsymbol{\beta}}_{max}, \widehat{\boldsymbol{\beta}}$ jsou odpovídající maximálně věrohodné odhady.

Lemma 2.7.7.
Nechť je dán model GLM a náhodný výběr $\mathbf{Y}$ se řídí tímto modelem. Dále nechť

(i) existují druhé parciální derivace hustoty $f(\mathbf{y};\boldsymbol{\beta})$ podle složek $\boldsymbol{\beta}$,

(ii) platí $E\left(\frac{f''_{\beta_i\beta_j}(\mathbf{Y};\boldsymbol{\beta})}{f(\mathbf{Y};\boldsymbol{\beta})}\right) = 0 \quad (i,j = 1,\dots,m)$

(iii) a existuje $\quad E\, l^*(\boldsymbol{\beta};\mathbf{Y})$.

Pak asymptoticky lze statistiku

$$D_m = 2\left[l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y}) - l^*(\boldsymbol{\beta};\mathbf{Y})\right]$$

aproximovat kvadratickou formou

$$W_m = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

a rozdělení statistiky D_m lze aproximovat rozdělením $\chi^2(m)$, přičemž $\widehat{\boldsymbol{\beta}}$ je maximálně věrohodný odhad parametru $\boldsymbol{\beta}$.

Důkaz. Aproximujme funkci $l^*(\boldsymbol{\beta};\mathbf{Y})$ Taylorovým rozvojem druhého řadu v bodě maximálně věrohodného odhadu $\widehat{\boldsymbol{\beta}}$

$$l^*(\boldsymbol{\beta};\mathbf{Y}) \approx l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y}) + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^{\mathrm{T}}\underbrace{\mathbf{U}_n^*(\widehat{\boldsymbol{\beta}})}_{=0} + \frac{1}{2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^{\mathrm{T}}\underbrace{\mathbf{U}_n^{*\prime}(\widehat{\boldsymbol{\beta}})}_{\to -\mathbf{J}_n(\boldsymbol{\beta})}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}),$$

kde $\mathbf{U}_n^*$ je skórový vektor.

Protože odhad $\widehat{\boldsymbol{\beta}}$ je maximálně věrohodný, pak platí

$$\mathbf{U}_n^*(\widehat{\boldsymbol{\beta}}) = 0.$$

Díky podmínce (ii) matice $\mathbf{U}_n^{*\prime}(\boldsymbol{\beta})$ konverguje skoro jistě k Fisherově informační matici $-\mathbf{J}_n(\boldsymbol{\beta})$.

Dále protože maximálně věrohodný odhad $\widehat{\boldsymbol{\beta}}$ konverguje podle věty 1.4.23 skoro jistě k $\boldsymbol{\beta}$, nahradíme při $n \to \infty$ matici $\mathbf{U}_n^{*\prime}(\widehat{\boldsymbol{\beta}})$ maticí $-\mathbf{J}_n(\boldsymbol{\beta})$, tedy

$$2\left[l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y}) - l^*(\boldsymbol{\beta};\mathbf{Y})\right] \approx W_m = (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta})(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}),$$

čímž je dokázáno první tvrzení lemma.

Dále podle věty 1.4.23 má statistika $W_m \overset{A}{\sim} \chi^2(m)$, proto platí druhá část lemma. □

Věta 2.7.8.
Mějme modely GLM a GLM_{max}. Nechť náhodný výběr $\mathbf{Y}$ se řídí modelem GLM_{max}. Dále nechť

(i) existují druhé parciální derivace hustoty $f(\mathbf{y};\boldsymbol{\beta}_{max})$ podle složek $\boldsymbol{\beta}_{max}$,

(ii) platí $E\left(\frac{f''_{\beta_{(max,i)}\beta_{(max,j)}}(\mathbf{Y};\boldsymbol{\beta}_{max})}{f(\mathbf{Y};\boldsymbol{\beta}_{max})}\right) = 0 \quad (i,j = 1,\ldots,n)$

(iii) a existuje $E\, l^*(\boldsymbol{\beta}_{max};\mathbf{Y})$.

*Platí-li hypotéza, že **model** GLM **je vhodný**, pak asymptoticky lze rozdělení škálové deviace modelu GLM aproximovat rozdělením $\chi^2(n-m)$, tj.*

$$\boxed{D \quad \overset{A}{\sim} \quad \chi^2(n-m)}.$$

Důkaz. Rozložme škálovou deviaci D modelu GLM na tři složky

$$D = \underbrace{2\left[l^*(\widehat{\boldsymbol{\beta}}_{max};\mathbf{Y}) - l^*(\boldsymbol{\beta}_{max};\mathbf{Y})\right]}_{D_{max}\overset{A}{\sim}\chi^2(n)\,-\,viz.lemma\ 2.7.7} - \underbrace{2\left[l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y}) - l^*(\boldsymbol{\beta};\mathbf{Y})\right]}_{D_m\overset{A}{\sim}\chi^2(m)\,-\,viz.lemma\ 2.7.7} + \underbrace{2\left[l^*(\boldsymbol{\beta}_{max};\mathbf{Y}) - l^*(\boldsymbol{\beta};\mathbf{Y})\right]}_{\approx 0},$$

Za předpokladu, že model GLM je vhodný, bude hodnota třetího členu kladná konstanta blízká nule.

Parametr $\boldsymbol{\beta}$ modelu GLM je podvektorem $\boldsymbol{\beta}_{max}$, proto existuje matice $\mathbf{C}_{n\times m},\ m < n$ taková že

$$\boldsymbol{\beta} = \mathbf{C}^{\mathrm{T}}\boldsymbol{\beta}_{max} \qquad \text{a} \qquad h(\mathbf{C}) = m.$$

Vektor $\mathbf{C}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{max}$ je lineární transformací vektoru $\widehat{\boldsymbol{\beta}}_{max}$, který má asymptoticky normální rozdělení: $\widehat{\boldsymbol{\beta}}_{max} \overset{A}{\sim} N_n(\boldsymbol{\beta}_{max}, \mathbf{J}_n(\boldsymbol{\beta}_{max})^{-1})$, proto také $\mathbf{C}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{max} \overset{A}{\sim} N_m(\mathbf{C}^{\mathrm{T}}\boldsymbol{\beta}_{max}, \underbrace{\mathbf{C}^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta}_{max})^{-1}\mathbf{C}}_{=\mathbf{V}})$. Podle lemma 2.7.7

$$D_{max} \approx W_{max} = (\widehat{\boldsymbol{\beta}}_{max} - \boldsymbol{\beta}_{max})^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta}_{max})(\widehat{\boldsymbol{\beta}}_{max} - \boldsymbol{\beta}_{max}) \overset{A}{\sim} \chi^2(n)$$

a také pro D_m platí (vyjádříme-li vektor $\boldsymbol{\beta}$ pomocí $\boldsymbol{\beta}_{max}$)

$$D_m \approx W_m = (\widehat{\boldsymbol{\beta}}_{max} - \boldsymbol{\beta}_{max})^{\mathrm{T}} \mathbf{C} \underbrace{\left(\mathbf{C}^{\mathrm{T}} \mathbf{J}_n(\boldsymbol{\beta}_{max})^{-1} \mathbf{C}\right)^{-1}}_{V^{-1}} \mathbf{C}^{\mathrm{T}} (\widehat{\boldsymbol{\beta}}_{max} - \boldsymbol{\beta}_{max}) \overset{A}{\sim} \chi^2(m)$$

neboť hodnost matice $h\left(\mathbf{C}^{\mathrm{T}} \mathbf{J}_n(\boldsymbol{\beta}_{max})^{-1} \mathbf{C}\right) = m$. Takže asympoticky lze aproximovat škálovou deviaci D rozdílem dvou kvadratických forem

$$\begin{aligned} D \approx & D_{max} - D_m = \\ & (\widehat{\boldsymbol{\beta}}_{max} - \boldsymbol{\beta}_{max})^{\mathrm{T}} \underbrace{\left(\mathbf{J}_n(\boldsymbol{\beta}_{max}) - \mathbf{C}\left(\mathbf{C}^{\mathrm{T}} \mathbf{J}_n(\boldsymbol{\beta}_{max})^{-1} \mathbf{C}\right)^{-1} \mathbf{C}^{\mathrm{T}}\right)}_{=\mathbf{G}>0} (\widehat{\boldsymbol{\beta}}_{max} - \boldsymbol{\beta}_{max}). \end{aligned}$$

Matice $\mathbf{G}$ je pozitivně definitní, neboť

$$\begin{aligned} \mathbf{G} = & \mathbf{J}_n - \mathbf{C}\left(\mathbf{C}^{\mathrm{T}} \mathbf{J}_n^{-1} \mathbf{C}\right)^{-1} \mathbf{C}^{\mathrm{T}} = \\ & \left(\mathbf{I}_n - \mathbf{C}\left(\mathbf{C}^{\mathrm{T}} \mathbf{J}_n^{-1} \mathbf{C}\right)^{-1} \mathbf{C}^{\mathrm{T}} \mathbf{J}_n^{-1}\right) \underbrace{\mathbf{J}_n}_{>0 \Rightarrow \mathbf{J}_n = \mathbf{B}\mathbf{B}^{\mathrm{T}}} \left(\mathbf{I}_n - \mathbf{C}\left(\mathbf{C}^{\mathrm{T}} \mathbf{J}_n^{-1} \mathbf{C}\right)^{-1} \mathbf{C}^{\mathrm{T}} \mathbf{J}_n^{-1}\right)^{\mathrm{T}} > 0 \end{aligned}$$

Platí-li hypotéza, že model GLM je vhodný, pak podle věty 2.9 v knize Seber, G.A.F: Lineární regresní analýza, str. 45, lze rozdělení deviace modelu GLM aproximovat rozdělením $\chi^2(n-m)$.□
<u>Poznámka</u> 2.7.9.
Škálovou deviaci D lze užít k testování hypotézy o vhodnosti modelu GLM. Jestliže platí

$$D > \chi^2_{1-\alpha}(n-m),$$

pak považujeme model GLM za nevhodný.

Věta 2.7.10.
Mějme základní model GLM s $\boldsymbol{\beta} \in \mathbb{R}^m$ a jeho submodel GLM_{sub} s $\boldsymbol{\beta}_{sub} \in \mathbb{R}^q$, přičemž $q < m < n$. Dále nechť náhodný výběr $\mathbf{Y}$ se řídí modelem GLM a platí

(i) existují druhé parciální derivace hustoty $f(\mathbf{y};\boldsymbol{\beta})$ podle složek $\boldsymbol{\beta}$,

(ii) platí $E\left(\frac{f''_{\beta_i\beta_j}(\mathbf{y};\boldsymbol{\beta})}{f(\mathbf{y};\boldsymbol{\beta})}\right) = 0 \qquad (i,j = 1,\dots,m)$

(iii) a existuje $\quad E\, l^*(\boldsymbol{\beta};\mathbf{Y})$.

*Platí-li hypotéza, že **submodel** $\mathbf{GLM_{sub}}$ **je vhodný**, pak asymptoticky lze rozdělení statistiky $\Delta D = D_{sub} - D$ aproximovat rozdělením $\chi^2(m-q)$, tj.* $\boxed{\Delta D = D_{sub} - D \overset{A}{\sim} \chi^2(m-q)}$.

Důkaz. Statistiku ΔD je možno rozepsat takto

$$\begin{aligned} \Delta D = & 2\left[l^*(\widehat{\boldsymbol{\beta}}_{max};\mathbf{Y}) - l^*(\widehat{\boldsymbol{\beta}}_{sub};\mathbf{Y})\right] - 2\left[l^*(\widehat{\boldsymbol{\beta}}_{max};\mathbf{Y}) - l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y})\right] = \\ & 2\left[l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y}) - l^*(\widehat{\boldsymbol{\beta}}_{sub};\mathbf{Y})\right] = \\ & \underbrace{2\left[l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y}) - l^*(\boldsymbol{\beta};\mathbf{Y})\right]}_{D_m \overset{A}{\sim} \chi^2(m)} - \underbrace{2\left[l^*(\widehat{\boldsymbol{\beta}}_{sub};\mathbf{Y}) - l^*(\boldsymbol{\beta}_{sub};\mathbf{Y})\right]}_{D_q \overset{A}{\sim} \chi^2(q)} + \underbrace{2\left[l^*(\boldsymbol{\beta};\mathbf{Y}) - l^*(\boldsymbol{\beta}_{sub};\mathbf{Y})\right]}_{\approx 0} \end{aligned}$$

Za předpokladu, že je submodel GLM_{sub} vhodný, bude hodnota posledního členu kladná konstanta blízká nule. Podobně jak v důkazu předchozí věty, lze první dva výrazy aproximovat jako rozdíl dvou kvadratických forem

$$\begin{aligned}\Delta D \approx &(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) - (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}}\mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta})^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \\ &(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}}\left(\mathbf{J}_n(\boldsymbol{\beta}) - \mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{J}_n(\boldsymbol{\beta})^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\right)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\end{aligned}$$

přičemž $\boldsymbol{\beta}_{sub}$ je podvektorem $\boldsymbol{\beta}$ a proto existuje matice $\mathbf{C}_{m\times q}$ taková, že $\boldsymbol{\beta}_{sub} = \mathbf{C}^{\mathrm{T}}\boldsymbol{\beta}$. Zbytek důkazu je analogický s předchozím důkazem. □

<u>POZNÁMKA</u> 2.7.11.

Statistiky ΔD se užívá při porovnání dvou modelů, přičemž jeden je submodelem druhého.

Definujme diagonální matici $\mathbf{C}$, na jejíž hlavní diagonále budou nuly a jedničky, tj. $c_{ii} \in \{0, 1\}$. Počet jedniček na hlavní diagonále je roven číslu $q = Tr(C) < m$. Předpokládáme, že základní model s vektorem neznámých parametrů, který tentokrát označíme $\boldsymbol{\beta}_1$, dobře popisuje chování náhodného výběru.

Uvažujme hypotézu

$$\boxed{H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_{sub} = \mathbf{C}^{\mathrm{T}}\boldsymbol{\beta}_1}$$

a alternativní hypotézu odpovídající základnímu modelu

$$\boxed{H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1}.$$

Test hypotézy H_0 vůči H_1 provedeme s využitím statistiky ΔD. Má-li statistika ΔD hodnotu

$$\Delta D > \chi^2_{1-\alpha}(m-q),$$

pak **považujeme submodel za nevhodný** a pro popis chování sledovaného náhodného výběru použijeme model základní. <u>V opačném případě</u> jsou oba dva modely dobré a **vybereme jednodušší model odpovídající** H_0.

Je nutné zdůraznit, že pojem þ*dobře popisuje chování náhodného výběru*ß je zde použit pouze relativně z důvodu srovnání dvou modelů, kdy je třeba zjistit, zda obecnější model je výrazně lepší, či jsou oba modely přibližně srovnatelné. Neznamená ještě, když přijmeme hypotézu H_0, že jsou oba modely pro daná data (absolutně) dobré, ale například jen to, že jsou oba stejně špatné.

Proto při praktických úlohách se volí matice plánu obecnějšího modelu se všemi potencionálními vysvětlujícími veličinami a předpokládá se, že tento model popisuje data dobře. Následuje odstraňování jednotlivých vysvětlujících veličin a testování, zda vzniklý jednodušší model nepopisuje data na určité hladině významnosti stejně dobře, jako složitější model základní.

ODHAD RUŠIVÉHO PARAMETRU ϕ POMOCÍ MOMENTOVÉ METODY

Mějme náhodný výběr $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ z rozdělení exponenciálního typu, který se řídí GLM modelem Předpokládejme, že pro hustotu exponenciálního typu platí:

$$\boxed{\psi_i(\phi) = \frac{\phi}{\omega_i}},$$

kde $\omega_i > 0$ jsou **známé apriorní váhy** a $\phi > 0$ je **neznámý rušivý parametr**, který se též nazývá **škálovým** či **rozptylovým** parametrem.

Pak škálovou deviaci můžeme vyjádřit takto

$$
\begin{aligned}
\boxed{D} &= 2\left[l^*(\widehat{\boldsymbol{\beta}}_{max};\mathbf{Y}) - l^*(\widehat{\boldsymbol{\beta}};\mathbf{Y})\right] \\
&= 2\sum_{i=1}^{n}\left\{\frac{\omega_i\left[Y_i\hat{\theta}_{i,max} - \gamma(\hat{\theta}_{i,max})\right]}{\phi} + d(Y_i,\phi) - \frac{\omega_i\left[Y_i\hat{\theta}_i - \gamma(\hat{\theta}_i)\right]}{\phi} - d(Y_i,\phi)\right\} \\
&= \frac{1}{\phi}2\sum_{i=1}^{n}\omega_i\left[Y_i(\hat{\theta}_{i,max} - \hat{\theta}_i) - \gamma(\hat{\theta}_{i,max}) + \gamma(\hat{\theta}_i)\right] \\
&= \boxed{\frac{1}{\phi}D^*}
\end{aligned}
$$

a D^* nazveme **neškálovanou deviací** (unscaled deviance). Protože platí

$$
D = \frac{1}{\phi}D^* \quad \overset{A}{\sim} \quad \chi^2(n-m) \qquad \Rightarrow \qquad ED = \frac{1}{\phi}ED^* \approx n - m,
$$

neboť střední hodnota u χ^2 je rovna počtu stupňů volnosti, pak

$$
\boxed{\hat{\phi}_{D^*} = \frac{D^*}{n-m}}.
$$

<u>Poznámka</u> 2.7.12.
Další často používanou mírou vhodnosti modelu je tzv. **zobecněná Pearsonova statistika**

$$
X^2 = \sum_{i=1}^{n}\frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad \overset{A}{\sim} \quad \chi^2(n-m)
$$

a proto dalším momentovým odhadem založeným na této statistice je

$$
\boxed{\hat{\phi}_{X^2} = \frac{X^2}{n-m}}.
$$

Přehled rozptylových parametrů a neškálovaných deviací pro rozdělení exponenciálního typu je dán v následující tabulce.

Rozdělení	ϕ	D^*
Normální rozdělení	σ^2	$\sum\limits_{i=1}^{n}(Y_i - \hat{\mu}_i)^2$
Poissonovo rozdělení	1	$2\sum\limits_{i=1}^{n}\left[Y_i ln\frac{Y_i}{\hat{\mu}_i} - (Y_i - \hat{\mu}_i)\right]$
Binomické rozdělení	1	$2\sum\limits_{i=1}^{n}\left[Y_i \ln\frac{y_i}{\hat{\mu}_i} + (n_i - Y_i)\ln\frac{n_i - Y_i}{n_i - \hat{\mu}_i}\right]$
Gamma rozdělení	$\frac{1}{\alpha}$	$2\sum\limits_{i=1}^{n}\left[\frac{Y_i - \hat{\mu}_i}{\hat{\mu}_i} - ln\frac{Y_i}{\hat{\mu}_i}\right]$

ANALÝZA REZIDUÍ

Analýza reziduí nám poskytuje nenahraditelnou informaci o vhodnosti použitého modelu a nelze ji opomíjet, i když analýza deviace ukazuje na vhodnost modelu.

Teprve zde se častokrát zjistí, že výchozí předpoklad o rozdělení náhodných chyb či tvaru linkovací funkce nebyl správný.

Analýza reziduí může odhalit body, jejichž reziduum je výrazně odlišné od ostatních pozorování, což může být způsobeno neobvyklou závislostí mezi vysvětlovanou a vysvětlujícími veličinami či prozaičtěji chybou měření, chybou přepisu hodnot do databáze apod.

Když se v grafu reziduí objeví určitá závislost reziduí na fitované hodnotě či vysvětlujících veličinách nebo třeba s rostoucí odhadnutou (fitovanou) hodnotou roste variabilita reziduí, pak je nutné celý model přehodnotit a případně jej začít vytvářet od začátku.

Uveďme nejznámější typy reziduí používaných v GLM :

(a) **Standardizovaná rezidua** (*linear*): též *Pearsonova*

$$\boxed{r_i^P = \frac{Y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}}.$$

Nevýhodou těchto reziduí je fakt, že pro nenormální rozdělení jsou značně zešikmená.

(b) **Standardizovaná transformovaná rezidua** (*transformed linear*)

$$\boxed{r_i^{Tr} = \frac{g(Y_i) - Eg(Y_i)}{\sqrt{Dg(Y_i)}}}.$$

Důvodem zavedení transformovací je snaha, aby transformovaná rezidua měla rozdělení, které se co nejvíce blíží normálnímu.

(1) **Anscombova rezidua** jsou založena na transformaci typu

$$g_A(\mu) = \int \frac{d\mu}{V^{\frac{1}{3}}(\mu)},$$

jejiž snahou je, aby transformovaná rezidua měla **nulovou šikmost**.

(2) **Rezidua stabilizující rozptyl** jsou založena na transformaci typu

$$g_S(\mu) = \int \frac{d\mu}{V^{\frac{1}{2}}(\mu)},$$

a cílem je, aby u transformovaných reziduí rozptyl nebyl funkcí střední hodnoty, ale konstantní.

Příklady standardizovaných transformovaných reziduí		
Rozdělení	Anscombova	stabilizující rozptyl
Binomické	$\frac{t\left(\frac{Y_i}{n_i}\right)-\left[t(\hat{\pi}_i)+\hat{\pi}_i^{-\frac{1}{3}}(1-\hat{\pi}_i)^{-\frac{1}{3}}(2\hat{\pi}_i-1)/6n_i\right]}{\hat{\pi}_i^{\frac{1}{6}}(1-\hat{\pi}_i)^{\frac{1}{6}}/n_i}$	$\frac{\arcsin\sqrt{\frac{Y_i}{n_i}}-\arcsin\sqrt{\hat{\pi}_i}}{1/(2\sqrt{n_i})}$
Poissonovo	$\frac{Y_i^{\frac{2}{3}}-\left(\hat{\mu}_i^{\frac{2}{3}}-\hat{\mu}_i^{-\frac{1}{3}}/9\right)}{\frac{2}{3}\hat{\mu}_i^{\frac{1}{6}}}$	$\frac{Y_i^{\frac{1}{2}}-\hat{\mu}_i^{\frac{1}{2}}}{\frac{1}{2}}$
Gamma	$\frac{(Y_i/\alpha)^{\frac{1}{3}}-\left[\hat{\mu}_i^{\frac{1}{3}}-\hat{\mu}_i^{-\frac{2}{3}}/9\right]}{\hat{\mu}_i^{-\frac{1}{6}}/3}$	neuvažujeme

kde pro binomické rozdělení: $\hat{\pi}_i = \hat{\mu}_i/n_i$ a $t(u) = \int_0^u s^{-\frac{1}{3}}(1-s)^{-\frac{1}{3}}ds$

(c) **Deviační rezidua** (*deviance residual*)

$$\boxed{r_i^D = sign(Y_i - \hat{\mu}_i)\sqrt{d_i}},$$

přičemž

$$D = 2[l^*(\mathbf{b}_{max}; \mathbf{Y}) - l^*(\mathbf{b}; \mathbf{Y})] = \sum_{i=1}^{n} d_i.$$

Ještě lepší vlastnosti mají tzv. **korigovaná deviační rezidua** (bias-adjusted deviance residual)

$$\boxed{r_i^{AD} = sign(Y_i - \hat{\mu}_i)\sqrt{d_i} + \rho_3(\hat{\mu}_i)/6},$$

kde

$$\rho_3(\mu) = E\left[\left(\frac{Y-\mu}{\sqrt{DY}}\right)^3\right].$$

2.8 TABULKY ROZDĚLENÍ EXPONENCIÁLNÍHO TYPU

2.8.1 Tabulka rozdělení exponenciálního typu

Příklady rozdělení exponenciálního typu

Rozdělení	$EY_i = \mu_i$	**Kanonická link-funkce** $\theta_i = \eta_i = g(\mu_i)$	$\gamma(\theta_i)$	**Rozptylová funkce** $V(\mu_i)$	$\psi_i(\phi) = \frac{\phi}{\omega_i}$	
					ϕ	ω_i
$Y_i \sim N(\mu_i, \frac{\sigma^2}{\omega_i})$	μ_i	μ_i	$\frac{1}{2}\theta_i^2$	1	σ^2	ω_i
$Y_i \sim Bi(n_i, \pi_i)$	$n_i\pi_i$	$\ln \frac{\mu_i}{n_i - \mu_i}$	$n_i \ln(1+e^{\theta_i})$	$\mu_i\left(1-\frac{\mu_i}{n_i}\right)$	1	1
$n_iY_i \sim Bi(n_i, \pi_i)$	π_i	$\ln \frac{\mu_i}{1-\mu_i}$	$\ln(1+e^{\theta_i})$	$\mu_i(1-\mu_i)$	1	n_i
$Y_i \sim Po(\mu_i)$	μ_i	$\ln \mu_i$	e^{θ_i}	μ_i	1	1
$Y_i \sim Ex(\mu_i)$	μ_i	$-\frac{1}{\mu_i}$	$-\ln(-\theta_i)$	μ_i^2	1	1
$Y_i \sim G(\alpha, \beta_i)$	$\alpha\beta_i$	$-\frac{1}{\mu_i}$	$-\ln(-\theta_i)$	μ_i^2	$\frac{1}{\alpha}$	1

2.8.2 Tabulka různých spojovacích funkcí

Rozdělení	**link-funkce**	$\eta_i = g(\mu_i)$	$\mu_i = g^{-1}(\eta_i)$	$\frac{\partial \eta_i}{\partial \mu_i}$
Normální, Gama	identity	μ_i	η_i	1
Exponenciální	log	$\ln \mu_i$	e^{η_i}	$\frac{1}{\mu_i}$
Poissonovo	power	μ_i^a	$e^{\frac{1}{a}\ln \eta_i}$	$a\mu_i^{a-1}$
	logit	$\ln \frac{\mu_i}{n_i - \mu_i}$	$n_i \frac{e^{\eta_i}}{1+e^{\eta_i}}$	$\frac{n_i}{\mu_i(n_i-\mu_i)}$
Binomické	probit	$\Phi^{-1}(\frac{\mu_i}{n_i})$	$n_i\Phi(\eta_i)$	$\frac{\sqrt{2\pi}}{n_i} e^{\frac{1}{2}\left(\Phi^{-1}(\frac{\mu_i}{n_i})\right)^2}$
	complement.log-log	$\ln\left(-\ln\left(1-\frac{\mu_i}{n_i}\right)\right)$	$n_i\left(1-e^{-e^{\eta_i}}\right)$	$\frac{1}{(\mu_i - n_i)\ln\left(1-\frac{\mu_i}{n_i}\right)}$
	log-log	$-\ln\left(-\ln\left(\frac{\mu_i}{n_i}\right)\right)$	$n_i e^{-e^{-\eta_i}}$	$-\frac{1}{\mu_i \ln\left(\frac{\mu_i}{n_i}\right)}$

kde Φ^{-1} je kvantilová funkce standardizovaného normálního rozdělení.

KLASICKÝ REGRESNÍ MODEL A GLM

ZNAČENÍ

Střední hodnoty jednotlivých pozorování $EY_i = \mu_i \quad (i = 1, \ldots, n)$

Hodnoty regresorů jednotlivých pozorování $\mathbf{x}_i^{\mathrm{T}} = (x_{i1}, \ldots, x_{im}) \quad (i = 1, \ldots, n)$

Vektor pozorovaných hodnot $\quad \mathbf{Y} = (y_1, \ldots, y_n)^{\mathrm{T}}$

Vektory neznámých parametrů $\quad (q < m < n)$

$$\boldsymbol{\beta} = \boldsymbol{\beta}_m = (\underbrace{\beta_1, \ldots, \beta_q}_{\boldsymbol{\beta}_q}, \underbrace{\beta_{q+1}, \ldots, \beta_m}_{\boldsymbol{\beta}_{m-q}})^{\mathrm{T}}; \qquad \boldsymbol{\beta}_{max} = (\underbrace{\beta_1, \ldots, \beta_q}_{\boldsymbol{\beta}_q}, \underbrace{\beta_{q+1}, \ldots, \beta_m}_{\boldsymbol{\beta}_{m-q}}, \underbrace{\beta_{m+1}, \ldots, \beta_n}_{\boldsymbol{\beta}_{n-m}})^{\mathrm{T}}$$

Matice plánu z hodnot regresorů $\quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$

Submodely $\quad (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{V} \end{pmatrix} \qquad h(\mathbf{X}^{\mathrm{T}}\mathbf{X}) = m;\ h(\mathbf{V}) = m - q;\ h(\mathbf{A}) = q$

MODEL

$EY_i = \mu_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ $Y_i \sim N(\underbrace{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}_{EY_i}, \underbrace{\sigma^2}_{DY_i})$	$EY_i = \mu_i \qquad \eta_i = g(\mu_i) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} \qquad \mu_i = g^{-1}(\eta_i)$ $Y_i \sim \mathcal{L}_{exp}(\underbrace{\gamma'(\theta_i) = g^{-1}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}_{EY_i}, \underbrace{\psi_i(\phi)\gamma''(\theta_i)}_{DY_i})$

ODHADY

Metoda nejmenších čtverců	Iterativni vážená metoda nejm. čtverců (max.věr.odh.)
$\widehat{\boldsymbol{\beta}} = \mathbf{b}_m = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$	$\widehat{\boldsymbol{\beta}}^{(s)} = \mathbf{b}_m^{(s)} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Z}^{(s-1)}$
	$\mathbf{Z}^{(s-1)} = (Z_1^{(s-1)}, \ldots, Z_n^{(s-1)})^{\mathrm{T}};\ Z_i^{(s-1)} = \hat{\eta}_i^{(s-1)} + (y_i - \hat{\mu}_i^{(s-1)})\frac{d\eta_i}{d\mu_i}$ $\hat{\eta}_i^{(s-1)} = \mathbf{x}_i^{\mathrm{T}}\mathbf{b}^{(s-1)}; \quad \hat{\mu}_i^{(s-1)} = g^{-1}(\hat{\eta}_i^{(s-1)})$ $w_i = \frac{1}{DY_i}\left(\frac{d\mu_i}{d\eta_i}\right)^2; \quad W = diag\{w_1, \ldots, w_n\}$

$$\mathbf{b}_{max} = (b_1, \ldots, b_n)^{\mathrm{T}} = (y_1, \ldots, y_n)^{\mathrm{T}}$$

VÝBĚROVÁ ROZDĚLENÍ odhadu $\mathbf{b}_m$

$\mathbf{b}_m \sim N_m(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1})$	$\mathbf{b}_m \overset{A}{\sim} N_m(\boldsymbol{\beta}, \mathbf{J}_n^{-1})$
$\frac{1}{\sigma^2}(\mathbf{b}_m - \boldsymbol{\beta}_m)^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\mathbf{b}_m - \boldsymbol{\beta}_m) \sim \chi^2(m)$	$(\mathbf{b}_m - \boldsymbol{\beta}_m)^{\mathrm{T}}\mathbf{J}_{\mathbf{n}}(\mathbf{b}_m - \boldsymbol{\beta}_m) \overset{A}{\sim} \chi^2(m) \quad$ *Waldova statistika*

VÝBĚROVÁ ROZDĚLENÍ škálové deviace $D = 2[l(\mathbf{b}_{max}; \mathbf{y}) - l(\mathbf{b}_m; \mathbf{y})]$

$D \sim \chi^2(n - m)$	$D \overset{A}{\sim} \chi^2(n - m)$
Normální rozdělení $D = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2 = \frac{SSE}{\sigma^2} = \frac{(n-m)s_m^2}{\sigma^2}$	**Poissonovo rozdělení**: $\quad D = 2\sum_{i=1}^{n}\left[y_i ln\frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i)\right]$
	Binomické rozdělení $\quad D = 2\sum_{i=1}^{n}\left[y_i ln\frac{y_i}{\hat{\mu}_i} + (n_i - y_i)\ln\frac{n_i - y_i}{n_i - \hat{\mu}_i}\right]$
	Gamma rozdělení $\quad D = 2\alpha\sum_{i=1}^{n}\left[\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} - ln\frac{y_i}{\hat{\mu}_i}\right]$

SUBMODELY

$\boxed{H_0} : \boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^{\mathrm{T}}$ proti $\boxed{H_1} : \boldsymbol{\beta} = (\beta_1, \ldots, \beta_q, \beta_{q+1}, \ldots, \beta_m)^{\mathrm{T}}$

nebo $\boxed{H_0} : (\beta_{q+1}, \ldots, \beta_m)^{\mathrm{T}} = \mathbf{0}$ proti $\boxed{H_1} : (\beta_{q+1}, \ldots, \beta_m)^{\mathrm{T}} \neq \mathbf{0}$

$F \sim F(m - q, n - m)$	$\Delta D \overset{A}{\sim} \chi^2(m - q)$
$F = \frac{1}{(m-q)s_m}\mathbf{b}_{m-q}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{b}_{m-q}$	$\Delta D = D_{sub} - D = 2[l(\mathbf{b}_m; \mathbf{y}) - l(\mathbf{b}_q; \mathbf{y})]$

2.9 ANALÝZA DEVIACE

Analýza deviace je obdobou analýzy rozptylu u klasických lineárních regresních modelů. Užívá se k **určení vlivu** kategorických proměnných či **faktorů na střední hodnotu náhodného vektoru**. Přitom se uvažuje pouze aditivní působení faktorů (ne interakce).

Každý z modelů (submodelů) odpovídá určitému uspořádání pokusu a předpokládanému vlivu faktorů na střední hodnoty pozorovaných náhodných veličin.

V analýze rozptylu jsou jednotlivé modely srovnávány pomocí součtu čtverců, neboť jejich parametry jsou odhadovány metodou nejmenších čtverců. Ze stejného důvodu je deviace kritériem vhodnosti zobecněného lineárního modelu. **Metoda maximální věrohodnosti totiž odpovídá hledání minima deviace modelu**.

Při analýze deviace používáme zobecněný lineární model, ve kterém vysvětlující proměnné jsou tzv. *dummy*, *fiktivní*, *latentní* proměnné, tj. vektory tvořené nulami a jedničkami.

2.9.1 Jednoduché třídění

Nejjednodušší formou analýzy deviace je analýza při jednoduchém třídění - jednofaktorový komplex. V tomto případě zkoumáme vliv jednoho faktoru na výsledky pokusu tak, že sledujeme výsledky pokusu při několika variantách (úrovních) tohoto faktoru. Naměřené hodnoty roztřídíme do skupin podle varianty (úrovně) faktoru.

Nechť náhodný vektor $\mathbf{Y}_n = (Y_1, \dots, Y_n)^{\mathrm{T}}$ se řídí zobecněným lineárním modelem s linkovací funkcí $\boxed{g}$ a uvažujme vliv jednoho faktoru $\boxed{\mathrm{A}}$.

Přeindexujeme

$$\begin{array}{lllll} Y_1, \dots, Y_n & \text{na} & Y_{jk}, & j = 1, \dots, I & I \ \dots \text{ počet úrovní faktoru A} \\ \varepsilon_1, \dots, \varepsilon_n & & \varepsilon_{jk}, & k = 1, \dots, n_j & n_j \ \dots \text{ počet případů } j\text{-té úrovně.} \end{array}$$

Předpokládejme, že uvnitř úrovní tvoří náhodné veličiny náhodný výběr, tj. že

$$EY_{j1} = \dots = EY_{jn_j} = \mu_j \qquad \text{pro} \qquad j = 1, \dots, I.$$

Při jednoduchém třídění testujeme hypotézu

$$H_0 : \mu_1 = \dots = \mu_I$$

proti alternativě, že ne všechny střední hodnoty jsou si rovny.

Jinými slovy hypotéza říká, že $\mathbf{Y}_n = (Y_1, \dots, Y_n)^{\mathrm{T}}$ tvoří jeden náhodný výběr, alternativa znamená, že táž pozorování představují obecně I náhodných výběrů lišících se střední hodnotou. To odpovídá situaci, kdy pozorování byla pořízena za I různých podmínek, jejichž vliv, pokud existuje, se dá vyjádřit aditivní změnou střední hodnoty.

Pokud hypotézu zamítneme, považujeme vliv zkoumaného faktoru za významný, v opačném případě za bezvýznamný.

K testování nulové hypotézy je možné použít různé formulace základního modelu, který vždy pak srovnáváme ze submodelem

$$\boxed{GLM_{sub} : \eta_{jk} = g(EY_{jk}) = \mu} \qquad j = 1, \ldots, I \quad k = 1, \ldots, n_j.$$

Vyjděme z tohoto základního modelu

$$\boxed{GLM_{*} : \eta_{jk} = g(EY_{jk}) = \mu + \alpha_j} \qquad j = 1, \ldots, I \quad k = 1, \ldots, n_j.$$

Matice plánu má v tomto případě tvar

$$\mathbf{X}_* = \begin{pmatrix} 1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{1} & \underline{1} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{0} \\ 1 & 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{1} & \underline{0} & \underline{1} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{0} \\ 1 & 0 & 0 & \cdots & \cdots & 0 & 1 & 0 \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{1} & \underline{0} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{0} & \underline{1} & \underline{0} \\ 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 & 1 \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \tag{2.8}$$

V tomto případě **matice plánu X není plné hodnosti** (první sloupec je součtem všech ostatních). Říkáme, že model je PŘEPARAMETRIZOVÁN (tzv. *"overparameterized model"*). Nezbývá než přidat další **reparametrizační podmínku** a podle toho, jakou podmínku přidáme, dojdeme k následujícím variantám

(1) Nejjednodušší variantou je přidání vedlejší podmínky $\boxed{\mu = 0}$.
V tomto případě představují parametry $\boxed{\alpha_1, \ldots, \alpha_I}$ **vlivy úrovní** faktoru A a z matice $\mathbf{X}_*$ vynecháme první sloupec. Pak

$$\mathbf{X}^{\mathrm{T}}\mathbf{W} = \begin{pmatrix} w_{11} & \cdots & w_{1n_1} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & w_{21} & \cdots & w_{2n_2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & & & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & w_{I-1,1} & \cdots & w_{I-1,n_{I-1}} & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & w_{I1} & \cdots & w_{In_I} \end{pmatrix}$$

$$\mathbf{J}_n = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X} = diag\{\sum_{k=1}^{n_1} w_{1k}, \ldots, \sum_{k=1}^{n_I} w_{Ik}\}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Z} = \left(\sum_{k=1}^{n_1} w_{1k} Z_{1k}, \ldots, \sum_{k=1}^{n_I} w_{Ik} Z_{Ik} \right)^{\mathrm{T}}$$

a odtud

$$\widehat{\boldsymbol{\beta}}^{(s+1)} = \left(\hat{\alpha}_1^{(s+1)}, \ldots, \hat{\alpha}_I^{(s+1)} \right)^{\mathrm{T}} = \left(\frac{\sum_{k=1}^{n_1} w_{1k}^{(s)} Z_{1k}^{(s)}}{\sum_{k=1}^{n_1} w_{1k}^{(s)}}, \ldots, \frac{\sum_{k=1}^{n_I} w_{Ik}^{(s)} Z_{Ik}^{(s)}}{\sum_{k=1}^{n_I} w_{Ik}^{(s)}} \right)^{\mathrm{T}}$$

kde

$$
\begin{aligned}
Z_{jk}^{(s)} &= \widehat{\boldsymbol{\beta}}_j^{(s)} + r_{jk}^{(s)} q_{jk}^{(s)} \\
r_{jk}^{(s)} &= Y_{jk} - \mu_{jk}^{(s)} \\
\boxed{\mu_{jk}^{(s)}} &= g^{-1}\left(\mathbf{x}_{jk}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}^{(s)}\right) = g^{-1}\left(\hat{\beta}_j^{(s)}\right) = \boxed{g^{-1}\left(\frac{\sum_{k=1}^{n_j} w_{jk}^{(s)} Z_{jk}^{(s)}}{\sum_{k=1}^{n_j} w_{jk}^{(s)}}\right)}.
\end{aligned}
$$

(2) Další variantou je položit

$$\boxed{\alpha_1 = 0}.$$

V tomto případě představuje parametr $\boxed{\mu}$ **účinek první úrovně** faktoru A a $\boxed{\alpha_j}$ $(j = 2, \dots, I)$ zastupuje **rozdíl mezi účinkem první a j-té úrovně** faktoru A. Matici plánu $\mathbf{X}$ dostaneme vynecháním druhého sloupce v matici $\mathbf{X}_*$.

$$
\mathbf{X}^{\mathrm{T}}\mathbf{W} = \begin{pmatrix}
w_{11} & \cdots & w_{1n_1} & w_{21} & \cdots & w_{2n_2} & \cdots & \cdots & \cdots & w_{I1} & \cdots & w_{In_I} \\
0 & \cdots & 0 & w_{21} & \cdots & w_{2n_2} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & & & & & & & & & & & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & w_{I-1,1} & \cdots & w_{I-1,n_{I-1}} & 0 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & w_{I1} & \cdots & w_{In_I}
\end{pmatrix}
$$

$$
\mathbf{J}_n = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X} = \begin{pmatrix}
\sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk} & \sum_{k=1}^{n_2} w_{2k} & \cdots & \cdots & \sum_{k=1}^{n_I} w_{Ik} \\
\sum_{k=1}^{n_2} w_{2k} & \sum_{k=1}^{n_2} w_{2k} & 0 & \cdots & 0 \\
\vdots & 0 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & 0 \\
\sum_{k=1}^{n_I} w_{Ik} & 0 & \cdots & 0 & \sum_{k=1}^{n_I} w_{Ik}
\end{pmatrix}
$$

$$
\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Z} = \left(\sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk}Z_{jk}, \sum_{k=1}^{n_2} w_{2k}Z_{2k}, \dots, \sum_{k=1}^{n_I} w_{Ik}Z_{Ik}\right)^{\mathrm{T}}
$$

Pokud systém rovnic $\quad \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Z} \quad$ rozepíšeme, dostaneme

$$
\begin{aligned}
\mu \sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk} + \alpha_2 \sum_{k=1}^{n_2} w_{2k} + \cdots + \alpha_I \sum_{k=1}^{n_I} w_{Ik} &= \sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk}Z_{jk} \\
\mu \sum_{k=1}^{n_j} w_{jk} + \alpha_j \sum_{k=1}^{n_j} w_{jk} &= \sum_{k=1}^{n_j} w_{jk}Z_{jk} \qquad j = 2, \dots, I
\end{aligned}
$$

Pokud od první rovnice odečteme ostatní, dokážeme vypočítat

$$\mu = \frac{\sum_{k=1}^{n_1} w_{1k}Z_{1k}}{\sum_{k=1}^{n_1} w_{1k}}$$

a odtud po dosazení do ostatních rovnic

$$\alpha_j = \frac{\sum_{k=1}^{n_j} w_{jk}Z_{jk}}{\sum_{k=1}^{n_j} w_{jk}} - \frac{\sum_{k=1}^{n_1} w_{1k}Z_{1k}}{\sum_{k=1}^{n_1} w_{1k}} \qquad j = 2, \dots, I.$$

Pak

$$\widehat{\boldsymbol{\beta}}^{(s+1)} = \left(\frac{\sum_{k=1}^{n_1} w_{1k}^{(s)} Z_{1k}^{(s)}}{\sum_{k=1}^{n_1} w_{1k}^{(s)}}, \frac{\sum_{k=1}^{n_2} w_{2k}^{(s)} Z_{2k}^{(s)}}{\sum_{k=1}^{n_2} w_{2k}^{(s)}} - \frac{\sum_{k=1}^{n_1} w_{1k}^{(s)} Z_{1k}^{(s)}}{\sum_{k=1}^{n_1} w_{1k}^{(s)}}, \ldots, \right.$$
$$\left. \ldots, \frac{\sum_{k=1}^{n_I} w_{Ik}^{(s)} Z_{Ik}^{(s)}}{\sum_{k=1}^{n_I} w_{Ik}^{(s)}} - \frac{\sum_{k=1}^{n_1} w_{1k}^{(s)} Z_{1k}^{(s)}}{\sum_{k=1}^{n_1} w_{1k}^{(s)}} \right)^{\mathrm{T}}$$

kde

$$\begin{aligned} Z_{jk}^{(s)} &= \widehat{\boldsymbol{\beta}}_j^{(s)} + r_{jk}^{(s)} q_{jk}^{(s)} \\ r_{jk}^{(s)} &= Y_{jk} - \mu_{jk}^{(s)} \\ \boxed{\mu_{jk}^{(s)}} &= g^{-1}\left(\mathbf{x}_{jk}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}^{(s)}\right) = g^{-1}\left(\hat{\beta}_j^{(s)}\right) = g^{-1}\left(\hat{\mu}^{(s)} + \hat{\alpha}_j^{(s)}\right) = \boxed{g^{-1}\left(\frac{\sum_{k=1}^{n_j} w_{jk}^{(s)} Z_{jk}^{(s)}}{\sum_{k=1}^{n_j} w_{jk}^{(s)}}\right)}. \end{aligned}$$

(3) Další možnou variantou je dodatečná podmínka

$$\boxed{\sum_{j=1}^{I} \alpha_j = 0} \qquad \Rightarrow \qquad \alpha_I = -\alpha_1 - \cdots - \alpha_{I-1}$$

(tzv. ***"σ-restricted"*** **kódování**) s vektorem neznámých parametrů

$$\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_{I-1})^{\mathrm{T}}$$

a s maticí plánu

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & 0 & \cdots & 0 & 1 \\ \vdots & -1 & \cdots & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & \cdots & \cdots & -1 \end{pmatrix}.$$

Ukážeme nyní, že v tomto případě představuje parametr $\boxed{\mu}$ **obecnou střední hodnotu** (je prostým průměrem jednotlivých úrovní) a parametry $\boxed{\alpha_1, \ldots, \alpha_I}$ představují **efekty** jednotlivých úrovní faktoru A.

$$\mathbf{X}^{\mathrm{T}}\mathbf{W} = \begin{pmatrix} w_{11} & \cdots & w_{1n_1} & w_{21} & \cdots & w_{2n_2} & \cdots & \cdots & \cdots & w_{I1} & \cdots & w_{In_I} \\ w_{11} & \cdots & w_{1n_1} & 0 & \cdots & 0 & 0 & \cdots & 0 & -w_{I1} & \cdots & -w_{In_I} \\ 0 & \cdots & 0 & w_{21} & \cdots & w_{2n_2} & 0 & \cdots & 0 & -w_{I1} & \cdots & -w_{In_I} \\ \vdots & & & \ddots & & & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & & & & \ddots & & & \ddots & 0 & \vdots & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & w_{I-1,1} & \cdots & w_{I-1,n_{I-1}} & -w_{I1} & \cdots & -w_{In_I} \end{pmatrix}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X} = \begin{pmatrix} \sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk} & \sum_{k=1}^{n_1} w_{1k} - \sum_{k=1}^{n_I} w_{Ik} & \cdots & \cdots & \sum_{k=1}^{n_{I-1}} w_{I-1,k} - \sum_{k=1}^{n_I} w_{Ik} \\ \sum_{k=1}^{n_1} w_{1k} - \sum_{k=1}^{n_I} w_{Ik} & \sum_{k=1}^{n_1} w_{1k} + \sum_{k=1}^{n_I} w_{Ik} & \sum_{k=1}^{n_I} w_{Ik} & \cdots & \sum_{k=1}^{n_I} w_{Ik} \\ \vdots & \sum_{k=1}^{n_I} w_{Ik} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sum_{k=1}^{n_I} w_{Ik} \\ \sum_{k=1}^{n_{I-1}} w_{I-1,k} - \sum_{k=1}^{n_I} w_{Ik} & \sum_{k=1}^{n_I} w_{Ik} & \cdots & \sum_{k=1}^{n_I} w_{Ik} & \sum_{k=1}^{n_{I-1}} w_{I-1,k} + \sum_{k=1}^{n_I} w_{Ik} \end{pmatrix}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Z} = \left(\sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk}Z_{jk}, \sum_{k=1}^{n_1} w_{1k}Z_{1k} - \sum_{k=1}^{n_I} w_{Ik}Z_{Ik}, \ldots, \sum_{k=1}^{n_{I-1}} w_{I-1,k}Z_{I-1,k} - \sum_{k=1}^{n_I} w_{Ik}Z_{Ik}\right)^{\mathrm{T}}$$

Pokud systém rovnic $\quad \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Z} \quad$ rozepíšeme, máme

$$\mu\sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk} + \alpha_1\left(\sum_{k=1}^{n_1} w_{1k} - \sum_{k=1}^{n_I} w_{Ik}\right) + \cdots + \alpha_{I-1}\left(\sum_{k=1}^{n_{I-1}} w_{I-1,k} - \sum_{k=1}^{n_I} w_{Ik}\right) =$$
$$= \sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk}Z_{jk}$$

a pro $j = 1, \ldots, I-1$

$$\mu\left(\sum_{k=1}^{n_j} w_{jk} - \sum_{k=1}^{n_I} w_{Ik}\right) + \alpha_1\sum_{k=1}^{n_I} w_{Ik} + \cdots + \alpha_j\left(\sum_{k=1}^{n_j} w_{jk} + \sum_{k=1}^{n_I} w_{Ik}\right) + \cdots + \alpha_{I-1}\sum_{k=1}^{n_I} w_{Ik} =$$
$$= \sum_{k=1}^{n_j} w_{jk}Z_{jk} - \sum_{k=1}^{n_I} w_{Ik}Z_{Ik}$$

Jestliže od první rovnice odečteme ostatní, dostaneme

$$\mu\left(\underbrace{\sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk} - \sum_{k=1}^{n_1} w_{1k} - \cdots - \sum_{k=1}^{n_{I-1}} w_{I-1,k}}_{=\sum_{k=1}^{n_I} w_{Ik}} + (I-1)\sum_{k=1}^{n_I} w_{Ik}\right) +$$
$$\alpha_1\left(\sum_{k=1}^{n_1} w_{1k} - \sum_{k=1}^{n_I} w_{Ik} - \sum_{k=1}^{n_1} w_{1k} - (I-1)\sum_{k=1}^{n_I} w_{Ik}\right) +$$
$$\vdots$$
$$\alpha_{I-1}\left(\sum_{k=1}^{n_{I-1}} w_{I-1,k} - \sum_{k=1}^{n_I} w_{Ik} - \sum_{k=1}^{n_{I-1}} w_{I-1,k} - (I-1)\sum_{k=1}^{n_I} w_{Ik}\right) =$$
$$= \underbrace{\sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk}Z_{jk} - \sum_{k=1}^{n_1} w_{1k}Z_{1k} - \cdots - \sum_{k=1}^{n_{I-1}} w_{I-1,k}Z_{I-1,k}}_{\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}} + (I-1)\sum_{k=1}^{n_I} w_{Ik}Z_{Ik},$$

tedy

$$I\sum_{k=1}^{n_I} w_{Ik}\left(\mu\underbrace{-\alpha_1-\cdots-\alpha_{I-1}}_{=\alpha_I}\right)=I\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}$$

a odtud

$$\boxed{\alpha_I=\frac{\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}}{\sum_{k=1}^{n_I} w_{Ik}}-\mu}.$$

Pro $j=1,\dots,I-1$ můžeme pak psát

$$\mu\left(\sum_{k=1}^{n_j} w_{jk}-\sum_{k=1}^{n_I} w_{Ik}\right)+\alpha_j\sum_{k=1}^{n_j} w_{jk}+\sum_{k=1}^{n_I} w_{Ik}\underbrace{(\alpha_1+\cdots+\alpha_{I-1})}_{=-\alpha_I}=\sum_{k=1}^{n_j} w_{jk}Z_{jk}-\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}$$

a po jednoduchých úpravách dostaneme

$$\boxed{\alpha_j=\frac{\sum_{k=1}^{n_j} w_{jk}Z_{jk}}{\sum_{k=1}^{n_j} w_{jk}}-\mu}.$$

Protože

$$0=\alpha_1+\cdots+\alpha_I=\frac{\sum_{k=1}^{n_1} w_{1k}Z_{1k}}{\sum_{k=1}^{n_1} w_{1k}}-\mu+\cdots+\frac{\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}}{\sum_{k=1}^{n_I} w_{Ik}}-\mu,$$

ihned máme

$$\boxed{\mu=\frac{1}{I}\left(\frac{\sum_{k=1}^{n_1} w_{1k}Z_{1k}}{\sum_{k=1}^{n_1} w_{1k}}+\cdots+\frac{\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}}{\sum_{k=1}^{n_I} w_{Ik}}\right)}.$$

Pak

$$\widehat{\boldsymbol\beta}^{(s+1)}=\left(\underbrace{\frac{1}{I}\left(\frac{\sum_{k=1}^{n_1} w_{1k}^{(s)}Z_{1k}^{(s)}}{\sum_{k=1}^{n_1} w_{1k}^{(s)}}+\cdots+\frac{\sum_{k=1}^{n_I} w_{Ik}^{(s)}Z_{Ik}^{(s)}}{\sum_{k=1}^{n_I} w_{Ik}^{(s)}}\right)}_{=\hat\mu^{(s)}},\underbrace{\frac{\sum_{k=1}^{n_1} w_{1k}^{(s)}Z_{1k}^{(s)}}{\sum_{k=1}^{n_1} w_{1k}^{(s)}}-\hat\mu^{(s)}}_{=\hat\alpha_1^{(s)}},\right.$$

$$\left.\dots,\underbrace{\frac{\sum_{k=1}^{n_I} w_{Ik}^{(s)}Z_{Ik}^{(s)}}{\sum_{k=1}^{n_I} w_{Ik}^{(s)}}-\hat\mu^{(s)}}_{=\hat\alpha_{I-1}^{(s)}}\right)^{\mathrm T}$$

$$Z_{jk}^{(s)}=\widehat{\boldsymbol\beta}_j^{(s)}+r_{jk}^{(s)}q_{jk}^{(s)}\qquad\text{a}\qquad r_{jk}^{(s)}=Y_{jk}-\mu_{jk}^{(s)}$$

$$\boxed{\mu_{jk}^{(s)}}=g^{-1}\left(\mathbf{x}_{jk}^{\mathrm T}\widehat{\boldsymbol\beta}^{(s)}\right)=$$

$$=\begin{cases} g^{-1}\left(\hat\mu^{(s)}+\hat\alpha_j^{(s)}\right)=\boxed{g^{-1}\left(\frac{\sum_{k=1}^{n_j} w_{jk}^{(s)}Z_{jk}^{(s)}}{\sum_{k=1}^{n_j} w_{jk}^{(s)}}\right)} & j\neq I\\ g^{-1}\left(\hat\mu^{(s)}-\hat\alpha_1^{(s)}-\cdots-\hat\alpha_{I-1}^{(s)}\right)=g^{-1}\left(\hat\mu^{(s)}+\hat\alpha_I^{(s)}\right)=\boxed{g^{-1}\left(\frac{\sum_{k=1}^{n_I} w_{Ik}^{(s)}Z_{Ik}^{(s)}}{\sum_{k=1}^{n_I} w_{Ik}^{(s)}}\right)} & j=I\end{cases}$$

(4) Jinou variantou je dodatečná podmínka

$$\boxed{\sum_{j=1}^{I} n_j\beta_j = 0} \qquad \Rightarrow \qquad \alpha_I = -\frac{n_1}{n_I}\alpha_1 - \cdots - \frac{n_{I-1}}{n_I}\alpha_{I-1}$$

s vektorem neznámých parametrů

$$\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_{I-1})^{\mathrm{T}}$$

a s maticí plánu

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & 0 & \cdots & 0 & 1 \\ \vdots & -\frac{n_1}{n_I} & \cdots & \cdots & -\frac{n_{I-1}}{n_I} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -\frac{n_1}{n_I} & \cdots & \cdots & -\frac{n_{I-1}}{n_I} \end{pmatrix}$$

Ukážeme nyní, že v tomto případě představuje parametr $\boxed{\mu}$ **obecnou střední hodnotu** (je váženým průměrem jednotlivých úrovní) a parametry $\boxed{\alpha_1, \ldots, \alpha_I}$ představují **efekty** jednotlivých úrovní faktoru A.

$$\mathbf{X}^{\mathrm{T}}\mathbf{W} =$$

$$\begin{pmatrix} w_{11} & \cdots & w_{1n_1} & w_{21} & \cdots & w_{2n_2} & \cdots & \cdots & \cdots & w_{I1} & \cdots & w_{In_I} \\ w_{11} & \cdots & w_{1n_1} & 0 & \cdots & 0 & 0 & \cdots & 0 & -\frac{n_1}{n_I}w_{I1} & \cdots & -\frac{n_1}{n_I}w_{In_I} \\ 0 & \cdots & 0 & w_{21} & \cdots & w_{2n_2} & 0 & \cdots & 0 & -\frac{n_2}{n_I}w_{I1} & \cdots & -\frac{n_2}{n_I}w_{In_I} \\ \vdots & & & \ddots & & & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & & & & \ddots & & & \ddots & 0 & \vdots & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & w_{I-1,1} & \cdots & w_{I-1,n_{I-1}} & -\frac{n_{I-1}}{n_I}w_{I1} & \cdots & -\frac{n_{I-1}}{n_I}w_{In_I} \end{pmatrix}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X} =$$

$$\begin{pmatrix} \sum\limits_{j=1}^{I}\sum\limits_{k=1}^{n_j} w_{jk} & \sum\limits_{k=1}^{n_1} w_{1k} - \frac{n_1}{n_I}\sum\limits_{k=1}^{n_I} w_{Ik} & \cdots & \cdots & \sum\limits_{k=1}^{n_{I-1}} w_{I-1,k} - \frac{n_{I-1}}{n_I}\sum\limits_{k=1}^{n_I} w_{Ik} \\ \sum\limits_{k=1}^{n_1} w_{1k} - \frac{n_1}{n_I}\sum\limits_{k=1}^{n_I} w_{Ik} & \sum\limits_{k=1}^{n_1} w_{1k} + \frac{n_1^2}{n_I^2}\sum\limits_{k=1}^{n_I} w_{Ik} & \frac{n_1 n_2}{n_I^2}\sum\limits_{k=1}^{n_I} w_{Ik} & \cdots & \frac{n_1 n_{I-1}}{n_I^2}\sum\limits_{k=1}^{n_I} w_{Ik} \\ \vdots & \frac{n_1 n_2}{n_I^2}\sum\limits_{k=1}^{n_I} w_{Ik} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sum\limits_{k=1}^{n_I} w_{Ik} \\ \sum\limits_{k=1}^{n_{I-1}} w_{I-1,k} - \frac{n_{I-1}}{n_I}\sum\limits_{k=1}^{n_I} w_{Ik} & \frac{n_1 n_{I-1}}{n_I^2}\sum\limits_{k=1}^{n_I} w_{Ik} & \cdots & \frac{n_{I-2} n_{I-1}}{n_I^2}\sum\limits_{k=1}^{n_I} w_{Ik} & \sum\limits_{k=1}^{n_{I-1}} w_{I-1,k} + \frac{n_{I-1}^2}{n_I^2}\sum\limits_{k=1}^{n_I} w_{Ik} \end{pmatrix}$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Z} = \left(\sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk}Z_{jk}, \sum_{k=1}^{n_1} w_{1k}Z_{1k} - \frac{n_1}{n_I}\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}, \ldots, \sum_{k=1}^{n_{I-1}} w_{I-1,k}Z_{I-1,k} - \frac{n_{I-1}}{n_I}\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}\right)^{\mathrm{T}}$$

Pokud systém rovnic $\quad \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Z} \quad$ rozepíšeme, máme

$$\mu \sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk} + \alpha_1 \left(\sum_{k=1}^{n_1} w_{1k} - \frac{n_1}{n_I}\sum_{k=1}^{n_I} w_{Ik}\right) + \cdots + \alpha_{I-1}\left(\sum_{k=1}^{n_{I-1}} w_{I-1,k} - \frac{n_{I-1}}{n_I}\sum_{k=1}^{n_I} w_{Ik}\right) = \\ = \sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk}Z_{jk}$$

a pro $j = 1, \ldots, I-1$

$$\mu \left(\sum_{k=1}^{n_j} w_{jk} - \frac{n_j}{n_I}\sum_{k=1}^{n_I} w_{Ik}\right) + \alpha_1 \frac{n_1 n_j}{n_I^2}\sum_{k=1}^{n_I} w_{Ik} + \cdots + \alpha_j \left(\sum_{k=1}^{n_j} w_{jk} + \frac{n_j^2}{n_I^2}\sum_{k=1}^{n_I} w_{Ik}\right) + \cdots + \\ \alpha_{I-1}\frac{n_j n_{I-1}}{n_I^2}\sum_{k=1}^{n_I} w_{Ik} = \sum_{k=1}^{n_j} w_{jk}Z_{jk} - \frac{n_j}{n_I}\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}$$

Jestliže od první rovnice, vynásobené výrazem n_I^2, odečteme ostatní, které jsou také vynásobené výrazem n_I^2, dostaneme

$$\mu \left(\underbrace{n_I^2 \sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk} - n_I^2 \sum_{k=1}^{n_1} w_{1k} - \cdots - n_I^2 \sum_{k=1}^{n_{I-1}} w_{I-1,k}}_{=n_I^2 \sum_{k=1}^{n_I} w_{Ik}} + n_I(n_1 + \cdots + n_{I-1})\sum_{k=1}^{n_I} w_{Ik} \right) + \\ \alpha_1 \left(n_I^2 \sum_{k=1}^{n_1} w_{1k} - n_1 n_I \sum_{k=1}^{n_I} w_{Ik} - n_I^2 \sum_{k=1}^{n_1} w_{1k} - n_1(n_1 + \cdots + n_{I-1})\sum_{k=1}^{n_I} w_{Ik}\right) + \\ \vdots \\ \alpha_{I-1}\left(n_I^2 \sum_{k=1}^{n_{I-1}} w_{I-1,k} - n_{I-1}n_I \sum_{k=1}^{n_I} w_{Ik} - n_I^2 \sum_{k=1}^{n_{I-1}} w_{I-1,k} - n_{I-1}(n_1 + \cdots + n_{I-1})\sum_{k=1}^{n_I} w_{Ik}\right) = \\ = \underbrace{n_I^2 \left(\sum_{j=1}^{I}\sum_{k=1}^{n_j} w_{jk}Z_{jk} - \sum_{k=1}^{n_1} w_{1k}Z_{1k} - \cdots - \sum_{k=1}^{n_{I-1}} w_{I-1,k}Z_{I-1,k}\right)}_{n_I^2 \sum_{k=1}^{n_I} w_{Ik}Z_{Ik}} + \\ + n_I(n_1 + \cdots + n_{I-1})\sum_{k=1}^{n_I} w_{Ik}Z_{Ik},$$

tedy

$$(n_1 + \cdots + n_I)\sum_{k=1}^{n_I} w_{Ik}\left(n_I\mu \underbrace{- n_1\alpha_1 - \cdots - n_{I-1}\alpha_{I-1}}_{=n_I\alpha_I}\right) = (n_1 + \cdots + n_I)n_I \sum_{k=1}^{n_I} w_{Ik}Z_{Ik}$$

a odtud

$$\boxed{\alpha_I = \frac{\sum_{k=1}^{n_I} w_{Ik}Z_{Ik}}{\sum_{k=1}^{n_I} w_{Ik}} - \mu}.$$

Pro $j = 1, \dots, I-1$ můžeme pak psát (pokud rovnice vynásobíme výrazem n_I^2)

$$\mu\, n_I \left(n_I \sum_{k=1}^{n_j} w_{jk} - n_j \sum_{k=1}^{n_I} w_{Ik} \right) + \alpha_j n_I^2 \sum_{k=1}^{n_j} w_{jk} + n_j \sum_{k=1}^{n_I} w_{Ik} \underbrace{(n_1\alpha_1 + \dots + n_{I-1}\alpha_{I-1})}_{=-n_I\alpha_I} =$$

$$= n_I^2 \sum_{k=1}^{n_j} w_{jk} Z_{jk} - n_I n_j \sum_{k=1}^{n_I} w_{Ik} Z_{Ik}$$

a po jednoduchých úpravách dostaneme

$$\boxed{\alpha_j = \frac{\sum_{k=1}^{n_j} w_{jk} Z_{jk}}{\sum_{k=1}^{n_j} w_{jk}} - \mu}.$$

Protože

$$0 = n_1\alpha_1 + \dots + n_I\alpha_I = n_1 \frac{\sum_{k=1}^{n_1} w_{1k} Z_{1k}}{\sum_{k=1}^{n_1} w_{1k}} + \dots + n_I \frac{\sum_{k=1}^{n_I} w_{Ik} Z_{Ik}}{\sum_{k=1}^{n_I} w_{Ik}} - \mu(n_1 + \dots + n_I),$$

ihned máme

$$\boxed{\mu = \frac{n_1 \frac{\sum_{k=1}^{n_1} w_{1k} Z_{1k}}{\sum_{k=1}^{n_1} w_{1k}} + \dots + n_I \frac{\sum_{k=1}^{n_I} w_{Ik} Z_{Ik}}{\sum_{k=1}^{n_I} w_{Ik}}}{n_1 + \dots + n_I}}.$$

Pak

$$\widehat{\boldsymbol\beta}^{(s+1)} = \left(\underbrace{\frac{n_1 \frac{\sum_{k=1}^{n_1} w_{1k} Z_{1k}}{\sum_{k=1}^{n_1} w_{1k}} + \dots + n_I \frac{\sum_{k=1}^{n_I} w_{Ik} Z_{Ik}}{\sum_{k=1}^{n_I} w_{Ik}}}{n_1 + \dots + n_I}}_{=\hat\mu^{(s)}}, \underbrace{\frac{\sum_{k=1}^{n_1} w_{1k}^{(s)} Z_{1k}^{(s)}}{\sum_{k=1}^{n_1} w_{1k}^{(s)}} - \hat\mu^{(s)}}_{=\hat\alpha_1^{(s)}}, \right.$$

$$\left. \dots, \underbrace{\frac{\sum_{k=1}^{n_I} w_{Ik}^{(s)} Z_{Ik}^{(s)}}{\sum_{k=1}^{n_I} w_{Ik}^{(s)}} - \hat\mu^{(s)}}_{=\hat\alpha_{I-1}^{(s)}} \right)^{\mathrm{T}}$$

$$Z_{jk}^{(s)} = \widehat{\boldsymbol\beta}_j^{(s)} + r_{jk}^{(s)} q_{jk}^{(s)} \qquad \text{a} \qquad r_{jk}^{(s)} = Y_{jk} - \mu_{jk}^{(s)}$$

$$\boxed{\mu_{jk}^{(s)}} = g^{-1}\left(\mathbf{x}_{jk}^{\mathrm{T}} \widehat{\boldsymbol\beta}^{(s)}\right) =$$

$$= \begin{cases} g^{-1}\left(\hat\mu^{(s)} + \hat\alpha_j^{(s)}\right) = \boxed{g^{-1}\left(\frac{\sum_{k=1}^{n_j} w_{jk}^{(s)} Z_{jk}^{(s)}}{\sum_{k=1}^{n_j} w_{jk}^{(s)}}\right)} & j \neq I \\ g^{-1}\left(\hat\mu^{(s)} - \frac{n_1\hat\alpha_1^{(s)} + \dots + n_{I-1}\hat\alpha_{I-1}^{(s)}}{n_I}\right) = g^{-1}\left(\hat\mu^{(s)} + \hat\alpha_I^{(s)}\right) = \boxed{g^{-1}\left(\frac{\sum_{k=1}^{n_I} w_{Ik}^{(s)} Z_{Ik}^{(s)}}{\sum_{k=1}^{n_I} w_{Ik}^{(s)}}\right)} & j = I \end{cases}$$

Závěrem je třeba zdůraznit, že **odhady neznámých parametrů**

$$\hat{\mu}^{(s)}, \hat{\alpha_1}^{(s)}, \ldots, \hat{\alpha_I}^{(s)}$$

jednotlivých modelů sice **závisí** na výběru vedlejších reparametrizačních podmínek (a mají jinou interpretaci), ale **odhady středních hodnot závislé proměnné**, tj.

$$\boxed{\widehat{EY_{jk}} = \hat{Y}_{jk}^{(s)} = \hat{\mu}_{jk}^{(s)}}, \qquad (j = 1, \ldots, I; \;\; k = 1, \ldots, n_j)$$

nezávisí na výběru vedlejších podmínek, což jsme také demonstrovali.

2.9.2 Dvojné třídění

Model dvojného třídění odpovídá uspořádání experimentu, v němž se sleduje vliv dvou faktorů A, B. Každý z faktorů se může vyskytovat na několika úrovních. Kombinace úrovní j, k určuje tzv. podtřídu s n_{jk} pozorováními náhodných veličin Y_{jki}, $j = 1, \ldots, I$; $j = 1, \ldots, J$. Vyjděme z tohoto základního modelu

$$\boxed{GLM_* : \eta_{jki} = g(EY_{jki}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}} \qquad j = 1, \ldots, I \quad k = 1, \ldots, J.$$

kde $\boxed{\mu}$ je **obecná střední hodnota**

$\boxed{\alpha_j}$ **efekt** j-té úrovně faktoru $\boxed{\text{A}}$

$\boxed{\beta_k}$ **efekt** k-té úrovně faktoru $\boxed{\text{B}}$

$\boxed{(\alpha\beta)_{jk}}$ **interakce** mezi j-tou úrovní faktoru $\boxed{\text{A}}$ a k-tou úrovní faktoru $\boxed{\text{B}}$

Výsledky se obvykle zapisují do tabulky, v níž řádky odpovídají úrovním faktoru A a sloupce úrovním faktoru B.

Matice plánu odpovídajícímu uvedenému modelu nemá plnou hodnost. Aby měly normální rovnice jediné řešení, kladou se opět na neznámé parametry další podmínky, které dohromady s normálními rovnicemi vytvoří soustavu, jejíž řešení je určeno jednoznačně.

Submodely uvažované k modelu dvojného třídění odpovídají hypotézám o nulovost efektů faktoru A nebo faktoru B nebo interakcí

$$H_A : \quad \alpha_1 = \cdots = \alpha_I = 0$$

$$H_B : \quad \beta_1 = \cdots = \beta_J = 0$$

$$H_{AB} : \quad (\alpha\beta)_{jk} = 0 \qquad j \in \{1, \ldots, I\} \quad k \in \{1, \ldots, J\}$$

Zatím jsme blíže nespecifikovali index i. Uvažujme proto následující dva modely, které závisí na výskytu interakcí mezi faktory A a B:

Křížové faktory (Crossed Factors)

Příklad: provádí se testování dvou léků ve třech nemocnicích, přičemž každý lék se testuje v všech nemocnicích.

Nemocnice	**Lék** L_1			**Lék** L_2		
	N_1	N_2	N_3	N_1	N_2	N_3
Pozorování	$Y_{11,1}$	$Y_{21,1}$	$Y_{31,1}$	$Y_{12,1}$	$Y_{22,1}$	$Y_{32,1}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$Y_{11,n_{11}}$	$Y_{21,n_{21}}$	$Y_{31,n_{13}}$	$Y_{12,n_{12}}$	$Y_{22,n_{22}}$	$Y_{32,n_{32}}$

Hiearchické faktory (Nested Factors)

Příklad: provádí se testování dvou léků v šesti nemocnicích, přičemž první lék se testuje v první až třetí nemocnici a druhý lék ve zbývajících nemocnicích.

Nemocnice	**Lék** L_1			**Lék** L_2		
	N_1	N_2	N_3	N_4	N_5	N_6
Pozorování	$Y_{11,1}$	$Y_{21,1}$	$Y_{31,1}$	$Y_{42,1}$	$Y_{52,1}$	$Y_{62,1}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	Y_{11,n_1}	Y_{21,n_2}	Y_{31,n_3}	Y_{42,n_4}	Y_{52,n_5}	Y_{62,n_6}

2.10 ANALÝZA KOVARIANCE

Při analýze kovariance se uplatňují tyto **proměnné**:

1. jeden či několik **faktorů** $\boxed{\text{A,B,}\ldots}$, přičemž stejně jako v analýze (deviace) rozptylu jde o nominální nebo alternativní proměnné;

2. jedna vysvětlovaná či závislá proměnná $\boxed{Y}$, na niž je při analýze soustředěna pozornost v tom smyslu, že chceme prokázat jejich závislost na faktoru či faktorech;

3. jedna nebo více **doprovodných proměnných** (také kontrolovaných proměnných) $\boxed{x_1, \ldots, x_p}$, které zahrnujeme do modelu a počítáme s nimi zejména proto, abychom **závislost vysvětlované proměnné na faktorech očistili od jejich vlivu**.

Jako příklad si uvedeme jednoduchý model s jedním faktorem $\boxed{\text{A}}$ a s jednou doprovodnou proměnnou $\boxed{x}$

$$\boxed{GLM_* : \eta_{jk} = g(EY_{jk}) = \mu + \gamma x_{jk} + \alpha_j} \qquad j = 1, \ldots, I \quad k = 1, \ldots, n_j.$$

a přidejme vedlejší podmínku:

$$\boxed{\alpha_1 = 0}.$$

Vektor neznámých parametrů je

$$\boldsymbol{\beta} = (\mu, \gamma, \alpha_2, \ldots, \alpha_I)^{\mathrm{T}}$$

a matice plánu má v tomto případě tvar

$$\mathbf{X} = \begin{pmatrix}
1 & x_{11} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\underline{1} & \underline{x_{1n_1}} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{0} \\
1 & x_{21} & 1 & 0 & \cdots & \cdots & \cdots & 0 \\
1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\underline{1} & \underline{x_{2n_2}} & \underline{1} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{0} \\
1 & \vdots & 0 & \cdots & \cdots & 0 & 1 & 0 \\
1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\underline{1} & \underline{\vdots} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{0} & \underline{1} & \underline{0} \\
1 & x_{I1} & 0 & \cdots & \cdots & \cdots & 0 & 1 \\
1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_{In_I} & 0 & \cdots & \cdots & \cdots & 0 & 1
\end{pmatrix}$$

Tento model odpovídá situaci, kdy náhodný výběr se rozpadá podle úrovní faktoru A na I náhodných výběrů, ve kterých regresní přímka (vzhledem k lineárnímu prediktoru η_{jk}) má

- stále **stejnou směrnici** $\boxed{\gamma}$ (tj. přímky jsou rovnoběžné) a
- **různé úseky** $\boxed{\mu, \mu + \alpha_1, \dots, \mu + \alpha_I}$.

Při jednoduchém třídění testujeme hypotézu

$$H_0 : \alpha_1 = \cdots = \alpha_I = 0$$

proti alternativě, že ne všechny úseky regresní přímky jsou si rovny a základní model srovnáváme ze submodelem

$$\boxed{GLM_{sub_1} : \eta_{jk} = g(EY_{jk}) = \mu + \gamma x_{jk}} \qquad j = 1, \dots, I \quad k = 1, \dots, n_j.$$

Chceme-li otestovat vliv doprovodné proměnné můžeme testovat další submodely

$$\boxed{GLM_{sub_2} : \eta_{jk} = g(EY_{jk}) = \mu + \alpha_j} \qquad j = 1, \dots, I \quad k = 1, \dots, n_j.$$

$$\boxed{GLM_{min} : \eta_{jk} = g(EY_{jk}) = \mu} \qquad j = 1, \dots, I \quad k = 1, \dots, n_j.$$

Pokud ovšem není splněna podmínka, že v každém náhodném výběru má regresní přímka stejnou směrnici, musíme uvažovat složitější model

$$\boxed{GLM_* : \eta_{jk} = g(EY_{jk}) = \mu + \gamma_j x_{jk} + \alpha_j} \qquad j = 1, \dots, I \quad k = 1, \dots, n_j.$$

a přidejme vedlejší podmínky:

$$\boxed{\alpha_1 = 0}.$$

Vektor neznámých parametrů je

$$\boldsymbol{\beta} = (\mu, \gamma_1, \dots, \gamma_I, \alpha_2, \dots, \alpha_I)^{\mathrm{T}}$$

a matice plánu má v tomto případě tvar

$$
\mathbf{X} = \begin{pmatrix}
1 & x_{11} & 0 & \cdots & \cdots & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\underline{1} & \underline{x_{1n_1}} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{0} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{0} \\
1 & 0 & x_{21} & 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\
1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\underline{1} & \underline{0} & \underline{x_{2n_2}} & \underline{0} & \underline{\cdots} & \underline{0} & \underline{1} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{\cdots} & \underline{0} \\
1 & 0 & \vdots & \vdots & \vdots & 0 & 0 & \cdots & \cdots & 0 & 1 & 0 \\
1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\underline{1} & \underline{0} & \underline{\vdots} & \underline{\vdots} & \underline{\vdots} & \underline{0} & \underline{0} & \underline{\cdots} & \underline{\cdots} & \underline{0} & \underline{1} & \underline{0} \\
1 & 0 & \cdots & \cdots & 0 & x_{I1} & 0 & \cdots & \cdots & \cdots & 0 & 1 \\
1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0 & \cdots & \cdots & 0 & x_{In_I} & 0 & \cdots & \cdots & \cdots & 0 & 1
\end{pmatrix}
$$

Tento model pak odpovídá situaci, kdy náhodný výběr se rozpadá podle úrovní faktoru A na I náhodných výběrů, ve kterých regresní přímka (vzhledem k lineárnímu prediktoru η_{jk}) má

- **různé směrnice** $\boxed{\gamma_1, \ldots, \gamma_I}$ (tj. přímky nejjsou rovnoběžné) a
- **různé úseky** $\boxed{\mu, \mu + \alpha_1, \ldots, \mu + \alpha_I}$.

Při jednoduchém třídění testujeme hypotézu

$$H_0 : \alpha_1 = \cdots = \alpha_I = 0$$

proti alternativě, že ne všechny úseky regresní přímky jsou si rovny a základní model srovnáváme ze submodelem

$$\boxed{GLM_{sub_1} : \eta_{jk} = g(EY_{jk}) = \mu + \gamma_j x_{jk}} \qquad j = 1, \ldots, I \quad k = 1, \ldots, n_j.$$

Pokud bychom chtěli vyloučit vliv faktorů nejen na úseky, ale i na směrnice, tj. popíráme i vliv doprovodné proměnné, můžeme základní model srovnat s modelem

$$\boxed{GLM_{min} : \eta_{jk} = g(EY_{jk}) = \mu} \qquad j = 1, \ldots, I \quad k = 1, \ldots, n_j.$$

Pokud chceme testovat pouze vliv doprovodné proměnné, můžeme základní model srovnat s modelem

$$\boxed{GLM_{sub_2} : \eta_{jk} = g(EY_{jk}) = \mu + \alpha_j} \qquad j = 1, \ldots, I \quad k = 1, \ldots, n_j.$$

2.11 Typy dat

Dříve než začneme členit GLM modely podle toho, jaké rozdělení má závisle proměnná Y, nejprve se obecně zamysleme nad typem proměnných (ať už závisle či nezávisle proměnných), a to z hlediska vztahů mezi dvěma hodnotami.

(1) **Nominální** proměnná je taková, o jejíž dvou hodnotách můžeme pouze říci, zda jsou stejné či různé (škola, fakulta, obor, krevní skupiny: A, B, O, A/B). Hodnotami mohou být texty (písmena), případně i číselné kódy.

(2) **Ordinální (pořadová)**, u jejíž dvou hodnot můžeme navíc určit pořadí (úroveň spokojenosti, vzdělání). Jako hodnoty lze použít text, datum, číslo.

(3) **Intervalová (rozdílová)** proměnná je taková, pro jejíž dvě hodnoty můžeme navíc (k možnostem ordinální proměnné) vypočítat, o kolik je jedna hodnota větší (resp. menší) než druhá (měsíční příjem domácnosti, počet dětí v rodině). Hodnotami jsou tedy čísla.

(4) **Poměrová (podílová)** proměnná je ta, pro jejíž dvě hodnoty můžeme navíc (k možnostem intervalové proměnné) vypočítat, kolikrát je jedna hodnota větší (resp. menší) než druhá, tzn. jedná se pouze o kladné hodnoty (počet členů domácnosti).

Nominální a *ordinální* proměnné jsou souhrnně označovány jako **kvalitativní**; *intervalové* a *poměrové* proměnné jsou souhrnně označovány jako **kvantitativní**.

Kvantitativní proměnné můžeme podle jiného hlediska dělit na

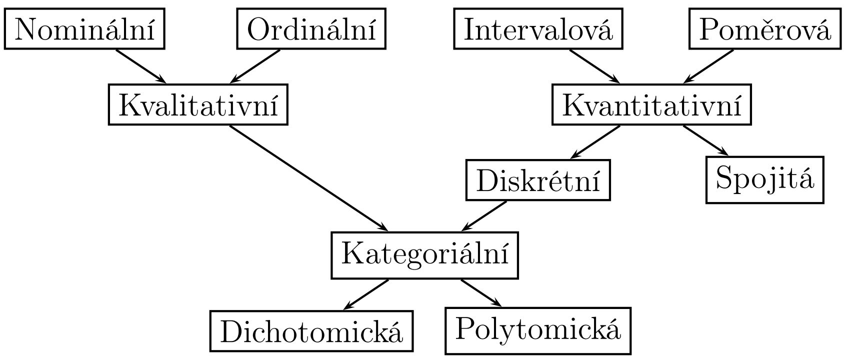
(a) **Diskrétní**, které nabývají pouze celočíselných obměn (počet válců automobilu), a

(b) **Spojité**, jež mohou nabývat libovolných hodnot z určitého intervalu (věk respondenta, cena výrobku, roční příjem domácnosti).

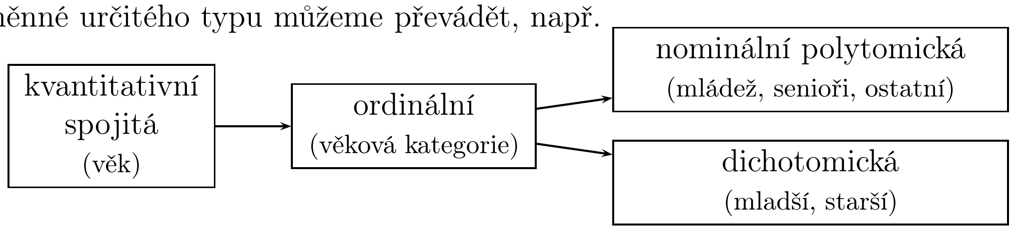
Nominální, *ordinální* a *kvantitativní diskrétní* proměnné můžeme souhrnně označit jako **kategoriální** (obměny těchto proměnných nazýváme kategoriemi). Podle jiného hlediska je můžeme dělit na

(i) **dichotomické (alternativní, binární)**, které nabývají pouze dvou kategorií (ekonomicky aktivní a neaktivní, kuřák a nekuřák), a

(ii) **polytomické (množné)**, jež nabývají více než dvou kategorií (rodinný stav, obor).



Proměnné určitého typu můžeme převádět, např.



2.12 ALTERNATIVNÍ A BINOMICKÁ DATA

2.12.1 Úvod

Předpokládejme, že sledovaná náhodná veličina U_i $(i = 1, \ldots, N)$ nabývá pouze dvou hodnot 0 a 1, tj. má alternativní rozdělení:

$$\boxed{U_i \sim A(\pi_i)} \sim \boxed{f_U(u)} = P(U_i = u) = \begin{cases} \pi_i & u = 1 \\ 1 - \pi_i & u = 0 \\ 0 & \text{jinak} \end{cases} = \boxed{\begin{cases} \pi_i^u (1-\pi_i)^{1-u} & u = 0, 1 \\ 0 & \text{jinak} \end{cases}}.$$

Předpokládejme, že náhodná veličina U_i závisí na m veličinách $x_{i1}, \ldots, x_{im}$, tzv. **kovariáty**. Data můžeme mít zadána různým způsobem:

- **jednotlivá pozorování** $\boxed{U_i}$:

hodnoty kovariát	pozorované binární veličiny
$x_{i1}, \ldots, x_{im}$	U_i

- **skupinově**, tj. pro každou kombinaci kovariát známe **absolutní četnosti** úspěchů $\boxed{Y_j}$ a celkový počet pokusů $\boxed{n_j}$, tedy máme k dispozici binomická data

$$\boxed{Y_j = \sum_{i=1}^{n_j} U_i \sim Bi(n_j, \pi_j) \sim f_Y(y) = P(Y_j = y) = \begin{cases} \binom{n_j}{y} \pi_j^y (1-\pi_j)^{n_j - y} & y = 0, 1, \ldots, n_j \\ 0 & \text{jinak} \end{cases}}$$

kde

$$j = 1, \ldots, n; \qquad N = n_1 + \cdots + n_n$$

a data můžeme zapsat formou tabulky

hodnota kovariát	počet úspěchů	počet pokusů
$x_{j1}, \ldots, x_{jm}$	Y_j	n_j

- **skupinově**, tj. pro každou kombinaci kovariát máme **relativní četnost** úspěchů $\boxed{Z_j = \frac{Y_j}{n_j}}$ a celkový počet pokusů $\boxed{n_j}$

$$\boxed{Z_j = \frac{Y_j}{n_j} = \frac{1}{n_j} \sum_{i=1}^{n_j} U_i \sim f_Z(y) = P(Z_j = y) = \begin{cases} \binom{n_j}{n_j y} \pi_j^{n_j y} (1-\pi_j)^{n_j - n_j y} & y = 0, \frac{1}{n_j}, \ldots, 1 \\ 0 & \text{jinak} \end{cases}}$$

kde

$$j = 1, \ldots, n; \qquad N = n_1 + \cdots + n_n$$

Data lze zapsat do tabulky

kovariáty	relativní úspěšnost	počet pokusů
$x_{j1}, \ldots, x_{jm}$	$Z_j = \frac{Y_j}{n_j}$	n_j

- pro **nominální** či **ordinální kovariáty** můžeme data psát do tzv. **kontingenčních tabulek.** Uvažujme jednoduchý příklad:

kovariáty		$U=0$	$U=1$
$x_1 = 1$	$x_2 = 1$	n_{110}	n_{111}
	$x_2 = 2$	n_{120}	n_{121}
$x_1 = 2$	$x_2 = 1$	n_{210}	n_{211}
	$x_2 = 2$	n_{220}	n_{221}

V dalším se soustřeďme na **relativní četnosti úspěchů**

$$\boxed{Z_i = \frac{Y_i}{n_i}}.$$

Hlavním úkolem statistické analýzy je pak nalézt vztah mezi Z_i, (tj. i Y_i) a $x_{i1}, \ldots, x_{im}$, tj. funkci

$$\boxed{\pi_i = \pi(\mathbf{x}_i) = \pi(x_{i1}, \ldots, x_{im})}.$$

Protože chceme použít GLM modely, modelujeme pravděpodobnosti π_i pomocí linkovacích funkcí

$$g(\pi_i) = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}.$$

Nejjednodušším modelem je **lineární model**

$$\boxed{\pi_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}}.$$

Avšak tento model má řadu nevýhod, především je třeba zajistit, aby $\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}$ nabývala hodnot mezi 0 a 1, tedy je třeba přidat nějaké dodatečné podmínky.

Proto, abychom tuto podmínku dodrželi, využijeme nějakou **distribuční funkci**

$$F(t) = \int_{-\infty}^{t} f(s)ds \qquad f(s) \geq 0 \qquad \int_{-\infty}^{\infty} f(s)ds = 1$$

s odpovídající **hustotou** $f(s)$, která se v tomto případě nazývá **toleranční funkce** (toleranční distribuce).

Nyní si ukážeme několik modelů, které využívají různé toleranční distribuce.

2.12.2 Modely dávka-odpověď

Typickým příkladem těchto modelů je vztah mezi dávkou toxické látky a odezvy (kladná-přežití, záporná-smrt) jedince na tuto dávku. Odezvy bývají obvykle udávány jako procenta kladné odezvy (quantal responses).

SYMETRICKÉ MODELY

Jestliže uvažujeme **toleranční distribuci** jako **rovnoměrně spojitou** na nějakém intervalu (a, b), tj

$$\boxed{f_0(s) \sim Rs(a,b) \qquad f(s) = \begin{cases} \frac{1}{b-a} & s \in (a,b) \\ 0 & \text{jinak} \end{cases}}$$

pak

$$\boxed{\pi_0(x) = F_0(x) = \int_a^x f(s)ds = \frac{x-a}{b-a} \quad \text{pro} \quad x \in (a,b)}$$

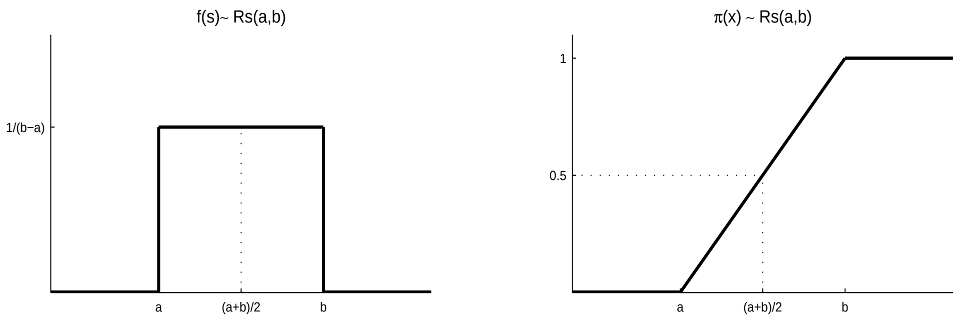
a tento model je **lineárním modelem**

$$\pi_0(x) = \frac{x-a}{b-a} = \beta_1 + \beta_2 x \qquad \text{tj.} \qquad \beta_1 = -\frac{a}{b-a} \quad \beta_2 = \frac{1}{b-a} > 0 \qquad \text{tedy } \boxed{\beta_2 > 0}$$

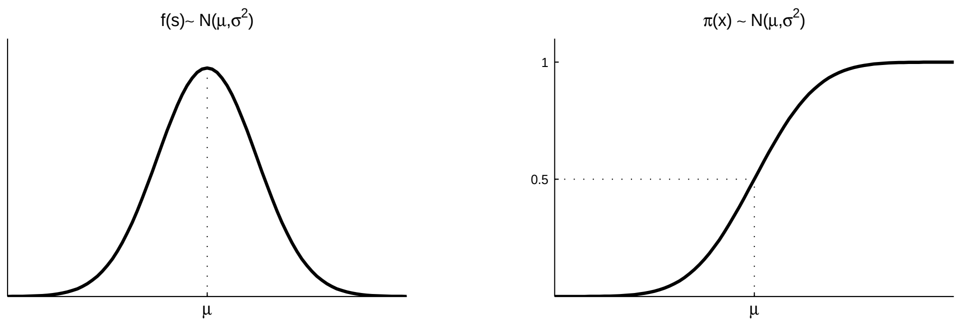
s identickou linkovací funkcí

$$\boxed{g_0(\pi) = \pi}.$$

V praxi tento model však nemá přílišné uplatnění.



Obrázek 2.6: Rovnoměrné rozdělení na (a, b).



Obrázek 2.7: Normální rozdělení $N(\mu, \sigma^2)$.

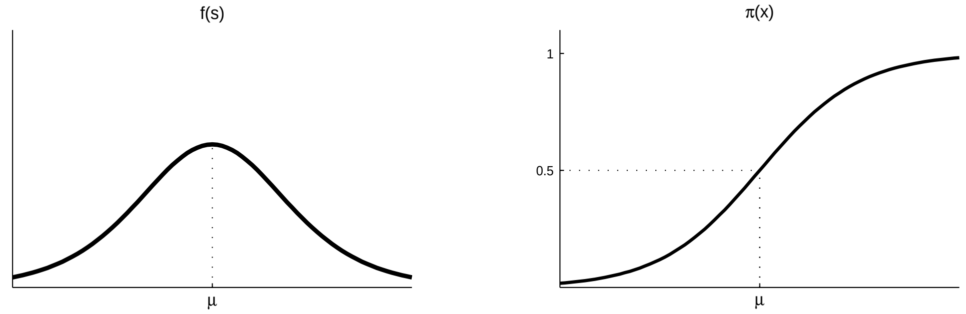
Další možností je vzít **normální hustotu** jako **toleranční funkci**. Připomeňme, že střední hodnota, medián i modus je roven parametru μ a rozptyl parametru $\sigma^2 > 0$. V tomto případě mluvíme o tzv. **PROBITOVÉM MODELU**:

$$\boxed{\pi_1(x) = F_1(x) = \int_{-\infty}^x f_1(s)ds = \int_{-\infty}^x \tfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2} ds = \Phi\left(\tfrac{x-\mu}{\sigma}\right)},$$

kde Φ je distribuční funkce standardizovaného normálního rozdělení. Pak tzv. **probitovou linkovací funkcí** je kvantilová funkce normálního rozdělení

$$\boxed{g_1(\pi) = \Phi^{-1}(\pi) = \tfrac{x-\mu}{\sigma} = \beta_1 + \beta_2 x} \qquad \text{tj.} \qquad \beta_1 = -\tfrac{\mu}{\sigma} \quad \beta_2 = \tfrac{1}{\sigma} > 0 \qquad \text{tedy } \boxed{\beta_2 > 0}.$$

Hodnota mediánu $x = \mu$ se nazývá **mediánová smrtící dávka** (*median lethal dose* - LD50) a odpovídá dávce, při které polovina jedinců má kladnou a polovina zápornou odezvu.



Obrázek 2.8: Logistické rozdělení.

Jiným velmi podobným modelem je tzv. **LOGISTICKÝ MODEL**, kde toleranční funkce je hustota **logistického rozdělení** (se střední hodnotou, mediánem i modusem μ a rozptylem $\frac{\pi^2}{3}\sigma^2 = 3.2899\sigma^2 = (1.8138\sigma)^2$)

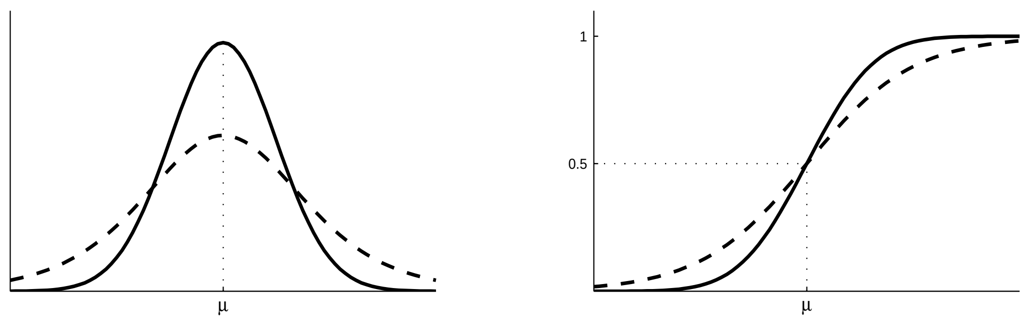
$$f_2(s) = \frac{1}{\sigma}\frac{\exp\left(\frac{s-\mu}{\sigma}\right)}{\left[1+\exp\left(\frac{s-\mu}{\sigma}\right)\right]^2} = \frac{1}{\sigma}\frac{\exp\left(-\frac{s-\mu}{\sigma}\right)}{\left[1+\exp\left(-\frac{s-\mu}{\sigma}\right)\right]^2},$$

takže

$$\boxed{\pi_2(x) = F_2(x) = \int_{-\infty}^{x} \frac{1}{\sigma}\frac{\exp\left(\frac{s-\mu}{\sigma}\right)}{\left[1+\exp\left(\frac{s-\mu}{\sigma}\right)\right]^2}\, ds = \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{1+\exp\left(\frac{x-\mu}{\sigma}\right)} = \frac{1}{1+\exp\left(-\frac{x-\mu}{\sigma}\right)}}$$

s tzv. **logit linkovací funkcí**

$$\boxed{g_2(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \frac{x-\mu}{\sigma} = \beta_1 + \beta_2 x} \qquad \text{tj.} \qquad \beta_1 = -\frac{\mu}{\sigma} \quad \beta_2 = \frac{1}{\sigma} > 0 \qquad \text{tedy } \boxed{\beta_2 > 0}.$$



Obrázek 2.9: Srovnání probitového a logistického (- - -) modelu při stejných parametrech μ a σ. Názorně je vidět, že logistický model má větší rozptyl a těžší konce.

ASYMETRICKÉ (EXTREMÁLNÍ) MODELY

Pokud za toleranční funkci zvolíme **Log-Weibullovo rozdělení** (*extreme-minimal-value distribution*) se střední hodnot

ve tvaru

$$f_3(s) = \frac{1}{\sigma}\exp\left(\frac{s-\mu}{\sigma}\right)\exp\left[-\exp\left(\frac{s-\mu}{\sigma}\right)\right],$$

pak

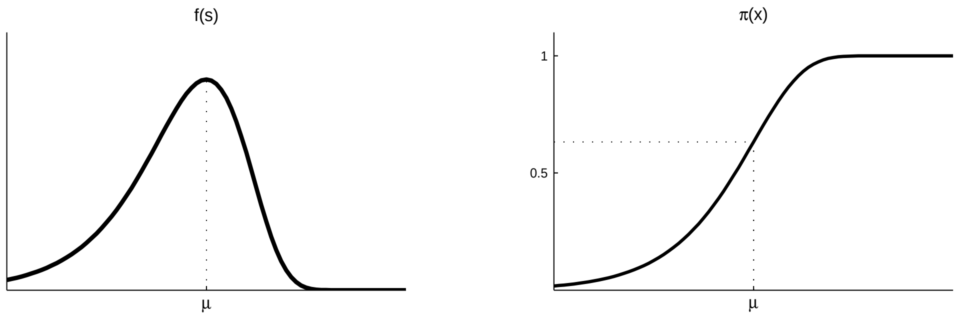
$$\boxed{\pi_3(x) = F_3(x) = \int_{-\infty}^{x} \frac{1}{\sigma}\exp\left(\frac{s-\mu}{\sigma}\right)\exp\left[-\exp\left(\frac{s-\mu}{\sigma}\right)\right]\, ds = 1 - \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right]}$$

s tzv. **komplementární log-log linkovací funkcí**

$$\boxed{g_3(\pi) = \log[-\log(1-\pi)] = \tfrac{x-\mu}{\sigma} = \beta_1 + \beta_2 x} \quad \text{tj.} \quad \beta_1 = -\tfrac{\mu}{\sigma} \quad \beta_2 = \tfrac{1}{\sigma} > 0 \quad \text{tedy } \boxed{\beta_2 > 0}.$$

Pak střední hodnota $= \mu - \gamma\sigma \doteq \mu - 0.57721\sigma$
medián $= \mu + \sigma\log(\log 2) \doteq \mu - 0.36651\sigma$
modus $= \mu$
rozptyl $= \frac{\pi^2}{6}\sigma^2 = 1.6449\sigma^2 = (1.2825\sigma)^2$,

kde $\gamma \doteq 0.57721$ je tzv. *Euler-Mascheroniho* konstanta.



Obrázek 2.10: Log-Weibullovo rozdělení.

Pokud jako toleranční funkci zvolíme **zobecněné Gumbelovo rozdělení** (*extreme-miaximal-value distribution*) ve tvaru

$$f_4(s) = \frac{1}{\sigma}\exp\left(-\frac{s-\mu}{\sigma}\right)\exp\left[-\exp\left(-\frac{s-\mu}{\sigma}\right)\right],$$

dostaneme

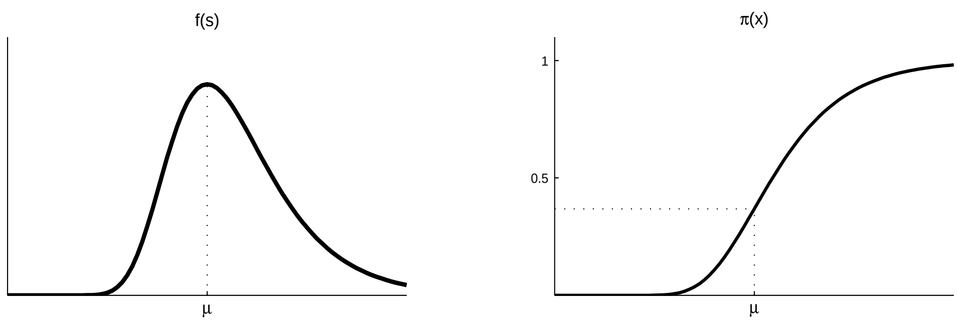
$$\boxed{\pi_4(x) = F_4(x) = \int_{-\infty}^{x} \frac{1}{\sigma}\exp\left[-\frac{s-\mu}{\sigma} - \exp\left(-\frac{s-\mu}{\sigma}\right)\right] ds = \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]}$$

s tzv. **log-log linkovací funkcí**

$$\boxed{g_4(\pi) = -\log[-\log(\pi)] = \tfrac{x-\mu}{\sigma} = \beta_1 + \beta_2 x} \quad \text{tj.} \quad \beta_1 = -\tfrac{\mu}{\sigma} \quad \beta_2 = \tfrac{1}{\sigma} > 0 \quad \text{tedy } \boxed{\beta_2 > 0}.$$

Pak střední hodnota $= \mu + \gamma\sigma \doteq \mu + 0.57721\sigma$
medián $= \mu - \sigma\log(\log 2) \doteq \mu + 0.36651\sigma$
modus $= \mu$
rozptyl $= \frac{\pi^2}{6}\sigma^2 = 1.6449\sigma^2 = (1.2825\sigma)^2$,

kde $\gamma \doteq 0.57721$ je tzv. *Euler-Mascheroniho* konstanta.



Obrázek 2.11: Zobecněné Gumbelovo rozdělení.

<u>Poznámka</u> 2.12.1.
Pokud náhodná veličina U má rozdělení **rovnoměrně spojité** na intervalu $(0,1)$, tj.

$$U \sim Rs(a,b),$$

pak náhodná veličina

$X = \mu + \sigma \log\left(\frac{U}{1-U}\right)$	má	logistické rozdělení s hustotou $f_2(x)$,
$X = \mu + \sigma \log(-\log(1-U))$		Log-Weibullovo rozdělení s hustotou $f_3(x)$,
$X = \mu - \sigma \log(-\log(U))$		Gumbelovo rozdělení s hustotou $f_4(x)$.

Těchto vztahů se využívá při generování pseudonáhodných čísel příslušných rozdělení.

2.12.3 Logistická regrese

Protože se v GLM modelech pro binární a binomická data nejčastěji používá **logit** linkovací funkce

$$g_2(\pi) = \log\left(\tfrac{\pi}{1-\pi}\right),$$

budeme se proto podrobněji věnovat interpretaci výsledků, které nabízí logistická regrese.

Předpokládejme, že závisle proměnná $\boxed{Y}$ je dichotomická (binární) proměnná, která nabývá hodnoty jedna, pokud sledovaný jev nastal, v opačném případě je rovna nule.

Protože jde o regresní model, bude nás zajímat vztah pravděpodobností úspěchu či neúspěchu k hodnotám vysvětlujících proměnných $\mathbf{x} = (x_1, \ldots, x_m)^{\mathrm{T}}$, tedy budeme zkoumat pravděpodobnosti

$$P(Y=1|x_1,\ldots,x_m) = \pi(\mathbf{x}) = \frac{\exp\{\eta(\mathbf{x})\}}{1+\exp\{\eta(\mathbf{x})\}} = \frac{1}{1+\exp\{-\eta(\mathbf{x})\}}$$

a

$$P(Y=0|x_1,\ldots,x_m) = 1-\pi(\mathbf{x}) = \frac{1}{1+\exp\{\eta(\mathbf{x})\}} = \frac{\exp\{-\eta(\mathbf{x})\}}{1+\exp\{-\eta(\mathbf{x})\}}.$$

Předpokládejme, že lineární prediktor je roven

$$\eta(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$$

a ukážeme, že absolutní člen má smysl uvádět samostatně.

Všimněme si nejprve, že podíl

$$odds(\mathbf{x}) = \frac{P(Y=1|x_1,\ldots,x_m)}{P(Y=0|x_1,\ldots,x_m)} = \frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})} = \exp(\beta_0 + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$$

má bezprostřední interpretaci. Porovnává pravděpodobnost jedničky (tj. výskyt sledovaného jevu při daných hodnotách kovairát) a nuly (nevýskyt sledovaného jevu při daných hodnotách kovairát).

Anglickému označení **odds** odpovídá české označení **šance**. Hodnota šance není shora ohraničená, zdola však nulou. Pokud zlogaritmujeme šanci, dostaneme **logit**, který nabývá hodnot od mínus do plus nekonečna.

Nyní budeme předpokládat, že máme jedinou kovariátu $\boxed{x}$, která je také dichotomická, takže nabývá dvou různých hodnot, které můžeme bez újmy na obecnosti označit jako 0 a 1.

V tom případě jde o kategoriální proměnnou, nebo–li x je umělá proměnná k dvouhodnotovému faktoru.

Za těchto podmínek je šance pro $x = 0$ rovna

$$odds(0) = \exp(\beta_0),$$

takže parametr β_0 je roven logitu pravděpodobnosti výskytu sledovaného jevu v bodě $x = 0$.

Pro $x = 1$ dostaneme

$$odds(1) = \exp(\beta_0 + \beta_1).$$

Poměr šancí (nebo také křížový poměr, anglicky *odds ratio*) pro dichotomická x je pak roven

$$OR = \frac{odds(1)}{odds(0)} = \exp(\beta_1),$$

takže parametr β_1 je roven logaritmu poměru šancí. Odtud tedy dostáváme, že pokud pravděpodobnost sledovaného jevu nezávisí na hodnotě proměnné x, je poměr šancí roven jedné, takže platí

$$\beta_1 = 0.$$

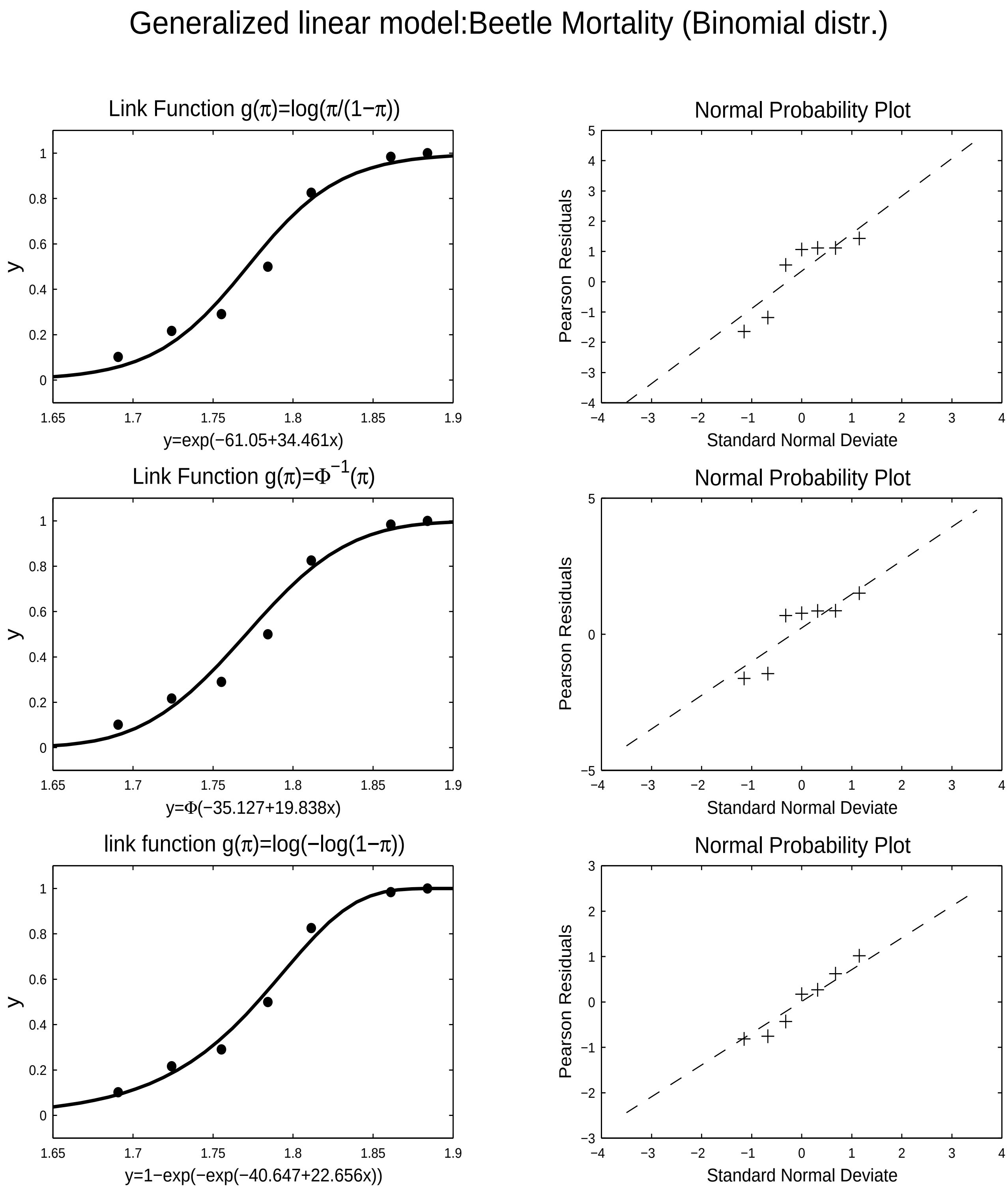
I v případě, že vysvětlující proměnná je **spojitá**, má zajímavou interpretaci především parametr β_1, neboť

$$\frac{odds(x+1)}{odds(x)} = \frac{\exp\{\beta_0 + \beta_1(x+1)\}}{\exp\{\beta_0 + \beta_1 x\}} = \exp\{\beta_1\},$$

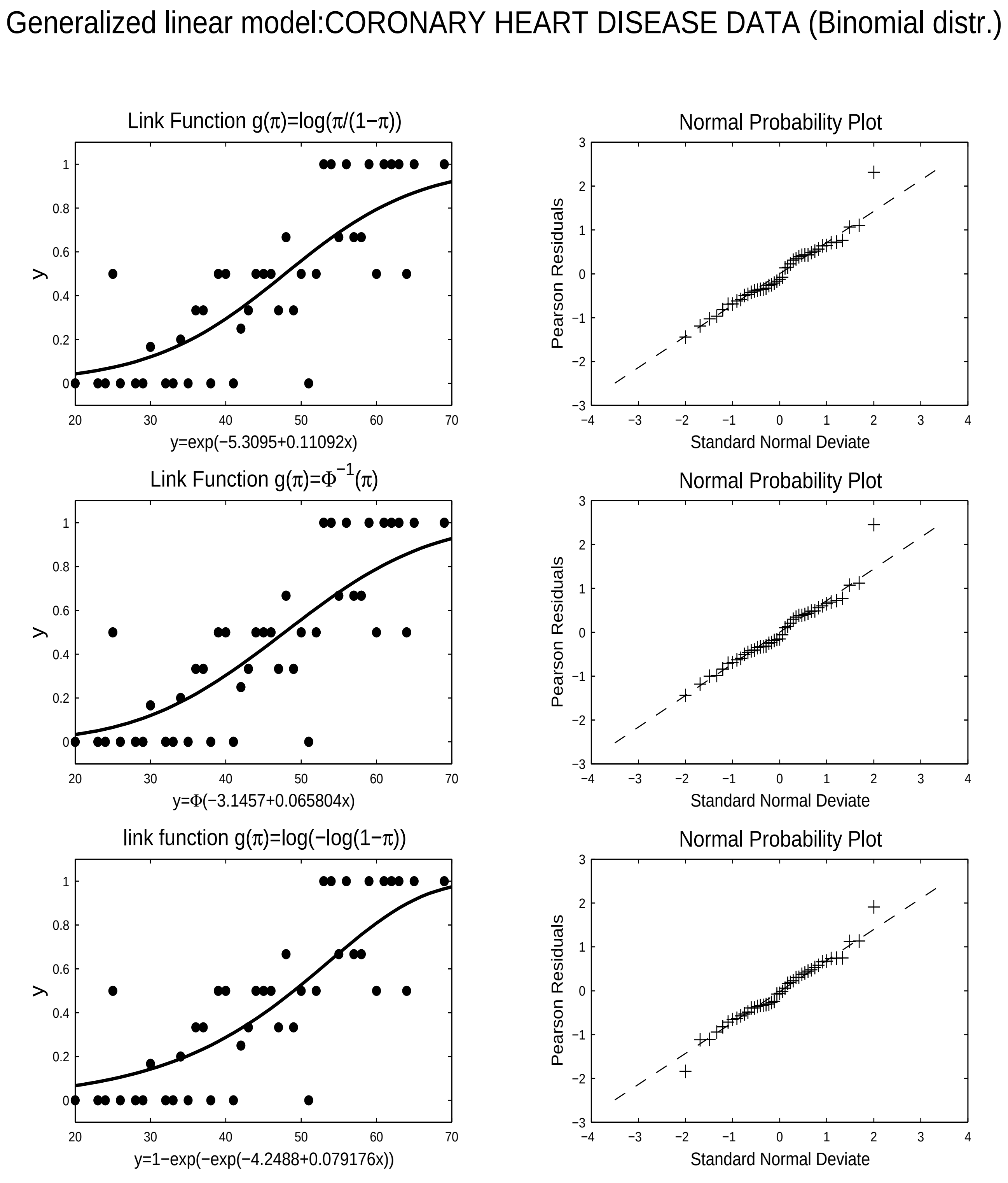
takže parametr β_1 vypovídá o změně vztažené k jednotkovému přírůstku nezávisle proměnné x, tentokrát je to změna logaritmu poměru šancí.

2.12.4 PŘÍKLADY

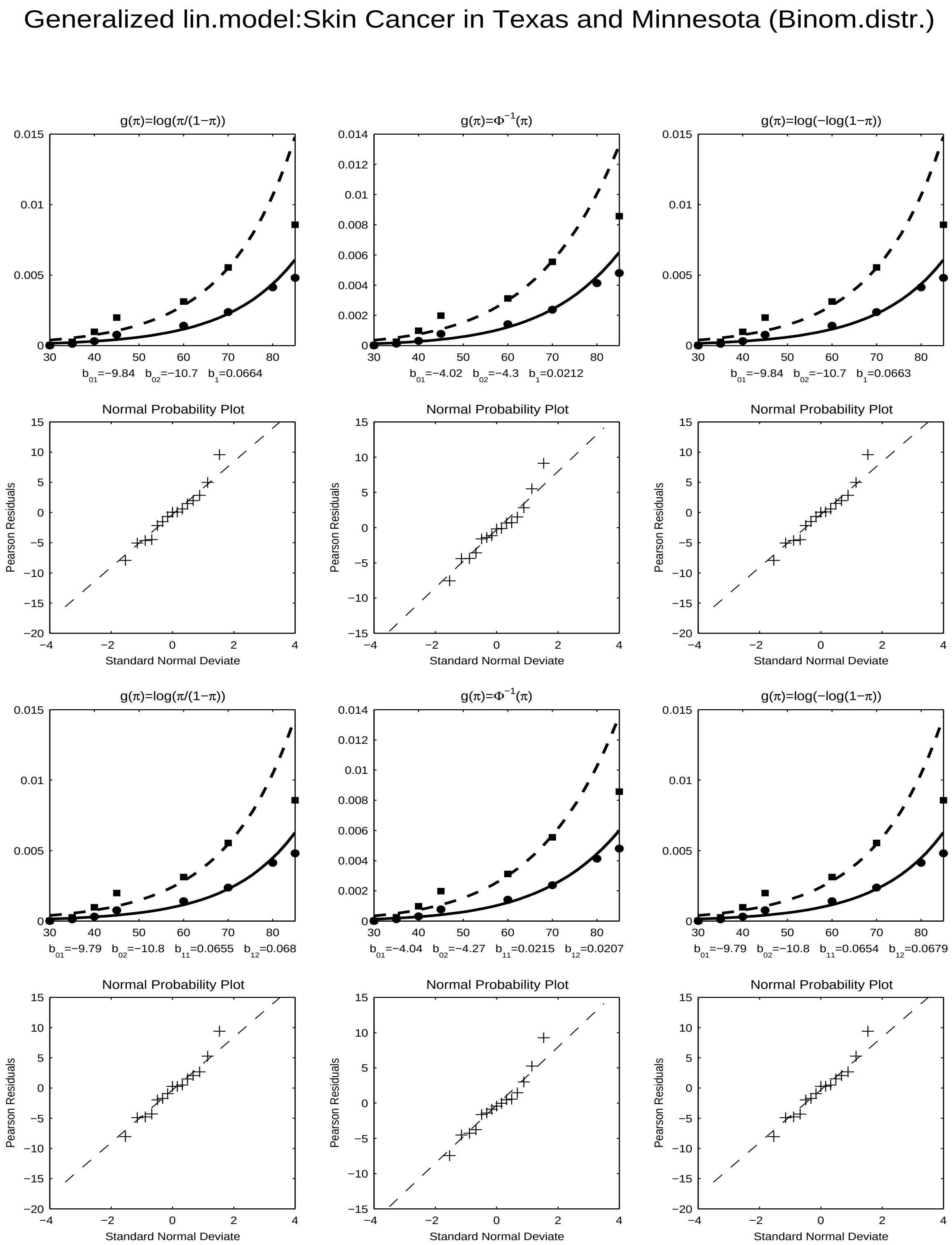
Uveďme nyní 4 aplikace pro binomická data.



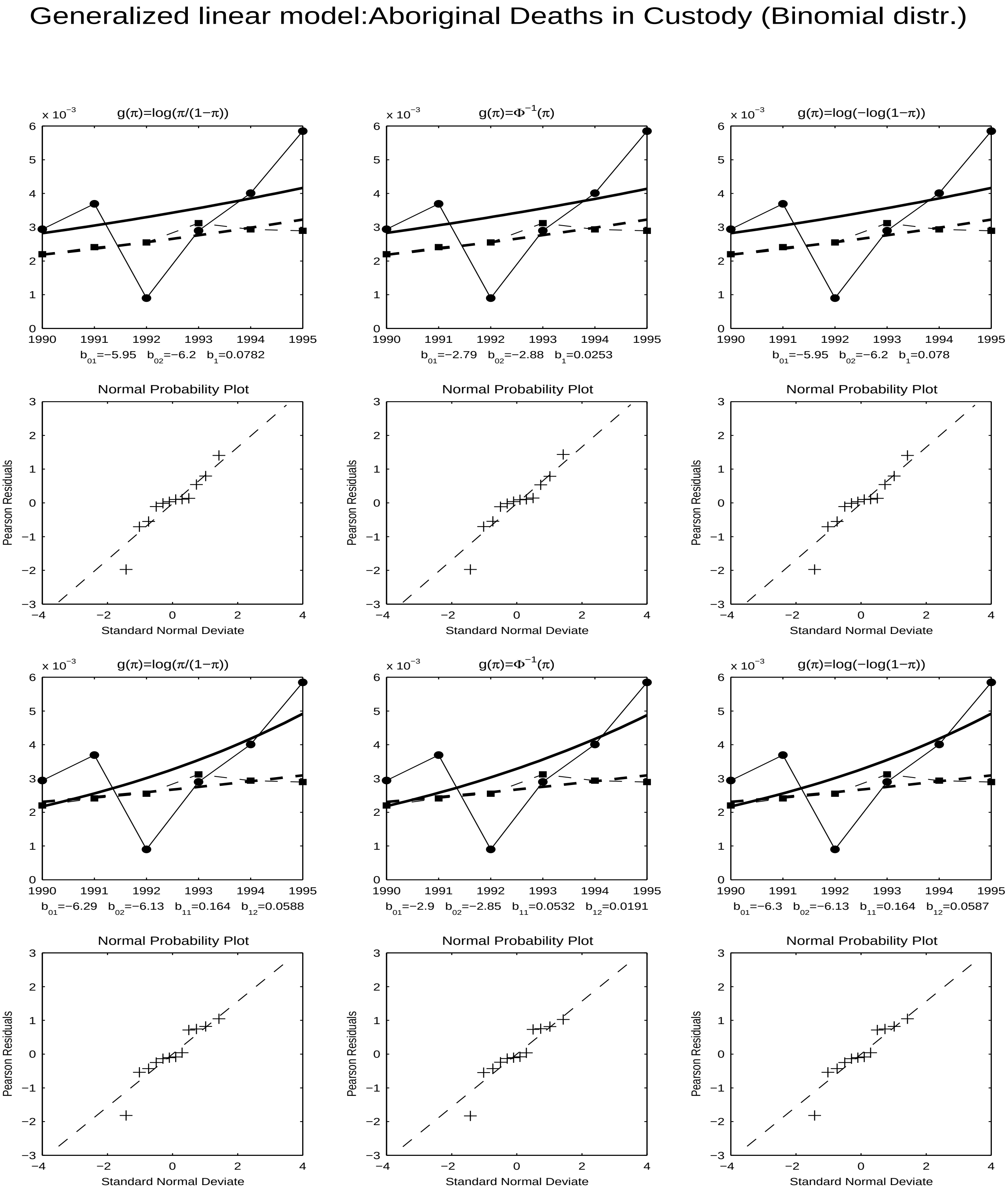
Obrázek 2.12: Beetle mortality data (Annette J. Dobson).



Obrázek 2.13: Coronary heart disease data (Hosmer and Lemeshow, 2000).



Obrázek 2.14: Skin Cancer in Texas and Minnesota.



Obrázek 2.15: Aboriginal Deaths in Custody.

2.13 MODELY PRO ORDINÁLNÍ DATA

Mějme tedy závisle proměnnou $\boxed{Y}$, která je ordinální s $\boxed{k}$ kategoriemi. Můžeme předpokládat, že existuje **latentní** (skrytá) **spojitá proměnná** $\boxed{Z}$ a že manifestní (pozorovatelná) ordinální proměnná Y vzniká diskretizací proměnné Z do k uspořádaných kategorií (skupin). Hodnoty na proměnné Z, které definují kategorie, jsou odhadovány pomocí prahů $\theta_1, \ldots, \theta_{k-1}$.

V tomto případě nepracujeme s pravděpodobnostmi

$$\pi_j = P(Y = j) \qquad (j = 1, \ldots, k) \qquad \sum_{j=1}^{k} \pi_j = 1,$$

ale s kumulovanými pravděpodobnostmi

$$\gamma_j = P(Y \leq j) \qquad (j = 1, \ldots, k-1) \qquad \gamma_k = 1;.$$

Paralelní modely. Jestliže prahová hodnota θ_j $(j = 1, \ldots, k-1)$ závisí pouze na tom, o jakou kategorii závisle proměnné Y jde, nikoliv však na hodnotách nezávisle proměnných, tj. regresorů $\mathbf{x} = (x_1, \ldots, x_m)^{\mathrm{T}}$, a regresní koeficienty $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^{\mathrm{T}}$ jsou stejné pro všechny kategorie, pak dostaneme nejjednodušší, **paralelní model**.

V tomto příkladě předpokládejme, že latentní proměnnou Z lze pomocí regresorů $\mathbf{x} = (x_1, \ldots, x_m)^{\mathrm{T}}$ vyjádřit ve tvaru $\boxed{Z = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} + \varepsilon}$, kde ε má jedno z následujících rozdělení s distribuční funkcí

	$F(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \Phi(x)$	normální rozdělení
rozdělení s distribuční funkcí	$F(x) = (1 + e^{-u})^{-1} = \frac{e^u}{1+e^u}$	logistické rozdělení
	$F(x) = 1 - \exp[-\exp(u)]$	Log-Weibullovo rozdělení
	$F(x) = \exp[-\exp(-u)]$	Gumbelovo rozdělení

které jsou ve standardizovanm tvaru vždy s nulovou střední hodnotou.

Položíme-li prahové hodnoty $\theta_0 = -\infty$ a $\theta_k = \infty$, pak pomocí prahových veličin $\theta_0 < \theta_1 < \cdots < \theta_{k-1} < \theta_k$ můžeme definovat

$$\boxed{Y = j, \quad \text{pokud} \quad \theta_{j-1} < Z \leq \theta_j} \qquad j = 1, \ldots, k.$$

Za těchto okolností hned ukážeme, že se lépe pracuje s kumulovanými pravděpodobnostmi,

$$\gamma_j(\mathbf{x}) = P(Y \leq j) = \sum_{i=1}^{j} P(Y = 1) = \sum_{i=1}^{j} \pi_i(\mathbf{x}) \qquad (j = 1, \ldots, k-1) \qquad \gamma_k = 1,$$

neboť

$$\boxed{\gamma_j(\mathbf{x})} = P(Y \leq j) = P(Z \leq \theta_j) = P(Z – \beta^{\mathrm{T}}\mathbf{x} \leq \theta_j – \beta^{\mathrm{T}}\mathbf{x}) = P(\varepsilon \leq \theta_j – \beta^{\mathrm{T}}\mathbf{x}) = \boxed{F(\theta_j - \beta^{\mathrm{T}}\mathbf{x})}$$

kdežto

$$\boxed{\pi_j(\mathbf{x})} = P(\theta_{j-1} < Z \leq \theta_j) = P(\theta_{j-1} – \beta^{\mathrm{T}}\mathbf{x} < Z \leq \theta_j – \beta^{\mathrm{T}}\mathbf{x}) = \boxed{F(\theta_j - \beta^{\mathrm{T}}\mathbf{x}) - F(\theta_{j-1} - \beta^{\mathrm{T}}\mathbf{x})}.$$

Za této situace lze kumulované pravděpodobnosti $\boxed{\gamma_j(\mathbf{x})}$ obecně vyjádřit takto

$$\gamma_j(\mathbf{x}) = g^{-1}(\eta_j(\mathbf{x})) = h(\eta_j(\mathbf{x})), \qquad \text{kde} \qquad \boxed{\eta_j(\mathbf{x}) = \theta_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}} \qquad j = 1, \ldots, k-1.$$

Za g volíme některou z linkovacích funkcí, takže dostáváme

$$\eta_j(\mathbf{x}) = g_1(\gamma_j(\mathbf{x})) = \Phi^{-1}(\gamma_j(\mathbf{x})) \qquad (j=1,\dots,k-1) \text{ probitový model} \tag{2.9}$$

$$\eta_j(\mathbf{x}) = g_2(\gamma_j(\mathbf{x})) = \log\left(\frac{\gamma_j(\mathbf{x})}{1-\gamma_j(\mathbf{x})}\right) \qquad (j=1,\dots,k-1) \text{ logistický model} \tag{2.10}$$

$$\eta_j(\mathbf{x}) = g_3(\gamma_j(\mathbf{x})) = \log[-\log(1-\gamma_j(\mathbf{x}))] \qquad (j=1,\dots,k-1) \text{ komplementární log-log model} \tag{2.11}$$

$$\eta_j(\mathbf{x}) = g_4(\gamma_j(\mathbf{x})) = -\log[-\log(\gamma_j(\mathbf{x}))] \qquad (j=1,\dots,k-1) \text{ log-log model} \tag{2.12}$$

Neznámé parametry $\theta_1 \le \theta_2 \le \cdots \le \theta_{k-1}$ se také nazývají **dělícími body** (*cutoff points, thresholds*).

Nejčastěji se používá druhý model s **logistickou linkovací funkcí**. Pokud si situace žádá rozlišení, používá se v případě binární proměnné místo názvu *logistická regrese* termín *binární logistická regrese* a v případě ordinální proměnné *ordinální logistická regrese*.

Ordinální logistický model se také nazývá **model s proporcionální šancí** (*proportional-odds model*), protože podíl šancí

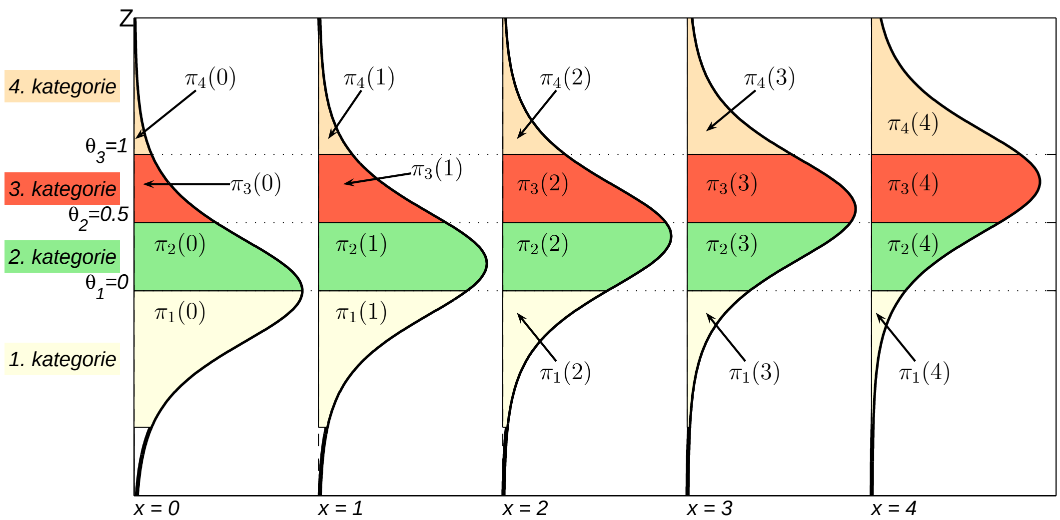
$$\frac{\gamma_j(\mathbf{x_1})/(1-\gamma_j(\mathbf{x_1}))}{\gamma_j(\mathbf{x_2})/(1-\gamma_j(\mathbf{x_2}))} = \exp\{-\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x_1}-\mathbf{x_2})\} \qquad \text{pro} \qquad j=1,\dots,k-1$$

nezávisí na výběru kategorie $\boxed{j}$.

Na následujícím obrázku se budeme snažit ilustrovat, jak se s měnícími hodnotami regresorů $\mathbf{x} = (x_1,\dots,x_m)^{\mathrm{T}}$ mění hodnoty $\pi_1(\mathbf{x}),\dots,\pi_k(\mathbf{x})$ a tím i hodnoty

$$\begin{aligned}
\gamma_1(\mathbf{x}) &= \pi_1(\mathbf{x}) \\
\gamma_2(\mathbf{x}) &= \pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) \\
&\vdots \\
\gamma_{k-1}(\mathbf{x}) &= \pi_1(\mathbf{x}) + \cdots + \pi_{k-1}(\mathbf{x}) \\
\gamma_k(\mathbf{x}) &= \pi_1(\mathbf{x}) + \cdots + \pi_k(\mathbf{x}) = 1
\end{aligned}$$

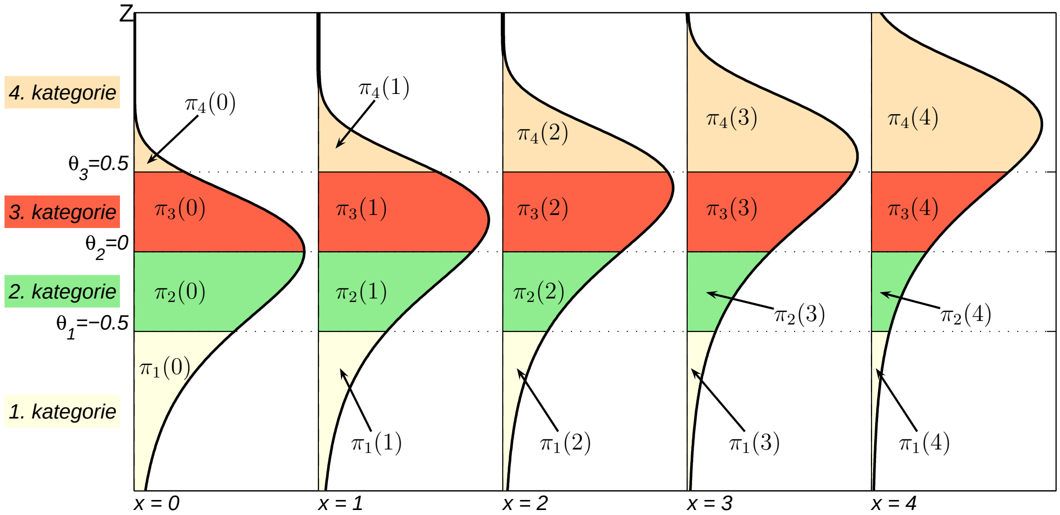
ORDINÁLNÍ LOGISTICKÝ MODEL



Hustoty latentní proměnné Z při různých hodnotách x v případě logistického rozdělení.

Druhým nejpoužívanějším modelem je model s **komplementární log-log linkovací funkcí**, který je známý také pod názvem *proportional hazard model*

ORDINÁLNÍ KOMPLEMENTÁRNÍ LOG-LOG MODEL



Hustoty latentní proměnné Z při různých hodnotách x v případě Log-Weibullova rozdělení.

Neparalelní modely. Pokud prahové hodnoty θ_j $(j = 1, \dots, k-1)$ budou záviset i na hodnotách regresorů $\mathbf{x} = (x_1, \dots, x_m)^{\mathrm{T}}$, pak dostaneme neparalelní model, který je tvaru

$$\gamma_j(\mathbf{x}) = g^{-1}(\eta_j(\mathbf{x})) = h(\eta_j(\mathbf{x})), \qquad \text{kde} \qquad \boxed{\eta_j(\mathbf{x}) = \theta_j - \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x}} \qquad j = 1, \dots, k-1.$$

2.14 MODELY PRO NOMINÁLNÍ DATA

Mějme závisle proměnnou $\boxed{Y}$, která je nominální s $\boxed{k}$ kategoriemi. Na rozdíl od ordinálního modelu budeme pracovat s pravděpodobnostmi

$$\pi_j = P(Y = j) \qquad (j = 1, \dots, k) \qquad \textstyle\sum_{j=1}^{k} \pi_j = 1.$$

Jednu kategorii je nutno zvolit za referenční, např. poslední, která je rovna $\boxed{\pi_k = 1 - \sum_{j=1}^{k-1} \pi_j}$. Protože jde o regresní modely, bude nás zajímat vztah těchto pravděpodobností k hodnotám regresorů (kovariát) $\mathbf{x} = (x_1, \dots, x_m)^{\mathrm{T}}$:

$$\pi_j(\mathbf{x}) = g^{-1}(\eta_j(\mathbf{x})) = h(\eta_j(\mathbf{x})), \qquad \text{kde} \qquad \boxed{\eta_j(\mathbf{x}) = \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x}} \qquad j = 1, \dots, k-1.$$

Za g volíme některou z linkovacích funkcí, takže dostáváme

$$\eta_j(\mathbf{x}) = g_1(\pi_j(\mathbf{x})) = \Phi^{-1}(\pi_j(\mathbf{x})) \qquad (j=1,\dots,k-1) \text{ probitový model} \tag{2.13}$$

$$\eta_j(\mathbf{x}) = g_2(\pi_j(\mathbf{x})) = \log\left(\frac{\pi_j(\mathbf{x})}{1-\pi_j(\mathbf{x})}\right) \qquad (j=1,\dots,k-1) \text{ logistický model} \tag{2.14}$$

$$\eta_j(\mathbf{x}) = g_3(\pi_j(\mathbf{x})) = \log[-\log(1-\pi_j(\mathbf{x}))] \qquad (j=1,\dots,k-1) \text{ komplementární log-log model} \tag{2.15}$$

$$\eta_j(\mathbf{x}) = g_4(\pi_j(\mathbf{x})) = -\log[-\log(\pi_j(\mathbf{x}))] \qquad (j=1,\dots,k-1) \text{ log-log model} \tag{2.16}$$

Opět nejčastěji se používá logistický model a tato regrese se nazývá **multinomickou logistickou regresí**.

2.15 POISSONOVSKÁ A MULTINOMICKÁ DATA

2.15.1 POISSONOVSKÁ DATA

Celočíselná data lze modelovat pomocí diskrétních rozdělení. V předchozí sekci jsme se zabývali alternativními a binomickými daty. Nyní soustřeďme pozornost na poissonovská data.

Předpokládejme, že náhodný výběr rozsahu n je z Poissonova rozdělení, tj

$$\boxed{Y_i \sim Po(\lambda_i) \sim f_Y(y) = P(Y_i = y) = \begin{cases} \frac{\lambda^y e^{-\lambda_i}}{y!} = & \lambda_i > 0;\ \ y = 0, 1, 2, \dots \\ 0 & \text{jinak} \end{cases}}$$

přičemž

$$EY_i = DY_i = \lambda_i.$$

Poznámka 2.15.1.
Poznamenejme, že rozdělením $\boxed{Po(\lambda)}$ lze dobře **aproximovat binomické rozdělení** $\boxed{Bi(n, \pi)}$ za podmínek, že

$$\boxed{n \to \infty \quad \wedge \quad \pi \to 0 \quad \wedge \quad n\pi = \lambda < \infty},$$

obvykle se doporučuje $n > 30$ a $\pi < 0.1$.

Poznámka 2.15.2.
Dále se tímto rozdělením řídí náhodná veličina, kterou je **počet výskytu sledovaného jevu v určitém časovém intervalu délky** t (nebo počet výskytu sledovaného jevu na ploše velikosti t apod.)

Jestliže jsou splněny následující podmínky

a) jev může nastat v kterémkoliv časovém okamžiku,

b) počet výskytů jevu během časového intervalu závisí jen na jeho délce a ne na jeho počátku ani na tom, kolikrát jev nastoupil před jeho počátkem,

c) pravděpodobnost, že jev nastoupí více než jednou v intervalu délky t, konverguje k nule rychleji než t,

d) $\boxed{\lambda}$ je **střední hodnota počtu výskytů jevu za časovou jednotku**

pak uvedená náhodná veličina má rozdělení $\boxed{Po(t\lambda)}$.

Náhodnou veličinou, která má Poissonovo rozdělení, je tedy např.

- počet vadných výrobků ve velké sérii, jestliže pravděpodobnost vyrobení vadného výrobku je velmi malá
- počet těžkých dopravních úrazů za den v určitém městě
- počet zákazníků v prodejně během nějakého časového intervalu
- počet částic v jednotce plochy nebo objemu, např. počet částic v zorném poli mikroskopu
- počet telefonních volání v časovém intervalu t
- počet létavic pozorovaných během intervalu délky t

Předpokládejme opět, že náhodná veličina Y_i závisí na m veličinách $x_{i1}, \ldots, x_{im}$, (tzv. **kovariáty**) a úkolem bude najít vztah mezi nimi, tj. hledáme funkci

$$\boxed{\lambda_i = \lambda(\mathbf{x}_i) = \lambda(x_{i1}, \ldots, x_{im})}.$$

Protože chceme použít GLM modely, modelujeme pravděpodobnosti λ_i pomocí linkovacích funkcí

$$g(\lambda_i) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}.$$

Nejjednodušším modelem je **lineární model** $\boxed{\lambda_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}}$.

Avšak tento model má řadu nevýhod, především je třeba zajistit, aby $\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ nabývala pouze kladných hodnot, nejčastěji se proto volí tyto dvě možnosti

$$EY_i = \mu_i = \boxed{\lambda_i = \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})} \qquad \Rightarrow \qquad g_1(\mu_i) = \eta_i = \boxed{g_1(\lambda_i) = \log(\lambda_i)} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$$

a mluvíme o tzv. **log-lineárních modelech**, nebo

$$EY_i = \mu_i = \boxed{\lambda_i = (\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})^2} \qquad \Rightarrow \qquad g_2(\mu_i) = \eta_i = \boxed{g_2(\lambda_i) = \sqrt{\lambda_i}} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$$

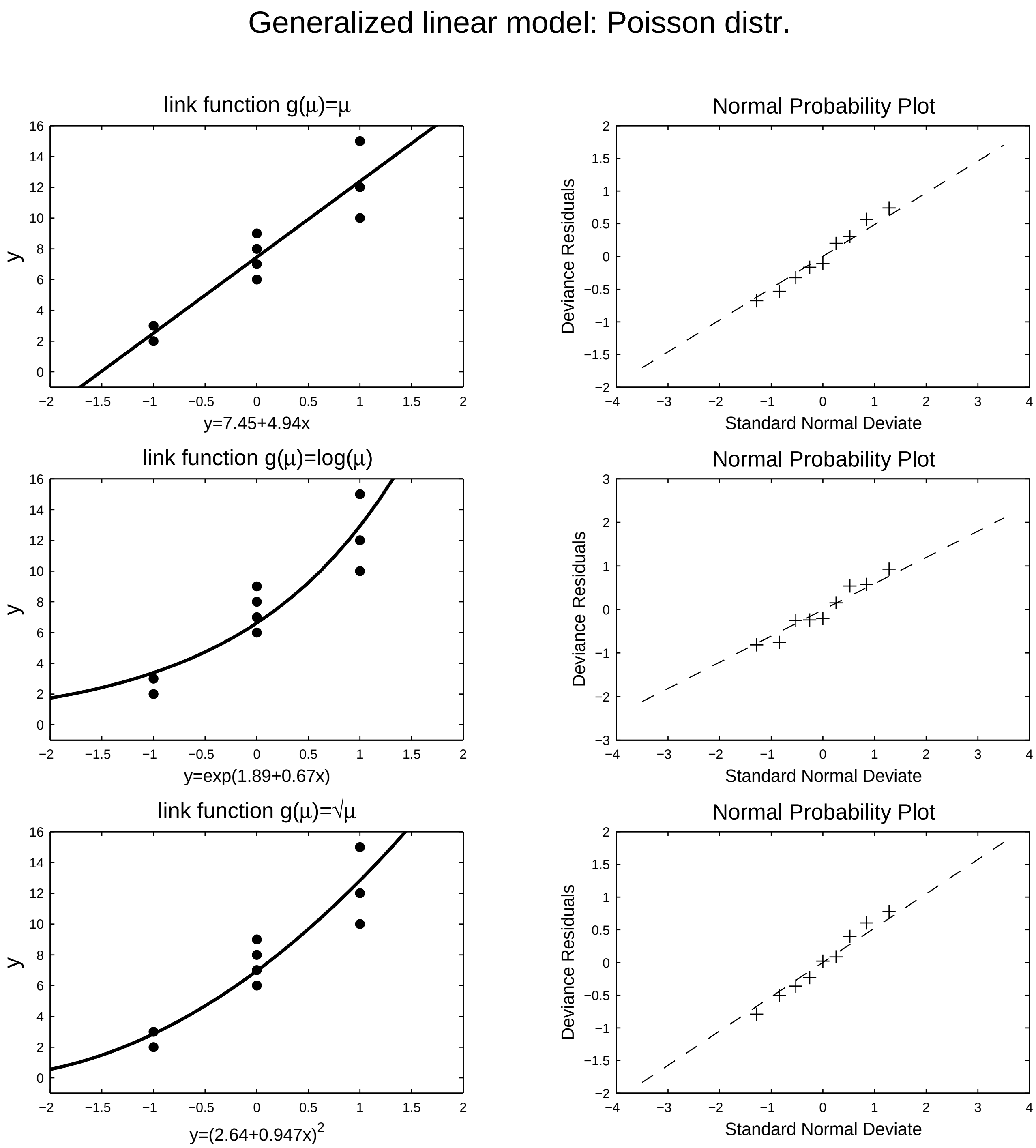
tzv. **odmocninový model** (*square-root-linear model*).

Poznámka 2.15.3.
Chceme-li aproximovat binomické rozdělení $\boxed{Bi(n_i, \pi_i)}$ (kde n_i jsou dostatečně velká a π_i dostatečně malá) pomocí rozdělení Poissonova $\boxed{Y_i \sim Po(\lambda_i = n_i\pi_i)}$ a přitom použijeme logaritmickou linkovací funkci, máme

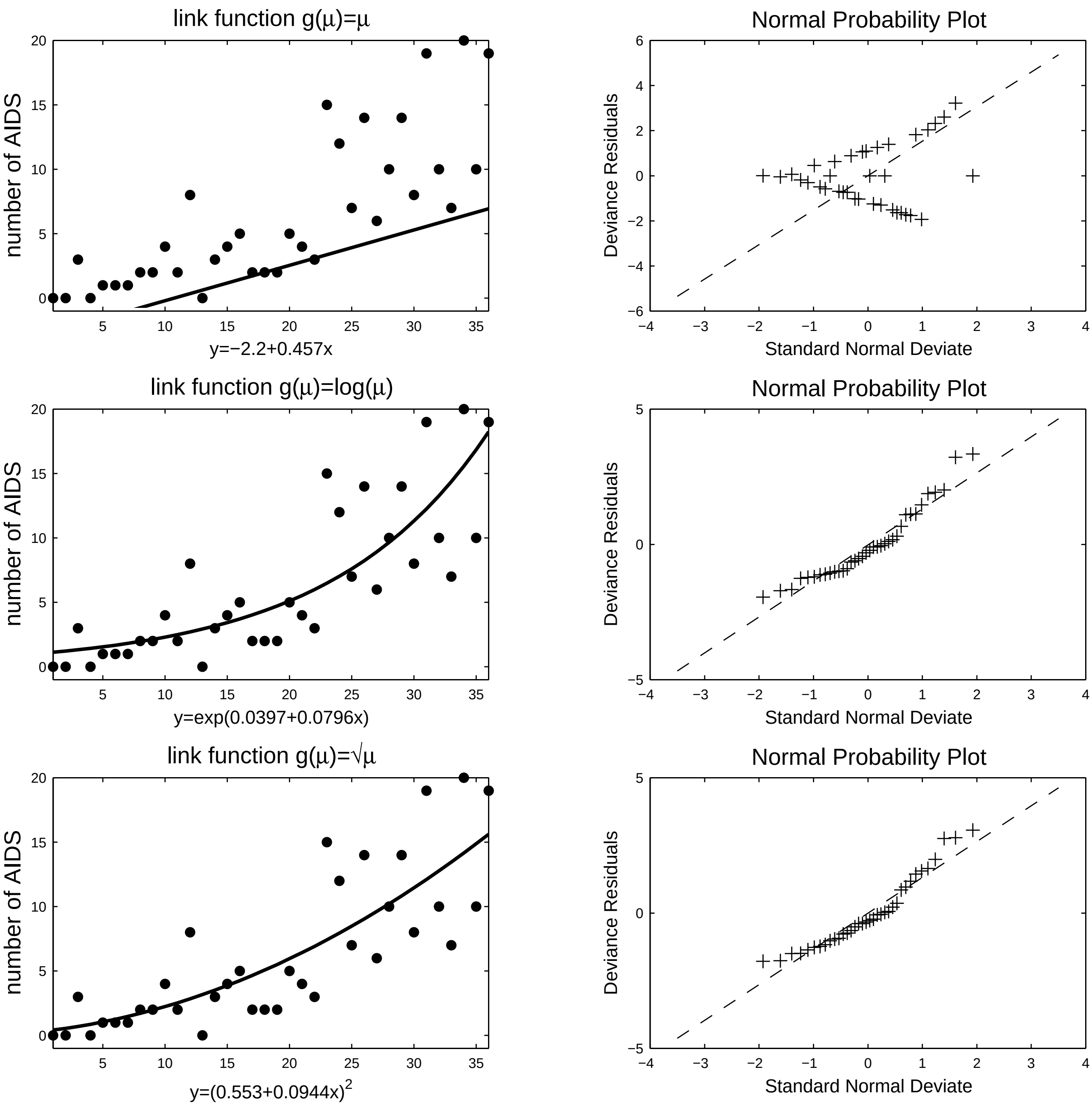
$$\lambda_i = n_i\pi_i = \exp(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \quad \Rightarrow \quad \log(\lambda_i) = \underbrace{\log(n_i)}_{\text{tzv. offset}} + \log(\pi_i) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}.$$

2.15.2 PŘÍKLADY

Uveďme nyní 2 aplikace pro poissonovská data.


Generalized linear model: Poisson distr.
link function g(μ)=μ
y
y=7.45+4.94x
Normal Probability Plot
Deviance Residuals
Standard Normal Deviate
link function g(μ)=log(μ)
y=exp(1.89+0.67x)
link function g(μ)=√μ
y=(2.64+0.947x)^2

AIDS in UK (Dec. 1982 − Nov. 1985): Poisson distr.



2.15.3 KONTINGENČNÍ TABULKY

Mějme náhodný výběr $\mathbf{Y} = \mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ rozsahu n, pro který $n = J \cdot K$, kde $J, K \in \mathbb{N}^+$ jsou kladná přirozená čísla, tj. náhodný výběr lze rozepsat takto

$$\mathbf{Y} = \mathbf{Y}_n = (Y_1, \ldots, Y_n)^{\mathrm{T}} = (Y_{11}, \ldots, Y_{1K}, \ldots, Y_{J1}, \ldots, Y_{JK})^{\mathrm{T}}.$$

Předpokládejme, že náhodný výběr $\mathbf{Y}$ je z Poissonova rozdělení, tj.

$$Y_{jk} \sim Po(\lambda_{jk}) \qquad j = 1, \ldots, J;\ k = 1, \ldots, K$$

s následujícími podmínkami

MODEL (1) s tzv. **celkovou** dodatečnou podmínkou

$$\boxed{N = \sum_{j=1}^{J} \sum_{k=1}^{K} y_{jk}} \qquad N \in \mathbb{N}^+,$$

kde y_{jk} jsou realizace náhodných veličin Y_{jk}.

Pak sdružená (nepodmíněná) pravděpodobnostní funkce náhodného vektoru $\mathbf{Y}$ je rovna

$$\begin{aligned} p_{\mathbf{Y}}(\mathbf{y}) &= P(Y_{11} = y_{11}, \ldots, Y_{1K} = y_{1K}, \ldots, Y_{J1} = y_{J1}, \ldots, Y_{JK} = y_{JK}) \\ &= \begin{cases} \prod_{j=1}^{J} \prod_{k=1}^{K} \frac{\lambda_{jk}^{y_{jk}} e^{-\lambda_{jk}}}{y_{jk}!} & y_{jk} = 0, 1, 2, \ldots;\ j = 1, \ldots, J;\ k = 1, \ldots, K, \\ 0 & \text{jinak.} \end{cases} \end{aligned}$$

Součet nezávislých náhodných veličin s Poissonovým rozdělením má opět Poissonovo rozdělení, tj.

$$Z_{..} = \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{jk} \sim Po\left(\lambda_{..} = \sum_{j=1}^{J} \sum_{k=1}^{K} \lambda_{jk}\right) \sim p_Z(z) = \begin{cases} \frac{\lambda_{..}^{z} e^{-\lambda_{..}}}{z!} & z = 0, 1, 2, \ldots, \\ 0 & \text{jinak.} \end{cases}.$$

Nás ovšem zajímá rozdělení náhodného vektoru $\mathbf{Y}$ za podmínky $Z_{..} = N$, (tj. že součet jeho složek je roven pevně danému kladnému přirozenému číslu N) s pravděpodobnostní funkcí $p_{\mathbf{Y}|Z_{..}=N}$, kterou lze snadno vypočítat ze vztahu

$$p_{\mathbf{Y}|Z_{..}=N}(\mathbf{y}) = \begin{cases} \frac{p_{\mathbf{Y}}(\mathbf{y})}{p_Z(N)} & p_{Z_{..}}(N) \neq 0, \\ 0 & jinak, \end{cases}$$

kterou s využitím vztahů

$$e^{-\lambda_{..}} = \prod_{j=1}^{J} \prod_{k=1}^{K} e^{-\lambda_{jk}}$$

$$\lambda_{..}^{N} = \lambda_{..}^{\sum_{j=1}^{J} \sum_{k=1}^{K} y_{jk}} = \prod_{j=1}^{J} \prod_{k=1}^{K} \lambda_{..}^{y_{jk}}$$

lze upravit takto

$$p_{\mathbf{Y}|Z_{..}=N}(\mathbf{y}) = \begin{cases} \frac{\prod_{j=1}^{J}\prod_{k=1}^{K}\frac{\lambda_{jk}^{y_{jk}}e^{-\lambda_{jk}}}{y_{jk}!}}{\frac{\lambda_{..}^{N}e^{-\lambda_{..}}}{N!}} = & \text{pro} \quad y_{jk}=0,1,\dots,N; \quad j=1,\dots,J; \\ & \qquad\qquad\qquad\qquad\qquad k=1,\dots,K, \\ = N!\prod_{j=1}^{J}\prod_{k=1}^{K}\frac{\left(\frac{\lambda_{jk}}{\lambda_{..}}\right)^{y_{jk}}}{y_{jk}!} & \qquad \sum_{j=1}^{J}\sum_{k=1}^{K} y_{jk} = N \\ 0 & \text{jinak.} \end{cases}$$

a položíme-li

$$\frac{\lambda_{jk}}{\lambda_{..}} = \pi_{jk} \qquad \text{pak} \qquad \sum_{j=1}^{J}\sum_{k=1}^{K}\pi_{jk} = 1.$$

Vidíme, že rozdělení náhodného vektoru $\mathbf{Y}$ za podmínky $Z_{..} = N$ je **multinomické** s pravděpodobnostní funkcí

$$\boxed{p_{\mathbf{Y}|Z_{..}=N}(\mathbf{y}) = \begin{cases} N!\prod_{j=1}^{J}\prod_{k=1}^{K}\frac{\pi_{jk}^{y_{jk}}}{y_{jk}!} & \textit{pro} \quad y_{jk}=0,1,\dots,N; \quad j=1,\dots,J; \\ & \qquad\qquad\qquad\qquad k=1,\dots,K, \\ & \quad \sum_{j=1}^{J}\sum_{k=1}^{K} y_{jk} = N \qquad \sum_{j=1}^{J}\sum_{k=1}^{K}\pi_{jk} = 1 \\ 0 & \text{jinak} \end{cases}},$$

tj.

$$\boxed{\mathbf{Y}|Z_{..} = N \quad \sim \quad Mn(N, \pi_{11}, \dots, \pi_{1K}, \dots, \pi_{J1}, \dots, \pi_{JK})},$$

přičemž

$$\begin{aligned} EY_{jk} =& N\pi_{jk} \\ DY_{jk} =& N\pi_{jk}(1-\pi_{jk}) \\ C(Y_{jk}, Y_{j'k'}) =& -N\pi_{jk}\pi_{j'k'} \end{aligned}$$

<u>Poznámka</u> 2.15.4.
Multinomické rozdělení popisuje situaci, kdy máme $n = J \times K$ neslučitelných jevů, které označme

$$(AB)_{11}, \dots, (AB)_{JK}.$$

Jednotlivé jevy mohou nastat v každém z N nezávislých pokusů s pravděpodobnostmi

$$\pi_{11}, \dots, \pi_{JK} \qquad \text{přičemž} \qquad \sum_{j=1}^{J}\sum_{k=1}^{K}\pi_{jk} = 1.$$

Multinomické rozdělení je zobecněním binomického rozdělení a je patrně nejdůležitějším diskrétním mnohorozměrným rozdělením. Svým významem by se dalo přirovnat k mnohorozměrnému normálnímu rozdělení, jemuž se podobá především díky dvěma vlastnostem: podmíněná i marginální rozdělení jsou opět multinomická. Realizace náhodných veličin i teoretické pravděpodobnosti lze uspořádat do tzv. **kontingenční tabulky**:

Kontingenční tabulka četností

faktor A	faktor B: B_1	B_2	$\cdots$	B_K	$\sum$
A_1	y_{11}	y_{12}	$\cdots$	y_{1K}	$N_{1.}$
A_2	y_{21}	y_{22}	$\cdots$	y_{2K}	$N_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_J	y_{J1}	y_{J2}	$\cdots$	y_{JK}	$N_{J.}$
$\sum$	$N_{.1}$	$N_{.2}$	$\cdots$	$N_{.K}$	$N = N_{..}$

Kontingenční tabulka pravděpodobností

faktor A	faktor B: B_1	B_2	$\cdots$	B_K	$\sum$
A_1	π_{11}	π_{12}	$\cdots$	π_{1K}	$\pi_{1.}$
A_2	π_{21}	π_{22}	$\cdots$	π_{2K}	$\pi_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_J	π_{J1}	π_{J2}	$\cdots$	π_{JK}	$\pi_{J.}$
$\sum$	$\pi_{.1}$	$\pi_{.2}$	$\cdots$	$\pi_{.K}$	$\pi_{..} = 1$

Čísla $\boxed{N_{j.}}$ a $\boxed{N_{.k}}$ se nazývají **marginální četnosti** a $\boxed{\pi_{j.}}$ a $\boxed{\pi_{.k}}$ jsou **marginální pravděpodobnosti**. Tabulky popisujeme slovně tak, že říkáme, že N jednotek bylo klasifikováno podle znaku A do J tříd a podle znaku B do K tříd. V praxi kontingenční tabulka vzniká tak, že na daných objektech sledujeme dva znaky (faktory). Vybereme-li náhodně N objektů, můžeme výsledky shrnout do kontingenční tabulky typu $J \times K$.

Nejčastěji se v kontingenčích tabulkách testuje hypotéza, že

faktory A a B jsou nezávislé

tj.

faktor A	faktor B: $\cdots$	B_k	$\cdots$	$\sum$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_j	$\cdots$	$\pi_{j.}\pi_{.k}$	$\cdots$	$\pi_{j.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sum$	$\cdots$	$\pi_{.k}$	$\cdots$	1

$$\boxed{\pi_{jk} = \pi_{j.}\pi_{.k}}, \text{ takže potom } \boxed{EY_{jk} = N\pi_{j.}\pi_{.k}}, \text{ přičemž } \sum_{j=1}^{J} \pi_{j.} = \sum_{k=1}^{K} \pi_{.k} = 1.$$

MODEL (2) Někdy se stává, že některé marginální četnosti (buď sloupcové nebo řádkové) jsou předem pevně stanoveny. Řekněme, že jsou fixovány řádkové marginální četnosti.

Uvažujme proto v tomto případě pro každý **řádek** $j = 1, \ldots, J$ dodatečnou podmínku

$$\boxed{N_{j.} = \sum_{k=1}^{K} y_{jk}} \qquad N_{j.} \in \mathbb{N}^+$$

kde y_{jk} jsou realizace náhodných veličin Y_{jk}. Označme novou náhodnou veličinu, která je součtem K náhodných poissonovských veličin takto

$$Z_{j.} = \sum_{k=1}^{K} Y_{jk} \sim Po\left(\lambda_{j.} = \sum_{k=1}^{K} \lambda_{jk}\right).$$

Pak každý řádek

$$\mathbf{Y}_{j.} = (Y_{j1}, \ldots, Y_{jK})^{\mathrm{T}}$$

za podmínky

$$Z_{j.} = N_{j.}$$

má **multinomické rozdělení** s pravděpodobnostní funkcí

$$p_{\mathbf{Y}_{j.}|Z_{j.}=N_{j.}}(\mathbf{y}) = \begin{cases} N_{j.}! \prod\limits_{k=1}^{K} \frac{\pi_{jk}^{y_{jk}}}{y_{jk}!} & \text{pro} \quad y_{jk} = 0, 1, \ldots, N_{j.}; \qquad k = 1, \ldots, K; \\ & \quad \sum\limits_{k=1}^{K} y_{jk} = N_{j.} \qquad j = 1, \ldots, J \\ & \quad \sum\limits_{k=1}^{K} \pi_{jk} = \sum\limits_{k=1}^{K} \frac{\lambda_{jk}}{\lambda_{j.}} = 1 \quad j = 1, \ldots, J \\ 0 & \text{jinak} \end{cases}$$

Označme náhodný vektor

$$\mathbf{Z}_r = (Z_{1.}, \ldots, Z_{J.})^{\mathrm{T}}.$$

Pak, protože jednotlivé řádky $\mathbf{Y}_{j.}$ jsou nezávislé, náhodný vektor $\mathbf{Y}$ za podmínky $\mathbf{Z}_r = (N_{1.}, \ldots, N_{J.})^{\mathrm{T}}$ má diskrétní rozdělení s pravděpodobnostní funkcí

$$p_{\mathbf{Y}|\mathbf{Z}_r=(N_{1.},\ldots,N_{J.})^{\mathrm{T}}}(\mathbf{y}) = \begin{cases} \prod\limits_{j=1}^{J} N_{j.}! \prod\limits_{k=1}^{K} \frac{\pi_{jk}^{y_{jk}}}{y_{jk}!} & \text{pro} \quad y_{jk} = 0, 1, \ldots, N_{j.} \quad k = 1, \ldots, K \\ & \quad \sum\limits_{k=1}^{K} y_{jk} = N_{j.} \qquad j = 1, \ldots, J \\ & \quad \pi_{j.} = \sum\limits_{k=1}^{K} \pi_{jk} = 1 \quad j = 1, \ldots, J \\ 0 & \text{jinak} \end{cases}$$

a toto rozdělení se nazývá **součinové multinomické rozdělení** (product multinomial distribution), což budeme značit

$$\boxed{\mathbf{Y}|\mathbf{Z}_r = (N_{1.}, \ldots, N_{J.})^{\mathrm{T}} \sim PMn(N_{1.}, \ldots, N_{J.}, \pi_{11}, \ldots, \pi_{1K}, \ldots, \pi_{J1}, \ldots, \pi_{JK})},$$

přičemž

$$\begin{aligned} EY_{jk} &= N_{j.}\pi_{jk} \\ DY_{jk} &= N_{j.}\pi_{jk}(1 - \pi_{jk}) \\ C(Y_{jk}, Y_{j'k'}) &= -N_{j.}\pi_{jk}\pi_{j'k'}. \end{aligned}$$

Pak pozorované četnosti jevů lze uspořádat do následujících kontingenčních tabulek

Kontingenční tabulka četností
(fixace řádků)

faktor	faktor B			
A	B_1	$\cdots$	B_K	$\sum$
A_1	y_{11}	$\cdots$	y_{1K}	$N_{1.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_J	y_{J1}	$\cdots$	y_{JK}	$N_{J.}$
$\sum$	$N_{.1}$	$\cdots$	$N_{.K}$	$N = N_{..}$

Kontingenční tabulka pravděp.
(fixace řádků)

faktor	faktor B			řádk.
A	B_1	$\cdots$	B_K	$\sum$
A_1	π_{11}	$\cdots$	π_{1K}	$\pi_{1.} = 1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_J	π_{J1}	$\cdots$	π_{JK}	$\pi_{J.} = 1$
marg.	$\pi_{.1}$	$\cdots$	$\pi_{.K}$	$\pi_{..} = 1$

a kontingenční tabulka obsahuje v každém řádku jiné multinomické rozdělení.

Poznamenejme, že $\pi_{.k}$ značí marginální rozdělení k-tého sloupce a $\pi_{.k}$ není součtem pravděpodobností $\pi_{1k}, \ldots, \pi_{Jk}$.

V modelu, kdy fixujeme **řádkové součty**, obvykle chceme testovat, zda máme

$$\boxed{\text{shodná řádková multinomická rozdělení}},$$

což lze formulovat i tak, že pravděpodobnost

$$\boxed{\pi_{jk} \text{ je stejná v celém sloupci } k.}$$

Tato hypotéza se nazývá **hypotéza homogenity** a lze ji formulovat takto

Fixace řádků

faktor A	faktor B $\cdots$	B_k	$\cdots$	$\sum$
$\vdots$	$\vdots$	$\pi_{.k}$	$\vdots$	1
A_j	$\vdots$	$\pi_{.k}$	$\vdots$	1
$\vdots$	$\vdots$	$\pi_{.k}$	$\vdots$	1
marg.	$\cdots$	$\pi_{.k}$	$\cdots$	1

$$\pi_{jk} = \pi_{.k}, \qquad \text{takže potom} \qquad \boxed{EY_{jk} = N_{j.}\pi_{.k}}, \qquad \text{přičemž} \qquad \sum_{k=1}^{K} \pi_{.k} = 1.$$

Analogický model týkající se fixace sloupců dostaneme, jestliže pro každý **sloupec** $k = 1, \ldots, K$ uvažujeme dodatečnou podmínku

$$N_{.k} = \sum_{j=1}^{J} y_{jk} \qquad N_{.k} \in \mathbb{N}^+$$

kde y_{jk} jsou realizace náhodných veličin Y_{jk}.

Vícerozměrné kontingenční tabulky V mnoha konkrétních případech nevystačíme pouze s dvourozměrnými kontingenčními tabulkami, proto uvažujme vícerozměrné kontingenční tabulky, konkrétně např. **třírozměrnou kontingenční tabulku** typu $J \times K \times L$, tj.

- s J *řádky,*
- s K *sloupci*
- s L *podtabulkami,*

Pokud **fixujeme totální součet celé tabulky**, tj. položíme-li

$$\boxed{N = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{L} y_{jkl}} \qquad N \in \mathbb{N}^+,$$

kde y_{jkl} jsou pozorování četností náhodných jevů Y_{jkl}, které mají odpovídající multinomické rozdělen, pak třírozměrné kontingenční tabulky absolutních a relativních četností lze psát takto:

Kontingenční tabulka 1 (četnosti)

Faktory		faktor B				
C	A	B_1	B_2	$\cdots$	B_K	$\sum$
C_1	A_1	y_{111}	y_{121}	$\cdots$	y_{1K1}	$N_{1.1}$
	A_2	y_{211}	y_{221}	$\cdots$	y_{2K1}	$N_{2.1}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	y_{J11}	y_{J21}	$\cdots$	y_{JK1}	$N_{J.1}$
	$\sum$	$N_{.11}$	$N_{.21}$	$\cdots$	$N_{.K1}$	$N_{..1}$
C_2	A_1	y_{112}	y_{122}	$\cdots$	y_{1K2}	$N_{1.2}$
	A_2	y_{212}	y_{222}	$\cdots$	y_{2K2}	$N_{2.2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	y_{J12}	y_{J22}	$\cdots$	y_{JK2}	$N_{J.2}$
	$\sum$	$N_{.12}$	$N_{.22}$	$\cdots$	$N_{.K2}$	$N_{..2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_L	A_1	y_{11L}	y_{12L}	$\cdots$	y_{1KL}	$N_{1.L}$
	A_2	y_{21L}	y_{22L}	$\cdots$	y_{2KL}	$N_{2.L}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	y_{J1L}	y_{J2L}	$\cdots$	y_{JKL}	$N_{J.L}$
	$\sum$	$N_{.1L}$	$N_{.2L}$	$\cdots$	$N_{.KL}$	$N_{..L}$
$\sum$	$\sum$	$N_{.1.}$	$N_{.2.}$	$\cdots$	$N_{.K.}$	N

Kontingenční tabulka 2 (pravděpodobnosti)

Faktory		faktor B				
C	A	B_1	B_2	$\cdots$	B_K	$\sum$
C_1	A_1	π_{111}	π_{121}	$\cdots$	π_{1K1}	$\pi_{1.1}$
	A_2	π_{211}	π_{221}	$\cdots$	π_{2K1}	$\pi_{2.1}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	π_{J11}	π_{J21}	$\cdots$	π_{JK1}	$\pi_{J.1}$
	$\sum$	$\pi_{.11}$	$\pi_{.21}$	$\cdots$	$\pi_{.K1}$	$\pi_{..1}$
C_2	A_1	π_{112}	π_{122}	$\cdots$	π_{1K2}	$\pi_{1.2}$
	A_2	π_{212}	π_{222}	$\cdots$	π_{2K2}	$\pi_{2.2}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	π_{J12}	π_{J22}	$\cdots$	π_{JK2}	$\pi_{J.2}$
	$\sum$	$\pi_{.12}$	$\pi_{.22}$	$\cdots$	$\pi_{.K2}$	$\pi_{..2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_L	A_1	π_{11L}	π_{12L}	$\cdots$	π_{1KL}	$\pi_{1.L}$
	A_2	π_{21L}	π_{22L}	$\cdots$	π_{2KL}	$\pi_{2.L}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	π_{J1L}	π_{J2L}	$\cdots$	π_{JKL}	$\pi_{J.L}$
	$\sum$	$\pi_{.1L}$	$\pi_{.2L}$	$\cdots$	$\pi_{.KL}$	$\pi_{..L}$
$\sum$	$\sum$	$\pi_{.1.}$	$\pi_{.2.}$	$\cdots$	$\pi_{.K.}$	$\pi_{...}$

Podrobně si však všimneme dvou konkrétních modelů:

MODEL (3) v každé podtabulce fixujeme **součty řádků**, tj. pro l-tou podtabulku a j-tý řádek položme

$$\boxed{N_{j.l} = \sum_{k=1}^{K} y_{jkl}} \qquad N_{j.l} \in \mathbb{N}^+ \quad j = 1, \ldots, J; \;\; l = 1, \ldots, L,$$

kde y_{jkl} jsou realizace náhodných veličin Y_{jkl}. Pak pozorované četnosti jevů a jejich teoretické pravděpodobnosti lze uspořádat do následující kontingenční tabulky, kde l-tá podtabulka vypadá takto

Kontingenční tabulka četností (fixace řádků)

Faktory		faktor B				řádkové
C	A	B_1	B_2	$\cdots$	B_K	$\sum$
C_l	A_1	y_{11l}	y_{12l}	$\cdots$	y_{1Kl}	$N_{1.l}$
	A_2	y_{21l}	y_{22l}	$\cdots$	y_{2Kl}	$N_{2.l}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	y_{J1l}	y_{J2l}	$\cdots$	y_{JKl}	$N_{J.l}$
	$\sum$	$N_{.1l}$	$N_{.2l}$	$\cdots$	$N_{.Kl}$	$N_{..l}$

Kontingenční tabulka pravděpodobností (fixace řádků)

Faktory		faktor B				řádkové
C	A	B_1	B_2	$\cdots$	B_K	$\sum$
C_l	A_1	π_{11l}	π_{12l}	$\cdots$	π_{1Kl}	$\pi_{1.l} = 1$
	A_2	π_{21l}	π_{22l}	$\cdots$	π_{2Kl}	$\pi_{2.l} = 1$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	π_{J1l}	π_{J2l}	$\cdots$	π_{JKl}	$\pi_{J.l} = 1$
	marg.	$\pi_{.1l}$	$\pi_{.2l}$	$\cdots$	$\pi_{.Kl}$	$\pi_{..l}$

Celá kontingenční tabulka obsahuje $K\cdot L$ multinomických rozdělení se sdruženou pravděpodobnostní funkcí

$$p_{\mathbf{Y}|\mathbf{Z}_r=(N_{1.1},\ldots,N_{1.L},\ldots,N_{J.1},\ldots,N_{J.L})^{\mathrm{T}}}(\mathbf{y}) =$$

$$= \begin{cases} \prod_{j=1}^{J}\prod_{l=1}^{L} N_{j.l}! \prod_{k=1}^{K} \frac{\pi_{jkl}^{y_{jkl}}}{y_{jkl}!} & \text{pro} \quad \begin{array}{ll} y_{jkl} = 0,1,\ldots,N_{j.l} & k=1,\ldots,K \\ \sum_{k=1}^{K} y_{jkl} = N_{j.l} & j=1,\ldots,J \\ & l=1,\ldots,L \\ \pi_{j.l} = \sum_{k=1}^{K} \pi_{jkl} = 1 & j=1,\ldots,J \\ & l=1,\ldots,L \end{array} \\ 0 & \text{jinak} \end{cases}$$

V modelu, kdy v každé podtabulce fixujeme **řádkové součty**, obvykle chceme testovat, zda máme

$\boxed{\text{shodná řádková multinomická rozdělení}}$,

což lze formulovat i tak, že pravděpodobnost

$\boxed{\pi_{jkl} \quad \text{v rámci podtabulky } l \text{ je stejná v celém sloupci } k}$

Tato hypotéza se nazývá **hypotéza homogenity** a lze ji formulovat takto

Fixace řádků l-té podtabulku

Faktory		faktor B			
C	A	$\cdots$	B_k	$\cdots$	$\sum$
C_l	A_1	$\vdots$	$\pi_{.kl}$	$\vdots$	1
	$\vdots$	$\vdots$	$\pi_{.kl}$	$\vdots$	1
	A_J	$\vdots$	$\pi_{.kl}$	$\vdots$	1
	$\sum$	$\cdots$	$\pi_{.kl}$	$\cdots$	$\pi_{..l}$

$$\pi_{jkl} = \pi_{.kl}, \quad \text{takže potom} \quad \boxed{EY_{jkl} = N_{j.l}\pi_{jkl} = N_{j.l}\pi_{.kl}}, \quad \text{přičemž} \quad \sum_{j=1}^{L}\sum_{k=1}^{K} \pi_{.kl} = 1.$$

MODEL (4) v každé podtabulce fixujeme **celkový součet podtabulky**, tj. pro l-tou podtabulku položme

$$\boxed{N_{..l} = \sum_{j=1}^{J}\sum_{k=1}^{K} y_{jkl}} \qquad N_{j.l} \in \mathbb{N}^+ \qquad l = 1,\ldots,L,$$

kde y_{jkl} jsou realizace náhodných veličin Y_{jkl}. Pak pozorované četnosti jevů lze uspořádat do následující kontingenční tabulky, kde l-tá podtabulka vypadá takto

Kontingenční tabulka četností
(fixace celkového součtu podtabulky)

Faktory		faktor B				
C	A	B_1	B_2	$\cdots$	B_K	$\sum$
C_l	A_1	y_{11l}	y_{12l}	$\cdots$	y_{1Kl}	$N_{1.l}$
	A_2	y_{21l}	y_{22l}	$\cdots$	y_{2Kl}	$N_{2.l}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	y_{J1l}	y_{J2l}	$\cdots$	y_{JKl}	$N_{J.l}$
	$\sum$	$N_{.1l}$	$N_{.2l}$	$\cdots$	$N_{.Kl}$	$N_{..l}$

Kontingenční tabulka pravděpodobností
(fixace celkového součtu podtabulky)

Faktory		faktor B				
C	A	B_1	B_2	$\cdots$	B_K	$\sum$
C_l	A_1	π_{11l}	π_{12l}	$\cdots$	π_{1Kl}	$\pi_{1.l}$
	A_2	π_{21l}	π_{22l}	$\cdots$	π_{2Kl}	$\pi_{2.l}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	π_{J1l}	π_{J2l}	$\cdots$	π_{JKl}	$\pi_{J.l}$
	$\sum$	$\pi_{.1l}$	$\pi_{.2l}$	$\cdots$	$\pi_{.Kl}$	$\pi_{..l} = 1$

Každá kontingenční tabulka má své multinomické rozdělení. Sdružená pravděpodobnostní funkce vektoru $\mathbf{Y}$ za podmínky $\mathbf{Z}_t = (N_{..1}, \dots, N_{..L})^{\mathrm{T}}$ je tvaru

$$p_{\mathbf{Y}|\mathbf{Z}_t=(N_{..1},\dots,N_{..L})^{\mathrm{T}}}(\mathbf{y}) = $$
$$= \begin{cases} \prod_{l=1}^{L} N_{..l}! \prod_{j=1}^{J} \prod_{k=1}^{K} \frac{\pi_{jkl}^{y_{jkl}}}{y_{jkl}!} & \text{pro} \quad y_{jkl} = 0, 1, \dots, N_{..l} \quad k = 1, \dots, K \\ & \qquad\qquad\qquad\qquad\qquad\qquad j = 1, \dots, J \\ & \qquad \sum_{j=1}^{J} \sum_{k=1}^{K} y_{jkl} = N_{..l} \quad l = 1, \dots, L \\ & \qquad \pi_{..l} = \sum_{j=1}^{J} \sum_{k=1}^{K} \pi_{jkl} = 1 \quad l = 1, \dots, L \\ 0 & \text{jinak.} \end{cases}$$

Nyní **hypotéza homogenity** zní

všechny podtabulky mají stejné multinomické rozdělení.

Tuto hypotézu lze formulovat také tak, že pravděpodobnost

π_{jkl} je stejná pro všechny podtabulky l

Fixace total-součtu l-té podtabulky

Faktory		faktor B			
C	A	B_1	$\cdots$	B_K	$\sum$
C_l	A_1	π_{11l}	$\cdots$	π_{1Kl}	$\pi_{1.l}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_J	π_{J1l}	$\cdots$	π_{JKl}	$\pi_{J.l}$
	$\sum$	$\pi_{.1l}$	$\cdots$	$\pi_{.Kl}$	$\pi_{..l}$=1

což lze psát takto

$$\pi_{jkl} = \pi_{jk.} \quad \text{takže potom} \quad \boxed{EY_{jkl} = N_{..l}\pi_{jkl} = N_{..l}\pi_{jk.}}, \quad \text{přičemž} \quad \sum_{j=1}^{J} \sum_{k=1}^{K} \pi_{jk.} = 1.$$

2.15.4 LOG-LINEÁRNÍ MODELY

Ve všech předchozích případech hypotézy nezávislosti a homogenity vedou k **multiplikativnímu modelu**, který logaritmováním lze převést na model **lineární** a odtud pramení všeobecně zažité pojmenování *LOG-LINEÁRNÍ MODELY*.

Nyní pro modely (1) až (5) hledejme odpovídající GLM modely:

(1) Pro model s **celkovou dodatečnou podmínkou** lze hypotézu o **nezávislosti dvou faktorů** definovat takto

$$\boxed{EY_{jk} = N\pi_{j.}\pi_{.k}}, \qquad \text{přičemž} \quad \sum_{j=1}^{J} \pi_{j.} = 1 \quad \text{a} \quad \sum_{k=1}^{K} \pi_{.k} = 1.$$

V GLM s **log-lineární linkovací funkcí** máme $\boxed{\eta_{jk} = \log EY_{jk} = \mathbf{x}_{jk}^{\mathrm{T}}\boldsymbol{\beta}}$, tedy

$$\eta_{jk} = \log EY_{jk} = \log(N\pi_{j.}\pi_{.k}) = \underbrace{\mu}_{=\log N} + \underbrace{\alpha_j}_{=\log \pi_{j.}} + \underbrace{\beta_k}_{=\log \pi_{.k}} .$$

Pokud bychom nepředpokládali nezávislost faktorů A a B, dostaneme **maximální model**

$$\eta_{jk} = \log EY_{jk} = \log(N\pi_{jk}) = \underbrace{\mu}_{=\log N} + \underbrace{\alpha_j + \beta_k + (\alpha\beta)_{jk}}_{=\log \pi_{jk}}$$

Vidíme, že základní i maximální modely jsou přeparametrizovány. Proto se musí upravit, například tak, že položíme

$$\alpha_1 = \beta_1 = 0 \quad \text{resp. navíc} \quad (\alpha\beta)_{11} = 0 \qquad\qquad \text{tzv. metoda horního rohu}$$

nebo

$$\sum_{j=1}^{J} \alpha_j = 1 \quad \text{a} \quad \sum_{k=1}^{K} \beta_k = 0 \quad \text{resp. navíc} \quad \sum_{j=1}^{J}\sum_{k=1}^{K} (\alpha\beta)_{jk} = 0.$$

Všimněme si **počtů parametrů** pro jednotlivé úrovně

μ	obecná střední hodnota	1
α	hlavní efekt	$J-1$
β	hlavní efekt	$K-1$
$\alpha\beta$	interakce prvního řádu	$(J-1)(K-1)$
	celkem	$n = JK$

Vidíme, že hypotéza nezávislosti dvou faktorů v kontingenčních tabulkách je ekvivalentní s hypotézou neexistence interakcí v analýze rozptylu (deviace), tj.

$$\boxed{H_0: \quad (\alpha\beta)_{jk} = 0} \qquad j = 1, \ldots, J; \quad k = 1, \ldots, K.$$

V log-lineárních modelech jsou obvykle výrazy vyšších řádů definovány jako odchylky od výrazů nižšího řádu. Tak například v základním modelu výraz α_j reprezentuje rozdíl efektu řádku j od obecné střední hodnoty μ. Takže model je **hiearchický** v tom smyslu, že výrazy vyšších řádů nejsou obsaženy ve výrazech nižších řádů.

Shrňme předchozí výsledky

$$\boxed{GLM_{max}}\ H_0: EY_{jk} = N\pi_{jk} \Rightarrow \eta_{jk} = \log EY_{jk} = \underbrace{\mu}_{\log N} + \underbrace{\alpha_j + \beta_k + (\alpha\beta)_{jk}}_{\log \pi_{jk}}$$

$$\boxed{GLM}\quad H_0: EY_{jk} = N_{j.}\pi_{.k} \Rightarrow \eta_{jk} = \log EY_{jk} = \underbrace{\mu}_{\log N} + \underbrace{\alpha_j + \beta_k}_{\log \pi_{.k}}$$

(2) V modelu, kdy fixujeme **řádkové součty**:

$$\boxed{GLM_{max}}\ H_0: EY_{jk} = N_{j.}\pi_{jk} \Rightarrow \eta_{jk} = \log EY_{jk} = \underbrace{\mu + \alpha_j}_{\log N_{j.}} + \underbrace{\beta_k + (\alpha\beta)_{jk}}_{\log \pi_{jk}}$$

$$\boxed{GLM}\quad H_0: EY_{jk} = N_{j.}\pi_{.k} \Rightarrow \eta_{jk} = \log EY_{jk} = \underbrace{\mu + \alpha_j}_{\log N_{j.}} + \underbrace{\beta_k}_{\log \pi_{.k}}$$

(3) Pro třírozměrnou kontingenční tabulku, kdy pro každou podtabulku fixujeme **součty řádků**

$$\boxed{GLM_{max}}\quad H_0: EY_{jkl} = N_{j.l}\pi_{jkl} \Rightarrow \eta_{jkl} = \log EY_{jkl} = \underbrace{\mu + \alpha_j + \gamma_l + (\alpha\gamma)_{jl}}_{\log N_{j.l}} + \underbrace{\beta_k + (\alpha\beta)_{jk} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl}}_{\log \pi_{jkl}}$$

$$\boxed{GLM}\quad H_0: EY_{jkl} = N_{j.l}\pi_{.kl} \Rightarrow \eta_{jkl} = \log EY_{jkl} = \underbrace{\mu + \alpha_j + \gamma_l + (\alpha\gamma)_{jl}}_{\log N_{j.l}} + \underbrace{\beta_k + (\beta\gamma)_{kl}}_{\log \pi_{.kl}}$$

(4) Pro třírozměrnou kontingenční tabulku, kdy pro každou podtabulku fixujeme **celkový součet podtabulky**

$$\boxed{GLM_{max}}\quad H_0: EY_{jkl} = N_{..l}\pi_{jkl} \Rightarrow \eta_{jkl} = \log EY_{jkl} = \underbrace{\mu + \gamma_l}_{\log N_{..l}} + \underbrace{\alpha_j + \beta_k + (\alpha\beta)_{jk} + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl}}_{\log \pi_{jkl}}$$

$$\boxed{GLM}\quad H_0: EY_{jkl} = N_{..l}\pi_{jk.} \Rightarrow \eta_{jkl} = \log EY_{jkl} = \underbrace{\mu + \gamma_l}_{\log N_{..l}} + \underbrace{\alpha_j + \beta_k + (\alpha\beta)_{jk}}_{\log \pi_{jk.}}$$