II.) Metoda maximální věrohodnosti

$X_1,\dots,X_n$ je náh. výběr s hustotou $f(x,\theta)$, tj. X_i má hustotu (pravd. funkci) $f(x,\theta)$, kde $\theta \in \mathbb{R}^p$ je neznámý parametr

sdružená hustota náh. vektoru $(X_1,\dots,X_n)^T$ je $L(\theta)=L(\theta,x_1,\dots,x_n)=\prod_{i=1}^{n} f(x_i,\theta)$ ... funkce θ (věrohodnostní funkce)

$\hat{\theta}=\hat{\theta}(X_1,\dots,X_n)$ je odhad parametru θ metodou max. věrohodnosti, jestliže $L(\hat{\theta})\geq L(\theta)$, $\forall\theta$, tj. $\hat{\theta}=\underset{\theta}{\operatorname{argmax}}\, L(\theta)$.

$\ell(\theta)=\log L(\theta)=\sum_{i=1}^{n}\log f(x_i,\theta)$ ... logaritmická věrohodnostní funkce $\qquad \hat{\theta}=\underset{\theta}{\operatorname{argmax}}\, \ell(\theta)$.

Příklad (motivace MLE)

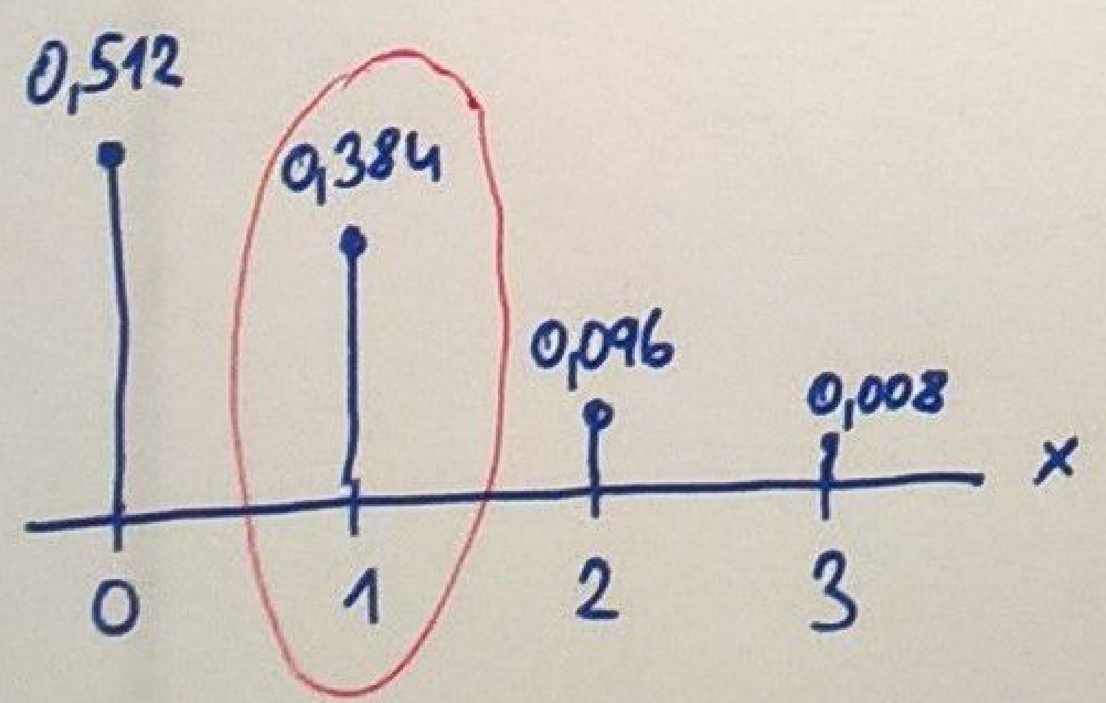
Hodíme 3× falešnou mincí. Víme, že pravd., že v daném hodu padne líc je buďto $p_1=0{,}2$ nebo $p_2=0{,}6$. Z těchto 3 hodů padl 1 líc.

Otázka: Je pravděpodobnost padnutí líce p_1 nebo p_2?

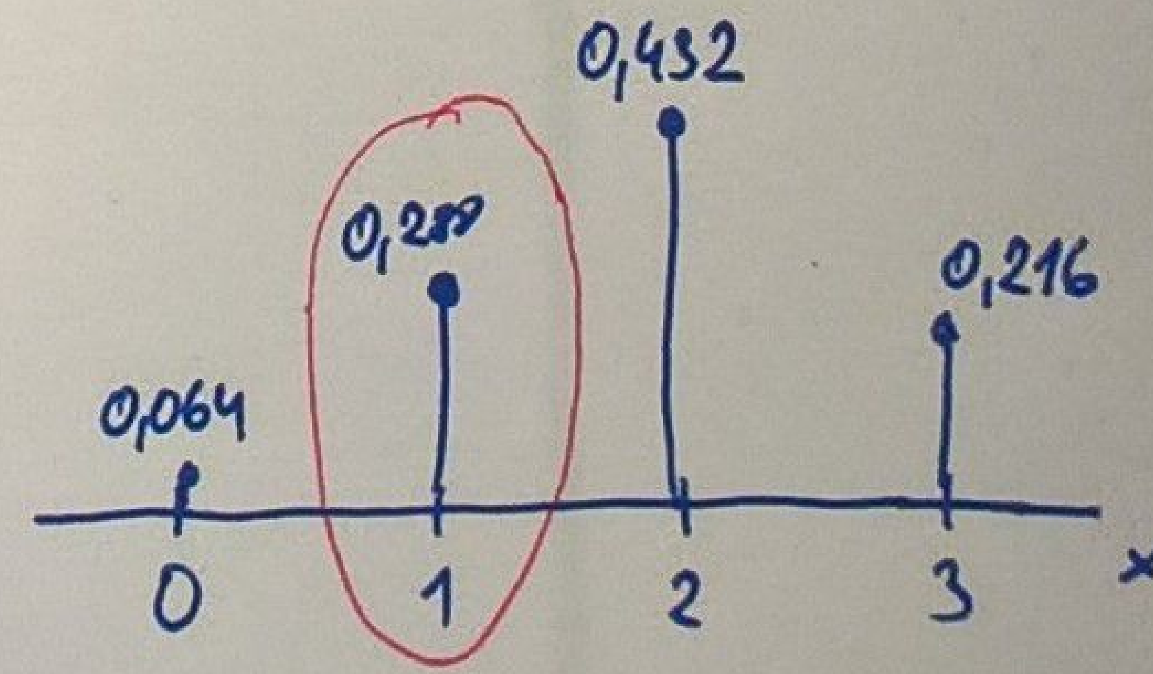
model: X = počet padnutých líců, $X\sim Bi(3,p)$, kde p = pravděpodobnost, že v 1 hodu padne líc $\quad \underline{p=p_1}$ nebo $\underline{p=p_2}$.

$$p(x)=P(X=x)=\binom{3}{x}p^x(1-p)^{3-x},\quad x=0,1,2,3$$

$P(x)$ pro $p_1=0{,}2$



$p(x)$ pro $p_2=0{,}6$



„MLE vezme pro danou realizaci náh. výběru nejpravděpodobnější hodnotu parametru", tedy MLE parametru p je $\hat{p}=p_1=0{,}2$.

Vlastnosti max. věrohodných odhadů

Za mírných předpokladů (podmínky regularity) jsou max. věrohodné odhady asymptoticky nestranné, konzistentní a asymptoticky normální:

$$\sqrt{n}\,(\hat{\theta}-\theta) \approx N_p\left(0, J^{-1}(\theta)\right), \text{ kde } J(\theta) = \left(E\, \frac{\partial \log f(X_i, \theta)}{\partial \theta_i} \cdot \frac{\partial \log f(X_i, \theta)}{\partial \theta_j} \right)_{i,j=1}^{p}$$

je Fisherova informační matice o parametru θ příslušná X_i

$$= \left\{ -E\left(\frac{\partial^2 \log f(X_i, \theta)}{\partial \theta_i\, \partial \theta_j} \right) \right\}_{i,j=1}^{p}.$$

Pozn.

$$\hat{\theta} \approx N_p\left(\theta, \frac{J^{-1}(\theta)}{n}\right).$$

Příklad (počet úmrtí při dopravních nehodách)

sledujeme počty smrtelných zranění při dopravních nehodách v ČR od 29. 6. 2015 do 28. 7. 2015 (celkem 30 dní)

počet úmrtí	0	1	2	3	4	5	6	7	8	9	10
počet dní	4	6	6	8	3	0	2	0	0	0	1

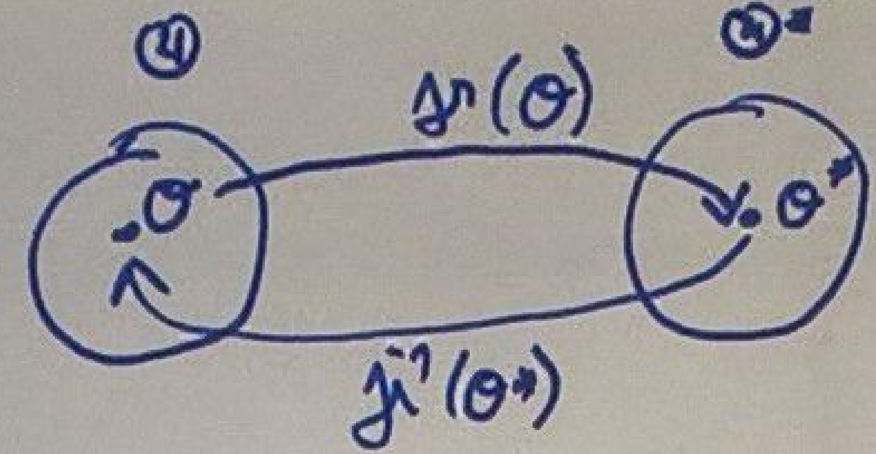
Chceme odhadnout pravděpodobnost, že se dnes nestane žádná smrtelná nehoda $= p$.

neparametrický model: empirický odhad $\hat{p} = \frac{\text{počet dní bez smrtelné nehody}}{\text{počet dní}} = \frac{4}{30} \doteq 0{,}13.$

parametrický model: X_i = počet smrtelných zranění i-tý den ; $X_1, \dots, X_{30}$ náhodný výběr z $Po(\lambda)$ λ = průměrný počet smrtelných zranění za 1 den

nás zajímá parametrická funkce $g(\lambda) = e^{-\lambda} = P(X_i = 0)$.

Maximálně věrohodný odhad parametrické funkce $g: \mathbb{R}^p \mapsto \mathbb{R}$.



Nechť $g: \overset{\subset \mathbb{R}^p}{\Theta} \mapsto \overset{\subset \mathbb{R}}{\Theta^*}$ je prostá funkce

Funkce $\tilde{L}(\theta^*) = \prod_{i=1}^{n} f(x_i, g^{-1}(\theta^*)) = L(g^{-1}(\theta^*))$ pro $\theta^* \in \Theta^*$ se nazývá věrohodnostní funkce indukovaná parametrickou funkcí g.

$\hat{\theta}^*$ je max. věrohodný odhad fce $g(\theta) = \theta^*$, jestliže $\tilde{L}(\hat{\theta}^*) \geq \tilde{L}(\theta^*), \forall \theta^* \in \Theta^*$.

Tvrzení (Zehnova věta, princip invariance MLE)

Je-li $\hat{\theta}$ max. věr. odhadem parametru θ, pak $g(\hat{\theta})$ je max. věr. odhadem parametrické funkce $g(\theta)$.

Věta (delta metoda)

Nechť $\{X_n\}_{n=1}^{\infty}$ je posloupnost p-rozměrných náh. vektorů takových, že $\sqrt{n}(X_n - \theta) \approx N_p(0, \Sigma)$. Dále buď $g: \mathbb{R}^p \to \mathbb{R}$ měřitelné zobrazení, které má totální diferenciál v bodě θ. Pak platí $\sqrt{n}(g(X_n) - g(\theta)) \approx N(0, \sigma^2)$, kde $\sigma^2 = \nabla g(\theta)^T \Sigma \nabla g(\theta)$ a

$\nabla g(\theta) = \left(\frac{\partial g(\theta)}{\partial \theta_1}, \dots, \frac{\partial g(\theta)}{\partial \theta_p}\right)^T$ je gradient funkce g v bodě θ.

Důsledek (pro MLE)

Za mírných předpokladů (podmínky regularity) a nechť $g: \mathbb{R}^p \mapsto \mathbb{R}$ je prostá. Pak $\sqrt{n}(g(\hat{\theta}) - g(\theta)) \approx N(0, \nabla g(\theta)^T J^{-1}(\theta) \nabla g(\theta))$.

Poznámka

$g(\hat{\theta})$ je as. nestranný, konzistentní odhad $g(\theta)$. Navíc $g(\hat{\theta}) \approx N\left(g(\theta), \frac{\nabla g(\theta)^T J^{-1}(\theta) \nabla g(\theta)}{n}\right)$

θ jednorozměrný parametr, tj. $p = 1$

$g(\hat{\theta}) \approx N\left(g(\theta), \frac{[g'(\theta)]^2}{n J(\theta)}\right)$.

Příklad (metody – pokračování)

$X_1,\dots,X_n$ náh. výběr z $Po(\lambda)$, $\lambda>0$ neznámý parametr

$$P(x,\lambda)=f(x,\lambda)=P(X=x)=e^{-\lambda}\cdot\frac{\lambda^x}{x!},\quad x=0,1,\dots$$

$$\log f(x,\lambda)=-\lambda+x\log\lambda-\log x!$$

$$\ell(\lambda)=\sum_{i=1}^n\log f(x_i,\lambda)=-n\lambda+\log\lambda\sum_{i=1}^n x_i-\sum_{i=1}^n\log x_i!$$

$$\ell'(\lambda)=-n+\frac{\sum_{i=1}^n x_i}{\lambda}\overset{!}{=}0$$

$$\lambda=\frac{1}{n}\sum_{i=1}^n x_i\implies\hat\lambda=\bar X \text{ je odhad parametru } \lambda$$

Vlastnosti $\hat\lambda$?

- podle CLV je $\hat\lambda\approx N\left(\lambda,\frac{\lambda}{n}\right)$
- podle teorie MLE

$$J(\lambda)=E\left(\frac{\partial\log f(X_i,\lambda)}{\partial\lambda}\right)^2=-E\,\frac{\partial^2\log f(X_i,\lambda)}{\partial\lambda^2}$$

$$\left.\begin{aligned}\frac{\partial\log f(x_i,\lambda)}{\partial\lambda}&=-1+\frac{x_i}{\lambda}\\ \frac{\partial^2\log f(x_i,\lambda)}{\partial\lambda^2}&=-\frac{x_i}{\lambda^2}\end{aligned}\right\}\ J(\lambda)=E\,\frac{X_i}{\lambda^2}=\frac{EX_i}{\lambda^2}=\frac{\lambda}{\lambda^2}=\frac{1}{\lambda}$$

$\hat\theta\approx N\left(\theta,\frac{1}{nJ(\theta)}\right)$; v našem případě $\hat\lambda\approx N\left(\lambda,\frac{\lambda}{n}\right)$.

Nyní již můžeme přejít k našemu problému, odhad parametrické funkce $g(\lambda)=e^{-\lambda}$. Její MLE je $e^{-\hat\lambda}=e^{-\bar X}$.

Spočítejme hodnotu odhadů pro naše data: $\hat\lambda=\bar x\doteq 2{,}53$ a $e^{-\hat\lambda}=\widehat{P(X_i=0)}\doteq e^{-2{,}53}\doteq 0{,}079$. Jak moc jsme si tím jistí? $\Rightarrow$ interval spolehlivosti

$[D(X_1,\dots,X_n),H(X_1,\dots,X_n)]$ je $100(1-\alpha)\%$ IS pro param. funkci $g(\theta)$, jestliže $P(D(X_1,\dots,X_n)\le g(\theta)\le H(X_1,\dots,X_n))=1-\alpha,\ \forall\theta$.

$g(\lambda)=e^{-\lambda}$

$g'(\lambda)=-e^{-\lambda}$

$[g'(\lambda)]^2=e^{-2\lambda}$

$J(\lambda)=\frac{1}{\lambda}$

$$e^{-\hat\lambda}\approx N\left(e^{-\lambda},\frac{e^{-2\lambda}}{n}\cdot\lambda\right)$$

$$\frac{e^{-\hat\lambda}-e^{-\lambda}}{\sqrt{\frac{\lambda\cdot e^{-2\lambda}}{n}}}\approx N(0,1)$$

– není pivotová statistika

Cramér–Slutsky

$$\frac{e^{-\hat\lambda}-e^{-\lambda}}{\sqrt{\frac{\bar X\cdot e^{-2\bar X}}{n}}}\approx N(0,1)$$

pivotová statistika

$\Rightarrow$ Nyní již můžeme asymptotický IS sestrojit.

$$\left[e^{-\bar{X}} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}\cdot e^{-2\bar{X}}}{n}}\ ,\ e^{-\bar{X}} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}\cdot e^{-2\bar{X}}}{n}}\right]$$ je hledaný $(1-\alpha)\cdot 100\ \%$ interval spolehlivosti pro $e^{-\lambda}$.

v našem případě: $[0{,}034;\ 0{,}125]$.

Na závěr výsledek porovnejme s neparametrickým modelem (bez předpokladu Poissonova rozdělení)

$$Y_i \begin{cases} 1 & i\text{-tý den se ~~stane~~ nestane žádná nehoda} \\ 0 & \text{jinak} \end{cases}$$

$Y_i \sim A(p)$, $p = P$(v daný den se nestane žádná nehoda), $p \in (0,1)$ je neznámý parametr

bodový odhad p je $\hat{p} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{4}{30} \doteq 0{,}13$.

$$\bar{Y} \approx N\left(p, \frac{p(1-p)}{n}\right)$$

$$\frac{\bar{Y} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1) \qquad \frac{\bar{Y} - p}{\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}}} \approx N(0,1) \xRightarrow{\text{hledaný IS je}} \left[\bar{Y} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}}\ ,\ \bar{Y} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{Y}(1-\bar{Y})}{n}}\right]$$

a konečně pro naše data: $[0{,}012;\ 0{,}255]$.