Метод максимального правдоподобия

θ

1) фонтан работает раз в год — $1/365$
2) фонтан работает только эту неделю — $7/365$
3) всегда — 1

x_1 = фонт. раб.

$$\mathbb{P}(x_1 = \text{фонт.} \mid \theta = \text{гипотеза})$$

Упражнение

$$\mathbb{P}(X=k) = \frac{e^{-\lambda}\cdot\lambda^k}{k!}$$

$$X_1,\ldots,X_n \sim \text{iid Poiss}(\lambda)$$

$$L(\lambda) = \mathbb{P}(X_1,\ldots,X_n \mid \lambda) = \mathbb{P}(x_1|\lambda)\cdot\ldots\cdot\mathbb{P}(x_n|\lambda) \to \max_{\lambda}$$

$$L(\lambda) = \frac{e^{-n\lambda}\cdot\lambda^{\sum x_i}}{x_1!\cdot x_2!\cdot\ldots\cdot x_n!} \to \max_{\lambda}$$

$$L(\lambda) \propto e^{-n\lambda}\cdot\lambda^{\sum x_i}$$

$$\ln L(\lambda) \propto -n\lambda + \sum x_i\cdot\ln\lambda$$

$$\frac{\partial \ln L(\lambda)}{\partial\lambda} = -n + \frac{\sum x_i}{\lambda} = 0 \Rightarrow \hat{\lambda}^{ML} = \bar{X}$$

$$\frac{\partial^2 \ln L(\lambda)}{\partial\lambda^2} = -\frac{\sum x_i}{\lambda^2}$$

$$\hat{\lambda}^{ML} \overset{asy}{\sim} \mathcal{N}\left(\lambda; \frac{\hat{\lambda}}{n}\right)$$

$$-\mathbb{E}\left(-\frac{\sum x_i}{\lambda^2}\right) = \frac{1}{\lambda^2}\mathbb{E}\left(\sum x_i\right) = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$

→ CI

→ гипотезы

Свойства:

1) $\hat{\Theta}^{ML} \xrightarrow{P} \Theta$ состоятельная
2) Оценка м.б. смещена, но она асимпт. несмещ.
3) Асимпт. эфф.
4) $\hat{\Theta}^{ML} \overset{asy}{\sim} N(\Theta, \hat{I}^{-1}(\Theta))$ инф. Фишер

верно при $\frac{n}{d} \to \infty$

$$I(\Theta) = -\mathbb{E}\left(\frac{\partial^2 \ln L}{\partial\Theta_i \partial\Theta_j}\right) = -\mathbb{E}(H)$$

5) $\Theta \quad \lambda = g(\Theta)$ инвариантность

$\uparrow$

$\hat{\Theta}^{ML} \quad \hat{\lambda}^{ML} = g(\hat{\Theta}^{ML})$

Это всё работает для распределений, у которых выполнены условия регулярности:

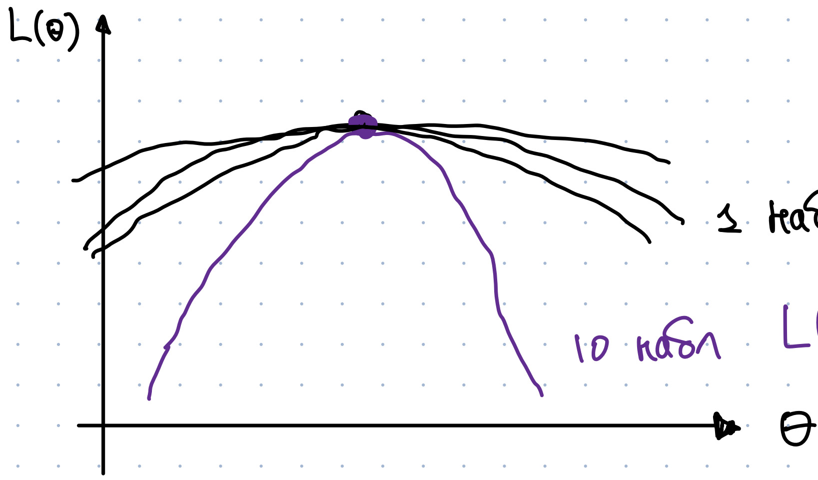
- Нормально берутся производные $f'_x(t)$
- Обл. опр. сл. вел. не зав. от параметра

Почему вторая производная отвечает за дисперсию?

$$\ln L(\Theta | x_1, \dots, x_n) = \sum_{i=1}^{n} \ln f(x_i | \Theta) \to \max_{\Theta}$$

$$\mathbb{P}(x_i | \Theta)$$

$$\frac{\partial \ln L}{\partial \Theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i | \Theta)}{\partial \Theta} = 0$$



$$L(\theta|x_1) = \ln f(x_1|\theta)$$

$$L(\theta|x_1, .., x_{10}) = \sum_{i=1}^{10} \ln f(x_i|\theta)$$

Мат. анализ:
$f''(x)$ отвечает за выпуклость

$$I_{obs.}(\theta) = -\frac{\partial^2 \ln L}{\partial \theta} = -H$$

наблюденная инф. Фишера

Чем выпуклее ↑, тем больше информации

$$I(\theta) = \mathbb{E}(-H)$$

ожидаемая инф. Фишера

$$\hat{I}(\theta)$$

- $\hat{I}(\theta) = -\mathbb{E}(H)\Big|_{\theta=\hat{\theta}}$
- $\hat{I}(\theta) = -H\Big|_{\theta=\hat{\theta}}$
- $\hat{I}(\theta) = \mathbb{E}\left(\left[\frac{\partial \ln L}{\partial \theta}\right] \cdot \left[\frac{\partial \ln L}{\partial \theta}\right]^T\right)\Big|_{\theta=\hat{\theta}}$

Если выполнены усл. регулярности:

$$-\mathbb{E}\left(\frac{\partial^2 L}{\partial \theta^2}\right) = \mathbb{E}\left(\left[\frac{\partial \ln L}{\partial \theta}\right] \cdot \left[\frac{\partial \ln L}{\partial \theta}\right]^T\right)$$

Упражнение

m&s

X	красный	желтый	другие
	p_1	p_2	$1-p_1-p_2$

$n = 100$

X_1 - кр.

X_2 - жёл

$$\boxed{\begin{aligned} &X_1 \sim Bin(p_1, n) \qquad X_1 + X_2 + X_3 = n \\ &X_2 \sim Bin(p_2, n) \\ &X_3 \sim Bin(p_3, n) \end{aligned}}$$

$$L \propto p_1^{x_1} \cdot p_2^{x_2} \cdot (1-p_1-p_2)^{n-x_1-x_2} \to \max_{p_1, p_2}$$

$$\ln L \propto x_1 \ln p_1 + x_2 \ln p_2 + (n-x_1-x_2)\ln(1-p_1-p_2)$$

$$\begin{cases} \dfrac{\partial \ln L}{\partial p_1} = \dfrac{x_1}{p_1} - \dfrac{n-x_1-x_2}{1-p_1-p_2} = 0 \\ \dfrac{\partial \ln L}{\partial p_2} = \dfrac{x_2}{p_2} - \dfrac{n-x_1-x_2}{1-p_1-p_2} = 0 \end{cases} \qquad \begin{cases} \hat{p}_1 = \dfrac{x_1}{n} \\ \hat{p}_2 = \dfrac{x_2}{n} \end{cases}$$

$$H = \begin{pmatrix} -\dfrac{x_1}{p_1^2} - \dfrac{n-x_1-x_2}{(1-p_1-p_2)^2} & -\dfrac{n-x_1-x_2}{(1-p_1-p_2)^2} \\ -\dfrac{n-x_1-x_2}{(1-p_1-p_2)^2} & -\dfrac{x_2}{p_2^2} - \dfrac{n-x_1-x_2}{(1-p_1-p_2)^2} \end{pmatrix} \quad \mathbb{E}(H) = \begin{pmatrix} -\dfrac{n}{p_1} - \dfrac{n}{1-p_1-p_2} & -\dfrac{n}{1-p_1-p_2} \\ -\dfrac{n}{1-p_1-p_2} & -\dfrac{n}{p_2} - \dfrac{n}{1-p_1-p_2} \end{pmatrix}$$

$$\mathbb{E}\left(-\frac{n-X_1-X_2}{(1-p_1-p_2)^2}\right) = -\frac{n-\mathbb{E}(X_1)-\mathbb{E}(X_2)}{(1-p_1-p_2)^2} = -\frac{n-np_1-np_2}{(1-p_1-p_2)^2} =$$

$$= -\frac{n\cancel{(1-p_1-p_2)}}{(1-p_1-p_2)^2}$$

$$\hat{Cov}(\hat{p}_1, \hat{p}_2) < 0$$

$$\hat{I}(p) = -\mathbb{E}(H)\Big|_{p=\hat{p}}$$

$$\hat{Var}(p) = \left[\hat{I}(p)\right]^{-1} = \begin{bmatrix} \hat{Var}(\hat{p}_1) & \hat{Cov}(\hat{p}_1,\hat{p}_2) \\ \hat{Cov}(\hat{p}_1,\hat{p}_2) & \hat{Var}(\hat{p}_2) \end{bmatrix}$$

$$\hat{p}_2 - \hat{p}_1 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\hat{Var}(\hat{p}_1) + \hat{Var}(\hat{p}_2) - 2\,Cov(\hat{p}_1, \hat{p}_2)}$$

можно даже использовать Δ-метод

$$\hat{\theta} \overset{asy}{\sim} N(\theta, \hat{Var}(\hat{\theta}))$$

$$g(\hat{\theta}) \overset{asy}{\sim} N\left(g(\theta);\ \nabla\hat{g}^T \cdot \hat{Var}(\hat{\theta}) \cdot \nabla\hat{g}\right)$$

А что делать, если условия регулярности сломаны?

$$X_1, \dots, X_n \sim iid\ U[0;\theta] \qquad f_X(x) = \begin{cases} \frac{1}{\theta}, & x \in [0;\theta] \\ 0 \end{cases}$$

$$L = \frac{1}{\theta}\cdot\frac{1}{\theta}\cdot\ldots\frac{1}{\theta} = 1/\theta^n \longrightarrow \max_{\theta}$$

$$x_1,\ldots,x_n \le \theta$$

$$\ln L = -n\ln\theta$$

$$(\ln L)'_{\theta} = -\frac{n}{\theta} = 0 \qquad \varnothing$$

$$x_{max} \le \theta$$

$$\theta\uparrow \;\Rightarrow\; L\downarrow$$

$$\hat\theta = x_{max}$$

А при чём тут МО?

x y

выборка

w-параметры

$$y = \langle w, x\rangle$$

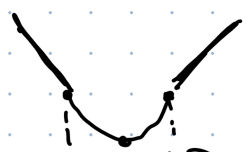
$$\hat y = \langle \hat w, x\rangle$$

$$Q = \frac{1}{n}\sum_{i=1}^{n} L(y_i,\hat y_i) \to \min_{w}$$

→ $L(y_i,\hat y_i) = (y_i-\hat y_i)^2$

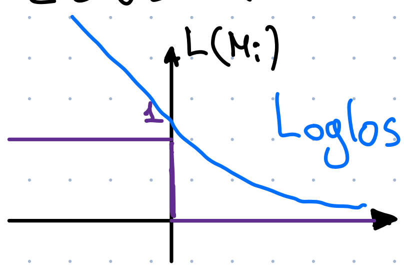
→ $L(y_i,\hat y_i) = |y_i-\hat y_i|$

→ ф. Хубер



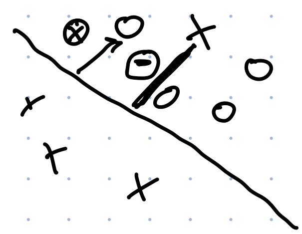
→ LogCosh

Error rate



$$M_i = y_i\cdot\langle w, x_i\rangle$$

$$\{-1, 1\}$$



Инженерный подход

Альтернатива: вероятностный подход

$y, x \qquad y = \langle w, x \rangle + \varepsilon$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\underset{\sim\sim\sim}{y|x} \sim \mathcal{N}(\langle w, x \rangle, \sigma^2)$$

На данные наложено вер. р.-е.

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$L(w|x,y) = \mathbb{P}(y|x,w) = \mathbb{P}(y_1|x_1,w) \cdot \ldots \cdot \mathbb{P}(y_n|x_n,w) =$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \cdot \exp\left(-\frac{\sum (y_i - \langle w, x_i \rangle)^2}{2\sigma^2}\right) \to \max_w$$

$$\ln L = -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \langle w, x_i \rangle)^2 \to \max_w$$

На σ^2 пофигу

$$-\sum_{i=1}^{n} (y_i - \langle w, x_i \rangle)^2 \to \max_w \qquad \left(\times -\frac{1}{n}\right)$$

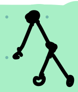
$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \langle w, x_i \rangle)^2 \to \min_w$$

$$-\ln L \propto MSE$$

Упражнение

$$\mathbb{P}(y_i \mid x_i) = \frac{e^{-\lambda(x_i)} \cdot \lambda(x_i)^{y_i}}{y_i!}$$

$\lambda(x_i) = \exp(\langle w, x_i \rangle)$ Пуассоновская регрессия

$\lambda(x_i) = \exp(a(x_i))$ $a(x_i) =$  $+ \gamma_n$ 

$$L(w) \propto e^{-\sum_i \lambda(x_i)} \cdot \lambda(x_1)^{y_1} \cdot \ldots \cdot \lambda(x_n)^{y_n}$$

$$\ln L(w) \propto -\sum_i \lambda(x_i) + \sum_i y_i \ln \lambda(x_i)$$

$$\text{PoissLoss} = \frac{1}{n} \sum_i \lambda(x_i) - y_i \ln \lambda(x_i) \to \min_w$$

$\frac{n}{d} \to \infty$

$<< 1$