Rachunek prawdopodobieństwa i statystyka

Wykład 14

Linie regresji II-go rodzaju

Liniami regresji II-go rodzaju zmiennej y (x) względem x (y) nazywamy zadane krzywe $y = g(x; \theta_1, \theta_2, \ldots)$ oraz $x = h(y; \theta_1, \theta_2, \ldots)$ gdy spełniają one odpowiednio warunki:

$$\mathcal{E}\left[\left(\mathrm{y} - g\left(\mathrm{x};\theta_1,\theta_2,\ldots\right)\right)^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(y - g\left(x;\theta_1,\theta_2,\ldots\right)\right)^2 f\left(x,y\right)dxdy = \min\left(\theta_1,\theta_2,\ldots\right)$$

$$\mathcal{E}\left[\left(\mathrm{x} - h(\left(\mathrm{y};\theta_1,\theta_2,\ldots\right)\right)^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x - h\left(y;\theta_1,\theta_2,\ldots\right)\right)^2 f\left(x,y\right)dxdy = \min\left(\theta_1,\theta_2,\ldots\right)$$

W szczególności proste regresji II-go rodzaju mają postać:

$$y = \rho\frac{\sigma_y}{\sigma_x}x - \rho\frac{\sigma_y}{\sigma_x}\mu_x + \mu_y$$

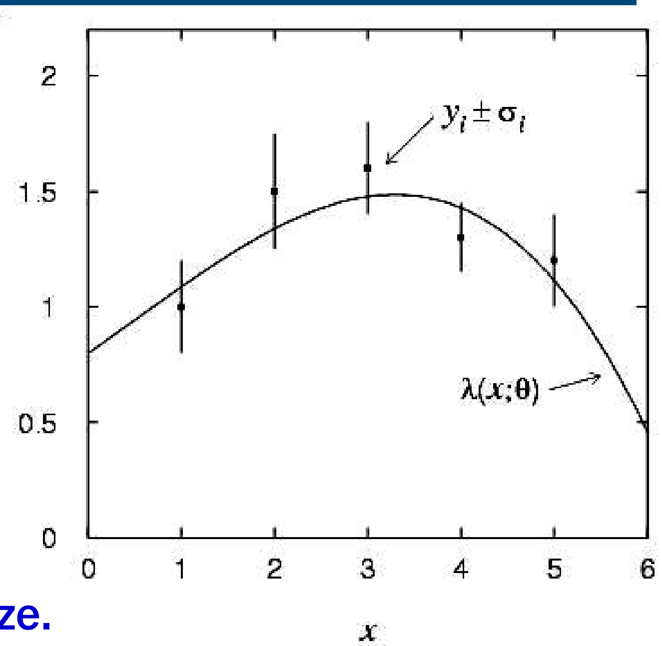
$$x = \rho\frac{\sigma_x}{\sigma_y}y - \rho\frac{\sigma_x}{\sigma_y}\mu_y + \mu_x \qquad \Rightarrow \qquad y = \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}x - \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}\mu_x + \mu_y$$

W praktyce zazwyczaj nie znamy łącznej funkcji rozkładu zmiennych, z której pochodzą zmienne (a więc także obu dyspersji i współczynnika korelacji), a krzywe regresji musimy określać z próbki dyskretnych danych.

Metoda najmniejszych kwadratów

Załóżmy, że mierzymy N niezależnych wartości $y_1,\ldots, y_n$ każda z rozkłądu Gaussa o nieznanej wartości oczekiwanej i znanej wariancji:

$$\mathcal{E}[y_i] = \eta(x_i;\vec{\theta}) \qquad \mathcal{V}[y_i] = \sigma_i^2$$



Wielkości x_i to tzw. zmienne kontrolowane, objaśniające, które zakładamy, że są znane dokładnie.

Chcemy oszacować nieznane parametry θ tak aby dopasowanie krzywej do punktów było możliwie najlepsze.

Konstruujemy łączną funkcję gęstości p-twa

$$g(y_1,\ldots,y_n;\eta_1,\ldots,\eta_n;\sigma_1^2,\ldots,\sigma_n^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i-\eta_i)^2}{2\sigma_i^2}\right)$$

Wówczas log funkcji wiarogodności ma postać (pomijamy wyrazy niezależne od parametrów):

$$\ln\mathcal{L}(\vec{\theta}) = -\frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - \eta(x_i;\vec{\theta})\right)^2}{\sigma_i^2} + \ldots$$

Metoda najmniejszych kwadratów

Zasada najmniejszych kwadratów głosi, że parametry θ_i należy dobrać tak, aby spełniały one warunek:

$$\mathcal{R}(\vec{\theta}) = \sum_{i=1}^{n} \frac{\left(y_i - \eta\left(x_i;\vec{\theta}\right)\right)^2}{\sigma_i^2} = \min(\vec{\theta})$$

W przypadku skorelowanych pomiarów korzystamy z wielowymiarowego rozkładu Gaussa ze znaną macierzą kowariancji $\mathbf{V}$:

$$g(\vec{x};\vec{\mu},\mathbf{V}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |\mathbf{V}|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^{\mathrm{T}} \mathbf{V}^{-1}(\vec{x}-\vec{\mu})\right)$$

Wtedy

$$\ln \mathcal{L}(\vec{\theta}) = -\frac{1}{2}\sum_{i,j=1}^{n}\left(y_i - \eta\left(x_i;\vec{\theta}\right)\right)\left(\mathbf{V}^{-1}\right)_{ij}\left(y_j - \eta\left(x_j;\vec{\theta}\right)\right) + \ldots$$

a zasada najmniejszych kwadratów sprowadza się do warunku

$$\mathcal{R}(\vec{\theta}) = \sum_{i,j=1}^{n}\left(y_i - \eta\left(x_i;\vec{\theta}\right)\right)\left(\mathbf{V}^{-1}\right)_{ij}\left(y_j - \eta\left(x_j;\vec{\theta}\right)\right) = \min(\vec{\theta})$$

Zależność funkcyjną $\eta\left(x_i;\hat{\vec{\theta}}\right)$ nazywamy krzywą regresji najlepszego dopasowania metody najmniejszych kwadratów. Parametry $\hat{\vec{\theta}}$ to estymatory MNK.

MNK - przypadek liniowy

Rozważmy sytuację kiedy związek pomiędzy parametrami θ_i, a wielkością mierzoną η przyjmuje postać liniową:

$$\eta(x;\vec{\theta}) = \varphi_1(x)\theta_1 + \varphi_2(x)\theta_2 + \ldots + \varphi_m(x)\theta_m$$

Dla n punktów pomiarowych (x_i,y_i) i = 1,...,n otrzymujemy układ n równań:

$$\begin{cases} \eta_1 = \varphi_1(x_1)\theta_1 + \varphi_2(x_1)\theta_2 + \ldots + \varphi_m(x_1)\theta_m \\ \eta_2 = \varphi_1(x_2)\theta_1 + \varphi_2(x_2)\theta_2 + \ldots + \varphi_m(x_2)\theta_m \\ \quad\vdots \\ \eta_n = \varphi_1(x_n)\theta_1 + \varphi_2(x_n)\theta_2 + \ldots + \varphi_m(x_n)\theta_m \end{cases} \quad \Leftrightarrow \quad \vec{\eta} = \Phi\vec{\theta}$$

gdzie:

$$\vec{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} \qquad \Phi = \begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_m(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_m(x_2) \\ \vdots & & \ddots & \vdots \\ \varphi_1(x_n) & \varphi_2(x_n) & \cdots & \varphi_m(x_n) \end{bmatrix} \qquad \vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{pmatrix}$$

Macierz Φ nazywamy macierzą konstrukcyjną.

MNK – przypadek liniowy

Zasada najmniejszych kwadratów przyjmuje postać (oznaczenie $\mathbf{Q} \equiv \mathbf{V}^{-1}$):

$$\mathcal{R} = \left(\vec{y} - \Phi\vec{\theta}\right)^{\mathrm{T}} \mathbf{Q}\left(\vec{y} - \Phi\vec{\theta}\right) = \left(\vec{y}^{\mathrm{T}} - \vec{\theta}^{\mathrm{T}}\Phi^{\mathrm{T}}\right)\mathbf{Q}\left(\vec{y} - \Phi\vec{\theta}\right) = \mathbf{min}\left(\vec{\theta}\right)$$

Wypiszemy wielkość $\mathcal{R}$ w jawnej postaci:

$$\mathcal{R} = \sum_{i,j=1}^{n}\left(y_i - \sum_{k=1}^{m}\varphi_k(x_i)\theta_k\right)Q_{ij}\left(y_j - \sum_{l=1}^{m}\varphi_l(x_i)\theta_l\right) = \sum_{i,j=1}^{n} y_i Q_{ij} y_j -$$

$$- \sum_{i,j=1}^{n}\sum_{k=1}^{m}\varphi_k(x_i)\theta_k Q_{ij} y_j - \sum_{i,j=1}^{n}\sum_{l=1}^{m} y_i Q_{ij}\varphi_l(x_j)\theta_l + \sum_{i,j=1}^{n}\sum_{k,l=1}^{m}\varphi_k(x_i)\theta_k Q_{ij}\varphi_l(x_j)\theta_l =$$

$$= \sum_{i,j=1}^{n} y_i Q_{ij} y_j - 2\sum_{i,j=1}^{n}\sum_{k=1}^{m}\varphi_k(x_i)\theta_k Q_{ij} y_j + \sum_{i,j=1}^{n}\sum_{k,l=1}^{m}\varphi_k(x_i)\theta_k Q_{ij}\varphi_l(x_j)\theta_l$$

Różniczkujemy względem parametru θ_p:

$$\frac{\partial\mathcal{R}}{\partial\theta_p} = -2\sum_{i,j=1}^{n}\sum_{k=1}^{m}\varphi_k(x_i)\delta_{kp} Q_{ij} y_j + \sum_{i,j=1}^{n}\sum_{k,l=1}^{m}\varphi_k(x_i)\delta_{kp} Q_{ij}\varphi_l(x_j)\theta_l +$$

$$+ \sum_{i,j=1}^{n}\sum_{k,l=1}^{m}\varphi_k(x_i)\theta_k Q_{ij}\varphi_l(x_j)\delta_{lp} =$$

$$= -2\sum_{i,j=1}^{n}\varphi_p(x_i) Q_{ij} y_j + \sum_{i,j=1}^{n}\sum_{l=1}^{m}\varphi_p(x_i) Q_{ij}\varphi_l(x_j)\theta_l + \sum_{i,j=1}^{n}\sum_{k=1}^{m}\varphi_k(x_i)\theta_k Q_{ij}\varphi_p(x_j)$$

MNK - przypadek liniowy

Zamieniając indeks l na k oraz zmieniając między sobą nazwy indeksów i oraz j otrzymujemy (macierz $\mathbf{Q}=\mathbf{V}^{-1}$ jest symetryczna):

$$\frac{\partial \mathcal{R}}{\partial \theta_p} = -2 \underbrace{\sum_{i,j=1}^{n} \varphi_p(x_i) Q_{ij} y_j}_{(\Phi^{\mathrm{T}}\,\mathbf{Q}\,y)_p} + 2 \underbrace{\sum_{i,j=1}^{n} \sum_{k=1}^{m} \varphi_p(x_i) Q_{ij} \varphi_k(x_j) \theta_k}_{(\Phi^{\mathrm{T}}\,\mathbf{Q}\,\Phi\vec{\theta})_p} = 0 \qquad p = 1, 2, \ldots, m$$

Otrzymaliśmy układ liniowych równań (zwanych normalnymi) na nieznane parametry θ_i w postaci macierzowej:

$$\Phi^{\mathrm{T}}\,\mathbf{Q}\,\Phi\vec{\theta} = \Phi^{\mathrm{T}}\,\mathbf{Q}\,y$$

o rozwiązaniach liniowo zależnych od mierzonych wielkości:

$$\hat{\vec{\theta}} = \left(\Phi^{\mathrm{T}}\,\mathbf{Q}\,\Phi\right)^{-1} \Phi^{\mathrm{T}}\,\mathbf{Q}\,\vec{y} = \underbrace{\left(\Phi^{\mathrm{T}}\,\mathbf{V}^{-1}\,\Phi\right)^{-1}}_{\equiv \mathbf{W}} \Phi^{\mathrm{T}}\,\mathbf{V}^{-1}\,\vec{y} = \underbrace{\mathbf{W}\,\Phi^{\mathrm{T}}\,\mathbf{V}^{-1}}_{\equiv \Psi}\,\vec{y} = \Psi\vec{y}$$

Jeśli wielkości mierzone są nieobciążone, to również nieobciążone są estymatory parametrów:

$$\mathcal{E}\left[\hat{\vec{\theta}}\right] = \mathcal{E}\left[\Psi\vec{y}\right] = \Psi\mathcal{E}\left[\vec{y}\right] = \Psi\vec{\eta} = \left(\Phi^{\mathrm{T}}\,\mathbf{V}^{-1}\,\Phi\right)^{-1} \Phi^{\mathrm{T}}\,\mathbf{V}^{-1}\,\Phi\vec{\theta} = \vec{\theta}$$

MNK – przypadek liniowy

Macierz kowariancji estymatorów parametrów ma postać:

$$\mathbf{V}\left[\hat{\vec{\theta}}\right]=\mathcal{E}\left[\left(\Psi(\vec{y}-\vec{\eta})\right)\left(\Psi(\vec{y}-\vec{\eta})\right)^{\mathrm{T}}\right]=\mathcal{E}\left[\Psi(\vec{y}-\vec{\eta})(\vec{y}-\vec{\eta})^{\mathrm{T}}\Psi^{\mathrm{T}}\right]=\Psi\mathcal{E}\left[(\vec{y}-\vec{\eta})(\vec{y}-\vec{\eta})^{\mathrm{T}}\right]\Psi^{\mathrm{T}}=$$

$$=\mathbf{W}\Phi^{\mathrm{T}}\mathbf{V}^{-1}\mathcal{E}\left[(\vec{y}-\vec{\eta})(\vec{y}-\vec{\eta})^{\mathrm{T}}\right]\mathbf{V}^{-1}\Phi\mathbf{W}=\mathbf{W}\Phi^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\Phi\mathbf{W}=\mathbf{W}\underbrace{\Phi^{\mathrm{T}}\mathbf{V}^{-1}\Phi}_{\mathbf{W}^{-1}}\mathbf{W}=\mathbf{W}$$

a więc: $\mathbf{V}\left[\hat{\vec{\theta}}\right]=\mathbf{W}=\left(\Phi^{\mathrm{T}}\mathbf{V}^{-1}\Phi\right)^{-1}$

Dysponując estymatami parametrów możemy wykorzystać je do interpolacji bądź ekstrapolacji, konstruując krzywą najlepszego dopasowania:

$$\hat{\eta}(x)=\varphi_1(x)\hat{\theta}_1+\varphi_2(x)\hat{\theta}_2+\ldots+\varphi_m(x)\hat{\theta}_m=\vec{\varphi}^{\mathrm{T}}(x)\hat{\vec{\theta}}$$

Błąd takiej operacji wynosi:

$$\mathcal{V}\left[\hat{\eta}(x)\right]=\mathcal{E}\left[(\hat{\eta}(x)-\eta(x))^2\right]=\mathcal{E}\left[\left(\vec{\varphi}^{\mathrm{T}}(x)(\hat{\vec{\theta}}-\vec{\theta})\right)\left(\vec{\varphi}^{\mathrm{T}}(x)(\hat{\vec{\theta}}-\vec{\theta})\right)^{\mathrm{T}}\right]=$$

$$=\mathcal{E}\left[\vec{\varphi}^{\mathrm{T}}(x)(\hat{\vec{\theta}}-\vec{\theta})(\hat{\vec{\theta}}-\vec{\theta})^{\mathrm{T}}\vec{\varphi}(x)\right]=\vec{\varphi}^{\mathrm{T}}(x)\mathbf{W}\vec{\varphi}(x)$$

Twierdzenie Gaussa-Markowa: Pośród wszystkich nieobciążonych estymatorów, które są liniowymi kombinacjami wielkości mierzonych, estymatory metody najmniejszych kwadratów mają najmniejszą wariancję.

Ocena jakości dopasowania

Ocena jakości dopasowania na podstawie resztkowej różnicy w minimum:

$$\vec{\varepsilon}_{\mathbf{min}} = \vec{y} - \hat{\vec{\eta}}$$

Wartość oczekiwana i macierz kowariancji są odpowiednio równe:

$$\mathcal{E}\left[\vec{\varepsilon}_{\mathbf{min}}\right] = \mathcal{E}\left[\vec{y} - \hat{\vec{\eta}}\right] = \mathcal{E}\left[\vec{y}\right] - \mathcal{E}\left[\hat{\vec{\eta}}\right] = \vec{\eta} - \mathcal{E}\left[\Psi\hat{\vec{\theta}}\right] = \vec{\eta} - \Psi\vec{\theta} = \vec{\eta} - \vec{\eta} = \mathbf{0}$$

$$\mathbf{V}\left[\vec{\varepsilon}_{\mathbf{min}}\right] = \mathcal{E}\left[\left(\vec{y} - \hat{\vec{\eta}}\right)\left(\vec{y} - \hat{\vec{\eta}}\right)^{\mathrm{T}}\right] = \ldots = \mathbf{V}\left[\vec{y}\right] - \mathbf{V}\left[\hat{\vec{\eta}}\right]$$

Ocena błędów systematycznych – wpływ: $\mathbf{z}_i = \dfrac{\mathbf{y}_i - \hat{\eta}_i}{\sqrt{\mathcal{V}[\mathbf{y}_i] - \mathcal{V}[\hat{\eta}_i]}}$

Ocena jakości dopasowania na podstawie wartości $\mathcal{R}_{\text{min}}$:

- porównanie modeli – im mniejsze $\mathcal{R}_{\text{min}}$ tym lepsza zgodność (uwaga: dopasowując wielomian n−1 stopnia do n punktów dostaniemy $\mathcal{R}_{\text{min}}=0$)
- gdy mierzone wielkości pochodzą z rozkładu normalnego, wówczas estymatory parametrów mają rozkłady normalne, natomiast wielkość $\mathcal{R}_{\text{min}}$ jest statystyką χ^2 wylosowaną z rozkładu o (n−m) stopniach swobody $\mathcal{R}_{\text{min}} = \chi^2_{\text{n-m}}$ o ile postulowana zależność $\eta(x; \theta_1,\ldots,\theta_m)$ jest słuszna.

(Prawda jeśli macierz $\mathbf{V}(\mathbf{y})$ jest macierzą kowariancji, a nie jej estymat.)

MNK – dopasowanie linii prostej

Rozważmy dopasowanie linii prostej do danych doświadczalnych:

$$\vec{\eta} = \Phi\vec{\theta} \qquad \Leftrightarrow \qquad \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

Zakładamy, że pomiary nie są skorelowane:

$$V = \begin{bmatrix} s_1^2 & 0 & \cdots & 0 \\ 0 & s_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_n^2 \end{bmatrix} \qquad \Rightarrow \qquad V^{-1} = \begin{bmatrix} \frac{1}{s_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{s_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{s_n^2} \end{bmatrix}$$

Wprowadzamy oznaczenia:

$$S = \sum_{i=1}^{n} \frac{1}{s_i^2} \quad S_x = \sum_{i=1}^{n} \frac{x_i}{s_i^2} \quad S_{xx} = \sum_{i=1}^{n} \frac{x_i^2}{s_i^2} \quad S_{xy} = \sum_{i=1}^{n} \frac{x_i y_i}{s_i^2}$$

$$S_y = \sum_{i=1}^{n} \frac{y_i}{s_i^2} \quad S_{yy} = \sum_{i=1}^{n} \frac{y_i^2}{s_i^2} \quad \Delta = \frac{1}{SS_{xx} - S_x^2}$$

MNK – dopasowanie linii prostej

Konstruujemy macierze W **i** Ψ:

$$\Phi^T V^{-1} \Phi = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \frac{1}{s_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{s_n^2} \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} \frac{1}{s_1^2} & \cdots & \frac{1}{s_n^2} \\ \frac{x_1}{s_1^2} & \cdots & \frac{x_n}{s_n^2} \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} S & S_x \\ S_x & S_{xx} \end{bmatrix}$$

$$W = \left(\Phi^T V^{-1} \Phi\right)^{-1} = \Delta \begin{bmatrix} S_{xx} & -S_x \\ -S_x & S \end{bmatrix}$$

$$\Psi = W \Phi^T V^{-1} = \Delta \begin{bmatrix} S_{xx} & -S_x \\ -S_x & S \end{bmatrix} \begin{bmatrix} \frac{1}{s_1^2} & \cdots & \frac{1}{s_n^2} \\ \frac{x_1}{s_1^2} & \cdots & \frac{x_n}{s_n^2} \end{bmatrix} = \Delta \begin{bmatrix} \frac{S_{xx} - x_1 S_x}{s_1^2} & \cdots & \frac{S_{xx} - x_n S_x}{s_n^2} \\ \frac{x_1 S - S_x}{s_1^2} & \cdots & \frac{x_n S - S_x}{s_n^2} \end{bmatrix}$$

MNK – dopasowanie linii prostej

Wyznaczamy estymatory parametrów linii prostej

$$\hat{\vec{\theta}} = \Psi \vec{y} \quad \Rightarrow \quad \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \Delta \begin{bmatrix} \dfrac{S_{xx} - x_1 S_x}{s_1^2} & \dots & \dfrac{S_{xx} - x_n S_x}{s_n^2} \\ \dfrac{S x_1 - S_x}{s_1^2} & \dots & \dfrac{S x_n - S_x}{s_n^2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \Delta \begin{bmatrix} S_{xx} S_y - S_{xy} S_x \\ S S_{xy} - S_x S_y \end{bmatrix}$$

Z macierzy $W = V[\hat{\vec{\theta}}]$ **odczytujemy:**

$$\mathcal{V}[\hat{\theta}_1] = S_{xx} \Delta \qquad \mathrm{cov}[\hat{\theta}_1, \hat{\theta}_2] = -S_x \Delta \qquad \rho(\hat{\theta}_1, \hat{\theta}_2) = -\frac{S_x}{\sqrt{S S_{xx}}}$$

$$\mathcal{V}[\hat{\theta}_2] = S \Delta$$

Obliczamy wartość statystyki $\mathcal{R}$ w minimum: $$\mathcal{R}_{\min} = S_{yy} - S_y \hat{\theta}_1 - S_{xy} \hat{\theta}_2$$

Ponieważ parametry są liniowymi funkcjami wielkości mierzonych, więc poziomice stałej wartości $\mathcal{R}$ (czyli χ^2) są elipsami:

$$\frac{1}{1-\rho^2} \left[\frac{\left(\theta_1 - \hat{\theta}_1\right)^2}{\mathcal{D}^2[\hat{\theta}_1]} + \frac{\left(\theta_2 - \hat{\theta}_2\right)^2}{\mathcal{D}^2[\hat{\theta}_2]} - 2\rho \left(\frac{\theta_1 - \hat{\theta}_1}{\mathcal{D}[\hat{\theta}_1]} \right) \left(\frac{\theta_2 - \hat{\theta}_2}{\mathcal{D}[\hat{\theta}_2]} \right) \right] = \mathcal{R} - \mathcal{R}_{\min}$$

Dopasowanie linii prostej - dygresja

Dopasowanie linii prostej z pominięciem rachunku macierzowego:

$$\eta(x) = f\left(x;\vec{\theta}\right) = \theta_1 + \theta_2 x$$

Znajdujemy wartości parametrów dla których $\mathcal{R}$ osiąga minimum:

$$\begin{cases} \dfrac{\partial \mathcal{R}}{\partial \theta_1} = -2\sum_{i=1}^{n}\left(\dfrac{y_i - \theta_1 - \theta_2 x_i}{s_i^2}\right) = -2\left(\sum_{i=1}^{n}\dfrac{y_i}{s_i^2} - \theta_1\sum_{i=1}^{n}\dfrac{1}{s_i^2} - \theta_2\sum_{i=1}^{n}\dfrac{x_i}{s_i^2}\right) \\ \dfrac{\partial \mathcal{R}}{\partial \theta_2} = -2\sum_{i=1}^{n} x_i\left(\dfrac{y_i - \theta_1 - \theta_2 x_i}{s_i^2}\right) = -2\left(\sum_{i=1}^{n}\dfrac{x_i y_i}{s_i^2} - \theta_1\sum_{i=1}^{n}\dfrac{x_i}{s_i^2} - \theta_2\sum_{i=1}^{n}\dfrac{x_i^2}{s_i^2}\right) \end{cases}$$

wprowadzając oznaczenia:

$$S = \sum_{i=1}^{n}\frac{1}{s_i^2} \quad S_x = \sum_{i=1}^{n}\frac{x_i}{s_i^2} \quad S_{xx} = \sum_{i=1}^{n}\frac{x_i^2}{s_i^2} \quad S_{xy} = \sum_{i=1}^{n}\frac{x_i y_i}{s_i^2}$$

$$S_y = \sum_{i=1}^{n}\frac{y_i}{s_i^2} \quad S_{yy} = \sum_{i=1}^{n}\frac{y_i^2}{s_i^2} \quad \Delta = \frac{1}{SS_{xx} - S_x^2}$$

dostajemy:

$$\begin{cases} S_y - \hat{\theta}_1 S - \hat{\theta}_2 S_x = 0 \\ S_{xy} - \hat{\theta}_1 S_x - \hat{\theta}_2 S_{xx} = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} \hat{\theta}_1 = \left(S_{xx} S_y - S_x S_{xy}\right)\Delta \\ \hat{\theta}_2 = \left(SS_{xy} - S_x S_y\right)\Delta \end{cases}$$

MNK – dopasowanie linii prostej

Linia prosta najlepszego dopasowania pozwala dokonać interpolacji lub ekstrapolacji:

$$\hat{\eta}(x) = \hat{\theta}_1 + \hat{\theta}_2 x$$

Błąd tej operacji dany jest przez:

$$\mathcal{V}[\hat{\eta}] = \varphi^T(x)\,\mathbf{W}\,\varphi(x) = \Delta\begin{bmatrix}1 & x\end{bmatrix}\begin{bmatrix}S_{xx} & -S_x \\ -S_x & S\end{bmatrix}\begin{bmatrix}1 \\ x\end{bmatrix} = \Delta\begin{bmatrix}S_{xx} - x S_x & -S_x + x S\end{bmatrix}\begin{bmatrix}1 \\ x\end{bmatrix} =$$

$$= \Delta(S_{xx} - x S_x - x S_x + x^2 S) = \Delta S\left(\frac{S S_{xx} - S_x^2}{S^2} + \left(x - \frac{S_x}{S}\right)^2\right) = \frac{1}{S} + \mathcal{V}[\hat{\theta}_2]\left(x - \frac{S_x}{S}\right)^2$$

Problem dopasowania do danych wielu nieliniowych funkcji można sprowadzić do problemu dopasowania linii prostej:

$$y = ab^x \qquad \Rightarrow \qquad y' = \ln a + x \ln b = a' + b'x$$

$$y = ax^b \qquad \Rightarrow \qquad y' = \ln a + b \ln x = a' + bx'$$

$$y = a\,e^{bx} \qquad \Rightarrow \qquad y' = \ln a + bx = a' + bx$$

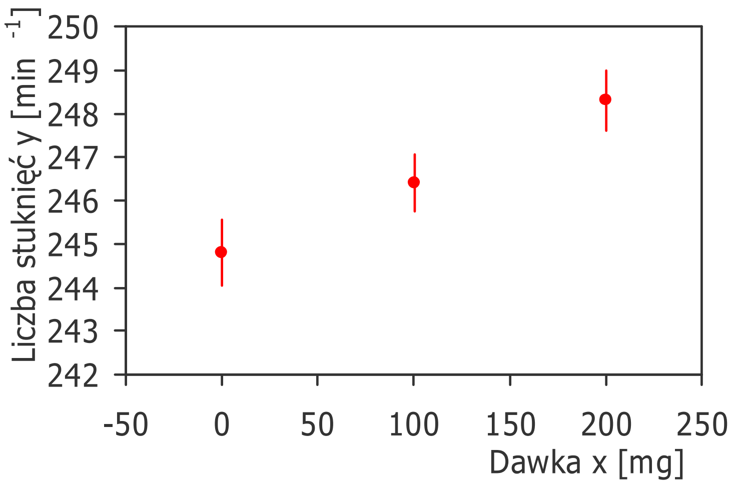
$$y = \frac{x}{a + bx} \qquad \Rightarrow \qquad y' = b\frac{1}{x} + a = a + bx'$$

Dopasowanie linii prostej

<u>Przykład:</u> Wykonano badania wpływu kofeiny na sprawność wykonywania prostych czynności manualnych. Trzydziestu studentów podzielono losowo na trzy grupy po dziesięć osób i każdej grupie podano różne ilości kofeiny: 0 mg, 100 mg i 200 mg. Po dwóch godzinach poproszono ich o wykonanie z maksymalną szybkością czynności polegającej na stukaniu palcami po stole. Wyniki pomiarów przedstawia tabela.

Dawka x [mg]	Liczba stuknięć palcami na minutę y									
0	**242**	**245**	**244**	**248**	**247**	**248**	**242**	**244**	**246**	**242**
100	**248**	**246**	**245**	**247**	**248**	**250**	**247**	**246**	**243**	**244**
200	**246**	**248**	**250**	**252**	**248**	**250**	**246**	**248**	**245**	**250**

$\bar{y} \pm \sigma_{\bar{y}}$
244.8 ± 0.76
246.4 ± 0.65
248.3 ± 0.70



Prosta dopasowana MNK:

$$y = \underbrace{(244.7 \pm 0.7)}_{\hat{\theta}_1 \pm \hat{\sigma}_{\theta_1}} + \underbrace{(0.017 \pm 0.005)}_{\hat{\theta}_2 \pm \hat{\sigma}_{\theta_2}} x$$

Kowariancja i współczynnik korelacji pomiędzy estymatorami parametrów:

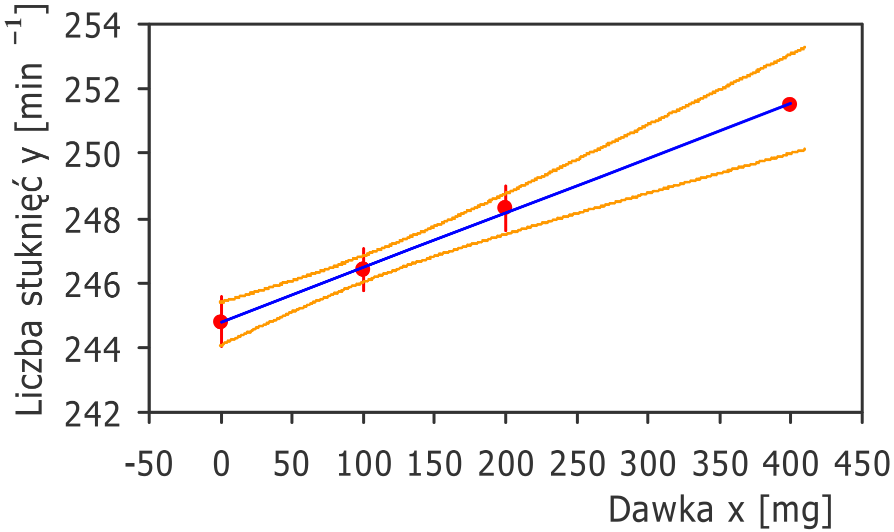
$$\mathrm{cov}[\hat{\theta}_1, \hat{\theta}_2] = -0.0028 \qquad \rho\left(\hat{\theta}_1, \hat{\theta}_2\right) = -0.8$$

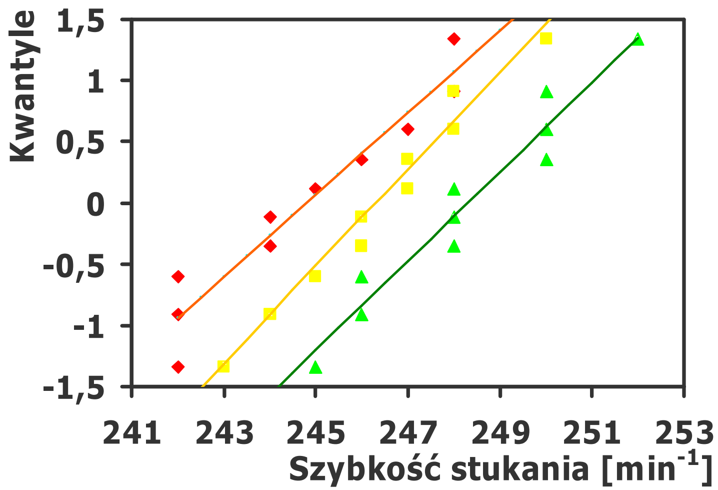
Dopasowanie linii prostej

$$S = 6.139 \qquad S_x = 644.85$$

$$\hat{\eta}(x) = \hat{\theta}_1 + \hat{\theta}_2 x \pm \sqrt{\frac{1}{S} + \mathcal{V}[\hat{\theta}_2]\left(x - \frac{S_x}{S}\right)^2}$$

$$\eta(400) = 251.5 \pm 1.5$$





$$\hat{F}(x) = \begin{cases} 0 & x < x_1 \\ \dfrac{i}{n+1} & x_i \le x < x_{i+1} \\ \dfrac{n}{n+1} & x \ge x_n \end{cases}$$

$$F(q_i) = \frac{i}{n+1} = \frac{i}{11}$$

$$q = \frac{x - \mu}{\sigma}$$

$F(q_i)$	0.091	0.182	0.273	0.364	0.455	0.546	0.636	0.727	0.818	0.909
q_i	-1.335	-0.908	-0.605	-0.349	-0.114	0.114	0.349	0.605	0.908	1.335

Dopasowanie linii prostej

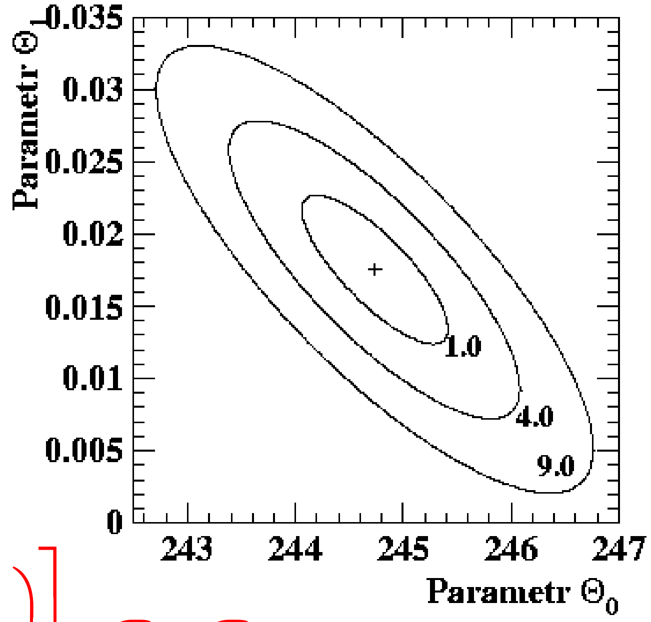
Jeśli hipoteza o zależności liniowej jest prawdziwa, to statystyka $\mathcal{R}$ ma rozkład χ^2 o $n-m = 3 - 2 = 1$ stopniach swobody.

$$\mathcal{R}_{\min} = S_{yy} - S_y \hat{\theta}_1 - S_{xy} \hat{\theta}_2 = 0.033 \quad \Rightarrow \quad P(\chi^2 > \mathcal{R}_{\min}) = \int_{\mathcal{R}_{\min}}^{\infty} \chi_1(u)du = 0.857$$

Jeśli zażądamy $P\left(\chi^2 > \mathcal{R}_{\min}\right) = \int_{\mathcal{R}_{\min}}^{\infty} \chi_1(u)du = 0.99 \quad \Rightarrow \quad \mathcal{R}_{\min} = 0.00015$

Oznacza to, że średnie odchylenie każdego punktu od dopasowanej linii wynosi w jednostkach typowej niepewności wynosi tylko

$$\sqrt{0.00015/3} \cong 0.007$$



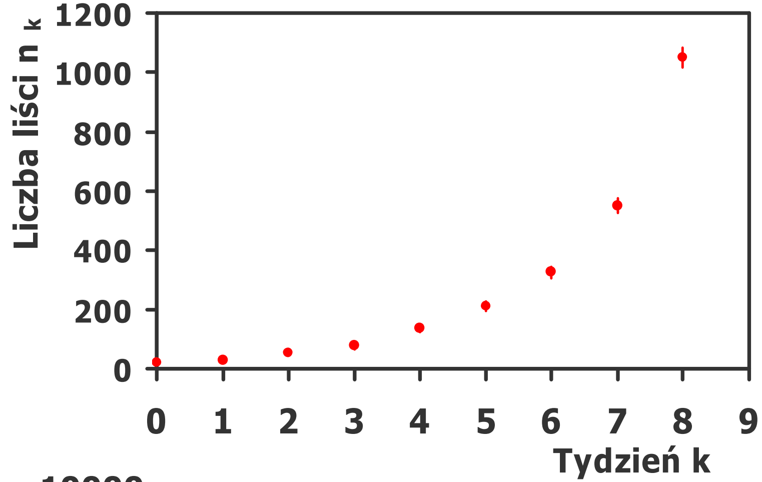
Ponieważ parametry są liniowymi funkcjami wielkości mierzonych, więc poziomice stałej wartości $\mathcal{R}$ (czyli χ^2) są elipsami:

$$\frac{1}{1-\rho^2}\left[\frac{\left(\theta_1-\hat{\theta}_1\right)^2}{\mathcal{D}^2[\hat{\theta}_1]} + \frac{\left(\theta_2-\hat{\theta}_2\right)^2}{\mathcal{D}^2[\hat{\theta}_2]} - 2\rho\left(\frac{\theta_1-\hat{\theta}_1}{\mathcal{D}[\hat{\theta}_1]}\right)\left(\frac{\theta_2-\hat{\theta}_2}{\mathcal{D}[\hat{\theta}_2]}\right)\right] = \mathcal{R} - \mathcal{R}_{\min}$$

Dopasowanie krzywej eksponencjalnej

<u>Przykład:</u> W tabeli podane są dane o liczbie n_k liści rzęsy wodnej jako funkcji czasu k liczonego w tygodniach.

k	0	1	2	3	4	5	6	7	8
n_k	20	30	52	77	135	211	326	550	1052

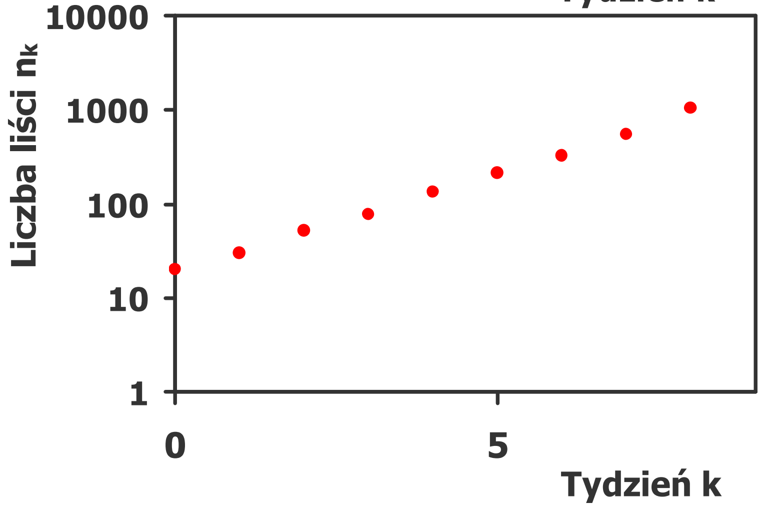


$$y = a\,e^{bx} \quad \Rightarrow \quad \underbrace{\ln y = \ln a + bx}_{z = a' + bx}$$

Uwagi:

- oznaczenia: $x \equiv k \quad y \equiv n_k$
- zaniedbujemy ewidentne korelacje,
- zakładamy, że niepewności zliczeń podlegają rozkładowi Poissona:

$$z = \ln y \quad \Rightarrow \quad \sigma_z^2 = \left(\frac{\partial z}{\partial y}\right)^2 \sigma_y^2 = \frac{\sigma_y^2}{y^2} = \frac{1}{y}$$



Dopasowanie krzywej eksponencjalnej

$$z = (\underbrace{2.81 \pm 0.08}_{a' \pm \sigma_{a'}}) + (\underbrace{0.51 \pm 0.01}_{b \pm \sigma_b})x$$

$$\mathrm{cov}[a', b] = -0.0009 \qquad \rho = -0.97$$

$$\mathcal{R}_{\min} = 11.83 \quad \Rightarrow \quad \mathrm{P}\left(\chi^2 > \mathcal{R}_{\min}\right) = 0.106$$

$$a' = \ln a \quad \Rightarrow \quad a = \mathrm{e}^{a'} \quad \Rightarrow \quad \sigma_a = \mathrm{e}^{a'}\sigma_{a'}$$

$$y = (16.5 \pm 1.3)\exp\left[(0.51 \pm 0.01)x\right]$$