

Statystyka i rachunek prawdopodobieństwa

Wykład 15: Prawo propagacji błędu

Mariusz Tarnopolski

Instytut Astronomii UMK

Statystyka ©2026



Mierzmy linijką (z podziałką o dokładności $\Delta d = 1$ mm) średnicę okręgu:
 $d \pm \Delta d$.

Promień to $r = \frac{d}{2}$; jaka jest niepewność tak obliczonego r ?

- a $\Delta r = \Delta d$?
- b $\Delta r = 2\Delta d$?
- c $\Delta r = \frac{1}{2}\Delta d$?

Mierzymy linijką (z podziałką o dokładności $\Delta d = 1$ mm) średnicę okręgu: $d \pm \Delta d$.

Promień to $r = \frac{d}{2}$; jaka jest niepewność tak obliczonego r ?

- a $\Delta r = \Delta d$?
- b $\Delta r = 2\Delta d$?
- c $\Delta r = \frac{1}{2}\Delta d$?

Ogólniej: jeśli mierzymy wielkość x z dokładnością/niepewnością Δx , to ile wynosi niepewność wielkości $y = f(x)$ wyznaczonej pośrednio? Tzn. jak się propaguje niepewność wyznaczenia x na niepewność wyznaczenia y ?

Mierzymy linijką (z podziałką o dokładności $\Delta d = 1$ mm) średnicę okręgu:
 $d \pm \Delta d$.

Promień to $r = \frac{d}{2}$; jaka jest niepewność tak obliczonego r ?

- a $\Delta r = \Delta d$?
- b $\Delta r = 2\Delta d$?
- c $\Delta r = \frac{1}{2}\Delta d$?

Ogólniej: jeśli mierzymy wielkość x z dokładnością/niepewnością Δx , to ile wynosi niepewność wielkości $y = f(x)$ wyznaczonej pośrednio? Tzn. jak się propaguje niepewność wyznaczenia x na niepewność wyznaczenia y ?

Jeszcze ogólniej: jeśli mierzymy $x_1 \pm \Delta x_1, x_2 \pm \Delta x_2, \dots, x_n \pm \Delta x_n$, to ile wynosi niepewność Δy wielkości $y = f(x_1, x_2, \dots, x_n)$?

Np. wyznaczenie przyspieszenia ziemskiego g ze wzoru $h = \frac{gt^2}{2}$ poprzez pomiar czasu spadku t z wysokości h .

Jeśli zmienne losowe X_i mają niewielkie odchylenia od swoich wartości oczekiwanych $\mu_i = \mathbb{E}(X_i)$ (co ma miejsce w większości laboratoryjnych przypadków, co sprowadza się do założenia o niewielkich błędach/niepewnościach pomiarów bezpośrednich), to zmienną losową Y będącą funkcją $Y = f(X_1, X_2, \dots, X_n)$ zmiennych losowych X_i można rozwinąć w szereg Taylora w pobliżu jej wartości oczekiwanej, określonej z kolei w przybliżeniu liniowym jako

$$\mathbb{E}(Y) \approx f(\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n))$$

Jeśli zmienne losowe X_i mają niewielkie odchylenia od swoich wartości oczekiwanych $\mu_i = \mathbb{E}(X_i)$ (co ma miejsce w większości laboratoryjnych przypadków, co sprowadza się do założenia o niewielkich błędach/niepewnościach pomiarów bezpośrednich), to zmienną losową Y będącą funkcją $Y = f(X_1, X_2, \dots, X_n)$ zmiennych losowych X_i można rozwinąć w szereg Taylora w pobliżu jej wartości oczekiwanej, określonej z kolei w przybliżeniu liniowym jako

$$\mathbb{E}(Y) \approx f(\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n))$$

czyli w okolicy punktu

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_n)), \text{ zatem } \mathbb{E}(Y) \approx f(\boldsymbol{\mu}).$$

Rozwijamy zatem w szereg Taylora ($f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$):

$$Y = \mathbb{E}(Y) + (X_1 - \mu_1) \left. \frac{\partial f}{\partial x_1} \right|_{\mathbf{x}=\boldsymbol{\mu}} + (X_2 - \mu_2) \left. \frac{\partial f}{\partial x_2} \right|_{\mathbf{x}=\boldsymbol{\mu}} + \dots$$

gdzie $\mathbb{E}(Y) \approx f(\boldsymbol{\mu})$. Oznaczmy $a_i = \left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}}$. Wtedy (opuszczając człony wyższych rzędów)

Rozwijamy zatem w szereg Taylora ($f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$):

$$Y = \mathbb{E}(Y) + (X_1 - \mu_1) \left. \frac{\partial f}{\partial x_1} \right|_{\mathbf{x}=\boldsymbol{\mu}} + (X_2 - \mu_2) \left. \frac{\partial f}{\partial x_2} \right|_{\mathbf{x}=\boldsymbol{\mu}} + \dots$$

gdzie $\mathbb{E}(Y) \approx f(\boldsymbol{\mu})$. Oznaczmy $a_i = \left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}}$. Wtedy (opuszczając czony wyższych rzędów)

$$\begin{aligned} V(Y) &\approx V\left(\sum_{i=1}^n (X_i - \mu_i) a_i\right) = V\left(\sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i \mu_i\right) = V\left(\sum_{i=1}^n a_i X_i\right) \\ &= \sum_{i=1}^n a_i^2 V(X_i) + 2 \sum_{i=1}^n \sum_{k=1}^{i-1} a_i a_k \text{Cov}(X_i, X_k) \end{aligned}$$

(Por. własności na slajdach 12 i 33 [wykładu 7.](#))

Otrzymujemy tedy **prawo propagacji błędu**:

$$\sigma_Y^2 \approx \sum_{i=1}^n a_i^2 \sigma_{X_i}^2 + 2 \sum_{i=1}^n \sum_{k=1}^{i-1} a_i a_k \rho(X_i, X_k) \sigma_{X_i} \sigma_{X_k}$$

gdzie

$$a_i = \left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}}$$

oraz skorzystaliśmy z tego, że (zob. slajd 2, [wykład 14](#))

$$\text{Cov}(X_i, X_k) = \rho(X_i, X_k) \sigma_{X_i} \sigma_{X_k}$$

Człony z kowariancją znikają gdy zmienne X_i są niezależne—typowa sytuacja przy dopasowywaniu funkcji do danych (por. [wykład 14](#)) lub robocze założenie w większości sytuacji gdy spodziewamy się, że niepewności badanych wielkości dominują nad ich kowariancjami. W takim wypadku prawo propagacji błędu upraszcza się do

$$\sigma_Y^2 \approx \sum_{i=1}^n \left(\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}} \right)^2 \sigma_{X_i}^2$$

Pierwiastek z wariancji to odchylenie standardowe, czyli dyspersja zmiennej losowej wokół jej wartości oczekiwanej:

$$\sigma_Y \approx \sqrt{\sum_{i=1}^n \sigma_{X_i}^2 \left(\left. \frac{\partial f}{\partial x_i} \right|_{x=\mu} \right)^2}$$

co w szczególnym przypadku gdy $y = f(x)$ sprowadza się do

$$\sigma_Y \approx \sqrt{\sigma_X^2 \left(\left. \frac{\partial f}{\partial x} \right|_{x=\mu} \right)^2} = \left| \left. \frac{\partial f}{\partial x} \right|_{x=\mu} \right| \sigma_X$$

co, poprzez utożsamienie niepewności standardowej z odchyleniem standardowym, prowadzi do powszechnego wyrażenia

$$\Delta y = \left| \left. \frac{\partial f}{\partial x} \right|_{\mathbb{E}(X)} \right| \Delta x$$

i podobnie dla poprzednich wzorów:

- gdy $y = f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$:

$$\Delta y = \sqrt{\sum_{i=1}^n \left(\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}} \right)^2} (\Delta x_i)^2$$

- oraz gdy kowariancje $Cov(X_i, X_k)$ są niepomijalne:

$$(\Delta y)^2 = \sum_{i=1}^n \left(\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}} \right)^2 (\Delta x_i)^2 +$$

$$2 \sum_{i=1}^n \sum_{k=1}^{i-1} \left(\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}} \right) \left(\left. \frac{\partial f}{\partial x_k} \right|_{\mathbf{x}=\boldsymbol{\mu}} \right) \rho(x_i, x_k) \Delta x_i \Delta x_k$$

Podkreślmy:

- rozwinięcie Taylora zostało wykonane do członów liniowych w wariancji/kowariancji
- $Cov(X_1, X_2) = \rho(X_1, X_2) \sigma_{X_1} \sigma_{X_2}$ może być dodatnia lub ujemna (por. slajdy 2–3 z [wykładu 14](#))—teoretycznie może się zdarzyć, że ujemna kowariancja skompensuje dodatnie wariancje w mało prawdopodobnym przypadku gdy fluktuacje wokół wartości są kompletnie skorelowane
- najpowszechniej używane prawo propagacji błędu:

$$\Delta y = \sqrt{\sum_{i=1}^n \left(\left. \frac{\partial f}{\partial x_i} \right|_{x=\mu} \right)^2 (\Delta x_i)^2}$$

jest przybliżeniem liniowym dla zmiennych losowych niezależnych—można je stosować jeśli kowariancje, w porównaniu z wariancjami, są pomijalnie małe

- W ogólności prawo propagacji błędu zawiera kowariancje, co bywa szczególnie istotne przy dopasowaniach funkcji do danych pomiarowych (zob. dalsze wykłady)

Podkreślmy:

- rozwinięcie Taylora zostało wykonane do członów liniowych w wariancji/kowariancji
- $Cov(X_1, X_2) = \rho(X_1, X_2) \sigma_{X_1} \sigma_{X_2}$ może być dodatnia lub ujemna (por. slajdy 2–3 z [wykładu 14](#))—teoretycznie może się zdarzyć, że ujemna kowariancja skompensuje dodatnie wariancje w mało prawdopodobnym przypadku gdy fluktuacje wokół wartości są kompletnie skorelowane
- najpowszechniej używane prawo propagacji błędów:

$$\Delta y = \sqrt{\sum_{i=1}^n \left(\left. \frac{\partial f}{\partial x_i} \right|_{x=\mu} \right)^2 (\Delta x_i)^2}$$

jest przybliżeniem liniowym dla zmiennych losowych niezależnych—można je stosować jeśli kowariancje, w porównaniu z wariancjami, są pomijalnie małe

- W ogólności prawo propagacji błędów zawiera kowariancje, co bywa szczególnie istotne przy dopasowaniach funkcji do danych pomiarowych (zob. dalsze wykłady)

Podkreślmy:

- rozwinięcie Taylora zostało wykonane do członów liniowych w wariancji/kowariancji
- $Cov(X_1, X_2) = \rho(X_1, X_2) \sigma_{X_1} \sigma_{X_2}$ może być dodatnia lub ujemna (por. slajdy 2–3 z [wykładu 14](#))—teoretycznie może się zdarzyć, że ujemna kowariancja skompensuje dodatnie wariancje w mało prawdopodobnym przypadku gdy fluktuacje wokół wartości są kompletnie skorelowane
- najpowszechniej używane prawo propagacji błędu:

$$\Delta y = \sqrt{\sum_{i=1}^n \left(\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}} \right)^2 (\Delta x_i)^2}$$

jest przybliżeniem liniowym dla zmiennych losowych niezależnych—można je stosować jeśli kowariancje, w porównaniu z wariancjami, są pomijalnie małe

- W ogólności prawo propagacji błędu zawiera kowariancje, co bywa szczególnie istotne przy dopasowaniach funkcji do danych pomiarowych (zob. dalsze wykłady)

Podkreślmy:

- rozwinięcie Taylora zostało wykonane do członów liniowych w wariancji/kowariancji
- $Cov(X_1, X_2) = \rho(X_1, X_2) \sigma_{X_1} \sigma_{X_2}$ może być dodatnia lub ujemna (por. slajdy 2–3 z [wykładu 14](#))—teoretycznie może się zdarzyć, że ujemna kowariancja skompensuje dodatnie wariancje w mało prawdopodobnym przypadku gdy fluktuacje wokół wartości są kompletnie skorelowane
- najpowszechniej używane prawo propagacji błędów:

$$\Delta y = \sqrt{\sum_{i=1}^n \left(\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}} \right)^2 (\Delta x_i)^2}$$

jest przybliżeniem liniowym dla zmiennych losowych niezależnych—można je stosować jeśli kowariancje, w porównaniu z wariancjami, są pomijalnie małe

- W ogólności prawo propagacji błędów zawiera kowariancje, co bywa szczególnie istotne przy dopasowaniach funkcji do danych pomiarowych (zob. dalsze wykłady)

Przykład. Jeśli $r = \frac{d}{2}$, to $\Delta r = \left| \frac{\partial r}{\partial d} \right| \Delta d = \frac{\Delta d}{2}$

Przykład. Jeśli $r = \frac{d}{2}$, to $\Delta r = \left| \frac{\partial r}{\partial d} \right| \Delta d = \frac{\Delta d}{2}$

Przykład. W ogólności, jeśli $y = ax + b$, to $\Delta y = \left| \frac{\partial y}{\partial x} \right| \Delta x = |a| \Delta x$

Przykład. Jeśli $r = \frac{d}{2}$, to $\Delta r = \left| \frac{\partial r}{\partial d} \right| \Delta d = \frac{\Delta d}{2}$

Przykład. W ogólności, jeśli $y = ax + b$, to $\Delta y = \left| \frac{\partial y}{\partial x} \right| \Delta x = |a| \Delta x$

Przykład. Ponadto, jeśli $z = ax + by$, to

$$\Delta z = \sqrt{\left(\frac{\partial z}{\partial x} \right)^2 (\Delta x)^2 + \left(\frac{\partial z}{\partial y} \right)^2 (\Delta y)^2} = \sqrt{a^2 (\Delta x)^2 + b^2 (\Delta y)^2}$$

Przykład. Jeśli $r = \frac{d}{2}$, to $\Delta r = \left| \frac{\partial r}{\partial d} \right| \Delta d = \frac{\Delta d}{2}$

Przykład. W ogólności, jeśli $y = ax + b$, to $\Delta y = \left| \frac{\partial y}{\partial x} \right| \Delta x = |a| \Delta x$

Przykład. Ponadto, jeśli $z = ax + by$, to

$$\Delta z = \sqrt{\left(\frac{\partial z}{\partial x} \right)^2 (\Delta x)^2 + \left(\frac{\partial z}{\partial y} \right)^2 (\Delta y)^2} = \sqrt{a^2 (\Delta x)^2 + b^2 (\Delta y)^2}$$

Jeśli jednak $\rho(x, y) \neq 0$ (czyli kowariancja nie znika), to

$$\Delta z = \sqrt{a^2 (\Delta x)^2 + b^2 (\Delta y)^2 + 2ab\rho(x, y) \Delta x \Delta y}$$

Przykład. Jeśli $h = \frac{gt^2}{2}$, to $g = \frac{2h}{t^2}$, zatem (zakładając niezależność, tj. h i t są mierzone niezależnie od siebie, np. różnymi przyrządami—linijką i stoperem; wtedy ich kowariancja jest dostatecznie dobrze przybliżana jako zerowa)

$$\begin{aligned}\Delta g &= \sqrt{\left(\frac{\partial g}{\partial h}\right)^2 (\Delta h)^2 + \left(\frac{\partial g}{\partial t}\right)^2 (\Delta t)^2} \\ &= \sqrt{\left(\frac{2}{t^2}\right)^2 (\Delta h)^2 + \left(\frac{-4h}{t^3}\right)^2 (\Delta t)^2} \\ &= \sqrt{g^2 \left(\frac{\Delta h}{h}\right)^2 + (2g)^2 \left(\frac{\Delta t}{t}\right)^2} \\ &= g \sqrt{\left(\frac{\Delta h}{h}\right)^2 + 4 \left(\frac{\Delta t}{t}\right)^2}\end{aligned}$$

Przykład. Jeśli obwód elektryczny spełnia prawo Ohma, to $R = \frac{U}{I}$.

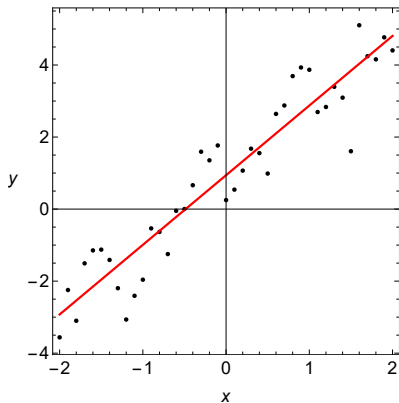
Wtedy

$$\begin{aligned}\Delta R &= \sqrt{\left(\frac{\partial R}{\partial U}\right)^2 (\Delta U)^2 + \left(\frac{\partial R}{\partial I}\right)^2 (\Delta I)^2} \\ &= \sqrt{\left(\frac{1}{I}\right)^2 (\Delta U)^2 + \left(-\frac{U}{I^2}\right)^2 (\Delta I)^2} \\ &= \sqrt{\left(\frac{R}{U}\Delta U\right)^2 + \left(\frac{RI}{I^2}\Delta I\right)^2} \\ &= \sqrt{R^2 \left(\frac{\Delta U}{U}\right)^2 + R^2 \left(\frac{\Delta I}{I}\right)^2} \\ &= R\sqrt{\left(\frac{\Delta U}{U}\right)^2 + \left(\frac{\Delta I}{I}\right)^2}\end{aligned}$$

Zakładamy, że niepewności pomiarów U i I są w przybliżeniu niezależne, więc pomijamy człon z kowariancją, tzn. mimo że U i I są fizycznie powiązane prawem Ohma, zakładamy tutaj, że błędy odczytu woltomierza i amperomierza są od siebie niezależne.

Przykład. Do danych pomiarowych $\{x_i, y_i\}$ dopasowano model $y = ax + b$ uzyskując najlepsze wartości parametrów $a = 1.56, b = 1.03$, oraz macierz kowariancji

$$C^{ab} = \begin{pmatrix} V(a) & Cov(a, b) \\ Cov(b, a) & V(b) \end{pmatrix} = \begin{pmatrix} 0.02 & -10^{-18} \\ -10^{-18} & 0.03 \end{pmatrix}$$



Niepewności dopasowanych parametrów to

$$\Delta a = \sqrt{V(a)} = \sqrt{0.02} \approx 0.14$$

$$\Delta b = \sqrt{V(b)} = \sqrt{0.03} \approx 0.17$$

oraz ewidentnie kowariancje są pomijalnie małe w porównaniu z wariancjami, tj. C^{ab} jest praktycznie diagonalna.

Niech będzie model $y = f(x; \theta)$ z parametrami $\theta = (\theta_1, \theta_2, \dots, \theta_p) \in \mathbb{R}^p$ dopasowywany do danych pomiarowych $\{x_i, y_i\}_{i=1}^n$; tedy $y_i = f(x_i; \theta) + \varepsilon_i$, gdzie $\varepsilon_i \sim \mathcal{N}(0, \sigma)$.

Niech będzie model $y = f(x; \theta)$ z parametrami $\theta = (\theta_1, \theta_2, \dots, \theta_p) \in \mathbb{R}^p$ dopasowywany do danych pomiarowych $\{x_i, y_i\}_{i=1}^n$; tedy $y_i = f(x_i; \theta) + \varepsilon_i$, gdzie $\varepsilon_i \sim \mathcal{N}(0, \sigma)$. Dopasowanie odbywa się poprzez minimalizację sumy kwadratów odchyleń danych od modelu:

$$S(\theta) = \sum_{i=1}^n r_i^2(\theta) \equiv \sum_{i=1}^n [y_i - f(x_i; \theta)]^2$$

Niech będzie model $y = f(x; \theta)$ z parametrami $\theta = (\theta_1, \theta_2, \dots, \theta_p) \in \mathbb{R}^p$ dopasowywany do danych pomiarowych $\{x_i, y_i\}_{i=1}^n$; tedy $y_i = f(x_i; \theta) + \varepsilon_i$, gdzie $\varepsilon_i \sim \mathcal{N}(0, \sigma)$. Dopasowanie odbywa się poprzez minimalizację sumy kwadratów odchyleń danych od modelu:

$$S(\theta) = \sum_{i=1}^n r_i^2(\theta) \equiv \sum_{i=1}^n [y_i - f(x_i; \theta)]^2$$

lub gdy każdemu punktowi przypisujemy wagi $w_i = \sigma_i^{-2}$:

$$S(\theta) = \sum_{i=1}^n \left(\frac{r_i(\theta)}{\sigma_i} \right)^2$$

Niech będzie model $y = f(x; \theta)$ z parametrami $\theta = (\theta_1, \theta_2, \dots, \theta_p) \in \mathbb{R}^p$ dopasowywany do danych pomiarowych $\{x_i, y_i\}_{i=1}^n$; tedy $y_i = f(x_i; \theta) + \varepsilon_i$, gdzie $\varepsilon_i \sim \mathcal{N}(0, \sigma)$. Dopasowanie odbywa się poprzez minimalizację sumy kwadratów odchyłeń danych od modelu:

$$S(\theta) = \sum_{i=1}^n r_i^2(\theta) \equiv \sum_{i=1}^n [y_i - f(x_i; \theta)]^2$$

lub gdy każdemu punktowi przypisujemy wagi $w_i = \sigma_i^{-2}$:

$$S(\theta) = \sum_{i=1}^n \left(\frac{r_i(\theta)}{\sigma_i} \right)^2$$

Zapisując macierzowo:

$$S(\theta) = r(\theta)^\top W r(\theta)$$

gdzie $W = \text{diag}(\sigma_i^{-2})$ jest diagonalną macierzą wag (zob. też slajdy **21–23**).

Niech $\hat{\theta}$ będzie najoptymalniejszym wektorem parametrów znalezionych powyższą metodą. Funkcję $f(x; \theta)$ można rozwinąć w szereg Taylora w pobliżu tego wektora:

$$f(x; \theta) = f(x; \hat{\theta}) + J (\theta - \hat{\theta})$$

gdzie J jest macierzą Jacobiego o elementach:

$$J_{ij} = \left(\frac{\partial f(x_i; \theta)}{\partial \theta_j} \Big|_{\theta = \hat{\theta}} \right)$$

Niech $\hat{\theta}$ będzie najoptymalniejszym wektorem parametrów znalezionych powyższą metodą. Funkcję $f(x; \theta)$ można rozwinąć w szereg Taylora w pobliżu tego wektora:

$$f(x; \theta) = f(x; \hat{\theta}) + J (\theta - \hat{\theta})$$

gdzie J jest macierzą Jacobiego o elementach:

$$J_{ij} = \left(\frac{\partial f(x_i; \theta)}{\partial \theta_j} \Big|_{\theta = \hat{\theta}} \right)$$

Macierz kowariancji $C(\hat{\theta})$ dopasowanych parametrów $\hat{\theta}$ to (wyprowadzenie poprzez informację Fishera—pomijamy; lecz por. slajd 6, wykład 9)

$$C(\hat{\theta}) \approx (J^T W J)^{-1}$$

Niech $\hat{\theta}$ będzie najoptymalniejszym wektorem parametrów znalezionych powyższą metodą. Funkcję $f(x; \theta)$ można rozwinąć w szereg Taylora w pobliżu tego wektora:

$$f(x; \theta) = f(x; \hat{\theta}) + J (\theta - \hat{\theta})$$

gdzie J jest macierzą Jacobiego o elementach:

$$J_{ij} = \left(\frac{\partial f(x_i; \theta)}{\partial \theta_j} \Big|_{\theta = \hat{\theta}} \right)$$

Macierz kowariancji $C(\hat{\theta})$ dopasowanych parametrów $\hat{\theta}$ to (wyprowadzenie poprzez informację Fishera—pomijamy; lecz por. slajd 6, wykład 9)

$$C(\hat{\theta}) \approx (J^T W J)^{-1}$$

To właśnie w ten sposób np. w PYTHONIE komenda `scipy.optimize.curve_fit` zwraca macierz kowariancji (zob. slajdy 13 i 16).

Przykład. Do danych pomiarowych $\{x_i, y_i\}$ dopasowano model $y = ax^b$ uzyskując najlepsze wartości parametrów $a = 4, b = 3$, oraz macierz kowariancji

$$C^{ab} = \begin{pmatrix} V(a) & Cov(a, b) \\ Cov(b, a) & V(b) \end{pmatrix} = \begin{pmatrix} 0.005 & 0.03 \\ 0.03 & 0.2 \end{pmatrix}$$

Niepewności dopasowanych parametrów to

$$\Delta a = \sqrt{V(a)} = \sqrt{0.005} \approx 0.07$$

$$\Delta b = \sqrt{V(b)} = \sqrt{0.2} \approx 0.45$$

Jako że $V(a) \ll Cov(a, b) \ll V(b)$, to żeby łatwiej ocenić siłę związku między parametrami a i b obliczymy macierz korelacji ρ (por. slajd 2, [wykład 14](#)):

$$\rho_{ij} = \frac{C_{ij}^{ab}}{\sqrt{C_{ii}^{ab} C_{jj}^{ab}}}$$

która dla powyższej C^{ab} wynosi

$$\rho_{ab} = \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}$$

Współczynnik korelacji $\rho(a, b) = 0.95$ implikuje, że korelacja między dopasowanymi parametrami jest bardzo silna—jeśli a wzrośnie, to b też wzrośnie *niemal deterministycznie*. Oznacza to, że różne pary (a, b) dają niemal takie same dopasowania funkcji $y = ax^b$ do danych.

Współczynnik korelacji $\rho(a, b) = 0.95$ implikuje, że korelacja między dopasowanymi parametrami jest bardzo silna—jeśli a wzrośnie, to b też wzrośnie *niemal deterministycznie*. Oznacza to, że różne pary (a, b) dają niemal takie same dopasowania funkcji $y = ax^b$ do danych. Jeśli model jest uzasadniony (np. teorią fizyczną), ale problemem są dane (np. duży rozrzut, co prowadzi do źle określonych parametrów), można np. przejść do innej bazy, w której macierz kowariancji C stanie się diagonalna:

$$C = P\Lambda P^{-1}$$

Współczynnik korelacji $\rho(a, b) = 0.95$ implikuje, że korelacja między dopasowanymi parametrami jest bardzo silna—jeśli a wzrośnie, to b też wzrośnie *niemal deterministycznie*. Oznacza to, że różne pary (a, b) dają niemal takie same dopasowania funkcji $y = ax^b$ do danych.

Jeśli model jest uzasadniony (np. teorią fizyczną), ale problemem są dane (np. duży rozrzut, co prowadzi do źle określonych parametrów), można np. przejść do innej bazy, w której macierz kowariancji C stanie się diagonalna:

$$C = P\Lambda P^{-1}$$

gdzie $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ jest macierzą wartości własnych, zaś P macierzą, której kolumny są wektorami własnymi (C jest symetryczna, więc $P^{-1} = P^T$).

W takiej bazie możemy zdefiniować nowe parametry α, β takie, że

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = P^{-1} \begin{pmatrix} a \\ b \end{pmatrix}$$

W takiej bazie możemy zdefiniować nowe parametry α, β takie, że

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = P^{-1} \begin{pmatrix} a \\ b \end{pmatrix}$$

Tutaj: $\alpha = 0.15a + 0.99b$, $\beta = 0.99a - 0.15b$ oraz

$$C^{\alpha\beta} = \begin{pmatrix} 0.2045 & 0 \\ 0 & 0.0005 \end{pmatrix}$$

W takiej bazie możemy zdefiniować nowe parametry α, β takie, że

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = P^{-1} \begin{pmatrix} a \\ b \end{pmatrix}$$

Tutaj: $\alpha = 0.15a + 0.99b$, $\beta = 0.99a - 0.15b$ oraz

$$C^{\alpha\beta} = \begin{pmatrix} 0.2045 & 0 \\ 0 & 0.0005 \end{pmatrix}$$

Nowe parametry (α, β) są liniowymi kombinacjami starych (a, b) , ale w nowej bazie macierz kowariancji staje się diagonalna, tj. α i β są niezależne. Ich niepewności to $\Delta\alpha = \sqrt{0.2045} = 0.45$,
 $\Delta\beta = \sqrt{0.0005} = 0.02$.

Możemy też odwrócić transformację:

$$\begin{pmatrix} a \\ b \end{pmatrix} = P \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

czyli $a = 0.15\alpha - 0.99\beta$, $b = 0.99\alpha + 0.15\beta$. Wtedy chcąc obliczyć np. $\Delta(ab)$ można wyrazić $ab = g(\alpha, \beta) = (0.15\alpha - 0.99\beta)(0.99\alpha + 0.15\beta) = 0.1485(\alpha^2 - \beta^2) - 0.9576\alpha\beta$, co pozwala obliczyć $\Delta(ab)$ jako $\Delta g(\alpha, \beta)$ bez użycia członów z kowariancjami.

Możemy też odwrócić transformację:

$$\begin{pmatrix} a \\ b \end{pmatrix} = P \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

czyli $a = 0.15\alpha - 0.99\beta$, $b = 0.99\alpha + 0.15\beta$. Wtedy chcąc obliczyć np. $\Delta(ab)$ można wyrazić $ab = g(\alpha, \beta) = (0.15\alpha - 0.99\beta)(0.99\alpha + 0.15\beta) = 0.1485(\alpha^2 - \beta^2) - 0.9576\alpha\beta$, co pozwala obliczyć $\Delta(ab)$ jako $\Delta g(\alpha, \beta)$ bez użycia członów z kowariancjami.

Podkreślmy: to nie usuwa korelacji między a i b , tylko obraca układ odniesienia (wybiera bazę) do takiego, w którym macierz kowariancji staje się diagonalna.

Możemy też odwrócić transformację:

$$\begin{pmatrix} a \\ b \end{pmatrix} = P \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

czyli $a = 0.15\alpha - 0.99\beta$, $b = 0.99\alpha + 0.15\beta$. Wtedy chcąc obliczyć np. $\Delta(ab)$ można wyrazić $ab = g(\alpha, \beta) = (0.15\alpha - 0.99\beta)(0.99\alpha + 0.15\beta) = 0.1485(\alpha^2 - \beta^2) - 0.9576\alpha\beta$, co pozwala obliczyć $\Delta(ab)$ jako $\Delta g(\alpha, \beta)$ bez użycia członów z kowariancjami.

Podkreślmy: to nie usuwa korelacji między a i b , tylko obraca układ odniesienia (wybiera bazę) do takiego, w którym macierz kowariancji staje się diagonalna.

Zastosowanie: to jest dokładnie to co leży u podstaw *analizy głównych składowych* (principal component analysis, PCA)—algorytmu redukcji wielowymiarowych przestrzeni parametrów do przestrzeni o niższym wymiarze, zachowującej maksymalną wariancję zawartą w oryginalnych parametrach.

Przykład. Niepewność średniej $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$:

$$\begin{aligned}\Delta \bar{x} &= \sqrt{\sum_{i=1}^n \left(\frac{\partial \bar{x}}{\partial x_i} \right)^2 (\Delta x_i)^2} = \sqrt{\sum_{i=1}^n \left(\frac{1}{n} \cdot 1 \right)^2 (\Delta x_i)^2} \\ &= \sqrt{\frac{1}{n^2} \sum_{i=1}^n (\Delta x_i)^2} = \frac{1}{n} \sqrt{\sum_{i=1}^n (\Delta x_i)^2}\end{aligned}$$

Przykład. Niepewność średniej $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$:

$$\begin{aligned}\Delta\bar{x} &= \sqrt{\sum_{i=1}^n \left(\frac{\partial\bar{x}}{\partial x_i}\right)^2 (\Delta x_i)^2} = \sqrt{\sum_{i=1}^n \left(\frac{1}{n} \cdot 1\right)^2 (\Delta x_i)^2} \\ &= \sqrt{\frac{1}{n^2} \sum_{i=1}^n (\Delta x_i)^2} = \frac{1}{n} \sqrt{\sum_{i=1}^n (\Delta x_i)^2}\end{aligned}$$

Jeśli $\forall i : \Delta x_i \equiv \Delta x$, to

$$\Delta\bar{x} = \frac{1}{n} \sqrt{\sum_{i=1}^n (\Delta x)^2} = \frac{\sqrt{n}\Delta x}{n} = \frac{\Delta x}{\sqrt{n}}$$

(Por. slajd 13, [wykład 8.](#))

Przykład. Niepewność średniej ważonej $\bar{x} = \sum_{i=1}^n x_i w_i$ z wagami

$$w_i \geq 0 : \sum_{i=1}^n w_i = 1:$$

$$\Delta \bar{x} = \sqrt{\sum_{i=1}^n \left(\frac{\partial \bar{x}}{\partial x_i} \right)^2 (\Delta x_i)^2} = \sqrt{\sum_{i=1}^n w_i^2 (\Delta x_i)^2}$$

Przykład. Niepewność średniej ważonej $\bar{x} = \sum_{i=1}^n x_i w_i$ z wagami

$$w_i \geq 0 : \sum_{i=1}^n w_i = 1:$$

$$\Delta \bar{x} = \sqrt{\sum_{i=1}^n \left(\frac{\partial \bar{x}}{\partial x_i} \right)^2 (\Delta x_i)^2} = \sqrt{\sum_{i=1}^n w_i^2 (\Delta x_i)^2}$$

Dla średniej ważonej odwrotnościami wariancji, tj. $w_i = \frac{\sigma_i^{-2}}{\sum_{j=1}^n \sigma_j^{-2}}$ (żeby

nadal zachodziło $\sum_{i=1}^n w_i = 1$):

$$\bar{x} = \frac{\sum_{i=1}^n x_i \sigma_i^{-2}}{\sum_{j=1}^n \sigma_j^{-2}}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i \sigma_i^{-2}}{\sum_{j=1}^n 1 \sigma_j^{-2}}$$

Wtedy prawo propagacji błędów daje:

$$\Delta \bar{x} = \sqrt{\sum_{i=1}^n \left(\frac{\sigma_i^{-2}}{\sum_{j=1}^n \sigma_j^{-2}} \right)^2 (\Delta x_i)^2} \stackrel{\sigma_i = \Delta x_i}{=} \sqrt{\sum_{i=1}^n \frac{\left(\frac{\Delta x_i}{(\Delta x_i)^2} \right)^2}{\left(\sum_{j=1}^n (\Delta x_j)^{-2} \right)^2}} = \sqrt{\frac{\sum_{i=1}^n (\Delta x_i)^{-2}}{\left(\sum_{j=1}^n (\Delta x_j)^{-2} \right)^2}} = \left(\sum_{j=1}^n (\Delta x_j)^{-2} \right)^{-1/2}$$

Wreszcie, jeśli $\forall i : \Delta x_i \equiv \Delta x$, to

$$\Delta \bar{x} = \left(\sum_{j=1}^n \frac{1}{(\Delta x)^2} \right)^{-1/2} = \left(\frac{n}{(\Delta x)^2} \right)^{-1/2} = \sqrt{\frac{(\Delta x)^2}{n}} = \frac{\Delta x}{\sqrt{n}}$$

Wreszcie, jeśli $\forall i : \Delta x_i \equiv \Delta x$, to

$$\Delta \bar{x} = \left(\sum_{j=1}^n \frac{1}{(\Delta x)^2} \right)^{-1/2} = \left(\frac{n}{(\Delta x)^2} \right)^{-1/2} = \sqrt{\frac{(\Delta x)^2}{n}} = \frac{\Delta x}{\sqrt{n}}$$

czyli to samo (por. slajd 20) co dla zwykłej średniej—która jest średnią ważoną z wagami $\forall i : w_i = w \Rightarrow \bar{x} = \sum_{i=1}^n w_i x_i = w \sum_{i=1}^n x_i$, gdzie $\sum_{i=1}^n w_i = nw \equiv 1 \Rightarrow w = \frac{1}{n}$, zatem $\sum_{i=1}^n w_i x_i = \frac{1}{n} \sum_{i=1}^n x_i$.