Poprzednim razem

- Metoda największej wiarygodności

Funkcja wiarygodności

$$\tilde{L}(\theta_1,\theta_2,\ldots,\theta_p)=\prod_{i=1}^{n} f(x_i;\theta_1,\theta_2,\ldots,\theta_p)$$

$$L(\theta_1,\theta_2,\ldots,\theta_p)=\ln\left(\tilde{L}(\theta_1,\theta_2,\ldots,\theta_p)\right)=\ln\left(\prod_{i=1}^{n} f(x_i;\theta_1,\theta_2,\ldots,\theta_p)\right)$$

Szukamy maksimum tej funkcji w przestrzeni parametrów

- Metoda najmniejszych kwadratów

$$Q^2=\sum_{i=1}^{n} w_i(y_i-f(x_i,\vec{\theta}))^2$$

Regresja I rodzaju E(Y|X)

Regresja II rodzaju

$$Q^2=(\vec{Y}-A\vec{\theta})^T B(\vec{Y}-A\vec{\theta})$$

Rozwiązanie (minimum Q²): $\vec{\theta}=(A^T BA)^{-1}A^T B\vec{Y}$

Przykład: regresja liniowa zwyczajna

- regresja liniowa ($f_1(x_i)=x_i$, $f_2(x_i)=1$); zakładamy $\sigma_1=\sigma_2=\ldots=\sigma_n\equiv\sigma_y$

$$f_1(x_i)=x_i \quad f_2(x_i)=1$$

$$y_i=\theta_1 x_i+\theta_2$$

$$\vec{\theta}=(A^T BA)^{-1}A^T B\vec{Y} \qquad \vec{Y}=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \vec{\theta}=\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

$$A_{ij}=\begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ f_1(x_2) & f_2(x_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ f_1(x_n) & \cdots & \cdots & f_m(x_n) \end{pmatrix} \rightarrow \underset{(n\times 2)}{A_{ij}}=\begin{pmatrix} x_1,1 \\ x_2,1 \\ \vdots\ \vdots \\ x_n,1 \end{pmatrix}$$

$$B_{ij}=\delta_{i,j}\frac{1}{(\sigma_i)^2}=\begin{pmatrix} (\sigma_1)^{-2} & 0 & \cdots & 0 \\ 0 & (\sigma_2)^{-2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & (\sigma_n)^{-2} \end{pmatrix} \rightarrow \underset{(n\times n)}{B_{ij}}=\frac{1}{\sigma_y^2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \vdots \\ 0 & 0 & \ddots & \\ 0 & \ldots & 0 & 1 \end{pmatrix}$$

Przykład: regresja liniowa zwyczajna

$$\vec{\theta}=(A^T BA)^{-1}A^T B\vec{Y}$$

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}=\left[\begin{pmatrix} x_1 & x_2\cdots \\ 1 & 1\cdots \end{pmatrix}\begin{pmatrix} x_1,1 \\ \vdots \\ x_n,1 \end{pmatrix}\right]^{-1}\begin{pmatrix} x_1,\ldots \\ 1\ldots \end{pmatrix}\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

W macierzy do odwracania: macierz jednostkową pomijamy

$$\begin{pmatrix} x_1 & x_2\cdots & \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} x_1,1 \\ \vdots \\ x_n,1 \end{pmatrix}=\begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}=\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$(A^T BA)^{-1}=\frac{\sigma_y^2}{\det A}\begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix}=\frac{\sigma_y^2}{n\sum x_i^2-(\sum x_i)^2}\begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix}$$

$$\det A^T 1A$$

Przykład: regresja liniowa zwyczajna

$$\vec{\theta}=(A^T BA)^{-1}A^T B\vec{Y}$$

Zatem $(A^T BA)^{-1}A^T B=\frac{\sigma_y^2}{n\sum x_i^2-(\sum x_i)^2}\begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix}\begin{pmatrix} x_1,\ldots & \\ 1\ldots & 1 \end{pmatrix}\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}=$

$$=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{pmatrix} nx_1-\sum x_i, & \ldots & nx_n-\sum x_i \\ -x_1\sum x_i+\sum x_i^2, & \ldots & {}_n\sum x_i+\sum x_i^2 \end{pmatrix}$$

i ostatecznie

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{pmatrix} nx_1-\sum x_i, & \ldots & nx_n-\sum x_i \\ -x_1\sum x_i+\sum x_i^2, & \ldots & {}_n\sum x_i+\sum x_i^2 \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{pmatrix} nx_1y_1-y_1\sum x_i+nx_2y_2-y_2\sum x_i+ & \ldots & -y_n\sum x_i \\ -x_1y_1\sum x_i+y_1\sum x_i^2-x_2y_2\sum x_i+y_2\sum x_i^2- & \ldots & \sum x_i+y_n\sum x_i^2 \end{pmatrix}=$$

$$=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{pmatrix} n\sum x_iy_i-\sum x_i\sum y_i \\ -\sum x_i\sum x_iy_i+\sum x_i^2\sum y_i \end{pmatrix}$$

Regresja krzywoliniowa – przykład liniowy

$$\vec{\theta}=(A^T BA)^{-1}A^T B\vec{Y}$$

- Wniosek: liniowy związek Y i Θ.
- Współczynniki Θ łatwe do wyliczenia (odwracanie macierzy)
- Przykład: regresja liniowa (m=2, $f_1(x_i)=x_i$, $f_2(x_i)=1$), dodatkowo zakładamy $\sigma_1=\sigma_2=\ldots=\sigma_n\equiv\sigma_y$

$$Y=\theta_1 f_1(X)+\theta_2 f_2(X)=\theta_1 X+\theta_2$$

$$\rightarrow \boxed{\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}=\frac{1}{n\sum_{i=1}^{n} x_i^2-\left(\sum_{i=1}^{n} x_i\right)^2}\begin{pmatrix} n\sum_{i=1}^{n} x_iy_i-\sum_{i=1}^{n} x_i\sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i^2\sum_{i=1}^{n} y_i-\sum_{i=1}^{n} x_i\sum_{i=1}^{n} x_iy_i \end{pmatrix}}$$

- Jest to **regresja zwyczajna**.
- Istnieją też inne odmiany regresji: **klasyczna** (nic nie wiemy o σ_i), **ważona** (różne σ_i), **efektywna** (uwzględnia także niepewność x_i).

Prawo przenoszenia błędów

- Dla liniowych związków pomiędzy Y i Θ znajdujemy macierz kowariancji C(Θ) (ćwiczenia)

$$\vec{\theta}=(A^T BA)^{-1}A^T B\vec{Y}\equiv T\vec{Y}$$

$$\theta_i=T_{ij}Y_j$$

$$E(\theta_i)=T_{ij}\,E(Y_j)$$

$$\boxed{C_{ij}(\vec{\theta})\equiv Cov(\theta_i,\theta_j)=\sum_{k=1}^{n}\sum_{l=1}^{n} T_{ik}T_{jl}Cov(Y_k,Y_l)}$$

$$C(\vec{\theta})=TC(\vec{Y})T^T$$

Prawo przenoszenia błędów

- W naszym przykładzie

$$C(\vec{Y}) = B^{-1} \quad \rightarrow$$

$$C(\vec{\theta}) = (A^T BA)^{-1} A^T BB^{-1}\left[(A^T BA)^{-1} A^T B\right]^T =$$

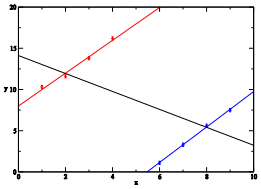
$$= (A^T BA)^{-1} A^T \left[A^{-1} B^{-1} (A^T)^{-1} A^T B\right]^T =$$

$$= (A^T BA)^{-1} A^T (A^{-1})^T = (A^T BA)^{-1} A^T (A^T)^{-1} = (A^T BA)^{-1}$$

$$C_{ij}(\vec{\theta}) = \begin{pmatrix} var(\theta_1) & cov(\theta_1, \theta_2) \\ cov(\theta_1, \theta_2) & var(\theta_2) \end{pmatrix} =$$

$$= \sigma_y^2 \left[A^T A\right]^{-1} = \frac{\sigma_y^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \begin{pmatrix} n & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

Paradoks Simpsona



y=1.99x+8.0

y=2.15x-11.75

y= -1.0883x+14.117

- Paradoks Simpsona – połączenie różnych grup może zmienić interpretację wyników. Prowadzi to do praktycznych problemów np. związanych z wyborem kuracji osoby chorej. Rozwiązanie problemu prowadzi do sieci Bayesowskich i teorii grafów i polega na prześledzeniu zależności pomiędzy zmiennymi X, Y i innymi zmiennymi, które mogą mieć wpływ na zmienne X i Y.

Miara jakości dopasowania – „test Chi-kwadrat"

Zał. σ_i mają rozkład normalny

$$M = \sum_{i=1}^{n} \frac{(y_i - \sum_{j=1}^{p} \theta_j f_j(x_i))^2}{\sigma_i^2}$$

- E(M)=n-p – dla rozkładu Chi-kwadrat o n-p stopniach swobody.
- Gdy $M > \chi_\gamma^{2,n-p}$ (n- liczba pomiarów, p liczba parametrów) to na poziomie ufności γ odrzucamy wynik dopasowania.
- Dla regresji liniowej mamy (p=2)

$$M = \sum_{i=1}^{n} \frac{(y_i - ax_i - b)^2}{\sigma_i^2}$$

- W praktyce sprawdzamy czy M/(n-2)≈1 – prosta dobrze przybliża.
- M/(n-2)<1 prosta, ale za duże σ_i; M/(n-2)>1 inna krzywa, za małe σ_i lub niektóre pomiary odstają od prostej.

Hipotezy statystyczne

- Testowanie hipotez statystycznych pozwala na sprawdzenie na podstawie wyników próby, przy zadanym poziomie ufności, czy jakieś twierdzenie (hipotezę) dotyczące populacji generalnej jest prawdziwe. Taką procedurę nazywamy **testem statystycznym**.
- Przykłady:
  1) Czy próbka pochodzi z rozkładu normalnego?
  2) Czy na podstawie próby można powiedzieć, że wartość oczekiwana jest większa niż pewna wybrana liczba?
  3) Czy wariancje dwóch rozkładów, z których mamy dwie różne próbki, są sobie równe czy nie?
- Jak widać mamy zawsze do czynienia z pewną **hipotezą zerową** H_0, którą sprawdzamy i **hipotezą alternatywną** H_1, którą przyjmiemy gdy w wyniku testu odrzucamy H_0.
- **Hipotezy** można podzielić też na **proste** (jednoznacznie określają rozkład, funkcję gęstości prawdopodobieństwa lub dystrybuantę zmiennej losowej) i **złożone** (pozostałe).
- Inny podział: **parametryczna** i **nieparametryczna**.

Testy statystyczne

- Schemat postępowania:
  1) Sformułowanie hipotezy zerowej H_0.
  2) Określenie rozmiaru próbki n.
  3) Ustalamy **poziom istotności α**, zwany również **poziomem błędu testu**. Jest to prawdopodobieństwo popełnienia **błędu I-go rodzaju**, a więc odrzucenia, w wyniku testu, hipotezy H_0, podczas gdy jest ona w rzeczywistości prawdziwa. Zazwyczaj α<0.1.
  4) Sformułowanie hipotezy alternatywnej H_1.
  5) Zakładamy chwilowo, że H_0 jest prawdziwa, i wybieramy statystykę testową δ, zależną od próby, o znanym rozkładzie prawdopodobieństwa.
  6) Wyznaczmy **obszar krytyczny W** dla wartości δ, zależny od α, H_1 i n, w taki sposób, że prawdopodobieństwo, że $\delta \in W$ wynosi α.
  7) Wyliczamy, otrzymaną na podstawie pobranej próby, wartość statystyki δ. **Gdy należy ona do obszaru krytycznego W to odrzucamy H_0, w przeciwnym razie twierdzimy, że nie ma podstaw do odrzucenia H_0.**

Testy statystyczne - uwagi

- Określenie obszaru krytycznego na podstawie α jest niejednoznaczne. Można to poprawić wprowadzając prawdopodobieństwo **błędu II-go rodzaju**: przyjęciu fałszywej H_0, gdy w rzeczywistości prawdziwa jest hipoteza alternatywna H_1. Prawdopodobieństwo to oznaczamy β, a 1-β nazywamy **mocą testu**.
- Jeżeli znamy $f(\delta|H_0)$ (α) i $f(\delta|H_1)$ (β) to obszar krytyczny wyznaczamy jako zakres wartości δ, który minimalizuje β przy zadanym α. Zazwyczaj znamy tylko $f(\delta|H_0)$ i nie znamy mocy testu co oznacza, że nie możemy twierdzić, że H_0 jest prawdziwa, a tylko, że nie odrzucamy H_0. Test statystyczny, przy nieznajomości β, nazywamy **testem istotności**.
- W praktyce opracowane są zestawy statystyk testowych i zakresy obszaru krytycznego dla różnych rodzajów hipotez H_0 i H_1. W większości testów zakładamy, że próbka pochodzi z rozkładu normalnego.