Modell illeszkedés II.

Kormányos Andor

Komplex Rendszerek Fizikája Tanszék

2022. február 7.

χ^2 eloszlás módszer

- $\chi^2(\mathbf{a}; x^{(i)}, y^{(i)})$ legkisebb négyzetes költségfüggvény minimalizálásával kapunk egy $\mathbf{a}_{fit}$ paramétervektort
- ha a mérést többször megismételnénk és mindig kiszámolnánk a $\chi^2(\mathbf{a}_{fit})$-t az adott mérési értékekkel, akkor a $\chi^2(\mathbf{a}_{fit})$-ekre kapnánk egy eloszlást

Bizonyítás nélkül: ha a modell lineárisan függ $a_0, a_1, \dots a_M$-tól, akkor a különböző $\chi^2(\mathbf{a}; x^{(i)}, y^{(i)})$ értékek eloszlása az összeg minimuma körül $\nu = N - M$ szabadsági fokú χ^2_ν eloszlást követ. Itt N a mérési pontok, és M az illesztett paraméterek száma.

- pl logisztikus regresszió esetén nem alkalmazható

χ^2_ν eloszlás

Figyelem: Ne keverjük össze a mérések és a modell segítségével számolt $\chi^2(\mathbf{a}; x^{(i)}, y^{(i)})$ összeget és a χ^2_ν eloszlást!

Definíció: ha $\xi_1, \xi_2, \dots \xi_\nu$ valószínűségi változók standard normál eloszlással, akkor $\xi_1^2 + \xi_2^2 + \dots + \xi_\nu^2$ eloszlása ν szabadsági fokú χ^2 eloszlás

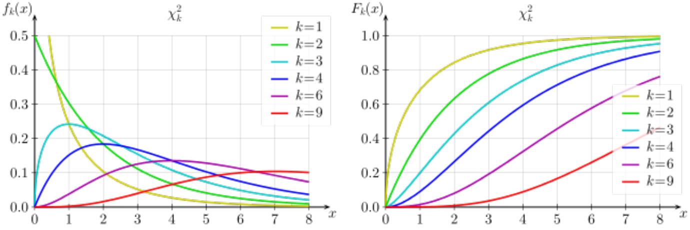

Figure: Különböző szabadsági fokú χ^2_ν sűrűség és eloszlás függvények

Mennyire illeszkedik jól a modell?

Megjegyzés:

- az $a_0, a_1, \ldots a_M$ értékeket úgy határoztuk meg, hogy χ^2 költségfüggény minimális legyen
- ezért $\sum_i \frac{\left[y^{(i)} - h(\mathbf{a}; x^{(i)})\right]^2}{\sigma_i^2}$-ben nem minden tag független
- $\Rightarrow \nu = N - M$ szabadsági fok

Mennyire illeszkedik jól a modell?

Megjegyzés:

- az $a_0, a_1, \ldots a_M$ értékeket úgy határoztuk meg, hogy χ^2 költségfüggény minimális legyen
- ezért $\sum_i \frac{\left[y^{(i)} - h(\mathbf{a}; x^{(i)})\right]^2}{\sigma_i^2}$-ben nem minden tag független
- $\Rightarrow \nu = N - M$ szabadsági fok

Mikor fogadhatjuk el a modellt?

- leosztjuk $\chi^2(\mathbf{a}_{fit}; x^{(i)}, y^{(i)})$-et $\nu = N - M$-mel
- a modell elfogadhatóan illeszkedik, ha $\frac{\chi^2(\mathbf{a}_{fit}; x^{(i)}, y^{(i)})}{\nu} \approx 1$
- pontosabban: $\chi^2(\mathbf{a}_{fit}; x^{(i)}, y^{(i)})$ statisztikus átlaga ν, ekörül a szórása $\sqrt{2\nu}$.

Példa: parabola illesztése 5 pontra

Emlékeztető: egy korábbi példában azt kaptuk, hogy:

$$\mathbf{a}_{fit} = \begin{bmatrix} 1.049 \\ -0.020 \\ 0.986 \end{bmatrix}$$

Az illesztés jósága ($\sigma_i = 0.1$):

$$\begin{aligned} \chi^2 &= (\mathbf{X}\mathbf{a}_{fit} - \mathbf{y}/\sigma_i)^2 \\ &= 4.11 \end{aligned}$$

$$\frac{\chi^2}{\nu} = \frac{4.11}{5-3} = 2.06$$

Ez még a $\sqrt{2\nu} = 2$ szóráson belül van.