

GA(M)M: Alapozó

Márton Sóskuthy, 2017: Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction.

<https://arxiv.org/pdf/1703.05339.pdf>

Martijn Wieling, 2018: Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English.

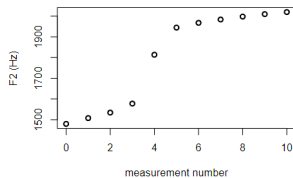
Journal of Phonetics 70, 86-116, ISSN 0095-4470,

<https://doi.org/10.1016/j.wocn.2018.03.002>

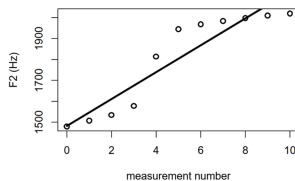
Generalized Additive (Mixed) Model
=
Általánosított additív (kevert) modell

- általánosított: a függő változót egy kapcsolati függvénnel (link function) köti a független változóhoz → magyarul nem csak lineáris kapcsolat adható meg
 - ilyen más stat. módszer is tud, de ott előre tudni kell, milyen összefüggést állítunk be, itt pedig ő kidolgozza
 - az általános része az általánosítottnak, nem összekeverendő

Mi ez? II

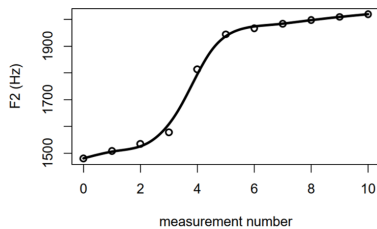


Adatok



Lineáris modell: $y = \alpha + \beta x_n$

Mi ez? III

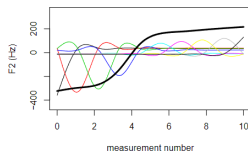
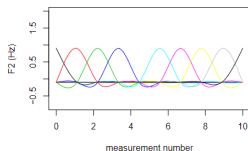


$$\text{GAM: } y = \alpha + f(x)$$

Mi ez? IV

- additív: az ún. kapcsolati függvény létrejöhet több függvény összeadásából, azaz "kikeveri" a végső összefüggés függvényét:

$$y = \alpha + \beta_1 f_1(x) + \beta_2 f_2(x) + \dots + \beta_n f_n(x)$$



Függvények összeadása

- (kevert): a szokásos, beépíthető random hatás
- modell: :o)

Mi ez? V

Hogyan adódik ez össze, mi a speckó?

- Dinamikus elemzések, pl.:
 - beszédbeni időben: formánskontúr, f_0 -görbe
 - létidőben: diakrónikus vagy longitudinális változás
 - távolságban: dialektus térképen elhelyezett értékek
 - térben: nyelvkontúr
- Skaláris független változó mentén, relatív pontokon \rightarrow normalizált pontokon mérek, azaz bármilyen skálát választok, az a teljes időtartamra/távolságra/időtávra vetemíttetik, pl.
 - magánhangzó időtartama \rightarrow normalizálom $[0, 1]$, $[0, 100]$, $[0, 11]$... értékartományra

Mi ez? VI

Ami egyből tűnjön szembe: a mérési pontok függenek egymástól. → Ezt több modell is tudja kezelni, amiket nem nagyon használunk nyelvészetben, de ez is, amit viszont használunk.

A statisztika az adatok egymástól való függetlenségét feltételezi, mivel az ún. autokorrelációt okoz. (Egy nyelvkontúr egyik pontjának pozíciója függ a másik pontétól, egy magánhangzóban mért formánsérték függ az előző értéktől, egy f_0 -érték függ az előző értéktől, stb. (Mérhetném más embernél a következő pontot, de nagyon macerás lenne, nyelvkontúrnál pl. kivitelezhetetlen is.)

Alapfüggvények (basic functions) I

Az alapfüggvények, amiket összead a modell, hogy elkészítse az adatainkat leíró függvényt. Az illesztett modell az alapfüggvények saját együtthatóikkal felszorzott összege:

$$y = \alpha + \beta_1 f_1(x) + \beta_2 f_2(x) + \dots + \beta_n f_n(x)$$

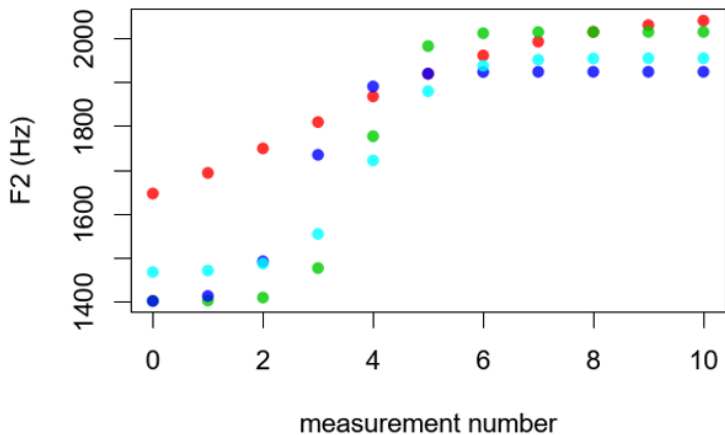
A függvények száma alapból 9, de állítható a knot (jele: k , csomópont?) értékével. Ez a függvények "találkozási pontja"-inak száma, azaz hány ponton kell konvergálni a függvényeknek (szabályos x -tengelyi távolságonként). $k-1$ alapfüggvényt használ, kivéve egy esetben, erre majd visszatérünk.

A k növelésével javítható a modell, de túlilleszteni nem szabad, mert akkor csak arra az adathalmazra lesz igaz az eredmény!!!!

Alapfüggvények (basic functions) II

Nem linearitási büntetés: (non-linearity penalty, m), bünteti a komplexebb függvényeket a túlillesztés elkerülése végett. Nem törődnek vele a nyelvészeti tutoriálok sokat, ahol szükséges, előkerül.

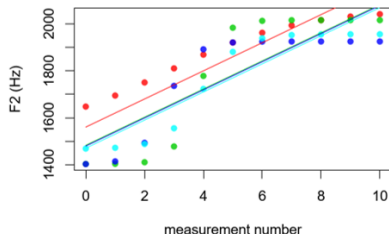
Randomitás I



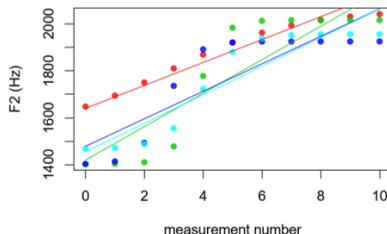
Randomitás II

1. Mi volt a randomitás lényege a lineáris kevert modelleknél?

LMER: random intercepts only



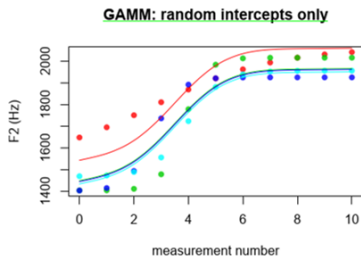
LMER: random intercepts + slopes



LMM: randomitás

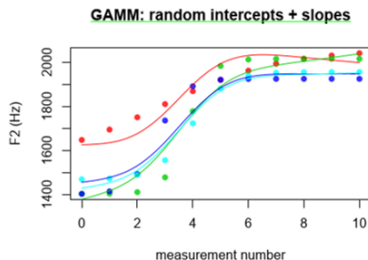
2. Mi lehet random a GAMM-ban?

Randomitás III



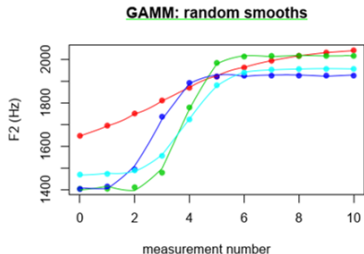
GAMM: random metszéspont (intercept)
Az illesztett görbe eltolása adott faktor értékei között (α változtatása).

Randomitás IV



GAMM: random metszéspont és görbe (intercept, slope)
Az illesztett görbe foroghat, nyúlhat adott faktor értékei között (β -együtthatók változtatása)

Randomitás V



GAMM: random simítás (smooth)

Eltérő függvényekkel is leírható a faktor értékeinként a függőváltozó.

A random smooth esetében az alapfüggvények száma k , nem pedig $k-1$.

Parametric and smooth terms I

(Parametrikus és simítási tagok?)

Feltehető kérdéstípusok:

- Van-e egy skaláris faktor mentén görbével, azaz simítással leírható összefüggés → pl. a magánhangzó időtartama alatt a formánsok értéke konstans, lineárisan vagy (egyéb) görbével írható-e le? → smooth term: $s()$ -ben adjuk meg
- Van-e egy faktor értékei mentén eltérés a mért menet átlagában → pl. az /i/ és az /ε/ 2. formánsa eltér-e? → parametrikus tag: csak beírom a faktort

Parametric and smooth terms II

- Van-e egy faktor értékei között egy skaláris (másik) faktor mentén leírható görbékben eltérés → pl. az az /i/, az /ε/ és /ɒ/ más F_2 -menetet mutat-e az egyes C-kontextusokban → a parametrikus és a simítási tagok interakciója
- Van-e eltérés a mért értékekre illeszthető függvénynek két skaláris faktor mentén → pl. a nyelvkontúr egyes pontjai (tér) és a magánhangzóbeli időpont (idő) hatása az adott nyelvpont pozíciójára → ez is interakció
- Kérdezhető interakció egy folytonos faktor és másik numerikus faktor között → a formáns értéke a magánhangzóbeli időpont és a teljes időtartam függvényében