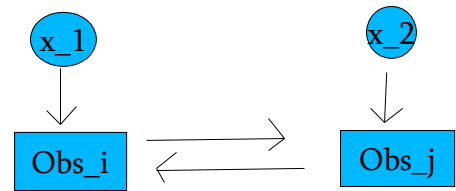


murresanamallista:

$$\text{Obs}_i \in \{0, 1\}$$

$$P(x, \text{obs}) = P(x)P(\text{obs}|x)$$



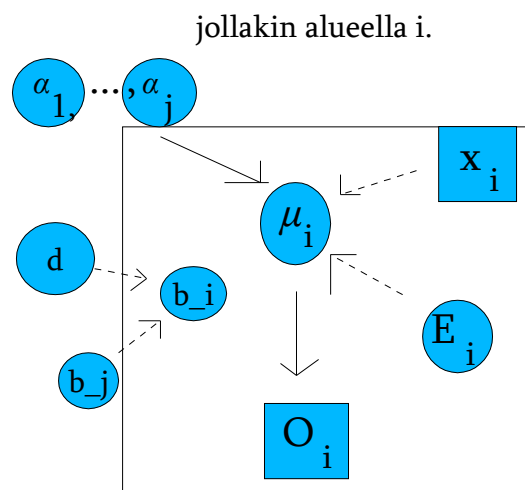
	Obs = 0	Obs = 1
X = 0	1	0
X = 1	a	1-a

$$a \sim \text{Beta}(\alpha, \beta)$$

Paikannimidata:

O_i : sanan osa esiintyy näin monta kertaa

$$O_i \sim \text{Poisson}(\mu_i)$$



Ainoastaan α_i vaikuttaa ruutuun i kaikista α .

$$\log \mu_i = \alpha_i x_i + \log E_i + d b_i$$

$$\alpha_i \sim \text{Norm}(a, b)$$

$$b_i \sim \text{Norm}\left(\frac{1}{n_{i \sim j}} \sum b_j, n_i\right)$$

$$d \sim \text{Gamma}(\cdot, \cdot)$$

$$\alpha_i \sim \text{Norm}(\cdot, \cdot)$$

Jos alueella on N_i paikannimeä yhteensä, tästä voidaan laskea jokin odotusarvo E_i tietyn sanan esiintymiselle. Tämä vaikuttaa μ_i :hin.

Kysymyksiä & Vastauksia vaatimuksiin:

Käyttötapauksista:

Kaikkia muuttujia saatava kytkeä pois.

Esim. Edellisen mallin yhtälöstä voisi kytkeä $d = 0$ kaikilla. Vertailla miten vaikuttaa.

Tai sitten voisi kytkeä $\log E_i$ -termin pois, ja verrata.

Summalausekkeisiin saatava 5-10 lisämuuttujaa, jos muuttujien lisääminen ei kuormita liikaa, ei määrää tarvitse rajoittaa. (olisi kiva saada lisätä niitä niin paljon kuin lystää)

Tulosteista: perustuloste: kaikkien parametrien arvot simulaation aikana: rivejä joilla on jokaisen muuttujan arvo, ja rivejä on monta.

Miten rajoitetaan: Käyttäjä saa kertoa mitkä muuttujat tulostetaan, tulostetaan sellaisenaan, ja lasketaan myöhemmin näistä jotain.

Oltava valinta - otetaanko joka 10, joka 100 parametririvi, koska useasti rivit muuttuvat hitaasti. Nämä tiheydet eri parametreille erikseen, niin että voidaan tulostaa eri aikoihin eri muuttujia.

Burn-in-pituus: kuinka monta otosta ohitetaan aluksi, että päästään oikeaan jakaumaan alkuarvoista.

Ohjelman olisi hyvä tulostaa ruudulle monesko otos on menossa (esim. 10% välein..)

Malli ja sim. tekn. param. vaikka samaan tiedostoon.

Voidaan olettaa vierusmatriisi erilliseksi tiedostoksi.

Mikä on malli? Muuttujien yhteistodennäköisyysjakauman esitys.

-Tästä käsitteellinen malli vaatimusmäärittelyyn olisi hyvä...

Malli pysyy samana datan vaihtuessa. Malli ottaa kantaa dataan vain jättämällä sen jakauman sanomatta, ja odottaa että se tulee jostain.

Halutaanko ohjelman osia käyttää uudelleen?

Halutaanko ohjelmaa laajentaa? - mahdollisesti.

GUI ja rinnakkaisuus optional.

NAG-kirjastosta löytyy jakaumia, etc. (mainintoja fortran-kirjallisuudessa)

Ehkä jokin muu vaihtoehto? Onko yliopistolla lisenssi tähän?

Vaihtuvaa kamaa simulaatioteknisissä:

-iteraatioiden lkm

-sisäänajokierrosten määrä

-ohennus (kuinka harvoin tulostetaan, joka 10s, joka 100s, eri muuttujille erillinen)

-ehdotusjakaumat: eri muuttujille annettava, annettava muuttujajoukoille helposti (jos ei annettu, käytetään mallin priorijakaumaa)

-päivitysstrategiat: päivitetään random muuttujaa, iteraatioiden lkm olisi päivitysten yhteinen

lkm.

-parametreille pitää voida antaa alkuarvot. (jos ei ole annettu, priorijakaumasta arvotaan)

-käyttäjän virheitä ei voida eliminoida... Ei tarvitse tarkistaa ovatko jakaumat järkeviä...

mallin kuvaus tekstitiedostossa voisi olla muotoa:

$O_i \sim \text{Poisson}(u_i)$

$\log u_i = \dots + \dots + \dots$

termien priorijakaumat...

spatiaalisuus on yksi termi näistä.

Jos priorijakaumaa ei ole, muuttujan löydyttävä datasta. Tai voidaan määrittää että mistä datatiedostosta ja miten muuttujan esiintymät löytyvät.