Bayes-päättely
Työterveyslaitos
8.-9.2.2018

Ville Hyvönen

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

CoSCo

Information, Complexity & Learning

2. Priorin valinta & konjugaattipriorit

1. Epäinformatiivinen priori
2. Informatiivinen priori
3. Epäoleellinen priori
4. Konjugaattipriorit
5. Miten valita priorijakauma??

Epäinformatiivinen priori

- **Epäinformatiivisella priorilla** tarkoitetaan priorijakaumaa, joka vaikuttaa mahdollisimman vähän posteriorijakaumaan.

Epäinformatiivinen priori

- **Epäinformatiivisella priorilla** tarkoitetaan priorijakaumaa, joka vaikuttaa mahdollisimman vähän posteriorijakaumaan.
- Tarkoitus antaa "aineiston puhua puolestaan."

Epäinformatiivinen priori

- **Epäinformatiivisella priorilla** tarkoitetaan priorijakaumaa, joka vaikuttaa mahdollisimman vähän posteriorijakaumaan.
- Tarkoitus antaa "aineiston puhua puolestaan."
- Epäinformatiivista prioria käytetään:

Epäinformatiivinen priori

- **Epäinformatiivisella priorilla** tarkoitetaan priorijakaumaa, joka vaikuttaa mahdollisimman vähän posteriorijakaumaan.
- Tarkoitus antaa "aineiston puhua puolestaan."
- Epäinformatiivista prioria käytetään:
  - Jos meillä ei ole mitään ennakkotietoa mallinnettavasta ilmiöstä, eli kuvaamaan suurta epävarmuutta parametrin todellisesta arvosta.

Epäinformatiivinen priori

- **Epäinformatiivisella priorilla** tarkoitetaan priorijakaumaa, joka vaikuttaa mahdollisimman vähän posteriorijakaumaan.
- Tarkoitus antaa "aineiston puhua puolestaan."
- Epäinformatiivista prioria käytetään:
  - Jos meillä ei ole mitään ennakkotietoa mallinnettavasta ilmiöstä, eli kuvaamaan suurta epävarmuutta parametrin todellisesta arvosta.
  - Tai jos halutaan, että ennakkokäsityksemme tutkittavasta ilmiöstä vaikuttavat päätelmiimme aineistosta mahdollisimman vähän $\rightarrow$ esimerkiksi jos halutaan testata antaako uusi tutkimus samankaltaisia tuloksia kuin vanhat.

Informatiivinen priori

- **Informatiivisella priorilla** tarkoitetaan priorijakaumaa, jonka tarkoitus on hyödyntää etukäteistietoa (esim. aiempia tutkimuksia) mallinnettavasta ilmiöstä, ja vaikuttaa selkeästi posteriorijakaumaan pienellä otoskoolla.

Informatiivinen priori

- **Informatiivisella priorilla** tarkoitetaan priorijakaumaa, jonka tarkoitus on hyödyntää etukäteistietoa (esim. aiempia tutkimuksia) mallinnettavasta ilmiöstä, ja vaikuttaa selkeästi posteriorijakaumaan pienellä otoskoolla.
- Informatiivista prioria käytetään:

Informatiivinen priori

- **Informatiivisella priorilla** tarkoitetaan priorijakaumaa, jonka tarkoitus on hyödyntää etukäteistietoa (esim. aiempia tutkimuksia) mallinnettavasta ilmiöstä, ja vaikuttaa selkeästi posteriorijakaumaan pienellä otoskoolla.
- Informatiivista prioria käytetään:
  - Jos halutaan järkevä estimaatti pienelläkin otoskoolla

Informatiivinen priori

- **Informatiivisella priorilla** tarkoitetaan priorijakaumaa, jonka tarkoitus on hyödyntää etukäteistietoa (esim. aiempia tutkimuksia) mallinnettavasta ilmiöstä, ja vaikuttaa selkeästi posteriorijakaumaan pienellä otoskoolla.
- Informatiivista prioria käytetään:
  - Jos halutaan järkevä estimaatti pienelläkin otoskoolla
  - Esim. vakuutusala, vedonlyönti, tekoälysovellukset.

Tasajakauma priorina

- Tasajakauma $U(0, 1)$ kuvaa suurinta mahdollista epävarmuutta parametrin arvosta?

Tasajakauma priorina

- Tasajakauma $U(0, 1)$ kuvaa suurinta mahdollista epävarmuutta parametrin arvosta?
  - Kaikkia parametrin arvoja pidetään etukäteen yhtä todennäköisinä.

Tasajakauma priorina

- Tasajakauma $U(0,1)$ kuvaa suurinta mahdollista epävarmuutta parametrin arvosta?
  - Kaikkia parametrin arvoja pidetään etukäteen yhtä todennäköisinä.
  - 2 pseudohavaintoa, joista 1 onnistuminen ja 1 epäonnistuminen $\rightarrow$ tasajakauma ei olekaan täysin epäinformatiivinen?

Tasajakauma priorina

- Tasajakauma $U(0,1)$ kuvaa suurinta mahdollista epävarmuutta parametrin arvosta?
  - Kaikkia parametrin arvoja pidetään etukäteen yhtä todennäköisinä.
  - 2 pseudohavaintoa, joista 1 onnistuminen ja 1 epäonnistuminen $\rightarrow$ tasajakauma ei olekaan täysin epäinformatiivinen?
- Priorin muoto riippuu parametrisoinnista: esimerkiksi jos parametrille tehdään ns. log-odds-muunnos:

$$\phi = \log \frac{\theta}{1-\theta},$$

ϕ ei olekaan enää tasaisesti jakautunut, vaikka θ olisi.

Tasajakauma priorina

- Tasajakauma $U(0,1)$ kuvaa suurinta mahdollista epävarmuutta parametrin arvosta?
  - Kaikkia parametrin arvoja pidetään etukäteen yhtä todennäköisinä.
  - 2 pseudohavaintoa, joista 1 onnistuminen ja 1 epäonnistuminen $\rightarrow$ tasajakauma ei olekaan täysin epäinformatiivinen?
- Priorin muoto riippuu parametrisoinnista: esimerkiksi jos parametrille tehdään ns. log-odds-muunnos:

$$\phi = \log \frac{\theta}{1-\theta},$$

  ϕ ei olekaan enää tasaisesti jakautunut, vaikka θ olisi.
- Tasajakauma rajaamattomalle parametrille, esim. normaalijakauman odotusarvolle $\mu \in \mathbb{R}$?

Priorin informaatio

- Kun priori on tasajakauma, posteriorijakauman $\mathrm{Beta}(y+1, n-y+1)$ odotusarvo on

$$\hat{\theta}_{\mathrm{Bayes}} = \frac{y+1}{n+2}.$$

Priorin informaatio

- Kun priori on tasajakauma, posteriorijakauman $\mathrm{Beta}(y+1, n-y+1)$ odotusarvo on
$$\hat{\theta}_{\text{Bayes}} = \frac{y+1}{n+2}.$$
- Jos havaitaan $y = 3$ kruunaa $n = 3$:lla heitolla, piste-estimaatti kruunan todennäköisyydelle on
$$\hat{\theta}_{\text{Bayes}} = \frac{3}{4}.$$

Priorin informaatio

- Kun priori on tasajakauma, posteriorijakauman $\mathrm{Beta}(y+1, n-y+1)$ odotusarvo on

$$\hat{\theta}_{\mathrm{Bayes}} = \frac{y+1}{n+2}.$$

- Jos havaitaan $y = 3$ kruunaa $n = 3$:lla heitolla, piste-estimaatti kruunan todennäköisyydelle on

$$\hat{\theta}_{\mathrm{Bayes}} = \frac{3}{4}.$$

- Tasajakauma $U(0,1)$ ei siis ollutkaan täysin epäinformatiivinen?

Kolikonheittoesimerkki

- Asetetaan priorijakauman parametreiksi esimerkiksi $\alpha = \beta = 0.01$

Kolikonheittoesimerkki

- Asetetaan priorijakauman parametreiksi esimerkiksi $\alpha = \beta = 0.01$
- Posteriorijakauman Beta$(y + 0.01, n - y + 0.01)$ odotusarvo:

$$\hat{\theta}_{\text{Bayes}} = \frac{y + 0.01}{n + 0.01} \approx \frac{y}{n}.$$

Epäoleellinen priori

- Voidaanko asettaa $\alpha = 0, \beta = 0$?
  - Kaikkein epäinformatiivisin priori?

Epäoleellinen priori

- Voidaanko asettaa $\alpha = 0, \beta = 0$?
  - Kaikkein epäinformatiivisin priori?
  - Betajakauman parametreille on pädettävä $\alpha > 0, \beta > 0$, koska $B(0,0)$ ei ole määritelty $\rightarrow$ Beta$(0,0)$ ei ole todennäköisyysjakauma!

Epäoleellinen priori

- Voidaanko asettaa $\alpha = 0, \beta = 0$?
  - Kaikkein epäinformatiivisin priori?
  - Betajakauman parametreille on pädettävä $\alpha > 0, \beta > 0$, koska $B(0,0)$ ei ole määritelty $\rightarrow$ Beta$(0,0)$ ei ole todennäköisyysjakauma!
- Kuitenkin: voidaan käyttää priorina, jos posteriori on kunnollinen tn-jakauma!

Epäoleellinen priori

- Voidaanko asettaa $\alpha = 0, \beta = 0$?
  - Kaikkein epäinformatiivisin priori?
  - Betajakauman parametreille on pädettävä $\alpha > 0, \beta > 0$, koska $B(0,0)$ ei ole määritelty $\rightarrow$ Beta$(0,0)$ ei ole todennäköisyysjakauma!
- Kuitenkin: voidaan käyttää priorina, jos posteriori on kunnollinen tn-jakauma!
  - Tällöin puhutaan *epäoleellisesta priorista* (improper prior)

Epäoleellinen priori

- Voidaanko asettaa $\alpha = 0, \beta = 0$?
  - Kaikkein epäinformatiivisin priori?
  - Betajakauman parametreille on pädettävä $\alpha > 0, \beta > 0$, koska $B(0,0)$ ei ole määritelty $\rightarrow$ Beta$(0,0)$ ei ole todennäköisyysjakauma!
- Kuitenkin: voidaan käyttää priorina, jos posteriori on kunnollinen tn-jakauma!
  - Tällöin puhutaan *epäoleellisesta priorista* (improper prior)
  - Käytetään verrannollisuusmerkintää:

$$p(\theta) \propto \theta^{-1}(1-\theta)^{-1}.$$

Epäoleellinen priori

- Epäoleellinen priori voidaan sijoittaa Bayesin kaavaan normaalisti priorin paikalle:

$$
\begin{aligned}
p(\theta) &\propto p(\theta)p(y|\theta) \\
&\propto \theta^{-1}(1-\theta)^{-1} \cdot \theta^{y}(1-\theta)^{n-y} \\
&= \theta^{y-1}(1-\theta)^{n-y-1}
\end{aligned}
$$

Epäoleellinen priori

- Epäoleellinen priori voidaan sijoittaa Bayesin kaavaan normaalisti priorin paikalle:

$$\begin{aligned} p(\theta) &\propto p(\theta)p(y|\theta) \\ &\propto \theta^{-1}(1-\theta)^{-1} \cdot \theta^{y}(1-\theta)^{n-y} \\ &= \theta^{y-1}(1-\theta)^{n-y-1} \end{aligned}$$

- Posterioriksi saadaan siis

$$\theta|y \sim \text{Beta(y, n-y)}.$$

Epäoleellinen priori

- Entä jos $y = 0$ tai $y = n$?

Epäoleellinen priori

- Entä jos $y = 0$ tai $y = n$?
- Epäoleellista prioria käytettäessä pitää aina muistaa tarkastaa, että posteriori on kunnollinen todennäköisyysjakauma!

Epäoleellinen priori

- Entä jos $y = 0$ tai $y = n$?
- Epäoleellista prioria käytettäessä pitää aina muistaa tarkastaa, että posteriori on kunnollinen todennäköisyysjakauma!
- Turvallisinta käyttää kunnollista prioria
  - Posteriorijakauma tällöin aina kunnollinen.

Epäoleellinen priori

- Epäoleelliset priorit saadaan usein raja-arvoina kunnollisista prioreista viemällä niiden parametrit määrittelyjoukkonsa rajoille.

Epäoleellinen priori

- Epäoleelliset priorit saadaan usein raja-arvoina kunnollisista prioreista viemällä niiden parametrit määrittelyjoukkonsa rajoille.
- Entä melkein epäoleelliset priorit, esim. $\mathrm{Beta}(0.001, 0.001)$?

Epäoleellinen priori

- Epäoleelliset priorit saadaan usein raja-arvoina kunnollisista prioreista viemällä niiden parametrit määrittelyjoukkonsa rajoille.
- Entä melkein epäoleelliset priorit, esim. $\mathrm{Beta}(0.001, 0.001)$?
- Tällaisia prioreja, erityisesti $\mathrm{Inv\text{-}Gamma}(0.001, 0.001)$ skaala-parametrille, näkee paljon vanhoissa BUGS-esimerkeissä.

Epäoleellinen priori

- Epäoleelliset priorit saadaan usein raja-arvoina kunnollisista prioreista viemällä niiden parametrit määrittelyjoukkonsa rajoille.
- Entä melkein epäoleelliset priorit, esim. $\mathrm{Beta}(0.001, 0.001)$?
- Tällaisia prioreja, erityisesti $\text{Inv-Gamma}(0.001, 0.001)$ skaala-parametrille, näkee paljon vanhoissa BUGS-esimerkeissä.
- Jos posteriori ei ole kunnollinen epäoleellisella priorilla, melkein epäoleellisella priorilla tulee helposti numeerisia ongelmia.

Konjugaattipriorit

- Jos jollekin otantajakaumalle posteriorijakauma on samaa muotoa kuin priorijakauma, tällöin sanotaan, että tämä priorijakauma on **konjugaattipriori** (conjugate prior) kyseiselle otantajakaumalle.

Konjugaattipriorit

- Jos jollekin otantajakaumalle posteriorijakauma on samaa muotoa kuin priorijakauma, tällöin sanotaan, että tämä priorijakauma on **konjugaattipriori** (conjugate prior) kyseiselle otantajakaumalle.
- Otantajakaumasta ja priorijakaumasta puhutaan tällöin **konjugaattiparina** (conjugate pair).

Konjugaattipriorit

- Jos jollekin otantajakaumalle posteriorijakauma on samaa muotoa kuin priorijakauma, tällöin sanotaan, että tämä priorijakauma on **konjugaattipriori** (conjugate prior) kyseiselle otantajakaumalle.
- Otantajakaumasta ja priorijakaumasta puhutaan tällöin **konjugaattiparina** (conjugate pair).
- Esimerkiksi beta-jakauma on konjugaattipriori binomijakaumalle, koska tälle mallille posteriorijakauma on myös beta-jakauma.

Konjugaattipriorit

- Konjugaattimalleille reunauskottavuus

$$p(\mathbf{y}) = \int_{\Omega} p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}) \mathrm{d}\,\boldsymbol{\theta}.$$

osataan ratkaista analyyttisesti.

Konjugaattipriorit

- Konjugaattimalleille reunauskottavuus

$$p(\mathbf{y}) = \int_{\Omega} p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}) \mathrm{d}\,\boldsymbol{\theta}.$$

  osataan ratkaista analyyttisesti.
  - Posteriorijakauma $p(\boldsymbol{\theta}|\mathbf{y})$ osataan siten ratkaista suljetussa muodossa.

Konjugaattipriorit

- Konjugaattimalleille reunauskottavuus

$$p(\mathbf{y}) = \int_{\Omega} p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}) \mathrm{d}\,\boldsymbol{\theta}.$$

  osataan ratkaista analyyttisesti.
  - Posteriorijakauma $p(\boldsymbol{\theta}|\mathbf{y})$ osataan siten ratkaista suljetussa muodossa.
- Muille kuin konjugaattimalleille reunauskottavuutta ei ole mahdollista ratkaista suljetussa muodossa (jos parametri θ on jatkuva)!

Konjugaattipriorit

- Konjugaattimalleille reunauskottavuus

$$p(\mathbf{y}) = \int_{\Omega} p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}) \mathrm{d}\,\boldsymbol{\theta}.$$

  osataan ratkaista analyyttisesti.
  - Posteriorijakauma $p(\boldsymbol{\theta}|\mathbf{y})$ osataan siten ratkaista suljetussa muodossa.
- Muille kuin konjugaattimalleille reunauskottavuutta ei ole mahdollista ratkaista suljetussa muodossa (jos parametri θ on jatkuva)!
  - Historiallisesti konjugaattimallit oleellisessa roolissa Bayes-päättelyssä.

Konjugaattipriorit

- Konjugaattimalleille reunauskottavuus

$$p(\mathbf{y}) = \int_{\Omega} p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}) \mathrm{d}\,\boldsymbol{\theta}.$$

  osataan ratkaista analyyttisesti.
  - Posteriorijakauma $p(\boldsymbol{\theta}|\mathbf{y})$ osataan siten ratkaista suljetussa muodossa.
- Muille kuin konjugaattimalleille reunauskottavuutta ei ole mahdollista ratkaista suljetussa muodossa (jos parametri θ on jatkuva)!
  - Historiallisesti konjugaattimallit oleellisessa roolissa Bayes-päättelyssä.
  - Nykyään kehittyneet MCMC-menetelmät ja tehokkaat tietokoneet mahdollistavat posteriorijakauman approksimoimisen simuloimalla myös ei-konjugaattimalleille $\rightarrow$ tähän palataan hetken päästä!

Miten valita priori??

- Hyvä kysymys: ehkä keskeisin este Bayes-päättelyn suuremmalle suosiolle soveltajien keskuudessa on se, että täytyy valita priori.
  - Ei "automaattista" pakettiratkaisua.

Miten valita priori??

- Hyvä kysymys: ehkä keskeisin este Bayes-päättelyn suuremmalle suosiolle soveltajien keskuudessa on se, että täytyy valita priori.
  - Ei "automaattista" pakettiratkaisua.
- Informatiivinen - epäinformatiivinen - jako ei formaalisti määritelty.

Miten valita priori??

- Konjugaattipriorit perinteisiä & niiden parametreilla näppärä tulkinta pseudo-havaintoina: nykyään MCMC-menetelmiä käytettäessä voidaan kuitenkin käyttää mitä tahansa prioria.

Miten valita priori??

- Konjugaattipriorit perinteisiä & niiden parametreilla näppärä tulkinta pseudo-havaintoina: nykyään MCMC-menetelmiä käytettäessä voidaan kuitenkin käyttää mitä tahansa prioria.
- Nyrkkisääntö: ellei halua / pysty käyttämään selkeästi informatiivista prioria, priori kannattaa valita aineiston oletetun skaalan mukaan, siten että suuruusluokka on suunnilleen oikea, mutta priori on kuitenkin sopivan epämääräinen

Miten valita priori??

- Konjugaattipriorit perinteisiä & niiden parametreilla näppärä tulkinta pseudo-havaintoina: nykyään MCMC-menetelmiä käytettäessä voidaan kuitenkin käyttää mitä tahansa prioria.
- Nyrkkisääntö: ellei halua / pysty käyttämään selkeästi informatiivista prioria, priori kannattaa valita aineiston oletetun skaalan mukaan, siten että suuruusluokka on suunnilleen oikea, mutta priori on kuitenkin sopivan epämääräinen
- OK jopa asettaa esim. (aineiston odotusarvoa estimoitaessa) normaalijakauma $N(\bar{y}, 5 \cdot s_y)$

Miten valita priori??

- Konjugaattipriorit perinteisiä & niiden parametreilla näppärä tulkinta pseudo-havaintoina: nykyään MCMC-menetelmiä käytettäessä voidaan kuitenkin käyttää mitä tahansa prioria.
- Nyrkkisääntö: ellei halua / pysty käyttämään selkeästi informatiivista prioria, priori kannattaa valita aineiston oletetun skaalan mukaan, siten että suuruusluokka on suunnilleen oikea, mutta priori on kuitenkin sopivan epämääräinen
- OK jopa asettaa esim. (aineiston odotusarvoa estimoitaessa) normaalijakauma $N(\bar{y}, 5 \cdot s_y)$
  - Tai vastaavasti normalisoida aineisto ja käyttää (aineiston odotusarvoa) estimoitaessa normaalijakaumaa $N(0, 5)$ priorina.

Miten valita priori??

- Konjugaattipriorit perinteisiä & niiden parametreilla näppärä tulkinta pseudo-havaintoina: nykyään MCMC-menetelmiä käytettäessä voidaan kuitenkin käyttää mitä tahansa prioria.
- Nyrkkisääntö: ellei halua / pysty käyttämään selkeästi informatiivista prioria, priori kannattaa valita aineiston oletetun skaalan mukaan, siten että suuruusluokka on suunnilleen oikea, mutta priori on kuitenkin sopivan epämääräinen
- OK jopa asettaa esim. (aineiston odotusarvoa estimoitaessa) normaalijakauma $N(\bar{y}, 5 \cdot s_y)$
  - Tai vastaavasti normalisoida aineisto ja käyttää (aineiston odotusarvoa) estimoitaessa normaalijakaumaa $N(0, 5)$ priorina.
- Joskus puhutaan tällöin **heikosti informatiivisesta priorista**.

Miten valita priori??

- Täältä löytyy hyviä priorisuosituksia Stanille (ja muutenkin): `https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations`

Miten valita priori??

- Täältä löytyy hyviä priorisuosituksia Stanille (ja muutenkin): `https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations`
- Jos aineistoa on paljon (huom. suhteessa malliin!), mallissa ei ole kovin paljon parametreja ja mallintamisen tarkoituksena selittäminen, ei priorin valinnalla monesti ole niin suurta merkitystä $\rightarrow$ monesti myös epäoleelliset tai hyvin heikosti informatiiviset priorit antavat hyvä tuloksia.