

Gaussian Process Surrogate Methods for Sample-Efficient Approximate Bayesian Computation

Gaussin prosessi -surrogaattimenetelmiä likimääräiseen
Bayesilaiseen päättelyyn

Marko Järvenpää
Department of Computer Science,
Aalto University

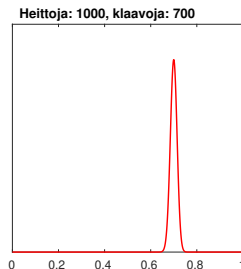
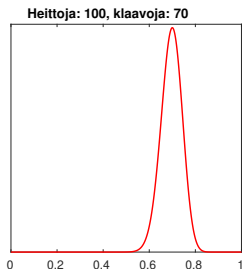
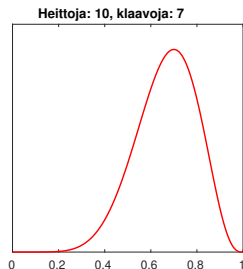
Lectio praecursoria, 25.9.2020

Tilastollinen päättely

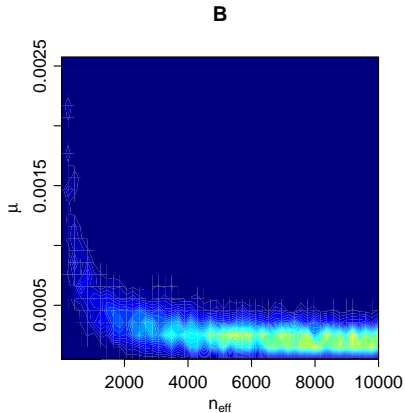
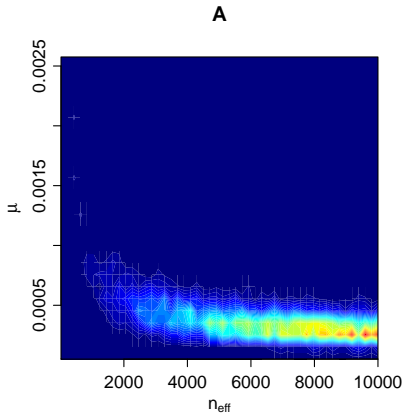
- ▶ **Todennäköisyyslaskenta:** esim. mikä on todennäköisyys että 10 kolikon heitolla saadaan 7 klaavaa?
- ▶ **Tilastollinen päättely:** esim. jos 10 kolikonheitolla saamme 7 klaavaa ja 3 kruunua, mitä voimme sanoa tämän perusteella kolikosta?
Entä jos saammekin 70 klaavaa ja 30 kruunua 100 heitolla?

Bayesilainen tilastollinen päättely 1/2

- **Posteriorijakauma** kertoo mitä mallin tuntemattomista parametreista tiedetään prioritiedon ja havaintojen perusteella
- Posteriorijakauma \propto priorijakauma \times uskottavuusfunktio



Bayesilainen tilastollinen päättely 2/2

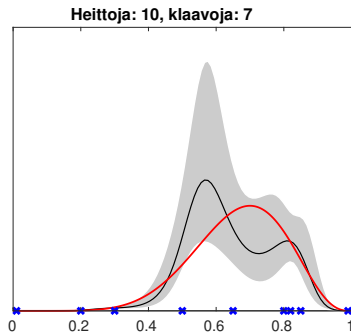
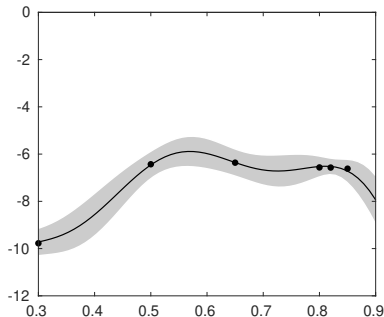


Bayesilainen tilastollinen päättely - laskennallisia haasteita

- ▶ Monimutkaisille malleille posteriorijakauman laskeminen tietokoneella voi olla hidasta
- ▶ Posteriorijakauman matemaattinen yhtälö voi olla liian hankala johtaa → **simulaatiopohjainen Bayesilainen tilastollinen päättely**
 - ▶ Voidaan tarvita esim. 100 000 simulointia. Jos jokainen simulointi vie esim. tunnin niin laskenta-aika on **11.4 vuotta**.
 - ▶ Myös muita haasteita (esim. johtaa likimääräiseen ratkaisuun)

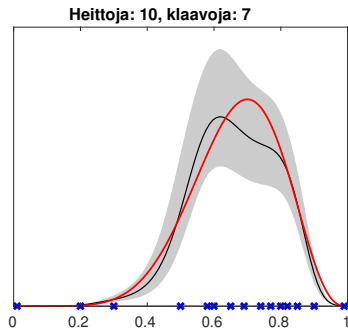
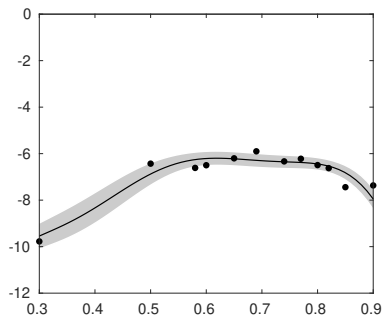
Surrogaattimallinnus

- ▶ Niin kutsutun surrogaattimallinnuksen avulla **voidaan vähentää tarvittavia posteriorijakauman laskuja tai simulaatioiden määrää**
- ▶ Tarkka ratkaisu korvataan likimääräisellä laskennan nopeuttamiseksi



Surrogaattimallinnus

- ▶ Niin kutsutun surrogaattimallinnuksen avulla **voidaan vähentää tarvittavia posteriorijakauman laskuja tai simulaatioiden määrää**
- ▶ Tarkka ratkaisu korvataan likimääräisellä laskennan nopeuttamiseksi



Väitöskirjan tutkimuskysymyksiä (hieman yksinkertaistettusti)

- ▶ Surrogaattimallinnuksen vaikutus likimääräisen posteriorijakauman tarkkuuteen
- ▶ Kuinka parametrit joilla mallia simuloidaan tulisi valita jotta laskennan määrä minimoituu?
- ▶ Kuinka simuloinnit saadaan rinnakkaistettua parhaalla mahdollisella tavalla?
- ▶ Sovellukset, erityisesti MRSA-sairaalabakteerin kolonisaation dynamiikan ymmärtäminen

Tutkimustuloksia: Laskennan rinnakkaistaminen

