



# DBN

## Mitä sillä tekee?

---

Dynaamisten Bayes-verkkojen  
määrittely aikasarja-analyysissä

Janne Toivola [jtoivola@iki.fi](mailto:jtoivola@iki.fi)



# Historiaa

- Bayesin kaavan hyödyntäminen
- BN-ohjelmistoja ollut ennenkin
- Tanskalaisten Hugin Expert
- Adnan Darwiche & co: SAMIAM
- Ei tarkoitettuja aikasarja-analyysiin



# Mitä iloa päättelystä

- Bayesilaisella päättelyllä tehdään optimaalisia johtopäätöksiä havaitsemattomista asioista
- Otetaan huomioon esim. aika, paikka, positio kromosomissa tms. konteksti
- Luokittelu, segmentointi jne. piilomuuttujien avulla
- Automaattinen puutteellisen datan käsittely



# BN: verkko

- Pohjana muuttujien riippuvuudet ilmaiseva suunnattu syklitön verkko DAG
- Mallin parametrit muotoa  $p(c|pa(c))$
- C on riippumaton muista kuin vanhemmistaan  $pa(C)$
- Muuttujia kuvaavien solmujen välillä kaaret vanhemmista lapsiin



# BN: liittymäpuu

- E.m. verkko soveltuu huonosti inferenssin vaatimiin laskutoimituksiin
- Täysi yhteis-tn-jakauma tuhlaa muistia
- Käytettävä vaihtoehtoista esitystapaa nimeltä liittymäpuu
- Puun solmut tallentavat nk. “perheiden” yhteistodennäköisyysjakaumia



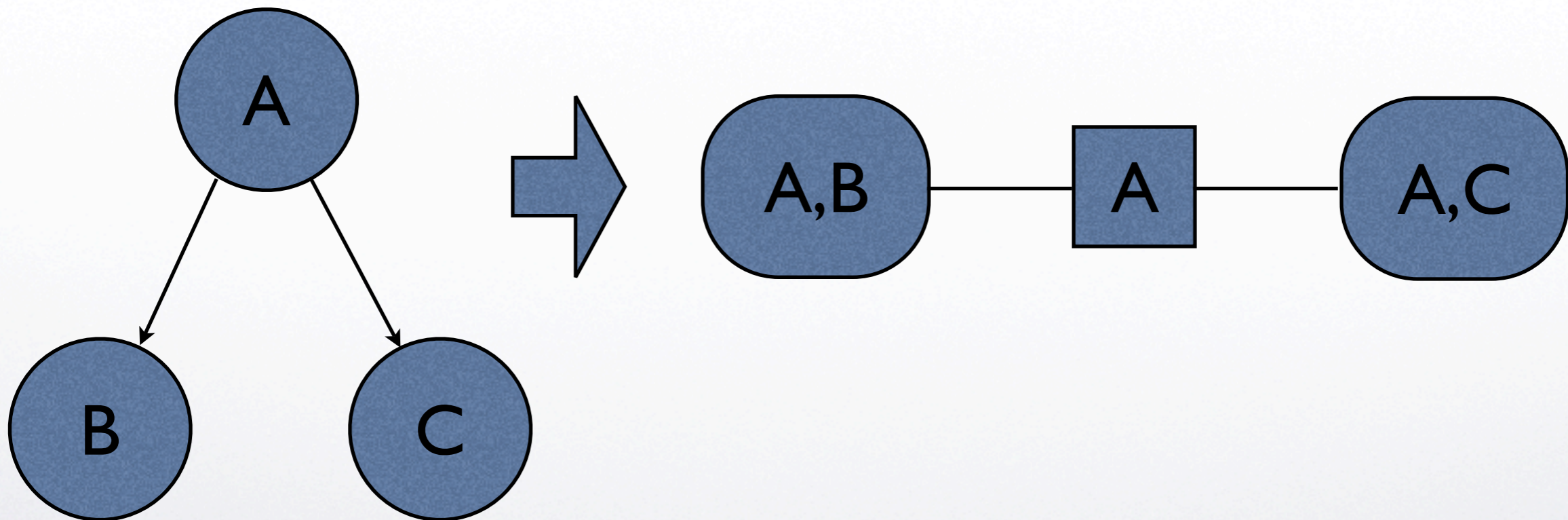
# BN: liittymäpuu

- Puuhun tallennetut jakaumat nimeltään potentiaaleja (mahd. normalisoimaton)
- Minimaalinen tapa esittää yhteis-tn-jakauma pienempien tulona
- Vrt.  $p(x, y, z) = p(x|y) * p(y|z) * p(z)$
- Myös ns. *running intersection property*



# BN: liittymäpuu

- Esimerkki graafista ja liittymäpuusta





# BN: päättely

- Päättely tapahtuu “levittämällä” potentiaaleja painottavia jakaumia naapurisolmuihin puussa
- Marginalisoimalla saadaan pienempi projektio potentiaalista:  $\phi(A) = \sum_B \phi(A, B)$
- Kertolaskulla päivitetään potentiaalia uuden evidenssin mukaan:  $\phi(A, C) := \phi(A, C) * \phi(A)$





# Lisää historiaa

- Kevin Murphy 2002
- Väitöskirja: *Dynamic Bayesian Networks: Representation, Inference & Learning*
- Esittää 1.5DBN käsitteen



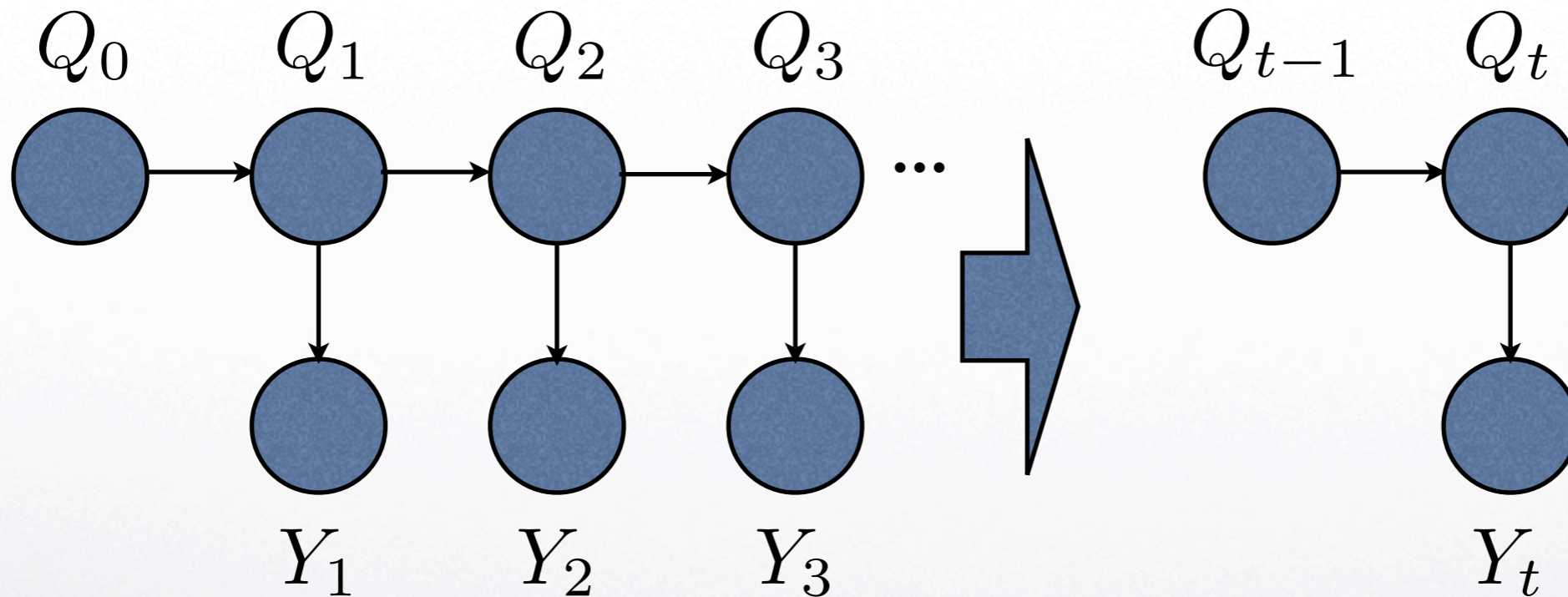
# DBN: aikaviipaleet

- Aikasarjoja varten toistuvarakenteiset mallit
- Esim. HMM toistaa pientä “verkkoa”
- Esitetään malli viipaleena, jota toistetaan
- Määritettävä toistuva osa liittymäpuusta laskutoimituksia varten
- Viipaleen käyttö päättelyssä “puoliaskelin”



# DBN: aikaviipaleet

- Esimerkki aikaviipaleesta: HMM



staattinen BN

1.5DBN



# Toteutus: yleistä

- C-kielinen ohjelmakirjasto
- Huginin net-tiedostojen jäsennys
- Muunnos graafeista liittymäpuuksi
- Aikasarjan lukeminen tiedostosta
- Edellä kuvattu inferenssi mahd. aikaviipalein
- EM-algoritmi parametrien oppimiseen



# Hugin net-formaatti

- Malli luetaan Huginin asettaman “standardin” mukaisesta tiedostosta
- node-määrittelyt kertovat muuttujat
- Oma *NIP\_next* kenttä aikaviipalemalleille
- potential-määrittelyt ehdollisille todennäköisyysjakaumille
- Graafin rakenne samassa implisiittisesti



# Graafista liittymäpuuksi

- NP-kova osuus liittymäpuun käytössä
- Muunnos heti net-tiedoston jäsennyksen jälkeen
- malli = liittymäpuu + muu kirjanpito

```
Nip model = parse_model("hmm.net");
```



# Datan luku tiedostosta

- Tiedoston 1. rivillä muuttujien nimet
- Loput rivit aikasarjan askelia
- Kullakin rivillä muuttujien arvojen *nimet*
- Luetaan data tiedostosta aikasarjaa esittäväksi tietorakenteeksi

```
n = read_timeseries(model, "data.txt", &n_timeseries);
```



# Päätely

- Inferenssialgoritmi antaa esim. “epävarman aikasarjan”  
UncertainSeries ucs =  
forward\_inference(ts, ts->hidden,  
ts->num\_of\_hidden);
- Annettujen muuttujien kunkin arvon todennäköisyys kullakin hetkellä
- Esim. yhteistodennäköisyydetkin mahdollisia, mutta implementoimatta





# EM-algoritmi

- Datan likelihoodin laskenta oli hankala
- Algoritmille annetaan opetettava malli, aikasarjoittain dataa ja kynnyisarvo lopetusehtoa varten
- Aikasarja sisältää viittauksen malliin

```
em_learn(n_timeseries, n, threshold, &learning_curve);
```



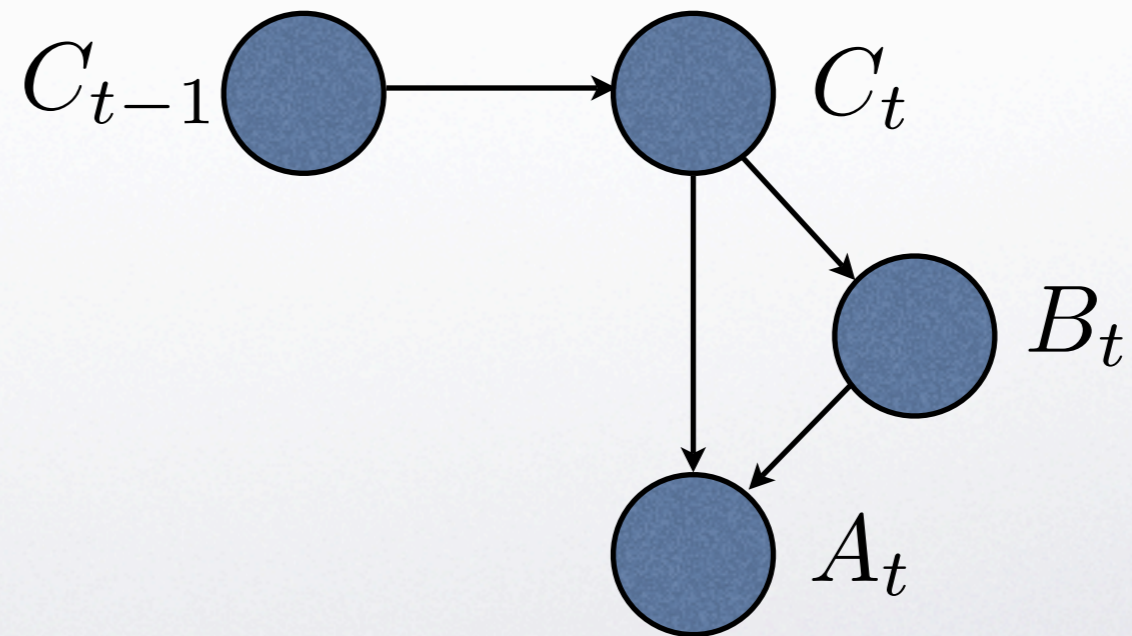
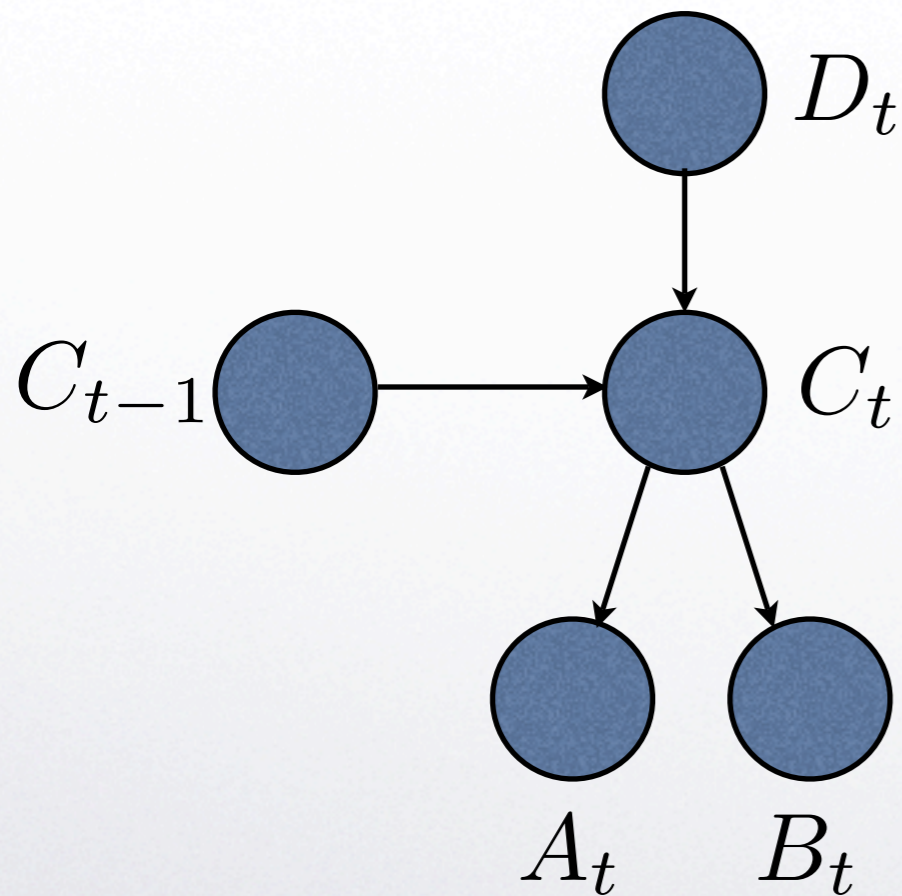
# Käytännön kokeilut

- Laboratoriotesti tunnetusta mallista generoidulla datalla
- - Havainnollistaa menetelmien toimintaa
- Kopiolukumuutosten aiheuttamat amplifikaatiot ihmisen perimässä
- - Osasyllisiä kasvaimiin (l. syöpään)



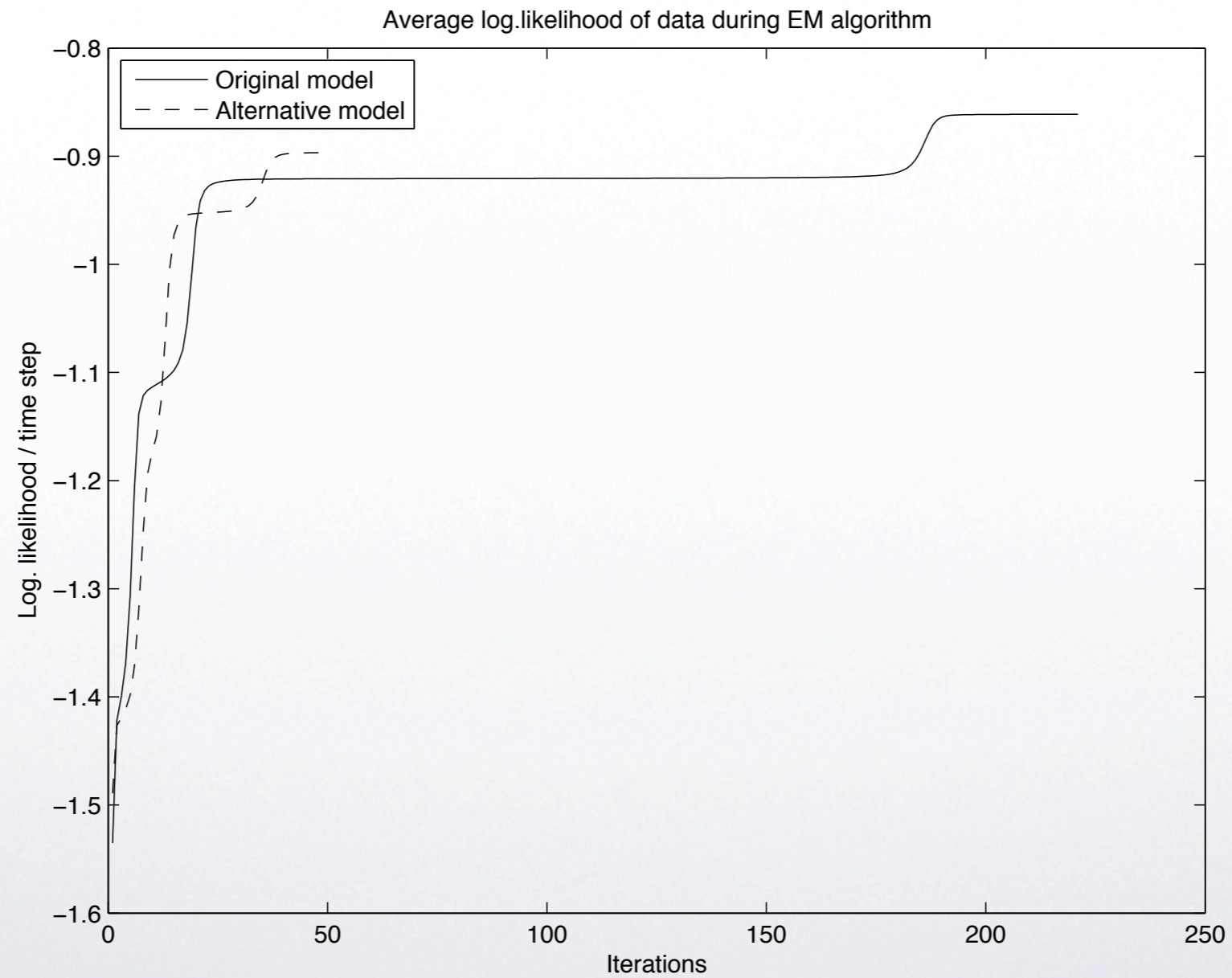
# Keinodatan mallit

- Alkuperäinen ja vaihtoehtoinen malli



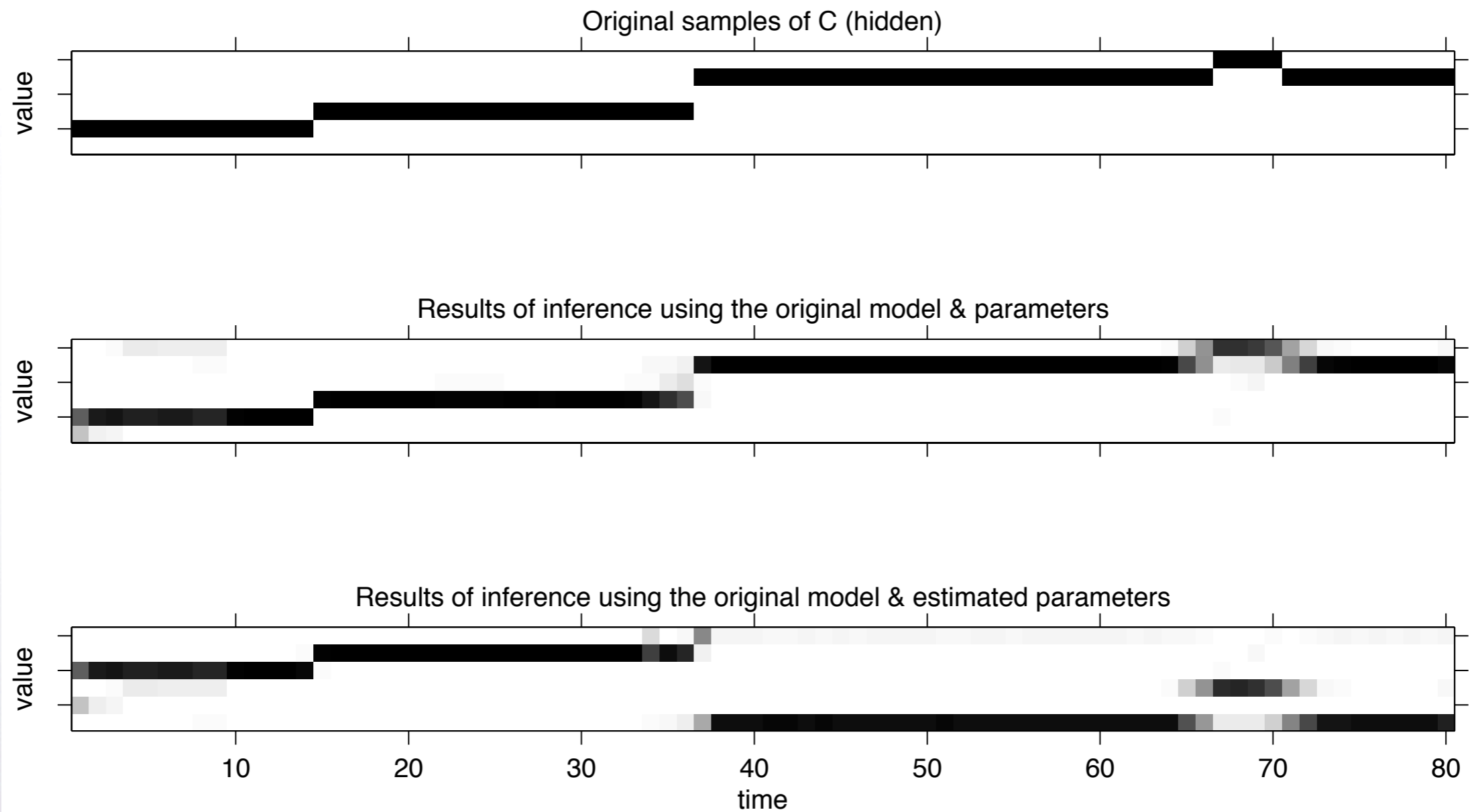


# Oppimiskäyrät



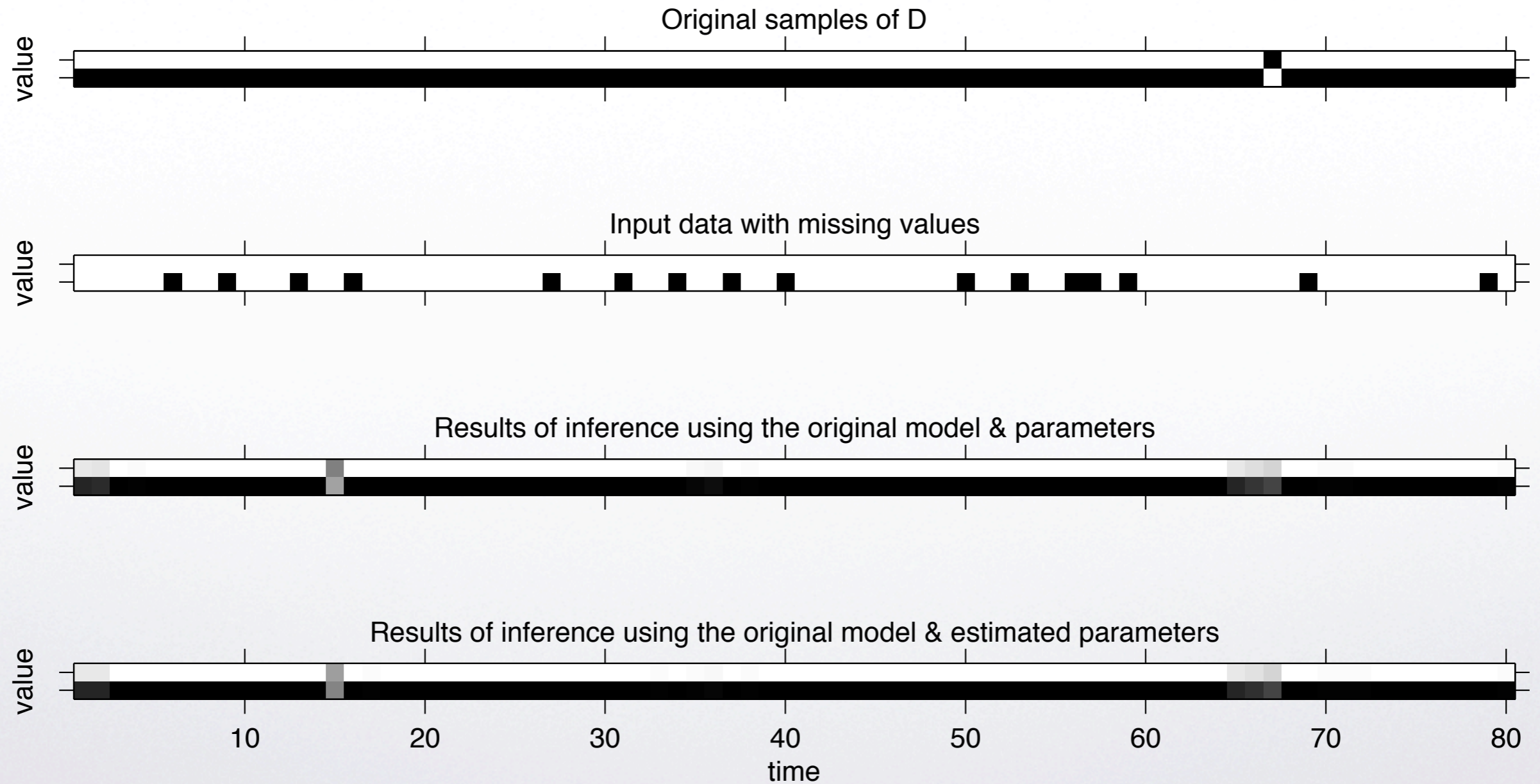


# Inferenssituloksia: C



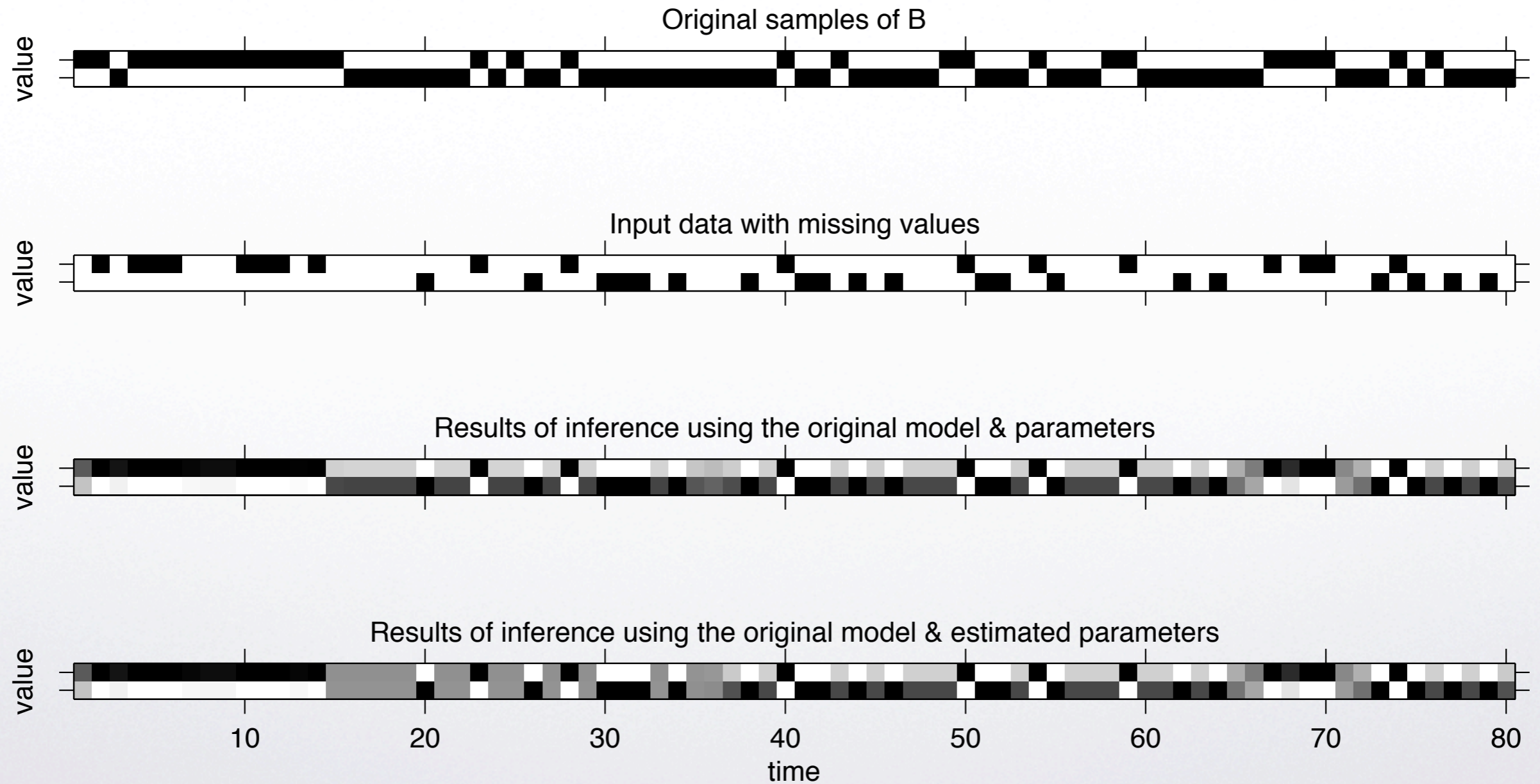


# Inferenssituloksia: D





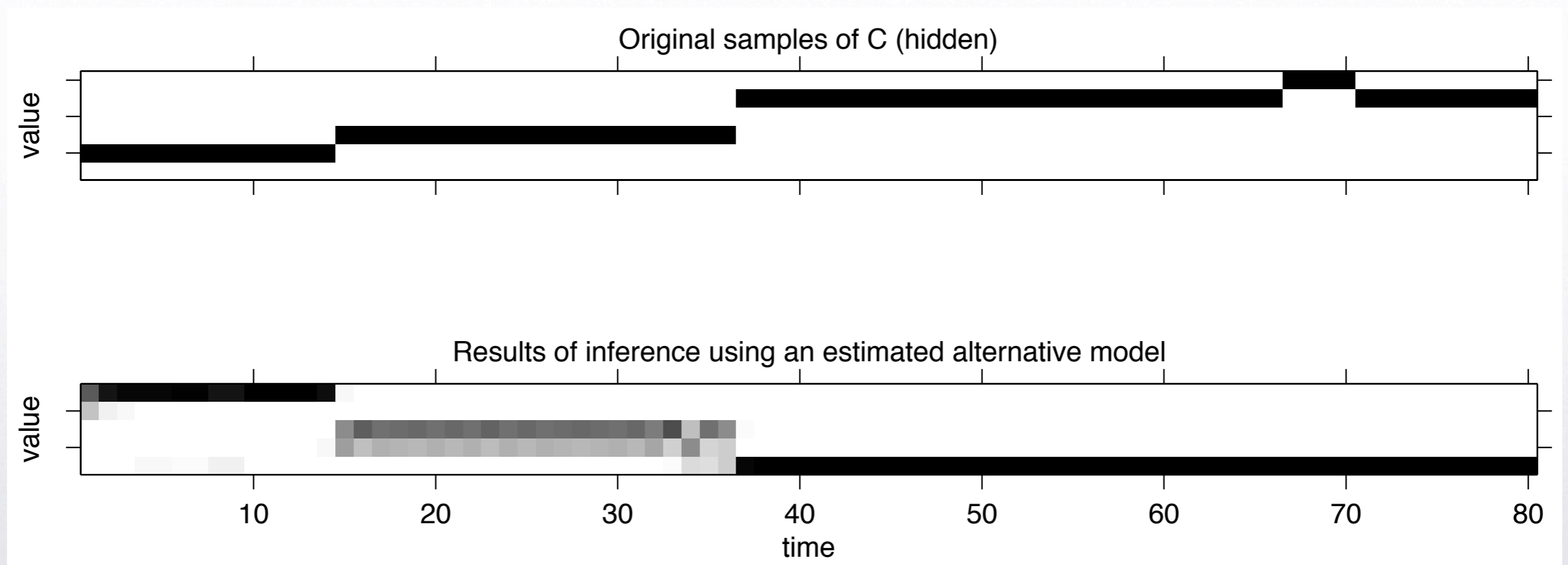
# Inferenssituloksia: B





# Inferenssituloksia: C

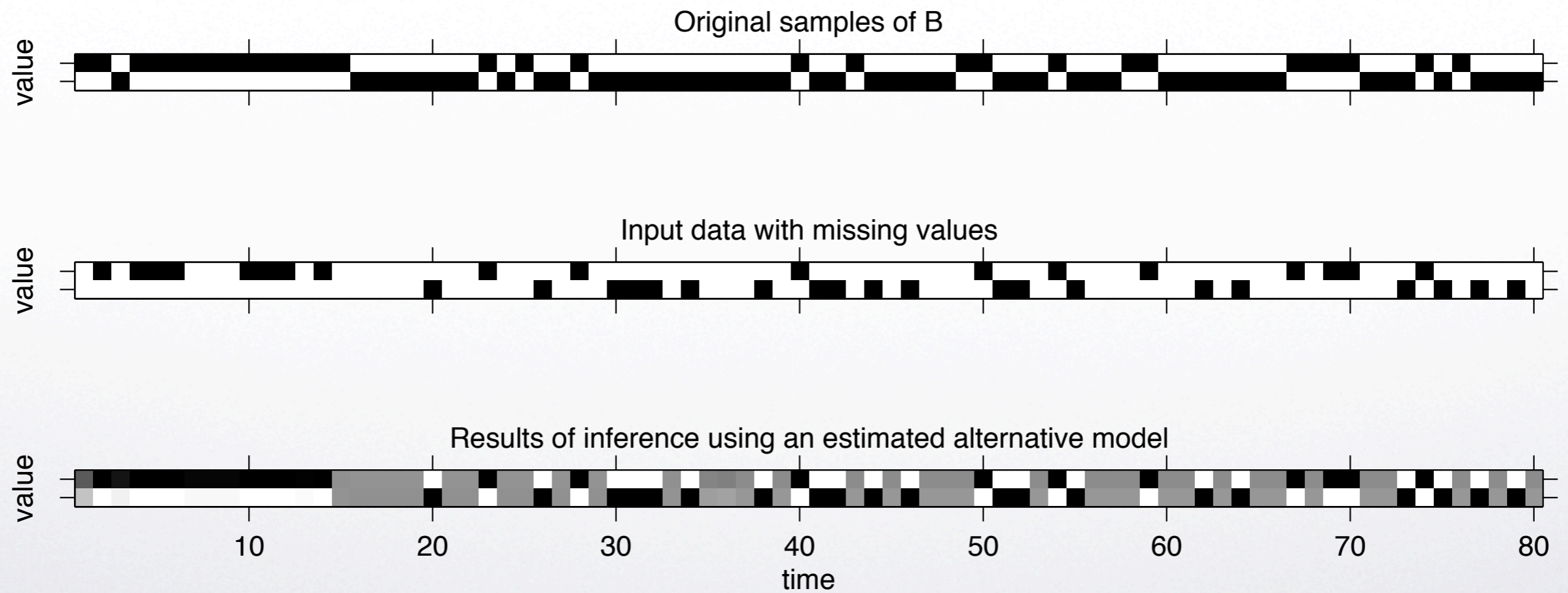
## Vaihtoehtoisen mallin piilotila







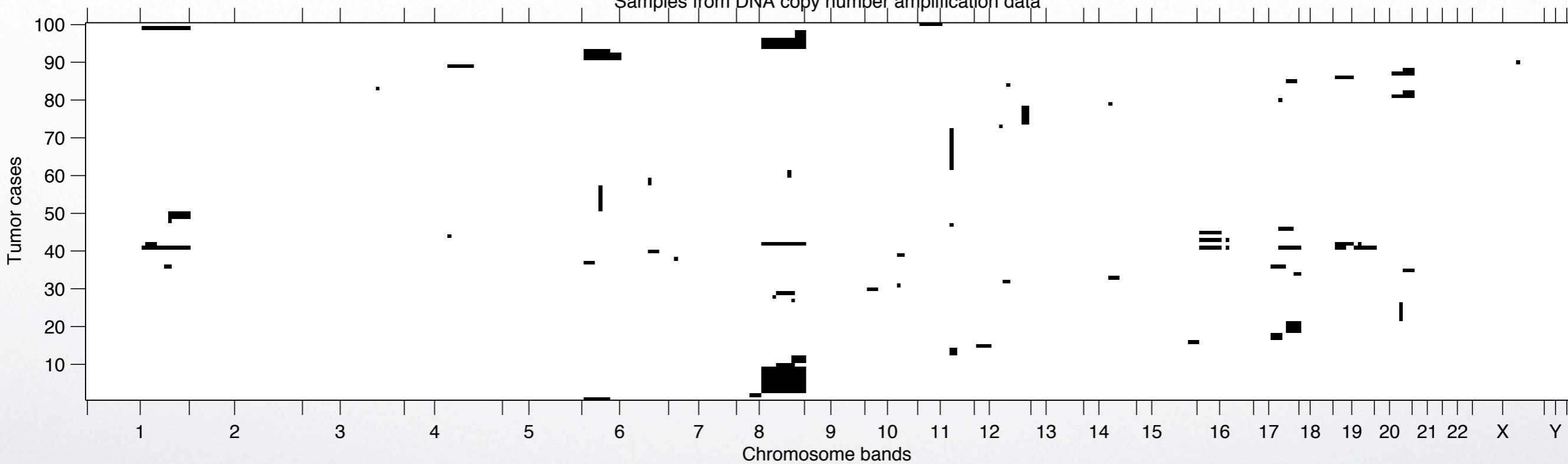
# Inferenssituloksia: B





# Amplifikaatiodata

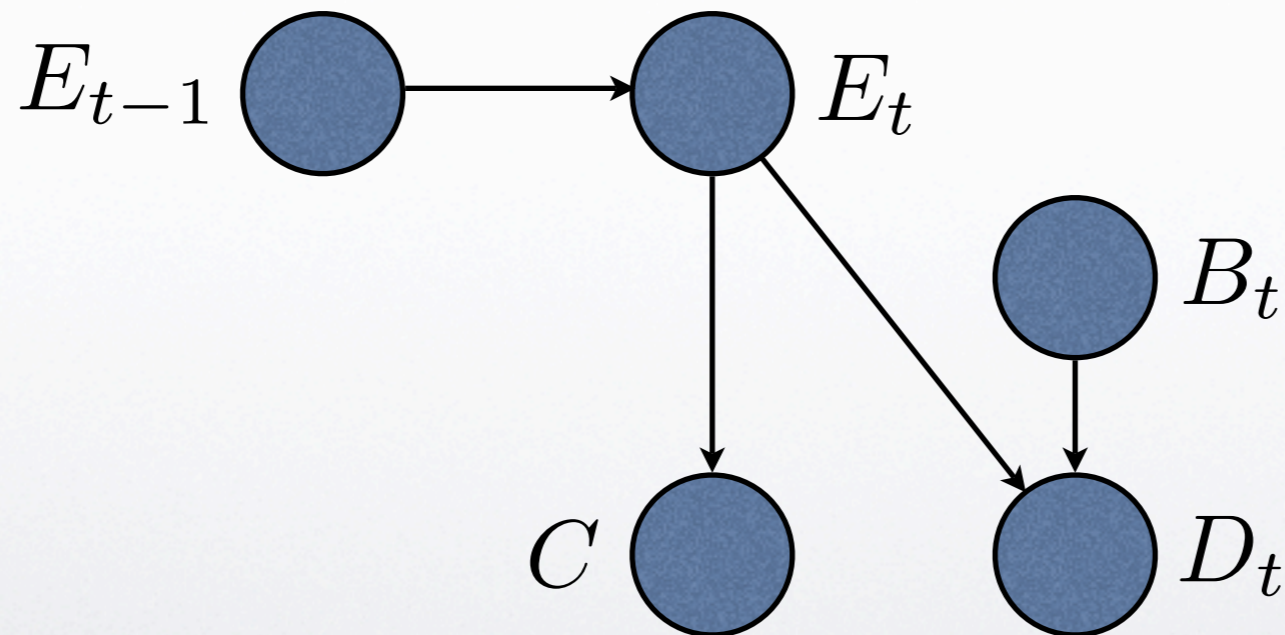
Samples from DNA copy number amplification data





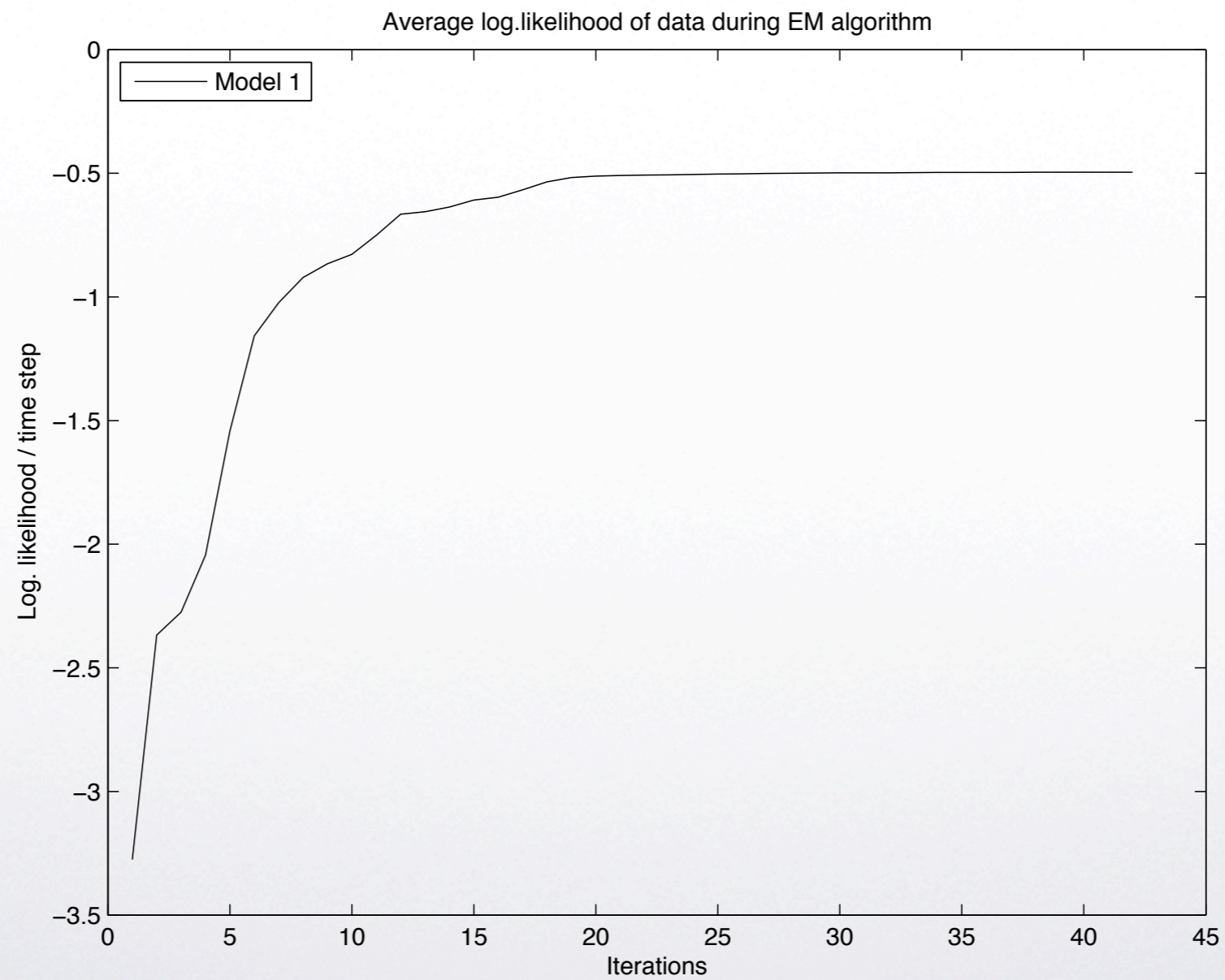
# Amplifikaatiomalli

- Kasvaintyyppi ja amplifikaatiot riippuvat piilomuuttujaketjusta





# Oppimiskäyrä







# Vielä muuta?

- Myös ns. “pehmeää evidenssiä” voisi syöttää havaintoina
- Jatkuvat muuttujat graafin lehtisolmuina?
- Paljon sovelluksia ja dataa?