- Ehkä 1 jäsen lisää?
  - 2. vuoden tutkijalinjalainen
- Esa Junttila – ”tekninen asiantuntija”
- Asiakas Salmenkivi
- Ehkä joku toinenkin asiakas Heikki Lokki?
- Wikla: ”Suunnitelkaa Javalla, tehkää Fortranilla!”
- Aiheesta:
  - mallintamista, prob, stat
  - päätellään kamaa aineistosta, ei visuaalista, vaan probalistista
  - esimerkki: kolikon tasapuolisuus. Data = ”1000” 1 on kr. p = ”tn että tulee 1”

  Pr (D | p = ½) = $\left(\tfrac{1}{2}\right)^4 = \frac{1}{16} = \frac{16}{256}$

  $$\Pr(D \mid p = \frac{1}{4}) = \frac{1}{4} \cdot \left(\frac{3}{4}\right)^3 = \frac{27}{256}$$ paras näistä kolmesta. Itseasiassa kaikista mahd.

  $$\Pr(D \mid p = \frac{3}{4}) = \frac{3}{4} \cdot \left(\frac{1}{4}\right)^3 = \frac{3}{256}$$

  Päätelläänkö että lantti on painotettu P(1) = ¼ ?
  Ei, olisi tarvittu lisää dataa.
  Otetaan aiempi tieto huomioon.
  Kaavio lantin kruunan tnästä: jakauma jossa piikki puolikkaan kohdalla.
  Tai oletetaan tasajakauma, kuten aluksi.

  $$\Pr(p \mid D) = \frac{\Pr(p) \cdot \Pr(D \mid p)}{\Pr(D)}$$ Missä Pr(p) on ennakkokäsitys jakaumasta.

  Ongelma on Pr(D). datan tn kaikkien yli... Mutta sitä ei tarvitse aina laskea, sillä

  $$\Pr(p \mid D) \cdot a = \Pr(p) \cdot \Pr(D \mid p)$$

  Käydään läpi p arvoilla 0.01, ... , 0.99
  Saadaan kuvaaja, suhteet kiinnostavat, eivät itsessään tnät.
  Kolikkotapauksessa saataisiin ehkä jakauma, joka olisi vähän lähempänä ¼ aa.
  Pr(D | p) sanotaan uskottavuudeksi (Likelihood)
  Pr(p | D) on meidän posteriori-tn.
  Entäs jos on paljon parametreja?
  Kaikkien kokeilu ei välttämättä ole järkevää 1000-ulotteisessa.
  Tehdään datasta päättelyä, approksimaatiota.
  Kun arvataan p jonka tn joka on suuri, kokeillaan sen vierestä... hill climbing
  - Idea: ei hirveän monimutkaisia konkreettisia malleja, jotka sopisivat moniin ongelmiin, ongelmat hyvin samanlaisia. Ohjelman pystyttävä ratkaisemaan pari yleistä mallia.
  - Hyvin dokumentoitu, voitaisiin käyttää näitä komponentteja tekemään kääntäjä joka laskee erilaisia ongelmia... epämääräistä.
  - Ongelma malleissa on kokeiltavien vaihtoehtojen määrä... mitä enemmän parametreja, sitä työläämpi, vaikeampi löytää hyviä p.
  - Läjä yhteentoimivia metodeja
  - malli syötteenä, ratkaistaan ongelma
  - periaatteessa ei väliä, mitkä ovat dataa...
  - ei-datan ylimmälle tasolle a priori-tn

Postscript-kalvoista levädatan arviot lämpötilasta
jakaumat N, Poisson

toinen malli jossa apumuuttuja r johon vaikuttavat opt_t, tol_t, scale ja temp.
R vaikuttaa abundanceen.
Idea: r = f(temp, opt_t, tol_t, scale)
ja r vaikuttaa stokastisesti abundanceen.
abundance ~ Poisson(r)

ja sitten laitetaan jotkin jakaumat opt_t ~N(10, 2) tai jotain... kaikille param annetaan jokin jakauma.

Kts. Postscriptista

tarkoitus saada selville muinaiset lämpötilat

tai vaikka toiseen suuntaan levän preferenssit.
Aina ei-datan posteriorijakauma(t).

integraalin approksimointia odotusarvon laskemiseksi

$$\int x\,\pi(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N} x_i$$ (monte carlo-integrointi) (myös postscriptissa)

jos ei voida generoida riippumattomia x,
generoidaan markovin ketjun muodostavia...

x_0, x_1, x_2, x_3 aina seuraava riippuu nykyisestä, ei edellisistä.

Entäs jos π on 100-dimensioinen?

Kun löydetään jokin iso tn, kokeillaan seuraavaksi sen läheltä. Hyväksytään testillä, tai hylätään.
Markovin ketju-tyyliin. (Metropolis-Hastings-algoritmi)

... lopulta valittujen x ketju noudattaa jakaumaa π .

x_0 on alkuarvo. Kokeillaan vaihtaa jokin x_0 :n parametri, hyväksytään riippuen sen

todennäköisyydestä ja siitä, onko saatu x_0' todennäköisempi kuin alkuarvo.

Esim. Jos nyt jotakin ehdotusta tässä ei hyväksyttäisi, toinen kandidaatti tilalle ja saadaan yksi parametri taas kohdalleen.

$$x_0 = \left(\theta_1^0, \theta_2^0, \theta_3^0\right)$$

$$\left(\theta_1^1, \theta_2^0, \theta_3^0\right)$$

$$\left(\theta_1^1, \theta_2^1, \theta_3^0\right)$$

$$x_1 = \left(\theta_1^1, \theta_2^1, \theta_3^1\right)$$