Matematiikan ja tilastotieteen laitos
Laskentaintensiiviset tilastolliset menetelmät 2006
Loppukoe 14.8.2008 (Kuulustelija: Petri Koistinen)

Huom. Laskutehtävissä annetaan pisteitä myös epätäydellisistä vastauksista, joissa selitetään, millä periaatteella täydelliseen vastaukseen tarvittavat suureet saataisiin laskettua.

1. Esitä konkreettinen simulointialgoritmi jakaumalle, jonka tiheysfunktio on normalisointivakiota vaille $h(x) = (2 + \sin(x) + \sin^2(3x)) \exp(-x^2/2)$. (Algoritmissa saa käyttää satunnaisarvoja standardijakaumista.)

2. Tahdomme arvion odotusarvolle $I = E|X|^{1.7}$, jossa $X \sim N(0,1)$. Selosta, miten I saadaan arvioitua käyttämällä Monte Carlo -integrointia ja kontrollimuuttujaa X^2. Selosta myös, miten I:lle saadaan laskettua luottamusväli.

3. Seuraavassa hierarkkisessa Bayes-mallissa Y_i kuvaa utaretulehdusten tapausten lukumäärää i:nnessä karjalaumassa. Suureet $\Theta_1, \ldots, \Theta_n$ ja $B_1, \ldots, B_n$ ovat parametreja ja luvut a, c ja d ovat tunnettuja vakioita.

Sillä ehdolla, että $\Theta_i = \theta_i, i = 1, \ldots, n$, ja $B_i = b_i$, $i = 1, \ldots, n$, ovat

$$Y_i \sim \text{Poisson}(\theta_i), \qquad i = 1, \ldots, n$$

keskenään riippumattomasti ja lisäksi

$$\Theta_i \sim \text{Gamma}(a, b_i), \qquad i = 1, \ldots, n$$

keskenään riippumattomasti. Lopulta

$$B_i \sim \text{Gamma}(c, d), \qquad i = 1, \ldots, n$$

riippumattomasti. Satunnaismuuttujien Y_i arvot y_i on havaittu, $i = 1, \ldots, n$.

Esitä Gibbsin otanta-algoritmi parametrien posteriorijakaumalle.

4. Tarkastelemme suurimman uskottavuuden estimointia tilanteessa, jossa satunnaismuuttujat X_i (selittäjät) noudattavat riippumattomasti välin $(0,1)$ tasajakaumaa, ja ehdolla $(X_1, \ldots, X_n)$ kukin satunnaismuuttujista Y_i (vasteet) noudattaa toisistaan riippumattomasti normaalijakaumaa $N(\beta X_i, 1)$, kun $i = 1, \ldots, n$. Kaikki vasteet Y_i on havaittu, mutta selittäjistä vain $X_2, \ldots, X_n$ on havaittu; ensimmäisen selittäjän arvoa ei tunneta. Johda EM-algoritmi parameterin β suurimman uskottavuuden estimaatin laskemiseksi.

Aputulos (jota ei tarvitse johtaa): jos V on välille (α, β) katkaistua normaalijakaumaa $N(\mu, \sigma^2)$ noudattava satunnaismuuttuja, niin sen momenttiemäfunktio on

$$M(t) = E\exp(tV) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)\, \frac{\Phi\left(\frac{\beta - \mu}{\sigma} - \sigma t\right) - \Phi\left(\frac{\alpha - \mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{\beta - \mu}{\sigma}\right) - \Phi\left(\frac{\alpha - \mu}{\sigma}\right)},$$

jossa Φ on jakauman $N(0,1)$ kertymäfunktio. V:n momentit saadaan momenttiemäfunktion derivaattojen arvoina pisteessä nolla, ts. $EV^k = M^{(k)}(0)$, kun $k = 1, 2, \ldots$

5. Vastaa, valintasi mukaan, **toiseen** seuraavista kysymyksistä.

a) Selosta, miten lasketaan studentisoitu saapasremmiluottamusväli.

b) Selosta, miten regressiofunktiota arvioidaan lokaalin lineaarisen regression avulla.

Liite: eräitä jakaumia

Eksponenttijakaumalla $\mathrm{Exp}(\lambda)$ parametrilla $\lambda > 0$ on tf

$$\lambda\, \mathrm{e}^{-\lambda x}, \quad x \geq 0.$$

Odotusarvo on λ^{-1} ja varianssi on λ^{-2}.

Gammajakaumalla $\mathrm{Gamma}(a, b)$ parametreillä $a > 0$ ja $b > 0$ on tf

$$\frac{b^a}{\Gamma(a)}\, x^{a-1} \mathrm{e}^{-bx}, \quad x > 0.$$

Odotusarvo on a/b ja varianssi ab^{-2}. $\Gamma(a)$ on gammafunktion arvo,

$$\Gamma(a) = \int_0^\infty x^{a-1} \mathrm{e}^{-x}\, \mathrm{d}x, \qquad a > 0.$$

Tunnetusti $\Gamma(a+1) = a\Gamma(a)$ kaikilla $a > 0$, ja $\Gamma(1) = 1$, joten $\Gamma(n) = (n-1)!$, kun $n = 1, 2, 3, \ldots$

Normaalijakaumalla $N(\mu, \sigma^2)$ (jossa $\sigma^2 > 0$) on tf

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right).$$

Odotusarvo on μ ja varianssi on σ^2.

Binomijakaumalla $\mathrm{Bin}(n, p)$ on ptnf

$$\binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Odotusarvo on np ja varianssi on $np(1-p)$.

Poissonin jakaumalla $\mathrm{Poisson}(\theta)$ parametrilla $\theta > 0$ on ptnf

$$\frac{1}{k!}\, \theta^k \mathrm{e}^{-\theta}, \quad k = 0, 1, 2, \ldots$$

Odotusarvo on θ ja varianssi on θ.