Johdatus tilastolliseen päättelyyn
Harjoitus 6 (3. 5.–7. 5. 2010)

1. Nastaa helistettiin lasipurkissa kymmenen kertaa. Se päätyi kolme kertaa selälleen ja seitsemän kertaa kyljelleen. Parametrina on todennäköisyys, jolla nasta päätyy selälleen. Priorijakauma on tasajakauma välillä $(0, 1)$. Laske posteriorijakauma sekä sen odotusarvo ja varianssi. (Sovella seuraavassa tehtävässä johdettavia betajakauman ominaisuuksia.)

2. (Betajakauman ominaisuuksia.) Olkoon satunnaismuuttujalla Y betajakauma parametreillä $\alpha, \beta > 0$, jolloin sen tiheysfunktio $f(y) = y^{\alpha-1}\,(1-y)^{\beta-1}/B(\alpha,\beta)$, kun $0 < y < 1$ (ja nolla muualla). Näytä, että Y:n odotusarvo $EY = \alpha/(\alpha+\beta)$ sekä laske Y:n toinen momentti $EY^2 = \int_0^1 y^2\, f(y)\, \mathrm{d}y$. Osoita kaavaa $\operatorname{Var} Y = EY^2 - (EY)^2$ käyttämällä, että

$$\operatorname{Var} Y = \frac{\alpha\,\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}.$$

   Opastus: tämä tehtävä ratkeaa soveltamalla kaavoja

$$B(x,y) = \int_0^1 t^{x-1}\,(1-t)^{y-1}\,\mathrm{d}t = \frac{\Gamma(x)\,\Gamma(y)}{\Gamma(x+y)}, \qquad \forall x, y > 0.$$
$$\Gamma(z+1) = z\,\Gamma(z), \qquad \forall z > 0.$$

   Näistä ensimmäinen kertoo yhteyden betafunktion ja gammafunktion

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

   välillä, ja toista kutsutaan gammafunktion funktionaaliyhtälöksi. (Tämän tehtävän kaavoja on tarpeetonta yrittää opetella ulkoa ainakaan tämän kurssin tenttiä varten.)

3. Olkoon satunnaismuuttujalla X Poissonin jakauma parametrilla θ, joka on Poissonin jakauman odotusarvo. Tulkitsemme tuntemattoman parametrin satunnaismuuttujaksi sekä kuvaamme sen arvoon liittyvää epävarmuutta käyttämällä priorijakaumaa $\mathrm{Gamma}(\alpha, \lambda)$, jossa gammajakauman tiheysfunktio on (normalisointivakiota vaille) muotoa

$$f(y) = c\,y^{\alpha-1}\,\mathrm{e}^{-\lambda y}, \qquad y > 0,$$

   (ja kirjoittamatta jätetty vakio on $c = \lambda^\alpha/\Gamma(\alpha)$). Tässä $\alpha > 0$ ja $\lambda > 0$ ovat (harkiten valittuja) vakioita.

   Saamme havainnon $X = x$ (jossa $x \geq 0$ on kokonaisluku). Osoita, että posteriorijakauma on tietty gammajakauma sekä laske kyseisen gammajakauman parametrit.

4. Erään huokean mutta epätarkan ajoitusmenetelmän mukaan tietyn kallion ikä on 370 ± 20 miljoonaa vuotta. Vuosia myöhemmin tehdyn tarkemman ajoituksen perusteella tämän kallion ikä on 421 ± 8 miljoonaa vuotta. Yhdistämme nämä mittaukset ajattelemalla, että ensimmäinen niistä muodostaa ennakkokäsityksemme kallion iälle: tarkemmin sanoen kuvaamme ennakkokäsitystämme normaalijakaumalla, jonka odotusarvo on 370 ja varianssi on 20×20. Toisen mittauksen X mallinnamme niin, että mikäli kallion ikä on todellisuudessa θ, niin X:n (ehdollinen) jakauma on normaalijakauma odotusarvolla θ ja varianssilla 8×8. Laske kallion iän posteriorijakauma. Laske myös sellainen väli, joka sisältää 95% posteriorijakaumasta, ja jonka kumpaankin häntään jää 2,5% posteriorijakaumasta. (Tämä on helppoa, koska posteriorijakauma on tietty normaalijakauma; tulos on esimerkki bayesläisestä väliestimaatista.)

(Tehtävän normaalijakaumaoletukset tehtiin laskujen helpottamiseksi: kallion ikä ei tietenäkään voi olla negatiivinen, vaikka normaalijakaumamallin mukaan tämä on mahdollista, joskin erittäin epätodennäköistä.)

5. Posteriorijakaumani välillä $(0, 1)$ määritellylle parametrille θ on betajakauma, jonka parametrit $\alpha_1, \beta_1 > 0$ olen selvittänyt. Seuraavaksi suunnittelen tekeväni uuden mittauksen X_{n+1}, jonka ehdollinen jakauma (ehdolla havainnot sekä parametri θ) on binomijakauma otoskoolla m sekä onnistumistodennäköisyydellä θ. (Esimerkiksi kyseessä voisi olla nasta lasipurkissa, ja voisin suunnitella palkkaavani jonkun muun ravistamaan purkkia m kertaa ja kertomaan minulle, kuinka monta kertaa nasta päätyy selälleen.)

   Laske uuden mittauksen ennustejakauman odotusarvo (opastus: tiedät kaavan binomijakauman odotusarvolle) sekä uuden mittauksen enustejakauman pistetodennäköisyysfunktio (opastus: lopputulos ei liene sinulle ennestään tuttu, mutta vastaus on helppo johtaa, ja siinä esiintyy binomikerroin sekä betafunktion arvoja).