Huomautus merkinnöistä. Eri satunnaismuuttujien tiheysfunktiot erotetaan usein toisistaan alaindeksillä, esim. f_X ja f_Y. Alaindeksillä merkitään myös ehdollista tiheysfunktiota, esim. $f_{X|Y}$.

Bayes-päättelyssä käytetään usein lyhennysmerkintää, jossa kaikkia *eri* tiheysfunktioita merkitään vain f:llä ilman alaindeksiä. Tällöin sulkujen sisällä olevasta argumentista on pääteltävä, mistä funktiosta on kysymys: esim. $f(x)$ tarkoittaa samaa kuin $f_X(x)$. Lyhennysmerkintää ei pidä käyttää jos se aiheuttaa sekaannusta, esim. $f(5)$ ei kerro onko kyse X:n vai Θ:n tiheysfunktiosta. Vrt. luento 5A.

(Luentomonisteessa käytetään f:ää datan tiheysfunktioille ja p:tä parametrin tiheysfunktioille, mutta tämä merkintä ei ole kovin yleisesti käytössä.)

5B Bayesläinen päättely

Tuntitehtävät

5B1 (Kadonnut lentokone) Seuraavantyyppistä tilastollista hakumenetelmää on käytetty mm. yritettäessä paikantaa 8.3.2014 kadonnutta Malaysia Airlinesin lentoa MH370. Etsintöjen kohteena oleva alue on jaettu neljään suureen ruutuun. Lennon taustatietojen pohjalta arvellaan lentokoneen sijaitsevan ruudussa 1 tn:llä 50%, ruudussa 2 tn:llä 30%, ruudussa 3 tn:llä 10% ja ruudussa 4 tn:llä 10%. Olosuhteista johtuen arvellaan, että yksittäiseen ruutuun kohdistettu etsintäyritys epäonnistuu löytämään ruudussa sijaitsevan koneen tn:llä 70%, aiemmista etsintäyrityksistä riippumatta.

(a) Tulkitaan lentokoneen sijainti satunnaismuuttujaksi (tuntemattomaksi parametriksi) $\Theta \in \{1, 2, 3, 4\}$. Esitä sijainnin priorijakauma.

(b) Etsintäryhmä tutkii ensiksi ruudun 1. Merkitään $X = 1$ jos kone löytyy ja $X = 0$ muuten. Määritä Θ:n uskottavuusfunktio havainnolle $X = 0$.

(c) Konetta ei löydetä ensimmäisellä hakuyrityksellä. Määritä koneen sijainnin posteriorijakauma tämän havainnon suhteen. Kannattaako ruudusta 1 etsiä uudelleen?

(d) Etsintäryhmä päättää tutkia ruudun 1 uudelleen. Merkitään tämän toisen haun tulosta satunnaismuuttujalla Y. Tämänkin haun tulos on, että konetta ei löytynyt. Määritä lentokoneen sijainnin posteriorijakauma tämän datan valossa. Mistä etsintäryhmän kannattaa seuraavaksi etsiä?

5B2 (Metron vuoroväli.) Metro kulkee säännöllisesti θ minuutin välein. Tuomas saapuu metrolaiturille joka päivä satunnaiseen aikaan, joten i:ntenä päivänä metron odotusajalla X_i on tasajakauma välillä $[0, \theta]$, tiheydellä

$$f(x_i \,|\, \theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x_i \le \theta, \\ 0, & \text{muuten.} \end{cases}$$

Koska Tuomas ei tunne θ:n arvoa, hän käsittelee sitä satunnaismuuttujana Θ. Yleistietonsa nojalla hän olettaa, että vuoroväli on ainakin 1 minuutti. Hän valitsee prioritiheydeksi

$$f_\Theta(\theta) = \begin{cases} 0.2\theta^{-1.2}, & \theta \ge 1, \\ 0, & \text{muuten.} \end{cases}$$

Viitenä päivänä Tuomas havaitsee odotusajat $\vec{x} = (7, 3, 2, 9, 6)$.

- Piirrä Tuomaksen prioritiheys välillä $[0, 20]$ käsin tai tietokoneella. Tarkista että kyseessä on todella tiheysfunktio, laskemalla sen integraali koko mahdollisella välillä $[1, \infty)$. Vihje: Vaikka eksponentti ei ole kokonaisluku, voit integroida funktion kuten polynomifunktion. Piirtämiseen riittää luultavasti funktion arvon laskeminen muutamassa pisteessä, jonka jälkeen näet funktion muodon suunnilleen.

- Laske posteriorijakauman tiheysfunktio (johda sen lauseke) ja piirrä se välillä $[0, 20]$. (Muista: priori kertaa uskottavuus, ja lopuksi normalisointi. Viiden havainnon uskottavuus on yksittäisten havaintojen uskottavuuksien tulo.)

- Tuomas päättää käyttää Θ:n *posteriorijakauman odotusarvoa* (huom. ei moodia) vuorovälin Θ piste-estimaattina. Laske kyseinen odotusarvo. (Tunnet posteriorijakauman edellisen kohdan perusteella, laske odotusarvo normaaliin tapaan.)

- Käyttäen posteriorijakaumaa, laske todennäköisyys (havaintojen perusteella) että $\Theta < 15$. Ilmaise sanallisesti, mitä tämä tarkoittaa.

Kotitehtävät

5B3 (Lehtien pituudet) Botanisti arvelee, että tietyn kasvilajin lehtien pituus (cm) noudattaa normaalijakaumaa odotusarvona Θ (tuntematon) ja keskihajontana $\sigma = 2$. Hän olettaa lisäksi, että tuntematon odotusarvo Θ noudattaa normaalijakaumaa odotusarvona $\mu_0 = 10$ ja keskihajontana $\sigma_0 = 1$. Botanisti mittasi viiden kyseisen lajin kasvin lehtien pituudet ja sai tulokseksi $\vec{x} = (10, 14, 11, 17, 8)$ (lehtien pituudet voidaan olettaa toisistaan riippumattomiksi). Määritä tehtyjen havaintojen valossa:

(a) Satunnaismuuttujan Θ posteriorijakauman odotusarvo.

(b) Väli, joka sisältää tuntemattoman odotusarvon todellisen arvon 90% todennäköisyydellä.

Vihje. Ylläolevin oletuksin tuntemattoman odotusarvon posteriorijakauma on normaalijakauma, jonka odotusarvo ja keskihajonta löytyvät luentomonisteen luvusta 9 / luennosta 5A. Normaalijakauman kertymäfunktioon voi käyttää taulukoita tai esim. R:n funktioita `pnorm` ja `qnorm`.

5B4 (Vaarallinen tie) Juuri avatulla tiellä kuukauden aikana sattuvien kolarien määrän oletetaan noudattavan Poisson-jakaumaa,

$$f(x \mid \theta) \;=\; e^{-\theta}\frac{\theta^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

tuntemattomalla odotusarvoparametrilla $\theta > 0$. Aiemman kokemuksensa perusteella bayesläinen insinööri arvelee, että tuntemattoman parametrin Θ arvo on 1 todennäköisyydellä 0.20, 2 todennäköisyydellä 0.60, ja 3 todennäköisyydellä 0.20. Koska parametrilla Θ on tässä vain kolme mahdollista arvoa, se on diskreetti satunnaismuuttuja.

(a) Ensimmäisenä kuukautena tiellä sattuu kaksi kolaria. Määritä odotusarvon $\Theta \in \{1, 2, 3\}$ posteriorijakauma tämän havainnon valossa. Havaintona on siis *yksi* luku, nimittäin 2, joka on peräisin Poisson-jakaumasta. Aloita selvittämällä uskottavuusfunktio.

(b) Toisena kuukautena tiellä ei satu yhtään kolaria. Määritä odotusarvon $\Theta \in \{1, 2, 3\}$ posteriorijakauma molempien kuukausien havaintojen valossa (oletetaan, että eri kuukausien kolarien lukumäärät ovat toisistaan riippumattomia). Havaintona on siis *kaksi* lukua, 2 ja 0, kumpikin peräisin samasta Poisson-jakaumasta.