

TILASTOLLISET MENETELMÄT TÄHTITIEESSÄ

IDL-harjoitus 2, Heikki Salo 6.3.2008

1. Luo N kappaleen otos välille $[0, 1]$ tasaisesti jakaantuneita satunnaismuuttujia X (merkitään $X_i \sim \text{TAS}(0, 1)$) käyttäen IDL:n `randomu`-funktioita.

- Plottaa `histo_f`-proseduurin avulla lukujen jakauma, jakamalla väli $[0, 1]$ esim. 10, 100, 1000, ... väliin. Piirrä samaan kuvaan tasaisen jakauman teoreettinen tiheysfunktio. Kokeile eri N :n arvoja.

- Miten luot satunnaislukuja välillä $[a, b]$, mikä on luomasi otoksen keskiarvo ja keskihajonta? Miten nämä suhtautuvat tasaisen jakauman teoreettiseen keskiarvoon ($\mu = \frac{b+a}{2}$) ja keskihajontaan ($\sigma = \frac{b-a}{\sqrt{12}}$). Vaikuttaako N otoksen keskihajontaan?

- Totea käyttämällä eri N :n arvoja, että otoksen keskiarvo lähestyy $\bar{X}$ tasaisen jakauman keskiarvoa N :n kasvaessa. Arvioi poikkeaman $\bar{X} - \mu$ keskihajontaa (= ”keskiarvon keskivirhe”) luomalla M kappaletta N :n luvun otoksia, ja laskemalla saaduista $\bar{X}_j$, $j = 1, \dots, M$ niiden otoshajonta σ_μ .

Miten σ_μ suhtautuu teoreettiseen arvioon $\sigma / \sqrt{N}$, jossa σ on tasaisen jakauman keskihajonta.

(Otoskeskiarvojen ja otoshajontojen laskemiseen IDL:ssä on `mean` ja `stdev` funktiot. Suositus on käyttää uudempaa `moment`-funktioita, joka palauttaa keskiarvon, varianssin, vinokkuuden ja huipukkuuden)

2. Luo N kappaleen otos normaalijakaumaa noudattavia satunnaismuuttujia ($X_i \sim N(0, 1)$) käyttäen IDL:n `randomn`-funktioita.

-Tarkista otoksen keskiarvo ja hajonta. Kuinka luot normaalijakauman, jonka keskiarvo on μ ja keskihajonta σ ?

-Plottaa saamasi jakauma eri N :n arvoilla, yhdessä teoreettisen Gaussisen tiheysfunktion kanssa.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right]$$

-Plottaa otoksesi kumulatiivinen tiheysjakauma. (Vihje: `sort`-funktio antaa taulukon alkioiden suuruusjärjestyksen, eli `x(sort(x))` järjestää x :n alkioit suuruusjärjestykseen; Kokeile mitä saat kun kirjoitat `plot, x(sort(x)), linden(n)/(float(n)-1)`, jossa `“n=n_elements(x)”` on otoksen alkioiden lukumäärä)

Vertaa Gaussisen jakauman teoreettiseen kumulatiiviseen tiheysfunktioon $\Phi(x)$ (Muista `gauss_pdf`-funktio.)

3. Normaalijakaantuneita satunnaismuuttujia voidaan luoda tasaisen jakauman avulla seuraavasti: Valitaan

$$x_1 \sim TAS(0, 1) \text{ ja}$$

$$x_2 \sim TAS(0, 1)$$

ja muodostetaan luvut

$$y_1 = \sqrt{-2 \log x_1} \cos 2\pi X_2 \text{ ja}$$

$$y_2 = \sqrt{-2 \log x_1} \sin 2\pi X_2$$

Totea että menetelmä toimii (myös IDL:n `randomn`-funktio käyttää tätä Box-Muller menetelmää).

4. Normaalijakaantuneita lukuja voidaan tuottaa myös keskeiseen raja-arvolauseeseen pohjautuen. Otetaan $N \gg 1$ tasaisesti jakaantunutta lukua $X_i \sim TAS(0, 1)$ ja muodostetaan niiden summa. Keskeisen raja-arvolauseen perusteella summan jakauma lähenee normaalijakaumaa $N(\mu, \sigma)$, jossa μ = summan odotusarvo (nyt $\frac{N}{2}$) ja σ^2 sen varianssin odotusarvo ($\frac{N}{12}$)

Erityisesti,

$$y = \left(\sum_{i=1}^{12} X_i \right) - 6 \sim N(0, 1) \text{ (Miksi?)}$$

Totea käytännössä millaisia normaalijakauman approksimaatioita saat vaihtelemalla summattavien lukujen määriä.

Huom: eo arvolla $N = 12$ saatujen lukujen arvoalue on $-6 \leq X \leq 6$. Arvioi kuinka suuri puute tämä on (eli miten suurella todennäköisyydellä oikea normaalijakauma $N(0,1)$ antaisi lukuja tämän valin ulkopuolella) Karkean arvion voit tehdä luomalla suuren määrän satunnaislukuja `randomn`-funktion avulla ja katsomalla moniko on tämän välin ulkopuolella.

5. Luodaan eksponentiaalisesti jakaantuneita satunnaislukuja tasaisesti jakaantuneiden lukujen $R \sim TAS(0, 1)$ avulla.

$$f(x) = ce^{-x/c} \text{ kun } x > 0 \quad (\bar{x} = c)$$

$$\text{Asetetaan } R_i = \int_0^{X_i} f(x)dx \Rightarrow \boxed{X_i = -c \log R_i}$$

Totea että menetelmä toimii, vertaamalla sitä IDL:n `randomu`-funktion avulla luotuihin eksponentiaalisestijakaantuneisiin lukuihin (kts `?randomu`), sekä teoreettiseen tiheysfunktioon.

6. Demonstroi keskeisen raja-arvolauseen pätevyys summaamalla eksponentiaalisesti jakaantuneita satunnaislukuja. Luennoilla annetussa esimerkissä (Wall-Jenkins Kuva 2.6) verrattiin 2, 4, ja 16 eksponenttijakaumasta otetun luvun keskiarvon jakaumia Gaussiseen jakaumafunktioon. Käytä näitä samoja arvoja.

7. Poisson jakauman luonti.

Poisson jakauma kuvaa prosessia, jossa aikayksikköä kohti tapahtuu keskimäärin λ tapahtumaa (raja-arvo binomijakaumasta, kun toistokokeen yksittäisessä yrityksessä on hyvin pieni onnistumistodennäköisyys p , mutta yritysten määrä N on suuri, $\lambda = pN$). Tällöin tapahtuneiden tapausten lukumäärä X noudattaa jakaumaa.

$$p(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, 2, \dots$$

Poisson prosessissa tapahtumien välinen odotusaika on eksponentiaalisesti jakaantunut. Koska tapahtumia on keskimäärin λ /aikayksikkö, niin odotusajan w keskiarvo on $\bar{w} = 1/\lambda$ aikayksikköä. Siten w :n jakauma on

$$f(w) = \bar{w} e^{-w/\bar{w}} = 1/\lambda e^{-\lambda w}$$

Poisson jakaantuneita satunnaislukuja voidaan tuottaa summaamalla eo. eksponentiaali-jakauman mukaisesti jakaantuneita odotusaikoja W_i , kunnes niiden summa ylittää aikayksikön. Merkitään summattujen odotusaikojen lukumäärää $X + 1$: // Tällöin tiedetään, että aikayksikön kuluessa tapahtui X tapausta. Näin luotujen X :n arvojen jakauma noudattaa Poisson-jakaumaa.

Tee ohjelma joka tuottaa Poisson jakautuneita satunnaislukuja käyttäen a) IDL random-ohjelmaa ja b) edellä kuvattua menetelmää. Vertaa teoreettiseen jakaumaan. Valitse esim. $\lambda = 5$.