

Generointi yksinkertaisista diskreeteistä jakaumista

- Seuraavassa $U, U_1, \dots, U_n$ tarkoittavat riippumattomia $U(0,1)$ -jakautuneita sm:jia
- $\text{int}(X) = \lfloor X \rfloor = X$:n kokonaisosa

Jakauma	Generointikaava
Symm. kaksiarvoinen $\{0,1\}$ jakauma $P\{X = 0\} = P\{X = 1\} = 0.5$	$\text{int}(2U)$ tai $\text{int}(U + 0.5)$
Kaksiarvoinen $\{0,1\}$ jakauma $P\{X = 0\} = 1 - p, P\{X = 1\} = p$	$\text{int}(U + p)$
Kaksiarvoinen $\{-1,1\}$ jakauma $P\{X = 0\} = 1 - p, P\{X = 1\} = p$	$2 \text{int}(U + p) - 1$
Kolmiarvoinen $\{0,1,2\}$ jakauma $\text{tn:t} = \{1 - p_1 - p_2, p_1, p_2\}$	$\text{int}(U + p_2) + \text{int}(U + p_1 + p_2)$
Tasainen diskreetti jakauma $\{0, 1, 2, \dots, n - 1\}$	$\text{int}(n U)$
Tasainen diskreetti jakauma $\{1, 2, 3, \dots, n\}$	$\text{int}(n U) + 1$
Binomijakauma $\text{Bin}(n, p)$	$\sum_{i=1}^n \text{int}(U_i + p)$

Generointi geometrisesta jakaumasta

- Geometrista jakaumaa $\text{Geom}(p)$ noudattavan diskreetin satunnaismuuttujan X pistetodennäköisyydet ovat

$$P\{X = i\} = p_i = p(1 - p)^i \quad i = 0, 1, 2, \dots$$

- Satunnaismuuttujan X arvojen generointiin on käytettävissä seuraava yksinkertainen menetelmä
- Algoritmi

$$X = \left\lfloor \frac{\log U}{\log(1 - p)} \right\rfloor$$

missä $U \sim U(0, 1)$

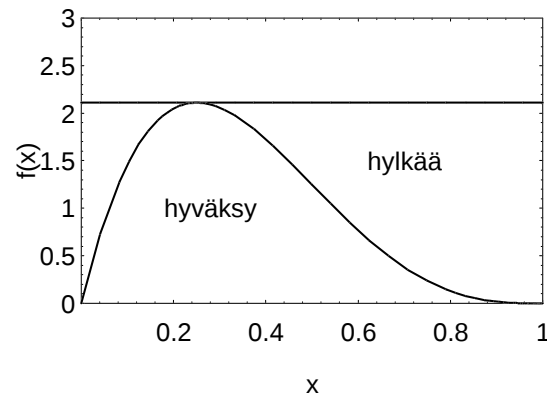
- Kyseessä on itse asiassa $\text{Exp}(-\log(1 - p))$ -jakaumaa noudattavan jatkuvan satunnaismuuttujan arvojen diskretointi lähinnä alempiin kokonaislukuarvoihin

Hylkäysmenetelmä

- Tehtävänä on generoida sm:n X arvoja jakaumasta, jolla on tiheysfunktio $f(x)$
- Olkoon $g(x)$ toinen tiheysfunktio ja c jokin vakio siten, että
 - $c g(x)$ antaa ylärajan $f(x)$:lle, ts. $c g(x) \geq f(x)$ X :n koko arvoalueella
 - tiheysfunktioita $g(x)$ noudattavan satunnaismuuttujan arvoja osataan generoida (helposti)
- Tällöin X :n generointi voidaan suorittaa seuraavasti
- Algoritmi
 - Generoi X tiheysfunktioilla $g(x)$
 - Generoi Y tasaisesta jakaumasta $U(0, c g(X))$
 - Jos $Y \leq f(X)$, niin hyväksy X
 - * muutoin generoi edelläolevan mukaisesti uusia arvoja X ja Y , kunnes löytyy pari, joka täyttää hyväksymisehdon; tulosta X
- Todistus:
 - $P\{X \in (x, x + dx) \text{ ja } Y \leq f(X)\} = g(x)dx \cdot f(x)/cg(x) = f(x)dx/c$
 - $P\{Y \leq f(X)\} = \int f(x)dx/c = 1/c$
 - $P\{X \in (x, x + dx) | Y \leq f(X)\} = (f(x)dx/c)/(1/c) = f(x)dx$

Hylkäysmenetelmä (esimerkki)

- Arvoalueen ollessa äärellinen väli (a, b) tiheysfunktioiksi $g(x)$ voidaan valita kyseisellä välillä tasanjakautuneen sm:n tiheysfunktio: $g(x) = 1/(b - a)$, kun $x \in (a, b)$



- Oletetaan, että $X \in (0, 1)$ noudattaa betajakaumaa $\beta(2, 4)$ eli tiheysfunktio on

$$f(x) = 20x(1 - x)^3, \quad 0 \leq x < 1$$

- Funktio on rajoitettu suorakulmiossa, jonka korkeus on 2.11
– valitaan $c = 2.11$ ja $g(x) = 1$, kun $0 \leq x < 1$

- Algoritmi toimii nyt seuraavasti
 - Generoi X tasaisesta jakaumasta $U(0, 1)$
 - Generoi Y tasaisesta jakaumasta $U(0, 2.11)$
 - Jos $Y \leq 20X(1 - X)^3$, hyväksy X ja lopeta, muutoin jatka alusta, kunnes löytyy hyväksyttävä pari
- Tässä tapauksessa generoitavat (X, Y) arvot edustavat tasanjakatunutta pistettä suorakulmiossa
 - on selvää, että hyväksytyjen arvojen $X = x$ osuus on verrannollinen $f(x)$:ään
 - hyväksytyjen X -arvojen tiheysfunktio on silloin $f(x)$

Yhdistelmämenetelmä (composition method)

- Oletetaan, että sen jakauman tiheysfunktio $f(x)$, josta X pitää arpoa, voidaan kirjoittaa (hajoittaa) muotoon

$$f(x) = \sum_{i=1}^r p_i f_i(x)$$

missä

- p_i :t muodostavat diskreetin todennäköisyysjakauman, $\sum_i p_i = 1$
- $f_i(x)$:t ovat tiheysfunktioita, $\int f_i(x) dx = 1$
- Tällaista jakaumaa kutsutaan sekoitusjakaumaksi (yhdistelmäjakauma, komposiittijakauma)
- Näytteenotto voidaan suorittaa seuraavasti
 - arvotaan ensin indeksi I jakaumasta $\{p_1, p_2, \dots, p_r\}$
 - arvotaan tämän jälkeen X käyttäen tiheysfunktioita $f_I(x)$

Yhdistelmämenetelmä (jatkoa)

- Esimerkiksi menetelmää voidaan käyttää jakamalla X :n vaihtelualue (a, b) erillisiin intervaleihin $(a_1, b_1), (a_2, b_2), \dots, (a_r, b_r)$

– p_i on tällöin todennäköisyys, että X osuu intervalliin i

$$p_i = \int_{a_i}^{b_i} f(x) dx$$

– $f_i(x)$ on ehdollinen tiheysjakauma intervallissa i

$$f_i(x) = \begin{cases} f(x)/p_i & x \in (a_i, b_i) \\ 0 & \text{muulloin} \end{cases}$$

Yhdistelmämenetelmä (esimerkki 1)

- Tehtävänä on generoida pisteitä X $\text{Exp}(1)$ -jakaumasta
- Jaetaan väli $(0, \infty)$ intervalleihin $(i, i + 1)$, $i = 0, 1, 2, \dots$
- Välien todennäköisyydet

$$p_i = P\{i \leq X < i + 1\} = e^{-i} - e^{-(i+1)} = e^{-i}(1 - e^{-1})$$

muodostavat geometrisen jakauman (alkaa 0:sta)

- Ehdolliset tiheysfunktiot ovat

$$f_i(x) = e^{-(x-i)} / (1 - e^{-1}) \quad i \leq x < i + 1$$

eli intervallissa i ($X - i$):llä on tiheysfunktio $e^{-x} / (1 - e^{-1})$, $0 \leq x < 1$

- Algoritmi

- arvotaan I geometrisesta jakaumasta $p_i = e^{-i}(1 - e^{-1})$, $i = 0, 1, 2, \dots$
- arvotaan Y tiheysfunktioilla $e^{-x} / (1 - e^{-1})$, $0 \leq x < 1$ (esim. hylkäysmenetelmällä)
- $X = I + Y$

- Etu: ei tarvita logaritmifunktiota niin kuin kertymäfunktion käännösmenetelmässä

Yhdistelmämenetelmä (esimerkki 2)

- Tiheysfunktioiden asemesta yhdistelmämenetelmässä voidaan yhtä hyvin toimia kertymäfunktioiden avulla
- Olkoon X :n kertymäfunktio

$$\begin{aligned} F(x) &= 1 - \alpha e^{-\beta_1 x} - (1 - \alpha) e^{-\beta_2 x} \\ &= \alpha(1 - e^{-\beta_1 x}) + (1 - \alpha)(1 - e^{-\beta_2 x}) \end{aligned}$$

- Algoritmi
 - valitse käytettävä jakauma I : $P\{I = 1\} = \alpha$, $P\{I = 2\} = 1 - \alpha$
 - arvo X asianomaisesta jakaumasta $F_I(x)$

$$F_1(x) = 1 - e^{-\beta_1 x} \quad F_2(x) = 1 - e^{-\beta_2 x}$$

- eli jos $I = 1$ niin $X = -\frac{1}{\beta_1} \log U$; jos $I = 2$ niin $X = -\frac{1}{\beta_2} \log U$
- Kertymäfunktion käännösmenetelmän käyttö olisi tässä tapauksessa hankalaa
 - käänteisfunktioita ei voi kirjoittaa analyyttisesti

Jakauman karakterisointiin perustuva menetelmä

- Monet jakaumat on määritelty muodossa X on jakautunut kuten n :n riippumattoman $sm:n$ summa, joista kukin noudattaa annettua jakaumaa “Dist”
- Tällöin X voidaan generoida kirjaimellisesti arpomalla riippumattomasti arvot n :lle muuttujalle Z_i jakaumasta “Dist”, jolloin $X = Z_1 + Z_2 + \dots + Z_n$ noudattaa haluttua jakaumaa
- Esimerkkejä tällaisista jakaumista ovat binomijakauma, gammajakauma (Erlangin jakauma) sekä χ^2 -jakauma

Jakauman karakterisointi (esim. binomijakauma)

- Binomijakauma $\text{Bin}(n, p)$ on n :n riippumattoman Bernoulli(p)-jakautuneen muuttujan summan jakauma

$$X = \sum_{i=1}^n B_i, \quad B_i \sim \text{Bernoulli}(p) \quad \Rightarrow \quad X \sim \text{Bin}(n, p)$$

- Bernoulli(p)-jakautunut muuttuja saa arvon 1 tn:llä p ja arvon 0 tn:llä $1 - p$
- $B_i = \text{int}(p + U_i) = \lfloor p + U_i \rfloor, \quad U_i \sim \text{U}(0, 1) \quad (\text{kokonaisosa})$

- Algoritmi

$$X = \sum_{i=1}^n \lfloor p + U_i \rfloor, \quad U_i \sim \text{U}(0, 1)$$

Jakauman karakterisointi (esim. gammajakauma)

- Kokonaislukuarvoilla n saadaan $\Gamma(n, \lambda)$ -jakauma n :n riippumattoman $\text{Exp}(\lambda)$ -jakautuneen muuttujan summan jakaumana

$$X = \sum_{i=1}^n Y_i, \quad Y_i \sim \text{Exp}(\lambda) \quad \Rightarrow \quad X \sim \Gamma(n, \lambda)$$

- Ottamalla huomioon, miten eksponenttijakautunut sm voidaan generoida tasanjakautuneen muuttujan U avulla, saadaan
- Algoritmi

$$X = -\frac{1}{\lambda} \log \prod_{i=1}^n U_i, \quad U_i \sim U(0, 1)$$

- Logaritmien summa kirjoitettu logaritmina tulosta
 - tämä on edullista, koska logaritmifunktio joudutaan laskemaan vain kerran

Jakauman karakterisointi (esim. χ^2 -jakauma)

- $\chi^2(\nu)$ -jakauma ν :llä vapausasteella (kokonaisluku) on ν :n $N(0, 1)$ -jakautuneen muuttujan summa

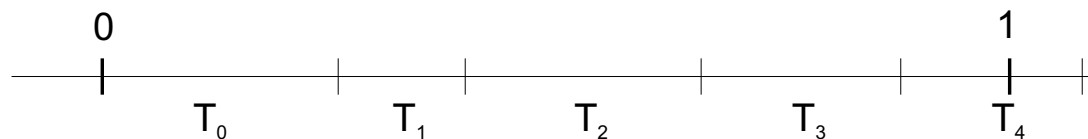
$$X = \sum_{i=1}^{\nu} Y_i, \quad Y_i \sim N(0, 1) \quad \Rightarrow \quad X \sim \chi^2(\nu)$$

Jakauman karakterisointi (esim. Poisson-jakauma)

- Toisen tyyppisen esimerkin karakterisoinnista tarjoaa Poisson-jakauma
- Poisson-prosessista (intensiteetti a) välille $(0, 1)$ tulevien saapumisten lukumäärä N on Poisson-jakautunut parametrilla a , $N \sim \text{Poisson}(a)$
- Arvotaan väliaikoja T_i , $i = 0, 1, 2, \dots$ $\text{Exp}(a)$ -jakaumasta, $T_i = -(1/a) \log U_i$
- N on välille $(0,1)$ “mahtuvien” väliaikojen lukumäärä eli $N = \min \left\{ n : \sum_{i=0}^n T_i > 1 \right\}$
- Algoritmi

$$N = \min \left\{ n : \prod_{i=0}^n U_i < e^{-a} \right\}$$

- kerrotaan lukuja $U_i \sim U(0, 1)$, $i = 0, 1, 2, \dots$ keskenään
- N on ensimmäinen i :n arvo, jolla tulo on pienempi kuin e^{-a}



Poisson-jakauma: numeerinen esimerkki

- Olkoon keskiarvoparametri $a = 0.2$
- Vertailuarvo $u = e^{-0.2} = 0.8187$

i	U_i	$U_0 \cdots U_i$	accept/continue	Tulos
0	0.4357	0.4357	$< u$, accept	$N = 0$
0	0.4146	0.4146	$< u$, accept	$N = 0$
0	0.8353	0.8353	$\geq u$, continue	
1	0.9952	0.8313	$\geq u$, continue	
2	0.8004	0.6654	$< u$, accept	$N = 2$

- Suurilla a :n arvoilla menetelmästä tulee hidas; N :n arvot ovat tällöin tyypillisesti suuria, ja joudutaan generoimaan monta lukua U_i
- Tällöin on parempi käyttää diskreetointimenetelmää (diskreetin kf:n käänös)
- Hyvin suurilla a :n arvoilla voidaan myös käyttää normaaliapproksimaatiota (merk. $Z \sim N(0, 1)$)

$$\text{Poisson}(a) \approx N(a, a) \Rightarrow N \approx \lceil a + \sqrt{a}Z - 0.5 \rceil$$

Jakauman karakterisointi (muita esimerkkejä)

- a :nneksi pienin luvuista $U_1, U_2, \dots, U_{a+b+1}$, missä U_i :t ovat riippumattomia tasanjakautuneita sm:ia, $U_i \sim U(0, 1)$, noudattaa $\beta(a, b)$ -jakaumaa
- Kahden $N(0, 1)$ -jakautuneen muuttujan suhde noudattaa $\text{Cauchy}(0, 1)$ -jakaumaa
- $\chi^2(\nu)$ -jakauma parillisilla vapausasteilla ν on sama kuin $\Gamma(\nu/2, 1/2)$ -jakauma
- Kahden riippumattoman gamma-jakautuneen muuttujan avulla voidaan konstruoida beta-jakautunut muuttuja

$$X_1 \sim \Gamma(b, a) \quad X_2 \sim \Gamma(c, a) \quad \Rightarrow \quad \frac{X_1}{X_1 + X_2} \sim \beta(b, c)$$

- Jos $X \sim N(0, 1)$ on normaalijakautunut sm, niin $e^{\mu + \sigma X}$ on lognormaalinen (μ, σ) sm

Generointi monidimensioisesta jakaumasta

- Tehtävä: generoi arvoja muuttujille $X_1, \dots, X_n$, joilla on yhteistiheysfunktio $f(x_1, \dots, x_n)$
- Kirjoitetaan tämä tf muotoon $f(x_1, \dots, x_n) = f_1(x_1)f_2(x_2 \mid x_1) \dots f_n(x_n \mid x_1, \dots, x_{n-1})$ missä $f_1(x_1)$ on X_1 :n reunajakauma ja $f_k(x_k \mid x_1, \dots, x_{k-1})$ on X_k :n ehdollinen jakauma ehdolla $X_1 = x_1, \dots, X_{k-1} = x_{k-1}$
- Ideana on generoida muuttujat yksi kerrallaan: ensin generoidaan X_1 oman reunajakaumansa mukaisesti, sitten generoidaan X_2 ehdollisesta jakaumasta käyttäen ehtona jo arvottua X_1 :n arvoa jne
- Merkitään F_k :lla ehdollista tf:ta f_k vastaavaa ehdollista kertymäfunktiota ja käytetään kertymäfunktion käännoäsmenetelmää
- Algoritmi
 - generoi sm:t $U_1, \dots, U_n$ tasaisesta jakaumasta $U(0, 1)$
 - ratkaise yhtälöt (käännä kertymäfunktiot)

$$\begin{aligned} F_1(X_1) &= U_1 \\ F_2(X_2 \mid X_1) &= U_2 \\ &\vdots \\ F_n(X_n \mid X_1, \dots, X_{n-1}) &= U_n \end{aligned}$$

Monidimensioinen jakauma: esimerkki

- Tehtävänä on generoida pisteitä (X, Y) yksikköneliössä, jonka vasen alakulma on origossa, käyttäen diagonaalin suunnassa kasvavaa tiheysfunktia (integraali neliön yli on ykkönen)

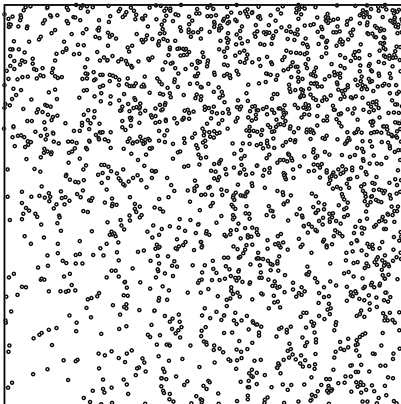
$$f(x, y) = x + y$$

- X :n reunajakauman tiheys ja kertymäfunktio ovat

$$f(x) = \int_0^1 f(x, y) dy = x + \frac{1}{2}, \quad F(x) = \int_0^x f(x') dx' = \frac{1}{2}(x^2 + x)$$

- Y :n ehdolliset tiheys- ja kertymäfunktio ovat

$$f(y | x) = \frac{f(x, y)}{f(x)} = \frac{x + y}{x + \frac{1}{2}}, \quad F(y | x) = \int_0^y f(y' | x) dy' = \frac{xy + \frac{1}{2}y^2}{x + \frac{1}{2}}$$



- Kertymäfunktioiden käännökset antavat laskentakaavat

$$X = \frac{1}{2}(\sqrt{8U_1 + 1} - 1)$$

$$Y = \sqrt{X^2 + U_2(1 + 2X)} - X$$

Generointi monidimensioisesta normaalijakaumasta

- Satunnaisvektorilla $\mathbf{X} = (X_1, \dots, X_n)$, joka noudattaa monidimensioista normaalijakaumaa (multinormaalijakauma) on tiheysfunktio

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \Sigma^{-1}(\mathbf{x}-\mathbf{m})}$$

missä $\mathbf{m}$ on keskiarvovektori ja Σ on kovarianssimatriisi

- Koska Σ on positiividefiniitti ja symmetrinen matriisi, voidaan aina löytää yksikäsitteinen alakolmiomatriisi (vaihtoehtoisesti symmetrinen matriisi) $\mathbf{C}$ siten, että $\Sigma = \mathbf{C}\mathbf{C}^T$
- Arvoja $\mathbf{X}$:lle voidaan nyt generoida seuraavasti
- Algoritmi

$$\mathbf{X} = \mathbf{C}\mathbf{Z} + \mathbf{m}$$

missä vektorin $\mathbf{Z} = (Z_1, \dots, Z_n)$ komponentit ovat riippumattomia normaalijakautuneita satunnaismuuttujia, $Z_i \sim N(0, 1)$

- kaavan oikeellisuus todetaan suorittamalla muuttujien vaihto tiheysfunktiossa, jolloin $\mathbf{Z}$:n tiheysfunktioksi saadaan $(2\pi)^{-n/2} e^{-\frac{1}{2}\mathbf{Z}^T \mathbf{Z}}$