Normálne rozdelenie

- $\varphi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{x^2}{2}}$
- $\Phi(x) = \int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x$
- $\Phi_0(x) = \int_{0}^{\infty} \varphi(x)\,\mathrm{d}x$

Platí:

- $\varphi(-x) = \varphi(x)$
- $\Phi(-x) = 1 - \Phi(x)$
- $\Phi(x) = \frac{1}{2} + \Phi_0(x)$
- $\Phi_0(x) = -\Phi_0(x)$

Pre malé x možno použiť

$$\Phi_0(x) = \varphi(x)\left(x + \frac{x^3}{3} + \frac{x^5}{3\cdot 5} + \frac{x^7}{3\cdot 5\cdot 7} + \dots\right)$$

Pre veľké x možno použiť

$$1-\Phi_0(x) = \varphi(x)\left(\frac{1}{x} - \frac{1}{x^3} + \frac{1\cdot 3}{x^5} - \frac{1\cdot 3\cdot 5}{x^7} + \frac{1\cdot 3\cdot 5\cdot 7}{x^9} - \dots\right)$$

kde $x > 0$.

Iná aproximácia

$$\Phi_0(x) = \frac{1}{2}\left[1 - \left(\sum_{i=0}^{6} b_i x^i\right)^{-16}\right],$$

kde $x > 0$ a

$$\begin{aligned} b_0 &= 1 \\ b_1 &= 0,049\ 867\ 347 \\ b_2 &= 0,021\ 141\ 006 \\ b_3 &= 0,003\ 277\ 626 \end{aligned}$$

Kvantilová funkcia

Prvotná aproximácia:

$$z = w - \frac{\sum\limits_{i=0}^{2} a_i w^i}{\sum\limits_{i=0}^{3} b_i w^i},$$

kde $\alpha \in (0, 0{,}5\rangle$, $w = \sqrt{-2\ln(1-\alpha)}$.

Taylorov polynóm má tvar

$$u_\alpha^{(n)} = z + \sum_{i=1}^{n} \frac{c_i(z)}{\mathrm{i}!}\left[\frac{\alpha - \Phi(z)}{\varphi(z)}\right]^i$$

kde

$$\begin{aligned} c_1(z) &= 1 \\ c_2(z) &= z \\ c_3(z) &= 2z^2 + 1 \\ &\vdots \\ c_{i+1}(z) &= \mathrm{i}\cdot z\cdot c_i(z) + \frac{\mathrm{d}}{\mathrm{d}z}c_i(z) \end{aligned}$$

Generátory náhodných čísiel

Lineárne rekurentné generátory

Nech $n = k, k+1, \dots, a_1, a_2, \dots, a_k, b$ sú nezáporné celé čísla a $X_0, \dots, X_{k-1}$ sú počiatočné hodnoty.

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + \dots + a_k X_{n-k} + b \pmod m.$$

Potom definujeme $U_i = \frac{X_i}{m}$, teda $U_i \in \langle 0, 1)$.

Najčastejšie používané

$$X_n = aX_{n-1} + b \pmod m \tag{1}$$

Tvrdenie 0.1 Nech b a m sú nesúdeliteľné, $a \equiv 1 \pmod p$ pre každý prvočiniteľ prvočíselného rozkladu m. Nech $a \equiv 4 \pmod 4$, ak m je násobok 4. Potom má generátor $X_n = aX_{n-1} + b \pmod m$ plnú periódu.

Tvrdenie 0.2 Nech b a m sú nesúdeliteľné, $a \equiv 1 \pmod p$ pre každý prvočiniteľ prvočíselného rozkladu m. Nech $a \equiv 4 \pmod 4$, ak m je násobok 4. Potom má generátor $X_n = aX_{n-1} + b \pmod m$ plnú periódu.

0.0.1 Seriálna korelácia

Pre generátor (1) s plnou periódou platí pre korelačný koeficient susedných čísiel

$$\rho_{X_{n-1}X_n} \doteq \frac{1}{a}$$

Maticové výpočty

Miera singularity

$$\mathrm{CN}_1(\mathrm{A}) = \|\mathrm{A}\| \cdot \|\mathrm{A}^{-1}\|,$$

kde $\|A\| = \max_j \sum_i |a_{ij}|$.

$$\mathrm{CN}_2(\mathrm{A}) = \frac{\lambda_{\max}(\mathrm{AA}^\top)}{\lambda_{\min}(\mathrm{AA}^\top)}$$

Väčší koeficient = matica je bližšie k singularite.

Inverzia symetrickej matice

- **inverzia trojuholníkovej matice** z rovnice

$$CC^{-1} = I$$

- **LU rozklad** $A = U^\top U$, kde U je horná trojuholníková matica. Potom

$$A^{-1} = U^{-1}(U^\top)^{-1}$$

- **Choleského metóda**. Rozložíme

$$A = U^\top DU$$

  Potom

$$A^{-1} = U^{-1}D^{-1}(U^\top)^{-1}$$

Symetrická sústava lineárnych rovníc

Riešime sústavu $Ax = b$, t. j.

$$x = A^{-1}b.$$

Rozložíme $A^{-1} = U^{-1}D^{-1}(U^\top)^{-1}$. Riešime dve sústavy

$$\begin{aligned} U^\top y &= b \\ Ux &= D^{-1}y \end{aligned}$$

Kontrola správnosti:
Nech $r = Ax_0 - b$ je reziduum približného riešenia sústavy. Spresnenie riešenia sa získa riešením sústavy $Ad = r$ a potom $x = x_0 - d$.

MP-inverzia

$$A^+ = A^\top U_r \Lambda_r^{-1} U_r^\top,$$

kde

$$AA^\top = (U_R \quad U_{m-r}) \begin{matrix} \Lambda_r 0 \\ 00 \end{matrix}$$

Podmienky:

- $AA^+A = A$
- $A^+AA^+ = A^+$
- $(AA^+)^\top = AA^+$
- $(A^+A)^\top = A^+A$

Alternatívne MP inverzia podľa matice $A^\top A$:

$$A^+ = U_r \Lambda_r^{-1} U_r^\top A^\top$$

Ostatné rozdelenia

Metóda inverznej transformácie

Nech $F(x)$ je distribučná funkcia, nech

$$F^{-1}(x) = \sup_{x \in (0,1)} \{x : F(x) \leq y\}$$

nech $\mathbb{Y} \sim R(0,1)$. Potom náhodná veličina $\mathbb{X} = F^{-1}(\mathbb{Y})$ má rozdelenie s distribučnou funkciou $F(x)$.

Zamietacia metóda

Nech pre hustotu pravdepodobnosti $f(x)$ platí

$$f(x) \begin{cases} = 0 & \text{ak } x \notin \langle a, b \rangle \\ \leq c & \text{pre } x \in \langle a, b \rangle \end{cases}$$

Nech $\mathbb{Y} \sim R(a,b)$ a $\mathbb{Z} \sim R(0,c)$ sú nezávislé náhodné veličiny. Potom náhodná veličiny $\mathbb{X} = \mathbb{Y}$ za podmienky, že $f(\mathbb{Y}) \geq \mathbb{Z}$ má rozdelenie s hustotou $f(x)$.
Problémy: predpoklad konečnosti nosiča, predpoklad ohraničenosti hustoty.

Algoritmus:

1. generuj nezávislé veličiny $\mathbb{U}_1, \mathbb{U}_2 \sim R(0,1)$
2. definuj $\mathbb{Y} = a + \mathbb{U}_1(b-a)$, $\mathbb{Z} = c\mathbb{U}_2$
3. ak $f(\mathbb{Y}) \geq \mathbb{Z}$, potom polož $\mathbb{X} = \mathbb{Y}$, inak zamietni dvojicu $\mathbb{U}_1, \mathbb{U}_2$.

Efektívnosť:

Všeobecne:

$$E = \frac{\text{počet nezamietnutých dvojíc}}{\text{počet všetkých dvojíc}}$$

Stredná hodnota efektívnosti:

$$\begin{aligned} E_0 &= \frac{\text{plocha pod hustotou}}{\text{plocha obdĺžnika}} = \frac{1}{c(b-a)} \\ &= P(\text{dvojicu nezamietame}) \end{aligned}$$

Všeobecná zamietacia metóda

Nech $\mathbb{Y}$ má rozdelenie s hustotou $g(x)$ a nech pre hustotu $f(x)$ existuje konštanta c, že

$$\forall x : f(x) \leq c \cdot g(x).$$

Nech podmienené rozdelenie náhodnej veličiny $\mathbb{Z}$ pri danom $\mathbb{Y}$ je $R(0, c \cdot g(\mathbb{Y}))$. Potom náhodná veličina $\mathbb{X} = \mathbb{Y}$ za podmienky, že $\mathbb{Z} \leq f(\mathbb{Y})$, má rozdelenie s hustotou $f(x)$.

Algoritmus:

1. generuj náhodné číslo $\mathbb{Y}$ z rozdelenia s hustotou $g(x)$.
2. generuj náhodné číslo $\mathbb{Z} \sim R(0, c \cdot g(\mathbb{Y}))$.
3. ak $f(\mathbb{Y}) \geq \mathbb{Z}$, potom polož $\mathbb{X} = \mathbb{Y}$, inak zamietni dvojicu $\mathbb{U}_1, \mathbb{U}_2$.

Podielová metóda

Nech $G = \left\{(y, z) : 0 \leq y \leq \sqrt{f\left(a + b\frac{z}{y}\right)}\right\}$, kde $f(x)$ je hustota pravdepodobnosti, $a \in \mathbb{R}$, $b > 0$. Nech $(\mathbb{Y}, \mathbb{Z})$ má rovnomerné rozdelenie na ohraničenej množine $H \supset G$.
Položme $\mathbb{W} = a + b\frac{\mathbb{Z}}{\mathbb{Y}}$. Potom náhodná veličina $\mathbb{X} = \mathbb{W}$, za podmienky, že $\mathbb{Y}^2 \leq f(\mathbb{W})$, má rozdelenie s hustotou $f(x)$.
Maximálne hodnoty súradníc má útvar na hranici

$$y = \sqrt{f\left(a + b\frac{z}{y}\right)}$$

Nech $t = a + b\frac{z}{y}$. Potom

$$y \in \left\langle 0, \sup \sqrt{f(t)} \right\rangle = \langle 0, y^* \rangle$$

Potom

$$z = \frac{t-a}{b} y = \frac{t-a}{b} \sqrt{f(t)},$$

teda

$$z = \left\langle \inf \frac{t-a}{b} \sqrt{f(t)}, \sup \frac{t-a}{b} \sqrt{f(t)} \right\rangle$$

teda

$$z \in \langle z_*, z^* \rangle.$$

Algoritmus:

1. generuj nezávislé veličiny $\mathbb{Y} \sim R(0, y^*)$ a $\mathbb{Z} \sim R(z_*, z^*)$.
2. definuj $\mathbb{W} = a + b\frac{\mathbb{Z}}{\mathbb{Y}}$
3. ak $\mathbb{Y}^2 \leq f(\mathbb{W})$, potom polož $\mathbb{X} = \mathbb{W}$, inak zamietni dvojicu $\mathbb{Y}, \mathbb{Z}$.

Kompozičná metóda

Nech $H(y)$ je distribučná funkcia a $\{g_y(x)\}$ systém hustôt pravdepodobností závislých od parametra y. Nech $\mathbb{Y}$ má rozdelenie s distribučnou funkciou $H(y)$ a nech podmienené rozdelenie náhodnej veličiny $\mathbb{X}$ pri danom $\mathbb{Y}$ má hustotu $g_{\mathbb{Y}}$. Potom $\mathbb{X}$ má rozdelenie s hustotou

$$f(x) = \int_{-\infty}^{\infty} g_y(x) \, \mathrm{d}H(y).$$

Špeciálne metódy

Generátor normálneho rozdelenia

Nech $\mathbb{U}_1, \mathbb{U}_2 \sim R(0, 1)$ sú nezávislé. Potom

$$\begin{aligned} \mathbb{X}_1 &= \sqrt{-2\ln(\mathbb{U}_1)} \sin(2\pi \mathbb{U}_2) \\ \mathbb{X}_2 &= \sqrt{-2\ln(\mathbb{U}_1)} \cos(2\pi \mathbb{U}_2) \end{aligned}$$

sú nezávislé náhodné veličiny s rozdelením $N(0, 1)$. Kvôli zníženiu závislosti je vhodné používať dva nezávislé generátory, napr:

$$\begin{aligned} U'_{n+1} &= 5U'_n \pmod{2^{35}} \\ U''_{n+1} &= 131U''_n \pmod{2^{35}} \end{aligned}$$

Gama rozdelenie

Generátor pre prirodzené p

Nech $\alpha > 0$, $p \in \mathbb{N}$ a $\mathbb{U}_1, \ldots, \mathbb{U}_n \sim R(0,1)$ sú nezávislé NV. Potom NV

$$\mathbb{X} = -\frac{1}{\alpha} \sum_{i=1}^{p} \ln \mathbb{U}_i$$

má rozdelenie $\Gamma(\alpha, p)$.

Generátor pre $p \in (0,1)$

Nech $\alpha > 0$, $p \in (0,1)$ a $\mathbb{Y} \sim B(p, 1-p)$ a $\mathbb{Z} \sim \Gamma(1,1)$ sú nezávislé NV. Potom NV

$$\mathbb{X} = \frac{1}{\alpha} \mathbb{Y}\mathbb{Z}$$

má rozdelenie $\Gamma(\alpha, p)$.

Vo všeobecnosti vieme $p \in \mathbb{R}^+$ rozdeliť na celú a zlomkovú časť

$$p = [p] + \{p\},$$

a využiť

$$\Gamma(\alpha, p) = \Gamma(\alpha, [p]) + \Gamma(\alpha, \{p\})$$

Generátor beta rozdelenia z $R(0,1)$

Nech $\alpha, \beta > 0$. Nech $\mathbb{Y}$, $\mathbb{Z} \sim R(0,1)$ sú nezávislé. Potom NV

$$\mathbb{X} = \frac{Y^{\frac{1}{\alpha}}}{Y^{\frac{1}{\alpha}} + Z^{\frac{1}{\alpha}}}$$

za podmienky, že $Y^{\frac{1}{\alpha}} + Z^{\frac{1}{\alpha}} < 1$ má rozdelenie $B(\alpha, \beta)$.

Beta rozdelenie

Generátor beta rozdelenia z Γ-rozdelenia

Nech $\alpha, \beta \in \mathbb{N}$. Nech $\mathbb{Y} \sim \Gamma(1, \alpha)$ a $\mathbb{Z} \sim \Gamma(1, \beta)$ sú nezávislé. Potom NV:

$$\mathbb{X} = \frac{\mathbb{Y}}{\mathbb{Y} + \mathbb{Z}}$$

má rozdelenie $B(\alpha, \beta)$.

Ostatné rozdelenia

χ^2-rozdelenie

Platí

$$\chi_n^2 = \Gamma\left(\frac{1}{2}, \frac{n}{2}\right)$$

$$\chi_n^2 = \sum_{1}^{n} (N(0,1))^2$$

Studentovo t-rozdelenie

V prípade

$$t_n = \frac{N(0,1)}{\sqrt{\chi_n^2}} \sqrt{n}$$

treba používať na čitateľa a menovateľa rôzne generátory.

Podielovou metódou s parametrami $a = 0$, $b = \frac{\sqrt{3}}{2}$ generujeme rozdelenie t_3 s hustotou

$$t_3(x) = \frac{2}{\pi\sqrt{3}} \left(1 + \frac{x^2}{3}\right)^{-2}.$$

V zovšeobecnenej zamietacej metóde hustota $t_3(x)$ dominuje pri koeficiente $c = \frac{8}{9}\sqrt{\frac{3\pi}{2\mathrm{e}}}$ rozdelenie

$$t_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

F-rozdelenie

Nech $\mathbb{Y} \sim B\left(\frac{\alpha}{2}, \frac{\beta}{2}\right)$. Potom NV

$$\mathbb{X} = \frac{\beta\mathbb{Y}}{\alpha(1 - \mathbb{Y})}$$

má rozdelenie $F(\alpha, \beta)$.

log-normálne rozdelenie

$$\mathrm{LN}(\mu, \sigma^2) = \mathrm{e}^{\mathrm{N}(\mu, \sigma^2)}$$

Simulácie

Nech $\mathbb{X}_1, \ldots, \mathbb{X}_n$ je náhodný výber z rozdelenia s distribučnou funkciou $F(x)$. Ak $\mathbb{X}_{(1)} \leq \mathbb{X}_{(2)} \leq \ldots \leq \mathbb{X}_{(n)}$ je prislúchajúci usporiadaný náhodný výber, potom definujeme empirickú distribučnú funkciu vzťahom

$$F_n(x) = \begin{cases} 0 & \text{ak } x \leq \mathbb{X}_{(1)} \\ \frac{k}{n} & \text{ak } \mathbb{X}_{(r)} < x \leq \mathbb{X}_{(k+1)} \\ & \quad\text{a} \\ & \mathbb{X}_{(r-1)} \leq \mathbb{X}_{(r)} = \ldots = \mathbb{X}_{(k)} < \mathbb{X}_{(k+1)} \\ 1 & \text{ak } \mathbb{X}_{(n)} < x \end{cases}$$

Približný výpočet integrálu

Majme funkciu $f(x)$, kde $\forall x : f(x) \geq 0$. Rátame

$$I = \int_0^1 f(x)\,\mathrm{d}x$$

Všeobecná metóda

Ak $\mathbb{U}_1, \ldots, \mathbb{U}_n \sim R(0,1)$ sú nezávislé NV, potom $f(\mathbb{U}_1), \ldots, f(\mathbb{U}_n)$ sú tiež nezávislé NV a platí

$$E(f(\mathbb{U}_i)) = I$$

Teda sú to nestranné odhady hodnoty I. Aritmetický priemer je potom tiež nestranným odhadom hodnoty I.

$$\theta_i = \frac{1}{n}\sum_{i=1}^{n} f(\mathbb{U}_i)$$

Zamietacia metóda

Nech $\forall x \in (0,1) : f(x) \leq c$. Generujeme náhodné body $(\mathbb{U}_1, \mathbb{V}_1), \ldots, (\mathbb{U}_n, \mathbb{V}_n)$ s rovnomerným rozdelením na $(0,1) \times (0,c)$. Teda $\mathbb{U}_i$ tvoria náhodný výber z $R(0,1)$ a $\mathbb{V}_j$ náhodný výber z $R(0,c)$. Definujeme funkciu

$$g(x,y) = \begin{cases} 0 & \text{ak } f(x) < y \\ 1 & \text{ak } f(x) \geq y \end{cases}$$

Potom nestranným odhadom I je

$$\theta_2 = c \cdot \frac{1}{n}\sum_{i=1}^{n} g(\mathbb{U}_i, \mathbb{V}_i) = c \cdot \frac{n_1}{n},$$

kde n_1 je počet vygenerovaných bodov ležiacich pod grafom $f(x)$.

Metóda výberu podľa dôležitosti

Nech $h(x)$ je hustota pravdepodobnosti taká, že $h(x) = 0$ akk $f(x) = 0$. Ak máme náhodný výber $\mathbb{X}_1, \ldots, \mathbb{X}_n$ z rozdelenia s distribučnou funkciou $H(x)$, potom nestranným odhadom integrálu I je

$$\theta_3 = \frac{1}{n}\sum_{i=1}^{n} \frac{f(\mathbb{X}_i)}{h(\mathbb{X}_i)}$$

Metóda stratifikovaného výberu

Majme interval $(0,1)$ rozdelený na k podintervalov

$$0 = a_0 < a_1 < \ldots < a_k = 1$$

Na i-tom podintervale generujme n_i čísiel s rovnomerným rozdelením. Ak $\sum_{i=1}^n n_i = n$ a $\mathbb{U}_1, \ldots, \mathbb{U}_n$ je náhodný výber z $R(0,1)$, potom definujeme

$$\mathbb{U}_{ij} = \mathbb{U}_{n_1 + \ldots + n_{i-1} + j}$$

a nestranný odhad I je

$$\theta_4 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(a_i - a_{i-1}) \cdot \frac{1}{n_i} f\Big(a_{i-1} + (a_i - a_{i-1})\mathbb{U}_{ij}\Big)$$

Metóda riadiacich premenných

Nech $g(x)$ je funkcia, ktorá dobre aproximuje $f(x)$ taká, že integrál

$$J = \int_0^1 g(x)\,\mathrm{d}x$$

vieme spočítať analyticky. Neznámym odhadom I je potom

$$\theta_5 = J + \frac{1}{n}\sum_{i=1}^{n}(f(\mathbb{U}_i) - g(\mathbb{U}_i)).$$

Metóda antitetických premenných

Nech pre $g(x)$ platí

$$\int_0^1 g(x)\,\mathrm{d}x = I$$

Ak $\mathbb{U}_1, \ldots, \mathbb{U}_n \sim R(0,1)$ sú nezávislé NV, potom nestranný odhad I je

$$\theta_6 = \frac{1}{2n}\sum_{i=1}^{n}(f(\mathbb{U}_i) + g(\mathbb{U}_i)).$$

Ak $f(x)$ je monotónna, potom vezmeme $g(x) = f(1-x)$ a

$$\theta_6 = \frac{1}{2n}\sum_{i=1}^{n}(f(\mathbb{U}_i) + f(1-\mathbb{U}_i)).$$

Bootstrap

Jackknife

Ak $\mathbb{X}_1, \dots, \mathbb{X}_n$ je náhodný výber a $S(\mathbb{X}_1, \dots, \mathbb{X}_n)$ je vypočítaná štatistika, potom séria odhadov

$$\begin{gathered} S_{n-1}(\mathbb{X}_2, \dots, \mathbb{X}_n), \\ S_{n-1}(\mathbb{X}_1, \mathbb{X}_3 \dots, \mathbb{X}_n), \\ \vdots \\ S_{n-1}(\mathbb{X}_1, \dots, \mathbb{X}_{n-1}) \end{gathered}$$

umožňuje vypočítať empirické rozdelenie štatistiky S (fakticky S_{n-1}).

Bootstrap / MOV

Nech $\mathbb{X}_1, \dots, \mathbb{X}_n$ je náhodný výber z neznámeho rozdelenia F. Nech $\mathbb{X} = (\mathbb{X}_1, \dots, \mathbb{X}_n)$ a $\mathbf{x} = (x_1, \dots, x_n)$ je realizácia náhodného výberu. Odhadnime rozdelenie $R = R(\mathbb{X}, F)$ náhodnej veličiny na základe pozorovaného vektora $\mathbf{x}$.
Použijeme štatistiku

$$R(\mathbb{X}, F) = t(x) - \theta(F),$$

kde $\theta(F)$ je parameter, ktorý nás zaujíma (napr. stredná hodnota) a $t(x)$ je jeho odhad (napr. aritmetický priemer).

- skonštruujeme výberové rozdelenie $\hat{F} = F_n(x)$, kde $F_n(x)$ je empirická distribučná funkcia
- pri pevnom $\hat{F}$ urobíme náhodný výber rozsahu n z $\hat{F}$. Označíme ho $\mathbb{X}^*$ a jeho realizáciu $\mathbf{x}^*$. Tento výber sa označuje ako MOV-výber.
- rozdelenie $R(\mathbb{X}, F)$ aproximujeme MOV-rozdelením $R^* = R(\mathbb{X}^*, \hat{F})$.

MOV-rozdelenie aproximujeme napr. empirickým rozdelením metódou Monte Carlo. Robíme opakované náhodné výbery $\mathbb{X}^*$ o rozsahu n z $\hat{F}$ a z ich realizácií $\mathbf{x}_1, \dots, \mathbf{x}_N$ zostrojíme empirické rozdelenie $G_N(R^*)$.