Centrální limitní věta a chyba metody Monte-Carlo

- metoda Monte-Carlo spočívá ve výpočtu určité veličiny (např. určitého integrálu) pomocí průměrování náhodných veličin X_i s určitými pravděpodobnostními rozděleními přičemž lze očekávat, že chyba bude s rostoucí počtem n náhodných veličin X_i klesat jako $\frac{1}{\sqrt{n}}$. To souvisí s centrální limitní větou z teorie pravděpodobnosti

- <u>základní pojmy</u>
  - <u>pravděpodobnostní funkce</u> náhodné diskrétní veličiny X

    $P[X=x] = P(x)$ udává pravděpodobnost, že náhodná veličina X bude mít hodnotu x

    musí tedy být $P(x) > 0$ pro $\forall x$ a $\sum_x P(x) = 1$
  - <u>distribuční funkce</u> náhodné diskrétní veličiny X

    $$F(x) = P[X \leq x] = \sum_{t \leq x} P(t) \quad , \quad P[x_1 \leq X \leq x_2] = F(x_2) - F(x_1)$$

    jde o neklesající funkci a $0 \leq F(x) \leq 1$
  - <u>rozdělení pravděpodobnosti spojité náhodné veličiny X</u>

    se určuje pomocí <u>hustoty pravděpodobnosti</u> $\varrho(x) > 0$

    splňující $\int_{\Omega} \varrho(x)\,dx = 1$, kde Ω je definiční obor X (obvykle určitý interval)

    pravděpodobnost, že X bude mít hodnotu v $\langle x_1, x_2 \rangle$ je

    $$P[x_1 \leq x \leq x_2] = \int_{x_1}^{x_2} \varrho(x)\,dx = F(x_2) - F(x_1)$$

    kde $F(x)$ je <u>distribuční funkce</u> spojité náh. veličiny X

    $$F(x) = P[X \leq x] = \int_{-\infty}^{x} \varrho(t)\,dt \quad \text{a tedy} \quad \varrho(x) = \frac{dF(x)}{dx}$$

    pravděpodobnost nalezení přesné hodnoty x pro spojitou náhodnou veličinu X je nulová

- <u>střední hodnota</u> (vážený průměr daného rozdělení)

$$E(X) = \sum_{i=1}^{n} x_i p_i \quad \text{pro diskrétní náhodnou veličinu}$$

$$= \int_{\Omega} x\, \varrho(x)\, dx \quad \text{pro spojitou}$$

- <u>rozptyl, neboli střední kvadratická odchylka</u> (též variance)

$$\sigma^2(X) = D(X) = var(X) =$$

člen $-2x_i E(x)$ dá $-2E(x)^2$

$$= \sum_{i=1}^{n} [x_i - E(X)]^2 p_i = \sum_i x_i^2 p_i - E(X)^2 \quad \text{(diskrétní)}$$

$$= \int_{\Omega} [x - E(X)]^2 \varrho(x)\, dx = \int_{\Omega} x^2 \varrho(x)\, dx - E(X)^2$$

$$= E(X^2) - E(X)^2$$

základní vlastnosti (pro dvě nezávislé veličiny X a Y)

$$E(X+c) = c + E(X) \qquad D(X+c) = D(X)$$

$$E(cX) = cE(X) \qquad D(cX) = c^2 D(X)$$

$$E(X+Y) = E(X) + E(Y) \qquad D(X+Y) = D(X) + D(Y)$$

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

např. $D(X+Y) = \iint (x+y-E(X+Y))^2 \varrho(x,y)\, dx\, dy =$ pro nezávislé X a Y

$$= \iint (x - E(X) + y - E(Y))^2 \varrho_x(x) \varrho_y(y)\, dx\, dy =$$

$$= \iint [(x-E(X))^2 + 2\underbrace{(x-E(X))(y-E(y))}_{0} + (y-E(Y))^2] \varrho_x(x) \varrho_y(y)\, dx\, dy$$

$$= D(X) + D(Y)$$

- <u>střední hodnota funkce náhodné veličiny</u>

pokud $Y = f(X)$ je náhodná veličina, která je funkcí náhodné veličiny X, pak

$$E(Y) = E(f(X)) = \int_{\Omega} f(x) \varrho(x)\, dx \quad \left(\text{nebo } = \sum_i f(x_i) p_i\right)$$

a obecně $E(f(X)) \neq f(E(X))$

- příklady rozdělení

  - rovnoměrné rozdělení na intervalu $\langle a,b\rangle$

$$\varrho(x) = \begin{cases} \frac{1}{b-a} & \text{na } \langle a,b\rangle \\ 0 & \text{jinde} \end{cases} \qquad (\text{pro interval } \langle 0,1\rangle \text{ je } \varrho(x)=1)$$

$$E(X) = \frac{a+b}{2}, \quad D(X) = \sigma^2(X) = \frac{(b-a)^2}{12}, \quad F(X) = \frac{x-a}{b-a} \text{ na } \langle a,b\rangle$$

  - normální rozdělení $N(\mu,\sigma^2)$ na $\langle -\infty,\infty\rangle$

$$\varrho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad E(X)=\mu, \quad D(X)=\sigma^2$$

distribuční funkce

$$F(x) = \int_{-\infty}^{x} \varrho(t)\,dt = \frac{1}{2}\left[1+\mathrm{Erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right]$$

kde $\mathrm{Erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$ (error function)

a tedy $P[x_1 \le X \le x_2] = \int_{x_1}^{x_2} \varrho(t)\,dt = \frac{1}{2}\left[\mathrm{Erf}\left(\frac{x_2-\mu}{\sqrt{2\sigma^2}}\right) - \mathrm{Erf}\left(\frac{x_1-\mu}{\sqrt{2\sigma^2}}\right)\right]$

pravidlo 3σ

$$P[\mu-3\sigma \le X \le \mu+3\sigma] = \mathrm{Erf}\left(\frac{3}{\sqrt{2}}\right) = 0{,}9973$$

a tedy náhodné měření téměř vždy leží v intervalu $\langle \mu-3\sigma, \mu+3\sigma\rangle$

- pravděpodobná chyba náhodné veličiny X

  - máme-li tedy jediné měření, pak s velkou pravděpodobností bude v intervalu $\langle \mu-3\sigma, \mu+3\sigma\rangle$ a absolutní chyba (od střední hodnoty μ) bude řádu $r \approx 0{,}67\,\sigma$

kde r je dáno $P[\mu-r \le X \le \mu+r] = 0{,}5$

r se nazývá pravděpodobná chyba měření X

- centrální limitní věta

  - máme-li n nezávislých náhodných veličin $X_1, \dots, X_n$ které mají stejné rozdělení pravděpodobnosti tj. jejich střední hodnoty $E(X_i) = m$ pro $i = 1, \dots, n$ a rozptyly $D(X_i) = b^2$, $i = 1, \dots, n$ jsou stejné

  pak pro jejich součet $S_n = X_1 + \dots + X_n$ (opět náhodná veličina) platí (viz $E(X+Y) = E(X) + E(Y)$ a $D(X+Y) = D(X) + D(Y)$)

$$E(S_n) = nm \quad \text{a} \quad D(S_n) = n b^2$$

  - nějme dále normální rozdělení $N(\mu, \sigma^2)$ kde $\mu = nm$ a $\sigma^2 = nb^2$ s hustotou pravděp. $\varrho_n(x)$

  - pak centrální limitní věta říká, že

$$P[s_1 \leq S_n \leq s_2] \simeq \int_{s_1}^{s_2} \varrho_n(x)\, dx$$

  pro libovolné s_1 a s_2 a všechna dostatečně velká n.

  - neboli rozdělení součtů S_n velkého množství nezávislých náhodných veličin je přibližně normální

Pozn: existují zobecnění pro mnohem slabší podmínky

  např. veličiny $X_1, \dots, X_n$ nemusí být nutně nezávislé a stejného rozdělení

  důležité je, aby jedna z veličin nehrála příliš důležitou roli

  díky tomu se normální rozdělení často objevuje v přírodě na výsledek má obvykle vliv řada náhodných faktorů

- to, že dostaneme normální rozdělení se $\sigma = \sqrt{n}\, b$ je důležité pro odhad chyby veličiny S_n v závislosti na n, pravděpodobná chyba S_n bude úměrná $\sqrt{n}$, a také pro odhad chyby MC

• odhad chyby metody Monte-Carlo

- počítáme-li určitou veličinu jako průměr náhodných veličin $X_1, \dots, X_n$, tj. $A_n = (X_1 + \dots + X_n)/n$, se středními hodnotami $E(X_i) = m$ a rozptyly $D(X_i) = b^2$ pak pravděpodobnostní rozdělení A_n bude přibližně normální s $E(A_n) = \sum_{i=1}^{n} \frac{E(X_i)}{n} = m$

  a $D(A_n) = \sum_{i=1}^{n} \frac{D(X_i)}{n^2} = \frac{b^2}{n}$ neboli $\sigma = \frac{b}{\sqrt{n}}$

  a tedy $P\left[m - \frac{3b}{\sqrt{n}} \leq A_n \leq m + \frac{3b}{\sqrt{n}}\right] \approx 0{,}9973$

  neboli $P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| < \frac{3b}{\sqrt{n}}\right] \approx 0{,}9973$

  pravděpodobná chyba A_n tedy je přibližně $0{,}67\,\sigma = 0{,}67\,\frac{b}{\sqrt{n}}$ a lze ji tedy očekávat řádu $\frac{1}{\sqrt{n}}$

- pozor! nejde o omezení chyby shora, jde o <u>pravděpodobnou</u> chybu, chyba může být ve skutečnosti větší, ale pravděpodobnost toho je malá

Integrace pomocí metody Monte-Carlo

- jak si ukážeme má tato metoda především velký význam pro <u>vícerozměrné integrály</u>, protože její chyba <u>nezávisí na dimenzi</u>, ale pouze na počtu použitých bodů N, tj. chybu lze vždy očekávat $\sim \frac{1}{\sqrt{N}}$ (zde budeme používat N pro celkový počet bodů, n bude počet bodů v 1dim.)
- pro $d \lesssim 4$ (dimenze integrálu) je obvykle mnohem přesnější použít kvadraturní vzorce, např. Gaussovy kvadratury nebo Newton-Cotesovy vzorce

- pro jednoduchost si však ilustrujme tuto metodu na jednorozměrném integrálu

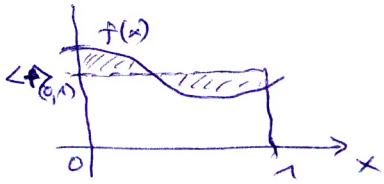
$$I = \int_0^1 f(x)\,dx$$

Pozn.: máme-li interval $\langle a,b\rangle$, lze v 1D použít lineární mapování $\int_a^b f(x)\,dx = (b-a)\int_0^1 \tilde{f}(y)\,dy$

kde $\tilde{f}(y) = f[(b-a)y - a]$

- tento integrál můžeme považovat za průměr funkce $f(x)$ na intervalu $\langle 0,1\rangle$ a aproximovat ho pomocí

$$I \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$



kde x_i jsou „rovnoměrně" (zatím) rozložené v intervalu $\langle 0,1\rangle$, tj. jde o náhodné veličiny X_i s rovnoměrným rozdělením na intervalu $\langle 0,1\rangle$

- $f(X_i)$ je tedy náhodná veličina s určitým rozdělením, její střední hodnota je ovšem

$$E(f(X_i)) = \int_0^1 f(x)\underbrace{g(x)}_{=1}\,dx = \int_0^1 f(x)\,dx = I$$

a rozptyl σ_f je tím menší, čím je $f(x)$ „hladší", méně se mění (pro konstantní $f(x)$ je $\sigma_f = 0$)

a díky centrální limitní větě dostaneme pro odhad chyby

$$\sigma_I \approx \frac{\sigma_f}{\sqrt{N}}$$

- v metodě MC se σ_f odhaduje pomocí

$$\sigma_f^2 \approx \langle f^2\rangle - \langle f\rangle^2 = \frac{1}{N}\sum_{i=1}^{N} f(x_i)^2 - \left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right)^2$$

- porovnání s lichoběžníkovým pravidlem s chybou $\mathcal{O}(h^2) \sim \mathcal{O}(n^{-2})$ v 1D

dimenze	chyba	celkový počet bodů
1	$\sim n^{-2} \approx N^{-2}$	$N = n$
2	$\sim n^{-2} \approx N^{-1}$	$N = n^2$
3	$\sim n^{-2} \approx N^{-2/3}$	$N = n^3$
4	$\sim n^{-2} \approx N^{-1/2}$	$N = n^4$

srovnatelné s MC

n je počet bodů kvadratury v jedné dimenzi

N je celkový počet bodů, kde je třeba vyčíslit $f(x)$

- ve vícerozměrném případu ovšem musíme občas řešit složité hranice oblasti, zde má metoda Monte Carlo velkou výhodu, neboť stačí pouze určit, zda zvolený bod patří či nepatří do oblasti přes kterou integrujeme

- protože chyba integrace pomocí metody Monte Carlo závisí na σ_f, je výhodné, aby σ_f byla co nejmenší
  - výsledek lze zpřesnit vhodnou váhovou funkcí $w(x)$ která je kladná a splňuje $\int_0^1 w(x)\,dx = 1$

  pak přepíšeme

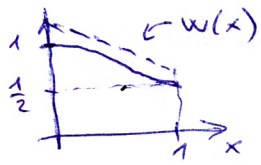
$$I = \int_0^1 dx\, w(x) \frac{f(x)}{w(x)} = \int_0^1 dy \frac{f(x(y))}{w(x(y))} = \int_0^1 dy\, g(y)$$

  neboť $y(x) = \int_0^x dx'\, w(x')$, $\frac{dy}{dx} = w(x)$

  $y(0) = 0$, $y(1) = 1$

  problém je ovšem s inverzí funkce $y(x)$, i když najdeme vhodné $w(x)$

<u>Příklad:</u> $f(x) = \frac{1}{1+x^2}$



lze např. vzít $w(x) = \frac{4}{3} - \frac{2}{3}x$ (koeficienty zvoleny tak, aby $\int_0^1 w(x)dx = 1$)

máme $y(x) = \frac{1}{3}x(4-x)$

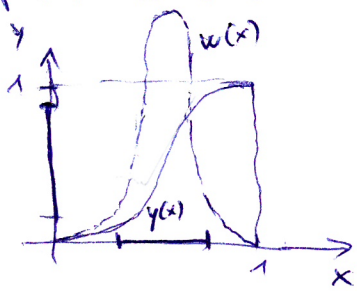
neboli $x(y) = 2 - \sqrt{4-3y}$

tato jednoduchá transformace dá o řád lepší výsledek!

- protože najít $x(y)$ je u vícerozměrných integrálů prakticky nemožné ($w(x)$ je Jacobián $\left|\frac{\partial \vec{y}}{\partial \vec{x}}\right|$)

  je výhodnější se na vztah $dy = w(x)\,dx$ dívat jako na změnu rovnoměrného rozdělení $\rho(y) = 1$ na rozdělení $w(x)$

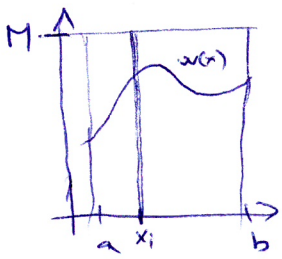
- je-li y rovnoměrné na $\langle 0,1\rangle$, pak x bude rozdělené podle $w(x)$ a tedy preferují se x tam, kde má $w(x)$ velké hodnoty neboť tam $y(x)$ rychle roste



- místo změny proměnných tedy budeme stále používat x, ale namísto rovnoměrného rozdělení pro x potřebujeme umět generovat náhodnou veličinu s daným rozdělením pravděpodobnosti $w(x)$

Generování náhodných veličin

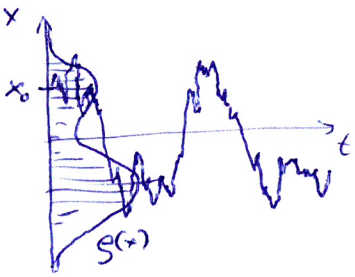
- v počítači máme většinou k dispozici pseudonáhodný generátor čísel na intervalu $\langle 0,1\rangle$ s rovnoměrným rozdělením
- existují speciální postupy pro určitá rozdělení
  např. pro normální rozdělení ve 2D přechodem k polárním souřadnicím (viz notebook v Mathematica)
- obecně lze použít von Neumannovu metodu či Metropolisův-Hastingsův algoritmus
- von Neumannova metoda
  - generujeme dvojici náhodných čísel s rovnom. rozdělením a podle $w(x)$ rozhodneme, zda vygenerované x_i použijeme



  - nejme $\xi_i, \eta_i \in \langle 0,1\rangle$ náhodná
    ξ_i namapujeme na $\langle a,b\rangle$, $x_i = a + \xi_i(b-a)$
    η_i namapujeme na $\langle 0,M\rangle$, $y_i = M\eta_i$
    (M zvoleno tak, aby $w(x) < M$ pro $\forall x \in \langle a,b\rangle$)
  - pokud $y_i \leq w(x_i)$, pak x_i použijeme
    pokud $y_i > w(x_i)$, pak x_i zahodíme
  - pravděpodobnost přijetí hodnoty v $\langle x, x+dx\rangle$ je úměrná $w(x)dx$ neboť ξ_i je rovnoměrně rozložené na $\langle a,b\rangle$

Metropolisův-Hastingsův algoritmus

- základní myšlenka: jít „náhodnou" procházkou v prostoru v závislosti na hustotě pravděpodobnosti, tj. jsme-li v bodě x_k, učiníme náhodný krok o délce $\delta x_k = \xi_k \delta$, kde $\xi_k \in \langle -1, 1 \rangle$ a δ je fixní parametr procházky (musí být vhodně nastaven, aby bylo případně možno překlenout oblasti s malou hustotou),



a tento krok učiníme, pokud $\varrho(x_k + \delta x_k) \geq \varrho(x_k)$, nebo je-li $\varrho(x_k + \delta x_k) < \varrho(x_k)$, pak vygenerujeme náhodné $\eta \in \langle 0, 1 \rangle$ a krok učiníme, pokud $\varrho(x_k + \delta x_k) / \varrho(x_k) > \eta$.

- pokud krok neuděláme, pak zůstaneme a $x_{k+1} = x_k$.
- v d-rozměrném prostoru bude algoritmus vypadat takto:

procedura vracející další bod, jsme-li v x(:)

```
r(:) = random(:)                    ← generátor na intervalu <0,1>
xtrial(:) = x(:) + δ(2r(:) - 1)     ← posun na <-1,1>
ratio = ϱ(xtrial) / ϱ(x)
if ( ratio ≥ 1 or ratio > random) then
        x = xtrial
return x(:)
```

- v limitě dlouhé náhodné procházky tento algoritmus opravdu generuje body rozložené podle $\varrho(x)$
  - označíme-li $N_n(x)$ hustotu náhodných nezávislých chodců startujících v různých bodech po n krocích, pak počet chodců jdoucích z x do y v následujícím kroku je

$$\Delta N(x) = N_n(x) P(x \to y) - N_n(y) P(y \to x) =$$

$$= N_n(y) P(x \to y) \left[\frac{N_n(x)}{N_n(y)} - \frac{P(y \to x)}{P(x \to y)} \right]$$

kde $P(x \to y)$ je pravděpodobnost, že chodec přejde z x do y

- rovnováha nastává pro

$$\frac{N_n(x)}{N_n(y)} = \frac{P(y \to x)}{P(x \to y)}$$

a lze ukázat, že pro velká n jde $N_n(x) \to N_e(x)$, což je rovnovážné rozdělení chodců v x

- pro M-H algoritmus je $N_e(x) \sim g(x)$, neboť $P(x \to y)$ lze vyjádřit jako

$$P(x \to y) = T(x \to y)\, A(x \to y)$$

kde $T(x \to y)$ je pravděpodobnost náhodného vykročení z x do y a $A(x \to y)$ je pravděpodobnost přijmutí tohoto kroku

- pokud tedy bude

$$\frac{g(x)}{g(y)} = \frac{P(y \to x)}{P(x \to y)} = \frac{T(y \to x)\, A(y \to x)}{T(x \to y)\, A(y \to x)}$$

pak $N_e(x)$ bude $\sim g(x)$

- pro vhodně zvolené δ, kdy lze y dosáhnout v jediném kroku, tj. $y \in \langle x-\delta, x+\delta \rangle$, pak $T(x \to y) = T(y \to x)$

a dále pro $g(x) \geq g(y)$ bude

$$A(y \to x) = 1 \quad \text{a} \quad A(x \to y) = \frac{g(y)}{g(x)}$$

a pro $g(x) \leq g(y)$ bude

$$A(x \to y) = 1 \quad \text{a} \quad A(y \to x) = \frac{g(x)}{g(y)}$$

tedy celkem v každém případě

$$\frac{A(y \to x)}{A(x \to y)} = \frac{g(x)}{g(y)}$$

a tedy vskutku $\dfrac{N_e(x)}{N_e(y)} = \dfrac{g(x)}{g(y)}$

Generátory (pseudo) náhodných čísel (random numbers)

- deterministické programy generující posloupnost čísel, které splňují jistá kritéria (testy) náhodnosti
- při stejných počátečních podmínkách generují tytéž náhodná čísla
  - poč. podm. = random seed (náhodné seminko)
- vzhledem ke konečné aritmetice počítačů a deterministickému chování nejde o náhodná čísla, ale o pseudonáhodná
- většinou generována jako posloupnost čísel $\{x_0, x_1, x_2 \dots\}$ → nutná periodicita, kvůli konečné aritmetice ⇒ perioda generátoru = nejmenší k takové, že $x_i = x_{i+k}$ pro $\forall i$

Základní generátory

- lineární kongruentní generátor (LCG) - velmi rychlý algoritmus

$$x_{i+1} = (a x_i + c) \bmod m$$

maximální (optimální) perioda ~ většinou menší

speciální případ - multiplikativní LCG, když $c = 0$ (navržen D.H. Lehmerem)

byl velmi často používán v historii, ale má často zásadní nedostatky pro aplikace ve více dimenzích, neboť n-tice $(x_i, x_{i+1}, \dots x_{i+n-1})$ se nacházejí na relativně malém počtu $(n-1)$-rozm. nadploch

Př. Nechvalně známý generátor RANDU

$a = 65539$, $c = 0$, $m = 2^{31}$ splňuje rekur. vztah $x_{i+2} = (2^{16}+3)x_{i+1} \bmod 2^{31}$

$$x_{i+2} = (2^{32} + 6 \cdot 2^{16} + 9) x_i \bmod 2^{31} = (6 x_{i+1} - 9 x_k) \bmod 2^{31}$$

a lze ukázat, že $\forall$ trojice bodů se nacházejí na 15 paralelních rovinách

Marsaglia: Random Numbers Fall Mainly in the Planes, PNAS 61 (1968) 25

Teorem pro MLGG: Pokud $c_1, \dots c_n$ jsou přirozená čísla taková, že

$$c_1 + c_2 a + c_3 a^2 + \dots c_n a^{n-1} = 0 \pmod m$$

pak $\forall$ body $P_1 = (u_1, u_2, \dots u_n)$, $P_2 = (u_2, \dots u_{n+1})$, ... kde $u_i = x_i/m$ leží v jedné z paralelních nadploch

$$c_1 z_1 + c_2 z_2 + \dots + c_n z_n = 0, \pm 1, \pm 2, \dots$$

Navíc nanejvýše $|c_1| + |c_2| + \dots + |c_n|$ těchto rovin protíná krychli $0 < z_1 < 1, \dots, 0 < z_n < 1$ a vždy lze vybrat taková $c_1, \dots c_n$, že $\forall$ body budou v méně než $(n!\,m)^{1/n}$ nadplochách.