Kapitola 5

Metody typu Monte-Carlo

Pojem metoda typu Monte-Carlo (často se používá zkratka MC) zahrnuje širokou třídu metod, které nějakým způsobem využívají generátorů náhodných čísel. Zde se budeme věnovat především využití metody Monte-Carlo k integraci funkcí mnoha proměnných a k řešení určitého typu okrajových úloh.

Jako jednoduchý motivační příklad si uvedeme metodu nalezení čísla π metodou Monte-Carlo.

PŘ 1: Budeme postupovat tak, že postupmě vygenerujeme N náhodných vektorů ze čtverce velikosti 2×2 (stačí generovat souřadnice x a y jako dvě náhodná čísla z intervalu $\langle -1, 1 \rangle$). Spočteme počet vektorů N_π, které leží v jednotkové kružnici, tj. splňují relaci $x^2 + y^2 < 1$. Podíl N_π/N se limitně blíží poměru plochy jednotkové kružnice a plochy čtverce 2×2, tj.

$$\pi \simeq= 4N_\pi/N.$$

Abychom porozuměli tomu, jaké chyby se doupouštíme touto aproximací, budeme potřebovat některé:

5.1 Základní pojmy matematické statistiky

Základním pojmem bude náhodná proměnná. Příkladem náhodné proměnné je číslo na horní ploše hrací kostky, které může nabývat hodnot 1,2,..,6 s pravděpodobnostmi 1/6 (pokud není kostkou podvodnou, pak můžou pravděpodobnosti nabývat šesti různých hodnot a půjde tedy o jinou náhodnou proměnnou). Obecně si náhodnou proměnnou můžete představovat jako nějaký generátor, který na požádání vrátí číslo. Přitom se pokusíme definovat pravidla jak budou tato čísla vypadat, když jich generujeme velké množství.

5.1.1 Diskrétní náhodná proměnná

Disktétní náhodná proměnná x je definovaná pomocí množiny hodnot $\{x_1, x_2, ..., x_n\}$, kterých může nabývat a pomocí pravděpodobností s jakou se každá z nich může objevit $\{p_1, p_2, ..., p_n\}$. Přitom pravděpodobnosti $p_i > 0$ a $p_1 + p_2 + ... + p_n = 1$. Pravděpodobnosti p_i můžeme chápat jako funkci přirozeného čísla i, které říkáme **distribuční funkce**.

Příkladem náhodné proměnné je číslo na horní ploše hrací kostky, které může nabývat hodnot 1,2,..,6 s pravděpodobnostmi 1/6 (pokud není kostkou podvodnou, pak můžou pravděpodobnosti nabývat šesti různých hodnot a půjde tedy o jinou náhodnou proměnnou).

Budeme předpokládat, že náhodnou proměnnou můžeme změřit kdykoli změřit (například hodit kostkou) čímž získáme některou z jejích možných hodnot $\xi \in \{x_1, ..., x_n\}$. Opakovaným měřením získáme soubor hodnot

$$S = \{\xi_1, \xi_2, ..., \xi_N\},$$

neboli, tzv. **statistický soubor**. To že statistický soubor byl získán měřením zadané náhodné proměnné se odráží v platnosti axiomu, nazývaného **zákon velkých čísel**. Označme N_i počet výskytů hodnoty x_i ve statistickém souboru S. Zákon velkých čísel potom říká, že limita podílu N_i/N pro $N \to \infty$ je dána pravděpodobností p_i.

Střední honotou náhodné proměnné x nazveme výraz

$$\langle x \rangle \equiv \sum_{i=1}^{n} x_i p_i.$$

V tomto výrazu pro střední hodnotu používáme veličin x_i, p_i definujících náhodnou proměnnou x. O tomto způsobu výpočtu budeme mluvit jako o *středování přes distribuční funkci*. S pomocí zákona velkých čísel můžeme výraz pro střední hodnotu přepsat jako *středování přes statistický soubor*:

$$\langle x \rangle = \lim_{N\to\infty} \sum_{i=1}^{n} x_i \frac{N_i}{N} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \xi_k.$$

Funkcí $f(x)$ **náhodné proměnné** x nazveme náhodnou proměnnou f, která nabývá hodnot $\{f(x_1), f(x_2), ..., f(x_n)\}$ s pravděpodobnostmi $\{p_1, ..., p_n\}$. Její střední hodnota potom je

$$\langle f \rangle = \sum_{i=1}^{n} f(x_i) p_i = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(\xi_k).$$

Důležitou charakteristikou náhodné proměnné x jsou její **momenty** μ_n, definované jako

$$\mu_n \equiv \langle x^n \rangle = \sum_n x_i^n p_i = \lim_{N\to\infty} \frac{1}{N} \sum_k \xi_k^n,$$

kde jsme opět použili středování přes distribuční funkci i přes statistický soubor.

Momenty náhodné proměnné pro malé n mají názorný význam. Nultý moment $\mu_0 = 1$ udává normování distribuční funkce p_i. První moment $\mu_1 = \langle x \rangle$ je prostě střední hodnota náhodné proměnné a pomocí druhého momentu určíme tzv. **varianci** neboli rozptyl náhodné proměnné

$$\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = \mu_2 - \mu_1^2,$$

která udává jak moc se typicky hodnoty ξ_i získané měřením náhodné veličiny odchylují od průměru $\langle x \rangle$.

5.1.2 Spojitá náhodná proměnná

Jednoduchým zobecněním můžeme zavést pojem spojité náhodné proměnné x. Ta je opět zadána množinou hodnot Ω (například interval $\langle a, b \rangle$), kterých nabývá s hustotou pravděpodobnosti danou distribuční (nezápornou) funkcí $p(x)$ definovanou na množině Ω. Pravěpodobnost, že náhodná veličina x nabude měřením hodnoty ξ z intervalu $I = \langle x_1, x_2 \rangle$ (obecně z podmnožiny $I \subset \Omega$) tedy je

$$P_I = \int_{x_1}^{x_2} p(x) dx$$

a speciálně

$$\int_\Omega p(x)dx = 1.$$

Poznamenejme, že náhodná proměnná může být dokonce vektorová veličina tj. $\Omega \subset \mathbf{R}^d$. Opakovaným měřením spojité náhodné proměnné můžeme opět vytvořit náhodný soubor $\{\xi_i\}_{i=1}^N$. Zákon velkých čísel pak formulujeme tak, že pro počet výskytů N_I hodnoty $\xi \in I$ ve statistickém souboru platí

$$\lim_{N\to\infty} \frac{N_I}{N} = P_I \equiv \int_I p(x)dx.$$

Podobně jako v předchozím případě můžeme definovat momenty náhodné proměnné

$$\mu_n \equiv \langle x^n \rangle = \int_\Omega x^n p(x)dx = \lim_{N\to\infty} \sum_k \xi_k^n,$$

její varianci σ_x, případně funkci f, jejíž střední hodnota je

$$\langle f(x) \rangle = \int_\Omega f(x)p(x)dx = \lim_{N\to\infty} \sum_k f(\xi_k).$$

Všiměte si, že vzorce vyjádřené pomocí středování přes statistický soubor jsou stejné pro diskrétní i pro spojitou náhonou proměnnou, jediné co se liší jsou hodnoty, jakých mohou nabývat prvky souboru.

5.1.3 Příklady náhodných proměnných a jejich vlastnosti

PŘ. 2: Diskrétní náhodná proměnná daná hodem hrací kostky má střední hodnotu $(1+2+3+4+5+6)/6 = 7/2$ a její druhý moment je dán součtem $(1+4+..+36)/6 = 91/6$ a variance tedy je $\sigma_x = \sqrt{91/6 - 49/4} \doteq 2.92$

PŘ. 3: Spojitá náhodná proměnná $x \in \langle 0, 1\rangle$ s rovnoměrným rozdělením $p(x) = 1$ má střední hodnotu a varianci

$$\langle x \rangle = \int_0^1 x\mathrm{d}x = \tfrac{1}{2}$$

$$\sigma_x^2 = \int_0^1 x^2\mathrm{d}x - (\tfrac{1}{2})^2 = \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{1}{12}$$

PŘ. 4: Normální (Gaussovská) náhodná proměnná může nabývat libovolné reálné hodnoty s hustotou pravděpodobnosti danou vzorcem

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

Konstanty jsou voleny tak, že střední hodnota této náhodné proměnné je $\langle x \rangle = a$ a její variance je $\sigma_x = \sigma$. Pravděpodobnost nalezení naměření hodnoty $\xi \in \langle x_1, x_2\rangle$ lze vyjádřit pomocí tzv. *"error function"* definované integrálem

$$Erf(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\mathrm{d}t,$$

takže

$$P_{\langle x_1,x_2\rangle} = Erf(...x_2) - Erf(...x_1).$$

Důležitými speciálními případy jsou, tzv. pravidlo 3σ:

$$P_{\langle a-3\sigma,a+3\sigma\rangle} = 0.997$$

a tzv. pravděpodobná chyba $r = 0.6745\sigma¿$:

$$P_{\langle a-r,a+r\rangle} = 0.5.$$

PŘ. 5 (Součet náhodných proměnných): Nechť x_1 a x_2 jsou dvě (nezávislé) náhodné proměnné s rovnoměrným rozdělením na intervalu $\langle 0,1\rangle$. Nyní můžeme definovat novou náhodnou proměnnou $x = x_1 + x_2$, jejíž hodnotu získáme tak, že generujeme hodnoty pomocí x_1 a x_2 a pak je sečteme. Výsledkem je náhodná proměnná, která nabývá hodnot z intervalu $\langle 0,2\rangle$ s pravděpodobností $p(x) = 1 - |x-1|$ (rozmyslete si proč).

5.1.4 Centrální limitní věta

Ekvivalence středování funkce náhodné proměnné přes distribuční funkci a přes statistický soubor lze využít k výpočtu integrálů. Pro praktické použití je třeba znát odhad chyby takového výpočtu a k tomu potřebujeme vědět jak rychle se limity

$$\lim_{N\to\infty} \frac{1}{N} \sum_k f(\xi_k) \to \langle f\rangle$$

ve vzorcích pro střední hodnoty nabývá. Na tuto otázku odpovídá

Centrální limitní věta: Mějme identické, statisticky nezávislé náhodné veličiny $x_1, \ldots, x_N$ charakterizované střední hodnotou $\langle x_1\rangle = ... = \langle x_N\rangle = \mu$ a variancí $\sigma_{x_1} = ... = \sigma_{x_N} = \sigma$. Pak náhodná proměnná $s_N \equiv x_1 + ... + x_N$ se pro $N \to \infty$ blíží normální proměnné se stření hodnotou $\langle s_N\rangle = N\mu$ a s variancí $\sigma_s^2 = N\sigma^2$.

Poznámky:

- Všiměte si, že centrální limitní věta je definována pomocí součtu náhodných proměnných (srovn. PŘ. 5).
- Věta používá pojem nezávilých náhodných veličin, k jehož přesnější definici se ještě vrátíme. Zatím si uveďme intuitivní definici: nezávislé náhodné veličiny jsou takové, že měření jedné z nich není nikterak ovlivněno hodnotou nalezenou měřením druhé. Můžeme si představit například dvě hrací kostky (nebo dva po sobě jdoucí hody touže kostkou), ale ne $x-$ovou složku a velikost r náhodného vektoru (neboť ty například splňují nerovnost $x \le r$).
- Důležitost věty pro nás spočívá v tom, že N po sobě jdoucích měření $\xi_1, ..., \xi_N$ proměnné x můžeme chápat jako nezávislé veličiny a výraz

  $$M_N = \frac{1}{N} \sum_{i=1}^{N} \xi_i$$

  podle centrální limitní věty představuje náhodnou proměnnou se střední hodnotou $\langle x\rangle$ a s variancí $\sigma_M = \sigma_x/\sqrt{N}$. Navíc pro velká N jde skoro o normální rozdělení, tj. např. pravděpodobná chyba je $0.67\sigma_x/\sqrt{N}$ a maximální (v 99.7% případů) chyba je $3\sigma_x/\sqrt{N}$.

Obecné schéma metody MC: Úlohu nalézt veličinu v formulujeme jako úlohu na nalezení střední hodnoty $v = \langle x \rangle$ vhodné náhodné veličiny x. Monte Carlo výpočet veličiny v pak spočívá v generování statistického souboru $\{\xi_1, ..., \xi_N\}$ náhodné veličiny x. Hodnota v pak získáme ze vzorce

$$v \simeq \frac{1}{N} \sum_{i=1}^{N} \xi_i$$

s chybou zmenšující se jako $1/\sqrt{N}$. Pokud známe varianci veličiny x víme, že nalezená hodnota leží vně intervalu $\langle v - 3\sigma/\sqrt{N}, v + 3\sigma/\sqrt{N} \rangle$ jen s pravděpodobností 0.3%.

PŘ. 1 (Pokračování): Formulaci problému nalezení čísla π, která se hodí k analýze pomocí centrální limitní věty provedeme následovné. Máme náhodnou vektorovou proměnnou (x, y). Ta nabývá hodnot ze čtverce $\langle -1, 1 \rangle \times \langle -1, 1 \rangle$ s hustotou pravděpodobnosti $p(x, y) = 1/4$. Definujeme funkci f této náhodné proměnné následovně

$$f = \begin{cases} 1 & \text{pro } x^2 + y^2 < 1, \\ 0 & \text{pro } x^2 + y^2 > 1. \end{cases}$$

Podíl N_π/N není nic jiného než střední hodnota této veličiny spočítaná průměrováním přes statistický soubor. Nyní si spočteme momenty náhodné veličiny f:

$$\begin{aligned} \mu_1 &= \langle f \rangle = \int_\square f(x,y)p(x,y)\mathrm{d}x\mathrm{d}y = \int_\bigcirc \tfrac{1}{4}\mathrm{d}x\mathrm{d}y = \frac{\pi}{4}, \\ \mu_2 &= \langle f^2 \rangle = \int_\square f(x,y)^2 p(x,y)\mathrm{d}x\mathrm{d}y = \int_\bigcirc \tfrac{1}{4}\mathrm{d}x\mathrm{d}y = \frac{\pi}{4}, \end{aligned}$$

a tedy $\sigma_f = \sqrt{\mu_2 - \mu_1^2} = \sqrt{4\pi - \pi^2}/4 \doteq 0.41$. Podle centrální limitní věty tedy podíl N_π/N jde ke střední hodnotě $\mu_1 = \pi/4$ přičemž chyba se škáluje jako $\sigma_f/\sqrt{N}$. Očekáváme tedy, že pro výpočet π na tři platné cifry budeme potřebovat generovat řádově milion náhodných bodů.

DEMO V PYTHONU: ukázat výsledky pro 2D příklad. Nechat někoho, ať ukáže jak je potřeba modifikovat pro nalezení objemu 13-ti dimenzionální koule. Po MC nalezení výsledku říct analytický vzorec a porovnat.

PŘ. 5 (zobecnění příkladu 1): Tentokrát budeme generovat náhodný dvousložkový vektor s rovnoměrným rozdělením na obrazci s plochou S_0 a budeme chtít zjistit plochu obrzce $S \subset S_0$ … analyzovat chybu a σ. Ukázat, že je vhodné, aby S nebylo moc větší než S_0.

5.2 Integrace metodou Monte-Carlo

Základní úlohou, které je vhodná pro výpočet metodou Monte-Carlo je integrace funkce. Začneme formulací pro integrování funkce jedné proměnné přes konečný interval $\langle a, b \rangle$

$$I \equiv \int_a^b f(x)\mathrm{d}x = |b-a| \int_a^b f(x)p dx = \langle f(x) \rangle. \tag{5.1}$$

Definici integrálu I jsme vynásobili a vydělili číslem $|b-a|$ a současně jsme si všimli, že $p = 1/|b-a|$ je distribuční funkce náhodné proměnné x s rovnoměrným rozdělením na intervalu $\langle a, b \rangle$, takže integrál lze chápat jako výpočet stření hodnoty funkce $f(x)$. Integrál I tudíž můžeme počítat středováním přes statistický soubor $\{\xi_i\}_{i=1}^N$ pro náhodnou proměnnou x

$$I_N = \frac{|b-a|}{N}\sum_{i=1}^{N} f(\xi_i). \tag{5.2}$$

Podle centrální limitní věty potom I_N konverguje k I s chybou úměrnou $\sigma_f/\sqrt{N}$. Je zajímavé porovnat tento vzorec s kvadraturními pravidly z minulého semestru, obecně

$$I_N = \sum_{i=0}^{N} w_i f(x_i), \tag{5.3}$$

kde x_i a w_i jsou uzly a váhy příslušné kvadratury. Například v případě (složeného) lichoběžníkového pravidla jsou $x_i = a + ih$, kde $w_i = h = |b-a|/N$ (pro všechna i kromě prvního a posledního). Dosazením těchto hodnot zjistíme, že vzorce (5.2) a (5.3) jsou velmi podobné. Jen v případě lichoběžníkového pravidla jsou uzly pravidelně uspořádány s rozestupem h, zatímco v metodě MC jde o náhodné body. Jaký má tento rozdíl vliv na přesnost kvadratury? V minulém semestru jsme si ukázali, že lichoběžníkové pravidlo má chybu $O(h^2) = O(1/N^2)$. To je mnohem rychlejší konvergence, ale výhody metody MC vyniknou ve vyšších dimenzích, neboť odhad chyby metody Monte-Carlo závisí jen na velikosti statistického souboru na ne nadimenzionalitě. V případě lichoběžníhového pravidla ve vyšších dimenzích (použijeme Fubiniovu větu a v každé dimenzi použijeme stejný kvadraturní vzorec) se chyba stále škáluje jako $O(h^2)$, ale počet kvadraturních bodů roste s dimenzí jako $N = 1/h^d$. Chyba lichoběžníkového pravidla pro výpočet integrálů funkcí d proměnných se dá tedy psát jako $O(h^2) = O(N^{-2/d})$. Pro $d = 4$ jsou výpočty metodou Monte-Carlo a lichoběžníkovým pravidlem řádově stejně efektivní, pro dimenze vyšší vyhrává MC. Jak jsme viděli v příkladě s výpočtem plochy kruhu, MC výpočet má další velkou výhodu. Nemusíme totiž složitě určovat meze každé proměnné (jako ve Fubiniově větě), ale stačí testovat, zda generovaný náhodný bod leží v integrační oblasti či ne.

5.2.1 Substituce v integraci metodou Monte-Carlo

Někdy je výhodné při integraci metodou Monte-Carlo zvolit trochu jinou formulaci. Podobně jako v minulém semestru při zavedení Gaussovy kvadratury můžeme integrál přepsat následovně (bez újmy na obecnsti volíme integraci přes interval $\langle 0,1\rangle$ čehož lze vždy dosáhnout lineární substitucí)

$$I = \int_0^1 f(x)\mathrm{d}x = \int_0^1 \frac{f(x)}{w(x)} w(x)\mathrm{d}x, \tag{5.4}$$

kde $w(x)$ je libovolná funkce definovaná na $x \in \langle 0,1\rangle$ nabývající nezáporných hodnot. Tím dostáváme jinou interpretaci výpočtu stejné veličiny I, neboť poslení uvedený výraz lze interpretovat jako střední hodnotu funkce $g(x) = f(x)/w(x)$ přes náhodnou proměnnou s distribuční funkcí $w(x)$ (předpokládáme, že $\int w(x) = 1$ jinak lze integrál přenormovat). Tato úprava může být výhodná, pokud má funkce $g(x)$ menší varianci σ_g.

PŘ. 6: Integrál z Koonina, kde dostáváme o řád menší chybu při sejné velikosti N statistického souboru.

Výše zmíněná metoda předpokládá, že umíme generovat náhodnou proměnnou s předem zadanou rozdělovací funkcí. V jedné dimenzi si můžeme pomoci následujícím trikem. Definujeme