

1 Teoria

Não temos no momento a definição exata de metodos de Monte Carlo. No nosso curso metodos de Monte Carlo vamos chamar *metodos computacionais de solução de problemas matemáticas usando a modelagem de variáveis aleatórias e estimação estatística de características deles*. Pela definição se-destaca:

1. métodos *computacionais*, e não analíticos
2. podemos solucionar problemas não somente probabilísticas, mas também qualquer

Com a essa definição para metodos de Monte Carlo temos que adicionar as técnicas de aproximação estocástica, ou a busca estocástica, que tradicionalmente formam um conjunto separado de métodos. Mas as vezes as especialistas dessas áreas chamam os seus métodos de Monte Carlo.

A data oficial de "nascimento" de metodos Monte Carlo é 1949, com artigo titulado "Metodo Monte Carlo"^{MC}_[1]

Um jeito importante de construção de método Monte Carlo - é representar o problema como cálculo de esperança matemática: para calcular um valor escalar a temos que achar uma variável aleatória ξ tal que $E(\xi) = a$; assim gerando N valores independentemente de ξ_1, \dots, ξ_N de variável ξ , podemos aproximar o valor de a como

$$a \approx \frac{\xi_1 + \dots + \xi_N}{N}. \quad (1) \quad \boxed{1}$$

existe infinitamente muito variáveis com $E\xi = a$. A teoria de metodos tem que dar seguintes respostas:

1. como escolher um "boa" variável ξ para calculo de problema
2. como gerar valores ξ_1, ξ_2, \dots de variável ξ qualquer

Exemplo. Exemplo de calculo de área de um volume G em um paralelepipedo Π . Neste caso variável é ξ que possua valor V_{Π} de o ponto caíu em G e $\xi = 0$ em caso contrário.

Nota: O cálculo que se a gente colocar uma grade regular com p pontos em cada dimensão, então esses pontos essencialmente estão distribuidos em superfície de cubo, e não em volume: $(\frac{p-1}{p})^N = (1 - \frac{1}{p})^N = e^{N \ln(1-2/p)} \approx e^{-2N/p} \rightarrow 0$, quando $N \rightarrow \infty$.

Convergencia de método

Sabemos que pelo Lei dos Grandes Números, se as variáveis ξ_i em fórmula $(\frac{1}{N})$ são independentes e identicamente distribuidos e com a esperança matemática finita, então a média aritmetica deles converge para a esperança a em probabilidade:

$$\bar{\xi}_N = \frac{\xi_1 + \dots + \xi_N}{N} \implies a, \text{ em probabilidade, quando } N \rightarrow \infty. \quad (2) \quad \boxed{2}$$

Assim, o estimador $\bar{\xi}_N$ podemos usar sempre quando existe esperança $E\xi = a$. (Lembramos que para esperança existir, tem que existir $E(|\xi|)$).

Erro do método

Supomos daqui para frente que existe a variância de ξ e ela é finita

$$Var(\xi) = E(\xi^2) - (E\xi)^2. \quad (3) \quad \boxed{3}$$

De curso de probabilidade sabemos que a sequencia de variáveis i.i.d. com variância finita obedeça o teorema central do limite: para quaisquer $x_1 < x_2$

$$\lim_{N \rightarrow \infty} P\left\{x_1 < \frac{1}{\sqrt{NVar(\xi)}} \sum_{i=1}^N (\xi_i - a) < x_2\right\} = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-t^2/2} dt. \quad (4) \quad \boxed{4}$$

escolhendo $x_2 = -x_1 = x$ obtemos

$$\lim_{N \rightarrow \infty} P\left\{\left|\frac{1}{N} \sum_{i=1}^N (\xi_i - a)\right| < x \sqrt{\frac{Var(\xi)}{N}}\right\} = 2\Phi(x) - 1.$$

Por isso, se N bastante grande, então

$$P\{|\bar{\xi}_N - a| < x\sqrt{Var(\xi)/N}\} \approx 2\Phi(x) - 1. \quad (5) \quad \boxed{5}$$

Dado $1 - \alpha/2 = \Phi(x)$, acharemos $z_{1-\alpha/2}$ e obtemos que a probabilidade da equação

$$|\bar{\xi}_N - a| < z_{1-\alpha/2}\sqrt{Var(\xi)/N} \quad (6) \quad \boxed{6}$$

aproximadamente igual à $1 - \alpha$ ($1 - \alpha$ é a probabilidade de confiança).

Exemplo

(Veja [2], p.70) Como a ilustração de convergência lenta do método de Monte Carlo consideramos o problema de estimação de número $\pi = 3.1415\dots$ usando estimador $Z := 4\mathbb{I}(U_1^2 + U_2^2 < 1)$, onde $U_i \sim Unif(0, 1)$. A variável $\mathbb{I}(U_1^2 + U_2^2 < 1)$ tem a distribuição de Bernoulli com a média $\pi/4$, assim a variância de Z é $4^2(\pi/4)(1 - \pi/4) \approx 2.70$. Assim, para obter o 3 em π , com a probabilidade, digamos de 95% precisaremos, $N \approx 1.96^2 2.70^2 / 0.5^2 \approx 112$, depois para o primeiro dígito "1", precisaremos $R \approx 11, 200$ e para segunda dígito "4" $R \approx 1, 120, 000$ ect.

Nota

Existe também os estimadores de velocidade de convergência para a distribuição normal em caso quando $\sigma^2 < \infty$ e existe o terceiro momento central $\beta_3 = E|\xi - a|^3 < \infty$:

$$\left| P(S_n - an > \sigma\sqrt{nx}) - \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du \right| \leq \frac{c_0\beta_3}{\sigma^3\sqrt{n}}$$

onde c_0 é um constante absoluto, é conhecido que $1/\sqrt{1\pi} \leq c_0 < 0.9051$

Erro provavel do metodo

Outro tipo de erro é

$$r_N = 0.6745\sqrt{\frac{Var(\xi)}{N}} \quad (7) \quad \boxed{e1}$$

o número é solução da equação $0.5 = \Phi(0.6745)$. O nome provável é porque

$$P\{|\bar{\xi}_N - a| < r_N\} \approx 1/2 \approx P\{|\bar{\xi}_N - a| > r_N\},$$

ou seja, os erros maior de que r_N e erros menor de que r_N são igualmente prováveis. Na prática r_N usa-se para estimar o ordem do erro. Assim, usando r_N nos estimamos a ordem de erro e usando (6) estimamos o limite superior do erro (com coeficiente de confiança de $1 - \alpha$).

estimação empírica de variância

em geral a variância $Var(\xi)$ é desconhecida. Usa-se estimador empírico para isso:

$$Var(\xi) \approx \frac{1}{N} \sum_{i=1}^N \xi_i^2 - \left(\frac{1}{N} \sum_{i=1}^N \xi_i \right)^2 \quad (8) \quad \boxed{8}$$

quando N pequeno usa-se estimador não viesado

$$Var(\xi) \approx \frac{1}{N-1} \sum_{i=1}^N \xi_i^2 - \left(\frac{\sum_{i=1}^N \xi_i}{N} \right)^2 \quad (9) \quad \boxed{8}$$

two-stage

(veja [2] p.71)

erro absoluto $|\bar{\xi}_N - a| < \varepsilon$;

erro relativo $|\bar{\xi}_N - a| < \varepsilon|a|$.

Dado erros calcularemos o tamanho de amostra q tem ser gerada:

$$N \approx \frac{z_{1-\alpha/2}^2 \sigma^2}{\varepsilon^2} \quad N \approx \frac{z_{1-\alpha/2}^2 \sigma^2}{\varepsilon^2 |a|^2}$$

1st estimamos σ desvio padrão de ξ usando primeiros N_{trial} (digamos $N_{trial} = 50$) ensaios, obtendo um estimador $\hat{\sigma}_{trial}$;

2nd estimamos usando resulatado do item anterior tamanho de amostra:

$$N \approx \frac{z_{1-\alpha/2}^2 \hat{\sigma}_{trial}^2}{\varepsilon^2} \quad N \approx \frac{z_{1-\alpha/2}^2 \hat{\sigma}_{trial}^2}{\varepsilon^2 \bar{\zeta}_{N_{trial}}^2}$$

estimação de erro sem cálculo de dispersão

Continuaremos a assumir que a variância é finita. Supomoes que não queremos (da causa da memoria) usar o estimador de variância empirica - não queremos calcular paralelamente a soma dos quadrados. Mesmo assim da para estimar o erro. Supomos que $N = mN_1$, onde $m \geq 3$ número inteiro não é grande. E supomos que N_1 grande bastante para usar o teorema central do limite para aritmética de N_1 variáveis. Calcularemos médias

$$\zeta_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} \xi_i, \quad \zeta_2 = \frac{1}{N_1} \sum_{i=1}^{N_1} \xi_{i+N_1}, \quad \dots, \quad \zeta_m = \frac{1}{N_1} \sum_{i=1}^{N_1} \xi_{i+N_1(m-1)}. \quad (10) \quad \boxed{9}$$

Usaremos o teorems (de Fisher?): se ζ_1, \dots, ζ_m são i.i.d. com distribuição normal e esperança a a variável

$$t = \sqrt{m-1} \frac{\bar{x} - a}{s}, \quad \text{onde } \bar{x} = \frac{1}{m} \sum_{k=1}^m \zeta_k, \quad s^2 = \frac{1}{m} \sum_{k=1}^m (\zeta_k - \bar{x})^2,$$

tem a distribuição de Student com $m-1$ graus de liberdades.

Em nosso caso $E\zeta_k = a$ e $\bar{x} = \bar{\xi}_N$ e $s^2 = (1/m) \sum_{k=1}^m (\zeta_k - \bar{\xi}_N)^2$. Logo obtemos

$$P\{|\bar{\xi}_N - a| < x \sqrt{s^2/(m-1)}\} \approx 2 \int_0^x p_{m-1}^{Student}(y) dy$$

assim a probabilidade de desigualdade

$$|\bar{\xi}_N - a| < t_{m-1, 1-\alpha} \sqrt{s^2/(m-1)} \quad (11) \quad \boxed{10}$$

aproximadamente igual à $1 - \alpha$.

Estimação (II) nos permite determinar o erro provável desse metodo:

$$r_N = t_{m-1, 0.5} \sqrt{s^2/(m-1)} \quad (12) \quad \boxed{11}$$

caso de variância infinita

A variância infinita não é um obstaculo para usar o lei dos grandes números, mas os estimadores de erro que nos apresentamos não funcionam mais. Caso quando a variância finita o erro diminua como $N^{-1/2}$. Os erros são piores em caso de variância infinita. A recomendação geral é não usar as vaiáveis com variância infinita.

2 Prática. Metodo de Monte Carlo para cálculo de integral.

simples Monte Carlo

Queremos calcular o integral sobre uma área $G \subset \mathbb{R}^2$, denotamos o ponto $P = (x, y)$ e uma área elemental $dP = dx dy$:

$$I = \int_G f(P) p(P) dP \quad (13) \quad \boxed{12}$$

onde $p(P)$ uma densidade qualquer de probabilidade definida em G , tal que $\int_G p(P) dP = 1$.

Notamos que integral qualquer em G limitada (área de G é finita) podemos considerar como um integral do tipo (II): considerando $f_1(P) = S_G \cdot f(P)$ e $p(P) = 1/S_G$ obtemos

$$\int_G f(P) dP = \int_G f_1(P) p(P) dP.$$

Para construir o Metodo de Monte Carlo para cálculo de integral do tipo $(\frac{12}{13})$, consideramos um ponto aleatório com a densidade $p(P)$ e variável aleatória $Z = f(P)$ cuja esperança é igual à integral

$$E(Z) = \int_G f(P)p(P)dP \quad (14) \quad \boxed{13}$$

Assim, se existe $E|Z|$ então existe a convergencia

$$\theta_N = \frac{1}{N} \sum_{i=1}^N f(P_i) \implies I, \text{ em probabilidade} \quad (15) \quad \boxed{14}$$

método geométrico

Supomos que $0 \leq f(P) \leq c$ para qualquer $P \in G$. Consideramos o volume $\tilde{G} = G \times (0, c)$. Ponto aleatório \tilde{Q} com a densidade $\tilde{p}(x, y, z) = p(x, y)/c$ em \tilde{G} . A projeção dele $Q = (\xi, \eta)$ é um ponto aleatório com a distribuição $p(x, y)$. Escolhendo N realizações independentes $\tilde{Q}_1, \dots, \tilde{Q}_N$ construímos estimador

$$\theta_N = c \frac{\nu}{N} \text{ onde } \nu = \#\{\text{vezes quando } \tilde{Q}_i \text{ ficou debaixo de superficie } f(P)\} \quad (16) \quad \boxed{15}$$

Exercicio

Provar que $\nu \sim B(N, p)$ onde $p = I/c$

Assim temos que $E\tilde{\theta}_N = I$. E convergencia $\tilde{\theta}_N \rightarrow I$ é um fato bem conhecido.

Notamos que o estimador pode ser representado como a media aritmetica

$$\tilde{\theta}_N = \frac{1}{N} \sum_{i=1}^N \tilde{Z}_i, \text{ onde } \tilde{Z} = \begin{cases} c, & \text{se } \zeta < f(\xi, \eta), \\ 0, & \text{se } \zeta \geq f(\xi, \eta) \end{cases}$$

comparação de precisão de métodos

Para primeiro método é necessário de existencia de disperção e para isso necessário e suficiente da existencia de segundo momento. Isso pode ser representado como uma condição de função pertencer à um conjunto de funções $L_2(G)$. Então a disperção

$$DZ = \int_G f^2(P)p(P)dP - I^2. \quad (17) \quad \boxed{d1}$$

Em método geometrico pegamos a média de variáveis tais que $E(\tilde{Z}^2) = c^2 P\{\zeta < f(\xi, \eta)\} = cI$, e a disperção

$$D\tilde{Z} = cI - I^2. \quad (18) \quad \boxed{d2}$$

Comparamos $(\frac{d1}{17})$ e $(\frac{d2}{18})$: se $0 \leq f(P) \leq c$, então

$$\int_G f^2(P)p(P)dP \leq c \int_G f(P)p(P)dP = cI, \implies DZ \leq D\tilde{Z}.$$

Exemplo

Calcular o integral

$$I = \int_0^1 e^x dx$$

estimações $(\frac{14}{15})$ e $(\frac{15}{16})$ são

$$\theta_N = \frac{1}{N} \sum_{i=1}^N e^{U_i}, \quad \tilde{\theta}_N = e \frac{\nu}{N},$$

onde ν é número de pares $(U_i, U'_i), \dots, (U_N, U'_N)$ tais que $e^{U'_i} < e^{U_i}$ (como sempre $U_1, \dots, U_N, U'_1, \dots, U'_N$ são i.i.d. uniformes em $(0, 1)$).

Usando $(\frac{d1}{17})$ calculamos a disperção $Var(Z)$:

$$Var(Z) = \int_0^1 e^{2x} dx - I^2 = \frac{1}{2}(e^2 - 1) - (e - 1)^2 = 0.2420$$

Usando $(\frac{d2}{18})$ calculamos disperção $Var(\tilde{Z})$:

$$Var(\tilde{Z}) = eI - I^2 = e - 1 = 1.7183$$

Custo computacional de Método de Monte Carlo

consideramos dois algoritmos Monte Carlo. Seja em um calcula-se a media de v.a. $\xi^{(1)} = \xi^{(1)}(U_1, \dots, U_{n_1})$ e em outro $\xi^{(2)} = \xi^{(2)}(U_1, \dots, U_{n_2})$ e staimadores são

$$\theta_N^{(1)} = \frac{1}{N} \sum_{i=1}^N \xi_i^{(1)} \quad \text{e} \quad \theta_N^{(2)} = \frac{1}{N} \sum_{i=1}^N \xi_i^{(2)}$$

Denotamos $D^{(1)} = Var(\xi^{(1)})$ e $D^{(2)} = Var(\xi^{(2)})$ e sejam $T^{(1)}, T^{(2)}$ são os tempos gastos em cálculo de um valor de $\xi^{(1)}$ e $\xi^{(2)}$ respectivamente. É natural supor um algoritmo mais eficiente se o tempo gasto em calculo de integral com dado erro provavel é menor.

Seja $r_N = \varepsilon$. Segundo (7) o tamanho de amostra tem que ser $N^{(1)} = D^{(1)}(0.6745/\varepsilon)^2$ e o tempo de cálculo igual à $T^{(1)}N^{(1)} = T^{(1)}D^{(1)}(0.6745/\varepsilon)^2$. e para o segundo $T^{(2)}N^{(2)} = T^{(2)}D^{(2)}(0.6745/\varepsilon)^2$.

Vamos chamar de *custo computacional* de um algoritmo Monte Carlo o produto $TVar(\xi)$. Neste caso O tempo gasto em cálculo de integral é proporcional ao laboriosidade de algoritmo.

Exemplo

Calcular o integral

$$I = \int_0^1 \sqrt[5]{x} dx = \frac{5}{6}.$$

Método simples. Podemos pensar que temos a densidade $p(x) \equiv 1$ de variável aleatória U em integral. Assim obtemos o estimador

$$I \approx \frac{1}{N} \sum_{i=1}^N \sqrt[5]{U_i}$$

A dispersão de variável $Z = \sqrt[5]{U}$ igual à

$$Var(Z) = \int_0^1 x^{2/5} dx - \left(\frac{5}{6}\right)^2 = \frac{5}{252}$$

Método geométrico. Como o $0 \leq x^{1/5} \leq 1$, então

$$I \approx \frac{1}{N} \sum_{i=1}^N \tilde{Z}_i,$$

onde

$$\tilde{Z}_i = \begin{cases} 1, & \text{se } U_i' < U_i^{1/5} \\ 0, & \text{caso contrário.} \end{cases}$$

a variância dela é

$$Var(\tilde{Z}) = \frac{5}{6} - \left(\frac{5}{6}\right)^2$$

Assim $Var(Z) < Var(\tilde{Z})$.

Calcularemos agora quantidade de operações para Z_i e \tilde{Z}_i .

- Para achar o número Z_i temos que calcular um U_i (três opeações) e depois calcular $U_i^{1/5}$ (número opeações aproximadamente 70)
- Para achar o número \tilde{Z}_i temos que calcular dois U_i e U_i' (seis operações) e verificar desigualdade $(U_i')^5 < U_i$ (aproximadamente 10 operações)

Assim $TVar(Z) \approx 1.4$ e $\tilde{T}Var(\tilde{Z}) \approx 1.4$ os dois métodos são computacionalmente iguais.

Mas podemos alterar o promeiro método

$$I \approx \sum_{i=1}^N \max\{U_i^{(1)}, U_i^{(2)}, U_i^{(3)}, U_i^{(4)}, U_i^{(5)}\}$$

neste caso $T \approx 20$ e $TVar(Z) \approx 0.40$, assim o custo computacional desse algoritmo ficará menor.

Referências

MC [1] Metropolis N., Ulam S.M. *The Monte Carlo method*. J. Amer. Statist. Assoc., **44**, N247, 335-341, 1949

As [2] S.Asmussen, P.W.Glynn *Stochastic Simulation. Algorithms and Analysis*. Springer. 2010

Ross [3] S.M.Ross (2004) *Simulation*.