Aula 4 (28/03/2011, 31/03/2011). Métodos de Monte Carlo II. Métodos de redução de variância.

1 Teoria

Na aula passada vimos que o erro de Monte Carlo é proprocional à $\sqrt{Var(\xi)/N}$. Melhorar o $\sqrt{N}$ nos não podemos, mas podemos diminuir a variância. Objetivo – achar contruir estimadores com a variância menor de que a variância sugerida pelo método simples de Monte Carlo.

1.1 Cálculo analitico parcial

se uma parte de integral conseguimos calcular analiticamente, entaõ isso pode diminuir a variância do método Monte Carlo. Devemos destacar que as vezes as abordagens desse tipo pode levar a custo maior computacional.

Extração de uma parte principal

Se conseguimos calcular analiticamente uma parte principal do integralo, então usaremos o Monte Carlo para calcular somente a "correção".

Queremos calcular

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

onde $f(P) \in L_2(G;p)$. Supomos que existe $h(P) \in L_2(G;p)$ que é "parecido"com $f(P)$ cujo integral

$$\int_G h(P)p(P)dP = C$$

é conhecido. Então em vez de usar o estimador

$$\theta_N = \frac{1}{N}\sum_{i=1}^{N} f(Q_i) \qquad (1)$$ smc

onde Q_i são pontos gerados pela distribuição $p(P)$, usaremos o estimador

$$\theta'_N = C + \frac{1}{N}\sum_{i=1}^{N}(f(Q_i) - h(Q_i)) = \frac{1}{N}\sum_{i=1}^{N} Z'_i. \qquad (2)$$ smc1

E a esperança de Z' é $E(Z') = C + I - C = I$. A variância

$$Var(Z') = \int_G (f(P) - h(P))^2 p(P)dP - (I - C)^2. \qquad (3)$$ var1

Se

$$\int_G (f(P) - h(P))^2 p(P)dP \le \varepsilon$$

então $Var(Z') \le \varepsilon$.

Exemplo

Consideramos um exemplo da aula passada

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

usaremos $e^x = 1 + x + \dots$, então usaremos $h(x) = x$ (constante 1 não influencia na variância). Temos estimador

$$\theta'_N = \frac{1}{2} + \frac{1}{N}\sum_{i=1}^{N}\Big(e^{U_i} - U_i\Big).$$

A variância de variável $Z' = 1/2 + e^U - U$ é

$$Var(Z') = \int_0^1 (e^x - x)^2 dx - (I - 1/2)^2 = (e-1)(5-e)/2 - 23/12 \approx 0.0487$$

muito menor de que a variância obtida na aula passada de 0.2420.

Integral pelo parte de área

Supomos que sabemos analiticamente calcular o integral em um subconjunto B de G:

$$\int_B f(P)p(P)dP = C, \qquad \int_B p(P)dP = c, \quad (0 < c < 1)$$

Assim o integral pode ser representado em forma seguinte

$$I = C + \int_{G \setminus B} f(P)p(P)dP. \tag{4}$$ int2

Seja Q' é distribuido em $G \setminus B$ com a densidade $p_1(P) = p(P)/(1-c)$. Consideramos a variável $Z' = C + (1-c)f(Q')$. Assim podemos approximar o integral pelo estimador

$$\theta'_N = C + \frac{1-c}{N}\sum_{i=1}^{N} f(Q'_i). \tag{5}$$ est2

t1 **Theorem 1** *Se existe a variância $Var(Z)$, então $Var(Z') \leq (1-c)Var(Z)$.*

Integração pelo subconjunto de variáveis

Ideia é simples – se analiticamente podemos calcular integral pelo parte das variáveis, e pela outra parte usaremos Metodo Monte Carlo. Neste caso a disperção diminua. Mas pode acontecer que a fórmula obtida depois de integração vai ser mais complicada ainda aumentadno custo computacional do método.

Queremos calcular $I = \int_G dP \int_{G'} f(P,P')p(P,P')dP'$, onde $p(P,P')$ é distribuição conjunta de pontos $Q \in G$ e $Q' \in G'$. Supomos que pelo P' podemos integrar e assim, podemos calcular a densidade de ponto Q: $p_1(P) = \int_{G'} p(P,P')dP'$, e também a função $f_1(P) = \int_{G'} f(P,P')p(P,P')dP'[p_1(P)]^{-1}$. Logo $I = \int_G f_1(P)p_1(P)dP$. Assim temos que fazer a média de variável $Z' = f_1(Q)$.

th2 **Theorem 2** *Se a variância $Var(Z')$ é finita, então $Var(Z') \leq Var(Z)$.*

1.2 Importance sampling

Ross
([3], capítulo 8.6) Queremos calcular o integral $I = \int_G f(P)dP$, onde G não é necessariamente limitada, e também o quadrado de $f(P)$ não é necessariamente integrável. Supomos somente que existe $\int_G |f(P)|dP > 0$.

Definição. A densidade $p(P)$ em G chamaremos *admissível em relação a $f(P)$*, se $p(P) > 0$ quando $f(P) \neq 0$.

Seja $G_0 = \{P : \ f(P) = 0\}$, $G^+ = G \setminus G_0$. Escolhemos uma densidade $p(P)$ admissível qualquer. Definimos

$$Z_0(P) = \begin{cases} f(P)/p(P), & \text{se } P \in G^+, \\ 0, & \text{se } P \in G_0 \end{cases}$$

Se Q um ponto aleatório em G com a densidade $p(P)$, então

$$E(Z_0(Q)) = \int_G Z_0(P)p(P)dP = \int_{G^+} f(P)dP = I_0$$

Método Monte Carlo construa estimador usando realizações independentes $Q_1, \dots, Q_N$

$$\theta_N = \frac{1}{N}\sum_{i=1}^{N} Z_0(Q_i).$$

erro depende da variância $Var(Z_0)$:

$$Var(Z_0) = \int_G Z_0^2(P)p(P)dP - I^2 = \int_{G^+} \frac{f^2(P)}{p(P)}dP - I^2. \tag{6}$$ is1

o valor depende da escolha $p(P)$. Peginta: escolher $p(P)$ que minimiza a varância $Var(Z_0)$.

is2 **Theorem 3** *A disperção minimal é atingível quando a densidade $p(P)$ proporcional a $|f(P)|$ e igual à*

$$Var(\hat{Z}_0) = \left(\int_G |f(P)|dP\right)^2 - I^2. \tag{7}$$ md

Prova. Se $p(P)$ é proporcional a $|f(P)|$ então

$$\hat{p}(P) = |f(P)|\Big(\int_G |f(P)|dP\Big)^{-1} \quad (8)$$

colocando em (6) temos $Var(Z_0) = Var(\hat{Z}_0)$.

Provaremos agora que para qualquer densidade admissível $p(P)$ a disperção dela $Var(Z_0) \geq Var(\tilde{Z}_0)$:

$$\Big(\int_G |f|dP\Big)^2 = \Big(\int_{G^+} |f|dP\Big)^2 = \Big(\int_{G^+} |f|p^{-1/2}p^{1/2}dP\Big)^2 \leq \int_{G^+} f^2p^{-1}dP \int_{G^+} p\ dP \leq \int_{G^+} f^2p^{-1}dP.$$

□

Nota

1. Se a função f não altera o seu sinal em G, então $Var(\hat{Z}_0) = 0$.
2. Usar a densidade (8) para cálculo de integral. Mas podemos inferir do teorema a seguinte conscusão. É bom escolher a densidade $p(P)$ mais proporcianal possível à fnção $|f|$. Esse método foi sugerido pelo Kahn H. [1].
3. Existe outra interpretação de razoavel escolha de $p(P)$ proporcional a $|f(P)|$: mais proxima $Z_0 = f(P)/p(P)$ a constante, menos variação $Var(Z_0(Q))$ ela tem.
4. Na prática quando $f(P)$ tem a singularidade tentam escolher a densidade $p(P)$ com o mesmo tipo de singularidade, mas com razão $f(P)/p(P)$ limitada. Esse método chama-se metodo de inclusão de singularidade em densidade.

Exemplo

Integral $\int_0^1 e^x dx$ podemos representar em forma seguinte $I = (3/2)\int_0^1 e^x(1+x)^{-1}p(x)dx$, onde $p(x) = (2/3)(1+x)$. Variável com densidade $p(x)$ podemos gerar como $\xi = \sqrt{1+3U} - 1$ (método inverso) e o estimador é

$$\theta_N = \frac{3}{2N}\sum_{i=1}^{N} e^{\xi_i}(1+\xi_i)^{-1}.$$

Neste exemplo a variância de $Z(\xi) = (3/2)e^{\xi}(1+\xi)^{-1}$ igual à $Var(Z) = (3/2)\int_0^1 e^{2x}(1+x)^{-1}dx - I^2 = 0.0269$

1.3 Variáveis antitéticas

([3], capítulo 8.1) A ideia é baseada na seguinte observação.

$$Var(\frac{X_1 + X_2}{2}) = \frac{1}{4}\Big(Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)\Big)$$

para diminuir a variância podemos escolher X_1, X_2 corelacionados negativamente. Como fazer isso? Supomos X_i são calculadas como $X_i = h(U_1, \ldots, U_m)$, então se a função h é monótona em cada componente, então $Y_i = h(1 - U_1, \ldots, 1 - U_m)$, são negativamente corelacionadas. Isso é um caso partigular de FKG desigualdade em Mecánica Estatística.

Construção do estimador. Geramos N variàveis $X_i = h(U_1, \ldots, U_m)$ e geramos paralelamente n variáveis $Y_i = h(1 - U_1, \ldots, 1 - U_m)$. O estimador $\theta_N^{(a)}$

$$\theta_N^{(a)} = \frac{1}{N}\sum_{1=1}^{N}(\frac{X_i + Y_i}{2})$$

é o estimador sugerido. Notamos que a esperança dele é $E(\theta_N^{(a)}) = a = E(X_1)$ e comparamos as variâncias com o estimador do método Monte Carlo comum: seja $Var(X_i) = \sigma^2$ e supomos também que N é um número par

$$Var(\theta_N) = \frac{\sigma^2}{N}, \quad Var(\theta_{N/2}^{(2)}) = \frac{\sigma^2}{N} + \frac{cov(X_1, Y_1)}{N} < Var(\theta_N)$$

Exemplo

([3], capítulo 8.1, exemplo 8d) Estimar o integral $I = \int_0^1 e^x dx$. Como o método comum gerariamos variáveis e^{U_i}. A função e^x é monotona, por isso

$$cov(e^U, e^{1-U}) = E(e^U e^{1-U}) - E(e^U)E(e^{1-U}) = e - (e-1)^2 = -0.2342$$

a varância

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

e

$$Var(\frac{e^{U_1}+e^{U_2}}{2}) = \frac{Var(e^U)}{2} = 0.1210 \qquad Var(\frac{e^U+e^{1-U}}{2}) = \frac{Var(e^U)}{2} + \frac{cov(e^U, e^{1-U})}{2} = 0.0039$$

1.4 Variável controle

([3], capírulo 8.2.) Queremos estimar $I = E(X)$. Supomos que sabemos a esperança de $\mu_Y = E(Y)$ de outra variàvel. Então para qualquer constante c, variável $Z = X + c(Y - \mu_Y)$ tambem é estimador de I, com a variância

$$Var(X + c(Y-\mu_Y)) = Var(X + cY) = Var(X) + c^2 Var(Y) + 2c \times cov(X,Y)$$

minimizando esse variância como uma função de c obtemos

$$c^* = -cov(X,Y)/Var(Y) \qquad \Longrightarrow Var(X + c^*(Y-\mu_Y)) = Var(X) - \frac{cov^2(X,Y)}{Var(Y)} \tag{9}$$

variável Y chama-se *variàvel controle* para estimação de X. Dividindo em (9) obtemos que a redução de variância é

$$\frac{Var(X + c^*(Y-\mu_Y))}{Var(X)} = 1 - \rho^2(X,Y)$$

Exemplo

Queremos estimar $I = \int_0^1 e^x dx = E(e^U)$. Uma escolha natural de variàvel controle é U.

$$cov(e^U, U) = E(Ue^U) - E(U)E(e^U) = \int_0^1 xe^x dx - \frac{e-1}{2} = 1 - \frac{e-1}{2} = 0.14086$$

lembrando que $Var(U) = 1/12$ obtemos

$$Var(Z^*) = Var(e^U + c^*(U - 1/2)) = Var(e^U) - 12(0.14086)^2 = 0.2420 - 0.2380 = 0.0039$$

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

Calculo analítico parcial

Exemplo. Integral parcial pela área

Calcular o volume V de figura limitada pelo superfície em coordinadas polares $r = a + hu(\phi,\theta)$, onde $-1 \le u(\phi,\theta) \le 1$, Denotamos V_0, V_1 volumes de espheras com radius $a-h$ e $a+h$ respectivamente. Usaremos o método geomêtrico. Sejam $Q_1, \ldots, Q_N$ são pontos uniformemente distribuidos em volume V_1, e seja ν número de pontos que cairam em volume V. Estimação de V é $\tilde{\theta}_N = V_1(\nu/N)$.

Extraimos agora o volume V_0. Para isso é suficiente escolher os pontos $Q'_1, \ldots, Q'_N$ que uniformemente distribuidos em camada $V_1 \setminus V_0$, Seja ν' número de pontos Q'_i que pertensem à volume V, então o estimador de V neste caso é $\tilde{\theta}'_N = V_0 + (V_1 - V_0)(\nu'/N)$. Comparamos as variâncias de $\tilde{\theta}_N$ e $\tilde{\theta}'_N$. ν e ν' são variáveis binomiais com parâmetros $p = V/V_1$ e $p' = (V - V_0)/(V_1 - V_0)$, então $Var(\nu) = Np(1-p), Var(\nu') = Np'(1-p')$. Temos

$$Var(\tilde{\theta}_N) = \frac{1}{N}V(V_1 - V) > Var(\tilde{\theta}'_N) = \frac{1}{N}(V - V_0)(V_1 - V)$$

Quando a razão $\varepsilon = h/a \ll 1$, então em estimador $\tilde{\theta}'_N$ nos na verdade extraimos a parte principal do problema, e diminuição da variància de $Var(\tilde{\theta}'_N)$ em comparação de $Var(\tilde{\theta}_N)$ tem que ser muito visível. Realmente, como

$V_0 = (4/3)\pi(a-h)^3, V_1 = (4/3)\pi(a+h)^3$, então existe u_0 tal que $|u_0| < 1$ e $V = (4/3)\pi(a+u_0h)^3$. E temos que o valor

$$V(V_1 - V) = [(4/3)\pi a^3]^2 3(1-u_0)\varepsilon + O(\varepsilon^2)$$

é proporcional à ε, e o valor

$$(V - V_0)(V_1 - V) = [(4/3)\pi a^3]^2 9(1-u_0^2)\varepsilon^2 + O(\varepsilon^2)$$

é da segunda ordem.

Agora mostraremos que modelando Q e Q' em coodrenadas polares, então os dois algoritmos são iguais computacionalmente. Realmente, in dois casos precisaremos 3 números aleatorios $U_i^{(1)}, U_i^{(2)}, U_i^{(3)}$ para gerar os pontos. Coordenadas Q_i e Q_i' podemos calcular pelas fórmulas

$$r_i = (a+h)\sqrt[3]{U_i^{(1)}}, \quad \phi_i = 2\pi U_i^{(2)}, \quad \cos\theta_i = 2U_i^{(3)}$$

ou

$$r_i = (a+h)\sqrt[3]{U_i^{(1)} + b(1-U_i^{(1)})}, \quad \phi_i = 2\pi U_i^{(2)}, \quad \cos\theta_i = 2U_i^{(3)}$$

onde $b = (a-h)^3/(a+h)^3$. Condição de pretença ao volume V verifica-se do mesmo modo para cada método: $r_i < a + hu(\phi_i, \theta_i)$. Se a condição vale então adicionamos uma unidade para contador ν (e ν' respectivamente).

Exemplo. Intagral pela parte de variáveis

.....

Importance sampling

Exemplo. Integral com singularidade.

G é a bola $x^2+y^2+z^2 \leq R^2$. Na física frequentemente podemos encontrar integralos do tipo $\int_G \int_G h(P,P')\rho^{-\alpha} dPdP'$, onde ρ é a distancia entre pontos P, P'. Consideramos neste exemplo tal integral

$$\int_G \int_G h(\rho)\rho^{-\alpha} dPdP', \tag{10}$$ int1

onde $h(P, P^prime) = h(|P - P'|)$ é limitada e $h(0) \neq 0$. Esse integral converge absolutamente quando $\alpha < 3$.

A. *Método simples de Monte Carlo.* Q, Q' uniformes independentes em G. Escolhendo $Z = V_G^2 h(\rho)\rho^{-\alpha}$, então $E(Z) = I$ e a disperção $Var(Z) = V_G^2 \int_G \int_G h^2(\rho)\rho^{-2\alpha} dPdP' - I^2$. Se $1.5 \leq \alpha < 3$ a variância é infinita.

B. *Importance Sampling.* Escolhemos densidade $p(P,P')$ de pontos Q, Q' com a mesma singularidade. Como $p(P,P') = p_Q(P)P_{Q'}(P' \mid P)$, podemos assumir Q uniforme em G. Para definir a densidade condicional, transferimos a origem em ponto P e escolhemos coordenadas polares r, θ, ϕ com a origem em P. A direção de ponto P denotamos ω (é um vetor unitario). A distancia conforme essa direção até a fronteira da bola denotamos $l(\omega)$. Seja

$$P_{Q'}(P' \mid P) = (4\pi)^{-1}(3-\alpha)r^{-\alpha}l^{-(3-\alpha)}, \tag{11}$$ f1

onde $r = |P' - P|$, e $\omega = (P' - P)/r$. A fórmula (f1 11) realmente determina a densidade. Para qualquer P

$$\int_G P_{Q'}(P' \mid P)dP' = \oint (4\pi)^{-1} d\omega \int_0^{l(\omega)} (3-\alpha)r^{2-\alpha}l^{\alpha-3} dr = 1.$$

Definimos variável Z:

$$Z(Q,Q') = 4\pi V_G(3-\alpha)^{-1}h(\rho)l^{3-\alpha}(\omega), \text{ onde } \rho = |Q - Q'|, \omega = (Q' - Q)/\rho.$$

e

$$E(Z) = \int_G \int_G Z p_Q(P) p_{Q'}(P' \mid P) dPdP' = \int_G \int_G h(r)r^{-\alpha} dPdP' = I$$

e a variância

$$Var(Z) = (4\pi V_G)^2(3-\alpha)^{-2} \int_G \int_G h^{2/6-2\alpha} p dPdP' - I^2$$

Como h é limitada, e l não exceda o diametro da bola, então o integral converge para todos $\alpha < 3$.

C. *Fórmulas para cálculo numerico.* Da causa da simetria podemos escolger o ponto Q em eixo $0z$. Então para Q definimos $x_Q = y_Q = 0$ e $z_Q = r_Q = R\sqrt[3]{U}$. Em método **A** o ponto Q' também é uniforme. Da causa da simetria podemos escolher ele em plano $\phi = 0$ e definir $x_{Q'} = r_{Q'}\sin\theta_{Q'}$, $y_{Q'} = 0$, $z_{Q'} = r_{Q'}\cos\theta_{Q'}$, onde $\cos\theta_{Q'} = 2U^{(1)} - 1, r_{Q'} = R\sqrt[3]{U^{(2)}}$. Obtemos seguinte algoritmo para cálculo de i com método **A**:

1. Fórmula para i-gêsimo experimento:

$$r_Q = R\sqrt[3]{U^{(1)}},\ \cos\theta_{Q'} = 2U^{(2)} - 1,\ r_{Q'} = R\sqrt[3]{U^{(2)}},\ \rho = \sqrt{r_{Q'}^2 - 2r_Q r_{Q'}\cos\theta_{Q'} + r_Q^2}.$$

2. se $\rho = \rho_i$ – valor de ρ obtido em i-gêsimo experimento, então

$$I_N^A = V_G^2 \frac{1}{N}\sum_{i=1}^{N} h(\rho_i)\rho_i^{-\alpha}. \tag{12}$$

Para o método **B** o ponto Q' com a densidade (11) podemos construir de jeito seguinte. De ponto Q escolhemos direção aleatória ω e depois nessa direção no raio $P' = Q + r'\omega$ escolhemos distancia ρ com distribuição $F_\rho(r) = (r/l)^{3-\alpha}$, onde $0 < r < l = l(\omega)$, então $Q' = Q + \rho\omega$.

Realmente, a densidade condicional de Q' (dado $Q = P$) em coordenadas polares r, θ, ϕ com centro P é igual à

$$p_{Q'}(P' \mid P) r^2 \sin\theta = [(4\pi)^{-1}\sin\theta][(3-\alpha)r^{2-\alpha}]l^{\alpha-2}.$$

primeiro fator em produto é densidade de direção aleatória, segundo – a densidade condicional da distancia em raio $P' = P + r\omega$.

As mesmas simetrias deixam escolher ω em $\phi = 0$. Isso se determina pelo um parâmetro $\mu = \cos\theta = 2U^{(1)} - 1$, calculando $l = |L - Q|$, obtemos o algoritmo seguinte:

1. Fórmula para i-gêsimo experimento

$$r_Q = R\sqrt[3]{U^{(1)}},\ \mu = 2U^{(2)} - 1,\ l = -\mu r_Q + \sqrt{R^2 - r_Q^2(1-\mu^2)},\ \rho = l(U^{(3)})^{1/(3-\alpha)};$$

2. se $l = l_i$ e $\rho = \rho_i$, então

$$I_N^B = 4\pi V_G(3-\alpha)^{-1}N^{-1}\sum_{i=1}^{N} h(\rho_i) l_i^{3-\alpha}.$$

Exemplo 8w [3]

X_i i.i.d. $X_i \sim N(\mu, 1)$, onde $\mu < 0$. Seja $S_n = \sum_{i=1}^n X_i$. E seja $N = \min\{n :\ \text{ ou } S_n < -A,\ \text{ ou } S_n > B\}$, onde A, B são positivos. Queremos estimar $\theta = P(S_N > B)$.

Importance sampling: Geramos $Y_i \sim N(-\mu, 1)$. Denotamos

$$I = \begin{cases} 1, & \text{se } \sum_{i=1}^N Y_i > B \\ 0, & \text{caso contrário} \end{cases}$$

O estimador de θ agora é

$$Z = I\prod_{i=1}^{N}\frac{f_\mu(Y_i)}{f_{-\mu}(Y_i)} = I\prod_{i=1}^{N} e^{2\mu\sum_{i=1}^N Y_i} < e^{2\mu B}$$

Notamos que obtemos um resultado teorico

$$P(\text{ atraversar } B \text{ antes de } -A) \le e^{2\mu B}$$

3 Lista de exercícios.

1. Provar Theorema 1.
2. Consideramos estimação pela parte de área: estimador (5). Vamos simular Q' usando o método de rejeição-aceitação. Seja t_0 o tempo gasto em verificação de condição $Q \in G \setminus B$, Seja t_Q o tempo gasto em calculo de um valor Q e seja t_f o tempo gasto em calculo de $f(Q)$. Achar o tempo médio $t_{Q'}$ em calculo de um ponto Q'. Provar que se $t_0 \le ct_f$, então o custo computacional de método (5) não é maior de que o custo computacional de estimador de método Monte Carlo simples.

   Note: A condição $t_0 \le ct_f$ é facil de verificar "na prática": a condição é vale quando a área B é "simples"e a função f é "complexa".
3. Provar Theorema 2
4. Exercício 3.1, página 66, [4]. Esse exercício é resolvido em [5], adiciona em item (b) o gragico de convergência como na Fig.3.3.
5. Exercício 3.4, pagina 70, [4]. (veja tambem [4] para exercícios parecidos)

Referências

Kahn [1] Kahn H. *Random sampling (Monte Carlo) techniques in neutron attenuation problems.* Nucleonics, **6**, N5, 27-33, 1950.

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

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

CG [4] C.P.Robert and G.Casella *Introducing Monte Carlo Methods with R.* Springer, 2010

CG1 [5] C.P.Robert and G.Casella *Introducing Monte Carlo Methods with R. Solutions to Odd-Numbered Exercises* arxive, 2010