Monte Carlo Method

Monte Carlo Simulation

Peter Frank Perroni

December 1, 2015

Histórico

- Técnica muito antiga porém somente recentemente oficializado como método estatístico.
- Foi muito importante nas simulações da bomba desenvolvida no Projeto Manhattan.
  - Stanislaw Ulam e John von Neumann, Projeto Manhattan (WW II), 1946.

Histórico

- Técnica muito antiga porém somente recentemente oficializado como método estatístico.
- Foi muito importante nas simulações da bomba desenvolvida no Projeto Manhattan.
  - Stanislaw Ulam e John von Neumann, Projeto Manhattan (WW II), 1946.
  - Desenvolvido durante uma tentativa de Stanislaw de estimar as chances de ganhar no jogo de cartas "*Paciência*".
  - Um colega (Nicholas Metropolis) sugeriu dar ao método o nome de um cassino de Mônaco chamado **Monte Carlo**, onde seu tio costumava perder todo seu dinheiro.

Definição

- Classe de algoritmo estocástico que utiliza amostras aleatórias para obter resultados numéricos.
- Utilizado quando os métodos matemáticos são ineficientes e/ou ineficazes.
- Aplicável a qualquer problema de interpretação probabilística.
  - Traduz incertezas do modelo de entrada em incertezas no modelo de saída.
  - Gera uma distribuição de probabilidade a partir de outra distribuição de probabilidade.

Utilização

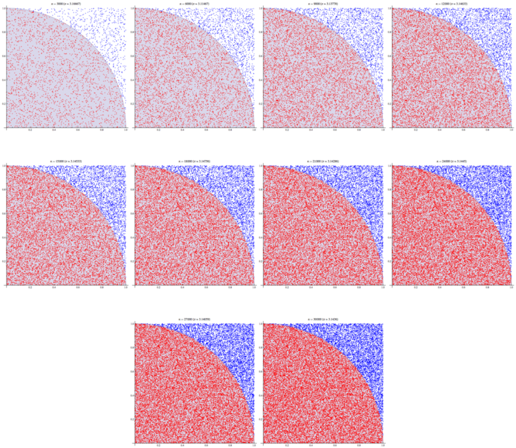
- Principais classes de problemas:
  - Visualização de Distribuições de Probabilidade com regras complexas
  - Otimização
  - Integração
- Aplicações em:
  - Análise de tráfego
  - Predição de bolsa de valores
  - Desenho de cirtuitos integrados
  - Desenvolvimento de modelos matemáticos de Astrofísica
  - Comunicação, Engenharias, Mecânica Quântica, etc.

Amostragem

- Valores amostrais calculados em pontos aleatórios da função sendo simulada são utilizados para estimar o valor real da função.
    - Quanto mais amostras, melhor a aproximação do valor real da função.

Amostragem

Cálculo de π através do Método Monte Carlo



Amostragem

- Considerando um caso hipotético em que
  - Para 1 única dimensão, 10 amostras são o suficiente para um bom resultado.
  - Então para 10 dimensões, 10^{10} amostras são necessárias.
    $\rightarrow$ Impraticável para 10 dimensões.
    $\rightarrow$ Não computável para 100 dimensões (10^{100}).

Amostragem

- Quantidade de amostras excessivamente reduzida afeta a qualidade dos resultados.
- A quantidade ideal de amostras é impraticável.
    - A Confiança é diretamente dependente do número de amostras utilizadas.
    - A seleção de bons conjuntos amostrais é crucial para resultado satisfatórios com reduzida quantidade de avaliações.

Amostragem

- Quantidade de amostras excessivamente reduzida afeta a qualidade dos resultados.
- A quantidade ideal de amostras é impraticável.
  - A Confiança é diretamente dependente do número de amostras utilizadas.
  - A seleção de bons conjuntos amostrais é crucial para resultado satisfatórios com reduzida quantidade de avaliações.

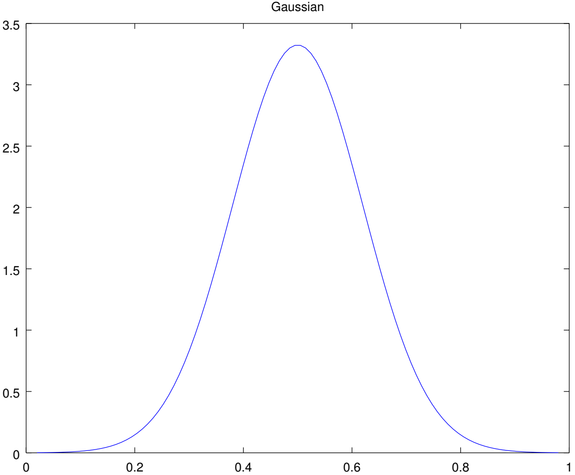
  $\rightarrow$ O gerador de números aleatórios é quem define quais amostras serão calculadas.

Distribuição de Probabilidade

O gerador de números aleatórios afeta diretamente na qualidade das amostras obtidas.

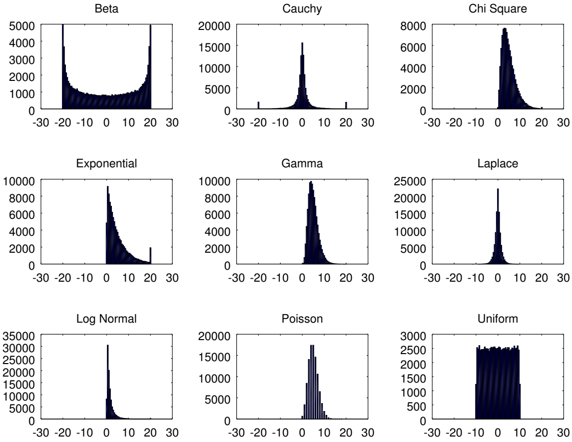
- Sua distribuição de probabilidade guia a região onde as amostras serão mais frequentemente calculadas.
- Deve ter boa aleatoriedade.

Distribuição Normal



Distribuição de Probabilidade

Outras Distribuições de Probabilidades

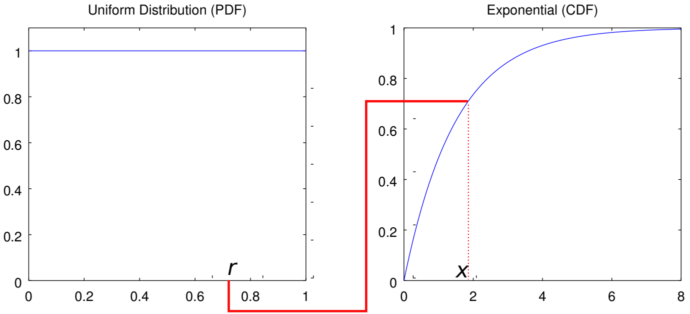


Distribuição de Probabilidade

Variáveis do problema podem ter características diferentes.

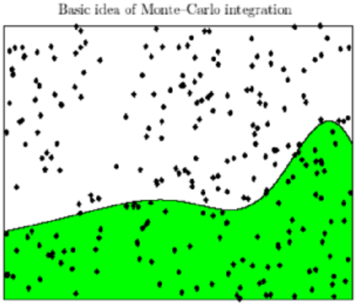
- Cada variável pode utilizar sua própria distribuição de probabilidades.

Método da Função Inversa



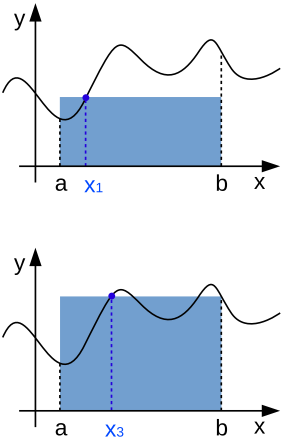
Integração pela Área Total

Aproximado pela área total multiplicado pela fração de pontos que ficaram abaixo da curva.



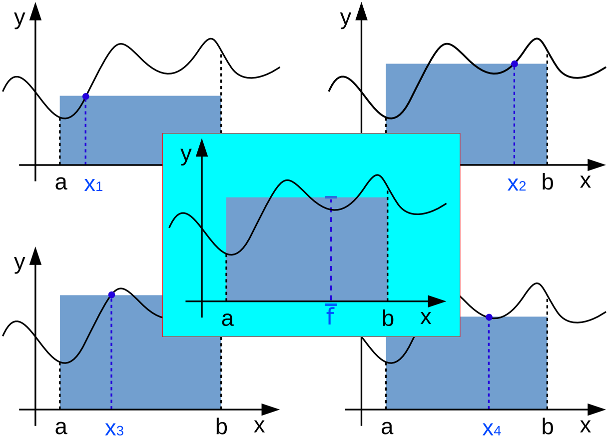
Integração pela Média

Calcula-se $f(x)$ para cada amostra aleatória.



Integração pela Média

Estima a integral pela média das amostras multiplicada pelo intervalo analisado.



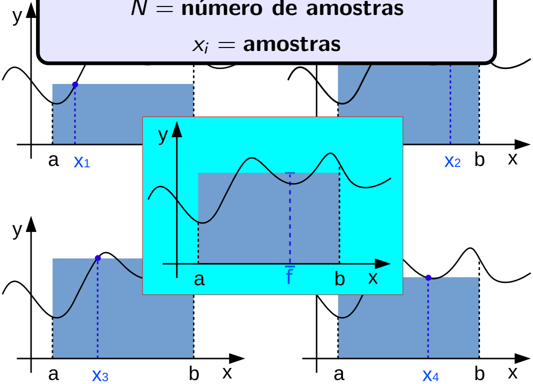
Integração pela Média

Estima a integral pela média das amostras multiplicada pelo intervalo analisado.

$$\int_a^b f(x)d(x) \approx (b-a)\frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

$N =$ **número de amostras**

$x_i =$ **amostras**



Números Aleatórios

Números Totalmente Aleatórios

- True Random Number Generator (TRNG)
- Não há como prever.
  - Experimentos tornam-se irreprodutíveis.
  - Nenhuma distribuição de probabilidades pode ser associada.
- Requer hardware especializado ou software específico.
  - /dev/random, /dev/urandom
  - rngd, entropy tools, etc.
  - i7, Babbler, GRANG, etc.

Números Pseudo-Aleatórios

Pseudo Aleatoriedade

- Números aparentemente aleatórios produzidos a partir de:
  - Uma semente.
  - Uma equação matemática.
- Possui período de ciclicidade definido pela equação utilizada.
- Segue uma distribuição probabilística.
  - Uniforme, Gaussiana, Exponencial, Cauchy, etc.

Números Pseudo-Aleatórios

Gerador de Números Pseudo-Aleatórios

- Pseudo Random Number Generator (PRNG).
- Deve possuir período suficientemente longo.
- Cálculo deve ser eficiente.
- Correlação entre os números gerados deve ser de difícil compreensão.
- Sementes diferentes devem produzir números diferentes.
  - Sementes iguais devem produzir sempre a mesma sequência.

Números Pseudo-Aleatórios

PRNG de Distribuição Uniforme

- Muito eficiente e de fácil implementação.
- Requer um módulo (m), um multiplicador (a) e um incremento (c).
  - Os valores utilizados para cada parâmetro afeta a qualidade e a periodicidade da sequência.

$$s_{n+1} = (a \times s_n + c) \mod m$$

Onde:

$$m > 0, \qquad 0 < a < m,$$
$$0 \leq c < m, \quad 0 \leq s_0 < m$$

Números Pseudo-Aleatórios

PRNG de Distribuição Uniforme

- Os parâmetros devem obedecer aos seguintes critérios para maximizar o período:
  - c e m devem ser primos entre si.
  - $(a-1)$ deve ser divisível por todos os fatores primos de m.
  - Se m for múltiplo de 4, $(a-1)$ também deve ser múltiplo de 4.

Números Pseudo-Aleatórios

PRNG de Distribuição Uniforme

- Os parâmetros devem obedecer aos seguintes critérios para maximizar o período:
  - c e m devem ser primos entre si.
  - $(a-1)$ deve ser divisível por todos os fatores primos de m.
  - Se m for múltiplo de 4, $(a-1)$ também deve ser múltiplo de 4.

Exemplos:

$$s_{n+1} = (3 \times s_n + 5) \mod 32, \text{ com } s_0 = 9$$

$$s_{n+1} = (1103515245 \times s_n + 12345) \mod 2^{31}$$

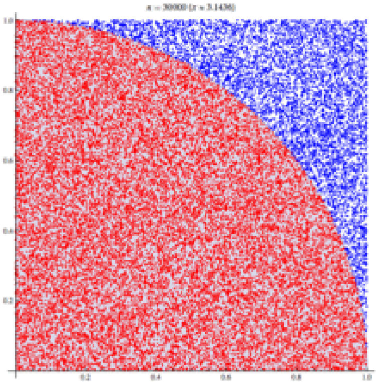


Figure : Cálculo de π através do Método de Monte Carlo.

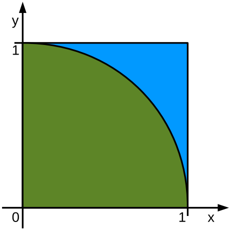


Figure : Quadrante do Círculo ($r = 1$).

Equação da Circunferência: $x^2 + y^2 = r^2 \quad \rightarrow \quad x^2 + y^2 = 1$
Área do Quadrante do Círculo: $A_c = \frac{\pi r^2}{4} = \frac{\pi}{4}$
Área do Quadrado: $A_q = r^2 = 1$
Portanto, $\frac{A_c}{A_q} = \frac{\frac{\pi}{4}}{1} = \frac{\pi}{4} \qquad \rightarrow \qquad \pi = 4\frac{A_c}{A_q}$

Exemplo

Cálculo de π

```c
// Outra forma de se ver o problema:
//    Área do círculo = pi * r^2
//    Para calcular pi, r = 1 --> pi = área do círculo.
#define RAIO_2 1

double pi(int n){
   int count = 0;
   double x, y;

   // Estima a área do Quadrante do círculo.
   for(int i=0; i < n; i++){
      // Utilizando Distribuição Uniforme.
      x = (double)rand() / RAND_MAX;
      y = (double)rand() / RAND_MAX;

      if(x * x + y * y <= RAIO_2) // Equação do círculo.
         count++; // Núm. incidências dentro do círculo.
   }

   // Estimativa da área do Quadrado = n.
   // Por regra de 3, Área do Quadrante do círculo = count / n.
   // Portanto, a área do Círculo = 4 * (count / n) = pi.
   return ((double) count / n) * 4.0;
}
```

Exemplo
Cálculo de Integral

```
double calcIntegral(double (*f)(double),
            double a, double b, int n){

    double x, sum = 0;

    for(int i=0; i < n; i++) {
        x = a + ((double)rand() / RAND_MAX) * (b - a);
        sum += f(x);
    }

    return (sum / n) * (b - a);
}
```

```c
double calcIntegral(double (*f)(double),
        double a, double b, int n){

    double x, sum = 0;

    for(int i=0; i < n; i++) {
        x = a + ((double)rand() / RAND_MAX) * (b - a);
        sum += f(x);
    }

    return (sum / n) * (b - a);
}
```

$$\boxed{(b-a)\frac{1}{N}\sum_{i=1}^{N} f(x_i)}$$

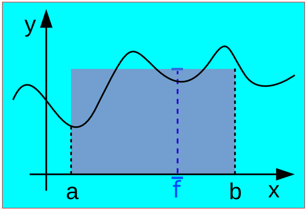
```
double calcIntegral(double (*f)(double),
          double a, double b, int n){

    double x, sum = 0;

    for(int i=0; i < n; i++) {
        x = a + ((double)rand() / RAND_MAX) * (b - a);
        sum += f(x);
    }

    return (sum / n) * (b - a);
}
```

$$\boxed{(b-a)\frac{1}{N}\sum_{i=1}^{N} f(x_i)}$$



Exemplo de Distribuição de Probabilidade corretamente aplicada a uma equação.

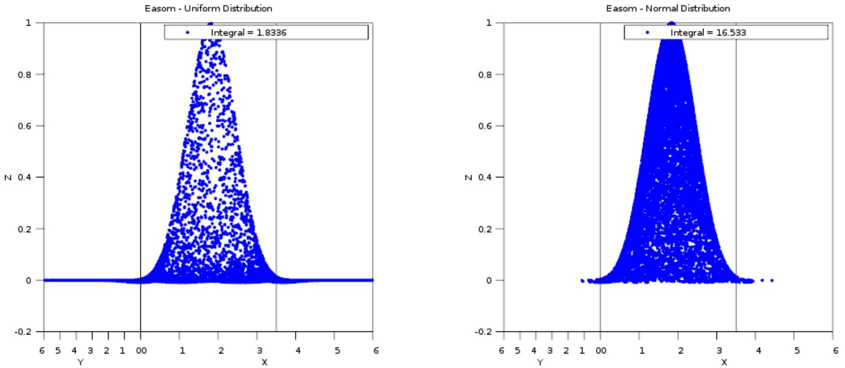
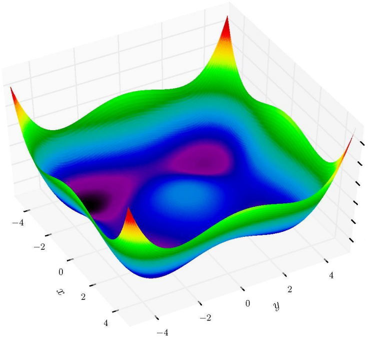


Figure : 10000 amostras no intervalo $[0, 6]$

```
$ wget http://www.inf.ufpr.br/pfperroni/montecarlo.tar.gz
$ tar -xzvf montecarlo.tar.gz
$ cd montecarlo
$ make
$ ./montecarlo pi 1000000
$ ./montecarlo integral square 1000000 -2 2
```

Função Styblinski-Tang

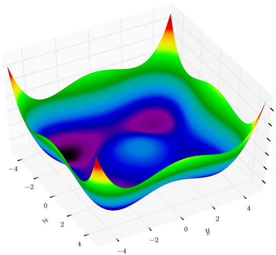
$$f(x) = \frac{\sum_{i=1}^{n} x_i^4 - 16x_i^2 + 5x_i}{2}$$
$$n \geq 1$$



Função Styblinski-Tang

- Calcule a área da função Styblinski-Tang através do método de integração de Monte Carlo pela média.
  - Calcule 10 milhões de amostras pseudo-aleatórias para duas variáveis ($n = 2$) no intervalo $[-4, 4]$.
    - n deve ser parâmetro.
  - Utilize como semente o valor 1234 (função srand()).
  - Não utilize otimizações neste código.

$$f(x) = \frac{\sum_{i=1}^{n} x_i^4 - 16x_i^2 + 5x_i}{2}$$



Função Styblinski-Tang

$$f(x) = \frac{\sum_{i=1}^{n} x_i^4 - 16x_i^2 + 5x_i}{2}$$

Qual o valor da integral?

Função Styblinski-Tang

$$f(x) = \frac{\sum_{i=1}^{n} x_i^4 - 16x_i^2 + 5x_i}{2}$$

Qual o valor da integral?

$\rightarrow$ Calcule agora para 4 dimensões.
Percentualmente, qual foi o aumento no tempo de processamento ao dobrar o número de dimensões?

Função Styblinski-Tang

$$f(x) = \frac{\sum_{i=1}^{n} x_i^4 - 16x_i^2 + 5x_i}{2}$$

Qual o valor da integral?

$\rightarrow$ Calcule agora para 4 dimensões.
Percentualmente, qual foi o aumento no tempo de processamento ao dobrar o número de dimensões?
$\rightarrow$ E ao quadruplicar as dimensões?

Distribuições Normal + Beta aplicadas à função Styblinski-Tang

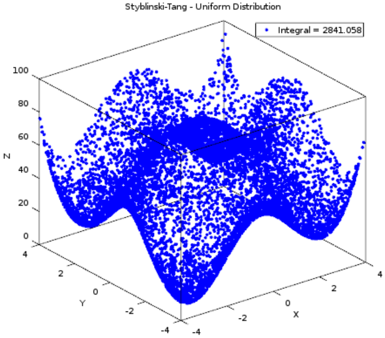


Figure : 10000 amostras no intervalo $[-4, 4]$

- Otimize a função implementada no exercício anterior utilizando as técnicas aprendidas durante as aulas.
  - Utilize Distribuição Uniforme.
  - Não é necessário deixar o código genérico.
  - Utilize as ferramentas de likwid para auxiliar na análise de desempenho.
  - Mantenha a versão anterior como base de comparação dos resultados.

Responda e Justifique

- Houve melhora ou piora na performance?
- Quais os motivos que levaram a este resultado?
- Se houve piora, pense quais os pontos que ainda podem ser ajustados no código para melhorar a performance.