

# UMA CONTRIBUIÇÃO PARA A UTILIZAÇÃO DA AMOSTRAGEM POR IMPORTÂNCIA EM CÁLCULOS MONTE CARLO

**Wilson J. Vieira**

Instituto de Estudos Avançados

Rodovia dos Tamoios, Km 5,5 12228-840 São José dos Campos, SP

[wjvieira@ieav.cta.br](mailto:wjvieira@ieav.cta.br)

## Resumo

O objetivo deste trabalho é mostrar um algoritmo computacional Monte Carlo, de caráter geral, mas capaz de ilustrar os conceitos teóricos básicos da amostragem por importância utilizando funções discretas. Esta técnica de redução da variância tem se tornado mais comum e isto se deve, em grande parte, à grande disseminação de planilhas de cálculo. Além disso, esta técnica está integrada em vários programas de cálculo Monte Carlo, disponibilizados aos profissionais da Pesquisa Operacional, Análise de Risco, Gerência de Projetos, Engenharia Nuclear e outras áreas científicas. Estes programas usam grandes vetores e matrizes de dados que podem ser facilmente trabalhados em planilhas. Esta é a motivação para o emprego de funções de importância discretas. Os resultados obtidos evidenciam a eficácia da técnica para redução da variância, que pode indicar reduções significantes no tempo computacional. Os programas de demonstração são comentados para ilustrar as aplicações.

**Palavras-Chaves:** Métodos Monte Carlo; Amostragem por Importância; Redução da Variância; Erro Circular Provável.

## Abstract

The objective of this work is to show a general Monte Carlo computational algorithm, to illustrate the basic theoretical concepts of importance sampling using discrete functions. This variance reduction technique has become more common because of the increase of the currently available powerful spreadsheet programs. Also, this technique is present in most of the Monte Carlo programs available to professionals of the Operational Research, Risk Analysis, Project Management, Nuclear Engineering, and other scientific areas. These programs use large data vectors and matrices, which can be easily handed by spreadsheets. This is the motivation to use discrete importance functions. The results show this technique efficacy in reducing the variance, which may indicate significant reduction in the computational time. The demonstrations programs are commented to illustrate the applications.

**Keywords:** Monte Carlo Methods; Importance Sampling; Variance Reduction; Circular Error Probable.

## 1. INTRODUÇÃO

A necessidade de reduzir os tempos de processamento em cálculos Monte Carlo foi discutida, em 1949, por H. Khan [1] e J. von Neumann [2], entre outros. A amostragem por importância, desde então, tem sido bastante utilizada [3, 4]. Neste caso, a utilização de funções discretas torna-se cada dia mais importante por causa da grande disseminação de programas com planilhas de cálculo, como por exemplo, o programa Excel [5] da Microsoft. Além disso, programas atuais, como por exemplo, os programas MCNP [6] e Crystal Ball [7], permitem grandes vetores (arrays) de dados que podem ou devem ser trabalhados em planilhas. Atualmente, é bastante comum os programas computacionais poderem acessar

dados diretamente destas planilhas. Portanto, existe uma grande motivação para a utilização de funções discretas e o propósito deste trabalho é mostrar um algoritmo computacional de caráter geral, mas capaz de ilustrar o conceito básico da amostragem por importância utilizando funções discretas, e ao mesmo tempo, apresentar o poder desta técnica na aplicação em um problema simplificado do cálculo do Erro Circular Provável (Circular Error Probable, CEP).

A Seção 2 mostra os fundamentos matemáticos básicos da amostragem por importância e um exemplo utilizando funções discretas. A Seção 3 mostra dois exemplos de aplicação para demonstrar a eficácia da técnica. Um exemplo, que trata da análise do Erro Circular Provável, as funções analíticas são transformadas em funções discretas para solução de um problema comum da área militar.

## 2. AMOSTRAGEM POR IMPORTÂNCIA

A amostragem por importância consiste em forçar a seleção de um maior número de amostras nas partes mais importantes do problema (eventos que mais contribuem para os parâmetros a serem inferidos). Esta distorção se faz introduzindo uma nova função distribuição e os valores devem ser corrigidos por um fator peso para não alterar os resultados esperados.

Uma maneira simplificada para explicar a amostragem por importância é através do conceito de integração via Monte Carlo. Seja a integral:

$$\Phi = \int_{-\infty}^{\infty} f(x)dx \tag{1}$$

A integração direta pode ser ineficiente por causa da forma da curva  $f(x)$ . Neste caso grandes variações nos valores de  $f(x)$  causam uma variância grande, o que significa dificuldades para a obtenção de uma variância da média de qualidade suficiente, visto que:

$$S_f^2 = \frac{1}{N} S_{f_i}^2 \tag{2}$$

O objetivo da amostragem por importância é concentrar a distribuição de amostras nas regiões da integral que são mais importantes em vez de amostrar uniformemente. Vale notar que a importância, de que se trata aqui, diz respeito ao que se quer calcular. No caso do cálculo de blindagens para radiação e do cálculo de pagamentos de seguro de vida, os eventos importantes são os mais raros. No caso de integrais queremos reduzir a variância do valor  $\Phi$ . O procedimento consiste em introduzir uma função importância  $i(x)$  da seguinte maneira:

$$\Phi = \int_{-\infty}^{\infty} \frac{f(x)}{i(x)} i(x)dx = \int_{-\infty}^{\infty} g(x) i(x)dx \tag{3}$$

A integral pode ser reescrita como:

$$\Phi = \int_{-\infty}^{\infty} g(x)dI(x); \tag{4}$$

$$\text{onde: } dI(x) = i(x)dx \text{ e } I(x) = \int_{-\infty}^x i(x')dx'$$

Portanto, o novo integrando  $g(x)$  poderá ser mais retangular que  $f(x)$  fazendo com que a variância seja menor por causa da menor variação nos valores da nova distribuição. A escolha de uma função  $i(x)$  bastante simples comparada com  $f(x)$  é suficiente para melhorar o resultado final. Porém, a função  $i(x)$  deverá ser escolhida adequadamente, pois do contrário os resultados podem piorar.

Para a solução de integrais, o procedimento para aplicação da amostragem por importância é o seguinte:

- I. Selecionar um número aleatório  $R_1$ .

$$\text{II. Fazer: } R_1 \cong I(x_1) = \int_{-\infty}^{x_1} i(x') dx' .$$

Esta equação determina o valor de  $x_1$ . É importante notar que  $x_1$  não é mais escolhido uniformemente, mas de acordo com  $i(x)$ .

$$\text{III. Calcular: } g(x_1) = \frac{f(x_1)}{i(x_1)} \equiv g_1 .$$

IV. Somar  $g_1$  no contador  $g_i$ ,  $i=1, N$ .

V. Repetir os passos acima  $N$  vezes.

Obtemos, então,

$$\bar{g} = \frac{1}{N} \sum_{i=1}^N g_i \quad (5)$$

A variância da função acima pode ser estimada por:

$$S_{g_i}^2 = \int_{-\infty}^{\infty} [g(x_i) - \bar{g}]^2 i(x) dx ; \text{ ou alternativamente por:} \quad (6)$$

$$S_{g_i}^2 \cong \frac{1}{N} \sum_{i=1}^N [g(x_i) - \bar{g}]^2$$

A variância da média é calculada por:

$$S_{\bar{g}}^2 = \frac{1}{N} S_{g_i}^2 \quad (7)$$

A variância da amostragem é diminuída, se a razão  $f(x)/i(x) = g(x)$  for, aproximadamente, constante, ou seja,  $g(x)$  aproxima-se de uma distribuição retangular.

Uma maneira mais formal de demonstrar o princípio da amostragem por importância é necessária para construção de um algoritmo de amostragem. Para isto, considere novamente a integral

$$\Phi = \int_{-\infty}^{\infty} f(x) dx \quad (8)$$

O procedimento consiste em introduzir uma função importância  $i(x)$  da seguinte maneira:

$$f(x) = \frac{f(x)}{i(x)} i(x) = \frac{\int_{-\infty}^{\infty} f(x) i(x) dx}{i(x)} \frac{f(x) i(x)}{\int_{-\infty}^{\infty} f(x) i(x) dx} \quad (9)$$

A distribuição modificada,  $g(x)$ , é dada por

$$g(x) = \frac{f(x) i(x)}{\int_{-\infty}^{\infty} f(x) i(x) dx} \quad (10)$$

ou,

$$f(x) = \frac{f(x)}{g(x)} g(x) \quad (11)$$

Onde  $f(x)/g(x)$  é o peso de correção. É necessário calcular também o fator de normalização,  $J$ , da nova função distribuição.

$$J = \int_{-\infty}^{\infty} f(x) i(x) dx \quad (12)$$

Portanto

$$g(x) = \frac{f(x)i(x)}{J} \tag{13}$$

O peso de correção pode ser escrito como

$$\frac{f(x)}{g(x)} = \frac{J}{i(x)} \tag{14}$$

O programa na Tabela 1, escrito em Scilab [6], mostra uma amostragem por importância utilizando uma função discreta. Este algoritmo pode ser utilizado, por exemplo, para amostragem da energia ou do ângulo de emissão de partículas, da análise de uma foto de satélite ou de uma cadeia de DNA. Para isto, basta o usuário colocar mais importância na região onde ele sabe que há mais eventos importantes. Este programa calcula a média do vetor  $f(x)$  que é igual a 50, utilizando a função importância  $i(x)$  em vez da distribuição uniforme.

A Fig. 1 ilustra o resultado de duas amostragens do programa na Tabela 1. Uma amostragem com a função importância uniforme e outra com dois valores de intervalos com importâncias maiores. Pode-se notar que os valores concentram-se nas partes de maior importância. No entanto, a média ( $f_{bar}$ ) continua aproximadamente igual a 50, que é a média da distribuição original.

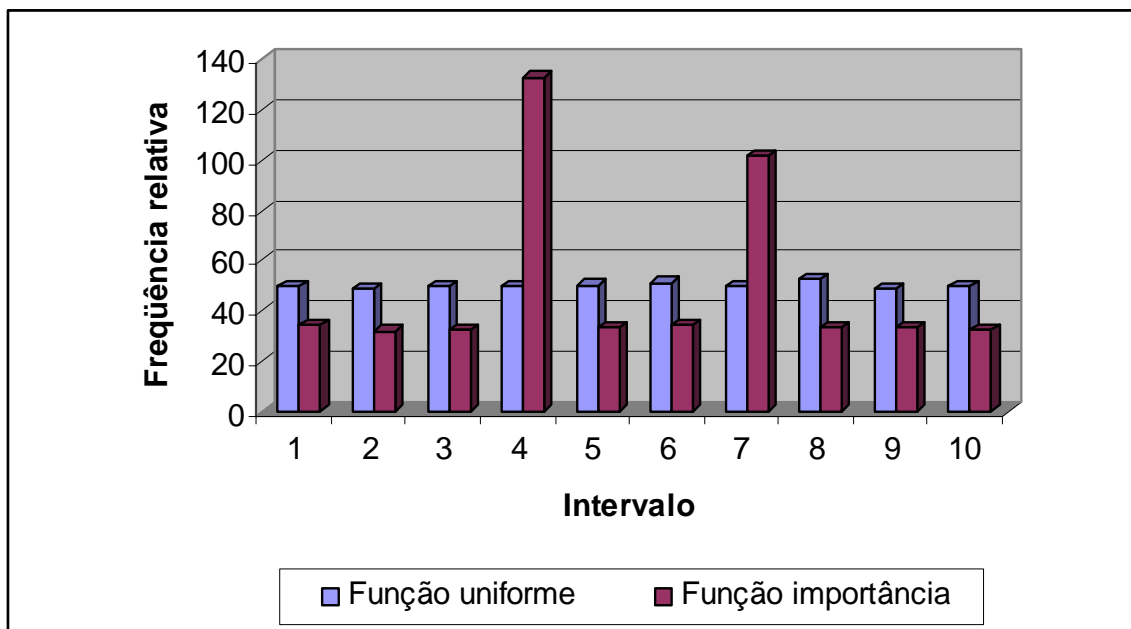


Figura 1. Gráfico de amostragem por importância de uma distribuição uniforme.

### 3. EXEMPLOS DE APLICAÇÃO

#### 3.1. CÁLCULO DO ERRO CIRCULAR PROVÁVEL (CEP)

Este problema tem relação com área militar, mais especificamente, sobre a probabilidade de sucesso de lançamento de bombas. Cada tipo de bomba, míssil ou ogiva tem associado um valor de CEP, que é definido como o raio de uma área que tem probabilidade de 50% de ser atingida. Para o caso da distribuição normal  $N(0,1)$  os resultados analíticos para o CEP e para o desvio padrão são, respectivamente, 1,18 e 1,41. Neste exemplo, o cálculo do valor do CEP é feito via integração Monte Carlo e as distribuições os eixos  $X$  e  $Y$  são independentes, sem covariância e não correlacionadas. A grande vantagem dos métodos Monte Carlo sobre outros métodos é que a introdução destas condições não acarreta em maiores dificuldades na solução do problema. A Tabela 2 mostra o programa Monte Carlo para o cálculo do CEP.

Tabela 1. Programa de amostragem por importância de uma distribuição uniforme utilizando funções discretas.

<b>//PROGRAMA</b>	<b>COMENTÁRIOS</b>
<i>rand('seed',0)</i>	<i>Garante a mesma sequência aleatória.</i>
<i>nhist = 10000;</i>	<i>Número total de histórias</i>
<i>counter = zeros(1,10);xinter =[1:1:10];</i>	<i>Contador e eixo para o gráfico</i>
<i>f = [50,50,50,50,50,50,50,50,50,50];</i>	<i>Distribuição original</i>
<i>ff = sum(f); f = f / ff;</i>	<i>Função distribuição de probabilidade original (fdp)</i>
<i>i = [1,1,1,4,1,1,3,1,1,1];</i>	<i>Distribuição da função importância</i>
<i>ii = sum(i); i = i / ii;</i>	<i>i tem que ser diferente de zero</i>
<i>J = sum(f.*i); J;</i>	<i>Fator de normalização</i>
<i>for n = 1:10</i>	
<i>fxi(1,n) = f(n) * i(n);</i>	
<i>peso(1,n) = J / i(n);</i>	<i>Criação da função peso para correção</i>
<i>end</i>	
<i>nfdp = fxi / J;</i>	<i>Nova distribuição fdp</i>
<i>for n = 1:10</i>	
<i>nfdc(1,n) = sum(nfdp(1:n));</i>	<i>Nova distribuição fdc</i>
<i>end</i>	
<i>for m = 1:nhist</i>	<i>Início do loop de amostragem</i>
<i>rn = rand();</i>	
<i>for n = 1:10</i>	
<i>if nfdc(n)&gt;rn then break end</i>	<i>Encontrando o intervalo da amostra</i>
<i>end</i>	
<i>ax(1,m) = f(n)*peso(n);</i>	<i>Corrigindo o valor amostrado</i>
<i>counter(n) = counter(n)+1;</i>	
<i>end</i>	<i>Fim do loop de amostragem</i>
<i>counter=counter/nhist*ff</i>	<i>Frequência de eventos da nova fdp</i>
<i>tcount=sum(counter)</i>	
<i>fbar = ff*sum(ax(1:nhist)/nhist);</i>	<i>Média da distribuição f</i>
<i>plot2d(xinter,counter,rect=[1,0,10,150]);fbar</i>	

Outro problema consiste em calcular a probabilidade de acerto em um alvo de raio igual a  $0,1$  localizado em  $(x=2,5$  e  $y=2,5)$ . A solução deste problema consiste em calcular o volume de um cilindro de base circular com raio  $0,1$  centrada em  $(x=2,5$  e  $y=2,5)$  e limitado pela superfície formada pelos pares  $(x,y)$  criados das distribuições normais. Este cálculo também é feito por integração Monte Carlo. O alvo está localizado a aproximadamente três CEPs (3,54) do centro, o que significa que apenas 0,2% das amostras caem além deste limite.

A Tabela 3 mostra o procedimento Monte Carlo para calcular a probabilidade de atingir o alvo em  $(x=2,5$  e  $y=2,5)$ . O programa consiste em gerar um grande número de amostras e verificar se elas estão dentro do pequeno círculo que compreende o alvo. Os dados iniciais deste programa estão na Tabela 2 para facilitar ao leitor testar os algoritmos.

É importante notar que se não for usado um número suficientemente grande de amostras, pode ocorrer de nenhuma amostra atingir o alvo.

Tabela 2. Programa Monte Carlo para cálculo do CEP.

<b>//PROGRAMA</b>	<b>COMENTÁRIOS</b>
<i>nb=100,nh=1000,</i>	<i>nb conjuntos de nh amostras</i>
<i>grand('setsd',0) // para facilitar debug</i>	<i>Sempre a mesma sequência aleatória</i>
<i>s1=0;s2=0;</i>	<i>Contadores de médias e desvios padrões</i>

//CEP Circular Error Probable	<b>Cálculo do CEP</b>
Av=0.0;Sd=1.0; // N(0,1)	Modificar o programa p/ outros valores.
for j=1:nb	Laço de conjuntos
x=grand(1,nh,'nor',Av,Sd);	nh valores de x
y=grand(1,nh,'nor',Av,Sd);	nh valores de y
for k=1:nh	Laço de amostragem
r(k)=sqrt(x(k)^2+y(k)^2);	Distâncias do ponto (0,0)
end	
rsort=gsort(r,'g','i');	Valores de r em ordem crescente
j50=nh/2;	Índice da amostra central (50%)
s1=s1+rsort(j50);	Distância da amostra central (50%)
s2=s2+rsort(j50)^2;	Para cálculo da variância
end	
cepm=s1/nb	CEP
sp2=(s2-s1^2/nh)/(nh-1);	Variância da amostragem
scepm=sqrt(sp2/nh)	Desvio padrão da média
//cepm = 1.1769448 confira	<b>Resultados obtidos após a rodada</b>
//scepm = 0.0111745	

Tabela 3. Probabilidade de acertar alvo na posição ( $x=2,5$  e  $y=2,5$ ).  
Sem amostragem por importância

//PROGRAMA	COMENTÁRIOS
//Prob. para alvo em (2,2,r=0.1)	<b>Sem amostragem por importância</b>
timer();	Contagem de tempo
na=nb*nh;xc=2.5;yc=2.5;raio=0.1;	Alvo em ( $x=2,5$ e $y=2,5$ ) e $r = 0.1$
x=grand(1,na,'nor',Av,Sd);	nb*nh amostras (x,y)
y=grand(1,na,'nor',Av,Sd);	
for j=1:na	
cir=(x(j)-xc)^2+(y(j)-yc)^2;	
if cir < raio then	Verifica se acertou o alvo
s3=s3+1;s4=s4+1;end	Acertou
end	
prob_area=s3/na	
sp2=(s4-s3^2/na)/(na-1)/na	
sprob_area=sqrt(sp2)	Probabilidade de acerto
sp2B=s3/na-(s3/na)^2 ;//Dist. de Bernoulli	
sprob_areaB=sqrt(sp2B/na)	Desvio padrão da média
tt=timer(),fomI=1/(sp2*tt)	Cálculo da eficiência computacional

A Figura 2 mostra o problema resolvido sem amostragem por importância. O círculo central representa o CEP calculado pelo programa mostrado na Tabela 2. O alvo está representado no canto superior direito da figura.

A região  $x > 2$  e  $y > 2$  contém o alvo e, portanto, a estatística desta região poderá ser aumentada utilizando a amostragem por importância. Para isto esta região é escolhida para compor uma função importância onde a região  $x < 2$  e  $y < 2$  tenha uma importância menor. Esta função é calculada a partir dos valores de área da distribuição normal para estes valores de  $x$  e de  $y$ . O algoritmo está descrito na Tabela 4. Note que as tabelas 2, 3 e 4 podem ser concatenadas em um só arquivo para rodar os programas. A Figura 3 revela que a amostragem por importância aumentou significativamente a estatística na região do alvo.

Os programas descritos nas tabelas 2, 3 e 4 facilitam o entendimento da aplicação da técnica e evitam que o leitor possa cometer erros na utilização dos conceitos. Embora a

simetria do problema deste exemplo facilite consideravelmente a amostragem, é importante notar que, para cada problema o algoritmo da amostragem deve ser construído e verificado.

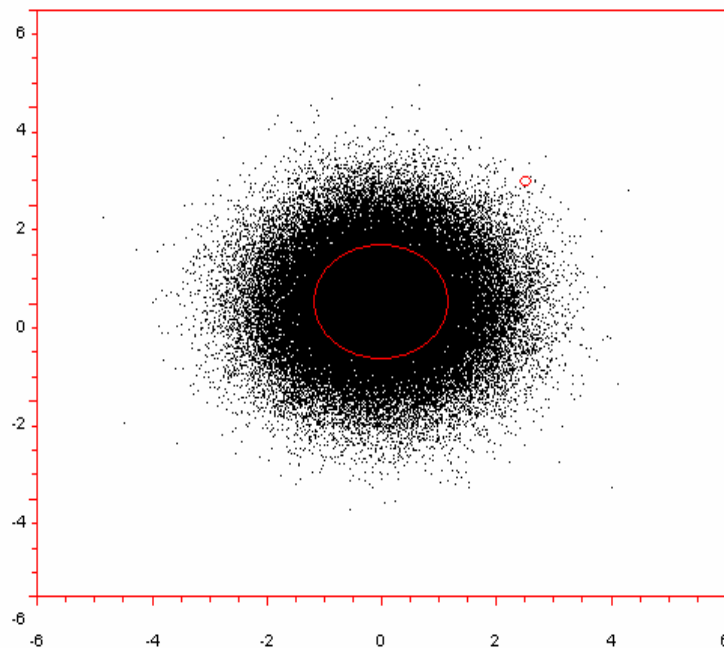


Figura 2. Amostras sem amostragem por importância.

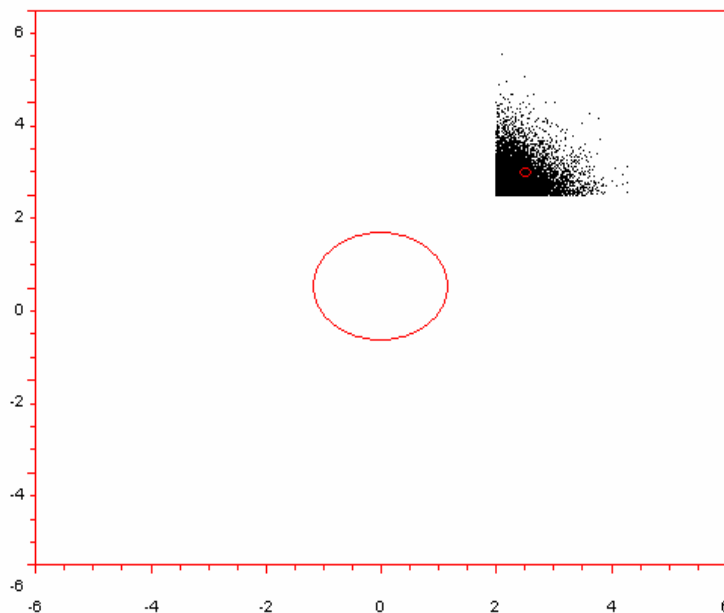


Figura 3. Amostras com amostragem por importância.

Tabela 4. Probabilidade de acertar alvo na posição  $(x=2,5$  e  $y=2,5)$ .  
Com amostragem por importância

<b>//PROGRAMA</b>	<b>COMENTÁRIOS</b>
<i>//Amostragem por importância em</i> <i>//x&gt;xylim e y&gt;xylim</i>	<b>Com amostragem por</b> <b>importância</b>
<i>xylim=2.0</i>	<i>Áreas da normal para se</i>
<i>[P,Q]=cdfnor("PQ",xylim,Av,Sd);</i>	<i>fazer a função de amostragem</i>

$f=[P,Q]$	<i>Função distribuição</i>
// $P=1/2*(1+erf(xylim/sqrt(2)))$ //prova	<i>Prova da cdfnor</i>
// $xylim=sqrt(2)*erfinv(2*P-1)$ //prova	<i>Prova da cdfnor</i>
$ff = sum(f); f = f / ff;$	<i>fdp</i>
$i = [1,100];$	<i>Função importância</i>
$ii = sum(i); i = i / ii;$	
$J = sum(f.*i);$	<i>Construção da fdp</i>
for $n=1:2$	<i>A região <math>x &gt; 2</math> e <math>y &gt; 2</math></i>
$fxi(1,n) = f(n)*i(n);$	<i>é considerada mais</i>
$peso(1,n) = J / i(n);$	<i>importante</i>
end	
$nfdp = fxi/J;$	
for $n = 1:2$	
$nfdc(1,n) = sum(nfdp(1:n));$	<i>Nova fdc</i>
end	
for $m = 1:nh$	
for $n = 1:2$	
if $nfdc(n) > rand()$ then break end	
end	
if $n==2$ then	<i>Amostra na região de interesse</i>
$Pr=P+(1-P)*grand(1,1,'def');$ $Qr=1-Pr;$	<i>(uniforme entre P e 1)</i>
$xi(m)=cdfnor("X",Av,Sd,Pr,Qr);$	
// $xi(m)=sqrt(2)*erfinv(2*Pr-1)$ //prova	
$Pr=P+(1-P)*grand(1,1,'def');$ $Qr=1-Pr;$	
$yi(m)=cdfnor("X",Av,Sd,Pr,Qr);$	
// $yi(m)=sqrt(2)*erfinv(2*Pr-1)$ //prova	
$cir=(xi(m)-xc)^2+(yi(m)-yc)^2;$	<i>Rejeição para o alvo.</i>
if $cir < raio$ then	
$s5 = s5 + f(n)*peso(n);$	
$s6 = s6 + (f(n)*peso(n))^2;$	
end	
end	
end	
$prob\_area\_imp=s5/nh$	<i>Probabilidade de acerto</i>
$sp2=(s6-s5^2/nh)/(nh-1)/nh$	
$sprob\_area\_imp=sqrt(sp2)$	<i>Desvio padrão da média</i>
$tt=timer(),fom2=1/(sp2*tt)$	<i>Figura de mérito</i>

A Figura 4 mostra o caso em que amostras na região menos importante devem ser consideradas para a solução do problema. Significa, por exemplo, que qualquer explosão tem alguma letalidade, mas o dano é pequeno para distâncias maiores do que 0.1 unidades do alvo.



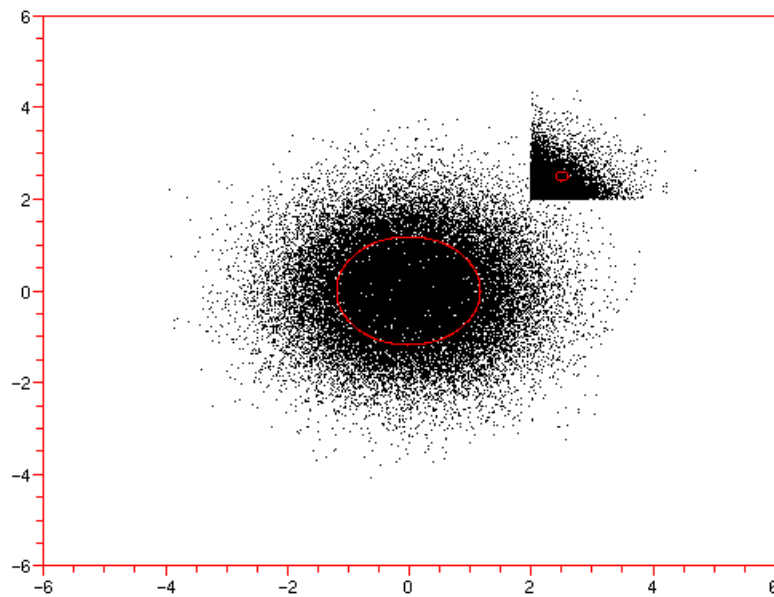


Figura 4. Amostras com amostragem por importância.  
Região menos importante também amostrada.

Os resultados obtidos são mostrados na Tabela 5. Foram utilizados valores de importância relativa de 10, 100 e 1000 vezes mais importantes para a região do alvo. A última linha da Tabela 5 é o resultado da amostragem de uma distribuição truncada na região  $2 < x < 3$  e  $2 < y < 3$  (Figura 5). Neste caso, a figura de mérito diminui por causa do aumento do esforço computacional e da pouca contribuição da cauda das normais. A Figura de Mérito (*FOM*) é definida sabendo que a variância da média é proporcional a  $1/N$  e o tempo computacional  $T$  é diretamente proporcional a  $N$ , onde  $N$  é o número de amostras.

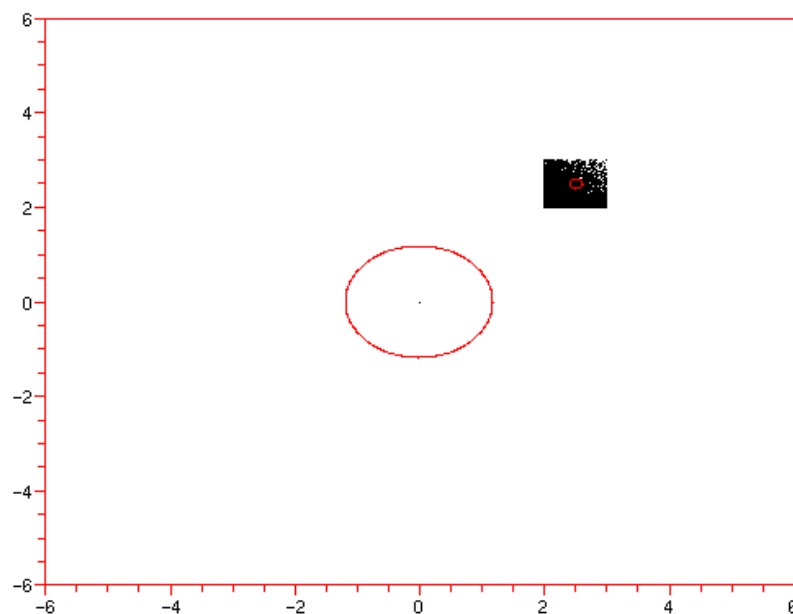


Figura 5. Amostras distribuição truncada:  $2 < x < 3$  e  $2 < y < 3$ .

Portanto,

$$FOM = \frac{1}{S_g^2 T} \quad (15)$$

Este parâmetro deve ser aproximadamente constante. Por esta razão, é também um indicador da acuidade dos valores das médias e seus desvios padrões além de indicar a eficiência computacional para comparação de algoritmos de redução da variância.

São mostrados os resultados para a média e desvio padrão. No entanto, as figuras de mérito para o caso de utilização de amostragem por importância são até 1396,8 vezes maiores. Para  $xylim=2,4$  ocorreu uma distorção na amostragem. Esta condição é um aviso para o usuário não forçar uma maior redução de variância. É importante notar que a utilização de amostragem por importância acarreta em um esforço computacional maior, que deve ser compensado pela redução do erro estatístico.

Tabela 5. Probabilidade de acerto ao alvo em ( $x=2,5$  e  $y=2,5$ ).

<i>xylim</i>	<i>Função Importância</i>	<i>Probabilidade ± Desvio Padrão</i>	<i>Tempo (s)</i>	<i>FOM</i>	<i>FOM normalizado</i>	<i>Número de amostras</i>
SAI*	Não tem.	0,00011 ± 0,0000332	1.65625	5.489D+08	1	10 <sup>5</sup>
SAI	Não tem.	0.000108 ± 0,0000104	16.109375	5.748D+08	1,047	10 <sup>6</sup>
-1	(1,10)**	Nenhuma amostra aceita	-	-	-	10 <sup>4</sup>
0	(1,10)	0.00011 ± 0.0000550	1.125	2.939D+08	0,5354	10 <sup>4</sup>
1	(1,10)	0.0001271 ± 0.0000221	0,90625	2.261D+09	4,119	10 <sup>4</sup>
2	(1,10)	0.0001094 ± 0,0000054	0.484375	7.173D+10	130,7	10 <sup>4</sup>
2	(1,100)	0.0001136 ± 0.0000027	0.96875	1.450D+11	264,16	10 <sup>4</sup>
2	(1,1000)	0.0001123 ± 0.0000022	1.1875	1.754D+11	319,55	10 <sup>4</sup>
2,2	(1,10)	0.0001054 ± 0,0000039	0.421875	1.541D+11	280,74	10 <sup>4</sup>
2,2	(1,100)	0.0001068 ± 0,0000015	0.890625	4.701D+11	856,44	10 <sup>4</sup>
2,2	(1,1000)	0.0001072 ± 0,0000010	1.21875	7.667D+11	1396,8	10 <sup>4</sup>
2,4	(1,10)	0.0000299 ± 0.0000016	0.375	1.048D+12	1909,00	10 <sup>4</sup>
>2 e <3	(1,10)	0,0001166 ± 0,0000053	1,375	2,560D+10	46,64	10 <sup>4</sup>

\* Sem amostragem por importância.

\*\* Índice 2 para a região do alvo.

### 3.2. EXEMPLO DE DISCRETIZAÇÃO GENERALIZADA

Neste exemplo, são formadas 100 regiões no quadrado de área unitária como pode ser visto na Figura 5. As regiões 23 e 75 têm peso 10 vezes maior que os pesos das áreas adjacentes que, por sua vez, têm peso 10 vezes maior que o restante das outras áreas. Este pode ser o caso da análise de imagens onde existe informação de que estas regiões são mais importantes para o parâmetro sendo estimado. Desta forma, a amostragem por importância

permite uma análise detalhada das regiões de interesse sem aumentar o tempo de computação. O resultado da amostragem pode ser visto nas figuras 6 e 7. A Tabela 6 mostra o algoritmo desta amostragem.

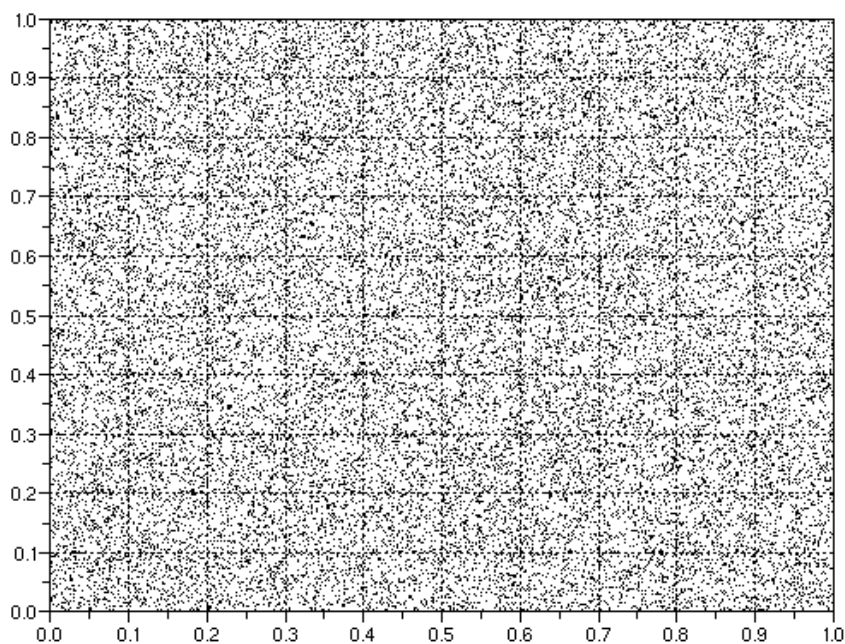


Figura 6. Amostras no quadrado unitário sem amostragem por importância.

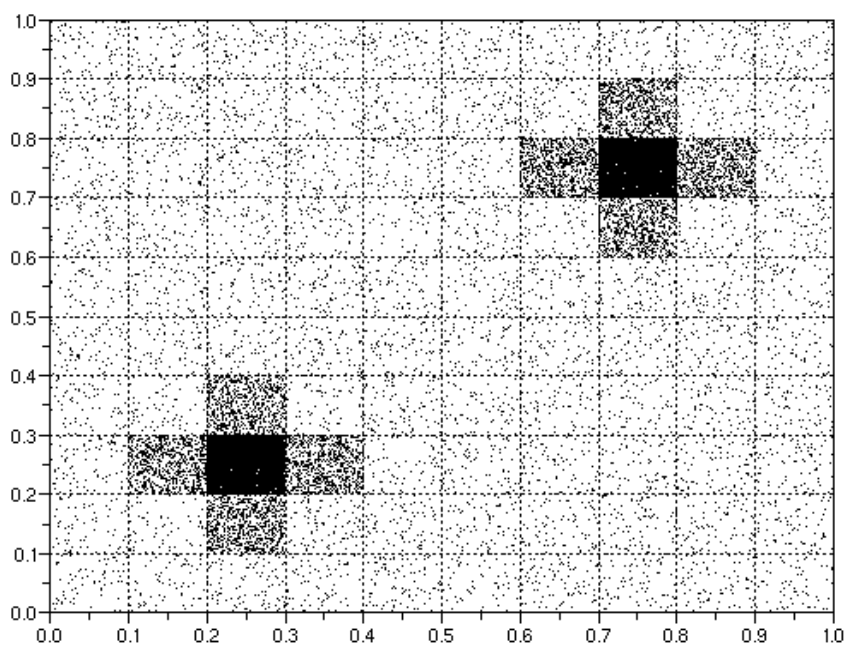


Figura 7. Amostras no quadrado unitário com amostragem por importância.

Tabela 6. Amostragem por importância. 100 regiões.

<i>for i=11:20, yinf(i)=0.1;end, for i=2:10:92, xinf(i)=0.1; end, for i=21:30, yinf(i)=0.2;end</i>
<i>for i=3:10:93, xinf(i)=0.2;end, for i=31:40, yinf(i)=0.3;end, for i=4:10:94, xinf(i)=0.3;end</i>

```

for i=41:50, yinf(i)=0.4;end, for i=5:10:95, xinf(i)=0.4;end, for i=51:60, yinf(i)=0.5;end
for i=6:10:96, xinf(i)=0.5; end, for i=61:70, yinf(i)=0.6;end, for i=7:10:97, xinf(i)=0.6;end
for i=71:80, yinf(i)=0.7;end, for i=8:10:98, xinf(i)=0.7;end, for i=81:90, yinf(i)=0.8;end
for i=9:10:99, xinf(i)=0.8;end, for i=91:100, yinf(i)=0.9;end, for i=10:10:100, xinf(i)=0.9;end
z=[xinf,yinf];nh=30000,s1=0;s2=0;nt=100;tt=timer(), x=grand(1,nh,'def'); y=grand(1,nh,'def');
set("current_figure",1),xgrid(1);xset("line style",0), plot2d(x,y,-0,'061','',[0,0,1,1]);tt=timer()
f=ones(1,nt); i=f;
i(78)=100;i(77)=10;i(79)=10;i(88)=10;i(68)=10; i(23)=100;i(22)=10;i(24)=10;i(33)=10;i(13)=10;
ff = sum(f); f = f/ff;ii = sum(i); i = i/ii; J = sum(f.*i); J;
for n = 1:nt, fxi(1,n) = f(n) * i(n); peso(1,n) = J / i(n); end
nfdp = fxi / J; for n = 1:nt, nfdc(1,n) = sum(nfdp(1:n)); end, for m = 1:nh, rn=grand(1,1,'def');
for n = 1:nt,if nfdc(n) > rn then break end, end
xi(m)=xinf(n)+0.1*grand(1,1,'def'); yi(m)=yinf(n)+0.1*grand(1,1,'def');end
set("current_figure",2), rect=[0,0,1,1];frameflag=1;xgrid(1);xset("line style",0)
plot2d(xi,yi,-0,'061','',[0,0,1,1]);tt=timer()

```

#### 4. CONCLUSÕES

Este trabalho mostra a teoria e aplicação da amostragem por importância utilizando funções discretas para redução da variância em cálculos Monte Carlo. A facilidade de aplicação desta técnica e sua disponibilidade em vários programas de simulação Monte Carlo justificam os argumentos teóricos e os detalhes computacionais aqui apresentados.

Os resultados obtidos mostraram que o emprego desta técnica pode reduzir consideravelmente a variância da amostragem e, conseqüentemente, o tempo computacional. No entanto, esta condição é dependente da experiência e conhecimento do usuário sobre o problema a ser solucionado. Neste ponto vale lembrar que a amostragem por importância é utilizada em problemas em que são os eventos raros [9] que definem a sua solução. Por exemplo, análise de risco, apólice de seguro, transporte de radiação, etc. Portanto, é importante garantir que não haja subamostragem e nem distorções causadas por funções importância muito pesadas.

A evolução dos computadores permitiu a utilização extensiva de Métodos Monte Carlo e, ao mesmo tempo, dispensou a utilização de técnicas de redução da variância. Vários problemas ficaram ao alcance de soluções sem necessidade de técnicas sofisticadas. No entanto, a cada dia surgem problemas mais complexos exigindo soluções mais detalhadas e mais realísticas, como por exemplo: análise de imagens, análise de cadeias de DNA, inferências sobre o aquecimento global, análise do fluxo de informações, estudos sobre custo de vida e níveis de consumo, simulações de mercado global, interpretação de sinais de sensores, etc. Um problema otimizado com a aplicação de amostragem por importância, com *FOM* relativa igual a 1400 e que leva um dia de tempo computacional, equivale a quase quatro anos para obtenção da solução do problema original com *FOM* relativa igual a 1. Por esta razão os problemas mencionados acima podem exigir supercomputadores e/ou a paralelização de um grande número de CPUs.

Outro aspecto importante da amostragem por importância é que embora elas tenham sido aplicadas extensamente na literatura e estejam disponíveis em vários programas Monte Carlo, a implantação dos algoritmos de amostragem pode não ser trivial, ou levar o usuário a soluções errôneas. Para a obtenção de grande economia computacional é necessário que o usuário tenha domínio da técnica e não utilize estes programas apenas como “caixas-pretas”.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] H. Khan, T. E. Harris, “Estimation of Particle Transmission by Random Sampling”. In: *Monte Carlo Method*, A. Householder, Editor, *Applied Mathematics Series* **12**, 1951.

- [2] J. von Neumann, “Various Techniques Used in Connection with Random Digits”. In: *Monte Carlo Method*, A. Householder, Editor, *Applied Mathematics Series* **12**, 1951.
- [3] L. L. Carter, E. D. Cashwell, “Particle Simulation with the Monte Carlo Method”, ERDA Critical Review Series TID-26607, 1975.
- [4] M. Kalos, P. A. Whitlock, “Monte Carlo Methods Volume I: Basics”, John Wiley & Sons, New York, 1986.
- [5] Microsoft Office Excel 2003. Copyright Microsoft Corporation.
- [6] X-5 Monte Carlo Team, “MCNP — A General Monte Carlo N-Particle Transport Code, Versão 5”, Los Alamos National Laboratory, April 24, 2003. (LA-CP-03-0245)
- [7] Crystal Ball <http://www.crystalball.com/> acesso em 12/06/2006
- [8] Scilab (c)INRIA-ENPC, [www.scilab.org](http://www.scilab.org) , acesso em 12/06/2006.
- [9] W. J. Vieira, P. N. Stevens, (1995). “Analysis of the sampling distribution in Monte Carlo radiation transport calculations”. *Ann. Nucl. Energy* 22:51–55.