Parte XI

Geração de variáveis aleatórias

- Como gerar números aleatórios com uma determinada distribuição, utilizando números aleatórios uniformes entre 0 e 1 ($U(0,1)$)?

- Assumimos que o gerador de números $U(0,1)$ é bom (testes estatísticos).

- Fatores que devem ser considerados:

Exatidão Alguém pode abrir mão da exatidão para ganhar eficiência. Ex.: somando 12 números $U(0,1)$ e subtraindo de 6 temos aproximadamente uma normal com média 0 e variância 1 (teorema central do limite).
Hoje dispomos de métodos eficientes e exatos para as principais distribuições.

Eficiência dentre as abordagens exatas para uma dada distribuições, escolher a mais eficiente em execução e/ou memória.

- Algumas abordagens tem tempo de inicialização longo.
- A estratégia deveria ser eficiente para todo parâmetro de entrada (robusto).
- Para apenas uma utilização do simulador, podemos abrir mão de eficiência para utilizar estratégias mais simples de implementar.

41 Métodos gerais

41.1 Transformação Inversa

- Gerar variável aleatória X contínua, onde a dist acumulada $F(x)$ é contínua e estritamente crescente.
  - Ou seja, se $x_1 < x_2$ então $0 < F(x_1) < F(x_2) < 1$.
  - Portanto, F admite inversa (denotada F^{-1}).

- Método:
  1. Gere $U \sim U(0,1)$.
  2. Retorne $X = F^{-1}(U)$.

- Prova: mostrar que para todo x, $\Pr[X \leq x] = F(x)$.

$$\begin{aligned}
\Pr[X \leq x] &= \Pr[F^{-1}(U) \leq x] \\
&= \Pr[F(F^{-1}(U)) \leq F(x)] \\
&= \Pr[U \leq F(x)] = F(x).
\end{aligned}$$

- Ex.: exponencial com parâmetro β.

$$F(x) = 1 - e^{-x/\beta}, \text{ para } x \geq 0.$$

$$U = 1 - e^{-F^{-1}(U)/\beta} \Rightarrow F^{-1}(U) = -\beta \ln(1-U).$$

  - Podemos usar $F^{-1}(U) = -\beta \ln(U)$.

- Quando X é discreta, pode assumir apenas os valores $x_1 < x_2 < \ldots$.
  1. Gere $U \sim U(0,1)$.
  2. Encontre o menor i tal que $U \leq F(x_i)$, e retorne $X = x_i$. Fig. L8.4.

- Algumas dist não têm forma fechada para a inversa. Ex.: normal e gama.
  - Neste caso, utilizar método numérico.

- Facilita a geração de dist truncadas entre a e b.

$$F^*(x) = \begin{cases} 0 & x < a \\ \frac{F(x)-F(a)}{F(b)-F(a)} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

  1. Gere $U \sim U(0,1)$.
  2. Seja $V = F(a) + (F(b) - F(a))U$.
  3. Retorne $X = F^{-1}(V)$.

```
u=rand(1,100); hist(u);
b=1; x=-b*log(u); hist(x); figure; ecdf(x);
   hold on; s=sort(x); plot(s,expcdf(s,b))
[h p] = chi2gof(x,'cdf',@(x)expcdf(x,b),'alpha',.05)
[h p] = kstest(x',[x' expcdf(x',b)],.05)
```

41.2 Composição

- Quando a dist F é combinação linear convexa das dist $F_1, F_2, \ldots$.

$$F(x) = \sum_{i=1}^{\infty} p_i F_i, \qquad \sum_{i=1}^{\infty} p_i = 1, \qquad p_i \geq 0 \ \ \forall i.$$

  - Ex.: dist com vários picos.

- Método:
  1. Gere uma va inteira positiva I tal que $\Pr[I = i] = p_i$ para todo i.
  2. Retorne X com dist F_i.

41.3 Convolução

- Quando a va pode ser escrita como a soma de n va independentes e identicamente distribuídas.

$$X = Y_1 + Y_2 + \cdots + Y_n.$$

  - Não é um caso particular de composição: não é combinação linear convexa.

- Ex.: binomial é a soma de bernoulli's.

- Método: seja F a dist de X, e G a dist de Y_i.

  1. Gere $Y_1, Y_2, \ldots, Y_n$, como observações independentes de G.
  2. Retorne $X = Y_1 + Y_2 + \cdots + Y_n$.

- Eficiente apenas quando n é pequeno. Caso contrário, utilizar outro método.

41.4 Aceita-rejeita

- Utilizada quando os outros métodos falham ou são ineficientes.

- Utiliza uma função $t(x)$ tal que $t(x) \geq f(x)$ para todo x.

- $t(x)$ não é uma função de densidade, pois

$$c = \int_{-\infty}^{\infty} t(x)dx \geq \int_{-\infty}^{\infty} f(x)dx = 1.$$

  - Entretanto, $r(x) = t(x)/c$ é função de densidade de probabilidade.
  - $t(x)$ deve ser escolhida de tal forma que seja fácil gerar uma va com densidade r.

- Método:

  1. Gere uma va Y com densidade r.
  2. Gere $U \sim U(0,1)$, independente de Y.
  3. Se $U \leq f(Y)/t(Y)$, retorne $X = Y$.
     Caso contrário, volte para o passo 1.

- Ex.: beta(4,3), para $0 \leq x \leq 1$,

$$f(x) = \frac{x^3(1-x)^2}{\int_0^1 x^3(1-x)^2 dx} = 60x^3(1-x)^2.$$

$$f'(x) = 5x^4 - 8x^3 + 3x^2 = 0 \Rightarrow x = 0,6.$$

$$t(x) = f(0,6) = 2,0736 \text{ para } 0 \leq x \leq 1.$$

$$c = \int_0^1 2,0736dx = 2,0736 \Rightarrow r(x) \sim U(0,1).$$

  - Geramos Y e U independentes com dist $U(0,1)$.
  - Aceitamos Y se

$$U \leq \frac{60Y^3(1-Y)^2}{2,0736}.$$

- Quanto maior a chance de rejeitar, mais lento o algoritmo.

  - A probabilidade de aceitar vale $1/c$.
  - Portanto, quanto mais próximo $t(x)$ for de $f(x)$, mais eficiente será o algoritmo.

- Uma estratégia comum é assumir que $r(x)$ é uma dist comum, como normal ou exponencial, e encontrar o menor c tal que $t(x) = cr(x) \geq f(x)$ para todo x.

```
n=1000; u=rand(1,n); y=rand(1,n); x=zeros(1,n); m=0;
for i=1:n; if u(i) <= 60*y(i)^3*(1-y(i))^2/2.0736;
   m=m+1; x(m)=y(i); end; end;
x=x(1:m); 1-m/n, hist(x); figure; ecdf(x); hold on;
   s=sort(x); plot(s,betacdf(s,4,3))
[h p]=chi2gof(x,'cdf',@(x)betacdf(x,4,3),'alpha',.05)
[h p]=kstest(x',[x' betacdf(x',4,3)],.05)
```

42 Métodos específicos

42.1 Uniforme

- Método:

  1. Gere $U \sim U(0,1)$.
  2. Retorne $X = a + (b-a)U$.

42.2 Exponencial

- Método: inversa

  1. Gere $U \sim U(0,1)$.
  2. Retorne $X = -\beta \ln(U)$.

42.3 Gama

- Não tem forma fechada para a inversa.

  - Podemos utilizar algum método numérico.
  - [Law] apresenta funções $r(x)$ apropriadas para aplicar o método aceita-rejeita.

42.4 Weibull

$$F(x) = 1 - e^{-(x/\beta)^\alpha} \Rightarrow F^{-1}(U) = \beta(-\ln(1-U)^{1/\alpha}).$$

- Método: inversa

  1. Gere $U \sim U(0,1)$.
  2. Retorne $X = \beta[-\ln(U)]^{1/\alpha}$.

42.5 Normal

- Podemos usar $X \sim N(0,1)$ para gerar $X' \sim N(\mu, \sigma^2)$, pois $X' = \mu + \sigma^2 X$.

- É importante ter uma função eficiente para gerar $X \sim N(0,1)$, pois esta dist é muito utilizada para gerar outras dist.

  - Não temos forma fechada para a inversa.
  - Existem métodos eficientes na literatura, não descritos aqui.

42.6 Lognormal

$$X \sim LN(\mu, \sigma^2) \quad \Leftrightarrow \quad \ln X \sim N(\mu, \sigma^2).$$

- Método:
  1. Gere $Y \sim N(\mu, \sigma^2)$.
  2. Retorne $X = e^Y$.

42.7 Beta

- $X' \sim beta(\alpha_1, \alpha_2)$ no intervalo $[a, b]$ pode ser obtida de $X \sim beta(\alpha_1, \alpha_2)$ no intervalo $[0, 1]$, fazendo $X' = a + (b - a)X$.
- Se $X_1 \sim gama(\alpha_1, \beta)$ e $X_2 \sim gama(\alpha_2, \beta)$ são va independentes, então $X_1/(X_1 + X_2) \sim beta(\alpha_1, \alpha_2)$.
- Método:
  1. Gere $Y_1 \sim gama(\alpha_1, 1)$ e $Y_2 \sim gama(\alpha_2, 1)$ independentes.
  2. Retorne $X = Y_1/(Y_1 + Y_2)$.

42.8 Empírica

- $x_1 < x_2 < \ldots < x_n$. $F(x_1) = 0, F(x_2) = 1/(n-1), \ldots, F(x_{n-1}) = (n-2)/(n-1), F(x_n) = 1$.
- Método: inversa
  1. Gere $U \sim U(0, 1)$.
  2. Seja $p = (n - 1)U$ e $i = \lfloor p \rfloor + 1$.
  3. Retorne $X = x_i + (p - i + 1)(x_{i+1} - x_i)$.

42.9 Bernoulli

- Método:
  1. Gere $U \sim U(0, 1)$.
  2. Se $U \leq p$, retorne $X = 1$.
     Caso contrário, retorne $X = 0$.

42.10 Uniforme discreto

- Método:
  1. Gere $U \sim U(0, 1)$.
  2. Retorne $X = i + \lfloor (j - i + 1)U \rfloor$.

42.11 Binomial

- Método:
  1. Gere $Y_1, Y_2, \ldots, Y_t$ como va de bernoulli independentes.
  2. Retorne $X = Y_1 + Y_2 + \cdots + Y_t$.
- Quando t é grande, é necessário utilizar abordagem mais eficiente.

42.12 Geométrica

$$F(x) = 1 - (1 - p)^{\lfloor x \rfloor + 1} \quad \Rightarrow \quad F^{-1}(U) = \lfloor \log_{1-p}(1 - U) \rfloor$$

- Método: inversa
  1. Gere $U \sim U(0, 1)$.
  2. Retorne $X = \lfloor \log_{1-p}(U) \rfloor$.

42.13 Binomial negativa

- Método:
  1. Gere $Y_1, Y_2, \ldots, Y_s$ como va $geo(p)$ independentes.
  2. Retorne $X = Y_1 + Y_2 + \cdots + Y_s$.
- Quando s é grande, é necessário utilizar abordagem mais eficiente.

42.14 Poisson

- Método:
  1. Faça $s \leftarrow 0$ e $i \leftarrow 0$.
  2. Enquanto $s \leq 1$,
     (a) Gere $Y \sim expo(1/\lambda)$.
     (b) $s \leftarrow s + Y$.
     (c) $i \leftarrow i + 1$
  3. Retorne i.

42.15 Discreta arbitrária

- Probabilidade: $p(x_1), p(x_2), \ldots, p(x_n)$.
- Podemos considerar que os possíveis valores são $1, 2, \ldots, n$, pois podemos indexar um vetor onde $v[1] = x_1, v[2] = x_2, \ldots, v[n] = x_n$.
- Digamos que as probabilidade tem 2 casas decimais (0,00 à 0,99).
- Inicializamos um vetor m com 100 elementos, colocando os primeiros $100p(1)$ valores em 1, os $100p(2)$ elementos seguintes com valor 2, assim sucessivamente.
- Método:
  1. Gere $i \sim DU(1, 100)$.
  2. Retorne $X = m[i]$.
- Desvantagem: armazenamento de 10 elevado ao número de casas decimais.
  - Existem estratégias mais sofisticadas para reduzir este armazenamento.