ACH2053 – Introdução à Estatística

Aula 08: Amostragem e Simulação

Valdinei Freire

valdinei.freire@usp.br

http://www.each.usp.br/valdinei

Escola de Artes, Ciências e Humanidades - USP

2025

Distribuição vs População

- Distribuição de probabilidade: descrita por meio de uma classe de funções parametrizada.
- População: cada indivíduo representa um resultado equiprovável.
- Dada uma população pode-se construir uma distribuição de probabilidade discreta.
  Exemplo: qual é a distribuição de idade nos alunos presentes na sala de aula?

Distribuição vs População

Dada uma distribuição de probabilidade ou população, pode-se calcular várias características:

Distribuição	População
$\mathsf{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$	$\mathsf{E}(X) = \frac{1}{n} \sum_{i=1}^{n} x_i$
$\mathsf{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$	$\mathsf{Var}(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$
$\Pr(X \leq x) = \int_{-\infty}^{x} f(t) dt$	$\Pr(X \leq x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i \leq x\}$
$\Pr(a \leq X \leq b) = \int_{a}^{b} f(t) dt$	$\Pr(a \leq X \leq b) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{a \leq x_i \leq b\}$

Calculo Numérico

Cálculo Numérico: solução **arbitrariamente aproximada** de problemas matemáticos (algébricos, diferenciais, lineares, não-lineares) por meio de **operações aritméticas algoritmizadas**.

Idéia: aproximar distribuição por população.

Como: Amostragem e Simulação.

Exemplo

Considere as variáveis aleatórias X e Y independentes e identicamente distribuídas (i.i.d.) com distribuição uniforme entre 0 e 1. Considere a variável aleatória $Z = X + Y$.

1. Calcule a probabilidade de $Z \leq 0.5$.
2. Calcule a probabilidade de $Z \geq 1.5$.
3. Calcule a c.d.f. de Z.
4. Calcule a p.d.f. de Z.
5. Calcule a esperança de Z.
6. Calcule a variância de Z.

Teorema

Para qualquer distribuição contínua com c.d.f. $F : \mathbb{R} \to [0, 1]$, se U é uma variável contínua com distribuição uniforme em $[0, 1]$, então $X = F^{-1}(U) = \inf\{x : F(x) \geq U\}$ segue a distribuição em F.

Exemplo: Especifique um procedimento para obter amostras da distribuição de Bernoulli.

Exemplo: Especifique um procedimento para obter amostras da distribuição exponencial $f(x) = \lambda e^{-\lambda x}$.

Exercícios

Exemplo: Especifique um procedimento para obter amostras da distribuição:

$$f(x) = \begin{cases} \frac{1}{9}x^2 & \text{para } 0 < x < 3, \\ 0 & \text{caso contrário} \end{cases}$$

Exemplo: Especifique um procedimento para obter amostras da distribuição:

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{para } 0 < x < 2, \\ 0 & \text{caso contrário} \end{cases}$$

Para cada um dos casos calcule por amostragem e analiticamente:

1. Calcule a esperança de Z.
2. Calcule a variância de Z.

Rejection Sampling

Teorema

Para qualquer distribuição contínua com p.d.f. $f : \mathbb{R} \to \mathbb{R}_+$, considere que exista uma variável aleatória Y com p.d.f. $g : \mathbb{R} \to \mathbb{R}_+$ tal que se possua um método para gerar amostras de Y e calcular $g(y)$. Considere que exista uma função $e : \mathbb{R} \to \mathbb{R}_+$ e $\alpha \geq 1$ tal que:

$$e(x) = \alpha g(x) \geq f(x), \text{ se } f(x) > 0.$$

Considere o seguinte procedimento:

1. Obtenha uma amostra y de Y
2. Obtenha uma amostra u da distribuição uniforme em $[0, 1]$
3. Rejeite y se $u > \frac{f(y)}{e(y)}$, e volte para o passo 1.
4. Caso contrário faça $x = y$.

O procedimento anterior produz uma amostra x que segue a p.d.f. $f(x)$.

Sampling Importance Resampling

Definition

Seja as p.d.f. multivariadas $f : \mathbb{R}^d \to \mathbb{R}_+$ e $g : \mathbb{R}^d \to \mathbb{R}_+$ tal que se $f(\mathbf{x}) > 0$ então $g(\mathbf{x}) > 0$. Considere o seguinte procedimento:

1. Obtenha uma amostragem $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M$ independente e identicamente distribuída de g
2. Calcule os pesos padronizados:

$$w_i = \frac{\frac{f(\mathbf{y}_i)}{g(\mathbf{y}_i)}}{\sum_{j=1}^{M} \frac{f(\mathbf{y}_j)}{g(\mathbf{y}_j)}}$$

3. Reamostre $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ de $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M$ com reposição e probabilidades $w_1, w_2, \ldots, w_M$.

Quando $m \to \infty$, a amostra $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ é equivalente a uma amostra vinda de f.