



Como sortear pontos uniformemente no disco

Matheus de Moraes

Universidade Federal de Uberlândia
Instituto de Matemática e Estatística
matheus.neves1@ufu.br

Pedro Franklin

Universidade Federal de Uberlândia
Instituto de Matemática e Estatística
pedrofranklin@ufu.br

Introdução

A motivação deste texto nasceu em sala: durante uma aula de *Probabilidade 3*, surgiu a pergunta “como sortear pontos de forma uniforme em um disco de raio R_0 ?”. Aqui, “uniforme” significa que regiões com a mesma *área* devem, em média, receber o mesmo número de pontos; isto é, a densidade é constante sobre o disco.

Uma abordagem intuitiva é gerar coordenadas polares (R, Θ) com $R \sim \text{Unif}(0, R_0)$ e $\Theta \sim \text{Unif}(0, 2\pi)$ e depois converter para coordenadas cartesianas considerando $X = R \cos \Theta$ e $Y = R \sin \Theta$. Contudo, isso não produz distribuição uniforme em área: como a área de um anel cresce linearmente com o raio, escolher R uniforme super-representa a região central e sub-representa a borda.

Neste trabalho apresentamos uma derivação elementar via mudança de variáveis, explicamos por que a escolha “ R uniforme” falha, e discutimos implementações práticas que geram pontos uniformes no disco de forma simples e eficiente. Usaremos o disco unitário para ilustrar as ideias, mas as conclusões se estendem a discos de raio arbitrário.

Como não sortear pontos uniformemente no disco

Considere o disco unitário $D = \{(x, y) : x^2 + y^2 \leq 1\}$. Defina o seguinte experimento: gere um valor de Θ uniforme em $[0, 2\pi]$; em seguida, gere um valor de R uniforme em $[0, 1]$; agora calcule $X = R \cos \Theta$ e $Y = R \sin \Theta$; repita o procedimento para obter muitas amostras (X, Y) .

A razão para o fracasso deste método pode ser explicada pelo seguinte argumento geométrico: embora a variável R seja uniforme, a *área* do disco não é uniformemente distribuída em relação a R . Ao sortear R de forma uniforme, o procedimento dá a mesma probabilidade para cada intervalo de R de igual comprimento, independentemente de sua localização. Por exemplo, o intervalo de R de $[0, 0.1]$ tem a mesma probabilidade de ser sorteado que o intervalo de $[0.9, 1.0]$. No entanto, o primeiro corresponde a um anel minúsculo próximo ao centro, enquanto o segundo corresponde a um anel muito maior na borda do disco. O resultado é que o mesmo número de pontos é gerado em uma área pequena e em uma área grande, levando a uma alta densidade de pontos no centro e a uma baixa densidade na borda. A distribuição, portanto, não é uniforme.

A Figura 1 apresenta uma simulação computacional do experimento descrito anteriormente com $R_0 = 1$: foram gerados 1000 valores de (R, Θ) tais que $R \sim \text{Unif}(0, 1)$ e $\Theta \sim \text{Unif}(0, 2\pi)$. Nesta figura, é possível ver que o centro do disco apresenta uma alta concentração de pontos, enquanto as regiões mais próximas à borda estão visivelmente mais esparsas.

Uma Abordagem Correta

Se o vetor aleatório (X, Y) possui distribuição uniforme no disco de raio $R_0 = 1$, então a densidade $f_{X,Y}(x, y)$ deste vetor é

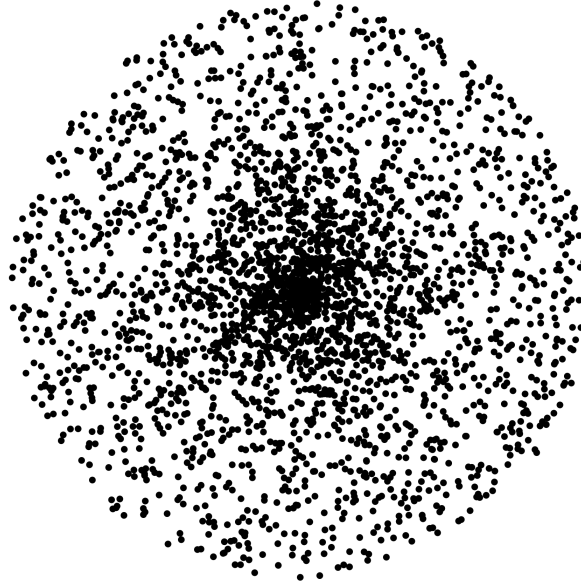


Figura 1: Amostragem *não* uniforme no disco ao tomar $R \sim \text{Unif}(0, 1)$ e $\Theta \sim \text{Unif}(0, 2\pi)$. O centro concentra mais pontos, enquanto a borda fica sub-representada.

$$f_{X,Y}(x, y) = \frac{\mathbb{1}_D(x, y)}{\text{área de } D} = \frac{\mathbb{1}_D(x, y)}{\pi R_0^2} = \frac{\mathbb{1}_D(x, y)}{\pi},$$

em que $\mathbb{1}_D(x, y)$ é a indicadora da região D . Por outro lado, considerando que $X = R \cos \Theta$ e $Y = R \sin \Theta$, temos que o jacobiano desta transformação é:

$$J = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$$

Logo, pelo método do jacobiano ([1]), segue que

$$f_{R,\Theta}(r, \theta) = f_{X,Y}(x, y)|J| = \frac{r}{\pi}, (x, y) \in D.$$

Agora, a partir das marginais, podemos obter as densidades de R e Θ . Vejamos:

$$f_R(r) = \int_0^{2\pi} f_{R,\Theta}(r, \theta) d\theta = 2r, \quad r \in (0, 1)$$

e

$$f_\Theta(\theta) = \int_0^1 f_{R,\Theta}(r, \theta) dr = \frac{1}{2\pi}, \quad \theta \in (0, 2\pi).$$

Particularmente, como a densidade de (R, Θ) pode ser escrita como o produto das densidades marginais, segue que R e Θ são variáveis independentes. Além disso, a partir da expressão da densidade de Θ , podemos afirmar que Θ segue uma distribuição uniforme em $[0, 2\pi]$.

Por fim, como temos a densidade de R e Θ , podemos usar estas densidades para gerar valores dessas duas variáveis. A variável Θ tem distribuição uniforme e, por isso, é fácil gerar valores dela. Para simular valores de R , recorreremos ao método da transformada inversa ([2]).

Teorema 1: Método da transformada inversa

Seja U uma variável aleatória uniforme em $(0, 1)$. Para qualquer função de distribuição contínua F , a variável aleatória X definida por

$$X = F^{-1}(U)$$

possui distribuição F .

Portanto, considerando o Teorema 1, se quisermos simularmos valores r de R , basta simularmos valores u de $U \sim U[0, 1]$ e definirmos $r = F_R^{-1}(u)$. Logo, resta apenas determinar F_R :

$$F_R(x) = \int_0^x 2r dr = x^2, x \in (0, 1).$$

Finalmente, um valor r de R poderá ser simulado tomando $r = \sqrt{u}$ em que u é um valor simulado de $U \sim [0, 1]$. Usando o R e a função `runif`, gerou-se, de forma independente, dois mil valores de $\Theta \sim U[0, 2\pi]$ e dois mil valores de $R \sim \sqrt{U}$. A Figura 2, apresenta os pontos sorteados uniformemente no disco D .

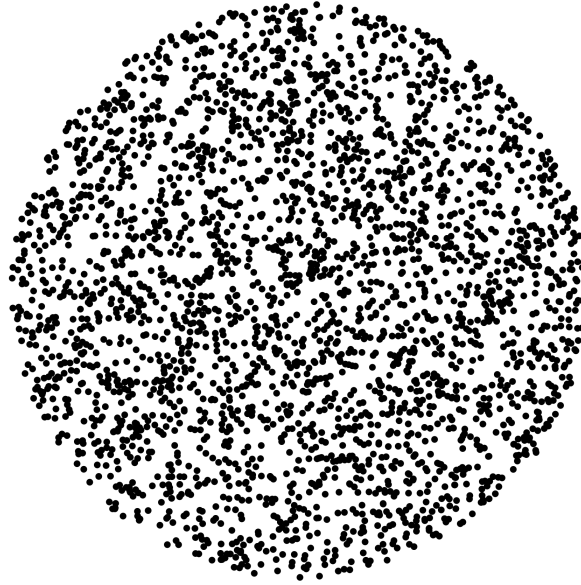


Figura 2: Amostragem uniforme no disco ao tomar $R \sim \sqrt{U}$, $U \sim U[0, 1]$, e $\Theta \sim \text{Unif}(0, 2\pi)$.

Referências

- [1] Barry James, *Probabilidade: um curso em nível intermediário*, IMPA, 1996.
- [2] Sheldon Ross, *Simulation*, Academic Press, 2010.