Variáveis aleatórias contínuas

Wagner H. Bonat
Elias T. Krainski
Fernando P. Mayer

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

18/04/2018

Sumário

Variáveis aleatórias

Em probabilidade, uma função X que associa a cada evento do espaço amostral um número real $X(\omega) \in \mathbb{R}$, é denominada uma **variável aleatória** (VA).

Uma variável aleatória pode ser classificada como **discreta** ou **contínua**, dependendo do domínio dos valores de X.

Exemplo: o número de alunos em uma sala é uma variável aleatória (discreta), denotada por X (maiúsculo). Uma observação dessa variável é denotada pela respectiva letra minúscula, *e.g.*, $x = 50$ alunos.

Em geral, denotamos a probabilidade de uma V.A. X assumir determinado valor x como

$$P[X] \quad \text{ou} \quad P[X = x]$$

Distribuições de probabilidade

Existem diversos *modelos probabilísticos* que procuram descrever vários tipos de variáveis aleatórias: são as **distribuições de probabilidade de variáveis aleatórias** (discretas ou contínuas).

A distribuição de probabilidades de uma VA X é, portanto, uma descrição das probabilidades associadas com os possíveis valores de X. Os valores que X assume determinam o **suporte** (S) da VA.

- **Variáveis discretas** $\rightarrow$ suporte em um conjunto de valores enumeráveis (finitos ou infinitos)
- **Variáveis contínuas** $\rightarrow$ suporte em um conjunto não enumerável de valores

Distribuições de probabilidade

Denomina-se de **distribuição de probabilidade** de alguma variável aleatória, a **regra** geral que define a

- **função de probabilidade** (fp) (V.A.s discretas), ou a
- **função densidade de probabilidade** (fdp) (V.A.s contínuas)

para a variável de interesse.

Existem muitas distribuições de probabilidade, mas algumas merecem destaque por sua importância prática.

Estas distribuições também são chamadas de **modelos probabilísticos**.

Variáveis aleatórias contínuas

Uma V.A. é classificada como contínua se assume valores em qualquer intervalo dos números reais, ou seja, um conjunto de valores não enumerável. Dessa forma, não é possível atribuir probabilidades para um ponto específico, apenas para intervalos da reta.

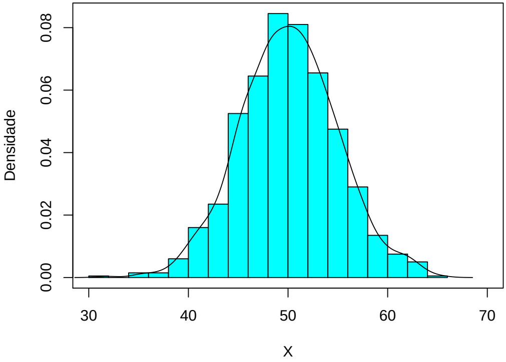
Exemplos:

- Peso de animais
- Tempo de falha de um equipamento eletrônico
- Altura da maré em uma hora específica
- Salinidade da água do mar
- Retorno financeiro de um investimento

Função densidade de probabilidade

Não podemos atribuir probabilidades à valores específicos, pois há uma quantidade **não enumerável** (infinita) de valores em um ponto.

Atribuimos probabilidades à intervalos de valores, por meio de uma **função**. Portanto, as probabilidades são representadas por áreas.



Função densidade de probabilidade

A **função densidade de probabilidade** (fdp) atribui probabilidades à intervalos de valores do tipo $[a, b]$, e é definida por

$$P[a < x < b] = \int_a^b f(x)dx$$

com as seguintes propriedades:

i. É uma função não negativa

$$f(x) \geq 0$$

ii. A área total sob a curva deve ser igual a 1

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

Função densidade de probabilidade

Observações:

- $P[X = x] = 0$, portanto:

$$P[a \leq X \leq b] = P[a < X \leq b] = P[a \leq X < b] = P[a < X < b]$$

- Qualquer função $f(\cdot)$ que seja não negativa e cuja área total sob a curva seja igual à unidade caracterizará uma VA contínua.
- $f(x)$ **não** representa a probabilidade de ocorrência de algum evento. A área sob a curva entre dois pontos é que fornecerá a probabilidade.

Exemplo

Seja a função:

$$f(x) = \begin{cases} 1,5x^2, & \text{se } -1 \leq x \leq 1 \\ 0, & \text{caso contrário} \end{cases}$$

- Verifique se essa função é uma fdp.
- Calcule:
  - $P[X > 0]$
  - $P[X > 0,5]$
  - $P[-0,5 \leq X \leq 0,5]$
  - $P[X < -2]$
  - $P[X < 0,5]$
  - $P[X < 0 \cup X > 0,5]$

Medidas de posição para VAs contínuas

- O valor esperado (ou média) da VA contínua X com função densidade $f(x)$, é dado pela expressão:

$$E(X) = \mu = \int_{-\infty}^{\infty} x\, f(x) dx.$$

- A mediana é o valor Md que tem a propriedade de

$$P(X \geq Md) \geq 0.5 \quad \text{e} \quad P(X \leq Md) \geq 0.5.$$

- A moda é o valor Mo tal que,

$$f(Mo) = \max_x f(x).$$

Variância para VAs contínuas

- Para uma VA X com densidade $f(x)$, a variância é dada por

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx.$$

- Expressão alternativa

$$Var(X) = \sigma^2 = E(X^2) - E^2(X).$$

onde

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx.$$

Exemplo

Seja a função:

$$f(x) = \begin{cases} 1,5x^2, & \text{se } -1 \leq x \leq 1 \\ 0, & \text{caso contrário} \end{cases}$$

Calcule $E(X)$, $Var(X)$, $DP(X)$.

Sumário

Modelo Uniforme contínuo

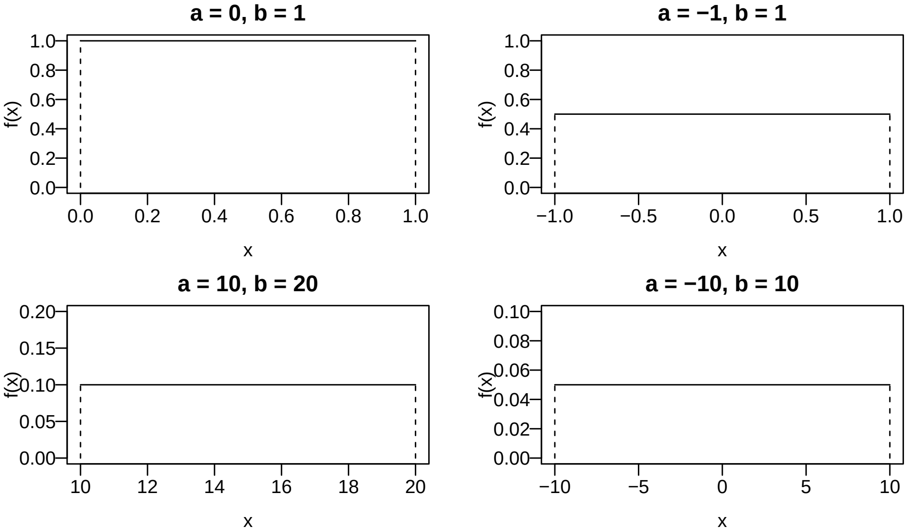
Definição: uma VA X tem distribuição Uniforme contínua no intervalo $[a, b]$, $a < b$, se sua função densidade de probabilidade é dada por

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{caso contrário} \end{cases}$$

Notação: $X \sim U(a, b)$

Esperança e variância: $E(X) = \frac{a+b}{2}$ e $Var(X) = \frac{(b-a)^2}{12}$.

Modelo Uniforme contínuo



Exemplo 6.5

- Com o objetivo de verificar a resistência à pressão de água, os técnicos de qualidade de uma empresa inspecionam os tubos de PVC produzidos. Os tudos inspecionados têm 6 metros de comprimento e são submetidos a grandes pressões até o aparecimento do primeiro vazamento, cuja distância a uma das extremidades (fixada à priori) é anotada para fins de analise. Escolhe-se um tubo ao acaso para ser inspecionado. Queremos calcular a probabilidade de que o vazamento esteja, a no máximo 1 metro das extremidades.

Modelo Exponencial

Definição: uma VA contínua X assumindo valores não negativos, segue o modelo exponencial com parâmetro $\alpha > 0$ se sua densidade é dada por

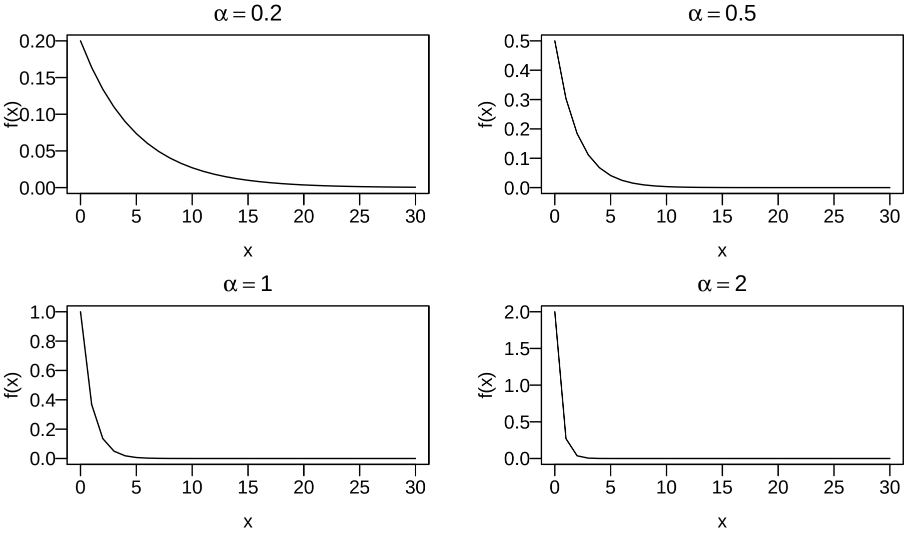
$$f(x) = \begin{cases} \alpha e^{-\alpha x} & \text{se } x \geq 0 \\ 0 & \text{caso contrário} \end{cases}$$

Notação: $X \sim \exp(\alpha)$

Esperança e variância: $E(X) = \mu = \frac{1}{\alpha}$ e $Var(X) = \frac{1}{\alpha^2}$.

Obs.: $P(a < X < b) = \int_a^b \alpha e^{-\alpha x}\, dx = e^{-\alpha a} - e^{-\alpha b}$.

Modelo exponencial



Exemplo 6.6

- Uma indústria fabrica lâmpadas especiais que ficam em operação continuamente. A empresa oferece a seus clientes a garantia de reposição, caso a lâmpada dure menos de 50 horas. A vida útil dessas lâmpadas é modelada através da distribuição Exponencial com parâmetro $1/8000$. Determine a proporção de troca por defeito de fabricação.

Exemplo 6.7

- O intervalo de tempo, em minutos, entre emissões consecutivas de uma fonte radioativa é uma variável aleatória com distribuição Exponencial de parâmetro $\alpha = 0,2$. Calcule a probabilidade de haver uma emissão em um intervalo inferior a 2 minutos.

Modelo Normal

Definição: Dizemos que uma VA X segue o modelo normal se sua fdp é a seguinte

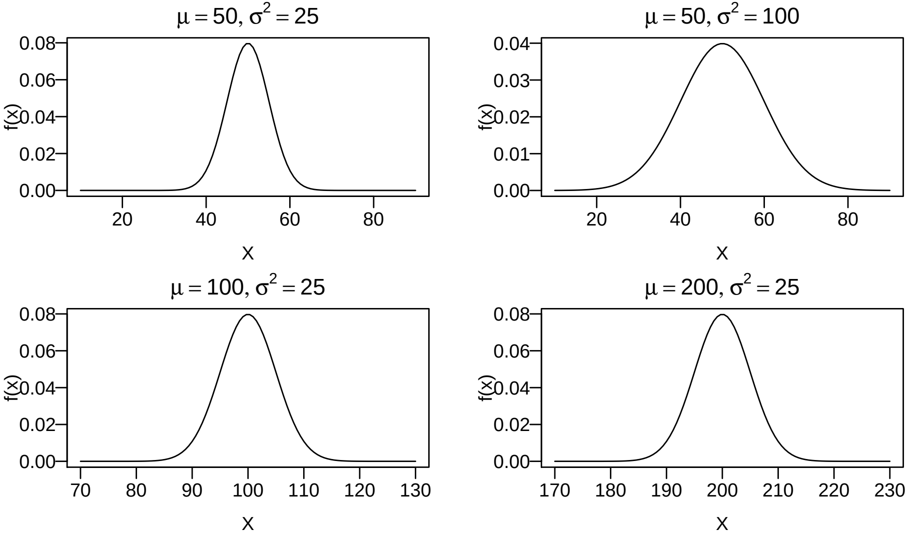
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty$$

onde $\mu \in \mathbb{R}$ é a média da população, $\sigma \in \mathbb{R}^+$ é o desvio-padrão populacional.

Notação: $X \sim \mathrm{N}(\mu, \sigma^2)$

Esperança e variância: $E(X) = \mu$ e $Var(X) = \sigma^2$

Modelo Normal



Modelo normal

Característcas da curva normal:

- É **simétrica** em relação à μ
- O ponto máximo (moda) de $f(x)$ é o ponto $x = \mu$
- Os pontos de inflexão da função são $\mu - \sigma$ e $\mu + \sigma$
- A área total sob a curva é 1 ou 100%
- A curva é **assintótica** em relação ao eixo x

Modelo normal

Para qualquer VA normal X, valem as seguintes relações:

$$\begin{aligned}
P[X > \mu] &= P[X < \mu] \\
P[\mu - \sigma < X < \mu + \sigma] &\approxeq 0,6827 \\
P[\mu - 2\sigma < X < \mu + 2\sigma] &\approxeq 0,9545 \\
P[\mu - 3\sigma < X < \mu + 3\sigma] &\approxeq 0,9973
\end{aligned}$$

Portanto, 6σ é frequentemente referida como a **largura** de uma distribuição normal.

Métodos mais avançados de integração podem ser utilizados para mostrar que a área sob a função densidade de probabilidade normal de $-\infty < x < \infty$ é igual a 1.

Modelo normal

Para obter uma probabilidade do modelo normal, devemos calcular a área entre os pontos a e b, ou seja,

$$P[a < X < b] = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

No entanto, a função da distribuição normal não possui forma fechada, portanto o cálculo de probabilidades não pode ser feito diretamente pela integral, apenas por aproximações numéricas.

Para contornar esse problema, os valores de probabilidade são obtidos para uma distribuição normal padrão (Z) com $\mu = 0$ e $\sigma^2 = 1$,

$$Z = \frac{X - \mu}{\sigma} \sim \mathrm{N}(0, 1)$$

que é o escore Z (número de desvios-padrões da média μ). A vantagem é que podemos fazer uma única tabela com as integrais aproximadas de Z, ao invés de uma tabela para cada par (μ, σ^2).

Modelo normal

Se $Z \sim \mathrm{N}(0, 1)$, então sua fdp é

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z)^2\right]$$

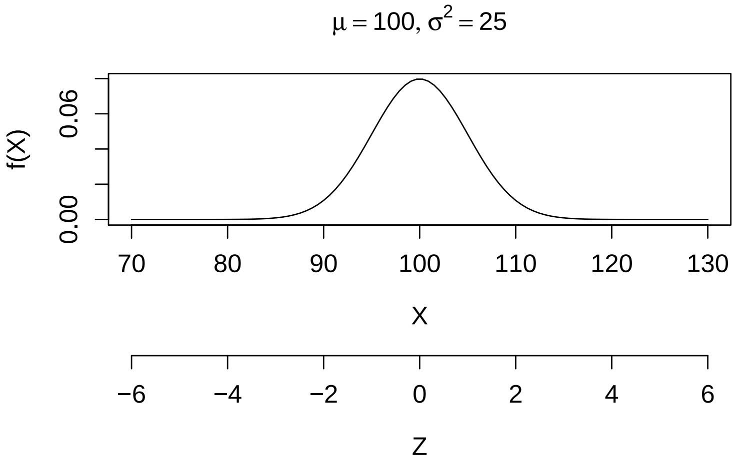
Para se obter a probabilidade de Z estar entre a e b,

$$P[a < Z < b] = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z)^2\right] dz$$

As integrais (áreas) para valores de Z entre 0,00 e 3,99 estão na tabela. Portanto, para qualquer valor de X entre a e b, podemos calcular a probabilidade correspondente através da transformação,

$$P[a < X < b] = P\left[\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right]$$

Modelo normal



Modelo normal

Exemplo de uso da tabela

- Calcule as probabilidades (áreas):
  - $P(0 < Z < 2)$
  - $P(Z > 2)$
  - $P(Z < -2)$
  - $P(2,0 < Z < 2,5)$
  - $P(-2,61 < Z < 2,43)$
  - $P(Z > -1,63)$

Exemplo 6.8

- Doentes sofrendo de certa moléstia são submetidos a um tratamento intensivo cujo tempo de cura foi modelado por uma densidade Normal de média 15 e desvio padrão 2 (em dias). Seja X o tempo de cura e, portanto temos $X \sim N(15, 4)$. Calcule a proporção de pacientes que demorão mais de 17 dias para se recuperar. Calcular a probabilidade um paciente ao acaso demorar menos de 20 dias para se recuperar. Qual o número esperado de dias para recuperação?

Normal como aproximação da binomial

Exemplo 6.9

- Estudo do Sindicato dos Bancários indica que cerca de 30% dos funcionários de banco têm problemas de estresse, provenientes das condições de trabalho. Numa amostra de 200 bancários, qual seria a probabilidade de pelo menos 50 com essa doença?

Combinação linear de Normais independentes

Exemplo 6.10

- Um serviço de fiscalização é criado para averiguar se garrafas de um certo refrigerante contém, de fato, o volume especificado pelo fabricante. Para tanto, 10 garrafas do produto são compradas no varejo, em várias regiões da cidade. Cada uma dessas garrafas é esvaziada e o volume de seu conteúdo, que denotaremos por V é aferido. Uma vez obtidos os 10 valores, a média aritmética M é calculada e, se $M < 290$ mililitros (ml), a companhia é multada. Estudos na linha de produção do fabricante mostraram que variações sempre ocorrem, mesmo se as especificações forem seguidas. Por essa razão, considera-se o volume do conteúdo das garrafas como seguindo o modelo Normal, com média $\mu = 300$ ml e desvio-padrão $\sigma = 25$ ml. Gostaríamos de calcular qual é a probabilidade de que o fabricante seja multado injustamente.

Exemplo 6.11

- Uma corretora negocia títulos na Bolsa de Valores e utiliza um modelo probabilístico para avaliar seus lucros. Suas aplicações financeiras de compra e venda atingem três áreas: agricultura, indústria e comércio. Admita que o seguinte modelo representa o comportamento do lucro diário da corretora (em milhares):

  $$L = 2L_A + 5L_I + 3L_C,$$

  onde L_A, L_I e L_C representam, os lucros diários nos setores de agricultura, indústria e comércio. As distribuições de probabilidades dessas variáveis aleatórias são $L_A \sim N(3, 4)$, $L_I \sim N(6, 9)$ e $L_C \sim N(4, 16)$. Supondo independência entre os três setores, qual será a probabilidade de um lucro diário acima de 50 mil.

Exercícios recomendados

- Seção 6.1 - $1, 2, 3, 4$ e 5.
- Seção 6.2 - 1 a 9.