Distribuições de probabilidade

Visão geral, principais modelos contínuos

Prof. Me. Lineu Alberto Cavazani de Freitas

Departamento de Estatística
Laboratório de Estatística e Geoinformação

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

Introdução

Introdução

- Vimos anteriormente conceitos a respeito de **variáveis aleatórias** e funções que atribuem probabilidades aos possíveis valores das variáveis (fp e fdp).

- Na prática temos a **evidência empírica**, isto é, o que o dado mostra.

- Com base na evidência empírica precisamos chegar a **funções** que atribuam probabilidades aos possíveis resultados das variáveis aleatórias.

- O processo para obtenção destas funções pode ser complexo.

Modelos de probabilidade

- Existe um conjunto de **distribuições de probabilidade** que podem ser utilizadas para descrever fenômenos: **os modelos**.
- De forma geral, os modelos são **comportamentos teóricos** que vão servir como instrumento para estudar fenômenos aleatórios com características comuns.
- A ideia é que em vez de construir a função de probabiliade ou densidade de probabilidade para o problema possamos usar uma expressão genérica.
- Tentamos obter a melhor combinação entre dado e modelo.

Modelos de probabilidade

- Um modelo possui **parâmetros**: quantidades desconhecidas que assumem valores dentro de um intervalo (espaço paramétrico) que definem características da distribuição.

- Estes parâmetros são estimados por meio dos dados.

- Se o modelo se adequar bem aos dados, utilizamos o modelo para determinar probabilidades, estimar parâmetros, testar hipóteses, avaliar efeito de outras variáveis, fazer predições, etc.

- Em alguns casos sabemos a priori o modelo que descreve bem o fenômeno.

- Em outros casos precisamos encontrar este modelo.

Modelos de probabilidade

- Existem diversos modelos disponíveis.
- Muitos destes modelos aplicáveis a problemas similares.
- E diferentes modelos podem apresentar vantagens e desvantagens.
- Veremos alguns dos principais modelos **discretos** e **contínuos** com foco na **definição** de cada um deles, suposições, fp ou fdp (expressão e comportamento), média, variância e também exemplos.

Modelos de probabilidade

Alguns dos modelos que serão discutidos:

- Principais modelos discretos:
  - Uniforme discreta.
  - Bernoulli.
  - Binomial.
  - Poisson.
  - Hipergeométrico.
- Principais modelos contínuos:
  - Uniforme contínua.
  - Normal.
  - Exponencial.

Modelo Uniforme Contínuo

Modelo Uniforme Contínuo

Definição

Uma variável aleatória Y segue o modelo Uniforme Contínuo se o resultado for um número real em um intervalo com limites conhecidos a e b, em que $a < b$ e todos os valores do domínio tem igual probabilidade de ocorrência.

Notação

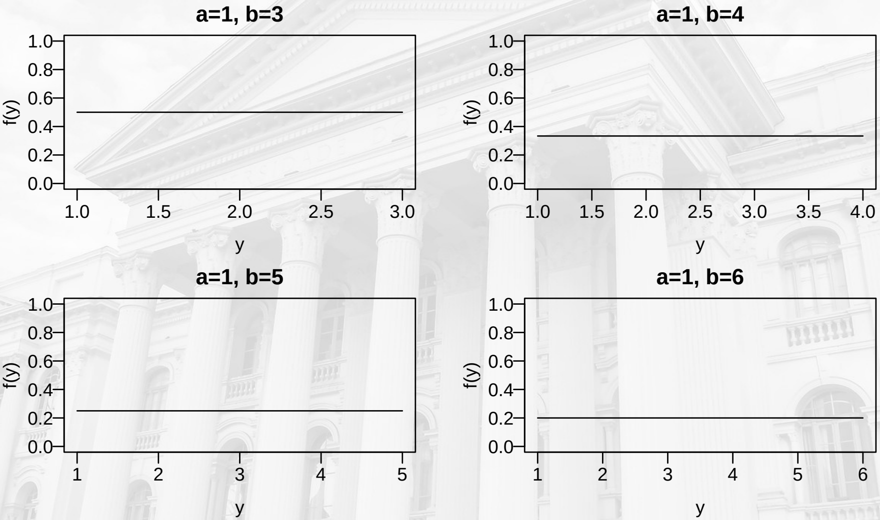
- $Y \sim \text{UC}(a,b)$

Função densidade de probabilidade

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

- Para calcular probabilidades não é necessário fazer uso de integrais, basta calcular a área sob o retângulo desejado.
  - $\mu = E(Y) = \frac{b+a}{2}$
  - $\sigma^2 = Var(Y) = \frac{(b-a)^2}{12}$

Modelo Uniforme Contínuo



Exemplo

Considere o experimento aleatório que consiste em verificar o surgimento de um defeito em uma pista de um trecho de rodovia com extensão de 20 km. Considere que existem razões para crer que a probabilidade da ocorrência de um defeito é constante para todo o trecho. Qual a probabilidade de observar um defeito entre o km 10 e 15?

Y : km em que ocorre o defeito.

$$Y \sim \text{UC}(p = 1/(20-0) = 1/20)$$

$$P(10 < Y < 15) = 0{,}25$$

Modelo Exponencial

Modelo Exponencial

Definição

- Distribuição usada para modelar variáveis aleatórias contínuas não negativas.
- Muito usada para modelar problemas que dizem respeito ao tempo até ocorrência de um evento.
- Tem como característica a falta de memória, isto é, a propensão à falha independe do tempo decorrido.
- A variável aleatória Y é contínua, não negativa e tem parâmetro $\alpha > 0$.

Modelo Exponencial

Notação

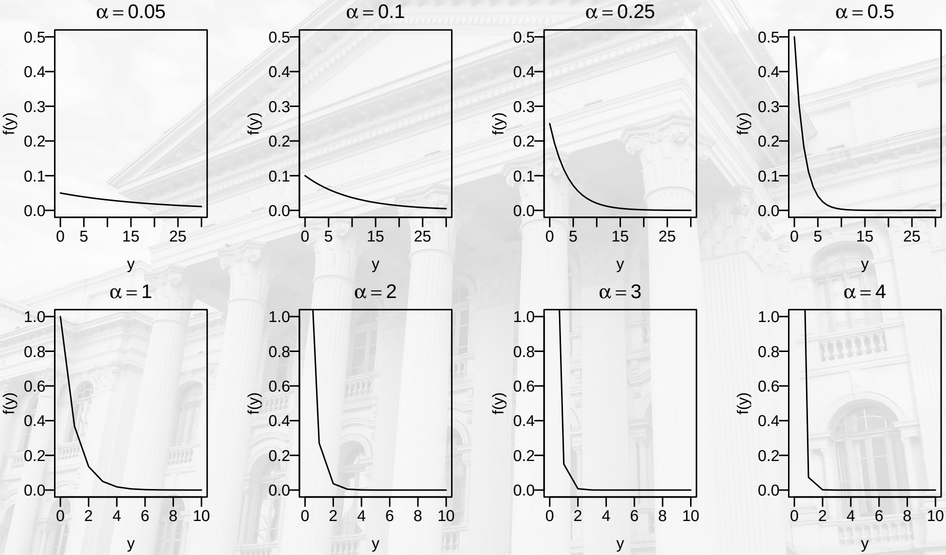
- $Y \sim \text{Exp}(\alpha)$

Função densidade de probabilidade

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

- $E(Y) = \mu = \frac{1}{\alpha}$.
- $Var(Y) = \frac{1}{\alpha^2}$.
- $P(a < Y < b) = \int_a^b \alpha e^{-\alpha y}\, dy = e^{-\alpha a} - e^{-\alpha b}$.

Modelo Exponencial



Exemplo

A duração do atendimento de cada cliente pelo sistema drive-thru de uma rede fast food tem distribuição Exponencial com tempo médio de atendimento de 10 minutos, o que implica em $\alpha = 1/10$.

1. Qual a probabilidade de um atendimento durar menos de 5 minutos?
2. Qual a probabilidade do atendimento ocorrer entre 4 e 6 minutos?

Exemplo

Y: Tempo de atendimento.

$$Y \sim \text{Exp}(\alpha = 1/10)$$

1. Qual a probabilidade de um atendimento durar menos de 5 minutos?
   - $P(Y < 5) = 0{,}3935$
2. Qual a probabilidade do atendimento ocorrer entre 4 e 6 minutos?
   - $P(4 < Y < 6) = 0{,}0638$

Modelo Normal

Modelo Normal

Definição

- É a mais importante distribuição contínua.
- Modela adequadamente a distribuição de um grande número de variáveis.
- Serve de aproximação para diversas outras distribuições.
- Tem papel central na Teoria Estatística, fundamentando a obtenção de inferências em diferentes contextos.
- A variável aleatória Y é contínua e assume valores de $-\infty$ até $+\infty$.

Modelo Normal

- Modela variáveis aleatórias contínuas não limitadas.
- Tem comportamento simétrico e em formato de sino.
- A função densidade de probabilidade é complexa.
- A integral da função não tem forma fechada.
- São necessários métodos numéricos ou a consulta a tabelas.
- Propriedades interessantes:
  - $P(\mu - \sigma < Y < \mu + \sigma) \approxeq 0{,}683$
  - $P(\mu - 2\sigma < Y < \mu + 2\sigma) \approxeq 0{,}954$
  - $P(\mu - 3\sigma < Y < \mu + 3\sigma) \approxeq 0{,}997$
  - Combinações lineares de variáveis aleatórias com distribuição Normal também têm distribuição Normal.

Modelo Normal

Notação

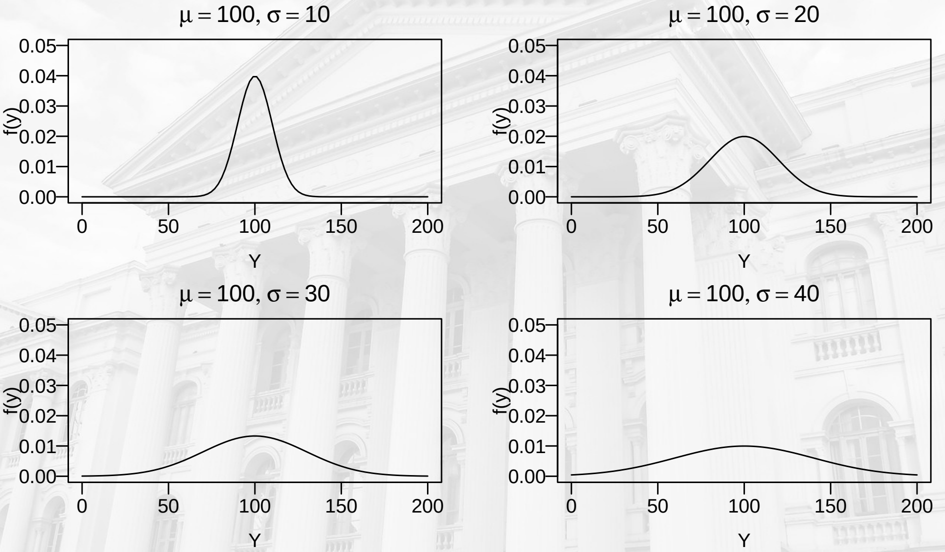
- $Y \sim \mathrm{N}(\mu, \sigma^2)$.

Função densidade de probabilidade

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

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

Modelo Normal



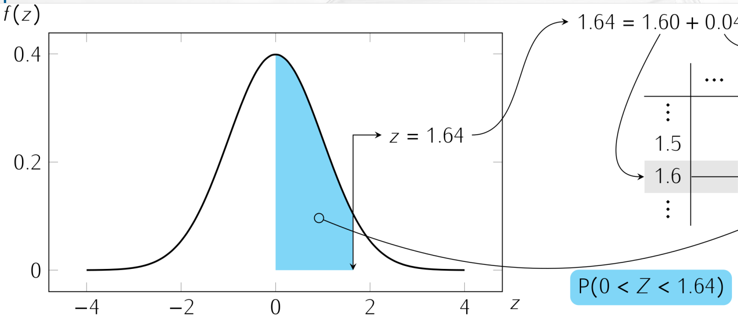
Normal padrão

- Para obter uma probabilidade do modelo normal, devemos calcular a área entre os pontos a e b.
- Contudo é difícil integrar uma função densidade como a da Normal.
- Devido à sua simetria, qualquer distribuição Normal pode ser padronizada de tal modo que $\mu = 0$ e $\sigma^2 = 1$.

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

- Esta distribuição é chamada normal padrão (Z).
- Esta transformação facilita o cálculo de probabilidades pois podemos usar uma única tabela de integrais.

Normal padrão



Probabilidades para a distribuição normal padrão.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.000000	0.003989	0.007978	0.011966	0.015953	0.019939	0.023922	0.027903	0.031881	0.035856
0.10	0.039828	0.043795	0.047758	0.051717	0.055670	0.059618	0.063559	0.067495	0.071424	0.075345
0.20	0.079260	0.083166	0.087064	0.090954	0.094835	0.098706	0.102568	0.106420	0.110261	0.114092
0.30	0.117911	0.121720	0.125516	0.129300	0.133072	0.136831	0.140576	0.144309	0.148027	0.151732
0.40	0.155422	0.159097	0.162757	0.166402	0.170031	0.173645	0.177242	0.180822	0.184386	0.187933
0.50	0.191462	0.194974	0.198468	0.201944	0.205401	0.208840	0.212260	0.215661	0.219043	0.222405
0.60	0.225747	0.229069	0.232371	0.235653	0.238914	0.242154	0.245373	0.248571	0.251748	0.254903

Normal padrão

- As integrais (áreas) para valores de Z entre 0,00 e 3,99 estão na tabela.
- Para qualquer valor de Y entre a e b, podemos calcular a probabilidade correspondente por meio da transformação.

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

Exemplo

Considere que altura de indivíduos de determinada modalidade esportiva se comporta como um modelo Normal com média 180 cm e variância de 7^2 cm.

1. Qual é a probabilidade de observar um indivíduo com mais de 1,90m?
2. Qual é a probabilidade de observar um indivíduo com menos de 1,60?
3. Qual é a probabilidade de observar um indivíduo entre 1,65 e 1,85?

Exemplo

Y: altura dos indivíduos.

$$Y \sim \mathrm{N}(\mu = 180, \sigma^2 = 7^2)$$

1. Qual é a probabilidade de observar um indivíduo com mais de 1,90m?
   - $P(Y > 190) = P\left(Z > \frac{190-180}{7}\right) = P(Z > 1{,}4286) = 0{,}0778$
2. Qual é a probabilidade de observar um indivíduo com menos de 1,60?
   - $P(Y < 160) = P\left(Z < \frac{160-180}{7}\right) = P(Z < -2{,}8571) = 0{,}0021$
3. Qual é a probabilidade de observar um indivíduo entre 1,65 e 1,85?
   - $P(165 < Y < 185) = P\left(\frac{165-180}{7} < Z < \frac{185-180}{7}\right) = P(-2.1428 < Z < 0.7143) = 0{,}7464$

Considerações finais

Considerações

- Existem muitos outros modelos na literatura.
  - Generalizações de modelos clássicos.
  - Modelos para outros fins.
- Devemos estar atentos aos pressupostos e parametrizações.
- Outros modelos discretos:
  - Geométrica.
  - Binomial Negativa.
- Outros modelos contínuos:
  - Lognormal.
  - Gama.
  - Weibull.
  - Beta.
- Modelos multivariados:
  - Distribuição multinomial.
  - Normal multivariada.
  - Distribuição de Dirichlet.

O que foi visto:

- Modelos de probabilidade.
- Alguns modelos contínuos.

Próximos assuntos:

- Inferência estatística.