

Curso de Verão em Estatística

Teoria da Probabilidade

Regis A. Ely
Mestrado em Economia Aplicada (PPGOM)
Universidade Federal de Pelotas (UFPel)

26 de fevereiro de 2016

Resumo

Este é a primeira aula do curso de verão em Estatística do Mestrado em Economia Aplicada do PPGOM. A referência principal são os capítulos de 1 a 5 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001*. Existe uma versão traduzida para o português deste livro, porém as notas de aula são baseadas na versão em inglês, então pode haver pequenas diferenças nos termos utilizados. Estas notas também contém exemplos de aplicação dos conceitos no R. Note que as aplicações no R, apesar de contribuírem para o entendimento do conteúdo, não substituem os exercícios do livro, que devem ser a fonte primária de preparação para as avaliações.

Sumário

1	Teoria de Probabilidade	2
1.1	Experimento, espaço amostral e eventos	2
1.2	Função de probabilidade	3
1.3	Probabilidade condicional e independência	6
1.4	Exemplo de aplicação no R	7
1.5	Variáveis aleatórias	9
1.6	Funções de distribuição	9
1.7	Funções de densidade e massa	10
2	Transformações em variáveis aleatórias	11
2.1	Distribuições de funções de variáveis aleatórias	11
2.2	Valor esperado	13
2.3	Variância e outros momentos	14
2.4	Função geradora de momentos	14
2.5	Exemplo no R com estatísticas descritivas	15
3	Distribuições de probabilidade	15
3.1	Distribuições discretas	15
3.2	Distribuições contínuas	15
3.3	Exemplos de distribuições de probabilidade no R	16
4	Variáveis aleatórias múltiplas	16
5	Propriedades de amostras aleatórias	16

1 Teoria de Probabilidade

1.1 Experimento, espaço amostral e eventos

Um dos principais objetivos da estatística é tirar conclusões sobre uma população de objetos conduzindo um experimento. Um *experimento aleatório* é sempre composto por uma ação e uma observação, embora às vezes a ação, ou até mesmo a observação, esteja subentendida na descrição do experimento. Se pudermos repetir um experimento um grande número de vezes então algumas regularidades poderão ser observadas, de modo a facilitar o processo de descrever o conjunto de resultados possíveis e as probabilidades associadas a eles.

Exemplo 1.1.1. (Experimentos aleatórios)

- i) Jogar dois dados e observar a soma dos números obtidos;
- ii) Jogar um dado justo (observação omitida: olhar para a face);
- iii) Observar o tempo que uma pessoa leva para ir da sua casa até o trabalho;
- iv) Observar o número de gols em uma partida de futebol;
- v) Observar o lucro de uma empresa no ano de 2010.

Definição 1.1.2. O conjunto S , de todos os possíveis resultados de um experimento é chamado *espaço amostral* do experimento.

Exemplo 1.1.3. (Espaços amostrais do Exemplo)

- i) $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$;
- ii) $S = \{H, T\}$;
- iii) $S = (0, \infty)$;
- iv) $S = \{0, 1, 2, 3, 4, \dots, 20\}$;
- v) $S = (-\infty, \infty)$.

Uma vez que definimos o espaço amostral, estamos interessados em calcular probabilidades de coleções de subconjuntos específicos deste espaço amostral (e.g. probabilidade dos dois dados somarem 12).

Definição 1.1.4. Um *evento* é qualquer coleção de possíveis resultados de um experimento, ou seja, qualquer subconjunto do conjunto S (incluindo o próprio S e o conjunto vazio \emptyset).

Exemplo 1.1.5. (Eventos)

- i) Suponha o experimento de selecionar uma carta de um baralho e verificar seu naipe: Ouros (O), Copas (C), Espadas (E) e Paus (P). O espaço amostral será $S = \{O, C, E, P\}$. Alguns eventos possíveis são $A = \{C, P\}$ e $B = \{C, E, O\}$. Logo, $A \cup B = \{O, C, E, P\}$, $A \cap B = \{C\}$ e $A^C = \{E, O\}$.

Dizemos que um evento A ocorre se o resultado de um experimento está no conjunto A . De-sejam atribuir probabilidades a eventos e não a experimentos ou espaços amostrais. Quando

atribuímos uma probabilidade a um evento chamamos ele de *evento aleatório*. Podemos realizar operações com eventos da mesma maneira que realizamos operações com conjuntos matemáticos.

Revisar teoria dos conjuntos:

1. Operações de união, interseção, complementar e produto cartesiano.
2. Propriedades comutativa, associativa, distributiva e leis de DeMorgan.
3. Conjuntos disjuntos (mutuamente excludentes) e partições.
4. Conjuntos finitos, enumeráveis e não-enumeráveis.

Quais as referências?

- Capítulo 1, seção 1.1 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*
- Capítulos 1 e 2 de *Elon, L. L. Curso de análise vol. 1. Impa, 2007.*
- Capítulo 1, seção 1.2 de *Meyer, P. L. Probabilidade: aplicações à estatística. Ed. LTC, 1983.*
- [Aula 5](#) de Métodos Estatísticos Básicos

1.2 Função de probabilidade

A noção de probabilidade pode ter várias interpretações distintas. Quando pensamos na probabilidade de sair um número 6 ao jogar um dado estamos utilizando a *abordagem clássica de probabilidade*, que corresponde ao número de resultados favoráveis ao nosso evento dividido pelo número de resultados possíveis (1/6). Essa abordagem é útil somente para espaços amostrais finitos e resultados igualmente verossímeis.

Outra abordagem possível é a *frequentista*, que nos diz que se repetimos um experimento um grande número de vezes podemos aproximar a probabilidade de um evento pelo número de vezes que ele ocorreu dividido pelo número de vezes que repetimos o experimento (tente jogar 100 vezes o dado e verifique quantas vezes saiu o 6). Essa é uma definição mais ampla de probabilidade, mas ainda assim, para calcularmos ela é necessário repetir o experimento um grande número de vezes, o que pode ser um empecilho.

Por outro lado, quando trabalhamos com espaços amostrais infinitos não-enumeráveis, é comum utilizarmos a *definição geométrica de probabilidade*, dada pela razão entre a área do evento A e a área do espaço amostral S .

Vimos que sempre atribuímos probabilidades a eventos, logo, se a probabilidade é uma função, o domínio dessa função deve ser uma coleção de conjuntos que correspondem a todos os possíveis eventos aleatórios que podem ser gerados a partir de um experimento.

Definição 1.2.1. Uma coleção de subconjuntos de um espaço amostral S é chamada de *sigma álgebra* (ou campo de Borel), e denotada por \mathcal{B} , se satisfaz as seguintes propriedades:

- a. $\emptyset \in \mathcal{B}$ (o conjunto vazio é um elemento de \mathcal{B}).
- b. Se $A \in \mathcal{B}$ então $A^C \in \mathcal{B}$ (\mathcal{B} é fechado sobre operações de complementar).

c. Se $A_1, A_2, \dots \in \mathcal{B}$, então $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ (\mathcal{B} é fechado sobre uniões contáveis).

Dizemos que o conjunto $\{\emptyset, S\}$ é uma *sigma álgebra trivial*. Estamos interessados na menor sigma álgebra que contém todos os subconjuntos de um dado espaço amostral. Se o conjunto S é finito ou infinito enumerável, \mathcal{B} será igual a todos os subconjuntos de S , que totalizarão 2^n conjuntos, onde n é o número de elementos de S . Quando S é não-enumerável é um pouco mais difícil descrever \mathcal{B} .

Exemplo 1.2.2. (Sigma álgebra)

- i) Se $S = \{1, 2, 3\}$, então \mathcal{B} terá $2^3 = 8$ conjuntos, dados por $\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}$.
- ii) Se $S = (-\infty, \infty) = \mathbb{R}$, então \mathcal{B} pode ser escolhido de modo a incluir os conjuntos $[a, b], (a, b], (a, b)$ e $[a, b)$, para todos os números reais a e b .

Agora que conhecemos o domínio da função de probabilidade, podemos defini-la.

Definição 1.2.3. Dado um espaço amostral S e um sigma álgebra associada \mathcal{B} , uma *função de probabilidade* é uma função P com domínio \mathcal{B} que satisfaz:

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$.
2. $P(S) = 1$.
3. Se $A_1, A_2, \dots \in \mathcal{B}$ forem disjuntos dois a dois, então $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

As propriedades acima são conhecidas como os *Axiomas de Probabilidade* ou *Axiomas de Kolmogorov*. Qualquer função que satisfaz esses axiomas é dita uma função de probabilidade. Como vimos, existem diversas interpretações de probabilidade, mas todas devem respeitar estes axiomas. Se quisermos rapidamente construir uma função de probabilidade, o teorema abaixo nos dá uma fórmula simples.

Teorema 1.2.4. Seja $S = \{s_1, \dots, s_n\}$ um conjunto finito e \mathcal{B} uma sigma álgebra de subconjuntos de S . Para quaisquer $A \in \mathcal{B}$, defina $P(A)$ como:

$$P(A) = \sum_{\{i: s_i \in A\}} p_i,$$

onde p_1, \dots, p_n são números não negativos que somam 1. Então P é uma função de probabilidade em \mathcal{B} . Esse teorema continua válido se S é um conjunto infinito enumerável.

Demonstração. Basta verificar a validade dos 3 Axiomas de Probabilidade. □

A partir dos Axiomas de Probabilidade podemos deduzir algumas propriedades úteis das funções de probabilidade.

Teorema 1.2.5. Se P é uma função de probabilidade e $A, B \in \mathcal{B}$, então:

- a. $P(\emptyset) = 0$, onde \emptyset é o conjunto vazio;
- b. $P(A) \leq 1$;
- c. $P(A^C) = 1 - P(A)$;
- d. $P(B \cap A^C) = P(B) - P(A \cap B)$;

- e. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- f. Se $A \subset B$, então $P(A) \leq P(B)$;
- g. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ para qualquer partição C_1, C_2, \dots ;
- h. $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ para quaisquer conjuntos A_1, A_2, \dots (*Desigualdade de Boole*)

Demonstração. Feita em aula. □

Note que se aplicarmos a desigualdade de Boole para A^C obtemos:

$$P(\cup_{i=1}^n A_i^C) \leq \sum_{i=1}^n P(A_i^C),$$

e usando o fato de que $\cup A_i^C = (\cap A_i)^C$ e $P(A_i^C) = 1 - P(A_i)$, então:

$$1 - P(\cap_{i=1}^n A_i) \leq n - \sum_{i=1}^n P(A_i).$$

Rearranjando temos:

$$P(\cap_{i=1}^n A_i) \geq \sum_{i=1}^n P(A_i) - (n - 1),$$

que é conhecida como a *Desigualdade de Bonferroni*. Esta desigualdade nos dá a possibilidade de limitar a probabilidade de um evento simultâneo (interseção) em termos das probabilidades individuais.

Nessa seção definimos os três elementos que compõem um *espaço de probabilidade* (S, \mathcal{B}, P) .

Definição 1.2.6. Um *espaço de probabilidade* (S, \mathcal{B}, P) é composto por:

- i) Um conjunto não-vazio S de todos os resultados possíveis de um experimento, chamado espaço amostral;
- ii) Uma sigma álgebra de eventos aleatórios \mathcal{B} , que contém todos os subconjuntos do espaço amostral;
- iii) Uma função de probabilidade P com domínio \mathcal{B} e que satisfaz os axiomas de probabilidade.

Este é o espaço matemático que trabalharemos a partir de agora.

Revisar análise combinatória:

1. Regra da multiplicação e da adição;
2. Permutações: ${}_nP_n = n!$;
3. Arranjos: ${}_nA_r = \frac{n!}{(n-r)!}$;
4. Combinações: $\binom{n}{r} = \frac{n!}{r!(n-r)!}$;
5. Permutação com elementos repetidos.

Quais as referências?

- Capítulo 1, seção 1.2.3 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*
- Capítulo 2, seção 2.3 de *Meyer, P. L. Probabilidade: aplicações à estatística. Editora LTC, 1983.*
- [Aula 8](#) de Métodos Estatísticos Básicos.

1.3 Probabilidade condicional e independência

Em muitos experimentos o espaço amostral pode mudar depois que obtemos nova informação. Se nosso experimento for retirar duas cartas de um baralho e estamos interessados no evento de obter dois Áses, ao retirar a primeira carta, o espaço amostral é reduzido, fazendo com que a probabilidade de se obter um Ás na segunda carta seja alterada. Nesse tipo de problema, precisamos utilizar o que chamamos de *probabilidade condicional*.

Definição 1.3.1. Se A e B são eventos em S , e $P(B) > 0$, então a *probabilidade condicional* de A dado B é:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Note que para calcularmos $P(A|B)$ devemos reduzir o espaço amostral S para B , e então calcularmos a probabilidade de A . No caso de eventos disjuntos, em que $P(A \cap B) = 0$, temos $P(A|B) = P(B|A) = 0$.

Como a equação da probabilidade condicional é simétrica, podemos derivar dela o seguinte resultado.

Teorema 1.3.2. (*Regra de Bayes*)

Seja A_1, A_2, \dots uma partição do espaço amostral, e B qualquer outro conjunto. Então, para cada $i = 1, 2, \dots$,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}.$$

Note que A_1, \dots, A_k é uma partição do espaço amostral S quando:

- i) $A_i \cap A_j = \emptyset$ para todo $i \neq j$;
- ii) $\cup_{i=1}^k A_i = S$;

iii) $P(A_i) > 0$ para todo i .

Nesse caso, quando o experimento é realizado, um e somente um dos eventos A_i ocorre. Logo, podemos deduzir que vale a *Lei das probabilidades totais*:

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)$$

A regra de Bayes utiliza este resultado no denominador.

Em alguns casos, a ocorrência do evento B não influencia a probabilidade do evento A ocorrer, de modo que $P(A|B) = P(A)$. Assim, pela regra de Bayes, $P(B|A) = P(B)$, e então, pela fórmula da probabilidade condicional, $P(A \cap B) = P(A)P(B)$. Nesse caso, dizemos que A e B são *eventos estatisticamente independentes*.

Teorema 1.3.3. *Se A e B são eventos independentes, então os seguintes eventos também são independentes:*

- a. A e B^C ,
- b. A^C e B ,
- c. A^C e B^C .

Demonstração. Feita em aula. □

Para mais de dois conjuntos devemos estender nossa definição de independência.

Definição 1.3.4. Uma coleção de eventos A_1, \dots, A_n é mutuamente independente se para qualquer subcoleção A_{i_1}, \dots, A_{i_k} , temos

$$P(\cap_{j=1}^k A_{i_j}) = \prod_{j=1}^k P(A_{i_j}).$$

Exemplo 1.3.5. (Probabilidade Condicional e Independência)

- i) **Probabilidade condicional:** ver exemplos 1.3.3 e 1.3.4 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001* e páginas 4 e 5 da [Aula 7](#) de Métodos Estatísticos Básicos.
- ii) **Regra de Bayes:** ver exemplo 1.3.6 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001* e página 8 da [Aula 7](#) de Métodos Estatísticos Básicos.
- iii) **Independência:** ver exemplos 1.3.8, 1.3.10, 1.3.11 e 1.3.13 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001* e página 12 da [Aula 7](#) de Métodos Estatísticos Básicos.

1.4 Exemplo de aplicação no R

O software estatístico [R](#) possui um pacote chamado [prob](#) em que podemos realizar aplicações de quase todos os conceitos aprendidos até aqui. Se você ainda não utilizou o R, um bom meio de começar é através do curso introdutório da [DataCamp](#). Você pode encontrar uma lista de cursos, documentação, livros, entre outros [neste post](#). Para ilustrar a aplicação no R considere

o seguinte exemplo.

Exemplo 1.4.1. (Aplicação das seções 1.1, 1.2 e 1.3 no R)

Suponha que um ex-presidente da república tenha ido ao cassino para tentar a sorte com o dinheiro de sua “aposentadoria”. Ele resolve jogar R\$100.000,00 em uma roleta americana, apostando R\$90.000,00 na cor vermelha e R\$10.000,00 no número 13. Sabendo que o pagamento se ele acertar o número é 35:1 e se acertar a cor é 1:1, resolva os itens abaixo.

- i) Defina o espaço amostral S .
- ii) Defina os eventos A em que ele acerta o número, e B em que ele acerta a cor.
- iii) Defina o evento E em que sai um número ímpar.
- iv) Defina os eventos $U = A \cup B$, $G = B \cap C$, $F = C - A$ e $H = (A \cup B)^C$.
- v) Qual a probabilidade de ele perder todo seu dinheiro?
- vi) Qual a probabilidade de ele dobrar seu dinheiro?
- vii) Qual a probabilidade de ele acertar o 13?
- viii) Qual é o valor esperado desta aposta?
- ix) Sabendo que um número ímpar saiu na roleta, qual a probabilidade de ele ter ganho algo?
- x) **Desafio:** Você é capaz de encontrar uma combinação que dê valor esperado positivo nessa roleta? Qual?

Solução:

```
1 #install.packages("prob")
2 library(prob)
3
4 # Item 1
5 S <- roulette(makespace = TRUE)
6
7 # Item 2
8 A <- subset(S, num == 13)
9 B <- subset(S, color == "Red")
10
11 # Item 3
12 numbers <- as.numeric(as.character(S[,1]))
13 E <- S[numbers %% 2 != 0,]
14
15 # Item 4
16 U <- union(A,B)
17 G <- intersect(B,C)
18 F <- setdiff(C,A)
19 H <- setdiff(S,U)
20
21 # Item 5, 6 e 7
22 Prob(H)
23 Prob(B)
24 Prob(A)
25
26 # Item 8
```



```

27 VE <- (90000.00 * 2 * Prob(B)) + (10000.00 * 36 * Prob(A)) - 100000.00
28
29 # Item 9
30 Prob(U, given = E)

```

1.5 Variáveis aleatórias

Muitas vezes estamos interessados em uma transformação do nosso espaço amostral. Se o nosso experimento for lançar duas moedas e observar as faces, sabemos que nosso espaço amostral será $S = \{(H, H), (H, T), (T, H), (T, T)\}$, mas se estivermos interessados no número de caras obtidas, podemos aplicar uma função $X(s)$ que nos retorne o número de caras para cada elemento do espaço amostral S . Então obteremos um novo espaço amostral $\mathcal{X} = \{0, 1, 2\}$.

Definição 1.5.1. Uma *variável aleatória* é uma função do espaço amostral S para os números reais.

A partir da função de probabilidade original, que tem como domínio eventos em S , podemos obter uma função de probabilidade induzida no novo espaço amostral X :

$$P_X(X = x_i) = P(\{s_j \in S : X(s_j) = x_i\})$$

A ideia fundamental por trás das variáveis aleatórias é criar funções mapeiam resultados de experimentos em números reais, de forma a facilitar o cálculo de probabilidades. Note que a partir de agora nos referimos a X como sendo a variável aleatória e x como sendo um valor específico da mesma.

Exemplo 1.5.2. (Variáveis aleatórias)

- i) No experimento de jogar duas moedas, X = número de caras obtidas;
- ii) No experimento de observar o lucro de uma empresa, X = margem do lucro, ou seja, lucro/vendas;

As variáveis aleatórias podem ser *discretas* ou *contínuas*.

Definição 1.5.3. (Variável aleatória discreta) Uma variável aleatória é discreta se toma um número finito ou enumerável de valores, isto é, se existe um subconjunto finito ou enumerável $\{x_1, x_2, \dots\} \in \mathbb{R}$ tal que $X(s) \in \{x_1, x_2, \dots\}$ para qualquer $s \in S$.

Definição 1.5.4. (Variável aleatória absolutamente contínua) Uma variável aleatória é absolutamente contínua se existe uma função $f : \mathbb{R} \rightarrow \mathbb{R}$ com $f(x) \geq 0$, tal que:

$$P(X \leq x) = \int_{-\infty}^x f(t)dx \quad \forall x \in \mathbb{R}.$$

1.6 Funções de distribuição

Podemos associar uma função de distribuição cumulativa a cada variável aleatória.

Definição 1.6.1. (Função de distribuição cumulativa) A função de distribuição cumulativa (cdf) de uma variável aleatória X , denotada $F_X(x)$, é definida por:

$$F_X(x) = P_X(X \leq x), \quad \forall x$$

Quando a variável aleatória X tem uma distribuição dada por $F_X(x)$ denotamos $X \sim F_X(x)$. Note que a cdf é definida em termos de probabilidades, logo, ela deverá respeitar algumas propriedades condizentes com os axiomas de probabilidade.

Teorema 1.6.2. *A função $F(x)$ é uma cdf se e somente se as seguintes condições são verdadeiras:*

- a. $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$.
- b. $F(x)$ é uma função não decrescente em x .
- c. $F(x)$ é contínua à direita, ou seja, para cada número x_0 , $\lim_{x \downarrow x_0} F(x) = F(x_0)$.

Demonstração. Feita em aula. □

O fato da cdf ser contínua ou possuir saltos está relacionado ao fato de a variável aleatória ser contínua ou discreta.

Definição 1.6.3. Uma variável aleatória X é contínua se $F_X(x)$ é uma função contínua de x . Uma variável X é discreta se $F_X(x)$ possui descontinuidades.

Se utilizarmos a menor sigma álgebra contendo todos os intervalos possíveis dos números reais, então se duas variáveis aleatórias tiverem a mesma cdf, elas terão as mesmas probabilidades para cada evento.

Definição 1.6.4. Duas variáveis aleatórias X e Y são identicamente distribuídas se, para cada conjunto $A \in \mathcal{B}$, $P(X \in A) = P(Y \in A)$.

Teorema 1.6.5. *As seguintes afirmações são equivalentes:*

- a. As variáveis aleatórias X e Y são identicamente distribuídas.
- b. $F_X(x) = F_Y(x)$ para cada x .

Demonstração. Feita em aula. □

1.7 Funções de densidade e massa

A cdf avalia $P(X \leq x)$, o que faz bastante sentido para variáveis aleatórias contínuas, mas se quisermos avaliar $P(X = x)$ então temos a *função densidade de probabilidade* (fdp) no caso contínuo, e a *função massa de probabilidade* no caso discreto.

Definição 1.7.1. A função massa de probabilidade (pmf) de uma variável aleatória discreta é dada por:

$$f_X(x) = P(X = x) \forall x$$

Definição 1.7.2. A função densidade de probabilidade (fdp) de uma variável aleatória discreta é a função $f_X(x)$ que satisfaz:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \forall x$$

Note que de acordo com a definição de fdp, também vale que:

$$\frac{d}{dx} F_X(x) = f_X(x)$$

Teorema 1.7.3. Uma função $f_X(x)$ é uma pdf ou pmf de uma variável aleatória X se e somente se:

- a. $f_X(x) \geq 0 \forall x$.
- b. $\sum_x f_X(x) = 1$ (pmf)
- c. $\int_{-\infty}^{\infty} f_X(x)dx = 1$ (pdf)

Demonstração. Feita em aula. □

Exemplo 1.7.4. (Funções de distribuição)

- i) Ver Exemplos 1.5.2, 1.5.4, 1.5.5, 1.5.6, 1.5.9, 1.6.2 e 1.6.4 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

2 Transformações em variáveis aleatórias

2.1 Distribuições de funções de variáveis aleatórias

Se X for uma variável aleatória com cdf $F_X(x)$, então qualquer função $g(X) = Y$ também será. Assim:

$$P(Y \in A) = P(g(X) \in A),$$

Note que $g(x)$ mapeia elementos no espaço amostral de X para elementos no novo espaço amostral de Y , $g(x) : \mathcal{X} \rightarrow \mathcal{Y}$. Associada a $g(x)$ também está a sua inversa, g^{-1} , denotada por:

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\}.$$

Voltando a noção de probabilidade, podemos ver que:

$$\begin{aligned} P(Y \in A) &= P(g(X) \in A) \\ &= P(\{x \in \mathcal{X} : g(x) \in A\}) \\ &= P(X \in g^{-1}(A)) \end{aligned}$$

É fácil checar que essas probabilidades satisfazem os axiomas de Kolmogorov.

Se X for uma variável aleatória discreta, então Y também será, e a pmf de Y poderá ser escrita como:

$$f_Y(y) = P(Y = y) = \sum_{x \in g^{-1}(y)} P(X = x) = \sum_{x \in g^{-1}(y)} f_X(x), \forall y \in \mathcal{Y}.$$

Note que no caso acima, $f_Y(y) = 0$ para $y \notin \mathcal{Y}$.

Exemplo 2.1.1. (Transformação da distribuição binomial)

- i) Ver Exemplo 2.1.1 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

Se X for uma variável aleatória contínua, então Y também será, e a cdf de Y poderá ser escrita como:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(\{x \in \mathcal{X} : g(x) \leq y\}) \\ &= \int_{x \in \mathcal{X} : g(x) \leq y} f_X(x) dx \end{aligned}$$

Exemplo 2.1.2. (Transformação da distribuição uniforme)

- i) Ver Exemplo 2.1.2 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

Teorema 2.1.3. Dada uma variável aleatória X com cdf $F_X(x)$, seja $Y = g(X)$, $\mathcal{X} = \{x : f_X(x) > 0\}$ e $\mathcal{Y} = \{y : y = g(x) \text{ para algum } x \in \mathcal{X}\}$, então:

- Se g é uma função crescente em \mathcal{X} , $F_Y(y) = F_X(g^{-1}(y))$ para $y \in \mathcal{Y}$.
- Se g é uma função decrescente em \mathcal{X} , e X é uma variável aleatória contínua, então $F_Y(y) = 1 - F_X(g^{-1}(y))$ para $y \in \mathcal{Y}$.

A partir deste teorema, podemos caracterizar também a pdf de Y .

Teorema 2.1.4. Seja X uma v.a. com pdf $f_X(x)$ e $Y = g(X)$, onde g é uma função monótona. E \mathcal{X} e \mathcal{Y} definidas como no teorema acima. Suponha que $f_X(x)$ é contínua em \mathcal{X} e que $g^{-1}(y)$ tem uma derivada contínua em \mathcal{Y} . Então a pdf de Y pode ser dada por:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{caso contrário} \end{cases}$$

Demonstração. Feita em aula. □

Note que este resultado se aplica apenas para funções monótonas, porém podemos estendê-lo para partes monótonas de funções que não são propriamente monótonas.

Teorema 2.1.5. Seja X uma v.a. com pdf $f_X(x)$ e $Y = g(X)$, onde g é uma função qualquer. E \mathcal{X} e \mathcal{Y} definidas como no teorema anterior. Suponha que existe uma partição A_0, A_1, \dots, A_k de \mathcal{X} tal que $P(X \in A_0) = 0$ e $f_X(x)$ é contínua em cada A_i . Ainda, suponha que exista funções $g_1(x), \dots, g_k(x)$, definidas em A_1, \dots, A_k respectivamente, que satisfazem:

- $g(x) = g_i(x)$, para $x \in A_i$,
- $g_i(x)$ é monótona em A_i ,
- o conjunto $\mathcal{Y} = \{y : y = g_i(x) \text{ para algum } x \in A_i\}$ é o mesmo para cada $i = 1, \dots, k$, e
- $g_i^{-1}(y)$ tem uma derivada contínua em \mathcal{Y} , para cada $i = 1, \dots, k$.

Então:

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{caso contrário} \end{cases}$$

Exemplo 2.1.6. (Distribuições gama invertida, não monótonas e qui-quadrada)

- i) Ver Exemplo 2.1.6 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*
- ii) Ver Exemplo 2.1.7 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*
- iii) Ver Exemplo 2.1.9 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

2.2 Valor esperado

Definição 2.2.1. (Valor esperado)

O valor esperado ou média de um variável aleatória $g(X)$, denotado $Eg(x)$, é dado por:

$$Eg(X) = \begin{cases} \int_{-\infty}^{\infty} g(x)f_X(x)dx & \text{se } X \text{ é contínua} \\ \sum_{x \in \mathcal{X}} g(x)f_X(x) = \sum_{x \in \mathcal{X}} g(x)P(X = x) & \text{se } X \text{ é discreta,} \end{cases}$$

desde que a integral ou a soma existam. Se $E|g(X)| = \infty$ dizemos que $Eg(X)$ não existe.

Note que o operador de expectativas é um operador linear, e por isso ele guarda as seguintes propriedades:

Teorema 2.2.2. *Seja X uma v.a. e a, b, c constantes. Então, para quaisquer funções $g_1(x)$ e $g_2(x)$ que possuem valores esperados, vale que:*

- a. $E(ag_1(x) + bg_2(X) + c) = aEg_1(x) + bEg_2(x) + c.$
- b. *Se $g_1(x) \geq 0$ para todo x , então $Eg_1(x) \geq 0.$*
- c. *Se $g_1(x) \geq g_2(x)$ para todo x , então $Eg_1(X) \geq Eg_2(X).$*
- d. *Se $a \leq g_1(x) \leq b$ para todo x , então $a \leq Eg_1(X) \leq b.$*

Demonstração. Feita em aula. □

Exemplo 2.2.3. (Valor esperado)

- i) Ver Exemplos 2.2.2, 2.2.3, 2.2.4, 2.2.6 e 2.2.7 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

2.3 Variância e outros momentos

Definição 2.3.1. (Momentos)

O n -ésimo momento da variável aleatória X , μ'_n é dado por:

$$\mu'_n = EX^n,$$

enquanto que o n -ésimo momento central de X é:

$$\mu_n = E(X - \mu)^n,$$

onde $\mu = \mu'_1 = EX$.

Depois da média, o momento mais importante de uma distribuição é o segundo momento central, conhecido como variância.

Definição 2.3.2. (Variância)

A variância de uma variável aleatória X é o segundo momento central, $VarX = E(X - EX)^2$. A raiz quadrada positiva da variância é conhecida como o desvio-padrão.

Teorema 2.3.3. *Se X é uma variável aleatória com variância finita, então para quaisquer constantes a e b ,*

$$Var(aX + b) = a^2 VarX$$

Demonstração. Feita em aula. □

Exemplo 2.3.4. (Variância)

- i) Ver Exemplos 2.3.3 e 2.3.5 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

2.4 Função geradora de momentos

Definição 2.4.1. (Função geradora de momentos)

Seja X uma variável aleatória com cdf F_X . A função geradora de momentos (mgf) de X , denotada $M_X(t)$, é dada por:

$$M_X(t) = Ee^{tX},$$

desde que este valor esperado exista para t em alguma vizinhança de 0.

Note que pela definição de valor esperado, temos que:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \text{ se } X \text{ é contínua,}$$
$$M_X(t) = \sum_x e_{tx} P(X = x) \text{ se } X \text{ é discreta}$$

Para entender como esta função gera momentos, observe o seguinte resultado:

Teorema 2.4.2. Se X tem mgf $M_X(t)$, então:

$$EX^n = M_X^{(n)}(0) = \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0}.$$

Ou seja, o n -ésimo momento é igual a n -ésima derivada de $M_X(t)$ avaliada em $t = 0$.

Demonstração. Feita em aula. □

2.5 Exemplo no R com estatísticas descritivas

3 Distribuições de probabilidade

A referência principal para esta parte é o capítulo 3, seções 3.1, 3.2, 3.3 e teorema 3.6.1 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

3.1 Distribuições discretas

Uma distribuição X é discreta se o conjunto X for enumerável. Vimos as seguintes distribuições discretas, além de seus valores esperados, variância, e em alguns casos a função geradora de momentos:

- i) Uniforme discreta
- ii) Hipergeométrica
- iii) Binomial e Bernoulli
- iv) Poisson
- v) Negativa binomial
- vi) Geométrica

Vimos também algumas relações entre variáveis aleatórias discretas. Uma delas foi o uso da Poisson como uma aproximação da binomial. A outra foi o fato de que a distribuição binomial negativa inclui a distribuição de Poisson como um caso limite.

3.2 Distribuições contínuas

Uma distribuição X é contínua se o conjunto X for infinito não-enumerável. Vimos as seguintes distribuições contínuas, além de seus valores esperados, variância, e em alguns casos a função geradora de momentos:

- i) Uniforme
- ii) Gama
- iii) Qui-quadrada
- iv) Exponencial
- v) Weibull
- vi) Normal
- vii) Lognormal

Também vimos algumas relações entre distribuições de probabilidade contínuas, como por exemplo o fato das distribuições qui-quadrada e exponencial serem casos especiais da distribuição gama.

Fechamos esta parte com a Desigualdade de Chebychev e alguns resultados associados a ela.

3.3 Exemplos de distribuições de probabilidade no R

4 Variáveis aleatórias múltiplas

- Distribuições conjuntas e marginais
- Distribuição condicional e independência
- Mistura de distribuições
- Esperança condicional
- Lei das expectativas totais e iteradas
- Variância condicional, decomposição de variância
- Covariância, correlação
- Previsão
- Exemplo no R com independência, estatísticas condicionais, covariâncias, correlação e previsão

5 Propriedades de amostras aleatórias