

Curso de Verão em Estatística

Teoria da Probabilidade

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

1 de março de 2016

Resumo

Estas são as notas de 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
3	Distribuições de probabilidade	15
3.1	Distribuições discretas	15
3.2	Distribuições contínuas	15
4	Variáveis aleatórias múltiplas	16
4.1	Distribuições conjuntas e marginais	16
4.2	Distribuição condicional e independência	17
4.3	Mistura de distribuições	19
4.4	Covariância e correlação	19

5	Propriedades de amostras aleatórias	20
5.1	Amostras aleatórias	20
5.2	Estatísticas	21
5.3	Amostragem da distribuição normal	22
5.4	Conceitos de convergência	23
6	Exemplos no R	24

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* (pdf) 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 (pdf) 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 pdf, 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. □

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.

4 Variáveis aleatórias múltiplas

4.1 Distribuições conjuntas e marginais

Uma variável aleatória é uma função do espaço amostral S para os números reais. Um vetor aleatório consiste em várias variáveis aleatórias.

Definição 4.1.1. (Vetor aleatório)

Um vetor aleatório n -dimensional é uma função do espaço amostral S para o \mathbb{R}^n .

Um vetor aleatório bidimensional é denotado (X, Y) , e pode ser discreto ou contínuo.

Definição 4.1.2. (Vetor aleatório bidimensional discreto)

Seja (X, Y) um vetor aleatório bidimensional discreto. Então a função $f_{X,Y}(x, y)$, do \mathbb{R}^2 no \mathbb{R} , definida por $f_{X,Y}(x, y) = P(X = x, Y = y)$ é chamada de função massa de probabilidade conjunta ou pmf conjunta de (X, Y) .

Note que esta função define completamente a distribuição de probabilidade do vetor aleatório (X, Y) . Dessa forma, podemos usá-la para calcular a probabilidade de qualquer evento definido em termos de (X, Y) , ou seja, para qualquer conjunto $A \subset \mathbb{R}^2$, temos:

$$P((X, Y) \in A) = \sum_{(x,y) \in A} f(x, y)$$

Seja $g(x, y)$ uma função definida em todos os valores (x, y) do vetor aleatório discreto (X, Y) , então $g(X, Y)$ também será um vetor aleatório discreto, e seu valor esperado será dado por:

$$Eg(X, Y) = \sum_{(x,y) \in \mathbb{R}^2} g(x, y) f(x, y)$$

Se $g_1(x, y)$ e $g_2(x, y)$ são funções e a, b, c constantes, então:

$$E(ag_1(X, Y) + bg_2(X, Y) + c) = aEg_1(X, Y) + bEg_2(X, Y) + c$$

O vetor bivariado discreto (X, Y) com pdf conjunta $f(x, y)$ deve ter certas propriedades:

- i) Para cada (x, y) , $f(x, y) \geq 0$.
- ii) $\sum_{(x,y) \in \mathbb{R}^2} f(x, y) = P((X, Y) \in \mathbb{R}^2) = 1$.

Teorema 4.1.3. (Função massa de probabilidade marginal)

Seja (X, Y) um vetor aleatório discreto com pmf conjunta $f_{X,Y}(x, y)$. Então as pmfs marginais de X e Y , $f_X(x) = P(X = x)$ e $f_Y(y) = P(Y = y)$, são dadas por:

$$f_X(x) = \sum_{y \in \mathcal{R}} f_{X,Y}(x, y)$$
$$f_Y(y) = \sum_{x \in \mathcal{R}} f_{X,Y}(x, y).$$

Demonstração. Feita em aula. □

Note que as distribuições marginais de X e Y não determinam univocamente a função de distribuição conjunta de (X, Y) , pois podem haver várias distribuições conjuntas com as mesmas distribuições marginais.

Definição 4.1.4. (Vetor aleatório bidimensional contínuo)

Uma função $f(x, y)$ do \mathcal{R}^2 para o \mathcal{R} é chamada de função densidade de probabilidade conjunta do vetor aleatório bidimensional contínuo (X, Y) se, para cada $A \subset \mathcal{R}^2$,

$$P((X, Y) \in A) = \int_A \int f(x, y) dx dy$$

Se $g(x, y)$ for uma função real, então o valor esperado de $g(X, Y)$ é definido como:

$$Eg(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$$

As pdfs marginais de X e Y são dadas por:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx, \quad -\infty < y < \infty$$

Note que para $f(x, y)$ ser uma pdf conjunta de um vetor aleatório bidimensional contínuo (X, Y) , ela deve satisfazer:

- a. $f(x, y) \geq 0$ para todo $(x, y) \in \mathcal{R}^2$;
- b. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

Ao invés das pdf conjunta, podemos caracterizar as probabilidades através da cdf conjuntas, que será dada por:

$$F(x, y) = P(X \leq x, Y \leq y) \quad \forall (x, y) \in \mathcal{R}^2$$

Para o caso contínuo, essa função será dada por:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds,$$

de modo que $\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$.

Exemplo 4.1.5. (Distribuições conjuntas e marginais)

- i) Ver Exemplos 4.1.2, 4.1.4, 4.1.5, 4.1.7, 4.1.9, 4.1.11 e 4.1.12 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

4.2 Distribuição condicional e independência

Definição 4.2.1. (Pmf e pdf condicional)

Seja (X, Y) um vetor aleatório bivariado discreto ou contínuo com pmf ou pdf conjunta $f(x, y)$ e marginais $f_X(x)$ e $f_Y(y)$. Para qualquer x tal que $P(X = x) = f_X(x) > 0$, a pmf ou pdf condicional de Y dado $X = x$ é a função de y denotada por $f(y|x)$ e definida por:

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}.$$

A pmf ou pdf condicional de X é definida de maneira análoga. Note que $f(y|x) \geq 0$ para todo y pois $f(x, y) \geq 0$ e $f_X(x) > 0$. Também note que, no caso discreto:

$$\sum_y f(y|x) = \frac{\sum_y f(x, y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$$

No caso contínuo essa igualdade também é válida, basta substituir a soma por uma integral. Se $g(Y)$ for uma função de Y , então o valor esperado condicional de $g(Y)$ dado $X = x$ é dado por:

$$E(g(Y)|x) = \sum_y g(y)f(y|x)$$

$$E(g(Y)|x) = \int_{-\infty}^{\infty} g(y)f(y|x)dy$$

Esses valores esperados herdam todas as propriedades do operador de expectativas. Note que $E(g(Y)|X)$ é uma variável aleatória que depende dos valores de X , enquanto que $E(g(Y))$ é apenas uma constante.

Definição 4.2.2. (Variáveis aleatórias independentes)

Seja (X, Y) um vetor aleatório bivariado com pdf ou pmf conjunta $f(x, y)$ e marginais $f_X(x)$ e $f_Y(y)$. Então X e Y são ditas variáveis aleatórias independentes se, para cada $x \in \mathcal{R}$ e $y \in \mathcal{R}$,

$$f(x, y) = f_X(x)f_Y(y).$$

Essa definição de independência pode ser aplicada para qualquer função das variáveis aleatórias, inclusive para o valor esperado. Note que se X e Y são independentes, então:

$$f(y|x) = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y)$$

Assim, o conhecimento de X não nos dá nenhuma informação adicional sobre Y . Uma maneira rápida de verificar a independência de duas variáveis aleatórias é através do seguinte teorema:

Teorema 4.2.3. *Seja (X, Y) um vetor aleatório bivariado com pdf ou pmf conjunta $f(x, y)$. Então X e Y são variáveis aleatórias independentes se e somente se existe funções $g(x)$ e $h(y)$ tais que, para cada $x \in \mathcal{R}$ e $y \in \mathcal{R}$,*

$$f(x, y) = g(x)h(y)$$

Teorema 4.2.4. *Seja X e Y variáveis aleatórias independentes com funções geradoras de momento $M_X(t)$ e $M_Y(t)$. Então a função geradora de momento da variável aleatória $Z = X + Y$ é dada por:*

$$M_Z(t) = M_X(t)M_Y(t)$$

Demonstração. $M_Z(t) = E e^{t(X+Y)} = E(e^{tX} e^{tY}) = (E e^{tX})(E e^{tY}) = M_X(t) M_Y(t)$ □

Teorema 4.2.5. *Sejam $X \sim N(\mu, \sigma^2)$ e $Y \sim N(\gamma, \tau^2)$ variáveis aleatórias independentes e normalmente distribuídas. Então a variável aleatória $Z = X + Y$ tem uma distribuição $N(\mu + \gamma, \sigma^2 + \tau^2)$.*

Exemplo 4.2.6. (Distribuição condicional e independência)

- i) Ver Exemplos 4.2.2, 4.2.4, 4.2.6, 4.2.8, 4.2.9, 4.2.11 e 4.2.13 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

4.3 Mistura de distribuições

Em geral uma variável aleatória tem apenas uma distribuição, porém em alguns casos temos variáveis aleatórias compostas por duas distribuições diferentes. Ao invés de elaborar uma função de distribuição complicada para modelar estes casos, podemos pensar o experimento em termos de hierarquia, o que nos leva a uma mistura de distribuições.

Definição 4.3.1. (Mistura de distribuições)

Uma variável aleatória X tem uma distribuição mista se X depende de um valor que também tem uma distribuição de probabilidade.

Um importante resultado associado a mistura de distribuições é a *lei das expectativas totais*.

Teorema 4.3.2. (*Lei das expectativas totais*)

Se X e Y são quaisquer duas variáveis aleatórias cujos valores esperados existem, então:

$$EX = E(E(X|Y))$$

Demonstração. Feita em aula. □

Teorema 4.3.3. (*Identidade da variância condicional*)

Para quaisquer duas variáveis aleatórias X e Y que possuem valores esperados bem definidos:

$$VarX = E(Var(X|Y)) + Var(E(X|Y))$$

Demonstração. Feita em aula. □

Exemplo 4.3.4. (Mistura de distribuições)

- i) Ver Exemplos 4.4.1, 4.4.2, 4.4.5, 4.4.6, 4.4.8, 4.2.11 e 4.2.13 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

4.4 Covariância e correlação

Se duas variáveis aleatórias guardam uma relação de dependência entre elas, seria interessante calcular estatísticas que nos mostrem a magnitude dessa dependência. Para isso, temos as noções de covariância e correlação.

Definição 4.4.1. (Covariância)

A covariância de X e Y é um número definido por:

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = EXY - \mu_X\mu_Y$$

Definição 4.4.2. (Correlação)

A correlação de X e Y (ou coeficiente de correlação) é um número definido por:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

Note que estamos denotando $EX = \mu_X$, $EY = \mu_Y$, $VarX = \sigma_X^2$ e $VarY = \sigma_Y^2$. Os valores da covariância e correlação estarão bem definidos desde que $0 < \sigma_X^2 < \infty$ e $0 < \sigma_Y^2 < \infty$. As duas medidas variam conjuntamente, porém a correlação nos dá um valor padronizado entre -1 e 1 , podendo ser mais amplamente utilizada para se comparar relações entre conjuntos diferentes de variáveis aleatórias.

Teorema 4.4.3. *Se X e Y são variáveis aleatórias independentes, então $Cov(X, Y) = 0$ e $\rho_{XY} = 0$.*

Demonstração. Feita em aula. □

Note o a volta deste teorema não é necessariamente válida, ou seja, $Cov(X, Y) = 0$ não implica independência.

Teorema 4.4.4. *Se X e Y são quaisquer duas variáveis aleatórias e a e b são constantes, então:*

$$Var(aX + bY) = a^2 VarX + b^2 VarY + 2ab Cov(X, Y).$$

Se X e Y são variáveis aleatórias independentes, então:

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

Demonstração. Feita em aula. □

Teorema 4.4.5. *Para quaisquer variáveis aleatórias X e Y ,*

- a. $-1 \leq \rho_{XY} \leq 1$.
- b. $|\rho_{XY}| = 1$ se e somente se existe números $a \neq 0$ e b tais que $P(Y = aX + b) = 1$. Se $\rho_{XY} = 1$, então $a > 0$, e se $\rho_{XY} = -1$, então $a < 0$.

5 Propriedades de amostras aleatórias

5.1 Amostras aleatórias

Definição 5.1.1.

 (Amostra aleatória)

As variáveis aleatórias X_1, \dots, X_n são chamadas de amostra aleatória de tamanho n de uma população dada por $f(x)$ se X_1, \dots, X_n são variáveis aleatórias mutuamente independentes e a pmf ou pdf marginal de cada X_i é a mesma função $f(x)$. Dizemos então que X_1, \dots, X_n são variáveis aleatórias independente e identicamente distribuídas (iid) com pmf ou pdf $f(x)$.

Se X_1, \dots, X_n são iid, então a pdf ou pmf conjunta será dada por:

$$f(x_1, \dots, x_n) = f(x_1)f(x_2) \dots f(x_n) = \prod_{i=1}^n f(x_i).$$

Se a pdf ou pmf populacional são paramétricas utilizamos a notação $f(x|\theta)$, e então podemos utilizar diferentes valores de θ para observar como a amostra aleatória se encaixa em diferentes distribuições populacionais.

Um meio de se construir uma amostra aleatória a partir de uma população finita é escolher valores com probabilidades iguais ($1/N$) utilizando reposição. Assim, podemos garantir que a nossa amostra é aleatória, visto que o processo de amostragem é independente. Se fizermos o mesmo processo mas sem reposição, então as variáveis aleatórias X_1, \dots, X_n não serão mutuamente independentes, porém para N grande essa amostra se aproximará de uma amostra aleatória.

5.2 Estatísticas

Definição 5.2.1. (Estatística)

Seja X_1, \dots, X_n uma amostra aleatória de tamanho n de uma população, e $T(x_1, \dots, x_n)$ uma função real cujo resultado é um número ou um vetor e o domínio inclui o espaço amostral de (X_1, \dots, X_n) . Então a variável ou vetor aleatório $Y = T(X_1, \dots, X_n)$ é chamada de uma estatística. A distribuição de probabilidade da estatística Y é chamada de distribuição amostral de Y .

Note que uma estatística é obtida a partir de uma amostra aleatória e não pode ser uma função dos parâmetros de uma distribuição. Duas estatísticas mais comumente utilizadas é a média amostral e a variância amostral.

Definição 5.2.2. A média amostral é a média aritmética dos valores de uma amostra aleatória, e denotada por:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Definição 5.2.3. A variância amostral é uma estatística definida por:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

O desvio-padrão amostral é definido por $S = \sqrt{S^2}$.

Teorema 5.2.4. *Seja x_1, \dots, x_n quaisquer números e $\bar{x} = (x_1 + \dots + x_n)/n$. Então:*

- a. $\min_a \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2$,
- b. $(n-1)s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$.

Demonstração. Feita em aula. □

Teorema 5.2.5. *Seja X_1, \dots, X_n uma amostra aleatória de uma população com média μ e variância $\sigma^2 < \infty$. Então:*

- a. $E\bar{X} = \mu$,
- b. $\text{Var}\bar{X} = \frac{\sigma^2}{n}$,
- c. $ES^2 = \sigma^2$.

Demonstração. Feita em aula. □

Os resultados dos itens *a* e *c* são exemplos de estatísticas não viesadas, ou seja, a estatística \bar{X} é um estimador não viesado de μ , e a estatística S^2 é um estimador não viesado de σ^2 . Note o uso de $n-1$ na variância amostral.

Teorema 5.2.6. *Seja X_1, \dots, X_n uma amostra aleatória de uma população com mgf $M_X(t)$. Então a mgf da média amostral é:*

$$M_{\bar{X}}(t) = [M_X(t/n)]^n.$$

5.3 Amostragem da distribuição normal

Teorema 5.3.1. *Seja X_1, \dots, X_n uma amostra aleatória de uma distribuição $N(\mu, \sigma^2)$, $\bar{X} = (1/n) \sum_{i=1}^n X_i$ e $S^2 = [1/(n-1)] \sum_{i=1}^n (X_i - \bar{X})^2$. Então:*

- a. \bar{X} e S^2 são variáveis aleatórias independentes,
- b. \bar{X} tem uma distribuição $N(\mu, \sigma^2/n)$,
- c. $(n-1)S^2/\sigma^2$ tem uma distribuição qui-quadrada com $n-1$ graus de liberdade.

Se X_1, \dots, X_n for uma amostra aleatória de uma $N(\mu, \sigma^2)$, sabemos que a quantia $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ é distribuída como $N(0, 1)$. Porém, na maioria das vezes, σ não é conhecida, de modo que gostaríamos de saber a distribuição de $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ para fazer inferência sobre μ . W. S. Gosset (Student) observou que:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}},$$

de modo que o numerador é uma variável aleatória com distribuição $N(0, 1)$, e o denominador corresponde a uma distribuição $\sqrt{\chi_{n-1}^2/(n-1)}$, independente do numerador.

Definição 5.3.2. *Seja X_1, \dots, X_n uma amostra aleatória com distribuição $N(\mu, \sigma^2)$. A quantia $(\bar{X} - \mu)/(S/\sqrt{n})$ tem uma distribuição t de Student com $n-1$ graus de liberdade. Em geral, uma variável aleatória $T \sim t_p$ tem pdf dada por:*

$$f_T(t) = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{(p\pi)^{1/2}} \frac{1}{(1 + t^2/p)^{(p+1/2)}}, \quad -\infty < t < \infty$$

Se T_p é uma variável aleatória com distribuição t_p , então:

$$ET_p = 0, \text{ se } p > 1,$$

$$VarT_p = \frac{p}{p-2}, \text{ se } p > 2.$$

De maneira análoga, a distribuição F aparece ao retirarmos razões de variâncias.

Definição 5.3.3. *Seja X_1, \dots, X_n uma amostra aleatória de uma população $N(\mu_x, \sigma_x^2)$, e seja Y_1, \dots, Y_m uma amostra aleatória de uma população independente $N(\mu_y, \sigma_y^2)$. A variável aleatória $F = (S_X^2/\sigma_x^2)/(S_Y^2/\sigma_y^2)$ tem uma distribuição F com $n-1$ e $m-1$ graus de liberdade. Em geral, se $F \sim \mathcal{X}(p, q)$, então:*

$$f_F(x) = \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})} \left(\frac{p}{q}\right)^{p/2} \frac{x^{(p/2)-1}}{[1 + (p/q)x]^{(p+q)/2}}, \quad -\infty < x < \infty$$

Se $X \sim t_q$, então $X^2 \sim F_{1,q}$. Note também que se $F \sim F_{n-1,m-1}$, então:

$$EF_{n-1,m-1} = \frac{m-1}{m-3}$$

5.4 Conceitos de convergência

Um dos artifícios mais utilizados em inferência é permitir que o tamanho da nossa amostra aleatória se aproxime do infinito e então investigar o comportamento de estatísticas que estamos interessados. Um dos conceitos mais fracos de convergência é a ideia de *convergência em probabilidade*.

Definição 5.4.1. (Convergência em probabilidade)

Uma sequência de variáveis aleatórias X_1, X_2, \dots , converge em probabilidade para a variável aleatória X se, para cada $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0,$$

ou equivalentemente,

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \epsilon) = 1.$$

Frequentemente estamos interessados em situações em que a variável aleatória limite é uma constante e a sequência de variáveis aleatórias são médias amostrais. Nesse caso, podemos utilizar a *lei fraca dos grandes números*.

Teorema 5.4.2. (*Lei fraca dos grandes números*)

Sejam X_1, X_2, \dots variáveis aleatórias iid com $EX_i = \mu$ e $VarX_i = \sigma^2 < \infty$. Defina $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$. Então, para cada $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1,$$

ou seja, \bar{X}_n converge em probabilidade para μ .

Demonstração. Feita em aula. □

Assim, a lei fraca dos grandes números nos diz que sob algumas condições, a média amostral se aproxima da média populacional quando $n \rightarrow \infty$. Essa propriedade é chamada de consistência do estimador da média.

Teorema 5.4.3. Suponha que X_1, X_2, \dots converge em probabilidade para uma variável aleatória X e que h é uma função contínua. Então $h(X_1), h(X_2), \dots$ converge em probabilidade para $h(X)$.

Definição 5.4.4. (Convergência quase certa)

Uma sequência de variáveis aleatórias X_1, X_2, \dots , converge quase certamente para uma variável aleatória X se, para cada $\epsilon > 0$,

$$P\left(\lim_{n \rightarrow \infty} |X_n - X| < \epsilon\right) = 1.$$

Teorema 5.4.5. (*Lei forte dos grandes números*)

Sejam X_1, X_2, \dots variáveis aleatórias iid com $EX_i = \mu$ e $VarX_i = \sigma^2 < \infty$. Defina $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$. Então, para cada $\epsilon > 0$,

$$P\left(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| < \epsilon\right) = 1,$$

ou seja, \bar{X}_n converge quase certamente para μ .

Definição 5.4.6. (Convergência em distribuição)

Uma sequência de variáveis aleatórias X_1, X_2, \dots converge em distribuição para uma variável aleatória X se:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

para todos pontos x onde $F_X(x)$ é contínua.

Teorema 5.4.7. *Se uma sequência de variáveis aleatórias X_1, X_2, \dots converge em probabilidade para uma variável aleatória X , a sequência também converge em distribuição para a variável aleatória X .*

Uma estatística cujo comportamento de amostras grandes é importante é a média amostral. Ao investigar a sua distribuição limite chegamos a duas versões do chamado *Teorema do Limite Central*.

Teorema 5.4.8. (Teorema do Limite Central)

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias iid com mgfs existentes na vizinhança de 0. Seja $EX_i = \mu$ e $VarX_i = \sigma^2 > 0$ (finitas pela existência da mgf). Defina $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$. Seja $G_n(x)$ a cdf de $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Então, para qualquer x , tal que $-\infty < x < \infty$,

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy,$$

ou seja, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ tem uma distribuição limite igual a normal padronizada.

Teorema 5.4.9. (Forma Forte do Teorema do Limite Central)

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias iid com $EX_i = \mu$ e $0 < VarX_i = \sigma^2 < \infty$. Defina $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$. Seja $G_n(x)$ a cdf de $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Então, para qualquer x , tal que $-\infty < x < \infty$,

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy,$$

ou seja, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ tem uma distribuição limite igual a normal padronizada.

Exemplo 5.4.10. (Amostra Aleatórias)

- i) Ver Exemplos 5.1.2, 5.1.3, 5.1.8, 5.3.5, 5.3.7, 5.5.3 e 5.5.16 de *Casella, G. e Berger, R. L. Statistical Inference. 2nd Edition. Duxbury Press, 2001.*

6 Exemplos no R

```
1 x = rnorm(100)
2
3 # Central measures
4 mean(x)
5 median(x)
6 mean(x, trim = 0.05)
7 summary(x)
8
9 # Spread measures
```



```

10 range(x)
11 sd(x)
12 var(x)
13 mad(x)
14 IQR(x)
15
16 # Other moments
17 #install.packages("e1071")
18 library(e1071)
19 skewness(x)
20 kurtosis(x)
21
22 # Graphs of distribution
23 plot(x)
24 boxplot(x)
25 hist(x)
26
27 # Multivariate data
28 y = 3*x + 4 + rnorm(100)
29 plot(x,y)
30 cov(x,y)
31 cor(x,y)
32 reg <- lm(y ~ x)
33 summary(reg)
34
35 # Discrete random variables
36 dices <- sample(6, size = 10, replace = TRUE) #uniform - tossing 10
37 dices
38 dbinom(2, size = 4, prob = 1/2) #pdf binomial
39 pbinom(2, size = 4, prob = 1/2) #cdf binomial
40 dhyper(3, 100, 200, 10) #pdf hypergeometric - type help(dhyper) for help
41 phyper(3, 100, 200, 10) #cdf hypergeometric
42 dpois(4, 5/3)
43 ppois(4, 5/3)
44 dnbinom(2, 5, 1/2) #prob de 2 fracassos ate obter 5 sucessos
45 pnbinom(2, 5, 1/2)
46 dgeom(2, 1/2) #prob de obter 2 fracassos ate que o primeiro sucesso
47 pgeom(2, 1/2)
48
49 # Continuous random variables
50 help(dunif)
51 help(dgamma)
52 help(dchisq)
53 help(dexp)
54 help(dweibull)
55 help(dnorm)
56 help(dlnorm)

```