Universidade Estadual de Campinas
Instituto de Matemática, Estatística e
Computação Científica

Diego Fernando de Bernardini

Distribuições Subexponenciais
Introdução e Exemplos

Campinas-SP

2007

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica

Diego Fernando de Bernardini

Distribuições Subexponenciais
Introdução e Exemplos

Monografia apresentada ao Instituto de Matemática, Estatística e Computação Científica, sob orientação da Prof. Dra. Laura Leticia Ramos Rifo, na conclusão do Bacharelado em Estatística.

Campinas-SP
2007

Prefácio

Esta monografia tem como objetivo principal reportar os estudos/avanços obtidos durante o desenvolvimento do projeto de Iniciação Científica, assim como fornecer ao leitor informação básica na área, e uma bibliografia completa e atualizada.

No primeiro capítulo, entregamos uma revisão dos conceitos fundamentais em Teoria de Medida e Integração, Funções de Distribuição e Variáveis Aleatórias. Finalizamos o capítulo apresentando os principais resultados em Convergência de Seqüências de Variáveis Aleatórias e princípios de simulação estocástica.

No Capítulo 2, fazemos o estudo aprofundado da distribuição exata e da distribuição assintótica do máximo de uma seqüência de variáveis aleatórias independentes e identicamente distribuídas. São abordadas as distribuições max-estáveis e o Teorema de Fisher-Tippett.

O Capítulo 3 é dedicado ao estudo da convolução de funções de distribuição, de modo a caracterizar a distribuição exata da soma de variáveis aleatórias independentes e identicamente distribuídas e seu comportamento assintótico. Discutimos o Teorema Central do Limite e apresentamos a classe das distribuições estáveis.

No quarto capítulo, definimos a família das distribuições subexponenciais e estudamos algumas de suas propriedades. Apresentamos ainda a família de distribuições de variação regular, identificando as diversas relações entre estas duas famílias e as distribuições estudadas nos dois capítulos anteriores.

Finalmente, no Capítulo 5, apresentamos duas aplicações estudadas ao longo da Iniciação Científica. A primeira delas está relacionada à inferência para valores extremos em dados sobre níveis de maré em Sheerness, Inglaterra, entre 1819 e 1991, extraído do artigo de Stephenson and Tawn [25]. A segunda aplicação está dedicada ao estudo das probabilidades de ruína em processos de risco de seguros, definidos pelo processo estocástico a tempo contínuo de Cramér-Lundberg; também analisamos uma aproximação a tempo discreto construída em Tang and Tsitsiashvili [29].

Este projeto teve financiamento da FAPESP durante um ano e meio, e foi escrito sob a orientação da Professora Laura L. R. Rifo, do Instituto de Matemática e Estatística da UNICAMP.

Gostaria de agradecer ao referee pela paciência de ler o trabalho minuciosamente em suas três versões preliminares e pelas correções sugeridas. O trabalho foi sem dúvida enriquecido pelas suas observações.

Sumário

Capítulo 1

Introdução

Este capítulo tem como objetivo geral apresentar alguns conceitos julgados importantes e pertinentes ao tema do projeto e ainda fornecer ao leitor os pré-requisitos básicos para um bom entendimento da monografia como um todo.

Na Seção 1.1 definimos uma σ-álgebra e uma σ-álgebra de Borel, conceitos freqüentemente utilizados neste texto. Também definimos formalmente uma medida, possibilitando a utilização deste conceito mais adiante na monografia.

Em seguida, na Seção 1.2, estudamos a Medida de Lebesgue definida para um conjunto linear de pontos e ainda a Integral de Lebesgue para funções de uma variável.

Na Seção 1.3 apresentamos um par de funções (função de distribuição e medida de probabilidade) univocamente associadas, de modo a caracterizar uma Distribuição de Probabilidade no espaço unidimensional.

A Integral de Lebesgue-Stieltjes para funções de uma variável é apresentada na Seção 1.4. Esta integral nos permite, mais adiante, na Seção 1.6, associar uma medida de probabilidade a uma função de densidade.

Definimos variável aleatória na Seção 1.5, e apresentamos a medida de probabilidade e a função de distribuição associadas a esta variável.

Na Seção 1.6 estudamos o conceito de Continuidade Absoluta e apresentamos a derivada de Radon-Nikodym, nos permitindo definir a função de densidade de uma variável aleatória como a derivada de sua função de distribuição.

Introduzimos o conceito de Convergência Fraca na Seção 1.7, nos levando a definição de Convergência em Distribuição para uma seqüência de variáveis aleatórias. Além disso, discutimos mais dois tipos de convergência de variáveis aleatórias: Convergência em Probabilidade e Convergência Quase Certa.

Na Seção 1.8 apresentamos três teoremas que tratam da geração de números pseudo-aleatórios com uma certa distribuição, a partir de números pseudo-aleatórios gerados de uma distribuição uniforme no intervalo $[0, 1]$.

1.1 σ-Álgebra e Medidas

1.1.1 σ-Álgebra

Definição 1 *Uma classe $\mathcal{A}$ de subconjuntos de um dado conjunto S é dita uma σ-álgebra se satisfaz as três seguintes condições:*

a) $S \in \mathcal{A}$.

b) *Para uma seqüência* $\{A_n\}_{n=1}^{\infty}$*, se* $\{A_n\}_{n=1}^{\infty} \in \mathcal{A}$ *então* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ *e* $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

c) *Se* A_1*,* $A_2 \in \mathcal{A}$ *e* $A_2 \subset A_1$*, então* $A_1 \cap A_2^C \in \mathcal{A}$*, onde* A_2^C *é o complemento de* A_2 *com respeito a* S.

Equivalentemente às condições enunciadas na Definição 1, e portanto também definindo uma σ-álgebra $\mathcal{A}$ de subconjuntos de S, temos as seguintes condições:

a*) $S \in \mathcal{A}$.

b*) Para uma seqüência $\{A_n\}_{n=1}^{\infty}$, se $\{A_n\}_{n=1}^{\infty} \in \mathcal{A}$ então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

c*) Se $A \in \mathcal{A}$ então $A^C \in \mathcal{A}$, onde A^C é o complemento de A com respeito a S.

Prova. Mostremos que as condições a), b) e c) são equivalentes às condições a*), b*) e c*).

Primeiramente suponhamos a), b) e c) verdadeiras. Desta maneira

$$\text{a) } verdadeira \Rightarrow \text{a*) } verdadeira$$

$$\text{b) } verdadeira \Rightarrow \text{b*) } verdadeira$$

Se em c) tomarmos $A_1 = S$ e $A_2 = A$ então $S \cap A^C \in \mathcal{A}$ pois, de a), $S \in \mathcal{A}$. Mas $S \cap A^C = A^C$ e portanto c*) é verdadeira.

Agora suponhamos a*), b*) e c*) verdadeiras. Assim

$$\text{a*) } verdadeira \Rightarrow \text{a) } verdadeira$$

De c*) sabemos que se $\{A_n\}_{n=1}^{\infty} \in \mathcal{A}$ então $\{A_n^C\}_{n=1}^{\infty} \in \mathcal{A}$. A partir de b*) temos que $\bigcup_{i=1}^{\infty} A_i^C \in \mathcal{A}$, e novamente de c*) temos $(\bigcup_{i=1}^{\infty} A_i^C)^C \in \mathcal{A}$. Mas $(\bigcup_{i=1}^{\infty} A_i^C)^C =$

$\bigcap_{i=1}^{\infty} A_i$ e portanto b) é verdadeira.

Se $A_1,\ A_2 \in \mathcal{A}$ e $A_2 \subset A_1$ então, de c*), temos que $A_2^C \in \mathcal{A}$. Seja a seqüência $\{B_n\}_{n=1}^{\infty}$ onde $B_1 = A_1,\ B_2 = A_2^C$ e $B_j = S$ para $j \geq 3$. Assim $\{B_n\}_{n=1}^{\infty} \in \mathcal{A}$ e acabamos de ver que $\bigcap_{i=1}^{\infty} B_i \in \mathcal{A}$. Mas $\bigcap_{i=1}^{\infty} B_i = A_1 \cap A_2^C$ e portanto c) é verdadeira.

$*$

Observamos que uma consequência da definição de uma σ-álgebra $\mathcal{A}$ é o fato de que podemos realizar as operações de união, intersecção e complementação com elementos de $\mathcal{A}$ um número finito ou enumerável de vezes, sempre encontrando como resultado um conjunto pertencente a $\mathcal{A}$.

Mostremos que uma σ-álgebra $\mathcal{A}$ de subconjuntos de um dado conjunto S é fechada por uniões e intersecções finitas.

A partir da Definição 1, item b), temos que, para uma seqüência $\{A_n\}_{n=1}^{\infty}$, se $\{A_n\}_{n=1}^{\infty} \in \mathcal{A}$ então $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ e $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

Suponhamos então que a seqüência $\{A_n\}_{n=1}^{\infty}$ seja tal que, para um $k \in \mathbb{N}$,

$$A_n = \emptyset \ \ para \ \ n > k.$$

Assim,

$$\{A_n\}_{n=1}^{\infty} \in \mathcal{A} \Longrightarrow \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{k} A_i \in \mathcal{A}.$$

Agora suponhamos que uma seqüência $\{B_n\}_{n=1}^{\infty}$ seja tal que, para um $j \in \mathbb{N}$,

$$B_n = S \ \ para \ \ n > j.$$

Então

$$\{B_n\}_{n=1}^{\infty} \in \mathcal{A} \Longrightarrow \bigcap_{i=1}^{\infty} B_i = \bigcap_{i=1}^{j} B_i \in \mathcal{A}.$$

Exemplo 2 *Seja um conjunto* $S \neq \emptyset$. *Então a classe* $\mathcal{A} = \{S, \emptyset\}$ *é uma* σ-*álgebra de subconjuntos de* S. *Neste caso é dita* σ-*álgebra trivial. Se realizamos as operações de união, intersecção ou complementação com* S *e/ou* $\emptyset$ *um número finito ou enumerável de vezes, sempre encontramos como resultado* S *ou* $\emptyset$.

Temos que a maior σ-álgebra de subconjuntos de um conjunto S é a classe de todos os subconjuntos de S, a qual é denotada por 2^S ou $\mathcal{P}(S)$.

Exemplo 3 *Sejam* $S \neq \emptyset$ *e* $A \subset S$. *Então* $\mathcal{A} = \{S, A, A^C, \emptyset\}$ *é uma* σ-*álgebra de subconjuntos de* S.

Exemplo 4 *Seja* $S = I\!N^*$ *o conjunto dos números naturais maiores ou igual a um. Sejam* $P = \{x \in I\!N^* : x \text{ é par}\}$ *e* $I = \{x \in I\!N^* : x \text{ é ímpar}\}$. *Então* $\mathcal{A} = \{S, P, I, \emptyset\}$ *é uma* σ-*álgebra de subconjuntos de* $S = I\!N^*$.

1.1.2 σ-Álgebra de Borel

Definição 5 [*Espaço Topológico*] *Seja* χ *um conjunto não-vazio. Uma classe* $\mathcal{D}$ *de subconjuntos de* χ *é chamada topologia se satisfaz as seguintes condições:*

a) $\emptyset \in \mathcal{D}$.

b) $\chi \in \mathcal{D}$.

c) *A intersecção de um número finito de elementos de* $\mathcal{D}$ *pertence a* $\mathcal{D}$.

d) *A união de um número arbitrário de elementos de* $\mathcal{D}$ *pertence a* $\mathcal{D}$.

Os conjuntos em $\mathcal{D}$ *são ditos conjuntos abertos e o par* $(\chi, \mathcal{D})$ *é dito espaço topológico.*

Exemplo 6 *Seja $\chi \neq \emptyset$. A classe $\mathcal{C} = \{\emptyset, \chi\}$ é a topologia trivial.*

Exemplo 7 *Sejam $\chi \neq \emptyset$ e $A \subseteq \chi$. A classe $\mathcal{C} = \{\emptyset, A, \chi\}$ é uma topologia.*

Exemplo 8 *Seja $\chi \neq \emptyset$. A classe $\mathcal{C} = \mathcal{P}(\chi)$ é chamada topologia[1] discreta.*

Proposição 9 *Para uma coleção arbitrária de topologias de subconjuntos de um conjunto χ, temos que a intersecção dessas topologias é também uma topologia de subconjuntos de χ.*

Exemplo 10 *Seja $\chi = \mathbb{R}$. A menor topologia que contém os intervalos abertos é chamada topologia usual e é definida pela intersecção de todas as topologias que contém tais conjuntos.*

Proposição 11 *Para uma coleção arbitrária de σ-álgebras de subconjuntos de um conjunto S, temos que a intersecção dessas σ-álgebras é também uma σ-álgebra de subconjuntos de S.*

Devido à Proposição 11 e ao fato de que 2^S é uma σ-álgebra temos que, para qualquer coleção $\mathcal{C}$ de subconjuntos de S, existe uma menor σ-álgebra $\mathcal{A}$ que contém $\mathcal{C}$, ou seja, a intersecção de todas as σ-álgebras que contém $\mathcal{C}$. Esta menor σ-álgebra é dita σ-álgebra *gerada* por $\mathcal{C}$.

Definição 12 *A σ-álgebra gerada pela coleção $\mathcal{C}$ de subconjuntos abertos de um espaço topológico $(\chi, \mathcal{C})$ é dita σ-álgebra de Borel. Um conjunto pertencente a esta σ-álgebra é dito um conjunto de Borel ou boreliano.*

[1] Outros exemplos de topologias podem ser encontrados em [8].

No caso $\chi = I\!\!R$, a σ-álgebra gerada pela topologia usual é chamada σ-álgebra de Borel e é denotada por $\mathcal{B}$.

Observemos que a σ-álgebra de Borel de $I\!\!R$ também é gerada, por exemplo, pelos intervalos da forma $[b, \infty)$, para $b \in I\!\!R$, já que $[b, \infty) = (-\infty, b)^C$.

Podemos também mostrar que a σ-álgebra de Borel de $I\!\!R$ é a menor σ-álgebra de $I\!\!R$ que inclui todos os intervalos.

Prova.

$i)$ Mostremos que todo intervalo pertence a $\mathcal{B}$.

Sejam a, $b \in I\!\!R$ tais que $a < b$. Assim

$[a, b] = (-\infty, a)^C \cap (b, \infty)^C \in \mathcal{B}$,

$[a, b) = (-\infty, a)^C \cap (a - 1, b) \in \mathcal{B}$,

$(a, b] = (a, b + 1) \cap (b, \infty)^C \in \mathcal{B}$,

$(-\infty, b] = (b, \infty)^C \in \mathcal{B}$,

$[a, \infty) = (-\infty, a)^C \in \mathcal{B}$.

$ii)$ Mostremos que $\mathcal{B}$ é a menor σ-álgebra que contém todos os intervalos.

Seja $\mathcal{B}' \subseteq \mathcal{B}$ uma σ-álgebra tal que $\mathcal{B}'$ contém todos os intervalos.

$\Rightarrow \mathcal{B}'$ contém todos os intervalos abertos

$\Rightarrow \mathcal{B} \subseteq \mathcal{B}'$ pela definição de σ-álgebra de Borel

$\Rightarrow \mathcal{B}' = \mathcal{B}$.

*

1.1.3 Medidas

A princípio, uma *medida* pode ser interpretada como uma forma de se atribuir um valor numérico ao "tamanho"de um conjunto, mas a fim de obtermos uma interpretação mais precisa a respeito, enunciamos a definição que segue.

Definição 13 *Seja $\mathcal{A}$ uma σ-álgebra de um conjunto S. Uma função $\mu : \mathcal{A} \to [0, \infty]$ é dita uma medida se*

- $\mu(\emptyset) = 0$,
- *dada uma seqüência $\{A_n\}_{n=1}^{\infty}$ de conjuntos mutuamente disjuntos em $\mathcal{A}$, $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$.*

A partir desta definição, observamos que o domínio de uma medida é uma coleção de conjuntos.

Exemplo 14 *Seja um conjunto $S \neq \emptyset$ e tomemos $A \subseteq S$. Definimos $\mu(A)$ como o número de elementos de A. Assim, $\mu(S) > 0$, $\mu(\emptyset) = 0$, e se $A_1 \cap A_2 = \emptyset$ então $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$. Notemos que $\mu(A) = \infty$ é possível desde que S tenha infinitos elementos. A medida μ descrita aqui é dita "medida de contagem em S"e pode ser definida em $\mathcal{P}(S)$.*

Exemplo 15 *Na σ-álgebra do Exemplo 3, podemos definir a medida*

$$\mu(A) = p = 1 - \mu(A^C), \text{ com } \mu(S) = 1.$$

1.2 Medida de Lebesgue e Integral de Lebesgue

1.2.1 Medida de Lebesgue de um conjunto linear de pontos

Definição 16 *Tomemos como espaço o conjunto* $\mathbb{R}$ *dos números reais. Qualquer subconjunto de* $\mathbb{R}$ *é dito um conjunto linear de pontos. Desta maneira, um intervalo é um caso simples de um conjunto deste tipo.*

Procederemos agora com a definição formal do *comprimento de um intervalo*, que será o ponto de partida para o desenvolvimento que faremos no decorrer desta Subseção.

Definição 17 *O comprimento de um intervalo finito* $i = (a, b)$ *é a quantidade não negativa* $b - a$*. Ainda, o comprimento assume o mesmo valor para um intervalo fechado, aberto ou semi-aberto com os mesmos extremos. Para um intervalo degenerado, isto é, para um ponto, o comprimento é* 0*. O comprimento de um intervalo infinito é definido como* $+\infty$*.*

Desta forma, associamos a todo intervalo $i = (a, b)$ um comprimento não negativo, o qual pode ser finito ou infinito. Podemos expressar isto dizendo que a função $L(i)$ que representa o comprimento do intervalo i é uma função não negativa deste intervalo.

Se um intervalo for dividido em um número finito ou enumerável de outros intervalos sem pontos em comum, o comprimento deste intervalo é igual à soma dos comprimentos destes últimos intervalos disjuntos. Expressamos esta propriedade dizendo que o comprimento $L(i)$ é uma função aditiva do intervalo i.

A fim de extendermos a Definição 17 para conjuntos mais genéricos ou mais complicados do que intervalos, enunciamos a seguinte definição, que trata da medida de uma união de intervalos.

Definição 18 *Consideremos uma classe de conjuntos $\mathcal{F}$ tal que um elemento desta classe seja uma união de uma seqüencia finita ou enumerável de intervalos.*

Seja $I \in \mathcal{F}$. Fazendo algumas transformações podemos escrever

$$I = \bigcup_{v=1}^{\infty} i_v$$

de tal forma que cada elemento da seqüencia $\{i_v\}_{v=1}^{\infty}$ seja um intervalo e ainda $i_u \cap i_v = \emptyset$ para $u \neq v$.

Então definimos a medida $L(I)$ como

$$L(I) = \sum_{v=1}^{\infty} L(i_v)$$

onde $L(i_v)$ denota o comprimento do intervalo i_v.

Podemos mostrar que a medida $L(I)$ está bem definida.

Proposição 19 *A medida $L(I)$ define para todo conjunto $I \in \mathcal{F}$ uma medida única, satisfazendo as seguintes condições:*

a) $L(I) \geq 0.$

b) *Se $I = \bigcup_{v=1}^{\infty} I_v$ onde $I_v \cap I_u = \emptyset$ para $v \neq u$, então $L(I) = \sum_{v=1}^{\infty} L(I_v)$.*

c) *No caso particular em que I é um intervalo, $L(I)$ é igual ao comprimento do intervalo.*

Ainda com o objetivo de extender as últimas definições para uma classe mais geral de conjuntos, e finalmente definir a medida de Lebesgue, introduziremos o conceito de duas funções auxiliares, chamadas *medida interna* e *medida externa*, definidas para todo subconjunto limitado de $I\!R$.

Dizemos que um subconjunto de $I\!R$ é limitado se seus limites superior e inferior são ambos finitos.

Definição 20 *Seja um conjunto $A \subset I\!R$, $A \neq \emptyset$. Dizemos que um ponto α é o limite inferior de A se este ponto for tal que, para todo $\varepsilon > 0$, existe ao menos um ponto de A pertencendo ao intervalo $[\alpha, \alpha + \varepsilon]$, enquanto que não existe nenhum ponto de A pertencendo ao intervalo $(-\infty, \alpha)$. Quando não existe um α finito com esta propriedade, dizemos que o limite inferior de A é $-\infty$.*

Analogamente, dizemos que um ponto β é o limite superior de A se este ponto for tal que, para todo $\varepsilon > 0$, existe pelo menos um ponto de A pertencendo ao intervalo $[\beta - \varepsilon, \beta]$ e nenhum ponto de A pertencendo ao intervalo (β, ∞). Da mesma forma, se não existir um β finito com esta propriedade, dizemos que o limite superior de A é $+\infty$.

Notemos que os pontos α e β podem ou não pertencer ao conjunto A.

Exemplo 21 *Seja o conjunto $A = (a, b] \cup [c, d] \subset I\!R$, com $a < b < c < d$. Então os limites inferior e superior de A são, respectivamente, a e d. Notemos que $d \in A$ mas $a \notin A$.*

Agora, dado $S \subseteq I\!R$ limitado, consideremos $I \in \mathcal{F}$ tal que

$$S \subset I \subset (a, b),$$

a, $b \in I\!R$, fixos. Pela Proposição 19, a medida $L(I)$ está bem definida.

Seja H o conjunto formado pelos números $L(I)$ correspondentes a todos os possíveis conjuntos I tais que $S \subset I \subset (a, b)$, a, $b \in I\!R$, fixos. Assim este conjunto H possui um limite inferior finito, pois $L(I) \geq 0$.

Definição 22 *A medida externa do conjunto S, denotada por $\overline{L}(S)$, é definida como o limite inferior do conjunto H citado acima. A medida interna de S denotada por $\underline{L}(S)$ é definida pela relação*

$$\underline{L}(S) = b - a - \overline{L}\left(S^C\right)$$

onde S^C representa o complemento de S com respeito ao intervalo (a, b).

Como todo conjunto S agora considerado é um subconjunto do intervalo (a, b), o qual pertence à classe $\mathcal{F}$, temos

$$0 \leq \overline{L}(S) \leq b - a$$

e

$$0 \leq \underline{L}(S) \leq b - a.$$

Diretamente da Definição 22 segue que as medidas $\overline{L}(S)$ e $\underline{L}(S)$ independem do intervalo (a, b) considerado e são ambas funções monótonas de S, isto é,

$$\overline{L}(S_1) \leq \overline{L}(S_2)$$

e

$$\underline{L}(S_1) \leq \underline{L}(S_2)$$

para $S_1 \subset S_2$.

Definição 23 *Um conjunto limitado S é dito mensurável se suas medidas interna e externa forem iguais. Esse valor comum será denotado por $\lambda(S)$ e chamado medida de Lebesgue de S:*

$$\overline{L}(S) = \underline{L}(S) = \lambda(S).$$

Um conjunto não limitado S é dito mensurável se a intersecção $i_x \cap S$, onde i_x denota o intervalo fechado $[-x, x]$, for mensurável para todo $x > 0$. Neste caso a medida $\lambda(S)$ fica definida pela relação

$$\lambda(S) = \lim_{x \to \infty} \lambda(i_x \cap S).$$

A partir da monotonicidade das medidas interna e externa, temos que $\lambda(i_x \cap S)$ é uma função não decrescente de x e portanto o último limite, que pode ser finito ou infinito, sempre existe.

Temos assim que a medida $\lambda(S)$ é também uma função monótona do conjunto S, ou seja,

$$\lambda(S_1) \leq \lambda(S_2)$$

para $S_1 \subset S_2$, S_1 e S_2 conjuntos mensuráveis.

Ainda, da Proposição 19, a medida de Lebesgue é a única função de conjuntos definida para todos os conjuntos de Borel em $\mathbb{R}$ que satisfaz as condições $a), b)$ e $c)$.

Definição 24 *Uma classe $\mathcal{A}$ de subconjuntos de um conjunto S é dita uma álgebra se*

- *$A \in \mathcal{A}$ implica $A^C \in \mathcal{A}$,*

- $A_1, A_2 \in \mathcal{A}$ *implica* $A_1 \cup A_2 \in \mathcal{A}$.

É interessante destacar que toda σ-álgebra é uma álgebra, mas nem toda álgebra é uma σ-álgebra . Vejamos o exemplo a seguir.

Exemplo 25 *Consideremos a classe de conjuntos* $\mathcal{S}$ *definida por*

$$\mathcal{S} = \{A \subseteq I\!R : A \text{ é uma união finita de intervalos } \}.$$

Claramente o complemento de uma união finita de intervalos também é uma união finita de intervalos, e portanto

$$A \in \mathcal{S} \Rightarrow A^C \in \mathcal{S}.$$

Ainda mais, uma união finita de uniões finitas de intervalos também é uma união finita de intervalos, e assim

$$\{A_i\}_{i=1}^k \in \mathcal{S} \Rightarrow \bigcup_{i=1}^k A_i \in \mathcal{S}, \quad k < \infty.$$

Fica claro, portanto, que a classe $\mathcal{S}$ *é uma álgebra.*

Por outro lado, se considerarmos a seqüência $\{A_n\}$ *de elementos em* $\mathcal{S}$ *tal que* $A_n = (n, n + \frac{1}{2})$, *para cada* $n \in I\!N$, *então temos que*

$$\bigcup_{n=1}^\infty A_n \notin \mathcal{S},$$

o que nos leva a concluir que a classe $\mathcal{S}$ *não constitui uma* σ*-álgebra .*

Podemos ainda utilizar o seguinte resultado, conhecido como Teorema de Extensão de Caratheodory, para garantir a existência da medida de Lebesgue a partir de sua definição na classe $\mathcal{F}$. Este resultado nos permite trabalhar com a medida de Lebesgue em estruturas mais simples, como a própria álgebra dos intervalos $\mathcal{F}$.

Teorema 26 *Seja μ uma função real extendida*[2] *definida em uma álgebra $\mathcal{C}$ de subconjuntos de um conjunto S, tal que μ é $\sigma - finita$ (ver Definição 44), não negativa, finitamente aditiva*[3] *e satisfaz $\mu(\emptyset) = 0$. Então existe uma única extensão de μ a uma medida em um espaço de medida $(S, \mathcal{A}, \mu^*)$ (ver Definição 37). Ou seja, para $\mathcal{A} \supseteq \mathcal{C}$, temos $\mu(A) = \mu^*(A)$ para todo $A \in \mathcal{C}$.*

Se tomarmos a medida L definida na álgebra $\mathcal{F}$ de subconjuntos de $I\!R$ (ver Definição 18) podemos, pelo Teorema 26, extendê-la a uma medida λ no espaço de medida $(I\!R, \mathcal{B}, \lambda)$, onde $\mathcal{B}$ é a σ-álgebra de Borel de subconjuntos de $I\!R$, já que $\mathcal{F} \subset \mathcal{B}$.

1.2.2 Integral de Lebesgue para funções de uma variável

Seja S um conjunto de Borel em $I\!R$ com medida de Lebesgue $\lambda(S)$ finita, e $g(x)$ uma função da variável real x definida para todos os valores de x pertencentes a S. Suponhamos que $g(x)$ é limitada em S, ou seja, os limites inferior e superior de $g(x)$ em S são finitos. Denotemos estes últimos limites por m e M respectivamente, e assim temos $m \leq g(x) \leq M$ para todo x pertencente a S.

Tomemos uma divisão do conjunto S em um número finito de partes

$$D = \{S_1, S_2, \ldots, S_n\}$$

[2]Assume valores em $\overline{\boldsymbol{R}} = \boldsymbol{R} \cup \{-\infty, +\infty\}$.

[3]Para uma seqüência de conjuntos $\{A_k\}_{k=1}^n \in \mathcal{C}$, n finito, $\mu(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n \mu(A_k)$ se $A_i \cap A_j = \emptyset$ para $i \neq j$.

de modo que

$$S = \bigcup_{i=1}^{n} S_i$$

com $S_u \cap S_v = \emptyset$ para $u \neq v$. No conjunto S_v, a função $g(x)$ possui um limite inferior m_v e um limite superior M_v, tais que $m \leq m_v \leq M_v \leq M$.

Definição 27 *Definimos as somas inferior e superior de Darboux, associadas à divisão de S citada, respectivamente, pelas relações*

$$z = \sum_{v=1}^{n} m_v \lambda(S_v)$$

e

$$Z = \sum_{v=1}^{n} M_v \lambda(S_v).$$

Temos, portanto, $m\lambda(S) \leq z \leq Z \leq M\lambda(S)$. Também segue diretamente da Definição 27 que qualquer divisão de S obtida a partir da subdivisão de algumas das partes S_v terá soma inferior ao menos igual à soma inferior da divisão original e soma superior no máximo igual à soma superior da divisão original.

Portanto, qualquer divisão de S em um número finito arbitrário de partes sem pontos em comum produz, de acordo com a Definição 27, uma soma inferior z e uma soma superior Z. Consideremos o conjunto de todas as possíves somas inferiores e o conjunto de todas as possíveis somas superiores, que serão chamados, respectivamente, $conjunto - z$ e $conjunto - Z$. Esses dois conjuntos são limitados, pois todos os possíveis valores de z e Z estão situados entre $m\lambda(S)$ e $M\lambda(S)$. Pode-se mostrar que o limite superior do $conjunto - z$ é no máximo igual ao limite inferior do $conjunto - Z$.

O limite superior do $conjunto - z$ será chamado integral inferior de $g(x)$ sobre S, enquanto que o limite inferior do $conjunto - Z$ será chamado integral superior de $g(x)$ sobre S.

Desta maneira escrevemos

$$\underline{\int_S} g(x)\, dx = \lim\sup\{z : \text{divisão } D\}$$

e

$$\overline{\int_S} g(x)\, dx = \lim\inf\{Z : \text{divisão } D\},$$

e segue que

$$m\lambda(S) \leq \underline{\int_S} g(x)\, dx \leq \overline{\int_S} g(x)\, dx \leq M\lambda(S).$$

Definição 28 *Se as integrais inferior e superior de $g(x)$ sobre S forem iguais, então $g(x)$ é dita ser integrável sobre S no sentido de Lebesgue, ou simplesmente integrável sobre S. O valor comum das duas integrais é chamado integral de Lebesgue de $g(x)$ sobre S, e escrevemos*

$$\underline{\int_S} g(x)\, dx = \overline{\int_S} g(x)\, dx = \int_S g(x)\, dx.$$

Uma condição necessária e suficiente para a integrabilidade de $g(x)$ sobre S é que, para todo $\epsilon > 0$, possamos encontrar uma divisão de S tal que a diferença $Z - z$ correspondente seja menor do que ϵ.

Observamos que a definição da integral de Lebesgue é perfeitamente análoga à definição da integral de Riemann. Neste caso, o conjunto S é um intervalo, que por sua vez é dividido em um número finito de sub-intervalos S_v, e as somas de Darboux, z e Z, são compostas da mesma maneira, onde $\lambda(S_v)$ denota o comprimento do sub-intervalo S_v. A diferença é que, no caso da integral de Lebesgue, consideramos uma classe de conjuntos mais geral do que intervalos quando tomamos S e as partes S_v como sendo quaisquer conjuntos de Borel em $I\!R$.

Quando estudamos funções integráveis no sentido de Riemann, e integrais de Lebesgue sobre um intervalo, não é necessário distinguir os dois tipos de integral, de Riemann e de Lebesgue.

A integral de Lebesgue sobre um intervalo (a, b) é usualmente escrita com a mesma notação da integral de Riemann:

$$\int_a^b g(x)\,dx.$$

Temos ainda que esta integral possui o mesmo valor quando consideramos o intervalo (a, b) fechado, aberto ou semi-aberto.

Ainda neste contexto em que consideramos funções limitadas sobre conjuntos de Borel em $I\!R$ com medida de Lebesgue finita, podemos enunciar as seguintes proposições, que são perfeitamente análogas às proposições correspondentes à integral de Riemann:

$$\int_S (g_1(x) + g_2(x))\,dx = \int_S g_1(x)\,dx + \int_S g_2(x)\,dx, \tag{1.1}$$

$$\int_S cg(x)\,dx = c\int_S g(x)\,dx, \tag{1.2}$$

$$\int_{S_1 \cup S_2} g(x)\, dx = \int_{S_1} g(x)\, dx + \int_{S_2} g(x)\, dx, \tag{1.3}$$

$$m\lambda(S) \leq \int_S g(x)\, dx \leq M\lambda(S), \tag{1.4}$$

onde c é uma constante arbitrária, os conjuntos S_1 e S_2 não possuem pontos em comum, e m e M denotam respectivamente os limites inferior e superior de $g(x)$ em S.

As proposições (1.1) e (1.3) podem ser imediatamente extendidas para um número finito arbitrário de termos. De (1.4) ainda segue que a integral de uma função $g(x)$ limitada sobre um conjunto com medida de Lebesgue igual a 0 é sempre igual a 0, e se $g(x)$ for identicamente igual a 1 sobre um conjunto S com medida de Lebesgue não necessariamente igual a 0, então

$$\int_S g(x)\, dx = \int_S dx = \lambda(S).$$

Consideremos agora o caso em que o conjunto de Borel S possui medida de Lebesgue $\lambda(S)$ finita mas a função $g(x)$ não é necessariamente limitada em S.

Tomemos quaisquer números finitos a e b tais que $a < b$ e coloquemos

$$g_{a,b}(x) = \begin{cases} a & \text{, se } g(x) < a \\ g(x) & \text{, se } a \leq g(x) \leq b \\ b & \text{, se } g(x) > b \end{cases} \tag{1.5}$$

Desta forma, $g_{a,b}(x)$ é limitada e integrável sobre S. Se o limite

$$\lim_{a \to -\infty, b \to +\infty} \int_S g_{a,b}(x) dx = \int_S g(x) dx$$

existir e possuir valor finito, dizemos que $g(x)$ é integrável sobre S. Este limite será por definição a integral de Lebesgue de $g(x)$ sobre S.

Neste contexto, em que $g(x)$ não é necessariamente limitada, ainda são válidas as Proposições (1.1), (1.2), (1.3) e (1.4). Com respeito à Proposição (1.4) devemos observar que um dos limites m e M, ou ambos, podem ser infinitos.

Consideremos o caso em que $g(x)$ não é necessariamente limitada e o conjunto de Borel S não mais possui medida de Lebesgue finita.

Denotemos por $S_{a,b}$ a intersecção de S com o intervalo fechado $[a, b]$, onde a e b são finitos. Assim, $S_{a,b}$ possui medida de Lebesgue finita.

Definição 29 *Seja a função $g(x)$ integrável sobre $S_{a,b}$ para todo a e todo b. Se o limite*

$$\lim_{a\to-\infty, b\to+\infty} \int_{S_{a,b}} |g(x)|dx = \int_S |g(x)|dx$$

existir e possuir valor finito, então dizemos que $g(x)$ é absolutamente integrável sobre S.

Teorema 30 *Se a função $g(x)$ for absolutamente integrável sobre S então $g(x)$ é integrável sobre S, ou seja, o limite*

$$\lim_{a\to-\infty, b\to+\infty} \int_{S_{a,b}} g(x)dx = \int_S g(x)dx$$

existe e possui valor finito.

Dizemos portanto que se $g(x)$ for absolutamente integrável sobre S, então a integral de Lebesgue de $g(x)$ sobre S é convergente e assume, por definição, o valor do limite enunciado no Teorema 30.

Segue diretamente da definição que, também neste caso, valem as proposições (1.1), (1.2) e (1.3). Ao invés de obtermos a proposição (1.4), temos simplesmente a seguinte relação:

$$\int_S g(x)dx \geq 0 \text{ se } g(x) \geq 0 \text{ para todo } x \in S.$$

Podemos ainda estudar a integral de Lebesgue como uma função não negativa e aditiva de conjuntos. Consideremos uma função não negativa fixa $f(x)$, integrável sobre qualquer intervalo finito, e um conjunto de Borel S em $I\!R$.

Definimos

$$P(S) = \begin{cases} \int_S f(x)dx & \text{, se } f(x) \text{ é integrável sobre } S \\ +\infty & \text{, caso contrário.} \end{cases}$$

Desta forma $P(S)$ é uma função não negativa e aditiva de conjuntos, definida para todos os conjuntos de Borel em $I\!R$.

No caso particular em que $f(x) = 1$, temos $P(S) = \lambda(S)$, a medida de Lebesgue de S. Outro importante caso particular ocorre quando $f(x)$ é integrável sobre todo o espaço $I\!R$ de modo que acabamos tendo $P(S)$ sempre finita e para qualquer conjunto de Borel S em $I\!R$

$$P(S) \leq \int_{-\infty}^{+\infty} f(x)dx.$$

1.3 Medida de probabilidade e função de distribuição

1.3.1 Funções não negativas e aditivas de conjuntos em $I\!R$

Na Subseção 1.2.1 determinamos a medida de Lebesgue $\lambda(S)$ para qualquer conjunto de Borel S em $I\!R$. Vimos que esta medida satisfaz as três condições que requerem que $\lambda(S)$ seja a) não negativa, b) aditiva e c) para qualquer intervalo, igual ao comprimento deste intervalo. Vimos finalmente que $\lambda(S)$ é a única função de conjuntos que satisfaz essas condições.

Por outro lado, se omitirmos a condição c) então $\lambda(S)$ não será mais a única função de conjuntos satisfazendo as condições restantes. Somos levados, portanto, ao conceito geral de uma função de conjuntos não negativa e aditiva como uma generalização natural da medida de Lebesgue.

Consideremos uma função de conjuntos $P(S)$ definida para todos os conjuntos de Borel S em $I\!R$ e satisfazendo as três seguintes condições:

A) $P(S)$ é não negativa: $P(S) \geq 0$.

B) $P(S)$ é aditiva:

$P(\bigcup_{i=1}^{\infty} S_i) = \sum_{i=1}^{\infty} P(S_i)$ para $S_u \cap S_v = \emptyset$ quando $u \neq v$.

C) $P(S)$ é finita para qualquer conjunto S limitado.

Assumimos que todas as funções de conjuntos tratadas nesta e na próxima subseção satisfazem essas condições.

Quando um conjunto S é constituído por todos os pontos ξ que satisfazem uma certa relação, denotamos o valor $P(S)$ simplesmente substituindo a letra S dentro dos parênteses pela relação em questão. Por exemplo, se S é o intervalo fechado $[a, b]$, escrevemos $P(S) = P(a \leq \xi \leq b)$. Se S é constituído pelo ponto $\xi = a$ então escrevemos $P(S) = P(\xi = a)$.

Chamamos $P(S)$ de função de conjuntos pois o argumento desta função é um conjunto. Para uma função $F(x_1, ..., x_n)$ de uma ou mais variáveis, o argumento pode ser considerado um ponto com as coordenadas $x_1, ..., x_n$, e nos referimos a tal função como uma função pontual.

Quando uma função de conjuntos $P(S)$ e uma constante k são dadas, definimos uma função pontual correspondente $F(x; k)$ por

$$F(x;k) = \begin{cases} P(k < \xi \leq x) & , \text{ para } x > k \\ 0 & , \text{ para } x = k \\ -P(x < \xi \leq k) & , \text{ para } x < k. \end{cases} \tag{1.6}$$

Qualquer que seja o valor do parâmetro constante k, obtemos para um intervalo finito $(a, b]$

$$F(b; k) - F(a; k) = P(a < \xi \leq b) \geq 0,$$

que mostra que F(x;k) é uma função não decrescente de x. Se nesta última relação consideramos a tendendo a $-\infty$ ou b tendendo a $+\infty$, ou ambos, segue que a mesma relação vale também para intervalos infinitos. No caso particular quando $P(S)$ é a medida de Lebesgue $\lambda(S)$, temos $F(x; k) = x - k$.

As funções $F(x; k)$ correspondentes a dois valores diferentes do parâmetro k diferem por uma quantidade independente de x. De fato, se $k_1 < k_2$, obtemos

$$F(x; k_1) - F(x; k_2) = P(k_1 < \xi \leq k_2).$$

Assim, se escolhermos um valor arbitrário k_0 para k e denotarmos a função correspondente $F(x; k_0)$ simplesmente por $F(x)$, qualquer outra $F(x; k)$ será da forma $F(x) + c$, onde c é uma constante independente de x.

Teorema 31 *Podemos assim dizer que para qualquer função de conjuntos $P(S)$ satisfazendo as condições $A)$, $B)$ e $C)$ descritas nesta subseção corresponde uma função pontual $F(x)$ não decrescente tal que, para qualquer intervalo (a, b) finito ou infinito, temos*

$$F(b) - F(a) = P(a < \xi \leq b).$$

$F(x)$ é univocamente determinada exceto por uma constante aditiva.

Escolhemos agora um valor arbitrário fixo para o parâmetro k e consideramos a função $F(x)$ correspondente. Como $F(x)$ é não decrescente, os limites pela direita e pela esquerda,

$$F(a + 0) = \lim_{x \to a^+} F(x) \text{ e } F(a - 0) = \lim_{x \to a^-} F(x),$$

existem para todo valor de a, e $F(a - 0) \leq F(a + 0)$. De acordo com o último teorema, temos para $x > a$

$$F(x) - F(a) = P(a < \xi \leq x). \tag{1.7}$$

Definição 32 *Seja uma seqüência $\{A_n\}_{n=1}^{\infty}$ de subconjuntos de $I\!R$. Definimos o limite desta seqüência por*

$$A = \{x \in I\!R : \exists n_0 \in I\!N \;\; tal \;\; que \;\; x \in A_n, \forall n \geq n_0\},$$

e denotamos $\lim_{n\to\infty} A_n = A$*, quando este conjunto* A *existir.*

Considerando a relação (1.7) para uma sequência decrescente de valores de x tendendo ao valor fixo a, temos que os intervalos semi-abertos $(a, x]$ correspondentes formam uma sequência decrescente de conjuntos cujo limite é o conjunto vazio e assim

$$F(a + 0) = F(a).$$

Por outro lado, se $x < a$ temos

$$F(a) - F(x) = P(x < \xi \leq a),$$

e um argumento similar mostra que

$$F(a - 0) = F(a) - P(\xi = a) \leq F(a).$$

Portanto a função $F(x)$ é sempre contínua à direita. Para todo valor de x tal que $P(\xi = x) > 0$, $F(x)$ possui uma descontinuidade com o salto $P(\xi = x)$. Para todo valor de x tal que $P(\xi = x) = 0$, $F(x)$ é contínua. Então qualquer x tal que $P(S) > 0$, para o conjunto S constituído somente pelo ponto x, é um ponto de descontinuidade de $F(x)$. Esses pontos são ditos pontos de descontinuidade também para $P(S)$, e qualquer ponto de continuidade de $F(x)$ é também dito um ponto de continuidade de $P(S)$. Os pontos de descontinuidade de $P(S)$ e $F(x)$ formam no máximo um conjunto enumerável.

Teorema 33 *Para qualquer função pontual não decrescente* $F(x)$*, finita para todo* x *finito e contínua à direita, corresponde uma função de conjuntos* $P(S)$ *univocamente determinada para todo conjunto de Borel* S *em* $\mathbb{R}$*, satisfazendo as condições*

A), B) e C) descritas, de modo que a relação

$$F(b) - F(a) = P(a < \xi \leq b)$$

seja válida para qualquer intervalo finito ou infinito (a, b).

Ainda, para duas funções $F_1(x)$ *e* $F_2(x)$ *corresponde a mesma* $P(S)$ *se e somente se a diferença* $F_1(x) - F_2(x)$ *for constante.*

Relativamente a este último teorema, temos que a função pontual não decrescente $F(x)$ determina uma função de intervalos não negativa $P(i)$ que pode ser definida como o crescimento de $F(x)$ sobre o intervalo i. Para algum intervalo semi-aberto definido por $a < x \leq b$, $P(i)$ assume o valor $P(a < x \leq b) = F(b) - F(a)$. Para os três outros tipos de intervalo com os mesmos extremos a e b, determinamos o valor de $P(i)$ por um processo limite simples e obtemos

$$P(a \leq x \leq b) = F(b) - F(a - 0),$$

$$P(a < x < b) = F(b - 0) - F(a),$$

$$P(a \leq x < b) = F(b - 0) - F(a - 0),$$

de modo que $P(i)$ fica completamente determinada para qualquer intervalo i.

O teorema afirma desta forma que é possível encontrar uma função de conjuntos não negativa e aditiva definida para todo conjunto de Borel S em $I\!R$ e igual a $P(i)$ no caso particular em que S é um intervalo i.

Comparando os dois últimos teoremas obtemos que, se duas funções $F_1(x)$ e $F_2(x)$ diferindo por uma constante forem consideradas a mesma, então há, de uma maneira geral, uma correspondência unívoca entre as funções de conjuntos $P(S)$ e as funções pontuais não decrescentes $F(x)$.

Definição 34 *Uma função de conjuntos $P(S)$ definida para um conjunto de Borel S em $I\!R$ e satisfazendo as condições A), B) e C) define uma $P-medida$ do conjunto S, a qual constitui uma generalização da medida de Lebesgue $\lambda(S)$. Como esta última, a $P-medida$ é não negativa e aditiva.*

De acordo com os últimos resultados, a $P - medida$ é univocamente determinada para qualquer conjunto de Borel S em $I\!R$ pela função pontual não decrescente $F(x)$ correspondente. Como $F(x)$ é contínua à direita, é suficiente conhecer $F(x)$ em seus pontos de continuidade.

Observemos que para qualquer conjunto de Borel S em $I\!R$ temos que $P(S) \leq P(I\!R)$ pois $S \subset I\!R$. Isto dá passo para a seguinte definição.

Definição 35 *Se $P(I\!R)$ for finita, dizemos que a função de conjuntos $P(S)$ é limitada.*

Quando $P(S)$ é limitada, sempre fixamos a constante aditiva na função pontual não decrescente $F(x)$ correspondente tomando $k = -\infty$ em (1.6), de modo que obtemos para todo valor de x

$$F(x) = P(\xi \leq x).$$

Quando x tende a $-\infty$ nesta relação, o conjunto de todos os pontos $\xi \leq x$ tende ao conjunto vazio e portanto $F(-\infty) = 0$. Por outro lado, quando x tende a $+\infty$, o conjunto dos pontos $\xi \leq x$ tende ao espaço $I\!R$ e portanto $F(+\infty) = P(I\!R)$. Como $F(x)$ é não decrescente, temos para todo x

$$0 \leq F(x) \leq P(I\!R).$$

1.3.2 Distribuições

Uma função de conjuntos não negativa e aditiva, $P(S)$, definida para todo conjunto de Borel S em $\mathbb{R}$, tal que $P(\mathbb{R}) = 1$, é obviamente limitada e a função pontual não decrescente $F(x)$ correspondente fica definida de modo que

$$F(x) = P(\xi \leq x),$$

$$0 \leq F(x) \leq 1,$$

$$F(-\infty) = 0 \text{ e } F(+\infty) = 1.$$

Um par de funções $P(S)$ e $F(x)$ deste tipo é concretamente interpretado de modo a representar uma "distribuição de massa sobre o espaço unidimensional $\mathbb{R}$". Imaginemos uma unidade de massa distribuída sobre $\mathbb{R}$ de modo que, para todo x, a quantidade de massa acumulada sobre o intervalo infinito $(-\infty, x]$ seja igual a $F(x)$. A interpretação da função de conjuntos $P(S)$ pode portanto ser colocada dizendo-se que qualquer conjunto de Borel S em $\mathbb{R}$ possui uma determinada quantidade de massa $P(S)$. A quantidade de massa total sobre o espaço $\mathbb{R}$ é $P(\mathbb{R}) = 1$.

Temos a liberdade de definir tal distribuição através da função de conjuntos $P(S)$ ou através da função pontual $F(x)$ correspondente, e dizemos que $P(S)$ é a medida de probabilidade da distribuição enquanto que $F(x)$ é dita a função de distribuição.

Portanto uma função de distribuição é uma função pontual não decrescente $F(x)$, contínua à direita e tal que $F(-\infty) = 0$ e $F(+\infty) = 1$. Ainda temos que qualquer $F(x)$ com estas propriedades determina uma única distribuição de probabilidade, possuindo $F(x)$ como sua função de distribuição.

Se x_0 é um ponto de descontinuidade de $F(x)$ com um salto igual a p_0 então a massa p_0 fica concentrada no ponto x_0, ou seja, $P(\xi = x_0) = p_0$. Por outro lado, se x_0 é um ponto de continuidade então $P(\xi = x_0) = 0$.

Podemos ainda extender a definição de uma medida de probabilidade para um contexto mais geral, no qual não necessariamente o espaço e a σ-álgebra considerados sejam o conjunto $I\!R$ dos números reais e a σ-álgebra de Borel, respectivamente.

Definição 36 *Sejam $\mathcal{A}$ uma σ-álgebra de subconjuntos de um dado conjunto Ω, e μ uma medida definida segundo a Definição 13, levando em conta a σ-álgebra $\mathcal{A}$. Então μ é dita uma medida de probabilidade se $\mu(\Omega) = 1$.*

Veremos mais adiante, na Seção 1.5, que é possível induzir uma medida de probabilidade definida para conjuntos de Borel em $I\!R$ a partir de uma medida de probabilidade como a definida acima, de modo que a essa medida induzida possamos associar univocamente uma função de distribuição.

1.4 Integral de Lebesgue-Stieltjes para funções de uma variável

Já vimos que a teoria da medida de Lebesgue pode ser generalizada com a introdução do conceito de uma $P-medida$ não negativa e aditiva. Veremos agora que uma generalização análoga pode ser aplicada à teoria da integral de Lebesgue.

Consideremos que uma $P-medida$ fixa é dada. Esta medida pode ser definida por uma função de conjuntos $P(S)$ não negativa e aditiva, ou pela função pontual não decrescente $F(x)$ correspondente. Já vimos que essas duas funções são perfeitamente equivalentes para o propósito de se definir a $P - medida$.

Consideremos ainda uma função $g(x)$, definida e limitada para todo x pertencente a um dado conjunto de Borel S em $I\!R$ com $P - medida$ finita. Da mesma forma como foi feito anteriormente, no desenvolvimento que levou à definição da integral de Lebesgue, dividimos S em um número arbitrário finito de partes $S_1, S_2, ..., S_n$, de modo que $S_i \cap S_j = \emptyset$, para $i \neq j$. Na Definição 27 das somas de Darboux agora substituímos a medida de Lebesgue pela $P - medida$ e então obtemos as somas de Darboux generalizadas

$$z = \sum_{v=1}^{n} m_v P(S_v)$$

e

$$Z = \sum_{v=1}^{n} M_v P(S_v),$$

onde, como anteriormente, m_v e M_v denotam respectivamente os limites inferior e superior de $g(x)$ em S_v.

O desenvolvimento seguinte é exatamente análogo ao desenvolvimento que levou à definição da integral de Lebesgue. O limite superior do conjunto de todos os possíveis valores z é chamado integral inferior de $g(x)$ sobre S com respeito à $P - medida$ dada, enquanto que o limite inferior do conjunto de todos os possíveis valores Z é a integral superior correspondente. Como anteriormente, temos que a integral inferior é no máximo igual à integral superior.

Se as integrais inferior e superior são iguais então $g(x)$ é dita integrável sobre S com respeito à $P - medida$ dada, e o valor comum das duas integrais é dito integral de Lebesgue-Stieltjes de $g(x)$ sobre S com respeito à $P - medida$ dada e é denotado por uma das duas expressões

$$\int_S g(x)dP(S) = \int_S g(x)dF(x).$$

Quando não houver risco de confusão, escreveremos simplesmente dP ou dF ao invés de $dP(S)$ ou $dF(x)$.

No caso particular em que $F(x) = x$ temos $P(S) = \lambda(S)$, e fica evidente que a definição da integral de Lebesgue-Stieltjes se reduz à definição da integral de Lebesgue. Assim afirmamos que a integral de Lebesgue-Stieltjes é obtida da integral de Lebesgue simplesmente substituindo, na definição desta última, a medida de Lebesgue pela $P - medida$.

Todas as propriedades da integral de Lebesgue já citadas são agora generalizadas para a integral de Lebesgue-Stieltjes. Para funções limitadas e conjuntos de Borel com $P - medida$ finita obtemos as seguintes generalizações das Proposições (1.1), (1.2), (1.3) e (1.4):

$$\int_S \left(g_1\left(x\right) + g_2\left(x\right)\right) dF = \int_S g_1\left(x\right) dF + \int_S g_2\left(x\right) dF, \tag{1.8}$$

$$\int_S cg\left(x\right) dF = c \int_S g\left(x\right) dF, \tag{1.9}$$

$$\int_{S_1 \cup S_2} g\left(x\right) dF = \int_{S_1} g\left(x\right) dF + \int_{S_2} g\left(x\right) dF, \tag{1.10}$$

$$mP\left(S\right) \leq \int_S g\left(x\right) dF \leq MP\left(S\right), \tag{1.11}$$

onde c é uma constante, m e M denotam os limites inferior e superior de $g(x)$ em S, e S_1 e S_2 são dois conjuntos de Borel em $I\!R$ sem pontos em comum.

As extensões da integral de Lebesgue para funções não limitadas e conjuntos de Borel com medida de Lebesgue infinita podem ser aplicadas, de maneira perfeitamente análoga, à integral de Lebesgue-Stieltjes.

Assim, uma função $g(x)$ não limitada é dita integrável com respeito a $P(S)$ (ou $F(x)$) sobre um conjunto de Borel S em $I\!R$ com $P-medida$ finita se o limite

$$\lim_{a\to-\infty,b\to+\infty}\int_S g_{a,b}(x)dP=\int_S g(x)dP=\int_S g(x)dF,$$

onde $g_{a,b}(x)$ é definida como em (1.5), existir e for finito.

Ainda mais, quando S possui $P-medida$ infinita, $g(x)$ é dita integrável com respeito a P (ou F) sobre S se $g(x)$ é integrável com respeito a P (ou F) sobre $S_{a,b}$ para todo a e todo b, onde $S_{a,b}$ é definido como na Subseção 1.2.2, e se o limite

$$\lim_{a\to-\infty,b\to+\infty}\int_{S_{a,b}}|g(x)|dP=\int_S|g(x)|dP=\int_S|g(x)|dF$$

existir e possuir valor finito. Neste caso, o limite

$$\lim_{a\to-\infty,b\to+\infty}\int_{S_{a,b}}g(x)dP=\int_S g(x)dP=\int_S g(x)dF$$

também existe e é finito, e dizemos que a integral de Lebesgue-Stieltjes de $g(x)$ com respeito a P (ou F) sobre o conjunto S é convergente e assume, por definição, o valor deste limite.

As propriedades (1.8), (1.9), (1.10) e (1.11) ainda são válidas quando estudamos funções não limitadas e conjuntos de Borel com $P-medida$ infinita, salvo que, neste último caso, devemos substituir a relação (1.11) por

$$\int_S g(x)dF\geq 0 \text{ se } g(x)\geq 0 \text{ para todo } x \text{ em } S.$$

1.5 Variáveis aleatórias

Nesta seção vamos proceder com a apresentação de algumas definições que serão úteis nas próximas seções e necessárias nesta para que possamos definir uma "variável aleatória".

Definição 37

- *Um par* $(S, \mathcal{A})$*, onde* S *é um conjunto e* $\mathcal{A}$ *é uma* σ*-álgebra de subconjuntos de* S*, é dito um "espaço mensurável".*
- *Se* P *é uma medida então a tripla* $(S, \mathcal{A}, P)$ *é dita um "espaço de medida".*
- *Se* $(S, \mathcal{A}, P)$ *é um espaço de medida e* P *é uma medida de probabilidade então a tripla* $(S, \mathcal{A}, P)$ *é dita um "espaço de probabilidade".*

Definição 38 *Suponhamos que* S *seja um conjunto com uma* σ*-álgebra* $\mathcal{A}$ *de subconjuntos e que* T *seja um outro conjunto com uma* σ*-álgebra* $\mathcal{C}$ *de subconjuntos. Suponhamos ainda que* $f : S \to T$ *seja uma função. Dizemos que* f *é "mensurável" se, para todo* $B \in \mathcal{C}$, $f^{-1}(B) \in \mathcal{A}$.

Teorema 39 *Uma função mensurável* f *de um espaço de medida* $(S_1, \mathcal{A}_1, \mu_1)$ *em um espaço mensurável* $(S_2, \mathcal{A}_2)$, $f : S_1 \to S_2$*, induz uma medida no conjunto* S_2*. Para cada* $A \in \mathcal{A}_2$*, define-se* $\mu_2(A) = \mu_1(f^{-1}(A))$*. Integrais com respeito a* μ_2 *podem ser escritas como integrais com respeito a* μ_1 *da seguinte maneira:*
Se $g : S_2 \to I\!R$ *for integrável, então*

$$\int g(y) d\mu_2(y) = \int g(f(x)) d\mu_1(x).$$

Definição 40 *A medida μ_2 definida no teorema anterior é dita a "medida induzida em $(S_2, \mathcal{A}_2)$ por f a partir de μ_1".*

Compreendemos agora alguns conceitos que se fazem necessários para se enunciar a definição que segue.

Definição 41 *Se $(S, \mathcal{A}, P)$ é um espaço de probabilidade, $(\chi, \mathcal{B})$ um espaço mensurável e $X : S \to \chi$ uma função mensurável, então X é dita uma "quantidade aleatória". Se $\chi = I\!R$ e $\mathcal{B}$ é a σ-álgebra de Borel então X é dita uma "variável aleatória". Seja P_X a medida de probabilidade induzida em $(\chi, \mathcal{B})$ por X a partir de P. Esta medida de probabilidade é dita a "distribuição de X". A distribuição de X é dita ser "discreta"se existir um conjunto contável $A \subseteq \chi$ tal que $P_X(A) = 1$.*

Quando induzimos uma medida de probabilidade em um espaço mensurável $(\chi, \mathcal{B})$ por uma quantidade aleatória, denotamos esta medida de probabilidade por Pr. Assim, por exemplo, na última definição da distribuição de uma quantidade aleatória X, se $B \in \mathcal{B}$ então $Pr(X \in B) = P(X^{-1}(B)) = P_X(B)$.

Proposição 42 *Se X é uma variável aleatória então a função $F_X(x) = Pr(X \leq x) = P_X(B)$, onde $B = (-\infty, x]$, é a função de distribuição univocamente associada à distribuição de X. Neste caso F_X é chamada a "função de distribuição de X" ou "função de distribuição acumulada de X".*

1.6 Continuidade absoluta e função de densidade

Definição 43 *Sejam μ_1 e μ_2 duas medidas em um mesmo espaço $(S, \mathcal{A})$. Suponhamos que, para todo $A \in \mathcal{A}$, $\mu_1(A) = 0$ implica que $\mu_2(A) = 0$. Então dizemos que μ_2 é absolutamente contínua com respeito a μ_1 e denotamos esta relação por $\mu_2 \ll \mu_1$. Quando $\mu_2 \ll \mu_1$, dizemos que μ_1 é uma medida dominante para μ_2.*

Um caso de particular interesse para o nosso estudo ocorre quando, na última definição, o espaço considerado é $(\chi, \mathcal{B})$, onde $\chi = I\!R$ e $\mathcal{B}$ é a σ-álgebra de Borel, e a medida dominante considerada é a medida de Lebesgue λ. Consideremos uma variável aleatória X com distribuição P_X e função de distribuição F_X, de modo que P_X seja uma medida em $(\chi, \mathcal{B})$. Desta forma $P_X \ll \lambda$ se, para qualquer $B \in \mathcal{B}$, $\lambda(B) = 0$ tivermos que $P_X(B) = 0$. Sabemos que $\lambda(B) = 0$, por exemplo, se $B = \emptyset$ ou se considerarmos B como sendo qualquer ponto $x \in I\!R$, ou um subconjunto dos racionais.

Definição 44 *Seja $(S, \mathcal{A}, \mu)$ um espaço de medida. Se existir uma partição enumerável do conjunto S tal que cada elemento da partição possui medida μ finita então dizemos que μ é $\sigma - finita$.*

Definição 45 *Consideremos o espaço de medida $(S, \mathcal{A}, \mu)$. Se E for alguma afirmação à respeito dos pontos de S e μ for uma medida em S, dizemos que E é verdadeira "em quase toda parte com respeito a μ" se o conjunto dos pontos de S para os quais E não é verdadeira estiver contido em um conjunto A com $\mu(A) = 0$.*

Exemplo 46 *É sabido que uma função não-decrescente pode ter no máximo um* *número enumerável de descontinuidades. Como conjuntos enumeráveis têm medida de Lebesgue igual a* 0 *dizemos que funções não-decrescentes são contínuas em quase toda parte com respeito à medida de Lebesgue.*

Teorema 47 *Sejam* μ_1 *e* μ_2 *duas medidas no espaço mensurável* $(S, \mathcal{A})$ *tais que* $\mu_2 \ll \mu_1$ *e* μ_1 *é* σ*-finita. Então existe uma função mensurável* $f : S \to [0, \infty]$ *tal que, para todo* $A \in \mathcal{A}$*,*

$$\mu_2(A) = \int_A f(x) d\mu_1(x).$$

Ainda, se $g : S \to I\!R$ *é integrável com respeito a* μ_2 *então*

$$\int g(x) d\mu_2(x) = \int g(x) f(x) d\mu_1(x).$$

A função f *é chamada "derivada de Radon-Nikodym de* μ_2 *com respeito a* μ_1*" e é única em quase toda parte com respeito a* μ_1*. Esta derivada é algumas vezes denotada por* $\frac{d\mu_2}{d\mu_1}(s)$*. Se* μ_2 *é* σ*-finita então* f *é finita em quase toda parte com respeito a* μ_1*.*

Este último teorema é conhecido como $teorema\ de\ Radon - Nikodym$.

Definição 48 *Sejam* $(S, \mathcal{A}, P)$ *um espaço de probabilidade e* $(\chi, \mathcal{B}, \nu)$ *um espaço de medida. Suponhamos que a função* $X : S \to \chi$ *seja mensurável. Seja* P_X *a medida induzida em* $(\chi, \mathcal{B})$ *por* X *a partir de* P*. Suponhamos ainda que* $P_X \ll \nu$*. Então a derivada de Radon-Nikodym* $f_X = \frac{dP_X}{d\nu}$ *é dita a "densidade de* X *com respeito a* ν*".*

Em particular, nos interessa o caso em que, na definição anterior, X é uma variável aleatória e a medida ν é a medida de Lebesgue denotada por λ. Neste caso, como vimos, a medida de probabilidade P_X é dita a distribuição de X e se associa univocamente à função de distribuição F_X de X. Devido a esta relação unívoca entre P_X e F_X podemos escrever

$$f_X(x) = \frac{dF_X}{dx}(x) \text{ para todo } x \in \chi = I\!R$$

e chamamos $f_X(x)$ de *densidade de X com respeito a* λ ou simplesmente *densidade de X*.

Proposição 49 *Sejam $(S, \mathcal{A}, P)$ um espaço de probabilidade, $(\chi, \mathcal{B}, \nu)$ um espaço de medida e $X : S \to \chi$ uma função mensurável. Seja ainda P_X a medida de probabilidade induzida em $(\chi, \mathcal{B})$ por X a partir de P e suponhamos que $P_X \ll \nu$. Se $h : X \to I\!R$ é mensurável e f_X é a densidade de X com respeito a ν então*

$$\int h(x)dF_X(x) = \int h(x)f_X(x)d\nu(x).$$

Com respeito a esta última proposição nos interessa, como anteriormente, o caso em que X é uma variável aleatória e $\nu = \lambda$, onde λ é a medida de Lebesgue. Desta forma, se consideramos ainda $h(x)$ identicamente igual a 1 sobre $I\!R$ então podemos escrever

$$P_X(A) = \int_A dF_X(x) = \int_A f_X(x)d\lambda(x) = \int_A f_X(x)dx$$

para todo A pertencente à σ-álgebra de Borel $\mathcal{B}$ em $I\!R$.

Exemplo 50 *Sejam X_1 e X_2 duas variáveis aleatórias e sejam P_{X_1} e P_{X_2} suas respectivas distribuições. Suponhamos que X_1 seja distribuída uniformemente no* *intervalo* $[0,10]$ *e que* X_2 *seja distribuída uniformemente no intervalo* $[0,5]$*, ou seja,* $X_1 \sim \mathcal{U}(0,10)$ *e* $X_2 \sim \mathcal{U}(0,5)$*. Então*

$$F_{X_1}(x) = \begin{cases} 0 & se \;\; x < 0 \\ \frac{x}{10} & se \;\; 0 \leq x \leq 10 \\ 1 & se \;\; x > 10 \end{cases} \tag{1.12}$$

e

$$F_{X_2}(x) = \begin{cases} 0 & se \;\; x < 0 \\ \frac{x}{5} & se \;\; 0 \leq x \leq 5 \\ 1 & se \;\; x > 5. \end{cases} \tag{1.13}$$

A densidade de X_1 *é*

$$f_{X_1}(x) = \frac{dF_{X_1}}{dx}(x) = \frac{1}{10} I_{[0,10]}(x).$$

Sabemos que $P_{X_2} \ll P_{X_1}$ *e então escrevemos, para todo* A *pertencente à* σ*-álgebra de Borel* $\mathcal{B}$ *em* $\mathbb{R}$*,*

$$P_{X_2}(A) = \int_A f(x) dP_{X_1}(x),$$

onde f *é a derivada de Radon-Nikodym de* P_{X_2} *com respeito a* P_{X_1}*.*

Seja $A = [0,x]$ *para algum* $x \in (0,5)$*, então*

$$P_{X_2}([0,x]) = \int_0^x f(y) dP_{X_1}(y) = \int_0^x f(y) f_{X_1}(y) dy.$$

Mas $P_{X_2}([0,x]) = \frac{x}{5}$*, então*

$$\frac{x}{5} = \int_0^x f(y) \frac{1}{10} I_{[0,10]}(y) dy = \int_0^x f(y) \frac{1}{10} dy.$$

Derivando com respeito a x *temos*

$$\frac{1}{5} = \frac{1}{10} f(x).$$

Portanto a derivada de Radon-Nikodym de P_{X_2} com respeito a P_{X_1} é

$$f(x) = 2I_{(0,5)}(x).$$

Consideremos agora uma variável aleatória X cuja distribuição P_X é discreta, ou seja, existe um conjunto contável $A \subseteq I\!\!R$ tal que $P_X(A) = 1$. Neste caso, também dizemos que a variável aleatória X é discreta. Desta maneira observamos que para alguns valores $x \in I\!\!R$ contáveis temos que $P_X(x) > 0$. Portanto a distribuição de uma variável aleatória discreta é um exemplo de uma medida não absolutamente contínua com respeito à medida de Lebesgue λ, uma vez que $\lambda(x) = 0$ para todo $x \in I\!\!R$.

1.7 Convergência

É interessante ressaltar que, nesta Seção, a grande maioria dos resultados é enunciada em termos de quantidades aleatórias, porém, como nosso grande interesse a essa altura é tratar da convergência de variáveis aleatórias devido a freqüência com a qual falamos delas em termos práticos, e como sabemos que o conceito de uma variável aleatória constitui uma particularização do conceito de uma quantidade aleatória, então devemos notar que, algumas vezes neste texto, estes resultados enunciados em termos que quantidades aleatórias são naturalmente extendidos em termos de variáveis aleatórias.

Introduziremos agora o conceito de ”convergência fraca”entre distribuições de probabilidade e em seguida trataremos de três tipos de convergência que consideramos para seqüências de quantidades aleatórias: convergência quase certa, convergência em probabilidade e convergência em distribuição.

Definição 51 *Consideremos uma seqüência $\{F_n\}$ de funções de distribuição e uma outra função de distribuição F. Dizemos que F_n converge fracamente a F, e denotamos $F_n \Rightarrow F$, se*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

para todo ponto de continuidade de F.

Exemplo 52 *Sejam as funções de distribuição $F_n(x)$ e $F(x)$ definidas como*

$$F_n(x) = \begin{cases} 0 & se \;\; x < \frac{1}{n} \\ 1 & se \;\; x \geq \frac{1}{n} \end{cases}$$

e

$$F(x) = \begin{cases} 0 & se \;\; x < 0 \\ 1 & se \;\; x \geq 0 \end{cases}$$

Notemos que $\lim_{n\to\infty} F_n(x) = F(x)$ exceto no ponto $x = 0$, o único ponto de descontinuidade de $F(x)$. Portanto $F_n \Rightarrow F$.

Afim de aprofundar nosso entendimento a respeito do conceito de "convergência fraca" consideremos as medidas de probabilidade P_n e P associadas às funções de distribuição F_n e F, respectivamente. Como já vimos, estas medidas de probabilidade, definidas para conjuntos de Borel em $\mathbb{R}$, são univocamente determinadas a partir das relações

$$P_n\{(-\infty, x]\} = F_n(x)$$

e

$$P\{(-\infty, x]\} = F(x).$$

Sabemos que a função de distribuição F é contínua no ponto x se e somente se o conjunto $\{x\}$ possuir $P-medida$ igual a 0. Desta forma $F_n \Rightarrow F$ significa que a relação

$$\lim_{n\to\infty} P_n\{(-\infty, x]\} = P\{(-\infty, x]\} \quad se \quad P(\{x\}) = 0 \tag{1.14}$$

é válida para todo x.

Denotemos por ∂A a fronteira de um conjunto de Borel A em $I\!R$. Assim, ∂A é formado por aqueles pontos que constituem limites[4] de seqüências de pontos em A e também limites de seqüências de pontos fora de A. Por exemplo, a fronteira[5] do conjunto $(-\infty, x]$ é constituída somente pelo ponto x, e então a relação (1.14) pode ser reescrita como

$$\lim_{n\to\infty} P_n(A) = P(A) \quad se \quad P(\partial A) = 0, \tag{1.15}$$

onde $A = (-\infty, x]$.

É possível mostrar que $F_n \Rightarrow F$ se e somente se a relação (1.15) for verdadeira para todo conjunto de Borel A em $I\!R$.

Definição 53 *Um conjunto de Borel A em $I\!R$ para o qual $P(\partial A) = 0$ é dito um conjunto P-contínuo.*

[4]Assumimos que convergência de números reais é conhecida pelo leitor.

[5]Uma definição mais precisa da fronteira de um conjunto em $\boldsymbol{R}$ pode ser encontrada em [8].

Dizemos, portanto, que P_n converge fracamente a P, e denotamos $P_n \Rightarrow P$, se $\lim_{n\to\infty} P_n(A) = P(A)$ para cada conjunto P-contínuo, ou seja, se a relação (1.15) é válida. Assim, $P_n \Rightarrow P$ se e somente se as funções de distribuição correspondentes satisfazem $F_n \Rightarrow F$.

Definição 54 *Seja S um conjunto arbitrário. Uma função $d : S \times S \to I\!R$, que associa um número $d(x, y) \in I\!R$ a cada par ordenado $(x, y) \in S \times S$, de modo que as condições adiante sejam satisfeitas para quaisquer $x, y, z \in S$, é dita uma métrica em S:*

a) $d(x, x) = 0$;

b) *Se $x \neq y$ então $d(x, y) > 0$;*

c) $d(x, y) = d(y, x)$;

d) $d(x, z) \leq d(x, y) + d(y, z)$.

A quantidade $d(x, y)$ é também chamada distância de x a y.

Notemos que, nesta última definição, as condições a) e b) afirmam que $d(x, y) \geq 0$ e ainda $d(x, y) = 0$ se, e somente se, $x = y$. A condição c) afirma que a distância $d(x, y)$ é uma função simétrica das variáveis x e y. E a condição d) é chamada "desigualdade do triângulo", tendo origem no fato de que, no plano euclidiano, o comprimento de um dos lados de um triângulo não supera a soma dos comprimentos dos outros dois lados.

Definição 55 *Um par (S, d), onde S é um conjunto e d é uma métrica em S, é dito um espaço métrico.*

Costumamos nos referir a um espaço métrico citando apenas o conjunto no par (S, d) e deixando subentendida a métrica d considerada. Por exemplo, diríamos apenas "o espaço métrico S" ao invés de dizer "o espaço métrico (S, d)".

Exemplo 56 *O conjunto $\mathbb{R}$ dos números reais constitui um exemplo bastante importante de espaço métrico. A distância entre dois pontos $x, y \in \mathbb{R}$ é definida por*

$$d(x, y) = |x - y|.$$

As condições a), b), c) e d) resultam diretamente das propriedades elementares do valor absoluto de números reais. Esta distância é chamada métrica usual da reta.

Teorema 57 *Sejam S e $\mathcal{B}$ um espaço métrico arbitrário e a σ-álgebra de Borel de subconjuntos de S, respectivamente, e consideremos ainda as medidas de probabilidade P_n e P definidas em $\mathcal{B}$. Então $P_n \Rightarrow P$ se e somente se*

$$\lim_{n\to\infty} \int f dP_n = \int f dP \quad se \quad f \in C(S),$$

onde $C(S)$ denota a classe de funções reais contínuas e limitadas em S.

Definição 58 *Dizemos que uma seqüência $\{X_n\}_{n=1}^{\infty}$ de quantidades aleatórias converge em distribuição a quantidade aleatória X, e denotamos*

$$X_n \xrightarrow{\mathcal{D}} X,$$

se as medidas de probabilidade P_n das quantidades aleatórias X_n convergem fracamente a medida de probabilidade P de X, ou seja, se $P_n \Rightarrow P$.

Definição 59 *Se X for uma variável aleatória com função de distribuição F_X, então o "valor esperado" (ou média, ou esperança) de X é definido e denotado por*

$$E(X) = \int x dF_X(x).$$

Teorema 60 *Se $X : S \to \chi$ for uma quantidade aleatória e $f : \chi \to I\!R$ for uma função mensurável, então*

$$E[f(X)] = \int f(x) dF_X(x),$$

onde F_X é a função de distribuição de X.

Prova. Se tomarmos $Y = f(X)$, então Y induz uma medida (com função de distribuição F_Y) em $(I\!R, \mathcal{B})$ de acordo com o Teorema 39. A definição de $E(Y)$ é

$$\int y dF_Y(y),$$

e o Teorema 39 também afirma que

$$\int y dF_Y(y) = \int f(x) dF_X(x).$$

$*$

Devido aos Teoremas 57 e 60 podemos enunciar a definição seguinte, que é equivalente à Definição 58.

Definição 61 *Seja $\{X_n\}_{n=1}^{\infty}$ uma seqüência de quantidades aleatórias e seja X uma outra quantidade aleatória, todas assumindo valores em um mesmo espaço* *topológico χ, levando em conta a σ-álgebra gerada pelos subconjuntos abertos de χ, ou seja, a σ-álgebra de Borel. Se $\lim_{n\to\infty} E(f(X_n)) = E(f(X))$, para toda função $f : \chi \to I\!R$ contínua e limitada, então dizemos que X_n converge em distribuição a X.*

Definição 62 *Se $\{X_n\}_{n=1}^{\infty}$ e X forem quantidades aleatórias em um espaço métrico com métrica d, e se, para todo $\epsilon > 0$,*

$$\lim_{n\to\infty} Pr(d(X_n, X) > \epsilon) = 0,$$

então dizemos que X_n converge em probabilidade a X, e escrevemos

$$X_n \xrightarrow{P} X.$$

Definição 63 *Seja um conjunto A pertencente a σ-álgebra de Borel de subconjuntos de $I\!R$ e seja X uma variável aleatória. Dizemos que o conjunto A é X-contínuo se*

$$Pr(X \in \partial A) = 0,$$

onde ∂A denota a fronteira de A.

Teorema 64 *Se $X_n \xrightarrow{P} X$, então*

$$\lim_{n\to\infty} Pr([X_n \in A]\Delta[X \in A]) = 0 \tag{1.16}$$

para todo conjunto A X-contínuo, de modo que $B\Delta C$ denota a diferença simétrica entre os conjuntos B e C.

Como (1.16) implica que

$$\lim_{n\to\infty} Pr(X_n \in A) = Pr(X \in A),$$

então do Teorema 64 segue que

$$X_n \xrightarrow{P} X \quad implica \quad X_n \xrightarrow{D} X.$$

Definição 65 *Dizemos que uma seqüência* $\{X_n\}_{n=1}^{\infty}$ *converge quase certamente a* X*, e escrevemos*

$$X_n \to X, q.c.,$$

se

$$\{s : X_n(s) \text{ não converge a } X(s)\} \subseteq E \text{ com } Pr(E) = 0,$$

ou seja,

$$Pr(\lim_{n\to\infty} X_n = X) = 1.$$

Pode-se mostrar que, se uma seqüência $\{X_n\}_{n=1}^{\infty}$ de variáveis aleatórias convergir quase certamente a uma outra variável aleatória X, então esta seqüência também converge em probabilidade a X.

O próximo exemplo ilustra a situação em que ocorre convergência em probabilidade mas não ocorre convergência quase certa.

Exemplo 66 *Seja uma variável aleatória* U *uniformemente distribuída no intervalo* $[0, 1]$*, ou seja,* $U \sim \mathcal{U}[0, 1]$*. Definamos uma seqüência* $\{X_{nk}\}$*,* $n = 1, \ldots, \infty$ *e*

$k = 1, \ldots, n$, de variáveis aleatórias,

$$\begin{array}{lllll} X_{11} & & & & \\ X_{21} & X_{22} & & & \\ X_{31} & X_{32} & X_{33} & & \\ \ldots & & & & \\ X_{n1} & X_{n2} & X_{n3} & \ldots & X_{nn} \\ \ldots, & & & & \end{array}$$

de modo que

$$X_{nk} = \begin{cases} 1 & \textit{se } U \in (\frac{k-1}{n}, \frac{k}{n}], \\ 0 & \textit{caso contrário,} \end{cases}$$

Desta forma, para qualquer $k \in \{1, \ldots, n\}$ fixo, temos que a variável aleatória X_{nk} possui distribuição Bernouli com parâmetro $\frac{1}{n}$, $X_{nk} \sim ber(\frac{1}{n})$, ou seja, assume o valor 1 com probabilidade $\frac{1}{n}$ e o valor 0 com probabilidade $1 - \frac{1}{n}$.

Consideremos agora a variável aleatória degenerada X, que assume o valor 0 com probabilidade 1, ou seja, $Pr(X = 0) = 1$.

Então, para qualquer $k \in \{1, \ldots, n\}$ fixo e um $\varepsilon > 0$,

$$\lim_{n\to\infty} Pr(|X_{nk} - X| > \varepsilon) = \lim_{n\to\infty} Pr(X_{nk} = 1) = \lim_{n\to\infty} \frac{1}{n} = 0,$$

e portanto $X_{nk} \xrightarrow{P} X$.

Por outro lado, para qualquer valor possível u de U e qualquer $n_0 \in I\!N$, existem $n > n_0$, e $k \in \{1, \ldots, n\}$ tais que $u \in (\frac{k-1}{n}, \frac{k}{n}]$ e, portanto, $X_{nk}(u) = 1$.

Assim,

$$P(\{u \in (0, 1] : \lim_{n\to\infty} X_{nk}(u) = 0\}) = P(\{\emptyset\}) = 0.$$

Então X_{nk} não converge quase certamente a X.

Teorema 67 *Seja $\{X_n\}_{n=1}^{\infty}$ uma seqüência de variáveis aleatórias e seja X uma outra variável aleatória. Se $X_n \xrightarrow{P} X$ então existe uma subseqüência $\{n_k\}_{k=1}^{\infty}$ tal que*

$$\lim_{k\to\infty} X_{n_k} = X, \quad q.c.$$

A demonstração do Teorema 67 está fora do escopo desta monografia, mas uma ilustração da implicação nele apresentada é dada no exemplo seguinte.

Exemplo 68 *Consideremos a seqüência $\{X_{nk}\}$ de variáveis aleatórias definida no exemplo 66. Sabemos que $X_{nk} \xrightarrow{P} X$. Agora, se tomarmos $k = 1$ teremos uma subseqüência $\{X_{n1}\}_{n=1}^{\infty}$ de tal forma que*

$$P(\{u \in (0,1] : \lim_{n\to\infty} X_{n1}(u) = 0\}) = P((0,1]) = 1,$$

e portanto $X_{n1} \to X$ quase certamente.

Nos exemplos a seguir são apresentados alguns resultados interessantes no que diz respeito a convergência de variáveis aleatórias. Tais resultados utilizam os conceitos discutidos nesta seção.

Exemplo 69 *Sejam $\{X_n\}_{n=1}^{\infty}$ variáveis aleatórias independentes e identicamente distribuídas com média μ finita. Então a Lei Fraca dos Grandes Números, um dos teoremas limite mais úteis, afirma que a variável aleatória $\overline{X}_n$, definida por*

$$\overline{X}_n = \sum_{i=1}^{n} \frac{X_i}{n},$$

converge em probabilidade a μ.

Exemplo 70 *Sejam* $\{X_n\}_{n=1}^{\infty}$ *variáveis aleatórias independentes e identicamente distribuídas com média* μ *finita e variância* σ^2 *também finita. Então o Teorema Central do Limite, outro importante teorema limite, afirma que*

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\mathcal{D}} Z,$$

onde Z *é uma variável aleatória com distribuição Normal com média* 0 *e variância* σ^2*, ou seja,* $Z \sim N(0, \sigma^2)$*.*

Lembremos que a derivada de Radon-Nikodym da distribuição $N(0, \sigma^2)$ *com respeito à medida de Lebesgue nos subconjuntos de Borel em* $I\!R$ *é dada por*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}x^2}.$$

Exemplo 71 *A Lei Forte dos Grandes Números afirma que, se* $\{X_n\}_{n=1}^{\infty}$ *for uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com média* μ*, então a variável aleatória*

$$\overline{X}_n = \sum_{i=1}^{n} \frac{X_i}{n},$$

converge quase certamente a μ*, ou seja,*

$$Pr(\lim_{n\to\infty} \overline{X}_n = \mu) = 1.$$

Utilizaremos, mais adiante, o conceito de convergência em distribuição para determinar distribuições limite para o máximo e para a soma de variáveis aleatórias independentes e identicamente distribuídas.

1.8 Simulação

Muitas vezes pode ser de interesse gerar observações que possuam uma determinada distribuição. Tais observações serão chamadas "números pseudo-aleatórios", porque as amostras aparentam possuir as propriedades de variáveis aleatórias, mas são na verdade geradas a partir de um processo determinístico.

Apresentaremos alguns teoremas que tratam da geração de números pseudo-a-leatórios com uma determinada distribuição, sob a suposição de que números pseudo-aleatórios com distribuição uniforme no intervalo $[0, 1]$ ($\mathcal{U}(0, 1)$) possam ser gerados.

Definição 72 *Seja F uma função de distribuição. Definimos a inversa de F por*

$$F^{-1}(q) = \begin{cases} inf\{x : F(x) \geq q\} & se \;\; q > 0, \\ inf\{x : F(x) > 0\} & se \;\; q = 0. \end{cases} \tag{1.17}$$

para $0 \leq q \leq 1$ e sob a convenção que $inf\{\emptyset\} = +\infty$.

Teorema 73 *Se a variável aleatória U possuir distribuição $\mathcal{U}(0, 1)$, então $X = F^{-1}(U)$ possuirá função de distribuição F.*

Prova. Determinemos $Pr(X \leq t)$ para todo t. Primeiramente, consideremos t como um ponto de continuidade de F. Então

$$Pr(X \leq t) = Pr(F^{-1}(U) \leq t) = Pr(U \leq F(t)) = F(t), \tag{1.18}$$

onde a segunda igualdade segue do fato que, em um ponto de continuidade t, $X \leq t$ se e somente se $U \leq F(t)$, e a terceira igualdade segue do fato que U possui distribuição $\mathcal{U}(0, 1)$. Finalmente, consideremos t como um ponto de descontinuidade de F e tomemos

$$F(t) - \lim_{x \to t^-} F(x) = c.$$

Então $X = t$ se e somente se $F(t) - c < U \leq F(t)$, e assim

$$Pr(X = t) = Pr(F(t) - c < U \leq F(t)) = c.$$

Desta forma, X possui função de distribuição F nos pontos de continuidade de F e sua distribuição possui os mesmos saltos que F, nos mesmos pontos. Então, a função de distribuição de X é F.

*

Notemos que as três igualdades em (1.18) também são válidas se t for um ponto de descontinuidade de F, pois os eventos $[F^{-1}(U) \leq t]$ e $[U \leq F(t)]$ são iguais para todo $t \in \mathbb{R}$. Isto simplifica a prova do Teorema 73.

Mostremos que os eventos $[F^{-1}(U) \leq t]$ e $[U \leq F(t)]$ são iguais para todo $t \in \mathbb{R}$. Sabemos que F e F^{-1} são não decrescentes e, a partir da definição de F^{-1}, temos que

$$F(F^{-1}(U)) \geq U, \quad para \;\; 0 \leq U \leq 1$$

e

$$F^{-1}(F(t)) \leq t, \quad para \;\; t \in \mathbb{R}.$$

Portanto, se $F^{-1}(U) \leq t$ então $F(F^{-1}(U)) \leq F(t)$. Mas $F(F^{-1}(U)) \geq U$ e então $F(t) \geq U$.

Por outro lado, se $U \leq F(t)$ então $F^{-1}(U) \leq F^{-1}(F(t))$. Mas como $F^{-1}(F(t)) \leq t$ então $F^{-1}(U) \leq t$.

Desta forma, o Teorema 73 nos leva a gerar números pseudo-aleatórios com função de distribuição F arbitrária, se pudermos encontrar F^{-1}.

Exemplo 74 *Seja uma variável aleatória Y com distribuição exponencial indexada por um parâmetro λ conhecido, ou seja, $Y \sim exp(\lambda)$. Então a função de distribuição de Y é*

$$F(y) = \begin{cases} 1 - e^{-\lambda y} & se \;\; y \geq 0, \\ 0 & se \;\; y < 0. \end{cases} \tag{1.19}$$

Desta maneira encontramos

$$F^{-1}(u) = \frac{-\ln(1-u)}{\lambda} \;\; para \;\; 0 \leq u < 1.$$

Notemos que

$$\lim_{u \to 1} F^{-1}(u) = +\infty.$$

Portanto, se U possuir distribuição $\mathcal{U}(0,1)$, $X = F^{-1}(U) = \frac{-\ln(1-U)}{\lambda}$ possuirá distribuição $exp(\lambda)$.

O próximo teorema nos leva a encontrar números pseudo-aleatórios com densidade f arbitrária se pudermos gerar números pseudo-aleatórios com uma outra densidade g tal que $f(x) \leq kg(x)$ para algum k e todo x.

Teorema 75 *Seja f uma função integrável não-negativa, e seja g uma função de densidade. Tomemos $k > 0$ e suponhamos que $f(x) \leq kg(x)$ para todo x. Suponhamos ainda que as variáveis aleatórias $\{Y_i\}_{i=1}^{\infty}$ e $\{U_i\}_{i=1}^{\infty}$ sejam todas independentes, e que Y_i tenha densidade g e U_i tenha distribuição $\mathcal{U}(0,1)$. Definamos $Z = Y_N$,* *onde*

$$N = \min\{i : U_i \leq \frac{f(Y_i)}{kg(Y_i)}\}.$$

Então Z possui densidade proporcional a f.

O teorema seguinte descreve o método conhecido como Razão de Uniformes.

Teorema 76 *Seja $f : \mathbb{R} \to [0, \infty)$ uma função integrável. Definamos*

$$A = \{(u, v) \in \mathbb{R}^2 : 0 \leq u \leq \sqrt{f(\frac{v}{u})}\}.$$

Se (U, V) possuir distribuição uniforme sobre o conjunto A, então $\frac{V}{U}$ possuirá densidade proporcional a f.

Capítulo 2

Distribuição do Máximo

Iniciamos o presente capítulo com uma discussão a respeito da distribuição exata do máximo de uma seqüência finita de variáveis aleatórias independentes e identicamente distribuídas.

Na Seção 2.2, definimos e estudamos as Distribuições Max-Estáveis, e vemos que tais distribuições são necessariamente do mesmo tipo que uma das três distribuições chamadas Distribuições de Valor Extremo. Vemos ainda que as distribuições max-estáveis, e somente tais distribuições, constituem limites para a distribuição de uma certa normalização do máximo de uma seqüência de variáveis aleatórias i.i.d., construída por uma seqüência de constantes. Com estes resultados, identificamos a distribuição assintótica desta normalização do máximo como sendo uma distribuição do mesmo tipo que uma das distribuições de valor extremo.

2.1 Distribuição Exata

Uma n-upla de números reais, $(x_1, x_2, \ldots, x_n)$, pode ser reordenada de maneira crescente, de modo a se obter uma nova n-upla

$$(x_{(1)}, x_{(2)}, \ldots, x_{(n)}), \text{ onde } x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}.$$

Esta operação aplicada a todos os pontos do espaço $I\!R^n$ induz n funções, que serão denotadas por $X_{(1)}, \ldots, X_{(n)}$. Se introduzirmos uma estrutura de probabilidade em $I\!R^n$, estas funções podem ser interpretadas como variáveis aleatórias.

Dizemos, neste caso, que a seqüência $X_1, \ldots, X_n$ de variáveis aleatórias forma uma amostra aleatória.

Estamos interessados no caso particular em que $X_1, \ldots, X_n$ são independentes e identicamente distribuídas (i.i.d.), formando o que chamamos amostra aleatória simples, ou amostra.

A variável $X_{(k)}$, definida acima, é chamada k-ésima estatística de ordem da amostra $X_1, \ldots, X_n$. Em particular, $X_{(1)}$ e $X_{(n)}$ são os extremos amostrais, ou mais especificamente, o mínimo e o máximo amostrais, respectivamente.

Seja $\{X_n\}_{n=1}^{\infty}$ uma seqüência infinita de variáveis aleatórias i.i.d., com função de distribuição F e função de densidade f.

Para todo $n \in I\!N$, consideremos a variável aleatória $X_{(n)}$ definida como

$$X_{(n)} = \max\{X_1, \ldots, X_n\},$$

ou seja, $X_{(n)}$ é o maior valor na amostra aleatória $X_1, \ldots, X_n$.

Determinemos a função de distribuição G_n e a função de densidade g_n desta variável aleatória $X_{(n)}$. Seja $-\infty < x < \infty$, então

$$\begin{aligned} G_n(x) &= Pr(X_{(n)} \leq x) = Pr(\max\{X_1, \dots, X_n\} \leq x) \\ &= Pr(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= Pr(X_1 \leq x) Pr(X_2 \leq x) \dots Pr(X_n \leq x) \\ &= F(x)F(x)\dots F(x) \\ &= [F(x)]^n. \end{aligned}$$

Assim, $G_n(x) = [F(x)]^n$. A terceira igualdade se deve à independência assumida para a seqüência.

A função de densidade de $X_{(n)}$, denotada por g_n, pode ser determinada a partir da diferenciação de G_n, nos pontos $x \in I\!R$ em que esta derivada existir,

$$g_n(x) = n[F(x)]^{n-1} f(x).$$

Temos ainda que, se a seqüência $\{X_n\}_{n=1}^{\infty}$ for multiplicada por uma constante positiva, então $X_{(n)}$ também será multiplicada pela mesma constante. Se uma constante qualquer for somada a esta seqüência, então $X_{(n)}$ terá seu valor acrescido da mesma quantidade. Observemos que, dado um valor $x \in I\!R$, a probabilidade de que $(X_{(n)} \leq x)$ diminui à medida que o tamanho da amostra aumenta. Assim, tanto os quantis[1] quanto a média de $X_{(n)}$ aumentam com n.

[1]Uma definição precisa a respeito dos quantis de uma variável aleatória pode ser encontrada em DeGroot [4].

2.2 Distribuição Assintótica

Suponhamos novamente que $\{X_n\}_{n=1}^{\infty}$ seja uma seqüência infinita de variáveis aleatórias i.i.d. com função de distribuição comum $F(x)$. Já sabemos que, dado $n \in I\!N$, $F^n(x) = [F(x)]^n$ é a função de distribuição de $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

Se tomarmos

$$x_0 = sup\{x : F(x) < 1\} \leq \infty,$$

então pode-se mostrar que

$$\lim_{n \to \infty} \uparrow X_{(n)} = x_0 \quad \text{quase certamente.} \tag{2.1}$$

A fim de proceder à prova deste último resultado, necessitamos do teorema a seguir, que trata da convergência de uma seqüência não decrescente de variáveis aleatórias.

Teorema 77 *Sejam $\{X_n\}_{n=1}^{\infty}$ uma seqüência não decrescente de variáveis aleatórias e X uma outra variável aleatória. Então*

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \to X, \quad q.c..$$

Mostremos que $\lim_{n \to \infty} \uparrow X_{(n)} = x_0$ quase certamente.

Observemos que, para $x < x_0$, temos $F(x) < 1$ e assim

$$Pr(X_{(n)} \leq x) = F^n(x) \to 0. \tag{2.2}$$

Por outro lado, para todo $x < x_0$,

$$Pr(X_{(n)} \leq x) = Pr(X_{(n)} - x_0 \leq x - x_0),$$

onde $X_{(n)} \leq x_0$ e $x - x_0 < 0$. Então,

$$Pr(X_{(n)} - x_0 \leq x - x_0) = Pr(|X_{(n)} - x_0| \geq |x - x_0|). \tag{2.3}$$

Seja $\varepsilon > 0$ e seja $x < x_0$ tal que $|x - x_0| > \varepsilon$. Então, de (2.2) e (2.3),

$$Pr(|X_{(n)} - x_0| > \varepsilon) \rightarrow 0,$$

quando $n \rightarrow \infty$.
Portanto $X_{(n)} \xrightarrow{P} x_0$. Como $\{X_{(n)}\}$ é uma seqüência não decrescente, então convergência em probabilidade implica em convergência quase certa, como queríamos demonstrar.

Para ilustrar a situação descrita pela relação (2.1) consideremos uma seqüência $\{X_n\}_{n=1}^{\infty}$ de variáveis aleatórias independentes com distribuição $\mathcal{U}(0,1)$. Neste caso $x_0 = 1$ e a função de densidade do máximo $X_{(n)}$ é dada por

$$f_{X_{(n)}}(x) = \begin{cases} nx^{n-1} & \text{se} \quad x \in [0,1], \\ 0 & \text{se} \quad x \notin [0,1]. \end{cases}$$

Na figura seguinte apresentamos os gráficos de $f_{X_{(n)}}(x)$ para $n = 50, 100, 1000$.

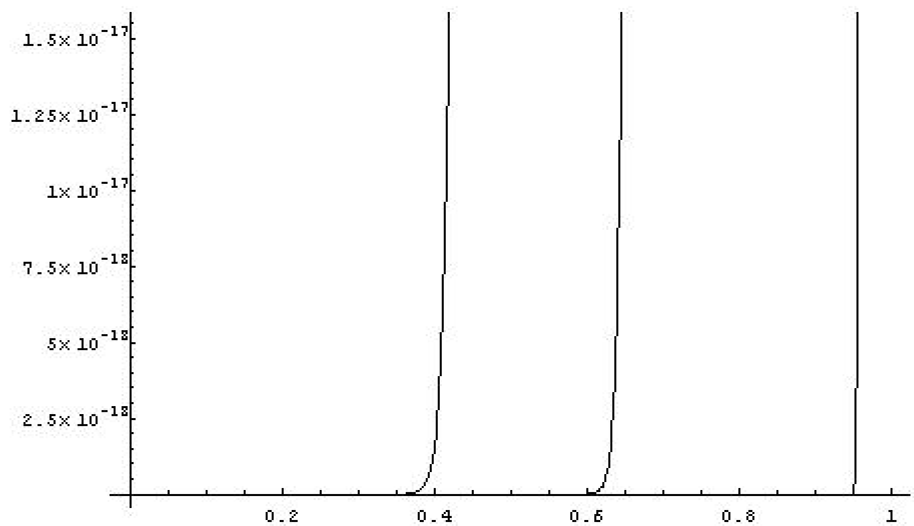

Figura 2.1: Função de densidade de $X_{(n)}$, para $n = 50, 100, 1000$ da esquerda para a direita.

Podemos observar que esta função de densidade tende a se concentrar em $x_0 = 1$.

Fica claro, a partir da relação (2.1), que uma distribuição limite não degenerada para $X_{(n)}$ não existirá a menos que normalizemos tal variável.

Mais precisamente, queremos estabelecer condições para $F(x)$ e para as seqüências $a_n > 0$, b_n tais que a distribuição limite de $a_n(X_{(n)} - b_n)$ seja uma distribuição G não degenerada:

$$Pr[a_n(X_{(n)} - b_n) \leq x] \rightarrow G(x) \tag{2.4}$$

para todo ponto de continuidade de G.

Em particular, estaremos interessados em determinar quais funções de distribuição G podem aparecer como o limite em (2.4). Veremos que estas funções de distribuição formam uma classe de distribuições.

Observemos que (2.4) pode ser escrita como

$$F^n(a_n^{-1}x + b_n) \to G(x), \tag{2.5}$$

para todo $x \in I\!R$, ponto de continuidade de G.

Definição 78 *Se (2.5) for válida para duas seqüências $\{a_n > 0\}$ e $\{b_n\}$, então dizemos que F pertence ao domínio de atração (para o máximo) de G e escrevemos $F \in D(G)$.*

2.2.1 Distribuições Max-Estáveis

Agora identificaremos as funções de distribuição G que são possíveis distribuições limite em (2.4) como a classe das Distribuições Max-Estáveis. Dizemos que uma função de distribuição G não degenerada é max-estável se, para cada $n = 2, 3, \ldots,$ existirem constantes $a_n > 0$ e b_n tais que

$$G^n(a_n x + b_n) = G(x), \quad para \;\; todo \;\; x \in I\!R.$$

Teorema 79 *(i) Uma função de distribuição G não degenerada é max-estável se e somente se existir uma seqüência F_n de funções de distribuição e constantes $a_n > 0$ e b_n tais que*

$$F_n(a_{nk}^{-1}x + b_{nk}) \to G^{1/k}(x) \quad quando \; n \to \infty, \tag{2.6}$$

para cada $k = 1, 2, \ldots,$ e para todo $x \in I\!R$ ponto de continuidade de G.

(ii) Em particular, se G é não degenerada, $D(G)$ é não vazio se e somente se G for max-estável. Então também $G \in D(G)$.

De acordo com este resultado, a classe de funções de distribuição G não degeneradas que aparecem como limites em (2.4) coincide com a classe de funções de distribuição max-estáveis.

A partir deste momento será necessário e conveniente utilizar a palavra *tipo* em um sentido técnico para este contexto. Segue portanto a próxima definição.

Definição 80 *Dizemos que duas funções de distribuição G_1 e G_2 são "do mesmo tipo"se*

$$G_2(x) = G_1(ax + b)$$

para duas constantes $a > 0$ e b.

Assim, a definição de distribuições max-estáveis pode ser reescrita como: "Uma função de distribuição G não degenerada é max-estável se, para cada $n = 2, 3, \ldots$, a função de distribuição G^n for do mesmo tipo que G."

Teorema 81 (Khintchine) *Seja $\{F_n\}$ uma seqüência de funções de distribuição e G uma função de distribuição não degenerada. Sejam $a_n > 0$ e b_n constantes tais que*

$$F_n(a_n x + b_n) \to G(x),$$

para todo $x \in I\!R$, ponto de continuidade de G. Então, para alguma função de distribuição não degenerada G_* e constantes $\alpha_n > 0$ e β_n, temos que

$$F_n(\alpha_n x + \beta_n) \to G_*(x)$$

se e somente se

$$a_n^{-1}\alpha_n \to a \quad e \quad a_n^{-1}(\beta_n - b_n) \to b$$

para algum $a > 0$ e algum b, e então

$$G_*(x) = G(ax + b).$$

O Teorema 81 mostra que se $\{F_n\}$ é uma seqüência de funções de distribuição com $F_n(a_n x + b_n) \to G_1(x)$ e $F_n(\alpha_n x + \beta_n) \to G_2(x)$ ($a_n > 0$ e $\alpha_n > 0$), então G_1 e G_2 são do mesmo tipo, desde que ambas sejam não degeneradas. Claramente, as funções de distribuição podem ser divididas em classes de equivalência (as quais chamamos *tipos*) dizendo que G_1 e G_2 são equivalentes se $G_2(x) = G_1(ax+b)$ para constantes $a > 0$ e b.

2.2.2 O Teorema dos Três Tipos

Mostraremos agora que uma função de distribuição é max-estável se e somente se ela for do mesmo tipo que uma das três possíveis distribuições chamadas Distribuições de Valor Extremo.

Uma função de distribuição do mesmo tipo que $G(x) = exp(-e^{-x})$ será dita do *Tipo I* ou Gumbel. Similarmente, diremos que uma função de distribuição é do *Tipo II* ou Fréchet (ou *Tipo III* ou Weibull negativa) se esta possuir a forma $G(ax + b)$, onde G é do *Tipo II* (ou *Tipo III*) listados no próximo teorema. Podemos nos referir ao *Tipo II* como *Tipo Cauchy* e ao *Tipo III* como *Tipo Limitada*.

Teorema 82 *Toda distribuição max-estável é do tipo de valor extremo, ou seja, é igual a $G(ax+b)$ para algum $a>0$ e algum b, onde para*

$$TipoI : G(x) = exp(-e^{-x}), \ \ -\infty < x < \infty; \tag{2.7}$$

$$TipoII : G(x) = \begin{cases} 0, & se \ \ x \leq 0, \\ exp(-x^{-\alpha}), & se \ \ x > 0; \end{cases} \tag{2.8}$$

$$TipoIII : G(x) = \begin{cases} exp(-(-x)^{\alpha}), & se \ \ x \leq 0, \\ 1, & se \ \ x > 0; \end{cases} \tag{2.9}$$

onde $\alpha > 0$ nos Tipos II e III.

Ainda, cada distribuição de valor extremo é max-estável.

Teorema 83 (Teorema dos Três Tipos) *Seja $X_{(n)} = \max\{X_1, \ldots, X_n\}$, onde X_i são variáveis aleatórias i.i.d.. Se para algumas constantes $a_n > 0$ e b_n tivermos*

$$Pr[a_n(X_{(n)} - b_n) \leq x] \to G(x) \tag{2.10}$$

para alguma função de distribuição G não degenerada, então G é do mesmo tipo que uma das três distribuições de valor extremo listadas no Teorema 82. Inversamente, cada função de distribuição G de valor extremo pode aparecer como um limite em (2.10), e de fato aparece quando G for a própria função de distribuição de cada X_i.

Prova. Se a relação (2.10) for verdadeira, então, segundo o Teorema 79, a função de distribuição G é max-estável e, a partir do Teorema 82, do mesmo tipo que uma das distribuições de valor extremo listadas. Inversamente, se G for uma distribuição de valor extremo, o Teorema 82 afirma que G é max-estável, e então, segundo o Teorema 79(ii), $G \in D(G)$.

$*$

O teorema 83 é também conhecido como Teorema de Fisher-Tippett, veja [13].

Teorema 84 *Seja* $\{X_n\}$ *uma seqüência de variáveis aleatórias i.i.d. com função de distribuição* F. *Seja* $0 \leq \tau \leq \infty$ *e suponhamos que* $\{u_n\}$ *seja uma seqüência de números reais tal que*

$$n(1 - F(u_n)) \to \tau \qquad \text{quando } n \to \infty. \tag{2.11}$$

Então

$$Pr(X_{(n)} \leq u_n) \to e^{-\tau} \qquad \text{quando } n \to \infty. \tag{2.12}$$

Inversamente, se (2.12) for válida para algum τ, $0 \leq \tau \leq \infty$, *então (2.11) também é válida.*

Exemplo 85 *Consideremos uma variável aleatória com distribuição exponencial com parâmetro unitário,* $X \sim exp(1)$. *Então*

$$F(x) = 1 - e^{-x}, \qquad x \geq 0.$$

Dado $\tau > 0$, *podemos escolher* u_n *tal que*

$$1 - F(u_n) = \frac{\tau}{n}$$

simplesmente tomando

$$u_n = -\log(\frac{\tau}{n}) = -\log\tau + \log n.$$

Então pelo Teorema 84,

$$Pr(X_{(n)} \leq -\log\tau + \log n) \rightarrow e^{-\tau}.$$

Tomando $\tau = e^{-x}$ *obtemos*

$$Pr[a_n(X_{(n)} - b_n) \leq x] \rightarrow G(x),$$

com

$$a_n = 1, \qquad b_n = \log n, \qquad G(x) = exp\{-e^{-x}\}.$$

Ou seja, $G(x)$ *é do Tipo I.*

Concluímos então que a variável aleatória $X_{(n)} - \log n$ *possui, assintoticamente, função de distribuição* $G(x) = exp\{-e^{-x}\}$*, que é do Tipo I.*

Teorema 86 *Dada uma função de distribuição* F*, seja* $x_F = \sup\{x; F(x) < 1\}$*. Condições necessárias e suficientes para que a função de distribuição* F *das variáveis aleatórias i.i.d. da seqüência* $\{X_n\}$ *pertença ao domínio de atração de cada um dos três tipos são (em ordem crescente de complexidade):*

Tipo II:

$$x_F = \infty \quad e \quad \lim_{t\rightarrow\infty} \frac{(1 - F(tx))}{(1 - F(t))} = x^{-\alpha}, \quad \alpha > 0, \quad para \quad cada \quad x > 0;$$

Tipo III:

$$x_F < \infty \quad e \quad \lim_{h\rightarrow 0^+} \frac{(1 - F(x_F - xh))}{(1 - F(x_F - h))} = x^{\alpha}, \quad \alpha > 0, \quad para \quad cada \quad x > 0;$$

Tipo I:

Existe alguma função $g(t)$ estritamente positiva tal que

$$\lim_{t \to x_F^-} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}, \quad para \;\; todo \;\; x \in I\!R.$$

Exemplo 87 *Consideremos a função de distribuição de uma variável aleatória com distribuição Pareto,*

$$F(x) = 1 - Kx^{-\alpha}, \quad \alpha > 0, \quad K > 0, \quad x \geq K^{1/\alpha}.$$

Então

$$\frac{(1 - F(tx))}{(1 - F(t))} = x^{-\alpha},$$

para cada $x > 0$, quando t é suficientemente grande, de modo que o Teorema 86 mostra que um limite do Tipo II se aplica. Ainda, para $u_n = (Kn/\tau)^{1/\alpha}$ temos

$$1 - F(u_n) = \frac{\tau}{n},$$

de modo que, pelo Teorema 84,

$$Pr[X_{(n)} \leq (\frac{Kn}{\tau})^{1/\alpha}] \to e^{-\tau}.$$

Tomando $\tau = x^{-\alpha}$, para $x \geq 0$, obtemos

$$Pr[(Kn)^{-1/\alpha} X_{(n)} \leq x] \to exp\{-x^{-\alpha}\},$$

e então um limite do Tipo II ocorre com

$$a_n = (Kn)^{-1/\alpha}, \qquad b_n = 0.$$

Portanto, a variável aleatória $(Kn)^{-1/\alpha}X_{(n)}$ *possui função de distribuição assintótica* $G(x) = exp\{-x^{-\alpha}\}$*.*

Ilustremos a convergência no caso particular em que $K = \alpha = 1$*, de modo que a função de densidade de* $\frac{X_{(n)}}{n}$ *é dada por*

$$f_n(x) = \begin{cases} [1 - (nx)^{-1}]^{n-1}x^{-2} & se \quad x \in [\frac{1}{n}, \infty), \\ 0 & se \quad x \notin [\frac{1}{n}, \infty). \end{cases}$$

Apresentamos os gráficos da função de densidade $f_n(x)$*, para* $n = 3, 5, 20$*, na figura 2.2.*

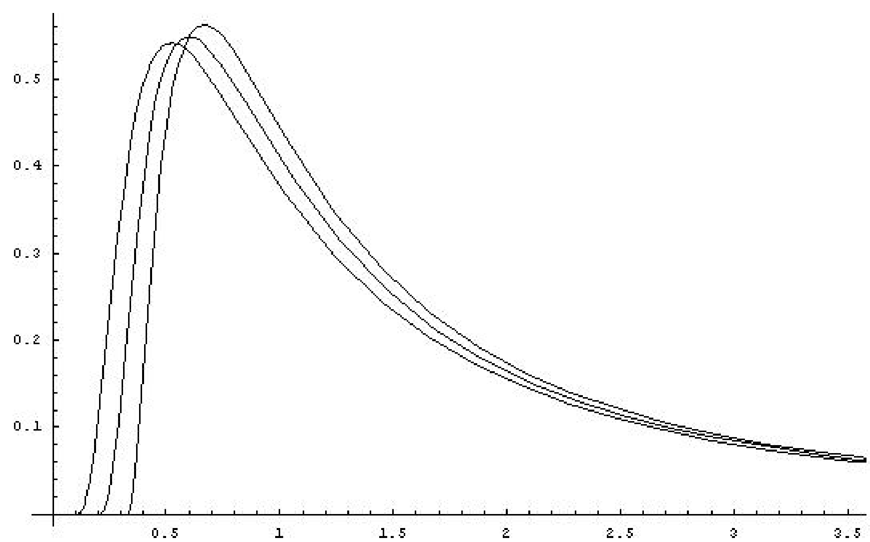

Figura 2.2: Função de densidade de $\frac{X_{(n)}}{n}$, para $n = 3, 5, 20$ da direita para a esquerda.

Neste caso, à medida que n *cresce, a função* $f_n(x)$ *se aproxima da função de densidade* $g(x) = G'(x) = exp\{-x^{-1}\}x^{-2}$*, definida em* $[0, \infty)$*.*

Exemplo 88 *Para uma variável aleatória com distribuição uniforme no intervalo* $[0,1]$ *temos*

$$F(x) = x, \qquad para \;\; 0 \leq x \leq 1.$$

Para $\tau > 0$ *e* $u_n = 1 - \tau/n$, *temos*

$$1 - F(u_n) = \frac{\tau}{n}, \qquad para \;\; n \geq \tau,$$

de modo que o Teorema 84 fornece

$$Pr[X_{(n)} \leq 1 - \frac{\tau}{n}] \to e^{-\tau}.$$

Assim, se tomamos $x < 0$ *e* $\tau = -x$ *obtemos*

$$Pr[n(X_{(n)} - 1) \leq x] \to e^{x},$$

o que nos leva a um limite do Tipo III com $\alpha = 1$ *e ainda*

$$a_n = n \qquad e \qquad b_n = 1.$$

Desta maneira, a variável $n(X_{(n)} - 1)$ *possui, assintoticamente, função de distribuição* $G(x) = e^{x}$.

Para ilustrar esta convergência apresentamos, para $n = 10, 20, 50$, *na figura 2.3, os gráficos da função de densidade de* $n(X_{(n)} - 1)$, *dada por*

$$f_n(x) = \begin{cases} \left(\frac{x}{n} + 1\right)^{n-1} & se \;\; x \in [-n, 0], \\ 0 & se \;\; x \notin [-n, 0]. \end{cases}$$

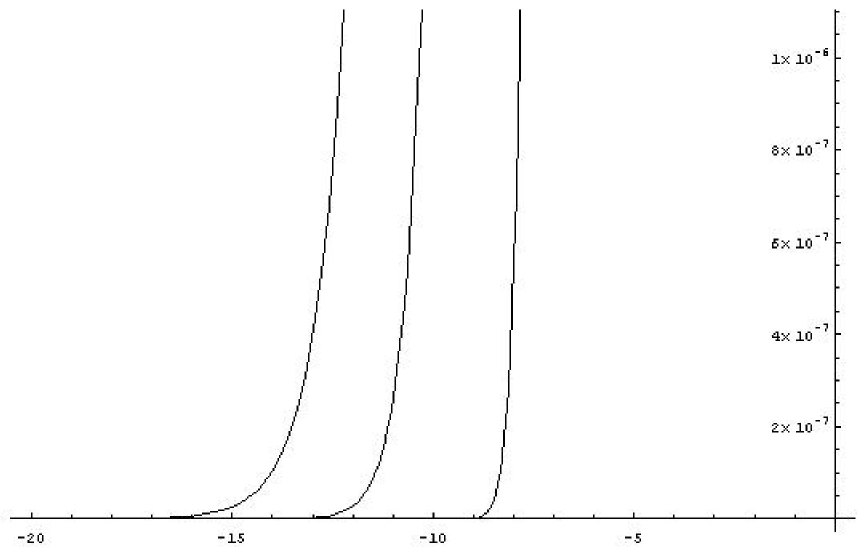

Figura 2.3: Função de densidade de $n(X_{(n)} - 1)$, para $n = 10, 20, 50$ da direita para a esquerda.

Podemos observar que, à medida que n cresce, a função de densidade $f_n(x)$ se aproxima de $g(x) = G'(x) = e^x$, definida no intervalo $(-\infty, 0]$.

Capítulo 3

Distribuição da Soma

Este capítulo se destina ao estudo da Distribuição da Soma de uma seqüência de variáveis aleatórias independentes e identicamente distribuídas.

Na Seção 3.1, introduzimos o conceito de Convolução entre funções de distribuição, de modo a caracterizar a Distribuição Exata da Soma de variáveis aleatórias i.i.d..

O Teorema Central do Limite é apresentado na Seção 3.2, e a partir dele identificamos a distribuição Normal como um limite para a distribuição de uma certa normalização da soma de variáveis aleatórias i.i.d..

As Distribuições Estáveis são estudas na Seção 3.3, e vemos que estas distribuições, e somente estas, constituem limites para a distribuição da soma normalizada, generalizando a discussão da Seção 3.2.

3.1 Convolução

Nos concentraremos agora acerca do estudo de uma ‘convolução’, basicamente em dois sentidos: a princípio como uma operação entre uma função de distribuição e uma função contínua, e posteriormente como uma operação entre funções de distribuição.

Sem perda de generalidade consideraremos apenas distribuições no espaço unidimensional $I\!R$.

Consideremos, assim, uma função de distribuição F e uma função pontual limitada φ (mais adiante veremos que φ poderá ser simplesmente uma função contínua ou uma função de distribuição). Definamos a função u como:

$$u(x) = \int_{-\infty}^{+\infty} \varphi(x-y)dF(y), \quad x \in I\!R. \tag{3.1}$$

Se F possuir uma função de densidade f, então a relação (3.1) se reduz a

$$u(x) = \int_{-\infty}^{+\infty} \varphi(x-y)f(y)dy. \tag{3.2}$$

Definição 89 *A convolução de uma função φ com uma função de distribuição F é a função definida por (3.1) e a denotamos por $u = F \star \varphi$. Se F possuir uma função de densidade f escrevemos, alternativamente, $u = f * \varphi$.*

Exemplo 90 *Tomemos F como a função de distribuição de uma variável aleatória X com distribuição uniforme no intervalo $[0, a]$, tal que*

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{a} I_{[0,a]}(x), \quad x \in \mathbb{R},$$

é sua função de densidade. Então, para uma função real φ,

$$u(x) = \int_{-\infty}^{+\infty} \varphi(x-y) f(y) dy = \int_{-\infty}^{+\infty} \varphi(x-y) \frac{1}{a} I_{[0,a]}(y) dy,$$

e fazendo $x - y = s$,

$$u(x) = \frac{1}{a} \int_{x-a}^{x} \varphi(s) ds$$

é a convolução de F com a função pontual φ.

Consideremos duas variáveis aleatórias X e Y independentes com funções de densidade $f(x)$ e $g(y)$, respectivamente. Então podemos escrever a função de densidade conjunta de X,Y, $f_{XY}(x, y)$, como

$$f_{XY}(x, y) = f(x)g(y), \quad \textit{para todo } (x, y) \in \mathbb{R}^2.$$

Esta relação nos permite determinar a medida de probabilidade sobre um conjunto de Borel em $\mathbb{R}^2$, digamos $A \subset \mathbb{R}^2$, como

$$P(A) = \int\int_A f(x)g(y) dx dy. \tag{3.3}$$

Se consideramos a variável aleatória $S = X + Y$, para X e Y independentes, então o evento $A = \{S \leq s\}$ é representado pelo semi-plano de pontos (x, y) tais que $x + y \leq s$. Denotemos a função de distribuição de Y por G, de modo que $g(y) = G'(y)$, onde esta derivada existir. Então, para obtermos a função de distribuição de $X+Y$ basta integrarmos em (3.3) sobre a região $A=\{(x,y)\in I\!R^2 : y\leq s-x\}$, de modo que

$$Pr(X+Y\leq s)=\int_{-\infty}^{+\infty}\int_{-\infty}^{s-x} g(y)f(x)dydx=\int_{-\infty}^{+\infty} G(s-x)f(x)dx.$$

Acabamos de mostrar o teorema seguinte.

Teorema 91 *Sejam X e Y variáveis aleatórias independentes com funções de distribuição F e G, respectivamente. Então*

$$Pr(X+Y\leq t)=\int_{-\infty}^{+\infty} G(t-x)dF(x). \tag{3.4}$$

Notemos que este último teorema nos fornece uma interpretação para $F\star\varphi$ quando φ é uma função de distribuição.

Por razões de simetria as posições ocupadas por F e G em (3.4) podem ser intercambiadas sem afetar o resultado. Assim,

$$Pr(X+Y\leq t)=\int_{-\infty}^{+\infty} G(t-x)dF(x)=\int_{-\infty}^{+\infty} F(t-y)dG(y).$$

Diferenciando a relação (3.4) observamos que a função de densidade de $X+Y$ é dada por uma das duas integrais na igualdade

$$\int_{-\infty}^{+\infty} f(t-y)g(y)dy=\int_{-\infty}^{+\infty} f(y)g(t-y)dy. \tag{3.5}$$

Definição 92 *A convolução de duas funções de densidade f e g é a função definida na relação (3.5). Denotamos tal convolução por $f*g$.*

Isto dá passo para o teorema seguinte.

Teorema 93 *Se as funções de distribuição F e G possuem funções de densidade f e g, respectivamente, então a convolução $F \star G$ possui função de densidade $h = f * g$ dada por*

$$h(x) = \int_{-\infty}^{+\infty} f(x-y)g(y)dy.$$

Ainda, de acordo com (3.5) temos $f * g = g * f$.

Dada uma terceira função de densidade u podemos escrever $(f * g) * u$ e esta é a função de densidade de uma soma $X + Y + Z$ de três variáveis aleatórias independentes com densidades f, g, u. Devido ao fato da soma de variáveis aleatórias ser comutativa e associativa, temos estas mesmas propriedades para convoluções, e então $f * g * u$ independe da ordem das operações.

É interessante notar que quando trabalhamos com variáveis aleatórias positivas, ou seja, quando nos referimos a funções de densidade concentradas no intervalo $[0, \infty)$, então a convolução, digamos $f * g$, se reduz a

$$f * g(s) = \int_0^s f(s-y)g(y)dy = \int_0^s f(x)g(s-x)dx.$$

Exemplo 94 *Sejam f e g duas funções de densidade concentradas em $[0, \infty)$ e definidas por*

$$f(x) = \alpha e^{-\alpha x} I_{[0,\infty)}(x) \qquad e \qquad g(x) = \beta e^{-\beta x} I_{[0,\infty)}(x).$$

Então, por definição,

$$f * g(x) \;=\; \int_{-\infty}^{\infty} f(x-y)g(y)dy$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \alpha e^{-\alpha(x-y)} I_{[0,\infty)}(x-y)\beta e^{-\beta y} I_{[0,\infty)}(y)dy \\
&= \int_{0}^{x} \alpha e^{-\alpha(x-y)}\beta e^{-\beta y}dy \\
&= \alpha\beta e^{-\alpha x}\int_{0}^{x} e^{y(\alpha-\beta)}dy.
\end{aligned}
$$

Portanto

$$f * g(x) = \frac{\alpha\beta}{\beta-\alpha}(e^{-\alpha x} - e^{-\beta x}), \qquad x > 0.$$

Freqüentemente uma soma $S_n = X_1 + X_2 + \ldots + X_n$ de n variáveis aleatórias mutuamente independentes com uma função de distribuição F comum pode ser de interesse. Neste caso, a distribuição de S_n é chamada convolução n-variada de F e é denotada por $F^{n\star}$. Então

$$F^{1\star} = F \qquad e \qquad F^{(n+1)\star} = F^{n\star} \star F.$$

Uma soma sem termos é convencionalmente interpretada como 0, e por consintência definimos $F^{0\star}$ como a distribuição degenerada concentrada na origem.

Se F possui função de densidade f então $F^{n\star}$ possui densidade $f * f * \ldots * f$ (n vezes) e denotamos esta densidade por f^{n*}.

Desta maneira podemos escrever

$$f^{n*}(x) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x - x_1 - \ldots - x_{n-1}) f(x_1) \ldots f(x_{n-1}) dx_1 \ldots dx_{n-1}. \quad (3.6)$$

Apresentaremos agora a definição da função geradora de momentos e então enunciaremos alguns teoremas relacionados a esta função. Tal discussão nos ajudará a determinar a distribuição da soma de variáveis aleatórias independentes de uma forma relativamente simples.

Definição 95 *Seja X uma variável aleatória. A esperança da variável aleatória e^{tX} é definida como a função geradora de momentos de X, se tal esperança existir para todo valor de t tal que $-h < t < h$, $h > 0$. Assim, a função geradora de momentos, denotada por $m_X(t)$ ou $m(t)$, é dada por*

$$m_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

se a variável aleatória X for contínua com função de densidade f, ou

$$m_X(t) = E(e^{tX}) = \sum_x e^{tx} Pr(X = x)$$

se a variável aleatória X for discreta.

O seguinte resultado estabelece uma relação biunívoca entre as funções de distribuição e as funções geradoras de momento.

Teorema 96 *Sejam X e Y duas variáveis aleatórias com funções de densidade f_X e f_Y, respectivamente. Suponhamos que as funções geradoras de momentos $m_X(t)$ e $m_Y(t)$ ambas existam e sejam iguais para todo $t \in (-h, h)$, para algum $h > 0$. Então as funções de distribuição das variáveis aleatórias X e Y, F_X e F_Y, são iguais.*

Teorema 97 *Se X e Y forem duas variáveis aleatórias independentes, e g_1 e g_2 forem duas funções reais, então*

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)].$$

Prova. Mostremos que $E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$ para duas variáveis aleatórias absolutamente contínuas e independentes, X e Y, com funções de densidade f_X e f_Y, respectivamente. A demonstração para variáveis aleatórias discretas ocorre de maneira análoga.

Sabemos que

$$E[g_1(X)g_2(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y)f_{XY}(x,y)dxdy.$$

Mas como X e Y são independentes, então

$$\begin{aligned} E[g_1(X)g_2(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y)f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} g_1(x)f_X(x)dx \int_{-\infty}^{\infty} g_2(y)f_Y(y)dy \\ &= E[g_1(X)]E[g_2(Y)]. \end{aligned}$$

$*$

Teorema 98 *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes, tais que cada uma de suas respectivas funções geradoras de momentos existam para todo $t \in (-h,h)$, para algum $h > 0$. Seja também a variável aleatória $S_n = \sum_{i=1}^n X_i$. Então a função geradora de momentos de S_n é dada por*

$$m_{S_n}(t) = E[exp(t\sum X_i)] = \prod_{i=1}^n m_{X_i}(t) \quad para \quad -h < t < h.$$

Prova. Mostremos que $m_{S_n}(t) = \prod_{i=1}^n m_{X_i}(t)$.

Temos que

$$m_{S_n}(t) = E[exp(t\sum X_i)] = E[\prod_{i=1}^n e^{tX_i}].$$

Mas, de acordo com o Teorema 97,

$$E[\prod_{i=1}^{n} e^{tX_i}] = \prod_{i=1}^{n} E[e^{tX_i}],$$

e então

$$m_{S_n}(t) = \prod_{i=1}^{n} E[e^{tX_i}] = \prod_{i=1}^{n} m_{X_i}(t).$$

*

A utilidade do Teorema 98 torna-se evidente quando recordamos o Teorema 96, o qual afirma que uma função geradora de momentos, quando existir, determinará a função de distribuição. Assim, se pudermos reconhecer $\prod_{i=1}^{n} m_{X_i}(t)$ como uma função geradora de momentos correspondente a uma distribuição em particular, então teremos encontrado a distribuição de $S_n = \sum_{i=1}^{n} X_i$. Os exemplos adiante ilustram este procedimento.

Exemplo 99 *Suponhamos que $X_1, \ldots, X_n$ sejam variáveis aleatórias independentes com distribuição Bernoulli indexada por um parâmetro p, $0 < p < 1$, isto é,*

$$Pr(X_i = 1) = p \quad e \quad Pr(X_i = 0) = 1 - p = q, \quad \forall i \in \{1, \ldots, n\}.$$

Desta forma

$$m_{X_i}(t) = E(e^{tX_i}) = \sum_{x=0}^{1} e^{tx} Pr(X_i = x) = q + pe^t,$$

e então

$$m_{S_n}(t) = \prod_{i=1}^{n} (q + pe^t) = (q + pe^t)^n.$$

Agora, se considerarmos uma variável aleatória $Y \sim Bin(n,p)$*, de modo que*

$$Pr(Y=y) = \left(\begin{array}{c} n \\ y \end{array} \right) p^y q^{n-y}, \quad y \in \{0,1,\ldots,n\},$$

então

$$m_Y(t) = \sum_{y=0}^{n} e^{ty} \left(\begin{array}{c} n \\ y \end{array} \right) p^y q^{n-y} = (q+pe^t)^n.$$

Do teorema 96, podemos concluir que a variável aleatória $S_n = \sum_{i=1}^{n} X_i$*, soma das* n *variáveis aleatórias independentes com distribuição Bernoulli com parâmetro* p*, possui distribuição Binomial com parâmetros* n *e* p*, ou seja,* $S_n \sim Bin(n,p)$*.*

Exemplo 100 *Seja* $\{X_i\}_{i=1}^{n}$ *uma seqüência de variáveis aleatórias independentes com distribuição Geométrica indexada por um parâmetro* p*,* $0 < p \le 1$*, ou seja,*

$$Pr(X_i = x) = p(1-p)^x = pq^x, \quad x \in \{0,1,\ldots\}, \quad \forall i \in \{1,\ldots,n\}.$$

Então temos

$$m_{X_i}(t) = E(e^{tX_i}) = \sum_{x=0}^{\infty} e^{tx} p q^x = p \sum_{x=0}^{\infty} (e^t q)^x = p \left(\frac{1}{1-e^t q} \right),$$

de modo que

$$m_{S_n}(t) = \prod_{i=1}^{n} \left(\frac{p}{1-e^t q} \right) = \left(\frac{p}{1-e^t q} \right)^n.$$

Por outro lado, se considerarmos uma variável aleatória Y *com distribuição Binomial Negativa indexada por parâmetros* n *e* p*, tal que*

$$Pr(Y=y) = \left(\begin{array}{c} n+y-1 \\ y \end{array} \right) p^n q^y, \quad y \in \{0,1,\ldots\},$$

então

$$m_Y(t) = \sum_{y=0}^{\infty} e^{ty} \begin{pmatrix} n+y-1 \\ y \end{pmatrix} p^n q^y = \left(\frac{p}{1-e^t q}\right)^n.$$

Concluímos então que a soma S_n de n variáveis aleatórias independentes com distribuição Geométrica com parâmetro p possui distribuição Binomial Negativa com parâmetros n e p, ou seja, $S_n \sim BNeg(n,p)$.

Exemplo 101 *Suponhamos que $\{X_i\}_{i=1}^n$ seja uma seqüência de variáveis aleatórias independentes tal que cada X_i possui distribuição Poisson com parâmetro $\lambda_i > 0$, ou seja,*

$$Pr(X_i = x) = \frac{e^{-\lambda_i}\lambda_i^x}{x!}, \quad x \in \{0, 1, \ldots\}.$$

Então

$$m_{X_i}(t) = E(e^{tX_i}) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda_i}\lambda_i^x}{x!} = exp\{\lambda_i(e^t - 1)\},$$

e portanto

$$m_{\sum X_i}(t) = \prod_{i=1}^{n} exp\{\lambda_i(e^t-1)\} = exp\{\sum_{i=1}^{n}\lambda_i(e^t-1)\} = exp\{(e^t-1)\sum_{i=1}^{n}\lambda_i\},$$

que pode ser reconhecida como a função geradora de momentos de uma variável aleatória com distribuição Poisson com parâmetro $\sum_{i=1}^n \lambda_i$. Portanto, a soma de variáveis aleatórias independentes com distribuição Poisson é novamente uma variável aleatória com distribuição Poisson, com parâmetro igual à soma dos parâmetros individuais.

Exemplo 102 *Seja $\{X_i\}_{i=1}^n$ uma seqüência de variáveis aleatórias independentes com distribuição Exponencial indexada por um parâmetro $\lambda > 0$, de modo que,* *para qualquer* $i \in \{1, \ldots, n\}$*, a função de densidade de* X_i *é dada por*

$$f_{X_i}(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x).$$

Desta maneira temos, para $t < \lambda$*,*

$$m_{X_i}(t) = E(e^{tX_i}) = \int_{-\infty}^{\infty} e^{tx} \lambda e^{-\lambda x} I_{[0,\infty)}(x) dx = \int_{0}^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t},$$

de modo que, para $S_n = \sum_{i=1}^{n} X_i$*,*

$$m_{S_n}(t) = \prod_{i=1}^{n} \frac{\lambda}{\lambda - t} = \left(\frac{\lambda}{\lambda - t} \right)^n.$$

Consideremos agora uma variável aleatória Y *com distribuição Gamma indexada pelos parâmetros* n *e* λ*. A função de densidade de* Y *é dada por*

$$f_Y(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} I_{[0,\infty)}(x),$$

e sua função geradora de momentos é

$$m_Y(t) = E(e^{tY}) = \int_{-\infty}^{\infty} e^{tx} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} I_{[0,\infty)}(x) dx =$$

$$\left(\frac{\lambda}{\lambda - t} \right)^n \int_0^{\infty} \frac{(\lambda - t)^n}{\Gamma(n)} x^{n-1} e^{-(\lambda - t)x} dx = \left(\frac{\lambda}{\lambda - t} \right)^n,$$

para $t < \lambda$*.*

Portanto, a variável aleatória S_n*, que constitui a soma das* n *variáveis aleatórias independentes e identicamente distribuídas* $X_i \sim Exp(\lambda)$*, possui distribuição Gamma com parâmetros* n *e* λ*. Notemos ainda que a distribuição Exponencial constitui uma particularização da distribuição Gamma ao tomarmos, nesta última, o parâmetro* $n = 1$*.*

Exemplo 103 *Consideremos as variáveis aleatórias* $X_1, \ldots, X_n$ *independentes tais que, para cada* $i \in \{1, \ldots, n\}$, $X_i \sim N(\mu_i, \sigma_i^2)$, *com* $\mu_i \in I\!R$ *e* $\sigma_i > 0$. *Assim, a função de densidade de* X_i *é dada por*

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}\sigma_i} exp\left\{\frac{-(x-\mu_i)^2}{2\sigma_i^2}\right\}, \quad x \in I\!R.$$

A função geradora de momentos de X_i *é*

$$m_{X_i}(t) = exp\left(\mu_i t + \frac{1}{2}\sigma_i^2 t^2\right),$$

e então a função geradora de momentos de $S_n = \sum_{i=1}^n X_i$ *é*

$$m_{S_n}(t) = \prod_{i=1}^n exp\left(\mu_i t + \frac{1}{2}\sigma_i^2 t^2\right) = exp\left(t\sum_{i=1}^n \mu_i + \frac{1}{2}t^2 \sum_{i=1}^n \sigma_i^2\right),$$

que pode ser reconhecida como a função geradora de momentos de uma variável aleatória normalmente distribuída com parâmetros $\sum_{i=1}^n \mu_i$ *e* $\sum_{i=1}^n \sigma_i^2$.
Portanto

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

No exemplo adiante determinamos a distribuição da soma entre uma variável aleatória com distribuição Bernoulli e uma variável aleatória normalmente distribuída.

Exemplo 104 *Sejam*

$$Y \sim Ber(p) \quad e \quad Z \sim N(\mu, \sigma^2).$$

A função de distribuição de $X = Y + Z$ é dada por

$$F_X(x) = Pr[X \leq x] = Pr[(Y = 0) \cap (Y + Z \leq x)] + Pr[(Y = 1) \cap (Y + Z \leq x)]$$

$$= Pr[Y + Z \leq x | Y = 0]Pr[Y = 0] + Pr[Y + Z \leq x | Y = 1]Pr[Y = 1]$$

$$= Pr[Z \leq x](1 - p) + Pr[Z + 1 \leq x]p$$

$$= F_Z(x)(1 - p) + F_{Z+1}(x)p,$$

onde $Z + 1 \sim N(\mu + 1, \sigma^2)$.

Derivando a última relação com respeito a x obtemos a função de densidade de X,

$$f_X(x) = f_Z(x)(1 - p) + f_{Z+1}(x)p, \qquad x \in I\!R,$$

que é dita uma mistura entre as funções de densidade de Z e $Z + 1$.

3.2 O Teorema Central do Limite

Consideremos, para cada $n \in I\!N$, uma seqüência

$$X_{n1}, \ldots, X_{nk_n}$$

de variáveis aleatórias independentes com média igual a 0 e variância σ_{nk}^2 finita com $k = 1, \ldots, k_n$, ou seja, para cada n temos uma seqüência diferente, com k_n termos.

Consideremos ainda a variável aleatória

$$S_n = X_{n1} + \ldots + X_{nk_n}$$

e suponhamos que sua variância

$$s_n^2 = \sigma_{n1}^2 + \ldots + \sigma_{nk_n}^2$$

seja positiva.

Denotemos por Z uma variável aleatória normalmente distribuída com média igual a 0 e variância igual a 1, ou seja, $Z \sim N(0,1)$. Com a notação anterior, enunciamos o seguinte teorema.

Teorema 105 (Teorema de Lindeberg) *Seja f_{nk} a função de densidade da variável aleatória X_{nk}. Se, dado $\varepsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon s_n\}} x^2 f_{nk}(x)dx = 0, \tag{3.7}$$

então

$$\frac{S_n}{s_n} \xrightarrow{\mathcal{D}} Z. \tag{3.8}$$

A condição (3.7) controla as variâncias σ_{nk}^2, à medida que n aumenta. Em particular, se considerarmos $k_n = n$ e $X_{nk} = X_k$ de modo que $\{X_k\}_{k=1}^n$ seja uma seqüência de variáveis aleatórias i.i.d. com média igual a 0, variância $\sigma^2 > 0$ finita e função de densidade f, então a soma em (3.7) será menor ou igual do que a integral de x_1^2/σ^2 sobre a região $\{|x_1| \geq \varepsilon\sigma\sqrt{n}\}$, ou seja,

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int_{\{|x| \geq \varepsilon s_n\}} x^2 f(x)dx \quad \leq \quad \int_{\{|x| \geq \varepsilon\sigma\sqrt{n}\}} \frac{x^2}{\sigma^2} f(x)dx \quad < \quad 1, \quad \forall n \in I\!N.$$

Portanto

$$\lim_{n\to\infty} \int_{\{|x|\geq \varepsilon\sigma\sqrt{n}\}} \frac{x^2}{\sigma^2} f(x)dx = 0,$$

satisfazendo a condição (3.7). Assim,

$$\frac{S_n}{s_n} = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} X_k \quad \xrightarrow{\mathcal{D}} \quad Z.$$

Acabamos de provar, a partir do Teorema de Lindeberg, o Teorema de Lindeberg-Lévy, também conhecido como *Teorema Central do Limite*:

Teorema 106 (Teorema Central do Limite) *Se $X_1, X_2, \ldots$ são variáveis aleatórias independentes e identicamente distribuídas com média igual a 0 e variância $\sigma^2 > 0$ finita, então*

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} X_k \quad \xrightarrow{D} \quad Z.$$

Se, por outro lado, a seqüência $\{X_k\}_{k=1}^{\infty}$ de variáveis aleatótias i.i.d. possuir média $\mu \neq 0$ finita e variância $\sigma^2 > 0$ também finita, então construímos uma nova seqüência $\{Y_k = X_k - \mu\}_{k=1}^{\infty}$ com média igual a 0 de modo que o teorema também se aplica, resultando

$$\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} Y_k = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu) \quad \xrightarrow{D} \quad Z.$$

3.3 As Distribuições Estáveis em $I\!R$

Antes de procedermos com a descrição das distribuições estáveis, necessitamos introduzir a notação

$$U \stackrel{d}{=} V$$

para indicar que as variáveis aleatórias U e V possuem a mesma distribuição. Assim,

$$U \stackrel{d}{=} aV + b$$

significa que as funções de distribuição de U e V diferem somente por uma reescalação linear do argumento, ou seja, são do mesmo tipo.

Consideremos agora uma seqüência $\{X_k\}_{k=1}^{\infty}$ de variáveis aleatórias independentes e identicamente distribuídas com função de distribuição F e ainda uma outra variável aleatória X independente desta seqüência, também com função de distribuição F. Denotemos por S_n a soma

$$S_n = X_1 + \ldots + X_n.$$

Definição 107 *A função de distribuição F é dita estável no sentido amplo se, para cada n, existirem constantes $c_n > 0$ e γ_n tais que*

$$S_n \stackrel{d}{=} c_n X + \gamma_n \tag{3.9}$$

e F não for concentrada na origem.

Ainda, F é dita estável no sentido estrito se (3.9) for válida para $\gamma_n = 0$.

Teorema 108 *Com respeito à Definição 107, somente são possíveis constantes normalizadoras c_n tais que*

$$c_n = n^{\frac{1}{\alpha}}.$$

A constante α é chamada 'expoente característico' da função de distribuição F e F pode também ser chamada α-estável. Pode-se mostrar que $0 < \alpha \leq 2$, mas isto foge ao escopo deste trabalho. A prova pode ser encontrada em [3].

Exemplo 109 *A distribuição Normal com média* 0 *e varância* σ^2*,* $N(0, \sigma^2)$*, caracterizada pela função de densidade*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \; e^{-\frac{x^2}{2\sigma^2}},$$

é estritamente estável com $c_n = \sqrt{n}$*, ou seja,* $\alpha = 2$*.*

Exemplo 110 *Seja* F *a função de distribuição*

$$F(x) = 2\left[1 - \phi\left(\frac{1}{\sqrt{x}}\right)\right], \quad x > 0,$$

onde ϕ *denota a função de distribuição da distribuição* $N(0, 1)$*. A função de densidade associada a* F *é dada por*

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi x^3}} \; e^{\frac{-1}{2x}}, & x > 0, \\ 0 & x \leq 0, \end{cases}$$

e esta distribuição é estritamente estável com $c_n = n^2$*, isto é,* $\alpha = \frac{1}{2}$*.*

Exemplo 111 *A distribuição Cauchy com parâmetros $\alpha \in \mathbb{R}$ e $\beta > 0$, caracterizada pela função de densidade*

$$f(x) = \frac{1}{\pi}\left(\frac{\beta}{\beta^2 + (x-\alpha)^2}\right),$$

é estável no sentido estrito com expoente característico $\alpha = 1$.

Lema 112 *Sejam F e G duas funções de distribuição. Se a e b são pontos de descontinuidade de F e G, respectivamente, então o ponto $a+b$ é um ponto de descontinuidade de $F \star G$. Ainda mais, todo ponto de descontinuidade de $F \star G$ é desta forma.*

Lema 113 *Toda função de distribuição estável no sentido amplo é contínua.*

Prova. Mostraremos que se uma função de distribuição não for contínua, então ela não é estável no sentido amplo.

Suponhamos que uma função de distribuição F possui um ou mais pontos de descontinuidade e denotemos por p o máximo dos valores dos saltos associados a estes pontos. Para distribuições estáveis, os pontos de descontinuidade de $F \star F$ diferem dos pontos de descontinuidade de F apenas pela locação, mas não por seus saltos. Assim, $F \star F$ possui um ponto de descontinuidade com um salto de tamanho p, o que contradiz o fato de que a convolução $F \star F$ deve possuir saltos estritamente menores do que p.

$*$

Utilizaremos o lema 112 para mostrar que a convolução $F \star F$ deve possuir saltos estritamente menores do que p.

Prova. Sejam X_1 e X_2 duas variáveis aleatórias independentes não degeneradas[1] com função de distribuição F. Denotemos por $a_1, a_2, \ldots,$ os pontos de descontinuidade de F com respectivos saltos $p_1, p_2, \ldots,$ e denotemos ainda por p o máximo de $p_1, p_2, \ldots.$ Mostraremos que a convolução $F \star F$ possui saltos estritamente menores do que p.

De acordo com o lema 112, um ponto de descontinuidade de $F \star F$ deve ser da forma $a_i + a_j,\ i, j \in I\!N$. Sabemos que o salto em $F \star F$ no ponto $a_i + a_j$ possui amplitude igual a

$$Pr(X_1 + X_2 = a_i + a_j) = Pr([X_1 = a_i] \cap [X_2 = a_j]), \quad \forall i, j \in I\!N.$$

Mas como X_1 e X_2 são independentes, então

$$Pr([X_1 = a_i] \cap [X_2 = a_j]) = Pr(X_1 = a_i)Pr(X_2 = a_j) = p_i p_j.$$

Como $0 < p_k < 1$, $\forall k \in I\!N$, então

$$p_i p_j < \max\{p_i, p_j\} \leq p, \quad \forall i, j \in I\!N.$$

Portanto

$$Pr(X_1 + X_2 = a_i + a_j) < p, \quad \forall i, j \in I\!N.$$

$*$

Teorema 114 *Se F é estritamente estável com expoente característico α, então, para constantes positivas arbitrárias s e t,*

$$s^{1/\alpha} X_1 + t^{1/\alpha} X_2 \stackrel{d}{=} (s + t)^{1/\alpha} X, \tag{3.10}$$

[1]Não concentram toda a massa em um único ponto.

onde X, X_1 e X_2 são variáveis aleatórias independentes com função de distribuição F comum.

De uma maneira geral, (3.10) implica que todas as combinações lineares da forma $a_1X_1 + a_2X_2$ são do mesmo tipo.

Devemos salientar que a restrição para distribuições estritamente estáveis é menos séria do que pode parecer, pois, de acordo com o teorema a seguir, toda distribuição estável com expoente característico $\alpha \neq 1$ pode ser centrada de modo a se tornar estritamente estável.

Teorema 115 *Se F é estável com expoente $\alpha \neq 1$, então existe uma constante b tal que $F(x+b)$ é estritamente estável.*

Consideremos a seqüência $X_1, \ldots, X_n$ de variáveis aleatórias independentes com função de distribuição comum F, agora possuindo média 0 e variância 1. Novamente denotemos $S_n = X_1 + \ldots + X_n$. O Teorema Central do Limite afirma que a variável aleatória $n^{-\frac{1}{2}}S_n$ converge em distribuição à variável aleatória Z com distribuição $N(0,1)$. Já para distribuições sem variância, as constantes normalizadoras para S_n podem ser encontradas de maneira diferente, mas um limite ainda pode existir. O interessante é que todas as distribuições estáveis, e somente estas, ocorrem como tais limites. A definição adiante nos ajudará nesta discussão.

Definição 116 *A função de distribuição F das variáveis aleatórias X_k, $k \in \mathbb{N}$, pertence ao domínio de atração de uma função de distribuição G se existirem constantes normalizadoras $a_n > 0$ e b_n tais que a função de distribuição de*

$$a_n^{-1}(S_n - b_n)$$

converge à função de distribuição G.

Podemos agora reformular nossa última afirmação dizendo que "uma função de distribuição G possui um domínio de atração se, e somente se, G for estável".

Finalmente fica caracterizada a importância das distribuições estáveis para o nosso contexto, uma vez que estas distribuições ocorrem como os limites das distribuições das somas S_n normalizadas.

Definição 117 *Uma função de distribuição F é dita infinitamente divisível se, para todo n, existir uma função de distribuição F_n tal que $F = F_n^{n\star}$. Ou seja, F é infinitamente divisível se, e somente se, para cada n, existirem variáveis aleatórias i.i.d., $X_{1,n}, \ldots, X_{n,n}$, com distribuição F_n tais que*

$$S_n = X_{1,n} + \ldots + X_{n,n}$$

tem distribuição F.

As distribuições estáveis são infinitamente divisíveis e se distinguem pelo fato de que F_n difere de F somente por uma reescalação linear do argumento.

Um estudo mais aprofundado das distribuições infinitamente divisíveis e sua relação com as distribuições α-estáveis pode ser encontrado em Feller, vol.II ([3]), capítulo 6.

Capítulo 4

Algumas distribuições de cauda pesada

Neste capítulo definimos as Distribuições Subexponenciais e estudamos algumas de suas propriedades. Estudamos ainda as Distribuições de Variação Regular, identificando a relação entre a propriedade de variação regular e a subexponencialidade.

4.1 Distribuições Subexponenciais

As distribuições subexponenciais compõem uma classe especial das chamadas distribuições de cauda pesada. Seu nome se deve a uma de suas propriedades: suas caudas decaem mais lentamente do que qualquer cauda exponencial. Isto implica que grandes valores podem ocorrer em uma amostra, com probabilidade não desprezível, tornando as distribuições subexponenciais candidatas a modelar situações nas quais podem ocorrer, em uma amostra, valores extremamente grandes quando comparados ao valor médio dos dados.

Assim, dados do mercado financeiro, intensidade de movimentos tectônicos em uma região, quantidade de chuva em uma cidade, temperaturas em um setor, etc., podem ser modelados por esta classe de distribuições.

Definição 118 *Seja* $\{X_n\}_{n=1}^{\infty}$ *uma seqüência de variáveis aleatórias i.i.d. positivas com função de distribuição comum* F*, tal que* $F(x) < 1$ *para todo* $x > 0$*. Denotemos por*

$$\overline{F}(x) = 1 - F(x), \quad x \geq 0,$$

a função de sobrevivência ou cauda de F*, e por*

$$\overline{F^{n\star}}(x) = 1 - F^{n\star}(x) = Pr(X_1 + \ldots + X_n > x)$$

a cauda da convolução n*-variada de* F*.*

Dizemos que F *é uma função de distribuição subexponencial, e denotamos* $F \in \mathcal{S}$*, se alguma das seguintes condições equivalentes for verdadeira:*

$$\lim_{x \to \infty} \frac{\overline{F^{n\star}}(x)}{\overline{F}(x)} = n \quad para \quad todo \quad n \geq 2, \tag{4.1}$$

$$\lim_{x \to \infty} \frac{Pr(X_1 + \ldots + X_n > x)}{Pr(\max(X_1, \ldots, X_n) > x)} = 1 \quad para \quad todo \quad n \geq 2. \tag{4.2}$$

Observações.

(i) É possível mostrar que o limite (4.1) é válido para todo $n \geq 2$ se, e somente se, for válido para $n = 2$. Pode-se mostrar ainda que (4.1) é válido para $n = 2$ se for válido para algum $n \geq 2$ (ver Chistyakov [17] e Embrechts and Goldie [18]).

Desta forma fica claro que, para uma certa função de distribuição ser caracterizada como subexponencial, basta satisfazer a relação (4.1) simplesmente para algum $n \geq 2$.

(ii) Uma prova da equivalência entre (4.1) e (4.2) segue de

$$Pr(\max(X_1, \ldots, X_n) > x) = 1 - F^n(x) = \overline{F}(x) \sum_{k=0}^{n-1} F^k(x) \sim n\overline{F}(x), \quad x \to \infty,$$

onde a notação $\sim$ significa que o quociente entre os termos da direita e da esquerda tende a 1. Assim,

$$\frac{Pr(X_1 + \ldots + X_n > x)}{Pr(\max(X_1, \ldots, X_n) > x)} \sim \frac{\overline{F^{n\star}}(x)}{n\overline{F}(x)}.$$

(iii) A relação (4.2) fornece uma interpretação física para a subexponenciabilidade: é plausível que a soma de n variáveis aleatórias subexponenciais i.i.d. seja grande se, e somente se, for plausível que o máximo destas n variáveis seja grande. Isto esclarece a ocorrência de valores grandes em uma amostra de uma distribuição subexponencial.

Definição 119 *Sejam F e G funções de distribuição com suporte ilimitado à direita. Dizemos que F e G são assintoticamente equivalentes se*

$$\lim_{x \to \infty} \frac{\overline{F}(x)}{\overline{G}(x)} = c, \quad c \in (0, \infty).$$

Notemos que esta última definição se aplica a (4.1) e (4.2), provando a observação (*iv*).

(iv) Devido à relação (4.1) e ao fato de que $\mathcal{S}$ é fechada com respeito à equivalência assintótica, concluímos que

$$F \in \mathcal{S} \Rightarrow F^{n\star} \in \mathcal{S} \ \ para \ \ todo \ \ n \in I\!N.$$

Ainda mais, devido à relação (4.2) e ao fato de que F^n é a função de distribuição do máximo de n variáveis aleatórias i.i.d. com função de distribuição F, obtemos que

$$F \in \mathcal{S} \Rightarrow F^{n} \in \mathcal{S} \ \ para \ \ todo \ \ n \in I\!N.$$

Portanto $\mathcal{S}$ é fechada com respeito à soma e ao máximo de variáveis aleatórias i.i.d..

(v) A relação (4.2) ilustra a característica das distribuições subexponenciais de possuirem caudas pesadas. Isto fica mais claro a partir das implicações

$$F \in \mathcal{S} \Longrightarrow \lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1 \quad \forall y \in I\!R \tag{4.3}$$

$$\Longrightarrow \int_0^\infty e^{\varepsilon x} dF(x) = \infty \quad \forall \varepsilon > 0 \tag{4.4}$$

$$\Longrightarrow \frac{\overline{F}(x)}{e^{-\varepsilon x}} \to \infty \quad \forall \varepsilon > 0. \tag{4.5}$$

Observemos que a cauda $\overline{G}(x)$ de uma distribuição exponencial $G(x)$ indexada por um parâmentro $\varepsilon > 0$ é dada por

$$\overline{G}(x) = 1 - G(x) = 1 - (1 - e^{-\varepsilon x}) = e^{-\varepsilon x}.$$

Assim a propriedade (4.5) esclarece o nome distribuição subexponencial: a cauda de F decai mais lentamente do que qualquer cauda exponencial.

$*$

Como toda função de distribuição subexponencial possui a propriedade (4.3) então a classe das distribuições que possuem tal propriedade fornece potenciais candidatas à subexponencialidade. Tal classe é definida a seguir.

Definição 120 *Seja F uma função de distribuição definida em $(0, \infty)$ tal que $F(x) < 1$ para todo $x > 0$. Então dizemos que F pertence à classe $\mathcal{L}$, e denotamos $F \in \mathcal{L}$, se*

$$\lim_{x \to \infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1, \quad \forall y \in \mathbb{R}.$$

Notemos que a classe $\mathcal{S}$ está contida na classe $\mathcal{L}$, ou seja, uma função de distribuição pertencente a $\mathcal{L}$ pode não ser subexponencial.

4.2 Distribuições de Variação Regular

Definição 121 *Uma função f positiva e mensurável é dita de variação regular com parâmetro α se*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\alpha}, \quad \forall t > 0,$$

e denotamos $f \in \mathcal{R}(\alpha)$ para $\alpha \in \mathbb{R}$. Uma função $f \in \mathcal{R}(0)$ é dita de variação lenta.

A definição desta classe de funções foi originalmente apresentada por Karamata [20]. Sua aplicação para distribuições de probabilidade foi estudada por Feller [22].

As funções de distribuição cujas caudas são de variação regular formam uma subclasse de $\mathcal{S}$.

Exemplo 122 *Seja F uma função de distribuição tal que*

$$\overline{F}(x) = x^{-\alpha} g(x), \quad x > 0,$$

para alguma $g \in \mathcal{R}(0)$ e $\alpha \geq 0$. Então

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = \lim_{x\to\infty} \frac{(tx)^{-\alpha} g(tx)}{x^{-\alpha} g(x)} = t^{-\alpha} \lim_{x\to\infty} \frac{g(tx)}{g(x)} = t^{-\alpha}, \quad \forall t > 0.$$

Portanto $\overline{F} \in \mathcal{R}(-\alpha)$.

Assim, por definição, a propriedade de variação regular independe do comportamento de f em intervalos finitos.

4.3 Exemplos

Como vimos anteriormente, a classe das funções de distribuição cujas caudas são de variação regular constitui um subconjunto da classe das distribuições subexponenciais, ou seja, uma função de distribuição com cauda de variação regular é subexponencial. Nos exemplos seguintes estudamos algumas distribuições conhecidas, no sentido de identificar se suas caudas ou mesmo as próprias funções de densidade são de variação regular ou não.

Exemplo 123 *Consideremos a distribuição Pareto com parâmetros $\alpha, k > 0$, cuja cauda é dada por*

$$\overline{F}(x) = \left(\frac{k}{k+x}\right)^{\alpha}.$$

Temos que

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = \lim_{x\to\infty} \left(\frac{k+x}{k+tx}\right)^{\alpha} = \lim_{x\to\infty} \frac{\frac{d^\alpha}{dx^\alpha}(k+x)^\alpha}{\frac{d^\alpha}{dx^\alpha}(k+tx)^\alpha} = \lim_{x\to\infty} \frac{1}{t^\alpha} = t^{-\alpha}.$$

Então a distribuição Pareto com parâmetros α e k possui cauda de variação regular com expoente $-\alpha$ e é, portanto, subexponencial.

Exemplo 124 *Seja F a distribuição Exponencial com parâmetro $\alpha > 0$. A cauda desta distribuição é dada por*

$$\overline{F}(x) = e^{-\alpha x}.$$

Para qualquer $t > 1$, temos

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = \lim_{x\to\infty} e^{x(\alpha-\alpha t)} = 0,$$

o que nos leva a concluir que a distribuição Exponencial não possui cauda de variação regular.

Exemplo 125 *A distribuição Cauchy indexada por parâmetros $\alpha \in I\!R$ e $\beta > 0$ possui função de densidade dada por*

$$f(x) = \frac{1}{\pi\beta\{1 + [(x-\alpha)/\beta]^2\}},$$

e cauda dada por

$$\overline{F}(x) = \frac{1}{2} - \frac{\arctan(\frac{x-\alpha}{\beta})}{\pi}.$$

Então

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = \lim_{x\to\infty} t \left[\frac{1 + (\frac{x-\alpha}{\beta})^2}{1 + (\frac{tx-\alpha}{\beta})^2} \right] = \lim_{x\to\infty} \frac{x-\alpha}{tx-\alpha} = t^{-1},$$

de onde concluímos que a distribuição Cauchy possui cauda de variação regular com expoente igual a -1.

Exemplo 126 *Consideremos agora a distribuição Uniforme no intervalo $(0,1)$, cuja função de densidade é dada por*

$$f(x) = I_{[0,1]}(x).$$

A cauda desta distribuição é

$$\overline{F}(x) = \begin{cases} 1 & se \quad x < 0, \\ 1 - x & se \quad 0 \le x \le 1, \\ 0 & se \quad x > 1, \end{cases}$$

que não satisfaz a condição $\overline{F}(x) < 1$, $\forall x \in I\!R$.

Portanto a distribuição $\mathcal{U}(0,1)$ não possui cauda de variação regular.

Exemplo 127 *A distribuição t de Student com k graus de liberdade possui função de densidade*

$$f(x) = \frac{\Gamma[(k+1)/2]}{\Gamma[k/2]} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+x^2/k)^{(k+1)/2}}.$$

Supondo k ímpar, temos

$$\lim_{x\to\infty} \frac{\overline{f}(tx)}{\overline{f}(x)} = \lim_{x\to\infty} \frac{\frac{d^{(k+1)/2}}{dx^{(k+1)/2}}(1+x^2/k)^{(k+1)/2}}{\frac{d^{(k+1)/2}}{dx^{(k+1)/2}}(1+(tx)^2/k)^{(k+1)/2}} = \frac{1}{(t^2)^{(k+1)/2}} = \frac{1}{t^{k+1}}.$$

Assim, a distribuição t de Student com k (ímpar) graus de liberdade possui função de densidade de variação regular com expoente $-(k+1)$.

Exemplo 128 *Por fim consideremos a distribuição Normal padrão caracterizada pela função de densidade*

$$f(x) = \frac{1}{\sqrt{2\pi}} exp\{-\frac{1}{2}x^2\}.$$

Para qualquer $t < -1$ *ou* $t > 1$*, temos*

$$\lim_{x\to\infty} \frac{\overline{f}(tx)}{\overline{f}(x)} = \lim_{x\to\infty} exp\{\frac{1}{2}x^2(1-t^2)\} = 0,$$

o que nos leva a concluir que a função de densidade da distribuição $N(0,1)$ *não é de variação regular, como já era de se esperar, pois seu decaimento é exponencial.*

Outros exemplos de distribuições de variação regular são:

- Distribuição Pareto com parâmetros $\alpha, k > 0$, caracterizada pela cauda

$$\overline{F}(x) = \left(\frac{k}{k+x}\right)^{\alpha}.$$

- Distribuição Log-gamma com parâmetros $\alpha > 1$ e $\beta > 0$, caracterizada pela função de densidade

$$f(x) = \frac{\alpha^{\beta}}{\Gamma(\beta)} (\ln x)^{\beta - 1} x^{-\alpha - 1}.$$

- Distribuição Burr com parâmetros $\alpha, k, \tau > 0$, caracterizada pela cauda

$$\overline{F}(x) = \left(\frac{k}{k + x^{\tau}} \right)^{\alpha}.$$

- Distribuição Beta Transformada indexada pelos parâmetros $a \in I\!R$ e $p, q > 0$, caracterizada pela função de densidade

$$f(x) = \frac{|a| x^{ap-1}}{B(p,q)(1 + x^{a})^{p+q}},$$

onde $B(p,q)$ representa a função Beta dada por

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1} dx, \quad para \;\; p, q > 0.$$

- Distribuição α-Estável Truncada indexada pelo parâmetro α, $0 < \alpha < 2$, caracterizada pela cauda

$$\overline{F}(x) = Pr(|X| > x),$$

onde X é uma variável aleatória estável indexada pelo parâmetro α.

- Distribuição Weibull com parâmetro τ, $0 < \tau < 1$, caracterizada pela cauda

$$\overline{F}(x) = e^{-x^{\tau}}.$$

- Distribuição Lognormal com parâmetros $\mu \in I\!R$ e $\sigma > 0$, caracterizada pela função de densidade

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right).$$

4.4 Convolução e Máximo

As próximas duas proposições ilustram aplicações típicas da propriedade de variação regular.

Antes de enunciarmos a primeira destas proposições consideremos o que segue:

Seja $\{X_k\}_{k=1}^{\infty}$ uma seqüência de variáveis aleatórias mutuamente independentes com função de distribuição F comum, e tomemos

$$X_{(n)} = \max\{X_1, \ldots, X_n\}.$$

Nos interessa estudar se existem constantes a_n tais que a variável aleatória $\frac{X_{(n)}}{a_n}$ possui uma função de distribuição limite não degenerada, digamos G.

Como já vimos no Capítulo 2, se o suporte da função de distribuição F possuir um limitante superior, digamos x_0, então a distribuição de $X_{(n)}$ claramente tende à distribuição degenerada concentrada em x_0, quando n tende a ∞. Por outro lado, ainda podemos escolher uma seqüência de constantes a_n crescendo tão rapidamente de modo que $\frac{X_{(n)}}{a_n}$ converge em probabilidade a 0.

O seguinte resultado estabelece uma relação entre a existência de um limite não-degenerado e variação regular.

Proposição 129 *Seja F a função de distribuição de cada variável aleatória na seqüência $\{X_k\}_{k=1}^{\infty}$, de modo que $F(x) < 1$ para todo x. A função de distribuição de $\frac{X_{(n)}}{a_n}$, digamos G_n, converge, segundo constantes a_n apropriadas, a uma função de distribuição G não degenerada se, e somente se, a cauda $\overline{F} = 1 - F$ for de* *variação regular com expoente $\alpha < 0$. Neste caso, para $c > 0$,*

$$G(x) = \begin{cases} exp\{-cx^{\alpha}\} & se \quad x > 0, \\ 0 & se \quad x < 0. \end{cases}$$

Exemplo 130 *Uma outra distribuição que pode ser identificada como subexponencial através da propriedade de variação regular de sua cauda é a distribuição max-estável do Tipo II (ou Tipo Cauchy). Vejamos:*

O Tipo II é caracterizado pela função de distribuição

$$G(x) = exp\{-x^{-\alpha}\}, \quad x > 0.$$

Temos que

$$\lim_{x\to\infty} \frac{\overline{G}(tx)}{\overline{G}(x)} = \lim_{x\to\infty} \frac{1 - exp\{-(tx)^{-\alpha}\}}{1 - exp\{-x^{-\alpha}\}} = t^{-(\alpha+1)},$$

ou seja, a distribuição max-estável do Tipo Cauchy possui cauda de variação regular com expoente igual a $-(\alpha+1)$, o que nos leva a concluir que tal distribuição é subexponencial.

A segunda proposição envolve o conceito de convolução.

Proposição 131 *Sejam F_1 e F_2 duas funções de distribuição tais que, para $x \to \infty$,*

$$1 - F_i(x) = a_i x^{-\alpha} L(x),$$

com $L(x)$ de variação lenta e $\alpha > 0$. Então a convolução $G = F_1 \star F_2$ possui cauda de variação regular de modo que

$$1 - G(x) \sim (a_1 + a_2) x^{-\alpha} L(x). \tag{4.6}$$

Prova. Sejam X_1 e X_2 variáveis aleatórias independentes com funções de distribuição F_1 e F_2, respectivamente. Consideremos $t' = (1+\delta)t > t$. Assim, se uma das variáveis for maior do que t' e a outra for maior do que $-\delta t$, então $X_1+X_2 > t$, ou seja, o evento

$$[(X_1 > t' \quad e \quad X_2 > -\delta t) \quad ou \quad (X_1 > -\delta t \quad e \quad X_2 > t')]$$

implica $X_1 + X_2 > t$. Ainda mais, sabemos que

$$\lim_{t\to\infty} Pr(X_i > -\delta t) = 1, \quad i = 1, 2.$$

Assim, para algum $\epsilon > 0$ e t suficientemente grande, temos a relação

$$Pr(X_1 + X_2 > t) \geq [Pr(X_1 > t') + Pr(X_2 > t')](1-\epsilon),$$

que pode ser reescrita como

$$1 - G(t) \geq [(1 - F_1(t')) + (1 - F_2(t'))](1-\epsilon). \tag{4.7}$$

Por outro lado, se tomarmos $t'' = (1-\delta)t$ com $0 < \delta < \frac{1}{2}$, então o evento $X_1 + X_2 > t$ implica

$$[(X_1 > t'' \quad ou \quad X_2 > t'') \quad ou \quad (X_1 > \delta t \quad e \quad X_2 > \delta t)].$$

Para $t \to \infty$, a probabilidade do evento $[X_i > \delta t]$ é desprezível quando comparada com a probabilidade do evento $[X_i > t'']$. Assim podemos escrever, para $\epsilon > 0$ e t suficientemente grande,

$$1 - G(t) \leq [(1 - F_1(t'')) + (1 - F_2(t''))](1+\epsilon). \tag{4.8}$$

Por hipótese temos, para $t \to \infty$,

$$(1 - F_1(t)) + (1 - F_2(t)) = (a_1 + a_2)t^{-\alpha}L(t).$$

Como ϵ e δ são arbitrariamente pequenos, então as relações (4.7) e (4.8) nos levam à afirmação feita em (4.6).

$*$

Por indução, provamos o seguinte corolário.

Corolário 132 *Seja F uma função de distribuição, de modo que $1 - F(x) \sim x^{-\alpha}L(x)$ com L de variação lenta. Então $1 - F^{m\star}(x) \sim mx^{-\alpha}L(x)$.*

Exemplo 133 *Seja X_1 uma variável aleatória com distribuição Pareto com parâmetros $\alpha = 1$ e $k = 1$ e com função de distribuição denotada por F_1. Seja X_2 uma variável aleatória independente de X_1, com distribuição Pareto com parâmetros $\alpha = 1$ e $k = 2$ e com função de distribuição denotada por F_2. Temos que*

$$\overline{F}_1(x) = x^{-1} \sim 1x^{-1}L(x) \quad e \quad \overline{F}_2(x) = 2x^{-1} \sim 2x^{-1}L(x),$$

onde $L(x)$ é de variação lenta.

Então a função de distribuição $G = F_1 \star F_2$ da variável aleatória $X_1 + X_2$ possui cauda de variação regular de modo que

$$\overline{G}(x) \sim 3x^{-1}L(x).$$

Capítulo 5

Aplicações

Neste capítulo apresentamos duas aplicações de tópicos estudados ao longo da elaboração desta monografia. Na Seção 5.1 é apresentada uma aplicação relacionada a inferência para valores extremos e na Seção 5.2 uma aplicação relacionada à área de seguros e probabilidade de ruína.

5.1 Máximos anuais de níveis de maré

Como anteriormente, consideremos uma seqüência $\{X_n\}_{n=1}^{\infty}$ de variáveis aleatórias independentes e identicamente distribuídas com função de distribuição F, e ainda a variável aleatória

$$X_{(n)} = \max\{X_1, \ldots, X_n\}.$$

Como vimos, podem existir seqüências de constantes normalizadoras $a_n > 0$ e b_n

tais que

$$\lim_{n\to\infty} Pr[a_n(X_{(n)} - b_n) \leq x] = \lim_{n\to\infty} F^n(a_n^{-1}x + b_n) = G(x),$$

para todo x ponto de continuidade de G, onde G é uma função de distribuição não-degenerada.

O Teorema dos Três Tipos afirma que, quando este último limite existir, a função de distribuição G será do mesmo tipo que uma das distribuições de valor extremo, a saber:

- Distribuição do Tipo I, ou Tipo Exponencial ou Tipo Gumbel;
- Distribuição do Tipo II, ou Tipo Cauchy ou Tipo Fréchet;
- Distribuição do Tipo III, ou Tipo Limitada ou Tipo Weibull negativa.

Em termos práticos, um interesse especial reside sobre o Tipo Gumbel, cuja função de distribuição é dada por

$$G(x) = exp\left\{-exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, \quad -\infty < x < \infty,$$

onde μ e $\sigma > 0$ são, respectivamente, os parâmetros ditos de locação e de escala. Este interesse existe devido ao fato desta função de distribuição possuir um domínio de atração consideravelmente vasto contendo muitas distribuições conhecidas, tais como as distribuições Normal, Lognormal, Exponencial e as distribuições Gamma, além de ser considerada uma aproximação bastante razoável em muitas áreas de aplicação.

Jenkinson [24] mostra que as distribuições dos três tipos podem ser combinadas de modo a se obter uma única família paramétrica, chamada Distribuição de Valor Extremo Generalizada (DVEG), dada por

$$G(x) = exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-1/\xi}\right\},$$

onde (μ, σ, ξ) são, respectivamente, os parâmetros ditos de locação, de escala e de forma, $\sigma > 0$ e $z_{+} = \max\{z, 0\}$. A partir desta família, obtemos as distribuições de valor extremo dos Tipos Fréchet, Gumbel e Weibull negativa, respectivamente quando $\xi > 0$, $\xi = 0$ e $\xi < 0$, de modo que o Tipo Gumbel é definido por continuidade.

Na prática, o modelo DVEG é ajustado a dados que possam ser observados como máximos de seqüências.

Stephenson and Tawn [25] cita duas estratégias distintas de inferência estatística que têm sido desenvoldidas para explorar este modelo DVEG, e identificam deficiências em cada uma das estratégias.

A primeira das estratégias consiste em se ajustar o modelo DVEG diretamente. Apesar de ser o mais usado, este método apresenta o inconveniente de nunca selecionar o Tipo Gumbel, pois este tipo se reduz a uma hipótese exata ($\xi = 0$).

Na segunda estratégia primeiramente seleciona-se o tipo extremo dentre os três possíveis modelos ($\xi < 0, \xi = 0, \xi > 0$) através de um procedimento de teste de hipóteses. Então é ajustado o modelo a partir do tipo selecionado. Este método seleciona *a priori* o modelo Gumbel como correto, ignorando a incerteza inerente à escolha do tipo na inferência realizada subseqüentemente.

No sentido de superar estas deficiências, Stephenson and Tawn sugerem a aplicação de técnicas de Inferência Bayesiana a um modelo extendido que explicitamente aloca uma probabilidade não nula ao Tipo Gumbel. Especificamente é atribuida uma probabilidade p_ξ ao subespaço paramétrico correspondente ao Tipo Gumbel, de modo a incorporar o conhecimento a respeito da estrutura do Teorema dos Três Tipos na inferência para o DVEG. Uma conseqüência desta análise é a obtenção de probabilidades posterioris para cada um dos três tipos extremos, ou seja, probabilidades posterioris para os eventos $\xi < 0$, $\xi = 0$ e $\xi > 0$. Segundo os autores, a aplicação deste procedimento sugere que a redução da incerteza constitui o maior efeito de se agregar o conhecimento a respeito do Teorema dos Tipos Extremos na inferência para valores extremos.

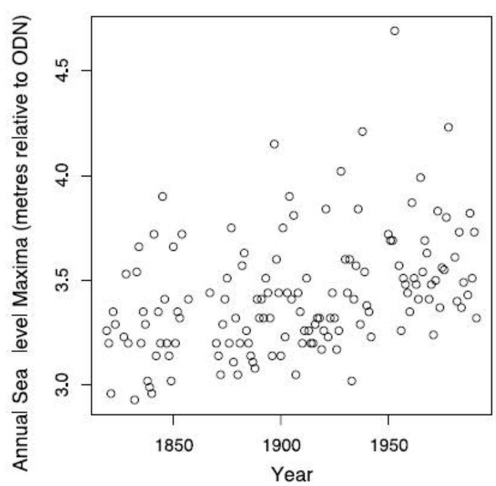

Figura 5.1: Máximos anuais dos níveis de maré em Sheerness, Inglaterra, entre 1819 e 1991, extraído de Stephenson and Tawn [25].

Estes mesmos autores analisam o conjunto de dados representado na figura 5.1, o qual se refere às observações de máximos anuais dos níveis de maré em Sheerness, na Inglaterra, entre os anos de 1819 e 1991. As observações são medidas em metros relativos a uma referência comum. Uma certa quantidade de dados foi perdida, incluindo as observações dos períodos entre 1858 e 1866 e entre 1943 e 1949.

5.2 Probabilidade de ruína

5.2.1 O modelo clássico

A área de seguros representa um dos campos de aplicação clássicos para as distribuições subexponenciais. Estas distribuições são usadas para modelar as indenizações (claim amount) em um seguro, as quais podem apresentar valores afastados dos valores centrais.

Em Mikosch [26], é definido o 'processo clássico de risco de seguro' como

$$R(t) = u + ct - \sum_{i=1}^{N_t} X_i, \quad t \geq 0, \tag{5.1}$$

onde

- $u \geq 0$ denota o capital inicial, uma quantia usualmente grande.
- $c > 0$ é a taxa de prêmio. Assim, de acordo com (5.1), há um prêmio linear em função do tempo.

- N_t é o número de indenizações até o instante t, definido por

$$N_t = \#\{i : T_i = Y_1 + \ldots + Y_i \leq t\}, \quad t \geq 0,$$

  onde $\{Y_i\}$ é uma seqüência de variáveis aleatórias i.i.d. não negativas, tais que Y_i representa o tempo decorrido entre as indenizações $i-1$ e i. Considera-se que $E(Y) = \frac{1}{\lambda} < \infty$ existe.

- $\{X_i\}$ são as indenizações. Considera-se que estas variáveis aleatórias são i.i.d. e não negativas, e possuem função de distribuição F. Ainda, $E(X_i) = \mu < \infty$. Notemos que a n-ésima indenização ocorre no instante T_n.

- As indenizações e os tempos de indenização são considerados independentes, ou seja, $\{X_i\}$ e $\{Y_i\}$ são independentes.

Com isto, é definido o evento

$$\{R(t) < 0 \ \ para \ \ algum \ \ t \ \ | \ \ u \geq 0\},$$

ao qual nos referimos como ruína, e a probabilidade

$$\psi(u) = Pr[R(t) < 0 \ \ para \ \ algum \ \ t \ \ | \ \ u \geq 0]$$

é então dita probabilidade de ruína.

Na área de seguros a probabilidade de ruína, como função de u, é uma importante medida de risco, e o processo de risco é chamado modelo de Cramér-Lundberg.

Seja $S(t) = \sum_{i=1}^{N_t} X_i$. Como a ruína pode ocorrer somente nos instantes de indenização T_n, então

$$\psi(u) = Pr\left[\inf_{t \geq 0}(u + ct - S(t)) < 0\right]$$

$$
\begin{aligned}
&= Pr\left[\inf_{n\geq 0}\left(u + cT_n - \sum_{i=1}^{n} X_i\right) < 0\right] \\
&= Pr\left[\inf_{n\geq 0}\sum_{i=1}^{n}(cY_i - X_i) < -u\right] \\
&= Pr\left[\sup_{n\geq 0}\sum_{i=1}^{n}(X_i - cY_i) > u\right].
\end{aligned}
$$

Mikosch [26] sugere que a probabilidade de não ruína, $1 - \psi(u)$, pode ser expressa como

$$1 - \psi(u) = \frac{\rho}{1+\rho}\sum_{n=0}^{\infty}(1+\rho)^{-n}F_I^{n\star}(u),$$

onde ρ é definido por

$$\rho = \frac{c}{\lambda\mu} - 1$$

e $F_I^{n\star}$ denota a função de distribuição de cauda integrada da convolução n-variada de F, de modo que a função de distribuição de cauda integrada é definida a seguir.

Definição 134 *Para uma função de distribuição F com média μ finita, definimos a função de distribuição de cauda integrada como*

$$F_I(x) = \frac{1}{\mu}\int_{-\infty}^{x}\overline{F}(y)dy, \quad x \in I\!R.$$

Neste contexto, segundo o mesmo autor, esta representação de $\psi(u)$ torna-se muito útil diante da dificuldade em se obter, para esta quantidade, certas estimativas assintóticas dadas por Cramér [27], as quais somente fazem sentido se $\overline{F}(x)$ decai segundo uma taxa exponencial.

Pode-se então escrever

$$\frac{\psi(u)}{\overline{F}_I(u)} = \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} (1+\rho)^{-n} \frac{\overline{F_I^{n\star}}(u)}{\overline{F}_I(u)}.$$

Se F_I for subexponencial, sabemos que

$$\overline{F_I^{n\star}}(u) \sim n\overline{F}_I(u),$$

e então pode-se mostrar que

$$\frac{\psi(u)}{\overline{F}_I(u)} \sim \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} (1+\rho)^{-n} n = \frac{1}{\rho}.$$

Esta solução para o problema de ruína para distribuições subexponenciais foi proposta por Embrechts and Veraverbeke [28].

5.2.2 Um modelo em tempo discreto

Em Tang and Tsitsiashvili [29], são apresentados resultados que levam a estimativas precisas para a probabilidade de ruína em tempo finito. Estes autores definem as variáveis aleatórias

$$S_m^\theta = \sum_{k=1}^{m} \theta_k X_k, \quad m \leq n, \quad e \quad M_n^\theta = \max_{1\leq m\leq n} S_m^\theta,$$

em que $\{X_k\}_{k=1}^n$ é uma seqüência de variáveis aleatórias i.i.d. com função de distribuição F e $\{\theta_k\}_{k=1}^n$ é uma seqüência de variáveis aleatórias positivas, de modo que $\{X_k\}_{k=1}^n$ e $\{\theta_k\}_{k=1}^n$ são mutuamente independentes. Desta forma, a variável aleatória S_m^θ é vista como uma soma aleatóriamente ponderada das variáveis X_k.

O principal resultado apresentado pelos autores se concentra basicamente no Teorema 136. Para enunciá-lo, no entanto, é necessária a apresentação da definição seguinte.

Definição 135 *Uma seqüência* $\{\xi_k\}_{k=1}^n$ *de variáveis aleatórias é dita 'limitada do tipo I' se*

$$Pr(a \leq \xi_k \leq b) = 1,$$

para a *e* b *tais que* $0 < a \leq b < \infty$ *e todo* $1 \leq k \leq n$.

Teorema 136 *Se a função de distribuição* F *das variáveis aleatórias na seqüência* $\{X_k\}_{k=1}^n$ *pertencer à classe das distribuições subexponenciais e a seqüência* $\{\theta_k\}_{k=1}^n$ *for limitada do tipo I, então*

$$Pr(M_n^\theta > x) \sim Pr(S_n^\theta > x) \sim Pr\left(\max_{1\leq k\leq n} \theta_k X_k > x\right) \sim \sum_{k=1}^{n} Pr(\theta_k X_k > x).$$

Consideremos agora a equação recursiva

$$S_0 = x, \quad S_n = \xi_n S_{n-1} + (\eta_n - Z_n), \quad n \geq 1, \tag{5.2}$$

que caracteriza o processo de superávit de uma companhia de seguros segundo um modelo em tempo discreto. Neste caso, $x \geq 0$ é o superávit inicial, Z_n e η_n são, respectivamente, o montante total referente às indenizações e o prêmio total da companhia durante o período de tempo n, e ξ_n é o coeficiente de inflação entre os tempos $n-1$ e n associado ao retorno de um investimento.

Suponhamos ainda que os pares (η_n, Z_n), $n \geq 1$, são i.i.d. e que as seqüências $\{(\eta_n, Z_n)\}_{n\geq 1}$ e $\{\xi_n\}_{n\geq 1}$ são mutuamente independentes.

Os resultados do Teorema 136 levam a estimativas para a probabilidade de ruína em tempo finito sob a suposição de uma estrutura de dependência arbitrária para a seqüência $\{\xi_n\}_{n\geq 1}$.

Especificamente, denotando por

$$\psi(x,n) = Pr\left(\min_{0\leq m\leq n} S_m < 0 \ \mid \ S_0 = x\right)$$

a probabilidade de ruína em um tempo n finito, então sob o modelo de risco (5.2) Tang and Tsitsiashvili mostra que se a função de distribuição F pertencer à classe das distribuições subexponenciais e a seqüência $\{\xi_k\}_{k=1}^{n}$ for limitada do tipo I, então

$$\psi(x,n) \sim \sum_{k=1}^{n} Pr\left(X_k \Pi_{i=1}^{k} Y_i > x\right),$$

onde $X_n = Z_n - \eta_n$, $n \geq 1$, são variáveis aleatórias i.i.d. com função de distribuição F e $Y_n = \xi_n^{-1}$, $n \geq 1$.

Referências Bibliográficas

[1] H. Cramér. *Mathematical Methods of Statistics*. Princeton Univ. Press, 1971.

[2] M. J. Schervish. *Theory of Statistics*. Springer, 1995.

[3] W. Feller. *An Introduction to Probability Theory an Its Applications*, volumes I, II. John Wiley & Sons, 1966.

[4] M. H. DeGroot. *Probability and Statistics*. Addison-Wesley Publishing Company, 1975.

[5] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 1968.

[6] P. Billingsley. *Probability and Measure*. John Wiley & Sons, 1979.

[7] E. L. Lima. *Espaços Métricos*. Instituto de Matemática Pura e Aplicada, 1983.

[8] E. L. Lima. *Curso de Análise*, volumes I, II. Instituto de Matemática Pura e Aplicada, 1981.

[9] O. H. Bustos & A. C. Frery. *Simulação Estocástica: Teoria e Algoritmos*. Instituto de Matemática Pura e Aplicada, 1992.

[10] M. R. Leadbetter, G. Lindgren & H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer, 1983.

[11] S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes.* Springer, 1987.

[12] E. J. Gumbel. *Statistics of Extremes*. Columbia Univ. Press, 1958.

[13] R. A. Fisher & L. H. C. Tippett. *Limiting forms of the frequency distribution of the largest or smallest member of a sample*. Proc. Cambridge Phil. Soc., 24:180-190, 1928.

[14] A. M. Mood, F. A. Graybill & D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, 1963.

[15] B. V. Gnedenko & A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley Publishing Company, 1968.

[16] C. Klüppelberg. *Subexponential Distributions*. Notas. http://www.ma.tum.de/stat/

[17] V. P. Chistyakov. *A theorem on sums of independent positive random variables and its applications to branching random processes*. Theory Probab. Appl. 9, 640-648, 1968.

[18] P. Embrechts & C. M. Goldie. *On convolution tails*. Stoch. Proc. Appl. 13, 263-278, 1982.

[19] J. D. O. dos Santos. *Discussões sobre a Relação entre Distribuições de Cauda Pesada e Conflitos de Informação em Inferência Bayesiana*. Dissertação de Mestrado. IMECC. UNICAMP.

[20] J. Karamata. *Sur un Mode de Croissance Régulière des Fonctions*. Mathematica (Cluj), V. IV, 38-53, 1930.

[21] J. Karamata. *Sur un Mode de Croissance Régulière. Théorèmes Fondamentaux*. Bulletin de la S. M. F. 61, 55-62, 1933.

[22] W. Feller. *On Regular Variation and Local Limit Theorems*. Proc. Fifth Berkeley Sympos. Math. Statist. and Probability, Vol II: Contributions to Probability Theory, Part 1, 373-388, 1967.

[23] N. H. Bingham, C. M. Goldie & J. L. Teugels. *Regular Variation*. Encyclopedia of Mathematics and its Applications, Vol. 27. Cambridge Univ. Press, 1987.

[24] A. F. Jenkinson. *The frequency distribution of the annual maximum (or minimum) of meteorological elements*. Quant. J. R. Met. Soc. 81, 158-171, 1955.

[25] A. Stephenson & J. Tawn. *Bayesian Inference for Extremes: Accounting for the Three Extremal Types*. Extremes 7, 291-307, 2004.

[26] T. Mikosch. *Regular Variation, Subexponentiality and Their Applications in Probability Theory*. Notas. University of Groningen.

[27] H. Cramér. *On the Mathematical Theory of Risk*. Skandia Jubilee Volume, Stockholm, 1930.

[28] P. Embrechts & N. Veraverbeke. *Estimates for the Probability of Ruin with Special Emphasis on the Possibility of Large Claims*. Insurance: Math. Econom. 1, 55-72, 1982.

[29] Q. Tang & G. Tsitsiashvili. *Randomly Weighted Sums of Subexponential Random Variables with Application to Ruin Theory*. Extremes 6, 171-188, 2003.