

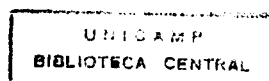
RESISTÊNCIA E ROBUSTEZ QUALITATIVAS EM PROCESSOS ESTACIONÁRIOS ERGÓDICOS

MARIO ANTONIO GNERI

Orientador: Prof. Dr. Oscar Humberto Bustos

IMECC - UNICAMP

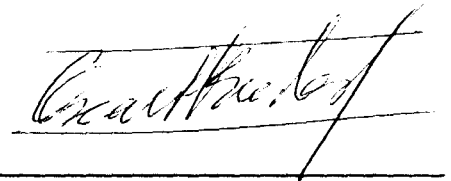
1995



RESISTÊNCIA E ROBUSTEZ QUALITATIVAS
EM PROCESSOS ESTACIONARIOS ERGÔDICOS

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida pelo Sr. Mario Antonio Gneri e aprovada pela Comissão Julgadora.

Campinas, 25 de agosto de 1995.



Prof. Dr. Oscar Humberto Bustos

Tese apresentada ao Instituto de Matemática, Estatística e Ciências da Computação, UNICAMP, como requisito parcial para obtenção do Título de Doutor em Ciências.

FICHA CATALOGRAFICA ELABORADA PELA
BIBLIOTECA DO INECC DA UNICAMP

Gneri, Mario Antonio

8553r Resistencia e robustez qualitativas em processos
estacionarios ergodicos/ Mario Antonio Gneri. -- Campinas,
1985 : s.n.l, 1985.

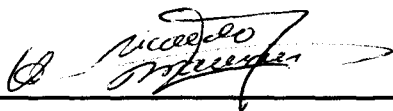
Orientador : Oscar Humberto Bustos

Tese (Doutorado) - Universidade Estadual de Campinas,
Instituto de Matematica, Estatistica e Ciencia da Computacao.

1. Estatistica robusta. 2. Processo estacionario. I. Bustos,
Oscar Humberto. II. Universidade Estadual de Campinas.
Instituto de Matematica, Estatistica e Ciencia da Computacao.
III. Titulo.

Tese defendida e aprovada em, 25 de 08 de 1995

Pela Banca Examinadora composta pelos Profs. Drs.



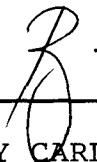
Prof(a). Dr(a). JACOB RICARDO FRAIMAN



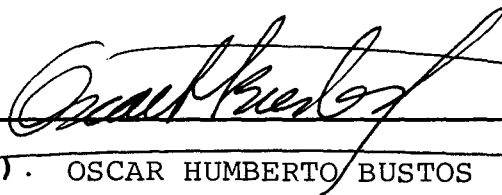
Prof(a). Dr(a). MANUEL FOLLEDO



Prof(a). Dr(a). ARNOLDO DE HOYOS GUEVARA



Prof(a). Dr(a). RODNEY CARLOS BASSANEZI



Prof(a). Dr(a). OSCAR HUMBERTO BUSTOS

"Las únicas causas perdidas son las causas abandonadas"
Hebe de Bonafini, presidente de Madres de Plaza de Mayo,
algún jueves, en la plaza.

A la memoria de mis viejos
A Paula y Maria Rosa
A mis amigos y familia
Aos 111 cidadãos assassinados no Carandirú
Aos trabalhadores petroleiros do Brasil 1995
A mis madres de la Plaza

Agradecimientos.

A Oscar, por su orientación culta, ofreciéndome enorme
variedad de opciones y aceptando mis elecciones.

A María Rosa, sin cuya ayuda competente y amorosa
hubiera sido imposible sacar este trabajo del vetusto
chi. writer 2 en que fue concebido.

A algunos amigos, cuyo incentivo constante me empujó a
encarar este trabajo, entre ellos Ana, María Rosa,
Regina, Alvaro, Armando, Manolo, Rodney,...

"y dale, y dale, y dale rojo dale"

Mario, autor de este trabajo, festejando una apilada de
Bocchini o, alternativamente, descubriendo el marxismo.

PRÓLOGO

Nesta tese nos demonstramos a equivalência entre os conceitos de resistência fraca e forte de Boente, Fraiman e Yohai, para estimadores invariantes por permutações das coordenadas, no contexto dos processos estacionários ergódicos, estendendo análogo resultado provado por Boente, Fraiman e Yohai (1987, *The Annals of Statistics*, Vol.15, No3, 1293-1312), no âmbito dos processos com observações independentes identicamente distribuídas (i.i.d.).

A invariância por permutações é uma restrição muito forte nos estimadores em processos não i.i.d., motivo que limita a aplicabilidade de nosso resultado.

O trabalho está organizado da seguinte maneira:

No capítulo 1 nos ocupamos da evolução histórica das idéias de robustez.

No capítulo 2 revisamos as propriedades gerais da topologia fraca no conjunto $\mathcal{M}$ das medidas finitas definidas no espaço mensurável $(\Omega, \mathcal{B})$, onde $\mathcal{B} = \{\text{borelianos do espaço métrico } (\Omega, d)\}$. Apresentamos aqui resultados conhecidos que usualmente se encontram restritos a medidas de probabilidade, agora com enunciados e demonstrações válidas em $\mathcal{M}$.

No capítulo 3 abordamos o estudo da métrica de Prohorov no espaço $\mathcal{M}$, realizando comparações entre as métricas de Levy e Prohorov no caso em que Ω é o conjunto dos números reais com a métrica usual. Revisamos o trabalho de Prohorov e as relações da topologia induzida pela métrica de Prohorov em $\mathcal{M}$ e a topologia fraca, em especial no caso (Ω, d) separável e completo. Provamos a equivalência de diferentes caracterizações da métrica de Prohorov e também a equivalência entre a completitude de $(\mathcal{M}, \Pi)$ e a de $(\mathcal{F}, \Pi/\mathcal{F})$, onde $\mathcal{F} = \{M \in \mathcal{M} : M(\Omega) = 1\}$.

No capítulo 4 fundamentamos formalmente a necessidade da estimação robusta em modelos paramétricos, mostrando que em geral o "modelo paramétrico" é um subconjunto magro do espaço das distribuições. São enunciadas e provadas algumas propriedades do espaço $(\mathcal{F}, \Pi)$ que justificam a escolha de Π , por sua capacidade de absorver os desvios das situações ideais. Também fundamentamos a escolha dos espaços métricos separáveis e completos como "espaços das observações".

No capítulo 5 estudamos a evolução dos conceitos de robustez qualitativa (Hampel / Papantoni-Kazacos e Gray / Boente, Fraiman e Yohai) e de resistência (Mosteller e Tukey/ Hampel/ Boente, Fraiman e Yohai). Finalmente demonstramos aqui a equivalência dos conceitos de resistência fraca e forte em estimadores invariantes por permutações das coordenadas em processos estacionários e ergódicos.

Segundo nosso ponto de vista, as contribuições deste trabalho são:

i) O Teorema 5.22, onde demonstramos a equivalência dos conceitos de resistência fraca e forte, para estimadores invariantes por permutações, em processos estacionários e ergódicos.

ii) As proposições 4.10 e 4.11 ((1), (3) e (4)), que dão um suporte formal a certa "sabedoria popular" em torno das limitações dos modelos paramétricos e o papel das técnicas robustas. Hampel (1986) afirma que: *a estatística paramétrica clássica considera somente um subconjunto (muito "magro") de baixa dimensão do conjunto de todas as probabilidades*. Na Proposição 4.10 provamos que, nos casos usuais, "o modelo" é um conjunto magro no sentido estrito (união enumerável de não densos) do conjunto de todas as probabilidades. Complementariamente, 4.11 diz respeito da adequacidade da métrica de Prohorov na descrição de uma "vizinhança real" de qualquer distribuição (nos provamos apenas (1), (3) e (4); (2) é devido a Prohorov).

iii) As propriedades da métrica de Prohorov da Proposição 3.10, que mostram a estrutura de cone de um espaço vetorial topológico do conjunto das medidas finitas. Utilizamos estas propriedades na demonstração da equivalência entre a completitude do espaço das medidas finitas e o das probabilidades (Lema 3.22).

iv) As demonstrações de equivalência de definições da métrica de Prohorov e comparações entre as métricas de Levy e Prohorov (no Capítulo 3), fatos conhecidos cujas demonstrações não se encontram disponíveis na literatura, segundo nosso conhecimento.

v) A generalização das propriedades da topologia fraca ao espaço das medidas finitas (Capítulo 2). Usualmente estes resultados (também conhecidos) só se encontram na literatura restritos ao espaço das probabilidades.

vi) A exposição numa linguagem corrente, sem cair em simplificações excessivas, da problemática da robustez objetivando possibilitar o acesso de um público vasto a estas idéias. Tentamos fazer isto no Capítulo 1.

ÍNDICE

CAPÍTULO 1: IDÉIAS GERAIS SOBRE ROBUSTEZ.	1.1
1.A. Um exemplo da vida real	1.1
1.B. Uma olhada crítica no estimador de máxima verossimilhança	1.5
1.C. O que significa robustez em Estatística?	
Quais os objetivos dos procedimentos robustos?	1.7
1.C.1. Algumas definições,	1.7
1.C.2. Relações entre robustez e rejeição de outliers,	1.8
1.C.3. Relações entre robustez e estatística não paramétrica,	1.10
1.C.4. Os objetivos e as propriedades desejáveis dos procedimentos robustos,	1.12
1.D. Um pouco de história	1.13
1.D.1. Os primeiros passos,	1.14
1.D.2. A média, a distribuição gaussiana e o autoritarismo,	1.15
1.D.3. Outras tentativas,	1.20
1.D.4. Os principais conceitos desenvolvidos no passado próximo,	1.22
 CAPÍTULO 2: CONVERGÊNCIA FRACA DE MEDIDAS EM ESPAÇOS MÉTRICOS. A TOPOLOGIA FRACA.	 2.1
2.A. Introdução e resumo do capítulo	2.1
2.B. Medidas nos borelianos de um espaço métrico	2.2
2.C. Convergência fraca de medidas definidas nos borelianos de um espaço métrico	2.6
2.D. Elementos aleatórios e convergência em distribuição	2.10
2.E. O caso particular de (R, e) : caracterização de medidas finitas e de convergência fraca via funções de distribuição	2.13
2.F. Topologia fraca no espaço das medidas finitas de um espaço métrico	2.16
2.G. Demonstrações	2.20

CAPÍTULO 3: A DISTÂNCIA DE PROHOROV E A METRIZAÇÃO DA TOPOLOGIA FRACA.	3.1
3.A. Introdução e resumo do capítulo	3.1
3.B. Medidas finitas nos $\mathcal{B}$ borelianos de (R, e) . A métrica de Levy L em $(\mathcal{M}^+(R, \mathcal{B}), L)$. Completitude de $(\mathcal{M}^+(R, \mathcal{B}), L)$	3.5
3.C. Medidas finitas nos borelianos $\mathcal{B}$ de um espaço métrico qualquer $(\Omega, \mathcal{B})$. A métrica de Prohorov em $\mathcal{M}^+(\Omega, \mathcal{B})$	3.9
3.D. Separabilidade e metrizabilidade de $\mathcal{M}^+(\Omega, \mathcal{B})$	3.13
3.E. Completitude e compacidade em $\mathcal{M}^+(\Omega, \mathcal{B})$	3.15
3.F. Demonstrações	3.26
 CAPÍTULO 4: FUNDAMENTAÇÃO FORMAL DA ESTIMAÇÃO ROBUSTA EM MODELOS PARAMÉTRICOS E JUSTIFICATIVA DA ESCOLHA DAS FERRAMENTAS MATEMÁTICAS ADEQUADAS.	 4.1
4.A. Introdução e resumo do capítulo	4.1
4.B. Estimadores definidos através de funcionais	4.2
4.C. Convergência das medida empíricas à distribuição subjacente. Necessidade de funcionais contínuos	4.7
4.D. Os funcionais lineares estatísticos e os estimadores a eles associados	4.10
4.E. O modelo como subconjunto magro de um espaço de distribuições, os desvios do modelo e a métrica de Prohorov	4.13
4.F. Derivada de von Mises e curva de influência de Hampel	4.18
4.G. Os estimadores robustos con viés minimax	4.25
4.H. Demonstrações	4.28
 CAPÍTULO 5: RESISTÊNCIA EM PROCESSOS ESTACIONÁRIOS ERGÓDICOS	 5.1
5.A. Introdução e resumo do capítulo	5.1
5.B. Questões preliminares e notacionais	5.5
5.C. Robustez e resistência qualitativa	5.8
5.C.1. Processos e estimadores, 5.8	
5.C.2. Robustez qualitativa, 5.8	
5.C.3. Resistência qualitativa, 5.10	
5.D. Equivalência entre resistência fraca e forte	5.15
5.E. Apêndice	5.21

REFERÊNCIAS.

- BILLINGSLEY, P.. (1968). Convergence of probability measures, Wiley, New York.
- BOENTE, G., FRAIMAN, R. and YOHAI, V.J. (1987). Qualitative robustness for stochastic processes. *The Annals of Statistics*, Vol. 15, No. 3, 1293-1312.
- BUSTOS, O.H. (1988). Outliers e robustez, *R. bras. Estat.*, Rio de Janeiro, 49(191):5-25, jan./jul..
- BUSTOS, O. H. (1981). Qualitative robustness for general processes. Informes de matemática. Série B-002/81. I.M.P.A., Rio de Janeiro.
- COX, D. (1981). Metrics on stochastic processes and qualitative robustness. Technical Report 3, Dept. Statistics, Univ. Washington.
- CRAMER, H. (1946). Mathematical Methods of Statistics, Princeton University Press.
- FERNLHOLZ, L.T. (1983). Von Mises Calculus for Statistical Functionals, Lecture Notes in Statistics, 19, Springer, New York.
- HALD, A. (1990). A history of probability and statistics and their applications before 1750, Wiley.
- HAMPEL, F.R. (1968). Contributions to the theory of robust estimation. Unpublished PhD thesis. University of California. Berkeley.
- HAMPEL, F.R. (1971). A general qualitative definition of robustness. *Ann. Math. Statisc.* 42, 1887-1896.
- HAMPEL, F.R. (1974). The influence curve and its role in robust estimation. *J. Am. Stat. Assoc.* 69, 383-393.

- HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J., STAHEL, W.A. (1986) ^{Robustness} The Approach Based on Influence Functions. Wiley, New York.
- HALMOS, P.R. (1950). Measure Theory. D. Van Nostrand. New York.
- HUBER, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* 35, 73-101.
- HUBER, P.J. (1972). Robust Statistics: a review. *Ann. Math. Statist.* 43, 1041-1067.
- HUBER, P.J. (1981). Robust Statistics. Wiley, New York.
- KELLEY, J.L. (1962). Topologia general. EUDEBA Editorial Universitaria de Buenos Aires.
- KENDALL, M.G. and BUCKLAND, W.R. (1971). A dictionary of statistical terms (prepared for the I.S.I.). Oliver and Boyd, Edinburg, 3rd. edition.
- MARTIN, R.D. and YOHAI, V.J. (1984). Robustness in time series and estimating ARMA models. Technical Report No. 50, Department of Statistics, University of Washington. Seattle, Washington.
- MARTIN, R.D., YOHAI, V.J. and ZAMAR, R.H. (1989). Min-max robust bias robust regression. *Ann. Statist.* 17, 1608-1630.
- MOSTELLER, F. and TUKEY, J.W. (1977). Data Analysis and Regression: A Second Course in Statistics. Addison-Wesley, Reading, Mass.
- PARTHASARATHY, K.R. (1967). Probability measures on metric spaces, New York: Academic Press.
- PROHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theor. Probability. Appl.* 1, 157-214.

- REY, W.J.J. (1978). Robust Statistical Methods. Lecture Notes in Mathematics N^o 690. Springer Verlag. Berlin.
- ROUSSEEUW, P.J. (1984). Least median of regression squares. *J. Amer. Statist. Assoc.* 79, 871-880.
- STIGLER, S.M. (1973). Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885-1920. *J. Am. Stat. Assoc.* 68, 872-879.
- STIGLER, S.M. (1975). Discussion of F.R. Hampel's paper "Beyond location parameters: Robust concepts and methods". Proc. 40th Session I.S.I., vol XLVI, Book 1, pp. 375-391.
- STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* 36, 423-439.
- TUKEY, J.W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics*, I. Olkin (ed.). Stanford University Press, Stanford, California, pp. 448-485.
- TUKEY, J.W. (1975). Instead of Gauss Markov least squares, what? *Applied Statistics*, R.P. Gupta (editor). North Holland Publishing Company, pp. 351-372.
- von MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* 18, 309-348.
- von MISES, R. (1964). *Mathematical Theory of Probability and Statistics*. Academic Press, New York.

CAPÍTULO 1

IDÉIAS GERAIS SOBRE ROBUSTEZ

1.A. UM EXEMPLO DA VIDA REAL.

Na edição de terça feira, 24 de janeiro de 1995, do jornal "Gazeta Mercantil", numa matéria intitulada "Razões para mudar câmbio", são citadas declarações do ex-presidente do Banco Central e eminente economista Affonso Celso Pastore, nas quais ele reconhece ter cometido um erro grosseiro em suas previsões a respeito do saldo da balança comercial.

O que teria levado tão distinto usuário de métodos estatísticos a semelhante erro? Será que a teoria estatística estaria errada?

O próprio Pastore, na matéria acima citada, tenta um esboço de resposta: "Uma coisa é uma equação baseada numa regressão matemática, outra é o mundo real". O que é uma verdade trivial que nada esclarece.

Não é muito difícil, entretanto, imaginar o que aconteceu. O Dr. Pastore utilizou um procedimento estatístico - estimou uma reta de

regressão (por quadrados mínimos) - a partir de dados do passado e fez sua previsão mediante extrapolação. Mas ele não considerou que o governo, através da anulação ou da diminuição drástica dos impostos de importação, tinha mudado totalmente as regras do jogo. É possível que o Dr. Pastore soubesse que os estimadores de quadrados mínimos são muito utilizados ou também que até são "ótimos" em algum sentido. Mas não considerou que eles foram concebidos para funcionar apenas sob determinadas condições.

Na falta de qualquer outra ferramenta teórica que lhe permitisse atacar a nova situação, teria sido preferível que o Dr. Pastore tentasse uma resposta na base do "chutômetro" a ficar preso a uma fórmula matemática inaplicável.

Pensemos numa situação equivalente: uma pessoa qualquer (portanto não necessariamente o Dr. Pastore) utiliza "o melhor martelo do mundo" para apertar um parafuso. Certamente o resultado não será muito bom.

Este tipo de erro, que consiste em fazer de conta que qualquer procedimento pode ser utilizado em qualquer circunstância, é habitual na aplicação das ferramentas matemáticas.

As causas deste uso inapropriado são muito profundas e (segundo nosso ponto de vista) tem a ver com as terríveis deficiências da educação matemática em todos os níveis (autoritária e centrada na inútil decoração de fórmulas). Raramente é explicado que todas as ciências trabalham com modelos aproximados da realidade e não com a realidade mesma, o que dificulta enormemente a compreensão tanto da realidade como dos modelos matemáticos que tentam descrevê-la.

Por exemplo, na escola primária, a chamada "regra de três" ocupa um espaço importantíssimo: se um pãozinho vale 5 centavos, então 8 pãezinhos valem 8 vezes 5 centavos. O problema não está no ensino da "receita", mas sim em lhe conferir o *status* de lei universal. Não demorariam em aparecer as contradições com "a realidade" (caso não fossem habilmente escapulidas), como é possível verificar através dos dois seguintes exemplos:

a) a regra diz: se um pãozinho vale 5 centavos, então 10.000 pãezinhos valem 10.000 vezes 5 centavos = 500 reais. A realidade ensina que a pessoa que comprasse 10.000 pãezinhos certamente obteria um abatimento no preço;

b) a regra diz que se 5 pedreiros demoram 100 dias para concluir a construção de uma pequena casa, então 50.000 pedreiros demoram $(5 \times 100 \text{ dias}) / 50.000 = 0,01 \text{ dia} = 14,4 \text{ minutos}$. A realidade diz que 50.000 pessoas nem caberiam no espaço físico destinado à construção.

Ou seja, não é dito que a regra de três é apenas uma receita que se aplica em algumas circunstâncias, como, por exemplo, nas compras diárias de uma família. Existe uma enorme variedade de opções para mostrar de forma simples que, muitas vezes, a receita não funciona. Por exemplo, se um rapaz de 12 anos tem 1,50 metro de estatura, isto não implica que a estatura aos 24 anos será de 3 metros.

O aluno mais esperto no uso da "regra de três" será, sem dúvidas, conceituado como um dos melhores da turma. Será que esta esperteza é um

sinal de "inteligência", ou denotará apenas falta de capacidade de observação e de espírito crítico?

Já na escola secundária, quando se trabalha com a função quadrática ou a função logarítmica, perde-se muitas vezes a oportunidade de mostrar que a interpolação linear funciona razoavelmente, desde que os dois pontos entre os quais se interpole não estejam "muito longe". Isto pode ser feito escolhendo três valores conhecidos "muito próximos", estimando o valor central por interpolação entre os dois extremos e comparando com o valor verdadeiro. Por outro lado, poderia ser estimado *logaritmo* (10) por interpolação entre *logaritmo* (1) e *logaritmo* (20), sendo trivial a verificação da inadequação deste procedimento. Ou, também, estimado *logaritmo* ($10^{-2}/2$) por interpolação entre *logaritmo* (10^{-20}) e *logaritmo* (10^{-2}), questionando o caráter absoluto das expressões "muito perto" ou "não muito longe", relacionando-as com a velocidade da mudança da função no conjunto de interesse e com a precisão desejada. Entretanto o aluno é "ensinado" a interpolar de forma mecânica, sem sequer ter pensado no papel de pressupostos de continuidade ou monotonia da função entre cujos valores pretende-se interpolar linearmente.

Sempre segundo nosso ponto de vista, na sociedade dividida em classes há mais interesse no "treino" do que na educação matemática (ou melhor, na educação em geral), já que é suficiente que alguns poucos pensem. Mas o furo é tão grande que até as pessoas destinadas a pensar acabam tendo também prejudicado o seu desenvolvimento, como acabamos de comprovar no caso do sempre ministeriável Dr. Pastore.

Deixando de lado as questões sociais, voltemos ao nosso exemplo. Qual é a relação dos fatos acima descritos com a problemática da robustez? Para responder a esta pergunta, deveremos previamente dar um significado a esta palavra (veja 1.C).

1.B. UMA OLHADA CRÍTICA NO ESTIMADOR DE MÁXIMA VEROSIMILHANÇA.

A idéia desta seção é colocar algumas interrogações na utilização costumeira das técnicas estatísticas. Dado que o estimador de máxima verossimilhança (e.m.v.) é um dos procedimentos estatísticos mais prestigiados e de uso mais divulgado, ele é um bom ponto de partida para colocar alguns questionamentos.

Uma de suas propriedades mais importantes é que o e.m.v., sob certas condições, é assintoticamente normal e eficiente (veja por exemplo Cramer, 1946). Isto significa que, pelo menos em termos assintóticos, ele é imbatível.

Analisemos, entretanto, a definição de e.m.v.: suponhamos ter uma amostra aleatória simples (observações independentes identicamente distribuídas) $x_1, x_2, \dots, x_n$ de uma variável aleatória cuja distribuição é dada por uma densidade f_θ , onde θ é o parâmetro a estimar. O e.m.v. é definido como o valor de θ que maximiza a expressão:

$$\prod_{j=1}^n f_\theta(x_j),$$

ou equivalentemente é a solução da equação de verossimilhança:

$$\sum_{j=1}^n \log(f'_{\theta}/f_{\theta})(x_j) = 0,$$

resultando evidente a dependência do estimador não apenas das observações como também de f_{θ} .

Então, mesmo deixando de lado a resolução da equação de verossimilhança, dado que:

- pode-se ter erro na medição dos x_j ;
- o modelo f_{θ} é apenas uma aproximação da distribuição de uma característica do fenômeno sob estudo;
- as "certas condições" sob as quais o estimador é ótimo raramente são verificadas, acontecendo o mesmo com a hipótese de independência;

é razoável se perguntar a respeito da estabilidade da solução da equação de verossimilhança sob desvios das condições ideais.

Infelizmente a resposta à pergunta acima não é, em geral, satisfatória, já que o comportamento dos estimadores e testes clássicos é altamente instável na presença de desvios em relação às hipóteses em que estes procedimentos se baseiam. A estatística robusta assume que as hipóteses são satisfeitas apenas aproximadamente e tenta construir procedimentos estatísticos que garantam alguma estabilidade neste contexto.

1.C. O QUE SIGNIFICA ROBUSTEZ EM ESTATÍSTICA?

QUAIS OS OBJETIVOS DOS PROCEDIMENTOS ROBUSTOS?

1.C.1. Algumas definições.

Kendall e Buckland (1971), no seu dicionário de termos estatísticos, dizem que:

"Um procedimento estatístico é chamado robusto se não é muito sensível a desvios das suposições sobre as quais se baseia".

Huber (1981, p. 1), afirma que:

"A inferência estatística se baseia apenas parcialmente nos dados. Uma parte igualmente importante é formada por supostos apriorísticos a respeito da situação subjacente. Ainda nos casos mais simples existem hipóteses explícitas ou implícitas acerca de aleatoriedade e independência, relativas a modelos distribucionais, talvez a distribuições a priori para alguns parâmetros desconhecidos, etc..

Estas hipóteses não são exatamente verdadeiras - elas são racionalizações matemáticas de conhecimentos freqüentemente difusos ou crenças. Como em qualquer outro ramo da matemática aplicada, tais racionalizações ou simplificações são vitais, sendo seu uso justificado pela invocação de um vago princípio de estabilidade ou continuidade: um

pequeno erro no modelo matemático causará apenas um pequeno erro nas conclusões finais.

Infelizmente isto nem sempre é verdadeiro."

Um pouco mais adiante esboça uma definição geral que coincide com a de Kendall e Buckland:

"Robustez significa insensibilidade a pequenos desvios dos supostos".

As expressões "insensibilidade" e "pequenos desvios" serão dotadas de um sentido mais preciso nos capítulos 4 e 5.

Muitas pessoas associam de forma confusa robustez com rejeição de "outliers" ou também com estatística não paramétrica. A seguir tentaremos esclarecer estas "confusões".

1.C.2.Relações entre robustez e rejeição de outliers.

Segundo Bustos (1988):

"quando realizamos uma experiência e medimos uma certa variável, nunca podemos estar absolutamente seguros de que um determinado dado seja realmente uma manifestação genuína do fenômeno que queremos estudar. Sem dúvida a confiabilidade provável de uma observação se baseia em sua relação com o conjunto das outras observações dessa variável realizadas em condições similares. Assim, um outlier dentro de um conjunto de dados será

uma observação (ou um conjunto de observações) que parece ser inconsistente com a maioria de dados do conjunto".

A detecção de outliers faz parte da estatística robusta e tem sido uma preocupação antiga dos cientistas. Neste item faz-se necessário o esclarecimento de alguns pontos.

(1) A detecção de outliers é apenas um dos objetivos da estatística robusta, porém importantíssimo, dado o enorme impacto dos outliers nos estimadores convencionais: média e variância amostral, coeficientes da reta de regressão amostral, coeficiente de correlação amostral, etc.

(2) É importante salientar que o objetivo não é a rejeição automática de outliers, mas sim sua detecção para um posterior tratamento em separado, que pode ser inclusive a rejeição. Este seria o caso de um físico que pretende estimar o comprimento de um objeto e obtém várias medições muito parecidas e uma única totalmente discordante: ele deve calcular a média de todas as observações exceto a discordante. Entretanto, muitas vezes o outlier é a observação mais importante e merece estudo especial: no ano de 1845 o astrônomo francês Urbano Le Verrier explicou o comportamento "anômalo" do planeta Urano pela existência de um planeta "transuraniano", calculando teoricamente sua órbita provável. No ano seguinte o astrônomo alemão Johannes Gottfried Galle descobriu o planeta Netuno seguindo as indicações de Le Verrier. Posteriormente, algumas anomalias ainda inexplicadas de Urano e as próprias do planeta Netuno levaram à descoberta de Plutão em 1930 no observatório Lowell, no Arizona (criado em 1905 com o objetivo principal da busca do planeta transnetuniano).

(3) Algumas pessoas, duvidando da necessidade de desenvolver procedimentos estatísticos robustos, se perguntam: porque após a detecção dos outliers não se os separa e se analisa as observações restantes segundo alguma técnica convencional? Esta conduta é inadequada mesmo no caso, relativamente simples, em que todas as observações seguem o modelo assumido, com exceção de um único outlier. Muitas vezes um outlier é reconhecido porque ele tem um resíduo muito grande em relação ao ajuste, mas, mesmo no caso simples acima descrito, isto nem sempre acontece. Tal é o caso dos outliers de alto "leverage" em regressão: estes pontos tem "tanta força" que fazem que a reta de regressão de mínimos quadrados passe bem "perto" deles, fazendo, em consequência, que vários pontos "bons" fiquem "longe" da reta (com resíduos muito grandes) e, portanto, sendo caracterizados como outliers. A caracterização dos outliers através de um resíduo grande deve ser feita somente após um "ajuste robusto". Além disso, a rejeição de outliers coloca o problema das falsas rejeições e falsas retenções, como também o da especificação do modelo a partir do qual se aceita ou rejeita uma observação.

1.C.3. Relações entre robustez e estatística não paramétrica.

A área de contato entre os procedimentos robustos e a estatística não paramétrica não é muito grande. Por exemplo, os estimadores não paramétricos da mediana e da média são respectivamente a mediana amostral e a média amostral, sendo que na presença de poucos outliers o primeiro é insensível e o segundo altamente influenciável.

De um modo geral, a estatística não paramétrica trabalha com o conjunto de todas as distribuições possíveis, que usualmente (por exemplo no caso de distribuições sobre o conjunto dos números reais) pode ser pensado como um espaço de dimensão infinita. Algumas vezes são colocadas algumas restrições (simetria, unimodalidade, etc.) que reduzem o espaço anterior em algumas dimensões.

A estatística robusta está intimamente relacionada à estatística paramétrica, cuja característica mais notável reside em que seus modelos reduzem a descrição de grandes quantidades de dados a uns poucos parâmetros, trabalhando em subespaços de dimensão finita do espaço das probabilidades (geralmente conjuntos magros). A estatística paramétrica clássica fornece procedimentos ótimos, desde que todos os pressupostos sejam estritamente satisfeitos. A estatística robusta toma um modelo paramétrico e assume apenas que os pressupostos ocorrem de uma forma "aproximada", ou seja, trabalha numa "vizinhança" do modelo paramétrico. A estatística robusta *"é a estatística dos modelos paramétricos aproximados"* (Hampel, 1986, p. 7).

Para esclarecer um pouco estes conceitos, estudaremos um exemplo simples em que o espaço de todas as probabilidades seja um subconjunto de um espaço de dimensão finita.

Consideremos o conjunto $\Omega = \{0,1,2\}$, seja $\mathcal{A}$ a σ -álgebra das partes de Ω e seja $\mathcal{F} = \{\text{medidas de probabilidade em } (\Omega, \mathcal{A})\}$. Neste caso $\mathcal{F}$ pode ser pensado como:

$$\mathcal{F} = \left\{ (x, y, z) \in \mathbb{R}^3 : \begin{array}{l} x+y+z = 1 \\ x \geq 0, y \geq 0, z \geq 0 \end{array} \right\}, \text{ ou equivalentemente}$$

$\mathcal{F} = \left\{ (x, y, z) \in \mathbb{R}^3; \begin{matrix} x+y+z \leq 1 \\ x \geq 0, \\ y \geq 0, \\ z \geq 0 \end{matrix} \right\}$, onde x, y e z representam as probabilidades de $\{0\}, \{1\}$ e $\{2\}$, respectivamente.

Seja agora $B(2, t) \subseteq \mathcal{F}$ o conjunto das probabilidades binomiais de parâmetros 2 e t ($0 \leq t \leq 1$), então

$$B(2, t) = \left\{ (x, y, z) \in \mathbb{R}^3; \begin{matrix} x=t^2, & y=2t(1-t), & z=t^2 \\ 0 \leq t \leq 1 \end{matrix} \right\}, \text{ ou equivalentemente}$$

$$B(2, t) = \left\{ (x, z) \in \mathbb{R}^2; \begin{matrix} x=t^2, & z=(1-t)^2 \\ 0 \leq t \leq 1 \end{matrix} \right\}.$$

Podemos informalmente dizer que a estatística não paramétrica trabalha no conjunto convexo $\mathcal{F}$, a estatística paramétrica na curva $B(2, t)$ (que é um subconjunto magro de $\mathcal{F}$), e a estatística robusta trabalha numa vizinhança do modelo $B(2, t)$.

O significado estrito das palavras "aproximado" e "vizinhança" será posteriormente esclarecido, como também as afirmações sobre dimensionalidade e conjuntos magros feitas acima (Veja capítulos 4 e 5).

1.C.4. Os objetivos e propriedades desejáveis dos procedimentos robustos.

Segundo Hampel (1986, p. 11), pensando em termos de análise de dados, os principais objetivos da estatística robusta são: -

- (1) *descrever estruturas que se ajustem à massa dos dados;*

(2) identificar os pontos excepcionais (outliers) e também as diferentes estruturas presentes (clusters, etc.) nos dados para um posterior tratamento diferenciado;

(3) identificar e ter presente, para o uso desta ou daquela ferramenta, os pontos que têm uma influência muito grande (pontos de alto "leverage") para o caso de estruturas de dados não balanceadas (por exemplo em regressão);

(4) considerar, no contexto de séries de tempo, os desvios das estruturas de correlação assumidas, particularmente a hipótese de independência.

Para Huber (1981, p. 5), trabalhando no contexto dos modelos paramétricos, um bom procedimento robusto deve ter as seguintes propriedades:

(1) eficiência quase ótima no modelo assumido;

(2) o seu desempenho deve ser pouco prejudicado por pequenos desvios dos supostos do modelo;

(3) desvios moderadamente grandes não devem causar catástrofes.

1.D. UM POUCO DE HISTÓRIA.

A seguir, será feita uma breve resenha histórica, que não pretende ser exaustiva, mas que irá apenas mostrar alguns dos fatos e trabalhos cujos conhecimentos nos surpreenderam. Alguns deles são as pedras fundamentais da estatística robusta posterior a 1960 (Huber 1964, Hampel 1971 e 1974), outros têm a virtude de serem muito claros e didáticos e, portanto, fundamentais na divulgação da problemática da robustez, como por exemplo Tukey (1960 e 1975) e Huber (1972).

O conteúdo desta seção está baseado nos trabalhos de Huber (1972), Stigler (1973), Rey (1978) e Hald (1990).

1.D.1. Os primeiros passos.

Devido em grande parte ao desenvolvimento da computação a área de robustez começou a receber muita atenção por parte dos estatísticos a partir de 1950 ou 1960. Mas a preocupação de muitos cientistas pela robustez das técnicas estatísticas é muito antiga.

Rey (1978) chama a atenção para um escrito de Tucídides do ano 428 antes de Cristo onde se reporta do uso da moda como estimador de locação, justamente por ser pouco sensível a certas alterações nos poucos dados disponíveis.

Já mais perto de nossa época, a prática de descartar outliers era de uso comum pelo menos 200 anos atrás. Stigler (1973) cita o caso de Legendre (1805, "On the method of least squares") que advoga explicitamente:

"Se entre estes erros estão alguns que parecem muito grandes para serem admissíveis, então aquelas equações que produziram estes erros serão rejeitadas, como provenientes de experimentos muito defeituosos...".

Huber cita um trabalho anônimo de 1821 (Dissertation sur la recherche du milieu le plus probable, *Ann. Math. Pures et Appl.* 12, 181-204), no qual afirma-se que:

"Há certas províncias na França onde, para determinar a produção média de uma propriedade rural, costuma-se observar esta produção durante vinte anos consecutivos, descartar a melhor e a pior produção e tomar a média dos anos restantes".

1.D.2. A média, a distribuição gaussiana e o autoritarismo.

Certamente a distribuição gaussiana, também chamada normal, é o modelo mais famoso e utilizado da estatística.

Mas, por que um modelo paramétrico seria tão importante do ponto de vista da robustez? Há várias razões, entre outras:

(1) porque, em geral, a estatística robusta trabalha numa vizinhança dos modelos paramétricos;

(2) porque, sendo o mais conhecido, é o mais utilizado na literatura, o que inclui a documentação de muitos casos de uso inadequado.

Huber (1972) popularizou o trabalho no qual Gauss introduz a distribuição normal (1821, "Göttingische gelehrte Anzeigen", p. 321-327).

Antes de analisar uma parte desta publicação, é bom esclarecer que Gauss estava trabalhando no modelo de medição: $X = \alpha + e$, onde a variável X representa as medições de uma distância α desconhecida e toda a variabilidade de X deve-se ao erro de medição e .

O modelo de medição era muito usado pelos astrônomos. Galileo Galilei, na sua obra "Diálogo" (1632, citada em Anders Hald, 1990) diz a respeito:

(1) *existe um único número que dá a distância da estrela até o centro da terra, a verdadeira distância,*

(2) *todas as observações têm erros, devido ao observador, aos instrumentos e demais condições observacionais,*

(3) *as observações estão distribuídas simetricamente em torno do verdadeiro valor, os erros estão distribuídos simetricamente em torno de zero,*

(4) *pequenos erros são mais freqüentes que grandes erros.*

Observa-se que os modelos de distribuição das observações dados por densidades em forma de sino satisfazem as condições de Galileo. Observa-se, também, que não há, aqui, evidências de preocupação com modelos.

Voltemos agora a Gauss (1821, op. cit.):

"O autor do presente tratado, quem, no ano de 1797, primeiro investigou este problema de acordo aos princípios da teoria das probabilidades, logo se deu conta que era impossível determinar o valor

mais provável da quantidade desconhecida, a menos que a função representante dos erros fosse conhecida. Mas como isto não acontecia, não houve outro recurso que supor tal função de uma maneira hipotética. Pareceu-lhe natural tomar o caminho oposto e buscar a função que deve ser tomada como base a fim de que, para o caso mais simples de todos, se obtenha um procedimento ou regra geralmente aceita como boa: simplesmente que o valor mais provável é a média aritmética de muitas observações de um mesmo fenômeno repetido em iguais condições. Isso implicou que a probabilidade de um erro deveria ser tomada como proporcional a uma expressão exponencial da forma $\exp(-h^2 x^2)$. Este método, que posteriormente, em particular desde 1801, ele teve oportunidade de usar quase diariamente em diversas computações astronômicas, e que entretanto também Legendre encontrou por acaso, agora é de uso geral sob o nome de método dos quadrados mínimos".

O fato mais notável deste trabalho é que Gauss introduz o modelo chamado normal a partir da aceitação axiomática da excelência da média amostral (que é o estimador de quadrados mínimos do parâmetro α do modelo de medição acima descrito) e deduzindo a partir deste fato a forma que deveria ter a função de densidade da distribuição dos erros.

Huber atribui a um mal entendimento do teorema de Gauss- Markov e do Teorema Central do Limite a confiança ilimitada na média amostral (e nos estimadores de quadrados mínimos em geral).

O primeiro afirma que se a variância populacional for finita a média amostral é o estimador de variancia mínima entre os estimadores lineares e não viciados da média populacional. O resultado vale em geral para o

estimador de quadrados mínimos de qualquer combinação linear das observações, em caso de variância finita da distribuição dos dados.

É muito interessante a forma em que Tukey (1972) detalha o enxuto enunciado do teorema de Gauss-Markov do parágrafo anterior agrupando os supostos em sete itens, que não são evidentes em uma leitura ingênua:

"(1) se o valor central de y para x 's dados é $\sum \beta_i x_i$, onde os β_i são completamente desconhecidos;

(2) se conhecermos os x 's sem erros;

(3) se a variância do j -ésimo y é σ^2/w_j , onde w_j é conhecido mas σ^2 não é;

(4) se as covariâncias dos diferentes y 's são zero, (é possível, obviamente, substituir zero por $h_{jk}\sigma^2$ com h_{jk} conhecidos complicando um pouco a resposta);

(5) se estivermos dispostos a permanecer confinados a estimadores dos β_j 's que sejam funções lineares dos y 's dados;

(6) se estivermos dispostos a considerar os valores centrais em (1) como sendo médias e nos restringirmos a considerar estimadores não viciados em média;

(7) se estivermos dispostos a julgar a qualidade de nossos estimadores por suas variâncias (para o conjunto dado de x 's)".

A conclusão é a usual: "então o melhor estimador de qualquer β ou qualquer combinação linear dos β deve ser obtido por quadrados mínimos".

Posteriormente Tukey analisa os perigos do afastamento destes pressupostos.

A versão mais popular do Teorema Central do Limite (T.C.L.) é a de Liapunov, que afirma que, quando o número de somandos tende para infinito, a distribuição limite da soma de variáveis aleatórias independentes identicamente distribuídas (i.i.d.) é normal se as variáveis tiverem variância finita. Uma leitura superficial (que em geral só lembra da tese dos teoremas) deste enunciado leva a interpretações erradas do seguinte tipo: "quase toda variável pode ser decomposta como soma de pequenos efeitos, portanto quase toda variável é aproximadamente normal".

Existem versões mais novas onde as hipóteses do T.C.L. (condições tipo Lindeberg, por exemplo) são menos restritivas, mas nem por isso de fácil verificação. Na maioria dos casos concretos, a invocação do T.C.L. para concluir sobre a normalidade de uma variável é incorreta. Isto ocorre porque, em qualquer versão do T.C.L., a verificação das hipóteses é quase impossível e certamente bem mais complicada do que o teste direto da normalidade da variável em questão.

Há também o resultado de que, sob normalidade, a média amostral é o estimador da média populacional de mínima variância entre os não viciados (e não apenas o melhor entre os estimadores não viciados e lineares). Também, sob normalidade, o estimador de quadrados mínimos coincide com o estimador de máxima verossimilhança.

Ou seja, existe uma série de resultados muito importantes que, quando considerados superficialmente, levam a argumentações circulares que

justificam a média pela normalidade e vice-versa. A confusão é tão grande que o físico francês Lippman, prêmio Nobel em 1908, afirmou:

"todos acreditavam na distribuição normal: os matemáticos porque pensavam que era um fato experimental, os experimentadores porque pensavam que era um teorema matemático..."

Os fatos acima expostos, o enorme prestígio de Gauss e o autoritarismo nos ambientes científicos, levaram ao uso pouco cuidadoso dos estimadores de quadrados mínimos e do modelo normal. Muitos conjuntos de dados foram considerados "normais" apesar de evidências claras em contrário. Outras vezes, conjuntos completos de dados reais foram desconsiderados porque não se adequavam à curva normal.

1.D.3. Outras tentativas.

Aparentemente (segundo Stigler (1973)) o primeiro trabalho matemático sobre estimação robusta seria o de Laplace (1818, Deuxieme Supplement a la Théorie Analitique des Probabilités), que estudou a mediana e, mais geralmente, funções lineares das estatísticas de ordem, dando inclusive condições necessárias e suficientes sobre a distribuição dos erros para que a mediana amostral tenha menor variância assintótica que a média amostral.

O primeiros trabalhos a respeito de critérios matemáticos sobre rejeição de outliers são devidos a Benjamin Peirce (1852, Criterion for

the Rejection of Doubtful Observations, *Astronomical Journal*, 2, 161-3) e Chauvenet (1863, *Manual of Practical Astronomy*, Philadelphia).

A filosofia por trás da rejeição também está muito ligada ao modelo normal e apresenta os seguintes problemas, como bem observa Huber (1972):

(1) *a classificação das observações em duas classes, as "normais" e os outliers, é artificial e não faz sentido se a distribuição subjacente tiver caudas mais pesadas do que a normal;*

(2) *pretende-se "limpar" os dados através da rejeição dos outliers, mas este processo leva a falsas retenções e falsas rejeições e portanto é errado assumir que os dados retidos são normais.*

A primeira tentativa de lidar com distribuições de caudas pesadas e abandonar o modelo normal deve-se a Newcomb (1886, *A generalized theory of the combination of observations so as to obtain the best result*, *Amer. J. Math.*, 8, 343-366). Sua proposta de modelo para a distribuição dos erros consiste numa mistura de normais (ele tinha observado que muitas vezes trabalhava com observações de diferentes graus de precisão).

Em 1912 Newcomb estimou a média populacional através de uma média ponderada das observações $x_1, x_2, \dots, x_n$ com os pesos respectivos $w_1, w_2, \dots, w_n$ dados por:

$$w_j = c / \text{máximo}\{|x_j - \bar{x}|, c\} \quad 1 \leq j \leq n, \text{ sendo } c \text{ uma constante positiva a determinar;}$$

é importante salientar que, chamando agora $\bar{x}$ a $\sum_{j=1}^n w_j x_j$ e iterando sucessivamente, o procedimento converge a o M-estimador definido por Huber em 1964 com função $\phi(t) = \text{máximo}\{-c, \text{mínimo}(t, c)\}$, que será visto em 1.D.4.b).

1.D.4. Os principais conceitos desenvolvidos no passado próximo.

Segundo Martin and Yohai (1984) os principais conceitos no contexto de estimação robusta (sob a hipótese de observações i.i.d.) ultimamente desenvolvidos são, em ordem de aparição: *robustez em eficiência*, *robustez minimax* e *robustez qualitativa*.

a) *Robustez em eficiência* (Tukey, 1960).

Seja X uma variável aleatória real cuja distribuição segue o modelo $\{P_\theta, \theta \in \Theta\}$, onde $\Theta \subseteq \mathbb{R}$. Seja $x_1, x_2, \dots, x_n$ uma amostra aleatória simples de X e $T_n(x_1, x_2, \dots, x_n)$ um estimador do parâmetro θ . Como usualmente define-se eficiência de T_n no modelo P_θ por:

$$\text{Eficiência}(T_n, P_\theta^n) =$$

$$\text{Var}(\text{estimador fixo ótimo}) / \text{Var}_{P_\theta^n}(T_n),$$

onde Var denota variância e P_θ^n a distribuição produto $(P_\theta)^n$.

Se o objetivo for a eficiência assintótica o estimador será denotado por T e a medida (produto) do processo $(x_n)_{n \in \mathbb{N}}$ por P_θ^∞ .

$$\text{Eficiência}(T, P_{\theta}^{\infty}) =$$

$$\text{Var assintótica (estimador fixo ótimo)} / V_{\infty}(T),$$

onde $V_{\infty}(T)$ = variância assintótica de $(n)^{1/2}T_n$ em P_{θ}^{∞} .

Seja $\mathcal{P} = \{P_{\theta,1}, P_{\theta,2}, \dots, P_{\theta,m}\}$ um conjunto de modelos distribucionais

onde $\theta_j \in R$, $1 \leq j \leq m$. Então T (ou T_n) é robusto em eficiência (para $\mathcal{P}$) se e somente se T tem alta eficiência para todos os $P_{\theta,j}$, $1 \leq j \leq m$.

Este conceito pode ser generalizado ao caso não i.i.d. (substituindo a medida produto pela distribuição conjunta) e também para parâmetros vetoriais mediante uma definição apropriada de eficiência multivariada.

b) *Robustez minimax (Huber, 1964).*

Huber (1964) introduziu os M-estimadores, dando uma contundente resposta ao desafio de como se construir procedimentos robustos. Se $x_1, x_2, \dots, x_n$ é uma amostra aleatória simples de uma variável aleatória X com distribuição dependente do parâmetro de locação θ , o M-estimador de θ é definido como o valor $\hat{\theta}$ que minimiza:

$$\sum_{j=1}^n \rho(x_j - \hat{\theta}), \text{ onde } \rho(0)=0, \rho \text{ é par, convexa e crescente em } (0, \infty);$$

ou, equivalentemente, $\hat{\theta}$ é a solução de:

$$\sum_{j=1}^n \psi(x_j - \hat{\theta}) = 0, \quad (*)$$

onde $\psi = \rho'$.

A equação (*) tem solução essencialmente única. Mais precisamente, Huber (op. cit.), provou que a convexidade de ρ implica em que o conjunto de soluções de (*) é um intervalo compacto de R e que, em caso de convexidade estrita de ρ , a solução é única.

Para conseguir que os M-estimadores sejam pouco sensíveis à presença de outliers deve-se acrescentar a condição de que ψ seja limitada.

Para torná-los equivariantes por mudanças na escala das observações a equação (*) é substituída por:

$$\sum_{j=1}^n \psi \left(s^{-1} (x_j - \hat{\theta}) \right) = 0$$
 , onde s é um estimador robusto de escala, por exemplo pode-se escolher

$$s = \text{mediana} \left\{ |x_j - \text{mediana}\{x_1, x_2, \dots, x_n\}| \right\}_{1 \leq j \leq n} .$$

Os M-estimadores formam uma classe muito rica, que entre outros contém o estimador de quadrados mínimos ($\rho(t)=t^2$, que é a média amostral), a mediana amostral ($\rho(t)=|t|$) e também o de máxima verossimilhança ($\rho(t)=-\logaritmo(f(t))$, onde f é a função de densidade das x).

Huber demonstrou a consistência e normalidade assintótica destes estimadores no modelo de posição sob condições bastante gerais. O leitor interessado pode consultar o trabalho citado de Huber ou também Huber (1981).

Apesar de terem sido concebidos no marco de estimação de um parâmetro de locação, a utilidade dos M-estimadores é muito grande, já que eles podem ser adaptados ao contexto de regressão (mediante sua aplicação

aos resíduos) através de convenientes modificações. Por exemplo, para garantir que o M-estimador seja pouco influenciado por outliers de alto leverage, costuma-se exigir que ρ também seja limitada. Neste caso ψ deve ser redescendente, com o que perde-se as propriedades de que ρ seja convexa e crescente com $|t|$ e conseqüentemente a "quase" unicidade da solução de (*).

Os M-estimadores são também aplicáveis ao contexto de séries de tempo com as modificações adequadas. Por exemplo é possível obter robustez em eficiência mediante o uso de M-estimadores nos modelos ARMA, mas estes estimadores não são qualitativamente robustos no sentido em que será definido Capítulo 5.

Posteriormente, para obter propriedades de robustez, foram introduzidos os *M-estimadores generalizados* ou *GM-estimadores*, que não dependem apenas dos resíduos das observações mas também do seu "leverage". Eles são definidos como o valor do parâmetro θ que minimiza

$$\sum_{j=1}^n \eta \left(s^{-1}(r_j(\theta)), \|x_j\| \right),$$

onde $\eta(r, x)$ é par e monótona não decrescente em r , x_j são as observações, $r_j(\theta)$ são os resíduos e $\| \cdot \|$ denota uma norma conveniente. Estes estimadores são qualitativamente robustos e resistentes (estes conceitos serão definidos no Capítulo 5) em estimação de processos AR se η for contínua e limitada. O trabalho de Martin and Yohai (1984) contém uma ~~resenha~~ destes assuntos.

Voltemos a Huber (1964). Neste famoso trabalho ele também introduziu o modelo dos *erros grosseiros* ou de *contaminação*, assumindo que a distribuição das observações era dada por:

$F = (1-\varepsilon)G + \varepsilon H$, onde a distribuição G representava o modelo, H a "contaminação" e ε a proporção de contaminação (G e ε fixos e F varia com H).

Huber pensou o problema de estimação como um jogo, onde a "natureza" escolhe uma F e o estatístico um M -estimador (dado por uma ψ). A função perda (do ponto de vista do estatístico) é dada por $V(\psi, F)$, a variância assintótica do estimador ψ na situação F . Huber provou que, sob condições bastante razoáveis, a variância assintótica tem um ponto de sela, isto é existe um par (ψ_0, F_0) tal que:

$$\sup_F V(\psi_0, F) = V(\psi_0, F_0) = \inf_{\psi} V(\psi, F_0),$$

Huber provou que, se o modelo G é normal, então o par (ψ_0, F_0) satisfaz:

$$\psi_0(t) = \text{máximo} \left(-k, \text{mínimo}(k, t) \right)$$

ou equivalentemente

$$\rho_0(t) = \begin{cases} (1/2)t^2_{se} & |t| < k \\ k|t| - (1/2)k^2_{se} & |t| \geq k \end{cases}$$

e F_0 tem densidade

$$f_0(t) = (1-\varepsilon)(2\pi)^{-1/2} e^{-\rho(t)},$$

onde $k > 0$ e ε e k estão relacionados através da fórmula:

$$\left((2-\varepsilon)/(2-2\varepsilon) \right)^{-1} = \phi(k) - k^{-1} \phi'(k),$$

onde ϕ é a função de distribuição acumulada da normal(0,1).

c) *Robustez qualitativa e conceitos associados (Hampel, 1968, 1971 e 1974).*

Hampel introduziu a chamada "*abordagem infinitesimal*" na teoria de robustez, baseando-se nos trabalhos de Volterra, Von Mises e Prohorov, cujas contribuições permitiram unificar a teoria assintótica e a de amostras finitas.

Esta abordagem está baseada em três conceitos fundamentais: robustez qualitativa, função de influência e ponto de ruptura, que, em um sentido geral, correspondem às noções de continuidade, primeira derivada e distância à singularidade mais próxima de uma função.

Huber (1972) estabeleceu a seguinte analogia dos conceitos acima e propriedades de estabilidade de uma estrutura mecânica: robustez qualitativa significa que pequenas perturbações produzem pequenos efeitos, a função de influência mede os efeitos de perturbações infinitesimais e o ponto de ruptura informa até quando é possível perturbar sem quebrar a estrutura.

A robustez qualitativa, que, segundo dito nos parágrafos acima, é uma propriedade relacionada com continuidade será abordada no Capítulo 5.

A seguir serão apresentados brevemente os conceitos de curva de influência e ponto de ruptura, sendo que a primeira será revista no Capítulo 4 num contexto mais geral.

Suponhamos ter uma amostra aleatória simples $(\omega_1, \omega_2, \dots, \omega_n)$ de uma variável aleatória cuja distribuição depende de um parâmetro desconhecido $\theta \in \mathbb{R}^k$.

Por definição, um estimador, é uma sequência $(T_n)_{n \in \mathbb{N}}$ de aplicações mensuráveis:

$$T_n: (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\mathbb{R}^k, \mathcal{B}^k) \quad \forall n \in \mathbb{N}, \text{ onde } \mathcal{B}^j = \{\text{borelianos de } \mathbb{R}^j\}.$$

Muitos estimadores podem ser definidos através processo descrito a seguir. Sejam:

i) $\mathcal{F} = \{\text{medidas de probabilidade em } (\mathbb{R}, \mathcal{B})\},$

ii) $\mathcal{C}$ subconjunto convexo de $\mathcal{F}$ tal que $\mathcal{C} \supseteq \left\{ \begin{smallmatrix} \text{distribuições} \\ \text{empíricas} \end{smallmatrix} \right\},$ ou seja

$$\mathcal{C} \supseteq \left\{ n^{-1} \sum_{j=1}^n 1_{\omega_j} : \omega_j \in \mathbb{R}, n \in \mathbb{N} \right\} \quad (1_{\omega} \text{ denota a medida de massa 1 em } \omega), \text{ e}$$

iii) T um funcional $T: \mathcal{C} \rightarrow \mathbb{R}^k$; nestas condições, T induz um estimador invariante por permutações das coordenadas definido por

$$T_n(\omega_1, \omega_2, \dots, \omega_n) = T\left(n^{-1} \sum_{j=1}^n 1_{\omega_j}\right) \quad \forall (\omega_1, \omega_2, \dots, \omega_n) \in \mathbb{R}^n \quad \forall n \in \mathbb{N},$$

caso as aplicações T_n acima definidas sejam mensuráveis $\forall n \in \mathbb{N}$.

Abusando um pouco da notação, denotaremos a este estimador por T .

Seja $F \in \mathcal{C}$, Hampel chamou curva de influência do estimador T na distribuição F à função $IC(T, F): \mathbb{R}^k \rightarrow \mathbb{R}^k$ definida por:

$$\begin{aligned} (IC(T, F))(\omega) &= \lim_{t \rightarrow 0^+} \left[T(F + t \cdot (1_\omega - F)) - T(F) \right] \cdot t^{-1} = \\ &= \lim_{t \rightarrow 0^+} \left[T((1-t)F + t \cdot 1_\omega) - T(F) \right] \cdot t^{-1}, \quad \omega \in \mathbb{R}, \end{aligned}$$

desde que o limite acima exista $\forall \omega \in \mathbb{R}$.

A $IC(T, F)$ fundamentalmente serve para:

i) avaliar a velocidade de mudança do valor de T quando o modelo F é contaminado por uma distribuição de massa 1 em ω ($\forall \omega \in \mathbb{R}$, a influência máxima neste aspecto é dada por:

$$GES(T, F) = \sup_{\omega \in \mathbb{R}} \| (IC(T, F))(\omega) \|,$$

chamada por Hampel *sensibilidade a erros grosseiros* (de T em F).

Será vista no Capítulo 4 a relação, sob certas condições, da $GES(T, F)$ com o vício assintótico de T e conseqüentemente a sua utilidade na definição dos estimadores com vício minimax, introduzidos por Huber (1964);

ii) calcular a variância assintótica de T , já que, sob determinadas condições, vale que esta é igual a $E_F \{ (IC(T, F))^2 \} = \int [(IC(T, F))(\omega)]^2 dF(\omega)$.

A curva de sensibilidade de Tukey, usualmente denotada $(SC(\omega_1, \omega_2, \dots, \omega_n))(\omega)$ ou $SC_n(\omega)$ e definida por:

$$(n+1) (T_{n+1}(\omega_1, \omega_2, \dots, \omega_n, \omega) - T_n(\omega_1, \omega_2, \dots, \omega_n)),$$

é uma versão amostral da $IC(T,F)$ que permite estudar a influência de uma nova observação ω relativa à massa introduzida $(1/n+1)$ no valor do estimador.

A curva de influência (e todas as medidas dela derivadas) da uma idéia do comportamento local de um estimador. Introduziremos agora o conceito de ponto de ruptura, que é uma medida da confiabilidade global do estimador.

Sejam (R,B) o espaço das observações, $(T_n)_{n \in \mathbb{N}}$ um estimador e $F \in \mathcal{F}$, Hampel (1971) definiu o "ponto de ruptura δ^* do estimador $(T_n)_{n \in \mathbb{N}}$ em F " por:

$$\begin{aligned} \delta^* &= \delta^* (\{T_n\}_{n \in \mathbb{N}}, F) = \\ &= \sup \left\{ \delta \leq 1 : \exists K(\delta) \text{ compacto em } R^k \text{ tal que} \right. \\ &\quad \left. \Pi(F, G) < \delta \Rightarrow \lim_n G(T_n \in K(\delta)) = 1 \right\}, \end{aligned}$$

onde $\Pi = \Pi_d$ denota a métrica de Prohorov em $\mathcal{F}$ associada a d . Esta métrica será definida no Capítulo 3, onde serão exibidas suas propriedades; aqui apenas basta saber que $\text{diâmetro}(\mathcal{F}, \Pi) \leq 1$ e que Π induz em $\mathcal{F}$ a topologia da convergência em distribuição.

Se $(T_n)_{n \in \mathbb{N}}$ for um estimador definido através de um funcional T com domínio em $\mathcal{F}$, ou seja se:

$$T_n(\omega_1, \omega_2, \dots, \omega_n) = T\left(n^{-1} \sum_{j=1}^n 1_{\omega_j}\right) \quad \forall (\omega_1, \omega_2, \dots, \omega_n) \in R^n \quad \forall n \in \mathbb{N},$$

o ponto de ruptura pode ser definido como em Hampel (1968):

$$\delta^*(T, F) = \sup \left\{ \delta : \text{diâmetro } T\{G : \Pi(F, G) < \delta\} < \infty \right\}.$$

Intuitivamente, o ponto de ruptura δ^* indica a que distância de F (o modelo) o estimador deixa de ter qualquer utilidade.

CAPÍTULO 2.

CONVERGÊNCIA FRACA DE MEDIDAS EM ESPAÇOS MÉTRICOS.

A TOPOLOGIA FRACA.

2.A. INTRODUÇÃO E RESUMO DO CAPÍTULO.

Neste capítulo serão transcritos os enunciados de alguns resultados fundamentais e bastante conhecidos referentes à topologia fraca no espaço M^+ das medidas finitas definidas nos borelianos de um espaço métrico (Ω, d) . As demonstrações, restritas a medidas de probabilidade, estão, com algumas exceções, nos textos clássicos de Billingsley (1968) ou Parthasarathy (1967). A extensão destes resultados ao âmbito mais geral das medidas finitas é, na maioria dos casos, trivial, não sendo, porém, facilmente encontrada na literatura. Por este motivo generalizaremos aqui esses resultados ao contexto das medidas finitas.

A Proposição 2.13 é de Prohorov (1958). Nossas contribuições são:

-o lema 2.11, onde provamos (2.11.a) que, com exceção de convergência à medida nula, a convergência fraca de medidas finitas se reduz à convergência fraca de medidas de probabilidade, em 2.11.b) damos uma caracterização simples da convergência à medida nula;

-a demonstração da Proposição 2.25.b) (separabilidade de $\mathcal{M}^+$ em caso de (Ω, d) separável).

A fim de que a exposição seja mais fluída, as demonstrações foram colocadas na última seção. Esta norma será adotada em forma geral.

2.B. MEDIDAS FINITAS NOS BORELIANOS DE UM ESPAÇO MÉTRICO.

NOTAÇÃO.

(Ω, d) denotará um espaço métrico, onde Ω é um conjunto não vazio e d uma métrica em Ω ;

$\mathcal{B}(\Omega, d) = \{\text{borelianos de } (\Omega, d)\}$, quando não haja possibilidade de confusão será denotado simplesmente por $\mathcal{B}$;

$\mathcal{P} = \{\text{medidas de probabilidade sobre } (\Omega, \mathcal{B})\}$;

$\mathcal{M}^+ = \{\text{medidas finitas sobre } (\Omega, \mathcal{B})\}$;

$\mathbb{R}$ denotará o conjunto dos números reais e $(\mathbb{R}, e)$ o espaço métrico onde e é a métrica euclideana usual em $\mathbb{R}$;

$$\mathcal{EL}(\Omega, d) = \left\{ f: (\Omega, d) \rightarrow (R, e) : \begin{array}{c} \text{contínua} \\ \text{limitada} \end{array} \right\} ;$$

$$\mathcal{UCL}(\Omega, d) = \left\{ f: (\Omega, d) \rightarrow (R, e) : \begin{array}{c} \text{uniformemente contínua} \\ \text{limitada} \end{array} \right\} .$$

PROPOSIÇÃO 2.1.

Seja (Ω, d) um espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$ e M uma medida finita em $(\Omega, \mathcal{B})$. Então:

(a) $\forall C \in \mathcal{B}$ vale que $A^- = \text{aderência}(C)$ é um G_δ (intersecção de uma família enumerável de abertos) e que $C^\circ = \text{interior}(C)$ é um F_σ (união de uma família enumerável de fechados), conseqüentemente todo conjunto fechado é um G_δ e todo aberto um F_σ ;

$$(b) \quad F \text{ fechado} \Rightarrow M(F) = \inf \left\{ M(A) : A \text{ aberto e } A \supseteq F \right\}$$

e também

$$G \text{ aberto} \Rightarrow M(G) = \sup \left\{ M(B) : B \text{ fechado e } B \subseteq G \right\};$$

(c) $\forall C \in \mathcal{B} \quad \forall \epsilon > 0$ existem F fechado e G aberto tais que:

$$F \subseteq C \subseteq G \quad \text{e} \quad M(G-F) < \epsilon;$$

(d) todo boreliano é "regular", isto é $\forall C \in \mathcal{B}$ vale que:

$$M(C) = \inf \left\{ M(G) : G \text{ aberto, } G \supseteq C \right\} = \sup \left\{ M(F) : F \text{ fechado, } F \subseteq C \right\}.$$

(e) M medida finita em $(\Omega, \mathcal{B})$, então M fica completamente determinada por $\{M(F) : F \text{ fechado}\}$ ou equivalentemente por $\{M(G) : G \text{ aberto}\}$.

Comentário: em Billingsley (1968, p. 7) está provada esta proposição para o caso em que M é uma probabilidade, a generalização a medidas finitas não oferece problemas. A afirmação a) é de natureza estritamente topológica. $\square$

OBSERVAÇÃO 2.2.

É fácil verificar que a hipótese de que Ω seja métrico em 2.1.(b,c,d,e) pode ser enfraquecida; basta que Ω seja um espaço topológico com a seguinte propriedade: todo fechado é um G_δ ou equivalentemente todo aberto é um F_σ .

PROPOSIÇÃO 2.3.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$.

(a) Sejam $F \in \mathcal{B}$, F fechado e $\varepsilon > 0$,

existe uma função $f \in UCL(\Omega, d)$ tal que:

$$0 \leq f \leq 1, \quad f/F = 1 \quad \text{e} \quad f/\{\omega \in \Omega : d(\omega, F) \geq \varepsilon\} = 0;$$

(b) M e N medidas finitas em $(\Omega, \mathcal{B})$, então:

$$\int f dM = \int f dN \quad \forall f \in UCL(\Omega, d) \Rightarrow M = N.$$

Comentário: a afirmação (a) é de caráter estritamente topológico e mostra que os espaços métricos são normais (ou seja que dados dos

conjuntos fechados disjuntos existem vizinhanças disjuntas desses conjuntos), a função f pode ser definida por:

$$f(\omega) = d(\omega, F) / [d(\omega, F) + d(\omega, A(\varepsilon))],$$

onde $A(\varepsilon) = \{\tau \in \Omega : d(\omega, F) \geq \varepsilon\}$;

a parte (b) é consequência de (a) e o teorema de convergência dominada de Lebesgue, válido para medidas em geral (inclusive não finitas).

Veja a demonstração completa em 2.G.□

DEFINIÇÃO 2.4.

$(\Omega, \mathcal{T})$ espaço topológico, $\mathcal{B} = \{\text{borelianos de } (\Omega, \mathcal{T})\}$, $\mathcal{M}^+ = \{\text{medidas finitas sobre } (\Omega, \mathcal{B})\}$, $M \in \mathcal{M}^+$.

M é tight $\stackrel{\text{DEF}}{\Leftrightarrow} \forall \varepsilon > 0 \exists K_\varepsilon$ compacto tal que $M(K_\varepsilon^c) < \varepsilon$.

Comentário:

1) no capítulo seguinte este conceito será estendido a qualquer subconjunto de $\mathcal{M}^+$ e será, depois, utilizado na caracterização dos seus subconjuntos relativamente compactos;

2) M é tight $\Leftrightarrow M$ tem algum suporte σ -compacto, onde $\text{suporte}(M)$ é qualquer boreliano B tal que $M(B^c) = 0$.□

PROPOSIÇÃO 2.5.

(Ω, d) espaço métrico separável e completo,

$\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

M uma medida finita sobre (Ω, d) ,

então: M é tight.

Comentário: veja Billingsley (1988, p.10), a generalização para qualquer medida finita é trivial e consta em 2.G. $\square$

2.C. CONVERGÊNCIA FRACA DE MEDIDAS FINITAS NOS BORELIANOS DE UM ESPAÇO MÉTRICO.

DEFINIÇÃO 2.6.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+ = \{\text{medidas finitas em } (\Omega, \mathcal{B})\}$, $M \in \mathcal{M}^+$ e $\{M_n\}_{n \in \mathbb{N}} \subseteq \mathcal{M}^+$:

M_n converge fracamente a $M \stackrel{\text{DEF}}{\Leftrightarrow} \int f dM_n \rightarrow \int f dM \quad \forall f \in CL(\Omega, d)$;

a convergência fraca de M_n a M sera denotada por:

$M_n \xrightarrow{w} M$ (w pela palavra inglesa "weak").

OBSERVAÇÃO 2.7.

a) quando existe convergência fraca, o limite é único, dado o item (b) da Proposição 2.3;

b) o conceito de convergência fraca depende apenas da topologia e não da métrica particular que gera a topologia, dado que a definição depende da classe das funções contínuas.

PROPOSIÇÃO 2.8.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{F} = \{\text{medidas de probabilidade sobre } (\Omega, \mathcal{B})\}$, $\{P_n\}_{n \in \mathbb{N}} \subseteq \mathcal{F}$, $P \in \mathcal{F}$. São equivalentes:

a) $P_n \xrightarrow{w} P$,

b) $\int f dP_n \rightarrow \int f dP$, para toda $f \in UCL(\Omega, d)$,

c) $\forall F$ fechado $\limsup_n P_n(F) \leq P(F)$,

d) $\forall G$ aberto $\liminf_n P_n(G) \geq P(G)$,

e) $\lim_n P_n(A) = P(A) \forall A$ tal que $P(\delta A) = 0$, onde δA significa fronteira de $A = \bar{A} - A^\circ$, sendo $\bar{A} = \text{aderência}(A)$ e $A^\circ = \text{interior}(A)$.

Comentário: a demonstração deste resultado está em Billingsley (1968, p. 11). $\square$

PROPOSIÇÃO 2.9.

Dados $\{M_n\}_{n \in \mathbb{N}} \subseteq \mathcal{M}^+$, $M \in \mathcal{M}^+$, as seguintes condições são equivalentes:

a) $M_n \xrightarrow{w} M$,

b) $\int f dM_n \rightarrow \int f dM$, para toda $f \in UCL(\Omega, d)$,

c) $M_n(\Omega) \rightarrow M(\Omega)$ e $\forall F$ fechado $\limsup_n M_n(F) \leq M(F)$,

d) $M_n(\Omega) \rightarrow M(\Omega)$ e $\forall G$ aberto $\liminf_n M_n(G) \geq M(G)$,

e) $\lim_n M_n(A) = M(A) \forall A$ tal que $M(\delta A) = 0$.

Comentário: a demonstração do resultado decorre da Observação 2.10 e do Lema 2.11. $\square$

OBSERVAÇÃO 2.10.

A Proposição 2.8 é a versão da 2.9 restrita a medidas de probabilidade; portanto é trivial verificar que 2.9 implica 2.8.

O seguinte Lema 2.11 permite mostrar uma quase recíproca do resultado acima, que poderia ser assim enunciada:

$$\left. \begin{array}{l} \left\{ M_n \right\}_{n \in \mathbb{N}} \subseteq \mathcal{M}^+ \text{ e } M_n(\Omega) \xrightarrow{e} 0 \Rightarrow M \xrightarrow{w} \text{medida nula} \\ \text{e} \\ 2.8 \end{array} \right\} \Rightarrow 2.9;$$

isto mostra que a convergência fraca em $\mathcal{M}^+$ é determinada pela convergência fraca em $\mathcal{F}$, com exceção do caso da convergência à medida nula.

A condição $M_n(\Omega) \xrightarrow{e} M(\Omega)$, que aparece em c) e d) da Proposição 2.9, impede o "escape de massa" e ela está implícita em b) (considere $f \equiv 1$) e em e) (pois $\delta(\Omega) = \phi$).

LEMA 2.11.

a) $\left\{ M_n \right\}_{n \in \mathbb{N}} \subseteq \mathcal{M}^+$, $M \in \mathcal{M}^+$ e $M(\Omega) \neq 0$, então:

$$M_n \xrightarrow{w} M \text{ see } \left\{ \begin{array}{l} M_n(\Omega) \xrightarrow{em(R,e)} M(\Omega) \\ M_n/M_n(\Omega) \xrightarrow{e,w} M/M(\Omega) \end{array} \right. ;$$

b) $M_n \xrightarrow{w} \text{medida nula}$ see $M_n(\Omega) \xrightarrow{e} 0$.

As duas proposições seguintes fornecem caracterizações da convergência fraca que serão utilizadas posteriormente.

PROPOSIÇÃO 2.12.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, M_n e M medidas finitas em $(\Omega, \mathcal{B})$. Sejam $M_n h^{-1}$ e $M h^{-1}$ as medidas finitas definidas nos borelianos de R por

$$M_n h^{-1}(B) = M_n(h^{-1}(B)) \text{ e } M h^{-1}(B) = M(h^{-1}(B)).$$

Nestas condições:

$$M_n \xrightarrow{w} M \text{ se e só se } M_n h^{-1} \xrightarrow{w} M h^{-1} \quad \forall h: (\Omega, d) \xrightarrow[\text{limitada}]{\text{contínua}} (R, e).$$

Comentário: a demonstração para o caso de medidas de probabilidade está em Billingsley (1968), em 2.G está a generalização para o caso de medidas finitas, que não oferece o menor problema. Também é possível estender o resultado em outros aspectos. Por exemplo, pode-se verificar que a hipótese de continuidade de h poderia ser substituída por $M\{\text{discontinuidades de } h\} = 0$ e também $h \in CL(\Omega, d)$ por $h \in UCL(\Omega, d)$ (veja a demonstração em 2.G). $\square$

PROPOSIÇÃO 2.13.

(Ω, d) espaço métrico, $\mathcal{B}(\Omega, d) = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+ = \{\text{medidas finitas sobre } (\Omega, \mathcal{B})\}$ e seja $\mathcal{U}$ subfamília de $\mathcal{B}(\Omega, d)$ com as seguintes propriedades:

- 1) $\Omega \in \mathcal{U}$,

$$2) A \in \mathcal{U}, B \in \mathcal{U} \Rightarrow A \cap B \in \mathcal{U},$$

$$3) V \text{ aberto em } (\Omega, d) \Rightarrow \exists \{U_n, n \in \mathbb{N}\} \subseteq \mathcal{U} \text{ tal que } V = \bigcup_{n \in \mathbb{N}} U_n.$$

Nestas condições, $\mathcal{U}$ é uma classe limite determinante, ou seja se $\{M_n : n \in \mathbb{N}\} \subseteq \mathcal{M}^+$ e $M \in \mathcal{M}^+$, vale que:

$$\lim_n M_n(U) = M(U) \quad \forall U \in \mathcal{U} \Rightarrow M_n \xrightarrow{w} M.$$

Comentário: este resultado figura em Prohorov (1956), em Billingsley encontra-se a versão restrita a medidas de probabilidade. $\square$

2.D. ELEMENTOS ALEATÓRIOS E CONVERGÊNCIA EM DISTRIBUIÇÃO.

DEFINIÇÃO 2.14.

Seja $(S, \mathcal{F}, P)$ um espaço de probabilidade, (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$ e $Y: S \rightarrow \Omega$,

Y é um elemento aleatório $\stackrel{\text{DEF}}{\Leftrightarrow} Y$ é $\mathcal{F}$ - $\mathcal{B}$ mensurável;

a medida de probabilidade PY^{-1} definida em $(\Omega, \mathcal{B})$ por:

$$PY^{-1}(B) = P(Y^{-1}(B)) \quad \forall B \in \mathcal{B}$$

é chamada a distribuição de Y , ou seja, dado um elemento aleatório Y este determina univocamente uma medida de probabilidade: PY^{-1} .

A afirmação acima admite uma especie de recíproca: seja (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$ e Q uma probabilidade em $(\Omega, \mathcal{B})$, então:

$Y: (\Omega, \mathcal{B}, Q) \rightarrow (\Omega, \mathcal{B})$ definida por $Y(\omega) = \omega \quad \forall \omega \in \Omega$,

satisfaz

$$QY^{-1} = Q,$$

ou seja, Q é a distribuição de um elemento aleatório.

É claro que não necessariamente haverá unicidade do elemento aleatório cuja distribuição é Q .

DEFINIÇÃO 2.15.

Sejam $(S_n, \mathcal{F}_n, P_n)$ espaços de probabilidade $\forall n \in \mathbb{N}$, (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$Y_n: (S_n, \mathcal{F}_n, P_n) \rightarrow (\Omega, \mathcal{B})$ elemento aleatório $\forall n \in \mathbb{N}$,

e também $Y: (S, \mathcal{F}, P) \rightarrow (\Omega, \mathcal{B})$ elemento aleatório.

Nestas condições:

Y_n converge em distribuição a $Y \stackrel{\text{DEF}}{\Leftrightarrow} P_{Y_n}^{-1} \xrightarrow{\text{fracamente}} P_Y^{-1}$, e será denotado por

$$Y_n \xrightarrow{\mathcal{D}} Y.$$

OBSERVAÇÃO 2.16.

a) Em muitos casos $\forall n \in \mathbb{N} \quad (S_n, \mathcal{F}_n, P_n) = (S, \mathcal{F}, P)$ e geralmente o S não é considerado, já que muitas vezes toda a informação relevante está nas $P_{Y_n}^{-1}$ (ou $P_n Y_n^{-1}$) e P_Y^{-1} ;

b) Dada a correspondência entre elementos aleatórios e distribuições comentada na Definição 2.14, todo resultado sobre convergência fraca tem uma versão em termos de convergência em distribuição, e viceversa. Por exemplo, na Proposição 2.17, a seguir, será reformulada a Proposição 2.8 em termos de convergência em distribuição.

c) Na literatura adota-se muitas vezes uma terminologia híbrida. Por exemplo, é comum o seguinte enunciado do teorema central do limite:

$$\left. \begin{array}{l} X_1, X_2, \dots, X_n, \dots \text{ i.i.d} \\ \sigma^2 = \text{Variância}(X) < \infty \end{array} \right\} \Rightarrow [n\sigma^2]^{-1} \left\{ \sum_{j=1}^n X_j - nE(X_1) \right\} \longrightarrow N(0,1),$$

onde as X_j são variáveis aleatórias (e portanto toda função das X_j é) e $N(0,1)$ denota uma medida de probabilidade).

d) também, por abuso de notação, a topologia fraca é chamada topologia da convergência em distribuição.

PROPOSIÇÃO 2.17.

Sejam (Ω, d) , $\mathcal{B}$, $(S_n, \mathcal{F}_n, P_n)$, $(S, \mathcal{F}, P)$, Y_n e Y como na Definição 2.15.

Então as seguintes afirmações são equivalentes.

a) $Y_n \xrightarrow{\mathcal{D}} Y;$

b) $\lim_n E(f(Y_n)) = \lim_n \int f dP_n Y_n^{-1} = \int f dP^{-1} = E(f(Y)) \quad \forall f \in UCL(\Omega, d);$

c) $\limsup_n P_n Y_n^{-1}(F) \leq P Y^{-1}(F) \quad \forall F \text{ fechado};$

d) $\liminf_n P_n Y_n^{-1}(G) \geq P Y^{-1}(G) \quad \forall G \text{ aberto};$

e) $\lim_n P_n Y_n^{-1}(A) = P Y^{-1}(A) \quad \forall A \text{ tal que } P Y^{-1}(\partial A) = 0.$

Comentário: a demonstração é trivial, dado o resultado 2.8 e a definição de convergência em distribuição. $\square$

2.E. O CASO PARTICULAR DE (R, e) : CARACTERIZAÇÃO DE MEDIDAS FINITAS E DE CONVERGÊNCIA FRACA VIA FUNÇÕES DE DISTRIBUIÇÃO.

DEFINIÇÃO 2.18.

Seja M uma medida finita definida nos borelianos do espaço métrico (R, e) , onde R é o conjunto dos números reais e $e(r, s) = |r - s|$ é a métrica usual em R .

Define-se $F_M: R \rightarrow R^+ = \{u \in R: u \geq 0\}$ por:

$$F_M(t) = M(-\infty, t], \text{ onde } (-\infty, t] = \{u \in R: u \leq t\};$$

F_M assim definida é chamada a função de distribuição de M , quando não haja possibilidade de confusão será denotada simplesmente por F .

Na Proposição 2.19, seguinte, será demonstrado que existe uma correspondência biunívoca entre as medidas finitas e as funções de distribuição em geral, cuja definição será dada a seguir. Em 2.20 será dada uma caracterização da convergência fraca de medidas através da convergência pontual das respectivas funções de distribuição.

PROPOSIÇÃO 2.19.

Seja M uma medida finita definida nos borelianos do espaço métrico (R, e) , onde R é o conjunto dos números reais e $e(r, s) = |r - s|$ é a métrica usual em R . Seja F_M a função de distribuição de M acima definida. Então:

a) F tem as seguintes propriedades:

- 1) F_M é não decrescente,
- 2) F_M é contínua a direita,
- 3) $\lim_{t \rightarrow -\infty} F_M(t) = 0$ e $\lim_{t \rightarrow \infty} F_M(t) < \infty$;

por outro lado qualquer F satisfazendo 1), 2) e 3) será definida como função de distribuição e será utilizada a seguinte notação:

$$F(-\infty) = \lim_{t \rightarrow -\infty} F(t) \text{ e } F(\infty) = \lim_{t \rightarrow \infty} F(t);$$

b) reciprocamente, se F for uma função satisfazendo 1), 2) e 3) acima, existe uma única medida finita definida nos borelianos de R tal que $M(-\infty, t] = F(t)$ (necessariamente esta medida satisfaz $M(R) = F(\infty)$).

Comentário: a demonstração de a) é de verificação simples. Uma demonstração de b) baseia-se na σ -aditividade da função de conjunto

$$M(a, b] = F(b) - F(a) \text{ sobre a o semianel de conjuntos}$$

$I = \{(a, b] : a \in R, b \in R, a < b\} \cup \{\emptyset\}$; este fato pode provar-se mediante uma generalização da demonstração usual da σ -aditividade da medida de

Lebesgue em I , utilizando a aditividade finita de M sobre I , a compacidade dos intervalos fechados e as propriedades de regularidade dos intervalos em geral; a posterior extensão de M a uma medida nos borelianos de R fundamenta-se no teorema de extensão de Caratheodory.

A demonstração não será dada aqui. Os elementos para desenvolver o caminho sugerido acima podem ser consultados em livros de teoria da medida de nível elementar, como por exemplo Halmos (1950).

Esta proposição admite uma óbvia generalização ao caso de R^P , com a métrica euclideana e os respectivos borelianos.

A identificação medida-função de distribuição e a validade dos argumentos subsistem se a condição 3) for trocada pela mais fraca "3') $\lim_{t \rightarrow -\infty} F(t) = 0$ ", ocorrendo neste caso uma correspondência biunívoca entre a classe de funções que satisfaz 1), 2) e 3') e o conjunto de medidas definidas nos borelianos de R (ou R^P) tais que cada compacto tem medida finita (como no caso da medida de Lebesgue). $\square$

PROPOSIÇÃO 2.20.

Sejam (R, e) o espaço métrico onde $R = \{\text{números reais}\}$ e e a métrica euclideana usual, $\mathcal{B} = \{\text{borelianos de } (R, e)\}$ e $\mathcal{M}^+ = \mathcal{M}^+(R, e) = \{\text{medidas finitas sobre } (R, \mathcal{B})\}$. Nestas condições, se $\{M_n : n \in \mathbb{N}\} \subseteq \mathcal{M}^+$, $M \in \mathcal{M}^+$, e $\{F_n : n \in \mathbb{N}\}$ e F são as respectivas funções de distribuição, tem-se que:

$$M_n \xrightarrow{w} M \text{ se e só se } \begin{cases} F_n(t) \rightarrow F(t) \quad \forall t \text{ ponto de continuidade de } F \\ F_n(\infty) \rightarrow F(\infty) \end{cases}$$

2.F. TOPOLOGIA FRACA NO ESPAÇO DAS MEDIDAS FINITAS DE UM ESPAÇO MÉTRICO.

PROPOSIÇÃO 2.21.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$ e $\mathcal{M}^+ = \{\text{medidas finitas em } (\Omega, d)\}$.

Consideremos agora, para cada $M \in \mathcal{M}^+$, a seguinte família de conjuntos:

$$W(M) = \left\{ W(M, f_1, f_2, \dots, f_k, \epsilon) : k \in \mathbb{N}, f_j \in CL(\Omega, d) \text{ se } 1 \leq j \leq k, \epsilon > 0 \right\},$$

onde

$$W(M, f_1, f_2, \dots, f_k, \epsilon) = \left\{ Q \in \mathcal{M}^+ : \left| \int f_j dQ - \int f_j dM \right| < \epsilon, 1 \leq j \leq k \right\}$$

(a) Para cada $M \in \mathcal{M}^+$, a família $W(M)$ tem as seguintes propriedades:

- 1) $\forall V \in W(M), M \in V$;
- 2) $U \in W(M), V \in W(M) \Rightarrow \exists W \in W(M) \text{ tal que } W \subset U \cap V$,
- 3) $\forall V \in W(M) \exists W \in W(M) \text{ tal que } W \subseteq V \text{ e } W \in W(Q) \forall Q \in W$,

que são satisfeitas por uma base de vizinanças de um ponto em qualquer espaço topológico.

Reciprocamente, se W é uma função que associa a cada $M \in \mathcal{M}^+$ uma família $W(M)$ satisfazendo 1) e 2), então a família

$$\mathcal{T} = \left\{ W \subseteq \mathcal{M}^+ : W \in W(M) \forall M \in W \right\}$$

é uma topologia em $\mathcal{M}^+$; se também 3) for satisfeita, então $\mathcal{W}(M)$ é uma base de vizinhanças de M respeito da topologia $\mathcal{T}$.

Em particular, a topologia $\mathcal{T}$ associada à família $\{\mathcal{W}(M) : M \in \mathcal{M}^+\}$ é designada pelo nome de topologia fraca.

(b) $(\mathcal{M}^+, \mathcal{T})$ é um espaço de Hausdorff, ou seja:

$$M \in \mathcal{M}^+, N \in \mathcal{M}^+ \text{ e } M \neq N \Rightarrow \exists U \in \mathcal{W}(M) \text{ e } \exists V \in \mathcal{W}(N) \text{ tais que } U \cap V = \emptyset.$$

Comentário: quanto a (a), é fácil comprovar que para cada $M \in \mathcal{M}^+$ a família $\mathcal{W}(M)$ satisfaz as condições 1), 2) e 3); a recíproca, que diz como construir uma topologia a partir de um sistema de vizinhanças de cada ponto é uma propriedade de validade geral, cuja fundamentação pode ser vista em Kelley (1962).

Já (b) é de verificação simples: sabe-se que se M e N são elementos de $\mathcal{M}^+$ e $\int f dM = \int f dN \quad \forall f \in CL(\Omega, d) \Rightarrow M = N$ (Proposição 2.3.(b)), portanto, dado que $M \neq N$, tem-se que: $\exists g \in CL(\Omega, d)$ tal que $\int g dM - \int g dN = \alpha \neq 0$,

em consequência:

$$\mathcal{W}(M, g, |\alpha|/2) \cap \mathcal{W}(N, g, |\alpha|/2) = \emptyset. \square$$

PROPOSIÇÃO 2.22.

$\{M_n, n \in \mathbb{N}\} \subseteq \mathcal{M}^+, M \in \mathcal{M}^+$. Então:

$$M_n \xrightarrow{\mathcal{W}} M \text{ se e só se } M_n \text{ converge a } M \text{ na topologia fraca.}$$

Comentário: a demonstração é consequência das definições. $\square$

PROPOSIÇÃO 2.23.

As seguintes famílias de conjuntos também são bases locais para a topologia fraca $\mathcal{T}$ acima definida. Para cada $M \in \mathcal{M}^+$, definem-se:

$$a) \mathcal{V}(M) = \left\{ V(M, F_1, F_2, \dots, F_k, \varepsilon) : \begin{array}{l} F_j, 1 \leq j \leq k, \text{ fechados em } (\Omega, d) \\ k \in \mathbb{N}, \varepsilon > 0 \end{array} \right\},$$

onde

$$V(M, F_1, F_2, \dots, F_k, \varepsilon) = \left\{ Q \in \mathcal{M}^+ : \begin{array}{l} Q(F_j) < M(F_j) + \varepsilon, \quad 1 \leq j \leq k \\ |Q(\Omega) - M(\Omega)| < \varepsilon \end{array} \right\};$$

$$b) \mathcal{U}(M) = \left\{ U(M, G_1, G_2, \dots, G_k, \varepsilon) : \begin{array}{l} G_j, 1 \leq j \leq k, \text{ abertos em } (\Omega, d) \\ k \in \mathbb{N}, \varepsilon > 0 \end{array} \right\},$$

onde

$$U(M, G_1, G_2, \dots, G_k, \varepsilon) = \left\{ Q \in \mathcal{M}^+ : \begin{array}{l} Q(G_j) > M(G_j) - \varepsilon, \quad 1 \leq j \leq k \\ |Q(\Omega) - M(\Omega)| < \varepsilon \end{array} \right\}$$

$$c) \mathcal{Y}(M) = \left\{ Y(M, C_1, C_2, \dots, C_k, \varepsilon) : k \in \mathbb{N}, M(\delta C_j) = 0, 1 \leq j \leq k, \varepsilon > 0 \right\},$$

onde

$$Y(M, C_1, C_2, \dots, C_k, \varepsilon) = \left\{ Q \in \mathcal{M}^+ : |Q(C_j) - M(C_j)| < \varepsilon, 1 \leq j \leq k \right\}.$$

Comentário: em c) não há necessidade de colocar explicitamente uma condição equivalente a $|Q(\Omega) - M(\Omega)| < \varepsilon$, já que ela está embutida na definição geral de vizinhança (considere $Y(M, \Omega, \varepsilon)$), como acontece na definição dada em 2.21 da topologia fraca (considere $W(M, 1_\Omega, \varepsilon)$; em

Billingsley (1968, p. 236) há uma demonstração restrita a medidas de probabilidade, veja generalização para o caso geral de medidas finitas em 2.G. $\square$

OBSERVAÇÃO 2.24.

Dadas as diversas caracterizações da topologia fraca dadas em 2.23 e tendo provado em 2.22 que $M_n \xrightarrow{w} M$ se e só se M_n converge a M na topologia fraca é bastante simples provar agora os resultados 2.8 e 2.9.

PROPOSIÇÃO 2.25.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos em } (\Omega, d)\}$, $\mathcal{M}^+ = \{\text{medidas finitas em } (\Omega, d)\}$. Consideremos a topologia fraca $\mathcal{T}$ em $\mathcal{M}^+$. Então:

- a) as medidas a suporte finito são densas em $\mathcal{M}^+$;
- b) (Ω, d) é separável $\Rightarrow (\mathcal{M}^+, \mathcal{T})$ é separável.

Comentário: a) está provado em Billingsley (1968, p. 237) para medidas de probabilidade, a generalização a medidas finitas é trivial (veja 2.G); quanto a b), provaremos que se $D^- = \Omega$, então

$$\mathcal{D}^* = \left\{ L \in \mathcal{M}^+ : \begin{array}{l} \text{suporte}(L) \subseteq D, \text{ suporte}(L) \text{ finito} \\ \text{átomos de } L \text{ têm massa racional} \end{array} \right\}$$

é denso enumerável em $(\mathcal{M}^+, \mathcal{T})$, veja 2.G). No capítulo seguinte (veja Proposição 3.12.2)) será provado que a separabilidade de (Ω, d) implica também que $(\mathcal{M}^+, \mathcal{T})$ é metrizável (além de separável). $\square$

2.G. DEMONSTRAÇÕES.

Demonstrações da Seção 2.B.

Proposição 2.1.

Seja (Ω, d) um espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$ e M uma medida finita em $(\Omega, \mathcal{B})$. Então:

(a) $\forall C \in \mathcal{B}$ vale que $C^- = \text{aderência}(C)$ é um G_δ e que $C^\circ = \text{interior}(C)$ é um F_σ , portanto todo conjunto fechado é um G_δ e todo aberto um F_σ ;

(b) F fechado $\Rightarrow M(F) = \inf\{M(A) : A \text{ aberto e } A \supseteq F\}$

e também

F aberto $\Rightarrow M(G) = \sup\{M(B) : B \text{ fechado e } B \subseteq G\}$;

(c) $\forall C \in \mathcal{B} \forall \varepsilon > 0$ existem F fechado e G aberto tais que:

$$F \subseteq C \subseteq G \quad \text{e} \quad M(G-F) < \varepsilon;$$

(d) todo boreliano é "regular", isto é $\forall C \in \mathcal{B}$ vale que:

$$M(C) = \inf\{M(G) : G \text{ aberto, } G \supseteq C\} = \sup\{M(F) : F \text{ fechado, } F \subseteq C\};$$

(e) M medida finita em $(\Omega, \mathcal{B})$, então M fica completamente determinada por $\{M(F) : F \text{ fechado}\}$ ou equivalentemente por $\{M(G) : G \text{ aberto}\}$.

Demonstração:

$$\forall C \in \mathcal{B} \text{ vale que } C^- = \left\{ \omega \in \Omega : d(\omega, C) = 0 \right\} = \bigcap_{n \in \mathbb{N}} G_n$$

onde os conjuntos $G_n = \left\{ \omega \in \Omega : d(\omega, C) < n^{-1} \right\}$ satisfazem:

$$G_n \text{ aberto } \forall n \in \mathbb{N},$$

$$G_n \supseteq G_{n+1} \quad \forall n \in \mathbb{N},$$

$$C^- = \bigcap_{n \in \mathbb{N}} G_n,$$

portanto:

C^- é G_δ (intersecção enumerável de abertos)

$$M(C^-) = \lim_n M(G_n) \text{ (por ser } M \text{ finita);}$$

por argumentos de complementação,

C^0 é um F_σ (união enumerável de fechados)

e sendo M finita, tem-se

$$M(C^0) = \lim_n M(F_n)$$

onde os conjuntos $F_n = \left\{ \omega \in \Omega : d(\omega, C^c) \geq n^{-1} \right\}$ satisfazem:

$$F_n \text{ fechado } \forall n \in \mathbb{N},$$

$$F_n \subseteq F_{n+1} \quad \forall n \in \mathbb{N},$$

$$C^{\circ} = \bigcup_{n \in \mathbb{N}} F_n;$$

dado que

$$C \text{ fechado} \Leftrightarrow C = C^{-}$$

e que

$$C \text{ aberto} \Leftrightarrow C = C^{\circ}$$

a demonstração de (a) está completa;

dado que (b) é uma consequência de (a), (e) é consequência de (d); e a finitude de M implica na equivalência de (c) e (d) é suficiente provar (c):

$$\text{seja então } \mathcal{C} = \left\{ C \in \mathcal{B}; \begin{array}{l} \forall \varepsilon > 0 \exists F \text{ fechado e } \exists G \text{ aberto} \\ \text{tais que } F \subseteq C \subseteq G \text{ e } M(G-F) < \varepsilon \end{array} \right\}, \text{ provaremos}$$

que $\mathcal{C}$ é uma σ -álgebra que contém os fechados (o que implicará que $\mathcal{C} \supseteq \mathcal{B}$):

seja então $\varepsilon > 0$, se C for fechado (b) implica que existe uma seqüência decrescente de abertos G_n tal que $M(G_n) \rightarrow M(C)$, portanto basta tomar G_m tal que $M(G_m) < M(C) + \varepsilon$ e definir $F = C$ e $G = G_m$, mostrando-se assim que $C \in \mathcal{C}$ (ou seja provou-se que $\{\text{fechados}\} \subseteq \mathcal{C}$);

provar que $\mathcal{C}$ é uma σ -álgebra equivale a provar que $\mathcal{C}$ é fechada para uniões enumeráveis e complementos:

$$\text{seja } \{C_n, n \in \mathbb{N}\} \subseteq \mathcal{C} \text{ e } \varepsilon > 0,$$

por definição de $\mathcal{C}$, $\forall n \in \mathbb{N}$ existem F_n fechado e G_n aberto satisfazendo

$$F_n \subseteq C_n \subseteq G_n \quad \text{e} \quad M(G_n - F_n) < \varepsilon/2^{n+1},$$

seja $m \in \mathbb{N}$ tal que

$$M\left(\bigcup_{n \in \mathbb{N}} F_n - \bigcup_{n \leq m} F_n\right) = M\left(\bigcup_{n \in \mathbb{N}} F_n\right) - M\left(\bigcup_{n \leq m} F_n\right) < \varepsilon/2,$$

então $F = \bigcup_{n \leq m} F_n$ e $G = \bigcup_{n \in \mathbb{N}} G_n$ verificam que:

$$F \text{ fechado, } G \text{ aberto, } F \subseteq \bigcup_{n \in \mathbb{N}} C_n \subseteq G \text{ e } M(G-F) < \varepsilon,$$

portanto $\bigcup_{n \in \mathbb{N}} C_n \in \mathcal{C}$ e $\mathcal{C}$ é fechada para uniões enumeráveis;

só falta verificar que $\mathcal{C}$ é fechada para complementos, mas isto é trivial já que:

$$F \subseteq C \subseteq G \quad \text{see} \quad G^c \subseteq C^c \subseteq F^c$$

F fechado e G aberto see G^c fechado e F^c aberto

$$G-F = F^c - G^c. \square$$

Proposição 2.3.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$.

(a) Sejam $F \in \mathcal{B}$, F fechado e $\varepsilon > 0$,

existe uma função $f \in UCL(\Omega, d)$ tal que:

$$0 \leq f \leq 1, \quad f|_F = 1 \quad \text{e} \quad f|_{\{\omega \in \Omega: d(\omega, F) \geq \varepsilon\}} = 0;$$

(b) M e N medidas finitas em $(\Omega, \mathcal{B})$, então:

$$\int f dM = \int f dN \quad \forall f \in UCL(\Omega, d) \Rightarrow M = N.$$

Demonstração:

Observe-se que se $C \neq \emptyset$, então $g(\omega) = d(\omega, C)$ é uniformemente contínua, já que, em consequência da desigualdade triangular:

$$d(\omega, C) \leq d(\omega, \tau) + d(\tau, C) \text{ e } d(\tau, C) \leq d(\tau, \omega) + d(\omega, C),$$

e portanto

$$|d(\omega, C) - d(\tau, C)| \leq d(\omega, \tau).$$

Agora, se A e B são dois conjuntos fechados de Ω tais que $A \cap B = \emptyset$, então $f: \Omega \rightarrow \mathbb{R}$ definida por:

$$f(\omega) = d(\omega, B) / [d(\omega, A) + d(\omega, B)]$$

satisfaz

$$0 \leq f \leq 1, \quad f|_A = 1, \quad f|_B = 0 \text{ e } f \text{ é contínua,}$$

(a continuidade decorre do fato de ser f quociente de funções contínuas com denominador não nulo);

além disso, acrescentando a hipótese de que $d(A, B) > 0$, tem-se que f é uniformemente contínua, já que o numerador $d(\omega, B)$ é uniformemente contínuo e o denominador $d(\omega, B) + d(\omega, A) \geq d(A, B) > 0$;

o resultado (a) é consequência do acima exposto, já que basta tomar $A = F$ e $B = \{\omega \in \Omega: d(\omega, F) \geq \varepsilon\}$ e observar que A é fechado (por hipótese), B é fechado (pois $d(\omega, F)$ é contínua) e verificar que $d(A, B) \geq \varepsilon > 0$;

devemos agora provar (b), sejam então M e N medidas finitas em $(\Omega, \mathcal{B})$ tais que:

$$\int f dM = \int f dN \quad \forall f: \Omega \rightarrow \mathbb{R} \text{ uniformemente contínua e limitada,}$$

seja F um fechado fixo, consideremos:

$$f_n(\omega) = d(\omega, B_n) / [d(\omega, F) + d(\omega, B_n)],$$

$$\text{onde } B_n = \{\omega \in \Omega : d(\omega, F) \geq n^{-1}\}$$

e sendo que as f_n são uniformemente contínuas e limitadas por (a), a hipótese assegura que

$$\int f_n dM = \int f_n dN \quad \forall n \in \mathbb{N},$$

mas também

$$0 \leq f_n \leq 1, \text{ e } \forall \omega \in \Omega \text{ vale que } \lim_{n \rightarrow \infty} f_n(\omega) = 1_F(\omega),$$

(onde 1_F é a função característica do conjunto F),

portanto, pelo teorema de convergência dominada, tem-se que:

$$M(F) = \int 1_F dM = \lim_n \int f_n dM = \lim_n \int f_n dN = \int 1_F dN = N(F);$$

este argumento vale para todo conjunto fechado F , ou seja, que M e N coincidem no conjunto dos fechados, portanto, pela proposição anterior $M = N$ em $\mathcal{B}$. $\square$

Proposição 2.5.

(Ω, d) espaço métrico separável e completo,

$\mathcal{B} = \{\text{borelianos de } (\Omega, d)\},$

M uma medida finita sobre $(\Omega, d),$

então: M é tight.

Demonstração:

seja $\varepsilon > 0$, provaremos que $\exists K_\varepsilon$ compacto com $M(K_\varepsilon) > M(\Omega) - \varepsilon;$

se $\omega \in \Omega$ e $r > 0$ seja $B(\omega, r) = \{\tau: d(\omega, \tau) < r\},$

se $n \in \mathbb{N}$ vale que $\Omega = \bigcup_{\omega \in \Omega} B(\omega, n^{-1}),$

sendo Ω métrico, Ω separável equivale a que Ω tem base de abertos enumerável e portanto Ω tem a propriedade de Lindeloff (i.e. toda cobertura de Ω por abertos tem uma subcobertura enumerável),

então existe uma sequência $\{\omega_j^n\} \subseteq \Omega$ tal que $\Omega \subseteq \bigcup_{j \in \mathbb{N}} B_j^n$, onde $B_j^n = B(\omega_j^n, n^{-1})$, seja j_n tal que $M\left\{\bigcup_{j \leq j_n} B_j^n\right\} > M(\Omega) - \varepsilon/2^n;$

o procedimento acima pode ser feito $\forall n \in \mathbb{N}$, neste caso o conjunto

$L_\varepsilon = \bigcap_{n \in \mathbb{N}} \bigcup_{j \leq j_n} B_j^n$ satisfaz $M(L_\varepsilon^c) < M(\Omega) - \varepsilon,$

além disso, L_ε é totalmente limitado, portanto, sendo (Ω, d) completo, $K_\varepsilon = \text{aderência}(L_\varepsilon)$ é compacto,

por outro lado, $M(K_\varepsilon) > M(\Omega) - \varepsilon. \square$

Demonstrações da Seção 2.C.

Lema 2.11.

a) $\{M_n\}_{n \in \mathbb{N}} \subseteq \mathcal{M}^+$, $M \in \mathcal{M}^+$ e $M(\Omega) \neq 0$, então:

$$M_n \xrightarrow{w} M \quad \text{see} \quad \begin{cases} M_n(\Omega) \xrightarrow{\text{em}(R,e)} M(\Omega) \\ M_n/M_n(\Omega) \xrightarrow{e, w} M/M(\Omega) \end{cases};$$

b) $M_n \xrightarrow{w} \text{medida nula}$ see $M_n(\Omega) \rightarrow 0$.

Demonstração:

a) seja $M_n \xrightarrow{w} M$, onde $M(\Omega) \neq 0$:

$$M_n \xrightarrow{w} M \Rightarrow M_n(\Omega) = \int_{\Omega} 1 dM_n \xrightarrow{w} \int_{\Omega} 1 dM = M(\Omega),$$

por outro lado

$$\int fd(M_n/M_n(\Omega)) = (M_n(\Omega))^{-1} \int fdM_n \xrightarrow{w} (M(\Omega))^{-1} \int fdM = \int fd(M/M(\Omega));$$

$$\text{reciprocamente, se } \begin{cases} M_n(\Omega) \xrightarrow{\text{em}(R,e)} M(\Omega) \\ M_n/M_n(\Omega) \xrightarrow{e, w} M/M(\Omega) \end{cases}, \text{ então tem-se que}$$

$$\int fdM_n = M_n(\Omega) \times \int fd(M_n/M_n(\Omega)) \xrightarrow{w} M(\Omega) \times \int fd(M/M(\Omega)) = \int fdM;$$

b) na parte inicial da demonstração de a) provou-se que se $M_n \xrightarrow{w} M \Rightarrow M_n(\Omega) \rightarrow M(\Omega)$, em particular se M é a medida nula tem-se que $M_n(\Omega) \rightarrow 0$;

reciprocamente se $M_n(\Omega) \rightarrow 0$:

$$\left| \int f dM_n \right| \leq \int |f| dM_n \leq \left(\sup_{\omega \in \Omega} |f(\omega)| \right) \times M_n(\Omega) \xrightarrow{n \rightarrow \infty} 0 = \int f d0, \text{ onde } 0$$

denota a medida nula. $\square$

Proposição 2.12.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, M_n e M medidas finitas em $(\Omega, \mathcal{B})$. Sejam $M_n h^{-1}$ e Mh^{-1} as medidas finitas definidas nos borelianos de R por

$$M_n h^{-1}(B) = M_n(h^{-1}(B)) \text{ e } Mh^{-1}(B) = M(h^{-1}(B)).$$

Nestas condições:

$$M_n \xrightarrow{w} P \text{ se e só se } M_n h^{-1} \xrightarrow{w} Mh^{-1}, \forall h: (\Omega, d) \xrightarrow[\text{limitada}]{\text{contínua}} (R, e).$$

Demonstração:

$(\Rightarrow)$ será provado que

$$\forall F \text{ fechado em } R \text{ vale que } \limsup_n M_n h^{-1}(F) \leq Mh^{-1}(F);$$

h contínua e F fechado $\Rightarrow h^{-1}(F)$ fechado em (Ω, d) , e, dado que por hipótese $M_n \xrightarrow{w} M$, tem-se que:

$$\limsup_n M_n(h^{-1}(F)) \leq M(h^{-1}(F)),$$

o que significa que

$$\limsup_n M_n h^{-1}(F) \leq Mh^{-1}(F);$$

$$(\Rightarrow) \text{ se } h: (\Omega, d) \xrightarrow[\text{limitada}]{\text{contínua}} (R, e), \text{ provaremos que } \int h dM_n \xrightarrow{n \rightarrow \infty} \int h dM,$$

dado que $M_n h^{-1} \xrightarrow{W} M h^{-1}$, vale que

$$\forall f \in CL(\Omega, d) \quad \int f dM_n h^{-1} \xrightarrow{} \int f dM h^{-1},$$

ou seja que

$$\forall f \in CL(R, e) \quad \int (f \circ h) dM_n \xrightarrow{} \int (f \circ h) dM,$$

dado que h é limitada, $\exists k > 0$ tal que $\sup\{|h(\omega)| : \omega \in \Omega\} \leq k$, seja então f_k definida por:

$$f_k(t) = \max\{\min\{t, k\}, -k\} = \min\{\max\{t, -k\}, k\},$$

$$f_k \in CL(R, e) \Rightarrow f_k \circ h = h \text{ satisfaz}$$

$$\int h dM_n \xrightarrow{} \int h dM. \square$$

Proposição 2.13.

(Ω, d) espaço métrico, $\mathcal{B}(\Omega, d) = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+ = \{\text{medidas finitas sobre } (\Omega, \mathcal{B})\}$ e seja $\mathcal{U}$ subfamília de $\mathcal{B}(\Omega, d)$ com as seguintes propriedades:

$$1) \Omega \in \mathcal{U},$$

$$2) A \in \mathcal{U}, B \in \mathcal{U} \Rightarrow A \cap B \in \mathcal{U},$$

$$3) V \text{ aberto em } (\Omega, d) \Rightarrow \exists \{U_n, n \in \mathbb{N}\} \subseteq \mathcal{U} \text{ tal que } V = \bigcup_{n \in \mathbb{N}} U_n.$$

Nestas condições, $\mathcal{U}$ é uma classe limite determinante, ou seja, se $\{M_n : n \in \mathbb{N}\} \subseteq \mathcal{M}^+$ e $M \in \mathcal{M}^+$, vale que:

$$\lim_n M_n(U) = M(U) \quad \forall U \in \mathcal{U} \Rightarrow M_n \xrightarrow{W} M.$$

Demonstração:

Dado que por hipótese $\mathcal{U}$ é fechada para intersecções finitas, tem-se que:

$$M_n\left(\bigcup_{i=1}^l U_j\right) = \sum_{i=1}^l M_n(U_i) - \sum_{i < j < l} M_n(U_i \cap U_j) + \sum_{i < j < k < l} M_n(U_i \cap U_j \cap U_k) - \dots$$

$$\xrightarrow{W} \sum_{i=1}^l M(U_i) - \sum_{i < j < l} M(U_i \cap U_j) + \sum_{i < j < k < l} M(U_i \cap U_j \cap U_k) - \dots = M\left(\bigcup_{i=1}^l U_j\right),$$

ou seja, $\mathcal{U}$ é também fechada para uniões finitas.

Seja agora G aberto e $\varepsilon > 0$,

por hipótese, existe $\{U_j : j \in \mathbb{N}\} \subseteq \mathcal{U}$ tal que $G = \bigcup_{j \in \mathbb{N}} U_j$, então:

$$\exists m \in \mathbb{N} \text{ tal que } M\left(\bigcup_{j \leq m} U_j\right) > M(G) - \varepsilon,$$

em consequência da afirmação da linha anterior e do fato de $\mathcal{U}$ ser fechada para uniões finitas:

$$M(G) - \varepsilon < M\left(\bigcup_{j \leq m} U_j\right) = \lim_n M_n\left(\bigcup_{j \leq m} U_j\right) = \lim_n \inf_{j \leq m} M_n(U_j) \leq \lim_n \inf_{j \leq m} M_n(G),$$

ou seja:

$$\forall G \text{ aberto } \forall \varepsilon > 0 \text{ tem-se que } M(G) - \varepsilon \leq \lim_n \inf_{j \leq m} M_n(G),$$

que por sua vez implica:

$\forall G$ aberto tem-se que $M(G) - \varepsilon \leq \liminf_n M_n(G)$;

dada a Proposição 2.9 e a afirmação da linha acima, para provar $M_n \xrightarrow{w} M$ falta verificar que $M_n(\Omega) \rightarrow M(\Omega)$, que é imediato pois $\Omega \in \mathcal{U}$ (propriedade 1)). $\square$

Demonstrações da Seção 2.E.

Proposição 2.20.

Sejam (R, e) o espaço métrico onde $R = \{\text{números reais}\}$ e e a métrica euclídeana usual, $\mathcal{B} = \{\text{borelianos de } (R, e)\}$ e $\mathcal{M}^+ = \mathcal{M}^+(R, e) = \{\text{medidas finitas sobre } (R, \mathcal{B})\}$. Nestas condições, se $\{M_n : n \in \mathbb{N}\} \subseteq \mathcal{M}^+$, $M \in \mathcal{M}^+$, e $\{F_n : n \in \mathbb{N}\}$ e F são as respectivas funções de distribuição, tem-se que:

$$M_n \xrightarrow{w} M \text{ se e só se } \begin{cases} F_n(t) \rightarrow F(t) \quad \forall t \text{ ponto de continuidade de } F \\ F_n(\infty) \rightarrow F(\infty) \end{cases}$$

Demonstração:

($\Rightarrow$) pela Proposição 2.9,

$$M_n \xrightarrow{w} M \text{ se e só se } \lim_n M_n(A) = M(A) \quad \forall A \text{ tal que } M(\delta A) = 0,$$

dado que:

$$M(\delta(-\infty, t]) = M\{t\} = \lim_{s \rightarrow t, s > t} F(s) - \lim_{s \rightarrow t, s < t} F(s) = F(t) - \lim_{s \rightarrow t, s < t} F(s)$$

$$t \text{ é ponto de continuidade de } F \Rightarrow M(\delta(-\infty, t]) = 0,$$

portanto, $M_n \xrightarrow{w} M$, t ponto de continuidade de F e a Proposição 2.9

implicam que:

$$M_n(-\infty, t] \xrightarrow{\quad} M(-\infty, t],$$

ou seja, se t é ponto de continuidade de F ,

$$F_n(t) \xrightarrow{\quad} F(t);$$

finalmente, dado que $\delta(\Omega) = \emptyset$ e que $M_n \xrightarrow{\quad} M$,

$$F_n(\infty) = M_n(\Omega) \xrightarrow{\quad} M(\Omega) = F(\infty);$$

($\Leftarrow$) se $\mathcal{B}$ denota a família dos borelianos de um espaço métrico Ω , M e $M_n, n \in \mathbb{N}$ medidas finitas sobre $\mathcal{B}$ e $\mathcal{U} \subseteq \mathcal{B}$ a Proposição 2.13 afirma que:

se $\mathcal{U}$ satisfaz

$$1) \Omega \in \mathcal{U},$$

$$A \in \mathcal{U}, B \in \mathcal{U} \Rightarrow A \cap B \in \mathcal{U},$$

$$3) V \text{ aberto em } (\Omega, d) \Rightarrow \exists \{U_n, n \in \mathbb{N}\} \subseteq \mathcal{U} \text{ tal que } V = \bigcup_{n \in \mathbb{N}} U_n.$$

então:

$\mathcal{U}$ é classe limite determinante, ou seja, que

$$\lim_n M_n(U) = M(U) \quad \forall U \in \mathcal{U} \quad \Rightarrow \quad M_n \xrightarrow{W} M,$$

sejam então $\mathcal{B} = \{\text{borelianos de } \mathbb{R}\}$, M_n e M medidas finitas sobre $\mathcal{B}$ e F_n e F suas respectivas funções de distribuição, por hipótese

$$F_n(t) \xrightarrow{\quad} F(t) \quad \forall t \text{ ponto de continuidade de } F \text{ e } F_n(\infty) \xrightarrow{\quad} F(\infty),$$

provaremos que $M_n \xrightarrow{W} M$ utilizando a Proposição 2.13;

seja $\mathcal{U} = \left\{ (a,b] : a \in \mathbb{R}, b \in \mathbb{R}, a < b, M\{a\} = M\{b\} = 0 \right\} \cup \{ \mathbb{R}, \emptyset \}$,

verifiquemos que $\mathcal{U}$ satisfaz as condições de 2.13:

1) $\Omega = \mathbb{R} \in \mathcal{U}$ por definição,

2) $(a,b] \cap (c,d] = (\max(a,c), \min(b,d)]$,

3) provar que todo aberto de $\mathbb{R}$ é união enumerável de conjuntos de $\mathcal{U}$ equivale a provar que todo intervalo aberto é união enumerável de elementos de $\mathcal{U}$,

$(a,b) = \bigcup_{n \in \mathbb{N}} (a_n, b_n]$, onde os extremos a_n e b_n são escolhidos de tal

forma que

$$a < a_n < a + n^{-1} \quad \text{e} \quad b + n^{-1} < b_n < b,$$

a escolha é possível, pois sendo M finita existe no máximo uma quantia enumerável de conjuntos disjuntos com medida estritamente positiva, já que $\forall n \in \mathbb{N}$ existe apenas um número finito de conjuntos disjuntos de medida maior que $1/n$ e

$$\left\{ A \in \mathcal{B} : M(A) > 0 \right\} = \bigcup_{n \in \mathbb{N}} \left\{ A \in \mathcal{B} : M(A) > 1/n \right\};$$

falta apenas provar que $M_n(U) \rightarrow M(U) \quad \forall U \in \mathcal{U}$:

seja $U = (a,b]$, onde $a < b$ e $M\{b\} = M\{a\} = 0$,

por hipótese: $F_n(b) \rightarrow F(b)$ e $F_n(a) \rightarrow F(a)$,

então,

$$M_n(U) = F_n(b) - F_n(a) \xrightarrow{w} F(b) - F(a) = M(a, b],$$

$$\text{se } U = \emptyset, \forall n \in \mathbb{N} \quad M_n(\emptyset) = M(\emptyset) = 0,$$

se $U = R$, por hipótese temos que $F_n(\infty) \xrightarrow{w} F(\infty)$, e dado que $F_n(\infty) = M_n(R)$ e $F(\infty) = M(R)$, tem-se que $M_n(R) \xrightarrow{w} M(R)$. $\square$

Demonstrações da Seção 2.F.

Proposição 2.22.

$\{M_n, n \in \mathbb{N}\} \subseteq M^+$, $M \in M^+$. Então:

$M_n \xrightarrow{w} M$ se e somente se M_n converge a M na topologia fraca.

Demonstração:

seja $M_n \xrightarrow{w} M$, consideremos $W(M, f_1, f_2, \dots, f_k, \varepsilon)$, onde $\varepsilon > 0$ e $f_j \in CL(\Omega, d)$, $1 \leq j \leq k$,

dado que $W(M, f_1, f_2, \dots, f_k, \varepsilon) = \bigcap_{j=1}^k W(M, f_j, \varepsilon)$, para provar a convergência de M_n a M na topologia fraca basta provar que:

dado $\varepsilon > 0$ e $f \in CL(\Omega, d)$ $\exists n_0$ tal que $n \geq n_0 \Rightarrow M_n \in W(M, f, \varepsilon)$,

o que é verdadeiro, dado que por definição de $\xrightarrow{w}$ deve ocorrer que $\int f dM_n \xrightarrow{w} \int f dM$;

reciprocamente, seja $f \in CL(\Omega, d)$ e $\varepsilon > 0$,

dado que por hipótese $M_n \xrightarrow{\text{fraca}} M$ na topologia fraca,

$\exists n_0$ tal que $n \geq n_0 \Rightarrow M_n \in W(M, f, \epsilon)$,

equivalentemente

$\exists n_0$ tal que $n \geq n_0 \Rightarrow |\int f dM_n - \int f dM| < \epsilon$. $\square$

Proposição 2.23.

As seguintes famílias de conjuntos também são bases locais para a topologia fraca $\mathcal{T}$ acima definida. Para cada $M \in \mathcal{M}^+$, definem-se:

$$a) \mathcal{V}(M) = \left\{ V(M, F_1, F_2, \dots, F_k, \epsilon) : \begin{array}{l} F_j, 1 \leq j \leq k, \text{ fechados em } (\Omega, d) \\ k \in \mathbb{N}, \epsilon > 0 \end{array} \right\},$$

onde

$$V(M, F_1, F_2, \dots, F_k, \epsilon) = \left\{ Q \in \mathcal{M}^+ : \begin{array}{l} Q(F_j) < M(F_j) + \epsilon, \quad 1 \leq j \leq k \\ |Q(\Omega) - M(\Omega)| < \epsilon \end{array} \right\};$$

$$b) \mathcal{U}(M) = \left\{ U(M, G_1, G_2, \dots, G_k, \epsilon) : \begin{array}{l} G_j, 1 \leq j \leq k, \text{ abertos em } (\Omega, d) \\ k \in \mathbb{N}, \epsilon > 0 \end{array} \right\},$$

onde

$$U(M, G_1, G_2, \dots, G_k, \epsilon) = \left\{ Q \in \mathcal{M}^+ : \begin{array}{l} Q(G_j) > M(G_j) - \epsilon, \quad 1 \leq j \leq k \\ |Q(\Omega) - M(\Omega)| < \epsilon \end{array} \right\}$$

$$c) \mathcal{Y}(M) = \left\{ Y(M, C_1, C_2, \dots, C_k, \epsilon) : k \in \mathbb{N}, M(\delta C_j) = 0, \quad 1 \leq j \leq k, \epsilon > 0 \right\},$$

onde

$$Y(M, C_1, C_2, \dots, C_k, \varepsilon) = \left\{ Q \in M^+ : |Q(C_j) - M(C_j)| < \varepsilon, 1 \leq j \leq k \right\}.$$

Demonstração: será provada a equivalência dos três sistemas de vizinhanças acima ao da topologia fraca $W(M)$ da seguinte forma: V equivale a W , V equivale a U e V equivale a $\mathcal{Y}$;

1) V equivale a W , parte $W \gg V$,

onde a notação $W \gg V$ significa que o sistema de vizinhanças de W é mais fino que o de V , ou seja, que $\forall V \in \mathcal{V}(M) \exists W \in \mathcal{W}(M)$ tal que $W \subset V$,

dado que $V(M, F_1, F_2, \dots, F_k, \varepsilon) = \bigcap_{j=1}^k V(M, F_j, \varepsilon)$, é suficiente considerar

$V = V(M, F, \varepsilon)$:

pela Proposição 2.3, dado $n > 0 \exists f_n \in CL(\Omega, d)$ tal que $f_n \equiv 1$ em F , $f_n \equiv 0$ em $\{\omega : d(\omega, F) \geq n^{-1}\}$ e $0 \leq f_n(\omega) \leq 1 \forall \omega \in \Omega$,

se $G_n = \{\omega : d(\omega, F) < n^{-1}\}$ vale que $G_{n+1} \subseteq G_n$ e que $F = \bigcap_{n \in \mathbb{N}} G_n$ (veja 2.1),

seja $n \in \mathbb{N}$ escolhido de tal forma que $M(G_n) < M(F) + \varepsilon/2$,

é fácil verificar que:

$$W(M, f, 1_\Omega, \varepsilon/2) \subseteq V(M, F, \varepsilon);$$

vejamos agora que $V \gg W$:

dado que $W(M, f_1, f_2, \dots, f_k, \epsilon) = \bigcap_{j=1}^k W(M, f_j, \epsilon)$ será considerado apenas $W = W(M, f, \epsilon)$,

também, para facilitar a notação, será suposto que

$$\forall \omega \in \Omega \quad 0 < f(\omega) < 1,$$

o que não implica perda de generalidade, já que $\forall f \in CL(\Omega, d)$, se $a < \inf\{f(\omega) : \omega \in \Omega\}$ e $b > \sup\{f(\omega) : \omega \in \Omega\}$ tem-se que $g = (b-a)^{-1}f$ satisfaz:

$$g \in CL(\Omega, d) \text{ e } 0 < g < 1 \text{ em } \Omega \text{ e } W(M, g, \epsilon) = W(M, f, (b-a)\epsilon),$$

seja então $W = W(M, f, \epsilon)$ com $0 < f < 1$:

sejam $\delta > 0$ a determinar posteriormente e $k \in \mathbb{N}$ tal que $k^{-1} < \delta$, defina-se

$$F_j = \{\omega : f(\omega) \geq j/k\}, \quad 0 \leq j \leq k \quad (\text{então, } F_0 = \Omega \text{ e } F_k = \emptyset),$$

considere-se agora

$$V(M, F_1, F_2, \dots, F_k, \delta),$$

se $\sum$ denota união disjunta, dado que $0 < f < 1$, tem-se que:

$$\Omega = \sum_{j=1}^k \left\{ \omega : (j-1)/k \leq f(\omega) \leq j/k \right\},$$

após algumas simplificações

$$\forall Q \in \mathcal{M}^+ \text{ vale } \sum_{j=1}^k Q(F_j) \leq \int f dQ < k^{-1} Q(\Omega) + k^{-1} \sum_{j=1}^k Q(F_j),$$

portanto, $\forall Q \in V(M, F_1, F_2, \dots, F_k, \delta)$ vale que

$$\text{ffd}Q < k^{-1}Q(\Omega) + k^{-1} \sum_{j=1}^k [M(F_j) + \delta] < k^{-1}Q(\Omega) + \text{ffd}M + \delta,$$

e como $k^{-1} < \delta$ e $Q \in V(M, F_1, F_2, \dots, F_k) \Rightarrow Q(\Omega) < M(\Omega) + \delta$,

$$\forall Q \in V(M, F_1, F_2, \dots, F_k, \delta) \text{ vale } \text{ffd}Q < \text{ffd}M + \delta(M(\Omega) + \delta + 1),$$

portanto basta escolher $\delta > 0$ tal que $\delta^2 + \delta(M(\Omega) + 1) < \varepsilon$, o que sempre é possível, já que $h(\delta) = \delta^2 + \delta(M(\Omega) + 1)$ é contínua, $h(\delta) > 0$ se $\delta \neq 0$ e $h(0) = 0$;

a desigualdade $\text{ffd}Q > \text{ffd}M - \varepsilon$ demonstra-se argumentando de forma similar com $1-f$ em lugar de f ;

2) $\mathcal{V}$ equivale a $\mathcal{U}$, decorre da seguinte afirmação de verificação simples:

$\forall \varepsilon > 0 \forall F$ fechado em (Ω, d) e $\forall G$ aberto em (Ω, d) vale que

$$V(M, G^c, \varepsilon) \subseteq U(M, G, 2\varepsilon) \text{ e } U(M, F^c, \varepsilon) \subseteq V(M, F, 2\varepsilon);$$

3) $\mathcal{V}$ equivale a $\mathcal{Y}$, parte $\mathcal{V} \gg \mathcal{Y}$,

é consequência da equivalência entre $\mathcal{V}$ e $\mathcal{U}$ e da afirmação:

se $M(\delta A) = 0$ tem-se que: $V(M, A^-, \varepsilon) \cap U(M, A^0, \varepsilon) \subseteq Y(M, A, \varepsilon)$, que demonstramos a seguir:

$$M(\delta A) = 0 \Rightarrow M(A^0) = M(A) = M(A^-),$$

portanto, dado que sempre $Q(A) \leq Q(A^-)$ e $Q(A) \geq Q(A^0)$, se $M(\delta A) = 0$ vale que:

$$\left\{ Q: Q(A^-) < M(A^-) + \varepsilon \right\} \cap \left\{ Q: Q(A^0) > M(A^0) + \varepsilon \right\} \subseteq \left\{ Q: |Q(A) - M(A)| < \varepsilon \right\},$$

ou seja, finalmente, se $M(\delta A) = 0$ tem-se que:

$$V(M, A^-, \varepsilon) \cap U(M, A^0, \varepsilon) \subseteq Y(M, A, \varepsilon);$$

falta ver que $V \ll Y$:

utilizaremos o seguinte resultado:

sejam M uma medida finita em Ω e $\{A_\lambda\}_{\lambda \in I}$ uma família de subconjuntos mensuráveis de Ω tal que:

$$\forall (\lambda, \tau) \in I \times I \quad \lambda \neq \tau \Rightarrow A_\lambda \cap A_\tau = \emptyset$$

então $\exists E \subseteq \Omega$, E enumerável tal que $M(A_\lambda) = 0 \quad \forall \lambda \in E^c$

($\forall n \in \mathbb{N}$ a finitude de M implica que $\{\lambda: M(A_\lambda) > n^{-1}\}$ é finito, em consequência $\{\lambda: M(A_\lambda) > 0\} = \bigcup_{n \in \mathbb{N}} \{\lambda: M(A_\lambda) > n^{-1}\}$ é enumerável);

para qualquer conjunto F , seja $A_\nu = \{\omega: d(\omega, F) = \nu\}$, $\nu \geq 0$,

os $A_\nu \in \mathcal{B}$ (são fechados) e satisfazem $\nu \neq \tau \Rightarrow A_\nu \cap A_\tau = \emptyset$,

também $\delta(F^\nu) \subseteq \{\omega: d(\omega, F) = \nu\} = A_\nu$,

então, dado o resultado acima:

$\exists \nu > 0$ tal que $M(\delta(F^\nu)) = 0$ e $M(F^\nu) < M(F) + \varepsilon$,

portanto,

dados F fechado e $\varepsilon > 0$ $\exists \nu > 0$ tal que:

$$\{Q: |Q(F^\nu) - M(F^\nu)| < \varepsilon\} \subseteq \{Q: Q(F) < M(F) + 2\varepsilon\}. \square$$

Proposição 2.25.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos em } (\Omega, d)\}$, $\mathcal{M}^+ = \{\text{medidas finitas em } (\Omega, d)\}$. Consideremos a topologia fraca $\mathcal{T}$ em $\mathcal{M}^+$. Então:

a) as medidas a suporte finito são densas em $\mathcal{M}^+$;

b) (Ω, d) é separável $\Rightarrow (\mathcal{M}^+, \mathcal{T})$ é separável.

Demonstração:

a) sejam $k \in \mathbb{N}$, $G_1, G_2, \dots, G_k$ abertos em (Ω, d) e $\varepsilon > 0$, considere-se a partição finita gerada por $G_1, G_2, \dots, G_k$, isto é a família de conjuntos:

$$\mathcal{A} = \left\{ G_1^* \cap G_2^* \cap \dots \cap G_k^* : G_j^* = G_j \text{ ou } G_j^* = G_j^c, 1 \leq j \leq k \right\};$$

se $\mathcal{A} = \{A_1, A_2, \dots, A_m\}$, para cada i , $1 \leq i \leq m$, escolhe-se um ponto qualquer $\omega_i \in A_i$ e define-se $L \in \mathcal{M}^+$ por:

$$L = \sum_{i=1}^m M(A_i) \times 1_{\omega_i},$$

onde 1_{ω_i} é a medida de probabilidade com massa 1 em ω_i ,

dado que $L(A_i) = M(A_i)$, $1 \leq i \leq m \Rightarrow L(G_j) = M(G_j)$, $1 \leq j \leq k$, tem-se que

$$L \in V(M, G_1, G_2, \dots, G_k, \varepsilon),$$

dado que $\#(\text{suporte } L) \leq m \leq 2^k$ e que, em espaços métricos qualquer ponto é fechado (portanto boreliano), o resultado está provado;

b) sejam (Ω, d) separável e D denso enumerável em (Ω, d) , na demonstração anterior foi provado que o conjunto:

$$\mathcal{D} = \{L \in \mathcal{M}^+; \text{suporte}(L) \text{ é finito}\} \text{ é denso em } \mathcal{M}^+,$$

seja agora

$$\mathcal{D}^* = \left\{ L \in \mathcal{M}^+; \begin{array}{l} \text{suporte}(L) \subseteq D, \text{ suporte}(L) \text{ finito} \\ \text{átomos de } L \text{ tem massa racional} \end{array} \right\},$$

é claro que $\mathcal{D}^* \subseteq \mathcal{D}$ e que $\mathcal{D}^*$ é enumerável,

provaremos que

$$\mathcal{D} \subseteq (\mathcal{D}^*)^- \text{ (ou seja que } \mathcal{D}^* \text{ é denso em } \mathcal{D}),$$

o que implicará (dado que $\mathcal{D}^- = \mathcal{M}^+$) que $\mathcal{D}^*$ é denso em $\mathcal{M}^+$ e portanto que $\mathcal{M}^+$ é separável (já que $\mathcal{D}^*$ é enumerável);

para facilitar a demonstração de " $\mathcal{D} \subseteq (\mathcal{D}^*)^-$ " utilizaremos ferramentas e resultados do capítulo seguinte:

consideremos a métrica de Prohorov definida no espaço $\mathcal{M}^+$ por $\Pi(M, L) = \inf \left\{ \varepsilon > 0; M(B) \leq L(B^\varepsilon) + \varepsilon \wedge L(B) \leq M(B^\varepsilon) + \varepsilon \forall B \in \mathcal{B} \right\},$

onde $B^\varepsilon = \{\omega \in \Omega : d(\omega, B) < \varepsilon\}$

(para maiores informações veja Definição 3.1),

provaremos que $\mathcal{D}^*$ é Π -denso em $\mathcal{D}$ e, dado que a Proposição 3.4 garante que Π induz em $\mathcal{M}^+$ uma topologia mais fina que $\mathcal{T}$, teremos então que $\mathcal{D}^*$ é $\mathcal{T}$ -denso em $\mathcal{D}$;

consideremos $\omega_1, \omega_2, \dots, \omega_n, \gamma_1, \gamma_2, \dots, \gamma_n$ elementos de Ω , $r_1, r_2, \dots, r_n, q_1, q_2, \dots, q_n$ reais positivos e τ definido por:

$$\tau = \max \left\{ \max\{d(\omega_j, \gamma_j) : 1 \leq j \leq n\}, n \cdot \max\{|r_j - q_j| : 1 \leq j \leq n\} \right\},$$

sejam M e N as medidas definidas por:

$$M = \sum_{j=1}^n r_j \cdot 1_{\{\omega_j\}}, \quad N = \sum_{j=1}^n q_j \cdot 1_{\{\gamma_j\}}$$

se $\varepsilon > \tau \Rightarrow \varepsilon > \max\{d(\omega_j, \gamma_j), 1 \leq j \leq n\}$, então:

$$\omega_j \in A \Rightarrow \gamma_j \in A^\varepsilon, \quad \gamma_j \in A \Rightarrow \omega_j \in A^\varepsilon, \quad 1 \leq j \leq n,$$

portanto

$$M(A) = \sum_{j=1}^n r_j \cdot 1_{\{\omega_j \in A\}} \leq \sum_{j=1}^n r_j \cdot 1_{\{\gamma_j \in A^\varepsilon\}}$$

e

$$N(A) = \sum_{j=1}^n q_j \cdot 1_{\{\gamma_j \in A\}} \leq \sum_{j=1}^n q_j \cdot 1_{\{\omega_j \in A^\varepsilon\}};$$

por outro lado, também $\varepsilon > \tau \Rightarrow \varepsilon > n \cdot \max\{|r_j - q_j| : 1 \leq j \leq n\}$,

portanto

$$\sum_{j=1}^n r_j \cdot 1_{\{\gamma_j \in A^\varepsilon\}} \leq \sum_{j=1}^n q_j \cdot 1_{\{\gamma_j \in A^\varepsilon\}} + \varepsilon = N(A^\varepsilon) + \varepsilon$$

e

$$\sum_{j=1}^n q_j \cdot 1_{\{\omega_j \in A^\varepsilon\}} \leq \sum_{j=1}^n r_j \cdot 1_{\{\omega_j \in A^\varepsilon\}} + \varepsilon = M(A^\varepsilon) + \varepsilon,$$

em consequência

$$\varepsilon > \tau \Rightarrow M(A) \leq N(A^\varepsilon) + \varepsilon \quad \text{e} \quad N(A) \leq M(A^\varepsilon) + \varepsilon,$$

que por sua vez implica que

$\Pi(M, N) \leq \tau$, por definição de Π ;

finalmente, seja $\varepsilon > 0$ fixo e consideremos uma medida qualquer $M \in \mathcal{D}$, M

é da forma: $M = \sum_{j=1}^n r_j \cdot 1_{\{\omega_j\}}$, com $\omega_j \in \Omega$, r_j real, $r_j > 0$, $1 \leq j \leq n$,

existe uma medida $N \in \mathcal{D}^*$ definida por:

$$N = \sum_{j=1}^n q_j \cdot 1_{\{\gamma_j\}}, \quad \text{com } \gamma_j \in D \text{ e } q_j \text{ racional, } q_j > 0 \quad 1 \leq j \leq n,$$

de tal forma que

$$\tau = \max \left\{ \max\{d(\omega_j, \gamma_j) : 1 \leq j \leq n\}, \quad n \cdot \max\{|r_j - q_j| : 1 \leq j \leq n\} \right\}$$

satisfaz $\tau < \varepsilon$ (pois $D^- = \Omega$ e $\{\text{racionais}\}^- = \{\text{reais}\}$),

pelas considerações anteriores

$\Pi(M, N) \leq \tau$, portanto D^* é Π -denso em D . $\square$

CAPÍTULO 3

A DISTÂNCIA DE PROHOROV E

A METRIZAÇÃO DA TOPOLOGIA FRACA.

3.A. INTRODUÇÃO E RESUMO DO CAPÍTULO.

Todo este capítulo está dedicado a um importantíssimo trabalho de Prohorov de 1956 (ver referências), do qual mencionaremos alguns dos principais resultados a seguir:

sejam:

(Ω, d) um espaço métrico separável e completo,

$\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$\mathcal{M}^+(\Omega, \mathcal{B}) = \{\text{medidas finitas sobre } (\Omega, \mathcal{B})\}$, e

Π_d definida em $\mathcal{M}^+ \times \mathcal{M}^+$ por:

$$\Pi_d(M, L) = \inf \left\{ \varepsilon > 0 : M(B) \leq L(B^\varepsilon) + \varepsilon \quad \wedge \quad L(B) \leq M(B^\varepsilon) + \varepsilon \quad \forall B \in \mathcal{B} \right\}$$

onde $B^\varepsilon = \{\omega \in \Omega : d(\omega, B) < \varepsilon\}$.

nestas condições:

1) Π_d é uma distância (hoje conhecida como "métrica de Prohorov") que metriza a topologia fraca $\mathcal{T}$ (definida em 2.21) no espaço $\mathcal{M}^+$, isto é a topologia induzida por Π_d em $\mathcal{M}^+$ é a topologia $\mathcal{T}$;

2) o espaço métrico $(\mathcal{M}^+, \Pi_d)$ é também separável e completo;

3) seja N subconjunto de $(\mathcal{M}^+, \Pi_d)$

N é relativamente compacto se e $\left\{ \begin{array}{l} N \text{ é tight} \\ e \\ \sup\{M(\Omega) : M \in N\} < \infty \end{array} \right\}$,

portanto, se $N \subseteq \mathcal{F}$, compacidade equivale a tightness.

Em particular, mediante a utilização do resultado 3) acima, Prohorov caracterizou os subconjuntos compactos de $(\mathcal{M}^+, \Pi_d)$ nos dois casos clássicos seguintes:

i) $\Omega = C[0,1] = \left\{ f: [0,1] \xrightarrow{\text{contínua}} (R, e) \right\}$, d é a métrica sup,

ii) $\Omega = D[0,1] = \left\{ f: [0,1] \xrightarrow[\text{da primeira espécie}]{\text{com discontinuidades}} (R, e) \right\}$, sendo d a

métrica de Skorohod.

Todos estes resultados, com excepção dos dois casos particulares acima referidos, serão provados aqui, inclusive com maior generalidade que no resumo acima.

Na Seção B abordaremos o caso em que $\Omega = R$, o conjunto dos números reais, $d = e$, a métrica euclidiana usual e $\mathcal{B}$ são os respectivos conjuntos borelianos. Neste caso será definida a métrica L de Levy em $(\mathcal{M}^+((R, \mathcal{B}), L)$ e

provado L metriza a topologia fraca (posteriormente, os resultados da seção D implicarão que a topologia fraca em $\mathcal{M}^+(\mathbb{R}, \mathcal{B})$ é também separável). Também será provado aqui que $(\mathcal{M}^+(\mathbb{R}, \mathcal{B}), L)$ é um espaço métrico completo, fato que será utilizado posteriormente na Seção E na demonstração da completitude $(\mathcal{M}^+(\Omega, \mathcal{B}), \Pi_d)$, onde (Ω, d) é um espaço métrico separável e completo e Π_d a métrica de Prohorov. As demonstrações desta seção são nossas.

Na Seção C será definida a métrica de Prohorov Π num contexto muito mais geral do que no caso anterior de L . Com, efeito, Π pode ser definida no conjunto $\mathcal{M}^+(\Omega, \mathcal{B})$, onde (Ω, d) é um espaço métrico qualquer. Entretanto, Prohorov definiu sua métrica certamente tomando a idéia essencial da métrica de Levy, daí a importância de L . Serão tratadas várias definições equivalentes de Π , e algumas de suas propriedades a respeito das aplicações:

$$\mathbb{R}^+ \times \mathcal{M}^+ \longrightarrow \mathcal{M}^+ \text{ definida por } (\lambda, M) \longrightarrow \lambda M, \text{ e}$$

$$\mathcal{M}^+ \times \mathcal{M}^+ \longrightarrow \mathcal{M}^+ \text{ definida por } (M, N) \longrightarrow M + N.$$

As demonstrações deste seção são nossas, com exceção da Proposição 3.9., parte c), que é uma adaptação da demonstração de Huber (1981), que foi feita para medidas de probabilidade. Alguns autores (por exemplo Huber, 1981) utilizam a definição alternativa da métrica de Prohorov dada na Proposição 3.9b), por esta causa e também para facilitar a comparação com a métrica de Levy no caso $\Omega = \mathbb{R}$, foi demonstrada a equivalência entre a definição original de Prohorov (que figura em 3.A) e a versão de Huber.

Na Seção D será provado que a topologia induzida por Π em $\mathcal{M}^+$ é mais fina que a topologia fraca e que as duas topologias são iguais quando $(\mathcal{M}^+, \mathcal{T})$ ($\mathcal{T}$ =topologia fraca) satisfaz o primeiro axioma de numerabilidade, isto é, quando existe base enumerável de vizinhanças de qualquer ponto em $(\mathcal{M}^+, \mathcal{T})$. Também provaremos aqui, que a separabilidade de (Ω, d) equivale à separabilidade de $(\mathcal{M}^+, \Pi)$, e portanto implica na metrizabilidade de $(\mathcal{M}^+, \Pi)$ (já que em espaços métricos separabilidade equivale à existência de base enumerável de abertos, propriedade que em qualquer espaço topológico implica o primeiro axioma de numerabilidade).

As demonstrações deste seção são adaptações das de Billingsley (1968, Apêndice III), que são restritas a medidas de probabilidade. A introdução da equivalência c) na Proposição 3.12 é nossa e ela implica que a separabilidade de $(\mathcal{M}^+, \Pi)$ equivale à do seu subespaço $(\mathcal{F}, \Pi/\mathcal{F})$, onde $\mathcal{F} = \{M \in \mathcal{M}^+ : M(\Omega) = 1\}$.

Na Seção E provaremos a completitude de $(\mathcal{M}^+(\Omega, \mathcal{B}), \Pi)$ sob a hipótese de completitude de (Ω, d) . Neste caso, também serão caracterizados os subconjuntos relativamente compactos de $(\mathcal{M}^+(\Omega, \mathcal{B}), \Pi)$.

As demonstrações nesta seção são essencialmente devidas a Prohorov (1956). A Proposição 3.16. parte a) e o Lema 3.20 estão apenas enunciados em Prohorov, as demonstrações são nossas. A Proposição 3.22 é nossa.

3.B. MEDIDAS FINITAS NOS BORELIANOS $\mathcal{B}$ de (R, e) . A MÉTRICA DE LEVY
 L EM $\mathcal{M}^+(R, \mathcal{B})$. COMPLETITUDE DE $(\mathcal{M}^+(R, \mathcal{B}), L)$,

DEFINIÇÃO 3.1.

Sejam R o conjunto dos números reais, e a métrica euclidiana usual
em R , $\mathcal{B} = \{\text{borelianos de } (R, e)\}$ e $\mathcal{M}^+(R, \mathcal{B}) = \{\text{medidas finitas em } (R, \mathcal{B})\}$.
Define-se $L: \mathcal{M}^+ \times \mathcal{M}^+ \rightarrow R^+$ por:

$$L(P, Q) = \inf \left\{ \varepsilon > 0 : G(t - \varepsilon) - \varepsilon \leq F(t) \leq G(t + \varepsilon) + \varepsilon, \forall t \in R \right\},$$

onde $P \in \mathcal{M}^+$, $Q \in \mathcal{M}^+$ e F e G são as respectivas funções de distribuição
(veja 2.19);

a função L assim definida é conhecida como a "métrica de Levy".

PROPOSIÇÃO 3.2.

a) a métrica de Levy L definida em 3.1 é uma distância propriamente
dita em $\mathcal{M}^+$, isto é, L satisfaz:

$$L \geq 0,$$

$$L(M, L) = 0 \text{ se } M = L,$$

$$L(M, L) = L(L, M),$$

$$L(M, N) \leq L(M, L) + L(L, N);$$

além disso vale que $L(M, L) \leq \text{máximo}\{M(\Omega), L(\Omega)\}$, portanto se $M \in \mathcal{F}$, $L \in \mathcal{F}$
 $\Rightarrow L(M, L) \leq 1$;

b) L pode ser alternativamente definida por:

$$L(P, Q) = \inf \left\{ \varepsilon > 0 : P(A) \leq Q((A^\varepsilon)^\sim) + \varepsilon \wedge Q(A) \leq P((A^\varepsilon)^\sim) + \varepsilon \quad \forall A \in \mathcal{A} \right\}, \text{ onde}$$

$$\mathcal{A} = \left\{ A \subseteq \mathbb{R} : \exists a \in \mathbb{R} \text{ tal que } A = (-\infty, a] \right\} \text{ e } (A^\varepsilon)^\sim = \{ \omega \in \Omega : e(\omega, A) \leq \varepsilon \}.$$

PROPOSIÇÃO 3.3.

Seja L a métrica de Levy definida em 3.1, então:

a) $L \leq K$, onde K é a distância de Kolmogoroff:

$$K(M, N) = \sup \{ |F(t) - G(t)| : t \in \mathbb{R} \},$$

F e G são as respectivas funções de distribuição de M e N ;

b) se Π_e é a métrica de Prohorov definida em 3.8 e e é a métrica euclidiana usual em $\mathbb{R}$, vale que

$$L \leq \Pi_e, \text{ além disso existem } P \text{ e } Q \text{ em } \mathcal{M}^+ \text{ tais que: } L(P, Q) < \Pi_e(P, Q).$$

PROPOSIÇÃO 3.4.

Se $\mathcal{B} = \{\text{borelianos de } (\mathbb{R}, e)\}$, a métrica de Levy L induz a topologia fraca $\mathcal{T}$ em $\mathcal{M}^+(\mathbb{R}, \mathcal{B})$.

Comentário: posteriormente, na Seção 3.D, provaremos que neste caso também a métrica de Prohorov Π_e induz $\mathcal{T}$, sendo portanto ambas métricas equivalentes. $\square$

DEFINIÇÃO 3.5.

$(\Omega, \mathcal{T})$ espaço topológico, $\mathcal{B} = \{\text{borelianos de } (\Omega, \mathcal{T})\}$, $\mathcal{M}^+ = \{\text{medidas finitas sobre } (\Omega, \mathcal{B})\}$, $\mathcal{N} \subseteq \mathcal{M}^+$.

$\mathcal{N}$ é tight $\stackrel{\text{DEF}}{\iff} \forall \varepsilon > 0 \exists K_\varepsilon$ compacto tal que $N(K_\varepsilon^c) < \varepsilon \forall N \in \mathcal{N}$.

PROPOSIÇÃO 3.6.

Consideremos a métrica de Levy L definida em $\mathcal{M}^+(R, \mathcal{B})$, onde $\mathcal{B} = \{\text{borelianos de } (R, e)\}$. Seja $\{M_n : n \in \mathbb{N}\} \subseteq \mathcal{M}^+(R, \mathcal{B})$.

Se $\{M_n : n \in \mathbb{N}\}$ for uma seqüência de Cauchy a respeito de L , então:

- a) $\{M_n : n \in \mathbb{N}\}$ é tight; b) $\{M_n(\Omega) : n \in \mathbb{N}\}$ é Cauchy em R .

Comentário:

a idéia da demonstração é simples: já que todas as M_n são tight, dado $\rho > 0$, existe para qualquer uma das M_n um compacto $K = K(\rho, n)$ tal que $M_n(K(\rho, n))^c < \rho$; pelo fato de $\{M_n, n \in \mathbb{N}\}$ ser uma seqüência de Cauchy é possível escolher um k_0 apropriado de tal forma que $\forall n \geq k_0$ $M_n(K^*(\rho, k_0))^c < h(\rho)$, onde $K^*(\rho, k_0)$ é um compacto obtido mediante uma expansão de $K(\rho, k_0)$ e $h(\rho)$ é alguma função contínua e positiva de ρ com $h(0) = 0$; restam ainda as medidas $\{M_n, n < k_0\}$, mas estas são em número finito e todas são tight;

este resultado implica que toda seqüência Π_e -Cauchy também é tight, já que toda seqüência Π_e -Cauchy é necessariamente L -Cauchy (dado que $L \leq \Pi_e$, por 3.3.b). $\square$

PROPOSIÇÃO 3.7.

Sejam $\mathcal{B} = \{\text{borelianos de } (R, e)\}$ e L métrica de Levy em $\mathcal{M}^+(R, \mathcal{B})$.
Então o espaço métrico $(\mathcal{M}^+(R, \mathcal{B}), L)$ é completo.

Comentário: o teorema de seleção de Helly que afirma que:

"sejam $F = \left\{ f: R \xrightarrow[\text{continua à direita}]{\text{não decrescente}} R \right\}$, $\{F_n: n \in \mathbb{N}\} \subseteq F$, λ e λ' reais tais
que $\lambda \leq F_n(t) \leq \lambda' \quad \forall t \in R \quad \forall n \in \mathbb{N}$,

nestas condições existem uma subsequência $\{F_{n_m}: m \in \mathbb{N}\}$ e uma $F \in F$ tais
que:

$$- \lambda \leq F(t) \leq \lambda' \quad \forall t \in R,$$

$$- \lim_{n \rightarrow \infty} F_{n_m}(t) = F(t) \quad \forall t \text{ ponto de continuidade de } F."$$

A versão do teorema de seleção de Helly aqui utilizada é levemente
mais geral que a que aparece em Billingsley (1968, Apêndice II, p, 226),
em que λ e λ' acima são 0 e 1, respectivamente. Entretanto a demonstração
de Billingsley permanece válida, já que ela está baseada em dois fatos:

i) toda $F \in F$ identifica-se a $(F(q): q \text{ racional}) \in R^{\infty}$, fato claramente
válido na versão Billingsley e na nossa,

ii) o critério de compacidade em R^{∞} : seja $r = (r_j: j \in \mathbb{N})$ um ponto de
 R^{∞} , então A^- é compacto se e só se $\forall j \in \mathbb{N} \{r_j: r \in A\}$ é um conjunto limitado de R , que
obviamente funciona em ambas as versões.

A demonstração desta proposição estará baseada no teorema de Helly:
se $\{M_n: n \in \mathbb{N}\}$ for uma sequência em $\mathcal{M}^+(R, \mathcal{B})$, a sequência $\{F_n: n \in \mathbb{N}\}$ das

respectivas funções de distribuição está contida em F ; se a seqüência for L -Cauchy necessariamente $\sup_n \sup\{|F_n(t)| : t \in \mathbb{R}\} < \infty$, portanto o teorema de Helly garante a existência da subsequência F_m e da $F \in F$ de tal modo que $\lim_m F_m(t) = F(t) \quad \forall t$ F -ponto de continuidade. Provaremos que a função de conjunto M definida por F é uma medida e que $M_m \xrightarrow{w} M$; dado que $(M^+(\mathbb{R}, \mathcal{B}), \mathcal{T})$ é um espaço de Hausdorff isto implicará que $M_n \xrightarrow{w} M$. $\square$

COROLÁRIO DAS PROPOSIÇÕES 3.4, 3.7 e 3.12.

$(M^+(\mathbb{R}, \mathcal{B}), \Pi_e)$ é completo.

Comentário: o resultado decorre de 3.7 e dos fatos:

i) $L \leq \Pi_e$, então $\{M_n : n \in \mathbb{N}\} \Pi_e$ -Cauchy $\Rightarrow \{M_n : n \in \mathbb{N}\} L$ -Cauchy,

ii) L e Π_e induzem em $M^+(\mathbb{R}, \mathcal{B})$ a topologia fraca $\mathcal{T}$ (veja proposições 3.4 e 3.12), portanto:

$\{M_n : n \in \mathbb{N}\}$ é Π_e -convergente se e só se $\{M_n : n \in \mathbb{N}\}$ é L -convergente.

Na Seção 3.E será generalizado o resultado acima, provando-se a completude de $(M^+(\Omega, \mathcal{B}), \Pi_d)$, onde (Ω, d) é um espaço métrico separável e completo e $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$. $\square$

3.C. MEDIDAS FINITAS NOS BORELIANOS $\mathcal{B}$ DE UM ESPAÇO MÉTRICO QUALQUER (Ω, d) . A MÉTRICA DE PROHOROV EM $M^+(\Omega, \mathcal{B})$,

DEFINIÇÃO 3.8.

Sejam:

(Ω, d) um espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+ = \{\text{medidas finitas não negativas sobre } (\Omega, \mathcal{B})\}$;

nestas condições define-se $\Pi_d: \mathcal{M}^+ \times \mathcal{M}^+ \rightarrow \mathbb{R}^+$ por:

$$\Pi_d(M, L) = \inf \left\{ \varepsilon > 0 : M(B) \leq L(B^\varepsilon) + \varepsilon \quad \wedge \quad L(B) \leq M(B^\varepsilon) + \varepsilon \quad \forall B \in \mathcal{B} \right\}$$

ou equivalentemente por:

$$\Pi_d(M, L) = \max \left\{ \begin{array}{l} \inf \{ \varepsilon > 0 : M(B) \leq L(B^\varepsilon) + \varepsilon \quad \forall B \in \mathcal{B} \} \\ \inf \{ \varepsilon > 0 : L(B) \leq M(B^\varepsilon) + \varepsilon \quad \forall B \in \mathcal{B} \} \end{array} \right\}$$

onde em todos os casos $B^\varepsilon = \{\omega \in \Omega : d(\omega, B) < \varepsilon\}$,

e também por

$$\Pi_d(M, L) = \inf \left\{ \varepsilon > 0 : M(B) \leq L((B^\varepsilon)^\sim) + \varepsilon \quad \wedge \quad L(B) \leq M((B^\varepsilon)^\sim) + \varepsilon \quad \forall B \in \mathcal{B} \right\},$$

onde $(B^\varepsilon)^\sim = \{\omega \in \Omega : d(\omega, B) \leq \varepsilon\}$;

Π_d assim definida é uma distância.

Se as medidas M e L têm a mesma massa total, ou seja, se $M(\Omega) = L(\Omega)$, o primeiro conjunto de desigualdades implica o segundo e vice-versa, portanto

$$\Pi_d(M, L) = \inf \left\{ \varepsilon > 0 : M(B) \leq L(B^\varepsilon) + \varepsilon \quad \forall B \in \mathcal{B} \right\}$$

$$\text{e também} \quad \Pi_d(M, L) = \inf \left\{ \varepsilon > 0 : L(B) \leq M(B^\varepsilon) + \varepsilon \quad \forall B \in \mathcal{B} \right\},$$

Quando não houver dúvidas a respeito do espaço (Ω, d) original, Π_d poderá ser denotada simplesmente por Π .

Dada a regularidade das medidas finitas definidas nos borelianos de um espaço métrico (Proposição A.1.1), em todas as fórmulas precedentes da definição de Π_d " $\forall B \in \mathcal{B}$ " pode ser substituído por " $\forall B$ fechado".

Estas afirmações a respeito da métrica de Π_d , conhecida como "métrica de Prohorov", serão formalmente demonstradas na seguinte proposição.

PROPOSIÇÃO 3.9.

Sejam (Ω, d) um espaço métrico, $\mathcal{B} = \{ \text{borelianos de } (\Omega, d) \}$,

$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$ e $\mathcal{F} = \{ F \in \mathcal{M}^+ : F(\Omega) = 1 \}$. Então:

a) a métrica de Prohorov é uma distância propriamente dita, isto é, $\Pi = \Pi_d$ satisfaz:

$$\Pi \geq 0,$$

$$\Pi(M, L) = 0 \text{ se } M = L,$$

$$\Pi(M, L) = \Pi(L, M),$$

$$\Pi(M, N) \leq \Pi(M, L) + \Pi(L, N);$$

além disso vale que $\Pi(M, L) \leq \text{máximo}\{M(\Omega), L(\Omega)\}$, portanto se $M \in \mathcal{F}$, $L \in \mathcal{F}$
 $\Rightarrow \Pi(M, L) \leq 1$;

b) a métrica de Prohorov Π_d pode ser alternativamente definida por:

$$\Pi_d(M, L) = \inf \left\{ \varepsilon > 0 : M(B) \leq L((B^\varepsilon)^\sim) + \varepsilon \wedge L(B) \leq M((B^\varepsilon)^\sim) + \varepsilon \quad \forall B \in \mathcal{B} \right\},$$

onde $(B^\varepsilon)^\sim = \{\omega \in \Omega : d(\omega, B) \leq \varepsilon\}$;

c) se $M \in \mathcal{M}$, $N \in \mathcal{M}$, $M(\Omega) = L(\Omega)$ e $\varepsilon > 0$, tem-se que:

$$\forall B \in \mathcal{B} \quad M(B) \leq L(B^\varepsilon) + \varepsilon \Rightarrow \forall B \in \mathcal{B} \quad L(B) \leq M(B^\varepsilon) + \varepsilon,$$

portanto na definição 3.8. da métrica de Prohorov o primeiro conjunto de desigualdades implica o segundo (e vice-versa, dada a simetria da situação), que então pode ser dispensado;

d) na definição da métrica de Prohorov " $\forall B \in \mathcal{B}$ " pode ser substituído por " $\forall B$ fechado".

PROPOSIÇÃO 3.10.

Sejam (Ω, d) um espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}, \quad \mathcal{F} = \left\{ F \in \mathcal{M}^+ : F(\Omega) = 1 \right\} \text{ e } \Pi_d = \Pi \text{ a métrica de}$$

Prohorov em $\mathcal{M}^+$. Então:

$$(a) \quad \forall \lambda \geq 0 \quad \forall (P, Q) \in \mathcal{M}^+ \times \mathcal{M}^+ : \Pi(\lambda P, \lambda Q) \leq (\lambda \vee 1) \Pi(P, Q),$$

$$(b) \quad \forall 0 \leq \lambda \leq 1 \quad \forall P \in \mathcal{M}^+ \text{ tal que } P(\Omega) \leq 1 \text{ vale que } \Pi(\lambda P, P) \leq 1 - \lambda,$$

$$(c) \quad \forall (P, Q, F, G) \in (\mathcal{M}^+)^4 : \Pi(P+Q, F+G) \leq 2\{\Pi(P, F) \vee \Pi(Q, G)\};$$

$$(d) \quad \text{as aplicações:} \quad (\lambda, P) \mapsto \lambda P \quad \text{e} \quad (P, Q) \mapsto P+Q,$$

de $\mathbb{R}^+ \times \mathcal{M}^+ \mapsto \mathcal{M}^+$ e de $\mathcal{M}^+ \times \mathcal{M}^+ \mapsto \mathcal{M}^+$, são contínuas, onde as topologias em $\mathbb{R}^+$ e $\mathcal{M}^+$ são a usual e a induzida por Π , respectivamente.

Comentário: no caso da topologia fraca, a continuidade das operações $+$ e $*$ é um fato conhecido. O resultado aqui apresentado é um pouco mais geral, já que a topologia induzida pela métrica de Prohorov é sempre mais fina que a topologia fraca. Além disso são de interesse as propriedades (a), (b) e (c), que serão utilizadas, por exemplo, na demonstração da Proposição 3.22. $\square$

3.D. SEPARABILIDADE E METRIZABILIDADE DE $\mathcal{M}^+$.

PROPOSIÇÃO 3.11.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas}^{\text{finitas}} \text{ não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}, \quad \mathcal{T} \text{ a topologia fraca e } \Pi_d \text{ a métrica de}$$

Prohorov em $\mathcal{M}^+$. Então:

- 1) $\mathcal{T}$ é sempre menos fina que Π_d ;
- 2) são equivalentes as seguintes afirmações:
 - a) $M \in \mathcal{M}^+$ e existe base enumerável de $\mathcal{T}$ -vizinhanças de M ;
 - b) $M \in \mathcal{M}^+$ e M tem suporte separável;
 - c) $\left\{ \mathcal{T}\text{-vizinhanças de } M \right\} = \left\{ \Pi_d\text{-vizinhanças de } M \right\}$.

Comentário: o resultado 2) refere-se a propriedades locais, já 1), o corolário seguinte e a Proposição 3.12 são de caráter global. $\square$

COROLÁRIO DA PROPOSIÇÃO 3.11.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}, \quad \mathcal{T} \text{ a topologia fraca e } \Pi_d \text{ a métrica de}$$

Prohorov em $\mathcal{M}^+$. Então:

$\mathcal{T}$ é metrizável se e a métrica de Prohorov Π_d metriza $\mathcal{T}$.

Comentário: é uma consequência de que em todo ponto de um espaço métrico existe base enumerável de vizinhanças, veja demonstração em 3.F. $\square$

PROPOSIÇÃO 3.12.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}, \quad \mathcal{T} \text{ a topologia fraca e } \Pi_d \text{ a métrica}$$

de Prohorov em $\mathcal{M}^+$.

Nestas condições são equivalentes as seguintes afirmações:

a) o espaço métrico (Ω, d) é separável;

b) $(\mathcal{M}^+, \mathcal{T}) = (\mathcal{M}^+, \Pi_d)$ e $\mathcal{T}$ é separável;

c) as topologias induzidas em $\mathcal{F} = \{M \in \mathcal{M}^+ : M(\Omega) = 1\}$ por $\mathcal{T}$ e Π_d coincidem e além disso $\mathcal{F}$ é $\mathcal{T}$ (e Π_d)-separável.

3.E. COMPLETITUDE E COMPACIDADE EM $\mathcal{M}^+$.

COMENTÁRIO 3.13.

Recapitularemos, aqui, alguns resultados e definições gerais sobre noções conexas com compacidade e completitude.

Seja $(Y, \mathcal{Y})$ um espaço topológico, $A \subseteq Y$:

$$A \text{ é compacto} \stackrel{\text{DEF}}{\Leftrightarrow} \begin{cases} A \subseteq \bigcup_{\lambda \in I} A_{\lambda}, A_{\lambda} \in \mathcal{Y} \ \forall \lambda \in I, \text{ então} \\ \exists F \text{ finito, } F \subseteq I, \text{ tal que } A \subseteq \bigcup_{\lambda \in F} A_{\lambda} \end{cases};$$

As seguintes propriedades são equivalentes:

a) A é compacto,

b) toda rede em A tem sub-rede convergente a um ponto de A ;

c) $\{F_{\lambda}\}_{\lambda \in I}$ família de subconjuntos fechados de A com a propriedade de intersecção finita (isto é J finito, $J \subseteq I \Rightarrow \bigcap_{\lambda \in J} F_{\lambda} \neq \emptyset$), então $\bigcap_{\lambda \in I} F_{\lambda} \neq \emptyset$.

A é relativamente compacto $\stackrel{\text{DEF}}{\Leftrightarrow} \bar{A} = \text{aderência}(A)$ é compacto.

Qualquer subconjunto fechado de um compacto é ele mesmo compacto, se o espaço topológico for de Hausdorff (como no caso dos espaços metrizáveis) compacto $\Rightarrow$ fechado.

A é seqüencialmente compacto $\stackrel{\text{DEF}}{\Leftrightarrow}$ toda seqüência de elementos de A contém uma subsequência convergente a um elemento de A ;

portanto vale que: A compacto $\Rightarrow A$ seqüencialmente compacto.

Seja $(Y, \mathcal{Y})$ espaço topológico:

$$(Y, \mathcal{Y}) \text{ é de Lindeloff} \stackrel{\text{DEF}}{\Leftrightarrow} \begin{cases} Y \subseteq \bigcup_{\lambda \in I} A_{\lambda}, A_{\lambda} \in \mathcal{Y} \quad \forall \lambda \in I, \text{ então} \\ \exists F \text{ enumerável}, F \subseteq I, \text{ tal que } A \subseteq \bigcup_{\lambda \in F} A_{\lambda}; \end{cases}$$

Em todo espaço de Lindeloff $(Y, \mathcal{Y})$ são equivalentes:

- a) A é seqüencialmente compacto,
- b) A é compacto;

esta equivalência vale também em espaços topológicos metrizáveis:

com efeito, se $(Y, \mathcal{Y})$ for metrizável e $A \subseteq Y$ o seu subespaço $(A, \text{topologia relativa})$ também é; também não é difícil provar que qualquer uma das condições a) e b) acima em espaços metrizáveis implica separabilidade, que por sua vez -sempre em espaços metrizáveis- equivale à existência de base de abertos enumerável, que em qualquer espaço topológico implica a propriedade de Lindeloff; portanto também vale que:

$(Y, \mathcal{Y})$ espaço topológico metrizável, são equivalentes:

- a) A é seqüencialmente compacto,
- b) A é compacto.

A partir daqui nosso espaço topológico será um espaço métrico (Ω, d) , e consideraremos a questão da completitude.

Seja $\{x_n, n \in \mathbb{N}\}$ uma seqüência, $\{x_n, n \in \mathbb{N}\} \subseteq \Omega$,

$\{x_n, n \in \mathbb{N}\}$ é de Cauchy $\stackrel{\text{DEF}}{\Leftrightarrow}$

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N}, \text{ tal que } n \geq n_0, m \geq n_0 \Rightarrow d(x_n, x_m) \leq \varepsilon.$$

Toda seqüência convergente de um espaço métrico é de Cauchy. Um espaço métrico é, por definição, completo, se toda seqüência de Cauchy for convergente. Um subconjunto de um espaço métrico completo é também completo se ele for fechado.

A completitude não é um invariante topológico, ela depende estritamente da métrica: já que o fato de uma seqüência ser convergente é invariante por homeomorfismos, mas o fato de uma seqüência ser de Cauchy não é.

Por exemplo, os conjuntos $(-1,1)$ e $\mathbb{R}$, ambos munidos da métrica euclidiana, são homeomorfos já que $h: \mathbb{R} \rightarrow (-1,1)$ definido por $h(t) = t/(1+|t|)$ é um homeomorfismo. Entretanto a seqüência $\{n: n \in \mathbb{N}\}$ não é Cauchy em $\mathbb{R}$, mas $\{h(n): n \in \mathbb{N}\} = \{n/(n+1): n \in \mathbb{N}\}$ é Cauchy em $(-1,1)$; é sabido que $(\mathbb{R}, e)$ é completo, o exemplo acima mostra que $((-1,1), e)$ não é, já que a seqüência $\{n/(n+1): n \in \mathbb{N}\}$ é de Cauchy mas não converge (em $(-1,1)$).

(Ω, d) métrico, se $\omega \in \Omega$ e $\delta > 0$ seja $B(\omega, \delta) = \{\omega' : d(\omega, \omega') < \delta\}$ e $A \subseteq \Omega$, por definição,

A é totalmente limitado see

$$\forall \delta > 0 \exists F = F(\delta) \text{ finito, } F \subseteq \Omega, \text{ tal que } A \subseteq \bigcup_{\omega \in F} B(\omega, \delta).$$

São verdadeiras as seguintes afirmações:

- a) A totalmente limitado e $B \subseteq A \Rightarrow B$ totalmente limitado;
 - b) união finita de conjuntos totalmente limitados é também totalmente limitado;
 - c) A totalmente limitado $\Leftrightarrow A^-$ totalmente limitado
- (($\Rightarrow$) decorre de $A \subseteq \bigcup_{\omega \in C} B(\omega, \delta)$, $C \subseteq \Omega$ (C não necessariamente finito) e $\alpha > 0 \Rightarrow A^- \subseteq \bigcup_{\omega \in C} B(\omega, \alpha + \delta)$; ($\Leftarrow$) é consequência de a)).

O seguinte resultado condensa algumas das relações entre compacidade e completitude:

(Ω, d) espaço métrico, $A \subseteq \Omega$, são equivalentes:

- a) A^- é compacto,
- b) toda seqüência de A tem subsequência convergente,
- c) A é totalmente limitado e A^- é completo;

portanto em (Ω, d) métrico, compacto implica completo.

Uma outra consequência do resultado acima é que se (Ω, d) for completo são equivalentes:

- a) A^- é compacto,
- b) A é totalmente limitado;

portanto em (Ω, d) completo:

a) A compacto see A fechado e totalmente limitado,

b) A relativamente compacto see A totalmente limitado.

Os fundamentos para provar todas as afirmações constantes neste comentário podem ser encontrados em livros de topologia geral, por exemplo, Kelley (1962).

A respeito do caso particular que nos ocupa, optamos aqui por seguir a linha de Prohorov, provando em primeira instância a completitude do espaço métrico (M^+, Π_d) e caracterizando posteriormente os seus subconjuntos compactos mediante a utilização dos resultados imediatamente acima.

Um outro caminho possível seria caracterizar os conjuntos relativamente compactos, provar que as seqüências de Cauchy são relativamente compactos e daí concluir a completitude.

LEMA 3.14.

Sejam (Ω, d) um espaço métrico ^{completo}, $\mathcal{B} = \{ \text{borelianos de } (\Omega, d) \}$,

$$M^+ = \left\{ \begin{array}{l} \text{medidas finitas não} \\ \text{negativas em } (\Omega, \mathcal{B}) \end{array} \right\}, \text{ e } N \subseteq M^+:$$

$$N \text{ é tight see } \left\{ \begin{array}{l} \forall \epsilon > 0 \quad \forall \delta > 0 \quad \exists F \subseteq \Omega, F = F(\epsilon, \delta) \text{ finito,} \\ \text{tal que } \forall M \in N \quad M \left(\left(\bigcup_{\omega \in F} B(\omega, \delta) \right)^c \right) < \epsilon \end{array} \right\},$$

$$\text{onde } B(\omega, \delta) = \left\{ \omega' \in \Omega : d(\omega', \omega) < \delta \right\}.$$

Comentário:

na Proposição 2.5 prova-se que toda medida finita em um espaço métrico completo é tight, a demonstração desse resultado sugere a utilidade da caracterização acima;

o resultado:

$$N \text{ é tight} \Rightarrow \left\{ \begin{array}{l} \forall \varepsilon > 0 \quad \forall \delta > 0 \quad \exists F \subseteq \Omega, \quad F = F(\varepsilon, \delta) \text{ finito,} \\ \text{tal que } \forall M \in N \quad M \left(\left(\bigcup_{\omega \in F} B(\omega, \delta) \right)^c \right) < \varepsilon \end{array} \right., \text{ onde } B(\omega, \delta) = \left\{ \omega' \in \Omega : d(\omega', \omega) < \delta \right\}, \text{ é válido em todo espaço métrico } (\Omega, d) \text{ (veja 3.F).} \square$$

LEMA 3.15.

(Ω, d) espaço métrico separável e completo,

$$\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}, \quad \mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas finitas não} \\ \text{negativas definidas em } (\Omega, \mathcal{B}) \end{array} \right\},$$

$\Pi = \Pi_d$ a métrica de Prohorov em $\mathcal{M}^+$. Nestas condições:

a) $\{M_n, n \in \mathbb{N}\}$ seqüência de Cauchy em $(\mathcal{M}^+, \Pi_d) \Rightarrow \{M_n, n \in \mathbb{N}\}$ é tight.

b) $\{M_n(\Omega), n \in \mathbb{N}\}$ é Cauchy em $\mathbb{R}$.

Comentário: Na Seção B provou-se este resultado para o caso particular $(\Omega, d) = (\mathbb{R}, e)$. $\square$

LEMA 3.16.

$$(\Omega, d) \text{ espaço métrico} \quad \mathcal{B} = \{\text{borelianos de } (\Omega, d)\}, \quad \mathcal{M}^+(\Omega, d) = \left\{ \begin{array}{l} \text{medidas finitas não} \\ \text{negativas definidas em } (\Omega, \mathcal{B}) \end{array} \right\},$$

$\Pi = \Pi_d$ a métrica de Prohorov em $\mathcal{M}^+(\Omega, d)$, $R = \{\text{reais}\}$, e a métrica euclidiana em R , L a métrica de Levy em $\mathcal{M}^+(R, e)$, $f: (\Omega, d) \rightarrow (R, e)$ contínua. Então:

a) K compacto em (Ω, d) e $\varepsilon > 0$, então $\exists \delta = \delta(\varepsilon, K)$ tal que:

$$x \in K^\delta, y \in K^\delta \text{ e } d(x, y) < \delta \Rightarrow e(f(x), f(y)) < \varepsilon;$$

b) $\{M_n, n \in \mathbb{N}\}$ sequência de Cauchy em $(\mathcal{M}^+(\Omega, d), \Pi_d)$

então: $\{M_n f^{-1}, n \in \mathbb{N}\}$ sequência de Cauchy em $(\mathcal{M}^+(R, e), L)$.

LEMA 3.17.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$\mathcal{M}^+(\Omega, \mathcal{B}) = \{\text{medidas finitas não negativas em } (\Omega, \mathcal{B})\}$, Π_d a métrica de Prohorov associada a d , $M \in \mathcal{M}^+(\Omega, \mathcal{B})$.

Seja $\{M_n, n \in \mathbb{N}\}$ uma sequência tight em $\mathcal{M}^+(\Omega, \mathcal{B})$

tal que $\forall f \in CL(\Omega, d) \exists J(f) = \lim_n \int f dM_n$ e $|J(f)| < \infty$,

Nestas condições:

a) $J: CL(\Omega, d) \rightarrow \mathbb{R}$, satisfaz: J é linear e que $f \geq 0 \Rightarrow J(f) \geq 0$;

b) existe uma única função de conjunto M definida na álgebra $\mathcal{B}^*$ gerada pelos abertos de (Ω, d) , finitamente aditiva e regular tal que:

$J(f) = \int f dM$, onde M regular significa que:

$$\forall B \in \mathcal{B}^* \quad M(B) = \sup\{M(F) : F \text{ fechado}, F \subseteq B\} = \inf\{M(G) : G \text{ aberto}, G \supseteq B\},$$

(é claro que M é finita, já que $M(\Omega) = \int 1 dM$);

c) $\forall \varepsilon > 0$ existe $K(\varepsilon)$ compacto tal que $M(K(\varepsilon)) \geq M(\Omega) - \varepsilon$;

d) $\forall B \in \mathcal{B}^* \quad M(B) = \sup\{M(K) : K \text{ compacto}, K \subseteq B\}$;

e) M é σ -aditiva em $\mathcal{B}^*$;

f) M tem única extensão a uma medida em $(\Omega, \mathcal{B})$.

PROPOSIÇÃO 3.18.

(Ω, d) espaço métrico separável e completo,

$\mathcal{B}_d = \{\text{borelianos de } (\Omega, d)\}$,

$\mathcal{M}^+(\Omega, d) = \{\text{medidas finitas não negativas sobre } (\Omega, \mathcal{B}_d)\}$, $\Pi = \Pi_d$ a métrica de Prohorov em $\mathcal{M}^+(\Omega, d)$.

Nestas condições, o espaço métrico $(\mathcal{M}^+, \Pi_d)$ é separável e completo.

LEMA 3.19.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$\mathcal{M}^+(\Omega, \mathcal{B}) = \{\text{medidas finitas não negativas em } (\Omega, \mathcal{B})\}$ e Π_d a métrica de Prohorov associada a d .

Sejam $M, f, \delta, \varepsilon$ e H tais que:

$M \in \mathcal{M}^+(\Omega, \mathcal{B})$, $f: (\Omega, \mathcal{B}) \xrightarrow{\text{mensurável}} (\Omega, \mathcal{B})$, $\delta > 0$, $H = \{\omega \in \Omega : d(\omega, f(\omega)) < \delta\}$ $\varepsilon > 0$ e $M(H^c) \leq \varepsilon$

então: $\Pi_d(M, Mf^{-1}) \leq \max\{\varepsilon, \delta\}$.

COROLÁRIO DO LEMA 3.19.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+(\Omega, \mathcal{B}) = \{\text{medidas finitas não negativas em } (\Omega, \mathcal{B})\}$, Π_d a métrica de Prohorov associada a d , $M \in \mathcal{M}^+(\Omega, \mathcal{B})$.

Dados $\{\omega_0, \omega_1, \omega_2, \dots, \omega_n\} \subseteq \Omega$ e $\delta > 0$, seja

$\psi = \psi(\omega_0, \omega_1, \omega_2, \dots, \omega_n, \delta)$, $\Psi: \Omega \rightarrow \Omega$ definida por:

$$\psi(\omega) = \omega_1 \text{ se } \omega \in B(\omega_1, \delta)$$

$$\psi(\omega) = \omega_j \text{ se } \omega \in B(\omega_j, \delta) - \bigcup_{i=1}^{j-1} B(\omega_i, \delta) \text{ se } 2 \leq j \leq n$$

$$\psi(\omega) = \omega_0 \text{ se } \omega \notin \bigcup_{i=1}^n B(\omega_i, \delta),$$

$$\text{onde } B(\omega, \delta) = \{\omega' \in \Omega: d(\omega', \omega) < \delta\},$$

nestas condições:

$$M\left(\left(\bigcup_{i=1}^n B(\omega_i, \delta)\right)^c\right) < \varepsilon \Rightarrow \Pi_d(M, M\psi^{-1}) < \text{máximo}\{\varepsilon, \delta\}.$$

LEMA 3.20.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+(\Omega, \mathcal{B}) = \{\text{medidas finitas não negativas em } (\Omega, \mathcal{B})\}$, Π_d a métrica de Prohorov associada a d , $M \in \mathcal{M}^+(\Omega, \mathcal{B})$.

O subconjunto finito de Ω e $a > 0$, nestas condições:

$$\mathcal{A} = \{M \in \mathcal{M}^+(\Omega, \mathcal{B}) : M(\Omega) \leq a \text{ e suporte}(M) \subseteq A\}$$

é um subconjunto totalmente limitado de $(M^+(\Omega, \mathcal{B}), \Pi_d)$.

PROPOSIÇÃO 3.21.

Sejam (Ω, d) espaço métrico separável e completo, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $M^+ = \left\{ \begin{array}{l} \text{medidas finitas não} \\ \text{negativas em } (\Omega, \mathcal{B}) \end{array} \right\}$, $\mathcal{T}$ a topologia fraca e Π_d a métrica de Prohorov em M^+ e $N \subseteq M^+$. Então:

$$N \text{ é } \mathcal{T}\text{-relativamente compacto} \iff \left\{ \begin{array}{l} N \text{ é tight} \\ \text{e} \\ \sup\{M(\Omega) : M \in N\} < \infty \end{array} \right\}.$$

Comentário: nossa demonstração da parte $(\Leftarrow)$ foi feita aqui segundo a versão de Prohorov e é bastante simples, já que sendo (Ω, d) separável e completo, $(M^+(\Omega, d), \Pi_d)$ também é (por Proposição 3.18), o que reduz a demonstração a provar que N é totalmente limitado.

A demonstração de $(\Leftarrow)$ que figura em Billingsley (1968), embora restrita a medidas de probabilidade, é mais complexa; porém tem o mérito ser válida em espaços métricos quaisquer, dispensando a hipótese de separabilidade e completude de (Ω, d) . A demonstração é feita em sucessivos passos: começa por R^k utilizando o teorema de seleção de Helly, a seguir reduz o caso R^∞ ao de R^k , mediante aplicação do teorema de consistência de Kolmogoroff. Depois passa a Ω σ -compacto e Ω qualquer mediante a aplicação dos seguintes resultados de fácil demonstração:

$$B_0 \in \mathcal{B}(\Omega, d), \mathcal{B}_0 = \{B \cap B_0 : B \in \mathcal{B}\}, M \in M^+(\Omega, \mathcal{B}), M^r = M/B_0,$$

$$M \in M^+(\Omega, \mathcal{B}_0), \text{ define-se } \forall A \in \mathcal{B} \quad M^e(A) = M(A \cap B_0);$$

nestas condições:

1) $\mathcal{N}$ tight em $(B_0, \mathcal{B}_0) \Rightarrow \mathcal{N}^e = \{M^e : M \in \mathcal{N}\}$ tight em $(\Omega, \mathcal{B})$,

$M_n \xrightarrow{w} M$ em $(B_0, \mathcal{B}_0) \Rightarrow M_n^e \xrightarrow{w} M^e$,

2) $M_n \xrightarrow{w} M$ em $(B, \mathcal{B})$ e $M_n(B_0) \rightarrow M(B_0) \Rightarrow M_n^r \xrightarrow{w} M^r$.

A generalização da demonstração de Billingsley acima mencionada a medidas finitas não oferece maiores dificuldades. $\square$

COROLÁRIO DA PROPOSIÇÃO 3.21.

Nas mesmas condições de 3.21, com a hipótese adicional $\mathcal{N} \subseteq \mathcal{F} = \{M \in \mathcal{M}^+ : M(\Omega) = 1\}$, vale que:

$\mathcal{N}$ é $\mathcal{F}$ -relativamente compacto se e só se $\mathcal{N}$ é tight.

PROPOSIÇÃO 3.22.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+ =$

$\left\{ \begin{array}{l} \text{medidas finitas} \\ \text{definidas em } (\Omega, d) \end{array} \right\}$, $\Pi_d = \Pi$ a métrica de Prohorov em $\mathcal{M}^+$ e $\mathcal{F} =$

$\{M \in \mathcal{M}^+ : M(\Omega) = 1\}$. Então:

$(\mathcal{M}^+, \Pi_d)$ é completo se e só se $(\mathcal{F}, \Pi/\mathcal{F})$ é completo.

Comentário: este resultado é bastante útil, já que em geral na literatura é tratado apenas o caso do espaço $\mathcal{F}$; também é possível deduzir da demonstração que a completitude de $(\mathcal{M}^+, \Pi_d)$ equivale a de $\{M \in \mathcal{M}^+ : M(\Omega) = r\}$, onde r é qualquer real positivo. A demonstração da parte $(\Rightarrow)$ é simples, já que $(\mathcal{F}, \Pi/\mathcal{F})$ é um subespaço fechado do espaço métrico completo $(\mathcal{M}^+, \Pi)$, portanto $(\mathcal{F}, \Pi/\mathcal{F})$ é completo. No sentido $(\Leftarrow)$, prova-se que se $\{M_n : n \in \mathbb{N}\}$ é uma sequência de Cauchy em $(\mathcal{M}^+, \Pi)$, então $\{M_n(\Omega) : n \in \mathbb{N}\}$ é uma sequência de

Cauchy em $\mathbb{R}$ e portanto $\exists \lambda \geq 0$ tal que $M_n(\Omega) \rightarrow \lambda$. Se λ for 0 prova-se que $M_n \rightarrow \theta$ (onde θ é a medida nula). Se $\lambda > 0$, $\exists n_0$ tal que $M_n(\Omega) > 0 \forall n \geq n_0$ e prova-se que a sequência de medidas de probabilidade $\{M_n/M_n(\Omega), n \geq n_0\}$ é de Cauchy e portanto, dada a hipótese da completitude de $\mathcal{F}$, convergente a uma medida de probabilidade P ; finalmente se verá que este fato implica que $M_n \rightarrow \lambda P$. $\square$

3.F. DEMOSTRAÇÕES.

Proposição 3.2.

a) a métrica de Levy L definida em 3.1 é uma distância propriamente dita em $\mathcal{M}^+$, isto é, L satisfaz:

$$L \geq 0,$$

$$L(M, L) = 0 \text{ se e só se } M = L,$$

$$L(M, L) = L(L, M),$$

$$L(M, N) \leq L(M, L) + L(L, N);$$

além disso vale que $L(M, L) \leq \max\{M(\Omega), L(\Omega)\}$, portanto se $M \in \mathcal{F}$, $L \in \mathcal{F}$ $\Rightarrow L(M, L) \leq 1$;

b) L pode ser alternativamente definida por:

$$L(P, Q) = \inf \left\{ \varepsilon > 0 : P(A) \leq Q((A^\varepsilon)^\sim) + \varepsilon \wedge Q(A) \leq P((A^\varepsilon)^\sim) + \varepsilon \forall A \in \mathcal{A} \right\}, \text{ onde}$$

$$\mathcal{A} = \left\{ A \subseteq \mathbb{R} : \exists a \in \mathbb{R} \text{ tal que } A = (-\infty, a] \right\} \text{ e } (A^\varepsilon)^\sim = \{ w \in \Omega : e(w, B) \leq \varepsilon \}.$$

Demonstração:

a) $L(M, N)$ está bem definida, dado que $\varepsilon = \max \{M(\Omega), N(\Omega)\}$

satisfaz $G(t-\varepsilon) - \varepsilon \leq F(t) \leq G(t+\varepsilon) + \varepsilon \quad \forall t \in \mathbb{R}$,

onde F e G são as acumuladas de M e N respectivamente,

o que prova também que $L(M, N) \leq \max \{M(\Omega), N(\Omega)\}$;

é trivial verificar que $L(M, N) \geq 0$;

vejam agora que $L(M, N) = L(N, M)$:

$G(t-\varepsilon) - \varepsilon \leq F(t) \quad \forall t \in \mathbb{R}$ see $G(u) \leq F(u+\varepsilon) + \varepsilon \quad \forall u \in \mathbb{R}$ (tome $u=t-\varepsilon$),

$F(t) \leq G(t+\varepsilon) + \varepsilon \quad \forall t \in \mathbb{R}$ see $F(u-\varepsilon) - \varepsilon \leq G(u) \quad \forall u \in \mathbb{R}$ (tome $u=t+\varepsilon$),

portanto:

$G(t-\varepsilon) - \varepsilon \leq F(t) \leq G(t+\varepsilon) + \varepsilon \quad \forall t \in \mathbb{R} \Leftrightarrow F(t-\varepsilon) - \varepsilon \leq G(t) \leq F(t+\varepsilon) + \varepsilon \quad \forall t \in \mathbb{R}$,

portanto a definição é simétrica;

seja $L(M, N) = 0$, então $\forall \varepsilon > 0$ vale que: $F(t) \leq G(t+\varepsilon) + \varepsilon, \quad \forall t \in \mathbb{R}$ e por simetria $G(t) \leq F(t+\varepsilon) + \varepsilon$,

tomando limites quando $\varepsilon \rightarrow 0$ e dado que F e G são contínuas à direita e não decrescentes:

$$F(t) \leq G(t^+) = \lim_{s \rightarrow t, s > t} G(s) = G(t) \quad \forall t \in \mathbb{R},$$

$$G(t) \leq F(t^+) = \lim_{s \rightarrow t, s > t} F(s) = F(t) \quad \forall t \in \mathbb{R},$$

portanto $F(t) = G(t) \quad \forall t \in \mathbb{R} \Rightarrow M = N$;

para verificar a desigualdade triangular é conveniente utilizar a segunda versão da definição de L , que será provada imediatamente depois:

$$L(M, N) \leq L(M, L) + L(L, N), \text{ pois}$$

$$\forall B \subseteq \Omega \quad \forall \varepsilon > 0 \quad \forall \tau > 0 \text{ vale que } (B^\varepsilon)^\tau \subseteq B^{\varepsilon+\tau}$$

(já que $\forall \omega \in \Omega \quad \forall \eta \in \Omega$ tem-se que $d(\omega, B) \leq d(\omega, \eta) + d(\eta, B)$);

vejamos agora a segunda versão da definição de L :

sejam M e N em M^+ com as respectivas funções de distribuição F e G , então:

$$G(t-\varepsilon)-\varepsilon \leq F(t) \leq G(t+\varepsilon)+\varepsilon \quad \forall t \in \mathbb{R}$$

equivale a (dada a simetria acima provada)

$$G(t-\varepsilon)-\varepsilon \leq F(t) \leq G(t+\varepsilon)+\varepsilon \wedge F(t-\varepsilon)-\varepsilon \leq G(t) \leq F(t+\varepsilon)+\varepsilon \quad \forall t \in \mathbb{R},$$

que por sua vez equivale a

$$F(t) \leq G(t+\varepsilon)+\varepsilon \wedge G(t) \leq F(t+\varepsilon)+\varepsilon \quad \forall t \in \mathbb{R},$$

(já que $F(t) \leq G(t+\varepsilon)+\varepsilon \quad \forall t \in \mathbb{R} \Leftrightarrow F(t-\varepsilon)-\varepsilon \leq G(t) \quad \forall t \in \mathbb{R}$,

e que $G(t) \leq F(t+\varepsilon)+\varepsilon \quad \forall t \in \mathbb{R} \Leftrightarrow G(t-\varepsilon)-\varepsilon \leq F(t) \quad \forall t \in \mathbb{R}$),

portanto L pode ser definida por:

$L(M,N) = \inf \left\{ \varepsilon > 0 : F(t) \leq G(t+\varepsilon) + \varepsilon \wedge G(t) \leq F(t+\varepsilon) + \varepsilon \quad \forall t \in \mathbb{R} \right\}$, ou também por

$$L(M,N) = \inf \left\{ \varepsilon > 0 : M(A) \leq N((A^\varepsilon)^\sim) + \varepsilon \wedge N(A) \leq M((A^\varepsilon)^\sim) + \varepsilon \quad \forall A \in \mathcal{A} \right\}, \text{ onde}$$

$$\mathcal{A} = \left\{ A \subseteq \mathbb{R} : \exists a \in \mathbb{R} \text{ tal que } A = (-\infty, a] \right\} \text{ e } (A^\varepsilon)^\sim = \{ \omega \in \Omega : e(\omega, A) \leq \varepsilon \}. \square$$

Proposição 3.3.

Seja L a métrica de Levy definida em 3.1, então:

a) $L \leq K$, onde K é a distância de Kolmogoroff:

$$K(M,N) = \sup \{ |F(t) - G(t)| : t \in \mathbb{R} \},$$

F e G são as respectivas funções de distribuição de M e N ;

b) se Π_e é a métrica de Prohorov definida em 3.8 e e é a métrica euclidiana usual em $\mathbb{R}$ vale que

$L \leq \Pi_e$, além disso existem P e Q em $\mathcal{M}^+$ tais que:

$$L(P,Q) < \Pi_e(P,Q).$$

Demonstração:

a) sejam F e G as respectivas funções de distribuição de M e N e $\tau = K(M,N)$,

então, dado que F e G são não decrescentes e a definição de K tem-se que:

$$G(t-\tau)-\tau \leq G(t)-\tau \leq F(t) \leq G(t) \leq G(t+\tau)+\tau \quad \forall t \in \mathbb{R}$$

portanto $L(M,N) \leq \tau$;

b) as caracterizações de L e Π_e dadas em 3.2.b) e em 3.9.b), respectivamente, permitem provar em forma direta que $L \leq \Pi_e$, pois $\mathcal{A} \subseteq \mathcal{B}$ e ao substituir $\mathcal{B}$ por $\mathcal{A}$ na definição de Π_e obtém-se a definição de L ;

vejamos agora que existem pares de medidas (P,Q) onde a desigualdade $L \leq \Pi_e$ ocorre em forma estrita:

seja $A = \{0,2\}$ e sejam P e Q duas medidas de probabilidades em $(\mathbb{R}, \mathcal{B})$ definidas por:

$$P\{0\}=\lambda \text{ e } P\{2\}=1-\lambda, \text{ onde } 0<\lambda<1, \lambda \text{ fixo e } Q\{1\}=1,$$

ou pelas respectivas funções de distribuição

$$F(t) = \begin{cases} 0 & \text{se } t < 0 \\ \lambda & \text{se } 0 \leq t < 2 \\ 1 & \text{se } 2 \leq t \end{cases} \quad \text{e} \quad G(t) = \begin{cases} 0 & \text{se } t < 1 \\ 1 & \text{se } t \geq 1 \end{cases},$$

$$\text{é claro que } P\{A\} = 1 \text{ e que } Q(A^\varepsilon) + \varepsilon = \begin{cases} \varepsilon & \text{se } \varepsilon \leq 1 \\ 1 + \varepsilon & \text{se } \varepsilon > 1 \end{cases},$$

portanto, sendo P e Q medidas de probabilidade,

$$1 \geq \Pi_e(M,N) = \inf \left\{ M(B) \leq N(B^\varepsilon) + \varepsilon \quad \forall B \in \mathcal{B} \right\} \geq 1 \Rightarrow \Pi_e(P,Q) = 1,$$

por outro lado, $L(P,Q) \leq K(P,Q) = \max\{\lambda, 1-\lambda\} < 1$, já que $0 < \lambda < 1$. $\square$

Proposição 3.4.

Se $\mathcal{B} = \{\text{borelianos de } (R, e)\}$, a métrica de Levy L induz a topologia fraca $\mathcal{T}$ em $\mathcal{M}^+(R, \mathcal{B})$.

Demonstração:

sendo (R, e) é um espaço métrico separável, os resultados da Seção 3.D implicam que Π_d induz a topologia $\mathcal{T}$ em $\mathcal{M}^+(R, \mathcal{B})$, por outro lado provou-se na proposição anterior que $L \leq \Pi_e$, que implica que a topologia induzida por L é menos fina que a induzida por Π_e ,

portanto esta proposição estará demonstrada se provarmos que $\mathcal{T}$ é menos fina que a topologia induzida pela métrica de Levy em $\mathcal{M}^+(R, \mathcal{B})$;

seja então $M \in \mathcal{M}^+(R, \mathcal{B})$, consideremos a vizinhança de M na topologia $\mathcal{T}$ definida por:

$$W(M, f, \varepsilon) = \left\{ N \in \mathcal{M}^+(R, \mathcal{B}) : \left| \int f dN - \int f dM \right| < \varepsilon \right\},$$

onde $\varepsilon > 0$ e $f: (R, e) \xrightarrow[\text{limitada}]{\text{contínua e}} (R, e)$ (veja 2.21),

sejam k, τ e λ números reais estritamente positivos a serem fixados posteriormente,

dado que f é contínua e que o intervalo $[-k, k]$ é compacto, f é uniformemente contínua em $[-k, k]$, então dado $\tau > 0$ existe $\delta = \delta(f, k, \tau)$ tal que:

$$s \in [-k, k], t \in [-k, k], |s - t| < \delta \Rightarrow |f(s) - f(t)| < \tau,$$

consideremos $Q \in \{N \in M^+(R, \mathcal{B}) : L(N, M) < \lambda\}$, então se F e G são as funções de distribuição de M e Q , respectivamente, tem-se:

$$F(t-\lambda) - \lambda \leq G(t) \leq F(t+\lambda) + \lambda \quad \forall t \in R$$

então

$$\int f(t) d(F(t-\lambda) - \lambda) \leq \int f(t) d(G(t)) \leq \int f(t) d(F(t+\lambda) + \lambda),$$

fazendo na integral de Stieljes da esquerda as mudanças de variável $u=t-\lambda$ e $t=u$, obtém-se:

$$\int f(t) d(F(t-\lambda) - \lambda) = \int f(t+\lambda) dF(t)$$

e fazendo na integral da direita as mudanças $u=t+\lambda$ e $t=u$, tem-se:

$$\int f(t) d(F(t+\lambda) + \lambda) = \int f(t-\lambda) dF(t),$$

então:

$$|\int f(t+\lambda) dF(t) - \int f(t-\lambda) dF(t)| \leq \int |f(t+\lambda) - f(t-\lambda)| dF(t),$$

se $\lambda < \delta/2$, tem-se que:

$$\int |f(t+\lambda) - f(t-\lambda)| dF(t) \leq$$

$$\tau \cdot M([-k, k]) + 2 \cdot \sup\{f(t) : t \in R\} \cdot M([-k, k]^c)$$

dado que M é tight, o número real $k=k(f, M, \epsilon)$ pode ser escolhido de tal forma que

$$2 \cdot \sup\{f(t) : t \in R\} \cdot M([-k, k]^c) < \epsilon/4,$$

a seguir, escolha-se $\tau = \tau(M, k)$ tal que

$$\tau \cdot M([-k, k]) < \varepsilon/4,$$

obtendo-se que:

$$|\int f(t+\lambda) dF(t) - \int f(t-\lambda) dF(t)| < \varepsilon/2,$$

que implica

$$|\int f(t+\lambda) dF(t) - \int f(t) dG(t)| < \varepsilon/2 \quad (*)$$

$$(\text{pois } \int f(t+\lambda) dF(t) \leq \int f(t) d(G(t)) \leq \int f(t-\lambda) dF(t)),$$

(observe-se que uma vez escolhidos k e τ como acima pode-se escolher

o $\delta = \delta(f, k, \tau)$ tal que

$$s \in [-k, k], \quad t \in [-k, k], \quad |s-t| < \delta \Rightarrow |f(s) - f(t)| < \tau,$$

e por fim definir $\lambda = \delta/2$);

por outro lado,

$$\begin{aligned} |\int f(t+\lambda) dF(t) - \int f(t) dF(t)| &\leq \int |f(t+\lambda) - f(t)| dF(t) \stackrel{\text{se } \lambda < \delta}{\leq} \\ &\leq \tau \cdot M([-k, k]) + 2 \cdot \sup\{f(t) : t \in \mathbb{R}\} \cdot M([-k, k]^c), \end{aligned}$$

que por sua vez ficará menor que $\varepsilon/2$ se k e τ forem os acima escolhidos e neste caso então

$$|\int f(t+\lambda) dF(t) - \int f(t) dF(t)| < \varepsilon/2 \quad (**);$$

finalmente, (*) e (**) implicam que

$$|\int f(t)dF(t) - \int f(t)dG(t)| < \epsilon, \text{ ou seja } Q \in W(M, f, \epsilon);$$

resumindo, provamos que dada a vizinhança $W(M, f, \epsilon)$ da medida M na topologia fraca, existe uma L -esfera de raio $\lambda = \lambda(W)$ tal que

$$\{N \in \mathcal{M}^+(R, \mathcal{B}) : L(N, M) < \lambda\} \subseteq W(M, f, \epsilon),$$

dado que as $W(M, f, \epsilon)$ formam uma subbase do sistema de vizinhanças de M da topologia $\mathcal{T}$, e que em qualquer espaço métrico a família das esferas com centro num mesmo ponto é base do sistema de vizinhanças do ponto, está demonstrado que $\mathcal{T}$ é menos fina que a topologia induzida por L em $\mathcal{M}^+(R, \mathcal{B})$. $\square$

Proposição 3.6.

Consideremos a métrica de Levy L definida em $\mathcal{M}^+(R, \mathcal{B})$, onde $\mathcal{B} = \{\text{borelianos de } (R, e)\}$. Seja $\{M_n : n \in \mathbb{N}\} \subseteq \mathcal{M}^+(R, \mathcal{B})$.

Se $\{M_n : n \in \mathbb{N}\}$ for uma seqüência de Cauchy a respeito de L , então

a) $\{M_n : n \in \mathbb{N}\}$ é tight e b) $\{M_n(\Omega) : n \in \mathbb{N}\}$ é Cauchy.

Demonstração:

a) denotaremos por F_n a função de distribuição de M_n , $\forall n \in \mathbb{N}$;

seja $\epsilon > 0$, ϵ arbitrário e fixo, sejam λ , τ e m números reais estritamente positivos a serem determinados posteriormente;

dado $\lambda \exists n_0 \in \mathbb{N}$ tal que $\forall n \geq n_0 \quad L(M_n, M_{n_0}) < \lambda$;

espaço métrico completo é tight, então F_{n_0} é, portanto, dado $\tau > 0 \exists K_\tau$ compacto tal que $M_{n_0}(K_\tau^c) < \tau$,

mas em R , todo compacto é limitado, então existe um intervalo $(a, b]$ tal que $K_\tau \subseteq (a, b]$, portanto vale que:

dado $\tau > 0 \exists (a, b]$ intervalo tal que $M_{n_0}(a, b] > M_{n_0}(\Omega) - \tau$;

por definição de L , se

$$m > \lambda \text{ e } n \geq n_0 \quad F_n(b+m) \geq F_{n_0}(b) - m \text{ e } F_n(a-m) \leq F_{n_0}(a) + m,$$

portanto, se $m > \lambda$ e $n \geq n_0$:

$$\begin{aligned} M_n[a-m, b+m] &\geq M_n(a-m, b+m) = F_n(b+m) - F_n(a-m) \geq \\ &\geq F_{n_0}(b) - m - (F_{n_0}(a) + m) = F_{n_0}(b) - (F_{n_0}(a) - 2m) = M_{n_0}(a, b] - 2m; \end{aligned}$$

por outro lado, dado que $\forall M \in \mathcal{M}^+(R, \mathcal{B}) \forall N \in \mathcal{M}^+(R, \mathcal{B})$ vale que $|M(\Omega) - N(\Omega)| \leq L(M, N)$, tem-se que

$$L(M_n, M_{n_0}) < \lambda \Rightarrow |M_n(\Omega) - M_{n_0}(\Omega)| < \lambda,$$

$$\text{portanto: } n \geq n_0 \Rightarrow |M_n(\Omega) - M_{n_0}(\Omega)| < \lambda \Rightarrow M_{n_0}(\Omega) > M_n(\Omega) - \lambda;$$

resumindo, se $m > \lambda$ e $n \geq n_0$:

$$M_n[a-m, b+m] \geq M_{n_0}(a, b] - 2m, \text{ e}$$

$$|M_n(\Omega) - M_{n_0}(\Omega)| < \lambda,$$

portanto, se $m > \lambda$ e $n \geq n_0$: $M_n[a-m, b+m] \geq M_{n_0}(a, b) - 2m \geq M_{n_0}(\Omega) - \tau - 2m > M_n(\Omega) - \lambda - \tau - 2m$,

a afirmação acima implica que, se $m > \lambda$ e $n \geq n_0$:

$$M_n[a-m, b+m] > M_n(\Omega) - \lambda - \tau - 2m,$$

portanto, dado que $m > \lambda$, escolhendo $\max\{m, \tau\} < \varepsilon/4$, tem-se que:

$$n \geq n_0 \Rightarrow M_n[a-m, b+m] > M_n(\Omega) - \varepsilon,$$

ou seja provou-se que:

dado $\varepsilon > 0 \exists K_\varepsilon = [a-m, b+m]$ compacto tal que $\forall n \geq n_0 M_n(K_\varepsilon)^c < \varepsilon$,

dado que as medidas $\{M_n, n < n_0\}$ são todas tight, dado $\varepsilon > 0$, é possível achar um intervalo $[c, d]$ de tal forma que $M_n([c, d])^c < \varepsilon \forall n < n_0$,

portanto, tomando $u = \min\{a-m, c\}$ e $v = \max\{a+m, d\}$, vale que:

$$\forall n \in \mathbb{N} \quad M_n([u, v])^c < \varepsilon,$$

resumindo, dado $\varepsilon > 0$, é possível achar um compacto $K_\varepsilon = [u, v]$, de tal forma que $\forall n \in \mathbb{N} \quad M_n([u, v])^c < \varepsilon$,

portanto a sequência de Cauchy $\{M_n : n \in \mathbb{N}\}$ é tight. $\square$

b) decorre de $|M(\Omega) - N(\Omega)| \leq L(M, N)$

Proposição 3.7.

Sejam $\mathcal{B} = \{\text{borelianos de } (R, e)\}$ e L métrica de Levy em $\mathcal{M}^+(R, \mathcal{B})$.

Então o espaço métrico $(\mathcal{M}^+(R, \mathcal{B}), L)$ é completo.

Demonstração: por definição de completitude, devemos provar que toda sequência de Cauchy é convergente;

seja $\{M_n, n \in \mathbb{N}\}$ uma sequência de Cauchy em $(M^+(R, \mathcal{B}), L)$ e $\{F_n, n \in \mathbb{N}\}$ a família das respectivas funções de distribuição,

se $\{M_n, n \in \mathbb{N}\}$ for sequência de Cauchy em $(M^+(R, \mathcal{B}), L)$, então $\{M_n(\Omega), n \in \mathbb{N}\}$ será sequência de Cauchy em $\mathbb{R}$ (pois $|M(\Omega) - N(\Omega)| \leq L(M, N)$), portanto $\{M_n(\Omega), n \in \mathbb{N}\}$ é limitada,

dado que $M_n(\Omega) = F_n(\infty) = \lim_{t \rightarrow \infty} F_n(t) = \sup\{F_n(t) : t \in \mathbb{R}\}$, que $\{M_n(\Omega), n \in \mathbb{N}\}$ é limitada e que as F_n são não decrescentes, tem-se que:

$$\exists \lambda > 0 \text{ tal que } \sup_n \sup\{F_n(t) : t \in \mathbb{R}\} \leq \lambda,$$

$$\text{e como também } 0 \leq F_n(t) \quad \forall t \in \mathbb{R} \quad \forall n \in \mathbb{N},$$

portanto o teorema de Helly garante a existência de uma F e de uma subsequência $\{F_m, m \in \mathbb{N}\}$ de $\{F_n, n \in \mathbb{N}\}$ tais que:

- $0 \leq F(t) \leq \lambda \quad \forall t \in \mathbb{R}$, F não decrescente e contínua à direita, e
- $\lim_n F_n(t) = F(t) \quad \forall t$ ponto de continuidade de F ,

seja M a função de conjunto definida nos intervalos $(a, b]$ de $\mathbb{R}$ por $M(a, b] = F(b) - F(a)$ e devidamente extendida aos borelianos de $\mathbb{R}$ (veja Proposição 2.19), provaremos que:

$$\text{a) } M \in M^+(R, \mathcal{B}), \text{ e}$$

b) $M_m \xrightarrow{W} M$, sendo que esta afirmação implica que $M_n \xrightarrow{W} M$, dado que $(M^+(R, \mathcal{B}), L)$ por ser métrico é um espaço de Hausdorff e portanto o limite é único;

$\{M_n(\Omega), n \in \mathbb{N}\}$ sequência L -Cauchy $\Rightarrow \{M_m(\Omega), m \in \mathbb{N}\}$ sequência de Cauchy em R , então, pela Proposição 3.6 acima, $\{M_m, m \in \mathbb{N}\}$ é tight, portanto,

dado $\varepsilon > 0 \exists K_\varepsilon$ compacto tal que $M_m(K_\varepsilon) > M_m(\Omega) - \varepsilon \forall m \in \mathbb{N}$,

também $\exists (a, b]$ intervalo tal que

a e b são pontos de continuidade de F e $K_\varepsilon \subseteq (a, b]$,

portanto

$M_m(a, b] > M_m(\Omega) - \varepsilon \forall m \in \mathbb{N}$,

em consequência, dado que $((a, b])^c = (-\infty, a] \cup (b, \infty)$,

$F_m(a) = M_m(-\infty, a] < \varepsilon \forall m \in \mathbb{N}$ e $M_m(\Omega) - F_m(b) = M_m(b, \infty) < \varepsilon \forall m \in \mathbb{N}$,

ou seja que:

dado $\varepsilon > 0$ existem a e b pontos de continuidade de F (*)

tais que $\forall m \in \mathbb{N}: F_m(a) < \varepsilon$ e $F_m(b) > M_m(\Omega) - \varepsilon$;

vejamos que M é uma medida:

dado (*) tem-se que $F(a) = \lim_m F_m(a) \leq \varepsilon$, portanto,

dado $\varepsilon > 0 \exists a \in R$ tal que $F(a) < \varepsilon$,

então, sendo $F(t) \geq 0 \forall t \in R$,

$$F(-\infty) = \lim_{t \rightarrow -\infty} F(t) = 0,$$

sendo M σ -aditiva na semi-álgebra dos intervalos $(a, b]$, sua extensão aos borelianos também é, a condição $F(-\infty)=0$ garante que M é uma medida (veja Proposição 2.19);

vejamos agora que $M_m \xrightarrow{w} M$:

dado que sabemos que

- $F_n(t) \rightarrow F(t) \quad \forall t$ ponto de continuidade de F ,

- $M \in \mathcal{M}^+(R, \mathcal{B})$, onde M é a função de conjunto M associada a F ,

para ver que $M_m \xrightarrow{w} M$ basta verificar (Proposição 2.20) que $F_m(\infty) \rightarrow F(\infty)$, ou equivalentemente que $M_m(\Omega) \rightarrow M(\Omega)$,

dado que $\{M_m : m \in \mathbb{N}\}$ é uma seqüência de Cauchy, $\{M_m(\Omega) : m \in \mathbb{N}\}$ também é (já que $|M(\Omega) - N(\Omega)| \leq L(M, N)$), portanto $\{M_m(\Omega) : m \in \mathbb{N}\}$ é convergente, queremos ver que o limite é $M(\Omega)$:

$\forall t$ ponto de continuidade de F vale que

$$F(t) = \lim_m F_m(t) \leq \lim_m M_m(\Omega)$$

mesmo sendo s uma descontinuidade de F , $F(s) \leq \lim_m M_m(\Omega)$, já que sempre $\exists t$ ponto de continuidade de F tal que $s < t$, neste caso seria

$$F(s) \leq F(t) \leq \lim_m M_m(\Omega)$$

portanto provou-se que:

$$F(t) \leq \lim_{m \rightarrow \infty} M_m(\Omega) \quad \forall t \in \mathbb{R},$$

em consequência

$$F(\infty) = M(\Omega) \leq \lim_{m \rightarrow \infty} M_m(\Omega);$$

por outra parte, por (*),

dado $\varepsilon > 0$ $\exists b$ ponto de continuidade de F tal que

$$\forall m \in \mathbb{N} \quad F_m(b) > M_m(\Omega) - \varepsilon;$$

portanto, provou-se que:

$$\text{dado } \varepsilon > 0 \quad \exists b \in \mathbb{R} \text{ tal que } F(b) = \lim_{m \rightarrow \infty} F_m(b) \geq \lim_{m \rightarrow \infty} M_m(\Omega) - \varepsilon,$$

então,

$$M(\Omega) = F(\infty) = \lim_{b \rightarrow \infty} F(b) \geq \lim_{m \rightarrow \infty} M_m(\Omega). \square$$

Proposição 3.9.

Sejam (Ω, d) um espaço métrico, $\mathcal{B} = \{ \text{borelianos de } (\Omega, d) \}$,

$$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\} \quad \text{e} \quad \mathcal{F} = \left\{ F \in \mathcal{M}^+ : F(\Omega) = 1 \right\}. \text{ Então:}$$

a) a métrica de Prohorov é uma distância propriamente dita, isto é $\Pi = \Pi_d$ satisfaz:

$$\Pi \geq 0,$$

$$\Pi(M, L) = 0 \text{ se e só se } M = L,$$

$$\Pi(M, L) = \Pi(L, M),$$

$$\Pi(M, N) \leq \Pi(M, L) + \Pi(L, N);$$

além disso vale que $\Pi(M, L) \leq \text{máximo}\{M(\Omega), L(\Omega)\}$, portanto se $M \in \mathcal{F}$, $L \in \mathcal{F}$
 $\Rightarrow \Pi(M, L) \leq 1$;

b) a métrica de Prohorov Π_d pode ser alternativamente definida por:

$$\Pi_d(M, L) = \inf \left\{ \varepsilon > 0 : M(B) \leq L((B^\varepsilon)^\sim) + \varepsilon \wedge L(B) \leq M((B^\varepsilon)^\sim) + \varepsilon \quad \forall B \in \mathcal{B} \right\},$$

onde $(B^\varepsilon)^\sim = \{\omega \in \Omega : d(\omega, B) \leq \varepsilon\}$;

c) se $M \in \mathcal{M}$, $N \in \mathcal{M}$, $M(\Omega) = L(\Omega)$ e $\varepsilon > 0$, tem-se que:

$$\forall B \in \mathcal{B} \quad M(B) \leq L(B^\varepsilon) + \varepsilon \Rightarrow \forall B \in \mathcal{B} \quad L(B) \leq M(B^\varepsilon) + \varepsilon,$$

portanto, na definição 3.8. da métrica de Prohorov o primeiro conjunto de desigualdades implica o segundo (e vice-versa, dada a simetria da situação), que então pode ser dispensado;

d) na definição da métrica de Prohorov " $\forall B \in \mathcal{B}$ " pode ser substituído por " $\forall B$ fechado".

Demonstração:

a) a definição é boa, já que todas as desigualdades são simultaneamente satisfeitas se $\varepsilon \geq \text{máximo}(M(\Omega), L(\Omega))$, o que por outro lado mostra que $\Pi \leq \text{máximo}(M(\Omega), L(\Omega))$, também $\Pi \geq 0$ por definição;

$$\Pi(M, M) = 0, \text{ dado que } \forall \varepsilon > 0 \quad \forall B \in \mathcal{B} \quad M(B) \leq M(B^\varepsilon) + \varepsilon,$$

reciprocamente se $\Pi(M,L)=0$, as desigualdades na definição de Π são satisfeitas $\forall \varepsilon > 0$ e portanto $\forall B$ boreliano vale que:

$$\lim_{\varepsilon \rightarrow 0} M(B) \leq \lim_{\varepsilon \rightarrow 0} L(B^\varepsilon) + \varepsilon \wedge \lim_{\varepsilon \rightarrow 0} L(B) \leq \lim_{\varepsilon \rightarrow 0} M(B^\varepsilon) + \varepsilon,$$

se B for fechado, sendo (Ω, d) métrico, $B = \bigcap_{n \in \mathbb{N}} (B^{1/n})$

então:

$$\lim_{\varepsilon \rightarrow 0} B^\varepsilon = \bigcap_{\varepsilon > 0} B^\varepsilon = B,$$

portanto:

$\forall B$ fechado vale $M(B) \leq L(B)$ e $L(B) \leq M(B)$,

em consequência

$\forall B$ fechado $M(B) = L(B)$,

ou seja, M e L são duas medidas finitas que coincidem sobre os conjuntos fechados, então (por Proposição 2.1.(e)):

$$\forall B \in \mathcal{B} \quad M(B) = L(B) ;$$

$\Pi(M, L) = \Pi(L, M)$ dada a simetria da definição de Π ;

$\Pi(M, N) \leq \Pi(M, L) + \Pi(L, N)$, pois

$$\forall B \subseteq \Omega \quad \forall \varepsilon > 0 \quad \forall \tau > 0 \text{ vale que } (B^\varepsilon)^\tau \subseteq B^{\varepsilon+\tau}$$

(já que $\forall \omega \in \Omega \quad \forall \eta \in \Omega$ tem-se que $d(\omega, B) \leq d(\omega, \eta) + d(\eta, B)$);

b) seja $\Pi^\sim = \Pi_d^\sim$ definida por

$$\Pi_{\sim}(M, L) = \inf \left\{ \varepsilon > 0 : M(B) \leq L((B^{\varepsilon})^{\sim}) + \varepsilon \wedge L(B) \leq M((B^{\varepsilon})^{\sim}) + \varepsilon \quad \forall B \in \mathcal{B} \right\},$$

$$\text{onde } (B^{\varepsilon})^{\sim} = \{\omega \in \Omega : d(\omega, B) \leq \varepsilon\};$$

e seja Π a métrica de Prohorov da definição 3.8, isto é

$$\Pi(M, L) = \inf \left\{ \varepsilon > 0 : M(B) \leq L(B^{\varepsilon}) + \varepsilon \wedge L(B) \leq M(B^{\varepsilon}) + \varepsilon \quad \forall B \in \mathcal{B} \right\},$$

$$\text{onde } B^{\varepsilon} = \{\omega \in \Omega : d(\omega, B) < \varepsilon\};$$

$$\Pi_{\sim} \leq \Pi, \text{ já que } (B^{\varepsilon})^{\sim} \supseteq B^{\varepsilon},$$

para provar que $\Pi_{\sim} \geq \Pi$ é suficiente verificar que:

$$\Pi_{\sim}(P, Q) < \lambda \text{ e } \varepsilon > \lambda \Rightarrow \Pi(P, Q) < \varepsilon;$$

$$\Pi_{\sim}(P, Q) < \lambda \Rightarrow \forall B \in \mathcal{B} \text{ vale } P(B) \leq Q(B^{\lambda})^{\sim} + \lambda \text{ e } Q(B) \leq P(B^{\lambda})^{\sim} + \lambda,$$

$$\varepsilon > \lambda \Rightarrow B^{\varepsilon} \supseteq (B^{\lambda})^{\sim}$$

portanto se

$$\Pi_{\sim}(P, Q) < \lambda \text{ e } \varepsilon > \lambda \Rightarrow \forall B \in \mathcal{B} \text{ vale } P(B) \leq Q(B^{\varepsilon}) + \varepsilon \text{ e } Q(B) \leq P(B^{\varepsilon}) + \varepsilon, \text{ então } \Pi(P, Q) < \varepsilon;$$

c) a seguinte demonstração é uma adaptação da encontrada em Huber (1981, pág.27):

$$\text{sejam } M(\Omega) = L(\Omega), \quad \varepsilon > 0 \text{ e } \forall B \in \mathcal{B}: M(B) \leq L(B^{\varepsilon}) + \varepsilon,$$

em particular $(B^{\varepsilon})^c$ satisfaz as desigualdades acima, então:

$$M((B^{\varepsilon})^c) \leq L(((B^{\varepsilon})^c)^{\varepsilon}) + \varepsilon,$$

equivalentemente

$$M(\Omega) - M(B^\varepsilon) \leq L(\Omega) - L(((B^\varepsilon)^c)^\varepsilon)^c + \varepsilon.$$

$$\text{e como } M(\Omega) = L(\Omega) \text{ e } B \subseteq \bar{B} = (((B^\varepsilon)^c)^\varepsilon)^c$$

$$L(B) \leq L(((B^\varepsilon)^c)^\varepsilon)^c \leq M(B^\varepsilon) + \varepsilon,$$

então:

$$L(B) \leq M(B^\varepsilon) + \varepsilon;$$

$$\text{d) se } \alpha = \inf \left\{ \varepsilon > 0 : M(B) \leq L(B^\varepsilon) + \varepsilon \wedge L(B) \leq M(B^\varepsilon) + \varepsilon \forall B \text{ fechado} \right\}$$

então $\alpha' \leq \Pi(M, L)$, por definição de Π ;

falta provar que $\Pi(M, L) \leq \alpha$, ou equivalentemente que

$$\forall h > 0 \forall B \in \mathcal{B} \text{ vale: } M(B) \leq L(B^{\alpha+h}) + \alpha + h \wedge L(B) \leq M(B^{\alpha+h}) + \alpha + h;$$

seja então $h > 0$ e $B \in \mathcal{B}$,

$$\text{dado que } \begin{cases} M(B) = \sup \{M(F) : F \text{ fechado}, F \subseteq B\} \\ L(B) = \sup \{L(F) : F \text{ fechado}, F \subseteq B\} \end{cases} \text{ (por A.1.1.e),}$$

existe uma seqüência crescente de fechados $\{F_n\}_{n \in \mathbb{N}}$ tal que:

$$M(B) = \lim_n M(F_n) \text{ e } L(B) = \lim_n L(F_n),$$

por definição de α e dado que os F_n são fechados:

$$\forall n \in \mathbb{N}: M(F_n) \leq L(F_n)^{\alpha+h} + \alpha + h \text{ e } L(F_n) \leq M(F_n)^{\alpha+h} + \alpha + h$$

e dado que $F_n \subseteq B$ e que $M(B) = \lim_n M(F_n)$ e $L(B) = \lim_n L(F_n)$,

tem-se que
$$\begin{cases} M(B) \leq \lim_n L(F_n^{\alpha+h}) + \alpha+h \leq L(B^{\alpha+h}) + \alpha+h \\ L(B) \leq \lim_n M(F_n^{\alpha+h}) + \alpha+h \leq M(B^{\alpha+h}) + \alpha+h \end{cases}$$

o que termina a demonstração. $\square$

Proposição 3.10.

Sejam (Ω, d) um espaço métrico, $\mathcal{B} = \{ \text{borelianos de } (\Omega, d) \}$,

$\mathcal{M} = \left\{ \begin{array}{l} \text{medidas finitas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$ e $\mathcal{F} = \{ F \in \mathcal{M}^+ : F(\Omega) = 1 \}$. Então:

(a) $\forall \lambda \geq 0 \forall (P, Q) \in \mathcal{M}^+ \times \mathcal{M}^+ : \Pi(\lambda P, \lambda Q) \leq (\lambda \vee 1) \Pi(P, Q),$

(b) $\forall 0 \leq \lambda \leq 1 \forall P \in \mathcal{M}^+$ tal que $P(\Omega) \leq 1$ vale que $\Pi(\lambda P, P) \leq 1 - \lambda,$

(c) $\forall (P, Q, F, G) \in (\mathcal{M}^+)^4 : \Pi(P+Q, F+G) \leq 2\{\Pi(P, F) \vee \Pi(Q, G)\};$

(d) as aplicações: $(\lambda, P) \mapsto \lambda P$ e $(P, Q) \mapsto P+Q,$

de $\mathbb{R}^+ \times \mathcal{M}^+ \mapsto \mathcal{M}^+$ e de $\mathcal{M}^+ \times \mathcal{M}^+ \mapsto \mathcal{M}^+$, são contínuas, onde as topologias em $\mathbb{R}^+$ e $\mathcal{M}^+$ são a usual e a induzida por Π , respectivamente.

Demonstração.

(a) é obviamente satisfeita se $\lambda = 0$, se $\lambda \neq 0$ vale que:

$$\Pi(\lambda P, \lambda Q) = \inf\{\varepsilon : P(B) \leq Q(B^\varepsilon) + \varepsilon/\lambda \wedge Q(B) \leq P(B^\varepsilon) + \varepsilon/\lambda \forall B \in \mathcal{B}\},$$

então se $\lambda < 1$

$$\Pi(\lambda P, \lambda Q) \leq \inf\{\varepsilon : P(B) \leq Q(B^\varepsilon) + \varepsilon \wedge Q(B) \leq P(B^\varepsilon) + \varepsilon \forall B \in \mathcal{B}\} = \Pi(P, Q)$$

e, se $\lambda \geq 1$

$$\Pi(\lambda P, \lambda Q) \leq \inf\{\varepsilon : P(B) \leq Q(B^{\varepsilon/\lambda}) + \varepsilon/\lambda \wedge Q(B) \leq P(B^{\varepsilon/\lambda}) + \varepsilon/\lambda \quad \forall B \in \mathcal{B}\}$$

$$= \lambda \Pi(P, Q);$$

$$(b) \quad \Pi(\lambda P, P) = \max \left\{ \begin{array}{l} \inf\{\varepsilon : \lambda P(B) \leq P(B^{\varepsilon}) + \varepsilon \quad \forall B \in \mathcal{B}\} \\ \inf\{\varepsilon : P(B) \leq \lambda P(B^{\varepsilon}) + \varepsilon \quad \forall B \in \mathcal{B}\} \end{array} \right\}$$

então se $0 \leq \lambda \leq 1$ e $P(\Omega) \leq 1$

$$\Pi(\lambda P, P) \leq 0 \vee \inf\{\varepsilon : P(B) \leq \lambda P(B^{\varepsilon}) + \varepsilon \quad \forall B \in \mathcal{B}\} \leq$$

$$\leq \inf\{\varepsilon : (1-\lambda)P(B) \leq \varepsilon \quad \forall B \in \mathcal{B}\} \leq 1-\lambda;$$

$$(c) \quad \Pi(P+Q, V+W) = \max \left\{ \begin{array}{l} \inf\{\varepsilon : (P+Q)(B) \leq (V+W)(B^{\varepsilon}) + \varepsilon \quad \forall B \in \mathcal{B}\} \\ \inf\{\varepsilon : (V+W)(B) \leq (P+Q)(B^{\varepsilon}) + \varepsilon \quad \forall B \in \mathcal{B}\} \end{array} \right\}$$

$\forall \varepsilon > 0$ e $\forall B \in \mathcal{B}$ vale que

$$P(B) \leq V(B^{\varepsilon/2}) + \varepsilon/2 \text{ e } Q(B) \leq W(B^{\varepsilon/2}) + \varepsilon/2 \Rightarrow (P+Q)(B) \leq (V+W)(B^{\varepsilon}) + \varepsilon$$

e também que

$$V(B) \leq P(B^{\varepsilon/2}) + \varepsilon/2 \text{ e } W(B) \leq Q(B^{\varepsilon/2}) + \varepsilon/2 \Rightarrow (V+W)(B) \leq (P+Q)(B^{\varepsilon}) + \varepsilon$$

portanto, tem-se que

$$\Pi(P+Q, V+W) \leq 2\{\Pi(P, V) \vee \Pi(Q, W)\};$$

(d) A continuidade de $+$ é consequência imediata de (c); a continuidade de $*$ será demonstrada em 3 etapas:

$$(i) \quad \forall \lambda \geq 0 \quad \forall (P, Q) \in \mathcal{M} \times \mathcal{M}: \Pi(P, Q) \longrightarrow 0 \Rightarrow \Pi(\lambda P, \lambda Q) \longrightarrow 0,$$

(ii) $\forall \alpha \geq 0 \quad \forall P \in \mathcal{M}: \lambda \longrightarrow \alpha \Rightarrow \Pi(\lambda P, \alpha P) \longrightarrow 0;$

(iii) * é contínua, ou seja

$\forall \alpha \geq 0 \quad \forall (P, Q) \in \mathcal{M} \times \mathcal{M}: \lambda \longrightarrow \alpha \text{ e } \Pi(P, Q) \longrightarrow 0 \Rightarrow \Pi(\lambda P, \alpha Q) \longrightarrow 0.$

Observação: em (ii) e (iii), se $\alpha = 0$, considere-se $\lambda \longrightarrow 0^+$.

(i): é consequência imediata de (a);

(ii): a partir de (b) conclui-se que:

$P(\Omega) \leq 1$ e $\lambda \longrightarrow 1^-$ então $\Pi(\lambda P, P) \longrightarrow 0;$ (I)

aplicando sucessivamente (a) e (b) tem-se:

se $\lambda > 1$: $\Pi(\lambda P, P) = \Pi(\lambda P, \lambda \lambda^{-1} P) \leq \lambda \Pi(P, \lambda^{-1} P)$

e se $P(\Omega) \leq 1$: $\Pi(P, \lambda^{-1} P) \leq 1 - \lambda^{-1},$

portanto

se $P(\Omega) \leq 1$ e $\lambda \longrightarrow 1^+$ então $\Pi(\lambda P, P) \longrightarrow 0$ (II)

conseqüentemente, dados (I) e (II),

$P(\Omega) \leq 1$ e $\lambda \longrightarrow 1 \Rightarrow \Pi(\lambda P, P) \longrightarrow 0;$ (III)

seja agora $P(\Omega) > 1$, define-se $Q(B) = P(B)/P(\Omega) \quad \forall B \in \mathcal{B},$

então

$\Pi(\lambda P, P) = \Pi(\lambda P(\Omega)Q, P(\Omega)Q) \leq P(\Omega)\Pi(\lambda Q, Q)$ (por (a)),

sendo $Q(\Omega) = 1$, (III) garante que se $\lambda \longrightarrow 1 \Rightarrow \Pi(\lambda Q, Q) \longrightarrow 0,$

obtendo-se

$$P(\Omega) > 1 \text{ e } \lambda \longrightarrow 1 \Rightarrow \Pi(\lambda P, P) \longrightarrow 0; \quad (\text{IV})$$

(III) e (IV) podem ser reunidos em

$$\text{se } \lambda \longrightarrow 1 \text{ então } \Pi(\lambda P, P) \longrightarrow 0;$$

finalmente seja $\alpha > 0$:

então $\lambda \longrightarrow \alpha \Rightarrow \lambda \cdot \alpha^{-1} \longrightarrow 1$ e sendo $\Pi(\lambda P, \alpha P) = \Pi(\lambda \alpha^{-1} \alpha P, \alpha P)$, obtem-se que:

$$\lambda > 0 \text{ e } \lambda \longrightarrow \alpha \text{ então } \Pi(\lambda P, \alpha P) \longrightarrow 0,$$

ou seja, que falta apenas provar (ii) para o caso $\alpha = 0$, que é trivial, já que, se θ for a medida nula, $\Pi(\lambda P, \theta) = \lambda P(\Omega)$.

(iii): pela desigualdade triangular

$$\Pi(\lambda P, \alpha Q) \leq \Pi(\lambda P, \lambda Q) + \Pi(\lambda Q, \alpha Q),$$

portanto, por (i) e (ii),

$$\text{se } \lambda \longrightarrow \alpha \text{ e } \Pi(P, Q) \longrightarrow 0 \text{ então } \Pi(\lambda P, \alpha Q) \longrightarrow 0,$$

ou seja $*: (\lambda, P) \longrightarrow \lambda P$ é continua. $\square$

Proposição 3.11.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$, $\mathcal{T}$ a topologia fraca e Π_d a métrica de

Prohorov em $\mathcal{M}^+$. Então:

1) $\mathcal{T}$ é sempre menos fina que Π_d ;

2) são equivalentes as seguintes afirmações:

a) $M \in \mathcal{M}^+$ e existe base enumerável de $\mathcal{T}$ -vizinhanças de M ;

b) $M \in \mathcal{M}^+$ e M tem suporte separável;

c) $\left\{ \mathcal{T}\text{-vizinhanças de } M \right\} = \left\{ \Pi_d\text{-vizinhanças de } M \right\}$.

Demonstração:

1) em primeiro lugar, veremos que a topologia fraca é menos fina que a definida pela métrica de Prohorov;

sejam $M \in \mathcal{M}^+$, F fechado em (Ω, d) e $\varepsilon > 0$,

dado que $F = \bigcap_{n \in \mathbb{N}} \left\{ \omega : d(\omega, F) < n^{-1} \right\}$, existe τ tal que:

$0 < \tau < \varepsilon$ e $M(F^\tau) < M(F) + \varepsilon/2$,

então vale que:

$$\left\{Q \in \mathcal{M}^+ : \Pi_d(Q, M) < \tau/2\right\} \subseteq \left\{Q \in \mathcal{M}^+ : \begin{array}{l} Q(F) < M(F^\varepsilon) + \varepsilon \\ |Q(\Omega) - M(\Omega)| < \varepsilon \end{array}\right\} = V(M, F, \varepsilon),$$

já que:

$$\Pi_d(Q, M) < \tau/2 \Rightarrow |Q(\Omega) - M(\Omega)| \leq \Pi_d(Q, M) < \tau/2 < \varepsilon, \quad e$$

$$\Pi_d(Q, M) < \tau/2 \Rightarrow Q(F) \leq M(F^{\tau/2}) + \tau/2 \leq M(F^\tau) + \tau/2 < M(F) + \varepsilon;$$

$$(\text{sendo } V(M) = \left\{V(M, F_1, F_2, \dots, F_k, \varepsilon); \begin{array}{l} F_j \text{ fechados, } 1 \leq j \leq k \\ k \in \mathbb{N}, \varepsilon > 0 \end{array}\right\}, \quad \text{onde}$$

$$V(M, F_1, F_2, \dots, F_k, \varepsilon) = \left\{Q \in \mathcal{M}^+ : \begin{array}{l} Q(F_j) < M(F_j) + \varepsilon, \quad 1 \leq j \leq k \\ |Q(\Omega) - M(\Omega)| < \varepsilon \end{array}\right\} \text{ é base de vizinhanças}$$

de M na topologia fraca $\mathcal{T}$ (veja Proposição 2.23), é suficiente considerar vizinhanças de M da forma $V(M, F, \varepsilon)$, isto é, uma sub-base da topologia, dado que

$$V(M, F_1, F_2, \dots, F_k, \varepsilon) = \bigcap_{j=1}^k V(M, F_j, \varepsilon), \quad e$$

$$\bigcap_{j=1}^k \left\{Q \in \mathcal{M}^+ : \Pi_d(Q, M) < \varepsilon_j\right\} = \left\{Q \in \mathcal{M}^+ : \Pi_d(Q, M) < \min\{\varepsilon_j : 1 \leq j \leq k\}\right\};$$

portanto fica provado que $\mathcal{T}$ é menos fina que a topologia induzida por Π_d em $\mathcal{M}^+$;

2) a) $\Rightarrow$ b): seja $M \in \mathcal{M}^+$ tal que existe base enumerável de vizinhanças de M ,

em todo espaço topológico vale que: existindo uma base enumerável de vizinhanças do ponto M , toda outra base de vizinhanças de M contém uma subfamília enumerável que também é base (se $\{B_n : n \in \mathbb{N}\}$ for a base

enumerável, existe uma B'_n vizinhança de M na outra base tal que $B'_n \subseteq B_n$, então $\{B'_n : n \in \mathbb{N}\}$ é também base), portanto:

sejam $\{B_n : n \in \mathbb{N}\}$ qualquer base enumerável de vizinhanças de M e $\{B'_n : n \in \mathbb{N}\} \subseteq \mathcal{V}(M)$, base de vizinhanças de M , então $\{V_n : n \in \mathbb{N}\}$, onde $V_n = \bigcap_{j=1}^n B'_j$, também é base;

na Proposição 2.25 provou-se que as medidas com suporte finito são densas em $\mathcal{M}^+$, portanto:

$\forall n \in \mathbb{N}$ existe M_n com suporte finito S_n tal que $M_n \in V_n$,

mas, dado que $\forall n \in \mathbb{N} \quad V_n \subseteq V_{n+1}$,

$\forall k \geq n: M_k \in V_n$ e em conseqüência $M_n \xrightarrow{w} M$;

falta ver que M tem suporte separável, é suficiente ver que S^- é um suporte de M , onde $S = \bigcup_{n \in \mathbb{N}} S_n$, pois sendo S enumerável S^- será separável:

seja $V = V(M, S^-, \epsilon)$, onde $\epsilon > 0$,

dado que $M_n \xrightarrow{w} M$, $\exists n_0$ tal que $n \geq n_0 \Rightarrow M_n \in V$,

portanto,

$\forall n \geq n_0: M_n(\Omega) = M_n(S^-) < M(S^-) + \epsilon$,

por outro lado, dado que $M_n \xrightarrow{w} M$,

$M_n(\Omega) \rightarrow M(\Omega)$,

então: $M(\Omega) \leq M(S^-) + \epsilon$,

uma vez que a afirmação acima vale $\forall \varepsilon > 0$, tem-se $M(\Omega) \leq M(S^-)$ e portanto que $M(\Omega) = M(S^-)$;

b) $\Rightarrow$ c): dado que em 1) provou-se que sempre $\mathcal{T}$ é menos fina que Π_d , é suficiente provar que, sob a hipótese de M ter suporte separável, Π_d é menos fina que $\mathcal{T}$, ou seja que:

basta provar que:

$$\forall \varepsilon > 0 \exists Y \mathcal{T}\text{-vizinhança de } M \text{ tal que } Y \subseteq \left\{ Q \in \mathcal{M}^+ : \Pi_d(Q, M) < \varepsilon \right\};$$

sejam $D = \{\omega_n : n \in \mathbb{N}\}$ denso em $S = \text{suporte}(M)$ e $a > 0$ a fixar posteriormente, dado que $D^- \subseteq S$ é claro que

$$S \subseteq \bigcup_{n \in \mathbb{N}} \left\{ Q \in \mathcal{M}^+ : d(\omega, \omega_n) < a \right\},$$

e também que se $a < b$, b a fixar posteriormente,

$$S \subseteq \bigcup_{n \in \mathbb{N}} \left\{ \omega \in \Omega : d(\omega, \omega_n) < a \right\} \subseteq \bigcup_{n \in \mathbb{N}} \left\{ \omega \in \Omega : d(\omega, \omega_n) < b \right\} \quad (*);$$

por outro lado, dado que

$$\left\{ \omega' \in \Omega : a < d(\omega', \omega) < b \right\} = \bigcup_{a < t < b} \left\{ \omega' \in \Omega : d(\omega', \omega) = t \right\}$$

e que, sendo M finita, existe no máximo uma quantidade enumerável de conjuntos disjuntos de medida estritamente positiva, é verdadeira a seguinte afirmação:

se $a < b$, $\forall \omega \in \Omega \exists s$ tal que: $a < s < b$ e $M\{\omega' : d(\omega, \omega') = s\} = 0$,

portanto, $\forall n \in \mathbb{N} \quad \exists t_n$ tal que $a < t_n < b$ e $M\{\omega' \in \Omega : d(\omega', \omega_n) = t_n\} = 0$,

se $B_n = \{\omega \in \Omega : d(\omega, \omega_n) < t_n\}$, sendo que

$\delta(B_n) \subseteq \{\omega \in \Omega : d(\omega, \omega_n) = t_n\}$ e dado (*), tem-se:

$$S \subseteq \bigcup_{n \in \mathbb{N}} B_n \quad \text{e} \quad \forall n \in \mathbb{N}: M(\delta(B_n)) = 0;$$

sejam $A_1 = B_1$, $A_n = B_n - \bigcup_{j=1}^{n-1} B_j$ se $n \geq 2$, então,

$$\delta(A_1) = \delta(B_1), \quad \forall n \geq 2: \delta(A_n) \subseteq \bigcup_{j=1}^n \delta(B_j) \quad \text{e} \quad \forall n \in \mathbb{N}: \text{diâmetro}(A_n) < b,$$

portanto:

$$S \subseteq \bigcup_{n \in \mathbb{N}} A_n \quad \text{e} \quad \forall n \in \mathbb{N}: M(\delta(A_n)) = 0 \quad \text{e} \quad \text{diâmetro}(A_n) < b,$$

dado que $M(\Omega) = M(S)$, $\exists k \in \mathbb{N}$ tal que $M(\bigcup_{j=1}^k A_j) > M(\Omega) - b$,

seja $C = (\bigcup_{j=1}^k A_j)^c$ (é claro que $M(\delta(C)) = 0$),

seja $Y = Y(M, A_1, A_2, \dots, A_k, C, b/k)$ definido por:

$$Y = \left\{ Q \in \mathcal{M}^+ : |Q(C) - M(C)| < b/k \quad \text{e} \quad |Q(A_j) - M(A_j)| < b/k, \quad 1 \leq j \leq k \right\}.$$

Y é uma vizinhança de M na topologia fraca $\mathcal{T}$ (veja Proposição 2.23),

veremos que, para b conveniente,

$$Y \subseteq \left\{ Q \in \mathcal{M}^+ : \Pi_d(Q, M) < \varepsilon \right\};$$

seja $\mathcal{A}$ = uniões finitas $\{A_1, A_2, \dots, A_k\}$,

dado $B \in \mathcal{B}$, sejam $I = \{j: 1 \leq j \leq k \text{ e } A_j \cap B \neq \emptyset\}$ e $A = \bigcup_{j \in I}^k A_j$ ($A \in \mathcal{A}$),

então, $B \subseteq A \cup \left(\bigcup_{j=1}^k A_j\right)^c$ e $A \subseteq B^b$ (pois $\text{diâmetro}(A_j) < b$, $1 \leq j \leq k$), portanto

se $Q \in Y$:

$$M(B) < M(A) + b < Q(A) + b + b \leq Q(B^b) + 2b < Q(B^{3b}) + 3b,$$

também

$$Q(B) \leq Q(A) + Q\left(\left(\bigcup_{j=1}^k A_j\right)^c\right) < M(A) + b + M\left(\left(\bigcup_{j=1}^k A_j\right)^c\right) + b/k < M(A) + b + b + b/k \leq M(B^b) + 3b \leq M(B^{3b}) + 3b,$$

portanto, $\forall B$ boreliano: $M(B) \leq M(B^{3b}) + 3b$ e $Q(B) \leq M(B^{3b}) + 3b$,

ou seja, provamos que:

$Y \subseteq \left\{Q \in \mathcal{M}^+ : \Pi_d(Q, M) < 3b\right\}$, portanto é suficiente tomar $a < b$ para

convalidar a construção anterior e $3b < \epsilon$ para obter $Y \subseteq \left\{Q \in \mathcal{M}^+ : \Pi_d(Q, M) < \epsilon\right\}$;

c) $\Rightarrow$ a): por hipótese,

$$c) \left\{\mathcal{T}\text{-vizinhanças de } M\right\} = \left\{\Pi_d\text{-vizinhanças de } M\right\}.$$

então é claro que: $\{B_n : n \in \mathbb{N}\}$, onde $B_n = \left\{Q \in \mathcal{M}^+ : \Pi_d(Q, M) < 1/n\right\}$,

é base enumerável de vizinhanças de M . $\square$

Corolário da Proposição 3.11.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$, $\mathcal{T}$ a topologia fraca e Π_d a métrica de

Prohorov em $\mathcal{M}^+$. Então:

$\mathcal{T}$ é metrizável se e só se a métrica de Prohorov Π_d metriza $\mathcal{T}$.

Demonstração:

seja $\mathcal{T}$ é metrizável, necessariamente existe $\forall M \in \mathcal{M}^+$ base enumerável de $\mathcal{W}(M)$, como, por exemplo, $[\{Q: D(Q, M) < n^{-1}\}]_{n \in \mathbb{N}}$, sendo D uma métrica que metriza $\mathcal{T}$; mas (pela Proposição 3.4.) a existência de base local enumerável ocorre se e só se $\mathcal{T}$ e Π_d se equivalem localmente e isto acontecendo $\forall M \in \mathcal{M}^+$ significa que os espaços topológicos $(\mathcal{M}^+, \mathcal{T})$ e $(\mathcal{M}^+, \Pi_d)$ são iguais;

no outro sentido a demonstração resultado é trivial. $\square$

Proposição 3.12.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas não negativas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$, $\mathcal{T}$ a topologia fraca e Π_d a métrica

de Prohorov em $\mathcal{M}^+$. Nestas condições são equivalentes as seguintes afirmações:

a) o espaço métrico (Ω, d) é separável;

b) $(\mathcal{M}^+, \mathcal{T}) = (\mathcal{M}^+, \Pi_d)$ e $\mathcal{T}$ é separável;

c) as topologias induzidas em $\mathcal{F} = \{M \in \mathcal{M}^+ : M(\Omega) = 1\}$ por $\mathcal{T}$ e Π_d coincidem e além disso $\mathcal{F}$ é $\mathcal{T}$ e Π_d -separável;

Demonstração:

provaremos : $a \Rightarrow b \Rightarrow c \Rightarrow b \Rightarrow a$);

$a) \Rightarrow b)$:

(Ω, d) separável , então toda medida tem suporte separável (o próprio Ω), portanto, pela proposição 3.11 vale que:

$$\forall M \in \mathcal{M}^+ : \left\{ \mathcal{T}\text{-vizinhanças de } M \right\} = \left\{ \Pi_d\text{-vizinhanças de } M \right\},$$

portanto $(\mathcal{M}^+, \mathcal{T}) = (\mathcal{M}^+, \Pi_d)$;

na demonstração da Proposição 2.25 b), provou-se que se D for denso enumerável em (Ω, d) :

$$\left\{ Q \in \mathcal{M}^+ : \begin{array}{l} \text{suporte}(Q) \text{ é finito} \\ \text{suporte}(Q) \subseteq D \\ \text{massas de cada átomo racionais} \end{array} \right\}$$

é um conjunto enumerável e denso em $(\mathcal{M}^+, \mathcal{T})$, portanto $(\mathcal{M}^+, \mathcal{T})$ é separável;

$b \Rightarrow c)$:

dado que $\mathcal{T}$ e Π_d coincidem em $\mathcal{M}^+$, eles necessariamente coincidem em $\mathcal{F}$; já a separabilidade de $(\mathcal{F}, \mathcal{T}/\mathcal{F})$ decorre dos seguinte fatos:

dado que $(\mathcal{M}^+, \Pi_d)$ é métrico e $\mathcal{F} \subseteq \mathcal{M}^+$ tem-se que: $(\mathcal{F}, \Pi_d/\mathcal{F})$ é métrico,

falta ver que $(\mathcal{F}, \Pi_d)$ é separável: neste sentido é interessante salientar que a propriedade de um espaço topológico ser separável não é hereditária, mas a propriedade de ter base de abertos enumerável é hereditária e em espaços métricos equivale à separabilidade (em geral ela implica a separabilidade), por estes motivos, sendo $(\mathcal{F}, \Pi_d)$ subespaço de (M^+, Π_d) (métrico e separável), ele é separável;

c) $\Rightarrow$ b):

no caso particular da topologia fraca vale o seguinte argumento:

seja $(\mathcal{F}, \mathcal{T}/\mathcal{F})$ separável e seja $\mathcal{D}$ denso enumerável em $(\mathcal{F}, \mathcal{T}/\mathcal{F})$, então:

$$Q \cdot \mathcal{D} = \left\{ q \cdot M : q \in Q \wedge M \in \mathcal{D} \right\}, \quad \text{onde } Q = \{\text{racionais}\},$$

é um conjunto enumerável e denso de (M^+, Π_d) , sendo que este argumento independe da metrizabilidade de $\mathcal{F}$ ou M^+ ;

seja $M \in M^+$, $M \neq 0$, consideremos uma vizinhança qualquer da base $\mathcal{V}(M)$ de $\mathcal{T}$ -vizinhanças de M , ou seja:

$$\mathcal{V}(M, F_1, F_2, \dots, F_k, \varepsilon) = \left\{ N \in M^+ : |N(\Omega) - M(\Omega)| < \varepsilon, N(F_j) < M(F_j) + \varepsilon, 1 \leq j \leq k \right\}$$

onde $\varepsilon > 0$ e F_j fechado, $1 \leq j \leq k$,

provaremos que $\exists M_0 \in Q \cdot \mathcal{D} \cap \mathcal{V}(M, F_1, F_2, \dots, F_k, \varepsilon)$:

se $M(\Omega) \in Q$, consideremos a vizinhança V' de $P = M/M(\Omega)$ na topologia relativa de $\mathcal{F}$ induzida por $\mathcal{T}$ definida por:

$$V' = \left\{ Q \in \mathcal{F} : Q(F_j) < P(F_j) + \varepsilon / M(\Omega), \quad 1 \leq j \leq k \right\},$$

dado que $\mathcal{D}$ é denso em $\mathcal{F}$, $\exists Q_0 \in \mathcal{D} \cap V'$, agora é simples verificar que:

$$M_0 = M(\Omega) \cdot Q_0 \in Q \cdot \mathcal{D} \cap V(M, F_1, F_2, \dots, F_k, \varepsilon) \quad (\text{veja que } M_0(\Omega) = M(\Omega));$$

se $M(\Omega) \notin Q$ sejam V' e Q_0 como acima, seja agora $q_0 \in Q$ tal que

$$M(\Omega) - \varepsilon < q_0 < M(\Omega),$$

é fácil verificar que $q_0 \cdot Q_0 \in Q \cdot \mathcal{D} \cap V(M, F_1, F_2, \dots, F_k, \varepsilon)$;

se $M=0$, seja Q_0 qualquer elemento de $\mathcal{F} \cap \mathcal{D}$, é muito simples verificar que: $\forall n \in \mathbb{N} \quad Q_n = n^{-1} \cdot Q_0 \in Q \cdot \mathcal{D}$ e que $Q_n \xrightarrow{w} 0$;

b) $\Rightarrow$ a):

$(\mathcal{F}, \Pi_d)$ é métrico separável see $(\mathcal{F}, \Pi_d)$ é métrico com base de abertos enumerável $\Rightarrow \Omega^* = \{1_\omega : \omega \in \Omega\}$ com a métrica induzida é métrico com base de abertos enumerável see (Ω^*, Π_d) é métrico separável,

por outro lado $\Pi_d(1_\omega, 1_\tau) = \min \{1, d(\omega, \tau)\} \quad \forall (\omega, \tau) \in \Omega \times \Omega$, o que implica que Ω^* é homeomorfo a Ω , então se $\mathcal{D}^*$ é enumerável denso em Ω^* a sua imagem homeomorfa será enumerável e denso em Ω , que será portanto separável. $\square$

Lema 3.14.

Sejam (Ω, d) um espaço métrico, $\mathcal{B} = \{ \text{borelianos de } (\Omega, d) \}$,

$$\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas finitas não} \\ \text{negativas em } (\Omega, \mathcal{B}) \end{array} \right\} \text{ e } \mathcal{N} \subseteq \mathcal{M}^+;$$

N é tight see $\left\{ \begin{array}{l} \forall \varepsilon > 0 \quad \forall \delta > 0 \quad \exists F \subseteq \Omega, F = F(\varepsilon, \delta) \text{ finito,} \\ \text{tal que } \forall M \in N \quad M \left(\left(\bigcup_{\omega \in F} B(\omega, \delta) \right)^c \right) < \varepsilon \end{array} \right.$

onde $B(\omega, \delta) = \{ \omega' \in \Omega : d(\omega', \omega) < \delta \}$.

Demonstração:

($\Rightarrow$) seja N tight, então $\forall \varepsilon > 0 \quad \exists K = K(\varepsilon)$ compacto tal que $M(K^c) < \varepsilon$,

$\forall \delta > 0$, é claro que $K \subseteq \bigcup_{\omega \in \Omega} B(\omega, \delta)$, então a compacidade de K implica que

$\exists F \subseteq \Omega, F = F(\delta)$ finito, tal que $K \subseteq \bigcup_{\omega \in F} B(\omega, \delta)$, sendo claro que

$$M \left(\left(\bigcup_{\omega \in F} B(\omega, 4\delta) \right)^c \right) \leq M(K^c) < \varepsilon;$$

observe-se que ($\Rightarrow$) é válido em todo espaço métrico (Ω, d) ;

($\Leftarrow$) seja $\varepsilon > 0$, pela hipótese, $\forall n \in \mathbb{N}$ existe F_n finito tal que $\forall M \in N$ vale: $M(B(n))^c < \varepsilon/2^n$, onde $B(n) = \bigcup_{\omega \in F_n} B(\omega, n^{-1})$,

consideremos o conjunto $B = \bigcap_{n \in \mathbb{N}} B(n)$, e claro que:

$\forall M \in N \quad M(B^c) < \varepsilon$ e que B é totalmente limitado, sendo (Ω, d) completo B^- é compacto e $\forall M \in N \quad M((B^-)^c) < \varepsilon$. $\square$

Lema 3.15.

(Ω, d) espaço métrico separável e completo, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $M^+(\Omega, d) = \left\{ \begin{array}{l} \text{medidas finitas não} \\ \text{negativas definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$, $\Pi = \Pi_d$ a métrica de Prohorov em M^+ . Nestas condições:

a) $\{M_n, n \in \mathbb{N}\}$ sequência de Cauchy em $\mathcal{M}^+(\Omega, d) \Rightarrow \{M_n, n \in \mathbb{N}\}$ é tight.

b) $\{M_n(\Omega) : n \in \mathbb{N}\}$ é sequência de Cauchy.

Demonstração:

Q) sejam $\varepsilon > 0$ e $\delta > 0$, provaremos que existe $F = F(\varepsilon, \delta)$ finito, $F \subseteq \Omega$, tal que:

$$\forall n \in \mathbb{N} \quad M_n \left(\left(\bigcup_{\omega \in F} B(\omega, \delta/2) \right)^c \right) < 2\varepsilon;$$

seja $\lambda = \min(\varepsilon/3, \delta/2)$ e seja n_0 tal que $n \geq n_0 \Rightarrow \Pi(M_{n_0}, M_n) < \lambda$, pela Proposição 2.5 do capítulo anterior, dado que (Ω, d) é separável e completo, M_{n_0} é tight, portanto:

$$\exists F \text{ finito, } F \subseteq \Omega, \text{ tal que } M_{n_0}(B^c) < \varepsilon/3, \text{ onde } B = \bigcup_{\omega \in F} B(\omega, \delta/3);$$

$$\text{portanto: } M_{n_0}(\Omega) = M_{n_0}(B) + M_{n_0}(B^c) < M_{n_0}(B) + \varepsilon/3,$$

2e, dado que $\Pi(M_{n_0}, M_n) < \lambda$ se $n \geq n_0$ e que $\lambda \leq \varepsilon/3$:

$$\forall n \geq n_0: M_{n_0}(\Omega) < M_{n_0}(B) + \varepsilon/3 \leq M_n(B^\lambda) + \lambda + \varepsilon/3 \leq M_n(B^\lambda) + 2\varepsilon/3,$$

também, $\lambda \leq \delta/2 \Rightarrow B^\lambda \subseteq B^{\delta/2} \subseteq \bigcup_{\omega \in F} B(\omega, \delta)$, então:

$$\forall n \in \mathbb{N}: M_n(B^\lambda) + 2\varepsilon/3 \leq M_n \left(\bigcup_{\omega \in F} B(\omega, \delta) \right) + 2\varepsilon/3,$$

$$\text{conseqüentemente, } \forall n \geq n_0: M_n \left(\bigcup_{\omega \in F} B(\omega, \delta) \right) > M_{n_0}(\Omega) - 2\varepsilon/3 \quad (*);$$

por outro lado, dado que $\forall (M, N) \in \mathcal{M}^+ \times \mathcal{M}^+ : |M(\Omega) - N(\Omega)| \leq \Pi_d(M, N)$,

$$\forall n \geq n_0: |M_n(\Omega) - M_{n_0}(\Omega)| \leq \Pi_d(M_n, M_{n_0}) < \lambda \leq \varepsilon/3 \quad (**);$$

então, dados (*) e (**), tem-se que:

$$\forall n \geq n_0: \quad M_n \left(\bigcup_{\omega \in F} B(\omega, \delta) \right) > M_n(\Omega) - \varepsilon;$$

dado que o conjunto $\{M_n, n < n_0\}$ é finito e que todos seus elementos são medidas tight:

$$\forall n < n_0 \exists F_n \text{ finito tal que } M_n \left(\bigcup_{\omega \in F_n} B(\omega, \delta) \right) > M_n(\Omega) - \varepsilon,$$

finalmente, se F_0 for definido por:

$$F_0 = F \cup \bigcup_{n < n_0} F_n,$$

então F_0 satisfará:

$$F_0 \text{ é finito e } \forall n \in \mathbb{N}: M_n \left(\bigcup_{\omega \in F_0} B(\omega, \delta) \right) > M_n(\Omega) - \varepsilon,$$

o resultado a) está demonstrado;

b) decorre de $|M(\Omega) - N(\Omega)| \leq \Pi(M, N)$. $\square$

Lema 3.16.

(Ω, d) espaço métrico $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+(\Omega, d) = \left\{ \begin{array}{l} \text{medidas finitas não} \\ \text{negativas definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$, $\Pi = \Pi_d$ a métrica de Prohorov em $\mathcal{M}^+(\Omega, d)$, $f: (\Omega, d) \rightarrow (R, e)$ contínua. Então:

a) K compacto e $\varepsilon > 0$, então: $\exists \delta = \delta(\varepsilon)$ tal que

$$x \in K^\delta, y \in K^\delta, d(x, y) < \delta \Rightarrow e(f(x), f(y)) < \varepsilon;$$

b) $R=\{\text{reais}\}$, é a métrica euclideana em R , L a métrica de Levy em $\mathcal{M}^+(R, e)$, $\{M_n, n \in \mathbb{N}\}$ seqüência de Cauchy em $(\mathcal{M}^+(\Omega, d), \Pi_d)$

então: $\{M_n f^{-1}, n \in \mathbb{N}\}$ seqüência de Cauchy em $(\mathcal{M}^+(R, e), L)$

Demonstração:

a) a demonstração será feita em dois passos:

i) mostraremos que existem V aberto em (Ω, d) e $\tau > 0$ tais que: $K \subseteq V$ e se $x \in V$, $y \in V$ e $d(x, y) < \tau \Rightarrow e(f(x), f(y)) < \varepsilon$;

ii) dados K compacto, V aberto e $K \subseteq V$ existe α tal que $0 < \alpha$ e $K^\alpha \subseteq V$, ou seja que os conjuntos da forma K^α são base de vizinhanças de qualquer compacto K de um espaço métrico;

de i) e ii) concluir-se-a que $\delta = \alpha \wedge \tau$ satisfaz:

se $x \in K^\delta$, $y \in K^\delta$ e $d(x, y) < \delta \Rightarrow e(f(x), f(y)) < \varepsilon$.

Dem de i): pela continuidade de f , $\forall \omega \in \Omega$ existe $\delta_\omega = \delta(\omega, \varepsilon)$ tal que se $d(x, \omega) < \delta_\omega \Rightarrow e(f(x), f(\omega)) < \varepsilon/4$;

é trivial verificar que $K \subseteq \bigcup_{\omega \in K} B(\omega, \delta_\omega)$, seja agora $\sigma > 0$ um número de Lebesgue de $\{K, \{B(\omega, \delta_\omega) : \omega \in K\}\}$, isto é:

$\forall \omega \in K$ existe $\omega' \in K$ tal que $B(\omega, \sigma) \subseteq B(\omega', \delta_{\omega'})$ (Kelley, 1962, p.178);

mas também vale que $K \subseteq \bigcup_{\omega \in K} B(\omega, (\sigma \wedge \delta_\omega)/3) = V$,

e, sendo K compacto, existe $\{\omega_1, \omega_2, \dots, \omega_n\} \subseteq K$ tal que:

$$K \subseteq \bigcup_{j=1}^n B(\omega_j, (\sigma \wedge \delta_{\omega_j})/3) = V,$$

seja $\tau = \min\{(\sigma \wedge \delta_{\omega_j})/3 : 1 \leq j \leq n\}$, será provado que:

$$\text{se } x \in V, y \in V \text{ e } d(x, y) < \tau \Rightarrow e(f(x), f(y)) < \varepsilon;$$

com efeito, sejam $x \in V, y \in V$ e $d(x, y) < \tau$, então existem ω_j e ω_k tais que: $d(x, \omega_j) < (\sigma \wedge \delta_{\omega_j})/3$ e $d(y, \omega_k) < (\sigma \wedge \delta_{\omega_k})/3$, que, juntamente com $d(x, y) < \tau$, implicam que $d(\omega_j, \omega_k) < \sigma$, e portanto, por definição de σ , que:

$$\exists \omega_0 \in K \text{ tal que } \omega_j \in B(\omega_0, \delta_{\omega_0}) \text{ e } \omega_k \in B(\omega_0, \delta_{\omega_0}), \text{ então}$$

$$e(f(\omega_j), f(\omega_k)) < 2\varepsilon/4 \quad (*)$$

mas também, $d(x, \omega_j) < (\sigma \wedge \delta_{\omega_j})/3$ e $d(y, \omega_k) < (\sigma \wedge \delta_{\omega_k})/3$ implicam que:

$$e(f(x), f(\omega_j)) < \varepsilon/4 \text{ e que } e(f(y), f(\omega_k)) < \varepsilon/4 \quad (**)$$

(*) e (**) implicam que $e(f(x), f(y)) < \varepsilon$, portanto, está provado que:

existem V aberto e $\tau > 0$ tais que:

$$K \subseteq V \text{ e se } x \in V, y \in V \text{ e } d(x, y) < \tau \Rightarrow e(f(x), f(y)) < \varepsilon;$$

Dem de ii): dados K compacto, V aberto e $K \subseteq V$, tem-se que:

K compacto, V^c fechado e $K \cap V^c = \emptyset$, o que implica que: $d(K, V^c) = r > 0$, portanto, se $\alpha < r$ vale que $K^\alpha \subseteq V$, o que termina a demonstração de ii).

b) dado que $\{M_n, n \in \mathbb{N}\}$ seqüência de Cauchy em $(M^+(\Omega, d), \Pi_d)$, por Lema 3.15,

$$\text{dado } \varepsilon > 0 \exists K_\varepsilon \text{ compacto tal que } M_n(K_\varepsilon) > M_n(K) - \varepsilon/2 \quad \forall n \in \mathbb{N},$$

seja δ tal que se

$$\omega \in (K_\varepsilon)^\delta, \omega' \in (K_\varepsilon)^\delta, d(\omega, \omega') < \delta \Rightarrow d(f(\omega), f(\omega')) < \varepsilon,$$

então $\exists n_0 \in \mathbb{N}$ tal que

$$\forall n \geq n_0, \forall m \geq n_0 \quad \supremo(\Pi_d(M_n, M_m)) < \lambda = \text{mínimo}\{\delta, \varepsilon/2\},$$

seja $\alpha > 0$,

$$\begin{aligned} M_n f^{-1}(-\infty, \alpha] &\leq M_n((f^{-1}(-\infty, \alpha]) \cap K_\varepsilon) + \varepsilon/2 \leq \\ &\leq M_m(((f^{-1}(-\infty, \alpha]) \cap K_\varepsilon)^\lambda) + \lambda + \varepsilon/2 \leq M_m f^{-1}(-\infty, \alpha + \varepsilon] + \varepsilon, \end{aligned}$$

portanto se $n \geq n_0$ e $m \geq n_0$ vale que $L(M_n f^{-1}, M_m f^{-1}) \leq \varepsilon$. $\square$

Lema 3.17.

(Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+(\Omega, \mathcal{B}) = \{\text{medidas finitas em } (\Omega, \mathcal{B})\}$, Π_d a métrica de Prohorov associada a d , $M \in \mathcal{M}^+(\Omega, \mathcal{B})$.

Seja $\{M_n : n \in \mathbb{N}\}$ uma sequência tight em $\mathcal{M}^+(\Omega, \mathcal{B})$

tal que $\forall f \in CL(\Omega, d) \exists J(f) = \lim_n \int f dM_n$ e $|J(f)| < \infty$,

Nestas condições:

a) $J: CL(\Omega, d) \rightarrow \mathbb{R}$, satisfaz

J é linear e $f \geq 0 \Rightarrow J(f) \geq 0$;

b) existe uma única função de conjunto M definida na álgebra $\mathcal{B}^*$ gerada pelos abertos de (Ω, d) , finitamente aditiva e regular tal que:

$J(f) = \int f dM$, onde M regular significa que:

$$\forall B \in \mathcal{B}^* \quad M(B) = \sup\{M(F) : F \text{ fechado}, F \subseteq B\} = \inf\{M(G) : G \text{ aberto}, G \supseteq B\},$$

é claro que M é finita, já que $M(\Omega) = \int 1 dM$;

c) $\forall \epsilon > 0$ existe $K(\epsilon)$ compacto tal que $M(K(\epsilon)) \geq M(\Omega) - \epsilon$; d) $\forall B \in \mathcal{B}^*$
 $M(B) = \sup\{M(K) : K \text{ compacto}, K \subseteq B\}$;

e) M é σ -aditiva em $\mathcal{B}^*$;

f) M tem única extensão a uma medida em $(\Omega, \mathcal{B})$.

Demonstração:

a) consideremos $J: CL(\Omega, d) \rightarrow \mathbb{R}$ definida acima, dado que *limite* é linear é fácil verificar que:

J é linear e que $f \geq 0 \Rightarrow J(f) \geq 0$;

b) seja $J: CL(\Omega, d) \rightarrow \mathbb{R}$, J linear e tal que $f \geq 0 \Rightarrow J(f) \geq 0$, nestas condições (veja Parthasaraty, 1967, p. 35) existe uma única função de conjunto M definida na álgebra $\mathcal{B}^*$ gerada pelos abertos de (Ω, d) , finitamente aditiva e regular tal que:

$$J(f) = \int f dM$$

(onde M regular significa que

$$\forall B \in \mathcal{B}^* \quad M(B) = \sup\{M(F) : F \text{ fechado}, F \subseteq B\} = \inf\{M(G) : G \text{ aberto}, G \supseteq B\},$$

é claro que M é finita, já que $M(\Omega) = \int 1 dM$;

c) dado que $\{M_n : n \in \mathbb{N}\}$ é 3.15, é tight, então dado $\varepsilon > 0$ existe $K(\varepsilon)$ compacto tal que

$$\forall n \in \mathbb{N} \quad M_n(K(\varepsilon)^c) < \varepsilon \quad (*),$$

pela regularidade de M ,

$$M(K(\varepsilon)) = \inf\{M(G) : G \text{ aberto}, G \supseteq K(\varepsilon)\},$$

dado $\delta > 0$, a igualdade acima e a finitude de M implicam que:

$$\exists G_0 = G(\delta) \text{ aberto tal que } K(\varepsilon) \subseteq G_0 \text{ e } M(G_0) < M(K(\varepsilon)) + \delta,$$

seja agora $g: \Omega \rightarrow \mathbb{R}$ definida por

$$g(\omega) = d(\omega, K(\varepsilon)) / (d(\omega, K(\varepsilon)) + d(\omega, (G_0)^c)),$$

dado que: $d(\omega, K(\varepsilon)) + d(\omega, (G_0)^c)$ é não nulo $\forall \omega \in \Omega$ e que

se B boreliano, $h(\omega) = d(\omega, B)$ é contínua em $(\Omega, \mathcal{B})$,

tem-se que g é contínua;

também g é limitada, mais precisamente $0 \leq 1_{K(\varepsilon)} \leq g \leq 1_{G_0} \leq 1$,

pela hipótese, sendo g contínua e limitada, tem-se:

$$\int g dM_n \rightarrow \int g dM, \text{ também: } M_n(K(\varepsilon)) = \int g dM_n \leq \int g dM \text{ e}$$

$$\int g dM = \int g dM \leq M(G_0) < M(K(\varepsilon)) + \delta,$$

as três últimas afirmações implicam que

$$\limsup_n M_n(K(\varepsilon)) \leq \limsup_n \int g dM_n = \lim_n \int g dM_n = \int g dM < M(K(\varepsilon)) + \delta,$$

e dado que δ é arbitrário

$$\limsup_n M_n(K(\varepsilon)) \leq M(K(\varepsilon))$$

também, $\forall n \in \mathbb{N} \quad M_n(K(\varepsilon)) > M_n(\Omega) - \varepsilon$ (por (*)),

então

$$M(K(\varepsilon)) \geq \limsup_n M_n(K(\varepsilon)) \geq \limsup_n M_n(\Omega) - \varepsilon =$$

$$\limsup_n \int 1 dM_n - \varepsilon = \lim_n \int 1 dM_n - \varepsilon = \int 1 dM - \varepsilon = M(\Omega) - \varepsilon,$$

provamos então que

$\forall \varepsilon > 0$ existe $K(\varepsilon)$ compacto tal que $M(K(\varepsilon)) \geq M(\Omega) - \varepsilon$;

d) sejam $\{K_n : n \in \mathbb{N}\}$ compactos tais que $M(K_n) \geq M(\Omega) - 1/n$,

seja F fechado e consideremos $F_n = F \cap K_n$, $n \in \mathbb{N}$,

dado que (Ω, d) é Hausdorff, K_n é fechado, então os F_n são compactos (por serem fechados contidos em compactos),

dado que $\forall n \in \mathbb{N} \quad M(K_n) \geq M(\Omega) - 1/n$, resulta

$$\forall n \in \mathbb{N} \quad M(F_n) \geq M(F) - 1/n,$$

que juntamente com $\forall n \in \mathbb{N} \quad F_n \subseteq F$,

implicam que $M(F) \leq \sup_n M(F_n) \leq \sup\{M(K) : K \text{ compacto}, K \subseteq F\}$

e dado que, M é monótona (por ser aditiva),

$$M(F) = \sup\{M(K) : K \text{ compacto}, K \subseteq F\};$$

devemos ainda provar que a propriedade acima, provada para qualquer fechado F , vale em geral para todo boreliano B , mas isto é consequência da propriedade acima e da regularidade de M , mais precisamente:

$$M(F) = \sup\{M(K) : K \text{ compacto}, K \subseteq F\} \quad \forall F \text{ fechado e}$$

$$M(B) = \sup\{M(F) : F \text{ fechado}, F \subseteq B\} \quad \forall B \text{ boreliano}$$

implicam que $M(B) = \sup\{M(K) : K \text{ compacto}, K \subseteq B\} \quad \forall B \text{ boreliano};$

e) vejamos que M é σ -aditiva em $\mathcal{B}^*$;

sendo M aditiva em $\mathcal{B}^*$, a σ -aditividade equivale à seguinte propriedade:

$$"\forall n \in \mathbb{N} \quad A_n \supseteq A_{n+1} \text{ e } \bigcap_{n \in \mathbb{N}} A_n = \emptyset \Rightarrow \lim_n M(A_n) = 0",$$

suponhamos, que $\forall n \in \mathbb{N} \quad A_n \supseteq A_{n+1} \text{ e } \bigcap_{n \in \mathbb{N}} A_n = \emptyset$, mas $\lim_n M(A_n) = \varepsilon > 0$,

então existe $\{K_n : n \in \mathbb{N}\}$ tal que:

$$\forall n \in \mathbb{N} : K_n \text{ compactos}, K_n \subseteq A_n \text{ e } M(A_n - K_n) < \varepsilon / 2^{n+1},$$

consideremos $K_n^* = \bigcap_{1 \leq j \leq n} K_j \quad \forall n \in \mathbb{N}$, $\{K_n^* : n \in \mathbb{N}\}$ satisfaz:

$$\forall n \in \mathbb{N} : K_n^* \supseteq K_{n+1}^* \text{ e } M(K_n^*) \geq \varepsilon / 2,$$

portanto $\{K_n^* : n \in \mathbb{N}\}$ é uma família de compactos com a propriedade de intersecção finita (p.i.f.), ou seja a intersecção de qualquer subfamília finita de $\{K_n^* : n \in \mathbb{N}\}$ é não vazia, e este fato implica (em qualquer espaço topológico) que $\bigcap_{n \in \mathbb{N}} K_n^* \neq \emptyset$ (se $\bigcap_{n \in \mathbb{N}} K_n^* = \emptyset \Rightarrow$ se $m \in \mathbb{N}$, $K_m^* \subseteq \Omega$

$= \bigcup_{n \in N} (K_n^*)^c$, então $K_m^* \subseteq \bigcup_{n \in F} (K_n^*)^c$ onde F é um subconjunto finito de N e $m \notin F$, portanto $K_m^* \cap \left(\bigcap_{n \in F} K_n^* \right) = \emptyset$ e $\{K_n^* : n \in N\}$ não teria p.i.f.),

mas $\bigcap_{n \in N} K_n^* \neq \emptyset$ contradiz a suposição $\bigcap_{n \in N} A_n = \emptyset$, portanto necessariamente $\lim_n M(A_n) = 0$;

f) acabamos de provar que M é σ -aditiva em $\mathcal{B}^*$; uma versão do teorema de extensão de Caratheodory diz que M admite uma única extensão a uma medida na σ -álgebra gerada por $\mathcal{B}^*$, que é a σ -álgebra $\mathcal{B}$ dos conjuntos borelianos de (Ω, d) . $\square$

Proposição 3.18.

Sejam (Ω, d) espaço métrico separável e completo,

$\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+(\Omega, d) = \{\text{medidas finitas sobre } (\Omega, \mathcal{B})\}$,

$\Pi = \Pi_d$ a métrica de Prohorov em $\mathcal{M}^+(\Omega, d)$.

Nestas condições o espaço métrico $(\mathcal{M}^+, \Pi_d)$ é separável e completo.

Demonstração:

a separabilidade de $(\mathcal{M}^+, \Pi_d)$, sob a hipótese da separabilidade, foi provada na Proposição 3.12; portanto só falta demonstrar a completitude;

sejam $\{M_n : n \in N\}$ uma seqüência de Cauchy em $(\mathcal{M}^+(\Omega, \mathcal{B}), \Pi_d)$ e $f \in CL(\Omega, d)$, isto é, $f : (\Omega, d) \rightarrow (R, e)$, contínua e limitada;

pelo Lema 3.16, $\{M_n f^{-1} : n \in \mathbb{N}\}$ é uma seqüência de Cauchy em $(M^+(R, e\text{-borelianos}), L)$, cuja completitude foi demonstrada na Proposição 3.7),

então $\exists M_f$ em $(M^+(R, e\text{-borelianos}), L)$ tal que $M_n f^{-1} \xrightarrow{w} M_f$,

então $\int (g \circ f) dM_n = \int g dM_n f^{-1} \xrightarrow{w} \int g dM_f \quad \forall g \in CL(R, e)$,

em particular para $g(t) = \max\{\min\{t, \sup f\}, \inf f\}$ vale a sentença acima, então, dado que $g \circ f = f$, tem-se que:

$$\int f dM_n \xrightarrow{w} \int f dM_f \stackrel{\text{DEF}}{=} J(f),$$

ou seja, provou-se que: $\forall f \in CL(\Omega, d)$ existe $J(f) = \lim_n \int f dM_n$, $|J(f)| < \infty$ (*);

como $\{M_n : n \in \mathbb{N}\}$ é seqüência de Cauchy, é tight (Lema 3.15), portanto estamos nas condições do Lema 3.17, que, entre outras coisas, garante que:

i) M é σ -aditiva em $\mathcal{B}^*$,

ii) M estende-se a uma única medida em $M^+(\Omega, \mathcal{B})$,

dados i), ii) e que:

$$\int f dM = J(f) = \lim_n \int f dM_n \quad \forall f \in CL(\Omega, d),$$

está provado que $M_n \xrightarrow{w} M$ e em conseqüência a completitude de $M^+(\Omega, \mathcal{B})$. $\square$

Lema 3.19.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,
 $\mathcal{M}^+(\Omega, \mathcal{B}) = \{\text{medidas finitas em } (\Omega, \mathcal{B})\}$ e Π_d a métrica de Prohorov associada
a d .

Sejam M, f, δ, ϵ e H tais que:

$M \in \mathcal{M}^+(\Omega, \mathcal{B})$, $f: (\Omega, \mathcal{B}) \xrightarrow{\text{mensurável}} (\Omega, \mathcal{B})$, $\delta > 0$, $H = \{\omega \in \Omega : d(\omega, f(\omega)) < \delta\}$ $\epsilon > 0$ e
 $M(H^c) \leq \epsilon$

então: $\Pi_d(M, Mf^{-1}) \leq \text{máximo}\{\epsilon, \delta\}$.

Demonstração:

$\forall B$ boreliano vale que: $B \cap H \subseteq f^{-1}(B^\delta)$ e $f^{-1}(B) \cap H \subseteq B^\delta$,

e dado que $M(H^c) \leq \epsilon$,

tem-se que: $M(B) = M(B \cap H) + M(B \cap H^c) \leq M(f^{-1}(B^\delta)) + \epsilon$, e também

$M(f^{-1}(B)) = M(f^{-1}(B) \cap H) + M(f^{-1}(B) \cap H^c) \leq M(B^\delta) + \epsilon$,

portanto $\Pi_d(M, Mf^{-1}) \leq \text{máximo}\{\epsilon, \delta\}$. $\square$

Corolário do Lema 3.19.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+(\Omega, \mathcal{B}) =$
 $\{\text{medidas finitas em } (\Omega, \mathcal{B})\}$, Π_d a métrica de Prohorov associada a d ,
 $M \in \mathcal{M}^+(\Omega, \mathcal{B})$.

Dados $\{\omega_0, \omega_1, \omega_2, \dots, \omega_n\} \subseteq \Omega$ e $\delta > 0$, seja

$\psi = \psi(\omega_0, \omega_1, \omega_2, \dots, \omega_n, \delta)$, $\Psi: \Omega \rightarrow \Omega$ definida por:

$$\psi(\omega) = \omega_1 \text{ se } \omega \in B(\omega_1, \delta)$$

$$\psi(\omega) = \omega_j \text{ se } \omega \in B(\omega_j, \delta) - \bigcup_{i=1}^{j-1} B(\omega_i, \delta) \text{ se } 2 \leq j \leq n$$

$$\psi(\omega) = \omega_0 \text{ se } \omega \notin \bigcup_{i=1}^n B(\omega_i, \delta),$$

$$\text{onde } B(\omega, \delta) = \left\{ \omega' \in \Omega : d(\omega', \omega) < \delta \right\},$$

nestas condições:

$$M\left(\left(\bigcup_{i=1}^n B(\omega_i, \delta)\right)^c\right) < \varepsilon \Rightarrow \Pi_d(M, M\psi^{-1}) < \max\{\varepsilon, \delta\}.$$

Demonstração:

$$H' = \bigcup_{i=1}^n B(\omega_i, \delta) \subseteq \{\omega \in \Omega : d(\omega, f(\omega)) < \delta\} = H,$$

$$\text{dado que: } M(H^c) \leq M(H'^c) < \varepsilon$$

$$\text{então, pelo Lema 3.18: } \Pi_d(M, M\psi^{-1}) < \max\{\varepsilon, \delta\}. \square$$

Lema 3.20.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+(\Omega, \mathcal{B}) = \{\text{medidas finitas em } (\Omega, \mathcal{B})\}$, Π_d a métrica de Prohorov associada a d , $M \in \mathcal{M}^+(\Omega, \mathcal{B})$.

O subconjunto finito de Ω e $a > 0$, nestas condições:

$$\mathcal{A} = \{M \in \mathcal{M}^+(\Omega, \mathcal{B}) : M(\Omega) \leq a \text{ e suporte}(M) \subseteq A\}$$

é um subconjunto totalmente limitado de $(\mathcal{M}^+(\Omega, \mathcal{B}), \Pi_d)$.

Demonstração:

seja $\varepsilon > 0$, prefixado,

se $A = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$, seja $n \in \mathbb{N}$ tal que $n^{-1} < \varepsilon/k$,

definamos agora $\mathcal{A}^* : \mathcal{A}^* = \left\{ N \in \mathcal{A} : N\{\alpha_j\} \text{ é múltiplo de } n^{-1}, 1 \leq j \leq k \right\} =$

$$= \left\{ N \in \mathcal{M}^+(\Omega, \mathcal{B}) : \begin{array}{l} M(\Omega) \leq a, \text{ suporte}(M) \subseteq A \\ N\{\alpha_j\} \text{ é múltiplo de } n^{-1}, 1 \leq j \leq k \end{array} \right\},$$

portanto, sendo $\mathcal{A}^*$ finito, para provar que $\mathcal{A}$ é totalmente limitado em $(\mathcal{M}^+(\Omega, \mathcal{B}), \Pi_d)$ é suficiente mostrar que:

$$\mathcal{A} \subseteq \bigcup_{N \in \mathcal{A}^*} B_{\Pi}^*(N, \varepsilon), \quad \text{onde } B_{\Pi}^*(N, \varepsilon) = \{Q \in \mathcal{M}^+(\Omega, \mathcal{B}) : \Pi_d(Q, N) < \varepsilon\},$$

ou equivalentemente $\forall M \in \mathcal{A} \exists M^* \in \mathcal{A}^*$ tal que $\Pi_d(M, M^*) < \varepsilon$:

seja então $M \in \mathcal{A}$, define-se $M^* \in \mathcal{A}^*$ por:

$$M^*\{\alpha_j\} = n^{-1}[n \cdot M\{\alpha_j\}], \quad 1 \leq j \leq k \text{ e } M^*(A^c) = 0,$$

onde $[t]$ denota a parte inteira de t , portanto:

$$\text{se } 1 \leq j \leq k: M^*\{\alpha_j\} \leq M\{\alpha_j\} \leq M^*\{\alpha_j\} + n^{-1} \text{ e } M^*(A^c) = M(A^c) = 0,$$

$$\text{então: } \forall B \in \mathcal{B}: M^*(B) \leq M(B) \leq M(B^T) + \tau \quad \forall \tau > 0, \quad \text{e}$$

$$\forall B \in \mathcal{B}: M(B) \leq M^*(B) + k/n \leq M^*(B^{k/n}) + k/n, \text{ em consequência,}$$

$$\text{dado que } k/n < \varepsilon: \quad \Pi_d(M, M^*) < \varepsilon. \quad \square$$

Proposição 3.21.

Sejam (Ω, d) espaço métrico separável e completo, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medidas finitas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$, $\mathcal{T}$ a topologia fraca e Π_d a métrica de Prohorov em $\mathcal{M}^+$ e $N \subseteq \mathcal{M}^+$. Então:

$$N \text{ é } \mathcal{T}\text{-relativamente compacto} \quad \text{see} \quad \left\{ \begin{array}{l} N \text{ é tight} \\ \text{e} \\ \sup\{M(\Omega) : M \in N\} < \infty \end{array} \right\}.$$

Demonstração:

($\Leftarrow$) seja agora $N \subseteq \mathcal{M}^+$ tal que N é tight e $\sup\{M(\Omega) : M \in N\} = \mu < \infty$; dada a completitude de $(\mathcal{M}^+, \Pi_d)$, para demonstrar que N é compacto é suficiente provar que N é totalmente limitado, ou seja, que $\forall \epsilon > 0$ existe uma família finita B de Π_d -esferas tal que $N \subseteq \bigcup_{B \in B} B$;

sendo N tight, dado $\epsilon > 0$, a Proposição 3.14 garante que:

$$\exists \{\omega_1, \omega_2, \dots, \omega_n\} \subseteq \Omega \text{ tal que } M\left(\bigcup_{j=1}^n (B(\omega_j, \epsilon/2))^c\right) < \epsilon/2 \quad \forall M \in N,$$

seja ω_0 um outro elemento de Ω , considere-se a função Ψ ,

$$\Psi = \psi(\omega_0, \omega_1, \omega_2, \dots, \omega_n, \epsilon/2)$$

definida no Corolário do Lema 3.19 e defina-se:

$$N^\Psi \stackrel{\text{DEF}}{=} \left\{ M\Psi^{-1} : M \in N \right\},$$

dados $M(\Psi^{-1}(\Omega)) = M(\Omega) \quad \forall M \in \mathcal{M}^+$ (já que $\forall M \in \mathcal{M}^+ : \Psi^{-1}(\Omega) = \Omega$), e a definição de Ψ , tem-se que:

$$\mathcal{N}^\Psi \subseteq \mathcal{A} = \left\{ M \in \mathcal{M}^+ : M(\Omega) \leq \mu \text{ e } \text{suporte}(M) \subseteq \{\omega_0, \omega_1, \omega_2, \dots, \omega_n\} \right\},$$

que é totalmente limitado (pelo Lema 3.20), em consequência $\mathcal{N}^\Psi$ é totalmente limitado, portanto

$$\text{existem } M_1, M_2, \dots, M_k \text{ em } \mathcal{M}^+ \text{ tais que } \mathcal{N}^\Psi \subseteq \bigcup_{j=1}^k B_\pi(M_j, \varepsilon/2),$$

$$\text{onde } B_\pi(M, \tau) = \left\{ N \in \mathcal{M}^+ : \Pi_d(N, M) < \tau \right\};$$

agora temos que: $\Pi_d(M, M\psi^{-1}) < \varepsilon/2 \quad \forall M \in \mathcal{M}^+$ (Corolário da Prop. 3.19),

$$\mathcal{N}^\Psi \stackrel{\text{DEF}}{=} \left\{ M\psi^{-1} : M \in \mathcal{N} \right\} \quad (\text{por definição}),$$

$$\mathcal{N}^\Psi \subseteq \bigcup_{j=1}^k B_\pi(M_j, \varepsilon/2) \quad (\text{por ser } \mathcal{N}^\Psi \text{ totalmente limitado}),$$

$$\text{o que permite concluir que: } \mathcal{N} \subseteq \bigcup_{j=1}^k B_\pi(M_j, \varepsilon),$$

sendo ε arbitrário, $\mathcal{N}$ resulta totalmente limitado e, sendo $(\mathcal{M}^+, \Pi_d)$ completo, $\mathcal{N}$ é relativamente compacto;

($\Rightarrow$) seja $\mathcal{N}$ relativamente compacto, para provar que $\mathcal{N}$ é tight, dada a Proposição 3.14, basta provar que dados ε e δ positivos existe $\{\omega_1, \omega_2, \dots, \omega_n\} \subseteq \Omega$ tais que:

$$M\left(\bigcup_{j=1}^n B_\pi(\omega_j, \delta)\right) > M(\Omega) - \varepsilon \quad \forall M \in \mathcal{N};$$

$\mathcal{N}$ é por hipótese relativamente compacto em $(\mathcal{M}^+, \Pi_d)$, então $\mathcal{N}$ é totalmente limitado, e então, dado $\lambda > 0$ (a fixar a posteriori), existe uma subfamília finita $\mathcal{N}' = \mathcal{N}'(\mathcal{N}, \lambda)$ de $\mathcal{N}$ tal que: $\mathcal{N} \subseteq \bigcup_{M' \in \mathcal{N}'} B_d(M', \lambda)$,

dada a completitude de (Ω, d) e o fato de $\mathcal{N}'$ ser finita $\exists \{\omega_1, \omega_2, \dots, \omega_n\} \subseteq \Omega$ ($\{\omega_1, \omega_2, \dots, \omega_n\}$ depende de $\mathcal{N}$ e λ) tal que:

$$M' \left(\bigcup_{j=1}^n B_d(\omega_j, \delta/2) \right) > M'(\Omega) - \lambda \quad \forall M' \in \mathcal{N}' \quad (*),$$

e dado que $\forall M \in \mathcal{N} \exists M' \in \mathcal{N}'$ tal que $\Pi(M, M') < \lambda$, existindo as seguintes relações entre M e M' :

$$M' \left(\bigcup_{j=1}^n B_d(\omega_j, \delta/2) \right) \leq M \left(\bigcup_{j=1}^n B_d(\omega_j, \delta/2) \right)^\lambda + \lambda = M \left(\bigcup_{j=1}^n B_d(\omega_j, \delta/2 + \lambda) \right) + \lambda,$$

$$\text{ou seja,} \quad M' \left(\bigcup_{j=1}^n B_d(\omega_j, \delta/2) \right) \leq M \left(\bigcup_{j=1}^n B_d(\omega_j, \delta/2 + \lambda) \right) + \lambda,$$

portanto, dados (*) e a desigualdade acima,

$$M \left(\bigcup_{j=1}^n B_d(\omega_j, \delta/2 + \lambda) \right) + \lambda > M'(\Omega) - 2\lambda \quad (**)$$

também, para o mesmo par (M, M') vale:

$$M'(\Omega) > M(\Omega) - \lambda \quad (***) \quad \text{pois } |M(\Omega) - M'(\Omega)| \leq \Pi_d(M, M'),$$

finalmente, dados (**) e2 (***), provou-se que:

$$\forall M \in \mathcal{N}: \quad M \left(\bigcup_{j=1}^n B_d(\omega_j, \delta/2 + \lambda) \right) > M(\Omega) - 3\lambda,$$

tomando $\lambda \leq \min\{\lambda/2, \varepsilon/3\}$, obtém-se

$$\forall M \in \mathcal{N} \quad M\left(\bigcup_{j=1}^n B_d(\omega_j, \delta)\right) > M(\Omega) - \varepsilon,$$

o que, dado que ε e δ são positivos e arbitrários, prova que $\mathcal{N}'$ é tight;

para terminar esta demonstração falta apenas ver que o conjunto

$$\left\{ M(\Omega) : M \in \mathcal{N} \right\} \text{ é limitado em } \mathbb{R},$$

mas já vimos que, sendo $\mathcal{N}$ totalmente limitado (por ser relativamente compacto), $\exists \lambda$ e $\mathcal{N}'$ subconjunto finito de $\mathcal{N}$ tais que $\mathcal{N} \subseteq \bigcup_{M' \in \mathcal{N}'} B_{\Pi_d}(M', \lambda)$,

$$\text{então } \forall M \in \mathcal{N} \exists M' \in \mathcal{N}' \text{ tal que } |M(\Omega) - M'(\Omega)| \leq \Pi_d(M, M') < \lambda,$$

$$\text{portanto } \sup\{M(\Omega) : M \in \mathcal{N}\} \leq \lambda + \sup\{M'(\Omega) : M' \in \mathcal{N}'\},$$

$$\text{e, sendo } \mathcal{N}' \text{ finito, } \sup\{M(\Omega) : M \in \mathcal{N}\} < \infty. \square$$

Proposição 3.22.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{M}^+ = \left\{ \begin{array}{l} \text{medi}_{2\text{das}} \text{ finitas} \\ \text{definidas em } (\Omega, \mathcal{B}) \end{array} \right\}$, $\Pi = \Pi_d$ a métrica de Prohorov em $\mathcal{M}^+$ e $\mathcal{F} = \{M \in \mathcal{M}^+ : M(\Omega) = 1\}$. Então: $(\mathcal{M}^+, \Pi_d)$ é completo see $(\mathcal{F}, \Pi/\mathcal{F})$ é completo.

Demonstração:

$(\Rightarrow)$: em um espaço métrico completo um subconjunto é completo see ele é fechado, portanto é suficiente provar que $\mathcal{F}$ é fechado:

sejam então $\{M_n, n \in \mathbb{N}\}$ seqüência em $\mathcal{F}$, $M \in \mathcal{M}^+$ e $\lim_n \Pi(M_n, M) = 0$, provaremos que $M \in \mathcal{F}$:

dado que $N \in \mathcal{M}^+, L \in \mathcal{M}^+ \Rightarrow \Pi(N, L) \geq |N(\Omega) - L(\Omega)|$,

obtem-se $1 = \lim_n M_n(\Omega) = M(\Omega)$; portanto $M \in \mathcal{F}$;

($\Leftarrow$) agora assume-se a completitude de $(\mathcal{F}, \Pi/\mathcal{F})$ e deve-se provar a completitude de $(\mathcal{M}^+, \Pi)$, o seja que se $\{M_n, n \in \mathbb{N}\}$ é uma seqüência de Cauchy em $\mathcal{M}^+$ ela será convergente:

em primeiro lugar provaremos que se $\{M_n, n \in \mathbb{N}\}$ é uma seqüência de Cauchy em $(\mathcal{M}^+, \Pi)$ então $\{M_n(\Omega), n \in \mathbb{N}\}$ é convergente em $\mathbb{R}$:

com efeito, lembrando mais uma vez que $N \in \mathcal{M}^+, L \in \mathcal{M}^+ \Rightarrow \Pi(N, L) \geq |N(\Omega) - L(\Omega)|$,

tem-se que se $\{M_n, n \in \mathbb{N}\}$ for seqüência de Cauchy em $(\mathcal{M}^+, \Pi)$, então $\{M_n(\Omega), n \in \mathbb{N}\}$ será Cauchy em $\mathbb{R}$, portanto, dado que $\mathbb{R}$ é completo existirá $\lambda \geq 0$ tal que $\lambda = \lim_n M_n(\Omega)$,

serão agora analisados em separado os casos $\lambda = 0$ e $\lambda > 0$;

caso $\lambda = 0$:

sabe-se que $\Pi(N, L) \leq \max\{N(\Omega), L(\Omega)\}$ (Proposição 3.9.a)), então se θ denota a medida nula: $\forall n \in \mathbb{N}: \Pi(M_n, \theta) \leq M_n(\Omega)$,

portanto $\theta = \lim_n M_n$;

caso $\lambda > 0$:

dado que $\lambda > 0$, $\exists n_0$ tal que $n \geq n_0 \Rightarrow M_n(\Omega) > 0$, provaremos que: (1) a sequência $\{M_n/M_n(\Omega), n \geq n_0\} \subseteq \mathcal{F}$ é de Cauchy e portanto a existência de $P \in \mathcal{F}$ tal que $\lim_n \Pi(M_n/M_n(\Omega), P) = 0$,

(2) $\lim_n \Pi(M_n, \lambda P) = 0$, o que completará a demonstração;

(1): a Proposição 3.10(a) afirma que

$$\tau \geq 0, N \text{ e } L \text{ em } \mathcal{M}^+ \Rightarrow \Pi(\tau L, \tau M) \leq (\tau \vee 1) \cdot \Pi(N, L),$$

então tem-se que: $\forall m \geq n_0, \forall n \geq n_0$:

$$\Pi(M_n/M_n(\Omega), M_m/M_m(\Omega)) \leq ([1/M_n(\Omega)] \vee 1) \cdot \Pi(M_n, [M_n(\Omega)/M_m(\Omega)]M_m),$$

dado que $\lim_n (1/M_n(\Omega)) = 1/\lambda$,

existe $k \in \mathbb{R}$ tal que $\forall n \geq n_0: 1/M_n(\Omega) \leq k$,

portanto:

$$\forall m \geq n_0, \forall n \geq n_0: \Pi(M_n/M_n(\Omega), M_m/M_m(\Omega)) \leq (k \vee 1) \cdot \Pi(M_n, [M_n(\Omega)/M_m(\Omega)]M_m) \quad (*),$$

sendo que $\lim_n M_n(\Omega) = \lambda > 0$ e dado que $\lim_n \alpha_n = \alpha$ e $\lim_n \beta_n = \beta$ e $\beta \neq 0$ implicam que $\lim_n \alpha_n/\beta_n = \alpha/\beta$, tem-se que:

dado $\varepsilon > 0 \exists n_1 = n_1(\varepsilon)$, $n_1 \geq n_0$ tal que: $\forall n \geq n_1: 1 - \varepsilon \leq M_n(\Omega)/M_m(\Omega) \leq 1 + \varepsilon$, (**)

por outro lado é fácil verificar que:

$$N \in \mathcal{M}^+, \varepsilon > 0 \Rightarrow \Pi(N, \nu N) \leq \varepsilon \cdot N(\Omega) \quad \forall \nu \text{ tal que } |\nu| \leq \varepsilon, \quad (***)$$

também, dado que $\{M_n(\Omega), n \in \mathbb{N}\}$ converge

$\exists h > 0$ tal que: $\forall n \in \mathbb{N}: M_n(\Omega) \leq h, \quad (****)$

finalmente, considerando (*), (**), (***) e (****), provou-se que:

dado $\varepsilon > 0 \exists n_1 = n_1(\varepsilon), n_1 \geq n_0$ tal que

$$m \geq n_1, \forall m \geq n_1 \quad \Pi(M_n/M_n(\Omega), M_m/M_m(\Omega)) \leq (k \vee 1) \cdot \varepsilon \cdot h$$

(sendo k e h constantes dependentes da sequência $\{M_n, n \in \mathbb{N}\}$),

portanto a sequência $\{M_n/M_n(\Omega), n \in \mathbb{N}\}$ de elementos de F é de Cauchy e portanto, dada a hipótese de completitude de $\mathcal{F}$, $\exists P \in \mathcal{F}$ tal que $\lim_n \Pi(M_n/M_n(\Omega), P) = 0$, então (1) está provado;

(2): dados os seguintes fatos:

$$\lim_n \Pi(M_n/M_n(\Omega), P) = 0, \quad \lim_n M_n(\Omega) = \lambda \text{ e que}$$

a aplicação $(\eta, N) \mapsto \eta N$ é contínua (Proposição 3.10(d))

$$\text{tem-se que: } \lim_n \Pi(M_n, \lambda P) = 0,$$

portanto a proposição está provada. $\square$

CAPÍTULO 4

FUNDAMENTAÇÃO FORMAL DA ESTIMAÇÃO ROBUSTA EM MODELOS PARAMÉTRICOS E JUSTIFICATIVA DA ESCOLHA DAS FERRAMENTAS MATEMÁTICAS ADEQUADAS.

4.A. INTRODUÇÃO E RESUMO DO CAPÍTULO.

O principal objetivo deste capítulo é explicar porque certos entes matemáticos foram escolhidos na formalização da teoria da robustez de Hampel. Estas idéias estão dispersas e nem sempre explícitas na literatura. Nosso trabalho aqui foi, fundamentalmente, de organização e explicitação.

Em função do objetivo acima enunciado, faremos uma breve revisão do conceito de estimador e trataremos de dar uma resposta às seguintes perguntas:

-Quais os inconvenientes dos métodos tradicionais de estimação em modelos paramétricos? Em que marco estão imersos esses modelos?

-Porque convém que o espaço das observações seja um espaço métrico separável e completo?

-Porque utilizar a métrica de Prohorov para medir distância entre medidas de probabilidade?

Também neste capítulo, completaremos algumas idéias que poderiam ter sido expostas no Capítulo 1, seção 1.D.4, mas que foram postergadas até aqui em razão de sua complexidade. Estes assuntos são tratados nas seções 4.F e 4.G. Nossa exposição foi baseada fundamentalmente em von Mises (1964), Fernholz (1983), Hampel (1968 e 1971), Huber (1964) e Martin, Yohai e Zamar (1989).

As demonstrações neste capítulo são poucas, nosso aporte aqui foi:

a) a generalização ao contexto dos processos estacionários ergódicos da demonstração de Parthasaraty da convergência das distribuições empíricas à distribuição subjacente no caso i.i.d. em espaços métricos separáveis;

b) a Proposição 4.10, onde provamos que na maioria dos casos usuais o "modelo" é um subconjunto magro do espaço $\mathcal{F}$ das probabilidades;

c) a formulação matemática das propriedades da métrica de Prohorov nos itens (1), (3) e (4) da Proposição 4.11.

4.B. ESTIMADORES DEFINIDOS ATRAVÉS DE FUNCIONAIS.

Em inferência estatística clássica, assume-se que o processo gerador das observações pode ser descrito por um modelo paramétrico. Isto é, observa-se um vetor aleatório Y a valores em R^p , cuja distribuição pertence à família de espaços de probabilidade $\mathcal{P} = \left\{ (\Omega, \mathcal{B}, P_\theta)_{\theta \in \Theta} \right\}$, onde $\Theta \subseteq R^k$. Nestas condições, pretende-se obter informação a respeito do parâmetro desconhecido θ a partir dos valores observados de Y .

É possível generalizar o esquema acima, substituindo R^p e R^k por espaços métricos separáveis e completos Ω e Λ , respectivamente. Para isso introduzimos a seguinte notação.

NOTAÇÃO.

- $(\Omega, \mathcal{B})$ espaço mensurável, onde (Ω, d) é um espaço métrico separável e completo e $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$ (Ω é o espaço das observações ou espaço amostral),

- $\mathcal{F} = \left\{ \text{medidas de probabilidade definidas em } (\Omega, \mathcal{B}) \right\}$,

- (Λ, λ) espaço métrico separável e completo, $\Theta \subseteq \Lambda$,

- $\mathcal{P} = \left\{ P_\theta, \theta \in \Theta \right\} \subseteq \mathcal{F}$ ($\mathcal{P}$ é o "modelo paramétrico"),

- Y elemento aleatório, ou seja, $Y: (S, \mathcal{Y}, P) \xrightarrow{\text{mensurável}} (\Omega, \mathcal{B})$, onde $(S, \mathcal{Y}, P)$ é um espaço de probabilidade e a distribuição de Y (ou seja a medida PY^{-1}) pertence à família $\mathcal{P}$,

- Ω^n e Ω^∞ espaços de n -uplas e de seqüências de elementos de Ω , respectivamente,

$\mathcal{B}^n$ e $\mathcal{B}^\infty$ as σ -álgebras produto em Ω^n e Ω^∞ (sendo (Ω, d) métrico separável coincidem com as σ -álgebras geradas pelas respectivas topologias produto).

A seguinte caracterização dos estimadores, devida ao matemático italiano Vito Volterra (1860-1940), possibilitou a utilização de elementos de análise funcional na formalização da teoria de robustez.

Seguiremos aqui a versão de von Mises (1964), que é a mesma de Hampel (1971).

DEFINIÇÃO 4.1.

Um estimador é uma sequência $\{T_n, n \in \mathbb{N}\}$ de aplicações

$$T_n: (\Omega^n, \mathcal{B}^n) \xrightarrow{\text{mensurável}} (\Lambda, \mathcal{A}), \quad \mathcal{A} = \{\text{borelianos de } (\Lambda, \lambda)\}.$$

Denotaremos por $w_1, w_2, \dots, w_n, \dots$ os sucessivos valores observados de Y . Também, como é costume, assume-se que os elementos aleatórios $\{Y_j = j\text{-sima realização de } Y, j \in \mathbb{N}\}$ são uma família independente. Em consequência, dada a amostra $w_1, w_2, \dots, w_n$ de tamanho n , o conjunto $\{w_1, w_2, \dots, w_n\}$ é uma estatística suficiente para θ (ou seja que a ordem dos w 's não contém informação adicional a respeito de θ , uma vez conhecido o conjunto $\{w_1, w_2, \dots, w_n\}$).

Portanto, nestas condições, basta considerar estimadores invariantes por permutações nos elementos da amostra, ou seja, que dependam apenas da medida empírica

$$F_n = n^{-1} \sum_{j=1}^n 1_{w_j},$$

onde 1_w denota a medida de probabilidade com massa 1 em w .

Observe-se que via identificação da amostra $w_1, w_2, \dots, w_n$ com a respectiva medida empírica F_n consegue-se que as "amostras" e a "distribuição subjacente" estejam no mesmo espaço $\mathcal{F}$.

$$\text{Seja } \mathcal{F}_n = \left\{ F \in \mathcal{F} : \begin{array}{l} \text{suporte de } F \text{ é finito} \\ \text{probabilidade dos átomos de } F \\ \text{é múltiplo de } 1/n \end{array} \right\}, n \in \mathbb{N};$$

em virtude dos argumentos acima expostos um estimador invariante por permutações pode ser pensado como uma seqüência de aplicações:

$$T_n: \mathcal{F}_n \longrightarrow \Lambda,$$

desde que os T_n satisfazam as seguintes condições de consistência:

$$\forall (n, m) \in \mathbb{N}^2 \quad \forall F \in \mathcal{F}_n \cap \mathcal{F}_m \quad \text{vale que } T_n(F) = T_m(F).$$

Dados $F \in \mathcal{F}$, $\{T_n, n \in \mathbb{N}\}$ estimador invariante por permutações e a hipótese de independência, fica univocamente determinada em (Λ, λ) a seqüência de medidas de probabilidade $(F^n T_n^{-1})_{n \in \mathbb{N}}$ ($F^n T_n^{-1}$ é usualmente denominada a distribuição de T_n sob F).

DEFINIÇÃO 4.2.

Um caso particular importante ocorre quando existem um subconjunto $\mathcal{D}$ de $\mathcal{F}$ e uma função T definida em $\mathcal{D}$ tais que:

$$\bigcup_{n \in \mathbb{N}} \mathcal{F}_n \subseteq \mathcal{D} \quad \text{e} \quad T: \mathcal{D} \longrightarrow \Theta,$$

então $\{T_n, n \in \mathbb{N}\}$ definido por $T_n = T/\mathcal{F}_n$ é um estimador invariante por permutações. A função T é dita um funcional estatístico com domínio $\mathcal{D}$.

EXEMPLOS.

Muitos estimadores clássicos podem ser definidos pelo procedimento acima, como o mostram os seguintes exemplos:

(1) dada $X: \Omega \rightarrow \mathbb{R}$ mensurável, $\forall F \in \mathcal{D} = \{F \in \mathcal{F}: \int |X| dF < \infty\}$ define-se $T(F) = \int X dF = \int X(w) dF(w)$; T é o funcional esperança ou média de X sob F , será especialmente estudado na seguinte seção e satisfaz:

$$F_n = n^{-1} \sum_{j=1}^n \delta_{w_j} \Rightarrow T(F_n) = n^{-1} \sum_{j=1}^n X(w_j), \text{ a média amostral;}$$

(2) mas, em geral, se $X: \Omega \rightarrow \mathbb{R}$ mensurável, define-se o momento de ordem m $T(F) = \int X^m dF \forall F \in \mathcal{D} = \{F \in \mathcal{F}: \int |X|^m dF < \infty\}$, então $T(F_n)$ é o m -simo momento amostral de X sob F ;

(3) se $X: \Omega \rightarrow \mathbb{R}$ mensurável, $m > 1$ e $\mathcal{D}$ é como no exemplo anterior, define-se $T(F) = \int |X - E_F(X)|^m dF$ (o momento central de ordem m), sendo $T(F_n)$ o m -simo momento central amostral; no caso de $m=2$ temos a variância de X sob F .

(4) Se $\psi: \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ mensurável, define-se $T(F)$ como a solução de $\int \psi(w, T(F)) dF(w) = 0$, se a solução da equação anterior existir e for única (em caso de várias soluções a unicidade pode ser forçada via restrições); portanto T_n (que corresponde a $F = F_n$) é definido como a solução de $\sum_{j=1}^n \psi(w_j, T_n) = 0$ (o M-estimador de Huber).

4.C. CONVERGÊNCIA DAS MEDIDAS EMPÍRICAS À DISTRIBUIÇÃO SUBJACENTE. NECESSIDADE DE FUNCIONAIS CONTÍNUOS.

OBSERVAÇÃO 4.3.

Dadas variáveis aleatórias i.i.d. com função de distribuição acumulada comum F , o teorema de Glivenko-Cantelli afirma que:

$\exists C \subseteq \mathbb{R}$ com $F(C)=1$ tal que $\forall \omega \in C$:

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |F_n(t, \omega) - F(t)| = 0,$$

onde $F_n(t, \omega)$ denota a distribuição empírica associada a uma amostra aleatória simples de F .

O resultado acima implica:

$\exists C \subseteq \mathbb{R}$ com $F(C)=1$ tal que $\forall \omega \in C$

$$\lim_{n \rightarrow \infty} F_n(t, \omega) = F(t) \quad \forall t \in \mathbb{R} \quad \left(\begin{array}{l} \text{em particular } \forall t \in \mathbb{R} \\ \text{onde } F \text{ é contínua} \end{array} \right),$$

ou seja que:

$$F_n(\omega) \xrightarrow{\text{em distribuição}} F \text{ com } F\text{-probabilidade } 1.$$

A última afirmação permanece válida num contexto mais geral: as variáveis aleatórias podem ser substituídas por elementos aleatórios a valores em qualquer espaço métrico separável (Ω, d) , sendo F qualquer

medida de probabilidade nos borelianos de (Ω, d) e F_n a medida empírica associada a uma amostra iid de F . Concretamente, tem-se o seguinte resultado:

Ω espaço métrico separável, $(S, \mathcal{P}, P)$ espaço de probabilidade,

$\forall j \in \mathbb{N}: Y_j: (S, \mathcal{P}, P) \xrightarrow{\text{mensurável}} (\Omega, \mathcal{B}), Y_1, Y_2, \dots, Y_n, \dots \text{ i.i.d. } (F),$

então vale que:

$$F_n(s) = \frac{1}{n} \sum_{j=1}^n Y_j(s) \xrightarrow[\text{com } F\text{-probabilidade } 1]{\text{em distribuição}} F.$$

Parthasaraty (1967) exhibe uma demonstração desta afirmação, baseada na existência, em todo espaço métrico separável, de classes limite-determinantes de cardinalidade enumerável. A argumentação de Parthasaraty pode ser extendida sem dificuldades para produzir uma versão do resultado acima válida para processos estacionários e ergódicos, como se verá na proposição 4.5.

DEFINIÇÃO 4.4.

$(H, \mathcal{X})$ espaço topológico, denotemos por $CL(H, \mathcal{X})$, ou por $CL(H)$, se não houver possibilidade de confusão, a

$$CL(H) = \{f: (H, \mathcal{X}) \xrightarrow[\text{limitada}]{\text{contínua}} \mathbb{R}\};$$

$F \in CL(H)$, F é, por definição, uma classe limite determinante se e somente se:

$\{\nu_n, n \in \mathbb{N}\}$ e ν medidas de probabilidade em $(H, \text{borel. de } \mathcal{X})$,

se $\text{ffd}_{\nu_n} \xrightarrow{\omega} \text{ffd}_{\nu} \quad \forall \epsilon \in F$ então $\nu_n \xrightarrow{\omega} \nu$.

PROPOSIÇÃO 4.5.

(Ω, d) espaço métrico separável, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

μ uma medida de probabilidade em $(\Omega^{\infty}, \mathcal{B}^{\infty})$, $K = \{1, 2, \dots, m\}$, μ_K a K -distribuição marginal de μ . Nestas condições a seqüência $(\mu_K(\omega^n))_{n \in \mathbb{N}}$ definida por

$$(\mu_K(\omega^n)) = (n-m+1)^{-1} \sum_{j=1}^{n-m+1} 1(\omega_j, \omega_{j+1}, \dots, \omega_{j+m-1})$$

satisfaz

$$\mu_K(\omega^n) \xrightarrow[n \rightarrow \infty]{\omega} \mu_K \text{ com } \mu_K\text{-probabilidade } 1.$$

Comentário: deve-se salientar que esta proposição também é válida para qualquer $K = \{k_1, k_2, \dots, k_n\}$ subconjunto finito dos naturais e $\Omega^K = \{(\omega_{k_1}, \omega_{k_2}, \dots, \omega_{k_m}) : \omega_{k_j} \in \Omega, 1 \leq j \leq m\}$; na demonstração em 4.F provamos apenas o caso $K = \{1, 2, \dots, m\}$ para evitar complicações na notação. $\square$

OBSERVAÇÃO 4.6.

Dado que F_n converge em distribuição a F com F -probabilidade 1 (Observação 4.3), tem-se que:

Ω espaço métrico separável,
 T funcional estatístico definido em F_0 ,
 T fracamente contínuo,

então:

$$T(F_n) \xrightarrow[n \rightarrow \infty]{\text{em distribuição com } F\text{-probabilidade } 1} T(F);$$

isto significa que, satisfeitas as condições acima, $T(F)$ poderá ser razoavelmente estimado pelo seu estimador natural $T(F_n)$.

Em processos estacionários ergódicos, segundo a Proposição 4.5, somente temos convergência das distribuições empíricas às marginais finito-dimensionais; em Papantoni-Kazacos e Gray (1979) pode ser visto que mesmo assim, se T for um funcional contínuo consistente Fisher com domínio nos processos estacionários e ergódicos de $(\Omega^\infty, \mathcal{B}^\infty)$, o estimador natural do parâmetro $T(\mu)$ a partir do vetor observado ω^n será $T(\mu_{\omega^n})$, onde μ_{ω^n} é o processo estacionário ergódico definido $\forall B \in \mathcal{B}^\infty$ por:

$$\mu_{\omega^n}(B) = n^{-1} \sum_{j=0}^{n-1} 1_B(F^j \omega(n)),$$

onde $\omega(n) \in \Omega^\infty$ é definido por:

$$(\omega(n))_i = (\omega^n)_{i-}, \quad 0 \leq i \leq n-1, \quad i- \text{ congruente } i \text{ mód}(n),$$

e $F: \Omega^\infty \rightarrow \Omega^\infty$ é definido por:

$$(F(\omega))_i = \omega_{i+1} \quad \forall i \text{ natural}.$$

4.D. OS FUNCIONAIS LINEARES ESTATÍSTICOS E OS ESTIMADORES A ELES ASSOCIADOS.

Entre os estimadores definidos através de funcionais merece um destaque especial a classe que definiremos a seguir.

DEFINIÇÃO 4.7.

Seja $X: \Omega \rightarrow \mathbb{R}$ mensurável, nestas condições um funcional linear estatístico é uma função $T: \mathcal{F} \rightarrow \mathbb{R}$ definida por:

$$T(F) = \int X dF = \int X(w) dF(w).$$

$T(F)$ está definida se e somente se X for F -integrável. Se $F_n = n^{-1} \sum_{j=1}^n \delta_{w_j}$, a medida empírica associada à amostra i.i.d. $\{\omega_1, \omega_2, \dots, \omega_n\}$, então $T(F_n) = n^{-1} \sum_{j=1}^n X(w_j)$.

Como é de praxe, assumiremos aqui que o processo gerador das observações é descrito por um elemento aleatório Y e que as sucessivas realizações de Y , denotadas por $\{Y_j, j \in \mathbb{N}\}$, são uma família independente. Nestas condições, os estimadores construídos via funcionais lineares estatísticos são muito úteis, em grande parte pelas propriedades condensadas na seguinte proposição:

PROPOSIÇÃO 4.8.

Seja T um funcional linear estatístico definido por $T(F) = \int X dF$, então T tem as seguintes propriedades:

(1) o domínio de T é um convexo de $\mathcal{F}$ que contém as distribuições de suporte finito;

(2) $T(F_n)$ é estimador não viesado de $T(F)$;

(3) Variância $(T(F_n)) = n^{-1}$ Variância X ;

(4) $T(F_n)$ é estimador consistente de $T(F)$ se a variância de X for finita;

(5) $T(F_n)$ é assintoticamente normal se a variância de X for finita.

OBSERVAÇÃO 4.9.

Von Mises (1947) definiu o conceito de derivada para funcionais estatísticos e desenvolveu expansões tipo Taylor para aproximá-los por funcionais mais simples. Mediante estas ferramentas estudou o comportamento assintótico dos estimadores definidos via funcionais estatísticos, em particular estendeu a normalidade assintótica dos estimadores definidos por funcionais lineares estatísticos a estimadores definidos por funcionais não lineares com convenientes propriedades de diferenciabilidade.

Estas boas propriedades dos estimadores definidos através de funcionais lineares estatísticos são, entretanto, insuficientes desde o ponto de vista da robustez, já que nos espaços de distribuições definidos sobre conjuntos de cardinalidade não finita, nem as propriedades de diferenciabilidade ou mesmo a linearidade garantem continuidade, que é uma condição que permite estimar $T(F)$ via $T(F_n)$.

Mais precisamente, o funcional linear

$$T(F) = \int X dF \text{ com } X: \Omega \rightarrow \mathbb{R} \text{ mensurável}$$

será fracamente contínuo se e somente se X for contínua e limitada (veja Huber, 1981, p. 21), condição não satisfeita, por exemplo, se $\Omega = \mathbb{R}$ e $X = \text{identidade}$.

Justamente a falta de robustez de muitos procedimentos da estatística clássica é consequência da não continuidade do funcional esperança, no qual ela está montada. Nas próximas seções serão analisadas formalmente as propriedades de continuidade dos estimadores definidos via funcionais.

Entretanto, a derivada de von Mises será considerada em 4.F, já que ela é importante na teoria de robustez pela sua relação com o conceito de função de influência de Hampel (1968 e 1974).

4.E. O MODELO PARAMÉTRICO COMO SUBCONJUNTO MAGRO DE UM ESPAÇO DE DISTRIBUIÇÕES, OS DESVIOS DO MODELO E A MÉTRICA DE PROHOROV.

Muitos procedimentos de estimação da estatística paramétrica clássica produzem estimadores ótimos em algum sentido, mas sempre sob estrita aderência ao modelo. Mas, como já temos observado reiteradamente, em qualquer situação real as hipóteses do modelo são satisfeitas apenas de uma forma aproximada, o que obriga a questionar a estabilidade das soluções obtidas pelos procedimentos tradicionais.

Como já temos dito a situação é a seguinte: observa-se um elemento aleatório $Y: (S, \mathcal{F}, P) \rightarrow (\Omega, \mathcal{B})$, onde $(S, \mathcal{F}, P)$ é um espaço de probabilidade, (Ω, d) é um espaço métrico separável e completo e $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$. A hipótese básica é que a distribuição de Y , ou seja a medida PY^{-1} em $(\Omega, \mathcal{B})$, está em

$$\mathcal{P} = \left\{ (\Omega, \mathcal{B}, P_\theta), \theta \in \Theta \right\} \text{ (o modelo),}$$

onde $\Theta \subseteq \Lambda$ e (Λ, λ) é espaço métrico separável e completo.

Hampel (1968 e 1971) considerou fundamental a apreciação das três seguintes fontes de desvio do modelo $\mathcal{P}$:

(1) erros de arredondamento, que são pequenos em valor absoluto e de alta frequência de ocorrência,

(2) erros grosseiros, de grande valor absoluto mas de rara ocorrência,

(3) os desvios produzidos pelo fato do modelo ser apenas uma simplificação do mecanismo gerador dos dados.

Para poder estudar os efeitos produzidos por estes desvios, Hampel (1968 e 1971) considerou o modelo como subconjunto do espaço $\mathcal{F} = \{\text{medidas de probabilidade sobre } (\Omega, \mathcal{B})\}$. $\mathcal{F}$ é um subconjunto gerador do espaço vetorial $\mathcal{M}$ das medidas finitas com sinal sobre $(\Omega, \mathcal{B})$. No caso usual de $\Omega = \mathbb{R}^p$, $\mathcal{M}$ é um espaço vetorial de dimensão infinita.

Assim, no caso usual $\Lambda = \mathbb{R}^k$, o modelo $\mathcal{P}$ pode ser pensado como a imagem de uma função $\tau: \Theta \rightarrow \mathcal{F}$, contínua respeito da topologia usual em $\mathbb{R}^k$ e alguma topologia razoável em $\mathcal{F}$. Isto implica que $\mathcal{P} = \tau(\Theta)$ é uma k -superfície em $\mathcal{F}$ e portanto em geral um conjunto magro, sendo que este fato subsiste em caso de se considerar um supermodelo (na Proposição 4.10 são formalizadas estas afirmações). Então, é conveniente assumir que a

distribuição de Y pertence a uma vizinhança de $\mathcal{P}$ e não necessariamente a $\mathcal{P}$.

Para estudar este problema, Hampel propôs munir o espaço $\mathcal{F}$ da topologia induzida pela métrica de Prohorov Π_d , já que esta possui as propriedades necessárias para levar em consideração os três tipos de desvio do modelo acima relacionados. Algumas dessas propriedades são mostradas na Proposição 4.11.

Deve-se salientar que as técnicas clássicas são resistentes apenas aos erros de arredondamento e muito sensíveis aos erros grosseiros e à imperfeição do modelo. De um modo geral, o uso de topologias muito finas (por exemplo, distância do supremo entre funções de distribuição acumulada) impede os erros grosseiros de "chegarem perto", enquanto que os desvios da situação ideal modelada são simplesmente desconsiderados, já que se procura a solução ótima dentro do próprio modelo.

PROPOSIÇÃO 4.10.

Sejam (Ω, d) espaço métrico separável e completo, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{F} = \{\text{medidas de probabilidade em } (\Omega, \mathcal{B})\}$ e Π a métrica de Prohorov em $\mathcal{F}$.

a) (Ω, d) não compacto e N compacto em $(\mathcal{F}, \Pi) \Rightarrow N$ é não denso, isto é, $\text{aderência}(N)$ tem interior vazio;

b) (Ω, d) não compacto, então qualquer conjunto σ -compacto em $(\mathcal{F}, \Pi)$ tem interior vazio.

Comentário:

Como consequência desta proposição, em quase todos os casos de uso corrente, o "modelo" será um subconjunto magro de $(\mathcal{F}, \Pi)$.

Com efeito, na maioria dos casos o conjunto Θ de todos os valores possíveis do parâmetro é subconjunto de R^k e $\Omega = R^p$. Também, a função $\tau: \Theta \rightarrow P_\Theta$ ($\Theta \rightarrow \mathcal{F}$), que associa a cada valor do parâmetro θ uma distribuição P_θ do modelo, deve ser contínua (caso contrário o modelo não prestaria).

Resumindo as condições acima, tem-se: $\tau(\Theta) \subseteq \mathcal{F}$, $\Theta \subseteq R^k$, τ contínua e $\Omega = R^p$. Dado que a imagem por uma função contínua de um compacto é compacto e que R^k é σ -compacto, $\tau(\Theta)$ está contido num conjunto σ -compacto de $(\mathcal{F}, \Pi)$. Por outro lado, sendo $\Omega = R^p$ - não compacto - a Proposição 4.10 garante que $\tau(\Theta)$ tem interior vazio. $\square$

PROPOSIÇÃO 4.11.

Sejam $(S, \mathcal{S}, P)$ espaço de probabilidade, (Ω, d) espaço métrico e $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$. Então:

$$(1) \quad \left\{ \begin{array}{l} Y, Z: (S, \mathcal{S}, P) \xrightarrow{\text{elementos aleatórios}} (\Omega, \mathcal{B}) \\ F \text{ e } G \text{ as respectivas distribuições} \\ P\{s: d(Y(s), Z(s)) \geq \varepsilon\} \leq \delta \end{array} \right\} \Rightarrow \Pi_d(F, G) \leq \varepsilon \delta;$$

(2) se (Ω, d) for separável, Π_d induz em $\mathcal{F}$ a topologia da convergência em distribuição;

(3) se $0 \leq \varepsilon \leq 1$, então $\Pi_d(F, (1-\varepsilon)F + \varepsilon H) \leq \varepsilon \quad \forall H \in \mathcal{F}$;

(4) as distribuições sem momentos finitos são densas em $(\mathcal{F}(R^k), \mathcal{B})$.

Comentários:

a) se pensarmos Y como a variável de interesse e Z como a realmente mensurada, a propriedade (1) diz respeito dos erros de "arredondamento" (incluídos em $\{s: d(Y(s), Z(s)) < \epsilon\}$) e dos erros "grosseiros" (incluídos em $\{s: d(Y(s), Z(s)) \geq \epsilon\}$;

b) a propriedade (2) permite a absorção dos desvios decorrentes do fato do modelo ser apenas uma aproximação da realidade sob estudo; lembre-se que a separabilidade é uma condição suficiente para que a topologia da convergência em distribuição seja induzida por Π_d e também para a convergência das distribuições empíricas à distribuição subjacente; além da separabilidade acostuma-se também adicionar a hipótese de completitude de (Ω, d) , pois esta última é condição suficiente para que estimadores induzidos por funcionais contínuos definidos em $\mathcal{F}$ sejam consistentes, veja por exemplo em Hampel (1971, lema 2); também nos espaços métricos separáveis e completos existem boas condições para o estudo da teoria assintótica, dado que, devido ao trabalho de Prohorov (1956), estão caracterizados os seus conjuntos relativamente compactos, como foi visto no Capítulo 3;

c) em (3), F pode ser interpretado como o modelo e H como qualquer contaminação, inclusive erros grosseiros e/ou uma distribuição sem momentos finitos; portanto (3) em geral e (4) no caso particular $\Lambda = \mathbb{R}^k$ ilustram os "perigos" que podem existir numa vizinhança da distribuição "melhor comportada". $\square$

4.F. DERIVADA DE VON MISES E CURVA DE INFLUÊNCIA DE HAMPEL.

4.F.1. Derivada de von Mises e curva de influência.

DEFINIÇÃO 4.12.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{F} = \{\text{medidas de probabilidade em } (\Omega, \mathcal{B})\}$, $\mathcal{C} \subseteq \mathcal{F}$, $\mathcal{C}$ convexo, $T: \mathcal{C} \rightarrow \mathbb{R}$, $F \in \mathcal{C}$. Por definição, T é von Mises diferenciável em F se e somente se:

$$a) \exists T'_F(G-F) = \lim_{t \rightarrow 0^+} \frac{T(F+t(G-F)) - T(F)}{t} \cdot t^{-1} \quad \forall G \in \mathcal{C},$$

e também

$$b) \exists \Phi_F: \Omega \xrightarrow{\text{mensurável}} \mathbb{R} \text{ tal que } T'_F(G-F) = \int \Phi_F(\omega) d(G-F)(\omega) \quad \forall G \in \mathcal{C}.$$

OBSERVAÇÕES 4.13.

a) sendo $\mathcal{C}$ convexo, $F \in \mathcal{C}$, $G \in \mathcal{C}$ e $0 \leq t \leq 1$, então

$$F+t(G-F) = (1-t)F + tG \in \mathcal{C};$$

b) dado que $(G-F)(\Omega) = 0$, Φ_F é única a menos de uma constante, a unicidade pode ser obtida impondo a condição $\int \Phi_F(\omega) dF(\omega) = 0$;

c) se T for von Mises diferenciável em F e $1_\omega \in \mathcal{C}$, vale que:

$$\Phi_F(\omega) = \lim_{t \rightarrow 0^+} \frac{T(F+t(1_\omega-F)) - T(F)}{t} \cdot t^{-1};$$

d) se definirmos $(A(F, G))(t) = T(F+t(G-F))$, $0 \leq t \leq 1$, e T for von Mises diferenciável, então

$$T'_F(G-F) = [\delta/\delta t(A(F,G))(t)]_{t=0};$$

mais ainda, von Mises considerou a expansão em série de Taylor de $(A(F,G))(t)$ em $t=0$, sendo a expansão de primeira ordem:

$$T(G) = T(F) + T'_F(G-F) \cdot t + \text{resto } (T-F);$$

(observe-se que a expansão de A induz uma expansão de T).

DEFINIÇÃO 4.14.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{F} = \{\text{medidas de probabilidade em } (\Omega, \mathcal{B})\}$, $\mathcal{C} \subseteq \mathcal{F}$, $\mathcal{C}$ convexo, $T: \mathcal{C} \rightarrow \mathbb{R}$, $F \in \mathcal{C}$. Hampel (1968, 1974) definiu a curva de influência do estimador T na distribuição F (denotada por $IC(T, F)$) por:

$$\begin{aligned} (IC(T, F))(\omega) &= \lim_{t \rightarrow 0} \frac{T(F + t(1_\omega - F)) - T(F)}{t} = \\ &= \lim_{t \rightarrow 0} \frac{T((1-t)F + t \cdot 1_\omega) - T(F)}{t}, \end{aligned}$$

se este limite existir $\forall \omega \in \Omega$.

OBSERVAÇÕES 4.15.

a) se T for von Mises diferenciável em F e $\{1_\omega : \omega \in \Omega\} \subseteq \mathcal{C}$, então existe $IC(T, F)$ (que é igual a $\Phi_F(\omega)$ de 4.13.c)) e portanto vale que

$$T'_F(G-F) = \int (IC(T, F))(\omega) d(G-F)(\omega) \quad \forall G \in \mathcal{C};$$

a recíproca não é verdadeira, ou seja, a existência da curva de influência não é condição suficiente para a diferenciabilidade de von Mises (veja contraexemplo em Fernholz, 1983);

b) como consequência da definição acima, em caso de existência, é claro que $IC(T,F): \Omega \rightarrow R$ (pois $T: \Omega \rightarrow R$);

entretanto não haveria problemas em generalizar a definição para $T: \Omega \rightarrow R^k$, sendo que neste caso

$$IC(T,F): \Omega \rightarrow R^k;$$

c) o valor $(IC(T,F))(\omega)$ informa a respeito da velocidade de mudança do valor de T quando o modelo F é contaminado por uma distribuição de massa 1 em ω ;

a curva de sensibilidade de Tukey, usualmente denotada $(SC(\omega_1, \omega_2, \dots, \omega_n))(\omega)$ ou $SC_n(\omega)$ e definida por:

$$(n+1)(T_{n+1}(\omega_1, \omega_2, \dots, \omega_n, \omega) - T_n(\omega_1, \omega_2, \dots, \omega_n)),$$

é uma versão amostral da $IC(T,F)$ que permite estudar a influência de uma nova observação ω relativa à massa introduzida $(1/n+1)$ no valor do estimador;

4.F.2. Medidas de robustez deduzidas da curva de influência.

DEFINIÇÃO 4.16.

A influência máxima que qualquer contaminação por uma distribuição de massa 1 em ω pode ter sobre o estimador é dada pelo supremo em Ω da função de influência, que foi denominado por Hampel (1968 e 1974) *sensibilidade a erros grosseiros*, cujas iniciais em inglês são GES:

$$GES(T,F) = \sup_{\omega \in \Omega} |(IC(T,F))(\omega)|;$$

caso $T: \Omega \rightarrow R^k$, $| \cdot |$ pode ser substituído pela norma do supremo em R^k .

OBSERVAÇÃO 4.17.

Serão consideradas aqui as relações entre a $IC(T,F)$ e a variância assintótica de T , onde T é um estimador definido através de um funcional.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{F} = \{\text{medidas de probabilidade em } (\Omega, \mathcal{B})\}$, $\mathcal{C} \subseteq \mathcal{F}$, $\mathcal{C}$ convexo, $\mathcal{C} \supseteq \{\text{distribuições empíricas}\}$, $T: \mathcal{C} \rightarrow R$. Também, como usualmente, assume-se que as observações são i.i.d..

Pode-se provar que, sob certas condições, a variância assintótica do estimador T sob F é igual a

$$E(IC(T,F))^2 = \int (IC(T,F))^2(\omega) dF(\omega);$$

com efeito, dado que $\int \phi_F(\omega) dF(\omega) = 0$, fazendo na expansão de primeira ordem de von Mises dada em 4.11.d as substituições $G = F_n = n^{-1} \sum_{j=1}^n \delta_{\omega_j}$ e $t=1$, tem-se:

$$\begin{aligned} T(F_n) &= T(F) + T'(F_n - F) + \text{resto}(F_n - F) = \\ &= T(F) + \int (IC(T,F))(\omega) d(F_n)(\omega) + \text{resto}(F_n - F) = \end{aligned}$$

$$= T(F) + n^{-1} \sum_{j=1}^n (IC(T, F))(w_j) + \text{resto}(F_n - F),$$

se também

$$E_F \{IC(T, F)\}^2 = \int (IC(T, F))^2 dF(\omega) < \infty \quad e$$

$$n^{1/2} \cdot \text{resto}(F_n - F) \xrightarrow[F\text{-probabilidade}]{em} 0,$$

o teorema central do limite garante que:

$$n^{1/2} \cdot (T(F_n) - T(F)) \xrightarrow[distribuição]{em} N(0, E_F \{IC(T, F)\}^2).$$

OBSERVAÇÃO 4.18.

Serão consideradas aqui as relações entre GES e viés assintótico.

Sejam (Ω, d) espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{F} = \{\text{medidas de probabilidade em } (\Omega, \mathcal{B})\}$, $\mathcal{C} \subseteq \mathcal{F}$, $\mathcal{C}$ convexo, $\mathcal{C} \supseteq \{\text{distribuições empíricas}\}$, $T: \mathcal{C} \rightarrow \mathbb{R}$ e $(T_n)_{n \in \mathbb{N}}$ o estimador induzido por T .

Freqüentemente estimadores são comparados mediante o critério do *erro quadrático médio* (EQM) definido por:

$$\text{EQM}(T_n, F) = E_F (T_n - \theta)^2 = \text{Variância}_{F_n} T + (\text{Viés}_{F_n} T)^2,$$

onde $\text{Viés}_{F_n} T = \|E_F(T_n) - \theta\|$, $\|\cdot\|$ é uma norma conveniente no espaço paramétrico e θ é o parâmetro a estimar;

na maioria dos estimadores razoáveis

$$\lim_n \text{Variância}_{F_n} T = 0 \quad e \quad \lim_n E_F(T_n) = T(F)$$

$$\lim_{n \rightarrow \infty} E_F(T_n) = T(F);$$

nestas condições é razoável julgar um estimador através do seu *viés assintótico* dado por:

$$v(T, F) = \|T(F) - \theta\|.$$

Considere-se o modelo de contaminação, isto é, assumamos que a verdadeira distribuição pertence à família:

$$Cont.(F_0, \varepsilon) = \{F : F = (1-\varepsilon)F_0 + \varepsilon M \quad \forall M \in \mathcal{M}\},$$

onde ε é fixo e F_0 pertence ao modelo paramétrico;

neste caso, uma medida da robustez do estimador será o máximo viés no conjunto $Cont.(F_0, \varepsilon)$, ou seja

$$v(T, F_0, \varepsilon) = \sup \left\{ \|T(F) - \theta_0\| : F \in Cont.(F_0, \varepsilon) \right\},$$

se T for consistente Fisher, isto é se $T(F)$ coincidir com o parâmetro a estimar:

$$v(T, F_0, \varepsilon) = \sup \left\{ |T(F) - T(F_0)| : F \in Cont.(F_0, \varepsilon) \right\};$$

por outro lado, se existir expansão de primeira ordem de von Mises terá-se:

$$|T(F) - T(F_0)| = |T'_{F_0}(F - F_0) \cdot \varepsilon + \text{resto}(F - F_0)|,$$

sob condições suficientes para desprezarmos o resto:

$$|T(F) - T(F_0)| \cong |T'_{F_0}(F - F_0) \cdot \varepsilon| = \varepsilon \cdot \left| \int (IC(F_0, T))(\omega) d(G - F)(\omega) \right| \leq \varepsilon \cdot \text{GES}(T, F_0),$$

portanto, sob convenientes condições de diferenciabilidade de T em F_0 vale que:

$$v(T, F_0, \varepsilon) \leq \varepsilon \cdot \text{GES}(T, F_0);$$

obviamente esta fórmula só terá utilidade se $\text{GES}(T, F_0) < \infty$.

DEFINIÇÃO 4.19.

A sensibilidade local a deslocamentos do estimador T na distribuição F é usualmente denotada por LSS e definida por:

$$LSS(T, F) = \sup_{\omega \neq \omega'} \{ |(IC(T, F))(\omega) - (IC(T, F))(\omega')| / |\omega - \omega'| \};$$

esta ferramenta tem a função de mensurar os efeitos produzidos sobre os estimadores quando os valores das observações são levemente modificados, como ocorre, por exemplo, com o arredondamento e/ou agrupamento dos dados.

DEFINIÇÃO 4.20.

Este conceito tem a ver com a idéia de rejeitar os dados extremamente afastados do centro da distribuição. Se $\Omega = \mathbb{R}^n$ e F for simétrica entorno de $0 \in \mathbb{R}^n$, define-se o ponto de rejeição de T em F como:

$$\inf \{ r > 0 : (IC(T, F))(\omega) = 0 \text{ se } \|\omega\| > r \},$$

onde $\|\cdot\|$ indica alguma das normas usuais em $\mathbb{R}^n$.

4.G. ESTIMADORES ROBUSTOS COM VIÉS MINIMAX.

Consideremos as condições de 4.18, ou seja que estejam dadas as condições para julgar um estimador dado por um funcional T definido em $\mathcal{F}$ através do seu viés assintótico.

DEFINIÇÃO 4.21.

Seja $\mathcal{E}$ uma classe de estimadores definidos através de funcionais. Um estimador T^* terá viés minimax na classe $\mathcal{E}$ para uma fração de contaminação ε se e somente se:

$$v(T^*, F_0, \varepsilon) \leq \min_{T \in \mathcal{E}} v(T, F_0, \varepsilon).$$

onde, da mesma forma que em 4.18,

$$v(T, F_0, \varepsilon) = \sup \left\{ v(T, F) : F \in \text{Cont.}(F_0, \varepsilon) \right\}, \text{ e}$$

$$\text{Cont.}(F_0, \varepsilon) = \{F : F = (1-\varepsilon)F_0 + \varepsilon M \ \forall M \in \mathcal{M}\}.$$

OBSERVAÇÃO 4.22.

Será visto aqui que, sob determinadas condições, é possível obter a $\text{GES}(T, F_0)$ e o uma versão do ponto de ruptura a partir de $v(T, F_0, \varepsilon)$.

Em 4.14 foi definida a curva de influência do estimador dado pelo funcional T na distribuição F_0 por:

$$(IC(T, F_0))(\omega) = \lim_{n \rightarrow 0^+} \left[T((1-t)F_0 + t.1_\omega) - T(F_0) \right] \cdot t^{-1},$$

também (Definição 4.16):

$$GES(T, F_0) = \sup_{\omega \in \Omega} |(IC(T, F_0))(\omega)|;$$

se T for consistente Fisher, ter-se-ia que:

$$(IC(T, F_0))(\omega) = \lim_{n \rightarrow 0^+} v(T, (1-t)F_0 + t.1_\omega) \cdot t^{-1};$$

dado que por definição

$$v(T, F_0, t) = \sup \left\{ v(T, F) : F \in \text{Cont.}(F_0, t) \right\},$$

é claro que

$$v(T, F_0, t) \geq \sup_{\omega \in \Omega} v(T, (1-t)F_0 + t.1_\omega),$$

então, dadas as condições que transformem a desigualdade acima numa igualdade e que também permitam trocar a ordem entre

$$\lim_{n \rightarrow 0^+} \text{ e } \sup_{\omega \in \Omega},$$

ter-se-ia que:

$$\begin{aligned} GES(T, F_0) &= \lim_{n \rightarrow 0^+} \left(\sup_{\omega \in \Omega} v(T, (1-t)F_0 + t.1_\omega) \cdot t^{-1} \right) = \\ &= \lim_{n \rightarrow 0^+} v(T, F_0, t) \cdot t^{-1} = \delta / \delta t \left(v(T, F_0, t) \right)_{t=0}, \end{aligned}$$

ou seja que sob determinadas condições:

$$GES(T, F_0) = \delta/\delta t \left(v(T, F_0, t) \right)_{t=0} ,$$

também é claro que é possível definir uma versão do ponto de ruptura através de:

$$\varepsilon^* = \infimo \{ \varepsilon : v(T, F_0, \varepsilon) = \infty \}.$$

As duas fórmulas acima mostram que, dadas as condições adequadas, a curva $v(T, F_0, \varepsilon)$ contém muita informação, que inclui GES e ε^* .

Alguns resultados sobre estimadores com viés minimax.

Huber (1964) provou que, no modelo de posição, para F_0 simétrica e unimodal e observações i.i.d., vale que:

$$v(\text{mediana}, F_0, \varepsilon) \leq \min_{T \in \mathcal{E}} v(T, F_0, \varepsilon),$$

sendo $\mathcal{E} = \{\text{estimadores invariantes por translações}\}.$

Martin, Yohai e Zamar (1989) acharam os estimadores de viés minimax na classe $\mathcal{E} = \{\text{M-estimadores}\}$ para o modelo de regressão sob os seguintes supostos:

i) a distribuição dos erros tem densidade simétrica e unimodal,

ii) distribuição das observações elíptica, isto é a distribuição das observações x depende de $\|x\|_A = x'Ax$, sendo a matriz A simétrica e definida positiva.

O estimador com viés minimax é obtido minimizando o quantil α dos valores absolutos dos resíduos, onde α depende da fração de contaminação

ϵ . Por este motivo ele é denotado por Q_α . Nos casos estudados, o valor de α é pouco sensível a alterações no valor de ϵ .

Caso $\alpha = 1/2$, Q_α coincide com o estimador de mínima mediana de quadrados proposto por Rousseau (1984).

Martin, Yohai e Zamar, no trabalho já citado, também obtiveram estimadores com viés minimax na classe dos GM estimadores para o modelo de regressão.

4.H. DEMONSTRAÇÕES.

Proposição 4.5.

(Ω, d) espaço métrico separável, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$,

μ uma medida de probabilidade em $(\Omega^\infty, \mathcal{B}^\infty)$, $K = \{1, 2, \dots, m\}$, μ_K a K -distribuição marginal de μ . Nestas condições a sequência $(\mu_K(\omega^n))_{n \in \mathbb{N}}$ definida por

$$(\mu_K(\omega^n)) = (n-m+1)^{-1} \sum_{j=1}^{n-m+1} 1(\omega_j, \omega_{j+1}, \dots, \omega_{j+n-1})$$

satisfaz

$$\mu_K(\omega^n) \xrightarrow[n \rightarrow \infty]{\omega} \mu_K \text{ com } \mu_K\text{-probabilidade } 1.$$

Demonstração:

a hipótese de estacionariedade implica que a aplicação:

$$\omega^n = (\omega_1, \omega_2, \dots, \omega_m) \rightarrow (\omega_2, \omega_3, \dots, \omega_{m+1})$$

preserva medida, então, dado que:

$$\int f d\mu_K(\omega^n) = (n-m+1)^{-1} \sum_{j=1}^{n-m+1} f(\omega_j, \omega_{j+1}, \dots, \omega_{j+m-1}),$$

seja $CL(\Omega^K) = \left\{ f: (\Omega^K, \text{topologia produto}) \xrightarrow[\text{limitada}]{\text{contínua}} \mathbb{R} \right\}$, o teorema ergódico

garante que

$$\forall f \in CL(\Omega^K) \quad \int f d\mu_K(\omega^n) \xrightarrow[n \rightarrow \infty]{w} \int f d\mu_K \text{ com } \mu_K\text{-probabilidade } 1,$$

ou seja

$\forall f \in CL(\Omega^K) \exists N(f)$ boreliano de Ω^K com $\mu_K(C(f))=1$ tal que

$$\int f d\mu_K(\omega^n) \xrightarrow[n \rightarrow \infty]{w} \int f d\mu_K \text{ em } C(f);$$

sendo (Ω, d) separável, Ω^K com a topologia produto também é, portanto existe uma família $F \subseteq CL(\Omega^K)$, F enumerável e F limite determinante,

se $A = \bigcup_{f \in F} (C(f))^c$ então $\mu_K(A) = 0$, por ser união enumerável de conjuntos de medida nula;

finalmente, tem-se que:

$$P(A^c)=1 \text{ e em } A^c \text{ vale que } \int f d\mu_K(\omega^n) \xrightarrow[n \rightarrow \infty]{w} \int f d\mu_K \quad \forall f \in F,$$

portanto o resultado está provado. $\square$

Proposição 4.8.

Seja T um funcional linear estatístico definido por $T(F) = \int X dF$, então T tem as seguintes propriedades:

- (1) o domínio de T é um convexo de $\mathcal{F}$ que contém as distribuições de suporte finito;
- (2) $T(F_n)$ é estimador não viesado de $T(F)$;
- (3) $\text{Variância}(T(F_n)) = n^{-1} \text{Variância } X$;
- (4) $T(F_n)$ é estimador consistente de $T(F)$ se a Variância de X é finita;
- (5) $T(F_n)$ é assintoticamente normal se $\text{Variância } X < \infty$.

Demonstração:

- (1), (2) e (3) são triviais;
- (4) decorre de (2) e (3) e da desigualdade de Tchebycheff;
- (5) é uma consequência do teorema central do limite, já que $F_n = n^{-1} \sum_{j=1}^n 1_{w_j} \Rightarrow T(F_n) = n^{-1} \sum_{j=1}^n X(w_j) = n^{-1} \sum (X \circ Y_j)(s)$ onde $\{X \circ Y_j, j \in \mathbb{N}\}$ é família independente, já que $\{Y_j, j \in \mathbb{N}\}$ também é (por hipótese). $\square$

Proposição 4.10.

Sejam (Ω, d) espaço métrico separável e completo, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$, $\mathcal{F} = \{\text{medidas de probabilidade em } (\Omega, \mathcal{B})\}$ e Π a métrica de Prohorov em $\mathcal{F}$.

a) (Ω, d) não compacto e $\mathcal{N}$ compacto em $(\mathcal{F}, \Pi) \Rightarrow \mathcal{N}$ é não denso, isto é, $\text{aderência}(\mathcal{N})$ tem interior vazio;

b) (Ω, d) não compacto, então qualquer conjunto σ -compacto em $(\mathcal{F}, \Pi)$ tem interior vazio.

Demonstração:

a) sendo $(\mathcal{F}, \Pi)$ métrico, $\mathcal{N}$ compacto $\Rightarrow \mathcal{N}$ fechado, portanto deve-se provar que $\text{interior}(\mathcal{N})$ é vazio, ou equivalentemente que $\forall \varepsilon > 0 \quad \forall F_0 \in \mathcal{N} \quad (V(F_0, \varepsilon, \Pi)) \cap \mathcal{N}^c \neq \emptyset$, onde $V(F_0, \varepsilon, \Pi)$ denota a esfera aberta $\{G \in \mathcal{F} : \Pi(F_0, G) < \varepsilon\}$;

sejam $F_0 \in \mathcal{N}$, $\varepsilon > 0$ e α tal que $0 < \alpha < \varepsilon$,

$\mathcal{N}$ compacto $\Rightarrow \mathcal{N}$ tight (por Proposição 3.21), ou seja,

$\exists K_\alpha$ compacto em (Ω, d) tal que $F(K_\alpha) > 1 - \alpha \quad \forall F \in \mathcal{N}$,

dado que (Ω, d) não é compacto $\exists \omega \in K_\alpha^c$, consideremos $G_0 \in \mathcal{F}$ definida por $G_0 = (1 - \alpha)F_0 + \alpha \cdot 1_\omega$,

então

i) $G_0(K_\alpha) \leq 1 - \alpha \Rightarrow G_0 \in \mathcal{N}^c$,

ii) por 4.11.(3), $\Pi(F_0, G_0) \leq \alpha < \varepsilon \Rightarrow G_0 \in V(F_0, \varepsilon, \Pi)$,

portanto: $(V(F_0, \varepsilon, \Pi)) \cap \mathcal{N}^c \neq \emptyset$. $\square$

b) pelo Lema 3.18, o espaço métrico $(\mathcal{F}, \Pi)$ é completo, pode-se então aplicar um conhecido teorema de Baire que afirma que em um espaço métrico completo toda união enumerável de conjuntos não densos tem interior vazio,

portanto, como consequência do teorema de Baire e de a), $\mathcal{N}$ tem interior vazio. $\square$

Proposição 4.11.

Sejam $(S, \mathcal{Y}, P)$ espaço de probabilidade, (Ω, d) espaço métrico e $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$. Então:

$$(1) \left\{ \begin{array}{l} Y, Z: (S, \mathcal{Y}, P) \xrightarrow{\text{elementos aleatórios}} (\Omega, \mathcal{B}) \\ F \text{ e } G \text{ as respectivas distribuições} \\ P\{s: d(Y(s), Z(s)) \geq \varepsilon\} \leq \delta \end{array} \right\} \Rightarrow \Pi_d(F, G) \leq \varepsilon \vee \delta;$$

(2) se (Ω, d) for separável, Π_d induz em $\mathcal{F}$ a topologia da convergência em distribuição;

(3) se $0 \leq \varepsilon \leq 1$, então $\Pi_d(F, (1-\varepsilon)F + \varepsilon H) \leq \varepsilon \quad \forall H \in \mathcal{F}$;

(4) as distribuições sem momentos finitos são densas em $\mathcal{F}(R^k, \mathcal{B})$.

Demonstração:

(1) $F = PY^{-1}$ e $G = PZ^{-1}$, seja $A = \{s: d(Y(s), Z(s)) < \varepsilon\}$, então,

por hipótese e por definição de A:

$$\forall B \in \mathcal{B}: \quad Y^{-1}(B) \cap A \subseteq Z^{-1}(B^c) \text{ e } P(A^c) \leq \delta,$$

em consequência:

$$\forall B \in \mathcal{B} \quad F(B) = P(Y^{-1}(B) \cap A) + P(Y^{-1}(B) \cap A^c) \leq P(Z^{-1}(B^\varepsilon)) + \delta = G(B^\varepsilon) + \delta,$$

ou seja tem-se que

$$\forall B \in \mathcal{B} \quad F(B) \leq G(B^\varepsilon) + \delta,$$

portanto, sendo F e G medidas de probabilidade,

$$\Pi_d(F, G) \leq \varepsilon \vee \delta;$$

(2) foi demonstrado no Capítulo 3;

(3) sejam $0 \leq \varepsilon \leq 1$, $H \in \mathcal{F}(\Omega, \mathcal{B})$ e $B \in \mathcal{B}$, então

$$F(B) = (1-\varepsilon)F(B) + \varepsilon F(B) \leq (1-\varepsilon)F(B^\varepsilon) + \varepsilon \leq (1-\varepsilon)F(B^\varepsilon) + \varepsilon + \varepsilon H(B^\varepsilon), \text{ isto é}$$

$$\forall B \in \mathcal{B}: \quad F(B) \leq ((1-\varepsilon)F + \varepsilon H)(B^\varepsilon) + \varepsilon,$$

$$\text{portanto} \quad \Pi_d(F, (1-\varepsilon)F + \varepsilon H) \leq \varepsilon \quad \forall H \in \mathcal{F};$$

(4) é decorrência de (3) e da existência de $H_0 \in \mathcal{F}(R^k, \mathcal{B})$ sem momentos finitos:

$$\forall F \in \mathcal{F}(R^k, \mathcal{B}) \quad \forall \varepsilon > 0 \text{ vale que:}$$

$$(1-\varepsilon)F + \varepsilon H_0 \text{ não tem momentos finitos}$$

$$\text{e } \Pi_d(F, (1-\varepsilon)F + \varepsilon H_0) \leq \varepsilon. \square$$

CAPÍTULO 5.

RESISTÊNCIA EM PROCESSOS ESTACIONÁRIOS ERGÓDICOS.

5.A. INTRODUÇÃO E RESUMO DO CAPÍTULO.

Neste capítulo será demonstrada a equivalência dos conceitos de resistência fraca e forte de Boente, Fraiman e Yohai para estimadores invariantes por permutações das coordenadas no contexto dos processos estacionários e ergódicos. Para tanto será feita a continuação numa resenha da evolução dos conceitos de resistência (Mosteller e Tukey/ Hampel/ Boente, Fraiman e Yohai) e robustez qualitativa (Hampel/ Papantoni-Kazacos e Gray/ Boente, Fraiman e Yohai).

Hampel (1971) introduziu a noção de robustez qualitativa em estimação paramétrica, de acordo com a seguinte idéia intuitiva: um estimador é robusto para um processo μ se para todo outro processo ν "perto" de μ a distribuição do estimador sob ν fica "perto" da do estimador sob μ , uniformemente no tamanho da amostra.

Hampel trabalhou com processos a tempos discretos definidos por medidas de probabilidade em $(\Omega^\infty, \mathcal{B}^\infty)$, onde (Ω, d) é um espaço métrico separável e completo e $\mathcal{B}$ denota a σ -álgebra dos conjuntos borelianos de (Ω, d) . Em princípio, Hampel somente considerou processos i.i.d., conseguindo, via esta restrição, uma simplificação na sua abordagem em dois aspectos fundamentais, que mencionaremos a seguir.

Em primeiro lugar, para definir a "proximidade" de dois processos bastou-lhe apenas considerar a "proximidade" das distribuições marginais unidimensionais (todas iguais) dos citados processos, dado que se F é uma marginal unidimensional de um processo i.i.d. a distribuição do processo é F^∞ . Ou seja, bastou-lhe trabalhar no espaço $\mathcal{F}$ das medidas de probabilidade sobre $(\Omega, \mathcal{B})$, medindo a "proximidade" entre dois processos i.i.d. pela distância de Prohorov entre as respectivas marginais unidimensionais.

Hampel observou que o uso desta métrica lhe permitia considerar os desvios causados por erros de arredondamento, erros grosseiros e desvios do modelo, concebido como representação aproximada da realidade.

Em segundo lugar, mediante argumentos de suficiência, bastou a Hampel apenas considerar estimadores invariantes por permutações das coordenadas, o que lhe permitiu (via a identificação amostra-distribuição empírica) considerar as "amostras" e a "distribuição subjacente" como elementos do mesmo espaço (isto é, como pontos de $\mathcal{F}$) e pensar os estimadores como funcionais definidos em $\mathcal{F}$.

Papantoni-Kazacos e Gray (1981) generalizaram a noção de robustez ao âmbito dos processos estacionários, medindo a distância entre dois

processos pela métrica de Orstein em $\mathcal{P}(\Omega^\infty, \mathcal{B}^\infty)$ e a distância entre as respectivas marginais n -dimensionais pela métrica de Wasserstein em $\mathcal{P}(\Omega^n, \mathcal{B}^n)$. A definição destas duas distâncias depende da métrica $a_n(u^n, v^n) = n^{-1} \left(\sum_{j=1}^n d(u_j, v_j) \right)$ em Ω^n através do funcional esperança, o que descaracteriza as idéias intuitivas de robustez, a menos que d seja limitada, caso no qual a métrica de Wasserstein equivale à de Prohorov P_{an} . Cox (1981) mostrou que, se d for a distância usual em R , a média amostral é robusta segundo este enfoque.

Boente, Fraiman e Yohai (1987) fizeram uma proposta diferente, baseada na generalização do conceito de Π -robustez (Hampel (1971)), mais adequado ao tratamento de observações dependentes.

Eles definiram a métrica

$$d_n \text{ em } \Omega^n \text{ por: } d_n(u^n, v^n) = \inf \{ \alpha : \# \{ i : d(u_i, v_i) \geq \alpha \} \leq n \}$$

e a métrica de Prohorov associada Π_{d_n} em $\mathcal{P}(\Omega^n, \mathcal{B}^n)$, sendo desta forma preservadas as noções de robustez. Através de d_n introduziram uma nova abordagem, mediante a generalização do conceito de resistência de Tukey através de uma versão da "condition B" de Hampel. O conceito de resistência fraca de Boente, Fraiman e Yohai equivale ao de Π_{d_n} -robustez se o espaço dos parâmetros for um subconjunto de R^k , porém de definição mais transparente, já que depende em forma direta do estimador (e não da distribuição do estimador, como no caso de robustez).

Boente, Fraiman e Yohai definiram resistência fraca e forte, provando a equivalência de ambos os conceitos no caso de estimadores

invariantes por permutações em processos i.i.d.. Neste capítulo estenderemos este resultado aos processos estacionários ergódicos. A seguir será descrita a organização do capítulo.

Na seção B são apresentadas questões de notação, definidas métricas em Ω^n e feitas considerações gerais sobre os espaços Ω^n e Ω^∞ . Na seção C são dadas as definições e é feita uma breve resenha da evolução dos conceitos de robustez qualitativa e resistência. Na seção D aborda-se a equivalência entre resistência qualitativa fraca e forte. A seção E é composta de alguns lemas que auxiliam as demonstrações da seção D.

Em forma diferente aos capítulos 2,3 e 4, em que as demonstrações foram colocadas para o final, aqui optamos por escrever as demonstrações no texto.

Nossa contribuição fundamental neste capítulo é o teorema 5.22 que demonstra a equivalência entre resistência fraca e forte para estimadores invariantes por permutações em processos estacionários ergódicos, generalizando análogo resultado, restrito a processos i.i.d., de Boente, Fraiman e Yohai (1987). Também são nossos os primeiros quatro lemas (5.24, 5.25, 5.27 e 5.29) da Seção 5.E. No lema 5.24 provamos que d_n induz a topologia produto em Ω^n . O lema 5.25 constata que a métrica d_n de Boente, Fraiman e Yohai absorve o sentido intuitivo de resistência. O lema 5.27 é uma modificação -mediante a introdução da métrica d_n - de um argumento utilizado por Hampel (1971) na sua demonstração de que "condition B" implica robustez qualitativa; o lema mostra que se dois processos i.i.d. μ e ν têm marginais unidimensionais vizinhas no sentido de Prohorov, então existe um processo τ em $\Omega^\infty \times \Omega^\infty$ tal que μ e ν são suas marginais em Ω^∞ e,

além disso, as distribuições amostrais de n -uplas produzidas por μ e ν são Prohorov-vizinhas com alta τ -probabilidade. O lema conjuntista 5.29 está implícito na demonstração do Teorema 4.1 de Boente, Fraiman e Yohai (1987), porém sem demonstração. O quinto e último lema da seção 5.E (5.31) é clássico.

5.B. QUESTÕES PRELIMINARES E NOTACIONAIS.

NOTAÇÃO 5.1.

Seja Ω um conjunto não vazio,

$$\Omega^n = \{\omega^n = (\omega_1, \omega_2, \dots, \omega_n) : \omega_j \in \Omega, 1 \leq j \leq n\},$$

$$\Omega^\infty = \{\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) : \omega_j \in \Omega \forall j \text{ natural}\}.$$

Seja (Ω, d) um espaço métrico, $A \subseteq \Omega$, $A \neq \emptyset$, $\alpha > 0$,

$$V(A, \alpha, d) = \{u \in \Omega : d(u, A) < \alpha\},$$

em particular se $A = \{\omega_0\}$, $V(\{\omega_0\}, \alpha, d)$ é a esfera de centro ω_0 e raio α em (Ω, d) , denotada em capítulos anteriores por $B(\omega_0, \alpha)$, sem menção explícita à métrica d .

OBSERVAÇÃO 5.2.

Seja (Ω, d) um espaço métrico, $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$ e $\mathcal{P}(\Omega, \mathcal{B}) = \{\text{medidas de probabilidade sobre } (\Omega, \mathcal{B})\}$. No Capítulo 3 foi definida a métrica de Prohorov Π_d $\forall (\mu, \nu) \in (\mathcal{P}(\Omega, \mathcal{B}))^2$ por:

$$\Pi d(\mu, \nu) = \inf \{ \alpha : \mu(B) \leq \nu(V(B, \alpha, d)) + \alpha, \forall B \in \mathcal{B} \}$$

A seguinte caracterização de Π , devida a Strassen (1965) e válida para (Ω, d) espaço métrico separável e completo é de especial utilidade no tratamento das relações entre resistência qualitativa e robustez qualitativa:

$$\Pi d(\mu, \nu) = \inf \left\{ \alpha : \begin{array}{l} \exists P \text{ probabilidade em } (\Omega \times \Omega, \mathcal{B} \times \mathcal{B}) \\ \text{com marginais } \mu \text{ e } \nu \text{ tal que} \\ P\{ (u_1, u_2) : d(u_1, u_2) \geq \alpha \} \leq \alpha \end{array} \right\}.$$

DEFINIÇÕES 5.3.

A métrica do supremo é definida por:

$$s_n(u^n, v^n) = \sup \{ d(u_j, v_j) : 1 \leq j \leq n \};$$

sabe-se que s_n define a topologia produto em Ω^n ,

de igual modo que a métrica tipo L_2 :

$$e_n(u^n, v^n) = \left(n^{-1} \cdot \sum_{j=1}^n d^2(u_j, v_j) \right)^{1/2},$$

e a métrica tipo L_1 :

$$a_n(u^n, v^n) = \left(n^{-1} \cdot \sum_{j=1}^n d(u_j, v_j) \right).$$

Hampel (1971) definiu a seguinte métrica que, diferentemente das três anteriores, é invariante por permutações das coordenadas:

$$h_n(u^n, v^n) = \Pi d \left(n^{-1} \cdot \sum_{i=1}^n \delta_{u_i}, n^{-1} \cdot \sum_{i=1}^n \delta_{v_i} \right),$$

onde δ_u denota a medida de probabilidade com massa 1 em u ;

OBSERVAÇÃO 5.4.

Boente, Fraiman e Yohai (1987) definiram a métrica d_n por:

$$d_n(u^n, v^n) = \inf \{ \alpha : \# \{ i : d(u_i, v_i) \geq \alpha \} \leq n\alpha \},$$

provando que $\Pi d(u^n, v^n) = \text{mínimo } d_n(u^n, p(v^n))$,

$$\text{onde } p(v^n) = \left(v_{p(1)}, v_{p(2)}, \dots, v_{p(n)} \right), \quad \text{e } p \in P = \left\{ \text{permutações de } \{1, 2, \dots, n\} \right\}.$$

Pode ser demonstrado (veja lema (5.24) no apêndice) que a métrica d_n também define em Ω^n a topologia produto, conseqüentemente H_n define em Ω^n a topologia produto módulo permutações das coordenadas.

OBSERVAÇÃO 5.5.

A menos que indicado de outra forma, (Ω, d) denotará sempre um espaço métrico separável e completo e $\mathcal{B}$ o conjunto dos borelianos de (Ω, d) . Ω^n e Ω^∞ denotarão respectivamente os espaços de n -uplas e de seqüências de elementos de Ω e $\mathcal{B}^n$ e $\mathcal{B}^\infty$ as respectivas σ -álgebras produto em Ω^n e Ω^∞ .

Sendo Ω separável as σ -álgebras produto $\mathcal{B}^n$ e $\mathcal{B}^\infty$ coincidem com as σ -álgebras geradas pelas respectivas topologias produto; portanto, neste caso, tem-se que $\mathcal{B}^n = \{\text{borelianos de } (\Omega^n, s_n)\}$, e que $\mathcal{B}^\infty$ é a σ -álgebra gerada por $\bigcup_{n=1}^{\infty} \{\text{abertos de } (\Omega^n, s_n)\}$. Estas afirmações permanecem válidas se s_n for substituída por e_n , a_n ou d_n , dado que todas estas métricas definem a topologia produto.

Também vale que a distância em Ω^∞ definida por

$$d^\infty(x, y) = \sum_{n=1}^{\infty} 2^{-n} d'(x_n, y_n) \text{ com } d' = \text{mínimo}(d, 1)$$

induz a topologia produto em Ω^∞ e que $(\Omega^\infty, d^\infty)$ é também separável e completo.

5.C. ROBUSTEZ E RESISTÊNCIA QUALITATIVAS.

5.C.1. Processos e estimadores.

Serão considerados, aqui, processos definidos por medidas de probabilidade em $(\Omega^\infty, \mathcal{B}^\infty)$, onde (Ω, d) é um espaço métrico separável e completo e $\mathcal{B} = \{\text{borelianos de } (\Omega, d)\}$.

NOTAÇÃO 5.6.

$\mathcal{P}(A, \mathcal{A})$ denotará o conjunto de medidas de probabilidade sobre $(A, \mathcal{A})$, onde $(A, \mathcal{A})$ é qualquer espaço mensurável.

Se $(T_n)_{n \in \mathbb{N}}$ é um estimador e $G \in \mathcal{P}(\Omega, \mathcal{B})$, neste caso $L(T_n, G)$ denotará a distribuição de T_n sob G .

5.C.2. Robustez Qualitativa.

Será feita agora uma rápida revisão do conceito de robustez qualitativa de um estimador.

No caso de observações independentes e identicamente distribuídas (i.i.d.), Hampel (1971) formulou a seguinte definição:

DEFINIÇÃO 5.7.

Seja $F \in \mathcal{P}(\Omega, \mathcal{B})$, um estimador $(T_n)_{n \geq n_0}$ é robusto em F se a sequência de aplicações:

$$\left(\mathcal{P}(\Omega, \mathcal{B}), \Pi_d \right) \longrightarrow \left(\mathcal{P}(\Lambda, \lambda), \Pi_\lambda \right)$$

definidas por $G \longrightarrow L(T_n, G)$, $n \geq n_0$, é equicontínua em F .

Equivalentemente:

$(T_n)_{n \geq n_0}$ é robusto em F se $\forall \varepsilon > 0 \exists \delta > 0$ tal que

$$\forall n \geq n_0, \forall Q \in \mathcal{P}(\Omega, \mathcal{B}): \Pi_d(F, G) < \delta \Rightarrow \Pi_\lambda \left(L(T_n, F), L(T_n, G) \right) < \varepsilon.$$

OBSERVAÇÃO 5.8.

No contexto da definição anterior, dado que as observações são i.i.d., é mais correto denotar a distribuição de T_n sob G por $L(T_n, G^n)$ em vez de $L(T_n, G)$. Assim sendo, ficará mais clara a definição seguinte de Π -robustez, também devida a Hampel (1971):

DEFINIÇÃO 5.9.

$(T_n)_{n \geq n_0}$ é Π -robusto em F se $\forall \varepsilon > 0 \exists \delta > 0$ tal que

$$\forall n \geq n_0, \forall Q_n \in \mathcal{P}(\Omega^n, \mathcal{B}^n): \Pi_{H_n}(F^n, Q_n') < \delta \Rightarrow \Pi_\lambda \left(L(T_n, F^n), L(T_n, Q_n) \right) < \varepsilon,$$

onde Q_n' a medida invariante por permutações das coordenadas induzida por Q_n em $(\Omega^n, \mathcal{B}^n)$.

OBSERVAÇÃO 5.10.

Para se trabalhar no contexto de séries de tempo e/ou processos estocásticos é de interesse a generalização do conceito de Π -robustez, cuja versão original já admite leves afastamentos da hipótese i.i.d..

É evidente que o conceito de Π -robustez admite várias generalizações para o caso de processos em geral. A seguir, apresenta-se a definição de Bustos (1981):

DEFINIÇÃO 5.11.

Sejam $\mu \in \mathcal{P}(\Omega^\infty, \mathcal{B}^\infty)$ com marginais $\mu_n = \mu/\Omega^n$ $n \geq n_0$, e ρ_n pseudométricas em $\mathcal{P}(\Omega^n, \mathcal{B}^n)$, $n \geq n_0$. Então $(T_n)_{n \geq n_0}$ é ρ_n -robusto em μ se e

dado $\varepsilon > 0$, $\exists \delta > 0$ tal que

$$\forall n \geq n_0, \forall \nu_n \in \mathcal{P}(\Omega^n, \mathcal{B}^n): \rho_n(\nu_n, \mu_n) < \delta \Rightarrow \Pi \lambda \left(L(T_n, \mu_n), L(T_n, \nu_n) \right) < \varepsilon.$$

No caso i.i.d., tomando $\rho_n(\mu_n, \nu_n) = \Pi H_n(\mu_n', \nu_n')$, obtém-se a definição de Hampel.

5.C.3. Resistência qualitativa.

A definição de resistência qualitativa está baseada no seguinte conceito de resistência (Mosteller e Tukey, 1977): pretende-se que o estimador varie pouco na presença dos dois seguintes tipos de modificações da amostra:

i) erros de arredondamento ou agrupamento em todas as observações, e /ou

ii) substituição de uma pequena fração da amostra por outliers arbitrários.

No trabalho já citado de Hampel (1971) esta idéia aparece sob o nome de "*condition B*".

DEFINIÇÃO 5.12.

$F \in \mathcal{P}(\Omega, \mathcal{B})$, $(T_n)_{n \geq n_0}$ estimador. Então F e $(T_n)_{n \geq n_0}$ satisfazem a "*condition B*" se

$$\forall \epsilon > 0 \quad \forall \eta > 0 \quad \exists \delta > 0 \quad \forall n \geq n_0 \quad \exists \mathcal{E}_n \subseteq \mathcal{F}_n:$$

$$\{F(\mathcal{E}_n) > 1 - \eta\} \wedge \{F_n \in \mathcal{E}_n \wedge G_n \in \mathcal{F}_n \wedge \Pi_d(F_n, G_n) < \delta\} \Rightarrow \lambda(T_n(F_n), T_n(G_n)) < \epsilon,$$

$$\text{onde } \mathcal{F}_n = \left\{ \begin{array}{l} \text{medidas discretas de probabilidade sobre } (\Omega, \mathcal{B}) \\ \text{cujos átomos tem massas múltiplos de } 1/n \end{array} \right\}$$

(veja que $\Pi_d/\mathcal{F}_n = H_n$).

OBSERVAÇÕES 5.13.

Esta definição preserva as intenções de Mosteller e Tukey, dado que Π_d absorve os dois tipos de desvio mencionados acima (veja observação 5.26, no apêndice). Entretanto é preciso salientar três aspectos, que ajudam a entender a evolução destas idéias em Boente, Fraiman e Yohay (1987):

a) nesta definição a idéia de resistência ocorre uniformemente no tamanho da amostra, sendo esta uma característica das definições de Hampel de robustez e Π -robustez, decorrente das propriedades dos estimadores induzidos por funcionais T definidos em $\mathcal{P}(\Omega, \mathcal{B})$ (isto é $T_n = T/\mathcal{F}_n$).

b) A introdução na definição dos conjuntos $\mathcal{E}_n$, de probabilidade uniformemente grande sob F , a torna menos restritiva; no trabalho de Hampel, de 1971, os $\mathcal{E}_n$ concretizam-se como $\{ F_n \in \mathcal{F}_n : \Pi_d(F, F_n) < \alpha \}$ para um certo $\alpha > 0$, isto é, como vizinhanças de F na topologia relativa dos $\mathcal{F}_n$ (que é onde interessa, na prática, que a propriedade funcione).

c) Hampel (1971, lema 4.1) provou que para processos i.i.d. "condition B" implica Π -robustez, sendo que uma parte importante da demonstração baseia-se no seguinte fato:

$$F \in \mathcal{P}(\Omega, \mathcal{B}), G \in \mathcal{P}(\Omega, \mathcal{B}), \epsilon > 0, \Pi_d(F, G) < \epsilon,$$

então existe $Q \in \mathcal{P}(\Omega \times \Omega, \mathcal{B} \times \mathcal{B})$ com marginais F e G tal que:

se $\{(x_j, y_j) : 1 \leq j \leq n\}$ é amostra i. i. d. de G então

$$Q \left\{ H_n(x^n, y^n) < \epsilon^{1/2} \right\} \geq 1 - \epsilon^{1/2},$$

é importante salientar que este resultado é ainda válido em uma forma mais forte, substituindo H_n por d_n (veja lema (5.27) no apêndice) e que Boente, Fraiman e Yohai (1987, teorema 4.2) utilizam argumentos semelhantes na sua demonstração de que resistência fraca implica π -robustez.

As seguintes definições são de Boente, Fraiman e Yohai (1987).

DEFINIÇÃO 5.14.

$$x^n \in \Omega^n, \delta > 0,$$

$$S_n(\delta, x^n) = \supremo \left\{ \lambda(T_n(y^n), T_n(z^n)) : \begin{array}{l} y^n \in V(x^n, \delta, d_n) \\ z^n \in V(x^n, \delta, d_n) \end{array} \right\}$$

ou também

$$S_n(\delta, x^n) = \text{diâmetro} \left(T_n(V(x^n, \delta, d_n)) \right) ;$$

DEFINIÇÃO 5.15.

$\mu \in \mathcal{P}(\Omega^\infty, \mathcal{B}^\infty)$, $(T_n)_{n \geq n_0}$ é fracamente resistente em μ se e

$$\forall n \geq n_0 \quad \mu \left(\left\{ x^n \in \Omega^n : S_n(\delta, x^n) \leq \varepsilon \right\} \right) \geq 1 - \varepsilon ;$$

DEFINIÇÃO 5.16.

$\mu \in \mathcal{P}(\Omega^\infty, \mathcal{B}^\infty)$, $(T_n)_{n \geq n_0}$ é assintoticamente fracamente resistente em μ

se e $\forall \varepsilon > 0 \exists \delta > 0$ e $n_1 = n_1(\varepsilon) \geq n_0$ tais que:

$$\forall n \geq n_1 \quad \mu_n \left(\left\{ x^n \in \Omega^n : S_n(\delta, x^n) \leq \varepsilon \right\} \right) \geq 1 - \varepsilon ;$$

DEFINIÇÃO 5.17.

$\tilde{x} \in \Omega^\infty$, $(T_n)_{n \geq n_0}$ é resistente em $\tilde{x}$ se e $\forall \varepsilon > 0 \exists \delta > 0$ tal que
 $\forall n \geq n_0 : S_n(\delta, x^{(n)}) \leq \varepsilon$.

DEFINIÇÃO 5.18.

$\mu \in \mathcal{P}(\Omega^\infty, \mathcal{B}^\infty)$, $(T_n)_{n \geq n_0}$ é fortemente resistente em μ se e

$$\mu\left(\left\{x \in \Omega^\infty : T_n \text{ é resistente em } x\right\}\right) = 1 ;$$

ou equivalentemente (Ver Boente, Fraiman e Yohai, 1987)

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ tal que } \mu\left(\bigcap_{n=n_0}^{\infty} \left\{x \in \Omega^\infty : S_n(\delta, x^n) \leq \varepsilon\right\}\right) \geq 1 - \varepsilon ;$$

OBSERVAÇÕES 5.19.

a) a definição de resistência fraca está claramente relacionada com a "condition B" de Hampel;

b) resistência forte $\Rightarrow$ resistência fraca (observe a segunda caracterização de resistência forte).

c) Boente, Fraiman e Yohai (1987, teorema 4.2) provaram as seguintes relações entre os conceitos de resistência fraca e robustez qualitativa:

Seja $\mu \in \mathcal{P}(\Omega^\infty, \mathcal{B}^\infty)$

(i) $(T_n)_{n \geq n_0}$ fracamente resistente em μ

$\Rightarrow (T_n)_{n \geq n_0}$ Π -robusto em μ ;

(ii) $(T_n)_{n \geq n_0}$ Π -robusto e consistente em μ

$\Rightarrow (T_n)_{n \geq n_0}$ assintoticamente fracamente resistente em μ .

OBSERVAÇÃO 5.20.

Robustez, Π -robustez ou resistência qualitativas não implicam consistência. Por exemplo, qualquer estimador que seja constante para cada tamanho de amostra ($T_n = \alpha_n \forall n \in \mathbb{N}$, onde $(\alpha_n)_{n \in \mathbb{N}}$ é qualquer sequência de

números reais) é qualitativamente Π -robusto (portanto robusto) e resistente. Entretanto a consistência de T_n implicaria na convergência da seqüência $(\alpha_n)_{n \in \mathbb{N}}$. Por exemplo, Hampel (1971) introduz a "*continuity condition*" para garantir consistência:

$(T_n)_{n \geq n_0}$ é contínuo em F se e

$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \exists n_1$ tais que $\forall n, m \geq n_1 \quad \forall F_n \in \mathcal{F}_n, F_m \in \mathcal{F}_m$:

$$\Pi(F_n, F) < \delta \wedge \Pi(F_m, F) < \delta \Rightarrow \lambda(T_n(F_n), T_m(F_m)) < \varepsilon.$$

É importante também salientar que estimadores qualitativamente π -robustos e/ou resistentes não têm o menor compromisso com eficiência (o estimador T_n acima definido jamais seria eficiente em condições não triviais).

5.D. EQUIVALÊNCIA ENTRE RESISTÊNCIA FRACA E FORTE.

A seguinte proposição é necessária para a demonstração da equivalência entre resistência fraca e forte nas condições do teorema 5.22. Ela está incluída na demonstração do teorema 4.1 de Boente, Fraiman e Yohai (1987).

PROPOSIÇÃO 5.21.

Seja (Ω, d) espaço métrico separável e completo. $\delta^* > 0$ e seja d_n a métrica em Ω^n definida em 5.B.

a) $\exists K \subseteq \Omega$, K compacto tal que se $K_0 = K^c$, $\mu_1(K_0) < \delta^*/8$ e existe $\{K_1, K_2, \dots, K_h\}$ partição de K com $\text{diâmetro}(K_i) < \delta^*/2$, $1 \leq i \leq h$,

b) dado $x^n = (x_1, x_2, \dots, x_n) \in \Omega^n$ definem-se $\text{Min}(x^n)$ e R_n por:

$$(1) \quad \text{Min}(x^n) = n^{-1} \cdot \#\{j: x_j \in K_i\}, \quad 0 \leq i \leq h,$$

$$(2) \quad R_n = \bigcap_{i=0}^h \left\{ x^n: \left| \text{Min}(x^n) - \mu_1(K_i) \right| \leq \delta^*/8h \right\};$$

dado $u^n = (u_1, u_2, \dots, u_n) \in R_n$, sejam $[1, n] = \{j \text{ natural}: 1 \leq j \leq n\}$ e $P(u^n)$ definido por:

$P(u^n) = \{x^n = (u_{p(1)}, u_{p(2)}, \dots, u_{p(n)}) : p \text{ permutação de } [1, n]\}$, então:

$$(3) \quad R_n \subseteq \mathcal{V}(P(u^n), \delta^*/2, d_n).$$

Demonstração:

a) se $\delta^* > 0$, a existência do compacto K tal que $\mu_1(K) > 1 - \delta^*/8$ é uma conhecida propriedade dos espaços métricos separáveis e completos denominada "tightness", cuja demonstração encontra-se, por exemplo, em Billingsley (1968);

por outro lado para qualquer conjunto K vale que

$$K \subseteq \bigcup_{\omega \in K} \mathcal{V}(\omega, \delta^*/4, d);$$

neste caso, sendo K compacto, existem h natural e $\omega_i \in K$ ($1 \leq i \leq h$) tais

$$\text{que: } K \subseteq \bigcup_{i=1}^h \mathcal{V}(\omega_i, \delta^*/4, d),$$

então os conjuntos K_i definidos por $K_i = K \cap \mathcal{V}(\omega_i, \delta^*/4, d)$, $1 \leq i \leq h$,

satisfazem $K = \sum_{i=1}^h K_i$ e $\text{diâmetro}(K_i) < \delta^*/2 \quad \forall 1 \leq i \leq h$.

b) seja $u^n \in R_n$ e $y^n \in R_n$ (portanto $|\text{Min}(u^n) - \text{Min}(y^n)| \leq \delta^*/4h$),

sejam Q_i e Q'_i definidos por

$$Q_i = \{j: y_j \in K_i\} \text{ e } Q'_i = \{j: u_j \in K_i\}, \quad 0 \leq i \leq h,$$

então existe uma partição $\{Q_0^*, Q_1^*, \dots, Q_h^*\}$ do intervalo $[1, n]$ de tal forma que

$$(4) \quad \#Q_i^* = n \cdot \text{Min}(u^n) = \#Q'_i, \quad 0 \leq i \leq h,$$

$$(5) \quad \text{e } \#(Q_i^* \Delta Q_i) \leq n \cdot \delta^*/4h, \quad 0 \leq i \leq h,$$

(veja o lema (5.29) no apêndice);

por (4) existe $v^n \in P(u^n)$ tal que $j \in Q_i^* \Rightarrow v_j \in K_i, \quad 0 \leq i \leq h$,

dado que se $1 \leq i \leq h$, $\text{diâmetro}(K_i) < \delta^*/2$, tem-se que

$$\{i: d(v_i, y_i) \geq \delta^*/2\} \subseteq \left(\bigcup_{i=1}^h (Q_i^* \Delta Q_i) \right) \cup Q_0^*;$$

finalmente: dados (1), (2) e $\mu_1(K^c) < \delta^*/8$, (4) e (5) implicam que:

$$\#\{i: d(v_i, y_i) \geq \delta^*/2\} \leq n \cdot \delta^*/4 + \#Q_0^* \leq 2 \cdot n \cdot \delta^*/4 = n \cdot \delta^*/2,$$

portanto $d_n(v^n, y^n) \leq \delta^*/2$ e então $R_n \subseteq V(P(u^n), \delta^*/2, d_n)$. $\square$

Boente, Fraiman e Yohai (1987, teorema 4.1) provaram que resistência fraca e forte são equivalentes para estimadores invariantes sob permutações no caso de processos i.i.d.. O teorema seguinte generaliza o

citado resultado para processos estacionários e ergódicos. A diferença básica na argumentação é a substituição da lei forte dos grandes números, utilizada em Boente, Fraiman e Yohai, pelo teorema ergódico.

TEOREMA 5.22.

Sejam μ uma probabilidade em $(\Omega^\infty, \mathcal{B}^\infty)$ correspondente a um processo estacionário ergódico e $(T_n)_{n \geq n_0}$ um estimador invariante por permutações das coordenadas e fracamente resistente em μ .

Então $(T_n)_{n \geq n_0}$ é também fortemente resistente.

Demonstração:

Seja $\varepsilon > 0$, será provado que

$$(6) \quad \exists \delta > 0 \text{ tal que } \forall n \geq n_0 \quad \mu \left(\bigcap_{n \geq n_0} \left\{ x \in \Omega^\infty : S_n(\delta, x^{(n)}) \leq \varepsilon \right\} \right) \geq 1 - \varepsilon.$$

Sendo $(T_n)_{n \geq n_0}$ fracamente resistente em μ , $\exists \delta^* > 0$ tal que

$$(7) \quad \forall n \geq n_0 \text{ vale } \mu_n \left(\left\{ x^n : S_n(\delta^*, x^n) \leq \varepsilon \right\} \right) \geq 1 - \varepsilon.$$

Considerando-se $\text{Min}(x^n)$ e R_n definidos em (1) e (2). O teorema ergódico afirma que

$\lim_{n \rightarrow \infty} \text{Min}(x^{(n)}) \rightarrow \mu_1(K_i)$ μ -quase certamente, $0 \leq i \leq h$, isto é

$$(8) \quad \mu \left(\bigcap_{i=0}^h \left\{ x : \text{Min}(x^{(n)}) \rightarrow \mu_1(K_i) \right\} \right) = 1,$$

o que implica que dado o ε de (6) $\exists n_1 \geq n_0$ tal que

$$(9) \quad \mu_n(R_n) = \mu \left(\bigcap_{i=0}^h \left\{ x : |\text{Min}(x^{(n)}) - \mu_1(K_i)| \leq \delta^*/8h \right\} \right) \geq 1 - \varepsilon \quad \forall n \geq n_1;$$

dados (6) e (9) (e supondo $\varepsilon \leq 1/2$):

$\forall n \geq n_1$ sempre existe $u^n \in R_n \cap \left\{ x^n : S_n(\delta^*, x^n) \leq \varepsilon \right\}$, o que implica que

$$(10) \quad P(u^n) \subseteq \left\{ x^n : S_n(\delta^*, x^n) \leq \varepsilon \right\}, \quad n \geq n_1,$$

pois $S_n(\delta, x^n)$ é também invariante por permutações das coordenadas (dado que $(T_n)_{n \in \mathbb{N}}$ é);

portanto, dados (3) e (10) tem-se que

$$(11) \quad R_n \subseteq V(P(u^n), \delta^*/2, d_n) \subseteq V \left(\left\{ x^n : S_n(\delta^*, x^n) \leq \varepsilon \right\}, \delta^*/2, d_n \right),$$

e sendo

$$(12) \quad V \left(\left\{ x^n : S_n(\delta^*, x^n) \leq \varepsilon \right\}, \delta^*/2, d_n \right) \subseteq \left\{ x^n : S_n(\delta^*/2, x^n) \leq \varepsilon \right\},$$

(11 e (12) implicam que

$$(13) \quad R_n = \bigcap_{i=0}^h \left\{ x^n : |\text{Min}(x^n) - \mu_1(K_i)| \leq \delta^*/8h \right\} \subseteq \left\{ x^n : S_n(\delta^*/2, x^n) \leq \varepsilon \right\}$$

(se $n \geq n_1$).

Além disso, sabe-se que (veja lema (5.31) no apêndice) se $(Z_n)_{n \geq 1}$ e Z são variáveis aleatórias então:

$$Z_n \xrightarrow{P} Z \quad \text{P-quase certamente} \quad \text{see} \quad \forall \alpha > 0 \quad P(\liminf \{|Z_n - Z| \leq \alpha\}) = 1,$$

ou equivalentemente see

$$(14) \quad \forall \alpha > 0 \quad \lim_{m \rightarrow \infty} P \left(\bigcap_{n=m}^{\infty} \left\{ |Z_n - Z| \leq \alpha \right\} \right) = 1;$$

portanto, (8), que decorre do teorema ergódico, equivale a afirmar

$$(15) \quad \forall \alpha > 0 \quad \lim_{m \rightarrow \infty} \mu \left(\bigcap_{n=m}^{\infty} \bigcap_{i=0}^h \left\{ x: |\text{Min}(x^{(n)}) - \mu_1(K_i)| \leq \alpha \right\} \right) = 1.$$

Reconsideremos R_n : (13) implica que, se $n \geq n_1$

$$(16) \quad \bigcap_{i=0}^h \left\{ x: |\text{Min}(x^{(n)}) - \mu_1(K_i)| \leq \delta^*/8h \right\} \subseteq \left\{ x: S_n(\delta^*/2, x^{(n)}) \leq \varepsilon \right\},$$

e em consequência,

$$(17) \quad \text{se } m \geq n_1 \quad \bigcap_{n=m}^{\infty} \bigcap_{i=0}^h \left\{ x: |\text{Min}(x^{(n)}) - \mu_1(K_i)| \leq \delta^*/8h \right\} \subseteq$$

$$\subseteq \bigcap_{n=m}^{\infty} \left\{ x: S_n(\delta^*/2, x^{(n)}) \leq \varepsilon \right\};$$

portanto, dados (15) e (17), tem-se que

$$\lim_{m \rightarrow \infty} \mu \left(\bigcap_{n=m}^{\infty} \left\{ x: S_n(\delta^*/2, x^{(n)}) \leq \varepsilon \right\} \right) = 1,$$

então $\exists n_2$ natural tal que

$$(18) \quad \forall m \geq n_2 \quad \mu \left(\bigcap_{n=m}^{\infty} \left\{ x: S_n(\delta^*/2, x^{(n)}) \leq \varepsilon \right\} \right) \geq 1 - \varepsilon/2.$$

Seja agora $n_3 = \text{máximo}\{n_1, n_2\}$. Pela hipótese (7) (que diz que $(T_n)_{n \in \mathbb{N}}$ é fracamente resistente), $\exists \delta \leq \delta^*/2$ tal que:

$$(19) \quad \forall n_0 \leq n \leq n_3 \text{ vale que: } \mu \left(\left\{ x: S_n(\delta, x^{(n)}) \leq \varepsilon \right\} \right) \geq 1 - \varepsilon/2(n_3 - n_0).$$

Finalmente, a resistência forte (condição (6)) pode ser deduzida de (18) e (19). $\square$

5.E. APÊNDICE.

NOTAÇÃO 5.23.

Seja Z o conjunto dos números inteiros, $a \in Z$, $b \in Z$, $a \leq b$. Denotaremos por $[a, b] = \{k \in Z : a \leq k \leq b\}$.

LEMA 5.24.

(Ω, d) espaço métrico, s_n e d_n métricas definidas em Ω^n por:

$$s_n(x^n, y^n) = \sup \{d(x_j, y_j) : 1 \leq j \leq n\} \quad e$$

$$d_n(x^n, y^n) = \inf \left\{ \varepsilon : \#\{i : d(x_i, y_i) \geq \varepsilon\} \leq n\varepsilon \right\},$$

onde $x^n = (x_1, x_2, \dots, x_n)$ e $y^n = (y_1, y_2, \dots, y_n)$.

Então:

$$a) \quad d_n(x^n, y^n) \leq s_n(x^n, y^n) \quad \forall (x^n, y^n) \in \Omega^n \times \Omega^n ;$$

b) $d_n(x^n, y^n) < n^{-1} \Leftrightarrow s_n(x^n, y^n) < n^{-1}$, além disso, neste caso, vale que $d_n(x^n, y^n) = s_n(x^n, y^n)$ (portanto s_n e d_n definem a mesma topologia);

$$c) \quad \forall \varepsilon > 0 \quad \forall x^n \in \Omega^n \quad V(x^n, \varepsilon, s_n) \subseteq V(x^n, \varepsilon, d_n \text{ (por a) }), \quad e$$

$$\forall \varepsilon < n^{-1} \forall x^n \in \Omega^n \quad V(x^n, \varepsilon, s_n) = V(x^n, \varepsilon, d_n) \quad (\text{por b}) .$$

Demonstração:

c) é trivial, só demonstraremos a) e b).

Prova de a): seja $s_n(x^n, y^n) = \alpha$ e seja $\varepsilon > \alpha$,

$$\text{então } \# \{i : d(x_i, y_i) \geq \varepsilon\} = 0 \leq n\varepsilon \Rightarrow d_n(x^n, y^n) \leq \alpha .$$

Prova de b): será feita em dois passos,

$$b1) \quad d_n(x^n, y^n) < n^{-1} \Rightarrow s_n(x^n, y^n) = d_n(x^n, y^n) \text{ e}$$

$$b2) \quad s_n(x^n, y^n) < n^{-1} \Rightarrow d_n(x^n, y^n) = s_n(x^n, y^n) ;$$

$$\text{se } d_n(x^n, y^n) < n^{-1} \text{ seja } \alpha \text{ tal que } d_n(x^n, y^n) < \alpha < n^{-1},$$

$$\therefore \# \{i : d(x_i, y_i) \geq \alpha\} \leq n\alpha < 1 \Rightarrow \# \{i : d(x_i, y_i) \geq \alpha\} = 0 \Rightarrow s_n(x^n, y^n) < \alpha,$$

$$\text{portanto } s_n(x^n, y^n) \leq d_n(x^n, y^n) \Rightarrow s_n(x^n, y^n) = d_n(x^n, y^n) \quad (\text{por a}) ;$$

$$\text{se } s_n(x^n, y^n) < n^{-1} \text{ então (por a) } d_n(x^n, y^n) < n^{-1} \text{ e volta-se ao caso}$$

acima. $\square$

LEMA 5.25.

Seja d_n como no lema 5.24 (definida em Observação 5.4).

d_n tem as seguintes propriedades:

$$a) \text{ sejam } \alpha > 0, \quad x^n \text{ e } y^n \text{ em } \Omega^n \text{ tais que } d(x_j, y_j) < \alpha \quad \forall j, \quad 1 \leq j \leq n,$$

$$\text{então } d_n(x^n, y^n) \leq \alpha;$$

b) sejam $\alpha > 0$, x^n e y^n em Ω^n , $I \subset [1, n]$ tais que
 $j \in I \Rightarrow d(x_j, y_j) \geq \alpha$, $j \in [1, n] - I \Rightarrow d(x_j, y_j) < \alpha$ e $\#(I) = k$,

então: $d_n(x^n, y^n) \leq \text{máximo}(\alpha, k/n)$.

Demonstração:

a) $\#\{j: d(x_j, y_j) \geq \alpha\} = 0 < n\alpha$, então $d_n(x^n, y^n) \leq \alpha$;

b) seja $\beta = \text{máximo}(\alpha, k/n)$,

se $\alpha > k/n \Rightarrow \beta = \alpha$ e $\#\{i: d(x_i, y_i) \geq \alpha\} = k < n\alpha$,

então $d_n(x_n, y_n) \leq \alpha$,

se $\alpha \leq k/n \Rightarrow \beta = k/n$ e $\#\{i: d(x_i, y_i) \geq k/n\} \leq k = n(k/n)$,

então $d_n(x_n, y_n) \leq k/n$. $\square$

OBSERVAÇÃO 5.26.

As duas propriedades acima garantem que d_n absorve os "erros de arredondamento ou agrupamento" (que na prática ocorrem com frequência relativa próxima a 1) e os "erros grosseiros" e *outliers* (idem com frequência relativa próxima de 0).

Dado que $H_n \leq d_n$, H_n obedece às duas propriedades acima.

LEMA 5.27.

(Ω, d) espaço métrico separável e completo,

$\mathcal{B} = \{ \text{borelianos de } (\Omega, d) \}$, $F \in \mathcal{P}(\Omega, \mathcal{B})$, $G \in \mathcal{P}(\Omega, \mathcal{B})$,

$\Pi d(F,G) < \epsilon$. Nestas condições, $\exists Q \in \mathcal{P}(\Omega \times \Omega, \mathcal{B} \times \mathcal{B})$ tal que:

F e G são as marginais de Q ,

$\{(x_j, y_j) : 1 \leq j \leq n\}$ amostra i. i. d. de Q, então:

$$(20) \quad Q\left\{\left\{d_n(x^n, y^n) \leq \epsilon^{1/2}\right\}\right\} \geq \epsilon^{1/2} \text{ . onde } x^n, y^n \text{ e } d_n \text{ são os mesmos}$$

do lema 5.24.

Demonstração:

Um resultado de Strassen (1965) diz que:

$\Pi d(F,G) < \epsilon$ se e só se $\exists Q \in \mathcal{P}(\Omega \times \Omega, \mathcal{B} \times \mathcal{B})$ tal que:

F e G são as marginais de Q e

$Q(D(\epsilon)) < \epsilon$, onde $D(\epsilon) = \{(\omega_1, \omega_2) : d(\omega_1, \omega_2) \geq \epsilon\}$,

portanto só falta demonstrar (20);

a) se $\{(x_j, y_j) : 1 \leq j \leq n\}$ amostra i. i. d. de Q vale que

$$Q\left\{\left\{n^{-1} \# \{i : d(x_i, y_i) \geq \epsilon\} \geq e^{1/2}\right\}\right\} < \epsilon^{1/2} \text{ .}$$

já que, dada a desigualdade de Tchebycheff

$(X \geq 0, \alpha > 0, P \text{ probabilidade, } \int X dP < \infty \Rightarrow P(\{X \geq \alpha\}) \leq \alpha^{-1} \int X dP)$, e que

$\# \{i : d(x_i, y_i) \geq \epsilon\}$ tem distribuição binomial(n, Q(D(ε)) ,

tem-se que

$$Q\left\{\left\{\# \{i: d(x_i, y_i) \geq \varepsilon\} \geq n\varepsilon^{1/2}\right\} \leq [nQ(D(\varepsilon))]/[n\varepsilon^{1/2}] < \varepsilon^{1/2};\right.$$

$$b) \text{ se } \varepsilon \leq 1 \Rightarrow n^{-1} \# \{i: d(x_i, y_i) \geq \varepsilon\} \leq \varepsilon^{1/2} \Rightarrow d_n(x^n, y^n) \leq \varepsilon^{1/2};$$

$$c) \text{ finalmente } Q\left\{\left\{d_n(x^n, y^n) \leq \varepsilon^{1/2}\right\}\right\} \geq 1 - \varepsilon^{1/2},$$

$$\text{dado que } Q\left\{\left\{d_n(x^n, y^n) \leq \varepsilon^{1/2}\right\}\right\} = 1 - Q\left\{\left\{d_n(x^n, y^n) > \varepsilon^{1/2}\right\}\right\} \geq$$

$$\underset{(\text{por } b)}{\geq} 1 - Q\left\{\left\{n^{-1} \# \{i: d(x_i, y_i) \geq \varepsilon\} > \varepsilon^{1/2}\right\}\right\} \underset{(\text{por } a)}{\geq} 1 - \varepsilon^{1/2}. \square$$

OBSERVAÇÃO 5.28.

Dado que $h_n \leq d_n$, o lema anterior permanece válido (em uma forma mais fraca) colocando h_n no lugar de d_n .

LEMA 5.29.

Sejam Ω conjunto, $\Omega = \sum_{i=0}^h \Omega_i$, onde $\Omega_i \neq \emptyset \quad \forall i, \quad 0 \leq i \leq h$ (ou seja $\{\Omega_0, \Omega_1, \Omega_2, \dots, \Omega_h\}$ é uma partição de Ω),

$$y^n \in \Omega^n, \quad Q_i = \{j: y_j \in \Omega_i\}, \quad 0 \leq i \leq h,$$

$$u^n \in \Omega^n, \quad Q'_i = \{j: u_j \in \Omega_i\}, \quad 0 \leq i \leq h;$$

$$(\text{é claro que } [1, n] = \sum_{i=0}^h Q_i = \sum_{i=0}^h Q'_i).$$

Então:

a) existem Q_i^* , $0 \leq i \leq h$, de maneira que

$$\forall 0 \leq i \leq h, \#Q_i^* = \#Q'_i, Q_i < Q_i^* \text{ ou } Q_i > Q_i^* \text{ e } [1, n] = \sum_{i=0}^h Q_i^* ;$$

$$b) \#(Q_i \Delta Q_i^*) = |\#Q_i - \#Q_i^*|, \forall 0 \leq i \leq h.$$

Demonstração:

é suficiente provar a), já que b) é consequência das relações de inclusão entre os Q_i e Q_i^* mencionadas em a).

$$[0, h] = \{i \in [0, h] : \#Q_i \geq \#Q'_i\} + \{i \in [0, h] : \#Q_i < \#Q'_i\},$$

se $\{i \in [0, h] : \#Q_i < \#Q'_i\} = \emptyset \Rightarrow \#Q_i = \#Q'_i, 1 \leq i \leq h$ e neste caso $Q_i^* = Q_i$ satisfazem a).

Seja agora $\{i \in [0, h] : \#Q_i < \#Q'_i\} \neq \emptyset$; é possível reordenar os índices $\{0, 1, \dots, h\}$ de tal forma que exista r natural, $0 < r < h$, tal que:

$$\{i \in [0, h] : \#Q_i \geq \#Q'_i\} = [0, r] \text{ e } \{i \in [0, h] : \#Q_i < \#Q'_i\} = [r+1, h],$$

Se $0 \leq i \leq r$ e $Q'_i = \emptyset$, define-se $Q_i^* = \emptyset$;

se $0 \leq i \leq r$ e $Q'_i \neq \emptyset$, seja $f_i: Q'_i \rightarrow Q_i$ qualquer função injetora, neste caso define-se

$$(21) \quad 0 \leq i \leq r, Q_i^* = f_i(Q'_i),$$

sendo claro que: $\forall 0 \leq i \leq r$ tem-se que $\#Q_i^* = \#Q'_i$ e $Q_i < Q_i^*$,

e também que: $\exists! f: \bigcup_{i=0}^r Q'_i \rightarrow \bigcup_{i=0}^r Q_i$ tal que $f|_{Q'_i} = f_i, 1 \leq i \leq r$

(f é injetora).

Será definido agora Q_i^* para $r+1 \leq i \leq h$, os Q_i^* deverão conter os Q_i e, eventualmente, se $\#Q_i' > \#Q_i$, elementos de $\bigcup_{i=1}^r Q_i - f\left(\bigcup_{i=1}^r Q_i'\right)$:

$$\text{seja } t = \sum_{i=r+1}^h \#Q_i' - \sum_{i=r+1}^h \#Q_i = \sum_{i=0}^r \#Q_i - \sum_{i=0}^r \#Q_i' =$$

$$= \sum_{i=0}^r \#Q_i - \sum_{i=0}^r \#Q_i^* = \sum_{i=0}^r \#(Q_i - Q_i^*),$$

$$\text{nestas condições } \exists g: [1, t] \xrightarrow{\text{bijetora}} \sum_{i=0}^r (Q_i - Q_i^*).$$

$$\text{Sejam } \alpha_i = \#Q_i' - \#Q_i, \forall r+1 \leq i \leq h \text{ e } \beta_i = \sum_{j=r+1}^i \alpha_j, \forall r+1 \leq i \leq h,$$

$$\text{sendo evidente que } \beta_h = \sum_{j=r+1}^h \alpha_j = t;$$

agora definem-se os Q_i^* ($r+1 \leq i \leq h$) da seguinte forma:

$$Q_{r+1}^* = Q_{r+1} + g([1, \dots, \beta_{r+1}]),$$

$$Q_{r+2}^* = Q_{r+2} + g([\beta_{r+1}+1, \dots, \beta_{r+2}]),$$

$$(22) \dots\dots\dots$$

$$\dots\dots\dots$$

$$Q_h^* = Q_h + g([\beta_{h-1}, \dots, \beta_h]).$$

Os Q_i^* , $1 \leq i \leq h$, definidos por (5.21) e (5.22) satisfazem as condições

a). $\square$

NOTAÇÃO 5.30.

Ω conjunto não vazio, $N = \{\text{números naturais}\}$, $\{A_n\}_{n \in N} \subset \mathcal{P}(\Omega)$,

$$\limsup A_n = \text{limite superior } \{A_n\}_{n \in N} = \bigcap_{n=1}^{\infty} \left(\bigcup_{m=n}^{\infty} A_m \right),$$

$$\liminf A_n = \text{limite inferior } \{A_n\}_{n \in N} = \bigcup_{n=1}^{\infty} \left(\bigcap_{m=n}^{\infty} A_m \right).$$

LEMA 5.31.

Sejam: $(\Omega, \mathcal{A}, P)$ espaço de probabilidade,

$\{X_n\}_{n \in N}$, X , variáveis aleatórias definidas em $(\Omega, \mathcal{A}, P)$.

$$\text{Então: } P\left(\left\{X_n \xrightarrow[n \rightarrow \infty]{} X\right\}\right) = 0 \quad \text{se } \forall \varepsilon > 0, P\left(\limsup \left\{|X_n - X| > \varepsilon\right\}\right) = 0.$$

Demonstração:

$$\text{Mostraremos agora } (\Leftarrow); \text{ seja então } P\left(\limsup \left\{|X_n - X| > \varepsilon\right\}\right) = 0 \forall \varepsilon > 0;$$

$$\text{Seja } C = \{X_n \xrightarrow[n \rightarrow \infty]{} X\}, \text{ é claro que } C = \bigcap_{\varepsilon > 0} \left(\bigcup_{n=1}^{\infty} \left(\bigcap_{m=n}^{\infty} \left\{|X_m - X| \leq \varepsilon\right\} \right) \right),$$

seja $(\varepsilon_k)_{k \in N}$, uma seqüência de números reais tal que

$$\varepsilon_k \geq \varepsilon_{k+1} > 0 \text{ e } \varepsilon_k \xrightarrow{k \rightarrow \infty} 0,$$

nestas condições:

$$C = \bigcap_{k=1}^{\infty} \left(\bigcup_{n=1}^{\infty} \left(\bigcap_{m=n}^{\infty} \left\{|X_m - X| \leq \varepsilon_k\right\} \right) \right) = \bigcap_{k=1}^{\infty} \left(\liminf \left\{|X_n - X| \leq \varepsilon_k\right\} \right),$$

portanto $C^c = \bigcup_{k=1}^{\infty} \left(\limsup \left\{ |X_n - X| > \varepsilon_k \right\} \right)$,

e sendo a união crescente

$$P(C^c) = \lim_{k \rightarrow \infty} P \left(\limsup \left\{ |X_n - X| > \varepsilon_k \right\} \right) = \sup_k P \left(\limsup \left\{ |X_n - X| > \varepsilon_k \right\} \right) = 0.$$

portanto $P(C) = 1$;

Provemos agora $(\Rightarrow)$, seja então $P(C)=1$;

se $A_n(\varepsilon) = \bigcap_{m=n}^{\infty} \left\{ |X_m - X| \leq \varepsilon \right\}$, $n \in \mathbb{N}$, é claro que $C \subset \bigcup_{n=1}^{\infty} A_n(\varepsilon)$,

portanto, dado que $P(C) = 1$, tem-se que

$$P \left(\bigcup_{n=1}^{\infty} A_n(\varepsilon) \right) = 1, \text{ mas como } \bigcup_{n=1}^{\infty} A_n(\varepsilon) = \liminf \left\{ |X_n - X| \leq \varepsilon \right\},$$

pode-se concluir que $P \left(\limsup \left\{ |X_n - X| > \varepsilon \right\} \right) = P \left(\bigcup_{n=1}^{\infty} A_n(\varepsilon) \right)^c = 0. \square$

OBSERVAÇÃO 5.32.

O lema 5.31 também pode ser enunciado assim:

$X_n \rightarrow X$ com probabilidade 1

$$\text{see } \lim_{n \rightarrow \infty} P \left(\bigcap_{m=n}^{\infty} |X_m - X| \leq \varepsilon \right) = 1$$

$$\text{see } \lim_{n \rightarrow \infty} P \left(\bigcup_{m=n}^{\infty} |X_m - X| > \varepsilon \right) = 0.$$