Universidade de São Paulo
Instituto de Física

Tópicos em Mecânica Estatística de Sistemas Complexos

Uma abordagem mecânico-estatística de dois tópicos de interesse em finanças, economia e sociologia.

Rafael S. Calsaverini

São Paulo
2013

Universidade de São Paulo - Instituto de Física

Tópicos em Mecânica Estatística de Sistemas Complexos

Uma abordagem mecânico-estatística de dois tópicos de interesse em finanças, economia e sociologia.

Rafael S. Calsaverini

Tese de doutorado apresentada ao Instituto de Física para a obtenção do título de Doutor em Ciências.

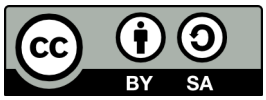
Orientador:
Prof. Dr. Nestor Felipe Caticha Alfonso

Co-Orientador:
Prof. Dr. Renato Vicente

São Paulo
2013

Este trabalho é licenciado pelo autor segundo uma licença Creative Commons 3.0, CC-by-sa, descrita nessa página: http://creativecommons.org/licenses/by-sa/3.0/br/. Isso significa que você está livre para usar este trabalho, criar trabalhos derivados, e compartilhar os resultados com quem quiser, desde que atribua corretamente os créditos ao autor e compartilhe qualquer trabalho derivado usando a mesma licença.

Ciência boa é compartilhada de forma livre, use Creative Commons!!!



Dedicatória

Dedico essa tese à minha filha Sônia Braga Calsaverini. Houve um tempo em que eu achei que esta tese seria a maior realização da minha vida por muitos anos. Não é. É absolutamente insignificante em comparação ao senso de realização que sinto quando vejo você.

Agradecimentos

Agradeço à minha esposa Ana Paola pela paciência para suportar minhas angústias, pela presença constante e apoio e por me proporcionar a maior alegria do universo, que é a minha filha. Agradeço à minha mãe Lourdes e ao meu pai José Arimatea por mais do que a simples existência, por uma existência plena de realizações que só foi possível por causa de vocês. Agradeço à minha sogra, Maria de Lourdes, por ter sido minha quase mãe nesses últimos 6 anos e por ter suportado toda a aporrinhação de ter um genro em casa por tanto tempo.

Agradeço aos professores Nestor Caticha e Renato Vicente pelos cinco anos de paciência e por proporcionar em nosso grupo de pesquisa uma ambiente intelectual que acredito ser ímpar no Instituto de Física e raro em toda a Universidade de São Paulo. Não tenho dúvida de que esse ambiente era o lugar certo para o meu doutorado e que a orientação desses dois professores foram peça chave da minha maturação intelectual.

Agradeço a todo resto da minha família: irmãos, irmãs, madrastas, avôs e avós, a todos os que estão aqui e que não estão mais. Nunca deixo de pensar em vocês.

Agradeço aos meus amigos Caio, Camila, Henrique e Andrea, que tão próximos estiveram e tanto fizeram por mim e minha pequena família nos últimos anos. Vocês fizeram por mim muito mais do que eu posso retribuir, oferecendo ajuda mesmo sem eu pedir.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) pelo auxílio financeiro, e à Universidade de São Paulo por ter sido minha casa pelos últimos 13 anos.

I've heard there was a secret chord
That David played, and it pleased the Lord
But you don't really care for music, do yah?
It goes like this:
The fourth, the fifth,
The minor fall, the major lift,
The baffled king composing Hallelujah.

Hallelujah, Hallelujah, Hallelujah, Hallelujah, ...

Your faith was strong but you needed proof.
You saw her bathing on the roof.
Her beauty in the moonlight overthrew yah.
She tied you to a kitchen chair,
She broke your throne, and she cut your hair,
And from your lips she drew the Hallelujah.

Hallelujah, Hallelujah, Hallelujah, Hallelujah

There was a time when you let me know
What's really going on below,
But now you never show it to me, do yah?
And remember when I moved in you
The holy dove was moving too
And every breath we drew was Hallelujah

Hallelujah, Hallelujah, Hallelujah, Hallelujah

Maybe there's a God above
But all I've ever learned from love
Was how to shoot at someone who outdrew yah.
It's not a cry you can hear at night,
It's not somebody who has seen the light,
It's a cold and it's a broken Hallelujah.

Hallelujah, Hallelujah, Hallelujah, Hallelujah

You say I took the name in vain.
I don't even know the name.
But if I did, well, really, what's it to yah?
There's a blaze of light in every word,
It doesn't matter which you heard:
The holy or the broken Hallelujah

Hallelujah, Hallelujah, Hallelujah, Hallelujah

I did my best, it wasn't much.
I couldn't feel, so I tried to touch.
I've told the truth, I didn't come to fool yah.
And even though it all went wrong
I'll stand before the Lord of Song,
With nothing on my tongue but Hallelujah

Hallelujah, Hallelujah, Hallelujah, Hallelujah...

—Hallelujah, música de Leonard Cohen

Resumo

No presente trabalho, exploramos dois temas de interesse em finanças, economia e antropologia social, através da aplicação de técnicas da teoria da informação e da mecânica estatística.

No primeiro tópico, estudamos a conexão entre teoria de dependência estatística, teoria de informação e teoria da cópulas. O conceito de distribuição cópula é revisto e aplicado em reformulação das definições de medida de dependência dadas por Rényi [1]. Em seguida, mostramos que a informação mútua satisfaz todos os requisitos para ser uma boa medida de dependência. Obtemos uma identidade entre a informação mútua e a entropia da distribuição cópula e uma decomposição mais específica da informação mútua de uma distribuição elíptica nas suas partes linear e não-linear. Avaliamos o risco de usar quantidades ingênuas como medidas de dependência estatística, mostrando que a correlação linear pode subestimar grosseiramente a dependência. Esses resultados são utilizados para desenvolver um método de detectação de desvios de dependência gaussiana em pares de variáveis aleatórias e aplicá-lo a séries temporais financeiras. Finalmente, discutimos um método para ajustar t-cópulas a dados empíricos [2] através da medida da informação mútua e do tau de Kendall.

No segundo tópico, desenvolvemos um modelo para o surgimento de autoridade em sociedades humanas pré-agrícolas. Discutimos motivações empíricas com raízes em neurociência, primatologia e antropologia para um modelo matemático capaz de explicar a ampla variabilidade de formas de organização social humana no eixo igualitário-hierárquico. O modelo resulta da aplicação de teoria da informação a uma hipótese sobre os custos evolutivos envolvidos na vida social. O modelo gera um diagrama de fase rico, com diferentes regimes que podem ser interpretados como diferentes tipos de organização social. Os parâmetros de controler do modelo estão ligados à capacidade cognitiva da espécie em questão, ao tamanho do grupo e a pressões ecológicas e sociais.

[1] A. Renyi. On measures of dependence. *Acta. Math. Acad. Sci. Hungar.*, 10:441–451, 1959

[2] R. S. Calsaverini and R. Vicente. An information-theoretic approach to statistical dependence: Copula information. *EPL (Europhysics Letters)*, 88(6): 68003, 2009. URL `http://stacks.iop.org/0295-5075/88/i=6/a=68003`

Abstract

In the present work we explore two topics of interest in finance, economics and social anthropology through the application of techniques from information theory and statistical mechanics.

In the first topic we study the connexion between statistical dependency theory, information theory and copula theory. The concept of copula distribution is reviewed and applied to the reformulation of the definition of dependency measures given by Rényi [3]. It is then shown that the mutual information satisfy all the requirements to be a good dependency measure. We derive an identity between mutual information and the entropy of the copula distribution and a more specific decomposition of the mutual information of an elliptical distribution into its linear and non-linear parts. We evaluate the risk of using naive measures as statistical dependency measures by showing that linear correlation can grossly underestimate dependency. Those results are used to develop a method to detect deviation from gaussian dependence in pairs of random variables and apply it to financial time series. Finally, we discuss a method to adjust t-copulas to empirical data[4] through the determination of the mutual information and Kendall's tau.

In the second topic we develop a model for the emergence of authority in pre-agricultural human societies. We discuss empirical motivations with roots in neuroscience, primatology and anthropology for a mathematical model able to explain the ample variability of forms of human social organization in the egalitarian-hierarchical axis. The model results from the application of information theory on a hypothesis about the evolutive costs involved in social life. It generates a rich phase diagram, with different regimes which can be interpreted as different types of societal organization, from egalitarian to hierarchical. The control parameters of the model are connected to the cognitive capacity of the species in question, the size of the group and ecological and social pressures.

[3] A. Renyi. On measures of dependence. *Acta. Math. Acad. Sci. Hungar.*, 10:441–451, 1959

[4] R. S. Calsaverini and R. Vicente. An information-theoretic approach to statistical dependence: Copula information. *EPL (Europhysics Letters)*, 88(6): 68003, 2009. URL `http://stacks.iop.org/0295-5075/88/i=6/a=68003`

Lista de Figuras

2.1 Representação pictórica da projeção de uma distribuição. 29
2.2 A correlação linear é subestimada no caso de marginais não-Gaussianas. 36
2.3 Estimativas para a informação mútua *vs.* tau de Kendall para séries financeiras. 42
2.4 Seleção de pares de ações. 43
2.5 Excesso de informação mútua na cópula t 47

3.1 Ilustração da história da organização social dos humanos e primatas pré-humanos. 51
3.2 Gráfico em escala di-logaritimica do tamanho médio do grupo em função da razão média entre o volume do neocórtex e o volume total do cérebro para diversas espécies de primatas. 52
3.3 Grafo totalmente conectado 55
3.4 Grafo estrela 55
3.5 Corte do diagrama de fases apresentando o valor esperado dos parâmetros de ordem. 59
3.6 Diagrama de fases apresentando a razão $\frac{d_{\text{avg}}}{d_{\text{max}}}$ em função de α e temperatura, para um número fixo de agentes $n = 16$. 60
3.7 Valor crítico do parâmetro a em função da temperatura para diferentes tamanhos do sistema. 61
3.8 Corte do diagrama de fase para vários tamanhos do grupo. 61
3.9 Taxa de aceitação do algoritmo de Monte Carlo — fração das propostas de mudanças no microestado do sistema que foram aceitas com probabilidade dada pelo fator de Gibbs, sampleada com $\beta = 7.0$ 63
3.10 Parâmetros de ordem associados à correlação entre grafos de diferentes agentes 64

Conteúdo

Resumo 3

Abstract 5

1 *Introdução* 13

1.1 Visão geral 13
1.2 Inferência, Probabilidades e Entropia 13
1.2.1 Probabilidades e Inferência 14
1.2.2 Informação e Máxima Entropia 16
1.2.3 Informação e Vínculos — atualização Bayesiana . 18
1.2.4 Informação e vínculos — distribuições de Gibbs . 19
1.2.5 Métodos de campo médio 20
1.3 Tópicos tratados na Tese 22
1.3.1 Dependência Estatística, Teoria de Cópulas e Teoria de Informação 22
1.3.2 Um modelo Mecânico-Estatístico para a emergência de autoridade 23

2 *Dependência Estatística, Teoria de Cópulas e Teoria de Informação* 25

2.1 Dependência estatística e medidas de dependência . . . 25
2.1.1 O que se deseja de uma medida de dependência? 27
2.2 Informação Mútua 28
2.3 Cópulas 30
2.3.1 Exemplos 31
2.3.2 Dependência extrema — limites de Frechet-Hoeffding 32
2.3.3 Medidas de dependência revisitadas 33
2.4 Entropia de Cópula 34

2.5 "Excesso" de informação mútua 35
2.5.1 Correlação Linear *vs.* parâmetro ρ 35
2.5.2 Informação mútua para cópulas elípticas 36
2.5.3 Excesso de Informação Mútua 40
2.5.4 Ajuste empírico de cópulas via informação mútua 41
2.5.5 Cópula t 45
2.6 Conclusões 47

3 *Um Modelo para emergência de autoridade em sociedades humanas.* 49

3.1 Introdução 49
3.2 Evidências empiricas 50
3.2.1 "U-shaped evolution" 50
3.2.2 Evolução do cérebro primata e a Teoria Maquiavélica 51
3.2.3 Teoria da Reversão de Dominância 52
3.3 Um modelo mecanico-estatístico baseado em agentes . . 53
3.3.1 Descrição dos agentes 53
3.3.2 Variáveis dinâmicas — grafos sociais 54
3.3.3 Custos 55
3.3.4 Dinâmica para agentes isolados — máxima entropia 56
3.3.5 Minimização do custo 56
3.3.6 Interpretação dos parâmetros α e β 57
3.3.7 Resultados numéricos para agentes isolados . . . 58
3.3.8 Interpretação parcial dos resultados 62
3.3.9 Dinâmica para muitos agentes e resultados numéricos 63
3.4 Sumarização e interpretação dos Resultados 65
3.4.1 Grupos pequenos e/ou alta capacidade cognitiva 65
3.4.2 Grupos de tamanho intermediário e/ou capacidade cognitiva intermediária 65
3.4.3 Grupos grandes e/ou capacidade cognitiva menor 66

4 *Conclusão e Observações Finais* 67

4.1 Tópicos tratados na Tese 67
4.1.1 Dependência Estatística, Teoria de Cópulas e Teoria de Informação 67
4.1.2 Um modelo Mecânico-Estatístico para a emergência de autoridade 68

5 *Bibliografia* 73

A *Teorema de Sklar* 79

B *Teoremas de Cox* 83

B.1 Primeiro teorema de Cox e a regra do produto 83
B.2 Valores extremos 84
B.3 Teorema de Bayes 84
B.4 Regra da soma 85

⌈1⌋ Introdução

Thesis est omnis divisa in partes tres, quarum unam introductionem opus est, aliam de theoria copula, tertiam de mathematicum formula ad originem auctoritate.

⌈1.1⌋ Visão geral

Thesis est omnis divisa in partes tres. Além desta introdução, este trabalho trata de dois tópicos — uma abordagem da teoria de dependência estatística e um modelo para a origem de estruturas sociais hierárquicas — sob o ponto de vista da mecânica estatística, da teoria de informação e da inferência estatística. A adoção desse ponto de vista norteia as estratégias de modelagem matemática aqui selecionadas, e de uma certa forma, são mais essenciais ao trabalho do que os específicos tópicos em si. Dessa forma, faz-se necessário explicitar e esclarecer o ponto de vista adotado antes que os tópicos específicos sejam apresentados.

⌈1.2⌋ Inferência, Probabilidades e Entropia

Adquirir informação e tomar decisões sob incerteza — dois pontos centrais em qualquer estudo quantitativo — são os temas centrais da teoria da inferência estatística. A tradição do uso da teoria de probabilidades como ferramenta de inferência é centenária e já observada em trabalhos como os de de Moivre [3] e Laplace [4, 5] sobre o conceito de probabilidades nos séculos XVIII e XIX. A relação entre o conceito de probabilidade e os problemas de inferência ficaram ainda mais fortes com os trabalhos de Cox [6][1] e Shannon [8][2], e as versões mais modernas desse paradigma[3] lançam luz sobre a natureza

[1] R. T. Cox. Probability, frequency, and reasonable expectation. *Am. Jour. Phys.*, 14:1–13, 1946; and R.T. Cox. *Algebra of Probable Inference*. Johns Hopkins University Press, 2001. ISBN 9780801869822. URL `http://books.google.com.br/books?id=dcNpAUU6ACgC`

[2] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 47:379–423 and 623–656, 1948

[3] E.T. Jaynes and G.L. Bretthorst. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. ISBN 9780521592710. URL `http://books.google.com.br/books?id=tTN4HuUNXjgC`; Ariel Caticha. Lectures on probability, entropy, and statistical physics. *arXiv:0808.0012*, 2008. URL `http://arxiv.org/abs/0808.0012`. Invited lecture at the 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (July 8-13, 2008, Boraceia Beach, Sao Paulo, Brazil); and Ariel Caticha. Quantifying rational belief. *arXiv:0908.3212*, 2009. URL `http://arxiv.org/abs/0908.3212`. Presented at the 29th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (July 5-10, 2009, Oxford, Mississippi, USA.) In

da física estatística e do conceito de entropia. Nesta introdução, pretendemos apresentar rapidamente o paradigma de inferência segundo o método de Máxima Entropia (ME) e suas relações com a mecânica estatística, que pensamos ser a linha unificadora que dá coerência à diversidade de temas abordados nesse trabalho.

Raciocínio sobre informação completa a respeito da veracidade ou não de um conjunto de proposições pode ser representado através da tradicional álgebra booleana. Se é conhecido o valor de verdade de uma certa proposição e como ela se relaciona com outras proposições, pode-se inferir o valor de verdade das proposições relacionadas através das regras bem definidas da álgebra de proposições. Por exemplo, se é sabido que $P_1 \Rightarrow P_2$, e há certeza de que P_1 é verdadeira, pode-se inferir imediatamente que P_2 é verdadeira. Da mesma forma, a certeza de que P_2 é falsa imediatamente implica na certeza de que P_1 é falsa. Em outras palavras, a hipótese $P_1 = V$ fornece *informação completa* a respeito de P_2, bem como a hipótese $P_2 = F$ fornece informação completa sobre P_1. Entretanto, a certeza a respeito da falsidade de P_1 não oferece conclusão alguma, dentro desse paradigma de inferência sobre informação completa, a respeito da veracidade de P_2. Não é difícil, porém, formular um exemplo em que a informação sobre a falsidade de P_1 fornece *alguma informação*, ainda que incompleta, sobre P_2.

Consideremos, em um exemplo simples, a hipótese de que a proposição P_1 ="vai chover" implica a proposição P_2 ="há nuvens de chuva". No *ambiente lógico* criado por essa hipótese, a observação de nuvens de chuva não leva à conclusão certa de que está chovendo, mas é uma decisão razoável carregar um guarda-chuvas ao se observar essas nuvens. De alguma forma, a observação de que há nuvens de chuva trouxe alguma informação ao observador a respeito da possibilidade de que chova. Construir um método de inferência capaz de levar em conta informação incompleta é o objetivo da teoria de probabilidades bayesiana e do método de máxima entropia.

⌈1.2.1⌉ Probabilidades e Inferência

Para derivar uma teoria coerente de inferência, devem ser estabelecidos alguns requisitos. Dadas duas proposições P e Q, postulamos uma medida $(P|Q) \in \mathbb{R}$ denominada *plausibilidade*[4,5] da proposição P *dada a proposição* Q. A plausibilidade $(P|Q)$ representa o *grau de confiança* de que P esteja correta, dada uma certa informação prévia Q. Postulamos ainda que, sempre que existam duas formas diferentes

[4] A justificativa para usar números reais vem de um argumento simples de transitividade — se a confiança na veracidade de P_1 é maior que na veracidade de P_2 e esta é maior que a confiança na veracidade de P_3, então, um requisito razoável é que a confiança em P_1 seja maior que em P_3. Dessa forma, $(P_1|Q) > (P_2|Q)$ e $(P_2|Q) > (P_3|Q)$ implica $(P_1|Q) > (P_3|Q)$. Isso é suficiente para mostrar que existe uma representação real para essas quantidades. Consequências interessantes de se relaxar o requisito de transitividade e considerar plausibilidades representadas por números complexos são discutidas em Goyal et al. [12]

[5] Philip Goyal, Kevin H. Knuth, and John Skilling. Origin of complex quantum amplitudes and Feynman's rules. *Phys. Rev. A*, 81:022109, Feb 2010. DOI: 10.1103/PhysRevA.81.022109. URL http://link.aps.org/doi/10.1103/PhysRevA.81.022109. arxiv:0907.0909

de calcular a mesma plausibilidade, o resultado deve ser idêntico. Esse requisito leva aos seguintes resultados[6,7]:

- A plausibilidade[8] de não-P dado Q é uma função monotônica e decrescente da plausibilidade de P dado Q:

$$(\bar{P}|Q) = F((P|Q)).$$

- A plausibilidade da conjunção "P_1 e P_2" ($P_1 \wedge P_2$) dado Q é uma função das plausibilidades de P_1 dado Q e de P_2 dado $P_1 \wedge Q$:

$$(P_1 \wedge P_2|Q) = G((P_1|Q), (P_2|Q \wedge P_2)).$$

Uma série de teoremas sobre a forma das funções $F(\cdot)$ e $G(\cdot,\cdot)$ pode ser demonstrada com o requisito de consistência e as regras básicas da álgebra booleana. Alguns dos principais teoremas, cujas provas se encontram no apêndice B, *Teoremas de Cox*, são:

1º teorema de Cox

Teorema 1. *Uma vez que uma representação consistente de plausibilidades $(P|Q)$ com um ordenamento bem definido foi encontrada, sempre é possível encontrar uma outra equivalente $\pi(P|Q)$, de forma que $G(u, v) = uv$, ou seja:*

$$\pi(P_1 \wedge P_2|Q) = \pi(P2|Q \wedge P_1)\pi(P_1|Q) \tag{1.1}$$

Valores limites para plausibilidades

Teorema 2. *Uma vez que uma representação consistente de plausibilidades $\tilde{\pi}(P|Q)$ que satisfaça a regra do produto, é sempre possível encontrar uma equivalente $\pi(P|Q)$ tal que:*

$$0 \leq \pi(P|Q) \leq 1 \tag{1.2}$$

de tal forma que $\pi(P|Q) = 0$ se, e somente se, P for uma proposição falsa dado Q e $\pi(P|Q) = 1$ se, e somente se, P for uma proposição verdadeira dado Q.

2º teorema de Cox

Teorema 3. *Uma vez que uma representação consistente de plausibilidades $\pi(P|Q)$ com um ordenamento bem definido for encontrada para a qual valha a regra do produto, sempre é possível encontrar outra equivalente $p(P|Q)$, que também satisfaça a regra do produto, de forma que $F(u) = 1 - u$, ou seja:*

$$p(\bar{P}|Q) + p(P|Q) = 1 \tag{1.3}$$

Tomados em conjunto, esses teoremas sugerem que as regras de uma álgebra de plausibilidades deve ser idêntica às conhecidas regras

[7] M. Tribus. *Rational descriptions, decisions, and designs*. Pergamon unified engineering series: engineering design section. Pergamon Press, 1969. URL `http://books.google.com.br/books?id=C4pRAAAAMAAJ`

[8] Os seguintes símbolos serão usados para as operações booleanas no presente capítulo:

Conjunção — $\wedge$: representa a conjunção "e" entre duas proposições: $P \wedge Q$, lido "p e q".

Disjunção — $\vee$: representa a disjunção "ou" entre duas proposições: $P \vee Q$, lido "p ou q".

Negação — $\bar{}$: representa a negação "não" de uma proposição: $\bar{P}$, lido "não-p".

da Teoria das Probabilidades. A partir de agora, portanto, daremos o nome "probabilidade" ao funcional $p(P|Q)$, e interpretaremos probabilidades como formas de codificar matematicamente graus de certeza a respeito de certas proposições. Quando essas proposições são afirmações sobre o valor de uma variável matemática x que toma valores sobre um conjunto $\mathcal{X}$, definem-se distribuições de probabilidade[9] sobre o valor dessas variáveis:

$$P(x \in s|Q) = \int_{x \in s} p(x|Q) \, \mathrm{d}x \tag{1.4}$$

para qualquer subconjunto $s \subset \mathcal{X}$. Um modelo matemático, nesse paradigma, é, portanto, uma atribuição de distribuições de probabilidade para as variáveis de interesse do modelo, indicando graus de confiança sobre os valores dessas variáveis sob certas condições.

[9] Note que a equação (1.3) implica que:

$$\int p(x|Q)\mathrm{d}x = 1$$

⌈1.2.2⌉ Informação e Máxima Entropia

Se modelos de inferência são atribuições de probabilidades sobre as variáveis de interesse, como é possível fazer isso a partir de informação pré-existente sobre o sistema sendo modelado? Ou, ainda, como é possível incorporar novas informações obtidas sobre o sistema? Eventualmente, o objetivo de realizar inferência é processar informação nova que nos é disponibilizada depois da realização de um certo experimento ou observação. No presente paradigma, isso significa atualizar nossa atribuição de probabilidades. Suponha a proposição P_s ="A variável $X \in \mathcal{X}$ tem seu valor no subconjunto $s \subset \mathcal{X}$". Suponha ainda que, inicialmente, haja um certo conjunto de informações *a priori*, representadas por I_0, que nos levam a crer que nossa atribuição inicial de probabilidades deve ser:

$$p(P_s|I_0) = \int_{x \in s} p(x) \, \mathrm{d}x.$$

Uma vez que se torne disponível uma nova informação, digamos I_1, que nos force a mudar de opinião a respeito da atribuição de probabilidades, qual deve ser a nova distribuição a ser utilizada? Vamos representar simbolicamente por $\mathbb{P}$ o conjunto de todas as distribuições sobre a variável x. Digamos ainda que fosse possível ordenar as distribuições $q(x) \in \mathbb{P}$ em ordem de preferência como nova distribuição a ser atribuida a P_x. Se essa preferência for transitiva, existe um funcional $S : \mathbb{P} \to \mathbb{R}$ que representa esse ordenamento e a nova distribuição será obtida através da maximização do funcional $S[\cdot]$:

$$P(P_s|I_1) = \int_{x \in s} q(x) \, \mathrm{d}x$$
$$q(x) = \underset{\tilde{q}(x) \in \mathbb{P}}{\arg\max} \, S[\tilde{q}(x)|p(x)],$$

onde a maximização deve ser submetida a quaisquer vínculos impostos pela nova informação I_1. É possível impor requisitos plausíveis sobre $S[\cdot]$ de forma a definir um único funcional compatível com uma atualização racional das probabilidades? De fato, é possível, impondo apenas um requisito: as atribuições de probabilidade devem ser atualizadas apenas até onde requerido pela nova informação disponível. Resumidamente — detalhes podem ser obtidos em Caticha [10] —, esse princípio leva às seguintes consequências:

Localidade — Se a nova informação diz respeito apenas a um subdomínio do espaço onde X toma valores, então a atribuição de probabilidades fora desse subdomínio não deve mudar. Isso implica que o funcional deve ser aditivo:

$$S[q|p] = \int_{\mathcal{X}} \mathrm{d}\mu(x) F(q(x), p(x), x)$$

onde $\mu(x)$ é uma medida de integração sobre $\mathcal{X}$.

Invariância por mudança de variáveis — Uma mudança de sistema de coordenadas não deve mudar a forma do funcional S nem a ordem das preferências. Isso implica que:

$$S[q|p] = \int_{\mathcal{X}} \mathrm{d}\mu(x) \Phi\left(\frac{p(x)}{\mu(x)}, \frac{q(x)}{\mu(x)}\right)$$

Ausência de nova informação — Se não há nova informação, não há razão para mudar de idéia e, portanto, o máximo sem vínculos de $S[\cdot|\cdot]$ deve ser a própria distribuição original $p(x)$. Isso implica que:

$$S[q|p] = \int_{\mathcal{X}} \mathrm{d}x\ p(x) \Phi\left(\frac{q(x)}{p(x)}\right)$$

Sistemas independentes Se a distribuição $p(x)$ contém informação de que dois subsistemas X_1 e X_2 são independentes, nova informação a respeito de um deles não deve afetar o outro. Esse princípio leva a uma equação funcional para $\Phi(x)$ que finalmente implica que:

$$S[q(x)|p(x)] = -\int_{\mathcal{X}} \mathrm{d}x\ q(x) \log \frac{q(x)}{p(x)} \tag{1.5}$$

Esse funcional é conhecido em Teoria de Informação[10] e é denominado "entropia relativa". Os passos acima levam à formulação do princípio de máxima entropia[11]:

Princípio da Máxima Entropia. *Dada uma atribuição inicial de probabilidades sobre uma variável $X \in \mathcal{X}$ dada por $p(x)$ (distribuição a priori, ou prévia), quando nova informação se torna disponível, uma nova distribuição de probabilidades deve ser atribuida de forma a maximizar a entropia relativa*

[10] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, 2006. ISBN 9780471748816. URL `http://books.google.com.br/books?id=EuhBluW31hsC`

[11] Os nomes atribuídos às distribuições são traduções do inglês *prior distribution* e *posterior distribution*

entre a antiga distribuição $p(x)$ e a nova $q(x)$ (distribuição posterior ou a posteriori), de forma a satisfazer os vínculos impostos pela nova informação.

$$q(x) = \arg\max_{q(x)} S[q(x)|p(x)] \tag{1.6}$$

$$F_i[q(x)] = 0\ ,\ i = 1, 2, \ldots, m \tag{1.7}$$

onde $F_i[\cdot]$ são funcionais que codificam os vínculos relacionados às informações disponíveis.

Uma vez que o funcional $S[q|p]$ é convexo, se os vínculos forem também convexos, o máximo será interior e único e pode ser encontrado pela técnica de multiplicadores de Lagrange, resolvendo a condição variacional de primeira ordem para $q(x)$:

$$\frac{\delta}{\delta q(x)} \left(S[q|p] - \sum_i \lambda_i F_i[q] \right) = 0 \tag{1.8}$$

A entropia relativa na eq.(1.5) também pode ser ligada a outro conceito corrente em teoria de informação e geometria de distribuições de probabilidade, denominado divergência de Kullback-Leibler[12]:

$$D[q(x)|p(x)] = \int_{\mathcal{X}} \mathrm{d}x\ q(x) \log \frac{q(x)}{p(x)} = -S[q(x)|p(x)]. \tag{1.9}$$

A divergência de Kullback-Leibler apresenta as propriedades de uma pré-métrica, ou seja: $D[q(x)|p(x)] \geq 0$ e $D[q(x)|p(x)] = 0$ se, e somente se, $q(x) = p(x)$, no sentido de igualdade de distribuições[13]. Nessa linguagem, o princípio de máxima entropia pode ser entendido como um princípio de "mínima distância" — a distribuição posterior deve ser tão próxima da distribuição *a priori* quanto permitido pelos vínculos impostos pela nova informação.

⌈1.2.3⌉ Informação e Vínculos — atualização Bayesiana

Um caso específico de aplicação do princípio de Máxima Entropia é o do ajuste de parâmetros teóricos de um modelo quando novos dados empíricos são coletados. Suponha que um par de variáveis X e Θ são considerados em um modelo M. A variável X é experimentalmente mensurável e a variável Θ é um parâmetro teórico do modelo. O modelo oferece informação prévia sob a forma de (1) uma distribuição *a priori* dos valores possíveis do parâmetro Θ, dada por $p(\theta|M)$ e (2) uma distribuição condicional *a priori* $p(x|\theta, M)$ que indica, dado um valor do parâmetro $\Theta = \theta$, os possíveis resultados para X. Eventualmente, o valor de X é medido, com resultado $X = x_0$. Como devemos atualizar nossa atribuição de probabilidades? O princípio da máxima

[12] S. Amari, H. Nagaoka, and D. Harada. *Jōhō Kika No Hōhō (Methods of Information Geometry)*. Translations of mathematical monographs v. 191. Amer Mathematical Society, 2000. ISBN 9780821843024. URL `http://books.google.com.br/books?id=vc2FWSo7wLUC`

[13] Entretanto, por não ser simétrica e não satisfazer a desigualdade do triângulo, $D[q|p]$ não é uma métrica legítima para o conjunto de distribuições de probabilidades. Se restrita a uma família paramétrica $\mathcal{P}_\theta$ de distribuições parametrizadas por um certo conjunto de parâmetros θ (denotada por $p(x|\theta)$), podemos escrever:

$$D[p(x|\theta + \mathrm{d}\theta), p(x|\theta)] = \\ = \frac{1}{2} \sum_{ij} g_{ij} \mathrm{d}\theta_i \mathrm{d}\theta_j + \mathrm{O}(\mathrm{d}\theta^3),$$

onde

$$g_{ij} = \left\langle \frac{\partial \log p(x|\theta)}{\partial \theta_i}\ \frac{\partial \log p(x|\theta)}{\partial \theta_j} \right\rangle$$

é a chamada métrica de Fisher-Rao, que provê uma estrutura métrica (veja Amari et al. [15]) ao conjunto $\mathcal{P}_\theta$.

entropia indica que a distribuição posterior $q(x, \theta)$ é aquela que maximiza o funcional da eq.(1.5) sob o vínculo de que conhecemos o valor de X. Ou seja, o vínculo é dado por[14]:

$$\int q(x, \theta) d\theta = q(x) = \delta(x - x_0). \tag{1.10}$$

[14] Note que essa equação implementa, na verdade, um número infinito de vínculos — um para cada valor de x.

A distribuição *a priori* sobre x e θ é dada por $p(x, \theta|M) = p(x|\theta, M)p(\theta|M)$. Finalmente, a minimização é dada por:

$$\frac{\delta}{\delta q}\left[S[q|p] + \lambda\left(\int \mathrm{d}x\mathrm{d}\theta\; q(x,\theta) - 1\right) + \int \mathrm{d}x\; \beta(x)\left(\int \mathrm{d}\theta q(x,\theta) - \delta(x - x_0)\right)\right] = 0$$

onde λ e $\beta(x)$ são multiplicadores de Lagrange que implementam, respectivamente, o vínculo de normalização de $q(x, \theta)$ e os vínculos impostos pela eq.(1.10). Executando essa diferenciação funcional e isolando $q(x, \theta)$, obtemos[15]:

[15] O segundo passo é uma aplicação a definição de distribuição condicional

$$q(x, \theta|M) = p(x, \theta|M)\frac{e^{\beta(x)}}{Z} = p(\theta|x, M)p(x|M)\frac{e^{\beta(x)}}{Z} \tag{1.11}$$

onde Z é uma constante de normalização. Impondo o vínculo eq.(1.10), finalmente se obtém:

$$\begin{aligned} q(x, \theta|M) &= q(x|M)q(\theta|x, M) = \delta(x - x_0)p(\theta|x, M) \\ &\Rightarrow q(\theta|x_0, M) = p(\theta|x_0, M) \end{aligned}$$

Ou seja, as distribuições condicionais de θ devem ser iguais antes e depois de receber a informação, pois apenas informação a respeito de x — informação marginal, que diz respeito apenas à distribuição marginal de x — foi recebida. Isso é consequência do princípio de mínima atualização usado na dedução do princípio de máxima entropia: apenas se deve atualizar as probabilidades quando isso é imposto pela nova informação recebida. Essa equação pode ser reescrita como:

$$q(\theta|x_0, M) = \frac{p(x_0|\theta, M)p(\theta, M)}{p(x_0|M)} \tag{1.12}$$

e esse é o teorema de Bayes como entendido em inferência bayesiana — como um princípio de atualização de graus de confiança quando um nova informação está disponível. A distribuição $q(\theta|x_0, M)$ incorpora informações a respeito do modelo original e do fato de que a variável X foi medida e vale x_0.

[1.2.4] Informação e vínculos — distribuições de Gibbs

Outro tipo possível de informação que pode ser incorporada à distribuição de probabilidades é a respeito do valor esperado de uma

certa função de x:

$$\langle E(x) \rangle = E_0.$$

Neste caso, a maximização da entropia será:

$$\begin{aligned} &\frac{\delta}{\delta q}\left[S[q|p] - \lambda\left(\int \mathrm{d}x\, q(x) - 1\right) - \beta\left(\int \mathrm{d}x\, q(x)E(x) - E_0\right)\right] = 0 \\ &= -\log\left(\frac{q(x)}{p(x)}\right) - 1 - \lambda - \beta E(x) = 0 \end{aligned}$$

e obtém-se, finalmente:

$$q(x) = \frac{1}{Z} p(x) e^{-\beta E(x)} \tag{1.13}$$

[16] Esse raciocínio é reminiscente do encontrado em textos clássicos de mecânica estatística como Landau et al. [16], onde a distribuição de probabilidades para um conjunto de partículas é encontrada maximizando a entropia sob vínculos associados a quantidades conservadas microscopicamente.

A distribuição (1.13) faz parte da classe de distribuições gibbsianas, comuns em mecânica estatística[16,17]. Uma forma alternativa dessa visão pode ser encontrada[18] em Caticha and Vicente [17]. Nesse trabalho, os autores consideraram um modelo em que há uma função das variáveis microscópicas cujo valor deve ser importante determinante para uma dinâmica microscópica desconhecida. Ainda que não haja indicação de que essa deve ser uma grandeza estritamente conservada pela dinâmica interna dos agentes, há uma indicação de que o valor dessa função oferece informação sobre esta dinâmica. Isso é o suficiente para que a incorporação do valor dessa função na distribuição de probabilidades através de um vínculo traga informação útil.

[17] L.D. Landau, E.M. Lifshitz, and L.P. Pitaevskii. *Statistical Physics Pt. 1*, volume 6 of *Course of Theoretical Physics*. Elsevier Science & Technology Books, 1980. ISBN 9780750626361. URL http://books.google.com.br/books?id=NaB7oAkon9MC

[18] Nestor Caticha and Renato Vicente. Agent-based social psychology: from neurocognitive processes to social data. *Advances in Complex Systems*, 14(05):711–731, 2011. DOI: 10.1142/S0219525911003190. URL http://www.worldscientific.com/doi/abs/10.1142/S0219525911003190

⌈1.2.5⌉ Métodos de campo médio

Eventualmente, podemos nos deparar com uma distribuição $p(x)$ que não conseguimos tratar analiticamente. Esse é frequentemente o caso das distribuições gibbsianas como as da eq.(1.13). Nesse caso, a interpretação do negativo da entropia relativa — a divergência de Kullback-Leibler — como espécie de distância pode oferecer um esquema aproximativo conveniente. Seja $\mathcal{P}_\Theta$ uma família de distribuições tratáveis analiticamente, parametrizada por um parâmetro Θ. Pode-se selecionar como aproximação uma das distribuições dessa família, maximizando a entropia relativa ou, de forma equivalente, minimizando a divergência de Kullback-Leibler. Assim, $p(x)$ será aproximada por $p(x|\tilde{\theta})$, onde:

$$\begin{aligned} \tilde{\theta} &= \arg\max_\theta D[p(x|\theta)|p(x)] \\ &= \arg\max_\theta \int p(x|\theta) \log \frac{p(x|\theta)}{p(x)}. \end{aligned}$$

No caso de uma distribuição gibbsiana como eq.(1.13) temos:

$$D[p(x|\theta)|p(x)] = \int \mathrm{d}x\; p(x|\theta) \log\left[p(x|\theta)\frac{Z}{e^{-\beta H(x)}}\right]$$

que pode ser ainda escrito como:

$$D[p(x|\theta)|p(x)] = -S(\theta) + \beta\,\langle H\rangle_\theta + \log(Z)$$

onde $S(\theta) = -\int \mathrm{d}x p(x|\theta)\log p(x|\theta)$ é a entropia de Shannon da distribuição $p(x|\theta)$. Usando ainda a definição de energia livre termodinâmica, essa equação pode ser escrita como:

$$D[p(x|\theta)|p(x)] = \frac{F(\theta) - F}{T}. \tag{1.14}$$

onde $T = \frac{1}{\beta}$ é uma quantidade similar à temperatura e

$$F(\theta) = \langle H\rangle_\theta - TS(\theta) \tag{1.15}$$

e $F = -T\log(Z)$ são funções similares às energias livres da termodinâmica. Essa função deve ser otimizada com relação aos parâmetros θ para obter uma distribuição aproximativa. Uma vez que F independe do parâmetro θ, basta então maximizar a energia livre $F(\theta)$:

$$p(x) \cong p(x|\tilde{\theta}),\ \text{onde: } \tilde{\theta} = \arg\max_\theta F(\theta) \tag{1.16}$$

Se a distribuição original $p(\boldsymbol{x})$ pertence ao conjunto $\mathcal{P}_\Theta$, as propriedades da divergência de Kullback-Leibler garantem que a solução do problema de otimização é a própria $p(\boldsymbol{x})$. Nesse caso, $D[p|p] = 0$ e a energia livre $F(\tilde{\theta})$ ótima é igual à energia livre exata. Caso contrário, a distribuição e energia livre ótimas são aproximações para a distribuição e energia livre exatas. Uma vez que a divergência de Kullback-Leibler nunca é negativa, a equação (1.14) resulta em:

$$F(\theta) \geq F, \tag{1.17}$$

mais uma versão da célebre desigualdade de Bogouliubov, familiar da teoria de aproximações de campo médio em mecânica estatística[19].

O exemplo clássico desse esquema aproximativo é a escolha da família de distribuições aproximativas como o conjunto de distribuições em que as variáveis $x_1, x_2, \ldots, x_N$ são independentes. Nesse caso, obtém-se o que tradicionalmente se chama em mecânica estatística uma aproximação de campo-médio. Diversos outros esquemas de aproximação, como a aproximação de Bethe-Peierls[20], podem ser colocadas no mesmo esquema escolhendo famílias de distribuições adequadas[21]. Mais sobre aproximações de campo médio sob este ponto de vista pode ser encontrado em Opper and Saad [19][22] e Mézard and Montanari [20][23].

[19] S.R. Salinas. *Introdução a Física Estatística*. EDUSP, 1997. ISBN 9788531403866. URL http://books.google.com.br/books?id=cAlP3sc0K44C

[20] Também conhecida como *"sum-product algorithm"* ou *"belief propagation"*

[21] Famílias de distribuições com vários graus diferentes de fatorização da dependência entre as variáveis.

[22] M. Opper and D. Saad. *Advanced Mean Field Methods: Theory and Practice*. Neural Information Processing Series. Mit Press, 2001. ISBN 9780262150545. URL http://books.google.com.br/books?id=cu0X8sCDeNAC

[23] M. Mézard and A. Montanari. *Information, Physics, and Computation*. Oxford Graduate Texts. OUP Oxford, 2009. ISBN 9780198570837. URL http://books.google.com.br/books?id=jhCM7i0a6UUC

⌈1.3⌉ Tópicos tratados na Tese

As idéias e pontos de vista desenvolvidos acima norteiam o desenvolvimento de dois temas independentes no presente trabalho, que passarão a ser descritos agora.

⌈1.3.1⌉ Dependência Estatística, Teoria de Cópulas e Teoria de Informação

A primeira parte da tese trata de uma análise da relação entre três campos isolados sobre os quais trabalhos recentes lançam luz — a teoria de informação, a teoria de medidas de dependência estatística e a teoria de cópulas. Inicialmente, discutimos uma definição de medidas de dependência e concordância devida a Renyi [1] e mostramos como essa definição pode ser reescrita usando o conceito de cópula. Em particular, mostramos que boas medidas de dependência devem ser funcionais da função cópula, e independentes das distribuições marginais.

A seguir, demonstramos que a Informação Mútua respeita todas os requisitos da definição de Renyi [1] e que, de fato, pode ser escrita como a entropia de Shannon da densidade de cópula, o que leva à decomposição da entropia conjunta das variáveis a uma parte relacionada às distribuições marginais e outra associada à dependência entre as variáveis. As consequências desse fato são discutidas para o caso comum do uso da correlação linear como medida de dependência, bem como os problemas associados a esse uso quando as distribuições marginais não são normais.

Posteriormente o caso particular da família de distribuições elípticas é estudado e é demonstrada uma decomposição da informação mútua dessa família de distribuições em duas partes — uma relacionada à parte linear da dependência e outra relacionada à parte não-linear. É então introduzido o conceito de "excesso de informação mútua" como medida do desvio da dependência com relação a uma dependência puramente normal e linear. Uma técnica para medida desse desvio para distribuições elípticas é introduzida e aplicada para o caso da dependência entre pares de ações que compõem o índice $S\&P500$, onde identificamos pares de ações com dependências que desviam fortemente da normalidade (ou linearidade).

Por fim, discutimos um método para ajuste empírico de cópulas e tes-

tamos esse método para uma particular sub-família das cópulas elípticas — as cópulas t, derivadas da distribuição t de Student.

[1.3.2] Um modelo Mecânico-Estatístico para a emergência de autoridade

A segunda parte da tese trata de um modelo mecânico-estatístico, de interesse em antropologia, sobre a emergência de autoridade em sociedades humanas primitivas. Inicialmente, fazemos uma breve revisão de fatos arqueológicos e antropológicos sobre organização social em primatas, em particular em humanos, e sobre achados empíricos sobre as pressões seletivas envolvidas na evolução do cérebro primata.

Em seguida, utilizamos insights obtidos nessa revisão para propor um modelo de agentes que representam as informações sociais que obtiveram sobre sua tribo em um grafo. Discutimos uma função custo sobre o grafo definido pelo agente composta de duas partes: um custo cognitivo, que representa o investimento de recursos cognitivos na obtenção e manutenção dessas informações sociais, e um custo social, que representa as consequências para o agente de erros de julgamento a respeito de relações sociais provocados pela falta de informações. São introduzidos também efeitos de interação entre os agentes, essenciais para a interpretação do resultado, através de um mecanismo de aprendizado social.

O diagrama de fases do modelo é obtido através de simulações de Monte Carlo. A competição entre as duas partes da função custo leva a uma série de fases interessantes no modelo, que podem ser interpretadas como diferentes tipos de organização social no grupo de agentes considerado.

[2] Dependência Estatística, Teoria de Cópulas e Teoria de Informação

[2.1] Dependência estatística e medidas de dependência

O conceito de dependência estatística é central à teoria de probabilidades. Fazer hipóteses a respeito da dependência estatística entre as variáveis de interesse em um modelo é comum em diversas áreas — da física à análise financeira. Não é óbvio, entretanto, como expressar esse conceito de maneira precisa. A formalização precisa desse conceito é um dos objetivos desse capítulo. De maneira informal e grosseira, dependência estatística diz respeito a quanta *informação* se obtém a respeito de uma variável quando o valor de outra variável é conhecido. Os dois casos extremos podem ser mais facilmente entendidos em primeira análise: o caso de completa independência e o caso de completa dependência. Duas variáveis são ditas independentes se sua distribuição conjunta pode ser fatorada em um produto[1]:

$$p_{X,Y}(x,y) = p_X(x)p_Y(y). \tag{2.1}$$

[1] Manteremos o foco de nossa atenção em distribuições bivariadas, uma vez que a generalização é imediata.

De maneira similar, pode-se dizer que duas variáveis são independentes quando a distribuição condicional $p(x|y)$ é idêntica à distribuição marginal de X - $p_X(x)$. Nessa situação, o conhecimento do valor da variável Y não fornece qualquer informação a respeito da variável X. Duas variáveis são ditas completamente dependentes quando uma pode ser escrita como função monotônica da outra:

$$x = F(y). \tag{2.2}$$

Nesse caso, o conhecimento de uma das variáveis determina completamente o valor de outra. Ou seja $p(x|y) = \delta(x - F(y))$, com $F(\cdot)$ uma função monotônica. Pode-se, dessa maneira, tentar introduzir alguma

forma concreta de medir a dependência estatística em um parâmetro que possa ser estimado e usado para caracterizar a dependência entre variáveis de forma mais concreta.

Um parâmetro comumente usado para esse fim é a chamada correlação linear[2]

$$\hat{\rho}_{XY} = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}. \tag{2.3}$$

[2] A rigor, o módulo da correlação linear. A correlação linear mede, além de dependência, concordância, ou seja, quanto duas variáveis reais apresentam variação coordenada de seus sinais. Essa informação extra não é captada apenas pelo conceito de dependência.

A correlação linear $\hat{\rho}$ é um número real no intervalo $[-1, 1]$, simétrica com relação à transposição de X e Y que sempre se anula quando duas variáveis são independentes. É certamente a medida mais popular de dependência utilizada em todo tipo de análise estatística. Entretanto, seria mais preciso dizer que a correlação linear, como a nomenclatura aqui empregada sugere, é apenas uma medida de dependência linear. Há diversas falhas dessa medida em capturar a completa dependência entre duas variáveis. Em particular é fácil notar que é possível obter duas variavéis com correlação linear nula em que, entretanto, haja forte dependência entre ambas. Um exemplo simples é o par de variáveis definido por:

$$Y = f(X) + \epsilon \tag{2.4}$$

em que X e ϵ sejam independentes e tenham distribuições simétricas em torno da origem e $f(\cdot)$ seja qualquer função par. Nesse caso temos:

$$\mathrm{Cov}(X,Y) = E[Xf(X)] + E[X]E[\epsilon] - E[X]E[f(X)]$$

Note que se a distribuição de X é uma função par, então o valor esperado de qualquer função ímpar de X é nulo, o que anula a expressão acima. Dessa forma temos $\hat{\rho}_{XY} = 0$. Entretanto, caso a distribuição de ϵ seja bastante concentrada em torno da origem, o conhecimento de X pode fornecer informação quase completa a respeito de Y. Essa informação não é capturada pela correlação linear. O epíteto "linear", usado nesse trabalho para descrever a correlação, pode ser melhor compreendido se notarmos uma propriedade interessante da correlação: ela é invariante por mudanças de escala lineares nas variáveis X e Y. Uma reparametrização do tipo:

$$X' = \alpha_X X + \beta_X$$
$$Y' = \alpha_Y Y + \beta_Y$$

não muda o valor da correlação linear: $\hat{\rho}_{X'Y'} = \hat{\rho}_{XY}$. Entretanto uma mudança mais geral de escala não preserva essa propriedade. Uma nova variável:

$$Y' = g(Y)$$

com $g(\cdot)$ monotônica, a princípio contém exatamente a mesma informação a respeito de X que a antiga variável Y. Entretanto não se garante que a correlação se mantenha invariante. Seria esperado, além disso, que a afirmação de que a correlação entre duas variáveis é máxima em módulo fosse uma indicação de que a dependência entre as duas variáveis é máxima. Entretanto, isso não é garantido.

⌈2.1.1⌉ O que se deseja de uma medida de dependência?

A digressão acima acerca da natureza da correlação linear imediatamente suscita a pergunta: que parâmetros são boas medidas de dependência e quais são suas características? Pode-se enumerar uma série de desejos a respeito dessas medidas que se baseiem na noção intuitiva de dependência como o conteúdo de informação de uma variável a respeito de outra. Explicitamente, esperamos que:

1. uma boa medida de dependência seja um funcional $R : P_2 \to \mathbb{R}$ da distribuição conjunta de probabilidades, bem definido para qualquer par de variáveis aleatórias X e Y;
2. o funcional seja invariante sob permutação das variávies X e Y: $R(X, Y) = R(Y, X)$;
3. o funcional seja nulo *se, e somente se* as variáveis forem estritamente independentes;
4. o funcional atinja um valor máximo *se, e somente se* as variáveis apresentem dependência máxima, ou seja, sejam funções monotônicas uma da outra;
5. o funcional seja invariante por escolhas de novas variáveis $(X, Y) \to (U = u(X), V = v(Y))$ desde que nenhuma informação seja perdida, ou seja, desde que $u(\cdot)$ e $v(\cdot)$ sejam funções bijetoras;
6. o funcional seja uma função monotônica e crescente do módulo da correlação linear para o caso de distribuições conjuntas gaussianas.

Essa série de requisitos, essencialmente[3] enumerados pela primeira vez [4] por Renyi [1] , não são suficientes para escolher um parâmetro único e são satisfeitos por uma grande variedade de diferentes parâmetros usados em estatística. Como exemplo podemos citar o τ de Kendall. Dados dois pares, (X_1, Y_1) e (X_2, Y_2), de pontos sorteados independentemente da distribuição $p_{XY}(x, y)$, o τ de Kendall é dado por:

$$\begin{aligned}\tau_{XY} &= \text{Prob}\{(X1 - X2)(Y1 - Y2) > 0\} - \text{Prob}\{(X1 - X2)(Y1 - Y2) < 0\} \\ &= 4\int F(x, y)dF(x, y) - 1,\end{aligned}$$

[3] Rényi exigia ainda que a medida tomasse valores no conjunto $[0, 1]$, o que dispensamos, uma vez que é sempre possível transpor uma medida no intervalo $[0, \infty]$ para esse conjunto através de uma função monotônica. Além disso há requisitos de continuidade e convergência para sequências convergentes de distribuições.

[4] A. Renyi. On measures of dependence. *Acta. Math. Acad. Sci. Hungar.*, 10:441–451, 1959; and B. Schweizer and E. F. Wolff. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):879 – 885, 1981

onde $F(x,y)$ é a distribuição cumulativa de X e Y. Outra medida que satisfaz esses requisitos é a correlação de postos de Spearman dada por:

$$\rho^S_{XY} = 12 \int \left(F(x,y) - F_X(x)F_Y(y)\right) dF_X(x) dF_Y(y), \tag{2.5}$$

onde $F_X(x)$ e $F_Y(y)$ são as distribuições cumulativas marginais de X e Y respectivamente.

[2.2] Informação Mútua

Do ponto de vista de teoria de informação a dependência mútua de um conjunto de variáveis pode ser quantificada através da mínima "distância na variedade estatística" (divergência de Kullback-Leibler) entre a distribuição conjunta dessas variáveis e a "sub-variedade" de distribuições independentes (veja figura 2.1) Isso pode ser escrito da forma:

$$I(X_1, \ldots, X_n) = \min_{\{\phi_i(x_i)\}} \int p(x_1, \ldots, x_n) \log \frac{p(x_1, \ldots, x_n)}{\prod_{i=1}^n \phi_i(x_i)} \prod_{i=1}^n \mathrm{d}x_i \tag{2.6}$$

Fazendo a derivada funcional da expressão a ser minimizada com relação às distribuições indeterminadas $\phi_k(x)$ obtemos a condição de extremo, com o vínculo de que as distribuições $\phi_j(x)$ sejam normalizadas:

$$\phi_k(x_k) = \int p(x_1, x_2, \ldots, x_n) \prod_{i \neq k} \mathrm{d}x_i = p_{X_k}(x_k) \tag{2.7}$$

Ou seja: as distribuições $\phi_j(x_j)$ devem ser as distribuição marginais da variáveis X_j, e podemos reescrever:

$$I(X_1, \ldots, X_n) = \int p(x_1, \ldots, x_n) \log \frac{p(x_1, \ldots, x_n)}{\prod_{i=1}^n p_{X_i}(x_i)} \prod_{i=1}^n \mathrm{d}x_i \tag{2.8}$$

Esse funcional que mede a dependência mútua entre grupos de variáveis é denominado Informação Mútua ou Correlação Total[5]. Interpretações para a o funcional podem ser obtidas notando que:

[5] Alguns autores reservam o nome Informação Mútua para o caso de duas variáveis

$$I(X_1, \ldots, X_n) = \sum_{i=1}^n H(X_i) - H(X_1, \ldots, X_n) \tag{2.9}$$

$$= H(X_j) - H(X_j | X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n) \tag{2.10}$$

Dessa forma, pode-se interpretar $I(X_1, \ldots, X_n)$ como a redução na incerteza a respeito da variável X_j proporcionada pelo conhecimento das variáveis $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n$. A informação mútua para duas variáveis, dada por:

$$I(X,Y) = \int \mathrm{d}x\, \mathrm{d}y\; p(x,y) \log \frac{p(x,y)}{p_X(x), p_Y(y))}, \tag{2.11}$$

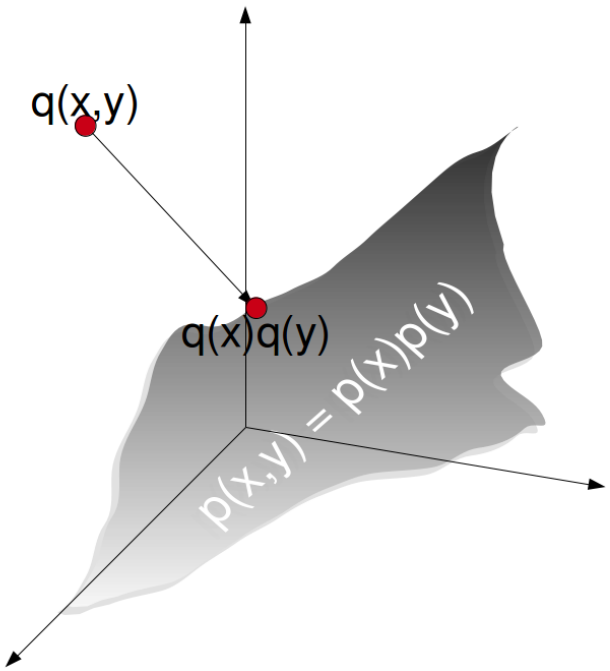


Figura 2.1 – Representação pictórica da projeção de uma distribuição em uma certa família de distribuições convenientes. O espaço representado na figura corresponde de forma pictórica ao espaço formado por todas as distribuições de probabilidade. A superfície corresponde a um sub-espaço, nesse caso, o sub-espaço de distribuições fatoráveis, correspondentes a variáveis independentes. A projeção de uma distribuição qualquer $q(x, y)$ sobre esse sub-espaço através da minimização da divergência de Kullback-Leibler resulta na distribuição dada pelo produto das distribuições marginais $q(x)q(y)$.

satisfaz todos os critérios discutidos na seção anterior para ser uma boa medida de dependência.

[2.3] Cópulas

A elucidação do conceito de dependência eventualmente se choca com a noção de cópula[6]. Um teorema devido a Sklar[7] permite separar a distribuição cumulativa conjunta — $F_{X,Y}(x,y)$ — de duas variáveis em duas partes: (a) as distribuições cumulativas marginais de cada variável — $F_X(x)$ e $F_Y(y)$, que trazem informação idiossincrática a respeito de cada uma das variáveis (b) e uma função cópula C(u,v), que traz informação sobre a dependência entre as variáves. De maneira geral o teorema de Sklar pode ser enunciado da seguinte forma:

[6] R.B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, 2006. ISBN 9780387286594. URL http://books.google.com.br/books?id=B3ONT5rBv0wC

[7] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959

Teorema 4 (Teorema de Sklar)**.** *Para toda distribuição cumulativa conjunta contínua de duas variáveis $F_{XY}(x,y)$, com distribuições cumulativas $F_X(x)$ e $F_Y(y)$, existe uma função cópula única $C(u,v)$ tal que:*

$$F_{XY}(x,y) = C(F_X(x), F_Y(y)). \tag{2.12}$$

Similarmente, dadas quaisquer duas distribuições cumulativas univariadas $F_X(x)$ e $F_Y(y)$ e uma função cópula $C(u,v)$, é possível construir uma distribuição conjunta dada por $F_{XY}(x,y) = C(F_X(x), F_Y(y))$.

A própria função cópula é uma legítima distribuição cumulativa conjunta, associada às variáveis $U = F_X(X)$ e $V = F_Y(Y)$:

$$F_{UV}(u,v) = C(u,v)$$

Dessa forma, podemos também definir a densidade de cópula, a densidade de probabilidade das variávies U e V:

$$p_{UV}(u,v) = c(u,v) = \frac{\partial^2 C}{\partial u \partial v}$$

Essa definição implica que a densidade de probabilidade conjunta das variáveis X e Y é dada por:

$$p_{XY}(x,y) = c(F_X(x), F_Y(y)) p_X(x) p_Y(y) \tag{2.13}$$

Dessa forma, o teorema de Sklar permite dividir a informação contida na distribuição conjunta em duas partes:

- uma parte que diz respeito apenas às propriedades de cada uma das variáveis — que é dada pelas distribuições marginais;
- e uma parte que diz respeito à dependência estatística entre as duas variáveis — que é dada pela função cópula.

⌈2.3.1⌉ Exemplos

Como exemplos de funções cópula temos as cópulas associadas às distribuições multivariadas mais comumente utilizadas. Qualquer família de distribuições multivariadas com um conjunto de parâmetros θ dada por $p(x, y|\theta)$ define uma função cópula dada por:

$$C(u, v|\theta) = \int_{-\infty}^{F_X^{-1}(u)} \int_{-\infty}^{F_X^{-1}(v)} \mathrm{d}x\mathrm{d}y p(x, y|\theta) \tag{2.14}$$

A cópula mais comumente empregada em todo tipo de análise estatística é a cópula normal ou gaussiana que, argumentaremos mais adiante, postula a mínima dependência linear entre duas variáveis:

$$N_\rho(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \mathrm{d}u\mathrm{d}v\, e^{-\frac{u^2+v^2-2uv\rho}{2(1-\rho^2)}} \tag{2.15}$$

onde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp -u^2/2\mathrm{d}u$ é a distribuição cumulativa normal padronizada. A família de cópulas normais depende apenas de um parâmetro $\rho \in [-1, 1]$ que, *no caso em que sejam inseridas marginais gaussianas para formar uma distribuição conjunta*, é igual à correlação entre as variáveis assim distribuidas. É importante notar que, para quaisquer outras marginais, a correlação poderá uma função do parâmetro ρ e de quaisquer outros parâmetros dessas distribuições marginais. Essa família contém a cópula de variáveis independentes $C(u, v) = uv$ quando $\rho = 0$.

Uma cópula ligeiramente mais complicada que a normal é a cópula associada à distribuição t de Student que, além da dependência linear descrita pelo parâmetro ρ, apresenta também dependência nas caudas da distribuição. A distribuição t de Student bivariada padrão[8] é dada por:

$$p_T(x, y \mid \rho, \nu) = \frac{\Gamma(1+\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})\pi\nu\sqrt{1-\rho^2}} \left[1 + \frac{q_\rho(x, y)}{\nu}\right]^{-(1+\frac{\nu}{2})}, \tag{2.16}$$

onde $q_\rho(x, y) = \frac{1}{1-\rho^2}\left[x^2 + y^2 - 2\rho xy\right]$. As marginais dessa distribuição são distribuições t univariadas:

$$p(x_i|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1 + \frac{x_i^2}{\nu}\right)^{-(\frac{\nu+1}{2})} \tag{2.17}$$

No limite $\nu \to \infty$ essa distribuição se reduz à distribuição normal bivariada padronizada, com parâmetro de correlação linear ρ, médias nulas e variâncias unitárias. Para ν finito as marginais adquirem caudas pesadas e a dependência entre as variáveis ganham uma componente

[8] Médias não-nulas e variância diferente da unitária podem ser trivialmente acrescentadas através de translações e mudanças de escala. Uma vez que essas transformações só afetam as marginais e mantém a cópula invariate — são trasformações inversíveis, coordenada a coordenada — não afetam a dependência

além da correlação linear — adicionando peso nas caudas da cópula. A cópula t pode ser obtida introduzindo essa distribuição na eq. (2.14):

$$C_T(u,v|\nu,\rho) = \frac{\Gamma(1+\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})\pi\nu\sqrt{1-\rho^2}} \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(u)} \mathrm{d}x\mathrm{d}y \left[1 + \frac{q_\rho(x,y)}{\nu}\right]^{-\frac{\nu+2}{2}} \tag{2.18}$$

onde $t_\nu(x)$ é a distribuição cumulativa associada à distribuição univariada em (2.17).

Uma terceira família de cópulas que convém citar são as cópulas arquimedianas, que podem ser escritas como:

$$C(u,v|\Psi(\cdot)) = \Psi^{-1}\left(\Psi(u) + \Psi(v)\right) \tag{2.19}$$

parametrizadas por uma função ϕ. Essas funções cópula existem desde que: $\Psi(1) = 0$, $\lim_{x\to 0}\Psi(x) = \infty$, $\Psi'(x) < 0$ e $\Psi''(x) > 0$. Essa família contém a cópula de variáveis independentes quando $\Psi(x) = -\log(x)$.

⌈2.3.2⌉ Dependência extrema — limites de Frechet-Hoeffding

A definição de cópula permite tornar mais preciso o conceito de dependência extrema. As equações (2.1) e (2.13) em conjunto nos permitem concluir que, para duas variáveis independentes:

$$c(u,v) = 1, \tag{2.20}$$
$$C(u,v) = uv, \tag{2.21}$$

Para o caso de dependência completa é possível mostrar que[9] toda cópula está limitada por duas funções que representam dependência máxima, denominadas limites de Frechet-Hoeffding. Essas funções são:

[9] R.B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, 2006. ISBN 9780387286594. URL http://books.google.com.br/books?id=B3ONT5rBv0wC

$$W(u,v) = \max(0, u+v-1) \tag{2.22}$$
$$M(u,v) = \min(u,v) \tag{2.23}$$

Essas duas funções são elas próprias cópulas e limitam por cima e por baixo todas as outras cópulas possíveis:

$$W(u,v) \leq C(u,v) \leq M(u,v) \tag{2.24}$$

para qualquer possível cópula $C(u,v)$. As densidades de cópula associadas a essas duas funções evidenciam que casos descrevem:

$$w(u,v) = \delta(u+v) \tag{2.25}$$
$$m(u,v) = \delta(u-v). \tag{2.26}$$

Inserindo marginais F(x) e G(y) quaisquer, nota-se o tipo de distribuições conjuntas que essas cópulas geram: ambas descrevem duas variáveis com dependência monotônica — crescente no caso de M e decrescente no caso de W.

⌈2.3.3⌉ Medidas de dependência revisitadas

Os "axiomas" de Rényi a respeito de medidas de dependência podem ser revisitados e tornados mais precisos com o conceito de cópula e cópulas extremas em mãos. Os itens 1, 2 e 5 ficam imediatamente satisfeitos se a medida de dependência em questão for funcional apenas da cópula e não das distribuições marginais. Além disso, os itens 3 e 4 podem ser reescritos em termos das cópulas extremas e da cópula independente. Podemos reescrever então esses requisitos da seguinte forma:

- Uma boa medida de dependência entre duas variáveis X e Y é um funcional $\mathcal{F} : C_2 \to \mathcal{R}$ que leva funções cópula $C_{XY}(\cdot,\cdot)$ em números reais e independe das distribuições marginais;
- atinge um valor mínimo, que será arbitráriamente escolhido como zero, se, e somente se, $C_{XY}(u,v) = uv$;
- atinge um valor máximo quando $C_{XY}(u,v) = W(u,v)$ ou $C_{XY}(u,v) = M(u,v)$.
- para $C_{XY}(u,v) = N_\rho(u,v)$, o funcional é um função monotônica crescente do parâmetro ρ.

A correlação linear falha em dois itens: é possível representar a correlação linear como função da cópula, mas não é possível eliminar sua dependência com as marginais:

$$\hat{\rho}_{X,Y} = \frac{1}{\sigma_X \sigma_Y} \int\int C(u,v) \mathrm{d}F_X^{-1}(u) \mathrm{d}F_X^{-1}(v).$$

Essa expressão depende das marginais explicitamente nas medidas de integração e implicitamente nas variâncias σ_i. Além disso, a correlação não atinge seus valor extremo sempre que a cópula escolhida como uma das cópulas de Frechet-Hoeffding — o valor assumido nesse caso depende das marginais específicas. Outras medidas apresentadas acima, o tau de Kendall (τ) e a correlação de postos de Spearman (ρ^S) podem ser facilmente escritas de forma a satisfazer todas os critérios acima:

$$\tau = 4 \int_0^1 \int_0^1 C(u,v) dC(u,v) - 1 \tag{2.27}$$

$$\rho^S = 12 \int_0^1 \int_0^1 [C(u,v) - uv] \, dudv \tag{2.28}$$

O último requisito pode ser verificado notando-se que, para o caso de cópulas normais, essas expressões se reduzem a[10]:

$$\tau = \frac{2}{\pi} \arcsin(\rho), \tag{2.29}$$

$$\rho^S = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right). \tag{2.30}$$

[10] A expressão para o tau de Kendall vale para toda distribuição elíptica, incluindo a distribuição t de Student

Além disso, a Informação Mútua, como mostraremos na próxima seção, também satisfaz os requisitos acima.

[2.4] Entropia de Cópula

Para escrever a informação mútua como um funcional da cópula basta recorrer à definição da densidade de cópula e a expressão da distribuição conjunta em termos desta, na eq. (2.13) que reproduzimos abaixo:

$$p_{XY}(x,y) = c(F_X(x), F_Y(y)) p_X(x) p_Y(y).$$

Usando essa expressão na definição de informação mútua eq. (2.11):

$$I(X,Y) = \int \mathrm{d}x\, \mathrm{d}y\; p(x,y) \log \frac{p(x,y)}{p_X(x), p_Y(y))},$$

temos:

$$\begin{aligned} I(X,Y) &= \int p_X(x)\mathrm{d}x\, p_Y(y)\mathrm{d}y\; c(F_X(x), F_Y(y)) \log\left[c(F_X(x), F_Y(y)\right] \\ &= \int \mathrm{d}u\mathrm{d}v\; c(u,v) \log c(u,v) \end{aligned}$$

e portanto:

$$I(X,Y) = \int\int \mathrm{d}u\mathrm{d}v\; c(u,v) \log c(u,v) = -S[c] \geq 0 \tag{2.31}$$

onde $S[c]$ é a entropia associada à distribuição conjunta $c(u,v)$, daqui por diante denominada *entropia de cópula* [11]. Esse resultado oferece ainda mais uma interpretação aos múltiplos significados da informação mútua — que ressoa diretamente com a definição de dependência apresentada no início deste capítulo — é o conteúdo de informação associado à dependência entre duas variáveis. A combinação desse resultado com a eq. (2.9) permite escrever[12]:

$$H(X_1, X_2, \ldots, X_N) = \sum_{i=1}^{N} H(X_i) + S[c] \tag{2.32}$$

decompondo então a entropia, o conteúdo informacional da distribuição conjunta, em parcelas devidas às características de cada uma das variáveis e uma parcela correspondente ao acoplamento entre essas

[11] Jian Ma and Zengqi Sun. Mutual information is copula entropy. *Tsinghua Science and Technology*, 16(1):51–54, 2011. ISSN 1007-0214. DOI: 10.1016/S1007-0214(11)70008-6. arXiv:0808.0845; and Rick L. Jenison and Richard A. Reale. The shape of neural dependence. *Neural Computation*, 16:665 – 672, 2004

[12] A expressão acima pode ser imediatamente generalizada para um número arbitrário de variáveis

variáveis. Eq. (2.31) também oferece um conveniente princípio para encontrar cópulas informacionalmente neutras segundo o princípio de máxima entropia[13]: a cópula menos informativa é a que postula a menor informação mútua possível entre as variáveis.

[13] E.T. Jaynes and G.L. Bretthorst. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. ISBN 9780521592710. URL http://books.google.com.br/books?id=tTN4HuUNXjgC

O ÚLTIMO PASSO para mostrar que a informação mútua satisfaz todos os critérios para ser uma boa medida de dependência é notar que, no caso de cópulas gaussianas:

$$I(X_1, X_2) = -\frac{1}{2}\log\left(1-\rho^2\right) \tag{2.33}$$

⌈2.5⌉ "EXCESSO" DE INFORMAÇÃO MÚTUA

⌈2.5.1⌉ CORRELAÇÃO LINEAR *vs.* PARÂMETRO ρ

A DISTRIBUIÇÃO DE MÁXIMA ENTROPIA que satisfaz vínculos de que correlação, médias, e variâncias de um par de variáveis X e Y assumam certos valores é a distribuição normal. Usando a decomposição da eq. (2.32), temos, uma vez que as entropias das marginais dependem apenas das variâncias e a informação mútua apenas da correlação[14]:

$$H(X_1, X_2) = h(\sigma_1) + h(\sigma_2) + \frac{1}{2}\log\left(1-\hat{\rho}^2\right)$$

onde $h(\sigma)$ é a entropia de uma distribuição normal univariada com variância σ^2. Se essa é a maior possível entropia dada a correlação e variâncias, e uma vez que a informação mútua independe das variâncias[15], então o valor:

$$I_0(\hat{\rho}) = -\frac{1}{2}\log\left(1-\hat{\rho}^2\right) \tag{2.34}$$

é um limite inferior para a informação mútua de qualquer par de variáveis que tenham correlação $\hat{\rho}$:

$$I_{XY} \geq I_0(\hat{\rho}_{XY}). \tag{2.35}$$

Dessa forma, representando formalmente em um plano I *vs.* $\hat{\rho}$ as possíveis distribuições conjuntas com cópula gaussiana[16], temos a figura 2.2.

NESSA FIGURA temos duas distribuições destacadas. Uma delas possui marginais gaussianas, sendo portanto uma distribuição normal bivariada e deve estar localizada sobre a curva $I_0(\hat{\rho})$, representada pela linha escura tracejada. Para essa distribuição o parâmetro ρ, uma boa medida de dependência se restrito a cópulas gaussianas, é exatamente

[14] Note a importância de separar conceitualmente o parâmetro correlação, que identifica uma certa distribuição na família de distribuições normais, do funcional são homônimos, o qual estamos chamado de "correlação linear" e que está definido para qualquer distribuição.

[15] Pois independe das marginais.

[16] Essa representação não é única. Está sendo empregada apenas como ilustração. Essas duas grandezas não formam um bom sistema de coordenadas para a variedade de distribuições com cópula gaussiana.

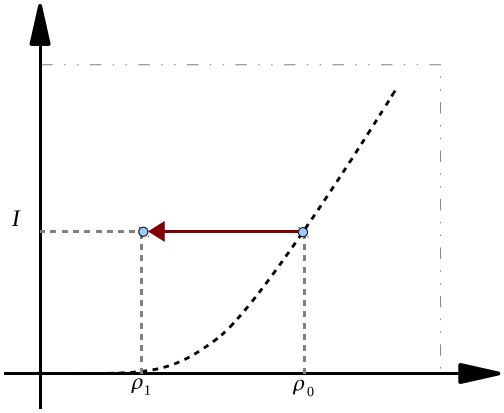


Figura 2.2 – A correlação linear é subestimada no caso de marginais não-Gaussianas. Se ambas as marginais são Gaussianas, a distribuição conjunta está localizada sobre o limite inferior para a informação mútua. Uma mudança nas marginais mantendo fixa a cópula, preserva a informação mútua, entretanto a correlação estimada deve se deslocar para valores de menor módulo.

igual à correlação linear $\hat{\rho}$. Se as marginais forem trocadas por marginais não-gaussianas, a informação mútua, independente das marginais, deve permanecer a mesma. Entretanto a correlação, como argumentado acima, deve mudar com a troca de marginais. Uma vez que uma mudança para valores maiores do módulo da correlação linear violaria a condição $I \geq I_0(\hat{\rho})$, a única alternativa é que o módulo da correlação diminua. Dessa forma, teríamos uma distribuição conjunta que tem cópula gaussiana com parâmetro ρ_0 e correlação linear $\hat{\rho} < \rho_0$. Uma tentativa de identificar o parâmetro ρ da cópula gaussiana com a correlação linear levaria a atribuir ao par de variáveis uma dependência menor — talvez muito menor — do que a real. Em muitas aplicações isso pode ser perigoso. Em particular em aplicações financeiras o risco de uma operação pode ser subestimado por se subestimar a frequencia de co-ocorrência de eventos negativos. É notório que séries temporais financeiras apresentam distribuições marginais com caudas pesadas — e portanto distantes de uma gaussiana. O uso, bastante difundido[17], de cópulas gaussianas para estimativa de risco combinado com o uso de estimadores baseados na correlação linear nessas condições pode significar que o risco de uma posição é substancialmente maior do que o estimado. Um dos requisitos para uma boa medida de dependência é que, para o caso de cópulas gaussianas, a medida em questão seja função monotônica e crescente do módulo do parâmetro ρ. Dessa forma, qualquer outra medida de dependência é mais adequada que a correlação linear para estimar a dependência de fato entre duas variáveis com dependência gaussiana.

[17] S. Jones. The formula that felled Wall St. *The Financial Times*, Abril 2009. URL `http://www.ft.com/cms/s/2/912d85e8-2d75-11de-9eba-00144feabdc0.html`

⌈2.5.2⌉ Informação mútua para cópulas elípticas

O tau de Kendall é um estimador ainda mais completo para ρ: ele é uma função monotônica do parâmetro ρ para qualquer cópula pertencente à família elíptica — da qual a gaussiana é um caso particular.

Uma distribuição é dita elíptica sempre que:

$$\left\langle e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \right\rangle = e^{i\boldsymbol{\mu}^T\boldsymbol{k}}\psi\left(i\frac{\boldsymbol{k}^T\Sigma\boldsymbol{k}}{2}\right), \tag{2.36}$$

onde Σ é a matriz de covariância e $\boldsymbol{\mu}$ o vetor de valores esperados de $\boldsymbol{x}$. Uma distribuição elíptica padronizada é aquela em que as médias são nulas e todas as variâncias unitárias, de modo que a matriz de covariâncias é igual à matriz de correlações $\Sigma_{ij} = \rho_{ij}$. Denotaremos essa família de distribuições por $p(\boldsymbol{x}|\Sigma, \psi(\cdot))$. No caso de $\Sigma_{ij} = \delta_{ij}$, a distribuição é invariante por rotações no vetor $\boldsymbol{x}$ e é dita uma *distribuição esférica* — na família $p(\boldsymbol{x}|\psi(\cdot))$.

É POSSÍVEL VER UMA VARIÁVEL DISTRIBUIDA com respeito a uma distribuição elíptica $\boldsymbol{x} \sim p(\boldsymbol{x}|\Sigma, \psi(\cdot))$ como uma transformação linear de variáveis distribuidas com relação à distribuição esférica com o mesma função $\psi(\cdot)$:

$$x_i = \sum_j A_{ij} y_j$$

Onde $\Sigma = A^T A$ e $\boldsymbol{y} \sim p(\boldsymbol{y}|\psi(\cdot))$. Cópulas elípticas são as cópulas associadas a essas distribuições e têm como parâmetros as correlações de pares — ρ_{ij} — e a função $\psi(\cdot)$. A eq.(2.36) é uma transformação de Fourier, que sempre pode ser invertida para obter a distribuição conjunta. No caso de distribuições elípticas padrão, temos:

$$p(\boldsymbol{x}|\Sigma, \psi(\cdot)) = \int \frac{\mathrm{d}^N k}{(2\pi)^N} e^{-i\boldsymbol{k}^T\boldsymbol{x}}\psi\left(i\frac{\boldsymbol{k}^T\Sigma\boldsymbol{k}}{2}\right)$$

podemos introduzir uma função delta e obter:

$$p(\boldsymbol{x}|\Sigma, \psi(\cdot)) = \int \mathrm{d}u \int \frac{\mathrm{d}^N k}{(2\pi)^N} e^{-i\boldsymbol{k}^T\boldsymbol{x}}\psi(u)\,\delta\left(u - i\frac{\boldsymbol{k}^T\Sigma\boldsymbol{k}}{2}\right)$$

e usando a representação integral da função delta:

$$\begin{aligned} p(\boldsymbol{x}|\Sigma, \psi(\cdot)) &= \int \frac{\mathrm{d}u\mathrm{d}\hat{u}}{2\pi} \int \frac{\mathrm{d}^N k}{(2\pi)^N} \psi(u) \exp\left[i\hat{u}\left(u - i\frac{\boldsymbol{k}^T\Sigma\boldsymbol{k}}{2}\right) - i\boldsymbol{k}^T\boldsymbol{x}\right] \\ &= \int \frac{\mathrm{d}u\mathrm{d}\hat{u}}{2\pi} e^{i\hat{u}u}\psi(u) \int \frac{\mathrm{d}^N k}{(2\pi)^N} \exp\left[\hat{u}\frac{\boldsymbol{k}^T\Sigma\boldsymbol{k}}{2} - i\boldsymbol{k}^T\boldsymbol{x}\right] \end{aligned}$$

A integral gaussiana sobre $\boldsymbol{k}$ pode ser feita e temos:

$$\begin{aligned} p(\boldsymbol{x}|\Sigma, \psi(\cdot)) &= \int \frac{\mathrm{d}u\mathrm{d}\hat{u}}{2\pi} e^{i\hat{u}u}\psi(u)\, N(x|\hat{u}\Sigma) \\ &= \int \mathrm{d}u\; p(u) N(x|u\Sigma) = \end{aligned}$$

onde $N(x|\Sigma)$ é a distribuição normal padronizada com matriz de correlação Σ e $p(u) = \int \mathrm{d}u e^{ivu}\psi(v)$ é uma certa função ligada à transformada de Fourier de $\psi(u)$. Essa representação para as distribuições

elípticas pode ser entendida da seguinte forma: $p(\boldsymbol{x}|\Sigma, \psi(\cdot))$ é a distribuição resultante quando se toma $\boldsymbol{x}$ de uma gaussiana de matriz de correlação $u\Sigma$ em que u é sorteado de acordo com uma distribuição $p(u)$. Dessa forma podemos, de forma alternativa, parametrizar a família elíptica pela distribuição $p(u)$ - $p(\boldsymbol{x}|\Sigma, p(\cdot))$. Também podemos escrever de forma mais explícita:

$$p(\boldsymbol{x}|\Sigma, p(\cdot)) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \int \mathrm{d}^N u \; \frac{1}{u^{N/2}} p(u) e^{-\frac{1}{2u}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}} \tag{2.37}$$

$$= \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\right) \tag{2.38}$$

onde definimos: $\Psi_j(q) = \int \mathrm{d}^j u \; \frac{1}{u^{j/2}} p(u) e^{-\frac{q}{u}}$. As marginais de uma distribuição elíptica podem ser facilmente calculadas notando-se que as marginais da normal padronizada são distribuições normais padronizadas sobre 1 variavel. Dessa forma:

$$p_j(x_j) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}u \; \frac{1}{\sqrt{u}} p(u) e^{-\frac{x_j^2}{2u}} = \frac{1}{\sqrt{2\pi}} \Psi_1\left(\frac{x_j^2}{2}\right) \tag{2.39}$$

Finalmente de posse dessas das eqs. (2.38) e (2.39) podemos mostrar a seguinte proposição.

Proposição 1 (Decomposição da informação mútua de uma cópula elíptica)**.** *Se $C(u, v|\Sigma, p(\cdot))$ é uma cópula elíptica com matriz de correlação Σ, então a informação mútua associada pode ser decomposta na forma:*

$$I(\Sigma, \psi(\cdot)) = I_0(\Sigma) + I[p(\cdot)], \tag{2.40}$$

onde $I_0(\Sigma) = -\frac{1}{2}\log\Sigma$ é a informação mútua de uma cópula gaussiana com matriz de correlação Σ e $I[p(\cdot)]$ é um funcional de $p(u)$ que é igual à informação mútua da distribuição esférica correspondente e independe da matriz de correlação.

Demonstração. Para mostrar essa proposição recorremos à eq.(2.9):

$$I = \sum_i^N H[X_j] - H[\boldsymbol{x}] = NH[X_1] - H[\boldsymbol{x}]$$

onde a segunda igualdade é obtida notando que as marginais $p_j(\cdot)$ são todas idênticas. A primeira parcela já é um funcional de p(u) que independe da matriz de correlação que pode ser escrito como:

$$NH[X_1] = N \int \mathrm{d}x \frac{1}{\sqrt{2\pi}} \Psi_1\left(\frac{x^2}{2}\right) \log\left[\frac{1}{\sqrt{2\pi}} \Psi_1\left(\frac{x^2}{2}\right)\right]$$

$$= \int \mathrm{d}^N x \frac{1}{\sqrt{(2\pi)^N}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right) \log\left[\frac{1}{\sqrt{(2\pi)^N}} \prod_j \Psi_1\left(\frac{x_j^2}{2}\right)\right]$$

A segunda parcela pode ser explicitamente escrita como:

$$H[\boldsymbol{x}] = -\int \mathrm{d}^N x \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\right) \log\left[\frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\right)\right]$$

A inversa da matriz de correlação é simétrica e quadrada, e portanto sempre pode ser escrita como $\Sigma^{-1} = U^T\Lambda U$, onde U é uma matriz unitária e $\Lambda_{ij} = \delta_{ij}\lambda_j$ é uma matriz diagonal dos autovalores de Σ^{-1}. Fazendo a mudança de variáveis $\boldsymbol{y} = U\boldsymbol{x}$ temos:

$$H[\boldsymbol{x}] = -\int \mathrm{d}^N y \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \Psi_N\left(-\frac{1}{2}\boldsymbol{y}^T\Lambda\boldsymbol{y}\right) \log\left[\frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \Psi_N\left(-\frac{1}{2}\boldsymbol{y}^T\Lambda\boldsymbol{y}\right)\right]$$

A matriz diagonal Λ pode ser escrita como $\Lambda = D^T D$, onde $D_{ij} = \delta_{ij}\sqrt{\lambda_j}$. Fazendo a mudança de variáveis[18] $\boldsymbol{x} = D\boldsymbol{y}$, temos:

$$H[\boldsymbol{x}] = -\int \mathrm{d}^N x \frac{1}{\sqrt{(2\pi)^N}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right) \log\left[\frac{1}{\sqrt{(2\pi)^N|\Sigma|}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right)\right]$$

Uma vez que $|D| = \prod_i \sqrt{\lambda_i} = \frac{1}{\sqrt{|\Sigma|}}$. O termo que contém $|\Sigma|$ dentro do logaritmo pode ser removido da integral[19] e ficamos com:

$$H[\boldsymbol{x}] = -\int \mathrm{d}^N x \frac{1}{\sqrt{(2\pi)^N}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right) \log\left[\frac{1}{\sqrt{(2\pi)^N}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right)\right] + \frac{1}{2}\log|\Sigma|$$

[18] Vamos renomear a nova variável de integração como $\boldsymbol{x}$ novamente por conveniência

[19] Note que a normalização de $\mathrm{p}(\boldsymbol{x})$ exige que $\int \mathrm{d}^N x \psi_N(\boldsymbol{x}^T\boldsymbol{x}) = \sqrt{(2\pi)^N}$

e finalmente podemos escrever[20]:

$$I = I[p(u)] - \frac{1}{2}\log|\Sigma| \tag{2.41}$$

onde:

$$I[p(u)] = \int \mathrm{d}^N x \frac{1}{\sqrt{(2\pi)^N}} \Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right) \log\left[\frac{\Psi_N\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{x}\right)}{\prod_j \Psi_1\left(\frac{x_j^2}{2}\right)}\right] \tag{2.42}$$

□

Essa proposição nos permite escrever a informação mútua em duas parcelas — uma dependente da estrutura linear de dependência, relacionada à matriz de correlações, e uma parcela que contém informações sobre dependências não-lineares entre as variáveis. Uma vez que a informação mútua é sempre positiva, esse resultado nos permite ainda escrever uma versão mais forte da desigualdade eq.(2.35) para o caso de cópulas elípticas:

$$I_{XY} \geq I_0(\Sigma_{XY}) \tag{2.43}$$

onde, nesse caso, Σ não é apenas a matriz de correlações lineares, mas o conjunto de parâmetros que identifica unicamente uma certa distribuição dentro de uma sub-família de cópulas elípticas com mesma função $\psi(\cdot)$.

⌈2.5.3⌉ Excesso de Informação Mútua

As inequações eq.(2.35) e eq.(2.43) permitem concluir que a cópula gaussiana é a cópula de menor entropia dado o vínculo de dependências lineares representados por $\hat{\rho}$ ou Σ. Em outras palavras, a cópula gaussiana é a cópula que assume que uma variável tem a menor quantidade possível de informação a respeito de outra que ainda explica a parte linear da dependência. Além disso, a cópula gaussiana possui apenas a parte linear da informação mútua, e portanto representa uma dependência apenas linear entre as variáveis. O uso dessa cópula portanto representa uma hipótese implícita de dependência linear e mínima entre as variáveis. Caso essa hipótese falhe, um termo adicional deve surgir na informação mútua, que diz respeito à dependência não-linear. Esse termo será chamado "excesso" de informação mútua — entenda-se excesso com relação à cópula linear. A observação de um excesso de informação mútua permite criar um diagnóstico de "gaussianidade" da dependência entre duas variáveis. Para tal é

[20] No caso bivariado essa expressão se reduz à expressão 2.34

necessário ser capaz de estimar o parâmetro ρ da distribuição e a informação mútua. Para estimar ρ, empregaremos o tau de Kendall — uma medida que independe das marginais e permite estimar:

$$\rho = \sin\left(\frac{\pi\tau}{2}\right). \tag{2.44}$$

O tau de Kendall pode ser estimado empiricamente com um algoritmo simples: dado um conjunto de pontos $\{\boldsymbol{x}_\mu \sim p(\boldsymbol{x})|\mu = 1, 2, \ldots, P\}$, temos:

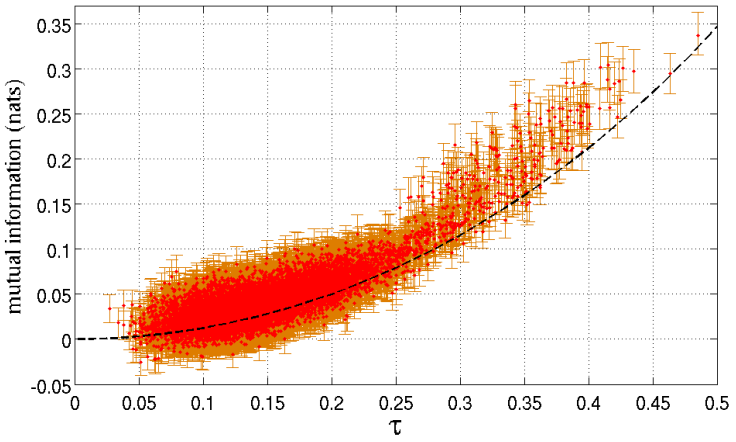
$$\tilde{\tau}[X_i, X_j] = \binom{N}{2}^{-1} \sum_{\mu<\nu} \operatorname{sign}(x^i_\mu - x^i_\nu)\operatorname{sign}(x^j_\mu - x^j_\nu) \tag{2.45}$$

onde a soma é feita sobre todos os pares (μ, ν). Um método para estimação da informação mútua foi recentemente publicado por Kraskov et al. [27], baseado em estatísticas de k-vizinhos. Medindo I e τ é possível diagnosticar um eventual "excesso" de informação mútua que, caso presente, indica que a dependência não é mínima e tem uma componente não-linear. Nesses casos a correlação linear não pode ser usada como medida de dependência. Como exemplo desse diagnóstico, mostramos na figura 2.3 estimativas para essas quantidades para séries temporais de uma seleção de 150 das 500 ações que compõe o índice S&P500, um índice de ações de alta capitalização negociadas em bolsas da NYSE Euronext e da NASDAQ OMX definido e mantido pela Standard & Poor's. As barras de erro para a informação mútua, que representam um intervalo de confiança de 90%, foram calculadas usando o método de bootstrap, repetindo a estimativa do algoritmo KSG[21] para diversas amostragens com repetição dos dados. O tau de Kendall foi também estimado usando procedimentos padrão como o da eq.(2.45). Observa-se que uma boa quantidade de pontos apresentam, dentro do intervalo de confiança, um valor para informação mútua não compatível com uma cópula gaussiana. Isso sugere que técnicas de avaliação e administração de risco baseadas no uso de cópulas gaussianas podem subestimar de forma substancial a dependência entre duas ações e o grau de co-movimento que elas apresentam. Na figura 2.4 por exemplo, mostramos alguns pares de ações com correlações muito pequenas ($\rho < 0.1$) e que no entanto apresentam informação mútua compativel dependência muito maior do que a capturada por essa medida linear.

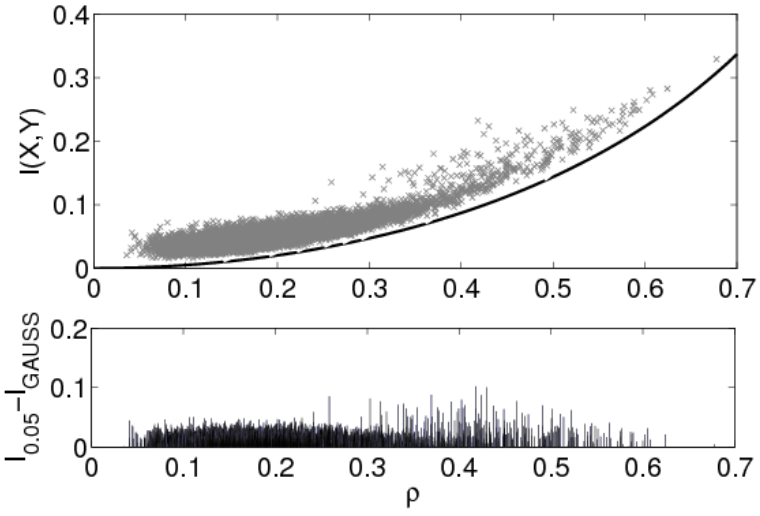
[21] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E*, 69:066138, 2004

⌈2.5.4⌉ Ajuste empírico de cópulas via informação mútua

A determinação de uma particular cópula para realizar essas medidas de risco poderia ser uma alternativa ao uso cego da cópula



(a) Todos os pares de ações.



(b) Seleção dos pares cujo limite inferior da barra de erro é maior que a informação mútua gaussiana — excesso de informação não nulo com 90% de confiança (as barras de erro foram removidas para melhor visualização).

Figura 2.3 – Estimativas para a informação mútua usando o algoritmo KSG contra o tau de Kendall (ou correlação medida via tau de Kendall) para pares de séries temporais de log-retornos diários $\log \frac{P_{\text{close}}}{P_{\text{open}}}$ (onde P_{close} e P_{open} são, respectivamente, preços dos ativos na abertura e fechamento diários do mercado) para 150 das ações mais negociadas que compõe o índice *S&P*500, no período de 02/01/1990 a 16/09/2008 (aproximadamente 4700 pontos por série). As barras de erro representam intervalos de confiança de 90% determinados segundo o procedimento de Bootstrap. Note que, nesse intervalo de confiança, um grande número de pares apresentam um excesso de informação mútua não-nulo com respeito à cópula gaussiana. As linhas tracejadas indicam o limite gaussiano para a informação mútua.

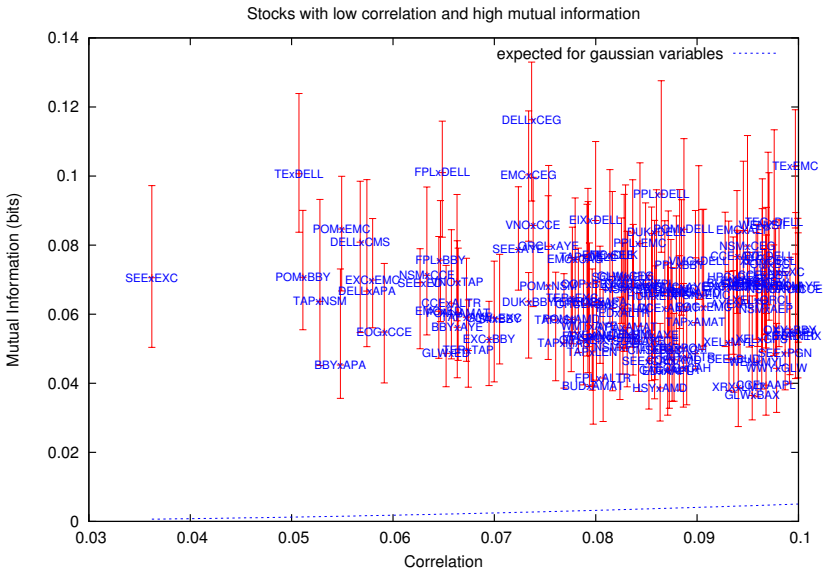


Figura 2.4 – Seleção de pares de ações com baixa correlação e grande excesso de informação mútua e pares compatíveis com uma distribuição gaussiana. Modelos de risco baseados em medidas de correlação linear devem ser adequados para o segundo grupo e devem falhar seriamente para o primeiro.

(a) Seleção de pares com baixa correlação e grande excesso de informação mútua

(b) Seleção de pares compatíveis com uma distribuição gaussiana.

gaussiana. Uma possível forma de realizar essa determinação é minimizar a divergência de Kullback-Leibler entre a cópula empírica e uma família $\mathcal{C}_\theta$ de cópulas parametrizadas por um conjunto de parâmetros θ:

$$D[C(\boldsymbol{u})||C(\boldsymbol{u}|\theta)] = \int \mathrm{d}^N u \; c(\boldsymbol{u}) \log \frac{c(\boldsymbol{u})}{c(\boldsymbol{u}|\theta)} \tag{2.46}$$

A minimização desse funcional com respeito a θ fornece uma possível cópula $c(\boldsymbol{u}|\theta^\star)$ que é a mais próxima possível da cópula real dentro dessa família. Essa expressão pode ser manipulada da seguinte forma:

$$D[\cdot|\cdot] = \int \mathrm{d}^N u \; c(\boldsymbol{u}) \log c(\boldsymbol{u}) - \int \mathrm{d}^N u \; c(\boldsymbol{u}) \log c(\boldsymbol{u}|\theta) \tag{2.47}$$

$$= I - L_\infty(\theta) \tag{2.48}$$

onde o primeiro termo é o negativo da entropia de cópula, igual à informação mútua, como mostrado anteriormente, e o segundo termo é o valor assintótico da log-verossimilhança quando o número de amostras é grande:

$$L_N(\theta) = \frac{1}{N} \sum_{\mu=1} N \log c(\boldsymbol{u}_\mu|\theta)$$

Uma vez que $D[|] \geq 0$ e $I \geq 0$, então minimizar a divergência de Kullback-Leibler, cujo menor valor possível é nulo, é equivalente a maximizar a log-verossimilhança com a informação mútua como limite. Se a log-verossimilhança fosse conhecida analíticamente isso poderia ser feito de maneira imediata resolvendo a equação:

$$L_\infty(\theta) = I$$

Para um I determinado empiricamente a partir dos dados,numericamente se necessário. Não há garantia alguma, no entanto de que há solução. Apenas haverá solução para essa equação se a própria cópula original fizer parte da família $\mathcal{C}_\theta$. Nesse caso a solução é única. Além disso não é possível conhecer L_∞ analiticamente sem conhecer a cópula e uma aproximação é necessária. Supondo que a família de cópulas $\mathcal{C}_\theta$ é suficientemente próxima da cópula original, podemos substituir $c((u))$ pela própria $c(\boldsymbol{u}|\theta)$ na integral e aproximar $L_\infty(\theta)$ pela informação mútua de $c(\boldsymbol{u}|\theta)$. Dessar forma ficamos com uma espécie de método de "correspondência de momentos" (moment matching): deve ser escolhida na família $\mathcal{C}_\theta$ a cópula que tem mesma informação mútua que a empiricamente obtida:

$$I(\theta) = I.$$

Novamente, não há garantia de solução, e agora nem mesmo da unicidade da solução. Mas a cópula escolhida certamente será capaz de descrever uma estrutura de dependência mais complexa do que a descrita pela cópula gaussiana. Se a família escolhida for um subconjunto

da famíla de cópulas elípticas, o procedimento pode ser ainda melhorado. Parte do conjunto de parâmetros θ corresponde às correlações Σ. Escrevendo $I(\theta) = I_0(\Sigma) + \Delta I(\theta')$ pode-se já eliminar a parte linear da informação mútua usando o tau de Kendall para calcular Σ e ajustar o excesso de informação mútua ao medido empiricamente.

⌈2.5.5⌉ Cópula t

Como exemplo desse procedimento vamos escolher uma subfamilia das cópulas elípticas, as cópulas t. Essas cópulas, como discutido anteriormente, são as cópulas que se originam da distribuição t de Student. Quando se permite que o parâmetro ν seja contínuo, essa é, em essência, a mesma distribuição obtida pela maximização da chamada entropia de Tsallis, e essa distribuição e sua cópula associada têm recebido certa atenção na literatura de análise financeira por seu bom ajuste empirico a dados de diversas naturezas[22]. O cálculo do excesso de informação mútua da distribuição t de Student se resume a calcular a eq.(2.42) — a informação mútua da distribuição t de Student esférica:

$$p(\mathbf{t} \mid \hat{\Sigma}, \nu) = \frac{1}{Z_N(\nu)} \left[1 + \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}{\nu}\right]^{-\frac{\nu+N}{2}} \tag{2.49}$$

onde a normalização é dada por[23]:

$$Z_N(\nu) = \int \mathrm{d}^N x \, \left[1 + \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}{\nu}\right]^{-\frac{\nu+N}{2}} = \frac{B(\frac{\nu}{2}, \frac{N}{2})\sqrt{(\pi\nu)^N}}{\Gamma(\frac{N}{2})} \tag{2.50}$$

Uma vez que todas as marginais são idênticas[24] a informação mútua se reduz a:

$$I(\nu) = NH_1(\nu) - H_N(\nu), \tag{2.51}$$

onde $H_n(\nu)$ é a entropia de uma distribuição de student n-dimensional. Para calcular H_n note que:

$$\begin{aligned}
H_n(\nu) &= -\int \mathrm{d}^n x \frac{1}{Z_n(\nu)} \left(1 + \frac{\boldsymbol{x}^T\boldsymbol{x}}{\nu}\right)^{-\frac{\nu+n}{2}} \left[-\frac{\nu+n}{2}\log\left(1 + \frac{\boldsymbol{x}^T\boldsymbol{x}}{\nu}\right) - \log Z_n(\nu)\right] \\
&= \log Z_n(\nu) + \frac{\nu+n}{2Z_n(\nu)} \int \mathrm{d}^n x \left(1 + \frac{\boldsymbol{x}^T\boldsymbol{x}}{\nu}\right)^{-\frac{\nu+n}{2}} \log\left(1 + \frac{\boldsymbol{x}^T\boldsymbol{x}}{\nu}\right) \\
&= \log Z_n(\nu) + \frac{(\nu+n)\Omega_n}{2Z_n(\nu)} \int_0^\infty \mathrm{d}r r^{n-1} \left(1 + \frac{r^2}{\nu}\right)^{-\frac{\nu+n}{2}} \log\left(1 + \frac{r^2}{\nu}\right) \\
&= \log Z_n(\nu) + \frac{(\nu+n)\Omega_n\sqrt{\nu^n}}{2Z_n(\nu)} \int_0^\infty \mathrm{d}u u^{n-1} \left(1 + u^2\right)^{-\frac{\nu+n}{2}} \log\left(1 + u^2\right) \\
&= \log Z_n(\nu) + \frac{(\nu+n)\Omega_n\sqrt{\nu^n}}{2Z_n(\nu)} R_n(\nu)
\end{aligned}$$

[22] Stefano Demarta and Alexander J. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1):111–129, 2005; and Lisa Borland. Option pricing formulas based on a non-gaussian stock price model. *Phys. Rev. Lett.*, 89(9): 098701, Aug 2002. DOI: 10.1103/PhysRevLett.89.098701

[23] $B(a, b)$ é a função beta, dada por:

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} \, \mathrm{d}t$$

[24] Todas são iguais à distribuição t de Student em uma dimensão com o mesmo parâmetro ν

Onde $\Omega_n = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}$ é a área de uma esfera unitaria em n dimensões e

$$R_n(\nu) = \int_0^\infty \mathrm{d}u u^{n-1} \left(1+u^2\right)^{-\frac{\nu+n}{2}} \log\left(1+u^2\right).$$

Essa integral pode ser feita com o auxílio de um truque similar ao truque de réplicas comum em mecânica estatística. Se notarmos que $\log(x) = \lim_{r\to 0} \frac{\partial}{\partial r} x^r$ podemos escrever[25]:

$$R_n(\nu) = \lim_{r\to 0} \frac{\partial}{\partial r} B\left(\frac{\nu}{2} - r, \frac{n}{2}\right) = -B\left(\frac{\nu}{2}, \frac{n}{2}\right)\left[\psi\left(\frac{\nu+n}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right]$$

onde $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ é a função beta e $\psi(x)$ é a função digamma, substituindo esse resultado na expressão original e escrevendo todos os termos explicitamente temos:

$$H_n(\nu) = \log\left[\frac{\sqrt{(\pi\nu)^n} B\left(\frac{\nu}{2}, \frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\right] + \frac{\nu+n}{2}\left[\psi\left(\frac{\nu+n}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right] \tag{2.52}$$

Para o caso $n = 1$ isso se reduz a:

$$H_1(\nu) = \log\left[\sqrt{\nu} B\left(\frac{\nu}{2}, \frac{1}{2}\right)\right] + \frac{\nu+1}{2}\left[\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right] \tag{2.53}$$

E a informação mútua $I = NH_1(\nu) - H_N(\nu)$ finalmente pode ser escrita:

$$I = \log\left\{\frac{\left[B\left(\frac{\nu}{2}, \frac{1}{2}\right)\right]^N \Gamma\left(\frac{N}{2}\right)}{\pi^{\frac{N}{2}} B\left(\frac{\nu}{2}, \frac{N}{2}\right)}\right\} - \frac{\nu(N-1)}{2}\psi\left(\frac{\nu}{2}\right) + \frac{N(\nu+1)}{2}\psi\left(\frac{\nu+1}{2}\right) - \frac{\nu+N}{2}\psi\left(\frac{\nu+N}{2}\right).$$

Para $N = 2$ isso se reduz a:

$$I(\nu) = 2\log\left(\sqrt{\frac{\nu}{2\pi}} B\left(\frac{\nu}{2}, \frac{1}{2}\right)\right) - \frac{2+\nu}{\nu} + (1+\nu)\left[\psi\left(\frac{\nu+1}{2}\right) - \psi\left(\frac{\nu}{2}\right)\right] \tag{2.54}$$

[25] Pois:
$\int_0^\infty \mathrm{d}u u^{n-1}\left(1+u^2\right)^{-\alpha} = B\left(\alpha - \frac{n}{2}, \frac{n}{2}\right)$,
$\frac{\partial}{\partial x} B(x,y) = -B(x,y)(\psi(x+y) - \psi(x))$

Finalmente, a expressão acima pode ser usada, segundo o método discutido na seção 2.5.4, para ajustar cópulas t a dados empíricos. Como ilustração, a figura 2.5 apresenta uma série de simulações de ajuste com dados sorteados de uma cópula t com diversos valores de correlação e ν conhecidos. O excesso de informação mútua é estimado usando o algoritmo KSG [26] e o tau de Kendall e plotado em função do ν conhecido. A linha cheia corresponde à eq.(2.54). Esse gráfico mostra que, exceto por um ponto que não pode ser recuperado[27], é possível recuperar cópulas t a partir de dados experimentais através desse procedimento.

[26] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E*, 69:066138, 2004

[27] Para um valor muito pequeno de ν, para o qual a distribuição t começa a apresentar diversas patologias, como variância infinita.

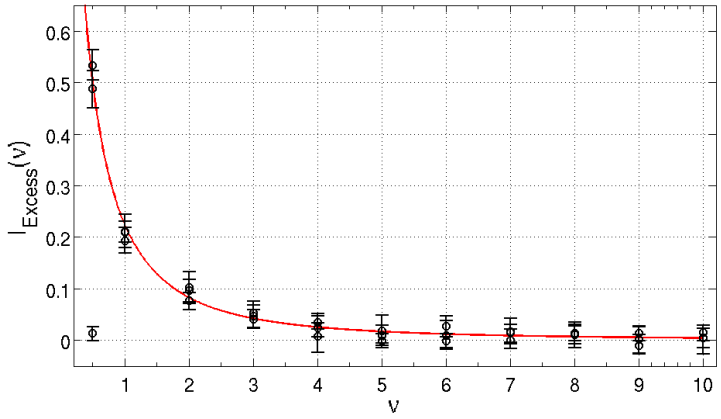


Figura 2.5 – Excesso de informação mútua na cópula t. $I(\nu)$, como dado na eq.(2.54). Círculos mostram estimativas para 20 amostragens de pontos de uma cópula t usando o método de "moment matching" para diversos valores de correlação e ν.

[2.6] Conclusões

A literatura em teoria de informação e teoria de cópulas e dependência estatística — ambas com décadas de existência — se desenvolveram em relativo isolamento, com apenas pontos muito recentes de contato. Neste trabalho tentamos discutir as consequencias de alguns desses pontos e conexões entre os dois tópicos. A teoria de cópulas pode ser usada para decompor as distribuições conjuntas em flutuações idiossincráticas das marginais de cada variável e flutuações devidas ao acoplamento entre as variáveis. A conexão com a teoria de informação permite levar adiante essa decomposição e isolar contribuições à informação total contida na distribuição em partes relativas às marginais, ao acoplamento de natureza linear e acoplamentos de mais alta ordem. Essa decomposição oferece testes simples a respeito da linearidade e "gaussianidade" do acoplamento e também sugere um método de ajuste de cópulas baseados no ajuste da informação mútua. Essa abordagem também clarifica os perigos do uso

da correlação linear como medida de dependência em séries financeiras para, por exemplo, estimativas de riscos de contratos complexos e otimização de carteiras, pois essa medida é fadada a subestimar a dependência em séries em que flutuações não-gaussianas são esperadas. Finalmente, pensamos que uma conexão entre essas duas áreas — teoria de informação e teoria de dependência estatística — pode ser útil em fornecer conceitos e técnicas novas para o estudo de sistemas complexos.

⌈3⌉ Um Modelo para emergência de autoridade em sociedades humanas.

> The resemblance between humans and chimpanzees is not only external. If we look straight and deep into a chimpanzee's eyes, an intelligent self-assured personality looks back at us. If they are animals, what must we be?
>
> — "Chimpanzee Politics: Power and Sex among Apes" de Frans de Waal[30]

> There isn't a sharp line dividing humans from the rest of the animal kingdom — it's a very wuzzy line — it's a very wuzzy line — and it's getting wuzzy-er all the time...
>
> — TED Talk "What separates us from chimpanzees" de Jane Goodall[31]

⌈3.1⌉ Introdução

Os grandes primatas, em particular os humanos, apresentam vidas sociais intensas. Atividades sociais, formação de coalizões, cooperação para realização de diversas tarefas, guerras, compartilhamento de alimentos, disputas por liderança *et cetera*, situações ubíquas em agrupamentos humanos, não são entretanto limitadas a essa espécie mas são pervasivas em todas as espécies desse grupo. A variedade das experiências sociais em espécies como chimpanzés e bonobos são comparáveis apenas à da espécie humana[1]. Entretanto a natureza dessa experiência social pode ser bem diferente entre humanos e outras espécies aparentadas. Há diferentes pontos de equilíbrio na divisão dos recursos e na tomada de risco por diferentes membros do grupo em

[1] F. B. M. de Waal. *Chimpanzee Politics: Power and Sex among Apes*. Chimpanzee Politics. Johns Hopkins University Press, 2007. ISBN 9780801886560. URL http://books.google.com.br/books?id=XsrhU2vV5PIC; and F. B. M. de Waal. *Peacemaking Among Primates*. Harvard University Press, 1990. ISBN 9780674659216. URL http://books.google.com.br/books?id=5at0TN8U6RMC

atividades coletivas, como coleta de alimentos, defesa contra predadores, cuidado com ninhadas, etc[2]. Há grupos — ditos igualitários — em que recursos são distribuidos de forma igualitária ou de maneira aproximadamente proporcional ao risco tomado, e há grupos — ditos hierárquicos, ou despóticos — em que poucos membros do grupo têm acesso desproporcional ao fruto dessas atividades, em detrimento dos outros membros. Na natureza se observa um contínuo de variações de comportamento entre esses dois campos extremos.

A MAIORIA DAS ESPÉCIES apresenta um comportamento social básico razoavelmente fixo, ainda que flexível [34]. Os grandes primatas tipicamente vivem em sociedades hierárquicas, marcadas por uma grande concentração dos usos de recursos energéticos e reprodutivos por parte de poucos membros do grupo. Por outro lado, humanos se destacam por uma grande variabilidade nesse espectro etológico social. Certos grupos humanos apresentam organização fortemente centralizada e hierárquica, com concentração de uso de recursos e riqueza. Outros grupos apresentam sociedades basicamente equalitárias, com compartilhamento de recursos e alimentos, ausência de distinções de status, autoridades ou concentração de riqueza[3]. O quebra-cabeças associado à origem dessa variabilidade é uma matéria de ativa discussão acadêmica e, a intenção deste trabalho é tentar lançar luz sobre o problema através de um modelo matemático com origem em certas observações empíricas recentes.

⌈3.2⌉ Evidências empiricas

⌈3.2.1⌉ "U-shaped evolution"

O REGISTRO ARQUEOLÓGICO REVELA uma dinâmica temporal na organização social dos humanos através da pré-história[4]. Os humanos, descendendo de primatas com provável organização social hierárquica, passaram por um período de grupos equalitários, sem autoridade central, com baixa densidade populacional. No neolítico houve uma transição para grupos fortemente hierárquicos, conforme a densidade populacional aumenta após a revolução agrícola. Esse quadro, ilustrado pela figura 3.1, é denominado em [35] *"a U-shaped evolution"*.

HOJE, SOCIEDADES HUMANAS preenchem todo o espectro de organização social e evidências etnográficas e arqueológicas também apontam para uma relação entre o tamanho dos grupos de humanos caçadores-coletores e suas formas de organização social[5]. Grupos pequenos de humanos tendem a apresentar organização equalitária, sem concen-

[2] Sandra L. Vehrencamp. A model for the evolution of despotic versus egalitarian societies. *Animal Behaviour*, 31(3):667 – 682, 1983. ISSN 0003-3472. DOI: 10.1016/S0003-3472(83)80222-X. URL http://www.sciencedirect.com/science/article/pii/S000334728380222X; and C Boehm. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Harvard University Press, 2001. ISBN 9780674006911. URL http://books.google.com.br/books?id=ljxS8gUlgqgC

[3] Note que nos interessamos neste trabalho por agrupamentos de humanos caçadores-coletores que podem trazer mais insights sobre humanos em seu contexto evolutivo.

[4] Bruce M. Knauft, Thomas S. Abler, Laura Betzig, Christopher Boehm, Robert Knox Dentan, Thomas M. Kiefer, Keith F. Otterbein, Tohn Paddock, and Lars Rodseth. Violence and sociality in human evolution [and comments and replies]. *Current Anthropology*, 32:391–428, 1991. URL http://www.jstor.org/stable/2743815; Joyce Marcus. The archeological evidence for social evolution. *Annu. Rev. Anthropol*, 37:25166, 2008; and R. Foley. *Another unique species: patterns in human evolutionary ecology*. Longman Scientific & Technical, 1987. ISBN 9780470207284. URL http://books.google.com.br/books?id=AjuAAAAAMAAJ

[5] Thomas E. Currie, Simon J. Greenhill, Russell D. Gray, Toshikazu Hasegawa, and Ruth Mace. Rise and fall of political complexity in island south-east asia and the pacific. *Nature*, 467(7317):801–804, Oct. 2010. DOI: 10.1038/nature09461; and Douglas J. Kennett and Bruce Winterhalder. *Islands of Inquiry: Colonisation, seafaring and the archaeology of maritime landscapes*, chapter Demographic expansion, despotism and the colonisationof East and South Polynesia, pages 87–96. In Clark et al. [57], 2008. ISBN 9781921313905. URL http://books.google.com.br/books?id=NZ0-IiAZ_ZwC

tração de poder. Grupos maiores tendem a apresentar organizações hierárquicas, concentração e hereditariedade de poder.

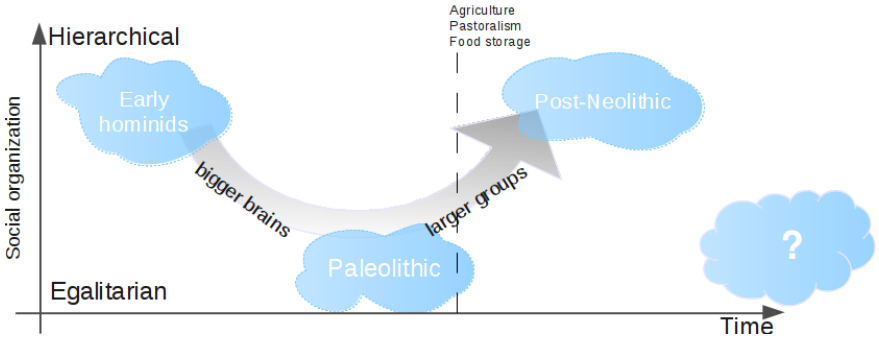


Figura 3.1 – Ilustração da história da organização social dos humanos e primatas pré-humanos.

[3.2.2] Evolução do cérebro primata e a Teoria Maquiavélica

Estudos recentes[6] relacionam o tamanho relativo de regiões do cérebro de diversas espécies de primatas a medidas relacionadas com a capacidade social da espécie, como tamanho dos grupos em que vivem, o tamanho de coalizões, número médio de indivíduos que interagem diretamente, etc. O que é típicamente encontrado é ilustrado na figura 3.2. Essa figura mostra um gráfico do tamanho médio do grupo em função da razão média entre o volume do neocórtex, região do cérebro envolvida no planejamento, raciocínio espacial, linguagem, entre outras funções cognitivas complexas, e o volume total do cérebro para diversas espécies de primatas. O gráfico sugere uma relação do tipo lei de potência entre as duas grandezas, similar a encontradas em diversas outras comparações desse tipo.

[6] T Sawaguchi and H Kudo. Neocortical development and social structure in primates. *Primates*, pages 283–90, 1990; Robin Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 20:469–93, 1992; R. I. M. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16: 681–694, 11 1993. ISSN 1469-1825. DOI: 10.1017/S0140525X00032325. URL `http://dx.doi.org/10.1017/S0140525X00032325`; Leslie C. Aiello and R. I. M. Dunbar. Neocortex size, group size, and the evolution of language. *Current Anthropology*, 34(2):184–193, 1993; and Tracey H. Joffe and R. I. M. Dunbar. Visual and socio–cognitive information processing in primate brain evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1386):1303–1307, 1997. DOI: 10.1098/rspb.1997.0180. URL `http://rspb.royalsocietypublishing.org/content/264/1386/1303.abstract`

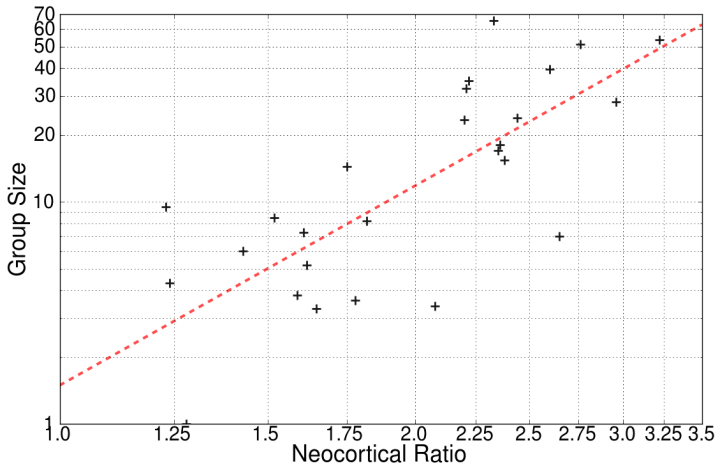


Figura 3.2 – Gráfico em escala dilogaritimica do tamanho médio do grupo em função da razão média entre o volume do neocórtex e o volume total do cérebro para diversas espécies de primatas. Dados disponíveis em [45].

Em essência, essa relação sugere que a capacidade cognitiva dos primatas está relacionada à sua necessidade de lidar com interações sociais complexas. Essa hipótese, denominada Teoria Maquiavélica ou Teoria do Cérebro Social, sugere um cenário em que o rápido crescimento na importância relativa do neocórtex é uma resposta a uma pressão seletiva associada a essa necessidade de interação social[7]. Ambientes sociais complexos favorecem reprodutivamente individuos com maior capacidade de processar informações sociais, que, por sua vez, torna as estratégias sociais ainda mais complexas em um "corrida armamentista" evolutiva. Essa hipótese tem recebido crescente suporte empírico e aprofundamento teórico[8].

[7] R W Byrne and A Whiten. Tactical deception in primates: the 1990 database. *Primate Report, Whole Volume*, 27:1–101, 1990; R W Byrne and A Whiten. Computation and mindreading in primate tactical deception. In A Whiten, editor, *Natural Theories of Mind*. Blackwells, Oxford, 1991; R W Byrne and A Whiten. Cognitive evolution in primates: evidence from tactical deception. *Man*, 27:609–627, 1992; and Andrew Whiten and David Erdal. The human socio-cognitive niche and its evolutionary origins. *Phil. Trans. R. Soc. B*, 367:2119–2129, 2012

[8] R. I. M. Dunbar. Cognitive constraints on the structure and dynamics of social networks. *Group Dynamics: Theory, Research, and Practice*, 12(1): 7–16, 2008; R. I. M. Dunbar. The social brain hypothesis and its implications for social evolution. *Annals of human biology*, 36(5):562–72, 2009. DOI: 10.1080/03014460902960289; and Susanne Shultz and Robin Dunbar. Encephalization is not a universal macroevolutionary phenomenon in mammals but is associated with sociality. *Proceedings of the National Academy of Sciences*, 107(50):21582–21586, 2010. DOI: 10.1073/pnas.1005246107. URL `http://www.pnas.org/content/107/50/21582.abstract`

[3.2.3] Teoria da Reversão de Dominância

Descrita por Boehm et al. [52] a partir de uma investigação antropológica e dados primatológicos, a Teoria da Reversão de Dominância (*Reverse Dominance Theory*) é um mecanismo explicativo para o estabelecimento de comportamento igualitário em espécies com forte comportamento de dominância. Segundo essa teoria, o comportamento igualitário surge em humanos caçadores-coletores não pela ausência de um mecanismo de dominância, mas através de uma resistência sistemática dos membros do grupo a comportamentos excessivamente dominantes por parte de membros do grupo. O comportamento igualitário é garantido por uma inversão do mecanismo de dominância: candidatos a líder são dominados pelos outros membros do grupo, que

limitam os mecanismos de ascensão através da expressão de opinião pública, crítica e ridicularização, disobediência e, em última instância, sanções extremas como assassinato[9].

[9] C Boehm. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Harvard University Press, 2001. ISBN 9780674006911. URL http://books.google.com.br/books?id=ljxS8gUlgqgC; and Christopher Boehm, Harold B. Barclay, Robert Knox Dentan, Marie-Claude Dupre, Jonathan D. Hill, Susan Kent, Bruce M. Knauft, Keith F. Otterbein, and Steve Rayner. Egalitarian behavior and reverse dominance hierarchy [and comments and reply]. *Current Anthropology*, 34(3): pp. 227–254, 1993. ISSN 00113204. URL http://www.jstor.org/stable/2743665

[3.3] Um modelo mecanico-estatístico baseado em agentes

[3.3.1] Descrição dos agentes

O objetivo dessa seção é descrever um modelo estatístico para a formação de estrutura social que seja tratável por técnicas comuns à mecânica estatística e teoria de informação e compatível com as informações experimentais descritas acima. O primeiro passo na descrição do modelo consiste na elaboração de uma dinâmica de comportamento para um conjunto de agentes hipotéticos, que será a dinâmica temporal microscópica que dará origem a um modelo mecânico-estatístico.

Considere um grupo de n agentes dotados de certa capacidade cognitiva limitada e engajados em atividades sociais. Cada agente carrega um registro mental da informação que possui a respeito das relações sociais entre os membros do seu grupo. Essa informação está relacionada a como se relaciona socialmente cada par de outros agentes do grupo. Essa informação deve responder perguntas como:

- qual a possibilidade de um certo par de individuos serem adversários em uma disputa ou aliados em uma coalizão?
- com que frequencia cooperam em uma atividade conjunta?
- como compartilham seus recursos um com o outro?
- etc...

Cada agente adquire essa informação através de mecanismos diversos: através da história do grupo, baseado no comportamento pregresso dos agentes; através de mecanismos de aprendizado social como "fofoca" (*gossip*) em que a comunicação com outros agentes permite que o agente aprenda sobre experiências de outros; etc. Uma vez adquirida, essa informação é critica para subsidiar decisões sociais a serem tomadas pelo agente: com que grupo de agentes formar uma coalizão, quando esperar cooperação de um certo indivíduo, quando esperar reciprocação ao compartilhar recursos etc. Cometer erros nessas decisões podem custar recursos e posição social, e influenciar negativamente a capacidade reprodutiva do agente. Portanto, espera-se que uma espécie de agentes que tenha surgido por evolução via seleção natural

tenha mecanismos cognitivos adequados para tentar minimizar esses erros em algum sentido.

A AQUISIÇÃO DESSE TIPO DE INFORMAÇÃO SOCIAL é uma atividade cognitivamente custosa. A capacidade limitada de processar essas informações implica que para mantê-las atualizadas, o indivíduo precisa desviar recursos que poderiam ser aplicados em outras atividades — coleta de alimentos, construção de abrigos, etc. O rastreamento de relações sociais demanda um considerável tempo dos individuos adultos em tribos de chimpanzés e humanos[10] e as evidências da Teoria do Cérebro Social de que a habilidade de navegar informações sociais consiste em uma pressão seletiva importante indica a intensidade desse custo. Conforme se aumenta o tamanho do grupo, esse custo cresce com o número de ligações sociais possíveis e, portanto, quadraticamente com o número de indivíduos. Isso pode pode tornar a estratégia de adquirir e manter informações sobre todas as ligações sociais possíveis no grupo pouco adaptativa. Pode ser preferível ao agente nesse caso obter apenas informação sobre certas ligações sociais importantes e fiar-se em heurísticas para inferir as outras relações sociais (regras como "amigo do amigo é amigo", etc...).

[10] F. B. M. de Waal. *Chimpanzee Politics: Power and Sex among Apes*. Chimpanzee Politics. Johns Hopkins University Press, 2007. ISBN 9780801886560. URL http://books.google.com.br/books?id=XsrhU2vV5PIC

DESSA FORMA, O MODELO CONSISTIRÁ dos seguintes elementos: muitos *agentes* que interagem entre si, cada um carregando uma *representação mental da estrutura social* do grupo a que pertence e individualmente tentando *minimizar custos* associados a carregar esass informações sociais. Abaixo discutiremos uma representação matemática de cada um desses elementos.

⌈3.3.2⌉ VARIÁVEIS DINÂMICAS — GRAFOS SOCIAIS

A REALIZAÇÃO MATEMÁTICA da representação mental da estrutura social que cada agente carrega será feita através de grafos. A cada agente está associado um grafo cujos nós representam todos os indivíduos do grupo e arestas representam as relações sociais sobre as quais o agente possui informação. A informação será considerada binária: o agente pode ter certeza sobre a relação social entre dois individuos, havendo portanto uma aresta ligada entre os nós correspondentes de seu grafo, ou não tem nenhuma informação direta sobre ela, caso em que não haverá uma aresta enter os nós correspondentes. As arestas do grafo portanto podem estar apenas ligadas ou desligadas, sem estados intermediários. As arestas do grafo são entidades dinâmicas, que podem ser criadas quando o agente adquire informação sobre uma relação social anteriormente desconhecida, ou destruídas quando o

agente, por alguma razão, desiste de continuar mantendo aquela informação. Dados os custos, que serão discutidos abaixo, o agente deverá decidir quais informações valem a pena ser guardadas ou não. Um agente que possui informação completa sobre todas as relações sociais do grupo tem um grafo totalmente conectado, como o representado na figura 3.3.

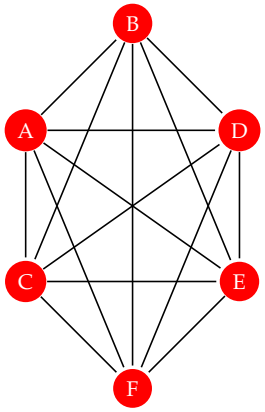


Figura 3.3 – Exemplo de grafo social — um grafo completamente conectado. Um agente com essa estratégia despende recursos para conhecer todas as relações sociais do grupo. Um grafo como esse possui $\frac{1}{2}n(n-1)$ arestas, onde n é o número de agentes.

Quando uma aresta é faltante no grafo carregado por um agente, a informação social correspondente à essa aresta é incompleta, e o agente deve recorrer a heuristicas para determinar quaisquer informações necessárias. Para tal, vamos considerar que o grafo deve ser conexo — deve ser possível, para todos os pares de nós, encontrar um *caminho* de arestas ligadas conectando os dois nós. Dessa forma, sempre é possível a um agente determinar alguma informação indireta entre dois nós desconexos, através de uma heuristica que utilize as outras arestas conhecidas. Um exemplo é o grafo da figura 3.4, que representa um grafo do tipo estrela.

Nesse grafo, não é possível conhecer diretamente todas as relações sociais pois apenas $n-1$ das $\frac{1}{2}n(n-1)$ arestas possíveis está presente. Mas todas as relações sociais podem ser indiretamente determinadas por heurísticas sobre caminhos de comprimento 2 ("amigo do amigo é amigo", etc...). Os grafos mais esparsos possíveis que ainda são conexos possuem $n-1$ arestas, no mínimo.

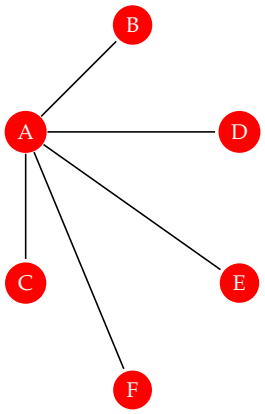


Figura 3.4 – Exemplo de grafo social — um grafo estrela. Um agente com essa estratégia despende recursos para conhecer apenas as relações envolvendo um certo indivíduo central (o nó A na figura). As outras relações são determinadas através de regras heurísticas. Esse grafo possui $n-1$ arestas.

[3.3.3] Custos

Em nosso modelo existem custos associados à manutenção de um certo grafo de informações sociais. Há dois tipos de custos:

1. O custo cognitivo de adquirir e manter informação dos diferentes pares de individuos. Como discutido anteriormente, se o agente tem recursos cognitivos limitados, manter essas informações é custoso. Se assumirmos que o investimento de recursos para obter informações sobre cada aresta do grafo é constante, o custo cognitivo total por agente deverá ser proporcional ao número de arestas do grafo:

$$H_{\text{cognitivo}} \propto n_e \tag{3.1}$$

   onde n_e é o número de arestas (*edges*).

2. O custo social de falhar em determinar corretamente a relação entre dois indivíduos. As heuristicas utilizadas pelo agente quando

ele não possui informação direta sobre uma certa relação social podem falhar e, nesse caso, o agente pode inferir erroneamente a relação social entre dois agentes. Ao tomar decisões baseadas nessa avaliação errônea, o agente incorre em custos — o agente pode avaliar incorretamente em que lado de uma disputa um individuo vai se posicionar, falhar em reconhecer uma coalizão em formação, etc, e ter prejuízos reais com uma situação social inesperada. As heuristicas que se valem de relações conhecidas para inferir relações desconhecidas serão tão mais confiáveis quanto menor o caminho a ser percorrido no grafo entre os dois nós em questão. Quanto mais longos os trajetos a serem percorridos, maior é a probabilidade de erro. Portanto, o custo social esperado deve ser proporcional à distância geodésica média entre os nós do grafo:

$$H_{\text{social}} \propto \frac{2}{n(n-1)} \sum_{i<j} L_{ij} \tag{3.2}$$

onde L_{ij} é a distância geodésica entre os nós i e j, e a soma é realizada sobre todos pares de agentes.

Assim, para cada agente, o custo total de se manter uma certa representação mental da rede social do grupo é dado, portanto, por:

$$H = \frac{n_e}{\alpha} + \bar{L} \tag{3.3}$$

Onde $\bar{L}$ é a distancia geodésica média do grafo, α é uma constante associada à importância relativa entre o custo cognitivo e o custo social (quanto maior α, menos importante é o custo cognitivo). Note que tipicamente n_e escala como uma fração do número total de arestas possíveis $n(n-1)/2$ e que $\bar{L}$ escala tipicamente com $\log(n)$ para grafos aleatórios. É, portanto, interessante, reescrever (3.3), a menos de uma constante multiplicativa, como:

$$H = \phi_e + a\bar{L} \tag{3.4}$$

onde $\phi_e = \frac{n_e}{n(n-1)/2}$ é a fração de arestas ocupadas e $a = \frac{2\alpha}{n(n-1)}$ é uma constante de acoplamento normalizada.

[3.3.4] Dinâmica para agentes isolados — máxima entropia

[3.3.5] Minimização do custo

Para que um modelo possa ser estabelecido, não basta a expressão para o custo, mas uma descrição de que ações o agente deverá tomar com base em sua aferição do custo. A estratégia do agente é

definida por uma certa escolha de arestas a investir. Em uma primeira abordagem, atribuimos ao agente a tendência a usar a estratégia que minimiza o custo total. Então deveriamos procurar pelo grafo definido por:

$$G_\star(\alpha) = \arg\min_G H(G, \alpha) \tag{3.5}$$

Se $\alpha \gg 1$, então as limitações cognitivas são menos e menos importantes. O grafo que minimiza o custo é o grafo que minimiza a distância geodésica média — um grafo totalmente conectado como o da figura 3.3. Se $\alpha \ll 1$, então as limitações cognitivas se tornam mais e mais importantes, e o grafo ótimo é aquele que minimiza o número de arestas, enquanto ainda mantendo uma distância média finita — um grafo em forma de estrela [11] como o da figura 3.4. Para valores intermediários de α, o grafo ótimo possui configurações intermediárias entre esses dois extremos.

[11] Estritamente para $\alpha = 0$ o grafo ótimo é o grafo sem aresta alguma. Não há prescrição canônica para a distância geodésica média de um grafo sem arestas. Adotaremos aqui a convenção de que se não é possível desenhar um caminho entre dois nós (em outras palavras: dois nós que pertencem a diferentes componentes do grafo), então a distância entre eles é infinita. Isso restringe nossa análise apenas a grafos conexos. Mesmo entre os grafos conexos, a estrela não é o único mínimo do custo acima para α estritamente zero — o grafo caminho também é possível. Entretanto, para o grafo caminho, $\bar{L}$ é proporcional a N, e portanto, para $\alpha \to +0$ o único mínimo existente quando $\alpha > 0$ tem grafo estrela como limite.

Podemos relaxar o vínculo de estrita minimização e propor o seguinte modelo: o agente decide sua estratégia através de uma dinamica interna que ocorre em seu cérebro, ditada por regras que levam em conta o custo total. Podemos então associar probabilidades às estratégias de acordo com nossa expectativa de que grafos devem surgir dessa dinâmica. Se não conhecemos detalhes da dinâmica, mas temos informação de que o valor esperado do custo total é uma variável importante, o procedimento bayesiano adequado é associar ao grafo uma distribuição de máxima entropia, restringida pelo valor do custo total. A distribuição resultante é a distribuição de Gibbs:

$$p(G) = \frac{q(G)}{Z} e^{-\beta H(G)} \tag{3.6}$$

onde β, uma espécie de inverso de temperatura, regula a importância relativa dos efeitos que o custo $H(G)$ e outros efeitos negligenciados pelo modelo, e $q(G)$ é a atribuição de probabilidades a priori. Assumiremos probabilidades uniformes a priori, com a restrição de que o grafo seja conexo (veja nota [11]).A atribuição de probabilidades da equação (3.6) torna o problema de determinar as propriedades macroscópicas desse sistema em um problema de mecânica estatística, cuja variável dinâmica é a matriz de adjacências do grafo G:

$$M_{ij} = \begin{cases} 1 & \text{se } i \text{ e } j \text{ estão ligados por uma aresta} \\ 0 & \text{outro caso} \end{cases} \tag{3.7}$$

⌈3.3.6⌉ Interpretação dos parâmetros α e β

O parâmetro α (ou sua versão normalizada, a) regula a importância relativa entre os dois custos — social e cognitivo. O valor de α é

regulado pela capacidade cognitiva dos agentes: quanto maior a capacidade de realizar cálculos sociais, menor é a importância do custo cognitivo, maior é o valor de α. No presente trabalho α será, portanto, interpretado como uma medida da capacidade cognitiva dos agentes.

O parâmetro β regula a escala em que flutuações no valor do custo total são toleradas. Para β grande, variações no custo total acima do custo mínimo são muito pouco prováveis. Para valores maiores de β as mesmas flutuações apresentam probabilidades maiores. Dessa forma, β regula o quão prováveis são configurações que consistem de flutuações em torno das configurações ótimas discutidas na seção *Minimização do custo*. Flutuações em torno do custo mínimo representam um dispêndio extra de energia e recursos que poderiam ser gastos em outras atividades, portanto β pode ser parcialmente entendido como uma variável ecológica — escassez de recursos ou ambientes sociais mais rigorosos implicam em menor tolerarância a flutuações do custo total. Assim, β é uma variável que regula a intensidade da pressão para otimização do custo total, que pode ter origem ambiental ou social. Essa interpretação encontra ressonância em Earle [53][12], que identifica fatores como pressão populacional, falta de recursos naturais, seca e guerras como fatores de mudança de organização social.

[12] T.K. Earle. *How Chiefs Come to Power: The Political Economy in Prehistory*. Antropology. Political science. Stanford University Press, 1997. ISBN 9780804728560. URL http://books.google.com.br/books?id=4AjAk15WSbQC

⌈3.3.7⌉ Resultados numéricos para agentes isolados

Simulações de Monte Carlo desse modelo foram feitas usando o algoritmo de Metropolis[13] Partindo de um grafo inicial aleatório (sorteado do ensemble de Erdos-Rényi, com fração de arestas $\frac{1}{2}$), a cada passo do algoritmo é proposta uma mudança em uma aresta do grafo —- adicionando uma aresta faltante ou removendo uma aresta existente. Caso a mudança não quebre a conectividade do grafo, ela será aceita com probabilidade dada pelo fator de Gibbs:

[13] D.P. Landau and K. Binder. *A Guide To Monte Carlo Simulations In Statistical Physics*. Cambridge University Press, 2005. ISBN 9780521842389. URL http://books.google.com.br/books?id=11tun6Y5tOAC

$$e^{-\beta(H(G')-H(G))}. \tag{3.8}$$

Na figura 3.5 temos um corte do diagrama de fase desse modelo com temperatura e número de agentes constante, variando o parâmetro α. Seja d_i o grau do i-ésimo nó do grafo de um agente escolhido ao acaso. No painel superior destacamos dois parâmetros de ordem: $d_{\max} = \frac{1}{N-1} \max(d_1, d_2, \ldots, d_N)$, em azul, é proporcional ao grau do nó mais bem conectado do grafo, e $d_{\text{avg}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} d_i$, em vermelho, é proporcional ao grau médio de todos os nós do grafo. Uma vez que o maior possível grau para um nó é $N-1$, esses parâmetros foram normalizados para que o valor máximo seja 1. As linhas sólidas

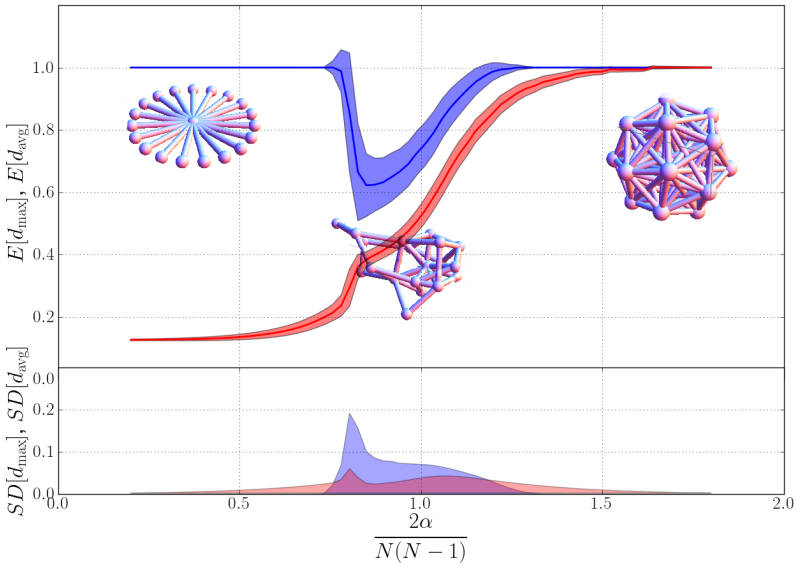


Figura 3.5 – Corte do diagrama de fases apresentando o valor esperado obtido via Monte Carlo dos parâmetros de ordem normalizados $d_{\max}$ e d_{avg} em função de α, com temperatura e número de agentes fixo, bem como seus desvios padrão. Sobreposto ao gráfico se observam exemplos de arquiteturas do grafo sorteadas da distribuição de equilibrio na região correspondente do diagrama. Os parâmetros de controle para essa simulação são $n = 16$ e $\beta = 7.0$.

correspondem aos valores esperados:

$$E\left[d_{\max}\right] = \sum_{G} P(G|\alpha, n, \beta) d_{\max}(G) \tag{3.9}$$

$$E\left[d_{\text{avg}}\right] = \sum_{G} P(G|\alpha, n, \beta) d_{\text{avg}}(G) \tag{3.10}$$

obtida através da simulação de Monte Carlo do modelo, enquanto o sombreado ao redor da linha representa o desvio padrão obtido da mesma forma. No painel inferior apresentamos novamente o desvio padrão para melhor visualização. Sobrepostas aos gráficos estão figuras representativas de grafos sorteados da distribuição de equilíbrio em pontos correspondentes do diagrama de fases.

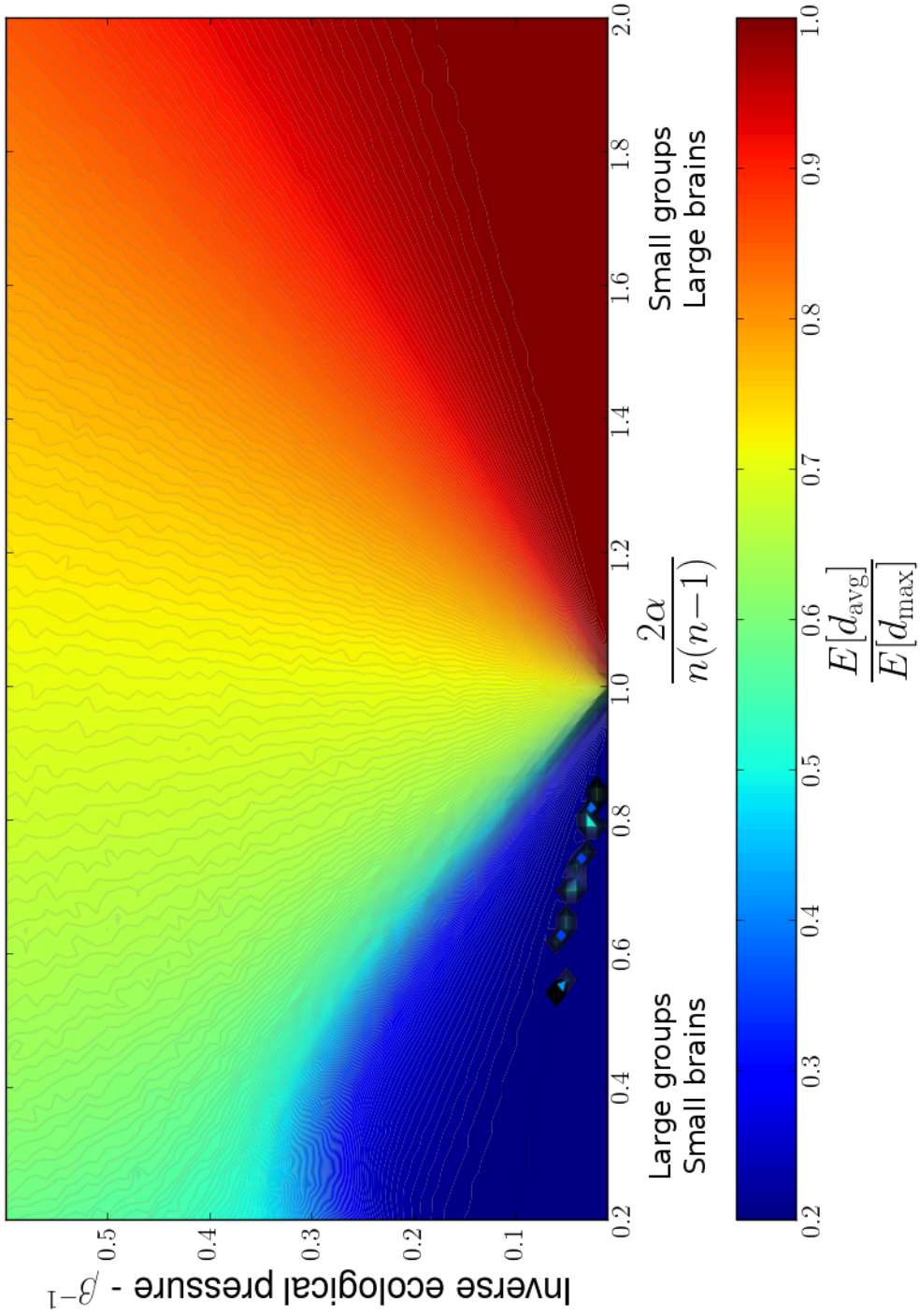


Figura 3.6 – Diagrama de fases apresentando a razão $\frac{d_{\mathrm{avg}}}{d_{\mathrm{max}}}$ em função de α e temperatura, para um número fixo de agentes $n = 16$.

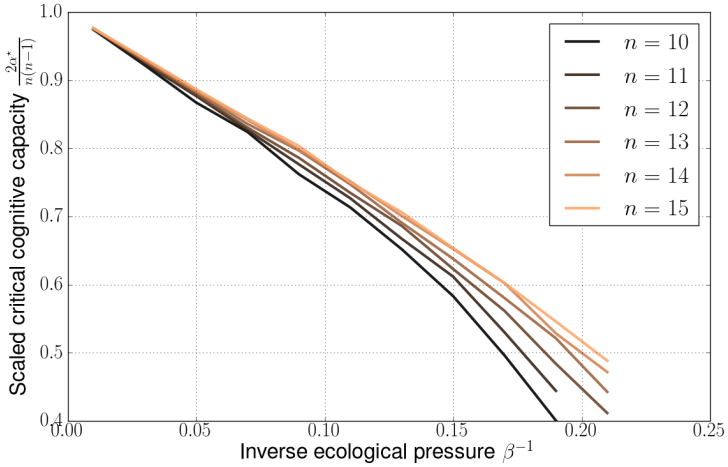


Figura 3.7 – Valor crítico do parâmetro a em função da temperatura para diferentes tamanhos do sistema.

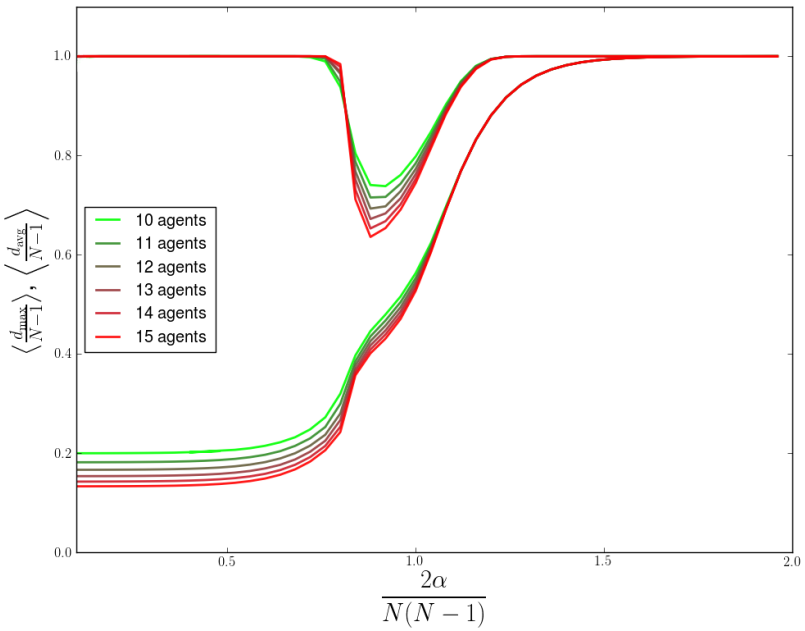


Figura 3.8 – Corte do diagrama de fase para vários tamanhos do grupo onde se pode ver que a variação dos parâmetros de ordem em torno do valor crítico de α cresce com o número de agentes. Simulações obtidas com $\beta = 7$.

Na figura3.6 temos um diagrama de fases completo variando α e a temperatura para um número fixo de agentes. A variável descrita no mapa de cores é a razão $\frac{d_{\text{avg}}}{d_{\text{max}}}$. Esse diagrama mostra uma linha de transição de fases entre a região azul escura — a região em que a organização do grafo é fortemente centralizada, com nós periféricos pouco conectados, e uma região em que a razão $\frac{d_{\text{avg}}}{d_{\text{max}}}$ é menos extrema. Acima de uma temperatura crítica essa fase não é mais observada. A região vermelho escura corresponde à fase totalmente conectada, ou situações bem próximas disso. Nessa região não há grandes saltos nos parâmetros de ordem, que mudam continuamente com a temperatura e a.

A linha de transição de fase pode ser observada para diferentes valores da temperatura na figura 3.7. Note que para $\beta^{-1} \to 0$, temos $\alpha^\star = 1$. Na figura 3.8 pode-se ver como o diagrama de fases varia com o número de agentes. Quanto maior o número de agentes, mais extremas as variações dos parametros de ordem no entorno da possível transição de fase indicando uma possível descontinuidade quando $n \to \infty$.

[3.3.8] Interpretação parcial dos resultados

O diagrama de fases apresenta três regimes. Para valores altos de $a = \frac{2\alpha}{N(N-1)}$, ou seja, alta capacidade cognitiva ou grupos com poucos agentes, todos os nós apresentam praticamente a mesma conectividade. O grafo é simétrico, com conectividade densa e bem próximo de totalmente conexo. As arestas são aleatóriamente distribuidas, de forma simétrica. Para valores intermediários de a, existe um nó com conectividade ligeiramente maior, mas existem flutuações grandes. A taxa de aceitação do algoritmo de monte carlo é alta (ver: figura 3.9). Nesse regime, o grafo é momentaneamente não-simétrico, mas estatísticamente qualquer nó pode ocupar a posição de conectividade maior e alterações desse nó central são frequentes. Há um pico nos desvios padrão dos parâmetros de ordem para um certo valor do parâmetro de controle $a^\star$, indicando um possível ponto crítico. Para valores mais baixos de a, a simetria é espontaneamente quebrada e apenas um nó ocupa uma posição central. Esse nó está conectado a todos os outros, que estão quase que exclusivamente conectados a ele. O grafo se torna uma estrela e não há flutuação observável na conectividade do nó central. Neste regime, a representação mental da rede de relacionamenos sociais construídas pelo agente é assimétrica e existe um único nó que serve como proxy para todas as relações sociais do grupo. Na representação mental do agente em questão, o status social de cada um dos

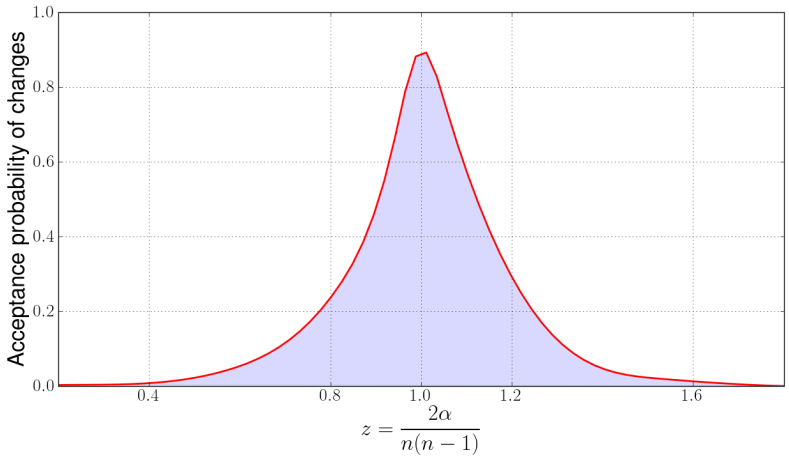


Figura 3.9 – Taxa de aceitação do algoritmo de Monte Carlo — fração das propostas de mudanças no microestado do sistema que foram aceitas com probabilidade dada pelo fator de Gibbs, sampleada com $\beta = 7.0$

outros nós é definido por como ele se relaciona a esse nó central.

[3.3.9] Dinâmica para muitos agentes e resultados numéricos

As figuras acima tratam de propriedades independentes da interação entre os agentes. Essa interação, como dito anteriormente, será introduzidas na forma de aprendizado social ("fofoca" ou *gossip*). Durante a simulação de Monte Carlo, duas possíveis fontes serão consideradas para a proposta de uma nova aresta no passo de Metropolis:

- Com probabilidade $1 - g$, um novo valor para a aresta (i, j) do agente k será sorteado ao acaso,
- Com probabilidade g, um novo valor para a aresta (i, j) do agente k será copiado da aresta (i, j) de um outro agente l sorteado ao acaso.

Essa proposta de novo valor de aresta será aceita com probabilidade proporcional ao fator de Gibbs (3.8). Esse procedimento visa imitar o aprendizado social observado em humanos [14]. Essa escolha de interação não altera os diagramas de fase já mostrados no capítulo anterior, mas introduz correlação entre os grafos de diferentes agentes.

Na figura 3.10 são exibidas duas grandezas que quantificam a correlação entre os grafos na fase estrela. Vamos denotar por c_i o label que identifica o nó central do grafo do i-ésimo agente. Para um certo número N de agentes temos então o conjunto $\{c_1, c_2, \ldots, c_N\}$. Seja a

[14] R. I. M. Dunbar. *How Many Friends Does One Person Need? Dunbar's Number and Other Evolutionary Quirks*. Harvard University Press, 2010. ISBN 9780674057166. URL `http://books.google.com.br/books?id=W-kkeffNx5MC`

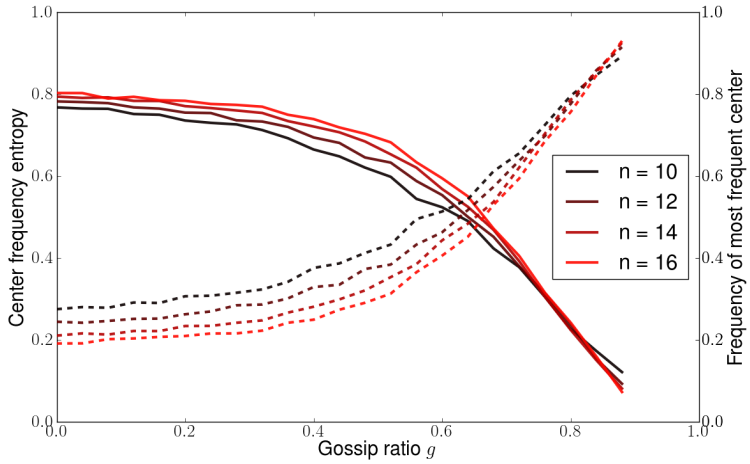


Figura 3.10 – Parâmetros de ordem associados à correlação entre grafos de diferentes agentes, calculados na fase em que os grafos apresentam estrutura de estrela. As curvas tracejadas correspondem à frequência do nó central mais frequente. As linhas tracejadas correspondem à entropia da distribuição de centros.

variável aleatória C definida como um valor sorteado ao acaso desse conjunto e seja:

$$p(c) = \text{Prob}\,\{C = c\}$$

a sua distribuição de probabilidades. O primeiro parâmetro de ordem, correspondente às linhas tracejadas, é dada por $E\,[\max_c p(c)]$, ou seja, o fração do número de agentes que possuem como nó central o nó que mais vezes aparece como nó central. Isso corresponde de forma grosseira a que fração dos agentes tem grafos estrela com o mesmo nó ocupando o centro da estrela. A segunda variável é, a menos de uma constante multiplicativa, simplesmente a entropia da distribuição de c: $S(c) = -\sum_c p(c) \log p(c)$. Ambas as grandezas são calculadas para grafos em forma de estrela, em função da probabilidade de encontro entre dois agentes dada por g, para valores fixos de temperatura, variando-se o número de agentes. O resultado mostra que, para baixos valores de g, a probabilidade de que um certo nó seja o centro de um agente tomado ao acaso é aproximadamente uniforme, e nenhum dos nós domina como centro de uma fração substancial de grafos. Para valores maiores de g, os grafos estrela tendem a se correlacionar e o mesmo nó pode ser central em uma grande fração de agentes. Dessa forma, é possível que no regime em que o grafo é uma estrela, o mesmo agente sirva como proxy para as relações sociais de todo o grupo para uma substancial maioria dos agentes.

[3.4] Sumarização e interpretação dos Resultados

Construímos um modelo para a organização social de uma sociedade de agentes que tentam representar mentalmente sua estrutura social. A representação mental tem um custo cognitivo, derivado da limitação cognitiva do agente, e um custo social, derivado da necessidade de se navegar corretamente as relações sociais. Esses custos levam a um modelo mecanico-estatístico que possui um diagrama de fases com alguns regimes interessantes, controlados pelos parâmetros a, que regula a capacidade cognitiva do agente e/ou o tamanho do grupo, g que controla a intensidade da interação social e β, que controla a tolerância a flutuações no custo total do agente. As fases observadas são:

[3.4.1] Grupos pequenos e/ou alta capacidade cognitiva

Para a grande, ou seja, grupos de tamanho pequeno ou agentes com grande capacidade cognitiva, os agentes possuem modelos mentais do panorama social do seu grupo em que nenhum agente em particular ocupa uma posição central. Em outras palavras, a representação mental das redes sociais nesse grupo são todas simétricas e nenhum agente se destaca. Nessa fase, em que nenhum agente se destaca como referência social para os outros, pode-se invocar a Teoria da Reversão de Dominância[15], que propõe um mecanismo de formação de grupos igualitários em que eventuais líderes de um grupo são dominados pelos membros do grupo, através de uma atitude de desaprovação de comportamento hierarquico e tentativas de dominação. À luz dessa teoria, pode-se concluir que, em situações em que a representação mental dos agentes é simétrica, tentativas de imposição sua autoridade ao grupo vão encontrar o tipo de resistência discutido por Boehm et al., que não será superada na ausência de um membro do grupo que possua suficiente capital social para tanto. Tal raciocínio implica na obtenção de uma estrutura igualitária nessa fase do modelo.

[15] C Boehm. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Harvard University Press, 2001. ISBN 9780674006911. URL http://books.google.com.br/books?id=ljxS8gUlgqgC; and Christopher Boehm, Harold B. Barclay, Robert Knox Dentan, Marie-Claude Dupre, Jonathan D. Hill, Susan Kent, Bruce M. Knauft, Keith F. Otterbein, and Steve Rayner. Egalitarian behavior and reverse dominance hierarchy [and comments and reply]. *Current Anthropology*, 34(3): pp. 227–254, 1993. ISSN 00113204. URL http://www.jstor.org/stable/2743665

[3.4.2] Grupos de tamanho intermediário e/ou capacidade cognitiva intermediária

Para valores intermediários de a, os agentes possuem representações fluidas do panorama social de seu grupo, com flutuações grandes. Há uma certa concentração na conectividade, que pode rapidamente flutuar entre um ou outro nó temporariamente central. Estatísticamente, o modelo ainda é simétrico, e ainda faz sentido invocar

o argumento acima para inferir que a estrutura deve ser aproximadamente igualitária, com eventual dominância temporária de agentes que ocupariam posições de *primus inter pares*.

⌈3.4.3⌉ Grupos grandes e/ou capacidade cognitiva menor

Quando a é pequeno, ou seja, para grupos grandes e/ou agentes com menor capacidade cognitiva, existe uma quebra de simetria: as representações mentais da rede social dos agentes é centralizada e congelada em um grafo em forma de estrela. Cada um dos agentes possui uma representação mental assimétrica da rede social. O nó central do grafo de um agente se torna a única referência para todos os cálculos sociais a serem realizado por ele e todas as suas decisões em jogos sociais são tomadas levando em conta a natureza da relação dos outros nós com esse nó central.

Quando g é pequeno (baixo aprendizado via *"gossip"*), entretanto, os nós centrais de cada agente são aleatórios e, de certa forma, apesar de haver uma quebra de simetria na representação mental que cada agente faz do grupo, a situação global ainda é simétrica. Para valores maiores de g, a maioria dos agentes possui o mesmo modelo mental: um grafo em forma de estrela, centrado em torno do mesmo agente específico. O grupo usa as conexões desse mesmo agente central como informação mais relevante na tomada de decisões em jogos sociais. É natural supor que o capital social que cada agente é capaz de recrutar em seu favor na formação de uma coalizão, por exemplo, deve ser proporcional ao número de vizinhos que possui em sua própria representação da rede social. Esses vizinhos são os indivíduos com relação aos quais o agente em questão conhece sua posição social. Se uma fração significante dessas relações resultam em coalizões, temos uma distribuição de capital social proporcional ao número de primeiros vizinhos.

Esse agente central está posicionado de forma privilegiada na solução de dilemas sociais, formação de coalizões e outras atividades sociais do grupo. Fazendo a hipótese de que o capital social derivado dessa posição quebra a simetria do resultado de jogos sociais de maneira vantajosa ao agente central[16], pode-se esperar que esse agente atinja um certo grau de proeminência ou dominância. A simetria assumida na Teoria da Reversão de Dominância é quebrada, pois existe um agente com vantagens sociais claras.

[16] A plausibilidade dessa conclusão pode ser analizada em um jogo social simples em que cada agente tenta recrutar outros agentes para uma coalizão. As regras são: (1) cada agente comece o recrutamento a partir dos outros indivíduos cuja relação com ele próprio é conhecida (primeiros vizinhos do grafo interno daquele agente), e (2) que o agente é capaz de recrutar uma fração de seus vizinhos. Nessa situação é evidente que um agente que ocupe uma posição central em quase todas as redes, será capaz de recrutar para sua coalizão uma fração macroscópica de todo o grupo.

⌈4⌉ Conclusão e Observações Finais

⌈4.1⌉ Tópicos tratados na Tese

Foram obtidos nesse trabalho resultados matemáticos concernentes a dois tópicos distintos — ambas beneficiadas por um olhar oriundo de uma formulação bayesiana da mecânica estatística e da teoria de informação. Os tópicos foram desenvolvidos usando conceitos comuns a esse paradigma. Abaixo apresentamos uma recapitulação dos resultados obtidos e uma discussão final.

⌈4.1.1⌉ Dependência Estatística, Teoria de Cópulas e Teoria de Informação

Nesse tópico foi discutida uma relação entre três campos de pesquisa relacionados à Estatística e à Teoria de Probabilidades: dependência estatística, teoria de cópulas e teoria de informação. Através dessa visão unificada foi possível reescrever axiomas devidos à Renyi [1] para a definição de uma boa medida de dependência da seguinte forma:

- Uma boa medida de dependência entre duas variáveis X e Y é um funcional $\mathcal{F} : C_2 \to \mathcal{R}$ que leva funções cópula $C_{XY}(\cdot, \cdot)$ em números reais e independe das distribuições marginais;
- atinge um valor mínimo, que será arbitráriamente escolhido como zero, se, e somente se, $C_{XY}(u, v) = uv$;
- atinge um valor máximo quando $C_{XY}(u, v) = W(u, v)$ ou $C_{XY}(u, v) = M(u, v)$.
- para $C_{XY}(u, v) = N_\rho(u, v)$, o funcional é um função monotônica crescente do parâmetro ρ.

Essa nova definição faz uso explicito do conceito de cópula, tornando mais simples demonstrar que uma certa estatística satisfaz todos os requisitos e conectando duas áreas relacionadas de investigação que até então não possuiam conexão até onde vai o conhecimento do autor.

Em seguida, o status da informação mútua como medida de dependência é explorado, demonstrando sua relação com a entropia de cópula na equação (2.31):

$$I(X,Y) = \int\int \mathrm{d}u\mathrm{d}v\; c(u,v)\log c(u,v) = -S[c] \geq 0. \tag{4.1}$$

Essa relação permite associar o principio de máxima entropia, comumente aplicado na obtenção de distribuições a priori não-informativas e em outros cálculos típicos de inferência bayesiana, com um princípio de mínima dependência (mínima informação mútua). Além disso, a relação entre a informação mútua e o parâmetro de correlação de cópulas gaussianas suscita uma discussão a respeito da validade e dos riscos do uso da correlação linear como medida de dependência, e uma demonstração gráfica de que a correlação linear sistematicamente subestima a dependência entre variáveis com acoplamento gaussiano mas distribuições marginais não-gaussianas.

O caso particular das cópulas elípticas é estudado e é obtida uma decomposição da dependência associada a essas cópulas em duas partes, associadas às partes linear e não-linear da dependência:

$$I(\Sigma, \psi(\cdot)) = I_0(\Sigma) + I[p(\cdot)]. \tag{4.2}$$

onde $I_0(\Sigma) = -\frac{1}{2}\log\det(\Sigma)$ é a parte linear da dependência e $I[p(\cdot)]$ é a informação mútua da cópula esférica padronizada correspondente à cópula elíptica em questão. Essa decomposição permite escrever um teste estatístico de desvio de gaussianidade que é então aplicado a um conjunto de dados de séries temporais de ações de alta liquidez negociadas em bolsas de valores americanas. Essa análise leva à conclusão de que são abundantes casos de forte dependência estatística entre pares de ações em que a correlação linear é praticamente nula. Por fim, um método para ajuste de cópulas elípticas foi desenvolvido e aplicado ao caso especial das cópulas t, de ampla aplicação em finanças.

⌈4.1.2⌉ Um modelo Mecânico-Estatístico para a emergência de autoridade

Neste tópico foi desenvolvido um modelo para a emergência de autoridade em sociedades humanas pré-agrícolas. O problema proposto é a questão da variabilidade dos tipos de organização social dos

humanos, que apresentam uma distribuição muito mais ampla no espectro etológico social do que a maioria das outras espécies de primatas. Foram revisados certos fatos empíricos e teóricos associados ao tema — a Hipótese do Cérebro Social, a Teoria da Reversão de Dominância e a observação da evolução temporal da organização social humana, denominada *"U-shaped evolution"*. Essas observações fornecem peças para o quebra-cabeça e subsidiam a criação de um modelo matemático atacando essa questão.

Através das pistas oferecidas por essa revisão, um modelo mecânico-estatístico é definido. Trata-se de um modelo de agentes em que cada um dos agentes de um grupo possui uma representação interna das relações sociais do grupo na forma de um grafo. As arestas do grafo representam relações sociais do grupo que aquele particular agente ativamente despende recursos para obter. São descritos os custos são envolvidos na manutenção de um certo conjunto de arestas: o custo associado às limitações cognitivas do agente, e os custos sociais associados a erros de inferência que podem ser cometidos pelo agente por manter apenas uma representação limitada do seu ambiente social. O método da máxima entropia, descrito na introdução, é usado para determinar uma distribuição de probabilidades para os grafos dos agentes, dando origem a um modelo mecânico-estatístico, com uma típica distribuição de Gibbs. O modelo é posteriormente extendido para incluir interações entre os agentes, através de um mecanismo de aprendizado social ou "fofoca".

O modelo é estudado através da técnica de Monte Carlo, sendo obtidos os parâmetros de ordem relevantes:

$$E\left[d_{\max}\right] = \sum_{G} P(G|\alpha, n, \beta, g) d_{\max}(G), \tag{4.3}$$

$$E\left[d_{\text{avg}}\right] = \sum_{G} P(G|\alpha, n, \beta, g) d_{\text{avg}}(G), \tag{4.4}$$

que são os valores esperados do grau médio e do grau máximo dos nós dos grafos de cada um dos agentes. Além disso é medido o grau de correlação entre os nós mais centrais dos grafos de cada um dos agentes. Os parâmetros de controle desse modelo são α, associado à capacidade cognitiva dos agentes, n, o número de agentes, β, uma pressão ambiental-social que controla a tolerância a flutuações dos custos descritos acima, e g, que controla a intesidade do aprendizado social.

O modelo apresenta três fases de interesse:

- A primeira ocorre quando α é grande ou n é pequeno. Nesse caso, os grafos são basicamente grafos aleatórios, com elevado grau de

conectividade. Os agentes são capazes de manter a maioria da informação social e nenhum dos agentes ocupa posições privilegiadas no sistema.

- A segunda fase interessante ocorre para valores intermediários de α e n e valores maiores de g, em que os grafos dos diferentes agentes se tornam fortemente correlacionados. Nessa situação, o sistema está em uma fase fluida, em que flutuações são altas, mas há uma certa ocorrência de nós temporariamente centrais, que ocupam uma posição de *hub* social por um certo tempo. Quanto maior a pressão ecológica β, menos pronunciada é essa fase, que ocorre para regiões cada vez menores do espaço de parâmetros conforme β aumenta.

- A terceira fase ocorre quando α é pequeno ou n é grande, também para maiores valores de g. Nesse caso, os grafos de todos os agentes se tornam grafos tipo estrela, com um nó central bem definido. Nessa fase as flutuações se tornam muito menores, e os grafos se congelam com um nó central específico. Se há forte correlação entre os grafos (g grande), uma grande parte dos grafos vão se organizar em torno do mesmo nó central, resultando em um status social diferenciado para um dos agentes.

O MODELO INDICA QUE APENAS LIMITAÇÕES COGNITIVAS são suficientes para produzir uma quebra de simetria na representação interna dos agentes das relações sociais do grupo, ainda que outras variáveis sociais sejam simétricas. Há evidência empírica[1] (Earle [53], Wiessner [56]) de correlação entre posição social percebida e o exercício factual de autoridade, indicando um quadro em que essa quebra de simetria na percepção dos agentes leva a uma quebra de simetria de fato na rede social. Além disso, a teoria da reversão de dominância sugere que, em uma representação simétrica, os grupos deveriam se organizar de forma igualitária devido à resistência oferecida pelo grupo a dominação por líderes emergentes. Na situação de quebra de simetria, o mecanismo de reversão de dominância deve perder eficiência diante do capital social acumulado do indivíduo que ocupa uma posição central na rede social inferida de uma maioria dos agentes.

[1] T.K. Earle. *How Chiefs Come to Power: The Political Economy in Prehistory*. Antropology. Political science. Stanford University Press, 1997. ISBN 9780804728560. URL `http://books.google.com.br/books?id=4AjAk15WSbQC`; and P. Wiessner. The vines of complexity: Egalitarian structures and the institutionalization of inequality among the Enga. *Current anthropology*, 43(2):233–269, 2002

ESSE QUADRO É COMPATÍVEL com o observado empiricamente na tese da *"u-shaped evolution"*. O aparecimento do gênero *Homo*, com elevada capacidade cognitiva, resultado da pressão seletiva associada à necessidade de cérebros cada vez maiores para lidar com ambientes sociais cada vez mais complexos, leva a uma transição de grupos hierárquicos (grandes primatas pré-humanos) para grupos igualitários. O posterior surgimento da agricultura eleva a concentração de indivíduos a níveis inéditos, sem tempo para ajuste evolutivo da capacidade cognitiva, o

que causa uma transição para uma sociedade hierárquica.

A utilidade analítica desse modelo deverá ser avaliada comparando suas previsões a dados empiricos. O modelo prevê uma relação entre o número de agentes e a organização social, o que é corroborado por evidências encontradas em Currie et al. [38][2] e Kennett and Winterhalder [39][3](em Clark et al. [57]). Além disso, o modelo prevê que para valores maiores do parâmetro β, o tamanho de grupo necessário para a transição para a fase "despótica" é menor. Em outras palavras, a fase mais fluida, em que há flutuações na organização social, é menos resiliente quanto maior o valor de β. O parâmetro β é interpretado como um mixto de pressão ecológica (interpretação corroborada por, por exemplo, Earle [53][4]) e pressão social de pares(ver: Caticha and Vicente [17]). A relação entre pressões ambientais e surgimento de estruturas despóticas encontra ressonância em dados empíricos, por exemplo em Kennett and Winterhalder [39], e a relação entre ambientes mais difíceis e surgimento de hierarquia é observada em diversos estudos revisados por Summers [58][5]. Análises adicionais, no entanto, são necessárias para determinar com maior grau de confiança a relação entre as previsões do modelo e dados empíricos, que podem ser obtidos em amplas pesquisas etnográficas.

[2] Thomas E. Currie, Simon J. Greenhill, Russell D. Gray, Toshikazu Hasegawa, and Ruth Mace. Rise and fall of political complexity in island south-east asia and the pacific. *Nature*, 467(7317):801–804, Oct. 2010. DOI: 10.1038/nature09461

[3] Douglas J. Kennett and Bruce Winterhalder. *Islands of Inquiry: Colonisation, seafaring and the archaeology of maritime landscapes*, chapter Demographic expansion, despotism and the colonisationof East and South Polynesia, pages 87–96. In Clark et al. [57], 2008. ISBN 9781921313905. URL `http://books.google.com.br/books?id=NZ0-IiAZ_ZwC`

[4] T.K. Earle. *How Chiefs Come to Power: The Political Economy in Prehistory*. Antropology. Political science. Stanford University Press, 1997. ISBN 9780804728560. URL `http://books.google.com.br/books?id=4AjAk15WSbQC`

[5] K. Summers. The evolutionary ecology of despotism. *Evolution and Human Behavior*, 26(1):106–135, 2005

[5] Bibliografia

[1] A. Renyi. On measures of dependence. *Acta. Math. Acad. Sci. Hungar.*, 10:441–451, 1959.

[2] R. S. Calsaverini and R. Vicente. An information-theoretic approach to statistical dependence: Copula information. *EPL (Europhysics Letters)*, 88(6):68003, 2009. URL http://stacks.iop.org/0295-5075/88/i=6/a=68003.

[3] A. de Moivre. *The Doctrine of Chances: Or, A Method of Calculating the Probability of Events in Play*. W. Pearson, 1718. URL http://books.google.com.br/books?id=3EPac6QpbuMC.

[4] P.S. Laplace. *Théorie analytique des probabilités*. Ve. Courcier, 1814. URL http://books.google.com.br/books?id=6MRLAAAAMAAJ.

[5] P.S. Laplace. *Essai philosophique sur les probabilités*. Bachelier, 1840. URL http://books.google.com.br/books?id=SAsyUDOK0mIC.

[6] R. T. Cox. Probability, frequency, and reasonable expectation. *Am. Jour. Phys.*, 14:1–13, 1946.

[7] R.T. Cox. *Algebra of Probable Inference*. Johns Hopkins University Press, 2001. ISBN 9780801869822. URL http://books.google.com.br/books?id=dcNpAUU6ACgC.

[8] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 47:379–423 and 623–656, 1948.

[9] E.T. Jaynes and G.L. Bretthorst. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. ISBN 9780521592710. URL http://books.google.com.br/books?id=tTN4HuUNXjgC.

[10] Ariel Caticha. Lectures on probability, entropy, and statistical physics. *arXiv:0808.0012*, 2008. URL http://arxiv.org/abs/0808.0012. Invited lecture at the 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (July 8-13, 2008, Boraceia Beach, Sao Paulo, Brazil).

[11] Ariel Caticha. Quantifying rational belief. *arXiv:0908.3212*, 2009. URL http://arxiv.org/abs/0908.3212. Presented at the 29th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (July 5-10, 2009, Oxford, Mississippi, USA.) In.

[12] Philip Goyal, Kevin H. Knuth, and John Skilling. Origin of complex quantum amplitudes and Feynman's rules. *Phys. Rev. A*, 81:022109, Feb 2010. DOI: 10.1103/PhysRevA.81.022109. URL http://link.aps.org/doi/10.1103/PhysRevA.81.022109. arxiv:0907.0909.

[13] M. Tribus. *Rational descriptions, decisions, and designs*. Pergamon unified engineering series: engineering design section. Pergamon Press, 1969. URL http://books.google.com.br/books?id=C4pRAAAAMAAJ.

[14] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, 2006. ISBN 9780471748816. URL http://books.google.com.br/books?id=EuhBluW31hsC.

[15] S. Amari, H. Nagaoka, and D. Harada. *Jōhō Kika No Hōhō (Methods of Information Geometry)*. Translations of mathematical monographs v. 191. Amer Mathematical Society, 2000. ISBN 9780821843024. URL http://books.google.com.br/books?id=vc2FWSo7wLUC.

[16] L.D. Landau, E.M. Lifshitz, and L.P. Pitaevskii. *Statistical Physics Pt. 1*, volume 6 of *Course of Theoretical Physics*. Elsevier Science & Technology Books, 1980. ISBN 9780750626361. URL http://books.google.com.br/books?id=NaB7oAkon9MC.

[17] Nestor Caticha and Renato Vicente. Agent-based social psychology: from neurocognitive processes to social data. *Advances in Complex Systems*, 14(05):711–731, 2011. DOI: 10.1142/S0219525911003190. URL http://www.worldscientific.com/doi/abs/10.1142/S0219525911003190.

[18] S.R. Salinas. *Introdução a Física Estatística*. EDUSP, 1997. ISBN 9788531403866. URL http://books.google.com.br/books?id=cAlP3sc0K44C.

[19] M. Opper and D. Saad. *Advanced Mean Field Methods: Theory and Practice*. Neural Information Processing Series. Mit Press, 2001. ISBN 9780262150545. URL http://books.google.com.br/books?id=cuOX8sCDeNAC.

[20] M. Mézard and A. Montanari. *Information, Physics, and Computation*. Oxford Graduate Texts. OUP Oxford, 2009. ISBN 9780198570837. URL http://books.google.com.br/books?id=jhCM7i0a6UUC.

[21] B. Schweizer and E. F. Wolff. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4):879 – 885, 1981.

[22] R.B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer, 2006. ISBN 9780387286594. URL http://books.google.com.br/books?id=B3ONT5rBv0wC.

[23] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.

[24] Jian Ma and Zengqi Sun. Mutual information is copula entropy. *Tsinghua Science and Technology*, 16(1):51–54, 2011. ISSN 1007-0214. DOI: 10.1016/S1007-0214(11)70008-6. arXiv:0808.0845.

[25] Rick L. Jenison and Richard A. Reale. The shape of neural dependence. *Neural Computation*, 16:665 – 672, 2004.

[26] S. Jones. The formula that felled Wall St. *The Financial Times*, Abril 2009. URL http://www.ft.com/cms/s/2/912d85e8-2d75-11de-9eba-00144feabdc0.html.

[27] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E*, 69:066138, 2004.

[28] Stefano Demarta and Alexander J. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1):111–129, 2005.

[29] Lisa Borland. Option pricing formulas based on a non-gaussian stock price model. *Phys. Rev. Lett.*, 89(9):098701, Aug 2002. DOI: 10.1103/PhysRevLett.89.098701.

[30] F. B. M. de Waal. *Chimpanzee Politics: Power and Sex among Apes*. Chimpanzee Politics. Johns Hopkins University Press, 2007. ISBN 9780801886560. URL http://books.google.com.br/books?id=XsrhU2vV5PIC.

[31] Jane Goodall. *What separates us from chimpanzees*. TED Talk, Março de 2002. URL http://www.ted.com/talks/jane_goodall_on_what_separates_us_from_the_apes.html.

[32] F. B. M. de Waal. *Peacemaking Among Primates*. Harvard University Press, 1990. ISBN 9780674659216. URL http://books.google.com.br/books?id=5at0TN8U6RMC.

[33] Sandra L. Vehrencamp. A model for the evolution of despotic versus egalitarian societies. *Animal Behaviour*, 31(3):667 – 682, 1983. ISSN 0003-3472. DOI: 10.1016/S0003-3472(83)80222-X. URL http://www.sciencedirect.com/science/article/pii/S000334728380222X.

[34] C Boehm. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Harvard University Press, 2001. ISBN 9780674006911. URL http://books.google.com.br/books?id=ljxS8gUlgqgC.

[35] Bruce M. Knauft, Thomas S. Abler, Laura Betzig, Christopher Boehm, Robert Knox Dentan, Thomas M. Kiefer, Keith F. Otterbein, Tohn Paddock, and Lars Rodseth. Violence and sociality in human evolution [and comments and replies]. *Current Anthropology*, 32:391–428, 1991. URL http://www.jstor.org/stable/2743815.

[36] Joyce Marcus. The archeological evidence for social evolution. *Annu. Rev. Anthropol*, 37:25166, 2008.

[37] R. Foley. *Another unique species: patterns in human evolutionary ecology*. Longman Scientific & Technical, 1987. ISBN 9780470207284. URL http://books.google.com.br/books?id=AjuAAAAAMAAJ.

[38] Thomas E. Currie, Simon J. Greenhill, Russell D. Gray, Toshikazu Hasegawa, and Ruth Mace. Rise and fall of political complexity in island south-east asia and the pacific. *Nature*, 467(7317):801–804, Oct. 2010. DOI: 10.1038/nature09461.

[39] Douglas J. Kennett and Bruce Winterhalder. *Islands of Inquiry: Colonisation, seafaring and the archaeology of maritime landscapes*, chapter Demographic expansion, despotism and the colonisationof East and South Polynesia, pages 87–96. In Clark et al. [57], 2008. ISBN 9781921313905. URL http://books.google.com.br/books?id=NZ0-IiAZ_ZwC.

[40] T Sawaguchi and H Kudo. Neocortical development and social structure in primates. *Primates*, pages 283–90, 1990.

[41] Robin Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 20:469–93, 1992.

[42] R. I. M. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16:681–694, 11 1993. ISSN 1469-1825. DOI: 10.1017/S0140525X00032325. URL http://dx.doi.org/10.1017/S0140525X00032325.

[43] Leslie C. Aiello and R. I. M. Dunbar. Neocortex size, group size, and the evolution of language. *Current Anthropology*, 34(2):184–193, 1993.

[44] Tracey H. Joffe and R. I. M. Dunbar. Visual and socio–cognitive information processing in primate brain evolution. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1386):1303–1307, 1997. DOI: 10.1098/rspb.1997.0180. URL http://rspb.royalsocietypublishing.org/content/264/1386/1303.abstract.

[45] R. I. M. Dunbar. The social brain hypothesis and its implications for social evolution. *Annals of human biology*, 36(5):562–72, 2009. DOI: 10.1080/03014460902960289.

[46] R W Byrne and A Whiten. Tactical deception in primates: the 1990 database. *Primate Report, Whole Volume*, 27:1–101, 1990.

[47] R W Byrne and A Whiten. Computation and mindreading in primate tactical deception. In A Whiten, editor, *Natural Theories of Mind*. Blackwells, Oxford, 1991.

[48] R W Byrne and A Whiten. Cognitive evolution in primates: evidence from tactical deception. *Man*, 27:609–627, 1992.

[49] Andrew Whiten and David Erdal. The human socio-cognitive niche and its evolutionary origins. *Phil. Trans. R. Soc. B*, 367:2119–2129, 2012.

[50] R. I. M. Dunbar. Cognitive constraints on the structure and dynamics of social networks. *Group Dynamics: Theory, Research, and Practice*, 12(1):7–16, 2008.

[51] Susanne Shultz and Robin Dunbar. Encephalization is not a universal macroevolutionary phenomenon in mammals but is associated with sociality. *Proceedings of the National Academy of Sciences*, 107(50):21582–21586, 2010. DOI: 10.1073/pnas.1005246107. URL http://www.pnas.org/content/107/50/21582.abstract.

[52] Christopher Boehm, Harold B. Barclay, Robert Knox Dentan, Marie-Claude Dupre, Jonathan D. Hill, Susan Kent, Bruce M. Knauft, Keith F. Otterbein, and Steve Rayner. Egalitarian behavior and reverse dominance hierarchy [and comments and reply]. *Current Anthropology*, 34(3):pp. 227–254, 1993. ISSN 00113204. URL http://www.jstor.org/stable/2743665.

[53] T.K. Earle. *How Chiefs Come to Power: The Political Economy in Prehistory*. Antropology. Political science. Stanford University Press, 1997. ISBN 9780804728560. URL http://books.google.com.br/books?id=4AjAk15WSbQC.

[54] D.P. Landau and K. Binder. *A Guide To Monte Carlo Simulations In Statistical Physics*. Cambridge University Press, 2005. ISBN

9780521842389. URL http://books.google.com.br/books?id=11tun6Y5tOAC.

[55] R. I. M. Dunbar. *How Many Friends Does One Person Need? Dunbar's Number and Other Evolutionary Quirks*. Harvard University Press, 2010. ISBN 9780674057166. URL http://books.google.com.br/books?id=W-kkeffNx5MC.

[56] P. Wiessner. The vines of complexity: Egalitarian structures and the institutionalization of inequality among the Enga. *Current anthropology*, 43(2):233–269, 2002.

[57] Geoffrey Clark, Foss Leach, and Sue O'Connor, editors. *Islands of Inquiry: Colonisation, seafaring and the archaeology of maritime landscapes*. Terra Australis. ANU E Press, 2008. ISBN 9781921313905. URL http://books.google.com.br/books?id=NZ0-IiAZ_ZwC.

[58] K. Summers. The evolutionary ecology of despotism. *Evolution and Human Behavior*, 26(1):106–135, 2005.

[59] L. Rüschendorf. On the distributional transform, sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927, 2009.

[60] Olivier P. Faugeras. Sklar's theorem by probabilistic continuation and two consistency results. Working Paper, 2012. URL http://hal.archives-ouvertes.fr/docs/00/78/00/82/PDF/sklar_by_convolution4_plain.pdf.

[61] J. Aczél and Z. Daróczy. *On Measures of Information and Their Characterizations*. Mathematics in science and engineering vol.115. Acad. Press, 1975. ISBN 9780120437603. URL http://books.google.com.br/books?id=9Xsj7AWmV8UC.

[A] Teorema de Sklar

Apresentamos neste apêndice uma demonstração do teorema de Sklar baseada nas referências [59][1] e [60][2]. Em primeiro lugar, o teorema é enunciado da seguinte forma:

Teorema 5 (Teorema de Sklar.)**.** *Sejam X e Y duas variáveis aleatórias com distribuição cumulativa conjunta $F(x,y) = \text{Prob}\{X < x, Y < y\}$ e distribuições marginais $F_x(X)$ e $F_y(Y)$ respectivamente. Então:*

1. *Existe uma única função $C : [0,1]^2 \to \mathbb{R}$, chamada função cópula, tal que:*

$$F(x,y) = C(F_X(x), F_y(y)). \tag{A.1}$$

   Ou, de forma recíproca:

$$C(u,v) = F(F_x^{-1}(u), F_y^{-1}(v)) \tag{A.2}$$

   A função C, denominada cópula, é uma distribuição cumulativa com suporte em $[0,1]$.

2. *De forma recíproca, se $C(u,v)$ é uma função cópula, ou seja, uma distribuição cumulativa com suporte em $[0,1]$, e se $F_X(x)$ e $F_Y(y)$ são distribuições cumulativas univariadas com suporte em $\mathcal{X}$ e $\mathcal{Y}$ respectivamente, então a função definida por $F(x,y) = C(F_X(x), F_Y(y))$ é uma distribuição com suporte em $\mathcal{X} \times \mathcal{Y}$ e distribuições marginais $F_X(x)$ e $F_Y(y)$.*

Demonstração: caso contínuo. Inicialmente note que, caso $F(x,y)$ seja contínua, (A.1) pode ser demonstrada de forma simples com a transformação de variáveis:

$$U = F_X(X) \text{ e } V = F_Y(Y).$$

Note que U e V tomam valores no intervalo $[0,1]$ e são uniformemente distribuídas. Para provar o teorema de Sklar no caso contínuo, note que:

$$\begin{aligned} F(x,y) =& \text{Prob}\{X < x, Y < y\} \\ =& \text{Prob}\{F_X(X) < F_X(x), F_Y(Y) < F_Y(y)\} \end{aligned} \tag{A.3}$$

[1] L. Rüschendorf. On the distributional transform, sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927, 2009

[2] Olivier P. Faugeras. Sklar's theorem by probabilistic continuation and two consistency results. Working Paper, 2012. URL `http://hal.archives-ouvertes.fr/docs/00/78/00/82/PDF/sklar_by_convolution4_plain.pdf`

uma vez que as funções distribuição $F_X(x)$ e $F_Y(y)$ são monotônicas e contínuas. Seja $C(u,v) = \text{Prob}\{U < u, V < v\}$ a distribuição conjunta de U e V. Dessa forma:

$$F(x,y) = \text{Prob}\{U < F_X(x), V < F_Y(y)\} = C(F_X(x), F_Y(y)),$$

demonstrando assim a primeira parte do teorema. A segunda parte é trivialmente demonstrada notando que os seguintes fatos:

- as distribuições marginais de $C(u,v)$ são uniformes no intervalo $[0,1]$, o que implica que $F(x,y) = C(F_X(x), F_Y(y))$ é uma distribuição cumulativa, uma vez que $0 \leq F_X(x) \leq 1$ e $0 \leq F_Y(y) \leq 1$ são elas próprias distribuições cumulativas
- como $C(u,v)$ é uma distribuição cumulativa, então $C(1,v) = v$ e $C(u,1) = u$, o que implica que as marginais de $F(x,y)$ são dadas por

  $$F_X(x) = \text{Prob}\{X < x\} = F(x,\infty) = C(F_X(x), 1) = F_X(x)$$

  e similarmente para $F_Y(y)$.

□

Demonstração: caso descontínuo. O ponto onde a demonstração anterior falha para o caso de descontinuidades[3] em $F(x,y)$ é na equação (A.3). Para contornar esse problema, vamos definir a seguinte transformação de variáveis:

[3] Por exemplo, para variáveis aleatórias que tomam valores em conjuntos discretos, entre outros casos.

Definição 1 (Transformação distribucional)**.** *Seja X uma variável aleatória com distribuição cumulativa $F(x) = \text{Prob}\{X \leq x\}$, com suporte em $\mathcal{X}$. Seja a função $f(x,\lambda)$ dada por:*

$$f(x,\lambda) = \text{Prob}\{X \leq x\} + \lambda \text{Prob}\{X = x\}.$$

Define-se a transformação distribucional de X como a variável aleatória:

$$U = f(X,Z)$$

onde Z é uma variável aleatória uniformemente distribuida no intervalo $[0,1]$.

Para o caso em que $F(x)$ é contínua, $f(x,u) = F(x)$ e $U = F(X)$ é idêntica à variável aleatória homônica definida na demonstração anterior. Para outros casos, essa identidade não se observa, porém ainda se sustentam as seguintes propriedades[4]:

[4] Uma demonstração desse fato pode ser encontrada em Rüschendorf [59]

- $U = F(X,z)$ é uniformemente distribuída no intervalo $[0,1]$.
- $X = F^{-1}(U)$, onde $F^{-1}(\cdot)$ é a função quantil, definida por:

  $$F^{-1}(u) = \inf\{x \in \mathbb{R} | F(x) \geq u\}$$

De posse dessa definição podemos prosseguir com uma demonstração mais robusta do teorema de Sklar. Sejam X e Y variáveis aleatórias com função distribuição conjunta dada por $F(x,y)$. Sejam Z uma variável uniformemente distribuídas no intervalo $[0,1]$. Finalmente, sejam $U = f_X(X,Z)$ e $V = f_Y(Y,Z)$ as transformações distribucionais associadas a X e Y. Com anteriormente dito, são válidas as identidades $X = F_X^{-1}(U)$ e $Y = F_Y^{-1}(V)$. Dessa forma, denotando por $C(u,v)$ a distribuição conjunta de U e V, temos:

$$\begin{aligned} F(x,y) &= \text{Prob}\{X < x, Y < y\} \\ &= \text{Prob}\left\{F_X^{-1}(U) < x, F_Y^{-1}(V)\right\} \\ &= \text{Prob}\{U < F_X(x), V < F_Y(y)\} \\ &= C(F_X(x), F_Y(y)) \end{aligned}$$

A demonstração da segunda parte do teorema segue similarmente ao caso contínuo. □

⌈B⌉ Teoremas de Cox

Este apêndice contém demonstrações dos teoremas exibidos na seção 1.2.1, *Probabilidades e Inferência*. O método de demonstração consiste em considerar casos em que há informação completa, ou seja, em que é válida a lógica tradicional, em que existam duas formas diferentes de determinar a mesma plausibilidade. Impondo consistência sobre essas situações, se obtém equações funcionais cujas soluções implicam no resultado desejado.

⌈B.1⌉ Primeiro teorema de Cox e a regra do produto

Teorema 6. *Uma vez que uma representação consistente de plausibilidades $(P|Q)$ com um ordenamento bem definido foi encontrada, sempre é possível encontrar uma outra equivalente $\pi(P|Q)$, de forma que:*

$$\pi(P_1 \wedge P_2|Q) = \pi(P2|Q \wedge P_1)\pi(P_1|Q) \tag{B.1}$$

Demonstração. A função $G(u, v)$ define a decomposição da plausibilidade de uma conjunção de duas proposições:

$$(P_1 \wedge P_2|Q) = G((P_1|Q), (P_2|Q \wedge P_2)).$$

A conjunção booleana ($\wedge$) é uma operação associativa, ou seja:

$$P_1 \wedge (P_2 \wedge P_3) = (P_1 \wedge P_2) \wedge P_3.$$

Isso implica também na associatividade da função $G(u, v)$, ou seja:

$$G(G(u, v), w) = G(u, G(v, w)).$$

Esse vínculo é satisfeito por infinitas possíveis funções $G(u, v)$, porém todas elas [1] têm a forma:

$$G(u, v) = g^{-1}(g(u)g(v))$$

com $g(\cdot)$ uma função monotônica. Sendo $g(u)$ monotônica, pode-se redefinir a atribuição de números reais às plausibilidades para $g(P|Q)$ sem perder o ordenamento de proposições segundo suas plausibilidades. Escrevendo portanto:

$$\pi(P|Q) = g((P|Q))$$

[1] J. Aczél and Z. Daróczy. *On Measures of Information and Their Characterizations*. Mathematics in science and engineering vol.115. Acad. Press, 1975. ISBN 9780120437603. URL http://books.google.com.br/books?id=9Xsj7AWmV8UC; and Ariel Caticha. Lectures on probability, entropy, and statistical physics. *arXiv:0808.0012*, 2008. URL http://arxiv.org/abs/0808.0012. Invited lecture at the 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (July 8-13, 2008, Boraceia Beach, Sao Paulo, Brazil)

se obtém a equação (B.1), que é reminiscente da regra do produto da Teoria das Probabilidades:

$$\pi(P_1 \wedge P_2|Q) = \pi(P2|Q \wedge P_1)\pi(P_1|Q)$$

□

⌈B.2⌉ Valores extremos

Teorema 7. *Sejam*[2] P_T *e* P_F *os valores associados à plausibilidade regraduada* $\pi(\cdot|\cdot)$ *de eventos sabidamente verdadeiros ou falsos, respectivamente*[3]*. Temos necessariamente que:*

$$P_T = 1 \text{ e } P_F = 0, \text{ ou} \tag{B.2}$$

$$P_T = 1 \text{ e } P_F = \infty \tag{B.3}$$

[2] Das palavras inglesas *"true"* e *"false"*, respectivamente.

[3] Requisitos de consistência exigem que sejam iguais para quaisquer proposições falsas ou verdadeiras

Demonstração. Se P é sabido verdadeiro, então a plausibilidade de que P e Q sejam simultaneamente verdadeiros é exatamente a plausibilidade de apenas Q ser verdadeiro, ou seja, $\pi(P \wedge Q|P) = \pi(Q|P)$. Mas, pela regra do produto:

$$\pi(P \wedge Q|P) = \pi(P|P)\pi(Q|P \wedge P) = P_T\pi(Q|P)$$

Dessa forma, $P_T\pi(Q|P) = \pi(Q|P)$, para quaisquer Q e P, o que implica em $P_T = 1$.

Da mesma forma, a plausibilidade de que simultaneamente P e $\bar{P}$ sejam verdadeiros, dada uma proposição Q qualquer, deve ser P_F, pois $P \wedge \bar{P}$ é uma contradição. Mas, pela regra do produto:

$$\pi(P \wedge \bar{P}|Q) = \pi(P|Q \wedge \bar{P})\pi(\bar{P}|Q)$$

Independentemente de Q, $\pi(P|Q \wedge \bar{P})$ deve ser também igual a P_F e, assim, $P_F = P_F\pi(\bar{P}|Q)$, para quaisquer P e Q. Duas soluções são possíveis: $P_F = 0$ ou $P_F = \infty$. Uma vez que quaisquer das soluções para P_F pode ser mapeada na outra por uma regraduação monotônica[4], pode-se arbitrariamente escolher $P_F = 0$, e assim limitar valores de $\pi(P|Q)$ no intervalo $[0, 1]$. □

[4] Consistindo da função $h(u) = \frac{1}{u}$.

⌈B.3⌉ Teorema de Bayes

Teorema 8. *Para qualquer conjunto de três proposições* P_1*,* P_2 *e* Q*, temos:*

$$\pi(P2|Q \wedge P_1) = \frac{\pi(P1|Q \wedge P_2)\pi(P_2|Q)}{\pi(P_1|Q)}. \tag{B.4}$$

Demonstração. Uma conseqüência imediata da regra do produto segue da seguinte observação. Uma vez que a conjunção $P \wedge Q$ é simétrica:

$$\pi(P_1 \wedge P_2|Q) = \pi(P_2 \wedge P_1|Q)$$

Aplicando a regra do produto em ambos os membros da equação acima, temos:

$$\pi(P2|Q \wedge P_1)\pi(P_1|Q) = \pi(P1|Q \wedge P_2)\pi(P_2|Q)$$

que implica trivialmente na equação (B.4) que é similar ao teorema de Bayes da Teoria de Probabilidades. □

⌈B.4⌉ Regra da soma

Para demonstrar a regra da soma são necessários os seguintes lemas:

Lema 1. *Dadas as proposições P, S e $Q = \overline{P \wedge S}$ temos que:*

$$P \wedge \bar{Q} = \bar{Q} \tag{B.5}$$

$$\bar{P} \wedge Q = \bar{P} \tag{B.6}$$

Demonstração. As equações decorrem diretamente das leis de De Morgan[5]. Note, em primeiro lugar, que:

$$P \wedge \bar{Q} = P \wedge (P \wedge S) = P \wedge S = \bar{Q}.$$

Note ainda que:

$$\overline{\bar{P} \wedge Q} = P \vee \bar{Q} = P \vee (P \wedge S) = P,$$

e portanto $\bar{P} \wedge Q = \bar{P}$ e $P \wedge \bar{Q} = \bar{Q}$. □

[5] Leis de De Morgan são as seguintes relações distributivas da negação:

$$\overline{A \wedge B} = \bar{A} \vee \bar{B},$$

e, similarmente,

$$\overline{A \vee B} = \bar{A} \wedge \bar{B},$$

válidas para quaisquer proposições A e B.

Teorema 9 (2º teorema de regraduação de Cox)**.** *Uma vez que uma representação consistente de plausibilidades $\pi(P|Q)$ com um ordenamento bem definido foi encontrada para a qual vale a regra do produto, sempre é possível encontrar uma outra equivalente $p(P|Q)$ tal que:*

$$p(P|Q) + p(\bar{P}|Q) = 1 \tag{B.7}$$

Demonstração. A função $F(u, v)$ define a relação entre a plausibilidade de uma proposição e a plausibilidade de sua negação:

$$(\bar{P}|Q) = F((P|Q)).$$

Considere a seguinte plausibilidade regraduada dada por

$$\pi(P \wedge Q|R) = \pi(P|R)\pi(Q|P \wedge R).$$

Note que a função $F(\cdot)$ deve ser idempotente, uma vez que $\bar{\bar{P}} = P$ implica que $F(F(u)) = u$. Portanto,

$$\pi(Q|S) = F(\pi(\bar{Q}|S))$$

para qualquer S. Assim, usando repetidamente a regra do produto:

$$\pi(P \wedge Q|R) = \pi(P|R)F\left(\pi(\bar{Q}|P \wedge R)\right) \tag{B.8}$$

$$= \pi(P|R)F\left(\frac{\pi(\bar{Q} \wedge P|R)}{\pi(P|R)}\right). \tag{B.9}$$

Mas a mesma operação pode ser feita em outra ordem — uma vez que a conjunção $P \wedge Q$ é simétrica na troca de P por Q e, portanto:

$$\pi(P \wedge Q|R) = \pi(P|R)F\left(\frac{\pi(\bar{Q} \wedge P|R)}{\pi(P|R)}\right) = \pi(Q|R)F\left(\frac{\pi(\bar{P} \wedge Q|R)}{\pi(Q|R)}\right)$$

para quaisquer P, Q e R. Em particular, deve valer para o caso particular em que $Q = \overline{P \wedge S}$. Nesse caso se aplicam os resultados do Lema 1 e, portanto, $P \wedge \bar{Q} = \bar{Q}$ e $\bar{P} \wedge Q = \bar{P}$. Isso permite escrever:

$$\pi(P|R)F\left(\frac{\pi(\bar{Q}|R)}{\pi(P|R)}\right) = \pi(Q|R)F\left(\frac{\pi(\bar{P}|R)}{\pi(Q|R)}\right)$$

e, finalmente, sendo arbitrárias as proposições P e S, isso significa que:

$$uF\left(\frac{F(v)}{u}\right) = vF\left(\frac{F(u)}{v}\right)$$

Novamente, há infinitas soluções $F(\cdot)$ para esse vínculo, mas todas elas satisfazem[6]:

$$F(u)^\alpha + u^\alpha = 1, \tag{B.10}$$

para algum $\alpha \in \mathbb{R}$. É possível, portanto, regraduar as plausibilidades novamente por uma transformação monotônica $p(P|Q) = \pi(P|Q)^\alpha$. Se $\alpha > 0$, essa regraduação preserva os valores $P_F = 0$ e $P_V = 1$ e o teorema de Bayes. Caso $\alpha < 0$, sempre é possível aplicar a regraduação $h(u) = \frac{1}{u}$, que também é monotônica, e preservar os valores limite. Esse resultado permite enunciar o segundo teorema de Cox:

$$p(P|Q) + p(\bar{P}|Q) = 1$$

□

[6] J. Aczél and Z. Daróczy. *On Measures of Information and Their Characterizations*. Mathematics in science and engineering vol.115. Acad. Press, 1975. ISBN 9780120437603. URL http://books.google.com.br/books?id=9Xsj7AWmV8UC; and Ariel Caticha. Lectures on probability, entropy, and statistical physics. *arXiv:0808.0012*, 2008. URL http://arxiv.org/abs/0808.0012. Invited lecture at the 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (July 8-13, 2008, Boraceia Beach, Sao Paulo, Brazil)