

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional

Distribuições Conjugadas e Aproximações

Maria Madalena de Freitas Malva

Doutoramento em Estatística e Investigação Operacional
(Probabilidades e Estatística)

Ano 2006

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional

Distribuições Conjugadas e Aproximações

Tese Orientada por: Professor Doutor Dinis Pestana
Professora Doutora Sandra Mendonça

Maria Madalena de Freitas Malva

Investigação subsidiada por

Dissertação apresentada à Faculdade de Ciências
da Universidade de Lisboa, para obtenção do grau
de Doutor em Probabilidades e Estatística.

Agradecimentos

Aos meus orientadores Professor Doutor Dinis Pestana e Professora Doutora Sandra Mendonça desejo expressar o meu reconhecimento por me terem proposto o tema deste trabalho, pela sua orientação, pela confiança, pelo apoio e pela permanente disponibilidade durante a sua realização. Sem a vossa amizade este trabalho teria sido sem dúvida mais difícil.

À Escola Superior de Tecnologia de Viseu, em especial ao Departamento de Matemática o meu muito obrigado por me terem disponibilizado as condições necessárias para a realização deste trabalho.

Aos elementos do Centro de Estatística e Aplicações do Departamento de Estatística e Investigação Operacional, o meu obrigado pelas facilidades concedidas. Esta dissertação foi apoiada por: FCT/POCTI/FEDER, Projecto VEXTRA.

Um obrigado especial ao meu colega de gabinete por ter aturado algumas das minhas dúvidas durante este período sempre com uma palavra de ânimo.

A todos os meus colegas, amigos e familiares desejo agradecer o interesse e a paciência demonstrados, em particular, aos meus Pais e ao Tito por terem sido quem mais de perto me acompanhou.

Um miau cheio de ternura para a Xica.

A todos a minha gratidão.

Maria Madalena de Freitas Malva

Viseu, Abril de 2006

Aos meus Pais

e ao

Tito

Resumo

Resultados assintóticos são inspiração e justificação de importantes aproximações usadas em modelação estatística, mesmo quando o intuito é analisar uma mão-cheia de dados. Há, de facto, situações em que a convergência é rápida, mas muitas outras são conhecidas em que a convergência é deveras lenta. Os estudos sobre velocidades de convergência são naturalmente uma parte essencial do *corpus* das convergências estocásticas.

Ao Teorema Limite Central, se o cerne da investigação não se situar nas questões de velocidade de convergência, basta existência de segundo momento (e mesmo esta exigência pode ser relaxada, usando o conceito de variação lenta de Karamata). Terceiro e quarto momentos — e portanto, indirectamente, assimetria e achatamento da população parente —, por outro lado, são os instrumentos adequados para a abordagem inicial das questões de velocidade de convergência.

Reexpressando assimetria e achatamento em termos dos cumulantes de Thiele–Fisher, e retomando ideias implícitas nas expansões de Gram–Charlier e de Edgeworth, por um lado, e de Cornish–Fisher por outro, reencontramos outros instrumentos preciosos, tais como as transformadas de Esscher e distribuições conjugadas de Cramér–Khintchine, e as aproximações usando ponto de sela (que são a base de toda a área de *small sample asymptotics*); as expansões de Edgeworth diferidas (por vezes apodadas de *tilted Edgeworth expansions*), são, de facto, o traço de união de todos estes resultados.

Por outro lado, aproximações excelentes são válidas em situações inesperadas, como demonstramos com o exemplo da estável de Lévy, que apesar de não ter sequer primeiro momento pode ser aproximada de forma muito adequada com aqueles instrumentos, que pareciam ter sido talhados para circunstâncias bem diversas.

A investigação de expansões de Edgeworth diferidas levou-nos naturalmente ao estudo de famílias exponenciais (e polinómios ortogonais associados), e particularmente às famílias naturais de Morris, com variância que é função quando muito quadrática do valor médio.

Por outro lado, o exemplo da Lévy levou-nos a concentrar algum esforço na aproximação por leis estáveis para somas, obtendo resultados parciais mais interessantes no caso de a distribuição parente estar no domínio de atracção *standard*.

Por outro lado, o recurso a *tilting* permite bons resultados na chamada zona de grandes

desvios, e abre a perspectiva de a velocidade de convergência no Teorema Limite Central melhorar de $O\left(\frac{1}{\sqrt{n}}\right)$ para $O\left(\frac{1}{n}\right)$, desde que se use devidamente a teoria de que Daniels foi pioneiro. Uma incursão pela teoria da informação, perspectivando resultados assintóticos como os que correspondem a entropia máxima em situações bem tipificadas, ajuda a compreender este progresso notável.

Palavras-Chave: Teorema limite central, polinômios ortogonais, famílias exponenciais naturais — função de variância quadrática, expansões assintóticas, teoria da informação, entropia.

Abstract

Asymptotic results are an inspiration and justification for important approximations used in statistical modulation. Even when the intention is to analyse a full-set of data there are situations where the convergence is fast, but many others are known where convergence is slow. Studies about convergence rates are naturally an essential part of the “corpus” of stochastic convergences.

In the Central Limit Theorem, if the main purpose of the investigation is not based on questions of convergence speed, the existence of a second moment is sufficient (and even this requirement can be relaxed, using the concept of Karamata’s regular variation). On the other hand, when questions of convergence speed are the main issue, the third and fourth moments — and hence, skewness and kurtosis of the parent population — are at the core of the matter.

Expressing skewness and kurtosis in terms of Thiele-Fisher cumulants, and retaking ideas implicit in Gram-Charlier series and the Edgeworth and the Cornish-Fisher expansions, we find other valuable instruments, such as Esscher transforms and Cramér-Khintchine tilted distributions, and saddlepoint approximations, which are the basis of the whole area of small asymptotics; the tilted Edgeworth expansions are, in fact, the line that unites all these results.

Nevertheless, excellent approximations are valid in some unexpected situations, as we demonstrate in the stable Lévy law. In spite it has not finite first moment, it can be adequately approximated by an appropriate Edgeworth expansion, a tool which seems tuned out for a wide variety of circumstances.

The investigation of tilted Edgeworth expansions brought us to the study of exponential families (and associated orthogonal polynomials), in particular the natural Morris families, whose variances are at most a quadratic function of their mean.

On the other hand, the Lévy example led us make some effort in approximating via stable laws for sums, thereby obtaining partial results showing that the approximations are neater when the parent distribution is in the *standard* attraction domain of a stable law.

On the other hand, resorting to tilting allows good results in the so-called great deviation region, and opens up the possibility of the convergence speed improving from $O\left(\frac{1}{\sqrt{n}}\right)$

to $O\left(\frac{1}{n}\right)$, as in Daniels pioneering work. New developments in information theory, and in particular the investigation of classes of maximal entropy laws under special assumptions — the Gaussian being the maximal entropy law under the finite variance assumption — sheds a new light in general forms of the central limit problem under which the $O\left(\frac{1}{n}\right)$ rate of convergence appears as what should be expected.

Keywords: Central limit theorem, orthogonal polynomials, natural exponential families — quadratic variance function, asymptotic expansions, information theory, entropy.

Conteúdo

Resumo	1
Abstract	3
1 Introdução	9
1.1 O Mundo Normal	9
1.2 Um Mundo Estável?	17
1.3 Um Mundo de Infinitamente Divisíveis e de Irredutíveis	19
1.4 Quando as Somas são Aleatórias	23
1.5 Uma Perspectiva dos Resultados Obtidos	24
2 Polinómios Ortogonais e Expansões Assintóticas	27
2.1 Transformadas Integrais. Momentos e Cumulantes	27
2.2 Exposição Breve sobre Famílias Exponenciais	32
2.2.1 Desigualdade de Cramér – Rao	34
2.2.2 Condições para atingir o Limite Inferior na Desigualdade de Cramér-Rao	37
2.3 Polinómios Ortogonais e Famílias Exponenciais	40
2.4 Famílias Exponenciais Naturais com Função de Variância Quadrática	41
2.4.1 Polinómios Ortogonais para NEF-QVF	46
2.5 Expansões Assintóticas	51
2.5.1 Expansões de Edgeworth	52
2.5.2 Expansões de Edgeworth Diferidas	57
2.5.3 Expansões Assintóticas de Distribuições Estáveis	61
2.5.4 Expansões de Cornish–Fisher e Cornish–Fisher Diferidas	64

2.5.5	Expansões com Base no Ponto Sela	66
2.6	Relações entre Distribuições Conjugadas e Somas Aleatórias	69
2.6.1	Poisson Dual e Filtragem Uniforme Discreta	74
2.6.2	Funções Características e Dualidade	76
2.6.3	Caracterização da Exponencial e Geométrica pela Auto-Dualidade	77
2.6.4	Fórmulas de Pollaczek–Khintchine e de Beckman e Interpretação Probabilística da Dualidade	79
3	O TLC no Contexto da Teoria da Informação e Entropia	83
3.1	Entropia e Distância de Kullback-Leibler	83
3.2	Distribuições com Entropia Máxima	90
3.3	Informação de Fisher	92
3.3.1	Distribuições com Informação de Fisher Mínima	99
3.4	Distância de Kullback-Leibler e Informação de Fisher — Algumas Aplicações	103
3.4.1	Desigualdade de Cramér-Rao Generalizada	105
4	Aproximações por Leis Limites Estáveis — Mais uns Passos em Volta	111
4.1	Teorema Limite Central e suas Generalizações	111
4.2	Domínios de Atracção e Leis Estáveis para Somas	114
4.2.1	Leis Infinitamente Divisíveis	114
4.2.2	Domínios de Atracção e Leis Estáveis	115
4.2.3	Funções de Variação Regular e Lenta	117
4.3	Velocidades de Convergência em Caudas com Condição de Regularidade de 2ª ordem	119
4.4	Convoluções de Leis Estáveis	122
	Conclusão	125
	A Alguns Cálculos	127
	B Mais Alguns Cálculos	129
	Bibliografia	131
	Índice Remissivo	138

Notação

Ao longo deste trabalho faremos uso da seguinte notação:

O símbolo/ A expressão	designa/abrevia/substitui
va	variável aleatória (<i>va's</i> no plural)
aa	amostra aleatória
fd	função de distribuição de probabilidade
fdp	função de densidade de probabilidade
fc	função característica
fgc	função geradora de cumulantes
fgm	função geradora de momentos
TLC	Teorema Limite Central
LGN	Lei dos Grandes Números
iid	independentes e identicamente distribuídas
F_X	função distribuição da <i>va</i> X
f_X	função densidade de probabilidade da <i>va</i> X
φ_X	função característica da <i>va</i> X
$\psi_X = K_X$	função geradora de cumulantes da <i>va</i> X
M_X	função geradora de momentos da <i>va</i> X
NEF	família exponencial natural
QVF	função de variância quadrática
k_i	i -ésimo cumulante de uma <i>va</i>
ρ_i	i -ésimo cumulante padronizado de uma <i>va</i>
$\hat{\theta}$	estimador
$\{X_i\}_{i \in \mathbb{N}}$	sucessão de <i>va's</i>
$X \sim Dist$	X com distribuição Dist
$X_k \stackrel{d}{=} Y$	igual em Dist
\equiv	equivalente
\star	convolução
S_n	soma de n <i>va's iid</i>
S_n^*	soma de n <i>va's iid</i> , do tipo $X^* = \frac{X - b}{c}$ com $c \neq 0$ e b constantes
\square	indica o final de uma demonstração
\blacksquare	indica o final de um exemplo

Capítulo 1

Introdução

1.1 O Mundo Normal

O *Ars Conjectandi* de J. Bernoulli⁽¹⁾, publicado postumamente em 1713 pelo seu sobrinho Nicolaus Bernoulli, é um salto qualitativo importante no desenvolvimento da Probabilidade. Apesar de o Livro I ser “apenas” uma reedição recheada de anotações de *De Ratiociniis in Ludo Aleae* (1657) de Huyghens, e de o Livro IV ter ficado incompleto, neste tratado, que no próprio título presta homenagem a Pascal⁽²⁾, transparece pela primeira vez a percepção de que a probabilidade pode medir a incerteza sobre asserções relativas a resultados de experiências nas quais não há a simetria inerente aos artefactos de jogo, conducente à equiprobabilidade dos acontecimentos elementares que tinha servido de referência aos seus predecessores; nele se encontra, por isso, o embrião do “princípio da razão insuficiente” que mais tarde Laplace (1814) há-de popularizar, e sobretudo uma justificação para as aplicações dos raciocínios probabilistas na construção das tabelas de mortalidade de Graunt e de Halley (1662): o resultado notabilíssimo, que Bernoulli chamava com ternura de seu “*theorema aureum*”, que posteriormente Poisson (1827) crismou como Lei dos Grandes Números (*LGN*), em que frequência relativa e probabilidade se abraçam. Pela mesma altura, John Arbuthnot (1710) inova também, com o que David e Edwards (2001) consideram o primeiro teste de hipóteses, mostrando que no estudo de fenómenos naturais faz mais sentido avaliar a probabilidade pela frequência relativa do que inventando equilíbrios que os factos teimam em negar.

Vejamos um pouco mais em detalhe a *LGN*. Em formulação moderna, sejam X_1, X_2, \dots variáveis aleatórias independentes e identicamente distribuídas (*iid*),

⁽¹⁾ Estas referências “históricas” não são anotadas na bibliografia. Consulte-se Pestana e Velosa (2006) para indicações sobre edições acessíveis.

⁽²⁾ Há a convicção generalizada de que Bernoulli escolheu este título por evocar o *Ars Cogitandi*, a *Lógica de Port-Royal*, em cuja parte III, onde muitos vêm a mão do jansenista Pascal, se defende que na tomada de decisões não interessam apenas presumíveis ganhos e perdas, há que atender também às respectivas probabilidades.

$\mathbb{E}(X_k) = \mu$, $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Como a existência de $\mathbb{E}(X_k) = \mu$ indica que $\mathbb{P}\left[\frac{X_{n+1}}{n+1} \geq \varepsilon\right] \xrightarrow{n \rightarrow \infty} 0$ segue-se que $\bar{X}_{n+1} = \frac{n}{n+1} \bar{X}_n + \frac{X_{n+1}}{n+1} \approx \bar{X}_n$ — por outras palavras, sob a condição de a soma das caudas ter um peso moderado, a média empírica tende a estabilizar; prova-se então facilmente (modernamente, recorrendo ao teorema da continuidade de Lévy, e ao desenvolvimento de ordem 1 da função característica de \bar{X}_n) que $\bar{X}_n \xrightarrow[n \rightarrow \infty]{d} \mu$.

A convergência da média amostral para o valor médio populacional estabelecida pela *LGN*, com hipóteses tão simples, necessitava evidentemente de um complemento: um resultado indicando a velocidade com que tal ocorre, e consequentemente a qualidade das aproximações que sugere. O Teorema Limite Central (*TLC*), cuja forma embrionária surgiu na segunda edição de *The Doctrine of Chances* de Abraham de Moivre, pode ser visto nessa perspectiva: com a hipótese suplementar de a população parente ter variância σ^2 , prova-se — modernamente também recorrendo à expansão em série de MacLaurin da função característica de $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$, desta feita até à ordem 2, e ao teorema da continuidade de Lévy — que $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} Z \sim \text{Gaussiana}(0, 1)$ ⁽³⁾.

Bernoulli, ao apodar o seu resultado de teorema de ouro, estava decerto consciente que tinha criado uma jóia da Teoria da Probabilidade (Hoffman-Jørgensen (1976) intitulou *The Two Pearls of Probability* um notável curso sobre a *LGN* e o *TLC*). No caso especial de $X_k \sim \text{Bernoulli}(p)$ serem as variáveis indicatrizes associadas a um acontecimento A , segue-se que a frequência relativa com que A é observado converge para a probabilidade de A , uma asserção que consola os frequentistas, e está na base do paradigma da Estatística com base na noção de *colectivo*, desenvolvida por von Mises (1928) nos anos 20 do século passado, veja-se também Ville (1939). O *TLC*, por seu lado, permitia em condições muito

⁽³⁾ Adoptamos a nomenclatura “lei gaussiana” em lugar da mais tradicional “lei normal”. Pearson (1920) escreveu:

“Many years ago I called the Laplace-Gaussian curve the normal curve, which name, while it avoids the international question of priority, has the disadvantage of leading people to believe that all other distributions of frequency curves are in one sense or another abnormal.”

De facto, já Peirce (1873) tinha usado a expressão “distribuição normal” antes de Pearson (1893), e o uso, tantas vezes acrítico, da aproximação pela lei “normal” invocando aditividade subjacente tornou-se norma na primeira metade do século XX. A partir dos anos 60 do século passado, a disponibilidade de novos meios computacionais permitiu o recurso rotineiro a modelos mais próximos dos fenómenos que se pretende estudar, e a escola de *Exploratory Data Analysis* liderada por Tukey reclamou, com algum sucesso, o retorno à nomenclatura que adoptamos.

A gaussiana foi de facto introduzida por de Moivre, que em 1733 obteve o resultado sobre a aproximação da binomial pelo integral de $e^{-\frac{x^2}{2}}$; ilustres historiadores da Probabilidade registam o “teorema da atribuição errada”: a nomenclatura usual dos grandes resultados não homenageia, em regra, os seus criadores. É em obediência a este teorema que não usamos “lei de de Moivre”, designação que aliás todos estranhariam.

satisfatórias para as aplicações práticas, recorrer à aproximação de \bar{X}_n por uma variável aleatória $Y \sim \text{Gaussiana}(\mu, \frac{\sigma}{\sqrt{n}})$. É a base, por exemplo, da construção de estimadores intervalares com probabilidade de cobertura aproximadamente $(1 - \alpha) \times 100\%$ para \bar{X}_n , instrumentais na construção de intervalos de confiança de parâmetros em populações Binomiais, geométricas, Poisson, exponenciais, etc.

O *TLC* começou a sua carreira de forma modesta, dando uma aproximação integral à soma de átomos de probabilidades binomiais. Em Laplace (1814) tem já uma formulação geral, para modelos aditivos, e ao longo do século XIX há um longo trabalho de relaxação das hipóteses sobre existência dos momentos, em que se baseia a sua demonstração clássica. O uso de funções características trouxe a possibilidade de invocar apenas a existência do momento de ordem 2, e a teoria das funções de variação regular de Karamata (1930) mostra que a existência de momento de segunda ordem é condição suficiente, mas não necessária. No entanto, na presente secção vamos admitir que as populações parentes têm variância (na linguagem da secção que se segue, estão no domínio de atracção *standard* da gaussiana). Se, quer do ponto de vista de elegância formal das demonstrações quer do ponto de vista de criação de novos instrumentos e metodologias, todo esse trabalho foi um progresso notável, veremos que, no que se refere às questões de velocidade de convergência, funções simples do terceiro e quarto momentos (mais precisamente, coeficientes de assimetria e de achatamento) dão informações cruciais a esse respeito, inclusivamente permitindo estabelecer convergências $O\left(\frac{1}{n}\right)$ em vez de $O\left(\frac{1}{\sqrt{n}}\right)$.

Liapunov (1901) demonstrou a convergência da soma, convenientemente centrada e reduzida, de variáveis independentes (não necessariamente identicamente distribuídas) para a gaussiana padrão, assumindo existência de terceiro momento. O resultado mais notável nesta direcção deve-se a Lindeberg (1920, 1922) que, assumindo que a sucessão das variâncias truncadas é uniformemente limitada, estabeleceu a generalização daquele resultado usando operadores de contracção, uma técnica que se adapta a demonstrações de extensões para variáveis dependentes, que não são abordáveis com o método das funções características de Lévy. Feller (1935, 1937) viria a demonstrar a necessidade das condições de Lindeberg, estabelecendo que são essencialmente equivalentes às de Liapunov.

Todos estes desenvolvimentos trouxeram o *TLC* para as luzes da ribalta, e Pólya (1920) refere-o pela primeira vez como “o teorema central” dos limites de somas, expressão que como se sabe foi afortunada.

Jacques Bernoulli e Abraham de Moivre inauguraram, assim, o filão dos resultados assintóticos. Gnedenko e Kolmogoroff (1954), no prefácio em si mesmo notável de um livro tão justamente célebre e perene, afirmam que o valor gnoseológico da Teoria da Probabilidade advém, sobretudo, dos resultados assintóticos reunidos sob o nome geral de *TLC* (veja-se as secções seguintes). Tal afirmação, que se justificava plenamente à época da publicação, anterior às possibilidades novas trazidas pelo cálculo automático, continua a fazer sentido: cita-se Tukey, porventura o maior impulsionador da mudança da Estatística numa perspectiva computacional, como afirmando que mais vale um resultado aproximado de um problema com formulação exacta do que um resultado exacto de um

problema formulado de forma meramente “aproximativa”.

Evidentemente — talvez não de imediato após os seus desenvolvimentos incipientes, mas actualmente decerto — tal como o *TLC* dá alguma indicação sobre a velocidade da convergência de \bar{X}_n para μ afirmada na *LGN*, torna-se também importante caracterizar a velocidade de convergência de $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ para a gaussiana limite, ou seja avaliar a qualidade da aproximação de \bar{X}_n por $Y \sim \text{Gaussiana}(\mu, \frac{\sigma}{\sqrt{n}})$. Sabe-se que essa convergência é rápida quando a população parente é simétrica, e resultados gerais como o teorema de Berry-Esséen (Berry, 1941; Esséen, 1945) estabelecem limites para $|F^{*n}(n\mu + \sigma\sqrt{nx}) - \Phi(x)|$, onde como é usual F^{*n} denota a convolução de ordem n da função de distribuição F de X — isto é, a função de distribuição de $\sum_{k=1}^n X_k$ —, e Φ denota a função de distribuição da gaussiana standard. Veremos que resultados anteriores, as expansões em série de Edgeworth (1905), que usaremos profusamente numa parte substancial desta dissertação, vieram a revelar-se um instrumento adequado para abordar a questão.

Mais modernamente, a Lei do Logaritmo Iterado (*LLI*) de Khintchine (1960) — que só uma vintena de anos mais tarde, num excepcional trabalho de Hartman e Winter, se tornaria uma condição necessária e suficiente — introduziu a questão dos grandes desvios.

Nesse contexto, Daniels (1956) — num trabalho republicado entre os “pioneiros” arrolados por Johnson and Kotz (1995) — conseguiu alguns resultados espectaculares importando as aproximações usando pontos de sela para a Estatística. A análise cuidada desses resultados leva-nos a admirar também o pioneirismo de Cramér (1928), que tinha introduzido a inovação de “atirar” a distribuição para um ponto junto daquele em que se pretendia aproximar, para obter resultados melhorados na região de grandes desvios — aquilo que actualmente se filia nas técnicas de *tilting* e *back-tilting* usando a expansão de Edgeworth “diferida”. Cramér credits Esscher com a ideia de usar “distribuições conjugadas” (o nome que ele usa para a *tilted distribution*); as “distribuições conjugadas” permitem de facto a demonstração elegante de resultados difíceis de estabelecer com técnicas clássicas, por exemplo a fórmula de Pollaczeck-Khinchine Malva e Sequeira, (2003).

A importância do *TLC* justifica plenamente a recorrente publicação de novas demonstrações. Um dos progressos realizados, como já referimos, foi assumir cada vez menos no que se refere a momentos. Porém esse progresso, que não contestamos, acabou por ter um efeito secundário pernicioso: a velocidade de convergência no *TLC* é $O(\frac{1}{\sqrt{n}})$, bastante lenta. Ora a demonstração do *TLC* clássico baseada na ideia de entropia máxima deixa aberta a possibilidade de a velocidade de convergência ser substancialmente melhorada, para $O(\frac{1}{n})$, pelo menos fora da zona de grandes desvios. E, no que se refere essa zona, alguma coisa se pode fazer com *tilting* e *back-tilting*, usando expansões de Edgeworth diferidas, que se mostra serem equivalentes às aproximações usando pontos de sela de Daniels, pelo menos no caso de assimetria e achatamento serem moderados. É um dos objectivos da presente dissertação, que abordaremos com mais detalhe em capítulos ulteriores.

Deixando embora a questão das expansões de Edgeworth e as expansões de Edgeworth diferidas para mais tarde, arrumamos de imediato algumas questões simples sobre momentos, transformadas integrais e cumulantes da distribuição gaussiana, relevantes em si mesmas, e essenciais para abordar posteriormente aquelas expansões:

Seja $Z \sim \text{Gaussiana}(0, 1)$, cuja função densidade de probabilidade $f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathbb{I}_{\mathbb{R}}(x)$. O cálculo dos momentos de ordem par obtém-se então facilmente recorrendo à função gama,

$$\mathbb{E}(Z^{2k}) = \int_{-\infty}^{\infty} x^{2k} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{2^k}{\sqrt{\pi}} \int_0^{\infty} y^{k-\frac{1}{2}} e^{-y} dy = \frac{2^k \Gamma(k + \frac{1}{2})}{\sqrt{\pi}} = \frac{(2k)!}{2^k k!}.$$

Segue-se naturalmente que todos os momentos de ordem ímpar existem; e como só podem ser o valor principal de Cauchy, da imparidade da função integranda deduz-se que $\mathbb{E}(Z^{2k+1}) = 0$.

Assim, $\mathbb{E}(Z) = 0$ e $\text{var}(Z) = 1$ — genericamente, se $\mu + \sigma Z = X \sim \text{Gaussiana}(\mu, \sigma)$, $\mathbb{E}(X) = \mu$ e $\text{var}(X) = \sigma^2$. E do desenvolvimento formal da função característica $\varphi_Z(t) = \mathbb{E}(e^{itX})$ em série de MacLaurin,

$$\varphi_Z(t) = \sum_{n=0}^{\infty} \frac{\mathbb{E}(Z^n) (it)^n}{n!} = \sum_{k=0}^{\infty} \frac{\left(-\frac{t^2}{2}\right)^k}{k!} = e^{-\frac{t^2}{2}},$$

e conseqüentemente a função característica de $\mu + \sigma Z = X \sim \text{Gaussiana}(\mu, \sigma)$ é $\varphi_X(t) = \exp\left(i\mu t - \frac{(\sigma t)^2}{2}\right)$, de que se deduz imediatamente que as variáveis aleatórias gaussianas são *estáveis*, no sentido em que a soma de gaussianas independentes tem o mesmo tipo, isto é, a menos de localização e escala, parcelas e soma têm a mesma distribuição.

É também imediato que a função geradora de momentos $M_X(t) = \mathbb{E}(\exp(tX))$ de uma variável $X \sim \text{Gaussiana}(\mu, \sigma)$ é $M_X(t) = \exp\left(\mu t + \frac{(\sigma t)^2}{2}\right)$, e conseqüentemente a função geradora de cumulantes $K_X(t) = \ln(M_X(t)) = \mu t + \frac{(\sigma t)^2}{2}$.

Quer isto dizer, no que se refere aos cumulantes de $X \sim \text{Gaussiana}(\mu, \sigma)$, que $K_1 = \mu$, $K_2 = \sigma^2$, e $K_n = 0$, $n \geq 3$. Os cumulantes — coeficientes da expansão em série de MacLaurin do logaritmo da função geradora de momentos — surgiram no fim do século XIX na obra de Thiele (1903), e foram redescobertos por Fisher (1934), que apreciou o seu enorme potencial no contexto da família exponencial; estão naturalmente relacionados polinomialmente com os momentos e com os momentos centrados, e embora menos intuitivos que estes são, em muitos problemas, matematicamente mais simples de tratar.

Usamos também os cumulantes padronizados $\rho_n = \frac{K_n}{\sigma^n}$. Note-se que, com a usual notação μ_n para o momento centrado de ordem n , $\rho_3 = \frac{\mu_3}{\sigma^3} = \gamma_1$, o coeficiente de assimetria, e que $\rho_4 = \frac{\mu_4}{\sigma^4} - 3 = \gamma_2$, o coeficiente de achatamento (curtose) que nos fornece uma

forma simples de comparar com a gaussiana a probabilidade acumulada nas caudas de distribuições com cumulante de ordem 4 (na gaussiana, $\gamma_2 = 0$; $\gamma_2 > 0$ denuncia caudas suportando probabilidade mais elevada dos que as caudas da gaussiana).

O *TLC*, por si mesmo, é uma razão de peso para a importância da família gaussiana em Teoria da Probabilidade. Mas outras razões existem para o protagonismo desta família:

Independência I — Na multigaussiana, se o coeficiente de correlação $\rho(X_i, X_j)$ for nulo então X_i e X_j são independentes. A equivalência entre independência e correlação nula é uma propriedade característica de populações gaussianas.

Independência II — A independência I permite estabelecer facilmente que se X_1, \dots, X_n for uma amostra aleatória de uma população $X \sim \text{Gaussiana}(\mu, \sigma)$, então $\bar{X}_n = \frac{1}{n} \sum_{k=0}^n X_k$ e $S_n^2 = \frac{1}{n-1} \sum_{k=0}^n (X_k - \bar{X}_n)^2$ são independentes. Também a independência de \bar{X}_n e S_n^2 é uma caracterização da população gaussiana.

Distribuição Amostral de Momentos Empíricos — Do que estabelecemos sobre somas de gaussianas independentes, é imediato que $\bar{X}_n \sim \text{Gaussiana}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. A independência entre \bar{X}_n e S_n^2 permite estabelecer, recorrendo à forma analítica da função característica das gamas, que $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{(n-1)}^2$. Definindo a variável $t_{(n)}$ como o quociente $\frac{X}{\sqrt{Y}}$, onde $X \sim \text{Gaussiana}(0, 1)$ e $Y \sim \chi_{(n)}^2$ são variáveis independentes, demonstra-se então que $\frac{\bar{X}_n - \mu}{S \sqrt{\frac{1}{n}}} \sim t_{(n-1)}$.

No caso de duas amostras de $X \sim \text{Gaussiana}(\mu_X, \sigma)$ e de $Y \sim \text{Gaussiana}(\mu_Y, \sigma)$, respectivamente, independentes, a extensão deste último resultado é imediato: $\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}$, onde $S_p^2 = \frac{(n_1-1)S_X^2 + (n_2-1)S_Y^2}{n_1+n_2-2}$ é a variância amostral ponderada. E no caso geral de duas amostras de $X \sim \text{Gaussiana}(\mu_X, \sigma_X)$ e de $Y \sim \text{Gaussiana}(\mu_Y, \sigma_Y)$, respectivamente, independentes,

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}} \approx t_\nu,$$

onde ν pode ser estimado por

$$\tilde{\nu} = \frac{\left(\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}\right)^2}{\frac{S_X^4}{n_1^2(n_1-1)} + \frac{S_Y^4}{n_2^2(n_2-1)}}.$$

Assumindo homocedasticidade mais simples — nomeadamente por não haver tabelas facilmente acessíveis para a distribuição t de Student com número de graus de liberdade fraccionário⁽⁴⁾ — é fácil testar a hipótese nula de as amostras provirem de distribuições com iguais variâncias, pois $\frac{S_X^2}{\sigma_X^2} / \frac{S_Y^2}{\sigma_Y^2} \sim F_{(n_1-1, n_2-1)}$.

A comparação dos valores médios de k populações gaussianas homocedásticas, com base em amostras independentes traz um progresso notável: enquanto na studentização se elimina o parâmetro perturbador desconhecido de escala por forma a dispor de uma variável fulcral (estimação intervalar) ou, sob validade de H_0 , de uma estatística de teste (testes de hipóteses), procede-se no caso geral à análise da variância (ANOVA), pela influência que uma “falsa posição” tem na avaliação dos momentos de segunda ordem.

De facto, é fácil provar que a variância é o menor dos momentos de segunda ordem. Decompondo então — limitamo-nos ao caso de ANOVA simples, por facilidade de exposição — a soma de quadrados que reflecte a variabilidade total (TSS) em soma dos quadrados que avalia a variabilidade entre amostras (BSS) e soma de quadrados totaliza as que avaliam a variabilidade dentro de cada amostra (WSS), é uma simples extensão dos resultados anteriores que BSS e WSS são transformações de escala de variáveis com distribuição de qui-quadrado, e independentes, pelo que o quociente entre elas divididas pelos respectivos graus de liberdade, $k - 1$ e $N - k$ respectivamente, tem distribuição F .

É ainda fácil de estabelecer que enquanto $\frac{WSS}{N-k}$ é um estimador centrado de σ^2 , sempre, $\frac{BSS}{k-1}$ apenas é centrado sob a hipótese nula de os k valores médios serem iguais, tendendo a sobrestimar a variância populacional se essa hipótese nula for falsa.

Extensões generalizando os resultados aproximados de Welch-Satterthwaite para a situação mais vasta de comparação de k gaussianas, não assumindo homocedasticidade, com base em amostras independentes, encontram-se sumariamente descritas em Oehlert (2001).

Como se vê, a inferência em populações gaussianas tem regras bem conhecidas. A análise das demonstrações — que se pode consultar em Pestana e Velosa (2001) — mostra que assentam na independência entre \bar{X}_n e S_n^2 , que como afirmámos é característica das populações gaussianas (também demonstrada em Pestana e Velosa, 2001).

Nas outras populações a dedução da distribuição amostral de variáveis fulcrais e de

⁽⁴⁾ Uma dificuldade que é, actualmente, facilmente ultrapassada: por exemplo, o *package* R, por defeito, implementa este *teste de Welch-Satterthwaite*, sendo necessário forçá-lo, com a explicitação de que é verdade que as variâncias das duas amostras são iguais, a realizar o teste exacto clássico para essa situação.

estatísticas de teste é em geral tão complexa que ou se argumenta com o *TLC* para usar a teoria da gaussiana (o que foi frontalmente questionado por Hotelling, 1961, com argumentos irrefutáveis) ou se prefere⁽⁵⁾ uma abordagem não paramétrica.

Auto-reciprocidade — A menos de uma constante multiplicativa, a função densidade de probabilidade e a função característica da gaussiana padrão têm a mesma expressão analítica. Pode parecer uma curiosidade, mas Pollard (2001) comenta explicitamente este facto e a sua relevância para o desenvolvimento da teoria das transformadas de Fourier. A demonstração da inversão de funções características absolutamente integráveis baseada na relação de Parseval é, provavelmente, a forma mais elementar de abordar a questão.

Regressão Linear — Seja $(X, Y) \sim \text{Gaussiana}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; álgebra elementar chega para mostrar que $f_{(X,Y)}(x, y) = f_X(x) \times f_{Y|X=x}(y)$, onde $X \sim \text{Gaussiana}(\mu_X, \sigma_X)$ e $Y|_{X=x} \sim \text{Gaussiana}\left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y \sqrt{1 - \rho^2}\right)$. (As notações que usamos são standard, pelo que nos dispensamos de comentar a parametrização adoptada.)

Assim, a regressão $\mathbb{E}(Y|_{X=x}) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ é linear. A expressão paraleliza obviamente a da recta dos mínimos quadrados quando se ajusta um modelo linear, com aquele critério, aos pontos $(x, \bar{y}|_x)$ em que a notação que adoptamos pretende sugerir que num verdadeiro problema de regressão (por contraste com um problema de correlação), as ordenadas são as médias da variável resposta nos diversos níveis considerados da variável controlada. Estes resultados são naturalmente extensíveis a vectores gaussianos em \mathbb{R}^n .

A linearidade da regressão não caracteriza a família das gaussianas, mas é evidentemente uma razão mais para suspirar por afinal o Mundo não ser normal!

Note-se que o *TLC*, no sentido restrito de convergência de somas de variáveis aleatórias apropriadamente centradas e com variância estabilizada para uma variável aleatória gaussiana, pode ser estabelecido em condições que relaxam de forma não trivial a hipótese de as parcelas serem independentes e identicamente distribuídas. De facto, as aplicações práticas levam, em muitas circunstâncias, a invocar o *TLC* quando as parcelas não são identicamente distribuídas; o que é válido, impondo que a sucessão de variâncias truncadas seja uniformemente limitada (Lindeberg–Feller), como já referimos. Anote-se que a demonstração pode ser feita, alternativamente, usando operadores contraentes, como alternativa à demonstração apresentada em Pestana e Velosa (2002); Rohatgi (1976), baseando-se numa ideia de Trotter (1959), que simplifica singularmente a apresentação misteriosa de Feller (1967); este método de demonstração presta-se à demonstração do *TLC* impondo condições de dependência fraca, nomeadamente com a estrutura de dependência estabelecida usando momentos, das martingalas.

⁽⁵⁾ Ou preferia, antes de as técnicas de computação intensiva abrirem novas perspectivas em Estatística.

Outra forma de dependência fraca de grande importância é a permutabilidade, essencial no tratamento de aproximações usando o *TLC* em populações finitas, imprescindível em questões incontornáveis em Amostragem.

1.2 Um Mundo Estável?

A convergência de somas para uma gaussiana dá-se quando nenhuma das parcelas tem um papel preponderante. Todavia, a situação altera-se drasticamente quando as estatísticas ordinais extremas têm algum protagonismo, o que acontece se o modelo distribucional de base tiver caudas pesadas. Por exemplo, a variável aleatória de Cauchy tem caudas muito pesadas — não existe valor médio — e a sucessão de somas padronizadas de variáveis de Cauchy é uma variável de Cauchy.

Surge assim o interesse em investigar modelos aditivos alternativos ao modelo gaussiano, em situações de distribuição de base com caudas pesadas. Mais concretamente, caracterizar as possíveis leis limites de somas padronizadas $\frac{S_n - b_n}{a_n}$ onde $S_n = \sum_{k=1}^n X_k$, com parcelas X_k *iid*, e as constantes normalizadoras ou constantes de atracção $a_n > 0$ e $b_n \in \mathbb{R}$. O papel das constantes normalizadoras é localizar⁽⁶⁾ a soma num valor finito, geralmente 0, e não deixar a dispersão aumentar indefinidamente, em geral forçando a escala unitária.

No caso de existirem constantes normalizadoras tais que a sucessão de somas centradas e reduzidas de réplicas *iid* $X_k \stackrel{d}{=} X$, $\frac{S_n - b_n}{a_n}$ converge para uma variável não degenerada Y , dizemos que X está no domínio de atracção de Y , e escrevemos $S \in \mathcal{D}(Y)$.

O *TLC* clássico estabelece portanto que $\text{var}(X) < \infty$ é uma condição suficiente para X estar no domínio de atracção de $Z \sim \text{Gaussiana}(0, 1)$.

Se $\{a_n\}_{n \geq 1}$, com $a_n > 0$ e $\{b_n\}_{n \geq 1}$, $b_n \in \mathbb{R}$, forem constantes de padronização, isto é se a sucessão de somas $\frac{S_n - b_n}{a_n}$ convergir em distribuição para uma variável aleatória não degenerada Y , também $\{a_n^*\}_{n \geq 1}$ e $\{b_n^*\}_{n \geq 1}$ tais que $\frac{a_n^*}{a_n} \xrightarrow{n \rightarrow \infty} 1$ e $\frac{b_n^* - b_n}{a_n} \xrightarrow{n \rightarrow \infty} 0$ são constantes normalizadoras.

De facto, estamos a tratar de convergência de classes, ou tipos de Khintchine⁽⁷⁾. Se, nas condições acima, $\frac{S_n - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} Y$ não degenerada, é sempre possível encontrar sucessões apropriadas de constantes normalizadoras tais que

⁽⁶⁾ Propositadamente usamos um termo “vago”, uma vez que 0 pode ser a extremidade do suporte, a média, ou a mediana — depende de o que permite uma parametrização mais simples. Adiante, escala aparece também como conceito vago.

⁽⁷⁾ Duas variáveis aleatórias X e Y são do mesmo tipo de Khintchine se e só se $Y = aX + b$, com $a > 0$ e $b \in \mathbb{R}$, isto é se apenas diferem em localização e escala, mantendo a forma.

- $\frac{S_n - B_n}{A_n} \xrightarrow[n \rightarrow \infty]{d} AY + B$
- $\frac{\tilde{S}_n - \tilde{B}_n}{\tilde{A}_n} \xrightarrow[n \rightarrow \infty]{d} Y$ com $\tilde{S}_n = \sum_{k=1}^n (\alpha X_k + \beta)$.

Neste sentido, não há perda de generalidade em escolher constantes de atracção que localizem a variável em 0 e lhe confirmem escala unitária.

Temos, assim, que precisar a frase acima em que dissemos que todas as variáveis com segundo momento finito estão no domínio de atracção de $Z \sim \text{Gaussiana}(0, 1)$: elas estão, de facto, no domínio de atracção da gaussiana, pois a escolha apropriada de constantes permite obter como lei limite qualquer variável do tipo gaussiano.

As possíveis variáveis aleatórias limites Y acima referidas têm de verificar uma equação de estabilidade, isto é, denotando $S_n = \sum_{k=1}^n Y_k$, têm que existir constantes normalizadoras $\alpha_n > 0$ e $\beta \in \mathbb{R}$ tais que, para $n = 1, 2, \dots$,

$$\frac{S_n - \beta_n}{\alpha_n} \stackrel{d}{=} Y,$$

onde os Y_k são réplicas independentes de Y . Por isso são chamadas leis estáveis para somas.

A equação de estabilidade — que se obtém “partindo aos bocados” uma sucessão $X_1, \dots, X_r, X_{r+1}, \dots, X_{2r}, \dots, X_{(\nu-1)r+1}, \dots, X_{\nu r}$, e que permite estabelecer Feller (1971), que as constantes normalizadoras são da forma $n^{\frac{1}{\alpha}}$, com $\alpha \in (0, 2]$ denominado expoente característico da lei estável Y — que por isso passaremos a explicitar como Y_α , — pode também exprimir-se em termos de funções características

$$\varphi_{\frac{S_n - \beta_n}{\alpha_n}}(t) = e^{-i \frac{\beta_n}{\alpha_n} t} \varphi_Y^n \left(\frac{t}{\alpha_n} \right) = \varphi_Y(t).$$

No que se refere às variáveis aleatórias estáveis Y_α , $\alpha_n = n^{\frac{1}{\alpha}}$, com $\alpha \in (0, 2]$. No que se refere às variáveis aleatórias X no domínio de atracção de Y_α , prova-se que $\alpha_n = L(n) n^{\frac{1}{\alpha}}$, onde L é uma função de variação lenta no sentido de Karamata, isto é

$$\frac{L(tx)}{L(x)} \xrightarrow{x \rightarrow \infty} 1$$

e $\alpha \in (0, 2]$ é o índice ou expoente característico da variável aleatória limite Y .

As variáveis aleatórias estáveis, ou seja os possíveis limites não degenerados Y de somas normadas de variáveis aleatórias

$$\frac{S_n - \beta_n}{\alpha_n} \xrightarrow[n \rightarrow \infty]{d} Y$$

surtem assim como modelos naturais no caso de estarmos a abordar fenômenos regulares em que exista um mecanismo aditivo.

Como veremos na seção seguinte, as variáveis aleatórias estáveis são infinitamente divisíveis. E portanto, a função característica de uma variável aleatória estável pode ser escrita usando a representação canônica de Lévy:

$$\varphi_X(t) = \exp \left\{ iat - c|t|^\alpha \left[1 + i\beta \frac{t}{|t|} w(t, \alpha) \right] \right\},$$

com localização $a \in \mathbb{R}$, escala $c > 0$, índice ou expoente característico $\alpha \in (0, 2]$, e coeficiente de assimetria $\beta \in [-1, 1]$, e

$$w(t, \alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \text{se } \alpha \neq 1 \\ \frac{2}{\pi} \log|t| & \text{se } \alpha = 1 \end{cases}.$$

Seríamos assim levados a pensar que os modelos estáveis assim descritos seriam ideais para abordar o estudo de fenômenos em que há subjacente algum mecanismo aditivo. De facto, isso permitiria que, estimado o valor do expoente característico α — que avalia o peso das caudas do modelo — e do coeficiente de assimetria β , tudo se reduzisse apenas a uma questão de localização/escala. Por outras palavras, mais informação poderia ser facilmente incorporada, mudando apenas a localização e a escala.

Infelizmente o uso destes modelos estáveis, na prática, está bastante limitado por duas razões:

- O único modelo estável para somas com variância finita é o modelo gaussiano.
- Não se consegue inverter as funções características estáveis a não ser em três casos especiais, e por isso as únicas funções densidade de probabilidades estáveis conhecidas correspondem a $\alpha = 2$ (Gaussiana), a $\alpha = 1$ e $\beta = 0$ (Cauchy), e a $\alpha = \frac{1}{2}$ e $\beta = 1$ (Lévy).

Notemos que as duas últimas distribuições não têm valor médio finito.

Concluimos, assim, que afinal o mundo não pode ser apenas estável!

1.3 Um Mundo de Infinitamente Divisíveis e de Irredutíveis

Em 1917 Paul Lévy foi convidado a fazer um ciclo de três conferências sobre o Teorema Limite central (ele trabalhava em Mecânica Racional, nessa época). Ficou insatisfeito com as restrições do TLC clássico: variáveis independentes, variáveis identicamente distribuídas e variância finita, e resolveu investigar o que acontecia se relaxasse as hipóteses.

Entre 1918 e 1924 debruçou-se sobre o relaxamento da hipótese sobre os momentos estudando a “equação de estabilidade”

$$\frac{S_n}{A_n} \stackrel{d}{=} X_k,$$

estabilidade em sentido estrito, pois só estabilizou a escala, não se preocupou com a localização. Provou em particular que $A_n \sim n^{1/\alpha}$, $\alpha \in (0, 2]$ com α o expoente característico.

A sùmula do seu trabalho é publicado em 1925 em “*Calcul des Probabilités*”.

Em 1924 Lévy tinha-se interessado pelas “somas infinitésimas”, somas de parcelas infinitamente pequenas associadas a processos estocásticos com incrementos independentes. Introduz a noção de divisibilidade infinita

$$X \text{ é infinitamente divisível} \Leftrightarrow X \stackrel{d}{=} Y_1 + Y_2 + \dots + Y_n, \forall n = 1, 2, \dots$$

com os Y_k independentes e identicamente distribuídos. Em termos da função característica

$$\varphi_X = \varphi_{Y_n}^n \Rightarrow \varphi_X^{1/n} = \varphi_{Y_n},$$

ou seja, X é infinitamente divisível sse $\varphi_X^{1/n}$ é função característica para qualquer $n = 1, 2, \dots$, disto decorre que φ^α é função característica para todo $\alpha > 0$.

Em 1930 de Finetti mostra que se conclui imediatamente dos trabalhos de Lévy que:

- φ não pode ter zero reais, e por isso faz sentido usar $\psi(t) = \ln(\varphi(t))$
- φ é infinitamente divisível se e só se for limite de Poisson compostas.

Uma variável aleatória infinitamente divisível pode ser construída de duas formas: ou como uma soma infinita de termos evanescentes (esquema de “triangular array”), ou a partir de somas estocásticas governadas por uma Poisson, neste caso o número de parcelas está com grande probabilidade entre $\lambda - 3\sqrt{\lambda}$ e $\lambda + 3\sqrt{\lambda}$.

Em 1930 Kolmogoroff dá uma “representação integral” de ψ no caso de $\mathbb{E}(X^2)$ existir.

Entre 1932 e 1935 Lévy desenvolve uma representação geral

$$\varphi(t) = \exp \left\{ i a t + c |t|^2 + \int_0^\infty \left(e^{iut} - 1 - \frac{iut}{1+u^2} \right) dM(u) + \int_0^\infty \left(e^{iut} - 1 - \frac{iut}{1+u^2} \right) dN(u) \right\}$$

onde $i a t$ corresponde à componente degenerada, $c |t|^2$ corresponde à componente gaussiana (se $c \neq 0$) e

$$\int_0^\infty \left(e^{iut} - 1 - \frac{iut}{1+u^2} \right) dM(u) + \int_0^\infty \left(e^{iut} - 1 - \frac{iut}{1+u^2} \right) dN(u),$$

correspondente à Poisson generalizada.

Em 1937 Khinchine apresenta uma representação alternativa

$$\varphi(t) = \exp \left\{ i a t + \int_{-\infty}^{+\infty} \left(e^{i t x} - 1 - \frac{i t x}{1 + x^2} \right) \frac{1 + x^2}{x^2} d\theta(x) \right\}.$$

No entanto, a construção das representações anteriores, pode ser vista como uma elaboração de trivialidades. De facto, em 1972 Burril mostra que as representações integrais não são mais do que elaborações do resultado

$$\prod_{k=1}^{\infty} a_k \sim e^{-\sum_{k=1}^{\infty} (1 - a_k)}.$$

Notemos que, $\prod_{k=1}^{\infty} a_k$ converge se e só se $\sum_{k=1}^{\infty} (1 - a_k)$ converge. Mas $\sum_{k=1}^{\infty} (1 - a_k) \sim \ln[1 - \prod_{k=1}^{\infty} a_k]$, pelo que

$$\ln \left(\prod_{k=1}^{\infty} a_k \right) \sim -\sum_{k=1}^{\infty} (1 - a_k).$$

Tomando exponenciais temos que

$$\prod_{k=1}^{\infty} a_k \sim e^{-\sum_{k=1}^{\infty} (1 - a_k)}.$$

Kendall em 1963, para variáveis positivas, e Johansen em 1966 — que conseguiu provar que o logaritmo de uma função característica infinitamente divisível é uma função definida positiva — mostram que as distribuições Gaussiana e Poisson são os pontos extremos da família convexa das infinitamente divisíveis, e que a representação de Lévy é uma representação integral de Choquet em termos de uma medida de probabilidade suportada pelo conjunto compacto dos pontos extremos. Kendall e Johansen dão assim, em conjunto, uma visão profunda das distribuições infinitamente divisíveis.

Seja

$$S_n \stackrel{d}{=} \frac{X_1 + X_2 + \cdots + X_n - B_n}{A_n} = \frac{X_1}{A_n} + \cdots + \frac{X_n}{A_n} - \frac{B_n}{A_n}.$$

Tem-se

$$\begin{aligned}\varphi_{S_n}(t) &= e^{-it\frac{B_n}{A_n}} \varphi^n\left(\frac{t}{A_n}\right) \\ \Rightarrow \varphi_{S_n}^{1/n} &= e^{-it\frac{B_n}{nA_n}} \varphi\left(\frac{t}{A_n}\right).\end{aligned}$$

Ou seja o conjunto das leis estáveis está contido no conjunto das leis infinitamente divisíveis, e portanto a sua função característica admite a representação de Lévy. Em 1937 Lévy reexpressa as leis estáveis no contexto (como corolário) das leis infinitamente divisíveis (*Théorie de L'Addition des Variables Aléatoires*).

Variáveis aleatórias como as de Poisson, as gamas e as gaussianas podem ser decompostas na soma de n parcelas *iid*, para todo o $n = 1, 2, \dots$ e são por esse facto chamadas de infinitamente divisíveis, e têm uma papel relevante na teoria da soma de variáveis aleatórias. Por outro lado, é impossível escrever a uniforme como soma de duas variáveis aleatórias *iid*, pelo que é chamada de irredutível.

Sabemos que o produto de duas funções características é sempre uma função característica. É assim óbvio que algumas funções características podem ser escritas como o produto de duas ou mais funções características.

Toda a função característica φ pode ser escrita como o produto de duas funções características $\varphi_1(t) = e^{imt}$ e $\varphi_2(t) = \varphi(t)e^{-imt}$, com m real. Dizemos que a representação de uma função característica como o produto de duas funções características é trivial se um dos factores tem a forma $\varphi_1(t) = e^{imt}$. Para evitar a representação trivial de produtos, vamos introduzir a seguinte definição.

A função característica φ é dita decomponível se puder ser escrita da seguinte forma:

$$\varphi = \varphi_1\varphi_2$$

onde φ_1 e φ_2 são ambas funções características não degeneradas. Neste caso dizemos que φ_1 e φ_2 são os factores de φ .

A função característica que apenas admite representação em produtos triviais é dita irredutível.

A factorização de uma função característica em factores irredutíveis é similar à factorização de números reais em factores primos. Esta é a razão porque a teoria da decomposição das funções características é muitas vezes chamada teoria aritmética das funções de distribuição. No entanto, esta analogia não vai muito longe; por exemplo a factorização de funções características em factores irredutíveis não é sempre única.

Khintchine demonstrou que qualquer variável aleatória pode ser decomposta na soma de variáveis aleatórias irredutíveis e infinitamente divisíveis sem factores irredutíveis, que é o máximo que se consegue estabelecer para este tipo de variáveis aleatórias.

Lukacs (1960) provou que a existência de distribuições infinitamente divisíveis com factores irreduzíveis não era uma ocorrência rara. Assim, provou que uma função característica infinitamente divisível φ , cuja representação canónica de Lévy fosse determinada pelas constantes $a = c = 0$ e pelas funções

$$N(u) = \begin{cases} k(u - a) & \text{se } 0 < u < a \\ 0 & \text{o.c.} \end{cases} \quad \text{e } M(u) = 0$$

onde $k > 0$ e $a > 0$ são constantes reais arbitrárias e tem sempre um factor irreduzível.

Além disso, se ϕ for uma função característica infinitamente divisível, e se as funções M e N , que ocorrem na representação canónica de Lévy, satisfazem a seguinte condição: Existem duas constantes positivas k e c tal que pelo menos uma das seguintes relações $M'(u) > k$ quase certamente em $(-c, 0)$ ou, $N'(u) > k$ quase certamente em $(0, c)$ se verifica, então ϕ tem um factor irreduzível.

Se φ é uma função característica infinitamente divisível, que satisfaz a condição anterior então é divisível pelo produto de uma sequência infinita de funções características irreduzíveis.

Notemos no entanto que, não existe um método geral para encontrar os factores irreduzíveis para uma dada função característica.

1.4 Quando as Somas são Aleatórias

A divisibilidade infinita advém, em termos de teoremas limites originando a teoria das representações integrais, de somarmos parcelas assintoticamente negligíveis. De facto, tal — ou, o que é equivalente, face à continuidade das funções características, ou o facto de se estar a considerar $\prod_{k=1}^{\infty} \varphi_k(t)$ com $\varphi_k(t) = 1 + o(t)$ — é uma condição necessária para assegurar convergência. Note-se que, em termos de produtos de funções características, e pelo que se sabe sobre critérios de convergência, é o mesmo que considerar $e^{-\sum_{k=0}^{\infty} (\varphi_k(t)-1)}$.

No entanto, o grande trabalho seminal de Bruno de Finetti (1930) mostra que qualquer variável aleatória infinitamente divisível é Poisson composta ou soma de Poissons compostas. Esta abordagem alternativa capitaliza no facto de haver uma soma de parcelas, em número que pode, com baixíssima probabilidade, exceder qualquer natural fixo, mas que, pela desigualdade de Chebycheff, com grande probabilidade está entre $\lambda - k\sqrt{\lambda}$ e $\lambda + k\sqrt{\lambda}$ (a probabilidade de exceder $\lambda + k\sqrt{\lambda}$ parcelas é limitada por $\frac{1}{k^2}$).

Parte do desenvolvimento da Probabilidade deriva de se aleatorizar qualquer coisa que antes era encarado como determinista. As somas de variáveis aleatórias seguem este padrão. As aplicações em teoria de seguros fazem considerar um processo de risco

$$Y_t = \sum_{k=0}^{N_t} X_k$$

onde X_k são os modelos para as indenizações e N_t é um processo pontual de contagem. Os modelos iniciais foram, naturalmente, os mais simples, aqueles em que as parcelas X_k são independentes e identicamente distribuídas — existe um modelo suficientemente fluído para ajustar todas as possíveis indenizações — e independentes da variável subordinadora N_t (a qual, por simplicidade, e também por realismo, foi originalmente considerada Poisson). Existe hoje uma rica literatura sobre “randomly stopped sums”, quer do ponto de vista teórico, quer num espírito aplicado à teoria do risco e muitas outras áreas da Probabilidade em que os modelos aditivos são o ponto de partida.

No que se refere a teoremas limite centrais, o mais interessante a referir é que se $N_t \sim \text{Geométrica}(p(t))$ e recordamos que o modelo geométrico pode ser construído por aleatorização exponencial do parâmetro da Poisson, pelo que há quem a refira como uma “Poisson mais dispersa” se constrói uma teoria da estabilidade semelhante à de Lévy (Velosa e Pestana (2003)), os trabalhos iniciais devem-se a Kovalenko (1956), e deram origem ao conceito de variáveis $N - \text{Gaussianas}$ (no caso de $N_t \sim \text{Geométrica}(p)$ a $N - \text{Gaussiana}$ é a lei de Laplace).

Assim, os problemas que abordamos podem ser contextualizados neste padrão mais vasto de aproximações e velocidades de convergência que as justificam, por leis gaussianas.

1.5 Uma Perspectiva dos Resultados Obtidos

O modelo gaussiano tem um protagonismo em Estatística que decorre de propriedades — algumas das quais características desse modelo, que portanto melhor seria apodado de “anormal” do que de normal — que permitem um tratamento matemático simples e elegante. Há uma tradição de o usar como aproximação, justificada pelo *TLC*. No entanto, a procura de demonstrações cada vez mais límpidas e gerais do *TLC*, exigindo cada vez menos condições sobre existência de momentos, levou a uma formulação em que a velocidade de convergência é apenas $O\left(\frac{1}{\sqrt{n}}\right)$.

Porém, as modernas abordagens ao *TLC* usando teoria da informação (capítulo 3) levam a pensar que seria possível (e mesmo natural) obter velocidades de convergência $O\left(\frac{1}{n}\right)$. Esta forma de equacionar o problema leva-nos naturalmente a estender a desigualdade de Cramér-Rao, e a caracterização das situações em que o limite inferior de Cramér-Rao é atingido é uma forma intuitiva de aceder à família exponencial de Fisher-Darmois-Koopman-Pitman.

Assim, os desenvolvimentos da teoria da informação que surgem como justificação adicional à *posteriori* do que fazemos no capítulo 2, invocam a necessidade de contextualizar uma abordagem moderna na teoria das famílias exponenciais, e de procurar caracterizar as circunstâncias em que a velocidade de convergência é melhorada para $O\left(\frac{1}{n}\right)$.

Trabalhos de Cramér (1963) e de Daniels (1956), o primeiro usando a ideia de distribuições conjugadas por *tilting* (uma ideia que atribui a Esscher) e o segundo usando

aproximações recorrendo a pontos de sela, têm um contributo notável para a teoria das aproximações, nomeadamente na denominada zona de grandes desvios. No capítulo 1 recorreremos às expansões de Edgeworth diferidas (isto é usando *tilting*) para fazer a ponte de união entre os resultados de Cramér e de Daniels.

Neste contexto, surgem naturalmente as famílias naturais de Morris (1982, 1983). Estas interessaram-nos por os seus membros absolutamente contínuos terem versões simetrizadas cujas funções densidade de probabilidade são definidas positivas, uma questão que foi abordada no trabalho científico de Malva e Sequeira (2003), e por seus membros discretos terem funções massa de probabilidade que verificam a expressão de recorrência de Panjer. No entanto, não conseguimos obter as relações estruturais que esperavamos. Anote-se, no entanto, que existem relações inesperadas entre todos estes temas e a extensão não trivial que é usar somas aleatórias: Malva e Sequeira (2003) estabeleceram que a famosa fórmula de Khinchine-Pollaczek para somas de variáveis aleatórias que ocorrem em teoria de filas de espera e em teoria de risco pode ser reobtida como subproduto da teoria dos pares conjugados de variáveis aleatórias.

No capítulo 2, completamos a exposição das aproximações dando, além das expressões de Edgeworth e Edgeworth diferidas, as expansões de Fisher-Cornish. Neste contexto, os cumulantes padronizados, embora não tenham a interpretação intuitiva que os momentos têm, trazem grande simplicidade à exposição. Os cumulantes padronizados de ordem três e quatro são os coeficientes de assimetria e de curtose que tanto relevo têm na organização das leis de Pearson, e conclui-se assim que quase-simetria e peso de caudas moderado são instrumentais numa boa velocidade de convergência no *TLC*, sendo possível usar *tilting* e *back-tilting* para conseguir estender os bons resultados nas zonas de grandes desvios.

Curiosamente, um exemplo usando a estável de Lévy mostra que as aproximações, apesar de concebidas para variáveis no domínio de atracção da gaussiana, podem ter um validade em casos inesperados⁽⁸⁾, tal leva-nos a tentar — o que fazemos no capítulo 4, dando os detalhes fastidiosos nos apêndices — a estudar velocidades de convergência quando a lei limite é estável não gaussiana, inspirando-nos de trabalhos em teoria dos valores extremos em que parâmetros de segunda ordem podem ter influência decisiva. Os resultados parcelares que obtivemos não constituem, porém, um progresso relevante em relação a resultados de Cramér (1962, 1963). Sendo consolador tê-los reobtido antes de conhecer os trabalhos de Cramér, a quem naturalmente atribuímos a prioridade, é no entanto desencorajante observar que a utilização de ferramentas analíticas maciças produz tão poucos resultados. Mais uma vez, parece chegar-se à (i)moralidade: “use a aproximação gaussiana ou vá antes dar uma curva” (não sabemos é qual, e por isso o despite está sempre eminente).

⁽⁸⁾Embora possa haver alguma excepcionalidade no exemplo, e a explicação provir da relação entre as estáveis com parâmetros característicos α e $\frac{1}{\alpha}$ estabelecida por Zolotarev (1986). Mas não conseguimos estabelecer se esta intuição é fundamentada.

Capítulo 2

Polinómios Ortogonais e Expansões Assintóticas

2.1 Transformadas Integrais. Momentos e Cumulantes

A função característica, assim como outras funções geradoras, foram originalmente desenvolvidas como uma ferramenta para encontrar soluções para problemas em teoria da probabilidade. Esta teoria tem muitas aplicações não só no ramo das probabilidades, mas também na estatística matemática.

Seja X uma variável aleatória (*va*) com função de distribuição (*fd*) F_X . A função geradora de momentos (*fgm*) define-se como sendo

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_{\mathbb{R}} e^{tx} dF_X(x).$$

O integral anterior é absolutamente convergente numa faixa vertical do plano complexo $\{t : t_1 < \Re(t) < t_2\}$ com $t_1, t_2 \geq 0$, e pode eventualmente degenerar no eixo imaginário. A função definida anteriormente permite, sob condições muito gerais, expressar os momentos da variável aleatória, caso existam, em termos das derivadas de M_X em zero,

$$\left. \frac{d^n M_X(t)}{dt^n} \right|_{t=0} = \int_{\mathbb{R}} x^n dF_X(x) = \mathbb{E}(X^n).$$

Por exemplo, se $Z \sim \text{Gaussiana}(0, 1)$, a sua função geradora de momentos é:

$$M_Z(t) = e^{\frac{t^2}{2}},$$

tendo-se

$$\frac{d^n M_X(0)}{dt^n} = \int_{\mathbb{R}} x^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \mathbb{E}(Z^n).$$

Uma outra transformada integral que vai ser utilizada ao longo deste trabalho é a função característica (fc), definida por

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx} dF_X(x).$$

Retomando, ainda, o exemplo, $Z \sim \text{Gaussiana}(0, 1)$, obtém-se para Z a seguinte função característica

$$\varphi_Z(t) = e^{-\frac{t^2}{2}}.$$

A função característica tem a vantagem de existir para todo o $t \in \mathbb{R}$, pois

$$|\varphi_X(t)| = \left| \int_{\mathbb{R}} e^{itx} dF_X(x) \right| \leq \int_{\mathbb{R}} |e^{itx}| dF_X(x) = \int_{\mathbb{R}} dF_X(x) = 1.$$

A função característica tem diversas propriedades úteis de entre as quais se destaca:

- Se X for simétrica, em torno de 0, então a função característica φ_X é real e par, pois

$$\begin{aligned} \varphi_X(t) &= \int_{\mathbb{R}} e^{itx} dF_X(x) = \int_{\mathbb{R}} \cos(tx) dF_X(x) + i \int_{\mathbb{R}} \sin(tx) dF_X(x) \\ &= \int_{\mathbb{R}} \cos(tx) dF_X(x) = \int_{\mathbb{R}} \cos(-tx) dF_X(x) = \varphi_X(-t) \end{aligned}$$

visto que

$$\int_{\mathbb{R}} |\sin(tx)| dF_X(x) \leq \int_{\mathbb{R}} dF_X(x) = 1,$$

e a função integranda em $\int_{\mathbb{R}} \sin(tx) dF_X(x)$ é ímpar, pelo que $\int_{\mathbb{R}} \sin(tx) dF_X(x) = 0$.

Prova-se que a recíproca também é verdadeira.

- Qualquer função característica é uniformemente contínua em \mathbb{R} .
- É óbvio que a função geradora de momentos calculada no eixo imaginário, e a função geradora de probabilidades calculada na circunferência que delimita o círculo unitário, é a função característica.
- Facilita o tratamento de somas de variáveis aleatórias independentes: se X e Y são *va's* independentes, também e^{itX} e e^{itY} são independentes, e então o valor médio do produto é o produto dos valores médios, tendo-se que

$$\varphi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX} e^{itY}) = \varphi_X(t) \varphi_Y(t).$$

Como consequência se $\{X_k\}_{k=1}^n$ são variáveis aleatórias independentes e identicamente distribuídas, com $X_k \stackrel{d}{=} X$, e $S_n = \sum_{k=1}^n X_k$, tem-se que $\varphi_{S_n}(t) = \varphi_X^n(t)$.

- Se existir $\mathbb{E}(X^k)$, tem-se que

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k), \quad k = 0, \dots, n.$$

- A função característica caracteriza a variável aleatória, no sentido em que

$$\varphi_X \equiv \varphi_Y \Rightarrow X \stackrel{d}{=} Y.$$

Existem, no entanto, duas propriedades da função característica que devido à sua importância e aplicabilidade que iremos salientar nos seguintes teoremas.

Teorema 2.1.1. Teorema da continuidade de Lévy-Cramér

Seja $\{X_k\}_{k \geq 1}$ uma sucessão de variáveis aleatórias, e suponha que

$$F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F_Y(x),$$

para todo o ponto x de continuidade de F_Y ; então $\varphi_{X_n}(x) \xrightarrow{n \rightarrow \infty} \varphi_Y(x)$, $x \in \mathbb{R}$. Por outro lado, se $\{\varphi_{X_n}\}_{n \geq 1}$ for uma sucessão de funções características tal que existe $\lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \psi(t)$, e esta função for contínua na vizinhança da origem, então ψ é a função característica de uma variável aleatória Y , e

$$F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F_Y(x),$$

para todo o ponto x de continuidade de F_Y .

A demonstração deste resultado encontra-se, por exemplo, em Lukacs (1970, pp. 54-55).

Exemplo

O teorema anterior permite esclarecer a possibilidade de aproximar $X \sim Binomial(n, p)$ por $Y \sim Poisson(np)$.

Seja $\{X_n\}_{n \geq 1}$ uma sucessão de variáveis aleatórias $X_n \sim Binomial(n, p_n)$ com $p_n \xrightarrow{n \rightarrow \infty} 0$ mas $np_n \xrightarrow{n \rightarrow \infty} \lambda > 0$. Então

$$X_n \xrightarrow[n \rightarrow \infty]{d} Y \sim Poisson(\lambda).$$

A função característica de X_n é $\left[1 + p_n (e^{it} - 1)\right]^n$, onde podemos escrever p_n como $\frac{\lambda + \varepsilon_n}{n}$, com $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$. Uma vez que

$$\left[1 + \frac{\lambda + \varepsilon_n}{n} (e^{it} - 1)\right]^n \xrightarrow{n \rightarrow +\infty} \exp[\lambda(e^{it} - 1)],$$

que obviamente é uma função contínua na vizinhança da origem, a qual reconhecemos ser a função característica de $Y \sim \text{Poisson}(\lambda)$. ■

Teorema 2.1.2. Teorema da inversão das funções características de L^1

Seja φ a função característica correspondente à função de distribuição F , e suponha-se que $\varphi \in L^1(-\infty, +\infty)$.

Então, F tem função densidade de probabilidade contínua e limitada, e

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi(t) dt.$$

A demonstração deste resultado pode ser consultada, por exemplo, em Pestana e Velosa (2002, pp. 799).

Exemplo

Sabemos que a distribuição Laplace é a simetrização da exponencial (cuja função característica é $\varphi_X(t) = \frac{1}{1-it}$), isto é, se $X_1 \stackrel{d}{=} X_2 \sim \text{Exponencial}(1)$ com X_1 e X_2 independentes, então $X = X_1 - X_2$ é Laplace padrão. Portanto

$$\varphi_X(t) = \varphi_{X_1}(t)\varphi_{-X_2}(t) = \frac{1}{1-it} \frac{1}{1+it} = \frac{1}{1+t^2}.$$

Usando agora o teorema da inversão, obtém-se:

$$\frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \frac{1}{1+t^2} dt = \frac{1}{2} e^{-|x|} \mathbf{I}_{\mathbb{R}}(x).$$

Mais, como $\varphi(t) = \frac{1}{1+t^2}$ é a menos de uma constante a função densidade de probabilidade da Cauchy padrão e sendo uma função par a parte imaginária do integral é nula,

$$\frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \frac{1}{1+t^2} dt = \frac{1}{2\pi} \int_{\mathbb{R}} e^{itx} \frac{1}{1+t^2} dt = \frac{1}{2} e^{-|x|} \mathbf{I}_{\mathbb{R}}(x).$$

Conclui-se assim que a função característica da Cauchy padrão, isto é, da variável aleatória Y com função densidade de probabilidade $f_Y(x) = \frac{1}{\pi} \frac{1}{1+x^2} \mathbf{I}_{\mathbb{R}}(x)$ é

$$\varphi_Y(t) = \int_{\mathbb{R}} e^{itx} \frac{1}{\pi} \frac{1}{1+x^2} dx = e^{-|t|}. \quad \blacksquare$$

À custa da função característica define-se a função geradora de cumulantes, à qual se dá algum destaque na secção seguinte.

Os momentos de uma distribuição são um conjunto de constantes descritivas da distribuição, que são úteis para quantificar algumas propriedades, e em certas circunstâncias, para especificar a própria distribuição. No entanto, não são o único conjunto de constantes com tal propósito nem o melhor. Outra série de constantes são os denominados cumulantes, que têm propriedades menos intuitivas mas muito melhores para alguns desenvolvimentos teóricos.

Para uma variável aleatória Y qualquer, com função característica φ , o j -ésimo cumulante k_j , de Y , é definido como sendo o coeficiente de $\frac{1}{j!}(it)^j$ no desenvolvimento em série de $\log(\varphi_Y(t))$

$$\begin{aligned}\log(\varphi_Y(t)) &= \log(\varphi_Y(0)) + it\mathbb{E}(Y) + \frac{(it)^2}{2!}(\mathbb{E}(Y^2) - (\mathbb{E}(Y))^2) + \\ &+ \frac{(it)^3}{3!}(\mathbb{E}(Y^3) - 3\mathbb{E}(Y^2)\mathbb{E}(Y) + 2(\mathbb{E}(Y))^3) + \dots \\ &= ti k_1 + \frac{(it)^2}{2!} k_2 + \frac{(it)^3}{3!} k_3 + \dots + \frac{(it)^j}{j!} k_j + \dots\end{aligned}$$

ou seja,

$$\varphi_Y(t) = \exp \left\{ ti k_1 + \frac{(it)^2}{2!} k_2 + \frac{(it)^3}{3!} k_3 + \dots + \frac{(it)^j}{j!} k_j + \dots \right\}.$$

Tem-se ainda que

$$\begin{aligned}\varphi_Y(t) &= \mathbb{E}(e^{itY}) = \mathbb{E} \left\{ 1 + itY + \frac{(it)^2}{2!} Y^2 + \frac{(it)^3}{3!} Y^3 + \dots + \frac{(it)^j}{j!} Y^j + \dots \right\} \\ &= 1 + \sum_{j \geq 1} \frac{(it)^j}{j!} \mathbb{E}(Y^j) \Rightarrow \log(\varphi_Y(t)) = \log \left(1 + \sum_{j \geq 1} \frac{(it)^j}{j!} \mathbb{E}(Y^j) \right).\end{aligned}$$

Resulta então que,

$$\sum_{j \geq 1} \frac{(it)^j}{j!} K_j = \log \left(1 + \sum_{j \geq 1} \frac{(it)^j}{j!} \mathbb{E}(Y^j) \right).$$

Atendendo ao facto, de para $|x| < 1$, a função $\ln(1+x)$ poder ser representada pela série de Maclaurin $\sum_{k \geq 1} (-1)^{k+1} \frac{x^k}{k}$, se nesta fórmula se tomar $x = \sum_{j \geq 1} \frac{(it)^j}{j!} \mathbb{E}(Y^j)$ ⁽¹⁾ obtém-se,

$$\log(\varphi_Y(t)) = \sum_{k \geq 1} (-1)^{k+1} \frac{1}{k} \left\{ \sum_{j \geq 1} \frac{1}{j!} \mathbb{E}(Y^j) (it)^j \right\}^k$$

⁽¹⁾Da expansão formal $\varphi(t) = \mathbb{E}(e^{itX}) = \mathbb{E} \left(\sum_{k=0}^{\infty} \frac{(itX)^k}{k!} \right)$, como $|\varphi(t)| \leq \varphi(0) = 1$ conclui-se que, exceptuando os pontos em que eventualmente $\varphi(t) = 1$ — apenas 0, no caso de variáveis aleatórias “regulares” — a expansão é válida.

ou seja,

$$\begin{aligned} \sum_{j \geq 1} \frac{(it)^j}{j!} K_j &= \log \left(1 + \sum_{j \geq 1} \frac{(it)^j}{j!} \mathbb{E}(Y^j) \right) \\ &= \sum_{k \geq 1} (-1)^{k+1} \frac{1}{k} \left\{ \sum_{j \geq 1} \frac{1}{j!} \mathbb{E}(Y^j) (it)^j \right\}^k. \end{aligned}$$

Assim, igualando os termos em $(it)^j$ tem-se

$$\begin{aligned} k_1 &= \mathbb{E}(Y) \\ k_2 &= \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \text{Var}(Y) \\ k_3 &= \mathbb{E}(Y - \mathbb{E}(Y))^3 \\ k_4 &= \mathbb{E}((Y - \mathbb{E}(Y))^4) - 3(\text{Var}(Y))^2. \end{aligned}$$

Note-se que k_j é um polinómio de grau j homogéneo nos momentos. Também os momentos podem ser expressos como polinómios homogéneos nos cumulantes.

A $\psi_X = \log(\varphi_X)$ chamamos segunda função característica da variável aleatória X . E $\psi_X(-itx) = \log(M_X(t))$ é a função geradora de cumulantes de X .

Da relação $\varphi_{aX+b}(t) = e^{itb} \varphi_X(at)$ resulta que os cumulantes de ordem $k \geq 2$ são invariantes por translações.

Para uma transformação linear do tipo $\xi = lx + m$, os cumulantes permanecem inalterados no que diz respeito à constante m e k_r vem multiplicado por l^r . A única excepção é o primeiro cumulante o qual é igual à média. Em particular, os cumulantes de uma distribuição estandardizada são iguais a $\rho_r = \frac{k_r}{\sigma^r}$, onde k_r são os cumulantes da distribuição original e σ é o desvio padrão, que se prova ser igual a $k_2^{1/2}$.

2.2 Exposição Breve sobre Famílias Exponenciais

Uma família de funções de densidade de probabilidade, ou de funções de massa de probabilidade, é dita uma família exponencial se pode ser expressa como

$$f(x|\theta) = h(x)c(\theta) \exp \left(\sum_{i=1}^k \omega_i(\theta) t_i(x) \right), \quad (2.1)$$

onde $h(x) \geq 0$ e $t_1(x), \dots, t_k(x)$ são funções reais (que não dependem de θ), $c(\theta) \geq 0$ e $\omega_1(\theta), \dots, \omega_k(\theta)$ são funções reais. A família anterior é muitas vezes designada por família exponencial de Fisher-Darmois.

Exemplo

Seja n um inteiro positivo e consideremos a família *Binomial*(n, p) com $0 < p < 1$. A função de massa de probabilidade desta família, para $x = 0, \dots, n$ e $0 < p < 1$, é

$$\begin{aligned} f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{n}{x} (1-p)^n \left(\frac{p}{1-p}\right)^x \\ &= \binom{n}{x} (1-p)^n \exp\left(\log\left(\frac{p}{1-p}\right) x\right). \end{aligned} \tag{2.2}$$

Defina-se

$$h(x) = \begin{cases} \binom{n}{x} & x = 0, \dots, n \\ 0 & \text{caso contrário} \end{cases}, \quad c(p) = (1-p)^n, \quad 0 < p < 1,$$

$$\omega_1(p) = \log\left(\frac{p}{1-p}\right), \quad 0 < p < 1, \quad \text{e} \quad t_1(x) = x.$$

Então

$$f(x|p) = h(x) c(p) \exp[\omega_1(p) t_1(x)], \tag{2.3}$$

é da forma de (2.1), com $k = 1$. Em particular, note-se que $h(x) > 0$ se $x = 0, \dots, n$ e $c(p)$ é definida apenas para $0 < p < 1$. Os valores do parâmetro $p = 0$ e $p = 1$ são muitas vezes incluídos no modelo binomial, mas neste exemplo não serão considerados, porque o conjunto de valores de x para os quais $f(x|p) > 0$ é diferente para $p = 0, 1$ e para os restantes valores de p . ■

A forma específica de (2.1) faz com que as famílias exponenciais tenham boas propriedades matemáticas. Mas mais importante para o modelo estatístico, é o modo rápido e expedito com que a forma (2.1) conduz a boas propriedades estatísticas. Vamos demonstrar o que atrás se disse, com o cálculo da média para o modelo do exemplo anterior. Para tal tenhamos em conta o seguinte teorema, (Casela and Berger, 2002, pp. 112).

Teorema 2.2.1. *Se X é uma variável aleatória com função densidade de probabilidade ou função massa de probabilidade da forma (2.1) então*

$$\mathbb{E}\left(\sum_{i=1}^n \frac{\partial \omega_i(\theta)}{\partial \theta_j} t_i(X)\right) = -\frac{\partial}{\partial \theta_j} \log[c(\theta)];$$

$$\text{Var}\left(\sum_{i=1}^n \frac{\partial \omega_i(\theta)}{\partial \theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial \theta_j^2} \log(c(\theta)) - \mathbb{E}\left(\sum_{i=1}^n \frac{\partial^2 \omega_i(\theta)}{\partial \theta_j^2} t_i(X)\right).$$

Assim, para o exemplo anterior tem-se

$$\frac{d}{dp} \omega_1(p) = \frac{d}{dp} \log \frac{p}{1-p} = \frac{1}{p(1-p)} \quad \text{e} \quad \frac{d}{dp} \log(c(p)) = \frac{d}{dp} n \log(1-p) = \frac{-n}{1-p},$$

e, conseqüentemente,

$$\mathbb{E} \left(\frac{1}{p(1-p)} X \right) = \frac{n}{1-p} \Leftrightarrow \mathbb{E}(X) = np.$$

De modo análogo se obteria a variância.

Uma família exponencial pode ser reparametrizada na forma

$$f(x|\theta) = h(x)c^*(\eta) \exp \left(\sum_{i=1}^k \eta_i t_i(x) \right), \quad (2.4)$$

onde as funções $h(x)$ e $t_i(x)$ são as mesmas da parametrização (2.1) e $\eta = (\eta_1, \dots, \eta_k)$. O conjunto $\mathcal{H} = \left\{ \eta = (\eta_1, \dots, \eta_k) : \int_{\mathbb{R}} h(x) \exp \left(\sum_{i=1}^k \eta_i t_i(x) \right) dx < \infty \right\}$ é chamado espaço natural dos parâmetros para a família. (O integral será substituído por uma soma para todos os valores de x para os quais $h(x) > 0$ se X for discreta.) Os valores de $\eta \in \mathcal{H}$, têm que ser tais que $c^*(\eta) = \left[\int_{\mathbb{R}} h(x) \exp \left(\sum_{i=1}^k \eta_i t_i(x) \right) dx \right]^{-1}$ para garantir que f é de facto uma função densidade de probabilidade.

Uma forma natural de abordar a família exponencial de Fisher-Darmois tem que ver com a função limite inferior de Cramér-Rao. Escolhemos apresentá-la desse modo, pois no capítulo em que abordamos o Teorema Limite Central na perspectiva da lei da entropia máxima, vamos estabelecer desigualdades de entre as quais a de Cramér-Rao é um caso especial.

2.2.1 Desigualdade de Cramér – Rao

Suponha-se que, para estimar o parâmetro $\tau(\theta)$ da distribuição $f(x|\theta)$, se pode especificar um limite inferior, digamos $B(\theta)$, para a variância de um qualquer estimador não enviesado de $\tau(\theta)$. Se se puder encontrar um estimador não enviesado W^* que satisfaça $\text{Var}_\theta(W^*) = B(\theta)$, encontramos o melhor estimador não enviesado. Este tipo de aproximação usa o limite inferior de Cramér-Rao.

Teorema 2.2.2. Teorema da Desigualdade de Cramér-Rao

Seja (X_1, \dots, X_n) uma amostra aleatória com função de densidade de probabilidade conjunta $f(\mathbf{x}|\theta)$, e seja $W(\mathbf{X}) = W(X_1, \dots, X_n)$ um estimador satisfazendo

$$\frac{d}{d\theta} \mathbb{E}_\theta[W(\mathbf{X})] = \int_{\mathcal{S}} \frac{\partial}{\partial \theta} [W(\mathbf{x}) f(\mathbf{x}|\theta)] d\mathbf{x} \quad \text{e} \quad \text{Var}_\theta[W(\mathbf{S})] < \infty. \quad (2.5)$$

Então

$$\text{Var}_\theta[W(\mathbf{X})] \geq \frac{\left\{ \frac{d}{d\theta} \mathbb{E}_\theta[W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta)) \right]^2 \right\}}. \quad (2.6)$$

Demonstração

A prova deste teorema resulta duma simples aplicação da desigualdade de Cauchy-Schwarz, isto é, do facto de para quaisquer duas variáveis aleatórias

$$[\text{cov}(X, Y)]^2 \leq [\text{Var}(X)][\text{Var}(Y)]. \quad (2.7)$$

De (2.7) resulta o limite inferior para a variância de X ,

$$\text{Var}(X) \geq \frac{[\text{cov}(X, Y)]^2}{\text{Var}(Y)}.$$

De seguida, basta substituir na desigualdade anterior X pelo estimador $W(\mathbf{X})$ e Y pela quantidade $\frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta))$, e aplicar a desigualdade de Cauchy-Schwarz.

Note-se primeiro que

$$\begin{aligned} \frac{d}{d\theta} \mathbb{E}_\theta[W(\mathbf{X})] &= \int_S W(\mathbf{x}) \left[\frac{\partial}{\partial \theta} f(\mathbf{x}|\theta) \right] d\mathbf{x} \\ &= \int_S W(\mathbf{x}) \left[\frac{\partial}{\partial \theta} f(\mathbf{x}|\theta) \right] \frac{f(\mathbf{x}|\theta)}{f(\mathbf{x}|\theta)} d\mathbf{x} \\ &= \mathbb{E}_\theta \left[W(\mathbf{X}) \frac{\frac{\partial}{\partial \theta} f(\mathbf{X}|\theta)}{f(\mathbf{X}|\theta)} \right] \\ &= \mathbb{E}_\theta \left[W(\mathbf{X}) \frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta)) \right], \end{aligned} \quad (2.8)$$

Para se determinar a covariância é necessário subtrair o produto dos valores esperados. Vamos então calcular $\mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta)) \right)$. Se em (2.8) fizermos $W(\mathbf{X}) = 1$, tem-se

$$\mathbb{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(\mathbf{X}|\theta)] \right\} = \frac{d}{d\theta} \mathbb{E}_\theta(1) = 0. \quad (2.9)$$

Assim, a $\text{Cov}_\theta \left\{ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log[f(\mathbf{X}|\theta)] \right\}$ é igual à esperança de produto. Então de (2.8) e (2.9) tem-se

$$\begin{aligned} \text{Cov}_\theta \left\{ W(\mathbf{X}), \frac{\partial}{\partial \theta} \log[f(\mathbf{X}|\theta)] \right\} &= \mathbb{E}_\theta \left\{ W(\mathbf{X}) \frac{\partial}{\partial \theta} \log[f(\mathbf{X}|\theta)] \right\} \\ &= \frac{d}{d\theta} \mathbb{E}_\theta[W(\mathbf{X})]. \end{aligned} \quad (2.10)$$

Além disso, como $\mathbb{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(\mathbf{X}|\theta)] \right\} = 0$ tem-se

$$\text{Var}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(\mathbf{X}|\theta)] \right\} = \mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta)) \right]^2 \right\}. \quad (2.11)$$

Usando a desigualdade de Cauchy-Schwarz, juntamente com (2.10) e (2.11), obtém-se

$$\text{Var}_\theta[W(\mathbf{X})] \geq \frac{\left\{ \frac{d}{d\theta} \mathbb{E}_\theta[W(\mathbf{X})] \right\}^2}{\mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta)) \right]^2 \right\}}.$$

□

Se se estiver a trabalhar sob a hipótese de amostras independentes, os cálculos para encontrar o limite inferior são bastantes simplificados.

Corolário 2.2.2.1. *Se as hipóteses do teorema anterior forem satisfeitas e se, adicionalmente, as variáveis aleatórias do vector \mathbf{X} , (X_1, \dots, X_n) forem independentes e identicamente distribuídas com função densidade de probabilidade $f(X|\theta)$, então*

$$\text{Var}_\theta[W(\mathbf{X})] \geq \frac{\left\{ \frac{d}{d\theta} \mathbb{E}_\theta[W(\mathbf{X})] \right\}^2}{n \mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(X|\theta)) \right]^2 \right\}}.$$

Demonstração

Neste caso, é apenas necessário provar que:

$$\mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta)) \right]^2 \right\} = n \mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(X|\theta)) \right]^2 \right\}.$$

Como X_1, \dots, X_n são independentes

$$\begin{aligned} \mathbb{E}_\theta \left[\left\{ \frac{\partial}{\partial \theta} \log[f(\mathbf{X}|\theta)] \right\}^2 \right] &= \mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log \prod_{i=1}^n f(X_i|\theta) \right]^2 \right\} \\ &= \mathbb{E}_\theta \left\{ \left[\sum_{i=1}^n \frac{\partial}{\partial \theta} \log(f(X_i|\theta)) \right]^2 \right\} \\ &= \sum_{i=1}^n \mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(X_i|\theta)) \right]^2 \right\} + \\ &+ \sum_{i \neq j} \mathbb{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(X_i|\theta)] \frac{\partial}{\partial \theta} \log[f(X_j|\theta)] \right\}. \end{aligned} \quad (2.12)$$

Para $i \neq j$, atendendo a (2.10) e ao facto dos X_i serem independentes tem-se

$$\mathbb{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(X_i|\theta)] \frac{\partial}{\partial \theta} \log[f(X_j|\theta)] \right\} = \mathbb{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(X_i|\theta)] \right\} \mathbb{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(X_j|\theta)] \right\} = 0,$$

Então a soma em (2.12) reduz ao primeiro termo, sendo este igual a

$$\sum_{i=1}^n \mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(X_i|\theta)) \right]^2 \right\} = n \mathbb{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(X|\theta)] \right\},$$

pois os X_i são identicamente distribuídos a X . Então

$$\mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta)) \right]^2 \right\} = n \mathbb{E}_\theta \left\{ \frac{\partial}{\partial \theta} \log[f(\mathbf{X}|\theta)] \right\},$$

e o corolário está provado. □

Embora se tenha mostrado os resultados anteriores para vectores absolutamente contínuos estes são válidos para outro tipo de vectores.

A quantidade $\mathbb{E}_\theta \left\{ \left[\frac{\partial}{\partial \theta} \log(f(\mathbf{X}|\theta)) \right]^2 \right\}$ é chamada informação de Fisher da amostra. Esta terminologia reflecte o facto de a informação de Fisher dar um limite inferior para a variância do melhor estimador não enviesado de θ . À medida que a informação de Fisher aumenta, tem-se mais informação sobre θ , o limite inferior é menor para a variância do melhor estimador não enviesado.

O termo “desigualdade de informação” é uma alternativa ao termo “desigualdade de Cramér-Rao”. A noção de informação de Fisher existe numa forma mais geral do que a aqui apresentada. A forma mais geral desta desigualdade é obtida quando as hipóteses efectuadas sobre os estimadores são substituídas por hipóteses sobre a distribuição subjacente. Mais à frente neste trabalho, voltar-se-á à desigualdade de Cramér-Rao na sua forma mais geral.

2.2.2 Condições para atingir o Limite Inferior na Desigualdade de Cramér-Rao

As condições para atingir o limite inferior da desigualdade de Cramér-Rao são bastante simples. Lembramos que a desigualdade de Cramér-Rao é consequência de uma aplicação da desigualdade de Cauchy-Schwarz, pelo que, as condições para atingir o limite da desigualdade de Cramér-Rao são as mesmas para atingir a igualdade na desigualdade de Cauchy-Schwarz.

Teorema 2.2.3. *Seja (X_1, \dots, X_n) uma amostra constituída por variáveis aleatórias independentes e identicamente distribuídas com função de densidade de probabilidade conjunta $f(\mathbf{X}|\theta)$, onde $f(\mathbf{X}|\theta)$ satisfaz as condições do teorema de Cramér-Rao. Seja $L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ a função de verosimilhança. Se $W(\mathbf{X}) = W(X_1, \dots, X_n)$ é um estimador não enviesado de $\tau(\theta)$ então $W(\mathbf{X})$ atinge o limite inferior de Cramér-Rao se e só se*

$$a(\theta) [W(\mathbf{X}) - \tau(\theta)] = \frac{\partial}{\partial \theta} \log[L(\theta|\mathbf{X})], \quad (2.13)$$

para alguma função $a(\theta)$.

Demonstração

A desigualdade de Cramér-Rao pode ser escrita como

$$\text{cov}_\theta \left[W(\mathbf{X}), \frac{\partial}{\partial \theta} \log \prod_{i=1}^n f(X_i|\theta) \right] \leq \text{Var}_\theta [W(\mathbf{X})] \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log \prod_{i=1}^n f(X_i|\theta) \right).$$

Recordando que $\mathbb{E}_\theta [W(\mathbf{X})] = \tau(\theta)$, $\mathbb{E}_\theta \left[\frac{\partial}{\partial \theta} \log \prod_{i=1}^n f(X_i|\theta) \right] = 0$, e atendendo às propriedades que dão a igualdade na desigualdade de Cramér-Rao, tem-se que a igualdade é atingida se e só se $W(\mathbf{X}) - \tau(\theta)$ é proporcional a $\frac{\partial}{\partial \theta} \log \prod_{i=1}^n f(X_i|\theta)$. Que é exactamente o que está expresso em (2.13). □

As famílias exponenciais surgem naturalmente quando se investigam as condições sob as quais é atingido o limite inferior na desigualdade de Cramér-Rao.

Consideremos uma função densidade de probabilidade que depende de modo regular do parâmetro θ e verifica as seguintes condições

$$\int_S \frac{\partial}{\partial \theta} f(x; \theta) dx = 0.$$

Consideremos as variáveis aleatórias X_1, \dots, X_n independentes e identicamente distribuídas com função densidade probabilidade conjunta $f(\mathbf{x}, \theta) = f(x_1; \theta) \cdots f(x_n; \theta)$. Seja $\theta^*(\mathbf{X})$ um estimador não enviesado de θ que satisfaz a condição

$$\int_{\mathbb{R}^n} \theta^*(\mathbf{X}) \frac{\partial}{\partial \theta} f(\mathbf{x}; \theta) d\mathbf{x} = 1.$$

Então para qualquer estimador T deste tipo verifica-se que pelo corolário 2.2.2.1

$$\text{Var}(T) \geq \frac{1}{n A(\theta)},$$

onde $A(\theta) = \int \frac{\left(\frac{\partial f(x, \theta)}{\partial \theta} \right)^2}{f(x, \theta)} dx$, é a chamada informação de Fisher.

Seja $D(x, \theta) = \frac{\partial}{\partial \theta} \ln[f(x, \theta)]$. Prova-se que

$$\int_{\mathbb{R}^n} (\theta^*(\mathbf{x}) - \theta) \left(\sum_{i=1}^n D(x_i, \theta) \right) f(\mathbf{x}, \theta) d\mathbf{x} = 1, \theta \in \Theta,$$

que definindo

$$\langle g_1, g_2 \rangle = \int g_1(\mathbf{x}) g_2(\mathbf{x}) \eta(\mathbf{x}) d\mathbf{x},$$

pode ser reescrito na forma

$$\langle g_1, g_2 \rangle = 1,$$

com $g_1(\mathbf{x}) = (\theta^*(\mathbf{x}) - \theta)$, $g_2(\mathbf{x}) = \sum_{i=1}^n D(x_i, \theta)$ e $\eta(\mathbf{x}) = f(\mathbf{x}, \theta) \geq 0$.

Para obter a igualdade na desigualdade de Cramér-Rao é necessário que as funções anteriores sejam linearmente dependentes: assim consideremos $g_1(x) = \lambda g_2(x)$. Então

$$\int g_1(x) g_2(x) \eta(x) dx = \lambda \int g_2^2(x) \eta(x) dx.$$

Mas, como neste caso $\int g_1(x) g_2(x) \eta(x) dx = 1$, tem-se que

$$\begin{aligned} \lambda \int g_2^2(x) \eta(x) dx = 1 &\Leftrightarrow \lambda = \left[\int g_2^2(x) \eta(x) dx \right]^{-1} \\ &\Leftrightarrow \lambda = \frac{1}{n A(\theta)}. \end{aligned}$$

Assim, obtém-se a relação

$$\theta^* - \theta = \frac{1}{n A(\theta)} \sum_{i=1}^n D(X_i, \theta),$$

ou

$$\theta^* = \theta + \frac{1}{n A(\theta)} \sum_{i=1}^n D(X_i, \theta) = \frac{1}{n} \sum_{i=1}^n \left[\theta + \frac{D(X_i, \theta)}{A(\theta)} \right].$$

Note-se que, o primeiro termo da equação acima não depende de θ . Ou seja, cada termo $\theta + \frac{D(X_i, \theta)}{A(\theta)}$ deve ser independente de θ ; se tomarmos $c(x) = \theta + \frac{D(X_i, \theta)}{A(\theta)}$, tem-se

$$D(X_i, \theta) = [c(x) - \theta] A(\theta), \theta \in \Theta \tag{2.14}$$

e o estimador θ^* tem a forma

$$\theta^*(X) = \frac{1}{n} \sum_{i=1}^n c(X_i). \tag{2.15}$$

Resolvendo a equação (2.14):

$$\frac{\partial}{\partial \theta} \log[f(x; \theta)] = A(\theta) (c(x) - \theta) ,$$

onde $A(\theta) = B''(\theta)$, obtém-se

$$\log[f(X|\theta)] = B'(\theta) (c(x) - \theta) + B(\theta) + H(x) .$$

Então

$$f(X|\theta) = \exp [B'(\theta) (c(x) - \theta) + B(\theta) + H(x)] . \quad (2.16)$$

Ou seja, a família anterior é uma família exponencial. Assim, conclui-se que a igualdade na desigualdade de Cramér-Rao é atingida se e só se a família $\{f(x; \theta), \text{ com } \theta \in \Theta\}$ é exponencial, isto é, dada pela equação (2.16), onde $B''(\theta) > 0$. Neste caso o estimador erro quadrático médio mínimo de θ é dado por (2.15), a sua variância é igual a $\frac{1}{nB''(\theta)}$, e a informação de Fisher iguala $B''(\theta)$.

2.3 Polinómios Ortogonais e Famílias Exponenciais

Morris (1982) estudou um conjunto de variáveis da família exponencial, contendo a gaussiana (μ, σ) e a Poisson (μ) , caracterizada por $\text{Var}(X) = v_0 + v_1\mu + v_2\mu^2$, onde $\mathbb{E}(X) = \mu$ e $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$, isto é, pela variância ser expressa como função do valor médio na forma de um polinómio de grau máximo dois. Para além da gaussiana e da Poisson demonstrou que as restantes soluções eram a gama, a binomial, a binomial negativa e a secante hiperbólica.

É interessante anotar que às medidas de probabilidade associadas a cada uma daquelas soluções, estão associados polinómios ortogonais da classe de Meixner. Estes polinómios são de grande importância, nomeadamente no contexto das expansões assintóticas, pois em muitos casos os polinómios p_j de grau j em x que aparecem, quer nas expansões de Edgeworth quer nas de Gram-Charlier, são ortogonais em relação à distribuição associada a f , que é, neste contexto, considerada a aproximação de primeira ordem.

Genericamente, pode-se definir polinómios ortogonais do seguinte modo. Seja μ uma medida na recta real tal que os momentos

$$c_n = \int_{\mathbb{R}} x^n d\mu(x), \quad n = 0, \dots$$

são finitos. O suporte de μ é definido como

$$S = \text{supp}(\mu) = \{x \in \mathbb{R} : \forall \varepsilon > 0, \mu[(x - \varepsilon, x + \varepsilon)] > 0\} .$$

Considere-se a sucessão de polinómios $\{P_k(x)\}_{k \geq 0}$, onde P_k é de grau exacto igual a k . A sucessão é dita ortogonal, com respeito à medida μ se e só se

$$\int_{\mathbb{R}} P_m(x) P_k(x) d\mu(x) = 0, \quad m \neq k.$$

Se o coeficiente director de $P_m(x)$ for um (para todo $m \geq 0$), então os polinómios são chamados de polinómios ortogonais mónicos com respeito à medida μ . Se além disso, a sucessão $\{P_k(x)\}_{k \geq 0}$ satisfizer adicionalmente

$$\int_{\mathbb{R}} [P_m(x)]^2 d\mu(x) = 1, \quad m \geq 0,$$

então a sucessão é chamada de ortonormada, com respeito à medida μ .

A função geradora dos polinómios ortogonais de Meixner é da forma

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} M_k(x) = g(z) \exp[x h(z)],$$

onde quer g quer h admitem expansões em série de potências. Por exemplo, no que se refere às famílias “naturais” de Morris temos as seguintes correspondências:

Distribuição	Polinómio
Gaussiana	Hermite
Poisson	Poisson-Charlier
Gama	Laguerre
Binomial	Krawtchouk
Binomial Negativa	Meixner
Secante Hiperbólica	Pollaczek

2.4 Famílias Exponenciais Naturais com Função de Variância Quadrática

Uma vez que nos interessámos por questões associadas a pares conjugados Malva e Sequeira, (2003) expomos a teoria NEF (*Natural Exponential Families*) em que a variância se pode exprimir polinomialmente, com grau inferior ou igual a dois, como função do valor médio, área em que o trabalho seminal de Morris (1982) levou a perceber a relevância dos polinómios ortogonais no tratamento expedito de problemas do tipo que nos ocuparam.

Neste contexto, surgem inesperadamente as famílias de distribuição secante hiperbólica, que já Lévy e Bass (1967) tinham anotado entre as expressões que são simultaneamente, a menos de uma constante multiplicativa adequada, função densidade de probabilidade e função característica.

Como já vimos, uma família de distribuições paramétricas com $\Theta \subset \mathbb{R}$, o conjunto de parâmetros naturais, é uma família exponencial univariada se as variáveis aleatórias Y com estas distribuições satisfazem

$$P_\theta(Y \in A) = \int_A \exp\{\theta T(y) - \psi(\theta)\} d\xi(y) \quad (2.17)$$

para alguma medida ξ não dependente de $\theta \in \Theta$, $A \subset \mathbb{R}$ um conjunto mensurável e T uma função de valores reais mensurável. Qualquer factor da densidade de Y que não dependa de θ é absorvido em ξ . Em (2.17), θ é chamado parâmetro natural e Θ , que é o maior conjunto para o qual (2.17) é finito quando $A \equiv \mathbb{R}$, é chamado espaço de parâmetros naturais. Muitas vezes θ é uma função não-linear de vários parâmetros naturais. A função $\psi(\theta)$ é determinada de modo que o integral em (2.17) seja igual a um para $A \equiv \mathbb{R}$. A observação natural em $A \equiv \mathbb{R}$ é $X = T(Y)$. A sua distribuição pertence à família exponencial natural (NEF),

$$P_\theta(Y \in A) = \int_A \exp\{\theta x - \psi(\theta)\} dF(x) \quad (2.18)$$

com F a medida de Stieltjes em \mathbb{R} . Se $0 \in \Theta$ e $\psi(0) = 0$ então F é função de distribuição. Caso contrário, para qualquer $\theta_0 \in \Theta$, seja $dF_0(x) = \exp\{\theta_0 x - \psi(\theta_0)\} dF$. Então F_0 é uma função distribuição e por (2.18) gera a mesma família exponencial que F . Assim, sem perda de generalidade, assume-se que F em (2.18) é uma função distribuição.

Toda a função de distribuição F_0 que possua função *fgm* numa vizinhança de zero gera uma NEF do seguinte modo. Defina-se a função geradora de cumulantes $\psi(\theta)$ em Θ , o maior intervalo para o qual a função geradora de cumulantes existe, por

$$\psi(\theta) = \log \left(\int \exp(\theta x) dF_0(x) \right), \quad \theta \in \Theta. \quad (2.19)$$

Então as funções de distribuição F_θ , $\theta \in \Theta$, definidas por

$$dF_\theta(x) = \exp\{\theta x - \psi(\theta)\} dF_0(x), \quad (2.20)$$

formam uma NEF, com F_θ função de distribuição. A NEF assim gerada é chamada família conjugada. Para qualquer $\theta^* \in \Theta$, F_{θ^*} gera a mesma NEF que (2.19) e (2.20). Ou seja o conjunto das distribuições NEF é fechado, pois qualquer elemento da família pode gerar a família da forma descrita.

Poucas NEF têm função de variância quadrática (QVF) dada por

$$\text{Var}(X) = V(\mu) = v_0 + v_1\mu + v_2\mu^2. \quad (2.21)$$

Se X tem uma distribuição NEF e variância, $V(\mu)$ quadrática, dizemos que X tem uma distribuição exponencial natural com função variância quadrática — NEF-QVF. As distribuições gaussianas, Poisson, gama, binomial e binomial negativa são exemplos de distribuições NEF-QVF.

Vimos anteriormente, que a distribuição gaussiana pertence à família exponencial. Além disso, sabemos que se $X \sim \text{Gaussiana}(\mu, \sigma)$ então $\text{Var}(X) = \mu$ ou seja, $\text{Var}(X) = v_0 + 0v_1 + 0v_2$ isto é, a gaussiana é uma NEF-QVF, sendo a função de variância constante.

Sejam $c \neq 0$ e b constantes, e seja $X^* = \frac{X - b}{c}$ uma transformação linear de X com média $\mu^* = \frac{\mu - b}{c}$. Então

$$V^*(\mu^*) \equiv \text{Var}(X^*) = \frac{\text{Var}(X)}{c^2} = \frac{V(\mu)}{c^2} = \frac{V(c\mu^* + b)}{c^2}.$$

Assim,

$$\begin{aligned} V^*(\mu^*) &= \frac{1}{c^2} V(\mu) = \frac{1}{c^2} [v_0 + v_1\mu + v_2\mu^2] \\ &\Leftrightarrow V^*(\mu^*) = \frac{1}{c^2} [v_0 + v_1(\mu^*c + b) + v_2(\mu^*c + b)^2] \\ &\Leftrightarrow V^*(\mu^*) = \frac{v_0 + v_1b + v_2b^2}{c^2} + \frac{v_1 + 2bv_2}{c}\mu^* + (\mu^*)^2v_2. \end{aligned}$$

Ora

$$V(b) = v_2b^2 + v_1b + v_0 \Rightarrow V'(b) = 2v_2b + v_1,$$

donde

$$V^*(\mu^*) = \frac{V(b)}{c^2} + \frac{V'(b)}{c}\mu^* + v_2(\mu^*)^2. \quad (2.22)$$

Ou seja, se considerarmos a variável aleatória X^* , resultante da transformação linear de X apresentada, então a variância desta nova variável aleatória é também quadrática, como função do valor médio de X^* , tendo coeficientes $v_2^* = v_2$, $v_1^* = \frac{V'(b)}{c}$ e $v_0^* = \frac{V(b)}{c^2}$.

Defina-se $d \equiv v_1^2 - 4v_0v_1$ como o discriminante de $V(\mu)$. Como

$$[V'(\mu)]^2 = (2v_2\mu + v_1)^2 = 4v_2^2\mu^2 + 4v_1v_2\mu + v_1^2 = 4v_2(v_2\mu^2 + v_1\mu + v_0) - 4v_2v_0 + v_1^2$$

então,

$$[V'(\mu)]^2 = 4v_2V(\mu) + d, \quad (2.23)$$

então o discriminante d^* de (2.22) é

$$d^* = \left[\frac{V'(b)}{c} \right]^2 - 4v_2 \frac{V(b)}{c^2} = \frac{4v_2}{c^2} V(b) + \frac{d}{c^2} - \frac{4v_2}{c^2} V(b) = \frac{d}{c^2},$$

ou seja, d é inalterável por translações, mas não por homotetias.

Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas com uma distribuição NEF-QVF, com $V(\mu)$ a satisfazer (2.21). Defina-se

$$X^* = \frac{X_1 + X_2 + \dots + X_n - nb}{c}.$$

Então X^* tem uma distribuição NEF-QVF com

$$\Theta^* = \left\{ Y : Y = \frac{n(X - b)}{c}, \text{ para } X \in \Theta \right\},$$

e

$$\mu^* = E(X^*) = \frac{n(\mu - b)}{c}$$

sendo,

$$\begin{aligned} V^*(\mu^*) &= Var(X^*) = Var \left[\frac{1}{c}(X_1 + \dots + X_n) - \frac{nb}{c} \right] \\ &\Leftrightarrow V^*(\mu^*) = \frac{n}{c^2} Var(X_1) = \frac{n}{c^2} V(\mu) = \frac{n}{c^2} (v_0 + v_1\mu + v_2\mu^2) \\ &\Leftrightarrow V^*(\mu^*) = \frac{n}{c^2} \left[v_0 + \frac{v_1c}{n}\mu^* + v_1b + v_2 \left(\frac{c^2(\mu^*)^2}{n^2} + 2\frac{cb}{n}\mu^* + b^2 \right) \right] \\ &\Leftrightarrow V^*(\mu^*) = \frac{n}{c^2} V(b) + \frac{1}{c} V'(b)\mu^* + \frac{v_2}{n} (\mu^*)^2. \end{aligned} \tag{2.24}$$

Assim, se se considerar a variável aleatória X^* , resultante da transformação linear de X apresentada, então a variância desta nova variável aleatória é também quadrática, como função do valor médio de X^* , tendo como coeficientes

$$v_0^* = \frac{n}{c^2} V(b), \quad v_1^* = \frac{1}{c} V'(b) \quad \text{e} \quad v_2^* = \frac{v_2}{n}, \tag{2.25}$$

e o discriminante

$$d^* = \frac{[V'(b)]^2}{c^2} - 4\frac{n}{c^2} V(b) \frac{v_2}{n} = \frac{1}{c^2} [4v_2 V(b) + d] - \frac{4v_2}{c^2} V(b) = \frac{d}{c^2},$$

isto é, $d^* = \frac{d}{c^2}$.

As fórmulas (2.24) e (2.25) mostram que a propriedade QVF é preservada para transformações lineares e convoluções e mostram as alterações que estas transformações e convoluções provocam na função variância.

De seguida, caracterizam-se todas as distribuições NEF-QVF estritamente quadráticas. Suponhamos que X tem uma distribuição NEF-QVF com função variância dada por

$$V(u) = v_0 + v_1u + v_2u^2 \quad \text{e} \quad v_2 \neq 0.$$

Defina-se para

$$a = \begin{cases} 1 & \text{se } d = 0 \\ |dv_2|^{-1/2} & \text{caso contrário} \end{cases},$$

$$\begin{aligned} X^* &= a V'(X) = a(v_1 + 2v_2 X) = 2a v_2 X - (-a v_1) \\ &= \frac{X - \frac{v_1}{2v_2}}{\frac{1}{2av_2}} = \frac{X - b}{c}, \end{aligned}$$

com $b = \frac{v_1}{2v_2}$ e $c = \frac{1}{2av_2}$.

Então X^* é uma função linear de X com função de variância dada por

$$V^*(\mu^*) = \frac{V(b)}{c^2} + \frac{V'(b)}{c} \mu^* + v_2(\mu^*)^2,$$

isto é,

$$V^*(\mu^*) = s + v_2(\mu^*)^2 \quad \text{com } s = -\text{sgn}(dv_2). \quad (2.26)$$

A equação (2.26) pode ser vista como o membro canónico para transformações lineares de NEF-QVF com $v_2 \neq 0$ ($v_1 = 0, s = 0, \pm 1$). Todas as outras funções de variância quadrática, isto é, todos os outros v_1 e v_0 podem ser obtidos a partir da forma canónica de (2.26) através da transformação linear $X = \left(\frac{X^*}{a} - v_1\right) \frac{1}{2v_2}$, a transformação inversa de $X^* = aV'(X)$.

Fazendo $v_2 \neq 0$ em (2.26) existem seis combinações possíveis para $V^*(\mu^*)$, correspondentes a $v_2 < 0$, $v_2 > 0$ e $s = -1, 0, 1$. Os casos $v_2 < 0$ e $s = -1$ ou $s = 1$ não podem ocorrer pois implicam $V^*(\mu^*) < 0$. O caso $v_2 < 0$ e $s = 1$ corresponde a transformações lineares da distribuição binomial, $v_2 > 0$ e $s = 0$ ou $s = -1$ correspondem, respectivamente, a transformações lineares das distribuições gama e binomial negativa.

Se $v_2 = 1$ e $s = 1$ a distribuição correspondente é a da observação natural para a família exponencial beta. Seja

$$Y \sim \text{beta} \left(\frac{1}{2} + \frac{\theta}{\pi}, \frac{1}{2} - \frac{\theta}{\pi} \right), \quad |\theta| < \frac{\pi}{2}.$$

A observação natural para esta família exponencial é

$$X = \frac{1}{\pi} \ln \frac{Y}{1 - Y},$$

que tem função densidade de probabilidade dada por

$$f(x) = \frac{\exp(\theta x + \ln(\cos(\theta)))}{2 \cosh(\frac{\pi x}{2})}. \quad (2.27)$$

Derivando a função geradora de cumulantes, $\psi(\theta) = -\log(\cos(\theta))$, obtém-se a média e a variância de (2.27)

$$E(X) = \mu = \psi'(\theta) = \tan(\theta), \quad \text{Var}(X) = V(\mu) = \psi''(\theta) = \text{cosec}^2(\theta) = 1 + \mu^2,$$

respectivamente.

À custa da variável base, usando mudanças de escala e convoluções, geramos todas as possíveis variáveis aleatórias com as propriedades requeridas, no que se refere à expressão da variância como função quadrática do valor médio, isto é, como

$$V(\mu) = r + \frac{\mu^2}{r}, \quad \text{para qualquer } r > 0.$$

Se na expressão (2.27) se fizer $\theta = 0$ obtemos a distribuição secante hiperbólica-HS, com função densidade de probabilidade dada por

$$f(x) = \frac{1}{2 \cosh\left(\frac{\pi x}{2}\right)}.$$

Analogamente, da distribuição anterior produzem-se as distribuições secante hiperbólica generalizada—GHS, que são distribuições simétricas. A família exponencial natural gerada pela distribuição secante hiperbólica generalizada será referida como distribuição NEF-GHS, enquanto a notação NEF-HS será reservada para (2.27).

Tem-se assim, todas as famílias exponenciais naturais univariadas com função variância quadrática: gaussiana com função variância constante, Poisson com função variância linear, e as distribuições gama, binomial, binomial negativa, NEF-GHS e transformações lineares destas.

Cada uma destas distribuições contém uma subfamília, a qual é chamada família de distribuições elementares, com coeficiente director em $V(\mu)$ igual a ± 1 . Estas são a distribuição gaussiana com variância unitária, ($v_0 = 1$), a Poisson usual ($v_1 = 1$), a exponencial ($v_2 = 1$ na *gama*), a Bernoulli ($v_2 = -1, r = 1$ na binomial), a geométrica ($v_2 = 1, r = 1$ na binomial negativa), e a NEF-HS (2.27).

2.4.1 Polinómios Ortogonais para NEF-QVF

Seja $f(x, \theta)$ uma densidade NEF-QVF proporcional a $f(x, \theta) = \exp(\theta x - \psi(\theta))$ relativamente a alguma medida. Defina-se

$$P_m(x, \mu) = \frac{V^m(\mu) \frac{\partial^m f(x, \theta)}{\partial \mu^m}}{f(x, \theta)} \quad \text{para } m = 0, 1, 2, \dots \quad (2.28)$$

Prova-se que

$$P_0(x, \mu) = 1, \quad P_1(x, \mu) = x - \mu \quad \text{e} \quad P_2(x, \mu) = (x - \mu)^2 - V'(\mu)(x - \mu) - V(\mu).$$

Mostra-se que $P_m(x, \mu)$ é um polinómio de grau m em x e μ e com termo principal x^m , e que P_m é uma família de polinómios ortogonais.

De (2.28) temos que (para simplificar a notação omitiremos os argumentos de P_m)

$$P_{m+1} = \frac{V^{m+1} \frac{\partial^{m+1} f(x, \theta)}{\partial \mu^{m+1}}}{f(x, \theta)},$$

como

$$\frac{\partial^{m+1} f(x, \theta)}{\partial \mu^{m+1}} = (P_m f V^{-m})',$$

então

$$\begin{aligned} P_{m+1} &= \frac{V^{m+1} d(P_m f V^{-m})}{f d\mu} = V^{m+1} f^{-1} (P_m' f V^{-m} + P_m f' V^{-m} + P_m f (V^{-m})') \\ &= P_m' f^{-1} f V^{m+1-m} + P_m f^{-1} (f)' V^{m+1-m} + P_m V^{m+1} f^{-1} f (V^{-m})' \\ &= P_m' V + P_m V f^{-1} \frac{P_1 f}{V} + P_m V^{m+1} V^{-m-1} (-V') m \\ &= P_m' V + P_m P_1 - m V' P_m \\ &= (P_1 - m V') P_m + V P_m' \end{aligned} \tag{2.29}$$

com

$$P_m' = \frac{\partial P_m(x, \mu)}{\partial \mu} \quad \text{e} \quad P_m^{(r)} = \frac{\partial^r P_m(x, \mu)}{\partial \mu^r}$$

Teorema 2.4.1. *O conjunto $\{P_m(x, \mu) : m = 0, 1, \dots\}$ é, para as famílias NEF-QVF, um sistema ortogonal de polinómios com respeito a $f(x, \theta) = \exp[x\theta - \psi(\theta)]$. $P_m(x, \theta)$ tem grau exacto igual a m tanto em μ com em x com termo director x^m e é gerado por*

$$P_{m+1} = (P_1 - m V') P_m - m[1 + (m-1)v_2] V P_{m-1} \tag{2.30}$$

para $m \geq 1$ com $P_0 = 1$, $P_1 = x - \mu$. Defina-se $a_0 = 1$ e para $m \geq 1$

$$a_m = m! \prod_{i=0}^{m-1} (1 + i v_2). \tag{2.31}$$

Então para $m \geq 1$, $r = 0, 1, \dots, m$, as derivadas em ordem a μ são

$$P_m^{(r)} = (-1)^r \frac{a_m}{a_{m-r}} P_{m-r}. \tag{2.32}$$

Finalmente, $\mathbb{E}_\theta(P_m) = 0$ para $m \geq 1$ e

$$\mathbb{E}_\theta(P_m P_n) = \delta_{mn} a_m V^m, \quad m, n \geq 0. \tag{2.33}$$

Demonstração

Defina-se $b_m = (m + 1)(1 + mv_2)$, e prove-se (2.31) para $r = 1$. Tem-se

$$\frac{a_m}{a_{m-1}} = \frac{m! \prod_{i=0}^{m-1} (1 + iv_2)}{(m-1)! \prod_{i=0}^{m-2} (1 + iv_2)} = m[1 + (m-1)v_2] = b_{m-1}.$$

Então,

$$P'_m = -\frac{a_m}{a_{m-1}}P_{m-1} \Leftrightarrow P'_m = -b_{m-1}P_{m-1}, \quad (2.34)$$

que para $m = 1$ é verdadeira pois $P'_1 = -b_0P_0$ e $P_0 = b_0 = 1$. Assuma-se que (2.30) é válida para $m \geq 1$ então

$$P_{m+1} = (P_1 - mV')P_m + VP'_m \Leftrightarrow P_{m+1} = (P_1 - mV')P_m - Vb_{m-1}P_{m-1}.$$

Diferenciando a igualdade anterior em relação a μ obtém-se

$$\begin{aligned} P'_{m+1} &= (P'_1 - mV'')P_m + (P_1 - mV')P'_m - b_{m-1}V'P_{m-1} - b_{m-1}VP'_{m-1} \\ &= (-1 - mV'')P_m - [(P_1 - mV') + V']b_{m-1}P_{m-1} - Vb_{m-1}P'_{m-1} \\ &= (-1 - mV'')P_m - b_{m-1}\{[P_1 - (m-1)V']P_{m-1} - VP'_{m-1}\} \\ &= -(1 + mV'')P_m - b_{m-1}\underbrace{\{[P_1 - (m-1)V']P_{m-1} + Vb_{m-2}P_{m-2}\}}_{P_m}. \end{aligned}$$

Ou seja,

$$\begin{aligned} P'_{m+1} &= -(1 + mV'' + b_{m-1})P_m = -\{1 + 2v_2m + m[1 + (m-1)v_2]\}P_m \\ &= -[1 + m + (1 + m)mv_2]P_m = -(1 + m)(1 + mv_2)P_m. \end{aligned}$$

Tem-se que

$$P'_{m+1} = -b_mP_m. \quad (2.35)$$

A indução sobre m prova que $P'_m = -b_{m-1}P_{m-1}$. Iterando (2.34) $(r-1)$ vezes obtém-se

$$P_{m+1}^{(r)} = (-1)^r (b_{m-1} \dots b_{m-r})P_{m-r} = (-1)^r \frac{a_m}{a_{m-r}}P_{m-r},$$

o que prova (2.32). A equação (2.30) resulta de (2.29) e (2.32) com $r = 1$. De facto, de (2.32) com $r = 1$ obtém-se

$$P_m^{(1)} = -\frac{a_m}{a_{m-1}}P_{m-1}$$

e de (2.29)

$$\begin{aligned} P_{m+1} &= (P_1 - mV')P_m + vP'_m = (P_1 - mV')P_m - \frac{a_m}{a_{m-1}}VP_{m-1} \\ &= (P_1 - mV')P_m - b_{m-1}VP_{m-1} \\ &= (P_1 - mV')P_m - m(1 + (m-1)v_2)P_{m-1}. \end{aligned}$$

Que $P_m(x, \mu)$ é um polinómio de grau exacto igual a m em x e μ com termo director x^m é imediato de (2.30).

Para $n < m$ tem-se

$$\begin{aligned} \mathbb{E}_\theta [X^n P_m(X, \mu)] &= \int_{\mathbb{R}} x^n P_m(X, \mu) f(x, \theta) dx = \int_{\mathbb{R}} x^n \frac{V^m(\mu)}{f(x, \theta)} f(x, \theta) \frac{d^m f(x, \theta)}{d\mu^m} dx \\ &= V^m(\mu) \int_{\mathbb{R}} x^n \frac{d^m f(x, \theta)}{d\mu^m} dx = V^m(\mu) \frac{d^m}{d\mu^m} \int_{\mathbb{R}} x^n f(x, \theta) dx \\ &= V^m(\mu) \frac{d^m}{d\mu^m} E_\theta(X^n) = 0, \end{aligned}$$

pois $\mathbb{E}_\theta(X^n)$ é um polinómio de grau quando muito igual a n em $\mu = \psi'(\theta)$. Logo para $m > n$, $\frac{d^m}{d\mu^m} \mathbb{E}_\theta(X^n) = 0$. Então P_m é ortogonal a todos os polinómios de grau inferior a m .

Para provar a igualdade anterior para $m = n \geq 1$, consideremos a igualdade (2.30). Multiplicando-a por P_{m-1} e tomando médias obtemos

$$P_{m+1}P_{m-1} = (P_1 - mV')P_mP_{m-1} - m[1 + (m-1)v_2]VP_{m-1}^2$$

e,

$$\begin{aligned} 0 &= \mathbb{E}(P_{m+1}P_{m-1}) = \mathbb{E}(P_1P_mP_{m-1}) - mV' \mathbb{E}(P_mP_{m-1}) - b_{m-1}V \mathbb{E}(P_{m-1}^2) \\ &\Leftrightarrow \mathbb{E}(P_1P_mP_{m-1}) = b_{m-1}V \mathbb{E}(P_{m-1}^2). \end{aligned}$$

Como

$$P_m = [P_1 - (m-1)V']P_{m-1} - (m-1)[1 + (m-2)v_2]VP_{m-2},$$

multiplicando por P_m e calculando médias obtém-se

$$\mathbb{E}(P_m^2) = \mathbb{E}(P_1P_mP_{m-1}).$$

Tem-se então que

$$\mathbb{E}(P_m^2) = \mathbb{E}(P_1P_mP_{m-1}) = b_{m-1}V \mathbb{E}(P_{m-1}^2),$$

isto é,

$$\mathbb{E}(P_m^2) = b_{m-1} V \mathbb{E}(P_{m-1}^2) .$$

Iterando obtém-se

$$\mathbb{E}(P_m^2) = b_{m-1} V b_{m-2} V \mathbb{E}(P_{m-2}^2) = b_{m-1} b_{m-2} \dots b_0 V^m \mathbb{E}(P_0^2) .$$

Mas $P_0 = 1 \Rightarrow \mathbb{E}(P_0^2) = 1$, ou seja,

$$\mathbb{E}(P_m^2) = \frac{a_m}{a_{m-1}} \frac{a_{m-1}}{a_{m-2}} \dots \frac{a_1}{a_0} V^m \Leftrightarrow \mathbb{E}(P_m^2) = a_m V^m .$$

Resulta assim que

$$\mathbb{E}_\theta(P_m P_n) = \delta_{mn} a_m V^m .$$

□

Os polinómios anteriores são conhecidos individualmente por polinómios de Hermite se a distribuição associada é a gaussiana, Poisson-Charlier se a distribuição associada é a Poisson, Laguerre generalizados se a distribuição associada é a gama, Krawtchouk se a distribuição associada é a binomial, Meixner se a distribuição associada é a binomial negativa e Pollaczek se a distribuição associada é a secante hiperbólica generalizada (GHS) (mas apenas para subfamílias simétricas). Note-se que os resultados (2.30), (2.31), (2.32) e (2.33) são válidos para membros não simétricos da família NEF-GHS.

O sistema $P_m(x, \mu)$, $m = 0, 1, \dots$ forma um conjunto de polinómios de grau m para qualquer família exponencial natural. No entanto, estes polinómios formam um sistema ortogonal apenas se a propriedade QVF se verificar.

Corolário 2.4.1.1. *Sejam $\mu, \mu_0 \in \Omega$ os valores médios dados de duas distribuições na mesma NEF-QVF. Então para qualquer $m = 0, 1, \dots$*

$$P_m(x, \mu_0) = a_m \sum_{r=0}^m \frac{(\mu - \mu_0)^{m-r}}{(m-r)!} \frac{1}{a_r} P_r(x, \mu), \quad (2.36)$$

com $a_m = m! \prod_{i=0}^{m-1} (1 + iv_2)$, tendo-se

$$\mathbb{E}_\mu [P_m(X, \mu_0)] = \frac{a_m}{m!} (\mu - \mu_0)^m . \quad (2.37)$$

Demonstração

Expandindo $P_m(x, \mu)$ em série de Taylor em torno de μ_0 obtém-se

$$P_m(x, \mu) = P_m(x, \mu_0) + P'_m(x, \mu_0)(\mu - \mu_0) + \frac{P''_m(x, \mu_0)}{2!} (\mu - \mu_0)^2 + \dots + \frac{P_m^{(m)}(x, \mu_0)}{m!} (\mu - \mu_0)^m ,$$

ou seja,

$$P_m(x, \mu) = \sum_{r=0}^m \frac{(\mu - \mu_0)^r}{r!} P_m^{(r)}(x, \mu_0).$$

Mas,

$$P_m^{(r)}(x, \mu) = (-1)^r \frac{a_m}{a_{m-r}} P_{m-r}(x, \mu_0)$$

então,

$$P_m(x, \mu) = \sum_{r=0}^m \frac{(\mu - \mu_0)^r}{r!} (-1)^r \frac{a_m}{a_{m-r}} P_{m-r}(x, \mu_0).$$

Se se trocar os papéis de r e $m - r$ tem-se

$$P_m(x, \mu) = a_m \sum_{r=0}^m \frac{(\mu - \mu_0)^{m-r}}{(m-r)!} \frac{1}{a_r} (-1)^{m-r} P_r(x, \mu_0).$$

Tem-se ainda que

$$\mathbb{E}_\mu (P_m(X, \mu)) = a_m \sum_{r=0}^m \frac{(\mu - \mu_0)^{m-r}}{(m-r)!} \frac{1}{a_r} \mathbb{E} (P_r(X, \mu_0) (-1)^{m-r}),$$

mas $\mathbb{E} (P_r(X, \mu_0)) = \delta_{r,0}$, então

$$\mathbb{E}_\mu (P_m(X, \mu)) = \frac{a_m}{m!} (\mu_0 - \mu)^m. \quad (2) \quad \square$$

2.5 Expansões Assintóticas

Seja $Z \sim \text{Gaussiana}(0, 1)$. A probabilidade da cauda, integrando por partes — usando em ambos os casos $g' = t \exp\left(-\frac{t^2}{2}\right)$, e usando $f = \frac{1}{t}$ no segundo caso — é dada por

$$\begin{aligned} \sqrt{2\pi} \mathbb{P}[Z > x] &= \int_x^\infty \exp\left(-\frac{t^2}{2}\right) dt = \frac{1}{x} \exp\left(-\frac{x^2}{2}\right) - \int_x^\infty \frac{1}{t^2} \exp\left(-\frac{t^2}{2}\right) dt = \\ &= \frac{1}{x} \exp\left(-\frac{x^2}{2}\right) - \frac{1}{x^3} \exp\left(-\frac{x^2}{2}\right) + \int_x^\infty \frac{3}{t^4} \exp\left(-\frac{t^2}{2}\right) dt. \end{aligned}$$

⁽²⁾Ao contrário do que Morris afirma na p. 77 do seu artigo de 1983, onde afirma que $\mathbb{E}_\mu (P_m(X, \mu)) = \frac{a_m}{m!} (\mu - \mu_0)^m$.

Obtém-se assim o enquadramento da cauda gaussiana

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \phi(x) \leq 1 - \Phi(x) \leq \frac{1}{x} \phi(x), \quad (x > 0)$$

que é bem conhecido, e base de desenvolvimento teóricos notáveis. Aquele enquadramento decorre da expansão de $\Phi(x)$ em termos de derivadas da função densidade de probabilidade gaussiana, ou, mais propriamente, usando polinómios de Hermite. Foi decerto uma fonte de inspiração para Gram, para Charlier e para Edgeworth.

Nesta secção apresentamos a teoria das expansões de Edgeworth (1905), que nos permitem enquadrar devidamente a investigação da velocidade de convergência no Teorema Limite Central. As expansões de Edgeworth correspondem a um progresso relativamente às séries de Gram-Charlier (1883), que foram posteriormente sujeitas a desenvolvimentos interessantes nas mãos de Cramér (1925) de Fisher e Cornish (1937).

Cramér, em particular, usando a ideia de “distribuições conjugadas” — que reporta a anteriores trabalhos de Esscheer (1932) — apresentou um desenvolvimento notável, por vezes referido com “expansões de Edgeworth diferidas”, que se verá serem correspondentes ao uso dos pontos sela por Daniels (1956).

Fisher e Cornish apresentam uma variante do problema das aproximações com expansões em série, que naturalmente ganhou a denominação de “expansões de Cornish-Fisher”.

A teoria, sendo entretanto mais geral, adapta-se perfeitamente ao tratamento de elementos da família exponencial.

2.5.1 Expansões de Edgeworth

Nesta secção as expansões de Edgeworth serão utilizadas como aproximações para as distribuições de estimadores de parâmetros desconhecidos.

Se $\hat{\theta}$ é um estimador construído a partir de uma amostra de tamanho n e se $\sqrt{n}(\hat{\theta} - \theta_0)$ tem distribuição assintótica gaussiana de média zero e variância σ^2 , então em muitos casos de interesse prático $\sqrt{n}(\hat{\theta} - \theta_0)$ pode ser expandido em série de potências em $n^{-1/2}$,

$$\mathbb{P}\left(\frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sigma} \leq x\right) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + \dots + n^{-j/2}p_j(x)\phi(x) + \dots \quad (2.38)$$

onde $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ é a função densidade de probabilidade da gaussiana padrão, $\Phi(x) = \int_{-\infty}^x \phi(u)du$ a respectiva função distribuição e p_j são polinómios com coeficientes dependendo dos cumulantes de $\hat{\theta} - \theta_0$. Na igualdade (2.38) vamos considerar

$n^{-1/2}p_1(x)\phi(x)$ como o primeiro termo da aproximação de $\mathbb{P}\left(\frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sigma} \leq x\right)$ pela distribuição Φ .

Invertendo a equação (2.38) é possível mostrar que a solução $x = u_\alpha$ da equação

$$\mathbb{P}\left(\frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sigma} \leq x\right) = \alpha$$

admite a expansão

$$u_\alpha = z_\alpha + n^{-1/2}p_{11}(z_\alpha) + n^{-1}p_{21}(z_\alpha) + \dots + n^{-j/2}p_{j1}(z_\alpha) + \dots, \quad (2.39)$$

onde p_{j1} são polinómios resultantes de p_j por derivação e z_α é solução da equação $\Phi(z_\alpha) = \alpha$. As expansões inversas do tipo (2.39) são denotadas por expansões de Cornish-Fischer, como veremos mais à frente com maior detalhe.

Em muitos casos de interesse a variância assintótica σ^2 é desconhecida e tem de ser estimada por uma função, $\hat{\sigma}^2$, dos dados. Neste caso a função de distribuição de interesse é a da variável studentizada $\frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\hat{\sigma}}$. Também neste caso é possível efectuar expansões do tipo (2.38) mas com diferentes polinómios p_j .

Sejam X_1, X_2, \dots , variáveis aleatórias independentes e identicamente distribuídas com média $\theta_0 = \mu$ e variância finita σ^2 . Um estimador centrado de θ_0 é

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i$$

com variância $\frac{1}{n}\sigma^2$. Aplicando o teorema limite central tem-se que

$$S_n = \frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sigma} \overset{\circ}{\sim} N(0, 1).$$

Uma maneira simples de descrever os erros na aproximação à gaussiana é através das funções características. Como S_n é assintoticamente gaussiana padrão a função característica φ_n , de S_n , converge para a função característica da gaussiana padrão, quando $n \rightarrow \infty$,

$$\varphi_n(t) = \mathbb{E}(e^{itS_n}) \xrightarrow{n \rightarrow \infty} \mathbb{E}(e^{itZ}) = e^{-t^2/2} \quad \text{com } Z \sim N(0, 1) \quad \text{e } t \in \mathbb{R}. \quad (2.40)$$

Mas,

$$\varphi_n(t) = \mathbb{E}(e^{itS_n}) = (\varphi(t/\sqrt{n}))^n, \quad (2.41)$$

com φ a função característica da variável aleatória $Y = \frac{X - \mu}{\sigma}$.

Se se padronizar Y para localização e escala tal que $\mathbb{E}(Y) = k_1 = 0$ e $\text{Var}(Y) = k_2 = 1$. Então por (2.41) e usando o facto de

$$\varphi(t) = \exp \left\{ k_1 it + \frac{1}{2} k_2 (it)^2 + \cdots + \frac{1}{j!} k_j (it)^j + \cdots \right\},$$

fazendo uma expansão em série da função exponencial obtém-se

$$\varphi_n(t) = \exp \left\{ \left[k_1 \frac{it}{\sqrt{n}} + \frac{(it)^2}{2(\sqrt{n})^2} k_2 + \frac{(it)^3}{3!(\sqrt{n})^3} k_3 + \cdots + \frac{(it)^j}{j!(\sqrt{n})^j} k_j + \cdots \right] \times n \right\}.$$

onde k_i é um cumulante de Y . Como $k_1 = 0$ e $k_2 = 1$ tem-se

$$\begin{aligned} \varphi_n(t) &= \exp \left\{ \frac{-t^2}{2} + \frac{(it)^3}{3!\sqrt{n}} k_3 + \cdots + n^{-(j-2)/2} \frac{(it)^j}{j!} k_j + \cdots \right\} \\ &= \exp \left\{ \frac{-t^2}{2} \right\} \exp \left\{ \frac{(it)^3 k_3}{3!\sqrt{n}} \right\} \exp \left\{ \frac{(it)^4 k_4}{4!n} \right\} \cdots \exp \left\{ \frac{(it)^j k_j}{j!n^{-(j-2)/2}} \right\} \cdots \end{aligned}$$

Fazendo uma expansão em série para cada um dos factores anteriores, excepto para o primeiro, multiplicado depois todos os desenvolvimentos obtém-se

$$\varphi_n(t) = e^{-\frac{1}{2}t^2} \left\{ 1 + \frac{1}{\sqrt{n}} r_1(it) + \frac{1}{n} r_2(it) + \cdots + \frac{1}{\sqrt{n}^j} r_j(it) + \cdots \right\}, \quad (2.42)$$

onde r_j é um polinómio de coeficientes reais, de grau $3j$, dependendo de k_3, \dots, k_{j+2} mas não de n .

O polinómio r_j é um polinómio de potências pares quando j é par, e de potências ímpares quando j é ímpar. De (2.42) tem-se

$$r_1(u) = \frac{1}{6} k_3 u^3 \quad (2.43)$$

e

$$r_2(u) = \frac{1}{24} k_4 u^4 + \frac{1}{72} k_3^2 u^6 \quad (2.44)$$

Se se reescrever (2.42) do seguinte modo

$$\varphi_n(t) = e^{-\frac{t^2}{2}} + n^{-1/2} r_1(it) e^{-\frac{t^2}{2}} + n^{-1} r_2(it) e^{-\frac{t^2}{2}} + \cdots + n^{-j/2} r_j(it) e^{-\frac{t^2}{2}} + \cdots \quad (2.45)$$

Como

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx} d\mathbb{P}(S_n \leq x)$$

e

$$e^{-\frac{t^2}{2}} = \int_{-\infty}^{\infty} e^{itx} d\Phi(x), \quad (2.46)$$

onde Φ é a função de distribuição da gaussiana padrão, tem-se que a função característica de S_n converge para a função característica da gaussiana padrão. Então pelo teorema da convergência da função característica tem-se que a função de distribuição de S_n converge para a função distribuição da gaussiana padrão, ou seja, (2.45) sugere a expansão inversa

$$\mathbb{P}(S_n \leq x) = \Phi(x) + n^{-1/2}R_1(x) + n^{-1}R_2(x) + \cdots + n^{-j/2}R_j(x) + \cdots, \quad (2.47)$$

onde $R_j(x) = r_j(it) e^{-\frac{t^2}{2}}$ denota a função cuja transformada de Fourier-Stieltjes é igual a $r_j(it) e^{-\frac{t^2}{2}}$, isto é,

$$\int_{-\infty}^{\infty} e^{itx} dR_j(x) = r_j(it) e^{-\frac{t^2}{2}}.$$

Se repetidamente se integrar por partes a fórmula (2.46) obtém-se

$$e^{-\frac{t^2}{2}} = (-it)^{-j} \int_{-\infty}^{\infty} e^{itx} d\Phi^{(j)}(x),$$

onde $\Phi^{(j)}(x) = \frac{d^j \Phi(x)}{dx^j}$.

Então,

$$\int_{-\infty}^{\infty} e^{itx} d\{r_j(-D)\Phi(x)\} = (it)^j e^{-\frac{t^2}{2}}, \quad (2.48)$$

onde D é o operador diferencial $\frac{d}{dx}$. Interpretando $r_j(-D)$ como um polinómio em D , então $r_j(-D)$ é ele próprio um operador diferencial. Por (2.48)

$$\int_{-\infty}^{\infty} e^{itx} d\{r_j(-D)\Phi(x)\} = r_j(it) e^{-\frac{t^2}{2}},$$

e a solução procurada é $r_j(-D\Phi(x))$, isto é,

$$R_j(x) = r_j\left(-\frac{d}{dx}\right)\Phi(x). \quad (2.49)$$

Para $j \geq 1$

$$(-D)^j \Phi(x) = -H_{e_{j-1}}(x)\phi(x),$$

onde $H_{e_{j-1}}$ são polinómios de Hermite.

Estes polinómios são ortogonais com respeito à função peso ϕ e são normalizados de modo que o coeficiente da maior potência em x é um. Note-se que $H_{e_j}(x)$ é de precisamente grau j e, conseqüentemente, é ímpar se j é ímpar, e par se j é par. Utilizando (2.43), (2.44) e (2.49) tem-se

$$R_1(x) = -\frac{1}{6}k_3(x^2 - 1)\phi(x)$$

e

$$R_2(x) = -x \left\{ \frac{1}{24}k_4(x^2 - 3) + \frac{1}{72}k_3^2(x^5 - 10x^3 + 15x) \right\} \phi(x).$$

De modo geral,

$$R_j(x) = p_j(x)\phi(x)$$

onde p_j é um polinómio de grau $(3j-1)$ par se j ímpar e ímpar se j par, tal é consequência do facto de r_j ser de grau $3j$ par se j par, e ímpar se j ímpar. Tem-se que

$$p_1(x) = -\frac{1}{6}k_3(x^2 - 1) \quad (2.50)$$

e

$$p_2(x) = -x \left\{ \frac{1}{24}k_4(x^2 - 3) + \frac{1}{72}k_3^2(x^4 - 10x^2 + 15) \right\}. \quad (2.51)$$

A fórmula (2.47) assume agora a forma

$$\mathbb{P}(S_n \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + \dots + n^{-j/2}p_j(x)\phi(x) + \dots \quad (2.52)$$

a que se chama expansão de Edgeworth para a função de distribuição $\mathbb{P}(S_n \leq x)$.

Os cumulantes k_3 e k_4 a assimetria e a curtose, respectivamente.

O termo de ordem de ordem $n^{-1/2}$ em (2.52) corrige, na aproximação “básica” feita à distribuição gaussiana, o principal efeito da assimetria enquanto o termo de ordem n^{-1} corrige o efeito principal da curtose e o efeito secundário da assimetria.

Exemplo

Sejam X_1, \dots, X_n variáveis aleatórias com distribuição exponencial de média 1. A função densidade exacta de $S_n^* = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}}$ é dada por

$$f_{S_n^*}(x) = \frac{\sqrt{n}(n + x\sqrt{n})^{n-1} \exp(-n - x\sqrt{n})}{(n-1)!}.$$

Tem-se que $\mathbb{E}(S_n) = n$, $\text{Var}(S_n) = n$, $\rho_3 = 2$ e $\rho_4 = 6$, onde ρ_i são os cumulantes padronizados. Então a expansão de Edgeworth para a função probabilidade de S_n^* é:

$$f_{S_n^*}(x) = \phi(x) \left\{ 1 + \frac{H_3(x)}{3\sqrt{n}} + \frac{H_4(x)}{3n} + \frac{H_6(x)}{18n} \right\} + O(n^{-3/2}),$$

onde os H_i são os polinómios de Hermite de grau 3, 4 e 5.

Na tabela seguinte compara-se, para $n = 5$, o valor exacto $f_{S_n^*}(x)$ com a aproximação dada pela gaussiana e com a expansão de Edgeworth obtida anteriormente. Considerou-se, ainda a expansão com o termo de ordem $O(n^{-1/2})$ e de seguida como os termos de ordem $O(n^{-1})$.

x	Valor exacto	Aprox. Gaussiana	$O(n^{-1/2})$	$O(n^{-1})$
-2	0.0043	0.0540	0.0379	0.0178
-1.5	0.1319	0.1295	0.1512	0.1480
-1.0	0.3428	0.2420	0.3141	0.3329
-0.5	0.4361	0.3521	0.4242	0.4335
0	0.3924	0.3989	0.3989	0.3922
1	0.1840	0.2420	0.1698	0.1887
2	0.0577	0.0540	0.0701	0.0500
3	0.0144	0.0044	0.0163	0.0181

Da tabela anterior, observa-se que para valores pequenos de n a expansão não funciona bem nas caudas da distribuição de S_n^* . Mas fora das caudas, a expansão funciona muito bem. Note-se que a expansão incluindo apenas os termos de $O(n^{-1})$ é superior à expansão com termos da ordem $O(n^{-1/2})$. ■

A expansão (2.52) raramente converge como série infinita. Se X tem uma distribuição absolutamente contínua, o tipo de condições que têm de ser impostas para garantir a convergência são da forma $\mathbb{E}(\exp(\frac{1}{4}X^2)) < \infty$, (Cramér, 1928), o que representa uma restrição severa nas caudas da distribuição de X . Normalmente (2.52) está apenas disponível como série assintótica

$$\mathbb{P}(S_n \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + \dots + n^{-j/2}p_j(x)\phi(x) + O(n^{-j/2}), \quad (2.53)$$

válida para j fixo e $n \rightarrow \infty$.

Condições suficientes de regularidade para (2.53), como o resto de uma ordem fixa uniformemente em todo o x , são

$$\mathbb{E}(|X|^{j+2}) < \infty \quad \text{e} \quad \limsup_{|t| \rightarrow \infty} |\varphi(t)| < 1.$$

A última restrição é chamada condição de Cramér, (cf. Hall, 1992, pag. 45), que se verifica quando a distribuição de X é não singular, ou equivalentemente, se a distribuição tem uma componente não degenerada absolutamente contínua — em particular, se X tem uma função de densidade “própria”.

Assim, a inversão formal da expansão da função característica (2.45) é válida se X tem um número suficientemente grande de momentos finitos e uma distribuição regular, e falha se a distribuição de X é singular.

2.5.2 Expansões de Edgeworth Diferidas

Sejam X_1, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade f_X , $\mathbb{E}(X) = \mu$, $\text{var}(X) = \sigma^2$, função geradora de

momentos M_X e função geradora de cumulantes $\psi_X = K_X = \log(M_X)$. Considere-se a família associada a f_X , no sentido de Cramér-Khintchine

$$f_{X_\lambda}(y) = f(y; \lambda) = \frac{e^{\lambda y} f_X(y)}{M_X(\lambda)} = e^{(\lambda y - K_X(\lambda))} f_X(y),$$

com $\lambda \in \mathbb{R}$. Muitas vezes este procedimento é referido com “exponential tilting” e gera uma família de distribuições associadas (cf. Esscher (1932), Cramér (1963), Khintchine(1960)).

Por exemplo, se $X \sim \text{Gaussiana}(\mu, \sigma)$ então $X_\lambda \sim \text{Gaussiana}(\mu + \lambda + \sigma^2, \sigma)$.

É imediato que, tomando como habitualmente $S_n = \sum_{i=1}^n X_i$ e $S_{n,\lambda} = \sum_{i=1}^n X_{\lambda_i}$ onde X_λ tem distribuição f_{X_λ} se tem

$$\begin{aligned} f_{S_{n,\lambda}}(x) &= f_{S_n}(x; \lambda) = e^{(\lambda x - nK_X(\lambda))} f_{S_n}(x; 0) \\ \Rightarrow f_{S_n}(x; 0) &= f_{S_n}(x) = e^{(-\lambda x + nK_X(\lambda))} f_{S_{n,\lambda}}(x). \end{aligned} \quad (2.54)$$

Escolhendo $\lambda = \hat{\lambda}$ tal que $\mathbb{E}(S_{n,\hat{\lambda}}) = x$

$$\begin{aligned} \mathbb{E}(S_{n,\hat{\lambda}}) &= n\mathbb{E}(X_{\hat{\lambda}}) = n \int_{\mathbb{R}} y \frac{e^{\hat{\lambda}y} f_X(y)}{M_X(\hat{\lambda})} dy \\ &= \frac{n}{M_X(\hat{\lambda})} \int_{\mathbb{R}} ye^{\hat{\lambda}y} f_X(y) dy = n \frac{M'_X(\hat{\lambda})}{M_X(\hat{\lambda})} \\ &= n \frac{\partial}{\partial \hat{\lambda}} \log(M_X(\hat{\lambda})) = nK'_X(\hat{\lambda}) = x \\ \Leftrightarrow K'_X(\hat{\lambda}) &= \frac{x}{n}. \end{aligned}$$

Relativamente à variância de $S_{n,\hat{\lambda}}$, tem-se

$$\text{Var}(S_{n,\hat{\lambda}}) = n\text{Var}(X_{\hat{\lambda}}) = nK''_X(\hat{\lambda}).$$

A soma padronizada é

$$S_{n,\hat{\lambda}}^* = \frac{S_{n,\hat{\lambda}} - \mathbb{E}(S_{n,\hat{\lambda}})}{\sqrt{\text{Var}(S_{n,\hat{\lambda}})}} = \frac{S_{n,\hat{\lambda}} - x}{\sqrt{nK''_X(\hat{\lambda})}},$$

donde nos pontos de continuidade da função $f_{S_{n,\hat{\lambda}}}$

$$\begin{aligned} f_{S_{n,\hat{\lambda}}}(y) &= \frac{d}{dy} F_{S_{n,\hat{\lambda}}}(y) = \frac{d}{dy} \mathbb{P} \left(S_{n,\hat{\lambda}} \leq y \right) \\ &= \frac{d}{dy} \mathbb{P} \left(\frac{S_{n,\hat{\lambda}} - \mathbb{E}(S_{n,\hat{\lambda}})}{\sqrt{\text{Var}(S_{n,\hat{\lambda}})}} \leq \frac{y - x}{\sqrt{nK_X''(\hat{\lambda})}} \right) \\ &= \frac{d}{dy} F_{S_{n,\hat{\lambda}}}^* \left(\frac{y - x}{\sqrt{nK_X''(\hat{\lambda})}} \right) \\ &= \frac{1}{\sqrt{nK_X''(\hat{\lambda})}} f_{S_{n,\hat{\lambda}}}^* \left(\frac{y - x}{\sqrt{nK_X''(\hat{\lambda})}} \right). \end{aligned}$$

Pela expansão de Edgeworth sabemos que

$$f_{S_{n,\hat{\lambda}}}^*(y) = \phi(y) \left\{ 1 + \frac{\rho_3}{6\sqrt{n}} H_3(y) + \frac{\rho_4}{24n} H_4(y) + \frac{\rho_3^2}{72n} H_5(y) \right\} + O(n^{-3/2})$$

onde ρ_i é o i -ésimo cumulante padronizado de X e os H_i ($i = 3, 4, 5$) são polinómios de Hermite. Temos então que

$$f_{S_{n,\hat{\lambda}}}^*(y) = \phi(y) \left\{ 1 + \frac{\rho_3}{6\sqrt{n}} H_3(y) \right\} + O(n^{-1})$$

e, conseqüentemente,

$$\begin{aligned} f_{S_{n,\hat{\lambda}}}(y) &= \frac{1}{\sqrt{nK_X''(\hat{\lambda})}} f_{S_{n,\hat{\lambda}}}^* \left(\frac{y - x}{\sqrt{nK_X''(\hat{\lambda})}} \right) \\ f_{S_{n,\hat{\lambda}}}(y) &= \frac{1}{\sqrt{nK_X''(\hat{\lambda})}} \left[\phi \left(\frac{y - x}{\sqrt{nK_X''(\hat{\lambda})}} \right) \left\{ 1 + \frac{\rho_3}{6\sqrt{n}} H_3 \left(\frac{y - x}{\sqrt{nK_X''(\hat{\lambda})}} \right) \right\} + O(n^{-1}) \right]. \end{aligned}$$

Para $y = x$ temos

$$f_{S_{n,\hat{\lambda}}}(x) = \frac{1}{\sqrt{2nK_X''(\hat{\lambda})}} [1 + O(n^{-1})].$$

De (2.54) resulta então

$$\begin{aligned} f_{S_n}(x) &= \exp \left(-\hat{\lambda}x + nK_X(\hat{\lambda}) \right) f_{S_n}(x; \hat{\lambda}) \\ &= \exp \left(-\hat{\lambda}x + nK_X(\hat{\lambda}) \right) \frac{1}{\sqrt{2nK_X''(\hat{\lambda})}} [1 + O(n^{-1})]. \end{aligned} \quad (2.55)$$

Assim, usando criteriosamente “tilting” e “back-tilting”, pode-se melhorar de forma não-trivial a qualidade da aproximação em zonas por vezes apodadas de “grandes desvios”.

Exemplo

Consideremos a variável aleatória X com distribuição de Lévy (estável de parâmetros ($\alpha = \frac{1}{2}$ e $\beta = 1$)) com função densidade de probabilidade

$$f_X(y) = \frac{\exp\left(-\frac{1}{2y}\right)}{\sqrt{2\pi y^3}} I_{(0,\infty)}(y).$$

Como é estável

$$f_{\frac{S_n}{n^2}}(x) = \frac{\exp\left(-\frac{1}{2x}\right)}{\sqrt{2\pi x^3}} I_{(0,\infty)}(x)$$

e

$$f_{S_n}(x) = f_{\frac{S_n}{n^2}}\left(\frac{x}{n^2}\right) \frac{1}{n^2} = n \frac{\exp\left(-\frac{n^2}{2x}\right)}{\sqrt{2\pi x^3}} I_{(0,\infty)}(x). \quad (2.56)$$

Por outro lado $K_X(\lambda) = -\sqrt{-2\lambda}$, definido para $\lambda < 0$. Tem-se que

$$K'_X(\lambda) = \frac{1}{\sqrt{-2\lambda}} \quad \text{e} \quad K''_X(\lambda) = \frac{-1}{2\lambda\sqrt{-2\lambda}}.$$

Donde,

$$K'_X(\hat{\lambda}) = \frac{x}{n} \Leftrightarrow \hat{\lambda} = -\frac{1}{2} \frac{n^2}{x^2}, \quad K''_X(\hat{\lambda}) = \frac{x^3}{n^3} \quad \text{e} \quad K_X(\hat{\lambda}) = -\frac{n}{x}, \quad \text{para } x > 0.$$

Substituindo em (2.55)

$$\begin{aligned} f_{S_n}(x) &= \exp\left(-\hat{\lambda}x + nK_X(\hat{\lambda})\right) \frac{1}{\sqrt{2\pi K''_X(\hat{\lambda})}} [1 + O(n^{-1})] \\ &\Leftrightarrow f_{S_n}(x) = \exp\left(\frac{1}{2} \frac{n^2}{x^2} x - n \frac{n}{x}\right) \frac{1}{\sqrt{2\pi n \frac{x^3}{n^3}}} [1 + O(n^{-1})] \\ &\Leftrightarrow f_{S_n}(x) = n \frac{\exp\left(-\frac{1}{2} \frac{n^2}{x}\right)}{\sqrt{2\pi x^3}} [1 + O(n^{-1})]. \end{aligned}$$

Compare-se este último resultado com (2.56). Merece destaque o facto de a distribuição de Lévy não ter momentos inteiros e, curiosamente, a aproximação dada por (2.55) funcionar bem. ■

2.5.3 Expansões Assintóticas de Distribuições Estáveis

Já vimos na secção anterior que se X, X_1, \dots, X_n são variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade f_X , $\mathbb{E}(X) = \mu$, $\text{var}(X) = \sigma^2$, função geradora de momentos M_X e função geradora de cumulantes $\psi_X = K_X = \log(M_X)$. A família exponencial associada a f_X é

$$f(y; \lambda) = \frac{e^{\lambda y} f_Y(y)}{M_X(\lambda)} = e^{(\lambda y - K_X(\lambda))} f_X(y),$$

com $\lambda \in \mathbb{R}$. E que

$$f_{S_n}(x) = \exp\left(-\hat{\lambda}x + nK_X(\hat{\lambda})\right) \frac{1}{\sqrt{2\pi K_X''(\hat{\lambda})}} [1 + O(n^{-1})], \quad (2.57)$$

onde $\hat{\lambda}$ é solução da equação $K_X'(\hat{\lambda}) = \frac{x}{n}$.

Suponhamos agora, que S_n resulta da soma de n variáveis aleatórias independentes e identicamente distribuídas estáveis. Ou seja, nesta secção X, X_1, X_2, \dots denotam variáveis aleatórias mutuamente independentes com distribuição comum F estável e $S_n = X_1 + \dots + X_n$.

Uma variável aleatória X é estável se a sua função característica φ é tal que

$$\ln(\varphi(t)) = i\gamma t - c|t|^\alpha \left[1 + i\beta \frac{t}{|t|} w(t, \alpha)\right],$$

onde $0 < \alpha \leq 2$, é o expoente característico da distribuição estável, β, γ, c são constantes reais, sendo $c \leq 0$ parâmetro de escala, $|\beta| \leq 1$ parâmetro de assimetria e

$$w(t, \alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \text{se } \alpha \neq 1 \\ \frac{2}{\pi} \log|t| & \text{se } \alpha = 1 \end{cases}.$$

De $K_X(t) = \ln(\varphi(-it))$ vem

$$K_X(t) = \gamma t - c|t|^\alpha \left[1 + \beta \frac{t}{|t|} w(it, \alpha)\right] = \begin{cases} \gamma t - c|t|^\alpha \left(1 + \beta \frac{t}{|t|} \tan\left(\frac{\pi\alpha}{2}\right)\right) & \text{se } \alpha \neq 1 \\ \gamma t - c|t|^\alpha \left(1 + \beta \frac{t}{|t|} \frac{2}{\pi} \log|t|\right) & \text{se } \alpha = 1 \end{cases}.$$

Vamos calcular (2.57) para $\alpha \neq 1$, supondo $\hat{\lambda} > 0$ e $\hat{\lambda} < 0$, e para $\alpha = 1$.

- Para $\alpha \neq 1$ temos

$$K_X(t) = \gamma t - c|t|^\alpha \left(1 + \beta \frac{t}{|t|} \tan\left(\frac{\pi\alpha}{2}\right)\right),$$

pelo que, para $t \neq 0$,

$$K'_X(t) = \gamma - c\alpha \frac{t}{|t|} |t|^{\alpha-1} \left(1 + \frac{t}{|t|} \beta \tan \left(\frac{\pi\alpha}{2} \right) \right)$$

e

$$K''_X(t) = -c\alpha(\alpha-1) |t|^{\alpha-2} \left(1 + \frac{t}{|t|} \beta \tan \left(\frac{\pi\alpha}{2} \right) \right).$$

De $\hat{\lambda}$ ser a solução de $K'_X(\hat{\lambda}) = \frac{x}{n}$ temos,

$$\begin{aligned} \gamma - c\alpha \frac{\hat{\lambda}}{|\hat{\lambda}|} \left(1 + \frac{\hat{\lambda}}{|\hat{\lambda}|} \beta \tan \left(\frac{\pi\alpha}{2} \right) \right) &= \frac{x}{n} \Rightarrow \\ 1 + \frac{\hat{\lambda}}{|\hat{\lambda}|} \beta \tan \left(\frac{\pi\alpha}{2} \right) &= \frac{\gamma - \frac{x}{n}}{c\alpha \frac{\hat{\lambda}}{|\hat{\lambda}|} |\hat{\lambda}|^{\alpha-1}} = \frac{\gamma - \frac{x}{n}}{c\alpha \hat{\lambda} |\hat{\lambda}|^{\alpha-2}} \Rightarrow \\ |\hat{\lambda}| &= \left[\frac{\hat{\lambda}}{|\hat{\lambda}|} \frac{1}{c\alpha} \frac{\gamma - \frac{x}{n}}{1 + \frac{\hat{\lambda}}{|\hat{\lambda}|} \beta \tan \left(\frac{\pi\alpha}{2} \right)} \right]^{\frac{1}{\alpha-1}} \Rightarrow \begin{cases} \hat{\lambda} = \left[\frac{1}{c\alpha} \frac{\gamma - \frac{x}{n}}{1 + \beta \tan \left(\frac{\pi\alpha}{2} \right)} \right]^{\frac{1}{\alpha-1}} & \text{se } \hat{\lambda} > 0 \\ \hat{\lambda} = \left[\frac{1}{c\alpha} \frac{\gamma - \frac{x}{n}}{1 - \beta \tan \left(\frac{\pi\alpha}{2} \right)} \right]^{\frac{1}{\alpha-1}} & \text{se } \hat{\lambda} < 0 \end{cases} \end{aligned}$$

pelo que

$$\begin{aligned} K''_X(\hat{\lambda}) &= -c\alpha(\alpha-1) \frac{\hat{\lambda}}{|\hat{\lambda}|} |\hat{\lambda}|^{\alpha-2} \left(1 + \frac{\hat{\lambda}}{|\hat{\lambda}|} \beta \tan \left(\frac{\pi\alpha}{2} \right) \right) \\ &= (1-\alpha) \frac{\gamma - \frac{x}{n}}{|\hat{\lambda}|} \end{aligned}$$

e

$$\begin{aligned} K_X(\hat{\lambda}) &= \gamma \hat{\lambda} - c |\hat{\lambda}|^\alpha \left(1 + \beta \frac{\hat{\lambda}}{|\hat{\lambda}|} \tan \left(\frac{\pi\alpha}{2} \right) \right) \\ &= \gamma \hat{\lambda} - c |\hat{\lambda}|^\alpha \frac{\gamma - \frac{x}{n}}{c\alpha \hat{\lambda} |\hat{\lambda}|^{\alpha-2}} = \frac{\hat{\lambda}}{\alpha} \left(\alpha\gamma - \gamma + \frac{x}{n} \right). \end{aligned}$$

- Para $\alpha = 1$ temos

$$K_X(t) = \gamma t - c |t|^\alpha \left(1 + \beta \frac{t}{|t|} \frac{2}{\pi} \log |t| \right),$$

pelo que, para $t \neq 0$,

$$K'_X(t) = \gamma - c|t|^{\alpha-1} \left(\alpha \frac{t}{|t|} + \alpha\beta \frac{2}{\pi} \log|t| + \alpha\beta \frac{2}{\pi} \right)$$

e

$$K''_X(t) = -c|t|^{\alpha-2} \frac{t}{|t|} \left[(\alpha - 1) \left(\frac{t}{|t|} \alpha + \alpha\beta \frac{2}{\pi} \log|t| + \beta \frac{2}{\pi} \right) + \alpha\beta \frac{2}{\pi} \right].$$

De $\hat{\lambda}$ ser a solução de $K'(\hat{\lambda}) = \frac{x}{n}$ temos,

$$\begin{aligned} \gamma - c|\hat{\lambda}|^{\alpha-1} \left(\alpha \frac{\hat{\lambda}}{|\hat{\lambda}|} + \alpha\beta \frac{2}{\pi} \log|\hat{\lambda}| + \beta \frac{2}{\pi} \right) &= \frac{x}{n} \Rightarrow \\ \alpha \frac{\hat{\lambda}}{|\hat{\lambda}|} + \alpha\beta \frac{2}{\pi} \log|\hat{\lambda}| + \beta \frac{2}{\pi} &= \frac{\gamma - \frac{x}{n}}{c|\hat{\lambda}|^{\alpha-1}} \Rightarrow \\ 1 + \frac{\hat{\lambda}}{|\hat{\lambda}|} \beta \frac{2}{\pi} \log|\hat{\lambda}| &= \frac{\hat{\lambda}}{|\hat{\lambda}|} \alpha \left(\frac{\gamma - \frac{x}{n}}{c|\hat{\lambda}|^{\alpha-1}} - \beta \frac{2}{\pi} \right) \end{aligned}$$

pelo que

$$\begin{aligned} K''_X(\hat{\lambda}) &= -c|\hat{\lambda}|^{\alpha-2} \frac{\hat{\lambda}}{|\hat{\lambda}|} \left[(\alpha - 1) \left(\frac{\hat{\lambda}}{|\hat{\lambda}|} \alpha + \alpha\beta \frac{2}{\pi} \log|\hat{\lambda}| + \beta \frac{2}{\pi} \right) + \alpha\beta \frac{2}{\pi} \right] \\ &= -c|\hat{\lambda}|^{\alpha-2} \frac{\hat{\lambda}}{|\hat{\lambda}|} \left[(\alpha - 1) \frac{\gamma - \frac{x}{n}}{c|\hat{\lambda}|^{\alpha-1}} + \alpha\beta \frac{2}{\pi} \right] \\ &= -\frac{(\alpha - 1)(\gamma - \frac{x}{n})}{\hat{\lambda}} - c|\hat{\lambda}|^{\alpha-3} \hat{\lambda} \alpha \beta \frac{2}{\pi} \end{aligned}$$

e

$$\begin{aligned} K_X(\hat{\lambda}) &= \gamma \hat{\lambda} - c|\hat{\lambda}|^\alpha \left(1 + \beta \frac{\hat{\lambda}}{|\hat{\lambda}|} \frac{2}{\pi} \log|\hat{\lambda}| \right) \\ &= \gamma \hat{\lambda} - c|\hat{\lambda}|^\alpha \frac{\hat{\lambda}}{|\hat{\lambda}|} \alpha \left(\frac{\gamma - \frac{x}{n}}{c|\hat{\lambda}|^{\alpha-1}} - \beta \frac{2}{\pi} \right) \\ &= \hat{\lambda} \left(\gamma - \frac{\gamma - \frac{x}{n}}{\alpha} + \frac{2c\beta|\hat{\lambda}|^{\alpha-1}}{\alpha\pi} \right). \end{aligned}$$

Substituindo as expressões anteriores em

$$f_{S_n}(x) = \exp\left(-\hat{\lambda}x + nK_X(\hat{\lambda})\right) \frac{1}{\sqrt{2\pi K_X''(\hat{\lambda})}} [1 + O(n^{-1})]$$

obtemos:

- Para $\alpha = 1$

$$f_{S_n}(x) = \frac{\exp\left[\hat{\lambda}\left[\left(1 - \frac{1}{\alpha}\right)(n\lambda - x) - \frac{2c\beta n|\hat{\lambda}|^{\alpha-1}}{2\pi}\right]\right]}{\sqrt{-2\pi\hat{\lambda}^{-1}(\alpha-1)\left(\gamma - \frac{x}{n}\right) - c|\hat{\lambda}|^{\alpha-3}\hat{\lambda}\alpha\beta 2\pi^{-1}}} [1 + O(n^{-1})].$$

- Para $\alpha \neq 1$

$$f_{S_n}(x) = \frac{\exp\left[\lambda\left(\frac{1}{\alpha} - 1\right)(x - n\gamma)\right] \sqrt{|\hat{\lambda}|}}{\sqrt{2\pi(1-\alpha)\left(\gamma - \frac{x}{n}\right)}} [1 + O(n^{-1})].$$

Obtivemos assim expressões formais para expansões de Edgeworth e expansões de Edgeworth diferidas, similares às conhecidas em situações mais regulares, e que mostram que o uso do *tilting* e *back tilting* tem um efeito positivo no contexto de grandes desvios — que no caso de leis estáveis não gaussianas têm que ser entendidas em termos de localização-escala mais gerais do que o par $\mathbb{E}(X)$, $\text{Var}(X)$ em variáveis no domínio de atracção da gaussiana.

2.5.4 Expansões de Cornish–Fisher e Cornish–Fisher Diferidas

As expansões de Cornish-Fisher são usadas para determinar numericamente as distribuições de probabilidades de estatísticas, quando as suas distribuições exactas são difíceis de calcular. Sejam Y e u_α definidos por

$$\mathbb{P}(Y \leq y_\alpha) = \Phi(u_\alpha) = 1 - \alpha,$$

onde Y_α é uma variável aleatória e $\alpha \in]0, 1[$.

As expansões de Cornish-Fisher são duas expansões assintóticas que relacionam os quantis y_α e u_α : uma expansão normalizadora que expressa u_α como função de y_α e a sua expansão inversa dando y_α em termos de u_α .

Expandindo $\Phi(u_\alpha)$ vem

$$\Phi(u_\alpha) = \Phi(y_\alpha + (u_\alpha - y_\alpha)) = \Phi(y_\alpha) + \sum_{r=1}^{\infty} \frac{(u_\alpha - y_\alpha)^r}{r!} D^r \Phi(y_\alpha),$$

com $D^r \Phi(y) = \frac{d^r \Phi(y)}{dy^r} = -H_{r-1}(y)\phi(y)$. Temos então,

$$\Phi(u_\alpha) = \Phi(y_\alpha) - \sum_{r=1}^{\infty} \frac{(u_\alpha - y_\alpha)^r}{r!} H_{r-1}(y_\alpha)\phi(y_\alpha). \quad (2.58)$$

De

$$\mathbb{P}(Y \geq y_\alpha) = 1 - \Phi(y_\alpha) + \phi(y_\alpha) \left\{ \frac{\rho_3}{6\sqrt{n}} H_2(y_\alpha) + \frac{\rho_4}{24n} H_3(y_\alpha) + \frac{\rho_3^2}{72n} H_5(y_\alpha) \right\} + O(n^{-1})$$

e $\mathbb{P}(Y \leq Y_\alpha) = 1 - \Phi(u_\alpha)$ e de (2.58), resulta

$$u_\alpha \approx p(y_\alpha) = y_\alpha - \frac{\rho_3}{6\sqrt{n}}(y_\alpha^2 - 1) + \frac{\rho_3^2}{36n}(4y_\alpha^2 - 7y_\alpha) - \frac{\rho_4}{24n}(y_\alpha^3 - 3y_\alpha). \quad (2.59)$$

Então, qualquer probabilidade $\mathbb{P}(Y \geq y_\alpha)$ até $O(n^{-1})$ é facilmente aproximada por $1 - \Phi(u_\alpha)$, com o quantil u_α dado por (2.59). Este procedimento é válido para qualquer estatística contínua padronizada que tenha terceiro e quarto cumulantes de ordens $O(n^{-1/2})$ e $O(n^{-1})$, respectivamente, e os demais cumulantes de ordem $O(n^{-1})$.

O objectivo da expansão inversa de Cornish-Fisher é expressar os quantis y_α de Y como função dos correspondentes quantis u_α da distribuição gaussiana reduzida. A inversão da expansão (2.59) para calcular y_α em termos do quantil u_α da gaussiana reduzida é feito através da fórmula geral de inversão de Lagrange, obtendo-se

$$y_\alpha \approx u_\alpha + \frac{\rho_3}{6\sqrt{n}}(u_\alpha^2 - 1) + \frac{\rho_3^2}{36n}(2u_\alpha^2 - 5u_\alpha) + \frac{\rho_4}{24n}(u_\alpha^3 - 3u_\alpha). \quad (2.60)$$

A importância da inversão da expansão de Cornish-Fisher (2.60) na inferência resulta do facto de possibilitar o cálculo dos quantis de estatísticas em termos dos quantis correspondentes da distribuição gaussiana reduzida.

Exemplo

Suponhamos que $X \sim \chi_n^2$ e seja $Y = \frac{X-n}{\sqrt{2n}}$ a variável aleatória qui-quadrado padronizada, cujos terceiro e quarto cumulantes padronizados são $\rho_3 = 2\sqrt{2}$ e $\rho_4 = 12$. Então $\mathbb{P}(X \leq z_\alpha) = \mathbb{P}\left(Y \leq \frac{z_\alpha - n}{\sqrt{2n}}\right)$, e portanto, considerando a expansão (2.60) até à ordem n^{-1} tem-se

$$y_\alpha \approx n + \sqrt{2n} \left\{ u_\alpha + \frac{\sqrt{2}}{3\sqrt{n}}(u_\alpha^2 - 1) + \frac{\sqrt{1}}{18n}(u_\alpha^3 - 7u_\alpha) \right\}. \quad \blacksquare$$

Já vimos que se X_1, \dots, X_n forem variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade f_X , $\mathbb{E}(X) = \mu$, $\text{Var}(X) = \sigma^2$, função

geradora de momentos M_X e função geradora de cumulantes $\psi_X = K_X = \log(M_X)$, a família associada a f_X , no sentido de Cramér-Khinchine, é

$$f_{X_\lambda}(y) = f(y; \lambda) = \frac{e^{\lambda y} f_Y(y)}{M_X(\lambda)} = e^{(\lambda y - K_X(\lambda))} f_X(y),$$

com $\lambda \in \mathbb{R}$. Em termos de função distribuição tem-se

$$F_{S_{n,\hat{\lambda}}}(y_\alpha) = F_{S_{n,\hat{\lambda}}}^* \left(\frac{y_\alpha - x}{\sqrt{nK_X''(\hat{\lambda})}} \right).$$

Pela formula (2.59) tem-se que

$$\begin{aligned} u_\alpha \approx & \frac{Y_\alpha - x}{\sqrt{nK_X''(\hat{\lambda})}} - \frac{\rho_3}{6\sqrt{n}} \left[\frac{(y_\alpha - x)^2}{nK_X''(\hat{\lambda})} - 1 \right] + \frac{\rho_3^2}{36n} \left[\frac{4(y_\alpha - x)^2}{nK_X''(\hat{\lambda})} - \frac{7(y_\alpha - x)}{\sqrt{nK_X''(\hat{\lambda})}} \right] - \\ & - \frac{\rho_4}{24n} \left[\left(\frac{y_\alpha - x}{\sqrt{nK_X''(\hat{\lambda})}} \right)^3 - \frac{3(y_\alpha - x)}{\sqrt{nK_X''(\hat{\lambda})}} \right], \end{aligned}$$

onde $\mathbb{E}(S_{n,\hat{\lambda}}) = x$ e ρ_3, ρ_4 são os cumulantes padronizados de X . ■

2.5.5 Expansões com Base no Ponto Sela

Estas expansões são muito importantes na teoria assintótica por aproximarem com grande precisão as funções de densidade e de distribuição, sendo facilmente deduzidas da função geradora de cumulantes correspondente.

Sejam Y_1, \dots, Y_n variáveis aleatórias contínuas independentes e identicamente distribuídas com função densidade f e funções geradoras de momentos e cumulantes M e K , respectivamente. Defina-se a família exponencial conjugada de f , indexada por um parâmetro λ , por

$$f(y; \lambda) = \exp\{\lambda y - K(\lambda)\} f(y). \quad (2.61)$$

A família exponencial (2.61) reproduz exactamente a função densidade f , postulada para os dados quando $\lambda = 0$. O divisor necessário para normalizar a expressão $\exp(\lambda y) f(y)$, é igual à função geradora de momentos M de Y . A função geradora de cumulantes $K(t; \lambda)$, correspondente a (2.61) é expressa em termos da função geradora de Y , $K(t)$, por

$$K(t; \lambda) = K(t + \lambda) - K(\lambda).$$

Sejam $f_{S_n}(s; \lambda)$ e $K_{S_n}(t; \lambda)$ as funções de densidade e geradora de cumulantes de S_n relativas à família (2.61). Tem-se

$$K_{S_n}(t; \lambda) = n K(t + \lambda) - n K(t)$$

e, por inversão, vem

$$f_{S_n}(s; \lambda) = \exp\{s \lambda - n K(\lambda)\} f_{S_n}(s), \quad (2.62)$$

sendo $f_{S_n}(s) = f_{S_n}(s; 0)$.

As funções de densidade de S_n e S_n^* correspondentes à família (2.61) estão relacionadas por

$$f_{S_n}(s; \lambda) = f_{S_n^*}(y; \lambda) \frac{1}{\sqrt{n K''(\lambda)}}, \quad (2.63)$$

onde $y = \frac{s - n K'(\lambda)}{\sqrt{n K''(\lambda)}}$. Aproxima-se $f_{S_n^*}$ pela expansão Edgeworth (2.52) escolhendo $y = 0$ para anular o termo $O(n^{-1/2})$. Esta escolha equivale a considerar a distribuição em (2.61) definida por $\hat{\lambda}$ que satisfaz a equação $K'(\hat{\lambda}) = \frac{s}{n}$. O estimador $\hat{\lambda}$ pode ser interpretado como o estimador de máxima verosimilhança de λ baseado numa única observação de s de (2.62). Logo,

$$f_{S_n}(s; \hat{\lambda}) = f_{S_n^*}(0, \hat{\lambda}) \frac{1}{\sqrt{n K(\hat{\lambda})}}.$$

Agora, $f_{S_n^*}(0, \hat{\lambda})$ vem de (2.52) observando que os cumulantes referentes a (2.61) são n vezes as derivadas de $K(\lambda)$, isto é,

$$f_{S_n^*}(0; \hat{\lambda}) = \frac{1}{\sqrt{2\pi}} \{1 + M(\hat{\lambda})\} + O(n^{-2}), \quad (2.64)$$

onde $M(\lambda)$ é um termo de ordem n^{-1} dado por

$$M(\lambda) = \frac{3\rho_4(\lambda) - 5\rho_3(\lambda)^2}{24n}, \quad (2.65)$$

sendo $\rho_j(\lambda) = \frac{K^{(j)}(\lambda)}{\sqrt{(K^{(2)}(\lambda))^j}}$ para $j = 3, 4$ e $K^{(j)}(\lambda) = \frac{d^j K(\lambda)}{d\lambda^j}$. Assim, $\rho_3(\lambda)$ e $\rho_4(\lambda)$ são os cumulantes padronizados que medem a assimetria e a curtose da distribuição (2.61). O erro em (2.64) é $O(n^{-2})$, pois o polinómio correspondente a $O(n^{-3/2})$ é de ordem ímpar e anula-se para zero.

Fazendo $\lambda = \hat{\lambda}$ em (2.62), explicitando $f_{S_n}(s)$ e usando (2.63) e (2.64) vem

$$\begin{aligned}
 f_{S_n}(s; \hat{\lambda}) &= \exp\{s \hat{\lambda} - n K(\hat{\lambda})\} f_{S_n}(s) \\
 &= \exp\{n K(\hat{\lambda}) - s \hat{\lambda}\} f_{S_n}(s; \hat{\lambda}) \\
 &= \exp\{n K(\hat{\lambda}) - s \hat{\lambda}\} f_{S_n}(0; \hat{\lambda}) \frac{1}{\sqrt{n K''(\hat{\lambda})}} \\
 &= \exp\{n K(\hat{\lambda}) - s \hat{\lambda}\} \frac{1}{\sqrt{2\pi}} \{1 + M(\hat{\lambda}) + O(n^{-2})\} \frac{1}{\sqrt{n K''(\hat{\lambda})}} \\
 &= \frac{\exp\{n K(\hat{\lambda}) - s \hat{\lambda}\}}{\sqrt{2n\pi K''(\hat{\lambda})}} \{1 + M(\hat{\lambda}) + O(n^{-2})\} \tag{2.66}
 \end{aligned}$$

A fórmula (2.66) para aproximar a função densidade de S_n é denominada expansão ponto sela da soma e produz aproximações precisas para funções densidade baseadas nas funções geradoras de cumulantes. Uma dedução alternativa pode ser feita através do integral de contorno que inverte a função geradora de momentos de S_n .

Exemplo

Consideremos o exemplo apresentado no ponto 1.5.1. no qual X_1, \dots, X_n têm distribuição exponencial de média um. S_n tem função de densidade $f_{S_n}(x) = \frac{x^{n-1} e^{-x}}{(n-1)!} I_{(0,+\infty)}$. Então, $M(\lambda) = \frac{1}{1-\lambda}$ e $K(\lambda) = -\log(1-\lambda)$, válida para $\lambda < 1$. O estimador de máxima verosimilhança de λ é $\hat{\lambda} = 1 - \frac{n}{x}$, $K(\hat{\lambda}) = -\log\left(\frac{n}{x}\right)$ e $M(\hat{\lambda}) = \frac{-1}{12n}$. Tem-se então que

$$f_{S_n}(x) = \frac{x^{n-1} e^{-x}}{e^{-n} n^{n-1/2} \sqrt{2\pi}} \left\{ 1 - \frac{1}{12n} + O(n^{-2}) \right\}$$

é a expansão ponto sela para a função densidade de f_{S_n} . ■

Observamos que o termo principal de (2.66) só depende da função geradora de cumulantes K de Y . Observe-se ainda que para usar (2.66) é necessário calcular, além de $\hat{\lambda}$, a função geradora de cumulantes e não apenas os primeiros quatro cumulantes. O termo principal de (2.66) não é a função de densidade da distribuição gaussiana $N(0, 1)$ e, embora seja sempre positivo, nem sempre o integral vale um (no entanto pode ser normalizado). A expansão de (2.66) é dada em potências de n^{-1} , enquanto que a expansão de Edgeworth é dada em potências de $n^{-1/2}$. Uma desvantagem de (2.66) é que nem sempre é fácil integrar o seu lado direito para obter uma aproximação para a função de distribuição de S_n . A expansão ponto sela constitui, em geral, uma melhor aproximação para a função densidade de S_n^* do que (2.52), pois o erro é $O(n^{-2})$ em vez de $O(n^{-3/2})$. A expansão ponto sela tem um erro multiplicativo enquanto na expansão de Edgeworth

o erro é aditivo. A fórmula (2.66) é satisfeita mesmo para regiões de grandes desvios da forma $|s - n \mathbb{E}(Y)| \leq b_n$, para b_n fixo, e em certos casos, mesmo para todos os valores de s .

A expansão ponto sela para S_n^* (soma padronizada) num ponto qualquer w tem uma expressão idêntica a de (2.66) com $n K'(0) + \sqrt{n} w K''(0)$ no lugar de s e o radicando substituído por $\frac{2\pi K''(\hat{\lambda})}{K''(0)}$.

A expansão para a função densidade da média amostral $\bar{Y}_n = \frac{S_n}{n}$ segue directamente de (2.66)

$$f_{\bar{Y}_n} = \left\{ \frac{n}{2\pi K''(\hat{\lambda})} \right\}^{1/2} \exp \left[n(K(\hat{\lambda}) - \hat{\lambda} y) \right] \left\{ 1 + M(\hat{\lambda}) + O(n^{-2}) \right\}, \quad (2.67)$$

onde M é obtido de (2.65). O termo principal de (2.67) é denominado aproximação ponto sela para $f_{\bar{Y}_n}$. Assim, basta conhecer a função geradora de cumulantes K comum às variáveis aleatórias independentes e identicamente distribuídas para se obter a aproximação ponto sela da função densidade da média amostral dessas variáveis.

2.6 Relações entre Distribuições Conjugadas e Somas Aleatórias

Comecemos por observar que se X for uma variável aleatória positiva com valor médio finito μ_X , função de distribuição F_X e função densidade de probabilidade f_X , então

$$f_X^*(x) = \frac{1 - F_X(x)}{\mu_X}$$

é uma função densidade de probabilidade, que denominamos *dual* de f_X ; analogamente, chamamos duais às correspondentes variáveis aleatórias, funções de distribuição, funções características, transformadas de Laplace, etc.

De facto, basta recordar que se $X > 0$, existindo $\mu_X = \mathbb{E}(X)$ este pode ser calculado como $\mu_X = \int_0^\infty (1 - F_X(x)) dx$. Esta expressão — que tem a vantagem de ser genérica, transformando o integral de Stieltjes num integral de Riemann — estabelece-se imediatamente usando integração por partes, uma vez que a existência de valor médio garante que $\lim_{x \rightarrow \infty} x [1 - F_X(x)] = 0$.

Analogamente, no caso de uma variável aleatória discreta com suporte natural, é óbvio que definindo $p_k^* = \frac{1 - F_X(k-1)}{\mu_X}$ se obtém uma função massa de probabilidade dual.

É imediato estabelecer que se $X \sim \text{Exponencial}(\delta)$ ou $Y \sim \text{Geométrica}(p)$, $f_X^* = f_X$ e $p_Y^* = p_Y$, são “auto-duais”. Adiante usaremos funções características para estabelecer

que se trata de uma caracterização da exponencial entre as absolutamente contínuas, e da geométrica entre as discretas, respectivamente. Tem, no entanto, interesse considerar relações menos fortes na família das paretos generalizadas e noutras famílias com esta relacionadas, que abordamos adiante.

No caso de $X \sim \text{Poisson}(\lambda)$, a conjugada X^* tem uma interpretação simples em termos de modelos hierárquicos: corresponde a uma filtragem (“*thinning*”) da Poisson por uma uniforme discreta. O objectivo é relacionar a dualidade, entendida na acepção que aqui apresentamos, com somas aleatórias, um tema de crescente importância em modelação estatística.

Alguns Exemplos — Densidades Duais

Apresentamos seguidamente, com os comentários adequados, algumas densidades (no sentido geral: função densidade de probabilidade no caso absolutamente contínuo, função massa de probabilidade no caso discreto) duais.

1. Família das Pareto

i. Pareto clássica

De seguida, vamos calcular a função dual para a distribuição Pareto. A função densidade de probabilidade da Pareto é dada por:

$$f(x) = \begin{cases} \frac{\beta\alpha^\beta}{x^{\beta+1}} & \text{se } x \geq \alpha \\ 0 & \text{se } x < \alpha \end{cases}$$

como $\alpha, \beta > 0$.

A média de uma variável aleatória com distribuição Pareto de parâmetros α e β é dada por $\mathbb{E}(X) = \frac{\beta\alpha}{\beta-1}$, a função de distribuição é dada por $F(x) = 1 - \frac{\alpha^\beta}{x^\beta}$.

Calculemos agora a função dual.

$$\begin{aligned} f^*(x) &= \frac{1 - 1 + \frac{\alpha^\beta}{x^\beta}}{\frac{\beta\alpha}{\beta-1}} = \frac{\beta-1}{x^\beta} \left(\frac{\beta\alpha^\beta}{\alpha} \right) \\ &= \frac{\beta-1}{x^\beta} \left(\frac{\alpha}{\beta^{-\frac{1}{\beta-1}}} \right)^{\beta-1} \end{aligned}$$

Ou seja, a variável X^* tem distribuição *Pareto* $\left(\beta-1, \frac{\alpha}{\beta^{-\frac{1}{\beta-1}}} \right)$.

ii. Pareto Generalizada

Consideremos agora uma variável aleatória X com distribuição de Pareto, isto é, $X \sim \text{Pareto}(\alpha)$. Consideremos a seguinte reparametrização da função de distribuição

$$F_X(x) = 1 - (1 + \gamma x)^{-1/\gamma} \text{ com } x > 0, \delta \in \mathbb{R} \text{ e } 1 + \gamma x > 0.$$

Vamos calcular $\mathbb{E}(X)$ recorrendo ao resultado para variáveis aleatórias positivas que afirma que

$$\mu_X = \int_0^\infty [1 - F_X(x)]dx.$$

Prova-se que a esperança de X só existe no caso em que $0 < \gamma < 1$, tendo-se neste caso

$$\begin{aligned} \mu &= \int_0^\infty (1 - \gamma x)dx = \frac{1}{\gamma - 1} [(1 + \gamma x)^{-1/\gamma}]_0^\infty \\ &\Leftrightarrow \mu = \frac{1}{1 - \gamma} \end{aligned}$$

Podemos agora calcular a função dual de f_X tendo-se

$$\begin{aligned} f_X^*(x) &= \frac{1 - F_X(x)}{\mu} = \frac{1 - 1 + (1 + \gamma x)^{-1/\gamma}}{\frac{1}{1 - \gamma}} \\ &\Leftrightarrow f_X^*(x) = (1 - \gamma)(1 + \gamma x)^{-1/\gamma} \end{aligned}$$

Quando $0 < x < \frac{1}{|\gamma|}$, as expressões que se obtêm para a esperança e função dual são iguais às obtidas para o caso $0 < \gamma < 1$.

No modelo Pareto generalizado apresentado anteriormente distinguem-se os seguintes três submodelos:

– $\gamma \rightarrow 0$ submodelo exponencial

$$\omega_0(X) = 1 - e^{-x}, \quad x \geq 0,$$

– $\gamma > 0$ a Pareto usual

$$\omega_{1, \alpha}(X) = 1 - x^{-\alpha}, \quad x \geq 1,$$

– $\gamma < 0$ uma subclasse da família beta

$$\omega_{2, \alpha}(X) = 1 - (-x)^{-\alpha}, \quad -1 \leq x \leq 0.$$

Note-se que $\alpha \in \mathbb{R}^+$ é o parâmetro de escala.

2. Geométrica

Consideremos o caso de uma variável aleatória com distribuição geométrica, isto é, $X \sim \text{Geométrica}(p)$. Se $X \sim \text{Geométrica}(p)$ então a função massa de probabilidade é dada por

$$\mathbb{P}(X = x) = p q^x, \quad x = 0, 1, 2, \dots$$

e a média da variável é $\mathbb{E}(x) = \frac{q}{p}$, onde p é a probabilidade de sucesso e $q = 1 - p$. A função de distribuição de X é dada por

$$F(x) = p \frac{1 - q^{x+1}}{1 - q}.$$

Calculando agora a função dual obtemos

$$\begin{aligned} f^*(x) &= \frac{1 - p \frac{1 - q^{x+1}}{1 - q}}{\frac{q}{p}} = \frac{p(p - p + pq^{x+1})}{pq} \\ &= \frac{p^2 q^{x+1}}{pq} \Leftrightarrow f^*(x) = p q^x. \end{aligned}$$

Também neste caso a função dual coincide com a função massa de probabilidade da variável aleatória.

3. Betas

Quando no caso 1(ii) trabalhamos com a Pareto generalizada surgiu-nos, quando $\gamma < 0$, uma subfamília importante — a família das betas. Neste ponto vamos calcular a função dual para duas betas particulares, o caso em que $\alpha = 1$ e o caso em que $\beta = 1$.

Se $X \sim \text{Beta}(\alpha, \beta)$ a função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} & \text{se } 0 < x < 1 \\ 0 & \text{outros casos} \end{cases},$$

com

$$B(\alpha, \beta) = \int_{0^+}^{1^-} x^{\alpha-1}(1-x)^{\beta-1} dx.$$

A função distribuição é dada por:

$$F(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ \frac{1}{B(\alpha, \beta)} \int_{0^+}^x y^{\alpha-1}(1-y)^{\beta-1} dy & \text{se } 0 < x < 1 \\ 1 & \text{se } x \geq 1 \end{cases},$$

sendo a média dada por $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$.

A função dual de f quando $\alpha = 1$ é

$$f(x) = \begin{cases} \frac{(1-x)^{\beta-1}}{B(1,\beta)} & \text{se } 0 < x < 1 \\ 0 & \text{outros casos} \end{cases},$$

e $\mathbb{E}(X) = \frac{1}{1+\beta}$. Como

$$B(1,\beta) = \int_0^1 (1-x)^{\beta-1} dx = \left[-\frac{1}{\beta}(1-x)^\beta \right]_0^1 \Leftrightarrow B(1,\beta) = \frac{1}{\beta},$$

tem-se

$$\begin{aligned} F(x) &= \beta \int_0^x (1-y)^{\beta-1} dy = \beta \left[-\frac{1}{\beta}(1-y)^\beta \right]_0^x \\ &\Leftrightarrow F(x) = 1 - (1-x)^\beta. \end{aligned}$$

Sendo a função dual dada por

$$f^*(x) = (\beta + 1)(1-x)^\beta.$$

Assim a função dual de uma função densidade de probabilidade de uma beta com $\alpha = 1$, é ainda uma beta, isto é,

$$X^* \sim \text{Beta}(1, \beta + 1).$$

De seguida vamos ver o que acontece quando $X \sim \text{Beta}(\alpha, 1)$.

Neste caso

$$f(x) = \begin{cases} \frac{x^{\alpha-1}}{B(\alpha,1)} & \text{se } 0 < x < 1 \\ 0 & \text{outros casos} \end{cases},$$

e $\mathbb{E}(X) = \frac{\alpha}{\alpha+1}$, tendo-se

$$B(\alpha, 1) = \frac{1}{\alpha},$$

e

$$F(x) = \alpha \int_0^x y^{\alpha-1} dy \Leftrightarrow F(x) = [y^\alpha]_0^x \Leftrightarrow F(x) = x^\alpha.$$

A função dual é dada por

$$f^*(x) = \frac{\alpha + 1}{\alpha} [(1-x^\alpha)^\alpha]^\frac{1}{\alpha}.$$

Neste caso podemos dizer que

$$X^* = (1 - X^\alpha)^\alpha \frown \text{Beta} \left(1, \frac{\alpha + 1}{\alpha} \right) .$$

Note-se que

$$0 < x < 1 \Rightarrow 0 < x^\alpha < 1 \Rightarrow -1 < -x^\alpha < 0 \Rightarrow 0 < 1 - x^\alpha < 1 \Rightarrow 0 < (1 - x^\alpha)^\alpha < 1 .$$

4. Família F de Fisher–Snedecor

Se $X \frown F(m, n)$, então a variável aleatória

$$Y = \frac{1}{1 + \frac{m}{n}X} \frown \text{Beta} \left(\frac{n}{2}, \frac{m}{2} \right) \text{ com } 0 < y < 1 ,$$

e conseqüentemente para cada $x > 0$,

$$F_X(x) = 1 - F_Y \left[\frac{1}{1 + \frac{m}{n}x} \right] .$$

Para $n > 2$ a esperança de X é dada por $\mu_X = \mathbb{E}(X) = \frac{n}{n-2}$. Calculando a função dual de X para $n > 2$, obtém-se

$$f_X^*(x) = \frac{n-2}{n} F_Y \left[\frac{1}{1 + \frac{m}{n}x} \right] \text{ para } x > 0 \text{ e } n > 2 .$$

A menos da constante $\frac{n-2}{n}$ a função dual de f_X é a função de distribuição de uma *beta* $\left(\frac{n}{2}, \frac{m}{2} \right)$ no caso em que $n > 2$. Note-se que no caso de $n \leq 2$, X não tem esperança.

2.6.1 Poisson Dual e Filtragem Uniforme Discreta

Consideremos a variável aleatória X com distribuição de *Poisson*(λ). A função massa de probabilidade é dada por :

$$\mathbb{P}(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, k = 0, 1, 2, \dots$$

A função de distribuição é dada por

$$F(x) = P(X \leq x) = \sum_{k=0}^x \frac{e^{-\lambda} \lambda^k}{k!},$$

e a esperança de X é dada por $\mathbb{E}(X) = \lambda$.

Consideremos a variável aleatória

$$Y|_{X=k} \sim \text{UniformeDiscreta no conjunto } \{0, 1, \dots, k\}.$$

Tem-se

$$\mathbb{P}(Y = j) = \frac{1}{\lambda} P(X > j),$$

isto é, a variável aleatória Y tem função de probabilidade dada pela expressão anterior com suporte de Y dado por $\{0, 1, \dots, k\}$. Ou seja,

$$\mathbb{P}(Y = j) = \frac{\text{cauda direita da Poisson}}{\text{valor médio da Poisson}}.$$

Pegando na igualdade anterior tem-se

$$p_j = \frac{1}{\lambda}(1 - F_X(j)) \Leftrightarrow \lambda p_j = 1 - F_X(j) \Leftrightarrow F_X(j) = 1 - \lambda p_j$$

onde $p_j = \mathbb{P}(Y = j)$.

A dual da função f é dada por

$$f^*(x) = \frac{1 - F_X(x)}{\mu} \Leftrightarrow f^*(x) = \frac{1 - 1 + \lambda p_x}{\lambda} \Leftrightarrow f^*(x) = p_x = \mathbb{P}(Y = x).$$

Mas,

$$\mathbb{P}(Y = x) = \frac{1}{\lambda} \mathbb{P}(X > x),$$

ou seja,

$$f^*(x) = \frac{\text{cauda direita da Poisson}}{\text{valor médio da Poisson}}.$$

Note-se que esta interpretação viabiliza o cálculo do valor médio e variância por condicionamento:

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}[Y|X]) = \mathbb{E}\left(\frac{X}{2}\right) = \frac{\lambda}{2}$$

e,

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E}(\text{Var}(Y|X)) + \text{Var}(\mathbb{E}(Y|X)) \\ \Leftrightarrow \text{Var}(Y) &= \mathbb{E}\left(\frac{(X+1)^2 - 1}{12}\right) + \text{Var}\left(\frac{X}{2}\right) \\ \Leftrightarrow \text{Var}(Y) &= \mathbb{E}\left(\frac{X^2}{12}\right) + \mathbb{E}\left(\frac{X}{6}\right) + \frac{1}{4}\text{Var}(X) \\ \Leftrightarrow \text{Var}(Y) &= \frac{(\lambda + \lambda^2) + 2\lambda + 3\lambda}{12} = \frac{\lambda^2 + 6\lambda}{12}. \end{aligned}$$

2.6.2 Funções Características e Dualidade

As variáveis duais também podem ser caracterizadas pelas suas funções características .

Seja $X \geq 0$ uma variável aleatória com função densidade F_X , função característica φ_X , e suponha-se que existe $\mu_X = \mathbb{E}(x) < \infty$. Então existe uma variável aleatória Y com função densidade de probabilidade

$$f_Y(x) = \frac{1 - F_X(y)}{\mu_X} I_{(0, \infty)}(x) .$$

Calculemos a função característica de Y .

$$\begin{aligned} \varphi_Y(t) &= \int_0^\infty e^{ity} \frac{1 - F_X(y)}{\mu_X} dy = \frac{1}{\mu_X} \left\{ \left[\frac{e^{ity}}{it} (1 - F_X(y)) \right]_0^\infty + \frac{1}{it} \int_0^\infty e^{ity} f_X(y) dy \right\} \\ &= \frac{1}{\mu_X} \left[-\frac{1}{it} + \frac{1}{it} \varphi_X(t) \right] \Leftrightarrow \varphi_Y(t) = \frac{\varphi_X(t) - 1}{it\mu_X} . \end{aligned}$$

De modo análogo, no caso discreto se X é uma variável aleatória positiva com $\mu_X = \mathbb{E}(X) < \infty$. Então existe uma variável aleatória com função massa de probabilidade dada por

$$p_k = p(x_k) = \frac{\mathbb{P}(X \geq x_k)}{\mu} = \frac{1 - F_X(x_{k-1})}{\mu} , \text{ com } x_k = 0, 1, 2, \dots .$$

Calculando a função característica de Y vamos obter

$$\varphi_Y(t) = \sum_{k=0}^{\infty} \frac{e^{itk}}{\mu} \mathbb{P}(X \geq K) = \sum_{k=0}^{\infty} a_k z^k ,$$

com

$$a_k = \frac{\mathbb{P}(X \geq k)}{\mu} \text{ e } z = e^{it} .$$

A série anterior é uma série de potências, logo os coeficientes a_k são únicos. Assim, os coeficientes a_k determinam completamente a distribuição de X através da relação

$$\mathbb{P}(X \geq k) = \mu a_k ,$$

supondo que X é discreta de suporte nos inteiros não negativos .

Se $X \sim \text{Exponencial}(\alpha)$ a sua função característica é dada por :

$$\varphi_X(t) = \frac{\alpha}{\alpha - it} .$$

A função característica de $X \sim \text{Geométrica}(p)$ é

$$\varphi_X(t) = p \sum_{k=0}^{\infty} e^{itk} (1-p)^k,$$

que é uma série de potências de razão igual a $e^{it}(1-p)$ e cujo primeiro termo é um. A soma da série é dada por

$$\varphi_X(t) = \frac{p}{1 - e^{it}(1-p)}.$$

Ou seja,

$$a_0 = p, a_1 = p(1-p), a_2 = p(1-p)^2, \dots$$

2.6.3 Caracterização da Exponencial e Geométrica pela Auto-Dualidade

Vamos provar que a exponencial é a única distribuição contínua tal que $f = f^*$. Consideremos uma variável aleatória X positiva, com função distribuição F e valor médio α , vamos provar que se $f(x) = \frac{1-F(x)}{\alpha}$ então $X \sim \text{Exponencial}(\alpha)$.

Tem-se

$$\frac{1-F(x)}{\alpha} = f(x) \Leftrightarrow \frac{1}{\alpha} = \frac{f(x)}{1-F(x)}.$$

Se integrarmos a expressão anterior obtemos

$$\begin{aligned} -\ln[1-F(x)] &= \frac{x}{\alpha} + c \Leftrightarrow \ln[1-F(x)] = -\frac{x}{\alpha} + c' \\ \Leftrightarrow 1-F(x) &= e^{-\frac{x}{\alpha} + c'} \Leftrightarrow F(x) = 1 - e^{-\frac{x}{\alpha}} e^{c'} \end{aligned}$$

Seja $k = e^{c'}$, temos que

$$F(x) = 1 - ke^{-\frac{x}{\alpha}}.$$

Mas $F(0) = 0$ donde

$$F(0) = 1 - k = 0 \Leftrightarrow k = 1,$$

ou seja,

$$F(x) = 1 - e^{-\frac{x}{\alpha}} I_{(0,+\infty)}(x),$$

isto é,

$$X \sim \text{Exponencial} \left(\frac{1}{\alpha} \right).$$

Provemos agora que a distribuição geométrica é a única distribuição discreta que verifica $f = f^*$. Para tal vamos recorrer a um teorema que diz o seguinte:

Teorema 2.6.1. *Seja X uma variável aleatória de valores inteiros não negativos que satisfaz*

$$\mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X \geq n)$$

para quaisquer m e n inteiros positivos. Então X tem distribuição geométrica.

A demonstração pode ser consultada, por exemplo, em Rohatgi (1976) pp. 190.

Tem-se que

$$f(x) = \frac{1 - F(x)}{\mu} \Leftrightarrow F(x) = 1 - \mu f(x) \Leftrightarrow \mathbb{P}(X \leq x) = 1 - \mu \mathbb{P}(X = x)$$

Por outro lado,

$$\begin{aligned} \mathbb{P}(X > m + n \mid X > m) &= \frac{\mathbb{P}(X > m + n \cap X > m)}{\mathbb{P}(X > m)} = \frac{\mathbb{P}(X > m + n)}{\mathbb{P}(X > m)} \\ &= \frac{1 - \mathbb{P}(X \leq m + n)}{1 - \mathbb{P}(X \leq m)} = \frac{1 - 1 + \mu \mathbb{P}(X = m + n)}{1 - 1 + \mu \mathbb{P}(X = m)} = \frac{p_{m+n}}{p_m}. \end{aligned}$$

Como,

$$\begin{aligned} p_k = P(X = k) &= \mathbb{P}(X > k - 1) - \mathbb{P}(X > k) \Leftrightarrow p_k = \sum_{i=k}^{\infty} \mathbb{P}(X = i) - \sum_{j=k+1}^{\infty} \mathbb{P}(X = j) \\ &\Leftrightarrow p_k = q_{k-1} - q_k \quad \text{com} \quad \mathbb{P}(X > m) = q_m = \sum_{k=m+1}^{\infty} p_k. \end{aligned}$$

Tem-se que

$$\mathbb{P}(X > m + n \mid X > m) = \frac{q_{m+n-1} - q_{m+n}}{q_{m-1} - q_m}.$$

Atendendo à notação introduzida anteriormente podemos escrever

$$\mathbb{P}(X > m + n \mid X > m) = \frac{\mathbb{P}(X > m + n)}{\mathbb{P}(X > m)} = \mathbb{P}(X \geq n) = \frac{q_{m+n}}{q_m}$$

donde

$$q_{m+n} = q_m q_{n-1} \text{ e } q_{m+1} = q_m q_0,$$

com

$$q_0 = \mathbb{P}(X > 0) = p_1 + p_2 + p_3 + \cdots = 1 - p_0 \Leftrightarrow q_0 = 1 - p_0.$$

Prova-se que

$$q_k = (1 - p_0)^k.$$

Assim,

$$q_{k-1} - q_k = (1 - p_0)^{k-1} - (1 - p_0)^k = (1 - p_0)^{k-1}(1 - 1 + p_0) ,$$

ou seja,

$$p_k = q_{k-1} - q_k = p_0(1 - p_0)^{k-1} .$$

Tem-se então que

$$\begin{aligned} \mathbb{P}(X > m + n \mid X > m) &= \frac{q_{m+n-1} - q_{m+n}}{q_{m-1} - q_m} = \frac{p_0(1 - p_0)^{m+n}}{p_0(1 - p_0)^m} = (1 - p_0)^n = q_{n-1} \\ \Leftrightarrow \mathbb{P}(X > m + n \mid X > m) &= \mathbb{P}(X > n) \end{aligned}$$

Ou seja, X^* tem distribuição geométrica.

2.6.4 Fórmulas de Pollaczek–Khintchine e de Beckman e Interpretação Probabilística da Dualidade

Seja $N \sim \text{Geométrica}(1 - p)$ com N a tomar os valores $0, 1, 2, \dots$ com probabilidade $p_k = (1 - p)p^k$ e $S_N = \sum_{k=0}^N y_k$, onde os y_k são réplicas independentes de Y , e N é independente dos y_k . Então

$$\varphi_S(t) = \mathbb{E}(e^{itS_N}) = \sum_{k=0}^N \mathbb{E}[e^{itS_N} \mid N = n](1 - p)p^n = \frac{1 - p}{1 - p \frac{\varphi_X(t) - 1}{i\mu_X t}} .$$

Esta é a célebre fórmula de Pollaczek–Khintchine por eles deduzida num contexto completamente diferente: seja o tempo de serviço numa fila de espera (FIFO) T , com $\mathbb{E}(T) = \mu_T < \infty$ e $n\mu_T = p$ o período de ocupação médio numa unidade de tempo. (Assume-se que $p < 1$, caso contrário a fila de espera cresce indefinidamente). Um cliente que se junte à fila de espera no instante $t > 0$, tem que esperar W (*waiting time*) até ser servido. A caracterização do tempo de espera W foi feita por Pollaczek–Khintchine em termos de função característica

$$\varphi_W(t) = \frac{1 - p}{1 - p \frac{\varphi_T(t) - 1}{i\mu_T t}} .$$

Assim, o que estabelecemos acima foi que o problema de Pollaczek–Khintchine tem a mesma solução que a soma aleatória (geométrica) de parcelas com distribuição conjugada no sentido de

$$f_Y(y) = \frac{1 - F_X(y)}{\mu_X} .$$

Considere-se agora, o processo de risco clássico

$$Z(t) = ct - \sum_{k=1}^{N(T)} X_K ,$$

com $Z(0) = 0$, $\sum_{k=1}^{\infty} X_K = 0$, $N(t)$ e os X_k independentes. Z é a reserva de uma companhia de seguros no instante t , os X_k são as indemnizações pagas em $[0, t]$. Note-se que $N(t)$ é um processo de Poisson homogéneo.

Consideremos X_k cópias independentes de X com função densidade F_X e $\mathbb{E}(X) = \mu < \infty$. Seja u o capital inicial da companhia de seguros. A probabilidade de ruína, como função de u , é

$$\psi(u) = \mathbb{P}[u + Z(t) < 0 \text{ para algum } t > 0].$$

Reescreva-se $c = (1 + \theta)\mu_X\lambda$ na definição $Z(t) = ct - \sum_{k=1}^{N(t)} X_K$, e defina-se

$$L = \sup_{t>0} \sum_{i=1}^{N(t)} -(1 + \theta)\mu_X t = \sup_{t>0} -Z(t),$$

que é a perda agregada máxima. Como $Z(0) = 0$, $L \geq 0$ quase certamente. Notando que

$$\begin{aligned} 1 - \psi(u) &= \mathbb{P}[u + Z(t) \geq 0 \text{ para algum } t > 0] \\ &= \mathbb{P}[u + (1 + \theta)\mu_X \lambda t - \sum_{i=1}^{N(t)} X_i \geq 0, \forall t > 0] \\ &= \mathbb{P}\left[\sum_{i=1}^{N(t)} X_i - (1 + \theta)\mu_X \lambda t \leq u, \forall t > 0\right] = \mathbb{P}[L \leq u]. \end{aligned}$$

A probabilidade de ruína pode então ser representada sob a forma

$$\mathbb{P}[L > u] = 1 - \frac{\theta}{1 + \theta} \sum_{n=0}^{\infty} \frac{H^{*n}(u)}{(1 + \theta)^n},$$

onde $H^{*n}(u) = H(u)$ é a função de Heaviside e para $n = 1, 2, \dots$

$$H^{*n} = \frac{1}{\mu_X} \int_0^u H^{*(n-1)}(u-x)[1 - F_X(x)]dx, \quad u \geq 0,$$

que é fórmula de Beckman para a probabilidade de ruína.

A fórmula de Beckman implica que

$$L = L_1 + \dots + L_N^* \text{ em distribuição,}$$

com N^*, L_1, \dots, L_N^* independentes e $N \sim \text{Geométrica}(p)$ com $p = \frac{\theta}{1+\theta}$, isto é,

$$N^* = \begin{cases} j & j=0,1,\dots \\ \frac{\theta}{(1+\theta)^{j+1}} & \end{cases},$$

onde as variáveis aleatórias L_K são independentes e identicamente distribuídas com densidade

$$f_L(x) = \frac{1 - F_X(x)}{\mu_X},$$

ou seja, L_K têm distribuição conjugada.

Capítulo 3

O TLC no Contexto da Teoria da Informação e Entropia

3.1 Entropia e Distância de Kullback-Leibler

A informação $\mathcal{I}(A)$ dada pelo acontecimento A é uma função decrescente de $\mathbb{P}(A)$. A definição proposta por Shannon (1948)

$$\mathcal{I}(A) = -\log_2 \mathbb{P}(A) .$$

Considere-se apenas variáveis aleatórias X puramente discretas com função massa de probabilidade $\{p_k\}_{k \in \mathcal{S}}$ e va 's puramente absolutamente contínuas com função densidade de probabilidade f_X , com suporte $\mathcal{S} = \mathcal{S}(X) = \{x_k : p_k = \mathbb{P}(X = x_k) > 0\}$, no caso discreto, e $\mathcal{S} = \mathcal{S}(X) = \{x_k : f_X(x) > 0\}$, no caso puramente absolutamente contínuo.

Tomando a variável aleatória definida por $\mathcal{I}_X(x_k) = \mathcal{I}(X = x_k)$, $k \in \mathcal{S}$ no caso em que a variável aleatória X é discreta, e a generalização para o caso em que X é absolutamente contínuo, $\mathcal{I}_X = -\log_2(f_X(X))$, e aceitando a convenção $0 \times \ln(0) = 0$, define-se entropia da variável aleatória X

$$\mathcal{H}(X) = \mathbb{E}[\mathcal{I}_X] = \begin{cases} -\sum_{k: x_k \in \mathcal{S}} \log_2(p_k) p_k & (X \text{ puramente discreta}) \\ -\int_{\mathcal{S}} \log_2[f_X(x)] f_X(x) dx & (X \text{ absolutamente contínuo}) \end{cases} .$$

Entropia é a medida de incerteza associada a uma variável aleatória (cf. Cover and Thomas) e mede-se em *bits*.

Alguns exemplos:

1. $X \sim \text{Geométrica}(p)$, $p \in (0, 1)$,

$$X = \begin{cases} k \\ p_k = p(1-p)^{k-1} \end{cases}, k = 1, 2, \dots$$

$$\begin{aligned} \mathcal{H}_X(X) = \mathbb{E}(\mathcal{I}_X) &= - \sum_{k=1}^{\infty} p(1-p)^{k-1} \log_2(p(1-p)^{k-1}) \\ &= - \log_2 p \sum_{k=1}^{\infty} p(1-p)^{k-1} - \log_2(1-p) \sum_{k=1}^{\infty} (k-1) p(1-p)^{k-1} \\ &= - \log_2 p + \frac{p-1}{p} \log_2(1-p). \end{aligned}$$

2. $X \sim \text{Exponencial}(\delta)$, $\delta > 0$

$$\begin{aligned} \mathcal{H}(X) &= - \int_0^{+\infty} \log_2(f_X(X)) f_X(x) dx = - \int_0^{+\infty} \log_2 \left(\frac{1}{\delta} e^{-\frac{x}{\delta}} \right) \frac{1}{\delta} e^{-\frac{x}{\delta}} dx \\ &= - \int_0^{+\infty} - \log_2 \delta \frac{1}{\delta} e^{-\frac{x}{\delta}} dx - \int_0^{+\infty} -\frac{x}{\delta} \log_2 e \frac{1}{\delta} x^{-\frac{x}{\delta}} dx \\ &= \log_2 \delta + \frac{\mathbb{E}(X)}{\delta} \log_2 e = \log_2(\delta e). \end{aligned}$$

3. $X \sim \text{Gaussiana}(\mu, \sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$

$$\begin{aligned} \mathcal{H}(X) &= \mathbb{E}[-\log_2(f_X(X))] = \mathbb{E} \left[-\log_2 \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} \right) \right] \\ &= \mathbb{E} \left[\frac{\log_2(2\pi\sigma^2)}{2} + \left(\frac{X-\mu}{\sigma} \right)^2 \frac{\log_2 e}{2} \right] \\ &= \frac{\log_2(2\pi\sigma^2)}{2} + \frac{\log_2 e}{2} = \frac{1}{2} \log_2(2\pi e\sigma^2), \end{aligned}$$

que, note-se, não depende do parâmetro μ .

As propriedades da entropia são substancialmente diferentes consoante X é discreta ou X é absolutamente contínua.

No caso discreto, $\mathcal{H}(X) \geq 0$, com $\mathcal{H}(X) = 0$ se e só se X for degenerada, enquanto no caso contínuo a entropia pode ser negativa. No caso discreto,

$$\mathcal{H}(\alpha X + \beta) = \mathcal{H}(X), \quad \alpha, \beta \in \mathbb{R} \text{ com } \alpha \neq 0$$

isto é, a entropia é invariante por translações e para mudanças de escala. No caso absolutamente contínuo, como $f_{\alpha X + \beta}(x) = \frac{1}{|\alpha|} f_X\left(\frac{x - \beta}{\alpha}\right)$, a entropia é invariante para translações mas não para mudanças de escala

$$\mathcal{H}(\alpha X + \beta) = \mathcal{H}(X) + \log_2(\beta).$$

Chamamos distância da entropia relativa, ou distância de Kullback-Leibler, ou ainda divergência de informação entre duas funções densidade de probabilidade f_X e f_Y

$$\mathcal{D}(f_X; f_Y) = \int_{S(X)} \log_2\left(\frac{f_X(x)}{f_Y(x)}\right) f_X(x) dx.$$

A aplicação \mathcal{D} que não é uma distância no sentido topológico, uma vez que não é simétrica e não obedece à desigualdade triangular. No entanto $\mathcal{D}(f_X; f_Y) \geq 0$, com $\mathcal{D}(f_X; f_Y) = 0$ se e só se $X \stackrel{d}{=} Y$. Note-se que a propriedade anterior é um caso especial da desigualdade de Gibbs. De facto, a desigualdade decorre imediatamente da desigualdade de Gibbs. Considerando $g : [0, \infty) \rightarrow \mathbb{R}$ tal que $g(x) \geq c(1 - \frac{1}{x})$, para algum $c > 0$ com igualdade se e só se $x = 1$, então

$$\begin{aligned} \int_S g\left(\frac{f_X(x)}{f_Y(x)}\right) f_X(x) dx &\geq c \int_{\mathbb{R}} \left(1 - \frac{f_Y(x)}{f_X(x)}\right) I_S(x) f_X(x) dx \\ &= c \left[\int_{\mathbb{R}} I_S(x) f_X(x) dx - \int_{\mathbb{R}} f_Y(x) I_S(x) dx \right] \\ &= c [1 - \mathbb{P}(Y \in S(X))] \geq 0. \end{aligned}$$

Mostra-se ainda que a igualdade ocorre se e só se $X \stackrel{d}{=} Y$.

Como $g(x) = \log_2(x)$ verifica as hipóteses acima, conclui-se que $\mathcal{D}(f_X; f_Y) \geq 0$, com $\mathcal{D}(f_X; f_Y) = 0$ se e só se $X \stackrel{d}{=} Y$.

Diz-se que $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Y$ se e só se $\mathcal{D}(f_X; f_Y) \xrightarrow[n \rightarrow \infty]{} 0$. Esta forma de convergência é mais forte do que a convergência em L^1 . Kullback mostrou que

$$\mathcal{D}(f_X; f_Y) \geq \frac{\log_2 e}{2} \left(\int_{\mathbb{R}} |f_X(x) - f_Y(x)| dx \right)^2.$$

Demonstramos agora que as variáveis aleatórias com distribuição exponencial e gaussiana têm notáveis propriedades de maximização da entropia, respectivamente na classe das

variáveis aleatórias com suporte numa semi-recta e valor médio finito, e na classe das variáveis aleatórias com suporte na recta e variância finita.

Teorema 3.1.1. *Seja f_X a função de densidade de probabilidade de uma variável aleatória X com $\mathbb{E}(X) = \mu$ e suporte $\mathcal{S} = (0, \infty)$, e seja $Y_\mu \sim \text{Exponencial}(\mu)$. Então $\mathcal{H}(X) \leq \mathcal{H}(Y_\mu)$, com igualdade se e só se X for uma variável aleatória exponencial.*

Demonstração

Temos que

$$\begin{aligned} \mathcal{D}(f_X; f_{Y_\mu}(x)) &= \int_{\mathbb{R}} \log_2 \left(\frac{f_X(x)}{f_{Y_\mu}(y)} \right) f_X(x) dx = \int_0^\infty [\log_2 f_X(x) - \log_2 f_{Y_\mu}(y)] f_X(x) dx \\ &= \int_0^\infty f_X(x) \log_2 f_X(x) dx + \log_2 \mu \int_0^\infty f_X(x) dx + \log_2 e \int_0^\infty \frac{x}{\mu} f_X(x) dx \\ &= -\mathcal{H}(X) + \log_2 \mu + \frac{\log_2 e}{\mu} \mu = -\mathcal{H}(X) + \log_2(\mu e) = -\mathcal{H}(X) + \mathcal{H}(Y_\mu). \end{aligned}$$

Como

$$D(f_X; f_{Y_\mu}) \geq 0 \Rightarrow -\mathcal{H}(X) + \mathcal{H}(Y_\mu) \geq 0 \Leftrightarrow \mathcal{H}(X) \leq \mathcal{H}(Y_\mu),$$

com igualdade se e só se $X \stackrel{d}{=} Y_\mu$. □

Note-se que, como referimos no capítulo 1, secção 4, a exponencial tem um papel semelhante à gaussiana na classe de Kovalenko e a Laplace tem um papel semelhante à lei gaussiana associada ao semi-grupo geométrico.

Teorema 3.1.2. *Seja f_X uma função de densidade de probabilidade de uma variável aleatória X com variância σ^2 e suporte $\mathcal{S} = \mathbb{R}$, e considere-se, sem perda de generalidade, que $\mathbb{E}(X) = 0$. Seja $Y_\sigma \sim \text{Gaussiana}(0, \sigma)$. Então $\mathcal{H}(X) \leq \mathcal{H}(Y_\sigma)$, com igualdade se e só se X for uma variável aleatória gaussiana.*

Demonstração

Temos que

$$\begin{aligned} \mathcal{D}(f_X; f_{Y_\sigma}) &= \int_{\mathbb{R}} \log_2(f_X(x)) f_X(x) dx - \int_{\mathbb{R}} \log_2(f_{Y_\sigma}(x)) f_X(x) dx \\ &= -\mathcal{H}(X) + \int_{\mathbb{R}} \frac{\log_2(2\pi\sigma^2)}{2} f_X(x) dx + \frac{\log_2 e}{2\sigma^2} \int_{\mathbb{R}} x^2 f_X(x) dx \\ &= -\mathcal{H}(X) + \frac{1}{2} \log_2(2\pi e \sigma^2) = -\mathcal{H}(X) + \mathcal{H}(Y_\sigma) \geq 0, \end{aligned}$$

e conseqüentemente $\mathcal{H}(Y_\sigma) \geq \mathcal{H}(X)$, com igualdade se e só se $X \stackrel{d}{=} Y_\sigma$. \square

Note-se que, na distância de Kullback-Leibler, uma variável aleatória X com valor médio $\mathbb{E}(X)$ e variância $\text{Var}(X)$ está mais próxima de $Y_{\mu,\sigma} \sim \text{Gaussiana}(\mathbb{E}(X), \sqrt{\text{Var}(X)})$ do que qualquer outra variável aleatória gaussiana. De facto,

$$\begin{aligned} \mathcal{D}(f_X; f_{Y_{\mu,\sigma}}) &= \int_{\mathbb{R}} f_X(x) \log_2(f_X(x)) dx - \int_{\mathbb{R}} f_X(x) \log_2(f_{Y_{\mu,\sigma}}(x)) dx \\ &= -\mathcal{H}(X) - \int_{\mathbb{R}} f_X(x) \log_2\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx \\ &= -\mathcal{H}(X) + \frac{\log_2(2\pi\sigma^2)}{2} \int_{\mathbb{R}} f_X(x) dx + \frac{\log_2 e}{2\sigma^2} \int_{\mathbb{R}} f_X(x) (x-\mu)^2 dx \\ &= -\mathcal{H}(X) + \frac{\log_2(2\pi\sigma^2)}{2} + \frac{\log_2 e}{2\sigma^2} \mathbb{E}((X-\mu)^2), \end{aligned}$$

pelo que a escolha óptima para μ é $\mathbb{E}(X)$. Por outro lado, minimizando $\mathcal{D}(f_X; f_{Y_{\mu,\sigma}})$ como função de σ , verifica-se que, para $\mu = \mathbb{E}(X)$,

$$\begin{aligned} \mathcal{D}(f_X; f_{Y_{\mu,\sigma}}) &= -\mathcal{H}(X) + \frac{\log_2(2\pi)}{2} + \frac{1}{2} \left[\log_2 \sigma^2 + \frac{\log_2 e}{\sigma^2} \text{Var}(X) \right] \\ \Rightarrow \frac{\partial \mathcal{D}(f_X; f_{Y_{\mu,\sigma}})}{\partial \sigma} &= \frac{1}{\sigma \log 2} - \log_2 e \frac{1}{\sigma^3} \text{Var}(X) = 0 \\ \Leftrightarrow \frac{\sigma^2}{\log 2} - \log_2 e \text{Var}(X) &= 0 \\ \Leftrightarrow \frac{\sigma^2}{\log 2} = \log_2 e \text{Var}(X) &\Leftrightarrow \sigma^2 = \log_e 2 \log_2 e \text{Var}(X) \Leftrightarrow \sigma^2 = \text{Var}(X), \end{aligned}$$

isto é, verifica-se que o mínimo é atingido quando $\text{Var}(X) = \sigma^2$.

Cálculo da Entropia para Distribuições Estáveis com Densidade Conhecida

Apenas em três casos a função de densidade de probabilidade associada a variáveis aleatórias estáveis para somas é conhecida: gaussiana, Cauchy e Lévy. Vimos já que se $X \sim \text{Gaussiana}(0, \sigma)$ então $\mathcal{D}(f_X; f_{Y_\sigma}) = -\mathcal{H}(X) + \mathcal{H}(Y_\sigma)$. Relativamente às outras duas distribuições conhecidas temos:

i. Cauchy

A função densidade de probabilidade é dada por

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2} \mathbf{I}_{\mathbb{R}}(x), \theta \in \mathbb{R}.$$

Uma vez que a entropia é invariante por translações, vamos calcular a entropia para o caso em que $\theta = 0$

$$\begin{aligned} \mathcal{H}(X_C) &= \mathbb{E}[-\log(f_{X_C}(x))] = - \int_{\mathbb{R}} \log_2 \left(\frac{1}{\pi} \frac{1}{1 + x^2} \right) \frac{1}{\pi} \frac{1}{1 + x^2} dx \\ &= \log_2(\pi) \int_{\mathbb{R}} \frac{1}{\pi} \frac{1}{1 + x^2} dx - \int_{\mathbb{R}} \log_2 \left(\frac{1}{1 + x^2} \right) \frac{1}{\pi} \frac{1}{1 + x^2} dx \\ &= \log_2(\pi) + \int_{\mathbb{R}} \log_2(1 + x^2) f_{X_C}(x) dx = \log_2(\pi) + \mathbb{E}[\log_2(1 + X_C^2)]. \end{aligned}$$

Dada uma variável aleatória X_C a sua distância de Kullback-Leibler a Y_C é

$$\begin{aligned} \mathcal{D}(f_{X_C}; f_{Y_C}) &= \int_{\mathbb{R}} \log_2 \left(\frac{f_{X_C}(x)}{f_{Y_C}(x)} \right) f_{X_C}(x) dx \\ &= \int_{\mathbb{R}} \log_2(f_{X_C}(x)) f_{X_C}(x) dx - \int_{\mathbb{R}} \log_2(f_{Y_C}(x)) f_{X_C}(x) dx \\ &= -\mathcal{H}(X_C) - \int_{\mathbb{R}} \log_2 \left(\frac{1}{\pi} \frac{1}{1 + y_C^2} \right) f_{X_C}(x) dx \\ &= -\mathcal{H}(X_C) + \log_2(\pi) \int_{\mathbb{R}} f_{X_C}(x) dx + \int_{\mathbb{R}} \log_2(1 + y_C^2) f_{X_C}(x) dx \\ &= -\mathcal{H}(X_C) + \mathcal{H}(Y_C) - \mathbb{E}[\log_2(1 + X_C^2)] + \mathbb{E}[\log_2(1 + Y_C^2)]. \end{aligned}$$

De $D(f_{X_C}; f_{Y_C}) \geq 0$ resulta

$$\mathcal{H}(Y_C) - \mathbb{E}[\log_2(1 + Y_C^2)] \geq \mathcal{H}(X_C) + \mathbb{E}[\log_2(1 + X_C^2)].$$

ii. Lévy

A distribuição de Lévy de parâmetros $(\frac{1}{2}, 1)$ é absolutamente contínua e tem função densidade probabilidade definida por

$$f_{X_L}(x) = \frac{\exp\left(-\frac{1}{2x}\right)}{\sqrt{2\pi} x^3} \mathbf{I}_{(0, +\infty)}(x).$$

A entropia desta distribuição é

$$\begin{aligned}
 \mathcal{H}(X_L) &= - \int_0^{\infty} \log_2(f_{X_L}(x)) f_{X_L}(x) dx \\
 &= - \int_0^{\infty} \log_2\left(\frac{\exp\left(-\frac{1}{2x}\right)}{\sqrt{2\pi} x^3}\right) f_{X_L}(x) dx \\
 &= - \int_0^{\infty} \left[\log_2 \exp\left(-\frac{1}{2x}\right) - \log_2(\sqrt{2\pi} x^3) \right] f_{X_L}(x) dx \\
 &= - \int_0^{\infty} \log_2 \exp\left(-\frac{1}{2x}\right) f_{X_L}(x) dx + \int_0^{\infty} \log_2(\sqrt{2\pi} x^3) f_{X_L}(x) dx \\
 &= \log_2 e \int_0^{\infty} \frac{1}{2x} f_{X_L}(x) dx + \frac{\log_2(2\pi)}{2} \int_0^{\infty} f_{X_L}(x) dx + \int_0^{\infty} \log_2(x^{3/2}) f_{X_L}(x) dx \\
 &= \frac{\log_2 e}{2} \int_0^{\infty} \frac{1}{x} \frac{\exp\left(-\frac{1}{2x}\right)}{\sqrt{2\pi} x^3} dx + \frac{\log_2(2\pi)}{2} + \int_0^{\infty} \frac{3}{2} \log_2(x) \frac{\exp\left(-\frac{1}{2x}\right)}{\sqrt{2\pi} x^3} dx \\
 &= \frac{\log_2 e}{2\sqrt{2\pi}} \int_0^{\infty} x^{-5/2} \exp\left(-\frac{1}{2x}\right) dx + \frac{1 + \log_2 \pi}{2} + \\
 &+ \frac{3}{2\sqrt{2\pi}} \int_0^{\infty} \log_2(x) x^{-3/2} \exp\left(-\frac{1}{2x}\right) dx.
 \end{aligned}$$

De

$$\begin{aligned}
 \int_0^{\infty} x^{-5/2} \exp\left(-\frac{1}{2x}\right) dx &= 2^{5/2-1} \int_0^{\infty} t^{5/2} \exp(-t) dt \\
 &= 2^{3/2} \Gamma\left(\frac{3}{2}\right) = 2^{3/2} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = 2^{3/2} \frac{1}{2} \sqrt{\pi} = \sqrt{2\pi},
 \end{aligned}$$

resulta que

$$\begin{aligned}
 \mathcal{H}(X_L) &= \frac{\log_2 \pi e + 1}{2} + \frac{3}{2\sqrt{2\pi}} \int_0^{\infty} \log_2(x) x^{-3/2} \exp\left(-\frac{1}{2x}\right) dx \\
 &= \frac{\log_2 \pi e + 1}{2} + I_1 = \log_2 \sqrt{\pi} + \frac{\log_2 e}{2} + \frac{1}{2} + I_1.
 \end{aligned}$$

Dada uma variável aleatória X com suporte em \mathbb{R}^+ , a sua distância de Kullback-Leibler a X_L é

$$\begin{aligned}
\mathcal{D}(f_X; f_{X_L}) &= \int_0^{+\infty} \log_2 \left[\frac{f_X(x)}{f_{X_L}(x)} \right] f_X(x) dx \\
&= -\mathcal{H}(X) - \int_0^{+\infty} \log_2 \left[\frac{\exp\left(-\frac{1}{2x}\right)}{\sqrt{2\pi x^3}} \right] f_X(x) dx \\
&= -\mathcal{H}(X) - \int_0^{+\infty} \log_2 \left[\exp\left(-\frac{1}{2x}\right) \right] f_X(x) dx + \int_0^{+\infty} \log_2 \left[\sqrt{2\pi x^3} \right] f_X(x) dx \\
&= -\mathcal{H}(X) + \frac{\log_2 e}{2} \int_0^{+\infty} \frac{1}{x} f_X(x) dx + \log_2 \sqrt{2\pi} + \int_0^{+\infty} \log_2 x^{3/2} f_X(x) dx \\
&= -\mathcal{H}(X) + \frac{\log_2 e}{2} \mathbb{E} \left(\frac{1}{X} \right) + \log_2 \sqrt{2\pi} + \mathbb{E}(\log_2 X^{3/2}) \\
&= -\mathcal{H}(X) + \frac{\log_2 e}{2} \mathbb{E} \left(\frac{1}{X} \right) + \mathcal{H}(X_L) - \frac{\log_2 e}{2} - I_1 + \mathbb{E}(\log_2 X^{3/2}).
\end{aligned}$$

onde $I_1 = \frac{3}{2\sqrt{2\pi}} \int_0^{\infty} \log_2(x) x^{-3/2} \exp\left(-\frac{1}{2x}\right) dx$.

De $\mathcal{D}(f_X; f_{X_L}) \geq 0$, resulta

$$\frac{\log_2 e}{2} \mathbb{E} \left(\frac{1}{X} \right) + \mathcal{H}(X_L) - \frac{\log_2 e}{2} - I_1 + \mathbb{E}(\log_2 X^{3/2}) \geq \mathcal{H}(X).$$

Embora nos dois casos apresentados anteriormente a distância de Kullback-Leiber não seja uma “expressão elegante”, não se conseguindo estabelecer para a Lévy e para a Cauchy resultados análogos aos da Gaussiana. Mas tem-se que

$$\mathcal{H}(Y_C) - \mathbb{E}[\log_2(1 + Y_C^2)] \geq \mathcal{H}(X_C) + \mathbb{E}[\log_2(1 + X_C^2)]$$

e

$$\frac{\log_2 e}{2} \mathbb{E} \left(\frac{1}{X} \right) + \mathcal{H}(X_L) - \frac{\log_2 e}{2} - I_1 + \mathbb{E}(\log_2 X^{3/2}) \geq \mathcal{H}(X).$$

3.2 Distribuições com Entropia Máxima

Enunciamos de seguida, sem demonstrar, dois lemas que serão necessários mais tarde (detalhes sobre a demonstração dos lemas podem ser encontrados em Linnik e Kagan, 1973, pp. 408.).

Lema 3.2.1. *Sejam $\sum a_i$ e $\sum b_i$ séries convergentes de termos positivos tais que $\sum a_i \geq \sum b_i$. Então*

$$\sum a_i \log \left(\frac{b_i}{a_i} \right) \leq 0.$$

Lema 3.2.2. *Sejam f e g funções integráveis não negativas com respeito à medida μ , e \mathcal{S} o conjunto onde $f > 0$. Se $\int_{\mathcal{S}} (f - g) d\mu \geq 0$, então*

$$\int_{\mathcal{S}} f \log \left(\frac{f}{g} \right) d\mu \geq 0, \tag{3.1}$$

com a igualdade a verificar-se se e só se $f = g$ quase certamente (qc) para a medida μ .

Seja X uma variável aleatória que toma valores num dado intervalo $]a, b[$, e com função densidade de probabilidade f_X . Recordemos que a entropia de X é por definição

$$- \int_a^b f_X(x) \log(f_X(x)) dx.$$

A sùmula dos resultados sobre entropia máxima encontram-se num teorema de Kagan-Linnik-Rao (1965) pp. 408-410 que abaixo transcrevemos:

Teorema 3.2.1. *Seja X uma variável aleatória com densidade*

$$f_X(x) > 0 \text{ para } x \in]a, b[\text{ e } f_X(x) = 0 \text{ caso contrário,} \tag{3.2}$$

com $a, b \in \overline{\mathbb{R}}$.

Sejam h_1, h_2, \dots funções integráveis em $]a, b[$ satisfazendo para as constantes dadas g_1, g_2, \dots as condições

$$\int_a^b h_i(x) f_X(x) dx = g_i, i = 1, 2, \dots \tag{3.3}$$

Então a entropia máxima é atingida pelas distribuições com densidade da forma

$$f_X(x) = \exp \{ a_0 + a_1 h_1(x) + a_2 h_2(x) + \dots \}, \tag{3.4}$$

(e só por elas), se existirem a_0, a_1, \dots tais que a densidade anterior satisfaz as condições (3.2) e (3.3).

Da aplicação do teorema anterior chegamos aos seguintes resultados:

Conjunto de valores da va	Restrições	Fdp correspondente à entropia máxima
$(0, 1)$		$f(x) = 1$ Uniforme
$(0, 1)$	$\mathbb{E}(\log(X)) = g_1$ $\mathbb{E}(\log(1 - X)) = g_2$	$f(x) = \frac{x^{m-1}(1-x)^{n-1}}{B(m,n)}$ Beta
$(0, +\infty)$	$\mathbb{E}(X) = g_1$	$f(x) = a e^{-ax}$ Exponencial
$(0, +\infty)$	$\mathbb{E}(X) = g_1$ $\mathbb{E}(\log(X)) = g_2$	$f(x) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}$ Gama
\mathbb{R}	$\mathbb{E}(X) = g_1$ $\mathbb{E}(X^2) = g_2$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ Gaussiana
\mathbb{R}	$\mathbb{E} X = g_1$	$f(x) = \frac{1}{2} e^{-a x }$ Laplace

A tabela anterior dá-nos uma série de distribuições de probabilidade com entropia máxima, mas de entre elas há uma que tem informação de Fisher mínima. Na secção seguinte vamos começar por definir informação de Fisher para de seguida ver qual, ou quais, têm informação de Fisher mínima.

3.3 Informação de Fisher

Um resultado de De Bruijn cf. Barron (1985) estabeleceu uma ligação notável entre entropia e informação de Fisher, nomeadamente permitindo observar que entropia máxima e informação mínima são atingidas sob as mesmas condições.

A informação de Fisher para uma família de distribuições $\{F_\theta, \theta \in \Delta\}$ no espaço $(\mathcal{X}, \mathcal{A})$ onde Δ é um intervalo no eixo real, com densidade $f(x, \theta)$ com respeito a alguma medida μ é definida como

$$\mathcal{J}(X, \theta) = \mathbb{E}_\theta \left(\frac{\partial \log f(X, \theta)}{\partial \theta} \right)^2 = \int_{\mathcal{X}} \left(\frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) d\mu$$

sob a hipótese que o lado direito existe.

Considere uma família de distribuições em \mathbb{R} , dependendo de um parâmetro de localização θ e dada pelas densidades $f(x - \theta)$ com respeito à medida de Lebesgue. Assuma-se que

- i. f tem derivadas contínuas;
- ii. $\int_{\mathbb{R}} x^2 f(x) dx < \infty$
- iii. $|x| f(x) \rightarrow 0$ quando $|x| \rightarrow \infty$.

A informação de Fisher para a família de densidades $f(x - \theta)$ é igual a

$$\mathcal{J}(X, \theta) = \int_{\mathbb{R}} \left(\frac{\partial \log f(x - \theta)}{\partial \theta} \right)^2 f(x - \theta) dx.$$

Facilmente se vê que

$$\mathcal{J}(X, \theta) = \int_{\{f(x) > 0\}} \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx = \mathcal{J}(0) := \mathcal{J}(X).$$

No caso de X ser uma variável aleatória cuja função densidade de probabilidade $f_X(x|\theta_1, \theta_2, \dots, \theta_n)$ é diferenciável em relação a $\theta_1, \theta_2, \dots, \theta_n$, a informação de Fisher é a matriz cuja entrada (i, j) é definida por:

$$\mathcal{J}_{ij}(X) = \int_{\mathbb{R}} \frac{1}{f_X(x)} \frac{\partial}{\partial \theta_i} f_X(x) \frac{\partial}{\partial \theta_j} f_X(x) dx, \quad i, j = 1, \dots, n,$$

e permite avaliar a que ponto são salientes os picos da função log-verossimilhança.

No caso particularmente importante de θ ser um parâmetro de localização, isto é, se $f_X(x|\theta) = f_X(x - \theta)$, a definição de informação de Fisher pode ser apresentada de uma forma mais manipulável:

Definição 3.3.1. *Seja X uma variável aleatória cuja função densidade de probabilidade $f_X(x|\theta) = f_X(x - \theta)$ tem derivada contínua. A função de pontuação (score) é*

$$\rho_X(X) = \frac{f'_X(x)}{f_X(x)} = \frac{d}{dx} \ln f_X(x)$$

e a informação de Fisher relativamente ao parâmetro de localização θ é

$$\mathcal{J}(X, \theta) = \mathbb{E} [\rho_X^2(X)].$$

Note-se que,

$$\begin{aligned} \mathcal{J}(X) &= \mathbb{E} [\rho_X^2(X)] = \int \frac{(f'_X(x))^2}{f_X(x)} dx \\ &= \int_{\mathbb{R}} \left(\frac{f'_X(x)}{f_X^{1/2}(x)} \right)^2 dx = 4 \int_{\mathbb{R}} \left(\frac{d}{dx} f_X^{1/2}(x) \right)^2 dx. \end{aligned}$$

Verificando-se que

$$\rho_{\alpha X}(X) = \frac{\frac{1}{\alpha^2} f' \left(\frac{X}{\alpha} \right)}{\frac{1}{\alpha} f \left(\frac{X}{\alpha} \right)} = \frac{1}{\alpha} \frac{f' \left(\frac{x}{\alpha} \right)}{f \left(\frac{x}{\alpha} \right)} = \frac{1}{\alpha} \rho_X \left(\frac{X}{\alpha} \right)$$

e que

$$\begin{aligned} \mathcal{J}(\alpha X) &= \mathbb{E} [\rho_{\alpha X}^2(X)] = \mathbb{E} \left[\left(\frac{1}{\alpha} \rho_X \left(\frac{X}{\alpha} \right) \right)^2 \right] \\ &= \frac{1}{\alpha^2} \mathbb{E} \left[\rho_X^2 \left(\frac{X}{\alpha} \right) \right] = \frac{1}{\alpha} \mathbb{E} [\rho_X^2(X)] = \frac{\mathcal{J}(X)}{\alpha}. \end{aligned}$$

Assim, o efeito de estandardizar uma variável com valor médio μ e variância σ^2 é na função *score*:

$$\rho_{\frac{X-\mu}{\sigma}}(X) = \sigma \rho_X(\mu + \sigma X).$$

Definição 3.3.2. A informação estandardizada de Fisher, se X tiver função *score* ρ_X , valor médio μ e variância σ^2 é:

$$\mathcal{J}_{st}(X) = \sigma^2 \mathbb{E} \left[\left(\rho_X(X) + \frac{X - \mu}{\sigma^2} \right)^2 \right].$$

A racionalidade desta definição ficará plenamente visível quando discutirmos a desigualdade de Cramér-Rao.

Exemplos

1. $X \sim \text{Gaussiana}(\mu, \sigma)$

Neste caso a função *score* é dada por

$$\begin{aligned} \rho_X(X) &= \frac{f'_X(x)}{f_X(x)} = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \left(-\frac{(x-\mu)^2}{2\sigma^2} \right)'}{\frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right)} \\ &= -\frac{x - \mu}{\sigma^2}. \end{aligned}$$

e a informação de Fisher — $J(X)$

$$\begin{aligned} \mathcal{J}(X) &= \mathbb{E}(\rho_X^2(X)) = \mathbb{E} \left[\frac{(X - \mu)^2}{\sigma^4} \right] = \frac{1}{\sigma^4} \mathbb{E}(X^2 - 2\mu X + \mu^2) \\ &= \frac{1}{\sigma^4} (\sigma^2 + \mu^2 - 2\mu^2 + \mu^2) = \frac{1}{\sigma^2}, \end{aligned}$$

assim, a informação de Fisher estandardizada é:

$$\mathcal{J}_{st}(X) = \sigma^2 \mathbb{E} \left[\left(-\frac{X - \mu}{\sigma^2} + \frac{X - \mu}{\sigma^2} \right)^2 \right] = 0.$$

Note-se que

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left(-\frac{(x - \mu)^2}{2\sigma^2} \right) \\ &= \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left(-\frac{\mu^2}{2\sigma^2} \right) \exp \left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} \right) \\ &= \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left(-\frac{\mu^2}{2\sigma^2} \right) \exp \left[\int_0^x -\frac{u - \mu}{\sigma^2} du \right] \\ &= f_X(0) \exp \left[\int_0^x \rho_X(U) du \right] \end{aligned}$$

e conseqüentemente a função *score* é linear se e só se X for gaussiana. Esta caracterização das populações gaussianas é de grande utilidade.

2. $X \sim \text{Gama}(n, \frac{1}{\theta})$

Tem-se que a função densidade de probabilidade é dada por:

$$f_X(x) = \frac{\theta^n x^{n-1} e^{-\theta x}}{\Gamma(n)} \mathbf{I}_{(0,+\infty)}(x)$$

com $\mathbb{E}(X) = \frac{n}{\theta}$ e $\text{Var}(X) = \frac{n}{\theta^2}$.

Então a função *score*, para $x > 0$ é dada por:

$$\begin{aligned} \rho_X(x) &= \frac{\theta^n (n-1) x^{n-2} e^{-\theta x} + \theta^n x^{n-1} (-\theta) e^{-\theta x}}{\Gamma(n)} \\ &= \frac{\theta^n x^{n-1} e^{-\theta x} [(n-1)x^{-1} - \theta]}{\Gamma(n)} \\ &= \frac{\Gamma(n) \theta^n e^{-\theta x} x^{n-1} [(n-1)x^{-1} - \theta]}{\Gamma(n) \theta^n x^{n-1} e^{-\theta x}} = \frac{n-1}{x} - \theta. \end{aligned}$$

A informação de Fisher - $\mathcal{J}(X)$ é

$$\begin{aligned}\mathcal{J}(X) &= \mathbb{E}[\rho_X^2(X)] = \mathbb{E}\left[\frac{(n-1)^2}{X^2} - 2\theta\frac{n-1}{X} + \theta^2\right] \\ &= (n-1)^2\mathbb{E}\left[\frac{1}{X^2}\right] - 2\theta(n-1)\mathbb{E}\left[\frac{1}{X}\right] + \theta^2,\end{aligned}$$

tendo em conta que a transformada de Mellin para a função gama é dada por

$$\mathbb{E}(X^s) = \frac{\Gamma(n+s)}{\theta^s\Gamma(n)}, s > -n$$

tem-se que

$$\begin{aligned}\mathcal{J}(X) &= (n-1)^2\frac{\Gamma(n-2)}{\theta^{-2}\Gamma(n)} - 2\theta(n-1)\frac{\Gamma(n-1)}{\theta^{-1}\Gamma(n)} + \theta^2 \\ &= (n-1)^2\theta^2\frac{(n-3)!}{(n-1)!} - 2\theta^2(n-1)\frac{(n-2)!}{(n-1)!} + \theta^2 \\ &= \theta^2\left(\frac{n-1}{n-2} - 1\right) = \frac{\theta^2}{n-2}.\end{aligned}$$

A informação de Fisher estandardizada é:

$$\begin{aligned}\mathcal{J}_{st}(X) &= \frac{n}{\theta^2}\mathbb{E}\left[\left(\frac{n-1}{X} - \theta + \frac{\theta^2 X - \theta n}{n}\right)^2\right] \\ &= \frac{n}{\theta^2}\mathbb{E}\left[\left(\frac{n-1}{X} - \left(2\theta - \frac{\theta^2 X}{n}\right)\right)^2\right] \\ &= \frac{n(n-1)^2}{\theta^2}\frac{\Gamma(n-2)}{\theta^{-2}\Gamma(n)} - 4\frac{n(n-1)}{\theta}\frac{\Gamma(n-1)}{\theta^{-1}\Gamma(n)} + \\ &= +6n - 2 - 4\theta\frac{\Gamma(n+1)}{\theta\Gamma(n)} + \frac{\theta^2}{n}\frac{\Gamma(n+2)}{\theta^2\Gamma(n)} \\ &= \frac{n^2 - n}{n-2} - n - 1 = \frac{2}{n-2}.\end{aligned}$$

De seguida vamos definir distância da informação de Fisher. Para tal notemos que

$$\begin{aligned}
 \frac{d}{dx} \sqrt{\frac{f_X(x)}{f_Y(x)}} &= \frac{(f'_X(x) f - Y(x) - f_X(x) f'_Y(x)) \sqrt{f_Y(x)}}{2(f_Y(x))^2 \sqrt{f_X(x)}} \\
 \Rightarrow \left(\frac{d}{dx} \sqrt{\frac{f_X(x)}{f_Y(x)}} \right)^2 &= \frac{(f'_X(x))^2 f_Y^2(x) - 2f'_X(x) f_Y(x) f_X(x) f'_Y(x) + f_X^2(x) (f'_Y(x))^2}{4(f_Y(x))^3 f_X(x)} \\
 \Rightarrow f_Y(x) \left(\frac{d}{dx} \sqrt{\frac{f_X(x)}{f_Y(x)}} \right)^2 &= \frac{1}{4} \frac{(f'_X(x))^2}{f_X(x)} - \frac{1}{2} \frac{f'_X(x) f'_Y(x)}{f_Y(x)} + \frac{1}{4} \frac{f_X(x) (f'_Y(x))^2}{(f_Y(x))^2} \\
 &= \frac{1}{4} f_X(x) \left[\frac{(f'_X(x))^2}{f_X^2(x)} - 2 \frac{f'_X(x) f'_Y(x)}{f_Y(x) f_X(x)} + \frac{(f'_Y(x))^2}{(f_Y(x))^2} \right] \\
 &= \frac{1}{4} f_X(x) \left(\frac{f'_X(x)}{f_X(x)} - \frac{f'_Y(x)}{f_Y(x)} \right)^2 \\
 &= \frac{1}{4} f_X(x) \left[\frac{d}{dx} \ln f_X(x) - \frac{d}{dx} \ln f_Y(x) \right]^2.
 \end{aligned}$$

Assim, a distância da informação de Fisher é definida por

$$\begin{aligned}
 \mathcal{J}(X, Y) = \mathcal{J}(f_X, f_Y) &= \int_{\mathbb{R}} f_X(x) \left(\frac{d}{dx} \ln f_X(x) - \frac{d}{dx} \ln f_Y(x) \right)^2 dx \\
 &= 4 \int_{\mathbb{R}} f_Y(x) \left(\frac{d}{dx} \sqrt{\frac{f_X(x)}{f_Y(x)}} \right)^2 dx.
 \end{aligned}$$

Se $\mathbb{E}(X) = \mu$ e $\text{Var}(X) = \sigma^2$, então $\mathcal{J}(X) \geq \frac{1}{\sigma^2}$, com $\mathcal{J}(X) = \frac{1}{\sigma^2}$ se e só se $X \sim \text{Gaussiana}(\mu, \sigma)$. Esta famosa desigualdade de Cramér-Rao pode ser demonstrada de forma muito simples usando as propriedades da função *score*. De facto, por simples integração por partes, e supondo que a função g satisfaz as condições de regularidade necessárias (existência de derivada e $\lim_{x \rightarrow \pm\infty} f_X(x) g(x) = 0$)

$$\begin{aligned}
 \mathbb{E}[\rho_X(X) g(X)] &= \int_{\mathbb{R}} \rho_X(x) g(x) f_X(x) dx \\
 &= \int_{\mathbb{R}} \frac{f'_X(x)}{f_X(x)} g(x) f_X(x) dx \\
 &= \int_{\mathbb{R}} f'_X(x) g(x) dx = - \int_{\mathbb{R}} f_X(x) g'(x) dx = -\mathbb{E}[g'(X)]
 \end{aligned}$$

A igualdade anterior é conhecida como identidade de Stein. As funções lineares $g(x) = ax + b$ satisfazem as condições de regularidade necessárias, e conseqüentemente

$$\mathbb{E}[\rho_X(X)(aX + b)] = -\mathbb{E}[(aX + b)'] = -a.$$

Conclui-se imediatamente que com $a = \frac{1}{\sigma^2}$,

$$\begin{aligned} 0 &\leq \mathbb{E} \left[\left(\rho_X(X) + \frac{X - \mu}{\sigma^2} \right)^2 \right] \\ &= \mathbb{E}[\rho_X^2(X)] + \mathbb{E} \left[2 \rho_X(X) \left(\frac{X - \mu}{\sigma^2} \right) \right] + \mathbb{E} \left[\frac{X^2 - 2\mu X + \mu^2}{\sigma^4} \right] \\ &= \mathcal{J}(X) - 2\mathbb{E} \left[\left(\frac{X - \mu}{\sigma^2} \right)' \right] + \frac{1}{\sigma^4} (\mathbb{E}(X^2) - \mu^2) \\ &= \mathcal{J}(X) - \frac{1}{\sigma^2} \quad \Rightarrow \quad \mathcal{J}(X) \geq \frac{1}{\sigma^2}. \end{aligned}$$

Como veremos na secção seguinte a igualdade dá-se se e só se X for gaussiana.

Tem-se assim que as populações gaussianas têm:

1. entropia máxima, entre todas as com suporte em toda a recta real e com variância σ^2 ;
2. informação de Fisher mínima, de entre todas as que têm variância σ^2 e suporte na recta real.

A relação entre estas duas propriedades é a identidade de Bruijn: Se X , com variância 1 e $Z_\sigma \sim \text{Gaussiana}(0, \sigma)$ forem independentes de facto

$$\begin{aligned} \mathcal{D}(f_X, f_{Z_\sigma}) &= \frac{\log_2 e}{2} \int_0^\infty \left(\mathcal{J}(X + Z_\sigma) - \frac{1}{1 + \sigma} \right) d\sigma \\ &= \frac{\log_2 e}{2} \int_0^\infty \frac{\mathcal{J}_{st}(X + Z_\sigma)}{1 + \sigma} d\sigma = 0. \end{aligned}$$

Notemos que

$$\begin{aligned}
 \mathcal{J}_{st}(X + Z_\sigma) &= (1 + \sigma) \mathbb{E} \left[\left(\rho_{X+Z_\sigma}(X + Z_\sigma) + \frac{X + Z_\sigma - \mathbb{E}(X)}{(1 + \sigma)} \right)^2 \right] \\
 &= (1 + \sigma) \left[\mathbb{E}(\rho_{X+Z_\sigma}^2(X + Z_\sigma)) + 2 \mathbb{E} \left(\rho_{X+Z_\sigma}(X + Z_\sigma) \frac{X + Z_\sigma - \mathbb{E}(X)}{1 + \sigma} \right) \right] \\
 &\quad + \frac{1 + \sigma}{(1 + \sigma)^2} \mathbb{E} [(X + Z_\sigma)^2 - 2(X + Z_\sigma) \mathbb{E}(X) + (\mathbb{E}(X))^2] \\
 &= (1 + \sigma) \left[\mathcal{J}(X + Z_\sigma) - \frac{2}{1 + \sigma} + \frac{1}{(1 + \sigma)^2} (1 + \sigma) \right] \\
 &= (1 + \sigma) \mathcal{J}(X + Z_\sigma) - 1.
 \end{aligned}$$

Donde

$$\frac{\mathcal{J}_{st}(X + Z_\sigma)}{1 + \sigma} = \mathcal{J}(X + Z_\sigma) - \frac{1}{1 + \sigma}.$$

logo

$$\frac{\log_2 e}{2} \int_0^\infty \left(\mathcal{J}(X + Z_\sigma) - \frac{1}{1 + \sigma} \right) d\sigma = \frac{\log_2 e}{2} \int_0^\infty \frac{\mathcal{J}_{st}(X + Z_\sigma)}{1 + \sigma}.$$

A função integranda é não negativa, e na demonstração da desigualdade de Cramér-Rao (Capítulo 2, secção 1.2) apontou-se que seria zero se e só se $X + Z_\sigma$ fosse gaussiana; ora sabe-se — teorema Lévy-Cramér sobre a decomposição da gaussiana em parcelas independentes — que isso acontece se e só se X for gaussiana.

3.3.1 Distribuições com Informação de Fisher Mínima

De seguida enunciamos e provamos dois teoremas que afirmam que as distribuições gaussianas e gama têm informação de Fisher mínima.

Teorema 3.3.1. *Na classe de todas as densidades com variância finita igual a σ^2 e satisfazendo as condições*

- i. f tem derivadas contínuas;*
- ii. $\int_{\mathbb{R}} x^2 f(x) dx < \infty$*
- iii. $|x| f(x) \rightarrow 0$ quando $|x| \rightarrow \infty$.*

A informação mínima é atingida para para distribuições gaussianas.

Demonstração

Sem perda de generalidade vamos tomar $\int_{\mathbb{R}} f(x) dx = 0$. Seja $A^+ = \{x : f(x) > 0\}$. Integrando por partes e tendo em conta as condições (i)-(iii) obtemos

$$\int_{A^+} x f'(x) dx = - \int_{A^+} f(x) dx$$

donde

$$\int_{A^+} x f'(x) dx = -1.$$

Pela desigualdade de Cauchy-Schwarz, tem-se

$$\begin{aligned} 1 &= \left(\int_{A^+} x f'(x) dx \right)^2 = \left(\int_{A^+} x \frac{f'(x)}{f(x)} f(x) dx \right)^2 \\ &\leq \int_{A^+} \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx \cdot \int_{A^+} x^2 f(x) dx \\ &\Rightarrow \int_{A^+} \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx \cdot \int_{A^+} x^2 f(x) dx \geq 1 \\ &\Leftrightarrow \int_{A^+} \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx \cdot \sigma^2 \geq 1 \\ &\Leftrightarrow \int_{A^+} (\log(f(x)))^2 f(x) dx \geq \frac{1}{\sigma^2} \\ &\Leftrightarrow \mathcal{J}(X) \geq \frac{1}{\sigma^2}, \quad \text{onde } X \text{ é uma variável aleatória com densidade } f. \end{aligned}$$

A igualdade é atingida quando para alguma constante c

$$\frac{f'(x)}{f(x)} = c, x \quad \text{qc em } A^+ \tag{3.5}$$

(com respeito à medida de Lebesgue). Por causa da continuidade de f' concluímos que $\frac{f'(x)}{f(x)} = cx$, para $x \in A^+$. Integrando (3.5) obtemos

$$\begin{aligned} \int \frac{f'(x)}{f(x)} dx &= \int cx dx \\ \Leftrightarrow \log(f(x)) &= \frac{cx^2}{2} + B \\ \Leftrightarrow f(x) &= A \exp\left(\frac{cx^2}{2}\right), \forall x \in A^+, \end{aligned} \tag{3.6}$$

onde $A = e^B$. Mas, $A \exp\left(\frac{cx^2}{2}\right) \neq 0$ para x real; então, a continuidade de f , juntamente com (3.5) implica

$$f(x) = A \exp\left(\frac{cx^2}{2}\right), \forall x$$

onde $c = -\frac{1}{\sigma^2}$ e $A = \frac{1}{\sigma \sqrt{2\pi}}$. □

Considere-se agora uma família de distribuições na semi-recta \mathbb{R}_+ dependendo de um parâmetro de escala $\sigma \in \mathbb{R}_+$ e dada pelas densidades $\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$ com respeito à medida de Lebesgue. Suponhamos que f satisfaz as condições (i) e (ii) enunciadas no teorema anterior e ainda a condição

iv. $x^2 f(x) \rightarrow 0$ quando $x \rightarrow \infty$, $x f(x) \rightarrow 0$ quando $x \rightarrow 0$.

A informação de Fisher associada às variáveis aleatórias com densidades dadas por $\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$ é igual a

$$\begin{aligned} \mathcal{J}(\sigma) &= \int_{\{f(x/\sigma) > 0\}} \left[\frac{\partial}{\partial \sigma} \log \left(\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) \right) \right]^2 \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) dx \\ &= \int_{\{f(x/\sigma) > 0\}} \left(\frac{-\frac{1}{\sigma^2} f\left(\frac{x}{\sigma}\right) - \frac{1}{\sigma} f'\left(\frac{x}{\sigma}\right) \frac{x}{\sigma^2}}{\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)} \right)^2 \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) dx \\ &= \int_{\{f(x/\sigma) > 0\}} \left(-\frac{1}{\sigma} - \frac{x}{\sigma^2} \frac{f'\left(\frac{x}{\sigma}\right)}{f\left(\frac{x}{\sigma}\right)} \right)^2 \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) dx \\ &= \int_{\{f(x/\sigma) > 0\}} \frac{1}{\sigma^2} \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) dx + \int_{\{f(x/\sigma) > 0\}} \frac{2x}{\sigma^3} \frac{f'\left(\frac{x}{\sigma}\right)}{f\left(\frac{x}{\sigma}\right)} \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) dx + \\ &\quad + \int_{\{f(x/\sigma) > 0\}} \left(\frac{x}{\sigma^2} \frac{f'\left(\frac{x}{\sigma}\right)}{f\left(\frac{x}{\sigma}\right)} \right)^2 \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) dx \\ &= \frac{1}{\sigma^2} \left[\int_{A^+} \left(y \frac{f'(y)}{f(y)} \right)^2 dx - 1 \right] \\ &= \frac{1}{\sigma^2} \mathcal{J}(1). \end{aligned}$$

onde $A^+ = \{x : f(x/\sigma) > 0\}$.

Teorema 3.3.2. *Na classe de todas as densidades com momentos $\alpha_i = \int_0^\infty x^i f(x) dx$, $i = 1, 2, 3$, satisfazendo as condições (i), (ii) e (iv), anteriormente referidas, a informação mínima é atingido pelas distribuições gama.*

Demonstração

Temos que

$$\int_{A^+} x f'(x) dx = -1 \quad \text{e} \quad \int_{A^+} x^2 f'(x) dx = -2\alpha_1.$$

Multiplicando a primeira relação por μ , a segunda por $-\lambda$ e somando obtém-se

$$\begin{aligned} \mu \int_{A^+} x f'(x) dx + \lambda \int_{A^+} x^2 f'(x) dx &= -\mu + 2\lambda\alpha_1 \\ \Leftrightarrow \int_{A^+} x(\mu - \lambda x) f'(x) dx &= (2\lambda\alpha_1 - \mu). \end{aligned}$$

Aplicando a desigualdade de de Cauchy-Schwarz, obtém-se

$$\begin{aligned} (2\lambda\alpha_1 - \mu)^2 &= \left\{ \int_{A^+} \frac{f'(x)}{f(x)} x(\mu - \lambda x) f(x) dx \right\}^2 \\ &\leq \int_{A^+} \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx \cdot \int_{A^+} x(\mu - \lambda x)^2 f(x) dx. \end{aligned}$$

Então, para quaisquer λ e μ

$$\begin{aligned} \int_{A^+} \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx &\geq \frac{(\mu - 2\lambda\alpha_1)^2}{\int_{A^+} x(\mu - \lambda x)^2 f(x) dx} = \frac{(\mu\alpha - 2\lambda\alpha_1)^2}{\mu^2\alpha_1 - 2\mu\lambda\alpha_2 + \lambda^3\alpha_3} \\ \Leftrightarrow \mathcal{I}(X) &\geq \frac{(\mu - 2\lambda\alpha_1)^2}{\mu^2\alpha_1 - 2\mu\lambda\alpha_2 + \lambda^2\alpha_3}, \end{aligned}$$

onde para λ e μ dados, a igualdade verifica-se quando para alguma constante c a seguinte relação

$$\frac{x f'(x)}{f(x)} = c(\mu - \lambda x)$$

se verifica quase certamente em A^+ . Por causa da continuidade de f a relação anterior é satisfeita para todo o x em A^+ . Absorvendo a constante c em μ e λ e integrando a igualdade anterior obtemos

$$f(x) = k e^{-\lambda x} x^\mu, \quad x \in A^+. \quad (3.7)$$

Como $e^{-\lambda x} x^\mu \neq 0, \forall x \in \mathbb{R}_+, \mu > 0$, a continuidade de f garante a validade de (3.7) para todo o x em \mathbb{R}_+ . Fazendo $\mu = \nu - 1$ e determinando k pelas condições de normalidade chegamos à forma usual da densidade Gama

$$f(x) = \frac{\lambda^\nu}{\Gamma(\nu)} e^{-\lambda x} x^{\nu-1}, x > 0.$$

Se tomarmos $\lambda = \frac{\alpha_1}{\alpha_2 - \alpha_1^2}$ e $\nu = \frac{\alpha_1^2}{\alpha_2 - \alpha_1^2}$ os momentos da distribuição são α_1 e α_2 . \square

3.4 Distância de Kullback-Leibler e Informação de Fisher — Algumas Aplicações

Denote-se $Z_\sigma \sim \text{Gaussiana}(0, \sigma)$, com função densidade de probabilidade

$$f_{Z_\sigma}(x, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma}} \mathbb{I}_{\mathbb{R}}(x).$$

Esta densidade verifica a equação diferencial do calor

$$\frac{\partial}{\partial \sigma} \phi(x, \sigma) = \frac{1}{2} \frac{\partial^2}{\partial x^2} \phi(x, \sigma).$$

Considerando $Y_t = X + Z_\sigma$, com X e Z_σ independentes, e considerando que

$$\mathcal{H}(Y_\sigma) = - \int_{\mathcal{S}} \log_2(f_{Y_\sigma}(y)) f_{Y_\sigma}(y) dy$$

temos que

$$f_{Y_\sigma}(x) = \int f_{Y_\sigma}(x - t, \sigma) f_X(t) dt$$

pelo que,

$$\begin{aligned} \frac{\partial}{\partial \sigma} \mathcal{H}(Y_\sigma) &= - \int_{\mathcal{S}} \frac{\partial}{\partial \sigma} [\log_2(f_{Y_\sigma}(y)) f_{Y_\sigma}(y)] dy \\ &= - \int_{\mathcal{S}} \frac{\frac{\partial}{\partial \sigma} f_{Y_\sigma}(y)}{\log_2 f_{Y_\sigma}(y)} f_{Y_\sigma}(y) dy - \int_{\mathcal{S}} \log_2(f_{Y_\sigma}(y)) \frac{\partial}{\partial \sigma} f_{Y_\sigma}(y) dy \\ &= - \frac{1}{2 \log 2} \int_{\mathcal{S}} \frac{\partial^2}{\partial y^2} f_{Y_\sigma}(y) dy - \int_{\mathcal{S}} \log_2(f_{Y_\sigma}) \frac{1}{2} \frac{\partial^2}{\partial y^2} f_{Y_\sigma}(y) dy \\ &= - \frac{1}{2} \int_{\mathcal{S}} \log_2(f_{Y_\sigma}) \frac{\partial^2}{\partial y^2} f_{Y_\sigma}(y) dy, \end{aligned}$$

(3.8)

onde \mathcal{S} é o suporte da função.

Da definição de \mathcal{J}

$$\begin{aligned}\mathcal{J}(Y_\sigma) &= \int_{\mathcal{S}} \left(\frac{f'_{Y_\sigma}(x)}{f_{Y_\sigma}(x)} \right)^2 f_{Y_\sigma}(x) dx = \int_{\mathcal{S}} \frac{f'_{Y_\sigma}(x)}{f_{Y_\sigma}(x)} \frac{f'_{Y_\sigma}(x)}{f_w(x)} f_{Y_\sigma}(x) dx \\ &= \int_{\mathcal{S}} \frac{d}{dx} \ln(f_{Y_\sigma}(x)) f'_{Y_\sigma}(x) dx,\end{aligned}$$

integrando por partes, e admitindo comportamento regular de f_{Y_σ} em $\pm\infty$

$$\mathcal{J}(Y_\sigma) = - \int \ln(f_{Y_\sigma}(x)) f''_{Y_\sigma}(x) dx = - \log 2 \int \log_2(f_{Y_\sigma}(x)) f''_{Y_\sigma}(x) dx. \quad (3.9)$$

Comparando (3.8) e (3.9), temos que

$$\frac{\partial}{\partial \sigma} \mathcal{H}(Y_\sigma) = \frac{1}{\log 2} \mathcal{J}(Y_\sigma) = \frac{1}{\log 2} \mathcal{J}(X + Z_\sigma).$$

Integrando entre zero e infinito

$$\begin{aligned}\int_0^\infty \frac{\partial}{\partial \sigma} \mathcal{H}(Y_\sigma) d\sigma &= \frac{1}{\log 2} \int_0^\infty \mathcal{J}(X + Z_\sigma) d\sigma \\ \Leftrightarrow \lim_{t \rightarrow \infty} \mathcal{H}(Y_t) - \mathcal{H}(Y_0) &= \frac{1}{\log 2} \int_0^\infty \mathcal{J}(X + Z_\sigma) d\sigma.\end{aligned}$$

Mas, $Y_0 = X + Z_0$ com Z_0 uma gaussiana com média e variância nula, logo é degenerada em zero e $Y_0 = X$.

Sejam agora $Z_t \sim \text{Gaussiana}(0, \sqrt{t})$ (\sqrt{t} desvio padrão de Z_t), e X com $\mathbb{E}(X) = 0$

e $\text{Var}(X) = 1$. Como já vimos

$$\begin{aligned}
\mathcal{D}(f_X, f_{Z_t}) &= \int \log_2 \left(\frac{f_X(x)}{f_{Y_t}(x)} \right) f_X(x) dx \\
&= \int \log_2(f_X(x)) f_X(x) dx - \int \log_2(f_{Y_t}(x)) f_X(x) dx \\
&= -\mathcal{H}(X) - \int \log_2 \left(\frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \right) f_X(x) dx \\
&= -\mathcal{H}(X) + \int \log_2(\sqrt{2\pi t}) f_X(x) dx - \int \log_2 e^{-\frac{x^2}{2t}} f_X(x) dx \\
&= -\mathcal{H}(X) + \frac{\log_2(2\pi t)}{2} + \log_2 e \int \frac{x^2}{2t} f_X(x) dx \\
&= -\mathcal{H}(X) + \frac{\log_2(2\pi t)}{2} + \frac{\log_2 e}{2t} \mathbb{E}(X^2) \\
&= -\mathcal{H}(X) + \frac{\log_2(2\pi t)}{2} + \frac{\log_2 e}{2t}, \quad \text{pois } \mathbb{E}(X^2) = 1.
\end{aligned}$$

Para $Y_t = X + Z_t$, que tem valor médio $\mathbb{E}(Y_t) = 0$ e variância $\text{Var}(Y_t) = 1 + t$, vem

$$\begin{aligned}
\mathcal{D}(f_{Y_t}, f_{Z_t}) &= -\mathcal{H}(Y_t) + \frac{\log_2(2\pi(1+t))}{2} + \frac{\log_2 e}{2(1+t)} \int_{\mathcal{S}} y^2 f_{Y_t}(y) dy \\
&= -\mathcal{H}(Y_t) + \frac{\log_2(2\pi(1+t))}{2} + \frac{\log_2 e}{2(1+t)}.
\end{aligned}$$

Comparando a expressão de $\mathcal{D}(f_X, f_{Z_t})$ com $\mathcal{D}(f_{Y_t}, f_{Z_t})$, verifica-se que apenas diferem nas parcelas em que intervém a variância $(1+t)$ de Y_t .

3.4.1 Desigualdade de Cramér-Rao Generalizada

Pretendemos através da informação de Fisher chegar a uma das formas da desigualdade de Cramér-Rao. Consideremos duas variáveis aleatórias X e Y , ambas com suporte $\mathcal{S} = \{x : f_X(x) > 0\}$, e respectivas funções de densidade de probabilidade, f_X e f_Y . Consideremos ainda K uma função regular e $\alpha \in \mathbb{R}$. Temos

$$\int_{\mathcal{S}} \left\{ \frac{d}{dx} \ln \left[\frac{f_X(x)}{f_Y(x)} \right] + \alpha K(x) \right\} f_X(x) dx \geq 0.$$

Recordando a definição de informação de Fisher (cf, pp. 12)

$$\mathcal{I}(f_X, f_Y) = \int_{\mathcal{S}} \left\{ \frac{d}{dx} \ln \left[\frac{f_X(x)}{f_Y(x)} \right] \right\}^2 f_X(x) dx$$

obtemos

$$\begin{aligned}
& \int_{\mathcal{S}} \left\{ \frac{d}{dx} \ln \left[\frac{f_X(x)}{f_Y(x)} \right] + \alpha K(x) \right\} f_X(x) dx \\
&= \int_{\mathcal{S}} \left\{ \frac{d}{dx} \ln \left[\frac{f_X(x)}{f_Y(x)} \right] + \right\}^2 f_X(x) dx + \int_{\mathcal{S}} 2 \frac{d}{dx} \ln \left[\frac{f_X(x)}{f_Y(x)} \right] \alpha K(x) f_X(x) dx + \int_{\mathcal{S}} [\alpha K(x)]^2 f_X(x) dx \\
&= \mathcal{I}(f_X, g_Y) + 2\alpha \int_{\mathcal{S}} \frac{d}{dx} \left[\frac{f_X(x)}{f_Y(x)} \right] K(x) f_X(x) dx + \alpha^2 \mathbb{E} \{ [K(X)]^2 \}.
\end{aligned} \tag{3.10}$$

Mas

$$\begin{aligned}
& \int_{\mathcal{S}} \frac{d}{dx} \left[\frac{f_X(x)}{f_Y(x)} \right] K(x) f_X(x) dx \\
&= \int_{\mathcal{S}} \frac{f'_X(x)}{f_X(x)} K(x) f_X(x) dx - \int_{\mathcal{S}} \frac{f'_Y(x)}{f_Y(x)} K(x) f_X(x) dx \\
&= \int_{\mathcal{S}} f'_X(x) K(x) dx - \int_{\mathcal{S}} \rho_Y(x) K(x) f_X(x) dx,
\end{aligned}$$

onde $\rho_Y = \frac{f'_Y}{f_Y}$, recorde-se, é a função score de Y . Integrando o primeiro integral por partes vem (da regularidade K)

$$\begin{aligned}
& \int_{\mathcal{S}} \frac{d}{dx} \left[\frac{f_X(x)}{f_Y(x)} \right] K(x) f_X(x) dx \\
&= [f_X(x) K(x)]_{\mathcal{S}} - \int_{\mathcal{S}} f_X(x) k'(X) dx - \mathbb{E}[\rho_Y(X) K(X)] \\
&= - \int_{\mathcal{S}} f_X(x) K'(X) dx - \mathbb{E}[\rho_Y(X) K(X)] \\
&= -\mathbb{E}[K'(X)] - \mathbb{E}[\rho_Y(x) K(X)].
\end{aligned}$$

Substituindo (3.10) obtemos

$$\begin{aligned}
& \int_{\mathcal{S}} \left\{ \frac{d}{dx} \ln \left[\frac{f_X(x)}{f_Y(x)} \right] + \alpha K(x) \right\}^2 f_X(x) dx \\
&= \mathcal{I}(f_X, f_Y) - 2\alpha \mathbb{E}[K'(X)] - 2\alpha \mathbb{E}[\rho_Y(X) K(X)] + \alpha^2 \mathbb{E}\{[K(X)]^2\},
\end{aligned}$$

de onde resulta

$$\mathcal{I}(f_X, f_Y) \geq 2\alpha \left\{ \mathbb{E}[K'(X)] + \mathbb{E}[\rho_Y(X) K(X)] \right\} - \alpha^2 \mathbb{E}\{[K(X)]^2\}.$$

O segundo membro é máximo para

$$\alpha = \frac{\mathbb{E}[K'(X)] + \mathbb{E}[\rho_Y(X) K(X)]}{\mathbb{E}\{[K(X)]^2\}}.$$

De facto,

$$\begin{aligned} \frac{d}{d\alpha} (2\alpha \{\mathbb{E}[K'(X)] + \mathbb{E}[\rho_Y(X) K(X)]\} - \alpha^2 \mathbb{E}\{[K^2(X)]\}) &= 0 \Leftrightarrow \\ 2\mathbb{E}[K'(X)] + 2\mathbb{E}[\rho_Y(X) K(X)] - 2\alpha \mathbb{E}\{[K(X)]^2\} &= 0 \Leftrightarrow \\ \frac{\mathbb{E}[K'(X)] + \mathbb{E}[\rho_Y(X) K(X)]}{\mathbb{E}\{[K(X)]^2\}} &= \alpha \end{aligned}$$

e

$$\frac{d^2}{d\alpha^2} (2\{\mathbb{E}[K'(X)] + \mathbb{E}[\rho_Y(X) K(X)]\} - \alpha^2 \mathbb{E}\{[K(X)]^2\}) = -2\mathbb{E}\{[K(X)]^2\} < 0.$$

Consequentemente,

$$\begin{aligned} \mathcal{I}(f_X, g_Y) &\geq 2 \frac{\left\{ \mathbb{E}[K'(X)] + \mathbb{E}[\rho_Y(X) K(X)] \right\}^2}{\mathbb{E}[K^2(X)]} - \left\{ \frac{\mathbb{E}[K'(X)] + \mathbb{E}[\rho_Y(X) K(X)]}{\mathbb{E}\{[K(X)]^2\}} \right\}^2 \mathbb{E}\{[K(X)]^2\} \\ &= \frac{\left\{ \mathbb{E}[K'(X)] + \mathbb{E}[\rho_Y(X) K(X)] \right\}^2}{\mathbb{E}\{[K(X)]^2\}}. \end{aligned}$$

Seja X uma variável aleatória com $\mathbb{E}(X) = 0$, $\text{Var}(X) = \sigma^2$ e denote-se $Z_\sigma \sim \text{Gaussiana}(0, \sigma)$. Tem-se então

$$\mathcal{J}_{st}(X) = \sigma^2 \mathcal{J}(X, Z_\sigma) \geq \frac{\sigma^2 \left(\mathbb{E} \left(K'(X) - \frac{X}{\sigma^2} K(X) \right) \right)^2}{\mathbb{E}(K^2(X))}, \quad (3.11)$$

pois $\rho_{Z_\sigma}(X) = -\frac{X}{\sigma^2}$.

Note-se que a desigualdade de Cramér-Rao é $\mathcal{J}(X) \geq \frac{1}{\sigma^2}$, ou seja,

$$\sigma^2 \mathcal{J}(X) - 1 \geq 0 \Rightarrow \sigma^2 \mathcal{J}(X) \geq 0 \Leftrightarrow \mathcal{J}_{st}(X) \geq 0.$$

No caso de tomarmos $K(X) = X$ vem

$$\begin{aligned} \mathcal{J}_{st}(X) &\geq \frac{\sigma^2 \left(\mathbb{E} \left(1 - \frac{X^2}{\sigma^2} \right) \right)^2}{\mathbb{E}(X^2)} = \frac{\sigma^2 \left(1 - \frac{1}{\sigma^2} \mathbb{E}(X^2) \right)^2}{\sigma^2} \\ &\Leftrightarrow \mathcal{J}_{st}(X) \geq \frac{\sigma^2 \left(1 - \frac{1}{\sigma^2} \sigma^2 \right)^2}{\sigma^2} = 0, \end{aligned}$$

isto é, a expressão (3.11) é uma forma generalizada da desigualdade de Cramér-Rao.

Seja $S_n = \sum_{k=1}^n X_k$, e denote-se $Y_n = \frac{S_n}{\sqrt{n}}$. Sem perda de generalidade, vamos supor que $\mathbb{E}(X_k) = 0$ e que $\text{Var}(X_k) = \sigma^2$, e em (3.11) vamos usar $K(Y_n) = Y_n^2 - \sigma^2$. Então, admitindo que $\mathbb{E}(X_k^4) < \infty$,

$$\begin{aligned} \mathbb{E} \left[K'(Y_n) - \frac{Y_n}{\sigma^2} K(Y_n) \right] &= \mathbb{E} \left[2Y_n - \frac{Y_n^3}{\sigma^2} + \frac{Y_n}{\sigma^2} \sigma^2 \right] \\ &= \mathbb{E} \left[Y_n - \frac{Y_n^3}{\sigma^2} \right] = \mathbb{E} \left[-\frac{Y_n^3}{\sigma^3} \right] = \frac{-m_3(Y_n)}{\sigma^2}, \end{aligned}$$

onde $m_k(X) = \mathbb{E}(Y_n^k)$. E,

$$\begin{aligned} \mathbb{E} [K^2(Y_n)] &= \mathbb{E} [Y_n^4 - 2\sigma^2 Y_n^2 + \sigma^4] = \mathbb{E}(Y_n^4) - 2\sigma^2 \mathbb{E}(Y_n^2) + \sigma^4 \\ &= m_4(Y_n) - 2\sigma^4 + \sigma^4 = m_4(Y_n) - \sigma^4, \end{aligned}$$

assim

$$\begin{aligned} \mathcal{J}_{st}(Y_n) &\geq \frac{\sigma^2 \left(-\frac{m_3(Y_n)}{\sigma^2} \right)^2}{m_4(Y_n) - \sigma^4} = \frac{\left(-\frac{m_3(Y_n)}{\sigma^2} \right)^2}{\frac{m_4(Y_n)}{\sigma^4} - 1} \\ &= \frac{\gamma_1^2(Y_n)}{\gamma_2(Y_n) + 2}. \end{aligned}$$

Como⁽¹⁾

$$m_3(Y_n) = \frac{m_3(X)}{\sqrt{n}} \quad \text{e} \quad m_4(Y_n) = \frac{m_4(X)}{n} + 3\sigma^4 \frac{n-1}{n}$$

vem

$$\begin{aligned} \mathcal{J}_{st}(Y_n) &\geq \frac{\left(\frac{m_3(X)}{\sigma^3 \sqrt{n}} \right)^2}{\frac{m_4(X)}{n\sigma^4} + \frac{3\sigma^4(n-1)}{\sigma^4 n} - 1} = \frac{\left(\frac{m_3(X)}{\sigma^3} \right)^2}{\frac{m_4(X)}{\sigma^4} + 3n - 3 - n} \\ \mathcal{J}_{st}(Y_n) &\geq \frac{\gamma_1^2(X)}{\gamma_2(X) + 2n} \\ n \mathcal{J}_{st}(Y_n) &\geq \frac{\gamma_1^2(X)}{\left(2 + \frac{\gamma_2(X)}{n} \right)}. \end{aligned}$$

⁽¹⁾Note-se que a curtose $\gamma_2(Y_n) = \frac{m_4}{m_2^2} - 3 \Leftrightarrow \frac{m_4}{m_2^2} = \gamma_2(Y_n) + 3$.

Como

$$\liminf_{n \rightarrow \infty} n \mathcal{J}_{st}(Y_n) \geq \frac{\gamma_1^2(X)}{2}.$$

O resultado acima mostra que, a menos que $\gamma_1(X) = 0$ (simetria) o melhor que há a esperar é uma velocidade de convergência $O\left(\frac{1}{n}\right)$.

Capítulo 4

Aproximações por Leis Limites Estáveis — Mais uns Passos em Volta

4.1 Teorema Limite Central e suas Generalizações

O Teorema Limite Central clássico, demonstrável com função característica, é o fecho de cúpula de muitos livros elementares de Probabilidade. Fizemos o seu elogio q.b. na Introdução.

Anote-se no entanto que esse teorema, tão apropriado para uma exposição didática a nível elementar quando olhado com espírito crítico mostra-se em certo sentido deploravelmente “matemático”. De facto:

1. a evolução da demonstração caminhou no sentido de aligeirar hipóteses sobre os momentos, obtendo-se uma demonstração de grande elegância, mas que esconde o facto de nas situações interessantes ser possível obter aproximações $O\left(\frac{1}{n}\right)$, enquanto na formulação habitual a velocidade de convergência é apenas de $O\left(\frac{1}{\sqrt{n}}\right)$.
2. o teorema assenta na hipótese irrealista de as parcelas serem independentes e identicamente distribuídas. Usando operadores de contracção, Lindeberg (1922) estabeleceu um teorema mais geral (equivalente, num sentido precisado em Pestana e Velosa (2002, pp. 996)) ao teorema com condições sobre a existência de $\mathbb{E}(X_k^3)$, de Liapuvov (1901). Por lado, Erdős and Rényi (1966) estabeleceram um resultado análogo sob condições de permutabilidade, essencial em Teoria da Amostragem, em que a estratégia mais comum é amostragem simples sem reposição, quando se aborda o problema da determinação da dimensão da amostra requerida para obter o grau de aproximação pretendido.

Por outro lado, as condições sobre existência de segundo momento são mais naturais (veja-se na introdução a discussão ligando o *TLC* à análise da velocidade de convergência na lei dos grandes números — *LGN*), mas é fácil exemplificar que se trata de uma condição suficiente, mas não necessária.

O tratamento geral fez-se (Feller, 1967) no contexto da variação regular de Karamata (1930). É aliás o contexto que, *a posteriori*, veio a revelar-se adequado para o estudo dos domínios de atracção para limites estáveis, quer em esquemas de somas quer em esquemas de extremos.

Expomos brevemente estes desenvolvimentos, e os resultados parcelares (mais encorajadores) que obtivemos na tentativa de construir uma teoria geral de leis limites lascada no critério de entropia máxima sob restrições específicas.

O Teorema Limite Central clássico afirma que se $\{X_k\}_{k \geq 1}$ é uma sucessão de variáveis aleatórias independentes e identicamente distribuídas, $X_k \stackrel{d}{=} X$, e $\mathbb{E}(X^2) < \infty$ então, denotando $S_n = \sum_{k=1}^n X_k$, tem-se

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \xrightarrow[n \rightarrow \infty]{d} Z \sim \text{Gaussiana}(0, 1).$$

Mas o teorema enunciado de modo informal anteriormente, tem subjacente hipóteses muito fortes — identidade distribucional das parcelas, independência das parcelas e caudas de peso moderado — tal não permite considerá-lo uma justificação para o uso generalizado do modelo Gaussiano.

Assim, o Teorema Limite Central clássico foi generalizado de várias maneiras relaxando as várias hipóteses. Começemos por ver as generalizações do Teorema Limite Central para variáveis não identicamente distribuídas.

Teorema 4.1.1. (Teorema Limite Central com a condição de Liapunov)

Seja $\{X_k\}_{k \in \mathbb{N}}$ uma sucessão de variáveis aleatórias independentes tais que $\mathbb{E}(X_k) = 0$, $\text{Var}(X_k) = \sigma_k^2 < \infty$ e $\mathbb{E}(|X_k|^3) < \infty$. Denote-se

$$\sigma_n^{*2} = \sigma_1^2 + \dots + \sigma_n^2.$$

Se

$$\frac{1}{\sigma_n^{*3}} \sum_{k=1}^n \mathbb{E}(|X_k|^3) \xrightarrow[n \rightarrow \infty]{} 0$$

então

$$\frac{1}{\sigma_n^*} \sum_{k=1}^n X_k = \frac{S_n}{\sigma_n^*} \xrightarrow[n \rightarrow \infty]{} Z \sim \text{Gaussiana}(0, 1).$$

A demonstração pode ser consultada em Pestana e Velosa, 2002, pp. 995-996.

Outra generalização do Teorema Limite Central para variáveis não identicamente distribuídas foi estabelecida por Lindeberg:

Teorema 4.1.2. (Teorema Limite Central com a condição de Lindeberg)

Seja $\{X_k\}_{k \in \mathbb{N}}$ uma sucessão de variáveis aleatórias independentes tais que $\mathbb{E}(X_k) = 0, \text{Var}(X_k) = \sigma_k^2 < \infty$, e denote-se

$$\sigma_n^{*2} = \sigma_1^2 + \dots + \sigma_n^2.$$

Se para todo $\varepsilon > 0$

$$\frac{1}{\sigma_n^{*2}} \sum_{k=1}^n \int_{|x| \geq \varepsilon \sigma_n^*} x^2 dF_{X_k}(x) \xrightarrow{n \rightarrow \infty} 0$$

então

$$\frac{1}{\sigma_n^*} \max_{1 \leq k \leq n} \sigma_k \xrightarrow{n \rightarrow \infty} 0$$

e

$$\frac{1}{\sigma_n^*} \sum_{k=1}^n X_k = \frac{S_n}{\sigma_n^*} \xrightarrow{n \rightarrow \infty} Z \sim \text{Gaussiana}(0, 1).$$

A demonstração do teorema anterior consiste na aplicação do Teorema de Liapunov às variáveis truncadas

$$\begin{cases} X_k & \text{se } |X_k| < \varepsilon_n \sigma_n^* \\ 0 & \text{se } |X_k| \geq \varepsilon_n \sigma_n^* \end{cases},$$

com $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$, e em verificar que o efeito da truncatura é negligível. A demonstração detalhada pode ser encontrada em Pestana e Velosa, 2002, pp. 996-1000.

Os teoremas anteriores são essencialmente equivalentes, apesar de se apoiarem em hipóteses diversas.

Feller (1935) demonstrou que as condições de Lindeberg são não só suficientes como também necessárias para convergência de somas normadas para o modelo gaussiano.

O interesse do Teorema de Lindeberg-Feller reside no facto de evidenciar que a convergência para o modelo Gaussiano resulta da contribuição de cada uma das parcelas ser negligível.

Teorema 4.1.3. (Lindeberg-Feller)

Seja $\{X_k\}_{k \geq 1}$ uma sucessão de variáveis aleatórias independentes tais que $\mathbb{E}(X_k) = 0, \text{Var}(X_k) = \sigma_k^2 < \infty$, e denote-se

$$\sigma_n^{*2} = \sigma_1^2 + \dots + \sigma_n^2.$$

Então

$$\frac{1}{\sigma_n^*} \sum_{k=1}^n X_k = \frac{S_n}{\sigma_n^*} \xrightarrow[n \rightarrow \infty]{d} Z \sim \text{Gaussiana}(0, 1).$$

(uniformemente em $x \in \mathbb{R}$) e

$$\frac{1}{\sigma_n^*} \max_{1 \leq k \leq n} \sigma_k \xrightarrow{n \rightarrow \infty} 0$$

se e só se para todo $\varepsilon > 0$

$$\frac{1}{\sigma_n^{*2}} \sum_{k=1}^n \int_{|x| \geq \varepsilon \sigma_n^*} x^2 dF_{X_k}(x) \xrightarrow{n \rightarrow \infty} 0.$$

Sob a hipótese de as parcelas serem independentes e identicamente distribuídas, mas relaxando a condição sobre a existência de momentos, o Teorema Limite Central, no sentido mais amplo de convergência de somas normadas de variáveis aleatórias, foi abordado com toda a generalidade por Paul Lévy, que começou a investigar os possíveis limites não degenerados de $\frac{S_n - a_n}{b_n}$, onde as “constantes de atracção” $a_n > 0$ e $b_n \in \mathbb{R}$ têm a função de estabilizar a soma.

4.2 Domínios de Atracção e Leis Estáveis para Somas

4.2.1 Leis Infinitamente Divisíveis

Variáveis aleatórias tal como as Poisson, as Gamas e as Gaussianas, podem ser decompostas na soma de n parcelas independentes e identicamente distribuídas, para todo o $n = 1, 2, \dots$, são apodadas por esse facto de infinitamente divisíveis, e têm um papel importante na teoria da adição de variáveis aleatórias. Por outro lado, é impossível escrever a uniforme como soma de duas variáveis independentes e identicamente distribuídas, pelo que é apodada de irreductível.

Khintchine demonstrou que qualquer variável aleatória pode ser decomposta na soma de variáveis aleatórias irreductíveis e infinitamente divisíveis sem factores irreductíveis, um teorema que paraleliza o teorema fundamental da aritmética racional, que mostra que todo natural pode ser decomposto em produto de naturais primos.

Lembramos que uma variável aleatória X é infinitamente divisível se e só se para todo o $n \geq 1$, existirem variáveis aleatórias Y_n tais que

$$X \stackrel{d}{=} \sum_{k=1}^n Y_{nk},$$

onde as variáveis Y_{nk} são réplicas independentes de Y_n . Por outras palavras, para todo $n \geq 1$

$$F_X = F_{Y_n}^{*n},$$

onde F^{*n} denota $F * F * \dots * F$, ou, em termos de funções características

$$\varphi_X = (\varphi_{Y_n})^n .$$

Ou ainda: X é infinitamente divisível se e só se, para todo $n \geq 1$, $\psi_n = (\varphi_X)^n$ for uma função característica.

Prova-se que qualquer variável aleatória infinitamente divisível ou é Poisson composta ou então limite de uma sucessão de variáveis aleatórias Poisson compostas, um notável resultado que se deve a de Finetti (1930).

Um resultado importante associado a funções características de variáveis infinitamente divisíveis é o seguinte:

Teorema 4.2.1. *Uma função característica infinitamente divisível não tem zeros reais.*

A demonstração pode ser consultada em Pestana e Velosa (2006, 2ª edição), ou em alternativa em Lukacs (1970).

O resultado anterior, tem como consequência o facto de qualquer função característica infinitamente divisível poder ser escrita na forma: $\varphi = e^{\psi}$, o que está na base das representações canónicas das funções características infinitamente divisíveis.

4.2.2 Domínios de Atracção e Leis Estáveis

Já referimos no capítulo 1, o porquê do interesse em investigar modelos aditivos alternativos ao modelo gaussiano, em situações em que distribuição de base tem caudas pesadas. Mais concretamente, caracterizar as possíveis leis limite (convergência fraca) de somas padronizadas $\frac{S_n - b_n}{a_n}$, onde $S_n = \sum_{k=1}^n X_k$, com parcelas X_k independentes e identicamente distribuídas, e as constantes normalizadoras ou constantes de atracção $a_n > 0$ e $b_n \in \mathbb{R}$. As constantes normalizadoras têm a função de “centrar” S_n subtraindo-lhe b_n por forma a localizá-la junto da origem, e “reduzir” dividindo por a_n para estabilizar a escala.

Também referimos, anteriormente, que as leis estáveis são distribuições infinitamente divisíveis. De facto,

$$e^{-i \frac{\beta_n}{\alpha_n} \frac{t}{n}} \varphi_Y \left(\frac{t}{\alpha_n} \right) = \varphi_Y^{1/n}(t),$$

e consequentemente para todo o $n \in \mathbb{N}$, $\varphi_Y^{1/n}(t) = \varphi_{\frac{nY - \beta_n}{n \alpha_n}}(t)$ é uma função característica.

Podemos assim, concluir que a gaussiana é estável com expoente característico 2, e que a Cauchy é a estável simétrica com expoente característico 1.

De seguida vamos enunciar, em forma de teorema algumas propriedades das leis estáveis.

Teorema 4.2.2. *Todas as distribuições estáveis são contínuas.*

Demonstração

Suponhamos que F tem um ou mais átomos e seja p o maior dos seus pesos. Para distribuições estáveis os átomos de $F \star F$ diferem dos de F apenas por localização e não pelo seu peso. Assim, $F \star F$ tem então um átomo de peso p , o que contradiz o facto de sob convoluções o peso máximo dos átomos decrescer. \square

Enunciamos sem demonstração o seguinte teorema:

Teorema 4.2.3. *Se F é estável com expoente $\alpha \neq 1$ existem constantes a e b tal que $F(ax + b)$ é estritamente estável.*

A restrição a distribuições estritamente estáveis é menos séria do que a primeira vista pode parecer, porque toda a distribuição estável com expoente $\alpha \neq 1$ pode ser centrada de modo a tornar-se estritamente estável.

Como já vimos, uma função característica infinitamente divisível não tem zeros reais. Com base neste resultado, Kolmogorov (1932), para o caso de variáveis infinitamente divisíveis com variância finita, Lévy (1934) e Khintchine (1935) descobriram representações integrais das respectivas funções características, que foram a base do desenvolvimento da “aritmética das variáveis aleatórias”.

Como qualquer variável estável é infinitamente divisível, a função característica de uma variável aleatória estável pode ser escrita usando a representação canónica de Lévy. A representação geral de uma função característica estável é:

$$\varphi(t) = \exp \left\{ i a t - c |t|^\alpha \left[1 + i \beta \frac{t}{|t|} w(t, \alpha) \right] \right\} \quad (4.1)$$

com localização $a \in \mathbf{R}$, escala $c > 0$, índice ou expoente característico $0 < \alpha \leq 2$, e coeficiente de assimetria $-1 \leq \beta \leq 1$, e

$$w(t, \alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \in (0, 1) \cup (1, 2] \\ \frac{2}{\pi} \ln(|t|) & \alpha = 1. \end{cases}$$

As caudas da função distribuição F de uma variável estável satisfazem a

$$x^\alpha [1 - F(x)] \longrightarrow A \frac{c_2}{c_1 + c_2} \frac{2 - \alpha}{\alpha} \quad \text{e} \quad x^\alpha F(-x) \longrightarrow A \frac{c_2}{c_1 + c_2} \frac{2 - \alpha}{\alpha}, \quad (A > 0),$$

com $c_1, c_2 \geq 0$ e $c_1 + c_2 > 0$, quando $x \rightarrow \infty$.

Excluindo o caso $\alpha = 2$, β qualquer, verifica-se que na expressão (4.1)

$$\beta = \frac{c_2 - c_1}{c_1 + c_2}$$

e o parâmetro de escala c é, para todo o α , ($0 < \alpha \leq 2$)

$$c = -(c_1 + c_2)Q(\alpha),$$

com

$$Q(\alpha) = \begin{cases} -\cos\left(\frac{\pi\alpha}{2}\right)\Gamma(1-\alpha) & \alpha \in (0, 1) \\ -\frac{\pi}{2} & \alpha = 1 \\ \cos\left(\frac{\pi\alpha}{2}\right)\frac{\Gamma(1-\alpha)}{\alpha-1} & \alpha \in (1, 2). \end{cases}$$

4.2.3 Funções de Variação Regular e Lenta

Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes e identicamente distribuídas com distribuição comum F , média zero e variância unitária. Seja $S_n = X_1 + \dots + X_n$. O Teorema Limite Central diz que a distribuição de $S_n n^{-1/2}$ converge em distribuição para uma variável aleatória gaussiana. Para distribuições sem variância as constantes normalizadoras têm que ser escolhidas de modo diferente, mas o limite pode ainda existir. Note-se que as distribuições estáveis são as únicas que ocorrem como tais limites.

Definição 4.2.1. *Uma função de valores reais L , definida sobre um intervalo $(a, +\infty)$ é dita de variação lenta se*

- i. $L(x) > 0, \forall x \in (a, +\infty)$*
- ii. $\frac{L(xy)}{L(y)} \xrightarrow{y \rightarrow \infty} 1$ para cada $x > 0$*
- iii. L é limitada para cada subintervalo limitado de (a, ∞) .*

Uma função φ definida sobre um intervalo (a, ∞) é dita de variação regular com expoente ρ se existe uma função de variação lenta L tal que $\varphi(x) \sim x^\rho L(x)$, $x \rightarrow \infty$.

Se L é uma função de variação lenta, um resultado a que se recorre frequentemente afirma que $xL(x) \xrightarrow{x \rightarrow \infty} \infty$. Este resultado não é necessariamente verdadeiro se (iii) na definição anterior for omitida. Se condição (iii) na definição anterior for substituída pela condição

- iii'. L é mensurável,*

a condição (iii) ainda se verifica (embora (iii) e (iii') não sejam equivalentes). Sob as condições (i), (ii) e (iii'), isto é, L mensurável e de variação regular, é válido o teorema de representação de Karamata

$$L(x) = c(x) \exp\left(\int_a^x \frac{\theta(t)}{t} dt\right),$$

onde $c(x)$ é uma função mensurável que satisfaz $c(x) \rightarrow c > 0$ (c constante) quando $x \rightarrow \infty$, onde $\theta(t)$ é integrável à Lebesgue sobre $(a, x) \forall x > a$ e $\theta(t) \rightarrow 0$ quando $t \rightarrow \infty$.

Prova-se que, se L_1 e L_2 são duas funções de variação lenta o seu produto também o é. Se, além disso, L_1 e L_2 forem mensuráveis e de variação lenta (isto é, (iii) substituído por (iii')), então o quociente entre L_1 e L_2 também é de variação lenta, o que não é necessariamente verdade se (iii) se verificar, mas (iii') não se verificar.

Autores como Loève particionaram os domínios de atracção em domínio *standard* (\mathcal{D}_S) e domínio de atracção não *standard* ($\mathcal{D}_{\bar{S}}$).

$$X \in \mathcal{D}_S(Y_\alpha) \quad \text{sse} \quad \frac{X_1 + \cdots + X_n}{A n^{1/\alpha}} \xrightarrow[n \rightarrow \infty]{} Y_\alpha$$

enquanto,

$$X \in \mathcal{D}_{\bar{S}}(Y_\alpha) \quad \text{sse} \quad \frac{X_1 + \cdots + X_n}{n^{1/\alpha} L(n)} \xrightarrow[n \rightarrow \infty]{} Y_\alpha,$$

com L de variação regular tal que $L(x) \xrightarrow[x \rightarrow \infty]{} c \in (0, \infty)$.

Por exemplo, todas as variáveis aleatórias com variância finita estão no domínio de atracção *standard* da gaussiana.

As condições de atracção podem ser expressas usando a teoria de variação regular de Karamata (1930):

Teorema 4.2.4. *A variável aleatória $X \in \mathcal{D}(Y_2)$ se e só se o seu segundo momento truncado, $s(x) = \int_{-x}^x w^2 dF_X(w)$, for uma função de variação lenta. No caso em que $0 < \alpha \leq 2$, $X \in \mathcal{D}(Y_\alpha)$ se e só se*

1. $\frac{1 - F_X(x)}{F_X(-x)} \xrightarrow[t \rightarrow \infty]{} k \in [0, +\infty)$.
2. $\frac{F_X(-tx) + 1 - F_X(tx)}{F_X(-x) + 1 - F_X(x)} \xrightarrow[t \rightarrow \infty]{} t^{-\alpha}$, ou seja, a soma das caudas $\mathbb{P}(|X| > x)$ é uma função de variação regular com expoente $-\alpha$.

A demonstração pode ser consultada em Galambos (1995, pp. 257-262).

Como se disse atrás, o domínio de atracção de uma lei estável pode ser particionado num “domínio de atracção *standard*” — em que as constantes de atracção são da forma $A n^{1/\alpha}$ onde α é o índice da estável —, e num “domínio de atracção não-*standard*”, em que as constantes de atracção são necessariamente da forma $L(n) n^{1/\alpha}$, com L uma função de variação lenta no sentido de Karamata, isto é,

$$L(n) \xrightarrow[n \rightarrow \infty]{} K \in]0, +\infty[$$

e $0 < \alpha \leq 2$ é o índice característico da variável aleatória limite Y .

4.3 Velocidades de Convergência em Caudas com Condição de Regularidade de 2ª ordem

Considere-se agora uma sucessão $\{X_k\}_{k \in \mathbb{N}}$ de variáveis aleatórias simétricas independentes e identicamente distribuídas no domínio de atracção não *standard* de uma estável simétrica Y_α , com função característica

$$\varphi_{Y_\alpha}(t) = \exp \{-c |t|^\alpha\}, \quad \alpha \in (0, 2) \quad \text{e} \quad c > 0. \quad (4.2)$$

Então, as caudas da função distribuição comum das variáveis $X_k, k = 1, 2, \dots$ têm o seguinte comportamento assintótico

$$F(-x) = \frac{c_1 + o(1)}{(-x)^\alpha} L(-x), \quad x < 0 \quad \text{e} \quad 1 - F(x) = \frac{c_2 + o(1)}{x^\alpha} L(x), \quad x > 0 \quad (4.3)$$

onde $c_1, c_2 \geq 0, c_1 + c_2 > 0$ e L é uma função de variação lenta em infinito. Como estamos a considerar variáveis aleatórias simétricas tem-se $c_1 = c_2$.

Calculando a função característica obtemos

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{\infty} e^{itx} f(x) dx = \int_{-\infty}^{\infty} \cos(tx) dF(x) + i \int_{-\infty}^{\infty} \sin(tx) dF(x) \\ &= 1 - 2t \int_0^{\infty} \sin(tx) [1 - F(x)] dx \end{aligned} \quad (4.4)$$

que é uma função real uma vez que F é simétrica.

Para qualquer distribuição no domínio de atracção de uma estável com expoente característico α é válido o seguinte resultado

Teorema 4.3.1. *Uma função distribuição com função característica φ , pertence ao domínio de atracção de uma estável cuja função característica tem logaritmo da forma*

$$-c |t|^\alpha \left[1 - i \beta \frac{t}{|t|} w(|t|, \alpha) \right]$$

onde $\alpha, \beta, c, w(|t|, \alpha)$ estão definidos em (4.1) se e só se numa vizinhança da origem se tem

$$\log(\varphi(t)) = i a t - c |t|^\alpha \tilde{L}(t) \left[1 - i \beta \frac{t}{|t|} w(|t|, \alpha) \right]$$

onde a é uma constante real, e \tilde{L} é uma função de variação regular lenta quando $t \rightarrow 0$.

A demonstração deste resultado utiliza as seguintes aproximações válidas quando $t \rightarrow 0$

(i) $0 < \alpha < 1 \wedge 1 < \alpha < 2$

$$\int_0^\infty \sin(x) \frac{L(x/t)}{x^\alpha} dx \approx \frac{\Gamma(2-\alpha)}{1-\alpha} \cos\left(\frac{\pi\alpha}{2}\right) L(1/t) \quad \text{e}$$

$$\int_0^\infty \frac{\sin(x)}{x^\alpha} dx \approx \Gamma(1-\alpha) \cos\left(\frac{\pi\alpha}{2}\right)$$

(ii) $\alpha = 1$

$$\int_0^\infty \sin(x) \frac{L(x/t)}{x} dx \approx \int_t^\infty \sin(x) \frac{L(x/t)}{x^\alpha} dx + O(t^2) =$$

$$= L(1/t) \int_0^\infty \frac{\sin(x)}{x} dx + O(t^2)$$

$$\text{e } \int_0^\infty \frac{\sin(x)}{x} dx = \frac{\pi}{2}$$

A demonstração do resultado encontra-se completa em Linnik (1959), pp. 85-86.

Considere-se uma sucessão $\{X_k\}_{k \in \mathbb{N}}$ de variáveis aleatórias independentes e identicamente distribuídas no domínio de atracção de uma estável de índice α , Y_α . Para simplificar vamos supor que estamos a trabalhar apenas com variáveis aleatórias simétricas.

O caso $\alpha = 2$ isto é, quando se está a trabalhar no domínio de atracção da gaussiana, foi já muito explorado, sabendo-se, por exemplo, que quando as variáveis aleatórias X_k têm terceiro momento finito, denotando por F_n a função distribuição das somas padronizadas e por Φ a função distribuição da gaussiana $(0, 1)$ se tem

$$\max_x |F_n(x) - \Phi(x)| = O(n^{-1/2}).$$

Quando $0 < \alpha < 1$ o problema torna-se um pouco mais complexo, mas existem alguns resultados no campo dos domínios de atracção *standard*. Por exemplo, num trabalho de Cramér de 1963 onde se mostra o seguinte resultado:

Teorema 4.3.2. *Seja F uma dada função distribuição, satisfazendo as seguintes condições:*

- i. F é simétrica,*
- ii. para $x \rightarrow \infty$ tem-se $1 - F(x) = \frac{p}{x^\alpha} + \frac{q}{x^\beta} + O(x^{-\beta})$, com $0 < \alpha < \beta < 2$, e q uma constante*
- iii. se $\beta \leq 1$, $O(x^{-\beta})$ é monótona para todo $x > 0$ suficientemente grande.*
- iv. $\limsup_{|t| \rightarrow \infty} |f(t)| < 1$.*

Então $F_n(x) = [F(n^{1/\alpha}(x))]^n \rightarrow G_\alpha(x)$, com G_α uma função distribuição estável e simétrica, com

$$\begin{cases} F_n(x) - G_\alpha(x) = -\frac{Q G_{\alpha,\beta}(x)}{n^{\beta/\alpha-1}} + o(n^{1-\beta/\alpha}) & \text{se } \beta < 2\alpha \\ F_n(x) - G_\alpha(x) = -\frac{P^2 G_{\alpha,2\alpha}(x)}{2n} + o(n^{-1}) & \text{se } 2\alpha < \beta < 2 \end{cases},$$

onde

$$P = 2p \int_0^\infty \frac{\sin(x)}{x^\alpha} dx > 0,$$

$$Q = 2q \int_0^\infty \frac{\sin(x)}{x^\beta} dx,$$

e

$$G_{\alpha,\beta} = \frac{1}{\pi} \int_0^{+\infty} t^{\beta-1} \exp(-Pt^\alpha) \sin(tx) dt.$$

No caso limite, quando $\beta = 2\alpha$, a última relação verifica-se com P^2 substituído por $P^2 + 2Q$.

Nos domínios de atracção não *standard*, refere-se aqui o trabalho de Iglésias Pereira *et al.* (1996) por ser especialmente interessante para nós: se $\{X_k\}_{k \in \mathbb{N}}$ for uma sucessão de variáveis aleatórias simétricas independentes e identicamente distribuídas, com função distribuição comum dada pela expressão (4.3) onde se substituiu L por \ln , então

$$1 - F(x) = \frac{k \ln(x)}{x^\alpha} + r(x), \quad \text{onde } r(x) = O(n^{-\alpha}), \quad x > 0.$$

A função característica de F é $\varphi(t) = 1 - 2t \int_0^{+\infty} \sin(tx)[1 - F(x)] dx$.

Usando as seguintes igualdades

$$\int_0^\infty \sin(x) \frac{L(x/t)}{x^\alpha} dx = \Gamma(1 - \alpha) \cos\left(\frac{\pi\alpha}{2}\right) L(1/t) + o[L(1/t)],$$

$$\int_0^\infty \frac{\sin(x)}{x^\alpha} dx = \Gamma(1 - \alpha) \cos\left(\frac{\pi\alpha}{2}\right),$$

tomando t na vizinhança de zero, e considerando $0 < \alpha < 2$, conclui-se que

$$\ln(\varphi(t)) = -c |t|^\alpha \tilde{L}(t),$$

onde $c = -2k \Gamma(1 - \alpha) \cos\left(\frac{\pi\alpha}{2}\right)$ e $\tilde{L}(t) = 1 - \ln |t|$.

Atendendo às propriedades do logaritmo pode-se escrever

$$\ln(\varphi(t)) \approx -c |t|^\alpha \left(\frac{1 - |t|^{\varepsilon_n}}{\varepsilon_n} + 1 \right),$$

onde resulta que a função característica de $S_n^* = \frac{X_1 + \dots + X_n}{A_n}$, tomando $A_n = \ln(n) n^{1/\alpha}$, é tal que

$$\ln(\varphi_{S_n^*}) \approx -c |t|^{\alpha_n}, \quad \alpha_n \xrightarrow{n \rightarrow \infty} \alpha^+.$$

Considere-se uma função de distribuição F de uma variável aleatória X no domínio de atracção (para somas) de uma estável de índice α . Suponha-se que X é simétrica, absolutamente contínua e que, para $x > 0$,

$$1 - F(x) = \frac{L(x)}{x^\alpha} + \frac{A}{x^\beta} + r(x)$$

com $0 < \alpha < 2, \alpha < \beta < \nu$, L uma função de variação lenta, $r(x) = o(x^{-\nu})$ e A uma constante. A função característica de X é

$$\varphi(t) = 1 - 2t \int_0^\infty \sin(tx) \left(\frac{L(x)}{x^\alpha} + \frac{A}{x^\beta} + r(x) \right) dx.$$

Fazendo $y = x |t|$ e efectuando algumas simplificações⁽¹⁾ tem-se

$$\begin{aligned} \varphi(t) &= 1 - 2t \int_0^\infty \sin(\operatorname{sgn}(t)y) \left(\frac{L(y/|t|)}{y^\alpha} |t|^\alpha + \frac{A}{y^\beta} |t|^\beta + r\left(\frac{y}{|t|}\right) \right) \frac{1}{|t|} dy \\ \varphi^n \left(\frac{t}{n^{1/\alpha} L_1(n)} \right) &= \exp \left\{ -\frac{|t|^\alpha a}{L_1^\alpha(n)} L \left(\frac{n^{1/\alpha} L_1(n)}{|t|} \right) \right\} \exp \left\{ n \rho_3 \left(\frac{t}{n^{1/\alpha} L_1(n)} \right) \right\}. \end{aligned}$$

O estudo de velocidades de convergência no que se refere ao domínio de atracção da gaussiana “parece” uma teoria com um alto grau de sofisticação e complexidade. Porém, quando se aborda o estudo da velocidade de convergência nos domínios de atracção de estáveis não gaussianas é que as dificuldades emergem em todo o seu esplendor.

Tentámos, com o uso de parâmetros de segunda ordem — a exemplo do que se tem feito nos domínios de atracção de estáveis para extremos — controlar a velocidade de convergência. Os resultados parcelares a que chegamos dependem de condições porventura artificiosas, e a sua extensão para domínios de atracção não-*standard* traz complexidades ainda maiores.

4.4 Convoluções de Leis Estáveis

No que respeita a convoluções de leis estáveis recordamos os seguintes resultados que julgamos ser mais relevantes para o nosso trabalho, e cujas demonstrações podem ser encontradas, por exemplo, em Feller, 1960, pp. 271.

⁽¹⁾A dedução sem omissão de passos é apresentada em apêndice.

Teorema 4.4.1. *Sejam F_1 e F_2 duas funções de distribuição tais que para $x \rightarrow \infty$*

$$1 - F_i(x) \sim \frac{a_i}{x^\rho} L(x),$$

com $L(x)$ uma função de variação lenta. Então a convolução $G = F_1 \star F_2$ tem uma cauda de variação regular tal que $1 - G(x) \sim \frac{a_1 + a_2}{x^\rho} L(x)$. Se $1 - F(x) \sim x^{-\rho} L(x)$ então $1 - F^{r\star} \sim r x^{-\rho} L(x)$.

Mais geralmente, é possível mostrar (cf. Tucker, 1968, pp. 1382-1383) o seguinte resultado:

Teorema 4.4.2. *Sejam F_1, \dots, F_m , $m \geq 2$, funções de distribuição tais que, para $x \rightarrow \infty$,*

$$1 - F_i(x) \sim \frac{L_i(x)}{x^{\rho_i}}, \quad 1 \leq i \leq m,$$

onde para algum $k \in \{1, \dots, m\}$ fixo, $0 < \rho_1 = \dots = \rho_k < \rho_{k+1} \leq \dots \leq \rho_m$, e onde L_1, \dots, L_m são funções mensuráveis de variação lenta. Então $1 - (F_1 \star \dots \star F_m)$ é de variação regular de expoente $-\rho_1$ e, para $x \rightarrow \infty$,

$$1 - (F_1 \star \dots \star F_m)(x) \sim x^{-\rho_1} \sum_{j=1}^k L_j(x).$$

Consideremos agora F e G duas funções de distribuição e suponhamos que

$$1 - F(x) \sim x^{-\alpha_1} + A_1 x^{-\mu_1} + O(x^{-\nu_1}), \quad 0 < \alpha_1 < \mu_1 < \nu_1 < 2$$

e

$$1 - G(x) \sim x^{-\alpha_2} + A_2 x^{-\mu_2} + O(x^{-\nu_2}), \quad 0 < \alpha_2 < \mu_2 < \nu_2 < 2,$$

com A_1 e A_2 constantes reais. Interessa-nos saber como se comportam as caudas da convolução de F com G isto é, como se comporta $1 - F \star G$.

Tem-se que

$$1 - F(x) \sim \frac{L_1(x)}{x^{\alpha_1}} \quad \text{e} \quad 1 - G(x) \sim \frac{L_2(x)}{x^{\alpha_2}},$$

onde

$$L_1(x) = 1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1}) \quad \text{e} \quad L_2(x) = 1 + A_2 x^{\alpha_2 - \mu_2} + O(x^{\alpha_2 - \nu_2}).$$

são funções de variação lenta⁽²⁾.

Como L_1 e L_2 são funções de variação lenta podemos aplicar o resultado do teorema (4.4.2) para $m = 2$, tendo-se que

$$1 - (F \star G)(x) \sim x^{-\min(\alpha_1, \alpha_2)} (L_1(x) + L_2(x)),$$

⁽²⁾Demonstração feita em apêndice.

com

$$L_1(x) + L_2(x) = 2 + A_1 x^{\alpha_1 - \mu_1} + A_2 x^{\alpha_2 - \mu_2} + O(x^{\alpha_1 - \nu_1}) + O(x^{\alpha_2 - \nu_2}).$$

Supondo, sem perda de generalidade, que $\alpha_1 < \alpha_2$ temos

$$1 - (F \star G)(x) \sim x^{-\alpha_1} (2 + A_1 x^{\alpha_1 - \mu_1} + A_2 x^{\alpha_2 - \mu_2}) + O(x^{-(\alpha_2 - \alpha_1 - \nu_2)}).$$

Conclusão

A teoria da adição de variáveis aleatórias moldou o desenvolvimento da teoria da probabilidade moderna. Mas só recentemente, com o advento de disponibilidade generalizada de meios computacionais eficientes, conceitos como o de modelo estável começaram a evidenciar potencialidades práticas efectivas. Mesmo assim, a modelação usando a lei “normal” (porventura a mais anormal de todas, tantas são as propriedades relevantes que dela são características) continua a dominar o panorama das aplicações, tal a facilidade e eficiência das metodologias estatísticas que lhe estão estreitamente associadas.

Importa por isso investigar condições gerais que esclareçam em que situações a aproximação por uma gaussiana — e, mais geralmente, por uma estável para somas — é aceitável. Condições genéricas, como as desenvolvidas por Berry e Esséen são bem conhecidas, e existe uma teoria coerente de aproximações na zona dos grandes desvios. Cramér, inspirando-se no uso de distribuições conjugadas de Esscher, e Daniels, importando as técnicas de aproximação usando pontos sela, introduziram ideias seminais na área de convergências estocásticas. Aquelas duas abordagens são afinal equivalentes, e em trabalho conjunto com Sequeira mostrámos que num nível mais profundo estão associadas à teoria das somas aleatórias de variáveis aleatórias, permitindo inclusivamente uma re-obtenção da fórmula de Khintchine-Pollaczek.

A apresentação coerente é facilitada usando expansões de Edgeworth e de Cornish-Fisher, a teoria dos polinómios ortogonais, por um lado, e restringindo a aplicação a distribuições da família exponencial, por outro (atendendo à relação que têm com a desigualdade de Cramér-Rao, a sua preponderância na teoria das aplicações de Estatística só pode ser considerada natural). A subfamília de Morris ganha, neste contexto, um papel de relevo, que porventura está relacionado com o facto de a simetrisação das densidades dessa família serem funções características (um ângulo que não tivemos oportunidade de explorar).

O aspecto mais notável da abordagem de aproximação na zona de grandes desvios usando distribuições conjugadas e pontos de sela é que se obtêm convergências $O\left(\frac{1}{n}\right)$, quando na abordagem clássica não se esperava melhor do que $O\left(\frac{1}{\sqrt{n}}\right)$. Este resultado por si só justifica a importância das expansões de Edgeworth diferidas, e de toda a área de *small sample asymptotics* que inspirou. Mostra, também, a relevância dos cumulantes de terceira e quarta ordem, ou seja o papel da assimetria e achatamento nas velocidades

de convergência.

A explicação deste espectacular progresso para convergências $O\left(\frac{1}{n}\right)$ encontra-se afinal na apresentação do Teorema Limite Central via teoria da informação. Anote-se, aliás, que a caracterização através da entropia máxima, que no caso de variáveis com suporte numa semi-recta e valor médio finito conduz à exponencial, cuja simetriação é a Laplace, que oportunamente referimos ser a \mathcal{N} -gaussiana no caso de somas aleatórias subordinadas por uma geométrica. Há, assim, explicações muito profundas, e que reclamam uma re-ligação da Probabilidade à Física, no que se refere a comportamentos assintóticos.

Curiosamente o uso de expansões de Edgeworth diferidas (*tilted*, isto é recorrendo a distribuições conjugadas) mostra que podemos encontrar excelentes aproximações nas situações mais inesperadas, nomeadamente no que se refer à estável de Lévy (uma possível explicação, que não conseguimos provar, terá porventura que ver com a relação entre as estáveis com expoente característico α e com expoente característico $1/\alpha$).

Mas o estudo genérico de aproximações de somas de variáveis aleatórias no domínio de atracção de uma estável com expoente característico α pela lei limite é mais complexo. Apresentamos alguns passos na tentativa de uma teoria geral de avaliar velocidades de convergência usando parâmetros reguladores de segunda ordem, restringindo assim a variação regular que caracteriza a soma das caudas das variáveis no domínio de atracção de uma estável; este trabalho inspira-se nesse aspecto, do que tem vindo a ser feito em teoria de valores extremos (nomeadamente entre nós por de Haan, Fraga Alves, Gomes e seus colaboradores), mas tem a dificuldade acrescida de no caso de estáveis para somas não haver forma explícita simples (salvo os casos excepcionais que são a Lévy, a Cauchy e a Gaussiana) para as densidades assintóticas.

Tal como Fisher e Tippett observaram para extremos, Iglésias Pereira, Oliveira e Pestana mostraram que há aproximações pré-assintóticas na teoria de somas, e que analogamente são regra nos domínios de atracção não-*standard*. Neste caso, o estudo de velocidades de convergência impondo parâmetros de segunda ordem conduz a expressões de grande complexidade. Mais uma vez, a anormal “normal” sai reforçada como modelo por excelência, uma vez que, como Draper tão justamente afirmou, os modelos são para ser usados, não são para se acreditar neles.

Apêndice A

Alguns Cálculos

Considere-se uma função de distribuição F de uma variável aleatória X no domínio de atracção de uma estável de índice α . Suponha-se que X é simétrica, absolutamente contínua e que para $x > 0$

$$1 - F(x) = \frac{L(x)}{x^\alpha} + \frac{A}{x^\beta} + r(x)$$

com $0 < \alpha < 2, \alpha < \beta < \nu$, L uma função de variação lenta, $r(x) = o(x^{-\nu})$ e A uma constante. A função característica de X é

$$\varphi(t) = 1 - 2t \int_0^\infty \sin(tx) \left(\frac{L(x)}{x^\alpha} + \frac{A}{x^\beta} + r(x) \right) dx.$$

Fazendo $y = x|t|$ tem-se

$$\begin{aligned} \varphi(t) &= 1 - 2t \int_0^\infty \sin(\operatorname{sgn}(t)y) \left(\frac{L(y/|t|)}{y^\alpha} |t|^\alpha + \frac{A}{y^\beta} |t|^\beta + r\left(\frac{y}{|t|}\right) \right) \frac{1}{|t|} dy \\ &= 1 - 2|t|^\alpha \int_0^\infty \frac{\sin(y) L(y/|t|)}{y^\alpha} dy - 2A|t|^\beta \int_0^\infty \frac{\sin(y)}{y^\beta} dy - 2 \int_0^\infty \sin(y) r\left(\frac{y}{|t|}\right) dy. \end{aligned}$$

Para $\alpha \neq 1$ tem-se

$$\begin{aligned} \varphi(t) &= 1 - 2|t|^\alpha \left\{ \frac{\Gamma(2-\alpha)}{1-\alpha} \cos\left(\frac{\pi\alpha}{2}\right) L\left(\frac{1}{|t|}\right) + o\left[L\left(\frac{1}{|t|}\right)\right] \right\} - 2A|t|^\beta \frac{\Gamma(2-\beta)}{1-\beta} \cos\left(\frac{\pi\beta}{2}\right) - \\ &\quad - \int_0^\infty \sin(y) r\left(\frac{y}{|t|}\right) dy \\ &= 1 - 2|t|^\alpha \left\{ \frac{\Gamma(2-\alpha)}{1-\alpha} \cos\left(\frac{\pi\alpha}{2}\right) L\left(\frac{1}{|t|}\right) + o\left[L\left(\frac{1}{|t|}\right)\right] \right\} - 2A|t|^\beta \frac{\Gamma(2-\beta)}{1-\beta} \cos\left(\frac{\pi\beta}{2}\right) + o(|t|^\nu) \\ &= 1 - |t|^\alpha a L\left(\frac{1}{|t|}\right) - |t|^\alpha r_1(t) - |t|^\beta b + r_2(t) \end{aligned}$$

com $a = 2 \frac{\Gamma(2-\alpha)}{1-\alpha} \cos\left(\frac{\pi\alpha}{2}\right)$ e $b = 2A \frac{\Gamma(2-\beta)}{1-\beta} \cos\left(\frac{\pi\beta}{2}\right)$, $r_1(t) = o\left[L\left(\frac{1}{|t|}\right)\right]$ e $r_2(t) = o(|t|^\nu)$.

Como

$$|t|^\alpha r_1(t) = o\left[|t|^\alpha L\left(\frac{1}{|t|}\right)\right], \quad r_2(t) = o\left[|t|^\alpha L\left(\frac{1}{|t|}\right)\right],$$

e

$$|t|^\beta b = o\left[|t|^\alpha L\left(\frac{1}{|t|}\right)\right]$$

resulta,

$$\varphi(t) = 1 - |t|^\alpha a L\left(\frac{1}{|t|}\right) + \rho_1(t)$$

onde

$$\rho_1(t) = o\left[|t|^\alpha L\left(\frac{1}{|t|}\right)\right].$$

Usando

$$\ln(1-x) = -x + O_{x \rightarrow 0}(x^2) = -x + o_{x \rightarrow 0}(x^\delta), \quad \forall \delta \in (0, 2),$$

vem, para t numa vizinhança de zero tal que $\varphi(t) > 0$

$$\ln(\varphi(t)) = -|t|^\alpha a L\left(\frac{1}{|t|}\right) + \rho_3(t)$$

onde $\rho_3(t) = \rho_1(t) + \rho_2(t) = o\left(|t|^\alpha L\left(\frac{1}{|t|}\right)\right)$. Substituindo t por $\frac{t}{B_n} = \frac{t}{n^{1/\alpha} L_1(n)}$, onde L_1 é uma função de variação lenta, e multiplicando por n ,

$$n \ln \varphi\left(\frac{t}{n^{1/\alpha} L_1(n)}\right) = -\frac{|t|^\alpha a}{L_1^\alpha(n)} L\left(\frac{n^{1/\alpha} L_1(n)}{|t|}\right) + n \rho_3\left(\frac{t}{n^{1/\alpha} L_1(n)}\right),$$

ou seja,

$$\varphi^n\left(\frac{t}{n^{1/\alpha} L_1(n)}\right) = \exp\left\{-\frac{|t|^\alpha a}{L_1^\alpha(n)} L\left(\frac{n^{1/\alpha} L_1(n)}{|t|}\right) + n \rho_3\left(\frac{t}{n^{1/\alpha} L_1(n)}\right)\right\}.$$

Para $\alpha = 1$ iríamos obter uma expressão análoga.

Apêndice B

Mais Alguns Cálculos

Vamos provar que $L_1(x) = 1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})$ é de variação lenta, onde $0 < \alpha_1 < \mu_1 < \nu_1 < 2$.

Tem-se

$$\lim_{x \rightarrow \infty} \frac{L_1(\lambda x)}{L_1(x)} = \lim_{x \rightarrow \infty} \frac{1 + A_1 (\lambda x)^{\alpha_1 - \mu_1} + O[(\lambda x)^{\alpha_1 - \nu_1}]}{1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})}.$$

Mas,

$$\lim_{x \rightarrow \infty} \frac{1}{1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})} = 1 \quad \text{pois} \quad \alpha_1 - \mu_1 < 0 \wedge \alpha_1 - \nu_1 < 0,$$

e

$$\lim_{x \rightarrow \infty} \frac{A_1 (\lambda x)^{\alpha_1 - \mu_1}}{1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})} = \lim_{x \rightarrow \infty} (\lambda x)^{\alpha_1 - \mu_1} = 0.$$

Além disso,

$$\lim_{x \rightarrow \infty} \frac{O(x^{\alpha_1 - \nu_1})}{1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})} = \frac{0}{1} = 0,$$

tendo em atenção que

$$\alpha_1 - \nu_1 < 0 \quad \text{e} \quad \alpha_1 - \mu_1 < 0.$$

Então,

$$\begin{aligned}
\lim_{x \rightarrow \infty} \frac{L_1(\lambda x)}{L_1(x)} &= \lim_{x \rightarrow \infty} \frac{1 + A_1 (\lambda x)^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})}{1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})} \\
&= \lim_{x \rightarrow \infty} \frac{1}{1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})} + \lim_{x \rightarrow \infty} \frac{A_1 (\lambda x)^{\alpha_1 - \mu_1}}{1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})} \\
&\quad + \lim_{x \rightarrow \infty} \frac{O(x^{\alpha_1 - \nu_1})}{1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})} \\
&= 1 + 0 + 0 = 1,
\end{aligned}$$

ou seja, $L_1(x) = 1 + A_1 x^{\alpha_1 - \mu_1} + O(x^{\alpha_1 - \nu_1})$ é de variação lenta, e por um processo análogo ao anterior, prova-se que também $L_2(x) = 1 + A_2 x^{\alpha_2 - \mu_2} + O(x^{\alpha_2 - \nu_2})$ é de variação lenta.

Bibliografia

- [1] Abramowitz, M. and Stegun, I. (1992). *Handbook of Mathematical Functions*, Dover, New York.
- [2] Alves, I. F., Haan, L. and Lin, T. (2003). Estimation of the parameter controlling the speed of convergence in extreme value theory. *Mathematical Methods of Statistics*, **12**, 155-176.
- [3] Arbuthnott, J. An argument for divine providence, taken from the constant regularity observed in the birth of both sexes, *Philosophical Transactions*, **40**, 90-186, reeditado em Kendall and Plackett (1977) e em David and Edwards (2001).
- [4] Anderson. C. W. (1978). Super-slowly varying functions in extreme value theory. *Journal of Royal Statistical Soc.*, **40**, 197-202.
- [5] Barron, R. A. (1985). Entropy and the central limit theorem, *Ann. Probab.* **1**, 336-342.
- [6] Bateman, H. (1954). *Tables of Integrals Transforms*, McGraw-Hill, New York.
- [7] Bartholomew, D. J. (1960). Sufficient conditions for a mixture of exponentials to be a probability density function, *Ann. Math. Probab.*, **40**, 2183-2188.
- [8] Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variables, *Trans. Amer. Math. Soc.*, **49**, 122-136.
- [9] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular Variation*, Cambridge Press, Cambridge.
- [10] Buhlmann, H., Delbaen, F. and Embrechts, P. (1998). On Esscher transforms in discrete finance models, *Astin Bull.*, **28**, 171-186.
- [11] Burrell, C. W. (1972). *Measure, Integration and Probability*, McGraw-Hill, New York.
- [12] Casella, G., Berger, R. L. (2002), *Statistical Inference*, Duxbury Advanced Series, Pacific Grove.
- [13] Charlier, C. V. L. (1906), Uber der Darstellung Willkürlicher Funktionen, *Ark. Mat. Astr. och Fys.* **2**, 1-35.

- [14] Cordeiro, M. G. (1999). *Introdução à Teoria Assintótica*, 22º Colóquio Brasileiro de Matemática, Rio de Janeiro.
- [15] Cover, T., Thomas, J. (1991). *Elements of Information Theory*, Willey, New York.
- [16] Cramér, H. (1928). On the composition of elementary errors, *Skand. Aktuarietidskr*, **11**, 13-74.
- [17] Cramér, H. (1928a). On the composition of elementary errors II: statistical applications, *Skand. Aktuarietidskr*, **11**, 141-180.
- [18] Cramér, H. (1939). On the representation of a function by certain Fourier integrals, *Trans. Amer. Math. Soc.*, **46**, 191-201.
- [19] Cramér, H. (1962). On the approximation to a stable probability distributions. In *Studies in Mathematical Analysis and Related Topics* (Szergo, G., ed.), Stanford Press, Califórnia, 70-76.
- [20] Cramér, H. (1962a). *Random Variables and Probability Distributions*, 3th ed., Cambridge Press, Cambridge.
- [21] Cramér, H. (1963). On asymptotic expansions for sums of independent random variables with limiting stable distributions, *Sankhya Ser.*, **25**, 12-24.
- [22] Cramér, H. (1979). Half a century with probability theory: some personal recollections, *Ann. Probab.*, **4**, 509-546.
- [23] Cover, T. M., Thomas, J. A. (1990). *Elements of information theory*, Wiley, New York.
- [24] Daniels, H. E. (1954). Saddlepoint approximations in Statistics, *Ann. Math. Ststist.*, **25**, 631-650.
- [25] Darmois, G. (1946). Sur les lois de probabilité, *C. R. Acad. Sci. Paris*, **222**, 164-165.
- [26] David, H.A., Edwards, A. W. F. (2001). *Annotated Readings in the History of Statistics*, Springer, New York.
- [27] Edgeworth, F. Y. (1905). The Law of Error, *Proc. Cambridge Philos. Soc.* **20**, 36-66.
- [28] Esscher, F. (1932). On the probability function in the collective theory of risk, *Skand. Aktuarietidskr*, **15**, 175-195.
- [29] Esséen, C. G. (1932). On the Liapounoff limit of error in the theory of probability, *Ark. Mat. Astr. Och. Fys.*, **28**, 1-19.
- [30] Esséen, C. G. (1945). Fourier analysis of distributions functions, *Acta Math.*, **77**, 1-125.

- [31] Feller, W. (1935). Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung, *Math. Z.*, **40**, 512-559.
- [32] Feller, W. (1937). Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung II, *Math. Z.*, **42**, 301-312.
- [33] Feller, W. (1943). Generalization of a probability limit theorem of Cramér, *Trans. Amer. Soc.*, **54**, 361-372.
- [34] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, vol. I, Wiley, New York.
- [35] Feller, W. (1967). *An Introduction to Probability Theory and its Applications*, vol. II, Wiley, New York.
- [36] Finetti, B. de (1930). Le funzione caratteristiche di legge istantanea. *Rend. Ac. Lenci*, (6) **12**, 278-282.
- [37] Fisher, R. A. (1934) Two new properties of mathematical likelihood, *Proc. Royal Soc.*, **144**, 295-307.
- [38] Galambos, J. (1954). *Advanced Probability Theory*, M. Dekker, New York.
- [39] Gerber, H., Shiu, E. (1994). Option pricing by Esscher transforms, *Transactions of Society of Actuaries*, **46**, 99-140.
- [40] Gnedenko, B. V. (1940). On the theory of domains of attraction of stable laws. *Uchemye Zapiski, Moskov. Gos. Univ.*, **30**, 61-72.
- [41] Gnedenko, B. V., Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Reading, Mass.
- [42] Gomes, I., Pestana, D. (1987). Nonstandard domains of attraction an rates of convergence. In *New Perspectives in Theoretical and Applied Statistics* (Puri, M. L., Vilaplana, J. P. e Wertz, W., eds), Wiley, New York, 467-477.
- [43] Graunt, J. (1662). *Natural and political observations made upon bills of mortality*, Roycroft, Londres.
- [44] Haan, L. de (1984). Slow variation and characterizations of domains of attraction. In *Statistical Extremes and Applications* (Oliveira, J. T., ed), D. Reidel Publishing Company, 31-48.
- [45] Haan, L. de (1996). Second order regular variation and rates of convergence in extreme value theory, *Ann. Probab.*, **24**, 97-124,
- [46] Hall, P. (1992). *The Bootstrap and Edgeworth Expansions*, Springer-Verlag, New York.

- [47] Hampel, F. R. (1973). Some small sample asymptotics, In *Proceedings of the Prague Symposium on Asymptotic Statistics*, **2**, 109-126.
- [48] Harkness, W. L., Harkness, M. L. (1974). Generalized hyperbolic secant distributions, *American Statistical Association Journal*, **63**, 329-337.
- [49] Hotelling, H. (1961). The behavior of some standard statistical tests under non-standard conditions, *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, **I**, California University Press, Berkeley, 319-359.
- [50] Hoffman-Jørgensen, J. (1976). *The Two Pearls of Probability Theory*, Springer, Berlin.
- [51] Kagan, A. M., Linnik, Y. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*, Wiley, New York.
- [52] Karamata, J. (1930). *Sur un mode de croissance régulière des fonctions*, *Mathematica (Cluj)*, **4**, 38-53.
- [53] Kendall, M. G., Plackett (1977). *Studies in History of Statistics and Probability*, Griffin, London.
- [54] Kendall, M. G., (1968). Studies in the history of probability and statistics. XIX. Francis Ysidro Edgeworth, *Biometrika*, **55**, 175-269.
- [55] Khintchine, A. Y. (1960). *Mathematical Methods in the Theory of Queueing*, Griffin, London.
- [56] Koopman, L. H. (1936). On distributions admitting a sufficient statistic, *Trans. Amer. Math. Soc.*, **39**, 399-409.
- [57] Kotz, S. (1966). Recent results in information theory, *J. Appl. Probab.*, **39**, 399-409.
- [58] Ibragimov, I. A., Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Netherlands.
- [59] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, vol I, Wiley, New York.
- [60] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, vol II, Wiley, New York.
- [61] Laha, R. G., Luckacs, E. (1960). On a problem connected with quadratic regression, *Biometrika*, **47**, 335-343.
- [62] Laplace, P. H. S. (1814). *Théorie Analytique des Probabilités*, Mme Veuvw Courcier, Paris.

- [63] La Cam, L. (1986). The central limit theorem around 1935, *Statistical Science*, **1**, 78-96.
- [64] Lévy, P. (1925). *Calcul des probabilités*. Gauthier-Villars, Paris.
- [65] Liapunov, A. M. (1901). Nouvelle forme du théorème sur le limite des probabilités, *Acad. Imp. Sci. St. Petersbourg*, **12**, 1-24.
- [66] Lindeberg, J. W. (1920). Über das exponentialgesetz in der wahrscheinlichkeitsrechnung, *Ann. Acad. Sci. Fennicae*, **16**, 1-23.
- [67] Lindeberg, J. W. (1922). Eine neue herleitung des exponentialgesetz in der wahrscheinlichkeitsrechnung, *Math. Z.*, **15**, 211-225.
- [68] Linnik, Y. V. (1959). An information - theoretic proof of the central limit theorem with Lindeberg condition, *Theory Probab. Appl.*, **3**, 288-298.
- [69] Loève, M. (1977). *Probability Theory*, vol. I, Springer-Verlag, New York.
- [70] Luckacs, E. (1970). *Characteristic Functions*, Griffin, London.
- [71] Magnus, E., Oberhettinger, T., Erdélyi, A. and Tricomi, F. G. (1953). *Higher Transcendental Functions*, McGraw-Hill, New York.
- [72] Malva, M. (2006) Domínios de atracção e velocidades de convergência, submetido para publicação nas actas do XIII congresso da SPE.
- [73] Malva, M., Mendonça, S. (2004). Distribuições conjugadas e somas de variáveis aleatórias, In *Estatística com Acaso e Necessidade* (Rodrigues, M. M., Rebelo, E. L. e Rosado, F., eds.), Edições SPE, Lisboa, 401-406.
- [74] Malva, M., Mendonça, S. (2005). Desenvolvimentos assintóticos e polinómios ortogonais, In *Estatística Jubilar* (Braumann, C., Infante, P., Oliveira, M. M., Jara, R. A., e Rosado, F., eds.), Edições SPE, Lisboa, 433-444.
- [75] Malva, M., Sequeira, F. (2003). Distribuições duais, sua génese e caracterizações, In *Literacia e Estatística* (Brito, P., Figueiredo, A., Sousa, F., Teles, P. e Rosado, F., eds.), Edições SPE, Lisboa, 369-380.
- [76] Mason, J. D. (1970). Convergence of distributions attracted to stable laws, *Ann. Math. Stat.*, **39**, 1381-1390.
- [77] Mason, J. D. (1970a). Convergence of sums to a convolution of stable laws, *Ann. Math. Stat.*, **41**, 1068-1070.
- [78] Mason, J. D. (1970b). Convolution of stable laws as limit distributions of partial sums, *Ann. Math. Stat.*, **41**, 101-114.

- [79] Mitrinovic, D. S. (1966). *Calculus of Residues*, Noordhoff, Netherlands.
- [80] Moivre, A. (1718). De mensura sortis, *Phil. Trans.*, **27**, 213-264
- [81] Morris, C. N. (1982). Natural exponential families with quadratic variance functions, *Ann. Stat.*, **10**, 65-80.
- [82] Morris, C. N. (1983). Natural exponential families with quadratic variance functions: statistical theory, *Ann. Stat.*, **2**, 515-529.
- [83] Neuman, C. (1693). An estimate of the degrees of the mortality of mankind, drawn from curious tables of birth and funerals at the city of Breslaw; with an attempt to ascertain the price of annuities upon lives, Breslaw.
- [84] Oliveira, O. (1997). Constantes de atracção e velocidades de convergência em situações pré-assintóticas, In *A Estatística a Decifrar o Mundo* (Vasconcelos, R., Alves, I. F., Castro, L. C., Pestana D., eds.), Edições SPE, Lisboa, 117-119.
- [85] Pearson, K. (1920). Notes on the history of correlation, *Biometrika*, **13**, 24-45.
- [86] Pereira. H. I. (1985). *Escolha Estatística em Modelos Estáveis*, Univ. Lisboa, Lisboa.
- [87] Pereira, H. I., Oliveira, O. e Pestana, D. (1997). Limites estáveis e comportamentos pré-assintóticos, In *A Estatística a Decifrar o Mundo* (Vasconcelos, R., Alves, I. F., Castro, L. C., Pestana D., eds.), Edições SPE, Lisboa, 109-1196.
- [88] Pestana, D., Velosa, S. (2003). Classes de leis N -infinidamente divisíveis, In *Literacia e Estatística* (Brito, P., Figueiredo, A., Sousa, F., Teles, P. e Rosado, F., eds.), Edições SPE, Lisboa, 511-522.
- [89] Pestana, D., Velosa, S. (2006, 2ª ed.). *Introdução à Probabilidade e Estatística*, vol. I, Fundação Calouste Gulbekian, Lisboa.
- [90] Pitman, E. J. G. (1967). On the behavior of the characteristic function of a probability distribution in the neighborhood of the origin, *J. Austral. Math. Soc.*, **8**, 423-443.
- [91] Pitman, E. J. G. (1936). Sufficient statistics and intrinsic accuracy, *Cambridge Phil. Soc.*, **32**, 567-579.
- [92] Poisson, S. D. (1827). *Sur la Probabilité des Résultats Moyens des Observations, Additions à la Connaissance des Temps pour L'année 1827*, Bachelier, Paris.
- [93] Pólya, G. (1920). Uber den zentralen grenzwetsatz der wahrscheinlickeitsrechnung und das moment problem, *Math. Z.*, **8**, 171-180.
- [94] Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*, Cambridge Press, Cambridge.

- [95] Plato, J. V. (2001). *Creating Modern Probability*, Cambridge Press, Cambridge.
- [96] Ramachandran, B. (1976). *Advanced Theory of Characteristic Functions*, Statistical Publishing Society, Calcutta.
- [97] Resnick, S. (1971). Tail equivalence and its applications, *Journal Appl. Probab.*, **8**, 136-156.
- [98] Rényi, S. (1970). *Probability Theory*, North-Holland, Amesterdan.
- [99] Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*, Wiley, New York.
- [100] Ross, S. (1998). *A First Course in Probability*, Prentice-Hall, London.
- [101] Shannon, C. E. (1948). A mathematical theory of commmunication. *Bell System Tech. J.*, **27**, 379-423, 623-656.
- [102] Steutel, F. W. (1967). Note on infinity divisibility of exponential mixtures, *Ann. Math. Probab.*, **4**, 1303-1305.
- [103] Stigler, S. M. (1980). *American Contributions to Mathematical Statistics in the Nineteenth Century*, vol. II, Arno, New York.
- [104] Thiele, T. N. (1903). *Theory of Observations*, Layton, London (republicado em *Ann. Math. Stat.*, **2**, 165-308, 1931).
- [105] Tucker, H. (1970). Convolutions of distributions attracted to stable laws, *Ann. Math. Stat.*, **5**, 1381-1390.
- [106] Von Misses, (1981). *Probability Statistics and Truth*, Reprinted Dover, New York.
- [107] Winter, A. (1941). The singularities of Cauchy distributions, *Duke Mathematical Journal*, **8**, 678-681.
- [108] Zolotarev, V. M. (1986). *One-dimensional Stable Distributions*, Amer. Math. Soc., Providence.
- [109] Zolotarev, V. M., Uchaikin, V. V. (1999). *Chance and Stability. Stable Distributions and their Applications*, VSP, Utrecht.

Índice

- Anova, 15
- Auto-duais, 69
- Coefficiente
 - achatamento (curtose), 13
 - assimetria, 13
 - correlação, 14
- Condição de Cramér, 57
- Constantes
 - atracção, 17
 - normalizadoras, 17
 - padronização, 17
- Cumulantes, 13, 31
 - padronizados, 13
- Desigualdade
 - Cauchy-Schwarz, 35
 - Cramér-Rao, 38, 40, 97, 108
 - Gibbs, 85
- Distância
 - informação de Fisher, 97
 - Kullback-Leibler, 85
- Divisibilidade infinita, 20
- Domínio
 - atracção, 17
 - standard*, 118
 - da Gaussiana, 18
 - não *standard*, 118
- Entropia, 83
 - máxima, 92
- Equação
 - diferencial do calor, 103
 - estabilidade, 20
- Expansões Assintóticas
 - Cornish–Fisher, 64
 - distribuições estáveis, 61, 64
 - Edgeworth, 52, 56
 - Edgeworth diferidas, 57
 - Ponto Sela, 66, 68
 - Cornish–Fisher diferidas, 64
- Expoente característico, 18, 61
- Exponential tilting, 58
- Fórmula
 - Beckman, 80
 - Pollaczek–Khintchine, 79
- Família
 - associada, 58
 - exponencial, 42
 - exponencial natural (NEF), 42
- Família exponencial, 32
- Função
 - característica, 28
 - decomponível, 22
 - irreduzível, 22
 - lei estável, 61
 - dual, 69
 - geradora
 - cumulantes, 13, 32
 - momentos, 13, 27
 - pontuação (score), 93
 - variância quadrática (QVF), 42
 - variação
 - lenta, 18, 117
 - regular, 117
- Identidade
 - Bruijn, 98
 - Stein, 98
- Infinitamente divisível, 114
- Informação, 83

- estandardizada de Fisher, 94
- Fisher, 37, 92
- Fisher mínima, 99

- Lei dos grandes números, 9
- Leis estáveis para somas, 18

- Natural Exponential Families, 41
- NEF-QVF, 42

- Polinómios
 - ortogonais, 40
 - ortogonais de Meixner, 41
 - ortogonais para NEF-QVF, 47
- Probabilidade, 9

- Regressão linear, 16
- Representação
 - canónica de Lévy, 19
 - integral, 20
- Representação de Karamata, 117

- Somas padronizadas, 17

- Teorema
 - continuidade de Lévy-Cramér, 29
 - desigualdade de Cramér–Rao, 34
 - inversão de f_c de L^1 , 30
 - Limite Central, 10, 112
 - condição de Liapunov, 112
 - condição de Lindeberg, 113
 - Lindeberg-Feller, 113
- Tipos de Khintchine, 17

- Variação regular de Karamata, 118