Maria de Fátima Almeida Brilhante

Inferência Estatística em Modelos Não Gaussianos Com Recurso A Spacings E Outras Funções De Estatísticas Ordinais

Universidade dos Açores
Departamento de Matemática
Outubro de 1999

Trabalho parcialmente suportado pela FCT / PRAXIS XXI / FEDER

Dissertação apresentada no âmbito das provas para obtenção do grau de Doutor em Matemática, especialidade de Probabilidades e Estatística.

Sob Orientação:

Professor Doutor Dinis Duarte Ferreira Pestana, Professor Catedrático da Faculdade de Ciências da Universidade de Lisboa.

Agradecimentos

O meu orientador, Professor Doutor Dinis Duarte Pestana (FCUL), tem contribuído, desde os meus tempos de aluna de mestrado, para o enriquecimento do meu conhecimento de Probabilidades e Estatística. A constante partilha de ideias de investigação foram o fulcro em torno do qual se articulou o meu trabalho, e sem este câmbio dificilmente teria concretizado esta dissertação. Por isso quero deixar expressa a minha gratidão. O Professor Dinis é, incontestavelmente, um modelo a seguir (só espero não encontrar pelo caminho muito ”ruído”que me afaste demasiado do ”sinal”).

Ao Departamento de Matemática estou grata por ter-me apoiado em alturas cruciais da minha carreira académica. Em particular, uma palavra de apreço vai para o meu colega Professor Doutor José Carlos Rocha pelo seu encorajamento desde de o início e a quem devo a minha incursão no mundo da Estatística.

À Universidade dos Açores agradeço as facilidades concedidas, principalmente no que refere dispensa de serviço docente, o que permitiu dedicar-me exclusivamente à investigação condutora dos resultados aqui apresentados.

Os meus agradecimentos vão igualmente para o Centro de Estatística e Aplicações da Universidade de Lisboa, pelos meios computacionais e documentação disponibilizados, e sem os quais não teria sido possível prosseguir.

O espírito não trabalha sem o corpo — e por isso o meu agradecimento mais fundo vai para a minha família, com natural destaque para a minha Mãe, que me tem propiciado as condições que faltam a quase todas as mulheres para terem uma carreira profissional. Estou a propor-me à obtenção de um grau, que também é dela.

Índice

Resumo

O **Capítulo 1** aborda as perspectivas enquadradoras do desenvolvimento das Probabilidades e Estatística no século XX.

O modelo aditivo merece especial destaque na *secção 1.1*, enquanto na *secção 1.2* o teorema limite central, em especial a sua versão clássica, é recordado.

Na *secção 1.3* revêem-se resultados de convergência de classes e leis limites fracas no esquema aditivo. O correspondente esquema para máximos é também explorado.

A redefinição de problemas em termos de estatísticas ordinais e/ou funções de estatísticas ordinais merece particular atenção na *secção 1.4*, onde alguns resultados fundamentais são revisitados.

As leis infinitamente divisíveis e resultados afins são abordados com algum pormenor na *secção 1.5*, e na *secção 1.6* estabelece-se um paralelismo entre leis aditivas e valores extremos.

Na *secção 1.7* indica-se os eixos do desenvolvimento da Estatística, com especial destaque para a studentização e a análise de variância clássica, enquanto questões modernas de resistência e robustez são abordadas na *secção 1.8*.

Ao longo do capítulo procura-se perspectivar o trabalho desenvolvido, e salientar o seu interesse na evolução da abordagem aos diversos problemas.

No **Capítulo 2** revêem-se propriedades fundamentais dos membros da família Pareto generalizada na *secção 2.1* e que serão utilizadas em secções subsequentes. Alguns resultados novos também são estabelecidos, nomeadamente (i) testes que propiciam inferência sobre a escala (localização) das Pareto tipo I e tipo II baseados nos estimadores de máxima verosimilhança dos parâmetros; (ii) a distribuição dos *spacings* de Pareto generalizadas com índice não nulo; (iii) a distribuição de um produto generalizado de Paretos; (iv) um teste para homogeneidade de escala (localização) como extensão natural da análise de escala com parente exponencial.

Na *secção 2.2* um estudo formal dos *spacings* de misturas de exponenciais é efectuado, e estabelece-se ainda a divisibilidade infinita desses *spacings*.

A distribuição dos *spacings* (multiplicativos) de misturas de Pareto generalizadas é obtida na *secção 2.3* na sequência dos resultados da secção 2.2. A divisibilidade infinita dos *spacings* multiplicativos de misturas de Pareto tipo I e a dos logaritmos dos simétricos dos *spacings* multiplicativos de misturas de Pareto tipo II são também determinadas.

Na *secção 2.4* obtém-se a distribuição dos *spacings* com parente Laplaceana, assim como a divisibilidade infinita e formas de simetria dos *spacings* são estabelecidas.

No **Capítulo 3** propõe-se uma estatística para testar exponencialidade versus Pareto generalizada definida por um quociente entre a dispersão quartal superior e a dispersão quartal inferior. A principal vantagem da estatística, quando comparada com estatísticas homólogas existentes, reside num maior limite de ruptura, ou maior resistência a observações perturbadoras.

A distribuição exacta da estatística é obtida na *secção 3.1*, enquanto na *secção 3.2* obtém-se uma distribuição aproximada baseada no facto de os quartos serem uma forma de quartil. Uma tabela de quantis é apresentada para fins de inferência.

Para grandes amostras obtém-se na *secção 3.3* a distribuição Gaussiana como distribuição assintótica.

O cerne da dissertação é precedido por uma referência sumária a resultados analíticos que oportunamente se recorre ao longo do texto. Em apêndice explicita-se algumas funções densidade de probabilidade usadas na determinação dos quantis apresentados.

Abstract

Chapter 1 focuses the framework on which Probability and Statistics are being developed in the 20th Century.

The additive model deserves a special reference in *section 1.1*, while in *section 1.2* the central limit theorem, in its classic version, is recalled.

In *section 1.3* convergence of classes and weak limit laws are reviewed. The analogous scheme for maxima is also explored.

The redefinition of problems in terms of order statistics and/or functions of order statistics is mentioned in *section 1.4*, where some basic results are revisited.

Infinitely divisible laws and related topics are analysed with some detail in *section 1.5*, and in *section 1.6* a parallelism between additive laws and extreme values is established.

The foundations of Statistics are mentioned in *section 1.7*, where special attention is paid to studentization and analysis of variance. Modern issues of robustness and resistance are referred to in *section 1.8*.

Throughout this chapter we try to focus in appropriate perspective our work and, at the same time, we try to explain our contribution to the solution of some problems.

In **Chapter 2** some fundamental properties of the members of the generalized Pareto family are reviewed in *section 2.1*, since they will be used in subsequent sections. New results are also established in this section such as: (i) a test for inference about the scale (location) of Paretos of type I and type II based on the maximum likelihood estimators of the parameters; (ii) the distribution of the spacings of generalized Paretos with index different from zero; (iii) the distribution of a generalized product of Paretos; (iv) a test for homogeneity of scale/location with a Paretian parent, as a natural extension of the analysis of scale with exponential parent.

In *section 2.2* a formal study of the spacings of exponential mixtures is done. We also establish the infinite divisibility of those spacings.

The results obtained in section 2.2 are used in *section 2.3*, where the distribution of (multiplicative) spacings of mixtures of generalized Pareto is determined. Some arithmetic properties are also established.

In *section 2.4* the Laplacean spacings are studied, and the results are used to establish infinite divisibility.

In **Chapter 3** a statistic defined in terms of a ratio of upper fourth dispersion and lower fourth dispersion is proposed for testing exponentiality versus generalized Pareto. The main feature of this statistic, when compared with existing homologous statistics, relies in the fact that it has stronger rupture limit, i.e. greater resistance to nuisance values.

Its distribution is obtained in *section 3.1*, and in *section 3.2* an approximate distribution is established based on the fact that the fourths are a form of quartile. We also present a table of quantiles for testing purposes. We establish, in *section 3.3*, a central limit theorem for this statistic.

Introdução

As palavras profetas de H. G. Wells sobre a importância crescente da Estatística devem ser encaradas com maior seriedade, e não é de estranhar que no decurso das últimas décadas o paradigma da Estatística tenha sofrido alterações importantes. A Estatística deixou de ser apenas um instrumento de descrição da realidade, tornando-se cada vez mais a **ferramenta** apropriada para tomar decisões sob incerteza.

Como qualquer outra ciência, a eficácia da Estatística advém da sua capacidade de simplificar o real, oferecendo à nossa reflexão **modelos** — como sempre, a rasoura de Occam a permitir o avanço intelectual.

Como Einstein lapidarmente afirmou, um modelo deve ser tão simples quanto possível, mas não mais simples do que isto.

O modelo por excelência em Estatística é o modelo aditivo, e no campo mais restrito da Estatística paramétrica, os modelos de localização e escala apresentam, porventura, os maiores atractivos. Não é pois de estranhar que o modelo normal ou, em terminologia mais moderna Gaussiano, ocupe uma posição de relevo incontestada, e os grandes avanços da inferência estatística a ele estejam associados. Refiro-me obviamente, a studentização e a análise de variância, as duas ideias seminais que estão na génese de toda a Estatística moderna.

Uma e outra são, na essência, metodologias que permitem realizar inferência sobre o parâmetro de localização, usando o parâmetro de escala, ou como parâmetro espúrio que é necessário eliminar, ou como parâmetro ”instável”, fortemente influenciado pelo verdadeiro valor do parâmetro de localização. A evolução da Estatística levou, naturalmente, à consideração de inúmeras famílias de distribuições, em geral multiparametradas, sendo em geral a tarefa preliminar da análise estatística a singularização dos valores dos parâmetros por forma a melhor ajustar o modelo à realidade. Em muitas situações, um dos parâmetros tem um papel preponderante — por exemplo o expoente característico de uma lei estável, ou o índice de uma gama ou de uma Pareto generalizada. Nesta situação, uma perspectiva simplificadora será tentar, com álgebra apropriada, eliminar os outros parâmetros, relegando-os para o estatuto de parâmetros perturbadores.

Não é irrelevante referir desde já que mesmo a situação de famílias de localização e escala, de tão elegante solução no caso Gaussiano, foi durante décadas considerado intratável; ver os recentes avanços, *e.g.* Pestana e Rocha (1995) e Brilhante (1996). Esta situação levou mesmo ao desenvolvimento de todo o ramo da Estatística não paramétrica, em que a noção de modelo é *vaga*, admitindo-se apenas a continuidade da distribuição subjacente. Mas toda a problemática associada aos novos conceitos de robustez e resistência, e nomeadamente a observação de que as características amostrais de localização e escala tradicionalmente usadas têm limite de ruptura zero, tornaram imprescindível um investimento coerente na investigação de situações mais gerais. Isto levou, naturalmente, a uma importância crescente de metodologias que recorrem a funções de estatísticas ordinais.

Usamos na nossa abordagem os grandes princípios de redução de informação em que se baseia a espinha dorsal da Estatística: verosimilhança, suficiência, invariância, condicionamento. Ainda que não seja o nosso objectivo uma avaliação dos fundamentos da Estatística, ponderamos o valor intelectual daqueles esteios da inferência estatística, e da contestação construtiva a que são periodicamente sujeitos. Embora novas abordagens, privilegiando por exemplo o princípio do qui-quadrado mínimo em detrimento do princípio da verosimilhança máxima, tenham facetas tentadoras, os critérios de eficiência levaram-nos a alinhar pela posição mais clássica, usando assim verosimilhança como ponto de partida.

Fixamos como objectivo uma abordagem coerente a questões que, em termos gerais, podemos apontar como associadas ao conceito de peso das caudas. Trabalho anterior nosso, no caso especial de parente exponencial, em que a caracterização usando a independência de *spacings* teve um papel importante, levou-nos naturalmente ao estudo formal dos *spacings* de parente Laplaceana, a simetrização da exponencial, e ao estudo das suas propriedades aritméticas, no âmbito da teoria de divisibilidade infinita. No que refere a problemática central dos pesos de caudas, focamos a nossa atenção no modelo Pareto generalizado (de que o modelo exponencial é uma situação limite, de fronteira). Recordamos que o modelo Pareto generalizado oferece um largo espectro de situações de variação regular, que podemos considerar as fundamentais quer em termos de "*tail equivalence"de Resnick (1971), quer em termos de domínios de atracção não standard de Gomes e Pestana (1987). Neste caso, usamos transformações logarítmicas que permitem uma referência ao modelo exponencial, com vantagem concomitante de o parâmetro de forma surgir, na varíavel transformada, como parâmetro de escala.*

Alguns comentários parecem oportunos. Por um lado, a opção de reportar o que é novo e conhecido a teorias bem estabelecidas é um método comprovado em investigação científica, como é reconhecido pelos filósofos que vêm reflectindo sobre o conhecimento e criatividade. No entanto, esse reporte está longe de ser imediato; note-se por exemplo, que o parâmetro de forma da Pareto generalizada contém informação sobre a localização. Estamos assim longe da clareza e simplicidade do modelo de localização e escala clássico.

Os parágrafos que precedem procuram apenas situar, de forma muito ligeira, o campo em que desenvolvemos o nosso trabalho. Aspectos mais concretos do que acima foi dito serão abordados com extensão apropriada no Capítulo 1, para o qual reclamamos alguma originalidade no que refere as perspectivas enquadradoras do que tem sido o desenvolvimento da Estatística.

Nos restantes capítulos procedemos à exposição dos resultados da nossa investigação. Parece-nos essencial tornar claro as contribuições originais que uma dissertação apresenta. Assim, procedemos aqui a um arrolamento sumário dos resultados originais que apresentamos como fruto do nosso trabalho:

- *Face às propriedades dos quocientes entre estatísticas ordinais consecutivas (*spacings *multiplicativos) do modelo Pareto tipo I e Pareto tipo II, fundamentamos, na primeira secção do* **Capítulo 2**, *o uso de transformações logarítmicas a fim de propiciar inferência sobre o parâmetro de escala (localização). Outras propriedades de relativo interesse são expostas na secção 2.1. Assim, na subsecção 2.1.2, dedicada exclusivamente ao modelo Pareto tipo I (Pareto clássico), temos: (i) a distribuição de um produto generalizado de Paretos; (ii) a distribuição dos* spacings*; (iii) teste para homogeneidade de escala (localização). Um esquema semelhante ao anterior foi seguido para o modelo Pareto tipo II, abordado na subsecção 2.1.3.*
  A importância das propriedades dos spacings *do modelo exponencial, como por exemplo na caracterização do modelo, levou-nos a estudar formalmente na secção 2.2 os* spacings *de misturas de exponenciais. Observamos que estes são também misturas de exponenciais, e como propriedades relevantes salientamos a divisibilidade infinita dos* spacings. *Por outro lado, algumas generalizações para* spacings *de misturas generalizadas de exponenciais são estabelecidas.*
  Os resultados da secção 2.2 têm uma tradução a nível dos spacings *multiplicativos de misturas de Pareto generalizadas (Pareto tipo I e Pareto tipo II), e que assinalámos na secção 2.3. Das propriedades observadas destacamos a divisibilidade infinita dos* spacings *multiplicativos de misturas de Pareto tipo I e a dos*

logaritmos dos simétricos de spacings *multiplicativos de misturas de Pareto tipo II. A homogeneidade de escala das componentes das misturas é um importante factor na conclusão dos resultados acima mencionados.*
Os spacings *com parente Laplaceana são formalmente estudados na secção 2.4 por a Laplace ser a simetrização da exponencial. Constatamos que os* spacings *são misturas generalizadas de duas exponenciais, à excepção do* spacing *central para uma amostra de dimensão par que é uma mistura de exponencial e gama. Como propriedades, evidenciamos a proveniência dos coeficientes das misturas do modelo binomial simétrico, a divisibilidade infinita e uma espécie de simetria no que refere identidade distribucional de* spacings *equidistantes.*

- *Como as estatísticas de que temos conhecimento para testar exponencialidade versus Pareto generalizada apresentam limite de ruptura zero (ou por outras palavras, não são resistentes a observações perturbadoras), propomos no* **Capítulo 3** *uma estatística de teste onde as letras-resumo usadas (quartos e mediana) são garante de uma maior resistência, ou maior limite de ruptura. Assim, na secção 3.1 obtemos a distribuição amostral da referida estatística, enquanto na secção 3.2 obtemos uma distribuição aproximada baseada no facto de os quartos serem uma forma de quartil. Para efeitos de inferência apresentamos uma tabela de quantis para valores de* $n = 3(1)30$*. Na secção 3.3 obtemos a distribuição Gaussiana como distribuição assintótica.*

Preliminares

Família de Distribuições Gama

A fórmula de Euler para a função gama *é*

$$\Gamma(z) = \lim_{n\to\infty} \frac{n!n^z}{z(z+1)\dots(z+n)}\,, \quad z \neq 0, -1, -2, \dots. \tag{0.0.1}$$

No entanto, a restrição da função gama aos reais positivos permite a representação integral

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt, \quad \alpha > 0 \tag{0.0.2}$$

conhecida também por integral de Euler de 2ª espécie.

Assim, de (0.0.2) *podemos obter a função densidade de probabilidade de uma* gama com índice α

$$f_X(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} I_{(0,\infty)} \tag{0.0.3}$$

e que denotamos por $X \frown G(\alpha)$.

Por outro lado, se considerarmos a transformação $X = \frac{Y-\lambda}{\delta}$ *com* $\lambda \in IR$ *e* $\delta > 0$, *obtemos uma gama com índice* α, *localização* λ *e escala* δ, $Y \frown G(\alpha, \delta, \lambda)$, *com função densidade de probabilidade*

$$f_Y(y) = \frac{(y-\lambda)^{\alpha-1}e^{-(x-\lambda)/\delta}}{\delta^\alpha\Gamma(\alpha)} I_{(\lambda,\infty)}. \tag{0.0.4}$$

Facilmente se verifica que se $Y \frown G(\alpha, \delta, \lambda)$, *então* $kY \frown G(\alpha, k\delta, k\lambda)$ *se* $k > 0$.

Como casos particulares da distribuição gama temos a distribuição exponencial para $\alpha = 1$, *e para* $\alpha = n/2$ *(*n *inteiro) e* $\delta = 2$ *a distribuição do qui-quadrado com* n *graus de liberdade,* $\chi^2_{(n)}$.

Ora se $X \frown \chi^2_{(m)}$ *e* $Y \frown \chi^2_{(n)}$ *forem v.a. independentes, então* $\frac{X/m}{Y/n} \frown F_{(m,n)}$, *ou seja, o quociente anterior tem distribuição* F *de Snedecor com* m *e* n *graus de liberdade.*

Por outro, se $Y_1, \ldots, Y_n$ *forem gamas independentes com* $Y_k \frown G(\alpha_k, \delta, \lambda_k)$, *então* $\sum_{k=1}^n X_k \frown G(\sum_{k=1}^n \alpha_k, \delta, \sum_{k=1}^n \lambda_k)$. *Em particular, a soma de* n *exponenciais independentes e identicamente distribuídas (i.i.d.) com escala* δ *é uma gama com índice* n *e escala* δ.

De (0.0.1) *é fácil verificar que*

$$\Gamma(z+1) = z\Gamma(z), \quad z \neq 0, -1, -2, \ldots \tag{0.0.5}$$

enquanto de (0.0.2) *é imediato que* $\Gamma(1) = 1$. *Assim, se* n *for um inteiro positivo, obtém-se após aplicação sucessiva de* (0.0.5), $\Gamma(n+1) = n!$. *Por analogia, podemos considerar*

$$\Gamma(z+1) = z!, \quad z \neq 0, -1, -2, \ldots$$

sendo a função gama também denominada função factorial.

Ora se $X \frown G(\alpha)$, *obtém-se*

$$E[X^n] = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} = (\alpha)_n \tag{0.0.6}$$

onde $(\alpha)_n = \alpha(\alpha+1)\cdots(\alpha+n-1)$ *é o símbolo de Pochhammer com a convenção* $(\alpha)_0 = 1$. *Assim,*

$$Y = \delta X + \lambda \frown G(\alpha, \delta, \lambda) \Rightarrow \begin{cases} E[Y] = \alpha\delta + \lambda \\ Var[Y] = \alpha\delta^2 \end{cases} .$$

Algumas transformadas integrais, em particular as do tipo $\int_{-\infty}^{\infty} e^{itx} dF_X(x) = E[e^{itX}]$, *têm sido ferramentas poderosas no estudo de funções* *de distribuição. Transformadas do tipo anterior são vulgarmente conhecidas por transformadas de Fourier. Porém, em probabilidades são apelidadas* funções características.

A função característica da distribuição gama com índice α é

$$\begin{aligned} E[e^{itX}] &= \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{itx} x^{\alpha-1} e^{-x} dx \\ &= \frac{1}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(1-it)x} dx \\ &= \frac{(1-it)^{-\alpha}}{\Gamma(\alpha)} \int_0^\infty t^{\alpha-1} e^{-t} dt \\ &= (1-it)^{-\alpha}. \end{aligned}$$

Assim, se um parâmetro de localização λ e um parâmetro de escala δ forem considerados, a função característica tem o aspecto

$$E[e^{itY}] = E[e^{it(\delta X+\lambda)}] = e^{i\lambda t}(1-i\delta t)^{-\alpha}. \tag{0.0.7}$$

As funções relacionadas com a função gama e que usaremos nesta dissertação são:

(i) Função gama incompleta

$$\Gamma(\alpha, x) = \int_x^\infty t^{\alpha-1} e^{-t} dt, \quad x > 0.$$

A função

$$\gamma(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} dt$$

também é conhecida por função gama incompleta, onde

$$F_X(x) = P\{X \leq x\} = \frac{\gamma(\alpha, x)}{\Gamma(\alpha)} \quad \textit{se } X \frown G(\alpha).$$

(ii) Função poligama de ordem n

$$\psi^{(n)}(z) = \frac{d^{n+1}}{dz^{n+1}} \ln \Gamma(z)$$

Se $n = 0$*, obtém-se a* função digama

$$\psi(z) = \frac{d \ln \Gamma(z)}{dz} = \frac{\Gamma'(z)}{\Gamma(z)}.$$

Também usaremos as relações

$$\psi(n) = -\gamma + \sum_{k=1}^{n-1} k^{-1}, \quad n = 2, 3, \ldots \tag{0.0.8}$$

onde γ *é a constante de Euler (*$\gamma = 0.5772156649...$*),*

$$\psi(n+z) = \sum_{k=1}^{n} \frac{1}{(n-k)+z} + \psi(z), \quad n = 1, 2, \ldots \tag{0.0.9}$$

e

$$\psi^{(m)}(n+1) = (-1)^m m! \left[-\zeta(m+1) + 1 + \frac{1}{2^{m+1}} + \cdots + \frac{1}{n^{m+1}} \right], \quad n, m = 1, 2 \ldots \tag{0.0.10}$$

onde ζ *é a* função zeta de Riemann*, isto é,* $\zeta(s) = \sum_{k=1}^{\infty} k^{-s},\ s > 1$.

Família de Distribuições Beta

Da função beta *ou* integral de Euler de 1ª espécie

$$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt, \quad p > 0,\ q > 0 \tag{0.0.11}$$

podemos definir a função densidade de probabilidade (f.d.p.) de uma beta standard

$$f_X(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} I_{(0,1)}. \tag{0.0.12}$$

Uma variável aleatória (v.a.) X com função densidade de probabilidade (0.0.12) *diz-se uma* beta standard *com parâmetros p e q, e denotamos por $X \frown \mathcal{B}e(p,q)$. Se considerarmos a transformação $X = \frac{Y-a}{b-a}$, situação que corresponde a considerarmos uma localização a e uma escala $(b-a)$, obtemos uma v.a. beta com parâmetros p, q, a e b com função densidade de probabilidade*

$$f_Y(y) = \frac{1}{B(p,q)} \frac{(y-a)^{p-1}(b-y)^{q-1}}{(b-a)^{p+q-1}} I_{(a,b)} \tag{0.0.13}$$

indicando-se por $Y \frown \mathcal{B}e(p,q,a,b)$.

Repare-se que a distribuição uniforme em (a,b), denotemos $U \frown U(a,b)$, é uma beta com $p = q = 1$. A uniforme em $(0,1)$, ou uniforme standard, é bastante importante em Estatística devido ao teorema da transformação uniformizante: se X for uma v.a. contínua com função de distribuição F_X, então $Y = F_X(X) \frown U(0,1)$. *As estatísticas ordinais da uniforme são betas; mais precisamente, se $U_{i:n}$ for a i-ésima estatística ordinal de n v.a. i.i.d. uniformes em $(0,1)$, então $U_{i:n}$ é uma beta standard com parâmetros i e $n-i+1$.*

Se em (0.0.12) *tomarmos $q = 1$, obtemos a função densidade de probabilidade da distribuição função potência standard com parâmetro p. E se em* (0.0.12) *efectuarmos a mudança de variável $t = \frac{x}{1-x}$, obtemos uma* beta de 2ª espécie *com parâmetros p e q, isto é, com f.d.p.*

$$f_T(t) = \frac{1}{B(p,q)} \frac{t^{p-1}}{(1+t)^{p+q}} I_{(0,\infty)} \tag{0.0.14}$$

denotando-se $T \frown \mathcal{B}e^*(p,q)$.

É fácil verificarmos que se $X \frown G(\alpha_1,\delta)$ e $Y \frown G(\alpha_2,\delta)$ forem v.a. independentes, então $Z = \frac{X}{Y} \frown \mathcal{B}e^*(\alpha_1,\alpha_2)$.

As funções beta e gama estão relacionadas entre si da forma

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}. \tag{0.0.15}$$

Assim, se $X \frown \mathcal{B}e(p,q)$, tem-se

$$E[X^n] = \frac{B(p+n,q)}{B(p,q)} = \frac{\Gamma(p+n)}{\Gamma(p)} \frac{\Gamma(p+q)}{\Gamma(p+q+n)} = \frac{(p)_n}{(p+q)_n}. \tag{0.0.16}$$

Consequentemente,

$$Y = (b-a)X + a \frown \mathcal{B}e(p,q,a,b) \Rightarrow \begin{cases} E[Y] = (b-a)\frac{p}{p+q} + a \\ Var[Y] = (b-a)^2 \frac{pq}{(p+q)^2(p+q+1)} \end{cases}.$$

As funções

$$B_x(p,q) = \int_0^x t^{p-1}(1-t)^{q-1}dt, \quad 0 < x < 1 \tag{0.0.17}$$

e

$$I_x(p,q) = \frac{B_x(p,q)}{B(p,q)}, \quad 0 < x < 1 \tag{0.0.18}$$

são conhecidas por funções beta incompletas.

Usando integração por partes, é fácil verificar que

$$I_p(a, n-a+1) = \sum_{k=a}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \tag{0.0.19}$$

Por outras palavras, se $X \frown \mathcal{B}e(a, n-a+1)$ *e* $Y \frown b(n,p)$ *(isto é, uma binomial com parâmetros* n *e* p*), então*

$$P\{X \leq p\} = I_p(a, n-a+1) = P\{Y \geq a\}.$$

No que refere a função característica de $X \frown \mathcal{B}e(p,q)$*, tem-se*

$$\begin{aligned} E[e^{itX}] &= \frac{1}{B(p,q)} \int_0^1 e^{itx} x^{p-1}(1-x)^{q-1}dx \\ &= \frac{1}{B(p,q)} \int_0^1 \sum_{n=0}^{\infty} \frac{(itx)^n}{n!} x^{p-1}(1-x)^{q-1}dx \\ &= \frac{1}{B(p,q)} \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \int_0^1 x^{p+n-1}(1-x)^{q-1}dx \\ &= \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \frac{B(p+n,q)}{B(p,q)} \\ &= \sum_{n=0}^{\infty} \frac{(p)_n}{(p+q)_n} \frac{(it)^n}{n!} = M(p, p+q, it) \end{aligned}$$

onde $M(a,b,z) = \sum\limits_{n=0}^{\infty} \frac{(a)_n}{(b)_n} \frac{z^n}{n!}$ *é a* função hipergeométrica confluente.

Nesta dissertação aparecerá outro tipo de função hipergeométrica, a função hipergeométrica de Gauss*, definida por*

$$ {}_2F_1(a,b;c;z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}. \tag{0.0.20}$$

Também necessitaremos da relação

$$\frac{\partial B(p,q)}{\partial p} = \int_0^1 \ln t \, t^{p-1}(1-t)^{q-1} dt = B(p,q)[\psi(p) - \psi(p+q)]. \tag{0.0.21}$$

Capítulo 1

O Desenvolvimento das Probabilidades e da Estatística no século XX

1.1 O Modelo Aditivo — uma primeira perspectiva sobre o desenvolvimento da Estatística

No magistral prefácio de um dos grandes clássicos da Probabilidade — Limit Distributions for Sums of Independent Random Variables *— Gnedenko e Kolmogoroff (1954) escrevem que o verdadeiro valor epistemológico da Teoria da Probabilidade advém dos teoremas limites.*

A afirmação, ao tempo, tinha todo o sentido: os meios de cálculo disponíveis nos meados do século XX levaram a que modelos tratáveis fossem em número reduzido e a sua aplicabilidade fortemente ligada a tratabilidade matemática, da função densidade de probabilidade ou de alguma transformada integral, nomeadamente a função característica. Volvidos 30 anos, as publicações de Exploratory Data Analysis *de Tukey (1977) e de* Data Analysis and Regression: A Second Course in Statistics *de Mosteller e Tukey (1977), apresentam um panorama totalmente diferente: os meios computacionais permitem o tratamento exacto de distribuições amostrais de pequenas amostras, e colocam os estudos de robustez e resistência, e mais geralmente de simulação estatística, no primeiro plano de investigação, e também das aplicações da Estatística.*

Esta revolução do pensamento estatístico merece alguns comentários. É importante anotar que o paradigma da Ciência vai mudando, por choques culturais, umas vezes de evoluções técnicas que revolucionam a nossa visão do Mundo, outras pelo desenvolvi-

mento de novos ramos da Ciência.

Exemplo do primeiro caso é o "método"de Descartes (e se o Discurso do Método *merece em geral prioridade na referência a Descartes, não deixa de valer a pena citar as suas* Meditações Metafísicas, *em que expressa desassombradamente a sua confiança absoluta no racionalismo: "os dados da razão pura não saberiam enganar-nos"). O método postulado por Descartes — se tem um problema complicado a resolver, "parta-o"em problemas mais simples, resolva-os e "cole"as soluções — invade todo o desenvolvimento do conhecimento, e em particular a Teoria da Probabilidade que então dá os primeiros passos. De facto, o teorema da probabilidade total mais não é que o método de Descartes vazado em expressão matemática. (E basta recordar algumas aplicações do teorema da probabilidade total, tão diversas quanto a fundamentação da amostragem estratificada e a resolução de problemas finos de processos estocásticos à custa de uma "partição"das trajectórias, para se avaliar o manancial de resultados advenientes da aplicação do "método".)*

Exemplo do segundo caso é o desenvolvimento de técnicas que permitem uma medição mais precisa. Basta referir a revolução que trouxeram à Ciência o microscópio electrónico ou os radio-telescópios para alicerçar a afirmação de que avanços tecnológicos numa área podem ter efeitos de revolução profunda em outras áreas da Ciência.

Como exemplo do terceiro caso, podemos referir os resultados, ainda incipientes, da Teoria da Probabilidade do século XIX, e especificamente o teorema limite central, na alteração do paradigma da Ciência. De facto, até ao século XIX, o desenvolvimento da Ciência evoluíu com a capacidade de medir, de forma cada vez mais precisa. Muitas vezes não se chama a devida atenção para a importância fundamental que teve para a evolução da Ciência a criação de medidas padrão, e o convencionamento de sistemas de medida adoptadas internacionalmente pelos cientistas. Sem capacidade de dividir cada vez mais a unidade (aproximando a realidade do ideal matemático do contínuo) não haveria física atómica. Assim, a corrente de desenvolvimento da Ciência apontava, decididamente, no sentido de afinar cada vez mais a capacidade de medir com precisão.

Ora o que o teorema limite central veio mostrar foi que, em muitas circunstâncias (e porque entretanto a evolução das ideias em Ciência tinha levado à percepção de que o próprio procedimento de medida comporta, inevitavelmente, erro), mais vale medir muitas vezes de forma imprecisa, e usar médias, de que medir apenas uma vez, com grande esforço de precisão.

De facto, uma das regras da Ciência ocidental é a rasoura de Occam, a adopção de modelos com representação simples da realidade inatingível (e cite-se, a propósito a *lapidar frase de Einstein, reformulação moderna da rasoura de Occam:* "make things as simple as possible; but not simpler". De entre todos os modelos, o mais universal é o modelo aditivo

$$x_k = M + \varepsilon_k$$

ou seja, cada "dado"ou "observação"x_k corresponde a uma corrupção do modelo M, devido a um "erro"aditivo, ou resíduo ε_k. Por outras palavras, há um *ruído* ε_k que perturba o *sinal* M que pretendemos conhecer.

Em condições muito genéricas — as do teorema limite central — a melhor forma de conhecer o modelo M é tomar médias de "muitos"dados x_k.

De facto, admitindo que o ruído ε_k flutua aleatoriamente em torno de 0 (zero) — mais especificamente, tem valor esperado 0 e variância finita σ^2 — a média

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k = M + \overline{\varepsilon}$$

onde $\overline{\varepsilon} = \frac{1}{n}\sum\limits_{k=1}^{n} \varepsilon_k \sim Gau\big(0, \frac{\sigma}{\sqrt{n}}\big)$, e assim, pela desigualdade de Chebycheff por exemplo, $P\{|\overline{\varepsilon}| \geq \delta\} \leq \frac{\sigma^2}{n\delta^2} \xrightarrow[n\to\infty]{} 0$, ou seja, $\overline{x}$ aproxima-se cada vez mais do modelo M. (A desigualdade acima é genérica, mas grosseira, e o teorema limite central é, nestas situações, um refinamento notável — de facto, podemos afirmar que $P\Big[M - 3\frac{\sigma}{\sqrt{n}} < \overline{x} \leq M + 3\frac{\sigma}{\sqrt{n}}\Big] \approx 99.7\%$.)

Este contributo notável da Teoria da Probabilidade para a evolução da Ciência tem, como todos, naturais limitações:

- a velocidade de convergência, da ordem de $n^{-1/2}$, é baixa.
- da anterior afirmação decorre, naturalmente, desconfiança sobre o recurso a resultados assintóticos, quando de facto as nossas amostras são sempre finitas.
- as condições do teorema limite central clássico (e nomeadamente condições sobre o 2º momento) nem sempre se verificam, ainda que nas ciências da natureza seja difícil de imaginar situações dessas.
- há situações experimentais tão caras (por exemplo, produção de elementos transuranianos em grandes aceleradores) em que dificilmente se aceita a repetição de experiências. Nesses casos, claro que há que investir no rigor de uma única medição, e não em medições repetidas mais grosseiras.

Apesar das limitações acima apontadas, o teorema limite central clássico foi, longamente, o esteio de toda a Estatística. Na primeira metade do nosso século, os meios computacionais não permitiam usar modelos sofisticados (por exemplo, modelos robustos, ou modelos ajustados por estimação não paramétrica ou semi-paramétrica), e passou a haver um recurso sistemático – e por vezes acrítico – ao modelo ”normal”ou Gaussiano, implicitamente justificado pelo teorema limite central.

Um dos grandes bioestatísticos dos anos 30, Winsor, considerou mesmo que ”na parte central todos os modelos são normais”. (Ainda hoje se chama winsorização à técnica que consiste em truncar as pontas extremas do modelo, e reescalá-lo por condicionamento, por forma a poder facilmente usar, para a parte central, uma aproximação Gaussiana.)

Mas como tão bem apontou o grande biólogo Huxley, em Ciência muitas grandes ideias nascem como heresia, morrem como dogma. As ideias de Winsor, que revolucionaram a aplicabilidade da Estatística quando ele as enunciou, tornaram-se obsoletas com o advento das modernas técnicas computacionais, que abriam o caminho para distribuições amostrais exactas e simulação com base na função de distribuição empírica, e todos os desenvolvimentos possíveis com estatística computacional.

Assim, nos anos sessenta o trabalho pioneiro de Hotelling (1961) sobre distribuições exactas faz uma crítica certeira às ideias de Winsor, e explicita as enormes limitações de recurso ao teorema limite central: a aproximação é em geral boa na parte central, mas má nas caudas, e afinal toda a inferência estatística assenta na avaliação de quantis situados nas caudas!

1.2 O Teorema Limite Central Clássico

Na secção anterior referimos por diversas vezes o teorema limite central, sem uma explicitação rigorosa. Legitima-nos o facto de ser porventura o resultado mais celebrado da Teoria da Probabilidade (ainda recentemente Hoffmann-Jorgensen *et al.* (1976) publicaram um curso intitulado *The Two Pearls of Probability Theory*, sendo as duas pérolas a lei dos grandes números e o teorema limite central).

Recordemos, de qualquer forma, este extraordinário resultado. Na sua formulação clássica, afirma muito simplesmente que a soma de variáveis aleatórias independentes e identicamente distribuídas com variância finita se aproxima de uma v.a. Gaussiana.

Começou por ser estabelecido, para distribuições de Bernoulli, e com exigências diversas sobre momentos, mas a afinação de instrumentos como as funções características permitiu enfim estabelecer, com base nas hipóteses acima, aquela condição suficiente.

De facto, se $S_n = \sum_{k=1}^{n} X_k$, e $Var(X_k) = \sigma^2 < \infty$, denotando $\mu = E(X_k)$, então $E(S_n) = n\mu$ e $Var(S_n) = n\sigma^2$. Então, denotando a função característica (f.c.) $\varphi_X(t) = E\big[e^{itX}\big]$

$$\begin{aligned}
\varphi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) &= \varphi_{\sum\left(\frac{X_k-\mu}{\sigma}\right)}(t/\sqrt{n}) \\
&= \varphi^n_{\frac{X_k-\mu}{\sigma}}(t/\sqrt{n}) \\
&= \left[1 - \frac{t^2/2}{n} + o\left(\frac{t^2}{n}\right)\right]^n \xrightarrow[n\to\infty]{} e^{-t^2/2}
\end{aligned}$$

e consequentemente, pelo teorema da continuidade de Cramér,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{w} Z \frown Gau(0,1).$$

Mesmo nesta formulação clássica simples, o resultado acima abraça um conjunto de corolários notáveis. O mais importante é decerto o que corresponde a observar que a média de v.a. i.i.d. com valor médio μ e desvio padrão σ é também uma soma de v.a. i.i.d. com variância finita, e consequentemente

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{w} Z \frown Gau(0,1).$$

Ou, por outras palavras, $\overline{X}_n \sim Gau\big(\mu, \frac{\sigma}{\sqrt{n}}\big)$, o que quer dizer que $\overline{X}_n$ se aproxima de μ, pois $\frac{\sigma}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0$. Assim, as experiências repetidas usadas nas ciências experimentais ganham um novo significado. Como atrás referimos, em condições "normais"a precisão do resultado pode ser atingida com "muitas"medições grosseiras, baratas, em vez de apostar na medição "cara"com um instrumento de precisão.

Claro que o preço a pagar é alto, no sentido em que $\frac{\sigma}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0$ à velocidade de $n^{-1/2}$. Assim, para reduzir σ pela escala 10, são necessárias 100 medições. Mas para reduzir de novo pelo factor $\frac{1}{10}$, são necessárias *mais* 9900 medições!

Chamamos a atenção para o facto de o teorema limite central clássico ser uma condição suficiente, mas não necessária. O aspecto mais espectacular é que qualquer das hipóteses pode ser relaxada, obtendo ainda como limite o modelo Gaussiano. Por exemplo, em vez de variância finita basta limitação das variâncias truncadas, as condições de identidade distribucional podem ser substituídas pelas condições de Lindeberg-Feller sobre limitação da sucessão de variâncias truncadas, e a hipótese de independência pode ser relaxada para diversas formas de dependência markoviana, estacionaridade ou dependência de martingalas. Os livros de Petrov (1995) ou de Ibragimov e Linnik (1971) são excelentes panorâmicas dos muitos desenvolvimentos do teorema limite central clássico.

1.3 Convergência de Classes e Leis Limites Fracas no Esquema Aditivo

A história da Teoria da Probabilidade mudou quando em 1917 um jovem professor de Mecânica Racional da École Polytechnique recebeu a encomenda de fazer 3 conferências sobre o teorema limite central clássico. Chamava-se ele Paul Lévy, e achou o teorema clássico muito insatisfatório — era uma condição suficiente, mas que lhe pareceu admitir relaxações não triviais nas hipóteses.

Esta insatisfação levou-o a ser o probabilista mais influente no desenvolvimento da Teoria da Probabilidade, trazendo as transformadas de Fourier para a Teoria da Probabilidade, criando a noção de estabilidade e de domínios de atracção (de 1917 a 1925, data de publicação do seu *Calcul des Probabilités*, com solução completa sobre os domínios de atracção feita pelo seu discípulo Doeblin em 1940), a inventar seguidamente as leis infinitamente divisíveis (cuja apresentação culmina no seu *Théorie de L'Addition des Variables Aléatoires*, de 1937, que Spitzer (1961) não hesitou em considerar o primeiro livro moderno de Teoria da Probabilidade), e pelo caminho vai produzindo estudos sobre o "espaço diferencial de Wiener"que transformam o movimento Browniano (de que o 1º modelo é o de Einstein e Schmoluvsky) no processo estocástico de Wiener-Lévy. O seu *Processus Stochastiques et Mouvement Brownien* (1965) é uma terceira obra-prima, culminando o trabalho de criação dos processos estocásticos — e Feller, no seu obituário de Lévy, faz um elogio à sua extraordinária capacidade criativa, dizendo que "era como se ele conseguisse seguir com os olhos as infinitas trajectórias dos processos estocásticos".

Abordemos então a primeira grande criação de Lévy nos processos aditivos, as leis estáveis e seus domínios de atracção.

No teorema limite central clássico, na expressão

$$\frac{S_n - n\mu}{\sigma\sqrt{n}}$$

as constantes de atracção $B_n = n\mu \in I\!R$ e $A_n = \sigma\sqrt{n} > 0$ têm como efeito "puxar"S_n

para 0, e estabilizar a dispersão usando a escala intrínseca $\sigma\sqrt{n}$, respectivamente.

Lévy interrogou-se, mais geralmente: Dadas v.a. i.i.d. X_k, e denotando $S_n = \sum_{k=1}^{n} X_k$, em que condições existem constantes de atracção centralizadoras $B_n \in IR$ e normalizadoras $A_n > 0$ tais que

$$\frac{S_n - B_n}{A_n} \xrightarrow[n\to\infty]{w} Y \quad (Y \text{ não degenerada}).$$

Paul Lévy começou por demonstrar que as possíveis leis limites são *estáveis*, no sentido em que, se Y_j forem réplicas independentes de Y, existem constantes $\alpha_n > 0$ e $\beta_n \in IR$ tais que

$$\frac{\sum_{j=1}^{n} Y_j - \beta_n}{\alpha_n} \stackrel{d}{=} Y.$$

As constantes α_n são necessariamente da forma $\alpha_n = kn^{1/\alpha}$, com $\alpha \in (0, 2]$. A $\alpha = 2$ corresponde a Gaussiana, a única lei estável com variância finita; as outras densidades estáveis conhecidas correspondem a $\alpha = 1$, simétrica, que é a Cauchy com f.d.p. $f(x) = \frac{1}{\pi}\frac{1}{1+x^2} I_{IR}$ e f.c. $\varphi(t) = e^{-|t|}$, e a $\alpha = \frac{1}{2}$ e suporte positivo, a Lévy com f.d.p. $f(x) = \frac{1}{\sqrt{2\pi x^3}} e^{-1/(2x)} I_{(0,\infty)}$ e f.c. $\varphi(t) = \exp\left\{-|t|^{1/2}\left(1 - i\frac{t}{|t|}\right)\right\}$.

Genericamente, as f.c. das leis estáveis são da forma

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

onde $a \in IR$ é um parâmetro de localização, $c > 0$ é um parâmetro de escala, $\alpha \in (0, 2]$ é o expoente característico, $\beta \in [-1, 1]$ é um parâmetro de assimetria — $\beta = 0$ corresponde a simetria — e

$$\omega(|t|, \alpha) = \begin{cases} \mathrm{tg}\left(\frac{\pi\alpha}{2}\right) & , \text{ se } \alpha \neq 1 \\ \frac{2}{\pi} \ln|t| & , \text{ se } \alpha = 1 \end{cases}.$$

As condições necessárias e suficientes para $\frac{\sum_{k=1}^{n} X_k - B_n}{A_n} \longrightarrow Y_\alpha$, onde Y_α denota uma estável com expoente característico α, tem que ver com a variação lenta da sequência de variâncias truncadas no caso $\alpha = 2$, e com variação regular da soma das caudas

$$\frac{F(-tx) + [1 - F(tx)]}{F(-x) + [1 - F(x)]} \xrightarrow[x\to\infty]{} t^{-\alpha}$$

no caso $\alpha \in (0, 2]$.

Para detalhes, cf. Gnedenko e Kolmogoroff (1954). Usando os teoremas tauberianos, as condições sobre caudas podem ser reformuladas em termos de variação regular da f.c. na vizinhança da origem, cf. Ibragimov e Linnik (1971).

O conceito de variação lenta e de variação regular, inventados por Karamata (1930) são hoje a base da formulação geral de condições de convergência, desde que Feller (1967) os usou para enunciar de forma mais concisa e elegante as condições de convergência de Doeblin (1940); para detalhes, cf. o tratado de Bingham, Goldie e Steutel (1987).

Esta formulação mais abrangente do teorema limite central clássico cedo produziu frutos: um discípulo de Lévy, Fréchet (1927), estudou uma "equação de estabilidade"no que refere o esquema de máximos, isto é, interrogou-se sobre a convergência de

$$\frac{\max(X_1, \ldots, X_n) - B_n}{A_n} \longrightarrow Y, \quad \text{não degenerada}$$

com $A_n > 0$ e $B_n \in I\!R$, concluíndo que os possíveis limites verificam a condição de estabilidade

$$\frac{\max(Y_1, \ldots, Y_n) - B_n}{A_n} \stackrel{d}{=} Y$$

e concluíu que as v.a. "de Fréchet", com função de distribuição do tipo

$$F_Y(y) = \exp\{-y^{-\alpha}\} I_{(0,\infty)}, \quad \alpha > 0$$

são estáveis para máximos.

Pouco depois Fisher e Tippett (1928) mostraram que as distribuições estáveis para máximos são do tipo Gumbel

$$\Lambda(x) = \exp\{-e^{-x}\} I_{I\!R}$$

ou do tipo Fréchet

$$\Phi_\alpha(x) = \begin{cases} 0 & , \ x \leq 0 \\ \exp\{-x^{-\alpha}\} & , \ x > 0 \end{cases}, \qquad \alpha > 0$$

ou do tipo Weibull

$$\Psi_\alpha(t) = \begin{cases} \exp\{-(-x)^\alpha\} & , \ x < 0 \\ 1 & , \ x \geq 0 \end{cases}, \quad \alpha > 0$$

(von Mises e Jenkinson propuseram uma reparametrização que engloba os três tipos). A demonstração geral de que são as únicas leis estáveis para máximos deve-se a Gnedenko (1943), que caracteriza os domínios de atracção (note-se que Gnedenko tinha independentemente de Doeblin, caracterizado os domínios de atracção no esquema aditivo).

Abordagens mais modernas do esquema de extremos reformulam as questões em termos de variação lenta e variação regular. O estudo das constantes de atracção levou naturalmente ao aprofundamento das excedências de níveis elevados, e isso naturalmente à abordagem POT (*peaks over thresholds*), em que as Pareto generalizadas aparecem como "estáveis"neste novo esquema. O conceito de "*tail equivalence", de Resnick (1971), é a parte que liga Pareto generalizadas e a expressão unificada de von Mises-Jenkinson.*

O conceito de variação lenta permitiu mais recentemente, a Urbanik (1973), estudar "slowly varying triangular arrays"e refinar a classe de leis auto-decomponíveis de Linnik, de que adiante falaremos.

Por outro lado, Bingham (1971) estudou em termos abstractos as convoluções generalizadas e seus domínios de atracção, obtendo de novo condições de convergência em termos de variação regular.

O esquema multiplicativo foi abordado por Zolotarev (1957) e por Athayde (1984).

Anotemos, ainda, que o que está em causa é uma convergência de classes, no sentido em que para cada n se "selecciona"um ponto da classe de todos os $\frac{S_n - B_n}{A_n}$, uma transformação linear dos S_n.

Assim, a formulação rigorosa de todas estas questões acabou por só ser possível após a definição de *tipo de Khinchine*, as transformações lineares com declive positivo. Consequentemente, as definições das leis limite são em termos de tipos.

Note-se que o esquema acima pode ser alargado não trivialmente, considerando para cada variável aleatória a sua própria normalização, isto é, estudando limites de

$$\sum_{k=1}^{n} \frac{X_k - B_{k,n}}{A_{k,n}}.$$

As classes de leis limites, que são convexas, estão a ser correntemente investigadas por Mendonça, que se ocupou do problema análogo no esquema de extremos.

Idealmente, em vez de nos restringirmos a transformações lineares seria bem mais interessante com transformações gerais passar da distribuição "parente"à distribuição limite.

Como Balkema observa, o teorema da transformação uniformizante permitiria, idealmente, fazer a passagem em dois passos

$$\text{parente } X \frown F_X(.) \longrightarrow U = F_X(X) \longrightarrow \text{limite } Y = F_Y^{-1}(U)$$

mas em modelação estatística as situações concretas são bem distintas desta situação ideal (a parente é em geral desconhecida).

Note-se ainda, lateralmente, que a abordagem clássica não comporta asserções sobre velocidades de convergência. Fisher e Tippett (1928) chamaram a atenção para o facto de a convergência poder ser muito lenta, exemplificando com parente Gaussiana cujos máximos convergem para a Gumbel: mesmo para $n = 10^{12}$ réplicas, a distribuição exacta do máximo ainda estava longe do limite Gumbel Λ, havendo uma melhor aproximação pré-assintótica (*"penultimate") usando uma Weibull Ψ_α, com α apropriado.*

O problema das aproximações pré-assintóticas no caso de limite Gumbel foi genericamente explicado por Gomes (1978), e Gomes e Pestana (1987) mostraram que no caso de "domínios de atracção não standard (isto é, com constantes normalizadoras $\alpha_n = n^{1/\alpha}L(n)$ com $L(n)$ de variação lenta e $L(n) \underset{n\to\infty}{\longrightarrow} 0$, ou $L(n) \underset{n\to\infty}{\longrightarrow} \infty$), as aproximações penultimate *são a regra, e não a excepção, no esquema de máximos. Iglésias Pereira, Oliveira e Pestana (1996) abordaram o problema de aproximações pré-assintóticas no esquema de somas.*

1.4 Estatísticas Ordinais e Funções de Estatísticas Ordinais

Seja $(X_1, X_2, \ldots, X_n)$ um vector aleatório com margens univariadas X_k i.i.d., com função de distribuição (f.d.) F_X. Reordenando os X_k em ordem crescente obtém-se o vector das estatísticas ordinais (e.o.) $X_{k:n}$

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{k:n} \leq \cdots \leq X_{n:n}$$

onde genericamente $X_{k:n}$ é a k-ésima e.o. (ascendente), $X_{1:n}$ o mínimo e $X_{n:n}$ o máximo.

É fácil provar que

$$F_{X_{k:n}}(x) = \sum_{j=k}^{n} \binom{n}{j} [F_X(x)]^j [1 - F_X(x)]^{n-j}$$

e no caso de continuidade absoluta

$$f_{X_{k:n}}(x) = \frac{1}{B(k, n-k+1)}[F_X(x)]^{k-1}[1-F_X(x)]^{n-k} f_X(x)$$

(note-se que k é o rank *ascendente e $n-k+1$ é o* rank *descendente de $X_{k:n}$), ou seja, $X_{k:n}$ é uma "beta transformada".*

De facto, sendo F_X crescente, preserva as ordens, e transforma as e.o. de X em e.o. da uniforme que são betas $(k, n-k+1)$. Aquele resultado mostra que a cauda direita de uma binomial $(n, p = F_X(x))$ é equivalente à cauda esquerda de uma beta.

Já referimos o esquema de estabilidade em máximos. Não é necessário estabelecer resultados para mínimos, uma vez que

$$\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n).$$

O esquema de máximos foi estendido para e.o. de topo por Smirnov (1952).

Queremos ainda anotar que são as e.o. extremais que influenciam o comportamento aditivo limite. *De facto,*

$$\sum_{k=1}^{n} X_k = \sum_{k=1}^{n} X_{k:n}$$

e Watts (1980) e Watts, Rootzén and Leadbetter (1982) refrasearam muitas questões em termos de estatísticas ordinais extremais e estatísticas ordinais centrais.

Considere-se $X_{k(n):n}$, numa sucessão $\{X_k\}_{k \in IN}$ de v.a. i.i.d..

- *Se $\frac{k(n)}{n} \longrightarrow 0$, $X_{k(n):n}$ é uma e.o. de mínimos;*
- *Se $\frac{k(n)}{n} \longrightarrow 1$, $X_{k(n):n}$ é uma e.o. de máximos;*
- *Se $\frac{k(n)}{n} \longrightarrow q \in (0,1)$, é uma e.o. central (*central order statistic*).*

Por exemplo, $X_{n-10:n}$ é uma e.o. extremal, de máximos, o 2^o mínimo $X_{2:n}$ é uma e.o. extremal de mínimos, e o quantil $X_{\left[\frac{9n}{10}\right]:n}$ ou a mediana $X_{\left[\frac{n+1}{2}\right]:n}$ são e.o. centrais.

As e.o. centrais são aproximáveis pela Gaussiana, mas o mesmo não se passa com as e.o. extremais. Os domínios de atracção no esquema aditivo têm que ver com o equilíbrio de mínimos e máximos, e com os resultados sobre dominância de variação *regular estabelecidos por Tucker (1968). Para mais pormenores, cf. Chow e Teugels (1979). A referência clássica sobre e.o. é David (1981).*

O carácter markoviano das e.o. no caso de parente exponencial — que goza, como se sabe , de "falta de memória", por a cauda e^{-x} *verificar a equação funcional de Cauchy* $f(x+y) = f(x)f(y)$ *— leva a que os* spacings $S_1 = X_{1:n}$, $S_2 = X_{2:n} - X_{1:n}$, $\ldots$, $S_k = X_{k:n} - X_{k-1:n}$, $\ldots$, $S_n = X_{n:n} - X_{n-1:n}$, *sejam* independentes.

Tal como a falta de memória é característica da exponencial (nos modelos contínuos, propriedade que é partilhada pela geométrica no que refere a contrapartida discreta) esta independência dos spacings *caracteriza a exponencial, cf. Galambos (1987).*

Este resultado foi usado com sucesso por Pestana e Rocha (1995) para abordarem genericamente a studentização interna no modelo exponencial, com refinamento posterior de Brilhante (1996). Pestana e Rocha (1993) usaram esta propriedade para lançarem uma "análise de escala"paralela, no caso de parente exponencial, à clássica análise da variância de Fisher para o modelo Gaussiano — a qual assenta como é conhecido, na independência entre estimadores $\overline{X}_n$ *da localização e* S_n *da escala, que no caso de parente Gaussiana são independentes (situação também característica da Gaussiana — teorema de Darmois).*

Note-se que no contexto do modelo de Pareto generalizado, o caso $\alpha = 0$ *corresponde, por limite, à exponencial, enquanto para* $\alpha > 0$ *temos os tipos Pareto clássicos, e para* $\alpha < 0$ *temos tipos beta em* $(-1, 0)$. *Aquela propriedade dos* spacings *da exponencial aparece, neste contexto, como caso especial da independência dos quocientes de e.o. consecutivas* $\frac{X_{k:n}}{X_{k-1:n}}$ *nas Paretos.*

Isto animou-nos a estudar propriedades de spacings *em modelos mais gerais — nomeadamente em misturas de exponenciais, de que a Laplace* $f(x) = \frac{1}{2}e^{-|x|}I_{IR}$ *é apenas um caso especial, correspondente a simetrização.*

Note-se os resultados incompletos apresentados por Rocha (1995) no que refere o modelo Laplace, o que bem atesta a complexidade analítica que a simples simetrização traz ao modelo exponencial.

Provamos alguns resultados aritméticos sobre spacings *de misturas exponenciais, e em particular a divisibilidade infinita dos* spacings *da Laplace.*

Abordamos ainda o estudo de propriedades de quocientes de e.o. consecutivas.

1.5 Divisibilidade Infinita, Auto-decomponibilidade, Aritmética de Leis de Probabilidades, Semi-grupos Délficos, Convoluções Generalizadas

A construção das leis estáveis no esquema aditivo, que culmina com a publicação do Calcul des Probabilités *não satisfez Paul Lévy (1925).*

De facto, como todas as grandes criações intelectuais, a solução de cada problema gerou novas questões mais profundas, e a maior sede de conhecimento.

A observação de que cada variável aleatória estável pode ser escrita como soma de tantas parcelas do mesmo tipo (isto é, idênticas à soma, a menos de localização e escala) levou naturalmente à conclusão de que pode ser escrita como uma soma de parcelas infinitamente pequenas, e tal, para uma inteligência criativa como a de Lévy, levou naturalmente à introdução de integrais estocásticos, neste caso integrais "cujas parcelas são elementos aleatórios homogéneos"(Lévy, 1934).

Colocou-se então a questão, mais geral, de investigar a questão de quais as v.a. que podem ser decompostas em parcelas tão pequenas quanto se queira, que Lévy apelidou de v.a. com "leis infinitamente divisíveis". O instrumento privilegiado para o seu estudo é naturalmente a função característica (e como à aditividade de v.a. independentes corresponde o produto das respectivas f.c., a expressão "infinitamente divisíveis"aparece como natural).

Desde logo se observa que as v.a. estáveis são infinitamente divisíveis, mas que há v.a. infinitamente divisíveis que não são estáveis. De facto, todas as v.a. estáveis são absolutamente contínuas, e é imediato observar que a v.a. de Poisson é infinitamente divisível. Recorde-se que a sua f.c. é

$$\varphi_X(t) = e^{\lambda(e^{it}-1)}, \quad \lambda > 0$$

e naturalmente

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

A definição de v.a. infinitamente divisível (I.D.) é usualmente

— φ_X é uma f.c. I.D. sse $\forall n \in IN$, $\psi_n = \varphi_X^{1/n}$ é uma f.c. de uma v.a. Y_n

(corresponde a dizer que

$$X \stackrel{d}{=} Y_{n1} + Y_{n2} + \cdots + Y_{nn}$$

onde os Y_{nk} são réplicas independentes de Y_n).

Da definição acima é imediato que uma f.c. I.D. não pode ter zeros reais, e consequentemente

$$\varphi \text{ f.c. infinitamente divisível } \Leftrightarrow \varphi = e^{\Psi}.$$

(Ψ é por vezes referida como 2^a função característica)

O grande resultado sobre divisibilidade infinita deve-se a de Finetti (1930):

X é uma v.a. I.D. sse X for o limite fraco de uma sucessão de v.a. Poisson compostas.

Recordamos que uma v.a. Poison composta é uma soma aleatória de v.a. i.i.d., em que o número de parcelas é uma Poisson independente das parcelas, ou seja,

$$X = \sum_{k=1}^{N} Y_k$$

onde os Y_k são v.a. i.i.d., independentes de $N \frown P(\lambda)$. Prova-se que se $\varphi(t) = E[e^{itY_k}]$ for a f.c. de cada uma das parcelas, a f.c. da Poisson composta X é

$$\varphi_X(t) = e^{\lambda[\varphi(t)-1]}.$$

(É interessante observar que a f.c. da $N \frown P(\lambda)$ é $\varphi_N(t) = e^{\lambda(e^{it}-1)}$, e que e^{it} é a f.c. da v.a. degenerada em 1. Assim, $P(\lambda)$ é uma soma aleatória de 1's — a trajectória de um processo de Poisson é, de facto, constituída por acréscimos de 1's em instantes aleatórios.)

Um outro passo importante na teoria da divisibilidade infinita foi o reconhecimento de que as leis infinitamente divisíveis estão naturalmente associadas a limites de "esquemas *triangulares assintóticamente negligíveis"(*null triangular arrays*)*

$$\begin{array}{llll} X_{11} & & & \\ X_{21} & X_{22} & & \\ X_{31} & X_{32} & X_{33} & \\ \multicolumn{3}{l}{\dots\dots\dots\dots\dots\dots} & \\ X_{n1} & X_{n2} & \dots\dots & X_{nn} \\ \multicolumn{4}{l}{\dots\dots\dots\dots\dots\dots\dots\dots\dots} \end{array}$$

onde $\sup P\{|X_{nk}| > \varepsilon\} \to 0$. *Gnedenko e Kolmogoroff (1954) chamam a atenção para o facto de as leis I.D. surgirem, na sua generalidade, associadas a sequências duplamente indexadas.*

As abordagens modernas à divisibilidade infinita fazem-se hoje sobretudo neste esquema duplamente indexado, ou no contexto de Poissons compostas. A abordagem original de Lévy, em termos de integrais, é em geral reservada para estudos de processos estocásticos com incrementos i.i.d., de que o processo de Poisson é o protótipo simples.

A observação de que se φ é uma f.c. I.D., então $\varphi = e^{\Psi}$ veio a ter consequências teóricas de grande importância. Kolmogoroff apresentou uma representação integral para a 2^a função característica Ψ no caso de X ter variância finita

$$\Psi(t) = i\gamma t + \int_{-\infty}^{\infty} (e^{itx} - 1 - itx) \frac{dK(x)}{x^2}$$

onde $\gamma \in IR$ e K é uma função limitada não decrescente tal que $K(-\infty) = 0$.

Por sua vez, Lévy dedicou alguns anos a procurar representações integrais genéricas; a representação integral de Lévy para a 2^a função característica é da forma

$$\Psi(t) = i\gamma t - ct^2 + \int_{-\infty}^{0^-} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) dM(u) + \int_{0^+}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) dN(u) \quad (1.5.1)$$

onde γ é uma localização, $c \geq 0$ é uma escala ($c = 0$ se X não tiver variância finita) e M e N — medida espectral de Lévy — que verificam:

(i) M e N são funções não decrescentes nos intervalos $(-\infty, 0)$ e $(0, \infty)$, respectivamente;

(ii) $M(-\infty) = N(+\infty) = 0$*;*

(iii) $\int_{\epsilon^-}^0 u^2 dM(u) < \infty$ *e* $\int_0^\epsilon u^2 dN(u) < \infty$ *para todo o* $\epsilon > 0$.

É desta expressão geral que se infere a representação das f.c. estáveis atrás apresentada.

Note-se que $e^{i\gamma t}$ *corresponde a uma degenerada,* e^{-ct^2} *corresponde a uma componente Gaussiana, e* $\int_{-\infty}^{0^-}\big(e^{itu} - 1 - \frac{itu}{1+u^2}\big)dM(u)$ *e* $\int_{0^+}^{\infty}\big(e^{itu} - 1 - \frac{itu}{1+u^2}\big)dN(u)$ *correspondem a componentes Poissonianas (com suporte reticulado não necessariamente inteiro) no semi-eixos negativo e positivo, respectivamente.*

A representação acima foi posteriormente modificada, e a representação integral hoje mais usada é a de Lévy-Khinchine

$$\Psi(t) = i\gamma t + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\frac{1+x^2}{x^2}d\theta(x) \tag{1.5.2}$$

onde $\gamma \in IR$ *e* θ *é uma função limitada não decrescente tal que* $\theta(-\infty) = 0$*. A função integranda está definida para* $x = 0$ *por continuidade sendo igual a* $-\frac{t^2}{2}$*. O "factor de alisamento"* $\frac{1+x^2}{x^2}$ *força a convergência do integral — cf. por exemplo a abordagem heurística em Burrill (1972), reveladora do que está em questão.*

Em 1963 Kendall abordou o estudo das transformadas de Laplace das v.a. I.D. com suporte positivo, mostrando que a representação integral é uma representação de Choquet. Recordamos que Minkovski tinha observado que qualquer ponto de um poliedro convexo é "baricentro"dos seus vértices, e que Krein e Milman (1940) mostraram, com generalidade, que em situações de regularidade qualquer ponto de um conjunto convexo pode ser escrito como um integral estendido ao conjunto dos seus "pontos extremos". Um ponto extremo é um ponto que pode ser retirado ao conjunto convexo sem que ele deixe de ser convexo. Exemplifiquemos:

Num segmento $[a, b]$*, apenas os pontos* a *e* b *são extremos. Qualquer ponto* $x \in [a, b]$ *pode ser escrito como "combinação convexa"*

$$x = \lambda a + (1-\lambda)b, \qquad 0 \le \lambda \le 1$$

ou por outras palavras

$$x = E[X]$$

onde $X = \begin{cases} a & b \\ \lambda & 1-\lambda \end{cases}$.

Num triangulo $[ABC]$*,* A*,* B *e* C *são os pontos extremos. Qualquer* x *ponto do triângulo pode ser escrito*

$$x = \lambda_1 A + \lambda_2 B + \lambda_3 C$$

ou $x = E[X]$*, com* $X = \begin{cases} A & B & C \\ \lambda_1 & \lambda_2 & \lambda_3 \end{cases}$ *, com* $0 \leq \lambda_k \leq 1$ *e* $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

A grande criação de Choquet (1960), na sequência do seu estudo sobre cargas e capacidades, uma generalização não trivial de probabilidades, foi a descoberta de que se $\mathcal{C}$ *for um conjunto convexo e compacto, com um conjunto de pontos extremos compacto, num espaço metrizável, então cada* $x \in \mathcal{C}$ *é o valor médio de uma medida de probabilidade suportada pelos pontos extremos de* $\mathcal{C}$*. (Posteriormente, estendeu a representação a cones convexos, com base nas geratrizes estremais.).*

Esta abordagem de Kendall (1963) revolucionou o pensamento sobre divisibilidade infinita, uma vez que permitiu uma compreensão estrutural da arquitectura das leis I.D.. Johansen (1966) estabeleceu que a 2^a *f.c. das I.D. é, tal como a própria f.c., uma função definida positiva, e com base nisso mostrou que os pontos extremos do conjunto convexo das f.c. I.D. são*

(i) f.c.'s degeneradas em $-\infty$ *e* $+\infty$

(ii) f.c. Gaussiana

(iii) f.c.'s Poissonianas generalizadas (em retículas não naturais)

(note-se que Johansen admitiu v.a. desonestas, com átomos em $-\infty$ *e/ou* $+\infty$*, a fim de conseguir a compactificação do conjunto dos pontos extremos), mostrando assim que a representação integral de Lévy (e a representação integral de Lévy-Khinchine) é uma representação integral de Choquet.*

Em certo sentido, toda esta questão pode parecer académica. De facto, para que uma soma convirja as suas parcelas têm que ser evanescente (donde o requisito $\sup P\{|X_{nk}| > \varepsilon\} \to 0$*, na abordagem com esquemas triangulares). Por outro lado, se uma v.a.* $X_{nk} \sim 0$*, a sua f.c.* $\varphi_{X_{nk}}(t) \sim 1$*, visto que* $\varphi(t) \equiv 1$ *é a f.c. da degenerada em 0, e sabemos que as f.c. são uniformemente contínuas. Também, para que um produto convirja, os factores têm que ser aproximadamente 1.*

Assim

$$\varphi_X(t) = \prod_k \varphi_{X_k}(t) \sim \exp\left\{\sum_k \left[\varphi_{X_k}(t) - 1\right]\right\}$$

se $\varphi_{X_k}(t) \sim 1$*. Neste sentido, as célebres representações integrais que tanta tinta fizeram correr são, em certo sentido, uma trivialidade.*

Mas por outro lado esta investigação suscitou a questão da "aritmética das leis da probabilidade". Genericamente, colocou-se a questão de saber quando é que X *é decomponível de forma não trivial, isto é,*

$$X \stackrel{d}{=} Y_1 + Y_2$$

com Y_1 *e* Y_2 *independentes, não degeneradas.*

Por exemplo, uma v.a. X *discreta com suporte em dois pontos é* indecomponível.

Uma questão mais específica ainda, com interesse na aritmética das leis da probabilidade, é investigar em que situações não há decomposições $X \stackrel{d}{=} Y_1 + Y_2$ *com* Y_1 *e* Y_2 *v.a. i.i.d.. Por exemplo, se* X *é uniforme em* $[0, 1]$*, não admite decomposição como soma de v.a. i.i.d. — mas por outro lado,* $X \frown U[0, 1]$ *admite a decomposição*

$$X \stackrel{d}{=} \sum_{k=0}^{\infty} \frac{Y_k}{2^k}$$

como os $Y_k \frown B\left(\frac{1}{2}\right)$ *independentes, ou seja,* $Y_k = \begin{cases} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{cases}$ *(basta pensar na expansão de* $x \in [0, 1]$ *na base 2), o que mostra que uma v.a. absolutamente contínua pode surgir como soma de um número infinito de discretas.*

As questões aritméticas no semi-grupo aditivo das v.a. estão cheias de patologias, tendo mesmo Lévy conseguido construir o exemplo de uma v.a. I.D. que pode ser decomposta como soma de um número finito de v.a. indecomponíveis!

Khinchine, e posteriormente Ostrovski e Ibragimov, deram-se conta da enorme importância da subclasse I_0 *da classe das I.D.;* I_0 *é a subclasse das leis I.D. que só admitem factores infinitamente divisíveis. Elementos célebres de* I_0 *são a Gaussiana, que só admite factores Gaussianos (teorema de Lévy-Cramér) e a Poisson, que só admite factores Poisson (teorema de Raikov). Note-se que estes resultados, que a seu tempo foram es-*

tabelecidos com um armorial analítico complexo[1] *são, à luz da identificação dos pontos extremos do conjunto de f.c.'s I.D. feita por Johansen (1966), absolutamente triviais.*

A definição da classe I_0 permitiu a Khinchine estabelecer o teorema fundamental da aritmética das leis de probabilidades:

— Qualquer f.c. φ pode ser escrita na forma

$$\varphi(t) = \psi(t) \prod_k \Theta_k(t)$$

onde $\psi \in I_0$, e os Θ_k são f.c. indecomponíveis.

Este teorema paraleliza o teorema fundamental da aritmética, de acordo com o qual qualquer $n \in I\!N$ pode ser escrito

$$n = 1 \times \prod_k p_k$$

onde os p_k são indecomponíveis ou primos. Note-se que 1 é o único elemento I.D. em $I\!N$.

Para mais detalhes, consultar o aliciante prefácio de Kendall e Harding (1973) à sua compilação Stochastic Analysis.

Kendall (1967) identificou os elementos algébricos fundamentais da aritmética das leis infinitamente divisíveis, e a partir daí definiu a estrutura de semi-grupos topológicos genéricos em que é possível estudar questões aritméticas, a que chamou semi-grupos délficos *(este nome sibilino deve-se a tê-los apresentado pela 1^a vez num congresso que decorria nas imediações de Delfos). O seu discípulo Davidson (1968) investigou com grande generalidade questões aritméticas nestes semi-grupos. Pestana (1984) estudou a estrutura délfica do semi-grupo das f.c. de Pólya, identificando a classe I_0 neste semi-grupo.*

Outra abordagem a questões do mesmo tipo surgiu da investigação de Kingman (1963) sobre passeios aleatórios em esferas, que o levou a definir convoluções generalizadas; seu discípulo Bingham (1971), como já atrás referimos, investigou as questões de estabilidade e domínios de atracção em convoluções generalizadas.

[1] O teorema de Lévy-Cramér é tão complexo, na abordagem clássica, que Lévy (1937) assentou nele os 3 últimos capítulos da sua *Théorie de L'Addition des Variables Aléatoires* sem o conseguir demonstrar. Só volvidos 12 anos Cramér o demonstrou.

Urbanik (1973) tem desenvolvido questões similares usando "funcionais característicos", e na sequência dos livros de Parthasarathy (1967) e de Heyer (1977) multiplicaram-se as investigações sobre divisibilidade infinita e decomponibilidade em espaços de Hilbert, espaços de Banach, etc. Em particular, Masani (1970) identificou as leis infinitamente divisíveis com hélices no espaço de Hilbert.

Outro rumo que as investigações têm tomado relaciona-se com a observação, feita por Zolotarev, de que num esquema triangular pode haver convergência sem que se exija negligibilidade assintótica. Esta observação originou um novo capítulo de "teoremas limite não clássicos", que nos últimos anos não tem conseguido grandes progressos. Consulte-se o capítulo escrito por Osipov no tratado de Linnik e Ostrovski (1977).

Refira-se ainda uma questão originada nos anos 30, e que recentemente se mostra de novo campo promissor de pesquisa.

Diz respeito a desenvolvimentos recentes da classe $\mathcal{L}$ ou de auto-decomponíveis, a classe de leis limites de somas normadas $\sum_k \frac{X_k - B_k}{A_k}$, $B_k \in I\!R$ e $A_k > 0$, quando as parcelas X_k são independentes mas não necessariamente identicamente distribuídas. Note-se que é necessário impor algumas condições de compacticidade do tipo $\frac{A_{k+1}}{A_k} \xrightarrow[k\to\infty]{} 1$, e que Kruglov (1972) e Sreehari (1970) investigaram situações em que $\frac{A_{k+1}}{A_k} \xrightarrow[k\to\infty]{} r$, respectivamente na situação $r > 1$ e $r \in (0,1)$.

Dizemos que a v.a. X é auto-decomponível*, ou da classe $\mathcal{L}$ de Khinchine, sse*

$$X \stackrel{d}{=} aX + Y_a$$

com $a \in (0,1)$, e X e Y_a independentes.

Prova-se que Y_a é infinitamente divisível, e que se X for auto-decomponível é também I.D., e admite a representação canónica

$$\varphi_a(t) = \exp\left\{ it\gamma_1 - c(1-a^2)t^2 + \int_{-\infty}^{0^-} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) d\left[M(u) - M\left(\frac{u}{a}\right)\right] \right.$$
$$\left. + \int_{0^+}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) d\left[N(u) - N\left(\frac{u}{a}\right)\right] \right\}$$

onde γ_1, M e N obedecem às mesmas condições de (1.5.1)*, e com $M(u) - M\left(\frac{u}{a}\right)$ e $N(u) - N\left(\frac{u}{a}\right)$ funções não decrescentes.*

Em 1973 Urbanik (que tem vindo a produzir um conjunto de resultados de grande profundidade usando técnicas de representações integrais suportadas por pontos extremos) importou as ideias de variação lenta para a definição de "esquemas triangulares com variação lenta" e que com base nesse esquema refinou a classe $\mathcal{L}$ de Khinchine.

Resumindo muito concisamente os seus resultados

1. *Renomeou $\mathcal{L}_1$, a classe $\mathcal{L}$ de Khinchine, e chamou $\mathcal{L}_0$ à classe das I.D.*

2. *Com arranjos triangulares baseados em $\mathcal{L}_1$, definiu $\mathcal{L}_2$, a classe das v.a. que admitem a decomposição*
$$X \stackrel{d}{=} aX + Y_a$$
com $a \in (0,1)$, X e Y_a independentes, provando que $Y_a \in \mathcal{L}_1$.

3. *Iterativamente, construíu as classes $\mathcal{L}_k$, $k = 1, 2, \ldots$, mostrando que $X \in \mathcal{L}_k$ sse*
$$X \stackrel{d}{=} aX + Y_a$$
com $a \in (0,1)$, X e Y_a independentes, o que arrasta $Y_a \in \mathcal{L}_{k-1}$.

4. *$\mathcal{L}_\infty = \bigcap_{k=0}^{\infty} \mathcal{L}_k$ é a classe das leis que aparecem como somas de v.a. independentes, com parcelas do tipo de um número finito ou numerável de tipos. Assim, $\mathcal{L}_\infty$ surge como a classe "natural" para modelar fenómenos aditivos, em que se admite independência das parcelas, mas nada se exige sobre identidade distribucional.*

1.6 Leis Aditivas e Valores Extremos

Dizia Camilo Castelo Branco que o caminho mais curto entre dois pontos não é uma recta, é a sinuosa curva da imaginação.

Também a invenção científica não segue uma linha recta — ou pelo menos esta quebra-se inesperadamente, com aparente prejuízo para o desenvolvimento harmonioso da Ciência.

Os estudos iniciais de Lévy levaram-no, entre 1917 e 1925, a desenvolver a teoria da estabilidade para somas. O seu discípulo Fréchet (1927) introduziu a equação da estabilidade para máximos, e com isso permitiu o desenvolvimento inicial da teoria dos valores extremos por Fisher e Tippett (1928).

Nos anos 30, Lévy e Khinchine desenvolveram a teoria das infinitamente divisíveis, e Doeblin (1940), um discípulo de Lévy, caracterizou os domínios de atracção para somas. Independentemente dele, num mundo conturbado pela guerra, um discípulo de Khinchine, B.V. Gnedenko (1940) publicou resultados semelhantes, e em 1943 publicou os alicerces da teoria dos domínios de atracção para as leis estáveis para máximos (o tratamento rigoroso do domínio de atracção da Gumbel apenas foi feito por de Haan (1970), volvidos alguns anos, e aproveitando a chamada de atenção de Feller (1967) para a teoria das funções de variação lenta).

Entretanto Khinchine tinha introduzido a classe $\mathcal{L}$, das auto-decomponíveis, e um discípulo do seu discípulo Gnedenko, D. Mejzler (1956) estudou a auto-decomponibilidade para máximos, introduzindo a classe das leis limite de máximos de v.a. independentes mas não necessariamente identicamente distribuídas. Este trabalho pioneiro, escrito em ucraniano, não teve descendência directa durante muitos anos até Pestana e Graça Martins (1985) o tomarem como base de refinamento da classe que, em honra de Mejzler, denotaram M_1 — refinamentos esses que levaram à caracterização da classe M_∞, símile para o esquema de máximos da classe $\mathcal{L}_\infty$ de Urbanik no esquema de somas.

Até os anos 80 as duas teorias divergiram, apesar de terem uma evidente base algébrica comum.

A que se deve este divórcio?

Pensamos que assenta no facto de a teoria aditiva ter tido como fulcro de desenvolvimento a noção de divisibilidade infinita de Lévy.

Ora à soma de v.a independentes corresponde o produto das f.c., e a noção de divisibilidade infinita exprime-se mais concisamente em termos destas transformadas integrais. De facto, a definição mais usual de divisibilidade infinita é:

— A f.c. φ é I.D. sse $\forall n \in IN$ existir uma f.c. ψ_n tal que $\varphi^{1/n} = \psi_n$.

Esta asserção não é trivial, pois as f.c. são as funções definidas positivas com $\varphi(0) = 1$, e nada obriga uma raíz de uma função definida positiva a ser definida positiva.

No que respeita máximos, sabemos que $F_{X_{n:n}}(x) = [F_X(x)]^n$, e consequentemente o papel desempenhado pelas f.c. no que refere somas é desempenhado por funções de distribuição no que refere máximos.

Ora uma f.d. é uma função não decrescente e contínua à direita, com $F_X(-\infty) = 0$ *e* $F_X(+\infty) = 1$, *(se a v.a.* X *for honesta) — e a raíz de índice* n *de uma tal função contínua a ter essas propriedades. Por outras palavras, qualquer f.d. (univariada) é trivialmente uma f.d. I.D.!*

Mas o que é trivial não excita a imaginação, e não suscita interesse.

Só quando Balkema e Resnick (1977) observaram que em IR^n, *com* $n \geq 2$, *tal já não acontece, é que o conceito de* ”max-infinite divisibility”começou a atrair as atenções. Mostraram aqueles autores, em particular, que as f.d. max-I.D. são as que aparecem naturalmente associadas a processos estocásticos extremais.

O fosso que entretanto se tinha criado entre o esquema aditivo e o esquema de máximos está longe de ter sido preenchido.

De facto, enquanto no que refere somas o interesse dos investigadores se centrou sobretudo em relaxar as hipóteses de identidade distribucional (problema que Gnedenko, com resultados parciais de Zolotarev e Koroljuk (1961), e os refinamentos à classe $\mathcal{L}$ feitos por Urbanik (1973) e a extensões multivariadas (Cuppens (1975); Giné (1980)) ou em espaços algébricos abstractos (Parthasarathy (1967), Heyer (1977), Jurek (1981)), no caso da teoria de valores extremos, que em geral surgem em *clusters* dependentes, o esforço principal foi feito no sentido de estabelecer resultados em situações de dependência fraca, e na extensão ao estudo das estatísticas ordinais de topo. A teoria multivariada tem apostado sobretudo no esquema de tomar limites componente a componente, usando a ”função de cópula”de Deheuvels (1978), ou no conceito de concomitantes de estatísticas ordinais.

Para detalhes consultar Galambos (1987), Leadbetter, Lindgren e Rootzén (1982) e Resnick (1987).

A presente dissertação pretende contribuir para preencher um pouco o fosso entre as duas teorias, na sequência dos trabalhos de Pestana (1978), Gomes e Pestana (1985), Graça Martins e Pestana (1987), Iglésias Pereira, Oliveira e Pestana (1996).

1.7 Eixos do Desenvolvimento da Estatística — uma segunda perspectiva

A Estatística é multifacetada; entre os seus aspectos mais salientes há o que tem como objectivo principal a descrição da realidade (e isto não se limita à clássica Estatística Descritiva, engloba os aspectos distribucionais e estruturais da Estatística), e o que tem como objectivo principal a predição e o controlo.

Naturalmente, nos alicerces probabilísticos da Estatística, encontramos como base daqueles aspectos a noção de independência, e como ponto de partida do segundo as ideias de condicionamento.

A expressão geral de probabilidade de uma cadeia

$$P\left(\bigcap_{k=1}^{n} A_n\right) = P(A_1)P(A_2|A_1)\cdots P\left(A_{n-1}\middle|\bigcap_{k=1}^{n-2} A_k\right)P\left(A_n\middle|\bigcap_{k=1}^{n-1} A_k\right)$$

simplifica-se de facto singularmente com a definição de independência, passando a ter o aspecto

$$P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P(A_k)$$

— é por isso que, caso controlemos a situação, usamos amostragem, sendo pertinente na definição clássica de amostra a exigência de independência entre as observações.

Já nos estudos mais sofisticados em que procuramos relacionar realidades diversas — por exemplo estudos regressivos — é a expressão geral de condicionamento, ou aspectos particulares mais simples mas que todavia não dispensam totalmente a dependência, que predominam.

Também no que respeita a recolha de informação, os nossos propósitos guiam os nossos procedimentos: estudos meramente observacionais, em que se procura observar sem alterar a realidade que se observa, os estudos experimentais, em que alteramos a realidade e observamos a juzante os efeitos dessa alteração.

Sendo as fronteiras ténues, distinguimos também entre *amostragem*, que é sobretudo *"design based", e o* planeamento de experiências*, que é mais* "model based".

Numa perspectiva actual tudo isto parece natural, mas no desenvolvimento efectivo da Estatística o progresso foi, naturalmente descontínuo, impulsionado por saltos de gigante transpostos por grandes génios.

E, naturalmente, nos desenvolvimentos iniciais, esteve o modelo "normal"ou Gaussiano, o que pragmaticamente era usável na primeira metade deste século.

Há excelentes razões para adorar o modelo Gaussiano:

1. É, privilegiadamente, o modelo de localização-escala.

2. Qualquer Gaussiana é, por estandartização, transformável na $Gau(0,1)$.

3. A localização da Gaussiana é estimada por $\overline{X}_n$. No caso de $X \frown Gau(\mu, \sigma)$, $\overline{X}_n \frown Gau\big(\mu, \frac{\sigma}{\sqrt{n}}\big)$ (o que é bonito mas não serve para nada, uma vez que o que interessa é estimar os parâmetros). Uma vez que a estandartização de $\overline{X}_n$ está em geral para além do que é acessível, usamos antes a *studentização*

$$\frac{\overline{X}_n - \mu}{S/\sqrt{n}} \frown t_{(n-1)}$$

onde $S^2 = \frac{1}{n-1} \sum\limits_{k=1}^{n} (X_k - \overline{X}_n)^2$ é a estatística que estima o desvio padrão σ.

4. De facto, $\frac{(n-1)S^2}{\sigma^2} \frown \chi^2_{(n-1)}$, e na observação acima sobre studentização está contido o que distribucionalmente se sabe sobre o quociente entre duas v.a. independentes, a saber a divisão de uma $Gau(0, 1)$ pela raíz quadrada de um qui-quadrado dividido pelo seu valor esperado (ou seja, o número de graus de liberdade (g.l.)).

5. A teoria distribucional dos momentos amostrais em universos Gaussianos permite-nos ainda cotejar duas estimativas independentes da variância:
Se tomarmos o quociente entre os ”quadrados médios”, que correspondem a dividir cada soma de quadrados $\sum\limits_{j=1}^{n_k} (X_{kj} - \overline{X}_k)^2$ pelo respectivo número de g.l., obtemos uma v.a. com distribuição F de Fisher-Snedecor.
Para além dos aspectos distribucionais acima referidos, outros factos há que tornam o modelo Gaussiano aliciante, quase irresistível.

6. O teorema limite central permite aproximar por uma Gaussiana apropriada (em geral apenas há que equacionar valor esperado e variância do modelo”exacto”e da aproximação Gaussiana) a maior parte dos modelos usuais (por exemplo, binomial, Poisson, geométrica e outras binomiais negativas, exponencial e qui-quadrado e outras gamas, Pareto, Gumbel, etc.)

7. Na multiGaussiana, a regressão de uma margem univariada noutra é *linear*.

Nos atractivos da Gaussiana acima descritos, referimos aspectos das distribuições amostrais que têm grande relevo: studentização e estimação de variância por somas de quadrados independentes (a base na análise da variância), que são base da inferência em populações Gaussianas.

Studentização e análise da variância derivam, uma e outra, de uma situação privilegiada da amostragem Gaussiana: $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ e $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2$ são independentes, propriedade que é característica da Gaussiana (teorema de Darmois).

A studentização foi uma ideia seminal de Gosset, publicada na *Biometrika* em 1908 sob o pseudónimo Student. Gosset apercebeu-se que quando queremos estandartizar, mas σ desconhecido é estimado or S, esta constante deve ser encarada como valor observado da v.a. S; consequentemente, desenvolveu a teoria amostral correspondente ao quociente entre Gaussiana e raíz quadrada de um qui-quadrado (dividido este pelo número de g.l., por ser mais conveniente).

Quando o advento de novos meios computacionais abriu caminho à utilização de modelos menos cómodos mas mais flexíveis, uma das dificuldades a enfrentar foi o de dependência entre estimadores da localização e escala nesses modelos mais manejáveis.

Por isso surgiu o conceito de studentização externa e de studentização interna (cf. David (1981)), passando a chamar-se studentização, em acepção geral, à divisão de um estimador de localização que depende de um parâmetro perturbador de escala por um estimador desse parâmetro de escala cuja f.d. não depende do parâmetro de localização. A studentização interna, em que os estimadores são dependentes, é obviamente de tratamento analítico difícil, e o problema, na sua generalidade, está londe de estar resolvido.

Pestana e Rocha (1992) abordaram a questão com grande generalidade, dando expressões aproximadas muito gerais em populações simétricas, e estudando com inversão de transformadas integrais as populações beta e gama. Rocha (1995) apresenta com detalhe uma abordagem muito geral a populações simétricas, obtendo resultados que admitem comparação com a integração numérica de Gauss, no sentido em que os ”pontos interpoladores”que usa o valor médio generalizado são totalmente independentes da população parente.

Brilhante (1996) afinou os resultados em populações exponencias, obtendo resultados exactos e assintóticos, quer à custa de funções especiais quer por inversão de transformadas de Laplace, fornecendo tabelas que tornam esses resultados efectivamente usáveis em inferência estatística. Diamantino e Pestana (1996) abordaram estudos de robustez, usando uma família para-Gaussiana, e pares de estimadores biponderados de localização/escala.

O passo gigantesco que se seguiu em inferência estatística foi a invenção da análise da variância e o decorrente planeamento de experiências, por Fisher (1925). Em certo sentido, foi um "*big bang"cuja repercussão continua a fazer-se sentir.*

A studentização, na sua forma original, tinha objectivos práticos inferir (estimação intervalar, testes de hipóteses) sobre o parâmetro de localização de uma população Gaussiana em que a média e variância eram desconhecidas, e também comparar as médias de duas populações Gaussianas com a mesma variância, ainda que desconhecida (e como a Gaussiana é um modelo completamente especificado pela média e a variância, uma vez que se partia de homocedasticidade estávamos, afinal, perante um estudo de homogeneidade de populações).

A questão que se punha como passo seguinte era, obviamente, a comparação de $k > 2$ *médias de Gaussianas, sob hipótese de homocedasticidade.*

A solução que parecia óbvia consistia em perfazer as $\binom{k}{2}$ *comparações possíveis. Fisher apercebeu-se que este procedimento é falacioso — se por exemplo estivermos a decidir admitindo* $\alpha = 5\%$ *de erro de* 1^a *espécie – em termos médios, em cada 20 testes rejeitar uma vez um hipótese nula que afinal de contas era verdadeira — no caso de* $k = 10$*, por exemplo, é praticamente certo cometer um erro de* 1^a *espécie numa das 45 comparações de pares!*

Esta observação foi importante não só para a evolução da Estatística como para a própria Filosofia da Ciência. Tal como o teorema da probabilidade total era a importação do "método"de Descartes para a Matemática, a posição crítica de Fisher é porventura a primeira objecção forte ao "erro de Descartes".

Isto levou Fisher (1925) a retomar alguns conceitos fundamentais. Nomeadamente

A— observar que a variância é influenciada pelo valor da média, no sentido preciso em que

$$f(A) = \int_{IR} (x - A)^2 dF(x)$$

é minimizada por $A = \int_{IR} x dF(x) = E(X)$*. Consequentemente, se tivermos várias "imagens"amostrais da média,* $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_n$*, analisando a variância através dos vários estimadores* $(n_k - 1)S_k^2 = \sum_{j=1}^{n_k} (X_{kj} - \overline{X}_k)^2$*, poderemos inferir se são todas elas perturbações da mesma média global.*

B— observar que a variância é um momento quadrático, e que a função $f(x) = x^2$ *é a* *única função não trivial tal que*

$$f\left(\frac{a+b}{2}\right)+f\left(\frac{a-b}{2}\right)=f(a)+f(b)$$

(ou seja, é possível eliminar o termo "rectangular").

Com base nestas duas ideias, pôde Fisher decompor a soma de quadrados

$$(n-1)S^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{\overline{X}}\right)^2$$

(onde $n=\sum_{i=1}^{k} n_i$, e $\overline{\overline{X}}$ é a média da "amostra combinada"ou amostra global), subtraindo e somando $\overline{X}_i$, desenvolvendo os quadrados e eliminando o termo rectangular:

$$\begin{aligned}(n-1)S^2&=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[\left(X_{ij}-\overline{X}_i\right)+\left(\overline{X}_i-\overline{\overline{X}}\right)\right]^2\\&=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2+\sum_{i=1}^{k}n_i\left(\overline{X}_i-\overline{\overline{X}}\right)^2.\end{aligned}$$

Por outras palavras, o estimador global da variância, com $\sum_{i=1}^{k} n_i-1$ graus de liberdade, foi decomposto em duas parcelas independentes, $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2$ *com $\sum_{i=1}^{k}(n_i-1)$ g.l. que estima a variância* dentro *das amostras, e* $\sum_{i=1}^{k}n_i\left(\overline{X}_i-\overline{\overline{X}}\right)^2$ *com $(k-1)$ g.l., que estima a variância* entre *as amostras.*

No caso de homogeneidade de localização (igualdade das médias) os "quadrados médios", que se obtêm dividindo as somas de quadrados pelo respectivo número de g.l., estão a estimar a mesma variância, pelo que o valor esperado do quociente é 1. A "perturbação"que sob validade da hipótese nula pode ocorrer é medida pela distribuição amostral daquele quociente de quis-quadrados independentes divididos cada qual pelo seu valor esperado — por outras palavras, a distribuição F de Fisher-Snedecor com os números de graus de liberdade apropriados.

Esta é, porventura, a melhor razão para termos um grande apego ao modelo Gaussiano. De facto, a teoria acima desenvolvida apenas se aplica ao modelo Gaussiano, pois apenas nesta decomposição é em estimadores independentes, com distribuições de qui-quadrado.

As tentativas de estender a análise de variância a outros modelos (por exemplo Barão (1992), em modelos extremais e usando com instrumento a razão de verosimilhanças) têm levado à introdução de diversas distribuições "pseudo-F". Em geral os resultados obtidos são assintóticos, o que de certo modo os torna irrelevantes: de facto, a grande importância da studentização e da análise da variância reside na possibilidade de perfazer inferência estatística com pequenas amostras!

Pestana e Rocha (1993), e posteriormente Rocha (1995) retomaram a filosofia inicial de Fisher, e foram procurar outras situações em que:

(i) houvesse um modelo de localização e escala, em que o estimador da escala fosse influenciado pela estimação da localização.

(ii) os estimadores fossem "regulares-– por exemplo, estimadores de máxima verosimilhança.

(iii) o estimador global da escala pudesse ser particionado em estimadores independentes *dentro e entre amostras.*

Uma situação simples em que tal acontece é o modelo exponencial com localização e escala, em que o estimador da localização é $X_{1:n}$ *e o estimador da escala é* $\frac{1}{n}\sum\limits_{k=1}^{n}(X_k - X_{1:n}) = \frac{1}{n}\sum\limits_{k=1}^{n}(X_{k:n} - X_{1:n})$*; de facto, é característico da exponencial a independência entre* $X_{1:n}$ *e os "*spacings*"* $X_{k:n} - X_{k-1:n}$.

Um dos objectivos que nos propusémos foi explorar outras situações não triviais em que o recurso a funções estatísticas ordinais nos levasse a encontrar as circunstâncias (i), (ii) e (iii) de Pestana e Rocha (1993).

Um caminho promissor é naturalmente a família das Pareto generalizadas, de que a exponencial é o caso limite (índice $\alpha = 0$*). De facto, a independência dos* spacings *da exponencial é apenas a "tradução"para aquele caso limite da independência entre quocientes de pares de e.o. consecutivas nas Pareto com índice* $\alpha \neq 0$.

1.8 Estatísticas Ordinais, Robustez e Resistência — uma terceira perspectiva sobre a história da Estatística

As especificidades do modelo Gaussiano, que o tornam matematicamente tratável, levaram — enquanto os recursos computacionais a isso obrigaram — à aceitação das ideias de Winsor durante cerca de meio século.

De facto, assim que se sai do modelo Gaussiano os desenvolvimentos analíticos exibem dificuldades quase inultrapassáveis.

Por isso, nos anos 50 a Estatística Aplicada balançava entre dois pólos: ou se assumia um modelo Gaussiano e se adoptava o "modelo linear"(recordamos que na multiGaussiana a regressão de uma margem noutra , ou na combinação linear de outras, é linear), ou se assumia apenas que as observações provinham de uma população contínua com distribuição F_X não especificada.

Esta aborgdagem, que recorre ou a contagens ou a argumentos combinatórios usando ranks *em lugar de magnitudes, chama-se* estatística não paramétrica.

O advento de novos meios computacionais veio revolucionar tudo (inclusive deslocar muitos métodos de estatística não paramétrica para a área de estatísticas ordinais ponderadas).

Com efeito, o modelo Gaussiano é extremamente concentrado em torno da média, na sua escala intrínseca — sabemos mesmo (cf. Graça Martins (1983)) que é a lei I.D. com caudas mais leves — pelo que em amostras Gaussianas não é de esperar observar outliers.

Ora o teste de Iglewicz (1981) sobre forma Gaussiana leva à conclusão que as amostras reais não são, em geral, Gaussianas.

Por outro lado, um célebre exemplo de Tukey (cf. Hoaglin, Mosteller e Tukey, 1992) mostra que os tradicionais estimadores de localização e escala — o par (média, desvio padrão empírico) — são muito sensíveis à presença de outliers*; o conceito de limite de ruptura esclarece bem a situação.*

Neste contexto, a observação de que a média minimiza o momento de 2^a *ordem pode levar-nos a considerar outras funções objectivo a optimizar — e os estimadores biponderados de Tukey, ou os estimadores de onda de Andrews têm-se revelado alternativas saudáveis.*

Tudo isto tem recolocado na linha da frente da inferência estatística o uso de estatísticas ordinais centrais e extremais, eventualmente ponderados, por forma a propiciar *métodos robustos e/ou resistentes, que acomodem erros de observação e/ou caudas mais pesadas do que o esperado.*

Por outro lado, estatísticas envolvendo e.o. extremais e centrais, como por exemplo $\frac{X_{n:n}-M}{M-X_{1:n}}$ *foram usadas com sucesso (cf. Gomes e Tiago de Oliveira (1984)) como decisores sobre o modelo apropriado.*

Uma vez que $X_{1:n}$ *e* $X_{n:n}$ *podem ocasionar problemas de robustez e/ou resistência, já no nosso estudo sobre studentização na exponencial (Brilhante, 1996) usámos estatísticas ordinais intermédias — situadas sensivelmente a* $\frac{1}{3}$ *dos extremos.*

Na presente dissertação, exploramos quocientes de spacings *generalizados como instrumento de decisão estatística.*

Capítulo 2

Inferência Estatística em Modelos Pareto Generalizados

Em populações exponenciais a independência dos spacings *constitui frequentemente um factor determinante na obtenção da distribuição de estatísticas studentizadas que propiciam inferência sobre o parâmetro de localização. As sedutoras propriedades do modelo exponencial fazem-no provavelmente um dos mais investigados actualmente em Estatística — veja-se por exemplo o volume editado por N. Balakrishnan e A. Basu (1995) sobre a distribuição exponencial com 33 capítulos e que constitui uma prova irrefutável da aplicabilidade estatística da exponencial, sobretudo na formulação de modelos de tempo de vida e em processos estocásticos em geral.*

Daí, ao explorarmos outros modelos, eventualmente com um parâmetro de forma, pode ser vantajoso reportarmos ao modelo exponencial. Esse reporte normalmente é feito sem grandes sobressaltos analíticos se o modelo exponencial estiver directamente relacionado com o modelo a investigar, como por exemplo nos casos em que a exponencial surge como caso particular.

Assim, neste capítulo focaremos essencialmente a nossa atenção na distribuição de Pareto generalizada, onde a distribuição exponencial surge como caso limite (índice $\alpha = 0$*). Tendo como ponto de partida os estimadores de máxima verosimilhança dos parâmetros, apresentaremos um teste que propicia inferência sobre o parâmetro de escala. Constataremos que o uso de transformações logarítmicas de quocientes entre estatísticas ordinais de Pareto generalizadas (índice* $\alpha \neq 0$*) pode ser uma solução viável para instrumento de decisão estatística. Por outro lado, um estudo formal dos* spacings *de misturas de Pareto generalizadas e dos* spacings *Laplaceanos também será efectuado.*

2.1 Modelo Pareto Generalizado

A distribuição de Pareto generalizada engloba três classes de distribuições:

$$F_{1,\alpha}(x) = \begin{cases} 0 & , \ x \leq 1 \\ 1 - x^{-\alpha} & , \ x > 1 \end{cases} \qquad \textit{Pareto} \qquad (2.1.1)$$

$$F_{2,\alpha}(x) = \begin{cases} 0 & , \ x \leq -1 \\ 1 - (-x)^{\alpha} & , \ -1 < x < 0 \\ 1 & , \ x \geq 0 \end{cases} \qquad \textit{Uniforme (-1,0), etc.} \qquad (2.1.2)$$

$$F_{3,0} \equiv F_3(x) = \begin{cases} 0 & , \ x \leq 0 \\ 1 - e^{-x} & , \ x > 0 \end{cases} \qquad \textit{Exponencial}$$

onde $\alpha > 0$*, sendo* (2.1.1) *e* (2.1.2) *denominadas Pareto tipo I e Pareto tipo II, respectivamente*[1]*. Nas distribuições anteriores podemos incluir parâmetros de localização e escala. No entanto, é mais usual considerar-se em* (2.1.1) *e* (2.1.2) *somente uma escala porque de certa forma esta funciona também como localização.*

As três classes de distribuições da Pareto generalizada estão relacionadas entre si por transformações simples. Ora se $Z \frown F_3 \equiv Exp(1)$*, facilmente se verifica que*

$$e^{Z/\alpha} \frown F_{1,\alpha} \qquad (2.1.3)$$

e

$$-e^{-Z/\alpha} \frown F_{2,\alpha} \qquad (2.1.4)$$

com $\alpha > 0$*.*

As relações (2.1.3) *e* (2.1.4) *permitem estender rapidamente muitas propriedades do modelo exponencial aos outros membros da Pareto generalizada. Como recorreremos frequentemente a essas propriedades, faremos em primeiro lugar um breve apontamento sobre o modelo exponencial.*

[1]Repare-se que a distribuição de Pareto tipo I corresponde à distribuição de Pareto clássica. Doravante, sempre que referirmos simplesmente "Pareto", subentenda-se "Pareto tipo I".

2.1.1 Modelo Exponencial

A distribuição exponencial é um caso particular da distribuição gama para $\alpha = 1$. *Assim, se* $X \frown G(1, \delta, \lambda) \equiv Exp(\delta, \lambda)$, *tem f.d.*

$$F_X(x) = \left[1 - \exp\left(-\frac{x-\lambda}{\delta}\right)\right] I_{(\lambda,\infty)} \tag{2.1.5}$$

e f.d.p.

$$f_X(x) = \frac{1}{\delta}\exp\left(-\frac{x-\lambda}{\delta}\right) I_{(\lambda,\infty)}. \tag{2.1.6}$$

Consequentemente,

$$E[X] = \delta + \lambda \quad e \quad Var[X] = \delta^2.$$

A distribuição exponencial possui, além de falta de memória, outras propriedades interessantes, sendo, por exemplo, infinitamente divisível. De facto, a função característica de $X \frown Exp(\delta, \lambda)$ *é*

$$\varphi_X(t) = e^{i\lambda t}(1 - i\delta t)^{-1}$$

e como $\varphi_X^{1/n}(t) = e^{i\frac{\lambda}{n}t}(1 - i\delta t)^{-1/n}$ *representa a função característica de uma gama com índice* $1/n$, *escala* δ *e localização* λ/n, *segue-se que a distribuição exponencial é I.D..*

Como estaremos essencialmente interessados nas propriedades dos spacings *da exponencial, considere-se o vector* $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ *das e.o. ascendentes que resulta da ordenação de* n *v.a.* X_k *i.i.d. com f.d.* (2.1.5), *e denotemos os* spacings *por*

$$S_k = X_{k:n} - X_{k-1:n}, \quad k = 1, \ldots, n \tag{2.1.7}$$

com a convenção usual $X_{0:n} \equiv \lambda$.

Ora o vector aleatório $(X_{1:n}, X_{2:n}, \ldots, X_{n:n})$ *tem f.d.p. conjunta*

$$f_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = \begin{cases} n! \prod_{k=1}^{n} \frac{1}{\delta}\exp\left(-\frac{x_k-\lambda}{\delta}\right) & , \ 0 < x_1 < x_2 < \cdots < x_n \\ 0 & , \ \textit{caso contrário} \end{cases}.$$

Se considerarmos a transformação definida em (2.1.7)*, segue-se que*

$$\begin{cases} X_{1:n} = S_1 + \lambda \\ X_{2:n} = S_1 + S_2 + \lambda \\ \quad\vdots \\ X_{n:n} = S_1 + S_2 + \cdots + S_n + \lambda \end{cases} \Rightarrow \left| \frac{\partial(X_{1:n}, X_{2:n}, \ldots, X_{n:n})}{\partial(S_1, S_2, \ldots, S_n)} \right| = 1$$

pelo que obtemos para f.d.p. conjunta de $(S_1, S_2, \ldots, S_n)$

$$\begin{aligned} f_{(S_1,S_2,\ldots,S_n)}(s_1, s_2, \ldots, s_n) &= \\ &= f_{1,2,\ldots,n:n}(s_1 + \lambda, s_1 + s_2 + \lambda, \ldots, s_1 + s_2 + \cdots + s_n + \lambda) \\ &= n! \frac{1}{\delta} \exp\Big(-\frac{s_1}{\delta}\Big) \frac{1}{\delta} \exp\Big(-\frac{s_1 + s_2}{\delta}\Big) \cdots \frac{1}{\delta} \exp\Big(-\frac{s_1 + s_2 + \cdots + s_n}{\delta}\Big) \\ &= \prod_{k=1}^{n} \frac{n-k+1}{\delta} \exp\Big(-\frac{n-k+1}{\delta} s_k\Big) \\ &= \prod_{k=1}^{n} f_{S_k}(s_k), \quad s_k > 0, \;\; k = 1, \ldots, n \end{aligned}$$

onde $f_{S_k}(s_k) = \frac{n-k+1}{\delta} \exp\big(-\frac{n-k+1}{\delta} s_k\big)$.

Assim, como a f.d.p. conjunta de $(S_1, S_2, \ldots, S_n)$ *permite a factorização anterior, podemos concluir que os* spacings $S_1, S_2, \ldots, S_n$ *são independentes. Repare-se que ao provarmos a indepêndencia dos* spacings*, provamos indirectamente a sua exponencialidade, isto é,*

$$X_{k:n} - X_{k-1:n} \frown Exp\left(\frac{\delta}{n-k+1}, 0\right) \equiv Exp\left(\frac{\delta}{n-k+1}\right), \quad k = 1, \ldots, n \quad (2.1.8)$$

e por conseguinte, a sua divisibilidade infinita.

É importante frisarmos que a exponencialidade e a independência de spacings *só ocorre no modelo exponencial, sendo por isso outra caracterização da distribuição exponencial (cf. Galambos (1987), teorema 1.6.3).*

Considere-se o modelo exponencial standard, isto é, $\delta = 1$ *e* $\lambda = 0$*. Ora é possível expressar cada e.o. com parente exponencial em função dos* spacings *normalizados, isto* *é, em função das v.a.*

$$Z_j = (n-j+1)(X_{j:n} - X_{j-1:n}) \frown Exp(1), \quad j = 1, \ldots, n \tag{2.1.9}$$

dado que

$$X_{k:n} \stackrel{d}{=} \sum_{j=1}^{k}(X_{j:n} - X_{j-1:n}) \stackrel{d}{=} \sum_{j=1}^{k} \frac{Z_j}{n-j+1}, \quad k = 1, \ldots, n. \tag{2.1.10}$$

O resultado anterior deve-se a Rényi (1953) e mostra que a k-ésima e.o. com parente exponencial é exprimível como combinação linear de k exponenciais standard i.i.d.. Esta propriedade tem sido basilar na resolução de muitos problemas relacionados com a distribuição exponencial como, por exemplo, na obtenção da distribuição assintótica de $X_{k:n}$ com k fixo.

Da igualdade (2.1.10)*, é imediato que*

$$E[X_{k:n}] = \sum_{j=1}^{k} \frac{1}{n-j+1} \;\; e \;\; Var[X_{k:n}] = \sum_{j=1}^{k} \frac{1}{(n-j+1)^2}$$

e

$$X_{j:n} - X_{i:n} \stackrel{d}{=} X_{j-i:n-i}, \quad 1 \le i < j \le n \tag{2.1.11}$$

obtendo-se como caso particular

$$X_{k:n} - X_{k-1:n} \stackrel{d}{=} X_{1:n-k+1}, \quad k = 1, \ldots, n.$$

Se os parâmetros δ e λ da exponencial forem desconhecidos, podemos usar em sua substituição as estimativas de máxima verosimilhança. Assim, se $(X_1, \ldots, X_n)$ for uma amostra aleatória (a.a.) proveniente de uma população exponencial com localização λ e escala δ, tem-se

$$\hat{\lambda} = X_{1:n} \quad e \quad \hat{\delta} = \frac{1}{n}\sum_{k=1}^{n}(X_k - X_{1:n}). \tag{2.1.12}$$

Ora $\hat{\lambda} = X_{1:n} \frown Exp\big(\frac{\delta}{n}, \lambda\big)$, *enquanto*

$$\hat{\delta} = \frac{1}{n}\sum_{k=1}^{n}(X_{k:n} - X_{1:n}) = \frac{1}{n}\sum_{k=2}^{n}(n-k+1)(X_{k:n} - X_{k-1:n}) \frown G\left(n-1, \frac{\delta}{n}\right).$$

Repare-se que a independência dos spacings *acarreta a independência dos estimadores* $\hat{\delta}$ *e* $\hat{\lambda}$*. Assim, podemos usar a estatística* studentizada externamente $\frac{\hat{\lambda}-\lambda}{\hat{\delta}}$ *para fazer inferências sobre o parâmetro* λ*, e que tem distribuição beta de* 2^a *espécie com parâmetros 1 e* $(n-1)$*. (Brilhante (1996) usou em alternativa a estatística studentizada externamente* $\frac{X_{1:n}-\lambda}{X_{n:n}-X_{1:n}}$ *de cálculo mais rápido.)*

Foi recordado na secção 0.1 que a soma de n *exponenciais i.i.d. com escala* δ *é uma gama com índice* n *e escala* δ*. No capítulo 3 usaremos uma generalização deste resultado onde se relaxa a condição de igualdade de escala.*

Lema 2.1.1 (Convolução Generalizada de Exponenciais) *Sejam* $X_1, X_2, \ldots, X_n$ n *v.a. independentes com f.d.p.*

$$f_{X_i}(x) = \theta_i e^{-\theta_i x}, \quad x > 0, \qquad i = 1, 2, \ldots, n$$

onde $\theta_i > 0$ *e* $\theta_i \neq \theta_j$ *se* $i \neq j$*; então a v.a.* $S_n = \sum\limits_{i=1}^{n} X_i$ *tem f.d.p.*

$$f_{S_n}(s) = \sum_{i=1}^{n} C_{i,n}\theta_i e^{-\theta_i s}, \quad s > 0 \tag{1}$$

em que $C_{i,n} = \prod\limits_{\substack{j=1\\ j\neq i}}^{n} \frac{\theta_j}{\theta_j - \theta_i}$.

Dem.

Demonstraremos o lema por indução matemática. Porém, antes de o fazermos, recordemos que a f.d.p. da convolução de duas variáveis aleatórias não negativas e independentes X *e* Y *é dada por*

$$f_{X+Y}(s) = \int_0^s f_X(x) f_Y(s-x)dx = \int_0^s f_X(s-x) f_Y(x)dx.$$

Para $n = 2$, tem-se

$$\begin{aligned}
f_{S_2}(s) &= \int_0^s f_{X_1}(x) f_{X_2}(s-x)dx \\
&= \theta_1\theta_2 e^{-\theta_2 s} \int_0^s e^{-(\theta_1-\theta_2)x} dx \\
&= \frac{\theta_1\theta_2}{\theta_1-\theta_2} e^{-\theta_2 s}\left[1 - e^{-(\theta_1-\theta_2)s}\right] \\
&= \frac{\theta_2}{\theta_2-\theta_1}\,\theta_1 e^{-\theta_1 s} + \frac{\theta_1}{\theta_1-\theta_2}\,\theta_2 e^{-\theta_2 s}, \quad s > 0
\end{aligned}$$

validando assim a expressão (1).

Admitamos para hipótese de indução

$$f_{S_{n-1}}(s) = \sum_{i=1}^{n-1} C_{i,n-1}\theta_i e^{-\theta_i s}, \quad s > 0$$

com $C_{i,n-1} = \prod_{\substack{j=1 \\ j\neq i}}^{n-1} \frac{\theta_j}{\theta_j-\theta_i}$.

Ora

$$\begin{aligned}
f_{S_n}(s) &= \int_0^s f_{S_{n-1}}(x) f_{X_n}(s-x)dx \\
&= \sum_{i=1}^{n-1} C_{i,n-1}\theta_i\theta_n e^{-\theta_n s} \int_0^s e^{-(\theta_i-\theta_n)x} dx \\
&= \sum_{i=1}^{n-1} C_{i,n-1}\frac{\theta_i\theta_n}{\theta_i-\theta_n}\, e^{-\theta_n s}\left[1 - e^{-(\theta_i-\theta_n)s}\right] \\
&= \sum_{i=1}^{n-1} C_{i,n-1}\frac{\theta_i}{\theta_i-\theta_n}\theta_n e^{-\theta_n s} + \sum_{i=1}^{n-1} C_{i,n-1}\frac{\theta_n}{\theta_n-\theta_i}\theta_i e^{-\theta_i s} \\
&= \sum_{i=1}^{n-1} C_{i,n-1}\frac{\theta_i}{\theta_i-\theta_n}\theta_n e^{-\theta_n s} + \sum_{i=1}^{n-1} C_{i,n}\theta_i e^{-\theta_i s}.
\end{aligned}$$

Seja $k_n = \sum\limits_{i=1}^{n-1} C_{i,n-1} \frac{\theta_i}{\theta_i - \theta_n}$*; então,*

$$f_{S_n}(s) = k_n \theta_n e^{-\theta_n s} + \sum_{i=1}^{n-1} C_{i,n} \theta_i e^{-\theta_i s}. \tag{2.1.13}$$

Também podemos obter a f.d.p. de S_n *atendendo a que*

$$f_{S_n}(s) = \int_0^s f_{X_1}(s-x) f_{X_2+\cdots+X_n}(x) dx.$$

Assim, seguindo um raciocínio similar ao utilizado para encontrar a expressão (2.1.13)*, vem*

$$f_{S_n}(s) = \sum_{i=2}^{n} C_{i,n-1} \frac{\theta_i}{\theta_i - \theta_1} \theta_1 e^{-\theta_1 s} + \sum_{i=2}^{n} C_{i,n} \theta_i e^{-\theta_i s}.$$

Se $k_1 = \sum\limits_{i=2}^{n} C_{i,n-1} \frac{\theta_i}{\theta_i - \theta_1}$*, então*

$$f_{S_n}(s) = k_1 \theta_1 e^{-\theta_1 s} + \sum_{i=2}^{n} C_{i,n} \theta_i e^{-\theta_i s}. \tag{2.1.14}$$

Atendendo às expressões (2.1.13) *e* (2.1.14)*, vem*

$$k_n \theta_n e^{-\theta_n s} + \sum_{i=1}^{n-1} C_{i,n} \theta_i e^{-\theta_i s} = k_1 \theta_1 e^{-\theta_1 s} + \sum_{i=2}^{n} C_{i,n} \theta_i e^{-\theta_i s}$$

e que por sua vez é equivalente a termos

$$k_n \theta_n e^{-\theta_n s} + C_{1,n} \theta_1 e^{-\theta_1 s} = k_1 \theta_1 e^{-\theta_1 s} + C_{n,n} \theta_n e^{-\theta_n s}.$$

Sem perda de generalidade, admitamos que $\theta_1 > \theta_n$ *(para tal basta rearranjarmos os* X_i*'s de forma a que tal ocorra). Se multiplicarmos ambos os membros da equação* *anterior por $e^{\theta_n s}$ e depois tomarmos limites quando s tende para infinito, vem $k_n = C_{n,n}$. Assim, substituindo o valor de k_n em* (2.1.13)*, segue-se que*

$$f_{S_n}(s) = \sum_{i=1}^{n} C_{i,n}\theta_i e^{-\theta_i s}, \quad s > 0.$$

OBSERVAÇÃO*: A variável S_n também é conhecida por* hipoexponencial.

A partir deste momento não teremos grandes dificuldades em estender propriedades do modelo exponencial aos restantes elementos da família Pareto generalizada. Veremos que essa extensibilidade vem acompanhada de certos ajustamentos, e é feita essencialmente dos spacings *(*spacings *aditivos) da exponencial aos quocientes entre estatísticas ordinais consecutivas das Pareto tipo I e tipo II. Doravante, chamaremos a esses quocientes "*spacings *multiplicativos".*

2.1.2 Modelo Pareto Tipo I

Seja W uma v.a. com distribuição Pareto tipo I com índice α e escala δ, $W \frown F_{1,\alpha}(\delta)$, isto é, com f.d.

$$F_{1,\alpha}(w) = \left[1 - \left(\frac{w}{\delta}\right)^{-\alpha}\right] I_{(\delta,\infty)} \tag{2.1.15}$$

e f.d.p.

$$f_{1,\alpha}(w) = \frac{\alpha}{\delta}\left(\frac{x}{\delta}\right)^{-\alpha-1} I_{(\delta,\infty)} \tag{2.1.16}$$

(quando $\delta = 1$, indicaremos a distribuição por $F_{1,\alpha}$).

No modelo Pareto clássico há restrições a impor ao parâmetro α por forma a garantir a existência de momentos. De facto,

$$E[W^n] = \frac{\alpha\delta^n}{\alpha - n} \quad \textit{se } \alpha > n.$$

Por conseguinte, somente as Pareto com índice superior a 1 têm valor médio, enquanto só as com índice superior 2 têm variância dada por Var[W]=$\frac{\alpha\delta^2}{(\alpha-2)(\alpha-1)^2}$.

A distribuição de Pareto é I.D. uma vez que a sua f.d.p. é completamente monótona — isto é, $f_{1,\alpha}$ possui derivadas de todas as ordens que verificam a desigualdade

$(-1)^n f_{1,\alpha}^{(n)}(x) \geq 0,\ x > \delta$ *(cf. Steutel (1969), Teorema 1)*[2].

No que refere propriedades das e.o. com parente Paretiana, considere-se, sem perda de generalidade, $Z_{k:n}$ e $W_{k:n}$ as k-ésimas e.o. com parentes $Exp(1)$ e $F_{1,\alpha}$, respectivamente.

De (2.1.3)*, vem*

$$W_{k:n} \stackrel{d}{=} e^{Z_{k:n}/\alpha}, \quad k = 1, \dots, n$$

donde

$$\frac{W_{k:n}}{W_{k-1:n}} \stackrel{d}{=} \exp\left(\frac{Z_{k:n} - Z_{k-1:n}}{\alpha}\right)$$

ou ainda,

$$\frac{W_{k:n}}{W_{k-1:n}} \stackrel{d}{=} \exp\left\{\frac{(n-k+1)(Z_{k:n} - Z_{k-1:n})}{\alpha(n-k+1)}\right\}, \quad k = 1, \dots, n$$

com $W_{0:n} \equiv 1$.

Assim, de (2.1.9) *e* (2.1.3) *vem*

$$\frac{W_{k:n}}{W_{k-1:n}} \frown F_{1,\alpha(n-k+1)}, \quad k = 1, \dots, n \tag{2.1.17}$$

ou seja, os spacings *multiplicativos com parente Paretiana ainda são Paretos (com índices diferentes) e, por conseguinte, são I.D.. Repare-se que* $W_{1:n} \frown F_{1,\alpha n}$.

Por outro lado, a independência dos spacings *do modelo exponencial implica a independência dos* spacings *multiplicativos, isto é, as v.a.*

$$W_{1:n}, \quad \frac{W_{2:n}}{W_{1:n}}, \ \cdots\cdots \ , \frac{W_{n:n}}{W_{n-1:n}}$$

são independentes. Assim, podemos exprimir cada e.o. com parente Paretiana como produto de Paretos independentes dado que

[2] O critério de monotonicidade para divisibilidade infinita só pode ser usado em distribuições com suporte em $[a, \infty)$ e $a > -\infty$.

$$W_{k:n} \stackrel{d}{=} \frac{W_{k:n}}{W_{k-1:n}} \frac{W_{k-1:n}}{W_{k-2:n}} \cdots \frac{W_{2:n}}{W_{1:n}} W_{1:n}, \quad k = 1, \ldots, n. \tag{2.1.18}$$

Consequentemente, de (2.1.18) *é imediato que*

$$E[W_{k:n}] = \prod_{j=1}^{k} \frac{\alpha(n-j+1)}{\alpha(n-j+1)-1} \ \textit{ se } \alpha > \frac{1}{n-k+1}.$$

No que refere o spacing *multiplicativo generalizado* $\frac{W_{k:n}}{W_{j:n}}$*, vem de* (2.1.3) *e* (2.1.11)

$$\frac{W_{k:n}}{W_{j:n}} \stackrel{d}{=} W_{k-j:n-j}, \quad 1 \leq j < k \leq n$$

obtendo-se como caso particular

$$\frac{W_{k:n}}{W_{k-1:n}} \stackrel{d}{=} W_{1:n-k+1}, \quad k = 1, \ldots, n.$$

Os estimadores de máxima verosimilhança de α *e* δ *são*

$$\hat{\delta} = W_{1:n} \quad \textit{e} \quad \hat{\alpha} = \frac{n}{\sum\limits_{k=1}^{n} \ln\left(\frac{W_k}{W_{1:n}}\right)}. \tag{2.1.19}$$

Como $\hat{\delta} = W_{1:n}$*, é imediato que* $\hat{\delta} \frown F_{1,\alpha n}(\delta)$.

Por outro lado, a independência dos spacings *multiplicativos facilita consideravelmente a identificação da distribuição de* $\hat{\alpha}$.

Ora

$$\hat{\alpha} = \frac{n}{\sum\limits_{k=1}^{n} \ln\left(\frac{W_k}{W_{1:n}}\right)} = \frac{n}{\sum\limits_{k=1}^{n} \ln\left(\frac{W_{k:n}}{W_{1:n}}\right)} = \frac{n}{\sum\limits_{k=2}^{n} (n-k+1) \ln\left(\frac{W_{k:n}}{W_{k-1:n}}\right)}. \tag{2.1.20}$$

Dado que a distribuição de $\hat{\alpha}$ não depende de δ, consideremos, sem perda de generalidade, $\delta = 1$. Assim, atendendo a (2.1.17) *e* (2.1.3)*, vem*

$$\ln\left(\frac{W_{k:n}}{W_{k-1:n}}\right) \frown Exp\left(\frac{1}{\alpha(n-k+1)}\right) \Rightarrow (n-k+1)\ln\left(\frac{W_{k:n}}{W_{k-1:n}}\right) \frown Exp\left(\frac{1}{\alpha}\right)$$

pelo que no denominador de (2.1.20) *temos uma soma de exponenciais i.i.d. com escala $1/\alpha$. Consequentemente,*

$$\sum_{k=2}^{n}(n-k+1)\ln\left(\frac{W_{k:n}}{W_{k-1:n}}\right) \frown G\left(n-1,\frac{1}{\alpha}\right).$$

Como

$$\frac{1}{n}\sum_{k=2}^{n}(n-k+1)\ln\left(\frac{W_{k:n}}{W_{k-1:n}}\right) \frown G\left(n-1,\frac{1}{\alpha n}\right)$$

então $\hat{\alpha}$ tem f.d.p.

$$f_{\hat{\alpha}}(x) = \frac{(\alpha n)^{n-1}x^{-n}e^{-\alpha n/x}}{\Gamma(n-1)}I_{(0,\infty)}. \tag{2.1.21}$$

Note-se também que $\hat{\alpha}$ e $\hat{\delta}$ são independentes.

Caso tivéssemos considerado para o modelo exponencial a reparametrização $\theta = 1/\delta$, teríamos obtido para distribuição de $\hat{\theta}$ o inverso de uma gama com índice $(n-1)$ e escala $1/\theta n$. De facto, podemos admitir que o parâmetro $1/\alpha$, estimado por $1/\hat{\alpha}$, tem um comportamento similar à da escala do modelo exponencial. Assim, se quisermos fazer inferências sobre δ no modelo Pareto, que também funciona como localização, podemos usar a estatística $\frac{\ln W_{1:n}-\ln\delta}{1/\hat{\alpha}}$ por a distribuição de $1/\hat{\alpha}$ não depender de δ e o quociente considerado permitir descartar o parâmetro perturbador α.

É fácil identificarmos a distribuição do quociente $\frac{\ln W_{1:n}-\ln\delta}{1/\hat{\alpha}}$ dado que

$$\ln W_{1:n} - \ln\delta = \ln\left(\frac{W_{1:n}}{\delta}\right) \frown Exp\left(\frac{1}{\alpha n}\right) \equiv G\left(1,\frac{1}{\alpha n}\right)$$

e

$$\frac{1}{\hat{\alpha}} = \frac{1}{n}\sum_{k=1}^{n} \ln\left(\frac{W_k}{W_{1:n}}\right) \frown G\left(n-1, \frac{1}{\alpha n}\right)$$

Como $(\ln W_{1:n} - \ln \delta)$ *e* $1/\hat{\alpha}$ *são gamas independentes com igual escala, segue-se que*

$$\frac{n(\ln W_{1:n} - \ln \delta)}{\sum_{k=1}^{n} \ln\left(\frac{W_k}{W_{1:n}}\right)} \frown \mathcal{B}e^*(1, n-1). \tag{2.1.22}$$

Podemos usar em alternativa para estatística de teste o quociente entre aquelas gamas, divididas pelos respectivos graus de liberdade, obtendo-se uma F *de Snedecor, ou seja,*

$$\frac{(\ln W_{1:n} - \ln \delta)\big/2}{\frac{1}{n}\sum_{k=1}^{n} \ln\left(\frac{W_k}{W_{1:n}}\right)\Big/2(n-1)} = \frac{n(n-1)(\ln W_{1:n} - \ln \delta)}{\sum_{k=1}^{n} \ln\left(\frac{W_k}{W_{1:n}}\right)} \frown F_{(2,2(n-1))}. \tag{2.1.23}$$

Ora o uso de $\hat{\alpha}$ *para inferir sobre* α *não é uma opção a considerar por a distribuição de* $\hat{\alpha}$ *depender de* α*. Por outro lado, o recurso a funções de* spacings *(aditivos) para efeitos de inferência sobre* δ *pode revelar-se uma abordagem pouco viável como atestará a distribuição dos* spacings.

Denotemos por

$$U_{k,n} = W_{k:n} - W_{k-1:n}, \quad k = 1, \dots, n$$

com a convenção $W_{0:n} \equiv 1$ *(estamos a supor* $\delta = 1$*).*

Assim,

$$\begin{aligned}
f_{U_{k,n}}(u) &= \int_{-\infty}^{\infty} f_{k-1,k:n}(w, w+u)dw \\
&= \frac{n!}{(k-2)!(n-k)!}\int_{-\infty}^{\infty} [F_{1,\alpha}(w)]^{k-2}[1 - F_{1,\alpha}(w+u)]^{n-k} f_{1,\alpha}(w) f_{1,\alpha}(w+u)dw \\
&= \alpha^2 C_{k,n}\int_{1}^{\infty} (1 - w^{-\alpha})^{k-2} w^{-\alpha-1}(w+u)^{-\alpha(n-k+1)-1}dw \\
&= \alpha^2 C_{k,n}\sum_{j=0}^{k-2}\binom{k-2}{j}(-1)^j\int_{1}^{\infty} w^{-\alpha(j+1)-1}(w+u)^{-\alpha(n-k+1)-1}dw
\end{aligned}$$

em que $C_{k,n} = \frac{n!}{(k-2)!(n-k)!}$.

Efectuando a mudança de variável $w = \frac{u}{x}$*, obtém-se*

$$f_{U_{k,n}}(u) = \alpha^2 C_{k,n} \sum_{j=0}^{k-2} \binom{k-2}{j} (-1)^j u^{-\alpha(n+j-k+2)-1} \int_0^u \frac{x^{\alpha(n+j-k+2)}}{(1+x)^{\alpha(n-k+1)+1}} dx$$

e atendendo à fórmula

$$\int_0^u \frac{x^{\mu-1}dx}{(1+\beta x)^{\nu}} = \frac{u^{\mu}}{\mu} {}_2F_1(\nu, \mu; 1+\mu; -\beta u), \quad \textit{Re}\ \mu > 0 \textit{ e } |\textit{arg}(1+\beta u)| < \pi \tag{2.1.24}$$

(cf. Gradshteyn et al. *(1994), p. 333), segue-se que*

$$\begin{aligned} f_{U_{k,n}}(u) = {} & \frac{\alpha^2 n!}{(k-2)!(n-k)!} \sum_{j=0}^{k-2} \binom{k-2}{j} (-1)^j \frac{1}{\alpha(n+j-k+2)+1} \\ & \times {}_2F_1(\alpha(n-k+1)+1, \alpha(n+j-k+2)+1; \alpha(n+j-k+2)+2; -u) \end{aligned} \tag{2.1.25}$$

se $u > 0$ *e* $k = 2, \ldots, n$*; enquanto o primeiro* spacing*,* $U_{1,n} = X_{1:n} - 1$*, tem f.d.p.*

$$f_{U_{1,n}}(u) = n\alpha(1+u)^{-n\alpha-1}, \quad u > 0 \tag{2.1.26}$$

ou seja, é uma Pareto de 2^a *espécie ou de Lomax.*

Como podemos observar, as f.d.p. dos spacings *com parente Paretiana envolvem combinações lineares de funções hipergeométricas de Gauss com* α *"entranhado"nessas funções.*

Tudo isto vem mostrar que há nítidas vantagens em considerarmos transformações logarítmicas de spacings *multiplicativos aquando da concepção de novas estatísticas de teste.*

Teste para homogeneidade de localização (escala) com parente Paretiana

Em populações Paretianas o parâmetro de escala também funciona como localização, de modo que se quisermos testar igualdade de localização (escala) de $k \geq 2$ *populações,* *convém usarmos para o efeito um teste que proporcione a comparação simultânea das localizações (escalas).*

Numa análise de escala[3]*, é usual admitir-se à partida que as escalas das populações são iguais, pelo que suporemos no caso de parente Paretiana a igualdade de forma (apesar de não podermos falar propriamente numa análise de escala neste caso, estamos a admitir que os parâmetros de forma das Paretos funcionam como escalas). De facto, podemos argumentar que o parâmetro α comporta-se como uma escala. Note-se que α aparece a multiplicar na identidade $Z \stackrel{d}{=} \alpha \ln W$ com $Z \frown Exp(1)$ e $W \frown F_{1,\alpha}$, sugerindo-nos que tem um comportamento idêntico ao do parâmetro de escala do modelo exponencial.*

No entanto, por forma a obtermos uma decomposição do estimador global de "escala"em dois estimadores independentes, um reflectindo a "variação"entre as amostras e o outro reflectindo a "variação"dentro das amostras, à semelhança do que acontece na análise de variância clássica e na análise de escala com parente exponencial (Pestana e Rocha (1993)), usaremos para "escala"o parâmetro $\alpha^* = 1/\alpha$.

Os estimadores considerados serão os de máxima verosimilhança.

Sejam

$$\begin{aligned} \underset{\sim}{W}_1 &= (W_{11}, W_{12}, \ldots, W_{1n_1}) \\ \underset{\sim}{W}_2 &= (W_{21}, W_{22}, \ldots, W_{2n_2}) \\ &\vdots \\ \underset{\sim}{W}_k &= (W_{k1}, W_{k2}, \ldots, W_{kn_k}) \end{aligned}$$

k *amostras aleatórias independentes, onde*

$$W_{ij} \frown F_{1,\alpha}(\delta_i), \quad j = 1, 2, \ldots, n_i;\ i = 1, 2, \ldots, k.$$

Denotemos por $N = \sum_{i=1}^{k} n_i$ *e por* $W_{1:N}$ *o mínimo da amostra combinada.*

Sob a hipótese

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_k = \delta$$

[3]Note-se que a análise de variância clássica mais não é que uma análise de escala com parente Gaussiana.

os estimadores de δ e α^* baseados na amostra combinada são $\hat{\hat{\delta}} = W_{1:N}$ e

$$\widehat{\alpha^*} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left(\frac{W_{ij}}{W_{1:N}}\right) = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left(\frac{W_{j[i]:n_i}}{W_{1:N}}\right)$$

onde $W_{j[i]:n_i}$ denota a j-ésima e.o. ascendente da i-ésima amostra.

Ora

$$\begin{aligned}
\widehat{\alpha^*} &= \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left(\frac{W_{j[i]:n_i}}{W_{1:N}}\right)\\
&= \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left(\frac{W_{j[i]:n_i}}{W_{1[i]:n_i}}\frac{W_{1[i]:n_i}}{W_{1:N}}\right)\\
&= \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left(\frac{W_{j[i]:n_i}}{W_{1[i]:n_i}}\right) + \frac{1}{N}\sum_{i=1}^{k} n_i \ln\left(\frac{W_{1[i]:n_i}}{W_{1:N}}\right)\\
&= \widehat{\alpha^*_D} + \widehat{\alpha^*_E}
\end{aligned}$$

*com $\widehat{\alpha^*_D}$ a reflectir a "variação"*dentro das amostras e $\widehat{\alpha^*_E}$ a "variação"*entre as amostras.*

Por um lado,

$$\widehat{\alpha^*_D} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\ln\left(\frac{W_{j[i]:n_i}}{W_{1[i]:n_i}}\right) = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=2}^{n_i}(n_i - j + 1)\ln\left(\frac{W_{j[i]:n_i}}{W_{(j-1)[i]:n_i}}\right). \quad (2.1.27)$$

Da independência dos spacings *multiplicativos com parente Paretiana e do facto de*

$$\ln\left(\frac{W_{j[i]:n_i}}{W_{(j-1)[i]:n_i}}\right) \frown Exp\left(\frac{\alpha^*}{n_i - j + 1}\right), \quad j = 1, \ldots, n_i$$

segue-se que

$$\sum_{j=2}^{n_i}(n_i - j + 1)\ln\left(\frac{W_{j[i]:n_i}}{W_{(j-1)[i]:n_i}}\right) \frown G(n_i - 1, \alpha^*), \quad i = 1, \ldots, k.$$

Donde,

$$\widehat{\alpha_D^*} \frown G\left(\sum_{i=1}^{k}(n_i-1), \frac{\alpha^*}{N}\right) \equiv G\left(N-k, \frac{\alpha^*}{N}\right). \tag{2.1.28}$$

(note-se que independentemente de H_0 ser verdadeira ou não, $\widehat{\alpha_D^*} \frown G(N-k, \alpha^*/N)$)

Por outro lado,

$$\begin{aligned}
\widehat{\alpha_E^*} &= \frac{1}{N}\sum_{i=1}^{k} n_i \ln\left(\frac{W_{1[i]:n_i}}{W_{1:N}}\right) \\
&= \frac{1}{N}\left(\sum_{i=1}^{k} n_i \ln W_{1[i]:n_i} - N \ln X_{1:N}\right) \\
&= \frac{1}{N}\left(\sum_{i=1}^{k} \ln W_{1[i]:n_i}^{n_i} - \ln W_{1:N}^{N}\right).
\end{aligned}$$

Dado que os quocientes $\frac{W_{1[i]:n_i}}{W_{1:N}}$ são distribucionalmente independentes de δ, podemos considerar, sem perda de generalidade, $\delta = 1$, de modo que sob a validade de H_0

$$W_{1[i]:n_i}^{n_i} \frown F_{1,\alpha} \quad e \quad W_{1:N}^{N} \frown F_{1,\alpha}.$$

Como $W_{1:N} = \min\limits_{1\le i\le k}\{W_{1[i]:n_i}\}$, temos que $\widehat{\alpha_E^*}$ é exprimível em função de combinações lineares de logaritmos dos spacings *multiplicativos de k v.a. i.i.d. com parente Paretiana com índice α. Por conseguinte,*

$$\widehat{\alpha_E^*} = \frac{1}{N}\sum_{i=1}^{k} n_i \ln\left(\frac{W_{1[i]:n_i}}{W_{1:N}}\right) \frown G\left(k-1, \frac{\alpha^*}{N}\right). \tag{2.1.29}$$

Ora a independência dos spacings *multiplicativos implica a independência de $\widehat{\alpha_E^*}$ e $\widehat{\alpha_D^*}$, de modo que obtivémos a decomposição do estimador global $\widehat{\alpha^*}$ com $2(N-1)$ g.l. em dois estimadores independentes, $\widehat{\alpha_E^*}$ com $2(k-1)$ g.l. e $\widehat{\alpha_D^*}$ com $2(N-k)$ g.l..*

Assim, se dividirmos aqueles estimadores pelos respectivos g.l., vem para estatística *de teste*

$$\frac{\widehat{\alpha_E^*}\big/(k-1)}{\widehat{\alpha_D^*}\big/(N-k)} = \frac{\sum_{i=1}^{k} n_i \ln\left(\frac{W_{1[i]:n_i}}{W_{1:N}}\right)\Big/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} \ln\left(\frac{W_{ij}}{W_{1[i]:n_i}}\right)\Big/(N-k)} \frown F_{(2(k-1),2(N-k))}. \qquad (2.1.30)$$

Consequentemente, teremos razões para rejeitar H_0 *com base nas amostras e a um nível de significância* $\gamma\%$ *se*

$$\frac{\widehat{\alpha_E^*}\big/(k-1)}{\widehat{\alpha_D^*}\big/(N-k)} \geq F_{1-\gamma;2(k-1),2(N-k)}$$

onde $F_{1-\gamma;2(k-1),2(N-k)}$ *denota o quantil de ordem* $1-\gamma$ *da distribuição* F *de Snedecor com* $2(k-1)$ *e* $2(N-k)$ *graus de liberdade.*

Note-se que a partir de (2.1.30)*, e atendendo à relação* (2.1.3)*, podemos obter os resultados de Pestana e Rocha (1993) relativos à análise de escala com parente exponencial.*

O lema 2.1.1 sobre convolução generalizada de exponenciais pode ser usado na obtenção da distribuição de um produto generalizado de Paretos, e que por analogia com o caso exponencial podemos designar hipopareto.

Lema 2.1.2 (Produto Generalizado de Paretos) *Se* $W_1,\ldots,W_n$ *forem* n *v.a. independentes com* $W_i \frown F_{1,\alpha_i}(\delta_i)$*, onde* $\alpha_i \neq \alpha_j$ *se* $i \neq j$*, então* $P_n = \prod_{i=1}^{n} W_i$ *tem f.d.p.*

$$f_{P_n}(x) = \sum_{i=1}^{n} C_{i,n}\frac{\alpha_i}{\delta}\left(\frac{x}{\delta}\right)^{-\alpha_i-1}, \quad x > \delta$$

onde $\delta = \prod_{j=1}^{n}\delta_j$ *e* $C_{i,n} = \prod_{\substack{j=1\\ j\neq i}}^{n} \frac{\alpha_j}{\alpha_j-\alpha_i}$.

Dem.

De (2.1.3)*, vem*

$$W_i \stackrel{d}{=} \delta_i e^{Z_i/\alpha_i}, \quad i = 1, \dots, n$$

onde $Z_1, \dots, Z_n$ são exponenciais standard i.i.d.. Assim,

$$P_n = \prod_{i=1}^{n} W_i \stackrel{d}{=} \delta \exp\left(\sum_{i=1}^{n} \frac{Z_i}{\alpha_i}\right)$$

com $\delta = \prod_{j=1}^{n} \delta_j$.

Como $\frac{Z_i}{\alpha_i} \frown Exp\big(\frac{1}{\alpha_i}\big)$, a f.d.p. de $S_n = \sum_{i=1}^{n} \frac{Z_i}{\alpha_i}$ é a convolução generalizada de n exponenciais, pelo que decorre do lema 2.1.1 que

$$f_{S_n}(s) = \sum_{i=1}^{n} C_{i,n} \alpha_i e^{-\alpha_i s}, \quad s > 0$$

com $C_{i,n} = \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{\alpha_j}{\alpha_j - \alpha_i}$. Logo, P_n tem f.d.p.

$$f_{P_n}(x) = \sum_{i=1}^{n} C_{i,n} \frac{\alpha_i}{\delta} \left(\frac{x}{\delta}\right)^{-\alpha_i - 1}, \quad x > \delta.$$

2.1.3 Modelo Pareto Tipo II

Seja V uma v.a. com distribuição Pareto tipo II com índice α e escala δ, $V \frown F_{2,\alpha}(\delta)$, isto é, com f.d.

$$F_{2,\alpha}(v) = \begin{cases} 0 & , \ v \leq -\delta \\ 1 - \left(-\frac{v}{\delta}\right)^{\alpha} & , \ -\delta < v < 0 \\ 1 & , \ v \geq 0 \end{cases} \tag{2.1.31}$$

e f.d.p.

$$f_{2,\alpha}(v) = \frac{\alpha}{\delta}\left(-\frac{v}{\delta}\right)^{\alpha-1} I_{(-\delta,0)}. \tag{2.1.32}$$

Ao contrário do modelo Pareto tipo I, não há restrições a considerar para a existência de momentos. De facto, o modelo Pareto tipo II possui momentos de todas as ordens, *sendo o de ordem n igual a*

$$E[V^n] = (-1)^n \frac{\alpha\delta^n}{\alpha + n}.$$

Como a distribuição $F_{2,\alpha}$ possui suporte limitado não é I.D.. No entanto, como logaritmos de betas são I.D. (cf. Shanbhag, Pestana e Sreehari (1977)), e $-V \frown \mathcal{B}e(\alpha, 1, 0, \delta)$ se $V \frown F_{2,\alpha}(\delta)$, então $\ln(-V)$ é I.D..

Considere-se, sem perda de generalidade, $V_{k:n}$ e $Z_{k:n}$ as k-ésimas e.o. com parentes $F_{2,\alpha}$ e $Exp(1)$, respectivamente. Decorre de (2.1.4) *que*

$$V_{k:n} \stackrel{d}{=} -e^{-Z_{k:n}/\alpha}, \quad k = 1, \dots, n$$

pelo que

$$\frac{V_{k:n}}{V_{k-1:n}} \stackrel{d}{=} \exp\left(-\frac{Z_{k:n} - Z_{k-1:n}}{\alpha}\right)$$

ou

$$-\frac{V_{k:n}}{V_{k-1:n}} \stackrel{d}{=} -\exp\left(-\frac{(n-k+1)(Z_{k:n} - Z_{k-1:n})}{\alpha(n-k+1)}\right).$$

Assim, de (2.1.9) *e* (2.1.4) *vem*

$$-\frac{V_{k:n}}{V_{k-1:n}} \frown F_{2,\alpha(n-k+1)}$$

ou

$$\frac{V_{k:n}}{V_{k-1:n}} \frown \mathcal{B}e(\alpha(n-k+1), 1), \quad k = 1, \dots, n.$$

Por outras palavras, os spacings *multiplicativos com parente $F_{2,\alpha}$ são betas, em particular, são funções potência standard (com parâmetros diferentes). Note-se que com a convenção usual $V_{0:n} = -1$, tem-se $-V_{1:n} \frown \mathcal{B}e(\alpha n, 1)$, ou $V_{1:n} \frown F_{2,\alpha n}$. Consequentemente, os logaritmos dos* spacings *multiplicativos com parente Paretiana tipo II são I.D., ou seja, $\ln(-V_{1:n})$ e $\ln\left(\frac{V_{k:n}}{V_{k-1:n}}\right)$, $k = 2, \dots, n$ são I.D. (note-se que bastaria termos justificado a divisibilidade infinita de tais logaritmos por serem exponenciais negativas).*

Tal como acontece no modelo Pareto clássico, a independência dos spacings *(aditivos) da exponencial causa a independência dos* spacings *multiplicativos*

$$-V_{1:n}, \quad \frac{V_{2:n}}{V_{1:n}}, \; \ldots\ldots \;, \frac{V_{n:n}}{V_{n-1:n}}.$$

Como podemos exprimir o simétrico da k-ésima e.o. com parente Paretiana tipo II como produto de k betas (funções potência standard) independentes, ou seja,

$$-V_{k:n} \stackrel{d}{=} \frac{V_{k:n}}{V_{k-1:n}} \frac{V_{k-1:n}}{V_{k-2:n}} \cdots \frac{V_{2:n}}{V_{1:n}} (-V_{1:n}), \quad k = 1, \ldots, n \tag{2.1.33}$$

é imediato que

$$E[V_{k:n}] = -\prod_{j=1}^{k} \frac{\alpha(n-j+1)}{\alpha(n-j+1)+1}.$$

Refira-se que a independência de spacings *multiplicativos de v.a. i.i.d. positivas ou negativas com f.d. comum F ocorre apenas se $F = F_{1,\alpha}$ ou $F = F_{2,\alpha}$ (cf. Galambos (1987)), constituíndo por isso uma caracterização dos modelos Pareto tipo I e tipo II.*

Recorde-se ainda que numa população $U(0,1)$ são independentes os quocientes

$$\frac{U_{n-1:n}}{U_{n:n}}, \frac{U_{n-2:n}}{U_{n-1:n}}, \ldots\ldots, \frac{U_{1:n}}{U_{2:n}}, U_{1:n}$$

(devendo-se este resultado a Malmquist (1950)), e não os quocientes do tipo $\frac{U_{k:n}}{U_{k-1:n}}$. De facto, os quocientes $\frac{Y_{k:n}}{Y_{k+1:n}}$, $k = 1, \ldots, n$, onde $Y_{k:n}$ é a k-ésima e.o. de n v.a. $Y_1, \ldots, Y_n$ i.i.d. não negativas com f.d. F, são independentes sse $F_Y(y) = Cy^\alpha$ para $y \in (0, A)$ com $C > 0$ e $\alpha > 0$ (cf. Galambos (1987), corolário 1.6.1).

No que respeita spacings *multiplicativos generalizados, segue de* (2.1.11)

$$\frac{V_{k:n}}{V_{j:n}} \stackrel{d}{=} -V_{k-j:n-j}, \quad 1 \le j < k \le n$$

obtendo-se como caso particular

$$\frac{V_{k:n}}{V_{k-1:n}} \stackrel{d}{=} -V_{1:n-k+1}, \quad k = 1, \ldots, n.$$

Os estimadores de máxima verosimilhança de δ e α são

$$\hat{\delta} = -V_{1:n} \quad e \quad \hat{\alpha} = -\frac{n}{\sum_{k=1}^{n} \ln\left(\frac{V_k}{V_{1:n}}\right)} \tag{2.1.34}$$

e a independência dos spacings *multiplicativos com parente $F_{2,\alpha}(\delta)$ também facilita a identificação da distribuição de $\hat{\alpha}$.*

Ora como $\hat{\delta} = -V_{1:n}$ e $V_{1:n} \frown F_{2,\alpha n}(\delta)$, tem-se $\hat{\delta} \frown \mathcal{B}e(\alpha n, 1, 0, \delta)$.

Quanto ao estimador $\hat{\alpha}$, tem-se

$$\hat{\alpha} = -\frac{n}{\sum_{k=1}^{n} \ln\left(\frac{V_k}{V_{1:n}}\right)} = -\frac{n}{\sum_{k=1}^{n} \ln\left(\frac{V_{k:n}}{V_{1:n}}\right)}$$

ou ainda,

$$\hat{\alpha} = \frac{n}{-\sum_{k=2}^{n} (n-k+1) \ln\left(\frac{V_{k:n}}{V_{k-1:n}}\right)}. \tag{2.1.35}$$

A distribuição de $\hat{\alpha}$ também é distribucionalmente independente de δ pelo que consideremos, sem perda de generalidade, $\delta = 1$.

Ora se $V \frown F_{2,\alpha}$, então $-V \frown \mathcal{B}e(\alpha, 1)$, seguindo-se de (2.1.4) *que $-\ln(-V) \frown Exp(1/\alpha)$. Assim,*

$$-\ln\left(\frac{V_{k:n}}{V_{k-1:n}}\right) \frown Exp\left(\frac{1}{\alpha(n-k+1)}\right) \Rightarrow -(n-k+1)\ln\left(\frac{V_{k:n}}{V_{k-1:n}}\right) \frown Exp\left(\frac{1}{\alpha}\right).$$

Consequentemente, no denominador de (2.1.35) *temos uma soma de exponenciais i.i.d. com escala $1/\alpha$, pelo que*

$$-\sum_{k=2}^{n} (n-k+1) \ln\left(\frac{V_{k:n}}{V_{k-1:n}}\right) \frown G\left(n-1, \frac{1}{\alpha}\right)$$

seguindo-se que

$$-\frac{1}{n}\sum_{k=2}^{n}(n-k+1)\ln\left(\frac{V_{k:n}}{V_{k-1:n}}\right) \frown G\left(n-1,\frac{1}{\alpha n}\right).$$

Donde, $\hat{\alpha}$ *tem f.d.p.* (2.1.21). *Repare-se que os estimadores* $\hat{\alpha}$ *e* $\hat{\delta}$ *também são independentes.*

As conclusões proferidas no modelo anterior, nomeadamente no que refere estratégias para escolha de estatísticas para inferência sobre δ*, são válidas para o modelo Pareto tipo II. No entanto, a estatística a considerar para inferir sobre* δ *tem a expressão* $\frac{-\ln(-V_{1:n})+\ln\delta}{1/\hat{\alpha}}$.

A independência entre numerador e denominador permite-nos afirmar que

$$\frac{-\ln(-V_{1:n})+\ln\delta}{1/\hat{\alpha}} = \frac{n(\ln(-V_{1:n})-\ln\delta)}{\sum_{k=1}^{n}\ln\left(\frac{V_k}{V_{1:n}}\right)} \frown \mathcal{B}e^*(1,n-1).$$

Em alternativa, podemos usar a estatística

$$\frac{n(n-1)[\ln(-V_{1:n})-\ln\delta]}{\sum_{k=1}^{n}\ln\left(\frac{V_k}{V_{1:n}}\right)} \frown F_{(2,2(n-1))}. \tag{2.1.36}$$

Se denotarmos por

$$U_{k,n} = V_{k:n} - V_{k-1:n}, \quad k=1,\ldots,n$$

com a convenção $V_{0:n} \equiv -1$ *(estamos a considerar, sem perda de generalidade,* $\delta = 1$*), tem-se*

$$\begin{aligned}
f_{U_{k,n}}(u) &= \int_{-\infty}^{\infty} f_{k-1,k:n}(v,v+u)dv \\
&= C_{k,n}\int_{-1}^{-u}[1-(-v)^{\alpha}]^{k-2}[-(v+u)]^{\alpha(n-k)}\alpha(-v)^{\alpha-1}\alpha[-(v+u)]^{\alpha-1}dv \\
&= \alpha^2 C_{k,n}\sum_{j=0}^{k-2}\binom{k-2}{j}(-1)^j\int_{-1}^{-u}(-v)^{\alpha(j+1)-1}[-(v+u)]^{\alpha(n-k+1)-1}dv
\end{aligned}$$

onde $C_{k,n} = \frac{n!}{(k-2)!(n-k)!}$ *e* $0 < u < 1$.

Procedendo à transformação $-v = \frac{u}{1-x}$, *segue-se que*

$$f_{U_{k,n}}(u) = \alpha^2 C_{k,n} \sum_{j=0}^{k-2} \binom{k-2}{j} (-1)^j u^{\alpha(n+j-k+2)-1} \int_0^{1-u} \frac{x^{\alpha(n-k+1)-1} dx}{(1-x)^{\alpha(n+j-k+2)}}$$

e atendendo a (2.1.24), *vem*

$$\begin{aligned} f_{U_{k,n}}(u) = & \frac{\alpha n!}{(n-k+1)(k-2)!(n-k)!} \sum_{j=0}^{k-2} \binom{k-2}{j} (-1)^j \\ & \times u^{\alpha(n+j-k+2)-1} (1-u)^{\alpha(n-k+1)} \\ & \times {}_2F_1(\alpha(n+j-k+2), \alpha(n-k+1); \alpha(n-k+1)+1; 1-u) \end{aligned} \tag{2.1.37}$$

com $0 < u < 1$ *e* $k = 2, \ldots, n$.

Quanto à f.d.p. de $U_{1,n} = X_{1:n} + 1$, *tem-se*

$$f_{U_{1,n}}(u) = n\alpha(1-u)^{n\alpha-1}, \quad 0 < u < 1 \tag{2.1.38}$$

ou seja, $U_{1,n} \frown \mathcal{B}e(1, \alpha n)$.

Teste para homogeneidade de localização (escala) com parente Paretiana tipo II

Na secção anterior avançamos com um teste que nos possibilita inferir sobre a homogeneidade de "localização"(escala) de $k \geq 2$ *populações Paretianas tipo I. Aqueles resultados são facilmente estendidos a populações Paretianas tipo II uma vez que o parâmetro de escala da Pareto tipo II também funciona como localização.*

Admitamos que

$$\begin{aligned} \underset{\sim}{V}_1 &= (V_{11}, V_{12}, \ldots, V_{1n_1}) \\ \underset{\sim}{V}_2 &= (V_{21}, V_{22}, \ldots, V_{2n_2}) \\ &\vdots \\ \underset{\sim}{V}_k &= (V_{k1}, V_{k2}, \ldots, V_{kn_k}) \end{aligned}$$

são k amostras aleatórias independentes, onde

$$V_{ij} \frown F_{2,\alpha}(\delta_i), \quad j = 1, 2, \ldots, n_i;\ i = 1, 2, \ldots, k.$$

Denotemos igualmente por $N = \sum\limits_{i=1}^{k} n_i$, $V_{1:N}$ *o mínimo da amostra combinada e* $\alpha^* = 1/\alpha$.

Supondo verdadeira a hipótese

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_k = \delta$$

os estimadores de máxima verosimilhança de δ *e* α^* *baseados na amostra combinada são* $\hat{\delta} = -V_{1:N}$ *e*

$$\widehat{\alpha^*} = -\frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln\left(\frac{V_{ij}}{V_{1:N}}\right) = -\frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln\left(\frac{V_{j[i]:n_i}}{V_{1:N}}\right)$$

com $V_{j[i]:n_i}$ *a denotar a* j*-ésima e.o. ascendente da amostra* i.

Assim,

$$\begin{aligned}
\widehat{\alpha^*} &= -\frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln\left(\frac{V_{j[i]:n_i}}{V_{1:N}}\right) \\
&= -\frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln\left(\frac{V_{j[i]:n_i}}{V_{1[i]:n_i}}\right) - \frac{1}{N} \sum_{i=1}^{k} n_i \ln\left(\frac{V_{1[i]:n_i}}{V_{1:N}}\right) \\
&= \widehat{\alpha^*_D} + \widehat{\alpha^*_E}
\end{aligned}$$

com $\widehat{\alpha^*_D}$ *e* $\widehat{\alpha^*_E}$ *a terem uma interpretação idêntica ao do caso Pareto tipo I.*

Ora

$$\widehat{\alpha^*_D} = -\frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln\left(\frac{V_{j[i]:n_i}}{V_{1[i]:n_i}}\right) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=2}^{n_i} (n_i - j + 1) \left[-\ln\left(\frac{V_{j[i]:n_i}}{V_{(j-1)[i]:n_i}}\right)\right]. \tag{2.1.39}$$

Atendendo a que

$$-\ln\left(\frac{V_{j[i]:n_i}}{V_{(j-1)[i]:n_i}}\right) \frown Exp\left(\frac{\alpha^*}{n_i - j + 1}\right), \quad j = 1, \ldots, n_i$$

e ao facto de os spacings *multiplicativos com parente Paretiana tipo II serem independentes, segue-se que*

$$\sum_{j=2}^{n_i}(n_i - j + 1)\left[-\ln\left(\frac{V_{j[i]:n_i}}{V_{(j-1)[i]:n_i}}\right)\right] \frown G(n_i - 1, \alpha^*), \quad i = 1, \ldots, k.$$

Consequentemente,

$$\widehat{\alpha_D^*} \frown G\left(N - k, \frac{\alpha^*}{N}\right). \tag{2.1.40}$$

(note-se também que independentemente da validade de H_0, $\widehat{\alpha_D^*} \frown G(N-k, \alpha^*/N)$)

Por outro lado,

$$\begin{aligned}
\widehat{\alpha_E^*} &= -\frac{1}{N}\sum_{i=1}^{k} n_i \ln\left(\frac{V_{1[i]:n_i}}{V_{1:N}}\right) \\
&= -\frac{1}{N}\left(\sum_{i=1}^{k} n_i \ln(-V_{1[i]:n_i}) - N\ln(-V_{1:N})\right) \\
&= -\frac{1}{N}\left(\sum_{i=1}^{k} \ln(-V_{1[i]:n_i})^{n_i} - \ln(-V_{1:N})^N\right).
\end{aligned}$$

Como os quocientes $\frac{V_{1[i]:n_i}}{V_{1:N}}$ são distribucionalmente independentes de δ, consideremos, sem perda de generalidade, $\delta = 1$. Assim, sob a hipótese nula

$$(-V_{1[i]:n_i})^{n_i} \frown \mathcal{B}e(\alpha, 1) \quad e \quad (-V_{1:N})^N \frown \mathcal{B}e(\alpha, 1)$$

pelo que podemos exprimir $\widehat{\alpha_E^*}$ em função de combinações lineares de spacings *multiplicativos de k v.a. i.i.d. com parente Paretiana tipo II com índice α.*

Por conseguinte,

$$\widehat{\alpha_E^*} = -\frac{1}{N}\sum_{i=1}^{k} n_i \ln\left(\frac{V_{1[i]:n_i}}{V_{1:N}}\right) \frown G\left(k-1, \frac{\alpha^*}{N}\right). \qquad (2.1.41)$$

A independência dos spacings *multiplicativos com parente Paretiana tipo II também implica a decomposição do estimador global $\widehat{\alpha^*}$ com $2(N-1)$ g.l. em dois estimadores independentes, $\widehat{\alpha_E^*}$ com $2(k-1)$ g.l. e $\widehat{\alpha_D^*}$ com $2(N-k)$ g.l..*

Da mesma forma, se dividirmos aqueles estimadores pelos respectivos g.l., obtemos para estatística de teste

$$\frac{\widehat{\alpha_E^*}\big/(k-1)}{\widehat{\alpha_D^*}\big/(N-k)} = \frac{\sum_{i=1}^{k} n_i \ln\left(\frac{V_{1[i]:n_i}}{V_{1:N}}\right)\Big/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i} \ln\left(\frac{V_{ij}}{V_{1[i]:n_i}}\right)\Big/(N-k)} \frown F_{(2(k-1),2(N-k))}. \qquad (2.1.42)$$

Rejeitaremos a hipótese nula ao nível de significância $\gamma\%$ com base nas amostras se

$$\frac{\widehat{\alpha_E^*}\big/(k-1)}{\widehat{\alpha_D^*}\big/(N-k)} \geq F_{1-\gamma;2(k-1),2(N-k)}.$$

Repare-se que as estatísticas (2.1.30) *e* (2.1.42) *estão definidas da mesma forma e têm a mesma distribuição amostral.*

O lema 2.1.1 permitem-nos ainda obter a distribuição de um produto de certas betas independentes.

Lema 2.1.3 *Se $X_1, \ldots, X_n$ forem n v.a. independentes tais que $X_i \frown \mathcal{B}e(\alpha_i, 1, 0, \delta_i)$, onde $\alpha_i \neq \alpha_j$ se $i \neq j$, então a v.a. $P_n = \prod_{i=1}^{n} X_i$ tem f.d.p.*

$$f_{P_n}(x) = \sum_{i=1}^{n} C_{i,n}\frac{\alpha_i}{\delta}\left(\frac{x}{\delta}\right)^{\alpha_i - 1}, \quad 0 < x < \delta$$

onde $\delta = \prod_{j=1}^{n} \delta_j$ e $C_{i,n} = \prod_{\substack{j=1\\ j\neq i}}^{n} \frac{\alpha_j}{\alpha_j - \alpha_i}$.

Dem.

Se $V_1, \ldots, V_n$ forem n v.a. independentes tais que $V_i \frown F_{2,\alpha_i}(\delta_i)$, tem-se

$$X_i \stackrel{d}{=} -V_i, \quad i = 1, \ldots, n$$

e como

$$-V_i \stackrel{d}{=} \delta_i e^{-Z_i/\alpha_i}, \quad i = 1, \ldots, n$$

com $Z_1, \ldots, Z_n$ v.a. exponenciais standard i.i.d., vem

$$P_n = \prod_{i=1}^{n} X_i \stackrel{d}{=} \delta \exp\left(-\sum_{i=1}^{n} \frac{Z_i}{\alpha_i} \right)$$

onde $\delta = \prod\limits_{j=1}^{n} \delta_j$.

Como a f.d.p. de $S_n = \sum\limits_{i=1}^{n} \frac{Z_i}{\alpha_i}$ é a convolução generalizada de n exponenciais independentes, segue do lema 2.1.1 que

$$f_{S_n}(s) = \sum_{i=1}^{n} C_{i,n} \alpha_i e^{-\alpha_i s}, \quad s > 0$$

onde $C_{i,n} = \prod\limits_{\substack{j=1 \\ j \neq i}}^{n} \frac{\alpha_j}{\alpha_j - \alpha_i}$. Assim, obtemos para f.d.p. de P_n

$$f_{P_n}(x) = \sum_{i=1}^{n} C_{i,n} \frac{\alpha_i}{\delta} \left(\frac{x}{\delta} \right)^{\alpha_i - 1}, \quad 0 < x < \delta.$$

2.2 Spacings de Misturas de Exponenciais

Frequentemente variações de uma distribuição "pura"devem ser consideradas para modelar situações reais. Por exemplo, considere-se uma fila de espera onde m tipos de serviço estão à disposição dos clientes que apenas optam por um tipo. Se o tempo necessário para completar um serviço do tipo k for exponencial com valor médio $1/\theta_k$, *e que o cliente opta com probabilidade p_k, então o tempo de serviço (independentemente do tipo escolhido) tem função de distribuição*

$$F(t) = \sum_{k=1}^{m} p_k(1 - e^{-\theta_k t}), \quad t > 0$$

ou seja, é uma mistura de distribuições exponenciais.

*Um forma de encarar o problema de misturas é pensarmos que temos uma população constituída por m subpopulações $\Pi_1, \Pi_2, \ldots, \Pi_m$ nas proporções $p_1, p_2, \ldots, p_m$ ($p_1 + \cdots + p_m = 1$). As misturas do tipo anterior são misturas finitas e na literatura da especialidade (*e.g.*, teoria de filas de espera, fiabilidade) são usualmente denominadas* hiperexponenciais.

As propriedades aritméticas de misturas de distribuições têm sido objecto de estudo de alguns estatísticos. Goldie (1967) provou que um produto de v.a. não negativas e independentes é I.D. se uma delas estiver distribuída exponencialmente; por outras palavras, provou que uma mistura de exponenciais é I.D.. Este resultado fornece-nos outro critério para a divisibilidade infinita, apesar de não ser dos mais usados.

Steutel (1967) generaliza o resultado de Goldie ao determinar uma condição suficiente para uma mistura generalizada de exponenciais, isto é, uma mistura que admita p_k negativos mas sujeitos à condição $\sum\limits_{k=1}^{m} p_k = 1$, ser I.D.; mais precisamente, que na sucessão dos p_k ordenados segundo os θ_k não haja mais de uma mudança de sinal.

A importância dos spacings *na caracterização do modelo exponencial levou-nos a estudar os* spacings *de misturas de exponenciais. Os resultados obtidos poderão estender-se aos* spacings *multiplicativos de misturas de Pareto generalizadas.*

Consideremos então

$$Z = \begin{cases} X_1 & X_2 & \cdots & X_m \\ p_1 & p_2 & \cdots & p_m \end{cases} \tag{2.2.1}$$

em que $\sum\limits_{k=1}^{m} p_k = 1$, $p_k > 0$ e $X_k \frown Exp(\delta_k)$.

Como a variável aleatória Z é uma mistura de v.a. exponenciais, tem f.d.

$$F_Z(z) = \sum_{k=1}^{m} p_k \left(1 - e^{-z/\delta_k}\right), \quad z > 0 \tag{2.2.2}$$

e f.d.p.

$$f_Z(z) = \sum_{k=1}^{m} p_k \, \frac{1}{\delta_k} \, e^{-z/\delta_k}, \quad z > 0. \tag{2.2.3}$$

Seja $(Z_1, \ldots, Z_n)$ *uma a.a. proveniente de uma população com distribuição* (2.2.2) *e seja* $(Z_{1:n}, \ldots, Z_{n:n})$ *o vector das e.o. ascendentes associado à amostra. Uma vez que estamos interessados nos* spacings *da mistura, denotemos por*

$$U_{i,n;m} = Z_{i:n} - Z_{i-1:n}, \quad i = 1, \ldots, n$$

com a convenção $Z_{0:n} \equiv 0$.

A f.d.p. de $U_{i,n;m}$ *é dada por*

$$f_{U_{i,n;m}}(u) = \int_{-\infty}^{\infty} f_{i-1,i:n}(x, x+u) dx, \quad u > 0$$

com $f_{i-1,i:n}(.,.)$ *a f.d.p. conjunta de* $(Z_{i-1:n}, Z_{i:n})$.

Assim sendo,

$$\begin{aligned} f_{U_{i,n;m}}(u) = C_{i,n} \int_0^{\infty} & \left(1 - \sum_{k=1}^{m} p_k \, e^{-x/\delta_k}\right)^{i-2} \left(\sum_{k=1}^{m} p_k \, e^{-x/\delta_k} e^{-u/\delta_k}\right)^{n-i} \\ & \times \left(\sum_{k=1}^{m} \frac{p_k}{\delta_k} \, e^{-x/\delta_k}\right) \left(\sum_{k=1}^{m} \frac{p_k}{\delta_k} \, e^{-x/\delta_k} e^{-u/\delta_k}\right) dx \end{aligned} \tag{2.2.4}$$

onde $C_{i,n} = \frac{n!}{(i-2)!(n-i)!}$.

Dado que

1.

$$\left(1-\sum_{k=1}^{m} p_k\, e^{-x/\delta_k}\right)^{i-2} = \sum_{t=0}^{i-2} \binom{i-2}{t} (-1)^t \sum_{j_1+\,\cdots\,+j_m=t} \frac{t!}{j_1!\cdots j_m!}\, p_1^{j_1}\cdots p_m^{j_m}$$
$$\times\,\exp\left[-\left(\frac{j_1}{\delta_1}+\cdots+\frac{j_m}{\delta_m}\right)x\right]$$

2.

$$\left(\sum_{k=1}^{m} p_k\, e^{-x/\delta_k} e^{-u/\delta_k}\right)^{n-i} = \sum_{k_1+\cdots+k_m=\,n-i} \frac{(n-i)!}{k_1!\ldots k_m!}\, p_1^{k_1}\cdots p_m^{k_m}$$
$$\times\,\exp\left[-\left(\frac{k_1}{\delta_1}+\cdots+\frac{k_m}{\delta_m}\right)x\right]\exp\left[-\left(\frac{k_1}{\delta_1}+\cdots+\frac{k_m}{\delta_m}\right)u\right]$$

3.

$$\left(\sum_{k=1}^{m} \frac{p_k}{\delta_k}\, e^{-x/\delta_k}\right)\left(\sum_{k=1}^{m} \frac{p_k}{\delta_k}\, e^{-x/\delta_k} e^{-u/\delta_k}\right) = \sum_{k=1}^{m}\sum_{j=1}^{m} \frac{p_k\, p_j}{\delta_k\, \delta_j}\, \exp\left[-\left(\frac{1}{\delta_k}+\frac{1}{\delta_j}\right)x\right] e^{-u/\delta_j}$$

obtemos após multiplicar as expressões anteriores e integrando em ordem a x (omitimos a constante $C_{i,n}$)

$$
\begin{aligned}
f_{U_{i,n;m}}(u) \propto & \sum_{k_1+\cdots+k_m=n-i} \frac{(n-i)!}{k_1!\cdots k_m!} \frac{\delta_2\,\delta_3\cdots\delta_m}{\delta_2\,\delta_3\cdots\delta_m(k_1+1)+\delta_1\,\delta_3\cdots\delta_m\,k_2+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}\,k_m} \\
& \times\left(\frac{k_1+1}{\delta_1}+\frac{k_2}{\delta_2}+\cdots+\frac{k_m}{\delta_m}\right)\exp\left[-\left(\frac{k_1+1}{\delta_1}+\frac{k_2}{\delta_2}+\cdots+\frac{k_m}{\delta_m}\right)u\right] \\
& \times\sum_{t=0}^{i-2}\binom{i-2}{t}(-1)^t\sum_{j_1+\cdots+j_m=t}\frac{t!}{j_1!\cdots j_m!} \\
& \times\left\{\frac{\delta_2\,\delta_3\cdots\delta_m\,p_1^{k_1+j_1+2}p_2^{k_2+j_2}\cdots p_m^{k_m+j_m}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1+2)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m)}\right. \\
& +\frac{\delta_1\,\delta_3\cdots\delta_m\,p_1^{k_1+j_1+1}p_2^{k_2+j_2+1}\cdots p_m^{k_m+j_m}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1+1)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2+1)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m)} \\
& +\cdots+ \\
& \left.+\frac{\delta_1\,\delta_2\cdots\delta_{m-1}\,p_1^{k_1+j_1+1}p_2^{k_2+j_2}\cdots p_m^{k_m+j_m+1}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1+1)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m+1)}\right\} \\
& +\cdots+ \\
& +\sum_{k_1+\cdots+k_m=n-i} \frac{(n-i)!}{k_1!\cdots k_m!} \frac{\delta_1\,\delta_2\cdots\delta_{m-1}}{\delta_2\,\delta_3\cdots\delta_m k_1+\delta_1\,\delta_3\cdots\delta_m\,k_2+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+1)} \\
& \times\left(\frac{k_1}{\delta_1}+\frac{k_2}{\delta_2}+\cdots+\frac{k_m+1}{\delta_m}\right)\exp\left[-\left(\frac{k_1}{\delta_1}+\frac{k_2}{\delta_2}+\cdots+\frac{k_m+1}{\delta_m}\right)u\right] \\
& \times\sum_{t=0}^{i-2}\binom{i-2}{t}(-1)^t\sum_{j_1+\cdots+j_m=t}\frac{t!}{j_1!\cdots j_m!} \\
& \times\left\{\frac{\delta_2\,\delta_3\cdots\delta_m\,p_1^{k_1+j_1+1}p_2^{k_2+j_2}\cdots p_m^{k_m+j_m+1}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1+1)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m+1)}\right. \\
& +\frac{\delta_1\,\delta_3\cdots\delta_m\,p_1^{k_1+j_1}p_2^{k_2+j_2+1}\cdots p_m^{k_m+j_m+1}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2+1)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m+1)} \\
& +\cdots+ \\
& \left.+\frac{\delta_1\,\delta_2\cdots\delta_{m-1}\,p_1^{k_1+j_1}p_2^{k_2+j_2}\cdots p_m^{k_m+j_m+2}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m+2)}\right\}.
\end{aligned}
$$

Simplificando a expressão anterior, obtemos para f.d.p. do i-*ésimo* spacing *de uma mistura de exponenciais*

$$\begin{aligned}
f_{U_{i,n;m}}(u) = {} & \frac{n!}{(i-2)!} \sum_{k_1+\cdots+k_m=n-i+1} \frac{1}{k_1!\cdots k_m!} \sum_{t=0}^{i-2} \binom{i-2}{t} (-1)^t \sum_{j_1+\cdots+j_m=t} \frac{t!}{j_1!\cdots j_m!} \\
& \times \left\{ \frac{\delta_2\,\delta_3\cdots\delta_m\,p_1^{k_1+j_1+1}\,p_2^{k_2+j_2}\cdots p_m^{k_m+j_m}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1+1)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m)} \right. \\
& + \frac{\delta_1\,\delta_3\cdots\delta_m\,p_1^{k_1+j_1}\,p_2^{k_2+j_2+1}\cdots p_m^{k_m+j_m}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2+1)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m)} \\
& + \cdots + \\
& \left. + \frac{\delta_1\,\delta_2\cdots\delta_{m-1}\,p_1^{k_1+j_1}\,p_2^{k_2+j_2}\cdots p_m^{k_m+j_m+1}}{\delta_2\,\delta_3\cdots\delta_m(k_1+j_1)+\delta_1\,\delta_3\cdots\delta_m(k_2+j_2)+\cdots+\delta_1\,\delta_2\cdots\delta_{m-1}(k_m+j_m+1)} \right\} \\
& \times \left(\frac{k_1}{\delta_1}+\frac{k_2}{\delta_2}+\cdots+\frac{k_m}{\delta_m}\right) \exp\left[-\left(\frac{k_1}{\delta_1}+\frac{k_2}{\delta_2}+\cdots+\frac{k_m}{\delta_m}\right)u\right], \quad u>0.
\end{aligned} \tag{2.2.5}$$

Como podemos verificar, não é possível obter a partir de (2.2.5) *a f.d.p. de* $U_{1,n;m}$ *uma vez que* $(-1)! = \Gamma(0)$ *e a função gama não está definida para inteiros não positivos. No entanto, se atendermos a*

$$f_{U_{1,n;m}}(u) = n[1-F_Z(u)]^{n-1} f_Z(u)$$

segue-se após algumas manipulações algébricas que

$$\begin{aligned}
f_{U_{1,n;m}}(u) = {} & \sum_{k_1+\cdots+k_m=n} \frac{n!}{k_1!\cdots k_m!}\, p_1^{k_1}\cdots p_m^{k_m} \\
& \times \left(\frac{k_1}{\delta_1}+\cdots+\frac{k_m}{\delta_m}\right) \exp\left[-\left(\frac{k_1}{\delta_1}+\cdots+\frac{k_m}{\delta_m}\right)u\right], \quad u>0.
\end{aligned} \tag{2.2.6}$$

Assim, chegamos a duas conclusões interessantes: a primeira é que os spacings *de misturas de exponenciais são ainda misturas de exponenciais, e por conseguinte, são I.D. (cf. Goldie (1967)); a segunda é que as proporções de mistura do mínimo de uma mistura de exponenciais são os coeficientes multinomiais de* $(p_1+\cdots+p_m)^n$.

Se as misturas forem de duas componentes, i.e.*, se* $m = 2$*, vem de* (2.2.5)

$$f_{U_{i,n;2}}(u) = \frac{n!}{(i-2)!(n-i+1)!} \sum_{k=0}^{n-i+1} \binom{n-i+1}{k} \sum_{t=0}^{i-2} \binom{i-2}{t} (-1)^t \sum_{j=0}^{t} \binom{t}{j}$$
$$\times \left\{ \frac{\delta_2 \, p_1^{n-i+1-(k+j)+t+1} \, p_2^{k+j}}{\delta_2[n-i+1-(k+j)+t+1] + \delta_1(k+j)} + \frac{\delta_1 \, p_1^{n-i+1-(k+j)+t} \, p_2^{k+j+1}}{\delta_2[n-i+1-(k+j)+t] + \delta_1(k+j+1)} \right\}$$
$$\times \left(\frac{n-i+1-k}{\delta_1} + \frac{k}{\delta_2}\right) \exp\left[-\left(\frac{n-i+1-k}{\delta_1} + \frac{k}{\delta_2}\right)u\right], \quad u > 0 \tag{2.2.7}$$

onde $p_1 + p_2 = 1$ *e* $i = 2, \ldots, n$*; enquanto de* (2.2.6) *é imediato que*

$$f_{U_{1,n;2}}(u) = \sum_{k=0}^{n} \binom{n}{k} p_1^{n-k} \, p_2^k \left(\frac{n-k}{\delta_1} + \frac{k}{\delta_2}\right) \exp\left[-\left(\frac{n-k}{\delta_1} + \frac{k}{\delta_2}\right)u\right], \quad u > 0. \tag{2.2.8}$$

Assim, se $Z = \begin{cases} X_1 & X_2 \\ 1/5 & 4/5 \end{cases}$*, com* $\delta_1 = 1$ *e* $\delta_2 = 2$*, tem-se por exemplo*

$$f_{U_{3,5;2}}(u) = \frac{241}{21875}\, e^{-3u} + \frac{76}{525}\, e^{-2.5u} + \frac{2784}{4375}\, e^{-2u} + \frac{20352}{21875}\, e^{-1.5u}, \;\; u > 0.$$

As expressões (2.2.5) *e* (2.2.6) *são ainda válidas se em vez de considerarmos uma mistura de exponenciais, considerarmos uma mistura generalizada de exponenciais, ou seja, se alguns* p_k *de* (2.2.1) *forem negativos. Neste caso, e de modo a garantirmos que (2.2.3) seja uma função densidade de probabilidade própria, os parâmetros* p_k *e* δ_k *devem satisfazer determinadas condições. Steutel (1967) obteve um conjunto de condições necessárias, enquanto Bartholomew (1969) e mais tarde Harris, Marchal e Botta (1992)* *obtiveram condições suficientes. No que refere as condições necessárias, se efectuarmos a reparametrização $\theta_k = 1/\delta_k$ e se, sem perda de generalidade, os rearranjarmos de forma a que $\theta_1 < \theta_2 < \cdots < \theta_m$ então deve verificar-se*

$$\sum_{k=1}^{m} p_k\theta_k \geq 0 \;\; e \;\; p_1 > 0.$$

Quanto às condições suficientes indicaremos a condição de Harris et al.. *Se escrevermos a f.d.p. da mistura de exponenciais na forma*

$$f(t) = f^{+}(t) + f^{-}(t) = \sum_{p_k>0} p_k\theta_k e^{-\theta_k t} + \sum_{p_k<0} p_k\theta_k e^{-\theta_k t}$$

a condição é

$$\{\min(\theta_k; p_k < 0) - \theta_1\} \ln\left(\frac{|f^{-}(0)|}{p_1\theta_1}\right) < 0.$$

Assim, considerando novamente o último exemplo, mas onde $p_1 = -1/5$ e $p_2 = 6/5$, obtém-se

$$f_{U_{3,5;2}}(u) = -\frac{141}{21875}\, e^{-3u} + \frac{23}{175}\, e^{-2.5u} - \frac{792}{875}\, e^{-2u} + \frac{46008}{21875}\, e^{-1.5u}, \;\; u > 0.$$

2.3 Spacings Multiplicativos de Misturas de Pareto Tipo I e Pareto Tipo II

Quando procurámos generalizar propriedades da exponencial à Pareto tipo I ou à Pareto tipo II, o parâmetro de forma destas comportava-se como a escala da exponencial. Esta ocorrência é consequência de o parâmetro α encontrar-se a multiplicar nas identidades $Z \stackrel{d}{=} \alpha \ln W$ e $Z \stackrel{d}{=} -\alpha \ln(-V)$ com $Z \frown Exp(1)$, $W \frown F_{1,\alpha}$ e $V \frown F_{2,\alpha}$. Assim, podemos usar as expressões (2.2.5) *e* (2.2.6) *para mostrar que os* spacings *multiplicativos de misturas de Pareto tipo I e de Pareto tipo II são misturas de Pareto tipo I e misturas de potências standard (betas), respectivamente. Esta propriedade mais não é que a consequência de os* spacings *(aditivos) de misturas de exponenciais serem ainda misturas de exponenciais. Todavia, algumas alterações terão que ser efectuadas.*

A parametrização que temos vindo a utilizar ao longo desta dissertação para a distribuição exponencial é

$$X \frown Exp(\delta) \Rightarrow f_X(x) = \frac{1}{\delta} e^{-x/\delta} I_{(0,\infty)}.$$

Mas como nas funções densidade $f_{1,\alpha}(x) = \alpha x^{-\alpha-1} I_{(1,\infty)}$ *e* $f_{2,\alpha}(x) = \alpha(-x)^{\alpha-1} I_{(-1,0)}$ *não surge o factor* $1/\alpha$*, teremos que substituir nas expressões* (2.2.5) *e* (2.2.6) δ_k *por* $1/\alpha_k$*.*

Consequentemente, se em (2.2.1) *os* $X_k \frown F_{1,\alpha_k}$*, segue que o* spacing *multiplicativo* $W_{i,n;m} = \frac{Z_{i:n}}{Z_{i-1:n}}$ *tem f.d.p.*

$$\begin{aligned} f_{W_{i,n;m}}(w) = \frac{n!}{(i-2)!} & \sum_{k_1+\cdots+k_m=n-i+1} \frac{1}{k_1!\cdots k_m!} \sum_{t=0}^{i-2} \binom{i-2}{t} (-1)^t \sum_{j_1+\cdots+j_m=t} \frac{t!}{j_1!\cdots j_m!} \\ & \times \left\{ \frac{\alpha_1\, p_1^{k_1+j_1+1}\, p_2^{k_2+j_2} \cdots p_m^{k_m+j_m}}{\alpha_1(k_1+j_1+1)+\alpha_2(k_2+j_2)+\cdots+\alpha_m(k_m+j_m)} \right. \\ & \quad + \frac{\alpha_2\, p_1^{k_1+j_1}\, p_2^{k_2+j_2+1} \cdots p_m^{k_m+j_m}}{\alpha_1(k_1+j_1)+\alpha_2(k_2+j_2+1)+\cdots+\alpha_m(k_m+j_m)} \\ & \quad + \cdots + \\ & \quad \left. + \frac{\alpha_m\, p_1^{k_1+j_1}\, p_2^{k_2+j_2} \cdots p_m^{k_m+j_m+1}}{\alpha_1(k_1+j_1)+\alpha_2(k_2+j_2)+\cdots+\alpha_m(k_m+j_m+1)} \right\} \\ & \times (\alpha_1 k_1 + \alpha_2 k_2 + \cdots + \alpha_m k_m)\, w^{-(\alpha_1 k_1+\alpha_2 k_2+\cdots+\alpha_m k_m)-1}, \qquad w > 1 \end{aligned} \tag{2.3.1}$$

para $i = 2, \ldots, n$*; enquanto*

$$\begin{aligned} f_{W_{1,n;m}}(w) = & \sum_{k_1+\cdots+k_m=n} \frac{n!}{k_1!\cdots k_m!} p_1^{k_1} \cdots p_m^{k_m} \\ & \times (\alpha_1 k_1 + \cdots + \alpha_m k_m) w^{-(\alpha_1 k_1+\cdots+\alpha_m k_m)-1}, \quad w > 1. \end{aligned} \tag{2.3.2}$$

Repare-se que misturas finitas de Paretos, isto é, misturas com f.d.p. do tipo

$$f(x) = \sum_{k=1}^{m} p_k \alpha_k x^{-\alpha_k-1}, \quad x > 1$$

onde $\alpha_k > 0$, são I.D. por terem também funções densidade de probabilidade completamente monótonas. Assim, como os spacings *multiplicativos de misturas de Pareto são ainda misturas de Pareto, então são I.D..*

Note-se que os resultados que temos vindo a obter para as misturas de Pareto tipo I e de Pareto tipo II pressupõem escalas iguais para as componentes.

A generalização dos resultados (2.3.1) *e* (2.3.2) *em termos de* spacings *multiplicativos de misturas de Pareto tipo II é agora imediata. Porém, não podemos afirmar que essas misturas são I.D.. No entanto, se X for uma v.a. com f.d.p.*

$$f_X(x) = \sum_{k=1}^{n} p_k \alpha_k (-x)^{\alpha_k - 1}, -1 < x < 0$$

isto é, se X for uma mistura de v.a. $X_k \frown F_{2,\alpha_k}$, então, por analogia com o caso simples, $\ln(-X)$ *deve ser I.D.. De facto, verifica-se que* $\ln(-X)$ *é uma mistura de exponenciais negativas, sendo I.D. por* $-\ln(-X)$ *ser uma mistura de exponenciais.*

Os spacings *aditivos de misturas de Pareto generalizadas são de tal modo complexos, mesmo para duas componentes, que não vemos qualquer utilidade no seu estudo. No entanto, indicamos a sua distribuição para o caso de misturas de duas componentes Paretianas tipo I. A demonstração dos resultados apresentados encontra-se no apêndice.*

Seja

$$Z = \begin{cases} X_1 & X_2 \\ p & q \end{cases}$$

com $X_1 \frown F_{1,\alpha}(\delta_1)$, $X_2 \frown F_{1,\alpha}(\delta_2)$ e $p + q = 1$.

A v.a. Z tem f.d.

$$F_Z(z) = p\left[1 - \left(\frac{z}{\delta_1}\right)^{-\alpha}\right] I_{(\delta_1,\infty)} + q\left[1 - \left(\frac{z}{\delta_2}\right)^{-\alpha}\right] I_{(\delta_2,\infty)}$$

e f.d.p.

$$f_Z(z) = p\frac{\alpha}{\delta_1}\left(\frac{z}{\delta_1}\right)^{-\alpha-1} I_{(\delta_1,\infty)} + q\frac{\alpha}{\delta_2}\left(\frac{z}{\delta_2}\right)^{-\alpha-1} I_{(\delta_2,\infty)}.$$

Sem perda de generalidade, considere-se $\delta_1 < \delta_2$; então,

$$F_Z(z) = \begin{cases} 0 & , \; z \leq \delta_1 \\ p\left[1 - \left(\frac{z}{\delta_1}\right)^{-\alpha}\right] & , \; \delta_1 < z \leq \delta_2 \\ 1 - \theta z^{-\alpha} & , \; z > \delta_2 \end{cases} \tag{2.3.3}$$

e

$$f_Z(z) = \begin{cases} 0 & , \; z \leq \delta_1 \\ p\frac{\alpha}{\delta_1}\left(\frac{z}{\delta_1}\right)^{-\alpha-1} & , \; \delta_1 < z \leq \delta_2 \\ \alpha\theta z^{-\alpha-1} & , \; z > \delta_2 \end{cases} \tag{2.3.4}$$

onde $\theta = p\delta_1^\alpha + q\delta_2^\alpha$.

Seja $(Z_1, Z_2, \ldots, Z_n)$ uma a.a. proveniente de uma população com distribuição (3.3.14)*, e denotemos por*

$$U_{k,n} = Z_{k:n} - Z_{k-1:n}, \quad k = 1, \ldots, n$$

com a convenção $Z_{0:n} = \delta_1$.

Ora o vector aleatório $(Z_{k-1:n}, Z_{k:n})$ tem f.d.p. conjunta

$$f_{k-1,k:n}(x,y) =$$
$$= \begin{cases} C_{k,n}\alpha^2\delta_1^{2\alpha}p^k\left[1 - \left(\frac{x}{\delta_1}\right)^{-\alpha}\right]^{k-2}\left[q + p\left(\frac{y}{\delta_1}\right)^{-\alpha}\right]^{n-k}x^{-\alpha-1}y^{-\alpha-1} & , \; \delta_1 < x < y \leq \delta_2 \\ C_{k,n}\alpha^2\delta_1^{\alpha}p^{k-1}\theta^{n-k+1}\left[1 - \left(\frac{x}{\delta_1}\right)^{-\alpha}\right]^{k-2}x^{-\alpha-1}y^{-\alpha(n-k+1)-1} & , \; \delta_1 < x \leq \delta_2 < y \\ C_{k,n}\alpha^2\theta^{n-k+2}(1 - \theta x^{-\alpha})^{k-2}x^{-\alpha-1}y^{-\alpha(n-k+1)-1} & , \; \delta_2 < x < y \end{cases}$$

onde $C_{k,n} = \frac{n!}{(k-2)!(n-k)!}$.

Como

$$f_{U_{k,n}}(u) = \int_{-\infty}^{\infty} f_{k-1,k:n}(x, x+u)dx, \quad u > 0$$

estabelece-se que

(i) se $0 < u \leq \delta_2 - \delta_1$,

$$
\begin{aligned}
f_{U_{k,n}}(u) = {} & \frac{\alpha^2 n!}{(k-2)!(n-k)!} \sum_{i=0}^{k-2} \binom{k-2}{i} (-1)^i \left\{ \sum_{j=0}^{n-k} \binom{n-k}{j} \frac{p^{n-j} q^j}{\delta_1 [\alpha(n-k-j+i+2)+1]} \right. \\
& \times \left[{}_2F_1\left(\alpha(n-k-j+1)+1, \alpha(n-k-j+i+2)+1; \alpha(n-k-j+i+2)+2; -\frac{u}{\delta_1} \right) \right. \\
& - \left(\frac{\delta_1}{\delta_2 - u} \right)^{\alpha(n-k-j+i+2)+1} \times \\
& \times {}_2F_1\left. \left(\alpha(n-k-j+1)+1, \alpha(n-k-j+i+2)+1; \alpha(n-k-j+i+2)+2; -\frac{u}{\delta_2 - u} \right) \right] \\
& + \frac{1}{\alpha(n-k+i+2)+1} \left[\frac{p^{k-1} \theta^{n-k+1} \delta_1^{\alpha(i+1)}}{(\delta_2 - u)^{\alpha(n-k+i+2)+1}} \right. \\
& \times {}_2F_1\left(\alpha(n-k+1)+1), \alpha(n-k+i+2)+1; \alpha(n-k+i+2)+2; -\frac{u}{\delta_2 - u} \right) \\
& + \frac{\theta^{n-k+1} [\theta^{i+1} - p^{k-1} \delta_1^{\alpha(i+1)}]}{\delta_2^{\alpha(n-k+i+2)+1}} \\
& \times {}_2F_1\left. \left. \left(\alpha(n-k+1)+1), \alpha(n-k+i+2)+1; \alpha(n-k+i+2)+2; -\frac{u}{\delta_2} \right) \right] \right\}
\end{aligned}
$$

(ii) se $u > \delta_2 - \delta_1$,

$$
\begin{aligned}
f_{U_{k,n}}(u) = {} & \frac{\alpha^2 n!}{(k-2)!(n-k)!} \sum_{i=0}^{k-2} \binom{k-2}{i} \frac{(-1)^i}{\alpha(n-k+i+2)+1} \left\{ \frac{p^{k-1} \theta^{n-k+1}}{\delta_1^{\alpha(n-k+1)+1}} \right. \\
& \times {}_2F_1\left(\alpha(n-k+1)+1, \alpha(n-k+i+2)+1; \alpha(n-k+i+2)+2; -\frac{u}{\delta_1} \right) \\
& + \frac{\theta^{n-k+1} [\theta^{i+1} - p^{k-1} \delta_1^{\alpha(i+1)}]}{\delta_2^{\alpha(n-k+i+2)+1}} \\
& \times {}_2F_1\left. \left(\alpha(n-k+1)+1, \alpha(n-k+i+2)+1; \alpha(n-k+i+2)+2; -\frac{u}{\delta_2} \right) \right\}
\end{aligned}
$$

com $\theta = p\delta_1^\alpha + q\delta_2^\alpha$ e $k = 2, \ldots, n$.

No que refere a distribuição de $U_{1,n}$, tem-se

$$
f_{U_{1,n}}(u) = \begin{cases} p\frac{n\alpha}{\delta_1} \left[q + p\left(1 + \frac{u}{\delta_1}\right)^{-\alpha} \right]^{n-1} \left(1 + \frac{u}{\delta_1}\right)^{-\alpha-1} & , \ 0 < u \leq \delta_2 - \delta_1 \\ \theta^n n\alpha (u + \delta_1)^{-n\alpha-1} & , \ u > \delta_2 - \delta_1 \end{cases} .
$$

2.4 Spacings do Modelo Laplaceano

No início do Capítulo 2 foi dito que quando se pretende investigar modelos alternativos ao modelo exponencial, tem interesse começar por modelos que com ele estejam directamente relacionados. Ora o modelo de Laplace surge como candidato natural dada a sua expressão simples obtida por simetrização de exponenciais i.i.d.; ou seja, se X_1 *e* X_2 *forem exponenciais i.i.d. com escala* δ*, a v.a.* $Y = X_1 - X_2$ *tem distribuição de Laplace, isto é, tem f.d.p.*

$$f_Y(y) = \frac{1}{2\delta} e^{-\left|\frac{y}{\delta}\right|} I_{IR}.$$

Como podemos observar, a distribuição de Laplace é simétrica (neste caso em torno da origem), e esta propriedade terá implicações directas na f.c., nomeadamente será uma função real e par. Ora se, sem perda de generalidade, $\delta = 1$*, tem-se*

$$\begin{aligned}
\varphi_Y(t) &= E[e^{itY}] = \varphi_{X_1}(t)\varphi_{X_2}(-t) \\
&= \frac{1}{1-it}\frac{1}{1+it} \\
&= \frac{1}{1+t^2}
\end{aligned}$$

confirmando-se assim a asserção anterior.

Por outro lado, como a f.c. admite uma factorização em f.c. I.D., então a Laplace também é I.D..

Nesta secção estudaremos formalmente os spacings *com parente Laplaceana, ao mesmo tempo que desvendaremos algumas das suas propriedades.*

Seja $(X_1, \ldots, X_n)$ *uma a.a. proveniente de uma população Laplaceana com parâmetro de localização* λ *e parâmetro de escala* δ*,* i.e.*, com f.d.p.*

$$f(x) = \frac{1}{2\,\delta} \exp\left(-\left|\frac{x-\lambda}{\delta}\right|\right) I_{IR} \tag{2.4.1}$$

e f.d.

$$F(x) = \frac{1}{2}\exp\left(\frac{x-\lambda}{\delta}\right) I_{(-\infty,\lambda]} + \left[1 - \frac{1}{2}\exp\left(-\frac{x-\lambda}{\delta}\right)\right] I_{(\lambda,\infty)}. \tag{2.4.2}$$

Seja ainda $(X_{1:n},\ldots,X_{n:n})$ *o vector das e.o. ascendentes associado à amostra. Sem perda de generalidade, considere-se o modelo Laplaceano standard,* i.e., $\lambda=0$ *e* $\delta=1$*, e denotemos por*

$$L_{i,n}=X_{i+1:n}-X_{i:n},\quad i=1,\ldots,n-1.$$

Como o vector aleatório $(X_{i:n},X_{i+1:n})$ *tem f.d.p. conjunta*

$$f_{i,i+1:n}(x,y)=\begin{cases}\frac{n!}{(i-1)!(n-i-1)!}\left(\frac{1}{2}\right)^{i+1}e^{ix}e^{y}\left(1-\frac{1}{2}e^{y}\right)^{n-i-1} & ,\ x<y\leq 0\\ [0.1cm]\frac{n!}{(i-1)!(n-i-1)!}\left(\frac{1}{2}\right)^{n}e^{ix}e^{-(n-i)y} & ,\ x\leq 0<y\\ [0.1cm]\frac{n!}{(i-1)!(n-i-1)!}\left(\frac{1}{2}\right)^{n-i+1}\left(1-\frac{1}{2}e^{-x}\right)^{i-1}e^{-x}e^{-(n-i)y} & ,\ 0<x<y\end{cases},$$

a f.d.p. de $L_{i,n}$ *é dada por*

$$f_{L_{i,n}}(v)=\int_{-\infty}^{-v}f_{i,i+1:n}(x,x+v)dx+\int_{-v}^{0}f_{i,i+1:n}(x,x+v)dx$$
$$+\int_{0}^{\infty}f_{i,i+1:n}(x,x+v)dx,\quad v>0.\quad (2.4.3)$$

Considerando $C_{i,n}=\frac{n!}{(i-1)!(n-i-1)!}$*, e calculando separadamente os integrais de* (2.4.3)*, vem:*

(a)

$$\int_{-\infty}^{-v}f_{i,i+1:n}(x,x+v)dx=C_{i,n}\left(\frac{1}{2}\right)^{i+1}e^{v}\int_{-\infty}^{-v}e^{(i+1)x}\left(1-\frac{1}{2}e^{x+v}\right)^{n-i-1}dx,$$

e efectuando a mudança de variável $t=\frac{1}{2}e^{x+v}$*, segue-se que*

$$\begin{aligned}\int_{-\infty}^{-v}f_{i,i+1:n}(x,x+v)dx&=C_{i,n}\left(\frac{1}{2}\right)^{i}\int_{0}^{1/2}(2te^{-v})^{i}(1-t)^{n-i-1}dt\\&=C_{i,n}e^{-iv}\int_{0}^{1/2}t^{i}(1-t)^{n-i-1}dt\\&=C_{i,n}e^{-iv}B_{1/2}(i+1,n-i).\end{aligned}\quad (2.4.4)$$

No que respeita o segundo integral, há que considerar duas situações distintas:

$(b.1)\quad n-2i\neq 0$

$$\begin{aligned}\int_{-v}^{0} f_{i,i+1:n}(x,x+v)dx &= C_{i,n}\left(\frac{1}{2}\right)^{n} e^{-(n-i)v}\int_{-v}^{0} e^{-(n-2i)x}dx \\ &= C_{i,n}\left(\frac{1}{2}\right)^{n}\frac{1}{n-2i}\left[e^{-iv}-e^{-(n-i)v}\right] \end{aligned} \tag{2.4.5}$$

$(b.2)\quad n-2i=0$

$$\int_{-v}^{0} f_{i,i+1:n}(x,x+v)dx = C_{i,n}\left(\frac{1}{2}\right)^{n} v e^{-nv/2}. \tag{2.4.6}$$

OBSERVAÇÃO*: A condição* $n-2i=0$ *ocorre somente se* n *par (*$i=\frac{n}{2}$*), ou seja, quando se considera o* spacing *central* $L_{\frac{n}{2},n}$*. Neste caso, obteremos uma situação diferente de os restantes* spacings *e de que falaremos mais adiante.*

(c)

$$\int_{0}^{\infty} f_{i,i+1:n}(x,x+v)dx = C_{i,n}\left(\frac{1}{2}\right)^{n-i+1} e^{-(n-i)v}\int_{0}^{\infty} e^{-(n-i+1)x}\left(1-\frac{1}{2}e^{-x}\right)^{i-1}dx,$$

e se procedermos à transformação $t=\frac{1}{2}e^{-x}$*, vem*

$$\begin{aligned}\int_{0}^{\infty} f_{i,i+1:n}(x,x+v)dx &= C_{i,n}\left(\frac{1}{2}\right)^{n-i} e^{-(n-i)v}\int_{0}^{1/2}(2t)^{n-i}(1-t)^{i-1}dt \\ &= C_{i,n}e^{-(n-i)v}\int_{0}^{1/2} t^{n-i}(1-t)^{i-1}dt \\ &= C_{i,n}e^{-(n-i)v}B_{1/2}(n-i+1,i). \end{aligned} \tag{2.4.7}$$

Substituindo as expressões (2.4.4)*,* (2.4.5) *e* (2.4.7) *em* (2.4.3)*, vem*

$$\begin{aligned} f_{L_{i,n}}(v) &= C_{i,n}\,e^{-iv}\,B_{1/2}(i+1,n-i) + C_{i,n}\left(\frac{1}{2}\right)^{n}\frac{1}{n-2i}\left[e^{-iv}-e^{-(n-i)v}\right] \\ &\quad + C_{i,n}\,e^{-(n-i)v}B_{1/2}(n-i+1,i). \end{aligned} \tag{2.4.8}$$

Atendendo a (0.0.18) *e* (0.0.19)*, segue-se que*

$$f_{L_{i,n}}(v) = ie^{-iv}\left\{\sum_{k=i+1}^{n}\binom{n}{k}\left(\frac{1}{2}\right)^n + \frac{(i+1)}{n-2i}\binom{n}{i+1}\left(\frac{1}{2}\right)^n\right\}$$
$$+ (n-i)e^{-(n-i)v}\left\{\sum_{k=n-i+1}^{n}\binom{n}{k}\left(\frac{1}{2}\right)^n - \frac{i}{n-2i}\binom{n}{i}\left(\frac{1}{2}\right)^n\right\}$$

ou

$$f_{L_{i,n}}(v) = ie^{-iv}\left\{\sum_{k=i+1}^{n}\binom{n}{k}\left(\frac{1}{2}\right)^n + \frac{n-i}{n-2i}\binom{n}{i}\left(\frac{1}{2}\right)^n\right\}$$
$$+ (n-i)e^{-(n-i)v}\left\{\sum_{k=0}^{i-1}\binom{n}{k}\left(\frac{1}{2}\right)^n - \frac{i}{n-2i}\binom{n}{i}\left(\frac{1}{2}\right)^n\right\}. \quad (2.4.9)$$

Se R for uma v.a. tal que $R \frown b\,(n, 0.5)$, então a expressão anterior pode escrever-se na forma

$$f_{L_{i,n}}(v) = ie^{-iv}\left[P\{R \geq i+1\} + \frac{n-i}{n-2i}P\{R=i\}\right]$$
$$+ (n-i)e^{-(n-i)v}\left[P\{R \leq i-1\} - \frac{i}{n-2i}P\{R=i\}\right] \quad (2.4.10)$$

com $v > 0$ e $i = 1, \ldots, n-1$.

Quando n é par, o spacing *central $L_{\frac{n}{2},n}$ tem, contrariamente dos restantes* spacings *cuja f.d.p. é dada por* (2.4.10)*, f.d.p.*

$$f_{L_{\frac{n}{2},n}}(v) = \frac{n}{2}e^{-nv/2}\left[2\sum_{k=n/2+1}^{n}\binom{n}{k}\left(\frac{1}{2}\right)^n\right] + \left(\frac{n}{2}\right)^2 ve^{-nv/2}\left[\binom{n}{\frac{n}{2}}\left(\frac{1}{2}\right)^n\right] \quad (2.4.11)$$

ou

$$f_{L_{\frac{n}{2},n}}(v) = \frac{n}{2}e^{-nv/2}\left[1 - P\left\{R = \frac{n}{2}\right\}\right] + \left(\frac{n}{2}\right)^2 ve^{-nv/2}\,P\left\{R = \frac{n}{2}\right\}, \quad v > 0. \quad (2.4.12)$$

Como podemos constatar a partir de (2.4.10)*, os* spacings *do modelo Laplaceano são misturas de exponenciais, e uma vez que as ÔÔproporçõesÕÕ têm sinais contrários, a mistura diz-se generalizada. De facto, se*

$$p_1 = P\{R \geq i+1\} + \frac{n-i}{n-2i}\, P\{R = i\} \;\; e \;\; p_2 = P\{R \leq i-1\} - \frac{i}{n-2i}\, P\{R = i\}$$

tem-se $p_1 > 0$ *e* $p_2 < 0$ *para* $i < \frac{n}{2}$*, enquanto* $p_1 < 0$ *e* $p_2 > 0$ *para* $i > \frac{n}{2}$.

Porém, o spacing *central* $L_{\frac{n}{2},n}$ *(*n *par) não partilha da mesma propriedade, ou seja, não é uma mistura generalizada de exponenciais. Neste caso, temos uma mistura, no sentido clássico, de exponencial e gama com proporções igualmente provenientes do modelo binomial simétrico.*

2.4.1 Propriedades

A simetria da distribuição de Laplace não é reproduzida na distribuição dos spacings. *No entanto, outros tipos de simetria podem observar-se. Por um lado, temos que os coeficientes das misturas provêm do modelo binomial simétrico, e esta propriedade resulta da Laplace poder obter-se por um processo de simetrização de exponenciais i.i.d.. Por outro lado, de* (2.4.10) *conclui-se que*

$$L_{i,n} \stackrel{d}{=} L_{n-i,n}\,, \quad i = 1, \ldots, n-1 \tag{2.4.13}$$

ou seja, spacings *equidistantes a partir do centro têm a mesma distribuição.*

Os spacings *da Laplace não são independentes, mas em contrapartida têm a seu favor o facto de serem misturas generalizadas de exponenciais, excepto no caso de* $L_{\frac{n}{2},n}$ *(*n *par) que é uma mistura de exponencial e gama. Por isso, a divisibilidade infinita do modelo de Laplace é extensível aos* spacings *do modelo. De facto, os* spacings *com f.d.p.* (2.4.10) *são I.D. por serem misturas generalizadas de duas exponenciais (cf. Steutel (1967), corolário 2). Para justificar a divisibilidade infinita de* $L_{\frac{n}{2},n}$ *não podemos usar os argumentos anteriores, mas podemos nos apoiar no lema que se segue.*

Lema 2.4.1 *Uma f.c.* ψ *é I.D. se* $\ln\psi$ *puder escrever-se na forma*

$$\ln\psi(t) = \gamma i t + \int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \theta(x) dx \tag{1}$$

onde $\gamma \in IR$, $\theta(x) \geq 0$ *e* $[x^2/(1+x^2)]\theta(x)$ *integrável em* $(0,\infty)$.

(o lema 2.4.1 é um caso particular da representação de Lévy-Khinchine para funções características I.D..)

OBSERVAÇÃO*: Segundo a notação do lema 2.4.1, tem-se para a distribuição exponencial com f.d.p.* $f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}$, $\gamma = \int_0^\infty (1+x^2)^{-1} e^{-\lambda x} dx$ *e* $\theta(x) = x^{-1} e^{-\lambda x}$.

Se n *for par,* $L_{\frac{n}{2},n}$ *tem f.c.*

$$\psi(t) = p\,\frac{n/2}{n/2 - it} + (1-p)\left(\frac{n/2}{n/2 - it}\right)^2$$

onde $p = 1 - P\{R = n/2\}$. *A expressão anterior admite por sua vez a factorização*

$$\psi(t) = \left(\frac{n/2}{n/2 - it}\right)^2 \left(\frac{n/2p - it}{n/2p}\right).$$

Aplicando logaritmos a ambos os membros, vem

$$\ln \psi(t) = 2 \ln\left(\frac{n/2}{n/2 - it}\right) - \ln\left(\frac{n/2p}{n/2p - it}\right).$$

Como no segundo membro da expressão anterior surgem logaritmos de f.c. de exponenciais, podemos usar o lema 2.4.1 aplicado a essas exponenciais.

Assim,

$$\begin{aligned}
\ln\left(\frac{n/2}{n/2 - it}\right) &= \gamma_1 it + \int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\theta_1(x)dx \\
\ln\left(\frac{n/2p}{n/2p - it}\right) &= \gamma_2 it + \int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\theta_2(x)dx
\end{aligned}$$

onde γ_1 *e* γ_2 *são números reais,* $\theta_1(x) = x^{-1} e^{-nx/2}$ *e* $\theta_2(x) = x^{-1} e^{-nx/2p}$.

Consequentemente,

$$\ln \psi(t) = (2\gamma_1 - \gamma_2)it + \int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)\theta(x)dx$$

onde $\theta(x) = 2\theta_1(x) - \theta_2(x)$.

Como $2\gamma_1 - \gamma_2$ *é um número real,* $\theta(x) \geq 0$, *pois* $p \geq 0.5$, *e* $[x^2/(1+x^2)]\theta(x)$ *é integrável em* $(0, \infty)$, *segue do lema 2.4.1 que a f.c.* ψ *é I.D.; ou seja, o* spacing *central de uma amostra de dimensão par proveniente de uma população Laplaceana também é I.D..*

Steutel (1968) provou que misturas com f.c.

$$\phi(t) = \int_0^\infty \frac{\lambda}{\lambda - h(t)} dF(\lambda), \quad \lambda > 0$$

são I.D. se $h(t)$ *for o logaritmo de uma f.c. I.D.. Este resultado permite afirmar que a divisibilidade infinita é preservada sob mistura de Laplaces (a f.c. da Laplace com parâmetro de escala* $\beta = 1/\delta$ *é*

$$\varphi(t) = \frac{\beta^2}{\beta^2 + t^2}$$

e se considerarmos $h(t) = -t^2 = \ln e^{-t^2}$, *temos que* e^{-t^2} *é a f.c. de uma Gaussiana com valor médio 0 e variância 2). O mesmo artigo permite ainda concluir que no que respeita a distribuição geométrica, contrapartida discreta da exponencial, a divisibilidade infinita é também preservada sob mistura.*

Capítulo 3

Exponencialidade versus Pareto Generalizada – uma estatística de teste resistente

É muito pouco provável que se saiba a priori *qual a classe de distribuições da Pareto generalizada $F_{1,\alpha}$, $F_{2,\alpha}$ ou F_3 que está subjacente às observações. Se for este o caso, há que testar hipóteses sobre o índice i ($i = 1, 2, 3$), e uma forma de fazê-lo é ter em conta a parametrização de von Mises-Jenkinson para a distribuição de Pareto generalizada; ou seja,*

$$F_\beta(x) = 1 - \left(1 + \beta\,\frac{x}{\delta}\right)^{-1/\beta}, \quad 1 + \beta\,\frac{x}{\delta} > 0, \quad x > 0 \tag{3.0.1}$$

onde δ é um parâmetro de escala e $\beta \in IR$ um parâmetro de forma.

Repare-se que testar hipóteses sobre o índice i equivale na parametrização de von Mises-Jenkinson a testar hipóteses sobre o parâmetro de forma β. De facto, se $\beta > 0$, $F_\beta\left(\frac{\delta(x-1)}{\beta}\right)$ é a distribuição de Pareto clássica com índice $\alpha = 1/\beta$, enquanto se $\beta < 0$, $F_\beta\left(\frac{\delta(x+1)}{|\beta|}\right)$ é a distribuição de Pareto tipo II com índice $\alpha = 1/|\beta|$. Note-se ainda que quando $\beta \to 0$, $\left(1 + \beta\frac{x}{\delta}\right)^{-1/\beta} \longrightarrow e^{-x/\delta}$, justificando-se assim o facto da exponencial ser um caso limite da Pareto generalizada para $\beta = 0$. Assim, se testarmos a hipótese $H_0 : \beta = 0$, estaremos a testar o comportamento exponencial dos dados.

Algumas estatísticas têm sido propostas para testar exponencialidade versus Pareto generalizada (cf. van Monfort e Witter (1985) e Gomes e van Monfort (1987)). Todavia, as estatísticas apresentadas são bastante sensíveis (ou pouco resistentes) a observações *perturbadoras. Como indicador da resistência ou sensibilidade de uma estatística face à presença de* outliers *na amostra temos o* limite de ruptura *de Hampel:*

Definição 3.0.1 *Uma estatística T tem limite de ruptura $\alpha \times 100\%$ se a fracção de observações da amostra que puderem crescer sem T divergir tender para α.*

A estatística que propomos para testar exponencialidade versus Pareto generalizada baseia-se na estatística $\frac{X_{n:n}-M}{M-X_{1:n}}$, onde $X_{n:n}$, M e $X_{1:n}$ são, respectivamente, o máximo, a mediana e o mínimo de uma a.a. de dimensão n, e que foi usada num contexto similar por Gomes (1982) para discriminar entre modelos extremais[1]*. Como a estatística é função das e.o. extremais máximo e mínimo, possui limite de ruptura 0. Assim, por forma a aumentarmos o limite de ruptura da estatística anterior, propomos a estatística*

$$\frac{F_U - M}{M - F_L} \tag{3.0.2}$$

onde F_U, M e F_L são, respectivamente, o quarto superior, mediana e quarto inferior da amostra[2]*.*

Os quartos fazem parte de um conjunto de letras-resumo que fornecem resumos rápidos de uma colecção de dados e são definidos à custa da noção de profundidade*. Ora cada observação ocupa numa colecção de dados ordenada duas posições a que chamamos ordens (ou* ranks*), uma ordem ascendente a contar do mínimo e uma ordem descendente a contar do máximo. A profundidade de uma observação mais não é que o mínimo das suas ordens ascendente e descendente. Repare-se que se i for a profundidade de uma observação x numa colecção de dimensão n, então se a sua ordem ascendente for i, a sua ordem descendente será $n - i + 1$, sendo a soma das duas ordens, ascendente e descendente, igual a $n + 1$.*

Por exemplo, a mediana é uma letra-resumo que se encontra a meio caminho dos extremos que são letras-resumo com profundidade 1. Assim, a mediana pode ser definida em termos de profundidade, ou seja,

$$\textit{profundidade } M = \frac{n+1}{2}.$$

[1]A estatística foi introduzida pela primeira vez por Gumbel (1965) com intuito de fornecer uma estimação rápida do parâmetro de forma num modelo Fréchet.

[2]A simbologia utilizada para os quartos vem do inglês, Upper-Fourth e Lower-Fourth.

Se n for par, $\frac{n+1}{2}$ envolve a proporção $\frac{1}{2}$, de modo que procedemos por convenção a uma interpolação; ou seja, se $x_1, x_2, \ldots, x_n$ for uma colecção de n dados, temos

$$M = \begin{cases} x_{\frac{n+1}{2}:n} & , \textit{ se } n \textit{ ímpar} \\ \frac{1}{2}\left(x_{\frac{n}{2}:n} + x_{\frac{n}{2}+1:n}\right) & , \textit{ se } n \textit{ par} \end{cases}.$$

Por sua vez, os quartos estão definidos em termos da profundidade da mediana, ou seja,

$$\textit{profundidade } F = \frac{[\textit{Profundidade } M] + 1}{2}$$

onde $[x]$ representa a parte inteira de x, procedendo-se igualmente a uma interpolação no caso de a fracção anterior envolver a proporção $\frac{1}{2}$.

Se, por exemplo, tivermos uma amostra constituída pelas observações $x_1, x_2, x_3, x_4, x_5, x_6, x_7$, então a mediana tem profundidade 4 e os quartos profundidade 2.5. Donde, $M = x_{4:7}$, $F_L = \frac{1}{2}(x_{2:7} + x_{3:7})$ e $F_U = \frac{1}{2}(x_{5:7} + x_{6:7})$.

Por outro lado, os quartos podem intervir em medidas de dispersão simples e resistentes, sendo a mais utilizada a dispersão-quartal *$d_F = F_U - F_L$. A dispersão-quartal em conjunto com os quartos podem ser usados na identificação de* outliers.

Dado que os quartos são na realidade uma forma de quartil, têm limite de ruptura aproximadamente igual a 25%, pelo que a estatística que propomos tem limite de ruptura próximo de 25%.

*Numa população Pareto generalizada, $\frac{F_U - M}{M - F_L} = \frac{2^\beta - 1}{1 - (2/3)^\beta}$, e se $\beta = 0$, então $\frac{F_U - M}{M - F_L} = \frac{\ln 2}{\ln(3/2)}$ (≈ 1.7095). Assim, se a hipótese nula $H_0 : \beta = 0$ (*i.e.*, exponencialidade) for verdadeira, a estatística deve assumir valores próximos de $\frac{\ln 2}{\ln(3/2)}$.*

3.1 Distribuição Exacta

Nesta secção obteremos a distribuição exacta de $\frac{F_U - M}{M - F_L}$ sob a hipótese $H_0 : \beta = 0$, isto é, supondo uma parente exponencial. Sublinhamos a palavra exacta porque na secção seguinte obteremos uma distribuição aproximada para a estatística sob a mesma hipótese *baseada no facto de os quartos serem uma forma de quartil.*

Seja então $(X_1, \ldots, X_n)$ *uma a.a. proveniente de uma população exponencial com parâmetro de escala* δ*,* i.e*, com f.d.p.*

$$f(x) = \frac{1}{\delta}\, e^{-x/\delta} I_{(0,\infty)}$$

e seja $(X_{1:n}, \ldots, X_{n:n})$ *o correspondente vector das e.o. ascendentes.*

A fim de simplificarmos a escrita, seja $T_n = \frac{F_U - M}{M - F_L}$*. Como a estatística* T_n *é independente do parâmetro de escala* δ*, considere-se, sem perda de generalidade,* $\delta = 1$*.*

A definição de mediana e quartos em termos da noção de profundidade obrigam-nos a considerar quatro casos, mais precisamente, as classes de congruência módulo quatro[3]*.*

♦ 1^o *Caso:* n *ímpar*

Se $n = 2k - 1$*, então*

$$\textit{profundidade } M = \frac{n+1}{2} = k$$
$$\textit{profundidade } F = \frac{k+1}{2}.$$

Consequentemente, $M = X_{k:n}$ *e*

$(i.1)$ *se* k *for ímpar [*$n \equiv 1 \pmod 4$*], tem-se*

$$F_L = X_{\frac{k+1}{2}:n} \quad \textit{e} \quad F_U = X_{n-\frac{k+1}{2}+1:n}.$$

Assim sendo,

$$T_n = \frac{X_{n-\frac{k+1}{2}+1:n} - X_{k:n}}{X_{k:n} - X_{\frac{k+1}{2}:n}}. \tag{3.1.1}$$

[3]Recorde-se que $a \equiv b \pmod m$ sse m divide $a - b$.

Decorre de (2.1.11) *que*

$$X_{n-\frac{k+1}{2}+1:n} - X_{k:n} \stackrel{d}{=} X_{n-(\frac{k+1}{2}+k)+1:n-k}$$
$$X_{k:n} - X_{\frac{k+1}{2}:n} \stackrel{d}{=} X_{k-\frac{k+1}{2}:n-\frac{k+1}{2}}$$

donde,

$$T_n \stackrel{d}{=} \frac{X_{n-(j+k)+1:n-k}}{X_{k-j:n-j}} \tag{3.1.2}$$

onde $j = \frac{k+1}{2}$.

A independência dos spacings *do modelo exponencial implica a independência do numerador e denominador de* (3.1.1) *e, por conseguinte, a independência do numerador e denominador de* (3.1.2).

Assim sendo,

$$f_{T_n}(t) = \int_{-\infty}^{\infty} f_{k-j:n-j}(x) f_{n-(j+k)+1:n-k}(tx)|x|dx$$

ou seja,

$$f_{T_n}(t) = \kappa_n \int_0^{\infty} x(1-e^{-x})^{k-j-1} \left(e^{-x}\right)^{n-k+jt+1} \left(1-e^{-tx}\right)^{n-j-k} dx$$

ou

$$f_{T_n}(t) = \kappa_n \sum_{i=0}^{n-j-k} \binom{n-j-k}{i} (-1)^i \int_0^{\infty} x \left(e^{-x}\right)^{n-k+(i+j)t+1} (1-e^{-x})^{k-j-1} dx.$$

onde $\kappa_n = \frac{(n-j)!}{(k-j-1)!(n-j-k)!(j-1)!}$.

Efectuando a mudança de variável $u = e^{-x}$*, segue-se que*

$$f_{T_n}(t) = -\kappa_n \sum_{i=0}^{n-j-k} \binom{n-j-k}{i} (-1)^i \int_0^1 \ln u \, u^{n-k+(j+i)t}(1-u)^{k-j-1} du$$
$$= \kappa_n \sum_{i=0}^{n-j-k} \binom{n-j-k}{i} (-1)^i \, B(n-k+(j+i)t+1, k-j)$$
$$\times \left[\psi(n-j+(j+i)t+1) - \psi(n-k+(j+i)t+1)\right].$$

Atendendo a (0.0.9)*, vem*

$$f_{T_n}(t) = \kappa_n \sum_{i=0}^{n-j-k} \binom{n-j-k}{i} (-1)^i \, B(n-k+(j+i)t+1, k-j)$$
$$\times \sum_{s=1}^{k-j} \frac{1}{n-j-s+1+(j+i)t}.$$

Como $k = \frac{n+1}{2}$ *e* $j = \frac{n+3}{4}$*, segue-se que*

$$f_{T_n}(t) = \frac{\left(\frac{3n-3}{4}\right)!}{\left(\frac{n-5}{4}\right)!^2 \left(\frac{n-1}{4}\right)!} \sum_{i=0}^{\frac{n-5}{4}} \binom{\frac{n-5}{4}}{i} (-1)^i \, B\left(\frac{n+1}{2} + \left(\frac{n+3}{4}+i\right)t, \frac{n-1}{4}\right)$$
$$\times \sum_{s=1}^{\frac{n-1}{4}} \frac{1}{\frac{3n+1}{4} - s + \left(\frac{n+3}{4}+i\right)t}, \quad t > 0. \quad (3.1.3)$$

$(i.2)$ *se* k *for par [*$n \equiv 3 \pmod 4$*], tem-se*

$$F_L = \frac{1}{2}\left(X_{\frac{k}{2}:n} + X_{\frac{k}{2}+1:n}\right) \quad e \quad F_U = \frac{1}{2}\left(X_{n-\frac{k}{2}:n} + X_{n-\frac{k}{2}+1:n}\right).$$

Consequentemente,

$$T_n = \frac{\frac{1}{2}\left(X_{n-\frac{k}{2}:n} + X_{n-\frac{k}{2}+1:n}\right) - X_{k:n}}{X_{k:n} - \frac{1}{2}\left(X_{\frac{k}{2}:n} + X_{\frac{k}{2}+1:n}\right)}$$

ou ainda,

$$T_n \stackrel{d}{=} \frac{X_{n-\frac{k}{2}:n} + X_{n-\frac{k}{2}+1:n} - 2X_{k:n}}{2X_{k:n} - X_{\frac{k}{2}:n} - X_{\frac{k}{2}+1:n}}. \tag{3.1.4}$$

Por forma a explicitarmos a f.d.p. teremos que ter em conta o lema 2.1.1 por nos ser possível exprimir a estatística definida em (3.1.4) *como quociente entre duas hipoexponenciais independentes; ou seja,*

$$T_n \stackrel{d}{=} \frac{\left(X_{n-\frac{k}{2}+1:n} - X_{n-\frac{k}{2}:n}\right) + 2\sum\limits_{i=k/2+1}^{n-k}(X_{n-i+1:n} - X_{n-i:n})}{2\sum\limits_{i=k/2+1}^{k-1}(X_{i+1:n} - X_{i:n}) + \left(X_{\frac{k}{2}+1:n} - X_{\frac{k}{2}:n}\right)}. \tag{3.1.5}$$

Sejam

$$N = \left(X_{n-\frac{k}{2}+1:n} - X_{n-\frac{k}{2}:n}\right) + 2\sum_{i=k/2+1}^{n-k}(X_{n-i+1:n} - X_{n-i:n})$$

$$D = 2\sum_{i=k/2+1}^{k-1}(X_{i+1:n} - X_{i:n}) + \left(X_{\frac{k}{2}+1:n} - X_{\frac{k}{2}:n}\right).$$

Se atendermos à reparametrização da distribuição exponencial usada no lema 2.1.1, os parâmetros de escala das exponenciais de N são

$$\theta_1 = k/2; \quad \theta_i = \frac{k/2+i-1}{2},\ i = 2,\ldots,k/2,$$

enquanto os parâmetros de escala das exponenciais de D são

$$\theta_i^* = \frac{3k/2-i-1}{2},\ i = 1,\ldots,k/2-1; \quad \theta_{k/2}^* = 3k/2-1.$$

Consequentemente, a f.d.p. de N é

$$f_N(x) = \sum_{i=1}^{k/2} C_{i,k/2}\theta_i\, e^{-\theta_i x}, \quad x > 0$$

com $C_{i,k/2} = \prod\limits_{\substack{j=1 \\ j\neq i}}^{k/2} \frac{\theta_j}{\theta_j - \theta_i}$*, enquanto a f.d.p. de* D *é*

$$f_D(x) = \sum_{i=1}^{k/2} C^*_{i,k/2} \theta^*_i \, e^{-\theta^*_i x}, \quad x > 0$$

onde $C^*_{i,k/2} = \prod\limits_{\substack{j=1 \\ j\neq i}}^{n} \frac{\theta^*_j}{\theta^*_j - \theta^*_i}$.

Tal como sucede no caso anterior, a independência dos spacings *do modelo exponencial implica a independência do numerador e denominador de* (3.1.4) *e, por conseguinte, a independência das hipoexponenciais* N *e* D*. Assim, a f.d.p. de* T_n *pode obter-se através da expressão*

$$f_{T_n}(t) = \int_{-\infty}^{\infty} f_D(x) f_N(tx)|x| dx$$

que no caso em questão traduz-se na forma

$$\begin{aligned} f_{T_n}(t) &= \int_0^{\infty} x \Bigg(\sum_{i=1}^{k/2} C^*_{i,k/2} \theta^*_i \, e^{-\theta^*_i x} \Bigg) \Bigg(\sum_{i=1}^{k/2} C_{i,k/2} \theta_i \, e^{-\theta_i t x} \Bigg) dx \\ &= \sum_{i=1}^{k/2} \sum_{j=1}^{k/2} C^*_{i,k/2} C_{j,k/2} \theta^*_i \theta_j \int_0^{\infty} x (e^{-x})^{\theta^*_i + \theta_j t} dx. \end{aligned}$$

Procedendo à substituição $u = x(\theta^*_i + \theta_j t)$*, vem*

$$\begin{aligned} f_{T_n}(t) &= \sum_{i=1}^{k/2} \sum_{j=1}^{k/2} C^*_{i,k/2} C_{j,k/2} \frac{\theta^*_i \theta_j}{(\theta^*_i + \theta_j t)^2} \int_0^{\infty} u e^{-u} du \\ &= \sum_{i=1}^{k/2} \sum_{j=1}^{k/2} C^*_{i,k/2} C_{j,k/2} \frac{\theta^*_i \theta_j}{(\theta^*_i + \theta_j t)^2}, \quad t > 0. \end{aligned}$$

Quando $n \equiv 3 \pmod 4$, $k = \frac{n+1}{2}$*; substituindo o valor de* k *na expressão anterior, vem*

$$f_{T_n}(t) = \sum_{i=1}^{\frac{n+1}{4}} \sum_{j=1}^{\frac{n+1}{4}} C^*_{i,\frac{n+1}{4}} C_{j,\frac{n+1}{4}} \frac{\theta_i^* \theta_j}{(\theta_i^* + \theta_j t)^2}, \quad t > 0. \tag{3.1.6}$$

Por exemplo, se $n = 7$*, tem-se*

$$\begin{array}{llll} \theta_1 = 2 & \theta_2 = 3/2 & C_{1,2} = -3 & C_{2,2} = 4 \\ \theta_1^* = 2 & \theta_2^* = 5 & C^*_{1,2} = 5/3 & C^*_{2,2} = -2/3 \end{array}$$

donde,

$$f_{T_7}(t) = -\frac{20}{(2+2t)^2} + \frac{20}{(2+3t/2)^2} + \frac{20}{(5+2t)^2} - \frac{20}{(5+3t/2)^2}, \quad t > 0.$$

♦ 2^o *Caso:* n *par*

Se $n = 2k$*, então*

$$\begin{aligned} \textit{profundidade M} &= \frac{n+1}{2} = k + \frac{1}{2} \\ \textit{profundidade F} &= \frac{k+1}{2}. \end{aligned}$$

Consequentemente, $M = \frac{1}{2}(X_{k:n} + X_{k+1:n})$*, e*

$(ii.1)$ *se* k *for ímpar [*$n \equiv 2 \pmod 4$*], tem-se*

$$F_L = X_{\frac{k+1}{2}:n} \quad e \quad F_U = X_{n-\frac{k+1}{2}+1:n}$$

donde

$$T_n = \frac{X_{n-\frac{k+1}{2}+1:n} - \frac{1}{2}(X_{k:n} + X_{k+1:n})}{\frac{1}{2}(X_{k:n} + X_{k+1:n}) - X_{\frac{k+1}{2}:n}}. \tag{3.1.7}$$

Por forma a simplificarmos a escrita, considere-se $j = \frac{k+1}{2}$ $(k = 2j-1)$*; então*

$$T_n \stackrel{d}{=} \frac{2X_{n-j+1:n} - X_{2j-1:n} - X_{2j:n}}{X_{2j-1:n} + X_{2j:n} - 2X_{j:n}}. \tag{3.1.8}$$

O numerador e o denominador de (3.1.8) *não são independentes, pelo que as técnicas usadas para* n *ímpar não podem aplicar-se neste caso.*

Ora o vector aleatório $(X_{j:n}, X_{2j-1:n}, X_{2j:n}, X_{n-j+1:n})$ *tem f.d.p. conjunta*

$$\begin{aligned} f(x_1,x_2,x_3,x_4) = C_{j,n}\,(1-e^{-x_1})^{j-1}(e^{-x_1}-e^{-x_2})^{j-2}(e^{-x_3}-e^{-x_4})^{n-3j} \\ \times\, e^{-x_1}e^{-x_2}e^{-x_3}(e^{-x_4})^{j}, \quad 0<x_1<x_2<x_3<x_4 \end{aligned}$$

em que $C_{j,n} = \frac{n!}{(j-1)!^2(j-2)!(n-3j)!}$.

Considerando a transformação

$$\begin{cases} X = X_{j:n} \\ Y = X_{2j-1:n} - X_{j:n} \\ Z = X_{2j:n} - X_{2j-1:n} \\ T_n = \frac{2X_{n-j+1:n} - X_{2j-1:n} - X_{2j:n}}{X_{2j-1:n} + X_{2j:n} - 2X_{j:n}} \end{cases}$$

obtém-se para f.d.p. conjunta de (X, Y, Z, T_n)

$$\begin{aligned} f(x,y,z,t) = C_{j,n}\left(y+\frac{z}{2}\right)(1-e^{-x})^{j-1}(e^{-x})^{n-j+1}(1-e^{-y})^{j-2} \\ \times\,(e^{-y})^{n+j(t-2)+2}(e^{-z})^{n+j(t-5)/2+1} \\ \times\left(1-e^{-ty}e^{-(t-1)z/2}\right)^{n-3j} \end{aligned} \tag{3.1.9}$$

se $x>0$, $y>0$, $z>0$ *e* $2ty+(t-1)z>0$.

- *Se* $0 < t < 1$, *então*

$$f_{T_n}(t) = C_{j,n}\int_0^{\infty}\int_{\frac{(1-t)z}{2t}}^{\infty}\int_0^{\infty}\left(y+\frac{z}{2}\right)(1-e^{-x})^{j-1}(e^{-x})^{n-j+1}(1-e^{-y})^{j-2}$$
$$\times\,(e^{-y})^{n+j(t-2)+2}(e^{-z})^{n+j(t-5)/2+1}$$
$$\times\left(1-e^{-ty}e^{-(t-1)z/2}\right)^{n-3j}dzdydx.$$

Assim sendo,

$$f_{T_n}(t) = C'_{j,n}\sum_{i=0}^{n-3j}\binom{n-3j}{i}(-1)^i\int_0^{\infty}\int_{\frac{(1-t)z}{2t}}^{\infty}\left(y+\frac{z}{2}\right)(1-e^{-y})^{j-2}$$
$$\times\,(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}\,dzdy$$

onde $C'_{j,n} = \frac{(n-j)!}{(j-1)!(j-2)!(n-3j)!}$*; ou ainda,*

$$f_{T_n}(t) = C'_{j,n}\sum_{i=0}^{n-3j}\binom{n-3j}{i}(-1)^i\Bigg\{\int_0^{\infty}\int_{\frac{(1-t)z}{2t}}^{\infty}y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}$$
$$\times\,(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy$$
$$+\frac{1}{2}\int_0^{\infty}\int_{\frac{(1-t)z}{2t}}^{\infty}z(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy\Bigg\}$$
$$(3.1.10)$$

Considerando o primeiro integral duplo de (3.1.10)*, tem-se*

$$\int_0^{\infty}\int_{\frac{(1-t)z}{2t}}^{\infty}y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy =$$
$$=\int_0^{\infty}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}\left(\int_{\frac{(1-t)z}{2t}}^{\infty}y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}dy\right)dz$$
$$=\sum_{s=0}^{j-2}\binom{j-2}{s}(-1)^s\int_0^{\infty}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}\left(\int_{\frac{(1-t)z}{2t}}^{\infty}y(e^{-y})^{n+(j+i)t-2j+s+2}dy\right)dz$$

e que depois de efectuarmos a mudança de variável $w = y[n+(j+i)t-2j+s+2]$ *obtém-se*

$$\int_0^\infty \int_{\frac{(1-t)z}{2t}}^\infty y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy =$$

$$= \sum_{s=0}^{j-2} \frac{\binom{j-2}{s}(-1)^s}{[n+(j+i)t-2j+s+s]^2} \int_0^\infty (e^{-z})^{n+(j+i)(t-1)/2-2j+1}$$

$$\times \Gamma\left(2, \frac{[n+(j+i)t-2j+s+2](1-t)}{2t} z\right) dz.$$

Atendendo à fórmula

$$\int_0^\infty e^{-\beta x}\Gamma(2,\alpha x)dx = \frac{2\alpha+\beta}{(\alpha+\beta)^2}, \quad \textit{se} \;\; \alpha+\beta > 0 \tag{3.1.11}$$

(cf. Gradshteyn et al. *(1994), p. 690), vem*

$$\int_0^\infty \int_{\frac{(1-t)z}{2t}}^\infty y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy =$$

$$= \sum_{s=0}^{j-2} \binom{j-2}{s}(-1)^s \frac{4t\,[n-2j+s+2-(s+1)t-(j+i)t(t-1)/2]}{[n-2j+s+2+(j+i)t]^2[n-2j+s+2+(n-2j-s)t]^2}. \tag{3.1.12}$$

Considerando agora o segundo integral de (3.1.10)*, tem-se*

$$\int_0^\infty \int_{\frac{(1-t)z}{2t}}^\infty z(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy =$$

$$= \int_0^\infty z(e^{-z})^{n+(j+i)(t-1)/2-2j+1}\left(\int_{\frac{(1-t)z}{2t}}^\infty (1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}dy\right)dz$$

$$= \sum_{s=0}^{j-2} \binom{j-2}{s}(-1)^s \int_0^\infty z(e^{-z})^{n+(j+i)(t-1)/2-2j+1}\left(\int_{\frac{(1-t)z}{2t}}^\infty (e^{-y})^{n+(j+i)t-2j+s+2}dy\right)dz.$$

Efectuando a mudança de variável $w = y[n+(j+i)t-2j+s+2]$, segue-se que

$$\int_0^\infty \int_{\frac{(1-t)z}{2t}}^\infty z(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy =$$

$$= \sum_{s=0}^{j-2} \frac{\binom{j-2}{s}(-1)^s}{n+(j+i)t-2j+s+2} \int_0^\infty z(e^{-z})^{n+(j+i)(t-1)/2-2j+1}$$
$$\times \Gamma\left(1, \frac{[n+(j+i)t-2j+s+2](1-t)}{2t}z\right)dz.$$

Dado que

$$\int_0^\infty x e^{-\beta x}\Gamma(1,\alpha x)dx = \frac{1}{(\alpha+\beta)^2}, \quad \textit{se} \;\; \alpha+\beta>0 \tag{3.1.13}$$

(cf. Gradshteyn et al. *(1994), p. 690), decorre que*

$$\int_0^\infty \int_{\frac{(1-t)z}{2t}}^\infty z(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy =$$
$$= \sum_{s=0}^{j-2}\binom{j-2}{s}(-1)^s \frac{4t^2}{[n-2j+s+2+(j+i)t][n-2j+s+2+(n-2j-s)t]^2}. \tag{3.1.14}$$

Substituindo (3.1.12) *e* (3.1.14) *em* (3.1.10)*, vem após simplificarmos a expressão*

$$f_{T_n}(t) = C'_{j,n}\sum_{i=0}^{n-3j}\sum_{s=0}^{j-2}\binom{n-3j}{i}\binom{j-2}{s}(-1)^{i+s}$$
$$\times \frac{2t[2(n-2j+s+2)+(n-j+i-s)t]}{[n-2j+s+2+(j+i)t]^2[n-2j+s+2+(n-2j-s)t]^2}.$$

Assim, se $n \equiv 2 \pmod 4$, $j = \frac{n+2}{4}$; donde,

$$f_{T_n}(t) = \frac{\left(\frac{3n-2}{4}\right)!}{\left(\frac{n-2}{4}\right)!\left(\frac{n-6}{4}\right)!^2}\sum_{i=0}^{\frac{n-6}{4}}\sum_{s=0}^{\frac{n-6}{4}}\binom{\frac{n-6}{4}}{i}\binom{\frac{n-6}{4}}{s}(-1)^{i+s}$$
$$\times \frac{2t\left[n+2s+2+\left(\frac{3n-2}{4}+i-s\right)t\right]}{\left[\frac{n+2}{2}+s+\left(\frac{n+2}{4}+i\right)t\right]^2\left[\frac{n+2}{2}+s+\left(\frac{n-2}{2}-s\right)t\right]^2}. \tag{3.1.15}$$

- *Se* $t \geq 1$*, tem-se*

$$f_{T_n}(t) = C_{j,n}\int_0^\infty\int_0^\infty\int_0^\infty \left(y+\frac{z}{2}\right)(1-e^{-x})^{j-1}(e^{-x})^{n-j+1}(1-e^{-y})^{j-2}(e^{-y})^{n+j(t-2)+2} \times (e^{-z})^{n+j(t-5)/2+1}\left(1-e^{-ty}e^{-(t-1)z/2}\right)^{n-3j}dzdydx.$$

Ora

$$\begin{aligned} f_{T_n}(t) = &C'_{j,n}\sum_{i=0}^{n-3j}\binom{n-3j}{i}(-1)^i\Big\{\int_0^\infty\int_0^\infty y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2} \\ &\times(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy \\ &+\frac{1}{2}\int_0^\infty\int_0^\infty z(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy\Big\} \end{aligned} \tag{3.1.16}$$

com $C'_{j,n}$ *definido de forma idêntica à da situação* $0<t<1$*.*

No que respeita o primeiro integral duplo de (3.1.16)*, tem-se*

$$\begin{aligned} &\int_0^\infty\int_0^\infty y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy = \\ &=\left(\int_0^\infty (e^{-z})^{n+(j+i)(t-1)/2-2j+1}dz\right)\left(\int_0^\infty y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}dy\right) \\ &=-\frac{1}{n+(j+i)\left(\frac{t-1}{2}\right)-2j+1}\int_0^1 \ln w\, w^{n+(j+i)t-2j+1}(1-w)^{j-2}dw \\ &=\frac{B(n+(j+i)t-2j+2,j-1)}{n+(j+i)\left(\frac{t-1}{2}\right)-2j+1}[\psi(n+(j+i)t-j+1)-\psi(n+(j+i)t-2j+2)]. \end{aligned}$$

De (0.0.9)*, segue-se que*

$$\begin{aligned} &\int_0^\infty\int_0^\infty y(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy = \\ &\quad=\frac{B(n+(j+i)t-2j+2,j-1)}{n+(j+i)\left(\frac{t-1}{2}\right)-2j+1}\sum_{s=1}^{j-1}\frac{1}{n-j-s+1+(j+i)t}. \end{aligned} \tag{3.1.17}$$

Quanto ao segundo integral, tem-se

$$\int_0^{\infty}\int_0^{\infty} z(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dzdy =$$
$$= \left(\int_0^{\infty} z(e^{-z})^{n+(j+i)(t-1)/2-2j+1}dz\right)\left(\int_0^{\infty}(1-e^{-y})^{j-2}(e^{-y})^{n+(j+i)t-2j+2}dy\right)$$
$$= \frac{1}{\left[n+(j+i)\left(\frac{t-1}{2}\right)-2j+1\right]^2}\, B(n+(j+i)t-2j+2, j-1). \tag{3.1.18}$$

Substituindo (3.1.17) *e* (3.1.18) *em* (3.1.16)*, obtemos após simplificarmos a expressão*

$$f_{T_n}(t) = C'_{j,n}\sum_{i=0}^{n-3j}\binom{n-3j}{i}(-1)^i\,\frac{2B(n+(j+i)t-2j+2,j-1)}{2n+(j+i)(t-1)-4j+2}$$
$$\times\left\{\sum_{s=1}^{j-1}\frac{1}{n+(j+i)t-j-s+1}+\frac{1}{2n+(j+i)(t-1)-4j+2}\right\}.$$

Como $j=\frac{n+2}{4}$*, vem*

$$f_{T_n}(t) = \frac{\left(\frac{3n-2}{4}\right)!}{\left(\frac{n-2}{4}\right)!\left(\frac{n-6}{4}\right)!^2}\sum_{i=0}^{\frac{n-6}{4}}\binom{\frac{n-6}{4}}{i}(-1)^i\,\frac{2B\left(\frac{n+2}{2}+\left(\frac{n+2}{4}+i\right)t,\frac{n-2}{4}\right)}{\frac{3n-2}{4}-i+\left(\frac{n+2}{4}+i\right)t}$$
$$\times\left\{\sum_{s=1}^{\frac{n-2}{4}}\frac{1}{\frac{3n+2}{4}-s+\left(\frac{n+2}{4}+i\right)t}+\frac{1}{\frac{3n-2}{4}-i+\left(\frac{n+2}{4}+i\right)t}\right\}. \tag{3.1.19}$$

$(ii.2)$ *se k for par [*$n\equiv 0 \pmod 4$*], tem-se*

$$F_L=\frac{1}{2}(X_{\frac{k}{2}:n}+X_{\frac{k}{2}+1:n}) \quad e \quad F_U=\frac{1}{2}(X_{n-\frac{k}{2}:n}+X_{n-\frac{k}{2}+1:n}).$$

Neste caso,

$$T_n = \frac{\frac{1}{2}(X_{n-\frac{k}{2}:n} + X_{n-\frac{k}{2}+1:n}) - \frac{1}{2}(X_{k:n} + X_{k+1:n})}{\frac{1}{2}(X_{k:n} + X_{k+1:n}) - \frac{1}{2}(X_{\frac{k}{2}:n} + X_{\frac{k}{2}+1:n})}$$

ou

$$T_n \stackrel{d}{=} \frac{X_{n-\frac{k}{2}:n} + X_{n-\frac{k}{2}+1:n} - X_{k:n} - X_{k+1:n}}{X_{k:n} + X_{k+1:n} - X_{\frac{k}{2}:n} - X_{\frac{k}{2}+1:n}}. \tag{3.1.20}$$

Como o numerador e o denominador de (3.1.20) *não são independentes, teremos que seguir um caminho análogo ao do caso* $n \equiv 2 \pmod 4$.

Ora o vector aleatório $\left(X_{\frac{k}{2}:n}, X_{\frac{k}{2}+1:n}, X_{k:n}, X_{k+1:n}, X_{n-\frac{k}{2}:n}, X_{n-\frac{k}{2}+1:n}\right)$ *tem f.d.p. conjunta*

$$f(x_1, x_2, x_3, x_4, x_5, x_6) = C_{k,n}(1 - e^{-x_1})^{k/2-1}(e^{-x_2} - e^{-x_3})^{k/2-2}(e^{-x_4} - e^{-x_5})^{n-3k/2-2} \\ \times e^{-x_1}e^{-x_2}e^{-x_3}e^{-x_4}e^{-x_5}(e^{-x_6})^{k/2}$$

se $0 < x_1 < x_2 < x_3 < x_4 < x_5 < x_6$*, e onde* $C_{k,n} = \frac{n!}{(k/2-1)!^2(k/2-2)!(n-3k/2-2)!}$.

Considerando a transformação

$$\begin{cases} X = X_{\frac{k}{2}:n} \\ Y = X_{\frac{k}{2}+1:n} - X_{\frac{k}{2}:n} \\ Z = X_{k:n} - X_{\frac{k}{2}+1:n} \\ W = X_{k+1:n} - X_{k:n} \\ U = X_{n-\frac{k}{2}:n} - X_{k+1:n} \\ T_n = \frac{X_{n-\frac{k}{2}:n} + X_{n-\frac{k}{2}+1:n} - X_{k:n} - X_{k+1:n}}{X_{k:n} + X_{k+1:n} - X_{\frac{k}{2}:n} - X_{\frac{k}{2}+1:n}} \end{cases}$$

obtemos para f.d.p. conjunta de (X, Y, Z, W, U, T_n)

$$
\begin{aligned}
f(x,y,z,w,u,t) = C_{k,n}(y+2z+w)(1-e^{-x})^{k/2-1}(e^{-x})^{n-k/2+1} \\
\times (e^{-y})^{n+(t-1)k/2}(1-e^{-z})^{k/2-2}(e^{-z})^{n+(t-1)k+1} \\
\times (e^{-w})^{n+(t-3)k/2}(1-e^{-u})^{n-3k/2-2}(e^{-u})^{1-k/2}
\end{aligned}
$$

se $x>0$, $y>0$, $z>0$, $w>0$, $u>0$ *e* $w+2u<t(y+2z+w)$.

Como a determinação da f.d.p. marginal de T_n envolve o cálculo de um integral quíntuplo cujos limites de integração não conseguimos identificar por o suporte da distribuição de (X,Y,Z,W,U,T_n) o impossibilitar, não nos é possível explicitar a f.d.p. de T_n quando n é múltiplo de 4. É claro que esta situação não invalida os resultados obtidos até aqui. Porém, na secção seguinte encontaremos uma solução aproximada, mas bastante satisfatória, para qualquer valor de n.

3.2 Distribuição Aproximada

A impossibilidade de explicitarmos a f.d.p. de T_n quando n é múltiplo de quatro deve-se em parte ao facto de termos que interpolar todas as letras-resumo que definem a estatística. Por forma a reduzirmos o número de interpolações, podemos atender a que os quartos são uma forma de quartil. Usaremos a definição de quantil amostral de ordem p ($0<p<1$) encontrada em Casella e Berger (1990) (p. 230); ou seja,

$$
\tilde{\xi}_p = \begin{cases} X_{\{np\}:n} & , \text{ se } p<0.5 \\ X_{n-\{n(1-p)\}+1:n} & , \text{ se } p>0.5 \end{cases}
$$

onde a notação $\{b\}$ define-se como o número b arredondado de forma usual ao inteiro mais próximo[4]. No que refere a mediana amostral ($p=0.5$), define-se como usualmente.

Repare-se que a definição anterior faz com que os quantis amostrais de ordem p e $1-p$ tenham igual profundidade ($\{np\}$ se $p<0.5$). Assim, o 1^o quartil é dado pela e.o. $X_{\{n/4\}:n}$, enquanto o 3^o quartil por $X_{n-\{n/4\}+1:n}$. Esta abordagem nos permitirá reduzir de quatro para dois o número de casos a estudar, sendo-nos possível explicitar a f.d.p. para qualquer n.

[4]Se i for um inteiro tal que $i-0.5\le b<i+0.5$, então $\{b\}=i$.

Denotemos a nova estatística por T_n^*. Tal como acontecia anteriormente, T_n^* é distribucionalmente independente de δ, de modo que consideremos, sem perda de generalidade, $\delta = 1$.

- *Se n for ímpar,*

$$T_n^* = \frac{X_{n-\{n/4\}+1:n} - X_{\frac{n+1}{2}:n}}{X_{\frac{n+1}{2}:n} - X_{\{n/4\}:n}} \tag{3.2.1}$$

e da propriedade (2.1.11)*, segue-se que*

$$T_n^* \stackrel{d}{=} \frac{X_{\frac{n-1}{2}-\{n/4\}+1:\frac{n-1}{2}}}{X_{\frac{n+1}{2}-\{n/4\}:n-\{n/4\}}}. \tag{3.2.2}$$

A independência dos spacings *do modelo exponencial implica a independência do numerador e denominador de* (3.2.1) *e, por conseguinte, a independência do numerador e denominador de* (3.2.2)*.*

Assim,

$$\begin{aligned} f_{T_n^*}(t) &= k_n \int_0^\infty x(1-e^{-x})^{(n+1)/2-\{n/4\}-1}(e^{-x})^{(n-1)/2+t\{n/4\}+1} \\ &\qquad \times (1-e^{-tx})^{(n-1)/2-\{n/4\}}dx \\ &= k_n \sum_{i=0}^{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}} \binom{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}}{i}(-1)^i \int_0^\infty x(e^{-x})^{(n-1)/2+t(i+\{n/4\})+1} \\ &\qquad \times (1-e^{-x})^{(n+1)/2-\{n/4\}-1}dx \end{aligned}$$

com $k_n = \frac{(n-\{n/4\})!}{((n-1)/2-\{n/4\})!^2(\{n/4\}-1)!}$.

Procedendo à transformação $w = e^{-x}$, *vem*

$$
\begin{aligned}
f_{T_n^*}(t) = &-k_n \sum_{i=0}^{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}} \binom{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}}{i} (-1)^i \int_0^1 \ln w \, w^{(n-1)/2+t(i+\{n/4\})} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times (1-w)^{(n+1)/2-\{n/4\}-1} dw \\
= &\, k_n \sum_{i=0}^{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}} \binom{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}}{i} (-1)^i B\left(\tfrac{n-1}{2}+t\left(i+\left\{\tfrac{n}{4}\right\}\right)+1, \tfrac{n+1}{2}-\left\{\tfrac{n}{4}\right\}\right) \\
&\times \left[\psi\left(n-\left\{\tfrac{n}{4}\right\}+t\left(i+\left\{\tfrac{n}{4}\right\}\right)+1\right)-\psi\left(\tfrac{n-1}{2}+t\left(i+\left\{\tfrac{n}{4}\right\}\right)+1\right)\right].
\end{aligned}
$$

De (0.0.9)*, segue-se que*

$$
\begin{aligned}
f_{T_n^*}(t) = k_n & \sum_{i=0}^{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}} \binom{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}}{i} (-1)^i B\left(\tfrac{n+1}{2}+t\left(i+\left\{\tfrac{n}{4}\right\}\right), \tfrac{n+1}{2}-\left\{\tfrac{n}{4}\right\}\right) \\
& \times \sum_{j=1}^{\frac{n+1}{2}-\left\{\frac{n}{4}\right\}} \frac{1}{n-\left\{\frac{n}{4}\right\}-j+1+t\left(i+\left\{\frac{n}{4}\right\}\right)}
\end{aligned}
$$

ou ainda,

$$
\begin{aligned}
f_{T_n^*}(t) = &\frac{\left(n-\left\{\frac{n}{4}\right\}\right)!}{\left(\left\{\frac{n}{4}\right\}-1\right)!\left(\frac{n-1}{2}-\left\{\frac{n}{4}\right\}\right)!^2} \sum_{i=0}^{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}} \binom{\frac{n-1}{2}-\left\{\frac{n}{4}\right\}}{i} (-1)^i \\
&\times B\left(\tfrac{n+1}{2}+\left(i+\left\{\tfrac{n}{4}\right\}\right)t, \tfrac{n+1}{2}-\left\{\tfrac{n}{4}\right\}\right) \\
&\times \sum_{j=1}^{\frac{n+1}{2}-\left\{\frac{n}{4}\right\}} \frac{1}{n-\left\{\frac{n}{4}\right\}+1-j+\left(i+\left\{\frac{n}{4}\right\}\right)t}, \quad t>0. \qquad (3.2.3)
\end{aligned}
$$

- *Se n for par,*

$$T_n^* = \frac{X_{n-\{n/4\}+1:n} - \frac{1}{2}\left(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}\right)}{\frac{1}{2}\left(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}\right) - X_{\{n/4\}:n}}$$

ou

$$T_n^* \stackrel{d}{=} \frac{2X_{n-\{n/4\}+1:n} - X_{\frac{n}{2}:n} - X_{\frac{n}{2}+1:n}}{X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n} - 2X_{\{n/4\}:n}}. \tag{3.2.4}$$

Neste caso, o numerador e o denominador não são independentes, pelo que as técnicas usadas para n ímpar não podem ser aplicadas.

Ora a f.d.p. conjunta de $(X_{\{n/4\}:n}, X_{\frac{n}{2}:n}, X_{\frac{n}{2}+1:n}, X_{n-\{n/4\}+1:n})$ é

$$\begin{aligned} f(x_1, x_2, x_3, x_4) = C_n &\left(1 - e^{-x_1}\right)^{\{n/4\}-1} \left(e^{-x_1} - e^{-x_2}\right)^{n/2-\{n/4\}-1} \\ &\times \left(e^{-x_3} - e^{-x_4}\right)^{n/2-\{n/4\}-1} e^{-x_1}\, e^{-x_2}\, e^{-x_3} \left(e^{-x_4}\right)^{\{n/4\}} \end{aligned} \tag{3.2.5}$$

se $0 < x_1 < x_2 < x_3 < x_4$ e com $C_n = \frac{n!}{(\{n/4\}-1)!^2 (n/2-\{n/4\}-1)!^2}$.

Efectuando a transformação

$$\begin{cases} X = X_{\{n/4\}:n} \\ Y = X_{\frac{n}{2}:n} - X_{\{n/4\}:n} \\ Z = X_{\frac{n}{2}+1:n} - X_{\frac{n}{2}:n} \\ T_n^* = \frac{2X_{n-\{n/4\}+1:n} - X_{\frac{n}{2}:n} - X_{\frac{n}{2}+1:n}}{X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n} - 2X_{\{n/4\}:n}} \end{cases}$$

obtemos para a f.d.p. conjunta de (X, Y, Z, T_n^*)

$$\begin{aligned} f(x, y, z, t) = C_n \left(y + \frac{z}{2}\right) &\left(1 - e^{-x}\right)^{\{n/4\}-1} \left(e^{-x}\right)^{n-\{n/4\}+1} \left(e^{-y}\right)^{n/2+t\{n/4\}+1} \\ &\times \left(1 - e^{-y}\right)^{n/2-\{n/4\}-1} \left(e^{-z}\right)^{n/2+\{n/4\}(t-1)/2} \\ &\times \left(1 - e^{-ty} e^{-(t-1)z/2}\right)^{n/2-\{n/4\}-1} \end{aligned} \tag{3.2.6}$$

se $x > 0$, $y > 0$, $z > 0$ e $2ty + (t-1)z > 0$.

(i) *Se* $0 < t < 1$, *segue de* (3.2.6) *que*

$$f_{T_n^*}(t) = C_n \int_0^\infty \int_{\frac{(1-t)z}{2t}}^\infty \int_0^\infty \left(y + \frac{z}{2}\right) \left(1 - e^{-x}\right)^{\{n/4\}-1} \left(e^{-x}\right)^{n-\{n/4\}+1}$$
$$\times \left(1 - e^{-y}\right)^{n/2-\{n/4\}-1} \left(e^{-y}\right)^{n/2+t\{n/4\}+1}$$
$$\times \left(e^{-z}\right)^{n/2+\{n/4\}(t-1)/2} \left(1 - e^{-ty} e^{-(t-1)z/2}\right)^{n/2-\{n/4\}-1} dz dy dx.$$

Consequentemente,

$$f_{T_n^*}(t) = C_n^* \sum_{i=0}^{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1} \binom{\frac{n}{2} - \left\{\frac{n}{4}\right\} - 1}{i} (-1)^i$$
$$\times \left\{ \int_{\frac{(1-t)z}{2t}}^\infty \int_0^\infty y(1 - e^{-y})^{n/2-\{n/4\}-1} (e^{-y})^{n/2+t(i+\{n/4\})+1} \right.$$
$$\times (e^{-z})^{n/2+(i+\{n/4\})(t-1)/2} dz dy$$
$$+ \frac{1}{2} \int_{\frac{(1-t)z}{2t}}^\infty \int_0^\infty z(1 - e^{-y})^{n/2-\{n/4\}-1} (e^{-y})^{n/2+t(i+\{n/4\})+1}$$
$$\left. \times (e^{-z})^{n/2+(i+\{n/4\})(t-1)/2} dz dy \right\}$$

em que $C_n^* = \frac{(n-\{n/4\})!}{(\{n/4\}-1)!(n/2-\{n/4\}-1)!^2}$.

No entanto, a expressão anterior é equivalente a

$$f_{T_n^*}(t) = C_n^* \sum_{i,j=0}^{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1} \binom{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1}{i}\binom{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1}{j}(-1)^{i+j}$$
$$\times\left\{\int_{\frac{(1-t)z}{2t}}^{\infty}\int_0^{\infty} y\left(e^{-y}\right)^{n/2+t(i+\{n/4\})+j+1}\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}dzdy\right.$$
$$\left.+\frac{1}{2}\int_{\frac{(1-t)z}{2t}}^{\infty}\int_0^{\infty} z\left(e^{-y}\right)^{n/2+t(i+\{n/4\})+j+1}\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}dzdy\right\}. \tag{3.2.7}$$

Calculando separadamente os integrais de (3.2.7)*, vem:*

(a)

$$\int_{\frac{(1-t)z}{2t}}^{\infty}\int_0^{\infty} y\left(e^{-y}\right)^{n/2+t(i+\{n/4\})+j+1}\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}dzdy =$$
$$= \int_0^{\infty}\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}\left(\int_{\frac{(1-t)z}{2t}}^{\infty} y\left(e^{-y}\right)^{n/2+t(i+\{n/4\})+j+1}dy\right)dz$$
$$= \frac{\int_0^{\infty}\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}\Gamma\left(2,\frac{(1-t)[n/2+t(i+\{n/4\})+j+1]}{2t}z\right)dz}{\left[\frac{n}{2}+t(i+\left\{\frac{n}{4}\right\})+j+1\right]^2}.$$

Atendendo à fórmula (3.1.11)*, segue-se que*

$$\int_{\frac{(1-t)z}{2t}}^{\infty}\int_0^{\infty} y\left(e^{-y}\right)^{n/2+t(i+\{n/4\})+j+1}\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}dzdy =$$
$$= \frac{2t\left\{n+2j+2+t\left[i+\left\{\frac{n}{4}\right\}-2j-2-\left(i+\left\{\frac{n}{4}\right\}\right)t\right]\right\}}{\left[\frac{n}{2}+j+1+\left(i+\left\{\frac{n}{4}\right\}\right)t\right]^2\left[\frac{n}{2}+j+1+\left(\frac{n}{2}-j-1\right)t\right]^2}. \tag{3.2.8}$$

(b)

$$\int_{\frac{(1-t)z}{2t}}^{\infty}\int_0^{\infty} z\left(e^{-y}\right)^{n/2+t(i+\{n/4\})+j+1}\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}dzdy =$$

$$= \frac{\int_0^{\infty} z\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}\Gamma\left(1,\frac{(1-t)[n/2+j+1+t(i+\{n/4\})]}{2t}z\right)dz}{\frac{n}{2}+j+1+(i+\left\{\frac{n}{4}\right\})t}. \qquad (3.2.9)$$

De (3.1.13)*, obtém-se*

$$\int_{\frac{(1-t)z}{2t}}^{\infty}\int_0^{\infty} z\left(e^{-y}\right)^{n/2+t(i+\{n/4\})+j+1}\left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2}dzdy =$$

$$= \frac{4t^2}{\left[\frac{n}{2}+j+1+t\left(i+\left\{\frac{n}{4}\right\}\right)\right]\left[\frac{n}{2}+j+1+t\left(\frac{n}{2}-j-1\right)\right]^2}. \qquad (3.2.10)$$

Substituindo (3.2.8) e (3.2.10) em (3.2.7), obtém-se

$$f_{T_n^*}(t) = \frac{(n-\left\{\frac{n}{4}\right\})!}{(\left\{\frac{n}{4}\right\}-1)!\left(\frac{n}{2}-\left\{\frac{n}{4}\right\}-1\right)!^2}\sum_{i,j=0}^{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1}\binom{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1}{i}\binom{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1}{j}$$

$$\times(-1)^{i+j}\frac{2(n+2j+2)t+(n+2i-2j-2+2\left\{\frac{n}{4}\right\})t^2}{\left[\frac{n}{2}+j+1+(i+\left\{\frac{n}{4}\right\})t\right]^2\left[\frac{n}{2}+j+1+\left(\frac{n}{2}-j-1\right)t\right]^2}. \qquad (3.2.11)$$

(ii) *Se* $t \geq 1$*, decorre de* (3.2.6) *que*

$$f_{T_n^*}(t) = C_n\int_0^{\infty}\int_0^{\infty}\int_0^{\infty}\left(y+\frac{z}{2}\right)\left(1-e^{-x}\right)^{\{n/4\}-1}\left(e^{-x}\right)^{n-\{n/4\}+1}\left(e^{-y}\right)^{n/2+t\{n/4\}+1}$$

$$\times\left(1-e^{-y}\right)^{n/2-\{n/4\}-1}\left(e^{-z}\right)^{n/2+\{n/4\}(t-1)/2}$$

$$\times\left(1-e^{-ty}e^{-(t-1)z/2}\right)^{n/2-\{n/4\}-1}dzdydx$$

ou

$$f_{T_n^*}(t) = C_n^* \sum_{i=0}^{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1} \binom{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1}{i} (-1)^i$$
$$\times \Bigg\{ \int_0^\infty \int_0^\infty y \left(1-e^{-y}\right)^{n/2-\{n/4\}-1} \left(e^{-y}\right)^{n/2+t(i+\{n/4\})+1}$$
$$\times \left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2} dzdy$$
$$+ \frac{1}{2} \int_0^\infty \int_0^\infty z \left(1-e^{-y}\right)^{n/2-\{n/4\}-1} \left(e^{-y}\right)^{n/2+t(i+\{n/4\})+1}$$
$$\times \left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2} dzdy \Bigg\}$$

com C_n^* *definido da mesma forma que no caso* $0 < t < 1$.
Por um lado, tem-se

$$\int_0^\infty \int_0^\infty y \left(1-e^{-y}\right)^{n/2-\{n/4\}-1} \left(e^{-y}\right)^{n/2+t(i+\{n/4\})+1} \left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2} dzdy =$$
$$= \frac{B\left(\frac{n}{2}+\left(i+\left\{\frac{n}{4}\right\}\right)t+1, \frac{n}{2}-\left\{\frac{n}{4}\right\}\right)}{\frac{n}{2}+\left(i+\left\{\frac{n}{4}\right\}\right)\left(\frac{t-1}{2}\right)}$$
$$\times \left[\psi\left(n-\left\{\frac{n}{4}\right\}+\left(i+\left\{\frac{n}{4}\right\}\right)t+1\right) - \psi\left(\frac{n}{2}+\left(i+\left\{\frac{n}{4}\right\}\right)t+1\right)\right]$$

e de (0.0.9), segue-se que

$$\int_0^\infty \int_0^\infty y \left(1-e^{-y}\right)^{n/2-\{n/4\}-1} \left(e^{-y}\right)^{n/2+t(i+\{n/4\})+1} \left(e^{-z}\right)^{n/2+(i+\{n/4\})(t-1)/2} dzdy =$$
$$= \frac{B\left(\frac{n}{2}+\left(i+\left\{\frac{n}{4}\right\}\right)t+1, \frac{n}{2}-\left\{\frac{n}{4}\right\}\right)}{\frac{n}{2}+\left(i+\left\{\frac{n}{4}\right\}\right)\left(\frac{t-1}{2}\right)} \sum_{j=1}^{\frac{n}{2}-\left\{\frac{n}{4}\right\}} \frac{1}{n-\left\{\frac{n}{4}\right\}-j+1+\left(i+\left\{\frac{n}{4}\right\}\right)t}.$$

Por outro lado, tem-se

$$\int_0^\infty \int_0^\infty z \left(1-e^{-y}\right)^{\frac{n}{2}-\{n/4\}-1} \left(e^{-y}\right)^{\frac{n}{2}+t(i+\{n/4\})+1} \left(e^{-z}\right)^{\frac{n}{2}+(i+\{n/4\})\left(\frac{t-1}{2}\right)} dzdy =$$
$$= \frac{B\left(\frac{n}{2}+\left(i+\left\{\frac{n}{4}\right\}\right)t+1, \frac{n}{2}-\left\{\frac{n}{4}\right\}\right)}{\left[\frac{n}{2}+\left(i+\left\{\frac{n}{4}\right\}\right)\left(\frac{t-1}{2}\right)\right]^2}.$$

Consequentemente,

$$f_{T_n^*}(t) = \frac{\left(n-\left\{\frac{n}{4}\right\}\right)!}{\left(\left\{\frac{n}{4}\right\}-1\right)!\left(\frac{n}{2}-\left\{\frac{n}{4}\right\}-1\right)!^2} \sum_{i=0}^{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1} \binom{\frac{n}{2}-\left\{\frac{n}{4}\right\}-1}{i} (-1)^i$$
$$\times \frac{2B\left(\frac{n}{2}+1+(i+\left\{\frac{n}{4}\right\})t, \frac{n}{2}-\left\{\frac{n}{4}\right\}\right)}{n-\left\{\frac{n}{4}\right\}-i+(i+\left\{\frac{n}{4}\right\})t}$$
$$\times \left\{ \sum_{j=1}^{\frac{n}{2}-\left\{\frac{n}{4}\right\}} \frac{1}{n-\left\{\frac{n}{4}\right\}-j+1+(i+\left\{\frac{n}{4}\right\})t} + \frac{1}{n-\left\{\frac{n}{4}\right\}-i+(i+\left\{\frac{n}{4}\right\})t} \right\}. \quad (3.2.12)$$

Na tabela 3.2.1 indicamos alguns quantis da distribuição de T_n^* para $n = 3(1)30$ para fins de inferência. Os cálculos foram feitos usando o package Mathematica *v. 2.2.*

3.3 Distribuição Assintótica

Na maioria das vezes as amostras de que dispomos são pequenas, mas na eventualidade de termos amostras de dimensão superior a 30, a tabela 3.2.1 não tem qualquer utilidade. Nestes casos, não nos resta outra alternativa a não ser recorrermos a aproximações assintóticas. Daí, mostraremos que sob a hipótese de parente exponencial a distribuição de T_n^* é aproximada pela distribuição Gaussiana para valores elevados de n. Para o efeito, usaremos o lema seguinte.

Lema 3.3.1 *Seja $X_{i:n}$ a i-ésima estatística ordinal de n v.a. exponenciais standard i.i.d. $X_1, \ldots, X_n$; então,*

$$P\left\{\tau_n^{-1} \sum_{i=1}^{n} a_{i:n}\left(X_{i:n}-\mu_{i:n}\right) \leq t\right\} \longrightarrow \Phi(t), \quad n \to \infty$$

Tabela 3.2.1 – Quantis da distribuição de T_n^*.

	p												
n	.0005	.001	.005	.01	.025	.05	.075	.1	.15	.2	.3	.4	.5
3	.0010005	.002002	.0100503	.020202	.0512821	.105263	.162162	.222222	.352941	.5	.857143	1.33333	2
4	.0339203	.0481962	.109953	.157895	.25753	.377631	.475966	.563427	.720465	.864021	1.14039	1.47214	1.89898
5	.0317789	.0452952	.104761	.152076	.253828	.38311	.494905	.599284	.799805	1	1.4264	1.9212	2.53199
6	.0261576	.0371939	.0851249	.122545	.200915	.296473	.375618	.446705	.576014	.69614	.924951	1.16228	1.4641
7	.0237974	.0339081	.0783175	.113572	.189165	.284822	.367234	.443957	.590826	.736881	1.04648	1.40391	1.84326
8	.0788097	.100698	.181106	.235947	.340696	.459053	.553215	.636308	.786431	.927271	1.20835	1.51498	1.87437
9	.0794793	.101789	.18461	.241868	.352935	.481009	.584734	.677508	.847698	1.00975	1.33643	1.69301	2.10936
10	.0688629	.0879961	.158306	.206277	.297931	.401498	.483861	.556487	.687467	.809912	1.05219	1.31383	1.61883
11	.0686749	.0879237	.159285	.208538	.303903	.413623	.502314	.581529	.726605	.864492	1.14186	1.44392	1.79591
12	.128054	.155508	.249264	.309436	.419674	.539709	.633041	.714271	.858967	.992674	1.25339	1.52804	1.83889
13	.130834	.159141	.256445	.31938	.435567	.563223	.663171	.750562	.906955	1.05204	1.33571	1.63483	1.9732
14	.116863	.141903	.227373	.28219	.382536	.491667	.576412	.650085	.781112	.901944	1.13691	1.38373	1.66245
15	.118703	.144346	.232398	.289277	.394155	.509217	.599167	.677804	.818337	.948571	1.20288	1.47067	1.77324
16	.175702	.207072	.309842	.373452	.486978	.607504	.699585	.77878	.917993	1.04476	1.28739	1.53729	1.8141
17	.179864	.212214	.318627	.384789	.503349	.62977	.726652	.810138	.95716	1.09124	1.34816	1.61291	1.90615
18	.163923	.19316	.288876	.348073	.45363	.565567	.650999	.724417	.853344	.970617	1.19477	1.4253	1.68035
19	.167212	.197241	.295934	.357238	.466993	.583903	.673422	.750517	.886192	1.00984	1.24655	1.49026	1.75998
20	.220092	.254261	.363179	.428946	.544165	.664212	.754687	.831774	.965854	1.08655	1.31433	1.5451	1.79682
21	.225101	.260251	.372584	.440598	.560022	.684738	.778878	.859162	.998925	1.12483	1.36256	1.60348	1.86627
22	.2081	.240369	.343161	.405183	.513759	.626783	.711899	.784379	.910364	1.02369	1.23738	1.45367	1.68942
23	.212339	.245448	.351184	.415157	.527406	.644537	.732896	.808217	.939275	1.05727	1.27994	1.50544	1.75128
24	.260952	.297159	.410317	.477406	.593295	.712291	.801015	.876047	1.00546	1.1209	1.33638	1.55192	1.78428
25	.266457	.303595	.41986	.488907	.60834	.731138	.822774	.900308	1.0341	1.15349	1.37641	1.59942	1.83983
26	.248963	.283462	.391224	.455071	.565294	.678392	.762668	.83391	.956728	1.06623	1.2705	1.47468	1.69468
27	.253797	.289122	.399648	.465247	.578652	.695179	.782094	.855609	.982416	1.09553	1.30662	1.51769	1.74513
28	.298486	.33618	.452277	.520129	.636044	.753679	.840628	.913714	1.03892	1.14978	1.35489	1.55796	1.77482
29	.304236	.342819	.461715	.531286	.650225	.771029	.860365	.935479	1.06419	1.17818	1.38912	1.59798	1.82102
30	.286611	.322775	.434061	.499067	.610061	.72264	.805815	.875704	.995389	1.10132	1.29722	1.49108	1.69802

Tabela 3.2.1 — (Continuação)

	p											
n	.6	.7	.8	.85	.9	.925	.95	.975	.99	.995	.999	.9995
3	3	4.66667	8	11.3333	18	24.6667	38	78	198	398	1998	3998
4	2.47723	3.32456	4.74597	5.94427	7.95445	9.64911	12.4919	18.9089	31.641	45.9898	106.545	151.919
5	3.33679	4.49353	6.40755	8.00951	10.6861	12.9372	16.708	25.2088	42.0613	61.0474	141.159	201.186
6	1.87298	2.47214	3.47723	4.32456	5.74597	6.94427	8.95445	13.4919	22.4949	32.641	75.4597	107.545
7	2.4201	3.24661	4.61053	5.75011	7.65208	9.25053	11.9268	17.9572	29.9083	43.3704	100.17	142.728
8	2.32418	2.93493	3.87688	4.61551	5.77373	6.69018	8.13465	11.0924	16.1972	21.2305	38.5345	49.3624
9	2.62809	3.32936	4.40656	5.249	6.56767	7.60979	9.25092	12.6082	18.3984	24.1052	43.7196	55.9919
10	1.99927	2.51454	3.30764	3.92876	4.90191	5.67146	6.88388	9.36528	13.6463	17.8665	32.3728	41.4495
11	2.23369	2.82459	3.73087	4.4389	5.54632	6.421	7.79787	10.6133	15.4669	20.2496	36.6851	46.9677
12	2.2145	2.70572	3.43077	3.97759	4.80453	5.43693	6.40129	8.27779	11.2915	14.0651	22.6738	27.5833
13	2.38161	2.91501	3.70115	4.29337	5.18821	5.87209	6.91445	8.94156	12.1954	15.1891	24.4785	29.7756
14	1.99866	2.43769	3.08486	3.57249	4.30944	4.87274	5.7314	7.40152	10.0826	12.5495	20.2047	24.57
15	2.13807	2.61408	3.31496	3.84257	4.63936	5.24806	6.17552	7.97851	10.8715	13.5326	21.7885	26.4957
16	2.14137	2.55938	3.15941	3.60077	4.25298	4.74103	5.46987	6.84343	8.95265	10.8126	16.2383	19.1608
17	2.25272	2.69515	3.32981	3.79638	4.48554	5.00105	5.77066	7.22054	9.44613	11.4083	17.1307	20.2126
18	1.98157	2.36597	2.91724	3.32247	3.92101	4.36873	5.03713	6.29635	8.22927	9.9334	14.9033	17.58
19	2.07853	2.48491	3.06745	3.49549	4.12748	4.60007	5.30544	6.63388	8.6724	10.4692	15.7086	18.53
20	2.08985	2.45792	2.97591	3.35025	3.89447	4.29553	4.88575	5.97359	7.59303	8.97966	12.8578	14.8666
21	2.17214	2.55624	3.09661	3.48699	4.05439	4.47244	5.08753	6.22094	7.90778	9.35186	13.3899	15.4814
22	1.96367	2.30792	2.79209	3.14181	3.65007	4.02451	4.57541	5.5905	7.10117	8.39438	12.0105	13.8834
23	2.03726	2.39621	2.90093	3.26541	3.79501	4.1851	4.75894	5.81607	7.38896	8.73523	12.4991	14.4483
24	2.05159	2.38312	2.84276	3.1705	3.6412	3.9841	4.48323	5.38809	6.70456	7.80754	10.7991	12.305
25	2.11639	2.45933	2.9347	3.27359	3.7602	4.11465	4.63053	5.56557	6.92568	8.06506	11.1549	12.7102
26	1.94765	2.26124	2.69579	3.00554	3.45024	3.77414	4.24552	5.09985	6.34247	7.38337	10.206	11.6268
27	2.00667	2.33085	2.78002	3.10014	3.55968	3.89435	4.38134	5.26384	6.54723	7.62215	10.5367	12.0035
28	2.02197	2.32542	2.74118	3.03452	3.45175	3.75298	4.18771	4.96571	6.07764	6.99369	9.42047	10.6156
29	2.07521	2.38725	2.81472	3.11629	3.54517	3.85477	4.30155	5.101	6.24341	7.18448	9.67729	10.9047
30	1.93377	2.2231	2.61938	2.89888	3.29635	3.58325	3.99723	4.73796	5.79636	6.66817	8.97737	10.1145

para todo o t, se, e só se,

$$\max_{1\le j\le n} \tau_n^{-1} |b_{j,n}| \longrightarrow 0, \quad n\to\infty$$

onde $\mu_{i:n} = E(X_{i:n})$, $b_{j,n} = (n-j+1)^{-1} \sum\limits_{i=j}^{n} a_{i:n}$ *e* $\tau_n^2 = \sum\limits_{i=1}^{n} b_{i,n}^2$.

Dem.
Se usarmos a representação de Rényi para as e.o. do modelo exponencial, tem-se

$$X_{i:n} \stackrel{d}{=} \sum_{j=1}^{i} \frac{X_j}{n-j+1} \tag{3.3.1}$$

seguindo-se que

$$\mu_{i:n} = \sum_{j=1}^{i} \frac{1}{n-j+1}. \tag{3.3.2}$$

Atendendo a (3.3.1) e (3.3.2), vem

$$\sum_{i=1}^{n} a_{i:n}(X_{i:n} - \mu_{i:n}) \stackrel{d}{=} \sum_{j=1}^{n} \frac{1}{n-j+1} \sum_{i=j}^{n} a_{i:n}(X_j - 1)$$

ou

$$\sum_{i=1}^{n} a_{i:n}(X_{i:n} - \mu_{i:n}) \stackrel{d}{=} \sum_{j=1}^{n} b_{j,n}(X_j - 1) \tag{3.3.3}$$

em que $b_{j,n} = (n-j+1)^{-1} \sum\limits_{i=j}^{n} a_{i:n}, \quad j = 1, \ldots, n$.

De (3.3.3) é imediato que

1. $E\left\{\sum\limits_{i=1}^{n} a_{i:n}(X_{i:n} - \mu_{i:n})\right\} = 0$
2. $Var\left\{\sum\limits_{i=1}^{n} a_{i:n}(X_{i:n} - \mu_{i:n})\right\} = \sum\limits_{j=1}^{n} b_{j,n}^2 = \tau_n^2$.

Aplicando o teorema limite central, segue-se que

$$P\left\{\tau_n^{-1}\sum_{j=1}^{n} b_{j,n}\left(X_j-1\right)\leq t\right\}\longrightarrow\Phi(t)\,,\quad n\rightarrow\infty$$

ou seja,

$$P\left\{\tau_n^{-1}\sum_{i=1}^{n} a_{i:n}\left(X_{i:n}-\mu_{i:n}\right)\leq t\right\}\longrightarrow\Phi(t),\quad n\rightarrow\infty \tag{3.3.4}$$

para todo o t*, se, e só se,*

$$\max_{1\leq j\leq n}\tau_n^{-1}\left|b_{j,n}\right|\longrightarrow 0\,,\quad n\rightarrow\infty. \tag{3.3.5}$$

A equivalência de (3.3.4) e (3.3.5) é um caso particular do teorema de Lindeberg-Lévy-Feller demonstrado em Chernoff et al. *(1967), Lema 1.*

• *Se* n *for ímpar, a estatística* T_n^* *está definida em* (3.2.1)*, e como podemos observar, o numerador é uma combinação linear de e.o. do modelo exponencial. Assim, usando a notação do lema 3.3.1, temos*

$$a_{i:n}=\begin{cases}-1 & ,\ i=\frac{n+1}{2}\\ 1 & ,\ i=n-\left\{\frac{n}{4}\right\}+1\\ 0 & ,\ \textit{caso contrário}\end{cases}\quad ,$$

$$b_{j,n}=\begin{cases}0 & ,\ 1\leq j\leq\frac{n+1}{2}\\ \frac{1}{n-j+1} & ,\ \frac{n+1}{2}+1\leq j\leq n-\left\{\frac{n}{4}\right\}+1\\ 0 & ,\ n-\left\{\frac{n}{4}\right\}+2\leq j\leq n\end{cases}$$

e

$$\tau_n^2=\sum_{j=\frac{n+1}{2}+1}^{n-\{n/4\}+1}\frac{1}{(n-j+1)^2}=\sum_{k=\{n/4\}}^{\frac{n-1}{2}}\frac{1}{k^2}. \tag{3.3.6}$$

Por outro lado,

$$X_{n-\{n/4\}+1:n} - X_{\frac{n+1}{2}:n} = \big(X_{n-\{n/4\}+1:n} - \mu_{n-\{n/4\}+1:n}\big) \\ - (X_{\frac{n+1}{2}:n} - \mu_{\frac{n+1}{2}:n}) + \lambda_n$$

onde

$$\lambda_n = \mu_{n-\{n/4\}+1:n} - \mu_{\frac{n+1}{2}:n} = \sum_{j=\frac{n+1}{2}+1}^{n-\{n/4\}+1} \frac{1}{n-j+1} = \sum_{k=\{n/4\}}^{\frac{n-1}{2}} \frac{1}{k}. \tag{3.3.7}$$

Aplicando o lema 3.3.1, vem

$$\tau_n^{-1}\left(X_{n-\{n/4\}+1:n} - X_{\frac{n+1}{2}:n} - \lambda_n\right) \xrightarrow[n\to\infty]{w} Z \frown Gau(0,1).$$

Por outro lado, como o denominador de (3.2.1) *converge em probabilidade para* $\ln(3/2)$*, segue-se que*

$$\frac{\ln(3/2)}{X_{\frac{n+1}{2}:n} - X_{\{n/4\}:n}} \xrightarrow{P} 1.$$

Aplicando o teorema de Slutsky, vem

$$\frac{\ln(3/2)}{X_{\frac{n+1}{2}:n} - X_{\{n/4\}:n}}(X_{n-\{n/4\}+1:n} - X_{\frac{n+1}{2}:n}) = \ln(3/2)T_n^* \xrightarrow[n\to\infty]{w} Y \frown Gau(\lambda_n, \tau_n).$$

Donde, para valores de n *ímpar elevados podemos usar a aproximação*

$$\ln(3/2)\tau_n^{-1}\left(\frac{X_{n-\{n/4\}+1:n} - X_{\frac{n+1}{2}:n}}{X_{\frac{n+1}{2}:n} - X_{\{n/4\}:n}} - \frac{\lambda_n}{\ln(3/2)}\right) \sim Z \frown Gau(0,1). \tag{3.3.8}$$

- *Se n for par, consideremos T_n^* definido em* (3.2.4). *Neste caso, temos*

$$a_{i:n} = \begin{cases} -1 & , \ i = \frac{n}{2}, \ \frac{n}{2}+1 \\ 2 & , \ i = n - \left\{\frac{n}{4}\right\} + 1 \\ 0 & , \ \textit{caso contrário} \end{cases} ,$$

$$b_{j,n} = \begin{cases} 0 & , \ 1 \leq j \leq \frac{n}{2} \\ \frac{2}{n} & , \ j = \frac{n}{2}+1 \\ \frac{2}{n-j+1} & , \ \frac{n}{2}+2 \leq j \leq n - \left\{\frac{n}{4}\right\} + 1 \\ 0 & , \ n - \left\{\frac{n}{4}\right\} + 2 \leq j \leq n \end{cases}$$

e

$$\tau_n^2 = 4\left[\frac{1}{n^2} + \sum_{j=\frac{n}{2}+2}^{n-\{n/4\}+1} \frac{1}{(n-j+1)^2}\right] = 4\left[\frac{1}{n^2} + \sum_{k=\{n/4\}}^{\frac{n}{2}-1} \frac{1}{k^2}\right]. \tag{3.3.9}$$

Como

$$\begin{aligned} 2X_{n-\{n/4\}+1:n} - X_{\frac{n}{2}:n} - X_{\frac{n}{2}+1:n} = 2&\left(X_{n-\{n/4\}+1:n} - \mu_{n-\{n/4\}+1:n}\right) \\ &- \left(X_{\frac{n}{2}:n} - \mu_{\frac{n}{2}:n}\right) - \left(X_{\frac{n}{2}+1:n} - \mu_{\frac{n}{2}+1:n}\right) \\ &+ \lambda_n^* \end{aligned}$$

onde

$$\lambda_n^* = 2\mu_{n-\{n/4\}+1:n} - \mu_{\frac{n}{2}:n} - \mu_{\frac{n}{2}+1:n} = 2\sum_{k=\{n/4\}}^{n/2} \frac{1}{k} - \frac{2}{n}. \tag{3.3.10}$$

Decorre do lema 3.3.1 que

$$\tau_n^{-1}\left(2X_{n-\{n/4\}+1:n} - X_{\frac{n}{2}:n} - X_{\frac{n}{2}+1:n} - \lambda_n^*\right) \xrightarrow[n\to\infty]{w} Z \frown Gau(0,1).$$

Dado que o denominador de (3.2.4) *converge em probabilidade para* $2\ln(3/2)$*, segue do teorema de Slutsky que*

$$2\ln(3/2)T_n^* \xrightarrow[n\to\infty]{w} Y \frown Gau(\lambda_n^*, \tau_n).$$

Consequentemente, para valores de n par elevados podemos usar a aproximação

$$\ln(3/2)\,\tau_n^{*^{-1}}\left(\frac{2X_{n-\{n/4\}+1:n}-X_{\frac{n}{2}:n}-X_{\frac{n}{2}+1:n}}{X_{\frac{n}{2}:n}+X_{\frac{n}{2}+1:n}-2X_{\{n/4\}:n}}-\frac{\lambda_n^*}{2\ln(3/2)}\right)\sim Z\frown Gau(0,1) \tag{3.3.11}$$

onde $\tau_n^{*2}=\frac{1}{4}\tau_n^2$.

Podemos, no entanto, considerar outras constantes de atracção.

No modelo exponencial standard sabe-se que é válida a identidade

$$\mu_{k:n}=\psi(n+1)-\psi(n-k+1),\quad k=1,\ldots,n$$

onde ψ é a função digama.

Como $\psi(n)$ é assintóticamente equivalente a $\ln n$ ($\psi(n)\sim\ln n$), então se n for ímpar

$$\lambda_n=\psi\left(\tfrac{n+1}{2}\right)-\psi\left(\left\{\tfrac{n}{4}\right\}\right)\sim\ln\left(\frac{n}{2\left\{\frac{n}{4}\right\}}\right)\sim\ln 2.$$

Por outro lado, usando a fórmula (0.0.10) *para $m=1$, obtém-se*

$$\tau_n^2=\sum_{k=\{n/4\}}^{\frac{n-1}{2}}\frac{1}{k^2}=\psi'\left(\left\{\tfrac{n}{4}\right\}\right)-\psi'\left(\tfrac{n+1}{2}\right),$$

e dado que $\psi'(z)\sim\frac{1}{z}$, vem

$$\tau_n^2\sim\frac{2}{n}.$$

Consequentemente,

$$\ln(3/2)\sqrt{\frac{n}{2}}\left(\frac{X_{n-\{n/4\}+1:n}-X_{\frac{n+1}{2}:n}}{X_{\frac{n+1}{2}:n}-X_{\{n/4\}:n}}-\frac{\ln 2}{\ln(3/2)}\right)\sim Z\frown Gau(0,1). \tag{3.3.12}$$

Se n for par, tem-se

$$\lambda_n^* = 2\left[\psi\left(\frac{n}{2}\right) - \psi\left(\left\{\tfrac{n}{4}\right\}\right)\right] + \frac{2}{n} \sim 2\ln\left(\frac{n}{2\left\{\frac{n}{4}\right\}}\right) \sim 2\ln 2$$

e, por sua vez,

$$\tau_n^{*2} = \frac{1}{n^2} + \sum_{k=\{n/4\}}^{\frac{n}{2}-1} \frac{1}{k^2} = \frac{1}{n^2} + \psi'\left(\left\{\tfrac{n}{4}\right\}\right) - \psi'\left(\frac{n}{2}\right) \sim \frac{2}{n}.$$

Donde,

$$\ln(3/2)\sqrt{\frac{n}{2}}\left(\frac{2X_{n-\{n/4\}+1:n} - X_{\frac{n}{2}:n} - X_{\frac{n}{2}+1:n}}{X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n} - 2X_{\{n/4\}:n}} - \frac{\ln 2}{\ln(3/2)}\right) \sim Z \frown Gau(0,1). \tag{3.3.13}$$

Com as novas constantes de atracção podemos aproximar a distribuição de T_n^* pela mesma Gaussiana independentemente de n ser ímpar ou par (recorde-se que o valor populacional de $\frac{F_U-M}{M-F_L}$ com parente exponencial é $\frac{\ln 2}{\ln(3/2)}$).

Refira-se que a aproximação pela Gaussiana não é das melhores mesmo para $n = 30$ por a distribuição ainda ser bastante assimétrica. No entanto, não nos foi possível obter quantis para valores de n superiores a 30 por termos começado a observar instabilidade nos cálculos dos quantis. Esta situação já nos tinha ocorrido em trabalho anterior (Brilhante, Pestana e Rocha (1996)), e pensamos que deve-se a situações modernamente abordadas no âmbito de investigação do caos.

Bibliografia

[1] *Abramowitz, M. and Stegun, I. A. (1970).* Handbook of Mathematical Functions. *Dover, New York.*

[2] *Athayde, M. Emília F. Q. (1984).* Estudos Sobre Multiplicação de Variáveis Aleatórias Independentes. *Dissertação de Doutoramento, FCUL, Lisboa.*

[3] *Balakrishnan, N. and Basu, A. P. (eds.) (1995).* The Exponential Distribution. Theory, Methods and Applications. *Gordon and Breach Publishers, Amsterdam.*

[4] *Balkema, A. A. and Resnick, S. I. (1977). Max-infinite divisibility.* J. Appl. Prob., **14**, *309–319.*

[5] *Barão, I. (1992). Comparação estatística de populações Gumbel. In M. A. A. Turkman e M. L. S. Carvalho (eds.)* Actas da 1ª Conferência em Estatística e Optimização. *CEAUL, Lisboa.*

[6] *Bartholomew, D. J. (1969). Sufficient conditions for a mixture of exponentials to be a probability density function.* Annals of Mathematical Statistics, **40**, *2183–2188.*

[7] *Bingham, N. H. (1971). Factorization theory and domains of attraction for generalized convolution algebras.* Proc. London Math. Soc. *(3)*, **23**, *16–30.*

[8] *Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987).* Regular Variation. *University Press, Cambridge.*

[9] *Bondessen L. (1979). A general result on infinite divisibility.* Annals of Probability, **6**, *965–979.*

[10] *Brilhante, M. F. (1996). Inferência sobre o parâmetro de localização de uma população exponencial I. Studentização Externa. In* A Estatística a Decifrar o Mundo, *47–55, Salamandra, Lisboa.*

[11] *Brilhante, M. F., Pestana, D. D. e Rocha, J. (1996). Inferência sobre o parâmetro de localização de uma população exponencial II. Studentização Interna. In* A Estatística a Decifrar o Mundo, *57–63, Salamandra, Lisboa.*

[12] *Brilhante, M. F. (1998). Spacings of a Laplace model as generalized mixtures of spacings of an exponential model.* Notas e Comunicações do CEAUL, *Nota $n\frac{1}{4}$ 17.*

[13] *Brilhante, M. F. (1999). A resistant test statistic for testing exponentiality versus generalized Pareto. In* Extreme Values and Additive Laws, *11–14, Lisboa.*

[14] *Burrill, C. W. (1972).* Measure, Integration and Probability, *Mcgraw-Hill, New York.*

[15] *Casella, G. and Berger, R. L. (1990).* Statistical Inference. *The Wadsworth & Brooks/Cole Publishing Company, California.*

[16] *Chernoff, H., Gastwirth, J. L. and Johns, M. V. (1967). Asymptotic distribution of linear combinations of order statistics with applications to estimation.* Ann. Math. Statist. **38***, 52–72.*

[17] *Choquet, G. (1960). Le théorème de la représentation intégrale dans les ensembles convexes compacts.* Ann. Inst. Fourier, **10***, 333–344.*

[18] *Chow, T. L. and Teugels, J. L. (1979). The sum and the maximum of iid random variables. In P. Mandl and M. Huskova (eds).* Asymptotic Statistics, Proc. 2nd Symp. NosHolland*, 81–92.*

[19] *Cuppens, R. (1975).* Decomposition of Multivariate Probabilities. *Academic Press, New York.*

[20] *David, H. A. (1981).* Order Statistics. *John Wiley & Sons, New York.*

[21] *Davidson, R. (1968). Arithmetic and other properties of certain Delphic semi-groups,* Z. Wahrsch. verw. Gebiete*,* **10***, 120–145.*

[22] *Deheuvels, P. (1978).* Caracterization compléte de loi extreme multivariée et de la convergence des tipes extremes. *Publication de L'Institut de Statistique de la Université de Paris,* **23***, 1–36.*

[23] *Diamantino, F. e Pestana, D. D. (1996). Perturbações da Gaussiana - sua influência em estatísticas studentizadas. In* A Estatística a Decifrar o Mundo*, 65–71, Salamandra, Lisboa.*

[24] *Doeblin, W. (1940). Sur l'ensemble des puissances d'une loi de probabilité.* Studia Math.*,* **9***, 71–96.*

[25] *Feller, W. (1967). On regular variation and local limit theorems.* Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, II*, Univ. Califórnia Press, Los Angeles, 373–378.*

[26] *Feller, W. (1971).* An Introduction to Probability Theory and Its Applications. *vol. II. John Wiley & Sons, New York.*

[27] *Finetti, B. de (1930). Le funzioni caratteristische di legge instantanea.* Rend. Ac. Lincei*,* **12***, 278–282.*

[28] *Fisher, R. A. (1925).* Statistical Methods for Research Workers. *Oliver and Boyd, Edinburgh.*

[29] *Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of frequency of the largest or smallest member of a sample.* Proc. Camb. Phil. Soc.*,* **24***, 180–190.*

[30] *Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum.* Ann. Soc. Polonaise Math.*,* **6***, 93–116.*

[31] *Galambos, J. (1987).* The Asymptotic Theory of Extreme Order Statistics. *Robert E. Krieger Publishing Company, INC, Florida.*

[32] *Giné, E. and Araújo, A. (1980).* The Central Limit Theorem for Real and Banach Valued Random Variables. *John Wiley & Sons, New York.*

[33] *Gnedenko, B. V. (1940). On the theory of domains of attraction of stable laws.* Uchemye Zapiski, Moskov Gos. Univ.*,* **30***, 61–72.*

[34] *Gnedenko, B. V. (1943). Sur la distribution limite du term maximum d'une série aléatoire.* Ann. Math.*,* **44***, 423–453.*

[35] *Gnedenko, B. V. and Kolmogoroff, A. N. (1954).* Limit Distributions for Sums of Independent Random Variables. *Addison-Wesley, Massachusetts.*

[36] *Goldie, C. M. (1967). A class of infinitely divisible distributions.* Proc. Cambridge Philos. Soc., **63**, *1141–1143.*

[37] *Gomes, M. I. (1978).* Some Probabilistic and Statistical Problems in Extreme Value Theory. *Ph D Thesis, University of Sheffield.*

[38] *Gomes, M. I. (1982). A note on statistical choice of extremal models.* Actas IX Jornadas Mat. Hispano-Lusas, *Salamanca, 653–655.*

[39] *Gomes, M. I. and Pestana, D. D. (1985). Domains of Attraction and Penultimate Behavior.* Abs. 16th European Meeting of Statisticians, *202–203.*

[40] *Gomes, M. I. and Pestana, D. D. (1987). Non-Standard Domains of Attraction and Rates of Convergence. In* New perspectives in Theoritical and Applied Statistics *(José P. Vilaplana, ed.), John Wiley & Sons, New York, 467–477.*

[41] *Gomes, M. I. and van Monfort, M. A. J. (1987). Exponentiality versus Generalized Pareto – Quick Tests.* Statistical Climatology, **87**, *185–195.*

[42] *Graça Martins, E. (1983).* Modelos Estáveis, I – Estudo Preliminar com Vista às Aplicações. *FCUL, Lisboa,*

[43] *Graça Martins, E. and Pestana, D. D. (1985). The extremal limit problem - extensions.* V Pannonian Symp. Math. Statist., *Viségrad.*

[44] *Graça Martins, E. and D. D. (1987). Nonstable Limit Laws in Extreme Value Theory. In* New perspectives in Theoritical and Applied Statistics *(José P. Vilaplana, ed.), John Wiley & Sons, New York, 449–457.*

[45] *Gradshteyn, I. S. and Ryzhik, I. M. (1994).* Table of Integrals, Series, and Products. *5th ed. Academic Press, California.*

[46] *Gumbel, E. J. (1965). A quick estimation of the parameters in Fréchet's distribution.* Review Intern. Statist. Inst., **33**, *349–363.*

[47] *Haan, L. de (1970). On regular variation and its application to the weak convergence of sample extremes.* Math. Centrum, *Amsterdam.*

[48] *Harris, C. M., Marchal, W. G. and Botta, R. F. (1992). A note on generalized hyperexponential distributions.* Communications in Statistics-Stochastic Models, **8**, *179–191.*

[49] *Heyer, H. (1977).* Probability Measures on Locally Compact Groups. *Springer-Verlag, Berlin.*

[50] *Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (1992).* Análise Exploratória de Dados. Técnicas Robustas - Um Guia. *Colecção Novas Tecnologias, Ed. Salamandra, Lisboa.*

[51] *Hoffmann-Jorgensen, J., Liggett, T. M. et Neveu, J. (1976).* École d'eté de probabilités de Saint-flour VI.

[52] *Hotelling, H. (1961). The behavior of some standard statistical tests under nonstandard conditions.* Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability, I, *319–359.*

[53] *Ibragimov, I. A. and Linnik, Yu. V. (1971).* Independent and Stationary Sequences of Random Variables. *Wolters-Noordhoff, Groningen.*

[54] *Iglésias Pereira, H., Oliveira, O. e Pestana, D. (1996). Limites estáveis e comportamentos pré-assintóticos. In* A Estatística a Decifrar o Mundo*, 109–116, Salamandra, Lisboa.*

[55] *Iglewicz, B. (1981). Robust scale estimators and confidence intervals for location. In D. Hoaglin, F. Mosteller and J. Tukey (eds)* Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, New York.*

[56] *Johansen, S. (1966). An application of extreme point methods to the representation of infinitely divisible distributions.* Z. Wahrsch. verw. Geb.*,* **5***, 304–316.*

[57] *Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994).* Continuous Univariate Distributions. *2nd ed. John Wiley & Sons, New York.*

[58] *Jurek, Z. J. (1981). The problem of B. V. Gnedenko on Banach spaces.* Bull. Acad. Polonaise Sc., Ser. Sc. Math.*,* **29***, 399–407.*

[59] *Karamata, J. (1930). Sur un mode de croissance régulière des fonctions.* Mathematica (Cluj), **4**, *38–53.*

[60] *Keilson, J. and Steutel, F. W. (1972). Families of infinitely divisible distributions closed under mixing and convolution.* Annals of Mathematical Statistics, **43**, *242–250.*

[61] *Kendall, D. G. (1963). Extreme point methods in stochastic analysis.* Z. Wahrsch. verv. Geb.*,* **1**, *295–300.*

[62] *Kendall, D. G. (1967). Delphic semi-groups* Bull. Ann. Math. Soc.*,* **73***, 120–121.*

[63] *Kendall, D. G. and Harding, E. F. (1973).* Stochastic Analysis. A tribute to the memory of Rolo Davidson. *John Wiley & Sons, New York.*

[64] *Kingman, J. F. C. (1963). Random walks with spherical symmetry.* Acta Mathematica*,* **109***, 11–53.*

[65] *Krein, M. and Milman, D. (1940). On the extreme points of regular convex sets.* Studia Mathematica*,* **9***, 133–138.*

[66] *Kruglov, V. M. (1972). On an extension of the class of stable distributions.* Theory and Probability Applications*,* **17***, 723–732.*

[67] *Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1982).* Extreme and Related Properties of Random Sequences and Processes. *Springer-Verlag, New York.*

[68] *Lévy, Paul (1925).* Calcul de Probabilités. *Gauthier-Villars, Paris.*

[69] *Lévy, Paul (1934). Sur les intégrales dont les élements sont des variables aléatoires independentes.* Ann. Scuola Norm. Pisa *(2),* **3***, 337–366.*

[70] *Lévy, Paul (1937).* Théorie de L'Addition des Variables Aléatoires. *Gauthier-Villars, Paris.*

[71] *Lévy, Paul (1965).* Processus Stochastiques et Mouvement Brownien. *Gauthier-Villars, Paris.*

[72] *Linnik, Yu. V. and Ostrovski, I. V. (1977).* Decomposition of Random Variables and Vectors. *American Mathematical, Rhode Island.*

[73] *Lukacs, E. (1970).* Characteristic Functions. *2nd ed. Griffin, London.*

[74] *Malmquist, S. (1950). On a property of order statistics from a rectangular distribution.* Skand. Aktuar.*,* **33***, 214–222.*

[75] *Masani, T. (1970). On helices in Hilbert spaces, I.* Theory of Probability and Its Applications*,* **15***, 1–20.*

[76] *Mejzler, D. (1956). On the problem of the limit distribution for the maximal term of a variational series.* L'vov Pol. Inst. Naucn Zp.*,* **38***, 90–109.*

[77] *Mosteller, F. and Tukey, J. W. (1977).* Data Analysis and Regression: a Second Course in Statistics. *Addison-Wesley, Massachusetts.*

[78] *Parthasarathy, K. R. (1967).* Probability Measures on Metric Spaces. *Academic Press, New York.*

[79] *Pestana, D. D. F. (1978).* Some Contributions to Unimodality, Infinite Divisibility, and Related Topics. *Ph D Thesis, University of Sheffield.*

[80] *Pestana, D. (1984). A note on Pólya's theorem.* Trabajos de Estadistica y de Investigacion 0perativa*,* **35***, 104–111.*

[81] *Pestana, D. e Rocha, J. (1992). Distribuições amostrais em situações não Gaussianas. In D. Pestana (ed.)* Estatística Robusta, Extremos e Mais Alguns Temas*, 73–80, Salamandra, Lisboa.*

[82] *Pestana, D. e Rocha, J. (1993). Análise de Escala I: Modelo Exponencial. In* A Estatística e o Futuro da Estatística*, Salamandra, Lisboa, 295–303.*

[83] *Pestana, D. D. and Rocha, J. (1995). Internally studentized statistics with exponential parent.* Notas e Comunicações do CEAUL

[84] *Pestana, D., Brilhante, F. and Rocha, J. (1999). The analysis of variance revisited. In* Extreme Values and Additive Laws*, 73–77, Lisboa.*

[85] *Petrov, V. V. (1995).* Limit Theorems of Probability Theory: Sequences of Independent Random Variables. *Clarendon, Oxford.*

[86] *Phelps, R. R. (1966).* Lectures on Choquet's Theorem. *D. van Nostrand Company, Inc., New York.*

[87] *Reiss, R-D. (1989).* Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics. *Springer-Verlag, New York.*

[88] *Rényi, A. (1953). On the theory of order statistics.* Acta Mathematica Academiae Scientarium Hungaricae*,* **4***, 191–231.*

[89] *Resnick, S. I. (1971). Tail equivalence and its applications.* J. Appl. Prob.*,* **8***, 135–156.*

[90] *Resnick, S. I. (1987).* Extreme Values, Regular Variation, and Point Processes. Applied Probability. *Springer-Verlag, New York.*

[91] *Rocha, J. (1995).* Localização e Escala em Situações não Clássicas. *Dissertação de Doutoramento, Universidade dos Açores, Ponta Delgada.*

[92] *Ross, S. M. (1997).* Introduction to Probability Models. *6th ed. Academic Press, San Diego.*

[93] *Shanbag, D. N., Pestana, D. and Sreehari, M. (1977). Some further results in infinite divisibility.* Math. Proc. Camb. Phil. Soc.*,* **82***, 289–295.*

[94] *Smirnov, N. V. (1952). Limit distributions for the terms of a variational series.* Amer. Math. Soc. Transl. Ser., **1**, *No. 67.*

[95] *Spitzer, F. (1961).* Random Walks. *van Nostrand.*

[96] *Sreehari, M. (1970). A limit theorem for the maximum of normalized partial sums.* Z. Wahrsch. verw. Geb, *22–28.*

[97] *Steutel, F. W. (1967). Note on the infinite divisibility of exponential mixtures.* Annals of Mathematical Statistics, **38**, *1303–1305.*

[98] *Steutel, F. W. (1968). A class of infinitely divisible mixtures.* Annals of Mathematical Statistics, **39**, *1153–1157.*

[99] *Steutel, F. W. (1969). Note on completely monotone densities.* Annals of Mathematical Statistics, **40**, *1130–1131.*

[100] *Steutel, F. W. (1972). Families of infinitely divisible distributions closed under mixing and convolution.* Annals of Mathematical Statistics, **43**, *242–250.*

[101] *Student (1908). On the probable error of the mean.* Biometrika, **6**, *1–25.*

[102] *Tiago de Oliveira, J. and Gomes, M. I. (1984). Two test statistics for choice of univariate extreme models. In* Statistical Extremes and Applications, *ed. J. Tiago Oliveira, 651–668. D. Reidel.*

[103] *Tucker, H. G. (1968). Convolutions of distributions attracted to stable laws.* Ann. Math. Statist., **39**, *1381–1390.*

[104] *Tukey, J. W. (1977).* Exploratory Data Analysis. *Addison-Wesley, Massachusetts.*

[105] *Urbanik, K. (1973).* Limit Laws for Sequences of Normed Sums Satisfying Some Stability Conditions, *Multivariate Analysis, III, Academic Press, New York, 225–237.*

[106] *van Monfort, M. A. J. and Witter, J. V. (1985). Testing exponentiality against Generalized Pareto distribution. Accepted by J. Hydrology.*

[107] *Watts, V. (1980). The almost sure representation of intermediate order statistics.* Z. Wahrsch. verw. Geb., **54**, *281–285.*

[108] *Watts, V., Rootzén, H. and Leadbetter, M. R. (1982). Convergence in distribution of intermediate order statistics from stationary sequences.* Ann. of Prob., **10**, *653–662.*

[109] *Zolotarev, V. M. (1957). Mellin-Stieljes transforms in probability theory.* Teor. Veroyatnost. i. Primenen, **2**, *444–469.*

[110] *Zolotarev, V. M. and Koroljuk, V. S. (1961). On the hypothesis proposed by B. V. Gnedenko.* Theor. Probab. Appl., **6**, *431–435.*

Apêndice

A.1 Selecção de Funções Densidade de Probabilidade

(funções densidade de probabilidade de $\frac{F_U - M}{M - F_L}$ para alguns valores de n)

♦ n ímpar

- $n = 3$

$$f(t) = \frac{2}{(2+t)^2}, \quad t > 0$$

- $n = 5$

$$f(t) = \frac{6t(888 + 1533t + 978t^2 + 273t^3 + 28t^4}{(2+t)^2(3+t)^2(4+t)^2(3+2t)^2}, \quad t > 0$$

- $n = 7$

$$f(t) = \frac{30t(2440 + 5445t + 4530t^2 + 1665t^3 + 228t^4}{(2+t)^2(5+2t)^2(4+3t)^2(5+3t)^2}, \quad t > 0$$

- $n = 9$

$$\begin{aligned} f(t) = [&70t^2(30443751450 + 184811950890t + 511471520591t^2 + 853294307652t^3 + \\ &955743364852t^4 + 757134957906t^5 + 434985020859t^6 + 182605087584t^7 + \\ &55589991928t^8 + 11967978528t^9 + 1729548464t^{10} + 150633216t^{11} + 5978880t^{12})] \\ &\div [(2+t)^2(3+t)^2(3+2t)^2(5+2t)^2(7+2t)^2(5+3t)^2(7+3t)^2(5+4t)^2(7+4t)^2], \quad t > 0 \end{aligned}$$

- $n = 11$

$$\begin{aligned}
f(t) = [&70t^2(49307892480 + 291713393664t + 773536802012t^2 + 1210612521844t^3 + \\
&1238257945135t^4 + 864848002420t^5 + 417686431126t^6 + 137724491788t^7 + \\
&29669377271t^8 + 3770389260t^9 + 214612200t^{10})] \\
&\div [(2+t)^2(3+2t)^2(7+3t)^2(8+3t)^2(7+4t)^2(6+5t)^2(7+5t)^2(8+5t)^2], \quad t > 0
\end{aligned}$$

- $n = 13$

$$\begin{aligned}
f(t) = [&126t^3(4217114514281227776000 + 5005778951404670866560 0t + \\
&2827414098703911028155520t^2 + 10109195155453957968795 08t^3 + \\
&25675188116358165389566604t^4 + 49271935522432293706001 21t^5 + \\
&7419178591128988034537918t^6 + 8984757052438139440348181t^7 + \\
&8898180715571069288724004t^8 + 7288537976238322350567994t^9 + \\
&4973848530008453133095860t^{10} + 2839488327315380841029542t^{11} + \\
&1357862525628154914462460t^{12} + 543196772611209789350869t^{13} + \\
&181030611743506904067486t^{14} + 49895564984227631625921t^{15} + \\
&11245460343378752797812t^{16} + 2038158870372226958616t^{17} + \\
&289808006019501018816t^{18} + 31131716953273548048t^{19} + \\
&2375128216945811520t^{20} + 114681064330425600t^{21} + 2634261025536000t^{22})] \\
&\div [(2+t)^2(3+t)^2(3+2t)^2(5+2t)^2(4+3t)^2(5+3t)^2(7+3t)^2(8+3t)^2(10+3t)^2 \\
&\times (7+4t)^2(9+4t)^2(7+5t)^2(8+5t)^2(9+5t)^2(7+6t)^2], \quad t > 0
\end{aligned}$$

- $n = 15$

$$
\begin{aligned}
f(t) = {} & [462t^3(564914779092140133888000 + 7365360345780762557971200t+ \\
& 45741695647458739170512160t^2 + 180006205843505321621624756t^3 \\
& 503711515958214295810881132t^4 + 1066144705345111032688058825t^5+ \\
& 1772446424844022283920780062t^6 + 2372347347356548383478616069t^7+ \\
& 2599459017902893230945165876t^8 + 2358241602848060893866014234t^9+ \\
& 1784283474142920531732519732t^{10} + 1130558257779370885805867558t^{11}+ \\
& 600688127235009692452637772t^{12} + 267269974549813839067984037t^{13}+ \\
& 99175523896001744949529470t^{14} + 30467296430213086044317873t^{15}+ \\
& 7661764895367915379207140t^{16} + 1551062734605534450178904t^{17}+ \\
& 246604801150152932069376t^{18} + 29651999824962747411344t^{19}+ \\
& 2534880820970105993280t^{20} + 137290804363988755200t^{21}+ \\
& 3541175129270016000t^{22})] \\
& \div [(2+t)^2(3+2t)^2(5+2t)^2(4+3t)^2(5+3t)^2(9+4t)^2(11+4t)^2(8+5t)^2 \\
& \quad \times (9+5t)^2(11+5t)^2(11+6t)^2(8+7t)^2(9+7t)^2(10+7t)^2(11+7t)^2], \quad t > 0
\end{aligned}
$$

- $n = 17$

$$
\begin{aligned}
f(t) = [6006t^4(&79170765197007884167458916186685724000000+\\
&15656989254661497278837637032505195816000 0t+\\
&150249403846137103894940615838553475684700 0t^2+\\
&932099505099389326683185627422988794294920 0t^3+\\
&420178002072765761456649855780827086122373 62t^4+\\
&146674763000728558000131042188529158330979 822t^5+\\
&412603338699904507362781354424786300718086691t^6+\\
&961053089350155533884581421602609506379981024t^7+\\
&1890040252590033658878201216118159059250951761t^8+\\
&3184389320286090215859698061015950683502411358t^9+\\
&4647860476665066605756142133357073640912312599t^{10}+\\
&5928162675068851161071867314372675179401705436t^{11}+\\
&6652548608482775186675169095162237941861785757t^{12}+\\
&6603702858779788844497724827842019841158733630t^{13}+\\
&5822920440257475588302266606318536055336549365t^{14}+\\
&4575592096477477940490346248454895677887541960t^{15}+\\
&3211756708054336160326511964614442345338052555t^{16}+\\
&2017159476750421300951086578261854643751710610t^{17}+\\
&1134639279785673879686566518910854812946196605t^{18}+\\
&571784806522391620071694945308691539202319820t^{19}+\\
&258063310084769567957654894182521609314753445t^{20}+\\
&104212524079274801539128317992512593264900580t^{21}+\\
&37592315228416505625393957768271573449766140t^{22}+\\
&12084448339823001388274839505353057778643360t^{23}+\\
&3450665060503452849233512972021840285723248t^{24}+\\
&871572526303228849783211996278115469999168t^{25}+\\
&193685569944265860222125787462077667050304t^{26}+\\
&37611622495562659310044421094250311008256t^{27}+\\
&6327155643360489472721897575606569432064t^{28}+\\
&911832545933848428539038513525688082432t^{29}+\\
&110947684947101995558260738016104759296t^{30}+\\
&11177510935546553831140135898697170944t^{31}+
\end{aligned}
$$

$n = 17$ *(continuação)*

$$
\begin{aligned}
&907428624550298731605347028227063808t^{32}+\\
&570446590864516595340959140020224000t^{33}+\\
&26056851732569606606179868344320000t^{34}+\\
&7694061494689818750398496768000000t^{35}+\\
&1102379183861653325846937600000t^{36})]\\
&\div [(2+t)^2(3+t)^2(3+2t)^2(5+2t)^2(5+3t)^2(5+4t)^2(9+4t)^2(11+4t)^2\\
&\quad \times (13+4t)^2(9+5t)^2(11+5t)^2(12+5t)^2(13+5t)^2(11+6t)^2(13+6t)^2\\
&\qquad \times (9+7t)^2(10+7t)^2(11+7t)^2(12+7t)^2(13+7t)^2(9+8t)^2(11+8t)^2(13+8t)^2] \;\; t > 0
\end{aligned}
$$

- $n = 19$

$$
\begin{aligned}
f(t) = [858t^4(&120730451911063409387040559497234708480000 0+\\
&26070912677989276828222300734671573667840000t+\\
&27332506695431521546461246258847945197388800 0t^2+\\
&18533998736765609162287544533066828761268480 00t^3+\\
&913693252296811348266738161959132732148023270 4t^4+\\
&348980805156178881063109151766322120454162385 92t^5+\\
&107467061493700377752081820977922891624792975360t^6+\\
&274159067508890014700742592357702340524481655040t^7+\\
&590817895468576045687595828790045135103824786208t^8+\\
&1091313398028142687268741100552953146244539222432t^9+\\
&1747145855052326037348826857057144385921201100344t^{10}+\\
&2445454380419337791747894124768248366240042240620t^{11}+\\
&3013006744244853653691169552739358567461932198927t^{12}+\\
&3285337367598977351870945499178450487709493766392t^{13}+\\
&3183611701133994137054661511585623206960411450644t^{14}+\\
&2750541926992765124690538175137866755987481635528t^{15}+\\
&2123776629508025442398714757226332481097862747136t^{16}+\\
&1467924428639806761821535442400482848444963434152t^{17}+\\
&909115591409183948505982602555253805489816910844t^{18}+\\
&504650619957912107352211920867024747266855156608t^{19}+\\
&251003025499491017111394738153520243139655567526t^{20}+\\
&111754026557486186857946802746850082421712312792t^{21}+
\end{aligned}
$$

$n = 19$ *(continuação)*

$$
\begin{aligned}
&4446593981912160556524047656763778405286296736 4t^{22}+\\
&157736939344781819316315866095865233050249804 08t^{23}+\\
&49725490369861162483512680928009963419876378 32t^{24}+\\
&13872012121063852579137102211701826582669426 80t^{25}+\\
&3406287343060270571511588097405772916188322 84t^{26}+\\
&7312081956898752254450082625799083473063790 8t^{27}+\\
&136034102737894916674406116875076832430924 03t^{28}+\\
&21689944579510962933203776358863238648829 60t^{29}+\\
&2921109921300757217899765869549384560031 60t^{30}+\\
&325868019750870802074916579802865691658 88t^{31}+\\
&29305911580152101543358691301359301372 64t^{32}+\\
&2041650865123450501948446335814273600 00t^{33}+\\
&103392310643000860753459521111798720 00t^{34}+\\
&3386073171662616634932752177356800 00t^{35}+\\
&538293529778795788399535508480000 0t^{36})]\\
&\div [(2+t)^2(3+2t)^2(4+3t)^2(5+3t)^2(7+3t)^2(5+4t)^2(7+4t)^2(11+5t)^2\\
&\quad \times (12+5t)^2(13+5t)^2(14+5t)^2(11+6t)^2(13+6t)^2(10+7t)^2(11+7t)^2(12+7t)^2\\
&\quad\quad \times (13+7t)^2(11+8t)^2(13+8t)^2(10+9t)^2(11+9t)^2(13+9t)^2(14+9t)^2], \quad t>0
\end{aligned}
$$

- $n = 21$

$$
\begin{aligned}
f(t) = [858t^5(&25644791417031985846352403685908753498408229721340005842944 0000+\\
&78092328501905682899374797231074896894640465995105245855744 00000t+\\
&11656491014768186321007421511016199208209553166929372491913 6256000t^2+\\
&11367402642094081241589975338648502592289788526637169384976 54784000t^3+\\
&81447932122549171798353908534461908913048488093437549298253 00611072t^4+\\
&45717826918467626286921857774683526999521981455049074850926 068760576t^5+\\
&20933043055297078514315011988270343786999343186453947486594 6499350528t^6+\\
&80384721767582008074418460089421244329431329747522523588105 2006236160t^7+\\
&26416474917448277209216213857244972880059931619753822049747 21118502912t^8+\\
&75436013251836954347784070157329976526934983024474656675245 086244116 48t^9+\\
&18944117048818348971029736922370370268599760162358696134508 953900599808t^{10}+
\end{aligned}
$$

$n = 21$ *(continuação)*

$$4223880122871193782196156079908268140057754243129597835736557176795526 4t^{11}+$$
$$842693454872691827460816597423603082796141280754029643282831689805584 00t^{12}+$$
$$1514047448579276199232709869047430918148032648540035013148238037518340 96t^{13}+$$
$$2462928599437882775837555711397331733365075724799478438453117632152276 80t^{14}+$$
$$3643922554326584164291206404117433264393801322848917361581602771200084 12t^{15}+$$
$$4922139169927943044120308535381977891292629983792003178608609613400121 96t^{16}+$$
$$6090033736249990307884933312796518167625324888510168994065511814890488 85t^{17}+$$
$$6921026595784580041805929111650935157535045058784539821685536320335816 34t^{18}+$$
$$7241540042286712039879265833745497253964719463201337439865012499317653 99t^{19}+$$
$$6989882350214037701171631826299504321587655871890219030395647180892472 80t^{20}+$$
$$6234751800949585332287505852065657708875390815066457072616439364794544 35t^{21}+$$
$$5146219912961319311991928034834852740873634703323973670883480182777878 90t^{22}+$$
$$3935283750683919363521926883593127041536443911623571708466990644678703 85t^{23}+$$
$$2790481507888856322302221776832448083615285763146234884395735634926442 48t^{24}+$$
$$1836113516072380547892403053193122157629926833259092818037384402953522 54t^{25}+$$
$$1121628937587569744326672160250823901475501717368248892936719408607216 52t^{26}+$$
$$636288094171279249100724446031653016914663031435947532609176250365670 30t^{27}+$$
$$335239007896749033620450404303030560689532903155914895506474847323011 68t^{28}+$$
$$164024603435105277975286872599037256784637265186243211558561778233889 02t^{29}+$$
$$74505840417678683988755396978589505049532641082110093495104025422674 12t^{30}+$$
$$31404022587731259055166688561530888384786569294389800931888290185994 06t^{31}+$$
$$12274131640104215278412353369768502156590334721281030530622617391287 40t^{32}+$$
$$4444355069461556889083740077886368602752418534687323110500779485392 49t^{33}+$$
$$1489152978652125955947502280170498285168650183537366074835854961895 30t^{34}+$$
$$461075209515623296397863923992756140368225491906377044040267203889 63t^{35}+$$
$$131696141953742189871670051193327804828129624454080296254656873207 04t^{36}+$$
$$34631638378211363225620123734599713828839017565261638244812968032 55t^{37}+$$
$$8364614642302572897264951194807752402969367899692639494195553958 66t^{38}+$$
$$1850489396181956867516828446706229501121084770890816106571915449 81t^{39}+$$
$$373749142672204620352731165679300308569376714752215377581630332 80t^{40}+$$
$$68653566309103513975376423265723706868027860466619846097580067 00t^{41}+$$

$n = 21$ *(continuação)*

$$
\begin{aligned}
&1141751501622190618145393746205865335347808922708773194322622000t^{42}+\\
&1709912319833256522161937671011948619267597193499140682642800000t^{43}+\\
&2291277338397411619145460973308387330069958644770558388720000 0t^{40}+\\
&2725854178567256529265770376311043610971084203370079610000000t^{45}+\\
&2851659987519913763695995790206355709357048850624612000000000t^{46}+\\
&2592224379556243804175989490472326313438614015772210000000 0t^{47}+\\
&2016432486995179111727304507803963855647958705112000000000t^{48}+\\
&1315418705684487750493589072994068457279368547900000000000t^{49}+\\
&6999375967251292253682675226552132019067773400000000000000t^{50}+\\
&2917520241573213651546775176866055183197700000000000000t^{51}+\\
&893439678118532534171335083118490251488000000000000000t^{52}+\\
&1787396004502553962190377885542397344000000000000000000t^{53}+\\
&175265194342094957439656887631846400000000000000000000t^{54})]
\end{aligned}
$$

$$
\begin{aligned}
\div\ &[(2+t)^2(3+t)^2(3+2t)^2(5+2t)^2(4+3t)^2(5+3t)^2(7+3t)^2(8+3t)^2\\
&\times(7+4t)^2(6+5t)^2(7+5t)^2(8+5t)^2(11+5t)^2(12+5t)^2(13+5t)^2(14+5t)^2\\
&\times(16+5t)^2(11+6t)^2(13+6t)^2(11+7t)^2(12+7t)^2(13+7t)^2(15+7t)^2(16+7t)^2\\
&\times(11+8t)^2(13+8t)^2(15+8t)^2(11+9t)^2(13+9t)^2(14+9t)^2(16+9t)^2(11+10t)^2\\
&\times(13+10t)^2],\quad t>0
\end{aligned}
$$

- $n = 23$

$$
\begin{aligned}
f(t) = [14586t^5(&98512972700269354271050514515207407117576645518901881969049600000+\\
&3177131376042332909470761747800213961647636760811131546422476800000t+\\
&50238756275635076183149091721227471995935148491768223618616524800000t^2+\\
&519144564875338146806522859281117530128673263768532556321875034112000t^3+\\
&3942530032653041072406164056434576750955880779439851156826386574868480t^4+\\
&23461746860781726905995522142908938613231842511443459905523804114518016t^5+\\
&113919735923790237007340982061499927250890368328949937607978401868414976t^6+\\
&464027857812316411717648402944809184339407127775362619138946517220835328t^7+\\
&1617936699130942170374748148210822035048743023273417861845084183231332352t^8+\\
&4903359954569473920311713582452774743365580722630773485554141656180931584t^9+\\
&13071642067546292133068228684769114252792279989678781767224820679690298880t^{10}+\\
&30947140103126091092746844874256890618592272712638133711809963728759286080t^{11}+
\end{aligned}
$$

$n = 23$ *(continuação)*

$$655758944810680258194712516464920683025512699291505342827752407337735945 60t^{12}+$$
$$1251677198827387633816863516115471580218535665683486012245138323019830844 00t^{13}+$$
$$2163692782932729276545428344510695445327725054838503790048348919738209531 20t^{14}+$$
$$3402644497286016697409196541723359385355496951312781329758930781617002666 20t^{15}+$$
$$4886724034741191270341965971150988197283343711691959062622365907034343202 20t^{16}+$$
$$6430030128113652995408174171818436451064201236348207375055790067374974610 05t^{17}+$$
$$7773313520274838261301544881162366512357280608494810797245013797889836929 30t^{18}+$$
$$8654101918365005385643381439999546161927213595362704486743097783536089963 75t^{19}+$$
$$8890566921161150968324619140306946956574171815038570771343845219937687656 40t^{20}+$$
$$8442293884795226868491432296835506294748701650293266554706891143171402420 75t^{21}+$$
$$7420353598366528347952836581876158805463441112962529637145978774425111030 10t^{22}+$$
$$6043946299471321488998750518427456174888974128776566783888262919027700912 65t^{23}+$$
$$4566099423770294259685503152390975833495831456096420520927174085374331488 40t^{24}+$$
$$3201849099209202588399466212865113214694576995610225003888226262877824944 78t^{25}+$$
$$2084966889968370125255840489174919711413230973689813359393747106189007241 48t^{26}+$$
$$1261147674002280294070485508081483795069082398621320093775350619744630944 54t^{27}+$$
$$708667475082297989435241174553665696685001285707988683159268002744821741 76t^{28}+$$
$$369901787348195507891158217463703997871522289032054001019525180419916179 42t^{29}+$$
$$179296144718613408298061457467324909285639601023631845819249111404088737 80t^{30}+$$
$$80664581013954358450767618417407837053568713381450516228340092183906362 70t^{31}+$$
$$33660391106356456286458371796651713830028352287825659561398349962827693 40t^{32}+$$
$$13016126552139966807591447506880387407932164507768294670777510212997028 25t^{33}+$$
$$4658765529073883051729726647606703050023953097189916649455660697834759 30t^{34}+$$
$$1541259488912264363791951646737753169988554457919785515498903567542771 55t^{35}+$$
$$470504147367480083720555923180045269707728124986017295049716242411898 80t^{36}+$$
$$132270758442317008581006314848057655056552168314854990110364822860284 95t^{37}+$$
$$34162619423245964084243812420226323170800528619105334412435173485673 70t^{38}+$$
$$8083887963821400144861097411182341128200800401191406407098898351035 25t^{39}+$$
$$1746853209696977359295791139043184422639485161475713685065717190289 60t^{40}+$$
$$343396537196867241268084953445964452652509518378043440442757356587 00t^{41}+$$
$$61132914111020102516215378172642542219985019888514270269123960914 40t^{42}+$$
$$9803071143025916982441411839132919124759862930760900889763406249 60t^{43}+$$

$n = 23$ *(continuação)*

$$
\begin{aligned}
&14069054021737844395465438287423811268102435153518286684169422208 0t^{44}+\\
&179309337893588076061902132752694838636952848541399616360261730 56t^{45}+\\
&201013493163969035222469985871993227061767602379965062824330137 6t^{46}+\\
&195858026500590399059555452462580675534851052464230330056817996 8t^{47}+\\
&163345779555169663398652823753816708064906251247909772679454 72t^{48}+\\
&114276735815930996480625605817484694527152578281385517808742 4t^{49}+\\
&652284180594264968576286145983821367978132269121812669030 40t^{50}+\\
&291735471204685991421691762759045659041641375634455449600 0t^{51}+\\
&958854140498842287079436854949108455014285576570470400 00t^{52}+\\
&205936894234257159967672331554015951927592736030720000 0t^{53}+\\
&216844904356445191686225158235441302194837782528000 00t^{54})]
\end{aligned}
$$

$$
\begin{aligned}
&\div [(2+t)^2(3+2t)^2(5+2t)^2(4+3t)^2(5+3t)^2(7+3t)^2(8+3t)^2(7+4t)^2\\
&\quad \times (6+5t)^2(7+5t)^2(8+5t)^2(13+6t)^2(17+6t)^2(12+7t)^2(13+7t)^2(15+7t)^2\\
&\qquad \times (16+7t)^2(17+7t)^2(13+8t)^2(15+8t)^2(17+8t)^2(13+9t)^2(14+9t)^2\\
&\qquad\quad \times (16+9t)^2(17+9t)^2(13+10t)^2(17+10t)^2(12+11t)^2(13+11t)^2(14+11t)^2\\
&\qquad\qquad \times (15+11t)^2(16+11t)^2(17+11t)^2], \quad t > 0
\end{aligned}
$$

♦ n par

- $n = 4$

$$
f(t) = \begin{cases} \frac{24t}{(t+3)^3} & , \ 0 < t < 1 \\ \frac{24}{(t+3)^3} & , \ t \geq 1 \end{cases}
$$

- $n = 6$

$$
f(t) = \begin{cases} \frac{12t}{(2+t)^3} & , \ 0 < t < 1 \\ \frac{12}{(2+t)^3} & , \ t \geq 1 \end{cases}
$$

- $n = 8$

$0 < t < 1$

$$
f(t) = \frac{40t^2(11+5t)(2745+4046t+1697t^2-108t^3-216t^4-36t^5)}{(2+t)^2(3+t)^3(5+2t)^2(5+3t)^3}
$$

$n = 8$ *(continuação)*

$t \geq 1$

$$f(t) = \frac{40(11+5t)(-900+1080t+3816t^2+3029t^3+986t^4+117t^5)}{(2+t)^2(3+t)^3(5+2t)^2(5+3t)^3}$$

- $n = 10$

$0 < t < 1$

$$f(t) = \frac{140t^2(13+7t)(5355+9707t+5405t^2+225t^3-648t^4-144t^5)}{(2+t)^2(3+2t)^3(7+3t)^3(7+4t)^2}$$

$t \geq 1$

$$f(t) = \frac{280(13+7t)(-882+756t+4185t^2+4075t^3+1591t^4+225t^5)}{(2+t)^2(3+2t)^3(7+3t)^3(7+4t)^2}$$

- $n = 12$

$0 < t < 1$

$$\begin{aligned} f(t) = [&420t^3(127205198209152 + 830799616655232t + 2478370256035679t^2 + \\ &4457651167858946t^3 + 5369630829283731t^4 + 4544996085788344t^5 + \\ &2755470100584478t^6 + 1198097443389324t^7 + 373477604229598t^8 + \\ &90099059782264t^9 + 24364976255715t^{10} + 9685841614466t^{11} + \\ &3616466569103t^{12} + 937130921280t^{13} + 154004378400t^{14} + \\ &14574816000t^{15} + 609120000t^{16})] \\ \div [&(2+t)^3(3+t)^3(7+3t)^2(8+3t)^2(7+4t)^2(7+5t)^3(8+5t)^2(9+5t)^2] \end{aligned}$$

$t \geq 1$

$$\begin{aligned} f(t) = [&420(12324108496896 + 24841789378560t - 28171024172928t^2 + \\ &24209228066688t^3 + 802028309134943t^4 + 2620893381510146t^5 + \\ &4598719171545747t^6 + 5326557118231480t^7 + 4401898325827870t^8 + \\ &2690121653573772t^9 + 1234136722389214t^{10} + 425384338104952t^{11} + \\ &108817580475363t^{12} + 20078354384258t^{13} + 2529902673551t^{14} + \\ &195057576000t^{15} + 6948295200t^{16})] \\ \div [&(2+t)^3(3+t)^3(7+3t)^2(8+3t)^2(7+4t)^2(7+5t)(8+5t)^2(9+5t)^2] \end{aligned}$$

- $n = 14$

$0 < t < 1$

$$\begin{aligned} f(t) = [&42t^3(10418362819200 + 65694410602380t + 186335185912447t^2+\\ &312474699352077t^3 + 341869454913160t^4 + 253307135536282t^5+\\ &127203719461519t^6 + 42007348140745t^7 + 8930493368394t^8+\\ &1848472402516t^9 + 894548862480t^{10} + 416232993600t^{11}+\\ &112169498400t^{12} + 16024608000t^{13} + 959040000t^{14})]\\ &\div [(2+t)^2(3+2t)^2(5+2t)^3(4+3t)^3(5+3t)^2(9+4t)^2(8+5t)^2(9+5t)^3] \end{aligned}$$

$t \geq 1$

$$\begin{aligned} f(t) = [&42(1016167680000 + 1825229376000t - 2901482143200t^2+\\ &2338534756800t^3 + 65703271291380t^4 + 199152358427047t^5+\\ &320795456986077t^6 + 334623491794360t^7 + 242560321799482t^8+\\ &125527361670319t^9 + 46483284491545t^{10} + 12078314774394t^{11}+\\ &2097456493516t^{12} + 219032836680t^{13} + 10417276800t^{14})]\\ &\div [(2+t)^2(3+2t)^2(5+2t)^3(4+3t)^3(5+3t)^2(9+4t)^2(8+5t)^2(9+5t)^2] \end{aligned}$$

- $n = 16$

$0 < t < 1$

$$\begin{aligned} f(t) = [&1848t^4(21+11t)(585005078036866148679600000+\\ &709682696061550919384054400t + 4099792492248496875898682220t^2+\\ &15002148707023523218584917076t^3 + 39016929116819426982470283375t^4+\\ &76708045024456943610778457412t^5 + 118380333484355695027231689954t^6+\\ &146999987005229717709936775644t^7 + 149386328387482391975559687681t^8+\\ &125723640518939328007961923984t^9 + 88373634784245638551583926288t^{10}+\\ &52211327780797052895408495152t^{11} + 26058227196203480938110423537t^{12}+\\ &11036158535882173521017238436t^{13} + 3980636264994775010721371882t^{14}+\\ &1221852257072663549485531660t^{15} + 313615032543093368271311183t^{16}+\\ &62401024445606807437097128t^{17} + 6547133390352681943189656t^{18}-\\ &1535331206662126881740332t^{19} - 1148905315541268362205776t^{20}-\\ &390134745258398679299360t^{21} - 91202094717382957056000t^{22}-\\ &15556863579453744811200t^{23} - 1919892892118148864000t^{24}- \end{aligned}$$

$n = 16$ *(continuação)*

$$
\begin{aligned}
&1633785072289574400000t^{25} - 86094504502272000000t^{26} - \\
&212242560768000000t^{27})] \\
&\div [(2+t)^2(3+t)^3(3+2t)^2(5+2t)^2(5+3t)^3(9+4t)^2(11+4t)^2(9+5t)^2 \\
&\quad \times (11+5t)^3(12+5t)^2(11+6t)^2(9+7t)^3(10+7t)^2(11+7t)^2(12+7t)^2]
\end{aligned}
$$

$t \geq 1$

$$
\begin{aligned}
f(t) = {} & [1848(21+11t)(-159596702189979436800000\,0- \\
& 9107882525812020825600000t - 14786168011598841802560000t^2 + \\
& 160577160072758204497920000t^3 + 1437613944472909063215828000t^4 + \\
& 7895326985901440375980140000t^5 + 40020766397798489863091548200t^6 + \\
& 147283318919263514749338056760t^7 + 388756620612695024782161390000t^8 + \\
& 769467300516357080250216873870t^9 + 1187837351410695898948239065040t^{10} + \\
& 1471239358491300215807699411940t^{11} + 1491270413247114378333542461560t^{12} + \\
& 1254020209346566417959472231090t^{13} + 882906227450268001555098015880t^{14} + \\
& 523453019178119591081830171520t^{15} + 262089707540476076735097609120t^{16} + \\
& 110861857214181961233494682610t^{17} + 39531743827548182822310576320t^{18} + \\
& 11826637531165661065653138100t^{19} + 2945146558372147864685325080t^{20} + \\
& 603345841487951947100062030t^{21} + 99954436014369379952890560t^{22} + \\
& 13059503762272820058381680t^{23} + 1295434413585258623146240t^{24} + \\
& 91690143647567342130400t^{25} + 4125489026465319744000t^{26} + \\
& 88679975215204800000t^{27})] \\
& \div [(2+t)^2(3+t)^3(3+2t)^2(5+2t)^2(5+3t)^3(9+4t)^2(11+4t)^2(9+5t)^2 \\
& \quad \times (11+5t)^3(12+5t)^2(11+6t)^2(9+7t)^3(10+7t)^2(11+7t)^2(12+7t)^2]
\end{aligned}
$$

- $n = 18$

$0 < t < 1$

$$\begin{aligned}
f(t) = [&-6006t^4(23+13t)(-380163348828799538714712000 0)- \\
&50166981676498565464570670400t - 3155501683434715772613224511 60t^2- \\
&12584471073244075921321725179 24t^3 - 357060612013529620480321274411 0t^4- \\
&76661052619650419419528991144 83t^5 - 129327773432457720829763536851 71t^6- \\
&17571948553968175298753674832221t^7 - 19555323417569114999383145958659t^8- \\
&18033658187842941247384582965346t^9 - 13892675579053310060866687981982t^{10}- \\
&8991182938663982608805155746478t^{11} - 4908523454748598685978986511108t^{12}- \\
&2267960833307497946766411046279t^{13} - 889563918958568317946385882383t^{14}- \\
&296643421800173026967423487985t^{15} - 83561362017712079649273953867t^{16}- \\
&19157867685806228142510841972t^{17} - 3040205332133907305214529304t^{18}- \\
&62926958813830164239765 92t^{19} + 2214123895529966991997457 44t^{20}+ \\
&980250782891149593587804 80t^{21} + 269680758147409828477440 00t^{22}+ \\
&5269674036574926480499200t^{23} + 736802820716914925568000t^{24}+ \\
&70618665978135797760000t^{25} + 4175856281950617600000t^{26}+ \\
&115225754370048000000t^{27})] \\
\div\, [&(2+t)^3(3+2t)^2(5+3t)^2((5+4t)^3(11+5t)^2(12+5t)^2(13+5t)^3(11+6t)^2 \\
&\times (13+6t)^2(10+7t)^2(11+7t)^3(12+7t)^2(13+7t)^2(11+8t)^2(13+8t)^2)]
\end{aligned}$$

$t \geq 1$

$$
\begin{aligned}
f(t) = [&6006(23+13t)(-1045261654135680833280000000)- \\
&708578000463944165068800000t - 16090631644525093515974400000t^2+ \\
&166064110115799275677440000t^3 + 10552632725207648126631808800t^4+ \\
&614574804579293970082001760000t^5 + 3146850755446360977593721497600t^6+ \\
&12332589676918288067498067905240t^7 + 35418181427794077626746754331100t^8+ \\
&76723702322567589114934945590830t^9 + 129771425198491193504992719509710t^{10}+ \\
&17608695395048078595262122932021t^{11} + 19545724521402905712590035149259t^{12}+ \\
&17992859296058889174413833852346t^{13} + 13864680150922644142910156213782t^{14}+ \\
&8995481827765596899058114283478t^{15} + 4928853442903850223219726274108t^{16}+ \\
&2281723697035439607913585338479t^{17} + 890590925246300111392886921383t^{18}+ \\
&291699685977014698580711587385t^{19} + 79548972809528185225296384067t^{20}+ \\
&17851439492959453109878546172t^{21} + 3240611770220222715140999704t^{22}+ \\
&464107245632469545488324592t^{23} + 50481297853354820198484656t^{24}+ \\
&3919440793192189842265920t^{25} + 193522381682495673830400t^{26}+ \\
&4566752333922562560000t^{27}] \\
\div [&(2+t)^3(3+2t)^2(5+3t)^2(5+4t)^3(11+5t)^2(12+5t)^2(13+5t)^3(11+6t)^2 \\
&\times (13+6t)^2(10+7t)^2(11+7t)^3(12+7t)^2(13+7t)^2(11+8t)^2(13+8t)^2)]
\end{aligned}
$$

- $n = 20$

$0 < t < 1$

$$
\begin{aligned}
f(t) = [(1716t^5(&515003012975963916186712697526079415157423360000+\\
&114609887383301092361212608502828867123585828800000t+\\
&1241972345937413644003657032916246620256233080640000t^2+\\
&8732370022362224012081990849046857239004425904220000t^3+\\
&44785292257729816981967489351795761426725746080795680t^4+\\
&178581280976602702235103859591213105106476436321871600t^5+\\
&576278286791720165818344593673110768430680370924310560t^6+\\
&1546721286129202627388892142510375276950773317794452070t^7+\\
&3521809617526548037412533851921442085801500073844771030t^8+\\
&6904786070289783299486180359543892853120934395144665400t^9+\\
&11791188820550612870305193943210659460731095227886795640t^{10}+\\
&17698211154677016058422134587482570473428991599931027020t^{11}+\\
&23519284291905150675413777362695881473735117184895206420t^{12}+\\
&27835493549703962221714656423114138946047325389525814040t^{13}+\\
&29480599958727657412937279810635121674795588467369388720t^{14}+\\
&28050114538969834794477702125464713789112805403388963610t^{15}+\\
&24053245860106652124276398156042235250078955089219647170t^{16}+\\
&18636399045867487729356766346341432278171376718676122560t^{17}+\\
&13072837457813440077680358170314210201956243325032035480t^{18}+\\
&8314769435391723462662150835946700928020408533801116200t^{19}+\\
&4800076439037731763957341790759491285999183944670918200t^{20}+\\
&2516511871694322343252489553640577545976247296558853600t^{21}+\\
&1198158868767899185449613846408499070856074268488531200t^{22}+\\
&517795041531794846263766300179194410555475454951056250t^{23}+\\
&202868580986112421745059897073605169293011735579142970t^{24}+\\
&71924396318336193367885712124636341309706245605220600t^{25}+\\
&23019193251989733315089011954369382576746729859227240t^{26}+\\
&6634715540388247859597180212965578056615061877359180t^{27}+
\end{aligned}
$$

$n = 20$ *(continuação)*

$$
\begin{aligned}
&172145745534806932317995884986005973642081245258322t^{28}+\\
&404841894673195619969691765348075882848321025202220t^{29}+\\
&88587469334932205222953791924973148870826577101136t^{30}+\\
&19211150380372787207703533802518716716939994223343t^{31}+\\
&4518704733762197285476459724053918084956014734833t^{32}+\\
&1194186058819486664738156805727855801025298019176t^{33}+\\
&330545236652397331703319333424874545694371267708t^{34}+\\
&86946486633499217876338732950343500556785102224t^{35}+\\
&20443782239803021675780355420684416643055960048t^{36}+\\
&4172678168597689030245795730474283176509258224t^{37}+\\
&728126564325138743438652660929773171329102272t^{38}+\\
&107372615114729126096483749207026870372480000t^{39}+\\
&13206465350651239989957325862310267152640000t^{40}+\\
&1331133752802259051532465423468327424000000t^{41}+\\
&107195139074613006094084227511120281600000t^{42}+\\
&6637314937037524215719718459801600000000t^{43}+\\
&296881277485609762264021686681600000000t^{44}+\\
&8543023665349022961304338432000000000t^{45}+\\
&118811194611790549915533312000000000t^{46})]\\
&\div [(2+t)^2(3+t)^3(3+2t)^3(5+2t)^2(4+3t)^2(5+3t)^2(7+3t)^3(7+4t)^2(11+5t)^2\\
&\quad \times (12+5t)^2(13+5t)^2(14+5t)^2(11+6t)^2(13+6t)^2(11+7t)^2(12+7t)^2(13+7t)^3\\
&\qquad \times (15+7t)^2(11+8t)^2(13+8t)^2(15+8t)^2(11+9t)^3(13+9t)^2(14+9t)^2)]
\end{aligned}
$$

$t \geq 1$

$$
\begin{aligned}
f(t) = [3432(&192373312456913188562363117409348695040000000000+\\
&2671507102083064904438651674502749969920000000000t+\\
&16428528321594357493451719376170003505088000000000t^2+\\
&56330069410628948941989985377141963932736000000000t^3+\\
&103496737049167070023674909887526897668572512000000t^4+\\
&53075108921608491383697980879177617621371120000000t^5+\\
&203479716465153633884842549121070540166987525200 0t^6+\\
&5242939590498497567686907843047283424275977147 2000t^7+\\
&42752320527770119505675308791362792205067543617 2784t^8+
\end{aligned}
$$

$n = 20$ *(continuação)*

$$2247471410906078626009663729717011883349187027760080t^{9}+$$
$$8964894404387167305773123752299522935985082311491778t^{10}+$$
$$28856186725121170452051574770516223603767637533134916t^{11}+$$
$$77335344377277747103433741352195950697676622429091114t^{12}+$$
$$176018984693931930377841976339545373949169317570708895t^{13}+$$
$$345140718448274011243733291198852973072363268244171532t^{14}+$$
$$589526131384695361317224559227748253391043369437542726t^{15}+$$
$$884991123591649132626963813703787643840062932191340196t^{16}+$$
$$1176101029114905397818456894200616900005390675954757952t^{17}+$$
$$1391852152129741193103963231177173541730037154232963186t^{18}+$$
$$1473986959884623704687885525695288105952050348077129618t^{19}+$$
$$1402389315044604846079913354077769383178003922803070046t^{20}+$$
$$1202572013517851890715077711321969959760544663496050503t^{21}+$$
$$931813138984149446363437744867151167869057080018037524t^{22}+$$
$$653697938840883185761340926428606544489040602805921560t^{23}+$$
$$415798912442641932216466862177367089140697431195977160t^{24}+$$
$$240029377562661069755202989742194065030475228073222680t^{25}+$$
$$125816583446950884181008046019780582286988388848170310t^{26}+$$
$$59886370372110968331628310070768931241285690686309000t^{27}+$$
$$25874146472802807651631838477756163342226030998450086t^{28}+$$
$$10138782565583500369357179729371515242372869585761405t^{29}+$$
$$3598431961081450870355446265202707806854613452112612t^{30}+$$
$$1154648772364117619449555630919745464025804668305334t^{31}+$$
$$334155018112668613516717746390828151967771181400036t^{32}+$$
$$86951418426656558781066616983398215041362697336360t^{33}+$$
$$20266313420198255904342888346404963974647385613318t^{34}+$$
$$4211009642914010818982371363495312840062069423234t^{35}+$$
$$775480950439079713317695057134416682636116601554t^{36}+$$
$$125655137997561446626815608822729829059337789213t^{37}+$$
$$17753138360712573981063033507000851139542504604t^{38}+$$
$$2162027220096560023524153875595735083811489612t^{39}+$$
$$223601972320325123407673395429151295809625024t^{40}+$$

$n = 20$ *(continuação)*

$$
\begin{aligned}
&19253905959028911023258931496545488592132912t^{41}+\\
&1343048470391556004853980678653393492563136t^{42}+\\
&729092348513511483735225373746883392720000t^{43}+\\
&28897688594525865562109974184068331520000t^{44}+\\
&743816566213225331761221566873932800000t^{45}+\\
&9330882817883115582921213233971200000t^{46})]\\
&\div [(2+t)^2(3+t)^3(3+2t)^3(5+2t)^2(4+3t)^2(5+3t)^2(7+3t)^3(7+4t)^2(11+5t)^2\\
&\quad \times (12+5t)^2(13+5t)^2(14+5t)^2(11+6t)^2(13+6t)^2(11+7t)^2(12+7t)^2\\
&\qquad \times (13+7t)^3(15+7t)^2(11+8t)^2(13+8t)^2(15+8t)^2(11+9t)^3(13+9t)^2\\
&\qquad\quad \times (14+9t)^2)]
\end{aligned}
$$

- $n = 22$

$0 < t < 1$

$$\begin{aligned}f(t) = [858t^5(&74848747472876662163195173289664234651648000 0+\\&159808573274035102464118681838823047178485760 00t+\\&16570701768599619252024559908258027751782481920 0t^2+\\&1111708643181064392622504584800754989261371146240t^3+\\&5424109299042791415593097519054124882398124941312t^4+\\&20510986208600035821951955394944312511967230103552t^5+\\&62557033683519841746031026940507392681301632384000t^6+\\&158121056282588151816311422048751375942567858478336t^7+\\&337764774807189675328760226209671280467980811529632t^8+\\&618720280414204044532507367698506029946660315223408t^9+\\&982879436515972581906270654686902628257783669960976t^{10}+\\&1365969673953228038654879667112703015094879608172376t^{11}+\\&1672366862272396421753297432614710917307843485145890t^{12}+\\&1813730128361054529686749980078210955071326609952763t^{13}+\\&1750156250492234175508616521803106004460260071448140t^{14}+\\&1507844390384601814700205391644443022550992121084965t^{15}+\\&1163019027411300605397264915642504326141747005684170t^{16}+\\&804725104890385471303596953970976097642675313541135t^{17}+\\&500204915169336431393931396011213162953620862412440t^{18}+\\&279533338027581622415317875948272126122522502572085t^{19}+\\&140466742141219291370498476629399059453949531281410t^{20}+\\&63434250614690824469202935595164404837873400068325t^{21}+\\&25707874538882056563433502596384667157713423430220t^{22}+\\&9326696480588233923060418534192154444225691357415t^{23}+\\&3018031258919141694387211531146757736089926164118t^{24}+\end{aligned}$$

$n = 22$ *(continuação)*

$$
\begin{aligned}
&867090143749678874717893411185009629901955174173t^{25}+\\
&220340572083467060316572113805699237172072008160t^{26}+\\
&49691679777430642034117857483196727989752161879t^{27}+\\
&10247839898822591588061875433781792432772120028t^{28}+\\
&2123494470675961881458910404671156852457763012t^{29}+\\
&512014195827722202932127457970945967244963104t^{30}+\\
&147287089166637828366765049098509361601004304t^{31}+\\
&43677945018887633519659811332700303987259840t^{32}+\\
&11718357648407142970627103934725567347453632t^{33}+\\
&2683575816135894488456906436030559882613760t^{34}+\\
&512663103213140792213306286295796906342400t^{35}+\\
&80556511504185576027446195888406555033600t^{36}+\\
&10245795056274735647169432887454234624000t^{37}+\\
&1030303407376868876893315446513592320000t^{38}+\\
&78961307346853030007320157680435200000t^{39}+\\
&4338794731264879458693746884608000000t^{40}+\\
&152384277188053611937624227840000000t^{41}+\\
&2571924077080670898251366400000000t^{42})]\\
\div\ &[(2+t)^2(3+2t)^2(5+2t)^2(4+3t)^2(5+3t)^2(7+3t)^2(8+3t)^3(7+4t)^3(6+5t)^3\\
&\times (7+5t)^2(8+5t)^2(13+6t)^2(12+7t)^2(13+7t)^2(15+7t)^3(16+7t)^2(13+8t)^2\\
&\quad\times (15+8t)^2(13+9t)^3(14+9t)^2(16+9t)^2(13+10t)^2)]
\end{aligned}
$$

$t \geq 1$

$$
\begin{aligned}
f(t) = [858(&563944187634423522568572776994570240000000+\\
&7326626993186960712795568591450669056000000t+\\
&41332667433762733493184371021428083916800000t^2+\\
&124622480348962601070393378199599505735680000t^3+
\end{aligned}
$$

$n = 22$ *(continuação)*

$$
\begin{aligned}
&171871239812375692232745948019902866522112000 0t^4-\\
&6008317091836510509605661106202788484874240 00t^5+\\
&547270353513481030776404462826449761629634560 0t^6+\\
&1472981240746441274078771961512686153700474880 00t^7+\\
&110652616762503855124439358550797081760855425024 0t^8+\\
&546205942874321863184711221118506329232851239731 2t^9+\\
&2058564165839387904829454771012052148533694207795 2t^{10}+\\
&6259684122240885355524286174540457031885340085760 0t^{11}+\\
&1580462153789716961578259485430411103202147746255 36t^{12}+\\
&3375987447599938832168515972637894370514948693856 32t^{13}+\\
&6186116446766857149500531613195343235194014458954 08t^{14}+\\
&9829605590445674073465083499291127553377430870425 76t^{15}+\\
&1366194825939407080184948242259155283223272790362776t^{16}+\\
&1672537167345857140239113407162216503107819126798690t^{17}+\\
&1813696221974735178995008773740617546858416779016763t^{18}+\\
&1749966228564210556489320139997647962917849089520140t^{19}+\\
&1507675678137667594416747552122802514700997679747365t^{20}+\\
&1162993871871901226221712706880908604490182053728970t^{21}+\\
&804818524880113474561557645044271022397458175633935t^{22}+\\
&500310353908190962077953803405594573346243872364440t^{23}+\\
&279576342936414822362766957882431512066761079324085t^{24}+\\
&140448452379112002892621602789027020926169127079810t^{25}+\\
&63396301090760482972325915621986979604275627841125t^{26}+\\
&25683067233267076464622399579861761490000466915020t^{27}+\\
&9321481135960109662605811578196203815706923517415t^{28}+\\
&3023306370691160087201158627022272649807528756118t^{29}+\\
&873344343540296156519899645355515497228136476573t^{30}+\\
&223729935384820878438797495447666268323539131360t^{31}+\\
&50550248111058711529786185337651783648221609879t^{32}+\\
&10004195888660096246645472822471050158758464028t^{33}+\\
&1719078899826516039069258523865373518447499012t^{34}+\\
&253623848107736643743845169844023355820987104t^{35}+
\end{aligned}
$$

$n = 22$ *(continuação)*

$$3166004986912083917831770886015312833167950 4t^{36}+$$

$$327913980811819444481836456590892738030944 0t^{37}+$$

$$274237297353388202058456746340208554493632t^{38}+$$

$$177943654473138563529803466916031914137 60t^{39}+$$

$$84057300253373925561011568750877240320 0t^{40}+$$

$$2571623948988489840504146537309798400 0t^{41}+$$

$$382450050627671095931062645678080000t^{42})]$$

$$\div [(2+t)^2(3+2t)^2(5+2t)^2(4+3t)^2(5+3t)^2(7+3t)^2(8+3t)^3(7+4t)^3(6+5t)^3$$
$$\times (7+5t)^2(8+5t)^2(13+6t)^2(12+7t)^2(13+7t)^2(15+7t)^3(16+7t)^2(13+8t)^2$$
$$\times (15+8t)^2(13+9t)^3(14+9t)^2(16+9t)^2(13+10t)^2)]$$

- $n = 24$

$0 < t < 1$

$$f(t) = [-(9724t^6(31+17t)$$
$$(-2402952132443362317560596724190137894204059765260894617347719954432 00000-$$
$$787245551811367720913803855050603581141303047374090644025071857041408 0000t-$$
$$126631230588348607348012616952117611951021976166813622965914568824963072000t^2-$$
$$133304113671197778828992300697125006058410885191498685477510821655163617280 0t^3-$$
$$103284312106914245864704911716015626984108973903610009132775735556565742694 40t^4-$$
$$628059535184614983087301958126324516856425947468315479058646980630825920130 56t^5-$$
$$312122925879416289799741197898622729946200955102214005656657486428612790366208t^6-$$
$$130344378323660453690828941115942335559410838963601808087998455544099694277 2224t^7-$$
$$466765935887089601490385433328762286394493088747842930691619422865414195260 3776t^8-$$
$$145553828152383551981011029939126645396125975039352018611488937571781208803 47168t^9-$$
$$400030281879871071939792836175741800834802217579297737225748749744192841127 73024t^{10}-$$
$$978357444211314525346697168576446419838815999219948743209371848005968701774 17680t^{11}-$$
$$214615528399515770662774101824744787158374337703582112258491382853015068581 250656t^{12}-$$
$$425032832432036892649142365249649461146075797046212434190345964059969553425 518603t^{13}-$$
$$764124327497850504501814204711448936308277471253618639232886306589199745764 445718t^{14}-$$
$$125286554129555782933935071226235653027641260983340580053164613416832239160 0980513t^{15}-$$
$$188090637776861261464974803888860062053640052007581249260446758130012200953 2682796t^{16}-$$

$n = 24$ *(continuação)*

$2594370066706650547290243763066784992044907058060630148297439096103977936760616327t^{17}-$
$3297421412414124367648818988084873513002664992751545288314144757401322496447908114t^{18}-$
$3871661043305385191805364781866084096485821008687044432851209168869161873524683493t^{19}-$
$4208759875219209171177748507797855495029225042401986582436202215601877841159904856t^{20}-$
$4243935229772319990551819451846583012232352788700559289164303357972710361629022167t^{21}-$
$3976037636054890176587335878100969427393432779280607344406190260793515036768594214t^{22}-$
$3465823611192776598259217380060514843261534850759280098403662644786419819653179613t^{23}-$
$2814175501017052572379816494038025865460390333356752553118393513492742643075516116t^{24}-$
$2130660575380348917614539452225834828133423876392194402884447024189561782879282715t^{25}-$
$1505380918863125817518298765533683981892153838897633903965094249057599221576200018t^{26}-$
$993170605504701005200419159586529916710972084389863035084165165899596464676325905t^{27}-$
$612141509711569149122751646447343571249630167903636794584188262332448218042589424t^{28}-$
$352587448669028484857465693333403982065545005517057279429454723930820926583611561t^{29}-$
$189818429781226108668631360523553935748926974079072620986325188193845677317055346t^{30}-$
$95514643257305170244028097083230340551120883367360875972632242249370441814149123t^{31}-$
$44915904032038513086910949451465503924445904381631477152067889138784530359390484t^{32}-$
$19733472865993944287534827590027191010286557933733660008261470728257802798728741t^{33}-$
$8096401145035889512463606536735284382532521775852623421402762930607591938714438t^{34}-$
$3100389779559468079380461022979936551831658222386240462068277748224552688706367t^{35}-$
$1107301601479423490575731182874768510683970864030822902522029748456117268209304t^{36}-$
$368522662415446603649783084669130574475346491306489316802058472301806539414573t^{37}-$
$114168698397003183389407953881119931247935666168425929424985373786056757271986t^{38}-$
$32877682083218361821115316896583227587882498748374465979691796411031762266607t^{39}-$
$8782231619087900308841496683652408229603119737789178847433533011437706264044t^{40}-$
$2167921061678426524396668485274284447739203732632629688587885204687692297433t^{41}-$
$490903751531231570521439815655997032793982634038230895992609452040683605686t^{42}-$
$100316838652021172381059233200545309968143738212224168883098011598734320187t^{43}-$
$17771385198040866062183515430053774079835831950923161625704820143880916544t^{44}-$
$2408836493597243868072061309589355346886839852022371862334090059792231304t^{45}-$
$98730734757499551409694503343932926143163154530386767916907250616990624t^{46}+$
$86445954296929850144715311644559459268620944726302855438503706786080304t^{47}+$
$42767909855226001727303519362267962235443149828758788916751741668665600t^{48}+$

$n = 24$ *(continuação)*

$$
\begin{aligned}
&1380836892277063184184206250364774829408149604010328323101715519458150 4t^{49}+\\
&36502756259511983914876707698653614798181927913427162746630059987123 20t^{50}+\\
&8357509175460611493005823329957425075653096102358526348533245518753 28t^{51}+\\
&1688726045514760718839709564632025161086621316635501083533731769958 40t^{52}+\\
&302892557194870478762904479388394407170323717127057728414873952317 44t^{53}+\\
&482226060152896757786179385313390225120803878363582665249495198105 6t^{54}+\\
&67919632015376213292921523469533777019191937019286878060435621888 0t^{55}+\\
&8414081357202324938519738073634071231828245844695828177414553600 0t^{56}+\\
&909380432980755703974987019410894094951777071448724605797335040 0t^{57}+\\
&84817921586973562065095462731286282632552669028142552278630400 0t^{58}+\\
&6729200061888131971136564861806779682551522399538807111680000 0t^{59}+\\
&445365934716652804139144340665420477112850232007390658560000 0t^{60}+\\
&23930856982746427246648704763458742561456451485591142400000 0t^{61}+\\
&1003134441180093323835764671463051079176722946457600000000 0t^{62}+\\
&30779051030513607511821886678854920421541297520640000000 0t^{63}+\\
&614932177688003684430633189128544806150917324800000000 0t^{64}+\\
&6004028712067283151033064191922268616523776000000000 0t^{65})]\\
&\div [(2+t)^3(3+t)^3(3+2t)^2(5+2t)^2(5+3t)^3(7+3t)^2(8+3t)^2(7+4t)^2(9+4t)^2\\
&\quad \times (7+5t)^3(8+5t)^2(9+5t)^2(13+6t)^2(17+6t)^2(13+7t)^2(15+7t)^2(16+7t)^2\\
&\qquad \times (17+7t)^3(18+7t)^2(13+8t)^2(15+8t)^2(17+8t)^2(13+9t)^2(14+9t)^2(16+9t)^2\\
&\qquad\quad \times (17+9t)^2(13+10t)^2(17+10t)^2(13+11t)^3(14+11t)^2(15+11t)^2(16+11t)^2\\
&\qquad\qquad \times (17+11t)^2(18+11t)^2)]
\end{aligned}
$$

$t \geq 1$

$$
\begin{aligned}
f(t) = {} & [9724(31+17t)\\
&(-48273535893675577718272738569031277889441849220359870480384000000000 0-\\
&10828945893222913076991377509959196312822586585683058139070464000000000t-\\
&1140856835619616720342333651489971989418700230502089899158536192000000000t^2-\\
&742942809180864381957048475168316712594436707577130332340323614720000000t^3-\\
&3295417968812000776028230876224197258081591465905329014777950765056000000t^4-\\
&10225781330072730415129644754915625050449556673446061259415825888051200000t^5-\\
&20962014011831958377521236004396877556533745399133905978725249146552320000t^6-\\
&137917239139286396320236725414935326655661279033535114212453577106391040000t^7+
\end{aligned}
$$

$n = 24$ *(continuação)*

$154082897994899306669083813069711123074175332436351481437720522696313241600t^{8}+$
$1497415706292512397519161843072824997088820843793369400353326555319768064000t^{9}+$
$10662335006700313938717007061542405946946060019549137566724809942671432861440t^{10}+$
$63097805925129312002340644204318601355661512345110076855160012233923256083456t^{11}+$
$311836617744791005116814756330880490794114269656256793759340379529511622007808t^{12}+$
$1302091681705611359287534439972472345627574818569608407599634563432900444183424t^{13}+$
$4665638503699513168856891778693769210369336649643740431063587375994976628636576t^{14}+$
$14555447097633673354777937659288296115654111012619520287255440048566783460153116 8t^{15}+$
$40005379665095663039365443795217373368745260873991820680346599834384862211212624t^{16}+$
$97841435782734370734100399148192277537458608048307684193318445269916433815615080t^{17}+$
$214620900741976184445996287873530551034440679847973642924273999574161544717299831t^{18}+$
$425032374269599463540146404873865677913259718432302768801379493212115530676389178t^{19}+$
$764115772678700333808257651908862568821939683863579021992101890012538925420916093t^{20}+$
$1252853603630048291702187245884579899865350018521002003448911410304965204192612288t^{21}+$
$1880900337700615483325630008584075718161606605656474288608600345973710564508018771t^{22}+$
$2594376029278094952856324194030398609996916770065654325604165430337877308867022102t^{23}+$
$3297436178998232772728676487801542832069261603323413241510608956174880165257786689t^{24}+$
$3871673986693637271924909947489428615644209772161031671929481050216580514016929868t^{25}+$
$4208761644800860780697736186110864043466251770149122054386275224286184794264522131t^{26}+$
$4243925322114165624924902505561064796382616718941229328906253287074957782270851642t^{27}+$
$3976024092353166456297475299155855834605209833487714323606403173838676489680822889t^{28}+$
$3465815822701136687252949617870112124540820476905066020587685528513936905737562488t^{29}+$
$2814177104331826715243536307562312410475765934904272941754396863361613969925803991t^{30}+$
$2130668164577818857893028842985925612104274649429934436588417108395195493834199590t^{31}+$
$1505388244783705902627611403928851911672486875686366562779175712130743294326420893t^{32}+$
$993173590206542621684932328855151293132692566105950028505291977230611642205005380t^{33}+$
$612140131514967234355329954119496316327473455053733120447860323450658122179876749t^{34}+$
$352584228203343278280867324973865184027880871423496669140256492232091806758922686t^{35}+$
$189815856204270053555344472930411499551373594025894895369602459815664064926341487 1t^{36}+$
$95513716069531672334165592010847154925453583217295947075041223938533947236472848t^{37}+$
$44916258229534958730802872825248367126772146263371008872003089652658442824672009t^{38}+$
$19734263439765828414247099416592790795197878874400190329989661996892318310170066t^{39}+$
$8097016808744732621775809046545376226820050289638645102563516803879780793252563t^{40}+$

$n = 24$ *(continuação)*

$$
\begin{aligned}
&3100650265171641314022217886213798210324131207234663922118519995394650084675892t^{41}+\\
&1107304021490636047852994989802203262892104921988439932732253979469362860094529t^{42}+\\
&368426615673078922729490966193217091915794998400291478893272057668088890112398t^{43}+\\
&114078407989470489692046953054449111798652769876009132928689222615187966983411t^{44}+\\
&32826852592395954476182574921219870481328411521125719524257948199641948810232t^{45}+\\
&8764560726929140248657630960330087784027151869453928737052652744820060332269t^{46}+\\
&2167167061129936062334340723598820629648026215517870518373015610586645429458t^{47}+\\
&495189266754824584865896094223101888653616734545273578901948546408733198911t^{48}+\\
&104296981728817825579877295502097992828949424756911228092413278111598326612t^{49}+\\
&20189457085063424258998074106575460657422643358356408830931085622009053544t^{50}+\\
&3579772644267175902163234358448411563308495319708496350730548063127011904t^{51}+\\
&579088488736742122125667963757975816466442013650962926228514743281595024t^{52}+\\
&85070217630692459264263013645613411636689470047786732304935565117288896t^{53}+\\
&11286855945525035508677770264036745892876321046039209977967407312678400t^{54}+\\
&1343673301995925208265026339369397142906161380546539003850500012586496t^{55}+\\
&142402203071952991835547332656571794141737028778381186813998373863680t^{56}+\\
&13306111163360435920336299069556122514337914180472333998817491356672t^{57}+\\
&1083101234017394498789054012920365672793583014944778375022602536960t^{58}+\\
&75629570223597845939773246281532354488768094940012284479766265856t^{59}+\\
&4439342126758771250641104269406704486315591756319015101723602944t^{60}+\\
&213041580931588994039231216550118910172351724055957820851896320t^{61}+\\
&8026820587710528584271898206450850235490142198221204425932800t^{62}+\\
&222672052077260699419429689527441163217234078691928342528000t^{63}+\\
&4044005534171336640739221142522595106312104695030087680000t^{64}+\\
&36072124591479675999210601047653793383137666557542400000t^{65})]\\
&\div [(2+t)^3(3+t)^3(3+2t)^2(5+2t)^2(5+3t)^3(7+3t)^2(8+3t)^2(7+4t)^2(9+4t)^2\\
&\quad \times (7+5t)^3(8+5t)^2(9+5t)^2(13+6t)^2(17+6t)^2(13+7t)^2(15+7t)^2(16+7t)^2\\
&\quad\quad \times (17+7t)^3(18+7t)^2(13+8t)^2(15+8t)^2(17+8t)^2(13+9t)^2(14+9t)^2\\
&\quad\quad\quad \times (16+9t)^2(17+9t)^2(13+10t)^2(17+10t)^2(13+11t)^3(14+11t)^2\\
&\quad\quad\quad\quad \times (15+11t)^2(16+11t)^2(17+11t)^2(18+11t)^2)]
\end{aligned}
$$

Nota: *Os interessados podem solicitar as expressões para valores de n não contemplados.*

A.2 Spacings de Misturas de Pareto Tipo I

Seja

$$Z = \begin{cases} X_1 & X_2 \\ p & q \end{cases}$$

com $X_1 \frown F_{1,\alpha}(\delta_1)$, $X_2 \frown F_{1,\alpha}(\delta_2)$ *e* $p + q = 1$.

A v.a. Z *tem f.d.*

$$F_Z(z) = p\left[1 - \left(\frac{z}{\delta_1}\right)^{-\alpha}\right] I_{(\delta_1,\infty)} + q\left[1 - \left(\frac{z}{\delta_2}\right)^{-\alpha}\right] I_{(\delta_2,\infty)}$$

e f.d.p.

$$f_Z(z) = p\frac{\alpha}{\delta_1}\left(\frac{z}{\delta_1}\right)^{-\alpha-1} I_{(\delta_1,\infty)} + q\frac{\alpha}{\delta_2}\left(\frac{z}{\delta_2}\right)^{-\alpha-1} I_{(\delta_2,\infty)}.$$

Sem perda de generalidade, considere-se $\delta_1 < \delta_2$*; então,*

$$F_Z(z) = \begin{cases} 0 & , \ z \leq \delta_1 \\ p\left[1 - \left(\frac{z}{\delta_1}\right)^{-\alpha}\right] & , \ \delta_1 < z \leq \delta_2 \\ 1 - \theta z^{-\alpha} & , \ z > \delta_2 \end{cases} \tag{3.3.14}$$

e

$$f_Z(z) = \begin{cases} 0 & , \ z \leq \delta_1 \\ p\frac{\alpha}{\delta_1}\left(\frac{z}{\delta_1}\right)^{-\alpha-1} & , \ \delta_1 < z \leq \delta_2 \\ \alpha\theta z^{-\alpha-1} & , \ z > \delta_2 \end{cases} \tag{3.3.15}$$

onde $\theta = p\delta_1^\alpha + q\delta_2^\alpha$.

Seja $(Z_1, Z_2, \ldots, Z_n)$ *uma a.a. proveniente de uma população com distribuição* (3.3.14)*, e denotemos por*

$$U_{k,n} = Z_{k:n} - Z_{k-1:n}, \quad k = 1, \ldots, n$$

com a convenção $Z_{0:n} = \delta_1$.

O vector aleatório $(Z_{k-1:n}, Z_{k:n})$ *tem f.d.p. conjunta*

$$f_{k-1,k:n}(x,y) =$$
$$= \begin{cases} C_{k,n}\alpha^2\delta_1^{2\alpha}p^k\big[1-\big(\frac{x}{\delta_1}\big)^{-\alpha}\big]^{k-2}\big[q+p\big(\frac{y}{\delta_1}\big)^{-\alpha}\big]^{n-k}x^{-\alpha-1}y^{-\alpha-1} & , \delta_1 < x < y \leq \delta_2 \\ C_{k,n}\alpha^2\delta_1^{\alpha}p^{k-1}\theta^{n-k+1}\big[1-\big(\frac{x}{\delta_1}\big)^{-\alpha}\big]^{k-2}x^{-\alpha-1}y^{-\alpha(n-k+1)-1} & , \delta_1 < x \leq \delta_2 < y \\ C_{k,n}\alpha^2\theta^{n-k+2}(1-\theta x^{-\alpha})^{k-2}x^{-\alpha-1}y^{-\alpha(n-k+1)-1} & , \delta_2 < x < y \end{cases}$$

onde $C_{k,n} = \frac{n!}{(k-2)!(n-k)!}$.

Ora

$$f_{U_{k,n}}(u) = \int_{-\infty}^{\infty} f_{k-1,k:n}(x, x+u)dx, \quad u > 0.$$

- *Se* $0 < u \leq \delta_2 - \delta_1$*, então*

$$f_{U_{k,n}}(u) = \int_{\delta_1}^{\delta_2-u} f_{k-1,k:n}(x,x+u)dx + \int_{\delta_2-u}^{\delta_2} f_{k-1,k:n}(x,x+u)dx$$
$$+ \int_{\delta_2}^{\infty} f_{k-1,k:n}(x,x+u)dx. \quad (3.3.16)$$

Calculando separadamente cada integral de (3.3.16), *vem*

(a)

$$\int_{\delta_1}^{\delta_2-u} f_{k-1,k:n}(x,x+u)dx =$$

$$= C_{k,n}\alpha^2\delta_1^{2\alpha}p^k\int_{\delta_1}^{\delta_2-u}\left[1-\left(\frac{x}{\delta_1}\right)^{-\alpha}\right]^{k-2}\left[q+p\left(\frac{x+u}{\delta_1}\right)^{-\alpha}\right]^{n-k}x^{-\alpha-1}(x+u)^{-\alpha-1}dx$$

$$= C_{k,n}\alpha^2\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\sum_{j=0}^{n-k}\binom{n-k}{j}p^{n-j}q^j\delta_1^{\alpha(n-k-j+i+2)}$$

$$\times\int_{\delta_1}^{\delta_2-u}x^{-\alpha(i+1)-1}(x+u)^{-\alpha(n-k-j+1)-1}dx$$

$$= C_{k,n}\alpha^2\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\sum_{j=0}^{n-k}\binom{n-k}{j}p^{n-j}q^j\delta_1^{\alpha(n-k-j+i+2)}$$

$$\times\left\{\int_{\delta_1}^{\infty}x^{-\alpha(i+1)-1}(x+u)^{-\alpha(n-k-j+1)-1}dx-\int_{\delta_2-u}^{\infty}x^{-\alpha(i+1)-1}(x+u)^{-\alpha(n-k-j+1)-1}dx\right\}.$$

Procedendo à mudança de variável $w=\frac{u}{x}$ *e aplicando (2.1.24), segue-se que*

$$\int_{\delta_1}^{\delta_2-u} f_{k-1,k:n}(x,x+u)dx =$$

$$= C_{k,n}\alpha^2\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\sum_{j=0}^{n-k}\binom{n-k}{j}p^{n-j}q^j\,\frac{\delta_1^{\alpha(n-k-j+i+2)}}{\alpha(n-k-j+i+2)+1}$$

$$\times\left\{\frac{{}_2F_1\left(\alpha(n-k-j+1)+1,\alpha(n-k-j+i+2)+1;\alpha(n-k-j+i+2)+2;-\frac{u}{\delta_1}\right)}{\delta_1^{\alpha(n-k-j+i+2)+1}}\right.$$

$$\left.-\frac{{}_2F_1\left(\alpha(n-k-j+1)+1,\alpha(n-k-j+i+2)+1;\alpha(n-k-j+i+2)+2;-\frac{u}{\delta_2-u}\right)}{(\delta_2-u)^{\alpha(n-k-j+i+2)+1}}\right\}.$$

(b)

$$\int_{\delta_2-u}^{\delta_2} f_{k-1,k:n}(x,x+u)dx =$$

$$= C_{k,n}\alpha^2\delta_1^{\alpha}p^{k-1}\theta^{n-k+1}\int_{\delta_2-u}^{\delta_2}\left[1-\left(\frac{x}{\delta_1}\right)^{-\alpha}\right]^{k-2} x^{-\alpha-1}(x+u)^{-\alpha(n-k+1)-1}dx$$

$$= C_{k,n}\alpha^2 p^{k-1}\theta^{n-k+1}\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\delta_1^{\alpha(i+1)}\int_{\delta_2-u}^{\delta_2} x^{-\alpha(i+1)-1}(x+u)^{-\alpha(n-k+1)-1}dx$$

$$= C_{k,n}\alpha^2 p^{k-1}\theta^{n-k+1}\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\delta_1^{\alpha(i+1)}$$

$$\times\left\{\int_{\delta_2-u}^{\infty} x^{-\alpha(i+1)-1}(x+u)^{-\alpha(n-k+1)-1}dx - \int_{\delta_2}^{\infty} x^{-\alpha(i+1)-1}(x+u)^{-\alpha(n-k+1)-1}dx\right\}.$$

Efectuando a transformação $x=\frac{u}{w}$ *e aplicando a fórmula (2.1.24), obtém-se*

$$\int_{\delta_2-u}^{\delta_2} f_{k-1,k:n}(x,x+u)dx =$$

$$= C_{k,n}\alpha^2 p^{k-1}\theta^{n-k+1}\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\frac{\delta_1^{\alpha(i+1)}}{\alpha(n-k+i+2)+1}$$

$$\times\left\{\frac{{}_2F_1\left(\alpha(n-k+1)+1,\alpha(n-k+i+2)+1;\alpha(n-k+i+2)+2;-\frac{u}{\delta_2-u}\right)}{(\delta_2-u)^{\alpha(n-k+i+2)+1}}\right.$$

$$\left.-\frac{{}_2F_1\left(\alpha(n-k+1)+1,\alpha(n-k+i+2)+1;\alpha(n-k+i+2)+2;-\frac{u}{\delta_2}\right)}{\delta_2^{\alpha(n-k+i+2)+1}}\right\}.$$

(c)

$$\int_{\delta_2}^{\infty} f_{k-1,k:n}(x,x+u)dx =$$
$$= C_{k,n}\alpha^2\theta^{n-k+2}\int_{\delta_2}^{\infty}(1-\theta x^{-\alpha})^{k-2}x^{-\alpha-1}(x+u)^{-\alpha(n-k+1)-1}dx$$
$$= C_{k,n}\alpha^2\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\theta^{n-k+i+2}\int_{\delta_2}^{\infty}x^{-\alpha(i+1)-1}(x+u)^{-\alpha(n-k+1)-1}dx.$$

Efectuando a substituição $x=\frac{u}{w}$*, e aplicando novamente (2.1.24), segue-se que*

$$\int_{\delta_2}^{\infty} f_{k-1,k:n}(x,x+u)dx =$$
$$= C_{k,n}\alpha^2\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\frac{\theta^{n-k+i+2}}{\alpha(n-k+i+2)+1}$$
$$\times\ \frac{{}_2F_1\left(\alpha(n-k+1)+1,\alpha(n-k+i+2)+1;\alpha(n-k+i+2)+2;-\frac{u}{\delta_2}\right)}{\delta_2^{\alpha(n-k+i+2)+1}}.$$

Substituindo as expressões obtidas, vem após simplificarmos

$$f_{U_{k,n}}(u) = \frac{\alpha^2 n!}{(k-2)!(n-k)!}\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\left\{\sum_{j=0}^{n-k}\binom{n-k}{j}\frac{p^{n-j}q^j}{\delta_1[\alpha(n-k-j+i+2)+1]}\right.$$
$$\times\left[{}_2F_1\left(\alpha(n-k-j+1)+1,\alpha(n-k-j+i+2)+1;\alpha(n-k-j+i+2)+2;-\frac{u}{\delta_1}\right)\right.$$
$$-\left(\frac{\delta_1}{\delta_2-u}\right)^{\alpha(n-k-j+i+2)+1}\times$$
$$\left.\times\ {}_2F_1\left(\alpha(n-k-j+1)+1,\alpha(n-k-j+i+2)+1;\alpha(n-k-j+i+2)+2;-\frac{u}{\delta_2-u}\right)\right]$$
$$+\frac{1}{\alpha(n-k+i+2)+1}\left[\frac{p^{k-1}\theta^{n-k+1}\delta_1^{\alpha(i+1)}}{(\delta_2-u)^{\alpha(n-k+i+2)+1)}}\right.$$
$$\times\ {}_2F_1\left(\alpha(n-k+1)+1),\alpha(n-k+i+2)+1;\alpha(n-k+i+2)+2;-\frac{u}{\delta_2-u}\right)$$
$$+\frac{\theta^{n-k+1}[\theta^{i+1}-p^{k-1}\delta_1^{\alpha(i+1)}]}{\delta_2^{\alpha(n-k+i+2)+1}}$$
$$\left.\left.\times\ {}_2F_1\left(\alpha(n-k+1)+1),\alpha(n-k+i+2)+1;\alpha(n-k+i+2)+2;-\frac{u}{\delta_2}\right)\right]\right\}$$

com $\theta = p\delta_1^\alpha + q\delta_2^\alpha$.

- *Se* $u > \delta_2 - \delta_1$, *então*

$$f_{U_{k,n}}(u) = \int_{\delta_1}^{\delta_2} f_{k-1,k:n}(x, x+u)dx + \int_{\delta_2}^{\infty} f_{k-1,k:n}(x, x+u)dx. \tag{3.3.17}$$

O segundo integral de (3.3.17) *foi calculado em* (c). *Quanto ao primeiro integral podemos usar o resultado obtido em* (b), *para tal basta substituirmos* $\delta_2 - u$ *por* δ_1.

Assim,
(d)

$$\int_{\delta_1}^{\delta_2} f_{k-1;k}(x, x+u)dx =$$

$$= C_{k,n}\alpha^2 p^{k-1}\theta^{n-k+1}\sum_{i=0}^{k-2}\binom{k-2}{i}(-1)^i\frac{\delta_1^{\alpha(i+1)}}{\alpha(n-k+i+2)+1}$$

$$\times\left\{\frac{{}_2F_1\left(\alpha(n-k+1)+1, \alpha(n-k+i+2)+1; \alpha(n-k+i+2)+2; -\frac{u}{\delta_1}\right)}{\delta_1^{\alpha(n-k+i+2)+1}}\right.$$

$$\left.-\frac{{}_2F_1\left(\alpha(n-k+1)+1, \alpha(n-k+i+2)+1; \alpha(n-k+i+2)+2; -\frac{u}{\delta_2}\right)}{\delta_2^{\alpha(n-k+i+2)+1}}\right\}.$$

Consequentemente, substituindo (c) *e* (d) *em* (3.3.17), *obtém-se após simplificação*

$$f_{U_{k,n}}(u) = \frac{\alpha^2 n!}{(k-2)!(n-k)!}\sum_{i=0}^{k-2}\binom{k-2}{i}\frac{(-1)^i}{\alpha(n-k+i+2)+1}\left\{\frac{p^{k-1}\theta^{n-k+1}}{\delta_1^{\alpha(n-k+1)+1}}\right.$$

$$\times {}_2F_1\left(\alpha(n-k+1)+1, \alpha(n-k+i+2)+1; \alpha(n-k+i+2)+2; -\frac{u}{\delta_1}\right)$$

$$+\frac{\theta^{n-k+1}[\theta^{i+1} - p^{k-1}\delta_1^{\alpha(i+1)}]}{\delta_2^{\alpha(n-k+i+2)+1}}$$

$$\left.\times {}_2F_1\left(\alpha(n-k+1)+1, \alpha(n-k+i+2)+1; \alpha(n-k+i+2)+2; -\frac{u}{\delta_2}\right)\right\}$$

com $\theta = p\delta_1^\alpha + q\delta_2^\alpha$.

Atendendo a que a f.d.p. de $Z_{1:n}$ *é dada por*

$$f_{1:n}(z) = n[1 - F_Z(z)]^{n-1} f_Z(z)$$

obtém-se para f.d.p. de $U_{1,n} = Z_{1:n} - \delta_1$

$$f_{U_{1,n}}(u) = \begin{cases} p\frac{n\alpha}{\delta_1}\left[q + p\left(1 + \frac{u}{\delta_1}\right)^{-\alpha}\right]^{n-1}\left(1 + \frac{u}{\delta_1}\right)^{-\alpha-1} & ,\ 0 < u \leq \delta_2 - \delta_1 \\ \theta^n n\alpha(u + \delta_1)^{-n\alpha-1} & ,\ u > \delta_2 - \delta_1 \end{cases} \quad (3.3.18)$$

onde $\theta = p\delta_1^\alpha + q\delta_2^\alpha$.