**Teste de credibilidade para a
determinação da região em que
as caudas de distribuições são
aproximadas por funções pareto,
com aplicações em análise de risco**

Francisco Martins Moreira

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO GRAU DE MESTRE
EM
CIÊNCIAS

Área de concentração: Matemática Aplicada
Orientador: Prof. Dr. Julio Michael Stern

Durante a elaboração deste trabalho o autor recebeu auxílio financeiro do CNPq.

São Paulo, 05 de fevereiro de 2007 .

Teste de credibilidade para a determinação
da região em que as caudas de distribuições
são aproximadas por funções pareto,
com aplicações em análise de risco

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e defendida por Francisco Martins Moreira,
aprovada pela Comissão Julgadora.

São Paulo, 05 de fevereiro de 2007 .

Banca examinadora:

- Prof. Dr. Julio Michael Stern (Presidente) - IME - USP
- Prof. Dr. Vladimir Belitsky - IME - USP
- Prof. Dr. Basilio Bragança Pereira - COPPE - UFRJ

Dedico esta tese
a meus pais: Eurides e Adonai,
e a meus irmãos Mônica, Adonai e Verônica,
pelo amor e incentivo a mim dedicados.

Agradecimentos

Ao meu orientador, Prof. Dr. Julio M. Stern, por sua orientação geral, empenho pessoal e encorajamento manifestado durante este trabalho.

Ao Prof. Dr. Vladimir Belitsky, IME-USP, em virtude de sua contribuição para o meu crescimento pessoal e acadêmico.

A todos aqueles amigos que de alguma forma, direta ou indiretamente, contribuíram para a conclusão desta dissertação.

RESUMO

A aproximação da cauda de uma distribuição a partir de sua amostra é um problema que tem suas raízes no fato de que a amostra, por si só, geralmente não fornece uma boa aproximação para a cauda; e são dois os motivos para isso: os poucos pontos que a amostra possui na cauda e a grande distância entre esses pontos. Um dos métodos mais utilizados na melhora desta aproximação chama-se Peaks Over Threshold (POT), que baseia-se na Teoria dos Valores Extremos e usa funções de Pareto para tal aproximação. Na execução do POT há uma parte delicada, que consiste em determinar a região onde a cauda pode ser de fato aproximada por uma função de Pareto. O presente trabalho se ocupa com a apresentação e a experimentação de um novo método para a determinação desta região, método esse que, implementado computacionalmente, faz uso da teoria das Misturas de Distribuições, do método de Monte Carlo, do Full Bayesian Significance Test (veja [16]) e do Teorema de Pickands (veja [18]). Seu desempenho foi testado para dois conjuntos de dados, um artificial, outro referente a um histórico de perdas do índice NASDAQ.

ABSTRACT

The approximation of a tail of a probability distribution from its sample is a problem which difficulty is caused basically by two facts: the fewness of the sample points that came from the tail and their sparceness. One of the popular methods that solves this problem is the so-called Peaks-over-threshold method; it is based on facts from the Extreme Value Theory and uses Pareto's distributions as approximators. A relatively complicated and delicate step of the execution of the Peaks-over-threshold method is the determination of the region where the distribution tail indeed can be approximated by a Pareto distribution. The present work constructs and tests a new approach for the region determination. The approach uses the Bayesian credibility test (see [16]), Pickands' theorem (see [18]), Monte-Carlo method and distribution mixtures. The efficiency of the approach is tested on two data bases: one has been created artificially, the other one is the daily returns of NASDAQ.

Conteúdo

1	Introdução	1
2	Funções Generalizadas de Pareto	3
2.1	Distribuições Padrão de Pareto generalizadas	3
2.2	Distribuições de Pareto generalizadas	7
2.3	Caudas finitas e infinitas	10
2.4	Influência dos parâmetros de uma GPD no seu formato e suas conseqüências na estimativa do risco	13
3	Teoria da distribuição assintótica de extremos	21
3.1	O problema da distribuição assintótica de extremos	21
3.2	A solução do problema da distribuição assintótica de extremos	27
3.3	Da relação entre um problema financeiro e o problema da distribuição assintótica de extremos	32
4	Resultado de Pickands	37
4.1	Preliminares	37
4.2	Resultado de Pickands	38
4.3	Da relação entre as EVD's e as GPD's	40

5	Breve exposição sobre MCMC, FBST e Misturas de Distribuições	45
5.1	Misturas de Distribuições	45
5.1.1	misturas paramétricas	46
5.1.2	estimativa dos parâmetros	46
5.2	MCMC	47
5.2.1	Cadeias de Markov	47
5.2.2	Integração de Monte Carlo	49
5.2.3	Metropolis-Hasting	52
5.3	FBST	53
6	Desenvolvimento do método	55
6.1	Fundamentos Conceituais	55
6.2	Estrutura do desenvolvimento	56
6.3	Exemplo1	57
6.3.1	dados iniciais	57
6.3.2	Determinação das prioris	58
6.3.3	Cálculo da evidência	62
6.3.4	Resultados do Exemplo1	65
6.4	Exemplo2	65
6.4.1	Obtenção das densidades marginais	67
6.4.2	Cálculo da evidência	69
6.4.3	Resultados do exemplo 2	70
6.5	Considerações finais	72

Capítulo 1

Introdução

Um dos assuntos que ocupa lugar importante na pesquisa em finanças está relacionado às grandes perdas. Questões como perda máxima em determinado período, ocorrências extremas de um ativo, e relação entre grandes oscilações de ativos são comuns no mundo financeiro, e nos remetem à teoria dos valores extremos; uma teoria que busca descrever o comportamento desses eventos raros e que pode ser encontrada, por exemplo, em Embrechts, Klüppelberg e Mikosch(1999) ou Coles(2001).

Em todas as situações mencionadas acima a cauda da distribuição desempenha papel fundamental; DuMouchel(1983) destaca os comportamentos distintos das partes central e caudal da distribuição; Smith(1987) menciona duas questões teóricas cruciais na modelagem de caudas. Uma a da escolha entre diferentes estimadores; outra a da escolha do limiar ótimo (que será tratada neste trabalho). O artigo de McNeil(1996) chama a atenção para a pequena quantidade de pontos disponíveis na análise dos valores extremos, e também ressalta a sensibilidade das inferências à introdução de novas perdas extremas (algo que lembra a teoria bayesiana). No mesmo trabalho há considerações sobre a influência do limiar u na modelagem das caudas de uma distribuição. Em Smith(1987) ou Danielsson e de Vries(1997) a abordagem é feita em torno da ponderação entre *bias* e variância, que aparece quando o limiar é tomado muito antes ou muito depois do ponto de limiar ótimo, respectivamente.

A questão principal a ser tratada nessa dissertação é a de obter informações sobre o limiar u que melhor modele os valores extremos de uma distribuição empírica, ou seja, um valor u onde os pontos da amostra maiores que u sejam satisfatoriamente modelados por uma GPD. Devido ao trabalho de Pickands(1975), as GPDs (Generalized Pareto Distributions) foram eleitas as melhores distribuições, sob certas condições, para a modelagem da cauda de uma distribuição; Embrechts *et al*(1997), Mcneil(1997) e Belitsky(2006) exibem métodos que exemplificam essa modelagem e que, num dado instante, têm de estimar um valor para o limiar ótimo. Existem várias maneiras de encontrar uma estimativa. Belitsky(2006) e

McNeil(1997) utilizam a análise gráfica, Smith(1987) chega a uma quantificação entre *bias* e variância para escolhas assintóticas ótimas do limiar, e finalmente Danielsson e de Vries(1997) procuram um comprimento ótimo para a cauda que visa minimizar a *mean-squared error(MSE)loss function*, para também produzir uma ponderação ótima entre variância e *bias*.

Este trabalho descreve um novo método que objetiva encontrar o início do comportamento caudal de uma distribuição via a junção entre o teste de significância FBST, desenvolvido por Stern e Pereira(1999), e o resultado de Pickands(1975) sobre o comportamento de uma distribuição de excedentes.

Capítulo 2

Funções Generalizadas de Pareto

Chama-se *distribuição de Pareto generalizada* (GPD¹ devido a seu equivalente em Inglês “Generalized Pareto Distribution”) a função que possui uma das formas (2.6), (2.7), (2.8) apresentadas em seguida. As funções das formas (2.6) e (2.7) foram estudadas por Pareto – daí o termo “distribuições de Pareto” no nome dado a elas. Já a forma (2.8) pode ser vista como uma generalização das formas (2.6) e (2.7) – daí o acréscimo do termo “generalizadas”.

Antecipando a definição de uma GPD, notamos que a expressão desta função possui três parâmetros, que serão denotados por ξ, β e s . No nosso ponto de vista a influência desses parâmetros no formato da GPD é mais simples de ser entendida quando a construção das GPDs é realizada em duas etapas: uma na qual contrói-se GPDs com $\beta = 1$ e $s = 0$ – tais GPDs são chamadas *GPDs padrão*; outra onde mostra-se como as GPDs padrão podem ser transformadas para se obter todas as outras possíveis GPDs. Este caminho foi adotado na nossa apresentação. A primeira etapa será realizada na seção 2.1 e a segunda etapa na seção 2.2.

2.1 Distribuições Padrão de Pareto generalizadas

As funções chamadas *distribuições padrão de Pareto generalizadas*² (abreviadas por *GPDs padrão* no texto) distinguem-se entre si por meio dos diferentes valores de um único parâmetro. Este parâmetro, designado por ξ , pode assumir qualquer valor real. A GPD

¹Observe que “distribuição de Pareto generalizada” será abreviada por GPD, enquanto que “distribuições de Pareto generalizadas” serão abreviadas por GPDs.

²*Standard generalized Pareto distributions* em Inglês.

Figura 2.1: O formato da GPD padrão com diversos valores do parâmetro ξ .

padrão com valor $\xi > 0$ possui a seguinte forma (veja Figura 2.1):

$$G_{\xi}(x) = \begin{cases} 0, & \text{para } x < 0 \\ 1 - (1 + \xi x)^{-1/\xi}, & \text{para } x \geq 0 \end{cases} \quad \left(\begin{array}{l} \text{esta fórmula} \\ \text{vale quando } \xi > 0 \end{array} \right) \quad (2.1)$$

A GPD padrão com valor $\xi = 0$ possui a seguinte forma³ (veja Figura 2.1):

$$G_0(x) = \begin{cases} 0, & \text{para } x < 0 \\ 1 - e^{-x}, & \text{para } x \geq 0 \end{cases} \quad \left(\begin{array}{l} \text{esta função} \\ \text{corresponde a } \xi = 0 \end{array} \right) \quad (2.2)$$

A GPD padrão com valor $\xi < 0$ possui a seguinte forma (veja Figura 2.1):

$$G_{\xi}(x) = \begin{cases} 0, & \text{para } x < 0 \\ 1 - (1 + \xi x)^{-1/\xi}, & \text{para } x \in [0, -\frac{1}{\xi}] \\ 1, & \text{para } x > -\frac{1}{\xi} \end{cases} \quad \left(\begin{array}{l} \text{esta fórmula} \\ \text{vale quando } \xi < 0 \end{array} \right) \quad (2.3)$$

Quanto às funções de densidade das GPDs padrão, a fórmula (2.4) dá suas expressões analíticas, enquanto que as Figuras 2.2 e 2.3 ilustram suas formas geométricas:

$$\begin{aligned} f_{\xi}(x) &= (\text{a função de densidade de } G_{\xi}(\cdot)) = \frac{dG_{\xi}(x)}{dx} = \\ &= \begin{cases} (1 + \xi x)^{-1-1/\xi} & \text{para } x \geq 0, \text{ e } 0 \text{ para outros } x, \text{ quando } \xi > 0 \\ e^{-x} & \text{para } x \geq 0, \text{ e } 0 \text{ para outros } x, \text{ quando } \xi = 0 \\ (1 + \xi x)^{-1-1/\xi} & \text{para } x \in [0, -1/\xi], \text{ e } 0 \text{ para outros } x, \text{ quando } \xi < 0 \end{cases} \quad (2.4) \end{aligned}$$

³Nos comentários 1 e 2 há uma discussão detalhada a respeito desse caso especial, uma vez que aqui a expressão da distribuição, G_0 , não depende do parâmetro ξ .

Figura 2.2: Funções de densidade de GPDs padrão com $\xi \geq 0$.

Comentário 1. Cada GPD padrão é “parente” de uma hipérbole, uma parábola, ou uma função exponencial.

Caso $\xi > 0$, o gráfico da GPD padrão $G_\xi(\cdot)$ no semi-eixo $[0; +\infty)$ é apenas ramo direito da hipérbole $1/(x^{1/\xi})$ que foi refletido, deslocado e re-escalado para que a hipérbole passasse pelo ponto $(0, 0)$ e se aproximasse do nível 1 quando $x \rightarrow \infty$.

No caso $\xi = 0$, a correspondente GPD padrão é exatamente a função de distribuição exponencial de parâmetro 1. Lembramos ao leitor a nomenclatura comumente usada para distribuições exponenciais: A função que vale 0 para $x < 0$ e vale $1 - e^{-\lambda x}$ para $x \geq 0$ denomina-se “distribuição exponencial de parâmetro λ ”. Quando $\lambda = 1$, esta função chama-se simplesmente “distribuição exponencial” – termo que usaremos. Apesar disso alguns autores preferem usar o nome “distribuição exponencial padrão” quando $\lambda = 1$.

O caso $\xi < 0$ é mais “complexo” que os casos $\xi > 0$ e $\xi = 0$, uma vez que a forma de $G_\xi(\cdot)$ é radicalmente transformada em consequência da passagem do valor de ξ de 0 a $-\infty$. A Figura 2.1 ilustra o caso $-1 < \xi < 0$, onde a função $G_\xi(\cdot)$ comporta-se no intervalo $[-1/\xi, 0]$ como função-potência de x (isto é, como x^a para $a > 0$). Ainda neste caso, $G_\xi(\cdot)$ aproxima-se suavemente o nível 1, isto é, sua derivada é 0 no ponto onde ela toca o nível 1. Esta propriedade da derivada está ilustrada na Figura 2.3. No caso $\xi = -1$, a correspondente

Figura 2.3: Funções de densidade de GPDs padrão com $\xi < 0$.

GPD padrão cresce linearmente entre as abscissas 0 e 1, ali a GPD padrão apresenta “quebra” no ponto onde toca o nível 1 (veja Figura 2.1). Finalmente, no caso $\xi < -1$, a GPD padrão se comporta como uma hipérbole no intervalo $[0; -1/\xi]$ (isto é, como x^a para $a < 0$). Neste caso também há “quebra” do gráfico da função no ponto em que o gráfico toca o nível 1 (veja Figura 2.1).

Observemos ainda que no caso $\xi < 0$ a função de densidade pode apresentar cinco padrões diferentes, apresentados na Figura 2.3.

Comentário 2. Depois do Comentário 1 esclarecer que a GPD padrão correspondente ao valor 0 para o parâmetro ξ é simplesmente a distribuição exponencial, qualquer leitor faria a seguinte pergunta: “Porque à distribuição exponencial foi atribuído valor 0 para o parâmetro ξ ? Ora, não há “0” ou “ ξ ” na fórmula desta função de distribuição!” O presente comentário responde esta pergunta.

Foi necessário colocar a função de distribuição exponencial no mesmo conjunto onde ficam as funções do tipo (2.1) e (2.3). A razão desta necessidade é o fato de tanto estas funções quanto a da distribuição exponencial servirem para a aproximação de caudas. Pela mesma razão, era desejável indexar todas as funções deste conjunto. A indexação das funções do tipo (2.1) e (2.3) é natural: o índice de qualquer uma destas funções é seu próprio valor do parâmetro ξ . Com isto o valor 0 para o parâmetro ξ torna-se livre (simplesmente porque não pode-se colocar $\xi = 0$ na expressão $1 - (1 + \xi x)^{-1/\xi}$ das fórmulas (2.1) e (2.3)). Este valor livre foi então atribuído ao novo membro do conjunto, à distribuição exponencial. Note que esta atribuição é puramente semântica, pois, de fato, não há ξ na fórmula da distribuição exponencial. Em palavras simples, a “atribuição semântica” se traduz no seguinte acordo:

se eu lhe pedir a GPD padrão correspondente, por exemplo, a $\xi = 2$, sua resposta será: “ $1 - (1 + 2x)^{-1/2}$ ”, enquanto que se eu lhe pedir a GPD padrão correspondente a $\xi = 0$, sua resposta será: “é a função de distribuição exponencial, quer dizer, $1 - e^{-x}$ ”.

O leitor, que concordou com a lógica dos argumentos do parágrafo anterior, pode agora argumentar: “São, na verdade, dois valores de ξ que não são “aproveitados” para a indexação das funções do tipo (2.1) e (2.3); estes são 0 e ∞ . Então existe razão para que seja 0 e não ∞ o índice atribuído à distribuição exponencial?” Sim, existe. Esta razão é o fato da função exponencial ser o limite das funções $1 - (1 + \xi x)^{-1/\xi}$ quando $\xi \rightarrow 0$.

2.2 Distribuições de Pareto generalizadas

As GPDs, Funções de Pareto Generalizadas, serão construídas nesta sub-seção a partir das GPDs padrão e, para que a essência desta construção fique bem clara, começaremos com dois fatos bem conhecidos.

Fato 1. Seja $f(x)$, $x \in \mathbb{R}$, uma função arbitrária qualquer, d e c duas constantes quaisquer, sendo que $c > 0$. Construiremos uma nova função $g(\cdot)$ via a igualdade $g(x) = f\left(\frac{x-d}{c}\right)$, $x \in \mathbb{R}$.

O objetivo do presente comentário é apresentar o seguinte fato: o gráfico da função $g(\cdot)$ pode ser obtido do gráfico da função $f(\cdot)$ via as duas transformações (veja na Figura 2.4 estas transformações para o caso particular quando $f(x) = e^x$):

- (1) esticar c vezes o gráfico de $f(\cdot)$ ao longo do eixo x ;
- (2) deslocar por d unidades à direita o gráfico obtido no item anterior.

Que (1)-(2) de fato transformam $f(\cdot)$ em $g(\cdot)$ pode ser justificado da seguinte maneira: Escolha um ponto (x, y) no gráfico de $f(\cdot)$ e pergunte: qual deve ser x^{novo} para que⁴ $g(x^{\text{novo}}) = y$? Suponha, para facilitar a conta, que $f(\cdot)$ é uma função um-a-um. Então, vale que: $g(x^{\text{novo}}) = y \Leftrightarrow f\left(\frac{x^{\text{novo}}-d}{c}\right) = y \Leftrightarrow f\left(\frac{x^{\text{novo}}-d}{c}\right) = f(x) \Leftrightarrow x^{\text{novo}} = c \cdot x + d$. A última igualdade diz então que x^{novo} é obtido quando x é multiplicado por c e acrescenta-se d ao resultado da multiplicação. Isto justifica que (1)-(2) transformam $f(\cdot)$ em $g(\cdot)$ (veja Figura 2.4).

⁴Ou seja, que mudanças é preciso realizar no eixo das abscissas para que as ordenadas dos gráficos de f e g coincidam?

Figura 2.4: $(x, f(x))$ corresponde ao gráfico de e^x ; $(x, f(x/2))$ ao gráfico de $e^{\frac{x}{2}}$ e $(x, f((x-3)/2))$ ao gráfico de $e^{\frac{x-3}{2}}$.

Fato 2. Considere $F_X(\cdot)$ a função de distribuição de uma variável aleatória arbitrária X , sendo d e $c > 0$ números reais quaisquer e Y uma nova variável aleatória definida por $Y = cX + d$. Denotaremos por $F_Y(\cdot)$ a função de distribuição de Y .

Segue das construções e definições agora feitas que

$$\begin{aligned} F_Y(x) &= IP[Y \leq x] = IP[cX + d \leq x] = \\ IP\left[X \leq \frac{x-d}{c}\right] &= F_X\left(\frac{x-d}{c}\right), \quad \forall x \in \mathbb{R} \end{aligned} \quad (2.5)$$

A relação (2.5) juntamente com o Fato 1 sustenta que se a função de distribuição de uma variável aleatória X for transformada via os passos (1)-(2) do Fato 1, o resultado será uma nova função de distribuição; ainda mais, se designarmos por Y a variável aleatória distribuída conforme esta nova função de distribuição, Y e X se relacionam via $Y = cX + d$.

Passaremos então a decrever a família de *distribuições de Pareto generalizadas* (abreviadas por *GPDs*). Neste caso a distinção entre as GPDs é feita por três parâmetros, ξ, β e s , sendo que ξ e s podem assumir qualquer valor real e β pode assumir qualquer valor real positivo. A GPD com $\xi > 0$ tem a forma:

$$G_{\xi, \beta, s}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-s}{\beta}\right)^{-1/\xi}, & \text{para } x \geq s \\ 0, & \text{para } x < s \end{cases} \quad (\text{quando } \xi > 0) \quad (2.6)$$

Enquanto que a GPD com $\xi = 0$ é dada por:

$$G_{0, \beta, s}(x) = \begin{cases} 1 - e^{-\frac{x-s}{\beta}}, & \text{para } x \geq s \\ 0, & \text{para } x < s \end{cases} \quad (\text{correspondente a } \xi = 0) \quad (2.7)$$

Finalmente a GPD que corresponde a $\xi < 0$:

$$G_{\xi,\beta,s}(x) = \begin{cases} 0, & \text{para } x < s \\ 1 - (1 + \xi \frac{x-s}{\beta})^{-1/\xi}, & \text{para } x \in [s, s - \frac{\beta}{\xi}] \\ 1, & \text{para } x > s - \frac{\beta}{\xi} \end{cases} \left(\begin{array}{l} \text{esta fórmula} \\ \text{vale quando } \xi < 0 \end{array} \right) \quad (2.8)$$

Já as funções de densidade relativas a (2.6), (2.7) e (2.8) são dadas pelas expressões

$$g_{\xi,\beta,s}(x) = \begin{cases} \frac{1}{\beta} \left(1 + \xi \frac{x-s}{\beta}\right)^{-1/\xi-1}, & \text{para } x \geq s \\ 0, & \text{para } x < s \end{cases} \left(\begin{array}{l} \text{lembre-se, isto vale} \\ \text{quando } \xi > 0 \end{array} \right) \quad (2.9)$$

$$g_{0,\beta,s}(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x-s}{\beta}}, & \text{para } x \geq s \\ 0, & \text{para } x < s \end{cases} \left(\text{correspondente a } \xi = 0 \right) \quad (2.10)$$

$$g_{\xi,\beta,s}(x) = \begin{cases} 0, & \text{para } x < s \\ \frac{1}{\beta} (1 + \xi \frac{x-s}{\beta})^{-1/\xi-1}, & \text{para } x \in [s, s - \frac{\beta}{\xi}] \\ 0, & \text{para } x > s - \frac{\beta}{\xi} \end{cases} \left(\begin{array}{l} \text{esta fórmula} \\ \text{vale quando } \xi < 0 \end{array} \right) \quad (2.11)$$

A transformação descrita nos passos (1)-(2) do Fato 1 deixa claro que qualquer GPD com parâmetros ξ, β e s é obtida da GPD padrão de mesmo valor ξ , com $c = \beta$ e $d = s$, ou, especificamente falando, o gráfico da GPD com parâmetros ξ, β e s é o gráfico da GPD padrão com parâmetro ξ após este ser esticado β vezes ao longo do eixo x e deslocado na direção horizontal por s unidades. Este fato justifica os nomes atribuídos aos parâmetros β e s : o parâmetro β chama-se *parâmetro de escala*, enquanto o parâmetro s é o *parâmetro de deslocamento* ou *de locação*. O parâmetro ξ leva o nome *parâmetro de forma*.

A relação entre a GPD com parâmetros ξ, β e s e a GPD padrão de mesmo parâmetro ξ pode ser vista, com a ajuda do conteúdo do Fato 2, em termos da relação entre as variáveis aleatórias correspondentes a estas distribuições: considerando Y a variável aleatória com distribuição $G_{\xi,\beta,s}(\cdot)$ e X a variável aleatória com distribuição $G_{\xi}(\cdot)$ então X e Y são tais que $Y = \beta X + s$. Esta relação pode ser útil a quem costuma pensar em uma função de distribuição em termos da variável aleatória que possui tal distribuição. A relação diz que se tomarmos X distribuída conforme uma GPD padrão com parâmetro ξ , multiplicarmos esta por β e acrescentarmos s , então a variável aleatória resultante terá distribuição GPD com parâmetros ξ, β e s .

2.3 Caudas finitas e infinitas

Na maior parte da exposição a seguir, e especialmente no cap. 6 (veja seção 6.1), onde o problema da determinação do limiar é abordado, assumiremos que a cauda da distribuição desconhecida, $F(\cdot)$, é infinita; uma aceitação que merece uma análise mais cuidadosa uma vez que *a priori* não foi dado nenhum motivo especial para tal aceitação. Começaremos então nossa análise com o conceito de cauda de uma distribuição.

Comentário 3. A palavra “cauda” possui duas interpretações. Na primeira a cauda de $F(\cdot)$ à direita de u é a parte do seu gráfico à direita da abscissa u . Se aceitarmos esta interpretação, então a cauda de qualquer função de distribuição só pode ser infinita, pois o gráfico de uma função de distribuição nunca “para”. A outra interpretação é que a cauda de $F(\cdot)$ à direita de u é a parte de plano delimitada pela reta $y = 1$ por cima, pelo gráfico de $F(\cdot)$ por baixo, e pela reta $x = u$ à esquerda. É esta interpretação que deve ter-se em mente quando diz-se “cauda finita” ou “cauda infinita”. Vejamos um exemplo.

Considere a função $G_\xi(\cdot)$ com $\xi = -1/4$ apresentada na Figura 2.1. Ela “toca” o nível 1 no ponto $x = 4$ e, após isso, é sempre igual a 1. Neste caso diz-se que a função de distribuição tem cauda direita finita. Usaremos o termo mais curto *cauda finita*, pois em todo o trabalho analisaremos somente caudas à direita. Para facilitar expressões formais, introduziremos o símbolo x_F para designar o *fim da cauda da distribuição* $F(\cdot)$, que é a abscissa do ponto onde $F(\cdot)$ toca em 1. Por exemplo, na Figura 2.1, temos que $x_{G_\xi} = -1/\xi$ para $\xi < 0$.

Observe agora a função da distribuição exponencial apresentada na Figura 2.1 (correspondente a $\xi = 0$). Ela nunca toca o nível 1 (simplesmente porque não existe x tal que $1 - e^{-x} = 1$). Neste caso a função de distribuição tem cauda direita infinita. Usaremos o termo simplificado *cauda infinita*, por razões já explicadas acima. A fim de aproveitar as mesmas fórmulas tanto para funções de cauda finita quanto para as de cauda infinita, assumiremos que x_F é igual a ∞ se $F(\cdot)$ possui cauda infinita.

Comentário 4. A cauda direita de uma função de distribuição tem relação com o fato dos valores da variável aleatória, que possui esta distribuição, serem limitados ou não: se $F(\cdot)$ possuir cauda direita finita, então a variável aleatória X distribuída conforme $F(\cdot)$, será limitada por cima e o limite superior será x_F , quer dizer, $\mathbb{P}[X \leq x_F] = 1$; já se $F(\cdot)$ possuir cauda direita infinita, então a variável aleatória X , distribuída conforme $F(\cdot)$, não será limitada por cima, quer dizer, não existe um número finito M tal que $\mathbb{P}[X \leq M] = 1$.

Como foi dito no início desta seção, quando apresentarmos o método da análise do limiar, no cap. 6, assumiremos o seguinte pressuposto⁵

⁵O surpreendente surgimento da discussão do Pressuposto 1 no capítulo destinado às GPDs tem duas justificativas: em primeiro lugar as GPDs servirão como exemplos ilustrativos para a explicação dos conceitos “cauda finita” e “cauda infinita”, necessários para um bom entendimento sobre as GPDs; em segundo lugar,

Pressuposto 1: A cauda de $F(\cdot)$ é infinita.

Dessa aceitação surgem diversas questões, das quais as três seguintes são, ao nosso ver, as principais – elas serão cuidadosamente respondidas no restante desta seção.

- (a) Existem procedimentos que indicam se a cauda é finita ou infinita?
- (b) Quais são as vantagens obtidas quando se assume o Pressuposto 1? Há também desvantagens? Se sim, qual é o prejuízo?
- (c) Porque escolhemos trabalhar com a caso de cauda infinita e não com o caso cauda finita?

A resposta à pergunta (a) é a mais fácil. Se temos uma amostra de uma função de distribuição $F(\cdot)$ desconhecida, então podemos inferir, a partir da amostra, se a cauda de $F(\cdot)$ é finita ou infinita. Essa tarefa é delegada à atual Teoria de Valores Extremos, que fornece diversos procedimentos estatísticos que fazem esta inferência. As referências [6] e [19] são fontes ricas para quem deseja conhecer e aplicar estes procedimentos.

Note-se que apresentar e discutir os procedimentos estatísticos mencionados no parágrafo acima não é o objetivo do nosso texto; se fosse, seu volume seria muito maior, sem haver bons motivos para tal. Esta é uma das razões de termos assumido o Pressuposto 1.

Uma outra razão vem da existência de pequenas diferenças na execução do método descrito no capítulo 6 dependendo de $F(\cdot)$ possuir cauda finita ou infinita. As diferenças são pequenas, mas a inclusão de ambos os casos seria desgastante e repetitiva. Por isto optamos por um só caso, preferindo nos concentrar naquele onde $F(\cdot)$ possui cauda infinita.

A última razão a ser considerada decorre do fato da teoria de ajuste de caudas via GPD usar unicamente GPDs com parâmetro ξ não-negativo para aproximar a cauda de $F(\cdot)$, quando esta possui cauda infinita. Isto ficará claro na formulação do Resultado de Pickands e na dedução da idéia do método a partir deste resultado (veja o comentário 15 do capítulo 4). Este fato nos permite concentrar a atenção nas GPDs de parâmetro ξ não-negativo, ou, em outras palavras, às GPDs definidas por (2.1) e (2.2), ou, ainda, às GPDs com caudas infinitas. Dessa forma o volume da nossa apresentação é reduzido. As três razões apresentadas acima respondem a questão (b).

Finalmente responderemos a pergunta (c). Observe a importância desta pergunta. Os argumentos acima indicam que para tornar a exposição mais concisa poderíamos concentrar nossa atenção tanto às funções com cauda finita quanto às funções com cauda infinita. Pergunta-se: Por que optamos pelas segundas? A “preferência” pelas caudas infinitas deve-se à combinação de dois fatores: o primeiro é que a maioria das aplicações referentes às idéias aqui abordadas se dedica à estimativa de caudas de distribuições de perdas de retornos de investimentos financeiros, perdas decorrentes de sinistros de seguradora, ou de perdas

a aceitação do Pressuposto 1 justificará, conforme a ser explicado logo em seguida, porque as GPDs com cauda infinita irão demandar maior atenção que as de cauda finita em nossa apresentação.

causadas por falhas operacionais de um banco ou uma indústria; o segundo fato reside na crença comum de que tais distribuições possuem cauda infinita. Isto então responde à pergunta (c), mas com um teor de insatisfação: porque o segundo fato foi colocado como “crença”? Sobre isto discutiremos agora.

Considere a variável aleatória “retorno (digamos, diário, para título de exposição) com sinal negativo de um investimento”. Observe que para que

$$\text{retorno diário com sinal negativo} = -\log \frac{\text{preço de hoje}}{\text{preço de ontem}} \quad (2.12)$$

seja ilimitado por cima, o “preço de hoje” deve assumir um valor arbitrariamente pequeno em relação ao “preço de ontem”. Se isto ocorre, de acordo com o comentário 4, a cauda direita de $F(\cdot)$ deve ser infinita. Acontece que a ilimitação alegada acima discorda com a realidade devido ao seguinte fato: o “preço de hoje” de uma ação nunca será menor que 1 centavo dividido pela quantidade de todas as ações da empresa, enquanto que o “preço de ontem” nunca pode ser maior que a quantidade de todo o dinheiro do mundo. Portanto, rigorosamente falando, a variável aleatória “retorno diário com sinal menos” está limitada por cima por um valor x_F e a função da distribuição desta variável aleatória salta para 1 no ponto x_F . Como então “crer que a cauda de $F(\cdot)$ deve ser infinita?”.

É fato, como vimos acima, que para a nossa $F(\cdot)$ existe x_F onde $F(x_F) = 1$; entretanto este limite x_F é imenso. Na literatura são relatadas diversas análises com base nas amostras de retornos de diversos ativos financeiros, e praticamente todas essas análises concordam que tais amostras provêm de uma distribuição com cauda infinita. Isto ocorre porque o ponto x_F onde a verdadeira distribuição de retornos toca ao nível 1 é tão distante, e a probabilidade de observar valores perto deste ponto é tão insignificante, que tais valores, determinantes na escolha de uma cauda finita, simplesmente não aparecem em nenhuma amostra real. Pergunta-se então: Vale a pena desconsiderar as caudas finitas quando o objeto de análise são amostras de retornos de ativos financeiros? Apenas as caudas infinitas modelam a cauda da distribuição de tais retornos? Infelizmente não há uma resposta única a esta questão, apropriada a todos os casos. Para justificar esta afirmação acompanhe o exemplo abaixo.

Seja $G_{fin}(\cdot)$ a função-aproximador da cauda no caso de impormos que a cauda deva ser finita, e $G_{inf}(\cdot)$ a função aproximador da cauda no caso de assumirmos que a cauda seja infinita; que fique claro: trata-se das aproximações fornecidos pelo método POT⁶ a partir da mesma amostra. Assim, as funções $G_{fin}(\cdot)$ e $G_{inf}(\cdot)$ serão relativamente afastadas uma da outra para aqueles valores muitíssimo próximos de onde $G_{fin}(\cdot)$ toca no nível 1. Caso tais valores sejam importantes para fins de modelagem da cauda, é preciso insistir que a cauda seja finita, caso contrário – não. Por exemplo, caso a finalidade seja a estimação de risco via o conceito VaR_p (a ser explicado na seção 2.4), então pode-se considerar a cauda como

⁶O POT é um método específico de ajuste de cauda, descrito em detalhes em [1]

sendo infinita, pois a diferença entre ela e a cauda finita afetaria VaR_p para os valores de p excessivamente próximos de 1, desinteressantes para o cálculo de VaR_p (veja o comentário 5 no final desta seção). Caso o leitor queira exemplos de onde a cauda não possa ser infinita, eis um: O VaR_p condicional de uma variável aleatória é a média de seus valores acima do limiar p .⁷ Esta média pode ser infinita caso a distribuição da variável aleatória possua cauda infinita. Se o risco de retorno for estimado por este VaR_p condicional, então a infinitude do resultado pode não ter interpretação prática. Portanto, neste caso faz sentido insistir que a cauda da função aproximador seja finita.

Portanto, devido à discussão acima, é razoável trabalhar com a “crença” de que retorno diário é uma variável aleatória ilimitada. Já se formos falar de $F(\cdot)$ como a distribuição de perdas de uma seguradora devido ocorrência de sinistros, ou de perdas de um banco devido ocorrências de falhas operacionais, seremos também obrigados a assumir que teoricamente a perda é limitada pelo volume de todo o dinheiro disponível no mundo. Porém este limite é tão imenso, que na prática a cauda de $F(\cdot)$ se comporta como se fosse ilimitada. E justifica-se este pressuposto pelo mesmo argumento utilizado acima para o caso de retornos.

Comentário 5. Nosso leitor pode estranhar o fato de uma função de cauda infinita ser uma boa aproximação da cauda finita de uma outra função. Eis o exemplo que mostra essa possibilidade. Volte a atenção à figura 2.1. Observe como são próximas as GPDs $G_{-1/4}(\cdot)$ e $G_0(\cdot)$ no intervalo $[0, 4]$. Isto deve convencer que as GPDs $G_{-1/100}(\cdot)$ e $G_0(\cdot)$ serão ainda mais próximas no intervalo $[0, 100]$. No ponto $x = 100$ a função $G_{-1/100}(\cdot)$ toca o nível 1, enquanto que o valor de $G_0(\cdot)$ nesse ponto é e^{-100} . Isto significa que em qualquer ponto à direita de $x = 100$ as duas funções não se distanciam por mais de e^{-100} . Portanto $G_0(\cdot)$ pode ser um bom aproximador para $G_{-100}(\cdot)$, apesar da cauda da função-aproximador ser infinita, enquanto que a da função aproximada é finita.

2.4 Influência dos parâmetros de uma GPD no seu formato e suas consequências na estimativa do risco

O assunto a ser discutido na presente seção é fundamental para aqueles interessados na resolução do problema de estimação de risco. Consideraremos aqui aqueles métodos onde a cauda da distribuição $F(\cdot)$ (não conhecemos a cauda da distribuição $F(\cdot)$) é aproximada via algum método de ajuste, tendo como objetivo uma estimativa para $F^{-1}(\cdot)$, onde o valor de p é algo próximo de 1, fixado de antemão. Se designarmos por $T(\cdot)$ a função de distribuição fornecida pelo procedimento escolhido, então $T^{-1}(p)$ será a resposta dada por esta abordagem. Acontece que a cauda de $T(\cdot)$ será construída da GPD $G_{\xi, \beta, s}(\cdot)$, cujos valores numéricos

⁷Não discutiremos e não usaremos VaR_p condicional no presente trabalho.

Figura 2.5: À esquerda: Se supusermos que $F_B^{-1}(p) \geq F_A^{-1}(p) \geq x_0$; para toda abscissa x maior que x_0 temos que $F_A(x) \geq F_B(x)$. À direita: Da mesma suposição anterior decorre também que $F_B^{-1}(p) = VaRpB \geq F_A^{-1}(p) = VaRpA$ para todo p maior que y_0 .

dos parâmetros ξ, β e s são estimados com base nos pontos caudais de $F(\cdot)$; e sabemos que toda estimativa carrega consigo alguma imprecisão⁸, um desvio em relação aos dados reais. No que se refere ao problema de estimação de risco, esta dúvida assume a seguinte forma: de que modo uma imprecisão nas estimativas de ξ, β e s afeta o valor de $G_{\xi, \beta, s}^{-1}(p)$? Para responder este problema, é preciso entender

(P4) como uma mudança nos valores de ξ, β e s afeta os valores de $G_{\xi, \beta, s}^{-1}(p)$ quando p está próximo de 1?

Este problema é a semente que originou a presente seção. Estudaremos as propriedades das GPDs que nos permitirão responder a pergunta (P4). Iniciaremos com a apresentação de um fato que facilitará a argumentação e permitirá a interpretação da resposta em termos de riscos de investimentos.

Fato 3. Imagine dois investimentos financeiros hipotéticos com horizonte de, digamos, um dia num mercado com taxa de juros livre de risco, e igual a 0. Chamaremos os investimentos de “A” e “B” e designaremos pelas variáveis aleatórias A e B , respectivamente, seus retornos diários com sinal “-”. Considere $F_A(\cdot)$ e $F_B(\cdot)$ as funções de distribuição das variáveis aleatórias A e B . Nessas condições as afirmações (a), (b), (c), (d) são equivalentes (veja figura 2.5):

⁸Essa imprecisão pode ter origem em vários lugares, na falta de representatividade da amostra, na escolha de um procedimento estatístico inadequado ou nas questões computacionais.

Figura 2.6: As funções de distribuição $F_A(\cdot)$ e $F_B(\cdot)$ apresentadas na Figura (a) satisfazem $F_A(x) \geq F_B(x)$ para todo $x \in \mathbb{R}$, e as apresentadas na Figura (b) satisfazem $F_A(x) \geq F_B(x)$ para todo $x \geq x_0$. Em ambos os casos as funções são iguais a 0 para todo $x \leq 0$; isto é assumido somente para facilitar a apresentação.

- (a) $F_A^{-1}(p) \leq F_B^{-1}(p)$ quando ambos são maiores que $x_0 \geq 0$;
- (b) o gráfico de $F_A(\cdot)$ fica à esquerda de $F_B(\cdot)$ para toda ordenada maior que $F_A(x_0) = F_B(x_0) = y_0 \geq 0$;
- (c) o gráfico de $F_A(\cdot)$ fica acima do de $F_B(\cdot)$ para toda abscissa maior que $x_0 \geq 0$;
- (d) o investimento A é menos arriscado que B para grandes valores de perdas, especificamente para perdas com retorno negativo maior em módulo que $x_0 \geq 0$.

(2.13)

Quanto à demonstração da alegada equivalência, notamos que (a) $\Leftrightarrow$ (b) segue imediatamente da definição de $F^{-1}(p)$; que (b) $\Leftrightarrow$ (c) é um fato genérico cuja derivação baseia-se no bom senso, e finalmente, que (a) $\Leftrightarrow$ (d) é a consequência das considerações

- (a) $F(\cdot)$ seja a função de distribuição do retorno com sinal “-” de um investimento; e
- (b) $F^{-1}(p)$ seja um valor positivo,

(2.14)

então $F^{-1}(p)$ pode ser interpretado como o risco ao nível p do investimento. Lembramos que é exigido pelo (b) de (2.14) que $F^{-1}(p)$ seja um valor positivo, pois isto permite interpretar $F^{-1}(p)$ como perda. A Figura 2.6 ajuda a visualizar as considerações feitas em (2.13).

Agora às propriedades das GPDs que esclarecerem a questão levantada em (P4). A primeira delas é:

$$\begin{aligned} &\text{Se } \xi_1 \leq \xi_2, \text{ então } G_{\xi_1, \beta, s}^{-1}(p) \leq G_{\xi_2, \beta, s}^{-1}(p) \\ &\text{para todo } p \in [0, 1], \text{ para quaisquer } \beta \text{ e } s \end{aligned} \quad (2.15)$$

Para demonstrar a propriedade (2.15), consideraremos primeiramente o caso particular onde $\beta = 1$ e $s = 0$. Já sabemos que a GPD $G_{\xi,1,0}(\cdot)$ é idêntica à GPD padrão $G_{\xi}(\cdot)$, logo temos que comparar GPDs padrão com diversos valores de ξ . Na Figura 2.1 é nítido que quanto menor o valor de ξ mais “alto” é o gráfico da GPD padrão $G_{\xi}(\cdot)$ no eixo⁹ x . Devido às equivalências (2.13) o fato anterior garante a validade da relação (2.15) quando $\beta = 1$ e $s = 0$. No caso genérico, o esboço da demonstração da relação (2.15) seria o seguinte: Sejam $\xi_1 < \xi_2$ e sejam β e s quaisquer. Sabemos que $G_{\xi_1}(\cdot)$ fica acima de $G_{\xi_2}(\cdot)$. Então, se esticarmos ambas as funções β vezes ao longo do eixo x e as deslocarmos por s unidades à direita, então a que ficava acima continuará acima. Mas sabemos que o esticamento por β e deslocamento por s faz $G_{\xi_1}(\cdot)$ tornar-se $G_{\xi_1,\beta,s}(\cdot)$, e faz $G_{\xi_2}(\cdot)$ tornar-se $G_{\xi_2,\beta,s}(\cdot)$. Portanto $G_{\xi_1,\beta,s}(\cdot)$ fica acima de $G_{\xi_2,\beta,s}(\cdot)$ em todo o eixo x , e isto é suficiente para que, devido a equivalência (2.13), a relação (2.15) seja válida para quaisquer β, s e $\xi_1 < \xi_2$. Fecharemos o presente parágrafo com a interpretação da propriedade (2.15) via o conceito de risco: se os valores dos parâmetros de escala e de locação permanecerem fixos, então o risco cresce com o aumento do valor do parâmetro de forma.

A segunda das propriedades é:

$$\begin{aligned} &\text{Se } \beta_1 \leq \beta_2, \text{ então } G_{\xi,\beta_1,s}^{-1}(p) \leq G_{\xi,\beta_2,s}^{-1}(p) \\ &\text{para todo } p \in [0, 1], \text{ para quaisquer } \xi \text{ e } s. \end{aligned} \quad (2.16)$$

Notamos que a demonstração dessa propriedade baseia-se nas equivalências (2.13) e no papel dos parâmetros β e s na construção da GPD $G_{\xi,\beta,s}(\cdot)$ a partir da GPD padrão $G_{\xi}(\cdot)$ (formulada no Fato 1). A Figura 2.7(a) ilustra a propriedade (2.16) para o caso particular em que $\xi = 1$ e $s = 0$. A interpretação da relação (2.16) via o conceito de risco é: se os valores dos parâmetros de forma e de locação estiverem fixados, então o risco cresce com o aumento do valor do parâmetro de escala.

A terceira propriedade:

$$\begin{aligned} &\text{Se } s_1 \leq s_2, \text{ então } G_{\xi,\beta,s_1}^{-1}(p) \leq G_{\xi,\beta,s_2}^{-1}(p) \\ &\text{para todo } p \in [0, 1], \text{ para quaisquer } \xi \text{ e } \beta. \end{aligned} \quad (2.17)$$

Assim como acontece para a propriedade (2.16), a demonstração da presente propriedade segue facilmente das equivalências (2.13) e do papel dos parâmetros β e s na construção da GPD $G_{\xi,\beta,s}(\cdot)$ a partir da GPD padrão $G_{\xi}(\cdot)$. A Figura 2.7(b) ilustra a propriedade (2.17) para o caso onde $\xi = 0$ e $\beta = 1$. Por fim, interpretaremos a relação (2.17) via o conceito de risco: se os valores dos parâmetros de forma e de escala não são alterados, o aumento do valor do parâmetro de locação implica no aumento do risco.

⁹Aqui abrimos mão do formalismo: o correto seria dizer que $G_{\xi_1}(x) > G_{\xi_2}(x)$ quando $\xi_1 < \xi_2 \forall x > 0$; e que $G_{\xi_1}(x) = G_{\xi_2}(x) \forall x \leq 0$, quaisquer que sejam ξ_1, ξ_2 .

Figura 2.7: A parte (a) mostra a relação entre os gráficos da GPD $G_{0,\beta,0}(\cdot)$ com diversos valores de β . A parte (b) mostra a relação entre os gráficos da GPD $G_{0,1,s}(\cdot)$ com diversos valores de s .

Há algo em comum entre as três propriedades apresentadas até agora. Em todas elas variamos apenas um dos parâmetros da GPD e vimos como essa variação afeta os valores de seus quantis.¹⁰ E quando se varia dois ou três parâmetros de uma só vez? O que acontece? Discutiremos isto agora.

Consideremos o caso $s = 0$, onde ξ e β podem variar. Sejam as GPDs $G_{0,2,0}(\cdot)$ e GPD $G_{1,1,0}(\cdot)$. Afirmamos que os gráficos destas se relacionam como na figura (2.8). De fato, nosso estudo das derivadas destas GPDs nos mostra que $g_{0,2,0}(0) = 1/2$ e $g_{1,1,0}(0) = 1$ (aqui g significa a derivada de G). Isto garante que a função $G_{0,2,0}(\cdot)$ “saí” do ponto $(0,0)$ mais “devagar” que a função $G_{1,1,0}(\cdot)$. Portanto $G_{0,2,0}(\cdot)$ está abaixo de $G_{1,1,0}(\cdot)$ para x positivos próximos de 0. Porém, a medida que x aumenta, a distância entre a GPD $G_{0,2,0}(\cdot)$ e o nível $y = 1$ diminui como $e^{-x/2}$, enquanto que a distância entre GPD $G_{1,1,0}(\cdot)$ e o nível $y = 1$ diminui como $1/(1+x)$. Para valores de x suficientemente grandes ocorre que $e^{-x/2} < 1/(1+x)$. Portanto $G_{0,2,0}(x)$ estará acima de $G_{1,1,0}(x)$ para valores de x maiores que um certo x_0 . É isto que está esquematicamente ilustrado na Figura 2.8. (A figura ainda mostra que há um só ponto de interseção entre $G_{0,2,0}(\cdot)$ e $G_{1,1,0}(\cdot)$; este fato pode se provado rigorosamente, mas não assumiremos esta tarefa.)

O caso discutido detalhadamente na parágrafo acima, assim como os casos (2.15), (2.16) e (2.17) discutidos anteriormente, são casos particulares da propriedade genérica formulada abaixo. A discussão destes casos dá a noção de como esta propriedade seria demonstrada; isto é importante pois não incluímos no nosso texto a demonstração da afirmação abaixo.

Afirmção. Sejam ξ_1, β_1, s_1 e ξ_2, β_2, s_2 quaisquer triplas de valores reais satisfazendo $\beta_1 > 0$ e $\beta_2 > 0$.

(a) Suponha que $\xi_1 < \xi_2$. Então existe $x_0 > 0$ tal que $G_{\xi_1, \beta_1, s_1}^{-1}(p) \leq G_{\xi_2, \beta_2, s_2}^{-1}(p)$ desde

¹⁰Lembre-se que $G^{-1}(p)$ se chama p -quantil da distribuição G .

Figura 2.8: O desenho esquemático da posição relativa dos gráficos das GPDs $G_{0,2,0}(\cdot)$ e $G_{1,1,0}(\cdot)$; a GPD $G_{0,1,0}(\cdot)$ (mais acima) está apresentada como a “referência” de posicionamento: sabemos que $G_{0,2,0}(\cdot)$ e $G_{1,1,0}(\cdot)$ ficam abaixo dela, e sabemos que $G_{0,2,0}(\cdot)$ é a função de referência esticada duas vezes. Temos que $x_0 \approx 2.51285$ e $y_0 = G_{0,2,0}(x_0) = G_{1,1,0}(x_0) \approx 0.71509$.

que ambos valores sejam maiores que x_0 ; o valor de x_0 depende de $\xi_1, \beta_1, s_1, \xi_2, \beta_2, s_2$.

(b) Suponha que $\xi_1 = \xi_2$ e designe ambos por ξ . Suponha que $\beta_1 < \beta_2$. Então existe $x_0 > 0$ tal que $G_{\xi, \beta_1, s_1}^{-1}(p) \leq G_{\xi, \beta_2, s_2}^{-1}(p)$ desde que ambos valores sejam maiores que x_0 ; o valor de x_0 depende de $\xi, \beta_1, \beta_2, s_1, s_2$.

(c) Suponha que $\xi_1 = \xi_2$ e designe ambos por ξ . Suponha que $\beta_1 = \beta_2$ e designe ambos por β . Suponha que $s_1 < s_2$. Então existe $x_0 > 0$ tal que $G_{\xi, \beta, s_1}^{-1}(p) \leq G_{\xi, \beta, s_2}^{-1}(p)$ desde que ambos valores sejam maiores que x_0 ; o valor de x_0 depende de ξ, β, s_1, s_2 .

O item (a) da Afirmação acima diz que entre os três parâmetros ξ, β, s da GPD $G_{\xi, \beta, s}(\cdot)$, o índice de cauda ξ é quem determina a grandeza dos valores de $G_{\xi, \beta, s}^{-1}(p)$ quando p está suficientemente próximo de 1. Usando a equivalência (a) $\Leftrightarrow$ (d) de (2.13), podemos chegar à seguinte

Interpretação em termos de risco: Se as perdas de duas atividades “A” e “B” forem distribuídas conforme as GPDs $G_{\xi_1, \beta_1, s_1}(\cdot)$ e $G_{\xi_2, \beta_2, s_2}(\cdot)$, então sempre existe um limiar x_0 tal que se formos considerar somente as perdas acima de x_0 , a atividade correspondente ao maior valor de ξ é mais ariscada que a outra.

Já usando a equivalência (a) $\Leftrightarrow$ (c) de (2.13), chegamos à seguinte

Interpretação do item (a) da Afirmação em termos da posição relativa das GPDs: Se considerarmos duas GPDs, sempre há um valor x_0 tal que a cauda à direita de x_0 da GPD com o maior valor do parâmetro ξ encontra-se abaixo da cauda à direita de x_0 da GPD com menor valor de ξ .

A última interpretação deu origem ao seguinte jargão: Entre duas GPDs aquela que tem maior valor do parâmetro ξ diz-se possuir cauda *mais pesada*. Ainda mais, a cauda da GPD com $\xi = 0$ chama-se *leve* ou *fina*, a cauda de GPD com $\xi > 0$ chama-se *pesada* ou *grossa*. Os argumentos da presente sub-seção deixam claro que este jargão é a gíria que expressa o papel “dominante” do parâmetro de forma ξ . Mais ainda, a equivalência das duas interpretações dadas acima permite ao nosso leitor traduzir a gíria em termos relacionados ao risco. Por exemplo, a frase: “A cauda das perdas do investimento A é mais leve que a do investimento B” significa “O investimento A é menos arriscado que B, sendo que talvez, trate-se da comparação de riscos de grandes perdas”.

Capítulo 3

Teoria da distribuição assintótica de extremos

Neste capítulo explicaremos o significado do termo “encontrar-se no domínio de uma distribuição de valores extremos”, que surgirá no resultado de Pickands (seção 4.2, cap. 4). Começaremos com a introdução do *problema da distribuição assintótica de extremos*, depois apresentaremos sua solução, onde surgem as funções de distribuição de valores extremos. Após isso especificaremos o significado da afirmação: “uma função de distribuição pertencer ao domínio de atração das distribuições de valores extremos”.

3.1 O problema da distribuição assintótica de extremos

Seja X uma variável aleatória qualquer e seja $F(\cdot)$ sua função de distribuição. Seja $X_1, X_2, \dots$ uma seqüência infinita de variáveis aleatórias independentes entre si, onde cada uma possui a mesma distribuição que X . Considere o seguinte problema:

(P1) encontrar duas seqüências numéricas $\{c_n\}$ e $\{d_n\}$ tais que a seqüência de variáveis aleatórias

$$c_n^{-1} [\max(X_1, \dots, X_n) - d_n], n = 1, 2, \dots \quad (3.1)$$

convirja em distribuição¹ para alguma variável aleatória não degenerada (o significado de "não degenerada" será explicado no comentário 7).

Este é o problema *da distribuição assintótica de extremos*.

O problema (P1) é estudado desde os anos 30 do século passado e, de lá para cá foi totalmente solucionado em uma seqüência de trabalhos matemáticos. Apresentaremos sucintamente esta resposta em (I)-(IV) abaixo. Porém, apesar dos nossos esforços em simplificá-la, ainda assim ela ficou extensa: é de sua natureza. E foi precisamente essa sua característica que nos levou a antecipar a resposta com uma série de exemplos; que expõem as sutilezas do problema (P1) e que preparam o leitor para uma melhor compreensão de sua solução.

Exemplo 1. Considere um dado, com os números de 1 a 6 nas suas faces, mas tal que os pesos das faces são diferentes. Ponhamos o peso $\varepsilon = 10^{-10}$ na face 6 e distribuamos o restante uniformemente pelas outras 5 faces. Vamos lançar este dado sem parar. Seja X_i o resultado do i -ésimo lançamento. Então $X_1, X_2, \dots$ são variáveis aleatórias independentes entre si, e cada uma delas tem a mesma distribuição:

$$\mathbb{P}[X_i = 1] = \dots = \mathbb{P}[X_i = 5] = (1 - \varepsilon)/5, \text{ e } \mathbb{P}[X_i = 6] = \varepsilon \quad (3.2)$$

Abordaremos o problema (P1) para esta seqüência.

Caso (a). Sejam $c_n = 1$ e $d_n = 0$ para todo n . Com esta escolha a seqüência das variáveis aleatórias do problema (P1) torna-se $\max\{X_1, \dots, X_n\}$, $n = 1, 2, \dots$. Vejamos agora como se comporta o limite desta seqüência. Para tanto, observe o seguinte fato: se formos lançar o dado sem parar, então cedo ou tarde conseguiremos a face 6, independentemente de sua pequena probabilidade de ocorrência. Este fato não contradiz a nossa intuição, e, em verdade, pode ser demonstrado rigorosamente (o que não faremos neste trabalho). Em termos matemáticos isso significa que

$$\max\{X_1, X_2, \dots, X_n\} \rightarrow 6 \text{ conforme } n \rightarrow \infty \quad (3.3)$$

que é a resposta do problema (P1) para este caso. Para que não reste dúvidas, lembramos ao leitor que a expressão em (3.3) diz que no presente caso as variáveis aleatórias (3.1) convergem para uma variável aleatória degenerada (que é a constante $c = 6$).

¹É importante recordar o conceito de convergência em distribuição usado no enunciado do problema: Sejam V e $V_1, V_2, \dots$ variáveis aleatórias quaisquer. Diz-se que a seqüência $\{V_n, n \geq 1\}$ converge em distribuição à variável aleatória V se $F_{V_n}(\cdot)$, a função da distribuição da variável aleatória V_n , converge à função da distribuição da variável aleatória V , $F_V(\cdot)$, em cada ponto $x \in \mathbb{R}$ onde $F_V(x)$ é contínua. No nosso caso específico as funções de distribuição limite, que podem surgir na solução do problema da distribuição assintótica de extremos, são todas funções contínuas. Veja também o comentário 6.

Caso (b). Vejamos agora como o limite (3.1) se comporta para a escolha $c_n = 2 + \frac{1}{n}$ e $d_n = 1 + \frac{1}{n}$ para cada $n \in \mathbb{N}$. Para tal fim, notamos que se $x_1, x_2, \dots$ for qualquer realização das variáveis aleatórias $X_1, X_2, \dots$, então um destes x 's é igual a 6 (devido ao fato explicado no Caso (a)). Portanto, $\max\{x_1, \dots, x_n\} = 6$ para todos n 's maiores que um certo número finito (o qual pode variar de uma realização para outra). Assim $c_n^{-1} [\max\{x_1, \dots, x_n\} - d_n] = [6 - 1 - \frac{1}{n}] / (2 + \frac{1}{n})$ para todos n 's suficientemente grandes. Isto garante que as variáveis aleatórias (3.1) convergem para a constante $c = 2.5$ no presente caso.

Caso (c). Tomaremos $c_n = \frac{1}{n}$ e $d_n = 0$ para cada $n \in \mathbb{N}$. Um raciocínio similar àquele usado no Caso (b) mostra facilmente que aqui a seqüência das variáveis aleatórias (3.1) tende a $+\infty$. Em linguagem vulgar pode-se dizer que esta seqüência "converge para $+\infty$ ", mas já que nenhuma variável aleatória pode assumir o valor $+\infty$ (pela própria definição do conceito de "variável aleatória"), seria mais correto dizer que neste caso a seqüência de variáveis aleatórias (3.1) não converge.

Caso (d). Façamos agora $d_n = 0$ para todo n , onde a seqüência $\{c_n\}$ assume a forma: $\frac{1}{2}, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{4}, \frac{1}{8}, -\frac{1}{8}, \dots$. Uma abordagem semelhante à utilizada no Caso (b) mostra facilmente que a seqüência

$$c_n^{-1} [\max\{x_1, \dots, x_n\} - d_n]$$

oscila entre $+n$ e $-n$ para todos os n 's suficientemente grandes. Estas oscilações garantem que a seqüência de variáveis aleatórias (3.1) não converge para as escolhas de c_n e d_n feitas neste caso.

Exemplo 2. Neste exemplo nos concentraremos na seqüência de variáveis aleatórias independentes $X_1, X_2, \dots$, onde cada X_i tem distribuição uniforme no intervalo $[0; 1]$, isto é, cada variável aleatória X_i tem distribuição $F(\cdot)$ onde

$$F(x) = x \text{ se } x \in [0, 1], \quad = 0 \text{ se } x < 0, \quad = 1 \text{ se } x > 1 \quad (3.4)$$

Caso (a). Seja $d_n = 0$ e $c_n = 1$. Para esta escolha temos:

$$\begin{aligned} \mathbb{P} [c_n^{-1} (\max\{X_1, \dots, X_n\} - d_n) \leq x] &= \mathbb{P} [\max\{X_1, \dots, X_n\} \leq x] \\ &= \mathbb{P} [X_1 \leq x] \times \dots \times \mathbb{P} [X_n \leq x] = \{F(x)\}^n \\ &= \begin{cases} x^n, & \text{se } 0 \leq x \leq 1 \\ 1^n, & \text{se } x > 1 \\ 0^n, & \text{se } x < 0 \end{cases} \xrightarrow{n \rightarrow \infty} \begin{cases} 0, & \text{se } x < 1 \\ 1, & \text{se } x \geq 1 \end{cases} \end{aligned}$$

É fácil perceber que neste caso a nossa seqüência de variáveis aleatórias $\max\{X_1, \dots, X_n\}$ $n = 1, 2, \dots$ converge para a distribuição degenerada (acompanhe a figura 3.1):

$$F(x) = \begin{cases} 0, & \text{para } x < 0 \\ 1, & \text{para } x \geq 0 \end{cases} \quad (3.5)$$

Caso (b). Considere $d_n = 1$ e $c_n = \frac{1}{n}$ para todo $n \in \mathbb{N}$. Para esta escolha, temos que (veja a figura 3.1)

$$\begin{aligned} IP \left[c_n^{-1} (\max\{X_1, \dots, X_n\} - d_n) \leq x \right] &= IP \left[\max\{X_1, \dots, X_n\} \leq 1 + \frac{x}{n} \right] \\ &= IP \left[X_1 \leq 1 + \frac{x}{n} \right] \times \dots \times IP \left[X_n \leq 1 + \frac{x}{n} \right] = \left\{ F \left(1 + \frac{x}{n} \right) \right\}^n \\ &= \begin{cases} \left(1 + \frac{x}{n} \right)^n, & \text{se } x \leq 0 \text{ e } n > x \\ 1^n, & \text{se } x \geq 0 \\ 0^n, & \text{se } x \leq 0 \text{ e } n \leq x \end{cases} \xrightarrow{n \rightarrow \infty} \begin{cases} e^x, & \text{se } x \leq 0 \\ 1, & \text{se } x \geq 0 \end{cases} \end{aligned}$$

Na última passagem dos cálculos acima o leitor pode sentir alguma dificuldade. Acontece que para facilitar a visualização dos cálculos, ali usamos o seguinte fato (cuja demonstração pode ser encontrada em qualquer livro-texto de Cálculo):

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n = e^x \text{ para qualquer } x \in \mathbb{R} \quad (3.6)$$

Neste caso(b) a sequência em 1 converge para a distribuição (veja a figura 3.1):

$$F(x) = \begin{cases} e^x, & \text{para } x < 0 \\ 1, & \text{para } x \geq 0 \end{cases} \quad (3.7)$$

Figura 3.1: As figuras acima apresentam, da esquerda para a direita, os gráficos de $\max\{X_1\}$, $\max\{X_1, \dots, X_4\}$ e $\max\{X_1, \dots, X_{10}\}$, onde as variáveis aleatórias são as do exemplo 1, caso (a). Note que para as constantes de normalização $c_n = 1$ e $d_n = 0$, a variável aleatória $\max\{X_1, \dots, X_n\}$ converge para uma variável aleatória degenerada (toda a distribuição se concentra no ponto 1).

Figura 3.2: Figura que corresponde à convergência da variável aleatória $Y_n = c_n^{-1} (\max\{X_1, \dots, X_n\} - d_n)$ do exemplo 1, caso (b). Da esquerda à direita temos a função de distribuição de Y_1 , Y_4 e Y_{50} . A linha tracejada é o gráfico da distribuição limite.

Exemplo 3. Aqui tomaremos a sequência de variáveis aleatórias independentes $X_1, X_2, \dots$ tal que cada X_i tem distribuição exponencial de parâmetro 1, ou seja,

$$IP[X_i \leq x] = F(x) = \begin{cases} 0, & \text{para } x < 0 \\ 1 - \exp\{-x\}, & \text{para } x \geq 0 \end{cases} \quad (3.8)$$

Caso (a). Usando a independência das variáveis aleatórias da sequência $X_1, X_2, \dots$ podemos concluir que

$$\begin{aligned} IP[\max\{X_1, \dots, X_n\} \leq x] &= IP[X_1 \leq x] \times \dots \times IP[X_n \leq x] \\ &= \begin{cases} 0, & x < 0 \\ (1 - \exp\{-x\})^n, & x \geq 0 \end{cases} \end{aligned} \quad (3.9)$$

Note agora que $1 - \exp\{-x\}$ é menor que 1 para todo $x \geq 0$. Este fato em conjunto com a igualdade (3.9) garante que $IP[\max\{X_1, \dots, X_n\} \leq x] \rightarrow 0$ conforme $n \rightarrow \infty$ para cada $x \in \mathbb{R}$. Já que não existe variável aleatória que assume todos os valores da reta com probabilidade 0 (note que nem uma variável aleatória é degenerada assim), então podemos concluir que

$$\begin{aligned} &\text{se } c_n = 1 \text{ e } d_n = 0 \text{ para cada } n, \text{ então a sequência } c_n^{-1} \\ &[\max\{X_1, \dots, X_n\} - d_n] \text{ não converge} \end{aligned} \quad (3.10)$$

Note que a conclusão (3.10) só é válida para uma escolha específica de sequências c_n e d_n . Que tal tentarmos outras escolhas? Vejamos o caso (b). Caso (b). Acompanhem o trecho abaixo:

$$\begin{aligned} IP[\max\{X_1, \dots, X_n\} - \ln(n) \leq x] &= IP[\max\{X_1, \dots, X_n\} \leq x + \ln(n)] \\ &= (1 - \exp\{-x - \ln(n)\})^n \text{ (quando } n \text{ é suficientemente grande)} \\ &= \left(1 - \frac{e^{-x}}{n}\right)^n = e^{-e^{-x}}, \text{ para todo } x \in \mathbb{R} \end{aligned} \quad (3.11)$$

(acima, na passagem da primeira para segunda linha, usamos a independência entre as variáveis da sequência e o fato de que para qualquer que seja x , a expressão $x + \ln(n)$ é positiva para n suficientemente grande, e portanto, para tal n , é válido que $\mathbb{P}[X_i \leq x + \ln(n)] = 1 - \exp\{-x - \ln(n)\}^2$; na última passagem usamos o fato de que $(1 - \frac{a}{n})^n \rightarrow e^{-a}$ para todo $a \in \mathbb{R}$ conforme $n \rightarrow \infty$). Segue do resultado (3.11) que

$$\text{se } c_n = 1 \text{ e } d_n = \ln(n) \text{ para cada } n, \text{ então } c_n^{-1} [\max\{X_1, \dots, X_n\} - d_n] \text{ converge para a distribuição } e^{-e^{-x}}, x \in \mathbb{R} \quad (3.12)$$

Caso (c). Com um raciocínio similar ao usado em (3.11) pode-se deduzir resultados tais como (3.12) para diversas escolhas de seqüências c_n e d_n . Por exemplo, vale que

$$\text{se } c_n = 3 \text{ e } d_n = \ln(n) + 2 \text{ para cada } n, \text{ então } c_n^{-1} [\max\{X_1, \dots, X_n\} - d_n] \text{ converge para a distribuição } \exp\{-e^{-(\frac{x+2}{3})}\}, x \in \mathbb{R} \quad (3.13)$$

Exemplo 4. Neste exemplo a seqüência de variáveis aleatórias independentes $X_1, X_2, \dots$ é tal que cada um dos X_i tem distribuição *lognormal* de parâmetros $\sigma = 1$ e $\mu = 0$, quer dizer,

$$\mathbb{P}[X_i \leq x] = F(x) = \begin{cases} \int_0^x \frac{1}{u\sqrt{2\pi}} e^{-\frac{(\ln u)^2}{2}} du, & \text{para } x > 0 \\ 0, & \text{para } x \leq 0 \end{cases} \quad (3.14)$$

Ilustramos na figura 3.3 o comportamento da variável aleatória $Y_n = c_n^{-1} (\max\{X_1, \dots, X_n\} - d_n)$ para $c_n^{-1} = 1$ e $d_n = 0$.

As realizações da variável aleatória $Y_n = c_n^{-1} (\max\{X_1, \dots, X_n\} - d_n)$ foram construídas da seguinte forma³: Primeiro tomamos o máximo de $x_1, x_2, \dots, x_n$, que correspondem a uma realização das variáveis aleatórias $X_1, X_2, \dots, X_n$. Uma vez encontrado o máximo, y_m , subtraímos d_n deste máximo e depois multiplicamos o resultado por c_n^{-1} para obtermos uma realização de Y_n , ou seja, o valor $y = c_n^{-1}(y_m - d_n)$.

Comentário 6. É importante ressaltar que na Teoria de Probabilidades existem outros tipos de convergência, e todos eles possuem suas aplicações: uma seqüência de variáveis aleatórias pode convergir a uma variável aleatória em probabilidade, pode convergir quase certamente, e também de outras maneiras; em particular pode convergir em distribuição – um conceito que foi explicado anteriormente e que será usado com exclusividade, entre os demais tipos de convergência, na nossa exposição a seguir.

Comentário 7. Por fim, apresentaremos o conceito de distribuição degenerada. Lembramos ao leitor que uma variável aleatória é classificada como *degenerada* se ela é uma constante,

²Veja na equação (3.14) que o expoente da distribuição exponencial deve ser não-positivo.

³Veja também o Método de Máximos por blocos apresentado na seção 3.3

Figura 3.3: Figuras que ilustram a divergência – dispersão e deslocamento – da variável aleatória $Y_n = c_n^{-1}(\max\{X_1, \dots, X_n\} - d_n)$ quando $c_n^{-1} = 1$ e $d_n = 0$. Cada X_i é distribuída de acordo com a distribuição lognormal. De cima para baixo temos os histogramas de $\max\{X_1\}$, $\max\{X_1, \dots, X_{50}\}$ e $\max\{X_1, \dots, X_{500}\}$.

isto é, se ela assume um único valor com a probabilidade total igual a 1 – daí o termo “degenerada”. Uma função de distribuição é *degenerada* se ela corresponde a uma variável aleatória degenerada. Todas as funções de distribuição degeneradas têm a mesma cara: elas são funções-escada que possuem um único pulo de tamanho um. De fato, se uma variável aleatória degenerada é igual a uma constante c , então sua função de distribuição é 0 à esquerda de c e é 1 em c e à direita dele. Geralmente as funções escada são inúteis quando desempenham o papel de função-limite em resultados assintóticos como aquele que queremos deduzir para distribuições de valores extremos.

3.2 A solução do problema da distribuição assintótica de extremos

A resposta ao problema da distribuição assintótica de extremos é o conteúdo dos itens (I)-(IV) apresentados abaixo. A divisão da resposta em itens não é um padrão, mas foi feita de

Figura 3.4: Distribuições de Fréchet, Weibull e Gumbel.

maneira que, ao nosso ver, facilita a compreensão.

- (I) Nem toda distribuição $F(\cdot)$ admite seqüências numéricas $\{c_n\}$ e $\{d_n\}$ que façam a seqüência de variáveis aleatórias (3.1) convergir para uma distribuição que seja não degenerada. Ou ainda, existem distribuições $F(\cdot)$ onde para quaisquer que sejam as seqüências $\{c_n\}$ e $\{d_n\}$ a variável aleatória (3.1) não converge (incluindo a possibilidade de divergir para $+\infty$ ou $-\infty$), ou, se converge, seu limite corresponde a uma variável aleatória degenerada.

Nota à parte: No exemplo 1 foi dado um exemplo destas distribuições. Na verdade ele não é o único exemplo. Assim será qualquer função $F(\cdot)$ que não satisfaça as condições a serem apresentadas no item (IV). Com base nestas condições você perceberá que é possível criar uma infinidade destas funções de distribuição atípicas.

- (II) Se para $F(\cdot)$ existir uma escolha de seqüências numéricas c_n e d_n onde a seqüência de variáveis aleatórias (3.1) converge para uma distribuição não degenerada, estas seqüências numéricas podem ser escolhidas de maneira tal que (3.1) tem como limite
- uma função da família chamada *distribuições de Fréchet*;
 - ou uma função da família chamada *distribuições de Weibull*;
 - ou uma função chamada *distribuição de Gumbel*.

Eis as formas destas funções e suas famílias:

a família de Fréchet constitui-se das distribuições $\Phi_\alpha(\cdot)$ com $\alpha > 0$, onde:

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

a família de Weibull constitui-se das distribuições $\Psi_\alpha(\cdot)$ com $\alpha > 0$, onde:

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & \text{se } x \leq 0 \\ 1, & \text{se } x > 0 \end{cases}$$

a distribuição de Gumbel é definida por:

$$\Lambda(x) = \exp\{-e^{-x}\}, \text{ para todo } x \in \mathbb{R}$$

Nota a parte: Lembre-se que nos exemplos 2 e 3, caso (a), tivemos o azar de escolher $\{c_n\}$ e $\{d_n\}$ de maneira que as variáveis aleatórias (3.1) não convergiam, ou que convergiam para uma variável aleatória degenerada. Note também que existiam outras escolhas (apresentadas nos casos (b)), para as quais havia a convergência das variáveis aleatórias (3.1). Observe que, para estas últimas escolhas, a distribuição do limite era a distribuição de Weibull com $\alpha = 1$ (no caso do Exemplo 2), e a de Gumbel (no caso do Exemplo 3). Estes fatos ilustram o conteúdo da afirmação (II).

Daqui adiante usaremos o termo *família de distribuições de valores extremos* para nos referir à união que contém a família de Fréchet, a família de Weibull e a distribuição de Gumbel. Qualquer membro desta união será chamado de *distribuição de valores extremos*, EVD (do Inglês *extreme value distribution*).

(III) Suponhamos que uma função de distribuição $F(\cdot)$ e duas seqüências numéricas $\{c_n\}$ e $\{d_n\}$ são tais que a seqüência das variáveis aleatórias (3.1) converge para uma EVD; simbolizada por $H(\cdot)$ (que pode ser, devido a nossa suposição, uma distribuição de Gumbel, uma distribuição da família Fréchet, ou uma distribuição da família Weibull). Sejam agora $\{c'_n\}$ e $\{d'_n\}$ duas outras seqüências numéricas quaisquer. Nessas condições pode ocorrer uma e somente uma das duas seguintes alternativas:

ou $(c'_n)^{-1} [\max\{X_1, \dots, X_n\} - d'_n]$ converge para uma variável aleatória cuja distribuição é $H(ax + b)$, $x \in \mathbb{R}$, para $a \in \mathbb{R}$, $a \neq 0$ e $b \in \mathbb{R}$ (cujos valores dependem das seqüências $\{c'_n\}$ e $\{d'_n\}$);

ou $(c'_n)^{-1} [\max\{X_1, \dots, X_n\} - d'_n]$ converge para uma distribuição degenerada ou nem sequer converge.

Para que este resultado importante seja adequadamente compreendido, ressaltamos a relação entre a função $H(ax + b)$ e a função $H(x)$, $x \in \mathbb{R}$: a primeira é obtida da segunda por re-escala e translação ao longo do eixo das abscissas. Dessa forma o resultado acima formulado pode ser interpretado da seguinte maneira:

Dada uma distribuição $F(\cdot)$, o limite (3.1), quando existe e não é uma variável aleatória degenerada, tem a forma $H(ax + b)$, $x \in \mathbb{R}$, onde $H(\cdot)$ é uma EVD que é específica para esta $F(\cdot)$ dada; ou, em outras palavras, é impossível que para dois pares de seqüências numéricas $\{c_n\}$, $\{d_n\}$ e $\{c'_n\}$, $\{d'_n\}$; $c_n^{-1} [\max\{X_1, \dots, X_n\} - d_n]$ convirja à distribuição não degenerada $H(\cdot)$, e $(c'_n)^{-1} [\max\{X_1, \dots, X_n\} - d'_n]$ convirja à distribuição não degenerada $K(\cdot)$, sem que $H(\cdot)$ e $K(\cdot)$ estejam relacionadas via⁴ $K(ax + b) = H(x) \quad \forall x \in \mathbb{R}$.

A propriedade discutida acima diz que cada função de distribuição $F(\cdot)$ é "fiel" a somente uma EVD. Podemos também analisar esta relação sob outra perspectiva: cada EVD possui um conjunto de funções de distribuição relacionado a ela. Este conjunto é o *domínio de atração* da EVD. Em termos precisos. Diz-se que uma função de distribuição $F(\cdot)$ está no *domínio de atração* de uma EVD $H(\cdot)$ se existem seqüências $\{c_n\}$ e $\{d_n\}$ tais que as variáveis aleatórias (3.1) convergem para $H(\cdot)$. O conceito "domínio de atração de uma EVD" será utilizado na formulação do item (IV) e, o que é mais importante, na formulação do teorema de Pickands (veja a seção 4.2).

- (IV) Uma função de distribuição $F(\cdot)$ está no domínio de atração da EVD de Fréchet com parâmetro α se, e somente se, $F(\cdot)$ tem cauda direita infinita⁵ e para valores suficientemente grandes de x ocorre que⁶ $\bar{F}(x) = x^{-\alpha}L(x)$, onde $L(\cdot)$ é uma função de variação lenta (veja o comentário 8 no final desta seção). Caso $F(\cdot)$ satisfaça esta condição, a escolha⁷ das seqüências $d_n = 0$ e $c_n = F^{\leftarrow}(1 - n^{-1})$ garante a convergência das variáveis aleatórias em (3.1) à distribuição de Fréchet com parâmetro α .

Uma função de distribuição $F(\cdot)$ está no domínio de atração da EVD de Weibull de parâmetro α se e somente se $F(\cdot)$ tem cauda direita finita⁸ e para valores suficientemente grandes de x ocorre que $\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$, onde x_F é a abscissa do fim da cauda direita, quer dizer, o menor valor onde $F(\cdot)$ assume 1, onde $L(\cdot)$ é uma função de variação lenta (veja o comentário 8 adiante). Se $F(\cdot)$ satisfizer esta condição, então

⁴A propriedade afirmada neste ítem foi observada no exemplo 3. Lembre-se que ali a distribuição limite obtida no caso (b) e a distribuição obtida no caso (c) possuem uma relação do tipo $K(ax+b) = H(x) \quad \forall x \in \mathbb{R}$.

⁵Vá à seção 2.3 do cap. 2 para rever o conceito de cauda infinita.

⁶Usaremos o símbolo $\bar{F}(\cdot)$, cuja definição é: $\bar{F}(x) = 1 - F(x)$, $x \in \mathbb{R}$.

⁷Usaremos o símbolo $F^{\leftarrow}(\cdot)$ para a função inversa de $F(\cdot)$ construída de maneira que $F(\cdot)$ possa ter discontinuidades; explicitamente: $F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}$, $t \in (0; 1)$.

⁸Vá à seção 2.3 do cap. 2 para rever o conceito de cauda finita.

a escolha das seqüências $d_n = x_F$ e $c_n = x_F - F^{*-1}(1 - n^{-1})$ garante a convergência das variáveis aleatórias em (3.1) à distribuição de Weibull com parâmetro α .

Uma função de distribuição $F(\cdot)$ está no domínio de atração da EVD de Gumbel se e somente se⁹ $x_F \leq \infty$ e existe $z < x_F$ tal que $F(\cdot)$ admita a seguinte representação:¹⁰

$$\bar{F}(x) = h(x) \exp \left\{ - \int_z^x \frac{g(t)}{a(t)} dt \right\}, \text{ para } x \in (z, x_F) \quad (3.15)$$

onde as funções $h(\cdot)$ e $g(\cdot)$ são tais que $g(x)$ converge a 1 e $h(x)$ converge a uma constante positiva conforme x se aproxima de x_F , a função $a(\cdot)$ é diferenciável e sua derivada $a'(\cdot)$ satisfaz $\lim_{x \uparrow x_F} a'(x) = 0$. Mais ainda, caso $F(\cdot)$ satisfaça esta condição, então a escolha $d_n = F^{*-1}(1 - n^{-1})$ e $c_n = a(d_n)$ garante a convergência das variáveis aleatórias (3.1) à distribuição de Gumbel.

O conteúdo do item (IV) supreende em dois aspectos. Em primeiro lugar, é de se suspeitar que apenas a cauda de $F(\cdot)$ determine o domínio de atração que contém $F(\cdot)$ (veja o comentário 9 abaixo). Em segundo lugar, é inacreditável a existência de fórmulas genéricas para as seqüências numéricas $\{c_n\}$ e $\{d_n\}$ que garantem a convergência das variáveis aleatórias (3.1).

Comentário 8. Começaremos com a observação de que caso uma função de distribuição $F(\cdot)$ satisfaça¹¹ $\bar{F}(x) \equiv 1 - F(x) = x^{-\alpha}$ para $\alpha > 0$ e x suficientemente grande, então pode-se provar que esta função de distribuição está no domínio de atração de uma EVD de Fréchet. Com esta mesma demonstração é possível provar a mesma propriedade para qualquer função de distribuição que satisfaça $\bar{F}(x) \equiv 1 - F(x) = Cx^{-\alpha}$, para uma constante C positiva. Ainda mais, ao analisar esta prova, podemos descobrir que o fator ao lado de $x^{-\alpha}$ não precisa ser obrigatoriamente uma constante, pode ser uma função cujo comportamento assintótico – quer dizer, o comportamento para todos os valores de x suficientemente grandes – é parecido com uma constante. Um exemplo de tal função é

$$\frac{x^2}{1 + x^2} \quad (3.16)$$

Esta descoberta nos motiva a procurar por condições mais genéricas sobre o fator pelo qual podemos multiplicar $x^{-\alpha}$ sem estragar sua atração pela EVD de Fréchet. Acontece que se

⁹A desigualdade $x_F \leq \infty$ é a forma matemática de dizer que a cauda direita de $F(\cdot)$ pode ser ou finita ($x_F < \infty$), ou infinita ($x_F = \infty$).

¹⁰A representação (3.15) não é única, pois as funções $h(\cdot)$, $g(\cdot)$ e $a(\cdot)$ não são especificadas de forma muito rígida. Isto faz com que possamos, por exemplo, modificar a função $g(\cdot)$ e compensar esta modificação com a mudança de $h(\cdot)$ de sorte que a expressão (3.15) preserve todos os seus valores. A compreensão detalhada da fórmula (3.15) é desnecessária.

¹¹Aqui estamos supondo que $x \in [0, +\infty)$, e que, portanto, a cauda é infinita

formos fazer isto, deduziremos que este fator pode ser qualquer função $L(x)$, $x > 0$, que satisfaça a seguinte condição

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1 \text{ para todo } t > 1 \quad (3.17)$$

Qualquer função $L(\cdot)$ que atenda esta condição é chamada *função de variação lenta*¹². Este nome é justificado pela seguinte interpretação da propriedade (3.17): Fixe uma seqüência crescente de valores $x_1, x_2, \dots$, e seja esta tal que seus valores são equiespaçados (o que não é obrigatório para a validade da propriedade (3.17), mas que ajuda na compreensão); então, conforme o índice i cresce, o par de valores x_i e tx_i desloca-se à direita, e a distância entre os pontos do par cresce (pois esta distância é $tx_i - x_i = x_i(t - 1)$); uma função $L(\cdot)$ será de variação lenta caso a razão entre os seus valores nos pontos de cada par convirja para 1. É fácil ver que a função em (3.16) é uma função de variação lenta. Porém, para que uma função goze desta propriedade, não é necessário que ela seja assintoticamente constante. Por exemplo, a função

$$\log(1 + x) \quad (3.18)$$

tem variação lenta, apesar de não ser assintoticamente constante, uma vez que não há constante C tal que

$$\lim_{x \rightarrow \infty} \frac{\log(1 + x)}{C} = 1$$

ou seja, a propriedade (3.17) é satisfeita por esta função.

Comentário 9. Com respeito a primeira “surpresa”, volte a analisar os Exemplos 1, 2 e 3. Estes indicam que de fato o que determina a distribuição do limite de máximos é o comportamento assintótico da cauda de $F(\cdot)$, isto é, o comportamento de $F(x)$ conforme x vai ao $+\infty$.

3.3 Da relação entre um problema financeiro e o problema da distribuição assintótica de extremos

Agora que toda a teoria sobre a distribuição assintótica de extremos foi apresentada, cabe uma pergunta: Em que situações práticas surge o problema mencionado em (P1)?

Imagine uma agência de um banco brasileiro lolizada no Japão (Os principais fregueses desta agência seriam os brasileiros que trabalham no Japão. Tais fregueses depositam parte de seus salários mensalmente numa poupança desta agência, e resgatam toda a poupança

¹²Na verdade esta é a variação lenta no infinito. Com o objetivo de facilitar a discussão sobre o assunto adotaremos esta nomenclatura

acumulada ao voltar para o Brasil). Considere o fluxo diário desta agência como sendo composto depósitos e saques desses fregueses especiais. Imagine que no caso do valor do fluxo ao final do dia ser positivo, a agência destine este valor ao empréstimo que oferece às empresas nipônicas (o que é natural, pois empréstimos têm altos juros), já no caso do valor acima ser negativo, ele deve ser compensado com parte de uma reserva mantida na agência. Admita que o banco adote a seguinte estratégia em relação a esta reserva: caso ela tenha sido usada, total ou parcialmente, deve ser restituída no dia seguinte via transferência da agência central do banco. Com esta estratégia, em cada manhã a agência tem o mesmo volume na reserva; designemos este volume por M . Finalmente, considere o caso em que o valor do fluxo de caixa no final do dia é negativo e maior em módulo que o volume da reserva. Nesta situação a agência toma o dinheiro faltante no mercado financeiro nipônico. Note uma particularidade em relacionada à manutenção da reserva: a transferência de dinheiro do Brasil para a agência não tem custo, pois é uma transação entre agências do mesmo banco, já o dinheiro tomado no mercado financeiro nipônico tem seu custo, pois é um empréstimo concedido por outro banco, que cobra certa taxa de juros. Esta nota explica a importância da seguinte pergunta em relação ao gerenciamento da reserva da agência:

- (P2) Fixado um horizonte de k dias (que corresponde, digamos, a um ano) qual é a probabilidade de que durante os k dias futuros nenhum saldo dos fluxos diários fique abaixo de um dado valor $-M$?

A situação acima descrita pode ser apresentada de outra forma, de modo a destacar e formalizar as características que são importantes para a compreensão da relação do problema (P2) com o problema (P1)

- (A) Há um fluxo de caixa (de uma instituição, como, digamos, uma agência de um banco), cujos saldos diários podem ser considerados variáveis aleatórias independentes e identicamente distribuídas. Designaremos por $F(\cdot)$ a função de distribuição do saldo diário de tal fluxo com sinal negativo¹³;
- (B) O mercado conhece os fatos descritos em (A) (isto é, que os saldos diários são independentes e identicamente distribuídos) mas não conhece a forma da função $F(\cdot)$;
- (C) Possui-se os valores dos saldos dos fluxos diários observados no passado, ou seja, assumiremos que foram monitorados n dias consecutivos. A sequência $x_1, \dots, x_n$ representa os saldos diários com sinal negativo observados nestes dias;

¹³Pela natureza da problema P2, o que nos interessa são os valores negativos dos saldos diários, mas a técnica a ser utilizada para a resolução deste problema foi desenvolvida para analisar valores positivos. Para adequar o problema à técnica, consideraremos o saldo diário com sinal negativo.

- (D) Fixa-se um horizonte futuro de k dias e um limiar positivo M com o objetivo de estimar a probabilidade de que nenhum saldo dos fluxos diários nos próximos k dias fique abaixo de $-M$.

A descrição ($\mathcal{A}$ -D) deixa claro que a formalização matemática do problema (P2) é:

- (P3) para um dado número natural k e o número real positivo M estimar

$$\mathbb{P} [\max\{X^{(1)}, \dots, X^{(k)}\} \leq M] \quad (3.19)$$

onde $X^{(1)}, X^{(2)}, \dots, X^{(k)}$ são variáveis aleatórias independentes, cada uma com distribuição $F(\cdot)$ desconhecida. A única informação que temos sobre $F(\cdot)$ é uma amostra $x_1, \dots, x_n$ de n observações originadas desta distribuição ¹⁴.

Para que os resultados (I)-(IV) sejam aplicáveis ao problema (P3), é preciso que a função de distribuição $F(\cdot)$ pertença ao domínio de atração de uma das EVDs; este é o primeiro passo. Se isto ocorre ou não pode ser verificado por métodos estatísticos aplicados à amostra $x_1, \dots, x_n$. A descrição destes métodos é feita em [6] e em [19], onde também há uma indicação sobre o tipo de EVD em cujo domínio de atração encontra-se $F(\cdot)$ ¹⁵. Nesse trabalho abdicaremos da aplicação destas análises estatísticas por falta de espaço e por não termos atribuído valores numéricos à amostra $x_1, \dots, x_n$. Assumiremos então, sem verificação, que a função $F(\cdot)$ do problema (P3) está no domínio de atração de uma das distribuições de Fréchet $\Phi_\alpha(\cdot)$. Note que o valor exato do parâmetro α é desconhecido, quer dizer, admitiremos que nesta etapa da solução não sabemos precisar exatamente a função $\Phi_\alpha(\cdot)$, mas apenas que ela é da família de Fréchet.

O pressuposto de que $F(\cdot)$ pertence ao domínio de atração da EVD $\Phi_\alpha(\cdot)$ implica, devido ao conteúdo de (I)-(IV), que existem duas seqüências numéricas c_i e d_i tais que¹⁶

$$\begin{aligned} &\text{a distribuição de } c_i^{-1} [\max\{X^{(1)}, \dots, X^{(i)}\} - d_i] \\ &\text{converge a } \Phi_\alpha(\cdot) \text{ quando } i \rightarrow \infty \end{aligned} \quad (3.20)$$

onde $X^{(1)}, X^{(2)}, \dots$ são variáveis aleatórias independentes e cada uma distribuída conforme $F(\cdot)$. Assumindo que o valor de k do problema (3) seja suficientemente grande, podemos deduzir da relação assintótica (3.20) que

$$\begin{aligned} &c_k^{-1} [\max\{X^{(1)}, \dots, X^{(k)}\} - d_k] \text{ tem aproximadamente} \\ &\text{distribuição } \Phi_\alpha(\cdot) \end{aligned} \quad (3.21)$$

¹⁴Já que em textos matemáticos a notação x_i é usualmente entendida como observação de X_i , preferimos a notação $X^{(1)}, X^{(2)}, \dots, X^{(k)}$ à notação $X_1, \dots, X_k$; isso porque a amostra $x_1, \dots, x_n$ não se refere às variáveis aleatórias $X^{(1)}, X^{(2)}, \dots, X^{(k)}$. Estas últimas estão relegadas às observações futuras.

¹⁵Novamente; dada a amostra $x_1, \dots, x_n$, é possível determinar, via análises estatísticas, se $F(\cdot)$ pertence ao domínio de atração de uma Fréchet, de uma Weibull ou de uma Gumbel. É certo que a precisão dessa resposta depende de várias questões como, por exemplo, o tamanho da amostra.

¹⁶Aqui usaremos o índice i porque o índice n foi reservado ao tamanho da amostra.

ou, equivalentemente, que¹⁷

$$\begin{aligned} \max\{X^{(1)}, \dots, X^{(k)}\} \text{ tem aproximadamente} \\ \text{distribuição } \Phi_\alpha(c_k^{-1}(x - d_k)), x \in \mathbb{R} \end{aligned} \quad (3.22)$$

Recorde que possuímos a amostra $x_1, \dots, x_n$, que corresponde às realizações de $F(\cdot)$. Vamos tomar esta amostra e dividi-la em blocos com k observações em cada bloco. Procuraremos o valor máximo em cada bloco. Este procedimento garante que os valores máximos obtidos são realizações independentes da variável aleatória $\max\{X^{(1)}, \dots, X^{(k)}\}$. Portanto, devido à relação (3.22), o conjunto destes valores forma uma amostra da distribuição $\Phi_\alpha(c_k^{-1}(x - d_k)), x \in \mathbb{R}$. Este fato permite empregar procedimentos de Estatística Paramétrica para deduzir, com base na amostra, estimativas para os valores de c_k , d_k e α desta distribuição. No nosso trabalho não abordaremos estes procedimentos. O leitor pode encontrá-los em [6] ou [19]. Estes procedimentos, juntamente com a construção da amostra de máximos por blocos, formam o *Método de Máximos por Blocos* ou *Método de Máximos Anuais*.

Segue-se então de (3.22) que

$$\begin{aligned} \mathbb{P}[\max\{X^{(1)}, \dots, X^{(k)}\} \leq M] &\approx \Phi_{\hat{\alpha}}(\hat{c}_k^{-1}(M - \hat{d}_k)) \\ \text{onde } \hat{c}_k, \hat{d}_k \text{ e } \hat{\alpha} &\text{ são as estimativas de } c_k, d_k \text{ e } \alpha, \\ &\text{obtidas da amostra } x_1, \dots, x_n \text{ pelo} \\ &\text{Método de Máximos por Blocos} \end{aligned} \quad (3.23)$$

Esta é a solução do problema (P3) obtida com auxílio dos resultados (I)-(IV).¹⁸

Como pode-se ver da apresentação acima, a contribuição dos resultados (I)-(IV) na solução do problema (3.19) materializou-se numa família a três parâmetros da distribuição de máximos. Isto significa que o nosso problema recai na inferência dos valores dos três parâmetros c_k , d_k e α . Quem conhece a Teoria Estatística sabe avaliar o quanto essa contribuição é valiosa. Para quem não conhece a Teoria, apresentamos abaixo o que seria a solução “direta”, isto é, a solução sem o auxílio teorias específicas. Na solução direta usamos a amostra $x_1, \dots, x_n$ para construir a função de distribuição amostral $\hat{F}(\cdot)$. Depois, assumindo que ela seja uma boa aproximação para $F(\cdot)$ e usando a identidade $\mathbb{P}[\max\{X^{(1)}, \dots, X^{(k)}\} \leq x] = \mathbb{P}[X^{(1)} \leq x] \times \dots \times \mathbb{P}[X^{(k)} \leq x]$, obtivemos o resultado:

$$\mathbb{P}[\max\{X_1, \dots, X_k\} \leq M] \approx (\hat{F}(M))^k \quad (3.24)$$

¹⁷Eis a conta que mostra que (3.22) deriva de (3.21): “ $F(x), x \in \mathbb{R}$, é a função de distribuição de $a(X - b)$ ”
 $\Leftrightarrow \mathbb{P}[a(X - b) \leq x] = F(x), \forall x \Leftrightarrow \mathbb{P}[X \leq x/a + b] = F(x), \forall x \Leftrightarrow \mathbb{P}[X \leq y] = F(a(y - b)), \forall y$.

¹⁸A EVD de Fréchet Φ_α apareceu na solução (3.23) devido a nosso pressuposto sobre $F(\cdot)$. É claro que dependendo do caso, esta EVD pode ser de Weibull ou de Gumbel.

Neste caso deve-se esperar uma significativa discordância entre os valores de $\hat{F}(M)$ e de $F(M)$ quando M é muito grande. Esta discordância aumenta quando ambos os valores são elevados à potência k . Conseqüentemente, espera-se que a aproximação em " $\approx$ " em (3.24) seja muito ruim, a desvantagem da solução "direta" do problema (3.19).

Capítulo 4

Resultado de Pickands

Já foi assinalado diversas vezes que o principal resultado desta dissertação, apresentado no cap. 6, tem como base o teorema de Pickands, que apareceu pela primeira vez em [18], mas que também pode ser encontrado facilmente em publicações mais recentes, pois ele é citado em praticamente todos livros que falam da Teoria de Valores Extremos e suas aplicações; por exemplo, em [6] ele aparece como o item (b) do Teorema 3.4.13 na página 165.

Este capítulo se dedica à apresentação do resultado atribuído à Pickands.

4.1 Preliminares

A presente seção serve de apoio aos argumentos das seções seguintes; aqui lembraremos duas notações e introduziremos um novo conceito.

As notações a serem lembradas são: $G_{\xi,\beta,u}(\cdot)$ que denota a função GPD cuja definição está na seção 2.2, e x_F que denota o “fim” da cauda direita da distribuição $F(\cdot)$ quando esta possui cauda finita, e que é igual a $+\infty$ quando $F(\cdot)$ possui cauda infinita. Notamos que x_F foi introduzido e explicado na seção 2.3.

Passaremos agora à introdução de um novo conceito necessário à formação do Resultado de Pickands. Suponha que é dada uma função de distribuição $F(\cdot)$ e designe por X a variável aleatória que possui esta distribuição. Suponha também que é dado um valor u , com $u < x_F$. Neste caso o evento $\{X > u\}$ ocorre com uma probabilidade não nula, e por isso é legítimo perguntar sobre a distribuição da variável aleatória $X - u$ dado que ocorreu

o evento $\{X > u\}$. Designaremos esta função de distribuição¹ por $F_u(\cdot)$, quer dizer,

$$F_u(x) = \mathbb{P}[X - u \leq x \mid X > u], \quad x \in \mathbb{R}, \text{ sendo que } u < x_F \quad (4.1)$$

A fórmula (4.1) é a definição formal da função $F_u(\cdot)$. Sua construção gráfica e intuitiva está ilustrada na figura 4.1, onde ela é essencial para a compreensão dos argumentos. Acrescentaremos agora que $F_u(\cdot)$ é chamada de *função de distribuição de excessos* da distribuição $F(\cdot)$ *acima de limiar* u ; e que o nome é motivado pelo fato de que quando X ultrapassa u então a diferença $X - u$ é usualmente chamada de *excesso* da variável aleatória X *acima do limiar* u . Acrescentaremos ainda, para completar a nomenclatura, que o valor da variável aleatória X que excede u chama-se *excedente de X acima de u* .

Comentário 10. De acordo com a definição (4.1) e a fórmula de probabilidade condicional deduz-se que

$$F_u(x) = \frac{\mathbb{P}[X - u \leq x, X > u]}{\mathbb{P}[X > u]} = \begin{cases} 0, & \text{se } x < 0 \\ \frac{\mathbb{P}[u < X \leq u+x]}{\mathbb{P}[X > u]}, & \text{se } x \geq 0 \end{cases} = \begin{cases} 0, & \text{para } x < 0 \\ \frac{F(x+u) - F(u)}{1 - F(u)}, & \text{para } x \geq 0 \end{cases} \quad (4.2)$$

A primeira e a última partes de (4.2) exibem a transformação que leva $F(\cdot)$ em $F_u(\cdot)$. Esta transformação pode ser visualizada na Figura 4.1 abaixo, em três passos, que descrevemos verbalmente em (a)-(c):

- (a) transportar a cauda de $F(\cdot)$ à direita de u de maneira o ponto $(u, F(u))$ coincida com $(0, 0)$ (o acréscimo “ $+u$ ” em $F(x+u)$ na fórmula (4.2) é o que transporta $F(\cdot)$ à esquerda por u unidades, enquanto que o termo $-F(u)$ é o que transporta $F(\cdot)$ para baixo por $F(u)$ unidades);
- (b) esticar a cauda transportada por fator $\frac{1}{1-F(u)}$ ao longo do eixo vertical (o fator $\frac{1}{1-F(u)}$ na fórmula (4.2) faz esta transformação; note que este fator é > 1 , razão pela qual usamos a palavra “esticar”);
- (c) acrescentar semi-eixo $(-\infty; 0]$ à cauda transportada e esticada (isto é necessário para que a função seja definida para todo x e possa ser chamada de função de distribuição).

4.2 Resultado de Pickands

Teorema (Pickands [18]). Se uma função de distribuição $F(\cdot)$ está no domínio de atração de

¹O correto seria dizer “função de distribuição condicional” assinalando com isto que trata-se da distribuição da variável aleatória $X - u$ condicionada ao evento $\{X > u\}$.

Figura 4.1: A transformação que leva $F(\cdot)$ em $F_u(\cdot)$.

uma das EVDs, e se definimos ξ de acordo com a seguinte regra:

$$\xi = \begin{cases} \alpha^{-1}, & \text{se } F(\cdot) \text{ estiver no domínio de atração de Fréchet } \Phi_\alpha(\cdot); \\ -\alpha^{-1}, & \text{se } F(\cdot) \text{ estiver no domínio de atração de Weibull } \Psi_\alpha(\cdot); \\ 0, & \text{se } F(\cdot) \text{ estiver no domínio de atração de Gumbel } \Lambda(\cdot), \end{cases} \quad (4.3)$$

então existe função positiva $\beta(\cdot)$ (cujo formato depende do formato de $F(\cdot)$) tal que

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi, \beta(u), 0}(x)| = 0 \quad (4.4)$$

A recíproca também é válida.

Comentário 11. O caso em que $F(\cdot)$ possui cauda infinita tem atenção especial no nosso trabalho. Por isso é interessante que o leitor tenha em mãos a versão da fórmula (4.4) para este caso. Eis esta:

$$\lim_{u \uparrow +\infty} \sup_{0 < x < +\infty} |F_u(x) - G_{\xi, \beta(u), 0}(x)| = 0 \quad (4.5)$$

Comentário 12. Se por algum motivo temos *a priori* a informação de que a cauda de $F(\cdot)$ é infinita, de acordo com o item (IV) da seção 3.2, $F(\cdot)$ está no domínio de atração de uma Fréchet $\Phi_\alpha(\cdot)$ ou no domínio de atração de Gumbel $\Lambda(x)$. No primeiro caso o teorema de Pickands (seção anterior) fornece $\xi = \alpha^{-1}$, com $\alpha > 0$, para o valor do parâmetro de forma

da GPD que aproxima a cauda de $F(\cdot)$. No segundo caso temos que $\xi = 0$. De qualquer forma, se sabemos que a cauda de $F(\cdot)$ é infinita, a GPD que aproxima a cauda de $F(\cdot)$ terá cauda infinita, além do índice de cauda, ξ , ser não negativo ($\xi \geq 0$).

Por outro lado, se a cauda de $F(\cdot)$ for finita, o mesmo item (IV) da seção 3.2 diz que $F(\cdot)$ está no domínio de atração de uma Weibull $\Psi_\alpha(\cdot)$ ou de uma Gumbel $\Lambda(x)$. Se estiver no domínio de atração de uma Weibull, a GPD $G_{\xi, \beta(u), 0}$ dada pelo teorema de Pickands (seção anterior) possui $\xi = -\alpha^{-1}$, com $\alpha > 0$. Para o caso de estar no domínio de atração de uma Gumbel, $\xi = 0$. Assim, se $F(\cdot)$ possui cauda finita, a GPD que aproxima a cauda de $F(\cdot)$, dada pelo POT, tem cauda finita ou exponencial (caso em que $\xi = 0$). Concluimos então que se a cauda de $F(\cdot)$ for finita, tem-se $\xi \leq 0$.

Comentário 13. Note que ao apresentar o Resultado de Pickands não especificamos como encontrar a função $\beta(\cdot)$ envolvida nele. É claro que se passássemos pela demonstração deste resultado, encontraríamos a expressão para $\beta(\cdot)$. Mas, esta expressão dependeria da forma da cauda de $F(\cdot)$; por exemplo, caso $F(\cdot)$ possua cauda infinita e $\bar{F}(x) = x^{-\alpha}L(x)$, onde $L(\cdot)$ é uma função de variação lenta (veja o primeiro parágrafo do item (IV) na sub-seção 3.2; recorde que esta condição garante que $F(\cdot)$ esteja no domínio de atração da distribuição Fréchet com parâmetro α), a expressão para $\beta(\cdot)$ depende de $L(\cdot)$ e de α . Recorde agora que a informação disponível sobre $F(\cdot)$ no problema de estimação de sua cauda está toda contida em sua amostra $x_1, \dots, x_n$. Acontece que esta amostra não carrega informação suficiente para podermos adivinhar a estrutura de $L(\cdot)$ com a precisão adequada para construção exata da função $\beta(\cdot)$. Esta é a razão de termos evitado falar sobre a estrutura de $\beta(\cdot)$.

Comentário 14. Já que não apresentamos a estrutura de $\beta(\cdot)$ envolvida no Resultado de Pickands, também não apresentaremos a fórmula para o valor da distância $\varepsilon(u)$ entre as funções $F_u(\cdot)$ e $G_{\xi, \beta(u), 0}(\cdot)$ da fórmula (4.4). A razão para isto é a mesma da apresentada no comentário 13: o desconhecimento da estrutura exata de $F(\cdot)$ e a impossibilidade de inferir sobre esta com a precisão adequada a partir da amostra de $F(\cdot)$.

4.3 Da relação entre as EVD's e as GPD's

Existe uma relação intrínseca entre as EVD's (as Distribuições de Valores Extremos) e as GPD's (as Distribuições de Pareto Generalizadas). O Resultado de Pickands, a base do método discutido neste texto, é uma consequência desta relação, ou, dependendo do ponto de vista, parte dela. A descrição completa desta relação não coube no nosso texto, mas de qualquer forma nos sentimos obrigados a avisar o leitor sobre a sua existência, e exibir duas das diversas formas nas quais ela se manifesta. Uma delas já foi exibida: a fórmula (4.3) do

Resultado de Pickands. Esta expõe o seguinte lado da relação entre EVD's e GPD's:

à cada EVD corresponde um conjunto de GPD's; a correspondência se dá da seguinte forma: cada EVD possui "seu" domínio de atração; considere então aquelas GPD's para as quais o método POT pode construir um aproximador para a cauda das funções deste domínio; este é o conjunto correspondente. O valor do parâmetro ξ das GPD's deste conjunto está rigidamente relacionado com o valor do parâmetro α da EVD correspondente; esta relação foi determinada pela fórmula (4.3). (4.6)

O fato (4.6) tem uma consequência muito importante na execução da idéia do método POT; esta consequência é o assunto do Comentário 15 abaixo.

A segunda forma da manifestação da relação entre EVD's e GPD's é a seguinte:

todas as EVD's e todas as GPD's estão numa relação um-a-um determinada pela seguinte fórmula:
a função de distribuição de EVD =
= exp { - a função de sobrevivência de GPD },
válida em um sub-conjunto do conjunto onde as EVD's estão definidas. (4.7)

Este fato será exibido no do Comentário 16 devido a sua ligação com a representação padronizada das EVD's, que gostaríamos de apresentar aos leitores, já que esta aparece com frequência na literatura. Não usaremos o fato acima nos argumentos do nosso texto.

Ao leitor, que queira entender a relação entre EVD's e GPD's em um nível mais profundo, sugerimos que comece com a leitura da Seção 3.4 de [6].

Comentário 15. Os próximos fatos foram mencionados e empregados na discussão da Subseção 2.3, e ainda serão frequentemente usados em nossos futuros argumentos:

- 1) se a cauda de $F(\cdot)$ é infinita, então a GPD, construída pelo Resultado de Pickands para aproximar a função de excessos, $F_u(\cdot)$, também terá cauda infinita;
- 2) se a cauda de $F(\cdot)$ é finita, então a GPD construída pelo Resultado de Pickands para aproximar a função de excessos, $F_u(\cdot)$, terá ou cauda finita ou a cauda da distribuição exponencial.

Mostrar como estes fatos provêm da fórmula (4.3) e da caracterização dos domínios de atração das EVD's. Note que o comentário expande o conteúdo dos fatos, indicando precisamente como o valor do parâmetro de forma ξ da função aproximador depende da cauda de $F(\cdot)$.

Se a cauda de $F(\cdot)$ é infinita, de acordo com o ítem (IV) da Sub-seção 3.2, $F(\cdot)$ está no domínio de atração de uma Fréchet $\Phi_\alpha(\cdot)$ com $\alpha > 0$, ou no domínio de atração de Gumbel $\Lambda(\cdot)$. No primeiro caso, a GPD que aproxima $F_u(\cdot)$ no Resultado de Pickands seu parâmetro $\xi = \alpha^{-1}$, com $\alpha > 0$. Isto é determinado pela fórmula (4.3). No segundo caso, tem-se $\xi = 0$. Em ambos os casos a GPD aproximadora possui cauda infinita.

Por outro lado, se a cauda de $F(\cdot)$ for finita, o mesmo ítem (IV) da Sub-seção 3.2 diz que $F(\cdot)$ está no domínio de atração de uma Weibull $\Psi_\alpha(\cdot)$ com $\alpha > 0$, ou de uma Gumbel $\Lambda(\cdot)$. Se estiver no domínio de atração de uma Weibull, então, de acordo com a fórmula (4.3), a GPD $G_{\xi, \beta(u), 0}$ dada pelo Resultado de Pickands possui $\xi = -\alpha^{-1}$, com $\alpha > 0$. Tal GPD possui cauda finita. Para o caso de estar no domínio de atração de uma Gumbel, $\xi = 0$. Tal GPD possui cauda infinita, a da distribuição exponencial. Então se a cauda de $F(\cdot)$ for finita, a GPD aproximador tem cauda finita ou exponencial.

Comentário 16. O conteúdo do presente comentário pode ser ignorado numa primeira leitura. Ele visa somente exibir uma forma comum de apresentar as EVD's e chamar a atenção para uma profunda relação entre EVD's e GPD's.

Volte à Sub-seção 3.2 e observe a família de distribuições de valores extremos (EVD's) ali apresentada. Recorde que esta constitui-se da família das distribuições de Fréchet, da família da distribuições de Weibull, e da distribuição de Gumbel. Observe que a família de Fréchet é indexada por um parâmetro que pode assumir qualquer valor positivo, e que a família de Weibull também é indexada por um parâmetro que igualmente pode assumir qualquer valor positivo. Logo se vê que na “indexação” das EVD's os valores positivos são duplamente utilizados, enquanto que os valores não positivos são ignorados. Isso sugere que deva existir uma maneira de indexar todas as EVD's por um parâmetro que corre de $-\infty$ a $+\infty$. Para tal fim é suficiente reescrever a família de Weibull de modo que seu índice corra nos valores negativos, postular que a distribuição de Gumbel corresponde ao valor 0 do índice, e deixar o que índice assuma os valores positivos para a família de Fréchet. Esta idéia é executada abaixo. A vantagem do resultado é puramente estética. A diferença é que as EVD's são apresentadas de maneira que dependam de apenas um parâmetro, chamado aqui de a , que pode assumir qualquer valor real:

ao valor 0 do parâmetro a corresponde a função

$$\exp\{-e^{-x}\} \text{ definida para todo } x \in \mathbb{R},$$

que é a função de distribuição de Gumbel;

a qualquer valor positivo do parâmetro a corresponde a função

$$\begin{aligned} &0, & \text{se } x \leq 0, \\ &\exp\{-x^{-a}\}, & \text{se } x > 0, \end{aligned}$$

que é a função de distribuição de Fréchet com parâmetro a ;

a qualquer valor negativo do parâmetro a corresponde a função

$$\begin{cases} \exp\{-[(-x)^{-a}]\}, & \text{se } x \leq 0, \\ 1, & \text{se } x > 0, \end{cases}$$

que é a função de distribuição de Weibull com parâmetro $-a$.

Acima, para cada família de distribuições, usamos sua fórmula. Aí então surge uma nova pergunta: “Há como apresentar todos os três casos numa mesma fórmula?” Por uma só não, mas por duas – grande economia – sim. Tal apresentação leva os nomes de Jenkinson e de von Mises. Aqui está:

$$H_\xi(x) = \begin{cases} \exp\left\{-(1 + \xi x)^{-1/\xi}\right\}, & \text{se } \xi \neq 0, \\ \exp\{-\exp(-x)\}, & \text{se } \xi = 0 \end{cases} \quad (4.8)$$

sendo que esta expressão vale para

$$\begin{aligned} x &> -\xi^{-1} && \text{quando } \xi > 0, \\ x &< -\xi^{-1} && \text{quando } \xi < 0, \\ x &\in \mathbb{R} && \text{quando } \xi = 0, \end{aligned} \quad (4.9)$$

e sendo que para x fora do conjunto determinado acima a função $H_\xi(\cdot)$ deve ser completada por valor 0 ou 1 da maneira que garanta que o resultado seja uma função de distribuição. A função $H_\xi(\cdot)$ é a *Distribuição padrão generalizada de Valores Extremos* (*standard generalized Extreme Value Distribution* em Inglês). Quando $\xi > 0$ ela corresponde à distribuição de Fréchet com parâmetro $\alpha = 1/\xi$; quando $\xi < 0$ ela corresponde à distribuição de Weibull com parâmetro $\alpha = -1/\xi$; quando $\xi = 0$ ela corresponde à distribuição de Gumbel. Note o uso da palavra “corresponde”. Com esta queremos indicar que $H_\xi(\cdot)$ não será idêntica a uma distribuição de Fréchet, de Weibull ou de Gumbel. Ela será uma dessas funções depois de reescalada e deslocada. Os valores das constantes de reescala e deslocamento podem ser facilmente calculadas a partir da expressão de $H_\xi(\cdot)$, usando as idéias apresentadas nos Fatos 1 e 2 da Sub-seção 2.2.

Há diversas razões e conseqüências do fato de $H_\xi(\cdot)$ ser uma EVD reescalada e deslocada. Uma dos lados positivos disso é que a expressão de $H_\xi(\cdot)$ revela claramente a relação (4.7). De fato, da definição (4.8) de $H_\xi(\cdot)$ e das definições (2.1), (2.2), (2.3) das GPD's padrão segue que

$$H_\xi(x) = \begin{cases} \exp\left\{-(1 - G_\xi(x))\right\}, & \text{para } x > -\xi^{-1} && \text{quando } \xi > 0, \\ \exp\left\{-(1 - G_\xi(x))\right\}, & \text{para } x \in [0, -\xi^{-1}] && \text{quando } \xi < 0, \\ \exp\left\{-(1 - G_0(x))\right\}, & \text{para } x > 0 && \text{quando } \xi = 0, \end{cases} \quad (4.10)$$

que é a expressão matematicamente rigorosa da relação (4.7). Esta relação e sua expressão (4.10) são dadas para indicar que há uma ligação íntima entre as EVD's e as GPD's. Infelizmente o nível de detalhamento do nosso texto não nos permite investigar esta ligação.

A relação (4.10) fez com que o parâmetro ξ das GPD's fosse chamado de *parâmetro de forma*. A razão é que, por um lado, o parâmetro ξ das GPD está relacionado com ξ de $H_\xi(\cdot)$ via (4.10), e, por outro lado, o nome natural para ξ de $H_\xi(\cdot)$ seria *parâmetro de forma*, já que este determina a forma da EVD que $H_\xi(\cdot)$ adquire.

Capítulo 5

Breve exposição sobre MCMC, FBST e Misturas de Distribuições

5.1 Misturas de Distribuições

Existem determinadas situações em que os elementos de uma amostra apresentam características distintas e deseja-se saber qual a origem de tais diferenças. Essas diferenças podem ser encaradas como uma mistura. Por exemplo, se temos uma urna com duas bolas brancas e oito bolas vermelhas, uma amostra de 100 sorteios nos fornece a estimativa de uma densidade discreta que é formada por uma ponderação de duas outras. Supondo que¹ $f_1 = \chi_{brancas}$ e $f_2 = \chi_{vermelhas}$ representam as funções características relacionadas às bolas brancas e bolas vermelhas, a densidade que originou os resultados é dada por $0.2f_1 + 0.8f_2$. Vejamos outro exemplo. Segundo a *quantal hypothesis* da neurifisiologia as transmissões entre células são realizadas em pacotes discretos, os quanta, e a quantidade de transmissores em cada pacote pode ser modelada por uma variável aleatória gaussiana. Usando a abordagem das misturas de distribuições é possível estimar o número de componentes dessa mistura (os quanta), que por sua vez podem ser interpretados como sendo a quantidade de vesículas pré-sinápticas. Existem ainda várias manipulações experimentais que modificam o processo de liberação das transmissões entre os neurônios; e informações relevantes sobre tais mudanças podem ser dadas pelos vários parâmetros de uma mistura (veja [20]). O campo de aplicações é amplo. Passa pela genética, sistemas anti-spam, velocidade das galáxias...

Uma mistura pode ser definida como sendo uma distribuição que é formada por ponderações de outras distribuições (as componentes), e tem a vantagem de permitir uma ma-

¹Se b representa o evento “foi sorteada uma bola branca”, e v representa o evento complementar, temos: $\chi_{brancas}(b) = 1, \chi_{brancas}(v) = 0$ e $\chi_{vermelhas}(b) = 0, \chi_{vermelhas}(v) = 1$.

leabilidade muito grande; com umas poucas componentes de mistura é possível conseguir modelagens interessantes(veja [10]).

Figura 5.1: Exemplos de modelagens com duas componentes normais.

Uma combinação convexa

$$\sum_{i=1}^k p_i f_i(x) \quad \sum_{i=1}^k p_i = 1 \quad 1 < k \leq +\infty$$

de um conjunto de distribuições $\{f_i\}_{1 \leq i \leq k}$ é uma mistura. Cada f_i representa uma componente da mistura.

5.1.1 misturas paramétricas

Quando $\{f_i\}$ provêm de uma família paramétrica, de parâmetros desconhecidos $\{\theta_i\}$, é natural analisarmos a mistura paramétrica:

$$\sum_{i=1}^k p_i f(x, \theta_i) \quad \sum_{i=1}^k p_i = 1 \quad 1 < k \leq +\infty$$

5.1.2 estimativa dos parâmetros

Foi dito que uma das grandes vantagens da modelagem via misturas de distribuições é sua maleabilidade. Contudo ela tem seu preço. A dificuldade surge porque a mesma quantidade

de parâmetros que torna o método maleável também é responsável pelas dificuldades na hora de sua estimativa. Utilizaremos as abordagens comumente citadas na literatura para realizar nossas estimativas.

Uma dessas abordagens é o *EM algorithm* (veja, *e.g.*, [2]), que tem suas limitações expostas em [10]. Pela via bayesiana o papel crucial na estimativa dos parâmetros está no cálculo de integrais multimensionais que relacionam os múltiplos parâmetros e, para que essas integrais sejam calculadas, é preciso lançar mão de métodos tais como *Gibbs sampler* e MCMC. Sendo este último o escolhido no desenvolvimento deste trabalho.

5.2 MCMC

O conjunto de métodos denominados por MCMC (*Monte Carlo Markov Chain*) é uma classe de algoritmos utilizada, por exemplo, para gerar amostras a partir de uma dada função – normalmente tais amostras são requeridas quando outros métodos de amostragem não produzem resultados satisfatórios, devido à complexidade das funções em questão. Nesse caso a principal etapa de do MCMC é a construção de uma Cadeia de Markov cuja distribuição estacionária coincida com a função de densidade pré-estabelecida, de forma que, a medida que aumenta o número de iterações, o estado da cadeia passe a ser cada vez mais uma melhor aproximação de uma amostra da função original. Uma aplicação importante dessas amostras é no cálculo de integrais (ver [9], [7], [17]), via método de Monte Carlo.

5.2.1 Cadeias de Markov

Dizemos que o processo discreto $X = (X_1, X_2 \dots)$ é uma cadeia de Markov se satisfaz a condição de Markov:

$$P(X_{t+1} = s | X_0 = x_0, X_1 = x_1, \dots, X_t = x_t) = P(X_{t+1} = s | X_t = x_t)$$

p/ todo $t \geq 1$ e $s, x_0, x_1 \dots x_t \in S$, onde S é o conjunto onde as variáveis aleatórias discretas X_t tomam seus valores (na presente discussão S será considerado enumerável).

evolução da cadeia

Na discussão que segue consideraremos apenas cadeias de Markov que tomam seus valores em espaços de estados enumeráveis, além de serem homogêneas em relação ao tempo. Essa simplificação pode parecer contradizente com os métodos de MCMC, que em vários casos usam cadeias de de Markov não-homogêneas e espaços de estado contínuos, mas em verdade

ela busca apenas uma rápida apresentação da teoria sobre as cadeias de Markov, isso sem perder a diretriz teórica de vista.

Nosso texto é baseado em [7] e [9], e uma discussão sobre não-homogeneidade em relação ao tempo pode ser encontrada em [15]. Em [14] há uma abordagem completa sobre estados de espaços gerais.

Como foi mencionado acima, nossa preocupação consiste em saber de que forma as cadeias de Markov nos auxiliarão na construção de amostras específicas, e nesse sentido a grande pergunta a ser formulada é:

“Dada a cadeia $(X_t)_{t \geq 0}$ podemos considerar, para grandes valores de t , que X_t segue uma distribuição π independente do ponto inicial x_0 ?”

Para responder a essa pergunta vamos primeiramente a algumas definições (veja [7]).

A passagem entre dois estados, x e y , é dada pelo núcleo de transição²:

$$K(x, y) = P(X_{t+1} = y | X_t = x)$$

que por sua vez define um novo conceito: Se π é tal que

$$\sum_y \pi(y) K_t(y, x) = \pi(x) \quad \forall x, \quad \pi \text{ é estacionária}$$

Uma cadeia é dita irredutível se todos os estados se intercomunicam, isto é, se para todo x, y existe algum $t > 0$ onde $P(X_t = y | X_0 = x) > 0$. E uma cadeia irredutível é denominada (fortemente) aperiódica se para algum estado x a probabilidade de permanência em x é estritamente positiva, $K(x, x) > 0$. Isto evita comportamentos cíclicos.

Vamos então responder à questão levantada acima:

Seja $X = (X_t)_{t \geq 0}$ uma cadeia de Markov com núcleo de transição K . Se X é irredutível, aperiódica e tem π como distribuição estacionária então:

$$P(X_t = x | X_0 = x_0) \xrightarrow{t \rightarrow \infty} \pi(x) \quad \forall x \quad (5.1)$$

e para todo estado inicial x_0

Este resultado diz que para construirmos uma cadeia que, a partir de determinado momento forneça amostras confiáveis de uma função de densidade π , basta encontrar um núcleo

²Como as cadeias são homogêneas, o núcleo de transição independe do tempo.

K que tenha π como distribuição estacionária e que dê a cadeia as propriedades de aperiódicidade e irreducibilidade. Apesar desse resultado somente ser válido nas condições estabelecidas inicialmente (homogeneidade da cadeia, estado de espaço discreto), alguns casos mais gerais seguem de perto este mesmo esquema teórico (veja [22], [14], [7]). Nessa mesma bibliografia são apresentadas em pormenor as questões relativas à convergência da cadeia, que não abordaremos neste trabalho.

5.2.2 Integração de Monte Carlo

Um dos grandes entraves da integração clássica é a impossibilidade prática de se realizar integrações em grandes dimensões. Os métodos comumente disponíveis funcionam bem para uma, duas, três dimensões. Acontece que em análise bayesiana é comum aparecer integrais de cinco, dez, ou mais dimensões. Isso nos conduzirá naturalmente à teoria de integração via Monte-Carlo (ver [7], [8], [9]).

A idéia chave da integração via Monte-Carlo é a de relacionar a integral com a esperança de uma variável aleatória; onde o domínio de integração é o conjunto onde a variável aleatória assume suas realizações. Vejamos como isto é feito.

Considerando a função integrável $f : [a, b] \rightarrow R$, temos que

$$I = \int_{[a,b]} f(x)dx = (b-a) \int_{[a,b]} f(x) \frac{1}{(b-a)} dx \quad \text{e} \quad (5.2)$$

$$\int_{[a,b]} f(x) \frac{1}{(b-a)} dx = E_{\pi}[f(X)] \quad (5.3)$$

onde $\pi \sim \text{unif}([a,b])$ e X é uma variável aleatória que assume valores em $[a, b]$.

Um estimador³ para a esperança, e, conseqüentemente, para a integral em (5.3) é dado por

$$I_n = \frac{1}{n} \sum_{i=1}^n f(x_i)$$

onde $\{x_i\}_{1 \leq i \leq n}$ é uma seqüência de realizações da variável aleatória X ; X segue a distribuição π . Dessa forma o valor da integral em (5.2) pode ser aproximado por $(b-a)I_n$, que converge para I . Se $m = E_{\pi}[f(x_i)]$, a variância do estimador vale

³Se $x = (x_1, x_2 \dots x_n)$ é uma seqüência de realizações de uma variável aleatória, dizemos que $t(x)$ é um estimador para o parâmetro θ se $\lim_{n \rightarrow \infty} t(x) = \theta$

$$\frac{1}{n} \int_{[a,b]} (f(x) - m)^2 \quad (5.4)$$

que é aproximada por

$$\frac{1}{n-1} \sum_{i=1}^n (f(x_i) - I_n)^2$$

Os principais entraves à integração de Monte Carlo estão relacionados com a variância, que influencia na taxa de convergência do estimador. Comumente uma convergência mais acelerada é conseguida quando se utiliza estimadores adequados, ou ainda formas mais inteligentes de amostragem. Os dois casos abaixo exploram superficialmente essas idéias. Para um tratamento mais adequado consulte [9] ou [8].

stratified samping

Em estatística *stratified samping* é um método de amostragem de uma população, utilizado quando as sub-populações variam consideravelmente entre si. Neste caso é vantajoso tomar amostras de cada subpopulação (partição) independentemente, porque a média ponderada obtida por intermédio das partições possui menor variação que a média aritmética produzida por uma amostragem usual da população.

Ao prosseguir com essa abordagem, procuraremos encontrar um estimador para a integral em (5.2) que possua uma variância menor que a dada pela seção anterior e que conseqüentemente forneça uma taxa de convergência mais apurada.

Seja $f : [a, b] \rightarrow R$ integrável. Via *stratified sampling* obtemos o seguinte estimador para a integral em (5.3).

$$I_n^* = \sum_{i=1}^M c_i I_{n_i} \quad \text{onde}$$

$$I_{n_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} f(x_{ij})$$

onde M é a quantidade de partições do intervalo $[a, b]$; $n = \sum n_i$; c_i é a medida de cada partição do intervalo e $\{x_{ij}\}$ é uma seqüência de n_i realizações da variável aleatória X na i -ésima partição (distribuída de acordo com a densidade π). A variância do estimador mencionado é

$$\text{var}(I_n^*) = \sum_{i=1}^M c_i^2 \text{var}(I_{n_i}) \quad (5.5)$$

Não apresentaremos aqui as condições nas quais a variância dada pelo *stratified sampling* é menor que aquela fornecida por (5.4), mas estas são detalhadamente descritas em [9], cap.3.

Importance Sampling

Vimos na exposição acima como uma escolha adequada para o estimador pode melhorar a estimativa para a integral em (3.2), mas também é possível melhorar a estimativa realizando amostragens mais inteligentes. Um exemplo de como isso é feito ocorre no *Importance Sampling*. A chave do método consiste em encontrar uma função auxiliar $g : [a, b] \rightarrow R$ que tenha aproximadamente a mesma forma da função $f : [a, b] \rightarrow R$, de modo que em $[a, b]$ o quociente f/g tenha pequena variância. Nesta situação algumas mudanças têm de ser feitas a fim de se evitar distorções do valor da integral.

Seja

$$I = \int_{[a,b]} f(x)dx = \int_{[a,b]} \frac{f(x)}{g(x)}g(x)dx = c \int_{[a,b]} \frac{f(x)}{g(x)}dG(x)$$

para funções g e G satisfazendo

$$\frac{dG(x)}{dx} = \frac{1}{c}g(x) \quad \text{em } [a, b]$$

sendo que

$$G(b) = \frac{1}{c} \int_{[a,b]} g(u)du = 1 \quad \text{e} \quad g \geq 0 \text{ em } [a, b]$$

As duas últimas condições garantem que G seja uma função de distribuição no intervalo $[a, b]$, e isso é necessário para que possamos obter uma amostra de G , que tem o mérito de dar maior relevância a regiões do domínio que verdadeiramente têm “peso” no cálculo da integral. Nesse processo o estimador e sua variância são dados pelas equações

$$I_n^{**} = \frac{1}{n} \sum_{i=1}^n h(x_i) \quad \text{com } h(x) = \frac{f(x)}{g(x)} \quad \text{e} \quad x_i \sim G$$

$$\text{var}(I_n^{**}) = \int_{[a,b]} (h(x) - m)^2 dG(x); \quad \text{onde } m = E_G[h(x_i)]$$

É fácil perceber que o quanto mais f/g se aproxima de uma constante menor é a variância do estimador, ou seja, melhor é seu desempenho.

5.2.3 Metropolis-Hasting

Na subseção anterior foi mencionado que em problemas práticos de integração multi-dimensional é impraticável usar a teoria clássica de integração para avaliar os valores dessas integrais. Como vimos, o método de Monte Carlo procura contornar esse problema relacionando a integral com a esperança de uma determinada variável aleatória. Vimos também que o método pode ser potencializado se pudermos obter uma amostra inteligente dessa variável aleatória. A fim de obtermos essa amostra inteligente utilizaremos um MCMC específico, o *Metropolis-Hastings*, que é definido pelos seguintes passos:

1. Um primeiro ponto, x_0 é escolhido para compor a amostra
2. Em um dado tempo t , um ponto y é gerado de acordo com a *density proposal* $q(x_t, \cdot)$, que é a função encarregada de gerar os candidatos a elemento da amostra. A função $q(x_t, \cdot)$ pode depender do último ponto aceito, x_t .
3. O ponto y é aceito ou rejeitado, de acordo com a probabilidade $\alpha(x_t, y)$

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right\}$$

onde π é a função alvo, a função da qual desejamos obter uma amostra.

4. No caso de y ser aceito, $x_{t+1} = y$; do contrário $x_{t+1} = x_t$ e a cadeia não se move.

Seria notável se nesse momento pudéssemos aplicar os resultados contidos na subseção (5.2.1) e concluir que o núcleo de transição do *Metropolis-Hastings*

$$K(x_t, x_{t+1}) = q(x_t, x_{t+1})\alpha(x_t, x_{t+1}) + \chi_{[x_{t+1}=x_t]} \left\{ 1 - \int q(x_t, w)\alpha(x_t, w)dw \right\}$$

satisfaz a condição de convergência dada em (5.1). Infelizmente, até onde sabemos, ainda não há resultados que garantam (5.1) para o caso geral, ou seja, para núcleos dependentes do tempo e com estados de espaços contínuos, (veja [7], [17]). Felizmente no nosso caso específico, onde usamos o *independent sampler* (veja seção 6.3.2), existem condições que garantem a convergência da cadeia (confira em [22]).

Outro tema que merece destaque é a taxa de convergência do *Metropolis-Hastings*. Digamos que a distribuição da cadeia $P(X_t = x_t | X_0 = x_0) = K(x_t, x_0)$ convirja para π . A partir

de que valor t podemos considerar a amostra gerada por $K(x_t, x_0)$ uma boa aproximação da de uma amostra verdadeira gerada por π ? Em [22] e [14] há abordagens teóricas sofisticadas sobre o tema. Neste trabalho, na subseção 6.3.2 do exemplo 1, seguiremos a perspectiva dada em [7].

5.3 FBST

Em determinada parte do nosso trabalho surge a questão de como avaliar se determinada hipótese possui evidência relevante ou não. No nosso caso específico se trata de saber qual seria a melhor escolha para um parâmetro específico – o limiar. Para a resolução desse problema, usaremos o teste bayesiano de significância denominado FBST (Full Bayesian Significance Test); dado em [16]. Nesta seção daremos uma idéia de como o FBST se apresenta.

Nossas considerações começam com a definição de espaço de parâmetro Θ , que é o conjunto onde é permitido ao parâmetro $\theta = (\theta_1, \theta_2 \dots \theta_n)$ tomar seus valores; já o conjunto de hipótese nula é dado por $\Theta_H \subset \Theta$, onde $H : \theta \in \Theta_H$ define a hipótese a ser analisada e, finalmente, $f(\theta|x)$ é a densidade *a posteriori*. A fim de nos limitarmos a hipóteses precisas, suporemos sempre que $\dim(\Theta_H) < \dim(\Theta)$.

Para obtermos o valor da evidencia associado ao conjunto Θ_H basta seguir os seguintes passos:

(1) Defina

$$T_\varphi = \{\theta \in \Theta | f(\theta) > \varphi\}$$

(2) A credibilidade de T_φ é dada por

$$\kappa = \int_{T_\varphi} f(\theta|x) d\theta = \int_{\Theta} f_\varphi(\theta|x) d\theta \quad (5.6)$$

onde $f_\varphi(y) = f(y)$ se $f(y) > \varphi$ e, caso contrário, $f_\varphi(y) = 0$.

(3) Defina

$$f^* = f(\operatorname{argmax}_{\theta \in \Theta_H} f(\theta)) \quad (5.7)$$

e tome $T^* = T_{f^*}$ como sendo o conjunto “tangente” relativo a hipótese nula, H , cuja credibilidade é κ^*

(4) O valor

$$1 - \kappa^*$$

corresponde à evidência a favor de H , $Ev(H)$.

A idéia do FBST é medir o “peso” da *posteriori* no subconjunto Θ_H . Primeiramente o processo de otimização fornece $f^* = f(\operatorname{argmax}_{\theta \in \Theta_H} f(\theta))$, que por sua vez entra no passo de integração para quantificar a relação entre o conjunto de hipótese nula e o espaço de parâmetros, dando assim a medida da evidência relativa a H .

Vários exemplos de aplicações do FBST são descritos em [16].

Capítulo 6

Desenvolvimento do método

Foi pontuado em outras passagens que em várias situações na natureza as ocorrências de determinados eventos raros sugerem comportamentos que destoam das distribuições clássicas. Normalmente estes eventos raros estão associados a situações onde há grandes perdas ou grandes ganhos. Por este e outros motivos vários trabalhos nos últimos 30 anos se dedicaram à análise dos grandes valores observados, ou seja, à análise da cauda de certas classes uma distribuição.

Um passo fundamental nessa análise é saber a partir de que momento os eventos observados passam a apresentar características distintas da parte mais concentrada da distribuição. E esse é um passo delicado. O presente capítulo trilha por tais caminhos, e se dedica a um novo método que procura obter um valor satisfatório para o limiar u , que pode ser visto com sendo o valor a partir do qual os eventos passam a apresentar característica caudal.

6.1 Fundamentos Conceituais

Antes de mais nada devemos ressaltar que toda a discussão futura se refere a distribuições com cauda infinita, e o motivo de tal premissa tem três causas principais: (1) Experimentalmente distribuições relacionadas a grandes perdas possuem cauda infinita, com a vantagem de que o trabalho é reduzido pela metade; (2) O caso cauda finita segue as mesmas diretrizes do caso cauda finita; (3) A simplificação computacional é maior quando se trata apenas do caso cauda infinita. Maiores detalhes a respeito de caudas finitas e caudas infinitas podem ser obtidos na seção 2.3 do capítulo 2.

Os pilares teóricos que dão sustentação ao método aqui apresentado podem ser resumidos em dois: A teoria das misturas de distribuições (cap. 5), e o teorema sobre o comportamento

assintótico de distribuições devido a Pickands (seção 4.2, cap. 4). Explicaremos, em linhas gerais, como estas duas teorias se relacionam com o nosso método.

Seja $F(\cdot)$ a distribuição da variável aleatória X , ou seja, a distribuição que originou nossa amostra. De acordo com a teoria desenvolvida por Pickands, quanto maior o avanço do limiar sobre a cauda da distribuição maior é a chance dos excedentes (valores maiores que o limiar) serem regidos por uma GPD. Ocorre que nossa amostra é finita, isto é, não existe a possibilidade de se aumentar o limiar indefinidamente, o que nos priva da abordagem assintótica. Por outro lado, o fato de as modelagens baseadas nos valores centrais não se adequarem bem aos valores caudais sugere que deve haver um limiar u permissível, onde o comportamento assintótico a favor da distribuição Pareto Generalizada seja considerável.

A discussão acima propõe naturalmente uma idéia simples. Por que não supor que exista um determinado limiar u a partir do qual nossa $F(\cdot)$ assume a forma de uma GPD, e antes do qual é regida por uma distribuição adequada¹? Nesse sentido $F(\cdot)$ seria escolhida dentre os elementos de uma família de misturas que modelam os pontos centrais da mistura via primeira componente e os pontos caudais via segunda componente. Os elementos dessa família têm a forma geral:

$$f(x|\theta_1, \theta_2) = \alpha f_1(x|\theta_1) + (1 - \alpha)f_2(x|\theta_2)$$

onde $\theta_1 = (\theta_{11}, \theta_{12} \dots \theta_{1k})$ corresponde aos parâmetros da primeira componente da mistura e $\theta_2 = (\theta_{21}, \theta_{22} \dots \theta_{2m})$ aos parâmetros da segunda componente; $(\theta_1, \theta_2) \in \Omega$, é o espaço de parâmetros, que no nosso caso é um subconjunto de $\mathbb{R}^{k+m}$, e $x = (x_1, x_2 \dots x_n)$ é amostra dada.

6.2 Estrutura do desenvolvimento

Como repetido inúmeras vezes este trabalho objetiva encontrar, por intermédio das informações disponíveis *a priori*, a partir de que ponto u do eixo das abscissas inicia a parte caudal da distribuição que rege a amostra. A grande questão pode ser resumida na seguinte frase: “Dado um intervalo $[a, b]$, qual é a evidência de que $u = u_0, u_0 \in [a, b]$?”. Para que possamos quantificar a evidência da hipótese relacionada a essa frase, a pergunta acima deve ser reformulada em termos compreensíveis ao FBST (seção 5.3, cap. 5). Esse é o papel do conjunto de hipótese nula Ω_0 (veja seção 6.3.2 adiante).

¹Uma pergunta que merece atenção é a de como modelar os valores centrais da distribuição. Devido à quantidade maior de pontos em relação à cauda, este problema pode ser resolvido pelo método que o analista julgar mais adequado, segundo, por exemplo, a teoria clássica. Apesar de trabalharmos com valores caudais, ficará evidente que a mistura que usaremos leva em conta a distribuição dos pontos centrais.

Por sua vez o FBST é baseado em duas etapas, uma de otimização e outra de integração. A etapa de otimização consiste em encontrar o ponto (θ_1^*, θ_2^*) que maximiza a densidade *a posteriori*

$$P[\theta_1, \theta_2 | x] \sim \prod_{i=1}^n f(x_i | \theta_1, \theta_2) \pi(\theta_1, \theta_2) \quad (6.1)$$

Neste trabalho a etapa de integração assume um formato particular.

Começaremos com a geração de uma amostra, via *Metropolis-Hastings*, que nos trará informações sobre π , a distribuição inicial dos parâmetros dada em (6.1). Após isso calcularemos a integral em (5.6) para concluir a segunda etapa do FBST e obter a evidência a favor da hipótese nula.

Os próximos exemplos detalham minuciosamente os passos acima, e objetivam descrever todo o processo, que começa com a amostra inicial e culmina com a evidência relativa aos candidatos a limiar da amostra.

6.3 Exemplo1

6.3.1 dados iniciais

Nossa amostra inicial, formada por 500 observações, foi obtida a partir da função de distribuição

$$F(x) = \begin{cases} N(x; 0, 1), & \text{para } x \leq c \\ 1 - (1 + \frac{\xi}{\beta}(x - c))^{-\frac{1}{\xi}} & \text{para } x > c \end{cases} \quad (6.2)$$

onde² $\xi = 0.7$, $u = 1$, $\beta = 1$ e $c = 1.28$. Também

$$N(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

representa a distribuição normal com parâmetros μ e $\tau = \sigma^{-2}$.

²Os valores de ξ e β dependem de c .

6.3.2 Determinação das priors

A densidade *a priori* π quantifica a informação inicial que temos com respeito aos parâmetros α , u , τ , ξ e β . No nosso exemplo essa informação é muito pequena, e nesse sentido esta subseção visa suprir essa deficiência com as densidades marginais da densidade *a posteriori* (veja o final desta seção), que dão uma idéia de como os parâmetros são distribuídos. Essas marginais foram inferidas por intermédio de uma amostra retirada da *posteriori*, tarefa a realizada pelo MCMC³.

Inicialmente assumimos que

$$\begin{aligned}\alpha &\sim \text{unif}([0, 1]) \\ u &\sim \text{unif}([0, 3]) \\ \tau &\sim \text{unif}([0.5, 1.5]) \\ \xi &\sim \text{unif}([0, 2]) \\ \beta &\sim \text{unif}([0, 3])\end{aligned}\tag{6.3}$$

Assim, cada parâmetro foi inicialmente distribuído em um intervalo adequado, altamente não informativo.

MISTURAS

A expressão da mistura que usamos para analisar os dados fornecidos por (6.2) considera que até o ponto u os pontos são regidos pela densidade normal, e que os valores maiores que u são regidos por uma GPD⁴. As componentes da mistura são dadas respectivamente por

$$f_1(x|u, \tau) = \begin{cases} \frac{\tau^{1/2} \exp(-\frac{x^2 \tau}{2})}{(2\pi)^{1/2} (N(u; 0, \tau^{-1/2}))} & \text{se } x \leq u \\ 0 & \text{caso contrário} \end{cases}\tag{6.4}$$

³O resultado seria mais sofisticado se fosse realizado um estudo da dependência entre os parâmetros (método de cópulas, *e.g.*).

⁴É claro que outras composições podem ser utilizadas noutros casos, ou mesmo em um só caso para fins de comprovação da análise.

e

$$f_2(x|u, \xi, \beta) = \begin{cases} \beta^{\frac{1}{\xi}}(\beta + \xi(x - u))^{-\frac{1}{\xi}-1} & \text{se } x \geq u \\ 0 & \text{caso contrário} \end{cases} \quad (6.5)$$

A mistura tem a forma

$$f(x|\alpha, u, \tau, \xi, \beta) = \alpha f_1(x|u, \tau) + (1 - \alpha) f_2(x|u, \xi, \beta) \quad (6.6)$$

Nesse ponto pode ser argumentado que a abordagem via misturas é irrelevante uma vez que f_1 e f_2 não têm domínio comum, ou seja, bastaria considerar o conjunto das distribuições paramétricas compostas pelas duas distribuições acima (distribuições contínuas). Mas em verdade as misturas possibilitam uma modelagem mais abrangente; por exemplo, as descontinuidades são permitidas – as distribuições contínuas formam um caso particular. Devemos também insistir que nosso interesse não reside em encontrar a melhor modelagem para os dados iniciais, mas sim na busca do ponto que determina o início de uma boa modelagem da cauda via GPD. Nesse sentido as descontinuidades podem ser muito interessantes (veja a seção 6.4).

MCMC

Um aspecto que tem de ser muito bem entendido é o fato da nossa amostra (daqui adiante esta palavra faz referência ao sampling dado pelo MCMC) ser formada pelos elementos de uma cadeia de Markov. E para que uma tal cadeia seja construída precisamos de um ponto inicial X_0 , da função limite P e do núcleo de transição $K(x, y)$. Explicaremos.

ponto inicial para o sampling

Foi dito na subseção (5.2.1) do capítulo 5 que sob determinadas condições o estado inicial X_0 não tem influência na convergência da cadeia $\{X_k\}_{k \geq 0}$, apesar de influenciar na taxa de convergência. De qualquer forma tomamos para estado inicial um ponto aleatório que segue as prioris mencionadas acima. Escolhemos o ponto⁵.

(0.8603, 1.04019, 0.9750, 1.1743, 0.8110)

posteriori

⁵Foram usados outros pontos iniciais com o propósito de se testar a estabilidade do método.

A menos de uma constante multiplicativa⁶, a nossa *posteriori* P tem a forma

$$P(\theta|x) \propto \prod_{i=1}^n [f(x|\theta)] \pi(\theta) \quad (6.7)$$

com $\theta = (\alpha, u, \tau, \xi, \beta)$, e sendo π a distribuição *a priori* dos parâmetros.

núcleo de transição

O núcleo de transição é a parte do MCMC (Metropolis-Hastings) que constrói a cadeia que irá convergir para nossa *posteriori*. Ele é formado por duas partes, o *proposal kernel* $q(x, y)$ e a *acceptance probability* $\alpha(x, y)$, descritos em (5.2.3).

$$K(x_t, x_{t+1}) = q(x_t, x_{t+1})\alpha(x_t, x_{t+1}) + \chi_{[x_{t+1}=x_t]} \left\{ 1 - \int q(x_t, w)\alpha(x_t, w)dw \right\}$$

No nosso exemplo concreto usamos, em um primeiro momento, uma *proposal*, como no caso das *prioris*, que é altamente não informativa e que conseqüentemente produz uma amostra da *posteriori* com muito custo. Esta $q(x, y)$ foi definida por

$$q((\alpha_t, u_t, \tau_t, \xi_t, \beta_t)) = q1(\alpha_t)q2(u_t)q3(\tau_t)q4(\xi_t)q5(\beta_t) \quad \text{e}$$

$$q1(\alpha_t) = \text{unif}([0, 1]);$$

$$q2(u_t) = \text{unif}([0, 3]);$$

$$q3(\tau_t) = \text{unif}([0.5, 1.5])$$

$$q4(\xi_t) = \text{unif}([0, 2]); \quad (6.8)$$

$$q5(\beta_t) = \text{unif}([0, 3]);$$

Este caso em que a *proposal* não depende do último ponto aceito é conhecido como *independent sampler*, nele a *acceptance probability* tem a expressão

⁶A busca dessa constante é o motivo de boa parte do trabalho.

$$\alpha(\theta_t, \theta_{t+1}) = \frac{P(\theta_{t+1})q(\theta_t)}{P(\theta_t)q(\theta_{t+1})}$$

onde $\theta_t = (\alpha_t, u_t, \tau_t, \xi_t, \beta_t)$.

PRIMEIROS RESULTADOS

No começo desta seção foi dito que seu objetivo era obter informações em relação à distribuição inicial dos parâmetros α , u , τ , ξ e β . Neste caso o MCMC forneceu amostras da *posteriori* P considerando-se *prioris* uniformes (6.3). A figura (6.1) apresenta a distribuição dos parâmetros segundo a *posteriori* P (amostral) e as respectivas modelagens pelas densidades $beta(26.69, 2.97)$, $0.90normal(1.26, 0.26) + 0.10normal(1.86, 0.06)$, $normal(1.04, 0.12)$, $gama(2.72, 3.83)$ e $gama(4.03, 4.31)$. A vantagem é que agora temos em mãos um conjunto de funções que nos auxiliará tanto na reconstituição das *prioris* quanto nos futuros cálculos de integrais que o método FBST exige.

As funções de densidade da figura (4.1) representam as marginais da *posteriori*:

$$\propto \prod_{i=1}^n [f(x|\alpha, u, \tau, \xi, \beta)] \pi_1(\alpha) \pi_2(u) \pi_3(\tau) \pi_4(\xi) \pi_5(\beta) \quad (6.9)$$

onde

$$\begin{aligned} \pi_1(\alpha) &= \frac{1}{B(26.69, 2.97)} \alpha^{26.69-1} (1-\alpha)^{2.97-1} \quad \text{para } \alpha \in [0, 1] \\ \pi_2(u) &= 0.90 \frac{1}{0.26\sqrt{2\pi}} e^{-\frac{(u-1.26)^2}{2(0.26)^2}} + 0.10 \frac{1}{0.06\sqrt{2\pi}} \exp^{-\frac{(u-1.86)^2}{2(0.06)^2}} \\ \pi_3(\tau) &= \frac{1}{0.12\sqrt{2\pi}} \exp^{-\frac{(\tau-1.057)^2}{2(0.12)^2}} \\ \pi_4(\xi) &= \xi^{2.72-1} \frac{3.83^{2.72} e^{-3.83\xi}}{\Gamma(2.72)} \quad \text{para } x \geq 0 \\ \pi_5(\beta) &= \beta^{4.03-1} \frac{4.31^{4.03} e^{-4.31\beta}}{\Gamma(4.03)} \quad \text{para } \beta \geq 0 \end{aligned} \quad (6.10)$$

Figura 6.1: Marginais referentes aos parâmetros α , u , τ , ξ e σ .

6.3.3 Cálculo da evidência

FBST

espaço de parâmetros

Quando chega o momento de definir que conjunto de valores pode assumir os parâmetros, e de que forma isso se dá, é preciso levar em conta duas questões conflitantes. A primeira diz respeito à velocidade de convergência do método, que é mais acentuada se há boas precisões em relação aos valores que os parâmetros podem assumir. A segunda trata da busca de bons estimadores pelo método, que pode deixar a desejar se certas regiões do domínio forem negligenciadas pelo espaço de parâmetros.

No nosso caso o espaço de parâmetros foi definido pelos intervalos relativos às *prioris* definidas em (6.10).

$$\Omega = \{(\alpha, \tau, u, \xi, \beta) \mid \alpha \in [0, 1], u \in [0, +\infty], \tau \in [0, +\infty], \xi \in [0, +\infty] \text{ e } \beta \in [0, +\infty]\}$$

conjunto de hipótese nula

O espaço de parâmetros Ω é formado pelo conjunto de todos valores possíveis aos parâmetros, enquanto que o conjunto de hipótese nula, Ω_0 , é um subconjunto de Ω definido por alguma condição sobre os parâmetros (A hipótese nula).

A hipótese nula adotada nesse exemplo será aquela que foi formulada na seção 6.2 – hipótese nula H : o limiar u é igual ao valor u_0 com u_0 em $[a, b]$ –, que em termos matemáticos é descrita por

$$\Omega_0 = \{(\alpha, \tau, u, \xi, \beta) \in \Omega \mid u = u_0\}$$

assim a evidência do conjunto Ω_0 é a evidência da hipótese H acima.

passo de otimização

Neste ponto do método utilizamos um dos códigos clássicos de otimização disponíveis para encontrar o ponto de máximo $\theta^* = (\alpha^*, u^*, \tau^*, \xi^*, \beta^*)$ sobre o conjunto de hipótese nula Ω_0 . Mais especificamente a função `optim` disponível no aplicativo (`scilab 3.0`). A fim de evitar regiões intransponíveis, executamos o código de otimização para um conjunto aleatório de pontos iniciais.

A função otimizada:

$$P(\theta|x) \propto \prod_{i=1}^n [f_1(x|\alpha, u) + f_2(x|\tau, \xi, \beta)] \pi(\theta)$$

$$\pi(\theta) = \pi_1(\alpha) \pi_2(u) \pi_3(\tau) \pi_4(\xi) \pi_5(\beta)$$

que equivale a otimizar

$$\ln P(\theta|x) = \sum_{i=1}^n \ln(f_1(x|\alpha, u) + f_2(x|\tau, \xi, \beta)) + \dots$$

$$\ln(\pi_1(\alpha)) + \ln(\pi_2(u)) + \ln(\pi_3(\tau)) + \ln(\pi_4(\xi)) + \ln(\pi_5(\beta))$$

tem como derivadas parciais:

$$\frac{\partial \ln P}{\partial \alpha} = \frac{n_1}{\alpha} - \frac{n_2}{1-\alpha} + \frac{a1-1}{\alpha} - \frac{b1-1}{1-\alpha};$$

$$\frac{\partial \ln P}{\partial u} = 0; \quad \text{pois } u \text{ é fixo em } \Omega_0; \quad u = u_0;$$

$$\frac{\partial \ln P}{\partial \tau} = \frac{n_1}{2\tau} - \sum_{i=1}^{n_1} \frac{x_{1i}^2}{2} - \frac{n_1 u_0 e^{-\frac{u_0^2 \tau}{2}}}{2\sqrt{2\pi\tau}N(u_0, 0, \sqrt{\tau^{-1}})} - \frac{(\tau - m1)}{sd1^2};$$

$$\begin{aligned} \frac{\partial \ln P}{\partial \xi} &= \frac{-n_2 \ln(\beta)}{\xi^2} + \xi^{-2} \sum_{i=1}^{n_2} \ln(\beta + \xi(x_{2i} - u_0)) - \\ &(\xi^{-1} + 1) \sum_{i=1}^{n_2} \frac{(x_{2i} - u_0)}{\beta + \xi(x_{2i} - u_0)} + \frac{a2-1}{\xi} - b2; \end{aligned}$$

$$\frac{\partial \ln P}{\partial \beta} = \frac{n_2}{\xi\beta} - (\xi^{-1} + 1) \sum_{i=1}^{n_2} \frac{1}{\beta + \xi(x_{2i} - u_0)} + \frac{a3-1}{\beta} - b3;$$

com $a1=26.69$, $b1=2.97$, $a2=2.72$, $b2=3.83$, $a3=4.03$, $b3=4.31$, $m1=1.04$ e $sd1=0.12$.

passo de integração

Definido o conjunto de hipótese nula Ω_0 e o número⁷ $P(\theta^*|x) = P^*$ dado pelo passo de otimização, o passo de integração produz a evidência relativa a Ω_0 por intermédio do número⁸

$$I = \int_{\Lambda} P(\theta|x) d\theta$$

onde

$$P(\theta|x) \propto \prod_{i=1}^n [f(x|\alpha, u, \tau, \xi, \beta)] \pi_1(\alpha) \pi_2(u) \pi_3(\tau) \pi_4(\xi) \pi_5(\beta)$$

e

$$\Lambda = \{(\alpha, \tau, u, \xi, \beta) \in \Omega \mid P(\alpha, \tau, u, \xi, \beta) > P^*\}$$

A subseção 5.2.2 mostra como é realizada a integração acima enquanto que as figuras abaixo ilustram o ganho na diminuição da variância proporcionado pelas *prioris* descritas em (6.10).

⁷O valor de P^* é o número representativo da atuação da *posteriori* no conjunto Ω_0 em relação ao espaço de parâmetros Ω .

⁸De acordo com o FBST o valor aqui encontrado, I , é a evidência contra a hipótese H . Assim $1 - I$ é a evidência a favor da hipótese H .

Figura 6.2: Na figura da esquerda temos a variação da *posteriori* sobre uma amostra de 10000 pontos, realizada sobre *prioris* uniformes (6.3); na figura da direita temos a variação agora sobre uma amostra feita com base nas *prioris* dadas em (6.10)

6.3.4 Resultados do Exemplo1

conjunto de hipótese nula	evidência	ponto de ótimo
$\Omega_0(u=0)$	0.01	(0.51, 0, 1.018, 0.16, 0.79)
$\Omega_0(u=0.2)$	0	(0.58, 0.2, 1.00, 0.24, 0.68)
$\Omega_0(u=0.4)$	0	(0.66, 0.4, 1.00, 0.30, 0.62)
$\Omega_0(u=0.6)$	0.30	(0.73, 0.6, 1.00, 0.42, 0.52)
$\Omega_0(u=0.8)$	0.24	(0.81, 0.8, 1.01, 0.49, 0.53)
$\Omega_0(u=1)$	0.65	(0.85, 1, 1.01, 0.67, 0.45)
$\Omega_0(u=1.2)$	0.95	(0.91, 1.2, 1, 0.56, 0.73)
$\Omega_0(u=1.4)$	0.81	(0.93, 1.4, 1.017, 0.47, 0.99)
$\Omega_0(u=1.6)$	0.17	(0.94, 1.6, 1.05, 0.48, 1.07)
$\Omega_0(u=1.8)$	0.14	(0.95, 1.8, 1.07, 0.46, 1.28)
$\Omega_0(u=2)$	0.02	(0.96, 2, 1.10, 0.57, 1.10)

6.4 Exemplo2

Motivados pela exposição detalhada feita no exemplo passado e pela semelhança entre os dois exemplos, apresentaremos aqui apenas os aspectos mais relevantes do processo. O interessante é que nesse segundo exemplo o método ataca uma amostra empírica. Os dados

Figura 6.3: Gráfico da evidência. Só é possível encontrar valores de u com uma evidência maior que 70% no intervalo $[1, 1.41]$. Ao que parece em nenhum outro intervalo a evidência supera 70%.

Figura 6.4: Modelagens da cauda. Na figura (a) foi utilizado na função de mistura o ponto de máximo (conjunto de parâmetros) referente a $u=0.2$, com $EV(\Omega_0(u=0.2)) = 0$; a figura (b) corresponde a $u=0.6$, onde $EV(\Omega_0(u=0.6)) = 0.30$; na figura (c) utilizamos o ponto de máximo $(0.914, 1.25, 1.00, 0.52, 0.82)$, sendo que $EV(\Omega_0(u=1.25)) = 0.97$ e na figura (d) temos $EV(\Omega_0(u=1.6)) = 0.17$.

foram coletados do índice nasdaq e correspondem a uma série de retornos positivos.

Figura 6.5: Histograma que corresponde aos retornos positivos do índice nasdaq.

6.4.1 Obtenção das densidades marginais

Prioris

Consideraremos que

$$\begin{aligned}
 \alpha &\sim \text{unif}([0, 1]) \\
 u &\sim \text{unif}([1.3, 4]) \\
 \tau &\sim \text{unif}([0.1, 2]) \\
 \xi &\sim \text{unif}([0, 2]) \\
 \beta &\sim \text{unif}([0, 2])
 \end{aligned} \tag{6.11}$$

onde $u \in [1.3, 4]$ significa que o início do comportamento caudal está entre 30% e 2.5% do final da amostra (ordenada).

Mistura

Uma vez que nossa amostra é formada apenas pelos retornos positivos do do índice nasdaq, a nossa mistura possui uma expressão um pouco distante daquela vista em (6.6);

$$f_1(x|u, \tau) = \begin{cases} 2 \frac{\tau^{1/2} \exp(\frac{-x^2 \tau}{2})}{(2\pi)^{1/2} (2N(u; 0, \tau^{-1/2}) - 1)} & \text{se } x \leq u \\ 0 & \text{caso contrário} \end{cases} \quad (6.12)$$

e

$$f_2(x|u, \xi, \beta) = \begin{cases} \beta^{\frac{1}{\xi}} (\beta + \xi(x - u))^{-\frac{1}{\xi} - 1} & \text{se } x \geq u \\ 0 & \text{caso contrário} \end{cases} \quad (6.13)$$

A mistura tem a forma

$$f(x|\alpha, u, \tau, \xi, \beta) = \alpha f_1(x|u, \tau) + (1 - \alpha) f_2(x|u, \xi, \beta)$$

Densidades Marginais

Como no exemplo 1 executamos o MCMC a fim de conseguirmos uma amostra que modele a *posteriori* em (6.14), e que assim nos traga informações sobre a distribuição dos parâmetros.

$$P((\alpha, u, \tau, \xi, \beta|x)) \propto \prod_{i=1}^n [f(x|\alpha, u, \tau, \xi, \beta)] \pi_1(\alpha) \pi_2(u) \pi_3(\tau) \pi_4(\xi) \pi_5(\beta) \quad (6.14)$$

com

$$\begin{aligned}
\pi_1(\alpha) &= \frac{1}{B(39.67, 10.12)} \alpha^{39.67-1} (1-x)^{10.12-1} \quad \text{para } \alpha \in [0, 1] \\
\pi_2(u) &= 0.77 \frac{1}{0.2\sqrt{2\pi}} e^{-\frac{(u-1.69)^2}{2(0.2)^2}} + 0.22 \frac{1}{0.41\sqrt{2\pi}} \exp^{-\frac{(u-2.22)^2}{2(0.41)^2}} \\
\pi_3(\tau) &= \tau^{29.44-1} \frac{32.36^{29.44} e^{-32.36\tau}}{\Gamma(29.44)} \quad \text{para } \tau \geq 0 \\
\pi_4(\xi) &= \xi^{1.85-1} \frac{7.56^{1.85} e^{-7.56\xi}}{\Gamma(1.85)} \quad \text{para } \xi \geq 0 \\
\pi_5(\beta) &= \beta^{38.82-1} \frac{40.63^{38.82} e^{-40.63\beta}}{\Gamma(38.82)} \quad \text{para } \beta \geq 0
\end{aligned} \tag{6.15}$$

6.4.2 Cálculo da evidência

O conjunto

$$\Omega = \{(\alpha, \tau, u, \xi, \beta) \mid \alpha \in [0, 1], u \in [0, +\infty], \tau \in [0, +\infty], \xi \in [0, +\infty] \text{ e } \beta \in [0, +\infty]\}$$

define o espaço de parâmetros. Já o conjunto de hipótese nula é dado por

$$\Omega_0 = \{(\alpha, \tau, u, \xi, \beta) \in \Omega \mid u = u_0\}$$

O próximo passo consiste em variar o limiar u no intervalo $[1.3, 4]$ e obter a evidência, por intermédio do FBST, de cada valor u fixado.

6.4.3 Resultados do exemplo 2

conjunto de hipótese nula	evidência	ponto de ótimo
$\Omega_0(u=1.3)$	0.11	(0.71, 1.3, 0.89, 0.13, 1)
$\Omega_0(u=1.35)$	0.12	(0.72, 1.35, 0.88, 0.12, 1.01)
$\Omega_0(u=1.4)$	0.15	(0.73, 1.4, 0.90, 0.13, 1.01)
$\Omega_0(u=1.6)$	0.52	(0.77, 1.6, 0.94, 0.15, 0.97)
$\Omega_0(u=1.8)$	0.99	(0.80, 1.8, 0.97, 0.20, 0.90)
$\Omega_0(u=2)$	0.38	(0.83, 2, 0.87, 0.20, 0.93)
$\Omega_0(u=2.2)$	0	(0.87, 2.2, 0.81, 0.21, 0.95)
$\Omega_0(u=2.4)$	0.06	(0.89, 2.4, 0.78, 0.23, 0.93)
$\Omega_0(u=2.6)$	0	(0.90, 2.6, 0.71, 0.23, 0.99)
$\Omega_0(u=2.8)$	0.07	(0.92, 2.8, 0.69, 0.25, 0.99)
$\Omega_0(u=3)$	0.02	(0.93, 3, 0.67, 0.28, 0.97)
$\Omega_0(u=3.3)$	0.03	(0.95, 3.3, 0.62, 0.28, 1.03)
$\Omega_0(u=3.6)$	0.07	(0.96, 3.6, 0.60, 0.26, 1.07)
$\Omega_0(u=4)$	0.03	(0.96, 4, 0.58, 0.26, 1.06)

Figura 6.6: Marginais referentes aos parâmetros α , u , τ , ξ e β .

Figura 6.7: Segundo a figura acima, para uma evidência maior que 70%, uma escolha razoável para o limiar u seria um valor entre 1.54 e 1.91.

Figura 6.8: A figura (c) que corresponde ao conjunto $\Omega_0(u=1.8)$ possui evidência igual a 0.99, enquanto que nas figuras (a), (b) e (d) as evidências relativas aos conjuntos $\Omega_0(u=1.3)$, $\Omega_0(u=1.6)$ e $\Omega_0(u=3)$ valem, respectivamente, 0.11, 0.52 e 0.02.

6.5 Considerações finais

No final de cada exemplo tratado anteriormente obtivemos a evidência relativa a cada conjunto de parâmetros $(\alpha, u, \tau, \xi, \beta)$. As figuras 4.4 e 4.8 revelam que o método teve o desempenho esperado, ou seja, a modelagem dos pontos extremos da distribuição por uma GPD passa a ter bastante evidência a partir de determinado limiar u . No caso específico do exemplo 2 o comportamento normal dos dados deixa de existir a certo ponto. Apesar dos resultados satisfatórios, a presente seção tem o objetivo de deixar claro como aperfeiçoar o método por intermédio de algumas novas abordagens. Começaremos com as várias possibilidades para as componentes de modelagem da mistura.

Como já foi destacado, nossa abordagem usa a idéia de que até certo limiar u a amostra segue uma determinada distribuição, daí adiante uma GPD. Essa determinada distribuição pode assumir inúmeras formas, inclusive a de uma mistura. Não podemos nos esquecer que cada caso é um caso, ou seja, a escolha depende do problema atacado. Entretanto algumas etapas do processo devem ser monitoradas, tais como a falta de representatividade da amostra coletada, ou a utilização de uma distribuição grosseira para a modelagem dos pontos centrais. Nos dois exemplos abordados assumimos que os pontos centrais da distribuição obedecem à distribuição normal. Vejamos outro tema importante.

Há certas situações em que o Metropolis-Hastings (o método de Monte Carlo utilizado na nossa análise) produz uma amostra de uma dada distribuição a um preço muito caro, e essa lentidão se deve à taxa de aceitação dos pontos y propostos. Uma vez que é o núcleo de transição $K(x, y)$ que gera os candidatos y , as principais estratégias de aperfeiçoamento do Metropolis-Hastings consistem em escolhas de núcleos $K(x, y)$ apropriados. Uma apresentação de como isso é feito pode ser encontrada em [7] ou [15]. Também podemos buscar em vez do aperfeiçoamento do Matropolis-Hastings, a substituição deste por um outro método de MCMC e comparar os desempenhos relativos. O *Gibbs Sampler* é um desses outros métodos amplamente divulgados na literatura sobre o assunto, veja [7].

Bibliografia

- [1] Belitsky, V. and Moreira, F., (2006) *Emprego do método "Peaks-Over-Threshold" na estimativa de risco; uma exposição abrangente, detalhada e simples* (a sair).
- [2] Bilmes, J.A. (1998) *A Gentle Tutorial of the E.M. Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*. International Computer Science Institute, Berkeley, april.
- [3] Coles, S., (2001) *An introduction to statistical modeling of extreme values*. Springer (Springer Series in Statistics).
- [4] Danielsson, J. and de Vries, C. G. (1987) *Tail index and quantile estimation with very high frequency data*. Journal of Empirical Finance, 241-257.
- [5] DuMouchel, W. H. (1987) *Estimating the stable index α in order to measure tail thickness: A critique*. The Annals of Statistics, vol. 11, No 4, december, pp. 1019-1031.
- [6] Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997) *Modeling Extremal Events for Insurance and Finance*. Springer, Berlin.
- [7] Gilks, W. R. Richardson, S. Spiegelhalter, D. J. (1996) *Markov Chain in Monte Carlo Practice*. Chapman and Hall.
- [8] Hammersley, J. M. Handscomb, D. C. (1964) *Monte Carlo Methods*. John Wiley & Sons.
- [9] Madras, N. (2002) *Lectures on Monte Carlo Methods*. American Mathematical Society.
- [10] Marin, J. M. Mengersen, K. Robert, C. P. (1993) *Bayesian Modeling and Inference on Mixtures of Distributions*. Springer-Verlag.
- [11] McNeil, A. J. (1996) *Estimating tail of loss Severity Distributions using Extreme Value Theory*. Departement Mathematik, ETH Zentrum, CH-8092 Zürich, december.

- [12] McNeil, A. and Saladin, (1997) *T. The Peaks over Thresholds Method for estimating high quantiles of loss Distributions*. Departement Matematik, ETH Zentrum, CH-8092 Zürich, april.
- [13] Mendes, B. V. M. (2004) *Introdução à Análise de Eventos Extremos*. Editora E-Papers, Rio de Janeiro.
- [14] Meyn, S. P. Tweedie, R. L. (1993) *Markov Chains and Stochastic Stability*. Springer-Verlag.
- [15] Neal, R. M. (1993) *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Technical Report CRG-TR-93-1. Department of Computer Science. University of Toronto, september .
- [16] Pereira, C. A. P. Stern, J. M. (1999) *Evidence and Credibility; Full Bayesian Test for Precise Hypotheses*. Entropy Journal & 1,69-80,
- [17] Ó Ruanaidh, J. J. K. Fitzgerald, W. J. (1996) *Numerical Bayesian Methods Applied to Signal Processing*. Springer.
- [18] Pickands, J. (1975) *Statistical inference using extreme order statistics*. The Annals of Statistics, **3**, 119-131.
- [19] Reiss, R. D. and Thomas, M. (2001) *Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields*. (2nd edition) Birkhäuser Verlag.
- [20] Smith, B. R. (1993) *Mixture Estimation and the Neurophysiological Quantal Hypothesis*. Biometrics, vol. 49, No. 2, 643-651.
- [21] Smith, R. L. (1987) *Estimating Tail of Probability Distributions*. The Annals of Statistics, vol. 15, No 3, september, pp. 1174-1207.
- [22] Tierney, L. (1994) *Markov Chains for Exploring Posterior Distributions*. The Annals of Statistics, vol. 22, No. 4, december, pp. 1701-1728.