

# Estudo de Estimadores de Impulsividade em Amostras de Sinais de Comunicações

E. Ferraz e R. Coelho

**Resumo**—Nota-se cada vez mais a presença de fenômenos impulsivos ou de cauda-pesada em diferentes áreas, o que sugere que estes não possam ser mais modelados como se fossem Gaussianos. Ao mesmo tempo, a modelagem destes processos como não-Gaussianos motiva a necessidade de uma melhor caracterização dos parâmetros que os definem. Neste trabalho, foram estudados e avaliados os métodos de McCulloch e o QQ plot que estimam impulsividade e índice de cauda. Além disso, foi proposto um método que transpõe as restrições dos dois primeiros. Os estimadores foram aplicados em seqüências amostrais obtidas de comunicações móveis e também de geração de amostras com distribuições Pareto e  $\alpha$ -estáveis.

**Palavras-Chave**—Processos de cauda-pesada, processos  $\alpha$ -estáveis, impulsividade, método de McCulloch, QQ plot.

**Abstract**—Heavy-tailed or impulsive phenomena are often noted in different areas, suggesting that a non-Gaussian assumption may be the most appropriate model for them. This motivates a better characterization of the parameters which define those phenomena. In this work, it was studied and evaluated the McCulloch method and the QQ plot which estimate impulsivity and tail index. Besides, it was proposed a new method which overruns some restrictions of the two first ones. The estimators were applied on both, sample sequences obtained from mobile communications and generated samples distributed as Pareto and  $\alpha$ -stable.

**Keywords**—Heavy-tail processes,  $\alpha$ -stable processes, impulsiveness, McCulloch method, QQ plot.

## I. INTRODUÇÃO

A crescente identificação de características impulsivas em sinais, fenômenos e aplicações oriundas de diferentes áreas, tais como comunicações, hidrologia, economia e geologia, sugere que estes não possam ser representados por distribuições Gaussianas [1]. As distribuições não-Gaussianas, estáveis, também designadas como cauda-pesada (*heavy-tail*), permitem a caracterização destes fenômenos impulsivos. A propriedade da estabilidade, o Teorema do Limite Central Generalizado [2] (TLCG) e a Distribuição Generalizada de Pareto [3] (DGP) possibilitam avaliar estas características em casos limites.

Neste trabalho, foram estudados e avaliados dois métodos de estimação, um cuja impulsividade é estimada a partir de distribuições  $\alpha$ -estáveis ( $\alpha$ -stable) a partir do método de McCulloch [4] e outro onde se considera o índice da cauda-pesada na obtenção da impulsividade, através do *QQ plot* [5].

Um novo método de estimação, denominado método da Soma Convergente, também é proposto neste artigo. Este método propõe-se a estimar o grau de impulsividade, após tratamento, em amostras com características de cauda-pesada.

O método demonstrou resultados melhores do que os clássicos para distribuições genéricas. Os estimadores foram aplicados em seqüências amostrais obtidas de comunicações móveis e também de geração de amostras com distribuições Pareto e  $\alpha$ -estáveis.

O restante deste trabalho está organizado da seguinte forma: a seção II mostra as definições e fornece uma breve explicação sobre impulsividade e cauda-pesada; os estimadores estão descritos em detalhes na seção III; na seção IV, os resultados são apresentados; e a seção V conclui o trabalho.

## II. ESTABILIDADE, IMPULSIVIDADE E CAUDA-PESADA: DEFINIÇÕES

A propriedade da estabilidade é responsável por grande parte da motivação de se modelar sinais com o uso de funções  $\alpha$ -estáveis e é enunciada como [2]:

*Teorema 1:* Uma variável aleatória (V.A.)  $X$  é  $\alpha$ -estável se, e somente se, para duas V.A.s independentes  $X_1$  e  $X_2$  com a mesma distribuição de  $X$ , e para quaisquer  $a_1$  e  $a_2$  reais, existem  $a$  e  $b$  tais que

$$a_1 X_1 + a_2 X_2 \stackrel{d}{\approx} aX + b \quad (1)$$

onde  $\stackrel{d}{\approx}$  significa semelhança em distribuição. Além disso, se  $X$  é estável, então sua função característica é definida por quatro parâmetros. O parâmetro  $a$  define a localização da distribuição. Tem-se que  $\gamma$  representa a dispersão da distribuição. O parâmetro  $\beta$  mede o nível de simetria da distribuição. Por fim,  $\alpha$  é o parâmetro que caracteriza a impulsividade da distribuição. A função característica desta família de distribuições é dada por:

$$\phi(t) = \exp\{jat - \gamma|t|[1 + j\beta\text{sign}(t)\omega(t, \alpha)]\}$$

onde

$$\omega(t, \alpha) = \begin{cases} \tan(\alpha\pi/2) & \text{para } \alpha \neq 1 \\ 2/\pi \log|t| & \text{para } \alpha = 1, \end{cases}$$

$$\text{sign}(t) = \begin{cases} 1 & \text{para } t < 0 \\ 0 & \text{para } t = 0 \\ -1 & \text{para } t > 0, \end{cases}$$

e

$$-\infty < a < \infty, \gamma > 0, 0 < \alpha \leq 2, -1 \leq \beta \leq 1.$$

Sinais ou amostras com características impulsivas possuem função densidade de probabilidade (fdp) com caudas mais pesadas que uma distribuição Exponencial [1], [2]. Como exemplos de amostras impulsivas, pode-se citar a distribuição de Pareto ou a  $\alpha$ -estável.

E. Ferraz e R. Coelho, Instituto Militar de Engenharia, Departamento de Engenharia Elétrica, Laboratório de Comunicações e Sistemas Ópticos (LaRSO), Emails:{ferraz,coelho}@ime.eb.br. Este trabalho foi parcialmente apoiado pela FAPERJ.

Uma V.A.  $X$  possui distribuição cauda-pesada (*heavy-tail distribution*, ou HTD) se obedece a condição

$$P(|X| > x) \approx \frac{L(x)}{x^\alpha} \quad (2)$$

onde  $L(x)$  é tal que  $\lim_{\tau \rightarrow \infty} L(\tau x)/L(\tau) = 1$ . Nota-se que uma constante qualquer  $c$  obedece a esta propriedade, o que permite escrever que V.A.s distribuídas como Pareto, com

$$P(|X| > x) = \frac{c}{x^\alpha}$$

são HTDs. Isso motiva o uso do seguinte teorema [3]:

*Teorema 2:* Sejam  $u$  e  $c$  constantes reais positivas,  $F_u(x) = P(X - u \leq x | X > u)$  e

$$G(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi} & \text{se } \xi \neq 0 \\ 1 - e^{-x/\beta} & \text{se } \xi = 0 \end{cases} \quad (3)$$

Então

$$\lim_{u \rightarrow \infty} F_u(x) = G(x)$$

Para valores de  $x$  muito grandes, tem-se que para  $1/\alpha = \xi \neq 0$

$$1 + \frac{x}{(\alpha\beta)} \approx \frac{x}{(\alpha\beta)}$$

e se  $c = (\alpha\beta)^\alpha$ :

$$P(X - u > x | X > u) \approx \frac{c}{x^\alpha} \quad (4)$$

Portanto, quando  $\xi > 0$ ,  $G(x)$  toma a forma de uma distribuição de Pareto. Assim, percebe-se que o valor de  $\xi$  permite também modelar um conjunto de momentos que sejam infinitos. Por exemplo, se  $\xi = 0,25$ , então o quarto momento, ou qualquer momento superior, não irá convergir. De forma semelhante, a variância diverge se  $\xi \geq 0,5$ . Com isso, pode-se aproximar as distribuições ou, ao menos, parte delas, por distribuições de Pareto. Assim, a impulsividade de processos de cauda-pesada e de processos estáveis está relacionada através de

$$\alpha = \frac{1}{\xi}$$

onde  $\alpha = 2$  é o valor da impulsividade de uma distribuição Gaussiana ou cauda exponencial. Amostras com  $\alpha < 2$  apresentam variâncias infinitas.

### III. ESTIMADORES DE IMPULSIVIDADE E ÍNDICE DE CAUDA

Os métodos de estimação do grau de impulsividade de processos cauda-pesada dividem-se em quatro categorias:

- aqueles que usam fractiles de amostras de dados empíricos [6] [4];
- aqueles que geram a função característica e fazem ajustes a partir da definição da distribuição estável;
- aqueles que utilizam momentos fracionais de baixa ordem;
- aqueles que verificam a taxa de decaimento das distribuições.

Nesta seção serão apresentados os quatro principais métodos disponíveis para a estimação da impulsividade (parâmetro

TABELA I  
INTERVALOS POSSÍVEIS DE ESTIMAÇÃO SEGUNDO [7]

Name	$\alpha$	$\beta$	$\gamma$	$a$
McCulloch	(0,6 , 2)*	(-1,1)*	(0, $\infty$ )*	( $-\infty$ , $\infty$ )*
Kogon	(0 , 2)*	(-1,1)*	(1)	(0)
Ma	(0 , 2)*	(0)	(0, $\infty$ )*	(0)
Tsihrintzis	(0 , 2)*	(0)	(0, $\infty$ )	( $-\infty$ , $\infty$ )*

$\alpha$ ) [7]: McCulloch, Kogon, Ma e Tsihrintzis. Cada método compreende diferentes intervalos de valores para os parâmetros de amostras com supostas distribuições estáveis. A Tab. I apresenta estes intervalos, com marcação de \* onde os valores podem ser de fato estimados. Portanto, nota-se que o único estimador capaz de estimar os quatro parâmetros é o McCulloch, sendo, desta forma, o estimador mais abrangente de todos. A precisão de todos os estimadores de impulsividade é discutida em [7].

#### A. Método de McCulloch

O método de McCulloch [4] é um aprimoramento do método proposto por Fama e Roll [6], e se baseia em características dos quantiles. Fama e Roll notaram que poderiam obter os parâmetros de dispersão  $\gamma$  e de posição  $a$ , independentemente dos outros parâmetros, pois certas combinações de quantiles se alteravam muito pouco (no máximo 4%) para variação dos valores de  $\alpha$  e  $\beta$ . McCulloch estendeu essa idéia, determinando os parâmetros  $\alpha$  e  $\beta$  sem depender dos outros dois parâmetros, assim como determinou os parâmetros  $\gamma$  e  $\beta$  sem o tendenciamento resultante no método predecessor. Se  $\hat{x}_q$  é o quantile  $q$  da distribuição e  $\hat{\nu}_\alpha$  e  $\hat{\nu}_\beta$  são funções dos quantiles para a determinação de  $\alpha$  e  $\beta$  definidas como:

$$\hat{\nu}_\alpha = \frac{\hat{x}_{0,95} - \hat{x}_{0,05}}{\hat{x}_{0,75} - \hat{x}_{0,25}}$$

$$\hat{\nu}_\beta = \frac{\hat{x}_{0,95} - \hat{x}_{0,05} - 2\hat{x}_{0,5}}{\hat{x}_{0,95} - \hat{x}_{0,05}}$$

Essas relações são biunívocas e funcionam sem tendenciamento na faixa indicada pela Tab. I, de forma que conhecendo  $\hat{\nu}_\alpha$  e  $\hat{\nu}_\beta$  determina-se unicamente  $\hat{\alpha}$  e  $\hat{\beta}$  para quaisquer  $\gamma$  e  $a$ . A dificuldade computacional deste método resume-se a ordenar as amostras para determinar os quantiles desejados. Diferentemente dos métodos que avaliam momentos fracionários e pesos das caudas, por depender de apenas 5 quantiles, este método tanto pode estimar impulsividade em amostras não impulsivas, quanto pode não detectar impulsividade em amostras impulsivas (como nas Figs. 1 e 2). Desta forma, o método pode falhar completamente para amostras que não sejam  $\alpha$ -estáveis.

#### B. Método QQ plot

O método *QQ plot* permite inferir o peso da cauda. Esta é uma ferramenta visual onde os quantiles da distribuição empírica são plotados sobre uma distribuição hipotética. Se a amostra provém desta distribuição ou de uma transformação linear desta, o *QQ plot* é linear. Para verificar-se o peso da cauda, considera-se um valor qualquer para  $\xi$  e compara-se com a distribuição obtida diretamente pela amostra. Se esta

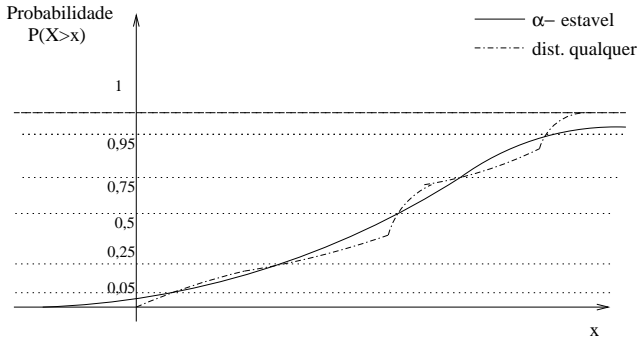


Fig. 1. Exemplo de uma distribuição não impulsiva estimada por McCulloch como impulsiva

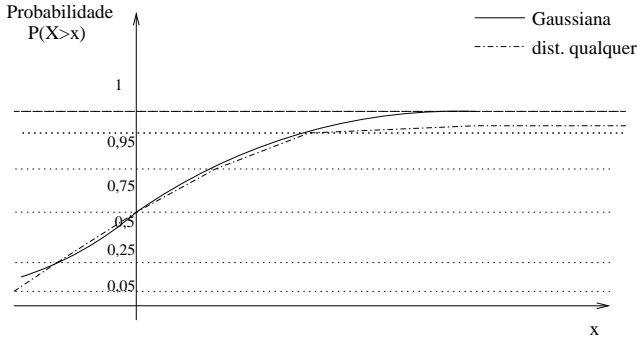


Fig. 2. Exemplo de uma distribuição impulsiva estimada por McCulloch como não impulsiva

última é côncava, então a cauda é mais pesada do que qualquer uma que possua  $\alpha = 1/\xi$ , enquanto que se ela é convexa, a cauda é mais leve de que  $1/\xi$ . Procura-se, então, por tentativa e erro, um valor de  $\xi$  que faça as caudas das duas distribuições (a experimental e a teórica) coincidirem (ou ao menos serem paralelas, pois o fator  $\beta$  de  $G(x)$  é capaz de corrigir isso).

Este método, entretanto, possui a desvantagem da análise por inspeção visual. Certos tipos de amostras são difíceis de serem analisadas pela irregularidade da cauda. Além disso, é comum não encontrar um valor exato, mas um valor de  $\alpha$  razoável para uma faixa de valores. A figura 3 mostra amostras  $\alpha$ -estáveis geradas com cauda de  $\alpha = 1,5$  e a comparação com uma cauda distribuída como  $G(x)$  com o mesmo  $\alpha$ .

C. Método da Soma Convergente (SC)

Como visto anteriormente, mesmo os métodos clássicos apresentam problemas para estimar a impulsividade de sinais com distribuições desconhecidas. Nesta seção, propõe-se um método mais genérico para tal estimação que use como únicas hipóteses amostras independentes e identicamente distribuídas.

Sejam  $X_1, X_2, \dots, X_n$  as amostras independentes e identicamente distribuídas de uma distribuição qualquer. Gera-se então um processo a partir deste dado pela soma de  $\lfloor fn \rfloor$  amostras em seqüência ( $f \in (0, 1)$ ), originando um processo com  $\lfloor 1/f \rfloor$  amostras ( $Y_1, Y_2, \dots, Y_{\lfloor 1/f \rfloor}$ ). Este procedimento será chamado de soma por grupos. Neste caso, nota-se que

$$Y_j = \sum_{i=(j-1)\lfloor fn \rfloor + 1}^{j\lfloor fn \rfloor} X_i$$

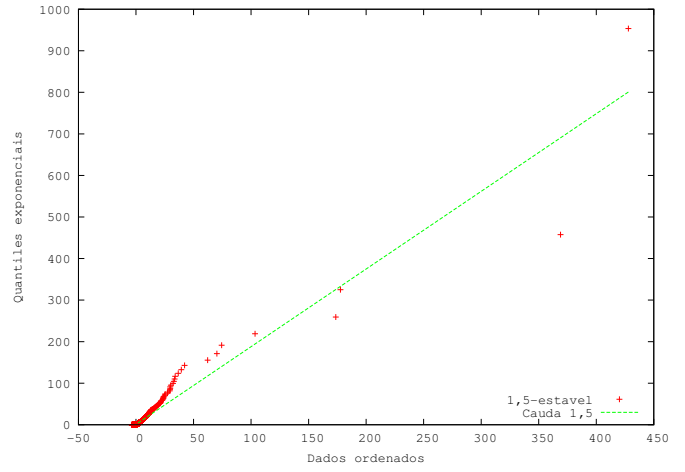


Fig. 3. Uso do QQ plot para comparar duas caudas de  $\alpha=1,5$ , que deveriam ser iguais

Para dar base a tal método, será considerado o conhecido teorema que, por si só, define as funções  $\alpha$ -estáveis:

**Teorema 3:** (Teorema do Limite Central Generalizado) Sejam  $X_1, X_2, \dots, X_n$  independentes e identicamente distribuídas com  $a_n \rightarrow \infty$ . Definindo  $S_n$  como

$$S_n = (X_1 + X_2 + \dots + X_n)/a_n - b_n \tag{5}$$

$X$  é  $\alpha$ -estável se e somente se  $X = \lim_{n \rightarrow \infty} S_n$ .

Pode-se usar o TLGC para provar a seguinte proposição:

**Proposição 1:** Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes e identicamente distribuídas, tais que para algum  $0 < \alpha < 2$ :

$$\mathbf{E}[|X_1|^p] = \infty, \text{ se } p \geq \alpha \tag{6}$$

$$\mathbf{E}[|X_1|^p] < \infty, \text{ se } 0 \leq p < \alpha \tag{7}$$

Então se  $X = \lim_{n \rightarrow \infty} \sum_{i=1}^n X_i$ ,  $X$  é  $\alpha$ -estável. Se  $\alpha \geq 2$ , a distribuição de  $X$  será Gaussiana.

**Demonstração:** Sejam  $R$  e  $S$  variáveis aleatórias identicamente distribuídas com fdps  $p_R$  e  $p_S$ , respectivamente, tais que  $\mathbf{E}[|R|^p]$  e  $\mathbf{E}[|S|^p]$  converjam apenas para  $p \leq \alpha < 2$ . Seja  $T = R + S$  com fdp  $p_T$ . Desta forma:

$$p_T(x) = \int_{-\infty}^{\infty} p_R(\lambda)p_S(x - \lambda)d\lambda$$

Primeiramente é provado que  $\mathbf{E}[|T|^p]$  converge se e somente se  $p \leq \alpha$ . O  $p$ -ésimo momento de  $T$  é dado por:

$$\mathbf{E}[|T|^p] = \int_{-\infty}^{\infty} x^p \int_{-\infty}^{\infty} p_R(\lambda)p_S(x - \lambda)d\lambda dx$$

Trocando a ordem de integração e fazendo a substituição de variáveis  $u = x - \lambda$ , obtém-se:

$$\mathbf{E}[|T|^p] = \int_{-\infty}^{\infty} p_R(\lambda) \int_{-\infty}^{\infty} (u + \lambda)^p p_S(u) du d\lambda$$

Para provar a ida, considera-se o caso em que  $p \leq \alpha$ . Se

$u \leq \lambda$ :

$$\int_{-\infty}^{\lambda} (u + \lambda)^p p_S(u) du < \int_{-\infty}^{\lambda} (2u)^p p_S(u) du =$$

$$2^p \int_{-\infty}^{\lambda} u^p p_S(u) du < 2^p \mathbf{E}[|S|^p]$$

E se  $u > \lambda$

$$\int_{\lambda}^{\infty} (u + \lambda)^p p_S(u) du < \int_{\lambda}^{\infty} (2\lambda)^p p_S(u) du =$$

$$(2\lambda)^p \int_{\lambda}^{\infty} p_S(u) du < (2\lambda)^p$$

Definindo  $C$  como

$$C = \int_{-\infty}^{\infty} p_R(\lambda) (2^p \mathbf{E}[|S|^p] + (2\lambda)^p) d\lambda =$$

$$2^p (1 + \mathbf{E}[|R|^p] \mathbf{E}[|S|^p])$$

tem-se que  $C$  converge. Como  $C > \mathbf{E}[|T|^p]$ , então  $\mathbf{E}[|T|^p]$  também deve convergir, o que prova a ida. Considerando agora  $p > \alpha$ , é verdade que

$$\int_{-\infty}^{\infty} (u + \lambda)^p p_S(u) du \geq \int_{-\infty}^{\infty} (u^p + \lambda^p) p_S(u) du =$$

$$\int_{-\infty}^{\infty} u^p p_S(u) du + \lambda^p$$

Como o termo direito da Eq. 8 diverge, o termo esquerdo também deve divergir, e por consequência,  $\mathbf{E}[|S|^p] = \infty$ . Invocando o TLCG, sabe-se que a soma de infinitas variáveis aleatórias com mesma distribuição deve ser uma variável aleatória  $\alpha$ -estável. Como foi mostrado que os momentos fracionários da soma só convergem para  $p \leq \alpha$ , então  $X$  só pode ser  $\alpha$ -estável. Para  $\alpha \geq 2$  aplica-se o Teorema Central do Limite. ■

Portanto, se  $n \rightarrow \infty$ , de acordo com a Proposição 1, haverá uma seqüência de  $\lfloor 1/f \rfloor$  amostras com distribuição  $\alpha$ -estável, ou se  $n$  é um número grande, haverá ao menos uma distribuição próxima a esta. A partir deste ponto, pode-se usar o método de McCulloch para estimar a impulsividade do processo.

#### IV. RESULTADOS

Para a avaliação dos métodos de estimação, foi considerado um conjunto de 346 seqüências amostrais, com aproximadamente 9000 amostras em cada, obtidas de medidas (*traces*) de uma rede sem fio baseado no padrão IEEE802.11 [8],[9]. Com o *QQ plot*, investigou-se algumas distribuições mais côncavas do que a função do tipo  $G(x)$  descrita na Eq. 3 com  $\xi \geq 0, 5^1$ . A partir deste conjunto, verificou-se o valor aproximado de  $\xi$  e consequentemente da cauda  $\alpha = 1/\xi \leq 2$ . Foram escolhidas 5 seqüências amostrais dentre as iniciais para as estimações. Para cada uma destas, foi estimada a impulsividade pelo método de McCulloch e pelo método proposto com  $fn = 50$  e  $fn = 100$ . Além disso, para cada uma dessas seqüências, foram geradas mais duas, uma distribuída como Pareto e outra

<sup>1</sup>A escolha de  $\xi \geq 0,5$  foi feita pois, de acordo com a Proposição 1,  $\xi < 0,5$  faria o método da SC gerar uma Gaussiana, estimando impulsividade 2.

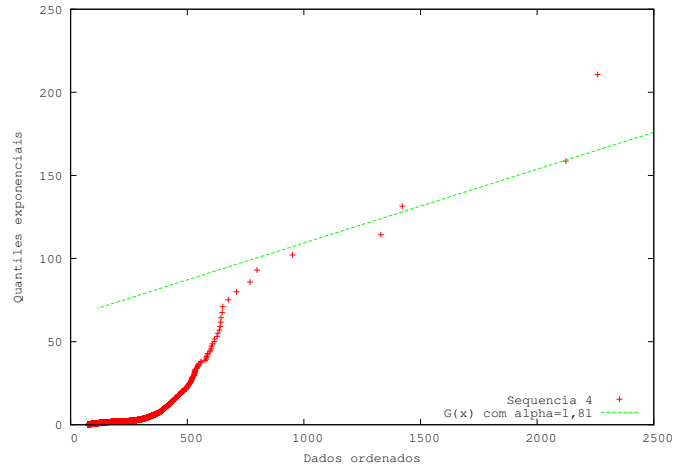


Fig. 4. Estimação da cauda da seqüência 1 a partir do QQ plot, onde  $\alpha = 1,81$

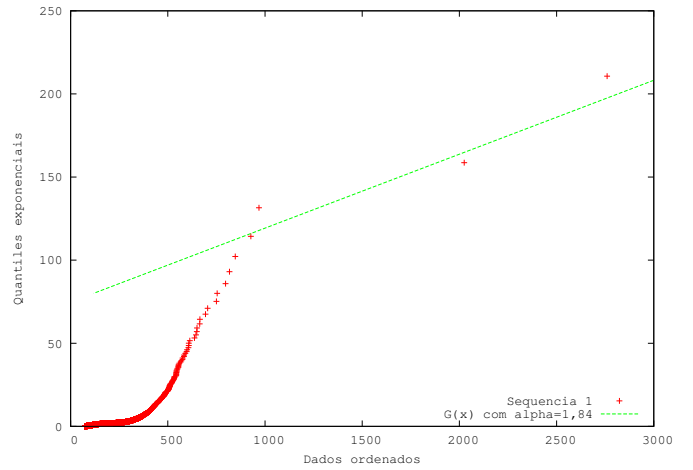


Fig. 5. Estimação da cauda da seqüência 2 a partir do QQ plot, onde  $\alpha = 1,84$

como  $\alpha$ -estável a partir do método descrito em [10] com mesmo número de amostras  $N$  e mesma cauda. Com essas duas séries geradas, foram novamente estimados a cauda e a impulsividade pelos mesmos métodos aplicados nas séries originais. Os resultados são os que se seguem.

##### A. Método QQ plot

As Figs. 4 a 8 apresentam as estimações das caudas das cinco seqüências escolhidas a partir do uso do *QQ plot*. Pode-se perceber a dificuldade de se escolher um possível melhor valor para  $\alpha$ , já que o método é visual e é difícil perceber o ponto exato onde a cauda começa. Nota-se que, exceto pela fig. 5, a reta foi escolhida descartando-se os últimos pontos. Incluí-los resultaria traçar a reta para os dois últimos pontos, o que descartaria todos os outros. Ao mesmo tempo, foram usados apenas 3 pontos para se escolher o  $\alpha$  das figs. 5 e 6.

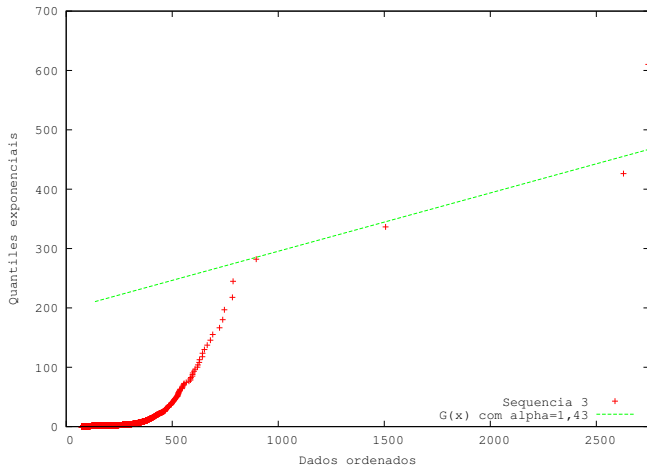
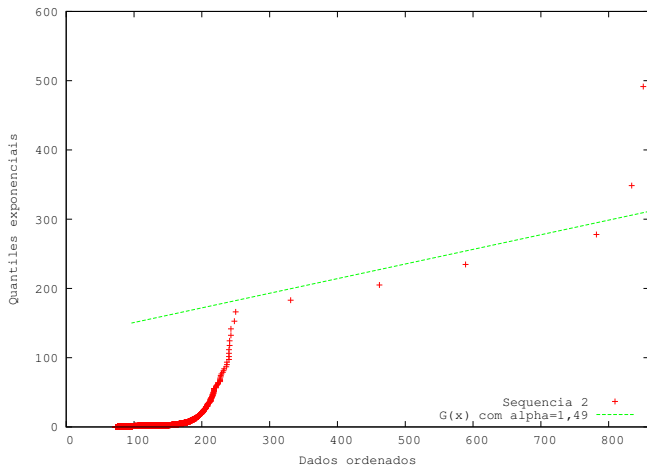
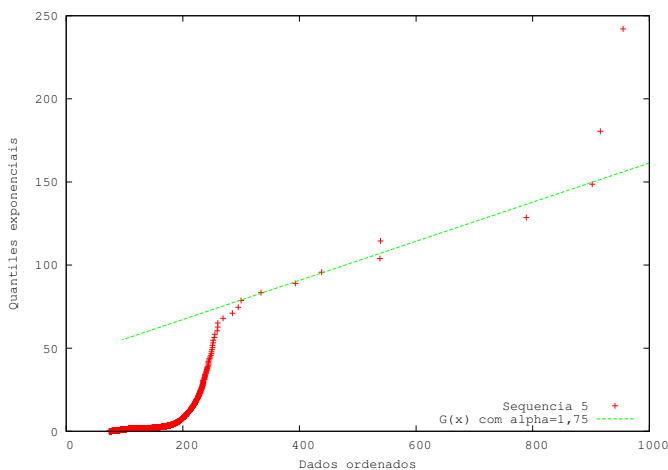
##### B. Comparação dos Diferentes Estimadores

A Tab. II mostra, para cada uma das seqüências amostrais, o número de amostras, o valor estimado pelo *QQ*

TABELA II

VALORES DE  $\alpha$  ESTIMADOS A PARTIR DAS SEQUÊNCIAS AMOSTRAIS ORIGINALIS POR DIFERENTES MÉTODOS

Seq.	N	QQ plot	Originais		
			McCulloch		
			Original	grupos de 50	grupos de 100
1	8621	1,81	2,00	1,90	1,57
2	8620	1,84	2,00	1,78	1,59
3	8622	1,43	2,00	1,64	1,52
4	8622	1,49	2,00	2,00	1,50
5	8622	1,75	2,00	1,76	1,89

Fig. 6. Estimação da cauda da sequência 3 a partir do QQ plot, onde  $\alpha = 1,43$ Fig. 7. Estimação da cauda da sequência 4 a partir do QQ plot, onde  $\alpha = 1,49$ Fig. 8. Estimação da cauda da sequência 5 a partir do QQ plot, onde  $\alpha = 1,75$ 

*plot* (supostamente o valor correto para o peso da cauda), o valor estimado por McCulloch diretamente na sequência e o valor estimado para o método proposto com  $f = 0,02$  e  $f = 0,01$ , ou seja, com a soma por 50 e 100 grupos de amostras, respectivamente. O valor do *QQ plot* foi considerado o mais consistente porque, como já visto, McCulloch pode perceber impulsividade em sequências amostrais quaisquer não impulsivas, da mesma forma que pode não perceber essa propriedade mesmo ela estando presente, ou seja, mesmo com toda a imprecisão do *QQ plot*, ele ainda serve como um bom indicativo para se perceber a presença de impulsividade. A primeira nota a se destacar é que o valor da impulsividade estimado por McCulloch, em todas as amostras, foi igual a 2, o que mostra que McCulloch não conseguiu perceber impulsividade em nenhuma das sequências de cauda-pesada. A única informação que se pode extrair disso é que as amostras não possuem distribuição  $\alpha$ -estável. Além disso, devido à soma por grupos das amostras como no método proposto, o método de McCulloch passa a estimar impulsividade diferente de 2. Entretanto, este valor pode variar muito para diferentes números de grupos. Isso porque o pequeno número inicial de amostras é ainda mais reduzido por este método, o que resultaria em uma estimação de  $\alpha$  imprecisa. Mesmo assim, percebe-se que se alguma das 3 colunas de estimação chegava perto do valor original, isso acontecia depois do tratamento pelo método proposto.

A Tab. III apresenta o comportamento dos estimadores quanto às distribuições de Pareto. Foram escolhidas distribuições com mesma cauda e mesmo número de amostras que as sequências originais. Neste caso, há um certo padrão: McCulloch subestima as impulsividades e este valor tende a aumentar à medida que se aumenta a quantidade de amostras por grupo. Lida-se, portanto, com a relação entre número de amostras e número de grupos: enquanto aumentar o número de grupos faz a amostra aproximar-se de uma  $\alpha$ -estável, há uma diminuição no número de amostras e, portanto, maior imprecisão para estimar com McCulloch. Além disso, para diminuir a dependência temporal, o método usado para a geração de grupos escolhia aleatoriamente as amostras formadoras dos grupos, ou seja, o método usado não é determinístico, pois sua aplicação repetida gera resultados diferentes em uma mesma amostra, o que pode ser facilmente contornado escolhendo um método de escolha das amostras formadoras dos grupos. Esses dois fatores (a troca entre um bom número de amostras por um bom número de grupos e o método usado para a escolha dos grupos) explicam a variação dos resultados, onde são obtidas melhores

TABELA III

VALORES DE  $\alpha$  ESTIMADOS A PARTIR DE SEQUÊNCIAS AMOSTRAIS DISTRIBUÍDAS COMO PARETO POR DIFERENTES MÉTODOS

Pareto				
$\alpha$ gerado	QQ plot	McCulloch		
		Original	grupos de 50	grupos de 100
1,81	1,81	1,17	1,70	1,62
1,84	1,67	1,11	1,36	1,87
1,43	1,45	0,96	1,11	1,43
1,49	1,53	1,08	1,34	1,25
1,75	1,67	1,10	1,37	1,75

TABELA IV

VALORES DE  $\alpha$  ESTIMADOS A PARTIR DE SEQUÊNCIAS AMOSTRAIS DISTRIBUÍDAS COMO  $\alpha$ -ESTÁVEL POR DIFERENTES MÉTODOS

$\alpha$ -estável				
$\alpha$ gerado	QQ plot	McCulloch		
		Original	grupos de 50	grupos de 100
1,82	2,00	1,81	1,64	1,66
1,84	2,00	1,83	1,47	1,72
1,43	1,82	1,42	1,75	1,70
1,49	1,67	1,46	1,63	1,67
1,75	2,00	1,75	1,78	1,64

estimativas com grupos de 100 amostras para três valores de  $\alpha$ . Ao mesmo tempo, os outros dois valores de  $\alpha$  contidos na tabela, 1,91 e 1,49, as melhores estimativas são obtidas para grupos de 50 amostras. Novamente, os resultados com o método proposto são mais precisos do que simplesmente aplicando McCulloch nas sequências. Mais ainda, *QQ plot* estima muito bem a impulsividade, o que era esperado, já que este foi modelado justamente para comparar distribuições cauda-pesada com a Pareto.

A Tab. IV é análoga à Tab. III, exceto que a distribuição gerada é agora  $\alpha$ -estável. O *QQ plot* não foi um bom estimador para a impulsividade, o que pode ser visto na primeira coluna da tabela. Como citado na Seção III, o método de McCulloch estima com ótima precisão a impulsividade de amostras  $\alpha$ -estáveis. Além disso, o método proposto obteve menor precisão do que o método de McCulloch. Isso é esperado, pois, como este procedimento objetiva criar uma distribuição que se aproxime ao máximo de uma  $\alpha$ -estável, ele não possui nenhum efeito sobre variáveis já distribuídas desta forma que não seja degradá-las pela diminuição do número de amostras. Fora isso, comentários análogos aos feitos na Tab. III valem para a Tab. IV no que se refere aos erros cometidos pelo método da soma convergente para grupos de 50 amostras e grupos de 100 amostras. Neste caso, apenas os valores de  $\alpha$  gerados 1,49 e 1,75 foram melhor estimados com grupos de 50 amostras.

## V. CONCLUSÕES

Neste trabalho foi apresentado um estudo sobre três métodos para a estimação do índice de cauda, dois deles clássicos e um proposto. Como apresentado, os estimadores de cauda existentes não são capazes de fazer a estimação em diferentes distribuições. O método proposto, mesmo com baixa precisão, chegou em todos os casos verificados mais próximo de estimar corretamente a cauda do que o pior estimador para cada caso.

Entretanto, este estimador ainda apresenta alguns problemas relevantes: ele foi feito baseado em um processo pseudo-aleatório (e portanto pode, para uma mesma amostra, estimar valores diferentes) e diminui o número de amostras dada uma sequência amostral. A primeira desvantagem pode ser facilmente vencida adotando um algoritmo fixo de escolha de elementos para realizar a soma. Para contornar-se o segundo problema, está em estudo um estimador de cauda, baseado neste trabalho, que passe a trabalhar não mais com as amostras, mas com a distribuição empírica gerada por elas. Isso poderia impedir a redução do número de amostras, permitindo a aplicação deste método com mais precisão onde ele é mais falho: em pequenas sequências amostrais.

## REFERÊNCIAS

- [1] O. Cappé, E. Moulines, J.-C. Pesquet, A. Petropulu, and ueshi Yang, "Long range dependence and heavy-tail modeling for teletraffic data," *IEEE Signal Processing*, pp. 14–27, May 2002.
- [2] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," in *Proceedings of the IEEE*, vol. 81, pp. 986–1010, July 1993.
- [3] A. A. Balkema and L. de Haan, "Residual lifetime at great age," *Annals of Probability*, vol. 2, pp. 792–804, 1974.
- [4] J. H. McCulloch, "Simple consistent estimators of stable distribution parameter," *Commun. Statist.-Simula*, pp. 1109–0036, 1986.
- [5] R. Gençay, F. Selçuk, and A. Ulugülyagci, "Evim: A software for extreme value analysis in matlab," in *Studies in Nonlinear Dynamics and Econometrics*, October 2001.
- [6] E. Fama and R. Roll, "Some properties of symmetric stable distributions," *American Statistical Association*, pp. 817–836, September 1968.
- [7] S. Bates and S. M. Laughlin, "The estimation of stable distribution parameter from teletraffic data," *IEEE Transactions on Signal Processing*, vol. 48, pp. 865–870, March 2000.
- [8] "http://nms.lcs.mit.edu/mbalazin/wireless," 2002.
- [9] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proc. of the First International Conference on Mobile Systems (MobiSys)*, 2003.
- [10] C. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Application*. New York: Wiley, 1995.