As Estáveis no R

Helena Iglésias Pereira, *CEAUL e Departamento de Estatística e Investigação Operacional, FCUL, hmpereira@fc.ul.pt*

Resumo: Neste trabalho é proposto um estimador do parâmetro de assimetria β de uma distribuição estável, e estuda-se a sua distribuição de amostragem com recurso ao software R.

Palavras– chave: distribuições estáveis, parâmetro de assimetria.

Abstract: An estimator of the skewness parameter β of a standard stable distribution with known characteristic exponent $\alpha \neq 1$, is proposed. We use the function *rstable* of the library *Rmetrics* to study his sampling distribution.

Key-Words: Stable distributions; skewness parameter.

1. INTRODUÇÃO

É comum em estatística aplicada, assumir que os fenómenos aleatórios observados são o efeito de um grande número de causas independentes e não observáveis que se adicionam resultando no fenómeno em estudo. Pelo teorema Limite Central a soma de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) convenientemente centrada e reduzida, tem distribuição assintótica normal. Mais geralmente, a soma de v.a.'s i.i.d., convenientemente normada tem ainda distribuição normal, desde que se imponham algumas condições no comportamento assintótico do segundo momento truncado das parcelas.
Pelo Teorema Limite Central Generalizado, se a soma de v.a.'s i.i.d. tem distribuição limite não degenerada esta distribuição tem de ser um elemento da classe das leis *estáveis*, de que a normal é o único elemento com variância finita.
A classe das leis estáveis é caracterizada por quatro parâmetros, usualmente designados por α, β, a e c, respectivamente o expoente característico (e.c.), parâmetro de assimetria, localização e escala.
Embora as variáveis aleatórias estáveis tenham propriedades aditivas interessantes e sejam absolutamente contínuas, somente se conhecem expressões analíticas das funções densidade de probabilidade (f.d.p.) correspondentes aos casos: α=2 (normal), α=1, β=0 (Cauchy) e α=1/2, β=1 (Lévy).
Este facto aliado à não existência de momentos para algumas destas distribuições, dificulta muito o problema da inferência estatística em modelos estáveis. Apesar das dificuldades apontadas vários métodos foram utilizados para estimar os quatro parâmetros das leis estáveis, ver por exemplo Iglésias Pereira, H. (1985) e Zolotarev, V.M. (1986).
Mais recentemente, a existência de programas de computador fiáveis permite calcular as funções densidade, as funções de distribuição e os quantis das distribuições estáveis.
Neste trabalho apresenta-se um estimador do parâmetro de assimetria sugerido pelo comportamento na origem da função de distribuição (f.d.) de uma lei estável, e estuda-se a sua distribuição de amostragem com recurso ao software R.

Investigação financiada por FCT/POCTI e POCI/FEDER

2. MODELOS ESTÁVEIS: DEFINIÇÕES E PROPRIEDADES

Seja $\{Y_i\}_{n \in \mathbb{N}}$ uma sucessão de v.a.'s i.i.d., a v.a. X diz-se *estável* sse para todo o $n \in \mathbb{N}$ existem constantes $a_n > 0$ e $b_n \in \mathbb{R}$ tais que

$$\frac{\sum_{i=1}^{n} Y_i - b_n}{a_n} \xrightarrow{d} X, \quad n \to +\infty \tag{2.1}$$

A classe das distribuições estáveis tem função característica da forma (A)

$$\phi(t) = exp\left\{ iat - c|t|^{\alpha}\left[1 - i\beta \frac{t}{|t|}\omega_A\left(|t|, \alpha\right)\right]\right\} \tag{2.2}$$

$$onde\ a \in \Re,\ c > 0,\ \alpha \in (0,2],\ \beta \in [-1,1] \quad e$$

$$\omega_A\left(|t|, \alpha\right) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right), & \alpha \neq 1 \\ -\frac{2}{\pi}\ln|t|, & \alpha = 1 \end{cases} \tag{2.3}$$

Ou ainda da forma (B)

$$\phi(t) = exp\left\{ iat - c|t|^{\alpha}\omega_B(|t|, \alpha)\right\}$$

onde $a \in \mathbb{R},\ c > 0,\ \alpha \in (0,2],\ \beta \in [-1,1]$ (2.4)

$$\omega_B(|t|, \alpha) = \begin{cases} exp\left(-i\left(\frac{\pi}{2}\right)\beta K(\alpha)sgn(t)\right), \alpha \neq 1 \\ \frac{\pi}{2} + i\beta\ log|t|\ sgn\,t\ , \alpha = 1 \end{cases} \tag{2.5}$$

$K(\alpha) = 1 - |1 - \alpha|\ , \alpha \neq 1$.

Os parâmetros estão relacionados através de:

$\beta_A = \beta_B, \quad c_A = \pi c_B / 2 \quad se\ \alpha = 1$ e

$$\left.\begin{array}{l} \beta_A = cot\,an\left(\frac{\pi\alpha}{2}\right) tan\left(\frac{\pi\beta_B K(\alpha)}{2}\right) \\ \qquad\qquad\qquad\qquad se\ \ \alpha \neq 1 \\ c_A = c_B\, cos\left(\frac{\pi}{2}\beta_B K(\alpha)\right) \end{array}\right. \tag{2.6}$$

O expoente característico α é o mesmo nas duas formas [6], p.12.

As distribuições estáveis têm propriedades interessantes:

I) As caudas da função de distribuição $F_X(x;\alpha,\beta)$ de uma v.a. X estável satisfazem

$$x^{\alpha}\left[1-F_X(x)\right]\to C\frac{k_2}{k_2+k_1}\frac{2-\alpha}{\alpha}$$

$$x^{\alpha}F_X(-x)\to C\frac{k_1}{k_2+k_1}\frac{2-\alpha}{\alpha} \qquad (2.7)$$

$$quando\ x\to+\infty \quad e \quad (C>0,\ k_1,k_2\geq 0\ \ e\ \ k_1+k_2>0)$$

Donde se conclui que o e. c. α está intimamente relacionado com o comportamento das caudas da f.d., sendo o peso destas tanto menor quanto maior o e.c. (a normal é a estável com caudas mais leves).

II) Toda a distribuição estável de expoente característico α ($0<\alpha<2$) tem momentos absolutos finitos de ordem $\gamma\in(0,\alpha)$ (esta propriedade é consequência da anterior).

III) Dado que a função característica ϕ(t) é absolutamente integrável, todas as leis estáveis são absolutamente contínuas.

Por outro lado, o parâmetro de assimetria β "compara" o peso da cauda direita com o peso da cauda esquerda, uma vez que se tem

$$\lim_{x\to+\infty}\frac{1-F(x;\alpha,\beta)}{1-F(x;\alpha,\beta)+F(-x;\alpha,\beta)}=\frac{k_2}{k_1+k_2}$$

$$\lim_{x\to+\infty}\frac{F(-x;\alpha,\beta)}{1-F(x;\alpha,\beta)+F(-x;\alpha,\beta)}=\frac{k_1}{k_1+k_2} \qquad (2.8)$$

$$tendo-se \quad \beta=\frac{k_2-k_1}{k_1+k_2}$$

E quando $\alpha\neq 1$, tem-se ainda a seguinte relação [6], p. 79.

$$F_X(0;\alpha,\beta_B)=\frac{1}{2}\left(1-\frac{K(\alpha)}{\alpha}\beta_B\right) \qquad (2.9)$$

isto é, $F_X(0;\alpha,\beta)$ decresce com β tanto mais rapidamente quanto menor for α.

Ou ainda atendendo a (2.6):

$$F_X(0;\alpha,\beta_A)=\frac{1}{2}-\frac{\arctan\left(\beta_A\tan\left(\frac{\pi\alpha}{2}\right)\right)}{\pi\alpha} \qquad (2.10)$$

Também para α=1 se pode verificar que $F_X(0;\alpha,\beta)$ decresce com β [1]. Podemos pois concluir que o parâmetro de assimetria está relacionado com as caudas da função distribuição e com o valor desta no ponto x=0.

Os gráficos das f.d.p. $f_X(x;\alpha,\beta)$ de Holt e Crow [2] e das f.d. $F_X(x;\alpha,\beta)$ de Graça Martins [1] ilustram as afirmações feitas e mostram que quando α está próximo de 2 a influência de β é quase imperceptível, facto este que também pode ser constatado em Nolan [5], pág.25.

A função **stableSlider** da biblioteca *Rmetrics* do software R 2.6.2, também permite visualizar este comportamento.

3. UM ESTIMADOR DO PARÂMETRO DE ASSIMETRIA

Designemos por $S(\alpha,\beta,a,c)$ uma v.a. estável de parâmetros (α,β,a,c), $\alpha \in (0,2],\ -1 \le \beta \le 1,\ a \in \mathbb{R}, c > 0$, todos conhecidos excepto o parâmetro de assimetria β. Recordemos que

$$F(0;\alpha,\beta_A) = \frac{1}{2} - \frac{\arctan\left(\beta_A \tan\left(\frac{\pi\alpha}{2}\right)\right)}{\pi\alpha} \tag{3.1}$$

E resolvendo em ordem a β_A obtém-se:

$$\beta_A = \frac{\tan\left(\pi\alpha\left(\frac{1}{2} - F_X(0;\alpha,\beta_A)\right)\right)}{\tan\left(\frac{\pi\alpha}{2}\right)} \tag{3.2}$$

A partir da relação anterior podemos obter um estimador do parâmetro β supondo α conhecido. Seja $(X_1,\ldots X_n)$ uma amostra aleatória de uma população estável padrão, i.e., $S(\alpha,\beta,0,1)$ ($a = 0, c = 1$, sem perda de generalidade)

$$\beta_n^* = \frac{\tan\left(\pi\alpha\left(\frac{1}{2} - F_n^*(0;\alpha,\beta_A)\right)\right)}{\tan\left(\frac{\pi\alpha}{2}\right)} \tag{3.3}$$

onde $F_n^*(x)$ é a função de distribuição empírica de uma amostra aleatória $(X_1,\ldots X_n)$.

A função `rstable(n,alpha,beta,gamma=1,delta=0,pm=c(0,1,2))` da biblioteca *Rmetrics* permite gerar amostras pseudo-aleatórias de distribuições estáveis, sendo gamma o parâmetro de escala *c*, delta o parâmetro de localização *a* e *pm* o parâmetro que indica a parametrização utilizada. No presente trabalho usámos a parametrização *pm=1*, que é baseada na representação (A) e para a qual se tem a igualdade (3.1). A título de exemplo apresentamos o programa utilizado para a obtenção de *m* valores de β_n^* a partir da geração de amostras de uma estável $S(0.9,0.5,0,1)$.

```
n<-500
m<-100
alpha<-0.9
beta<-0.5
s<-matrix(0,nrow=n,ncol=m)
freq<-matrix(0,nrow=1,ncol=m)
F0_emp<-matrix(0,nrow=1,ncol=m)
estbeta<-matrix(0,nrow=1,ncol=m)
set.seed(011)
```

```
for(j in 1:m)
{s[,j]<-rstable(n,alpha,beta,1,0,1)
freq[,j]<-length(subset(s[,j],s[,j]<=0))
F0_emp[,j]<-freq[,j]/n  # f.d. empírica no ponto 0
estbeta[,j]<-tan(alpha*pi*(0.5- F0_emp[,j]))/tan(pi*alpha/2)}
F0_emp
estbeta
round(mean(estbeta),4)
```

[1] 0.5046

```
var<-matrix(0,nrow=1,ncol=m)
for(j in 1:m)
{var[,j]<-((estbeta[,j]-mean(estbeta)))^2/(m-1)}
round(sqrt(sum(var)),4)
```

[1] 0.0472

Nas tabelas seguintes registaram-se os valores das médias e variâncias empíricas das estimativas obtidas. Excluem-se os casos $\alpha<1$, $\beta=1$ uma vez que correspondem às distribuições estáveis positivas.

Tabela 1**:** *Média e (variância) de m=100 observações de* β_n^*, *n=500*

	$\beta=0.2$	$\beta=0.3$	$\beta=0.5$	$\beta=0.8$	$\beta=1.0$
$\alpha=0.1$	0.2017	0.3034	0.4993	0.7998	
	(0.0414)	*(0.0428)*	*(0.0379)*	*(0.0293)*	
$\alpha=0.2$	0.1993	0.2926	0.4989	0.7996	
	(0.0414)	*(0.0424)*	*(0.0377)*	*(0.0304)*	
$\alpha=0.5$	0.2043	0.3001	0.4992	0.8010	
	(0.0354)	*(0.0342)*	*(0.0309)*	*(0.0267)*	
$\alpha=0.8$	0.1961	0.3011	0.4987	0.8027	
	(0.0223)	*(0.0292)*	*(0.0405)*	*(0.0429)*	
$\alpha=0.9$	0.2027	0.2983	0.5046	0.8082	
	(0.0218)	*(0.0305)*	*(0.0472)*	*(0.0718)*	
$\alpha=1.1$	0.203	0.3101	0.5186	0.8681	1.1519
	(0.0284)	*(0.0514)*	*(0.1072)*	*(0.3076)*	*(0.4713)*
$\alpha=1.2$	0.2021	0.3043	0.5067	0.8416	1.0189
	(0.0357)	*(0.0483)*	*(0.0734)*	*(0.167)*	*(0.2442)*
$\alpha=1.5$	0.1978	0.3248	0.4993	0.8261	1.0162
	(0.1134)	*(0.1229)*	*(0.130)*	*(0.1762)*	*(0.2226)*
$\alpha=1.8$	0.2113	0.3426	0.493	0.7678	1.0005
	(0.360)	*(0.454)*	*(0.381)*	*(0.4328)*	*(0.3973)*
$\alpha=1.9$	0.2239	0.1408	0.5723	0.7305	1.0006
	(0.845)	*(0.9539)*	*(0.7909)*	*(0.9151)*	*(0.8833)*

Considerando agora amostras de dimensão *n=1000* de estáveis de expoente característico α perto de 2, calculámos m =100 valores de β_n^*. As respectivas médias e variâncias destas estimativas estão registadas na tabela seguinte.

Tabela 2: *Média e (variância) de m observações de* β_n^*

	β=0.2	β=0.3	β=0.5	β=0.8	β=1.0
α=1.8	0.1861	0.2999	0.4951	0.7811	1.0429
	(0.2731)	*(0.2704)*	*(0.2927)*	*(0.3163)*	*(0.2878)*
α=1.9	0.3000	0.2855	0.4798	0.8830	1.0348
	(0.6354)	*(0.6017)*	*(0.564)*	*(0.5621)*	*(0.6489)*

4. Conclusão

Pela observação das tabelas anteriores concluímos que o estimador β_n^* é mais preciso quando α<1, vindo o seu desempenho a piorar à medida que α se aproxima de 2, pois como já tínhamos referido anteriormente, a influência de β nestes casos é quase imperceptível.

5. Agradecimentos

O autor agradece ao Professor John P. Nolan os seus valiosos esclarecimentos relativamente à parametrização a utilizar na função *rstable*.

Referências

[1] Graça Martins, E. (1983). *Modelos Estáveis*, vol. I e II. Dissertação de Doutoramento. DEIOC, FCUL.

[2] Holt, D. R. and Crow, E. L. (1973). Tables and graphs of the stable probability density functions. *J. Res. Nat. Bur. Standards*, 77B, 143-198.

[3] Iglésias Pereira, H. (1985). *Escolha Estatística em Modelos Estáveis*. Tese de Doutoramento. DEIOC, FCUL.

[4] Nolan, John P. (1998). Parameterizations and models of stable distributions. *Statistics and Probability Letters*, 38, p. 187-195.

[5] Nolan, John P. (2008). *Stable Distributions- Models for Heavy Tailed Data*, Capítulo 1, http://academic2.american.edu/~jpnolan/stable/stable.html.

[6] Zolotarev, V.M. (1986). One-dimensional Stable Distributions. *Am. Math. Society*. Providence, R.I.