



A Randomness Test for Stable Data

Adel Mohammadpour^{†,*}, Ali Mohammad-Djafari[‡],
and John P. Nolan^{*}

[†] Amirkabir University of Technology and
Statistical Research and Training Center, Iran

[‡] CNRS-Supélec-University Paris 11

^{*} American University

Abstract. In this paper, we propose a new method for checking randomness of non-Gaussian stable data based on a characterization result. This method is more sensitive with respect to non-random data compared to the well-known non-parametric randomness tests.

Keywords. stable distributions; randomness tests (test for i.i.d.); characterization.

1 Introduction

Let x_1, \dots, x_d be observed values of the univariate random variables X_1, \dots, X_d . We say that the observations are random if the X_i 's ($i = 1, \dots, d$) are independent and identically distributed (i.i.d.) random variables. There are several non-parametric ways for checking the randomness of a data set, i.e., testing the following hypotheses (e.g. Brockwell and Davis, 1996; Knuth, 1981):

$$\begin{cases} H_0 : X_i \text{'s are i.i.d.} \\ H_1 : \text{Negation of } H_0 \end{cases} \quad (1)$$

* Corresponding author

These randomness tests are neither powerful nor sensitive with respect to i.i.d. assumption, usually useful for rejecting the null hypothesis and each one of them revealing a particular type of imperfection in a data set. In the Gaussian case, there exists some parametric tests based on characterization results (Kagan et al., 1973). For stable distributions we can use characterization results for constructing more sensitive tests with respect to non-random data. To the authors knowledge, for the non-Gaussian stable case there is no such tests.

In this paper, we first present a new property of spectral measure of i.i.d. stable random variables which combines two previously established properties of spectral measure of independent and permutation-symmetric stable random variables (Mohammadpour and Soltani, 2000). Then we propose a method for checking randomness of univariate stable data based on this characterization property. In the following section we recall the definition of stable random variables and vectors. In Section 3 we present this new characterization result for i.i.d. non-Gaussian stable random variables and its geometric interpretation and then use this result for proposing a method for checking randomness of univariate stable data.

2 Stable Distributions

A random variable (r.v.) X is said to be stable if there are parameters $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma > 0$, and $\delta \in \mathbb{R}$ such that, its characteristic function (c.f.) has the form

$$\phi_X(t) = \begin{cases} \exp \left\{ -\gamma^\alpha |t|^\alpha \left(1 - i\beta \operatorname{sgn}(t) \tan\left(\frac{\pi\alpha}{2}\right) \right) + i\delta t \right\}, & \alpha \neq 1, \\ \exp \left\{ -\gamma |t| \left(1 + i\beta \frac{2}{\pi} \operatorname{sgn}(t) \ln |t| \right) + i\delta t \right\}, & \alpha = 1, \end{cases} \quad (2)$$

where $\operatorname{sgn}(u) = -1, 0$, or 1 if $u <, =$, or > 0 , respectively. We will denote stable distributions by $S(\alpha, \beta, \gamma, \delta)$ and write

$$X \sim S(\alpha, \beta, \gamma, \delta)$$

to indicate that X has the stable distribution with the characteristic index (exponent) α , skewness β , scale γ , and location δ (Samorodnitsky and Taqqu, 1994; Nolan, 1998). When $\gamma = 1$ and $\delta = 0$ the distribution is called standard stable and will be denoted by $S(\alpha, \beta)$.

A random vector (r.v.), $\mathbf{X} = (X_1, \dots, X_d)'$, is said to be stable or α -stable (α -S), in \mathbb{R}^d , if there are parameters $\alpha \in (0, 2]$, $\Delta = (\delta_1, \dots, \delta_d)' \in \mathbb{R}^d$, positive definite matrix A of order d , and a finite measure Γ on the unit sphere $\mathbb{S}_d =$

$\{\mathbf{s} = (s_1, \dots, s_d)' | \mathbf{s}'\mathbf{s} = 1\}$ of \mathbb{R}^d such that

$$\phi_{\mathbf{X}-\Delta}(\mathbf{t}) = \begin{cases} \exp \left\{ - \int_{\mathbb{S}_d} |\mathbf{t}'\mathbf{s}|^\alpha \left(1 - i \operatorname{sgn}(\mathbf{t}'\mathbf{s}) \tan\left(\frac{\pi\alpha}{2}\right) \right) \Gamma(d\mathbf{s}) \right\}, & \alpha \neq 1, \\ \exp \left\{ - \int_{\mathbb{S}_d} |\mathbf{t}'\mathbf{s}| \left(1 + i \frac{2}{\pi} \operatorname{sgn}(\mathbf{t}'\mathbf{s}) \ln |\mathbf{t}'\mathbf{s}| \right) \Gamma(d\mathbf{s}) \right\}, & \alpha = 1, \\ \exp \{-\mathbf{t}'\mathbf{A}\mathbf{t}\}, & \alpha = 2, \end{cases} \quad (3)$$

where $\mathbf{t} = (t_1, \dots, t_d)'$.

When Γ is a discrete spectral measure with a finite number of point masses, i.e.,

$$\Gamma(\cdot) = \sum_{k=1}^n \gamma_k \delta_{\mathbf{s}_k}(\cdot),$$

where γ_k 's are the weights, and $\delta_{\mathbf{s}_k}$'s are point masses (Dirac measures of mass 1) at the points $\mathbf{s}_k = (s_{1k}, \dots, s_{dk})' \in \mathbb{S}_d, k = 1, \dots, m$, (3) becomes

$$\phi_{\mathbf{X}-\Delta}(\mathbf{t}) = \begin{cases} \exp \left\{ - \sum_{k=1}^n |\mathbf{t}'\mathbf{s}_k|^\alpha \left(1 - i \operatorname{sgn}(\mathbf{t}'\mathbf{s}_k) \tan\left(\frac{\pi\alpha}{2}\right) \right) \gamma_k \right\}, & \alpha \neq 1, \\ \exp \left\{ - \sum_{k=1}^n |\mathbf{t}'\mathbf{s}_k| \left(1 + i \frac{2}{\pi} \operatorname{sgn}(\mathbf{t}'\mathbf{s}_k) \ln |\mathbf{t}'\mathbf{s}_k| \right) \gamma_k \right\}, & \alpha = 1. \end{cases} \quad (4)$$

If \mathbf{X} is symmetric around Δ , symmetric α -stable (S α S), then equivalently Γ is symmetric on \mathbb{S}_d , and (3) reduces to $\exp \left\{ - \int_{\mathbb{S}_d} |\mathbf{t}'\mathbf{s}|^\alpha \Gamma(d\mathbf{s}) \right\}, \alpha \neq 2$.

Samorodnitsky and Taqqu (1994) is our reference for basic properties of stable r.v.s. Some applications of stable r.v.s and r.v.s are given by Nikias and Shao (1995) and Nolan (2007).

3 Randomness Test

3.1 A Characterization

In this section we introduce a characterization result and use this result for proposing a graphical method for testing randomness of data.

Lemma 1 *Let $\alpha \in (0, 2)$. A non-degenerate α -S r.v. \mathbf{X} , with c.f. (3), has independent components if and only if the spectral measure Γ of \mathbf{X} is concentrated on the intersection of the axes with the sphere \mathbb{S}_d .*

Proof: See, e.g., Samorodnitsky and Taqqu (1994), p. 68.

A r.v. \mathbf{X} is said to be permutation-symmetric (or finitely exchangeable) if for every one of the $d!$ permutations (π_1, \dots, π_d) of $(1, \dots, d)$, \mathbf{X} and $\pi\mathbf{X} = (X_{\pi_1}, \dots, X_{\pi_d})'$ have the same distribution.

Lemma 2 *Let \mathbf{X} be a d -dimensional α -S r.v., $\alpha \in (0, 2)$, with spectral measure Γ and location parameter Δ . Then \mathbf{X} is permutation-symmetric if and only if Δ has the same components, and $\Gamma(E) = \Gamma(\pi E)$ for every permutation π and Borel set $E \subset \mathbb{S}_d$, where $\pi E = \{\mathbf{s} : \pi^{-1}\mathbf{s} \in E\}$, and $\pi\mathbf{s} = (s_{\pi^{-1}1}, \dots, s_{\pi^{-1}d})'$. We call such a spectral measure a permutation-symmetric spectral measure.*

Proof: See Mohammadpour and Soltani (2000).

Lemma 3 *Let $\alpha \in (0, 2]$. A d -dimensional α -S r.v. \mathbf{X} , with c.f. (3), has i.i.d. components if and only if the spectral measure \mathbf{X} is permutation-symmetric and concentrated on the intersection of the axes with the sphere \mathbb{S}_d , and the location parameter Δ has the same components.*

Proof: Suppose r.v. \mathbf{X} has i.i.d. components. Then

$$\phi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^d \phi_{X_k}(t_k) \quad \text{and} \quad \phi_{X_k}(t) = \phi_{X_l}(t), \quad k, l \in \{1, \dots, d\},$$

and so, for each permutation $\pi = (\pi_1, \dots, \pi_d)$,

$$\phi_{\mathbf{X}}(\mathbf{t}) = \prod_{k=1}^d \phi_{X_k}(t_k) = \prod_{k=1}^d \phi_{X_{\pi_k}}(t_k) = \phi_{\pi\mathbf{X}}(\mathbf{t}).$$

Therefore, Γ is concentrated on the intersection of the axes with the sphere \mathbb{S}_d (by Lemma 1), and Γ is permutation-symmetric and the location parameter Δ has the same components (by Lemma 2). By noting that a permutation-symmetric r.v. has identical components, the converse of lemma follows by Lemmas 1 and 2.

3.2 Proposed Method

In bivariate case we can give a geometric interpretation for the spectral measure in Lemma 3. Without loss of generality, we assume that location parameter is zero. Spectral measure of a bivariate stable distribution is concentrated on the unite circle, \mathbb{S}_2 , and so the spectral measure of a bivariate stable distribution with independent components is concentrated on the points in the set

$$\{(1, 0)', (0, 1)', (-1, 0)', (0, -1)'\}. \quad (5)$$

On the other hand, by Lemma 3, spectral measure of a bivariate permutation-symmetric stable distribution is symmetric about the line $x = y$, i.e., $s_1 = s_2$. Therefore, spectral measure of a bivariate stable distribution with i.i.d. components is symmetric about the line $x = y$, and so, is concentrated on 4 or 2 points in the set (5). Figure 1 shows all possible cases for spectral measures of a bivariate stable distributions with i.i.d. components. Each “•” shows a point mass of spectral measure.

Only the first graph, (i), of the spectral measures, is symmetric with respect to (w.r.t.) the origin, and so the spectral measure of a bivariate SaS distribution with i.i.d. components, has one possible form. On the other hand, the graphs (iii) and (v), of the spectral measures, can be obtained by rotation from (ii) and (iv) respectively. Therefore, the spectral measure of a bivariate non-SaS distribution with i.i.d. components, has two basic forms, i.e., graphs (ii) and (iv). Based on these two properties of spectral measure of a bivariate SaS and non-SaS distributions with i.i.d. components, we propose a method for testing (1).

We assume that the number of data is even and propose the following method.

1. Estimate parameters of univariate stable distribution, called $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\delta}$, from the data. Under H_0 , in (1), the observations are i.i.d. and there are several ways for estimating these parameters, e.g. Nolan (2001). Without loss of generality, we assume that $\hat{\delta} = 0$ in the following steps.
2. Randomly pair the data and plot their bivariate empirical density.
3. Compute parameters of spectral measure of the bivariate stable distribution with independent components, $\{\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4\}$, based on estimated parameters in step 1. The relations are given in Table 1.

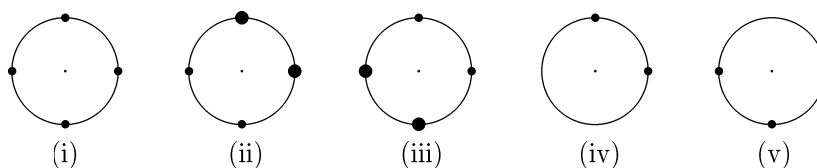


Figure 1. Spectral measures of bivariate stable distributions with i.i.d. components

Table 1. Relation between the estimated parameters of i.i.d. stable r.v. and the spectral measure of a bivariate distribution with i.i.d. components. These relations are obtained using the c.f.s (2) and (4) and the uniqueness of the spectral measure

$\Gamma = \gamma_1 \delta_{(1,0)'} + \gamma_2 \delta_{(0,1)'} + \gamma_3 \delta_{(-1,0)'} + \gamma_4 \delta_{(0,-1)'}$	
Spectral Measures Types (in Figure 1) Based on $\hat{\beta}$	Estimated Weights Based on $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$
$\hat{\beta} = 0 \longrightarrow$ (i)	$\hat{\gamma}_1 = \hat{\gamma}_2 = \hat{\gamma}_3 = \hat{\gamma}_4 = \hat{\gamma}^{\hat{\alpha}}/2$
$\hat{\beta} > 0 \longrightarrow$ (ii)	$\hat{\gamma}_1 = \hat{\gamma}_2 = \hat{\gamma}^{\hat{\alpha}}(1 + \hat{\beta})/2, \hat{\gamma}_3 = \hat{\gamma}_4 = \hat{\gamma}^{\hat{\alpha}}(1 - \hat{\beta})/2$
$\hat{\beta} < 0 \longrightarrow$ (iii)	$\hat{\gamma}_1 = \hat{\gamma}_2 = \hat{\gamma}^{\hat{\alpha}}(1 + \hat{\beta})/2, \hat{\gamma}_3 = \hat{\gamma}_4 = \hat{\gamma}^{\hat{\alpha}}(1 - \hat{\beta})/2$
$\hat{\beta} = 1 \longrightarrow$ (iv)	$\hat{\gamma}_1 = \hat{\gamma}_2 = \hat{\gamma}^{\hat{\alpha}}, \hat{\gamma}_3 = \hat{\gamma}_4 = 0$
$\hat{\beta} = -1 \longrightarrow$ (v)	$\hat{\gamma}_1 = \hat{\gamma}_2 = 0, \hat{\gamma}_3 = \hat{\gamma}_4 = \hat{\gamma}^{\hat{\alpha}}$

4. Calculate and plot the theoretical bivariate stable density with i.i.d. components using the spectral measure obtained in step 3. We can use the method in Nolan (1997).
5. Compare the produced density in steps 2 and 4 either visually or using one of the goodness of fit criteria such as Pearson's χ^2 test statistic (Bishop et al., 1975) or Kullback-Leibler divergence (Kullback and Leibler, 1951).
6. Reject H_0 if two densities have not the same structures or if one of the selected criteria is larger than a threshold.

3.3 Numerical Experiments

Figure 2-left shows the contours of bivariate stable densities with independent components for $\alpha = 0.7, 1, 1.3$, and $\Delta = \mathbf{0}$. The first row shows bivariate SaS with i.i.d. components (spectral measure (i) in Figure 1), and the two other rows show non-SaS with i.i.d. components with 4 and 2 point masses (with spectral measures (ii) and (iv) in Figure 1) respectively. Figure 2-right shows a set of empirical density contours of non i.i.d. bivariate stable distributions for $\alpha = 1.3$ and $\Delta = \mathbf{0}$. The number of point masses is 4 for the first column, 3 for the second column, and 2 for the last column. The first line shows the graphs of independent non-identically distributed (i.n.i.d.) components, the second row shows non-independent i.d. (ni.i.d.) components, and the last row shows ni.n.i.d. components. Each graph is generated by simulating 10^6 bivariate stable r.v. with the same parameters. The location of point masses in degrees and their corresponding weights are denoted under each graph.

One can distinguish bivariate stable distribution with ni. and/or ni.d. components in Figure 2-right from stable distribution with i.i.d. components (last column of Figure 2-left). From these graphs we can see that, except the case of the two last graphs in the first row of Figure 2, it is easy to visually compare the empirical and theoretical graphs to make a decision for rejecting H_0 . In the cases that we do not have a large set of data and/or it is difficult to reject H_0 by comparing these graphs visually, we can use an appropriate test statistic for testing (1). To make our point clear we give the following example.

Example 1. Let $X_i = (1 - \tau)Z_i + \tau Z_0$; $Z_0, Z_i \sim S(1, 0)$ (standard S1S distribution), $i = 1, \dots, 600$ and $\tau \in [0, 1)$. It can be shown that X_i 's are identically distributed, and (X_i, X_j) , $i \neq j$, has a spectral measure plotted in Figure 3.

To test the sensitivity of the proposed method, we generate different sets of data from this distribution with different values of τ .

When $\tau = 0.1$ and the number of data is 600 we cannot use the proposed graphical method. Instead of plotting the empirical and theoretical densities in steps 2 and 4 in the proposed method, we construct a 2-dimensional

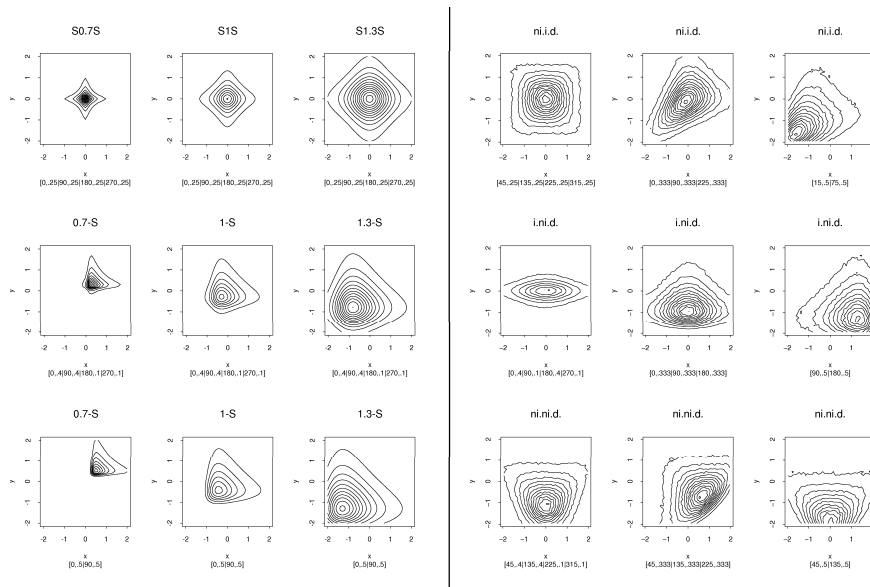


Figure 2. Left: Density contour of bivariate α -S distributions, with i.i.d. components. Right: Empirical density contour of bivariate 1.3-S distributions, with ni.i.d., i.i.d., and ni.ni.d. components

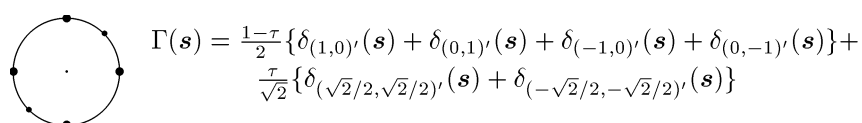


Figure 3. Spectral measure of the exchangeable stable random variables in Example 1

Table 2. Left: 2-dimensional frequency table for 300 pairs of observations in Example 1 for $\tau = 0.1$; Right: corresponding expected values

Number of observed values					Number of expected values				
	(-2,-1]	(-1,0]	(0,1]	(1,2]		(-2,-1]	(-1,0]	(0,1]	(1,2]
(-2,-1]	8	19	11	4	(-2,-1]	3.09	7.75	7.75	3.09
(-1,0]	11	25	13	3	(-1,0]	7.87	21.01	21.01	7.87
(0,1]	8	31	10	5	(0,1]	7.75	19.47	19.47	7.75
(1,2]	1	6	6	3	(1,2]	3.09	7.75	7.75	3.09
Total = 164					Total = 152.23				

frequency or contingency table (empirical density) for observations (Table 2-left) and calculate the expected number for each cell based on estimated distribution function corresponding to the estimated parameters in steps 1 and 3 (Table 2-right). The statistic $\chi^2 = 48.40$ for this data and we reject H_0 at the significance level 0.05 (95th percentile of χ^2 distribution with 9 degrees of freedom is $\chi_{0.95,9}^2 = 21.67$). On the other hand the Z statistic of the runs test (test of randomness, e.g., Gibbons, 1971), is $|z| = 0.41$ and we cannot reject H_0 at significance level 0.05, ($z_{0.975} = 1.96$).

In Figure 4 we plot the empirical mean of $\ln(\chi^2)$ and $3.5|z|$ w.r.t. $\tau \in [0, 1)$.

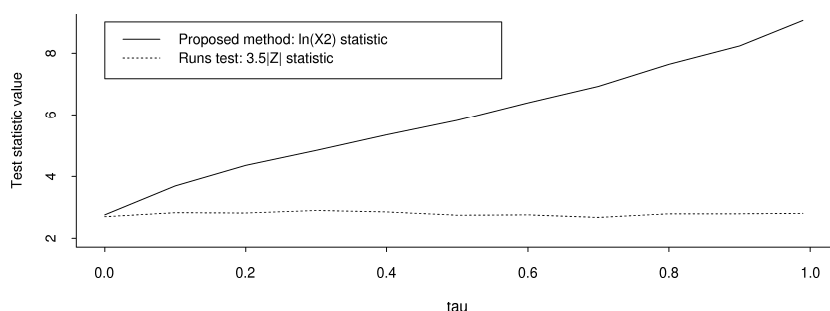


Figure 4. Sensitiveness of proposed method and runs test w.r.t. ni.i.d. data

We see that, the proposed method is more sensitive with increasing τ . This is likely a general result: the test is likely to be more sensitive when the spectral measure is far away from the independent case.

Example 2. In this example we consider a network data set which have been introduced in Leland et al. (1994). The data set contains a million packet arrivals. It has been measured in external Ethernet traffic. Xiaohu et al. (2004) show that some subsets of this data set have stable distribution. We would like to test the following hypotheses.

$$\begin{cases} H_0 : \text{Observations are i.i.d. stable} \\ H_1 : \text{Negation of } H_0 \end{cases}$$

We test the hypotheses in the following steps.

1. Take a random sample of size 1,000,000 from the data set;
2. Subtract the median of random sample from each observation;
3. Paired observations i and $i + 500000$, $i = 1, \dots, 500000$, and plot bivariate empirical density;
4. Calculate maximum likelihood estimator of parameters under H_0 , generate a random sample of size 1,000,000 with these parameters, paired the generated data, and plot bivariate empirical density.

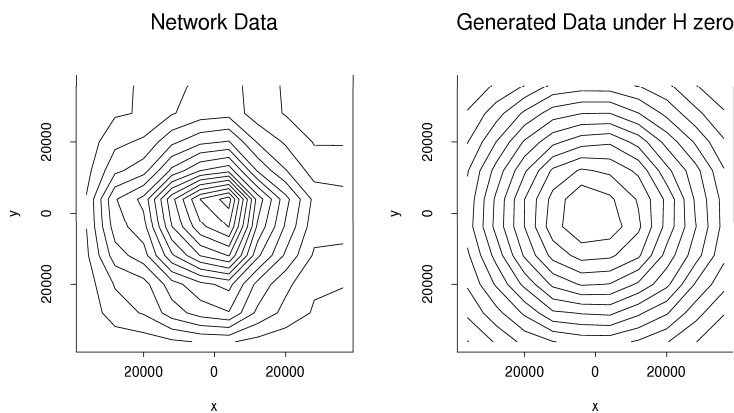


Figure 5. Contours of empirical density of network data and its corresponding generated empirical density based on H_0

Contours of empirical density of network data and its corresponding generated empirical density based on H_0 are plotted in Figure 5. It is easy to see that they are not similar and we can reject H_0 .

4 Conclusion

In this work, we first presented a new property of spectral measure of i.i.d. stable r.v.s, which combines two previously established properties of spectral measure of independent and permutation-symmetric stable r.v.s. Then we presented a randomness test for univariate stable data, which is based on this particular property. The main test process is to compare the empirical distribution of pairwised data to the theoretical bivariate stable distribution with i.i.d. components. This comparison is made either visually just by comparing the structures of these two distributions or by computing a goodness of fit criterion between them. We showed the performance of this new test via some graphs of i.i.d. and non i.i.d. bivariate densities and an example of simulated univariate data which compares the sensitiveness of the proposed test to the classical non-parametric runs test. Finally we used the suggested method for testing randomness of ethernet traffic data, and rejected this hypothesis.

Acknowledgment

The authors would like to thank a referee for his helpful suggestion for adding a real example. The first author is also grateful to Supélec and Statistical Research and Training Center of Iran, Grant No. 8601, for their supports.

References

- Bishop, Y.M.M.; Fienberg, S.E.; Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA.
- Brockwell, P.J.; Davis, R.A. (1996). *Time Series: Theory and Methods*. Springer, New York.
- Gibbons, J.D. (1971). *Nonparametric Statistical Inference*. McGraw-Hill, New York.
- Kagan, A.M.; Linnik, Y.V.; Rao, C.R. (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.
- Knuth, D. (1981). *The Art of Computer Programming*, vol. 2, 2nd ed. Addison-Wesley, Reading, MA.

- Kullback, S.; Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22** 79-86.
- Leland, W.E.; Taqqu, M.S.; Willinger, W.; Wilson, D.V. (1994). On the selfsimilar nature of ethernet traffic (extended version). *IEEE Trans. Networking* **2**, 1-15.
- Mohammadpour, A.; Soltani, A.R. (2000). Exchangeable stable random vectors and their simulations. *Comput. Statist.* **15**, 205-217.
- Nikias C.L.; Shao, M. (1995). *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley, New York.
- Nolan, J.P. (1997). Numerical calculation of stable densities and distribution functions: Heavy tails and highly volatile phenomena. *Comm. Statist. Stoch. Models* **13**, 759-774.
- Nolan, J.P. (1998). Parameterizations and modes of stable distributions. *Statist. Probab. Lett.* **38**, 187-195.
- Nolan, J.P. (2001). Maximum likelihood estimation and diagnostics for stable distributions. In *Lévy Processes: Theory and Applications*, O.E. Barndorff-Nielsen, T. Mikosch, and S.I. Resnick, eds., Birkhäuser, Boston, pp. 379-400.
- Nolan, J.P. (2007). *Stable Distributions: Models for Heavy Tailed Data*. In progress, Chapter 1 online at academic2.american.edu/~jpnolan.
- Samorodnitsky, G.; Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York.
- Xiaohu, G.; Guangxi, Z.; Yaoting, Z. (2004). On the testing for alpha-stable distributions of network traffic. *Comput. Comm.* **27**, 447-457.

Adel Mohammadpour

Department of Statistics,
Faculty of Mathematics and
Computer Sciences,
Amirkabir University of Technology
(Tehran Polytechnic),
424 Hafez Ave.,
Tehran, Iran.
e-mail: adel@aut.ac.ir

Ali Mohammad-Djafari

Laboratoire des signaux et systèmes (L2S),
Supélec-3 rue Joliot-Curie,
91192 Gif-sur-Yvette cedex,
France.
e-mail: djafari@lss.supelec.fr

John P. Nolan

Mathematics and Statistics Department,
American University,
4400 Massachusetts Avenue,
NW Washington, DC 20016,
USA.
e-mail: jpnolan@american.edu

