A Generator of Cauchy-distributed Time Series with Specific Hurst Index

Leopoldo Estrada-Vargas
Centro de Investigación y de Estudios Avanzados - CINVESTAV
Guadalajara, Jalisco, México
lestrada@gdl.cinvestav.mx

Homero Toral-Cruz
Universidad de Quintana Roo
Chetumal, Quintana Roo, México
Instituto Tecnológico Superior de Las Choapas
Las Choapas, Veracruz, México
htoral@uqroo.mx

Al-Sakib Khan Pathan
Department of Computer Science
International Islamic University Malaysia - IIUM
Kuala Lumpur, Malaysia
sakib@iium.edu.my

Abstract

A generator of artificial Cauchy-distributed time series is presented. This generator transforms any random time series, e.g., standardized fractional Gaussian noise (FGN), into a Cauchy-distributed series with specific location and scale parameters and correlation structure, determined by the Hurst index. The proposed algorithm consists of an inverse cumulative distribution function (ICDF) transformation, a wavelet-analysis synthesis and, finally, a linear transformation. The resulting Cauchy-distributed series has approximately the desired location and scale parameters and exactly the desired Hurst index. The performance of the proposed generator is evaluated by estimating the location, scale and Hurst parameters from artificial time series and by calculating the mean squared error (MSE) of their cumulative distribution function (CDF). The input location, scale and Hurst index used in the simulations are taken from jitter samples of monitored Voice over Internet Protocol (VoIP) calls, which have been proved to be adequately modeled with these processes under some circumstances.

This work was supported by CONACYT grant 166532 and PROMEP - UQROO-PTC-110.

1. Introduction

The analysis and statistical study of time series, that represent network characteristics, are often used for the design of communications systems. These studies are used for the design and testing of communication improvements. Also, the Quality of Service (QoS) of network communication can be estimated or predicted from these measurements [1].

There are two main issues that must be conducted in order to obtain data samples: 1) configuration of the measurement setting, e.g., phones, gatekeepers, traffic monitors and network interfaces, and 2) the design and realization of the measurement protocols. In addtion to these works, there is also certain amount of time necessary to capture a data sample of certain size, e.g., the duration of an Internet call. Artificial data generators are then used in order to gather a large volume of data without conducting the mentioned issues and saving a lot of effort and time. The artificial time series must produce time series that are representative, in the statistical sense, of the characteristics of the communication systems, e.g., their distribution or correlation structure.

In this work, we propose a method to simulate Cauchy-type processes that represent the delay jitter (or jitter) series of a Voice over Internet Protocol (VoIP) call. This generator produces a random sequence of Cauchy-distributed observations whose correlation is determined by the self-similarity parameter, namely the Hurst index (H). This proposed method demonstrates that it is possible to synthesize artificial time series with both a specific distribution and Hurst index, even in the infinite variance case (a study of the finite-variance case is presented in [2]).

The rest of this paper is structured as follows: The mathematical background is presented in Section 2. A proposed wavelet-based synthesis of self-similar time series is described in Section 3. An H.323 zone and how jitter measurements are obtained are described in Section 4. The proposed generator is described in Section 5 and the evaluation of its performance is summarized in section 6. The main findings and conclusions are presented in Section 7.

2. Mathematical Backgrounds and Preliminaries

A. The Inverse Cumulative Distribution Function Transformation

The inverse cumulative distribution function (ICDF) transformation produces a time series Y_t with cumulative distribution function (CDF) $F_Y(y)$ from a random time series X_t with CDF $F_X(x)$ by applying the following sample-to-sample formula:

$$y_t = F_Y^{-1}\big(F_X(x_t)\big). \tag{1}$$

If X_t is uniformly distributed between 0 and 1, as $F_X(x_t) = x_t$, then (1) simplifies to:

$$y_t = F_Y^{-1}(x_t). \tag{2}$$

B. The Gaussian and Cauchy Distributions

The CDF function of a Gaussian RV is defined by the mean (μ) and variance (σ^2) of the distribution:

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\lambda-\mu)^2}{2\sigma^2}} d\lambda. \tag{3}$$

In the absence of a closed form for (3), numerical approximations are used to estimate it.

A Cauchy RV has the probability density function (PDF):

$$f_X(x) = \frac{1}{\pi s\left[1+\left(\frac{x-c}{s}\right)^2\right]}, \tag{4}$$

where c, the location parameter, defines the location of the peak of the distribution and s, the scale parameter, specifies the half-width at half-maximum. A standardized Cauchy RV has $c = 0$ and $s = 1$. An approximated standardization for a Cauchy RV is:

$$X_{(0,1)} \sim s^{-1}\big(X_{(s,c)} - c\big). \tag{5}$$

The inverse of (5) produces a Cauchy distributed sample of certain location and scale parameters from a standardized Cauchy sample, i.e.,

$$X_{(c,s)} \sim sX_{(0,1)} + c. \tag{6}$$

Unlike a Gaussian RV, there exist closed forms for the CDF and ICDF of a Cauchy RV, which are:

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-c}{s}\right), \tag{7}$$

and

$$F_X^{-1}(p) = c + s \cdot tan\left[\pi\left(x - \frac{1}{2}\right)\right]. \tag{8}$$

C. Discrete Self-similarity

When considering discrete stochastic time series the definition of self-similarity is given in terms of the aggregated processes. From a discrete time series $\{X_t; t \in N\}$, others series can be obtained by aggregation. The aggregated time series is a sequence defined by (9):

$$X(m) = \left\{X_k^{(m)}; k \in N\right\}, \tag{9}$$

where each term $X_k^{(m)}$ is obtained as:

$$X_k^{(m)} = \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i \,; k \in N, \tag{10}$$

and where m represents the aggregation level. That is, each new time series is obtained by partitioning the original time series into non-overlapping blocks of size m and then averaging each block to obtain its respective values.

Let X_t be a covariance stationary discrete time series with mean $\mu_X = 0$, variance σ_X^2 and auto-covariance function (ACvF) $\gamma_X(k)$, and $X_k^{(m)}$ its aggregated series. Then it is said that X_t is exactly self-similar (H-SS) if (11) holds [3]:

$$X_k^{(m)} \sim m^{H-1} \cdot X_t, \tag{11}$$

where ~ means equality in distribution.

A stochastic process is second-order self-similar (H-SOSS) if the variance and covariance of the aggregated time series are defined, respectively, by (12) and (13):

$$var\left(X_k^{(m)}\right) = m^{2H-2} \cdot var(X_t), \tag{12}$$

$$\gamma_X^{(m)}(k) = \frac{\sigma_X^2}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right]. \tag{13}$$

Obviously, an H-ss process is also H-soss. Note that the correlation structure of an H-soss process is defined by the Hurst index.

D. Logscale Diagram of Self-similar Processes

The wavelet decomposition consists of a transformation of a signal X_t into a set of orthogonal components, i.e.,

$$X(t) = \sum_{j=i}^{J} \sum_{k=1}^{2^j} d_X(j,k)\psi_{j,k}(t), \tag{14}$$

where each function $\psi_{j,k}(t)$ is derived from a basis function $\psi_0(t)$, namely the *mother wavelet*, by scaling and displacement as follows:

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi_0(2^{-j}t - k), \tag{15}$$

and $d_X(j,k) = \langle X(t), \psi_{j,k}(t)\rangle$ is the k^{th} coefficient at scale j.

The statistic $S_2(j)$ defined as:

$$S_2(j) = E|d_X(j,\cdot)|^2, \tag{16}$$

calculated from a H-ss or, at least, H-soss process is related to H as follows [4]:

$$S_2(j) = c_f C 2^{j(2H-1)}, \tag{17}$$

where the quantity $c_f C$, related to the power of the process, is considered a constant.

The plot $log_2[S_2(j)]$ vs. j forms the widely known *Logscale Diagram* described by D. Veitch and P. Abry [4], and it is a straight line for H-soss, but for real-world or artificial time series, the estimation of the *Logscale Diagram* may differ from the ideal, linear model.

3. Wavelet-Based Synthesis of Self-Similar Time Series

We propose a method to synthesize H_1-SOSS from almost any time series regardless of whether it is or not self-similar or its marginal distribution. This method consists of adjust the sum expressed by (14) with a set of weights, i.e.,

$$X_{H_1}(t) = \sum_{j=i}^{J} w_j \sum_{k=1}^{2^j} d_X(j,k)\psi_{j,k}(t). \tag{18}$$

The weights w_j of (18) are defined as:

$$w_j = \sqrt{\frac{\widehat{c_f C} \cdot 2^{j(2H_1-1)}}{\hat{S}_2(j)}}, \tag{19}$$

where $\hat{S}_2(j)$ and $\widehat{c_f C}$ are the respective estimations of $S_2(j)$ and $c_f C$ (the associated power parameter [4]) from X_t and H_1 is the desired Hurst index of the new synthetic series. Even though this synthesis is independent of the original Hurst index and its only theoretical restriction is: $\hat{S}_2(j) > 0 \forall j = 1, \ldots, J$, the best results are obtained when the input signal (X_t) is already self-similar (e.g., FGN) and the desired Hurst index of the output signal is close to that of X_t. Pathological behavior can be produced in the output series for some critical conditions, e.g., noticeable steps, which may invalidate the hypothesis of stationarity, result of the transforming a SRD or uncorrelated input signal to a LRD output with H close to 1. Impulses of very large magnitude (outliers) can be also be produced when $\hat{S}_2(j)$ is close to zero.

This wavelet synthesis coincides exactly with the wavelet estimator proposed by D. Veitch and P. Abry [4], i.e., the estimated H of the synthesized series is unbiased ($E(\hat{H}) = H$) and has zero variance ($var(\hat{H}) = 0$).

4. Delay Jitter Measurements

The one way delay (OWD) of the k^{th} datagram is defined as the difference (variation) of its reception (R_k) and sending (S_k) times, i.e.,

$$O_k = R_k - S_k;\ k = 1, \ldots, N. \tag{20}$$

The jitter is defined as the difference of the OWD of consecutive datagrams, i.e.,

$$J_{k+1,k} = OWD_{k+1} - OWD_k; k = 1, \dots, N-1. \tag{21}$$

In order to obtain the jitter sequences, from which the statistical parameters are estimated, a number of VoIP calls was established within an $H.323$ zone, that consists of the endpoints $A1$, $A2$, $A3$ and $A4$ located in the local area network (LAN) A, the gatekeeper and the endpoints $B1$, $B2$, $B3$ and $B4$, both located in the LAN B. Each endpoint has an Alliance FXS PCI Voice Card developed at CTS CINVESTAV and a conventional cord phone. The measurements were monitored at LAN A using the network protocol analyzer *Wireshark* [5]. The measurement setting is shown in Figure 1.

From the measurement protocol which is more detailed in [6], [7], 96 VoIP test calls, each one with duration of 1-hour, were monitored. The captured RTP streams from these calls were processed with *Wireshark* and filtered with a script in order to obtain the jitter time series.

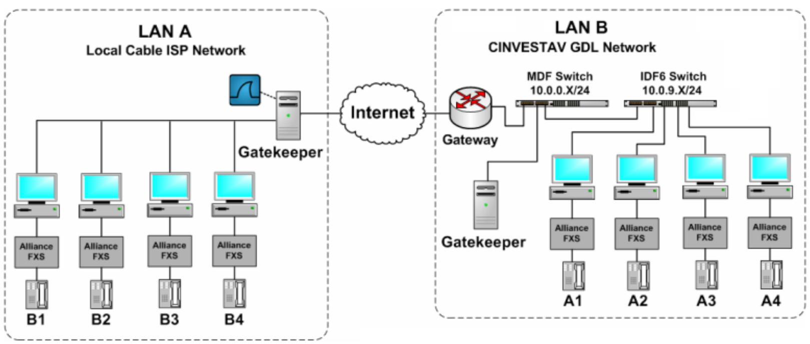

Figure 1: Measurement setting

5. Generator of Cauchy-type Time Series With Specific Hurst Index

The proposed algorithm for generating artificial Cauchy time series with defined size, location, scale and Hurst index consists of the following four stages:

1) Generate of a random sample of certain size, e.g., an uniformly distributed series or standardized FGN.

2) Convert the random sample into a Cauchy-distributed series, with a specific c and s, by means of the inverse CDF transformation, i.e., applying equations (3) and (8).

3) Adjust the Hurst index of the series by means of the weighted synthesis described in Section 3.

4) Adjust the location and scale parameters by means of the linear transformations (5) and (6), which do not alter the estimation of the Hurst index.

In Figure 2 shows the algorithm, where X_t is a measured jitter sample from which the location and scale parameters are estimated and Z_t is the output. Y_t is the resulting series of the ICDF transformation alone.

Note that the wavelet-based synthesis is not strictly a linear transformation (it is only for multiplication by a constant), consequently, the output of the generator (Z_t) may be not

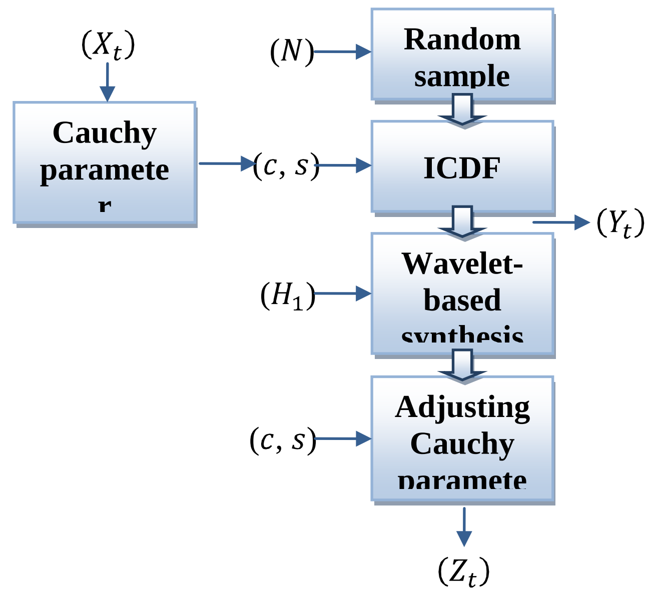

Figure 2: Proposed algorithm

exactly Cauchy-distributed, but it is actually very close if the Hurst index adjustment is not abrupt. Further work can be developed to describe mathematically how the distribution is affected by this non-linear transformation.

6. Simulation and Results

A. Simulation Protocol

The proposed algorithm is evaluated by generating a set of 96 artificial jitter samples. For each jitter sample described in Section 4, a new artificial series was generated with the same length, location, scale and Hurst parameters than the original. The efficiency of the generation was evaluated by estimating the c, s and H parameters and by calculating the square root of the MSE (SMSE) of the CDF from the new series.

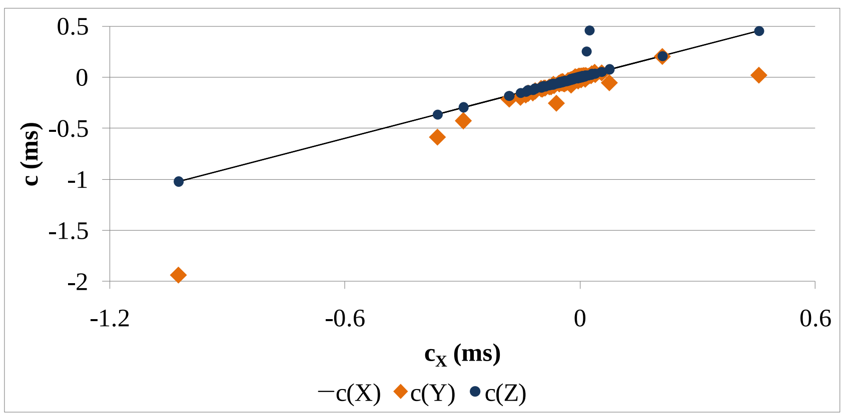

Figure 3: Estimations of the location parameter

The random samples consisted of FGN series generated with an implementation of the algorithm proposed by R. B. Davies and D. S. Harte [8], with Hurst index equal to that of the original jitter series. The location and scale parameters were estimated by applying a MLE estimator [9]. The Hurst index was estimated with an implementation of the Haar-wavelet based estimator, as described in [4].

B. Results

The respective estimation of the c and s parameters for the series X_t, Y_t and Z_t, described in Section 5, are shown in Figure 3 and Figure 4. The estimated values for Y_t and Z_t are very close to that of X_t. The location parameter is in practice very close to zero, e.g., $c \in (-0.5, 0.5)$ for most traces. And the estimated scale parameter is in the range between 0 and 5 for all studied samples. Also note that the ICDF transformation does not depend on whether or not the original series (X_t) is adequately modeled with a Cauchy distribution, but only in the estimation of c and s.

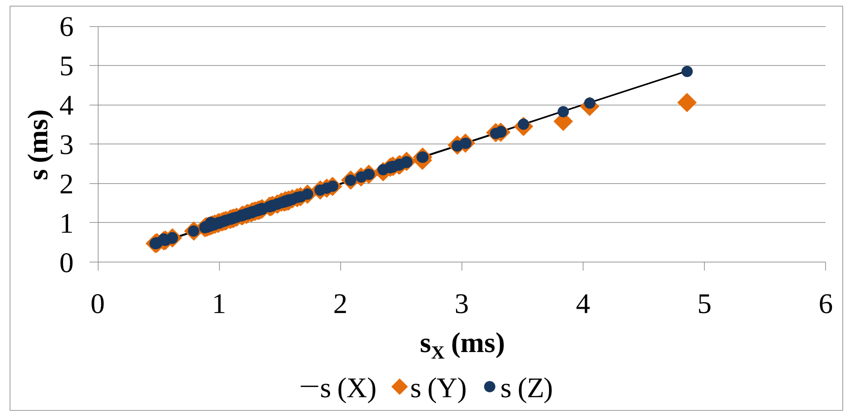

Figure 4: Estimation of the scale parameter

Figure 5 shows the estimated H for the three series X_t, Y_t and Z_t. It can be observed that the correlation structure is somehow altered because of the ICDF transformation, i.e., the estimated H for Y_t is different of that of X_t and closer to 0.5. But the weighted wavelet-based synthesis, described in Section 3, makes the Hurst index of Z_t match that of X_t (the estimated difference $|H_Z - H_X|$ is lower than 10^{-14}).

Figure 6 shows the respective SMSE of the CDF for Y_t and Z_t with respect to a Cauchy CDF with location and scale parameters c_X and s_X (estimated from X_t). It is observed that, although the wavelet-synthesis increases the SMSE of Z_t, it is very close to that of Y_t, i.e., Z_t is very close to Cauchy-distributed if Y_t so is. This is an indication that the efficiency of the proposed generator is very close the ICDF method, but with the advantage that the correlation structure is also adjusted.

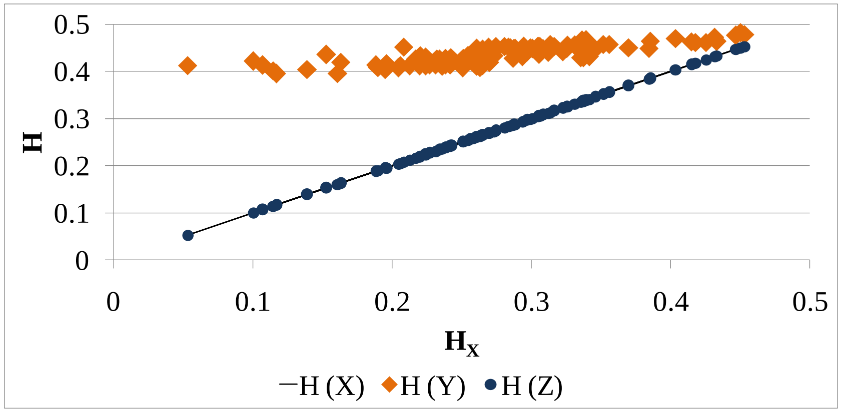

Figure 5: Estimation of the Hurst index

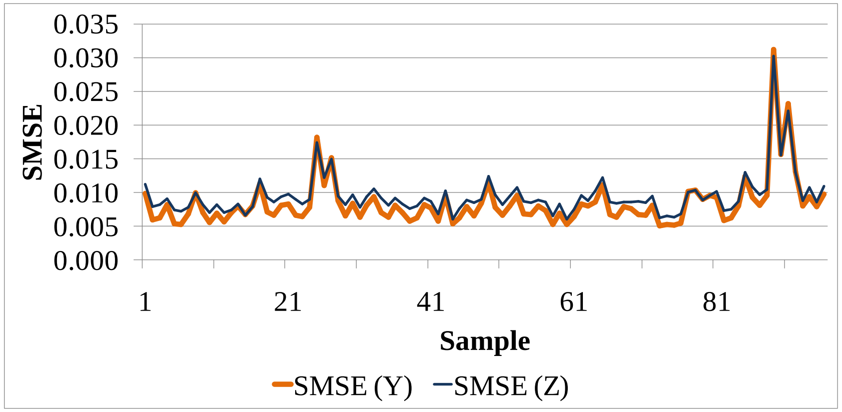

Figure 6: SMSE of the CDF

7. Conclusion

We have proposed a method to generate artificial Cauchy-distributed time series with specific correlation structure, measured by the Hurst index. The generation consists of a sequential implementation of four stages: random sample generation, ICDF transformation, wavelet-based synthesis and a linear transformation. The resulting Cauchy-distributed series has approximately the desired location and scale parameters and exactly the desired Hurst index.

The random sample generation can be implemented by practically any method, but the FGN generation was used in this study so that the Hurst index of the random series is close to the original jitter measurements and to guarantee that pathologic situations, e.g., the statistic $S_2(j) = 0$ for some j, are avoided.

The wavelet-based synthesis proposed in Section 3 produces exact H-soss samples, according to the wavelet-based estimator. The estimated *Logscale Diagram* of the synthesized samples is a perfect straight line, which implies that the wavelet-based estimator applied to these samples is not only unbiased, but also has zero variance. Then, these samples can be applied to evaluate the performance of other estimators.

The applied linear transformations, corresponding to the fourth stage of the proposed generator, adjust approximately the location and scale parameters but leave the Hurst index unchanged. Non-linear transformations may be applied to adjust exactly the location and scale parameters, but in that case the Hurst index may not be exactly the same as desired.

The evaluation of the performance of the generator threw the following results: the artificial samples are Cauchy-distributed and have approximately the desired location and scale parameters and exactly the desired Hurst index. The SMSE of the CDF of the produced series is slightly higher than that obtained with the ICDF method, but it has a specific correlation structure, which cannot be achieved with the ICDF method alone.

8. Acknowledgment

Leopoldo Estrada thanks CONACYT for the financial support. Grant 166532.
Homero Toral tanks PROMEP - UQROO-PTC-110 for partial financial support.

9. References

[1] L. Rizo, D. Torres, D. Munoz and C. Vargas, “Jitter in IP Networks: A Cauchy Approachs”, *IEEE Communication Letters*, Vol. 14, No. 2, 2010, pp. 190-192.

[2] H. J. Jeong, J. R. Lee, and H. Park, “Teletraffic Generation of Self-Similar Processes with Arbitrary Marginal Distributions for Simulation: Analysis of Hurst Parameters”, ICCSA 2004, LNCS 3045, 2004, pp. 827–836.

[3] Lee I. W. C., Fapojuwo A. O., “Stochastic processes for computer network traffic modeling”, *Computer Communications*, Vol. 29, No. 1, 2005, pp. 1-23.
[4] D. Veitch D. and P. Abry, “A wavelet based joint estimator of the parameters of long-range dependence”, *IEEE Transactions on Information Theory*, special issue *Multiscale statistical signal analysis and its applications*, Vol. 45, No. 3, 1999, pp. 878-897.
[5] "Wireshark: A Network Protocol Analyzer”, http://www.wireshark.org/.
[6] L. Estrada, D. Torres and H.Toral, “Analytical Investigation of the Performance of Packet-level FEC Techniques in VoIP Communications”, *WSEAS Transactions on Communications*, issue 9, Vol. 8, 2009.
[7] H. Toral, D. Torres and L. Estrada, “Simulation and Modeling of Packet Loss on VoIP Traffic: A Power-Law Model”, *WSEAS Transactions on Communications*, issue 10, Vol. 8, 2009.
[8] R. B. Davies and D. S. Harte, “Test for Hurst effect”, *Biometrika* 74 (1987), pp.95-101.
[9] P. Axensten, Cauchy CDF, PDF, inverse CDF, parameter fit and random generator.
http://www.mathworks.com/matlabcentral/fileexchange/1174.