Research Article

Heavy-Tailed Prediction Error: A Difficulty in Predicting Biomedical Signals of $1/f$ Noise Type

Ming Li,¹ Wei Zhao,² and Biao Chen²

¹School of Information Science & Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China
²Department of Computer and Information Science, University of Macau, Padre Tomás Pereira Avenue, Taipa, Macau

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 31 October 2012; Accepted 20 November 2012

1. Introduction

Signals of $1/f$ noise type are widely observed in biomedical engineering, ranging from heart rate to DNA and protein, see, for example, [1–37], just to cite a few. Predicting such a type of signals is desired in the field [38–43]. A fundamental issue in this regard is whether a biomedical signal of $1/f$ noise type to be predicted is predicable or not.

The predictability of signals of non-$1/f$ noise type is well studied [44–48]. However, the predictability of $1/f$ noise is rarely reported, to our best knowledge. Since many phenomena in biomedical engineering are characterized by $1/f$ noise [1–37], the predictability issue of $1/f$ noise is worth investigating.

Let $x(t)$ be a biomedical signal in the class of $1/f$ noise. Then, its PDF is heavy-tailed, and it is LRD, see, for example, Adler et al. [69], Samorodnitsky and Taqqu [70], Mandelbrot [71], Li and Zhao [72]. Due to that, here and below, the terms, $1/f$ noise, LRD random function, and heavy-tailed random function are interchangeable.

Let $p(x)$ be the PDF of a biomedical signal $x(t)$ of $1/f$ noise type. Then, its variance is expressed by

$$\text{Var}[x(t)] = \int_{-\infty}^{\infty} (x - \mu_x)^2 p(x)dx,$$

where $\mu_x$ is the mean of $x(t)$ if it exists. The term of heavy tail in statistics implies that $\text{Var}[x(t)]$ is large. Theoretically speaking, in general, we cannot assume that $\text{Var}[x(t)]$ always exists [72]. In some cases, such as the Pareto distribution, the Cauchy distribution, $\alpha$-stable distributions [72], $\text{Var}[x(t)]$ may be infinite. That $\text{Var}[x(t)]$ does not exist is particularly true for signals in biomedical engineering and physiology, see Bassingthwaighte et al. [33] for the interpretation of this point of view.

Recall that a prediction error is a random function as we shall soon mention below. Therefore, whether the prediction error is of $1/f$ noise, or equivalently, heavy-tailed, turns to be a crucial issue we need studying. We aim at, in this research, exhibiting that prediction error of $1/f$ noise is heavy-tailed
and accordingly is of $1/f$ noise. Thus, generally speaking, the variance of a prediction error of a biomedical signal $x(t)$ of $1/f$ noise type may not exist or large. That is a reason why predicting biomedical signals of $1/f$ noise type is difficult.

The rest of this paper is organized as follows. Heavy-tailed prediction errors occurring in the prediction of biomedical signals of $1/f$ noise type are explained in Section 2. Discussions are in Section 3, which is followed by conclusions.

2. Prediction Errors of $1/f$ Noise Type

We use $x(n)$ to represent a biomedical signal in the discrete case for $n \in \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers. Let $x_N(n)$ be a given sample of $x(n)$ for $n = 0, 1, \ldots, N - 1$. Denote by $x_M(m)$ the predicted values of $x(n)$ for $m = N, N + 1, N + M - 1$. Then, the prediction error denoted by $e(m)$ is given by

$$e(m) = \sum_{m=N}^{N+M-1} x(m) - x_M(m). \quad (2)$$

If one uses the given sample of $x(n)$ for $n = N, N + 1, \ldots, 2N - 1$ to obtain the predictions denoted by $x_M(m)$ for $m = 2N, 2N + 1, 2N + M - 1$, the error is usually different from (2), which implies that the error $e(m)$ is a random variable. Denote by $p(e)$ the PDF of $e(m)$. Then, its variance is expressed by

$$\text{Var}[e(m)] = \sum_{m=N}^{N+M-1} (e - \mu_e)^2 p(e), \quad (3)$$

where $\mu_e$ is the mean of $e(m)$.

Let $P$ be the operator of a predictor. Then,

$$x_M(m) = Px_N(n). \quad (4)$$

A natural requirement in terms of $P$ is that $\text{Var}[e(m)]$ should be minimized. Thus, the premise that $\text{Var}[e(m)]$ can be minimized is that it exists.

It is obviously seen that $\text{Var}[e(m)]$ may be large if $p(e)$ is heavy tailed. In a certain cases, $\text{Var}[e(m)]$ may not exist. To explain the latter, we assume that $e(m)$ follows a type of heavy-tailed distribution called the Pareto distribution.

Denote by $p_{\text{Pareto}}(e)$ the PDF of the Pareto distribution. Then [73], it is in the form

$$p_{\text{Pareto}}(e) = \frac{ab^a}{e^{a+1}}, \quad (5)$$

where $e \geq b$, $a > 0$, and $b > 0$. The mean and variance of $e(m)$ are, respectively, expressed by

$$\mu_e = \frac{ab}{a - 1}, \quad (6)$$

$$\text{Var}(e) = \frac{ab^2}{(a - 1)^2(a - 2)}.$$

The above exhibits that $\text{Var}[e(m)]$ does not exist if $a = 1$ or $a = 2$ and if $e(m)$ follows the Pareto distribution.

Note that the situation that $\text{Var}[e(m)]$ does not exist may not occur if $e(m)$ is light-tailed. Therefore, the question in this regard is whether $e(m)$ is heavy-tailed if a biomedical signal $x(n)$ is of $1/f$ noise. The answer to that question is affirmative. We explain it below.

Theorem 1. Let $x(n)$ be a biomedical signal of $1/f$ noise type to be predicted. Then, its prediction error is heavy-tailed. Consequently, it is of $1/f$ noise.

Proof. Let $r_{xx}(k)$ be the autocorrelation function (ACF) of $x(n)$. Then,

$$r_{xx}(k) = E[x(n)x(n + k)], \quad (7)$$

where $k$ is lag and $E$ the mean operator. Let $r_{MM}(k)$ be the ACF of $x_M(m)$. Then,

$$r_{MM}(k) = E[x_M(m)x_M(m + k)]. \quad (8)$$

Let $r_{ee}(k)$ be the ACF of $e(m)$. Then,

$$r_{ee}(k) = E[e(m)e(m + k)]. \quad (9)$$

Note that

$$r_{ee}(k) = E[e(m)e(m + k)]$$

$$= E[\{x(m) - x_M(m)\}\{x(m + k) - x_M(m + k)\}]$$

$$= E[x_M(m)x_M(m + k) + x_M(m)x_M(m + k) - x_M(m)x_M(m + k) - x_M(m)x_M(m + k)]$$

$$= r_{xx}(k) + r_{MM}(k) - r_{Mx}(k) - r_{Mx}(k). \quad (10)$$

In the above expression, $r_{Mx}(k)$ is the cross-correlation between $x_M(m)$ and $x(m)$. On the other side, $r_{Mx}(k)$ is the cross-correlation between $x(m)$ and $x_M(m)$. Since $r_{Mx}(k) = r_{Mx}(k)$, we have

$$r_{ee}(k) = r_{xx}(k) + r_{MM}(k) - 2r_{Mx}(k). \quad (11)$$

Recall that $x(m)$ is $1/f$ noise. Thus, it is heavy-tailed and hence LRD. Consequently, for a constant $c_1 > 0$, we have

$$r_{xx}(k) \sim c_1k^{-a} \quad (k \to \infty) \text{ for } 0 < \alpha < 1. \quad (12)$$

On the other hand, the predicted series $x_M(m)$ is LRD. Thus, for a constant $c_2 > 0$, the following holds:

$$r_{MM}(k) \sim c_2k^{-\beta} \quad (k \to \infty) \text{ for } 0 < \beta < 1. \quad (13)$$

In (11), if $r_{Mx}(k)$ is summable, that is, it decays faster than $r_{x}(k)$ or $r_{Mx}(k)$, it may be ignored for $k \to \infty$. In this case, $r_{ee}(k)$ is still non-summable. In fact, one has

$$r_{ee}(k) \sim \begin{cases} c_1k^{-\alpha}, & 0 < \alpha < \beta < 1, \\ c_2k^{-\beta}, & 0 < \beta < \alpha < 1, \quad (k \to \infty), \\ (c_1 + c_2)k^{-\beta}, & \alpha = \beta. \end{cases} \quad (14)$$
On the other side, when \( r_{xM}(k) \) is non-summable, \( r_{e}(k) \) is non-summable too. In any case, we may write \( r_{e}(k) \) by

\[
r_{e}(k) \sim k^{-\gamma} \quad (k \to \infty) \quad \text{for} \quad 0 < \gamma < 1. \tag{15}
\]

Therefore, the prediction error \( e(m) \) is LRD. Its PDF \( p(e) \) is heavy-tailed according to the Taqqu’s law. Following [72], therefore, \( e(m) \) is a \( 1/f \) noise. This completes the proof. \( \square \)

### 3. Discussions

The present result implies that cautions are needed for dealing with prediction errors of biomedical signals of \( 1/f \) noise type. In fact, if specific biomedical signals are in the class of \( 1/f \) noise, the variances of their prediction errors may not exist or large [72]. Tucker and Garway-Heath used to state that their prediction errors with either prediction model they used are large [74]. The result in this paper may in a way provide their research with an explanation.

Due to the fact that a biomedical signal may be of \( 1/f \) noise, PDF estimation is suggested as a preparatory stage for prediction. As a matter of fact, if a PDF estimation of biomedical signal is light-tailed, its variance of prediction error may not exist or large [72]. Tucker and Garway-Heath used to state that their prediction errors with either prediction model they used are large [74]. The result in this paper may in a way provide their research with an explanation.

### 4. Conclusions

We have explained that the prediction error \( e(m) \) in predicting biomedical signals of \( 1/f \) noise type is usually LRD. This implies that its PDF \( p(e) \) is heavy-tailed and \( 1/f \) noise. Consequently, \( \text{Var}[e(m)] \) may in general be large. In some cases [72], \( \text{Var}[e(m)] \) may not exist, making the prediction of biomedical signals of \( 1/f \) noise type difficult with the way of minimizing \( \text{Var}[e(m)] \).

### Acknowledgments

This work was supported in part by the 973 plan under the Project Grant no. 2011CB302800 and by the National Natural Science Foundation of China under the Project Grant no. 61272402, 61070214, and 60873264.

### References


[59] R. S. Deo, “Improved forecasting of autoregressive series by weighted least squares approximate REML estimation,”


Submit your manuscripts at http://www.hindawi.com