國立交通大學

電子工程學系 電子研究所

博士論文

自我類化訊流的合成及分散式辨識

Synthesization and Decentralized Identification of Self-Similar Processes

研 究 生：姚 建

指導教授：蔣 迪 豪 博 士

陳 伯 寧 博 士

中 華 民 國 九 十 七 年 七 月

自我類化訊流的合成及分散式辨識

Synthesization and Decentralized Identification of Self-Similar Processes

研 究 生：姚 建　　Student：Chien Yao

指導教授：蔣迪豪　　Advisor：Tihao Chiang

陳伯寧　　Po-Ning Chen

國 立 交 通 大 學

電子工程學系 電子研究所

博 士 論 文

A Dissertation

Submitted to Department of Electronics Engineering and

Institute of Electronics

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Electronics Engineering

July 2007

Hsinchu, Taiwan

2008 年 7 月

自我類化訊流的合成及分散式辨識

學生：姚 建　　　　　　　　　　　　　　　　　　　　　　　指導教授：蔣迪豪

陳伯寧

國立交通大學電子工程學系

摘　　要

在本論文的第一部分，我們提出了一個濾波器型式的自我類化訊流合成器。此合成器能生成可調控變異性及相關性的長程相依訊流，同時也只需很少的輸入參數。與既有的其它自我類化訊流合成器（如 RMD 方法和 Paxson 的 IFFT 方法）相比，我們提出的濾波器型式的自我類化訊流合成器具有能即時生成訊流以及生成之訊流恆不為負值之優點。我們接著研究了相關係數（只能反映線性的相依關係）和交互訊息（能測量一般的相依關係）兩者之間的蘊含關係。本研究的結果建議，對於弱相關的隨機變數，如一個自我類化過程中具有長的時間差的不同二個時刻的值，相關係數的平方值的一半似可作為交互訊息的一個合理的近似。

基於長尾分佈和自我類化訊流之間的存在的基本關係，我們進而研究了此類訊流的分散式辨識問題。我們發現若干有趣的結果。首先，我們確證了全同感測器系統在指數分佈族的參數辨識問題上的最佳性。在此研究方向上的一個相關的結果是，在指數分佈族辨識問題上，串接式兩感測器系統和平行式兩感測器系統具有相同的最佳性能。這多少是令人感到意外的，因為一般認為串接式兩感測器系統比平行式兩感測器系統有更好的性能。

其次，對於更一般類別分佈族的參數辨識問題，我們提出數個命題可用來檢證全同感測器系統的最佳性。如採取直捷的手法來檢證全同感測器系統，通常將導致在被一組非線性聯立方程所定義的解空間之中搜尋所有的局域最小值。然而這種方法在一些情況下會是不可行的，而我們提出的命題可作為一個較佳的替代方案。此外，我們的研究也可應用到其它的分散式檢測問題上，如在倖存分析及損壞時間分析的壽命問題上，或是如在利用在地理上分散設備，對不同連結上的封包到達時間間隔加以量測，來決定整個網路的自我類化參數的問題上。

最後，藉由對函數及方程式的數值計算結果，我們確認了，在加法性高斯雜訊下二元信號的檢測問題上全同感測器系統的最佳性。

Synthesization and Decentralized Identification of Self-Similar Processes

student：Chien Yao

Advisors：Dr. Tihao Chiang
Dr. Po-Ning Chen

Department of Electronics Engineering and Institute of Electronics
National Chiao Tung University

ABSTRACT

In the first part of this dissertation, we propose a filter-based generator for the synthesization of *self-similar* traffics. It can produce long range dependent traffics with adjustable levels of bustiness and correlation, and is parsimonious in the number of model parameters. By comparing it with existing self-similar traffic synthesizers, e.g., the RMD and the Paxson IFFT algorithms, the proposed filter-based synthesizer has the advantages that the synthetic traffic can be generated on the fly, and always produces non-negative-valued traffic. The implications between the correlation coefficient (a quantity that only measures the *linear* dependence) and mutual information (a quantity that can represent the *general* dependence) is subsequently investigated. The obtained results suggest that for weakly correlated random variables such as two instances of a self-similar process with a long time lag, half the square of the correlation coefficients might be a reasonable approximation to the mutual information.

Continuing from the synthesization of processes with heavy tails, we turn to study the impact of such processes on decentralized detection. Several interesting results are found. Firstly, the optimality of identical sensor system for the exponential distribution family has been verified. A side result along this research line is that the optimal performance of the serial two-sensor system is the same as that of the parallel two-sensor system for exponential sources. This is somewhat surprising because it is generally considered that the serial two-sensor system has better performance than the parallel two-sensor system.

Secondly, for a more general class of distribution families, we propose several propositions on the optimality of the identical system. A straightforward approach to test the optimality of identical sensor system often results in searching all local

minimums in the solution space that is defined through a set of nonlinear equations. However, this approach is not tractable in certain situations. Our propositions then provide an alternative for optimality test of identical sensor system. Besides, they can be applied to other decentralized detection problems like the detection of lifetime encountered in survival analysis and failure time analysis or the determination of the degree of self-similarity of the whole network system based on geographically dispersed measurements of the packet inter-arrival times on different links.

Finally, with the help of numerical study on functions and equations, we analytically confirm the optimality of identical sensor system over Gaussian sources.

誌　　　　謝

首先我要向兩位指導教授蔣迪豪教授和陳伯寧教授表示我最深的謝意，謝謝您們多年來給我的指導和對我提供的特別的幫助，您們的耐心、寬容、及對學生的真誠關懷是無法比擬的。我還要感謝台北大學通訊所韓永祥教授、中山大學通訊所王藏億教授、及 Syracuse University 的 Varshney 教授對我的協助。感謝花凱龍及陳金源對本論文第一部份的重要貢獻。感謝交大電子系對我的栽培與照顧。

最後，我想感謝我的家人給我的愛與支持，你們是我前進的動力。

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Stationary random processes, according to their autocorrelation functions, can be classified as either *short-range* or *long-range* dependence. The former have summable autocorrelation functions, while the latter have non-summable autocorrelation functions. The simulations of the short-range dependent random processes have attracted attention for years, and have found many applications such as the traffic model of telecommunication systems [6]. However, researchers had recently found that the traffic in many modern communications, such as the world wide web [4, 8, 14, 18, 20] and variable-bit-rate (VBR) video transmission [10], is significantly different from the conventional short-range dependent traffic models, and exhibit the renowned self-similar nature. This arouse the demand for the synthesization of processes with long-range dependence.

In literature, there have been several approaches proposed for the synthesization of long-range-dependent self-similar traffics. They include methods based on fractional Gaussian noise [14], $M/G/\infty$ queue model [12], autoregressive processes [3], wavelet [2], $\ldots$, etc. These synthesizers can be roughly divided into two categories: approaches derived from "time-domain" aspect and ones developed from "frequency-domain" standpoint. An example for the former is the random-midpoint displacement (RMD) algorithm proposed by Lau et al. [13], while the spectrum fitting to the fractional Gaussian noise, as proposed by Paxson [19], can be a typical synthesizer for the latter.

The procedures of the RMD algorithm is to recursively subdivide the present time intervals, and generate in each subdivision a new mid-point traffic data based on the end-point data obtained in the previous subdivision. This method can efficiently generate a well-approximated fractal Brownian motion (FBM) sequence. It however comes with the drawbacks that only the FBM traffics can be synthesized, and the desired amount of traffic has to be specified in advance.

Based on the power spectrum fitting to the fractional Gaussian noise (FGN), Paxson proposed a fast self-similar traffic generator using the inverse discrete-time Fourier transform (IDTFT), which is usually referred as the FFT method. By using an approximate form of the spectrum density of fractal Gaussian noises (FGN), a random sequence is formed in frequency domain. An inverse Fourier transformation (IFFT) is then performed to transform the sequence from the frequency domain to the time domain. The FFT algorithm improves the RMD algorithm in speed. In particular, the FFT algorithm only takes half time of the RMD algorithm for the same sequence length. Again, its drawback is that the traffic sequence cannot be generated on the fly. In addition, the simplified form of the FGN spectrum causes the resultant degree of self-similarity deviated from the target one.

In applying the aforementioned approaches to the generation of self-similar traces, several problems can be encountered. Firstly, the required length (i.e., amount) of traffic data must be priorly determined; hence, when a longer traffic sequence is required, one has to drop the existing data, and re-generate a completely new trace of the required length. Secondly, the required traffic data must be generated in an *off-line* fashion before they can be put to use. This somewhat restricts their usage in situation where *on-the-fly* traffic synthesizers are needed. Thirdly, these traffic generators may produce negative number, which is an undesired value for, say, packet-train arrivals. The direct elimination of these negative-valued data however may make the degree of self-similarity of the generated trace deviating from the target one.

In this work, we propose a model that can produce long-range dependent sequences with adjustable levels of bustiness and correlation. When it is compared to the two known self-similar traffic generators—the RMD and the Paxson FFT, our model provides additional advantages that the synthetic traffic can be generated on the fly, and is always non-negative. Although the variance-time analysis shows that the filter length W limits the valid aggregation size of self-similarity, this phenomenon turns out to match the measured behavior of true network traffic, where the self-similar nature only lasts beyond a practically manageable range, but disappears as the considered aggregated window is much further extended, e.g., Beran et al. [4, Fig. 2].

The relationship between the second-order statistics (which are used in the measurement of the self-similarity in the network traffic) and the quantities in the information theory is also an interesting topic. Since one might expect that the self-similar traffic has some special characteristics that can be easily identified in the information processing of the measured data, we discuss the relationship between the correlation coefficients and mutual information in Chapter 3.

For the practical control of network traffic, one might need to test whether its self-similarity is weak or strong enough that the long-range dependence could or could not be ignored. To reduce the response time and to alleviate the load of network, a decentralized scheme for the detection of the self-similarity might be useful. In this work, we consider the decentralized detection, especially on the optimal design of the local decision rules and the fusion rule for the classification of exponential sources. It turns out that the optimal strategy is to use identical sensors and k-out-of-n fusion rule. We also show for such classification problem that the optimal performance of the serial two-sensor system is the same as the optimal parallel two-sensor system. In addition, we address a set of propositions on the optimality of the identical sensor system, which can be verified without much difficulty. Some generalizations are further established and remarked for the decentralized detection of Gaussian sources, and for the determination of degree of self-similarity via the local measurements of packet inter-arrival durations.

1.1 Definitions of Self-Similar Processes

Self-similar processes were first introduced by Mandelbrot and his co-workers in 1968 [15, 16, 17]. These processes were thereafter found applications to many fields such as astronomy, chemistry, economics, engineering, mathematics, physics, statistics, etc. Recently, measurement studies have shown that the actual traffic from computer networks is long-range dependent [14, 18, 8, 4, 20], and thus another new application of self-similar processes was initiated.

Assume a second-order stationary real-valued stochastic process $\boldsymbol{Y} \triangleq \{Y_i\}_{i\in I_1}$ with finite marginal mean μ and marginal variance σ^2, where $I_j \triangleq \{j, j+1, j+2, \ldots\}$. Denote by $\boldsymbol{Y}^{(m)} \triangleq \{Y_i^{(m)}\}_{i\in I_1}$ the m-averaged process of $\boldsymbol{Y}$, where for $m, i \in I_1$,

$$Y_i^{(m)} \triangleq \frac{1}{m}\sum_{j=1}^{m} Y_{m(i-1)+j}.$$

Let the autocovariance and autocorrelation coefficient function of the m-averaged process $\boldsymbol{Y}^{(m)}$ be denoted by $C_m(k) \triangleq \mathrm{Cov}\{Y_i^{(m)}, Y_{i+k}^{(m)}\}$ and $\rho_m(k) \triangleq C_m(k)/C_m(0)$, respectively. For notational convenience, the subscript of $\rho_m(\cdot)$ will be dropped when $m = 1$. Then, several variants of self-similarities can be defined as follows.

Definition 1.1. [24] A strictly stationary process $\mathbf{Y}$ is called *strictly self-similar* with parameter $H = 1 - (\beta/2)$, where $0 < \beta < 1$, if

$$m^{1-H}\mathbf{Y}^{(m)} \stackrel{d}{=} \mathbf{Y} \quad \text{for } m \in I_1 \tag{1.1}$$

where "$\stackrel{d}{=}$" means that the equality is taken in the sense of finite-dimensional distributions.

Definition 1.2. [24] A second-order stationary process $\mathbf{Y}$ is called *exactly second-order self-similar* with parameter $H = 1 - (\beta/2)$, where $0 < \beta < 1$, if either of the following conditions holds:

$$\rho(k) = \frac{1}{2}[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}],\ \ k \in I_1 \tag{1.2}$$

$$C_m(k) = C_1(k)m^{-\beta},\ \ k \in I_0, m \in I_1 \tag{1.3}$$

Notably, (1.2) and (1.3) are indeed equivalent. Also note that (1.2) implies that $\rho_m(k) = \rho_1(k)$ for $m \in I_1$.

Definition 1.3. [24] A second-order stationary process $\mathbf{Y}$ is called *asymptotically second-order self-similar* with parameter $H = 1 - (\beta/2)$, where $0 < \beta < 1$, if

$$\lim_{k\to\infty} \rho_m(k) = \frac{1}{2}[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}],\ \ m \in I_1. \tag{1.4}$$

The parameter H in the above definitions is usually referred to as the Hurst parameter. For other variant definitions of self-similar processes, see [24] and [25].

1.2 Properties of Self-Similar Processes

In this section, we summarize the statistical properties of self-similar processes that are of use in this work.

1.2.1 Range of Dependence

Random processes can be classified into two groups: *short-range dependence* (SRD) and *long-range dependence* (LRD). Their formal definitions that have been appeared in the literature are given below.

Definition 1.4. [24] [7] A process $\mathbf{Y}$ is said to be *short-range dependent*, if

$$\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty \tag{1.5}$$

Definition 1.5. [24] [7] A process $\mathbf{Y}$ is said to be *long-range dependent*, if

$$\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty \tag{1.6}$$

A variant definition of long-range dependence is defined as follows.

Definition 1.6. [22] A process $\mathbf{Y}$ is said to be *long-range dependent*, if

$$\lim_{k\to\infty} \frac{\rho(k)}{L(k)k^{2H-2}} = 1, \tag{1.7}$$

where $L(k)$ is a slowly varying function at infinity, defined by

$$\lim_{k\to\infty} \frac{L(kx)}{L(k)} = 1 \text{ for all } x > 0. \tag{1.8}$$

For an exact second-order self-similar process $\mathbf{Y}$, its autocorrelation coefficient function is given by equation (1.2), i.e.,

$$\rho(k) = \frac{1}{2}[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}], \;\; k \in I_1.$$

Using Taylor expansion, we obtain

$$\rho(k) = H(2H-1)k^{2H-2} + o(k^{2H-2}), \;\; k \in I_1, \;\; 0.5 < H < 1.$$

Therefore, an exact second-order self-similar process is indeed long-range dependent in the sense of Definition 1.6.

1.2.2 $1/f$-Noise

$1/f$-noise is the term used to present a sharp divergence in the power spectral density around the origin. The exact definition of $1/f$-noise is in the following.

Definition 1.7. [22] A stationary process $\mathbf{Y}$ is said to present $1/f$*-noise*, if its power spectral density $S(\omega)$ satisfies:

$$\lim_{\omega \to 0} \frac{S(\omega)}{L(1/\omega)\omega^{1-2H}} = 0, \tag{1.9}$$

where $L(k)$ is a slowly varying function at infinity (cf. (1.8)), and Hurst parameter H is in the range of $(0.5, 1)$.

It has been proven that the long-range dependence in the sense of Definition 1.6 is equivalent to $1/f$-noise [3, pp. 53].

1.2.3 Slowly Decaying Variance of Self-Similar Processes

In the case of short-range dependence or independence, the variance of m-averaged process decreases as the reciprocal of the average size, m. However, by equation (1.3),

$$\mathrm{Var}\{X^{(m)}\} = C_m(0) = C_1(0)m^{-\beta} = \mathrm{Var}\{X\}/m^{2-2H}, \tag{1.10}$$

and the variance of m-averaged processes decreases more slowly than the reciprocal of the average size, m, for long-range dependent processes. In fact, (1.10) indicates that $\mathrm{Var}\{X^{(m)}\}$ decreases as a slop of $(2H-2)$ in log-log plot against m.

1.2.4 Heavy-Tailed Distribution

Definition 1.8. A random variable Y is said to be *heavy-tailed* with parameter $\alpha \geq 0$, if

$$\lim_{y \to \infty} \frac{\Pr\{Y > y\}}{L(y)y^{-\alpha}} = 1, \tag{1.11}$$

where $L(x)$ is a slowly varying function at infinity (cf. (1.8)).

Here, we only concern the cases of $1 < \alpha < 2$, i.e., the mean of random variable Y is finite, and its variance is infinite. The infinite variance can be regarded as an extremely variable phenomenon. This kind of heavy-tailed random variable has been used to model the inter-arrival time of network packets. It has been shown [11] that if the packet inter-arrival process is modelled as i.i.d. Pareto random variables,[1] the packet counting process is asymptotically second-order self-similar process with $H = (3 - \alpha)/2$, where parameter α is in the range of $(0, 1)$ and $(1, 2)$.

1.2.5 Hurst Effect

Historically, self-similar processes are marked because these processes provide an elegant interpretation of the empirical phenomenon, usually referred to as the *Hurst Effect*.

Given a series of observations $Y_1, Y_2, Y_3, \cdots$ with sample mean $\mu(n) = (1/n)\sum_{j=1}^{n} Y_j$ and sample variance

$$S(n) = \frac{1}{n}\sum_{j=1}^{n}[Y_j - \mu(n)]^2,$$

the re-scaled adjusted range (or conventionally, the R/S statistics) is defined as

$$\frac{R(n)}{S(n)} = \frac{\max_{1\le k\le n}\left[\sum_{j=1}^{k} Y_j - k\mu(n)\right] - \min_{1\le k\le n}\left[\sum_{j=1}^{k} Y_j - k\mu(n)\right]}{S(n)}. \tag{1.12}$$

Hurst [5] found that many naturally occurring time sequences could be well characterized by

$$\lim_{n\to\infty}\frac{E[R(n)/S(n)]}{cn^H} = 1 \tag{1.13}$$

with c being a finite positive constant and Hurst parameter in the range of $(0.5, 1)$. This is therefore termed the Hurst Effect.

Additionally, Mandelbrot and Van Ness [16] showed that if the observation sequences are short range dependent, then

$$\lim_{n\to\infty}\frac{E[R(n)/S(n)}{cn^{0.5}} = 1. \tag{1.14}$$

[1] Pareto distribution is a heavy-tailed distribution with probability density function $f(x) = ak^a/y^{a+1}$ for $a > 0$, $k > 0$ and $y \ge k$. The cumulative distribution function of Pareto is $1 - (k/y)^a$.

1.3 Decentralized Detection

A decentralized detection system consists of n sensors, sometimes geographically dispersed, and a remote fusion center. Each of the sensor observes a phenomenon (often modeled as a random variable X_i), summarizes it into a single bit u_i, and then transmits u_i to the fusion center uncooperatively. Based on received $\{u_i\}_{i=1}^n$, the fusion center determines whether these $\{X_i\}_{i=1}^n$ are drawn from null distribution $P(\cdot|H_0)$ or alternative distribution $P(\cdot|H_1)$.

Tenney and Sandell [35] are the first to bring attention to such a detection framework. Despite that it has an apparent handicap on the performance, the decentralized detection system requires much smaller bandwidth between the observers and the global decision maker than its centralized counterpart. This is a significant benefit when the system is required to operate in a harsh environment. The workload of information processing is also distributed from the decision-making center to the local observers; therefore the overall complexity of the classification system can be reduced. Furthermore, allotting many measurement devices and local data processers instead of one central unit can also partly ensure the reliability, even when some of the sensors malfunctions. All of these motivate distributed detection systems to rival with conventional centralized detection systems, especially for applications where the measurements have to be geographically dispersed, and have to be collected by remote sensors.

Contrast to the advantages from the operational aspects above, the optimal design of distributed detection systems is, however, far more difficult than centralized ones. This comes from the decisions of local processers entangle with each other for the contributions to the correctness of overall decision. Accordingly, the optimal design involves the joint optimization of local processers and fusion center. Such optimization problem has been studied for its different facets in the literature. Hereafter, we only mention those most related to the theme in this dissertation.

Tsitsiklis [37] investigated the error performance of decentralized systems with a large number of sensors in terms of error exponents. He showed that a system design with identical sensors are asymptotically optimal. This result was further extended by Chen and Papamarcou [36] by showing that the ratios of error probabilities between the best identical sensor system and the absolutely optimal system are bounded from both above and below. Irving and Tsitsiklis [34] found that for the detection of signals in Gaussian noises, the absolutely optimal two-sensor system should equip identical sensors. Zhang *et al.* [38] concerned the performance of identical sensor systems, and showed that the probability of error is a quasi-convex function of the likelihood ratio test thresholds of local sensors.

1.4 Synopsis of the Dissertation

The materials in this dissertation are arranged into two parts. The first part consisting of Chapters 2 and 3 focuses on the self-similar traffic synthesizer, while the second part extends the focus to decentralized detection in Chapter 4. The general facts about self-similar processes, heavy-tailed distributions, and long-range dependence have already been covered in Section 1.1. The background of decentralized detection required for Chapter 4 is contained in Section 1.4. In Chapter 2, a filter-based self-similar trace synthesizer is proposed, and the degree of its self-similarity is examined in terms of variance-time analysis. The effect due to filter truncation and filter output rounding is subsequently investigated. Comparison between the use of the forward filter and that of the reverse filter is also provided in Chapter 2. The relationship between the second-order statistics and the correlation coefficients is investigated in Chapter 3. The optimal design of the decentralized detection system is the focus of Chapter 4, where the optimality of identical sensor systems is built in an analytical way for exponential distributed hypotheses, and the extension to Gaussian sources follows.

For the general detection problem, a set of propositions on the optimality of the identical sensor system is addressed. Finally, in the same chapter, we indicate at the end that the decentralized detection framework we considered can be applied to other situations such as the detection of lifetime encountered in survival analysis and failure time analysis or the determination of the degree of self-similarity of the whole network system based on geographically dispersed measurements of the packet inter-arrival times on different links. The final comments appear in Chapter 5.

Chapter 2

A FILTER-BASED SELF-SIMILAR TRACE SYNTHESIZER

Recent empirical studies have shown that the modern computer network traffic is much more appropriately modelled by long-range dependent self-similar processes than traditional short-range dependent processes such as Poisson. Thus, if self-similar nature is not considered in the synthesization of experimental network data, incorrect performance assessments for network system may be resulted. This arises the need of a well self-similar trace synthesizing algorithm with long-range dependence. In this chapter, we proposed and examined the feasibility of a filter-based method for the synthesization of self-similar network traces. The proposed approach can alleviate the problems encountered by the conventional synthesizers, such as *random midpoint displacement* and *Paxson's spectrum fitting*, which cannot generate self-similar traces on the fly and may give negative numbers. Additionally, the extended range of self-similarity of the filtered approach can be well manageable by the filter truncation window; therefore, a trace that faithfully matches the measured behavior of true network traffic, where the self-similar nature only lasts beyond a certain range but disappears as the considered aggregated window is much further extended, can be generated.

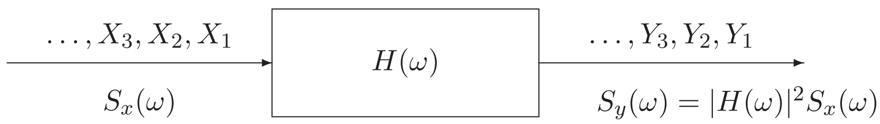

Figure 2.1: Relation between the power spectral densities of the filter input and filter output random processes.

2.1 Filter-Based Asymptotic Self-Similar Traffic Synthesizer

In this section, we proposed and proved that an asymptotic self-similar traffic can be theoretically synthesized through filter technique with prohibitively simple transfer function of infinite order. In its feasible realization, the filter of *infinite* order has to be truncated to a *finite* impulse response (FIR) filter. The resultant degradation due to filter truncation in asymptotic self-similar degree is subsequently examined.

2.1.1 Transfer Function In Self-Similar Traffic Synthesizer

Let $S_y(\omega)$ denote the power spectrum of discrete random process $\mathbf{Y}$ obtained by passing the random process $\mathbf{X}$ with power spectrum $S_x(\omega)$ through a filter with transfer function $H(\omega)$ as shown in Fig. 2.1. An elementary filtering theory immediately gives that $S_y(\omega) = |H(\omega)|^2 S_x(\omega)$. Accordingly, if $\mathbf{X}$ is i.i.d., and $|H(\omega)|^2$ well-approximates the power spectrum of an asymptotic self-similar traffic, then the filter output straightforwardly become self-similar, and can be obtained through $Y_n = X_n * h[n]$, where "$*$" denotes the convolution operator.

By Definition 1.3, the ultimate autocorrelation coefficient function of an asymptotic second-order self-similar process with parameter H equals $\frac{1}{2}[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}]$ for $k \in I_1$, which gives a power spectrum $\sin(\pi H) \cdot \Gamma(2H+1) \cdot |1 - e^{-j\omega}|^2 \sum_{k=-\infty}^{\infty} |\omega + 2\pi k|^{-1-2H}$

for $-\pi \leq \omega < \pi$. Since the asymptotic self-similar behavior of a process is only sensitive to the vicinity of those ω values around the origin [19], we can replace the above infinite sum by its main term at $k = 0$, and yield $\sin(\pi H) \cdot \Gamma(2H+1) \cdot |1 - e^{-j\omega}|^2 \cdot |\omega|^{-1-2H}$ for $-\pi \leq w < \pi$. We then observe that $|\omega|$ can be well-approximated by $|1 - e^{-j\omega}|$ when $|\omega|$ is small. As a consequence, our proposed filter output spectrum becomes $S_y(\omega) = |1 - e^{-jw}|^{1-2H}$ for $-\pi \leq w < \pi$, where the coefficients, $\sin(\pi H) \cdot \Gamma(2H + 1)$, is removed for analytical simplicity.

One may question that such an extensive simplification to the target second-order self-similar spectrum may already remove its self-similar nature. However, it can be derived from Theorem 2.1(ii) in [3] and from the below equation,

$$\lim_{|\omega|\downarrow 0} \frac{S_y(\omega)}{|\omega|^{1-2H}} = \lim_{|\omega|\downarrow 0} \frac{|1 - e^{-jw}|^{1-2H}}{|\omega|^{1-2H}} = \lim_{|\omega|\downarrow 0} \frac{(2\,|\sin(w/2)|)^{1-2H}}{|\omega|^{1-2H}} = 1,$$

that the autocorrelation function $C_1(k)$ of the filter output process $\mathbf{Y}$ with power spectrum $S_y(\omega) = |1 - e^{-jw}|^{1-2H}$ satisfies

$$\lim_{k\to\infty} \frac{C_1(k)}{2\,\Gamma(2-2H)\sin(\pi H - \pi/2)k^{2H-2}} = 1.$$

Thus, from [24, Thm. 3(2)], the marginal variance $C_m(0)$ of the m-averaged process of the filter output process satisfies

$$\lim_{m\to\infty} \frac{C_m(0)}{C_1(0)m^{2H-2}} = \frac{2\,\Gamma(2-2H)\sin(\pi H - \pi/2)}{H(2H-1)}.$$

This implies that for m large, $\log[C_m(0)/C_1(0)]$ behaves asymptotically as $(2H-2)\log(m) + \log[2\Gamma(2-2H)\sin(\pi H-\pi/2)/(H(2H-1))]$. Therefore, the filter output process is asymptotic self-similar with parameter H from the aspect of variance-time analysis, when the average window m is large.

A somewhat surprising result is that the designed filter output process $\mathbf{Y}$ is also quite "self-similar" for *small* m. In other words, $\mathbf{Y}$, in spite of its simple power spectrum formula,

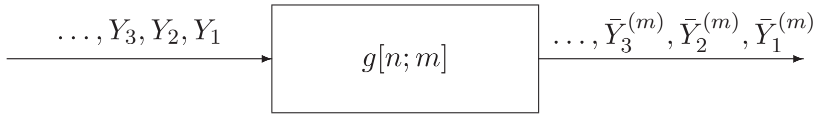

Figure 2.2: The variance-equivalent m-averaged process.

behaves close to an *exact* self-similar process from the aspect of variance-time analysis. This can be numerically verified as follows.

The self-similar nature of the filter output process at small m can be established by analyzing the marginal variance of its *variance-equivalent m-averaged process.* A *variance-equivalent m-average process* $\bar{Y}_1^{(m)}, \bar{Y}_2^{(m)}, \bar{Y}_3^{(m)}, \ldots$ of a random process $Y_1, Y_2, Y_3, \ldots$ is its output process through the filter $g[n;m] \doteq (1/m) \cdot \mathbf{1}\{0 \le n < m\}$, where $\mathbf{1}\{\cdot\}$ is the set indicator function (cf. Fig. 2.2). It is named the *variance-equivalent m-averaged process* because its marginal variance is equal to that of the m-average process $\mathbf{Y}^{(m)}$.

The autocovariance function $\bar{C}_m(k)$ of the variance-equivalent m-averaged process can be given by:

$$\begin{aligned} \bar{C}_m(k) &= E\left[\bar{Y}_{i+k}^{(m)} \bar{Y}_i^{(m)}\right] \\ &= E\left[\left(\frac{Y_{(i+k)+1} \cdots Y_{(i+k)+m}}{m}\right)\left(\frac{Y_{i+1} \cdots Y_{i+m}}{m}\right)\right] \\ &= \sum_{i=-\infty}^{\infty} \bar{C}_1(i) \cdot \pi(k-i), \end{aligned}$$

where

$$\pi(i) \doteq \frac{m - |i|}{m^2} \cdot \mathbf{1}\{|i| \le m\}.$$

Thus, the power spectrum of the variance-equivalent m-averaged process is equal to

$$S_y(\omega) \frac{\sin^2(m\omega/2)}{m^2 \sin^2(\omega/2)},$$

and the variance of the m-averaged process of $\mathbf{Y}$ is given by:

$$C_m(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_y(\omega)\frac{\sin^2(m\omega/2)}{m^2\sin^2(\omega/2)}d\omega = \frac{2^{2-2H}}{\pi}\int_0^{\pi/2}\frac{\sin^2(m\omega)}{m^2\sin^{2H+1}(\omega)}d\omega,$$

which immediately gives:

$$\log\frac{C_m(0)}{C_1(0)} = \log\frac{\displaystyle\int_0^{\pi/2}\frac{\sin^2(m\omega)}{m^2\sin^{2H+1}(\omega)}d\omega}{\displaystyle\int_0^{\pi/2}\sin^{1-2H}(\omega)d\omega} = \log\frac{2\Gamma(1.5-H)\displaystyle\int_0^{\pi/2}\frac{\sin^2(m\omega)}{m^2\sin^{2H+1}(\omega)}d\omega}{\Gamma(1-H)\sqrt{\pi}},$$

where $\Gamma(\cdot)$ is the Euler gamma function defined as $\Gamma(n) \doteq \int_0^\infty t^{n-1}e^{-t}dt$. Based on the above formula, we depict the relation between $\log[C_m(0)/C_1(0)]$ and $\log(m)$ in Fig. 2.3, and observe a perfect self-similarity from the aspect of variance-time analysis even for very small m.

In fact, we can analytically obtain a lower and an upper bounds that hold for every m for $\log[C_m(0)/C_1(0)]$ through two inequalities:

$$\int_0^{\pi/2}\frac{\sin^2(m\omega)}{m^2\sin^{2H+1}(\omega)}d\omega \geq m^{2H-2}\frac{(2/\pi)^{2H}}{2(1-H)}$$

and

$$\int_0^{\pi/2}\frac{\sin^2(m\omega)}{m^2\sin^{2H+1}(\omega)}d\omega \leq m^{2H-2}\frac{(1+2H\pi)[2^{-2H}-(1-H)]\pi^2}{8H^2(2H-1)(1-H)},$$

and they again confirm the almost perfect self-similarity of the filter output process (cf. Fig. 2.4).

After the verification of self-similarity of the filter output process, it remains to design a filter whose output spectrum due to an i.i.d. input of unity power spectrum equals $S_y(\omega)$, or specifically, $|H(\omega)|^2 = |1-e^{-jw}|^{1-2H}$. We note that the z-transforms, $X(z)$ and $Y(z)$, of the filter input and output can be characterized by $(1-z^{-1})^{-a}X(z) = Y(z)$, where $a \doteq (2H-1)/2$. By Taylor's expansion, we obtain:

$$(1-z)^{-a} = 1 + \frac{a}{1!}z + \frac{a(a+1)}{2!}z^2 + \cdots = \sum_{n=0}^{\infty}\frac{\Gamma(n+a)}{\Gamma(n+1)\Gamma(a)}z^n.$$

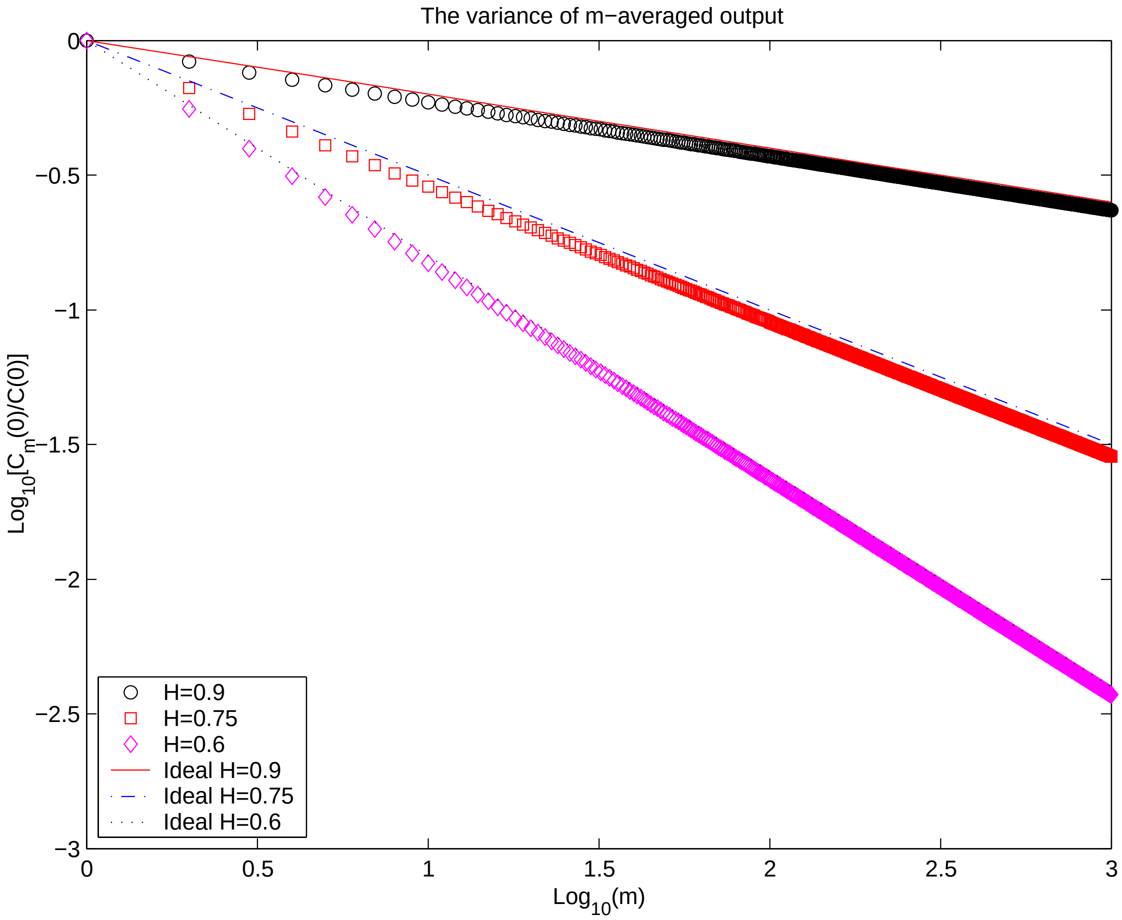

Figure 2.3: The variance-time analysis of the filter output process.

Therefore, the outputs $y[1], y[2], y[3] \ldots$ can be obtained through

$$y[n] \quad = \quad \sum_{k=0}^{\infty} \frac{\Gamma(k+a)}{\Gamma(k+1)\Gamma(a)} x[n-k] = \sum_{k=0}^{\infty} h[k] \cdot x[n-k],$$

where

$$h[n] \doteq \frac{\Gamma(n+a)}{\Gamma(n+1)\Gamma(a)} = \frac{\Gamma(n+H-0.5)}{\Gamma(n+1)\Gamma(H-0.5)} \qquad \text{for } k \geq 0.$$

Two problems will be encountered when one wishes to synthesize a self-similar network packet-arrival traffic in terms of the proposed filter system. Firstly, it is of infeasibly infinite length. Secondly, the filter outputs are in general non-integer-values even if the filter inputs are integer-values. Modifications such as filter truncation to finite length and rounding to

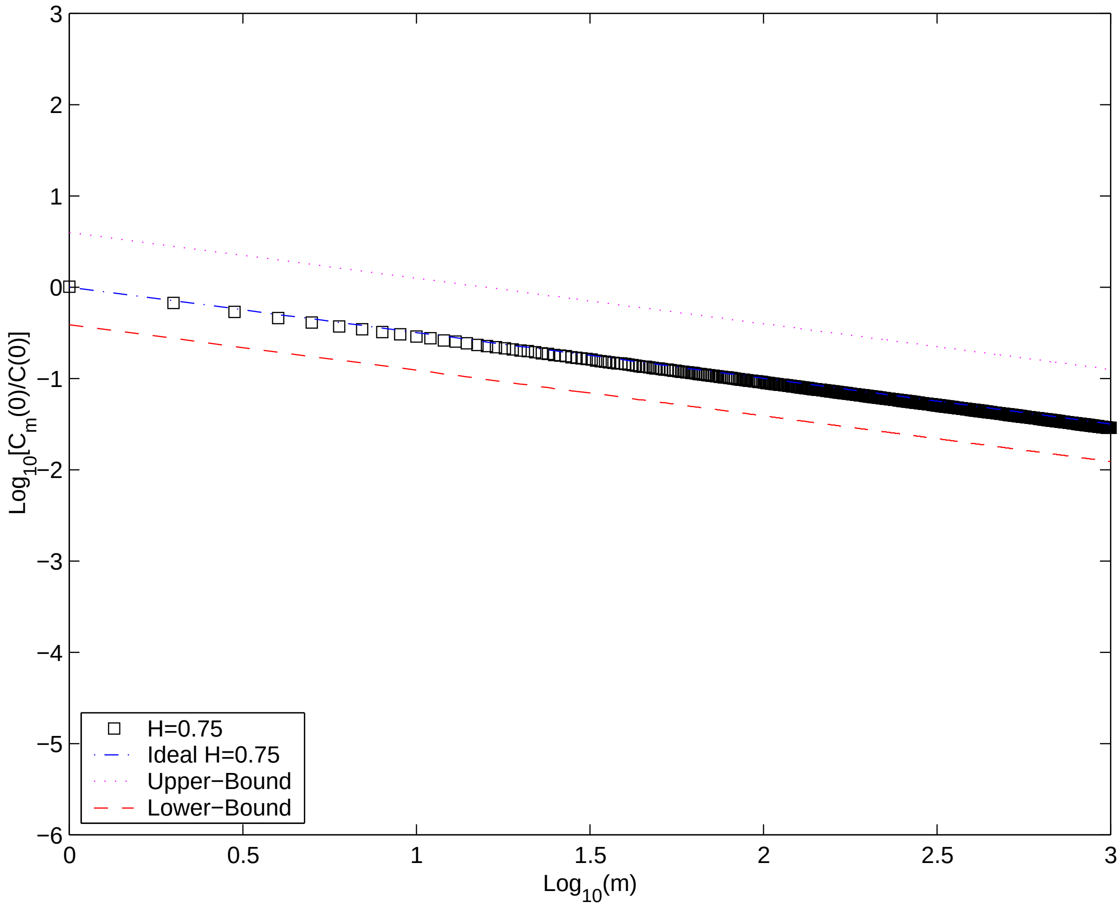

Figure 2.4: The lower and the upper bounds of $\log[C_m(0)/C_1(0)]$.

the nearest integers are therefore necessary. We will numerically examine the impact on self-similarity due to filter truncation and output rounding in later subsections.

2.1.2 Impact On Self-Similarity Due To Filter Truncation

Define $h[k;W] \doteq h[k] \cdot \mathbf{1}\{0 \leq k < W\}$. Then, the impact of the truncation window size W on the degree of self-similarity of the filter output process can be characterized through the derivation of the marginal variance $C_m(0;W)$ of the respective m-averaged filter output process. Again, we derive $C_m(0;W)$ through the help of the technique of the variance-equivalent m-average process.

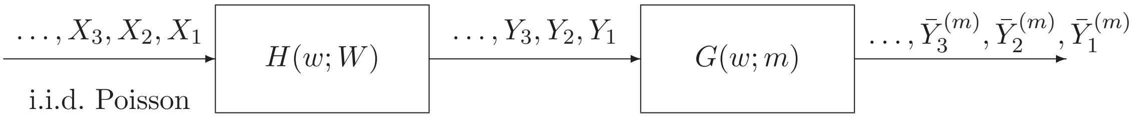

Figure 2.5: The *variance-equivalent m-averaged process* of the truncated filter output process.

Let $G(\omega;m)$ be the transfer function of the filter $g[n;m]$, and let $L(\omega;W,m) \doteq H(w;W)G(w;m)$. Then,

$$\ell[n;W,m] = \sum_{i=0}^{n} g[i;m] \times h[n-i;W] = \frac{1}{m} \sum_{i=\max\{0,n-W+1\}}^{\min\{n,m-1\}} h[n-i].$$

By letting $S_y(w;W)$ be the truncated counterpart of $S_y(\omega)$, we obtain:

$$\begin{aligned} C_m(0;W) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} S_y(w;W) d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\sum_{n=0}^{\infty} \ell[n] e^{-jnw} \right] \left[\sum_{n=0}^{\infty} \ell[n] e^{jnw} \right] d\omega \\ &= \sum_{n=0}^{\infty} |\ell[n]|^2 \\ &= \frac{1}{m^2} \left\{ \sum_{l=0}^{W-m} \left(\sum_{n=l}^{l+m-1} h[n] \right)^2 + \sum_{l=0}^{m-2} \left[\left(\sum_{n=0}^{l} h[n] \right)^2 + \left(\sum_{n=W-1-l}^{W-1} h[n] \right)^2 \right] \right\}. \end{aligned}$$

Based on the above formula, we numerically depict $\log_{10}[C_m(0;W)/C_1(0;W)]$ versus $\log_{10}(m)$ in Fig. 2.6, and observe that there are two apparent different self-similar behaviors for different m values. The resultant degree of self-similarity is close to the target one when $m \leq W$, but the slope of the variance-time curve quickly turns to a non-self-similar value, -1, once $m > W$.

2.1.3 Impact On Self-Similarity Due To Output Rounding

In this subsection, we further empirically examine the output rounding effect on self-similarity. Table 2.1 lists the resultant Hurst parameter of the trace synthesized according to the sys-

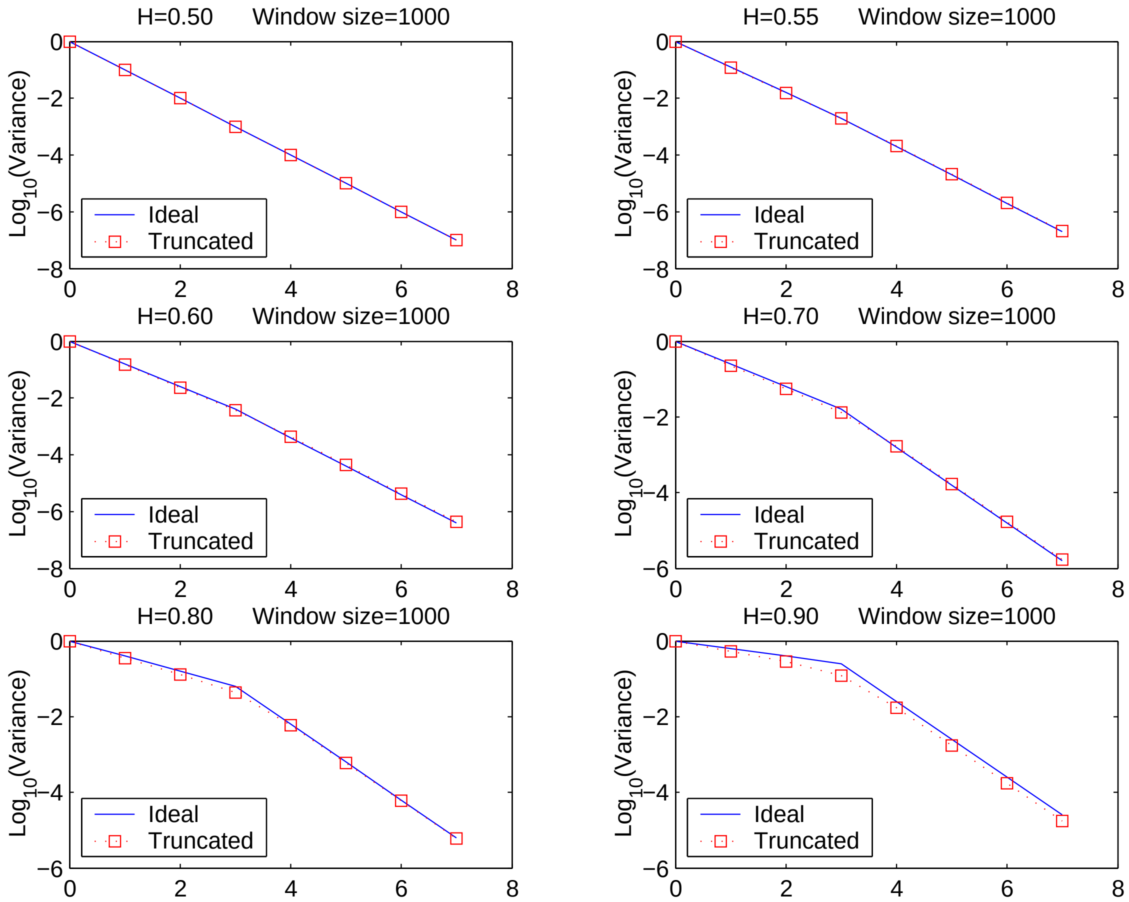

Figure 2.6: Variance-time analysis ($\log_{10}$ scale) for the truncated-filter output with truncation window $W = 10^3$. The slope of the solid line is equal to $2H - 2$ for $m \leq W$, and -1 for $m > W$.

tem in Fig. 2.7. It indicates that the rounding-to-the-nearest-integer operation on the filter output will have "unstable" impact on the degree of self-similarity of the output trace. Our simulations suggests that such an unstable impact can be neglected if the ratio of the maximal rounding error (i.e., 0.5) against the input mean λ is made less than 5%.

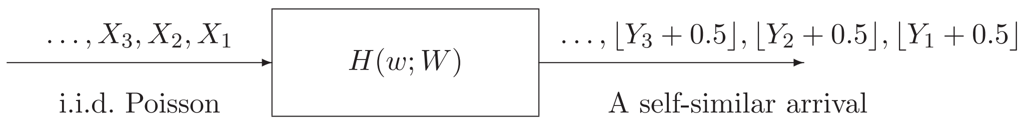

Figure 2.7: The proposed asymptotic self-similar traffic synthesizer. $H(w;W)$ represents a truncated version of $H(\omega)$ with truncation window W. The quantity $\lfloor Y_i + 0.5\rfloor$ equals the closest integer to Y_i.

Table 2.1: Comparison between the resultant Hurst parameters of the traces synthesized by the filter-based algorithm and the targeted ideal Hurst parameters.

Window size= 10000		
Ideal H	V-T($\lambda = 1$)	V-T($\lambda = 10$)
0.5001	0.4898783	0.5064982
0.55	0.5504289	0.5344366
0.6	0.6413529	0.5641452
0.7	0.4775099	0.7013537
0.8	0.5399816	0.7799114
0.9	0.5958403	0.8716414

2.2 The Reverse Filter Versus The Forward Filter

It can be easily seen that the z-transforms, $X(z)$ and $Y(z)$, of the filter input and output can be re-characterized by $(1-z^{-1})^a Y(z) = X(z)$. Again, by Taylor's expansion,

$$(1-z^{-1})^a = 1 + \frac{-a}{1!}z^{-1} + \frac{-a(1-a)}{2!}z^{-2} + \cdots = 1 - a\sum_{n=1}^{\infty}\frac{\Gamma(n-a)}{\Gamma(n+1)\Gamma(1-a)}z^{-n}.$$

Hence, the outputs $y[1], y[2], y[3]\ldots$ can be also obtained through an infinite impulse response (IIR) filter as:

$$y[n] = x[n] + a\sum_{k=1}^{\infty}\frac{\Gamma(k-a)}{\Gamma(k+1)\Gamma(1-a)}y[n-k] = x[n] + \sum_{k=1}^{\infty}h'[k]\cdot y[n-k],$$

where

$$h'[n] \doteq \frac{a\cdot\Gamma(n-a)}{\Gamma(n+1)\Gamma(1-a)} = \frac{(H-0.5)\cdot\Gamma(n-H+0.5)}{\Gamma(1.5-H)\Gamma(n+1)} \quad \text{for } k \geq 1.$$

We refer $h[\cdot]$ as the *forward filter* and $h'[\cdot]$ as the *reverse filter*, since the latter has a feedback or reverse path. Both $h[\cdot]$ system and $h'[\cdot]$ system can generate a *true* self-similar process in response to, say, an i.i.d. Poisson input; however, unlike the forward filter, the reverse filter gives an infinite impulse response filter (IIR) even if a finite truncation on $h'[\cdot]$ is applied. This may give a false impression that the reverse system equipped with an infinite impulse response (IIR) filter of *finite* number of coefficients can synthesize a more self-similar trace than the forward system with truncated forward filter of the same computational complexity (or more specifically, the same truncation window). Our simulations, however, indicate that the effective range of both filters are actually similar (cf. Fig. 2.8).

2.3 Concluding Remarks

In this chapter, a new model is proposed for the synthesization of self-similar traffics based on the filter technique. The synthesized trace can be made long-range dependent with adjustable levels of bustiness and correlation. Only three parameters need to be specified in our model: H is the targeted self-similar parameter that controls the bustiness and correlation of the synthetic traffic, λ defines the mean of the synthesized traffic, and W determines not only the length of the filter (which in turns determines the algorithmic complexity) but also the valid aggregation size of self-similar nature from the aspect of variance-time analysis.

When being compared to the two known self-similar traffic synthesizers—*random midpoint displacement* and *Paxson's spectrum fitting*, our model provides advantages that the synthetic traffic can be generated on the fly, and is always non-negative. The algorithmic complexity of *Paxon's spectrum fitting* was shown to be less than the random midpoint displacement, and is given by $(n/2)\log_2(n+2)$, where n is the length of the synthetic trace. The complexity of our model, however, is dependent on W, and is equal to $n \times W$. Hence, when the valid aggregation size of self-similar nature is specified, the complexity of our model only

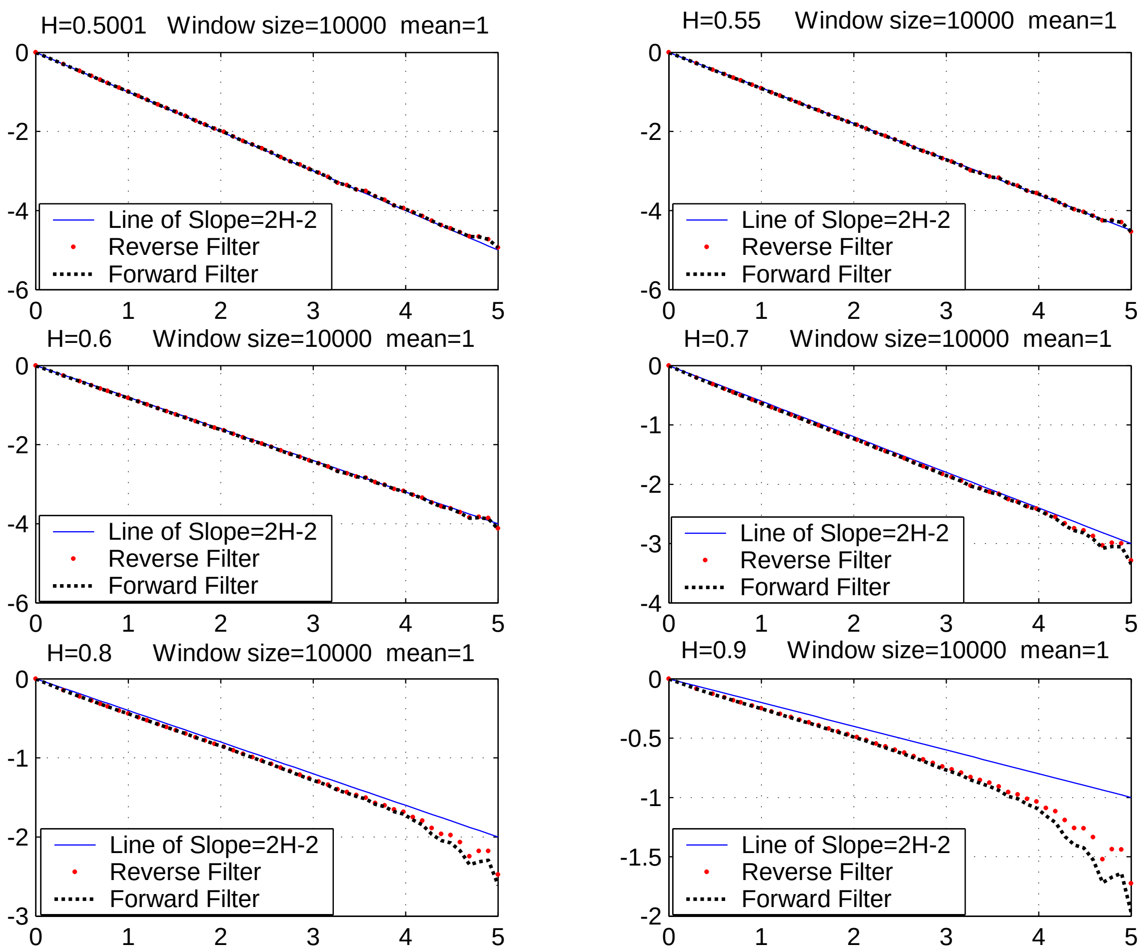

Figure 2.8: Variance-time plots ($\log_{10}$ scale) for the two filter-based synthetic arrivals with truncation window 10^4 and mean rate 1.

grows linearly with the trace size.

Chapter 3

Correlation Approximation to the Mutual Information of Self-Similar Processes

3.1 Introduction

Mutual information and correlation coefficient are both used as measures of dependance between random sources [27]. Generally speaking, the correlation coefficient only measures the *linear* dependance, while mutual information can represent the *general* dependance [28]. Thus, in the sense of generality, mutual information is a somewhat better quantity to measure the dependance than the correlation coefficient. However, estimating the mutual information function is much more difficult than estimating the correlation coefficient, as it requires a complete knowledge about the distributions.

In this chapter, we focus on the following question: Given the correlation coefficients of random sources, what is the minimum possible value of mutual information? An upper bound and a lower bound of this minimum possible value were established in situation where the correlation coefficients are small. It was subsequently shown that both bounds can be approximated by half the square of the correlation coefficient when the two random sources are both one dimensional. When the random sources are multidimensional, we found that this minimum mutual information function can be approximated by half the square of the Frobenius norm of the cross correlation coefficient matrix. We also address some examples to show the accuracy of these approximative bounds.

3.2 Definitions and Notations

Definition 3.1. Given two random sources X and Y (not necessarily random variables or random vectors), the mutual information function is defined as:

$$I\left(P_{X,Y}\right) \text{ or } I(X;Y) \triangleq \sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{\left(\sum_z P_{X,Y}(z,y)\right)\left(\sum_w P_{X,Y}(x,w)\right)},$$

where $P_{X,Y}(x,y)$ is the probability of the event $(X,Y)=(x,y)$.

Definition 3.2. The divergence function of P_X against Q_X is defined as:

$$D\left(P_X \| Q_X\right) \triangleq \sum_x P_X(x) \log \frac{P_X(x)}{Q_X(x)},$$

where $P_X(x)$ and $Q_X(x)$ are two probability mass functions, and the support of P_X is contained in the support of Q_X.

A straightforward consequence of the above definitions is that:

$$I\left(P_{X,Y}\right) = D\left(P_{X,Y} \| P_X \times P_Y\right),$$

where $P_X(x) \triangleq \sum_w P_{X,Y}(x,w)$ and $P_Y(y) \triangleq \sum_z P_{X,Y}(z,y)$.

Definition 3.3. (Minimum mutual information of a probability set) The minimum mutual information function with respect to a set $\mathcal{S}$ of probabilities is defined as:

$$I_{\min}\left(\mathcal{S}\right) \triangleq \min_{P_{X,Y} \in \mathcal{S}} I\left(P_{X,Y}\right),$$

where $P_{X,Y}$ is the probability mass function of X and Y chosen from the set $\mathcal{S}$.

Definition 3.4. (Minimum divergence function of a probability set and two marginal distributions) The minimum divergence function with respect to a set $\mathcal{S}$ of probabilities and two marginal distributions, P_X and P_Y, is defined as:

$$D_{\min}\left(S, P_X, P_Y\right) \triangleq \min_{Q_{X,Y}\in S} D\left(Q_{X,Y} \| P_X \times P_Y\right),$$

where $Q_{X,Y}$ is the probability mass function of X and Y chosen from the set $\mathcal{S}$.

Definition 3.5. (Correlation coefficient matrix) Given two random vectors, $\vec{X} = (X_1, ..., X_n)$ and $\vec{Y} = (Y_1, ..., Y_m)$, the (i, j)-component of the correlation coefficient matrix of $\vec{X}$ and $\vec{Y}$ is defined as:

$$C_{\vec{X},\vec{Y}}(i, j) \triangleq E[\hat{X}_i \hat{Y}_j],$$

where $(\hat{X}_1, ..., \hat{X}_n)$ is the Karhunen-Loeve transformation of $\vec{X}$, and each of $(\hat{X}_1, ..., \hat{X}_n)$ has zero mean and unity variance and is uncorrelated to the others, and $(\hat{Y}_1, ..., \hat{Y}_n)$ is defined similarly with respect to $\vec{Y}$.

Since Karhunen-Loeve transformation is invertible,

$$I\left(X_1, ..., X_n; Y_1, ..., Y_m\right) = I(\hat{X}_1, ..., \hat{X}_n; \hat{Y}_1, ..., \hat{Y}_n).$$

To simplify the proof in later section, we will assume that the considered $\vec{X}$ and $\vec{Y}$ are already their Karhunen-Loeve transformation counterparts that satisfy the conditions of uncorrelatedness, zero-mean and unity variance.

Definition 3.6. (Frobenius norm) The Frobenius norm of a matrix C is defined as:

$$\|C\| \triangleq \left(\sum_{i,j} C^2(i,j)\right)^{1/2},$$

where $C(i, j)$ is the (i, j)-component of the matrix C.

3.3 Main Theorems

Theorem 3.1. For two bounded[1] random variables X and Y respectively with marginal distributions P_X and P_Y,

$$I_{\min}(\mathcal{S}_\rho) = \frac{\rho^2}{2} + o(\rho^2), \text{ as } \rho \to 0,$$

where

$$\mathcal{S}_\rho \triangleq \{Q_{X,Y} : Q_X = P_X, Q_Y = P_Y \text{ and } E_Q[XY] = \rho\},$$

Q_X and Q_Y are the marginal distributions of $Q_{X,Y}$, $E_Q[\cdot]$ denotes that the expectation value is calculated according to distribution $Q_{X,Y}$, and $o(\cdot)$ is the little-o notation [29, pp. 286].

Note that $\mathcal{S}_\rho = \mathcal{S}_\rho(P_X, P_Y)$ is actually a function of the marginal distributions of P_X and P_Y. For convenience, we drop "(P_X, P_Y)" in the notation, and reserve P_X and P_Y to always denote given marginals.

It can be shown that $I_{\min}(\mathcal{S}_\rho)$ is a convex function of ρ [26]. Specifically,

$$\begin{aligned}
I_{\min}\left(\mathcal{S}_{\lambda\rho_1+(1-\lambda)\rho_2}\right) &\le I_{\min}(\lambda\mathcal{S}_{\rho_1} + (1-\lambda)\mathcal{S}_{\rho_2}) \\
&= D_{\min}(\lambda\mathcal{S}_{\rho_1} + (1-\lambda)\mathcal{S}_{\rho_2}, P_X, P_Y) \\
&= \min_{\substack{Q^{(1)}_{X,Y}\in\mathcal{S}_{\rho_1} \\ Q^{(2)}_{X,Y}\in\mathcal{S}_{\rho_2}}} D\left(\lambda Q^{(1)}_{X,Y} + (1-\lambda)Q^{(2)}_{X,Y} \middle\| P_X \times P_Y\right) \\
&\le \min_{Q^{(1)}_{X,Y}\in\mathcal{S}_{\rho_1}, Q^{(2)}_{X,Y}\in\mathcal{S}_{\rho_2}} \left[\lambda D\left(Q^{(1)}_{X,Y} \middle\| P_X \times P_Y\right)\right. \\
&\quad \left. +(1-\lambda)D\left(Q^{(2)}_{X,Y} \middle\| P_X \times P_Y\right)\right] \\
&= \lambda I_{\min}(\mathcal{S}_{\rho_1}) + (1-\lambda)I_{\min}(\mathcal{S}_{\rho_2}).
\end{aligned}$$

We now proceed to prove the theorem.

---

[1] By "boundedness", we mean that there exists $B > 0$ such that $P_X[x \in \Re : |x| < B] = P_Y[y \in \Re : |y| < B] = 1$.

Proof. In this proof, we first find a lower bound of $I_{\min}(\mathcal{S}_\rho)$. Then we use a specific distribution contained in $\mathcal{S}_\rho$ to form an upper bound of $I_{\min}(\mathcal{S}_\rho)$. The theorem is then proved since both bounds have the form $\rho^2/2 + o(\rho^2)$ as $\rho \to 0$.

Define a set $\mathcal{T}_\rho$ as: $\mathcal{T}_\rho \triangleq \{Q_{X,Y} : E_Q[XY] = \rho\}$. From the standpoint of mutual information and Karhunen-Loeve transformation, we can assume without loss of generality that both P_X and P_Y have zero marginal mean and unity marginal variance, and they are uncorrelated.

Since $\mathcal{S}_\rho \subset \mathcal{T}_\rho$, $I_{\min}(\mathcal{S}_\rho) = D_{\min}(\mathcal{S}_\rho, P_X, P_Y) \geq D_{\min}(\mathcal{T}_\rho, P_X, P_Y)$. Now, we apply the Lagrange multiplier method to evaluate $D_{\min}(\mathcal{T}_\rho, P_X, P_Y)$, i.e., to minimize

$$F(Q_{X,Y}) = D(Q_{X,Y} \| P_X \times P_Y) - \beta\left(\sum_{x,y} xyQ_{X,Y}(x,y) - \rho\right) - \theta\left(\sum_{x,y} Q_{X,Y}(x,y) - 1\right),$$

subject to the following restrictive conditions:

$$\sum_{x,y} xyQ_{X,Y}(x,y) = \rho, \tag{3.1}$$

and

$$\sum_{x,y} Q_{X,Y}(x,y) = 1.$$

We then take the derivative of $F(Q_{X,Y})$ with respect to $Q_{X,Y}(x,y)$, and obtain

$$\frac{\partial F(Q_{X,Y})}{\partial Q_{X,Y}(x,y)} = 1 + \log\frac{Q_{X,Y}(x,y)}{P_X(x)P_Y(y)} - \beta xy - \theta.$$

Letting the above derivative be zero, we have that the optimal $Q_{X,Y}$ must satisfy:

$$Q_{X,Y}(x,y) = \frac{P_X(x)P_Y(y)\exp\{\beta xy\}}{\sum_{u,v} P_X(u)P_Y(v)\exp\{\beta uv\}} = \frac{P_X(x)P_Y(y)\exp\{\beta xy\}}{M(\beta)},$$

where

$$M(\beta) = \sum_{u,v} P_X(u)P_Y(v)\exp\{\beta uv\}.$$

Taking the above result into $D_{\min}(\mathcal{T}_\rho, P_X, P_Y)$ yields

$$D_{\min}(\mathcal{T}_\rho, P_X, P_Y) = \beta\rho - C(\beta),$$

where $C(\beta) = \log M(\beta)$. Denote $f(\beta) = \partial C(\beta)/\partial\beta$, and observe that

$$f(\beta) = \sum_{x,y} xyQ_{X,Y}(x,y) = E_Q[XY].$$

Hence, the restrictive condition in (3.1) can be written as $f(\beta) = \rho$.

Since X and Y are bounded with respect to distributions P_X and P_Y, respectively, the moment generating function $M(\beta)$ is defined throughout an interval $(-\beta_0, \beta_0)$ for some $\beta_0 > 0$, which implies that moments of all orders are finite (namely, $\sum_{x,y} x^i y^j P_X(x) P_Y(y) < \infty$ for $i \geq 1$ and $j \geq 1$), and $M(\beta)$ has a Taylor expansion about origin with positive radius of convergence [30, pp. 278], and so do $C(\beta)$ and $f(\beta)$. Using the Taylor expansions of $f(\beta)$ and $C(\beta)$, we have

$$\begin{aligned} f(\beta) &= f(0) + f'(0)\beta + \frac{f''(0)}{2!}\beta^2 + o(\beta^2) \\ &= \beta + \frac{\gamma}{2}\beta^2 + o(\beta^2), \end{aligned}$$

and

$$\begin{aligned} C(\beta) &= C(0) + C'(0)\beta + \frac{C''(0)}{2!}\beta^2 + \frac{C'''(0)}{3!}\beta^3 + o(\beta^3) \\ &= \frac{\beta^2}{2} + \frac{\gamma}{6}\beta^3 + o(\beta^3), \end{aligned}$$

where $\gamma = \sum_{x,y} x^3 y^3 P_X(x) P_Y(y)$. Accordingly, $\beta = \rho + o(\rho^2)$, and $C(\beta) = \rho^2/2 + o(\rho^3)$. This immediately concludes $D_{\min}(\mathcal{T}_\rho, P_X, P_Y) = \beta\rho - C(\beta) = \rho^2/2 + o(\rho^3)$.

Now, we turn to the task of finding an upper bound. It suffices to use a trial distribution in $\mathcal{S}_\rho$ to form an upper bound. Define this trial distribution as:

$$J_{X,Y}(x,y) \triangleq P_X(x)P_Y(y)(1 + \rho xy),$$

and examine[2] that $\sum_{x,y} J_{X,Y}(x,y) = 1$, $\sum_y J_{X,Y}(x,y) = P_X(x)$, $\sum_x J_{X,Y}(x,y) = P_Y(y)$, and $\sum_{x,y} xyJ_{X,Y}(x,y) = \rho$. Using the inequality

$$x - \frac{x^2}{2} + \frac{x^3}{3} \geq \log(1+x),$$

for $|x| < 1$, we have

$$\begin{aligned} I(J_{X,Y}) &= \sum_{x,y} P_X(x)P_Y(y)(1+\rho xy)\log(1+\rho xy) \\ &\leq \frac{\rho^2}{2} - \frac{\gamma}{6}\rho^3 + \frac{\eta}{3}\rho^4, \end{aligned}$$

where $\eta = \sum_{x,y} x^4 y^4 P_X(x)P_Y(y)$. Since $I(J_{X,Y}) \geq I_{\min}(\mathcal{S}_\rho) \geq D_{\min}(\mathcal{T}_\rho, P_X, P_Y)$, we have

$$I_{\min}(\mathcal{S}_\rho) = \frac{\rho^2}{2} + o(\rho^2).$$

□

Theorem 3.2. Consider two bounded random vectors $\vec{X} = (X_1, ..., X_n)$ and $\vec{Y} = (Y_1, ..., Y_m)$. If the correlation coefficient matrix C satisfies $|C(i,j)| < \rho$ for each $1 \leq i \leq n$ and $1 \leq j \leq m$, then

$$I_{\min}(\mathcal{S}_C) = \frac{1}{2}\|C\|^2 + o(\rho^2), \text{ as } \rho \to 0,$$

where

$$\mathcal{S}_C \triangleq \{Q_{\vec{X},\vec{Y}} : Q_{\vec{X}} = P_{\vec{X}}, Q_{\vec{Y}} = P_{\vec{Y}} \text{ and } E_Q[X_i Y_j] = C(i,j)\},$$

$Q_{\vec{X}}$ and $Q_{\vec{Y}}$ are the marginal distributions of $Q_{\vec{X},\vec{Y}}$, $E_Q[\cdot]$ denotes that the expectation value is calculated according to distribution $Q_{\vec{X},\vec{Y}}$, and $o(\cdot)$ is the little-o notation [29, pp. 286].

Proof. The proof is similar to the previous theorem; hence, it is omitted. □

[2]Notably, following footnote 1, we can guarantee that $0 \leq J_{X,Y}(x,y) \leq 1$ when $|\rho| \leq 1/B^2$,

3.4 Examples

To discuss the accuracy of the upper and lower bounds of the minimum mutual information function, we first consider the simple case that both random variables X and Y are binary random variables, each taking values from $\{0, 1\}$. In this case, given the correlation coefficient ρ and marginal mean, $a = E[X]$ and $b = E[Y]$, one can determine the joint distribution of X and Y, i.e.,

$$\begin{aligned} P_{X,Y}(0,0) &= (1-a)(1-b)+r \\ P_{X,Y}(0,1) &= (1-a)b-r \\ P_{X,Y}(1,0) &= a(1-b)-r \\ P_{X,Y}(1,1) &= ab+r \end{aligned}$$

where $r = E[XY] - E[X]E[Y] = \rho[a(1-a)b(1-b)]^{1/2}$. The mutual information $I(X;Y)$ can be written as

$$I(X;Y) = H_b(b) - aH_b\left(b+\frac{r}{a}\right) - (1-a)H_b\left(1-b+\frac{r}{1-a}\right),$$

where $H_b(b) = -b\log(b) - (1-b)\log(1-b)$ is the binary entropy function. Thus, in binary case, $I_{\min}(\mathcal{S}_\rho) = I(X;Y)$. We then take the uniform marginal distributions as an example, i.e., $a = \frac{1}{2}$ and $b = \frac{1}{2}$, and obtain $D_{\min}(\mathcal{T}_\rho, P_X, P_Y) = \rho\tanh^{-1}(\rho) + \frac{1}{2}\log(1-\rho^2) = I(X;Y)$. Therefore, the lower bound used in the proof coincides with the minimum mutual information function. Notably, the simple binary case has already been examined in [28].

A good example that meets the boundedness assumption of our theorem is the Morgenstern distribution [31] that has the density of

$$p(x,y) = 1 + \alpha(2x+1)(2y+1),$$

where $0 \le x, y \le 1$, and its correlation coefficient equals $C_{x,y} = \alpha/3$. Its asymptotic mutual information with respect to the correlation coefficient can be obtained easily as:

$$I(X;Y) = \frac{\alpha^2}{18} + \frac{\alpha^4}{300} + o(\alpha^5) = \frac{\rho^2}{2} + o(\rho^2).$$

An example that can be used to show that $I_{\min}(\mathcal{S}_\rho)$ is indeed a lower bound to the mutual information of $P_{X,Y} \in \mathcal{S}_\rho$ is the bivariate density $p(x,y) = p_{Y|X}(y|x)p_X(x)$, where $p_X(x) = \frac{1}{2a} \cdot \mathbf{1}[|X| \le a]$ and $p_{Y|X}(y|x) = \frac{1}{2b} \cdot \mathbf{1}[|Y - \alpha X| \le b]$, which exactly define a *uniform diagonal strip*. The asymptotic mutual information of the uniform diagonal strip can be derived easily from [31] as $\frac{|\rho|}{2} - \frac{|\rho|^3}{4} + o(|\rho|^4)$. This indicates that in some situations, $I(X;Y) > I_{\min}(\mathcal{S}_\rho)$.

The validity of the theorem statement can be extended to the (unbounded) case that P_X and P_Y are both Gaussian distributed. In this case, the minimum value of mutual information can be achieved by a jointly Gaussian distributed $Q_{X,Y}$. One can derive that for Gaussian P_X and P_Y, $I_{\min}(\mathcal{S}_\rho) = -\frac{1}{2}\log(1-\rho^2)$. The lower bound, however, is given by:

$$D_{\min}(\mathcal{T}_\rho, P_X, P_Y) \;\; = -\frac{1}{2} + \left(\frac{1}{4} + \rho^2\right)^{\frac{1}{2}} + \frac{1}{2}\log\left(\frac{-\frac{1}{2} + (\frac{1}{4}+\rho^2)^{\frac{1}{2}}}{\rho^2}\right),$$

and is smaller than the simple hyperbolic approximation of $I_{\min}(\mathcal{S}_\rho) \approx \frac{\rho^2}{2}$. In addition, the "upper bound" used in the proof $\frac{\rho^2}{2} + \rho^4$ may become smaller than $I_{\min}(\mathcal{S}_\rho)$ at large $|\rho|$. Since we only use the upper bound under $|\rho| \ll 1$, we would not expect it to be useful outside the concerned range.

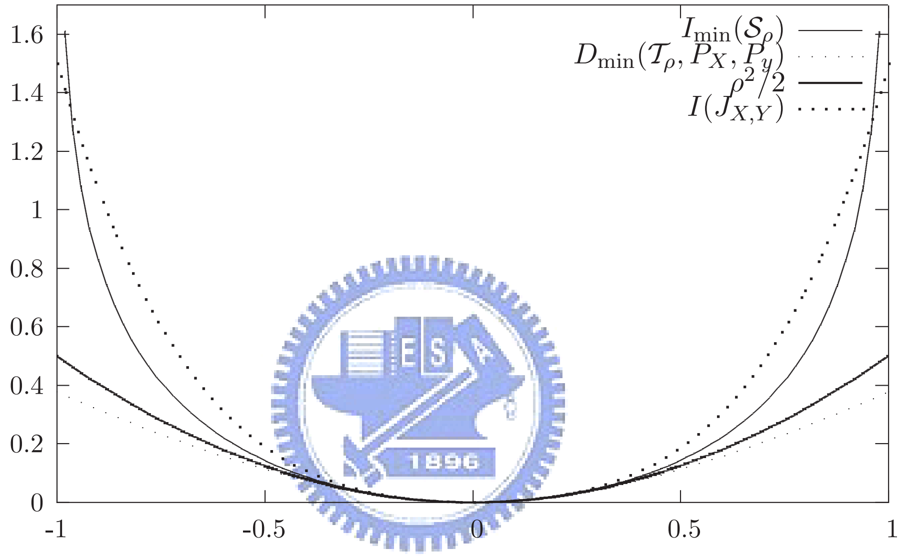

Figure 3.1: The bounds and minimum mutual information for Gaussian distributed P_X and P_Y.

Chapter 4

Bayesian Decentralized Detection for Exponential Distributions

A decentralized detection system consists of n sensors, sometimes geographically dispersed, and a remote fusion center. Each of the sensor observes a phenomenon (often modeled as a random variable X_i), summarizes it into a single bit u_i, and then transmits u_i to the fusion center uncooperatively. Based on received $\{u_i\}_{i=1}^n$, the fusion center determines whether $\{X_i\}_{i=1}^n$ are drawn from the null distribution $P(\cdot|H_0)$ or the alternative distribution $P(\cdot|H_1)$.

Decentralized detection, despite that it has a simple scenario, and has been studied extensively for more than two decades, still has many unsolved issues in the fundamental level. One of these unsolved issues concerns the global optimal strategy for the design of sensors and the fusion center. The difficulties comes from several points. Firstly, only the necessary conditions for the optimal strategy are known; hence, one have to search all the solutions to the equations of the necessary conditions in order to determine the global optimum. Moreover, these equations are coupled and nonlinear, and hence, to solve them is proved to be a hard mission [32]. The knowledge about the global optimal strategy is so little that there are almost no analytical results for the system with more than two sensors. The asymptotic results, however, had been found more pleasantly: the system with identical sensors has the same exponents of error probabilities as the optimal system [37]; the ratio of error probabilities between these two systems are shown bounded from above and from below [36]. Yet the exact and analytical results for the system with some finite $n > 2$ are still absent, although such results will give us more insight about the global optimum than the asymptotic results.

In this chapter, we analyze the decentralized classification problem for exponential sources for $n > 2$, and validate an intuition that the optimal system is the system with identical sensors. To our knowledge, there is no similar analytical result for the global optimum for the system with more than two sensors.

4.1 Preliminaries

Definition 4.1. If X is a random variable with an exponential distribution, then the probability that X is greater than some number x is given by

$$1 - F_X(x) = \Pr(X > x) = e^{-\alpha x}$$

for $x \geq 0$, where α is a positive parameter, and $F_X(x)$ is the cumulative distribution function (CDF) of X.

It follows that the probability density function (PDF) of an exponential distribution has the form

$$f_X(x) = \alpha e^{-\alpha x},$$

for $x \geq 0$.

In this chapter, we concern the following binary hypothesis testing problem for exponential distributions:

$$H_1 : f_X(x_i) = \beta e^{-\beta x_i}$$

versus

$$H_0 : g_X(x_i) = \gamma e^{-\gamma x_i},$$

or equivalently,

$$H_1 : F_X(x_i) = 1 - e^{-\beta x_i}$$

versus

$$H_0 : G_X(x_i) = 1 - e^{-\gamma x_i},$$

for $i = 1, 2, \ldots, n$, $\beta < \gamma$, and for $x_i \geq 0$, where x_i is the observed value of the random variable X_i at the i-th sensor. We assume that $\{X_i\}_{i=1}^{n}$ form a set of independent and identically distributed (i.i.d.) random variables. The prior probabilities of H_1 and H_0 are denoted as r_1 and r_0 or simply r and $1 - r$, respectively. For a fixed fusion rule, it is known that the optimal local decision rule for each sensor is the local likelihood ratio test (LLRT), i.e.,

$$\frac{f_X(x_i)}{g_X(x_i)} = \frac{\beta e^{-\beta x_i}}{\gamma e^{-\gamma x_i}} \underset{u_i=0}{\overset{u_i=1}{\gtrless}} \lambda_i$$

or equivalently,

$$x_i \underset{u_i=0}{\overset{u_i=1}{\gtrless}} t_i$$

for $i = 1, 2, \ldots, n$, where u_i is the decision of i-th sensor, $t_i = \frac{1}{\gamma-\beta} \log(\frac{\lambda_i}{\xi})$ is some constant threshold to be decided, and $\xi = \frac{\beta}{\gamma}$. Let $P_D(\lambda_i)$ and $P_F(\lambda_i)$ denote respectively the detection probability and the false alarm probability for the i-th sensor, where

$$P_D(\lambda_i) = \mathtt{Prob}(u_i = 1|H_1) = \frac{1}{\tilde{\lambda}_i^{\tilde{\beta}}}$$

and

$$P_F(\lambda_i) = \mathtt{Prob}(u_i = 1|H_0) = \frac{1}{\tilde{\lambda}_i^{\tilde{\gamma}}}.$$

Both are functions of the LLRT threshold λ_i as $\tilde{\lambda}_i = \frac{\lambda_i}{\xi}$, $\tilde{\beta} = \frac{\beta}{\gamma-\beta}$ and $\tilde{\gamma} = \frac{\gamma}{\gamma-\beta}$. Notably, $1 \leq \hat{\lambda} \leq \infty$, $\hat{\gamma} = \hat{\beta} + 1$ and $\frac{dP_D}{dP_F} = \xi\hat{\lambda}$. Moreover, we can rewrite P_D and λ_i as functions of

P_F, i.e.,

$$P_D(P_F(i)) = P_F(i)^{\xi}$$

and

$$\lambda_i = \frac{dP_D(P_F(i))}{dP_F(i)} = \xi \frac{P_D(P_F(i))}{P_F(i)},$$

where we abuse the notations to let $P_F(i)$ and $P_D(P_F(i))$ represent the false alarm probability and the detection probability of the i-th sensor, respectively. The graph consists of all (P_F, P_D) pairs is referred to as Receiver Operating Characteristics (ROC curve), which is identical for all sensors since the statistics of their observations are all the same.

The sensors transmit their decisions $\{u_i\}_{i=1}^{n}$ to the fusion center that makes the final decision u_0, which equals ℓ when the fusion center favors H_ℓ. Once the fusion rule is fixed, we can then evaluate the system detection probability $Q_D(\lambda_1, \cdots, \lambda_n) = \texttt{Prob}(u_0 = 1|H_1)$, the system false alarm probability $Q_F(\lambda_1, \cdots, \lambda_n) = \texttt{Prob}(u_0 = 1|H_0)$ and the probability of error $P_e^{(n)}(\lambda_1, \cdots, \lambda_n) = r(1 - Q_D(\lambda_1, \cdots, \lambda_n)) + (1 - r)Q_F(\lambda_1, \cdots, \lambda_n)$ as functions of the local thresholds $\lambda_1, \cdots, \lambda_n$.

It is known from classical detection theory that the fusion center should make the overall decision u_0 based on the likelihood ratio test of received $u_1, , u_2 \ldots, u_n$. Therefore, the error probability can be expressed as

$$\begin{aligned} P_e^{(n)}(\lambda_1, \cdots, \lambda_n) \quad &= \sum_{u^n \in \{0,1\}^n} \min \left[r \left(1 - \prod_{i=1}^{n} P_D(\lambda_i)^{u_i} (1 - P_D(\lambda_i))^{1-u_i} \right), \right. \\ &\qquad \left. (1 - r) \prod_{i=1}^{n} P_F(\lambda_i)^{u_i} (1 - P_F(\lambda_i))^{1-u_i} \right]. \end{aligned}$$

The above formula, however, is in general not differentiable, and could give us little insight into the optimal choice of LLRT thresholds $(\lambda_1, \cdots, \lambda_n)$.

For identical sensor system design, it is known that the optimal fusion rule should be a

k-out-of-n rule,

$$u_0 = \begin{cases} 1, & \text{if } u_1 + \cdots + u_n \geq k \\ 0, & \text{if } u_1 + \cdots + u_n < k, \end{cases}$$

where k is some positive integer smaller or equal to n. However, to our knowledge, the validity of the converse statement, i.e., for any k-out-of-n fusion rule, the optimal strategy is to apply identical local decision rules for all sensors, is still unknown.

Now, let us define a function $A(\lambda)$, and prove a relevant lemma that is useful in the subsequent sections. Define a function $A(\lambda)$ as

$$A(\lambda) = \log \frac{\Pr(u = 1|H_1)}{\Pr(u = 1|H_0)} - \log \frac{\Pr(u = 0|H_1)}{\Pr(u = 0|H_0)}.$$

Then we have the following result.

Lemma 4.1. $A(\lambda)$ is a positive and monotonically increasing function of λ.

Proof. Firstly,

$$A(\lambda) = \log \left(\frac{P_D(\lambda)}{P_F(\lambda)} \right) - \log \left(\frac{1 - P_D(\lambda)}{1 - P_F(\lambda)} \right),$$

is positive because for the ROC curve,

$$\frac{P_D(\lambda)}{P_F(\lambda)} > \frac{1 - P_D(\lambda)}{1 - P_F(\lambda)}.$$

Taking derivative of $A(\lambda)$ with respect to λ, we obtain

$$\begin{aligned} A'(\lambda) &= (-P'_F(\lambda)) \left(-\frac{\partial A(\lambda)}{\partial P_F} \right) \\ &= (-P'_F(\lambda)) \left(-\left(\frac{\lambda}{P_D} - \frac{1}{P_F} \right) + \left(\frac{-\lambda}{1 - P_D} - \frac{-1}{1 - P_F} \right) \right) \\ &= \frac{-P'_F(\lambda)}{P_D(1 - P_D)} \left(\frac{P_D(1 - P_D)}{P_F(1 - P_F)} - \lambda \right) \\ &= \frac{-P'_F(\lambda)}{P_F(1 - P_D)} \left(\frac{1 - P_D}{1 - P_F} - \xi \right) \\ &\geq 0, \end{aligned}$$

where in the above derivation, we use $\lambda = \xi \frac{P_D}{P_F}$, $\frac{1 - P_D}{1 - P_F} \geq \xi$, and $P_F(\lambda)$ is a monotonically decreasing function of λ. $\square$

In terms of $A(\lambda)$ defined above, it can be shown that the likelihood ratio test of $u_1, , u_2 \ldots, u_n$ at the fusion center is equivalent to

$$\sum_{i=1}^{n} A(\lambda_i) u_i \underset{u_0=0}{\overset{u_0=1}{\gtrless}} \log\left(\frac{1-r}{r}\right) - \sum_{i=1}^{n} \frac{1-P_D(\lambda_i)}{1-P_F(\lambda_i)}. \tag{4.1}$$

4.2 System with one sensor

We start our analysis from the simplest case: A system with only one sensor. In such case, the only possible fusion rule is $u_0 = u_1$. This leads to that the system detection probability and the system false alarm probability are $Q_D(\lambda_1) = P_D(\lambda_1) = (\frac{\xi}{\lambda_1})^\beta$ and $Q_F(\lambda_1) = P_F(\lambda_1) = (\frac{\xi}{\lambda_1})^\gamma$. As a result, the system probability of error is given by

$$\begin{aligned} P_e^{(1)}(\lambda_1) &= r(1-P_D(\lambda_1)) + (1-r)P_F(\lambda_1) \\ &= r\left(1-\left(\frac{\xi}{\lambda_1}\right)^{\frac{\beta}{\gamma-\beta}}\right) + (1-r)\left(\frac{\xi}{\lambda_1}\right)^{\frac{\gamma}{\gamma-\beta}}, \end{aligned}$$

where $\lambda_1 \geq \xi$ is the LLRT threshold of the first (and only) sensor. Taking the derivative of both sides of the above formula with respect to λ_1 and using $\frac{dP_D}{dP_F} = \lambda_1$, we obtain

$$\begin{aligned} \frac{dP_e^{(1)}}{d\lambda_1} &= (r\lambda_1 - (1-r))(-P_F'(\lambda_1)) \\ &= r(-P_F'(\lambda_1))\left(\lambda_1 - \frac{1-r}{r}\right). \end{aligned}$$

Notably, the false alarm probability P_F decreases as the LLRT threshold λ_1 increases, i.e., $P_F'(\lambda_1) = \frac{dP_F}{d\lambda_1} \leq 0$. We then have

$$\frac{dP_e^{(1)}}{d\lambda_1} \begin{cases} \leq 0, & \text{for } \dfrac{1-r}{r} \geq \lambda_{\max} \\ \geq 0, & \text{for } \dfrac{1-r}{r} \leq \lambda_{\min} \\ = 0, & \text{for some } \lambda^* = \dfrac{1-r}{r} \end{cases}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ is the maximum and minimum values of the threshold λ_1. Consequently, $P_e^{(1)}$ has an interior minimum only when

$$\lambda_{\min} \leq \frac{1-r}{r} \leq \lambda_{\max}.$$

For $\frac{1-r}{r} \leq \lambda_{\min}$ and $\frac{1-r}{r} \geq \lambda_{\max}$, we respectively obtain $P_e^{(1)}(\lambda_{\min}) = 1-r$ and $P_e^{(1)}(\lambda_{\max}) = r$.

For the specific hypothesis distributions of exponential, we have $\lambda_{\min} = \xi$ and $\lambda_{\max} = +\infty$. Hence, the optimal probability of error for the single sensor system becomes

$$P_e^{(1)} = \begin{cases} P_e^{(1)}(\lambda^*), & \text{if } r \leq \dfrac{1}{1+\xi} \\ 1-r, & \text{if } r > \dfrac{1}{1+\xi}, \end{cases}$$

where $\lambda^* = \frac{1-r}{r}$ satisfies

$$\frac{rP_D(\lambda^*)}{(1-r)P_F(\lambda^*)} = \frac{1}{\xi}.$$

4.3 Parallel Two-sensor System

We now turn to a true decentralized system, i.e., a system with two sensors. Hence, there will be two sensors' decisions, u_1 and u_2, available at the fusion center.

The two-sensor system has been discussed extensively in literature, since for systems with more than two sensors, it seems to be hard in the investigation of the optimal performance. For a two-sensor system, only three fusion rules are available: **OR**, **AND**, and **XOR**. Since we assume that the sensors' observations are conditionally independent given the hypothesis, the **XOR** fusion rule can not be a likelihood ratio test of u_1 and u_2 at the fusion center; thus, it should be excluded in the optimal design. Since **AND** and **OR** fusions are symmetric in the sense that u_1 and u_2 can be complemented before transmission, we will focus only on the **OR** fusion in the next subsection. Specifically, we will show that the identical sensor system is optimal for the exponential hypothesis distributions considered.

For a two-sensor system with LLRT thresholds λ_1 and λ_2 under the **OR** fusion rule, the formula of error probability is given by

$$\begin{aligned} P_{e,\mathrm{OR}}^{(2)}(\lambda_1,\lambda_2) &= r(1-P_D(\lambda_1))(1-P_D(\lambda_2)) + (1-r)(1-(1-P_F(\lambda_1))(1-P_F(\lambda_2))) \\ &= r\left(1-\left(\frac{\xi}{\lambda_1}\right)^{\frac{\beta}{\gamma-\beta}}\right)\left(1-\left(\frac{\xi}{\lambda_2}\right)^{\frac{\beta}{\gamma-\beta}}\right) \\ &\quad +(1-r)\left(1-\left(1-\left(\frac{\xi}{\lambda_1}\right)^{\frac{\gamma}{\gamma-\beta}}\right)\left(1-\left(\frac{\xi}{\lambda_2}\right)^{\frac{\gamma}{\gamma-\beta}}\right)\right). \end{aligned}$$

We then have the next lemma.

Theorem 4.1. For the decentralized detection of exponential sources, the optimal strategy of a two-sensor system given the **OR** fusion rule is to have two identical sensors.

Proof. Taking the derivative of $P_{e,\mathrm{OR}}^{(2)}$ with respect to λ_i, we have

$$\begin{aligned} \frac{\partial P_{e,\mathrm{OR}}^{(2)}}{\partial \lambda_i} &= (-P_F'(\lambda_i))\left(-\frac{\partial P_{e,\mathrm{OR}}^{(2)}}{\partial P_F(i)}\right) \\ &= (-P_F'(\lambda_i))(r\lambda_i(1-P_D(\lambda_{3-i}))-(1-r)(1-P_F(\lambda_{3-i}))) \\ &= (-P_F'(\lambda_i))r(1-P_D(\lambda_{3-i}))\left(\lambda_i-\frac{1-r}{r}\frac{1-P_F(\lambda_{3-i})}{1-P_D(\lambda_{3-i})}\right) \\ &= (-P_F'(\lambda_i))r(1-P_D(\lambda_{3-i}))\left(\xi\frac{P_D(\lambda_i)}{P_F(\lambda_i)}-\frac{1-r}{r}\frac{1-P_F(\lambda_{3-i})}{1-P_D(\lambda_{3-i})}\right), \end{aligned}$$

for $i=1,2$. Thus, if there exists a (λ_1,λ_2) such that $\frac{\partial P_{e,\mathrm{OR}}^{(2)}}{\partial\lambda_1}=0$ and $\frac{\partial P_{e,\mathrm{OR}}^{(2)}}{\partial\lambda_2}=0$, then this (λ_1,λ_2) must also satisfies

$$A(\lambda_1)=A(\lambda_2),$$

which from Lemma 4.1, can be valid only when $\lambda_1=\lambda_2$.

It remains to validate the existence of such (λ_1,λ_2). From the above derivation, their existence relies on the claim that

$$\lambda=\frac{1-r}{r}\frac{1-P_F(\lambda)}{1-P_D(\lambda)} \tag{4.2}$$

has solutions for $\lambda \geq \xi$. Since $\frac{1-P_F(\lambda)}{1-P_D(\lambda)}$ decreases monotonically, and λ increases monotonically, (4.2) has an unique root if $\frac{1-r}{r} \geq \xi^2$, or equivalently, $r \leq \frac{1}{1+\xi^2}$. For the case that $r > \frac{1}{1+\xi^2}$, we derive

$$\lambda_i \geq \xi > \frac{1-r}{r}\frac{1}{\xi} \geq \frac{1-r}{r}\frac{1-P_F(\lambda_i)}{1-P_D(\lambda_i)},$$

i.e., $\frac{\partial P^{(2)}_{e,\mathrm{OR}}}{\partial \lambda_1} > 0$ and $\frac{\partial P^{(2)}_{e,\mathrm{OR}}}{\partial \lambda_2} > 0$. Thus, the optimal strategy is still to adopt the identical LLRT thresholds, i.e., $\lambda_1 = \lambda_2 = \xi$. Hence, for both $r \leq \frac{1}{1+\xi^2}$ and $r > \frac{1}{1+\xi^2}$, the optimal strategy is to adopt identical sensors. □

In summary of the above theorem, the optimal error probability for the **OR** fusion rule is given by

$$P_e^{(2)} = \begin{cases} P^{(2)}_{e,\mathrm{OR}}(\lambda^*, \lambda^*) = (1-r)\left(1-\left(1-(\frac{\xi}{\lambda^*})^{\frac{\gamma}{\gamma-\beta}}\right)^2\right) + r\left(1-(\frac{\xi}{\lambda^*})^{\frac{\beta}{\gamma-\beta}}\right)^2, & \text{if } r \leq \frac{1}{1+\xi^2} \\ 1-r, & \text{if } r > \frac{1}{1+\xi^2}, \end{cases}$$

where λ^* is the solution of the equation

$$\lambda^* = \frac{1-r}{r}\frac{1-P_F(\lambda^*)}{1-P_D(\lambda^*)},$$

or equivalently, is the solution of

$$\frac{r}{1-r}\frac{P_D(\lambda^*)(1-P_D(\lambda^*))}{P_F(\lambda^*)(1-P_F(\lambda^*))} = \frac{1}{\xi}.$$

We end the above discussion by noting that the **OR** fusion is simply the 1-out-of-2 fusion rule, for which it is possible that for systems with more sensors, k-out-of-n fusion rule may be an optimal choice.

4.4 The Parallel Sensor System with an Additional Broadcast Sensor

Now, we temporarily turn our attention to a system with a different configuration from the parallel system in the previous sections, that is, a parallel sensor system with $n-1$ "ordinary" sensors and an additional special sensor. For convenience, we will index the special sensor as the n-th sensor. In operation, the broadcast sensor will broadcast its local decision to all other sensors and the fusion center before each of them makes its own decision. More precisely, the broadcast sensor makes its decision u_n based on its own observation X_n first, and then sends u_n to the fusion center and the remaining $n-1$ ordinary sensors. The i-th (ordinary) sensor afterwards makes its decision u_i based on its own observation X_i and the received u_n, and then conveys u_i to the fusion center individually. Once all of $\{u_i\}_{i=1}^{n}$ are received, the fusion center performs a likelihood ratio test of $(u_1, \ldots, u_{n-1}, u_n)$, and decide whether the hypothesis H_1 or the hypothesis H_0 is true. In subsequent discussions, we restrict ourselves to the special case that the $n-1$ "ordinary" sensors are all identical, and only the broadcast sensor can have a different local decision rule.

It can be shown that the likelihood ratio test of received $(u_1, \ldots, u_{n-1}, u_n)$ at the fusion center still results in a majority voting fusion rule, i.e., a k-out-of-$(n-1+m)$ fusion rule with the broadcast sensor has m ballots, while each of other "ordinary" sensors has only one ballot. For conciseness, we will refer the conventional parallel n-sensor system as system Γ_n and the system described above (with the identical ordinary sensors) as system Ξ_n hereafter.

Before we research on the general Ξ_n system, let us take a look at the simplest kind of it, i.e., system Ξ_2. It turns out that system Ξ_2 is equivalent to the decentralized 2-sensor tandem (serial) system in literature. Since the only non-broadcast sensor in system Ξ_2 has acquired all necessary information in making its own decision, we can just let the first sensor in Ξ_2 be integrated with the fusion center, and take $u_0 = u_1$.

In the following, we shows that for the classification of exponential sources problems, the optimal serial two-sensor strategy is to adopt identical local decision rules for both sensors, and an **OR** fusion rule at the first sensor.

4.4.1 The Serial Two-sensor System

A serial two-sensor system operates equivalently as system Ξ_2. The second sensor makes its decision u_2 according to its observation X_2, and then conveys u_2 to the first sensor. The first sensor then makes the overall decision based on the received u_2 and its own observation X_1. It is known that for the tandem configuration of two-sensor network, the optimal local decision rules are [32] that for the first sensor, two local likelihood ratio thresholds are required (one for $u_2 = 0$ and the other for $u_2 = 1$), while for the second sensor, only one the local likelihood ratio threshold is sufficient.

Denote the LLRT threshold of the second sensor by η. Let the LLRT threshold of the first sensor for $u_2 = 0$ as θ_0, and that for $u_2 = 1$ as θ_1. Then, the probability of error can be written straightforwardly as

$$\begin{aligned} P_e(\eta, \theta_1, \theta_0) &= rP_D(\eta)(1 - P_D(\theta_1)) + (1-r)P_F(\eta)P_F(\theta_1) \\ &+ r(1 - P_D(\eta))(1 - P_D(\theta_0)) + (1-r)(1 - P_F(\eta))P_F(\theta_0). \end{aligned}$$

Taking the derivatives of P_e with respect to η, θ_1, and θ_0, we have

$$\begin{aligned} \frac{\partial P_e}{\partial \eta} &= (-P_F'(\eta))\left(-\frac{\partial P_e}{\partial P_F(\eta)}\right) \\ &= (-P_F'(\eta))(r\eta(P_D(\theta_1) - P_D(\theta_0)) - (1-r)(P_F(\theta_1) - P_F(\theta_0))), \end{aligned}$$

$$\frac{\partial P_e}{\partial \theta_1} = (-P_F'(\theta_1))\left(-\frac{\partial P_e}{\partial P_F(\theta_1)}\right) = (-P_F'(\theta_1))(r\theta_1 P_D(\eta) - (1-r)P_F(\eta)),$$

and

$$\frac{\partial P_e}{\partial \theta_0} = (-P_F'(\theta_0))\left(-\frac{\partial P_e}{\partial P_F(\theta_0)}\right) = (-P_F'(\theta_0))(r\theta_0(1 - P_D(\eta)) - (1-r)(1 - P_F(\eta))).$$

Equating the above derivatives with zero, we obtain the necessary conditions for the optimal error probability as

$$\eta \frac{P_D(\theta_1) - P_D(\theta_0)}{P_F(\theta_1) - P_F(\theta_0)} = \frac{1-r}{r}, \tag{4.3}$$

$$\theta_1 \frac{P_D(\eta)}{P_F(\eta)} = \frac{1-r}{r}, \tag{4.4}$$

and

$$\theta_0 \frac{1-P_D(\eta)}{1-P_F(\eta)} = \frac{1-r}{r}. \tag{4.5}$$

Since for the ROC curve, $\frac{P_D(\eta)}{P_F(\eta)} > \eta > \frac{1-P_D(\eta)}{1-P_F(\eta)}$, we have $\theta_1 < \frac{P_D(\theta_1)-P_D(\theta_0)}{P_F(\theta_1)-P_F(\theta_0)} < \theta_0$.

Let us take a look at two extreme cases, namely, $\theta_0 = \infty$ and $\theta_1 = 0$. It turns out that[1] the cases of $\theta_0 = \infty$ and $\theta_1 = 0$ are equivalent to that the first sensor makes a local decision u_1 according to its own observation X_1 only, and then applies the **AND** and **OR** fusion rules, respectively, to decide the overall output u_0.

Lemma 4.2. For the serial two-sensor system with $\theta_0 = \infty$, the optimal strategy is to let the first and the second sensors make their local decisions u_1 and u_2 according to the LLRTs of their own observations X_1 and X_2, respectively, and then apply the **AND** fusion rule at the output of the first sensor, i.e., $u_0 = u_1 \otimes u_2$.

Proof.

$$\begin{aligned} P_e(\eta, \theta_1, \infty) &= rP_D(\eta)(1-P_D(\theta_1)) + (1-r)P_F(\eta)P_F(\theta_1) \\ &+ r(1-P_D(\eta))(1-P_D(\infty)) + (1-r)(1-P_F(\eta))P_F(\infty) \\ &= rP_D(\eta)(1-P_D(\theta_1)) + (1-r)P_F(\eta)P_F(\theta_1) + r(1-P_D(\eta)) \\ &= r(1-P_D(\eta)P_D(\theta_1)) + (1-r)P_F(\eta)P_F(\theta_1), \end{aligned}$$

where we have used a property of the ROC: $P_D(\infty) = P_F(\infty) = 0$. □

Lemma 4.3. For the serial two-sensor system with $\theta_1 = 0$, the optimal strategy is to let the first and the second sensors make their local decisions u_1 and u_2 according to the LLRTs of their own observations X_1 and X_2, respectively, and then apply the **OR** fusion rule at the output of the first sensor, i.e., $u_0 = u_1 \oplus u_2$.

[1] Here, with a slight abuse of notations, we let the intermediate product, i.e., the result of the LLRT at the first sensor, be denoted by u_1, and let the ultimate output of the first sensor be denoted by u_0).

Proof.

$$
\begin{aligned}
P_e(\eta, 0, \theta_0) &= rP_D(\eta)(1 - P_D(0)) + (1-r)P_F(\eta)P_F(0) \\
&+ r(1 - P_D(\eta))(1 - P_D(\theta_0)) + (1-r)(1 - P_F(\eta))P_F(\theta_0) \\
&= (1-r)P_F(\eta) + r(1 - P_D(\eta))(1 - P_D(\theta_0)) + (1-r)(1 - P_F(\eta))P_F(\theta_0) \\
&= r(1 - P_D(\eta))(1 - P_D(\theta_0)) + (1-r)(1 - (1 - P_F(\eta))(1 - P_F(\theta_0))),
\end{aligned}
$$

where we have used a property of the ROC: $P_D(0) = P_F(0) = 1$. □

In both of the above cases, the serial two-sensor systems function exactly like the parallel two-sensor system with corresponding fusion rules. In general, the optimal serial two-sensor system uses two finite and nonzero LLRT thresholds at the first sensor, and therefore does not necessarily reduce to some equivalent parallel two-sensor system. In the next theorem, we show that for the considered classification problem of exponential sources, the optimal serial two-sensor system is one of the above extreme cases. More precisely, the optimal serial two-sensor system is equivalent to the optimal parallel two-sensor system.

Theorem 4.2. For the classification problem of exponential sources, the optimal strategy for the serial two-sensor system is to let the first and the second sensors make their local decisions u_1 and u_2 according to the LLRTs of their own observations X_1 and X_2 with the two equal thresholds θ_0 and η, respectively, and then apply either **AND** or **OR** fusion rules at the output of the first sensor.

Proof. Firstly, define a function $B(\lambda) = \frac{\lambda}{\frac{P_D(\lambda)}{P_F(\lambda)}}$. Then, (4.3) becomes

$$
B(\eta)\frac{P_D(\eta)}{P_F(\eta)}\frac{P_D(\theta_1) - P_D(\theta_0)}{P_F(\theta_1) - P_F(\theta_0)} = \frac{1-r}{r}.
$$

Combining the above equation with the (4.4), we establish

$$
\theta_1 = B(\eta)\frac{P_D(\theta_1) - P_D(\theta_0)}{P_F(\theta_1) - P_F(\theta_0)}.
$$

By using the identity $\theta_1 = B(\theta_1)\frac{P_D(\theta_1)}{P_F(\theta_1)}$, the above equation can be written as

$$\frac{\frac{P_D(\theta_1)}{P_D(\theta_0)}}{\frac{P_F(\theta_1)}{P_F(\theta_0)}} = \frac{B(\eta)}{B(\theta_1)} \frac{\frac{P_D(\theta_1)}{P_D(\theta_0)} - 1}{\frac{P_F(\theta_1)}{P_F(\theta_0)} - 1}. \tag{4.6}$$

In (4.6), if $\frac{B(\eta)}{B(\theta_1)} \geq 1$, then we immediately have

$$\frac{\frac{P_D(\theta_1)}{P_D(\theta_0)}}{\frac{P_F(\theta_1)}{P_F(\theta_0)}} \geq \frac{\frac{P_D(\theta_1)}{P_D(\theta_0)} - 1}{\frac{P_F(\theta_1)}{P_F(\theta_0)} - 1},$$

which leads to

$$\frac{P_D(\theta_1)}{P_F(\theta_1)} \geq \frac{P_D(\theta_0)}{P_F(\theta_0)}.$$

Since $\frac{P_D(\lambda)}{P_F(\lambda)}$ increases monotonically with respect to λ, we have

$$\theta_1 \geq \theta_0,$$

which contradicts the aforementioned proposition: $\theta_0 \geq \theta_1$. Hence, $\frac{B(\eta)}{B(\theta_1)} < 1$. Yet, for the classification of exponential sources problem, we have $B(\eta) = B(\theta_1) = \xi$; thus, the optimal $(\eta, \theta_1, \theta_0)$ must lie on the boundary, i.e., the two extreme cases. □

Remark 4.1. For the classification problem of the additive Gaussian sources, $B(\lambda)$ is a monotonically increasing function of λ; thus, $\eta < \theta_1 < \theta_0$.

Corollary 4.1. For the classification problem of exponential sources, the optimal performance of the serial two-sensor system is equal to the optimal performance of the parallel two-sensor system.

4.5 The Ξ_n System

We now turn to the Ξ_n system, that is, the system with $n-1$ ordinary sensors, one broadcast sensor, and a fusion center with $n \geq 3$. It is easy to see that for the optimal system, each of the ordinary sensors still uses the joint LLRT of its own low observation and the received

u_n to determine its output u_i; however, it is not clear whether the optimal local decision rule of the broadcast sensor is still a LLRT on its own observation. Nonetheless, we will only discuss the case that the broadcast sensor adopts the LLRT as its local decision rule in this dissertation. Note that the extension of the result in this section to the optimal parallel system in the following sections is not affected by this restriction.

Denote the LLRT threshold of the n-th sensor as η. Denote the common LLRT thresholds of the $(n-1)$ ordinary sensors as θ_0 for $u_n = 0$, and θ_1 for $u_n = 1$. Put the fusion rule as $u_0 = \Upsilon(u_1, \ldots, u_n)$, or simply Υ. One can then decompose the fusion rule Υ as

$$\Upsilon(u_1, \ldots, u_n) = u_n \Upsilon_1(u_1, \ldots, u_{n-1}) + (1 - u_n)\Upsilon_0(u_1, \ldots, u_{n-1}),$$

where $\Upsilon_1(u_1, \ldots, u_{n-1}) = \Upsilon(u_1, \ldots, u_{n-1}, 1)$ and $\Upsilon_0(u_1, \ldots, u_{n-1}) = \Upsilon(u_1, \ldots, u_{n-1}, 0)$ correspond to the "conditional" fusion rules on $(u_1, \ldots, u_{n-1})$ conditioning on $u_n = 1$ and $u_n = 0$, respectively.

The probability of error then can be expressed as

$$\begin{aligned} \tilde{P}_e^{(n)} &= r(1 - P_D(\eta))(1 - R_D^{<0>}(\theta_0)) + (1 - r)(1 - P_F(\eta))R_F^{<0>}(\theta_0) \\ &+ rP_D(\eta)(1 - R_D^{<1>}(\theta_1)) + (1 - r)P_F(\eta)R_F^{<1>}(\theta_1), \end{aligned}$$

where $P_D^{<0>}(\theta_0)$ and $P_F^{<0>}(\theta_0)$ are the detection probability and the false alarm probability of the parallel $(n-1)$-sensor system with the common LLRT threshold θ_0 and fusion rule $\Upsilon_0(u_1, \ldots, u_{n-1})$, respectively. $P_D^{<1>}(\theta_1)$ and $P_F^{<1>}(\theta_1)$ are defined similarly.

Despite of some normalization constants, one can easily verify that in the formula of $\tilde{P}_e^{(n)}$, the first two terms can be regarded as the probabilities of error of the parallel $(n-1)$-sensor system with the common LLRT threshold θ_0, the fusion rule $\Upsilon_0(u_1, \ldots, u_{n-1})$ and the prior probability $\Pr\{H_1\} = r(1 - P_D(\eta))$. Likewise, the last two terms can be treated as the probability of error of the parallel $(n-1)$-sensor system with the common LLRT threshold

θ_1, the fusion rule $\Upsilon_1(u_1, \ldots, u_{n-1})$, and the prior probability $\Pr\{H_1\} = rP_D(\eta)$. These two probabilities of errors will be referred as $R_e^{<0>}(\theta_0)$ and $R_e^{<1>}(\theta_1)$, respectively.

Now, since we assume that the ordinary sensors are all identical, the fusion rules $\Upsilon_1(u_1, \ldots, u_{n-1})$ and $\Upsilon_0(u_1, \ldots, u_{n-1})$ must have the forms of the k-out-of-$(n-1+m)$ rules. For $1 < k < n$, the probability of error can then be expressed as

$$\begin{aligned}
\tilde{P}_{e,k}^{(n)}(\eta, \theta_0, \theta_1) &= r(1 - P_D(\eta))(1 - Q_{D,k}^{(n-1)}(\theta_0)) + (1-r)(1 - P_F(\eta))Q_{F,k}^{(n-1)}(\theta_0) \\
&+ rP_D(\eta)(1 - Q_{D,k-m}^{(n-1)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,k-m}^{(n-1)}(\theta_1),
\end{aligned}$$

where

$$Q_{D,k}^{(n-1)}(\theta_0) = \sum_{l=k}^{n-1} C_l^{n-1} P_D(\theta_0)^l (1 - P_D(\theta_0))^{n-1-l},$$

$$Q_{F,k}^{(n-1)}(\theta_0) = \sum_{l=k}^{n-1} C_l^{n-1} P_F(\theta_0)^l (1 - P_F(\theta_0))^{n-1-l},$$

$$Q_{D,k-m}^{(n-1)}(\theta_1) = \sum_{l=k-m}^{n-1} C_l^{n-1} P_D(\theta_1)^l (1 - P_D(\theta_1))^{n-1-l},$$

and

$$Q_{F,k-m}^{(n-1)}(\theta_1) = \sum_{l=k-m}^{n-1} C_l^{n-1} P_F(\theta_1)^l (1 - P_F(\theta_1))^{n-1-l}.$$

For $k = 1$ and $k \geq n$, the probabilities of errors are given by

$$\tilde{P}_{e,1}^{(n)}(\eta, \theta_0) = r(1-P_D(\eta))(1-P_D(\theta_0))^{n-l} + (1-r)(1-P_F(\eta))(1-(1-P_F(\theta_0))^{n-l}) + (1-r)P_F(\eta),$$

and

$$\tilde{P}_{e,k}^{(n)}(\eta, \theta_0, \theta_1) = r(1 - P_D(\eta)) + rP_D(\eta)(1 - Q_{D,k-m}^{(n-1)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,k-m}^{(n-1)}(\theta_1).$$

Let us take a look at a special case of $\tilde{P}_{e,k}^{(n)}$, i.e., $m = 1$. For this case, the overall fusion rule becomes

$$\Upsilon(u_1, \ldots, u_n) = \begin{cases} 1, & \text{if } u_1 + \cdots + u_n \geq k \\ 0, & \text{if } u_1 + \cdots + u_n < k, \end{cases}$$

i.e., the k-out-of-n fusion rule. The necessary conditions for the achievement of the optimal error are given by

$$\begin{aligned} &\log\left(\frac{r}{1-r}\right) + \log\left(\frac{1-P_D(\eta)}{1-P_F(\eta)}\right) + \log(\theta_0) \\ &\quad +(k-1)\log\left(\frac{P_D(\theta_0)}{P_F(\theta_0)}\right) + (n-1-k)\log\left(\frac{1-P_D(\theta_0)}{1-P_F(\theta_0)}\right) = 0, \end{aligned} \tag{4.7}$$

$$\begin{aligned} &\log\left(\frac{r}{1-r}\right) + \log\left(\frac{P_D(\eta)}{P_F(\eta)}) + \log(\theta_1\right) \\ &\quad +(k-2)\log(\frac{P_D(\theta_1)}{P_F(\theta_1)}) + (n-k)\log(\frac{1-P_D(\theta_1)}{1-P_F(\theta_1)}) = 0, \end{aligned} \tag{4.8}$$

$$\log\left(\frac{r}{1-r}\right) + \log(\eta) + \log\left(\frac{Q_{D,k}^{(n-1)}(\theta_0) - Q_{D,k-1}^{(n-1)}(\theta_1)}{Q_{F,k}^{(n-1)}(\theta_0) - Q_{F,k-1}^{(n-1)}(\theta_1)}\right) = 0 \text{ for } 1 < k < n, \tag{4.9}$$

$$\eta\left(\frac{1-P_D(\theta_0)}{1-P_F(\theta_0)}\right)^{n-1} = \frac{1-r}{r}, \tag{4.10}$$

$$\theta_0\frac{1-P_D(\eta)}{1-P_F(\eta)}\left(\frac{1-P_D(\theta_0)}{1-P_F(\theta_0)}\right)^{n-2} = \frac{1-r}{r} \text{ for } k=1, \tag{4.11}$$

$$\eta(\frac{P_D(\theta_1)}{P_F(\theta_1)})^{n-1} = \frac{1-r}{r}, \tag{4.12}$$

and

$$\theta_1\frac{P_D(\eta)}{P_F(\eta)}\left(\frac{P_D(\theta_1)}{P_F(\theta_1)}\right)^{n-2} = \frac{1-r}{r} \text{ for } k=n. \tag{4.13}$$

Remark 4.2. The above equations are coupled and nonlinear. Therefore, it is difficult to trace all possible solutions. It is however easy to show a property of the solutions, i.e., either $\theta_i \geq \eta$ for $i = 0, 1$ or $\theta_i \leq \eta$ for $m = 1$, $i = 0, 1$ and $1 < k < n$. This can be proved as follows.

Proof. Assume $\theta_0 > \eta > \theta_1$. Since both $\frac{P_D(\lambda)}{P_F(\lambda)}$ and $\frac{1-P_D(\lambda)}{1-P_F(\lambda)}$ increase monotonically with respect to λ, we have from (4.7) and (4.8) that

$$\begin{aligned} &\log\left(\frac{r}{1-r}\right) + \log\left(\frac{1-P_D(\theta_1)}{1-P_F(\theta_1)}\right) + \log(\theta_0) \\ &\quad +(k-1)\log\left(\frac{P_D(\theta_0)}{P_F(\theta_0)}\right) + (n-1-k)\log\left(\frac{1-P_D(\theta_0)}{1-P_F(\theta_0)}\right) < 0, \end{aligned} \tag{4.14}$$

and

$$\log\left(\frac{r}{1-r}\right)+\log\left(\frac{P_D(\theta_0)}{P_F(\theta_0)}\right)+\log(\theta_1)$$
$$+(k-2)\log(\frac{P_D(\theta_1)}{P_F(\theta_1)})+(n-k)\log(\frac{1-P_D(\theta_1)}{1-P_F(\theta_1)})>0. \tag{4.15}$$

Combining the above two inequalities yields

$$\begin{aligned}
& \log(\theta_1)+(k-2)\log\left(\frac{P_D(\theta_1)}{P_F(\theta_1)}\right)+(n-1-k)\log\left(\frac{1-P_D(\theta_1)}{1-P_F(\theta_1)}\right) \\
\geq\ & \log(\theta_0)+(k-2)\log\left(\frac{P_D(\theta_0)}{P_F(\theta_0)}\right)+(n-1-k)\log\left(\frac{1-P_D(\theta_0)}{1-P_F(\theta_0)}\right).
\end{aligned}$$

Since both sides of the above inequality are monotonically increasing functions of θ_1 and θ_0, we have $\theta_1>\theta_0$, which results in a contradict. Similarly, one can show that the alternative assumption $\theta_1>\eta>\theta_0$ also results in a contradiction. Hence, η has to be either no less or no greater than both θ_0 and θ_1. $\square$

In light of the above necessary conditions, one can obtain the following two lemmas.

Lemma 4.4. If $\frac{\lambda}{\frac{1-P_D(\lambda)}{1-P_F(\lambda)}}$ is monotonic with respect to λ, then the necessary conditions for $k=1$ and $m=1$, i.e., (4.10) and (4.11), have nontrivial solutions only when $\theta_0=\eta$.

Lemma 4.5. If $\frac{\lambda}{\frac{P_D(\lambda)}{P_F(\lambda)}}$ is monotonic with respect to λ, then the necessary conditions for $k=n$ and $m=1$, i.e., (4.12) and (4.13), have nontrivial solutions only when $\theta_1=\eta$.

The proofs of the above two lemmas are straightforward, and hence, we omit it.

4.6 Optimal Parallel Systems

In this section, we discuss the relationship between optimal parallel Γ_n system and the optimal Ξ_n system through the following propositions $\mathcal{S}(n)$, $\mathcal{T}(n)$, and $\mathcal{V}(n)$.

Proposition 4.1. $\mathcal{S}(n)$: For the parallel n-sensor system Γ_n and for arbitrary prior $P(H_1) = r$, the optimal error probability is achieved by identical sensors and k-out-of-n fusion rules.

Proposition 4.2. $\mathcal{T}(n)$: For the n-sensor Ξ_n system, if the broadcast sensor has $m \geq 1$ ballots in the voting fusion, then the optimal fusion rules are k-out-of-n fusion rules.

Proposition 4.3. $\mathcal{V}(n)$: For the n-sensor Ξ_n system with a fixed k-out-of-n fusion rule and for arbitrary prior $P(H_1) = r$, the optimal error probability is achieved by identical sensors, i.e., the $(n-1)$ ordinary sensors ignore the received decision of the broadcast sensor and use local decision rules that are the same as the broadcast sensor one, and that are based on their own observations only.

Lemma 4.6. If the proposition $\mathcal{S}(2)$ holds, then the proposition $\mathcal{T}(3)$ holds.

Proof. As discussed in the preceding section, the optimal fusion rule for the Ξ_n system is k-out-of-$(n-1+m)$ fusion rules for some $m \geq 0$; hence, it suffices to show that for $m \geq 1$, the optimal $m = 1$.

Assume that $\mathcal{S}(2)$ holds. Let us consider the case of $m = 1$ and $1 \leq k \leq 3$. These are obviously k-out-of-3 fusion rules. Now we consider the case of $m = 2$ and $1 \leq k \leq 4$, for which the fusion rule can be expressed as

$$u_0 = \begin{cases} 1, & \text{if } u_1 + u_2 + 2u_3 \geq k \\ 0, & \text{if } u_1 + u_2 + 2u_3 < k. \end{cases}$$

Observe from the above discussion that for the case of $m = 2$ and $k = 1$ and the case of $m = 2$ and $k = 4$, the fusion rules are equivalent to the 1-out-of-3 fusion rule and the

3-out-of-3 fusion rule, respectively. For the case of $m = 2$ and $k = 2$, we have

$$\begin{aligned}
\min_{\eta,\theta_0,\theta_1}(\tilde{P}_e^{(3)}) &= \min_{\eta,\theta_0}(r(1-P_D(\eta))(1-Q_{D,2}^{(2)}(\theta_0)) + (1-r)(1-P_F(\eta))Q_{F,2}^{(2)}(\theta_0) \\
&+ (1-r)P_F(\eta)) \\
&\geq \min_{\eta,\theta_0}(r(1-P_D(\eta))(1-Q_{D,2}^{(2)}(\theta_0)) + (1-r)(1-P_F(\eta))Q_{F,2}^{(2)}(\theta_0) \\
&+ \min_{\eta,\theta_1}(rP_D(\eta)(1-Q_{D,1}^{(2)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,1}^{(2)}(\theta_1))) \\
&= \min_{\eta,\theta_0,\theta_1}(r(1-P_D(\eta))(1-Q_{D,2}^{(2)}(\theta_0)) + (1-r)(1-P_F(\eta))Q_{F,2}^{(2)}(\theta_0) \\
&+ rP_D(\eta)(1-Q_{D,1}^{(2)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,1}^{(2)}(\theta_1)) \\
&= \min_{\eta,\theta_0,\theta_1}(\tilde{P}_{e,2}^{(3)}(\eta,\theta_0,\theta_1)).
\end{aligned}$$

Thus, the minimum error probability corresponding to the 2-out-of-3 fusion rule is no greater than that for the case of $m = 2$ and $k = 2$.

For the case $m = 2$ and $k = 3$, we have

$$\begin{aligned}
\min_{\eta,\theta_0,\theta_1}(\tilde{P}_e^{(3)}) &= \min_{\eta,\theta_1}(r(1-P_D(\eta)) \\
&+ rP_D(\eta)(1-Q_{D,1}^{(2)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,1}^{(2)}(\theta_1)) \\
&\geq \min_{\eta,\theta_1}(\min_{\eta,\theta_0}(r(1-P_D(\eta))(1-Q_{D,2}^{(2)}(\theta_0)) + (1-r)(1-P_F(\eta))Q_{F,2}^{(2)}(\theta_0)) \\
&+ rP_D(\eta)(1-Q_{D,1}^{(2)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,1}^{(2)}(\theta_1)) \\
&= \min_{\eta,\theta_0,\theta_1}(r(1-P_D(\eta))(1-Q_{D,2}^{(2)}(\theta_0)) + (1-r)(1-P_F(\eta))Q_{F,2}^{(2)}(\theta_0) \\
&+ rP_D(\eta)(1-Q_{D,1}^{(2)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,1}^{(2)}(\theta_1)) \\
&= \min_{\eta,\theta_0,\theta_1}(\tilde{P}_{e,2}^{(3)}(\eta,\theta_0,\theta_1)),
\end{aligned}$$

Thus, the minimum error probability of the 2-out-of-3 fusion rule is no greater than that for the case of $m = 2$ and $k = 3$.

Note that the fusion rule for the case of $m = 3$ and $k \leq 2$ is equivalent to the fusion rule for the case of $m = 2$ and $k \leq 2$. For the case of $m = 3$ and $k = 3$, the fusion rule is $u_0 = u_3$, and its minimum error probability is equal to $P_{e,min}^{(1)}$, i.e., the minimum error probability of the single sensor; hence, it can not be the optimal fusion rule. The fusion rule for the case of $m = 3$ and $k = 4, 5$ is equivalent to the fusion rule for the case of $m = 2$ and $k = 3, 4$. Finally, it is easy to see that the analysis for the case of $m \geq 3$ is identical to the analysis for the case of $m = 3$.

The lemma is then substantiated since we have shown that for all $m > 0$ fusion rules, there are some k-out-of-n fusion rules that have error probabilities no greater than the original ones. $\square$

Lemma 4.7. For $n \geq 3$, if

1. conditions (4.7), (4.8) and (4.9) are satisfied with $\theta_0 = \theta_1 = \eta$ for $1 < k < n$;
2. either $\frac{\lambda}{\frac{P_D(\lambda)}{P_F(\lambda)}}$ is monotonic, or $\tilde{P}_{e,n}^{(n)} > \min_{\lambda_1,\ldots,\lambda_n} (P_e^{(n)})$;
3. either $\frac{\lambda}{\frac{1-P_D(\lambda)}{1-P_F(\lambda)}}$ is monotonic, or $\tilde{P}_{e,1}^{(n)} > \min_{\lambda_1,\ldots,\lambda_n} (P_e^{(n)})$,

then proposition $\mathcal{V}(n)$ holds.

Proof. The proof is straightforward; hence, we omit it. $\square$

Lemma 4.8. If propositions $\mathcal{S}(2)$ and $\mathcal{V}(3)$ hold, then proposition $\mathcal{S}(3)$ holds.

Proof. For a fixed fusion rule, we have from the formula of the error probability of the parallel system

$$
\begin{aligned}
\min_{\lambda_1,\lambda_2,\lambda_3}(P_e^{(3)}) &= \min_{\lambda_1,\lambda_2,\lambda_3}(r(1-P_D(\lambda_3))(1-R_D^{<0>}(\lambda_1,\lambda_2)) + (1-r)(1-P_F(\lambda_3))R_F^{<0>}(\lambda_1,\lambda_2) \\
&+ rP_D(\lambda_3)(1-R_D^{<1>}(\lambda_1,\lambda_2)) + (1-r)P_F(\lambda_3)R_F^{<1>}(\lambda_1,\lambda_2)) \\
&\geq \min_{\lambda_3}(\min_{\lambda_1,\lambda_2}(r(1-P_D(\lambda_3))(1-R_D^{<0>}(\lambda_1,\lambda_2)) + (1-r)(1-P_F(\lambda_3))R_F^{<0>}(\lambda_1,\lambda_2)) \\
&+ \min_{\lambda_1,\lambda_2}(rP_D(\lambda_3)(1-R_D^{<1>}(\lambda_1,\lambda_2)) + (1-r)P_F(\lambda_3)R_F^{<1>}(\lambda_1,\lambda_2))) \\
&= \min_{\lambda_3}(\min_{\theta_{01},\theta_{02}}(r(1-P_D(\lambda_3))(1-R_D^{<0>}(\theta_{01},\theta_{02})) + (1-r)(1-P_F(\lambda_3))R_F^{<0>}(\theta_{01},\theta_{02})) \\
&+ \min_{\theta_{11},\theta_{12}}(rP_D(\lambda_3)(1-R_D^{<1>}(\theta_{11},\theta_{12})) + (1-r)P_F(\lambda_3)R_F^{<1>}(\theta_{11},\theta_{12}))) \\
&\geq \min_{\eta,\theta_0,\theta_1}(r(1-P_D(\eta))(1-Q_{D,k}^{(3-1)}(\theta_0)) + (1-r)(1-P_F(\eta))Q_{F,k}^{(3-1)}(\theta_0) \\
&+ rP_D(\eta)(1-Q_{D,k-m}^{(3-1)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,k-m}^{(3-1)}(\theta_1)) \\
&\geq \min_{\eta,\theta_0,\theta_1}(r(1-P_D(\eta))(1-Q_{D,k}^{(3-1)}(\theta_0)) + (1-r)(1-P_F(\eta))Q_{F,k}^{(3-1)}(\theta_0) \\
&+ rP_D(\eta)(1-Q_{D,k-1}^{(3-1)}(\theta_1)) + (1-r)P_F(\eta)Q_{F,k-1}^{(3-1)}(\theta_1)) \\
&= \min_{\eta,\theta_0,\theta_1}(\tilde{P}_{e,k}^{(3)}) \\
&= \min_{\lambda_1=\lambda_2=\lambda_3}(P_e^{(3)}) \\
&\geq \min_{\lambda_1,\lambda_2,\lambda_3}(P_e^{(3)})
\end{aligned}
$$

where $P_D^{<0>}(\lambda_1,\lambda_2)$ and $P_F^{<0>}(\lambda_1,\lambda_2)$ are the detection probability and the false alarm probability of the parallel 2-sensor system with the LLRT thresholds λ_1 and λ_2 and the fusion rule $\Upsilon_0(u_1,u_2)$, respectively. $P_D^{<1>}(\lambda_1,\lambda_2)$ and $P_F^{<1>}(\lambda_1,\lambda_2)$ are defined similarly. Note that in the above derivation, the second inequality comes from proposition $\mathcal{S}(2)$, the third inequality follows from proposition $\mathcal{T}(3)$, and the last equality comes from proposition $\mathcal{V}(3)$.

Hence, we have $\min_{\lambda_1,\lambda_2,\lambda_3}(P_e^{(3)}) = \min_{\lambda_1=\lambda_2=\lambda_3}(P_e^{(3)})$, and the lemma is proved.

$\square$

Theorem 4.3. If proposition $\mathcal{S}(2)$ holds, and if for $n=3$ the conditions in Lemma (4.7) hold, then proposition $\mathcal{S}(3)$ holds.

Proof. The theorem can be easily obtained from the above lemmas. □

Theorem 4.4. If propositions $\mathcal{S}(2)$, $T(l)$ and $V(l)$ hold for $n \geq l \geq 3$, then proposition $S(l)$ holds for $n \geq l \geq 1$.

Proof. The theorem can be easily obtained from the above lemmas. □

4.7 The Parallel Three-sensor System

Now we turn to the classification of exponential sources problem. Although we only discuss the optimal performance of the parallel three-sensor system, similar approach can be applied for the analysis of the system with more than three sensors.

For the classification problem of exponential sources, we can rewrite (4.8) as

$$\begin{aligned}\log\left(\frac{r}{1-r}\right) + \log\left(\frac{P_D(\theta_1)}{P_F(\theta_1)}\right) + \log(\eta)\\ +(k-2)\log\left(\frac{P_D(\theta_1)}{P_F(\theta_1)}\right) + (n-k)\log\left(\frac{1-P_D(\theta_1)}{1-P_F(\theta_1)}\right) = 0.\end{aligned} \tag{4.16}$$

Combining (4.16) and (4.9), we obtain

$$\frac{Q_{D,k}^{(n-1)}(\theta_0) - Q_{D,k-1}^{(n-1)}(\theta_1)}{P_D(\theta_1)^{k-1}(1-P_D(\theta_1))^{n-k}} = \frac{Q_{F,k}^{(n-1)}(\theta_0) - Q_{F,k-1}^{(n-1)}(\theta_1)}{P_F(\theta_1)^{k-1}(1-P_F(\theta_1))^{n-k}}, \tag{4.17}$$

which is equivalent to

$$\frac{Q_{D,k}^{(n-1)}(\theta_0) - Q_{D,k}^{(n-1)}(\theta_1)}{P_D(\theta_1)^{k}(1-P_D(\theta_1))^{n-k}} = \frac{Q_{F,k}^{(n-1)}(\theta_0) - Q_{F,k}^{(n-1)}(\theta_1)}{P_F(\theta_1)^{k}(1-P_F(\theta_1))^{n-k}}. \tag{4.18}$$

Denoting $a = P_F(\theta_0)$ and $b = P_F(\theta_1)$, and putting $P_D(\theta_i) = P_F(\theta_i)^\xi$ for $i = 0, 1$, we can rewrite the above equation as

$$J_k^{(n)}(\xi) = J_k^{(n)}(1), \tag{4.19}$$

where

$$J_k^{(n)}(\xi) = \frac{\sum_{l=k}^{n-1} C_l^{n-1} a^{l\xi}(1-a^\xi)^{n-1-l} - \sum_{l=k}^{n-1} C_l^{n-1} b^{l\xi}(1-b^\xi)^{n-1-l}}{b^{(k-1)\xi}(1-b^\xi)^{n-k}}.$$

Some examples of $J_k^{(n)}(\xi)$s are listed here for reference.

$$J_2^{(3)}(\xi) = \frac{a^{2\xi} - b^{2\xi}}{b^{\xi}(1-b^{\xi})},$$

$$J_2^{(4)}(\xi) = \frac{2(a^{2\xi} - b^{2\xi}) - 3(a^{3\xi} - b^{3\xi})}{b^{\xi}(1-b^{\xi})^2},$$

and

$$J_3^{(4)}(\xi) = \frac{a^{3\xi} - b^{3\xi}}{b^{2\xi}(1-b^{\xi})}.$$

Taking the derivative of $J_k^{(n)}(\xi)$ with respect to ξ, we can show numerically that for $a \neq b$, $J_k^{(n)}(\xi)$ is either monotonically increasing or monotonically decreasing for $0 < \xi < 1$ and $1 < k < n$; hence, for $1 < k < n$, (4.19) is satisfied if, and only if, $a = b$, i.e., $\theta_0 = \theta_1$. In other words, conditions (4.7), (4.8) and (4.9) are satisfied, only if $\theta_0 = \theta_1$. Moreover, from Lemma 4.1, (4.7) and (4.8), we learn that $\theta_0 = \theta_1$ results in $\theta_0 = \theta_1 = \eta$, i.e., conditions (4.7), (4.8) and (4.9) are satisfied, only when $\theta_0 = \theta_1 = \eta$ for $1 < k < n$.

In the above discussion, we know numerically that the first condition in the Lemma 4.7 holds. In addition, the third condition holds since

$$\frac{\lambda}{\frac{1-P_D(\lambda)}{1-P_F(\lambda)}} = \xi \frac{\frac{P_D(\lambda)}{P_F(\lambda)}}{\frac{1-P_D(\lambda)}{1-P_F(\lambda)}} = \exp(A(\lambda))$$

is monotonic. As for the second condition in Lemma 4.7, since $\frac{\lambda}{\frac{P_D(\lambda)}{P_F(\lambda)}} = \xi$ is not monotonic, we need to verify

$$\min_{\eta,\theta_0,\theta_1} (\tilde{P}_{e,n}^{(n)}) > \min_{\lambda_1,\ldots,\lambda_n} (P_e^{(n)}).$$

From the formula of $\tilde{P}_{e,n}^{(n)}$, we have

$$\begin{aligned}
\tilde{P}_{e,n}^{(n)}(\eta,\theta_0,\theta_1) &= r(1-P_D(\eta)) + rP_D(\eta)(1-P_D(\theta_1)^{n-1}) + (1-r)P_F(\eta)P_F(\theta_1)^{n-1} \\
&= r(1-P_D(\eta)P_D(\theta_1)^{n-1}) + (1-r)P_F(\eta)P_F(\theta_1)^{n-1} \\
&= r\left(1 - P_D\left(\eta\left(\frac{\theta_1}{\xi}\right)^{n-1}\right)\right) + (1-r)P_F\left(\eta\left(\frac{\theta_1}{\xi}\right)^{n-1}\right).
\end{aligned}$$

Hence,

$$\min_{\eta,\theta_0,\theta_1}(\tilde{P}_{e,n}^{(n)}) = \min_{\lambda_1}(P_e^{(1)}) > \min_{\lambda_1,\ldots,\lambda_n}(P_e^{(n)}),$$

Now, from Lemma 4.7, we notice that proposition $\mathcal{V}(n)$ holds. Moreover, if proposition $T(l)$ holds for $n \geq l \geq 3$, then proposition $S(l)$ also holds for $n \geq l \geq 1$, i.e., the optimal performance is achieved by the parallel systems with identical sensors.

The above arguments are, however, built partly based on numerical results. Nonetheless, for the relatively small numbers of sensors, we can show the same results analytically. In the following, we show the optimality of identical sensors on the classification of exponential sources for the parallel three-sensor system. Firstly, we will show the monotonicity of $J_2^{(3)}(\xi)$. One can show the monotonicity of other $J_k^{(n)}(\xi)$ for relatively small n in the same way.

Lemma 4.9. *$J_2^{(3)}(\xi)$ is a monotonic function for $a \neq b$.*

Proof. Denote the ratio between a and b as $\rho = \frac{a}{b}$. Then $J_2^{(3)}(\xi)$ can be expressed as

$$J_2^{(3)}(\xi) = \frac{a^{2\xi} - b^{\xi}}{b^{\xi}(1-b^{\xi})} + 1 = 1 - \frac{a^{-\xi} - \rho^{\xi}}{a^{-\xi} - \rho^{-\xi}}.$$

Taking the derivative of the above formula with respect to ξ, we have

$$J_2^{'(3)}(\xi) = \frac{\Omega(\xi)}{(a^{-\xi} - \rho^{-\xi})^2},$$

where

$$\Omega(\xi) = (-a^{-\xi}\log(a) + \rho^{-\xi}\log(\rho))(a^{-\xi} - \rho^{\xi}) - (a^{-\xi} - \rho^{-\xi})(-a^{-\xi}\log(a) - \rho^{\xi}\log(\rho)).$$

Now taking the derivative of $\Omega(\xi)$ with respect to ξ, we have

$$\Omega^{'}(\xi) = ((\log(\rho))^2 - (\log(a))^2)a^{-\xi}(\rho^{\xi} - \rho^{-\xi}).$$

Thus, either $\Omega^{'}(\xi) > 0$ for $\xi > 0$ or $\Omega^{'}(\xi) < 0$ for $\xi > 0$. Moreover, $\Omega(0) = 0$; hence, either $\Omega(\xi) > 0$ for $\xi > 0$ or $\Omega(\xi) < 0$ for $\xi > 0$. Thus, we have confirmed that $J_2^{'(3)}(\xi)$ is either positive for all $\xi > 0$ or negative for all $\xi > 0$, i.e., $J_2^{(3)}(\xi)$ is a monotonic function for $1 > \xi > 0$. $\square$

Theorem 4.5. For the classification of exponential sources in the parallel three-sensor system, the optimal local decision rules are identical for all sensors.

Proof. From Theorem 4.3, proposition $\mathcal{S}(2)$ holds as shown in Section 4.3. Also, from the above discussion, the conditions in Lemma 4.7 are satisfied. Thus, the theorem is valid. $\square$

4.8 Problems with Similar ROCs

As long as the hypothesis testing problem has similar ROCs as those discussed for exponential hypothesis sources, their performances should be able to be evaluated in similar manner. Here, we illustrate two such examples. Classification problems of this sort might be encountered in survival analysis and failure time analysis.

The first example considers the following hypothesis testing problem:

$$H_1 : X_i = \min(Z_{i,1}, \cdots, Z_{i,\beta})$$

versus

$$H_0 : X_i = \min(Z_{i,1}, \cdots, Z_{i,\gamma})$$

for $i = 1, 2, \ldots, n$ and $\beta < \gamma$, where X_i is the observation of i-th sensor, and $\{Z_{i,j}\}$ are independent and identically distributed random variables with the associated PDF $w_Z(z)$ and CDF $W_Z(z)$. Thus, X_i has CDF $F_X(x_i) = 1 - \Pr(\min(Z_{i,1}, \cdots, Z_{i,\beta}) > x_i) = 1 - (1 - W_Z(x_i))^\beta$ and PDF $f_X(x_i) = \beta(1 - W_Z(x_i))^{\beta-1} w_Z(x_i)$ when H_1 is true, and has CDF $G_X(x_i) = 1 - \Pr(\min(Z_{i,1}, \cdots, Z_{i,\gamma}) > x_i) = 1 - (1 - W_Z(x_i))^\gamma$ and PDF $g_X(x_i) = \gamma(1 - W_Z(x_i))^{\gamma-1} w_Z(x_i)$ when H_0 is true.

From the discussion in the previous sections, we know that each sensor should apply local likelihood ratio tests as the local decision rules, and the local likelihood ratio test threshold for i-th sensor is given by

$$\lambda_i = \frac{f_X(x_i)}{g_X(x_i)} = \xi(1 - W_Z(x_i))^{\beta-\gamma},$$

where $\xi = \frac{\beta}{\gamma}$. The detection probability P_D and the false alarm probability P_F for i-th sensor are therefore

$$P_D(\lambda_i) = \left(\frac{\xi}{\lambda_i}\right)^{\tilde{\beta}}$$

and

$$P_F(\lambda_i) = \left(\frac{\xi}{\lambda_i}\right)^{\tilde{\gamma}},$$

where $\tilde{\beta} = \frac{\beta}{\gamma-\beta}$ and $\tilde{\gamma} = \frac{\gamma}{\gamma-\beta}$. We also have

$$P_D(P_F(i)) = P_F(i)^{\xi}$$

and

$$\lambda_i = \frac{dP_D(P_F(i))}{dP_F(i)} = \xi \frac{P_D(P_F(i))}{P_F(i)},$$

where by abusing the notations, $P_F(i)$ and $P_D(P_F(i))$ are the false alarm probability and the detection probability for the i-th sensor, respectively. Hence, the ROC of this classification problem is of the same form as the aforementioned classification of exponential sources problem. The discussion for the classification of exponential sources can accordingly be well-fit to this problem.

The second example is an analogue of the first example. Consider the following hypothesis testing problem:

$$H_1 : X_i = \max(Z_{i,1}, \cdots, Z_{i,\beta})$$

versus

$$H_0 : X_i = \max(Z_{i,1}, \cdots, Z_{i,\gamma})$$

for $i = 1, 2, \ldots, n$ and $\beta < \gamma$, where X_i is the observation of i-th sensor, $\{Z_{i,j}\}$ are independent and identically distributed random variables with the associated PDF $w_Z(z)$ and CDF $W_Z(z)$. Thus, X_i has CDF $F_X(x_i) = \Pr(\max(Z_{i,1}, \cdots, Z_{i,\beta}) \leq x_i) = W_Z(x_i)^\beta$ and PDF $f_X(x_i) = \beta W_Z(x_i)^{\beta-1} w_Z(x_i)$ when H_1 is true, and has CDF $G_X(x_i) = \Pr(\max(Z_{i,1}, \cdots, Z_{i,\gamma}) \leq x_i) = W_Z(x_i)^\gamma$ and PDF $g_X(x_i) = \gamma W_Z(x_i)^{\gamma-1} w_Z(x_i)$ when H_0 is true.

The local likelihood ratio test threshold for i-th sensor is

$$\lambda_i = \frac{f_X(x_i)}{g_X(x_i)} = \xi W_Z(x_i)^{\beta-\gamma},$$

where $\xi = \frac{\beta}{\gamma}$. The detection probability P_D and the false alarm probability P_F for i-th sensor are equal to

$$P_D(\lambda_i) = 1 - \left(\frac{\xi}{\lambda_i}\right)^{\tilde{\beta}}$$

and

$$P_F(\lambda_i) = 1 - \left(\frac{\xi}{\lambda_i}\right)^{\tilde{\gamma}},$$

where $\tilde{\beta} = \frac{\beta}{\gamma-\beta}$ and $\tilde{\gamma} = \frac{\gamma}{\gamma-\beta}$. By the above setting, we immediately have

$$1 - P_D(P_F(i)) = (1 - P_F(i))^\xi$$

and

$$\lambda_i = \frac{dP_D(P_F(i))}{dP_F(i)} = \xi \frac{1 - P_D(P_F(i))}{1 - P_F(i)},$$

where $P_F(i)$ and $P_D(P_F(i))$ are again the false alarm probability and the detection probability for the i-th sensor, respectively. Hence, the ROC of this classification problem is also a mirror of the ROC of the classification of exponential sources. Consequently, with slight modification, we can have similar results as the classification of exponential sources problem.

4.8.1 Decentralized Classification of Heavy-tailed Sources Problems

The heavy-tailed distributions, specifically the Pareto distribution, are related to the self-similar phenomena in a way that if the packet inter-arrival process is modelled as i.i.d. Pareto random variables, the packet counting process is asymptotically second-order self-similar process with $H = (3 - \alpha)/2$, where α is the Pareto parameter.

In practical control of the network traffic, one might need to test whether its self-similarity is weak or strong to determine whether the long-range dependence can or cannot be ignored. To reduce the response time and to alleviate the load of network, a decentralized scheme for the detection of the self-similarity might be useful. As a result, one might need to consider the following binary hypothesis testing problem:

$$H_1 : f_X(x_i) = \beta \frac{1}{x_i^{\beta+1}}$$

versus

$$H_0 : g_X(x_i) = \gamma \frac{1}{x_i^{\gamma+1}},$$

or equivalently,

$$H_1 : F_X(x_i) = 1 - \frac{1}{x_i^{\beta}}$$

versus

$$H_0 : G_X(x_i) = 1 - \frac{1}{x_i^{\gamma}}$$

for $i = 1, 2, \ldots, n$, $\beta < \gamma$ and $x_i \geq 1$, where x_i is the observed value of the random variable X_i with the associated PDF $f_X(x_i)$ and CDF $F_X(x_i)$. Here, we assume that $\{X_i\}$ form a set of independent and identically distributed random variables. For a fixed fusion rule, the local likelihood ratio tests are

$$\frac{\frac{\beta}{x_i^{\beta+1}}}{\frac{\gamma}{x_i^{\gamma+1}}} \gtreqless \lambda_i,$$

or equivalently,

$$x_i \gtrless t_i$$

for $i = 1, 2, \ldots, n$, where λ_i and t_i are some constants to be decided, and $t_i = (\lambda_i \frac{\gamma}{\beta})^{\frac{1}{\gamma-\beta}}$.

It turns out that $P_D(\lambda)$ and $P_F(\lambda)$ of the above testing problem have the same forms as the classification of exponential sources problem in Section 4.1; hence, the previous result can be applied to the testing problem for the Pareto distributions directly.

4.9 Gaussian Classification Problems

All the previous parts in this chapter discuss mainly on the classification of exponential sources problem (or problems with the same ROC). In this section, we briefly discuss another classification problem that has drawn more attention among researchers, i.e., the classification of Gaussian sources problem.

Let us introduce some notations first.

Definition 4.2. If X is a Gaussian random variable with mean μ and variance σ^2, then it has a probability density function

$$c_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{\phi(\frac{x-\mu}{\sigma})}{\sigma},$$

and a distribution function

$$C_X(x; \mu, \sigma) = \int_{-\infty}^{x} c_X(x; \mu, \sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the probability density function and cumulant distribution function of the standard normal distribution, i.e., the Gaussian distribution with $\mu = 0$ and variance $\sigma = 1$.

We then concern the following binary hypothesis testing problem for Gaussian distributions:

$$H_1 : P(x_i) = c(x_i; \mu, 1)$$

versus

$$H_0 : P(x_i) = c(x_i; -\mu, 1)$$

for $i = 1, 2, \ldots, n$ and $\beta < \gamma$, where x_i is the observation of i-th sensor, and without loss of generality, we assume $\sigma = 1$. For a fixed fusion rule, it is known that the optimal local decision rules are local likelihood ratio tests, namely,

$$\frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+\mu)^2}{2\sigma^2}}} \gtreqless \lambda_i$$

or equivalently,

$$x_i \gtreqless \eta_i$$

for $i = 1, 2, \ldots, n$, where λ_i and η_i are some constants to be decided.

The optimal strategy of the two-sensor system under this setting has been solved in [34], in which they showed analytically that the identical local decision rules are optimal. However, for the system with more than two sensors, the desired result that identical sensor system is optimal is still absent. Here, we offer an alternative argument that is partly built on numerical results. Firstly, let us examine the conditions in Lemma 4.7. For Gaussian classification problem, it is easy to show analytically that the second and third conditions are satisfied. As for the first condition, we can show numerically that it is valid for relatively small n. Thus, we can conclude from Theorem 4.3 that the optimal three-sensor system still employ identical sensors.

Chapter 5

Conclusions

5.1 Self-Similar Traffic Generators

In the first part of this dissertation, we propose a filter-based generator for the synthesization of self-similar traffics. It can produce long range dependent traffics with adjustable levels of bustiness and correlation, and is parsimonious in the number of model parameters. Precisely, only three input parameters are required, i.e., the self-similar parameter H (which controls the bustiness and autocorrelation of the synthesized traffic), the mean of the traffic λ, and the length of the filter W (which also determines the effective aggregation size in the variance-time analysis). Despite the finite time scales of the self-similar phenomenon in the synthesized traffic, it actually agrees with the measured behavior of true network traffic, i.e., the self-similar nature only lasts beyond a practically manageable range, but disappears as the considered aggregated window is much further extended [4, Fig. 2]. When it is compared with exiting self-similar traffic synthesizers, e.g., the RMD and the Paxson IFFT algorithm, the proposed filter-based synthesizer has the advantages that the synthetic traffic can be generated on the fly, and always produces non-negative valued traffic.

Comparisons of the complexities of self-similar traffic generators are as follows. Given that the length of the synthesized traffic is n, the number of complex multiplications required for the Paxson IFFT method [19] is about $(n/2)(\log_2 n + 2)$. Our filter-based approach, on the other hand, requires $n \times W$ complex multiplications, where W represents the truncation window size. After analytically analyzing our approach based on variance-time test, we conclude that our synthesizer guarantees the generation of a traffic with desired degree of self-similarity beyond the intended range.

5.2 Correlation Approximation to the Mutual Information of Self-Similar Processes

We discuss the implications between the correlation coefficient (a quantity that only measures the *linear* dependance) and mutual information (a quantity that can represent the *general* dependance) in Chapter 3. We focus on the question that given the correlation coefficients of random sources, what is the minimum possible value of mutual information? Theorem 3.1 then suggests that for weakly correlated random variables, such as two instances of a self-similar process with a long time lag, half the square of the correlation coefficients is a reasonable approximation to the mutual information, provided they are also weakly dependent in a general sense.

5.3 Bayesian Decentralized Detection

Our investigation of the optimal decentralized system has yielded some interesting results. Firstly, for the classification of exponential sources problem, the optimality of identical sensor system has been proved for $n = 2$ and $n = 3$. For $n > 3$, we have to rely partly on numerical examination. A byproduct is that for the classification of exponential sources problem, the optimal performance of the optimal serial two-sensor system is the same as the optimal parallel two-sensor system. It is somewhat surprising since it is known that the serial two-sensor system in general has better performance than the parallel two-sensor system [32].

For the general classification problem, we propose a set of propositions on the optimality of the identical system. These propositions can be verified without much difficulty. Moreover, we point out that some classification problems encountered in the survival analysis and failure time analysis, as well as the decentralized detection for the self-similarity via the local measurements of the packet inter arrival times, can be manipulated in the same way. Finally, for the Gaussian classification problem, we conclude the optimality of identical sensors partly numerically.

Bibliography

[1] Allan Theodor Andersen, *Modeling of Packet Traffic with Matrix Analytic Methods*, PhD Thesis, Technical University of Denmark, Denmark, September 1995.

[2] P. Arby and D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 2-15, 1998.

[3] J. Beran, *Statistics for Long-Memory Processes*, New York: Chapman and Hall, 1994.

[4] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 2–4, pp. 1566–1579, February/March/April 1995.

[5] R. N. Bhattacharya, V. K. Gupta and E. Waymire, "The Hurst effect under trends," *Journal of Applied Probability*, vol. 20, pp. 649-662, 1983.

[6] S. Bose, *An Introduction to Queueing Systems*, Springer, USA, 2001.

[7] D. R. Cox, "Long-range dependence: A review," *Statistics: An Apprisal*, H. A. David and H. T. David, Eds., The Iowa State Univ. Press, Ames, Iowa, pp. 55–74, 1984.

[8] D. E. Duffy, A. A. Mclntosh, M. Rosenstein and W. Willinger, "Statistical analysis of CCSN/SS7 traffic data from working subnetworks," *IEEE J. Select. Areas Commun.*, 12, pp. 544–551, June 1994.

[9] P. Flandrin, “Wavelet qnalysis and wynthesis of rractional Brownian motion,” *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 910–917, March 1992.

[10] M. W. Garrett and W. Willinger, “Analysis, modeling and generation of self-similar VBR video traffic,” *Proceedings of the ACM SIGCOMM ’94 Conference on Communications Architectures, Protocols and Applications*, pp. 269-280, London, UK, 1994

[11] J. Gordon, “Pareto process as a model of self-similar packet trafficc,” *In IEEE Proc. GLOBECOM’95*, vol. 3, pp. 2232–2236, Singapore, November 1995.

[12] M. Krunz and A. Makowski, “A source model for VBR video traffic based on the $M/G/\infty$ input processes,” *Proceedings IEEE INFOCOM ’98, the Conference on Computer Communications*, pp. 1441-1449, San Francisco, CA, USA, 1998.

[13] W.-C. Lau, A. Erramilli, J. Wang and W. Willinger. “Self-similar traffic generation: The random midpoint eisplacement qlgorithm and its properties,” *IEEE Proc. ICC’95*, pp. 466–472, Seattle, WA, 1995.

[14] W. Leland, M. Taqqu, W. Willinger and D. Wilson, “On the self-similar nature of ethernet traffic (Extended version),” *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1–15, February 1994.

[15] B. B. Mandelbrot, “A fast fractional Gaussian noise generator,” *Water Resources Research*, vol. 7, pp. 543–553, 1971.

[16] B. B. Mandelbrot and J. W. Van Ness, “Fractional Brownian motions, fractional noises and applications,” *SIAM Rev.*, vol. 10, pp. 422–437, 1968.

[17] B. B. Mandelbrot and J. R. Wallis, “Computer experiments with fractional Gaussian noises,” *Water Resources Research*, vol. 5, pp. 228–267, 1969.

[18] K. Meier-Hellstern, P. E. Wirth, Y. L. Yan and D. A. Hoeflin, "Traffic models for ISDN data users: Office automation application," *Proc. 13th ITC*, Copenhagen, Denmark, pp. 167–172, 1991.

[19] V. Paxson, "Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic," *Computer Communication Review*, vol. 27, no. 5, pp. 5–18, Oct. 1997.

[20] V. Paxson and S. Floyd, "Wide-area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Networking*, vol. 3, no. 3, pp. 226–244, June 1995.

[21] Jon M. Peha, "Protocols can make traffic appear self-similar," *Proc. the 1997 IEE/ACM/SCS Communication Networks and Distributed Systems Modeling and Simulation Conference*, 1997.

[22] P. Pruthi, *An Application of Chaotic Maps to Packet Traffic Modeling*, PhD Thesis, Royal Institute of Technology, Stockholm, Sweden, October 1995.

[23] William Stallings, *High-Speed Networks: TCP/IP and ATM Design Principles*, Prentice-Hall International, Inc., pp. 181–208, 1998.

[24] B. Tsybakov and N. D. Georganas, "Self-similar processes in communications networks," *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1713–1725, September 1998.

[25] B. Tsybakov and N. D. Georganas, "On self-similar traffic in ATM queues: Definitions, overflow probability bound, and cell delay distribution," *IEEE/ACM Trans. Networking*, vol. 5, no. 3, pp. 397–409, June 1997.

[26] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, 1991.

[27] G. A. Darbellay, "The mutual information as a measure of statistical dependence," *Intl. Symp. on Information Theory*, pp. 405, June 1997.

[28] W. Li, "Mutual information functions versus correlation functions," *J. Stat. Phys., 60*, pp. 823–837, 1990.

[29] T. M. Apostol, *Calculus (Volumn I): One-Variable Calculus with an Introduction to Linear Algebra*, 2nd edition, 1967.

[30] P. Billingsley, *Probability and Measure*, 3rd edition, Wiley, 1995.

[31] G. A. Darbellay and I. Vajda, "Entropy expressions for multivariate continuous distributions," *Research Report no 1920*, UTIA, Academy of Sciences, Prague, February 1998. (http://siprint.utia.cas.cz/darbellay/)

[32] P. K. Varshney, *Distributed Detection and Data Fusion*, 1st edition, Springer-Verlag, 1996.

[33] P. Willett and D. Warren, *The suboptimality of randomized tests in distributed and quantized detection system*, *IEEE Trans. Inform. Theory*, vol. 38, pp. 355-361, March 1992.

[34] W. W. Irving and J. N. Tsitsiklis, *Some properties of optimal thresholds. in decentralized detection*, *IEEE Trans. Autom. Control*, vol. 39, pp. 835-838, April 1994.

[35] R. R. Tenney and N. R. Sandell, *Detection with distributed sensors*, *IEEE Trans. Aerospace Electron. Syst.*, vol. 17, pp. 501X510, July 1981.

[36] P.-N. Chen and A. Papamarcou, *New asymptotic results in parallel distributed detection*, *IEEE Trans. Inform. Theory*, vol. 39, no. 6, pp. 1847-1863, November 1993.

[37] J. N. Tsitsiklis, *Decentralized detection by a large number of sensors*, *Math. Contr. Signals Syst.*, vol. 1, no. 2, pp. 167V182, 1988.

[38] Q. Zhang, P. K. Varshney and R. D. Wesel, *Optimal bi-level quantization of i.i.d. sensor observations for binary hypothesis testing*, *IEEE Trans. Inform. Theory*, vol. 48, pp. 2105-2111, July 2002.

Vita

Chien Yao

Date of Birth

August 5, 1975

Place of Birth

Chung-li, Taiwan

Educations

- Ph.D, Institute of Electronics, National Chiao-Tung University, Taiwan, **2008**.
  Advisor: Professor Tihao Chiang and Professor Po-Ning Chen

- Field: Self-Similar Traffic Model, Distributed Detection, Information Theory.

- Dissertation: “Synthesization and Decentralized Identification of Self-Similar Processes”

- M.S., Institute of Electronics, National Chiao-Tung University, Taiwan, **1999**.
  - Field: Solid State Physics, Surface Science, Plasma Physics.
  - Master Thesis: “Effects of Surface Excitations on the Energy Loss of Moving Charge”

- B.S., Dept. of Electronics Engineering, National Chiao-Tung University, Taiwan, **1997**.

Publications

- Thesis

  1. Ph.D Dissertation: "Synthesization and Decentralized Identification of Self-Similar Processes", Institute of Electronics, National Chiao-Tung University, Taiwan, 2008.

  2. Master Dissertation: "Effects of Surface Excitations on the Energy Loss of Moving Charge", Institute of Electronics, National Chiao-Tung University, Taiwan, 1999.

- Journal Papers

  1. Chien Yao, Po-Ning Chen, Tsang-Yi Wang, Yunghsiang S. Han and Pramod K. Varshney, "Performance analysis and code design for minimum hamming distance fusion in wireless sensor networks," *IEEE Trans. Inform. Theory*, vol. 53, no. 5, pp. 1716-1734, May 2007.

  2. Chien Yao, Kai-Lung Hua, Po-Ning Chen, Jin-Yuan Chen and Tihao Chiang, "A filter-based self-similar trace synthesizer," *Journal of the Chinese Institute of Engineers (JCIE)*, vol. 30, no. 3, pp. 379-387, May 2007.

  3. C. M. Kwei, J. J. Chou, J. Yao and C. J. Tung, "Electronic stopping powers of solids for slow atoms," *Phys. Rev. A* 64, pp. 0429011-0429015, October 2001.

- Conference Papers

1. Sou-Yu Tseng, Po-Ning Chen, Yunghsiang S. Han, Tsang-Yi Wang and Chien Yao, "Equalization of the correlated additive Gaussian noises on the wireless links of sensor networks," *National Symposium on Telecommunications*, Wu-Sung, Kaohsiung County, Taiwan (Cheng Shiu Univ.), December 1-2, 2006.

2. Po-Ning Chen, Tsang-Yi Wang, Yunghsiang S. Han, Pramod K. Varshney and Chien Yao, "Fault-tolerance analysis of a wireless sensor network with distributed classification codes," *IEEE International Symposium on Information Theory*, Seattle, Washington, USA, July 9-July 14, 2006.

3. Po-Ning Chen, Tsang-Yi Wang, Yunghsiang S. Han, Pramod K. Varshney and Chien Yao, "Asymptotic analysis for minimum-Hamming-distance fusion," *IEEE International Conference on Speech, Acoustics, and Signal Processing* (ICASSP 2005), Philadelphia, PA, USA, March 19-23, 2005.

4. Chien Yao, Po-Ning Chen and Tihao Chiang, "Correlation approximation to the mutual information of self-similar processes," *IEEE International Symposium on Information Theory and its Applications*, Parma, Italy, October 10-13, 2004.

5. Chien Yao and Po-Ning Chen, "Shannon theory for hypothesis testing based on smallest decision set," *Asia Pacific Conference on Multimedia Technology and Applications (APCMTA)*, Kaohsiung, Taiwan, December 17-19, 2000.