Naval Research Laboratory
Washington, DC 20375-5320

UNCLASSIFIED

NRL/FR/5591--94-9735

On Commonalities in Signal Design for Non-Gaussian Channels

NHI-ANH CHU

Center for Computational Science
Information Technology Division

NIRMAL WARKE
GEOFFREY C. ORSAK

Department of Electrical and Computer Engineering
George Mason University

August 26, 1994

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 26, 1994	Final

4. TITLE AND SUBTITLE

On Commonalities in Signal Design for Non-Gaussian Channels

5. FUNDING NUMBERS

PE - 63011F

6. AUTHOR(S)

Nhi-Anh Chu, Nirmal Warke,* and Geoffrey C. Orsak*

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5320
George Mason University, EE Dept.
Fairfax, VA 22030-4444

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/FR/5591--94-9735

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Research Laboratory (Edison Memorial Scholarship)
Nat'l Science Foundation (Under following grants: NCR 9109858 and NCR 9109858-01)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

*Department of Electrical and Computer Engineering
George Mason University

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We derive for additive non-Gaussian white noise channels signal waveforms that are simultaneously optimal with respect to the minimum probability of error, the mini-max, and the Neyman-Pearson criteria. We show that for a large class of non-Gaussian statistics, there exist only two asymptotically optimal signal waveforms; one impulsive while the other is constant in amplitude. The impulsive waveform is optimal when the tails of the noise density fall off faster than the tails of the Gaussian density.

We show that under each of the three optimality criteria the asymptotic performance for small signals is essentially determined by the signal energy, while for large signals the performance is determined by a non-Euclidean metric that varies with respect to the tails of the noise density function. To support these results, we offer simulations for a variety of non-Gaussian channels. In each case, the asymptotic theory holds strikingly well even for decidedly nonasymptotic regimes.

14. SUBJECT TERMS

Optimal detectors
Non-Gaussian
Information theory

15. NUMBER OF PAGES

46

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18
298-102

UNCLASSIFIED

CONTENTS

ON COMMONALITIES IN SIGNAL DESIGN FOR NON-GAUSSIAN CHANNELS

1. INTRODUCTION

Signal design is an important aspect in the overall design of a communication system. Ideally, the optimal signal set, subject to certain constraints such as bandwidth or energy, minimizes either the probability of error (P_e) or the Neyman-Pearson performance (NP). Even in the fortuitous case that the channel is modeled as an additive Gaussian noise channel (possibly colored), there are few analytic results. Moreover, if the noise happens to be non-Gaussian in nature, the design problem as described above becomes for all practical purposes analytically intractable.

To address this in this report we apply and extend results from Large Deviation Theory (LDT) to the problem of signal design for non-Gaussian channels. Originally, Johnson and Orsak considered this approach in Ref. 1 where they focused on the design of signal waveforms which were asymptotically optimal with respect to the Neyman-Pearson criterion. We seek to generalize these results to determine signal sets that are simultaneously asymptotically optimal with respect to the minimum probability of error (P_e), the mini-max, and the Neyman-Pearson criteria.

One of the main issues addressed in this research is best summarized by the following question: Are signal sets operating in non-Gaussian environments that are optimal with respect to the Neyman-Pearson criterion also optimal with respect to the minimum P_e and the mini-max criteria? Through this work we are able to conclusively answer "yes," provided that the length of the signal vector grows without bound.

In LDT, the $min\, P_e$ and mini-max criteria are associated with the Chernoff Distance, and the NP criterion with the Kullback-Leibler distance. We are able to establish that a unique optimal signal (if exists) maximizes both of these distances. Significantly, we show that this maximality extends over the whole class of Ali-Silvey distances.

We show within this report that if the background noise is accurately modeled as a discrete-time generalized Gaussian random process, then there are only two optimal signal sets with respect to all of the above optimality criteria. If the tail of the noise distribution diminishes faster than that of the Gaussian, then the optimal signal waveform subject to an energy constraint is an impulse, that is, all of the energy is contained in a single sample of the signal waveform. Conversely, if the tail diminishes slower than that of the Gaussian, then the optimal signal has constant amplitude over the waveform. Only in the case of additive Gaussian noise is a time-varying signal (except for a purely impulsive signal) potentially[1] optimum. So, as a by-product, this work implies that sinusoidal waveforms can only be optimal for the purely Gaussian channel.

Manuscript approved June 21, 1994

[1] In fact, we know that in the case of an AWGN channel, only the total energy of the signal determines performance.

In addition to determining the optimal signal waveform, we have been able to analytically compare the relative performance of these designs with respect to the three optimality criteria of interest. Results will show that for "small" signal energies, the error exponents associated with the minimum P_e and mini-max criteria are one fourth of the error exponent of the miss probability (P_M) in the Neyman-Pearson criterion for all non-Gaussian channels. Thus, when signal energies are small, four times as much energy or four times as much data are required to achieve the same error exponent in the minimum P_e/mini-max performance as that required for the NP performance.

Conversely, for "large" signal energies, we have shown that the error exponents for the minimum P_e and mini-max criteria are no more than one half the error exponent for P_M under Neyman-Pearson criterion. This is to be expected since the Neyman-Pearson detector need only minimize P_M while the minimum P_e detector must simultaneously minimize P_F (false alarm rate) and P_M and therefore can commit no more than one half of the computation capability of the likelihood ratio test to either of the two error probabilities.

Even stronger results are obtained for the case of large signal energies when the background noise is assumed to be from the Generalized Gaussian family with decay rate r, i.e., when the noise density is modeled as $p_\eta(x) = K_1 \exp(-K_2|x|^r)$. If $r \geq 1$, then we have shown that the error exponent of the minimum P_e or mini-max performance is $1/2^r$ of the error exponent of P_M under Neyman-Pearson constraints. However, if $r < 1$, then the minimum P_e/mini-max error exponent is precisely one half of the error exponent of P_M under the Neyman-Pearson criterion. Therefore, as in the small energy case for $r > 1$, to equate the error exponents, one must utilize precisely four times as much energy in the minimum P_e/mini-max detection scheme as that used in the NP scheme. However, if $r < 1$, then one is required to utilize $1/2^{2/r}$ times more energy under minimum P_e/mini-max consideration as that used in NP considerations.

To support this theory, we have included Monte Carlo simulations. These results show that the asymptotic results hold with striking precision even when in decidedly non-asymptotic regimes.

2. PREVIOUS WORK

As described in the introduction, Johnson and Orsak [1] were apparently the first to use Large Deviation[2] approaches to design signal waveforms for the non-Gaussian channel.

The results in this report were based upon a generalization of Stein's lemma first offered by Kullback [3]. It was shown that under Neyman-Pearson optimality criterion, the error exponent of the miss probability is asymptotically equivalent to the average Kullback-Leibler distance (also known as the divergence) between the probability measures corresponding to the two hypotheses. From this, the "optimal" signal waveform in an additive non-Gaussian channel was determined by maximizing the Kullback-Leibler distance subject to an energy constraint on the signal waveform.

It should be pointed out that others have also considered maximizing the divergence (or other specific statistical distance measures) [4, 5, 6] between hypotheses as a means of designing "good"

[2]Large Deviation Theory (LDT) is used to estimate the probabilities of rare events [11]. For a binary detection problem, we are in the regime of large deviation when the separation between the probabilities of the two hypotheses is sufficiently large [2].

signal waveforms. Grettenberg [7] first proposed the maximum divergence criterion for the Gaussian channel based upon a duality result originating with the work by Bradt and Karlin [8] where it was shown that the maximum divergence criterion rendered the minimum probability of error signal waveform for *some* a priori probabilities on the hypotheses. Based upon this result, Grettenberg was able to establish that the Simplex Conjecture must hold for some set of input a priori probabilities.

Unfortunately, as pointed out by Kailath [9], there is no guarantee that the true a priori probabilities will match those required by this duality principle. In addition, in this work Kailath offered an alternate statistical distance measure known as the Bhattacharyya distance as a means of determining the optimal energy allocation in a Gaussian environment. Empirical results seemed to suggest that the Bhattacharyya distance offered solutions that were more consistent with those derived by considering the probability of error as an optimality criterion. Nevertheless, as in the case of the maximum divergence criterion, the maximum Bhattacharyya distance waveforms were not guaranteed to be optimum for the true a priori probabilities.

In this work, we generalize the results offered in Ref. 1 to consider not only the NP criterion, but also the minimum P_e and mini-max criterion for the non-Gaussian environment. The signal waveforms obtained from this analysis will asymptotically minimize the desired performance for every set of a priori probabilities and therefore will not suffer from the same kinds of theoretical limitations as those in Refs. 7 and 9.

3. RELATING PERFORMANCE TO CERTAIN STATISTICAL DISTANCE MEASURES

Consider the following binary detection problem where an N-dimensional vector is transmitted through an additive iid noise channel:

$$\begin{aligned} H_0 &: X_i = n_i \\ H_1 &: X_i = n_i + s_i, \quad n_i \overset{iid}{\sim} p_n. \end{aligned} \tag{1}$$

Absolute signal location does not determine performance when the noise density is symmetric. As such, without loss of generality, we have considered an "on-off" signaling scheme[3] where the observation and the signal vectors of interest will be denoted to as X^N and s^N respectively. We will assume throughout that the density function of the noise p_n is symmetric and monotonically decreasing.

The optimal detector computes the log-likelihood-ratio test (LLRT) based on the aggregate of N samples and compares the output to threshold γ:

$$\mathcal{L}(X^N) = \frac{1}{N}\sum_{i=1}^{N} \log \frac{p_n(x_i - s_i)}{p_n(x_i)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma\,, \tag{2}$$

[3]Consider for example sending binary symbols $W = \{0, 1\}$ in a quadrature phase modulation system. Symbol $W = 1$ selects a set of signal samples $s_i, i = 1, ..., N$ and generates a phase signal $s(t)$ where $s(iT_s) = s_i$ and T_s is the sampling period. Symbol $W = 0$ generates $s(t) = 0$ for the symbol interval. Two passband waveforms are generated for each symbol interval: $f_q(t) = \sin(\omega_c t + s(t))$ and $f_i(t) = \cos(\omega_c t + s(t))$. At the receiver, the quadrature-modulated signals are demodulated, and combined to recover a noisy version of the sequence $\hat{s}_i = s_i + n_i$.

where the threshold is chosen to optimize some performance measure. The false alarm and miss probabilities that arise from the N-dimensional hypothesis testing problem are defined as $\alpha_N = \Pr\{say\ H_1|H_0\}$ and $\beta_N = \Pr\{say\ H_0|H_1\}$, respectively [10].

In this report, we seek to determine the signal waveform s^N that simultaneously optimizes each of the following three criteria:

1 *(Neyman-Pearson)* minimize β_N such that $\alpha_N \le \alpha$.

2 *(minimum P_e)* minimize $\pi_0\alpha_N + \pi_1\beta_N$ where $\pi_i = Pr[H_i]$.

3 *(mini-max)* minimize maximum $\{\alpha_N, \beta_N\}$

For the general non-Gaussian channel, the optimal signal waveform under any of the above optimality criteria is analytically intractable. However, if we allow the length of the signal vector to grow without bound, we may readily relate the above performance criteria to information theoretic quantities that are more amenable to analysis. This is accomplished through the application of results from LDT to the detection problem.

In the case of the Neyman-Pearson criterion, we have via a generalized version of Stein's lemma [2, 11] the following asymptotic result:

Theorem 1 *Let α_N satisfy $\alpha_N \le \alpha$ where $\alpha > 0$. Then*

$$\lim_{N\to\infty} \frac{1}{N} \log \min \beta_N = -\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} d_{KL}(s_i) \tag{3}$$

where $d_{KL}(s_i)$ is the Kullback-Leibler distance (or divergence) between $p_n(x)$ and $p_n(x - s_i)$, i.e., $\int \log \frac{p_n(x)}{p_n(x-s_i)} p_n(x)dx$.

This result demonstrates that with respect to the NP criterion, the asymptotic error exponent is determined by the average divergence across the data vector. As such, the asymptotically optimal signal waveform must maximize this average divergence. It was this result that was used in Ref. 1 to design signal waveforms that are optimal for applications where the Neyman-Pearson criterion is appropriate, e.g., radar applications.

However, in most communication applications, one prefers to use either the minimum P_e or mini-max criteria. We may obtain an analogous asymptotic result by offering the following generalization of Sanov's theorem (or sometimes referred to as Chernoff's theorem)[2, 11]. The proof requires establishing asymptotically tight upper bound and lower bound on the error exponent that asymptotically converge to the *Chernoff bound.* The proof, as shown in Appendix A, extends standard versions based on *i.i.d.* random variables to the current problem where individual samples X_i of the vector X^N are independently distributed according to known translations of the noise density, *i.e* $X_i \sim p_{n-s_i}$.

Theorem 2 *Let $P_e^N = \pi_0 \alpha_N + \pi_1 \beta_N$. Then*

$$
\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} \log \min P_e^N &= \lim_{N\to\infty} \frac{1}{N} \log \min \max \{\alpha_N, \beta_N\} && (4) \\
&= -\lim_{N\to\infty} \frac{1}{N} \max_{\lambda}\{-\sum_{i=1}^{N} \log \int p_n^{\lambda}(x) p_n^{1-\lambda}(x - s_i) dx\} && (5) \\
&= -\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} d_C(s_i), && (6)
\end{aligned}
$$

where the so-called Chernoff distance $d_C(s)$ is given by $\max_\lambda\{-\log \int p_n^{\lambda}(x) p_n^{1-\lambda}(x-s)dx\}$

As opposed to NP considerations, in this case, the asymptotic error exponent for both the minimum P_e and the mini-max detectors is determined by the average Chernoff distance across the data vector. Thus, under these optimality criteria, the asymptotically optimal signal waveform must maximize the average Chernoff information.

Summarizing, under the consideration that the length of the signal vector grows without bound, the NP performance is determined by the statistical distance measure d_{KL}, whereas the minimum P_e/mini-max performance is determined by the statistical distance measure d_C. This observation clearly begs the question: How are d_{KL} and d_C related in the non-Gaussian environment? To address this, in the following section we consider two signal regimes, one being the case where the admissible signal energy is small and the other being the case where the admissible energy is large.

4. RELATION BETWEEN THE CHERNOFF INFORMATION AND THE DIVERGENCE

The Chernoff Information and Kullback-Leibler distances are computed directly (see Appendix B) for a selection of pdfs as shown in Table 1. One can observe that for the non-Gaussian noise models considered, there is very little functional similarity between d_C and d_{KL}. However, we

Table 1 — Chernoff Information and Kullback-Leibler Distance

Noise	$P_n(x)$	$d_C(s)$	$d_{KL}(s)$
Gaussian	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$	$\frac{s^2}{8\sigma^2}$	$\frac{s^2}{2\sigma^2}$
Cauchy	$\frac{1}{\pi\sigma} \frac{1}{1+(\frac{x}{\sigma})^2}$	$-\log(\frac{2}{\pi} \frac{1}{\sqrt{\sqrt{1+\frac{s^2}{2}+\frac{s^4}{16}}}} E_k[\frac{1}{2} - \frac{(\frac{1}{2}-\frac{s^2}{8})}{\sqrt{1+\frac{s^2}{2}+\frac{s^4}{16}}}])$	$\log(1 + \frac{s^2}{4\sigma^2})$
Laplacian	$\frac{1}{\sigma\sqrt{2}} e^{\frac{-\lvert x\rvert\sqrt{2}}{\sigma}}$	$\frac{\lvert s\rvert}{\sigma\sqrt{2}} - \log(1 + \frac{\lvert s\rvert}{\sigma\sqrt{2}})$	$\frac{\lvert s\rvert\sqrt{2}}{\sigma} + e^{-\frac{\sqrt{2}\lvert s\rvert}{\sigma}} - 1$
Gen.Gauss4	$\frac{1}{2\Gamma(1+\frac{1}{r})A(r)} \exp\left\{-\left[\frac{\lvert x\rvert}{A(r)}\right]^r\right\}$ $A(r) = [\sigma^2 \frac{\Gamma(1/r)}{\Gamma(3/r)}]^{\frac{1}{2}} (r = 4)$	$-\frac{7}{32}\frac{s^4}{A^4} - \log \frac{\sqrt{\frac{3}{32}\lvert s\rvert}}{\Gamma(\frac{5}{4})A} - \log K_{\frac{1}{4}}(\frac{9}{32}\frac{s^4}{A^4})$	$\frac{\Gamma^2(\frac{3}{4})}{\Gamma^2(\frac{1}{4})}[6\frac{s^2}{\sigma^2} + \frac{s^4}{\sigma^4}]$

Note: E_k is the complete elliptical integral of the first (real argument) and second (imaginary argument) kind. $K_{1/4}$ is the modified Bessel function of the second kind of fractional order 1/4.

will be able to analytically demonstrate that there are quite strong commonalities for these two distance measures when one considers both small and large signal displacements s.

Before proceeding, we wish to make an interesting observation regarding the structural relationship between d_C and d_{KL} and establish conditions for the reduction of d_C to the negative exponent of the Bhattacharyya distance.

4.1. Bhattacharyya Distance, Kullback-Leibler Distance and Chernoff Information

Let $d_C(\lambda) = -\log \int_{-\infty}^{\infty} p^{\lambda}(x)p^{1-\lambda}(x-s)dx$ then

Fact 1 *For a fixed value of s, the derivative of the Chernoff Information at $\lambda = 0, 1$ is equivalent to the Kullback-Leibler distance.*

Proof:

$$\begin{aligned} \frac{\partial}{\partial \lambda} d_C(\lambda) &= \frac{\int p^{\lambda}(x)p^{1-\lambda}(x-s)\log \frac{p(x)}{p(x-s)}dx}{\int p^{\lambda}(x)p^{1-\lambda}(x-s)dx} \\ &= \begin{cases} +d_{KL}(p(x-s),p(x)) & @\ \lambda = 0 \\ -d_{KL}(p(x),p(x-s)) & @\ \lambda = 1 \end{cases} \end{aligned} \tag{7}$$

For symmetric pdfs, $d_{KL}(p(x),p(x-s)) = d_{KL}(p(x-s),p(x))$.□

Fact 2 *For symmetric densities, $d_C(\lambda)$ is both symmetric and concave in λ.*

Proof:

$$\frac{\partial^2}{\partial \lambda^2} d_C(\lambda) = \frac{\left[\int p^{\lambda}(x)p^{1-\lambda}(x-s)\log \frac{p(x)}{p(x-s)}dx\right]^2 - \int p^{\lambda}(x)p^{1-\lambda}(x-s)dx \int p^{\lambda}(x)p^{1-\lambda}(x-s)\log^2 \frac{p(x)}{p(x-s)}dx}{\left[\int p^{\lambda}(x)p^{1-\lambda}(x-s)dx\right]^2} \tag{8}$$

which is negative by Schwarz integral inequality.□

Fact 1 shows that $d_C(\lambda)$ initially increases and eventually decreases with λ at the same rate. Combined with Fact 2 it establishes that there must be a maximum at the center of symmetry $\lambda = \frac{1}{2}$. Figure 1 illustrates this.

Since the Bhattacharyya distance between $p_0(x)$ and $p_1(x)$ is defined as $d_B = \int p_0^{1/2}(x)p_1^{1/2}(x)dx$, we have established that

Fact 3 *For the translation problem of distinguishing between $p(x)$ and $p(x-s)$, the Chernoff Information is equivalent to the negative of the exponent of the Bhattacharyya distance for symmetric densities.*

We use $\lambda = \frac{1}{2}$ and use $d_C = -\log d_B = -\log \int \sqrt{p_n(x)p_n(x-s)}dx$ as the Chernoff Information throughout this report.

Figure 1 also suggests a correlation between the slopes at end points $\pm d_{KL}$ and the maximum value d_C. Namely:

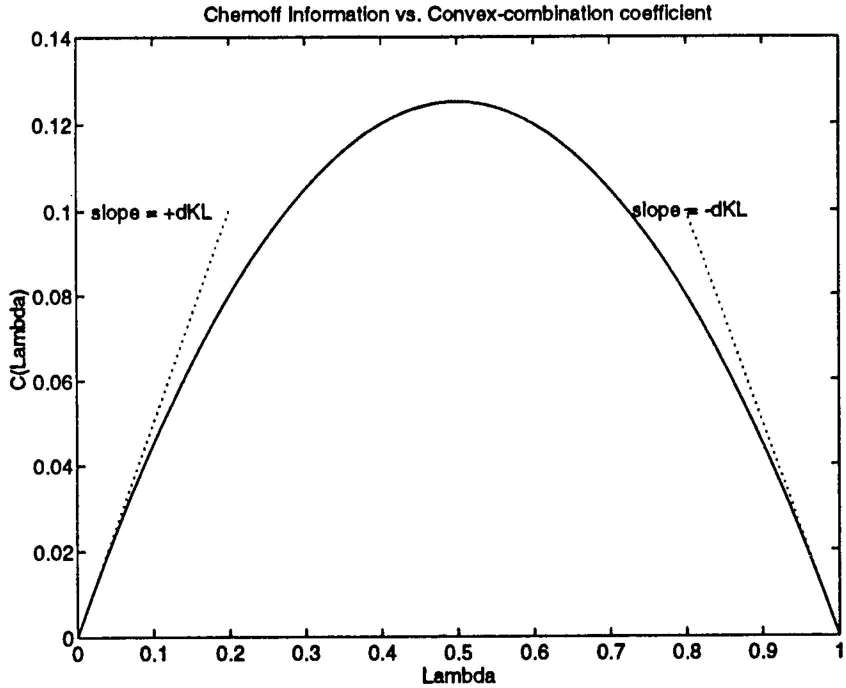

Fig. 1 — Chernoff Information as a function of λ for symmetric noise density functions

Conjecture 3 *It is plausible that for some symmetrical density functions, maximizing the Chernoff Information also maximizes the Kullback-Leibler distances.*

This observation is certainly true for the set of general exponential densities evaluated in this report, as shown in the analyses of small- and large-signal behaviors of the distances in Sections 4.3. and 4.4. We first offer a more general theorem using the concept of Asymptotic Most Favorable Statistics advanced by Orsak and Paris in Ref. 6.

4.2. Asymptotically Most Favorable Signal

Definition 1 *Signal s_*^N is called asymptotically most favorable (AMF) if and only if for any other signal s^N*

$$\lim_{N\to\infty} \frac{\pi_0\alpha(s_*^N) + \pi_1\beta(s_*^N)}{\pi_0\alpha(s^N) + \pi_1\beta(s^N)} \geq 1 \text{ for arbitrary } (\pi_0, \pi_1) \tag{9}$$

where α and β are the false alarm and miss probabilities defined for the binary hypothesis problem, and π_0, π_1 the a priori probabilities.

For optimal detectors, each of the terms of the fraction is the minimum probability of error $\min P_e(s)$ as a function of the signal s.

Lemma 4 *Any real convex function $C(x)$ of real argument x may be expressed in the following form*

$$C(x) = \phi(x) + \sum_{i=1}^{\infty} |(1-\pi_{1,i}) - \pi_{1,i}x| \tag{10}$$

where ϕ is some linear function and $\pi_{1,i} \in (0,1)$, $i = 1, ..., \infty$ are real.

Proof: Let $f(x)$ be $\sum_{i=1}^{\infty} |(1-\pi_{1,i}) - \pi_{1,i}x|$. For $i = 1$, $f(x)$ is a convex function with one degree of freedom controlled by $\pi_{1,1}$, which determines the location of the breakpoint and the slopes. For each additional value of i, the sum continues to be a convex function with an additional degree of freedom. In the limit, $f(x)$ constructs a convex function completely specified by the infinite set of $\{\pi_{1_i}\}$. The linear function $\phi()$ specifies the location of the minimum of the convex function. □

Theorem 5 *The asymptotically most favorable signal s_*^N asymptotically maximizes any f-divergence $d(p_0(s_*^N), p_1(s_*^N))$. Conversely, any signal s^N that asymptotically maximizes any f-divergence is the asymptotically most favorable signal.*

Proof: For any f-divergence of the Ali-Silvey class [4]

$$d(p_0, p_1) = h[\mathcal{E}_{p_0} C(\frac{dp_1}{dp_0})] \tag{11}$$

where h is a real, increasing function, and C is a convex function over $[0, \infty)$. By Lemma 4

$$d(p_0, p_1) = h[\mathcal{E}_{p_0} \phi(\frac{dp_1}{dp_0}) + \sum_{i=1}^{\infty} \mathcal{E}_{p_0} |(1-\pi_{1,i}) - \pi_{1_i} \frac{dp_1}{dp_0}|] \tag{12}$$

Using the identity $\min(a, b) = \frac{1}{2}|a+b| + \frac{1}{2}|a-b|$, with $a = \pi_0 = 1 - \pi_1$ and $b = \pi_1 \frac{dp_1}{dp_0}$, it can be shown that each of the terms $\mathcal{E}_{p_0}|(1-\pi_{1,i}) - \pi_{1,i}\frac{dp_1}{dp_0}|$ is precisely $(1 - 2\min P_{e,i})$, where $\min P_e = \pi_0\alpha + \pi_1\beta$, and $\pi_0 = 1 - \pi_1$.

Therefore,

$$\lim_{N\to\infty} \frac{d(p_0(s_*^N), p_1(s_*^N))}{d(p_0(s^N), p_1(s^N))} = \lim_{N\to\infty} \frac{h[\mathcal{E}_{p_0}\phi(\frac{dp_1(s_*^N)}{dp_0(s_*^N)}) + \sum_{i=1}^{\infty}(1 - 2\min P_{e,i}(s_*^N))]}{h[\mathcal{E}_{p_0}\phi(\frac{dp_1(s^N)}{dp_0(s^N)}) + \sum_{i=1}^{\infty}(1 - 2\min P_{e,i}(s^N))]} \tag{13}$$

The AFM condition of Theorem 5 is satisfied for s_*^N if

$$\lim_{N\to\infty} \frac{\min P_{e,i}(s^N)}{\min P_{e,i}(s_*^N)} \geq 1 \tag{14}$$

for each set of priors $(\pi_{0,i}, \pi_{1,i})$.

Applying this condition, we have:

$$\lim_{N\to\infty} \frac{d(p_0(s_*^N), p_1(s^N*))}{d(p_0(s^N), p_1(s^N))} = \lim_{N\to\infty} \frac{\phi(1)+a+\epsilon}{\phi(1)+a} \geq 1 \text{ for some positive } a, \epsilon \tag{15}$$

The converse may be proved by reversing the steps above. □

Corollary 6 *The signal s^N that maximizes the Chernoff distance between $\{p_0(s^N), p_1(s^N)\}$ also maximizes any f-divergence between these distributions.*

Proof: We have shown in this report that the signal s^N that maximizes the Chernoff distance is optimal under the $\min P_e$ criterion. This criterion is precisely the condition required by Theorem 5. □

4.3. Small Signal Behavior

It was shown [1] that for small signal values, the divergence is locally proportional to an ℓ_2 distance metric (e.g. s^2) where the multiplicative constant is one half of Fisher's information for location. Mathematically this is best stated as

$$\lim_{s\to 0} \frac{d_{KL}(s)}{\frac{s^2\mathcal{I}}{2}} = 1 \tag{16}$$

where $\mathcal{I}$ is Fisher's Information for location. A similar result for the Chernoff distance can be easily established.

Proposition 7 *For diminishingly small values of s, $d_C(s)$ is locally an ℓ_2 distance metric with multiplicative constant $\frac{\mathcal{I}}{8}$, i.e.,*

$$\lim_{s\to 0} \frac{d_C(s)}{\frac{s^2}{8}\mathcal{I}} = 1. \tag{17}$$

Proof: Let $\dot{p}_n(x) \stackrel{\text{def}}{=} \frac{\partial}{\partial x} p_n(x)$ and $\ddot{p}_n(x) \stackrel{\text{def}}{=} \frac{\partial}{\partial x}\dot{p}_n(x)$. Then $d_C(s) = -\log\int p_n(x)^{\frac{1}{2}} p_{n-s}(x)^{\frac{1}{2}} dx$ and the first and second derivatives are

$$\begin{aligned} \frac{\partial}{\partial s} d_C(s) &= \frac{\partial}{\partial s} - \log\int p_n(x)^{\frac{1}{2}} p_{n-s}(x)^{\frac{1}{2}} dx = \frac{\frac{1}{2}\int p_n(x)^{\frac{1}{2}} p_{n-s}(x)^{-\frac{1}{2}} \dot{p}_{n-s}(x) dx}{\int p_n(x)^{\frac{1}{2}} p_{n-s}(x)^{\frac{1}{2}} dx} \\ \frac{\partial^2}{\partial s^2} d_C(s) &= \frac{\frac{1}{2}\int p_n(x)^{\frac{1}{2}} p_{n-s}(x)^{\frac{1}{2}} dx \int p_n(x)^{\frac{1}{2}} (p_{n-s}(x)^{-\frac{1}{2}} \ddot{p}_{n-s}(x) + \frac{1}{2} p_{n-s}(x)^{-\frac{3}{2}} \dot{p}_{n-s}(x)^2) dx}{[\int p_n(x)^{\frac{1}{2}} p_{n-s}(x)^{\frac{1}{2}} dx]^2} + \\ &\quad \frac{\frac{1}{4}\int p_n(x)^{\frac{1}{2}} p_{n-s}(x)^{-\frac{1}{2}} \dot{p}_{n-s}(x) dx}{[\int p_n(x)^{\frac{1}{2}} p_{n-s}(x)^{\frac{1}{2}} dx]^2} \end{aligned} \tag{18}$$

For symmetric and decreasing densities, $\int_{-\infty}^{\infty} \ddot{p}_n(x)dx = 0$ and $\int_{-\infty}^{\infty} \dot{p}_n(x)dx = 0$, therefore

$$-\lim_{s\to 0} \frac{\partial^2}{\partial s^2} d_C = \frac{1}{4}\int \frac{\dot{p}_n(x)^2}{p_n(x)} dx = \frac{1}{4}\mathcal{I}(s) \tag{19}$$

where $\mathcal{I}(s)$ is the Fisher Information of the location s. $d_C(s)$ may be written in terms of a second order Taylor series expansion around $s = 0$ where the constant and the first order terms are zero, and the second order term is given in Eq. 19.□

Thus, these two facts (Eqs. 16 and 17) together show that both the average d_{KL} and average d_C are locally Euclidean metrics. However we recognize that for small signal amplitudes, the error exponent for the minimum P_e/mini-max detector is one fourth of the error exponent of the NP detector. Thus, for small signal energies, this then requires that the minimum probability of error detector utilize four times as much energy as the Neyman-Pearson detector to obtain the same performance (as measured by the error exponent.)

This small signal generalization implies that the local performance under the three criteria of interest depends only on the signal energy and not on the specific waveform. As such, in some sense, this result demonstrates that non-Gaussian environments behave as Gaussian environments when the signal energy is small. One might claim that this is merely the case because we are allowing the length of the data vector to grow without bound and as such the Central Limit Theorem would apply. However, we will show conclusively that this is not the case for all energy constraints.

4.4. Large Signal Behavior

As opposed to the results in the small signal case, we will show that the large signal performance depends explicitly upon the tail of the noise distribution of choice. To begin, as was shown in Ref. 1 that for large signals, the Kullback-Leibler distance is well approximated by the negative of the logarithm of the density function. To be more precise, it was shown that

$$\lim_{s\to\infty} \frac{d_{KL}(s)}{-\log p_n(s)} = 1$$

for noise densities satisfying some very general conditions. Similar large signal results can be demonstrated for the Chernoff distance. To begin, we supply an upper bound which holds for the same general class of distributions considered in Ref. 1.

Fact 4

$$\lim_{s\to\infty} \frac{d_C(s)}{d_{KL}(s)} = \lim_{s\to\infty} \frac{d_C(s)}{-\log p_n(s)} \leq \frac{1}{2}.$$

The proof is based on Jensen's inequality.

This fact suggests that for large signals, the error exponent for minimum P_e/mini-max detectors can be no bigger than one half of the error exponent for NP detectors. As described in the introduction, this should be the case since the NP detector need only minimize P_M while the minimum P_e detector must simultaneously minimize P_F and P_M and therefore can commit no more than one half of the computation capability of the likelihood ratio test to either of the two error probabilities.

If we limit our consideration to the class of generalized Gaussian density functions, i.e., densities of the form

$$p_n(x) = \frac{1}{2\Gamma(1+\frac{1}{r})A(r)} \exp\left\{-\left[\frac{|x|}{A(r)}\right]^r\right\},$$

then we may offer much stronger results.

Fact 5 *Let the noise be modeled as an arbitrary generalized Gaussian density function. Then*

$$\lim_{s\to\infty} \frac{d_C(s)}{-\log p_n(s)} = \begin{cases} \frac{1}{2^r} & \textit{if } r \geq 1 \\ \frac{1}{2} & \textit{if } r \leq 1 \end{cases}$$

The proof is shown in Appendix C.

Hence, for this case we may consider large signal approximations to $d_C(s)$ as:

$$d_C(s) = \begin{cases} -\frac{\log p_n(s)}{2^r} & \text{if } r \geq 1 \text{ and } s \gg 1 \\ -\frac{\log p_n(s)}{2} & \text{if } r \leq 1 \text{ and } s \gg 1 \end{cases}$$

By comparing the large signal results presented here to those derived in [1], we observe that the error exponent (performance) for all three optimality criteria is determined by the quantity $-\log p_n(s)$, which in the generalized Gaussian environment is equivalent to an ℓ_r metric of value $|s|^r$. Thus, for both large and small signal energies, the d_C and d_{KL} are identical up to a multiplicative constant. However, it should be pointed out that this constant diminishes exponentially fast as the decay rate of the density increases. This implies that for a fixed energy level, the relative performance of the minimum P_e/mini-max detector as compared with the NP detector falls off at a rate of $1/2^r$ in the *performance exponent* as decay rate increases. Nonetheless, if we combine this result with the small signal results, we see that in both regimes any signal that optimizes the Neyman-Pearson performance also optimizes the performance as measured by minimum P_e and mini-max criteria.

5. SIGNAL WAVEFORM DESIGN FOR THE NON-GAUSSIAN CHANNEL

For the family of generalized Gaussian noise models indexed by r, we consider the practical problem of allocating the available energy on the samples of the signal s^N so as to minimize the three performance criteria of interest. Without any available analytic solutions, we rely upon the asymptotic relations presented in the generalizations of Stein's Lemma and Sanov's Theorem. To accommodate this, we pose the signal design problem in the following way: Let s^N be a length N signal vector. Further, let $\tilde{s}_M$ be the $N \times M$ length signal formed by repeating s^N precisely M times. We seek to determine the signal waveform $\tilde{s}_M$ or equivalently s^N subject to an energy constraint such that the three performance measures of interest are minimized as $M \to \infty$. Note that we have moved from the original problem of an signal of finite energy to a power signal, so that the energy constraint E on s^N becomes the power contraint on s^M. For simplicity we use the same notation E for the power constraint.

Based upon this formulation, we know from our asymptotic analysis that the optimal signal subject to the NP criterion is determined by maximizing the quantity

$$\frac{1}{N}\sum_{i=1}^{N} d_{KL}(s_i) \;\; s.t. \; \sum_{i=1}^{N} s_i^2 \leq E.$$

Alternatively, we have shown that the optimal signal waveform subject to the minimum P_e and mini-max criteria is determined by maximizing the quantity

$$\frac{1}{N}\sum_{i=1}^{N} d_C(s_i) \;\; s.t. \; \sum_{i=1}^{N} s_i^2 \leq E.$$

When the signal energy is small, we know from our analysis in Section 4 that any waveform satisfying the energy constraint will be optimal for all three optimality criteria. However, if we consider the large signal regime, we show in this work that the optimal signal must be either fully impulsive, i.e., $s_1 = \sqrt{E}, s_i = 0$ for $i = 2, ..., N$ or constant for all i, that is $s_i = \sqrt{E/N}$ for all i.

The optimal signal s^N must maximize the Lagrangian

$$\mathbf{J} = \frac{1}{N}\sum_{i=1}^{N} d(s_i) + \rho(E - \sum_{i=1}^{N} s_i^2), \tag{20}$$

where the subscript on $d(s_i)$ has been left ambiguous to account for both d_C and d_{KL}. To find the maximizing $\{s_i\}$ for $\mathbf{J}$, we set its gradient w.r.t. s^N to zero.

$$0 = [\nabla \mathbf{J}]_i = \frac{1}{N}\dot{d}(s_i) - 2\rho s_i, i = 1, ...N \tag{21}$$

where $\dot{d}(s_i)$ is the derivative of $d(s_i)$ with respect to s_i.

Equation 21 can be shown to have only two solutions, namely $s_i = 0$ and $s_i = s^*, i = 1, ..., N$, for some $s^* \neq 0$, as demonstrated in Fig. 2 for the limited set of pdfs for which we can numerically evaluate the distances.

The maximizing signal set s^N therefore must have the form $s_i = \sqrt{E/L}$ for $i = 1, ..., L$ and zero for $i = L+1, ..., N$ for some L satisfying $1 \leq L \leq N$.

We pause for a moment to consider the practical implications of this model. First for the samples $L+1, ..., N$, the received signal is identical under either H_0 or H_1, and more importantly, this fact is known to the detector that will disregard samples in this interval. That raises the second question, namely that if only $L < N$ samples are ever used to represent the binary symbol, why not omit N from the problem? The answer is to vary N would amount to changing the symbol rate of the problem, and in turn the power E of the signal, thus making any performance comparison meaningless.

The solution to the minimization problem of Eq. 20 is most succinctly stated through the following proposition and its validity may be readily verified for the Generalized Gaussian family of pdfs, but it is rather complex for the general case.

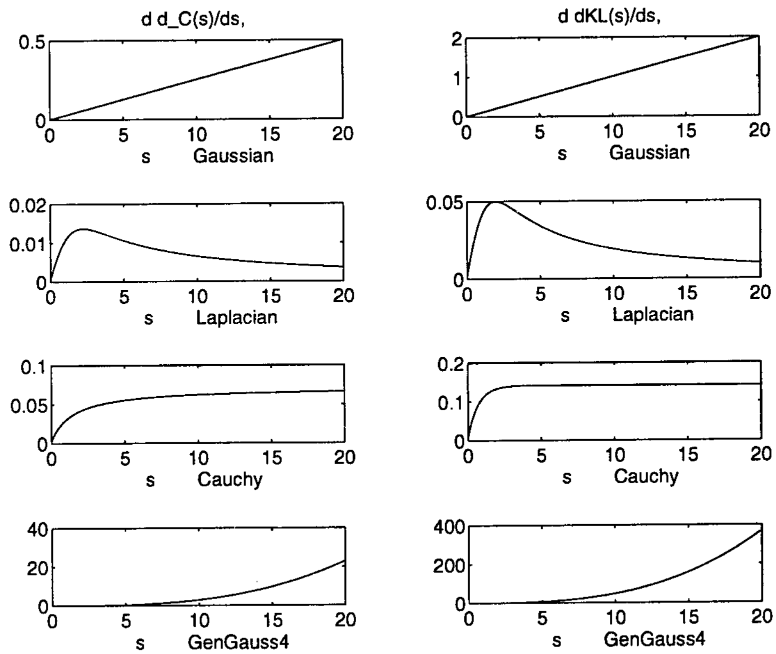

Fig. 2 — Derivative of Chernoff and KL distances w.r.t. displacement s vs s intersects any linear function through the origin at most one more point

Proposition 8 *Let $p_n(x)$ be an element of the generalized Gaussian family of density functions. Then $L = N$ for $r < 2$ and $L = 1$ for $r > 2$ under each of the three optimality criteria of interest.*

To demonstrate this, consider Figs. 3 and 4. In these two figures we have plotted the average Chernoff distance and the average Kullback-Leibler distance, respectively, as a function of L for various levels of power $E \in [0.1, 100]$ for the case of $N = 20$. When the noise is Laplacian ($r = 1$) or Cauchy, the maximum occurs when $L = N$, while for generalized Gaussian noise with $r = 4$, the optimal choice is $L = 1$. We see that for small signal energies, as for Gaussian noise, the average divergence and Chernoff distance are essentially invariant to the choice of L. This is to be the case since in this energy regime, only the energy and not the choice of L determines performance. In addition, one can observe that aside from the scale ($1/2^r$ for large E, $1/4$ for small E) the "shape" of these error exponents is essentially identical; this verifies the strong similarities established in the previous sections.

Since any waveform with a given energy is optimal for small amounts of energy, we then have arrived at the following optimal signal design procedure for the generalized Gaussian channel: If the tails of the noise density fall off faster than Gaussian tails, the optimal length N signal in both the large and small energy regimes is an impulse with amplitude $\sqrt{E}$. If, however, the tails fall off slower than Gaussian tails (e.g., Laplacian noise) then the optimal length N signal is constant with amplitude $\sqrt{E/N}$.

To demonstrate the generality of these results, we consider the Cauchy density as an alternate statistical model for the noise. As is well known, this model has polynomial tails and as such differs significantly from the generalized Gaussian model considered in this report. In addition, in this case the tail of the Cauchy density diminishes significantly slower than those of the Gaussian.

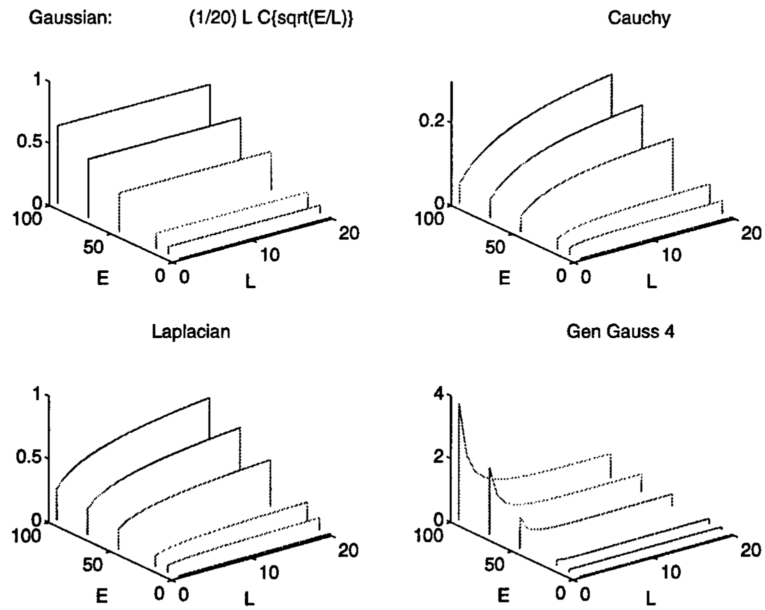

Fig. 3 — Average Chernoff Information vs the number of nonzero signal samples L for a variety of energy levels and $N = 20$

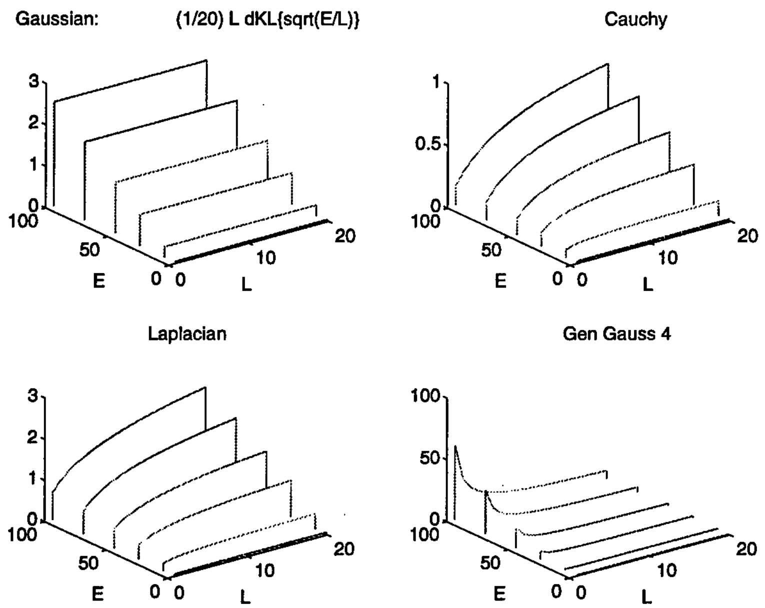

Fig. 4 — Kullback-Leibler distance vs the number of nonzero signal samples L for a variety of signal energies and $N = 20$

Thus, our theory would suggest that constant amplitude signal waveforms should be optimal. If one considers Figs. 3 and 4, then one can see that this is in fact the case for the error exponent with respect to both d_C and d_{KL}.

It should be stressed again that these length N signals are optimal only when the waveform s^N is repeated an infinite number of times. Of course, this is never the case in practice. Therefore, we are obliged to consider the performance of these signal waveforms in practical settings.

In Fig. 5, we have plotted the error exponent derived from simulations for the minimum P_e detector as a function of L for various levels of energy where one period ($M = 1$) of the signal is transmitted in an $N = 20$ sample waveform. The simulation was performed on a Connection Machine CM-5 for the present set of density functions. The simulation details are documented in Section 6.

From an asymptotic standpoint, this case of $M = 1$ is a worst case scenario. Nevertheless, as one can see, Fig. 5 is strikingly similar to both Figs. 3 and 4 even though the functional forms of d_C, d_{KL} and P_e are drastically different from one another. This seems to suggest that the similarity between d_C, d_{KL} and the minimum P_e is much stronger than that offered by the duality principle of Bradt and Karlin as discussed in Section 2 of this report.

6. SIMULATION

The objective is to estimate the detection error rate for our binary detection problem (Eq. 1) for equal priors, and when only one period of the optimal signal is used. For power E and number of nonzero samples L, a noise sequence is randomly generated according to density distribution

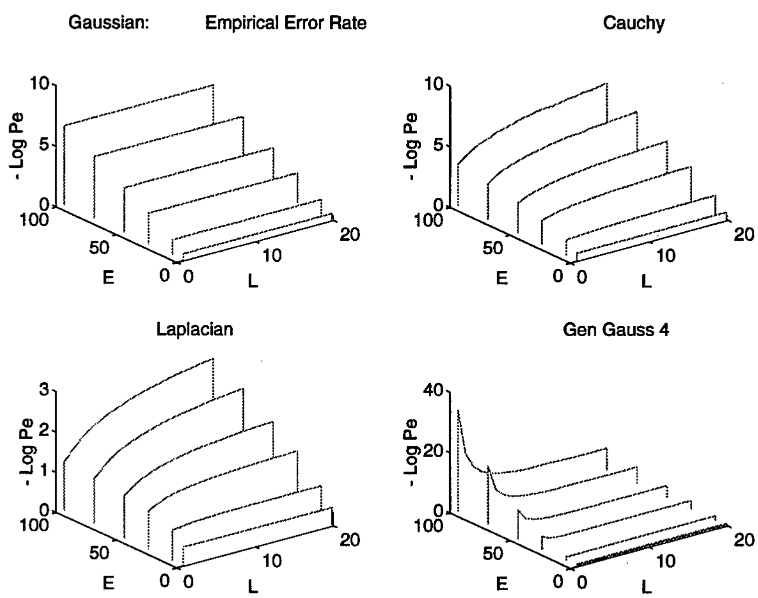

Fig. 5 — Empirical Probability of Error vs the number of nonzero samples of the signal waveform, L, for a variety of signal to noise ratios, $N = 20$ and $M = 1$

function p_n of one of four varieties, namely Gaussian, Laplacian, Cauchy, and Generalized Gaussian with decay rate 4. In the basic Monte Carlo simulation, the following is repeated until the desired accuracy is reached:

- Each of the L samples of the sequence are examined. This sequence is equivalent to a $L - sample$ waveform received at the output of a noisy channel under the null hypothesis H_0 (zero signal energy).
- Each received sample is transformed by the log-likelihood function $\log \frac{p_{n-s}}{p_n}$. Here p_n is a formula that describes a probability density function. p_{n-s} is another formula for the alternate hypothesis pdf, translated from p_n by the amount $s = \sqrt{\frac{E}{L}}$.
- All L transformed values are added and the sum is compared to a threshold 0. If positive, the error count is incremented by 1. (This is because the present sample is transmitted under H_0.)

The empirical probability of error is the error count divided by the number of experiments.

Importance Sampling

As the empirical probability (error rate) becomes very small, Monte Carlo simulation requires an excessive sample size—in the order of the inverse of the error rate. Instead of using the sample mean and $s = 0$ signal amplitude for H_0, we use the Importance Sampling technique. For a rigorous treatment of the subject, the reader is referred to the book by Bucklew [2], and the paper by Orsak[12].

The algorithm is as follows:

- L samples of randomly generated sequence n_i are added with a bias, $x_i = n_i + s_i$, $s_i = \frac{1}{2}\sqrt{\frac{E}{L}}, i = 1, ..., L$.
- Each (biased) received sample x_i is transformed by the usual $\log \frac{p_{n-s}}{p_n}$ where $s = \sqrt{\frac{E}{L}}$.
- For each sample, a counter-bias weight w_i is also calculated. $w_i = \log \frac{p_{n+\frac{s}{2}}}{p_n}(x_i)$.
- All L transformed values are added and the sum compared to a threshold 0. If positive, the error count is incremented by an amount $\exp(W)$. The exponent W is the sum of the w_i of each of the sequence sample.

Even though the sequence still represents the waveform received under H_0, the bias places the distribution precisely at the threshold of the log-likelihood-ratio test, such that an "error" occurs with a probability in the vicinity of $\frac{1}{2}$.

Random Number Generators

The simulation runs on a Connection Machine CM5 and uses the uniform random number generator native to this parallel architecture. The Gaussian generator uses a simplified version of the polar method [13]. The Laplacian and Cauchy generators use the transform method[13].

The Generalized Gaussian index 4 generator uses in addition to the uniform number generator a $Gamma(\frac{1}{4}, 2)$ generator. The latter uses Berman's algorithm[14].

Simulation Codes are documented in Appendix D.

7. CONCLUSION

We have considered the problem of waveform design in a non-Gaussian environment for communication applications. As is well known, optimal Neyman-Pearson, minimum error rate, and optimal mini-max solutions are analytically unavailable when the background noise deviates from Gaussian. In an effort to obtain "good" solutions, we have adapted Large Deviation based approaches to determine the asymptotically optimal signal waveform.

The principal result contained in this report was to show that for a given statistical model of the background noise, one signal waveform is optimal with respect to each of the three optimality criteria described above. This result holds for both large and small signal energies (amplitudes). Moreover, we have been able to obtain the precise waveform for a wide class of non-Gaussian statistics of much current interest. In addition to the above, we have demonstrated the following specific results:

- extended Chernoff's theorem for the non i.i.d. problem where the translation amount is known;
- demonstrated that the Chernoff Information and Kullback-Leibler distances determine the error exponent in the problems of interest;
- calculated large and small signal approximations for each of these statistical distance measures;
- calculated formulae for d_C and d_{KL} for a variety of non-Gaussian statistics;
- showed that for the generalized Gaussian class of pdf, maximizing d_{KL} also maximizes d_C, and vice versa;
- showed the maximizing signal of these two distances is an asymptotically most favorable signal, in the sense that it also asymptotically maximizes all distances in the Ali Silvey class;
- compared energy required to maintain same error exponent under NP and Bayes criteria;
- designed the optimal signal waveform under NP, mini-max and Bayes optimal criteria;
- confirmed by simulation the effectiveness of the asymptotic theory as applied to the finite-sample problem.

REFERENCES

1. D. H. Johnson and G. C. Orsak, "Relation of Signal Set to the Performance of Optimal Non-Gaussian Detectors," *IEEE Trans. Commun.*, vol. 41, pp. 1319–1328, sep 1993.

2. J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation.* Wiley-Interscience, 1990.

3. S. Kullback, *Information Theory and Statistics.* New York, NY: John Wiley and Sons, 1959.

4. S. M. Ali and D. Silvey, "A General Class of Coefficients of Divergence of One Distribution from Another," *J. Royal Stat Soc.*, vol. 28, pp. 131–142, 1966.

5. I. Csiszar, "Information-Type Measures of Difference of Probability Distributions and Indirect Observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 299–318, 1967.

6. G. C. Orsak and B.-P. Paris, "On the Relationship between Measures of Discrimination and the Performance of Suboptimal Detectors," *submitted to IEEE Trans. Inform. Theory*, November 1993.

7. T. L. Grettenberg, "Signal Selection in Communication and Radar Systems," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 265–275, Oct. 1963.

8. R. N. Bradt and S. Karlin, "On the Design and Comparison of Certain Dichotomous Experiments," *Ann. Math. Stat.*, vol. 27, pp. 390–409, 1956.

9. T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," *IEEE Trans. Commun. Tech.*, vol. COM-15, no. 1, pp. 52–60, Feb. 1967.

10. H. L. van Trees, *Detection, Estimation, and Modulation Theory, Part I.* New York: Wiley, 1968.

11. T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley-Interscience, 1991.

12. G. C. Orsak, "A Note on Estimating False Alarm Rates via Importance Sampling," *IEEE Trans. Commun.*, vol. 41, pp. 1275–1277, sep. 1993.

13. D. E. Knuth, *The Art of Computer Programming, Vol. 2.* Addison Wesley, 1969.

14. L. Devroye, *Non-Uniform Random Variate Generation.* Springer-Verlag, 1986.

Appendix A

DETAILED PROOF OF CHERNOFF THEOREM

In this appendix we present a detailed proof of Chernoff Theorem extended to the case of independent but not identically distributed samples.

Consider the following binary detection problem in which an N-vector is sent through additive iid noise channel in one case with zero signal, and in the other with some known signal $s^N = \{s_i, i = 1, \ldots, N\}$.

$$\begin{array}{rcl} H_0 : X_i & = & n_i \\ H_1 : X_i & = & n_i + s_i, \quad n_i \overset{iid}{\sim} p_n \end{array} \tag{A-1}$$

The optimal detector computes the LLRT based on the aggregate of N samples and compares it to some threshold γ:

$$\mathcal{L}(X^N) = \frac{1}{N}\sum_{i=1}^{N}\log(\frac{p(x_i - s_i)}{p(x_i)}) \underset{H_0}{\overset{H_1}{\gtrless}} \gamma \tag{A-2}$$

The decision regions are denoted Z_0^N and Z_1^N:

$$\begin{array}{rcl} Z_1^N & \stackrel{\text{def}}{=} & \{X^N | \mathcal{L}(X^N) \geq \gamma\} \\ Z_0^N & \stackrel{\text{def}}{=} & \overline{Z_1^N} \end{array} \tag{A-3}$$

The false alarm and miss probabilities are defined as $\alpha_N = \Pr\{say\ H_1 | H_0\}$ and $\beta_N = Pr\{say\ H_0 | H_1\}$, respectively. Define the normalized convex combination distribution p_λ^s as

$$p_\lambda^s(X^N) \stackrel{\text{def}}{=} \frac{p_0^{1-\lambda}(X^N) p_1^\lambda(X^N)}{\int p_0^{1-\lambda}(Z^N) p_1^\lambda(Z^N) dZ^N} \tag{A-4}$$

For the additive iid noise of Eq.(A-1), under either hypotheses, the difference between the received signal and the transmitted signal $(X^N - s^N) = \{x_i - s_i | i = 1, \ldots, N\}$. Each term $x_i - s_i$ is $\overset{iid}{\sim} p_n$. Thus

$$\begin{array}{rcl} p_\lambda^s(X^N) & = & \prod_{i=1}^N p_\lambda^i(x_i) = \prod_{i=1}^N \frac{p_0^{1-\lambda}(x_i) p_1^\lambda(x_i)}{J_\lambda(s_i)} = \prod_{i=1}^N \frac{p_n^{1-\lambda}(x_i) p_n^\lambda(x_i - s_i)}{J_\lambda(s_i)} \\ where & & \\ J_\lambda(s_i) & \stackrel{\text{def}}{=} & \int p_n^{1-\lambda}(z_i) p_n^\lambda(z_i - s_i) dz_i \end{array} \tag{A-5}$$

The following statistics may be defined in terms of the divergence between the distributions of the hypotheseses p_0 and p_1 and the normalized-convex distribution p_λ^i:

$$\begin{array}{rcl} G_{0,\lambda}(X^N) & \stackrel{\text{def}}{=} & \frac{1}{N}\sum_{i=1}^N \log \frac{p_n(x_i)}{p_\lambda^i(x_i)} \\ G_{1,\lambda}(X^N) & \stackrel{\text{def}}{=} & \frac{1}{N}\sum_{i=1}^N \log \frac{p_n(x_i - s_i)}{p_\lambda^i(x_i)} \end{array} \tag{A-6}$$

p_0, p_1 may be expressed in terms of G and p^i_λ as follows:

$$\begin{aligned} p(X^N) &= \exp(NG_{0,\lambda}(X^N))p^s_\lambda(X^N) \\ &= \exp(NG_{0,\lambda}(X^N))\textstyle\prod_i p^i_\lambda(x_i) \\ p(X^N - s^N) &= \exp(NG_{1,\lambda}(X^N))p^s_\lambda(X^N) \\ &= \exp(NG_{1,\lambda}(X^N))\textstyle\prod_i p^i_\lambda(x_i) \end{aligned} \tag{A-7}$$

Observe that under $p^s_\lambda(X^N)$, $\mathcal{E}\{\log \frac{p(x_i)}{p^i_\lambda(x_i)}\}$ is identified as the negative of the Kullback-Leibler distance d_{KL}, so that:

$$\begin{aligned} \mathcal{E}\{G_{0,\lambda}(X^N)\} &= \tfrac{1}{N}\textstyle\sum_{i=1}^N \mathcal{E}\{\log \frac{p_n(x_i)}{p^i_\lambda(x_i)}\} &= \tfrac{1}{N}\textstyle\sum_{i=1}^N -d_{KL}(p^i_\lambda, p_n) \\ \mathcal{E}\{G_{1,\lambda}(X^N)\} &= \tfrac{1}{N}\textstyle\sum_{i=1}^N \mathcal{E}\{\log \frac{p_n(x_i - s_i)}{p^i_\lambda(x_i)}\} &= \tfrac{1}{N}\textstyle\sum_{i=1}^N -d_{KL}(p^i_\lambda, p_{n-s_i}) \end{aligned} \tag{A-8}$$

By the Strong Law of Large Numbers, each of the statistics, Eq.(A-6) approaches its average defined in terms of the Kullback-Leibler distances. Therefore, the sample average of the statistics must approach the sample average of the KL distances.

$$\begin{aligned} \lim_{N\to\infty} G_{0,\lambda} &= \lim_{N\to\infty} \tfrac{1}{N}\textstyle\sum_{i=1}^N -d_{KL}(p^i_\lambda, p_n) \quad a.s. \\ \lim_{N\to\infty} G_{1,\lambda} &= \lim_{N\to\infty} \tfrac{1}{N}\textstyle\sum_{i=1}^N -d_{KL}(p^i_\lambda, p_{n-s_i}) \quad a.s. \end{aligned} \tag{A-9}$$

Define the ϵ -neighborhoods A^N of p_0 and B^N of p_1 in terms of the statistics G and their averages:

$$\begin{aligned} A^N &\stackrel{\text{def}}{=} \{X^N : |G_{0,\lambda}(X^N) + \tfrac{1}{N}\textstyle\sum_{i=1}^N d_{KL}(p^i_\lambda, p_n)| < \epsilon\} \\ B^N &\stackrel{\text{def}}{=} \{X^N : |G_{1,\lambda}(X^N) + \tfrac{1}{N}\textstyle\sum_{i=1}^N d_{KL}(p^i_\lambda, p_{n-s_i})| < \epsilon\} \end{aligned} \tag{A-10}$$

As $N \to \infty$, by (A-9) we must have that $\forall\epsilon, \delta > 0, \exists N_0$ such that for $\forall N > N_0$,

$$\begin{aligned} \Pr(A^N | X^N \sim p^s_\lambda(X^N)) &\geq 1 - \delta \\ \Pr(B^N | X^N \sim p^s_\lambda(X^N)) &\geq 1 - \delta \end{aligned} \tag{A-11}$$

We can now relate the false alarm and miss rates to the ϵ-neighborhoods of p_0 and p_1 as functions of ϵ and the Kullback-Leiber distances under probability measure p^s_λ.

$$\begin{aligned} \alpha_N &= \textstyle\int_{Z_1^N} p^s_\lambda(X^N) e^{NG_{0,\lambda}(X^N)} dX^N \\ &\geq \textstyle\int_{Z_1^N \cap A^N} p^s_\lambda(X^N) e^{NG_{0,\lambda}(X^N)} dX^N \\ &\geq \textstyle\int_{Z_1^N \cap A^N} p^s_\lambda(X^N) e^{-N\epsilon - \sum_{i=1}^N d_{KL}(p^i_\lambda, p_n)} dX^N \end{aligned} \tag{A-12}$$

As $N \to \infty$, we already showed that $G_{0,\lambda}(X^N)$ approaches $-\frac{1}{N}\sum_{i=1}^N d_{KL}(p^i_\lambda, p_n)$, i.e:

$$\epsilon < |G_{0,\lambda}(X^N) + \frac{1}{N}\sum_{i=1}^N d_{KL}(p^i_\lambda, p_n)| < \epsilon \tag{A-13}$$

Continuing developing inequality (A-12), we have

$$\begin{aligned} \alpha_N &\geq e^{-N\epsilon - \sum_{i=1}^{N} d_{KL}(p_\lambda^i, p_n)} p_\lambda^i (Z_1^N \cap A^N) \\ &\geq e^{-N\epsilon - \sum_{i=1}^{N} d_{KL}(p_\lambda^i, p_n)} [p_\lambda^s (Z_1^N) + p_\lambda^i (A^N) - 1], \end{aligned} \quad \text{(A-14)}$$

where the last inequality is from $P(Z \cap A) = 1 - P(\overline{Z} \cup \overline{A}) \geq 1 - P(\overline{Z}) - P(\overline{A}) = 1 - (1 - P(Z)) - (1 - P(A)) = P(Z) + P(A) - 1$, and the assumption that $P(Z_1^N(X^N)) \geq \frac{1}{2}$ with X^n distributed under P_λ^s .

The first Neyman Pearson lower bound is then

$$\begin{aligned} \alpha_N &\geq e^{-N\epsilon - \sum_{i=1}^{N} d_{KL}(p_\lambda^i, p_n)} [\tfrac{1}{2} + 1 - \delta - 1] \\ &\Rightarrow \\ \tfrac{1}{N} \log \alpha_N &\geq -\epsilon - \tfrac{1}{N} \textstyle\sum_{i=1}^{N} d_{KL}(p_\lambda^i, p_n) + \tfrac{1}{N} \log(\tfrac{1}{2} - \delta) \\ &\Rightarrow \\ \lim_{N\to\infty} \tfrac{1}{N} \log \alpha_N &\geq \lim_{N\to\infty} -\tfrac{1}{N} \textstyle\sum_{i=1}^{N} d_{KL}(p_\lambda^i, p_n) \end{aligned} \quad \text{(A-15)}$$

On the other hand, if $P(Z_1^N(X^N)) < \frac{1}{2}$ under this distribution then we have $P(Z_0^N(X^N)) \geq \frac{1}{2}$ and a similar result holds for the miss probability as follows.

$$\lim_{N\to\infty} \tfrac{1}{N} \log \beta_N \geq \lim_{N\to\infty} -\tfrac{1}{N} \textstyle\sum_{i=1}^{N} d_{KL}(p_\lambda^i, p_{n-s_i}) \quad \text{(A-16)}$$

Combining Eqs. (A-15) and (A-16), we have the following inequality that can be minimized with respect to λ to achieve a minimax solution for P_e.

$$\lim_{N\to\infty} \frac{1}{N} \log(\max(\alpha_N, \beta_N)) \geq \lim_{N\to\infty} -\frac{1}{N} \min(\sum_{i=1}^{N} d_{KL}(p_\lambda^*, p_n), \sum_{i=1}^{N} d_{KL}(p_\lambda^*, p_{n-s_i})) \quad \text{(A-17)}$$

Now, the minimum probability of error the best achievable exponent is $\sum_{i=1}^{N} d_{KL}(p_{\lambda_o}, p_n)$, where λ_o satisfies :

$$\lambda_o = arg_{(0\leq\lambda\leq 1)} \left\{ \sum_{i=1}^{N} d_{KL}(p_\lambda^i, p_n) = \sum_{i=1}^{N} d_{KL}(p_\lambda^i, p_{n-s_i}) \right\}$$

Hence,

$$\lim_{N\to\infty} minP_e \geq \lim_{N\to\infty} e^{-\sum_{i=1}^{N} d_{KL}(p_{\lambda_o}^i, p_n)} \quad \text{(A-18)}$$

Before proceeding any further we must make sure that the assumptions made in the derivation of equations (A-15) and (A-16) still hold simultaneously.

Lemma 9

$$arg_\lambda \left\{ d_{KL}(p_\lambda^i(X^N), p(X^N)) = d_{KL}(p_\lambda^i(X^N), p(X^N - s^N)) \ such\ that \lim_{N\to\infty} \mathcal{E}_\lambda^i [I_{Z_1^N}(X^N)] = \frac{1}{2} \right\}$$

$$= arg_\lambda \left\{ d_{KL}(p_\lambda^i(X^N), p(X^N)) = d_{KL}(p_\lambda^i(X^N), p(X^N - s^N)) \right\}$$

Proof: For minimum P_e define:

$$Z_o^N = \left\{X^N | \pi_o p_o(X^N) \geq \pi_1 p_1(X^N)\right\}$$

and

$$Z_1^N = \left\{X^N | \pi_1 p_1(X^N) \geq \pi_o p_o(X^N)\right\}.$$

The likelihood ratio is,

$$\mathcal{L}(X^N) = \frac{1}{N}\sum_{i=1}^{N} \log\left\{\frac{\pi_1 p_1(x_i)}{\pi_o p_o(x_i)}\right\} \underset{H_0}{\overset{H_1}{\gtrless}} 0.$$

Therefore,

$$\lim_{N\to\infty} \mathcal{L}(X^N) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \log\left\{\frac{p_1(x_i)}{p_o(x_i)}\right\} \underset{H_0}{\overset{H_1}{\gtrless}} 0.$$

Therefore for large N,

$$\begin{aligned}\mathcal{E}_*[I_{Z_o^N}(X^N)] &= P[X^N \in Z_o^N | X^N \sim p_\lambda^i] \\ &= P[X^N | \frac{1}{N}\sum_{i=1}^{N} \log\left\{\frac{p_1(x_i)}{p_o(x_i)}\right\} \leq 0, X^N \sim p_\lambda^s].\end{aligned}$$

and

$$\mathcal{E}_*[I_{Z_1^N}(X^N)] = P[X^N | \frac{1}{N}\sum_{i=1}^{N} \log\left\{\frac{p_1(x_i)}{p_o(x_i)}\right\} \geq 0, X^N \sim p_\lambda^s].$$

Now by the strong law of large numbers,

$$\lim_{N\to\infty} \tfrac{1}{N}\textstyle\sum_{i=1}^N \log\left\{\frac{p_1(x_i)}{p_o(x_i)}\right\} = \lim_{N\to\infty} \tfrac{1}{N}\textstyle\sum_{i=1}^N \int \log\left\{\frac{p_1(x_i)}{p_o(x_i)}\right\} p_{\lambda_o}^i(x_i)dx_i \quad a.s. \tag{A-19}$$

Also for $\lambda = \lambda_o$:

$$\begin{aligned}\sum_{i=1}^{N}\int \log\left\{\frac{p_1(x_i)}{p_o(x_i)}\right\} p_{\lambda_o}^i(x_i)dx_i &= \sum_{i=1}^{N}\int \log\left\{\frac{p_1(x_i)}{p_{\lambda_o}^i(x_i)}\frac{p_{\lambda_o}^i(x_i)}{p_o(x_i)}\right\} p_{\lambda_o}^i(x_i)dx_i \\ &= -\sum_{i=1}^{N} d_{KL}(p_{\lambda_o}^i, p_o) + \sum_{i=1}^{N} d_{KL}(p_{\lambda_o}^i, p_1) \\ &= 0\end{aligned}$$

Therefore,

$$\lim_{N\to\infty} \tfrac{1}{N}\textstyle\sum_{i=1}^N \log\left\{\frac{p_1(x_i)}{p_o(x_i)}\right\} = 0 \quad a.s. \tag{A-20}$$

Hence,

$$\mathcal{E}_*[I_{Z_o^N}(X^N)] = \mathcal{E}_*[I_{Z_1^N}(X^N)] = \frac{1}{2}.$$

□

The lemma shows that we still satisfy the assumptions in arriving at Eqs. (A-15) and (A-16). Hence from Eq. (A-18) ,

$$\lim_{N\to\infty} \frac{1}{N} \log\min P_e^N \geq \lim_{N\to\infty} -\frac{1}{N}\sum_{i=1}^{N} d_{KL}(p_{\lambda_o}^*, p_n). \tag{A-21}$$

Chernoff Information

Lemma 10

$$-\frac{1}{N}\sum_{i=1}^{N} d_{KL}(p^*_{\lambda_o}, p_n) = \min_{\lambda} \frac{1}{N}\log J_\lambda(s^N).$$

Proof :

$$\log J_\lambda(s^N) = \log \int p^{1-\lambda}(X^N)p^\lambda(X^N - s^N)dX^N.$$

Since the components of X^N are independently distributed,

$$\begin{aligned}
\log J_\lambda(s^N) &= \sum_{i=1}^{N} \log \int p^{1-\lambda}(x_i)p^\lambda(x_i - s_i)dx_i \\
&= \sum_{i=1}^{N} \log \int e^{(1-\lambda)\log p(x_i)+\lambda \log p(x_i - s_i)}dx_i
\end{aligned}$$

Maximizing J_λ wrt λ:

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\log J_\lambda(s^N) &= \sum_{i=1}^{N} \frac{1}{J_\lambda(s_i)} \int p^{1-\lambda}(x_i)p^\lambda(x_i - s_i)[\log p(x_i - s_i) - \log p(x_i)]dx_i. \\
&= \sum_{i=1}^{N} \frac{1}{J_\lambda(s_i)} \int p^{1-u}(x_i))p^\lambda((x_i - s_i)\log \frac{p(x_i - s_i)}{p(x_i)}dx_i \\
&= \sum_{i=1}^{N} \int p_\lambda(x_i)\log \frac{p(x_i - s_i)}{p(x_i)}dx_i \\
&= \sum_{i=1}^{N}[-d_{KL}(p^i_\lambda, p_{n-s_i}) + d_{KL}(p^i_\lambda, p_n)].
\end{aligned}$$

which must be equal to 0 for $\min \log J_\lambda(X^N)$. Hence the minimum λ_o is given by,

$$\sum_{i=1}^{N} d_{KL}(p^i_{\lambda_o}, p_n) = \sum_{i=1}^{N} d_{KL}(p^i_{\lambda_o}, p_{n-s_i}).$$

Now to get $\min_\lambda \frac{1}{N}\log J_\lambda(X^N)$, we must proceed as follows. Consider:

$$\log p^s_{\lambda_o}(X^N) = (1-\lambda)\log p(X^N) + \lambda \log p(X^N - s^N) - \log J_{\lambda_o}(s^N).$$

Therefore,

$$\log J_{\lambda_o}(s^N) = (1-\lambda)\log \frac{p(X^N)}{p^s_{\lambda_o}(X^N)} + \lambda \log \frac{p(X^N - s^N)}{p^s_{\lambda_o}(X^N)}.$$

Taking the expectation value w.r.t.$p^s_{\lambda_o}(X^N)$:

$$\begin{aligned}
\log J_{\lambda_o}(s^N) &= -(1-\lambda)d_{KL}(p^s_{\lambda_o}(X^N), p(X^N)) - \lambda d_{KL}(p^s_{\lambda_o}(X^N), p(X^N - s^N)) \\
&= -\sum_{i=1}^{N} d_{KL}(p^i_{\lambda_o}(x_i), p(x_i)) \qquad \text{(A-22)}
\end{aligned}$$

by making use of the fact that λ_o results in the minimum value for $J_\lambda(s^N)$ .Hence we have,

$$-\log J_{\lambda_o}(s^N) = \sum_{i=1}^{N} d_{KL}(p^i_\lambda, p_n) = \sum_{i=1}^{N} d_{KL}(p^i_\lambda, p_{n-s_i}).$$

Hence the result.□

From Eq. (A-21) we have,

$$\lim_{N\to\infty} \frac{1}{N} \log \min P_e^N \geq \lim_{N\to\infty} \frac{1}{N} \log J_{\lambda_o}(s^N)$$

which is the desired lower bound for $\min P_e$.

Upper bound, Chernoff

We will now show an upper bound for $minP_e$ which approaches the value of the lower bound, thus establishing the asymptotic equivalence of the Chernoff Information d_C to the best annihilating rate for P_e.

$$\begin{array}{rcl} \min P_e^N & = & \int \min\left(\pi_0 p_n(X^N), \pi_1 p_n(X^N - s^N)\right) dX^N \\ & \leq & \int (\pi_0 p_n(X^N))^{1-\lambda} (\pi_1 p_n(X^N - s^N))^{\lambda} dX^N \\ & = & \pi_0^{1-\lambda}\pi_1^{\lambda} \prod_{i=1}^{N} \int p_n^{1-\lambda}(x_i) p_n^{\lambda}(x_i - s_i) dx_i \\ & = & \pi_0^{1-\lambda}\pi_1^{\lambda} \exp\left(\sum \log \int p_n^{1-\lambda}(x_i) p_n^{\lambda}(x_i - s_i) dx_i\right) \\ \frac{1}{N}\log \min P_e^N & \leq & \frac{1}{N}\log(\pi_0^{1-\lambda}\pi_1^{\lambda}) + \frac{1}{N}\sum \log \int p_n^{1-\lambda}(x_i) p_n^{\lambda}(x_i - s_i) dx_i. \end{array} \quad \text{(A-23)}$$

Taking the limit:

$$\begin{array}{rcl} \lim_{N\to\infty} \frac{1}{N} \log \min P_e^N & \leq & \lim_{N\to\infty} \min_{\lambda} \frac{1}{N} \sum_{i=1}^{N} \log \int p_n^{1-\lambda}(x_i) p_n^{\lambda}(x_i - s_i) dx_i \\ & = & \lim_{N\to\infty} \min_{\lambda} \frac{1}{N} \log J_\lambda(s^N) \\ & = & \lim_{N\to\infty} \frac{1}{N} \log J_{\lambda_o}(s^N). \end{array} \quad \text{(A-24)}$$

Take the upper and lower bounds together, we have the limit.

Appendix B

DIRECT CALCULATION OF CHERNOFF INFORMATION

In this appendix we derive the Chernoff Information for the Gaussian, Cauchy, Laplacian and Generalized Gaussian of index 4 densities. The results are summarized in Table 1.

Gaussian

Probability density function: $\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}$

$$\begin{aligned} C_G(s,\sigma) &= -\log \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(x^2 + (x-s)^2)dx \\ &= -\log \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(x - \frac{s}{2})^2 + (\frac{s}{2})^2)dx \\ &= \frac{s^2}{8\sigma^2}. \end{aligned}$$

Cauchy

Probability density function: $\frac{1}{\pi\sigma}\frac{1}{1+(\frac{x}{\sigma})^2}$

$$\begin{aligned} C_C(a,\sigma) &= -\log \frac{1}{\pi\sigma} \int_{-\infty}^{\infty} \frac{1}{\sqrt{(1+\frac{x^2}{\sigma^2})(1+\frac{(x-s)^2}{\sigma^2})}} dx \\ &= -\log \frac{2}{\pi\sigma} \int_0^{\infty} \frac{1}{\sqrt{\frac{y^4}{\sigma^4}+(\frac{2}{\sigma^2}-\frac{s^2}{2\sigma^2})y^2+(\frac{s^2}{2\sigma^2}+\frac{s^4}{16\sigma^4}+1)}} dy \\ &= -\log \frac{2}{\pi\sigma} \int_0^{\infty} \frac{1}{\sqrt{y^4+(2\sigma^2-\frac{s^2}{2})y^2+(\frac{s^2}{2}\sigma^2+\frac{s^4}{16}+\frac{1}{\sigma^4})}} dy \end{aligned} \quad \text{(B-1)}$$

From Ref. B1, Formula 266.00, and definitions and *modulus* properties of $\mathcal{F}$ in pages xv, 12:

$$\begin{aligned} D(a,\sigma) &= \int_0^{\infty} \frac{1}{\sqrt{y^4+\beta y^2+\alpha^4}} dy = \frac{1}{2\alpha}\mathcal{F}(\pi,k) \\ &\textit{where} \\ k^2 &= \frac{2\alpha^2-\beta}{4\alpha^2} \\ \mathcal{F}(\pi,k) &= 2\mathcal{F}(\frac{\pi}{2},k) \end{aligned} \quad \text{(B-2)}$$

$\mathcal{F}(\frac{\pi}{2},k)$ is the complete elliptic integral of the first kind that we denote in Table 1 of the report as E_k.

The application of Formula 266.00 requires that

$$2\sigma^2 - \frac{s^2}{2} < \sqrt{\frac{s^2}{2}\sigma^2 + \frac{s^4}{16} + \frac{1}{\sigma^4}}$$

This can be shown to be true for all positive values of a and σ.

Apply Eqs. (B-2) to (B-1), we have the substitutions

$$\begin{aligned} \beta &= 2\sigma^2 - \tfrac{s^2}{2} \\ \alpha^4 &= \tfrac{s^2}{2}\sigma^2 + \tfrac{s^4}{16} + \tfrac{1}{\sigma^4} \\ \Rightarrow k &= \sqrt{\tfrac{1}{2} - \frac{\frac{\sigma^2}{2} - \frac{s^2}{8}}{\sqrt{\frac{s^2}{2}\sigma^2 + \frac{s^4}{16} + \frac{1}{\sigma^4}}}} \end{aligned} \qquad \text{(B-3)}$$

so that

$$\begin{aligned} C(s,\sigma) &= -\log \tfrac{2\sigma}{\pi} D(s,\sigma) \\ &= -\log \tfrac{2\sigma}{\pi} \frac{1}{(\frac{s^2}{2}\sigma^2 + \frac{s^4}{16} + \frac{1}{\sigma^4})^{\frac{1}{4}}} \mathcal{F}(\tfrac{\pi}{2}, k) \end{aligned} \qquad \text{(B-4)}$$

where k is defined in Eq. (B-3).

Laplacian

Probability distribution function: $\frac{1}{\sigma\sqrt{2}} e^{\frac{-|x|\sqrt{2}}{\sigma}}$

$$\begin{aligned} C_L(s,\sigma) &= -\log \int \tfrac{1}{\sqrt{2}\sigma} \exp - \tfrac{|x|}{2\sigma/\sqrt{2}} - \tfrac{|x-s|}{2\sigma/\sqrt{2}} dx \\ &= -\log[\int_{-\infty}^{0} \tfrac{1}{\sqrt{2}\sigma} e^{\frac{1}{2\sigma/\sqrt{2}}(x+x-s)} dx + \int_0^s \tfrac{1}{\sqrt{2}\sigma} e^{\frac{1}{2\sigma/\sqrt{2}}(-x+x-s)} dx + \int_s^{\infty} \tfrac{1}{\sqrt{2}\sigma} e^{\frac{1}{2\sigma/\sqrt{2}}(-x-x+s)} dx] \\ &= -\log[e^{\frac{s}{2\sigma/\sqrt{2}}} + \tfrac{s}{\sqrt{2}\sigma} e^{-\frac{s}{2\sigma/\sqrt{2}}}] \\ &= \tfrac{s}{2\sigma/\sqrt{2}} - \log(1 + \tfrac{s}{\sqrt{2}\sigma}). \end{aligned}$$

Generalized Gaussian index 4

Probability distribution function: $\frac{1}{2\Gamma(1+\frac{1}{r})A(r)} \exp -[\frac{|x|}{A(r)}]^4$, where $A(r) = [\sigma^2 \frac{\Gamma(1/r)}{\Gamma(3/r)}]^{\frac{1}{2}}$.

$$\begin{aligned} C_{GG4}(s,\sigma) &= -\log \int_{-\infty}^{+\infty} p_n^{\frac{1}{2}}(x) p_{n-s}^{\frac{1}{2}} x\, dx \\ &= -\log \tfrac{1}{2\Gamma(\frac{5}{4})A_4} \int_{-\infty}^{+\infty} \exp - \tfrac{1}{2A_4^4}(x^4 + (x-s)^4) dx \\ &\overset{y=x-\frac{s}{2}}{=} -\log \tfrac{1}{2\Gamma(\frac{5}{4})A_4} 2 \int_0^{+\infty} \exp - \tfrac{1}{2A_4^4}(2y^4 + 3s^2y^2 + \tfrac{s^4}{8}) dy \\ &= \log \Gamma(\tfrac{5}{4}A_4) + \tfrac{s^4}{16A_4^4} - \log \int_0^{\infty} \exp \tfrac{1}{A_4^4}(y^4 + \tfrac{3}{2}s^2y^2) dy \end{aligned} \qquad \text{(B-5)}$$

The argument of the logarithm of the third term in Eq.(B-5) is evaluated using an identity in Ref. B2. The identity is

$$\int_0^\infty \exp -\beta^2 x^4 - 2\gamma^2 x^2 dx = 2^{-\frac{3}{2}} \frac{\gamma}{\beta} e^{\frac{\gamma^4}{2\beta^2}} K_{\frac{1}{4}}\left(\frac{\gamma^4}{2\beta^2}\right) \tag{B-6}$$

where $K_{\frac{1}{4}}$ is the modified Bessel function with fractional order $\frac{1}{4}$. We apply identity Eq. (B-6) with $\beta = \frac{1}{A_4^2}$ and $\gamma = \sqrt{\frac{3}{4}\frac{a}{A_4^2}}$.

The integration is then

$$\int_0^\infty \exp \frac{1}{A_4^4}\left(y^4 + \frac{3}{2}s^2 y^2\right)dy = \sqrt{\frac{3}{32}} s e^{\frac{9}{32}\frac{s^4}{A_4^4}} K_{\frac{1}{4}}\left(\frac{9}{32}\frac{s^4}{A_4^4}\right). \tag{B-7}$$

Substituting

$$C(a,\sigma) = -\frac{7}{32}\frac{s^4}{A^4} - \log \frac{\sqrt{\frac{3}{32}} s}{\Gamma(\frac{5}{4})A} - \log K_{\frac{1}{4}}\left(\frac{9}{32}\frac{s^4}{A^4}\right) \tag{B-8}$$

where $A = A_4 = \sigma\sqrt{\frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{3}{4})}}$.

REFERENCES

B1. P.F.Byrd and M.D.Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists.* Springer-Verlag, 1971.

B2. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products.* Academic Press, 1980.

Appendix C

LARGE SIGNAl APPROXIMATION FOR CHERNOFF INFORMATION FOR AN EXPONENTIAL FAMILY OF PROBABILITY DENSITIES

In this appendix we detail proofs for Fact 5 in the main text that provide upper bounds to the Chernoff Information for large signal. It is repeated here:

Consider densities of the form $p_n(x) = \frac{1}{2\Gamma\{1+\frac{1}{r}\}\mathcal{A}(r)} \exp\left\{-\left\{\frac{|x|}{\mathcal{A}(r)}\right\}^r\right\}$, where $\mathcal{A}(r) = \left\{\frac{\sigma^2\Gamma(1/r)}{\Gamma(3/r)}\right\}^{1/2}$
Then if $\mathcal{Z}$ denotes the following limit:

$$\mathcal{Z} = \lim_{s\to\infty} \frac{d_C(s)}{-\log p_n(s)} = \lim_{s\to\infty} \frac{-\log\{\int p_n^{1/2}(x)p_n^{1/2}(x-s)dx\}}{-\log p_n(s)}$$

it must be that :

$$\mathcal{Z} = \frac{1}{2^r},\ r \geq 1$$
$$\mathcal{Z} = \frac{1}{2},\ r \leq 1.$$

Proof: Probability densities of this form are known as Generalized Gaussian density with decay rate r and variance σ^2. We begin with the $r \geq 1$ case.

1 **For** $r \geq 1$ **:**

Dividing the limits of the integral involved in C(s) as follows :

$$\begin{aligned}\int_{-\infty}^{\infty} p_n^{1/2}(x)p_n^{1/2}(x-s)dx &= \int_{-\infty}^{0} p_n^{1/2}(x)p_n^{1/2}(x-s)dx + \int_0^s p_n^{1/2}(x)p_n^{1/2}(x-s)dx \\ &+ \int_s^{\infty} p_n^{1/2}(x)p_n^{1/2}(x-s)dx \qquad \text{(C-1)}\end{aligned}$$

. Now denoting $b = \mathcal{A}(r)^r$ and $\mathcal{K} = \frac{1}{2\Gamma\{1+\frac{1}{r}\}\mathcal{A}(r)}$ the second term on the RHS of the above equation simplifies to ,

$$\begin{aligned}\int_0^s p_n^{1/2}(x)p_n^{1/2}(x-s)dx &= \mathcal{K}\int_0^s \exp\frac{-\{|x|^r + |x-s|^r\}}{2b}dx \\ &\leq \mathcal{K}\int_0^s \exp\frac{-\{|x| + |x-s|\}^r}{2^r b}dx\end{aligned}$$

where the inequality was introduced by using the relation:(Ref. C1)

$$|x|^r + |x-s|^r \geq \frac{\{|x| + |x-s|\}^r}{2^{r-1}} \qquad \text{(C-2)}$$

which holds for all $r \geq 1$ and for all x and s. Thus, by recognizing that $|x| + |x - s| = s$ over the region of integration, we have that

$$\begin{aligned} \int_0^s p_n^{1/2}(x) p_n^{1/2}(x-s)dx &\leq \mathcal{K} \int_0^s \exp \frac{-|s|^r}{2^r b} dx \\ &= \mathcal{K} s \exp \frac{-|s|^r}{2^r b}. \end{aligned} \tag{C-3}$$

Proceeding, we turn our attention to the first term in Eq. C-1 to establish an upper bound for the integral as follows :

$$\int_{-\infty}^0 p_n^{1/2}(x) p_n^{1/2}(x-s)dx = \mathcal{K} \int_{-\infty}^0 \exp \frac{-|x|^r}{2b} \exp \frac{-|x-s|^r}{2b} dx.$$

Now the term $\left\{\exp \frac{-|x-s|^r}{2b}\right\}$ has its maximum value at $x = 0$ for $x \in [-\infty, 0]$, hence we may upper bound the integral by,

$$\begin{aligned} \int_{-\infty}^0 p_n^{1/2}(x) p_n^{1/2}(x-s)dx &\leq \mathcal{K} \int_{-\infty}^0 \exp \frac{-|x|^r}{2b} \exp \frac{-|s|^r}{2b} dx \\ &= \mathcal{K} \exp \frac{-|s|^r}{2b} \int_{-\infty}^0 \exp \frac{-|x|^r}{2b} dx \\ &= \mathcal{K} \exp \left\{ \frac{-|s|^r}{2b} \right\} \frac{2^{1/r}}{2\mathcal{K}} \end{aligned} \tag{C-4}$$

where the latter integral was evaluated by using the fact that $\mathcal{K} \exp \frac{-|x|^r}{b}$ is a valid density function. Similarly we may upper bound the third term in Eq. C-1 as follows:

$$\int_s^\infty p_n^{1/2}(x) p_n^{1/2}(x-s)dx = \mathcal{K} \int_s^\infty \exp \frac{-|x|^r}{2b} \exp \frac{-|x-s|^r}{2b} dx$$

Here the term $\left\{\exp \frac{-|x|^r}{2b}\right\}$ has its maximum value at $x = s$ for $x \in [s, \infty]$. Hence we can upper bound the integral by,

$$\begin{aligned} \int_s^\infty p_n^{1/2}(x) p_n^{1/2}(x-s)dx &\leq \mathcal{K} \exp \frac{-|s|^r}{2b} \int_s^\infty \exp \frac{-|x-s|^r}{2b} dx \\ &= \mathcal{K} \exp \left\{ \frac{-|s|^r}{2b} \right\} \frac{2^{1/r}}{2\mathcal{K}} \end{aligned} \tag{C-5}$$

where the latter integral was evaluated as before. Therefore by combining the upper bounds on the terms in the RHS of Eq.(C-1) we arrive at the following upper bound ,

$$\int_{-\infty}^\infty p_n^{1/2}(x) p_n^{1/2}(x-s)dx \leq 2^{1/r} \exp \left\{ \frac{-|s|^r}{2b} \right\} + \mathcal{K} s \exp \frac{-|s|^r}{2^r b}. \tag{C-6}$$

Thus it must be that,

$$\frac{-\log\{\int_{-\infty}^{\infty} p_n^{1/2}(x)p_n^{1/2}(x-s)dx\}}{-\log p_n(s)} \geq \frac{-\log\{2^{1/r}\exp\frac{-|s|^r}{2b} + \mathcal{K}s\exp\frac{-|s|^r}{2^r b}\}}{-\log\mathcal{K} + |s|^r/b}$$

Therefore from Eq. C ,

$$\begin{aligned}
\mathcal{Z} &= \lim_{s\to\infty} \frac{-\log\{\int p_n^{1/2}(x)p_n^{1/2}(x-s)dx\}}{-\log p_n(s)} \\
&\geq \lim_{s\to\infty}\left\{\frac{-\log\left\{\exp\left[-\frac{|s|^r}{2^r b}\right]\right\}}{|s|^r/b} - \frac{\log\left\{\mathcal{K}s + 2^{1/r}\exp\left[\frac{-|s|^r}{2b} + \frac{|s|^r}{2^r b}\right]\right\}}{|s|^r/b}\right\}.
\end{aligned}$$

Now as $s\to\infty$: $\exp\left[\frac{-|s|^r}{2b} + \frac{|s|^r}{2^r b}\right] \to 0$ since $r \geq 1$, so the above limit reduces to,

$$\begin{aligned}
\mathcal{Z} &\geq \lim_{s\to\infty}\left\{\frac{|s|^r/2^r b}{|s|^r/b} - \frac{\log\{\mathcal{K}s\}}{|s|^r/b}\right\} \\
&= \frac{1}{2^r}
\end{aligned} \tag{C-7}$$

thus establishing a lower bound on $\mathcal{Z}$. Now we seek to bound $\mathcal{Z}$ from above by the same quantity and we proceed as follows.

We know that :

$$|x|^r + |x-s|^r \leq s^r - \left\{\frac{s^r - \frac{s^r}{2^{r-1}}}{\frac{s}{2}}\right\}x$$

for $x \in [0, s]$. Hence,

$$\begin{aligned}
\int_0^s p_n^{1/2}(x)p_n^{1/2}(x-s)dx &= 2\int_0^{s/2} p_n^{1/2}(x)p_n^{1/2}(x-s)dx \\
&\geq 2\int_0^{s/2} \mathcal{K}e^{-\frac{\{s^r - x[s^{r-1}(2-\frac{1}{2^{r-2}})]\}}{2b}}dx \\
&= 2\mathcal{K}e^{-\frac{s^r}{2b}}\int_0^{s/2} e^{\frac{[s^{r-1}(2-\frac{1}{2^{r-2}})]x}{2b}}dx \\
&= \frac{2\mathcal{K}e^{-\frac{s^r}{2b}}}{\frac{[s^{r-1}(2-\frac{1}{2^{r-2}})]}{2b}}\left\{e^{\frac{s^r[1-\frac{1}{2^{r-1}}]}{2b}} - 1\right\}.
\end{aligned}$$

Therefore we have,

$$\begin{aligned}
\int_{-\infty}^{\infty} p_n^{1/2}(x)p_n^{1/2}(x-s)dx &\geq \int_0^s p_n^{1/2}(x)p_n^{1/2}(x-s)dx. \\
&\geq \frac{2\mathcal{K}e^{-\frac{s^r}{2b}}}{\frac{[s^{r-1}(2-\frac{1}{2^{r-2}})]}{2b}}\left\{e^{\frac{s^r[1-\frac{1}{2^{r-1}}]}{2b}} - 1\right\}
\end{aligned}$$

Therefore,

$$\mathcal{Z} \leq \lim_{s\to\infty}\frac{\left\{-\log\left\{\frac{2\mathcal{K}}{\frac{[s^{r-1}(2-\frac{1}{2^{r-2}})]}{2b}}\right\} - \log e^{-\frac{s^r}{2b}} - \log\left\{e^{\frac{s^r[1-\frac{1}{2^{r-1}}]}{2b}}\right\}\right\}}{\frac{s^r}{b}}$$

$$
\begin{aligned}
&= \lim_{s\to\infty} \frac{\frac{s^r}{2b} - s^r[1 - \frac{1}{2^{r-1}}]}{\frac{s^r}{b}} \\
&= \frac{1}{2^r} \qquad\qquad \text{(C-8)}
\end{aligned}
$$

Hence combining the upper and lower bounds of Eqs. C-7 and C-8 we get the desired result.□

2 **For** $r \leq 1$ **:**

Let us start out by establishing a general upper bound valid for all probability densities. We shall first upper bound the Chernoff information C(s) using Jensen's inequality :

$$
\begin{aligned}
d_C(s) &= -\log\{\int p_n^{1/2}(x)p_n^{1/2}(x-s)dx\} \\
&= -\log\left\{E_x\left\{\frac{p_n^{1/2}(x-s)}{p_n^{1/2}(x)}\right\}\right\} \\
&\leq E_x\left\{-\log\left\{\frac{p_n^{1/2}(x-s)}{p_n^{1/2}(x)}\right\}\right\} \\
&= \frac{1}{2}d_{KL}(s) \qquad\qquad \text{(C-9)}
\end{aligned}
$$

Thus it is sufficient to show that,

$$
\lim s \to \infty \frac{d_{KL}(s)}{-\log p_n(s)} = 1. \qquad\qquad \text{(C-10)}
$$

Proceeding: we may write the above ratio as

$$
\frac{d_{KL}(s)}{-\log p_n(s)} = \frac{\int\{\log\{p_n(x) - \log p_n(x-s)\}\}\, p_n(x)dx}{-\log p_n(s)}.
$$

The final result is readily established by making the following assumptions

(a) $p_n(x)$ is symmetric in the tails,

(b) $\lim_{s\to\infty}\left\{\frac{\log p_n(x)}{\log p_n(s)}\right\} = 0$ for all x,

(c) $\lim_{s\to\infty}\left\{\frac{\log p_n(x-s)}{\log p_n(s)}\right\} = 1$ for all x.

to arrive at:

$$
\mathcal{Z} \leq \lim s \to \infty \left\{\frac{1}{2}\frac{d_{KL}(s)}{\{-\log p_n(s)\}}\right\} = \frac{1}{2}
$$

Thus establishing the desired result. As before, we shall lower bound $\mathcal{Z}$ by the same quantity namely $\frac{1}{2}$. Here we make use of the fact that :

$$
|x|^r + |x-s|^r \geq |s|^r
$$

for $x \in [0, s]$. Therefore,

$$
\begin{aligned}
\int_0^s p_n^{1/2}(x)p_n^{1/2}(x-s)dx &\leq \mathcal{K}\int_0^s e^{\{-\frac{s^r}{2b}\}}dx \\
&\leq \mathcal{K}se^{\{-\frac{s^r}{2b}\}}
\end{aligned}
$$

also from Eqs. (C-4) and (C-5),

$$\int_{-\infty}^{0} p_n^{1/2}(x) p_n^{1/2}(x-s)dx \leq \frac{2^{1/r}}{2} e^{\left\{\frac{-|s|^r}{2b}\right\}}$$

and,

$$\int_{s}^{\infty} p_n^{1/2}(x) p_n^{1/2}(x-s)dx \leq \frac{2^{1/r}}{2} e^{\left\{\frac{-|s|^r}{2b}\right\}}.$$

Therefore,

$$\begin{aligned}\int_{-\infty}^{\infty} p_n^{1/2}(x) p_n^{1/2}(x-s)dx &\leq \mathcal{K} s e^{\left\{-\frac{s^r}{2b}\right\}} + 2^{1/r} e^{\left\{\frac{-|s|^r}{2b}\right\}} \\ &= e^{\left\{\frac{-|s|^r}{2b}\right\}} \{2^{1/r} + \mathcal{K} s\}.\end{aligned}$$

Hence,

$$\begin{aligned}\mathcal{Z} &\geq \lim_{s\to\infty} \left\{ \frac{-\log\{2^{1/r} + \mathcal{K} s\} + \frac{s^r}{2b}}{\frac{s^r}{b}} \right\} \\ &= \frac{1}{2}.\end{aligned}$$

By combining the upper and lower bounds together we get the desired result.
If we consider only the Generalized Gaussian density function, we must have:
For $r \geq 1$:

$$d_C(s) = \frac{-\log p_n(s)}{2^r} \qquad \text{(C-11)}$$

and for $r \leq 1$:

$$d_C(s) = \frac{-\log p_n(s)}{2} \qquad \text{(C-12)}$$

for large 's'.

REFERENCE

C1. A. N. Shiryayev, *Probability*. New York: Springer-Verlag, 1984.

Appendix D

CM FORTRAN SIMULATION CODES

```
        program simulation
        integer NtMax, NMAX, MMAX, EMAX, SMAX
        parameter (NMAX= 256, N2MAX= 256, MMAX=20, EMAX= 7)
C       NtMAX: max number of blocks of time series data
C       NMAX: time series length; actually a slice small enough to fit machine
C       N2MAX: second time series length to do multiple slices in parallel
C       MMAX: maximum number of channels
C       EMAX: maximum number of Energy levels
        Real p0,p1                ! apriori prob of H0 to H1
        Real, array(EMAX), DATA ::EE = [0.1, 1.0, 10.0, 30.0, 50.0, 75.0, 100.0]
        Real N(NMAX, N2MAX, MMAX)! time series of noise, variance 1 noise
                                 !  iid among samples and among channels
        Real W(NMAX, N2MAX)      ! time series of weighted received signal
        Real WW(NMAX, N2MAX)     ! time series of weighted received signal
        Real R(NMAX, N2MAX)      ! time series of received data
        Real L(NMAX, N2MAX)      ! time series of Bayesian statistic
        Real P(MMAX, EMAX)       ! prob. of error
        Real Ptemp(MMAX, EMAX)   ! prob. of error
        Real a, temp, s, noise
        character*10 PDF, myPDF
CMF$    LAYOUT N(:serial, :news, :news), L(:serial, :news)
CMF$    LAYOUT R(:serial, :news)
CMF$    LAYOUT W(:serial, :news)
CMF$    LAYOUT WW(:serial, :news)
CMF$    LAYOUT P(:news, :news)
CMF$    LAYOUT Ptemp(:news, :news)
        Integer ns, ne, i, im, m, Ei, ii, j, jj
        Integer thePDF
        include 'Random.h'
        include 'LogLRT.h'
        include 'angle.comments'

        p0 =0.5
        p1 =0.5
10      print*,'Which noise PDF ?
        read*,PDF
        print*,PDF
        if (whichPDF(PDF) .EQ. 0) then
           print*,'hey'
```

```
          go to 10
        end if
        print*,whichPDF(PDF)
11      print*,'Which detector PDF ? '
        read*,myPDF
        print*,myPDF
        if (whichPDF(myPDF) .EQ. 0) then
          print*,'hey'
          go to 11
        end if
        print*,'How many slices (', NMAX*N2MAX,' samples each)?'
        read*,NtMAX
        P(1:MMAX,1:EMAX) = 0.0
        do Nt =1, NtMAX         !for all time slices Nt = 1: NtMax
           do im = 1, MMAX
             N(1:NMAX, 1:N2MAX, im) = MyRand(NMAX,N2MAX, PDF)
           end do
           Ptemp(1:MMAX, 1:EMAX) = 0.0
           do Ei = 1, EMAX
              E0 = EE(Ei)
              do m=1, MMAX      ! m is the grouping index
                 L(:,:) = 0.0   ! init the statistic array
                 a = Sqrt(E0/m)
                  W(:,:) = 0.0
                 do i=1, m      ! channel i
C                     R = a/2.0 + N(:,:,i)   !this line for Importance Sampling
                      R = N(:,:,i)           !this line for Monte Carlo
C                     W = W + LogLRT(NMAX, N2MAX, R, a/2.0, PDF, 1) !for IS
                      L = L + LogLRT(NMAX, N2MAX, R, a, myPDF, 1)   !for both
                 end do
                 WW(:,:) = 0.0
                 forall (ii=1:NMAX, jj=1:N2MAX, L(ii,jj).LT.Log(p1/p0))
     1               WW(ii,jj) = 1.0     !for Monte Carlo
C     1                      WW(ii,jj) = Exp(W(ii,jj))  !for IS
                 temp = sum(WW)
                 Ptemp(m, Ei) = temp
C                 Accumulate the empirical prob  of error across time slice
                 P(m,Ei) = P(m,Ei) + temp
              end do ! channel grouping
           end do ! energy
        end do !time slices
        P= P/(NMAX*N2MAX*NtMAX)
        print*,PDF,'-',myPDF
        write(*,98)EE(1:EMAX)
98        format(1x,' m|, E->',7(F16.4, '  '))
        do m= 1,MMAX
          write(*,99)m, P(m, 1:EMAX)
99          format(1x,I4,TR1,7(G20.8, ' '))
```

```
      end do
      end

      include 'Random.fcm'
      include 'LogLRT.fcm'
C.....................................................................
C   File LogLRT.fcm

      function LogLRT(NN,N2,  R, a, PDF, s)
      Integer NN, N2          ! vector length, number of channels
      Real LogLRT(NN,N2)
      Real R(NN, N2), a, s    ! received vector, translation, std dev
      character*10 PDF        ! which pdf
CMF$  LAYOUT LogLRT(:serial, :news), R(:serial, :news)

      integer i
      Real a2, s2
      Real AA2, AA4, GQuarter, GFiveQ, GThreeQ
      interface
        integer function whichPDF(PDF)
        character*10 PDF
      end interface

C      calculates the statistic for use in the Optimal test
      a2 = a*a
      s2 = s*s
      select case (whichPDF(PDF))
      case (:0)
         LogLRT(1:NN, 1:N2) = 0.0
      case (1)        !Gaussian
         LogLRT(1:NN, 1:N2) =-(a/s2)*R(1:NN, 1:N2) + a2/(2.0*s2)
      case (2)        !Laplacian
         LogLRT(1:NN, 1:N2) =
     1          -(1/s)*Sqrt(2.0)*(Abs(R(1:NN,1:N2)) - Abs(R(1:NN,1:N2) -a))
      case (3)        !Cauchy
         LogLRT(1:NN,1:N2) =
     1        Log((1.0 + ((R(1:NN,1:N2) -a)/s)**2)/(1+ (R(1:NN,1:N2)/s)**2))
      case (4:)       !GenGasuss4
         GQuarter = 3.6256099082
         GThreeQ = 1.2254167024
         GFiveQ   = GQuarter/4.0
         AA2 = s2*GQuarter/GThreeQ
         AA4 = AA2*AA2
         LogLRT(1:NN,1:N2) = (1.0/AA4)*((R(1:NN,1:N2)-a)**4 -R(1:NN,1:N2)**4)
      end select
      return
      end function
C.....................................................................
```

```
C     File Random.fcm

        Function RandUniform(N,N2)
        Integer N,N2
        Real RandUniform(N,N2)
CMF$    LAYOUT RandUniform(:serial,:news)

100     call cmf_random(RandUniform(1:N,1:N2))
        if ((minval(RandUniform) <= 0.0) .OR. (maxval(RandUniform) >= 1.0))then
           goto 100
        end if
        end

        Function RandGaussian(N,N2)
        Integer N,N2
        Real RandGaussian(N,N2)
        Real Temp(N,N2)
        Logical ISHALF(N, N2)
CMF$    LAYOUT RandGaussian(:serial, :news)
CMF$    LAYOUT Temp(:serial, :news)
CMF$    LAYOUT ISHALF(:serial, :news)
        integer i, j
        real twopi

        twopi = 4.0*acos(0.0)
100     call cmf_random(Temp(1:N,1:N2))
        RandGaussian(1:N,1:N2) = SQRT(-2.0*Log(Temp(1:N,1:N2)))
200     call cmf_random(Temp(1:N,1:N2))
        RandGaussian = RandGaussian*cos(TwoPi*Temp)
        end

        Function RandLaplacian(N,N2)
        Integer N,N2
        Real RandLaplacian(N, N2)
        Real Temp(N, N2)
CMF$    LAYOUT RandLaplacian(:serial, :news), Temp(:serial, :news)
        Real s2
        Integer i, j

100     call cmf_random(Temp(1:N,1:N2))
        if ((minval(Temp) <= 0.0) .OR. (maxval(Temp) >= 1.0))  then
          goto 100
        end if
        s2 = 1.0 /Sqrt(2.0)
        RandLaplacian(1:N,1:N2) = s2*Log(Temp(1:N,1:N2))
        call cmf_random(Temp(1:N,1:N2))
        forall (i=1:N, j=1:N2, Temp(i,j) > 0.5)
     1          RandLaplacian(i,j) = - RandLaplacian(i,j)
```

```
      end

      Function RandCauchy(N,N2)
      Integer N,N2
      Real RandCauchy(N,N2)
      Real Temp(N,N2)
CMF$  LAYOUT RandCauchy(:serial, :news), Temp(:serial, :news)
      Integer i, j
      Pi = 2.0*acos(0.0)

100      call cmf_random(Temp(1:N,1:N2))
      if ((minval(Temp) <= 0.0) .OR. (maxval(Temp) >= 1.0))  then
C           print*,'bad'
          goto 100
      end if
      RandCauchy(1:N,1:N2) = tan(Pi*(Temp(1:N, 1:N2) - 0.5))
      end

      Include 'Gamma.fcm'

      Function RandGenGauss4(N,N2)
      Integer N,N2
      Real RandGenGauss4(N,N2)
      Real Temp(N,N2), X(N,N2)
CMF$  LAYOUT RandGenGauss4(:serial, :news), Temp(:serial, :news)
CMF$  LAYOUT X(:serial, :news)
      Real A, A4, GQuarter, GFiveQ, GThreeQ, alpha
      Integer i, j, ii, jj
      real twopi
      interface
        function RandGamma1(N,N2,aa)
        integer N, N2
        real RandGamma1(N,N2), aa
CMF$    LAYOUT RandGamma1(:serial, :news)
      end interface

      twopi = 4.0*acos(0.0)
      alpha = 1.0/4.0
      GQuarter = 3.6256099082
      GThreeQ = 1.2254167024
      GFiveQ   = GQuarter/4.0
      A = Sqrt(GQuarter/GThreeQ)
100      call cmf_random(Temp(1:N,1:N2))
      if ((minval(Temp) <= 0.0) .OR. (maxval(Temp) >= 1.0))  then
C           print*,'bad'
          goto 100
      end if
      X(1:N, 1:N2) = RandGamma1(N, N2, alpha)
```

```
      RandGenGauss4 = A*SQRT(SQRT(X))
      call cmf_random(Temp(1:N,1:N2))
      forall (ii=1:N, jj=1:N2, Temp(ii,jj)>0.5)
     1        RandGenGauss4(ii,jj) = - RandGenGauss4(ii,jj)
      end function

      Function MyRand(N,N2, PDF)
      Integer N, N2
      Real MyRand(N, N2)
      Character*10 PDF
      Real Temp(N, N2)
CMF$      LAYOUT MyRand(:serial, :news), Temp(:serial, :news)

      interface
        integer function whichPDF(PDF)
        character*10 PDF
      end interface

      interface
        function RandUniform (N,N2)
        integer N,N2
        real RandUniform(N,N2)
CMF$    LAYOUT RandUniform(:serial, :news)
      end interface

      interface
        function RandGaussian (N,N2)
        integer N,N2
        real RandGaussian(N,N2)
        real Temp(N,N2)
CMF$    LAYOUT RandGaussian(:serial, :news), Temp(:serial, :news)
      end interface
      interface
        function RandLaplacian (N,N2)
        integer N,N2
        real RandLaplacian(N,N2)
        real Temp(N,N2)
CMF$    LAYOUT RandLaplacian(:serial, :news), Temp(:serial, :news)
      end interface
      interface
        function RandCauchy (N, N2)
        integer N,N2
        real RandCauchy(N, N2)
        real Temp(N, N2)
CMF$    LAYOUT RandCauchy(:serial, :news), Temp(:serial, :news)
      end interface
      interface
        function RandGenGauss4 (N,N2)
```

```
          integer N,N2
          real RandGenGauss4(N,N2)
          real Temp(N,N2)
CMF$      LAYOUT RandGenGauss4(:serial, :news), Temp(:serial, :news)
        end interface

        select case (whichPDF(PDF))
        case (:0)
           MyRand = 0.0
        case (1)
           MyRand = RandGaussian(N,N2)
        case (2)
           MyRand = RandLaplacian(N,N2)
        case (3)
           MyRand = RandCauchy(N,N2)
        case (4)
           MyRand = RandGenGauss4(N,N2)
        case (5)
           MyRand = RandUniform(N,N2)
        case (6:)
           MyRand = 0.0
        end select
        return
        end function

        Function whichPDF(PDF)
        integer whichPDF
        character*10 PDF

        if(index(PDF,'Normal').NE.0) then
           whichPDF = 1
        else if(index(PDF,'normal').NE.0) then
           whichPDF = 1
        else if(index(PDF,'gaussian').NE.0) then
           whichPDF = 1
        else if(index(PDF,'Gaussian').NE.0) then
           whichPDF = 1
        else if(index(PDF,'Laplacian').NE.0) then
           whichPDF = 2
        else if(index(PDF,'Laplace').NE.0) then
           whichPDF = 2
        else if(index(PDF,'laplacian').NE.0) then
           whichPDF = 2
        else if(index(PDF,'laplace').NE.0) then
           whichPDF = 2
        else if(index(PDF,'Cauchy').NE.0) then
           whichPDF = 3
        else if(index(PDF,'cauchy').NE.0) then
```

```
            whichPDF = 3
         else if(index(PDF,'GenGauss4').NE.0) then
            whichPDF = 4
         else if(index(PDF,'gengauss4').NE.0) then
            whichPDF = 4
         else if(index(PDF,'Gengauss4').NE.0) then
            whichPDF = 4
         else if(index(PDF,'uniform').NE.0) then
            whichPDF = 5
         else if(index(PDF,'Uniform').NE.0) then
            whichPDF = 5
         else
            whichPDF = 0
         end if
         return
         end function

         Function erf(NMAX, N2MAX, R, x)
         real erf
         integer NMAX, N2MAX
         real R(NMAX, N2MAX)
         real x
CMF$       LAYOUT R(:serial, :news)
         real One(NMAX, N2MAX)
CMF$       LAYOUT One(:serial, :news)
         integer ii,jj

         One(:,:) = 0.0
         forall(ii=1:NMAX, jj=1:N2MAX, R(ii,jj).LT.real(x))
     1           One(ii,jj)=1.0
         erf = sum(One)/real(N2MAX*NMAX)
         return
         end function

C.................................................................................
C     File Gamma.fcm

         Function RandGamma1(N, N2, aa)
         Integer N,N2
         Real RandGamma1(N,N2)
         Real aa
         Real Temp(N,N2)
         Real Temp2(N,N2)
         Real X(N,N2)
         Real Y(N,N2)
         Real Z(N,N2)
CMF$       LAYOUT RandGamma1(:serial, :news), Temp(:serial, :news)
CMF$       LAYOUT Temp2(:serial, :news)
```

```
CMF$        LAYOUT X(:serial, :news)
CMF$        LAYOUT Y(:serial, :news)
        Integer i, j, retry
        Real b

        retry = 0
C        print*,'aa', aa
        b = 1.0/(1.0-aa)
        call cmf_random(Temp(1:N, 1:N2))
        X(:, :) = Temp(:, :)**(1.0/aa)
        call cmf_random(Temp(1:N, 1:N2))
        Y(:, :) = X(:, :) + Temp(:,:)**b
100       if (maxval(Y).GT.1.0) then
          retry = retry + 1
          if (mod(retry,100) .EQ.0 ) then
C             print*,'retry =', retry
          endif
          call cmf_random(Temp2(1:N, 1:N2))
          call cmf_random(Temp(1:N, 1:N2))
          where (Y(:,:).GT.1.0)
            X(:, :) = Temp(:, :)**(1.0/aa)
            Y(:, :) = X(:, :) + Temp2(:,:)**b
          end where
          goto 100
        end if
C        print*,'retry = ', retry
101       call cmf_random(Temp(1:N,1:N2))
        if (minval(Temp) <= 0.0) then
C           print*,'bad'
          goto 101
        end if
102       call cmf_random(Temp2(1:N, 1:N2))
        if (minval(Temp2)<= 0.0) then
C           print*,'bad'
          goto 102
        end if
        RandGamma1 = X*(-log(Temp*Temp2))
        end
...............................................................................
C       File Random.h

        interface
          function myRand(N, N2, PDF)
          integer N, N2
          character*10 PDF
          real myRand(N,N2)
          real Temp(N,N2)
CMF$      LAYOUT myRand(:serial, :news), Temp(:serial, :news)
```

```
        end interface

        interface
          integer function whichPDF(PDF)
          character*10 PDF
        end interface

        interface
          function RandUniform (N,N2)
          integer N,N2
          real RandUniform(N,N2)
CMF$      LAYOUT RandUniform(:serial, :news)
        end interface

        interface
          function RandGaussian (N,N2)
          integer N,N2
          real RandGaussian(N,N2)
          real Temp(N,N2)
CMF$      LAYOUT RandGaussian(:serial, :news), Temp(:serial, :news)
        end interface
        interface
          function RandLaplacian (N,N2)
          integer N,N2
          real RandLaplacian(N,N2)
          real Temp(N,N2)
CMF$      LAYOUT RandLaplacian(:serial, :news), Temp(:serial, :news)
        end interface
        interface
          function RandCauchy (N, N2)
          integer N,N2
          real RandCauchy(N, N2)
          real Temp(N, N2)
CMF$      LAYOUT RandCauchy(:serial, :news), Temp(:serial, :news)
        end interface
        interface
          function RandGenGauss4 (N,N2)
          integer N,N2
          real RandGenGauss4(N,N2)
          real Temp(N,N2)
CMF$      LAYOUT RandGenGauss4(:serial, :news), Temp(:serial, :news)
        end interface

        interface
           function erf(NMAX, N2MAX, R, x)
           real erf
           integer NMAX, N2MAX
           real R(NMAX, N2MAX)
```

```
          real x
CMF$          LAYOUT R(:serial, :news)
        end interface
.................................................................................
C        File LogLRT.h

        interface
        function LogLRT(NN,N2,  R, a, PDF, s)
        Real LogLRT(NN,N2)
        character*10 PDF        ! which pdf
        Integer NN, N2          ! vector length, number of channels
        Real R(NN, N2), a, s
CMF$        LAYOUT LogLRT(:serial, :news), R(:serial, :news)
        end interface
```