The Statistical Characteristics of Power-Spectrum Subband Energy Ratios under Additive Gaussian White Noise

Han Li 1,2, Yanzhu Hu 1, Song Wang 1*, and Zhen Meng 1
1 School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876 China; lihan_60@bupt.edu.cn(H.L.); huyanzhu2019@163.com(Y.H.); wongsangwongsang@163.com(S.W.); zhenmeng2019@163.com(Z.M.)
2 School of Intelligent Equipment, Shandong University of Science and Technology, Taian 271019 China; lihan_60@bupt.edu.cn(H.L.)
* Correspondence: wongsangwongsang@163.com; Tel.: 00-86-010-13366036076

Abstract: The power-spectrum subband energy ratio (PSER) has been applied in a variety of fields, but reports on its statistical properties have been limited. As such, this study investigates these characteristics in the presence of additive Gaussian white noise for both pure noise and mixed signals. By analyzing the probability and independence of power spectrum bins, and the relationship between the F and beta distributions, we developed a probability distribution for the PSER. Results showed that in the case of pure noise, the PSER followed a beta distribution. In addition, the probability density function and the quantile exhibited no relationship with the noise variance, only with the number of lines in the power spectrum, that is, PSER was not affected by noise. When Gaussian white noise was mixed with the known signal, the resulting PSER followed a doubly non-central beta distribution and had relationship with the noise variance. In this case, it was difficult to identify the quantile, as the probability density and cumulative distribution functions were represented by infinite series. However, when a spectral bin did not contain the power spectrum of the known signal, an approximated quantile was found. This quantile was strictly proved to be in agreement with the quantile in the case of pure noise, which may offers a convenient methodology for identifying valid signals.

Keywords: Power-spectrum subband energy ratio; beta distribution distribution; doubly non-central beta; quantile

1. Introduction

The power-spectrum subband energy ratio (PSER) is a common metric used to represent the proportion of signal energy in a spectral band. The PSER is derived from spectral analysis and has been extensively applied in the fields of remote communication [1, 2], earthquake modeling [3], machine design [4], and geological engineering [5]. For example, Preisig et al. used the PSER to study signal detection in underwater acoustic environments [6]. Wang et al. described the distribution of acoustic emission signals in the fracturing of fine sandstone [7]. Xu and Guan extracted the characteristics of underground micro-seismic signals collected by a distributed optical fiber using the PSER [8]. Huang et al. investigated the effects of various dual-peak spectral parameters on dispersive coefficients [9]. Yang et al. used PSER as signal feature in coal gangue recognition [10]. Lu et al. studied micro-seismic signals in dual-layer igneous strata [11]. Kong et al. described the feature of electromagnetic radiation under coal oxidation [12].

There exists a direct relationship between the amplitude spectrum, the power spectrum, and the power-spectrum energy ratio. The additive white noise in the mixed signal can be expressed as:

$$s(n) = x(n) + z(n), n = 0,1, \cdots, N - 1.$$  \hspace{1cm} (1)

where \( s(n) \) is the noisy signal, \( x(n) \) is the effective signal, and \( z(n) \) is the Gaussian white noise with a mean of zero and a variance of \( \sigma^2 \). The pure noise case occurs when \( x(n) \) is zero and the signal \( s(n) \) contains only noise. The mixed case occurs when \( x(n) \) is not zero and
$s(n)$ includes both noise and the effective signal. The primary objective of this study is to derive statistical characteristics for the PER in each of these two cases.

The spectrum of $s(n)$ is given by:

$$\tilde{S}(k) = \sum_{n=0}^{N-1} s(n)e^{-j\frac{2\pi kn}{N}}, k = 0, 1, \cdots, N - 1.$$  \hspace{1cm} (2)

where $N$ is the number of spectral lines, $j$ is the imaginary unit, and the arrow superscript denotes a complex function. The $k^{th}$ line in the power spectrum of $s(n)$ can be expressed as:

$$P(k) = \frac{1}{N}|\tilde{S}(k)|^2.$$  \hspace{1cm} (3)

The definition of PSER for $P(k)$ is: the adjacent $d$ spectral bins in the power spectrum are combined, and the energy proportion in the whole power spectrum is calculated, that is

$$r(k) = \frac{\sum_{l=0}^{d-1} P(kd + l)}{\sum_{i=0}^{N-1} P(i)}, 1 \leq d < N, k = 0, 1, \cdots, N' - 1,$$ \hspace{1cm} (4)

where $\sum_{i=0}^{N-1} P(i)$ represents the total energy in the power spectrum, $\sum_{l=0}^{d-1} P(kd + l)$ represents the total energy in the $k^{th}$ power spectrum subband. There are $N$ spectral bins in the power spectrum, and $\lfloor N/d \rfloor$ subfrequency bands are formed after the combination. The symbol $\lfloor \rfloor$ stands for rounding down. If the remaining spectral bins is less than $d$, they will not be merged, and directly discarded. Let $N' = \lfloor N/d \rfloor$, then $k = 0, 1, \cdots, N' - 1$.

The frequency resolution of PSER is inferior to that of the power spectrum, but the advantage of PSER is that it can reduce the data quantity greatly. Power spectrum is a common feature in the process of signal recognition. However, if power spectrum is directly used, the characteristic dimension is too large. Therefore, several adjacent spectral bins are often combined into a sub-frequency band. That is, the subsequent training and recognition work will be carried out after the reduction of the feature dimension. For example, if the signal's time series contains 1024 sampling points, the power spectrum contains 1024 spectral bins. If four adjacent spectral bins are combined into one sub-frequency band, the 1024 spectral bins will be transformed into 256 sub-frequency bands, and the remaining data will be only a quarter of the original.

The statistical characteristics of frequency and power spectra have been extensively studied for Gaussian white noise. The real and imaginary parts of the spectrum follow a Gaussian distribution [2] and the amplitude conforms to a Rayleigh distribution [13]. Groth analyzed the statistical characteristics of power spectrum with white noise [14]. Johnson et al. studied the distribution characteristics of power spectra acquired using a periodogram averaging technique [15]. However, little has been reported on the statistical properties of the power-spectrum energy ratio. As such, this study systemically investigates the PSER in the presence of additive Gaussian white noise, to provide a theoretical basis for signal detection and recognition.

The remainder of this paper is organized as follows. Section II discusses the statistical characteristics of the frequency and power spectra for Gaussian white noise. Section III develops a basic solution for the probability distribution of the PSER. A statistical analysis of the PSER is conducted in Section IV for pure noise signals. A comparable investigation is conducted for mixed noise signals in Section V and Section VI provides additional details concerning the derivation process.

2. Some Useful Statistical Characteristics

2.1 The Statistical Characteristics of Gaussian White Noise

The statistical properties of white noise in power spectrum estimation are of significant importance for PER analysis, which was developed in this study as follows. The term $z(n)$ is assumed to be a stationary Gaussian noise sequence. The following postulates then hold for this set:
1) For $\forall n \in \{0, 1, 2, \ldots, N-1\}$, $z(n)$ follows a normal distribution with a mean of zero and a variance of $\sigma^2$.

2) For $\forall n_1, n_2 \in \{0, 1, 2, \ldots, N-1\}$ and $n_1 \neq n_2$, $z(n_1)$ and $z(n_2)$ are mutually independent random variables. As such,

$$E\{z(n_1)z(n_2)\} = E\{z(n_1)\}E\{z(n_2)\} = 0 .$$

(5)

The Gaussian white noise spectrum, calculated using a discrete Fourier transform (DFT) is given by:

$$\tilde{Z}(k) = \sum_{n=0}^{N-1} z(n) \left[ \cos\left( \frac{2\pi}{N} kn \right) - j \sin\left( \frac{2\pi}{N} kn \right) \right], \ k = 0, 1, 2, \ldots, N - 1 .$$

(6)

Let the real and imaginary parts of $\tilde{Z}(k)$ be $Z_r(k)$ and $Z_i(k)$, respectively. Then $Z_r(k)$ and $Z_i(k)$ are mutually independent, and they all follow the Gaussian distribution [14], that is

$$Z_r(k) \sim \mathcal{N}(0, \frac{N\sigma^2}{2}) , \ Z_i(k) \sim \mathcal{N}(0, \frac{N\sigma^2}{2}) .$$

(7)

The power spectrum for the Gaussian white noise term $\tilde{Z}(k)$ is defined as:

$$P_z(k) = \frac{1}{N} |\tilde{Z}(k)|^2 = \frac{1}{N} \left[ Z_r^2(k) + Z_i^2(k) \right] , \ k = 0, 1, \ldots, N - 1 .$$

(8)

$P_z(k)$ is also called a power spectrum bin. Any two white noise power spectral lines are independent[2]. $P_z(k)$ follows a chi-squared distribution with two degrees of freedom[2]. The probability density function (PDF) for $P_z(k)$ is:

$$f_{P_z(k)}(x) = \begin{cases} \lambda e^{-\frac{x}{\lambda}} , & x > 0 \\ 0 , & x \leq 0 \end{cases} , \ \lambda = \frac{1}{N\sigma^2} > 0 .$$

(9)

The term $\sum_{i=0}^{d-1} P_z(kd + i)$, the sum of $d$ Gaussian white noise power spectral bins, follows a chi-squared distribution with $2d$ degrees of freedom.

2.2 The Statistical Characteristics of Signal Mixed with Gaussian White Noise

(1) Power Spectrum

When the effective signal in Equation (1) is not zero, its discrete Fourier transform can be expressed as:

$$\tilde{S}(k) = \tilde{X}(k) + \tilde{Z}(k) , \ k = 0, 1, \ldots, N - 1 .$$

(10)

The power spectrum for $\tilde{X}(k)$ and $\tilde{Z}(k)$ are then respectively given by

$$P_x(k) = \frac{1}{N} \left[ X_r^2(k) + X_i^2(k) \right] , \ k = 0, 1, \ldots, N - 1 ,$$

(11)

$$P_z(k) = \frac{1}{N} \left[ Z_r^2(k) + Z_i^2(k) \right] , \ k = 0, 1, \ldots, N - 1 .$$

(12)

The power spectrum for $\tilde{S}(k)$ is

$$P_s(k) = \frac{1}{N} \left\{ (X_r(k) + Z_r(k))^2 + (X_i(k) + Z_i(k))^2 \right\} , \ k = 0, 1, \ldots, N - 1 .$$

(13)

Let $S_r(k) = X_r(k) + Z_r(k)$, $S_i(k) = X_i(k) + Z_i(k)$, then

$$P_s(k) = \frac{1}{N} \left[ S_r^2(k) + S_i^2(k) \right] , \ k = 0, 1, \ldots, N - 1 .$$

(14)

Since $\tilde{X}(k)$ represents the spectrum of a known signal, the real and imaginary parts of the
spectrum can be assumed to be constant (i.e., \( a_k = X_k(k) \) and \( b_k = X_1(k) \)) [14]. As a result,
\[
S_k(k) \sim \mathcal{N}(a_k, \frac{N\sigma^2}{2}), \quad S_1(k) \sim \mathcal{N}(b_k, \frac{N\sigma^2}{2}).
\]
(15)

(2) The statistical Characteristics of the sum of multiple power spectra bins for mixed signals

For the convenience of description, the power spectrum line \( P_s(k) \) in the next section is replaced by the random variable \( X_k \) (i.e., \( X_k = P_s(k) \)). The sum of \( d \) power spectrum lines \( \sum_{t=0}^{d-1} P(kd + t) \) is replaced by the random variable \( X_k' \), i.e.
\[
X_k' = P(kd) + P(kd + 1) + \ldots + P(kd + d - 1)
\]
\[
= \frac{1}{N} \sum_{t=0}^{d-1} \left[ S_k^2(kd + t) + S_1^2(kd + t) \right].
\]
(16)

Equation (15) then reduces to the following definitions:
\[
\sqrt{\frac{2}{N\sigma^2}} S_k(kd + t) \sim \mathcal{N}(a_{kd+t}, 1), \quad \sqrt{\frac{2}{N\sigma^2}} S_1(kd + t) \sim \mathcal{N}(b_{kd+t}, 1).
\]

And
\[
\frac{2}{\sigma^2} X_k' = \frac{2}{N\sigma^2} \sum_{t=0}^{d-1} S_k^2(kd + t) + \frac{2}{N\sigma^2} \sum_{t=0}^{d-1} S_1^2(kd + t).
\]

Let \( \lambda_0 = \frac{2(a_i^2 + b_i^2)}{N\sigma^2}, \quad \lambda_i' = \sum_{t=0}^{d-1} \lambda_{kd+t} \). It can be shown that \( 2\frac{X_k'}{\sigma^2} \) follows a non-central chi-square distribution with \( 2d \) degrees of freedom and a non-centrality parameter \( \lambda_i' \) [16]. As such:
\[
\frac{2}{\sigma^2} X_k' \sim \chi_2^2(\lambda_i').
\]
(17)

3. Identifying the Probability Distribution for the Power-Spectrum Subband Energy Ratio

The PSER is denoted as \( C_N(k) \), i.e.
\[
C_N(k) = \frac{X_k'}{\sum_{i=0}^{N-1} X_i} = \frac{X_{kd} + X_{kd+1} + \ldots + X_{kd+d-1}}{X_0 + X_1 + \ldots + X_{N-1}}, \quad k = 0, 1, \ldots, N' - 1.
\]
(18)

Note that \( X_k' \) is included in \( \sum_{i=0}^{N-1} X_i \). As such, the numerator and denominator of \( C_N(k) \) are not independent and the probability distribution for \( C_N(k) \) cannot be calculated from the probability distribution for \( X_k' \) and \( \sum_{i=0}^{N-1} X_i \). It is also clear that \( C_N(k) \) spans a range of \([0, 1]\).

The random variable \( C_N'(k) \) represents the ratio of \( X_k' \) to the sum of the remaining \( N - d \) variables that do not contain:
\[
C_N'(k) = \frac{X_k'}{\sum_{i=0}^{N-1} X_i - X_k'}, \quad k = 0, 1, \ldots, N' - 1.
\]
(19)

It is evident that when \( N \geq d + 1 \), the denominator of \( C_N'(k) \) is not 0. In addition, \( C_N'(k) \) spans a range of \([0, +\infty)\). Note that the numerator \( X_k' \) is not included in the denominator \( \sum_{i=0}^{N-1} X_i - X_k' \), indicating the numerator and denominator of \( C_N'(k) \) are independent. The probability distribution for \( C_N'(k) \) can then be calculated from the probability distributions of \( X_k' \) and \( \sum_{i=0}^{N-1} X_i - X_k' \). As a result, \( C_N(k) \) and \( C_N'(k) \) exhibit the following relationship:
Using this equation, the distribution of $\mathcal{C}_N(k)$ can be calculated indirectly from the distribution of $\mathcal{C}_N'(k)$. The range of values varies from $[0,1]$ to $(0,1)$, for consistency with the beta distribution included later in the study.

4. Statistical Characteristics of PSER for Gaussian White Noise

When there is only noise and no effective signal in $s(n)$, its power spectrum is a white noise power spectrum.

4.1 The Probability Distribution for $\mathcal{C}_N'(k)$

The numerator $X_k'$ in $\mathcal{C}_N'(k)$ follows a chi-squared distribution with $2d$ degrees of freedom and the denominator $\sum_{i=0}^{N-1} X_i - X_k'$ follows the same distribution with $2N - 2d$ degrees of freedom. The product $(N - d)\mathcal{C}_N'(k)/d$ follows an $F$ distribution with $2d, 2N - 2d$ degrees of freedom:

$$\frac{(N - d)}{d} \mathcal{C}_N'(k) = \frac{X_k'/2d}{\left(\sum_{i=0}^{N-1} X_i - X_k'\right)/(2N - 2d)} ~ F(2d, 2N - 2d).$$

According to the probability density function (PDF) of $F$ distribution, the PDF of $\mathcal{C}_N'(k)$ can be obtained

$$f_{\mathcal{C}_N'(k)}(x) = \begin{cases} \frac{1}{B(d, N-d)} \frac{x^{d-1}}{(1+x)^N}, & x > 0, N \geq d + 1, \\ 0, & x \leq 0 \end{cases}$$

where $B(d, N-d) = \frac{\Gamma(d) \Gamma(N-d)}{\Gamma(N)}$ is beta function.

4.2 The Probability Distribution for $\mathcal{C}_N(k)$

The PDF of $\mathcal{C}_N(k)$ can be determined by taking the derivative of the PDF of $\mathcal{C}_N'(k)$:

$$f_{\mathcal{C}_N(k)}(x) = \frac{1}{(1-x)^2} f_{\mathcal{C}_N'(k)}\left(\frac{x}{1-x}\right), \quad 0 < x < 1.$$ (23)

Substituting Equation (22) into (23) produces:

$$f_{\mathcal{C}_N(k)}(x) = \begin{cases} \frac{1}{B(d, N-d)} x^{d-1} (1-x)^{N-d-1}, & 0 < x < 1, N \geq d + 1, \\ 0, & x \leq 0 \text{ or } x \geq 1 \end{cases}$$ (24)

The probability density plot for $\mathcal{C}_N(k)$ is shown in Figure 1.
parameters \( d, N - d \), i.e.
\[
C_N(k) \sim \beta(d, N - d).
\]
When
\[
x = \frac{N - d - 1}{N - 2d - 2},
\]
\( f_{C_N(k)}(x) \) reaches its maximum. The cumulative distribution function (CDF) for \( C_N(k) \) is
\[
F_{C_N(k)}(x) = \begin{cases} 
I_x(d, N - d), & 0 < x < 1 \\
0, & x \leq 0 \text{ or } x \geq 1, N \geq d + 1.
\end{cases}
\]
where \( I_p(a, b) \) is the incomplete beta function.

**Figure 1** The probability density plot for \( C_n(k) \), where \( N=128, d \) is 2, 4 and 16 respectively.

It can be shown that the PDF and CDF of the PSER for Gaussian white noise have no relationship to the noise variance and are only related to the number of power spectrum bins in a subband \( d \) and the number of bins in total power spectrum \( N \). This means that PSER is not affected by noise and has very ideal robustness.

The expectation and variance of the beta distribution \( C_N(k) \) is then given by:
\[
E(C_N(k)) = \frac{d}{N},
\]
\[
Var(C_N(k)) = \frac{Nd - d^2}{N^2(N + 1)}.
\]

4.3 The Quantile of \( C_N(k) \)

The probability that the PSER \( C_N(k) \) is greater than \( F_\alpha \) is equal to \( \alpha \). As a result,
\[
\Pr(C_N(k) > F_\alpha) = \alpha
\]
\[
\Leftrightarrow 1 - \alpha = \Pr(C_N(k) \leq F_\alpha)
\]
\[
\Leftrightarrow 1 - \alpha = I_{\alpha}(d, N - d)
\]
\[
\Leftrightarrow F_\alpha = I^{-1}(1 - \alpha; d, N - d),
\]
where \( I^{-1}(1 - \alpha; d, N - d) \) is the quantile of \( C_N(k) \), and \( I^{-1}(1 - \alpha; d, N - d) \) is the Inverse
function of \( I_{F_d}(d, N - d) \). The quantile has no relationship to the noise variance, therefore it is not affected by changes in the noise intensity. \( I^{-1}(x; a, b) \) can be solved by referring to the method provided in [17, 18].

5. Statistical Characteristics for the PSER in Mixed Cases

In order to distinguish the pure noise from the mixed signal, \( C_{S,N}(k) \) denotes the PSER in mixed case, and \( C'_{S,N}(k) \) denotes the ratio of \( X'_k \) to the sum of the remaining \( N - d \) variables that do not contain in mixed case.

5.1 The Probability Distribution for \( C'_{S,N}(k) \)

The term \( C'_{S,N}(k) \) can be represented as:

\[
C'_{S,N}(k) = \frac{X'_k}{\sum_{i=0}^{N-1} X_i - X'_k} = \frac{2/\sigma^2 X'_k}{2/\sigma^2 \sum_{i=0}^{N-1} X_i} ,
\]

its numerator follows a non-central chi-square distribution with \( 2d \) degrees of freedom and a non-centrality parameter \( \lambda'_k \), and its denominator follows a non-central chi-square distribution with \( 2N - 2d \) degrees of freedom and a non-centrality parameter \( \sum_{i=0}^{N-1} \lambda_i - \lambda'_k \). Therefore, \( C'_{S,N}(k) \) is a ratio of two non-central chi-square distributions which is referred to as \( G \) distribution.

Let \( \delta_1 = \lambda'_k, \delta_2 = \sum_{i=0}^{N-1} \lambda_i - \lambda'_k \). According to the definition of the \( G \) distribution [16], we deduce the PDF of \( C'_{S,N}(k) \) as:

\[
f_{C'_{S,N}(k)}(x;2d,2N-2d;\delta_1,\delta_2) = e^{-\frac{\delta_1 + \delta_2}{2}} \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \frac{\left( \frac{\delta_1}{2} \right)^j \left( \frac{\delta_2}{2} \right)^l}{j!l!} I_{(1+t)}(j + d, N + l - d) \times B(d, j, N - d + l),
\]

where \( I_{p}(a,b) \) is the incomplete beta function defined by:

\[
I_{p}(a,b) = \int_{0}^{p} t^{a-1}(1-t)^{b-1}dt / B(a,b).
\]

5.2 The Probability Distribution for \( C_{S,N}(k) \)

The term \( C_{S,N}(k) \) can be represented as:

\[
C_{S,N}(k) = \frac{X'_k}{\sum_{i=0}^{N-1} X_i} = \frac{2/\sigma^2 X'_k}{2/\sigma^2 \sum_{i=0}^{N-1} X_i} , k = 0, 1, \ldots, N' - 1.
\]
distributions. The numerator is included in denominator the denominator, therefore $X'_k$ and $\sum_{i=0}^{N-1} X_i$ is not independence. According to the definition of the doubly non-central beta distribution\cite{16}, $C_N(k)$ follows a doubly non-central beta distribution with parameters $2d, 2N-2d$ and non-centrality parameters $\lambda'_k, \sum_{i=0}^{N-1} \lambda_i - \lambda'_k$, which can be denoted as:

$$C_{S,N}(k) \sim \beta_{2d,2N-2d} (\lambda'_k, \sum_{i=0}^{N-1} \lambda_i - \lambda'_k).$$

(36)

According to Equation\eqref{eq:20}, the CDF of $C_{S,N}(k)$ can be derived from the CDF of $C'_{S,N}(k)$:

$$F_{C_{S,N}(k)}(x) = \left\{ \begin{array}{ll} e^{-\frac{\delta_1 + \delta_2}{2} \sum_{j=0}^{x} \sum_{i=0}^{x} \left( \frac{\delta_1}{2} \right)^{j} \frac{\left( \frac{\delta_2}{2} \right)^{i}}{j!} I_x(j + d, N + l - d) & , x < 1 \end{array} \right. , \quad 0 < x < 1 .$$

(37)

The derivative of Equation\eqref{eq:20} is the PDF of $C_{S,N}(k)$:

$$f_{C_{S,N}(k)}(x) = \left\{ \begin{array}{ll} e^{-\frac{\delta_1 + \delta_2}{2} \sum_{j=0}^{x} \sum_{i=0}^{x} \left( \frac{\delta_1}{2} \right)^{j} \frac{\left( \frac{\delta_2}{2} \right)^{i}}{j!} (1-x)^{N+d-3} B(j+d, N-d+l)} & , x < 1 \end{array} \right. , \quad 0 < x < 1 .$$

(38)

Since $\delta_1$ and $\delta_2$ have relation with the noise variance $\sigma^2$, therefore $C_{S,N}(k)$ is affected by noise. Both the PDF and CDF of $C_{S,N}(k)$ are represented by infinite series, so their values only be obtained through numerical computation. When $N$ is $128, d = 2, 4, 16$ and $\delta_1 = 5.6, \delta_2 = 20$, the density probability plot of $C_{S,N}(k)$ is as follows:

![Figure 2](image_url)

Figure 2 The probability density plot for $C_{S,N}(k)$, where $N$ is $128, d = 2, 4, 16$ and $\delta_1 = 5.6, \delta_2 = 20$.

5.3 The Quantile of $C_{S,N}(k)$

The quantile for $C_{S,N}(k)$, representing the probability that $C_{S,N}(k)$ is greater than $F_\alpha$, is equal to $\alpha$, i.e.
The quantile \( F_{\alpha} \) plays a very important role in spectral signal detection. However, it is not feasible to directly calculate \( F_{\alpha} \), even when \( \alpha, \delta_1 \) and \( \delta_2 \) are known. Therefore we had to look for an approximate method for computing the quantile.

5.4 The Situation of Power Spectrum Bins Without the Effective Signal

In the spectral analysis, it is very common that some spectral bins do not contain effective signals. For example, in the narrow-band signal, the value of most spectral bins is zero. When there is only noise present in \( X'_k \) (\( \delta_1 = 0 \)), the \( j = 0 \) term in the CDF of \( C_{S,N}(k) \) is not zero (though the remaining terms are zero). In this case, the CDF of \( C_{S,N}(k) \) is given by:

\[
F_{C_{S,N}(k)}(x; \delta_1 = 0) = e^{-\frac{\delta_2}{2}} \sum_{l=0}^{\infty} \frac{\left(\frac{\delta_2}{2}\right)^l}{l!} I_x(d, N+l-d) .
\]

Assuming the probability of the PER being higher than \( F_{\alpha,\delta_1=0} \) is less than or equal to \( \alpha \) yields:

\[
\Pr(C_{S,N}(k) > F_{\alpha,\delta_1=0}; \delta_1 = 0) \leq \alpha
\]

\[
\Leftrightarrow 1 - \Pr(C_{S,N}(k) \leq F_{\alpha,\delta_1=0}; \delta_1 = 0) \leq \alpha
\]

\[
\Leftrightarrow 1 - \alpha \leq \Pr(C_{S,N}(k) \leq F_{\alpha,\delta_1=0}; \delta_1 = 0) .
\]

In order to find an approximate quantile, we extend the range of the quantile, therefore \( F_{\alpha,\delta_1=0} \) is less than or equal to \( \alpha \), but not equal to \( \alpha \). Substituting Equation (40) into (41) produces:

\[
F_{C_{S,N}(k)}(F_{\alpha,\delta_1=0}; \delta_1 = 0) \geq 1 - \alpha
\]

\[
\Leftrightarrow e^{-\frac{\delta_2}{2}} \sum_{l=0}^{\infty} \frac{\left(\frac{\delta_2}{2}\right)^l}{l!} I_x(d, N+l-d) \geq 1 - \alpha
\]

\[
\Leftrightarrow \sum_{l=0}^{\infty} \frac{\left(\frac{\delta_2}{2}\right)^l}{l!} I_x(d, N+l-d) \geq \sum_{l=0}^{\infty} \frac{\left(\frac{\delta_2}{2}\right)^l}{l!} (1 - \alpha) .
\]

It can be difficult to solve this equation and identify the quantile. However, an approximate quantile can easily be found. We first demonstrate the following theorem.

**Theorem 1:** If both \( d \) and \( n \) are positive integers, \( n > d \) and \( 0 < x < 1 \), then when \( x \geq d/(n+1) \),

\[
I_x(d, n-d+1) < I_x(d, n-d+2) < \cdots < I_x(d, n-d+i) .
\]

**Proof** The incomplete beta function has the following formula [19, 20]

\[
I_x(d, n-d+1) = \sum_{j=d}^{n} \binom{n}{j} x^j (1-x)^{n-j} , 1 \leq d \leq n .
\]

Similarly,
\[ I_s(d, n-d+2) = \sum_{j=0}^{n+1} \binom{n+1}{j} x^j (1-x)^{n-j+1}. \]

Because
\[ 1 = (x + (1-x))^n = \sum_{j=0}^{n} \binom{n}{j} x^j (1-x)^{n-j}, \]
therefore
\[ I_s(d, n-d+1) = 1 - \sum_{j=0}^{n-1} \binom{n}{j} x^j (1-x)^{n-j}, \]
\[ I_s(d, n-d+2) = 1 - \sum_{j=0}^{n+1} \binom{n+1}{j} x^j (1-x)^{n-j}. \]

Then
\[ I_s(d, n-d+1) - I_s(d, n-d+2) \]
\[ = 1 - \sum_{j=0}^{n-1} \binom{n}{j} x^j (1-x)^{n-j} - 1 + \sum_{j=0}^{n+1} \binom{n+1}{j} x^j (1-x)^{n-j} \]
\[ = \sum_{j=0}^{n} \left[ \binom{n+1}{j} x^j (1-x)^{n-j} - \binom{n}{j} x^j (1-x)^{n-j} \right]. \]

where
\[ \binom{n+1}{j} x^j (1-x)^{n-j} - \binom{n}{j} x^j (1-x)^{n-j} = \frac{n!}{j!(n-j)!} x^j (1-x)^{n-j} \cdot \frac{j-(n+1)x}{n+1-j}. \]

Because \( 0 \leq j \leq d-1 \) and \( n > d \), so \( n+1-j > 0 \). Because \( j-(n+1)x < d-(n+1)x \) and \( x \geq d/(n+1) \), so \( d-(n+1)x \leq 0, \quad j-(n+1)x \leq 0 \). Therefore
\[ \binom{n+1}{j} x^j (1-x)^{n-j} - \binom{n}{j} x^j (1-x)^{n-j} < 0. \]

and
\[ I_s(d, n-d+1) < I_s(d, n-d+2). \]

Similarly,
\[ I_s(d, n-d+i) - I_s(d, n-d+i+1) \]
\[ = \sum_{j=0}^{d-1} \left[ \binom{n+i+1}{j} x^j (1-x)^{n+i-j} - \binom{n+i}{j} x^j (1-x)^{n+j} \right]. \]

where
\[ \binom{n+i+1}{j} x^j (1-x)^{n+i-j} - \binom{n+i}{j} x^j (1-x)^{n+i-j} = \frac{(n+i)!}{j!(n+i-j)!} x^j (1-x)^{n+i-j} \cdot \frac{j-(n+i+1)x}{n+i+1-j}. \]

Because \( x \geq d/(n+1) > d/(n+i+1) \) (i.e. \( d-(n+i+1)x \leq 0 \)), so \( j-(n+i+1)x \leq 0 \) and \( I_s(d, n-d+i) < I_s(d, n-d+i+1) \).

The above theorem shows that in the mixed case of no effective signal and only noise in the sub-frequency band, when PSER is greater than or equal to \( d/N \),
\[ I_{F_{d,i}}(d, N-d) < I_{F_{d,i}}(d, N-d+1) < \cdots I_{F_{d,i}}(d, N-d+l) \cdots \] (45)

If \( I_{F_{d,i}}(d, N-d) \geq 1-\alpha \) and \( l \geq 1 \), then \( I_{F_{d,i}}(d, N-d+l) > 1-\alpha \) and the following inequality must be true.
\[
\sum_{l=0}^{\infty} \frac{\left(\frac{\alpha}{2}\right)^l}{l!} I_{R,\alpha=\alpha}(d, N+d) \geq \sum_{l=0}^{\infty} \frac{\left(\frac{\alpha}{2}\right)^l}{l!} (1-\alpha)
\]  
(46)

According \( I_{R,\alpha=\alpha}(d, N+d) \geq 1 - \alpha \), we can derive:
\[
F_{\alpha,\delta=0} \geq I^{-1}(1-\alpha;d, N-d). 
\]  
(47)

\( I^{-1}(1-\alpha;d, N-d) \) is the approximate quantile. This quantile has no relationship to the SNR, and is related to the number of power spectrum bins in a subband \( d \) and the number of bins in total power spectrum \( N \).

The necessary condition that the approximate quantile is valid is \( I^{-1}(1-\alpha;d, N-d) \geq d/N \). Therefore, \( I^{-1}(1-\alpha;d, N-d) \) should be calculated first, and then determine whether \( I^{-1}(1-\alpha;d, N-d) \) is greater than or equal to \( d/N \). If \( I^{-1}(1-\alpha;d, N-d) \geq d/N \), then \( I^{-1}(1-\alpha;d, N-d) \) is valid, otherwise \( I^{-1}(1-\alpha;d, N-d) \) is invalid.

For example, assuming \( d = 2 \), \( N = 128 \), when \( \alpha = 0.05 \), \( F_{0.05,\delta=0} = I^{-1}(0.95; 2, 126) = 0.036 \). Because \( d/N = 0.016 \), so \( F_{0.05,\delta=0} > d/N \), and 0.036 is valid. The means is that the probability of the PSER of a subband more than 3.6% is less than 5%.

When \( \alpha = 0.01 \), then \( F_{0.01,\delta=0} = 0.050 \). Because \( F_{0.01,\delta=0} > d/N \), so 0.050 is valid. The means is that the probability of the PSER of a subband more than 5.0% is less than 1%.

When \( \alpha = 0.5 \), then \( F_{0.5,\delta=0} = 0.013 \). Because \( F_{0.5,\delta=0} < d/N \), so 0.013 is invalid. The means is that the probability of the PSER of a subband more than 1.3% is not less than 50%. But in hypothesis testing, it doesn’t make sense to set \( \alpha \) to 0.5.

The approximate quantile is equivalent to the quantile of \( C_N(k) \) described in Equation(30), which suggests it is possible to indirectly determine whether there are valid signals through \( I^{-1}(1-\alpha;d, N-d) \).

6. Discussion

6.1 Rationale for Rectangular Window Function Selection

In spectral analysis, it is necessary to add a symmetric window function to a sampling sequence, in order to suppress side lobe interference and improve accuracy. This symmetric window function can be expressed as:
\[
w_N(k) = a - (1 - a) \cos(2\pi k / N), k = 0, 1, \ldots, N-1
\]  
(48)

A value of \( a = 1 \) corresponds to a rectangular window, \( a = 0.5 \) indicates a Hanning window, and \( a = 0.54 \) implies a Hamming window. In this study, a rectangular window was used to calculate the spectrum with Equation (2). Rectangular windows exhibit inherent limitations but, unlike other window functions, can ensure that spectral lines remain mutually independent after a discrete Fourier transform of the signal. The use of other window functions often produces a correlation between the power spectrum lines, which makes any subsequent calculation of probability distributions extremely difficult.

6.2 Significance of the approximate Quantile in the Case of Mixed Signals

In contrast, the quantile \( F_a \) for \( C_{S,N}(k) \), which is expressed in Equation(39), is difficult to be calculated directly. However, the quantile has important role in spectral signal detection, therefore solving \( F_a \) is the key to apply PSER in practice.
This paper identified an approximate quantile in mixed case with no effective signal, which was expressed by \( I^{-1}(1-\alpha; d, N-d) \). This approximate quantile has two advantages:

1) Easy to calculate. This quantile has no relationship to the noise variance, and is related to the number of power spectrum bins in a subband and the number of bins in total power spectrum. This means that the quantile can be calculated without knowing the noise variance and the power spectrum of the effective signal.

2) Consistent with the quantile in the pure noise case. This means that the approximate quantile can be used to determine whether the subband contains valid signals without knowing whether the power spectrum is in pure noise or mixed case. If the PSER is greater than \( I^{-1}(1-\alpha; d, N-d) \), it can be concluded that the subband is more likely to contain an effective signal.

This approximate quantile provides a methodology for determining whether power-spectrum contains a valid signal. However, a simple method to reduce the probability of an effective signal being misclassified as noise has not been found and further study is needed.

6.3 A special case of \( d=1 \)

In order to reduce the amount of data, the condition \( d \geq 2 \) is often required when PSER is used. When \( d \) is 1, it is a special case of PSER, and its statistical characteristics have many particularities.

In pure noise case, we can derive from Equation (25)

\[
C_N(k) \sim \beta(1, N-1).
\]

The expectation and variance are

\[
E(C_N(k)) = \frac{1}{N}, \quad \text{(50)}
\]

\[
Var(C_N(k)) = \frac{N-1}{N^2(N+1)}. \quad \text{(51)}
\]

The quantile is

\[
F_\alpha = I^{-1}(1-\alpha; 1, N-1).
\]

Through the following derivation, we can get

\[
1-\alpha = I_{F_\alpha}(1, N-1)
\]

\[
\begin{align*}
\int_0^{F_\alpha} (1-t)^{N-2} dt & = 1-\alpha \\
B(1, N-1) & = 1-\alpha \\
(\underbrace{N-1}) \frac{1-(1-F_\alpha)^{N-1}}{N-1} & = 1-\alpha \\
F_\alpha & = 1-\alpha^{\frac{1}{N-1}}.
\end{align*}
\]

When \( d \) is equal to 1, its quantile can be solved directly, but when \( d \) is greater than 1, its quantile can only be solved by numerical approximation.

In pure noise case, we can derive from Equation (36)

\[
C_{S,N}(k) \sim \beta_{2,2N-2}(\hat{\lambda}_k, \sum_{j=0}^{N-1} \lambda_j - \hat{\lambda}_k).
\]

When there is no effective signal in the power spectrum subband, the approximate quantile of the PSER is

\[
F_{\alpha, \hat{\lambda}_k=0} \geq 1-\alpha^{\frac{1}{N-1}}.
\]

The above properties can be summarized in the following table:

| \( d \geq 2 \) | \( d = 1 \) |
|----------------|------------------|
| Pure Noise     | Mixed            |
| Pure Noise     | Mixed            |
7. Conclusion

In this paper, a systematic investigation of the statistical characteristics for power-spectrum subband energy ratios was presented in the cases of pure noise and mixed signals. The statistical characteristics of the PSER provided a theoretical foundation for the use of PSER in practical applications. Results demonstrated the PSER of the Gaussian white noise and mixed signals followed beta and doubly non-central beta distributions, respectively. Although the probability distributions for the two cases are distinct, the upper percentiles are equivalent when the power spectral bins do not contain effective signals. As such, the proposed method could be a simple and easy way to detect the presence of effective signals.

References

1. ZL, H.; YC, L.; YH, C.; YH, W.; AA, C., Social-Aware Peer Discovery for Energy Harvesting-Based Device-to-Device Communications. Sensors-Basel 2019, 19, (10).
2. Pei-Han, Q.; Zan, L.; Jiang-Bo, S.; Rui, G., A robust power spectrum split cancellation-based spectrum sensing method for cognitive radio systems. Chinese Phys B 2014, 23, (12), 537-547.
3. Akram, J.; Eaton, D.W., A review and appraisal of arrival-time picking methods for downhole microseismic data. Geophysics 2016, 81, (2), KS71-KS91.
4. Islam, M.R.; Uddin, J.; Kim, J., Acoustic Emission Sensor Network Based Fault Diagnosis of Induction Motors Using a Gabor Filter and Multiclass Support Vector Machines. Ad Hoc Sens Wirel Net 2016, 34, (1), 273-287.
5. Mei, F.; Hu, C.; Li, P.; Zhang, J., Study on main Frequency precursor characteristics of Acoustic Emission from Deep buried Dali Rock explosion. Arabian Journal of Geosciences 2019, 12, (21).
6. Preisig, J.C.; Johnson, M.P., Signal detection for communications in the underwater acoustic environment. Icee J Oceanic Eng 2001, 26, (4), 572-585.
7. Wang, C.; Chang, X.; Liu, Y.; Chen, S., Mechanistic Characteristics of Double Dominant Frequencies of Acoustic Emission Signals in the Entire Fracture Process of Fine Sandstone. Energies 2019, 12, (20).
8. Xu, C.; Guan, J.; Bao, M.; Lu, J.; Ye, W., Pattern recognition based on enhanced multifeature parameters for vibration events in q-OTDR distributed optical fiber sensing system. Microwave & Optical Technology Letters 2017, 59, (12), 3134-3141.
9. Huang, G.; Law, A.W.; Guo, X., Taylor Dispersion of Contaminants by Dual-peak Spectral Random Waves. China Ocean Eng 2019, 33, (5), 537-543.
10. Yang, Y.; Zeng, Q.; Yin, G.; Wan, L., Vibration Test of Single Coal Gangue Particle Directly Impacting the Metal Plate and the Study of Coal Gangue Recognition Based on Vibration Signal and Stacking Integration. Icee Access 2019, 7, 106783-106804.
11. Lu, C.; Liu, Y.; Wang, H.; Liu, P., Microseismic signals of double-layer hard and thick igneous strata separation and fracturing. Int J Coal Geol 2016, 160, 28-41.
12. Kong, B.; Wang, E.; Li, Z.; Lu, W., Study on the Feature of Electromagnetic Radiation Under Coal Oxidation and Temperature Rise Based on Multifractal Theory. Fractals 2019, 27, (3).
13. Koopmans, L.H.; Press, A., The Spectral analysis of time series. 1974.
14. Groth, E.J., Probability distributions related to power spectra. In 1975; Vol. 29, pp 285-302.
15. Johnson, P.E.; Long, D.G., The probability density of spectral estimates based on modified periodogram averages. Ieee T Signal Proces 1999, 47, (5), P.1255-1261.
16. Berk, R.H., Continuous Univariate Distributions, Volume 2. Technometrics 1996, 38, (2), 189-189.
17. Gil, A.; Segura, J.; Temme, N.M., Efficient algorithms for the inversion of the cumulative central beta distribution. Numer Algorithms 2017, 74, (1), 77-91.
18. Temme, N.M., Asymptotic inversion of the incomplete beta function. J Comput Appl Math 1992, 41, (s 1–2), 145–157.
19. Johnson, N.L.; Kemp, A.W.; Kotz, S., Univariate Discrete Distributions, Third Edition. Wiley: 2005.
20. Abramowitz, M.; Stegun, I.A., Handbook of mathematical functions with formulas, graphs and mathematical tables / edited by Milton Abramowitz and Irene A. Stegun. US GPO: 1965.