LETTER TO THE EDITOR

Received 30 April 1971 15.2

Analysis of the Difference Between Log Mean and Mean Log Averaging

ROBERT L. HERSHEY

Department of Civil and Mechanical Engineering, Catholic University of America, Washington, D. C. 20017

The acoustic data analysis procedures at various installations differ in the intervals at which readings are converted from volts squared to decibels. There has been much speculation as to what would be the decibel difference between the results obtained for the same data, i.e., the difference between the log of the average and the average of the logs. In this analysis, the difference is treated as a random variable. By making the practical assumption that the distribution of the samples is Gaussian with mean μ and standard deviation σ, the expected value of the difference is found, as well as its appropriate confidence limits, for the special case of small σ/μ. These results are found to depend on the ratio σ/μ and on n, the number of samples averaged. It is found that the expected value of the difference is approximately equal to $2.17[(n-1)/n](\sigma/\mu)^2$ and that the standard deviation of the difference is equal to $[2/(n-1)]^{\frac{1}{2}}$ times this expected value. For practical values of σ/μ the expected value of the difference is found to be less than 0.2 dB and the Chebyshev 95% upper confidence limit of the difference is found to be less than 0.5 dB.

INTRODUCTION

In the types of acoustic analysis systems discussed here, each sample y_i is an average value from some fixed time interval of voltage squared. One system immediately converts each sample to decibels, while the second finds the average for a group of n samples from contiguous intervals and converts this average to decibels. It is well known that the average of the logs (first system) will always be less than or equal to the log of the average (second system). Thus, the decibel difference ϕ, as defined below, will always be positive:

$$\phi = 10\log\left(\frac{y_1+y_2+y_3+\cdots y_n}{n}\right) - \frac{10\log y_1 + 10\log y_2 + \cdots 10\log y_n}{n},$$

$$= 10\log\text{ (arithmetic mean)} - 10\log\text{ (geometric mean)}. \quad (1)$$

In general, greater spreads in values of the y_i yield larger values of ϕ and narrower spreads yield smaller values. For the case where all the y_i are identical, $\phi=0$.

The first quantitative treatment of the problem was an analysis of the upper limit of the difference by Cox.[1] His analysis showed that the worst case value of the difference increased as a function of K, the ratio of the highest reading to the lowest reading, increased. He found the upper limit of the difference to be independent of the number or readings, as long as all readings were at the upper and lower limits. His results showed that for moderate values of K the worst-case values of the difference can be several decibels.

Mitchell[2] analyzed the problem with a random variable approach, finding the means of the intensity and decibel distributions. He examined four different distributions, obtaining progressively smaller values of the difference as the distributions became narrower. His results are valid only for statistically large samples, where $n\rightarrow\infty$.

In a more recent paper Dyer[3] treated an associated log averaging problem with a random variable approach to transmission loss. He derived the statistics of various signal and noise combinations and showed how these statistics changed in transformation to decibels.

In the present analysis, we examine the small sample case for a Gaussian distribution of intensities. The mean of ϕ is derived and a set of practical confidence limits is determined.

I. MEAN VALUE OF THE DIFFERENCE

Assume that each sample y_i is itself an average of a large number of independent estimates of voltage squared from a stationary process. Then by the central limit theorem we can expect the distribution of y_i to approach Gaussian. We further assume each sample y_i to be independent of every other sample. In general, this will be true for large time-bandwidth products involved in practical systems.

For this Gaussian distribution of y_i, let the expected value be μ and the standard deviation be σ. We shall denote the ratio σ/μ by C. Our attention here is restricted to practical, moderately tight distributions for which $C\leq\frac{1}{4}$.

Next, we define the new random variables x_i as fluctuations about the expected value μ, of y_i:

$$\begin{aligned} x_i &= y_i - \mu, \qquad & i &= 1, \cdots, n, \\ E(x_i) &= 0, & \mathrm{Var}(x_i) &= \sigma^2. \end{aligned} \tag{2}$$

Then from Eqs. 1 and 2,

$$\phi = 10 \log\left(\frac{\mu + x_1 + \mu + x_2 + \cdots + \mu + x_n}{n}\right)$$
$$-\frac{10\log(\mu + x_1) + 10\log(\mu + x_2) + \cdots + 10\log(\mu + x_n)}{n}$$
$$= 10\log\left[1 + \frac{1}{\mu n}\sum_{i=1}^{n} x_i\right] - \frac{1}{n}\sum_{i=1}^{n} 10\log\left[\frac{1 + x_i}{\mu}\right]$$
$$= 4.34\left\{\ln\left[1 + \frac{1}{\mu n}\sum_{i=1}^{n} x_i\right] - \frac{1}{n}\sum_{i=1}^{n}\ln\left[1 + \frac{x_i}{\mu}\right]\right\}, \tag{3}$$

where ln denotes the natural logarithm.

Because of our assumption that $C \leq \frac{1}{4}$, the probability that $x_i/\mu \geq \frac{1}{2}$ is only 0.05. For these small values of x_i/μ, it is reasonable to approximate the natural log by the first two terms of the following series.

$$\ln(1+\xi) = \xi - \frac{\xi^2}{2} + \frac{\xi^3}{3} - \frac{\xi^4}{4} + \cdots. \tag{4}$$

Substituting the appropriate two-term expansions for each of the terms of Eq. 3, we obtain:

$$\phi \simeq \frac{4.34}{\mu n}\left(\sum_{i=1}^{n} x_i - \frac{1}{2\mu n}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j - \sum_{i=1}^{n} x_i + \frac{1}{2\mu}\sum_{i=1}^{n} x_i^2\right), \tag{5}$$

and from combining terms

$$\phi \simeq \frac{4.34}{2\mu^2 n}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j\right)$$
$$\simeq \frac{4.34}{2\mu^2 n}\sum_{i=1}^{n}(x_i - m)^2, \tag{6}$$

where $m = (1/n)\sum x_i$. It is apparent here that ϕ is proportional to a variance estimate.

The mean and variance of ϕ are most readily found by putting Eq. 6 in matrix form,

$$\phi = (2.17/\mu^2 n)\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}, \tag{7}$$

where $\mathbf{X}$ is the column vector of the x_is, $\mathbf{X}^{\mathrm{T}}$ is its transpose and $\mathbf{A}$ is the n by n matrix given by:

$$\mathbf{A} = \mathbf{I} - \frac{1}{n}\mathbf{U}\mathbf{U}^{\mathrm{T}}. \tag{8}$$

In the above form, $\mathbf{I}$ is the identity matrix having 1's on the main diagonal and zeroes elsewhere. With $\mathbf{U}$ as a column vector of 1's, $\mathbf{U}\mathbf{U}^{\mathrm{T}}$ is an n by n matrix of 1's.

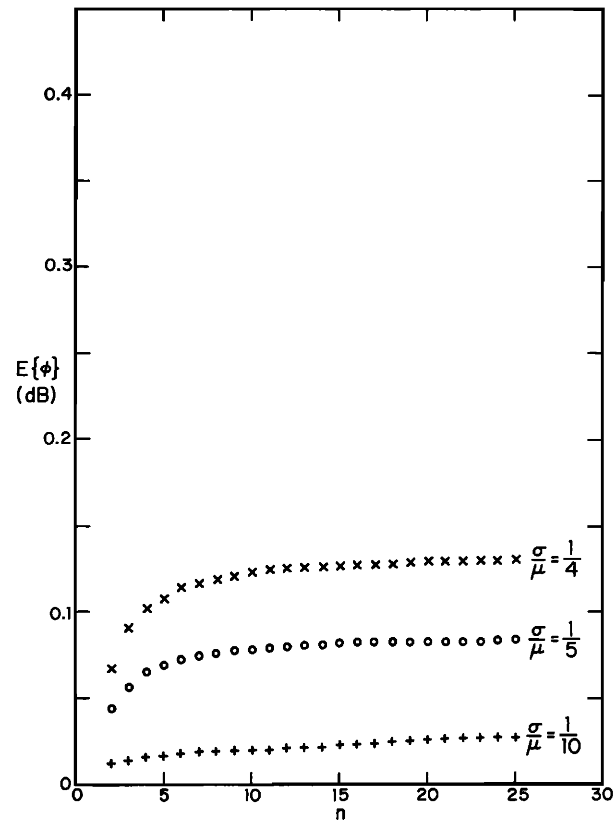

FIG. 1. Expected value of the difference $E\{\phi\}$ for various values of σ/μ with $2 \leq n \leq 25$.

Since all the x_i's are independent and have the same variance, the covariance matrix is simply:

$$\mathbf{V} = \sigma^2\mathbf{I}. \tag{9}$$

The expected value and variance of ϕ can now be found from the traces of appropriate matrices, i.e., the sums along their main diagonals.[4] By well-known results, we thus have:

$$E\{\phi\} \simeq \frac{2.17}{\mu^2 n}\,\mathrm{trace}\,\{\mathbf{A}\mathbf{V}\}. \tag{10}$$

Since

$$\mathbf{A}\mathbf{V} = \sigma^2[\mathbf{I} - (1/n)\mathbf{U}\mathbf{U}^{\mathrm{T}}], \tag{11}$$

then

$$E\{\phi\} \simeq 2.17\frac{(n-1)}{n}\frac{\sigma^2}{\mu^2}. \tag{12}$$

Thus, in terms of C,

$$E\{\phi\} \simeq 2.17\left(\frac{n-1}{n}\right)C^2. \tag{13}$$

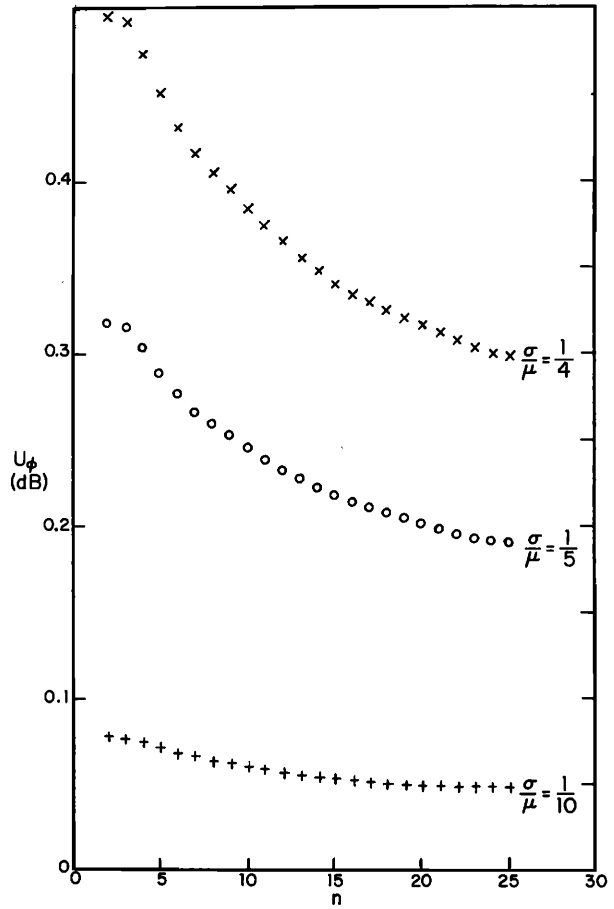

FIG. 2. Chebyshev 95% upper confidence limit of the difference U_ϕ for various values of σ/μ with $2 \leq n \leq 25$.

The expected value of the difference shown in Eq. 13 increases with the square of the σ/μ ratio. For smaller values of C, i.e., narrower distributions, the expected value of the difference approaches zero. For $C=0$, all the readings are the same and $E\{\phi\}=0$. The value of $E\{\phi\}$ from Eq. 13 is plotted in Fig. 1 for various values of C and n. Note that for large n

$$E\{\phi\} \simeq 2.17C^2. \qquad (14)$$

II. VARIANCE OF THE DIFFERENCE

Similarly the variance of ϕ is found from:

$$\sigma_\phi{}^2 = \left(\frac{2.17}{\mu^2 n}\right)^2 2 \text{ trace } \{(\mathbf{AV})^2\}. \qquad (15)$$

Since

$$(\mathbf{AV})^2 = \sigma^4[\mathbf{I}-(1/n)\mathbf{UU}^\mathrm{T}], \qquad (16)$$

then

$$\sigma_\phi{}^2 = \left(\frac{2.17}{\mu^2 n}\right)^2 2\sigma^4(n-1), \qquad (17)$$

or in terms of C

$$\sigma_\phi{}^2 = 2(2.17)^2 \frac{n-1}{n^2} C^4. \qquad (18)$$

Thus from Eqs. 12 and 18 we have

$$\sigma_\phi = [2/(n-1)]^{1/2} E\{\phi\} \qquad (19)$$

Although the mean and variance of ϕ have been found here directly, the probability density of ϕ is much more difficult to determine. In general, it is quite complicated, changing in shape with C and n. Equation 7 describes a "generalized χ^2 distribution"[5] whose characteristic function is easily found, but whose probability density can only be determined by numerical methods. It has been suggested that a simple closed form of the characteristic function is:

$$M_\phi(\theta) = E\{\exp(i\theta\phi)\}$$
$$= \prod_{j=1}^{n} E\{\exp(i\lambda_j \theta z_j{}^2)\}, \qquad (20)$$

where z_j is obtained from x_j by a linear transformation.

III. CONFIDENCE LIMITS

For most practical purposes the mean and variance will be sufficient, since an upper confidence limit on ϕ can be determined which is tight enough. Selecting 95% as the required level of confidence, we wish to find that value U_ϕ such that

$$P[\phi \leq U_\phi] \geq 0.95,$$
$$P[\phi > U_\phi] < 0.05.$$

We find the upper limit from Chebyshev's theorem[6]:

$$P[|\phi - E\{\phi\}| > \lambda\sigma_\phi] < 1/\lambda^2. \qquad (21)$$

From Eqs. 14 and 19, with $\lambda = (20)^{1/2}$, we obtain U_ϕ as a function of C:

$$U_\phi = E\{\phi\} + (20)^{1/2}\sigma_\phi$$
$$= [1+2(10)^{1/2}(n-1)^{-1/2}]E\{\phi\}$$
$$\simeq [1+6.32(n-1)^{-1/2}](2.17)\left(\frac{n-1}{n}\right)C^2. \qquad (22)$$

The values of U_ϕ are graphed in Fig. 2 for selected values of C.

IV. CONCLUSIONS

The expected value of the difference between the log of the average and the average of the logs will generally be less than 0.2 dB under the practical assumptions discussed here. The expected value and standard deviation of the difference are found to be proportional to the square of the σ/μ ratio. The Chebyshev 95%

upper confidence limit derived from the statistics of the difference is found to be less than 0.5 dB.

ACKNOWLEDGMENTS

The author wishes to thank Cdr. Henry Cox of the Office of the Secretary of Defense for suggesting the simplified matrix derivation of the result, and Dr. William D. Mark of Bolt Beranek and Newman, Inc., for offering various other suggestions. The support of this work by the Office of Naval Research is gratefully acknowledged.

[1] H. Cox, "Linear versus Logarithmic Averaging," J. Acoust. Soc. Amer. **39**, 688–690 (1966).

[2] S. Mitchell, "Comment on 'Linear versus Logarithmic Averaging,'" J. Acoust. Soc. Amer. **41**, 863–864 (1967).

[3] I. Dyer, "Statistics of Sound Propagation in the Ocean," J. Acoust. Soc. Amer. **48**, 337–345 (1970).

[4] H. Cox, "Interrelated Problems in Estimation and Detection I," NATO Advanced Study Inst. on Signal Processing with Emphasis on Underwater Acoustics (1968).

[5] D. Middleton, *An Introduction to Statistical Communication Theory* (McGraw–Hill, New York, 1960).

[6] H. Cramer, *Mathematical Methods of Statistics* (Princeton U. P., Princeton, N. J., 1946).