The Central Limit Theorem: Use with Caution

R. L. Fante, Fellow IEEE
The MITRE Corporation
Bedford, MA 01730

Abstract

The validity of the central limit theorem for the sum of N, K-distributed random phasors is investigated. It is demonstrated that the number of phasors that must be summed to obtain an amplitude distribution that can be approximated by a Rayleigh depends strongly on the underlying K-distribution, and is of order 200 for Weibull statistics.

It is quite often stated that when N independent, random phasors are summed (e.g., N clutter cells) the amplitude of the sum is approximately Rayleigh (i.e., in-phase and quadrature components are each gaussian) if N is sufficiently large. The question we examine here is how large is "sufficiently large"? In order to answer this, consider the normalized sum of N independent, random phasors

$$S = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} A_n e^{i\phi_n} \tag{1}$$

where all ϕ_n are independent and randomly distributed over $0 \le \phi_n \le 2\pi$, and the probability density of each of the amplitudes A_n is governed by a K-distribution:

$$p_A(A) = \frac{2b}{\Gamma(\nu)} \left(\frac{bA}{2} \right)^{\nu} K_{\nu-1}(bA) \tag{2}$$

where b is a constant that is related to the variance of A, K_ν is the modified Bessel function and Γ is the gamma function. The K-distribution in (2) is equal to the Weibull distribution for $\nu = 1/2$ approaches a Rayleigh as $\nu \to \infty$ and becomes log-normal-like (i.e., the pdf has a very long tail, and decays very slowly towards zero as $A \to \infty$) as $\nu \to 0$. Thus, it encompasses virtually all of the types of probability densities that are attributed to sea clutter, ground clutter and noise. It can be readily shown [1, 2] that when N K-distributed phasors are summed the amplitude $R = |S|$ also satisfies a K-distribution, given by

$$p_R(R) = C R^{\alpha} K_{\alpha-1}\left(bR\sqrt{N}\right) \tag{3}$$

where $\alpha = N\nu$ and $C = 2b(b/2)^{a}\left(\sqrt{N}\right)^{a+1}/\Gamma(a)$. From (2) we can readily show that $\sigma^2 = \langle R^2 \rangle = 4\nu/b^2$, where < > denotes an expectation.

One measure of how well Equation (2) approximates a Rayleigh distribution (which is the limit when $N \to \infty$) is to calculate the probability that R exceeds $\beta\sigma$, where β is an arbitrary multiplier, and then compare this with the Rayleigh result: $P_{ray}(R > \beta\sigma) = \exp(-\beta^2)$ From Equation (3) we have

$$P(R > \beta\sigma) = \frac{2\beta^{\alpha}\alpha^{\alpha/2}}{\Gamma(\alpha)} K_{\alpha}\left(2\beta\alpha^{1/2}\right) \tag{4}$$

where, on the right-hand side of (4), σ has been replaced by $4\nu^{1/2}/b$. As an example, we now choose $\beta = 3$ and examine how large N must be for $P(R > \beta\sigma)$ to be within 1 dB of the Rayleigh result. In Figure 1 we show the probability that $R > 3\sigma$ for different values of the order ν of the K-distribution satisfied by the phasor amplitude A_n. Also shown is the Rayleigh limit ($N \to \infty$). The error in using the central limit theorem for the sum of N phasors is the difference between the curve for each ν and the Rayleigh limit (-39 dB). Note that for a Weibull-distributed random variable ($\nu = 1/2$) the value of $P(R > 3\sigma)$ for $N = 10$ is more than 10 dB different from the Rayleigh limit. In fact, from Figure 1 (and additional results not shown) a good "rule of thumb" for the validity of the central limit theorem is that $(0 < \nu < 11)$

$$N \geq \frac{100}{\nu} \ . \tag{5}$$

Thus, for Weibull statistics ($\nu = 1/2$) we require $N \geq 200$. For more log-normal-like statistics (e.g., $\nu = 1/10$) we require N of order of thousands. Although we have chosen $\beta = 3$ the calculations are readily performed for other values of β.

The above analysis provides a measure of how many K-distributed random phasors must be summed together before the probability density of the amplitude of the sum can be approximated by a Rayleigh. Thus, if the clutter in a resolution cell of area A_c is K-distributed with shape parameter ν, one cannot use the Central Limit Theorem to argue that the probability density for the amplitude of the clutter from a larger resolution cell of area NA_c, where $N >> 1$, is Rayleigh unless N satisfies Equation (5).

References

1. E. Jakeman, "On the statistics of K-distributed noise", J. Physics A: Math. Gen., vol. 13, pp. 31-48, 1980.
2. R. Fante, "Detection of multiscatter targets in K-distributed clutter", IEEE Trans AP-32, pp. 1358-1363, 1984.

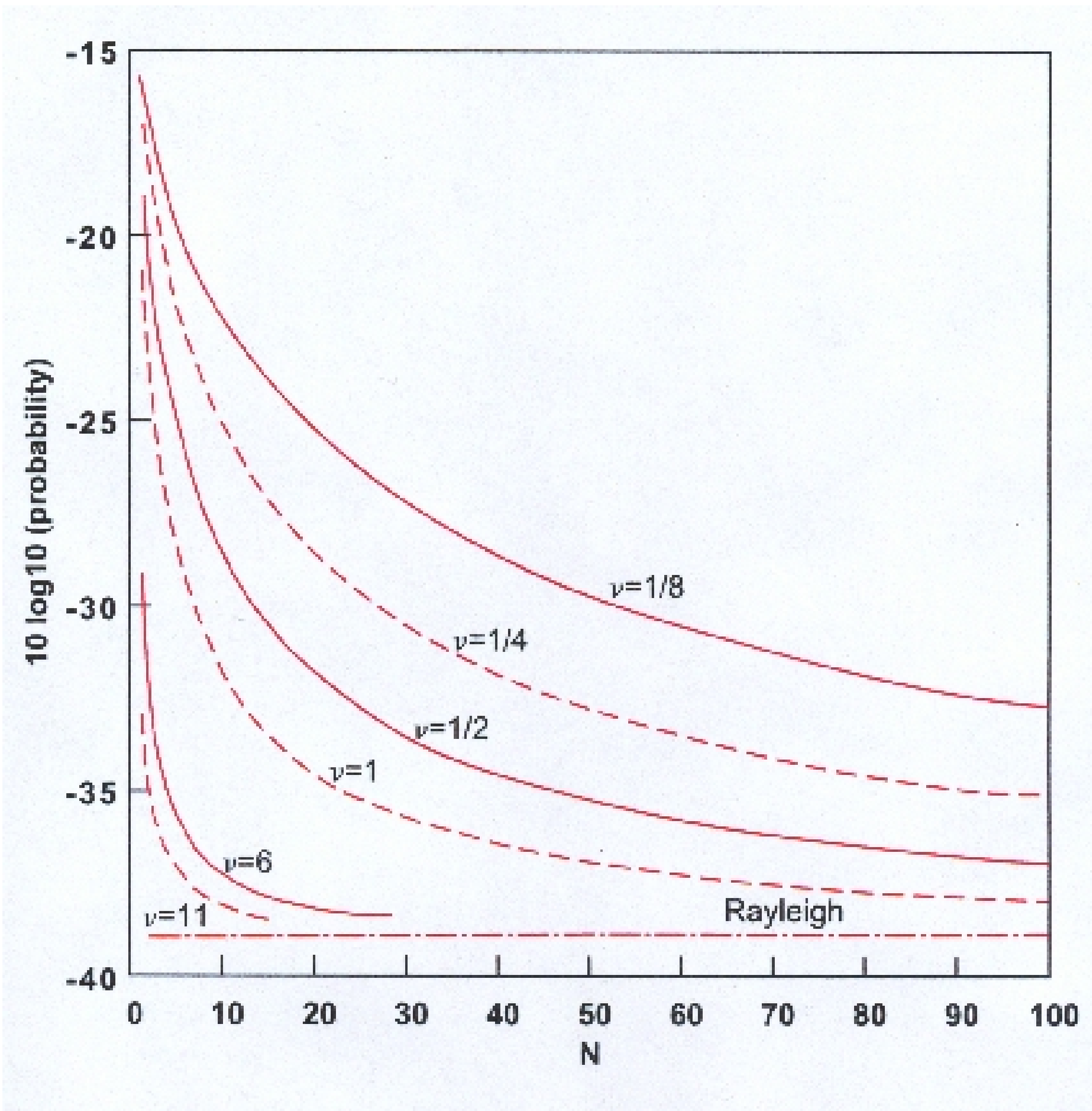

Figure 1. Probability that R > 3σ for Different K-Distributions

Response to Reviewers (AES-00-R037)

Thank you for your helpful comments. The Correspondence has been modified as follows:

1. The abstract has been expanded, and made more precise (all 4 reviewers requested this).
2. The index ν has been modified so that ν now runs from 0 to ∞ rather than -1 to ∞ (reviewer 3).
3. I now explain what log-normal like means (reviewer 4).
4. I now explain why the rhs of Equation (4) appears to be independent of σ (reviewer 4).
5. I have replaced "for convenience" by "As an example" (reviewer 4).
6. I no longer claim $N > 10$ is a so called "rule of thumb" (reviewers 3 & 4).
7. I have expanded the summary paragraph at then end of the Correspondence (reviewer 4).