XIV. PROCESSING AND TRANSMISSION OF INFORMATION*

Prof. F. C. Hennie III
Prof. R. M. Kennedy
Prof. C. E. Shannon
Prof. J. M. Wozencraft
Dr. R. E. Kahn
D. Chase
J. R. Colton
P. M. Ebert
D. D. Falconer
E. F. Ferretti
G. D. Forney, Jr.
C. J. Johnson
J. Max
R. F. McCann
C. W. Niessen
R. E. Olsen
R. Pilc
J. T. Pinkston III
R. N. Spann
W. R. Sutherland
M. G. Taylor

A. LOWER BOUNDS ON THE TAILS OF PROBABILITY DISTRIBUTIONS

Many problems in the Transmission of Information involve the distribution function of the sums of many random variables evaluated far from the mean. In these situations, a direct application of the Central Limit Theorem is virtually useless as an estimate of the distribution function. The so-called Chernov bound,[1] derived here in Eq. 13, turns out to be much more useful both as an upper bound and as an estimate on the far tails of the distribution function. We shall be primarily concerned, however, with deriving lower bounds and asymptotic estimates for the far tails of the distribution function of the sum of independent random variables. A number of the present results, particularly the asymptotic expressions, Eqs. 54 and 61, are due to C. E. Shannon.[2] They are reproduced because of their inaccessibility. The idea of the lower bound in Eq. 74 is also due to Shannon, although the result is stronger here in that it applies to nonidentically distributed variables. Another lower bound to the tail of a distribution has been given by Fano.[3] Fano's approach is to bound the multinomial coefficients for a sum of discrete finite random variables. Our results are more general than Fano's, since they are not restricted to discrete finite variables. On the other hand, in some situations, Fano's bound is tighter than our bounds.

Let ξ be a random variable with the distribution function $F(x) = P(\xi \leq x)$. We shall derive lower bounds to $1 - F(x)$ for $x > \bar{\xi}$, where $\bar{\xi}$ is the expectation of ξ. The bounds will be given in terms of the semi-invariant moment-generating function of ξ,

$$\mu(s) = \ln \int_{-\infty}^{\infty} \exp(sx)\, dF(x) = \ln \overline{\exp s\xi}, \tag{1}$$

in which the bar again denotes expectation.

The bounds will be useful primarily in situations for which ξ is the sum of a sequence of independent random variables, $\xi = \sum_{n=1}^{N} \xi_n$, where each ξ_n has a semi-invariant moment-generating function $\mu_n(s)$. Then

*This work was supported in part by the National Science Foundation (Grant GP-2495), the National Institutes of Health (Grant MH-04737-04), and the National Aeronautics and Space Administration (Grants NsG-334 and NsG-496).

$$\mu(s) = \ln \overline{\exp s \sum_{n=1}^{N} \xi_n} = \ln \overline{\prod_{n=1}^{N} \exp s\xi_n} \tag{2}$$

$$= \ln \prod_{n=1}^{N} \overline{\exp s\xi_n} = \sum_{n=1}^{N} \mu_n(s). \tag{3}$$

In going from Eq. 2 to Eq. 3 we have used the fact that for statistically independent random variables, the product of the averages is equal to the average of the product. Equation 3 will allow us to express bounds involving $\mu(s)$ in terms of the $\mu_n(s)$ without explicitly finding the distribution function $F(x)$. The semi-invariant moment-generating function exists for any random variable that takes on only a finite number of values and for any random variable ξ whose probability density drops off faster than exponentially as $\xi \to +\infty$ and as $\xi \to -\infty$. If the probability density drops off only exponentially, then $\mu(s)$ will exist only for a range of s. In the sequel, we assume an $F(x)$ for which $\mu(s)$ exists. If $\mu(s)$ exists only in a region, we consider only values of s in the interior of that region.

In order to find a lower bound to $1 - F(x)$, it is convenient to define a random variable ξ_s with the probability distribution function

$$F_s(x) = \frac{\int_{-\infty}^{x} \exp(sx')\, dF(x')}{\int_{-\infty}^{\infty} \exp(sx')\, dF(x')}. \tag{4}$$

The function $F_s(x)$ is generally called a tilted probability distribution, since it "tilts" the probability assigned by $F(x)$ by the factor e^{sx}. We now show that the mean and variance of the random variable ξ_s for a given s are given by the first and second derivatives of $\mu(s)$ evaluated at the same s. By direct differentiation of Eq. 1, we get

$$\mu'(s) = \frac{\int_{-\infty}^{\infty} x \exp(sx)\, dF(x)}{\int_{-\infty}^{\infty} \exp(sx')\, dF(x')} = \int_{-\infty}^{\infty} x\, dF_s(x) \tag{5}$$

$$\mu''(s) = \frac{\int_{-\infty}^{\infty} x^2 \exp(sx)\, dF(x)}{\int_{-\infty}^{\infty} \exp(sx')\, dF(x')} - [\mu'(s)]^2. \tag{6}$$

Thus

$$\mu'(s) = \overline{\xi_s}; \quad \mu''(s) = \overline{\xi_s^2} - \overline{\xi_s}^2. \tag{7}$$

Since $\mu''(s)$ is a variance and thus strictly positive for nontrivial distributions, $\bar{\xi}_s$ is an increasing function of s.

Using Eq. 4, we now have, for any given s,

$$dF_s(x) = \exp[-\mu(s)+sx]\, dF(x) \tag{8}$$

$$1 - F(x) = \int_{x'>x}^{\infty} dF(x') = \int_{x'>x}^{\infty} \exp[\mu(s)-sx']\, dF_s(x'). \tag{9}$$

It will now be instructive to find an upper bound to 1 - F(x) before proceeding to our major objective of lower-bounding 1 - F(x). For $s \geq 0$, we can upper-bound exp(-sx') in Eq. 9 by exp(-sx), and thus obtain

$$\begin{aligned} 1 - F(x) &\leq \exp[\mu(s)-sx] \int_{x'>x}^{\infty} dF_s(x') \\ &\leq \exp[\mu(s)-sx]\, [F_s(\infty)-F_s(x)] \\ &\leq \exp[\mu(s)-sx]; \quad s \geq 0. \end{aligned} \tag{10}$$

Since Eq. 10 is valid for any $s \geq 0$, we can get the best bound by minimizing $\mu(s) - sx$ with respect to s; if a solution exists for $s \geq 0$, it is

$$\mu'(s) = x. \tag{11}$$

Since $\mu''(s) \geq 0$, Eq. 11 does indeed minimize $\mu(s) - sx$. Finally, since $\mu'(s)$ is a continuous increasing function of s, we see that a solution will exist for s if

$$\bar{\xi} = \mu'(0) \leq x < \lim_{s\to\infty} \mu'(s). \tag{12}$$

Also, it can be seen from Eq. 5 that either $\lim_{s\to\infty} \mu'(s) = \infty$ or $\lim_{s\to\infty} \mu'(s)$ is the smallest x for which F(x) = 1, that is, the largest value taken on by the random variable ξ. Substituting Eq. 11 in 10, we get the well-known Chernov bound,[1] given in parametric form,

$$1 - F[\mu'(s)] \leq \exp[\mu(s)-s\mu'(s)] \qquad s \geq 0. \tag{13}$$

The exponent in Eq. 13, $\mu(s) - s\mu'(s)$, is zero for s = 0 and has a derivative of $-s\mu''(s)$ with respect to s. Thus for nontrivial distributions, the exponent is negative for s > 0. Figure XIV-1 gives a graphical interpretation of the terms in Eq. 13 for a typical random variable.

If we substitute Eq. 3 in Eq. 13, we obtain

$$1 - F\left[\sum_{n=1}^{N} \mu_n'(s)\right] \leq \exp \sum_{n=1}^{N} [\mu_n(s)-s\mu_n'(s)]. \tag{14}$$

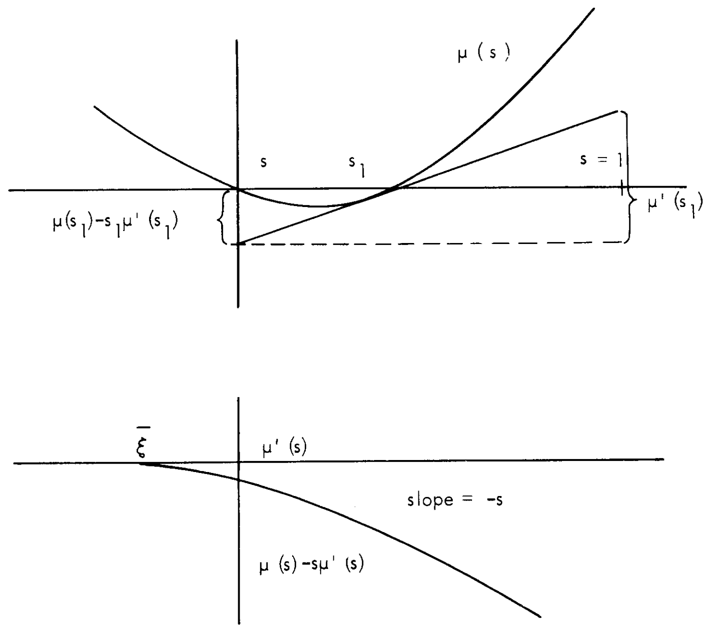

Fig. XIV-1. Graphical interpretations of Eq. 13.

If the ξ_n are all identically distributed, then the argument on the left and the exponent on the right are linear in N.

We next turn our attention to finding a lower bound to 1 - F(x). Since x is arbitrary in (9), let us substitute $\mu'(s) - A$ for x, where A is an arbitrary positive number to be chosen later.

$$1 - F[\mu'(s)-A] = \int_{x'>\mu'(s)-A}^{\infty} \exp[\mu(s)-sx'] \, dF_s(x')$$

$$\geq \int_{-A<x-\mu'(s)<A} \exp[\mu(s)-sx] \, dF_s(x). \tag{15}$$

Here, we have lower-bounded the left side by reducing the interval of integration. Restricting s to be non-negative, we observe that exp(-sx) is decreasing with x and is lower-bounded in (15) by $\exp[-s\mu'(s)-sA]$.

$$1 - F[\mu'(s)-A] \geq \exp[\mu(s)-s\mu'(s)-sA] \int dF_s(x) \qquad -A < x - \mu'(s) < A. \tag{16}$$

Recalling that ξ_s has a mean $\mu'(s)$ and a variance $\mu''(s)$, we can lower-bound the integral in (16) by the Chebyshev inequality,

$$\int dF_s(x) \geq 1 - \frac{\mu''(s)}{A^2} \qquad -A < x - \mu'(s) < A. \tag{17}$$

Choosing $A = \sqrt{2\mu''(s)}$, for simplicity, and substituting (17) in (16), we get

$$1 - F\left[\mu'(s)-\sqrt{2\mu''(s)}\right] \geq \frac{1}{2}\exp\left[\mu(s)-s\mu'(s)-s\sqrt{2\mu''(s)}\right]. \tag{18}$$

It is convenient to simplify the left-hand side of (18) at the expense of the right-hand side. Define s_1 to satisfy

$$\mu'(s_1) = \mu'(s) - \sqrt{2\mu''(s)}. \tag{19}$$

Expanding $\mu(s)$ in a Taylor expansion around $\mu(s_1)$, we get

$$\mu(s) = \mu(s_1) + (s-s_1)\mu'(s_1) + \frac{(s-s_1)^2}{2}\mu''(r); \qquad s_1 \leq r \leq s$$

$$\mu(s) \geq \mu(s_1) + (s-s_1)\mu'(s_1). \tag{20}$$

Substituting (19) and (20) in (18), we have

$$1 - F[\mu'(s_1)] \geq \frac{1}{2}\exp\left[\mu(s_1)-s_1\mu'(s_1)-2s\sqrt{2\mu''(s)}\right], \tag{21}$$

where $s \geq 0$, and s_1 is related to s through Eq. 19. Observe that Eqs. 13 and 21 are quite closely related. They differ primarily in the term $2s\sqrt{2\mu''(s)}$. When ξ is the sum of independent random variables, we see from Eq. 3 that $\sqrt{2\mu''(s)}$ is proportional to the square root of the number of random variables, whereas $\mu(s_1)$ and $s_1\mu'(s_1)$ are directly proportional to the number of variables. Thus, in some sense, $2s\sqrt{2\mu''(s)}$ should be unimportant for large N. Unfortunately, giving a precise meaning to this is somewhat involved as the next theorem illustrates.

THEOREM 1: Let $\xi_1, \xi_2, \ldots$ be an infinite sequence of random variables with semi-invariant moment-generating functions $\mu_1(s), \mu_2(s), \ldots$. For any positive number A assume that positive numbers L(A) and U(A) exist such that

$$L(A) \leq \frac{1}{N}\sum_{n=1}^{N} \mu_n''(s) \leq U(A); \qquad \text{for all } N \geq 1 \text{ and all } s,\ 0 \leq s \leq A. \tag{22}$$

Then for any $s_1 \geq 0$,

$$\lim_{N\to\infty} \frac{1}{N}\ln\left\{1-F[\mu'(s_1)]\right\} = \lim_{N\to\infty} \frac{[\mu(s_1)-s_1\mu'(s_1)]}{N}, \tag{23}$$

where $\xi = \sum_{n=1}^{N} \xi_n$, and F(x) and $\mu(s)$ are the distribution function and semi-invariant moment-generating function of ξ.

DISCUSSION: The condition in Eq. 22 is broad enough to cover situations in which each of the ξ_n has one of a finite set of distribution functions, each with nonzero variance and a semi-invariant moment-generating function. The number A is brought in to avoid ruling out the broad class of random variables for which $F_n(x) = 1$ for some finite x and thus $\lim_{s\to\infty} \mu_n''(s) = 0$.

PROOF: It follows immediately from (14) that the left side of (23) is less than or equal to the right side. Also, from (21), for any given N, we have

$$\frac{1}{N} \ln \{1-F[\mu'(s_1)]\} \geq \frac{\mu(s_1) - s_1\mu'(s_1)}{N} - \frac{\ln 2}{N} - \frac{2s\sqrt{2\mu''(s)}}{N}. \tag{24}$$

In Eq. 24 s and s_1 are related by Eq. 19. Also, from the mean value theorem,

$$\mu'(s_1) = \mu'(s) + (s_1-s)\mu''(s_2); \quad \text{for some } s_2,\ s_1 \leq s_2 \leq s. \tag{25}$$

Combining Eqs. 19 and 25, we have

$$s - s_1 = \frac{\sqrt{2\mu''(s)}}{\mu''(s_2)}. \tag{26}$$

For any given $A > 0$ and all $s \leq A$, we can upper-bound $\mu''(s)$ and lower-bound $\mu''(s_2)$ by Eq. 22.

$$s - s_1 \leq \frac{\sqrt{2NU(A)}}{NL(A)} = \frac{1}{\sqrt{N}} \frac{\sqrt{2U(A)}}{L(A)}. \tag{27}$$

Next, let ϵ be an arbitrary positive number and restrict N to satisfy

$$N \geq \frac{2U(A)}{[L(A)]^2\epsilon^2}. \tag{28}$$

From Eqs. 27 and 28,

$$0 \leq s - s_1 \leq \epsilon. \tag{29}$$

Since s_1 is a continuous function of s, s_1 will take on all values from 0 to $A - \epsilon$ as s goes from 0 to A. Thus for any s_1, $0 \leq s_1 \leq A - \epsilon$, and for any N satisfying (28), we have $s \leq A$, and therefore from Eq. 22

$$\frac{2s\sqrt{2\mu''(s)}}{N} \leq 2s\sqrt{\frac{2U(A)}{N}} \xrightarrow[N\to\infty]{} 0. \tag{30}$$

But, since A and ϵ are arbitrary, Eq. 30 is valid for any s_1. Thus, from (24),

$$\lim_{N\to\infty} \frac{1}{N} \ln \{1-F[\mu'(s_1)]\} \geq \lim_{N\to\infty} \frac{\mu(s_1) - s_1\mu'(s_1)}{N}, \tag{31}$$

thereby completing the proof.

It is frequently convenient to have a specific upper bound on $2s\sqrt{2\mu''(s)}$, since it is the essential difference between our upper and lower bounds. The following result is quite crude and applies only to discrete random variables taking on a maximum value. Let the random variable ξ_n take on the values $x_{n1} \geq x_{n2} \geq x_{n3} \geq \ldots$ with probabilities $p_{n1}, p_{n2} \ldots$. From Eq. 6 we know that $\mu_n''(s)$ is the variance of a random variable taking on the values $x_{n1}, x_{n2}, \ldots$ with probabilities $p_{n1}e^{sx_{n1}} \Big/ \sum_i p_{ni}e^{sx_{ni}}, \ldots$. Since the variance of a random variable is upper-bounded by the second moment around any value, we have

$$\mu_n''(s) \leq \sum_k (x_{nk}-x_{n1})^2 \frac{p_{nk}e^{sx_{nk}}}{\sum_i p_{ni}e^{sx_{ni}}}. \tag{32}$$

Multiplying numerator and denominator by $e^{-sx_{n1}}$ and defining η_{nk} by $s(x_{nk}-x_{n1})$, we have

$$s^2\mu_n''(s) \leq \sum_k \eta_{nk}^2 \frac{p_{nk}e^{\eta_{nk}}}{\sum_i p_{ni}e^{\eta_{ni}}}. \tag{33}$$

Now, $\eta_{n1} = 0$ and $\eta_{nk} \leq 0$ for $s \geq 0$. Thus the denominator in (33) can be lower-bounded by p_{n1}. Furthermore, $\eta_{nk}^2 e^{\eta_{nk}} \leq (2/e)^2$ for any $\eta_{nk} \leq 0$. Incorporating these results in (33) yields

$$s^2\mu_n''(s) \leq (2/e)^2\left[\frac{1-p_{n1}}{p_{n1}}\right], \tag{34}$$

where p_{n1} is the probability of the largest value taken on by ξ_n. Let $p_1 = \min_n p_{n1}$, and use Eq. 3, then, we get

$$s^2\mu''(s) \leq N(2/e)^2\left[\frac{1-p_1}{p_1}\right]. \tag{35}$$

Substituting (35) in (21), we get

$$1 - F[\mu'(s_1)] \geq \frac{1}{2}\exp\left[\mu(s_1) - s_1\mu'(s_1) - \frac{4}{e}\sqrt{\frac{2N(1-p_1)}{p_1}}\right]. \tag{36}$$

The previous results have all been derived through the use of the Chebyshev inequality and are characteristically simple and general but weak. We now turn to the use of the Central Limit Theorem to get tighter results. As before, the random variable ξ is the sum of N independent random variables, $\xi = \sum_{n=1}^{N} \xi_n$. The ξ_n have the distribution functions $F_n(x)$ and the semi-invariant moment-generating functions $\mu_n(s)$. Now, let us define the tilted random variables $\xi_{n,s}$ with the distribution functions

$$F_{n,s}(x) = \frac{\int_{-\infty}^{x} \exp(sx')\, dF_n(x')}{\int_{-\infty}^{\infty} \exp(sx')\, dF_n(x')}. \tag{37}$$

The semi-invariant moment-generating function of $\xi_{n,s}$ is

$$h_{n,s}(r) = \int_{-\infty}^{\infty} \exp(rx)\, dF_{n,s}(x) \tag{38}$$

$$= \mu_n(s+r) - \mu_n(s), \tag{39}$$

where Eq. 39 follows from substituting (37) in (38).

If ξ_s is now defined as $\xi_s = \sum_{n=1}^{N} \xi_{n,s}$, the semi-invariant moment-generating function of ξ_s is

$$h_s(r) = \sum_{n=1}^{N} h_{n,s}(r) = \mu(s+r) - \mu(s). \tag{40}$$

Thus, if we work backwards, the distribution function of ξ_s is given by Eq. 4.

Now, let $x = \mu'(s)$ in Eq. 9,

$$1 - F[\mu'(s)] = \exp[\mu(s)-s\mu'(s)] \int_{x>\mu'(s)} \exp\{s[\mu'(s)-x]\}\, dF_s(x). \tag{41}$$

We shall assume, temporarily, that the ξ_n are all identically distributed and non-lattice. (A lattice distribution is a distribution in which the allowable values of ξ_n can be written in the form $x_k = hk + a$, where h and a are arbitrary numbers independent of the integer k. The largest h for which the allowable values of ξ_n can be expressed in this way is called the span of the distribution.) Then ξ_s has the mean $\mu'(s) = N\mu_n'(s)$ and the variance $\mu''(s) = N\mu_n''(s)$, and for $s > 0$, the terms in the integral of (41) have the appearance shown in Fig. XIV-2. Observe that $F_s(x)$ is approximately a Gaussian distribution function, but the exponential term is changing much more rapidly than $F_s(x)$ for large N. Let η_s be the normalized random variable.

$$\eta_s = \frac{\xi_s - \mu'(s)}{\sqrt{\mu''(s)}}. \tag{42}$$

Let $G_s(y)$ be the distribution function of η_s,

$$G_s(y) = F_s[y\sqrt{\mu''(s)}+\mu(s)]. \tag{43}$$

Transforming the variable of integration in (41) by (42), we have

$$1 - F[\mu'(s)] = \exp[\mu(s)-s\mu'(s)] \int_{z>0} \exp[-s\sqrt{\mu''(s)}\ z]\ dG(z). \tag{44}$$

Assuming $s > 0$, we can use integration by parts on the integral in (44), to obtain

$$\int_{z>0} \exp[-s\sqrt{\mu''(s)}\ z]\ dG(z) = s\sqrt{\mu''(s)} \int_{z=0}^{\infty} [G(z)-G(0)]\exp[-s\sqrt{\mu''(s)}\ z]\ dz. \tag{45}$$

Equation 45 is now in a form suitable for application to the central limit theorem. Since the exponential term is decaying so rapidly, we need a particularly strong central

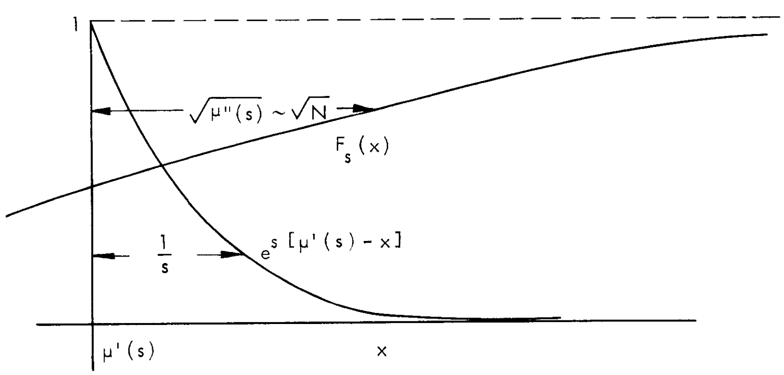

Fig. XIV-2. Sketch of terms in Eq. 41.

limit theorem. The appropriate theorem is due to Esseen[4] and is given by Gnedenko and Kolmogoroff.[5] Under conditions less restrictive than those that we have already assumed, the theorem states

$$G(z) = \Phi(z) + \frac{Q_1(z)\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi N}} + o\left(\frac{1}{\sqrt{N}}\right) \tag{46}$$

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \tag{47}$$

$$Q_1(z) = Q_1(0)\ (1-z^2) \tag{48}$$

$$Q_1(0) = \frac{\mu_n'''(s)}{6[\mu_n''(s)]^{3/2}} \tag{49}$$

and $o\left(\frac{1}{\sqrt{N}}\right)$ is a function approaching 0, uniformly in z, faster than $\frac{1}{\sqrt{N}}$; that is, $f(N) = o\left(\frac{1}{\sqrt{N}}\right)$ if $\lim_{N\to\infty} \sqrt{N}\ f(N) = 0$.

Substituting (46) and (48) in (45), we have

$$\int_{z>0} \exp[-s\sqrt{\mu''(s)}\ z]\ dG(z)$$

$$= s\sqrt{\mu''(s)} \int_{z=0}^{\infty} [\Phi(z)-\Phi(0)] \exp[-s\sqrt{\mu''(s)}\ z]\ dz$$

$$+ \frac{s\sqrt{\mu''(s)}\ Q_1(0)}{\sqrt{N}} \int_0^{\infty} \left[1 - \exp\left(-\frac{z^2}{2}\right)\right] \exp[-s\sqrt{\mu''(s)}\ z]\ dz$$

$$+ \frac{s\sqrt{\mu''(s)}\ Q_1(0)}{\sqrt{N}} \int_0^{\infty} z^2 \exp\left(-\frac{z^2}{2}\right) \exp[-s\sqrt{\mu''(s)}\ z]\ dz$$

$$+ s\sqrt{\mu''(s)} \int_0^{\infty} o\left(\frac{1}{\sqrt{N}}\right) \exp[-s\sqrt{\mu''(s)}\ z]\ dz. \tag{50}$$

The first integral on the right-hand side of (50) can be integrated by parts, and then by completing the square in the exponent. This yields

$$s\sqrt{\mu''(s)} \int_0^{\infty} [\Phi(z)-\Phi(0)] \exp[-s\sqrt{\mu''(s)}\ z]\ dz = \exp\left[\frac{s^2\mu''(s)}{2}\right] [1-\Phi(s\sqrt{\mu''(s)})] \tag{51}$$

Using standard inequalities on the normal distribution function (see Feller[6]), we obtain

$$\left[1 - \frac{1}{s^2\mu''(s)}\right] \frac{\exp\left[-\frac{s^2\mu''(s)}{2}\right]}{\sqrt{2\pi s^2\mu''(s)}} \leq 1 - \Phi(s\sqrt{\mu''(s)}) \leq \frac{\exp\left[-\frac{s^2\mu''(s)}{2}\right]}{\sqrt{2\pi s^2\mu''(s)}}. \tag{52}$$

Recalling that $\mu''(s) = N\mu_n''(s)$, we see that the first integral in (50) is equal to $\frac{1}{\sqrt{2\pi N s^2 \mu_n''(s)}} + o\left(\frac{1}{\sqrt{N}}\right)$.

A similar integration on the second integral in (50) shows that it can be represented by $o(1/\sqrt{N})$. If we upper-bound $\exp[-z^2/2]$ in the third integral of (50) by 1, it follows that it also is $o(1/\sqrt{N})$. By using the uniform convergence in z of the $o(1/\sqrt{N})$ in the fourth integral, we see that it too is $o(1/\sqrt{N})$. Thus,

$$\int_{z>0}^{\infty} \exp[-s\sqrt{\mu''(s)}\ z]\, dG(z) = \frac{1}{\sqrt{2\pi N s^2 \mu_n''(s)}} + o\left(\frac{1}{\sqrt{N}}\right). \tag{53}$$

Substituting (53) in (44), we see that for identically distributed nonlattice variables with $s > 0$,

$$1 - F[\mu'(s)] = \left[\frac{1}{\sqrt{2\pi N s^2 \mu_n''(s)}} + o\left(\frac{1}{\sqrt{N}}\right)\right] \exp[\mu(s)-s\mu'(s)]. \tag{54}$$

We shall now derive a relationship similar to Eq. 54 for the lattice case. Let the ξ_n be independent and identically distributed and take on only the values

$$x_k = a + hk, \tag{55}$$

where a and h are arbitrary numbers independent of the integer k, and one is the greatest common divisor of the integers k for which x_k has nonzero probability. The random variables ξ and ξ_s also can take on only the values Na + hk for integer k. Let Δ be the magnitude of the difference between $\mu'(s)$ and the smallest value of Na + hk larger than $\mu'(s)$. Define

$$p_s(j) = P(\xi_s=\mu'(s)+\Delta+hj). \tag{56}$$

We can now apply a central limit theorem for lattice distributions (see Gnedenko and Kolmogoroff[7]). This theorem states, in effect, that for any $\epsilon > 0$, there exists an N_o such that for $N \geq N_o$, we have

$$\left| p_s(j) - \frac{h}{\sqrt{2\pi\mu''(s)}} \exp\left[\frac{-(\Delta+jh)^2}{2\mu''(s)}\right] \right| \leq \frac{\epsilon}{\sqrt{\mu''(s)}}. \tag{57}$$

Bounding the exponential term, we can rewrite this as

$$-\frac{\epsilon}{\sqrt{\mu''(s)}} + \frac{h}{\sqrt{2\pi\mu''(s)}}\left[1 - \frac{(\Delta+jh)^2}{2\mu''(s)}\right] \leq p_s(j) \leq \frac{\epsilon}{\sqrt{\mu''(s)}} + \frac{h}{\sqrt{2\pi\mu''(s)}}. \tag{58}$$

Equation 41 can now be rewritten in terms of the $p_s(j)$ to yield

$$1 - F[\mu'(s)] = \exp[\mu(s)-s\mu'(s)] \sum_{j=0}^{\infty} p_s(j) \exp[-s(\Delta+jh)] \tag{59}$$

$$1 - F[\mu'(s)] \leq \exp[\mu(s)-s\mu'(s)]\left[\frac{\epsilon}{\sqrt{\mu''(s)}} + \frac{h}{\sqrt{2\pi\mu''(s)}}\right]\frac{\exp(-s\Delta)}{1-\exp(-sh)}. \tag{60}$$

Here, we have upper-bounded (59) by (58) and summed over j.

Equation 58 can also be used to lower-bound (59). If we define

$$A = \sum_{j=0}^{\infty} (\Delta+jh)^2 \exp[-s(\Delta+jh)],$$

then

$$1 - F[\mu'(s)] \geq \frac{\exp[\mu(s)-s\mu'(s)]}{\sqrt{2\pi\mu''(s)}}\left[(h-\epsilon\sqrt{2\pi})\frac{\exp(-s\Delta)}{1-\exp(-sh)} - \frac{hA}{\mu''(s)}\right].$$

Observing that ϵ can be made to approach 0 with increasing N and that A is bounded independently of N, we have

$$1 - F[\mu'(s)] = \left[\frac{h\exp(-s\Delta)}{2\pi N\mu_n''(s)[1-\exp(-sh)]} + o\left(\frac{1}{\sqrt{N}}\right)\right]\exp[\mu(s)-s\mu'(s)]. \tag{61}$$

Equation 61 is valid for any $s > 0$ for independent, identically distributed lattice variables if $\mu_n(s)$ exists. Note, however, that Δ will fluctuate between 0 and h as a function of N.

Equations 54 and 61 are not applicable in general to nonidentically distributed random variables. In some cases, however, Eqs. 54 and 61 can be made to apply, first, by grouping the variables to make them identically distributed. For example, for N variables, if N/2 variables have one distribution and N/2 have another distribution, then we can form N/2 identically distributed variables, each of which is the sum of a pair of the original variables.

In the sequel, we shall take a different approach to nonidentically distributed variables and derive a lower bound for $1 - F[\mu'(s)]$ by using a different form of the central limit theorem. This new result will be more complicated than (54) and (61), but will have the advantage of providing a firm lower bound to $1 - F[\mu'(s)]$ and of being applicable to nonidentically distributed variables. It will only be stronger than Eq. 21 for large N. We start with Eqs. 44 and 45, which are still valid for nonidentically distributed independent variables. Then the Berry theorem[8] states

$$|G(z)-\Phi(z)| \leq \frac{C\rho_{3,N}}{\sqrt{N}}, \tag{62}$$

where

$$\rho_{3,N} = \frac{\frac{1}{N}\sum_{n=1}^{N} \beta_{3,n}}{\left[\frac{1}{N}\sum_{n=1}^{N} \mu_n''(s)\right]^{3/2}} \tag{63}$$

$$\beta_{3,n} = \int_{-\infty}^{\infty} |x-\mu_n'(s)|^3 \, dF_{s,n}(x) \tag{64}$$

and C is a constant. Esseen[4] has shown that C may be taken to be 7.5, but no example has ever been found in which C need be larger than 0.41. The constant must be at least 0.41 to cope with a sum of binary random variables, each of which takes on one value with the probability $(\sqrt{10}-2)/2$.

From Eq. 62, and from the fact that G(z) is a distribution function, we have

$$G(z) - G(0) \geq \Phi(z) - \Phi(0) - \frac{2C\rho_{3,N}}{\sqrt{N}} \tag{65}$$

$$\geq 0; \quad z \geq 0. \tag{66}$$

Let z_o be that value of z for which the right-hand side of (65) is 0.

$$\Phi(z_o) = \frac{1}{2} + \frac{2C\rho_{3,N}}{\sqrt{N}}. \tag{67}$$

Observe that if N is not sufficiently large, (67) will have no solution. More precisely, N must be greater than $[4C\rho_{3,N}]^2$ for Eq. 67 to have a solution. For smaller values of N, we must use Eq. 21 to lower-bound $1 - F[\mu'(s)]$. Because of the importance of $\rho_{3,N}$ here, it is sometimes convenient to have a bound on $\rho_{3,N}$ in terms of $\mu(s)$. Using the theorem of the means, we have

$$\beta_{3,N} \leq \left[\int_{-\infty}^{\infty} [x-\mu_n'(s)]^4 \, dF_{n,s}(x)\right]^{3/4} = \{\mu_n''''(s)+3[\mu_n''(s)]^2\}^{3/4} \tag{68}$$

$$\rho_{3,N} \leq \frac{\frac{1}{N}\sum_{n=1}^{N} \{\mu_n''''(s)+3[\mu_n''(s)]^2\}^{3/4}}{\left[\frac{1}{N}\sum_{n=1}^{N} \mu_n''(s)\right]^{3/2}}. \tag{69}$$

Using Eq. 66 for $z < z_o$, we find that Eq. 45 becomes

$$\int_{z>0} \exp[-s\sqrt{\mu''(s)}\, z] \, dG(z) \geq s\sqrt{\mu''(s)} \int_{z_o}^{\infty} \left[\Phi(z) - \Phi(0) - \frac{2C\rho_{3,N}}{\sqrt{N}}\right] \exp[-s\sqrt{\mu''(s)}\, z] \, dz. \tag{70}$$

Integrating by parts and using Eq. 67, we get

$$\int_{z>0} \exp[-s\sqrt{\mu''(s)}\, z]\, dG(z) \geq \int_{z_o}^{\infty} \exp[-s\sqrt{\mu''(s)}\, z]\, d\Phi(z) \tag{71}$$

$$\geq \exp\left[\frac{s^2\mu''(s)}{2}\right] 1 - \Phi(z_o+s\sqrt{\mu''(s)}) \tag{72}$$

$$\geq \frac{\exp\left[-z_o s\sqrt{\mu''(s)} - \frac{z_o^2}{2}\right]}{\sqrt{2\pi}\,[z_o+s\sqrt{\mu''(s)}\,]} \left\{1 - \frac{1}{[z_o+s\sqrt{\mu''(s)}\,]^2}\right\}. \tag{73}$$

Equation 71 was integrated by completing the square in the exponent, and in Eq. 73 we used the bound on $\Phi(x)$ given by Eq. 52. If we define B as the right-hand side of Eq. 73, then from Eq. 44, we have

$$1 - F[\mu'(s)] \geq B \exp[\mu(s)-s\mu'(s)]. \tag{74}$$

It is instructive to estimate B for very large N, under the assumption that $\mu''(s)$ grows in some sense linearly with N. Under these circumstances, from Eq. 67,

$$z_o \approx \frac{2C\rho_{3,N}}{\sqrt{2\pi N}} \tag{75}$$

$$B \approx \frac{\exp\left[-2Cs\rho_{3,N}\sqrt{\sum_{n=1}^{N} \frac{\mu_n''(s)}{N}}\right]}{\sqrt{2\pi s^2 \sum_{n=1}^{N} \mu_n''(s)}}. \tag{76}$$

We see that for large N, Eqs. 54 and 74 differ by the numerator of Eq. 76. This term is essentially independent of N, but is typically very small relative to 1.

All of the results thus far are concerned with the upper tail of a distribution function, $1 - F(x)$, for $x > \bar{\xi}$. We can apply all of these results to the lower tail of a distribution, $F(x)$, for $x < \bar{\xi}$, simply by considering the random variable $-\xi$ rather than ξ. Since the semi-invariant moment-generating function of $-\xi$ is related to that of ξ through a change in sign of s, we can write the results immediately in terms of $\mu(s)$ for $s < 0$. Equation 13 becomes

$$F[\mu'(s)] \leq \exp[\mu(s)-s\mu'(s)]. \tag{77}$$

(Actually $F[\mu'(s)] = P[\xi \leq \mu'(s)]$, whereas the counterpart of Eq. 13 treats $P[\xi < \mu'(s)]$. A trivial modification of Eqs. 9-13 establishes the stronger result.)

Upon recognizing that $F[\mu'(s)] \geq P[\xi < \mu'(s)]$, Eqs. 18, 19, and 21 become

$$F[\mu'(s)+2\sqrt{\mu''(s)}] \geq \frac{1}{2} \exp[\mu(s)-s\mu'(s)+s\sqrt{2\mu''(s)}] \tag{78}$$

$$\mu'(s_1) = \mu'(s) + \sqrt{2\mu''(s)} \tag{79}$$

$$F[\mu'(s_1)] \geq \frac{1}{2} \exp[\mu(s_1)-s_1\mu'(s_1)+2s\sqrt{2\mu''(s)}]. \tag{80}$$

Equation 54, for identically distributed nonlattice variables, is

$$F[\mu'(s)] = \left[\frac{1}{\sqrt{2\pi N s^2 \mu_n''(s)}} + o\left(\frac{1}{\sqrt{N}}\right)\right] \exp[\mu(s)-s\mu'(s)]. \tag{81}$$

Equation 61, for identically distributed lattice variables, is

$$F[\mu'(s)] = \left\{\frac{h \exp(s\Delta)}{\sqrt{2\pi N \mu_n''(s)}\,[1-\exp(sh)]} + o\left(\frac{1}{\sqrt{N}}\right)\right\} \exp[\mu(s)-s\mu'(s)], \tag{82}$$

where Δ is the interval between $\mu'(s)$, and the largest value of Na + kh less than or equal to $\mu'(s)$. Finally, Eq. 74 becomes

$$F[\mu'(s)] \geq B \exp[\mu(s)-s\mu'(s)] \tag{83}$$

$$B = \frac{\exp\left[z_o s\sqrt{\mu''(s)} - \frac{z_o^2}{2}\right]}{\sqrt{2\pi}\,[z_o - s\sqrt{\mu''(s)}]} \left\{1 - \frac{1}{[z_o - s\sqrt{\mu''(s)}]^2}\right\}, \tag{84}$$

where z_o is given by Eq. 67, and Eqs. 77-84 are valid for $s < 0$.

R. G. Gallager

References

1. H. Chernov, A measure of asymptotic efficiency for tests of an hypothesis based on the sum of observations, Ann. Math. Statist. 23, 493 (1952).

2. C. E. Shannon, Seminar Notes for Seminar in Information Theory at M. I. T., 1956 (unpublished).

3. R. M. Fano, Transmission of Information (The M. I. T. Press, Cambridge, Mass., and John Wiley and Sons, Inc., New York, 1961); see Chap. 8.

4. C. G. Esseen, Fourier analysis of distribution functions: A mathematical study of the Laplace-Gaussian law, Acta Math. 77, 1-125 (1945).

5. B. V. Gnedenko and A. N. Kolmogoroff, Limit Distributions for Sums of Independent Random Variables (Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954); see p. 210.

6. W. Feller, Probability Theory and Its Applications (John Wiley and Sons, Inc., New York, 1950); see Chap. 7.

7. B. V. Gnedenko and A. N. Kolmogoroff, op. cit., p. 233.

8. A. C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates, Trans. Am. Math. Soc. 49, 122-136 (1941).

B. ERROR BOUNDS FOR GAUSSIAN NOISE CHANNELS

Considerable work on estimating the achievable probability of error in communication over a wide variety of channels has indicated that for time-discrete channels the minimum achievable error probability has the form

$$P_e \approx e^{-NE(R)},$$

where N is the block length on which coding and decoding operations are carried out, R is the transmission rate, in nats per channel use, and E(R) is a function of R and of the channel but is independent of N. The approximation is usually close only for large values of N. Thus the estimation of the function E(R) amounts to estimating

$$\lim_{N\to\infty} -\frac{\ln P_e}{N}.$$

Usually it is hard to calculate E(R) exactly, and bounds for the function are found instead. The upper and lower bounds are defined as follows: For any $\epsilon > 0$ and sufficiently large N there exists a code for which

$$P_e \leqslant e^{-N[E_U(R)-\epsilon]}$$

and there exists no code for which:

$$\overline{P_e} \leqslant e^{-N[E_L(R)+\epsilon]}.$$

Note that $E_U(R)$ is a lower bound to E(R), but it arises in upper-bounding the achievable P_e, and $E_L(R)$ is an upper bound to E(R) used in lower-bounding P_e.

Gallager[1,2] has found a number of these upper and lower bounds for the discrete memoryless channel and also for the time-discrete Gaussian noise channel with a power constraint. Shannon[3] found some of the same bounds for the bandlimited white Gaussian noise channel, except that in his case the block length is replaced by the time duration over which coding and decoding operations take place

$$P_e \leqslant e^{-TE(R)},$$

and the rate R is in nats per second. If one takes the limit of these bounds for $N \to \infty$, one finds that all of the bounds have some properties in common. The limits of the upper and lower bounds coincide for R = 0 and $R \geqslant R_{crit}$, where R_{crit} is a function of the channel and lies between 0 and capacity. The bounds are decreasing convex downward functions of R, and all become zero at R = Capacity.

The channel model that is analyzed here consists of a number of time discrete channels each disturbed by an independent Gaussian noise. Each channel is to be used only once, and the total energy used in signaling over the channels is constrained. There is no limitation on the number of channels, nor is the value of the noise power required to be the same in all channels. This model, therefore, represents one channel used N times, a set of Q parallel channels with arbitrary noise power in each used N times, or parallel channels, all having different noise powers each of which is used once. This model also represents the colored Gaussian noise channel with an average power constraint. One takes a Karhunen-Loeve expansion of the noise over a T-second interval. Each of the eigenfunctions of the noise autocorrelation function so obtained is considered as one of the component channels in the model. When the noise is Gaussian, the Karhunen-Loeve theorem states that the noise power in each of the eigenfunctions is independent, which is exactly what is needed for the model. As T is made large, the distribution of the noises in the eigenfunction channels approaches the power density spectrum of the noise, and the resulting E(R) function can be expressed in terms of this spectrum. In this case, the energy constraint is PT, where P is just the power available.

Techniques similar to those used by Gallager in obtaining upper and lower bounds for the discrete memoryless channel can be applied here, except that now there is some added freedom; the energies distributed to each of the component channels of the model are subject only to the constraint that they be positive and add up to NP on the average. With this freedom comes the new problem of determining the optimum distribution of energy to the component channels.

When the various bounds are evaluated a remarkable phenomenon appears. As might be expected, only the component channels with noise power below a threshold (N_b) are to be used for communication, but the value of the threshold over most of the parameter range is dependent only on the rate, and is independent of the power available or of the probability of error desired.

1. Lower Bound on E(R)

Since our model consists of a number of parallel channels, each of which is used only once, it has an implicit block length of one, and the resulting bound is of the form

$$P_e \leq e^{-E^*(R^*)+\epsilon},$$

where now R^* is the nats per block. The quantity ϵ will be discussed in more detail. Suffice it to say now that when we have Q channels, each used N times, $\frac{\epsilon}{E^*(R^*)}$ will go to zero with increasing N. Also, when the channels come from the eigenfunctions of a Karhune-Loeve expansion, $\frac{\epsilon}{E^*(R^*)} \to 0$ as $T \to \infty$.

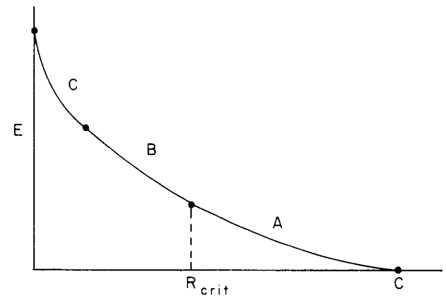

Fig. XIV-3. $E_U(R)$ for fixed power. (Because of the similar relations among R^* and R, E^* and E, and S and P, the curves are drawn for E, R, and P.)

The lower bound takes on three different forms as shown in Fig. XIV-3. Let μ_n be the noise power in the n^{th} channel, and let S be the total energy constraint on the inputs. Then in region A, the relations are

$$R^* = \frac{1}{2} \sum_{N_n \leq N_b} \ln \frac{N_b}{N_n}, \qquad \rho \leq 1 \tag{1}$$

$$E^* = \frac{\rho S}{(1+\rho)\, 2N_b} - \frac{1}{2} \sum_{N_n \leq N_b} \ln \left(1 + \rho - \rho \frac{N_n}{N_b}\right), \tag{2}$$

where

$$S = (1+\rho) \sum_{N_n \leq N_b} \frac{N_b - N_n}{1 - \frac{\rho}{1+\rho} \frac{N_n}{N_b}}. \tag{3}$$

For a given rate R^*, and a given energy S, we observe that N_b is defined by Eq. 1, ρ by Eq. 3, and E^* by Eq. 2; the bound is valid in the region where the resultant ρ lies in the interval (0, 1).

The form of Eqs. 1, 2, and 3 is somewhat different from Gallager and Shannon's, in that it has two parameters (ρ, N_b) rather than the usual one. This is not a serious problem if one approaches it in a slightly different manner. Instead of specifying the energy and then finding R^* and E^* as functions of ρ, we first specify the rate R^*. R^* determines N_b; although this is not a simple relation, it is a one-to-one relation. Once N_b is determined, one has E^* and S as functions of the parameter ρ.

To see how the probability of error goes to zero with increasing N for the Q-channel case, we note that the number of component channels in the model with a given value of N_n is a multiple of N, and consequently

$$\frac{1}{2}\sum_{N_n\le N_b}\frac{N_b-N_n}{1-\frac{\rho}{1+\rho}\frac{N_n}{N_b}}=\frac{N}{2}\sum_{N_q\le N_b}\frac{N_b-N_q}{1-\frac{\rho}{1+\rho}\frac{N_q}{N_b}}$$

in which the sum on the right is just over those of the Q original channels that have noise power equal to or less than N_b. The same thing is true of the other sums, and we can write

$$R^*=NR=\frac{N}{2}\sum_{N_q\le N_b}\ln\frac{N_b}{N_q}$$

$$E^*=NE=\frac{N\rho P}{2(1+\rho)\,N_b}-\frac{N}{2}\sum_{N_q\le N_b}\ln\left(1+\rho-\rho\frac{N_q}{N_b}\right), \tag{4}$$

where

$$S=NP=(1+\rho)N\sum_{N_n\le N_b}\frac{N_b-N_q}{1-\frac{\rho}{1+\rho}\frac{N_q}{N_b}}.$$

For the colored-noise case, as $T\to\infty$ it can be proved that

$$\frac{1}{2}\sum_{N_n\le N_b}\ln\frac{N_b}{N_n}\to T\cdot\frac{1}{2}\int_{N(f)\le N_b}\ln\frac{N(f)}{N_b}\,df+0(T),$$

where $\frac{0(T)}{T}\to 0$, and N(f) is the power density spectrum of the noise. The same thing is true of the other summations, and relations similar to (4) can be written. The boundary of region A is set by $\rho = 1$.

In region C we have

$$R^*=\frac{1}{2}\sum_{N_n\le N_b}\ln\frac{N_b}{N_n}-\frac{1}{2}\sum_{N_n\le N_b}\ln\left(2-\frac{N_n}{N_b}\right) \tag{5}$$

$$E^*=\frac{S}{4N_b}.$$

For completeness we write

$$S=4\rho\sum_{N_n\le N_b}\frac{N_b-N_n}{2-\frac{N_n}{N_b}}\qquad \rho\ge 1.$$

In both regions $\frac{\partial E^*}{\partial R^*} = \frac{\partial E}{\partial R} = -\rho$.

The functions in region B are found to be

$$S = 4 \sum_{N_n \leq N_b} \frac{N_b - N_n}{2 - \frac{N_n}{N_b}}$$

$$E^* = \frac{S}{4N_b} + \frac{1}{2} \sum_{N_n \leq N_b} \ln \frac{N_b}{N_n} - \frac{1}{2} \cdot \sum_{N_n \leq N_b} \ln \left(2 - \frac{N_n}{N_b}\right) - R^*.$$

In this region ρ is held at 1, and consequently $\frac{\partial E^*}{\partial R} = -1$; thus only the variable N_b is left to adjust the trade-off between E^* and S. In this region N_b is not a function of R^*, but of S. The separations of regions A, B, and C can be made by examining the value of ρ. Once the values of N_b and ρ have been determined, it is a simple matter to design the signals that will produce a probability of error as small as the upper-bound exponent indicates.

The form of Eqs. 4 lends itself to a presentation of E vs P for fixed R. This presentation is shown in Fig. XIV-4. The regions are labeled the same as those in

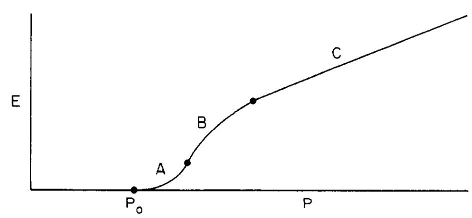

Fig. XIV-4. $E_U(P)$ for fixed R.

Fig. XIV-3. The value P_o is the minimum power needed for reliable transmission at the given rate; and for any power less than P_o, the exponent is zero. In region A we have

$$\frac{\partial E}{\partial P} = \frac{\rho}{2N_b(1+\rho)}.$$

At $\rho = 1$ we go into region B, and here ρ remains constant at 1 but N_b increases, thereby increasing the bandwidth. In this region

$$\frac{\partial E}{\partial P} = \frac{1}{4N_b}.$$

In region C we again find that N_b remains fixed, as in region A, except now at a larger value. We have

$$\frac{\partial E}{\partial P} = \frac{1}{4N_b}$$

which is just a straight line.

This curve gives an attainable exponent for given rate and power. According to the derivation of this bound the power used in each component channel is

$$(1+\rho)\frac{N_b - N_n}{1 - \frac{\rho}{1+\rho}\frac{N_n}{N_b}},$$

and the total energy used for each block transmission must lie within a shell

$$N(P_{av}-\delta) \leq S \leq NP_{av}.$$

2. Upper Bound on E(R)

An upper bound on the exponent can be obtained by a sphere-packing type of arguement. This bound has the same form as the lower-bound exponent in region A, Eq. 4, except that now ρ can take on any positive value. A typical curve is shown in Fig. XIV-5.

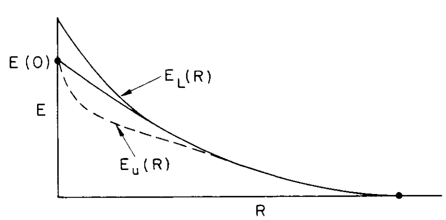

Fig. XIV-5. $E_L(P)$ for fixed P.

There are two small reductions that can be made in this bound. First, the value of E(0) for the lower-bound exponent is $\frac{P}{4N_{min}}$ (see Eq. 5), which Shannon[3] has shown is also the upper bound of the white Gaussian channel exponent at zero rate. The upper bound applies to the model also, since it will certainly be inferior to the white channel which has all of its noise power equal to the minimum of the noises in the model. Once the E(0) is reduced, one can produce a straight-line bound through E(0) tangent to the old upper bound by Shannon and Gallager's[2] technique of breaking the channel up into two parts and looking at the best list-decoding bound on one part and the zero-rate bound on the other part. Then the probability of error can be shown to be greater than one-fourth the product of the probability of error for each of these steps. Figure XIV-6 shows the sphere-packing and the zero-rate bounds (not the tangent bound) presented as E vs P. Curves A are the sphere-packing bounds for several rates and have slopes $\frac{\rho}{2N_b(1+\rho)}$;

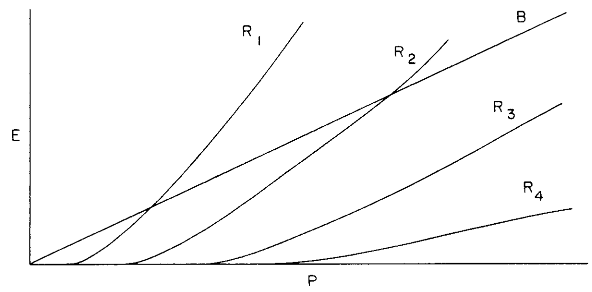

Fig. XIV-6. $E_L(P)$ for several fixed values of R.

thus for large ρ their slopes approach $\frac{1}{2N_b}$. Curve B has slope $\frac{1}{4N_{min}}$ and is independent of rate. It merely states that, no matter what the rate, the exponent is less than the zero-rate exponent. If the rate is large enough, curve A is always below curve B, and thus the tighter bound. The effect of the tangent line is not shown in Fig. XIV-6, since its effect can only be found graphically once the E vs R curves are plotted for all values of P. It is known that this tangent line has no effect for sufficiently large R.

In the region where the upper and lower bounds agree, between R_{crit} and capacity, one can make definite statements about the nature of the optimum signals. It has been found that the signals must lie entirely within those channels, or that part of the spectrum, where the noise power is below the threshold, N_b, and that N_b is determined by the relation

$$R^* = \frac{1}{2} \sum_{N_n \leq N_b} \ln \frac{N_b}{N_n}.$$

In the other regions where the two bounds do not agree, one cannot say anything about the signals with certainty, except that the average power required for a given rate and exponent is less than that given by Fig. XIV-4 and greater than that given in Fig. XIV-6.

3. Outline of Proofs

a. Lower Bound on E(R)

We obtain an upper bound to the probability of error by the standard random-coding argument.[1] The only difference from the standard procedure in this particular case lies in defining the ensemble from which the random code words are chosen. A code word consists of a set of x_n that are to be transmitted through the component channels in the model. The ensemble of codes is defined by taking each x_n in each code word from a Gaussian distribution with zero mean and variance

$$\overline{x}_n^2 = (1+\rho)\frac{N_b - N_n}{1 - \frac{\rho}{1+\rho}\frac{N_n}{N_b}}; \qquad N_n \leq N_b$$

$$= 0; \qquad N_n > N_b$$

and then rejecting all except those code words for which

$$NP - \delta < \sum_n x_n^2 \leq NP; \qquad \delta > 0.$$

In other words, the energy in each code word is required to be in a small shell. One then writes, following Gallager,[1]

$$P_e \leq e^{-N[E_o(\rho)-\rho R]},$$

where

$$E_o(\rho) = -\ln \int \left[\int p(\underline{x})\, p(\underline{y}/\underline{x})^{\frac{1}{1+\rho}}\, d\underline{x}\right]^{1+\rho} d\underline{y}, \qquad 0 \leq \rho \leq 1.$$

In this equation, $p(\underline{y}/\underline{x})$ is known to be a product of Gaussian distributions. $p(\underline{x})$ is given above but can be bounded for any $r > 0$ by

$$p(\underline{x}) \leq q^{-1} e^{r \sum_n x_n^2 - rNP + r\delta} \frac{e^{-\sum_n \frac{x_n^2}{2\sigma_n^2}}}{\prod_n \sqrt{2\pi\sigma_n^2}},$$

where

$$\sigma_n^2 = (1+\rho)\frac{N_b - N_n}{1 - \frac{\rho}{1+\rho}\frac{N_n}{N_b}}; \qquad N_n \leq N_b$$

$$r = \frac{\rho}{2N_b(1+\rho)^2}$$

$$q = \int_{NP-\delta \leq \sum_n x_n^2 \leq NP} \frac{e^{-\sum_n \frac{x_n^2}{2\sigma_n^2}}}{\prod_n \sqrt{2\pi\sigma_n^2}}\, d\underline{x}.$$

When

$$NP = (1+\rho) \sum_n \frac{N_b - N_n}{1 - \frac{\rho}{1+\rho} \frac{N_n}{N_b}},$$

q only decreases algebraically with N and consequently does not affect the exponent. Evaluating the expression for $E_o(\rho)$ and maximizing over ρ, we obtain the expressions for R and E. The expurgated bound is found by the same method. Now the variances used are

$$\sigma_n^2 = 4\rho \frac{N_b - N_n}{2 - \frac{N_n}{N_b}}; \quad N_n \leq N_b$$

and, following Gallager's expurgated bound,

$$P_e \leq 4\rho e^{-N[E_o(\rho)-\rho R]},$$

where now

$$E_o(\rho) = -\ln \left[\int_{\underline{x}'} \int_{\underline{x}} p(x)\, p(\underline{x}') \left(\int_{\underline{y}} \sqrt{p(\underline{y}/\underline{x})\, p(\underline{y}/\underline{x}')}\, d\underline{y} \right)^{1/\rho} d\underline{x}d\underline{x}' \right]^{\rho}, \quad \rho \geq 1.$$

This can be evaluated in the same manner, and the expurgated relations can be obtained.

b. Upper Bound on E(R)

The lower bound on the probability of error is obtained by using a method of Gallager,[2] based on the Chebychev inequality. The theorem used here states: Define $\mu_m(s)$ as

$$\mu_m(s) = \ln \int_{\underline{y}} f(\underline{y})^s\, p(\underline{y}/\underline{x}_m)^{1-s}\, d\underline{y},$$

where $f(\underline{y})$ is an arbitrary probability density on $\underline{y}$. Then if

$$\int_{Y_m} f(\underline{y})\, d\underline{y} \leq \frac{1}{4} \exp[\mu_m(s)+(1-s)\mu_m'(s)-(1-s)\sqrt{2\mu_m''(s)}],$$

where Y_m is that set of output sequences decoded as m, it follows that

$$P_{em} \geq \frac{1}{4} \exp[\mu_m(s)-s\mu_m'(s)-s\sqrt{2\mu_m''(s)}]. \quad (6)$$

Since the sets Y_m are disjoint, $\int_{Y_m} f(\underline{y})\, d\underline{y}$ cannot be large for all transmitted signals and, in fact, there must be one signal with $\int_{Y_m} f(\underline{y})\, d\underline{y} \leq \frac{1}{M} = e^{-NR}$. Thus when s is chosen so that

$$NR \geq -\mu_m(s) - (1-s)\mu_m'(s) + (1-s)\sqrt{2\mu''(s)} + \ln 4, \tag{7}$$

Eq. 6 gives us a lower bound on P_{em}.

We choose $f(\underline{y})$ to be

$$f(\underline{y}) = \prod_n \frac{e^{-\frac{y_n^2}{2\sigma_n^2}}}{\sqrt{2\pi\sigma_n^2}},$$

where

$$\sigma_n^2 \begin{cases} = N_n & \text{if } N_n \geq N_b \\ = \dfrac{N_b - s_1 N_n}{1 - s_1} & \text{if } N_n < N_b \end{cases}.$$

We shall set s_1 equal to s, but μ_m' and μ_m'' are understood to be the partial derivatives of μ_m with $f(\underline{y})$ fixed. If we set

$$R^* = \frac{1}{2} \sum_{N_n \leq N_b} \ln \frac{N_b}{N_n},$$

and then select s to meet Eq. 7 it turns out that the exponential behavior of the lower bound on P_e is the same as the upper bound. One point that needs to be enlarged upon is that $\mu_m(s)$ depends on the m for which Eq. 6 is satisfied; then it depends on $f(\underline{y})$, which in turn depends on s. If one is to choose s to meet Eq. 7, it looks as if an endless circle of dependencies will arise. It turns out that the $\sqrt{2\mu''(s)}$ becomes negligible for large block length, and that the $-\mu_n(s) - (1-s)\mu_m'(s)$ depends on m only through three sums:

$$\sum_{N_n \leq N_b} x_{mn}^2,$$

$$\sum_{N_n \leq N_b} x_{mn}^2 N_n,$$

$$\sum_{N_n > N_b} x_{mn}^2 / N_n.$$

We therefore restrict ourselves to a small fraction of the m. First consider only those m for which

$$\sum_n x_{mn}^2 \leq \frac{NP}{1-\alpha};$$

this will be at least αM of the signals. Now the three sums are bounded and can be subdivided into a finite number of intervals, each of length $N\delta$. There must be some triplet of intervals which contains at least $\frac{\delta^3(1-\alpha)^3}{\rho^3} \alpha M$ of the input signals. We consider only this set, and note that reducing the set of input symbols by a fixed fraction only reduces the rate by

$$\frac{-\ln \frac{\delta^3(1-\alpha)^3}{\rho^3} \alpha}{N}$$

which approaches zero for large N. Once one knows that Eq. 7 can be achieved within $N\delta$ of equality, substitution in Eq. 6 gives the sphere-packing exponent for the probability of error.

P. M. Ebert

References

1. R. G. Gallager, A simple derivation of the coding theorem and some applications (to be published in IEEE Trans., PTGIT).

2. C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, Lower bounds to error probability for coding on discrete memoryless channel (unpublished).

3. C. E. Shannon, Probability of error for optimum codes in a Gaussian channel, Bell System Tech. J. 38, 611 (1959).