Notes on Information Theory II and the Geometric Interpretation of the Shannon Channel Capacity

H.L. Rappaport

October 2014

Continuous Random Variables and Differential Entropy

The differential entropy of the continuous random variable X is defined [1]

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x)\, dx, \tag{1}$$

where $f_X(x)$ is the probability density function for X. The joint differential entropy of two random variables X, Y possessing a joint density function $f_{XY}(x, y)$ and marginal density functions $f_X(x)$ and $f_Y(y)$ is defined [2]

$$h(X, Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y) \log_2 f_{XY}(x, y)\, dx\, dy, \tag{2}$$

while the individual (or marginal) differential entropies are defined

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x)\, dx, \tag{3}$$

$$h(Y) = -\int_{-\infty}^{\infty} f_Y(y) \log_2 f_Y(y)\, dy, \tag{4}$$

and the conditional entropies are defined

$$h(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y) \log_2 f_X(x|y)\, dx\, dy, \tag{5}$$

$$h(Y|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y) \log_2 f_Y(y|x)\, dx\, dy. \tag{6}$$

In Eqs. (5) and (6) [3]

$$f_X(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}, \qquad f_Y(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}. \tag{7}$$

Key Properties of Conditional Distributions

In analogy with the discrete variable case, substitution of Eqs. (7) into

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = f_X(x), \tag{8}$$

yields

$$\int_{-\infty}^{\infty} f_Y(y|x)\,dy = 1, \tag{9}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_X(x|y)\,f_Y(y)\,dy, \tag{10}$$

and substitution of Eqs. (7) into

$$\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = f_Y(y), \tag{11}$$

yields

$$\int_{-\infty}^{\infty} f_X(x|y)\,dx = 1, \tag{12}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_Y(y|x)\,f_X(x)\,dx. \tag{13}$$

Continuous Random Variables and Mutual Information [1]

The mutual information for a memoryless channel with channel input described by a continuous random variable X and channel output described by a continuous random variable Y is given by

$$I(X;Y) = h(Y) - h(Y|X). \tag{14}$$

From Eqs. (4) and (6) then

$$I(X;Y) = -\int_{-\infty}^{\infty} f_Y(y)\log_2 f_Y(y)\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_Y(y|x)\,dx\,dy, \tag{15}$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_Y(y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_Y(y|x)\,dx\,dy, \tag{16}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 \frac{f_Y(y|x)}{f_Y(y)}\,dx\,dy, \tag{17}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 \frac{f_{XY}(x,y)}{f_X(x)\, f_Y(y)}\, dx\, dy, \tag{18}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 \frac{f_X(x|y)}{f_X(x)}\, dx\, dy, \tag{19}$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_X(x)\ dx\, dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_X(x|y)\ dx\, dy, \tag{20}$$

$$= -\int_{-\infty}^{\infty} f_X(x)\log_2 f_X(x)\ dx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_X(x|y)\ dx\, dy, \tag{21}$$

$$= h(X) - h(X|Y). \tag{22}$$

Thus

$$I(X;Y) = I(Y;X). \tag{23}$$

An expression for the mutual information in terms of the joint differential entropy can be derived by starting with Eq. (18)

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 \frac{f_{XY}(x,y)}{f_X(x)\, f_Y(y)}\, dx\, dy, \tag{24}$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_X(x)\ dx\, dy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_Y(y)\ dx\, dy$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_{XY}(x,y)\ dx\, dy, \tag{25}$$

$$= -\int_{-\infty}^{\infty} f_X(x)\log_2 f_X(x)\ dx - \int_{-\infty}^{\infty} f_Y(y)\log_2 f_Y(y)\ dy$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log_2 f_{XY}(x,y)\ dx\, dy, \tag{26}$$

or

$$I(X;Y) = h(X) + h(Y) - h(X,Y). \tag{27}$$

It turns out that the problem of finding the distribution function $f_X(x)$ that maximizes the differential entropy $h(X)$ of Eq. (1) subject to certain constraints is quite important. This is undertaken next.

Maximizing Differential Entropy

Following Haykin [1] we wish to maximize the differential entropy $h\left(X\right)$ of Eq. (1) subject to the constraints

$$\int_{-\infty}^{\infty} f_X\left(x\right)\,dx = 1, \tag{28}$$

to keep the total probability equal to one and

$$\int_{-\infty}^{\infty} \left(x-\bar{x}\right)^2 f_X\left(x\right)\,dx = \sigma^2 = \text{constant}, \tag{29}$$

where

$$\bar{x} = \int_{-\infty}^{\infty} x f_X\left(x\right)\,dx, \tag{30}$$

to fix the variance of $f_X\left(x\right)$. Thus an extremum in the functional

$$\mathcal{F} = -\int_{-\infty}^{\infty} f_X\left(x\right)\log_2 f_X\left(x\right)\,dx + \lambda_1\int_{-\infty}^{\infty} f_X\left(x\right)\,dx + \lambda_2\int_{-\infty}^{\infty}\left(x-\bar{x}\right)^2 f_X\left(x\right)\,dx, \tag{31}$$

where λ_1 and λ_2 are Lagrange multipliers is desired. This is found by requiring that the first variation of $\mathcal{F}$ vanish, e.g. $\delta\mathcal{F} = 0$ or that

$$-\delta\int_{-\infty}^{\infty} f_X\left(x\right)\log_2 f_X\left(x\right)\,dx + \lambda_1\delta\int_{-\infty}^{\infty} f_X\left(x\right)\,dx + \lambda_2\delta\int_{-\infty}^{\infty}\left(x-\bar{x}\right)^2 f_X\left(x\right)\,dx = 0. \tag{32}$$

To evaluate this variation it is helpful to note that

$$\delta\log_2 f_X\left(x\right) = \frac{\delta f_X\left(x\right)}{f_X\left(x\right)\,\ln 2} = \delta f_X\left(x\right)\frac{\log_2 e}{f_X\left(x\right)}, \tag{33}$$

and Eq. (32) becomes

$$\int_{-\infty}^{\infty} \delta f_X\left[-\log_2 f_X - \log_2 e + \lambda_1 + \lambda_2\left(x-\bar{x}\right)^2\right]\,dx = 0. \tag{34}$$

Now in order for Eq. (34) to be satisfied with $\delta f_X\left(x\right)$ arbitrary

$$-\log_2 f_X - \log_2 e + \lambda_1 + \lambda_2\left(x-\bar{x}\right)^2 = 0, \tag{35}$$

must be satisfied. Solving Eq. (35) for f_X yields

$$f_X = \exp\left[-1 + \frac{\lambda_1}{\log_2 e} + \frac{\lambda_2}{\log_2 e}\left(x-\bar{x}\right)^2\right], \tag{36}$$

showing that necessarily $\lambda_2 < 0$ so that integrals such as found in Eq. (28) converge. Using the integrals [4]

$$\int_{-\infty}^{\infty} \exp\left(-ax^2\right)\,dx = \sqrt{\frac{\pi}{a}}, \qquad \text{and} \qquad \int_{-\infty}^{\infty} x^2 \exp\left(-ax^2\right)\,dx = \frac{\sqrt{\pi}}{2a^{3/2}}, \tag{37}$$

substitution of Eq. (36) into Eq. (28) yields

$$\exp\left(-1 + \frac{\lambda_1}{\log_2 e}\right)\sqrt{\frac{\pi \log_2 e}{-\lambda_2}} = 1, \tag{38}$$

while substitution of Eq. (36) into Eq. (29) yields

$$\exp\left(-1 + \frac{\lambda_1}{\log_2 e}\right)\frac{\sqrt{\pi}}{2}\left(-\frac{\log_2 e}{\lambda_2}\right)^{3/2} = \sigma^2. \tag{39}$$

Combining Eqs. (38) and (39) yields

$$\lambda_2 = -\frac{\log_2 e}{2\sigma^2}, \tag{40}$$

$$\exp\left(-1 + \frac{\lambda_1}{\log_2 e}\right) = \frac{1}{\sigma\sqrt{2\pi}}, \tag{41}$$

so that Eq. (36) becomes

$$f_X\left(x\right) = f_{X\text{norm}}\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}\left(x - \bar{x}\right)^2\right], \tag{42}$$

the standard formula for a normal or Gaussian distribution [5].

The next step in the calculation is to actually evaluate the extremal entropy by substituting Eq. (42) into Eq. (1)

$$h = -\int_{-\infty}^{+\infty} f_X \log_2 f_X \, dx. \tag{43}$$

Now

$$\log_2 f_X = -\log_2\left(\sigma\sqrt{2\pi}\right) + \left(\log_2 e\right)\left[-\frac{1}{2\sigma^2}\left(x - \bar{x}\right)^2\right], \tag{44}$$

and

$$f_X \log_2 f_X = \frac{1}{\sigma\sqrt{2\pi}}\left[-\log_2\left(\sigma\sqrt{2\pi}\right) - \frac{\left(\log_2 e\right)\left(x - \bar{x}\right)^2}{2\sigma^2}\right]\exp\left[-\frac{1}{2\sigma^2}\left(x - \bar{x}\right)^2\right]. \tag{45}$$

So

$$h = \frac{1}{\sigma\sqrt{2\pi}}\log_2\left(\sigma\sqrt{2\pi}\right)\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2\sigma^2}\left(x - \bar{x}\right)^2\right]dx$$

$$+\frac{1}{\sigma\sqrt{2\pi}}\frac{\log_2 e}{2\sigma^2}\int_{-\infty}^{\infty}(x-\bar{x})^2\exp\left[-\frac{1}{2\sigma^2}(x-\bar{x})^2\right]dx, \tag{46}$$

Using the integrals given in (37) shows

$$h=\frac{\log_2\left(\sigma\sqrt{2\pi}\right)}{\sigma\sqrt{2\pi}}\sqrt{2\pi\sigma^2}+\frac{1}{\sigma\sqrt{2\pi}}\frac{\sqrt{\pi}\log_2 e}{4\sigma^2}\left(2\sigma^2\right)^{3/2}. \tag{47}$$

$$=\log_2\left(\sigma\sqrt{2\pi}\right)+\frac{\log_2 e}{2}=\frac{1}{2}\log_2\left(2\pi e\sigma^2\right), \tag{48}$$

a useful result.

Finally, to show that the entropy of Eq. (48) is truly a maximum, we follow a procedure outlined in the references [6, 7]. Consider an arbitrary probability density function (p.d.f.) $f_{X\text{arb}}(x)$ satisfying the constraints Eqs. (28) and (29).

From Eq. (44)

$$-\int_{-\infty}^{\infty}f_{X\text{arb}}(x)\log_2\left[f_{X\text{norm}}(x)\right]dx=\log_2\left(\sigma\sqrt{2\pi}\right)+\frac{1}{2}\log_2 e, \tag{49}$$

$$=\frac{1}{2}\log_2\left(2\pi e\sigma^2\right), \tag{50}$$

where the two constraint Eqs. (28) and (29) have been used. Since the results of Eqs. (50) and (48) are the same we have

$$\int_{-\infty}^{\infty}f_{X\text{arb}}(x)\log_2\left[f_{X\text{norm}}(x)\right]dx=\int_{-\infty}^{\infty}f_{X\text{norm}}(x)\log_2\left[f_{X\text{norm}}(x)\right]dx. \tag{51}$$

Equation (51) is an intermediate result needed to show the normal distribution yields the maximum entropy below.

The difference between the entropy of the normal distribution and an arbitrary distribution satisfying the constraints is given by

$$h_{\text{norm}}(X)-h_{\text{arb}}(X)=-\int_{-\infty}^{\infty}f_{X\text{norm}}\log_2 f_{X\text{norm}}\,dx+\int_{-\infty}^{\infty}f_{X\text{arb}}\log_2 f_{X\text{arb}}\,dx, \tag{52}$$

or using Eq. (51)

$$h_{\text{norm}}(X)-h_{\text{arb}}(X)=-\int_{-\infty}^{\infty}f_{X\text{arb}}\log_2 f_{X\text{norm}}\,dx+\int_{-\infty}^{\infty}f_{X\text{arb}}\log_2 f_{X\text{arb}}\,dx, \tag{53}$$

$$= \int_{-\infty}^{\infty} f_{X\text{arb}} \log_2 \left(\frac{f_{X\text{arb}}}{f_{X\text{norm}}} \right) \, dx \geq 0, \tag{54}$$

the result that was to be proven. The inequality in Eq. (54) follows from the discussion given in the appendix. Note that equality in (54) holds if $f_{X\text{arb}} = f_{X\text{norm}}$ as expected.

Our goal here will be to compute the channel capacity of a band-limited, power-limited Gaussian channel. The next step in the computations to find the conditional differential entropy of such a channel.

Conditional Differential Entropy of the Gaussian Channel

Consider a channel in which the input signal is represented by the random variable X and the output signal is represented by the random variable Y and

$$Y = X + N, \tag{55}$$

where N is a random variable representing the noise signal that is added to the signal and X and N are statistically independent. Distribution functions involving the random variable N can be defined in the usual way. For example, the probability that N is between n and $n+dn$ is $f_N(n)\, dn$.

From Eqs. (6) and (7)

$$h(Y|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\left[\log_2 f_{XY}(x, y) - \log_2 f_X(x)\right] dy\, dx. \tag{56}$$

so that when the y integration is performed x is held constant. Writing $y = x + n$ Eq. (56) becomes

$$h(Y|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, x+n)\left[\log_2 f_{XY}(x, x+n) - \log_2 f_X(x)\right] dn\, dx. \tag{57}$$

However,

$$f_{XY}(x, x+n) = f_{XN}(x, n). \tag{58}$$

There are a number of ways to obtain Eq. (58). One way is to invoke the random variable transformation theorem of Gillespie [8]. If

$$X' = X, \qquad N = Y - X, \tag{59}$$

then Gillespie's theorem gives

$$f_{X'N}(x', n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, \delta(x' - x)\, \delta[n - (y - x)]\; dx\, dy. \tag{60}$$

The integral in (60) is readily evaluated showing

$$f_{X'N}\left(x',n\right) = f_{XY}\left(x',x'+n\right), \tag{61}$$

yielding Eq. (58).

Substitution of Eq. (58) into Eq. (57) yields

$$h\left(Y|X\right) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XN}\left(x,n\right)\left[\log_2 f_{XN}\left(x,n\right) - \log_2 f_X\left(x\right)\right]\, dn\, dx. \tag{62}$$

Since X and N are statistically independent

$$f_{XN}\left(x,n\right) = f_X\left(x\right) f_N\left(n\right), \tag{63}$$

Eq. (62) becomes

$$h\left(Y|X\right) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X\left(x\right)\; f_N\left(n\right)\log_2 f_N\left(n\right)\; dn\, dx, \tag{64}$$

or

$$h\left(Y|X\right) = h\left(N\right). \tag{65}$$

This is a useful result and will assist with computing the channel capacity.

Computing Channel Capacity

The goal of these computations is to find the channel capacity in bits/s. However, the first step is to compute the capacity, or maximum of the mutual information, for a single channel use [1] or a single sample time of the input signal. The Shannon maximum sample rate is $2B$ samples per second where B is the signal bandwidth. Thus the capacity in bits/s is the capacity in bits per sample times $2B$.

Now the channel capacity is the maximum of the mutual information that can be obtained for any possible distribution of the input variable X holding the signal power constant. This can be written

$$C = \max\left[I\left(X;Y\right)\right], \tag{66}$$

where from Eq. (14)

$$I\left(X;Y\right) = h\left(Y\right) - h\left(Y|X\right). \tag{67}$$

For a Gaussian channel Eq. (65) yields

$$I(X;Y) = h(Y) - h(N), \tag{68}$$

so

$$C = \max[h(Y)] - h(N). \tag{69}$$

Now in Eq. (42) we found that the maximum differential entropy for a random variable subject to the constraints given in Eqs. (28) and (29) occurs when the random variable is Gaussian distributed and the maximum entropy is then given by Eq. (48). In addition, the noise N is also assumed to be Gaussian distributed so Eq. (48) gives $h(N)$ in Eq. (69). In other words, Eq. (48) gives results for *both terms* in Eq. (69). Thus

$$C = \frac{1}{2}\log_2\left(2\pi e\sigma_y^2\right) - \frac{1}{2}\log_2\left(2\pi e\sigma_n^2\right), \tag{70}$$

where σ_y^2 is the variance of Y and σ_n^2 is the variance of N. Of course Eq. (70) may be written

$$C = \frac{1}{2}\log_2\frac{\sigma_y^2}{\sigma_n^2}. \tag{71}$$

Assuming the signal X and the noise N are uncorrelated

$$\sigma_y^2 = \sigma_x^2 + \sigma_n^2, \tag{72}$$

where σ_x^2 is the variance of X. So, Eq. (71) becomes

$$C = \frac{1}{2}\log_2\left(1 + \frac{\sigma_x^2}{\sigma_n^2}\right). \tag{73}$$

This may be simplified by assuming that X and N are voltages measured across the same impedance so that

$$\frac{\sigma_x^2}{\sigma_n^2} = \frac{P_x}{P_n}, \tag{74}$$

where P_x/P_n is the ratio of the signal power to the noise power and Eq. (73) becomes

$$C = \frac{1}{2}\log_2\left(1 + \frac{P_x}{P_n}\right). \tag{75}$$

As discussed in the beginning of this section Eq. (75) gives the capacity for a single channel use or capacity in bits per sample. The capacity per unit time (or capacity in bits/s) C_t is given by $2B$ samples per second times C

$$C_t = 2BC = B\log_2\left(1 + \frac{P_x}{P_n}\right), \tag{76}$$

which is the famous result given by Shannon.

The Geometric Interpretation of the Shannon Channel Capacity

Papoulis [9] shows that for a band-limited signal $f(t)$

$$f(t) = \sum_{j=-\infty}^{\infty} f_j \frac{\sin(2\pi Bt - j\pi)}{2\pi Bt - j\pi}. \tag{77}$$

where f_j is the value of $f(t)$ at the j^{th} sample time and B is the bandwidth, e.g., the signal spectrum vanishes for frequencies outside the domain $[-B, B]$. Reza [10] considers such a signal that is approximately limited to a time interval of duration T, e.g., the signal vanishes for times t outside the domain $[-T/2, T/2]$. If $BT \gg 1$ the duration of the signal is given approximately as the time between the first and last non-zero samples with spacing $1/2B$ and

$$f(t) \approx \sum_{j=-BT}^{BT} f_j \frac{\sin(2\pi Bt - j\pi)}{2\pi Bt - j\pi}. \tag{78}$$

Next, the average power in the signal P_s is computed

$$P_s = \frac{1}{T}\int_{-T/2}^{T/2} [f(t)]^2\, dt \approx \frac{1}{T}\int_{-\infty}^{\infty} [f(t)]^2\, dt. \tag{79}$$

$$= \frac{1}{T}\sum_{j=-BT}^{BT}\sum_{j'=-BT}^{BT} f_j f_{j'} \int_{-\infty}^{\infty}\left[\frac{\sin(2\pi Bt - j\pi)}{2\pi Bt - j\pi}\right]\left[\frac{\sin(2\pi Bt - j'\pi)}{2\pi Bt - j'\pi}\right] dt. \tag{80}$$

One way to solve the integral in Eq. (80)

$$W_{\text{int}} = \int_{-\infty}^{\infty}\left[\frac{\sin(2\pi Bt - j\pi)}{2\pi Bt - j\pi}\right]\left[\frac{\sin(2\pi Bt - j'\pi)}{2\pi Bt - j'\pi}\right] dt, \tag{81}$$

$$= \frac{1}{2B}\int_{-\infty}^{\infty}\left[\frac{\sin[\pi(t'-j)]}{\pi(t'-j)}\right]\left[\frac{\sin[\pi(t'-j')]}{\pi(t'-j')}\right] dt', \tag{82}$$

is to introduce the expressions

$$\frac{\sin[\pi(t'-j)]}{\pi(t'-j)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp[i\omega(t'-j)]\, d\omega, \tag{83}$$

$$\frac{\sin[\pi(t'-j')]}{\pi(t'-j')} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp[i\omega'(t'-j')]\, d\omega', \tag{84}$$

under the integral sign in Eq. (82) yielding

$$W_{\text{int}} = \frac{1}{(2\pi)^2 2B} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} \exp\left[i\omega\left(t' - j\right) + i\omega'\left(t' - j'\right)\right] \, dt' \, d\omega' \, d\omega, \tag{85}$$

$$= \frac{1}{4\pi B} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \delta\left(\omega + \omega'\right) \exp\left(-i\omega j - i\omega' j'\right) \, d\omega' \, d\omega, \tag{86}$$

$$= \frac{1}{4\pi B} \int_{-\pi}^{\pi} \exp\left[-i\omega\left(j - j'\right)\right] \, d\omega = \begin{cases} 0 & \text{if } j \neq j', \\ 1/2B & \text{if } j = j', \end{cases} \tag{87}$$

or

$$W_{\text{int}} = \frac{1}{2B} \delta_{jj'}. \tag{88}$$

Substitution of Eq. (88) into Eq. (80) yields

$$P_s = \frac{1}{2BT} \sum_{j=-BT}^{BT} f_j^2, \tag{89}$$

The result of Eq. (88) shows that the coefficient functions in Eq. (77) form an orthogonal set. Thus Reza argues that the sum in Eq. (89) can be viewed as the norm squared of a vector in a $2BT$ dimensional vector space. The vector coordinates are given by the f_j and $f_0 = 0$ is assumed. If the length of the vector is d_s then from Eq. (89)

$$d_s = \sqrt{2BTP_s}. \tag{90}$$

Thus in the Gaussian noise channel with input X, output Y and noise N satisfying

$$Y = X + N, \tag{91}$$

the input signal is represented by a point a distance

$$d_x = \sqrt{2BTP_x}, \tag{92}$$

from the origin. The output signal is represented by a point a distance

$$d_y = \sqrt{2BT\left(P_x + P_n\right)}, \tag{93}$$

from the origin, and the noise by a point a distance

$$d_n = \sqrt{2BTP_n}, \tag{94}$$

from the origin.

The requirement for transmission of signals without noise is that the allowable signal points in the $2BT$ dimensional space must be separated a distance given by twice the length of the noise vector. Each of the received signals is represented by a point on a sphere with radius d_y in $2BT$ dimensional space. The question is now how many distinct signals (points on the sphere) can be allowed while keeping the separation between the points equal to $2d_n$? Enforcing this requirement permits decoding this signal without ambiguity.

Alternatively, one can ask how many non-overlapping noise spheres can be embedded in the surface of the output signal's sphere? Each noise sphere has radius d_n and has it's center on the surface of the output signal's sphere. Reza argues this problem is equivalent to asking how many spheres of radius d_n can be placed within the sphere of radius d_y because for $2BT$ very large most of the volume of a sphere is close to its surface.

According to these prescriptions the number of allowed signals M is given by

$$M \approx \frac{\text{Volume of sphere with radius } d_y}{\text{Volume of sphere with radius } d_n}. \tag{95}$$

since volume of an p dimensional sphere is proportional to r^p,

$$M \approx \left(\frac{d_y}{d_n}\right)^{2BT} = \left(\frac{P_x + P_n}{P_n}\right)^{BT}. \tag{96}$$

The number of bits be sent by these M allowed signals is just

$$\text{number of bits} = \log_2 M = BT \log_2 \left(1 + \frac{P_x}{P_n}\right), \tag{97}$$

and so the channel capacity C_t in bits/s is just

$$C_t = \frac{1}{T} \log_2 M = B \log_2 \left(1 + \frac{P_x}{P_n}\right), \tag{98}$$

in agreement with Eq. (76).

It is worthwhile to notes that in using $P_y = P_x + P_n$ in Eq. (93) the assumption the input signal and noise signal are uncorrelated has been made. Goldman [11] points out that in this case the vector representing the input signal and the vector representing the noise signal must be orthogonal. So if the input and noise vector have $2BT$ dimensions the output vector must reside on a space having $2BT - 1$ dimensions. This should be of no consequence because $2BT \gg 1$ is assumed.

Appendix

Some needed inequalities involving the natural log, e.g. $\ln(x)$, and log base 2, e.g. $\log_2(x)$, functions are now derived. To begin with

$$1 > \frac{1}{x}, \qquad \text{for } x > 1, \tag{99}$$

so

$$\int_1^x dx' > \int_1^x \frac{1}{x'}\, dx' \qquad \text{for } x > 1, \tag{100}$$

or

$$x - 1 > \ln x \qquad \text{for } x > 1. \tag{101}$$

Similarly,

$$1 < \frac{1}{x}, \qquad \text{for } 0 < x < 1, \tag{102}$$

so

$$\int_x^1 dx' < \int_x^1 \frac{1}{x'}\, dx', \qquad \text{for } 0 < x < 1. \tag{103}$$

This is the same as

$$\int_1^x dx' > \int_1^x \frac{1}{x'}\, dx', \qquad \text{for } 0 < x < 1, \tag{104}$$

or

$$x - 1 > \ln x, \qquad \text{for } 0 < x < 1. \tag{105}$$

Combining (101) and (105) shows

$$x - 1 \geq \ln x, \qquad \text{for } x > 0, \tag{106}$$

with equality holding only at $x = 1$. This relation is given in Haykin [1].

We can extend this result by replacing x with $1/x$ yielding

$$\frac{1}{x} - 1 \geq \ln \frac{1}{x} = -\ln x, \qquad \text{for } x > 0, \tag{107}$$

or

$$\ln x \geq 1 - \frac{1}{x}, \qquad \text{for } x > 0, \tag{108}$$

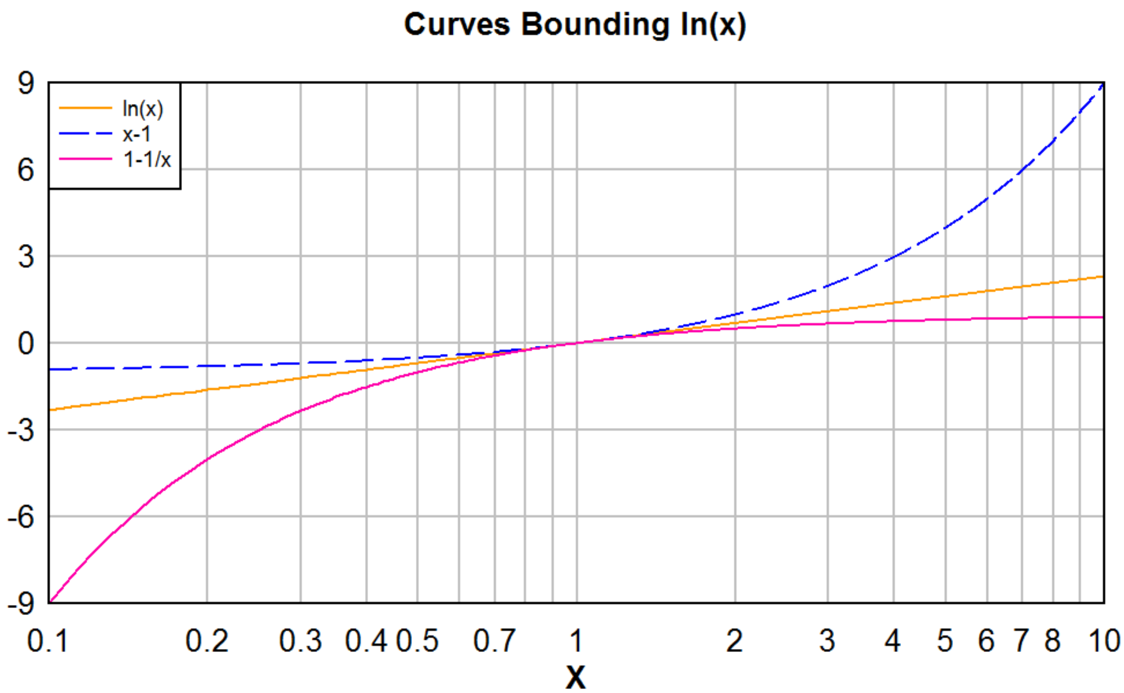

Figure 1: Graph showing $x-1 \geq \ln x \geq 1-1/x$, for $x>0$ with equality of all three curves at $x=1$.

with equality only at $x=1$. Combining (106) and (108) yields

$$x-1 \geq \ln x \geq 1-\frac{1}{x}, \qquad \text{for } x>0. \tag{109}$$

This result is shown graphically in Fig. 1. When $x=1$ all three curves equal zero.

Now if $f(x)$ and $g(x)$ are probability density functions and everywhere satisfy

$$f(x)/g(x)>0, \qquad g(x)/f(x)>0, \tag{110}$$

and assuming the integrals converge,

$$\int_{-\infty}^{\infty} f \log_2\left(\frac{f}{g}\right)\, dx = \log_2(e) \int_{-\infty}^{\infty} f \ln\left(\frac{f}{g}\right)\, dx, \tag{111}$$

$$\geq \log_2(e) \int_{-\infty}^{\infty} f\left(1-\frac{g}{f}\right)\, dx, \tag{112}$$

$$= \log_2(e) \int_{-\infty}^{\infty} f-g\ dx = 0, \tag{113}$$

the result used in Eq. (54). Notice that (112) uses (109) and Eq. (113) uses

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{\infty} g(x)\, dx = 1, \tag{114}$$

satisfied because f and g are probability density functions. Equality holds in (112) if $f = g$. To see the extent to which it may be said that equality holds *only* if $f = g$ consider that equality in (112) implies

$$\int_{-\infty}^{\infty} f \underbrace{\left[\ln\left(\frac{f}{g}\right) - \left(1 - \frac{g}{f}\right)\right]}_{\text{must vanish}} dx = 0. \tag{115}$$

This last relation states that the expected value of the bracketed quantity vanishes. Neither f nor the bracketed quantity can ever be negative. Thus the integrand must vanish everywhere. If f vanishes then g must vanish to be consistent with the inequalities (110). If f does not vanish then necessarily $f = g$.

References

[1] S. Haykin, *Digital Communications*, Wiley, N.Y. (1988), chapter 2.

[2] F. M. Reza, *An Introduction to Information Theory*, Dover, N.Y.,(1994),McGraw-Hill (1961), chapter 8.

[3] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, N.Y. (1965), p. 176.

[4] W.H. Beyer, CRC Standard Mathematical Tables, 25^{th} Edition, CRC Press, Boca Raton, Fl., (1979), p. 380.

[5] Papoulis (1965), p. 100.

[6] I. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," Euro. Trans. Telecommun., 10 (6), (1999), pp. 585 - 595.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, N.Y., (1991), p. 234.

[8] D. T. Gillespie, "A Theorem for Physicists in the Theory of Random Variables," Am. J. Phys. 51 (6), (1983), p. 520.

[9] A. Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, N.Y. (1962), p. 52.

[10] Reza, pp. 318 - 320.

[11] S. Goldman, *Information Theory*, Prentice-Hall, Englewood Cliffs, N.J., (1953), chapter 6.