This material is at the heart of the very major change I made to FT in first week of August (today Aug 8). It resulted in a *complete rewrite* of Section 35 on the subject of statistical pulse trains. Prior to this, I could not justify the Xiong type formulas for a single pulse train involving the $\langle ... \rangle_1$ type averages, but now this is just my "autocorrelation method". The big mystery was making the connection to <....> where you were then allowed to have those random variables Y_n. You just cannot have such things for a single pulse train!!!! The bottom line is that statistics applies to <...> with such things as μ^2 and σ^2 and never to the average $\langle ... \rangle_1$. But when you show these are the same thing, THEN you get Xiong's stuff. Whenever you use a thing like μ^2 or σ^2, you are really talking about an ensemble. For a single long train, this would be what you get chopping it into strings over time and lining them up -- an ensemble! Even B&D talk about ensemble if you look hard enough! I think I have really cleaned this up now in Section 35, maybe I am the only author who has ever done that in readable form.

Meaning of Statistical Pulse Train PhL 7.30.13

I think the phrase "statistical pulse train" may have no meaning! How would I define it? It is the average of the pulse trains in the ensemble? That might be a pulse train with all zero amplitudes, so that certainly is not what I mean. Maybe I should get rid of this phrase. How often used? how about

statistical pulse train → pulse train ensemble.

Lets see how that fits in various places in FT.

A **statistical pulse train** is one in which only certain statistical properties are known about the amplitudes of the pulse train. Given a statistical ensemble of pulse trains, one can determine these statistical properties, such as $\langle y_m \rangle$ and $\langle y_m y_n \rangle$. When we say that a statistical pulse train has a certain spectral power density, for example, we are referring to $\langle \mathcal{P}(\omega) \rangle$ which is the ensemble average of the $\mathcal{P}(\omega)$ over the ensemble. A statistical pulse train therefore is not a specific pulse train whose amplitudes can be written down. It is not any member of the pulse train ensemble, and it is certainly not the average of all the pulse trains in the ensemble (which might be a pulse train with all 0 amplitudes!). Thus one could say that a statistical pulse train is in fact just a set of pulse train ensemble averages of quantities of interest.

A **statistical pulse train** is one whose amplitudes form a **statistical sequence**. In more normal parlance, one would replace the word statistical with random: a **random pulse train is** one whose amplitudes form a **random sequence**. We avoid this terminology only because the word "random" seems to imply a flat probability distribution. For example, one might naturally think of a random sequence of 0's and 1's as one having the probability of a one being p = 1/2. But if p = 1/3, this would still be a statistical sequence. The same subtle implication of a flat distribution (like p(0) = 1/2 and p(1) = 1/2) serves to confuse the notion of a "random variable" as discussed in Appendix G. For us, a statistical sequence is not any particular HOLD HERE

This is getting into the <> versus $<>_1$ vertical/horizontal issue. I think this is problem lurking in the foundation of my FT doc, I had better get it cleaned up.

Major Confusion: If you were handed a long sequence of ones and zeros, you could compute the horizontal average and determine $\mu = p = \langle y_m \rangle_1$ and $\langle y_m y_{m+k} \rangle_1$ from that one pulse train. There is no need for an ensemble. You might find p = 1/2 or p = 1/3 for this "statistical pulse train". So why do I introduce the ensemble idea?

This statistical pulse train has some length N, say, which is relatively long. This is just one member of my ensemble. So that would be a simple definition. But this is just a single NRZ stream from a single input data stream. One should be interested in lots of pulse trains since *any particular one* might be weird in its statistics due to the underlying data stream. It might not have p = 1/2.

35. Statistical Pulse Trains

Definition: **Statistical Pulse Train**

In the following long-winded definition, one can replace the phrase "pulse train" with the word "sequence" where the sequence is the sequence of amplitudes of the pulses in the pulse train.

Suppose we intercept a 1 GHz digital signal (zeros and ones) for 1 msec and thus grab a sequence of N = 1,000,000 amplitude symbols y_m (perhaps using a logic analyzer or digital scope). We could compute the average value for this sequence and we might find $\langle y_m \rangle_1 = 0.55$. Here $\langle ... \rangle_1$ refers to a "horizontal" average over a single pulse train amplitude sequence. By its definition, this average cannot depend on the index m, we are just doing an average of the y_m values in the sequence,

$$\langle y_m \rangle_1 = (1/N) \Sigma_{m=1}^{N} y_m \ .$$

Now it might be that during this 1 msec, there was something unusual about the data being sent, so that this measurement of $\langle y_m \rangle_1 = 0.55$ is not really representative over a longer term. One alternative would be to capture a much longer sequence. Rather than do this, our approach will be to grab very many 1 msec pulse trains and form an ensemble of these representative pulse trains. Perhaps we do this for a week, so the ensemble then has a huge number of pulse trains. We then write down these amplitude sequences in a vertical list on a very tall piece of paper, one sequence below the next. We then number the positions in the sequences 1 to N and we then define $\langle y_m \rangle$ as the *vertical* average through this ensemble of the sequences in position m. This vertical average is written

$$\langle y_m \rangle = (1/I) \Sigma_{i=1}^{I} y_m^{(i)}$$

and in theory this could be different for different columns m. For example, consider this ensemble which has N = 6 and I = 3:

```
1  2  3  4  5  6
a  b  c  d  e  f        sequence #1
```

```
a' b' c' d' e' f'        sequence #2
a" b" c" d" d" f"        sequence #3
```

We would have

$\langle y_m \rangle_1 = (a+b+c+d+e+f)/6.$ // for the first sequence
$\langle y_3 \rangle = (c + c' + c")/3$ // for column 3

Our pulse train has a random variable Y_m associated with each horizontal pulse train position, m = 1,2...N, and the "vertical" average $\langle y_m \rangle$ is the expected value of Y_m over the ensemble, normally written $\langle y_m \rangle = E(Y_m)$.

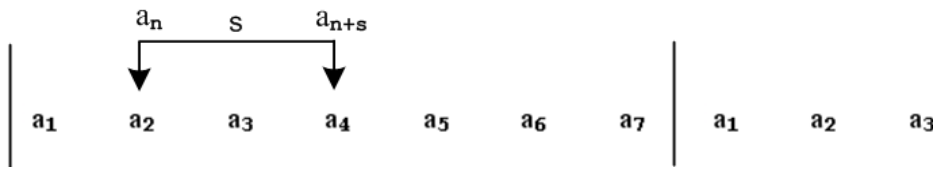
The ensemble is characterized by various statistical properties, such as $E(Y_m)$ or $E(Y_n Y_m Y_k)$, each being a vertical average down through the ensemble. Our **statistical pulse train** is an idealized pulse train whose amplitudes form a **statistical sequence** whose statistics (like $E(Y_m)$ and $E(Y_n Y_m Y_k)$) exactly match those of the ensemble. It is not any particular member of the ensemble, since any member might deviate somehow. And it is certainly not the average of the pulse trains in the ensemble, since this average pulse train would have amplitudes $y_m = \langle y_m \rangle$ which has nothing to do with anything (and also would have illegal symbols, eg, $y_m = 1/2$). If N were very large, a pulse train of the ensemble might come close to being a statistical pulse train.

More normal parlance would refer to our statistical pulse train as a **random pulse train** whose sequence of amplitudes is a **random sequence**. The problem with this terminology is that the word "random" suggests for example that $\langle y_m \rangle$ = the mean value of Y_m in the ensemble = 1/2. That is to say, the word random suggests a flat distribution where p(1) = 1/2 and p(0) = 1/2. But if p(1) = 1/3 and p(0) = 2/3, our ensemble would still be associated with a statistical pulse train, Any value of p ≡ p(1) would describe a statistical pulse train, along with the other statistical properties. This same subtle implication of the word random appears in Appendix G concerning the subject of "random variables".

In what follows, we shall only be concerned with the first and second order statistics of the statistical pulse train, which are $\langle y_m \rangle = E(Y_m)$ and $\langle y_m y_n \rangle = E(Y_m Y_n)$.

****

Plan B. I just am not happy about these two different averages floating around. The 1 average seems more appropriate for talking about correlation in a sequence. You scan the sequence with your pincers to measure $\langle a_m a_{m+k} \rangle_1$. For a GHz digital transmission, THAT is what you really want to know. You examine this long stream and look for correlation. Maybe every 3rd symbol is always a 1. This is the autocorrelation function processing.



But "where are the random variables?" A given position Y_m has only one value $y_m = 2$; there ARE no random variables. So why should you say $\langle a_m a_{m+k} \rangle_1 = \langle a_m \rangle_1 \langle a_{m+k} \rangle_1 = \langle a_m \rangle_1^2$ for uncorrelated? I

don't even know the definition of "uncorrelated" in this picture. Maybe r_s has the shape high in the middle and perfectly flat in the wings and THAT is what no correlation means.

Suppose you define Y_n and Y_m as little toll booths the pulse train passes through as it moves (like a train) from left to right. Then these are random variables, maybe. This is just the above pincer scan but the pincer stays fixed. There is a distribution of y_m in this case. You could run a finite or an infinite pulse train through a series of toll booths Y_1 ..... Y_K where K has nothing to do with the pulse train length N. In this case you would say E(Y_m) = indep of position since the same train runs through all toll booths.

Now flip to the other side of the coin, We have our power formula $\mathcal{P}(\omega) = \Sigma$ ....$y_n y_m$. This certainly exists for a single pulse train, there is some spectrum for sure. And we do know that

$$\mathcal{P}(\omega) \;=\; \mathcal{P}_{pulse}(\omega)\, R''(z)$$

where R"(z) is Z-Transform of r_s of the y_n. So we DO have a way to compute $\mathcal{P}(\omega)$ in an extremely simple manner that makes use of the $\langle a_m a_{m+k}\rangle_1$ object. I did this in App F for a P-repeat sequence.

So why bother with any kind of ensemble? Maybe the step $\langle\mathcal{P}(\omega)\rangle = \Sigma$ ....$\langle y_n y_m\rangle$ was a wrong step to take so early in this doc. This step requires an ensemble. It is like <h> = ... for height of men in ensemble of men. I have to bend over backwards to make a useful ensemble. Maybe my connection between an NRZ sequence and "ensemble" is a complete stupid thing to do.

Am I alone on this, or do other authors use the ensemble idea?

Bennett and Davey? In (19-15) I see something like my power formula $\mathcal{P}(\omega) = \Sigma$ ....$a_n a_m$. They are dealing with a pulse train made of pulse shapes g_1 and g_2 which occur with p and 1-p. Yes, they have subtracted out the discrete part of the spectrum in some tricky way, leaving the continuous part U. Then you see the magic word "**expectation**" and you see $\langle \mathbf{a}_m \mathbf{a}_n\rangle$ in 19-19, but the word "ensemble" does not appear. They refer to a_n and a_m as "variables". But wait! Third line up bott page 316: They talk about **"two ways".** He has Parseval sitting there and $\mathcal{P}(\omega) = w_s(f,T)$ where T is pulse train length. In method 1 you look at this thing integrated over some $\Delta\omega$ and then divide by $\Delta\omega$ to get $\mathcal{P}(\omega)$. I think they are noting that the total integral over $d\omega$ is infinite. The whole matter here is how to take $T \to \infty$, something I have tangled with in many ways. The **second method**, they say, is to "use the concept of an **ensemble** of signal waves". So yes, at least *somebody* is thinking about an ensemble. Now are they going to use this method or not?? They set up a wave type which is "uncorrelated" so that adjacent boxcars are independent, though that is not clarified. They have two pulse options g_1 and g_2. They even talk about the possibility of overlapping pulses. The train runs from -N to N just as I like to do it. So on the next page we get that word "expectation" and we see $\langle a_m a_n\rangle$, so they must be using the "ensemble method". I see this subject arises again on p 221 with a different kind of signal type. But again they use <...> as their ensemble average notation. Warning: with just two variables, they say uncorrelated does NOT imply independence. Maybe they have <xy> = <x><y> but some $\langle x^2 y\rangle$ does not factor? I guess I have to clean this up. I never got it right. But read on p 334. Also, corr lies in -1 to 1, something I missed. Now, in 19-77 they use the <...> notation for cross correlation, and they use it again for autocorrelation. They are talking random variable x(t as function of time, so $R_{\Delta t}(t) = \langle x(t)x(t+\Delta t)\rangle$ which is my $\langle\ldots\rangle_1$ thing. They do use the word autocorrelation. Then in 19-103 sure enough you see $\langle a_n a_m\rangle$ but now it means $\langle a_n a_m\rangle_1$!! There is no ensemble here I don't think! But bottom 337 and the ensemble is back and they make a very obscure remark trying to connect things. Very weak I think. Random phase angle, etc. They might in fact be talking about $\langle \langle a_m a_{m+s}\rangle_1\rangle = \langle r_s\rangle$. They have the phrase "averaging the autocorrelation function over a signaling interval" which makes no sense to me at all? You could average it over an ensemble as I

have just written. But it is already an average. Maybe you are supposed to chop your long signal into finite intervals and make an ensemble of that. That is certainly where I am headed. Then my double average makes sense. That is to say:

$$\langle a_m a_{m+s} \rangle_{1,\text{whole signal}} = (1/I) \Sigma_{i=1}^{I} \langle a_m a_{m+s} \rangle_{1,\text{interval } i} = \langle \langle a_m a_{m+s} \rangle_1 \rangle$$

On p 338 they start their AMI (alternate bipolar, pseudoternary are their terms) line code analysis. Right off the bat they are talking "ensemble" , boom! By the way, 19-26 is the closest they come to my α β formula.

Comments: So my very major source B&D in fact do use ensembles of signals and that is one meaning of their <...> averages. But they also talk about my $\langle ... \rangle_1$ averages without using a different notation, which I find confusion. It is just a general "expectation". They suggest the sausage machine idea of creating an ensemble.

My other author is Xiong. I have his entire book downloaded, so what does he say about ensembles? Line codes page 22. On following pages he discusses each code type, but no calculations. The general calc starts page 28. He then quotes the formula $\mathcal{P}(\omega) = \mathcal{P}_{\text{pulse}}(\omega)\, R''(z)$ from his own appendix. He writes this as

$$\Psi_s(f) = \frac{|G(f)|^2}{T} \sum_{n=-\infty}^{\infty} R(n) e^{-jn\omega T} \qquad (2.12)$$

function of random sequence $\{a_k\}$, defined as $R(n) = E\{a_k a_{k+n}\}$, where $E\{x\}$ is the probabilistic average of x. Equation (2.12) shows that the PSD of a digitally

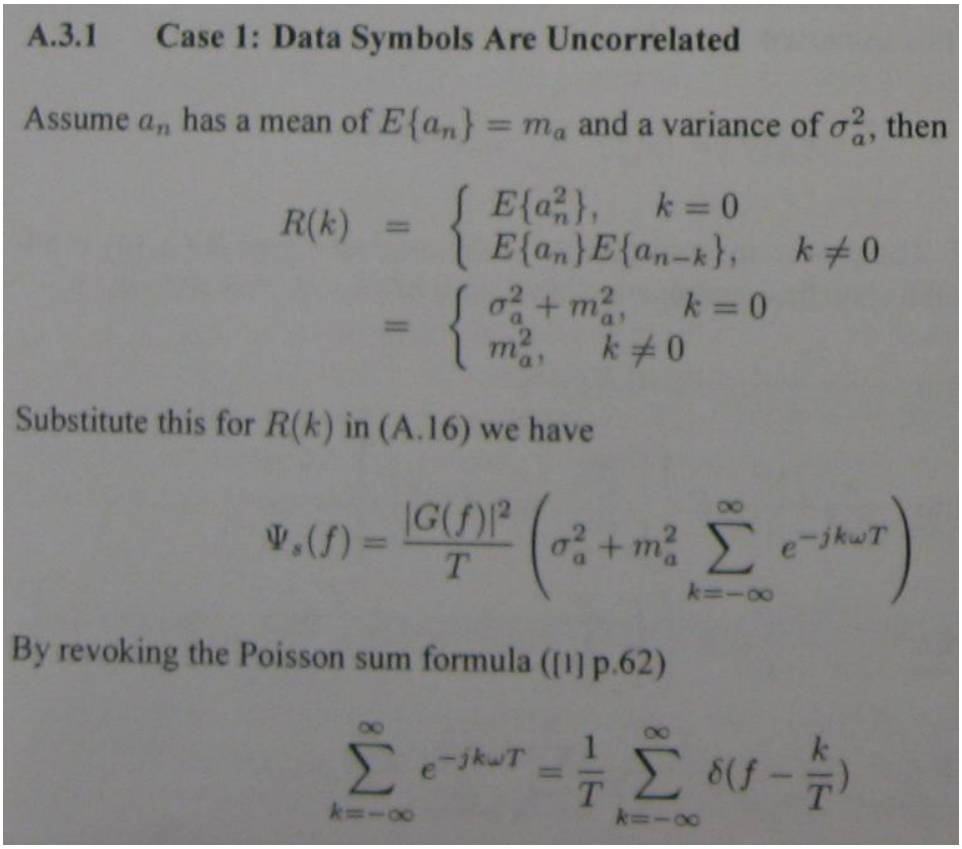
So his R(n) is my r_s and the sum shown is my R"(z). He then has a sort of α,β formula and then gets the classic result which is my α,β formula for $\mathcal{P}(\omega)$. He does NOT use any ensemble!!

$$\Psi_s(f) = \frac{|G(f)|^2}{T} \left[\sigma_a^2 + m_a^2 R_b \sum_{n=-\infty}^{\infty} \delta(f - nR_b) \right] \qquad (2.14)$$

Now let's look at his first use of this which is bipolar NRZ. Square pulse used. G(f) is my $X_{\text{pulse}}(\omega)$. How does he calculate r_s??? He does not use any kind of $\langle ... \rangle_1$ notation, but in effect that is how he computes things. $\langle a_k^2 \rangle = r_0 = p_1(1)^2 + p_0(-1)(-1)^2 = 1$, just as I would do it. But when I do it, I am using an ensemble! But our methods are very close, I will have to ponder that. I think the first time I saw Xiong I was not aware of the WK thing $\mathcal{P}(\omega) = \mathcal{P}_{\text{pulse}}(\omega)\, R''(z)$ so I did not know what he was doing. Xiong has no need for any kind of ensemble. Now let's go to his Appendix A on PSD. I never use "stationary random process", maybe should work that in. Start p 573. He derives my Z Transform W-K. Then there is a drop out in my copy on page 574, but he does say (bike trip was 5.11.13)

$$R(k) = \begin{cases} E\{a_n^2\}, & k = 0 \\ E\{a_n\}E\{a_{n-k}\}, & k \neq 0 \end{cases}$$

Luckily I can fill in with a photo I took



A.3.1 Case 1: Data Symbols Are Uncorrelated

Assume a_n has a mean of $E\{a_n\} = m_a$ and a variance of σ_a^2, then

$$R(k) = \begin{cases} E\{a_n^2\}, & k = 0 \\ E\{a_n\}E\{a_{n-k}\}, & k \neq 0 \end{cases} = \begin{cases} \sigma_a^2 + m_a^2, & k = 0 \\ m_a^2, & k \neq 0 \end{cases}$$

Substitute this for $R(k)$ in (A.16) we have

$$\Psi_s(f) = \frac{|G(f)|^2}{T}\left(\sigma_a^2 + m_a^2 \sum_{k=-\infty}^{\infty} e^{-jk\omega T}\right)$$

By revoking the Poisson sum formula ([1] p.62)

$$\sum_{k=-\infty}^{\infty} e^{-jk\omega T} = \frac{1}{T}\sum_{k=-\infty}^{\infty} \delta(f - \frac{k}{T})$$

So here you see him using $E\{a_n^2\}$ for what I would call $\langle a_n^2 \rangle_1$. Now here is another nearby comment

In digital communications, the most common case is that the data bits are binary (± 1), equiprobable, stationary, and uncorrelated. That is

$$s(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT)$$

where $a_n = +1$ or -1 with equal probabilities, and $E\{a_n a_m\} = E\{a_n\}E\{a_m\}$ for $n \neq m$. It is easy to see that its mean is zero:

where I presume these are my $\langle ... \rangle_1$ averages, somebody has finally said it. And there is that word "stationary" again. Here is further verification

$$
\begin{aligned}
R(k) &= E\{a_n a_{n-k}\} \\
&= \begin{cases} E\{a_n^2\} = \sigma_a^2 = 1, & k = 0 \\ E\{a_n a_{n-k}\} = E\{a_n\}E\{a_{n-k}\} = 0 \cdot 0 = 0, & k \neq 0 \end{cases}
\end{aligned}
$$

Here is yet another deal,

$$
A_N(f) = G(f) \sum_{n=-N}^{N} a_n e^{-j\omega nT}
$$

$$
\begin{aligned}
\Psi_s(f) &= \lim_{N\to\infty} \frac{1}{(2N+1)T} E\{|A_N(f)|^2\} \\
&= \lim_{N\to\infty} \frac{|G(f)|^2}{(2N+1)T} E\left\{\left|\sum_{n=-N}^{N} a_n e^{-j\omega nT}\right|^2\right\} \\
&= \lim_{N\to\infty} \frac{|G(f)|^2}{(2N+1)T} E\left\{\left(\sum_{n=-N}^{N} a_n e^{-j\omega nT}\right)\left(\sum_{n=-N}^{N} a_n e^{-j\omega nT}\right)^*\right\} \\
&= |G(f)|^2 \lim_{N\to\infty} \frac{1}{(2N+1)T} \sum_{n=-N}^{N} \sum_{m=-N}^{N} E\{a_n a_m\} e^{j(m-n)\omega T} \\
&= |G(f)|^2 \lim_{N\to\infty} \frac{1}{(2N+1)T} \sum_{n=-N}^{N} \sum_{k=n+N}^{n-N} R(k) e^{-jk\omega T}
\end{aligned}
$$

What is he doing here?? The first line is my $X(\omega) = X_{\mathtt{pulse}}(\omega)\ Y''(z)$ as in (34.6). The next line says to apply some kind of E to both sides, and left side is then $\Psi_{\mathtt{s}} = E(|X(\omega)|^2$. He then factors out the pulse and we have $E(|X''(z)|^2$. This is exactly like mine. I get

$$
\mathcal{P}(\omega) = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \left| Y''(z) \right|^2 = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} y_m^* \, y_n \, e^{i\omega(m-n)T_1}
$$

My expression has no expectation anything. So I have to interpret his $\Psi_{\mathtt{s}}(f)$ as an expected value of something! Whatever his E thing does. So how does he define this $\Psi_{\mathtt{s}}(f)$?

The PSD of a signal is defined as the signal power in a unit frequency bandwidth (i.e., watts per Hz). The PSD definition is

$$\Psi_s(f) \triangleq |S(f)|^2$$

Note that PSD is always a real function. Since $|S(f)|$ is the amplitude spectrum and

I see no average here. So I claim he is doing a fudge! The bottom line, however, is that Xiong does all his stuff without using an ensemble, it is all that other <..>$_1$ stuff! The word "ensemble" appears nowhere in his book except page 518 where he says, on another topic,

quite differently. To derive a value of B_c, we define an envelope correlation coefficient between two signals as

$$\rho(\Delta f, \Delta t) = \frac{E\{a_1 a_2\} - E\{a_1\}E\{a_2\}}{\sqrt{(E\{a_1^2\} - (E\{a_1\})^2)\,(E\{a_2^2\} - (E\{a_2\})^2)}}$$

where $E\{\cdot\}$ denote the ensemble average and a_1 and a_2 represent the amplitudes of signals at frequencies f_1 and f_2, respectively, and at times t_1 and

****

OK, these two books were good to look at. I will now attempt to assimilate this into some combined theory of the two methods.

First, what does "stationary" mean?

In mathematics, a **stationary process** (or **strict(ly) stationary process** or **strong(ly) stationary process**) is a stochastic process whose joint probability distribution does not change when shifted in time or space. Consequently, parameters such as the mean and variance, if they are present, also do not change over time or position. As a result, the mean and the variance of the process do not follow trends.

In probability theory, a **stochastic process** /stoʊˈkæstɪk/, or sometimes **random process** (*widely used*) is a collection of random variables; this is often used to represent the evolution of some random value, or system, over time. This is the probabilistic counterpart to a deterministic process (or deterministic system). Instead of describing a process which can only evolve in one way (as in the case, for example, of solutions of an ordinary differential equation), in a stochastic or random process there is some indeterminacy: even if the initial condition (or starting point) is known, there are several (often infinitely many) directions in which the process may evolve.

So here is the "time evolution" element which I don't seem to have in my doc. I see what they are talking about, though. Maybe I have a "space evolution" instead of time evolution, and space is then distance across the pulse train. But that is time evolution as well in my toll booth model. But I don't really think

that is right, because I think things are random at any time t which would not apply to me. Maybe my probability distribution for Y_n would change in time as boxcars keep passing by.

Weak or wide-sense stationarity [edit]

A weaker form of stationarity commonly employed in signal processing is known as **weak-sense stationarity**, **wide-sense stationarity** (WSS) or **covariance stationarity**. WSS random processes only require that 1st moment and covariance do not vary with respect to time. Any strictly stationary process which has a mean and a covariance is also WSS.

So, a continuous-time random process $x(t)$ which is WSS has the following restrictions on its mean function

$$\mathbb{E}[x(t)] = m_x(t) = m_x(t+\tau) \quad \text{for all} \quad \tau \in \mathbb{R}$$

and autocovariance function

$$\mathbb{E}[(x(t_1)-m_x(t_1))(x(t_2)-m_x(t_2))] = C_x(t_1,t_2) = C_x(t_1+(-t_2),t_2+(-t_2)) = C_x(t_1-t_2,0).$$

The first property implies that the mean function $m_x(t)$ must be constant. The second property implies that the covariance function depends only on the *difference* between t_1 and t_2 and only needs to be indexed by one variable rather than two variables. Thus, instead of writing,

$$C_x(t_1 - t_2, 0)$$

the notation is often abbreviated and written as:

$$C_x(\tau) \text{ where } \tau = t_1 - t_2.$$

This also implies that the autocorrelation depends only on $\tau = t_1 - t_2$, since

$$R_x(t_1,t_2) = R_x(t_1 - t_2).$$

The main advantage of wide-sense stationarity is that it places the time-series in the context of Hilbert spaces. Let H be the Hilbert

OK, I think I get the **stationary idea**.

$$E = \int_{-\infty}^{\infty} dt\, p(t) = \int_{-\infty}^{\infty} d\omega\, \mathcal{E}(\omega) \tag{32.6}$$

$$E/T = (1/T) \int_{-\infty}^{\infty} dt\, p(t) = (1/T) \int_{-\infty}^{\infty} d\omega\, \mathcal{E}(\omega) = \int_{-\infty}^{\infty} d\omega\, [\mathcal{E}(\omega)/T] = \int_{-\infty}^{\infty} d\omega\, \mathcal{P}(\omega) = P$$

I claim that E/T = P is the "average power of the pulse train" (averaged over time), whereas p(t) is the "instantaneous power in the pulse train" at some specific time. Then $\mathcal{P}(\omega)$ is the spectrum of the time-averaged power in the pulse train, while P is the time-averaged power in the pulse train.

Now we consider this thing again

$$\mathcal{P}(\omega) \equiv \frac{\mathcal{E}(\omega)}{T} = (1/2\pi T)\, |\, X_{\mathbf{pulse}}(\omega)\, |^2 |\, Y''(z)\, |^2$$

$$= (1/2\pi T)\, |\, X_{\mathbf{pulse}}(\omega)\, |^2 \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} y_m^*\, y_n\, e^{i\omega(m-n)T_1} \tag{34.12}$$

I could ensemble-average both sides and that is what I do. But instead, I know that

$$|\, Y''(z)\, |^2 = \frac{T}{T_1}\, R''(z) \qquad // \; r_s = \langle y_m^*\, y_{m+s} \rangle_1$$

$$= \frac{T}{T_1} \sum_{s=-\infty}^{\infty} r_s z^{-s} = \frac{T}{T_1} \sum_{s=-\infty}^{\infty} \langle y_m^* y_{m+s} \rangle_1$$

So I then get this new result I never really looked at before

$$\mathcal{P}(\omega) = (1/2\pi T) \, | X_{pulse}(\omega) |^2 \frac{T}{T_1} \sum_{s=-\infty}^{\infty} \langle y_m^* y_{m+s} \rangle_1$$

$$= (1/2\pi T_1) \, | X_{pulse}(\omega) |^2 \sum_{s=-\infty}^{\infty} \langle y_m^* y_{m+s} \rangle_1$$

$$= \mathcal{P}_{pulse}(\omega) \sum_{s=-\infty}^{\infty} \langle y_m^* y_{m+s} \rangle_1$$

So compare my two formulas first (34.12)

$$\langle \mathcal{P}(\omega) \rangle = \mathcal{P}_{pulse}(\omega) \frac{T_1}{T} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \langle y_m^* y_n \rangle e^{i\omega(m-n)T_1}$$

$$\mathcal{P}(\omega) = \mathcal{P}_{pulse}(\omega) \sum_{s=-\infty}^{\infty} \langle y_m^* y_n \rangle_1$$

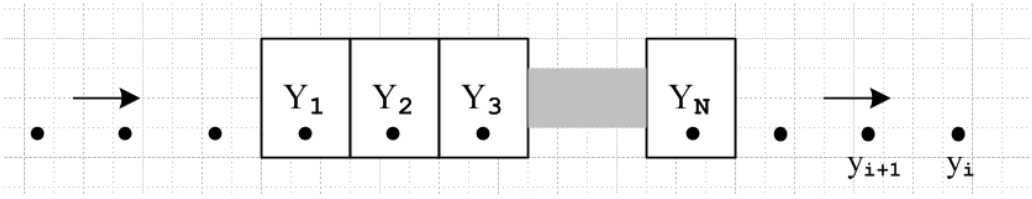
The second form requires no ensemble at all. Are there any random variables involved? Is there a product rule for non-correlation?

Here is a way I think I can make random variables appear.

Scenario #1

Imagine a *very* long sequence whose elements are symbols in some set. Perhaps this sequence is one billion symbols long. The sequence elements might be thought of as amplitudes of a pulse train (a signal) which is marching along in time left to right.

We run the sequence through a machine that looks like this:



The boxes are a sort of cooperating set of toll booths. Rather than take money, they calculate things based on the data of the passing sequence. Each box Y_n separately makes this calculation based on the symbols that pass through its one-symbol-wide "work station",

$$E(Y_n) = (1/K)\,\Sigma_{i=1}^{K}\, y_i$$

This is the mean of the y_i after K symbols have passed through. Another calculation is performed by a cooperative effort of the Y_1 and Y_3 boxes:

$$E(Y_1 Y_3) = (1/K)\,\Sigma_{i=1}^{K}\, y_i y_{i+2} \quad .$$

Box 1 also takes note of the passing y_i values and creates a distribution for the passing symbols by just binning values in bins and dividing on the fly by the number of symbols so far passed by,

$$p_{Y1}(x) \;=\; (1/K)\,\Sigma_{i=1}^{K}\,(\, y_i = x\,) \qquad \text{for all x values in the symbol set .}$$

Meanwhile, box 2 and 3 compute this joint distribution

$$p_{Y2Y3}(x,y) \;= (1/K)\,\Sigma_{i=1}^{K}\;(\,[\, y_i, y_{i+1}] = [x,y]\,)$$

All the boxes cooperate on this large N^{th} order joint distribution calculation

$$p_{Y2Y3\ldots YN}(x,y,z...) \;= (1/K)\,\Sigma_{i=1}^{K}\;(\,[\, y_i, y_{i+1}, \ldots y_{i+N}] = [x,y,z.....]\,)\;.$$

When the long sequence has finally passed through all the boxes, we have accumulated lots of data.

We would like to associate with box Y_1 a random variable we shall call Y_1. The values this variable takes are the values over time of the symbols which sat in its work station position as the sequence passed through. These values, which we assume are real numbers, have a distribution function $p_{Y1}(x)$, so indeed, Y_1 is a "random variable". In fact we have a set of N random variables Y_1 through Y_N and the machine has been computing various expected values of these random variables as shown above.

Perhaps we assume that the sequence is all 0's before and after its official starting and ending points. Then one thing is very clear from the above picture. After the entire sequence has passed through,

$E(Y_n)$ is the same for all n

$E(Y_n Y_{n+s})$ is the same for all n

This is because each box sees the exact same stream of symbols pass by.

If the sequence is some large number I = MN long, then box Y_n and Y_{n+s} compute the following object

$$E(Y_n Y_{n+s}) = (1/I)\,\Sigma_{i=1}^{I}\, y_i y_{i+s}\;.$$

For each value of s, this is makes use of a different pair of boxes. If the data is assembled for a set of s values, the result is the autocorrelation sequence for those s values

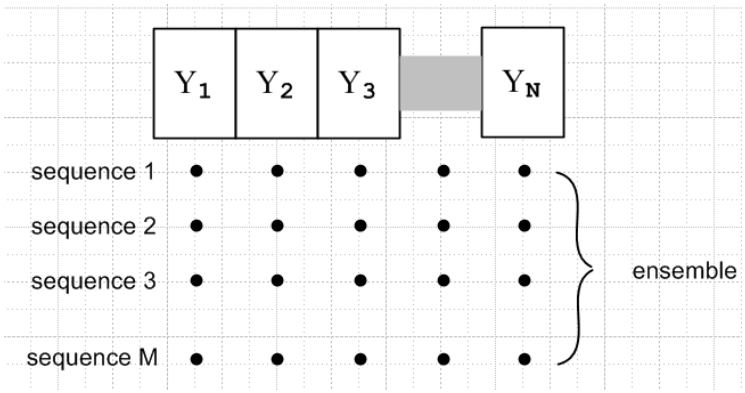
$$r_s \;= E(Y_n Y_{n+s}) \;= (1/I)\,\Sigma_{i=1}^{I}\, y_i y_{i+s} \;\equiv \langle y_n y_{n+s} \rangle_1$$

Here we use the notation $\langle ... \rangle_1$ to indicate an expected value for one sequence. If the random variables Y_n and Y_{n+s} are independent, then

$$E(Y_n Y_{n+s}) = E(Y_n)\, E(Y_n) \qquad \text{or} \qquad \langle y_n y_{n+s} \rangle_1 = \langle y_n \rangle_1 \langle y_{n+1} \rangle_1 = [\langle y_n \rangle_1]^2$$

Scenario #2

Imagine that the very long sequence of Scenario #1 is snipped into M smaller sequences each of length N which then form an ensemble of sequences of length N. Once the ensemble is collected, we consider the action of some new boxes sitting at the top of each column of symbols in the ensemble



With all the data just sitting there, the boxes perform certain calculations. Each box is now associated with the column of symbols directly below it. Here is a typical calculation done by box Y_1 :

$$p_{Y1}(x) = (1/M) \Sigma_{i=1}^{M} (y_i = x) \quad = \text{a vertical column average of values below box } Y_1$$

Since Y_1 takes real values and has an associated probability distribution, it is a random variable, and so are all the other Y_n for similar reasons. Meanwhile, box Y_n computes

$$E(Y_n) = (1/M) \Sigma_{i=1}^{M} y_i \equiv \langle y_i \rangle$$

while boxes n and n+s compute

$$E(Y_n Y_{n+s}) = (1/M) \Sigma_{i=1}^{M} y_i y_{i+s} = \langle y_i y_{i+s} \rangle \ .$$

Each of these "ensemble column expectation values" is indicated by notation <.....> with no label, to distinguish it from the expected value objects $\langle ... \rangle_1$ which appeared in Scenario #1.

Notice that $\langle y_i y_{i+s} \rangle$ is not the autocorrelation function for any of the M sequences. It is a different animal from $\langle y_i y_{i+s} \rangle_1$.

Power Formulas

For Scenario #1, we have this formula for the spectral power density of the long pulse train

$$\mathcal{P}(\omega) = \mathcal{P}_{\mathtt{pulse}}(\omega) \sum_{s=-\infty}^{\infty} \langle y_m^* y_{m+s} \rangle_1 \qquad (*)$$

where the sum is in fact finite since y_m is non-vanishing only in a finite range which specifies our long pulse train studied in this scenario.

For Scenario #2 we have this formula for the power density of just one of the sequences in the ensemble (each sequence is M symbols long)

$$\mathcal{P}(\omega) = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} y_m^* \, y_n \, e^{i\omega(m-n)T_1}$$

The infinite sums are really finite since again y_n is only non-zero for a finite range, T is finite as well, being the duration of the pulse train. We now apply the ensemble average <...> to both sides to get

$$\langle \mathcal{P}(\omega) \rangle = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \langle y_m^* \, y_n \rangle \, e^{i\omega(m-n)T_1}$$

If we now let n = m+s and change from sum on n to sum on s, this becomes

$$\langle \mathcal{P}(\omega) \rangle = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \sum_{s=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \langle y_m^* \, y_{m+s} \rangle \, e^{-i\omega s T_1}$$

What comes next? We cannot claim that $\langle y_m^* \, y_{m+s} \rangle$ is independent of m for the ensemble as we have constructed it! We are not allowed to suddenly use my "useful ensemble" because the ensemble is whatever it is from snipping the original long pulse train.

Suppose M is quite large. Maybe we can then make this claim

$$\langle y_m^* y_{m+s} \rangle_{1,\mathtt{orig\ PT}} \approx \langle y_m^* y_{m+s} \rangle_{1,\mathtt{\ any\ M\ seq}} \quad \text{for } |s| \le 2M$$

where I have now suddenly changed the M pulse train to be a 2M+1 pulse train -M to M. The restriction on s means I cannot replace this in (*) above.

Maybe here is the point.

Claim 1: When I snip the original single long sequence into a set of M still-long sequences, all sequences are "long enough" to have the same "statistics". In this case, statistics for the original sequence will be the same as statistics for the ensemble as a whole, or for any sequence in the ensemble.

So, what exactly do I mean by "statistics" ?

$\langle y_m \rangle_{1,orig} \approx \langle y_m \rangle_{1,seq} = \langle y_m \rangle$

$\langle y_m * y_{m+s} \rangle_{1,orig} \approx \langle y_m * y_{m+s} \rangle_{1,seq} \approx \langle y_m * y_{m+s} \rangle$ for

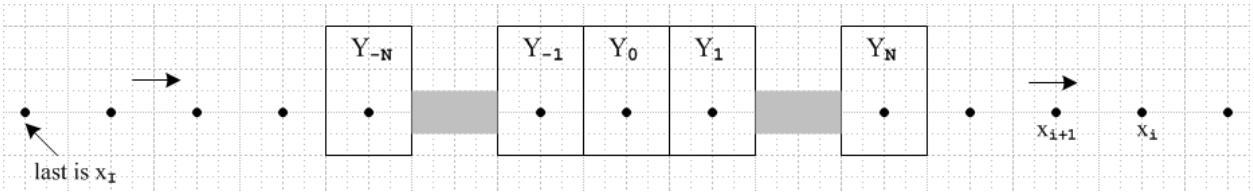
****

Scenario #1

Imagine a *very* long sequence whose elements are symbols in some set. Perhaps this sequence is a billion symbols long. The sequence elements might be thought of as amplitudes of a pulse train (a signal) which is marching along in time left to right. The sequence elements are labeled x_i for i = -I ...+I where I is very large, and temporally the first sequence element is x_{-I} and the last is x_I. The sequence can be written

$[x_{-I}, x_{-I+1}..... x_{-2}, x_{-1}, x_0, x_1, x_2.....x_I \}$

We run the sequence through a machine that looks like this:



The sequence values x_i are represented by the dots which move left to right in time. Each box can hold one sequence value at a time. The boxes cooperatively calculate certain statistical quantities of interest.

Each box Y_n separately makes this calculation based on the symbols that pass through,

$E(Y_n) = (1/K) \Sigma_{i=-I}^{K} x_i$.

This is the mean of the x_i after K symbols have passed through. Another calculation is performed by a cooperative effort of the Y_1 and Y_3 boxes:

$E(Y_1 Y_3) = (1/K) \Sigma_{i=-I}^{K} x_i x_{i+2}$.

Box 1 also takes note of the passing x_i values and creates a distribution for the passing symbols by just binning values in bins and dividing on the fly by the number of symbols so far passed by,

$p_{Y1}(x) = (1/K) \Sigma_{i=1}^{K} (x_i = x)$ for all x values in the symbol set .

Meanwhile, box 2 and 3 compute this joint distribution

$p_{Y2Y3}(x,y) = (1/K) \Sigma_{i=-I}^{K} ([x_i, x_{i+1}] = [x, y])$

N boxes cooperate on this N^{th} order joint distribution calculation

$$p_{Y2Y3\ldots YN}(x,y,z...) = (1/K)\,\Sigma_{i=-I}^{K}\ ([\ x_i, x_{i+1}, \ldots x_{i+N}] = [x,y,z.....]) .$$

When the long sequence has finally passed through all the boxes, we have accumulated lots of data.

We would like to associate with box Y_1 a random variable we shall call Y_1. The values y_i this variable takes are the values over time of the symbols x_i which sat in its work station position as the sequence passed through. These values, which we assume are real numbers, have a distribution function $p_{Y1}(x)$, so indeed, Y_1 is a "random variable". In fact we have a set of 2N+1 random variables Y_{-N} through Y_N and the machine has been computing various expected values of these random variables as shown above.

One thing is very clear from the above picture. After the entire sequence has passed through,

$E(Y_n)$ is the same for all n

$E(Y_nY_{n+s})$ is the same for all n

This is because each box sees the exact same stream of symbols pass by. When the entire sequence has passed through all the boxes, we have for example

$$E(Y_nY_{n+s}) = \frac{1}{2I+1}\ \Sigma_{i=-I}^{I}\ x_ix_{i+s}\ .$$

For each value of s, this makes use of a different pair of boxes. If the data is assembled for a set of s values, the result is the autocorrelation sequence for those s values

$$r_s = E(Y_nY_{n+s}) = (1/I)\,\Sigma_{i=-I}^{I}\ x_ix_{i+s} \equiv \langle y_ny_{n+s}\rangle_1$$

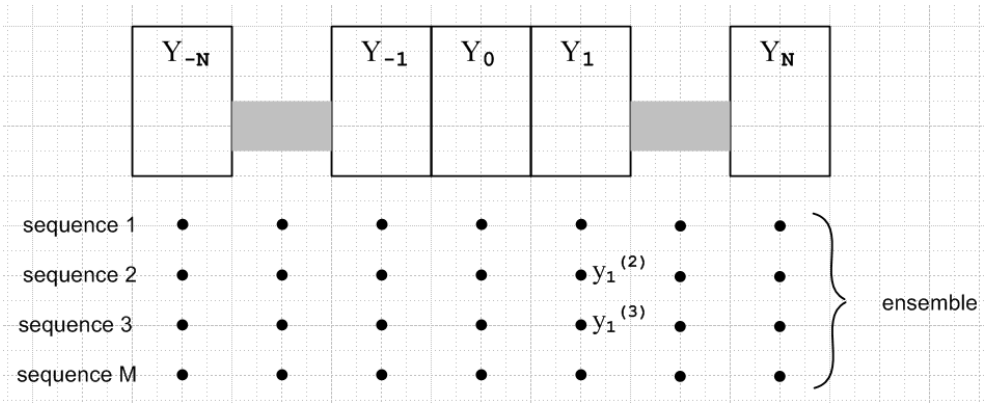
Since there are only 2N+1 boxes, we find that $r_s = 0$ for $|s| > 2N$, so our machine fails to compute autocorrelation sequence elements r_s for $|s| > 2N$.

Here we use the notation $\langle\ldots\rangle_1$ to indicate an expected value for one sequence. If the random variables Y_n and Y_{n+s} are independent, then

$$E(Y_nY_{n+s}) = E(Y_n)\,E(Y_n) \qquad \text{or} \qquad \langle y_ny_{n+s}\rangle_1 = \langle y_n\rangle_1\langle y_{n+1}\rangle_1 = [\langle y_n\rangle_1]^2$$

Scenario #2

Imagine that the very long sequence of Scenario #1 is snipped into M smaller sequences each of length 2N+1 which then form an ensemble of sequences of length 2N+1. Once the ensemble is collected, we consider the action of some new boxes sitting at the top of each column of symbols in the ensemble



With all the data just sitting there, the boxes perform certain calculations. Each box is now associated with the column of symbols directly below it. Here is a typical calculation done by box Y_1 :

$p_{Y1}(x) = (1/M) \Sigma_{i=1}^{M} (y_1^{(i)} = x)$ = a vertical column average of values below box Y_1

Since Y_1 takes real values and has an associated probability distribution, it is a random variable, and so are all the other Y_n for similar reasons. Meanwhile, box Y_n computes

$$E(Y_n) = (1/M) \Sigma_{i=1}^{M} y_n^{(i)} \equiv \langle y_n \rangle$$

while boxes n and n+s compute

$$E(Y_n Y_{n+s}) = (1/M) \Sigma_{i=1}^{M} y_n^{(i)} y_{n+s}^{(i)} = \langle y_n y_{n+s} \rangle \ .$$

Each of these "ensemble column expectation values" is indicated by notation <.....> with no label, to distinguish it from the expected value objects $\langle ... \rangle_1$ which appeared in Scenario #1.

Notice that $\langle y_n y_{n+s} \rangle$ is not the autocorrelation function for any of the M sequences. It is a different animal from $\langle y_n y_{n+s} \rangle_1$.

Claim 1

When we snip the original single long sequence into a set of M still-long sequences, all sequences are "long enough" to have the same "statistics". In this case, statistics for the original sequence will be the same as statistics for the ensemble as a whole, or for any sequence in the ensemble. What exactly do we mean by "statistics" ? Here are some examples,

$$\langle y_m \rangle_{1,orig} \approx \langle y_m \rangle_{1,seq} = \langle y_m \rangle$$

$$\langle y_m * y_{m+s} \rangle_{1,orig} \approx \langle y_m * y_{m+s} \rangle_{1,seq} \approx \langle y_m * y_{m+s} \rangle \qquad |s| < 2N$$

From these identifications, we conclude that $\langle y_m \rangle$ and $\langle y_m * y_{m+s} \rangle$ do not depend on m.

Power Formulas

For Scenario #1, we have this formula for the spectral power density of the long pulse train

$$\mathcal{P}(\omega) = \mathcal{P}_{\mathtt{pulse}}(\omega) \sum_{s=-\infty}^{\infty} \langle y_m^* y_{m+s} \rangle_1 \qquad (*)$$

where the sum is in fact finite since y_m is non-vanishing only in a finite range which specifies our long pulse train studied in this scenario.

For Scenario #2 we have this formula for the power density of just one of the sequences in the ensemble (each sequence is 2N+1 symbols long and $T = (2N+1)T_1$)

$$\mathcal{P}(\omega) = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \sum_{n=-N}^{N} \sum_{m=-N}^{N} y_m^* \, y_n \, e^{i\omega(m-n)T_1}$$

We now apply the ensemble average <...> to both sides to get

$$\langle \mathcal{P}(\omega) \rangle = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \sum_{n=-N}^{N} \sum_{m=-N}^{N} \langle y_m^* \, y_n \rangle \, e^{i\omega(m-n)T_1}$$

If we now let n = m+s and change from sum on n to sum on s, this becomes

$$\langle \mathcal{P}(\omega) \rangle = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \sum_{s=-M}^{M} \sum_{m=-M}^{M} \langle y_m^* \, y_{m+s} \rangle \, e^{-i\omega s T_1}$$

According to claim **, quantity $\langle y_m^* \, y_{m+s} \rangle$ is independent of m, so the sum can be written

$$\langle \mathcal{P}(\omega) \rangle = \mathcal{P}_{\mathtt{pulse}}(\omega) \frac{T_1}{T} \sum_{s=-M}^{M} \langle y_m^* \, y_{m+s} \rangle \, e^{-i\omega s T_1} \sum_{m=-M}^{M} 1$$

$$= \mathcal{P}_{\mathtt{pulse}}(\omega) \sum_{s=-M}^{M} \langle y_m^* \, y_{m+s} \rangle \, e^{-i\omega s T_1}$$

which is then the same as Scenario #1 result since using ***.

Question regarding AMI. We found there that

$$\alpha_{m,n} \equiv \langle y_m y_n \rangle = (-p^2/a) \, a^{|m-n|} \qquad m \neq n \qquad // \; a \equiv (1-2p)$$

$$\beta \equiv \langle y_n^2 \rangle = p \; . \qquad (37.8)$$

where p is the probability of a 1 being coded. If p = 1/2, then a = 0 and we get

$$\alpha_{m,n} \equiv \langle y_m y_n \rangle = (-(1/4)/0)\, 0^{|m-n|} \qquad m \neq n$$
$$\alpha_{m,n} \equiv \langle y_m y_{m+s} \rangle = (-(1/4)/0)\, 0^{|s|} \qquad s \neq 0$$

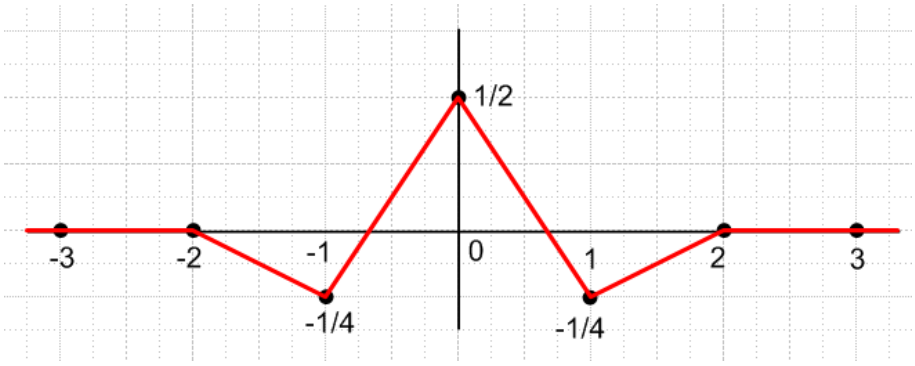
For s = ± 1 we get the zeros exactly cancelling and

$$\langle y_m y_{m+s} \rangle = -1/4$$

For larger s we get

$$\langle y_m y_{m+s} \rangle = 0$$

and for z = 0 we get $\langle y_n^2 \rangle = p = 1/2$. Let's plot this thing!



This is the first I ever recall seeing this thing. What happens if p = 0.8? Then a = 1-1.6 = -0.6 and so

$$\alpha_{m,m+s} \equiv \langle y_m y_{m+s} \rangle = (-(0.8)^2\, (-0.6)^{|s|\, -\, 1} \qquad s \neq 0$$

$$\beta \equiv \langle y_n^2 \rangle = p = 0.8$$

then $\langle y_m y_{m+s} \rangle = -\, 0.64 * (-0.6)^{|s|\, -\, 1}$

Maple on this says

```
for s from -4 to 4 do r[s] = -0.64*(-1.6)^abs(s) od;
```

$$r_{-4} = -4.194304$$

$$r_{-3} = 2.62144$$

$$r_{-2} = -1.6384$$

$$r_{-1} = 1.024$$

$$r_0 = -.64$$

$$r_1 = 1.024$$

$$r_2 = -1.6384$$

$$r_3 = 2.62144$$

$$r_4 = -4.194304$$

I thing this r_s is growing with s in both directions, something very new to me. Do I have a normalization problem?