This is a rewrite of temp55 that crashed and burned early on and led to a massive update of FT.

2.5 Basic Probability Theory, MLS Spectral Power Density, and Autocorrelation

(a) Basic Probability Theory

The probability of *events* A and B both being true is denoted by P(A∩B), or P(A,B). This is known as the joint second-order **probability density function (pdf).** A related concept is P(A|B) meaning the conditional probability that A is true given that B is true. Here are the relations:

$$P(A,B) = P(A\cap B) = P(A|B)\,P(B) = P(B|A)\,P(A) \ . \tag{2.5.1}$$

If events A and B are *statistically independent* in second order, one gets:

$$P(A,B) = P(A\cap B) = P(A)\,P(B) \quad \Leftrightarrow \quad P(A|B) = P(A), \quad P(B|A) = P(B) \ . \tag{2.5.2}$$

The result P(A|B) = P(A) means that the probability of A being true is not influenced by whether or not B is true.

Similarly one can define joint pdf's of third order and beyond. For example, statistical independence for a third-order pdf is indicated by,

$$P(A,B,C) = P(A\cap B\cap C) = P(A)P(B)P(C) \ . \tag{2.5.3}$$

A possible "event" is that a "random variable" X takes some value x. In this case, instead of writing things like P(A) = P(X = x), we abbreviate and say P(X=x) = p(x). Similarly P(X=x, Y=y) = p(x,y), so the position of the argument of p(....) indicates which random variable one is talking about. This concept applies for the values x being either continuous (such as the reals) or discrete (such as GF(p)). The joint pdfs are now written p(x,y) and p(x,y,z) and so on. If the random variables X and Y are **statistically independent**, then *all* the pdf's must completely "factor". For example, p(x,y,z) = p(x)p(y)p(z).

Any pdf is **normalized** to 1 since the probability of all possible events is 1. Thus,

$$\int\int \ldots .\ p(x,y\ldots .)\,dx\,dy\ldots \ = 1 \qquad \text{or} \qquad \Sigma_{ij\ldots .}\ p(x_i,y_j\ldots .) \ = 1 \ . \tag{2.5.4}$$

An example is that $\int p(x)dx = 1$ or $\Sigma_i p(x_i) = 1$.

If f(X,Y...) is some arbitrary function of some random variables X.Y..., one can define the **expectation value** of f as follows:

$$E(f(X,Y,...)) = \int\int .... f(x,y...)\, p(x,y....)\, dx\, dy.... \text{ or } \Sigma_{ij....} f(x_i,y_j...)\, p(x_i,y_j....) . \quad (2.5.5)$$

Cases of interest to us will be the following

$$E(X) = \int x\, p(x)\, dx \text{ or } \Sigma_i x_i\, p(x_i) = \text{"the } \mathbf{mean}\text{", often written } \mu_x \quad (2.5.6)$$

$$E((X-\mu_x)(Y-\mu_y)) = \int\int (x-\mu_x)(y-\mu_y)\, p(x,y)\, dx\, dy \text{ or } \Sigma_{i,j} (x_i-\mu_x)(y_j-\mu_y)\, p(x_i,x_j)$$

$$= \text{"the } \mathbf{covariance}\text{ of X and Y"} = \text{cov}(X,Y) . \quad (2.5.7)$$

If the joint second-order pdf factors, so that $p(x,y) = p(x)p(y)$, or equivalently $p(x_i,x_j) = p(x_i)p(y_j)$, then one says that the random variables X and Y are **uncorrelated**. Since the higher joint pdf's might not factor, this does not imply statistical independence. Thus we have only one direction here:

Fact: X and Y statistically independent $\Rightarrow$ X and Y are uncorrelated (2.5.8)

We can now identify three different statements:

Fact: X and Y are uncorrelated $\Leftrightarrow$ $E(XY) = E(X)E(Y)$ $\Leftrightarrow$ $\text{cov}(X,Y) = 0$. (2.5.9)

Proof: We show this in the continuous notation but everything applies also in the discrete notation. X and Y uncorrelated means $p(x,y) = p(x)p(y)$ which at once means $E(XY) = E(X)E(Y)$ from the definitions shown above:

$$E(XY) = \int\int x\, y\, p(x,y)\, dx\, dy = \int\int x\, y\, p(x)p(y)\, dx\, dy = [\int x\, p(x)\, dx]\, [\int y\, p(y)\, dy] = E(X)E(Y) .$$

The covariance is

$$\text{cov}(X,Y) = \int\int (x-\mu_x)(y-\mu_y)\, p(x,y)\, dx\, dy = \int\int (x-\mu_x)(y-\mu_y)\, p(x)p(y)\, dx\, dy$$

$$= [\int (x-\mu_x)\, p(x)\, dx\,] [\int (y-\mu_y)\, p(y)\, dy\,] .$$

But each factor vanishes for the reason shown below, and thus $\text{cov}(X,Y) = 0$:

$$[\int (x-\mu_x)\, p(x)\, dx\,] = \int x\, p(x)\, dx - \mu_x \int p(x)\, dx = \mu_x - \mu_x * 1 = 0 . \quad \mathbf{QED}$$

Now suppose X and Y are the *same* random variable. Then we have, for example,

$$p(x,y) = p(x)\delta(x-y) \qquad \text{or} \qquad p(x_i,y_j) = p(x_i)\delta_{i,j} \ . \tag{2.5.10}$$

The reason is that if X and Y are the same random variable, they cannot take different values. If X,Y and Z are all the same variable, then we would have

$$p(x,y,z) = p(x)\delta(x-y)\delta(x-z) \qquad \text{or} \qquad p(x_i,y_j,z_k) = p(x_i)\,\delta_{i,j}\,\delta_{i,k} \tag{2.5.11}$$

Note that $\delta(x-y)\delta(x-z) = \delta(x-y)\delta(y-z) = \delta(x-z)\delta(y-z)$ and $\delta_{i,j}\,\delta_{i,k} = \delta_{i,j}\,\delta_{j,k} = \delta_{i,k}\,\delta_{j,k}$.

The main application of this concept relates to the **variance** and **standard deviation** of X:

$$\text{cov}(X,X) = \int\int (x-\mu_x)\,(y-\mu_x)\,p(x,y)\,dx\,dy = \int\int (x-\mu_x)\,(y-\mu_x)\,p(x)\delta(x-y)\,dx\,dy$$

$$= \int (x-\mu_x)^2\,p(x)\,dx \qquad \text{or} \qquad \Sigma_i\,(x_i-\mu_x)^2\,p(x_i)$$

$$\equiv \text{var}(X) = \text{"the variance of X"}$$

$$= [\sigma(X)]^2 = \text{"the standard deviation squared"} \ . \tag{2.5.12}$$

Notice that this quantity could in theory vanish if $p(x) = \delta(x - x_1)$ so that the entire pdf is concentrated at a single value.

$$\text{var}(X) = \int (x-\mu_x)^2\,p(x)\,dx = \int (x-\mu_x)^2\,\delta(x-x_1)\,dx = (x_1-\mu_x)^2 = (x_1-x_1)^2 = 0 = \sigma(X) \ .$$

A pdf of this form is not very interesting and normally one has $\sigma(X) > 0$.

Fact: $\text{var}(X) = \sigma^2(X) = E(X^2) - \{E(X)\}^2$ (2.5.13)

Proof: $E(X^2) - \{E(X)\}^2 = \int x^2 p(x)\,dx - \{\int x\,p(x)\,dx\}^2 = \int x^2 p(x)\,dx - \mu_x^2$

$$\text{var}(X) = \int (x-\mu_x)^2\,p(x)\,dx = \int x^2 p(x)\,dx - 2\mu_x \int x\,p(x)\,dx + \mu_x^2 \int p(x)\,dx$$

$$= \int x^2 p(x)\,dx - 2\mu_x\,\mu_x + \mu_x^2\,1 = \int x^2 p(x)\,dx - \mu_x^2 \ . \qquad \textbf{QED}$$

There is one more object of interest which is the **correlation function** of X and Y

$$\text{corr}(X,Y) \equiv \text{cov}(X,Y) / [\sigma(X)\,\sigma(Y)] \ . \tag{2.5.14}$$

This is just a rescaled version of the covariance. We may now extend our fact above :

Fact: X and Y are uncorrelated $\Leftrightarrow$ $E(XY) = E(X)E(Y)$ $\Leftrightarrow$ $\text{cov}(X,Y) = 0$ $\Leftrightarrow$ $\text{corr}(X,Y) = 0$ (2.5.15)

To be consistent with notation in other documents, we shall use this alternate **<...> notation** for expectation values,

$$< f(x,y...)> \equiv E(f(X,Y,...)) \ . \qquad (2.5.16)$$

Normally the notation <...> means an average of something, and expectation values are averages.

(b) Application to Sequences

We shall now apply these probability concepts to a **sequence** $\{a_n\}$. A single such sequence, whether finite or infinite, has some expectation value for a_n. We could count the number of occurrences of each symbol in the sequence $N(a_n)$ and divide that by the length of the sequence N, to get $p(a_n) = N(a_n)/N$ as the probability of occurrence of symbol a_n. Then

$$<a_n>_1 \equiv \Sigma_n a_n \, p(a_n) = (1/N) \, \Sigma_n \, a_n N(a_n) = E_1(a_n)$$

would be the mean value of the symbols in this one sequence. The notation $<..>_1$ is used to indicate this kind of average.

We think of each position in the sequence as corresponding to a random variable A_i which can take values in some set of values, and the value that A_i takes is called a_i. Then, for example,

$$<a_n> \equiv E(A_n) = \Sigma_n a_n \, p(a_n) \qquad (2.5.17)$$

$$<a_n^2> \equiv E(A_n^2) = \Sigma_{n,m} \, a_n^2 \, p(a_n)$$

$$<a_n a_m> \equiv E(A_n A_m) = \Sigma_{n,m} \, a_n \, a_m \, p(a_n, a_m)$$

$$<a_n a_m> = <a_n><a_m> = <a_n>^2 \qquad \text{// from (2.5.15), only if } A_n \text{ and } A_m \text{ are uncorrelated}$$

where all sums are over the entire sequence. And Fact (2.5.13) says

$$\text{var}(A_n) = \sigma^2(A_n) = E(X^2) - \{E(X)\}^2$$

$$= <a_n^2> - <a_n>^2 \equiv \sigma_n^2 \qquad (2.5.18)$$

Question: Does $\langle a_n \rangle$ or $\langle a_n^2 \rangle$ depend on n ? Recall that $p(a_n) = P(A_n = a_n)$ so *in theory* we could have $P(A_3 = a_n) \neq P(A_5 = a_n)$, say. In our treatment of sequences we assume that there is nothing distinguishing about a particular symbol position other than its value, so we would have $P(A_3 = a_n) = P(A_5 = a_n)$ for positions 3 and 5 in any sequence. So the answer is no, $\langle a_n \rangle$ and $\langle a_n^2 \rangle$ do not depend on n, they are just averages over the sequence. On the other hand, when we write $\langle a_n a_m \rangle = \langle a_n a_{n+k} \rangle$, this could depend on the distance k between two symbol positions, as it does in the AMI line code (see FT Section 37), but $\langle a_n a_{n+k} \rangle$ does not depend on n. When the A_n and A_m variables are uncorrelated, then $\langle a_n a_m \rangle = \langle a_n \rangle \langle a_m \rangle = \langle a_n \rangle^2$ and in this case we find that $\langle a_n a_m \rangle$ depends on neither n nor m, and $\langle a_n a_{n+k} \rangle$ depends on neither n nor k.