Aalto University

T-61.3040 Statistical Signal Modeling

Autumn 2011

Amaury Lendasse & Yoan Miche

Today's Menu (27.9)

- Appetizer: Reminder about mean, (auto)correlation, density function, jointly distributed RVs...
- Starter: On Autocorrelation
- Main course: Stationarity and Wide Sense Stationarity
- Dessert: Ergodicity

Cumulative distribution function, etc.

- Let's have the usual real-valued random variable (RV) x
- Cumulative distribution function (cdf): $F_x(a) = P(x \leq a)$
- If F_x der. and the int. is F_x, then probability density function (pdf): $f_x(a) = \frac{d}{da} F_x(a)$

Mean, Variance, etc.

- For a discrete RV, Mean or expected value: $E(x) = \sum_k a_k P(x = a_k)$
- If the pdf exists, then also: $E(x) = \int_{-\infty}^{\infty} a f_x(a) da$
- Variance is the mean-square value of $x - E(x)$:

$$\mathrm{Var}(x) = E\left[(x - E(x))^2\right] = \int_{-\infty}^{\infty} [a - E(x)]^2 f_x(a) da$$

- Also: $\mathrm{Var}(x) = E(x^2) - E^2(x)$

Jointly distributed RV, moments, etc.

- With two RVs $x(1)$ and $x(2)$, joint distribution function is $F_{x(1),x(2)}(a,b) = P(x(1) \leq a, x(2) \leq b)$
- And joint density function $f_{x(1),x(2)}(a,b) = \frac{\partial^2}{\partial a \partial b} F_{x(1),x(2)}(a,b)$
- Correlation: $r_{xy} = E(xy^*)$
- Covariance: $c_{xy} = \mathrm{Cov}(x,y) = E\left[(x-m_x)\left(y-m_y\right)^*\right] = E(xy^*) - m_x m_y^*$
- Correlation coefficient: $\rho_{xy} = \frac{E[(x-m_x)(y-m_y)^*]}{\sigma_x \sigma_y} = \frac{E(xy^*) - m_x m_y^*}{\sigma_x \sigma_y}$

Independence, uncorrelation, etc.

- Two RVs are statistically independent if $f_{x,y}(a,b) = f_x(a) f_y(b)$
- They are uncorrelated (not independent!) if $E(xy^*) = E(x)E(y^*)$. Also then $r_{xy} = m_x m_y^*$ and then $c_{xy} = 0$

On to the Autocorrelation

For discrete RV $x(n)$

- Mean: "Average value of the process as a function of n": $m_x(n) = E(x(n))$
- Variance: "Average square deviation of the process away from the mean": $\sigma_x^2(n) = E\left[|x(n) - m_x(n)|^2\right]$
- Autocovariance: $c_x(k,l) = E\left[(x(k) - m_x(k))(x(l) - m_x(l))^*\right]$
- Autocorrelation: $r_x(k,l) = E\left[x(k)x^*(l)\right]$

On to the Autocorrelation

- If $k = l$, autocovariance is variance
- Autocovariance and autocorrelation are related:

$$c_x(k,l) = r_x(k,l) - m_x(k)m_x^*(l)$$

- So, if our two processes $x(k)$ and $x(l)$ are zero mean, autocovariance=autocorrelation
- If we talk of two different processes $x(k)$ and $y(l)$, then we talk of cross-covariance and cross-correlation (defined similarly)

Some important property of the autocorrelation

- Say you have a nice signal $x(n)$ that you measure
- Your instruments are nowehere near perfect
- You will have some noise $w(n)$ in your measurements
  $y(n) = x(n) + w(n)$
- Assume that noise is well-behaving (usual assumption), i.e.
  $E(w(n)) = 0, \forall n$ and $E(x(k)w^*(l)) = 0$

Some important property of the autocorrelation

- Then, the autocorrelation of $y(n)$ is the sum of the autocorrelations of $x(n)$ and $w(n)$
- $r_y(k,l) = E\left[y(k)y^*(l)\right] = E\left[\left(x(k)+w(k)\right)\left(x(l)+w(l)\right)^*\right]$
- Meaning

$$\begin{aligned} r_y(k,l) &= E\left[x(k)x^*(l)\right] + E\left[w(k)w^*(l)\right] \\ &\quad + E\left[x(k)w^*(l)\right] + E\left[w(k)x^*(l)\right] \\ &= r_x(k,l) + r_w(k,l) \end{aligned}$$

- Autocorrelation and autocovariance tell you about the degree of linear dependence between your two RVs

An example: The Harmonic Process

- Found often in radar and sonar signal processing
- Real-valued harmonic process $x(n) = A \sin(n\omega_0 + \phi)$ with A and $\omega_0 \in \mathbb{R}$ and ϕ uniformly distributed over $[-\pi, \pi]$
- That is,

$$f_\phi(a) = \begin{cases} 1/2\pi & \text{if } -\pi \leq a \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

An example: The Harmonic Process

- Mean of the process $x(n)$ is

$$m_x(n) = \int_{-\infty}^{\infty} A \sin(n\omega_0 + a) f_\phi(a) da = \int_{-\pi}^{\pi} \frac{1}{2\pi} A \sin(n\omega_0 + a) da = 0$$

- Autocorrelation $r_x(k,l) = E\left[x(k)x^*(l)\right]$ is (using trigonometric transformations)

$$r_x(k,l) = \frac{1}{2} A^2 E\left[\cos\left((k-l)\omega_0\right)\right] - \frac{1}{2} A^2 E\left[\cos\left((k+l)\omega_0 + 2\phi\right)\right]$$

- Therefore $r_x(k,l) = \frac{1}{2} A^2 \cos\left((k-l)\omega_0\right)$

An example: The Harmonic Process

- So, the mean is a constant, and for the autocorrelation we have $r_x(k, l) = r_x(k - l, 0)$
- This means that mean and autocorrelation do not change if we shift $x(n)$ in time
- This is called a *Wide Sense Stationary* process

Stationarity: some definitions

- Stationarity: "Statistical time-invariance"
- L-th order stationarity: "$x(n)$ and $x(n+k)$ have the same L-th order joint density function"
- For example: $x(n)$ is first order stationary if $f_{x(n)}(a) = f_{x(n+k)}(a)$
- In this case, first order statistics will be independent of time, i.e. $m_x(n) = m_x$ and $\sigma_x^2(n) = \sigma_x^2$

Stationarity: some definitions

- Second order stationarity: "The second order joint density function $f_{x(n_1),x(n_2)}(a_1, a_2)$ only depends on $n_2 - n_1$ and not on n_1 and n_2 individually"
- $x(n)$ is second order stationary if

$$f_{x(n_1),x(n_2)}(a_1, a_2) = f_{x(n_1+k),x(n_2+k)}(a_1, a_2)$$

- Second order stationary $\Longrightarrow$ first order stationary
- Second order stationarity: second order statistics invariant to time shifts

Stationarity: some definitions

- Example with the autocorrelation

$$\begin{aligned} r_x(k,l) &= \int_{-\infty}^{\infty} ab f_{x(k),x(l)}(a,b)\,da\,db \\ &= \int_{-\infty}^{\infty} ab f_{x(k+n),x(l+n)}(a,b)\,da\,db \\ &= r_x(k+n, l+n) \end{aligned}$$

- So $r_x(k,l) = r_x(k-l,0)$. $k-l$ is called the *lag*
- We write then $r_x(k-l,0)$ as $r_x(k-l)$
- If $\forall L > 0$, $x(n)$ is L-th order stationary, $x(n)$ is said to be *stationary in the strict sense*

Stationarity: some definitions

- *Wide-Sense Stationarity* (WSS) (for a random process $x(n)$) requires all three:
  - Mean of the process is a constant, $m_x(n) = m_x$
  - Autocorrelation $r_x(k, l)$ depends only on the difference $k - l$ (and not k and l separately)
  - The variance $c_x(0)$ is finite
- Wide-Sense Stationarity (WSS) is *weaker* than second-order stationarity (constraints on moments, not on density functions directly)
- For a Gaussian process, wide-sense=strict sense
- *An interesting property*: If the input $v(n)$ of a system is WSS, then the output process $y(n)$ is WSS if σ_y^2 is finite

Joint stationarity

- As for single processes, we can define stationarity for two or more processes
- Two processes $x(n)$ and $y(n)$ are jointly WSS if $x(n)$ and $y(n)$ are WSS and if the cross-correlation $r_{xy}(k, l)$ only depends on the difference $k - l$

Autocorrelation of WSS processes

Properties of the autocorrelation of a WSS process:

- *Symmetry*: $r_x(k) = r_x^*(-k)$
- *Mean-square value*: $r_x(0) = E\left[|x(n)|^2\right] \geq 0$ (the autocorrelation at lag 0 is equal to the Mean-square value of the process)

Autocorrelation of WSS processes

Properties of the autocorrelation of a WSS process (cont.):

- *Maximum value*: $|r_x(k)| \leq r_x(0)$ (magnitude of autocorrelation at lag k is upper bounded by the value at lag 0)
- *Periodicity*: If $r_x(k_0) = r_x(0)$ for some k_0, then $r_x(k)$ is *periodic* with period k_0. Also $E\left[|x(n) - x(n-k_0)|^2\right] = 0$ and $x(n)$ is said *mean-square periodic*

Autocorrelation of WSS processes

Example of periodic process:

- The previous $x(n) = A\cos(n\omega_0 + \phi)$
- Autocorrelation sequence is $r_x(k) = \frac{1}{2}A^2\cos(k\omega_0)$
- So if we have $\omega_0 = 2\pi/N$
  - $r_x(k)$ is N-periodic and
  - $x(n)$ is mean-square periodic

Handy formulation: Using matrices

- With a WSS random process $x(0), x(1), \ldots, x(N)$
- Note it $\mathbf{x} = [x(0), x(1), \ldots, x(N)]^T$ (^{T} for the transpose, ^{H} for the Hermitian transpose)
- Now we can have the outer product

$$\mathbf{x}\mathbf{x}^H = \begin{bmatrix} x(0)x^*(0) & x(0)x^*(1) & \cdots & x(0)x^*(N) \\ x(1)x^*(0) & x(1)x^*(1) & \cdots & x(1)x^*(N) \\ \vdots & \vdots & \vdots & \vdots \\ x(N)x^*(0) & x(N)x^*(1) & \cdots & x(N)x^*(N) \end{bmatrix}$$

Handy formulation: Using matrices

- With which we define the autocorrelation matrix $\mathbf{R}_x = E\left[\mathbf{x}\mathbf{x}^H\right]$ which looks like

$$\mathbf{R}_x = \begin{bmatrix} r_x(0) & r_x^*(1) & r_x^*(2) & \cdots & r_x^*(N) \\ r_x(1) & r_x(0) & r_x^*(1) & \cdots & r_x^*(N-1) \\ r_x(2) & r_x(1) & r_x(0) & \cdots & r_x^*(N-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x(N) & r_x(N-1) & r_x(N-2) & \cdots & r_x(0) \end{bmatrix}$$

- Similarly, the autocovariance matrix $\mathbf{C}_x$ is defined as

$$\mathbf{C}_x = E\left[(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^H\right]$$

Handy formulation: Using matrices

- Which means we have the relationship

$$\mathbf{C}_x = \mathbf{R}_x - \mathbf{m}_x \mathbf{m}_x^H$$

- Who knows what a Toeplitz matrix is?
- $\mathbf{R}_x$ is a Hermitian Toeplitz matrix: These have interesting properties (understand "nice for some calculations")

About Toeplitz matrices

- Have the form

$$
\begin{aligned}
\mathbf{M} &= \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-N+1} \\ a_1 & a_0 & a_{-1} & \cdots & a_{-N+2} \\ a_2 & a_1 & a_0 & \cdots & a_{-N+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N-1} & a_{N-2} & a_{N-3} & \cdots & a_0 \end{bmatrix} \\
&= \text{Toep}(a_0, \dots, a_{N-1})
\end{aligned}
$$

- That is, $M_{i,j} = M_{i+1,j+1}$
- Only has $2N - 1$ degrees of freedom

About Toeplitz matrices

- Solving a linear system $\mathbf{Mx} = \mathbf{b}$ is much easier (and faster) than in the general case
- Matrix decompositions are easier (LU, QR...)
- Additions of Toeplitz matrices in $O(N)$...
- Some even nicer properties, given Toeplitz+other properties
- Not covered here, see Wikipedia or any good linear algebra book :)

Back to autocorrelation matrix

- So we have for a WSS process x:
  - $\mathbf{R}_x$ is *Hermitian Toeplitz*
  - If x is real-valued, $\mathbf{R}_x$ is *symmetric Toeplitz*
  - $\mathbf{R}_x$ is *nonnegative definite*
  - The eigenvalues of $\mathbf{R}_x$ are *real-valued and nonnegative*
- These conditions allow "quick" checking whether a process is not WSS

Let's take an example

- From previously, the Harmonic process, we saw that the autocorrelation sequence is

$$r_x(k) = \frac{1}{2}A^2 \cos(k\omega_0)$$

- So, the autocorrelation matrix is

$$\mathbf{R}_x = \frac{1}{2}A^2 \begin{bmatrix} 1 & \cos(\omega_0) \\ \cos(\omega_0) & 1 \end{bmatrix}$$

Let's take an example

- And eigenvalues of $\mathbf{R}_x$ are $\lambda_{1,2} = 1 \pm \cos(\omega_0) \geq 0$
- And determinant of $\mathbf{R}_x$ is

$$\det(\mathbf{R}_x) = 1 - \cos^2 \omega_0 = \sin^2 \omega_0 \geq 0$$

- $\mathbf{R}_x$ is nonnegative definite, and if $\omega_0 \neq 0$ or π, then $\mathbf{R}_x$ positive definite

Ergodicity: Motivation

- Ensemble averages such as mean and autocorrelation give informations on the properties of the process being observed
- Important to be able to estimate them properly
- Might be difficult if only low number of sample realizations of the process available
- Ergodicity: “Each member of the process has the same statistical behavior (over time) as the entire process”

Ergodicity: Motivation

- For example, say we want to estimate the mean $m_x(n)$ of a random process $x(n)$
- If we have a large number L of realizations of the process, $\{x_1(n), x_2(n), \dots, x_L(n)\}$, we can consider the average, to estimate the mean:

$$\hat{m}_x^{(1)}(n) = \frac{1}{L} \sum_{i=1}^{L} x_i(n)$$

- Unlikely that we get enough realizations of the process to get this average. . .

Ergodicity: Motivation

- If we have only one realization $x(n)$, how about using the sample mean (average over time)

$$\hat{m}_x^{(2)}(N) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)$$

- Fine, but requires conditions on $x(n)$

Ergodicity: An example

- Take, for example $x(n) = A$, with A a RV that can take values -1 or 1 with equal probability
- The true mean is $m_x = 0$:

$$E\left[x(n)\right] = E\left[A\right] = 0$$

- But, the sample mean is (with equal probability for each)

$$\hat{m}_x^{(2)}(N) = \pm 1$$

Ergodicity in the mean

- So, $\hat{m}_x(N)$, the sample mean, will not converge to the true mean
- Now this is why we want ergodicity of our process: will give the insurance that we can use averages over time to estimate mean, autocorrelation and such
- Ergodicity: "Each member of the process has the same statistical behavior (over time) as the entire process"
- We will be looking at ergodicity of WSS processes only

Ergodicity in the mean

- *Ergodicity in the mean*:
  - If the sample mean $\hat{m}_x(N)$ of a WSS process converges to the real mean m_x in the mean-square sense, then the process is *ergodic in the mean* and we have

$$\lim_{N\to\infty} \hat{m}_x(N) = m_x$$

- *Converging in the mean square sense* for the mean requires:
  - $\lim_{N\to\infty} E\left[\hat{m}_x(N)\right] = m_x$ and
  - $\lim_{N\to\infty} \mathrm{Var}\left[\hat{m}_x(N)\right] = 0$

Ergodicity in the mean: Theorems

Two theorems to check ergodicity in the mean:

1. *Mean Ergodic Theorem 1*: With $x(n)$ a WSS process of autocovariance $c_x(k)$. It is necessary and sufficient for $x(n)$ to be ergodic in the mean that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} c_x(k) = 0$$

2. *Mean Ergodic Theorem 2*: With $x(n)$ a WSS process of autocovariance $c_x(k)$. It is sufficient for $x(n)$ to be ergodic in the mean that $c_x(0) < \infty$ and that

$$\lim_{k\to\infty} c_x(k) = 0$$

Ergodicity

An example with the previous $x(n) = A$, with $P(A = 1) = 0.5$ and $P(A = -1) = 0.5$

- Then $\mathrm{Var}(A) = 1$ and $c_x(k) = 1$
- Therefore
$$\frac{1}{N}\sum_{k=0}^{N-1} c_x(k) = 1$$
and this process is *not ergodic in the mean*

Ergodicity

Another example:

- The previous $x(n) = A\sin(n\omega_0 + \phi)$
- If $\omega_0 \neq 0$, autocovariance is $c_x(k) = \frac{1}{2}A^2\cos(k\omega_0)$
- Using trigonometric transformations, we can get

$$\frac{1}{N}\sum_{k=0}^{N-1} c_x(k) = \frac{A^2}{2N}\frac{\sin(N\omega_0/2)}{\sin(\omega_0/2)}\cos\left[(N-1)\omega_0/2\right] \longrightarrow 0$$

  for $N \to \infty$

- If $\omega_0 = 0$, then $x(n) = A\sin\phi$ which is *not ergodic in the mean*, because $c_x(k) = \frac{A^2}{2}$

Autocorrelation Ergodicity

Finally, *autocorrelation ergodicity*:

- With the same idea, a process is said *autocorrelation ergodic* if

$$\lim_{N\to\infty} E\left[\left|\hat{r}_x(k) - r_x(k)\right|^2\right] = 0$$

  that is, the estimate of the autocorrelation $\hat{r}_x(k)$ (from the sample mean estimate e.g.) converges in the mean square sense towards the real autocorrelation value.
- For *WSS Gaussian processes*, it is necessary and sufficient that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} c_x^2(k) = 0$$

  for the process to be autocorrelation ergodic

Ergodicity: In practice

In practice:

- Ergodicity is not exactly easy nor practical to assess in real-life cases
- Usually, we will assume ergodicity if we need to estimate mean, autocorrelation, and such, using time averages
- Once we can check if the results are proper, we can assess if that assumption was appropriate or not

Next time

- What happens to the autocorrelation for AR, MA and ARMA processes?
- What is filtering and how do we do that?
- The power spectrum
- Spectral factorization (maybe. . .)