12.714 Computational Data Analysis

Alan Chave (alan@whoi.edu)
Thomas Herring (tah@mit.edu),

http://geoweb.mit.edu/~tah/12.714

Stationary Stochastic Processes

- Class Today:
  - Stochastic Processes
  - Notation
  - Basic Theory
  - Real-valued stationary processes
  - Complex-valued stationary processes
  - Discrete Parameter Stationary Processes
  - Continuous parameter processes
  - Use as models of data

Stochastic Processes

- A stochastic process is a family of random variables indexed by t where t belongs to an index set T.
- A stochastic process is one which a particular sample, called a *realization*, from all possible realizations, called the *ensemble* is selected.
- A stochastic is denoted by: $\boxed{\{X(t) : t \in T\}}$
- The index t is often time but can be other quantities such as distance.
  - If t is continuous, the process is called *continuous parameter (or continuous time)*
  - If t is discrete, the process is called *discrete parameter (or discrete time)*.

Notation

- Notation to be used (and used in PW)
- $\{X_t\}$ discrete parameter, $\{X(t)\}$ continuous parameter
- If a range for t is not specified, then all integers for discrete, and all real axis for continuous
- Multiple stochastic processes will be given separate names (e.g. $\{Y(t)\}$) or an additional index e.g $\{X_{t,j}\}$ and $\{X_{t,k}\}$ or $\{X(t,j)\}$ and $\{X(t,k)\}$ for discrete and continuous parameters.
- Symbol Z will be reserved for complex process such as $Z_t = X_{t,1} + iX_{t,2}$ and i=sqrt(-1). $X_{t,1} + iX_{t,2}$ are real valued stochastic processes.

Basic Theory

- Since a stochastic process is composed of a set of random variables, we can define at a specific t the cumulative probability distribution function (CPDF) $F_t(a) = P[X_t \le a]$, then

$$E\{X_t\} = \int_{-\infty}^{\infty} x dF_t(x) \equiv \mu_t$$
$$\mathrm{var}\{X_t\} = \int_{-\infty}^{\infty} (x-\mu_t)^2 dF_t(x) \equiv \sigma_t^2$$

- This called the Riemann-Stieltjes integral. This form has the advantage that continuous and discrete probability distributions can be used depending on the continuity of the derivatives of $F_t(a)$
- The full stochastic process can be defined by the n-dimensional cumulative probability distribution function

$$F_{t_1,t_2,t_3,\ldots,t_n}(a_1,a_2,a_3,\ldots,a_n) = P[X_{t_1} \le a_1, X_{t_2} \le a_2, X_{t_3} \le a_3 \ldots X_{t_n} \le a_n]$$

Real Value Stationary Processes

- Stationary processes have properties that are invariant with time: Two types
  - *Complete stationarity* (*strong/strict sense*): If the CPDF for all n≥1 and for any $t_1,t_2,\ldots,t_n$ in the index set, and for any τ such that $t_1+\tau,\ldots,t_n+\tau$ are in the index set

$$F_{t_1,t_2,\cdots,t_3}(a_1,a_2,\cdots,a_n) = F_{t_1+\tau,t_2+\tau,\cdots,t_3+\tau}(a_1,a_2,\cdots,a_n)$$

  - *Second-order stationarity* (*weak/wide-sense/ covariance*) all joint first and second moments exist and are the same for the conditions above (ie., expectation and variance are the same).

Stationary processes

- Second-order stationarity: From the definition we have

$$E\{X_t\} \equiv \mu \quad E\{X_t^2\} \equiv \mu'_2 \quad \Rightarrow \text{var}\{X_t\} = \mu'_2 - \mu^2 \equiv \sigma^2$$
constant and independent of t. With shift $\tau = t_1 - t_2$
$$\text{E}\{X_{t_1} X_{t_2}\} = E\{X_0 X_{t_1 - t_2}\}$$

- We define the autocovariance sequence (acvs) for a discrete process as

$$\text{Cov}\{X_{t_1}, X_{t_2}\} = E\{[X_{t_1} - \mu][X_{t_2} - \mu]\} = E\{X_{t_1} X_{t_2}\} - \mu^2$$
$$s_\tau = \text{Cov}\{X_t, X_{t+\tau}\} = \text{Cov}\{X_0, X_\tau\}$$

Stationary procsses: Continuous

- For a continuous process the autocovariance function (acvf) is

$$s(\tau) \equiv \text{cov}\{X(t), X(t+\tau)\} = \text{cov}\{X(0), X(\tau)\}$$
τ is called the lag and is the $|t_1 - t_2|$

- *Some Properties of acvs and acvf*
  - The variance of the process is given by s_0 or S(0)
  - The autocorrelation sequence (acs) and autocorrelation function (acf) are $\rho_\tau = s_t/s_0$ and $\rho(\tau) = s(\tau)/s(0)$ (Difference in engineering literature).
  - ρ_τ and $\rho(\tau)$ are bounded between -1 and +1.
  - Sequence $\{s_\tau\}$ is positive semi-definite and therefore $\{s_\tau\}$ can not be an arbitrary sequence even if all the correlations are bounded in -1 to 1. Strict limits on negative correlations.

Properties of acvs

- The two dimensional variance-covariance matrix for a contiguous portion of the process is called a *Toeplitz* matrix because the elements depend only on the row and column difference. Since the absolute value of the difference is not important, it is a symmetric *Toeplitz* matrix.
- For a process with a Gaussian cpdf, wide-sense stationarity implies strict sense stationarity because the cpdf is defined totally by the first and second moments.

Complex Stationary processes

- Complete processes are similar to real valued ones with some exceptions (actually all the complex formulas are valid for real valued ones but not visa-versa.).
- The differences arise from the definition of covariance.

$$Z_t = X_{t,1} + iX_{t,2} \quad E\{Z_t\} = \mu_1 + i\mu_2 = \mu$$

$$\text{cov}\{Z_{t_1}, Z_{t_2}\} = E\{[Z_{t_1} - \mu]^*[Z_{t_2} - \mu]\} \text{ (* is complex conjucate)}$$

$$\text{cov}\{Z_t, Z_{t+\tau}\} = E\{[Z_t - \mu]^*[Z_{t+\tau} - \mu]\} = s_\tau$$

- Now $s_{-\tau} = s_\tau^*$ and for continuous $s(-\tau) = s^*(\tau)$

Complex processes

- The positive semi-definite condition becomes

$$\boxed{\mathrm{var}\left\{\sum_{j=1}^{n} c_j Z_{t_j}\right\} = \sum_{j=1}^{n}\sum_{k=1}^{n} s_{t_j - t_k} c_j c_k^* \geq 0}$$

- Notice the conjugate on the covariance element
- For complex processes, the covariance matrix is Hermitian Toeplitz because the off-diagonal terms are conjugates.
- Z is Complex Gaussian is its real and imaginary parts are defined by a bi-variate Gaussian distribution

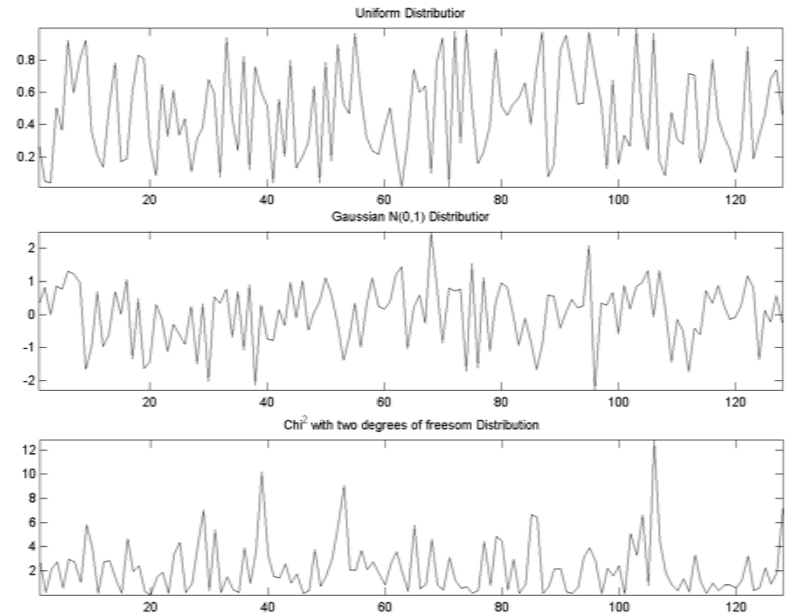
Examples of Stationary processes

- White noise process: $E\{X_t\}=\mu$ and $\mathrm{var}\{X_t\}=\sigma^2$ for all t. Uncorrelated means $\mathrm{cov}\{X_t,T_{t+\tau}\}=0$ for all t and $\tau\neq0$. The acvs is

$$\boxed{s_\tau = \begin{cases} \sigma^2, & \text{if } \tau = 0; \\ 0, & \text{if } \tau \neq 0 \end{cases} \quad \rho_\tau = \begin{cases} 1 \\ 0 \end{cases}}$$

- White noise processes are useful for forming other processes. Uncorrelated noise samples are easy to generate and can be used to generate many other processes.

Examples of White noise



Moving average process

- A process $\{X_t\}$ is called a *q*th order moving average process, MA(q), if it can be expressed as

$$\boxed{X_t = \mu - \theta_{0,q}\varepsilon_t - \theta_{1,q}\varepsilon_{t-1} - \cdots - \theta_{q,q}\varepsilon_{t-q}}$$

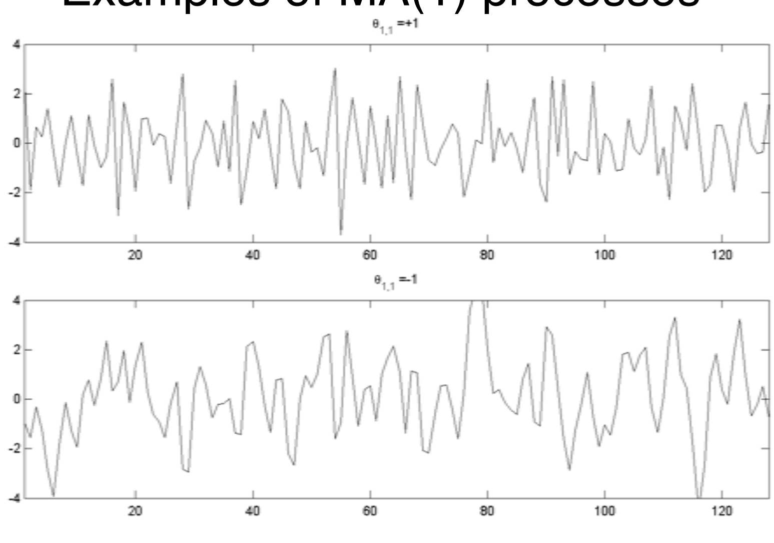
- where μ and $\theta_{j,q}$ are constants ($\theta_{0,q}$=-1 and $\theta_{q,q}\neq 0$) and $\{\varepsilon_t\}$ is a white noise process with zero mean and variance σ_ε^2. Expectation of $\{X_t\}$ is μ. Assume μ=0.

$$\boxed{\operatorname{cov}\{X_t, X_{t+\tau}\} = \sum_{j=0}^{q}\sum_{k=0}^{q}\theta_{j,q}\theta_{k,q}E\{\varepsilon_{t-j}\varepsilon_{t+\tau-k}\} = \sigma_\varepsilon^2\sum_{j=0}^{q-\tau}\theta_{j,q}\theta_{j+\tau,q} \equiv s_\tau}$$

Moving Average Process

- No special restrictions on $\theta_{j,q}$ to ensure stationarity.
- Variance of process given by $\boxed{s_0 = \sigma_e^2 \sum_{j=0}^{q} \theta_{j,q}^2}$
- Examples on next slide show $\theta_{0,1}$ = -1 and $\theta_{1,1}$ = +1 and -1. For these two case $r_1 = -\theta_{1,1} /(1+ \theta^2_{1,1})$ which becomes -1/2 and 1/2

Examples of MA(1) processes



As expected -1 values looks more correlated

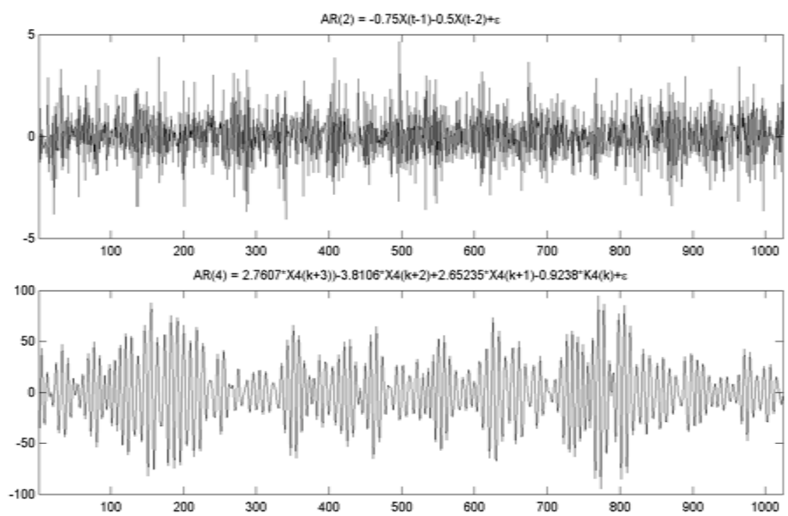
Autoregressive processes

- A process {Xt} with zero mean is *p*th order autoregressive process. AR(p), if it satisfies

$$\boxed{X_t = \phi_{1,p} X_{t-1} + \phi_{2,p} X_{t-2} + \cdots + \phi_{p,p} X_{t-p} + \varepsilon_t}$$

- where $\phi_{1,p},\ldots,\phi_{p,p} \neq 0$ are constants and ε_t is a zero mean white noise sequence with variance σ_ε^2. X_t is a linear combination of previous values plus white noise.
- Not all choices for $\phi_{1,p},\ldots,\phi_{p,p}$ lead to stationary processes.
- If AR(p) is stationary and non-deterministic, then it can expressed as an infinite order MA process. The MA coefficients can be determined from the AR coefficients.

Examples of AR(2) and AR(4)



Noise generated with N(0,1)

Other processes

- Autoregressive moving average: ARMA(p,q) is combination of the two processes. This combination can provide a rich variety process characteristics.
- *Harmonic process*
  - A process $\{X_t\}$ is harmonic if it can be written as

$$X_t = \mu + \sum_{l=1}^{L} D_l \cos(2\pi f_l t + \phi_l)$$

where μ, D_l, f_l are real-valued constants and
ϕ_l are independent rv's with uniform distribution $[-\pi, \pi]$

Harmonic Process

- The $E\{X_t\}=\mu$; The covariance is given by

$$\operatorname{cov}\{X_t, X_{t+\tau}\} = \sum_{l=1}^{L} D_l^2 \cos(2\pi f_l \tau)/2 \equiv s_\tau$$

$$s_0 = \sum_{l=1}^{L} D_l^2 / 2 \equiv \sigma^2$$

- Note that S_τ does not damp as τ goes to infinity.
- Since the phases are fixed once generated, any segment of a harmonic process with enough data to determine all of the parameters, can be used to determine the complete realization.
- All stationary processes can be written as a harmonic process with infinite numbers of terms. (Essential for spectral representation)

Harmonic Processes

- Harmonic processes are only stationary if the phases are independent random variables. (Assumption may be violated by ocean tide studies).
- The $D_l\cos(2\pi f_l t+\phi_l)$ can be separated into a $A_l\cos(2\pi f_t)$ and $B_l\sin(2\pi f_t)$. For this form to represent the same process there are restrictions on the generation of the A and B coefficients. Specially, random generation of A and B will be different because the amplitude will not be constant. The A and B formulation is more general.
- Harmonic processes are the justification for stationary periodic processes

Continuous Parameter Processes

- Processes (except harmonic ones) can be constructed by taking (possible infinite) linear combinations of discrete white noise (see matlab codes for this class).
- When continuous processes are generated with continuous white noise there is a problem: Continuous white noise can not exist.
- Continuous white noise requires infinite power in the process. In practice, a Dirac-delta function provide a useful approximation. The area under the Dirac-delta function is the variance of the white noise. (As the width of the Dirac-delta function goes to zero, its amplitude must go to infinity.

Stationary processes as Models for Data

- Care should taken in making conclusions based on stationarity assumptions when this assumption maybe suspect for a data set.
- Some common problems can corrected.
- *Linear trends*: Given a process $X_t=\alpha+\beta t+Y_t$, stationary process can be generated with:
  - Residuals to least squares fit to line
  - Using differences (mean is then estimate of β). Later we will see that spectral properties of Y_t can be determined for spectum of $Y_{t+1}-Y_t$.
- Always problem here with the lowest frequencies in Y_t being lost in fitting or differencing

Models for data

- Periodic signals: Again fitting can be used if frequencies are know. (Periodic signals do not necessarily violate stationarity so can be often ignored)
- Use of differences (separated by period) can also be used. If not an exact multiple of period, then differencing between several points can be used.
- Some non-stationary cases can be treated by dividing data to segments; variance changes (if known) handled by normalizing data with standard deviation.
- Formulation with multiple noise processes sometimes can be used, if non-stationary ones can be removed.

Summary

- Today we covered:
  - Stochastic Processes
  - Notation
  - Basic Theory
  - Real-valued stationary processes
  - Complex-valued stationary processes
  - Discrete Parameter Stationary Processes
  - Continuous parameter processes
  - Use as models of data
- Next class: Deterministic Spectral Analysis. Chapter 3 of PW