Classic Time Series Analysis

Concepts and Definitions

- Let Y be a random number with PDF *f*

$$Y(t) \sim f(\theta, t)$$

- Define

$$\mu(t) = E[Y(t)]$$

  - $\mu(t)$ is known as the ***trend***

- Define the autocovariance

$$\begin{aligned}\gamma(t,s) &= COV[Y(t), Y(s)] \\ &= E[(Y(t)-\mu(t))\cdot(Y(s)-\mu(s))]\end{aligned}$$

Stationarity

- Strict stationarity: joint PDF of Y does not change depending upon time

$$f(\theta, t_i + s) = f(\theta, t_j + s)$$

- Second-order stationarity

$$\mu(t) = \mu$$

No trend

$$\gamma(t, s) = \gamma(|t - s|)$$

Covariance only a function of difference in time

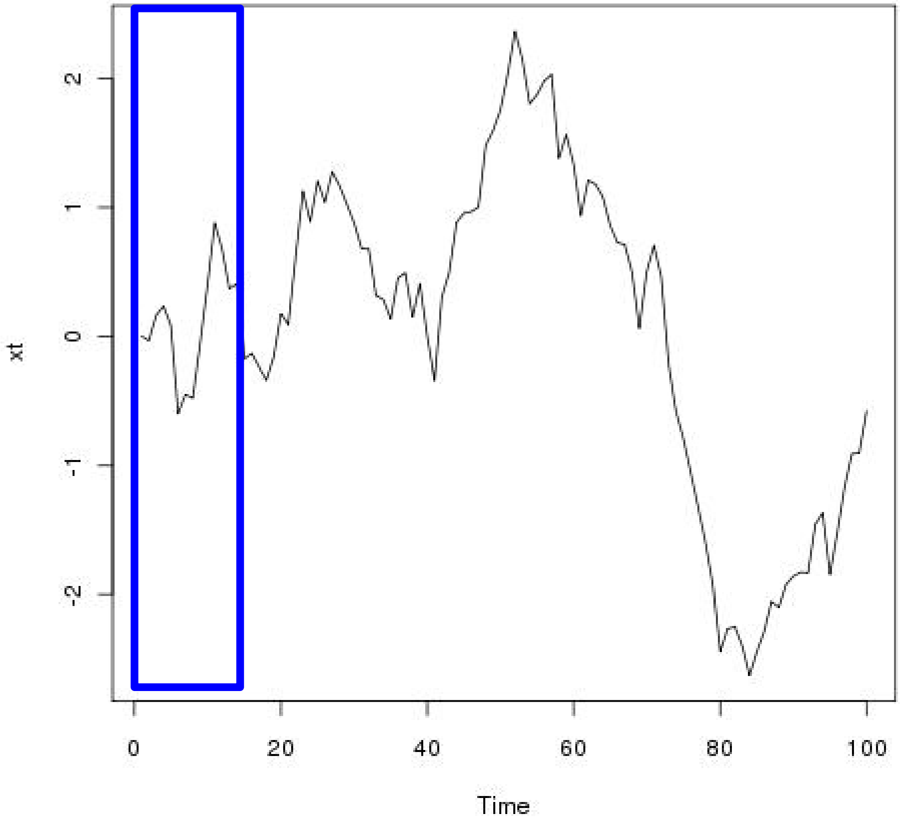
Most common assumption

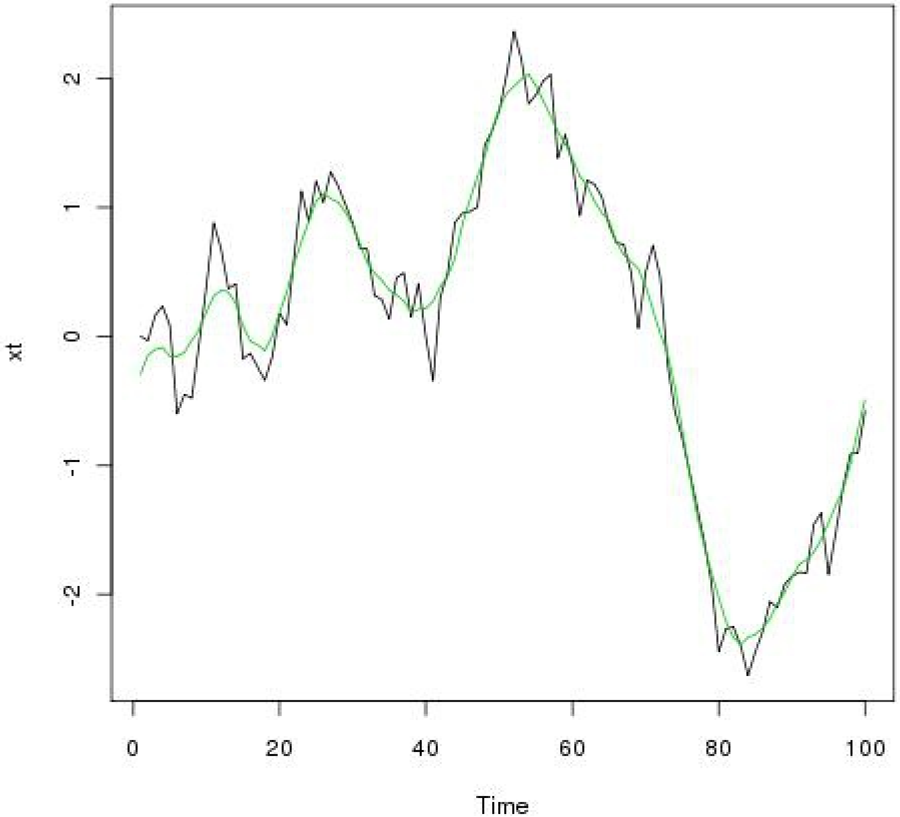
Descriptive Approaches

- Smoothing
- Detrending
- Differencing
- Autocorrelation
- Spectral decomposition (not covered)
  - Power spectra / Fourier transform
  - Wavelet

Smoothing

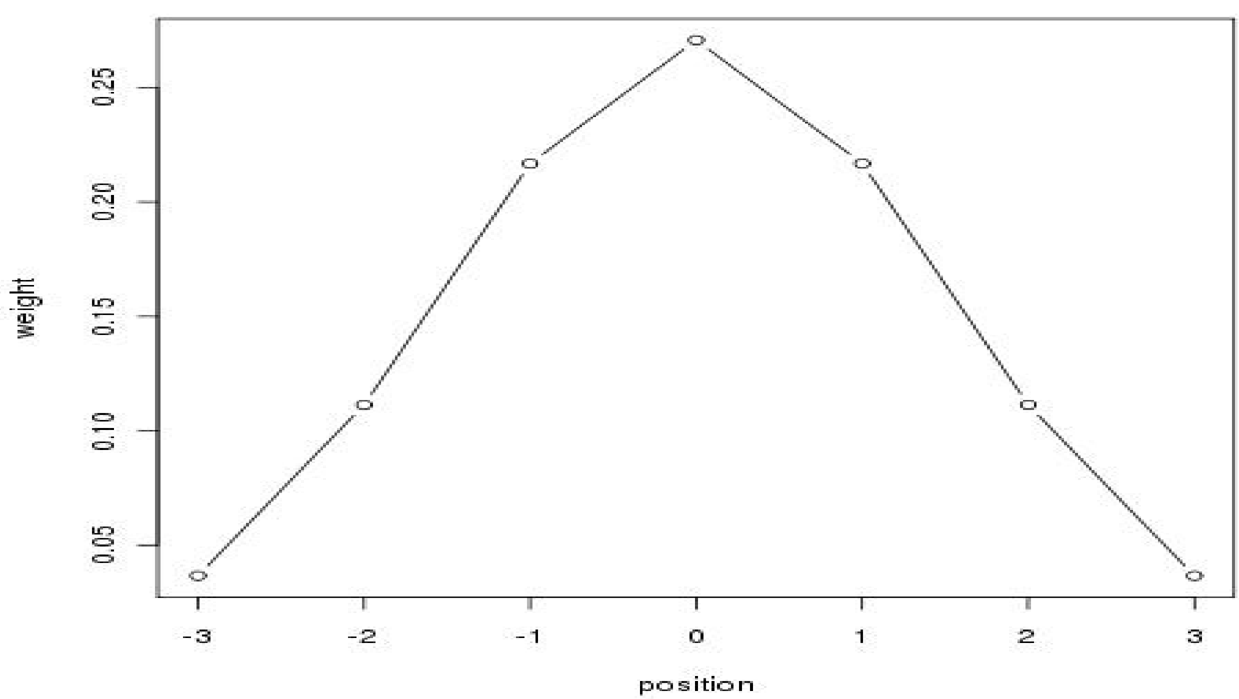
- Moving average
  - Calculated mean within a window





Smoothing

- Weighted moving average
  - Assign weights to different points within window
  - Weights should be symmetric and sum to 1

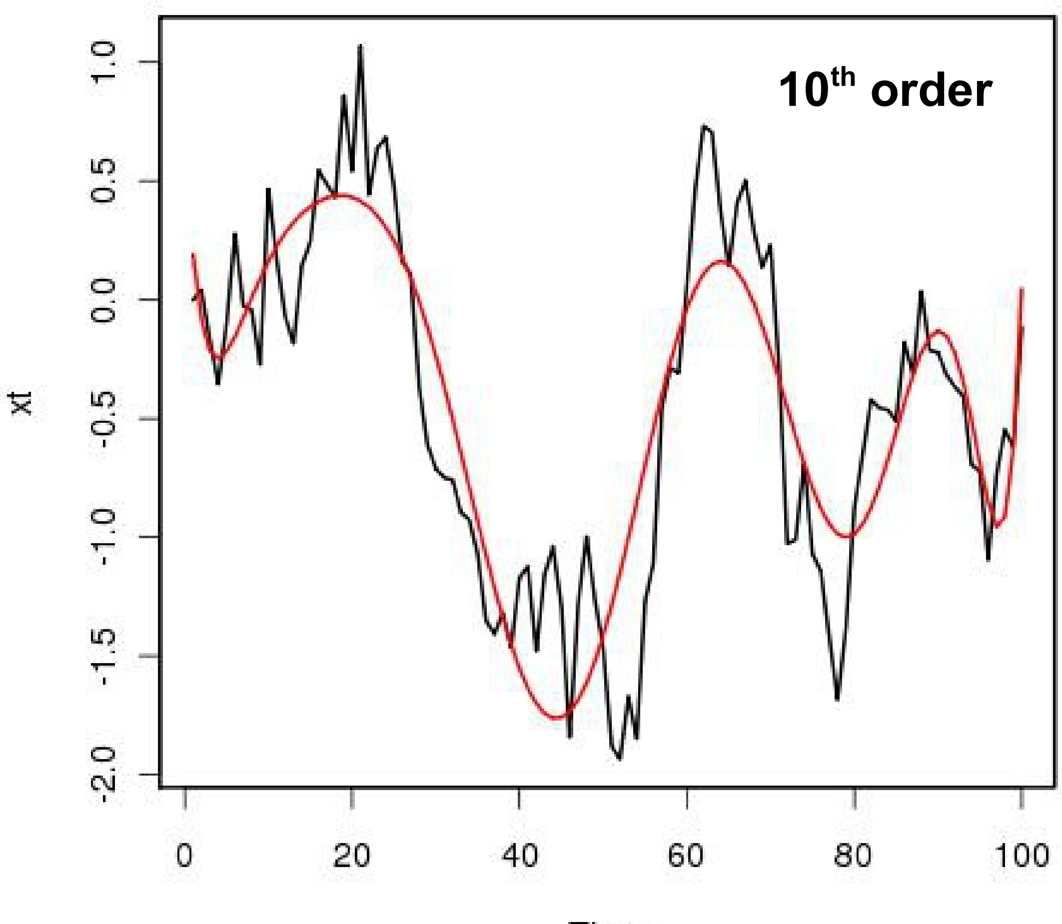


Smoothing

- Filtering
  - Assign weights to different points within window
  - Weights <u>NEED NOT</u> be symmetric and sum to 1
  - Generalization of Weighted Moving Average
  - R function: **filter(x,k)**
    - X = data
    - K = vector of weights (a.k.a. kernel)
      - e.g. k = c(0.1, 0.2, 0.4, 0.2, 0.1)

Smoothing

- Polynomial Regression $$Y(t) = \sum_{0}^{k} \beta_i t^i + \epsilon_t$$
  - R: **lm(X ~ t + t^2 + t^3...)**



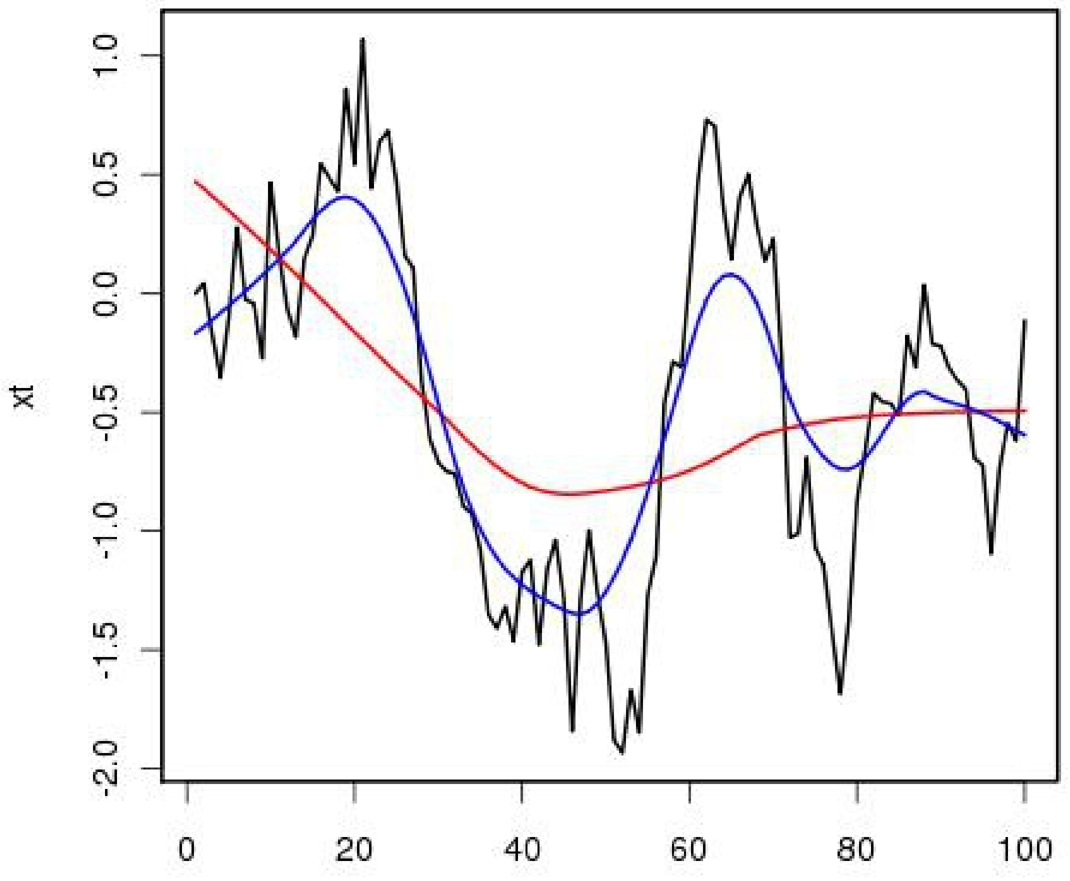
Smoothing

- LOESS / LOWESS (R: **lowess(x)**)
  - Local regression within a moving window
    - W = weighted

Window size

66%

25%

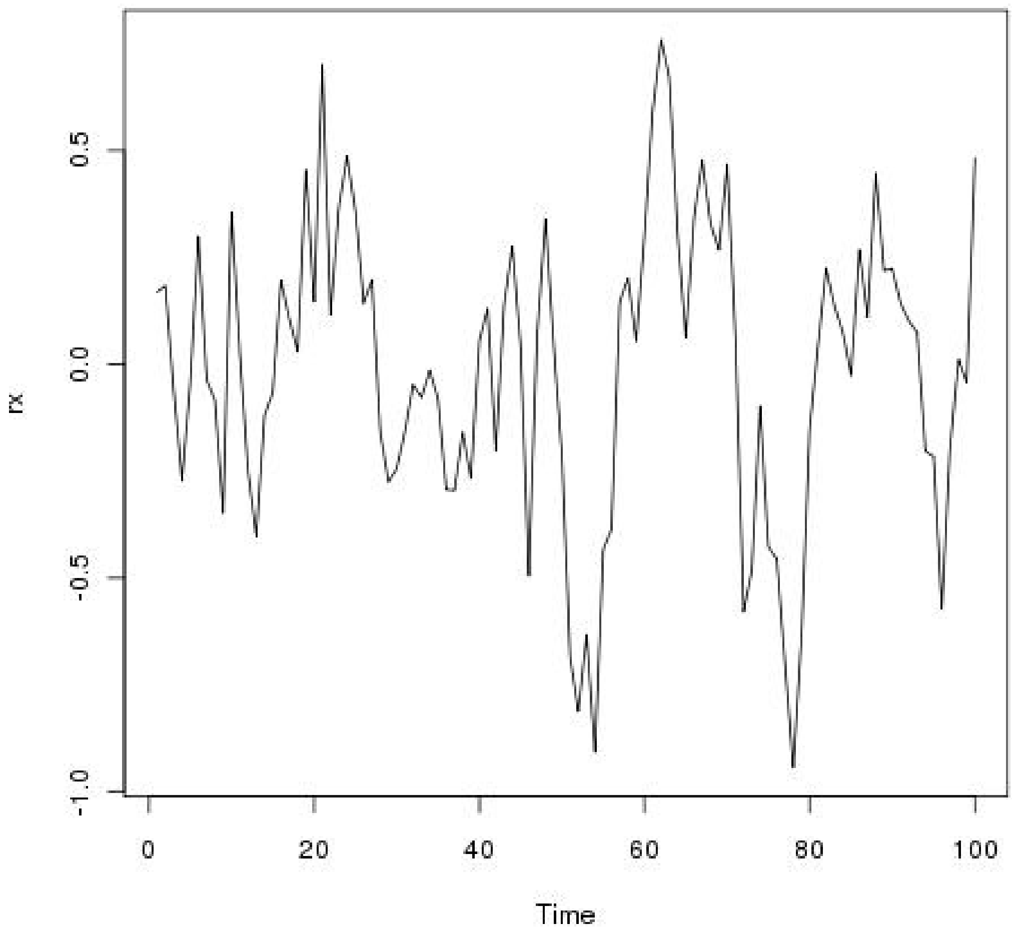


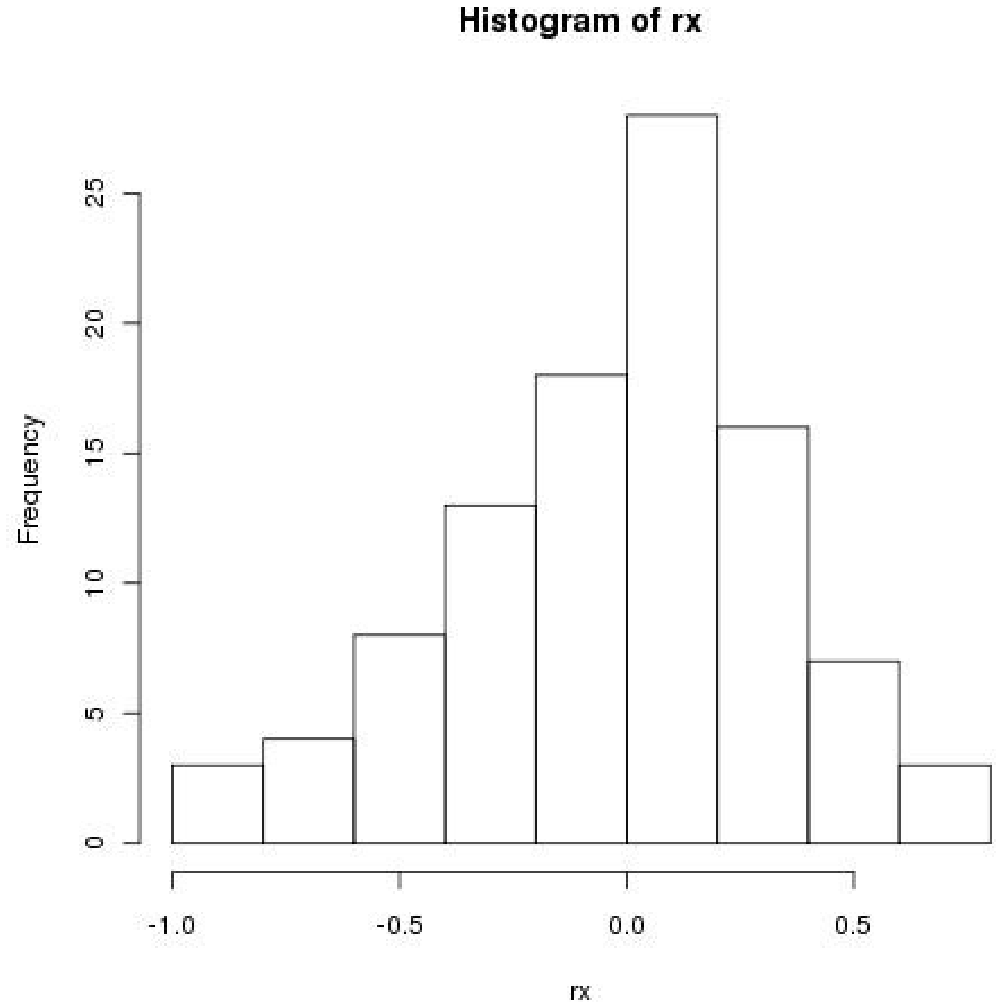
Detrending

- To meet the assumption of stationarity the trends in data need to be removed
- For exploratory purposes:
  - Estimate trend (smoothing)
  - Calculate residuals
  - Analyze residuals as a time-series
- For Analysis:

$$Y(t) = \underbrace{\hat{\mu}(t)}_{\text{Trend}} + \underbrace{\Omega(t)}_{\text{Autocorrelated Error}}$$

LOWESS residuals

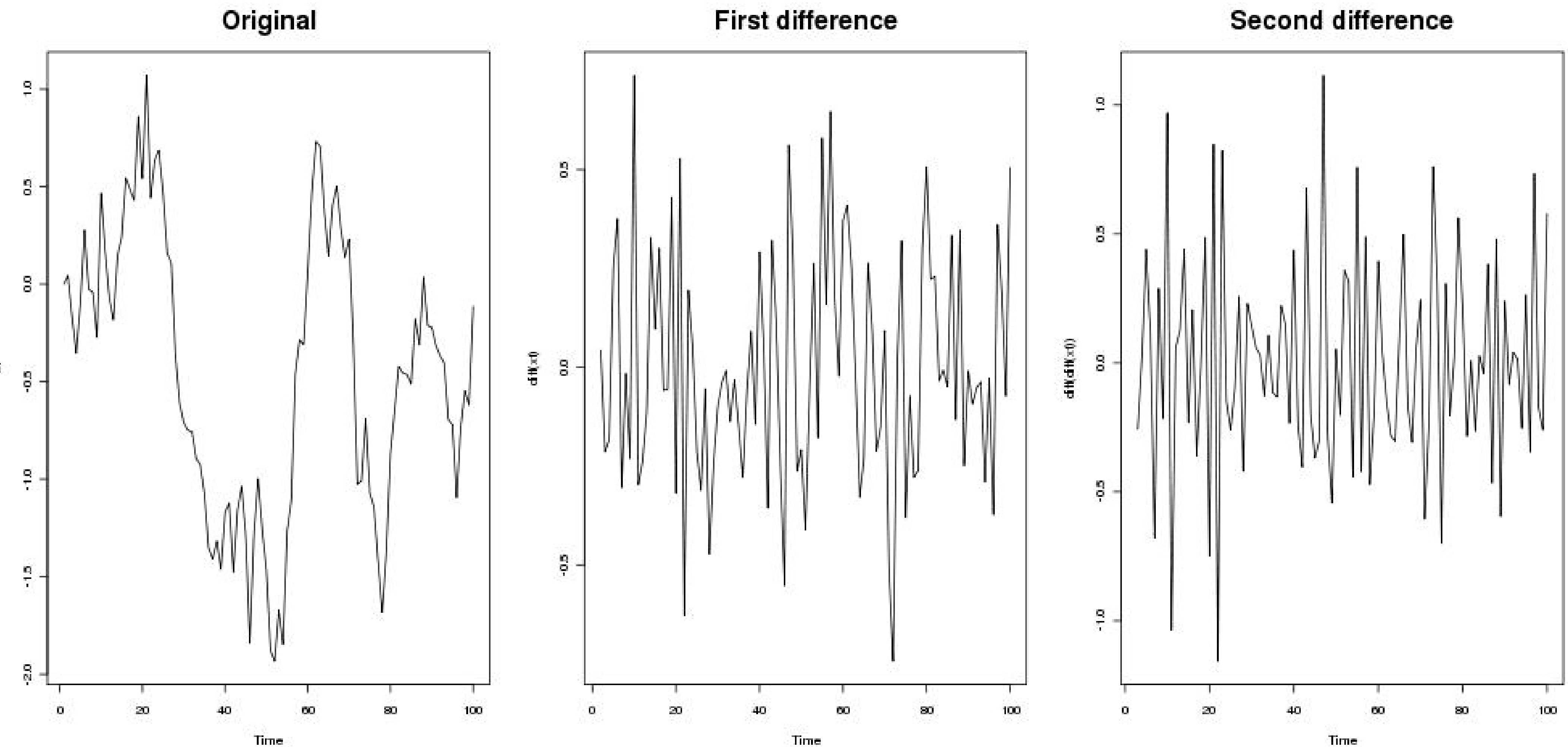




- Goal: stationarity
  - Normal with mean=0
  - Homoskedastic
  - Still autocorrelated is OK

Differencing

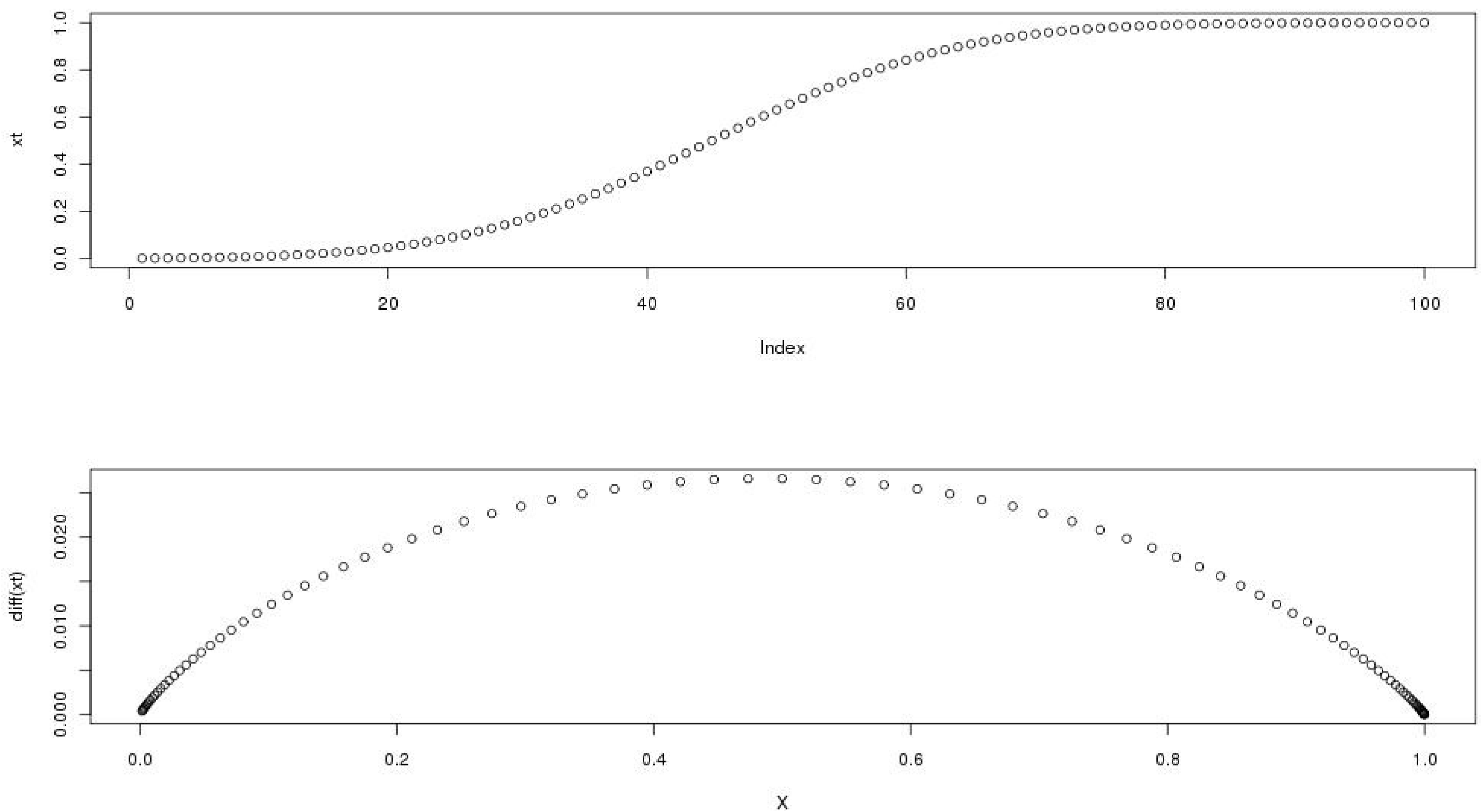
- Can help detrend, increase stationarity
- Can increase understanding of process
  - Often model change in X rather than X(t)
- Sometimes not autocorrelated (Markov)
- Discrete approx to derivative
- First difference $\Delta X = X_t - X_{t-1}$
- Lagged difference $\Delta X = X_t - X_{t-n}$
- Second difference $\Delta X_t - \Delta X_{t-1}$


Original
First difference
Second difference
diff(x)
diff(diff(x))
Time
Time
Time

- R: **diff(x)**

Density Dependence

- dN/dt changes with N
- Plot first difference vs N



Autocovariance

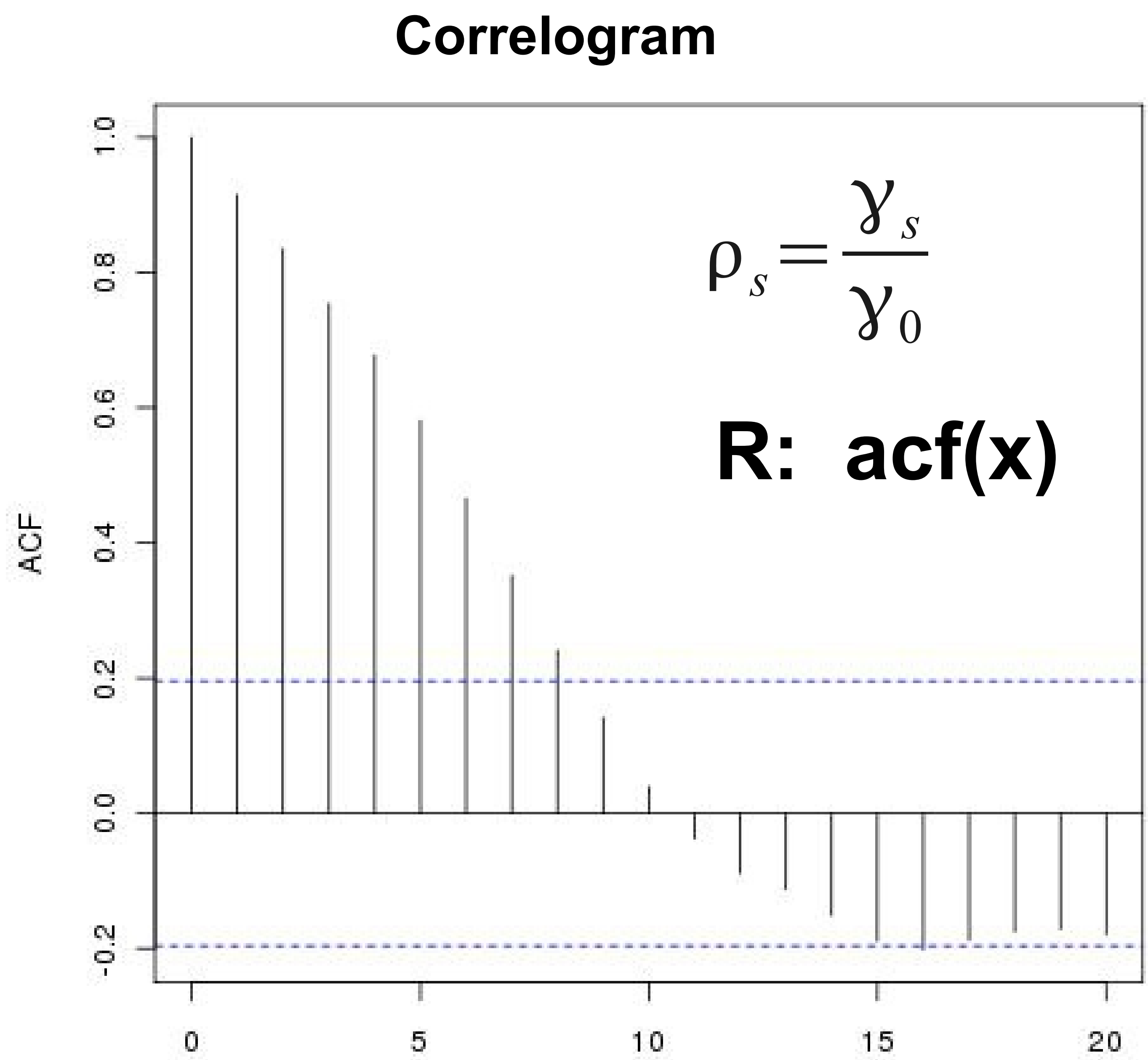
- Covariance between time-series and itself
- If 2nd order stationary, just a function of the lag

$$\gamma(t,s) = \gamma(|t-s|)$$

- Often written as $\gamma_s = Cov[Y_t, Y_s]$
- Special case: $\gamma_0 = Cov[Y_t, Y_t] = Var[Y_t]$

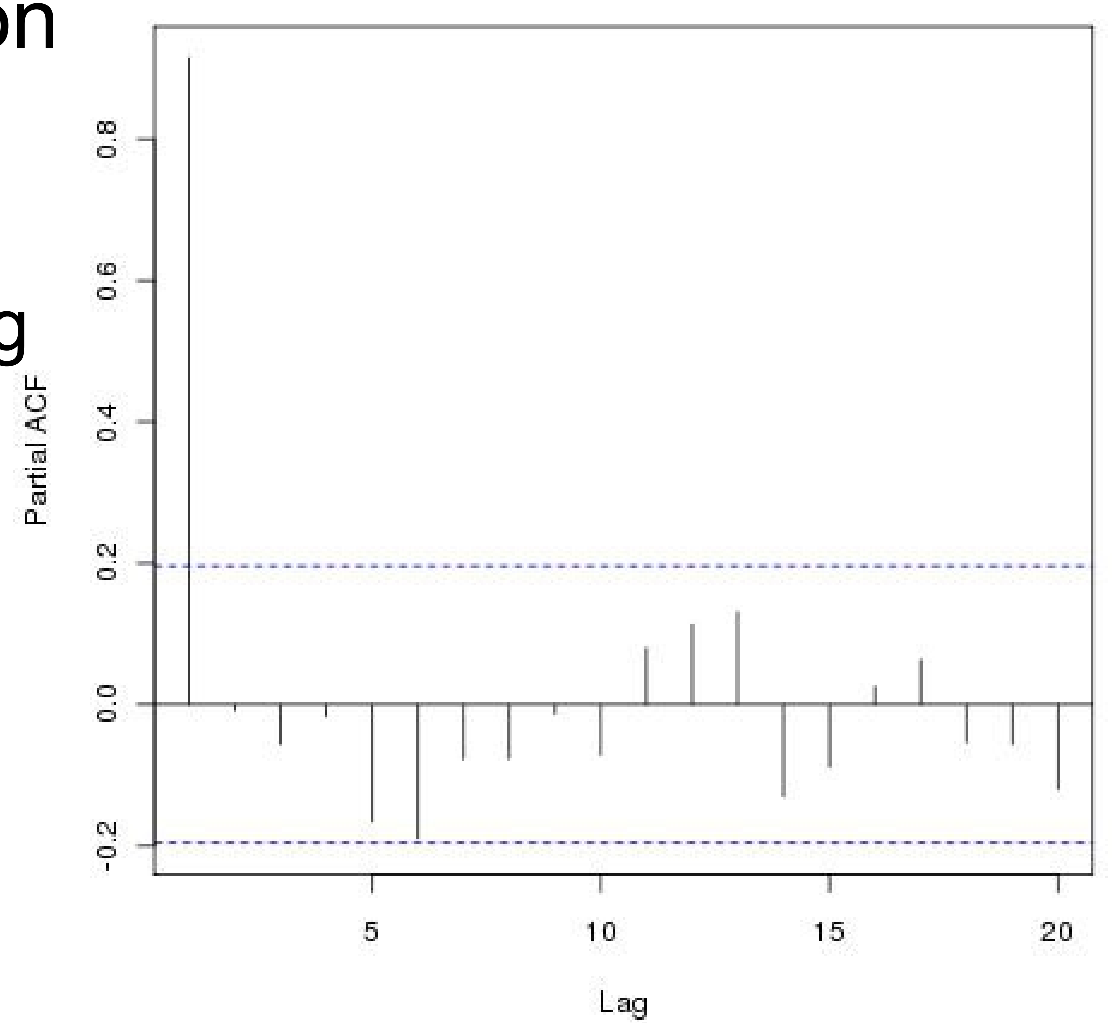
- Two time series can be related by their cross correlation: $\gamma_{XY,s} = Cov[X_t, Y_{t-s}]$

Autocorrelation



Partial and Cross-correlation

- Partial autocorrelation (R: **pacf(x)**)
  - Autocorrelation at lag s after accounting for correlation in lags up to s-1
- Cross-correlation (R: **ccf(x,y)**)
  - Correlation between X(t) and Y(t-s)



Descriptive Approaches

- Smoothing
- Detrending
- Differencing
- Autocorrelation
- Spectral decomposition
  - Power spectra / Fourier transform
  - Wavelet

Classic Time Series Models

- Account for lack of independence in observations
  - Model the temporal error structure
- Make forecasts that account for autocorrelation

$$Y(t) = \hat{\mu}(t) + \Omega(t)$$

Trend **Autocorrelated error**

- Note: remaining slides assume trend = 0

ARIMA model
Autoregressive Integrated Moving Average

- General case for classic frequentist time series
- Contains a number of important special cases
  - AR : Autoregressive models (p)
  - I : Integrated models (d)
  - MA : Moving average models (q)
  - ARMA: Autoregressive moving average
  - Gaussian white noise
- Models are named based on the *order* of the three terms

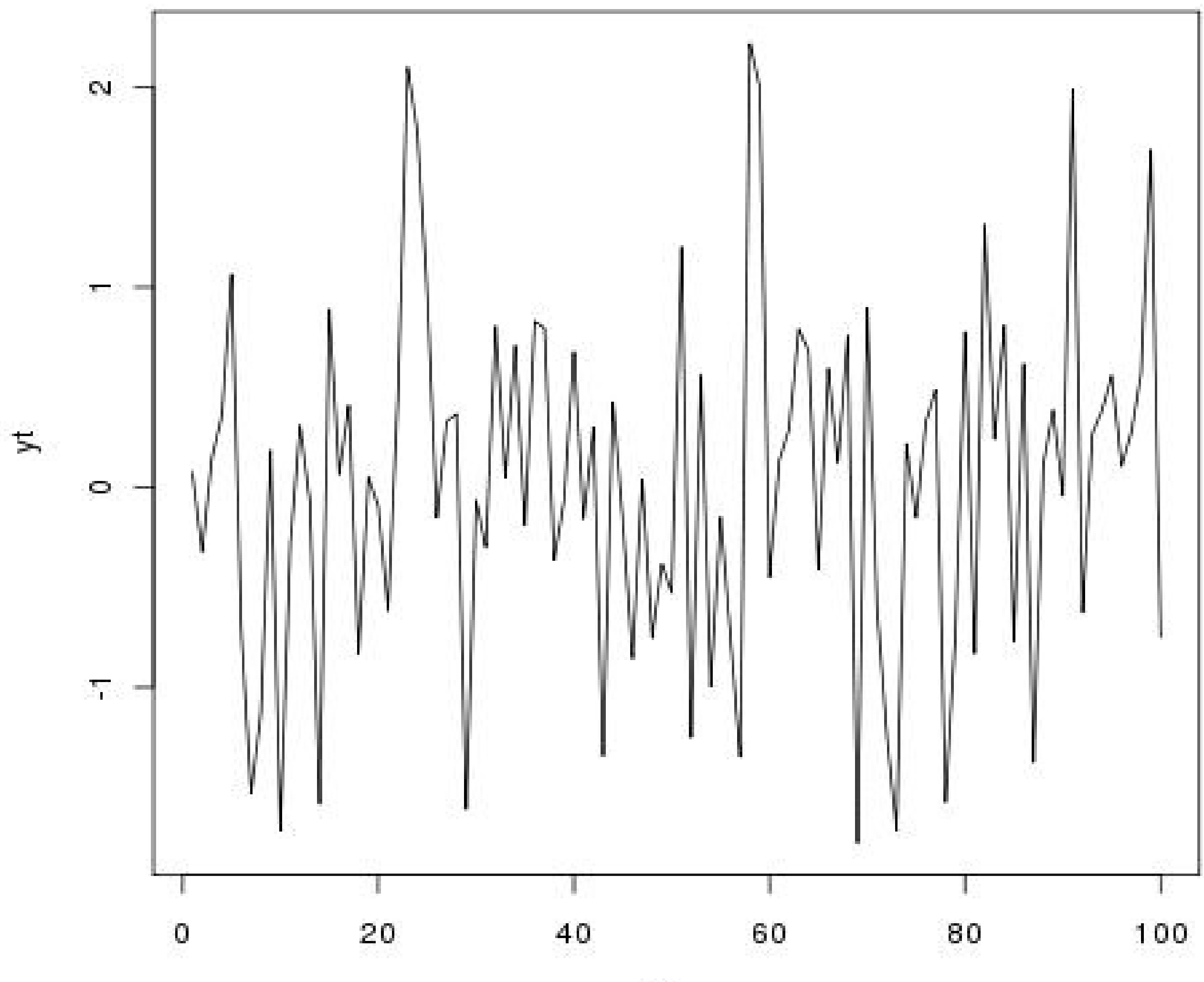
ARIMA(p,d,q)

Gaussian white noise

$$Y_t = \epsilon_t$$

- Mean 0
- Constant variance
- No autocorrelation

ARIMA(0,0,0)



Autoregressive Models: AR(p)

$$Y_t = \sum_{i=1}^{p} \rho_i Y_{t-i} + \epsilon_t$$

- Conceptually like fitting a linear regression against the last *p* values
- AR(1) = first-order Markov process = ARIMA(1,0,0)

$$E[Y_t] = Y_0 \rho^t$$

$$Var[Y_t] = \sigma^2 \sum_{i=0}^{t} \rho^{2i}$$

- If ρ = 1, AR(1) is a random walk
- If ρ = 0, AR(1) = AR(0) = white noise

Covariance matrices

- If |ρ| < 1 then as t -> ∞

$$E[Y_t] = Y_0 \rho^t \rightarrow 0$$

$$Var[Y_t] = \sigma^2 \sum_{i=0}^{t} \rho^{2i} \rightarrow \frac{\sigma^2}{1-\rho^2}$$

- The covariance at lag s then becomes

$$\gamma_s = \sigma^2 \frac{\rho^s}{1-\rho^2}$$

Covariance matrices

- If we have a time series
  $\ldots Y_{t-2}\ Y_{t-1}\ Y_t\ Y_{t+1}\ Y_{t+2} \ldots$
- The covariance with Y_t is
  $$\cdots \gamma_2 \quad \gamma_1 \quad \gamma_0 \quad \gamma_1 \quad \gamma_2 \cdots$$
  $$\frac{\sigma^2}{1-\rho^2}[\ldots \rho^2 \quad \rho \quad 1 \quad \rho \quad \rho^2 \ldots]$$
- Can do the same calculation for every Y_t

Covariance matrix

$$\Sigma = \sigma^2 R$$

$$R = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{t-1} \\ \rho & 1 & \rho & & \\ \rho^2 & \rho & 1 & & \\ \vdots & & & \ddots & p \\ \rho^{t-1} & & & p & 1 \end{bmatrix}$$

AR(1) in models

$$Y_t = \rho Y_{t-1} + \epsilon_t$$
$$\epsilon_t \sim N(0, \sigma^2)$$

$$Y_t = \epsilon_t$$
$$\epsilon_t \sim N(0, \Sigma)$$

$$\mu = 0$$
$$Y_t \sim N(\mu, \Sigma)$$

Process Model [trend]

Data Model [AR(1)]

$$\mu = X\beta + \alpha$$
$$Y_t \sim N(\mu, \Sigma)$$

$$\mu = f(X, \theta)$$
$$Y_t \sim N(\mu, \Sigma)$$

Moving Average Models: MA(q)

$$Y_t = \sum_{j=1}^{q} a_j \epsilon_{t-j} + \epsilon_t$$

- Lags on the errors instead of the Y's
- Equivalent to regression on the residuals
- Is related to the weighted moving average approach to smoothing
  - Coefficients fit rather than assumed
- MA(q) = ARIMA(0,0,q)

ARMA(p,q)

$$Y_t = \sum_{i=1}^{p} \rho_i Y_{t-i} + \epsilon_t + \sum_{j=1}^{q} a_j \epsilon_{t-j}$$

- Combines both Autoregressive and Moving Average components
- ARMA(p,q) = ARIMA(p,0,q)

Integrated Model: I(d)

$$\Delta^d Y_t = \epsilon_t$$

- Models the dth difference of Y rather than modeling Y
- Simplest case assumes dth difference is stationary (mean 0, constant variance)
- As mentioned before
  - Differences approximate derivatives
  - Biologically may expect these to follow some process model (e.g. Density dependence)
- I(d) = ARIMA(0,d,0)

ARIMA(p,d,q)
Autoregressive Integrated Moving Average

- General case for classic frequentist time series, work just as well in Bayesian context
- Extensible to dealing with autocor in data models
- Contains a number of important special cases
  - AR(p) = ARIMA(p,0,0)
  - MA(q) = ARIMA(0,0,q)
  - I(d) = ARIMA(0,d,0)
  - ARMA(p,q) = ARIMA(p,0,q)
  - Gaussian white noise = ARIMA(0,0,0)

How do you set p,d,q

- Exploratory analyses
  - Partial Autocorrelation function (pacf)
  - Differencing (diff)
  - Weighted moving average smoothing (filter)

- Model Selection
  - AIC, LRT, DIC, etc.
  - R function **arima(X,c(p,d,q))** returns AIC
  - R function **ar(X)** automatically finds the p with the lowest AIC