Introduction

Overview

1. Data
2. Concepts
3. Models, methods

Text Examples of Time Series

JJ earnings Choice between complicated polynomial with changing variance versus percentage change modeled as constant.

Transformations often simplify a model. The example shows that it can be very easy to confuse model specification error for dependence. The Durbin-Watson statistic is quite significant, but should we blame dependence?

Temperature Trend is the key here. Is the trend real, and is it short-term or long-term? What are the driving factors? Example expands the data used in the text.

Causation? Later example with glacial varve (sediments left when a glacier melts, page 63) as a proxy series; Adi Wyner has done related work.

Speech is highly *periodic*, with an echo in the periodogram. What does it mean to be periodic? Not just a sine wave. Perhaps changes state.

Long term dependence produces less precise (*i.e.*, higher standard error) estimates of statistics like the mean since

$$
\begin{aligned}
\mathrm{Var}(\overline{Y}) &= \tfrac{1}{n^2}(n\sigma^2 + \sum_{t \neq s} \mathrm{Cov}(Y_t, Y_s)) \\
&= \frac{\sigma^2}{n} + \frac{2}{n}\sum_{h=1}^{n-1}(1 - h/n)\,\mathrm{Cov}(Y_{t+h}, Y_t)
\end{aligned}
$$

The second line requires that the covariance only dependent on the separation rather than position, that is, stationarity. The covariances must decay to zero in order for $\overline{Y}$ to be a consistent estimator.

Finance such as the NYSE stock returns is more the domain of Steele and financial time series course; time series without signal versus time series with signal.

El nino and fish give example of trying to find a leading indicator, and the science suggests which is which.

The complicated cross-correlations have a possibly simple explanation. If $Y_t = \beta X_{t-\ell} + \epsilon_t$, then the cross-correlations are just a shifted version of the correlations in X_t itself (note the notation)

$$\mathrm{Cov}(Y_{t+h}, X_t) = \beta\,\mathrm{Cov}(X_{t+h-\ell}, X_t) = \beta\gamma_X(h-\ell)\ .$$

Hence, if X_t has a complex structure, then so do the cross-correlations, even though the lead-lag association is very simple.

Earthquakes and explosions illustarates *classification.* (See the pop-science overview in *Scientific American*, 2009.)

Random walk simulation shows the dangers of ignoring, incorrectly modeling dependence, as well as the difficulty of recognizing the presence of dependence.

Perspective, concepts

Variety of objectives Predict, classify, associate (as in regression). All are helped by forming a model, a way of describing the data as a mixture of pattern (which is predictable, regular) and noise.

$$\text{Model} = \text{Pattern} \quad + \quad \text{Noise}$$

Pattern in a time series captures/describes the association between values observed over time, concepts such as *autocorrelation.*

Time series is a stochastic process, a sequence of random variables defined on a common probability space.

We will stick to *discrete* index sets, not *continuous.* After all, data is always discrete.

Stationarity Does a time series define a lot or rows or a lot of columns? Sometimes its both (*e.g.*, longitudinal data), but we'll emphasize the analysis of a univariate series. Stationarity allows averaging.

Basic Models, Methods

Smoothing as a way to remove "noise" and emphasize the signal, improve the signal-to-noise ratio. Depends on what you mean by noise, as in the example of low-pass and high-pass filtering used in telephone lines.

Moving averages as a common way to remove the noise, taking local averages of the data to produce a new, "smoother" series. More descriptive, but important to appreciate how such methods influence the dependence as well.

Autoregression Recursive filtering (or insering feedback) is common; predict the future as a weighted average of recent values ($\mathbb{E}\, \epsilon_t = 0$)

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad \mathrm{Var}(\epsilon_t) = \sigma^2. \tag{1}$$

Backsubstitution shows the role of noise as a forcing function; the current value is just a weighted average of prior noise terms.

$$Y_t = \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + ... + \phi^{t-1}\epsilon_1 + \phi^t Y_0 \,, \quad t = 1, 2, \dots \tag{2}$$

Notice that:

1. *Markovian* if ϵ_t is independent of past, $f(Y_t \mid Y_{t-1}, Y_{t-2}, \dots, Y_0) = f(Y_t \mid Y_{t-1})$.
2. Asymptotically stationary with dependence on "boundary" conditions, here in the form of Y_0.
3. What's it mean as $t \to \infty$; need the notion of an infinite sequence of random variables? Most easily handled as a *Hilbert space*.
4. Random walk if $\phi = 1$.
5. Hints at the form of the Volterra expansion.

Sinusoids Decompose time series as a collection of orthogonal variables, as in

$$Y_t = \sum_j A_{j1} \cos 2\pi jt/n + A_{j2} \sin 2\pi jt/n, \quad j = 0, 1, \ldots, n/2. \tag{3}$$

Note that this is just an orthogonal transformation, from one coordinate system determined by time order to another determined by frequency, $A = MY$. Its not a model in our sense; there's no noise in this or a description of signal. *Wavelets* define another orthogonal transformation of the series.

Hidden state Models with a latent variable, a hidden state, such as those modeled by a Kalman filter (linear) or a hidden Markov model (HMM).