STAT 485/685
Basics

Richard Lockhart

Simon Fraser University

STAT 485/685 — Fall 2017

Purposes of These Notes

- Describe notation.
- Describe mean, autocovariance, autocorrelation
- Review Mathematical properties of these
- Do some examples
- Working on 2.1, 2.2, maybe 2.3 in the text.

Mathematical Framework

- Data is a sequence of numbers; usually $y_1, \ldots, y_T$
- T is number of time units.
- Values usually evenly spaced in time.
- Call it a *discrete time* time series; not *continuous time*.
- Plots are *scatter plots* of t (on x-axis) vs y_t.
- On plots try to use real values of time t – dates if that makes sense.
- Dots are connected by straight line segments.
- Join $(1, y_1)$ to $(2, y_2)$ then to $(3, y_3)$ and so on.

Probability Models

- Treat y_ts as observed values of sequence of random variables
- Notation $\ldots, Y_{-1}, Y_0, Y_1, \ldots$.
- Or $Y_1, \ldots, Y_T$ or $\{Y_t; t \in \mathcal{T}\}$.
- Looks like notation for random sample.
- But variables usually *not independent*.
- Key question: does collecting more data help?

Statistical Goals

- Predict future using past, current, data.
- Understand *dynamics*: what influences the next observation?
- Describe Ys using a *statistical model* or a *stochastic process model*.
- *Stochastic processes* are subject of STAT 380.
- In this course a few examples: linear time series models mostly.

This course

- Choose a time series model: *model identification*.
- Fit a time series model: *estimate* its *parameters*
- Diagnose model fit; check assumptions.
- *Forecast* future values; quantify likely forecast error.

Chapter 2: Fundamentals

- Complete stochastic model specifies joint distribution or density of $\{Y_t, t \in \mathcal{T}\}$.
- Most important summaries: means, standard deviations, correlations.
- Or means, variances, covariances.
- All you need for *Gaussian* (aka *normal*) data.
- SDs, correlations more interpretable.
- Variance and covariance formulas simpler.

Mean and autocovariance functions

- Mean denoted μ_t:
$$\mu_t = \mathrm{E}(Y_t).$$
- In general μ_t depends on t; called *mean function*.
- *Autocovariance function*: $\gamma_{t,s}$ defined by

$$\begin{aligned}
\gamma_{t,s} &= \mathrm{Cov}(Y_t, Y_s) \\
&= \mathrm{E}\left\{(Y_t - \mu_t)(Y_s - \mu_s)\right\} \\
&= \mathrm{E}\left\{Y_t Y_s - \mu_t Y_s - \mu_s Y_t + \mu_t \mu_s\right\} \\
&= \mathrm{E}\left\{Y_t Y_s\right\} - \mathrm{E}\left\{\mu_t Y_s\right\} - \mathrm{E}\left\{\mu_s Y_t\right\} + \mathrm{E}\left\{\mu_t \mu_s\right\} \\
&= \mathrm{E}\left\{Y_t Y_s\right\} - \mu_s \mu_t - \mu_s \mu_t + \mu_s \mu_t \\
&= \mathrm{E}\left\{Y_t Y_s\right\} - \mu_s \mu_t
\end{aligned}$$

Autocorrelation functions

- *Correlation* between Y_t and Y_s is unitless quantity.
- $\rho_{t,s}$ defined by

$$\begin{aligned}\rho_{t,s} &= \frac{\mathrm{Cov}(Y_t, Y_s)}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_s)}} \\ &= \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}\end{aligned}$$

- Remember

$$\mathrm{Var}(Y_t) = \mathrm{Cov}(Y_t, Y_t) = \mathrm{E}\left\{(Y_t - \mu_t)^2\right\} = \gamma_{t,t}.$$

Mathematical rules

- For (almost) *any* rvs U and V and constants a, b:

$$\mathrm{E}(aU + bV) = a\mathrm{E}(U) + b\mathrm{E}(V).$$

- Correlations are between -1 and 1:

$$-1 \le \rho_{t,s} \le 1 \text{ or } |\rho_{t,s}| \le 1 \text{ and for every } t \ \rho_{t,t} = 1$$

- Correlations and covariances are symmetric:

$$\gamma_{t,s} = \gamma_{s,t} \text{ and } \rho_{t,s} = \rho_{s,t}.$$

Mathematical rules 2

- Cov is *linear* in both its *arguments*:

$$\begin{aligned}\mathrm{Cov}(c_1X_1 + c_2X_2, d_1Y_1 + d_2Y_2) \\ &= c_1\mathrm{Cov}(X_1, d_1Y_1 + d_2Y_2) + c_2\mathrm{Cov}(X_2, d_1Y_1 + d_2Y_2) \\ &= c_1d_1\mathrm{Cov}(X_1, Y_1) + c_1d_2\mathrm{Cov}(X_1, Y_2) \\ &+ c_2d_1\mathrm{Cov}(X_2, Y_1) + c_2d_2\mathrm{Cov}(X_2, Y_2)\end{aligned}$$

Mathematical rules 3

- General formula we will use

$$\mathrm{Cov}\left(\sum_{i=1}^{s} c_i Y_i, \sum_{j=1}^{t} d_j Y_j\right) = \sum_{i=1}^{s}\sum_{j=1}^{t} c_i d_j \mathrm{Cov}(Y_i, Y_j)$$

- Applied to variances

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{T} c_i Y_i\right) &= \mathrm{Cov}\left(\sum_{i=1}^{T} c_i Y_i, \sum_{j=1}^{T} c_j Y_j\right) \\ &= \sum_{i=1}^{T}\sum_{j=1}^{T} c_i c_j \mathrm{Cov}(Y_i, Y_j) \\ &= \sum_{i=1}^{T} c_i^2 \mathrm{Var}(Y_i) + 2\sum_{i=2}^{T}\sum_{j=1}^{i-1} c_i c_j \mathrm{Cov}(Y_i, Y_j)\end{aligned}$$

Example Processes

- In following $\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots$ are iid, mean 0, variance σ^2.
- $Y_t = \epsilon_t$ defines *white noise* process.
- Mean of Y_t is $\mu_t = 0$.
- Autocovariance is

$$\gamma_{t,s} = \mathrm{Cov}(\epsilon_t, \epsilon_s) = \begin{cases} 0 & t \neq s \\ \sigma^2 & t = s \end{cases}$$

- Autocorrelation is

$$\rho_{t,s} = \mathrm{Corr}(\epsilon_t, \epsilon_s) = \begin{cases} 0 & t \neq s \\ 1 & t = s \end{cases}$$

Moving Averages

- $Y_t = \epsilon_t + \frac{1}{2}\epsilon_{t-1}$ defines *Moving Average* process.
- Mean of Y_t is $\mu_t = 0$.
- Autocovariance: 3 cases — $s = t$, $s = t + 1$ or $t = s + 1$, $|s - t| \ge 2$.
- First $s = t$:

$$\begin{aligned}
\gamma_{t,t} &= \mathrm{Cov}(\epsilon_t + \frac{1}{2}\epsilon_{t-1}, \epsilon_t + \frac{1}{2}\epsilon_{t-1}) \\
&= \mathrm{Cov}(\epsilon_t, \epsilon_t) + \frac{1}{2}\frac{1}{2}\mathrm{Cov}(\epsilon_{t-1}, \epsilon_{t-1}) \\
&= \frac{5}{4}\sigma^2.
\end{aligned}$$

Moving Averages Continued

- For $s = t + 1$:

$$\gamma_{t,t+1} = \mathrm{Cov}(\epsilon_t + \frac{1}{2}\epsilon_{t-1}, \epsilon_{t+1} + \frac{1}{2}\epsilon_t) = \frac{1}{2}\mathrm{Cov}(\epsilon_t, \epsilon_t) = \frac{1}{2}\sigma^2.$$

- For $s = t + 2$ (or bigger s)

$$\gamma_{t,t+2} = \mathrm{Cov}(\epsilon_t + \frac{1}{2}\epsilon_{t-1}, \epsilon_{t+2} + \frac{1}{2}\epsilon_{t+1}) = 0.$$

- So autocorrelation is

$$\rho_{t,s} = \begin{cases} 0 & |t = s| > 1 \\ \frac{2}{5} & |t - s| = 1 \\ 1 & t = s \end{cases}$$

Random Walk

- $Y_t = \epsilon_1 + \cdots + \epsilon_t$ defines *Random Walk*.
- Mean of Y_t is $\mu_t = 0$.
- Autocovariance: do case $s > t$:

$$\begin{aligned}\gamma_{t,s} &= \mathrm{Cov}(\epsilon_1 + \cdots + \epsilon_t, \epsilon_1 + \cdots + \epsilon_s) \\ &= \mathrm{Cov}(\epsilon_1, \epsilon_1) + \cdots + \mathrm{Cov}(\epsilon_t, \epsilon_t) = t\sigma^2.\end{aligned}$$

- So autocorrelation is

$$\rho_{t,s} = \frac{\min\{t,s\}\sigma^2}{\sqrt{t\sigma^2}\sqrt{s\sigma^2}} = \frac{\min t, s}{\sqrt{st}} = \min\left\{\sqrt{\frac{t}{s}}, \sqrt{\frac{s}{t}}\right\}.$$