Univariate time series

Adam Boult (www.bou.lt)

July 11, 2025

Contents

Preface

This is a live document, and is full of gaps, mistakes, typos etc.

Part I

Univariate stochastic processes

Chapter 1

Stochastic processes and their moments

1.1 Introduction to processes

1.1.1 Stochastic processes

In a stochastic process we have a mapping from a variable (time) to a random variable.

Discrete and continuous time

Time could be discrete, or continuous.

Temperature over time is a stochastic process, as is the number of cars sold each day.

Discrete and continous state space

The state space for temperature is continous, the number of people on the moon is discrete.

1.1.2 Stochastic evolution

We can describe processes by their evolution.

$p(x_t|x_{t-1}...)$

1.1.3 Gaussian processes

1.1.4 Moments of stochastic processes

1.1.5 Autocovariance and autocorrelation

Autocovariance

$AC(a,b) = cov(X_a, X_b)$

Autocorrelation

The autocorrelation between two time periods is their covariance, normlised by their variances

$AC(a,b) = \frac{E[(X_a - \mu_a)(X_b - \mu_b)]}{\sigma_a \sigma_b}$

This is also called serial correlation.

Chapter 2

White noise, and weak- and wide-sense stationarity

2.1 Stationarity

2.1.1 Weak- and wide-sense stationarity

Unconditional probabilities don't change over time.

So GDP would not be stationary, but random noise would. A random walk is not stationary, because the variance increases over time.

2.1.2 Weak-sense stationary

Mean and autocovarinance don't change over time.

2.1.3 Wide-sense stationary

All moments are the same.

2.1.4 Unit roots

2.2 Introduction

2.2.1 White noise

Variables at each time are indepdendent.

Chapter 3

Random walks

3.1 Random walks

3.1.1 Random walks

Chapter 4

Martingale processes

4.1 Introduction

4.1.1 Martingale property

For a process with the Martingale property, the expected value of all future variables is the current state.

This only restricts expectations.

$E(X_{n+1}|X_0, ..., X_n) = X_n$

Chapter 5

Markov processes

5.1 Introduction

5.1.1 Markov property

For a process with the Markov property, only the current state matters for all probability distributions.

$P(x_{t+n}|x_t) = P(x_{t+n}|x_t, x_{t-1}...)$

5.2 Markov chains

5.2.1 Finite state Markov chains

Transition matrices

This shows the probability for moving between discrete states.

We can show the probability of being in a state by multiplying the vector state by the transition matrix.

Mv

Time-homogenous Markov chains

For time-homogenous Markov chains the transition matrix is independent of time.

For these we can calculate the probability of being in any given state in the future:

$M^n v$

This becomes independent of v as we tend to infinity. The initial starting state does not matter for long term probabilities.

How to find steady state probability?

$Mv = v$

The eigenvectors! With associated eigenvector 1. There is only one eigenvector. We can find it by iteratively multiplying any vector by M.

5.2.2 Infinite state Markov chains

Markov model description We can represent the transition matrix as a series of rules to reduce the number of dimensions $P(x_t|y_{t-1}) = f(x, y)$

can represent states as number, rather than atomic. could be continuous, or even real.

in more complex, can use vectors.

5.3 Hidden Markov Models

5.3.1 Introduction

As well as the Markov process X, we have another process Y which depends on X.

5.4 Dynamic Bayesian networks

5.4.1 Introduction

Chapter 6

Survival functions

6.1 Introduction

6.1.1 Survival functions

Part II

Univariate Continuous-time stochastic processes

Chapter 7

Wiener processes and Brownian motion

7.1 Wiener proceses

7.1.1 Independent increments

The changes in any non-overlapping time increments are independent.

Formally:

$t_0 < t_1 < t_2 < ... < t_m$

With X_t

$X_{t_1} - X_{t_0}$ is indepentent from $X_{t_2} - Xt_1$ etc.

7.1.2 Wiener processes

A Wiener process is a process W_t with independent increments, which: + Is continuous + Has normally distributed increments.

Can be constructed as limit of random walk. Can also be constructed as integral of Gaussian noise?

7.2 Brownian motion

7.2.1 Brownian motion

brownian motion in stats. given we start at a, what is chance be end up at b? normal. do 1d then multi d

Chapter 8

Stochastic differential equations

Part III

Univariate discrete-time stochastic processes

Chapter 9

Orders of integration

9.1 Introduction

9.1.1 Orders of integration

How many diffs do you need to do to get a stationary process?

If something is first order integrated it is $I(1)$.

9.1.2 Trend stationary

If we can remove the trend as a function, eg linear or non-linear growth, and the rest is stationary, then the process is trend stationary

9.1.3 Seasonal and non-seasonal trends

We can model the process as:

$y_t = \mu_t + f(t) + \epsilon_t$

9.1.4 Cyclical fluctuations

We can have shocks having effects over time.

This is separate to trends.

Chapter 10

Auto-Regressive processes, Moving-Average processes and Wold's theorem

10.1 Autoregressive model

10.1.1 Autoregressive models (AR)

AR(1)

Our basic model was:

$x_t = \alpha + \epsilon_t$

We add an autoregressive component by adding a lagged observation.

$x_t = \alpha + \beta x_{t-1} + \epsilon_t$

AR(p)

AR(p) has p previous dependent variables.

$x_t = \alpha + \sum_{i=1}^{p} \beta_i x_{t-i}$

Propagation of shocks

A shock bumps up the output variable, which bumps up output variables forever, at a decreasing rate.

10.1.2 Testing for stationarity with Dickey-Fuller (DF) and Augmented Dicky-Fuller (ADF)

Stationarity

Unit roots

Integration order

Dickey-Fuller

The Dickey-Fuller test tests if there is a unit root.

The AR(1) model is:

$y_t = \alpha + \beta y_{t-1} + \epsilon_t$

We can rewrite this as:

$\Delta y_t = \alpha + (\beta - 1) y_{t-1} + \epsilon_t$

We test if $\beta - 1) = 0$.

If the coefficient on the last term is 1 we have a random walk, and the process is non-stationary.

If the last term is < 1 then we have a stationary process.

Variation: Removing the drift

If our model has no intercept it is:

$y_t = \beta y_{t-1} + \epsilon_t$

$\Delta y_t = (\beta - 1) y_{t-1} + \epsilon_t$

Variation: Adding a deterministic trend

If our model has a time trend it is:

$y_t = \alpha\beta y_{t-1} + \gamma t + \epsilon_t$

$\Delta y_t = \alpha + (\beta - 1) y_{t-1} + \gamma t + \epsilon_t$

Augmented Dickey-Fuller

We include more lagged variables.

$y_t = \alpha + \beta t + \sum_i^p \theta_i y_{t-i} + \epsilon_t$

If no unit root, can do normal OLS?

10.1.3 Autoregressive Conditional Heteroskedasticity (ARCH)

Variance of the AR(1) model

The standard AR(1) model is:

$y_t = \alpha + \beta y_{t-1} + \epsilon_t$

The variance is:

$Var(y_t) = Var(\alpha + \beta y_{t-1} + \epsilon_t)$

$Var(y_t)(1 - \beta^2) = Var(\epsilon_t)$

Assuming the errors are IID we have:

$Var(y_t)) = \frac{\sigma^2}{1-\beta^2}$

This is independent of historic observations, which may not be desirable.

Conditional variance

Consider the alternative formulation:

$y_t = \epsilon_t f(y_{t-1})$

This allows for conditional heteroskedasticity.

10.2 Moving average models

10.2.1 Moving Average models (MA)

We add previous error terms as input variables

MA(q) has q previous error terms in the model

Unlike AR models, the effects of any shocks wear off after q terms.

This is harder to fit the OLS, the error terms themselves are not observed.

10.3 Autoregressive Moving Average models

10.3.1 Autoregressive Moving Average models (ARMA)

We include both AR and MA

Estimted using Box-Jenkins

10.3.2 Autoregressive Integrated Moving Average models (ARIMA)

Uses differences to remove non statiority

Also estiamted with box-jenkins

10.3.3 Seasonal ARIMA

10.4 Wold's theorem

10.4.1 Introduction

Part IV

Sampling

Chapter 11

Markov chain Monte Carlo sampling

11.1 Markov Chain Monte Carlo (MCMC) methods

11.2 Metropolis-Hastings algorithm

11.2.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm creates a set of samples x such that the distribution of the samples approaches the goal distribution.

Initialisation

The algorithm takes an arbitrary starting sample x_0. It then must decide which sample to consider next.

Generation

It does this using a Markov chain. That is, there is a map $g(x_j, x_i)$.

This distribution is generally a normal distribution around x_i, making the process a random walk.

Acceptance

Now we have a considered sample, we can either accept or reject it. It is this step that makes the end distribution approximage the function.

We accept if $\frac{f(x_j)}{f(x_i)} > u$, where u is a random variable between 0 and 1, generated each time.

We can calculate this because we know this function.

Properties

11.3 Gibb's sampling

11.3.1 Gibb's sampling

Introduction

As with Metropolis-Hastings, we want to generate samples for $P(X)$ and use this to approximate its form.

We do this by using the conditional distribution. If X is a vector then we also have:

$P(x_j|x_0, ..., x_{j-1}, x_{j+1}, ..., x_n)$

We use our knowledge of this distribution.

Start with vector x_0.

This has components $x_{0,j}$

To form the next vector x_1 we loop through each component.

$P(x_{1,0}|x_{0,0}, x_{0,1}, ..., x_{0,n})$

We use this to form $x_{1,0}$

However after th the first component we update this so it uses the updated variables.

$P(x_{1,k}|x_{1,0}, ..., x_{1,k-1}, x_{0,k}, ..., x_{0,n}$

This means we only need to know the conditional distributions.

Chapter 12

Sampling from processes

12.1 Introduction

Chapter 13

Forecasting stochastic processes

13.1 Forecasting

13.1.1 Introduction to forecasting

We observe a series of observations:

$x_1, x_2, ..., x_t)$

What can we say about x_{t+1}?

If the data was drawn iid then the past data then we would just want to identify moments.

However if the data is not iid, for example because it is increasing in time, then this is not the best way.

Regression formation

We can model

$x_t = \alpha + \epsilon_t$

13.1.2 Monte carlo simulations

13.1.3 N-step ahead

13.1.4 Consensus forecasting

Part V

Signal processing

Chapter 14

Quantisation and sample rates

14.1 Introduction

14.1.1 Quantisation

14.1.2 Sample rate

Chapter 15

Discrete Fourier Transform

15.1 Introduction

15.1.1 Discrete Fourier Transform

Chapter 16

Down sampling

16.1 Introduction

16.1.1 Down sampling

Chapter 17

Fast Fourier Transform

17.1 Introduction

17.1.1 Fast Fourier Transform

Chapter 18

Noisy networks

18.1 Introduction

18.1.1 Noisy networks

Part VI

Advanced inference (time series univariate)

Chapter 19

Imputing missing data for time series

19.1 Time series

19.1.1 ARIMA interpolation

19.1.2 Last Observation Carried Forward (LOCF)

19.1.3 Next Observation Carried Backward (NOCB)

19.1.4 Other

Multi period averages for imputation on time series.

Part VII

Estimating time series models

Chapter 20

Estimating Markov chains

20.1 Estimating Markov chains

20.1.1 Estimating the Markov chain stochastic matrix

Introduction

Given a sequence: $x_1, ...x_n$.

The likelihood is:

$L = \prod_{i=2}^n p_{x_{i-1},x_i}$

If there are k states we can rewrite this as:

$L = prod_{i=1}^k \prod_{j=1}^k n_{ij} p_{ij}$

Where p_{ij} is the chance of moving from state i to state j, and n_{ij} is the number of transtions between i and j.

The log likelihood is:

$\ln L = \sum_{i=1}^k \sum_{j=1}^k n_{ij} \ln p_{ij}$

Constrained optimisation

Not all parameters are free. All probabilities must sum to 1.

$\ln L = \sum_{i=1}^k \sum_{j=1}^k n_{ij} \ln p_{ij} - sum_{i=1} \lambda_i (\sum_{j=1} p_{ij} - 1)$

This gives us:

$\hat{p}_{ij} = \frac{n_{ij}}{\sum_k n_{ik}}$

20.1.2 Estimating infinite state Markov chains

We can represent the transition matrix as a series of rules to reduce the number of dimensions

$P(x_t|y_{t-1}) = f(x, y)$

can represent states as number, rather than atomic. could be continuous, or even real.

in more complex, can use vectors.

20.2 Ergodic processes

20.2.1 Ergodic processes

Sample moments must converge to generating momements. Not guaranteed.

Eg process with path dependence. 50

Generating average is £50, but sample will only convergen to £100 or £0

Chapter 21

Estimating Hidden Markov Models (HMMs)

21.1 Estimating Hidden Markov Models (HMMs)

21.1.1 Recap of Hidden Markov Models (HMMs)

We don't see state

Each state produces a visible output. this output is drawn from a distribution for each state.

We observe a sequence of outputs, not states.

21.1.2 Estimating HMMs with the Viterbi algorithm

Assume we know transition matrix. and starting probls

Given we observe sequence of outputs, what were most likely actual paths?

Virbiti returns this

21.1.3 Estimating HMMs with the forward algorithm

Given we have observed outputs, what is the chance of being in a certain state at a certain time?

21.1.4 Estimating HMMs with the forward-backward algorithm

We calculate state x at time t given all obs.

21.1.5 Baum-Welch algorithm

21.1.6 Kalman filters

Chapter 22

Univariate forecasting

22.1 Introduction

22.1.1 Seasonal and non-seasonal trends

We can model the process as:

$y_t = \mu_t + f(t) + \epsilon_t$

22.1.2 Identifying the order of integration using Augmented Dickey-Fuller

The Dickey-Fuller test with deterministic time trend was:

$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \epsilon_t$

The Augmented Dickey-Fuller model adds lags for the differences.

$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_i^p \delta_i \Delta y_{t-i} + \epsilon_t$

22.1.3 Cyclical fluctuations

We can have shocks having effects over time.

This is separate to trends.

22.1.4 Identifying serial correlation using the Durbin-Watson statistic

22.1.5 Introduction to forecasting

We observe a series of observations:

$x_1, x_2, ..., x_t)$

What can we say about x_{t+1}?

If the data was drawn iid then the past data then we would just want to identify moments.

However if the data is not iid, for example because it is increasing in time, then this is not the best way.

Regression formation

We can model

$x_t = \alpha + \epsilon_t$

22.2 Autoregressive model

22.2.1 Autoregressive models (AR)

AR(1)

Our basic model was:

$x_t = \alpha + \epsilon_t$

We add an autoregressive component by adding a lagged observation.

$x_t = \alpha + \beta x_{t-1} + \epsilon_t$

AR(p)

AR(p) has p previous dependent variables.

$x_t = \alpha + \sum_{i=1}^p \beta_i x_{t-i}$

Propagation of shocks

A shock bumps up the output variable, which bumps up output variables forever, at a decreasing rate.

22.2.2 Testing for stationarity with Dickey-Fuller (DF) and Augmented Dicky-Fuller (ADF)

Stationarity

Unit roots

Integration order

Dickey-Fuller

The Dickey-Fuller test tests if there is a unit root.

The AR(1) model is:

$y_t = \alpha + \beta y_{t-1} + \epsilon_t$

We can rewrite this as:

$\Delta y_t = \alpha + (\beta - 1)y_{t-1} + \epsilon_t$

We test if $\beta - 1) = 0$.

If the coefficient on the last term is 1 we have a random walk, and the process is non-stationary.

If the last term is < 1 then we have a stationary process.

Variation: Removing the drift

If our model has no intercept it is:

$y_t = \beta y_{t-1} + \epsilon_t$

$\Delta y_t = (\beta - 1)y_{t-1} + \epsilon_t$

Variation: Adding a deterministic trend

If our model has a time trend it is:

$y_t = \alpha\beta y_{t-1} + \gamma t + \epsilon_t$

$\Delta y_t = \alpha + (\beta - 1)y_{t-1} + \gamma t + \epsilon_t$

Augmented Dickey-Fuller

We include more lagged variables.

$y_t = \alpha + \beta t + \sum_i^p \theta_i y_{t-i} + \epsilon_t$

If no unit root, can do normal OLS?

22.2.3 Autoregressive Conditional Heteroskedasticity (ARCH)

Variance of the AR(1) model

The standard AR(1) model is:

$y_t = \alpha + \beta y_{t-1} + \epsilon_t$

The variance is:

$Var(y_t) = Var(\alpha + \beta y_{t-1} + \epsilon_t)$

$Var(y_t)(1 - \beta^2) = Var(\epsilon_t)$

Assuming the errors are IID we have:

$Var(y_t)) = \dfrac{\sigma^2}{1 - \beta^2}$

This is independent of historic observations, which may not be desirable.

Conditional variance

Consider the alternative formulation:

$y_t = \epsilon_t f(y_{t-1})$

This allows for conditional heteroskedasticity.

22.3 Moving average models

22.3.1 Moving Average models (MA)

We add previous error terms as input variables

MA(q) has q previous error terms in the model

Unlike AR models, the effects of any shocks wear off after q terms.

This is harder to fit the OLS, the error terms themselves are not observed.

22.4 Autoregressive Moving Average models

22.4.1 Autoregressive Moving Average models (ARMA)

We include both AR and MA

Estimted using Box-Jenkins

22.4.2 Autoregressive Integrated Moving Average models (ARIMA)

Uses differences to remove non statiority

Also estiamted with box-jenkins

22.4.3 Seasonal ARIMA

22.5 Forecasting

22.5.1 Monte carlo simulations

22.5.2 N-step ahead

22.5.3 Consensus forecasting

22.6 Other

22.6.1 Identifying the order of integration using Augmented Dickey-Fuller

The Dickey-Fuller test with deterministic time trend was:

$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \epsilon_t$

The Augmented Dickey-Fuller model adds lags for the differences.

$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_i^p \delta_i \Delta y_{t-i} + \epsilon_t$

22.6.2 Identifying serial correlation using the Durbin-Watson statistic

Chapter 23

Inference with time series

23.1 OLS on time series data

23.1.1 Bias of static models and spurious correlations

Static models

Static models are of the form:

$y_t = \alpha + \beta x_t + \epsilon_t$

These have no lagged variables or difference operators.

Bias of static models

23.1.2 Heteroskedasticity and Autocorrelation (HAC) adjusted standard errors

23.2 Time series

23.2.1 Taking differences

What we use should depend on I(1), I(0) etc from ADF

if we're missing time invariant data, we can do first differences and this isn't a problem if we do diff in diff this removes trends?

page on first difference estiamtion? OLS on first differences. No other lags page on first difference ESTIMATOR

23.2.2 Discontinuity

Create a dummy for before/after a date.

23.3 Panel data

23.3.1 Difference-in-difference

Consider the grouped linear model:

$y_{ij} = \mu + \tau_i + X_j\theta + \epsilon_{ij}$

By taking differences with another observation in the same group we remove the average terms.

$y_{ij} - y_{ik} = (\mu + \tau_i + X_j\theta + \epsilon_{ij}) - (\mu + \tau_i + X_k\theta + \epsilon_{ik})$

$y_{ij} - y_{ik} = (X_j\theta - X_k) + (\epsilon_{ij} - \epsilon_{ik})$

diff in diff: control group and treated group. page on leakiness? are control affected too? Assumption: in absense of treatment, price would have evolved like control

23.3.2 Controlled experiments

23.3.3 Natural experiments

23.3.4 Structural breaks

Testing for structural breaks with the Chow test.

23.3.5 Dynamic or lagged independent variables

Static panel data: No lags of independent variables. Dynamic panel data: Lags of independent variables.

OLS is consistent for static panel data, not for dynamic This results in Nickell's bias for dynamic panel data

Dynamic panel data: y_{t-1} is a regressor Panel data estimation: LSDV. Least squares dummy variable estimator arnello bond

Chapter 24

Survival analysis

24.1 Introduction

24.1.1 Cox-hazard