**Week 1 16 July 2018**

Meeting with Thomas✓

* Thomas busy, 10 Aug 10:30am for next meeting

Scholar articles

- Wickramasuriya et al 2017, Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization ✓

- Ghosh & Steorts 2013, Two-stage Benchmarking as Applied to Small Area Estimation ✓

**Week 2 23 July 2018**

Stats 726 Time series **✓**

STATS 380 Statistical Computing ✓

Latex thesis draft

* Thesis structure **✓**
* CV **✓**

**Week 3 30 July 2018**

Stats 331 Introduction to Bayesian

* Bit too hard, may need to revisit textbook

Analysing multivariate data – Norman Cliff Cha 1 Ch2 ✓

robjhyndman.com

* Forecasting hierarchical time series <https://robjhyndman.com/talks/Hierarchical.pdf>✓
* Time series graphics https://robjhyndman.com/medascin/1-time-series-graphics.pdf✓
* Forecasting: Principles and Practice ch10 <https://otexts.org/fpp2/hierarchical.html> ✓

Other Materials

* Introduction to Bayesian Data Analysis with R. <http://www.sumsar.net/files/academia/user_2015_tutorial_bayesian_data_analysis_short_version.pdf> ✓(make notes later)

Latex

* How to draw up a hierarchical tree diagram for taxonomic classification <https://tex.stackexchange.com/questions/19029/how-to-draw-up-a-hierarchical-tree-diagram-for-taxonomic-classification> ✓
* Tools for tree diagram <https://tex.stackexchange.com/questions/50418/tools-for-tree-diagram> ✓
* Other tools? What tool for large tables and tree diagram with circles

Calendar 50 week plan ✓

Scholar articles

* Talagara et al 2018, Meta-learning how to forecast time series ✓
* Wickramasuriya et al 2015, Forecasting hierarchical and grouped time series through trace minimization ✓
* Shang & Hyndman 2016, Grouped functional time series forecasting, an application to age-specific mortality rates ✓ good

ICD 10

* U.S and NZ had different classifications, change of standard over the years etc
* Confirm with Thomas is it worth to look into ICD 10 and ICD 10 cm

Week 4 6 August 2018

Thesis Draft✓

Stats 331 Introduction to Bayesian Lecture 1- 17 ✓

Analysing multivariate data – Norman Cliff Cha 3 -5 ✓

Handbook of Discrete-Valued Time Series Davis et al section 1.1-1.5 ✓

robjhyndman.com

- R package for hierarchical time series <https://cran.r-project.org/web/packages/hts/hts.pdf>

- Time series all <https://robjhyndman.com/teaching/>

Start ppt, long version, able to describe all my work in 2 hours of talking.

Start a poster ?

Scholar articles

* Crystal Foy NZ Household Food Purchases in 2012 – 2016 ✓
* Mugdha Manda Creating synthetic datasets using multiple imputation and stochastically assigned synthetic individuals ✓

Meeting with Thomas

* Clarify what Result is expected from me, ie structure of thesis,
* Clarify what software to use , data, simulation, graphs figures etc
* Tell him problem about notations, seeing new concept and had to always go back and search for it
* Was reading and learning various related concepts
* Probably need to focus on matrix algebra for now
* You tube: Brandon 331, 3blue1brown
* Tom Eilliot Auckland Bus
* Graphuz draw tree plots
* Buds JAGS Stan(Faster?)
* **April** is the deadline to start Thesis
* Bus data, MMIC, NMD nz, Census, genetic data

Week 5 13th August 2018

Draft Research Proposal ✓ Ask Thomas when he want my proposal, end of semester?

Possibility of going to a conference? Just to see what it is like.

Stats 331, Online Lectures on Youtube, start doing notes

Latex thesis drafts ✓

Latex Tutorials

* <https://www.sharelatex.com/blog/2013/08/02/thesis-series-pt1.html>✓
* <http://www.rpi.edu/dept/arc/docs/latex-thesis/rpithesis-doc.pdf>✓

Baysian Analysis for financial time series Andreas Berg PhD ✓

Curriculum Development Project For 773 Clinical Trials Msc ✓

Bayesian Hierarchical Grouping: perceptual grouping as mixture estimation✓ Good

Hts package <http://pkg.earo.me/hts/>

Stats 782 Assignments

<https://www.clinicalstudies.+co.nz/participant-info/current-clinical-trials/>

* Field experience with medical trials? See how a trial is done?

Coastline\_paradox

* <https://en.wikipedia.org/wiki/Coastline_paradox>

Meeting with Thomas

* Clarify when the proposal is due and what is expected
* Clarify definition of hierarchical and grouped data
* Conference @palmerston North, End of November
* Clinical Meeting Februaryish, Thomas happy to take me
* 1 page proposal by end of break

E-mail: vfroyen@sunyopt.edu

Week 6 20 Aug

Stats 331, Online Lectures on Youtube ✓

Stats 210 probability theory ✓

Bayesian Hierarchical Grouping

Tue to Sat need to take dad to hospital

Week 7 27 Aug

Stats 210 probability theory ✓

Not going so well, oh well, family matter before academic progress.

Future aint looking bright for you mate

Week 8 3 Sep

Depression, need to get back on track Jung

Anomalies, YouTube ✓ see anomaly.docx for notes

Week 9 10 Sep

What are anomalies? What kind of anomaly are we interested in.

* Definition for epidemiologic outbreak?
  + Person (characteristics, ethnicity, agegroup?)?
  + Place (Cluster /abnormal concentration of cases in a location)
  + Time (Seasonal trend, cycle etc )
  + Clinical features (abnormal amount of certain symptoms)
  + What are clinical significant? (Maybe reach critical proportion, inequality etc)

What are the significant units of interest?

Count? Rate? Proportion?

Poisson distribution Y0 ~ poisson (Lambda, Ei) Y0 = Y00+Y01 E0 = E00+E01

Use Influenza as example <https://www.icd10data.com/ICD10CM/Codes/J00-J99/J09-J18>

First 2 level of ICD 10 have low clinical associative significance, so can ignore.?

Shall we look into different main categories rather than all?

Simple tree

<https://tex.stackexchange.com/questions/5447/how-can-i-draw-simple-trees-in-latex>

Delayed meeting with Thomas, not a good habit to start.

Need to speed up progress.

Twitters anomaly detection <https://anomaly.io/anomaly-detection-twitter-r/> good

**Week 10 17 Sep**

Need to speed up progress, week 10 already

Anomaly journals, see anomaly.docx

Start a journal note, where ideas can be organised and referenced later, just like what you did for EDUC 351 essay, you got a A+, remember that?

Meeting with Thomas

* Ask him to improve my proposal.
  + Cannot find suitable journal, need abstract for online search engine
  + Add word Bayesian to my title?
  + What am I trying to achieve?
* Clarify on anomalies
  + What are domain of application, and detection requirement?
  + Who started to focus on hierarchical structure, and anomaly detection? there seem to be stuff all over the place and no clear major author
  + Looking at false positive, false negative rate on simulation?
  + And expand on thing such as anomaly type, ie
    - global and local,
    - number of brunch
    - size of category
    - Compare different disease category?
  + And then add anomaly to existing data and see if we can detect it?
* Bayesian
  + not suitable for small or clustered data, Poisson not good
  + What to do with strong seasonal trends and overall trends? Anomaly detection have hard time detecting/distinguishing these
  + ED data not a simple model, what if we over model?
  + Need a text book or other resource on Bayesian model
* Needs help on Structure of thesis, to get a clear direction
* Concern on when to start to apply for data, when I have a relatively good simulated results?
* Social media as learning tool and build academic relation? how long/week should be spent
* Haven’t found specific articles, Is this a new topic or someone had done something similar? People seem to be more focused on timeliness
* How thesis is graded
* Machine learning (data science vs statistics?)
* Possibility of publishing on a journal
* Any good text book on Bayesian regression, Poisson model

Notes:

* Focus on getting the estimations first
* Bench marking Constraint/estimation proper hierarchical modelling
* Motivatd by Don Barry 2004
* More a estimation problem, not a testing proplem
* Poisson distribution with varying mean (extension)
* Consistent set of estimate rather than ……in hierarchy. Automatically?
* Anomaly objective, easy to detect, understanding
* Social media: blogs, twitter github.
* Infectious disease may appear in cluster, problem for poisson
* <http://notstatschat.tumblr.com/>
* <https://www.statschat.org.nz/>
* <https://github.com/tslumley/notstatschat/blob/master/content/about.md>
* <https://robjhyndman.com/>

**Week 11 24 Sep**

Thesis Writing

* Grammatical errors of Chinese writers <http://blog.sciencenet.cn/blog-1232242-904021.html>
* Search 写作 in PhD Circles miniblog

Ggplot2

* Data visualisation <http://r4ds.had.co.nz/data-visualisation.html>
* Example <http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html>
* Example <https://r4stats.com/examples/graphics-ggplot2/>

Anomaly detection

* Anomaly Detection with the Poisson Distribution <https://anomaly.io/anomaly-detection-poisson-distribution/>

Simulation

* Problem with random time selection, first and last day only half counts
* How to improve effeciency
* Improve dummy variable, and plots with forloop? Write a function?
* More than a million?
* Try to get estimates next
* How to run simulation, examples?
* <https://www.youtube.com/watch?v=hgv5RNgFqDQ> ?

BUGS, JAGS, STAN

* See BUGS JAGS STAN Notes.docx

Poisson distribution

* Poisson Notes.docx

Sentiment analysis Good stuff

* <https://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/>
* <https://notstatschat.rbind.io/2018/09/27/how-to-write-a-racist-ai-in-r-without-really-trying/>
* https://nlp.stanford.edu/projects/glove/

Meeting with Thomas

Simulation

* is one simulation enough or we need 1000 repeats? 1 million point good enough?
* How to make coding efficient, avoid for loop? How to do vectorised value assignment?
* Problem with random time selection, first and last day only half counts
* How to add demographic, seasonal trend etc with random simulation process?
* Quickly explain or show me how do people usually run simulation?
* Should get estimates by next week
* Start estimates
* Write models

**Week 12 1 Oct**

Organising Notes and readings

* Such a hideous task

STATS 782 Assignment 3

LaTeX

* Appendix 1 glossary codes not working?
* Appendix 2 softwares

Simulations

* See simulationV2.R

BUGS, JAGS, STAN

* Installed and able to run BUGS STAN

Model

* Unsure what it is like
* See draft paper

Note: [Stat Seminar] : Version control: An introduction to git and GitHub using RStudio 10Oct

**Meeting with Thomas**

* How to run r codes in a box on a latex file? Eg Barnette 2017, or <http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/>
* How to model, need some help
* Hierarchical Poisson model
* Terminology of trees
* Report on simulation
* How to show which software best suits hierarchical models??? How to compare performance and efficiency?
* What packages had Thomas used?
* Todos: try barry and hierarchical model, write on latex, run on jags

**Week 13 8 Oct**

|  |  |
| --- | --- |
| **Days** | **Plans / tasks of the day** |
| Sat | Catch up on 782 Lectures |
| Sun | Catch up on 782 Lectures |
| Mon | Organise file and Notes  Take notes on Bayesian models  Organise latex structure |
| Tue | Notes on anomaly  Simulation bugs, ask Thomas use 2,2 model |
| Wed | Add ggplot2 for sim.data  782 Assignment 4 |
| Thu | JAGS and STAN codes for models  SM Barry Bayes chapter  Write out bayesian models |
| Fri | Meet Thomas  782 Assignment 4 |

Meeting with Thomas

* How would we do a hierarchical model with frequentist approach

  + ddd
* What are some of the advantages for using a Bayesian approach for hierarchical models?
  + Flexibility for complex models, more realistic
  + Use of posterior, simulation friendly?
* Poisson models
  + Able to run example JAGS codes but not sure what model to use, not sure if my data structure will fit
  + What does tau eta pi tend to represent??
  + is it okay to add a comma before t?
  + Question on Barry models, clarify
  + Is there a way to write a DAG to describe our model?
  + Show Thomas what I have now, what other equations I should add?
* Simulations
  + Debug colsum issue use trunk(), use ifelse
  + Where to do 1000 repeats? Raw data, get 1000 diff prior?
* Anomaly?
  + What kind of anomaly will I be interested in? anomaly in counts in %?? , anomaly in change in rates, anomaly in covariates??? Just look at positive increase in counts, it is what cause conjestion, negative anomaly, anomaly In rates can be looked into later
* Going to GP next week, may get a surgery in worst case scenario, will inform Thomas if I cannot make the next meeting

**Week 14 15 Oct**

|  |  |
| --- | --- |
| **Days** | **Plans / tasks of the day** |
| Sat | - Take bus to Tok town  - 782 S4 |
| Sun | - gardening |
| Mon | - See GP :3 |
| Tue | - Take bus to Auckland  - 782 Assignment 4 |
| Wed | - 782 Assignment 4  - Clean up diary, and notes  - Hierarchical Models - Statistical Methods Filippi 2016 <http://www.stats.ox.ac.uk/~filippi/Teaching/msc_hierarchicalmodels_2016.html>  - Forecasting hierarchical time series <https://robjhyndman.com/talks/Hierarchical.pdf>  - Introduction to Bayesian Data Analysis with R (Baath 2015)  <http://www.sumsar.net/files/academia/user_2015_tutorial_bayesian_data_analysis_short_version.pdf> |
| Thu | - Latex  - Simulation |
| Fri | - Meet Thomas??? Not much will be done this week  Rbind.io stuff  <https://github.com/rbind>  <https://alison.rbind.io/post/up-and-running-with-blogdown/>  <https://github.com/rbind/yihui>  <https://github.com/rbind/robjhyndman.com> |

Week 15 22 Oct

|  |  |
| --- | --- |
| **Days** | **Plans / tasks of the day** |
| Sat | World of Tank Blitz Event |
| Sun | World of Tank Blitz Event  Gardening |
| Mon | Stats 782 Lecture 1-6 |
| Tue | Stats 782 Lecture 7-12 |
| Wed | Stats 782 Lecture 13-18  Stats 782 Lab 1, 2  Data Analysis Using Regression and multilevel/hierarchical Models ch12  Need to get the jargons right, Multilevel, Grouped, Nested, hierarchical |
| Thu | Stats 782 Lecture 19 – 24  Craving for KFC, don’t know why  Accommodation stuff  Summer School stuff |
| Fri | Ask Thomas on 27 Nov Conference stuff  Seminar "Techniques for detecting anomalies in purchasing card transactions”  Date: Friday, 26/10/2018 Time: 14:00-15:00 Room: 303-G14  Create GitHub account, experimenting on rbind.io, start repository for Notes,  Blog with blogdown  <https://bookdown.org/yihui/blogdown/structure-of-the-book.html> |

Week 16 27 Oct

|  |  |
| --- | --- |
| **Days** | **Plans / tasks of the day** |
| Sat |  |
| Sun |  |
| Mon | 782 Revision |
| Tue | 782 Revision |
| Wed | 782 Revision |
| Thu | 782 Revision |
| Fri | 782 Revision |

Week 17 3 Nov

|  |  |
| --- | --- |
| **Days** | **Plans / tasks of the day** |
| Sat |  |
| Sun |  |
| Mon | 782 Revision |
| Tue | 782 Revision |
| Wed | 782 Revision |
| Thu | 782 Exam |
| Fri | Recover |

Week 18 10 Nov

|  |  |
| --- | --- |
| **Days** | **Plans / tasks of the day** |
| Sat | Recover |
| Sun | Tokoroa |
| Mon | Tokoroa |
| Tue | Tokoroa |
| Wed | Back to Auckland |
| Thu | PG talk  Github |
| Fri | PG talk |

**Question for Thomas**

- Is it okay to upload thesis material on GitHub? I want to try out version control and need a secure data storage, and you can see my work when you needed to. <https://github.com/jungxue>

- What is the dress code for the NZSA conference? Is black jean + navy blue shirt + grey blazer okay and not overdressed?

- Currently focused on **SM Berry (Clinical trial),** **Andrew Gelman (Hierarchical Bayesian)**  and **Rob J Hyndman (Hierarchical time series)** ? I think they are the big guys in the field of hierarchical models.

- write 3 sections on their models, evaluate, and write a section on my own?

- Set up my model are so it is best suited for applications in ED arrival studies? I.e set up assumptions that reflect and suited for what we usually see in ED arrival data.

- Pooled vs. unpooled methods

- Hierarchical model is somewhere in between?

- Pooling under fit and non-pooling (separate models) over fit

- pooled = variance are equal (Var < 2)

- Study Plan:

|  |  |
| --- | --- |
| **November** | Produce estimates,  finish Xue.rbind.io, <https://github.com/jungxue/Xue.rbind.io>  Clean up and save 6 ½ year of my university files   * Back up 1 * Back up 2 * Laptop   Study link ✓  Short term plan for my thesis see to do list in readme.md : <https://github.com/jungxue/research_masters_Jung/blob/master/README.md>  Finish Notes on PG talk and NZSA seminars ✓ |
| **December** | wisdom tooth extraction, a painful Xmas holiday✓  work on latex, need to finish 1/3 (15000 word) by end of December  more reading, focus on Gelman, Hydnman, Berry ✓  Polish my LinkedIn, check out jobs |
| **January** | Play around with Simulation, start to email Thomas for questions  - See how size of groups, size of anomaly, distance of anomaly and level  that is being modelled, affect anomaly detection, group vs point anomaly  - Use **Cross validation** to demonstrate strength and weakness of  hierarchical model, and evaluate anomaly detections  work on latex  GitHub  Read stats 732, prepare for next semester |

12 Feb

|  |  |
| --- | --- |
| **Days** | **Plans / tasks of the day** |
| Mon |  |
| Tue |  |
| Wed | **Surgery 3:50** |
| Thu |  |
| Fri | **Back to Uni Next Meeting Mon 9:30**  Allenby et al 2005  Gelman 2006 |
| Sat | Latex - inline R codes  -  Readings - Salway et al 2017  Montgomery 1997  Berry & Berry 2004 |
| Sun | Latex - Working on Models Ch6  - landscape pages  - APA format tables  Readings - Cavadino 2017  Berry et al 2013 |

Meeting with Thomas

Question on models, try to make simple Poisson model right

Go with time series model as end point???

Motivation (see if this makes sense)

1. ↑pop↑age = ↑arrival
2. ↑arrivals = ↑ congestion and problem
3. Hospital need to operate at near full capacity due to limitation with resources
4. more detailed understanding about arrival required, abnormal high arrival cause congestion, maybe understanding of abnormality will help with prediction
5. with better prediction, it will help with strategy, ie know when to use plan B send more doctor to a location when it is predicted to have high arrival on the day

O’Sullivan also doing ED, maybe ask him about what his doing? And details on how government is planning to tackle the problem, what research is there? Will my study possible help?

Is bookdown a good way to write notes? Or even thesis???

Tried it, manage to get it working and connected to github, but how to create an website using bookdown??? So can see my notes from a url?

|  |  |
| --- | --- |
| **Days** | **Plans / tasks of the day** |
| Mon | Models |
| Tue |  |
| Wed |  |
| Thu |  |
| Fri |  |
| Sat |  |
| Sun |  |

Work on Model and JAGS codes

Work on 732 Book down notes

|  |  |
| --- | --- |
|  | **Plans / tasks of the day** |
| Mon | simulation |
| Tue | Dentist |
| Wed | Stats 732  Github |
| Thu | simulation  Recruitment skills workshop |
| Fri | simulation  **Meeting with Thomas 1:30** |
| Sat |  |
| Sun |  |

Add seasonality and trend

However Seasonality and trend affect anomaly detection,

So don’t add this now???

Add point and period anomaly to simulation so we knew that there is definitely abnormal counts within our simulation data??? Should see a high >0 probability on the abnormal day

Are we testing for possibility that obs > predicted on certain day or time

If greater than certain threshold we know its probably going to have congestions

Is this Anomaly detection?

So Berry model Xij Y ij change to Yij and Yij hat?

Do we need to compare the posterior for different priors? And figure out what is best??? How are we going to do that what numbers are we looking for?

Grouping by area instead of ICD10?

Run 1 to 560

Problem at 436 to 473, most likely at 457

Sankey diagram