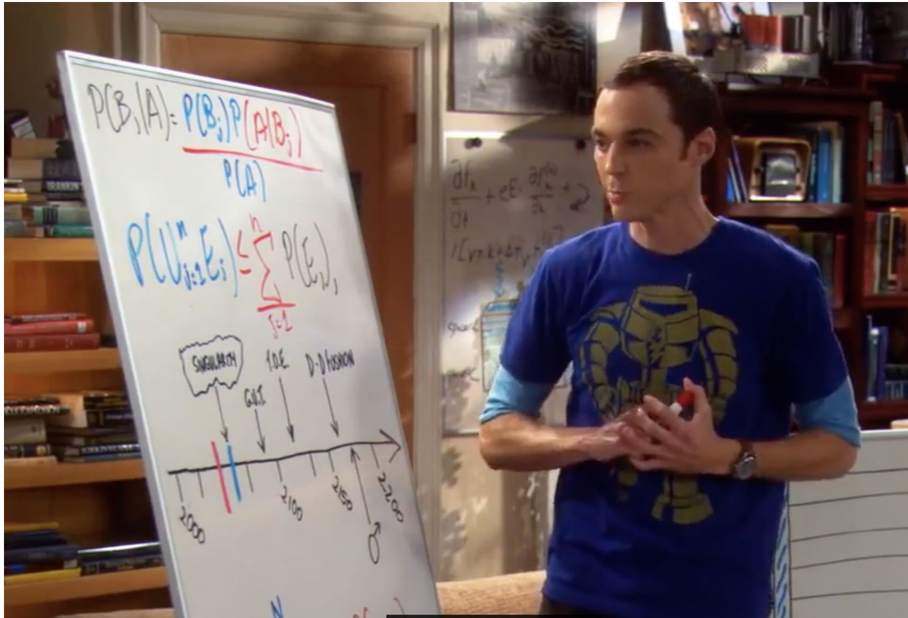
Bayesian Linear regression In Time Series

Lisha Li, Aditya Garg, Jiayi Li, Van Li

Agenda - what to expect?

- Bayesian theorem
- Bayesian linear regression model
- Working example in R
- Summary
- Questions



Bayesian Theorem

Bentley University → Downtown Boston

30 minutes

Friday Afternoon? P(L|T)

Sunday Midnight? P(L|T')

- Previous knowledge/belief
- Based on limited known data
- Educated guess!

Bayesian Theorem

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Late caused by Traffic　　Probability of Late

Late| Traffic $P(L|T) = \dfrac{P(T|L) \; * \; P(L)}{P(T)}$

Probability of Traffic

Posterior $\propto$ likelihood * Prior

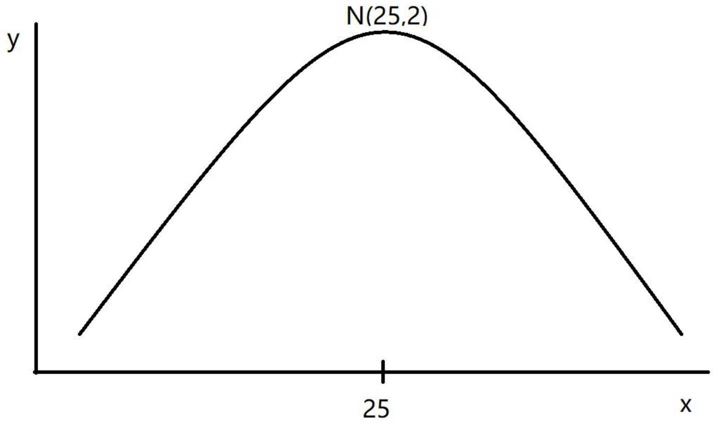
Example!

Bentley University → Downtown Boston

N(25,2) Normal Dist. (Prior)

3 Measurements - 24, 27, 40 (Likelihood)

P(Time|Measurement) $\propto$ P(M|T)*P(T)



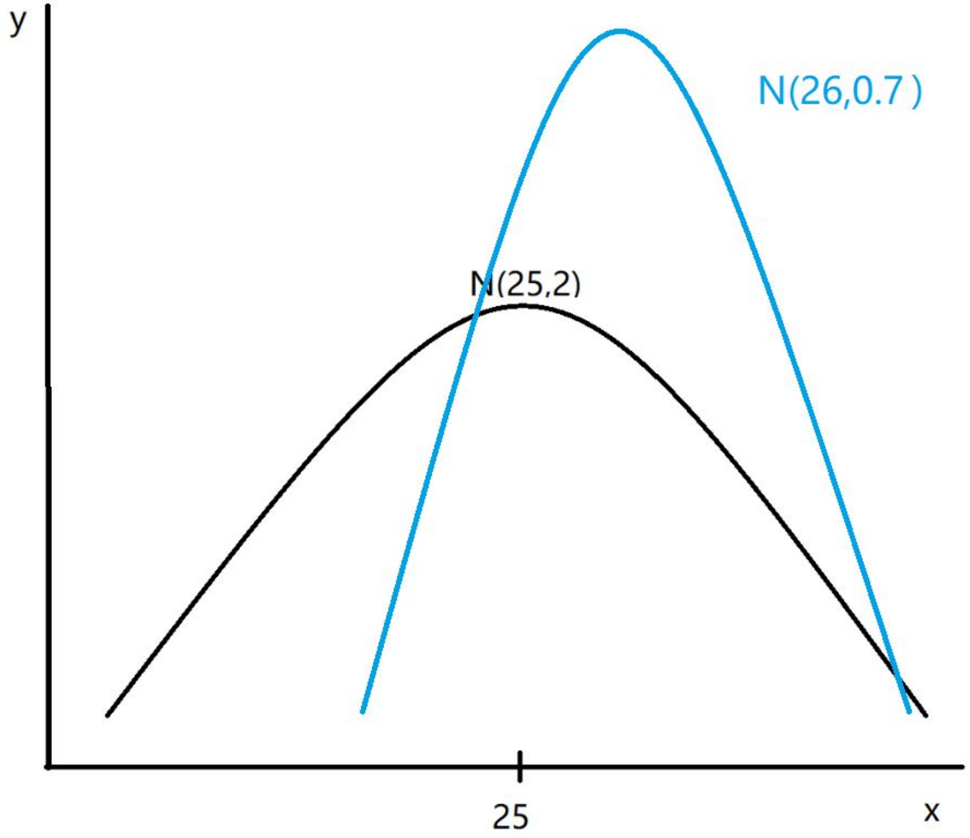
Example!

(Prior N(30))

P(T=30|M=24,27,40)

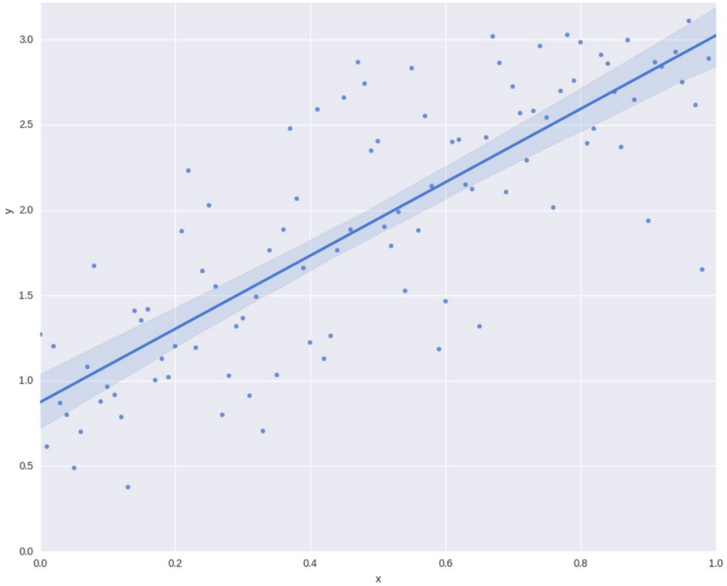
(Prior N(27))

P(T=27|M=24,27,40)



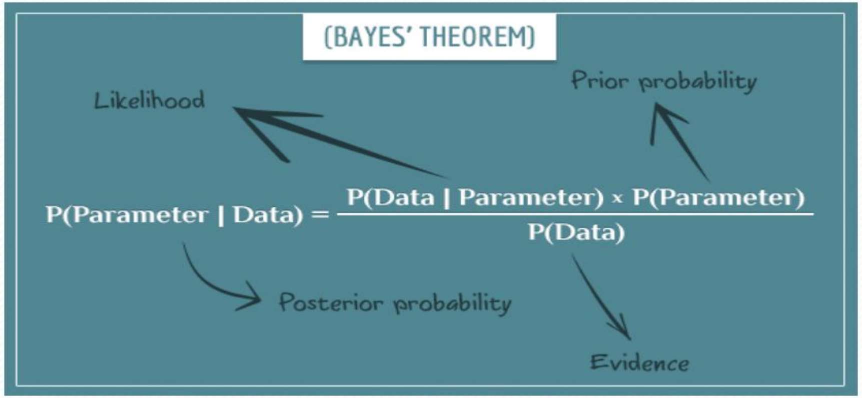
Why Bayesian- Bayesian vs. frequentist

- Data prediction
    - Frequentist: Predict certain data points (most likely value)
    - Bayesian: Full posterior distribution over possible parameter values
- General idea
    - Incorporate uncertainties into the predictions (credible interval)



Why bayesian - Bayesian vs. frequentist (continued)

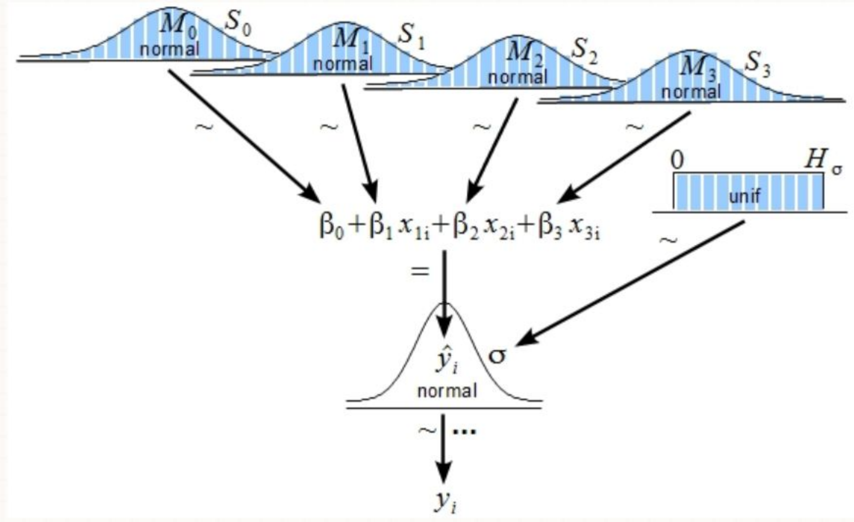
- Definition of probability
  - Frequentist: Long term frequencies for repeatable random events
  - Bayesian: Assign probability to uncertainty (limited data)
- Parameter estimation
  - Frequentist: Collect data and estimate the mean (confidence interval)
  - Bayesian: Define probability distribution over possible values of mean (credible interval)
    - Use sample data to update the distribution (Bayes' Theorem)



$$posterior \propto likelihood * prior$$

Bayesian Linear Regression Model

- Basic model form $y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n,$
- Assumptions
  - Residuals with independence, normality, constant variance, mean zero
- Parameter estimation
  - Assume a prior distribution for parameters (Example: N~(0,variance)) and then use Bayes Theorem
  - Use newly observed data (Sampling method) to get the likelihood
  - Obtain posterior distribution of the parameters
- Prediction: Posterior distribution



Likelihood function Prior belief

$$P(A \mid B) = \frac{P(B \mid A)\boxed{P(A)}}{P(B)}$$

Posterior / new belief

Evidence / marginal likelihood

Bayesian linear Regression model

Parameters Estimation

- MCMC with Gibbs sampling to replicate the posterior distribution
- MCMC - Markov Chain (a sequential process) Monte Carlo (simulation)
- Gibbs Sampler

Gibbs sampler

Starting point (random assignment)

	x_1	x_2	x_3
t_1	a	b	c
t_2	a_1	b	c
t_3	a_1	b_1	c
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t_n	a_n	b_n	c_n

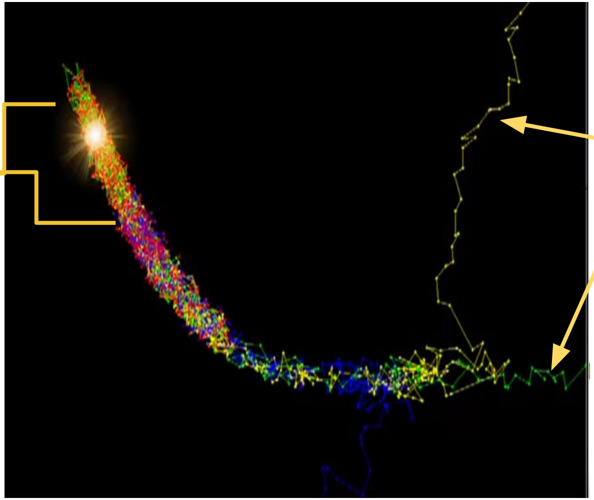
Sample 1

Sample 2

Autocorrelation...Thinning!

Burn in period + Convergence

Likely region



Burn in period

Bayesian linear Regression model

Diagnostics

- Satisfy the linear regression assumptions
- Satisfy the characteristics of MCMC chains
  - Trace plots should show no trends (just consistent random noise around a stable baseline)
  - Density plots follow the same shape as of prior distribution
  - MCMC samples are independent of each other

Forecast

- Using the mean of each parameter distribution (could use median as well)

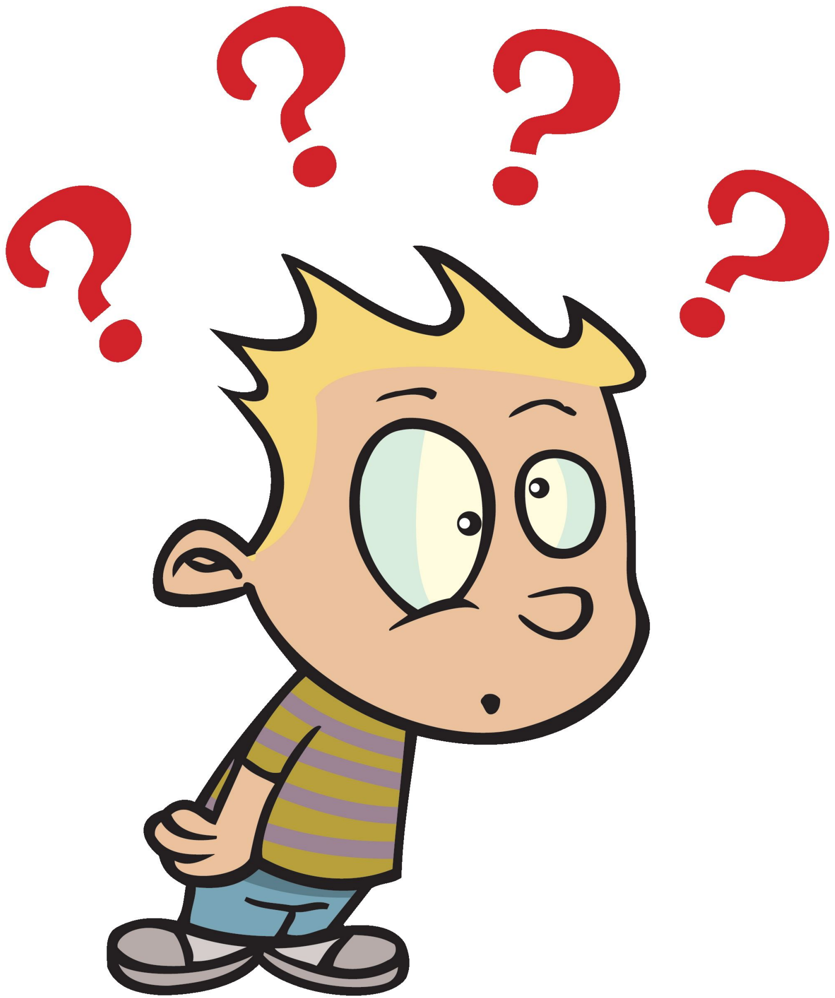
Let's see an Example - sales data for a Product

Data: product.Rdata (credit to Prof. Woolford)

- Number of units sold (in 1,000's) each month
- from May, 1994 to February, 2015

Side notes: go to pdf document

LinkToDocument



References

http://www.flutterbys.com.au/stats/tut/tut7.2b.html

http://www.flutterbys.com.au/stats/tut/tut4.3.html#h3_3

http://www.indiana.edu/~kruschke/BMLR/

http://web2.uconn.edu/cyberinfra/module3/Downloads/Day%202%20-%20Bayes%20Intro.pdf

https://pdfs.semanticscholar.org/4d42/58ef118cf066bb1c2b4c9491286335653b72.pdf

http://mc-stan.org/bayesplot/index.html