



UPPSALA
UNIVERSITET

• Introduction

Bayesian Statistics and Data Analysis

Lecture 1

Måns Magnusson
Department of Statistics, Uppsala University



UPPSALA
UNIVERSITET

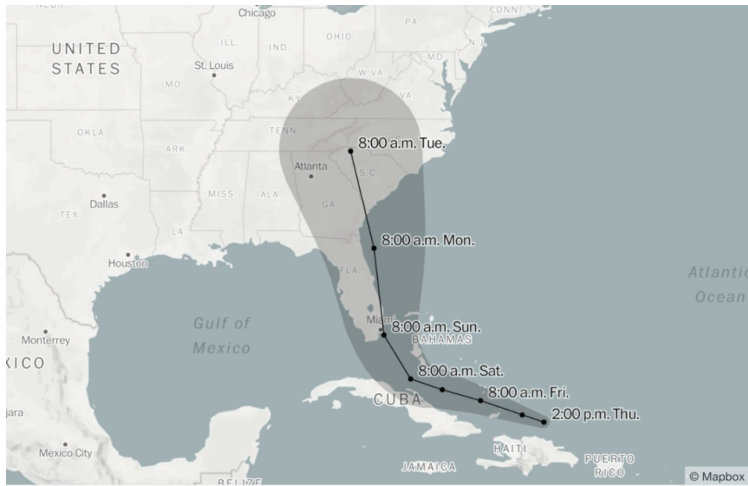
● Introduction

Section 1

Introduction



Decision making in case of uncertainties





• Introduction

- Based on Bayesian probability theory
 - uncertainty is presented with probabilities
 - probabilities are updated based on new information
- Thomas Bayes (170?–1761)
 - English nonconformist, Presbyterian minister, mathematician
 - considered the problem of *inverse probability*
 - significant part of the Bayesian theory
- Bayes did not invent all, but was first to solve problem of inverse probability in special case
- Modern Bayesian theory with rigorous proofs developed in 20th century



Term Bayesian used first time in mid 20th century

- Earlier there was just "probability theory"
 - concept of the probability was not strictly defined, although it was close to modern Bayesian interpretation
 - in the end of 19th century there were increasing demand for more strict definition of probability (mathematical and philosophical problem)
- In the beginning of 20th century frequentist view gained popularity
 - accepts definition of probabilities only through frequencies
 - does not accept inverse probability or use of prior
 - gained popularity due to apparent objectivity and "cook book" like reference books
- R. A. Fisher used in 1950 first time term "Bayesian" to emphasize the difference to general term "probability theory"
 - term became quickly popular, because alternative descriptions were longer



UPPSALA
UNIVERSITET

Uncertainty and probabilistic modeling

- Introduction

- Two types of uncertainty: aleatoric and epistemic
- Representing uncertainty with probabilities
- Updating uncertainty



Two types of uncertainty

- Introduction

- Aleatoric uncertainty due to randomness
 - we are not able to obtain observations which could reduce this uncertainty
- Epistemic uncertainty due to lack of knowledge
 - we are able to obtain observations which can reduce this uncertainty
 - two observers may have different epistemic uncertainty



Updating uncertainty

- Probability of red $\frac{\#red}{\#red + \#yellow} = \theta$
- $p(y = \text{red}|\theta) = \theta$ aleatoric uncertainty
- $p(\theta)$ epistemic uncertainty
- Picking many chips updates our uncertainty about the proportion
- $p(\theta|y = \text{red, yellow, red, red, } \dots) = ?$
- Bayes rule $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$

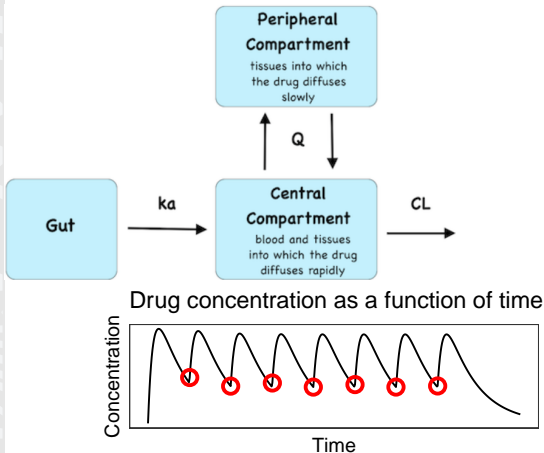


- Bayes rule $p(\theta|y) \propto p(y|\theta)p(\theta)$
- Model: $p(y|\theta)$ as a function of y given fixed θ describes the aleatoric uncertainty
- Likelihood: $p(y|\theta)$ as a function of θ given fixed y provides information about epistemic uncertainty, but is not a probability distribution
- Bayes rule combines the likelihood with prior uncertainty $p(\theta)$ and transforms them to updated posterior uncertainty



Example application: Drug dosage for liver transplant¹

- Everolimus is immunosuppressant to prevent rejection of organ transplants
- Pharmacokinetic model of drug and body, optimal dosage depends on weight





The art of probabilistic modeling

- The art of probabilistic modeling is to describe in a mathematical form (model and prior distributions) what we already know and what we don't know
- “Easy” part is to use Bayes rule to update the uncertainties
 - computational challenges
- Other parts of the art of probabilistic modeling are, for example,
 - model checking: is data in conflict with our prior knowledge?
 - presentation: presenting the model and the results to the application experts



- Galaxy clusters for cosmology
- Coagulation of blood
- Gene regulation
- Pharmacokinetics and -dynamics
- Decision support
- Effects of nutrition for diabetes
- Evolutionary anthropology
- Clinical trial designs
- Daily demand for gas
- Brain structure trees
- School enrollment
- Sports
- Product demand
- Cocoa bean fermentation
- Marine propulsion power
- Alcohol consumption trends
- Flood probability
- Instantaneous heart rate distributions
- Drug dosing regimens in pediatrics
- Human T stem cell memory cells
- Fairness in university admission policies
- Destruction of bacteria and bacterial spores under heat



- Introduction

- Treatment/control
 - randomize patients to treatment or control
 - is the treatment effective?
- Continuous valued treatment
 - randomize patients with different dosages
 - which dosage is sufficient without too many side effects?
- Different effects for different patients?
 - Is the treatment effect different for male/female, child/adult, light/heavy, ...



UPPSALA
UNIVERSITET

Bayesian approach

- Introduction

- Benefits of Bayesian approach
 - integrate over uncertainties to focus to interesting parts
 - use relevant prior information
 - hierarchical models
 - model checking and evaluation



Computation

We need to be able to compute expectations with respect to posterior distribution $p(\theta|y)$

$$E_{\theta|y} [g(\theta)] = \int p(\theta|y)g(\theta)d\theta$$

- Analytic
 - only for very simple models
- Monte Carlo, Markov chain Monte Carlo
 - generic
- Distributional approximations
 - e.g. Laplace, variational, expectation propagation
 - less generic, but can be much faster with sufficient accuracy



UPPSALA
UNIVERSITET

● Introduction

Probabilistic programming



Enables agile workflow for developing probabilistic models

language – automated inference – diagnostics



mc-stan.org



UPPSALA
UNIVERSITET

Binomial model for treatment/control comparison

- Introduction

- Two groups of patients: treatment and control
 - Binary outcome
 - Is the treatment useful?



Binomial model for treatment/control comparison

```
data {  
  int<lower=0> N1;  
  int<lower=0> y1;  
  int<lower=0> N2;  
  int<lower=0> y2;  
}  
parameters {  
  real<lower=0,upper=1> theta1;  
  real<lower=0,upper=1> theta2;  
}  
model {  
  theta1 ~ beta(1,1);  
  theta2 ~ beta(1,1);  
  y1 ~ binomial(N1,theta1);  
  y2 ~ binomial(N2,theta2);  
}  
generated quantities {  
  real oddsratio;  
  oddsratio = (theta2/(1-theta2))/(theta1/(1-theta1));  
}
```



Binomial model for treatment/control comparison

- Introduction

RStanARM

```
fit_bin2 <- stan_glm(y/N ~ grp2, family = binomial(),  
                    data = d_bin2, weights = N)
```



- Introduction

- Drop a ball from different heights and measure time
 - Newton
 - air resistance, air pressure, shape and surface structure of the ball
 - relativity
- Taking into account the accuracy of the measurements, how accurate model is needed?
 - often simple models are adequate and useful
 - *All models are wrong, but some of them are useful,* George P. Box



Reminder: Uncertainty and probabilistic modeling

- Introduction

- Two types of uncertainty: aleatoric and epistemic
- Representing uncertainty with probabilities
- Updating uncertainty



Questions

- Pick a number between 1–5
 - raise as many fingers
 - is the number of fingers raised random (by you or by others)?
- If we build a robot with very fast vision which can observe the rotating coin accurately, is the throw random for the robot?
- Is the quantum uncertainty aleatoric or epistemic?
- What is your own example with both aleatoric and epistemic uncertainty?



UPPSALA
UNIVERSITET

Rest of the course

- Introduction

- Basic models which can be used as building blocks
- Basic computation
- Typical simple scientific data analysis cases
 - e.g. comparison of treatments
- Presentation of the results



UPPSALA
UNIVERSITET

● Introduction

Some important terms

- probability
- probability density
- probability mass
- probability density function (pdf)
- probability mass function (pmf)
- probability distribution
- discrete probability distribution
- continuous probability distribution
- cumulative distribution function (cdf)
- likelihood



Ambiguous notation in statistics

In $p(y|\theta)$

- y can be variable or value
we could clarify by using $p(Y|\theta)$ or $p(y|\theta)$
- θ can be variable or value
we could clarify by using $p(y|\Theta)$ or $p(y|\theta)$
- p can be a discrete or continuous function of y or θ
we could clarify by using P_Y , P_Θ , p_Y or p_Θ
- $P_Y(Y|\Theta = \theta)$ is a probability mass function, sampling distribution, observation model
- $P(Y = y|\Theta = \theta)$ is a probability
- $P_\Theta(Y = y|\Theta)$ is a likelihood function (can be discrete or continuous)
- $p_Y(Y|\Theta = \theta)$ is a probability density function, sampling distribution, observation model
- $p(Y = y|\Theta = \theta)$ is a density
- $p_\Theta(Y = y|\Theta)$ is a likelihood function (can be discrete or continuous)
- y and θ can also be mix of continuous and discrete
- Due to the sloppiness sometimes likelihood is used to refer $P_{Y,\theta}(Y|\Theta)$, $p_{Y,\theta}(Y|\Theta)$