STATS 551

Basics of Bayesian Inference

Yang Chen

Department of Statistics

University of Michigan

ychenang@umich.edu

January 13, 2020

Overview

1. Bayesian Data Analysis
2. Three Steps of Bayesian Analysis
3. Examples

Bayesian Data Analysis

Quantities we observe

Data.

Bayesian Data Analysis

Quantities we observe

Data.

Quantities we wish to learn

Parameters.

Bayesian Data Analysis

Quantities we observe	Quantities we wish to learn
Data.	Parameters.

- Statistical inference:

  estimate unknown (parameters) from known (data).

Bayesian Data Analysis

Quantities we observe	Quantities we wish to learn
Data.	Parameters.

- Statistical inference:

  estimate unknown (parameters) from known (data).

- Bayesian methods:

  quantify uncertainty in statistical inferences.

Bayesin Inference

- The process of ‘inductive thinking’ via Bayes’ rule.

Bayesin Inference

- The process of ‘inductive thinking’ via Bayes’ rule.
- Bayesian methods provide
  - models for rational, quantitative learning
  - estimators that work for small and large sample sizes
  - methods for generating statistical procedures in complicated problems

Bayesian Data Analysis: Example

A clinical trial of a new cancer drug.

Quantities we observe

Patients' outcome under
new drug or standard treatment

Bayesian Data Analysis: Example

A clinical trial of a new cancer drug.

Quantities we observe	Quantities we wish to learn
Patients' outcome under new drug or standard treatment	5-year survival probability of new drug v.s. standard treatment

Bayesian Data Analysis: Example

A clinical trial of a new cancer drug.

Quantities we observe	Quantities we wish to learn
Patients' outcome under new drug or standard treatment	5-year survival probability of new drug v.s. standard treatment

- Statistical inference:

  estimate survival probabilities from patient outcomes.

Bayesian Data Analysis: Example

A clinical trial of a new cancer drug.

Quantities we observe	Quantities we wish to learn
Patients' outcome under new drug or standard treatment	5-year survival probability of new drug v.s. standard treatment

- Statistical inference:

  estimate survival probabilities from patient outcomes.

- Bayesian methods:

  quantify uncertainty in the estimated survival probabilities.

Three Steps of Bayesian Analysis

- Setting up a full probability model
    - Joint probability distribution of all observed & unobserved quantities

Three Steps of Bayesian Analysis

- Setting up a full probability model
  - Joint probability distribution of all observed & unobserved quantities
- Conditioning on observed: posterior distribution
  - Conditional probability distribution of unobserved given observed

Three Steps of Bayesian Analysis

- Setting up a full probability model
  - Joint probability distribution of all observed & unobserved quantities
- Conditioning on observed: posterior distribution
  - Conditional probability distribution of unobserved given observed
- Evaluating fitting & Interpreting posterior distributions.

Three Steps of Bayesian Analysis

- Setting up a full probability model
  - Joint probability distribution of all observed & unobserved quantities

Three Steps of Bayesian Analysis

- Setting up a full probability model
  - Joint probability distribution of all observed & unobserved quantities

Notation

Greek letters: parameters, e.g. θ.
Roman letters: observations, e.g. y_i, $1 \leq i \leq n$.

Three Steps of Bayesian Analysis

- Setting up a full probability model
  - Joint probability distribution of all observed & unobserved quantities

Notation

Greek letters: parameters, e.g. θ.
Roman letters: observations, e.g. y_i, $1 \le i \le n$.

Clinical trial example

For $1 \le i \le n$, $y_i = 1$ if alive and 0 otherwise. θ is probability of survival.

Three Steps of Bayesian Analysis

- Setting up a full probability model
  - Joint probability distribution of all observed & unobserved quantities

Notation

Greek letters: parameters, e.g. θ.
Roman letters: observations, e.g. y_i, $1 \leq i \leq n$.

Clinical trial example

For $1 \leq i \leq n$, $y_i = 1$ if alive and 0 otherwise. θ is probability of survival.

Mathematically

$$y_i \stackrel{i.i.d.}{\sim} p(\cdot|\theta), 1 \leq i \leq n; \quad \theta \sim p(\cdot).$$

$p(\cdot|\theta)$: conditional probability density (distribution); $p(\cdot)$: marginal distribution. Same notation for continuous & discrete densities.

Three Steps of Bayesian Analysis

Numerical formulation of joint beliefs about y and θ:

Three Steps of Bayesian Analysis

Numerical formulation of joint beliefs about y and θ:

- For each θ, our prior distribution $p(\theta)$ describes our belief that θ represents the true population characteristics.

Three Steps of Bayesian Analysis

Numerical formulation of joint beliefs about y and θ:

- For each θ, our prior distribution $p(\theta)$ describes our belief that θ represents the true population characteristics.
- For each θ and y, our sampling model $p(y|\theta)$ describes our belief that y would be the outcome of our study if we knew θ to be true.

Three Steps of Bayesian Analysis

Numerical formulation of joint beliefs about y and θ:

- For each θ, our prior distribution $p(\theta)$ describes our belief that θ represents the true population characteristics.
- For each θ and y, our sampling model $p(y|\theta)$ describes our belief that y would be the outcome of our study if we knew θ to be true.

Once we obtain data y, the last step is to update our beliefs about θ:

$$\cdots\cdots$$

Three Steps of Bayesian Analysis

- Conditioning on observed: posterior distribution
    - Conditional probability distribution of unobserved given observed

Three Steps of Bayesian Analysis

- Conditioning on observed: posterior distribution
  - Conditional probability distribution of unobserved given observed

Bayes Rule

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)},$$

where $p(y) = \sum_\theta p(\theta)p(y|\theta)$, $y = \{y_1, y_2, \cdots\}$.

Three Steps of Bayesian Analysis

- Conditioning on observed: posterior distribution
    - Conditional probability distribution of unobserved given observed

Bayes Rule

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)},$$

where $p(y) = \sum_\theta p(\theta)p(y|\theta)$, $y = \{y_1, y_2, \cdots\}$.

Unnormalized posterior density

$$p(\theta|y) \propto p(\theta)p(y|\theta).$$

Three Steps of Bayesian Analysis

- Conditioning on observed: posterior distribution
  - Conditional probability distribution of unobserved given observed

Bayes Rule

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)},$$

where $p(y) = \sum_\theta p(\theta)p(y|\theta)$, $y = \{y_1, y_2, \cdots\}$.

Unnormalized posterior density

$$p(\theta|y) \propto p(\theta)p(y|\theta).$$

normalizing constant: $p(y)$ — y is observed.

Three Steps of Bayesian Analysis

- Conditioning on observed: posterior distribution
  - Conditional probability distribution of unobserved given observed

Bayes Rule

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)},$$

where $p(y) = \sum_\theta p(\theta)p(y|\theta)$, $y = \{y_1, y_2, \cdots\}$.

Unnormalized posterior density

$$p(\theta|y) \propto p(\theta)p(y|\theta).$$

normalizing constant: $p(y)$ — y is observed.
likelihood: $p(y|\theta)$ — clinical trial example, $p(y|\theta) = \prod_{i=1}^n p(y_i|\theta)$.

Three Steps of Bayesian Analysis: Summary

Clinical trial for a new cancer drug example.

Three Steps of Bayesian Analysis: Summary

Clinical trial for a new cancer drug example.

1. Setting up a full probability model.

$$y_i \overset{i.i.d.}{\sim} p(\cdot|\theta), 1 \leq i \leq n; \quad \theta \sim p(\cdot).$$

Three Steps of Bayesian Analysis: Summary

Clinical trial for a new cancer drug example.

1. Setting up a full probability model.

$$y_i \stackrel{i.i.d.}{\sim} p(\cdot|\theta), 1 \leq i \leq n; \quad \theta \sim p(\cdot).$$

2. Obtaining posterior distribution.

$$p(\theta|\{y_1, y_2, \cdots, y_n\}) \propto p(\theta) \left\{ \prod_{i=1}^{n} p(y_i|\theta) \right\}.$$

Three Steps of Bayesian Analysis: Summary

Clinical trial for a new cancer drug example.

1. Setting up a full probability model.

$$y_i \overset{i.i.d.}{\sim} p(\cdot|\theta), 1 \le i \le n; \quad \theta \sim p(\cdot).$$

2. Obtaining posterior distribution.

$$p(\theta|\{y_1, y_2, \cdots, y_n\}) \propto p(\theta) \left\{ \prod_{i=1}^{n} p(y_i|\theta) \right\}.$$

3. Evaluating fitting & Interpreting posterior distributions.

Three Steps of Bayesian Analysis

Numerical formulation of joint beliefs about y and θ:

- For each θ, our prior distribution $p(\theta)$ describes our belief that θ represents the true population characteristics.
- For each θ and y, our sampling model $p(y|\theta)$ describes our belief that y would be the outcome of our study if we knew θ to be true.

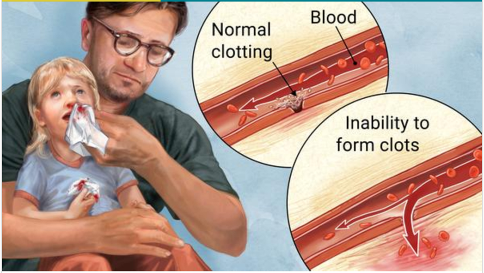
Once we obtain data y, the last step is to update our beliefs about θ:

- For each value of θ, our posterior distribution $p(\theta|y)$ describes our belief that θ is the true value, after having observed data y.

Example: inference about a genetic status

Hemophilia

ABOUT SYMPTOMS TREATMENTS

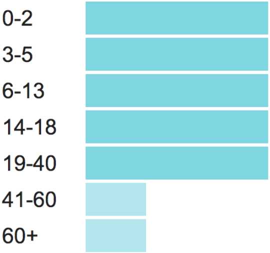


A disorder in which blood doesn't clot normally.

Rare

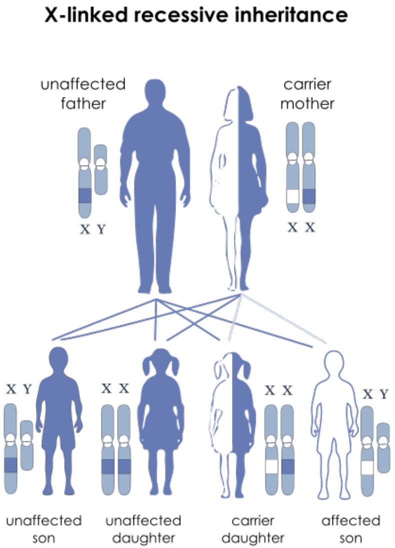
Fewer than 200,000 US cases per year

Ages affected



Genders affected

Example: inference about a genetic status



Humans

Male: XY chromosome.

Female: XX chromosome.

Hemophilia

- Male with the disease-causing gene on X: affected.
- Female with the disease-causing gene on one of two X: not affected.
- Female with the disease-causing gene on both two X: affected.

Example: inference about a genetic status

What is the genetic status of an unaffected woman with an affected brother, an unaffected father and an unaffected mother?

Example: inference about a genetic status

What is the genetic status of an unaffected woman with an affected brother, an unaffected father and an unaffected mother?

Only two possibilities: a carrier of the gene ($\theta = 1$) or not ($\theta = 0$).

Example: inference about a genetic status

What is the genetic status of an unaffected woman with an affected brother, an unaffected father and an unaffected mother?

Only two possibilities: a carrier of the gene ($\theta = 1$) or not ($\theta = 0$).

Prior Distribution

- Affected brother + Unaffected mother $\Rightarrow$
  Mother carries one 'good' and one 'bad' gene.

Example: inference about a genetic status

What is the genetic status of an unaffected woman with an affected brother, an unaffected father and an unaffected mother?

Only two possibilities: a carrier of the gene ($\theta = 1$) or not ($\theta = 0$).

Prior Distribution

- Affected brother + Unaffected mother $\Rightarrow$
  Mother carries one 'good' and one 'bad' gene.
- + Unaffected father $\Rightarrow$
  She has fifth-fifty chance of having the gene.

Example: inference about a genetic status

What is the genetic status of an unaffected woman with an affected brother, an unaffected father and an unaffected mother?

Only two possibilities: a carrier of the gene ($\theta = 1$) or not ($\theta = 0$).

Prior Distribution

- Affected brother + Unaffected mother $\Rightarrow$
  Mother carries one 'good' and one 'bad' gene.
- + Unaffected father $\Rightarrow$
  She has fifth-fifty chance of having the gene.
- Prior distribution for θ:

$$P(\theta = 1) = P(\theta = 0) = \frac{1}{2}.$$

Example: inference about a genetic status

Data

Neither of her two sons is affected ($y_1 = y_2 = 0$).

- Two sons: independent and not identical twins.

Example: inference about a genetic status

Data

Neither of her two sons is affected ($y_1 = y_2 = 0$).

- Two sons: independent and not identical twins.

Likelihood

Example: inference about a genetic status

Data

Neither of her two sons is affected ($y_1 = y_2 = 0$).

- Two sons: independent and not identical twins.

Likelihood

$$P(y_1 = y_2 = 0|\theta = 1) = 0.5 \times 0.5 = 0.25,$$
$$P(y_1 = y_2 = 0|\theta = 0) = 1 \times 1 = 1.$$

Example: inference about a genetic status

Posteiror Distribution

Let $y = (y_1, y_2)$, then Bayes' rule gives

Example: inference about a genetic status

Posteiror Distribution

Let $y = (y_1, y_2)$, then Bayes' rule gives

$$
\begin{aligned}
P(\theta = 1|y) &= \frac{P(y|\theta = 1)P(\theta = 1)}{P(y|\theta = 1)P(\theta = 1) + P(y|\theta = 0)P(\theta = 0)} \\
&= \frac{0.25 * 0.5}{0.25 * 0.5 + 1 * 0.5} = 0.2,
\end{aligned}
$$

which is smaller than 0.5 (given by the prior).

Example: inference about a genetic status

Posteiror Distribution

Let $y = (y_1, y_2)$, then Bayes' rule gives

$$P(\theta = 1|y) = \frac{P(y|\theta = 1)P(\theta = 1)}{P(y|\theta = 1)P(\theta = 1) + P(y|\theta = 0)P(\theta = 0)}$$
$$= \frac{0.25 * 0.5}{0.25 * 0.5 + 1 * 0.5} = 0.2,$$

which is smaller than 0.5 (given by the prior).

Adding more data

Another unaffected son.

Example: inference about a genetic status

Posteiror Distribution

Let $y = (y_1, y_2)$, then Bayes' rule gives

$$P(\theta = 1|y) = \frac{P(y|\theta = 1)P(\theta = 1)}{P(y|\theta = 1)P(\theta = 1) + P(y|\theta = 0)P(\theta = 0)}$$
$$= \frac{0.25 * 0.5}{0.25 * 0.5 + 1 * 0.5} = 0.2,$$

which is smaller than 0.5 (given by the prior).

Adding more data

Another unaffected son.

$$P(\theta = 1|y_1, y_2, y_3) = \frac{0.5 * 0.2}{0.5 * 0.2 + 1 * 0.8} = 0.111.$$

Example: spelling correction

Suppose someone types “radom”. How should that be read?

Example: spelling correction

Suppose someone types "radom". How should that be read?

- misspelling or mistyping of "random" or "radon"?
- intentional typing of "radom"

Example: spelling correction

Suppose someone types "radom". How should that be read?

- misspelling or mistyping of "random" or "radon"?
- intentional typing of "radom"

What is the probability that "radom" actually means "random"?

Example: spelling correction

Suppose someone types “radom”. How should that be read?

- misspelling or mistyping of “random” or “radon”?
- intentional typing of “radom”

What is the probability that “radom” actually means “random”?

Label y as the data and θ as the word that the person was intending to type, then

$$Pr(\theta|y = “radom'') \propto p(\theta)Pr(y =' radom'|\theta).$$

This product is the unnormalized posterior density.

Example: spelling correction

Let (θ_1 = random, θ_2 = radon, and θ_3 = radom), then

$$p(random|``radom") = \frac{p(\theta_1)p(``radom"|\theta_1)}{\sum_{j=1}^{3} p(\theta_j)p(``radom"|\theta_j)}.$$

Example: spelling correction

Let (θ_1 = random, θ_2 = radon, and θ_3 = radom), then

$$p(random|``radom") = \frac{p(\theta_1)p(``radom"|\theta_1)}{\sum_{j=1}^{3} p(\theta_j)p(``radom"|\theta_j)}.$$

Specify "Priors" $p(\theta)$ and "Likelihood" $p(``radom"|\theta)$.

Example: spelling correction

Let (θ_1 = random, θ_2 = radon, and θ_3 = radom), then

$$p(random|\text{``}radom'') = \frac{p(\theta_1)p(\text{``}radom''|\theta_1)}{\sum_{j=1}^{3} p(\theta_j)p(\text{``}radom''|\theta_j)}.$$

Specify "Priors" $p(\theta)$ and "Likelihood" $p(\text{``}radom''|\theta)$.

Prior: relative word frequency

θ	random	radon	radom
$p(\theta)$	7.60×10^{-5}	6.05×10^{-6}	3.12×10^{-7}

Example: spelling correction

Likelihood: Google spelling & tying error model

θ	random	radon	radom
p("radom" $\|\theta$)	0.00193	0.000143	0.975

Example: spelling correction

Likelihood: Google spelling & tying error model

θ	random	radon	radom
p(“radom” $\vert\theta$)	0.00193	0.000143	0.975

Posterior distribution.

θ	random	radon	radom
$p(\theta)$ p(“radom” $\vert\theta$)	1.47×10^{-7}	8.65×10^{-10}	3.04×10^{-7}
p($\theta\vert$ “radom”)	0.325	0.002	0.673

Example: estimating percentage of Dunkin' lovers

Prior Knowledge + Data = Current Knowledge