Bayesian Data Analysis

Sourish Das[1]

[1]Mathematics,
Chennai Mathematical Institute, INDIA

Introduction

CHENNAI MATHEMATICAL INSTITUTE

Introduction

Bayesian Analysis

Bayesian Analysis is a statistical methodology in which Bayes' Theorem is used to estimate the probability of a hypothesis as data is observed.

CHENNAI MATHEMATICAL INSTITUTE

Plan for This Course

1 Introduction

2 Binomial-Beta Model

3 Gausian-Gausian Model

4 Bayesian Regression Model
Case: Captal Asset Pricing Model

5 Monte Carlo Simulation

Plan for This Course

6 Bayesian Machine Learning - Regression

7 Bayesian Approach to Classification

8 Bayesian Analysis of Count Data with Poisson Model

9 Bayesian Hierarchical Model
Case: Dynamic Pricing with Hierarchical Regression

10 Gaussian Process Regression

11 Bayesian Optimization

Introduction

Agenda:

1. Objective and Subjective definition of probability
2. Axiomatic Definition of Probability
3. Bayes Theorem
4. Applications of Bayes Theorem

The objective (aka. frequency) interpretation

The probability of some particular outcome is the relative frequency with which the outcome would be obtained; if the process were repeated many times under similar conditions.

Example

The probability a head in a fair coin toss is 0.5 because the relative frequency of heads should be approximate 0.5 if we flip the coin many times.

The frequency interpretation

When we make statistical inferences from the frequentist perspective, we assume that our data is a sample from an entire population.

1. The population is described by the population mean and the population variance that are unknown.
2. The sample is described by the sample mean and the sample variance.
3. The sample mean and variance provide estimates about the mean and variance of the entire population.
4. Importantly, these estimates are known only with some uncertainty.

Our uncertainty about a statistic like the mean is summarized by its sampling distribution.

The frequency interpretation

The sampling distribution

Sampling Distribution

The sampling distribution is a probability distribution of all possible values of a statistic of interest for samples of size N that could be formed for a given population.

- The observed sample mean is just one realization.

CHENNAI MATHEMATICAL INSTITUTE

The frequency interpretation

The sampling distribution

Problem

Beyond some text book cases finding the exact sampling distribution is difficult task.

Solution

Frequentist approach to the problem is to approximate the sampling distribution by known distribution like Gaussian or t distribution under the assumption like sample size N is large.

Critique 1

Needless to say, this is a theoretical construct since, with a large population, there will be billions of unique samples and it would be superior to simply sample the entire population.

The frequency interpretation

The sampling distribution

Problem

Beyond some text book cases finding the exact sampling distribution is difficult task.

Solution

Frequentist approach to the problem is to approximate the sampling distribution by known distribution like Gaussian or t distribution under the assumption like sample size N is large.

Critique 2

P-values refer to the proportion of hypothetical draws from the sampling distribution that are consistent with the null hypothesis. As p-values are based on the concept of a sampling distribution, do they make sense if our data contains the almost entire population?

The Classical Interpretation of Probability

1. The classical interpretation is based on the concept of equally likely outcomes.
2. If the outcome of some process must be one of n different outcomes, and if these outcomes are equally likely to occur, then the probability of each outcome is $\frac{1}{n}$.

CHENNAI MATHEMATICAL INSTITUTE

The Classical Interpretation of Probability

1. If we flip a fair coin, the probability of a head would be $\frac{1}{2}$ because head and tail are equally likely outcomes.
2. The classical approach offers an appealing summary of uncertainty in a one-shot situation.

CHENNAI MATHEMATICAL INSTITUTE

Problems with the Classical Interpretation

The drawback of the classical interpretation is that the concept of equally likely outcomes is itself probabilistic.

1. In a sense, this makes the classical definition of probability circular.
2. Furthermore, the concept begins to break down in contexts other than gambling when events are not equally likely.

The classical response is...

1. Laplace's Rule of Insufficient Reason: in the absence of compelling evidence to the contrary, we should assume that events are equally likely.
2. This concept response is actually more useful to Bayesians when defending their priors.

The Subjective Interpretation of Probability

1. The probability that a person assigns to a possible outcome of some process represents his or her own judgment of the likelihood that the outcome will be obtained.

2. In contrast to the classical and frequentist interpretations of probability, this means that different individuals could have different probability judgments.

Example

If we flip a fair coin, the probability of a head could be $\frac{3}{4}$ because, for some reason, we think that *God wants it to be a head*.

CHENNAI MATHEMATICAL INSTITUTE

The Subjective Interpretation of Probability

1. Is subjective probability theory really that *ad-hoc*?

2. Not Necessarily...Bayesian methodology elicit priors in a manner that ensures coherence.

3. The main objective of this course is to develop understanding about how to elicit priors that regularize the solution coherently.

CHENNAI MATHEMATICAL INSTITUTE

Axiomatic Definition of Probability

A probability distribution on a sample space S is a specification of numbers $Pr(A_i)$ which satisfy:

Axiom 1

For any outcome A_i, $Pr(A_i) \geq 0$.

Axiom 2

$Pr(S) = 1$.

Axiom 3

For a sequence of disjoint events $A_1, A_2, ...$

$$Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(A_i)$$

It turns out that each of these three axioms can be justified using the coherence criterion.

Theorems of Probability

1. $P(\Phi) = 0$
2. For any finite sequence of disjoint events $\{A_1, A_2, ..., A_n\}$

$$Pr(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pr(A_i)$$

3. For any event A, $Pr(A^c) = 1 - Pr(A)$
4. For any event A, $0 \leq Pr(A) \leq 1$
5. For any two events A and B, $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

Independence Events

Idea of Independence

Two events A and B are independent if the occurrence or non-occurrence of one of the events has no influence on the occurrence or non-occurrence of the other event.

Independence Mathematical Definition

Two events A and B are independent if

$$Pr(A \cap B) = Pr(A)Pr(B)$$

Independent Events

Example of Independence

Are 'Smoking' and 'Lung Cancer' independent?

Suppose

1. $Pr(Smoking) = 0.4$,
2. $Pr(Lung\ \ Cancer) = 0.5$,
3. $Pr(Lung\ \ Cancer\ \ \cap\ \ Smoking) = 0.35$

Clearly, $Pr(Lung\ \ Cancer) \times Pr(Smoking) = 0.4 \times 0.5 = 0.2 \neq 0.35 = Pr(Lung\ \ Cancer\ \ \cap\ \ Smoking)$

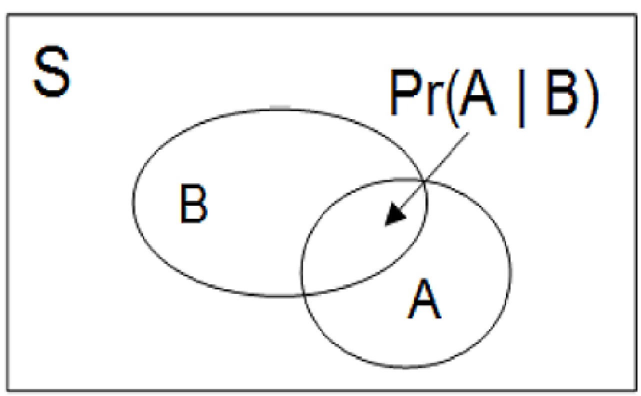
Conditional Probability

Conditional probabilities allow us to understand how the probability of an event A changes after it has been learned that some other event B has occurred.

- The key concept for thinking about conditional probabilities is that the occurrence of B reshapes the sample space for subsequent events.
- That is, we begin with a sample space S
- A and $B \in S$
- The conditional probability of A given that B looks just at the subset of the sample space for B.

CHENNAI MATHEMATICAL INSTITUTE

Conditional Probability



CHENNAI MATHEMATICAL INSTITUTE

Conditional Probability

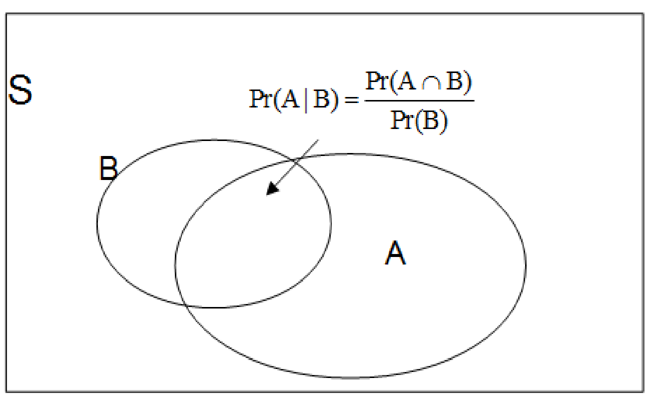
1. The conditional probability of A given B is denoted $Pr(A|B)$.

2. Importantly, according to Bayesian orthodoxy, all probability distributions are implicitly or explicitly conditioned on the model.

3. By definition: If A and B are two events such that $Pr(B) > 0$, then
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

CHENNAI MATHEMATICAL INSTITUTE

Conditional Probability



CHENNAI MATHEMATICAL INSTITUTE

Conditional Probability

Example

1. What is the $Pr(Lung\ Cancer|Smoking)$?
2. $Pr(Lung\ Cancer \cap Smoking) = 0.35$
3. $Pr(Smoking) = .4$
4. Thus, $Pr(Lung\ Cancer|Smoking) = .35/.4 = .875$

Properties of Conditional Probability

1. **The Conditional Probability for Independent Events**: If A and B are independent then $P(A|B) = P(A)$
2. **The Multiplication Rule for Conditional Probabilities**: In an experiment involving two non-independent events A and B, the probability that both A and B occurs can be found in the following two ways:

   $$Pr(A \cap B) = Pr(B)Pr(A|B)$$

   or

   $$Pr(A \cap B) = Pr(A)Pr(B|A)$$

3. The set of events $\{A_1, ..., A_n\}$ are partition of sample space S, where $\cup_{i=1}^{n} = S$, then

   $$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Bayes' Theorem

Bayes' Theorem

Let events $A_1, ..., A_k$ form a partition of the space S such that $Pr(A_j) > 0$ for all j and let B be any event such that Pr(B) > 0. Then for i = 1,..,k:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Bayes' Theorem

1. Bayes' Theorem is just a simple rule for computing the conditional probability of events A_i given B from the conditional probability of B given each event A_i and the unconditional probability of each A_i
2. $P(A_i)$ is the prior distribution of A_i.
3. $P(B|A_i)$ is the the conditional probability of B given A_i
4. $\sum_{i=1}^{n} P(B|A_i)P(A_i)$ is the normalizing constant
5. $P(A_i|B)$ is the posterior probability of A_i given B

Causation with Conditional Probability

1. If A and B are two events such that $Pr(B) > 0$, then
$$P(A|B) = \frac{P(A \ \ and \ \ B)}{P(B)}$$
2. What is the probability of A when we know the event B has happened?

Causation with Conditional Probability

Example

1. What is the $Pr(Lung \ \ Cancer|Smoking) = \frac{Pr(Lung \ \ Cancer \ \ and \ \ Smoking)}{Pr(Smoking)} =?$
2. What is the probability that some one will develop Lung Cancer when we know the person is smoker?

Causation with Conditional Probability

Example

1. What is the $Pr(Lung\ \ Cancer|Smoking) = \frac{Pr(Lung\ \ Cancer\ \ \ and\ \ \ Smoking)}{Pr(Smoking)} = ?$
2. $Pr(Lung\ \ Cancer\ \ \ and\ \ \ Smoking) = !$
3. $Pr(Smoking) = !$

CHENNAI MATHEMATICAL INSTITUTE

Causation with Conditional Probability

Example

$$\begin{aligned} Pr(Lung\ \ Cancer|Smoking) &= \frac{Pr(Lung\ \ Cancer\ \ \ and\ \ \ Smoking)}{Pr(Smoking)} \\ Pr(LC \mid Sm) &= \frac{Pr(LC\ \ \ and\ \ \ Sm)}{Pr(Sm)} \end{aligned}$$

Most of the time - we cannot calculate these probabilities directly.

CHENNAI MATHEMATICAL INSTITUTE

Bayes Theorem

$$Pr(LC \mid Sm) = \frac{Pr(Sm \mid LC) \times Pr(\ LC)}{Pr(Sm \mid LC) \times Pr(\ LC) + Pr(Sm \mid LC^c) \times Pr(\ LC^c)}$$

1. About 1 out of 15 people develop LC in USA -so $Pr(LC) \approx 1/15 = 0.067$ and $Pr(LC^c) \approx 14/15 = 0.933$
2. "As many as 20% of people who die from lung cancer in the United States every year have never smoked" - so $Pr(Sm^c|LC) = 0.2$ - that means $Pr(Sm \mid LC) = 0.8$

   $Pr(Sm \mid LC^c) \approx 0.2$

   Source: `https://www.cancer.org/latest-news/why-lung-cancer-strikes-nonsmokers.html`

Probable Cause

$$\begin{aligned} Pr(LC \mid Sm) &= \frac{Pr(Sm \mid LC) \times Pr(\ LC)}{Pr(Sm \mid LC) \times Pr(\ LC) + Pr(Sm \mid LC^c) \times Pr(\ LC^c)} \\ &= \frac{0.8 \times 0.067}{0.8 \times 0.067 + 0.2 \times 0.933} = 0.22 \end{aligned}$$

$$Pr(\ LC) = 0.067 \quad \& \quad Pr(LC \mid Sm) = 0.22$$

Bayes probability clearly indicates that if we know somebody is smoker the chance of having Lung Cancer increses by 3 times

Therefore 'smoking' is definitely a very strong 'probable cause' of lung cancer.

Independence

If

$$Pr(A \mid B) = Pr(A)$$

then we can say additional information about B has no impact on A - hence A and B are two independent events.

$$\begin{aligned} \frac{Pr(A \text{ and } B)}{Pr(B)} &= Pr(A) \\ Pr(A \text{ and } B) &= Pr(A).Pr(B) \end{aligned}$$

Rev. Thomas Bayes

Thank You

sourish@cmi.ac.in

www.cmi.ac.in/~sourish