MTH225 Spring 2016

E. Quinn

Contents

CHAPTER 1

Introduction

1.1. Statistical Inference

Statistical inference refers to a collection of techniques used to obtain information from data.

Probability theory is the foundation of statistical inference. Probability theory is based on the notion of a *probability experiment*:

Definition 1.1.1 (probability experiment). *A probability experiment is a repeatable procedure that produces exactly one of a set of possible outcomes each time it is performed.*

Common examples of probability experiments include:

- Tossing a coin
- Rolling a die
- Drawing a card from a shuffled deck
- Picking a number between 0 and 1
- Tossing a coin until the first 'heads' occurs

The important characteristics of a probability experiment are:

- The outcomes must be distinct and well-defined
- Each time the experiment is performed, exacly one outcome occurs
- The experiment can be repeated any number of times

The set of possible outcomes of an experiment is called the *sample space*.

Examples of sample spaces:

Tossing a coin: the sample space is:

$$\Omega = \{\text{heads,tails}\}$$

Rolling a die: the sample space is:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Picking a card from a shuffled deck of 52 standard playing cards: the sample space has 52 elements, each representing one of the 52 cards.

Picking a number between zero and one: the sample space is the set of real numbers between 0 and 1

$$\Omega = \{x : 0 < x < 1\}$$

Tossing a coin until it comes up heads:

$$\Omega = \{[\text{heads}], [\text{tails,heads}], [\text{tails,tails,heads}], [\text{tails,tails,tails,heads}], \ldots\}$$

The objects of study in probability theory, called *random variables*, are just real-valued functions defined on a sample space.

If the elements of a sample space happen to be numbers, there is an obvious way to do this:

- Rolling a die: The value of the random variable is the number on the top face
- Pick a number between zero and one: The value of the random variable is the number chosen

If the outcomes are not numbers, a random variable is defined by assigning a real number to each outcome:

In the coin toss experiment, he most common way to define a random variable is:

- If 'heads', the value of the random variable is 1
- If 'tails', the value of the random variable is 0

This is not the only possibility, we could pick any two distinct numbers to assign to 'heads' and 'tails'.

We could even assign the same number to both 'heads' and 'tails', although the resulting random variable would not be very interesting because it would take the same value regardless of the outcome of the experiment.

In some experiments, such as rolling a pair of dice, it is common to assign the same number to more than one outcome. Each of the outcomes

$$(1,5),(2,4),(3,3),(4,2),(5,1)$$

are typically assigned a value of '6' when defining a random variable on this sample space.

1.2. Random Variables

Random variables are numerical values that represent the outcomes of probability experiments.

There are two distinct types of random variables:

- *Discrete* random variables assume a (possibly infinite) discrete set of values, usually integers.
- *Continuous* random variables assume all values on an interval of the real line.

There are countless examples of both types of random variables. We will examine a few of the most commonly used ones.

1.2.1. Discrete Random Variables.

EXAMPLE 1.2.1. *Bernoulli random variable: The Bernoulli experiment has two outcomes, "success" and "failure". The value assigned to the Bernoulli random variable is:*

Outcome	*Value*
"success"	*1*
"failure	*0*

EXAMPLE 1.2.2. *Geometric random variable: In the geometric experiment we conduct independent Bernoulli trials with a constant probability of success until the first success occurs. The value assigned to*

the geometric random variable is the number of failures preceding the first success:

Outcome	*Value*
H	*0*
TH	*1*
TTH	*2*
TTTH	*3*
TTTTH	*4*
⋮	⋮

EXAMPLE 1.2.3. *Binomial random variable: In the binomial experiment we conduct a predetermined number n of independent Bernoulli trials with constant probability of success. The value assigned to the binomial random variable is the number of successes out of n trials:*

Outcome	*Value*
0 successes	*0*
1 success	*1*
2 successes	*2*
⋮	⋮
n successes	*n*

EXAMPLE 1.2.4. *Poisson random variable: The Poisson experiment produces a count that may be zero or a positive integer. The value assigned to the Poisson random variable is the same as the count.*

Count	*Value*
0	*0*
1	*1*
2	*2*
3	*3*
⋮	⋮

1.2.2. Continuous Random Variables.

EXAMPLE 1.2.5. *Uniform random variable: A uniform random variable assumes values between zero and one:*

$$X \in [0,1] \quad or \quad 0 \leq x \leq 1$$

EXAMPLE 1.2.6. *Normal random variable: A normal random variable random assumes any real number:*

$$X \in (-\infty, \infty) \quad or \quad -\infty < x < \infty$$

EXAMPLE 1.2.7. *Beta random variable: A beta random variable random assumes a value in the interval* $[0,1]$*:*

$$X \in [0,1] \quad or \quad 0 \leq x \leq 1$$

EXAMPLE 1.2.8. *Gamma random variable: A gamma random variable random assumes a value in the interval* $(0,\infty)$*:*

$$X \in (0,\infty) \quad or \quad 0 < x < \infty$$

The exponential distribution is a special case of the gamma distribution.

CHAPTER 2

Probability Distributions

2.1. Probability Distributions

A *probability distribution* associates probabilities with the values of a random variable.

2.1.1. Discrete probability distributions, parameters and probability mass functions. A *discrete probability distribution* associates values of discrete random variables with probabilities through a *probability mass function.*

EXAMPLE 2.1.1. *A Bernoulli random variable takes only two values,* 0 *and* 1*, associated with "failure" and "success", respectively.*

The single parameter θ *represents the probability of "success".*

The probability mass function is defined by:

$$p(x) = \theta^x(1-\theta)^{1-x} \quad \textit{for} \quad x = 0, 1$$

EXAMPLE 2.1.2. *A Poisson random variable takes values* $0, 1, 2, 3, \ldots$.

The single parameter λ *must be greater than zero.*

The probability mass function is defined by:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, 3, \ldots$$

EXAMPLE 2.1.3. *A binomial distribution has two parameters,* n *representing the number of trials, and* p *representing the probability of success on each trial.*

The probability mass function is defined by:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \textit{for} \quad x = 0, 1, 2, \ldots, n$$

2.1.2. Continuous probability distributions, parameters and probability density functions. A *continuous probability distribution* associates ranges of values of a random variable with probabilities through a *probability density function.*

EXAMPLE 2.1.4. *A uniform random variable has two parameters a and b. Both must be greater than zero, and a must be less than b.*

The probability density function is defined by:

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

If a and b are not specified, they are assumed to be zero and one, respectively:

$$f(x) = 1, \quad 0 \leq x \leq 1$$

EXAMPLE 2.1.5. *A normal or Gaussian random variable has two parameters μ (the mean or location) and sigma (the standard deviation or dispersion). μ can take any value, but σ must be positive.*

The probability density function of the normal distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty$$

If $\mu = 0$ and $\sigma = 1$, x is said to have a standard normal distribution. In this case the density function is:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad -\infty < x < \infty$$

EXAMPLE 2.1.6. *A gamma random variable has two parameters α (shape) and beta (scale). Both must be positive.*

The gamma distribution takes values on the interval $(0, \infty)$.

EXAMPLE 2.1.7. *A beta random variable has two parameters* α *(shape) and beta (scale). Both must be positive.*

The beta distribution takes values on the interval $(0, 1)$.

CHAPTER 3

Frequentist and Bayesian Inference

3.1. Frequentist and Bayesian Inference

There are two types of statistical inference, classical or *frequentist* and *Bayesian*.

3.1.1. Frequentist Inference. In classical or frequentist inference,

- Data values are considered to be random variables.
- Parameters are assumed to be fixed, unknown constants.
- Inference is based on hypothetical replication of the experiment that produced the data many times.

3.1.2. Bayesian Inference. In Bayesian inference,

- Data values are considered to be known values, observed from a probability distribution with associated *likelihood function*
- Parameters are assumed to be random variables.
- Each parameter has a *prior* distribution representing the experimenter's belief about the likely values of the parameter, expressed in the form of a probability distribution.
- Inference is based on the *posterior* distribution, which is a combination of the likelihood and the prior distribution, in effect a compromise.

A fully specified Bayesian model consists of:

- A likelihood function for the data
- A prior distribution for the parameter(s)
- A posterior distribution, which is a combination of the prior and the likelihood

The steps in Bayesian inference are:

- Define a prior distribution for each parameter
- Define a likelihood function for the data
- Use software such as STAN to draw a sample from the posterior distribution
- Use the draw from the posterior distribution to estimate quantities of interest like means, standard deviations, and percentiles.