Applied Statistics with R

Working with Probabilities and Probability Distributions

Kaloyan Ganev

2022/2023

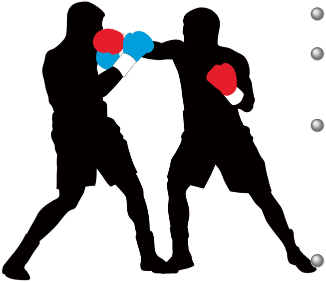
Lecture Contents

Probability

Probability Basics: What is Probability?

- No unified definition!
- Many definitions possible, but in the broad sense it relates to chance (or randomness)
- Randomness itself means that outcomes are unpredictable, i.e. you do not know the exact outcome beforehand
- At the same time, often there are patterns/trends, i.e. some outcomes are more frequent than others – this is what makes the study of probability interesting
- We live in a world full of randomness, so many phenomena are chance outcomes
- Examples: weather, tossing a coin, rolling dice, sales, end-of-year profit, engine malfunctions, etc.

Probability Basics: Frequentist vs. Bayesian



- Still, two major types of definitions co-exist
- The two views are sometimes thought of as opposing each other, but it's not like in the picture :)
- Frequentist: we record many outcomes of random phenomena, group outcomes by some rule, and then measure the relative frequency of each group in the total
- Bayesian: we form some prior knowledge or belief about chance and then use this belief to estimate probability; useful when we do not have the option to collect many observations on a phenomenon

We will stick to the frequentist approach

Probability Basics: A Simple Example

- Let's illustrate how frequentists calculate probability
- A die is cast 5300 times, and you get 1484 sixes. What is the probability of getting a six?

$$P(six) = \frac{1484}{5300} = 28\%$$

- If you were a Bayesian, what would you think of probability?
- What would you think of the die based on the above counts? Is it a fair one?

Probability Basics: Some Historical Names

- Elements of probability theory were developed as early as XVII-XVIII centuries
- Jacob Bernoulli, Geloramo Cardano, Pierre de Fermat, Blaise Pascal, Abraham de Moivre, Thomas Bayes, and Joseph Lagrange developed the first formulas and techniques
- Marquis de Laplace constructed the first general theory of probability in XIX century
- In the beginning of the XX century, Andrey Kolmogorov laid the foundations of modern probability theory
- Applications are wide ranging, practically in all areas of analysis

Probability Basics: Elementary Concepts

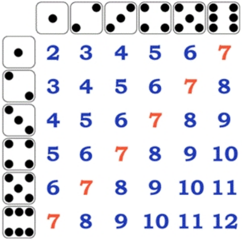
- Back to essential stuff: probability is the chance that something will happen
- Probability is always a number between 0 and one:

$$0 \leq P(X) \leq 1$$

- If X could never happen (example?), then $P(X) = 0$; if X is sure to happen (example?), then $P(X) = 1$
- In probability theory, we call the various possible outcomes of random mechanisms events
- For example, heads or tails in tossing a coin are events
- What are all possible events of tossing two coins? Two dice?

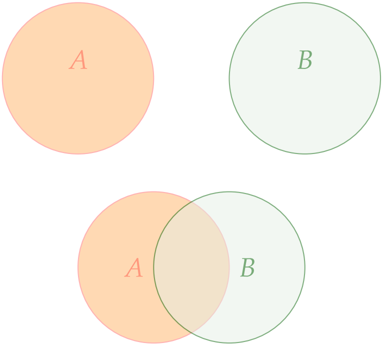
Probability Basics: Elementary Concepts (2)

- The set of all possible events is called the sample space
- Mutually exclusive events are events such that only one of them can occur at a time
- Events that are not mutually exclusive are defined by complementarity
- A list of possible events is collectively exhaustive if it includes every possible outcome



Probability Basics: Rules

- Recall we denoted the probability of X by $P(X)$
- This is called marginal/unconditional probability
- As you can guess, there are also other concepts: joint probability and conditional probability
- To illustrate rules for the two cases of mutual exclusion and non-exclusion, we will us the so-called Venn diagrams



Probability Basics: Addition Rule

- Addition rule: when we are interested in whether one event or the other will occur
- For mutually exclusive events:

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$$

- A and $\overline{A}$ are mutually exclusive, so:

$$P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1$$

- For events that are not mutually exclusive:

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probability Basics: Addition Rule (2)

Two examples:

- What is the probability of winning a five or a six from Sport Toto's 6/49 game?
- You have a regular deck of 52 cards. What is the probability of drawing an ace or a spade?

Probability under Statistical Independence

- Relates to two or more events
- Take the simplest case: two events only, A and B
- The outcome of the first event may or may not affect the outcome of the second one
- Thus, the two events can either be dependent or independent from each other
- Under independence, three types of probabilities are of interest:
    1. Marginal
    2. Joint
    3. Conditional

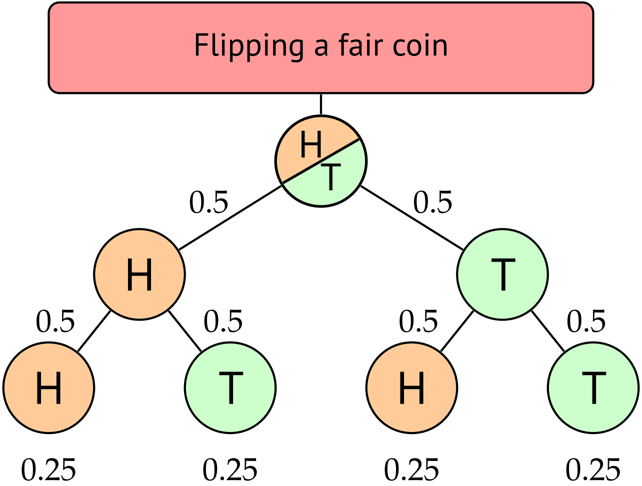
Probability under Statistical Independence (2)

- Marginal (unconditional) probabilities already introduced
- Joint probability: the probability of two or more events happening simultaneously or sequentially
- Calculated as:

$$P(A \text{ and } B) = P(A \cap B) = P(A)P(B)$$

- Example: What is the probability of getting three sixes in three consecutive rolls?

Probability Trees Under Independence



Conditional Probability Under Independence

- Gives the probability of event B given that event A has already occurred
- Mathematical notation: $P(B|A)$
- Under independence:

$$P(B|A) = P(B)$$

  i.e. the conditional probability equals the marginal (unconditional) one
- Example: What is the probability that you get a six in the second dice roll if in the first roll you got five?

Conditional Probability Under Dependence

- Again, three types of probabilities: conditional, joint, and marginal
- We'll consider them in turns
- Start with conditional probability

Example 1

There are ten balls in a box: 3 coloured (C) and dotted (D), 1 coloured and striped (S), 2 white (W) and dotted, 4 white and striped. You draw a ball and it turns out to be coloured. What is the probability that it is dotted? What is the probability that it is striped?

	Coloured	White
Dotted	3	2
Striped	1	4

Conditional Probability Under Dependence (2)

- We proceed as follows:

$$\begin{aligned} P(D|C) &= \frac{3}{4}, \quad P(S|C) = \frac{1}{4}, \quad P(D|C) + P(S|C) = 1 \\ P(D|W) &= \frac{1}{3}, \quad P(S|W) = \frac{2}{3}, \quad P(D|W) + P(S|W) = 1 \end{aligned}$$

- Note that 'coloured' or 'white' determines the probability of getting 'dotted' or 'striped', so we have dependence
- Now, the conditional probabilities can also be written as:

$$P(D|C) = \frac{P(D \cap C)}{P(C)}, \quad P(S|C) = \frac{P(S \cap C)}{P(C)}, \quad \text{etc.}$$

- To sum up, if we have two mutually dependent events A and B:

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

Joint Probability Under Dependence

- From the conditional probability formula, it follows directly that:

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

- Note it is not the same as in the independence case
- But since $P(B \cap A) = P(A \cap B)$, this is also equal to:

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

- How would you calculate $P(CD)$ in the white and coloured balls example?

Marginal Probability Under Dependence

- Marginal (unconditional) probabilities are calculated as the sum of all joint events in which the simple event occurs
- Take again the example with white and coloured balls
- The probability of a coloured ball is

$$P(C) = P(CD) + P(CS)$$

- The probability of a white ball is

$$P(W) = P(WD) + P(WS)$$

- $P(D)$ and $P(S)$ are calculated analogically

Probability Distributions: An Introduction

What Is A Probability Distribution?

Probability Distribution

A probability distribution is a mathematical function that associates the possible outcomes with their respective probabilities, i.e. it specifies the (theoretical) frequencies of outcomes.

An alternative definition would be:

Probability Distribution

A probability distribution is a mathematical function that describes fully the variation of outcomes a random variable (vector).

- As probability distributions provide expected frequencies of outcomes, they are very useful in making inference when there is uncertainty
- A clear distinction should be made between theoretical and empirical distributions!

Types of Probability Distributions

- Two major types: discrete and continuous
- In discrete distributions the number of possible outcomes is **countable**
- Examples: cards, dice, coins, etc.
- In continuous distributions the number of possible outcomes is **infinite** within a given range
- Examples: height, age, temperature, etc.

Random Variables

Random Variable

A variable that takes on different values as a result of a random experiment/phenomenon.

- Also classified as discrete and continuous
- This is because each random variable is associated with a probability distribution
- This distribution could either be known or unknown
- Each random variable is characterized with:
  - A probability mass function ($p_X(x)$) if it is discrete, or
  - A probability density function ($f_X(x)$) if it is continuous

Expected Value of A Random Variable

- If it exists, equals the sum of 'values times probabilities'

$$\mu = \mathsf{E}(X) = \sum_{i=1}^{n} p(x_i)x_i$$

- For continuous variables, sums are replaced with integrals

$$\mu = \mathsf{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

- The mean describes the centre of variation (central tendency) of a random variable

Variance and Standard Deviation

- If it exists, equals the average of the squared (why?) variation (distance) from the mean value
- For discrete variables:

$$\sigma^2 = \text{Var}(X) = \sum_{i=1}^{n} p(x_i)(x_i - \mu)^2$$

- For continuous variables, sums are replaced with integrals

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x_i - \mu)^2 f(x)\, dx$$

- Standard deviation: square root (why?) of variance:

$$\sqrt{\sigma^2} = \sigma$$

Quantiles

- Cut-off points denoting the values of a random variable below which a certain percentage of all values lie
- Also known as percentiles
- Quartiles: divide the distribution in four parts, each part containing a quarter of all values
- Median: equals the second quartile; divides the distribution in two equal halves

Bernoulli Distribution

- A discrete distribution
- Results from the experiment known as the Bernoulli trial
- Two outcomes are possible: success (1) or failure (0)
- Each outcome is assigned a probability, respectively p for success, and $q = 1 - p$ for failure
- Example: tossing a coin
- Probability mass function:

$$f(x) = p^x (1 - p)^{1-x}, \quad x = 0, 1$$

Binomial Distribution

- A sequence of n independent Bernoulli trials
- Naturally, it is also discrete
- Used to model the number of successes in sampling with replacement
- pmf:

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

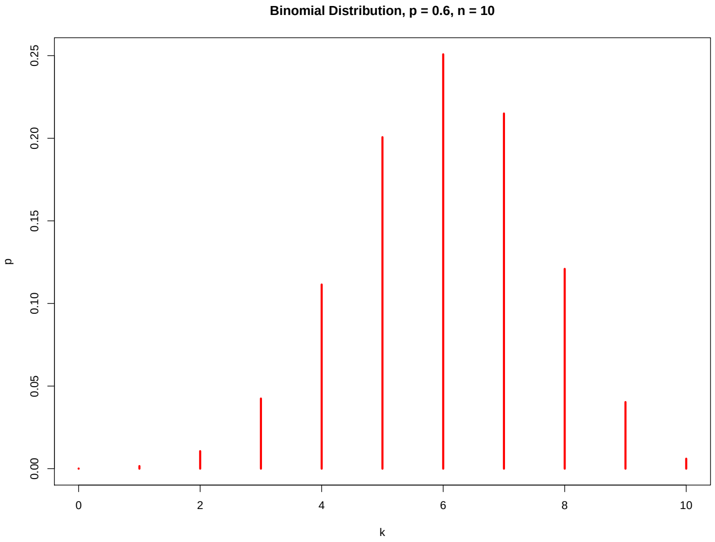
- Mean:

$$\mu = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} = np$$

- Variance:

$$\sigma^2 = np(1-p)$$

Binomial Distribution (2)



Poisson Distribution

- Discrete, named after Siméon Denis Poisson
- Counts the number of times an event happens during a specified time interval
- Examples: number of customers per hour; number of phone calls per minute; number of patients arriving per day in a hospital, number of cars passing through a tollbooth per hour, etc.
- pmf:

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

  where λ is the mean count per period
- Mean and variance:

$$\mu = \sigma^2 = \lambda$$

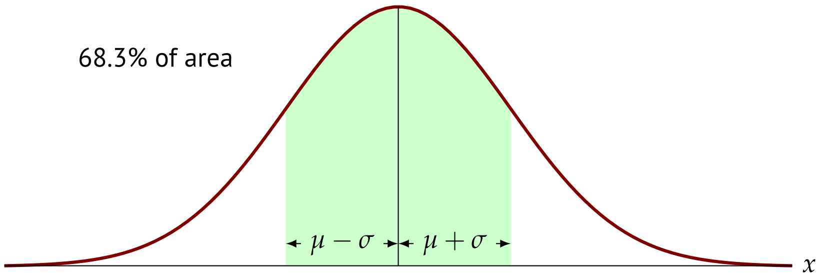
Normal Distribution

- Also known as the Gaussian distribution
- A continuous one, one of the most important and widely used
- Why? Because it successfully describes many phenomena (human weight, height, IQ, production yields, etc.)
- pdf:

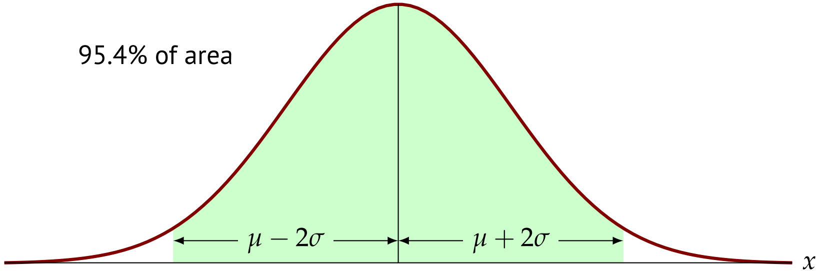
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x-\mu}{2\sigma^2})$$

- The mean and the variance are respectively μ and σ^2
- Standard normal distribution: $\mu = 0$, $\sigma^2 = 1$

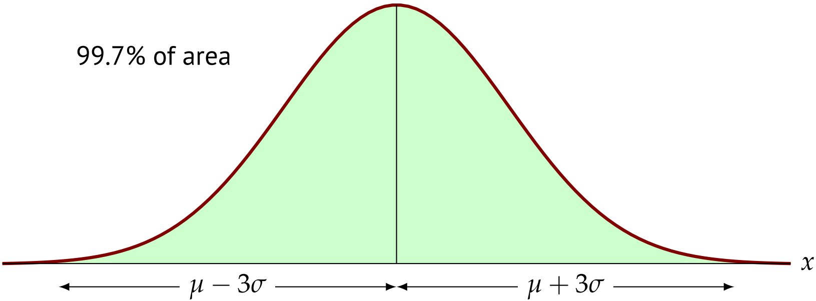
Normal Distribution (2)



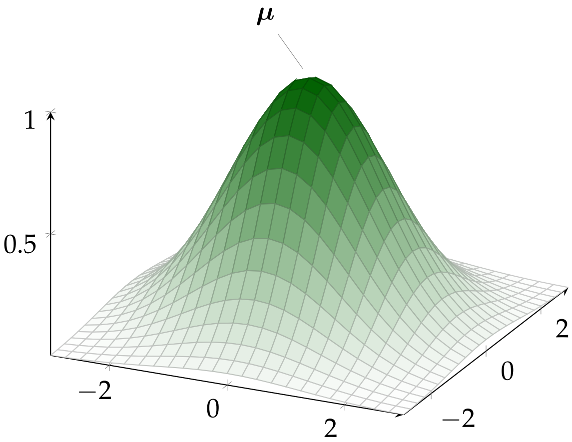
Normal Distribution (3)



Normal Distribution (4)

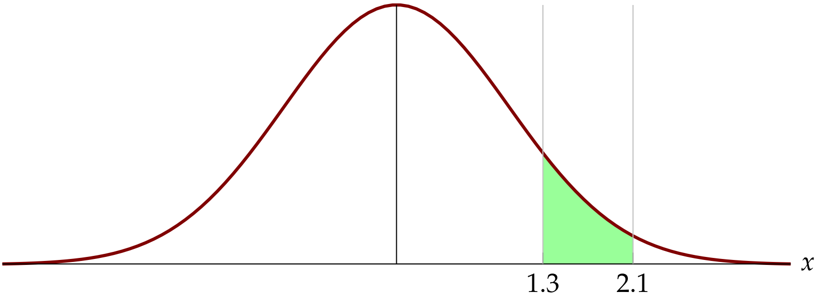


Normal Distribution (5)



Normal Distribution (6)

- Take the standard normal distribution
- How to find the probability that X is between 1.3 and 2.1?



- How do we solve this? Integrate? Tables? R?

Normal Distribution (7)

- As already seen clearly, it is a symmetrical distribution
- It is also unimodal
- The mean, the median, and the mode are equal among each other
- The tails of the curve extend to $\pm\infty$ and never cross the horizontal axis
- If you want to standardize a normal random variable:

$$z = \frac{x - \mu}{\sigma}$$