STA 237 NOTES

Author
Hao Hua He

Contents

Chapter 1: Week 1

1 Lec 1: Outcomes, Events. and Probability

1.1 Introduction

Definitions

1. **Probability**
   (a) numeric value of certainty/uncertainty
2. **(Random) Experiment**
   (a) mechanism/phenomenon that results in random or unpredictable outcomes
3. **Sample Space**
   (a) Set of all outcomes from an experiment
   (b) denoted Ω
4. **Event**
   (a) Subset of Sample Space
   (b) Relations between events
      i. INtersect
      ii. Union
      iii. Complement

Example 1.1. Neither A not B is denoted $(A \cup B)^c \Rightarrow A^c \cap B^c$

Theorem 1.2. De Morgan's Law sates for any events A and B

1. $(A \cup B)^c = A^c \cap B^c$
2. $(A \cap B)^c = A^c \cup B^c$

Example 1.3. Exactly one of A and B is denoted as

$$A \cup B \cap (A \cup B)^c = A \cup B \cap (A^c \cup B^c)$$

More Definitions:

1. **Disjoint(mutually exclusive)**
   (a) $A \cap B = \ = \varnothing$
2. A implies B
   (a) $A \subset B$
   (b) $A \cap B = A$

1.2 Probability Function

1.2.1 Definition

Definition 1.4. Probability func P defined on a finite sample space Ω assigns each event $A \in \Omega$ a number $P(A)$ s.t.

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. $P(A \cup B) = P(A) + P(B)$
   (a) if A and B disjoint.

where $P(A)$ is the probability that event A occurs.

1.2.2 Calculating by Counting

Calculating by counting only applies when

1. All outcomes of Ω are equally likely
   (a) Ω is finite

Then,

$$P(A) = \frac{\text{number of outcomes belonging to } A}{\text{Total number of outcomes in } \Omega}$$

1.2.3 Product of sample space

In general, given sample spaces Ω_1 and Ω_2, we have

$$\Omega = \omega_1 \times \omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2)\}$$

1.2.4 Permutation and Combination

Permutations and combinations are ways of grouping elements of a set into a subset, where permutations are ordered and combinations are unordered.

Theorem 1.5. The number of possible permutations of size n from N objects

$$_N P_n = \frac{N!}{(N-n)!}$$

Theorem 1.6. The number of possible combinations of size n from N objects

$$\binom{N}{n} = \frac{N!}{(N-n)!}$$

2 Lec 2: Conditional Probability and Independence

2.1 Conditional Probability

Conditional probability of an event A given event C is, for any C with $P(C) > 0$,

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

or alternatively, with the multiplication rule

$$P(A \cap C) = P(A|C) \cdot P(C)$$

2.1.1 Law of Total Probability

Suppose events $C_1, \ldots, C_2$ are disjoint s.t. $C \cup \cdots \cup C_m = \Omega$, for any arbitrary A,

$$P(A) = \sum_{i=1}^{m} [P(A|C_i)P(C_i)]$$

Thus,

$$\begin{aligned} P(C_i|A) &= \frac{P(C_i) \cap A}{P(A)} \\ &= \frac{P(A|C_i)P(C_i)}{\sum_{i=1}^{m}[P(A|C_i)P(C_i)]} \end{aligned}$$

2.1.2 Baye's Rule

Suppose $C_1, \ldots, C_m$ are disjoing s.t. $C \cup \cdots \cup C_m = \Omega$, the conditional probability of C_i given any arbitrary A is

$$P(C_i|A) = \frac{P(A|C_i) \cdot P(C_i)}{\sum_{i=1}^{m}[P(A|C_i)P(C_i)]}$$

2.2 Independence

2.2.1 Definition

Definition 2.1. Events A and B are independent if

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

That is to say that B occuring or not does not change the probability of A

2.2.2 Implications

1. Suppose A and B with $P(A|B) = P(A)$
   (a) Complements
      i. $P(A^c|B) = P(A^c)$
   (b) Mulplication Rule
      i. $P(A \cap B) = P(A)P(B)$
   (c) Mutal Property
      i. $P(B|A) = P(B)$

To show that A and B are independent, one of the following must be proved, otherwise A and B are dependent.

$$\begin{aligned} P(A^c|B) = P(A^c) \iff \\ P(A \cap B) = P(A)P(B) \iff \\ P(A|B) = P(B) \end{aligned}$$

Chapter 2: Week 2

3 Lec 3: Discrete Random Variables

3.1 Discrete Random Variables

Let Ω be a sample space. A **random variable** X is a function that maps Ω on to a real number ($\mathbb{R}$)

Definition 3.1. When a *random variable* X takes a countable number of values, it is called a **discrete random variable**

3.2 Probability mass function

Definition 3.2. The **probability mass function** p of a discrete random variable X is the finction

$$p : \mathbb{R} \to [0, 1]$$

defined by

$$p(k) = P(X = k)$$
$$\text{for } -\infty < k < \infty$$

The probability mass function **uniquely** defines the behaviour of a random variable.

3.3 Cumulative distribution function

Definition 3.3. The **cumulative distrubution function**, or **distribution function** F of a random variable X is the function

$$F : \mathbb{R} \to [0, 1]$$

defined by

$$F(A) = P(X \leq a)$$
$$\text{for } -\infty < a < \infty$$

3.4 Bernoulli distribution

A disc

4 Lec 4: Continuous Random Variables

Definition 4.1. A random variable X is **continuous** if for some function

$$f : \mathbb{R} \to \mathbb{R}$$

and any numbers a and b with $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

. Note: the function has to satisfy

1. $f(x) \geq 0$ for all x and
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

We call this function f the **probability density fuction** of X and the value(s) of $f(x)$ is the **probability density** of X at x

1. $f(x)$ is **not** probability
2. Both a pmf and pdf uniquely defines a random variable, but a pmf maps to $[0, 1]$ and a pdf to $[0, 1)$
3. $f(x)$ can be interpreted as a relative measure of likelihood around X

Definition 4.2. The **cumulative distrubution function of** F of a random variable X is the function

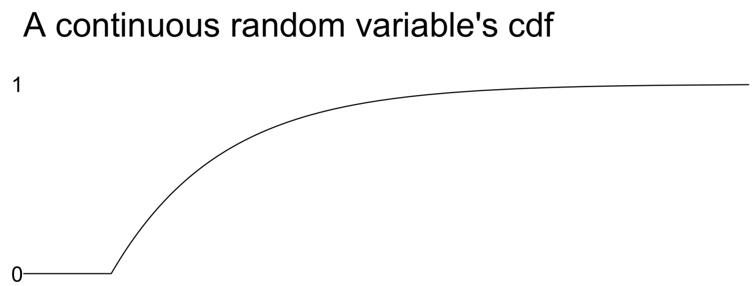
$$F : \mathbb{R} \to [0, 1]$$

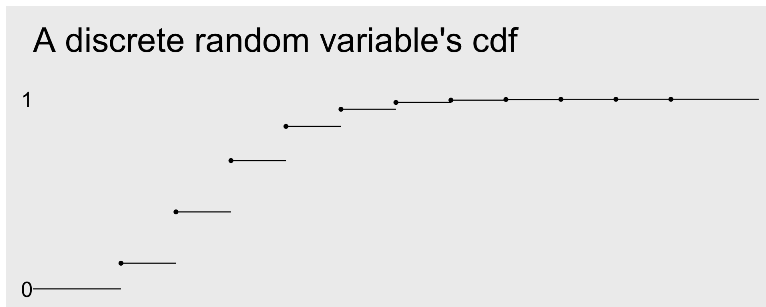
defined by

$$F(x) = P(X \leq a)$$
$$\textbf{for } -\infty < a < \infty$$

Note: the definition of *cdf* is the same for both discrete and continuous random variables

4.1 Property of cumulative distribution functions





Any cdfs are

1. Non-decreasing
2. Right continuous and
3. (Approaching) 0 on the left end and to 1 on the right end

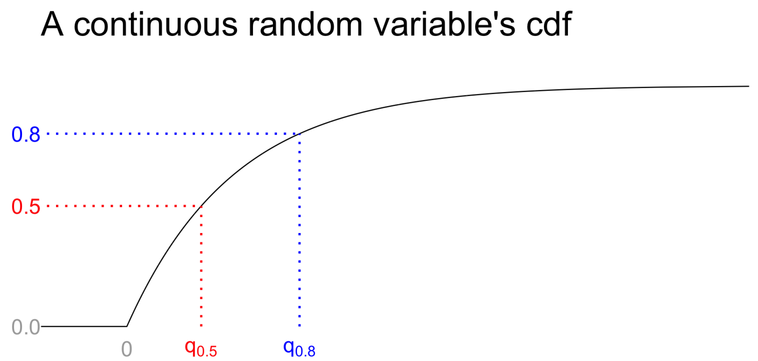
4.2 Quantile, Percentile, and Median

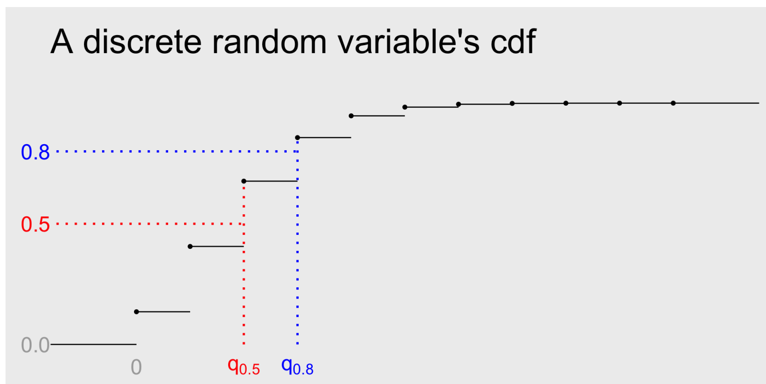
Definition 4.3. Let X be a random variable with cumulative distribution function F. Then the **quantile funtion** of X is the function F^{-1} defined by

$$F^{-1}(t) = \min(x : F(x) \geq t)$$

for $0 < t < 1$

4.2.1 Quantile function for continuous vs discrete





4.3 Common Continuous Distributions

4.3.1 Uniform distribution

Definition 4.4. A continuous random variable has a **unitiorm distribution** on interval $[\alpha, \beta]$ if its probability density funtion f is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \leq x \leq \beta \\ 0 & otherwise \end{cases} \tag{1}$$

4.3.2 Exponential Distribution

Definition 4.5. A continuous random variable has an **exponential distribution** with parameter $\lambda, \lambda > 0$ if its probability density function f is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases} \tag{2}$$

Exponential random variables are often used to model time until the next event in a Poisson process. λ is the expected rate of events.

Memoryless Property of Exponential Random Variables

4.3.3 Gamma Distribution

Definition 4.6. A continuous random variable has a **gamma distribution** with parameyers α and β with $\alpha > 0$ and $\beta > 0$ it its propability density f is given by

$$f(x) = \frac{1}{\Gamma(\alpha)}\beta^{\alpha}x^{\alpha-1}e^{-\lambda x} \quad \text{for } x > 0 \tag{3}$$

We denote this distrubution Gamma(α, β)

4.3.4 Normal Distribution

Definition 4.7. A continuous random variable has a **normal distribution** with parameter μ and σ^2 with $\sigma^2 > 0$ it its probability density function f is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\} \tag{4}$$

Normal distribution, or Gaussian distribution, ins central in probability theory and statistics.

It is often used to model observational errors.

Normal distributions have a symettric shape around its centre.

μ controls the center of the distribution (location) while the σ controls the spread of the distribution (shape)

Definition 4.8. A normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is called the **standard normal distribution**.

We often denote a srandard normal random variable by Z, Z $N(0,1)$ is its pdf with ϕ, and its cdf with Φ

Chapter 3: Week3

5 Lec 5: Expectation and Variance

5.1 Expectation

5.1.1 Expectration of DRV

Definition 5.1. The expectation of DRV X taking values $x_1, \dots, x_2$ with probability mass function p is given by

$$E[X] = \sum_{i\in\{1,2,\dots\}} x_i p(x_i)$$

5.1.2 Expectation of CRV

Definition 5.2. The **Expectation** of a CRV X with pdf f is given by

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

5.2 Variance

Definition 5.3. The **Variance** of a random variable X is the number defined by

$$\mathrm{Var}(X)x = E[(X - E[X])^2]$$

Chapter 4: Summation of Definitions and theorems

6 Definitions

1. **Event**
   (a) Subset of Sample Space
   (b) Relation between events
      i. Intersect
         A. denoted $A \cap B$
      ii. Union
         A. denoted $A \cup B$
      iii. Complement
         A. denoted A^c
2. **Event**
   (a) Subset of sample space
3. **(Random) Experiment**
   (a) Mechanism/Phenomenon that results in random or unpredictable outcomes
4. **Probability**
   (a) Numeric Value of certainty/uncertainty
5. **Sample Space**
   (a) Set of all possible outcomes (from experiment)
   (b) Denoted Ω

7 Theorems