Probability and statistics 1 - Concise Notes

MATH40001

Term 1 Content

Louis Gibson

Colour Code - **Definitions** are **green** in these notes, **Consequences** are **red** and **Causes** are **blue**

Mathematics
Imperial College London
United Kingdom
January 31, 2022

Contents

1 Sample sapces and iterpretations of probability

1.1 Notation

1.2 The sample Spae Ω

Definition 1.2.1. Sample space

The sample space Ω is defined as the sset of all possible outscomes of an experiment. The elements of Ω are typically denoted by ω and called sample points

1.2.1 Cardinality

Definition 1.2.2. Cardinality

For any set A, we define the cardinality of A as the number of elements in A.

Definition 1.2.3. Two sets have the same cardinality if there is a bijection between the two sets.

1.3 Interpretation of probability

1.3.1 Naive definition of probability

Definition 1.3.1. Naive definition of probability

Suppose that sample space Ω is finite. and consider an event $A\Omega$. Then the naive definition of A is defined as

$$P_{Naive}(A) = \frac{card(A)}{card(\Omega)}$$

In addition this definition assumes all the event outcomes are equally likely.

1.3.2 Limiting frequency

COnsidser n_{total} replication of an experiment and n_A denote the numbe of times event A occurs. Then we could interpret the probabnility of event A occurring as

$$P(A) = \lim n_{total} \to \infty \frac{n_A}{n_{total}}$$

2 Counting

2.1 The multiplication principle

Theorem 2.1. The multiplication principle

Consider two experiments: Experiment A has a possible outcomes, Experiment B has b possible outcomes, Then the compound experiment of performing Experiment A and B (in any order) has ab possible outcomes

2.2 Power Sets

A power set of a set A denoted $P(A)$ is defined as the set of all possible subsets of A including $\emptyset and \Omega$

Theorem 2.2. $card(\mathcal{P}(A)) = 2^{card(\Omega)}$

2.3 Sampliing with and without replacement

2.3.1 Sampling with replacement - ordered

Theorem 2.3. Sampling with replacement

In the case of sampling k balls with replacement from urn containing n balls as desdcribed above, there are $card(\Omega) = n^k$ possible outcomes whenb the order of the objets matters

2.3.2 Sampling without replacement - ordered

Theorem 2.4. Sampling without replaement

IIn the scase of sampling k balls without replacement from an urn containing n balls as described abovem there are $card(\Omega) = n(n-1)\dots(n-(k-1)) = (n)_k$

2.3.3 The birthday problem

2.3.4 Sampling without replacement - unordered

Definition 2.3.1. Binomial coefficiant

For any $k, n \in \mathbb{N} \cup 0$, the *binomial coefficient* is defined as the number of subsetz of size k for a set of size n.

2.3.5 Sampling with replacement - unordered

As before, let $k, n \in \mathbb{N}$ Let us consider the case that we have an urn with n balls with labels in $1, 2, \dots, n$and we want to choose k balls one after the other with replacement. Assuming the order of the balls does not matter, how many possible outcomes are there?

Theorem 2.5. Sampling with replacement when the order does not matter

In the sampling with replaement problem as described above and assuming that the order of the balls does not matterm we have $card(\Omega) = \binom{n+k-1}{n} = \binom{n+k-1}{n-1} possibilities$

3 Axiomatic definition of probability

3.1 The event spae $\mathcal{F}$

Definition 3.1.1. Algebra and σ algebra

1. (a) $\emptyset \in \mathcal{F}$
   (b) $\mathcal{F}$ is closed under complements $A \in \mathcal{F} \longrightarrow A^c \in \mathcal{F}$ and
   (c) $\mathcal{F}$ is closed under unions of paires of members, $A_1, A_2 \in \mathcal{F} \longrightarrow A_1 \cup A_2 \in \mathcal{F}$
2. an $sigma$-algebra if
   (a) $\emptyset \in \mathcal{F}$
   (b) $\mathcal{F}$ is closed under complements $A \in \mathcal{F} \longrightarrow A^c \in \mathcal{F}$ and
   (c) $\mathcal{F}$ is closed under countable union, $A_1, A_2, \dots \in \mathcal{F} \longrightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$

3.2 Definition of probabilities and basic properties

3.2.1 PRobability measure and probability space

Definition 3.2.1. Probalility measure

A mapping $P : \mathcal{F} \to \mathbb{R}$ is called a probablility measure on $(\Omega, \mathcal{F})$ if it satifsdies the three conditions,

roman* $P(A) \geq 0$ for all events $A \in \mathcal{F}$

roman* $P(\Omega) = 1$

roman* For any countable sequence of disjoint events $(A_i)_{i \in \mathcal{I}}$ with $A_i \in \mathcal{F}$ for all $i \in \mathcal{I}$, we have

$$P(\textstyle\bigcup_{i \in \mathcal{I}} A_i) = \sum_{i \in \mathcal{I}} P(A_i)$$

Definition 3.2.2. Probability Space

We define a probability space as the triplet $(\Omega, \mathcal{F}, P)$ where Ω is a $set (the sample sapce), \mathcal{F}$ is a σ-algebra consistingof subsets of Ω and P is a probability measure on $(\Omega, \mathcal{F})$

3.2.2 Basic properties of the probability measure

Theorem 3.1.

Consider a probability space $(\Omega, \mathcal{F}, P)$. Then m for any events $A, B \in \mathcal{F}$, we have

1. $P(A^c) = 1 - P(A)$
2. AB then $P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

4 Conditional Probablilites

4.1 Definition

Definition 4.1.1. Conditional probability

Consider a probablility space $(\Omega, \mathcal{F}, P)$.Consider $A, B \in \mathcal{F}$ with $P(B) > 0$. Then the conditional probability of A given B denoted by $P(A|B)$ is defined as

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

Conditional probability

Let $B \in \mathcal{F}$ with $P(B) > 0$ define $\mathcal{Q} : \mathcal{F} \to \mathbb{R}$ $\mathcal{Q} = P(A|B)$. Then $(\Omega, \mathcal{F}, \mathcal{Q})$ is a probability space.

4.2 Example

4.3 Multipliation rule

Theorem 4.1. Multiplication Rule

Let $n \in \mathbb{N}$ then for any events $A_1, \ldots\ldots A_n$ with $P(A_2 \cap \ldots \cap A_n) > 0$. We have

$$P(A_1 \cap \ldots \cap A_n) = P(A_1|A_2 \cap \ldots \cap A_n)P(A_2|A_3 \cap \ldots \cap A_n)(A_{n-2}|A_{n-1} \cap A_n)P(A_{n-1}|A_n)P(A_n)$$

4.4 Baye's rule and law of total probability

4.4.1 Bayes' rule

Theorem 4.2.

Let $A, B \in \mathcal{F}$ with $P(A) > 0$, $P(B) > 0$. Then

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

4.4.2 Law of total probability

Theorem 4.3. Partition

A partition of the sample space Ω is a collection $B_i : i \in \mathcal{I}$ of disjoint events such that their countable union equals Ω

Theorem 4.4. Law of total probability

Let $B_i : i \in \mathcal{I}$ denote a partition of Ω with $P(B_i) > 0$ for all $i \in \mathcal{I}$. Then for all $A \in \mathcal{F}$

$$P(A) = \sum_{i \in \mathcal{I}} P(A \cap B_i) = \sum_{i \in \mathcal{I}} P(A|B_i)P(B)$$

4.4.3 General's Bayes' Rule

Theorem 4.5.

Consider a partition $B_i : i \in \mathcal{I}$ of Ω with $P(B_i) > 0$ for all $i \in \mathcal{I}$, then for any event $A \in \mathcal{F}$ with $P(A) > 0$, we have

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{k \in \mathcal{I}} P(A|B_k)P(B_k)}$$

4.4.4 Baye's rule and law of total probability with additional conditioning

Theorem 4.6. Baye's rule with extra conditioninng

For events A, B, E with $P(A \cap E), P((B \cap E) > 0$ we have

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

Theorem 4.7. Law of total probability with additional conditioning

Consider events A, E with $P(E) > 0$ and let $B_i : i \in \mathcal{I}$ denote a partition of Ω with $P(B_i) > 0$ for all $i \in \mathcal{I}$. The

$$P(A|E) = \sum_{i \in \mathcal{I}} \frac{P(A \cap B_i \cap E}{P(E)} = \sum_{i \in \mathcal{I}} P(A|B_i \cap E)P(B_i|E)$$

4.5 Examples

5 Independence

5.1 Independene of events

Definition 5.1.1. Independent events

The event's A, B are independent if

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

and dependent otherwise

Theorem 5.1.

If the events A and B are independent then the same is true for eah of the pairs A^c and B, A and B^c, and A^c and B^c

Definition 5.1.2. Independence of events(general case)

1. A finite ollection of events $A_1, A_2, \ldots, A_n)$ is defined to be independent if

   $$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k})$$

   for every subcollection $\{i_1, \ldots, i_k\}$ of $\{1, 2, \ldots, n\}, k = 1, 2, \ldots, n$

2. A countable or or uncountably infinite collection of evenets is defined to be independent if each subcollection is independent

5.1.1 Conditional independence of events

Definition 5.1.3. Conditional independence of events

Consider evenets $A, B, C \in \mathcal{F}$ with $P(C) > 0$. Then we say that A and B are conditionally independent given C if $P(A|C) = P(A|C)P(B|C)$ If we in addition assume that $P(B \cap C) > 0$ then the above equation is equivalent ato the condition

$$P(A|B \cup C) = P(A|C)$$

5.1.2 Continuity of the probability measure and product rule

Definition 5.1.4.

The set difference between two sets $A, B\Omega$ denoted by $A\ B$ is defined as $A\ B = \omega \in \Omega : \omega \in A and \omega \notin BA \cap B^c$ **Lemma 6.1.11** Any countable union can be written as a countable union of disjoint sets. Let $A_1, A_2, \cdots \in \mathcal{F}$ and define $D_1 = A_1, D_2 = A_2 \sim A_1, D_3 = A_3 \sim (A_1 \cup A_2), \cdots$ then D_i us a collection of disjoint sets and their countable unions are equal.

Definition 5.1.5. Increasing and decreasing sets

A sequence of sets $(A_i)_{i=1}^{\infty}$ is said to increase to A if $A_i \uparrow A, A_1 A_2 A_3 \ldots$ and their infinite union equal A Same for decrease

Theorem 5.2.

If $A_1, A_2, \dots \in \mathcal{F}$ and $A_i \uparrow A$ or $A \downarrow A$ then

$$\lim_{i\to\infty} P(A_i) = P(A)$$

The above theorem states that for increasing or decreasing sets, we can intervchange the limit operation and the probability measure, i.e. we have

$$lim_{i\to\infty} P(A_i) = P(\lim_{i\to\infty} A_i)$$

where the set limit on the right hand side needs to be understood as taking an infinite union or intersectgion for increasiong and decreasing sequencs repectively.

Theorem 5.3.

If $A_1, A_2, \dots$ is a countable infinite set of independent events, then

$$P(\bigcap_{i=1}^{\infty} A_i) = \prod_{i=1}^{\infty} P(A_i)$$

6 Discrete random variables

6.1 Pre-images and their properties

Definition 6.1.1.

Consider a function with domain $\mathcal{X}$ and co-domain $\mathcal{Y}$ i.e. $f : \mathcal{X} \to \mathcal{Y}$

- For any subset $A\mathcal{X}$, we define the image of A under f as

$$f(A) = y \in \mathcal{Y} : x \in A : f(x) = y = f(x) : x \in A$$

  If $A = \mathcal{X}$ then we all $f(\mathcal{X} = Imf$ the image of f

- For any subset $B\mathcal{Y}$ we define the pre-image of B under f as

$$f^{-1}(B) = x \in \mathcal{X} : f(x) \in B$$

The definition of pre-image implies that

$$x \in f^{-1}(B) \iff f(x) \in B$$

Lemma 7.1.2 For any ollection of subsets $B_i\mathcal{Y}, i \in \mathcal{I}$ where I debnotes an arbitrary index setm we have thatm

$$f^{-1}(\bigcup_{i\in\mathcal{I}} B_i) = \bigcup_{i\in\mathcal{I}} f^{-1}(B_i)$$

6.2 Random variables

s a function from the sample space to the real number $\mathbb{R}$ i.e.

$$X : \Omega \to \mathbb{R}$$

The function needs to satisfy some properties which we introduce in the formal definition below. Note that

- Despite the name, a random variable is a function and not a variable
- We typically use capital letters such that X, Y, Z to denote random variables
- The value of the random variable X at the sample point ω is given by $X(\omega)$ and is called a realisation of X
- The randomness stems from $\omega \in \omega$(we dont know which outcome ω appears in the random experiment) the mapping itslef given by X is deterministic

6.3 Discrete random variables and probability distributions

Definition 6.3.1. Discrete random variables

A discrete random variable on the probability space $\Omega, \mathcal{F}, P$ is defined as a mapping $X : \Omega \to \mathbb{R}$ such that

- the image/range of Ω under X denoted by $ImX = X(\omega) : \omega \in \Omega$ is a countable subset of $\mathbb{R}$
- $X^{-1}(x) = \omega \in \Omega : X(\omega) = x \in \mathcal{F}$ for all $x \in \mathbb{R}$

Definition 6.3.2. Probability mass function

Tje probability mass function of the discrete random variable X is defined as the function $p_X : \mathbb{R} \to [0,1]$ given by

$$p_X(X) = P(\omega \in \Omega : X(\omega) = x) = P(X^{-1}(x))$$

Note that this definition implies

$$p_X(x) = 0 \text{ if } x \notin ImX$$

Note that for $x_1, x_2 \in ImX$ with $x_1 \neq x_2$ then

$$X^{-1}(x_1) \cap X^{-1}(x_2) = \omega \in \Omega : X(\omega) = x_1 \text{ and } X(\omega) = x_2 = \emptyset$$

Using Axiom iii we can derive

$$\textstyle\sum_{x\in\mathbb{R}} p_X(x) = 1$$

Theorem 6.1.

Let $\mathcal{I}$ denote a countable index set. SUppose that $S = s_i : i \in \mathcal{I}$is a countable set of distinctreal numbers and $\pi_i : i \in \mathcal{I}$ is a collection of numbers satisfying

$$\pi_i \geq 0 \text{ for all } i \in \mathcal{I} \text{ and } \textstyle\sum_{i\in\mathcal{I}} \pi_i = 1$$

then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a discrete random variable X on that probability space such thatits probabilit mass function is given by

$$p_X(s_i) = \pi_i \; foralli \in \mathcal{I}$$
$$p_X(s) = 0, \; s \notin S$$

6.4 Comm discrete distribution

6.4.1 Bernoulli distribution

Definition 6.4.1. Bernoulli distribution

A discrete random variable X is saif to have Bernoulli distribution with parameter $p \in (0,1)$ if X can only take two possible values 0 and 1. i.e. $ImX = 0, 1$
$p_X(1) = P(X = 1) = p$ $p_X(0) = 1 - p$ We write $X\tilde{B}ern(p)$

Definition 6.4.2. Indicator variable

Consider an event $A \in \mathcal{F}$ we denote by

$$\mathbb{I}_A(\omega) = 1 \text{ if } \omega \in A \text{ and } 0 \text{ if not}$$

the indicator variable of the event A

6.4.2 Binomial distribution

A discrete random variable X is said to follow the binomial distribution with parameters $n \in \mathbb{N}$and $p \in (0,1)$ if $ImX = 0, 1, 2 \ldots, n$ and

$$P(X = x) = \tbinom{n}{x} p^x (1-p)^{n-x}$$

We write $X\tilde{B}in(n, p)$

6.4.3 Hypergeomtric distribution

Definition 6.4.3. Hypergeometric distribution

A discrete random variable X is said to follow the hypergeometric distribution with three parameter $N \in \mathbb{N} \cup 0, K, n \in 0, 1, 2, \dots N$ if $ImX = 0, 1, 2, \dots, min(n, K)$and

$$P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \text{ for } x \in 0, 1, 2, \dots K \text{ and } n - x \in 0, 1, \dots N - K$$

and $P(X = x) = 0$ otherwise. We write $X\tilde{H}Geom(N, K, n$
Lemma 7.4.6 Vandermonde's identity For $k, m, n \in \mathbb{N}, k + m$

$$\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}$$

6.4.4 Discrete uniform distribution

Definition 6.4.4. Discrete uniform distribution

Let C denote a finite nonempty set of numbers. We say that a discrete random variable X follows the discrete uniform distribution on C, i.e. $X\tilde{D}Unif(C)$ if $ImX = C$ and

$$P(X = x) = \frac{1}{card(C)}$$

for $x \in C$ and $P(X = x) = 0$

6.4.5 Poisson distribution

Definition 6.4.5. Poission distribution

A discrete random variable X is said to followpw the Poisson distribution with parameter > 0 i.e. $X\tilde{P}oi()$ if $ImX = 0, 1, 2, \dots = \mathbb{N} \cup 0$ and

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, for x = 01, 2, 3, \dots$$

6.4.6 Geometric distribution

Definition 6.4.6. Geometric distribution

a discrete random variable X is said to follow the geomtric distribution with parameter $p \in (0, 1)$ i.e. $X\tilde{G}eom(p)$ if $ImX = \mathbb{N}$ and

$$P(X = x) = (1 - p)^{x-1} p \text{for } x = 1, 2, \dots$$

6.4.7 Negative Binomial distribution

Definition 6.4.7. NEgative binomial distribution

A discrete random variable X is said to follow thenegative binomial distrivution with parameters $r \in \mathbb{N}$ and $p \in (0, 1)$ written $X\tilde{N}Bin(r, p)$

$$P(X = x) = \binom{x+r-1}{r-1} p^r (1 - p)^x \text{ for } for x = 0, 1, 2, \dots$$

Definition 6.4.8.

For $\alpha \in \mathbb{C}, k \in \mathbb{N}$ we define

$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!}$$

The generalised binomial formula is then giben by

$$(1 + x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \text{ for } |X| < 1$$

Lemma 7.14.15 For $x \in \mathbb{N} \cup 0, r \in \mathbb{N}$

$$\binom{x+r-1}{r-1} = (-1)^x \binom{-r}{x}$$

7 Continuous random variables

7.1 Random variables and their distributions

Definition 7.1.1. Random variable

Arandomvariable on the probability space $(\Omega, \mathcal{F}, P)$ is defined as the mapping $X : \Omega\mathbb{R}$ which satisfies:

$$X^{-1}((-\infty, x]) = \omega \in \Omega : X(\omega) \leq x \in \mathcal{F} \text{ for all } x \in \mathbb{R}$$

Definition 7.1.2. Cumulative distribution function

Suppose that X is a random variable on $(\Omega, \mathcal{F}, P)$, then the cumulative distribution function of X is defined as the mapping $F_X : \mathbb{R} \to [0, 1]$ given by

$$F_X(x) = P(\omega \in \Omega : X(\omega) \leq x) = P(X^{-1}((-\infty, x])),$$

which is typically abbreviated to $F_X(x) = P(X \leq x)$

Theorem 7.1. Properties of cdf

- F_X is monotonically decreasing
- F_X is right-contiunous
- $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$

Theorem 7.2.

For $a < b4$ we have $P(a < X \leq b) = F_X(b) - F_X(a)$

7.2 Continuous random variables and probability density funcntion

Definition 7.2.1. Continuious random variables and probability density funtion

A random variable X is alled continuous if its cdf can be written as

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)du \text{ for all } x \in \mathbb{R}$$

where the function $f_X : \mathbb{R} \to \mathbb{R}$ satisfies

1. $f_X(u) \geq 0$ for all $u \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f_X(u)du = 1$

Theorem 7.3.

for a continuous random variable X with density f_X we have

$$P(X = x) = 0 \text{ for all } x \in \mathbb{R}$$

and

$$P(a \leq X \leq b) = \int_{a}^{b} f_X(u)du \text{ for all } a, b \in \mathbb{R} with a \leq b$$

7.3 Common continuous distributions

7.3.1 Uniform

Definition 7.3.1. Uniform distribution

A continuous random variable X is said to have the uniform distribution on the interval (a, b) for $a < b$ i.e. $X\tilde{U}(a, b)$ if its density function is given by

$$f_X(x) = \frac{1}{b-a} \text{ if } a < x < b \text{ and } = 0 \text{ otherwise}$$

Its cumulative distribution function is given by

$$F_X(x) = \frac{x-a}{b-a} \text{ if } a < x < b = 0 \text{ if } x \leq a \text{ and } = 1 \text{ if } x \geq b$$

7.3.2 Exponential

Definition 7.3.2. Exponential distribution

A continuous random variable X is said to have the expoenential distribution with parameter $\lambda > 0$ i.e. $X\tilde{E}xp(\lambda)$ if its density function is given by

$$f_X(x) = \lambda e^{-\lambda x} \text{ if } x > 0 \text{ and } = 0 \text{ otherwise}$$

Its cumulative distribution function is given by

$$F_X(x) = 0 \text{ if } x \leq 0 \text{ and } = 1 - e^{-\lambda x} \text{ if } x > 0$$

7.3.3 Gamma distribution

Definition 7.3.3. Gamma distribution

A continuous random variable X is said to have the Gamma distribution with shape parameter > 0 and rate parameter $\beta > 0$ i.e. $X\tilde{G}amma(,\beta)$ if its density function is given by

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \text{ only for strictly positive } x$$

Its cumulative distribution function is not available in close form

7.3.4 Chi-squared distribution

Definition 7.3.4. Chi-squared distribution

A continuous random variable X is said to have the Chi-squared distribution with $n \in \mathbb{N}$ degrees of freedom i.e. $X\tilde{\chi}^2(n)$, if its density function is given by

$$f_X(x) = \frac{1}{2\Gamma(n/2)} (\frac{x}{2})^{n/2-1} e^{-x/2} \text{ if } x \text{ is strictly positive}$$

Its cumulative distribution function is not abailable in closed form Note that $\chi^2(n)$ distribution is the same as the $Gamma(n/2.1/2)$ distribution

7.3.5 F-distribution

Definition 7.3.5. F-distribution

A continuous random variable X is said to have the $F-distribution$ with $d_1, d_2 > 0$ of freedom i.e. $X - \tilde{(}d_1, d_2)$ if its density function is given by

bruh i aint typing this shit out. google it. theres no way you need to know this

we note that the positive parameters d_1, d_2 are not restricted to be integers. Note that if we have independent random varriavles $X_1\tilde{\chi}^2_n$ and $X_2 - tilde\chi^2_m$ then the random variable

$$X = \frac{X_1/n}{X_2/m}\tilde{F}_{n,m}$$

7.4 Beta distribution

For $\alpha, \beta > 0$ denoted by

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

the so called beta function

Definition 7.4.1. A continuous random variable X is said to have the Beta distribution with parameters $\alpha, \beta > 0$ i.e. $X\tilde{B}eta(\alpha, \beta)$ if its density function is given by

$$f_X(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ if } x \text{ is between 1 and 0 inclusive}$$

7.4.1 Normal distribution

Definition 7.4.2. Standard normal distribution

A random variable X has the standard Gaussian distribution if it has the density function $f(x) = \phi(x)$ with

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ for}$$

Note that we typically write $X\tilde{N}(0,1)$ since a standard normal random variable has mean zero and variance one. The cdf is the ndenoted by $F(x) = \Phi(x)$ with

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \text{ for } x \in \mathbb{R}$$

Definition 7.4.3. Normal distribution

Let μ denote a real number and let $\sigma > 0$. A random variable X has the normal distribution with mean μ and variance σ^2 if it has density function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ for $x \in \mathbb{R}$ Note we typically write $X\tilde{N}(\mu, \sigma^2)$

7.4.2 Cauchy distribution

Definition 7.4.4. Cauchy distribution

A continuous random variable X is said to have the Cauchy distribution if its density function is given by

$$f_X(x) = \frac{1}{\pi(1+x^2)} \text{ for } x \in \mathbb{R}$$

Its cumulative distribution function is given by

$$F_X(x) = \frac{1}{\pi} arctan(x) + \frac{1}{2} \text{ for } x \in \mathbb{R}$$

We note that if we have two independent standard normal random variables $X, Y\tilde{N}(0,1)$ then their ration $Z = X/Y$ follows the Cauchy distribution

7.4.3 Student t-distribution

Definition 7.4.5. Student t-distribution

A continuous random variable X is said to have the student t-distribution with $\nu > 0$ degrees if freedin uf uts debsuty function is given by

$$f_X(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1 + \frac{x^2}{\nu})^{-\frac{\nu_1}{2}}$$

Its cumuylatige distribution function is not available in closed form

8 Transformations of random variables

8.1 The discrete case

Theorem 8.1.

Let X be a discrete random variable on $(\Omega, \mathcal{F}, P)$ and let $g : \mathbb{R} \to \mathbb{R}$ denote a deterministic funtion. THen $Y = g(X)$ is a discrete random variable with probability mass function given by

$$p_Y(y) = \sum_{x:g(x)=y} P(X = x)$$

for all $y \in ImY$

8.2 The continuous case

For the continuous recall that $Y = g(X)$ is only a random variable if Y satisfies the following condition

$$\omega \in \Omega : Y(\omega) \leq y \in \mathcal{F} \text{ for all } y \in \mathbb{R}$$

Theorem 8.2.

Suppose that X is a continuous random variable with density f_X and $g : \mathbb{R} \to \mathbb{R}$ is strictly increasing/decreasing and differentiable with inverse function denoted by g^{-1} the $Y = g(X)$ has density

$$f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}[g^{-1}(y)]| \text{ for all } y \in \mathbb{R}$$

9 Expectation of random variables

9.1 Definition of the expectation

Definition 9.1.1. Expectation of discrete random variable

Let X denote a discrete random variable, then the expectation of X is defined as

$$E(X) = \sum_{x \in ImX} xP(X = x)$$

whever the sum on the right hand side converges absolutely, i.e. when we have $\sum_{x \in ImX} |x|P(X = x) < \infty$

Definition 9.1.2. Expectation of a continuous random variable

For a continuous random variable X with density f_X we define the expectation of X as

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$$

provided that $\int_{-\infty}^{\infty} |x|f_X(x)dx < \infty$

9.2 Law of the unconcious statistician(LOTUS)

Theorem 9.1. LOTUS

Let X be a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$ then

$$E(g(X)) = \sum_{x \in ImX} g(x)P(X = x)$$

Theorem 9.2. LOTUS Continuous case

Let X be a continuous random variable with f_X consider a function $g : \mathbb{R} \to \mathbb{R}$ then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

provided that $\int_{-\infty}^{\infty} |g(x)|f_X(x)dx < \infty$

Theorem 9.3.

Consider a discrete/continuous random variable X with finite expectation

1. X is a non-negative, then $E(X) \geq 0$
2. $a, b \in \mathbb{R}$ then $E(aX + b) = aE(X) + b$

9.3 Variance

Definition 9.3.1. Variance

Let X be a discrete/continuous random variable. Then its variance is defined as $Var(X) = E[(X - E(X))^2]$ provided that it exists. Often we write $sigma^2 = Var(X)$

Theorem 9.4.

For a discrete/continuous random variable with finite variance we ahve that

$$Var(X) = E(X^2) - [E(X)]^2$$

Theorem 9.5.

Let X be a discrete/continuous random variable with finite variancve and consider deterministic constants $a, b \in \mathbb{R}$. Then

$$Var(aX + b) = a^2 Var(X)$$

10 Bridging lecture: Multivariate calculus

10.1 Partial derivatives

Partial derivatives: you just differentiate wrt to whatever letter you have at the bottom of the $\frac{\partial}{\partial x}$(so in this case it would be x) and just treat the other variables as constants. This is basically the same for multivariate integrals. Also in general:

$$\frac{\partial^2 f(x,y)}{\partial x \partial y} = \frac{\partial^2 f(x,y)}{\partial y \partial x}$$

Jacobian
The Jacobian of the transformation is defeind as the 2×2 matrix of all possible first order partial derivatives.

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix}$$

10.2 Change of variables formula

As in the univariate casem a change of variables formula also exist for the multivariate case. In that formula, a so-called Jacobian appears. we will explain the key ideas again in the bivariate setting,

$$\text{Let } f : \mathbb{R}^2 \to \mathbb{R}. \text{ we define the mapping } T : \mathbb{R}^2 \to \mathbb{R}^2 \text{ by}$$
$$T(x, y) = (u(x, y), \mu(x, y))$$

and assume that T is a bijetion from the domain $D\mathbb{R}^2$ to some range $S\mathbb{R}^2$. Then we can write $T^{-1} : S \to D$ for the inverse mapping of T, i.e. $(x, y) = T^{-1}(u, \nu)$. For the first component we write $x = x(u, \nu)$ and for the second $y = y(u, \nu)$. The Jacobian determinant of T^{-1} is defined as the determinant

$$J(u, \nu) = det(\frac{\partial(x,y)}{\partial(u,\nu)}) = det \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial \nu} - \frac{\partial x}{\partial \nu}\frac{\partial y}{\partial u}$$

The change of variable formula states that $(undermildconditions)$

$$\int \int_D f(x, y)dxdy = \int \int_S f(x(u, \nu), y(u, \nu))|J(u, \nu)|dud\nu$$

11 Multovariate random variables

11.1 Multivariate distributions

11.1.1 The bivariate case

Definition 11.1.1. Joint distirbution function

The joint distribution function of the random vector (X, Y) is defined as the mapping $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ given by

$$F_{X,Y}(x, y) = P(\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq y) \text{ for any } x, y \in \mathbb{R}$$

Using our shortened notationm we typically write

$$F_{x,Y}(x_1, y_1) = P(X \leq x, Y \leq y) \text{ for any } x, y \in \mathbb{R}$$

- $F_{X,Y}$ is non-decreasing in each variable, meaning that

$$F_{X,Y](x_1,y_1) \leq F_{X,Y}(x_2,y_2)} \text{ if } x_1 \leq x_2 \text{ and } y_1 \leq y_2$$

- $F_{X,Y}$ is continuous from above(the multivariate version of right-continuity), i.e. for two sequences $(x_n), (y_n)$ which approach x and y from the rights $n \to \infty$ we get that $F_{x,Y}(x_n, y_n) \to F$