Applied Probability

Lecture 1 - Probability

Guido Montúfar
montufar@math.ucla.edu

STATS 200A - Fall 2023

UCLA

Probability space

Definition 1 (Probability space)

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is composed of

- the set Ω of all possible outcomes of a random experiment,
- a σ-algebra $\mathcal{F}$ defined on Ω, and
- a probability $\mathbb{P}$ defined on the measurable space $(\Omega, \mathcal{F})$.

Definition 2 (Random experiment)

A random experiment is any procedure that can be repeated and has a well-defined set of possible outcomes Ω, known as the sample space of the experiment.

Definition 2 (Random experiment)

A random experiment is any procedure that can be repeated and has a well-defined set of possible outcomes Ω, known as the sample space of the experiment.

Example 3

A coin toss is a random experiment with sample space $\Omega = \{H, T\}$.

Example 4 (Experiment: roll a die)

Possible outcomes: ⚀, ⚁, ⚂, ⚃, ⚄, ⚅
Sample space: $S = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$
Event example: “Larger than 4” $E = \{$⚄, ⚅$\}$

Definition 5 (The sample space)

The sample space of an experiment is the collection of all possible outcomes of the experiment. The sample space S of an experiment can be thought of as a set.

Definition 6 (Event)

An event is a well-defined set of possible outcomes of the experiment. We say that an event E has occurred if the outcome s of the experiment is in E.

Note: Not every set of possible outcomes will be called an event. Instead, we will require that the collection of events is a σ-algebra.

Sigma algebra

Probability is not defined for all subsets of S. It is defined for subsets that belong to a meaningful collection called a σ-algebra.

Definition 7 (Sigma algebra)

A σ-algebra (sigma-algebra) on a set Ω is a nonempty collection $\mathcal{F}$ of subsets of Ω that is closed under complement, countable unions, and countable intersections.

1. $\emptyset \in \mathcal{F}$ (the empty set is an element of $\mathcal{F}$)
2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (closed under complementation)
3. If $A_1, A_2, \ldots \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$ (closed under unions)

Since $(\cup_i A_i)^c = \cap_i A_i^c$, we also have

3'. If $A_1, A_2, \ldots \in \mathcal{F}$, then $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$ (closed under intersections)

Note: It is an algebra because it is closed under complement, union and intersection. It is a σ-algebra because the union and intersection can be countable union and countable intersection.

Example 8

If $\Omega = \{a, b, c, d\}$, a possible σ-algebra is

$$\mathcal{F} = \{\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}.$$

Example 8

If $\Omega = \{a, b, c, d\}$, a possible σ-algebra is

$$\mathcal{F} = \{\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}.$$

Closed under complement:

$$\{a, b, c, d\} \setminus \emptyset = \{a, b, c, d\} \in \mathcal{F},$$
$$\{a, b, c, d\} \setminus \{a, b\} = \{c, d\} \in \mathcal{F},$$
$$\{a, b, c, d\} \setminus \{c, d\} = \{a, b\} \in \mathcal{F},$$
$$\{a, b, c, d\} \setminus \{a, b, c, d\} = \emptyset \in \mathcal{F}, \ldots$$

Closed under countable unions:

$$\{a, b\} \cup \{c, d\} = \{a, b, c, d\} \in \mathcal{F},$$
$$\{a, b\} \cup \emptyset = \{a, b\} \in \mathcal{F}, \ldots$$

Closed under countable intersections:

$$\{a, b, c, d\} \cap \emptyset = \emptyset \in \mathcal{F},$$
$$\{a, b, c, d\} \cap \{c, d\} = \{c, d\} \in \mathcal{F},$$
$$\{a, b\} \cap \{c, d\} = \emptyset \in \mathcal{F}, \ldots$$

Definition 9 (Measurable space)

The pair $(\Omega, \mathcal{F})$ is called a measurable space.

Definition 9 (Measurable space)

The pair $(\Omega, \mathcal{F})$ is called a measurable space.

Example 10

The pair $(\Omega, \mathcal{F})$ with sample space $\Omega = \{H, T\}$ and σ-algebra $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ is a measurable space.

In a given experiment, we assign to each event A in the sample space Ω a number $\Pr(A)$ that indicates its probability.

Definition 11 (Probability measure)

A probability is a non-negative measure $\mathbb{P}\colon \mathcal{F} \to \mathbb{R}_{\geq 0}$ with total mass equal to one. That is,

- $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$,
- $\mathbb{P}(\Omega) = 1$,
- for any sequence $A_n, n = 1, 2, \ldots$, of pairwise disjoint events,

$$\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n).$$

Events with probability zero are said to be *null*.
A property satisfied with probability one is *almost sure (a.s.)*.

Generated sigma algebra

A σ-algebra can be generated by a set of basic events by combining them by complement, countable union and intersection. This creates the smallest σ-algebra containing these simple events.

Definition 12 (σ-algebra generated by a family)

Let F be an arbitrary family of subsets of Ω. Then the σ-algebra generated by F, denoted $\sigma(F)$, is the smallest σ-algebra which contains every set in F.

If F is empty, then $\sigma(F) = \{\emptyset, \Omega\}$. Otherwise, $\sigma(F)$ consists of all the subsets of Ω that can be constructed from elements of F by a countable number of complement, union and intersections.

Example 13

Let $\Omega = \{a, b, c, d\}$. The σ-algebra generated by $F = \{\emptyset, \{a, b\}\}$ is $\sigma(F) = \{\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}$.

Example 14

If $\Omega = [0, 1]$, the basic events could be the intervals $(a, b) \subseteq [0, 1]$. The σ-algebra generated by the intervals is a Borel σ-algebra.

Definition 15 (Borel σ-algebra)

The Borel σ-algebra of $\mathbb{R}$, denoted $\mathcal{B}(\mathbb{R})$ is the σ-algebra generated by the intervals $(-\infty, x]$ for $x \in \mathbb{R}$.

Example 16

Let Ω be all infinite sequences of heads and tails. The simple events could be all the events about the outcome of the first n tosses.

Definition 17 (σ-algebra generated by a function)

- Let $f\colon \Omega \to \Omega'$ be a function from a set Ω to a set Ω', and let $\mathcal{F}'$ be a σ-algebra of subsets of Ω'.
- Then the σ-algebra generated by f, denoted $\sigma(f)$, is the collection of inverse images $f^{-1}(S)$ of the sets $S \in \mathcal{F}'$. That is,

$$\sigma(f) = \{f^{-1}(S)\colon S \in \mathcal{F}'\}.$$

- A function $f\colon \Omega \to \Omega'$ is measurable with respect to a σ-algebra $\mathcal{F}$ of subsets of Ω if and only if $\sigma(f)$ is a subset of $\mathcal{F}$.

Why a sigma algebra?

Why do we need to consider a σ-algebra?
On one hand $\mathcal{F}$ includes any countable unions of elements in $\mathcal{F}$.
On the other hand, elements not in $\mathcal{F}$ have no defined probability.
Why do we include / exclude so many sets?

Example 18 (Use of countable unions)

In order to formulate the strong law of large numbers we consider countable unions and intersections of events.

Example 19 (Taking arbitrary sets gives problems)

If we admit arbitrary sets, then we can end up assigning different probabilities to the same event, resulting in contradictions.
This is illustrated by the Banach-Tarski paradox.

Banach-Tarski paradox

- A ball can be cut into a finite number of disjoint pieces which can be reassembled into two identical copies of the original ball.


(1)

- Also, a small ball can be cut into pieces which can be reassembled into a large ball.

(1) An illustration of the effects of the Banach–Tarski Paradox. Public domain, accessed via Wikimedia Commons

Conditional probability, independent events

Conditional probability

Definition 20 (Conditional probability)

The conditional probability of the event A given that the event B has occurred is denoted $\Pr(A|B)$. If $\Pr(B) > 0$,

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

The conditional probability $\Pr(A|B)$ is not defined if $\Pr(B) = 0$.

Theorem 21 (Multiplication rule)

- *Let A and B be events. If* $\Pr(B) > 0$, *then*

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

  And if $\Pr(A) > 0$, *then*

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

- *Let $A_1, \ldots, A_n$ be events. If* $\Pr(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$, *then*

$$\begin{aligned}\Pr(A_1 \cap A_2 \cap \cdots \cap A_n) = &\Pr(A_1) \cdot \Pr(A_2|A_1) \cdot \Pr(A_3|A_1 \cap A_2) \\ &\cdots \\ &\Pr(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).\end{aligned}$$

Independent events

Definition 22 (Independence)

- Two events A and B are independent if

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

- If $\Pr(A) > 0$ and $\Pr(B) > 0$, then A and B are independent iff

$$\Pr(A|B) = \Pr(A) \quad \text{and} \quad \Pr(B|A) = \Pr(B).$$

Definition 23 (Mutually independent events)

The k events $A_1, \dots, A_k$ are independent (or mutually independent) if, for every subset $A_{i_1}, \dots, A_{i_j}$, $j \leq k$,

$$\Pr(A_{i_1} \cap \dots \cap A_{i_j}) = \Pr(A_{i_1}) \cdots \Pr(A_{i_j}).$$

Theorem 24 (Independence and conditional probability)

Let $A_1, \dots, A_k$ *be events such that* $\Pr(A_1 \cap \dots \cap A_k) > 0$. *Then* $A_1, \dots, A_k$ *are independent iff for every two disjoint subsets* $\{i_1, \dots, i_m\}$ *and* $\{j_1, \dots, j_l\}$ *of* $\{1, \dots, k\}$, *we have*

$$\Pr(A_{i_1} \cap \dots \cap A_{i_m} | A_{j_1} \cap \dots \cap A_{j_l}) = \Pr(A_{i_1} \cap \dots \cap A_{i_m}).$$

Learning that some of the events occur does not change the probability that any combination of the other events occurs.

Conditionally indepenent events

Conditional probability and independence combine into one of the most versatile models of data collection.

Definition 25 (Conditional independence)

We say that $A_1, \ldots, A_k$ are conditionally independent given B if, for every subset $A_{i_1}, \ldots, A_{i_j}$, $j \leq k$,

$$\Pr(A_{i_1} \cap \cdots \cap A_{i_j} | B) = \Pr(A_{i_1} | B) \cdots \Pr(A_{i_j} | B).$$

Note: Even if $A_1, \ldots, A_k$ are conditionally independent given B, they need not be conditionally independent given B^c.

Bayes' theorem

Theorem 26 (Bayes Theorem)

Let $B_1, ..., B_k$ form a partition of the sample space Ω such that $\Pr(B_j) > 0$ for $j = 1, ..., k$, and let A be an event with $\Pr(A) > 0$. Then, for $i = 1, ..., k$,

$$\Pr(B_i|A) = \frac{\Pr(B_i \cap A)}{\Pr(A)} = \frac{\Pr(B_i)\Pr(A|B_i)}{\sum_j \Pr(B_j)\Pr(A|B_j)}.$$

Note: The key point here is that $\Pr(B_i|A)$ is computed in terms of $\Pr(B_i)$ and conditional probabilities of the form $\Pr(A|B_i)$.

Counting techniques

- A sample space containing n outcomes $s_1, ..., s_n$ is called a simple sample space if $p_i = 1/n$ for each $i = 1, \ldots, n$.
- If an event A in this sample space has m outcomes, then

$$\Pr(A) = m/n.$$

Thus one is interested in the cardinality of events.

Definition 27 (Permutations, $P_{n,k}$)

Given a set of n elements, select k of them without replacement. Each such outcome is called a permutation of n elements taken k at the time. The number of distinct permutations is denoted $P_{n,k}$.

Definition 28 (Combinations, $C_{n,k}$)

Consider a set of n elements. Each subset of size k is called a combination of n elements taken k at the time. The number of distinct such combinations is denoted $C_{n,k}$.

ordered sampling with replacement	n^k
ordered sampling without replacement	$P_{n,k} = \frac{n!}{(n-k)!}$
unordered sampling without replacement	$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$
unordered sampling with replacement	$\binom{n+k-1}{k}$

Measure

A few logical steps to develop a theory of measure:

- Start from simple shapes, such as Jordan simple shapes, i.e., finite unions of nonoverlapping rectangles.
- For each set A, define its outer measure as

$$\mu^*(A) = \inf_{A \supseteq S} \mu(S),$$

  where the inf is over simple shapes that cover A.
- Similarly, define the inner measure

$$\mu_*(A) = \sup_{S \subseteq A} \mu(S),$$

  where the sup is over simple shapes inside A.
- The set A is measurable if $\mu^*(A) = \mu_*(A)$, then denoted $\mu(A)$.

Lebesgue measure

- A box in $\mathbb{R}^n$ is a set of the form $B = \times_{i=1}^n [a_i, b_i]$. The volume is $\mathrm{vol}(B) = \prod_{i=1}^n (b_i - a_i)$.
- For any subset A of $\mathbb{R}^n$, define its outer measure by

$$\lambda^*(A) = \inf_{\mathcal{C}} \sum_{B \in \mathcal{C}} \mathrm{vol}(B),$$

  where $\mathcal{C}$ is any countable collection of boxes whose union covers A.
- The set A is Lebesgue measurable if, for every $S \subseteq \mathbb{R}^n$,

$$\lambda^*(S) = \lambda^*(S \cap A) + \lambda^*(S \setminus A).$$

- For Lebesgue measurable A, Lebesgue measure is $\lambda(A) = \lambda^*(A)$.
- The Lebesgue measurable sets form a σ-algebra.
- The Borel measure agrees with the Lebesgue measure on those sets for which it is defined; however, there are more Lebesgue-measurable sets than there are Borel measurable sets.

Law of large numbers

We can take a look at the law a large numbers from a geometric perspective using the Lebesgue measure. Let $X_1, \ldots, X_n$ be iid uniform on $[0, 1]$. The mean is $\mu = \frac{1}{2}$.

- The law of large numbers states that

$$\text{the sample mean } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ “converges” to the mean } \mu.$$

- For the weak law of large numbers, convergence in probability, aka weak convergence, $\overline{X}_n \xrightarrow{p} \mu$, meaning that

$$\text{for each } \epsilon > 0, \qquad \lim_{n\to\infty} \Pr(|\overline{X}_n - \mu| < \epsilon) = 1.$$

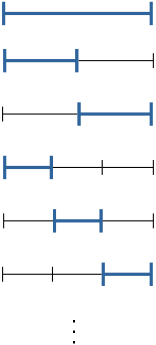
- For the strong law of large numbers, convergence almost surely, aka strong convergence, $\overline{X}_n \xrightarrow{as} \mu$, meaning that

$$\text{for each } \epsilon > 0, \qquad \Pr(\lim_{n\to\infty} |\overline{X}_n - \mu| < \epsilon) = 1.$$

Example 29 (of weak and strong LLN)

Let s be uniform on $[0,1]$, let $X(s) = s$ and

$$\begin{aligned}
X_1(s) &= s + I_{[0,1]}(s)\\
X_2(s) &= s + I_{[0,\frac{1}{2}]}(s)\\
X_3(s) &= s + I_{[\frac{1}{2},1]}(s)\\
X_4(s) &= s + I_{[0,\frac{1}{3}]}(s)\\
X_5(s) &= s + I_{[\frac{1}{3},\frac{2}{3}]}(s)\\
X_6(s) &= s + I_{[\frac{2}{3},1]}(s)\\
&\vdots
\end{aligned}$$



Then $X_n(s) = X(s)$ except for s in an interval that gets narrower as n increases, and $\Pr(|X_n - X| < \epsilon)$ tends to one.

However, there is no N with $|X_n - X| < \epsilon$ for all $n \geq N$.

Thus $X_n \xrightarrow{p} X$, but $X_n \overset{as}{\not\to} X$.

Geometric interpretation of weak LLN

In example with $X_1, \ldots, X_n$ iid uniform on $[0,1]$, the weak law of large numbers interpreted geometrically says the volume of

$$A_{n,\epsilon} = \left\{ (x_1, \ldots, x_n) \in [0,1]^n \colon \frac{1}{n} \sum_{i=1}^{n} x_i \in [\tfrac{1}{2} - \epsilon, \tfrac{1}{2} + \epsilon] \right\}$$

approaches one, occupying almost all volume of the n-cube $[0,1]^n$.

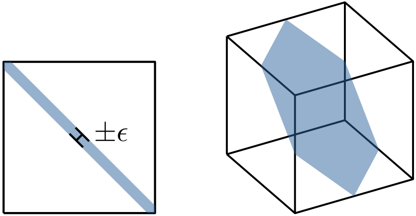


Figure 1: The ϵ thickened diagonal slice contains most volume of the n-cube as n tends to infinity. Shown are $n = 2, 3$.

Geometric interpretation of strong LLN

In the example with $X_1, \ldots, X_n$ iid uniform on $[0,1]$, the strong law of large numbers interpreted geometrically says A has volume 1. Here we consider the set of convergent sequences

$$A = \left\{(x_1, x_2, \ldots) \in [0,1]^\infty \colon \frac{1}{n} \sum_i x_i \to \tfrac{1}{2}\right\},$$

which is

$$A = \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} A_{n,\epsilon=1/k},$$

and may be regarded as an arbitrarily thin diagonal slice of $[0,1]^\infty$.

Concentration of measure

The law of large numbers results from a phenomenon called concentration of measure.

- Hoeffding's inequality, a concentration inequality, states

$$\mu(A^c_{n,\epsilon}) \leq 2\exp(-2n\epsilon^2).$$

  This directly implies the weak LLN.

- For the strong LLN,[2]

$$A^c = \cup_{k=1}^{\infty} \cap_{N=1}^{\infty} \cup_{n=N}^{\infty} A^c_{n,\epsilon=1/k}$$

  For fixed ϵ, $\mu(\cap_{N=1}^{\infty} \cup_{n=N}^{\infty} A^c_{n,\epsilon}) \leq \mu(\cup_{n=N}^{\infty} A^c_{n,\epsilon}) \leq \sum_{n=N}^{\infty} \mu(A^c_{n,\epsilon})$ for any fixed N. The LHS goes to 0 as $N \to \infty$, since

$$\sum_{n=1}^{\infty} \mu(A^c_{n,\epsilon}) \leq \sum_{n=1}^{\infty} 2\exp(-2n\epsilon^2) < \infty.$$

  Thus $\mu(A^c) = 0$ and $\mu(A) = 1$.

[2] More precisely, we use $\tilde{A}_{n,\epsilon} \subseteq [0,1]^{\infty}$, $\tilde{A}_{n,\epsilon} = A_{n,\epsilon} \times [0,1] \times [0,1] \times \cdots$

Above we encountered a technical step that is known as

Lemma 30 (Borel-Cantelli Lemma)

If the sum of the probabilities of the events $\{E_n\}$ *is finite,*

$$\sum_{n=1}^{\infty} \Pr(E_n) < \infty,$$

then the probability that infinitely many of them occur is 0*, that is,*

$$\Pr(\limsup_{n\to\infty} E_n) = 0.$$

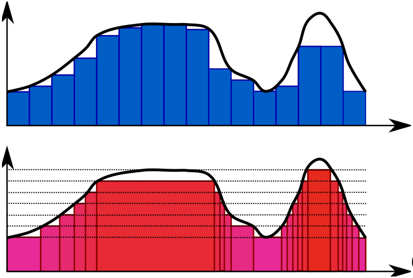
Here the lim sup denotes the set of outcomes that occur infinitely many times within the sequence of events,

$$\limsup_{n\to\infty} E_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} E_k.$$

Thus, the lemma states that if the sum of probabilities of the events E_n is finite, then the set of outcomes that are "repeated" infinitely many times must have probability zero.

Lebesgue integral

- For a continuous random variable X with pdf $f(x)$, we have $\Pr(X \in A) = \int_A f(x)dx$ and $\mathbb{E}(h(X)) = \int h(x)f(x)dx$. In measure theoretical probability these are Lebesgue integrals.
- Whereas the Riemann integral discretizes the domain of the function, the Lebesgue integral discretizes the range. Why?


(3)

(3) Riemann and Lebesgue Integral. CC0 1.0 Public Domain, accessed via Wikimedia Commons

- If we have a sequence of functions with $f_k(x) \to f(x)$ for x in the domain, then with mild conditions we may obtain $\int f_k \to \int f$ for the Lebesgue integral but not Riemann.
- For a positive f, we may discretize the range into bins $[y_i, y_{i+1} = y_i + \Delta y)$ and then

$$\int f d\mu = \lim_{\Delta y \to 0} \sum_i \mu(\{x \colon f(x) > y_i\})\Delta y.$$

  For this we need to be able to compute the Lebesgue measure of $\{x \colon f(x) > y\}$ for each y. Such f is called a *measurable*. Roughly, it means that the superlevel sets can be broken into countably many disjoint simple pieces.
- One may also define the Lebesgue integral as $\int f = \sup_{s \le f} \int s$, for simple functions $s(x) = \sum_i a_i \mathbf{1}_{S_i}(x)$, for which $\int s d\mu = \sum_i a_i \mu(S_i)$.

Convergence theorems

Theorem 31 (Monotone convergence)

If $f_k(x) \to f(x)$ for each x monotonically, i.e., $f_k(x) \leq f_{k+1}(x)$ for each x, and each f_k is measurable, then

$$\int f_k \to \int f.$$

To prove this, one observes first $\int f_k \leq \int f$ due to monotonicity. Then, for each simple function $s(x) = \sum_i a_i \mathbf{1}_{S_i}(x)$ below f, eventually $\int f_k \geq \int s$ since $f_k(x) \to f(x)$ for each x and $s(x) \leq f(x)$.

M.H. DeGroot and M.J. Schervish.
Probability and Statistics, Fourth Edition.
Pearson. Pearson Education, 2012.
http://bio5495.wustl.edu/Probability/Readings/DeGroot4thEdition.pdf.

Random variable

Definition 32 (real random variable)

- A real random variable is a measurable function $X\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, such that $X^{-1}(B) \in \mathcal{F}$ for all Borel sets $B \in \mathcal{B}(\mathbb{R})$.
- The distribution $\mathbb{P}_X$ is the image of $\mathbb{P}$ by X, which is characterized by the values of the distribution function $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X((-\infty, x])$ for $x \in \mathbb{R}$.
- The set $X^{-1}(\mathcal{B}(\mathbb{R}))$ is a σ-algebra on Ω, called the σ-algebra generated by X and denoted by $\sigma(X)$. It is the smallest σ-algebra that makes X measurable.

Definition 33 (real random vector)

A random vector $(X_1, \ldots, X_d)\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, with dimension d, is a finite family of d real random variables, or a random variable taking values in $\mathbb{R}^d$.

Its distribution is characterized by the values

$$\mathbb{P}_{(X_1,\ldots,X_d)}((-\infty, x_1] \times \cdots \times (-\infty, x_d]), \quad \text{for } (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

Definition 34 (absolutely continuous)

Let μ and ν be two nonnegative measures on a measure space $(\Omega, \mathcal{F})$. Then μ is said to be absolutely continuous with respect to ν, denoted $\mu \ll \nu$, if $\mu(A) = 0$ for all $A \in \mathcal{F}$ such that $\nu(A) = 0$.

Theorem 35 (Radon-Nikodym)

Let μ and ν be two σ-finite measures on $(\Omega, \mathcal{F})$. If $\mu \ll \nu$, then there exists a nonnegative Borel function $f : (\Omega, \mathcal{F}, \nu) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that

$$\mu(A) = \int_A d\mu = \int_A f \, d\nu, \quad A \in \mathcal{F}.$$

The ν-a.e. defined function f is unique. It is referred to as the density (or Radon-Nikodym derivative) of μ with respect to ν and denoted by $f = d\mu/d\nu$.

Finite: $\mu(A) < \infty$ for all $A \in \mathcal{F}$.
Borel function: If $\mathcal{F}$ is the Borel σ-algebra of Ω, a $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{R}$ is called a Borel function.

Definition 36 (density function)

A d-dimensional random vector X is absolutely continuous if its distribution is absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}^d$. More on the Lebesgue measure below.

This means that there exists a nonnegative Borel function $f_X : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ such that $d\mathbb{P}_X(x) = f_X dx$, or

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x) f_X(x) dx,$$

for all bounded Borel functions h. The function f_X is called the probability density function of the random variable X.