# Department of Economics, University of Warwick
# Mathematics for Economics
# Lecture 9: Probability

Andrés Carvajal, with revisions by Peter Hammond

September, 2011; revised 2014

# Outline

# Power Sets

Fix an abstract set $S \neq \emptyset$.

In case it is finite, its cardinality, denoted by $\#S$,
is the number of distinct elements of $S$.

The power set of $S$ is the family $\mathcal{P}(S) := \{ T \mid T \subseteq S \}$
of all subsets of $S$.

Sometimes the power set is denoted by $2^S$,
perhaps for two reasons.

1. One can define the bijection

$$\mathcal{P}(S) \ni T \mapsto f(T) \in \{0,1\}^S := \{(x_s)_{s \in S} \mid \forall s \in S : x_s \in \{0,1\}\}$$

by $f(T)_s = 1_T(s) = \begin{cases} 1 & \text{if } s \in T \\ 0 & \text{if } s \notin T \end{cases}$.

2. So $\#\mathcal{P}(S) = 2^{\#S}$.

The family $\mathcal{A} \subseteq \mathcal{P}(S)$ is a Boolean algebra on $S$ just in case

1. $\emptyset \in \Sigma$;
2. $A \in \Sigma$ implies $S \setminus A \in \Sigma$;
3. if $A, B$ lie in $\mathcal{A}$, then $A \cup B \in \mathcal{A}$.

The family $\Sigma \subseteq \mathcal{P}(S)$ is a $\sigma$-algebra just in case
it is a Boolean algebra with the stronger property:
if $(A_n)_{n=1}^{\infty}$ is a countably infinite family of sets in $\Sigma$,
then $\cup_{n=1}^{\infty} A_n \in \Sigma$.

The pair $(S, \Sigma)$ is a measurable space just in case $\Sigma$ is a $\sigma$-algebra.

# Exercise on Boolean Algebras and Sigma-Algebras

## Exercise

1. *Let $\mathcal{A}$ be a Boolean algebra on $S$.*
   *Prove that if $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$.*
2. *Let $\Sigma$ be a Boolean algebra on $S$.*
   *Prove that if $(A_n)_{n=1}^\infty$ is a countably infinite family of sets in $\Sigma$,*
   *then $\cap_{n=1}^\infty A_n \in \Sigma$.*

**Hint**

1. For part 1, use de Morgan's law

$$S \setminus (A \cap B) = (S \setminus A) \cup (S \setminus B)$$

2. For part 2, use the infinite extension of de Morgan's law.

# Generating a Sigma-Algebra

### Theorem

*Let $\{\Sigma_i \mid i \in I\}$ be any indexed family of $\sigma$-algebras.*

*Then the intersection $\Sigma^\cap := \cap_{i \in I} \Sigma_i$ is also a $\sigma$-algebra.*

Proof left as an exercise.

Let $X$ be a space, and $\mathcal{F} \subset 2^X$ any family of subsets.

Since $2^X$ is obviously a $\sigma$-algebra,
there exists a non-empty set $\mathcal{S}(\mathcal{F})$ of $\sigma$-algebras that include $\mathcal{F}$.

Let $\sigma(\mathcal{F})$ denote the intersection $\cap\{\Sigma \mid \Sigma \in \mathcal{S}(\mathcal{F})\}$;
it is the smallest $\sigma$-algebra that includes $\mathcal{F}$.

### Exercise

*Let $X$ be any uncountably infinite set, and let $\mathcal{F} := \{\{x\} \mid x \in X\}$ denote the family of all singleton subsets of $X$.*

*Show that $\sigma(\mathcal{F})$ consists of all subsets of $X$ that are either countable, or co-countable (i.e., have a countable complement).*

# Topological Spaces

Given a set $X$, a topology $\mathcal{T}$ on $X$
is a family of open subsets $U \subseteq X$ satisfying:

1. $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$;
2. if $U, V \in \mathcal{T}$, then $U \cap V \in \mathcal{T}$;
3. if $\{U_\alpha \mid \alpha \in A\}$ is any family of open sets in $\mathcal{T}$, then the union $\cup_{\alpha \in A} U_\alpha \in \mathcal{T}$.

Thus, finite intersections and arbitrary unions of open sets are open.

A topological space $(X, \mathcal{T})$ is any set $X$ together with a topology $\mathcal{T}$ that consists of all the open subsets of $X$.

# The Metric Topology

Let $(X, d)$ be any metric space.

The open ball of radius $r$ centred at $x$ is the set

$$B_r(x) := \{y \in X \mid d(x, y) < r\}$$

The metric topology $\mathcal{T}_d$ of $(X, d)$ is the smallest topology that includes the entire family $\{B_r(x) \mid x \in X \ \& \ r > 0\}$ of all open balls in $X$.

# Borel Sigma-Algebra

Let $(X, \mathcal{T})$ be any topological space.

Its Borel $\sigma$-algebra is defined as $\sigma(\mathcal{T})$
— i.e., the smallest $\sigma$-algebra containing every open set of $X$.

Suppose the topological space is a metric space $(X, d)$
with its metric topology $\mathcal{T}_d$.

Then the Borel $\sigma$-algebra is generated
by all the open balls $B_r(x) := \{x' \in X \mid d(x, x') < r\}$ in $X$.

For the case of the real line when $X = \mathbb{R}$,
its Borel $\sigma$-algebra is generated by all the open intervals of $\mathbb{R}$.

# Outline

Let $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty + \infty\} = [-\infty, +\infty]$
denote the extended real line which, at each end,
has an endpoint added at infinity.

Let $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{+\infty\} = [0, +\infty]$ be the non-negative part of $\bar{\mathbb{R}}$.

Any family $\mathcal{F}$ of subsets $A \subseteq X$ is said to be pairwise disjoint
just in case $A \cap B = \emptyset$ whenever $A, B \in \mathcal{F}$ with $A \neq B$.

A mapping $\mu : \Sigma \to \bar{\mathbb{R}}_+$ is said to be additive or finitely additive
just in case, for any pair $\{A, B\}$ of disjoint sets in $\Sigma$,
one has $\mu(A \cup B) = \mu(A) + \mu(B)$;

For any finite collection $\{A_n\}_{n=1}^k$ of pairwise disjoint sets in $\Sigma$,
note how finite additivity implies that

$$\mu\left(\cup_{n=1}^k A_n\right) = \sum\nolimits_{n=1}^k \mu(A_n)$$

# Measure as a Countably Additive Set Functions

Let $(X, \Sigma)$ be a measurable space.

A set function $\mu : \Sigma \to \bar{\mathbb{R}}_+$ is said
to be $\sigma$-additive or countably additive just in case,
for any countable collection $\{A_n\}_{n=1}^\infty$
of pairwise disjoint sets in $\Sigma$, one has

$$\mu\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n)$$

A measure on a measurable space $(X, \Sigma)$
is a countably additive set function $\mu : \Sigma \to \bar{\mathbb{R}}_+$
satisfying the requirement that $\mu(\emptyset) = 0$.

A measure space is a triple $(X, \Sigma, \mu)$ where

1. $\Sigma$ is a $\sigma$-algebra on $X$;
2. $\mu$ is a measure on $(X, \Sigma)$.

Example

A prominent example of a measure space is $(\mathbb{R}, \mathcal{B}, \ell)$ where:

1. $\mathcal{B}$ is the Borel $\sigma$-algebra
   induced by the open sets of the real line $\mathbb{R}$;
2. the measure $\ell(J)$ of any interval $J \subset \mathbb{R}$ is its length,
   defined by $\ell([a, b]) = \ell([a, b)) = \ell((a, b]) = \ell((a, b)) = b - a$;
3. $\ell$ is extended to all of $\mathcal{B}$ to satisfy countable additivity
   (it can be shown that this extension is unique).

A set $N \subset \mathbb{R}$ is null just in case
there exists a Borel subset $B \in \mathcal{B}$ with $\ell(B) = 0$
such that $N \subset B$.

This is possible for some non-Borel subsets of $\mathbb{R}$.

Let $\mathcal{N}$ denote the family of null subsets of $\mathbb{R}$.

These null sets can be used to generate the Lebesgue $\sigma$-algebra
of Lebesgue measurable sets, which is $\sigma(\mathcal{B} \cup \mathcal{N})$.

The symmetric difference of any two sets $S$ and $B$ is defined
as the set

$$S \triangle B := (S \setminus B) \cup (B \setminus S) = (S \cup B) \setminus (S \cap B)$$

of elements $s$ that belong to one of the two sets, but not to both.

One can show that $S \in \sigma(\mathcal{B} \cup \mathcal{N})$ if and only if
there exists a Borel set $B \in \mathcal{B}$ such that $S \triangle B \in \mathcal{N}$
— i.e., $S$ differs from a Borel set only by a null set.

# The Lebesgue Real Line

There is a well-defined function $\lambda : \sigma(\mathcal{B} \cup \mathcal{N}) \to \bar{R}_+$
that satisfies $\lambda(S) := \ell(B)$ whenever $S \triangle B \in \mathcal{N}$.

Moreover, one can prove
that the function $S \mapsto \lambda(S)$ is countably additive.

This makes $\lambda$ a measure, called the Lebesgue measure.

The associated measure space $(\mathbb{R}, \sigma(\mathcal{B} \cup \mathcal{N}), \lambda)$
is called the Lebesgue real line.

# Outline

## Simple Functions

Let $(X, \Sigma, \mu)$ be a measure space.

Given any set $E \in \Sigma$, the indicator function of $E$ is defined by

$$X \ni x \mapsto 1_E(x) := \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}$$

The finite or countably infinite collection $\{E_k | k \in K\}$ of pairwise disjoint sets $E_k \in \Sigma$ is a partition of $X$ just in case $\cup_{k \in K} E_k = X$.

The function $f : X \mapsto \mathbb{R}$ is simple just in case there exist a partition $\{E_k | k \in K\}$ of $X$ together with a corresponding collection $(a_k)_{k \in K}$ of real constants such that $f(x) = \sum_{k \in K} a_k 1_{E_k}(x)$.

Note that the range $\{y \in \mathbb{R} \mid \exists x : y = f(x)$ of this step function is the precisely the set $\{a_k | k \in K\}$ of real constants.

Let $\mathcal{F}(X, \Sigma)$ denote the set of all simple functions on the measurable space $(X, \Sigma)$; in fact it is a real vector space.

# Integrating Simple Functions

Given a function $f : X \mapsto \mathbb{R}$, whenever possible
we want to define the integral $\int_X f(x)\, d\mu = \int_X f(x)\, \mu(dx)$.

The simple function $f(x) = \sum_{k \in K} a_k \, 1_{E_k}(x)$ is $\mu$-integrable
just in case one has $\sum_{k \in K} |a_k| \, \mu(E_k) < +\infty$.

In particular, when $K$ is infinite, this requires
the infinite series $\sum_{k \in K} a_k \, \mu(E_k)$ to be absolutely convergent.

Then we can define the integral $\int_X f(x)\, \mu(dx)$
of the $\mu$-integrable simple function $f(x) = \sum_{k \in K} a_k \, 1_{E_k}(x)$
as the real number $\sum_{k \in K} a_k \, \mu(E_k)$.

# Upper and Lower Bounds

Given the measure space $(X, \Sigma, \mu)$,
the function $f : X \to \mathbb{R}$ is measurable just in case,
for every Borel set $B \subset \mathbb{R}$, its inverse image satisfies

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

Note that we have defined a simple function to be measurable.

Given any function $f : X \to \mathbb{R}$, define the two sets

$$\begin{aligned} \mathcal{F}^*(f; X, \Sigma) &:= \{f^* \in \mathcal{F}(f; X, \Sigma) \mid \forall x \in X : f^*(x) \geq f(x)\} \\ \mathcal{F}_*(f; X, \Sigma) &:= \{f_* \in \mathcal{F}(f; X, \Sigma) \mid \forall x \in X : f_*(x) \leq f(x)\} \end{aligned}$$

of simple functions that are respectively upper or lower bounds for the function $f$.

The integral $\int_X f^*(x)\,\mu(dx)$
of each simple function $f^* \in \mathcal{F}^*(f; X, \Sigma)$,
is an over-estimate of the true integral of $f$.

But the integral $\int_X f_*(x)\,\mu(dx)$
of each simple function $f_* \in \mathcal{F}_*(f; X, \Sigma)$,
is an under-estimate of the true integral of $f$.

Define the upper integral and lower integral of $f$ as, respectively

$$\begin{array}{rrcl} & I^*(f) & := & \inf_{f^* \in \mathcal{F}^*(f;X,\Sigma)} \int_X f^*(x)\,\mu(dx) \\ \text{and} & I_*(f) & := & \sup_{f_* \in \mathcal{F}_*(f;X,\Sigma)} \int_X f_*(x)\,\mu(dx) \end{array}$$

Of course, in case $f$ is itself a simple function,
one has $I^*(f) = I_*(f) = \int_X f(x)\,\mu(dx)$.

### Definition

The function $f : X \to \mathbb{R}$ is integrable just in case it is measurable
and also the upper integral $I^*(|f|)$
of the function $x \mapsto |f(x)|$ is defined
(because $|f|$ is bounded above by an integrable simple function).

### Theorem

*If the function $f : X \to \mathbb{R}$ is integrable,*
*then its upper and lower integrals $I^*(f)$ and $I_*(f)$ are equal.*

So if $f : X \to \mathbb{R}$ is integrable,
then we can define its integral $\int_X f(x)\,\mu(dx)$
as the common value of $I^*(f)$ and $I_*(f)$.

# Outline

# Probability Measure and Probability Space

Fix a measurable space $(S, \Sigma)$,
where $S$ is a set of unknown states of the world.

Then $\Sigma$ is a $\sigma$-algebra of unknown events.

A probability measure on $(S, \Sigma)$ is a measure $\mathbb{P} : \Sigma \to \bar{R}_+$ satisfying the requirement that $\mathbb{P}(S) = 1$.

Countable additivity implies that $\mathbb{P}(E) + \mathbb{P}(E^c) = 1$ for every event $E \in \Sigma$, where $E^c := S \setminus E$.

It follows that $\mathbb{P}(E) \in [0, 1]$ for every $E \in \Sigma$.

# Properties of Probability

**Theorem**

*Let $(S, \Sigma, \mathbb{P})$ be a probability space.*

*Then the following hold for all $\Sigma$-measurable sets $E, E'$ etc.*

1. $\mathbb{P}(E) \leq 1$ *and* $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$;
2. $\mathbb{P}(E' \cap E^c) = \mathbb{P}(E') - \mathbb{P}(E' \cap E)$ *and*
   $\mathbb{P}(E \cup E') = \mathbb{P}(E) + \mathbb{P}(E') - \mathbb{P}(E \cap E')$;
3. *for every partition* $\{E_n\}_{n=1}^N$ *of* $S$ *into pairwise disjoint sets, one has* $\mathbb{P}(E) = \sum_{n=1}^N \mathbb{P}(E \cap E_n)$;
4. $\mathbb{P}(E \cap E') \geq \mathbb{P}(E) + \mathbb{P}(E') - 1$.
5. $\mathbb{P}(\cup_{n=1}^\infty E_n) \leq \sum_{n=1}^\infty \mathbb{P}(E_n)$.

# Two Limiting Properties

## Theorem

*Let $(S, \Sigma, \mathbb{P})$ be a probability space,*
*and $(E_n)_{n=1}^N$ an infinite sequence of $\Sigma$-measurable sets.*

1. *If $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$,*
   *then $\mathbb{P}(\cup_{n=1}^\infty E_n) = \lim_{n\to\infty} \mathbb{P}(E_n)$.*
2. *If $E_n \supseteq E_{n+1}$ for all $n \in \mathbb{N}$,*
   *then $\mathbb{P}(\cap_{n=1}^\infty E_n) = \lim_{n\to\infty} \mathbb{P}(E_n)$.*

# Proving the Limiting Properties

Proof.

1. Because $E_n \subseteq E_{n+1}$ for all $n \in \mathbb{N}$, one has

$$\begin{array}{rcl} E_n & = & E_1 \cup [\bigcup_{k=2}^n (E_k \setminus E_{k-1})] \\ \text{and} \quad \cup_{n=1}^\infty E_n & = & E_1 \cup [\bigcup_{k=2}^\infty (E_k \setminus E_{k-1})] \end{array}$$

where the sets $E_1$ and $\{E_k \setminus E_{k-1} \mid k = 2, 3, \ldots\}$ are all pairwise disjoint. Hence

$$\begin{array}{rcl} \mathbb{P}(E_n) & = & \mathbb{P}(E_1) + \sum_{k=2}^n \mathbb{P}(E_k \setminus E_{k-1}) \\ \mathbb{P}(\cup_{n=1}^\infty E_n) & = & \mathbb{P}(E_1) + \sum_{k=2}^\infty \mathbb{P}(E_k \setminus E_{k-1}) \\ & = & \lim_{n\to\infty} [\mathbb{P}(E_1) + \sum_{k=2}^n \mathbb{P}(E_k \setminus E_{k-1})] \\ & = & \lim_{n\to\infty} \mathbb{P}(E_n) \end{array}$$

2. Apply part 1 to the complements of the sets $E_n$.

□

# Conditional Probability: First Definition

Let $E^* \in \Sigma$ be such that $\mathbb{P}(E^*) > 0$.

The conditional probability measure on $E^*$ is the mapping

$$\Sigma \ni E \mapsto \mathbb{P}(E|E^*) := \frac{\mathbb{P}(E \cap E^*)}{\mathbb{P}(E^*)} \in [0,1]$$

The triple $(E^*, \Sigma(E^*), \mathbb{P}(\cdot|E^*))$ with

$$\Sigma(E^*) := \{E \cap E^* \mid E \in \Sigma\} = \{E \in \Sigma \mid E \subset E^*\}$$

is then a conditional probability space given the event $E^*$.

# Conditional Probability: Two Key Properties

**Theorem**

*Provided that* $\mathbb{P}(E) \in (0,1)$, *one has*

$$\mathbb{P}(E') = \mathbb{P}(E)\mathbb{P}(E'|E) + (1 - \mathbb{P}(E))\mathbb{P}(E'|E^c)$$

**Theorem**

*Let* $(E_k)_{k=1}^n$ *be any finite list of sets in* $\Sigma$.
*Provided that* $\mathbb{P}(\cap_{k=1}^{n-1} E_k) > 0$, *one has*

$$\mathbb{P}(\cap_{k=1}^n E_k) = \mathbb{P}(E_1)\,\mathbb{P}(E_2|E_1)\,\mathbb{P}(E_3|E_1 \cap E_2)\,\ldots\,\mathbb{P}(E_n| \cap_{k=1}^{n-1} E_k)$$

# Independence

The finite or countably infinite family $\{E_k\}_{k \in K}$ of events in $\Sigma$ is:

- pairwise independent if $\mathbb{P}(E \cap E') = \mathbb{P}(E)\,\mathbb{P}(E')$ whenever $E \neq E'$;
- independent if for any finite subfamily $\{E_k\}_{k=1}^n$, one has $\mathbb{P}(\cap_{k=1}^n E_k) = \prod_{k=1}^n \mathbb{P}(E_k)$.

## Exercise

*Let $S$ be the set $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and $\mathbb{P}$ the probability measure on $2^S$ satisfying $\mathbb{P}(\{s\}) = 1/9$ for all $s \in S$.*

*Consider the three events*

$$E_1 = \{1, 2, 7\}, \; E_2 = \{3, 4, 7\} \; \textit{and} \; E_3 = \{5, 6, 7\}$$

*Are these event pairwise independent? Are they independent?*

## Exercise

*Prove that if $\{E, E'\}$ is independent, then so is $\{E^c, E'\}$.*

# Random Variable

- The function $X : S \to \mathbb{R}$ is $\Sigma$-measurable just in case for every $x \in \mathbb{R}$ one has

$$X^{-1}(-\infty, x) := \{s \in S \mid X(s) \leq x\} \in \Sigma$$

- A random variable (with values in $\mathbb{R}$) is a $\Sigma$-measurable function $X : S \to \mathbb{R}$.
- The distribution function or cumulative distribution function (cdf) of $X$ is the mapping $F_X : \mathbb{R} \to [0, 1]$ defined by

$$x \mapsto F_X(x) = \mathbb{P}(\{s \in S \mid X(s) \leq x\})$$

# Outline

### Theorem

*The CDF of any random variable $s \mapsto X(s)$ satisfies:*

1. $\lim_{x \to -\infty} F_X(x) = 0$ *and* $\lim_{x \to +\infty} F_X(x) = 1$.
2. $x \geq x'$ *implies* $F_X(x) \geq F_X(x')$.
3. $\lim_{h \downarrow 0} F_X(x+h) = F_X(x)$.
4. $\mathbb{P}(\{s \in S \mid X(s) > x\}) = 1 - F_X(x)$.
5. $\mathbb{P}(\{s \in S : x < X(s) \leq x'\}) = F_X(x') - F_X(x)$ *whenever* $x < x'$,
6. $\mathbb{P}(\{s \in S : X(s) = x\}) = F_X(x) - \lim_{h \uparrow 0} F_X(x+h)$.

Is it always true that $\lim_{h \uparrow 0} F_X(x+h) = F_X(x)$?

CDFs are sometimes said to be càdlàg,
which is a French acronym for *continue à droite, limite à gauche*
(continuous on the right, limit on the left).

- A random variable $X$ is
  1. continuous if $F_X$ is continuous;
  2. absolutely continuous if there exists a density $f_X : \mathbb{R} \to \mathbb{R}_+$ such that $F_X(x) = \int_{-\infty}^{x} f_X(u) \mathrm{d}u$
- The support of $X$ is the closure of the set on which $F_X$ is strictly increasing.

### Example

The uniform distribution on a closed interval $[a, b]$ of $\mathbb{R}$ has density function $f$ and distribution function $F$ given by

$$f_X(x) := \frac{1}{b-a} 1_{[a,b]}(x) \quad \text{and} \quad F_X(x) := \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } x \in [a, b] \\ 1 & \text{if } x > b \end{cases}$$

### Example

The standard normal distribution on $\mathbb{R}$
has density function $f$ given by

$$f_X(x) := ke^{-\frac{1}{2}x^2}$$

where $k := 1/\sqrt{2\pi}$ is chosen so that $\int_{-\infty}^{+\infty} ke^{-\frac{1}{2}x^2} dx = 1$.
Its mean and variance are

$$\begin{aligned}
\int_{-\infty}^{+\infty} kxe^{-\frac{1}{2}x^2} dx &= \lim_{a\to\infty} \int_{-a}^{+a} kxe^{-\frac{1}{2}x^2} dx \\
&= \lim_{a\to\infty} [-\int_0^a kxe^{-\frac{1}{2}x^2} dx + \int_0^a kxe^{-\frac{1}{2}x^2} dx] \\
&= 0 \\
\int_{-\infty}^{+\infty} kx^2 e^{-\frac{1}{2}x^2} dx &= 1
\end{aligned}$$

Define $I(a) := \int_{-a}^{+a} e^{-\frac{1}{2}x^2} dx$ for each $a \in \mathbb{R}$. Then

$$\begin{aligned}
[I(a)]^2 &= \left(\int_{-a}^{+a} e^{-\frac{1}{2}x^2} dx\right)\left(\int_{-a}^{+a} e^{-\frac{1}{2}y^2} dy\right) \\
&= \int_{-a}^{+a} \left(\int_{-a}^{+a} e^{-\frac{1}{2}y^2} dy\right) e^{-\frac{1}{2}x^2} dx \\
&= \int_{-a}^{+a} \int_{-a}^{+a} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}y^2} dxdy
\end{aligned}$$

Define also

$$\begin{aligned}
J(b) &:= 2\pi \int_0^b re^{-\frac{1}{2}r^2} dr \;=\; 2\pi|_0^b[-e^{-\frac{1}{2}r^2}] \\
&= 2\pi(1 - e^{-\frac{1}{2}b^2})
\end{aligned}$$

Let $S(a) := [-a, a]^2$ denote the solid square subset of $\mathbb{R}^2$ that is centred at the origin and has sides of length $2a$.

Let $D(b) := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq b^2\}$ denote the disk of radius $b$ centred at the origin.

Consider the transformation $(r, \theta) \mapsto (x, y) = (r \cos \theta, r \sin \theta)$ from polar to Cartesian coordinates.
The Jacobian matrix of this transformation is

$$\begin{vmatrix} \partial x/\partial r & \partial x/\partial \theta \\ \partial y/\partial r & \partial y/\partial \theta \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r(\cos^2 \theta + \sin^2 \theta) = r$$

It follows that changing to polar coordinates in the double integral $\int_{D(b)} e^{-\frac{1}{2}(x^2+y^2)} dxdy$ transforms it to

$$\int_0^b \int_0^{2\pi} r e^{-\frac{1}{2}r^2} dr d\theta = 2\pi \int_0^b r e^{-\frac{1}{2}r^2} dr = J(b)$$

Note that $D(a) \subset S(a) \subset D(a\sqrt{2})$ and so

$$\begin{aligned} \int_{D(a)} e^{-\frac{1}{2}(x^2+y^2)} dx\, dy &\leq \int_{S(a)} e^{-\frac{1}{2}(x^2+y^2)} dx\, dy \\ &\leq \int_{D(a\sqrt{2})} e^{-\frac{1}{2}(x^2+y^2)} dx\, dy \end{aligned}$$

One can show that

$$J(a) \leq [I(a)]^2 = \int_{S(a)} e^{-\frac{1}{2}(x^2+y^2)} dxdy \leq J(a\sqrt{2})$$

and so $2\pi(1 - e^{-\frac{1}{2}a^2}) \leq [I(a)]^2 \leq 2\pi(1 - e^{-a^2})$.
Taking limits as $a \to \infty$ one has $2\pi(1 - e^{-\frac{1}{2}a^2}) \to 2\pi$ and also $2\pi(1 - e^{-a^2}) \to 2\pi$, implying that $[I(a)]^2 \to 2\pi$.

### Theorem

*The Gaussian integral* $\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx$ *equals* $\sqrt{2\pi}$.

# Outline

## Expectation

Let $g : \mathbb{R} \to \mathbb{R}$ be any Borel function,
and $x \mapsto f_X(x)$ the density function of the random variable $X$.
Whenever the integral $\int_{-\infty}^{\infty} |g(x)| f_X(x) \mathrm{d}x$ exists,
the expectation of $g \circ X$ is defined as

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

### Theorem

*Let $g_1, g_2 : \mathbb{R} \to \mathbb{R}$ and $a, b, c \in \mathbb{R}$. Then:*

1. $\mathbb{E}(a g_1(X) + b g_2(X) + c) = a\mathbb{E}(g_1(X)) + b\mathbb{E}(g_2(X)) + c$.
2. *If* $g_1 \geq 0$, *then* $\mathbb{E}(g_1(X)) \geq 0$.
3. *If* $g_1 \geq g_2$, *then* $\mathbb{E}(g_1(X)) \geq \mathbb{E}(g_2(X))$.

: Fix $g : \mathbb{R} \to \mathbb{R}_+$ with $\mathbb{E}(g(X)) \in \mathbb{R}$.
Then $\mathbb{P}(g(X) \geq r) \leq \frac{1}{r}\mathbb{E}(g(X))$.

### Theorem

*For any random variable* $S \ni s \mapsto X(s) \in Z$*,*
*fix any measurable function* $g : Z \to \mathbb{R}_+$ *with* $\mathbb{E}[g(X(s))] < +\infty$*.*
*Then for all* $r > 0$ *one has* $\mathbb{P}(g(X) \geq r) \leq \frac{1}{r}\mathbb{E}[g(X)]$*.*

### Proof.

The two indicator functions $s \mapsto 1_{g(X) \geq r}(s)$ and $s \mapsto 1_{g(X) < r}(s)$ satisfy $1_{g(X) \geq r}(s) + 1_{g(X) < r}(s) = 1$ for all $s \in S$.

Because $g(X(s) \geq 0$ for all $s \in S$, one has

$$\begin{array}{rcl} \mathbb{E}[g(X)] & = & \mathbb{E}[1_{g(X) \geq r}(s)\, g(X(s))] + \mathbb{E}[1_{g(X) < r}(s)\, g(X(s))] \\ & \geq & r\, \mathbb{E}[1_{g(X) \geq r}(s)] \;=\; r\, \mathbb{P}(g(X) \geq r) \end{array}$$

Dividing by $r$ implies that $\frac{1}{r}\mathbb{E}[g(X)] \geq \mathbb{P}(g(X) \geq r)$. □

# Moments and Central Moments

For a random variable $X$:

- its $k^{th}$ (noncentral) moment is $\mathbb{E}(X^k)$;
- its $k^{th}$ central moment is $\mathbb{E}((X - \mathbb{E}(X))^k)$, assuming that $\mathbb{E}(X)$ exists in $\mathbb{R}$.
- its variance, $\mathsf{V}(x)$, is its second central moment.

# Central Moments of the Gaussian Distribution

Let $m_n := \int_{-\infty}^{+\infty} k x^n e^{-\frac{1}{2}x^2} dx$ denote the $n$th central moment of the standard normal distribution.

Because $\frac{d}{dx} e^{-\frac{1}{2}x^2} = -x\, e^{-\frac{1}{2}x^2}$, integrating by parts gives

$$
\begin{aligned}
m_n &= \int_{-\infty}^{+\infty} k x^n e^{-\frac{1}{2}x^2} dx \\
&= -\int_{-\infty}^{+\infty} k x^{n-1} \left( \frac{d}{dx} e^{-\frac{1}{2}x^2} \right) dx \\
&= -|_{-\infty}^{+\infty} k x^{n-1} e^{-\frac{1}{2}x^2} + \int_{-\infty}^{+\infty} k\, n - 1 x^{n-2} e^{-\frac{1}{2}x^2} dx \\
&= (n-1) m_{n-2}
\end{aligned}
$$

So $m_{2r-1} = 0$ for odd integers, whereas for even integers

$$
\begin{aligned}
m_{2r} &= (2r-1)(2r-3) \cdots 5 \cdot 3 \cdot 1 \\
&= \frac{2r(2r-1)(2r-2)(2r-3) \cdots 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2r(2r-2)(2r-4) \cdots 6 \cdot 4 \cdot 2} = \frac{(2r)!}{2^r r!}
\end{aligned}
$$

Let $S \ni s \mapsto \mathbf{X}(s) = (X_n(s))_{n=1}^N$
be an $N$-dimensional vector of random variables
defined on the probability space $(S, \Sigma, \mathbb{P})$.

- Its joint distribution function is the mapping defined by

$$\mathbb{R}^N \ni \mathbf{x} \mapsto F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\{s \in S \mid \mathbf{X}(s) \leqq \mathbf{x}\})$$

- The random vector $\mathbf{X}$ is absolutely continuous
  just in case there exists a density function $f_{\mathbf{X}} : \mathbb{R}^N \to \mathbb{R}_+$
  such that

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbf{u} \leqq \mathbf{x}} f_{\mathbf{X}}(\mathbf{u}) \, d\mathbf{u}$$

Let $\mathbf{X}$ be an $N$-dimensional vector valued random variable.

- If $\mathbf{X}$ is absolutely continuous,
  the marginal density $\mathbb{R} \ni x \mapsto f_{X_n}(x)$ of its $n$th component $X_n$
  is defined as the $N-1$-dimensional iterated integral

$$f_{X_n}(x) = \int \cdots \int f_{\mathbf{X}}(x_1, \ldots, x_{n-1}, x, x_{n+1}, \ldots, x_N) \mathrm{d}x_1 \ldots \mathrm{d}x_N$$

- The $N$ components of $\mathbf{X}$ are independent just in case
  $f_{\mathbf{X}} = \prod_{n=1}^{N} f_{X_n}$.
- The infinite sequence $(X_n)_{n=1}^{\infty}$ of random variables
  is independent just in case
  every finite subsequence $(X_n)_{n \in K}$ ($K$ finite) is independent.

# Expectations

Let $\mathbf{X}$ be an $N$-dimensional vector valued random variable, and $g : \mathbb{R}^N \to \mathbb{R}$ a measurable function.

The expectation of $g(\mathbf{X})$ is defined as the $N$-dimensional integral

$$\mathbb{E}[g(\mathbf{X})] := \int_{\mathbb{R}^N} g(\mathbf{u}) f_{\mathbf{X}}(\mathbf{u})\, d\mathbf{u}$$

## Theorem

*If the collection $(X_n)_{n=1}^N$ of random variables is independent, then*

$$\mathbb{E}\left[\prod_{n=1}^N X_n\right] = \prod_{n=1}^N \mathbb{E}(X_n)$$

## Exercise

*Prove that if the pair $(X_1, X_2)$ of r.v.s is independent, then its covariance satisfies*

$$\mathrm{Cov}(X_1, X_2) := \mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] = 0$$

## Marginal and Conditional Density

Fix the pair $(X_1, X_2)$ of random variables.

- The marginal density of $X_1$ is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x_1, x_2) \mathrm{d}x_2.$$

- At points $x_1$ where $f_{X_1}(x_1) > 0$,
  the conditional density of $X_2$ given that $X_1 = x_1$ is

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{(X_1,X_2)}(x_1, x_2)}{f_{X_1}(x_1)}$$

### Theorem

*If the pair $(X_1, X_2)$ is independent and $f_{X_1}(x_1) > 0$, then*

$$f_{X_2|X_1}(x_2|x_1) = f_{X_2}(x_2)$$

# Conditional Expectations

Fix the pair $(X_1, X_2)$ of random variables.

- The conditional expectation of $g(X_2)$ given that $X_1 = x_1$ is

$$\mathbb{E}[g(X_2)|X_1 = x_1] = \int_{-\infty}^{\infty} g(x_2) f_{X_2|X_1}(x_2|x_1) \mathrm{d}x_2.$$

- Given any measurable function $(x_1, x_2) \mapsto g(x_1, x_2)$, the law of iterated expectations states that

$$\mathbb{E}_{f_{(X_1,X_2)}}[g((X_1, X_2)(s))] = \mathbb{E}_{f_{X_1}}[\mathbb{E}_{f_{X_2|X_1}}[g((X_1, X_2)(s))]]$$

Proof.

$$\begin{aligned} \mathbb{E}_{f_{(X_1,X_2)}}[g] &= \int_{\mathbb{R}^2} g(x_1, x_2)\ f_{(X_1,X_2)}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} g(x_1, x_2)\ f_{X_2|X_1}(x_2|x_1)\, \mathrm{d}x_2\right]\ f_{X_1}(x_1)\, \mathrm{d}x_1 \\ &= \mathbb{E}_{f_{X_1}}[\mathbb{E}_{f_{X_2|X_1}} g(x_1, x_2)]] \end{aligned}$$

□

# Outline

The sequence $(X_n)_{n=1}^{\infty}$ of random variables:

- converges in probability to $X$ (written as $X_n \xrightarrow{p} X$) just in case, for all $\epsilon > 0$ one has

$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| < \epsilon) = 1.$$

- converges in distribution to $X$ (written as $X_n \xrightarrow{d} X$) just in case, for all $x$ at which $F_X$ is continuous,

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x).$$

# The Law of Large Numbers

- The sequence $(X_n)_{n=1}^\infty$ of random variables is i.i.d.
  — i.e., independently and identically distributed
  — just in case
  1. it is independent, and
  2. for every Borel set $D \subseteq \mathbb{R}$, one has $\mathbb{P}(X_n \in D) = \mathbb{P}(X_{n'} \in D)$.
- The weak law of large numbers:
  Let $(X_n)_{n=1}^\infty$ be i.i.d. with $\mathbb{E}(X_n) = \mu$.
  Define the sequence

$$(\bar{X}_n)_{n=1}^\infty := \left( \frac{1}{n} \sum\nolimits_{k=1}^{n} X_k \right)_{n=1}^\infty$$

  of sample means. Then, $\bar{X}_n \xrightarrow{p} \mu$.

# The Meaning of Probability

Prove the following:

Let $\gamma = p(X \in \Omega) \in (0, 1)$. Consider the following experiment:
"$n$ realizations of $X$ are taken independently."
Let $G_n$ be the relative frequency with which a realization in $\Omega$ is obtained in the experiment. Then, $G_n \xrightarrow{p} \gamma$.

# The Central Limit Theorem

- The central limit theorem:
  Let $(X_n)_{n=1}^{\infty}$ be i.i.d. random variables
  with $\mathbb{E}(X_n) = \mu$ and $\mathsf{V}(X_n) = \sigma^2$. Then,

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} \mathrm{d}u.$$

Let $(X_n)_{n=1}^{\infty}$ be i.i.d., with $\mathbb{E}[X_n] = \mu$ and $\mathbf{V}(X_n) = \sigma^2$. Then:

- by the law of large numbers,

$$\bar{X}_n \xrightarrow{p} \mu;$$

  so

$$\bar{X}_n \xrightarrow{d} \mu;$$

- but by the central limit theorem,

$$\frac{\bar{X}_n - \mu}{(\sigma/\sqrt{n})} \xrightarrow{d} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} \mathrm{d}u.$$