

# Lecture Notes 9: Measure and Probability

## Part C: Probability as Measure

Peter Hammond  
(loosely based on early notes by Andrés Carvajal)

latest revision 2025 September 29th,  
typeset from `measProbC25.tex`

# Outline

Kolmogorov's Definition of Probability

Random Variables and Their Distribution and Density Functions

Expected Values

Joint Probability Distributions

Limit Theorems

Convergence Results

Non-integrability for Macroeconomists

A Continuum of Independent Random Variables

# Probability Measure

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  $\mathbb{P} : \Sigma \rightarrow \bar{\mathbb{R}}_+$   
satisfying the additional requirement that  $\mathbb{P}(S) = 1$ .

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

For all  $E \in \Sigma$ , because  $\mathbb{P}(E) \geq 0$ , it follows that  $\mathbb{P}(E) \in [0, 1]$ .

# Probability Space

Following Kolmogorov (1933), a **probability space** is a triple  $(S, \Sigma, \mathbb{P})$  where:

1.  $S$  is the state space;
2.  $\Sigma$  is a  $\sigma$ -algebra of measurable **events**, making  $(S, \Sigma)$  a measurable space;
3.  $\Sigma \ni E \mapsto \mathbb{P}(E) \in [0, 1]$  is a **probability measure** on  $(S, \Sigma)$ ,

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

## Proof.

We leave the routine proof as an exercise. □

# Trivial and Minimal Probability Spaces

## Exercise

Given any non-empty set  $S$ , show that the triple  $(S, \{\emptyset, S\}, \mathbb{P})$  in which  $\Sigma$  consists of only two sets is a probability space, called the **trivial** probability space, just in case  $\mathbb{P}(\emptyset) = 0$  and  $\mathbb{P}(S) = 1$ .

## Exercise

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

Given any event  $E \in \Sigma$  with  $\emptyset \subsetneq E \subsetneq S$ , show that the  $\sigma$ -algebra  $\sigma(\{E\})$  generated by  $\{E\}$  is the Boolean algebra  $\{\emptyset, E, S \setminus E, S\}$ .

Show too that  $(S, \sigma(\{E\}), \mathbb{P}_E)$  is a probability space, called a **minimal non-trivial** probability space, provided that, for any  $\mathbb{P}_E(E) \in [0, 1]$ , we take:

$$\mathbb{P}_E(\emptyset) = 0, \quad \mathbb{P}_E(S \setminus E) = 1 - \mathbb{P}_E(E), \quad \text{and} \quad \mathbb{P}_E(S) = 1$$

# Two Limiting Properties

## Theorem

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

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

# Proving the Two Limiting Properties

Proof.

1. Because  $E_n \subseteq E_{n+1}$  for all  $n \in \mathbb{N}$ , it follows that

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

Note that  $E_1$  and the sets  $E_k \setminus E_{k-1}$  for  $k = 2, 3, \dots$  are all pairwise disjoint.

Because probabilities are non-negative, additivity and countable additivity imply that

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

2. Apply part 1 to the complements  $S \setminus E_n$  of the sets  $E_n$ . □

## Conditional Probability and an Extension

Let  $(S, \Sigma, \mathbb{P})$  be any probability space with  $\sigma$ -algebra  $\Sigma \subseteq 2^S$ .

Let  $E^* \in \Sigma$  be any measurable event satisfying  $\mathbb{P}(E^*) > 0$ .

### Exercise

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

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

Prove that:

1.  $\Sigma(E^*)$  is a  $\sigma$ -algebra;
2. the mapping  $E \mapsto \mathbb{P}(E|E^*)$  is a probability measure, called the **conditional probability measure** given the event  $E^*$ , defined on the measurable space  $(S, \Sigma(E^*))$ ;
3. the mapping  $\Sigma \ni E \mapsto \mathbb{P}(E|E^*) := \frac{\mathbb{P}(E \cap E^*)}{\mathbb{P}(E^*)} \in [0, 1]$  is an **extended conditional probability measure** given  $E^*$ , defined on the whole of  $\Sigma$  while satisfying  $\mathbb{P}(E^*|E^*) = 1$  as well as  $\mathbb{P}(E \setminus E^*|E^*) = 0$  for all  $E \in \Sigma$ .

# The Conditional Probability Space and Bayes' Rule

## Definition

Let  $(S, \Sigma, \mathbb{P})$  be any probability space with  $\sigma$ -algebra  $\Sigma \subseteq 2^S$ .

Let  $E^* \in \Sigma$  be any measurable event satisfying  $\mathbb{P}(E^*) > 0$ .

Then the **conditional probability space given  $E^*$**  is the triple  $(E^*, \Sigma(E^*), \mathbb{P}[\cdot|E^*])$  where:

1. the **conditional probability  $\sigma$ -algebra given  $E^*$**  is  $\Sigma(E^*) := \{E \in \Sigma \mid E \subseteq E^*\}$
2. the **conditional probability measure  $\mathbb{P}(\cdot|E^*)$  given  $\Sigma(E^*)$**  is the mapping  $\Sigma(E^*) \ni E \mapsto \mathbb{P}(E|E^*)$

whose value is given by **Bayes' rule**  $\mathbb{P}(E|E^*) = \frac{\mathbb{P}(E \cap E^*)}{\mathbb{P}(E^*)}$ .

# Conditional Probability: The Law of Total Probability

## Proposition

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

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

## Proof.

The extended definition of conditional probability implies that

$$\begin{aligned} & \mathbb{P}(E)\mathbb{P}(E'|E) + (1 - \mathbb{P}(E))\mathbb{P}(E'|E^c) \\ &= \mathbb{P}(E)\frac{\mathbb{P}(E' \cap E)}{\mathbb{P}(E)} + \mathbb{P}(E^c)\frac{\mathbb{P}(E' \cap E^c)}{\mathbb{P}(E^c)} \\ &= \mathbb{P}(E' \cap E) + \mathbb{P}(E' \cap E^c) \end{aligned}$$

But  $(E' \cap E) \cap (E' \cap E^c) = \emptyset$  and  $(E' \cap E) \cup (E' \cap E^c) = E'$ ,  
so  $\mathbb{P}(E' \cap E) + \mathbb{P}(E' \cap E^c) = \mathbb{P}(E')$ . □

# Conditional Probability: Multiplicative Rule

## Proposition

Let  $(E_k)_{k=1}^n$  be any finite list of events in the probability space  $(S, \Sigma, \mathbb{P})$ .

Provided that  $\mathbb{P}(\cap_{k=1}^{n-1} E_k) > 0$ , one has

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

## Proof.

By induction,  
using the extended definition of conditional probability.

Details are left as an exercise. □

# Independent Events

The finite or countably infinite family  $\{E_k\}_{k \in K}$  of events in the probability space  $(S, \Sigma, \mathbb{P})$  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  $\mathbb{P}(\cap_{k=1}^n E_k) = \prod_{k=1}^n \mathbb{P}(E_k)$ .

# Pairwise Independence Does Not Imply Independence

## Example

Consider the probability space  $(S, 2^S, \mathbb{P})$  where  $S = \mathbb{N}_9$  and  $\mathbb{P}(\{s\}) = 1/9$  for all  $s \in S$ .

Consider the three events

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

which all have probability  $\frac{1}{3}$ .

Note that for each pair  $i, j \in \mathbb{N}_3$  with  $i \neq j$  one has  $E_i \cap E_j = \{7\}$  and so  $\mathbb{P}(E_i \cap E_j) = \mathbb{P}(\{7\}) = \frac{1}{9} = \mathbb{P}(E_i)\mathbb{P}(E_j)$ .

Thus, the three events are pairwise independent.

Yet  $E_1 \cap E_2 \cap E_3 = \{7\}$

so  $\mathbb{P}(E_1 \cap E_2 \cap E_3) = \frac{1}{9} \neq \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3) = \left(\frac{1}{3}\right)^3 = \frac{1}{27}$ ,

implying that the three events are not independent.

# Implications of Independence

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

## Notation

Given any  $E \subseteq S$ , let  $E^c := S \setminus E$  denote the *complementary event*.

## Exercise

Show that, if the two events  $E$  and  $\tilde{E}$  in  $\Sigma$  are independent, then:

1. the pairs  $\{E^c, \tilde{E}\}$  and  $\{E, \tilde{E}^c\}$  are both independent;
2. provided that  $\mathbb{P}(E)$  and  $\mathbb{P}(\tilde{E})$  are both positive, the conditional probabilities satisfy:
  - ▶  $\mathbb{P}(E|\tilde{E}) = \mathbb{P}(E \cap \tilde{E})/\mathbb{P}(\tilde{E}) = \mathbb{P}(E)$ , independent of  $\tilde{E}$ ;
  - ▶  $\mathbb{P}(\tilde{E}|E) = \mathbb{P}(E \cap \tilde{E})/\mathbb{P}(E) = \mathbb{P}(\tilde{E})$ , independent of  $E$ .

# The Measurable Product Space

## Definition

Let  $\langle (S_k, \Sigma_k) \rangle_{k=1}^n$  be a finite list of  $n$  measurable spaces.

Then the measurable space  $(S, \Sigma)$

is the **product** of these  $n$  measurable spaces just in case:

1. the state space  $S$  is the Cartesian product  $\prod_{k=1}^n S_k$  of the individual state spaces;
2. the  $\sigma$ -algebra  $\Sigma$  on  $S = \prod_{k=1}^n S_k$  is the **measurable product**  $\bigotimes_{k=1}^n \Sigma_k$  of the individual  $\sigma$ -algebras, defined as the  $\sigma$ -algebra  $\sigma(\prod_{k=1}^n \Sigma_k)$  generated by all **measurable rectangles**  $\prod_{k=1}^n E_k$  satisfying  $E_k \in \Sigma_k$  for all  $k \in \mathbb{N}_n$ .

Then  $(\prod_{k=1}^n S_k, \bigotimes_{k=1}^n \Sigma_k)$  is the **measurable product** of the list  $\langle (S_k, \Sigma_k) \rangle_{k=1}^n$  of measurable spaces.

## Joint and Marginal Measures

Consider any **joint** probability measure  $\mathbb{P}$  defined on the measurable product space  $(\prod_{k=1}^n S_k, \otimes_{k=1}^n \Sigma_k)$ .

Then, for each  $k \in \mathbb{N}_n$ , there is a corresponding **marginal** probability space  $(S_k, \Sigma_k, \mathbb{P}_k)$  in which the probability of each event  $E_k \in \Sigma_k$  is given by

$$\begin{aligned}\mathbb{P}_k(E_k) &= \mathbb{P}(S_1 \times \dots \times S_{k-1} \times E_k \times S_{k+1} \times \dots \times S_n) \\ &= \mathbb{P}\left(E_k \times \prod_{j \in \mathbb{N}_n \setminus \{k\}} S_j\right) = [\text{marg}_{S_k} \mathbb{P}](E_k)\end{aligned}$$

### Exercise

*Verify that  $\text{marg}_{S_k}$  is a probability measure on the measurable space  $(S_k, \Sigma_k)$ .*

# The Product of a Finite List of Probability Spaces

## Definition

Let  $\langle (S_k, \Sigma_k, \mathbb{P}_k) \rangle_{k=1}^n$  be a finite list of  $n$  probability spaces.

Then the probability space  $(S, \Sigma, \mathbb{P})$

is the **product** of these  $n$  probability spaces just in case:

1. the measurable space  $(S, \Sigma)$   
is the measurable product  $(\prod_{k=1}^n S_k, \otimes_{k=1}^n \Sigma_k)$   
of the  $n$  measurable spaces  $(S_k, \Sigma_k)$ ;
2. the probability measure  $\mathbb{P}$  is the product measure  $\otimes_{k=1}^n \mathbb{P}_k$ ,  
defined as the unique extension  
to the product  $\sigma$ -algebra  $\otimes_{k=1}^n \Sigma_k$  of the function that,  
for each product  $\prod_{k=1}^n E_k$  of measurable rectangles,  
satisfies  $\otimes_{k=1}^n \mathbb{P}_k (\prod_{k=1}^n E_k) = \prod_{k=1}^n \mathbb{P}_k(E_k)$ .

## Independence for Marginal Probabilities

Given the product probability space  $(\prod_{k=1}^n S_k, \otimes_{k=1}^n \Sigma_k, \otimes_{k=1}^n \mathbb{P}_k)$ , for any  $k \in \mathbb{N}_n$  and any event  $E_k \in \Sigma_k$ , there is a corresponding product measurable **marginal event**  $\prod_{i=1}^{k-1} S_i \times E_k \times \prod_{j=k+1}^n S_j$  whose probability is  $\mathbb{P}_k(E_k)$ .

Then  $\mathbb{P}_k(E_k)$  is the **marginal probability**  $[\text{marg}_{S_k} \mathbb{P}](E_k)$ .

The above definitions imply that,

whenever  $k, \ell \in \mathbb{N}_n$  with  $k < \ell$  and also  $E_k \in \Sigma_k$ ,  $E_\ell \in \Sigma_\ell$ , then the two marginal events  $\prod_{i=1}^{k-1} S_i \times E_k \times \prod_{j=k+1}^n S_j$  and  $\prod_{i=1}^{\ell-1} S_i \times E_\ell \times \prod_{j=\ell+1}^n S_j$  are independent, because their intersection

$$\prod_{h=1}^{k-1} S_h \times E_k \times \prod_{i=k+1}^{\ell-1} S_i \times E_\ell \times \prod_{j=\ell+1}^n S_j$$

has probability  $\mathbb{P}_k(E_k) \mathbb{P}_\ell(E_\ell)$ .

# The Product of a Sequence of Probability Spaces

## Definition

Let  $\langle (S_k, \Sigma_k) \rangle_{k \in \mathbb{N}}$  be an infinite sequence of probability spaces.

Then the measurable space  $(S, \Sigma, \mathbb{P})$  is the **product**

$$\left( \prod_{k \in \mathbb{N}} S_k, \bigotimes_{k \in \mathbb{N}} \Sigma_k, \bigotimes_{k \in \mathbb{N}} \mathbb{P}_k \right)$$

of all these probability spaces just in case:

1. the state space  $S$  is the Cartesian product  $\prod_{k \in \mathbb{N}} S_k$  of all the individual state spaces;
2. the  $\sigma$ -algebra  $\Sigma$  on  $S = \prod_{k \in \mathbb{N}} S_k$  is the  $\sigma$ -algebra  $\sigma(\cup_{n \in \mathbb{N}} \bigotimes_{k=1}^n \Sigma_k)$  generated by the union of all the finite product  $\sigma$ -algebras  $\bigotimes_{k=1}^n \Sigma_k$ ;
3. the probability  $\mathbb{P}$  is the **product probability measure**  $\bigotimes_{k \in \mathbb{N}} \mathbb{P}_k$ , defined as the unique measure that, for each infinite product  $\prod_{k \in \mathbb{N}} E_k$  of measurable rectangles, satisfies  $\bigotimes_{k \in \mathbb{N}} \mathbb{P}_k \left( \prod_{k \in \mathbb{N}} E_k \right) = \inf_{n \in \mathbb{N}} \bigotimes_{k=1}^n \mathbb{P}_k \left( \prod_{k=1}^n E_k \right)$ .

# Outline

Kolmogorov's Definition of Probability

Random Variables and Their Distribution and Density Functions

Expected Values

Joint Probability Distributions

Limit Theorems

Convergence Results

Non-integrability for Macroeconomists

A Continuum of Independent Random Variables

# Random Variables

## Definition

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

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

$$x \mapsto F_X(x) = \mathbb{P}(\{s \in S \mid X(s) \leq x\}) = \mathbb{P}(X^{-1}((-\infty, x]))$$

# Properties of Distribution Functions, I

## Theorem

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

1.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  and  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .
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.  $\mathbb{P}(\{s \in S : X(s) = x\}) = F_X(x) - \lim_{h \uparrow 0} F_X(x + h)$ .

Because of Properties 3 and 6 in particular, 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).

# Properties of Distribution Functions, II

## Definition

A **continuity point** of the CDF  $F_X : \mathbb{R} \rightarrow [0, 1]$  is an  $\bar{x} \in \mathbb{R}$  at which the mapping  $x \mapsto F_X(x)$  is continuous.

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

## Exercise

Let  $F_X : \mathbb{R} \rightarrow [0, 1]$  be the CDF of any random variable  $S \ni s \mapsto X(s) \rightarrow \mathbb{R}$ , and  $\bar{x} \in \mathbb{R}$  any point.

*Prove that the following three conditions are equivalent:*

1.  $\bar{x}$  is a continuity point of  $F_X$ ;
2.  $\mathbb{P}(\{s \in S \mid X(s) = \bar{x}\}) = 0$ ;
3.  $\lim_{h \uparrow 0} F_X(\bar{x} + h) = F_X(\bar{x})$ .

# Continuous Random Variables

## Definition

- ▶ A random variable  $S \ni s \mapsto X(s) \rightarrow \mathbb{R}$  is
  1. **continuously distributed** just in case  $x \mapsto F_X(x)$  is continuous;
  2. **absolutely continuous** just in case there exists a **density function**  $\mathbb{R} \ni x \mapsto f_X(x) \rightarrow \mathbb{R}_+$  such that  $F_X(x) = \int_{-\infty}^x f_X(u) du$  for all  $x \in \mathbb{R}$ .
- ▶ The **support** of the random variable  $S \ni s \mapsto X(s) \rightarrow \mathbb{R}$  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

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

# The Standard Normal or Gaussian Distribution

## Example

The **standard normal distribution** on  $\mathbb{R}$

has density function  $f$  given by  $f_X(x) := ke^{-\frac{1}{2}x^2}$

where the normalizing constant  $k$  must be chosen

so that  $\int_{-\infty}^{+\infty} ke^{-\frac{1}{2}x^2} dx = 1$ .

Make the substitution  $y = x/\sqrt{2}$ ,

implying that  $y^2 = \frac{1}{2}x^2$  and  $dx = \sqrt{2} dy$ .

Using the rule for integration by substitution, for each  $b \in \mathbb{R}$

one has  $\int_{-b}^{+b} ke^{-\frac{1}{2}x^2} dx = \int_{-b/\sqrt{2}}^{+b/\sqrt{2}} k\sqrt{2}e^{-y^2} dy$ .

Taking limits as  $b \rightarrow \infty$ , we see that  $\int_{-\infty}^{+\infty} ke^{-\frac{1}{2}x^2} dx = 1$

only if  $\int_{-\infty}^{+\infty} k\sqrt{2}e^{-y^2} dy = 1$ .

But the Gaussian integral is  $\int_{-\infty}^{+\infty} e^{-y^2} dy = \sqrt{\pi}$  and so  $k\sqrt{2\pi} = 1$

implying that  $k = 1/\sqrt{2\pi}$  and so  $f_X(x) := (1/\sqrt{2\pi})e^{-\frac{1}{2}x^2}$ .

# Outline

Kolmogorov's Definition of Probability

Random Variables and Their Distribution and Density Functions

Expected Values

Joint Probability Distributions

Limit Theorems

Convergence Results

Non-integrability for Macroeconomists

A Continuum of Independent Random Variables

## Expectation w.r.t. a Probability Measure

Given the probability space  $(S, \Sigma, \mathbb{P})$ ,  
consider the  $\Sigma$ -measurable random variable  $S \ni s \mapsto X(s) \in \mathbb{R}$ .

Provided that  $S \ni s \mapsto |X(s)| \in \mathbb{R}_+$  is integrable,  
with  $\int_S |X(s)| \mathbb{P}(ds) < +\infty$ ,  
we can define the **expectation** or **expected value**  
of the random variable  $S \ni s \mapsto X(s) \in \mathbb{R}$   
as the Lebesgue integral  $\int_S X(s) \mathbb{P}(ds)$ .

## Expectation w.r.t. a Density Function

Let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be any Borel measurable function, and  $x \mapsto f_X(x)$  the density function of the random variable  $X$ .

Whenever the integral  $\int_{-\infty}^{\infty} |g(x)|f_X(x)dx$  exists, the **expectation** of  $g \circ X$  is defined as

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

### Theorem

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

1.  $\mathbb{E}(ag_1(X) + bg_2(X) + c) = a\mathbb{E}(g_1(X)) + b\mathbb{E}(g_2(X)) + c$ .
2. If  $g_1 \geq 0$ , then  $\mathbb{E}(g_1(X)) \geq 0$ .
3. If  $g_1 \geq g_2$ , then  $\mathbb{E}(g_1(X)) \geq \mathbb{E}(g_2(X))$ .

# Chebychev's Inequality: Statement

## Theorem

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

## Chebyshev's Inequality: Proof

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

$$\begin{aligned}\mathbb{E}[g(X)] &= \mathbb{E}[\{1_{g(X) \geq r}(s) + 1_{g(X) < r}(s)\} g(X(s))]\ \\ &= \mathbb{E}[1_{g(X) \geq r}(s) g(X(s))] + \mathbb{E}[1_{g(X) < r}(s) g(X(s))]\ \\ &\geq \mathbb{E}[1_{g(X) \geq r}(s) g(X(s))] \quad \text{because } g(X(s)) \geq 0 \\ &\geq r \mathbb{E}[1_{g(X) \geq r}(s)] \quad \text{when } g(X(s)) \geq r \\ &= r \mathbb{P}(g(X) \geq r)\end{aligned}$$

Dividing by  $r$ , which is positive,

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

# Moments and Central Moments

For a random variable  $X$  and any  $k \in \mathbb{N}$ :

- ▶ its  **$k$ th (noncentral) moment** is  $\mathbb{E}[X^k]$   
(where  $X^k(s)$  denotes the  $k$ th power of the random variable  $X(s)$ );
- ▶ its  **$k$ th central moment** is  $\mathbb{E}[(X - \mathbb{E}[X])^k]$ ,  
assuming that  $\mathbb{E}[X]$  exists in  $\mathbb{R}$ ;
- ▶ its **variance**,  $\text{Var } X$ , is its second central moment.

## Odd Central Moments of the Gaussian Distribution

Given any  $n \in \mathbb{N}$  and any  $a > 0$ ,  
define  $m_n(a) := \int_{-a}^{+a} \frac{1}{\sqrt{2\pi}} x^n e^{-\frac{1}{2}x^2} dx$ .

When  $n$  is odd, one has  $(-x)^n = -x^n$ , so

$$\begin{aligned} m_n(a) &= \int_{-a}^{+a} \frac{1}{\sqrt{2\pi}} x^n e^{-\frac{1}{2}x^2} dx \\ &= \int_{-a}^0 \frac{1}{\sqrt{2\pi}} x^n e^{-\frac{1}{2}x^2} dx + \int_0^{+a} \frac{1}{\sqrt{2\pi}} x^n e^{-\frac{1}{2}x^2} dx \\ &= - \int_0^{+a} \frac{1}{\sqrt{2\pi}} x^n e^{-\frac{1}{2}x^2} dx + \int_0^{+a} \frac{1}{\sqrt{2\pi}} x^n e^{-\frac{1}{2}x^2} dx \\ &= 0 \end{aligned}$$

This allows us to define  $m_n := \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^n e^{-\frac{1}{2}x^2} dx$   
as the  $n$ th central moment of the standard Gaussian distribution,  
and to assert that  $m_n = 0$  when  $n$  is odd.

## Even Central Moments of the Gaussian Distribution, I

Now suppose  $n = 2r$ , where  $r \in \mathbb{N}$ .

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

$$\begin{aligned} & \int_{-a}^{+a} \frac{1}{\sqrt{2\pi}} x^n e^{-\frac{1}{2}x^2} dx \\ &= \int_{-a}^{+a} \frac{1}{\sqrt{2\pi}} x^{n-1} \left( x e^{-\frac{1}{2}x^2} \right) dx \\ &= \int_{-a}^{+a} \frac{1}{\sqrt{2\pi}} x^{n-1} \left( -\frac{d}{dx} e^{-\frac{1}{2}x^2} \right) dx \\ &= - \left|_{-a}^{+a} \frac{1}{\sqrt{2\pi}} x^{n-1} e^{-\frac{1}{2}x^2} + \int_{-a}^{+a} \frac{1}{\sqrt{2\pi}} (n-1) x^{n-2} e^{-\frac{1}{2}x^2} dx \right. \\ &= - \frac{1}{\sqrt{2\pi}} [a^{n-1} - (-a)^{n-1}] e^{-\frac{1}{2}a^2} + \int_{-a}^{+a} \frac{1}{\sqrt{2\pi}} (n-1) x^{n-2} e^{-\frac{1}{2}x^2} dx \end{aligned}$$

Taking the limit as  $a \rightarrow \infty$ , the first non-integral term tends to 0, so one obtains  $m_n = (n-1)m_{n-2}$ .

## Even Central Moments of the Gaussian Distribution, II

For each  $r \in \mathbb{N}$ , let  $m_{2r}$  denote the  $(2r)$ th central moment of a Gaussian Distribution.

We have shown that  $m_0 = 1$  and that  $m_{2r} = (2r - 1)m_{2r-2}$  for all  $r \in \mathbb{N}$ .

It follows that

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

# Outline

Kolmogorov's Definition of Probability

Random Variables and Their Distribution and Density Functions

Expected Values

**Joint Probability Distributions**

Limit Theorems

Convergence Results

Non-integrability for Macroeconomists

A Continuum of Independent Random Variables

## Bivariate Distribution

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

Suppose that  $S \ni s \mapsto (X(s), Y(s)) \in \mathbb{R}^2$   
is a pair of  $\Sigma$ -measurable functions.

The **bivariate probability distribution function** is the mapping defined by  $\mathbb{R}^2 \ni (x, y) \mapsto F_{X,Y}(x, y) \in [0, 1]$   
where  $F_{X,Y}(x, y) := \mathbb{P}(\{s \in S \mid X(s) \leq x \text{ and } Y(s) \leq y\})$ .

There are two separate **marginal** distributions  $x \mapsto F_X(x)$  and  $y \mapsto F_Y(y)$   
of the two random variables  $X(s)$  and  $Y(s)$  given by

$$F_X(x) := \mathbb{P}(\{s \in S \mid X(s) \leq x\}) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$$

$$F_Y(y) := \mathbb{P}(\{s \in S \mid Y(s) \leq y\}) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y)$$

# Multiple Random Variables

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

$$\mathbb{R}^N \ni \mathbf{x} \mapsto F_{\mathbf{X}}(\mathbf{x}) := \mathbb{P}(\{s \in S \mid \mathbf{X}(s) \leq \mathbf{x}\}) \in [0, 1]$$

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

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbf{u} \leq \mathbf{x}} f_{\mathbf{X}}(\mathbf{u}) \, d\mathbf{u} \quad \text{for all } \mathbf{x} \in \mathbb{R}^N$$

# Independent Random Variables

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

$$\int \cdots \int f_{\mathbf{X}}(x_1, \dots, x_{n-1}, x, x_{n+1}, \dots, x_N) dx_1 \dots dx_{n-1} dx_{n+1} dx_N$$

in which every random variable except  $X_n$  gets “integrated out”.

- ▶ The  $N$  components of  $\mathbf{X}$  are **independent** just in case:
  1. the joint density  $f_{\mathbf{X}}$  is the product  $\prod_{n=1}^N f_{X_n}$  of the marginal densities;
  2. the joint CDF  $F_{\mathbf{X}}$  is the product  $\prod_{n=1}^N F_{X_n}$  of the marginal CDFs.
- ▶ 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 of a Function of $N$ Random Variables

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

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

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

when this integral exists.

### 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*

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

# Zero Covariance Does Not Imply Independence

## Example

1. Consider a bivariate distribution  $\frac{1}{4}(\delta_{-1,0} + \delta_{0,-1} + \delta_{0,1} + \delta_{1,0})$  of the two random variables  $X$  and  $Y$  giving equal weight  $\frac{1}{4}$  to each of the set of four corners of a diamond shape in  $\mathbb{R}^2$ .
2. Both  $X$  and  $Y$  have marginal distributions on  $\mathbb{R}$  given by  $\frac{1}{4}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_1$ , with expectations  $\mathbb{E}X = \mathbb{E}Y = 0$ .
3. Note that the product  $XY$  has the degenerate distribution  $\delta_0$ , so the expectation  $\mathbb{E}[XY] = 0$ .
4. The covariance of  $X$  and  $Y$  is  $\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = 0$ .
5. The conditional distributions  $\mathbb{P}_{X|Y}$  of  $X$  given the three different possible values of  $Y$  are  $\mathbb{P}_{X|-1} = \mathbb{P}_{X|1} = \delta_0$  and  $\mathbb{P}_{X|0} = \frac{1}{2}(\delta_{-1} + \delta_1)$ , and similarly for  $\mathbb{P}_{Y|X}$  for each  $X \in \{-1, 0, 1\}$ .
6. Evidently  $X$  and  $Y$  are not independent.

## Marginal and Conditional Density

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

- ▶ The **marginal density** of  $X_1$  is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{(X_1, X_2)}(x_1, x_2) dx_2.$$

- ▶ 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

$$\mathbb{E}[g(X_2)|X_1 = x_1] = \int_{-\infty}^{\infty} g(x_2) f_{X_2|X_1}(x_2|x_1) dx_2.$$

- ▶ 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.

$$\begin{aligned}\mathbb{E}_{f_{(X_1, X_2)}}[g] &= \int_{\mathbb{R}^2} g(x_1, x_2) f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 \\ &= \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} g(x_1, x_2) f_{X_2|X_1}(x_2|x_1) dx_2 \right] f_{X_1}(x_1) dx_1 \\ &= \mathbb{E}_{f_{X_1}}[\mathbb{E}_{f_{X_2|X_1}}[g(x_1, x_2)]] \quad \square\end{aligned}$$

# Outline

Kolmogorov's Definition of Probability

Random Variables and Their Distribution and Density Functions

Expected Values

Joint Probability Distributions

**Limit Theorems**

Convergence Results

Non-integrability for Macroeconomists

A Continuum of Independent Random Variables

# Convergence of Random Variables

The sequence  $(X_n)_{n=1}^{\infty}$  of random variables on the probability space  $(S, \Sigma, \mathbb{P})$ :

- ▶ **converges in probability** to the random variable  $X$  on the probability space  $(S, \Sigma, \mathbb{P})$  (written as  $X_n \xrightarrow{p} X$ ) just in case, for all  $\epsilon > 0$ , one has

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

- ▶ **converges in distribution** to the random variable  $X$  on the probability space  $(S, \Sigma, \mathbb{P})$  (written as  $X_n \xrightarrow{d} X$ ) just in case, for all  $x$  at which  $F_X$  is continuous, one has

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

# Definition of Weak Convergence

## Definition

Let  $(X, \Sigma, \mathbb{P})$  be any probability space where  $(X, \mathcal{T})$  is a topological space with the property that  $\Sigma$  is the **Borel  $\sigma$ -algebra** generated by the open sets that constitute  $\mathcal{T}$ .

Then a **continuity set** of  $(X, \Sigma, \mathbb{P})$  is any set  $B \in \Sigma$  whose boundary  $\partial B$  satisfies  $\mathbb{P}(\partial B) = 0$ .

## Definition

Let  $(X, d)$  be a metric space with its Borel  $\sigma$ -algebra  $\Sigma$  — i.e., the  $\sigma$ -algebra generated by the open sets of  $(X, d)$ .

A sequence  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  of probability measures on the measurable space  $(X, \Sigma)$  **converges weakly** to the probability measure  $\mathbb{P}$ , written  $\mathbb{P}_n \Rightarrow \mathbb{P}$ , just in case  $\mathbb{P}_n(B) \rightarrow \mathbb{P}(B)$  as  $n \rightarrow \infty$  for any continuity set  $B$  of  $(X, \Sigma, \mathbb{P})$ .

# Portmanteau Theorem

## Theorem

Let  $\mathbb{P}$  and  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  be probability measures on the measurable space  $(X, \Sigma)$ .

Then  $\mathbb{P}_n \Rightarrow \mathbb{P}$  as  $n \rightarrow \infty$  if and only if:

1. for all bounded continuous functions  $f : X \rightarrow \mathbb{R}$ , one has:

$$\int_X f(x) \mathbb{P}_n(dx) \rightarrow \int_X f(x) \mathbb{P}(dx)$$

2.  $\limsup_{n \rightarrow \infty} \mathbb{P}_n(C) \leq \mathbb{P}(C)$  for every closed subset  $C \subset X$ ;
3.  $\liminf_{n \rightarrow \infty} \mathbb{P}_n(U) \geq \mathbb{P}(U)$  for every open set  $U \subset X$ .

## Remark

Often it is easiest to check the first condition involving integrals of bounded continuous functions.

## Convergence of Probabilities: Warning

The following example shows that it is not very sensible to say that the sequence of probability measures  $\mathbb{P}_n$  ( $n \in \mathbb{N}$ ) on a measurable space  $(X, \Sigma)$  converges to  $\mathbb{P}$  just in case  $\mathbb{P}_n(E) \rightarrow \mathbb{P}(E)$  for all  $E \in \Sigma$ , even when  $E$  is not a continuity set.

### Exercise

1. Suppose that for each  $n \in \mathbb{N}$  the probability measure  $\mathbb{P}_n$  on the Borel real line corresponds to the uniform distribution on the interval  $I_n := (-\frac{1}{n}, \frac{1}{n})$ .

Show that  $\mathbb{P}_n \Rightarrow \delta_0$ , even though  $\mathbb{P}_n(\{0\}) = 0$  for all  $n \in \mathbb{N}$ .

Verify that 0 is not a continuity point of  $\delta_0$ .

2. Show that if the sequence  $x^\mathbb{N} = \langle x_n \rangle_{n \in \mathbb{N}}$  in  $\mathbb{R}$  converges to  $\hat{x}$ , then  $\delta_{x_n} \Rightarrow \delta_{\hat{x}}$ .

# Convergence of Distribution Functions

## Theorem

Let  $F$  and  $(F_n)_{n \in \mathbb{N}}$  be cumulative distribution functions on  $\mathbb{R}$  with associated probability measures  $\mathbb{P}$  and  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  on the Lebesgue real line that satisfy

$$F(x) = \mathbb{P}((-\infty, x]) \quad \text{and} \quad F_n(x) = \mathbb{P}_n((-\infty, x]) \quad (n \in \mathbb{N})$$

on the measurable space  $(X, \Sigma)$ .

Then  $\mathbb{P}_n \Rightarrow \mathbb{P}$  if and only if  $F_n(x) \rightarrow F(x)$  for all  $x$  at which  $F$  is continuous.

# The Prokhorov Metric on Probability Measures

The space  $\mathcal{P}(X, \Sigma)$  of all probability measures on the Borel  $\sigma$ -algebra of a metric space  $(X, d)$  is **metrizable**.

This means that the space  $\mathcal{P}(X, \Sigma)$  of probability measures can be given a metric  $\rho$ , or a mapping

$$\mathcal{P}(X, \Sigma) \times \mathcal{P}(X, \Sigma) \ni (\mu, \nu) \mapsto \rho(\mu, \nu) \in \mathbb{R}_+$$

with the key property that  $\mu_n \Rightarrow \mu \iff \rho(\mu_n, \mu) \rightarrow 0$ .

A  $\rho$  that works is the **Prokhorov metric**, which we define below.

First, given any  $A \subseteq \mathcal{P}(X, \Sigma)$  and any  $\epsilon > 0$ , define the  $\epsilon$ -neighbourhood of  $A$  by

$$N_\epsilon(A) := \{x \in X \mid \exists y \in A : d(x, y) < \epsilon\} = \cup_{y \in A} B_\epsilon(y)$$

Then let

$$\rho(\mu, \nu) := \inf_{\epsilon} \{\epsilon > 0 \mid \forall A \subset \mathcal{P}(X, \Sigma) : \\ \mu(A) \leq \nu(N_\epsilon(A)) + \epsilon \text{ and } \nu(A) \leq \mu(N_\epsilon(A)) + \epsilon\}$$

# The Weak Law of Large Numbers

- ▶ The sequence  $(X_n)_{n=1}^{\infty}$  of random variables on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is **i.i.d.**
  - i.e., independently and identically distributed
  - just in case
    1. the random variables are independent;
    2. for every Borel set  $D \subseteq \mathbb{R}$  and every  $n \in \mathbb{N}$ , the probability  $\mathbb{P}(X_n \in D)$  does not depend on  $n$ .

- ▶ **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

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

of **sample means**.

Then  $\bar{X}_n \xrightarrow{P} \mu$  — i.e.,  $\bar{X}_n$  converges in probability to the common mean  $\mu$ .

## A “Frequentist” Interpretation of Probability

### Exercise

*Suppose an experimenter observes  $n$  independent realizations of the same random variable  $X$  on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .*

*Let  $E \in \mathcal{F}$  denote a fixed event.*

*Let  $G_n$  be the relative frequency with which the event  $E$  is observed in the experiment.*

*Let  $\gamma = \mathbb{P}(X \in \Omega) \in (0, 1)$ .*

*Prove that  $G_n \xrightarrow{P} \gamma$  as  $n \rightarrow \infty$ .*

# The Central Limit Theorem

► **The central limit theorem:**

Let  $(X_k)_{k=1}^{\infty}$  be an infinite sequence of i.i.d. random variables with common mean  $\mathbb{E}(X_k) = \mu$  and variance  $V(X_k) = \sigma^2$ .

For each  $n \in \mathbb{N}$ , define  $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$

as the sample average of  $n$  observations.

1. Then  $\mathbb{E}(\bar{X}_n) = \mu$  and  $V(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n V(X_k) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$ ;
2. Suppose that, for each  $n \in \mathbb{N}$ ,

the random variable  $Z_n := \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$

is **standardized** in the sense that  $\mathbb{E}[Z_n] = 0$  and  $\mathbb{E}[Z_n^2] = 1$ .

Then one has  $Z_n \xrightarrow{d} Y$  where  $Y$  has the **standard normal** cdf given by  $F_Y(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}u^2} du$  for all  $x \in \mathbb{R}$ .

In particular,  $\mathbb{E}(Y) = 0$  and  $\mathbb{E}(Y^2) = 1$ .

# The Fundamental Theorems

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, one has  $\bar{X}_n \xrightarrow{P} \mu$   
and so  $\bar{X}_n \xrightarrow{d} \mu$ ;
- ▶ but by the central limit theorem,

$$Z_n := \frac{\bar{X}_n - \mu}{(\sigma/\sqrt{n})} \xrightarrow{d} Z \text{ where } F_Z(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}u^2} du$$

## Example

In case each  $X_n$  is Gaussian, it can be shown that the linear combination  $Z_n$  is Gaussian.

But  $\mathbb{E}[Z_n] = 0$  and  $\mathbf{V}(Z_n) = 1$ ,  
so each  $Z_n$  is exactly Gaussian with mean 0 and variance 1.

# Concepts of Convergence, I

## Definition

Say that the sequence  $X_n$  of random variables converges **almost surely** or **with probability 1** (or **strongly**) towards  $X$  just in case, for every  $\epsilon > 0$ , one has

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| < \epsilon\}) = 1$$

Hence, the random values of  $X_n$  approach those of  $X$ , in the sense that the event that  $X_n(\omega)$  does not converge to  $X(\omega)$  has probability 0.

Almost sure convergence is often denoted by  $X_n \xrightarrow[P\text{-a.s.}]{} X$ , with “ $P$ -a.s.” under the arrow indicating  $\mathbb{P}$ -almost sure convergence.

Of course, the concept of almost sure convergence depends on the probability measure being used.

## Concepts of Convergence, II

For generic random elements  $X_n$  on a general metric space  $(S, d)$ , almost sure convergence is defined similarly, replacing the absolute value  $|X_n(\omega) - X(\omega)|$  by the distance  $d(X_n(\omega), X(\omega))$ .

Almost sure convergence implies convergence in probability, and *a fortiori* convergence in distribution.

It is the notion of convergence used in the strong law of large numbers.

# The Strong Law of Large Numbers (or SLLN)

## Definition

The **strong law of large numbers** (or SLLN) states that the sample average  $\bar{X}_n$  of a sequence of random variables on the sample space  $(S, \Sigma, \mathbb{P})$  converges  $\mathbb{P}$ -almost surely to the expected value  $\mu = \mathbb{E}X$ .

It is this law (rather than the weak LLN) that justifies the intuitive interpretation of the expected value of a random variable as its “long-term average when sampling repeatedly.”

# Differences Between the Weak and Strong Laws

The **weak** law states that for a specified large  $n$ , the average  $\bar{X}_n$  is likely to be near  $\mu$ .

This leaves open the possibility that  $|\bar{X}_n - \mu| \geq \epsilon$  happens an infinite number of times, although at increasingly infrequent intervals.

The **strong** law shows that this almost surely will not occur.

In particular, it implies that with probability 1, for any  $\epsilon > 0$  there exists  $n_\epsilon$  such that  $|\bar{X}_n - \mu| < \epsilon$  holds for all  $n > n_\epsilon$ .

# Outline

Kolmogorov's Definition of Probability

Random Variables and Their Distribution and Density Functions

Expected Values

Joint Probability Distributions

Limit Theorems

**Convergence Results**

Non-integrability for Macroeconomists

A Continuum of Independent Random Variables

# Moment-Generating Functions

## Definition

The  $n$ th **moment** about the origin is defined as  $m_n := \mathbb{E}[X^n]$ .

This may not exist for large  $n$  unless the random variable  $X$  is **essentially bounded** both above and below, meaning that there exists an upper bound  $\bar{x}$  on the modulus such that  $\mathbb{P}(\{\omega \in \Omega \mid |X(\omega)| \leq \bar{x}\}) = 1$ .

## Definition

The **moment-generating function** of a random variable  $X$  is

$$\mathbb{R} \ni t \mapsto M_X(t) := \mathbb{E}[e^{tX}]$$

wherever this expectation exists.

At  $t = 0$ , of course,  $M_X(0) = 1$ .

For  $t \neq 0$ , however, unless  $X$  is essentially bounded, the expectation may not exist because  $e^{tX}$  can be unbounded.

## The Gaussian Case

For a normal or Gaussian distribution  $N(\mu, \sigma^2)$ , even though the random variable is unbounded, the tails of the distribution are thin enough to ensure that the moment generating function exists and is given by

$$\begin{aligned}M(t; \mu, \sigma^2) &= \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{tx - (x-\mu)^2/2\sigma^2} dx\end{aligned}$$

Now make the substitution  $y = (x - \mu - \sigma^2 t)/\sigma$ , implying that  $dx = \sigma dy$  and that

$$tx - \frac{(x - \mu)^2}{2\sigma^2} = -\frac{1}{2}y^2 + \mu t + \frac{1}{2}\sigma^2 t^2$$

Because of the Gaussian integral, this transforms the integral to

$$M(t; \mu, \sigma^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} e^{\mu t + \frac{1}{2}\sigma^2 t^2} dy = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

# From Moment-Generating Functions to Moments

Note that

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \cdots + \frac{t^n X^n}{n!} + \cdots = \sum_{n=0}^{\infty} \frac{t^n X^n}{n!}$$

Taking the expectation term by term  
and then using the definition of the moments of the distribution,  
one obtains

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= 1 + t\mathbb{E}[X] + \frac{t^2}{2!}\mathbb{E}[X^2] + \cdots + \frac{t^n}{n!}\mathbb{E}[X^n] + \cdots \\ &= 1 + tm_1 + \frac{t^2}{2!}m_2 + \cdots + \frac{t^n}{n!}m_n + \cdots \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!}m_n \end{aligned}$$

## Derivatives of the Moment-Generating Function

Consider the  $n$ th derivative of  $M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} m_k$  with respect to  $t$ .

One can easily prove by induction on  $n$  that

$$\frac{d^n}{dt^n} t^k = k(k-1)(k-2)\dots(k-n+1)t^{k-n} = \frac{k!}{(k-n)!} t^{k-n}$$

So differentiating  $t \mapsto M_X(t)$  term by term  $n$  times, one obtains

$$\begin{aligned} M_X^{(n)}(t) &= \mathbb{E} \left[ \frac{d^n}{dt^n} e^{tX} \right] = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} \frac{t^{k-n}}{k!} m_k \\ &= \sum_{k=n}^{\infty} \frac{t^{k-n}}{(k-n)!} m_k \end{aligned}$$

Putting  $t = 0$  yields the equality  $M_X^{(n)}(0) = \frac{t^0}{0!} m_n = m_n$ .

In this sense, the derivatives of the moment-generating function do “generate exponentially” the moments of the probability distribution.

## Definition of Characteristic Functions

The moment-generating function may not exist because the expectation need not converge absolutely.

By contrast, the expectation of the bounded function  $e^{itX}$  always lies in the unit disc of the complex plane  $\mathbb{C}$ .

So the characteristic function that we are about to introduce always exists, which makes it more useful in many contexts.

### Definition

For a scalar random variable  $X$  with CDF  $x \mapsto F_X(x)$ , the **characteristic function** is defined as the (complex) expected value of  $e^{itX} = \cos tX + i \sin tX$ , where  $i = \sqrt{-1}$  is the imaginary unit, and  $t \in \mathbb{R}$  is the argument of the characteristic function:

$$\mathbb{R} \ni t \mapsto \phi_X(t) = \mathbb{E}e^{itx} = \int_{-\infty}^{+\infty} e^{itx} dF_X(x) \in \mathbb{C}$$

## Gaussian Case

Consider a normally distributed random variable  $X$  with mean  $\mu$  and variance  $\sigma^2$ .

Its characteristic function can be found by replacing  $t$  by  $it$  in the expression for the moment

$$M(t; \mu, \sigma^2) = \mathbb{E}[e^{tX}] = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Recalling that  $(it)^2 = -t^2$ , the result is

$$\varphi(t; \mu, \sigma^2) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$$

In the **standard normal** or  $N(0, 1)$  case, when  $\mu = 0$  and  $\sigma^2 = 1$ , one has  $\varphi(t; 0, 1) = e^{-\frac{1}{2}t^2}$ .

## Linear Combinations of Gaussian Random Variables

Suppose that the two independent random variables  $X$  and  $Y$  have respective means  $m_X, m_Y$  and variances  $v_X, v_Y$ .

Consider the linear combination  $Z := \alpha X + \beta Y$  where  $\alpha, \beta \in \mathbb{R}$ .

### Exercise

Show that  $Z$  has mean  $m_Z = \alpha m_X + \beta m_Y$  and variance  $v_Z = \alpha^2 v_X + \beta^2 v_Y$ .

### Proposition

If  $X$  and  $Y$  are both Gaussian, then so is  $Z$ .

### Proof.

Because  $X$  and  $Y$  are independent and Gaussian, the characteristic function  $\varphi_Z(t) = \mathbb{E}[e^{itZ}]$  of  $Z = \alpha X + \beta Y$  satisfies

$$\varphi_Z(t) = \mathbb{E}[e^{it(\alpha X + \beta Y)}] = \mathbb{E}[e^{i(\alpha t)X}] \mathbb{E}[e^{i(\beta t)Y}] = \varphi_X(\alpha t) \varphi_Y(\beta t)$$

But  $X$  and  $Y$  are Gaussian

with respective means  $m_X, m_Y$  and variances  $v_X, v_Y$ .

## Proof Continued

But  $X$  and  $Y$  are Gaussian  
with respective means  $m_X$ ,  $m_Y$  and variances  $v_X$ ,  $v_Y$ .

So  $\varphi_X(t) = \exp[im_X t - \frac{1}{2}v_X t^2]$  and  $\varphi_Y(t) = \exp[im_Y t - \frac{1}{2}v_Y t^2]$

Because  $X$  and  $Y$  are independent and Gaussian,  
the char. function  $\varphi_Z(t) = \mathbb{E}[e^{itZ}]$  of  $Z = \alpha X + \beta Y$  satisfies

$$\varphi_Z(t) = \mathbb{E}[e^{it(\alpha X + \beta Y)}] = \mathbb{E}[e^{i(\alpha t)X}] \mathbb{E}[e^{i(\beta t)Y}] = \varphi_X(\alpha t) \varphi_Y(\beta t)$$

It follows that  $\varphi_Z(t) = \exp[i(\alpha m_X + \beta m_Y)t - \frac{1}{2}(\alpha^2 v_X + \beta^2 v_Y)t^2]$ .

But  $\exp[i(\alpha m_X + \beta m_Y)t - \frac{1}{2}(\alpha^2 v_X + \beta^2 v_Y)t^2]$   
is the characteristic function  $\exp[im_Z t - \frac{1}{2}v_Z t^2]$

of a Gaussian random variable

with mean  $m_Z = \alpha m_X + \beta m_Y$

and variance  $v_Z = \alpha^2 v_X + \beta^2 v_Y$ . □

# Use of Characteristic Functions

Characteristic functions can be used to give superficially simple proofs of both the LLN and the classical central limit theorems.

The following merely sketches the argument.

For much more careful detail, see Richard M. Dudley's major text, *Real Analysis and Probability*.

A key tool is **Lévy's continuity theorem**.

For a sequence of random variables, this connects convergence in distribution to pointwise convergence of their characteristic functions.

# Statement of Lévy's Continuity Theorem

## Theorem

*Suppose  $(X_n)_{n=1}^{\infty}$  is a sequence of random variables, not necessarily sharing a common probability space, with the corresponding sequence*

$$\mathbb{R} \ni t \mapsto \varphi_n(t) = \mathbb{E}e^{itX_n} \in \mathbb{C} \quad (n \in \mathbb{N})$$

*of complex-valued characteristic functions.*

*If  $X_n$  converges in distribution to the random variable  $X$ , then  $t \mapsto \varphi_n(t)$  converges pointwise to  $t \mapsto \varphi(t) = \mathbb{E}e^{itX}$ , the characteristic function of  $X$ .*

*Conversely, if  $t \mapsto \varphi_n(t)$  converges pointwise to a function  $t \mapsto \varphi(t)$  which is continuous at  $t = 0$ , then  $t \mapsto \varphi(t)$  is the characteristic function  $\mathbb{E}e^{itX}$  of a random variable  $X$ , and  $X_n$  converges in distribution to  $X$ .*

## Linear Approximation to the Characteristic Function

Suppose that the random variable  $X$  has a mean  $\mu_X := \mathbb{E}X = \int_{-\infty}^{\infty} x dF(x)$ .

One can then differentiate within the expectation to obtain

$$\frac{d}{dt} \mathbb{E}e^{itX} = \mathbb{E} \left[ \frac{d}{dt} e^{itX} \right] = \mathbb{E}[iXe^{itX}]$$

Consider the linear approximation

$$\mathbb{E}e^{ihX} = 1 + i[\mu + \xi(h)]h \quad \text{where} \quad \xi(h) := (\mathbb{E}e^{ihX} - 1 - ih\mu)/h$$

By l'Hôpital's rule, one has

$$\lim_{h \rightarrow 0} \xi(h) = \text{"0/0"} = \lim_{h \rightarrow 0} (\mathbb{E}[iXe^{ihX}] - i\mu)/1 = \mathbb{E}[iX] - i\mu = 0$$

## Quadratic Approximation to the Characteristic Function

Next, suppose that the random variable  $X$  has not only a mean  $\mu_X := \int_{-\infty}^{\infty} x dF(x)$ , but also a variance  $\sigma_X^2 := \int_{-\infty}^{\infty} (x - \mu)^2 dF(x)$ .

One can then differentiate twice within the expectation to obtain

$$\frac{d^2}{dt^2} \mathbb{E} e^{itX} = \mathbb{E} \left[ \frac{d^2}{dt^2} e^{itX} \right] = \mathbb{E} [(iX)^2 e^{itX}] = -\mathbb{E}[X^2 e^{itX}]$$

Consider the quadratic approximation

$$\mathbb{E} e^{ihX} = 1 + i\mu h - \frac{1}{2}[\sigma^2 + \mu^2 + \eta(h)]h^2$$

where  $\eta(h) := (1/h^2)[\mathbb{E} e^{ihX} - 1 - ih\mu] + \frac{1}{2}(\sigma^2 + \mu^2)$ .

Applying l'Hôpital's rule twice, one has

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h^2} [\mathbb{E} e^{ihX} - 1 - ih\mu] &= \text{"0/0"} = \lim_{h \rightarrow 0} \frac{1}{2h} (\mathbb{E}[iX e^{ihX}] - i\mu) \\ &= \text{"0/0"} = \lim_{h \rightarrow 0} \frac{1}{2} \mathbb{E} [(iX)^2 e^{ihX}] = -\frac{1}{2} \mathbb{E} X^2 = -\frac{1}{2}(\sigma^2 + \mu^2) \end{aligned}$$

implying that  $\eta(h) \rightarrow 0$  as  $h \rightarrow 0$ .

# A Helpful Lemma

## Lemma

Suppose that  $\mathbb{R} \ni h \mapsto \zeta(h) \in \mathbb{C}$  satisfies  $\zeta(h) \rightarrow 0$  as  $h \rightarrow 0$ .

Then for all  $z \in \mathbb{C}$ , one has  $\{1 + \frac{1}{n}[z + \zeta(1/n)]\}^n \rightarrow e^z$  as  $n \rightarrow \infty$ .

For a sketch proof, first one can show that

$$\lim_{n \rightarrow \infty} \left\{ 1 + \frac{1}{n} [z + \zeta(1/n)] \right\}^n = \lim_{n \rightarrow \infty} \left( 1 + \frac{1}{n} z \right)^n$$

Second, in case  $z \in \mathbb{R}$ , putting  $h = 1/n$  and taking logs gives

$$\begin{aligned} \ln \left[ \lim_{n \rightarrow \infty} \left( 1 + \frac{1}{n} z \right)^n \right] &= \ln \left[ \lim_{h \rightarrow 0} (1 + hz)^{1/h} \right] \\ &= \lim_{h \rightarrow 0} \frac{1}{h} [\ln(1 + hz) - \ln 1] = \left. \frac{d}{dh} \ln(1 + hz) \right|_{h=0} = z \end{aligned}$$

implying that  $(1 + \frac{1}{n}z)^n \rightarrow e^z$  as  $n \rightarrow \infty$ .

Dealing with the case when  $z$  is complex is more tricky.

## Sketch Proof of the Weak LLN, I

Consider now any infinite sequence  $X_1, X_2, \dots$  of observations of IID random variables drawn from a common CDF  $F(x)$  on  $\mathbb{R}$ , with common characteristic function  $t \mapsto \varphi_X(t) = \mathbb{E}[e^{itX}]$ .

For each  $n \in \mathbb{N}$ , let  $\bar{X}_n := \frac{1}{n} \sum_{j=1}^n X_j$  denote the random variable whose value is the sample mean of the first  $n$  observations.

This sample mean has its own characteristic function

$$\varphi_{\bar{X}_n}(t) := \mathbb{E}[e^{it\bar{X}_n}] = \mathbb{E}\left[\prod_{j=1}^n e^{itX_j/n}\right]$$

Then

$$\varphi_{\bar{X}_n}(t) = \prod_{j=1}^n \mathbb{E}[e^{itX_j/n}] = \left(\mathbb{E}[e^{itX/n}]\right)^n$$

because the random variables  $X_j$  are respectively independently and identically distributed.

## Sketch Proof of the Weak LLN, II

Suppose we take the linear approximation

$$\mathbb{E}e^{ihX} = 1 + i[\mu + \xi(h)]h \quad \text{where} \quad \xi(h) \rightarrow 0 \quad \text{as} \quad h \rightarrow 0$$

and replace  $h$  by  $t/n$  to obtain

$$\mathbb{E}[e^{it\bar{X}_n}] = \{1 + (it/n)[\mu + \xi(t/n)]\}^n$$

Because  $\xi(t/n) \rightarrow 0$  as  $n \rightarrow \infty$  and so  $h = t/n \rightarrow 0$ , one has

$$\lim_{n \rightarrow \infty} \{1 + \frac{1}{n}it[\mu + \xi(t/n)]\}^n = \lim_{n \rightarrow \infty} (1 + \frac{1}{n}it\mu)^n = e^{it\mu} = \mathbb{E}[e^{itY}]$$

where  $\mathbb{E}[e^{itY}]$  is the characteristic function of a degenerate random variable  $Y$  which is equal to  $\mu$  with probability 1.

Using the Lévy theorem, it follows that the distribution of  $\bar{X}_n$  converges to this degenerate distribution, implying that  $\bar{X}_n$  converges to  $\mu$  in probability.

## Sketch Proof of the Classical CLT, I

For each  $j \in \mathbb{N}$ , let  $Z_j$  denote the **standardized** value  $(X_j - \mu)/\sigma$  of  $X_j$ , defined to have the property that  $\mathbb{E}Z_j = 0$  and  $\mathbb{E}Z_j^2 = 1$ .

Now define  $\bar{Z}_n := \sum_{j=1}^n \frac{Z_j}{\sqrt{n}}$ .

This is called the **standardized mean** because:

1. linearity implies that  $\mathbb{E}\bar{Z}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbb{E}Z_j = 0$ ;
2. independence implies that  $\mathbb{E}\bar{Z}_n^2 = \frac{1}{n} \sum_{j=1}^n \mathbb{E}Z_j^2 = 1$ .

Putting  $\mu = 0$  and  $\sigma^2 = 1$  in the quadratic approximation

$$\mathbb{E}e^{ihX} = 1 + i\mu h - \frac{1}{2}[\sigma^2 + \mu^2 + \eta(h)]h^2$$

implies  $\mathbb{E}e^{ihZ} = 1 - \frac{1}{2}[1 + \eta(h)]h^2$  where  $\eta(h) \rightarrow 0$  as  $h \rightarrow 0$ .

Replacing  $hX$  by  $tZ_j/\sqrt{n}$  in this quadratic approximation yields

$$\mathbb{E}[e^{itZ_j/\sqrt{n}}] = 1 - \frac{1}{2} \frac{t^2}{n} [1 + \eta(t/n)]$$

## Sketch Proof of the Classical CLT, II

Now independence implies that

$$\mathbb{E}[e^{it\bar{Z}_n}] = \mathbb{E}\left[\exp\left(it\frac{1}{\sqrt{n}}\sum_{j=1}^n Z_j\right)\right] = \prod_{j=1}^n \mathbb{E}[e^{itZ_j/\sqrt{n}}]$$

Hence, another careful limiting argument shows that

$$\mathbb{E}[e^{it\bar{Z}_n}] = \left\{1 - \frac{1}{2}\frac{t^2}{n}[1 + \eta(t/n)]\right\}^n \rightarrow e^{-\frac{1}{2}t^2} \text{ as } n \rightarrow \infty$$

But we showed that this limit  $e^{-\frac{1}{2}t^2}$  is precisely the characteristic function of a standard normal distribution  $N(0, 1)$ .

Because  $t \mapsto e^{-\frac{1}{2}t^2}$  is continuous at  $t = 0$ , the central limit theorem follows from the Lévy continuity theorem, which confirms that the convergence of characteristic functions implies convergence in distribution.

# Outline

Kolmogorov's Definition of Probability

Random Variables and Their Distribution and Density Functions

Expected Values

Joint Probability Distributions

Limit Theorems

Convergence Results

**Non-integrability for Macroeconomists**

A Continuum of Independent Random Variables

# Economic Models with a Continuum of Agents, I

Harold Hotelling (1929) “Stability in Competition”  
*Economic Journal* 39 (153): 41–57.

William S. Vickrey (1953) “Measuring Marginal Utility  
by Reactions to Risk” *Econometrica* 13: 319–33.

James A. Mirrlees (1971)  
“An Exploration in the Theory of Optimum Income Taxation”  
*Review of Economic Studies* 38: 175–208.

Robert J. Aumann (1964) “Markets with a Continuum of Traders”  
*Econometrica* 32: 39–50.

Robert J. Aumann (1966) “Existence of Competitive Equilibria  
in Markets with a Continuum of Traders” *Econometrica* 34: 1–17.

Werner Hildenbrand (1974) *Core and Equilibria  
of a Large Economy* (Princeton University Press).

# Economic Models with a Continuum of Agents, II

Truman Bewley (1987) “Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consumers” in W. Hildenbrand and A. Mas-Colell (eds.) *Contributions to Mathematical Economics in Honor of Gerard Debreu* (North Holland) ch. 5.

# A Continuum of Independent Random Variables

Hotelling had a continuum of ice cream buyers distributed along a finite beach on a hot day.

Vickrey and Mirrlees considered optimal income taxation when a population of workers have continuously distributed skills.

Aumann and Hildenbrand modelled a perfectly competitive market system with a continuum of traders who each have negligible individual influence over market prices.

Market clearing in the economy as a whole requires, for each separate commodity, equality between:  
(i) mean demand per trader; and (ii) mean endowment per trader.

Bewley considered a continuum of consumers with “independently fluctuating” random endowments.

Then, is mean endowment in the population even defined?

# A Process with a Continuum of IID Random Variables

Let  $\mathcal{L}$  denote the Lebesgue  $\sigma$ -field on  $\mathbb{R}$ , and let  $I$  denote the unit interval  $[0, 1]$ .

## Definition

A **process** with a **continuum of iid random variables** on the Lebesgue unit interval  $(I, \mathcal{L}, \lambda)$  involves:

- ▶ a sample probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ ;
- ▶ a mapping  $I \times \Omega \ni (i, \omega) \mapsto f(i, \omega) \in \mathbb{R}$  satisfying

$$\begin{aligned} & \mathbb{P}(\cap_{n \in \mathbb{N}} \{\omega \in \Omega \mid f(i_n, \omega) \in B_n\}) \\ &= \prod_{n \in \mathbb{N}} \mathbb{P}(\{\omega \in \Omega \mid f(i_n, \omega) \in B_n\}) \end{aligned}$$

for every countable collection  $(i^{\mathbb{N}}, B^{\mathbb{N}})$   
of pairs  $(i, B) \in I \times \mathcal{B}$ . □

## Difficulty Illustrated

In the following graphs, think of the horizontal axis as the Lebesgue unit interval, indicating something like a U.S. social security number (SSN) ( $\times 10^{-9}$ ).

Think of the vertical axis as the Lebesgue unit interval, indicating something like an individual's height, measured as a percentile.

Assume that SSN gives no information about height.

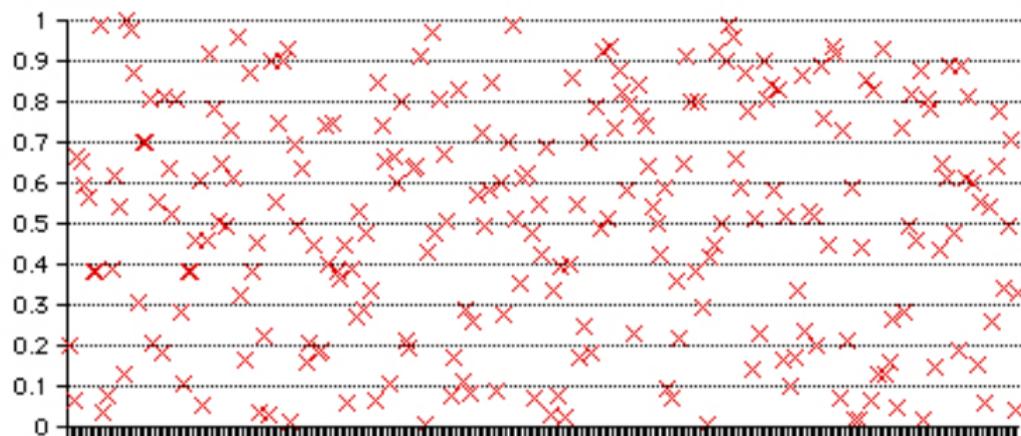
Then the heights of, approximately, a continuum of individuals may be regarded as statistically IID random variables.

## Difficulty Illustrated: 25 Random Draws



The points have  $x \in \{(1/25)(n - \frac{1}{2}) \mid n \in \{1, \dots, 25\}\}$   
and  $y$  pseudo-randomly drawn from a uniform distribution on  $[0, 1]$ .  
Finding the mean of  $y$  is becoming messy.

## Difficulty Illustrated: 200 Random Draws



The points have  $x \in \left\{ \left( \frac{1}{200} \right) \left( n - \frac{1}{2} \right) \mid n \in \{1, \dots, 200\} \right\}$   
and  $y$  pseudo-randomly drawn from a uniform distribution on  $[0, 1]$ .  
Finding the mean of  $y$  is becoming impossible.

# Essential Supremum and Infimum

Recall that the supremum is the least upper bound, and the infimum is the greatest lower bound.

Define the **essential** supremum and infimum of each random variable  $\omega \mapsto f(i, \omega)$  as:

$$\begin{aligned} \text{ess sup } f(i, \omega) &:= \inf \{b \in \mathbb{R} \mid \mathbb{P}(\{\omega \in \Omega \mid f(i, \omega) \leq b\}) = 1\} \\ \text{ess inf } f(i, \omega) &:= \sup \{a \in \mathbb{R} \mid \mathbb{P}(\{\omega \in \Omega \mid f(i, \omega) \geq a\}) = 1\} \end{aligned}$$

These differ from the supremum and infimum by allowing one to disregard an event which has probability zero.

# A Theorem on Non-Measurable Sample Paths

For all the continuum of IID random variables  $\omega \mapsto f(i, \omega)$ , let  $a \in \mathbb{R}$  denote the common value of  $\text{ess inf } f(i, \omega)$  and  $b \in \mathbb{R}$  the common value of  $\text{ess sup } f(i, \omega)$ .

## Theorem

*Whenever  $a < b$ , the sample path  $I \ni i \mapsto f(i, \omega)$  is  $\mathbb{P}$ -a.s. non measurable.*

Of course, when  $a = b$ , then  $\mathbb{P}$ -a.s. one has  $f(i, \omega) = a = b$  — i.e., the process  $(i, \omega) \mapsto f(i, \omega)$  is **essentially constant**.

The key idea in the proof is to show that the sample path has a lower integral  $a$  and an upper integral  $b$ .

## Monte Carlo Integration

Because of the strong law of large numbers, here is one way to approximate numerically the integral  $\int_K f(\mathbf{x}) \mu(d\mathbf{x})$  of a complicated function of  $\ell$  variables, where  $K \subset \mathbb{R}^\ell$  has an  $\ell$ -dimensional Lebesgue measure  $\mu(K) < +\infty$ .

1. First, choose a large sample  $\langle \mathbf{x}^r \rangle_{r=1}^n$  of  $n$  points that are independent and identically distributed random draws from the set  $K$ , with common probability measure  $\pi$  satisfying  $\pi(B) = \mu(B)/\mu(K)$  for all Borel sets  $B \subseteq K$ .
2. Second, calculate the sample average function value

$$M^n(\langle \mathbf{x}^r \rangle_{r=1}^n) := \frac{1}{n} \sum_{r=1}^n f(\mathbf{x}^r)$$

3. Third, observe that, by the strong law of large numbers, the sample average  $M^n(\langle \mathbf{x}^r \rangle_{r=1}^n)$  converges almost surely as  $n \rightarrow \infty$  to the theoretical mean

$$\mathbb{E}_\pi[f(\mathbf{x}^k)] = \int_K f(\mathbf{x}) \pi(d\mathbf{x}) = \frac{1}{\mu(K)} \int_K f(\mathbf{x}) \mu(d\mathbf{x})$$

# The Monte Carlo Integral: Rescuing Macroeconomics

## Definition

Given the process  $I \times \Omega \ni (i, \omega) \mapsto f(i, \omega) \in \mathbb{R}$   
with a continuum of iid random variables,  
define the **Monte Carlo integral** as the random variable

$$\Omega \ni \omega \mapsto {}_{\text{MC}} \int_I f(i, \omega) \lambda(di) \in \mathbb{R}$$

as the almost sure limit as  $n \rightarrow \infty$  of the average  $\frac{1}{n} \sum_{k=1}^n f(i_k, \omega)$   
when the  $n$  points  $\langle i_k \rangle_{k=1}^n$  are independent draws  
from the Lebesgue unit interval  $(I, \mathcal{L}, \lambda)$ .

Then, even though the Lebesgue integral  $\int_I f(i, \omega) \lambda(di)$   
is almost surely undefined, the strong law of large numbers  
implies that the Monte Carlo integral  ${}_{\text{MC}} \int_I f(i, \omega) \lambda(di)$   
is well defined as a degenerate random variable  $\Omega \ni \omega \mapsto \delta_m(\omega)$   
that attaches probability one  
to the common theoretical mean  $m := \int_{\Omega} f(i, \omega) P(d\omega)$ .