

# Searching for Universal Truths

## Measure-theoretic Treatment of Statistics

**Sunghee Yun**  
sunghee.yun@gmail.com

# Navigating Mathematical and Statistical Territories

- Notations & definitions & conventions
  - notations - 2
  - some definitions - 6
  - some conventions - 7
- Measure-theoretic treatment of probabilities - 8
  - probability measure - 9
  - random variables - 22
  - convergence of random variables - 42
- Proof & references & indices
  - selected proofs - 56
  - references - 58
  - index - 60

## Notations

- sets of numbers
  - $\mathbf{N}$  - set of natural numbers
  - $\mathbf{Z}$  - set of integers
  - $\mathbf{Z}_+$  - set of nonnegative integers
  - $\mathbf{Q}$  - set of rational numbers
  - $\mathbf{R}$  - set of real numbers
  - $\mathbf{R}_+$  - set of nonnegative real numbers
  - $\mathbf{R}_{++}$  - set of positive real numbers
  - $\mathbf{C}$  - set of complex numbers
- sequences  $\langle x_i \rangle$  and the like
  - finite  $\langle x_i \rangle_{i=1}^n$ , infinite  $\langle x_i \rangle_{i=1}^\infty$  - use  $\langle x_i \rangle$  whenever unambiguously understood
  - similarly for other operations, *e.g.*,  $\sum x_i$ ,  $\prod x_i$ ,  $\cup A_i$ ,  $\cap A_i$ ,  $\times A_i$
  - similarly for integrals, *e.g.*,  $\int f$  for  $\int_{-\infty}^\infty f$
- sets
  - $\tilde{A}$  - complement of  $A$

- $A \sim B$  -  $A \cap \tilde{B}$
- $A \Delta B$  -  $(A \cap \tilde{B}) \cup (\tilde{A} \cap B)$
- $\mathcal{P}(A)$  - set of all subsets of  $A$
- sets in metric vector spaces
  - $\overline{A}$  - closure of set  $A$
  - $A^\circ$  - interior of set  $A$
  - **relint**  $A$  - relative interior of set  $A$
  - **bd**  $A$  - boundary of set  $A$
- set algebra
  - $\sigma(\mathcal{A})$  -  $\sigma$ -algebra generated by  $\mathcal{A}$ , *i.e.*, smallest  $\sigma$ -algebra containing  $\mathcal{A}$
- norms in  $\mathbf{R}^n$ 
  - $\|x\|_p$  ( $p \geq 1$ ) -  $p$ -norm of  $x \in \mathbf{R}^n$ , *i.e.*,  $(|x_1|^p + \cdots + |x_n|^p)^{1/p}$
  - *e.g.*,  $\|x\|_2$  - Euclidean norm
- matrices and vectors
  - $a_i$  -  $i$ -th entry of vector  $a$
  - $A_{ij}$  - entry of matrix  $A$  at position  $(i, j)$ , *i.e.*, entry in  $i$ -th row and  $j$ -th column
  - $\text{Tr}(A)$  - trace of  $A \in \mathbf{R}^{n \times n}$ , *i.e.*,  $A_{1,1} + \cdots + A_{n,n}$

- symmetric, positive definite, and positive semi-definite matrices
  - $\mathbf{S}^n \subset \mathbf{R}^{n \times n}$  - set of symmetric matrices
  - $\mathbf{S}_+^n \subset \mathbf{S}^n$  - set of positive semi-definite matrices;  $A \succeq 0 \Leftrightarrow A \in \mathbf{S}_+^n$
  - $\mathbf{S}_{++}^n \subset \mathbf{S}^n$  - set of positive definite matrices;  $A \succ 0 \Leftrightarrow A \in \mathbf{S}_{++}^n$
- sometimes, use Python script-like notations (with serious abuse of mathematical notations)
  - use  $f : \mathbf{R} \rightarrow \mathbf{R}$  as if it were  $f : \mathbf{R}^n \rightarrow \mathbf{R}^n$ , *e.g.*,

$$\exp(x) = (\exp(x_1), \dots, \exp(x_n)) \quad \text{for } x \in \mathbf{R}^n$$

and

$$\log(x) = (\log(x_1), \dots, \log(x_n)) \quad \text{for } x \in \mathbf{R}_{++}^n$$

which corresponds to Python code `numpy.exp(x)` or `numpy.log(x)` where `x` is instance of `numpy.ndarray`, *i.e.*, numpy array

- use  $\sum x$  to mean  $\mathbf{1}^T x$  for  $x \in \mathbf{R}^n$ , *i.e.*

$$\sum x = x_1 + \dots + x_n$$

which corresponds to Python code `x.sum()` where `x` is numpy array

- use  $x/y$  for  $x, y \in \mathbf{R}^n$  to mean

$$\begin{bmatrix} x_1/y_1 & \cdots & x_n/y_n \end{bmatrix}^T$$

which corresponds to Python code `x / y` where `x` and `y` are 1-d numpy arrays

- use  $X/Y$  for  $X, Y \in \mathbf{R}^{m \times n}$  to mean

$$\begin{bmatrix} X_{1,1}/Y_{1,1} & X_{1,2}/Y_{1,2} & \cdots & X_{1,n}/Y_{1,n} \\ X_{2,1}/Y_{2,1} & X_{2,2}/Y_{2,2} & \cdots & X_{2,n}/Y_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m,1}/Y_{m,1} & X_{m,2}/Y_{m,2} & \cdots & X_{m,n}/Y_{m,n} \end{bmatrix}$$

which corresponds to Python code `X / Y` where `X` and `Y` are 2-d numpy arrays

## Some definitions

**Definition 1. [infinitely often - i.o.]** *statement  $P_n$ , said to happen infinitely often or i.o. if*

$$(\forall N \in \mathbf{N}) (\exists n > N) (P_n)$$

**Definition 2. [almost everywhere - a.e.]** *statement  $P(x)$ , said to happen almost everywhere or a.e. or almost surely or a.s. (depending on context) associated with measure space  $(X, \mathcal{B}, \mu)$  if*

$$\mu\{x | P(x)\} = 1$$

*or equivalently*

$$\mu\{x | \sim P(x)\} = 0$$

## Some conventions

- (for some subjects) use following conventions

- $0 \cdot \infty = \infty \cdot 0 = 0$

- $(\forall x \in \mathbf{R}_{++})(x \cdot \infty = \infty \cdot x = \infty)$

- $\infty \cdot \infty = \infty$



# **Measure-theoretic Treatment of Probabilities**

# **Probability Measure**

## Measurable functions

- denote *n-dimensional Borel sets* by  $\mathcal{R}^n$
- for two measurable spaces,  $(\Omega, \mathcal{F})$  and  $(\Omega', \mathcal{F}')$ , function,  $f : \Omega \rightarrow \Omega'$  with

$$(\forall A' \in \mathcal{F}') \left( f^{-1}(A') \in \mathcal{F} \right)$$

said to be *measurable with respect to  $\mathcal{F} / \mathcal{F}'$*  (thus, measurable functions defined on page ?? and page ?? can be said to be measurable with respect to  $\mathcal{B} / \mathcal{R}$ )

- when  $\Omega = \mathbf{R}^n$  in  $(\Omega, \mathcal{F})$ ,  $\mathcal{F}$  is assumed to be  $\mathcal{R}^n$ , and sometimes drop  $\mathcal{R}^n$ 
  - thus, *e.g.*, we say  $f : \Omega \rightarrow \mathbf{R}^n$  is measurable with respect to  $\mathcal{F}$  (instead of  $\mathcal{F} / \mathcal{R}^n$ )
- measurable function,  $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$  (*i.e.*, measurable with respect to  $\mathcal{R}^n / \mathcal{R}^m$ ), called *Borel functions*
- $f : \Omega \rightarrow \mathbf{R}^n$  is measurable with respect to  $\mathcal{F} / \mathcal{R}^n$  *if and only if* every component,  $f_i : \Omega \rightarrow \mathbf{R}$ , is measurable with respect to  $\mathcal{F} / \mathcal{R}$

## Probability (measure) spaces

- set function,  $P : \mathcal{F} \rightarrow [0, 1]$ , defined on algebra,  $\mathcal{F}$ , of set  $\Omega$ , satisfying following properties, called *probability measure* (refer to page ?? for resemblance with measurable spaces)
  - $(\forall A \in \mathcal{F})(0 \leq P(A) \leq 1)$
  - $P(\emptyset) = 0, P(\Omega) = 1$
  - $(\forall \text{ disjoint } \langle A_n \rangle \subset \mathcal{F})(P(\bigcup A_n) = \sum P(A_n))$
- for  $\sigma$ -algebra,  $\mathcal{F}$ ,  $(\Omega, \mathcal{F}, P)$ , called *probability measure space* or *probability space*
- set  $A \in \mathcal{F}$  with  $P(A) = 1$ , called *a support of  $P$*

## Dynkin's $\pi$ - $\lambda$ theorem

- class,  $\mathcal{P}$ , of subsets of  $\Omega$  closed under finite intersection, called  $\pi$ -system, i.e.,
  - $(\forall A, B \in \mathcal{P})(A \cap B \in \mathcal{P})$
- class,  $\mathcal{L}$ , of subsets of  $\Omega$  containing  $\Omega$  closed under complements and countable disjoint unions called  $\lambda$ -system
  - $\Omega \in \mathcal{L}$
  - $(\forall A \in \mathcal{L})(\tilde{A} \in \mathcal{L})$
  - $(\forall \text{ disjoint } \langle A_n \rangle)(\bigcup A_n \in \mathcal{L})$
- class that is both  $\pi$ -system and  $\lambda$ -system is  $\sigma$ -algebra
- *Dynkin's  $\pi$ - $\lambda$  theorem* - for  $\pi$ -system,  $\mathcal{P}$ , and  $\lambda$ -system,  $\mathcal{L}$ , with  $\mathcal{P} \subset \mathcal{L}$ ,

$$\sigma(\mathcal{P}) \subset \mathcal{L}$$

- for  $\pi$ -system,  $\mathcal{P}$ , two probability measures,  $P_1$  and  $P_2$ , on  $\sigma(\mathcal{P})$ , agreeing  $\mathcal{P}$ , agree on  $\sigma(\mathcal{P})$

## Limits of Events

**Theorem 1. [convergence-of-events]** *no for sequence of subsets,  $\langle A_n \rangle$ ,*

$$P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n)$$

- *for  $\langle A_n \rangle$  converging to  $A$*

$$\lim P(A_n) = P(A)$$

**Theorem 2. [independence-of-smallest-sig-alg]** *no for sequence of  $\pi$ -systems,  $\langle \mathcal{A}_n \rangle$ ,  $\langle \sigma(\mathcal{A}_n) \rangle$  is independent*

## Probabilistic independence

– given probability space,  $(\Omega, \mathcal{F}, P)$

- $A, B \in \mathcal{F}$  with

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

said to be *independent*

- indexed collection,  $\langle A_\lambda \rangle$ , with

$$(\forall n \in \mathbf{N}, \text{ distinct } \lambda_1, \dots, \lambda_n \in \Lambda) \left( P \left( \bigcap_{i=1}^n A_{\lambda_i} \right) = \prod_{i=1}^n P(A_{\lambda_i}) \right)$$

said to be *independent*

## Independence of classes of events

- indexed collection,  $\langle \mathcal{A}_\lambda \rangle$ , of classes of events (*i.e.*, subsets) with

$$(\forall A_\lambda \in \mathcal{A}_\lambda) (\langle A_\lambda \rangle \text{ are independent})$$

said to be *independent*

- *for independent indexed collection,  $\langle \mathcal{A}_\lambda \rangle$ , with every  $\mathcal{A}_\lambda$  being  $\pi$ -system,  $\langle \sigma(\mathcal{A}_\lambda) \rangle$  are independent*
- for independent (countable) collection of events,  $\langle \langle A_{ni} \rangle_{i=1}^\infty \rangle_{n=1}^\infty$ ,  $\langle \mathcal{F}_n \rangle_{n=1}^\infty$  with  $\mathcal{F}_n = \sigma(\langle A_{ni} \rangle_{i=1}^\infty)$  are independent



## Borel-Cantelli lemmas

- **Lemma 1. [first Borel-Cantelli]** *for sequence of events,  $\langle A_n \rangle$ , with  $\sum P(A_n)$  converging*

$$P(\limsup A_n) = 0$$

- **Lemma 2. [second Borel-Cantelli]** *for independent sequence of events,  $\langle A_n \rangle$ , with  $\sum P(A_n)$  diverging*

$$P(\limsup A_n) = 1$$

## Tail events and Kolmogorov's zero-one law

- for sequence of events,  $\langle A_n \rangle$

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(\langle A_i \rangle_{i=n}^{\infty})$$

called *tail  $\sigma$ -algebra associated with  $\langle A_n \rangle$* ; its elements are called *tail events*

- *Kolmogorov's zero-one law* - for independent sequence of events,  $\langle A_n \rangle$  every event in tail  $\sigma$ -algebra has probability measure either 0 or 1

## Product probability spaces

- for two measure spaces,  $(X, \mathcal{X}, \mu)$  and  $(Y, \mathcal{Y}, \nu)$ , want to find product measure,  $\pi$ , such that

$$(\forall A \in \mathcal{X}, B \in \mathcal{Y}) (\pi(A \times B) = \mu(A)\nu(B))$$

- e.g., if both  $\mu$  and  $\nu$  are Lebesgue measure on  $\mathbf{R}$ ,  $\pi$  will be Lebesgue measure on  $\mathbf{R}^2$
- $A \times B$  for  $A \in \mathcal{X}$  and  $B \in \mathcal{Y}$  is *measurable rectangle*
- *$\sigma$ -algebra generated by measurable rectangles* denoted by

$$\mathcal{X} \times \mathcal{Y}$$

- thus, *not* Cartesian product in usual sense
- generally *much larger* than class of measurable rectangles

## Sections of measurable subsets and functions

for two measure spaces,  $(X, \mathcal{X}, \mu)$  and  $(Y, \mathcal{Y}, \nu)$

- sections of measurable subsets
  - $\{y \in Y \mid (x, y) \in E\}$  is *section of  $E$  determined by  $x$*
  - $\{x \in X \mid (x, y) \in E\}$  is *section of  $E$  determined by  $y$*
- sections of measurable functions - for measurable function,  $f$ , with respect to  $\mathcal{X} \times \mathcal{Y}$ 
  - $f(x, \cdot)$  is *section of  $f$  determined by  $x$*
  - $f(\cdot, y)$  is *section of  $f$  determined by  $y$*
- sections of measurable subsets are measurable
  - $(\forall x \in X, E \in \mathcal{X} \times \mathcal{Y}) (\{y \in Y \mid (x, y) \in E\} \in \mathcal{Y})$
  - $(\forall y \in Y, E \in \mathcal{X} \times \mathcal{Y}) (\{x \in X \mid (x, y) \in E\} \in \mathcal{X})$
- sections of measurable functions are measurable
  - $f(x, \cdot)$  is measurable with respect to  $\mathcal{Y}$  for every  $x \in X$
  - $f(\cdot, y)$  is measurable with respect to  $\mathcal{X}$  for every  $y \in Y$

## Product measure

for two  $\sigma$ -finite measure spaces,  $(X, \mathcal{X}, \mu)$  and  $(Y, \mathcal{Y}, \nu)$

- two functions defined below for every  $E \in \mathcal{X} \times \mathcal{Y}$  are  $\sigma$ -finite measures
  - $\pi'(E) = \int_X \nu\{y \in Y | (x, y) \in E\} d\mu$
  - $\pi''(E) = \int_Y \mu\{x \in X | (x, y) \in E\} d\nu$
- for every measurable rectangle,  $A \times B$ , with  $A \in \mathcal{X}$  and  $B \in \mathcal{Y}$

$$\pi'(A \times B) = \pi''(A \times B) = \mu(A)\nu(B)$$

(use conventions in page 7 for extended real values)

- indeed,  $\pi'(E) = \pi''(E)$  for every  $E \in \mathcal{X} \times \mathcal{Y}$ ; let  $\pi = \pi' = \pi''$
- $\pi$  is
  - called *product measure* and denoted by  $\mu \times \nu$
  - $\sigma$ -finite measure
  - *only* measure such that  $\pi(A \times B) = \mu(A)\nu(B)$  for every measurable rectangle

## Fubini's theorem

- suppose two  $\sigma$ -finite measure spaces,  $(X, \mathcal{X}, \mu)$  and  $(Y, \mathcal{Y}, \nu)$  - define
  - $X_0 = \{x \in X \mid \int_Y |f(x, y)| d\nu < \infty\} \subset X$
  - $Y_0 = \{y \in Y \mid \int_X |f(x, y)| d\mu < \infty\} \subset Y$
- *Fubini's theorem* - for nonnegative measurable function,  $f$ , following are measurable with respect to  $\mathcal{X}$  and  $\mathcal{Y}$  respectively

$$g(x) = \int_Y f(x, y) d\nu, \quad h(y) = \int_X f(x, y) d\mu$$

and following holds

$$\int_{X \times Y} f(x, y) d\pi = \int_X \left( \int_Y f(x, y) d\nu \right) d\mu = \int_Y \left( \int_X f(x, y) d\mu \right) d\nu$$

- for  $f$ , (not necessarily nonnegative) integrable function with respect to  $\pi$ 
  - $\mu(X \setminus X_0) = 0, \nu(Y \setminus Y_0) = 0$
  - $g$  and  $h$  are finite measurable on  $X_0$  and  $Y_0$  respectively
  - (above) equalities of *double integral* holds

# Random Variables

## Random variables

- for probability space,  $(\Omega, \mathcal{F}, P)$ ,
- measurable function (with respect to  $\mathcal{F}/\mathcal{R}$ ),  $X : \Omega \rightarrow \mathbf{R}$ , called *random variable*
- measurable function (with respect to  $\mathcal{F}/\mathcal{R}^n$ ),  $X : \Omega \rightarrow \mathbf{R}^n$ , called *random vector*
  - when expressing  $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ ,  $X$  is measurable *if and only if* every  $X_i$  is measurable
  - thus,  $n$ -dimensional random vector is simply  $n$ -tuple of random variables
- smallest  $\sigma$ -algebra with respect to which  $X$  is measurable, called  *$\sigma$ -algebra generated by  $X$*  and denoted by  $\sigma(X)$ 
  - $\sigma(X)$  consists exactly of sets,  $\{\omega \in \Omega | X(\omega) \in H\}$ , for  $H \in \mathcal{R}^n$
  - random variable,  $Y$ , is measurable with respect to  $\sigma(X)$  *if and only if* exists measurable function,  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  such that  $Y(\omega) = f(X(\omega))$  for all  $\omega$ , *i.e.*,  $Y = f \circ X$



## Probability distributions for random variables

- probability measure on  $\mathbf{R}$ ,  $\mu = PX^{-1}$ , i.e.,

$$\mu(A) = P(X \in A) \text{ for } A \in \mathcal{R}$$

called *distribution* or *law* of random variable,  $X$

- function,  $F : \mathbf{R} \rightarrow [0, 1]$ , defined by

$$F(x) = \mu(-\infty, x] = P(X \leq x)$$

called *distribution function* or *cumulative distribution function (CDF)* of  $X$

- Borel set,  $S$ , with  $P(S) = 1$ , called *support*
- random variable, its distribution, its distribution function, said to be *discrete* when has *countable* support

## Probability distribution of mappings of random variables

- for measurable  $g : \mathbf{R} \rightarrow \mathbf{R}$ ,

$$(\forall A \in \mathcal{R}) \left( \mathbf{Prob} (g(X) \in A) = \mathbf{Prob} \left( X \in g^{-1}(A) \right) = \mu(g^{-1}(A)) \right)$$

hence,  $g(X)$  has distribution of  $\mu g^{-1}$

## Probability density for random variables

- Borel function,  $f : \mathbf{R} \rightarrow \mathbf{R}_+$ , satisfying

$$(\forall A \in \mathcal{R}) \left( \mu(A) = P(X \in A) = \int_A f(x) dx \right)$$

called *density* or *probability density function (PDF)* of random variable

- above is equivalent to

$$(\forall a < b \in \mathbf{R}) \left( \int_a^b f(x) dx = P(a < X \leq b) = F(b) - F(a) \right)$$

(refer to statement on page 12)

- note, though,  $F$  does not need to differentiate to  $f$  everywhere; only  $f$  required to integrate properly
- if  $F$  does differentiate to  $f$  and  $f$  is continuous, *fundamental theorem of calculus* implies  $f$  indeed is density for  $F$

## Probability distribution for random vectors

- (similarly to random variables) probability measure on  $\mathbf{R}^n$ ,  $\mu = PX^{-1}$ , *i.e.*,

$$\mu(A) = P(X \in A) \text{ for } A \in \mathcal{B}^k$$

called *distribution* or *law* of random vector,  $X$

- function,  $F : \mathbf{R}^k \rightarrow [0, 1]$ , defined by

$$F(x) = \mu S_x = P(X \preceq x)$$

where

$$S_x = \{\omega \in \Omega | X(\omega) \preceq x\} = \{\omega \in \Omega | X_i(\omega) \leq x_i\}$$

called *distribution function* or *cumulative distribution function (CDF)* of  $X$

- (similarly to random variables) random vector, its distribution, its distribution function, said to be *discrete* when has *countable* support

## Marginal distribution for random vectors

- (similarly to random variables) for measurable  $g : \mathbf{R}^n \rightarrow \mathbf{R}^m$

$$(\forall A \in \mathcal{R}^m) \left( \mathbf{Prob} (g(X) \in A) = \mathbf{Prob} \left( X \in g^{-1}(A) \right) = \mu(g^{-1}(A)) \right)$$

hence,  $g(X)$  has distribution of  $\mu g^{-1}$

- for  $g_i : \mathbf{R}^n \rightarrow \mathbf{R}$  with  $g_i(x) = x_i$

$$(\forall A \in \mathcal{R}) (\mathbf{Prob} (g(X) \in A) = \mathbf{Prob} (X_i \in A))$$

- measure,  $\mu_i$ , defined by  $\mu_i(A) = \mathbf{Prob} (X_i \in A)$ , called *(i-th) marginal distribution of  $X$*
- for  $\mu$  having density function,  $f : \mathbf{R}^n \rightarrow \mathbf{R}_+$ , density function of marginal distribution is

$$f_i(x) = \int_{\mathcal{R}^{n-1}} f(x_{-i}) d\mu_{-i}$$

where  $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$  and similarly for  $d\mu_{-i}$

## Independence of random variables

- random variables,  $X_1, \dots, X_n$ , with independent  $\sigma$ -algebras generated by them, said to be *independent*

(refer to page 15 for independence of collections of subsets)

- because  $\sigma(X_i) = X_i^{-1}(\mathcal{R}) = \{X_i^{-1}(H) | H \in \mathcal{R}\}$ , independent *if and only if*

$$(\forall H_1, \dots, H_n \in \mathcal{R}) \left( P(X_1 \in H_1, \dots, X_n \in H_n) = \prod P(X_i \in H_i) \right)$$

*i.e.,*

$$(\forall H_1, \dots, H_n \in \mathcal{R}) \left( P\left(\bigcap X_i^{-1}(H_i)\right) = \prod P\left(X_i^{-1}(H_i)\right) \right)$$

## Equivalent statements of independence of random variables

- for random variables,  $X_1, \dots, X_n$ , having  $\mu$  and  $F : \mathbf{R}^n \rightarrow [0, 1]$  as their distribution and CDF, with each  $X_i$  having  $\mu_i$  and  $F_i : \mathbf{R} \rightarrow [0, 1]$  as its distribution and CDF, following statements are *equivalent*
  - $X_1, \dots, X_n$  are independent
  - $(\forall H_1, \dots, H_n \in \mathcal{R}) \left( P \left( \bigcap X_i^{-1}(H_i) \right) = \prod P \left( X_i^{-1}(H_i) \right) \right)$
  - $(\forall H_1, \dots, H_n \in \mathcal{R}) \left( P(X_1 \in H_1, \dots, X_n \in H_n) = \prod P(X_i \in H_i) \right)$
  - $(\forall x \in \mathbf{R}^n) \left( P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod P(X_i \leq x_i) \right)$
  - $(\forall x \in \mathbf{R}^n) \left( F(x) = \prod F_i(x_i) \right)$
  - $\mu = \mu_1 \times \dots \times \mu_n$
  - $(\forall x \in \mathbf{R}^n) \left( f(x) = \prod f_i(x_i) \right)$

## Independence of random variables with separate $\sigma$ -algebra

- given probability space,  $(\Omega, \mathcal{F}, P)$
- random variables,  $X_1, \dots, X_n$ , each of which is measurable with respect to each of  $n$  independent  $\sigma$ -algebras,  $\mathcal{G}_1 \subset \mathcal{F}, \dots, \mathcal{G}_n \subset \mathcal{F}$  respectively, are independent



## Independence of random vectors

- for random vectors,  $X_1 : \Omega \rightarrow \mathbf{R}^{d_1}, \dots, X_n : \Omega \rightarrow \mathbf{R}^{d_n}$ , having  $\mu$  and  $F : \mathbf{R}^{d_1} \times \dots \times \mathbf{R}^{d_n} \rightarrow [0, 1]$  as their distribution and CDF, with each  $X_i$  having  $\mu_i$  and  $F_i : \mathbf{R}^{d_i} \rightarrow [0, 1]$  as its distribution and CDF, following statements are *equivalent*
  - $X_1, \dots, X_n$  are independent
  - $(\forall H_1 \in \mathcal{R}^{d_1}, \dots, H_n \in \mathcal{R}^{d_n}) (P(\bigcap X_i^{-1}(H_i)) = \prod P(X_i^{-1}(H_i)))$
  - $(\forall H_1 \in \mathcal{R}^{d_1}, \dots, H_n \in \mathcal{R}^{d_n}) (P(X_1 \in H_1, \dots, X_n \in H_n) = \prod P(X_i \in H_i))$
  - $(\forall x_1 \in \mathbf{R}^{d_1}, \dots, x_n \in \mathbf{R}^{d_n}) (P(X_1 \preceq x_1, \dots, X_n \preceq x_n) = \prod P(X_i \preceq x_i))$
  - $(\forall x_1 \in \mathbf{R}^{d_1}, \dots, x_n \in \mathbf{R}^{d_n}) (F(x_1, \dots, x_n) = \prod F_i(x_i))$
  - $\mu = \mu_1 \times \dots \times \mu_n$
  - $(\forall x_1 \in \mathbf{R}^{d_1}, \dots, x_n \in \mathbf{R}^{d_n}) (f(x_1, \dots, x_n) = \prod f_i(x_i))$

## Independence of infinite collection of random vectors

- infinite collection of random vectors for which every finite subcollection is independent, said to be *independent*

- for independent (countable) collection of random vectors,  $\langle \langle X_{ni} \rangle_{i=1}^{\infty} \rangle_{n=1}^{\infty}$ ,  $\langle \mathcal{F}_n \rangle_{n=1}^{\infty}$  with  $\mathcal{F}_n = \sigma(\langle X_{ni} \rangle_{i=1}^{\infty})$  are independent

## Probability evaluation for two independent random vectors

**Theorem 3. [Probability evaluation for two independent random vectors]** *for independent random vectors,  $X$  and  $Y$ , with distributions,  $\mu$  and  $\nu$ , in  $\mathbf{R}^n$  and  $\mathbf{R}^m$  respectively*

$$\left( \forall B \in \mathcal{R}^{n+m} \right) \left( \mathbf{Prob}((X, Y) \in B) = \int_{\mathbf{R}^n} \mathbf{Prob}((x, Y) \in B) d\mu_X \right)$$

*and*

$$\left( \forall A \in \mathcal{R}^n, B \in \mathcal{R}^{n+m} \right) \left( \mathbf{Prob}(X \in A, (X, Y) \in B) = \int_A \mathbf{Prob}((x, Y) \in B) d\mu_X \right)$$

## Sequence of random variables

**Theorem 4. [sequence of random variables]** *for sequence of probability measures on  $\mathcal{R}$ ,  $\langle \mu_n \rangle$ , exists probability space,  $(X, \Omega, P)$ , and sequence of independent random variables in  $\mathbf{R}$ ,  $\langle X_n \rangle$ , such that each  $X_n$  has  $\mu_n$  as distribution*

## Expected values

**Definition 3. [expected values]** for random variable,  $X$ , on  $(\Omega, \mathcal{F}, P)$ , integral of  $X$  with respect to measure,  $P$

$$\mathbf{E} X = \int X dP = \int_{\Omega} X(\omega) dP$$

called **expected value of  $X$**

- $\mathbf{E} X$  is
  - always defined for nonnegative  $X$
  - for general case
    - defined, or
    - $X$  has an expected value if either  $\mathbf{E} X^+ < \infty$  or  $\mathbf{E} X^- < \infty$  or both, in which case,  $\mathbf{E} X = \mathbf{E} X^+ - \mathbf{E} X^-$
- $X$  is integrable *if and only if*  $\mathbf{E} |X| < \infty$
- limits
  - if  $\langle X_n \rangle$  is dominated by integrable random variable or they are uniformly integrable,  $\mathbf{E} X_n$  converges to  $\mathbf{E} X$  if  $X_n$  converges to  $X$  in probability

## Markov and Chebyshev's inequalities

**Inequality 1. [Markov inequality]** for random variable,  $X$ , on  $(\Omega, \mathcal{F}, P)$ ,

$$\text{Prob}(X \geq \alpha) \leq \frac{1}{\alpha} \int_{X \geq \alpha} X dP \leq \frac{1}{\alpha} \mathbf{E} X$$

for nonnegative  $X$ , hence

$$\text{Prob}(|X| \geq \alpha) \leq \frac{1}{\alpha^n} \int_{|X| \geq \alpha} |X|^n dP \leq \frac{1}{\alpha^n} \mathbf{E} |X|^n$$

for general  $X$

**Inequality 2. [Chebyshev's inequality]** as special case of Markov inequality,

$$\text{Prob}(|X - \mathbf{E} X| \geq \alpha) \leq \frac{1}{\alpha^2} \int_{|X - \mathbf{E} X| \geq \alpha} (X - \mathbf{E} X)^2 dP \leq \frac{1}{\alpha^2} \text{Var } X$$

for general  $X$

## Jensen's, Hölder's, and Lyapunov's inequalities

**Inequality 3. [Jensen's inequality]** for random variable,  $X$ , on  $(\Omega, \mathcal{F}, P)$ , and convex function,  $\varphi$

$$\varphi(\mathbf{E} X) \mathbf{Prob}(X \geq \alpha) \leq \frac{1}{\alpha} \int_{X \geq \alpha} X dP \leq \frac{1}{\alpha} \mathbf{E} X$$

**Inequality 4. [Holder's inequality]** for two random variables,  $X$  and  $Y$ , on  $(\Omega, \mathcal{F}, P)$ , and  $p, q \in (1, \infty)$  with  $1/p + 1/q = 1$

$$\mathbf{E} |XY| \leq (\mathbf{E} |X|^p)^{1/p} (\mathbf{E} |Y|^q)^{1/q}$$

**Inequality 5. [Lyapunov's inequality]** for random variable,  $X$ , on  $(\Omega, \mathcal{F}, P)$ , and  $0 < \alpha < \beta$

$$(\mathbf{E} |X|^\alpha)^{1/\alpha} \leq (\mathbf{E} |X|^\beta)^{1/\beta}$$

- note Hölder's inequality implies Lyapunov's inequality

## Maximal inequalities

**Theorem 5. [Kolmogorov's zero-one law]** *if  $A \in \mathcal{F} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots)$  for independent  $\langle X_n \rangle$ ,*

$$\mathbf{Prob}(A) = 0 \vee \mathbf{Prob}(A) = 1$$

– define  $S_n = \sum X_i$

**Inequality 6. [Kolmogorov's maximal inequality]** *for independent  $\langle X_i \rangle_{i=1}^n$  with  $\mathbf{E} X_i = 0$  and  $\mathbf{Var} X_i < \infty$  and  $\alpha > 0$*

$$\mathbf{Prob}(\max S_i \geq \alpha) \leq \frac{1}{\alpha} \mathbf{Var} S_n$$

**Inequality 7. [Etemadi's maximal inequality]** *for independent  $\langle X_i \rangle_{i=1}^n$  and  $\alpha > 0$*

$$\mathbf{Prob}(\max |S_i| \geq 3\alpha) \leq 3 \max \mathbf{Prob}(|S_i| \geq \alpha)$$



## Moments

**Definition 4. [moments and absolute moments]** for random variable,  $X$ , on  $(\Omega, \mathcal{F}, P)$ , integral of  $X$  with respect to measure,  $P$

$$\mathbf{E} X^n = \int x^n d\mu = \int x^n dF(x)$$

called  $k$ -th moment of  $X$  or  $\mu$  or  $F$ , and

$$\mathbf{E} |X|^n = \int |x|^n d\mu = \int |x|^n dF(x)$$

called  $k$ -th absolute moment of  $X$  or  $\mu$  or  $F$

- if  $\mathbf{E} |X|^n < \infty$ ,  $\mathbf{E} |X|^k < \infty$  for  $k < n$
- $\mathbf{E} X^n$  defined only when  $\mathbf{E} |X|^n < \infty$

## Moment generating functions

**Definition 5. [moment generating function]** for random variable,  $X$ , on  $(\Omega, \mathcal{F}, P)$ ,  $M : \mathbf{C} \rightarrow \mathbf{C}$  defined by

$$M(s) = \mathbf{E} \left( e^{sX} \right) = \int e^{sx} d\mu = \int e^{sx} dF(x)$$

called moment generating function of  $X$

- $n$ -th derivative of  $M$  with respect to  $s$  is  $M^{(n)}(s) = \frac{d^n}{ds^n} F(s) = \mathbf{E} \left( X^n e^{sX} \right) = \int x e^{sx} d\mu$
- thus,  $n$ -th derivative of  $M$  with respect to  $s$  at  $s = 0$  is  $n$ -th moment of  $X$

$$M^{(n)}(0) = \mathbf{E} X^n$$

- for independent random variables,  $\langle X_i \rangle_{i=1}^n$ , moment generating function of  $\sum X_i$

$$\prod M_i(s)$$

# **Convergence of Random Variables**

## Convergences of random variables

**Definition 6. [convergence with probability 1]** *random variables,  $\langle X_n \rangle$ , with*

$$\mathbf{Prob}(\lim X_n = X) = P(\{\omega \in \Omega \mid \lim X_n(\omega) = X(\omega)\}) = 1$$

*said to converge to  $X$  with probability 1 and denoted by  $X_n \rightarrow X$  a.s.*

**Definition 7. [convergence in probability]** *random variables,  $\langle X_n \rangle$ , with*

$$(\forall \epsilon > 0) (\lim \mathbf{Prob}(|X_n - X| > \epsilon) = 0)$$

*said to converge to  $X$  in probability*

**Definition 8. [weak convergence]** *distribution functions,  $\langle F_n \rangle$ , with*

$$(\forall x \text{ in domain of } F) (\lim F_n(x) = F(x))$$

*said to converge weakly to distribution function,  $F$ , and denoted by  $F_n \Rightarrow F$*

**Definition 9. [converge in distribution]** When  $F_n \Rightarrow F$ , associated random variables,  $\langle X_n \rangle$ , said to **converge in distribution** to  $X$ , associated with  $F$ , and denoted by  $X_n \Rightarrow X$

**Definition 10. [weak convergence of measures]** for measures on  $(\mathbf{R}, \mathcal{R})$ ,  $\langle \mu_n \rangle$ , associated with distribution functions,  $\langle F_n \rangle$ , respectively, and measure on  $(\mathbf{R}, \mathcal{R})$ ,  $\mu$ , associated with distribution function,  $F$ , we denote

$$\mu_n \Rightarrow \mu$$

if

$$(\forall A = (-\infty, x] \text{ with } x \in \mathbf{R}) (\lim \mu_n(A) = \mu(A))$$

- indeed, if above equation holds for  $A = (-\infty, x)$ , it holds for many other subsets

## Relations of different types of convergences of random variables

**Proposition 1. [relations of convergence of random variables]** *convergence with probability 1 implies convergence in probability, which implies  $X_n \Rightarrow X$ , i.e.*

$X_n \rightarrow X$  a.s., i.e.,  $X_n$  converge to  $X$  with probability 1

$\Rightarrow X_n$  converge to  $X$  in probability

$\Rightarrow X_n \Rightarrow X$ , i.e.,  $X_n$  converge to  $X$  in distribution,

## Necessary and sufficient conditions for convergence of probability

$X_n$  converge in probability

*if and only if*

$$(\forall \epsilon > 0) (\mathbf{Prob} (|X_n - X| > \epsilon \text{ i.o.}) = \mathbf{Prob} (\limsup |X_n - X| > \epsilon) = 0)$$

*if and only if*

$$\begin{aligned} & (\forall \text{ subsequence } \langle X_{n_k} \rangle) \\ & \left( \exists \text{ its subsequence } \langle X_{n_{k_l}} \rangle \text{ converging to } f \text{ with probability } 1 \right) \end{aligned}$$

## Necessary and sufficient conditions for convergence in distribution

$$X_n \Rightarrow X, \text{ i.e., } X_n \text{ converge in distribution}$$

*if and only if*

$$F_n \Rightarrow F, \text{ i.e., } F_n \text{ converge weakly}$$

*if and only if*

$$(\forall A = (-\infty, x] \text{ with } x \in \mathbf{R}) (\lim \mu_n(A) = \mu(A))$$

*if and only if*

$$(\forall x \text{ with } \mathbf{Prob}(X = x) = 0) (\lim \mathbf{Prob}(X_n \leq x) = \mathbf{Prob}(X \leq x))$$



## Strong law of large numbers

– define  $S_n = \sum_{i=1}^n X_i$

**Theorem 6. [strong law of large numbers]** *for sequence of independent and identically distributed (i.i.d.) random variables with finite mean,  $\langle X_n \rangle$*

$$\frac{1}{n} S_n \rightarrow \mathbf{E} X_1$$

*with probability 1*

- strong law of large numbers also called *Kolmogorov's law*

**Corollary 1. [strong law of large numbers]** *for sequence of independent and identically distributed (i.i.d.) random variables with  $\mathbf{E} X_1^- < \infty$  and  $\mathbf{E} X_1^+ = \infty$  (hence,  $\mathbf{E} X = \infty$ )*

$$\frac{1}{n} S_n \rightarrow \infty$$

*with probability 1*

## Weak law of large numbers

– define  $S_n = \sum_{i=1}^n X_i$

**Theorem 7. [weak law of large numbers]** *for sequence of independent and identically distributed (i.i.d.) random variables with finite mean,  $\langle X_n \rangle$*

$$\frac{1}{n} S_n \rightarrow \mathbf{E} X_1$$

*in probability*

- because convergence with probability 1 implies convergence in probability (Proposition 1), strong law of large numbers implies weak law of large numbers

## Normal distributions

- assume probability space,  $(\Omega, \mathcal{F}, P)$

**Definition 11. [normal distributions]** *Random variable,  $X : \Omega \rightarrow \mathbf{R}$ , with*

$$(A \in \mathcal{R}) \left( \mathbf{Prob}(X \in A) = \frac{1}{\sqrt{2\pi}\sigma} \int_A e^{-(x-c)^2/2} d\mu \right)$$

*where  $\mu = PX^{-1}$  for some  $\sigma > 0$  and  $c \in \mathbf{R}$ , called **normal distribution** and denoted by  $X \sim \mathcal{N}(c, \sigma^2)$*

- note  $\mathbf{E} X = c$  and  $\mathbf{Var} X = \sigma^2$
- called **standard normal distribution** when  $c = 0$  and  $\sigma = 1$

## Multivariate normal distributions

– assume probability space,  $(\Omega, \mathcal{F}, P)$

**Definition 12. [multivariate normal distributions]** *Random variable,  $X : \Omega \rightarrow \mathbf{R}^n$ , with*

$$(A \in \mathcal{R}^n) \left( \mathbf{Prob}(X \in A) = \frac{1}{\sqrt{(2\pi)^n} \sqrt{\det \Sigma}} \int_A e^{-(x-c)^T \Sigma^{-1} (x-c)/2} d\mu \right)$$

where  $\mu = PX^{-1}$  for some  $\Sigma \succ 0 \in \mathbf{S}_{++}^n$  and  $c \in \mathbf{R}^n$ , called *(n-dimensional) normal distribution*, and denoted by  $X \sim \mathcal{N}(c, \Sigma)$

– note that  $\mathbf{E} X = c$  and covariance matrix is  $\Sigma$

## Lindeberg-Lévy theorem

– define  $S_n = \sum^n X_i$

**Theorem 8. [Lindeberg-Levy theorem]** *for independent random variables,  $\langle X_n \rangle$ , having same distribution with expected value,  $c$ , and same variance,  $\sigma^2 < \infty$ ,  $(S_n - nc)/\sigma\sqrt{n}$  converges to standard normal distribution in distribution, i.e.,*

$$\frac{S_n - nc}{\sigma\sqrt{n}} \Rightarrow N$$

*where  $N$  is standard normal distribution*

– Theorem 8 implies

$$S_n/n \Rightarrow c$$

## Limit theorems in $\mathbf{R}^n$

**Theorem 9. [equivalent statements to weak convergence]** *each of following statements are equivalent to weak convergence of measures,  $\langle \mu_n \rangle$ , to  $\mu$ , on measurable space,  $(\mathbf{R}^k, \mathcal{R}^k)$*

- $\lim \int f d\mu_n = \int f d\mu$  for every bounded continuous  $f$
- $\limsup \mu_n(C) \leq \mu(C)$  for every closed  $C$
- $\liminf \mu_n(G) \geq \mu(G)$  for every open  $G$
- $\lim \mu_n(A) = \mu(A)$  for every  $\mu$ -continuity  $A$

**Theorem 10. [convergence in distribution of random vector]** *for random vectors,  $\langle X_n \rangle$ , and random vector,  $Y$ , of  $k$ -dimension,  $X_n \Rightarrow Y$ , i.e.,  $X_n$  converge to  $Y$  in distribution if and only if*

$$\left( \forall z \in \mathbf{R}^k \right) \left( z^T X_n \Rightarrow z^T Y \right)$$

## Central limit theorem

– assume probability space,  $(\Omega, \mathcal{F}, P)$  and define  $\sum^n X_i = S_n$

**Theorem 11. [central limit theorem]** *for random variables,  $\langle X_n \rangle$ , having same distributions with  $\mathbf{E} X_n = c \in \mathbf{R}^k$  and positive definite covariance matrix,  $\Sigma \succ 0 \in \mathcal{S}_k$ , i.e.,  $\mathbf{E}(X_n - c)(X_n - c)^T = \Sigma$ , where  $\Sigma_{ii} < \infty$  (hence  $\Sigma \prec M I_n$  for some  $M \in \mathbf{R}_{++}$  due to Cauchy-Schwarz inequality),*

$$(S_n - nc)/\sqrt{n} \text{ converges in distribution to } Y$$

where  $Y \sim \mathcal{N}(0, \Sigma)$

(proof can be found in [Proof 1](#))

## Convergence of random series

- for independent  $\langle X_n \rangle$ , probability of  $\sum X_n$  converging is either 0 or 1
- below characterize two cases in terms of distributions of individual  $X_n$

**Theorem 12. [convergence with probability 1 for random series]** *for independent  $\langle X_n \rangle$  with  $\mathbf{E} X_n = 0$  and  $\mathbf{Var} X_n < \infty$*

$$\sum X_n \text{ converges with probability 1}$$

**Theorem 13. [convergence conditions for random series]** *for independent  $\langle X_n \rangle$ ,  $\sum X_n$  converges with probability 1 if and only if they converges in probability*

- define truncated version of  $X_n$  by  $X_n^{(c)}$ , i.e.,  $X_n I_{|X_n| \leq c}$

**Theorem 14. [convergence conditions for truncated random series]** *for independent  $\langle X_n \rangle$ ,  $\sum X_n$  converge with probability 1 if all of  $\sum \mathbf{Prob}(|X_n| > c)$ ,  $\sum \mathbf{E}(X_n^{(c)})$ , and  $\sum \mathbf{Var}(X_n^{(c)})$  converge for some  $c > 0$*



# **Selected Proofs**

## Selected proofs

- **Proof 1.** (*Proof for “central limit theorem” on page 54*)

Let  $Z_n(t) = t^T(X_n - c)$  for  $t \in \mathbf{R}^k$  and  $Z(t) = t^TY$ . Then  $\langle Z_n(t) \rangle$  are independent random variables having same distribution with  $\mathbf{E} Z_n(t) = t^T(\mathbf{E} X_n - c) = 0$  and

$$\mathbf{Var} Z_n(t) = \mathbf{E} Z_n(t)^2 = t^T \mathbf{E}(X_n - c)(X_n - c)^T t = t^T \Sigma t$$

Then by Theorem 8  $\sum^n Z_i(t)/\sqrt{nt^T \Sigma t}$  converges in distribution to standard normal random variable. Because  $\mathbf{E} Z(t) = 0$  and  $\mathbf{Var} Z(t) = t^T \mathbf{E} Y Y^T t = t^T \Sigma t$ , for  $t \neq 0$ ,  $Z(t)/\sqrt{t^T \Sigma t}$  is standard normal random variable. Therefore  $\sum^n Z_i(t)/\sqrt{nt^T \Sigma t}$  converges in distribution to  $Z/\sqrt{t^T \Sigma t}$  for every  $t \neq 0$ , thus,  $\sum^n Z_i(t)/\sqrt{n} = t^T(\sum^n X_i - nc)/\sqrt{n}$  converges in distribution to  $Z(t) = t^TY$  for every  $t \in \mathbf{R}$ . Then Theorem 10 implies  $(S_n - nc)/\sqrt{n}$  converges in distribution to  $Y$ . ■

# References

## References

- [Bil95] Patrick Billingsley. *Probability and Measure*. A Wiley-Interscience Publication, 605 Third Avenue, New York, NY 10158-0012, USA, 3rd edition, 1995.

# Index

$\lambda$ -system, [12](#)

$\pi$ - $\lambda$  theorem, [12](#)

Dynkin, Eugene Borisovich, [12](#)

$\pi$ -system, [12](#)

$\sigma$ -algebra

generated by random variables, [23](#)

a.e.

almost everywhere, [6](#)

a.s.

almost surely, [6](#)

absolute moments

random variables, [40](#)

almost everywhere, [6](#)

almost everywhere - a.e., [6](#)

almost surely, [6](#)

Borel functions, [10](#)

Borel sets

multi-dimensional, [10](#)

Borel, Félix Édouard Justin Émile

Borel-Cantelli lemmas, [16](#)

functions, [10](#)

Borel-Cantelli lemmas, [16](#)

first, [16](#)

second, [16](#)

boundary

set, [3](#)

Cantelli, Francesco Paolo

Borel-Cantelli lemmas, [16](#)

CDF, [24](#), [27](#)

central limit theorem, [54](#)

Chebyshev's inequality, [37](#)  
random variables, [37](#)

Chebyshev, Pafnuty  
Chebyshev's inequality  
random variables, [37](#)

closure  
set, [3](#)

complement  
set, [2](#)

complex number, [2](#)

converge in distribution, [44](#)

convergence  
in distribution, [44](#)  
in probability, [43](#)  
necessary and sufficient conditions for  
convergence in distribution, [47](#)  
necessary and sufficient conditions for  
convergence in probability, [46](#)  
of distributions, [43](#)  
of random series, [55](#)  
of random variables, [43–47](#)  
relations of, [45](#)  
weak convergence of distributions, [43](#)  
weak convergence of measures, [44](#)  
with probability 1, [43](#)

convergence conditions for random series, [55](#)

convergence conditions for truncated random series, [55](#)

convergence in distribution of random vector, [53](#)

convergence in probability, [43](#)

convergence with probability 1, [43](#)

convergence with probability 1 for random series, [55](#)

convergence-of-events, [13](#)

corollaries

strong law of large numbers, [48](#)

cumulative distribution function (CDF), [24](#), [27](#)

definitions

almost everywhere - a.e., [6](#)

converge in distribution, [44](#)

convergence in probability, [43](#)

convergence with probability 1, [43](#)

expected values, [36](#)

infinitely often - i.o., [6](#)

moment generating function, [41](#)

moments and absolute moments, [40](#)

multivariate normal distributions, [51](#)

normal distributions, [50](#)

weak convergence, [43](#)

weak convergence of measures, [44](#)

density, [26](#)

difference

set, [3](#)

distribution

probability, [24](#), [27](#)



distribution functions

probability, [24](#), [27](#)

Dynkin's  $\pi$ - $\lambda$  theorem, [12](#)

Dynkin, Eugene Borisovich

$\pi$ - $\lambda$  theorem, [12](#)

equivalent statements to weak convergence, [53](#)

Etemadi's maximal inequality, [39](#)

random variables, [39](#)

Etemadi, Nasrollah

Etemadi's maximal inequality, [39](#)

expected values, [36](#)

random variables, [36](#)

finite sequence, [2](#)

first Borel-Cantelli, [16](#)

Fubini's theorem

product probability spaces, [21](#)

Fubini, Guido

Fubini's theorem

product probability spaces, [21](#)

generated by

$\sigma$ -algebra

by random variables, [23](#)

product probability spaces

$\sigma$ -algebra by measurable rectangles, [18](#)

Hölder's inequality, [38](#)

random variables, [38](#)

Hölder, Ludwig Otto

Hölder's inequality, [38](#)

random variables, [38](#)

Holder's inequality, [38](#)

i.o.

infinitely often, [6](#)

independence

probability spaces, [14](#), [15](#)

random variables, [29–31](#)

infinitely many, [33](#)

random vectors, [32](#)

infinitely many, [33](#)

independence-of-smallest-sig-alg, [13](#)

inequalities

Chebyshev's inequality, [37](#)

Etemadi's maximal inequality, [39](#)

Holder's inequality, [38](#)

Jensen's inequality, [38](#)

Kolmogorov's maximal inequality, [39](#)

Lyapunov's inequality, [38](#)

Markov inequality, [37](#)

infinite sequence, [2](#)

infinitely often, [6](#)

infinitely often - i.o., [6](#)

integer, [2](#)

interior

set, [3](#)

Jensen's inequality, [38](#)

for random variables, [38](#)

Jensen, Johan Ludwig William Valdemar

Jensen's inequality

for random variables, [38](#)

Kolmogorov's law

random variables, [48](#)

Kolmogorov's maximal inequality, [39](#)

random variables, [39](#)

Kolmogorov's zero-one law, [17](#), [39](#)

random variables, [39](#)

Kolmogorov, Andrey Nikolaevich

Kolmogorov's law, [48](#)

Kolmogorov's maximal inequality, [39](#)

Kolmogorov's zero-one law, [17](#), [39](#)

Lévy, Paul

Lindeberg-Lévy theorem, [52](#)

lemmas

first Borel-Cantelli, [16](#)

second Borel-Cantelli, [16](#)

limit theorems

random variables, [53](#)

limits

events, [13](#)

Lindeberg, Jarl Waldemar

Lindeberg-Lévy theorem, [52](#)

Lindeberg-Lévy theorem, [52](#)

Lindeberg-Levy theorem, [52](#)

Lyapunov's inequality, [38](#)

random variables, [38](#)

Lyapunov, Aleksandr

Lyapunov's inequality  
random variables, [38](#)

marginal distribution  
random vectors, [28](#)

Markov inequality, [37](#)  
random variables, [37](#)

Markov, Andrey Andreyevich  
Markov inequality  
random variables, [37](#)

matrix  
positive definite, [4](#)  
positive semi-definite, [4](#)  
symmetric, [4](#)  
trace, [3](#)

maximal inequalities, [39](#)

measurable functions  
abstract measurable spaces, [10](#)

moment generating function, [41](#)

moment generating functions  
random variables, [41](#)

moments  
random variables, [40](#)

moments and absolute moments, [40](#)

multivariate normal distributions, [51](#)

natural number, [2](#)

norm  
vector, [3](#)

normal distributions, [50](#)

random variables, [50](#)

number

complex number, [2](#)

integer, [2](#)

natural number, [2](#)

rational number, [2](#)

real number, [2](#)

PDF, [26](#)

positive definite matrix, [4](#)

positive semi-definite matrix, [4](#)

probability

Kolmogorov's zero-one law, [17](#)

probability (measure) spaces, [11](#)

probability (measure) spaces, [11](#)

probability density function (PDF), [26](#)

probability distribution, [27](#)

probability distribution functions, [24](#), [27](#)

Probability evaluation for two independent random vectors, [34](#)

probability spaces, [11](#)

independence, [14](#), [15](#)

of collection of classes of events, [15](#)

of collection of events, [14](#)

of two events, [14](#)

Kolmogorov's zero-one law, [17](#)

limits

events, [13](#)

probability measure, [11](#)

product spaces, [18](#)

support, [11](#)

tail  $\sigma$ -algebra, [17](#)

tail events, [17](#)

product measure

product probability spaces, [20](#)

product probability spaces, [18](#)

$\sigma$ -algebra generated by measurable rectangles,  
[18](#)

Fubini's theorem, [21](#)

measurable rectangles, [18](#)

product measure, [20](#)

sections of measurable functions, [19](#)

sections of measurable subsets, [19](#)

propositions

relations of convergence of random variables, [45](#)

random variables, [23](#)

$\sigma$ -algebra generated by, [23](#)

absolute moments, [40](#)

CDF, [24](#)

central limit theorem, [54](#)

Chebyshev's inequality, [37](#)

convergence, [43](#)

convergence in distribution, [44](#)

convergence in probability, [43](#)

convergence with probability 1, [43](#)

cumulative distribution function (CDF), [24](#)

density, [26](#)

discrete, [24](#)

distribution, [24](#)

distribution functions, [24](#)

mappings, [25](#)

expected values, [36](#)

Hölder's inequality, [38](#)

- independence, [29–31](#)
  - equivalent statements, [30](#)
  - infinitely many, [33](#)
- Jensen's inequality, [38](#)
- Kolmogorov's law, [48](#)
- law, [24](#)
- limit theorems, [53](#)
- Lindeberg-Lévy theorem, [52](#)
- Lyapunov's inequality, [38](#)
- Markov inequality, [37](#)
- moment generating functions, [41](#)
- moments, [40](#)
- multivariate normal distributions, [51](#)
- necessary and sufficient conditions for  
convergences in distribution, [47](#)
- necessary and sufficient conditions for  
convergences in probability, [46](#)
- normal distributions, [50](#)

- PDF, [26](#)
- probability density function (PDF), [26](#)
- random vectors, [23](#)
- relations of convergences, [45](#)
- standard normal distribution, [50](#)
- strong law of large numbers, [48](#)
- support, [24](#)
- weak convergence of distributions, [43](#)
- weak convergence of measures, [44](#)
- weak law of large numbers, [49](#)

- random vectors, [23](#)
  - CDF, [27](#)
  - central limit theorem, [54](#)
  - cumulative distribution function (CDF), [27](#)
  - discrete, [27](#)
  - distribution, [27](#)
  - distribution functions, [27](#)

- independence, [32](#)
  - equivalent statements, [32](#)
  - infinitely many, [33](#)
- law, [27](#)
- marginal distribution, [28](#)
- rational number, [2](#)
- real number, [2](#)
- relations of convergence of random variables, [45](#)
- relative interior
  - set, [3](#)
- second Borel-Cantelli, [16](#)
- sequence, [2](#)
  - finite sequence, [2](#)
  - infinite sequence, [2](#)

- set
  - boundary, [3](#)
  - closure, [3](#)
  - complement, [2](#)
  - difference, [3](#)
  - interior, [3](#)
  - relative interior, [3](#)
- smallest  $\sigma$ -algebra containing subsets, [3](#)
- sequence of random variables, [35](#)
- standard normal distribution, [50](#)
- strong law of large numbers, [48](#)
  - random variables, [48](#)
- symmetric matrix, [4](#)
- tail  $\sigma$ -algebra, [17](#)



tail events, [17](#)

theorems

central limit theorem, [54](#)

convergence conditions for random series, [55](#)

convergence conditions for truncated random series, [55](#)

convergence in distribution of random vector, [53](#)

convergence with probability 1 for random series, [55](#)

convergence-of-events, [13](#)

equivalent statements to weak convergence, [53](#)

independence-of-smallest-sig-alg, [13](#)

Kolmogorov's zero-one law, [39](#)

Lindeberg-Levy theorem, [52](#)

Probability evaluation for two independent random vectors, [34](#)

sequence of random variables, [35](#)

strong law of large numbers, [48](#)

weak law of large numbers, [49](#)

trace

matrix, [3](#)

vector

norm, [3](#)

weak convergence, [43](#)

weak convergence of measures, [44](#)

weak law of large numbers, [49](#)

random variables, [49](#)

ZZ-todo

0 - apply new comma conventions, [0](#)

1 - convert bullet points to proper theorem, definition, lemma, corollary, proposition, etc., [0](#)

CANCELED - < 2024 0421 - python script  
extracting important list, [0](#)

CANCELED - 2025 0414 - 2 - diagram for  
convergence of random series, [55](#)

DONE - 2024 0324 - change tocpageref and  
funpageref to hyperlink, [0](#)

DONE - 2024 0324 - python script extracting  
figure list → using “list of figures”  
functionality on doc, [0](#)

DONE - 2024 0324 - python script extracting  
theorem-like list → using “list of theorem”  
functionality on doc, [0](#)

DONE - 2024 0324 - python script for converting  
slides to doc, [0](#)

DONE - 2025 0414 - 1 - change mathematicians’  
names, [0](#)