Probability 1

김지수 (Jisu KIM)

확률론 2 (Probability Theory 2), 2025 2nd semester

Probability Spaces

A probability space is a triple $(\Omega, \mathcal{F}, P)$ where Ω is a set of "outcomes," $\mathcal{F}$ is a set of "events," and $P : \mathcal{F} \to [0, 1]$ is a function that assigns probabilities to events.

Definition. Let Ω be a set. A nonempty collection $\mathcal{F}$ of subsets of Ω is called algebra (or field) if
(i) if $A \in \mathcal{F}$ then $\Omega \backslash A \in \mathcal{F}$, and
(ii) if $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$.
$\mathcal{F}$ is called a σ-algebra (or σ-field)
if (i) (ii) and
(iii) $A_1, A_2, \cdots \in \mathcal{F} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Example. $\mathcal{F} = \{\phi, \Omega\}$ trivial $\sigma-$field
$\mathcal{F} = 2^{\Omega} = \{A |\ A \subset \Omega\}$: power set $\Longrightarrow \sigma-$field

Example ([1, Example 1.1.6])**.** $\Omega = \mathbb{Z}$ ={integers}, $\mathcal{F} = \{A \subset \mathbb{Z} |\ |A| < \infty \text{ or } |A^c| < \infty\}$. Then $\mathcal{F}$ is a field but not a $\sigma-$field.

Without P, $(\Omega, \mathcal{F})$ is called a measurable space, i.e., it is a space on which we can put a measure.

Definition. A measure is a nonnegative countably additive set function; that is, for an σ-algebra $\mathcal{F}$, a function $\mu : \mathcal{F} \to [0, \infty]$ is a measure if
(i) $\mu(A) \geq \mu(\phi) = 0$ for all $A \in \mathcal{F}$, and
(ii) For $A_1, A_2, \cdots \in \mathcal{F}$ with $A_i \cap A_j = \phi$ for any $i \neq j$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Definition. (1) $\mu(\Omega) < \infty \Longrightarrow$finite measure
(2) $\mu(\Omega) = 1 \Longrightarrow$probability measure
(3) $\exists$a partition $A_1, A_2, \cdots$ with $\bigcup_{i=1}^{\infty} A_i = \Omega$ and $\mu(A_i) < \infty \Longrightarrow \sigma-$finite measure

Theorem ([1, Theorem 1.1.1])**.** *Let μ be a measure on $(\Omega, \mathcal{F})$.*
(i) Monotonicity. If $A \subset B$ then $\mu(A) \leq \mu(B)$.
(ii) Subadditivity. If $A \subset \bigcup_{i=1}^{\infty} A_i$ then $\mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i)$.
(iii) Continuity from below. $A_n \uparrow A$ (i.e. $A_1 \subset A_2 \subset \cdots$ and $A = \bigcup_{i=1}^{\infty} A_i$) then $\mu(A_i) \uparrow \mu(A)$.
(iv) Continuity from above. $A_n \downarrow A$ (i.e. $A_1 \supset A_2 \supset \cdots$ and $A = \bigcap_{i=1}^{\infty} A_i$) with $\mu(A_1) < \infty$ then $\mu(A_i) \downarrow \mu(A)$.

Definition. Let $\mathcal{A}$ be a class of subsets of Ω. Then $\sigma(\mathcal{A})$ denotes the smallest $\sigma-$algebra that contains $\mathcal{A}$.

For any any $\mathcal{A}$, such $\sigma(\mathcal{A})$ exists and is unique: this is by the following:

(i) If $\mathcal{F}_i, i \in I$ are $\sigma-$fields, then $\bigcap_{i\in I} \mathcal{F}_i$ is a $\sigma-$field.

(ii) If we are given a collection $\mathcal{A}$ of subsets of Ω, then there is a smallest σ-field containing $\mathcal{A}$.

Definition. Borel $\sigma-$field on $\mathbb{R}^d$, denoted by $\mathcal{R}^d$, is the smallest $\sigma-$field containing all open sets.

Theorem ([1, Theorem 1.1.4])**.** *There is a unique measure μ on $(\mathbb{R}, \mathcal{R})$ with*

$$\mu((a, b]) = b - a.$$

Such measure is called Lebesgue measure.

Distribution and Random Variables

Definition. Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ are measurable spaces. A mapping $X : \Omega \to S$ is a measurable map from $(\Omega, \mathcal{F})$ to $(S, \mathcal{S})$ if

$$\text{for all } B \in \mathcal{S},\ X^{-1}(B) := \{\omega \in \Omega :\ X(\omega) \in B\} \in \mathcal{F}.$$

If $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{R}^d)$ (where $\mathcal{R}^d = \mathcal{B}(\mathbb{R}^d)$) and $d > 1$ then X is called a random vector. If $d = 1$, X is called a random variable.

For convenience, we sometimes replace $\mathbb{R}^d$ by $(\mathbb{R}^*)^d = [-\infty, \infty]^d$ and $(\mathcal{R}^*)^d = \mathcal{B}((\mathbb{R}^*)^d)$ and still say random vector (or random variable).

Example. A trivial but useful example of a random variable is indicator function 1_A of a set $A \in \mathcal{F}$:

$$1_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A. \end{cases}$$

If X is a random variable, then X induces a probability measure on $\mathbb{R}$.

Definition. The probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined as $\mu(A) = P(X \in A)$ for all $A \in \mathcal{B}(\mathbb{R})$ is called the distribution of X.

Remark. The distribution can be defined similarly for random vectors.

The distribution of a random variable X is usually described by giving its distribution function.

Definition. The distribution function $F(x)$ of a random variable X is defined as $F(x) = P(X \le x)$.

Theorem ([1, Theorem 1.2.1])**.** *Any distribution function F has the following properties:*
(i) F is nondecreasing.
(ii) $\lim_{n\to\infty} F(x) = 1,\quad \lim_{n\to-\infty} F(x) = 0.$
(iii) F is right continuous. i.e. $\lim_{y\downarrow x} F(y) = F(x)$.
(iv) $P(X < x) = F(x-) = \lim_{y\uparrow x} F(x)$.
(v) $P(X = x) = F(x) - F(x-)$.

Theorem ([1, Theorem 1.2.2])**.** *If F satisfies (i) (ii) (iii) in [1, Theorem 1.2.1], then it is the distribution function of some random variable. That is, there exists a triple $(\Omega, \mathcal{F}, P)$ and a random variable X such that $F(x) = P(X \le x)$.*

Theorem. *If F satisfies (i) (ii) (iii), then there uniquely exists a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that for all $a < b$,*

$$\mu((a,b]) = F(b) - F(a).$$

Definition. If X and Y induce the same distribution μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we say X and Y are equal in distribution. We write

$$X \stackrel{d}{=} Y.$$

Definition. When the distribution function $F(x) = P(X \le x)$ has the form $F(x) = \int_{-\infty}^{x} f(y)dy$, then we say X has the density function f.

Remark. f is not unique, but unique up to Lebesgue measure 0.

Remark. For defining a density function of a given probability measure, we can use the Radon-Nykodym Theorem.

Definition. If F has a property that there exists f such that $F(x) = \int_{-\infty}^{x} f(y)dy$, we call F is absolutely continuous.

Example ([1, Example 1.2.7])**.** Uniform distribution on the Cantor set

Let $C_0 := [0,1]$, and let C_{n+1} be defined by removing middle third open interval of each interval that remains. For example, $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$.

Let $\mathcal{C} := \bigcap_{n=0}^{\infty} C_n$: Cantor set.

Let $F_0 : [0,1] \to \mathbb{R}$ by $F_0(x) = x$

Let $F_{n+1} : [0,1] \to \mathbb{R}$ by $F_{n+1}(x) = \begin{cases} \frac{1}{2}F_n(x) & 0 \le x \le \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{1}{2} + \frac{1}{2}F_n(3x-2) & \frac{2}{3} \le x \le 1 \end{cases}$

Then, $\{F_n\}$ are uniformly Cauchy, so it converges to a continuous function $F : [0,1] \to \mathbb{R}$

F is constant on each excluded middle interval, so if $\exists f$ s.t. $\int_0^x f(t)dt = F(x)$, $f = 0$ on $\mathcal{C}^C$

This is impossible because Cantor set has measure 0

Example ([1, Example 1.2.9])**.** Dense discontinuities. Let $q_1, q_2, \cdots$ be an enumeration of the rationals. Let $\alpha_i > 0$ with $\sum_{i=1}^{\infty} \alpha_i = 1$ and let

$F(x) = \sum_{i=1}^{\infty} \alpha_i 1_{[q_i, \infty)}$

Exercise ([1, Exercise 1.2.3])**.** Show that the number of jumps of distribution is at most countable.

Theorem ([1, Theorem 1.3.1])**.** *Let $\mathcal{A}$ be a collection of sets in $\mathcal{S}$ such that*

(i) $\{\omega \mid X(\omega) \in A\} \in \mathcal{F}$ for all $A \in \mathcal{A}$

(ii) $\sigma(\mathcal{A}) = \mathcal{S}$

Then X is measurable.

Remark. Note that $\{\{X \in B\} \mid B \in \mathcal{S}\}$ is a σ-field. It is the smallest σ-field on Ω that makes X a measurable map.

It is called the σ-field generated by X and denoted by $\sigma(X)$

Example ([1, Exercise 1.3.2])**.** $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{R})$. Possible choices of $\mathcal{A}$ in Theorem 1.3.1 are

$$\mathcal{A} = \begin{cases} \{(-\infty, x] \mid x \in \mathbb{R}\} \\ \{(-\infty, x) \mid x \in \mathbb{R}\} \\ \{(-\infty, x] \mid x \in \mathbb{Q}\} \\ \{(-\infty, x) \mid x \in \mathbb{Q}\} \end{cases}$$

Example ([1, Exercise 1.3.3]). $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{R}^d)$. A useful choice of $\mathcal{A}$ in Theorem 1.3.1 is
$\{(a_1, b_1) \times \cdots \times (a_d, b_d) \mid -\infty < a_i < b_i < \infty\}$ (set of open rectangles)

Theorem ([1, Theorem 1.3.4]). *If $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ are measurable maps, then $f(X)$ is measurable.*

Theorem. *$f : (S, \mathcal{S}) \to (T, \mathcal{T})$ and suppose $\mathcal{S} = \sigma(\text{open sets})$, $\mathcal{T} = \sigma(\text{open sets})$. Then, if f is continuous then f is measurable.*

Theorem ([1, Theorem 1.3.5]). *If $X_1, \cdots, X_n$ are random variables and $f : (\mathbb{R}^n, \mathcal{R}^n) \to (\mathbb{R}, \mathcal{R})$ is measurable, then $f(X_1, \cdots, X_n)$ is a random variable.*

Theorem ([1, Theorem 1.3.6]). *If $X_1, \cdots, X_n$ are random variables then $X_1 + \cdots + X_n$ is a random variable.*

Remark. If X, Y are random variables, then

$$cX \ (c \text{ is scalar}),\ X \pm Y,\ XY,\ \sin(X),\ X^2,\ \cdots,$$

are all random variables.

Theorem ([1, Theorem 1.3.7]). *$\inf_n X_n,\ \sup_n X_n,\ \limsup_n X_n,\ \liminf_n X_n$ are random variables.*

Integration and Expectation

Let μ be a σ-finite measure on $(\Omega, \mathcal{F})$.

Definition. For any predicate $Q(\omega)$ defined on Ω, we say Q is true $(\mu-)$almost everywhere (or a.e.) if $\mu(\{\omega : Q(\omega)\ is\ false\}) = 0$

Step 1.

Definition. φ is a simple function if $\varphi(\omega) = \sum_{i=1}^{n} a_i 1_{A_i}$ with $A_i \in \mathcal{F}$
If φ is a simple function and $\varphi \geq 0$, we let

$$\int \varphi d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$$

Step 2.

Definition. If f is measurable and $f \geq 0$ then we let

$$\int f d\mu = \sup\{\int \varphi d\mu :\ 0 \leq \varphi \leq f \text{ and } \varphi \text{ simple}\}$$

We define the integral of f over the set E:

$$\int_E f d\mu := \int f \cdot 1_E d\mu$$

Step 3.

Definition. We say measurable f is integrable if $\int |f| d\mu < \infty$. Let

$$f^+(x) := f(x) \vee 0, \qquad \text{and} \qquad f^-(x) := (-f(x)) \vee 0,$$

where $a \vee b = \max(a, b)$. We define the integral of f by

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

we can also define $\int f d\mu$ if $\int f^+ d\mu = \infty$ and $\int f^- d\mu < \infty$, or $\int f^+ d\mu < \infty$ and $\int f^- d\mu = \infty$

Theorem ([1, Theorem 1.4.7])**.** *Suppose f and g are integrable.*

(i) If $f \geq 0$ a.e. then $\int f d\mu \geq 0$
(ii) $\forall a \in \mathbb{R}$, $\int af d\mu = a \int f d\mu$
(iii) $\int f + g d\mu = \int f d\mu + \int g d\mu$
(iv) If $g \leq f$ a.e. then $\int g d\mu \leq \int f d\mu$
(v) If $g = f$ a.e. then $\int g d\mu = \int f d\mu$
(vi) $|\int f d\mu| \leq \int |f| d\mu$

Definition. $\|f\|_p = \left(\int |f|^p d\mu\right)^{1/p}$ for $0 < p < \infty$
$\|f\|_\infty = \inf\{M : \ \mu(\{x : \ |f(x)| > M\}) = 0\}$
Notice that $\forall c \in \mathbb{R}$, $\|cf\|_p = |c| \cdot \|f\|_p$

Theorem ([1, Theorem 1.5.1])**.** *Jensen's inequality. Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ is convex, that is, for all $\lambda \in [0, 1]$ and $x, y \in \mathbb{R}$,*

$$\lambda\varphi(x) + (1 - \lambda)\varphi(y) \geq \varphi(\lambda x + (1 - \lambda)y).$$

If μ is a probability measure, and f and $\varphi(f)$ are integrable, then

$$\varphi\left(\int f d\mu\right) \leq \int \varphi(f) d\mu.$$

Theorem ([1, Theorem 1.5.2])**.** *Holder's inequality. If $p, q \in [1, \infty]$ with $1/p + 1/q = 1$, then*

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$

Remark. The special case $p = q = 2$ is called the Cauchy-Schwarz inequality

Definition. We say $f_n \to f$ in measure if $\forall \epsilon > 0$, $\lim_{n\to\infty} \mu\{x : \ |f_n(x) - f(x)| > \epsilon\} = 0$

Theorem ([1, Theorem 1.5.3])**.** *Bounded convergence theorem. Let E be a set with $\mu(E) < \infty$. Suppose f_n vanishes on $E^{\complement}$, $|f_n(x)| \leq M$, and $f_n \to f$ in measure. Then*

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Theorem ([1, Theorem 1.5.5])**.** *Fatou's lemma. If $f_n \geq 0$ then*

$$\liminf_n \int f_n d\mu \geq \int \left(\liminf_n f_n\right) d\mu.$$

Theorem ([1, Theorem 1.5.7])**.** *Monotone convergence theorem. If $f_n \geq 0$ and $f_n \uparrow f$ then*

$$\int f_n d\mu \uparrow \int f d\mu.$$

Theorem ([1, Theorem 1.5.8])**.** *Dominated convergence theorem. If $f_n \to f$ a.e., $|f_n| \leq g$ for all n, and g is integrable, then*

$$\int f_n d\mu \to \int f d\mu.$$

Definition. If X is a random variable on $(\Omega, \mathcal{F}, P)$, we define its expected value to be $\mathbb{E}(X) = \int_\Omega X dP$

We also write $\mathbb{E}(X; A) = \int_A X dP$.

Theorem ([1, Theorem 1.6.4])**.** *Chebyshev's inequality. Suppose* $\varphi : \mathbb{R} \to \mathbb{R}$ *has* $\varphi \geq 0$*, let* $A \in \mathcal{R}$ *and let* $i_A = \inf\{\varphi(y) : \ y \in A\}$

$$i_A P(X \in A) \leq \mathbb{E}(\varphi(X); \ X \in A) \leq \mathbb{E}E\varphi(X)$$

Remark. if $\varphi(x) = x^2$ and $A = \{x : \ |x| \geq a\}$:

$$a^2 P(|X| \geq a) \leq \mathbb{E}X^2.$$

Several techniques of integration

- The pushforward measure of a transformation T is $T_*\mu := \mu(T^{-1}(A))$. The change of variables formula for pushforward measures is

$$\int_\Omega f \circ T d\mu = \int_{T(\Omega)} f dT_*\mu.$$

  Now, consider a probability space $(\Omega, \mathcal{F}, P)$, and consider a measurable map $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ as a transformation. Then the distribution measure μ_X of X is in fact the pushforward measure $\mu_X(A) = P(X \in A) = P(X^{-1}(A))$, and hence the change of variable formula becomes

$$\mathbb{E}_P\left[f(X)\right] = \int_\Omega f(X(\omega)) dP(\omega) = \int_{X(\Omega)} f(x) d\mu_X(x).$$

- For Lebesgue measure λ and Riemann integrable function f, $\int_{[a,b]} f d\lambda$ is the same as the Riemann integral $\int_a^b f(x)dx$.

- $\int f d\delta_x = f(x)$, where δ_x is the Dirac-delta measure, i.e., $\delta_x(A) = I(x \in A)$.

- For a random variable $X \geq 0$,

$$\begin{aligned}
\mathbb{E}_P\left[X\right] &= \int_\Omega X(\omega) dP(\omega) = \int_\Omega \int_{[0, X(\omega)]} dt dP(\omega) \\
&= \int_{\{(\omega,t) \in \Omega \times [0,\infty) : 0 \leq t \leq X(\omega)\}} dt \times dP(\omega) \\
&= \int_0^\infty \int_{\{\omega \in \Omega : X(\omega) \geq t\}} dP(\omega) dt \\
&= \int_0^\infty P(X \geq t) dt.
\end{aligned}$$

Product Measures, Fubini's Theorem

Let $\{(\Omega_i, \mathcal{F}_i, \mu_i)\}_{i=1}^n$ be a sequence of σ-finite measure spaces.

Let $\Omega = \Omega_1 \times \cdots \times \Omega_n = \{(\omega_1, \cdots, \omega_n) | \ \omega_i \in \Omega_i\}$

Let $\mathcal{F} = \mathcal{F}_1 \times \cdots \times \mathcal{F}_n$ =the $\sigma-$field generated by $A_1 \times \cdots \times A_n$, where $A_i \in \mathcal{F}_i$

Then there exists a unique measure μ on $\mathcal{F}$ with

$$\mu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n).$$

Let $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ be two σ-finite measure spaces. Let $\mu = \mu_1 \times \mu_2$.

Theorem ([1, Theorem 1.7.2])**.** *Fubini's theorem. Let f be a mesaurable function. If $f \geq 0$ or $\int |f| d\mu < \infty$, then*

$$\int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = \int_{X\times Y} f d\mu = \int_Y \int_X f(x,y)\mu_1(dx)\mu_2(dy)$$

Example ([1, Example 1.7.5])**.** Let $X = Y = \{1, 2, \cdots\}$ with $\mathcal{A} = \mathcal{B} =$all subsets and $\mu_1 = \mu_2 =$counting measure. let

$$f(m,n) = \begin{cases} 1 & m = n \\ -1 & m = n+1 \\ 0 & o.w \end{cases}$$

Then,

$$\int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = 0 \qquad \text{but} \qquad \int_Y \int_X f(x,y)\mu_1(dx)\mu_2(dy) = 1.$$

$$\begin{array}{cccccc} & \vdots & \vdots & \vdots & \vdots & \\ & 0 & 0 & 0 & 1 & \cdots \\ \uparrow & 0 & 0 & 1 & -1 & \cdots \\ n & 0 & 1 & -1 & 0 & \cdots \\ & 1 & -1 & 0 & 0 & \cdots \\ & & & m & \rightarrow & \end{array}$$

Example ([1, Example 1.7.6])**.** Let $X = (0,1)$, $Y = (1,\infty)$, both equipped with the Borel sets and Lebesque measure. Let $f(x,y) = e^{-xy} - 2e^{-2xy}$

$$\int_0^1 \int_1^\infty f(x,y)dydx = \int_0^1 x^{-1}(e^{-x} - e^{-2x})dx > 0,$$
$$\int_0^1 \int_1^\infty f(x,y)dydx = \int_0^1 x^{-1}(e^{-2x} - e^{-x})dx < 0.$$

Example ([1, Example 1.7.7])**.** Let $X = (0,1)$ with $\mathcal{A} =$the Borel sets and $\mu_1 =$Lebesque measure. Let $Y = (0,1)$ with $\mathcal{B} =$all subsets and $\mu_2 =$counting measure. Let $f(x,y) = 1$ if $x = y$ and 0 otherwise

$$\int_Y f(x,y)\mu_2(dy) = 1 \text{ for all } x \text{ so } \int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = 1,$$
$$\int_X f(x,y)\mu_1(dx) = 0 \text{ for all } y \text{ so } \int_Y \int_X f(x,y)\mu_2(dy)\mu_1(dx) = 0.$$

Independence (독립)

Definition. Let $(\Omega, \mathcal{F}, P)$ be probability space. Two events $A, B \in \mathcal{F}$ are independent (독립) if

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

Two random variables X and Y are independent if for all $C, D \in \mathcal{R}$,

$$P(X \in C,\ Y \in D) = P(X \in C)P(Y \in D).$$

Two σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ ($\subset \mathcal{F}$) are independent if for all $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$, A and B are independent.

Remark. An infinite collection of objects ($\sigma-$fields, random variables, or sets) is said to be independent if every finite subcollection is.

Definition. $\sigma-$fields $\mathcal{F}_1, \cdots, \mathcal{F}_n$ are independent if for all $A_i \in \mathcal{F}_i$,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i),$$

Random variables $X_1, \cdots, X_n$ are independent if for all $B_i \in \mathcal{R}$,

$$P\left(\bigcap_{i=1}^{n} \{X_i \in B_i\}\right) = \prod_{i=1}^{n} P(X_i \in B_i).$$

Sets $A_1, \cdots, A_n$ are independent if for all $I \subset \{1, \cdots, n\}$,

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

Remark. the definition of independent events is not enough to assume pairwise independent, which is $P(A_i \cap A_j) = P(A_i)P(A_j),\ i \neq j$. It is clear that indenendent events are pairwise independent, but converse is not true.

Example. Let $X_1,\ X_2,\ X_3$be independent random variables with $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$ Let $A_1 = \{X_2 = X_3\},\ A_2 = \{X_3 = X_1\}$ and $A_3 = \{X_1 = X_2\}$. These events are pairwise independent but not independent.

Theorem ([1, Theorem 2.1.7])**.** *Suppose $\mathcal{A}_1, \cdots, \mathcal{A}_n$ are independent and $\mathcal{A}_i$ are π-systems. Then $\sigma(\mathcal{A}_1), \cdots, \sigma(\mathcal{A}_n)$ are independent.*

Theorem ([1, Theorem 2.1.8])**.** *$(X_1, \cdots, X_n)$ are independent if and only if for all $x_i \in (-\infty, \infty]$,*

$$P(X_1 \leq x_1, \cdots, X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i)$$

Theorem ([1, Theorem 2.1.9])**.** *Suppose $\mathcal{F}_{i,j},\ 1 \leq i \leq n,\ 1 \leq j \leq m(i)$ are independent and let $\mathcal{G}_i = \sigma(\bigcup_j \mathcal{F}_{i,j})$. Then $\mathcal{G}_1, \cdots, \mathcal{G}_n$ are independent.*

Theorem ([1, Theorem 2.1.10])**.** *If for $1 \leq i \leq n,\ 1 \leq j \leq m(i),\ X_{i,j}$ are independent and $f_i : \mathbb{R}^{m(i)} \to \mathbb{R}$ are measurable then $f_i(X_{i,1}, \cdots, X_{i,m(i)})$ are independent.*

Theorem ([1, Theorem 2.1.11])**.** *Suppose $X_1, \cdots, X_n$ are independent random variables and X_i has distribution μ_i. Then $(X_1, \cdots, X_n)$ has distribution $\mu_1 \times \cdots \times \mu_n$.*

Theorem ([1, Theorem 2.1.12])**.** *Suppose X and Y are independent and have distribution μ and ν. If $h : \mathbb{R}^2 \to \mathbb{R}$ is a measurable function with $h \geq 0$ or $E|h(X, Y)| < \infty$, then*

$$\mathbb{E}h(X, Y) = \int\int h(x, y) d\mu(x) d\nu(y).$$

In particular, when $h(x, y) = f(x)g(y)$ with $f, g \geq 0$ or $\mathbb{E}\,|f(X)|\,, \mathbb{E}\,|g(Y)| < \infty$, then

$$\mathbb{E}_P[f(X)g(Y)] = \mathbb{E}_P[f(X)]\mathbb{E}_P[g(Y)].$$

Theorem ([1, Theorem 2.1.13])**.** *If $X_1, \cdots, X_n$ are independent and have (a) $X_i \geq 0$ for all i, or $\mathbb{E}|X_i| < \infty$ for* *all i, then*

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}X_i.$$

Example ([1, Exercise 2.1.14])**.** It can happen that $\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$ with X and Y are dependent. Suppose joint distribution of X and Y is given by the following table:

		Y		
		1	0	−1
X	1	0	a	0
	0	b	c	b
	−1	0	a	0

where $a, b > 0,\ c \geq 0$ and $2a + 2b + c = 1$. Then $E(XY) = 0 = EXEY$ but

$$P(X = 1, Y = 1) = 0 < ab = P(X = 1)P(Y = 1).$$

Definition. Two random variables X and Y with $\mathbb{E}X^2,\ \mathbb{E}Y^2 < \infty$ that have $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$ are said to be uncorrelated.

Theorem ([1, Theorem 2.1.15])**.** *If X and Y are independent, $F(x) = P(X \leq x)$, and $G(y) = P(Y \leq y)$, then*

$$P(X + Y \leq z) = \int F(z - y)dG(y)$$

The integral on the right-hand side is called the convolution of F and G and is denoted $F * G(z)$

Theorem ([1, Theorem 2.1.16])**.** *Suppose X with density f and Y with distribution function G are independent. Then $X + Y$ has density*

$$h(x) = \int f(x - y)dG(y)$$

When Y has density g, the last formula can be written as

$$h(x) = \int f(x - y)g(y)dy$$

Now, we consider constructing independent random variables.

[1] finite many random variables

Objective : Construct n many independent random variables whose distributions are $F_i,\ i = 1, \cdots, n$

Let $\Omega = \mathbb{R}^n,\ \mathcal{F} = \mathcal{R}^n$ and $X_i(\omega) = X_i(\omega_1, \cdots, \omega_n) = \omega_i$. Then we let

$$P([a_1, b_1] \times \cdots \times [a_n, b_n]) = \prod_{i=1}^{n}(F_i(b_i) - F_i(a_i)).$$

[2] Countably many random variables

Notation. $\Omega = \mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, \cdots) |\ \omega_i \in \mathbb{R}\}$

$\mathcal{B}(\mathbb{R}^{\mathbb{N}}) = \mathcal{R}^{\mathbb{N}}$: the smallest σ-fields generated by collection of finite dimensional rectangles $\{\omega |\ \omega_i \in B_i,\ B_i \in \mathcal{R},\ i = 1, \cdots, n\}$ $n = 1, 2, \cdots$

We want to specify P on $(\mathbb{R}^{\mathbb{N}},\ \mathcal{R}^{\mathbb{N}})$ by specifying P on finite dimensional rectangles

Theorem ([1, Theorem 2.1.21]). *Kolmogorov's extension theorem. Suppose we are given probability measures* μ_n *on* $(\mathbb{R}^n, \mathcal{R}^n)$ *that are consistent, that is,*

$$\mu_{n+1}((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Then, there is a unique probability measure P *on* $(\mathbb{R}^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}})$ *with*

$$P(\omega : \ \omega_i \in (a_i, b_i], \ 1 \le i \le n) = \mu_n((a_1, b_1] \times \cdots \times (a_n, b_n]).$$

Weak laws of large numbers (큰 수의 약법칙)

Various modes of convergence

$\{X_n\}$ and X are random variables defined on $(\Omega, \mathcal{F}, P)$

Definition. $X_n \to X$ almost surely (a.s.) (with probability 1(w.p. 1), almost everywhere(a.e.)) if $P\{\omega : \ X_n(\omega) \to X(\omega)\} = 1$

Equivalent definition : $\forall\epsilon, \ \lim_{m\to\infty} P\{\omega : \ |X_n(\omega) - X(\omega)| \le \epsilon \ \forall n \ge m\} = 1$
or $\forall\epsilon, \ \lim_{m\to\infty} P\{\omega : \ |X_n(\omega) - X(\omega)| > \epsilon \ \forall n \ge m\} = 0$

Definition. $X_n \to X$ in probability (확률수렴) (in pr, $\xrightarrow{p}$) if $\lim_{n\to\infty} P\{|X_n - X| > \epsilon\} = 0$

Theorem. $X_n \to X$ *a.s.* $\Longrightarrow X_n \xrightarrow{p} X$

Remark. $X_n \xrightarrow{p} X \not\Rightarrow X_n \to X$ a.s.

Definition. $X_n \to X$ in L_p, $0 < p < \infty$
if $\lim_{n\to\infty} E(|X_n - X|^p) = 0$ provided $E|X_n|^p < \infty$, $E|X|^p < \infty$.

Theorem. $X_n \to X$ *in* $L_p \Longrightarrow X_n \xrightarrow{p} X$

Theorem. *(Chebyshev inequality, 체비셰프 부등식)*

$$P(|X| \ge \epsilon) \le \frac{E|X|^p}{\epsilon^p}$$

Remark. $X_n \xrightarrow{p} X \not\Rightarrow X_n \to X$ in L_p

Example. $\Omega = [0, 1], \ \mathcal{F} = \mathcal{B}[0, 1], \ P = Unif[0, 1]$
$X(\omega) = 0, \ X_n(\omega) = nI(0 \le \omega \le \frac{1}{n})$
Then $P\{|X_n(\omega) - X(\omega)| > \epsilon\} = P\{0 \le \omega \le \frac{1}{n}\} = \frac{1}{n} \to 0$
But $\mathbb{E}|X_n - X| = E|X_n| = 1$

Theorem. $X_n \xrightarrow{p} X$ *and there exists a random variables* Z *s.t.*
$|X_n| \le Z$ *and* $\mathbb{E}|Z|^p < \infty$
Then $X_n \to X$ *in* L_p.

Remark. If $\mathbb{E}|X| < \infty$, then
$\lim_{n\to\infty} \int_{A_n} |X| dP \to 0$ whenever $P(A_n) \to 0$

L_2 weak law

Theorem ([1, Theorem 2.2.3]). *Let $X_1, X_2, \cdots$ be uncorrelated random variables with $EX_i = \mu$ and $Var(X_i) \le C < \infty$ Let $S_n = \sum_{i=1}^{n} X_i$. Then $\frac{S_n}{n} \to \mu$ in L_2 and also in probability.*

Theorem ([1, Theorem 2.2.14]). *Weak law of large numbers (큰 수의 약법칙, 대수의 약법칙)*

Let $X_1, X_2, \cdots$ be i.i.d. random variables with $E|X_i| < \infty$. Let $S_n = X_1 + \cdots + X_n$ and let $\mu = EX_1$. Then $\frac{S_n}{n} \to \mu$ in probability.

Borel-Canteli lemma

Let $\{A_n\}$ be a sequence of subsets of Ω.

Definition. $\limsup A_n = \lim_{m\to\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega$ that are in infinitely many $A_n\} = \{A_n\ i.o.\}$

$\liminf A_n = \lim_{m\to\infty} \bigcap_{n=m}^{\infty} A_n = \{\omega$ that are in all but finite $A_n\} = \{A_n\ a.b.f.\}$

Theorem ([1, Theorem 2.3.1]). *Borel-Canteli lemma (보렐-칸텔리 보조정리)*

If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then

$$P(A_n\ i.o.) = 0.$$

Theorem ([1, Theorem 2.3.2]). *$X_n \to X$ in probability (확률수렴) if and only if for every subsequence $X_{n(m)}$, there is a further subsequence $X_{n(m_k)}$ that converges almost surely to X.*

Theorem ([1, Theorem 2.3.3]). *For a given sequence $\{y_n\}$ of a topological space, if any subsequence $y_{n(m)}$ has a convergent subsequence $y_{n(m_k)}$ which converges to y, then $y_n \to y$*

Theorem ([1, Theorem 2.3.4]). *If f is continuous and $X_n \to X$ in probability (확률수렴), then $f(X_n) \to f(X)$ in probability (확률수렴). If in addition f is bounded, then $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$.*

Theorem ([1, Theorem 2.3.5]). *Let $\{X_n\}$ be i.i.d. random variables with $E(X_n) = \mu$ and $EX_1^4 < \infty$. If $S_n = \sum_{i=1}^{n} X_i$ then $\frac{S_n}{n} \to \mu$ a.s..*

Example ([1, Example 2.3.6]). $\Omega = [0,1]$ $\mathcal{F} = \mathcal{B}[0,1]$, $P \sim unif(0,1)$. Let $A_n = (0, \frac{1}{n})$, then $P(A_n\ i.o.) = 0$ but $\sum P(A_n) = \infty$.

Theorem ([1, Theorem 2.3.7]). *The second Borel Cantelli lemma*

If A_n are independent, then $\sum P(A_n) = \infty$ implies that

$$P(A_n\ i.o.) = 1.$$

Theorem ([1, Theorem 2.3.8]). *Let X_n be i.i.d. random variables with $E|X_1| = \infty$, then $P\{|X_n| \ge n\ i.o\} = 1$. So if $S_n = X_1 + \cdots + X_n$, then*

$$P\left(\lim \frac{S_n}{n} \in (-\infty, \infty)\right) = 0.$$

Strong Law of Large Numbers (큰 수의 강법칙)

Theorem ([1, Theorem 2.4.1]). *Let $X_1, X_2, \cdots$ be pairwise independent and identically distributed random variables with $E|X_1| < \infty$. Let $\mu = E(X_1)$ and $S_n = \sum_{i=1}^{n} X_i$. Then $\frac{S_n}{n} \to \mu$ a.s..*

Theorem ([1, Theorem 2.4.5]). *Let $X_1, X_2, \cdots$ be i.i.d. with $EX_i^+ = \infty$ and $EX_1^- < \infty$. If $S_n = X_1 + \cdots + X_n$ then $\frac{S_n}{n} \to \infty$ a.s.*

Example ([1, Example 2.4.8]). Empirical distribution functions

Let $X_1, X_2, \cdots \overset{iid}{\sim} F$, and let

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x).$$

1) For given x, $E(F_n(x)) = F(x)$ unbiased
2) For given x, $F_n(x) \to F(x)$ consistency
3) asymptotic efficient?

Theorem ([1, Theorem 2.4.9]). *Glivenko-Cantellli theorem (글리벤코-칸텔리 정리)*

$$\sup_x |F_n(x) - F(x)| \to 0 \text{ a.s. as } n \to \infty.$$

Weak Convergence

We define weak convergence for random variables, but most of the results can be generalized to measurable maps $X_n, X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$, where S is equipped with a metric ρ.

Definition. A sequence of random vectors $\{X_n\}$ converges weakly or converges in distribution (분포수렴) to a limit X ($X_n \Rightarrow X,\ X_n \xrightarrow{w} X,\ X_n \xrightarrow{d} X$) if

$$\mathbb{E}_P\left[g(X_n)\right] \to \mathbb{E}_P\left[g(X)\right], \qquad \text{for all } g \in C_b(\mathbb{R}),$$

where $C_b(\mathbb{R})$ is a set of continuous and bounded functions. We analogously define $P_n \xrightarrow{d} P$ for probability measures $\{P_n\}$ and P, i.e., $\int g(x)dP_n(x) \to \int g(x)dP(x)$ for all $g \in C_b(\mathbb{R})$. We also analogously define $F_n \xrightarrow{d} F$ ($F_n \Rightarrow F,\ F_n \xrightarrow{w} F$) for distribution functions $\{F_n\}$ and F, i.e., $\int g(x)dF_n(x) \to \int g(x)dF(x)$ for all $g \in C_b(\mathbb{R})$.

Theorem ([1, Theorem 3.2.9]). *A sequence of distribution function F_n converges weakly to a limit F if and only if $F_n(y) \to F(y)$ for all continuity points of F.*

Example ([1, Example 3.2.1]). Let $X_1, X_2, \cdots$ be i.i.d. with $P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2}$, and let $S_n = X_1 + \cdots + X_n$. Then

$$F_n(y) = P(S_n/\sqrt{n} \leq y) \to \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \ \forall y \in \mathbb{R}.$$

That is, $F_n \xrightarrow{w} \mathcal{N}(0,1)$.

Example ([1, Example 3.2.3]). Let $X \sim F$ and $X_n = X + \frac{1}{n}$. Then

$$F_n(x) = P(X_n \leq x) = F(x - \frac{1}{n}) \to F(x-).$$

Hence $F_n(x) \to F(x)$ only when $F(x) = F(x-)$, i.e. only if x is a continuity point of F. Still, $X_n \xrightarrow{d} X$.

Example ([1, Example 3.2.4]). Let $X_p \sim \mathit{Geo}(p)$, i.e. $P(X_p \geq m) = (1-p)^{m-1}$. Then

$$P(X_p > \frac{x}{p}) = (1-p)^{\frac{x}{p}} \to e^{-x}, \qquad \text{as } p \to 0.$$

In words, pX_p converges weakly to an exponential distribution.

Theorem. *Scheffe's theorem. Let $\{f_n\}$ be a sequence of densities and let f_∞ be a density. If $f_n \to f_\infty$ pointwisely, then*

$$\|\mu_n - \mu_\infty\|_{TV} := \sup_B |\mu_n(B) - \mu_\infty(B)| \to 0,$$

when μ_n and μ_∞ are probability measure corresponding to f_n and f_∞.

$\|\mu_n - \mu_\infty\|_{TV}$ is called the total variation norm. If $\mu_n \to \mu_\infty$ in the total variation norm, then $\mu_n \xrightarrow{w} \mu_\infty$ (i.e. $F_n \xrightarrow{w} F_\infty$) However, the converse is not true.

Theorem ([1, Theorem 3.2.8])**.** *(Skorohod representation theorem)*

Suppose $F_n \xrightarrow{d} F$. Then, there exists a probability space $(\Omega, \mathcal{F}, P)$, a sequence of random variables $\{Y_n\}$ and a random variables Y on $(\Omega, \mathcal{F}, P)$ so that $Y_n \sim F_n$, $Y \sim F$, and $Y_n \to Y$ a.s..

Theorem ([1, Theorem 3.2.10])**.** *Continuous mapping theorem.*

Let g be a measurable function and $D_g = \{x : g \text{ is continuous at } x\}$. If $X_n \xrightarrow{d} X$ and $P(X \in D_g) = 0$, then $g(X_n) \xrightarrow{d} g(X)$. If in addition g is bounded, then $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$.

Theorem ([1, Theorem 3.2.11])**.** *The following statements are equivalent*

(i) $X_n \xrightarrow{d} X$

(ii) $\forall$open set G, $\liminf P(X_n \in G) \geq P(X \in G)$

(iii) $\forall$closed set G, $\limsup P(X_n \in F) \leq P(X \in F)$

(iv) For all set A with $P(X \in \partial A) = 0$, $\lim P(X_n \in A) = P(X \in A)$, where $\partial A = clA - intA$.

Theorem ([1, Theorem 3.2.12])**.** *Helly's selection theorem*

For every sequence F_n of distribution functions, there exists a subsequence $F_{n(k)}$ and a right continuous nondecreasing function F so that

$$F_{n(k)}(y) \to F(y), \qquad \text{for all continuity points } y \text{ of } F.$$

Remark. The limit may not be a distribution function.

Theorem ([1, Theorem 3.2.13])**.** *Every subsequential limit of Helly's selection theorem is a distribution function if and only if the sequence F_n is tight, i.e., for all $\epsilon > 0$ there exists $M_\epsilon > 0$ so that*

$$\limsup_{n\to\infty} \{1 - F_n(M_\epsilon) + F_n(-M_\epsilon)\} \leq \epsilon.$$

.

Theorem ([1, Theorem 3.2.14])**.** *If there is a $\varphi \geq 0$ so that $\varphi(x) \to \infty$ as $|x| \to \infty$ and*

$$C = \sup_n \int \varphi(x) dF_n(x) < \infty,$$

then F_n is tight.

Exercise ([1, Theorem 3.2.15])**.** Lévy metric for cumulative distribution functions is

$$\rho(F, G) = \inf \{\epsilon : F(x - \epsilon) - \epsilon \leq G(x) \leq F(x + \epsilon) + \epsilon \text{ for all } x \in \mathbb{R}\}.$$

Then $\rho(F_n, F) \to 0$ if and only if $F_n \xrightarrow{d} F$. So, convergence in distribution can be thought as convergence in metric space.

The fact that convergence in distribution comes from a metric immediately implies

Theorem ([1, Theorem 3.2.15]). *If each subsequence of X_n has a further subsequence that converges to X then $X_n \xrightarrow{d} X$.*

Characteristic Functions

Definition. The characteristic function (ch.f.) of a random variable X is defined by

$$\varphi(t) = \mathbb{E}e^{itX} = E(\cos(tX)) + iE(\sin(tX))$$

Theorem ([1, Theorem 3.3.1]). *All characteristic functions have the following properties:*

(a) $\varphi(0) = 1$

(b) $\varphi(-t) = \overline{\varphi(t)}$, *where* $\bar{z} = a - bi$ *if* $z = a + bi$

(c) $|\varphi(t)| \leq 1$

(d) $\varphi(t)$ *is uniformly continuous on* $(-\infty, \infty)$

(e) $\mathbb{E}e^{it(aX+b)} = e^{itb}\varphi(at)$

Theorem ([1, Theorem 3.3.2]). *If X_1 and X_2 are two independent random variables with the ch.f. φ_1 and φ_2, then $X_1 + X_2$ has the ch.f. $\varphi_1(t) \cdot \varphi_2(t)$*

Theorem ([1, Theorem 3.3.11]). *Inversion formula.*

Let $\varphi(t) = \int e^{itX}\mu(dx)$, where μ is a probability measure. If $a < b$, then

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)dt = \mu(a, b) + \frac{1}{2}\mu\{a, b\}.$$

Theorem ([1, Theorem 3.3.14]). *If $\int |\varphi(t)|dt < \infty$, then μ has bounded continuous density f so that*

$$f(y) = \frac{1}{2\pi} \int e^{-ity}\varphi(t)dt.$$

Theorem ([1, Theorem 3.3.17]). *Let μ_n be a sequence of probability measures with the ch.f.s $\{\varphi_n\}$.*

(i) If $\mu_n \Rightarrow \mu_\infty$, then $\varphi_n(t) \to \varphi_\infty(t)$ for all t

(ii) Suppose $\varphi_n(t) \to \varphi(t)$ pointwisely. If φ is continuous at 0, then the associated distributions μ_n is tight, and converges weakly to the probability measure μ_∞ with the ch.f. φ.

Remark. The continuity of φ at 0 implies that that μ_∞ is a probability measure.

Central Limit Theorem (중심극한정리)

Theorem ([1, Theorem 3.4.1]). *Let $X_1, X_2, \cdots$ be i.i.d. with $\mathbb{E}X_i = \mu$ and $Var(X_i) = \sigma^2 > 0$. If $S_n = X_1 + \cdots + X_n$, then*

$$(S_n - n\mu)/(\sqrt{n}\sigma) \xrightarrow{d} \mathcal{N}(0, 1).$$

Theorem ([1, Theorem 3.4.10]). *Lindeberg-Feller theorem*

For each n, let $X_{n,m}$, $1 \leq m \leq n$, be independent random variables with $EX_{n,m} = 0$. Suppose

(i) $\sum_{m=1}^{n} EX_{n,m}^2 \to \sigma^2 > 0$,

(ii) $\forall \epsilon > 0$, $\lim_{n\to\infty} \sum_{m=1}^{n} E(|X_{n,m}|^2 I(|X_{n,m}| > \epsilon)) = 0$.

Then $S_n = X_{n,1} + \cdots + X_{n,n} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ *as* $n \to \infty$.

Remark. We can prove the first CLT using the Lindeberg-Feller theorem.

Exercise ([1, Exercise 3.4.12])**.** Lyapunov's theorem

Let $\{X_{n,m}\}$ be a triangular array of independent random variables satisfying

(i) $\mathbb{E}\,|X_{n,m}|^{2+\delta} < \infty$ for some $\delta > 0$

(ii) $\lim_{n\to\infty} \sum_{m=1}^{n} \mathbb{E}\,|X_{n,m} - \mathbb{E}\,|X_{n,m}||^{2+\delta}\,/s_n^{2+\delta} = 0$, where $s_n^2 = Var(S_n)$

Then $(S_n - \mathbb{E}S_n)/\sqrt{s_n^2} \xrightarrow{d} \mathcal{N}(0,1)$.

Theorem. *(Feller)*

Let $\{X_{n,k}\}$ be an array of independent random variables.

Lindeberg's condition holds if and only if CLT holds and $\max_{1\le k\le n} \sigma_{nk}^2/s_n^2 \to 0$ as $n \to \infty$.

Theorem ([1, Theorem 3.4.14])**.** *Let $X_1, X_2, \cdots$ be i.i.d. and $S_n = X_1 + \cdots + X_n$. Then there exist $a_n,\ b_n > 0$ so that $(S_n - a_n)/b_n \xrightarrow{d} \mathcal{N}(0,1)$ if and only if*

$$y^2 P(|X_1| > y)/\mathbb{E}(|X_1|^2;\ |X_1| \le y) \to 0$$

Theorem ([1, Theorem 3.4.17])**.** *Berry-Essen theorem*

Let $X_1, X_2, \cdots$ be i.i.d. with $EX_i = 0$, $EX_i^2 = \sigma^2$ and $E|X_1|^3 = \rho < \infty$. Let $F_n(x)$ be the distribution function of $(X_1 + \cdots + X_n)/(\sigma\sqrt{n})$ and $\Phi(x)$ be the standard normal distribution. Then

$$\sup_x |F_n(x) - \Phi(x)| \le 3\rho/(\sigma^3\sqrt{n}).$$

Stochastic Order Notation

The classical order notation should be familiar to you already.

1. We say that a sequence $a_n = o(1)$ if $a_n \to 0$ as $n \to \infty$. Similarly, $a_n = o(b_n)$ if $a_n/b_n = o(1)$.
2. We say that a sequence $a_n = O(1)$ if the sequence is eventually bounded, i.e. for all n large, $|a_n| \le C$ for some constant $C \ge 0$. Similarly, $a_n = O(b_n)$ if $a_n/b_n = O(1)$.
3. If $a_n = O(b_n)$ and $b_n = O(a_n)$ then we use either $a_n = \Theta(b_n)$ or $a_n \asymp b_n$.

When we are dealing with random variables we use stochastic order notation.

1. We say that $X_n = o_P(1)$ if for every $\epsilon > 0$, as $n \to \infty$
$$\mathbb{P}(|X_n| \ge \epsilon) \to 0,$$
i.e. X_n converges to zero in probability.
2. We say that $X_n = O_P(1)$ if for every $\epsilon > 0$ there is a finite $C(\epsilon) > 0$ such that, for all n large enough:
$$\mathbb{P}(|X_n| \ge C(\epsilon)) \le \epsilon.$$

The typical use case: suppose we have $X_1, \ldots, X_n$ which are i.i.d. and have finite variance, and we define:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

1. $\hat{\mu} - \mu = o_P(1)$ (Weak Law of Large Number)

2. $\hat{\mu} - \mu = O_P(1/\sqrt{n})$ (Central Limit Theorem)

Proposition. *1.* $X_n \xrightarrow{P} X$ *implies* $X_n \xrightarrow{d} X$*, and this implies* $X_n = O_p(1)$*. Also,* $X_n = o_p(1)$ *implies* $X_n = O_p(1)$.

2. *(a)* $O_p(1) + O_p(1) = O_p(1)$

(b) $O_p(1) + o_p(1) = O_p(1)$

(c) $o_p(1) + o_p(1) = o_p(1)$

(d) $O_p(1) \cdot O_p(1) = O_p(1)$

(e) $O_p(1) \cdot o_p(1) = o_p(1)$

(f) $o_p(1) \cdot o_p(1) = o_p(1)$

References

[1] Rick Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019. Fifth edition.