

## Handout 1: Basic probability tools in asymptotics

Lecturer &amp; author: Georgios Karagiannis

georgios.karagiannis@durham.ac.uk

References: [2, 1]

# 1 Modes of convergence and their relations

## Set-up and notation:

Consider a probability triplet  $(\Omega, \mathcal{F}, P)$ .

Consider random variable  $X : \Omega \rightarrow \mathbb{R}^d$ , where for simplicity we will denote the  $d$ -dimensional random vector as  $X := X(\omega)$ ,  $\forall \omega \in \Omega$ .

Likewise, we define a sequence of random variables  $X_n : \Omega \rightarrow \mathbb{R}^d$ , and for simplicity denote  $X_n := X_n(\omega)$ , for  $n = 1, 2, \dots$ , and  $\forall \omega \in \Omega$ .

The distribution function of r.v.  $X$  is denoted as

$$F_X(x) = P(X \leq x) = P(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Hereafter, the norm  $|\cdot|$  refers to the Euclidean norm; i.e.  $|X| = \sqrt{\sum_{j=1}^d X_j^2}$ , however the results can be generalized.

## Definitions of modes of convergence:

Some modes of convergence are defined below.

**Definition 1.**  $X_n$  converges in distribution to  $X$ , symb. as  $X_n \xrightarrow{D} X$ , iff

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

for all points  $x \in \mathbb{R}^d$  at which  $F_X(x)$  is continuous.

- Other names: converges in law, and weak convergence

**Definition 2.**  $X_n$  converges in probability to  $X$  iff for every  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0 \tag{1.1}$$

It is symbolized as  $X_n \xrightarrow{P} X$ .

- It means: for any  $\epsilon > 0$ , and for any  $\delta > 0$ , there exists  $N_{\epsilon, \delta} > 0$ , where  $P(|X_n - X| < \epsilon) < \delta$

**Definition 3.**  $X_n$  converges in almost surely to  $X$  iff for every  $\epsilon > 0$

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1 \quad (1.2)$$

It is symbolized as  $X_n \xrightarrow{a.s.} X$ .

- Other names: converges with probability 1, and strong convergence

**Definition 4.**  $X_n$  converges in the  $r$ -th mean to  $X$  iff for every  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} E|X_n - X|^r = 0$$

where  $r \in \{1, 2, \dots\}$ . It is symbolized as  $X_n \xrightarrow{r} X$ .

**Definition 5.**  $X_n$  converges in quadratic mean to  $X$  iff

$$\lim_{n \rightarrow \infty} E|X_n - X|^2 = 0 \quad (1.3)$$

It is symbolized as  $X_n \xrightarrow{qm} X$

### Convergence in probability versus almost surely:

To better understand the difference/connection between the  $\xrightarrow{P}$  and  $\xrightarrow{a.s.}$ , we restate the definitions in words.

**convergence in probability  $\xrightarrow{P}$ :** it requires that for every  $\epsilon > 0$  the probability that  $X_n$  is within  $\epsilon$  of  $X$  to tend to 1 as  $n$  tends to infinity

**convergence almost surely  $\xrightarrow{a.s.}$ :** it requires that for every  $\epsilon > 0$  the probability that  $X_k$  STAYS within  $\epsilon$  of  $X$  for all  $k \geq n$  to tend to 1 as  $n$  tends to infinity

The following Lemma shows the distinction between  $\xrightarrow{P}$  and the  $\xrightarrow{a.s.}$ .

**Lemma 6.**  $X_n \xrightarrow{a.s.} X$  iff for every  $\epsilon > 0$

$$P(|X_k - X| < \epsilon, \forall k \geq n) \rightarrow 1, \quad \text{as } n \rightarrow \infty$$

*Proof.* Given as Exercise 8 in the Exercise sheet. □

### Relations between convergence modes:

**Theorem 7.** *Relations between/among different modes of convergence*

$$1. X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X$$

$$2. X_n \xrightarrow{r} X, \text{ for some } r > 0 \implies X_n \xrightarrow{P} X$$

$$3. X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X$$

*Proof.* Given as Exercise 9 in the Exercise sheet. □

**Example.** (★) Consider  $Z \sim U(0, 1)$ , and  $X_n = 2^n 1_{[0, 1/n)}(Z)$ . Check if  $X_n \xrightarrow{r} 0$ ,  $X_n \xrightarrow{a.s.} 0$ , or  $X_n \xrightarrow{P} 0$

**Solution.** It is  $E|X_n|^r = \frac{1}{n} 2^{nr} \rightarrow \infty$ , so  $X_n \not\xrightarrow{r} 0$ . It is  $P(\{\lim X_n = 0\}) = P(\{Z > 0\}) = 1$ , so  $X_n \xrightarrow{a.s.} 0$ . It is  $P(\{|X_n| \geq \epsilon\}) = P(\{X_n = 2^n\}) = P(Z \in [0, 1/n)) = 1/n \rightarrow 0$ , so  $X_n \xrightarrow{P} 0$ .

**Definition 8.** Consider a constant vector  $c \in \mathbb{R}^d$ . We say that  $X$  is a degenerate random variable/vector identically equal to  $c \in \mathbb{R}^d$ , iff  $X(\omega) = c, \forall \omega \in \Omega$  (for every element of the sampling space).

*Note 9.* Mostly, we will use the symbol  $c \in \mathbb{R}^d$  to denote the constant point  $c$ , as well as the degenerate random vector identically equal to  $c$ .

**Proposition 10.** The distribution function of a degenerate random variable  $X$  equal to  $c$  is

$$F_X(x) = \begin{cases} 1 & , x \geq c \\ 0 & , \text{else} \end{cases}$$

*Note 11.* The Theorem 12, together with Theorem 7, implies that  $X_n \xrightarrow{D} c \iff X_n \xrightarrow{P} c$ , if  $c$  is constant.

**Theorem 12.** If  $c \in \mathbb{R}^d$  is a constant, then  $X_n \xrightarrow{D} c \implies X_n \xrightarrow{P} c$

*Proof.* Given as Exercise ?? in the Exercise sheet. □

Exercise sheet (for practice)

Exercises: 8 ; 9 ; ?? ; 3 ; 4 ; 5

## 2 Taylor expansion

We revise the Taylor expansion in many dimensions. For more details see [1].

*Notation 13.* Derivative notation:

- If  $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ , then

$$\dot{f}(x) = \frac{d}{dx} f(x) = \nabla_x f(x)$$

is a  $d \times k$  matrix whose  $(i, j)$ th element is  $\frac{d}{dx_j} f_i(x)$ .

- If  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , then

$$\ddot{f}(x) = \frac{d}{dx} \dot{f}(x)^T$$

is a  $d \times d$  matrix whose  $(i, j)$ th element is

$$[\ddot{f}(x)]_{i,j} = \frac{d^2}{dx_i dx_j} f(x)$$

**Fact 14.** If  $f : \mathbb{R}^d \rightarrow \mathbb{R}^s$ ,  $g : \mathbb{R}^s \rightarrow \mathbb{R}^k$ , and  $h(x) = g(f(x))$  then

$$\dot{h}(x) = \dot{g}(f(x)) \dot{f}(x) \quad (2.1)$$

**Fact 15.** If  $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ ,  $g : \mathbb{R}^s \rightarrow \mathbb{R}^k$ , and  $h(x) = f^T(x)g(x)$  then

$$\dot{h}(x) = g(x)^T \dot{f}(x) + f(x)^T \dot{g}(x)$$

**Theorem 16.** [The Mean Value Theorem] If  $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$  and if  $\dot{f}(x)$  is continuous in the ball  $B_r(x_0) = \{x \in \mathbb{R}^d : |x - x_0| < r\}$ , then for  $|t| < r$ ,

$$f(\underbrace{x_0 + t}_x) = f(x_0) + \left( \int_0^1 \dot{f}(x_0 + ut) du \right) t$$

*Proof.* Let  $h(u) = f(x_0 + ut)$ , so that  $\dot{h}(u) = \dot{f}(x_0 + ut)t$  (from (2.1)). Then,

$$\int_0^1 \dot{f}(x_0 + ut)t du = \int_0^1 \dot{h}(u) du = h(1) - h(0) = f(x_0 + t) - f(x_0)$$

□

**Theorem 17.** [The Taylor's formula (2nd order)] Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  and if  $\ddot{f}(x)$  is continuous in the ball  $B_r(x_0) = \{x \in \mathbb{R}^d : |x - x_0| < r\}$ . Then for  $x = x_0 + t$  where  $|t| < r$ :

$$f(x) = f(x_0) + \dot{f}(x_0)t + t^T \left( \int_0^1 \int_0^1 u \ddot{f}(x_0 + uvt) du dv \right) t$$

*Proof.* [FYI:] Same trick as above by using  $g(v) = t^T \left( \int_0^1 \dot{f}(x_0 + uvt) du \right) \dots$

□

**Notation 18.** If  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , then we denote the partial derivatives

$$\partial_{\underbrace{i_1 \dots i_k}_{\#k}}^{(k)} f(x_0) = \left. \frac{d^k}{dx_{i_1} \dots dx_{i_k}} f(x) \right|_{x=x_0}$$

**Notation 19.** If  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ , we denote as  $f^{(k)}(x; h)$ :

$$f^{(k)}(x; h) = \underbrace{\sum_{i_1=1}^d \dots \sum_{i_k=1}^d}_{\#k} \partial_{i_1 \dots i_k}^{(k)} f(x) \underbrace{h_{i_1} \dots h_{i_k}}_{\#k}$$

E.g.:  $\partial_{i,j}^{(2)} f(x) = \frac{d^2}{dx_1 dx_2} f(x) \Big|_{x=x_0}$  and  $f^{(k)}(x; h) = \sum_{i=1}^d \sum_{j=1}^d \partial_{i,j}^{(2)} f(x) h_i h_j$ .

**Theorem 20.** [The Taylor's formula] Let function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  with continuous partial derivatives of order  $n$  in the ball  $B_r(x_0) = \{x \in \mathbb{R}^d : |x - x_0| < r\}$ . Then the  $n - 1$  order Taylor expansion of  $f(x)$  around  $x_0$  where  $d = x - x_0$  when  $|h| < r$  is

$$f(x) = f(x_0) + \sum_{k=1}^{n-1} f^{(k)}(x_0; h) + R_n(x_0)$$

where the remainder is given in Lagrange's form by

$$R_n(x_0) = \frac{1}{n!} f^{(n)}(x_0 + th; h), \text{ for some } t \in (0, 1)$$

or equivalently in the integral form by

$$R_n(x_0) = \int_0^1 \frac{(1-t)^{n-1}}{(n-1)!} f^{(n)}(x_0 + th; h) dt$$

*Remark 21.* Regarding Theorem 20, if  $\partial_{i_1 \dots i_n}^{(n)} f(x) \leq M$  for  $x \in B_r(x_0)$  and some finite  $M > 0$  it is

$$R_n(x_0) \leq \frac{M}{n!} \|h\|^n$$

and hence the remainder is of order  $R_n(x_0) = O(\|h\|^n)$  or  $R_n(x_0) = o(\|h\|^{n-1})$ .

Exercise sheet (for practice)

Exercises: # 6, 7

### 3 Characteristic functions & other transformations

Characteristic functions provide an alternative way to the probability function for describing a random variable. In fact, it completely determines (see Theorem 23(8)) the behavior and properties of the probability distribution of the random variable  $X$ .

**Definition 22.** The characteristic function of a  $d$  dimensional random variable  $X$  is

$$\varphi_X(t) = E(e^{it^T X})$$

for  $t \in \mathbb{R}^d$ , where  $e^{it^T X} = \cos(t^T X) + i \sin(t^T X)$ .

**Theorem 23.** Some properties of characteristic functions

1.  $\varphi_X(t)$  exists for all  $t \in \mathbb{R}^d$  and is continuous
2.  $\varphi_X(0) = 1$  and  $|\varphi_X(t)| \leq 1$  for all  $t \in \mathbb{R}^d$
3.  $\varphi_{A+BX}(t) = e^{it^T A} \varphi_X(B^T t)$  if  $A \in \mathbb{R}^d$  and  $B \in \mathbb{R}^{k \times d}$  are constants

4.  $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$  iff  $X$  and  $Y$  are independent
5. if  $E|X| < \infty$ , then  $\dot{\varphi}_X(t)$  exists, it is continuous, and  $\dot{\varphi}_X(0) = iE(X)^T$
6. if  $E|X|^2 < \infty$ , then  $\ddot{\varphi}_X(t)$  exists, it is continuous, and  $\ddot{\varphi}_X(0) = -E(X^T X)$
7. if  $X$  is degenerate at  $c \in \mathbb{R}^d$  then  $\varphi_X(t) = e^{it^T c}$
8.  $F_Y(t) = F_X(t) \iff \varphi_Y(t) = \varphi_X(t)$ , for any  $t \in \mathbb{R}^d$
9. if  $X \sim N(\mu, \Sigma)$  then  $\varphi_X(t) = \exp(it^T \mu - \frac{1}{2}t^T \Sigma t)$

*Proof.* Straightforward from the Definition 22. □

**Theorem 24.** [Continuity theorem] Let  $X, X_1, X_2, \dots$  random vectors

$$X_n \xrightarrow{D} X \iff \varphi_{X_n}(t) \rightarrow \varphi_X(t), \text{ for any } t \in \mathbb{R}^d$$

**Example 25.** (★) Show that if  $X \sim \text{Ex}(\lambda)$  then  $\varphi_X(t) = \frac{\lambda}{\lambda - it}$ .

**Solution.** It is

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itX} \underbrace{\lambda e^{-\lambda x} 1(X > 0)}_{=f_{\text{Ex}}(x|\lambda)} dx = \lambda \int_{-\infty}^{\infty} e^{-x(\lambda - it)} dx = \frac{\lambda}{\lambda - it}$$

**Example 26.** (★)

1. Find  $\varphi_X(t)$  if  $X \sim \text{Br}(p)$ .
2. Find  $\varphi_Y(t)$  if  $Y \sim \text{Bin}(n, p)$

**Solution.**

1. It is

$$\varphi_X(t) = \sum_{x=0,1} e^{itX} P(X = x) = e^{it0}(1-p) + e^{it1}p = (1-p) + pe^{it}$$

2. Because Binomial r.v. results as a summation of  $n$  IID Bernoulli r.v., it is  $Y = \sum_{i=1}^n X_i$ , where  $X_i \sim \text{Br}(p)$   $i = 1, \dots, n$  and IID. Then

$$\varphi_Y(t) = \varphi_{\sum X_i}(t) = \prod_{i=1}^n \varphi_{X_i}(t) = ((1-p) + pe^{it})^n$$

## Other Integral transforms

**Definition 27.** The moment generation function of a  $d$  dimensional random variable  $X$  is

$$M_X(t) = \mathbb{E}(e^{t^T X})$$

for  $t \in \mathbb{R}^d$ .

*Remark 28.* It is  $M_X(t) = \phi_X(-it)$ . Hence, its properties can be easily derived.

**Definition 29.** The Cumulant generating function of a  $d$  dimensional random variable  $X$  is the natural logarithm of the moment-generating function

$$K_X(t) = \log(M_X(t)) = \log\left(\mathbb{E}(e^{t^T X})\right)$$

for  $t \in \mathbb{R}^d$ .

*Remark 30.* Properties of the Cumulant generating functions can be easily derived, e.g.  $K_{X+Y}(t) = K_X(t) + K_Y(t)$  iff  $X$  and  $Y$  are independent, etc...

*Note 31.* Some books refer to the Cumulant generating function as the log of the Characteristic function– we do not do this here.

### Exercise sheet (for practice)

Exercises: #10.

For more practice see the examples from

- <https://www.statlect.com/fundamentals-of-probability/characteristic-function>
- <https://www.statlect.com/fundamentals-of-probability/joint-characteristic-function>

## References

- [1] Tom M Apostol. *Mathematical analysis; 2nd ed.* Addison-Wesley Series in Mathematics. Addison-Wesley, Reading, MA, 1974. URL <https://cds.cern.ch/record/105425>.
- [2] Robert J Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.