

# The portmanteau lemma

Alex Fu

2023-01-12

## Lemma 1 (Portmanteau lemma).

Let  $(S, d)$  be a metric space,  $\Sigma$  the Borel  $\sigma$ -algebra induced by  $d$ , and  $\mu_1, \mu_2, \dots, \mu$  probability measures on  $(S, \Sigma)$ . Then the following are equivalent.

- $\mu_n$  converges weakly to  $\mu$ .
- For all bounded continuous functions  $g: S \rightarrow \mathbb{R}$ ,  $\int g d\mu_n \rightarrow \int g d\mu$ .
- For all bounded *Lipschitz* functions  $g: S \rightarrow \mathbb{R}$ ,  $\int g d\mu_n \rightarrow \int g d\mu$ .
- For all open sets  $O$ ,  $\liminf_{n \rightarrow \infty} \mu_n(O) \geq \mu(O)$ .
- For all closed sets  $C$ ,  $\limsup_{n \rightarrow \infty} \mu_n(C) \leq \mu(C)$ .
- For all continuity sets  $A$ ,  $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ . (The *boundary* of a set  $\partial A := \bar{A} - A^\circ$  is its closure minus its interior.  $A$  is a *continuity set* if its boundary is a  $\mu$ -null set:  $\mu(\partial A) = 0$ .)
- For all bounded measurable functions  $f$ , if the set of discontinuity points  $Df$  has zero  $\mu$ -measure, i.e.  $\mu(Df) = 0$ , then  $\int f d\mu_n \rightarrow \int f d\mu$ .

The portmanteau lemma is a useful bag of assorted conditions equivalent to **weak convergence**, or convergence *in distribution*, one of the modes of convergence central to probability theory.

## Definition 1 (Convergence in distribution).

Let  $X_1, X_2, \dots, X$  be random variables with cumulative distribution functions  $F_n, F$  and distributions  $\mu_n, \mu$  respectively. Then  $X_n$  converges **in distribution** to  $X$ , denoted  $X_n \xrightarrow{d} X$ ,  $F_n \xrightarrow{d} F$ , or  $\mu_n \xrightarrow{d} \mu$ , if  $F_n(x) \rightarrow F(x)$  pointwise at all continuity points  $x$  of  $F$ ,

Definition 1 works for  $\mathbb{R}$ - or  $\mathbb{R}^d$ -valued random variables, so it is sufficient for a great deal of purposes. However, other spaces do not necessarily have a notion of cumulative distribution functions, but we may still want to study distributional convergence on those spaces. After all,  $\mathbb{R}^d$  is only one class of particularly nice complete separable metric spaces. Also, almost sure convergence can be defined in general, and convergence in probability for general metric spaces, so it would be strange if the convergence of *distributions*, which exist in general probability spaces, could only be defined for real-valued random variables. As such, we have the following generalization.

**Definition 2** (Weak convergence).

Let  $(S, d)$  be a metric space, and let  $X_n, X$  be  $S$ -valued random variables with distributions  $\mu_n, \mu$  respectively.  $X_n$  converges **weakly** to  $X$ , denoted  $X_n \xrightarrow{w} X$  or  $\mu_n \xrightarrow{w} \mu$ , if for all bounded continuous functions  $g: S \rightarrow \mathbb{R}$ ,  $\mathbb{E}(g(X_n)) \rightarrow \mathbb{E}(g(X))$ .

This may seem like a strange definition at first, but it turns out to be the right definition.

**Proposition 1** (Distributional and weak convergence).

Convergence in distribution and weak convergence coincide for  $\mathbb{R}^d$ -valued random variables.

*Proof.* We will only prove the case of  $d = 1$ . Let  $\mu_n, \mu$  be probability measures (distributions) on  $\mathbb{R}$ . The forward direction is easy given Skorohod's representation theorem: there exist some  $X_n \sim \mu_n, X \sim \mu$  such that  $X_n \rightarrow X$  almost surely, and  $g(X_n) \rightarrow g(X)$  on the same almost sure set by the continuity of  $g$ . By bounded convergence,  $\mathbb{E}(g(X_n)) \rightarrow \mathbb{E}(g(X))$ . The same holds for any  $X_n \stackrel{d}{=} Y_n \sim \mu_n$  and  $Y \sim \mu$ .

For the converse direction, we want to show that  $F_n(x) = \mathbb{E}(\mathbb{1}_{X_n \leq x}) \rightarrow \mathbb{E}(\mathbb{1}_{X \leq x}) = F(x)$ , where the indicators are bounded but not continuous, so we need to approximate the step function  $g(z) = \mathbb{1}_{z \leq x}$  via continuous ones. Let  $g_\varepsilon$  be  $g$ , only linear from  $(x, 1)$  to  $(x + \varepsilon, 0)$ , so that  $g_\varepsilon$  is continuous (piecewise-linear), and  $\mathbb{E}(g_\varepsilon(X_n)) \rightarrow \mathbb{E}(g_\varepsilon(X))$  by hypothesis. As  $g \leq g_\varepsilon$ , we have that

$$\limsup_{n \rightarrow \infty} \mathbb{E}(g(X_n)) \leq \mathbb{E}(g_\varepsilon(X)).$$

To be continued.

□

■