Appendix F

Elements of Empirical Processes

This appendix introduces the empirical measure and the empirical process, and states the most important results on these objects.

The *empirical measure* of the random elements $X_1, \ldots, X_n$ in a sample space $(\mathfrak{X}, \mathscr{X})$ is the random measure given by

$$\mathbb{P}_n(C) = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}(C) = \frac{1}{n}\#\{1 \le i \le n: X_i \in C\}.$$

For a given class $\mathcal{F}$ of measurable functions $f: \mathfrak{X} \to \mathbb{R}$, we also identify the empirical measure with the a map $f \mapsto \mathbb{P}_n(f)$. The corresponding *empirical process* is defined as

$$\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P[f]) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (f(X_i) - P[f]).$$

Here it is assumed that $X_1, \ldots, X_n$ have a common marginal distribution P, and that $P[f]$ exists for every $f \in \mathcal{F}$.

If $X_1, \ldots, X_n$ are i.i.d., then $\mathbb{P}_n(f) \to P[f]$ a.s. by the strong law of large numbers for every f with $P[|f|] < \infty$; if also $P[f^2] < \infty$, then also $\mathbb{G}_n(f) \rightsquigarrow \mathrm{Nor}(0, P\big[(f - Pf)^2\big])$ by the central limit theorem. A class $\mathcal{F}$ for which these theorems are true "uniformly in $\mathcal{F}$" is called Glivenko-Cantelli or Donsker, respectively.

More precisely, a class $\mathcal{F}$ of functions is called *Glivenko-Cantelli* if[1]

$$\sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \to 0, \qquad \text{a.s.}$$

The classical Glivenko-Cantelli theorem is the special case that $X_1, \ldots, X_n$ are real-valued, and $\mathcal{F}$ is the class of all indicator functions of cells $(-\infty, t] \subset \mathbb{R}$.

For the Donsker property it is assumed that the sample paths $f \mapsto \mathbb{G}_n(f)$ are uniformly bounded, and the empirical process is viewed as a map into the metric space $\mathfrak{L}_\infty(\mathcal{F})$ of all uniformly bounded functions $z: \mathcal{F} \to \mathbb{R}$, metrized by the uniform norm. Then $\mathcal{F}$ is defined to be *Donsker* if $\mathbb{G}_n$ tends in distribution in $\mathfrak{L}_\infty(\mathcal{F})$ to a tight, Borel measurable random element in $\mathfrak{L}_\infty(\mathcal{F})$.[2] The limit $\mathbb{G}$ in the Donsker theorem can be identified

[1] In the case that the suprema S_n are not measurable, the almost sure convergence is interpreted in the sense $\mathrm{P}^*\big(\max_{m \ge n} |S_m| > \epsilon\big) \to 0$, for every $\epsilon > 0$, where the asterisk denotes outer probability.

[2] If the empirical process is not Borel measurable, convergence in distribution is understood as convergence of the *outer* expectations $\mathrm{E}^*\psi(\mathbb{G}_n)$ of bounded, measurable functions $\psi: \mathfrak{L}_\infty(\mathcal{F}) \to \mathbb{R}$. cf. Section 1.3 of van der Vaart and Wellner (1996) for details.

from its "marginals" $(\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))$, which are the weak limits of the random vectors $(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k))$. By the multivariate central limit theorem $(\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))$ must be normally distributed with mean vector 0 and covariance kernel

$$\mathrm{cov}\,(\mathbb{G}(f_i), \mathbb{G}(f_j)) = P[f_i f_j] - P[f_i]P[f_j].$$

Thus $\mathbb{G}$ is a Gaussian process; it is known as a *P-Brownian bridge*. The classical Brownian bridge is the special case where P is the uniform measure on the unit interval.

Sufficient conditions for the Glivenko-Cantelli and Donsker properties can be given in terms of the entropy of the class $\mathcal{F}$ and integrability of its envelope function. A measurable function $F: \mathfrak{X} \to \mathbb{R}$ is called an *envelope function* for $\mathcal{F}$ if $|f(x)| \leq F(x)$ for all $f \in \mathcal{F}$ and $x \in \mathfrak{X}$.

The entropy conditions use either $\mathbb{L}_r(P)$-bracketing numbers, or $\mathbb{L}_r(Q)$-covering numbers, for Q ranging over discrete measures. In the latter case it is also necessary to make some measurability assumptions. A class of functions $\mathcal{F}$ is called *P-measurable* if the supremum $\sup_{f\in\mathcal{F}} |\sum_{i=1}^n \xi_i f(X_i)|$ is a measurable variable for every $(\xi_1, \ldots, \xi_n) \in \{0, 1\}^n$. Write $N(\epsilon, \mathcal{F}, \mathbb{L}_r)$ for the *uniform covering number* of a class $\mathcal{F}$:

$$N(\epsilon, \mathcal{F}, \mathbb{L}_r) = \sup_Q N(\epsilon \|F\|_{Q,r}, \mathcal{F}, \mathbb{L}_r(Q)),$$

where the supremum is taken over all finitely discrete probability measures $Q \in \mathfrak{M}(\mathfrak{X})$. For proofs of the following theorems, see e.g. van der Vaart and Wellner (1996), Sections 2.4, 2.5 and 2.6.

Theorem F.1 (Glivenko-Cantelli) *A P-measurable class $\mathcal{F} \subset \mathbb{L}_1(P)$ is Glivenko-Cantelli if and only if $P\|f - Pf\|_{\mathcal{F}} < \infty$ and $n^{-1} \log N(\epsilon, \{f\mathbb{1}_{F\leq M}: f \in \mathcal{F}\}, \mathbb{L}_1(\mathbb{P}_n)) \to_p 0$ for every $\epsilon > 0$ and $M \in (0, \infty)$. Either one the following two conditions is sufficient:*

(i) $N_{[\,]}(\epsilon, \mathcal{F}, \mathbb{L}_1(P)) < \infty$*, for every* $\epsilon > 0$.
(ii) $N(\epsilon, \mathcal{F}, \mathbb{L}_1) < \infty$ *for every* $\epsilon > 0$, $PF < \infty$ *and* $\mathcal{F}$ *is P-measurable.*

Theorem F.2 (Donsker) *Either one of the two conditions is sufficient for $\mathcal{F}$ to be Donsker:*

(i) $\int_0^\infty \sqrt{\log N_{[\,]}(\epsilon, \mathcal{F}, \mathbb{L}_2(P))}\, d\epsilon < \infty$.
(ii) $\int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}, \mathbb{L}_2)}\, d\epsilon < \infty$, $PF^2 < \infty$ *and* $(\mathcal{F} - \mathcal{F})^2$ *and* $\mathcal{F}_\delta := \{f - g: f, g \in \mathcal{F}, \|f - g\|_{P,2} < \delta\}$ *are P-measurable for every* $\delta > 0$.

Another mode of convergence than that provided by the Glivenko-Cantelli theorem is convergence in distribution. In this case the empirical measure is viewed as a random element with values in the space of probability measures $(\mathfrak{M}, \mathscr{M})$ on $(\mathfrak{X}, \mathscr{X})$ equipped with the weak topology. On a Polish sample space the convergence is true without any additional condition.

Proposition F.3 (Weak convergence) *If $(\mathfrak{X}, \mathscr{X})$ is a Polish space with the Borel σ-field, then the empirical distribution $\mathbb{P}_n$ of a random sample from P converges weakly to P, almost surely.*

Proof The Borel measurability of P_n follows from Proposition A.5.

The weak convergence $P_n \rightsquigarrow P$ of a general sequence of measures P_n can be described by the convergence $P_n[f] \to P[f]$ of expectations, for every f in a suitable countable class of bounded, continuous functions. (See e.g. van der Vaart and Wellner 1996, Theorem 1.12.2.) For the empirical measure and a single function this convergence follows from the strong law of large numbers, and hence the proposition follows by countably many applications of the strong law. □

The rate of convergence of maximum likelihood estimators can be characterized through the bracketing entropy integral relative to the Hellinger distance; see Wong and Shen (1995) or van der Vaart and Wellner (1996, 2017), Chapter 3.4 for a proof, and other results of a similar nature.

Theorem F.4 (MLE) *If $\hat{p}_n$ is the maximizer of $p \mapsto \prod_{i=1}^n p(X_i)$ over a collection $\mathcal{P}_n$ of probability densities, for a random sample $X_1 \ldots, X_n \overset{iid}{\sim} p_0 \in \mathcal{P}_n$, then $d_H(\hat{p}_n, p_0) = O_{P_0}(\epsilon_n)$ for any ϵ_n satisfying*

$$\int_0^{\epsilon_n} \sqrt{\log N_{[\,]}(u, \mathcal{P}_n, d_H)}\, du \lesssim \sqrt{n}\epsilon_n^2. \tag{F.1}$$