0.1 Weak Convergence

Beginning of Oct. 19, 2022

Given a sequence of distribution functions F_n, we say $F_n \to F$ **weakly** if $F_n(x) \to F(x)$ at all continuity points of F.

Why continuity points only? Consider $F_n(x) = 1_{[1/n,\infty)}$, which should converge to a point mass at 0, with $F(x) = 1_{[0,\infty)}$. However, $F_n(0) = 0$.

We say $X_n \to X$ **in distribution** if the distribution functions converge weakly. Our previous De Moivre-Laplace theorem then states that $S_n/\sqrt{n}$ converges in distribution to $\mathcal{N}(0,1)$.

Note that this simply says $\mathbb{P}(X_n \in (-\infty, x]) \to \mathbb{P}(X \in (-\infty, x])$ for continuity points x, but does not require $\mathbb{P}(X_n \in A) \to \mathbb{P}(X \in A)$ for all Borel A. We will discuss more on what A satisfies such limit equation.

Example: Geometric r.v.'s. Let X_p be such that $\mathbb{P}(X_p = n) = (1-p)^{n-1}p$. What happens when $p \to 0$?
First note that $\mathbb{E}X_p = 1/p$, so $\mathbb{E}(pX_p) = 1$. Natural question: does pX_p has a limit in distribution?
For fixed x, $x/p \sim \lfloor x/p \rfloor$ (meaning ratio $\to 1$) as $p \to 0$. What about $\mathbb{P}(pX_p > x)$?
First, $\mathbb{P}(X_p > n) = (1-p)^n$ (i.e., first n all tails). Therefore,

$$\mathbb{P}(pX_p > x) = \mathbb{P}(X_p > x/p) = \mathbb{P}(X_p > \lfloor x/p \rfloor) = (1-p)^{\lfloor x/p \rfloor}.$$

Taking log, we obtain

$$\log(1-p)^{\lfloor x/p \rfloor} = \left\lfloor \frac{x}{p} \right\rfloor \log(1-p) \sim \frac{x}{p}(-p) = -x.$$

Therefore $\mathbb{P}(pX_p > x) \to e^{-x}$, an exponential with parameter 1.

Example: Density functions. If $F_n \to F$ weakly, it is not necessarily true that their derivatives $f_n \to f$ weakly.
Consider $f_n = 2$ on $(j - 1/2^n, j/2^n]$ for odd j and 0 for even j. Then F_n almost looks like diagonal and in fact it converges to $F(x) = x$. But clearly $f_n \not\to f \equiv 1$.

Proposition: Scheffe's Theorem

If f_n, f are densities of μ_n and μ, and if $f_n \to f$ pointwise, then $\sup_{B \in \mathcal{B}} |\mu_n(B) - \mu(B)| \to 0$.

Proof. Let $B_n := \{x : f_n(x) > f(x)\}$. Then

$$\sup_{B \in \mathcal{B}} (\mu_n(B) - \mu(B)) = \mu_n(B_n) - \mu(B_n) = \int (f_n - f)^+ \, \mathrm{d}x.$$

Similarly,

$$\sup_{B \in \mathcal{B}} (\mu(B) - \mu_n(B)) = \int (f_n - f)^- \, \mathrm{d}x.$$

Since f_n and f are densities, the two lines above are equal. It suffices to show $\int (f_n - f)^- \, \mathrm{d}x \to 0$ as $n \to \infty$. To do so we use DCT: $(f_n - f)^- \to 0$ a.s. and is bounded by f, so by DCT, the integral converges to 0. ☐

Lemma

For all distribution function F, there exists a random variable Y on $([0,1], \mathcal{B}, \mathbb{P})$ with $\mathbb{P}$ uniform, such that Y has distribution function F.

Proof. If F is continuous and strictly increasing, let $Y(\omega) = F^{-1}(\omega)$. Then $Y(\omega) \leqslant t$ iff $\omega \leqslant F(t)$ iff $\omega \in [0, F(t)]$, so $\mathbb{P}(Y \leqslant t) = \mathbb{P}(Y \in [0, F(t)]) = F(t)$.

More generally, let $Y(\omega) := \sup\{y : F(y) < \omega\}$. Then $Y(\omega) \leqslant t$ iff $\omega \leqslant F(t)$ iff $w \in [0, F(t)]$, and we are done. □

If X_n and Y_n have the same distribution, X and Y have the same distribution, and $X_n \to X$ a.s., is it true that $Y_n \to Y$ a.s.? The answer is no.

Example. Let $X \sim \mathcal{N}(0,1)$, and let $X_n = X$ for all n. Then $X_n \to X$ trivially. Let Y_n be i.i.d. standard normals, and clearly $Y_n \nrightarrow \mathcal{N}(0,1) := Y$.