~~Osama Mahmood~~
Osama Mahmood

D. Weak Convergence

Def. Given Borel probability distributions $\mu, \mu_1, \mu_2, \mu_3 \ldots$ on $\mathbb{R}$ we shall write $\mu_n \Rightarrow \mu$ and say that $\{\mu_n\}$ converges weakly to μ, if $\int_{\mathbb{R}} f \, d\mu_n \to \int_{\mathbb{R}} f \, d\mu$ for all bounded continuous functions $f: \mathbb{R} \to \mathbb{R}$.

Theorem D1: The following statements are equivalent;

(1) $\mu_n \Rightarrow \mu$ ($\{\mu_n\}$ converges weakly to μ)

(2) $\mu_n(A) \to \mu(A)$ for all measurable sets A such that $\mu(\partial A) = 0$

(3) $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ for all $x \in \mathbb{R}$ such that $\mu\{x\} = 0$

(4) (Skorohod's Theorem) there are random variables $Y, Y_1, Y_2 \ldots$ defined jointly on some probability triple, with $\mathcal{L}(Y) = \mu$ and ~~$\mathcal{L}_n(Y$~~ $\mathcal{L}(Y_n) = \mu_n$ for each $n \in \mathbb{N}$ such that $Y_n \to Y$ with probability 1.

(5) $\int_{\mathbb{R}} f \, d\mu_n \to \int_{\mathbb{R}} f \, d\mu$ for all bounded Borel-measurable functions $f: \mathbb{R} \to \mathbb{R}$ such that $\mu(D_f) = 0$, where D_f is the set of points where f is discontinuous.

Proof:

(5) $\to$ (1): follows quickly from the definition of weak convergence.

(5) $\Rightarrow$ (2): This follows by setting $f = 1_A$ so that we get $D_f = \partial A$. Thereby $\mu(D_f) = \mu(\partial A) = 0$. Hence we get that

$$\mu_n(A) = \int_R f d\mu_n \longrightarrow \int_R f d\mu = \mu(A).$$

(1) $\Rightarrow$ (3): Let $\epsilon > 0$ and let f be the function defined by $f(t) = 1$ for $t \le x$ and $f(t) = 0$ for $t \ge x + \epsilon$. And f is linear on the interval $(x, x+\epsilon)$. Therefore we have $1_{(-\infty,x]} \le f \le 1_{(-\infty,x+\epsilon]}$.

Hence,

$$\limsup_{n\to\infty} \mu_n((-\infty,x]) \le \limsup_{n\to\infty} \int f d\mu_n = \int f d\mu \le \mu((-\infty,x+\epsilon]).$$

Since this is true for any $\epsilon > 0$ we conclude that

$$\limsup_{n\to\infty} \mu_n((-\infty,x]) \le \mu((-\infty,x]).$$

Similarly, if we let f be the function defined by $f(t) = 1$ for $t \le x - \epsilon$ and $f(t) = 0$, for $t \ge x$, with f linear on the interval $(x-\epsilon, x)$. Then we will get $1_{(-\infty,x-\epsilon]} \le f \le 1_{(-\infty,x]}$ and we get,

$$\liminf_{n\to\infty} \mu_n((-\infty,x_n]) \ge \liminf_{n\to\infty} \int f d\mu_n = \int f d\mu \ge \mu((-\infty,x-\epsilon]).$$

Since this is true for all $\epsilon > 0$, we get

$$\liminf_{n\to\infty} \mu_n((-\infty,x_n]) \ge \mu((-\infty,x)).$$

Since $\mu\{x\} = 0 \Rightarrow \mu((-\infty,x]) = \mu((-\infty,x))$. We get as desired,

$$\limsup_{n\to\infty} \mu_n((-\infty,x_n]) = \liminf_{n\to\infty} ((-\infty,x_n]) = \mu((-\infty,x]).$$

(3) $\Rightarrow$ (4):

We first define the cummulative distribution functions, by $F_n(x) = \mu_n((-\infty, x])$ and $F(x) = ((-\infty, x])$. Then we let $(\Omega, \mathcal{F}, P)$ be a Lebesgue measure on $[0,1]$ and let $Y_n(w) = \inf\{x ; F_n(x) \geq w\}$ and $Y(w) = \inf\{x ; F(x) \geq w\}$, then as in Theorem (A1) we have $\mathcal{L}(Y_n) = \mu_n$ and $\mathcal{L}(Y) = \mu$.

So, $F(z) < a \implies z \leq Y(a)$

$F(z) \geq b \implies z \geq Y(b)$

Since $\{F_n\} \to F$ at most points it seems that $\{Y_n\} \to Y$ at most points. We will prove that $\{Y_n\} \to Y$ at the points at continuity of Y. Since Y is non-decreasing we can have at-most a countable number of discontinuities: indeed it well at most $m(Y(n+1) - Y(n)) < \infty$ discontinuities of size $\geq 1/m$ within the interval $(n, n+1]$. However since countable sets have measure 0 this implies that $\{Y_n\} \to Y$ with probability 1.

Suppose then that Y is continuous at w and let $y = Y(w)$. For any $\epsilon > 0$ we claim that $F(y-\epsilon) < w < F(y+\epsilon)$. If we had $F(y-\epsilon) = w$ then setting $w = y - \epsilon$ and $b = w$, this would imply that $Y(w) \leq y - \epsilon = Y(w) - \epsilon$, a contradiction. Or if we had $F(y+\epsilon) = w$ then setting $z = y+\epsilon$ and $a = w + \delta$ would imply that $Y(w+\delta) \geq y + \epsilon = Y(w) + \epsilon$ for all $\delta > 0$ contradicting the continuity of Y at w. So we must have $F(y-\epsilon) < w < F(y+\epsilon)$, $\forall \epsilon > 0$.

Now, given that $\epsilon > 0$ we find $\epsilon' > 0$ $0 < \epsilon' < \epsilon$ such that $\mu\{y-\epsilon'\} = \mu\{y+\epsilon'\} = 0$. Then $F_n(y-\epsilon') \to F(y-\epsilon')$ and $F_n(y+\epsilon') \to F(y+\epsilon')$, so $F_n(y-\epsilon') < w < F_n(y+\epsilon')$ for all sufficiently large n. This will imply that $y-\epsilon' \leq Y_n(\omega) \leq y+\epsilon'$ i.e. $|Y_n(\omega) - Y(\omega)| \leq \epsilon' < \epsilon$ for all sufficiently large n.

Hence $Y_n(\omega) \longrightarrow Y(\omega)$.

(4) $\Rightarrow$ (5): We know that if f is continuous at x and we have $\{x_n\} \to x$, then $f(x_n) \to f(x)$. Hence if $\{Y_n\} \to Y$ and $Y \notin D_f$ then $\{f(Y_n)\} \to f(Y)$.

Then we can say that $P[\,f(Y_n) \to f(Y)] \geq P[Y_n \to Y, Y \notin D_f]$.

But by assumption we have $P(Y_n \to Y) = 1$ and $P(Y \notin D_f) = P(Y \in D_f^c) = \mu(D_f^c) = 1$. So $P(f(Y_n) \to f(Y)) = 1$. If f is bounded then by Bounded Convergence Theorem, we get ~~E((f~~ $E[f(Y_n)] \to E[f(Y)]$ i.e. $\int f\, d\mu_n \to \int f\, d\mu$. $\square$

Example, We suppose that $X_1, X_2, X_3 \ldots$ are i.i.d random variables with a finite mean m, and $S = \frac{1}{n}(X_1 + X_2 + X_3 + \ldots + X_n)$. Then the Weak Law of Large Numbers says that $P(S_n \leq m - \epsilon) \to 0$ and $P(S_n \leq m + \epsilon) \to 1$ as $n \to \infty$. This states that $\mathcal{L}(S_n) \Rightarrow \delta_m(\cdot)$, a point mass at m.

Theorem 2: If $\{X_n\} \to X$ in probability, then $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$.

Proof: For any $\epsilon > 0$, if we have $X > z + \epsilon$ and $|X_n - X| < \epsilon$, then we must have $X_n > z$. That is, $\{X > z + \epsilon\} \cap \{|X_n - X| < \epsilon\} \subseteq \{X_n > z\}$. Then taking complements of the statement we get $\{X \leq z + \epsilon\} \cup \{|X_n - X| \geq \epsilon\} \supseteq \{X_n \leq z\}$. Hence we get,

$$P[X_n \leq z] \leq P[X \leq z + \epsilon] + P[|X_n - X| \geq \epsilon].$$

Since $\{X_n\} \to X$ in probability we get that,

$\limsup_{n\to\infty} P(X_n \le z) \le P(X \le z+\epsilon)$. Letting $\epsilon \to 0$ gives

$$\limsup_{n\to\infty} P(X_n \le z) \le P(X \le z).$$

Similarly interchanging X and X_n and replacing z with $z-\epsilon$, in the above gives $P(X \le z-\epsilon) \le P(X_n \le z) + P(|X_n - X| \ge \epsilon)$, or $P(X_n \le z) \ge P(X \le z-\epsilon) - P(|X_n - X| \ge \epsilon)$, so we get $\liminf P(X_n \le z) \ge P(X \le z-\epsilon)$. Letting $\epsilon \to 0$ gives us

$$\liminf_{n\to\infty} P(X_n \le z) \ge P(X < z).$$

If $P(X=z)=0$, we must have $P(X \le z) = P(X < z)$. So should get $\limsup_{n\to\infty} P(X_n \le z) = \liminf_{n\to\infty} (X_n \le z) = P(X \le z)$ □

<u>Theorem 13:</u> Suppose that $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ with $X_n \ge 0$. Then $E(X) \le \liminf E(X_n)$.

<u>Proof:</u> By Skorohod's Theorem, we can find random variables Y_n and Y with $\mathcal{L}(Y_n) = \mathcal{L}(X_n)$, $\mathcal{L}(Y) = \mathcal{L}(X)$ and $Y_n \to Y$ with probability 1. Then by Fatou's Lemma,

$$E(X) = E(Y) = E(\liminf Y_n) \le \liminf[E(Y_n)] = \liminf[E(X_n)] \square$$