Sia μ una misura di probabilità su $\mathbb{R}$. Se consideriamo che, indicando con $I_n = [-n,n]^c =]-\infty,n] \cup]n,+\infty[$

$$\bigcap_{n=1}^{\infty} I_n = \emptyset$$

deduciamo che (è essenziale la σ-additività)

$$\lim_{n\to\infty} \mu(I_n) = 0$$

In altre parole per ogni $\varepsilon > 0$ esiste un K_ε tale che

$$\mu(\{x \in \mathbb{R} : |x| > K_\varepsilon\}) < \varepsilon$$

Nel linguaggio delle variabili casuali possiamo dire che, data una variabile casuale X, per ogni $\varepsilon > 0$ esiste un K_ε per il quale si ha

$$\mathbb{P}(|X| > K_\varepsilon) < \varepsilon$$

Supponiamo che $X_n \xrightarrow{P} X$, cioè $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ per $n \to \infty$. Allora da

$$\kappa < |X_n| \le |X_n - X| + |X|$$

deduciamo che, non potendo essere contemporaneamente $|X_n - X| \le \frac{\kappa}{2}$ e $|X| \le \frac{\kappa}{2}$, abbiamo

$$(|X_n| > \kappa) \subset (|X_n - X| > \tfrac{\kappa}{2}) \cup (|X| > \tfrac{\kappa}{2})$$

da cui

$$\mathbb{P}((|X_n| > \kappa) \le \mathbb{P}(|X_n - X| > \tfrac{\kappa}{2}) + \mathbb{P}(|X| > \tfrac{\kappa}{2}).$$

Da questa formula ricaviamo che fissato un $\varepsilon > 0$, esiste, per l'argomento precedente, un K_ε per il quale si ha $\mathbb{P}(|X| > K_\varepsilon) < \varepsilon$, e quindi scegliendo $\kappa = \kappa_\varepsilon = 2K_\varepsilon$ ed n sufficientemente grande, diciamo $n \ge n_\varepsilon$, in modo che

$$\mathbb{P}(|X_n - X| > \tfrac{\kappa_\varepsilon}{2}) < \varepsilon$$

otteniamo che, per $n \ge n_\varepsilon$,

$$\mathbb{P}((|X_n| > \kappa_\varepsilon) < 2\varepsilon, \qquad \mathbb{P}(|X| > \kappa_\varepsilon) \le \mathbb{P}(|X| > \tfrac{\kappa_\varepsilon}{2}) < \varepsilon.$$

Sia $f\colon \mathbb{R} \to \mathbb{R}$ una funzione continua e limitata ($|f(x)| \leq M$). Supponiamo che $X_n \xrightarrow{P} X$. Allora per ogni $\varepsilon > 0$, consideriamo n_ε e κ_ε come nel punto precedente possiamo considerare

$$\Omega = (|X| > \kappa_\varepsilon) \cup (|X_n| > \kappa_\varepsilon, |X| \leq \kappa_\varepsilon) \cup (|X_n| \leq \kappa_\varepsilon, |X| \leq \kappa_\varepsilon)$$

e quindi

$$\mathbb{E}(|f(X_n) - f(X)|) = \int_\Omega |f(X_n) - f(X)|\, d\mathbb{P} =$$

$$\int\limits_{(|X|>\kappa_\varepsilon)} |f(X_n) - f(X)|\, d\mathbb{P} + \int\limits_{(|X_n|>\kappa_\varepsilon, |X|\leq\kappa_\varepsilon)} |f(X_n) - f(X)|\, d\mathbb{P} +$$

$$+ \int\limits_{(|X_n|\leq\kappa_\varepsilon, |X|\leq\kappa_\varepsilon)} |f(X_n) - f(X)|\, d\mathbb{P}$$

Per il primo integrale possiamo maggiorare con

$$\int\limits_{(|X|>\kappa_\varepsilon)} |f(X_n) - f(X)|\, d\mathbb{P} \leq 2M\ \mathbb{P}(|X| > \kappa_\varepsilon) < 2M\varepsilon.$$

Per il secondo in modo analogo, per $n \geq n_\varepsilon$

$$\begin{aligned}\int\limits_{(|X_n|>\kappa_\varepsilon, |X|\leq\kappa_\varepsilon)} |f(X_n) - f(X)|\, d\mathbb{P} &\leq 2M\ \mathbb{P}(|X_n| > \kappa_\varepsilon, |X| \leq \kappa_\varepsilon) \\ &\leq 2M\ \mathbb{P}(|X_n| > \kappa_\varepsilon) < 2M\varepsilon.\end{aligned}$$

Per il terzo, osserviamo che la funzione f è considerata nel compatto $[-k_\varepsilon, k_\varepsilon]$ dove è anche uniformemente continua: quindi possiamo trovare un δ_ε tale che per $|x - y| \leq \delta_\varepsilon$ si ha $|f(x) - f(y)| \leq \varepsilon$. Da cui

$$\int\limits_{(|X_n|\leq\kappa_\varepsilon, |X|\leq\kappa_\varepsilon)} |f(X_n) - f(X)|\, d\mathbb{P} =$$

$$\int\limits_{(|X_n|\leq\kappa_\varepsilon, |X|\leq\kappa_\varepsilon, |X_n - X|\leq\delta_\varepsilon)} |f(X_n) - f(X)|\, d\mathbb{P} + \int\limits_{(|X_n|\leq\kappa_\varepsilon, |X|\leq\kappa_\varepsilon, |X_n - X|>\delta_\varepsilon)} |f(X_n) - f(X)|\, d\mathbb{P}$$

$$\begin{aligned}&\leq \varepsilon\mathbb{P}((|X_n| \leq \kappa_\varepsilon, |X| \leq \kappa_\varepsilon, |X_n - X| \leq \delta_\varepsilon)+ \\ &\qquad + 2M\ \mathbb{P}((|X_n| \leq \kappa_\varepsilon, |X| \leq \kappa_\varepsilon, |X_n - X| > \delta_\varepsilon) \\ &\leq \varepsilon + 2M\ \mathbb{P}((|X_n - X| > \delta_\varepsilon)\end{aligned}$$

Scegliendo n ancora più grande di n_ε, se necessario abbiamo che

$$\mathbb{P}((|X_n - X| > \delta_\varepsilon) < \varepsilon.$$

e quindi possiamo concludere che per n sufficientemente grande si ha

$$\mathbb{E}(|f(X_n) - f(X)|) \leq (6M + 1)\varepsilon.$$

In altre parole $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X)))$, cioè la convergenza in legge delle rispettive distribuzioni di probabilità.