Probabilities II / Τσέτσας

Example: Let $X \sim N(0,1)$ and $Y = e^X$ (λογαριθμοκανονική). Show that:

i) $E(Y^k) < +\infty \quad \forall k \geq 1$

ii) $M_Y(t) < +\infty \Leftrightarrow t \leq 0$ [we say that $Y \sim \log N(0,1)$]

So it is an example of a distribution which has moments of any order but the moment generating function does not characterize its distribution (it is not finite on a neighbourhood of zero): In fact, if f_{X_0} is the density of $\log N(0,1)$. Then:

$$f_{X_a}(x) = f_{X_0}(x)\left(1 + a\sin(2\pi \log x)\right), \quad x > 0 \quad \text{with } |a| \leq 1.$$

They are different and:

$$E(X_a^n) = E(X_0^n), \text{ so } M_{X_a}(t) \text{ have the same } M_X(t).$$

<u>Solution</u>

$Y = e^X$.

$E(Y^k) = E\left((e^X)^k\right) = E\left[e^{k \cdot X}\right]$, where $X \sim N(0,1)$

But: $M_X(t) = E\left[e^{tX}\right] = e^{t^2/2}, \ \forall t \in \mathbb{R}$,

So: $E(Y^k) = M_X(k) = e^{k^2/2} \quad \forall k = 1, 2, \dots$ and $E(Y^k) < +\infty \quad \forall k = 1, 2, \dots$

ii) $[M_Y(t) < +\infty \Leftrightarrow t \leq 0]$

Let $\underline{t \leq 0}$. $Y = e^X > 0$. So:

$$M_Y(t) = E(e^{t \cdot Y}) \leq E(e^0) = 1 \text{ for all } t \leq 0.$$

Since $t \cdot Y \leq 0 \Rightarrow e^{tY} \leq e^0 = 1$.

So for $t \leq 0$, $M_Y(t) < +\infty$.

Now take $t > 0$ and assume that $M_Y(t) < +\infty$.

Then by the Theorem for $M_Y(t)$ ($< +\infty$, in a neighbourhood of zero)

$$M_Y(t) = \sum_{k=0}^{+\infty} \frac{E(Y^k)}{k!} t^k \overset{i)}{=\!=} \sum_{k=0}^{+\infty} \frac{e^{k^2/2}}{k!} t^k < +\infty$$

But:

$$\frac{e^{k^2/2} t^k}{k!} \geq \frac{e^{k/2} t^k}{k^k} = \left(\frac{e^{k/2} t}{k}\right)^k \quad \text{and} \quad \frac{e^{k/2} t}{k} \xrightarrow[k \to +\infty]{} +\infty$$

$$\Rightarrow \frac{e^{k^2/2} t}{k!} \xrightarrow[k \to +\infty]{} +\infty$$ <u>contradiction!</u> (should $\xrightarrow[k \to +\infty]{} 0$)

Convergence in Distribution (in law)

It is the weakest one, but the most important!

<u>Definition:</u> Let $(\mu_n)_{n \geq 1}$, μ probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We say that $(\mu_n)_{n \geq 1}$ converges weakly (συγκλίνει ασθενώς στο μέτρο μ) to μ and we write:

$\mu_n \Rightarrow \mu$, if $\mu_n((-\infty, x]) \longrightarrow \mu((-\infty, x]) \ \forall x \in \mathbb{R} : \mu(\{x\}) = 0.$

<u>Remark</u>: If the limit exists, then it is unique.

<u>Proof</u>

Assume that $\mu_n \Rightarrow \mu$ and $\mu_n \Rightarrow \nu$. We will show that $\mu = \nu$ (so uniqueness of the limit).

Set $C_\mu = \{x \in \mathbb{R} : \mu(\{x\}) = 0\}$ (continuity set)

$C_\nu = \{x \in \mathbb{R} : \nu(\{x\}) = 0\}$

$\mu_n \Rightarrow \mu$, so $\lim \mu_n(-\infty, x]) = F_{\mu_n}(x) \longrightarrow F_\mu(x) = \mu((-\infty, x]) \ \forall x \in C_\mu$.

$\mu_n \Rightarrow \nu$, so $\lim \mu_n((-\infty, x]) = F_{\mu_n}(x) \longrightarrow F_\nu(x) = \nu((-\infty, x]) \ \forall x \in C_\nu$.

So:

$\forall x \in C_\mu \cap C_\nu$, we have $F_\mu(x) = F_\nu(x)$.

But the discontinuity points of a distribution function ~~form~~ form a countable set. So:

$$(C_\mu \cap C_\nu)^c = \underbrace{C_\mu^c}_{disc\ \mu} \cup \underbrace{C_\nu^c}_{disc\ \nu} \text{ is countable} \rightarrow C_\mu \cap C_\nu \text{ is a dense subset of } \mathbb{R}$$

So: $\forall x \in (C_\mu \cap C_\nu)^c$; $\rightarrow$ σημεία ασυνέχειας

$$F_\mu(x) = \lim_{\substack{x_n \to x^+ \\ x_n \in C_\mu \cap C_\nu}} F_\mu(x_n) \overset{F_\nu = F_\mu \text{ in } C_\mu \cap C_\nu}{=} \lim_{\substack{x_n \to x^+ \\ x_n \in C_\mu \cap C_\nu}} F_\nu(x_n) = F_\nu(x)$$

$\Rightarrow \forall x \in \mathbb{R}: F_\mu(x) = F_\nu(x) \Rightarrow \boxed{\mu = \nu}$

Definition: Let $(X_n)_{n\geq 1}$, X real random variables. We say that (X_n) converges in distribution (or in Law) to X and we write $X_n \xrightarrow{d} X$ (or $X_n \xrightarrow{\mathcal{L}} X$), if $P_{X_n} \Rightarrow P_X$

Remark (Uniqueness of limit in the sense of Distributions): If $X_n \xrightarrow{d} X$ and $X_n \xrightarrow{d} Y \Rightarrow X \stackrel{d}{=} Y$

Theorem: If $(X_n)_{n\geq 1}$, X real random variables

$$X_n \xrightarrow{d} X \Leftrightarrow F_{X_n}(x) \longrightarrow F_X(x) \quad \forall x \text{ continuity point of } F_X.$$

(obvious from the previous statements)

Remark: It would be very strong to demand that this convergence holds for all $x \in \mathbb{R}$.

e.g. $X_n(\omega) = \frac{1}{n} \quad \forall \omega \in \mathbb{R}$

$X(\omega) = 0 \quad \forall \omega \in \mathbb{R}$

We have:

$$\frac{1}{n} \xrightarrow[n\to\infty]{} 0, \text{ so } X_n \xrightarrow{p.w.} X.$$

a.s.

Take $x = 0$ and

$$F_{X_n}(0) = 0 \quad \forall n \geq 1 \text{ and } F_X(0) = 1, \text{ so } F_{X_n}(0) \not\to F_X(0)$$

But for $x \neq 0$, $F_{X_n} \longrightarrow F_X(x)$ and thus $X_n \xrightarrow{d} X$.

<u>Remark</u>: $(X_n)_{n \geq 1}$ and X is not necessary to be defined on the same probability space in contrast to $\left(\xrightarrow{a.s.}, \xrightarrow{P}, \xrightarrow{\mathcal{L}^p} \right)$.

<u>Exercise</u>: Let $(X_n)_{n \geq 1}$ be i.i.d. with $X_1 \sim Exp(1)$. Then if $W_n = \max(X_1, \dots, X_n)$ and $Z_n = W_n - \log n$ $\forall n \geq 1$. Show that:

$Z_n \xrightarrow{d} Z \sim$ standard Gumbel (εφαρμογές σε ασφαλιστικά όταν έχουμε ακραία ενδεχόμενα). with $F_Z(x) = e^{-e^{-x}}$ $\forall x \in \mathbb{R}$.

In general,

$$F(x, \mu, \beta) = e^{-e^{-\frac{x-\mu}{\beta}}}$$ (applications in extreme value theory)

<u>Solution</u>

We will show that: $F_{Z_n}(x) \longrightarrow F_Z(x)$ $\forall x \in \mathbb{R}$, since $F_Z(x)$ is continuous

Let $x \in \mathbb{R}$.

$$F_{Z_n}(x) = P(Z_n \leq x) = P(W_n - \log n \leq x) = P(W_n \leq x + \log n) =$$

$$= P(\max\{X_1, \dots, X_n\} \leq x + \log n) = P(X_1 \leq x + \log n, \dots, X_n \leq x + \log n) \overset{\text{ind}}{=}$$

$$= \prod_{i=1}^{n} P(X_i \leq x + \log n) \overset{iid}{=} P^n(X_1 \leq x + \log n) = F^n_{Exp(1)}(x + \log n)$$

$\overset{x+\log n>0}{\Rightarrow} F_{Z_n}(x) = \left(1 - e^{-1\cdot(x+\log n)}\right)^n$, where $F_{Exp(1)}(x) = 1 - e^{-x}$, $x>0$.

$$= \left(1 - \frac{e^{-x}}{n}\right)^n \xrightarrow[n\to\infty]{} e^{-e^{-x}}$$

<u>Exercise</u>: If $(X_n)_{n\geq 1}$ with $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ with $\mu_n \to \mu \in \mathbb{R}$ and $\sigma_n^2 \to 0$. Show that $X_n \xrightarrow{d} \mu$ ($X = \mu$ with probability 1).

(Εφόσον εκφυλ. είναι 0,1 έχει άλμα ασυνέχειας στο μ)

($F_X(x) = \mathbb{1}_{[\mu, +\infty)}(x)$ check στα continuity points)

<u>Theorem</u>: (characterization of weak convergence): Let $(\mu_n)_{n\geq 1}$, μ probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then:

$$\mu_n \to \mu \leftrightarrow \int f(x)\,d\mu_n(x) \to \int f(x)\,d\mu(x)$$

$\forall f \in C^B(\mathbb{R})$ (continuous and bounded)

<u>Theorem</u>: Let $(X_n)_{n\geq 1}$, X real random variables. Then:

$$X_n \xrightarrow{d} X \Leftrightarrow E[f(X_n)] \to E[f(X)] \quad \forall f \in C^B(\mathbb{R})$$

<u>Example</u>: If $(X_n)_{n\geq 1}$ with $X_n \sim$ discrete uniform $(\{1, 2, \dots, n\})$. Then:

$\frac{X_n}{n} \xrightarrow{d} U \sim Unif(0,1)$.

<u>Proof</u>

<u>Reminder</u>: $\frac{X_n}{n} \sim$ discr Unif $\left(\left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}\right)$

$\frac{X_n}{n} \xrightarrow{d} U \Leftrightarrow E[f(\frac{X_n}{n})] \to E[f(U)] \quad \forall f \in C^B(0,1)$

Let $f \in C^B$. Then: $E[f(\frac{X_n}{n})] = \sum_{x=1}^{n} f(\frac{k}{n}) \cdot P(X_n = x) = \sum_{k=1}^{n} f(\frac{k}{n}) \cdot \frac{1}{n} =$

= ἀθροισμα (sum) Riemann της f στο $(0,1)$ με λεπτότητα $\frac{1}{n}$ με τιμές στο δεξί άκρο $\Rightarrow$

$$E[f(\frac{X_n}{n})] \xrightarrow[n\to\infty]{} \int_0^1 f(u)\,du = \int_0^1 f(u) \cdot 1\,du = E[f(U)].$$

Riemann integral

density of $Unif(0,1)$

<u>Theorem</u>: Let $(\mu_n)_{n\geq 1}, \mu$ be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then

$\mu_n \Rightarrow \mu \Leftrightarrow \lim \mu_n(A) = \mu(A) \quad \forall A \in \mathcal{B}(\mathbb{R}): \mu(\partial A) = 0$, where

σύνορο

$\partial A = \bar{A} \setminus A^\circ$

<u>Proof</u>

"$\Rightarrow$" Notes > Appendix

"$\Leftarrow$" if $x: \mu(\{x\}) = 0$, take $A = (-\infty, x]$. Then $\partial A = \bar{A} \setminus A^\circ = (-\infty, x] \setminus (-\infty, x)$

$\Rightarrow \partial A = \{x\}$.

By assumption: $\mu_n(A) \to \mu(A)$ since $\mu(\partial A) = \mu(\{x\}) = 0$. So:

$\mu_n((-\infty, x]) \to \mu((-\infty, x]) \quad \forall x: \mu(\{x\}) = 0$ and thus: $\mu_n \Rightarrow \mu$.

Theorem: Let $(X_n)_{n\geq 1}$, X be real random variables. Then:

$$X_n \xrightarrow{d} X \Leftrightarrow P_n(X_n \in A) \longrightarrow P(X \in A) \quad \forall A : P(X \in \partial A) = 0$$

(reminder ότι μπορεί να μην ορίζονται στους ίδιους χώρους κλπ)

Theorem: Let $(X_n)_{n\geq 1}$, X real random variables defined on a common prob. space $(\Omega, \mathcal{A}, P)$. Then:

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X.$$

Proof

Assume that $X_n \xrightarrow{P} X$ and $X_n \not\xrightarrow{d} X$.

Recall that: $X_n \xrightarrow{d} X \Leftrightarrow E[f(X_n)] \to E[f(X)] \quad \forall f \in C^B(\mathbb{R})$.

So: $X_n \not\xrightarrow{d} X \Rightarrow \exists f \in C^B(\mathbb{R})$, $\varepsilon > 0$ and an subsequence $(Y_n) \subset (X_n)$ such that:

$$|E[f(Y_n)] - E[f(X)]| \geq \varepsilon \quad \forall n \geq 1. \quad (*)$$

But $X_n \xrightarrow{P} X \Rightarrow Y_n \xrightarrow{P} X \underset{CMT}{\overset{f\text{ cont}}{\Rightarrow}} f(Y_n) \xrightarrow{P} f(X) \underset{\text{BCT (special case of DCT)}}{\overset{f\text{ bounded}}{\Longrightarrow}}$

ounded

$\Rightarrow E[f(Y_n)] \to E[f(X)]$ contradiction by $(*)$

Remark: If (Z_n) subsequence of (Y_n) that satisfies $(*)$ has the property that $Z_n \not\xrightarrow{d} X$.

Conclusion: if $X_n \not\xrightarrow{d} X$, then:

$\exists (Y_n)$ subsequence of (X_n) : $\forall (Z_n) \subset (Y_n)$ we have : $Z_n \not\xrightarrow{d} X$.

So we have a characterization of convergence in distribution.

$$X_n \xrightarrow{d} X \iff \forall (Y_n) \underset{\text{subseq.}}{\subset} (X_n),\ \exists (Z_n) \subset (Y_n) : Z_n \xrightarrow{d} X.$$

Proposition: $X_n \xrightarrow{P} c \iff X_n \xrightarrow{d} c$

c constant

Proof

"$\Rightarrow$" (previous Theorem)

"$\Leftarrow$" let $\varepsilon > 0$. We want to show that:

$$P(|X_n - c| \leq \varepsilon) \xrightarrow[n\to\infty]{} 1.$$

$$P(|X_n - c| \leq \varepsilon) = P(X_n \in [c-\varepsilon, c+\varepsilon]) = P(X_n \leq c+\varepsilon) - P(X_n < c-\varepsilon) =$$

$$= F_{X_n}(c+\varepsilon) - F_{X_n}[(c-\varepsilon)^-]$$

But

$$X_n \xrightarrow{d} c \Rightarrow F_{X_n}(x) \longrightarrow F_c(x) \quad \forall x \neq c.$$

So

$$F_{X_n}(c+\varepsilon) \xrightarrow[n\to\infty]{} F_c(\underset{\substack{\vee \\ c}}{c+\varepsilon}) = 1 \quad \text{and} \quad F_{X_n}((c-\varepsilon)^-) \leq F_{X_n}(c-\varepsilon) \xrightarrow[n\to\infty]{} F_c(c-\varepsilon) = 0$$

Therefore:

$$P(|X_n - c| \leq \varepsilon) \to 1 - 0 = 1.$$

Extension: Convergence in distribution extends also for sequences of random vectors

$$\underset{\mathbb{R}^s}{\underset{\cap}{X_n}} \xrightarrow{d} \underset{\mathbb{R}^s}{\underset{\cap}{X}} \overset{\text{definition}}{\Longleftrightarrow} P(X_n \in A) \longrightarrow P(X \in A) \quad \forall A \in \mathcal{B}(\mathbb{R}^s) : P(X \in \underset{\bar{A} \setminus A^\circ}{\underset{\shortparallel}{\partial A}}) = 0$$

$$X_n \xrightarrow{d} X \Leftrightarrow E[f(X_n)] \to E[f(X)], \ \forall f \in C^{\mathcal{B}}(\mathbb{R}^s; \mathbb{R})$$

Continuous Mapping Theorem ($\xrightarrow{d}$)

$X_n \xrightarrow{d} X \Rightarrow f(X_n) \xrightarrow{d} f(X)$, if f is continuous (the result also holds for random vectors)

Proof

Let $h \in C^{\mathcal{B}}$, we need to show that $E[h(f(X_n))] \to E[h(f(X))]$.

Indeed,

$E[h(f(X_n))] = E[h \circ f(X_n)] \to E[h \circ f(X)]$ because $h \circ f$ is continuous and bounded and $X_n \xrightarrow{d} X$.

Remark: CMT holds also for f to be continuous only on a set $B \in \mathcal{B}(\mathbb{R}^s)$: $P(X \in B) = 1$.

Discussion: $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$. The converse is not true.

Counterexample:

$X_n = Z \sim N(0,1)$

$X = -Z \sim N(0,1)$.

Then $X_n \xrightarrow{d} X$ but $X_n \not\xrightarrow{P} X$, since if $\varepsilon > 0$

$$P(|X_n - X| > \varepsilon) = P(|Z - (-Z)| > \varepsilon) = P(|2Z| > \varepsilon) = P(|\underset{\sim N(0,1)}{Z}| > \tfrac{\varepsilon}{2}) \not\to 0$$

$$\| \\ c > 0$$

<u>Theorem (Slutsky)</u>: If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ constant, then:

i) $X_n + Y_n \xrightarrow{d} X + c$

ii) $X_n \cdot Y_n \xrightarrow{d} c \cdot X$

iii) $\dfrac{X_n}{Y_n} \xrightarrow{d} \dfrac{X}{c}$ (when defined) $c \neq 0$.

(οποτεδήποτε έχουν νόημα αυτοί οι πράξεις μπορεί να ισχύει)

All these results hold for vectors and matrices whenever well defined.

<u>Tightness</u> (Σφιχτότητα)

tight $\equiv$ bounded in probability (φραγμένη κατά πιθανότητα)

<u>Definition</u>: Let $\{\mu_i\}_{i \in I}$ a family of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. It is called <u>tight</u>, if

$$\forall \varepsilon > 0 \;\; \exists M > 0 : \mu_i\left([-M, M]^c\right) < \varepsilon \quad \underline{\forall i \in I}$$

→ ομοιόμορφα.

↳ άρα και αντίστοιχο supremum $< \varepsilon$

Να συγκεντρώσω όλη μάζα πιθανότητας σε ένα οσοδήποτε μεγάλο και αν ήταν, με την έννοια του $1-\varepsilon$.

Θα είχε ενδιαφέρον για εφαρμογή. Μ μπορεί πολύ μεγάλο: τότε μάζα οσοδήποτε μικρή έξω από αυτό το διάστημα ή οσοδήποτε μεγάλη μέσα.

$\Leftrightarrow \mu_i([-M,M]) > 1-\varepsilon \quad \forall i \in I$

So they concentrate uniformly their mass on a certain interval $[-M,M]$ arbitrarily close to 1.

We can say equivalently

$$\lim_{M\to\infty} \sup_{i\in I} \mu_i([-M,M]^c) = 0$$

<u>Definition</u>: Let $\{X_i\}_{i\in I}$ be a family of real random variables. The family is called <u>tight</u> if the family $\{P_{X_i}\}_{i\in I}$ is tight.

<u>Remarks</u>

(1) $\{X_i\}_{i\in I}$ is tight $\Leftrightarrow \lim\limits_{M\to\infty} \sup\limits_{i\in I} P(|X_i|>M) = 0$

$\Leftrightarrow \forall \varepsilon>0, \exists M>0 : P(|X_i|>M) < \varepsilon$

(2) Every real random variable X is tight:

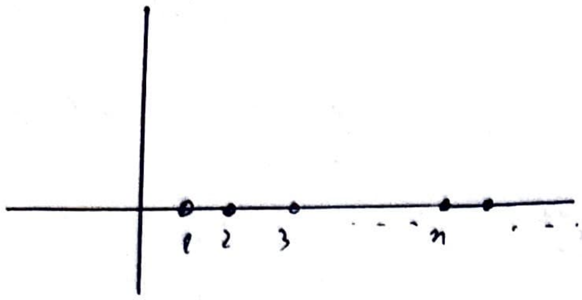
$$\lim_{M\to\infty} P(|X|>M) = \lim_{M\to+\infty} P(X>M) + \lim_{M\to+\infty} P(X<-M)$$

$$= 0 + 0 \quad (1-F(+\infty) + F(-\infty)) = 0$$

(όσο μικρό και αν είναι ε, βρίσκω ώστε $P(|X|>M) < \varepsilon$. $\Leftrightarrow$
$P(|X|\leq M) > 1-\varepsilon$)

(3) If $X_1, X_2, \dots, X_n$ are n real random variables then also $\{X_1,\dots,X_n\}$ is a tight family (Show it).

(4) Of course $X_n = n,\ n \geq 1$ is not tight

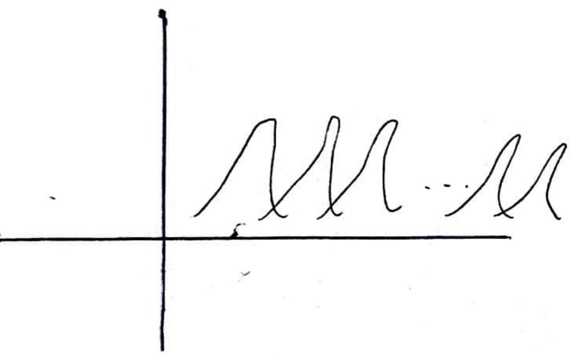


$$\sup_{n\geq 1} \underbrace{P(|X_n| > M)}_{\substack{\| \\ 1\ \forall n > M}} = 1$$

So: $\lim\limits_{M\to\infty} (\ldots) = 1$ (not tight)

Για να έχω σφιχτότητα πρέπει να έχω κάποιου είδους συγκέντρωση των κατανομών

(5) $X_n \sim N(n,1)$



it is not tight (Homework)

Σύγκλιση κατά κατανομή συνεπάγεται την κατανομή

<u>Proposition</u>: Let $\{\mu_n\}_{n\geq 1}, \mu$ probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

$\mu_n \Rightarrow \mu$, then $\{\mu_n\}$ is tight.

For random variables:

$X_n \xrightarrow{d} X \Rightarrow (X_n)_{n\geq 1}$ is tight

<u>Theorem</u> (Prokhorov): Let $(X_n)_{n\geq 1}$ be a sequence of real random variables.

$(X_n)_{n\geq 1}$ tight $\Rightarrow \exists (Y_n) \underset{\text{subs}}{\subset} (X_n)$ and Y real random variable: $Y_n \xrightarrow{d} Y$

(Αντίστοιχο θεωρ Weierstrass)

(it is the analogue of Bolzano-Weierstrass for random variables)

<u>Exercise</u> (characterization of tightness)

$(X_n)_{n\geq 1}$ is tight $\Longleftrightarrow$ every subsequence $(Y_n) \subset (X_n)$ has a convergent subsequence $(Z_n) \subset (Y_n)$ ($Z_n \xrightarrow{d} Z$, for some Z real r.v.)

(Homework)

↳ See notes 2017 Lecture 22

Weak Convergence of Probability Measures · Bellingsley