Il teorema del Limite Centrale

Teorema 1 (Portmanteau Theorem)**.** *Se* $\mathbb{P}_1, \mathbb{P}_2, \dots$ *e* $\mathbb{P}$ *sono probabilità su* $\mathbb{R}$*, le seguenti sono equivalenti:*

(1) $\mathbb{P}_n((-\infty, x]) \to \mathbb{P}((-\infty, x])$ *per ogni* x *punto di continuità della funzione di ripartizione di* $\mathbb{P}$

(2) $\int f\, d\mathbb{P}_n \to \int f\, d\mathbb{P}$ *per ogni* $f : \mathbb{R} \to \mathbb{R}$ *continua limitata*

(3) $\int f\, d\mathbb{P}_n \to \int f\, d\mathbb{P}$ *per ogni* $f : \mathbb{R} \to \mathbb{R}$ *Lipschitziana limitata*

(4) $\liminf_{n\to\infty} \int f\, d\mathbb{P}_n \geq \int f\, d\mathbb{P}$ *per ogni* $f : \mathbb{R} \to \mathbb{R}$ *continua nonnegativa*

(5) $\liminf_{n\to\infty} \mathbb{P}_n(G) \geq \mathbb{P}(G)$ *per ogni* G *aperto*

(6) $\limsup_{n\to\infty} \mathbb{P}_n(F) \leq \mathbb{P}(F)$ *per ogni* F *chiuso*

(7) $\mathbb{P}_n(B) \to \mathbb{P}(B)$ *per ogni boreliano* B *con* $\mathbb{P}(\partial B) = 0$*.*

Dimostrazione. Sia F la funzione di ripartizione di $\mathbb{P}$.
(5) $\Leftrightarrow$ (6): si ottiene passando al complementare.
(5) $\Rightarrow$ (7): sappiamo che vale la (6) e quindi $\limsup_{n\to\infty} \mathbb{P}_n(\overline{B}) \leq \mathbb{P}(\overline{B}) = \mathbb{P}(B)$. Ma è anche $\liminf_{n\to\infty} \mathbb{P}_n(\mathring{B}) \geq \mathbb{P}(\mathring{B}) = \mathbb{P}(b)$ e quindi $\mathbb{P}(B) \leq \liminf_{n\to\infty} \mathbb{P}_n(B) \leq \limsup n \to \infty \mathbb{P}_n(B) \leq \mathbb{P}(B)$.
(7) $\Rightarrow$ (1): se x è di continuità per F è $\mathbb{P}(\{x\}) = 0$ e $\{x\} = \partial(-\infty, x]$.
(1) $\Rightarrow$ (2): i punti di discontinuità di F sono al più numerabili per cui, fissato $\varepsilon > 0$, esistono $a < b$ punti di continuità tali che $\mathbb{P}((-\infty, a]) < \varepsilon$ e $\mathbb{P}((b, \infty)) < \varepsilon$ e questo sarà vero anche per le $\mathbb{P}_n$ per n sufficientemente grande. Per l'uniforme continuità di f in $[a, b]$ esistono $a = a_0 < a_1 < \dots < a_k = b$ tali che per ogni $j \leq k-2$ e per ogni $x, y \in [a_j, a_{j+2}]$ si abbia $|f(x) - f(y)| < \varepsilon$, quindi possiamo prendere $c_0 = a$, $c_{k+1} = b$ e $c_j \in (a_{j-1}, a_j)$ di continuità per F (per $1 \leq j \leq k$) e una funzione g costante sugli intervalli $(c_j, c_{j+1}]$ (e nulla al di fuori di $(a, b]$) in modo che si abbia $|f(x) - g(x)| < \varepsilon$ su tutto $(a, b]$. Ora, siccome per la (1) vale $\mathbb{P}_n((c_j, c_{j+1}]) \to \mathbb{P}((c_j, c_{j+1}])$ per $0 \leq j \leq k$, se poniamo $M = \|f\|_\infty$ abbiamo $\left|\int f\, d\mathbb{P} - \int g\, d\mathbb{P}\right| \leq 2\varepsilon M + \int_{(a,b]} |f - g|\, d\mathbb{P} \leq 2\varepsilon M + \varepsilon$ e lo stesso per le $\mathbb{P}_n$ con n abbastanza grande.
Infine $\left|\int g\, d\mathbb{P} - \int g\, d\mathbb{P}_n\right| \leq \sum_j |u_j| \cdot |\mathbb{P}((c_j, c_{j+1}]) - \mathbb{P}_n((c_j, c_{j+1}])| \to 0$ dove u_j è il valore di g su $(c_j, c_{j+1}]$.
(2) $\Rightarrow$ (3): ovvio.
(2) $\Rightarrow$ (4): per ogni $m > 0$ intero abbiamo $\liminf_{n\to\infty} \int f\, d\mathbb{P}_n \geq \liminf_{n\to\infty} \int \min(f, m)\, d\mathbb{P}_n = \int \min(f, m)\, d\mathbb{P}$ e per il teorema di convergenza monotona $\int \min(f, m)\, d\mathbb{P} \to \int f\, d\mathbb{P}$ per $m \to \infty$.
(3) $\Rightarrow$ (1): se x è di continuità per F e $\varepsilon > 0$ prendo f lineare a tratti che vale 1 su $(-\infty, x - \varepsilon]$ e 0 su $[x, \infty)$. Ho $\liminf_{n\to\infty} \mathbb{P}_n((-\infty, x]) \geq \liminf_{n\to\infty} \int f\, d\mathbb{P}_n = \int f\, d\mathbb{P} \geq \mathbb{P}((-\infty, x - \varepsilon])$. Analogamente prendendo f che vale 1 su $(-\infty, x]$ e 0 su $[x + \varepsilon, \infty)$ si ottiene $\limsup_{n\to\infty} \mathbb{P}_n((-\infty, x]) \leq \mathbb{P}((-\infty, x + \varepsilon])$. Dato che ε era arbitrario otteniamo la tesi.
(4) $\Rightarrow$ (1): esattamente come prima osservando che se $0 \leq f \leq 1$ vale $1 -$

$\limsup_{n\to\infty}\int f\,d\mathbb{P}_n = \liminf_{n\to\infty}\int(1-f)\,d\mathbb{P}_n \geq \int(1-f)\,d\mathbb{P} = 1-\int f\,d\mathbb{P}$, per cui anche $\limsup_{n\to\infty}\int f\,d\mathbb{P}_n \leq \int f\,d\mathbb{P}$, cioè $\lim_{n\to\infty}\int f\,d\mathbb{P}_n = \int f\,d\mathbb{P}$.
Dunque (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4).
(2) $\Rightarrow$ (5): sia $f_k(x) = \min\left(kd(x, G^c), 1\right)$ ($d(\cdot,\cdot)$ è la distanza) per k intero positivo. $0 \leq f_k \leq 1$ e la successione delle f_k tende puntualmente all'indicatrice di G. Ora $\int f_k\,d\mathbb{P} = \liminf_{n\to\infty}\int f_k\,d\mathbb{P}_n \leq \liminf_{n\to\infty}\mathbb{P}_n(G)$ e per $k\to\infty$ $\int f_k\,d\mathbb{P} \to \mathbb{P}(G)$ per il teorema di convergenza monotona. □

Definizione 2. *Una successione di misure* $(\mathbb{P}_n)$ *su* $\mathbb{R}$ converge in distribuzione *a* $\mathbb{P}$ *se vale una delle precedenti.*

Teorema 3 (di Prokhorov)**.** *Se* $\mathbb{P}_1, \mathbb{P}_2, \ldots$ *è una successione di probabilità su* $\mathbb{R}$ *tali che* $\lim_{M\to\infty}\sup_n \mathbb{P}_n\left([-M,M]^c\right) = 0$ *allora esiste* $\mathbb{P}$ *e una sottosuccessione convergente a* $\mathbb{P}$ *in distribuzione.*

Dimostrazione. Chiamando F_n le funzioni di ripartizione delle $\mathbb{P}_n$, troviamo una sottosuccessione (F_{n_k}) di (F_n) convergente su $\mathbb{Q}$ (enumeriamo i razionali e estraiamo una sottosuccessione convergente sul primo, da questa ne estraiamo un'altra convergente sul secondo e così via; la sottosuccessione cercata si ottiene con un procedimento diagonale), diciamo a H (definita solo su $\mathbb{Q}$). Poniamo $F(x) = inf_{y>x}H(y)$ $(y \in \mathbb{Q})$ e osserviamo che F è continua a destra ed è debolmente crescente. Inoltre, fissato $\varepsilon > 0$, esiste per ipotesi un M tale che $\sup_n \mathbb{P}_n\left([-M,M]^c\right) < \varepsilon$ e quindi per ogni $y < -M$ è $H(y) \leq \varepsilon$. Dunque $F(-M-1) \leq \varepsilon$ e analogamente $F(M) \geq 1-\varepsilon$. Quindi F è la funzione di ripartizione di qualche probabilità $\mathbb{P}$. Infine se x è un punto di continuità per F abbiamo $\lim_{k\to\infty}F_{n_k}(x) = F(x)$: essendo H debolmente crescente, per ogni ε esistono $\alpha < x$ e $\beta > x$ tali che $H(\alpha) > F(x)-\varepsilon$ e $H(\beta) < F(x)+\varepsilon$. Dunque definitivamente $F(x)-\varepsilon < F_{n_k}(\alpha) \leq F_{n_k}(x) \leq F_{n_k}(\beta) < F(x)+\varepsilon$. □

Definizione 4. *La* funzione caratteristica *di* $\mathbb{P}$ *(probabilità su* $\mathbb{R}$*) è* $\phi(t) = \int e^{itx}\,d\mathbb{P}(x)$.

Teorema 5. *La funzione caratteristica* ϕ *determina univocamente* $\mathbb{P}$.

Dimostrazione. Chiaramente basta ottenere che per ogni $a < b$ è
$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)\,dt = \frac{1}{2}\mathbb{P}(\{a,b\}) + \mathbb{P}\left((a,b)\right)$ (perché a quel punto $\mathbb{P}$ è univocamente determinata sugli intervalli aperti).
L'integrale si può fare perché $\phi(\cdot)$ è continua (per il teorema di convergenza dominata). Il teorema di Fubini dà $\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\left(\int e^{itx}\,d\mathbb{P}(x)\right)dt$
$= \int\left(\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}\,dt\right)d\mathbb{P}(x)$.
Ma ora $\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}\,dt = \int_{-T}^{T}\frac{\sin(t(x-a))-\sin(t(x-b))}{t}\,dt = 2\int_0^T\frac{\sin(t(x-a))}{t}\,dt$
$-2\int_0^T\frac{\sin(t(x-b))}{t}\,dt$ e (per ogni x) $2\int_0^T\frac{\sin(tx)}{t}\,dt = 2\,\mathrm{sgn}(x)\int_0^{T|x|}\frac{\sin u}{u}\,du \to \pi\,\mathrm{sgn}(x)$ (per $T\to\infty$). Essendo $\int_0^{T|x|}\frac{\sin u}{u}\,du$ limitato al variare di x e T possiamo applicare il teorema di convergenza dominata ottenendo $\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)\,dt =$
$\frac{1}{\pi}\int\left(\int_0^T\frac{\sin(t(x-a))}{t}\,dt - \int_0^T\frac{\sin(t(x-b))}{t}\,dt\right)d\mathbb{P}(x) \to$
$\to \frac{1}{2}\int\left(\mathrm{sgn}(x-a)-\mathrm{sgn}(x-b)\right)d\mathbb{P}(x)$, che è la tesi. □

Teorema 6 (di continuità di Lévy)**.** *Date $\mathbb{P}_1, \mathbb{P}_2, \dots$ probabilità su $\mathbb{R}$ e chiamate ϕ_n le loro funzioni caratteristiche, esiste $\mathbb{P}$ tale che $\mathbb{P}_n \to \mathbb{P}$ in distribuzione se e solo se $\phi_n \to \phi$ puntualmente (per qualche ϕ) con ϕ continua in 0. In tal caso ϕ è la funzione caratteristica di $\mathbb{P}$.*

Dimostrazione. Se $\mathbb{P}_n \to \mathbb{P}$ allora per il Teorema **1** $\phi_n \to \phi$ (segue dal punto (2)).
Viceversa se $\phi_n \to \phi$ basta vedere che $\lim_{M\to\infty} \sup_n \mathbb{P}_n\left([-M, M]^c\right) = 0$: fatto questo, per il Teorema **3** esistono $(\mathbb{P}_{n_k})$ sottosuccessione e $\mathbb{P}$ tali che $\mathbb{P}_{n_k} \to \mathbb{P}$ in distribuzione, da cui ϕ è la funzione caratteristica di $\mathbb{P}$. Se ora per assurdo non fosse $\mathbb{P}_n \to \mathbb{P}$ esisterebbero un'altra sottosuccessione $(\mathbb{P}_{m_k})$ e un x (di continuità per la funzione di ripartizione di $\mathbb{P}$) con $|\mathbb{P}_{m_k}((-\infty, x]) - \mathbb{P}((-\infty, x])| \geq \varepsilon$ per tutti i k e a meno di passare a un'ulteriore sottosuccessione potremmo supporre $\mathbb{P}_{m_k} \to \mathbb{P}'$ e ϕ sarebbe la funzione caratteristica di $\mathbb{P}'$. Per il Teorema **5** avremmo $\mathbb{P} = \mathbb{P}'$, cioè $\mathbb{P}_{m_k} \to \mathbb{P}$, assurdo.
Sia perciò $\delta > 0$: $\chi_{\{|\delta x|>2\}} \leq 2\left(1 - \frac{\sin(\delta x)}{\delta x}\right) = \frac{1}{\delta}\int_{-\delta}^{\delta}(1 - cos(tx))\, dt$. Dunque $\int \chi_{\{|\delta x|>2\}}\, d\mathbb{P}_n \leq \frac{1}{\delta}\int_{-\delta}^{\delta}\left(\int (1 - cos(tx))\, d\mathbb{P}_n(x)\right) dt$ (per il teorema di Fubini) e quest'ultimo è $\frac{1}{\delta}\int_{-\delta}^{\delta}(1 - \Re\phi_n(t))\, dt$. Per il teorema di convergenza dominata abbiamo $\frac{1}{\delta}\int_{-\delta}^{\delta}(1 - \Re\phi_n(t))\, dt \to \frac{1}{\delta}\int_{-\delta}^{\delta}(1 - \Re\phi(t))\, dt$. Essendo $\phi(0) = 1$ e ϕ continua in 0, per ogni ε possiamo trovare δ tale che il limite sia minore di ε e quindi definitivamente $\mathbb{P}_n(\{|\delta x| > 2\}) = \int \chi_{\{|\delta x|>2\}}\, d\mathbb{P}_n < \varepsilon$, cioè ponendo $M = \frac{2}{\delta}$ vale $\mathbb{P}_n([-M, M]^c) < \varepsilon$ definitivamente. □

Teorema 7 (del Limite Centrale)**.** *Dato uno spazio di probabilità Ω, se $X_1, X_2, \dots$ sono variabili aleatorie (reali) indipendenti, equidistribuite e con $\mathbf{E}[X_n] = 0$ e $Var(X_n) = 1$, posto $Y_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n)$ abbiamo $\mathbb{P}_{Y_n} \to N(0,1)$ in distribuzione.*

Dimostrazione. Se ϕ_n è la funzione caratteristica di $\mathbb{P}_{Y_n}$ e ψ quella di $\mathbb{P}_{X_1}$, $\phi_n(t) = \mathbf{E}\left[\exp\left(\frac{it}{\sqrt{n}}\sum_{i=1}^n X_i\right)\right] = \psi\left(\frac{t}{\sqrt{n}}\right)^n$ (per l'indipendenza delle X_i).
Ora, derivando sotto il segno di integrale, $\psi'(t) = \mathbf{E}\left[iX_1 \exp(itX_1)\right]$ e $\psi''(t) = \mathbf{E}\left[-X_1^2 \exp(itX_1)\right]$ (grazie all'ipotesi che $\mathbf{E}[X_1^2] < \infty$) e quindi, fissato t, $\psi\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 + \frac{it}{\sqrt{n}}\mathbf{E}[X_1] - \frac{t^2}{2n}\mathbf{E}[X_1^2] + o\left(\frac{1}{n}\right)\right)^n$ (sviluppo di Taylor con resto di Peano), cioè $\phi_n(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to \exp\left(-\frac{t^2}{2}\right) = \phi(t)$.
Dunque per il Teorema **6** esiste una probabilità $\mathbb{P}$ cui convergono le $\mathbb{P}_{Y_n}$ in distribuzione e la sua funzione caratteristica è ϕ. Ma la funzione caratteristica della distribuzione $N(0,1)$ è $\gamma(t) = \int_{-\infty}^{\infty} e^{itx}\frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}\, dx$ e $\gamma'(t) = \int_{-\infty}^{\infty} ixe^{itx}\frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}\, dx = -t\int_{-\infty}^{\infty} e^{itx}\frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}\, dx = -t\gamma(t)$ (integrando per parti), quindi per qualche C è $\gamma(t) = Ce^{-\frac{t^2}{2}}$ e per $t = 0$ otteniamo $C = 1$.
Perciò $\phi = \gamma$ e il Teorema **5** dà la tesi. □