LEGGE DEBOLE DEI GRANDI NUMERI – APPLICAZIONI 1

Note Title 13/03/2018

TEO Sia $(\Omega, \mathcal{E}, \mathbb{P})$ spazio probabilizzato e sia $\{X_n\}_{n\in\mathbb{N}}$ successione di v.a. su tale spazio t.c.

1) $\mathbb{E}[X_n] = E \in \mathbb{R} \quad \forall n \in \mathbb{N}$
2) $Cov(X_n, X_k) = 0 \quad \forall k, n \in \mathbb{N}$ e, $k \neq n$
3) $Var[X_n] \leq C^2 \quad \forall n \in \mathbb{N}$

Le X_n sono i.i.d. e integrabili

$$S_n := \sum_{k=1}^{n} X_k \qquad \bar{X}_n := \frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} X_k$$

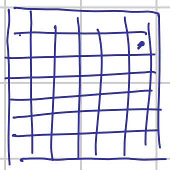
$\bar{X}_n$ converge in L^2 e in probabilità alla v.a costante E è quasi certamente

— o —

METODO MONTECARLO PER IL CALCOLO APPROSSIMATO DEGLI INTEGRALI

Sia $Q = [0,1]^N \subseteq \mathbb{R}^N$ $\quad f: Q \to \mathbb{R}$ continua

Voglio valutare $\int_Q f(x)\,dx$



$$\sum_{i=1}^{m} f(x_i)\, m(Q_i)$$

Sia $f: \mathbb{R}^N \to \mathbb{R}$ limitata $\quad |f(x)| \leq M \quad \forall x \in \mathbb{R}^N$
misurabile secondo Borel

$(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mathbb{P})$ spazio probabilizzato

$$\bar{f} = \int_{\mathbb{R}^N} f(x)\, \mathbb{P}(dx) \simeq \frac{1}{n} \sum_{i=1}^{n} f(x_i)$$

Modellare la scelta dei pti $x_1 — x_n$ come se fossero realizzazioni v.a su uno spazio $(\Omega, \mathcal{E}, \tilde{\mathbb{P}})$, $X_1 — X_n$

T.c. $\mathbb{P}_{X_i} = \mathbb{P}$ e $X_1 \dots X_n$ indipendenti

$$x_1 \;\_\_\_\_\; x_n \qquad \omega \in \Omega$$
$$\;\,\shortparallel \qquad\qquad \shortparallel$$
$$X_1(\omega) \qquad X_n(\omega)$$

Anche $f(X_1), f(X_2) \;\_\_\; f(X_n)$ sono indipendenti e identicamente distribuite

$$\mathbb{E}[\,|f(X_i)|\,] = \mathbb{E}[\,|f(X_1)|\,] = \int_{\mathbb{R}^N} |f(x)|\,\mathbb{P}(dx) \le$$

$$\le \int_{\mathbb{R}^N} M\,\mathbb{P}(dx) = M \cdot \mathbb{P}(\mathbb{R}^N) = M \cdot 1 = M$$

$$\mathbb{E}[f(X_i)] = \int_{\mathbb{R}^N} f(x)\,\mathbb{P}(dx) = \bar{f} \quad \Rightarrow \bar{f} \in \mathbb{R}$$

$$\mathrm{Var}[f(X_i)] = \mathrm{Var}[f(X_1)] = \mathbb{E}\left[(f(X_1) - \bar{f})^2\right] =$$

$$= \int_{\mathbb{R}^N} (f(x) - \bar{f})^2\,\mathbb{P}(dx) = \int_{\mathbb{R}^N} \left(f^2(x) - 2\bar{f} f(x) + \bar{f}^2\right)\mathbb{P}(dx)$$

$$\le \int_{\mathbb{R}^N} \left(f^2(x) + 2\underbrace{|\bar{f}|}_{\le M}|f(x)| + \bar{f}^2\right)\mathbb{P}(dx)$$

$$\le \int_{\mathbb{R}^N} \left(M^2 + 2M \cdot M + M^2\right)\mathbb{P}(dx) = 4M^2 \overbrace{\mathbb{P}(\mathbb{R}^N)}^{=1} = 4M^2$$

Applico legge debole dei grandi numeri a $\{f(X_i)\}_{i=1}^{\infty}$

$\frac{1}{n}\sum_{i=1}^{n} f(X_i)$ converge in probabilità, L^2 e quasi certamente a $\bar{f}$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \bar{f}\right| \ge \delta\right) \le \frac{4M^2}{n\delta^2} \qquad \forall \delta > 0$$

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[f(X_i)] \le \frac{1}{n^2}\, n\, 4M^2 = \frac{4M^2}{n}$$

ESEMPIO $\quad f: \mathbb{R}^N \to \mathbb{R} \qquad |f(x)| \le 10$

Voglio valutare $\overline{f}$ con errore $\delta = 10^{-2}$ e voglio essere sicuro al 90% di aver fatto questa valutazione

$$\mathbb{P}\left(\left| \frac{1}{n}\sum_{i=1}^{n} f(X_i) - \overline{f} \right| > 10^{-2} \right) \leq \frac{1}{10}$$

Basta che sia $\dfrac{4M^2}{n\delta^2} \leq \dfrac{1}{10}$ $\qquad \dfrac{4 \cdot 10^2}{n \cdot 10^{-4}} \leq 10^{-1}$

$n \cdot 10^{-4} \geq 4 \cdot 10^{3} \qquad n \geq 4 \cdot 10^{7}$

— o —

FUNZIONE DI RIPARTIZIONE EMPIRICA

$(\Omega, \mathcal{E}, \mathbb{P}) \qquad X : \Omega \to \mathbb{R} \qquad X(\Omega) = \{0, 1\}$

$\mathbb{P}(X = 1) = p$ $\qquad$ p parametro della distribuzione di Bernoulli

$\mathbb{P}(X = 0) = 1 - p$

La distribuzione si indica $B(p)$.

$\mathbb{E}[X] = 0 \cdot \mathbb{P}(X=0) + 1 \cdot \mathbb{P}(X=1) = p$

$\mathbb{E}[X^2] = 0 \cdot \mathbb{P}(X^2=0) + 1 \cdot \mathbb{P}(X^2=1) = p$

$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$

— o —

Sia X v.a. su $(\Omega, \mathcal{E}, \mathbb{P})$ e sia $\{X_n\}_{n\in\mathbb{N}}$ successione di v.a. i.i.d. t.c. $\mathbb{P}_{X_n} = \mathbb{P}_X$ $\forall n \in \mathbb{N}$.

$T_{n,t}(\omega) = \#$ di v.a. X_j di indice $j \leq n$ t.c. $X_j(\omega) \leq t$

cioè

$$T_{n,t}(\omega) = \sum_{j=1}^{n} \mathbb{1}_{(-\infty,t]}(X_j(\omega)) \qquad \omega \in \Omega$$

$\omega \in \Omega$, $n \in \mathbb{N}$, $t \in \mathbb{R}$

$Y_j(\omega) = \mathbb{1}_{(-\infty,t]}(X_j(\omega))$ $\qquad$ Le Y_j sono i.i.d. con $\mathbb{P}_{Y_j} = B(p)$

$$p = \mathbb{P}(Y_j = 1) = \mathbb{P}\left(\mathbb{1}_{(-\infty,t]}(X_j) = 1\right) = \mathbb{P}(X_j \leq t) = F_{X_j}(t) = F_X(t)$$

$$\{ \mathbb{1}_{(-\infty,t]}(X_j) = 1 \} = \{ X_j \in (-\infty, t] \} = \{ X_j \leq t \}$$

$$\mathbb{P}_{Y_j} = B(F_X(t)) \Rightarrow \mathbb{E}[Y_j] = F_X(t)$$

$$Var[Y_j] = F_X(t)(1 - F_X(t)) \leq \frac{1}{4}$$

$p(1-p) \quad p \in [0,1]$

$$p(1-p) = p - p^2 = \frac{1}{4} - \frac{1}{4} + p - p^2 = \frac{1}{4} - \left(\frac{1}{4} - p + p^2\right) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leq \frac{1}{4}$$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - F_X(t)\right| > \delta\right) \leq \frac{1/4}{n\delta^2} = \frac{1}{4n\delta^2} \qquad \forall \delta > 0 \quad \forall n \in \mathbb{N} \quad \forall t \in \mathbb{R}$$

$$\mathbb{P}\left(\left|\frac{1}{n} T_{n,t} - F_X(t)\right| > \delta\right) \leq \frac{1}{4n\delta^2}$$

La v.a. $\frac{1}{n} T_{n,t}(\omega)$ si indica $F_n(\omega, t)$ e si dice FUNZIONE DI RIPARTIZIONE EMPIRICA DI X

TEOREMA DI GLIVENKO - CANTELLI (NO DIM)

$$\lim_{n \to \infty} \sup_{t \in \mathbb{R}} \left| F_n(\omega, t) - F_X(t) \right| = 0 \qquad \mathbb{P}\text{-q.c. } \omega \in \Omega$$

— o —

ENTROPIA

$E = \{a_1, a_2, \ldots, a_q\}$ Posso sempre supporre $E = \{1, \ldots, q\}$

$(E, \mathcal{B}(E), \mu)$ $\quad \mu(\{i\}) = p_i \qquad p_1, \ldots, p_q \geq 0$

$$\sum_{i=1}^{q} p_i = 1$$

$x_1, \ldots, x_n \quad x_i \in E$

$X_1, \ldots, X_n \quad (\Omega, \mathcal{E}, \mathbb{P}) \quad X_1, \ldots, X_n$ i.i.d

$$\mathbb{P}_{X_i} = \mu$$

$(X_1(\omega), \ldots, X_n(\omega)) = i_1 i_2 \ldots i_n \qquad i_j \in E$

$$\mathbb{P}\left((X_1, \ldots, X_n) = (i_1, i_2, \ldots, i_n)\right) = \prod_{j=1}^{n} \mathbb{P}(X_j = i_j) = \prod_{j=1}^{n} \underbrace{\mu(\{i_j\})}_{p_{i_j}}$$

$$Y_n(\omega) = -\ln\left(P_{X_n(\omega)}\right) \qquad n \in \mathbb{N}$$

Le Y_n sono ancora v.a. su $(\Omega, \mathcal{E}, \mathbb{P})$

Sono i.i.d perché lo sono le X_n.

$$\sum_{k=1}^{n} Y_n(\omega) = \sum_{k=1}^{n} -\ln P_{X_k(\omega)} = -\sum_{k=1}^{n} \ln P_{X_k(\omega)} =$$

$$= -\ln \prod_{k=1}^{n} P_{X_k(\omega)}$$

$$\mathbb{E}[Y_n] = \mathbb{E}\left[-\ln P_{X_n}\right] = \mathbb{E}\left[-\ln P_{X_1}\right] = \int_\Omega -\ln P_{X_n(\omega)} \mathbb{P}(d\omega)$$

$$= \int_{\mathbb{R}} -\ln P_x \, \mu(dx) = \sum_{i=1}^{q} -\ln p_i \, \mu(\{i\}) =$$

$$= \sum_{i=1}^{q} -p_i \ln p_i = -\sum_{i=1}^{q} p_i \ln p_i$$