Media empirica / 07-04

$X \quad \{x_1, \dots, x_n\} \qquad X, Y \quad \{x_1, \dots, x_n\} \; \{y_1, \dots, y_m\}$

$p(x_1) \dots p(x_n) \qquad p(x_1, x_n) \; \dots \; p(x_n, y_m)$

$g(X, Y) \qquad g : \{x_1, \dots x_n\} \times \{y_1, \dots, y_m\} \to \mathbb{R}$

$$\mathbb{E}[g(X,Y)] = \sum_{i=1}^{n} \sum_{j=1}^{m} g(x_i, y_j)\, p(x_i, y_j)$$

(di per sé $\nexists$ $\mathbb{E}$ di una coppia)

$g(X,Y) = X + Y \qquad E[X+Y] = E[X] + E[Y]$

$g(X,Y) = X \cdot Y \qquad E[X \cdot Y] = E[X] \cdot E[Y]$

se X e Y indipendenti

Somme di var aleatorie

$$\sum_{i=1}^{N} X_i$$

X_i sono identiche e indipendenti
(es. più lanci di una stessa moneta o di più monete uguali)

es.

Se X è binomiale allora $X = \sum_{i=1}^{N} X_i$ dove X_i è Bernoulli

$E[X_i] = p$

$$E[X] = E\left[\sum_{i=1}^{N} X_i\right] = Np$$

$$E[X + X + \dots + X]$$

- $$\mu_N = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad E[X_i] = \mu \quad (\text{non più Bernoulli})$$

media empirica (eventi equiprobabili)

Valore atteso della media empirica

$$E[\mu_N] = \frac{N\mu}{N} = \mu$$

$$Var(\mu_N) = E\left[\left(\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\right)^2\right]$$

$$\mu = \frac{1}{N}\sum_{i=1}^{N}\mu$$

l'idea è più tentativi faccio meno varianza ho (errore relativo più piccolo)

$$= E\left[\left(\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)\right)^2\right]$$

$$= \frac{1}{N^2} E\left[\sum_{i=1}^{N}(X_i - \mu)^2\right]$$

E si spezza con le somme

se $N=2$: $E\left[\left(\underbrace{(X_1-\mu)}_{a} + \underbrace{(X_2-\mu)}_{b}\right)^2\right] =$

$$= E[(X_1-\mu)^2] + E[(X_2-\mu)^2] + 2E[(X_1-\mu)(X_2-\mu)]$$

$X_1 = X_2$ ma shiftate di μ

sono indip quindi $E[X_1 - \mu] \cdot E[X_2 - \mu]$

$= 0$

$*\ \sigma^2 = Var(X_i)$

perché $E[X+\mu] = E[X] + E[\mu] = E[X] - \mu = \mu - \mu = 0$

ma allora il tutto e' uguale a $2\sigma^2$ (per $N=2$)

Quindi $\frac{1}{N^2} E\left[\sum_{i=1}^{N}(x_i-\mu)^2\right] = \frac{\cancel{N}\sigma^2}{N^{\cancel{2}}} = \sigma^2/N$ (diminuisce)

La varianza migliora all'aumentare dei tentativi

↓

la varianza della media empirica e' la varianza diviso N

La varianza di una somma di var aleatorie INDIP e' la somma delle loro varianze | *

$$E[(X-\mu)^2] + E[(Y-\rho)^2]$$
$$Var(X) + Var(Y)$$

Variabili non indipendenti

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

$$E[XY] + E[X]E[Y] +$$

$-E[X E[Y]]$ se X e' un numero $\rightarrow -E[Y]E[X]$

$-E[Y E[X]] \rightarrow -E[X]E[Y]$

ma allora $Cov(X,Y) = E[XY] - E[X]E[Y]$

Se X e Y indip $E[XY] = E[X]E[Y]$ e $Cov(X,Y) = 0$

mentre in generale Cov puo' essere positiva o negativa
ma non e' normalizzata

↓

Correlazione

una specie di normalizzazione

stabilisco dei termini per la Cov e capire quanto sono correlate X e Y

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

↳ deviazioni standard

$\sigma_X^2 = Var(X)$

$\sigma_Y^2 = Var(Y)$

se occorre X allora occorre Y

se X e Y indip $Cov = 0$ e $\rho = 0$

In generale, essendo normalizzata, $-1 \leq \rho(X,Y) \leq 1$
(non vale per Cov)

Dimostrazione

- $Cov(aX, Y) = E[(aX - E[X])(Y - E[Y])] = a\,Cov(X,Y)$

- $Var(aX + Y) = E[(aX - E[X])^2] + E[(Y - E[Y])^2]$

$$+ 2_a E[(aX - E[aX])(Y - E[Y])]$$

$\neq 0$ non sono indipendenti $= Cov$

$$= a^2 Var(X) + Var(Y) + 2\,Cov(X,Y)$$

formula generale, se X e Y indipendenti $Cov = 0$
e viene la sola somma di Var X e Var Y

Se $Var(aX + Y) = a^2 Var(X) + Var(Y) + 2a\, Cov(X,Y)$

stesso se aggiungo un b

$$\cdot\ Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = \frac{Var(X)}{\sigma_X^2} + \frac{Var(Y)}{\sigma_Y^2} + 2\frac{1}{\sigma_X}\frac{1}{\sigma_Y} Cov(X,Y)$$

$0 \leq$

$$= 1 + 1 + 2\rho(X,Y)$$

$$= 2(1 + \rho(X,Y))$$

ma allora $\rho(X,Y) \geq -1 \rightarrow$ (valore minimo per fare zero)

se $Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 2(1 - \rho(X,Y))$ e $\rho(X,Y) \leq 1$