1 Momenti, varianza, covarianza di una v.a.

Definizione
Sia X una v.a; diciamo che X ha momento di ordine k finito se la v.a. X^k è tale che $E(|X^k|) < +\infty$.
In tal caso $E(X^k)$ si chiama *momento di ordine* k di X. Analogamente, se $E(|X - E(X)|^k) < +\infty$, la quantità $E[(X - E(X))^k]$ si chiama *momento centrato di ordine* k.
Se X è una v.a. che assume valori $\{x_1, x_2, \ldots, x_n\}$, con probabilità $p_i = P(X = x_i)$, $i = 1, 2, \ldots, n$, si ha:

$$E(X^k) = \sum_{i=1}^{n} x_i^k p_i$$

e

$$E[(X - E(X))^k] = \sum_{i=1}^{n} (x_i - E(X))^k p_i = \sum_{i=1}^{n} (x_i - m)^k p_i,$$

avendo posto $m = E(X)$.

Varianza di X
E' il momento centrato di ordine 2 di X :

$$Var(X) = E[(X - E(X))^2] = E[X^2 - 2E(X)X + (E(X))^2) = E(X^2) - E^2(X) \tag{1.1}$$

(attenzione alle parentesi: $E^2(X)$ è un abbreviazione per $(E(X))^2$, da non confondersi con $E(X^2)$).
La varianza di X, ovvero $Var(X)$, è una misura della dispersione di X attorno alla sua media. Infatti, se X prende valori lontani dalla sua media, allora la v.a. $(X - E(X))^2$ assumerà valori grandi e, di conseguenza $Var(X)$ sarà grande. Viceversa, se X prende solo il valore $m = E(X)$, allora $(X - E(X))^2 = 0$ con probabilità 1 e, di conseguenza $Var(X) = 0$.
Osserviamo subito che, dalla definizione (1.1) segue che $Var(X) \geq 0$, essendo la media di $[X - E(X)]^2$, che è una v.a. non negativa, e che, da quanto detto, risulta $Var(X) = 0$ se e solo se X assume, con probabilità 1, un unico valore $m = E(X)$. Inoltre, sempre da (1.1) segue che deve sempre essere, per ogni v.a. X :

$$E(X^2) - E^2(X) \geq 0, \text{ ovvero } E(X^2) \geq E^2(X). \tag{1.2}$$

Esempio Siano X e Y v.a. tali che:
$X \in \{-1, 1\}$ con probabilità $P(X = -1) = P(X = 1) = \frac{1}{2}$;
$Y \in \{-100, 100\}$ con probabilità $P(Y = -100) = P(Y = 100) = \frac{1}{2}$.
Si ha $E(X) = (-1) \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0$, come pure $E(Y) = (-100) \cdot \frac{1}{2} + 100 \cdot \frac{1}{2} = 0$.
Per quanto riguarda le varianze, si ha:

$$Var(X) = E[(X - E(X))^2] = E(X^2) - E^2(X) = E(X^2) = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1;$$

$$Var(Y) = E[(Y - E(Y))^2] = E(Y^2) - E^2(Y) = E(Y^2) = (-100)^2 \cdot \frac{1}{2} + (100)^2 \cdot \frac{1}{2} = 10000.$$

Come si vede, la varianza di Y è molto più grande di quella di X, pur essendo zero entrambe le loro medie; questo succede perché Y assume valori più lontani dalla sua media, di quanto non faccia X.

Disuguaglianza di Chebychev
Se X è una v.a. discreta con media e varianza finite, $\forall \varepsilon > 0$ si ha:

$$P(|X - E(X)| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}. \tag{1.3}$$

Dim. Posto $m = E(X)$, risulta:

$$P(|X - m| > \varepsilon) = \sum_{x:|x-m|>\varepsilon} P(X = x);$$

siccome, per i valori di x considerati nella somma risulta $\frac{(x-m)^2}{\varepsilon^2} > 1$ (visto che $|x - m| > \varepsilon$), la somma risulta

$$\leq \sum_{x:|x-m|>\varepsilon} \frac{(x-m)^2}{\varepsilon^2} \cdot P(X = x)$$

$$\leq \frac{1}{\varepsilon^2} \sum_{\text{tutte le } x \text{ possibili}} (x-m)^2 \cdot P(X = x) = \frac{1}{\varepsilon^2} Var(X).$$

La disuguaglianza di Chebychev (1.3) mette in relazione la dispersione della v.a X con la $Var(X)$. Più la varianza è piccola, minore è la probabilità che X assume valori lontani dalla sua media. Naturalmente, la disuguaglianza di Chebychev è significativa solo se $\frac{Var(X)}{\varepsilon^2} < 1$, poiché, altrimenti, è una disuguaglianza banale, in quanto ogni probabilità è ovviamente ≤ 1.

Si chiama *deviazione standard* o *scarto quadratico medio* di X la quantità $\sigma = \sqrt{Var(X)}$ (si ricordi che $Var(X) \geq 0$).

Proprietà della varianza
Se $a \in \mathbb{R}$, risulta:
(i) $Var(aX) = a^2 Var(X)$;
(ii) $Var(a + X) = Var(X)$.
Infatti:
(i) $Var(aX) = E[(aX)^2] - E^2[aX] = E[a^2X^2] - [E(aX)]^2 = a^2E(X^2) - a^2E^2(X)$
$= a^2 Var(X)$;
(ii) $Var(a+X) = E[(a+X-E(a+X)]^2 = E[a+X-a-E(X)]^2 = E[(X-E(X))^2] = Var(X)$.

Inoltre, se X e Y sono v.a. con varianze finite, si ha:
(iii) $Var(X + Y) = Var(X) + Var(Y) + 2cov(X, Y)$, dove abbiamo posto

$$cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

Intanto, mostriamo che l'ultima uguaglianza segue svolgendo i calcoli, infatti:

$$E[(X - E(X))(Y - E(Y))] = E[XY - XE(Y) - E(X)Y + E(X)E(Y)]$$

$$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y).$$

Ora, dimostriamo (iii). Si ha:

$$Var(X+Y) = E[(X+Y)^2] - E^2(X+Y) = E[X^2 + Y^2 + 2XY] - [E(X+Y)]^2$$
$$= E(X^2) + E(Y^2) + 2E(XY) - [E(X) + E(Y)]^2$$
$$= E(X^2) + E(Y^2) + 2E(XY) - E^2(X) - E^2(Y) - 2E(X)E(Y)$$
$$= [E(X^2) - E^2(X)] + [E(Y^2) - E^2(Y)] + 2[E(XY) - E(X)E(Y)]$$
$$= Var(X) + Var(Y) = 2cov(X,Y).$$

Se X e Y sono v.a. stocasticamente indipendenti, risulta $E(XY) = E(X)E(Y)$, ovvero $cov(X,Y) = 0$ e la (iii) diventa

(iii') $Var(X+Y) = Var(X) + Var(Y)$ (X e Y indipendenti).
Se $cov(X,Y) = 0$, le v.a. si dicono *scorrelate*, quindi X e Y indipendenti $\Rightarrow$ X e Y scorrelate, ma non è vero il viceversa, come approfondiremo in seguito.

Esempio 1 (Varianza di $X \sim Uni(\{x_1, x_2, \dots, x_n\})$)
Si ha $m = E(X) = (x_1 + x_2 + \cdots + x_n)/n$ e $E(X^2) = \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2)$, per cui

$$Var(X) = E(X^2) - E^2(X) = \frac{1}{n^2}[(x_1^2 + x_2^2 + \cdots + x_n^2) - (x_1 + x_2 + \cdots + x_n)^2].$$

Esempio 2 (Varianza di una v.a. di Bernoulli)
Se $X \sim B(1,p)$, $p \in (0,1)$, si ha

$$Var(X) = E(X^2) - E^2(X) = p - p^2 = p(1-p).$$

Esempio 3 (Varianza di una v.a. Binomiale)
Se $X \sim B(n,p)$, $p \in (0,1)$, si può scrivere $X = X_1 + X_2 + \cdots + X_n$, dove X_i è indipendente da X_j per $i \neq j$ e $X_i \sim B(1,p)$. Allora,

$$Var(X) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = np(1-p).$$

Esempio 4 (Varianza di una v.a. di Poisson)
Se $X \sim Poisson(\lambda)$, $\lambda > 0$, per $k = 0, 1, \dots$ si ha $P(X = k) = e^{-\lambda}\lambda^k/k!$ e $E(X) = \lambda$; allora

$$E(X^2) = e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} =$$

(ponendo $i = k - 1$)

$$e^{-\lambda}\lambda\sum_{i=0}^{\infty}(i+1)\frac{\lambda^i}{i!} = e^{-\lambda}\left[\lambda\sum_{i=0}^{\infty} i\frac{\lambda^i}{i!} + \lambda\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}\right] = \lambda E(X) + \lambda e^{-\lambda}e^{\lambda} = \lambda(\lambda+1)$$

(essendo $E(X) = \lambda$); quindi

$$E(X^2) = \lambda(\lambda+1) = \lambda^2 + \lambda.$$

Allora

$$Var(X) = E(X^2) - E^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

La v.a. di Poisson ha la proprietà di avere media e varianza uguali (ciò non avviene per tutte le v.a.).

Esempio 5 (Varianza di una v.a. Geometrica X e di una Geometrica modificata T)
Osserviamo che, potendosi porre $X = T - 1$, le v.a. X e T hanno la stessa varianza. Calcoliamo, ad esempio $Var(T)$. Se T è una v.a. con distribuzione Geometrica modificata di parametro $p \in (0,1)$, risulta $P(T = k) = p(1-p)^{k-1},\ k = 1, \dots$ e $E(T) = \frac{1}{p}$, come abbiamo già visto. Allora:

$$S := E(T^2) = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} =$$

(posto $k - 1 = j$)

$$\sum_{j=0}^{\infty}(j+1)^2 p(1-p)^j = \sum_{j=0}^{\infty} j^2 p(1-p)^j + 2\sum_{j=0}^{\infty} jp(1-p)^j + \sum_{j=0}^{\infty} p(1-p)^j$$

$$= (1-p)\sum_{j=1}^{\infty} j^2 p(1-p)^{j-1} + 2(1-p)\sum_{j=1}^{\infty} jp(1-p)^{j-1} + 1$$

(essendo $\sum_{j=0}^{\infty} p(1-p)^j = 1$, perché è la somma di tutte le probabilità di assunzione dei valori di una v.a. Geometrica). La prima serie nella formula di sopra è uguale a S, mentre la seconda è uguale a $E(T)$, dunque:

$$S = E(T^2) = (1-p)S + 2(1-p)E(T) + 1 = (1-p)S + 2(1-p)/p + 1,$$

da cui, risolvendo l'equazione rispetto a S, si ottiene

$$S = E(T^2) = \frac{2-p}{p^2};$$

infine:

$$Var(T) = E(T^2) - E^2(T) = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

Esempio 6 (Varianza di una v.a. Ipergeometrica)
Se $X \sim Hyper(r, b, n)$, si ha:

$$Var(X) = \frac{rbn}{(b+r)^2} \cdot \frac{b+r-n}{b+r-1}.$$

Infatti, abbiamo già visto che si può scrivere $X = X_1 + X_2 + \cdots + X_n$, dove $X_i \sim B(1, r/(b+r))$, ma non sono indipendenti. Comunque, essendo le X_i di Bernoulli, si ha:

$$E(X_i) = \frac{r}{b+r} \text{ e } Var(X_i) = \frac{r}{b+r}\left(1 - \frac{r}{b+r}\right).$$

Siccome le v.a. X_i non sono indipendenti, $Var(X) \neq \sum_{i=1}^n Var(X_i)$. Osserviamo però che:

$$1)\ E(X_1X_2) = P(X_1 = 1, X_2 = 1) = \frac{r}{b+r} \cdot \frac{r-1}{b+r-1};$$

2) $E(X_1X_3) = P(X_1 = 1, X_3 = 1) = P(X_1 = 1, X_2 = 0, X_3 = 1) + P(X_1 = 1, X_2 = 1, X_3 = 1)$

$$= \left(\frac{r}{b+r} \cdot \frac{b}{b+r-1} \cdot \frac{r-1}{b+r-2}\right) + \left(\frac{r}{b+r} \cdot \frac{r-1}{b+r-1} \cdot \frac{r-2}{b+r-2}\right) = \frac{r(r-1)}{(b+r)(b+r-1)}.$$

In maniera analoga, si ottiene, per $i < j$:

$$E(X_iX_j) = \frac{r(r-1)}{(b+r)(b+r-1)}.$$

Allora

$$cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) = \frac{r(r-1)}{(b+r)(b+r-1)} - \frac{r^2}{(b+r)^2} = -\frac{rb}{(b+r)^2(b+r-1)}.$$

Dunque:

$$Var(X) = Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + 2\sum_{i,j} cov(X_i, X_j),$$

dove l'ultima somma è estesa a tutti gli indici $i < j$ con $i, j = 1, 2 \ldots, n$; essi sono esattamente $C_{n,2} = n(n-1)/2$, per cui, riprendendo il calcolo, si ottiene infine

$$Var(X) = n \cdot \frac{r}{b+r}\left(1 - \frac{r}{b+r}\right) + n(n-1)\left(\frac{-rb}{(b+r)^2(b+r-1)}\right)$$

$$= \frac{nr}{(b+r)^2}\left[b + r - r + \frac{(n-1)(-b)}{b+r-1}\right] = \frac{rbn}{(b+r)^2} \cdot \frac{b+r-n}{b+r-1}.$$

Se $b + r \to \infty$, si riottiene la varianza di una Binomiale (schema di prove indipendenti), poiché $Var(X) \to \frac{rbn}{(b+r)^2}$.

Esempio 8 (Varianza di una v.a. di Pascal)
Se T_k è l'istante del $k-$mo successo in una successione di prove di Bernoulli, indipendenti, in ognuna delle quali la probabilità del successo è costante, ed uguale a $p \in (0, 1)$, abbiamo già calcolato $E(T_k)$, che è k/p. Con calcoli un po' più laboriosi, si può mostrare che

$$Var(T_k) = \frac{k(1-p)}{p^2}.$$

Notare che, per $k = 1$, ritorna la varianza di T_1, istante di primo successo.

2 Proprietà della covarianza

Variabili aleatorie scorrelate e v.a. indipendenti
Veniamo ora a studiare che relazione c'è tra v.a. indipendenti e v.a. scorrelate. Ricordiamo che si chiama covarianza delle v.a. X e Y la quantità:

$$cov(X, Y) = E(XY) - E(X)E(Y).$$

Se $cov(X, Y) = 0$, X e Y si dicono scorrelate; abbiamo già visto che, se X e Y sono stocasticamente indipendenti, allora esse sono scorrelate.
Il viceversa però non è vero, come mostra il seguente esempio.

Esempio di due v.a. scorrelate, ma non indipendenti
Sia $X \sim Uni(\{-1, 0, 1\})$ e poniamo $Y = X^2$; la v.a. Y assume valori 0 e 1 con $P(Y = 0) = 1/3$ e $P(Y = 1) = 2/3$. Risulta anche che $E(X) = 0$. Inoltre

$$E(XY) = E(X^3) = (-1)^3 \cdot \frac{1}{3} + 0^3 \cdot \frac{1}{3} + 1^3 \cdot \frac{1}{3} = 0,$$

per cui $cov(X, Y) = E(XY) - E(X)E(Y) = 0$. Ora osserviamo che, ad esempio, $P(X = 0, Y = 1) = 0$, poiché se $X = 0$, Y non può essere uguale a 1. Da ciò segue che X e Y non sono indipendenti, poiché:

$$0 = P(X = 0, Y = 1) \neq P(X = 0)P(Y = 1) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}\ .$$

$cov(X, Y)$ viene spesso usata come una misura di quanto X e Y si influenzino tra loro; se $cov(X, Y)$ è diversa da zero, ma è piccola, allora X e Y sono *poco correlate*, anche se non si può dire che sono indipendenti. Esiste una classe particolare di v. a. continue, che vedremo in seguito, le cosiddette v.a. Normali o Gaussiane, per cui vale:
se X e Y sono v.a. Gaussiane e scorrelate, allora X e Y sono indipendenti.
In generale, però, se X e Y sono scorrelate, non è detto che siano indipendenti.

Vale la seguente proprietà:

$$[E(XY)]^2 \leq E(X^2)E(Y^2). \tag{2.1}$$

Dim. Per ogni $\theta \in \mathbb{R}$ si ha che:

$$0 \leq E[\theta|X| + |Y|]^2 = \theta^2 E(X^2) + 2\theta E[|XY|] + E(Y^2) := f(\theta);$$

dunque, il trinomio di secondo grado $f(\theta)$ è sempre ≥ 0, il che implica che il suo discriminante deve essere ≤ 0, ovvero

$$E[|XY|^2] - E(X^2)E(Y^2) \leq 0,$$

che prova (2.1).

Si definisce *coefficiente di correlazione* la quantità definita da:

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}.$$

E' facile mostrare che, se $a, b > 0$:

$$\rho_{aX,bY} = \rho_{X,Y}$$

(la dim. è lasciata per esercizio).
Inoltre:

$$-1 \leq \rho_{X,Y} \leq 1. \tag{2.2}$$

Per verificare (2.2), basta osservare che, applicando la proprietà (2.1) alle v.a. $X' = X - E(X)$ e $Y' = Y - E(Y)$, si ottiene

$$[cov(X,Y)]^2 \leq Var(X)Var(Y),$$

da cui segue che $\rho^2_{X,Y} \leq 1$, ovvero vale (2.2).

Se $\rho_{X,Y} = -1$, le v.a. X e Y si dicono massimamente negativamente correlate, se $\rho_{X,Y} = 1$, le v.a. X e Y si dicono massimamente positivamente correlate. Si può dimostrare che:

$$\rho^2_{X,Y} = 1 \Leftrightarrow \text{ esistono } a \neq 0, b \in \mathbb{R} : \; P(Y = aX + b) = 1;$$

in particolare, $\rho_{X,Y} = 1$, se $a > 0$, $\rho_{X,Y} = -1$, se $a < 0$. Ciò significa che, con probabilità 1, la v.a. Y è una funzione lineare di X. Si osservi che, se fosse $a = 0$, si avrebbe $P(Y = b) = 1$, cioè Y sarebbe quasi certamente costante e uguale a $b = E(Y)$; allora, sarebbe $cov(X,Y) = E(XY) - E(X)E(Y) = E(bX) - E(X)b = bE(X) - bE(X) = 0$ e $\rho^2_{X,Y}$ varrebbe 0, non 1.