

## PROBABILITÀ – SCHEDA N. 4

### INDICI DI POSIZIONE E DI DISPERSIONE DI VARIABILI ALEATORIE

#### 1. Valore atteso (o media) e varianza di una variabile aleatoria discreta

Siamo interessati a individuare due indici numerici che descrivano sinteticamente una variabile aleatoria: la *media* ossia il numero che “meglio approssima” la variabile  $X$  e la *varianza* che fornisce una valutazione dell'errore che si commette sostituendo alla variabile  $X$  il valor medio.

- Il **valore medio** di  $X$  (indicato con  $E(X)$ ) si definisce come

$$E(X) = x_1 f_X(x_1) + x_2 f_X(x_2) + \dots = \sum_i x_i f_X(x_i)$$

Si usano, con lo stesso significato, anche i termini **valore atteso** o **speranza** di  $X$ . Il termine valore atteso non deve indurre a pensare che la media di  $X$  sia un valore che si spera di ottenere (infatti, come abbiamo già visto in statistica descrittiva, il valor medio non è necessariamente un valore assunto dalla variabile  $X$ ) e neanche che sia il valore più probabile. La definizione coincide con quella di media (aritmetica o pesata) o baricentro dei valori assunti da  $X$ , ciascuno pesato con la sua probabilità.

- La **varianza** di  $X$  (indicata con  $Var(X)$ ) si definisce come

$$Var(X) = E((X - E(X))^2) = \sum_i (x_i - E(X))^2 f_X(x_i)$$

ovvero

$$Var(X) = E(X^2) - E(X)^2$$

La varianza è la media del quadrato degli scarti dal baricentro dei possibili valori assunti dalla  $X$ , dove ogni valore è pesato dalla sua probabilità. La varianza di  $X$  è un indice (positivo) di dispersione, in quanto la varianza di  $X$  è nulla solo se la variabile aleatoria è costante; più la  $X$  assume valori molto dispersi intorno al valore atteso più la varianza è grande.

- La **deviazione standard (o anche scarto quadratico medio)** di  $X$  è definito come la radice quadrata positiva della varianza.

**Nota:** Se  $X$  non assume un numero finito di valori le definizioni precedenti necessitano di condizioni aggiuntive.

Il fatto che la varianza misura la distanza dei valori assunti dalla  $X$  dal valore atteso di  $X$  è espresso dalla seguente disuguaglianza, detta **disuguaglianza di Chebychev**.

$$\forall \delta > 0, \quad P(|X - E(X)| > \delta) \leq \frac{Var(X)}{\delta^2}$$

Questa proprietà afferma che la probabilità che la variabile  $X$  assuma valori che distano dal valore atteso più di un numero  $\delta$ , è minore o uguale del rapporto fra la varianza e il quadrato di  $\delta$ . In particolare la disuguaglianza di Chebychev afferma che, se la varianza è piccola, allora la probabilità che  $X$  assuma valori lontani dal valore atteso è anch'essa piccola.

#### ESEMPIO 1.

Supponiamo di avere un'urna con 7 palline, di cui 2 rosse e 5 verdi. Estraiamo dall'urna due palline. Indichiamo con  $X$  il numero di palline rosse estratte. La variabile aleatoria  $X$  assume il valore 0 se abbiamo estratto 2 volte una pallina verde;  $X$  assume il valore 1 se abbiamo ottenuto verde alla prima estrazione e rosso alla seconda oppure rosso alla prima e verde alla seconda;  $X$  assume il valore 2 se abbiamo estratto 2 volte una pallina rossa.

a) Consideriamo prima il caso con reinserimento nell'urna (ossia il caso in cui, una volta estratta, la pallina viene reinserita nell'urna prima di effettuare la seconda estrazione).

$$P(X=0) = 5/7 \times 5/7 = 25/49,$$

$$P(X=1) = 5/7 \times 2/7 + 2/7 \times 5/7 = 2 \times 5/7 \times 2/7 = 20/49,$$

$$P(X=2) = 2/7 \times 2/7 = 4/49$$

Allora:  $E(X) = 0 \times 25/49 + 1 \times 20/49 + 2 \times 4/49 = 28/49 = 4/7$ .

b) Consideriamo ora il caso senza reinserimento (ossia il caso in cui, una volta estratta, la pallina NON viene reinserita nell'urna).

Indichiamo con  $Y$  il numero di palline rosse con questo tipo di estrazione. I valori assunti da  $Y$  sono gli stessi di quelli assunti da  $X$  ma la densità è diversa.

Abbiamo infatti che

$$P(X=0) = 5/7 \times 4/6 = 20/42,$$

$$P(X=1) = 5/7 \times 2/6 + 2/7 \times 5/6 = 2 \times 5/7 \times 2/6 = 20/42,$$

$$P(X=2) = 2/7 \times 2/6 = 2/42$$

Il valore atteso è  $E(Y) = 20/42 + 20/42 + 2/42 = 42/42 = 1$ .

Osserviamo che i valori attesi di  $X$  e di  $Y$  sono uguali. Possiamo dedurre che anche le varianze di  $X$  e  $Y$  coincidono?

Si può dimostrare che, in generale:

- il valore atteso della variabile aleatoria che indica il numero di "successi" in estrazioni *con reinserimento* della pallina è uguale al valore atteso della variabile aleatoria che indica il numero di "successi" in estrazioni *senza reinserimento* della pallina.
- la varianza del caso senza reinserimento è sempre più piccola di quella del caso con reinserimento.

### ESEMPIO 2. "Testa o croce"

Consideriamo un gioco tra due giocatori A e B che consiste nel lanciare una moneta e vincere una certa somma dipendente dal risultato del lancio, secondo determinate regole. Supponiamo che A scelga di puntare su "testa". Se esce "testa" B paga ad A una somma  $s_1$ ; se esce "croce" A paga a B una somma  $s_2$ .

Quando il gioco è equo? Quando il valore atteso della vincita (o della perdita) è uguale per ciascuno dei due.

Se la moneta è equilibrata il valore atteso della vincita per A è:

$$\frac{1}{2} s_1 - \frac{1}{2} s_2$$

Il gioco è equo se  $\frac{1}{2} s_1 - \frac{1}{2} s_2 = 0$ , ossia se  $s_1 = s_2$ .

La moneta può anche non essere equilibrata.

In generale, se indichiamo con  $p$  la probabilità che esca "testa", la variabile aleatoria che indica quanto si "intasca", ha la distribuzione indicata nella seguente tabella:

	somma intascata	probabilità
esce "testa"	$s_1$	$p$
esce "croce"	$- s_2$	$1-p$

Quindi il suo valore atteso è:

$$s_1 p - s_2 (1 - p)$$

Cambiamo la regole. Se la regola è che tu vinci  $3s$  nel caso in cui esca testa e che tu paghi  $5s$  nel caso in cui esca croce, allora, affinché il gioco sia equo, la probabilità di ottenere testa non può essere uguale a quella di ottenere croce. Quanto dovrebbe essere? Basta impostare l'equazione

$$3sp - 5s(1-p) = 0$$

Eseguendo i calcoli si ottiene:  $8ps - 5s = 0$  da cui:  $p = \frac{5}{8}$

In questo caso, affinché il gioco sia equo, la probabilità dell'evento "esce testa" dovrebbe essere  $\frac{5}{8}$ , mentre la probabilità dell'evento "esce croce" dovrebbe essere  $\frac{3}{8}$ .

## 2. Valore atteso e varianza di una variabile aleatoria ottenuta come trasformazione lineare di una variabile aleatoria discreta

Come variano valore atteso e varianza quando trasformiamo la variabile X?

Se  $Y = aX + b$ , ovvero se Y si ottiene da X *tramite una trasformazione lineare*, allora

$$E(Y) = a E(X) + b \quad \text{Var}(Y) = a^2 \text{Var}(X)$$

Queste uguaglianze valgono **solo** per trasformazioni lineari. *Non* è vero, per esempio, che  $E(X^2) = (E(X))^2$  (da ciò seguirebbe che la varianza è sempre nulla!!!). In generale, abbiamo che

$$E(g(X)) = \sum_i g(x_i) f_x(x_i)$$

$$\text{Var}(g(X)) = E(g(X) - E(g(X)))^2 \quad \text{o, equivalentemente,} \quad \text{Var}(g(X)) = E(g(X)^2) - E(g(X))^2.$$

### ESEMPIO 5

Riprendiamo l'esempio 1 a) per verificare che se la trasformazione non è lineare il valore atteso e la varianza non si trasformano in modo semplice. Consideriamo la variabile casuale X che assume i valori -1, 0, 2 con probabilità rispettivamente 0.4, 0.1, 0.5. La variabile aleatoria  $Y = X^2$  assume valori 0, 1, 4 con probabilità rispettivamente 0.1, 0.4, 0.5.

Abbiamo quindi:

$$E(X) = -1 \times 0.4 + 0 \times 0.1 + 2 \times 0.5 = 0.6$$

$$\text{Var}(X) = [(-1)^2 \times 0.4 + 0^2 \times 0.1 + 2^2 \times 0.5] - 0.6^2 = 2.04$$

$$E(Y) = 1 \times 0.4 + 0 \times 0.1 + 4 \times 0.5 = 2.4$$

$$\text{Var}(Y) = [1^2 \times 0.4 + 0^2 \times 0.1 + 4^2 \times 0.5] - 2.4^2 = 2.64$$

Osserviamo appunto che  $E(X^2) \neq (E(X))^2$ .