1 Valore di aspettazione di una v.a. discreta

Sia $X : \Omega \longrightarrow \mathbb{R}$ una v.a. discreta, che assume valori $x_1, x_2, \ldots$ con probabilità $p_i = P(X = x_i),\ i = 1, 2, \ldots$

Definizione
Si dice che X ha valore di aspettazione finito se

$$\sum_i |x_i| p_i < +\infty.$$

In tal caso, la quantità

$$E(X) = \sum_i x_i p_i = \sum_i x_i P(X = x_i),$$

si chiama *valore di aspettazione di* X*, o media probabilistica, media pesata, o semplicemente media di* X. Naturalmente se X assume solo un numero finito di valori, la serie è una somma finita, e non c'è bisogno di imporre la convergenza assoluta.

Esempio 1 Supponiamo di estrarre una pallina da un'urna contenente 5 palline numerate da 1 a 5. Nell'ipotesi di estrazioni indipendenti, ed eventi elementari equiprobabili, possiamo identificare lo spazio degli eventi con $\Omega = \{1, 2, 3, 4, 5\}$ e considerare la v.a. discreta $X : \Omega \longrightarrow \{1, 2, 3, 4, 5\}$, che indica la pallina estratta. Qui, la v.a. X assume 5 valori $\{x_1, x_2, x_3, x_4, x_5\} = \{1, 2, 3, 4, 5\}$ e il range di X è $X(\Omega) = \{x_1, x_2, x_3, x_4, x_5\} \equiv \Omega$. Per $i = 1, 2, 3, 4, 5$, abbiamo:

$$p_i = P(\{\omega \in \Omega : X(\omega) = i\})$$

oppure, più semplicemente, omettendo $\omega \in \Omega :$

$$p_i = P(X = i) = \frac{1}{5}\ .$$

Si dice che la v.a. X ha distribuzione uniforme sull'insieme $\{1, 2, 3, 4, 5\}$. Allora:

$$E(X) = \sum_{i=1}^{5} i\ p_i = \sum_{i=1}^{5} i \cdot \frac{1}{5} = \frac{1}{5}(1 + 2 + 3 + 4 + 5) = \frac{15}{5} = 3.$$

In effetti, la media dei numeri estratti è 3, come uno si aspetta; in questo caso, infatti, $E(X)$ coincide con la *media aritmetica* dei valori $\{1, 2, 3, 4, 5\}$. Ciò è dovuto al fatto che risulta $p_i = \frac{1}{5}$ per ogni valore di $i \in \Omega$, ma naturalmente non sempre è cosi'.
Il valore di aspettazione $E(X) = \sum_i^n x_i p_i$ si dice anche media pesata con pesi p_i, per distinguerla dalla media aritmetica con pesi $p_i = \frac{1}{n}$, tutti uguali (nell'esempio di sopra $n = 5$ e $p_i = \frac{1}{5}$).
Per una v.a. con distribuzione uniforme sull'insieme $\{x_1, x_2, \ldots, x_n\}$, si ha

$$E(X) = \frac{x_1 + x_2 + \cdots + x_n}{n} \equiv \bar{X},$$

dove $\bar{X}$ denota la media aritmetica.
Modificando un po' l'esempio, supponiamo ora che l'estrazione della pallina non sia *veramente a caso*, ma che la pallina 5 sia privilegiata, rispetto alle altre, nel senso che essa viene

estratta più spesso delle altre; supponiamo, ad esempio, che $p_5 = \frac{2}{5}$ e $p_i = \frac{3}{20}$, $i = 1, 2, 3, 4$. Naturalmente, come deve essere, risulta $p_1 + p_2 + p_3 + p_4 + p_5 = 4 \cdot \frac{3}{20} + \frac{2}{5} = \frac{3}{5} + \frac{2}{5} = 1$. Se X indica la pallina estratta, si ha ora:

$$E(X) = \frac{3}{20} \cdot (1 + 2 + 3 + 4) + \frac{2}{5} \cdot 5 = \frac{3}{20} \cdot 10 + \frac{2}{5} \cdot 5 = \frac{3}{2} + 2 = \frac{7}{2} = 4.5$$

Come si vede, ora la media è più spostata verso il valore più probabile, ovvero 5. Se fosse stato più probabile estrarre la pallina 1, invece che le altre, avremmo ottenuto un valore più spostato verso il valore 1.

Sia $\underline{X} = (X_1, X_2, \dots, X_n)$ un vettore aleatorio con densità discreta $p(\underline{x}) = p(x_1, x_2, \dots, x_n)$, e $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ una funzione di n variabili reali; consideriamo la v.a. definita da $Y = g(\underline{X})$.

Definizione
Si pone

$$E(Y) = E(g(\underline{X})) = \sum_{x_1, x_2, \dots, x_n} g(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n),$$

se la serie converge assolutamente.

1. Per $\lambda \in \mathbb{R}$, sia $g : \mathbb{R} \longrightarrow \mathbb{R}$ definita da $g(x) = \lambda x$; allora si ha:

$$E(\lambda X) = \lambda E(X).$$

Infatti:

$$E(\lambda X) = E(g(X)) = \sum_x g(x)p(x) = \sum_x \lambda x p(x) = \lambda \sum_x x p(x) = \lambda E(X).$$

2. Se $\underline{X} = (X_1, X_2)$ è una v.a. bidimensionale con densità discreta $p(x_1, x_2)$, e $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ è definita da $g(x_1, x_2) = x_1 + x_2$, si ottiene:

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

Infatti, sotto la condizione che le serie scritte convergano assolutamente (se $\underline{X}$ assume solo un numero finito di coppie, le serie sono somme finite e non c'è bisogno di questa condizione), si ha:

$$E(X_1 + X_2) = \sum_{x_1} \sum_{x_2} (x_1 + x_2) p(x_1, x_2)$$

$$= \sum_{x_1} x_1 \sum_{x_2} p(x_1, x_2) + \sum_{x_2} x_2 \sum_{x_1} p(x_1, x_2) = \sum_{x_1} x_1 p^{X_1}(x_1) + \sum_{x_2} x_2 p^{X_2}(x_2) = E(X_1) + E(X_2),$$

dove $p^{X_1}(x_1) = \sum_{x_2} p(x_1, x_2)$ è la densità marginale di X_1 e $p^{X_2}(x_2) = \sum_{x_1} p(x_1, x_2)$ è la densità marginale di X_2.

3. Sia ora $g(x_1, x_2) = x_1 x_2$ e supponiamo che X_1 e X_2 siano v.a. indipendenti. Allora:

$$E(X_1 X_2) = E(X_1) E(X_2).$$

Infatti, si ha:

$$E(X_1X_2) = \sum_{x_1}\sum_{x_2} x_1x_2p(x_1,x_2) =$$

(siccome X_1 e X_2 sono indipendenti, ovvero $p(x_1,x_2) = p^{X_1}(x_1)p^{X_2}(x_2)$)

$$\left[\sum_{x_1} x_1p^{X_1}(x_1)\right]\left[\sum_{x_2} x_2p^{X_2}(x_2)\right] = E(X_1)E(X_2).$$

In generale, se X_1 e X_2 non sono indipendenti, risulta $E(X_1X_2) \neq E(X_1)E(X_2)$, anche se, in casi speciali, pur essendo X_1 e X_2 dipendenti, la media del loro prodotto può uguagliare il prodotto delle medie (*v.a. scorrelate*).

Proposition 1.1 Siano X e Y v.a. con medie $E(X), E(Y)$ finite.
(i) Se $P(X \geq Y) = 1$, allora $E(X) \geq E(Y)$ e l'uguaglianza è possibile se e solo se $P(X = Y) = 1$;
(ii) $|E(X)| \leq E(|X|)$.

Dim. (i) Sia $Z = X - Y$; per ipotesi risulta $P(Z \geq 0) = 1$, e ciò implica che i valori z_i assunti da Z sono tutti non negativi. Dunque $E(Z) = \sum_{z_i} z_iP(Z = z_i) \geq 0$ e quindi $E(X - Y) = E(X) - E(Y) \geq 0$, ovvero $E(X) \geq E(Y)$.
Inoltre, $E(X) = E(Y)$ se e solo se $E(Z) = 0$, ovvero $P(Z = z_i) = 0$ per ogni $z_i \neq 0$, e quindi $P(Z = 0) = 1$.
(ii) Poiché $-|X| \leq X \leq |X|$, per (i) si ha $-E(|X|) \leq E(X) \leq E(|X|)$, ovvero $|E(X)| \leq E(|X|)$.

Osservazione Dai punti precedenti e dalla Proposizione 1.1, segue che $E(X)$ è un funzionale lineare e positivo, cioè, per ogni $\alpha, \beta \in \mathbb{R}$, risulta $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$, qualora esistano finite le medie di X e Y; positivo significa che $E(X) \geq 0$ se X è una v.a. non negativa.

Esempio 2 (media di v.a. $0-1$, di Bernoulli)
Sia $X \in \{0, 1\}$, con $P(X = 1) = p \in (0, 1)$ e $P(X = 0) = 1 - p$. Allora $E(X) = p$.
Infatti:

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Esempio 3 (media di v.a. Binomiale)
Sia $X \sim B(n, p)$, allora $E(X) = np$.
Infatti:

$$p_k = P(X = k) = \binom{n}{k} p^k(1 - p)^{n-k},\ k = 0, 1, \dots, n$$

Quindi:

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k(1 - p)^{n-k}.$$

Invece di calcolare questa somma, ragioniamo come segue: si può scrivere $X = X_1 + X_2 + \cdots + X_n$ ove le v.a. X_i sono indipendenti e di Bernoulli, cioè $X_i \sim B(1,p)$. Allora:

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np.$$

Esempio 4 (media di v.a. di Poisson)

Sia X una v.a. di Poisson, di parametro $\lambda > 0$. Allora $E(X) = \lambda$.
Infatti:

$$p_k = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, \ldots,$$

quindi:

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

$$= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

Esempio 5 (media di v.a. Ipergeometrica)
Se X è una v.a. Ipergeometrica di parametri (r, b, n) si ha $E(X) = \frac{nr}{b+r}$. Possiamo pensare, ad esempio, che X conti il numero di palline rosse estrattte in n estrazioni senza rimpiazzo da un'urna contenente r palline rosse e b bianche. Allora, come è noto:

$$p_k = P(X = k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}$$

e

$$E(X) = \sum_k k\frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}.$$

Invece di calcolare questa somma, ragioniamo come segue: si può scrivere ancora $X = X_1 + X_2 + \cdots + X_n$ ove le v.a. X_i sono di Bernoulli, cioè $X_i \sim B(1,p)$, con $p = \frac{r}{b+r}$, ma **dipendenti**. Quindi:

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np = \frac{nr}{b+r}\ .$$

Osserviamo che, nonostante ora le v.a. X_i non siano indipendenti, abbiamo usato che la media di una somma di v.a uguaglia la somma delle medie.
Inoltre, il numero medio di palline rosse estratte è lo stesso di quello che si otterrebbe in n estrazioni con rimpiazzo; infatti, se le estrazioni avvenissero con rimpiazzo, la probabilità di estrarre una pallina rossa in ogni estrazione sarebbe $p = \frac{r}{b+r}$ e ponendo ancora $X = \#$ di palline rosse estratte, si avrebbe $X \sim B(n,p)$, e quindi $E(X) = np = \frac{nr}{b+r}$.

Esempio 6 (media di v.a. Geometrica e di una v.a. Geometrica modificata)
Se $X \sim Geom(p),\ p \in (0,1)$, si ha $E(X) = (1-p)/p$. Infatti

$$E(X) = \sum_{k=0}^{\infty} k \cdot p(1-p)^k = p \sum_{k=1}^{\infty} k(1-p)^k;$$

posto $k-1=j$, si ottiene:

$$E(X) = p \sum_{j=0}^{\infty} (j+1)(1-p)^{j+1} = p(1-p) \left[\sum_{j=0}^{\infty} j(1-p)^j + \sum_{j=0}^{\infty} (1-p)^j \right]$$

$$= (1-p) \sum_{j=0}^{\infty} j \cdot p(1-p)^j + p(1-p) \frac{1}{1-(1-p)} = (1-p)E(X) + 1 - p.$$

Quindi, risolvendo l'equazione rispetto a $E(X)$, si ottiene infine:

$$E(X) = (1-p)/p = 1/p - 1.$$

Se T è l'istante di primo successo (Geom modificata), si può scrivere $T-1 = X \sim Geom(p)$; dunque

$$E(T-1) = E(X), \text{ ovvero } E(T) - 1 = E(X) \ \Rightarrow E(T) = E(X) + 1 = \frac{1}{p} - 1 + 1 = \frac{1}{p}\ .$$

La media di T è inversamente proporzionale a p; infatti, se p è piccola, occorrerà in media un tempo grande affinché si verifichi l'evento successo, se p è grande, occorrerà in media un tempo piccolo.

Esempio 7 (media di v.a. di Pascal, o istante di $k-$mo successo)
Sia $T_k,\ k = 1, 2, \dots$ 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 $p \in (0,1)$; per $k=1$, si ottiene T_1, ovvero l'istante di primo successo, che abbiamo indicato precedentemente semplicemente con T, e che ha densità Geometrica modificata, di parametro p. Ricordiamo che la densità di T_k è:

$$P(T_k = i) = \binom{i-1}{k-1} p^k (1-p)^{i-k},\ i = k, k+1, \dots$$

Per $k=1$, ritorna la densità della Geometrica modificata, cioè

$$P(T_1 = i) = p(1-p)^{i-1},\ i = k, k+1, \dots$$

Si ha:

$$E(T_k) = \frac{k}{p}\ .$$

Notare che, per $k=1$, ritorna la media dell'istante di primo successo T_1.
Verifichiamo ora questo risultato. Per definizione, la media di T_k è:

$$E(T_k) = \sum_{i=k}^{\infty} i \binom{i-1}{k-1} p^k (1-p)^{i-k}$$

$$= \sum_{i=k}^{\infty} \frac{i!}{(k-1)!(i-k)!} p^k (1-p)^{i-k} =$$

$$= k \sum_{i=k}^{\infty} \binom{i}{k} p^k (1-p)^{i-k};$$

posto $i = h - 1$, l'ultima quantità si può scrivere:

$$\frac{k}{p} \sum_{h=k+1}^{\infty} \binom{h-1}{k} p^{k+1} (1-p)^{h-k-1} = \frac{k}{p},$$

visto che l'ultima serie ha per somma 1; infatti, se consideriamo l'istante del $k+1-$esimo successo, ovvero T_{k+1}, si ha:

$$P(T_{k+1} = h) = \binom{h-1}{k} p^{k+1} (1-p)^{h-k-1}, \ h = k+1, k+2, \ldots$$

e naturalmente

$$\sum_{h=k+1}^{\infty} P(T_{k+1} = h) = \sum_{h=k+1}^{\infty} \binom{h-1}{k} p^{k+1} (1-p)^{h-k-1} = 1,$$

come deve essere per ogni v.a.