Azzolini Riccardo 2020-03-24

Variabili aleatorie discrete

1 Problema: evento che è il risultato di un calcolo

Problema: Da un'urna contenente sei palline, numerate da 1 a 6, se ne estraggono due senza rimpiazzo (cioè senza reinserire ciascuna pallina estratta nell'urna, in modo che le due palline estratte siano sicuramente diverse). Qual è la probabilità che i numeri estratti differiscano al più di 2?

Indicando con Y_1 e Y_2, rispettivamente, i valori della prima e della seconda pallina estratta, e con $Y = |Y_1 - Y_2|$ la loro differenza, si è interessati ai casi in cui $Y \leq 2$, ovvero $|Y_1 - Y_2| \leq 2$. Un modo per determinare la probabilità di questi casi è considerare tutte le coppie $(Y_1 = y_1, Y_2 = y_2)$ di valori delle palline tali che $|Y_1 - Y_2| \leq 2$, e sommare le probabilità di ciascuna di esse (ciò è ammesso perché le coppie sono eventi disgiunti, in quanto elementari, quindi la probabilità dell'unione è data dalla somma delle probabilità).

Siccome il numero totale di coppie $(Y_1 = y_1, Y_2 = y_2)$ che possono essere estratte è

$$D_{6,2} = 6 \cdot 5 = 30$$

ed è ragionevole supporre che la probabilità sia uniforme, si può affermare che ciascuna coppia ha probabilità $\frac{1}{30}$.

L'ultimo passo è contare le coppie che verificano la proprietà $|Y_1 - Y_2| \leq 2$. Non essendo molto numerose, le si può contare semplicemente enumerandole tutte: esse sono

$$\begin{array}{llll}
(1,2) & (1,3) & & \\
(2,1) & (2,3) & (2,4) & \\
(3,1) & (3,2) & (3,4) & (3,5) \\
(4,2) & (4,3) & (4,5) & (4,6) \\
(5,3) & (5,4) & (5,6) & \\
(6,4) & (6,5) & &
\end{array}$$

cioè 18 coppie di palline.

Infine, per calcolare la probabilità che valga $|Y_1 - Y_2| \leq 2$, è sufficiente sommare le probabilità di queste 18 coppie, e, siccome esse sono equiprobabili, ciò equivale a moltiplicare per 18 la probabilità di una singola coppia:

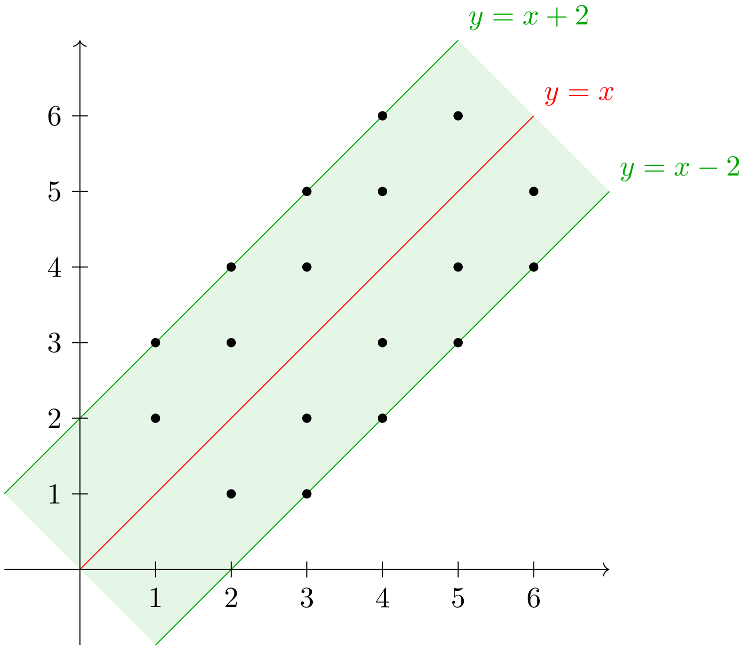
$$P(|Y_1 - Y_2| \leq 2) = 18 \cdot \frac{1}{30} = \frac{18}{30} = \frac{3}{5}$$

Le variabili Y_1, Y_2 e $Y = |Y_1 - Y_2|$ sono esempi di *variabili aleatorie*, cioè, informalmente, variabili che assumono valori numerici in funzione di un evento aleatorio.

Osservazione: Le coppie $(Y_1 = y_1, Y_2 = y_2)$ possono essere interpretate come punti in $\mathbb{R}^2$, e quindi rappresentate nel piano. Allora, riscrivendo la relazione $|Y_1 - Y_2| \le 2$ come

$$\begin{gathered} -2 \le Y_1 - Y_2 \le 2 \\ Y_2 - 2 \le Y_1 \le Y_2 + 2 \end{gathered}$$

si osserva che i punti da considerare sono quelli nella "striscia" compresa tra le rette $y = x - 2$ e $y = x + 2$ (includendo i punti direttamente su tali rette, ma escludendo quelli sulla diagonale $y = x$, che corrispondono all'estrazione "impossibile" di due palline uguali):



2 Problema: evento che è un confronto

Problema: Una moneta e un dado vengono lanciati insieme ripetutamente. Qual è la probabilità che la moneta dia testa prima che il dado dia 6?

Innanzitutto, si definiscono due variabili aleatorie S e T:

- S è il numero del lancio nel quale il dado dà 6 per la prima volta;
- T è il numero del primo lancio nel quale la moneta dà testa.

Ad esempio, se si verificasse la sequenza di lanci

Lancio	1	2	3	4	5
Dado	3	1	6	5	6
Moneta	croce	testa	croce	testa	testa

si avrebbero $S = 3$ e $T = 2$.

Siccome i lanci del dado sono tra loro indipendenti (in particolare, costituiscono uno schema successo-insuccesso), la probabilità $P(S = x)$ che il primo lancio a dare un 6 sia l'x-esimo è data dal prodotto delle probabilità che $x - 1$ lanci diano un risultato diverso da 6 ($\frac{5}{6}$) e che un lancio dia risultato 6 ($\frac{1}{6}$):

$$P(S = x) = \left(\frac{5}{6}\right)^{x-1} \frac{1}{6}$$

Analogamente, i lanci della moneta sono indipendenti, quindi:

$$P(T = y) = \left(\frac{1}{2}\right)^{y-1} \frac{1}{2} = \left(\frac{1}{2}\right)^{y}$$

L'evento richiesto dal problema è $T < S$. La sua probabilità, $P(T < S)$, può essere riformulata come $P((S, T) \in A)$, cioè come la probabilità che la coppia (T, S) appartenga a un opportuno insieme A, contenente tutte le coppie di numeri tali che, appunto, $T < S$. Dato che le coppie di numeri sono eventi elementari, e quindi disgiunti, $P((S, T) \in A)$ corrisponde alla somma delle probabilità delle singole coppie appartenenti ad A:

$$P(T < S) = P((S, T) \in A) = \sum_{(x,y) \in A} P(x, y)$$

Considerando che anche i lanci del dado e quelli della moneta sono indipendenti gli uni dagli altri, ovvero che S e T sono indipendenti, la probabilità di una di queste coppie $P(x, y)$, è uguale al prodotto di $P(S = x)$ e $P(T = y)$, purché x e y siano numeri naturali maggiori di 0 (in quanto rappresentano un numero di lanci):

$$P(x, y) = \begin{cases} \left(\frac{5}{6}\right)^{x-1} \frac{1}{6} \left(\frac{1}{2}\right)^{y} & \text{se } x, y = 1, 2, \ldots \\ 0 & \text{altrimenti} \end{cases}$$

A questo punto, rimane solo da calcolare la somma delle probabilità di tutte le coppie corrispondenti all'evento $T < S$:

$$\begin{aligned} P(T < S) = P((S, T) \in A) &= \sum_{(x,y) \in A} P(x, y) \\ &= \sum_{(x,y) \in A} \left(\frac{5}{6}\right)^{x-1} \frac{1}{6} \left(\frac{1}{2}\right)^{y} \end{aligned}$$

Le coppie $(x, y) \in A$ sono quelle con $y \geq 1$ e $x \geq y + 1$ (perché $y < x$):

$$= \sum_{y=1}^{\infty} \sum_{x=y+1}^{\infty} \left(\frac{5}{6}\right)^{x-1} \frac{1}{6} \left(\frac{1}{2}\right)^{y}$$

$$= \sum_{y=1}^{\infty} \sum_{x=y+1}^{\infty} \left(\frac{5}{6}\right)^{x} \underbrace{\left(\frac{5}{6}\right)^{-1} \frac{1}{6}}_{=\frac{1}{5}} \left(\frac{1}{2}\right)^{y}$$

$$= \frac{1}{5} \sum_{y=1}^{\infty} \sum_{x=y+1}^{\infty} \left(\frac{5}{6}\right)^{x} \left(\frac{1}{2}\right)^{y}$$

Per semplificare i calcoli, conviene riportare a 0 l'inizio della serie più interna:

$$= \frac{1}{5} \sum_{y=1}^{\infty} \sum_{x=0}^{\infty} \left(\frac{5}{6}\right)^{x+y+1} \left(\frac{1}{2}\right)^{y}$$

$$= \frac{1}{5} \sum_{y=1}^{\infty} \sum_{x=0}^{\infty} \left(\frac{5}{6}\right)^{x} \left(\frac{5}{6}\right)^{y+1} \left(\frac{1}{2}\right)^{y}$$

I fattori senza la x possono essere portati fuori dalla serie interna:

$$= \frac{1}{5} \sum_{y=1}^{\infty} \left(\left(\frac{5}{6}\right)^{y+1} \left(\frac{1}{2}\right)^{y} \sum_{x=0}^{\infty} \left(\frac{5}{6}\right)^{x} \right)$$

La serie interna è una geometrica di ragione $q = \frac{5}{6}$, perciò converge a $\frac{1}{1-q} = 6$:

$$= \frac{1}{5} \sum_{y=1}^{\infty} \left(\frac{5}{6}\right)^{y+1} \left(\frac{1}{2}\right)^{y} \cdot 6$$

$$= \frac{6}{5} \sum_{y=1}^{\infty} \left(\frac{5}{6}\right)^{y+1} \left(\frac{1}{2}\right)^{y}$$

Infine si riporta a 0 anche l'inizio della serie esterna, e si applica ancora la regola della serie geometrica:

$$\begin{aligned}
&= \frac{6}{5}\sum_{y=0}^{\infty}\left(\frac{5}{6}\right)^{y+2}\left(\frac{1}{2}\right)^{y+1} \\
&= \frac{6}{5}\sum_{y=0}^{\infty}\left(\frac{5}{6}\right)^{y}\left(\frac{5}{6}\right)^{2}\left(\frac{1}{2}\right)^{y}\frac{1}{2} \\
&= \frac{6}{5}\left(\frac{5}{6}\right)^{2}\frac{1}{2}\sum_{y=0}^{\infty}\left(\frac{5}{6}\cdot\frac{1}{2}\right)^{y} \\
&= \frac{5}{12}\sum_{y=0}^{\infty}\left(\frac{5}{12}\right)^{y} \\
&= \frac{5}{12}\cdot\frac{1}{1-\frac{5}{12}} = \frac{5}{12}\cdot\frac{12}{7} = \frac{5}{7}
\end{aligned}$$

Ricapitolando, la soluzione del problema è $P(T < S) = \frac{5}{7}$.

3 Variabile aleatoria

Definizione: Dato uno spazio di probabilità $(\Omega, \mathscr{A}, P)$, si dice **variabile aleatoria** (o **casuale**) un'applicazione $X : \Omega \to \mathbb{R}$ tale che, per ogni $t \in \mathbb{R}$, l'insieme $\{\omega \mid X(\omega) \leq t\}$ sia un evento, ovvero $\{\omega \mid X(\omega) \leq t\} \in \mathscr{A}$.

L'insieme $\{\omega \mid X(\omega) \leq t\}$ contiene tutti gli eventi elementari $\omega \in \Omega$ in corrispondenza dei quali la variabile aleatoria X assume un valore $\leq t$. Essendo tale insieme un evento, è possibile calcolarne la probabilità, $P(\{\omega \mid X(\omega) \leq t\})$. In altre parole, la definizione di variabile aleatoria richiede che sia possibile calcolare la probabilità che il valore assunto da tale variabile sia minore o uguale a un qualsiasi valore $t \in \mathbb{R}$.

In realtà, questa definizione è sufficiente per calcolare anche le probabilità di $\{\omega \mid X(\omega) > t\}$, $\{\omega \mid a < X(\omega) \leq b\}$ e $\{\omega \mid X(\omega) = x\}$. Infatti, sfruttando gli assiomi della σ-algebra $\mathscr{A}$, si dimostra che tutti questi tipi di insiemi sono eventi:

1. $\{\omega \mid X(\omega) > t\}$ è il complemento di $\{\omega \mid X(\omega) \leq t\}$;
2. $\{\omega \mid a < X(\omega) \leq b\}$ corrisponde a un'intersezione di eventi:

$$\{\omega \mid a < X(\omega) \leq b\} = \{\omega \mid X(\omega) > a\} \cap \{\omega \mid X(\omega) \leq b\}$$

3. $\{\omega \mid X(\omega) = x\}$ può essere ottenuto mediante un'intersezione infinita numerabile di eventi:
$$\{\omega \mid X(\omega) = x\} = \bigcap_{n=1}^{\infty} \left\{\omega \;\middle|\; x - \frac{1}{n} < X(\omega) \leq x\right\}$$

In generale, con questa definizione, $\{\omega \mid X(\omega) \in A\}$ *non* è un evento per qualsiasi tipo di insieme $A \subset \mathbb{R}$, ma i casi in cui si dimostra che lo sia sono solitamente sufficienti.

Nota: Spesso, per semplificare la notazione, gli eventi $\{\omega \mid X(\omega) \leq t\}$, $\{\omega \mid X(\omega) \in A\}$, ecc. vengono indicati con $\{X \leq t\}$, $\{X \in A\}$, ecc., e si scrive la probabilità $P(\{X \leq t\})$ come $P\{X \leq t\}$ o $P(X \leq t)$, omettendo le parentesi tonde oppure le graffe.

4 Distribuzione

Si dice **distribuzione** o **legge** di una variabile aleatoria X la "regola" che associa a ogni sottoinsieme $A \subset \mathbb{R}$ (tale che $\{X \in A\}$ sia un evento) la probabilità che X assuma valori appartenenti ad A, cioè $P\{X \in A\}$.

In termini formali, la distribuzione di X è l'applicazione $A \mapsto P\{X \in A\}$.

5 Funzione di ripartizione

Definizione: Data una variabile aleatoria X, si chiama **funzione di ripartizione** la funzione $F_X : \mathbb{R} \to [0, 1]$ definita da
$$F_X(t) = P\{X \leq t\}$$

6 Variabile aleatoria discreta

Definizione: Una variabile aleatoria X si dice **discreta** se il suo insieme immagine
$$\mathrm{Im}(X) = \{y \in \mathbb{R} \mid \exists \omega \in \Omega \;\; X(\omega) = y\}$$
ha cardinalità numerabile (cioè finita o infinita numerabile).

In altre parole, una variabile aleatoria è discreta se assume un numero al più infinito numerabile di valori distinti.

6.1 Densità discreta e funzione di ripartizione

Definizione: Data una variabile aleatoria discreta X, si dice **densità (di probabilità) discreta** la funzione

$$p(x) = P\{X = x\}$$

Nota: Talvolta, la densità discreta viene indicata con $f(x)$, invece che con $p(x)$.

La densità discreta $p(x) = P\{X = x\}$ può essere usata per calcolare la funzione di ripartizione $F_X(t) = P\{X \le t\}$. Infatti, siccome la variabile aleatoria è discreta, l'evento $\{X \le t\}$ può essere riscritto come un'unione numerabile di eventi elementari $\{X = x\}$,

$$\{X \le t\} = \bigcup_{x \le t} \{X = x\}$$

e tali eventi (essendo appunto elementari) sono disgiunti, perciò la probabilità della loro unione è data dalla somma delle singole probabilità:

$$F_X(t) = P\{X \le t\} = \sum_{x \le t} P\{X = x\} = \sum_{x \le t} p(x)$$

Viceversa, data la funzione di ripartizione, si può ricavare la densità discreta:[1]

$$F_X(x) - F_X(x-1) = P\{x - 1 < X < x\} = P\{X = x\} = p(x)$$

6.2 Costruzione di uno spazio di probabilità

Data una variabile aleatoria discreta X con densità $p(x)$, è sempre possibile costruire un corrispondente spazio di probabilità $(\Omega, \mathscr{A}, P)$. Infatti, è sufficiente:

- considerare come elementi (eventi elementari) dello spazio campionario Ω direttamente i valori assunti dalla variabile aleatoria, cioè scegliere $\Omega = \mathrm{Im}(X)$, ovvero $X(\omega) = \omega$;
- prendere la σ-algebra delle parti, $\mathscr{A} = \mathscr{P}(\Omega)$;
- definire la mappa di probabilità P in modo che corrisponda alla densità discreta di X:
$$P(\{\omega\}) = P\{X = \omega\} = p(\omega)$$

Quindi, affrontare un problema mediante una variabile aleatoria discreta equivale a ragionare direttamente su uno spazio di probabilità: cambia solo il linguaggio utilizzato.

[1] Questa formula vale solo se X assume valori interi. Per ottenere una formula che sia invece applicabile a qualsiasi variabile aleatoria discreta, sarebbe necessario porre $\mathrm{Im}(X) = \{x_1, x_2, \dots\}$, con $x_1 < x_2 < \cdots$, e scrivere la differenza $F_X(x_i) - F_X(x_{i-1})$.

7 Variabile aleatoria n-dimensionale discreta

Nei problemi dell'urna e dei lanci presentati prima, si considerano le coppie (Y_1, Y_2) e (S, T) di variabili aleatorie discrete. Per formalizzare la soluzione di tali problemi, è allora necessario dare un significato preciso alle coppie, o, più in generale, alle n-uple, di variabili aleatorie discrete.

Definizione: Una **variabile aleatoria n-dimensionale discreta** (o **vettore aleatorio discreto**) è un'applicazione

$$X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$$

tale che le applicazioni $X_1, \dots, X_n$ siano della variabili aleatorie discrete.

Per ogni n-upla di numeri reali $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, l'evento $\{X = x\}$ si verifica quando le singole variabili $X_1, \dots, X_n$ assumono rispettivamente i valori $x_1, \dots, x_n$:

$$\begin{aligned} \{X = x\} &= \{X_1 = x_1,\ \dots,\ X_n = x_n\} \\ &= \{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\} \end{aligned}$$

Allora, la funzione $p(x) = P\{X = x\}$ può essere scritta come

$$p(x) = P\{X = x\} = P(\{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\})$$

ed è chiamata **densità** (**di probabilità**) **congiunta** delle variabili aleatorie $X_1, \dots, X_n$.

Invece, l'evento $\{X_i = x_i\}$ fissa solo il valore x_i dell'i-esimo elemento della n-upla x, e la funzione $p_i(x_i) = P\{X_i = x_i\}$ che ne esprime la probabilità prende il nome di **densità marginale** della variabile aleatoria X_i. Essa può essere ricavata dalla densità congiunta $p(x)$, perché le n-uple che verificano l'evento $\{X_i = x_i\}$ sono tutte quelle ottenute al variare dei valori delle altre variabili,

$$\{X_i = x_i\} = \bigcup_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} \{X_1 = x_1,\ \dots,\ X_{i-1} = x_{i-1},\ \underbrace{X_i = x_i}_{\text{unico valore fissato}},\ X_{i+1} = x_{i+1},\ \dots,\ X_n = x_n\}$$

e ciascuna di queste n-uple è diversa dalle altre (poiché cambia sempre il valore di almeno una delle variabili), quindi si ha un'unione di eventi disgiunti, la cui probabilità è data dalla somma delle probabilità delle singole n-uple, ovvero da una somma sulla densità congiunta:

$$\begin{aligned} p_i(x_i) &= P\{X_i = x_i\} \\ &= \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} P\{X_1 = x_1,\ \dots,\ X_{i-1} = x_{i-1},\ X_i = x_i,\ X_{i+1} = x_{i+1},\ \dots,\ X_n = x_n\} \\ &= \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} p(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \end{aligned}$$

Invece, non è possibile, in generale, ricostruire la densità congiunta dalle densità marginali di $X_1, \dots, X_n$ (infatti, a densità congiunte diverse possono corrispondere le stesse densità marginali).

8 Variabili aleatorie indipendenti

Definizione: Le variabili aleatorie discrete $X_1, \dots, X_n$ sono **indipendenti** se e solo se, per ogni scelta di insiemi $A_1, \dots, A_n \subset \mathbb{R}$, si ha

$$P\{X_1 \in A_1,\ \dots,\ X_n \in A_n\} = P\{X_1 \in A_1\} \cdots P\{X_n \in A_n\}$$

Il significato intuitivo dell'indipendenza tra variabili aleatorie è analogo a quello dell'indipendenza tra eventi: la conoscenza dei valori assunti da alcune delle variabili non dà alcuna informazione sui valori delle altre.

Come caso particolare, se si considerano insiemi costituiti da singoli elementi ($A_i = \{x_i\}$), l'uguaglianza diventa

$$P\{X_1 = x_1,\ \dots,\ X_n = x_n\} = P\{X_1 = x_1\} \cdots P\{X_n = x_n\}$$

e, ponendo $x = (x_1, \dots, x_n)$, ciò si può riscrivere come

$$p(x) = p_1(x_1) \cdots p_n(x_n)$$

In altre parole, se le variabili sono indipendenti, la densità congiunta può essere calcolata da quelle marginali (mentre, come già detto, ciò non è vero in generale).

Inoltre, si può dimostrare che, viceversa, quando è verificata l'uguaglianza

$$p(x) = p_1(x_1) \cdots p_n(x_n)$$

allora vale anche

$$P\{X_1 \in A_1,\ \dots,\ X_n \in A_n\} = P\{X_1 \in A_1\} \cdots P\{X_n \in A_n\} \quad \forall A_1, \dots, A_n \subset \mathbb{R}$$

ovvero che le variabili $X_1, \dots, X_n$ sono indipendenti se e solo se la densità congiunta è il prodotto delle densità marginali. In questo modo, è possibile verificare facilmente se delle variabili aleatorie discrete siano indipendenti o meno.[2]

9 Densità condizionale

Definizione: Date due variabili aleatorie X e Y con densità congiunta p, la **densità condizionale** di X dato $Y = y$ è

$$\bar{p}_{X|Y}(x \mid y) = \frac{p(x, y)}{p_Y(y)}$$

[2] In particolare, per fare ciò, è sufficiente conoscere la densità congiunta delle variabili, dalla quale possono infatti essere ricavate le densità marginali.

dove p_Y è la densità marginale di Y.

Osservazione: Questa definizione è analoga a quella della probabilità condizionata tra eventi:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- la densità congiunta (al numeratore) corrisponde alla probabilità dell'intersezione $A \cap B$;
- la densità marginale p_Y (al denominatore) corrisponde alla probabilità dell'evento B sul quale è condizionata la probabilità $P(A \mid B)$.

10 Esercizio di notazione

Siano $X_1, \ldots, X_m, Y_1, \ldots, Y_k$ delle variabili aleatorie discrete indipendenti, e siano $\phi : \mathbb{R}^m \to \mathbb{R}$ e $\psi : \mathbb{R}^k \to \mathbb{R}$ delle applicazioni. Verificare che $\phi(X_1, \ldots, X_m)$ e $\psi(Y_1, \ldots, Y_k)$ sono indipendenti.

Innanzitutto, per brevità, si pongono $X = (X_1, \ldots, X_m)$ e $Y = (Y_1, \ldots, Y_k)$.

L'obiettivo è verificare che, per ogni $z, w \in \mathbb{R}$, valga la seguente uguaglianza:

$$P\{\phi(X) = z,\ \psi(Y) = w\} = P\{\phi(X) = z\}\, P\{\psi(Y) = w\}$$

L'evento $\{\phi(X) = z,\ \psi(Y) = w\}$ può essere riscritto come l'appartenenza di X e Y alle controimmagini $\phi^{-1}(z)$ e $\psi^{-1}(w)$:[3]

$$\begin{aligned} \{\phi(X) = z,\ \psi(Y) = w\} &= \{X \in \phi^{-1}(z),\ Y \in \psi^{-1}(w)\} \\ &= \bigcup_{\substack{x \in \phi^{-1}(z) \\ y \in \psi^{-1}(w)}} \{X = x,\ Y = y\} \end{aligned}$$

Questa è un'unione di eventi disgiunti, e la probabilità di ciascuno di questi eventi è data dalla densità congiunta $p(x, y)$. Allora:

$$\begin{aligned} P\{\phi(X) = z,\ \psi(Y) = w\} &= P\{X \in \phi^{-1}(z),\ Y \in \psi^{-1}(w)\} \\ &= \sum_{\substack{x \in \phi^{-1}(z) \\ y \in \psi^{-1}(w)}} p(x, y) \end{aligned}$$

[3] Queste controimmagini sono insiemi, e non singoli valori, perché non si sa se le applicazioni ϕ e ψ siano iniettive, quindi possono esistere, ad esempio, diversi elementi $x \in \mathbb{R}^m$ aventi la stessa immagine $\phi(x) = z$ (e lo stesso vale per ψ).

Per ipotesi, X e Y sono indipendenti, quindi $p(x,y) = p_X(x)\, p_Y(y)$:

$$= \sum_{\substack{x \in \phi^{-1}(z) \\ y \in \psi^{-1}(w)}} p_X(x)\, p_Y(y)$$

Gli indici della sommatoria possono essere separati:

$$\begin{aligned}
&= \sum_{x \in \phi^{-1}(z)} \sum_{y \in \psi^{-1}(w)} p_X(x)\, p_Y(y) \\
&= \sum_{x \in \phi^{-1}(z)} \left(p_X(x) \sum_{y \in \psi^{-1}(w)} p_Y(y) \right) \\
&= \left(\sum_{x \in \phi^{-1}(z)} p_X(x) \right) \left(\sum_{y \in \psi^{-1}(w)} p_Y(y) \right)
\end{aligned}$$

Infine, si eseguono in ordine inverso i passaggi iniziali, tornando dalle unioni di eventi elementari $\{X = x\}$ e $\{Y = y\}$ ai corrispondenti eventi di appartenenza alle controimmagini, e poi a $\{\phi(X) = z\}$ e $\{\psi(Y) = w\}$:

$$\begin{aligned}
&= P\{X \in \phi^{-1}(z)\}\; P\{Y \in \psi^{-1}(w)\} \\
&= P\{\phi(X) = z\}\; P\{\psi(Y) = w\} \qquad \square
\end{aligned}$$