Densità di probabilità del prodotto di due variabili casuali distribuite uniformemente

In questa dispensa, che presentiamo a semplice titolo di esercizio e applicazione di proprietà note, ci proponiamo di calcolare la densità di probabilità del prodotto di due variabili casuali indipendenti, distribuite in modo uniforme fra 0 e 1. Ciascuna delle due variabili, x_1 e x_2[1], ha una funzione di distribuzione data da:

$$h(x) = \begin{cases} 0 & \text{se } x < 0 \text{ o } x > 1 \\ 1 & \text{se } 0 \leq x \leq 1 \end{cases} \tag{1}$$

Siamo interessati a determinare la funzione densità di probabilità del prodotto $\mathsf{x}_1\mathsf{x}_2$. Questo risultato si può ottenere con due metodi.

1. Determinazione diretta della densità di probabilità

Se consideriamo una variabile casuale vettoriale di dimensione n: $\boldsymbol{\mathsf{x}} = \{\mathsf{x}_1, \mathsf{x}_2 \ldots \mathsf{x}_n\}$ la cui densità di probabilità è data da $f(x_1, x_2 \ldots x_n)$ e una nuova variabile di dimensione n funzione della prima $\boldsymbol{\mathsf{y}} = H(\boldsymbol{\mathsf{x}})$, la densità di probabilità di $\boldsymbol{\mathsf{y}}$ può essere espressa come

$$g(\boldsymbol{y}) = f\left(H^{-1}(\boldsymbol{y})\right) \cdot \left| J\left(\frac{\boldsymbol{x}}{\boldsymbol{y}}\right) \right| \tag{2}$$

dove J rappresenta il determinante della matrice jacobiana:

$$J\left(\frac{\boldsymbol{x}}{\boldsymbol{y}}\right) = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix} \tag{3}$$

Questo metodo si può applicare se esiste la funzione inversa tale che $\boldsymbol{\mathsf{x}} = H^{-1}(\boldsymbol{\mathsf{y}})$, ossia se esiste una relazione *biunivoca* fra $\boldsymbol{\mathsf{x}}$ e $\boldsymbol{\mathsf{y}}$.

Nel nostro caso identificheremo la nuova variabile $\boldsymbol{\mathsf{y}}$ con una variabile a 2 dimensioni, contenente come primo elemento $\mathsf{y}_1 = \mathsf{x}_1\mathsf{x}_2$, e sceglieremo il secondo elemento in modo da

[1] Per convenzione utilizziamo la grafia "x" per distinguere una variabile casuale da una variabile normale, per cui useremo la grafia "x".

verificare le condizioni di biunivocità. Possiamo ad esempio prendere $y_2 = x_1$. Si determina anche facilmente la relazione inversa:

$$\begin{cases} y_1 = x_1 x_2 \\ y_2 = x_1 \end{cases} \Rightarrow \begin{cases} x_1 = y_2 \\ x_2 = \frac{y_1}{y_2} \end{cases} \tag{4}$$

Si vede che anche le nuove variabili assumono valori compresi fra 0 e 1. Il determinante jacobiano è dato da

$$J\left(\frac{\boldsymbol{x}}{\boldsymbol{y}}\right) = \det\begin{pmatrix} 0 & 1 \\ \frac{1}{y_2} & -\frac{y_1}{y_2^2} \end{pmatrix} = -\frac{1}{y_2} \tag{5}$$

Possiamo scrivere dunque la (2) come

$$g(y_1, y_2) = h(y_2)\, h\left(\frac{y_1}{y_2}\right) \frac{1}{y_2} \tag{6}$$

A questo punto non ci resta che calcolare dalla (6) la densità di probabilità marginale di y_1 data da

$$g(y_1) = \int_{-\infty}^{\infty} h(y_2)\, h\left(\frac{y_1}{y_2}\right) \frac{1}{y_2}\, dy_2 \tag{7}$$

dove i limiti di integrazione infiniti hanno valore simbolico, dato che la variabile y_2 ha densità di probabilità non nulla solo nell'intervallo $[0, 1]$. La presenza di due funzioni h, come definite nella (1) implica che l'integrale dà contributi non nulli solo se

$$\begin{cases} 0 \leq y_2 \leq 1 \\ y_2 \geq y_1 \end{cases} \tag{8}$$

La (7) si riduce quindi a

$$g(y_1) = \int_{y_1}^{1} \frac{1}{y_2}\, dy_2 = -\log y_1 = \log\left(\frac{1}{y_1}\right) \tag{9}$$

La funzione trovata è sempre positiva nell'intervallo $(0, 1)$; verifichiamo anche la normalizzazione ossia calcoliamo

$$\int_0^1 -\log y\, dy = \lim_{\epsilon\to 0}\int_\epsilon^1 -\log y\, dy = \lim_{\epsilon\to 0}(1+\epsilon\,\log\epsilon-\epsilon) = 1 \tag{10}$$

Come esercizio, ripetiamo lo stesso procedimento con una scelta diversa per la variabile $\boldsymbol{y}$:

$$\begin{cases} y_1 &= x_1x_2 \\ y_2 &= \frac{x_2}{x_1}\end{cases} \quad\Rightarrow\quad \begin{cases} x_1 &= \sqrt{\frac{y_1}{y_2}} \\ x_2 &= \sqrt{y_1y_2}\end{cases} \tag{11}$$

In questo caso y_2 spazia da 0 a infinito. Per il determinante jacobiano si ha:

$$J\left(\frac{\boldsymbol{x}}{\boldsymbol{y}}\right) = \det\begin{pmatrix} \frac{\frac{1}{y_2}}{2\sqrt{\frac{y_1}{y_2}}} & \frac{-\frac{y_1}{y_2^2}}{2\sqrt{\frac{y_1}{y_2}}} \\ \frac{y_2}{2\sqrt{y_1y_2}} & \frac{y_1}{2\sqrt{y_1y_2}}\end{pmatrix} = \frac{1}{2y_2} \tag{12}$$

La densità di probabilità congiunta vale

$$g(y_1,y_2) = \frac{1}{2}\,h\left(\sqrt{\frac{y_1}{y_2}}\right)\,h\left(\sqrt{y_1y_2}\right)\,\frac{1}{y_2} \tag{13}$$

Dato che $0\le y_1\le 1$ e $y_2>0$ il prodotto delle funzioni h è non nullo se

$$\begin{cases} y_2 \ge y_1 \\ y_2 \le \frac{1}{y_1}\end{cases} \tag{14}$$

Conseguentemente

$$g(y_1) = \frac{1}{2}\int_{y_1}^{\frac{1}{y_1}}\frac{1}{y_2}\,dy_2 = \frac{1}{2}\left[\log\left(\frac{1}{y_1}\right)-\log y_1\right] = \log\left(\frac{1}{y_1}\right) \tag{15}$$

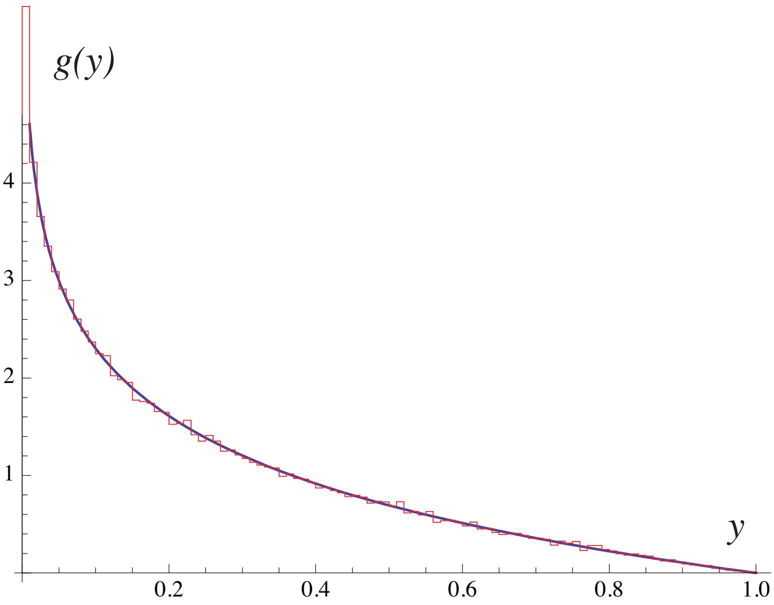

Fig. 1: densità di probabilità del prodotto di due variabili distribuite uniformemente fra 0 e 1.

In figura è rappresentata la densità di probabilità del prodotto, $g(y)$, in blu e una simulazione Monte–Carlo (istogramma in rosso) effettuata estraendo 100000 coppie di numeri pseudo–casuali.

2. Determinazione tramite la funzione di distribuzione

Si può calcolare la densità di probabilità partendo dalla funzione di distribuzione. Nel nostro caso la funzione è data da

$$G(y) = P(x_1 x_2 < y) \tag{16}$$

dove x_1 e x_2 sono distribuite secondo la (1). Il calcolo si esegue facilmente se si considera lo schema in figura 2.

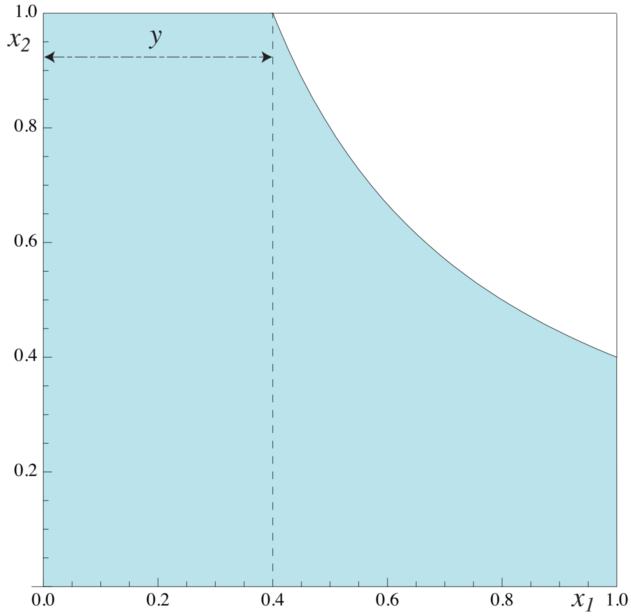

Fig. 2: zona per il calcolo della funzione di distribuzione del prodotto di due variabili distribuite uniformemente.

Il quadrato di lato unitario in figura rappresenta i possibili punti individuati da una coppia di valori casuali $\{x_1, x_2\}$. Data la distribuzione uniforme delle probabilità per le due variabili, la probabilità che un punto cada in una qualsiasi zona interna al quadrato è semplicemente data dall'area della zona stessa. Il tratto di curva disegnato è la funzione $x_2 = y/x_1$ e delimita la parte del quadrato per cui il valore del prodotto $x_1 x_2$ è minore o maggiore di y. La probabilità espressa dalla (16) è quindi data dall'area tinteggiata in figura. Essa è data da

$$G(y) = y + \int_y^1 \frac{y}{x_1}\, dx_1 = y - y \log y \tag{17}$$

Derivando la (17) si ottiene la densità di probabilità, che coincide ovviamente con la (9).

3. Estensione a intervalli diversi

Una volta ottenuta la (9), non è difficile estendere il risultato a situazioni in cui gli intervalli delle variabili sono diversi. Ad esempio, consideriamo il caso semplice in cui x_1 e

x_2 sono distribuite uniformemente fra -1 e 1, per cui la loro densità di probabilità vale:

$$h(x) = \begin{cases} \frac{1}{2} & \text{se } -1 \leq x \leq 1 \\ 0 & \text{se } x < -1 \text{ o } x > 1 \end{cases} \tag{18}$$

Possiamo notare che:

a. Il prodotto delle variabili è ora distribuito fra -1 e 1

b. Le probabilità che il prodotto sia positivo o negativo sono entrambe 1/2

c. Le probabilità di ottenere un prodotto y o un prodotto $-y$ sono uguali

d. Se consideriamo i casi $y > 0$, l'andamento della densità di probabilità è lo stesso che nella situazione studiata precedentemente, dato che la densità superficiale nel piano (x_1, x_2) è ancora costante, salvo per l'integrale che vale 1/2.

   Da queste considerazioni risulta immediatamente

$$g(y) = \frac{1}{2}\log\left(\frac{1}{|y|}\right) \tag{19}$$

In figura 3 è mostrata la densità di probabilità e l'istogramma ottenuto con metodo Monte-Carlo su 100000 coppie di valori.

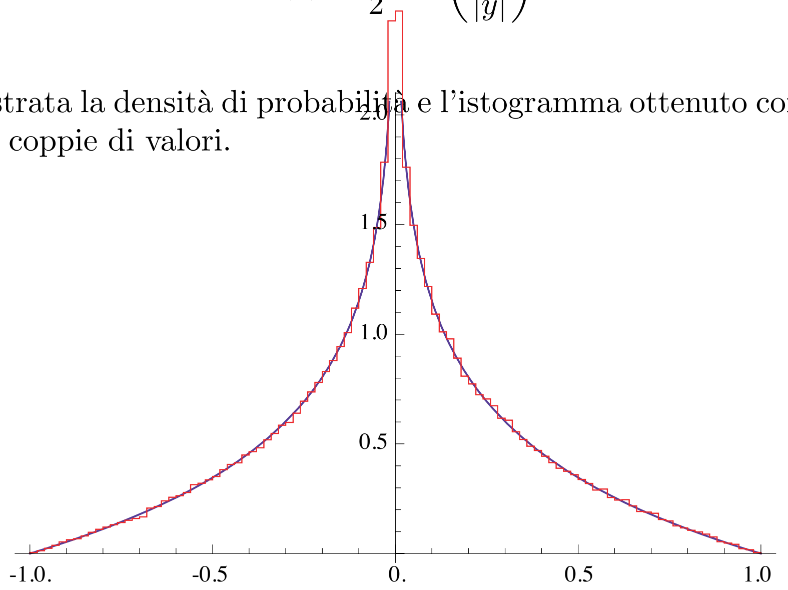

Fig. 3: densità di probabilità del prodotto di due variabili distribuite uniformemente fra -1 e 1.