

CALCOLO DELLE PROBABILITÀ E STATISTICA 2021/2022

**VARIABILI ALEATORIE
CONTINUE**



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

1 Introduzione

Le variabili aleatorie discrete descrivono quantità che possono assumere un numero finito o al più un'infinità numerabile di valori. Ci sono però quantità che assumono un'infinità *continua* di valori.

Esempio 1.1. *Si consideri un componente elettronico. Sia*

$$X = \text{“tempo di vita del componente”}.$$

In tal caso è naturale supporre che X possa assumere qualsiasi valore maggiore o uguale a zero, quindi “ $\mathcal{S}_X = [0, +\infty)$ ”.

Per descrivere tali quantità introduciamo le *variabili aleatorie continue*. Prima di fornire la definizione precisa di variabile aleatoria continua, premettiamo alcune osservazioni facendo riferimento alla variabile aleatoria X dell'Esempio 1.1.

Innanzitutto, poiché X può assumere un'infinità continua di valori, l'eventualità che ne assuma esattamente uno in particolare (ad esempio il numero $x = 3.45362$) è praticamente impossibile. Quindi, si dovrà avere che

$$p_X(x) = \mathbb{P}(X = x) = 0, \quad \forall x \in \mathbb{R}.$$

In altri termini, la densità discreta p_X per una variabile aleatoria continua è sempre identicamente uguale a zero, non gioca dunque alcun ruolo.

Tuttavia, se invece di considerare un singolo valore x si considera un intervallo di valori $[a, b] \subset \mathbb{R}$, con $a < b$, allora ci aspettiamo che

$$\mathbb{P}(a \leq X \leq b) > 0, \quad \text{se } [a, b] \subset [0, +\infty),$$

dove la richiesta $[a, b] \subset [0, +\infty)$ deriva dal fatto che nell'Esempio 1.1 la v.a. X è un “tempo”, quindi “ $\mathcal{S}_X = [0, +\infty)$ ”.

In conclusione, facendo sempre riferimento alla v.a. X dell'Esempio 1.1, ci troviamo di fronte al problema di conciliare queste due ragionevoli richieste:

- (i) $p_X(x) = \mathbb{P}(X = x) = 0$, per ogni $x \in \mathbb{R}$;
- (ii) $\mathbb{P}(a \leq X \leq b) > 0$, per ogni $[a, b] \subset [0, +\infty)$, con $a < b$.

Tale problema è risolto introducendo il concetto di variabile aleatoria continua. Infatti, se X è una variabile aleatoria continua allora esiste una funzione $f_X \geq 0$, chiamata *densità continua* di X , tale che

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx.$$

Nel caso dell'Esempio 1.1 una possibile scelta di f_X è la seguente:

$$f_X(x) = \begin{cases} 0, & x < 0, \\ e^{-x}, & x \geq 0. \end{cases}$$

È allora chiaro che le proprietà (i) e (ii) sono verificate. Infatti, ad esempio, per quanto riguarda la proprietà (i) si ha che

$$p_X(x) = \mathbb{P}(X = x) \stackrel{\substack{= \\ \uparrow \\ \mathbb{P}(X=x)=\mathbb{P}(x \leq X \leq x)}}{=} \int_x^x f_X(y) dy = 0, \quad \forall x \in \mathbb{R}.$$

2 Definizioni di densità continua e v.a. continua

Definizione 2.1. Una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ si dice **densità (continua) o funzione di densità di probabilità o PDF^a** se:

- $f(x) \geq 0$, per ogni $x \in \mathbb{R}$;
- $\int_{-\infty}^{+\infty} f(x) dx = 1$.

^aDall'inglese *probability density function*.

OSSERVAZIONE. Al contrario della densità discreta p_X , che deve necessariamente verificare le disuguaglianze

$$0 \leq p_X(x) \leq 1, \quad \forall x \in \mathbb{R},$$

la densità continua verifica in generale solo la prima disuguaglianza:

$$0 \leq f(x), \quad \forall x \in \mathbb{R}.$$

Al contrario, si può benissimo avere $f(x) > 1$ per qualche $x \in \mathbb{R}$. L'importante è che $\int_{-\infty}^{+\infty} f(x) dx = 1$.

La definizione di *variabile aleatoria continua* fa intervenire la densità continua.

Definizione 2.2. Sia (Ω, \mathbb{P}) uno spazio di probabilità e X una variabile aleatoria. Si dice che X è una **variabile aleatoria continua** (in breve **v.a.c.**) se esiste una densità continua, indicata con f_X , tale che

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx, \quad \forall [a, b] \subset \mathbb{R}$$

e, più in generale,

$$\mathbb{P}(X \in B) = \int_B f_X(x) dx, \quad \forall B \subset \mathbb{R}. \quad (2.1)$$

OSSERVAZIONE 1. Esempi di utilizzo della formula (2.1) sono i seguenti:

$$\begin{aligned} \mathbb{P}(a < X < b) &= \int_a^b f_X(x) dx, & (B = (a, b)), \\ \mathbb{P}(a \leq X < b) &= \int_a^b f_X(x) dx, & (B = [a, b)), \end{aligned}$$

$$\begin{aligned}
\mathbb{P}(a < X \leq b) &= \int_a^b f_X(x) dx, & (B = (a, b]), \\
\mathbb{P}(X < x) &= \int_{-\infty}^x f_X(x) dx, & (B = (-\infty, x)), \\
\mathbb{P}(X \leq x) &= \int_{-\infty}^x f_X(x) dx, & (B = (-\infty, x]), \\
\mathbb{P}(X > x) &= \int_x^{+\infty} f_X(x) dx, & (B = (x, +\infty)), \\
\mathbb{P}(X \geq x) &= \int_x^{+\infty} f_X(x) dx, & (B = [x, +\infty)).
\end{aligned}$$

Altri casi in cui utilizzeremo la formula (2.1) sono quelli in cui B è unione di più intervalli. Ad esempio, se $B = [-1, 0] \cup (2, 3] \cup [5, +\infty)$ allora

$$\mathbb{P}(X \in B) = \int_{-1}^0 f_X dx + \int_2^3 f_X dx + \int_5^{+\infty} f_X dx.$$

Notiamo in particolare che **la probabilità che una variabile aleatoria continua X assuma valori in un intervallo non dipende dal fatto che gli estremi dell'intervallo siano inclusi o esclusi**, contrariamente a quanto accade per le variabili aleatorie discrete.

OSSERVAZIONE 2. È importante sottolineare che, se X è una variabile aleatoria continua, la sua densità f_X **non è unica**. Infatti, se f_X è una densità di X , allora ogni funzione g per cui

$$\int_a^b g(x) dx = \int_a^b f_X(x) dx, \quad \forall [a, b] \subset \mathbb{R}, \quad (2.2)$$

è una densità di X . Ad esempio, se g è ottenuta da f_X modificandone il valore in un numero finito (o infinito numerabile) di punti, allora (2.2) vale e dunque g è una densità di X .

Questa ambiguità nella nozione di densità di una variabile aleatoria continua non costituisce di norma un problema. Come vedremo, in molti casi esiste una versione “canonica” della densità che è regolare, ad esempio continua. Spesso diremo, impropriamente, che una certa funzione f è **la** densità di X .

OSSERVAZIONE 3. Al contrario di quanto visto con le variabili aleatorie discrete, non definiremo il **supporto** di una variabile aleatoria continua. Infatti il supporto dovrebbe essere definito in termini della densità, ma la densità non è unica (si veda l'osservazione precedente). Più precisamente, la definizione di supporto dovrebbe essere la seguente:

$$\mathcal{S}_X = \{x \in \mathbb{R} : f_X(x) > 0\}. \quad (2.3)$$

Nonostante la non unicità di f_X , in molti casi esiste una versione “canonica” della densità (si veda l'osservazione precedente) e dunque esiste anche una versione “canonica” di supporto, definito come in (2.3) scegliendo come f_X la versione canonica. In questi casi diremo, impropriamente, che l'insieme \mathcal{S}_X è **il** supporto di X .

Dalla definizione di variabile aleatoria continua seguono le seguenti proprietà.

Teorema 2.1. Sia X una variabile aleatoria continua con densità f_X .

1) La densità discreta di X è identicamente uguale a zero:

$$p_X(x) = 0, \quad \forall x \in \mathbb{R}.$$

2) La funzione di ripartizione di X è data da

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad \forall x \in \mathbb{R}.$$

Quindi F_X è continua (e non solo continua a destra).

Dimostrazione.

1) Dalla definizione di p_X , abbiamo che

$$p_X(x) = \mathbb{P}(X = x) = \int_x^x f_X(y) dy = 0, \quad \forall x \in \mathbb{R},$$

che dimostra il punto 1).

2) Dalla definizione di F_X , si ha che

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(y) dy, \quad \forall x \in \mathbb{R}.$$

Resta da dimostrare che F_X è continua e non solo continua a destra. Quindi resta da dimostrare che

$$F_X(x) = F_X(x-), \quad \forall x \in \mathbb{R}.$$

Ricordando che $F_X(x-) = \mathbb{P}(X < x)$, si ha

$$F_X(x-) = \mathbb{P}(X < x) = \int_{-\infty}^x f_X(y) dy = F_X(x), \quad \forall x \in \mathbb{R}.$$

□

OSSERVAZIONE 1. Dal Teorema 2.1 segue che la funzione di ripartizione di una variabile aleatoria continua è una **funzione integrale**. Le funzioni integrali costituiscono una particolare sotto-famiglia delle funzioni continue. Esse sono anche dette **funzioni assolutamente continue**. Per tale ragione le variabili aleatorie continue sono anche dette **variabili aleatorie assolutamente continue**.

OSSERVAZIONE 2. Grazie alla continuità della funzione di ripartizione F_X , notiamo che le probabilità

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b),$$

che come abbiamo già visto sono tutte uguali, sono in particolare date da

$$F_X(b) - F_X(a) = \int_a^b f_X(x) dx.$$

Analogamente

$$\mathbb{P}(X < a) = \mathbb{P}(X \leq a) = \int_{-\infty}^a f_X(x) dx = F_X(a)$$

e

$$\mathbb{P}(X > a) = \mathbb{P}(X \geq a) = \int_a^{+\infty} f_X(x) dx = 1 - F_X(a).$$

Possiamo dunque riassumere nel seguente schema le differenze principali tra variabili aleatorie discrete e continue:

VARIABILI ALEATORIE DISCRETE	VARIABILI ALEATORIE CONTINUE
densità discreta p_X	densità continua f_X
$\mathbb{P}(X \in B) = \sum_{x_i \in B} p_X(x_i)$	$\mathbb{P}(X \in B) = \int_B f_X(x) dx$
F_X è costante a tratti: $F_X(x) = \sum_{x_i \leq x} p_X(x_i)$	F_X è una funzione integrale o, equivalentemente, è una funzione assolutamente continua: $F_X(x) = \int_{-\infty}^x f_X(y) dy$

Si noti infine che queste due classi di variabili aleatorie non esauriscono tutte le possibilità, infatti ci sono ad esempio le variabili aleatorie *miste*. Quest'ultime hanno come "supporto" un insieme infinito più che numerabile, ad esempio un intervallo $[a, b]$, ma all'interno del supporto esistono un numero finito (o al più infinito numerabile) di valori con probabilità strettamente positiva. Per descrivere la legge di una variabile aleatoria mista è necessario utilizzare sia la densità discreta che la densità continua. Un esempio di variabile aleatoria mista è il seguente.

Esempio 2.1. *Si consideri un componente elettronico. Sia*

$$X = \text{“tempo di vita del componente”}.$$

In tal caso può essere ragionevole supporre che^a

$$\mathbb{P}(X = 0) \in (0, 1)$$

mentre

$$\mathbb{P}(X = x) = 0, \quad \forall x \neq 0.$$

^aProbabilità che il componente sia rotto, a causa di un difetto di fabbricazione.

Si noti che:

- X non è una v.a. discreta, infatti

$$p_X(x) = \begin{cases} \mathbb{P}(X = 0), & x = 0, \\ 0, & x \neq 0. \end{cases}$$

Quindi non vale che $\sum_i p_X(x_i) = 1$, dato che $\sum_i p_X(x_i) = \mathbb{P}(X = 0) < 1$;

- X non è una v.a. continua, infatti $\mathbb{P}(X = 0) > 0$, in contraddizione con quanto affermato nel Teorema 2.1.

2.1 Dalla funzione di ripartizione alla densità continua

Nella sezione precedente abbiamo visto come si passa dalla densità continua alla funzione di ripartizione. In particolare, ricordiamo dal Teorema 2.1 che vale la formula seguente:

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad \forall x \in \mathbb{R}.$$

Supponiamo ora di conoscere la funzione di ripartizione F_X di una variabile aleatoria X .

- 1) Che proprietà¹ deve verificare F_X affinché X sia una variabile aleatoria *continua*?
- 2) Una volta stabilito che X è una variabile aleatoria continua, come si trova f_X a partire da F_X ?

Per quanto riguarda il punto 1), come abbiamo visto nella sezione precedente, si ha che

$$X \text{ è una v.a. } \textit{continua} \iff \begin{array}{l} F_X \text{ è una } \textit{funzione integrale} \\ \text{(o, equivalentemente, } \textit{assolutamente continua}) \end{array}$$

Nella pratica, non è semplice verificare² che F_X è una funzione integrale. Per tale ragione, nel seguito studieremo solo casi in cui sarà già noto che X è una v.a. *continua*. Per quanto riguarda invece il punto 2), vale il seguente risultato.

¹Ricordiamo che, per quanto riguarda le variabili aleatorie discrete, vale il seguente risultato:

$$X \text{ è una v.a. } \textit{discreta} \iff F_X \text{ è una } \textit{funzione costante a tratti}$$

²Segnaliamo però che una *condizione sufficiente* affinché F_X sia una funzione integrale è la seguente condizione di facile verifica: F_X è C^1 **a tratti**, cioè

- F_X è continua in ogni punto $x \in \mathbb{R}$;
- esiste un sottoinsieme *finito* $D \subset \mathbb{R}$ tale che F_X è derivabile in ogni punto $x \in \mathbb{R} \setminus D$;
- la derivata F'_X è una funzione continua in ogni punto $x \in \mathbb{R} \setminus D$.

Proposizione 2.1. Sia X una variabile aleatoria e indichiamo con F_X la sua funzione di ripartizione. Supponiamo di sapere già che X è una **variabile aleatoria continua** (quindi sappiamo già che F_X è una **funzione integrale**). Allora la sua densità f_X è data da

$$f_X(x) = F'_X(x), \quad \forall x \text{ in cui } F_X \text{ è derivabile.}$$

Nei punti in cui F_X non è derivabile, f_X è definita in modo arbitrario.

Esercizio 2.1. Sia X una variabile aleatoria continua^a con funzione di ripartizione

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ (1 - e^{-x})^2, & x \geq 0. \end{cases}$$

Determinare:

- (a) la densità di X ,
- (b) $\mathbb{P}(X > 1)$,
- (c) $\mathbb{P}(1 < X < 2)$.

^aSi noti che il testo dell'esercizio specifica già che X è *continua*, non serve dimostrarlo. Non serve quindi mostrare che F_X è una *funzione integrale*. Notiamo anche che in tutti i casi che considereremo F_X sarà sempre una funzione C^1 a tratti, quindi, automaticamente, una funzione integrale (si veda a tal proposito la *nota 2*).

Soluzione.

- (a) Poiché la variabile aleatoria X è continua, possiamo applicare la Proposizione 2.1, da cui si ha che

$$f_X(x) = F'_X(x), \quad \forall x \text{ in cui } F_X \text{ è derivabile,}$$

altrimenti f_X è definita in modo arbitrario dove F_X non è derivabile. Dall'espressione di F_X si vede che F_X è derivabile ovunque tranne al più nel punto $x = 0$. Quindi

$$f_X(x) = \begin{cases} 0, & x < 0, \\ 2(1 - e^{-x})e^{-x}, & x > 0. \end{cases}$$

Abbiamo dunque specificato f_X per ogni $x \in \mathbb{R} \setminus \{0\}$. Nel punto $x = 0$ non è necessario verificare se F_X è derivabile, infatti possiamo comunque definire f_X in modo arbitrario, ad esempio ponendo $f_X(0) = 0$. Ciò è una conseguenza dell'osservazione in cui compare l'uguaglianza (2.2). In tale osservazione si dice infatti che è possibile modificare il valore di f_X in un numero finito (o infinito numerabile) di punti. Assegnando dunque il valore arbitrario $f_X(0) = 0$, otteniamo

$$f_X(x) = \begin{cases} 0, & x \leq 0, \\ 2(1 - e^{-x})e^{-x}, & x > 0. \end{cases}$$

(b)

$$\mathbb{P}(X > 1) = 1 - F_X(1) = 1 - (1 - e^{-1})^2 \simeq 0.6.$$

(c)

$$\mathbb{P}(1 < X < 2) = F_X(2) - F_X(1) = (1 - e^{-2})^2 - (1 - e^{-1})^2 \simeq 0.348.$$

□

Esercizio 2.2. Sia X una variabile aleatoria continua con funzione di ripartizione

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ x, & 0 \leq x \leq \frac{1}{4}, \\ (x - \frac{1}{4})^2 + \frac{1}{4}, & \frac{1}{4} \leq x \leq 1, \\ \frac{3}{16}(1 - e^{-(x-1)}) + \frac{13}{16}, & x \geq 1. \end{cases}$$

Determinare la densità di X .

Soluzione. Per la Proposizione 2.1 si ha che

$$f_X(x) = \begin{cases} 0, & x < 0, \\ 1, & 0 < x < \frac{1}{4}, \\ 2(x - \frac{1}{4}), & \frac{1}{4} < x < 1, \\ \frac{3}{16}e^{-(x-1)}, & x > 1. \end{cases}$$

Abbiamo dunque specificato f_X per ogni $x \in \mathbb{R} \setminus \{0, \frac{1}{4}, 1\}$. Nei punti $x = 0$, $x = \frac{1}{4}$ e $x = 1$ possiamo invece definire f_X in modo arbitrario, ad esempio ponendo:

$$f_X(x) = \begin{cases} 0, & x \leq 0, \\ 1, & 0 < x \leq \frac{1}{4}, \\ 2(x - \frac{1}{4}), & \frac{1}{4} < x \leq 1, \\ \frac{3}{16}e^{-(x-1)}, & x > 1. \end{cases}$$

□

2.2 Funzioni di variabili aleatorie continue

Siano $h: \mathbb{R} \rightarrow \mathbb{R}$ una qualunque funzione e X una variabile aleatoria continua. Poniamo

$$Y = h(X).$$

Ricordiamo che quando X è discreta, Y è necessariamente anch'essa una variabile aleatoria discreta. Al contrario, quando X è continua, non possiamo dire nulla su Y . In particolare, Y potrebbe essere discreta, continua, mista.

La situazione più semplice si ha quando la variabile aleatoria Y è discreta. Si è in tale situazione quando Y assume un numero finito o al più infinito numerabile di valori. Ad esempio, se $h(x) = 1_{\{x>10\}}$ allora

$$Y = 1_{\{X>10\}} = \begin{cases} 1, & \text{se } X > 10, \\ 0, & \text{se } X \leq 10. \end{cases}$$

In tal caso Y è discreta in quanto assume solo due valori: 0 e 1. In particolare, Y ha distribuzione di Bernoulli di parametro $p = \mathbb{P}(X > 10)$:

$$Y \sim B(p).$$

Supponiamo invece che Y non assuma un numero finito o al più numerabile di valori. Supponiamo inoltre di sapere già³ che Y è una variabile aleatoria *continua*. Come si trovano funzione di ripartizione e/o densità di Y ?

Per risolvere questo problema, si procede determinando innanzitutto la *funzione di ripartizione* di Y . Nel caso in cui siamo interessati alla *densità* di Y , la otteniamo successivamente derivando la funzione di ripartizione F_Y , applicando dunque la Proposizione 2.1.

Per trovare la funzione di ripartizione di Y , i primi passaggi che si fanno sono sempre i seguenti:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(h(X) \leq y).$$

Si cerca dunque di esprimere $\mathbb{P}(h(X) \leq y)$ in termini della funzione di ripartizione di X . Per fare questo è necessario risolvere la disuguaglianza $h(X) \leq y$. Vediamo un esempio.

Esempio 2.2. Sia X una variabile aleatoria continua con funzione di ripartizione

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-x}, & x \geq 0. \end{cases}$$

Qual è la densità della variabile aleatoria continua^a $Y = e^X$?

^aSi noti che il testo dell'esercizio specifica già che Y è *continua*, non serve dimostrarlo.

Soluzione. Iniziamo col determinare la funzione di ripartizione di Y :

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(e^X \leq y).$$

Dobbiamo risolvere la disuguaglianza $e^X \leq y$. Distinguiamo due casi: $y \leq 0$ e $y > 0$.

Primo caso: $y \leq 0$. Se $y \leq 0$, la disuguaglianza $e^X \leq y$ non è mai verificata. Questo significa che

$$\{e^X \leq y\} = \emptyset.$$

³Se invece non sapessimo ancora se Y è *continua*, per stabilirlo dovremmo prima determinare F_Y e poi mostrare che F_Y è una *funzione integrale*, ad esempio mostrando che è C^1 a tratti.

Quindi $\mathbb{P}(e^X \leq y) = \mathbb{P}(\emptyset) = 0$. Perciò

$$F_Y(y) = 0 \quad \forall y \leq 0.$$

Secondo caso: $y > 0$. Se $y > 0$, allora la disuguaglianza $e^X \leq y$ è verificata se e solo se $X \leq \log y$, cioè

$$e^X \leq y \quad \iff \quad X \leq \log y.$$

Quindi

$$F_Y(y) = \mathbb{P}(X \leq \log y) = \mathbb{P}(X \leq \log y) = F_X(\log y)$$

Utilizzando l'espressione di F_X , si ottiene dunque:

- se $\log y \leq 0$, cioè $0 < y \leq 1$, allora $F_Y(y) = F_X(\log y) = 0$;
- se $\log y > 0$ (cioè $y > 1$), allora $F_Y(y) = F_X(\log y) = 1 - e^{-\log y} = 1 - \frac{1}{y}$.

In conclusione, mettendo insieme primo e secondo caso, otteniamo

$$F_Y(y) = \begin{cases} 0, & y \leq 1, \\ 1 - \frac{1}{y}, & y > 1. \end{cases}$$

Applicando la Proposizione 2.1, si ricava la densità di Y derivando F_Y :

$$f_Y(y) = \begin{cases} 0, & y \leq 1, \\ \frac{1}{y^2}, & y > 1, \end{cases}$$

dove abbiamo posto, in modo arbitrario, $f_Y(1) = 0$. □

3 *Indici di sintesi di una distribuzione: μ e σ^2*

Come per le variabili aleatorie discrete, definiamo valore atteso e varianza per variabili aleatorie continue.

3.1 *Media o valore atteso*

Definizione 3.1. *Sia X una variabile aleatoria continua. La **media** (o **valore atteso**) di X è data da*

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

La media si indica anche con μ oppure μ_X .

Nel seguito capiterà spesso di dover calcolare il valore atteso di una funzione di X :

$$Y = h(X).$$

Risulta dunque particolarmente utile il seguente risultato.

Teorema 3.1. *Sia X una variabile aleatoria continua. Inoltre, siano $h: \mathbb{R} \rightarrow \mathbb{R}$ e $Y = h(X)$. Allora*

$$\mathbb{E}[Y] = \mathbb{E}[h(X)] = \int_{-\infty}^{+\infty} h(x) f_X(x) dx.$$

Ricordiamo infine la proprietà di linearità del valore atteso, già dimostrata nel caso discreto (la dimostrazione segue dalla formula del Teorema 3.1).

Teorema 3.2 (Linearità del valore atteso). *Sia X una variabile aleatoria continua. Inoltre, siano $h: \mathbb{R} \rightarrow \mathbb{R}$, $g: \mathbb{R} \rightarrow \mathbb{R}$ e $a, b \in \mathbb{R}$ costanti. Allora*

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

e, più in generale,

$$\mathbb{E}[ah(X) + bg(X)] = a\mathbb{E}[h(X)] + b\mathbb{E}[g(X)].$$

3.2 Varianza

Definizione 3.2. *Sia X una variabile aleatoria continua. La **varianza** di X è data da*

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f_X(x) dx.$$

La varianza si indica anche con σ^2 oppure σ_X^2 .

La radice quadrata della varianza si chiama **deviazione standard** (o **scarto quadratico medio** o anche **scostamento quadratico medio**) e si indica con σ oppure σ_X .

Per calcolare la varianza di una variabile aleatoria è utile, come nel caso discreto, la seguente formula.

Teorema 3.3. *Sia X una variabile aleatoria continua. Vale che*

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \mathbb{E}[X]^2.$$

Ricordiamo infine le seguenti proprietà, già dimostrate nel caso discreto.

Teorema 3.4 (Proprietà della varianza). Siano X una variabile aleatoria continua e $a, b \in \mathbb{R}$ costanti. Allora

1) $\text{Var}(X) \geq 0$.

2) $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

4 Distribuzioni continue notevoli

In questa sezione vediamo le principali distribuzioni continue.

Distribuzione uniforme (continua). Diciamo che X ha **distribuzione uniforme (continua)** su⁴ (a, b) se X è una variabile aleatoria continua con densità

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{altrimenti.} \end{cases}$$

In tal caso scriviamo

$$X \sim \text{Unif}(a, b).$$

Si noti che

$$F_X(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x \geq b. \end{cases}$$

Inoltre

$$\begin{aligned} \mathbb{E}[X] &= \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}, \\ \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_a^b \frac{x^2}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}. \end{aligned}$$

Distribuzione esponenziale. Sia $\lambda > 0$. Diciamo che X ha **distribuzione esponenziale di parametro** λ se X è una variabile aleatoria continua con densità

$$f_X(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0. \end{cases}$$

In tal caso scriviamo

$$X \sim \text{Exp}(\lambda).$$

⁴Essendo X una variabile aleatoria continua, non fa alcuna differenza includere o escludere gli estremi dell'intervallo. Più precisamente, sono identiche le distribuzioni (e dunque le v.a.) uniformi su (a, b) oppure $[a, b)$ oppure $(a, b]$ oppure $[a, b]$. Quindi scriveremo semplicemente $X \sim \text{Unif}(a, b)$.

Si noti che

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Inoltre

$$\begin{aligned} \mathbb{E}[X] &= \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}, \\ \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$

La distribuzione esponenziale si usa ad esempio per descrivere il tempo di vita di un macchinario oppure di un componente elettronico (si veda l'Esempio 1.1).

Distribuzione normale (o gaussiana). Siano $\mu \in \mathbb{R}$ e $\sigma > 0$. Diciamo che X ha **distribuzione normale (o gaussiana) di media μ e varianza σ^2** se X è una variabile aleatoria continua con densità

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad \text{per ogni } x \in \mathbb{R}.$$

In tal caso scriviamo

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Diciamo inoltre che X ha **distribuzione normale standard** se

$$X \sim \mathcal{N}(0, 1),$$

ovvero X ha distribuzione normale di media $\mu = 0$ e varianza $\sigma^2 = 1$. In tal caso, la densità di X è data da

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \text{per ogni } x \in \mathbb{R}.$$

La distribuzione normale ha un ruolo fondamentale in Probabilità e Statistica, come conseguenza del *Teorema centrale del limite* che vedremo in seguito.

NOTAZIONE. Una variabile aleatoria con distribuzione normale standard viene generalmente indicata con la lettera Z anziché X . Inoltre densità e funzione di ripartizione di Z si indicano con φ e Φ anziché f_Z e F_Z .

OSSERVAZIONE. Si noti che non esiste un'espressione esplicita della funzione di ripartizione della distribuzione normale. Ad esempio, nel caso della distribuzione normale standard, Φ è data da

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy, \quad \text{per ogni } x \in \mathbb{R}.$$

Tale integrale non ammette un'espressione esplicita in termini di funzioni note. Di conseguenza, i valori di Φ vengono calcolati per via numerica, approssimando il valore dell'integrale.

Anche se non è possibile calcolare esplicitamente l'integrale $\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$ per un generico valore x , è possibile farlo in alcuni casi particolari. Ad esempio, quando $x = +\infty$, come affermato nel seguente lemma.

Lemma 4.1 (Integrale di Gauss). *Vale che*

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}. \quad (4.1)$$

OSSERVAZIONE. *Come conseguenza del Lemma 4.1 segue che la funzione*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad \text{per ogni } x \in \mathbb{R},$$

*è effettivamente una densità, ovvero*⁵

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

Infatti, con un cambio di variabili, si ha che

$$\int_{-\infty}^{+\infty} f_X(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \stackrel{\substack{\uparrow \\ z=\frac{x-\mu}{\sigma\sqrt{2}}}}{=} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-z^2} dz = 1.$$

Dimostrazione del Lemma 4.1. Indichiamo con I il valore dell'integrale di Gauss, quindi

$$I = \int_{-\infty}^{+\infty} e^{-x^2} dx.$$

Poiché la funzione $x \mapsto e^{-x^2}$ è pari, si ha che

$$I = 2 \int_0^{+\infty} e^{-x^2} dx.$$

Dimostrare che $I = \sqrt{\pi}$ è equivalente a mostrare che $I^2 = \pi$. Dimostriamo dunque che $I^2 = \pi$. Si ha che

$$I^2 = 4 \left(\int_0^{+\infty} e^{-x^2} dx \right) \left(\int_0^{+\infty} e^{-x^2} dx \right) = 4 \left(\int_0^{+\infty} e^{-x^2} dx \right) \left(\int_0^{+\infty} e^{-y^2} dy \right),$$

dove la seconda uguaglianza discende dal fatto che, essendo la variabile d'integrazione *muta*, possiamo indicarla con un'altra lettera, ad esempio y , anziché x . Il prodotto dei due ultimi integrali è in effetti uguale all'integrale doppio

$$\left(\int_0^{+\infty} e^{-x^2} dx \right) \left(\int_0^{+\infty} e^{-y^2} dy \right) = \int_0^{+\infty} \int_0^{+\infty} e^{-(x^2+y^2)} dx dy.$$

⁵È chiaro che $f_X(x) \geq 0$ per ogni $x \in \mathbb{R}$.

Quindi

$$I^2 = 4 \int_0^{+\infty} \int_0^{+\infty} e^{-(x^2+y^2)} dx dy.$$

Integrando prima rispetto a y e poi rispetto a x , possiamo riscrivere I^2 come segue:

$$I^2 = 4 \int_0^{+\infty} \left(\int_0^{+\infty} e^{-(x^2+y^2)} dy \right) dx.$$

Nell'integrale interno (in cui x compare come un parametro fissato maggiore di zero),

$$\int_0^{+\infty} e^{-(x^2+y^2)} dy,$$

eseguiamo il cambio di variabile $y = xz$, con $x > 0$ fissato. Quindi $dy = x dz$, perciò

$$\int_0^{+\infty} e^{-(x^2+y^2)} dy = \int_0^{+\infty} e^{-x^2(1+z^2)} x dz.$$

Di conseguenza, I^2 diventa

$$I^2 = 4 \int_0^{+\infty} \left(\int_0^{+\infty} e^{-x^2(1+z^2)} x dz \right) dx.$$

Scambiando l'ordine di integrazione

$$I^2 = 4 \int_0^{+\infty} \left(\int_0^{+\infty} e^{-x^2(1+z^2)} x dx \right) dz.$$

Si noti che ora è possibile calcolare esplicitamente l'integrale interno, infatti

$$\int_0^{+\infty} e^{-x^2(1+z^2)} x dx = -\frac{1}{2(1+z^2)} [e^{-x^2(1+z^2)}]_0^{+\infty} = -\frac{1}{2(1+z^2)} [0 - 1] = \frac{1}{2(1+z^2)}.$$

In conclusione, si ottiene

$$I^2 = 4 \int_0^{+\infty} \frac{1}{2(1+z^2)} dz = 2 [\arctan z]_0^{+\infty} = \pi.$$

□

Proposizione 4.1. Siano $\mu \in \mathbb{R}$, $\sigma > 0$ e $X \in \mathcal{N}(\mu, \sigma^2)$.

- 1) $\mathbb{E}[X] = \mu$.
- 2) $\text{Var}(X) = \sigma^2$.

Dimostrazione.

1) Dalla definizione di valore atteso, abbiamo che

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \\ &= \int_{-\infty}^{+\infty} (\mu + z\sigma\sqrt{2}) \frac{1}{\sqrt{\pi}} e^{-z^2} dz \\ &\quad \uparrow \\ &\quad z = \frac{x-\mu}{\sigma\sqrt{2}} \\ &= \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-z^2} dz + \sigma\sqrt{2} \int_{-\infty}^{+\infty} z \frac{1}{\sqrt{\pi}} e^{-z^2} dz \\ &= \mu + \sigma\sqrt{2} \int_{-\infty}^{+\infty} z \frac{1}{\sqrt{\pi}} e^{-z^2} dz, \\ &\quad \uparrow \\ &\quad \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-z^2} dz = 1\end{aligned}$$

dove nell'ultima uguaglianza abbiamo usato (4.1). Infine, essendo $z \mapsto z \frac{1}{\sqrt{\pi}} e^{-z^2}$ una funzione dispari, si ha che

$$\int_{-\infty}^{+\infty} z \frac{1}{\sqrt{\pi}} e^{-z^2} dz = 0.$$

Quindi $\mathbb{E}[X] = \mu$.

2) Dalla definizione di varianza, abbiamo che

$$\begin{aligned}\text{Var}(X) &= \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx = \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \\ &= \sigma^2 \int_{-\infty}^{+\infty} z^2 \frac{2}{\sqrt{\pi}} e^{-z^2} dz. \\ &\quad \uparrow \\ &\quad z = \frac{x-\mu}{\sigma\sqrt{2}}\end{aligned}$$

Integrando per parti, si ottiene

$$\text{Var}(X) = \sigma^2 \left\{ \left[-z \frac{1}{\sqrt{\pi}} e^{-z^2} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-z^2} dz \right\} = \sigma^2 \{ [0 - 0] + 1 \} = \sigma^2,$$

dove nella seconda uguaglianza abbiamo usato (4.1).

□

La distribuzione normale standard ha un ruolo fondamentale nello studio della distribuzione normale. Ciò deriva dal fatto che qualunque variabile aleatoria normale diventa, eseguendo un riscalamento e una traslazione (in una parola, *standardizzando*), una variabile aleatoria normale standard.

Proposizione 4.2 (Standardizzazione). *Siano $\mu \in \mathbb{R}$, $\sigma > 0$ e $X \in \mathcal{N}(\mu, \sigma^2)$. Allora*

$$Z = \frac{X - \mu}{\sigma}$$

è una variabile aleatoria normale standard, quindi $Z \sim \mathcal{N}(0, 1)$.

Dimostrazione. Mostriamo che Z ammette densità continua data da (ricordiamo che generalmente la densità della variabile aleatoria normale standard si indica con φ anziché f_Z)

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \forall x \in \mathbb{R}.$$

Iniziamo col determinare la funzione di ripartizione di Z :

$$F_Z(x) = \mathbb{P}(Z \leq x) = \mathbb{P}\left(\frac{X - \mu}{\sigma} \leq x\right) = \mathbb{P}(X \leq \mu + \sigma x) = F_X(\mu + \sigma x).$$

Derivando, si ottiene

$$f_Z(x) = \sigma f_X(\mu + \sigma x).$$

Sapendo che $X \sim \mathcal{N}(\mu, \sigma^2)$, si ha che

$$\sigma f_X(\mu + \sigma x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(\mu + \sigma x - \mu)^2}{\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Quindi abbiamo dimostrato che Z ha densità φ , ovvero $Z \sim \mathcal{N}(0, 1)$. □

La funzione di ripartizione di una variabile aleatoria normale standard Z , indicata generalmente con Φ anziché F_Z , è data per definizione da

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy, \quad \forall x \in \mathbb{R}.$$

Come abbiamo già notato, non esiste una formula esplicita per Φ . Tuttavia Φ possiede le seguenti utili proprietà.

Proposizione 4.3 (Proprietà di Φ). *Sia*

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy, \quad \forall x \in \mathbb{R}$$

la funzione di ripartizione di una variabile aleatoria normale standard. Φ verifica le seguenti proprietà:

- 1) $\Phi(0) = \frac{1}{2}$.
- 2) $\Phi(-x) = 1 - \Phi(x)$, per ogni $x \geq 0$.

OSSERVAZIONE. *La proprietà 2) è particolarmente utile in quanto permette di calcolare Φ per valori negativi una volta nota Φ per valori positivi.*

Dimostrazione della Proposizione 4.3.

1) Si noti che la proprietà 1) segue dalla 2) scegliendo $x = 0$. Infatti si ottiene

$$\Phi(0) = 1 - \Phi(0),$$

che ha come soluzione $\Phi(0) = \frac{1}{2}$. Resta dunque da dimostrare la proprietà 2).

2) Dato che $y \mapsto \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}$ è una funzione pari, si ha che

$$\int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2} dy = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2} dy, \quad \forall x > 0.$$

Questo fornisce l'uguaglianza voluta, infatti il primo integrale è $\Phi(-x)$, mentre il secondo integrale è pari a $1 - \Phi(x)$. \square

Esercizio 4.1. *Un apparecchio dosatore riempie delle provette da 10 cl. Assumiamo che la quantità di liquido versata in una provetta (misurata in cl), indicata con X , abbia una distribuzione $\mathcal{N}(9.99, (0.012)^2)$, ovvero X è una variabile aleatoria continua con densità*

$$f_X(x) = \frac{1}{0.012\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-9.99)^2}{(0.012)^2}}, \quad \forall x \in \mathbb{R}.$$

- (a) *Trovare la percentuale di provette fatte traboccare dal dosatore. [Si esprima il risultato nella forma^a $1 - \Phi(x)$, per qualche $x > 0$]*
- (b) *Determinare ℓ in modo tale che la percentuale di provette che contengono una quantità di liquido inferiore a ℓ sia pari al 10% delle provette. [Si usi^b che $\Phi^{-1}(0.1) \simeq -1.282$, dove Φ^{-1} denota la funzione inversa di Φ]*

^aUn'approssimazione di $\Phi(x)$ può essere ottenuta utilizzando WolframAlpha, si veda www.wolframalpha.com, tramite il comando `CDF[NormalDistribution[0,1],x]`.

^bIl valore di $\Phi^{-1}(0.1)$ è stato ottenuto con WolframAlpha tramite il comando `InverseCDF[NormalDistribution[0,1],0.1]`.

Soluzione.

(a) Si noti che l'evento "il dosatore fa traboccare la provetta" è dato da

$$\{X > 10\}.$$

Si noti inoltre che la v.a.

$$Z = \frac{X - 9.99}{0.012}$$

ha legge normale standard. Quindi

$$\begin{aligned} \mathbb{P}(X > 10) & \stackrel{\substack{\uparrow \\ \text{standardizzazione}}}{=} \mathbb{P}\left(\frac{X - 9.99}{0.012} > \frac{10 - 9.99}{0.012}\right) \\ & \simeq \mathbb{P}(Z > 0.833) = 1 - \mathbb{P}(Z \leq 0.833) = 1 - \Phi(0.833) \simeq 20.24\%. \end{aligned}$$

(b) Dobbiamo trovare ℓ tale che

$$\mathbb{P}(X < \ell) = 10\%.$$

Standardizzando, possiamo riscrivere questa uguaglianza in termini di Z :

$$\mathbb{P}\left(Z < \frac{\ell - 9.99}{0.012}\right) = 0.1.$$

Quest'ultima uguaglianza può essere riscritta in termini di Φ , la funzione di ripartizione di Z :

$$\Phi\left(\frac{\ell - 9.99}{0.012}\right) = \mathbb{P}\left(Z \leq \frac{\ell - 9.99}{0.012}\right) = \mathbb{P}\left(Z < \frac{\ell - 9.99}{0.012}\right) = 0.1,$$

dove la seconda uguaglianza deriva dal fatto che Z è una variabile aleatoria continua. Quindi

$$\frac{\ell - 9.99}{0.012} = \Phi^{-1}(0.1) \simeq -1.282,$$

da cui si ottiene $\ell \simeq 9.9746$.

□

5 Generatori aleatori*

In diverse aree della matematica applicata, ad esempio nell'ambito della simulazione numerica o in crittografia, si richiede a un computer di produrre una sequenza di numeri casuali con distribuzione assegnata. La maggior parte dei software di calcolo scientifico (come ad esempio MATLAB oppure R) possiede comandi (o, meglio, *generatori aleatori*, anche detti *generatori di numeri casuali*) che forniscono tali sequenze di numeri casuali, almeno per le distribuzioni più comuni. In C, ad esempio, si trova un generatore aleatorio che fornisce una sequenza di numeri casuali con distribuzione uniforme su $[0, \text{RAND_MAX}]$, dove RAND_MAX è una costante che, secondo gli standard di C, deve essere maggiore o uguale di 32767. In questa sezione affronteremo le due seguenti questioni.

- 1) Come si genera una variabile aleatoria con distribuzione uniforme?
- 2) Come si genera una variabile aleatoria con distribuzione non necessariamente uniforme?

5.1 Simulare la distribuzione uniforme

Possiamo suddividere i generatori aleatori di numeri casuali con distribuzione uniforme in due categorie.

- **Generatori fisici.** I generatori fisici non sempre generano direttamente un numero con distribuzione uniforme. Tale numero va dunque trasformato opportunamente (procedendo in modo simile a come si fa per affrontare il problema 2 riportato sopra) per ottenere un numero con distribuzione uniforme. Un esempio famoso di generatore fisico era stato progettato dalla RAND Corporation alla fine degli anni '40 del Novecento. Il risultato di questo esperimento è riportato nel libro "*A Million Random Digits with 100000 Normal Deviates*", pubblicato dalla RAND Corporation nel 1955. Tale libro contiene appunto un milione di numeri generati casualmente da una simulazione elettronica di una roulette con numeri da 00000 a 99999.

- **Generatori algoritmici.** Un generatore algoritmico consiste in un algoritmo (completamente deterministico) in grado di produrre lunghe sequenze di numeri apparentemente casuali. Tali numeri si chiamano *pseudo-casuali*. Fornendo all'algoritmo lo stesso valore iniziale (*seme*) si ottiene dunque la medesima sequenza. Tra i generatori algoritmici più conosciuti troviamo i **generatori lineari congruenziali (LCG)**. In tal caso, la sequenza di numeri pseudo-casuali è in generale data da un'espressione del tipo seguente:

$$x_n = (a x_{n-1} + c) \pmod{m},$$

dove a , c , m sono dei parametri da fissare. Dunque x_n è un numero intero compreso tra 0 ed $m - 1$. Ad esempio, una nota combinazione di parametri è $a = 7^5 = 16807$, $c = 0$, $m = 2^{31} - 1$.

5.2 Simulare una distribuzione non necessariamente uniforme

Consideriamo una distribuzione non necessariamente uniforme, di cui si conosce la funzione di ripartizione F . Supponiamo che $U \sim \text{Unif}(0, 1)$. Come si genera un numero aleatorio X con funzione di ripartizione F a partire da U ?

Per rispondere a questa domanda, conviene studiare separatamente il caso in cui X sia una v.a. *continua* oppure *discreta*. Prima però ricordiamo che $U \sim \text{Unif}(0, 1)$ significa che U è una variabile aleatoria continua con densità

$$f_U(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{altrimenti} \end{cases}$$

e funzione di ripartizione

$$F_U(x) = \begin{cases} 0, & x \leq 0, \\ x, & 0 \leq x \leq 1, \\ 1, & x \geq 1. \end{cases}$$

Distribuzione continua (ed F invertibile). Supponiamo che F sia la funzione di ripartizione di una v.a. *continua* e inoltre che F sia *invertibile*⁶. Si noti che F è invertibile se e solo se F è *strettamente crescente*. Dato che F è invertibile, esiste la sua inversa $F^{-1}: (0, 1) \rightarrow \mathbb{R}$. Poniamo

$$X = F^{-1}(U).$$

Allora X è la variabile aleatoria che cerchiamo, ossia X è una v.a. *continua* con funzione di ripartizione data proprio da F :

$$X \sim F.$$

Per dimostrarlo, notiamo innanzitutto che vale la seguente proprietà:

$$u \leq F(x) \iff F^{-1}(u) \leq x. \quad (5.1)$$

⁶ F è invertibile ad esempio quando X è una variabile aleatoria *normale*.

Allora

$$\begin{aligned}
 F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) \\
 &\stackrel{(5.1)}{=} \mathbb{P}(U \leq F(x)) = F_U(F(x)) \stackrel{0 \leq F(x) \leq 1}{=} F(x).
 \end{aligned}$$

Distribuzione discreta. Ricordiamo che F è la funzione di ripartizione di una v.a. *discreta* se e solo se F è *costante a tratti*. Quindi, a differenza del caso precedente, F non è strettamente crescente, dunque non è invertibile. Nonostante ciò, in tal caso ci sono vari modi per definire X a partire da U . Generalmente si definisce X tramite la cosiddetta *inversa generalizzata* di F . Vediamo con un esempio come si procede.

Esempio 5.1. Si consideri la funzione di ripartizione

$$F(x) = \begin{cases} 0, & x < 2, \\ 1/5, & 2 \leq x < 3, \\ 1/2, & 3 \leq x < 5, \\ 1, & x \geq 5. \end{cases}$$

Trovare una funzione $G: (0, 1) \rightarrow \mathbb{R}$ tale che la variabile aleatoria

$$X = G(U)$$

ha funzione di ripartizione data proprio da F .

Soluzione. Una possibile scelta di G è la seguente:

$$G(u) = \begin{cases} 2, & 0 < u \leq \frac{1}{5}, \\ 3, & \frac{1}{5} < u \leq \frac{1}{2}, \\ 5, & \frac{1}{2} < u < 1. \end{cases}$$

Infatti, sia

$$X = G(U).$$

Allora $\mathcal{S}_X = \{2, 3, 5\}$ e

X	2	3	5
p_X	$\mathbb{P}(G(U) = 2) = \frac{1}{5}$	$\mathbb{P}(G(U) = 3) = \frac{3}{10}$	$\mathbb{P}(G(U) = 5) = \frac{1}{2}$

Dunque F_X coincide con la funzione F . □

In generale, sia p_X la densità discreta associata alla funzione di ripartizione F . Allora, F è data da

$$F(x) = \begin{cases} 0, & x < x_1, \\ p_X(x_1), & x_1 \leq x < x_2, \\ p_X(x_1) + p_X(x_2), & x_2 \leq x < x_3, \\ p_X(x_1) + p_X(x_2) + p_X(x_3), & x_3 \leq x < x_4, \\ \dots & \\ p_X(x_1) + \dots + p_X(x_{n-1}), & x_{n-1} \leq x < x_n, \\ 1, & x \geq x_n. \end{cases}$$

In tal caso, una possibile scelta per la funzione $G: (0, 1) \rightarrow \mathbb{R}$ è la seguente:

$$G(u) = \begin{cases} x_1, & 0 < u \leq p_X(x_1), \\ x_2, & p_X(x_1) < u \leq p_X(x_2), \\ x_3, & p_X(x_2) < u \leq p_X(x_3), \\ x_4, & p_X(x_3) < u \leq p_X(x_4), \\ \dots & \\ x_{n-1}, & p_X(x_{n-1}) < u \leq p_X(x_n), \\ x_n, & p_X(x_n) < u < 1. \end{cases}$$