

**Appunti del corso**  
**Laboratorio di Calcolo 1**

Probabilità e simulazione

A.A. 2001/2002

LUCIA CARAMELLINO

[http://www.mat.uniroma2.it/~caramell/LC1\\_0102/LC1.htm](http://www.mat.uniroma2.it/~caramell/LC1_0102/LC1.htm)

[ultima versione: 27/09/2002]



# Indice

<b>1</b>	<b>Richiami di Probabilità</b>	<b>1</b>
1.1	Variabili aleatorie discrete . . . . .	1
1.2	Media, varianza e covarianza . . . . .	5
1.3	La Legge dei Grandi Numeri . . . . .	10
1.4	Variabili aleatorie assolutamente continue . . . . .	14
1.5	Il Teorema del Limite Centrale . . . . .	21
1.6	Appendice: retta di regressione ed interpretazione della co- varianza . . . . .	28
1.7	Appendice: complementi sulla legge gaussiana. . . . .	30
<b>2</b>	<b>Simulazione di variabili aleatorie discrete</b>	<b>31</b>
2.1	Schema di Bernoulli e conseguenze . . . . .	32
2.2	Simulazione di variabili aleatorie finite . . . . .	43
2.3	Metodo Monte Carlo per il calcolo numerico di integrali ed aree	46
2.4	Lista degli esercizi/problemi relativi al Capitolo 2 . . . . .	58
<b>3</b>	<b>Simulazione di variabili aleatorie continue</b>	<b>59</b>
3.1	Simulazione di esponenziali e conseguenze . . . . .	60
3.2	Simulazione di variabili aleatorie di Cauchy e gaussiane . . .	65
3.3	Simulazione di una linea tramviaria . . . . .	69
3.4	Appendice: complementi sulla legge esponenziale e la legge di Weibull . . . . .	75
3.5	Lista degli esercizi/problemi relativi al Capitolo 3 . . . . .	77



## Ringraziamenti

La prima stesura di questi appunti non era esente da errori di vario tipo; se ora cominciano ragionevolmente a diminuire, gran parte del merito spetta ad una studentessa scrupolosa e paziente. Ringrazio quindi Sara Belocchi per la preziosa collaborazione.



# Capitolo 1

## Richiami di Probabilità

In questo Capitolo vengono richiamati alcuni risultati principali di Calcolo delle Probabilità di cui si farà uso in seguito. Si tratta di una breve introduzione piuttosto discorsiva, per la formalizzazione e le dimostrazioni si rimanda al libro di testo di P. Baldi [1].

### 1.1 Variabili aleatorie discrete

Sia dato uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbb{P})$ , dove:

- $\Omega$  è un generico insieme, detto **spazio campionario**;
- $\mathcal{F}$  è una  $\sigma$ -algebra<sup>1</sup> di  $\Omega$  e ogni elemento di  $\mathcal{F}$  è detto un **evento**;
- $\mathbb{P}$  è una **misura di probabilità** su  $(\Omega, \mathcal{F})$ , cioè  $\mathbb{P}$  è un'applicazione  $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}_+$  tale che:
  - $\mathbb{P}(\Omega) = 1$ ;
  - se  $\{A_i\}_{i \in I} \subset \mathcal{F}$ , con  $I$  insieme finito o numerabile, e  $A_i \cap A_j = \emptyset$  per ogni  $i, j \in I$ , con  $i \neq j$ , allora

$$\mathbb{P}(\cup_{i \in I} A_i) = \sum_{i \in I} \mathbb{P}(A_i).$$

---

<sup>1</sup>Ricordiamo che una  $\sigma$ -algebra su  $\Omega$  è una classe di sottoinsiemi di  $\Omega$  contenente  $\Omega$  e chiusa sotto operazione di complementare e di unione o intersezione numerabile, cioè  $\mathcal{F} \subset \mathcal{P}(\Omega)$  (=insieme delle parti di  $\Omega$ ) e

- $\Omega \in \mathcal{F}$ ;
- se  $A \in \mathcal{F}$  allora  $A^c = \Omega \setminus A \in \mathcal{F}$ ;
- se  $\{A_i\}_{i \in I} \subset \mathcal{F}$ , con  $I$  insieme finito o numerabile, allora  $\cup_{i \in I} A_i \in \mathcal{F}$ .

[**Esercizio:** dimostrare che se  $\mathcal{F}$  è una  $\sigma$ -algebra allora: se  $\{A_i\}_{i \in I} \subset \mathcal{F}$ , con  $I$  insieme finito o numerabile, allora  $\cap_{i \in I} A_i \in \mathcal{F}$ .]

**Definizione 1.1.1.** Una **variabile aleatoria** (v.a.)  $X$  su  $(\Omega, \mathcal{F}, \mathbb{P})$  è una funzione

$$X : \Omega \longrightarrow \mathbb{R}$$

tale che per ogni  $x \in \mathbb{R}$ ,

$$\{X < x\} = \{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}(-\infty, x] \in \mathcal{F}.$$

Quindi una v.a.  $X$  è semplicemente una funzione su  $\Omega$ , a valori in  $\mathbb{R}$ , tale che gli insiemi di livello  $X^{-1}(-\infty, x]$  sono eventi.

Una v.a.  $X$  è detta **discreta** se l'insieme dei valori che assume  $E_X = X(\Omega) \equiv \{x \in \mathbb{R} : x = X(\omega), \text{ per qualche } \omega \in \Omega\}$  è un insieme discreto, cioè finito o numerabile. In particolare, diremo che  $X$  è una v.a. **finita** se  $E_X$  è un insieme finito.

Il legame tra variabili aleatorie e probabilità è descritto dalla distribuzione, definita come segue. Sia data una v.a. discreta  $X$  sullo spazio di probabilità  $(\Omega, \mathcal{F}, \mathbb{P})$  e sia  $E_X$  lo spazio dei valori che  $X$  assume. La **distribuzione (discreta)** (anche detta **legge** o **densità discreta**) di  $X$  è la funzione

$$\begin{aligned} p_X : E_X &\longrightarrow [0, 1] \\ x &\mapsto p_X(x) \stackrel{def}{=} \mathbb{P}(X = x) \end{aligned}$$

dove si è posto  $\{X = x\} = \{\omega : X(\omega) = x\} = X^{-1}(\{x\})$ . In particolare,  $p_X$  è una funzione non negativa tale che

$$\sum_{x \in E_X} p_X(x) = 1;$$

inoltre, per ogni  $A \subset \mathbb{R}$ ,

$$\mathbb{P}(X \in A) = \sum_{x : x \in A \cap E_X} p_X(x),$$

dove, come usuale, si è posto  $\{X \in A\} = X^{-1}(A) = \{\omega : X(\omega) \in A\}$  e  $\sum_{x : x \in \emptyset} (\cdot) \stackrel{def}{=} 0$ .

Com'è noto, nella pratica il formalismo appena introdotto non viene usato. Infatti, quando si lavora con variabili aleatorie in realtà si perde di vista qual è l'originario spazio di probabilità  $(\Omega, \mathcal{F}, \mathbb{P})$  e  $X$  vista come *applicazione* da  $\Omega$  su  $\mathbb{R}$ , anzi spesso tali ingredienti neanche si conoscono. Piuttosto, si lavora con i valori  $E_X$  che la v.a.  $X$  può assumere e con la sua distribuzione  $p_X$ <sup>2</sup>.

---

<sup>2</sup>Si può infatti dimostrare che fissati dei valori  $x_1, x_2, \dots$  (in numero finito o numerabile) e dei numeri  $p_1, p_2, \dots$  non negativi e tali che  $\sum_i p_i = 1$  allora esiste uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbb{P})$  sul quale si può definire una v.a.  $X$  tale che  $E_X = \{x_1, x_2, \dots\}$  e  $p_X(x_i) = p_i$  per ogni  $x_i \in E_X$ .



Siano ora  $X$  e  $Y$  due v.a., con spazio degli stati, rispettivamente,  $E_X$  e  $E_Y$ . Le rispettive distribuzioni  $p_X$  e  $p_Y$  descrivono, come abbiamo visto, il comportamento probabilistico delle due v.a. prese singolarmente. Per la descrizione probabilistica del comportamento congiunto occorre introdurre il concetto di **distribuzione congiunta** di  $X$  e  $Y$ , data dalla funzione

$$p_{X,Y}(x, y) : E_X \times E_Y \longrightarrow [0, 1]$$

$$(x, y) \mapsto p_{X,Y}(x, y) \stackrel{def}{=} \mathbb{P}(X = x, Y = y),$$

dove, come usuale, si è posto  $\{X = x, Y = y\} = \{\omega : X(\omega) = x \text{ e } Y(\omega) = y\} = X^{-1}(\{x\}) \cap Y^{-1}(\{y\})$ . Analogamente al caso unidimensionale, valgono le seguenti proprietà:

$$\sum_{x: x \in E_X} \sum_{y: y \in E_Y} p_{X,Y}(x, y) = 1$$

$$\mathbb{P}(X \in A, Y \in B) = \sum_{x: x \in A \cap E_X} \sum_{y: y \in B \cap E_Y} p_{X,Y}(x, y).$$

Inoltre, una volta nota la distribuzione congiunta, sono note anche le due distribuzioni marginali della  $X$  e della  $Y$ , date da:

$$p_X(x) = \sum_{y: y \in E_Y} p_{X,Y}(x, y), \quad x \in E_X$$

$$p_Y(y) = \sum_{x: x \in E_X} p_{X,Y}(x, y), \quad y \in E_Y.$$

Di più: nota la distribuzione congiunta è possibile (almeno in via teorica) calcolare anche la distribuzione di una v.a. funzione sia della  $X$  che della  $Y$ . Ad esempio, è facile vedere che la distribuzione della somma è

$$p_{X+Y}(z) = \sum_{x \in E_X} p_{X,Y}(x, z - x) = \sum_{y \in E_Y} p_{X,Y}(z - y, y), \quad (1.1)$$

per ogni  $z \in E_Z = \{x + y : x \in E_X, y \in E_Y\}$ .

È evidente che il concetto di distribuzione congiunta si può estendere al caso di  $n$  v.a.  $X_1, X_2, \dots, X_n$ : essa sarà data da

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n),$$

con  $x_i \in E_{X_i}$ ,  $i = 1, 2, \dots, n$ , e per calcolare la distribuzione (congiunta, eventualmente) di un qualsiasi sottoinsieme  $X_{i_1}, \dots, X_{i_k}$  delle v.a.  $X_1, \dots, X_n$  basta “saturare” (=sommare) la distribuzione congiunta rispetto alle variabili che non interessano:

$$p_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k}) = \sum_{j \notin \{i_1, \dots, i_k\}} \sum_{x_j \in E_{X_j}} p_n(x_1, \dots, x_n).$$

Ricordiamo poi che due v.a.  $X$  e  $Y$  si dicono **indipendenti** se

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B), \quad \text{per ogni } A, B.$$

È facile vedere che

$$X \text{ e } Y \text{ sono indipendenti se e solo se } p_{X,Y}(x, y) = p_X(x)p_Y(y),$$

cioè, se e solo se la distribuzione congiunta si fattorizza nel prodotto delle due distribuzioni marginali. Nel caso di v.a. indipendenti, la (1.1) diviene quindi

$$p_{X+Y}(z) = \sum_{x \in E_X} p_X(x)p_Y(z-x) = \sum_{y \in E_Y} p_X(z-y)p_Y(y), \quad (1.2)$$

per ogni  $z \in E_Z = \{x+y : x \in E_X, y \in E_Y\}$ , dunque la distribuzione di  $X+Y$  è data dalla “convoluzione discreta” delle distribuzioni marginali.

La definizione di indipendenza si estende in modo naturale ad un numero qualsiasi di v.a.:  $n$  v.a.  $X_1, \dots, X_n$  si dicono **indipendenti** se

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i), \quad \text{per ogni } A_1, \dots, A_n$$

o, equivalentemente, se e solo se

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n).$$

Infine, ricordiamo la **formula di Bayes**: fissato un evento  $\Lambda$  tale che  $\mathbb{P}(\Lambda) > 0$ , per  $A \in \mathcal{F}$  la probabilità di  $A$  dato  $\Lambda$  è data da

$$\mathbb{P}(A | \Lambda) = \frac{\mathbb{P}(A \cap \Lambda)}{\mathbb{P}(\Lambda)}.$$

Si noti che  $\mathbb{P}(A | \Lambda) = \mathbb{P}(A)$  se e solo se  $\mathbb{P}(A \cap \Lambda) = \mathbb{P}(A)\mathbb{P}(\Lambda)$ , cioè il verificarsi di  $\Lambda$  non dà informazioni sulla probabilità del verificarsi anche di  $A$  se e solo se  $\mathbb{P}(A \cap \Lambda) = \mathbb{P}(A)\mathbb{P}(\Lambda)$ . Infatti, in tal caso gli eventi  $A$  e  $\Lambda$  sono indipendenti. Inoltre, è immediato verificare che

$$\mathbb{Q}_\Lambda : A \ni \mathcal{F} \rightarrow \mathbb{Q}_\Lambda(A) = \mathbb{P}(A | \Lambda) \in [0, 1]$$

è una probabilità su  $(\Omega, \mathcal{F})$  detta probabilità condizionata. Tale fatto consente di introdurre la distribuzione condizionata: date due v.a.  $X$  e  $Y$  con distribuzione congiunta  $p_{X,Y}$  e distribuzioni marginali  $p_X$  e  $p_Y$  rispettivamente, fissato  $y \in E_Y$  tale che  $p_Y(y) > 0$  si definisce la **distribuzione di  $X$  condizionata a  $Y$**  come la funzione

$$p_{X|Y}(x | y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad x \in E_X$$

(sottolineiamo che perché sia una distribuzione dev'essere vista come una funzione di  $x$ , per ogni  $y \in E_Y$  fissato).  $p_{X|Y}(x|y)$  dà quindi la probabilità che  $X$  assuma il valore  $x$  una volta noto che  $Y = y$ . Osserviamo che  $X$  e  $Y$  sono indipendenti se e solo se

$$p_{X|Y}(x|y) = p_X(x), \quad x \in E_X,$$

quindi il verificarsi di  $\{Y = y\}$  non aggiunge né toglie informazioni sul verificarsi di  $\{X = x\}$ .

## 1.2 Media, varianza e covarianza

Proseguiamo nella rassegna degli ingredienti necessari per il prosieguo.

Fissato un intero positivo  $k$ , si dice che  $X$  **ha momento  $k$ -esimo finito** se

$$\sum_{x \in E_X} |x|^k p_X(x) < \infty.$$

In particolare, se  $E_X$  è finito, la serie sopra scritta è in realtà somma di un numero finito di termini, per ogni fissato  $k$ , quindi ogni v.a. finita ha momento finito di qualsiasi ordine. Se  $X$  ha momento  $k$ -esimo finito, di definisce **momento  $k$ -esimo di  $X$**  la quantità<sup>3</sup>

$$\sum_{x \in E_X} x^k p_X(x).$$

In particolare, si parla di **media** quando  $k = 1$ . Quindi, riassumendo, una v.a. discreta  $X$  ha media se e solo se

$$\sum_{x \in E_X} |x| p_X(x) < \infty$$

e la **media della v.a.  $X$**  (anche detta **speranza matematica** o **aspettazione** o **valore atteso**) è data da:

$$\mathbb{E}(X) = \sum_{x \in E_X} x p_X(x).$$

Ora, siano date una v.a.  $X$  ed una funzione  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Allora la funzione composta  $g \circ X(\omega) = g(X(\omega))$  manda  $\Omega$  su  $\mathbb{R}$  ed è ancora (sotto buone condizioni su  $g$ , per noi sempre verificate) una variabile aleatoria, discreta se  $X$  è discreta. Denotiamo quindi con  $Y$  tale v.a.:  $Y = g(X)$ .

---

<sup>3</sup>Ricordiamo che una serie assolutamente convergente è convergente, quindi la definizione di momento  $k$ -esimo è ben posta.

Potremmo pensare, ad esempio, di calcolarne la media, se esiste. Per quanto appena detto, dovremmo dapprima calcolarne la distribuzione  $p_Y$ , studiare la convergenza della serie  $\sum_{y \in E_Y} |y| p_Y(y)$  e, se tale serie converge, calcolare infine  $\mathbb{E}(Y) = \sum_{y \in E_Y} y p_Y(y)$ . Oppure, un modo più veloce (che, se non altro, elimina il problema della calcolo della distribuzione  $p_Y$  di  $Y$ ) è quello di usare il seguente

**Teorema 1.2.1.**  $Y = g(X)$  ha media se e solo se

$$\sum_{x \in E_X} |g(x)| p_X(x) < \infty$$

e in tal caso

$$\mathbb{E}(Y) = \sum_{x \in E_X} g(x) p_X(x) < \infty.$$

Per la dimostrazione, si rimanda al Teorema 2.29 in [1]. Osserviamo che, in virtù del Teorema precedente, se una v.a. ha momento  $k$ -esimo finito allora la v.a.  $g(X) = X^k$  ha media data da

$$\mathbb{E}(X^k) = \sum_{x \in E_X} x^k p_X(x),$$

quindi il momento  $k$ -esimo non è altro che la media della v.a.  $X^k$ .

Utilizzando anche il Teorema precedente, è possibile dimostrare alcune proprietà fondamentali della media, qui di seguito riassunte (per la dimostrazione si rimanda a [1], Proposizioni 2.30, 2.31, 2.37):

**Proposizione 1.2.2.** *Siano  $X$  e  $Y$  due v.a. aventi media.*

1. (Omogeneità) Per ogni  $c \in \mathbb{R}$ , la v.a.  $cX$  ha media e  $\mathbb{E}(cX) = c\mathbb{E}(X)$ .
2. (Additività)  $X + Y$  ha media e  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ .
3. (Positività) Se  $\mathbb{P}(X \geq 0) = 1$  (ovvero  $E_X \subset [0, +\infty)$ ) allora  $\mathbb{E}(X) \geq 0$  e l'uguaglianza è possibile se e solo se  $\mathbb{P}(X = 0) = 1$  (ovvero  $E_X = \{0\}$ ).
4. (Monotonia) Se  $\mathbb{P}(X \geq Y) = 1$  allora  $\mathbb{E}(X) \geq \mathbb{E}(Y)$  e l'uguaglianza è possibile se e solo se  $\mathbb{P}(X = Y) = 1$ .
5.  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ .
6. Fissato  $k > 1$ , se  $X$  ha momento  $k$ -esimo finito allora ha momento  $j$ -esimo finito per ogni  $1 \leq j \leq k$ .
7. Fissato  $k > 1$ , se  $X$  e  $Y$  hanno momento  $k$ -esimo finito allora  $X + Y$  ha momento  $k$ -esimo finito.

Un altro strumento di cui faremo largo uso è la **varianza**. Per definirla, ricordiamo che una v.a. ha momento secondo finito se  $\sum_{x \in E_X} x^2 p_X(x) < \infty$  e in tal caso il **momento secondo** è dato da

$$\mathbb{E}(X^2) = \sum_{x \in E_X} x^2 p_X(x).$$

Ora, se una v.a.  $X$  ha momento secondo, allora anche la v.a.  $X - \mathbb{E}(X)$  ha momento secondo. Allora, per v.a. con momento secondo definiamo la **varianza** come il momento secondo di  $X - \mathbb{E}(X)$ :

$$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

(la seconda uguaglianza segue dalla linearità e omogeneità della media).

A differenza della media, la varianza non è in generale lineare, né omogenea. Studieremo fra poco l'eventuale linearità; per quanto riguarda l'omogeneità, essendo la varianza definita come un momento secondo, cioè come media di qualcosa di quadratico, è immediato verificare che per ogni  $c \in \mathbb{R}$  e  $X$  con momento secondo, allora la varianza di  $cX$  esiste e vale

$$\text{Var}(cX) = c^2 \text{Var}(X),$$

così come vale

$$\text{Var}(X + c) = \text{Var}(X)$$

cioè la varianza è invariante per traslazioni (deterministiche).

Prima di proseguire nella rassegna dei risultati, cerchiamo di capire meglio qual è il significato probabilistico della media e della varianza di una v.a.

Sia  $X$  una v.a. In particolare quindi,  $X$  non si conosce ma si conosce la "misura", cioè la probabilità, che tale v.a. assuma un valore  $x \in E_X$ , data appunto da  $p_X(x)$ . Ora, supponiamo di voler approssimare  $X$  con una quantità deterministica, cioè nota<sup>4</sup>,  $c$ . Come fare? Un modo potrebbe essere quello di considerare la distanza tra  $X$  e  $c$ , data da

$$|X - c|.$$

Com'è noto, lavorare con i moduli non è molto comodo, quindi è meglio considerare il quadrato della distanza:

$$|X - c|^2 = (X - c)^2.$$

---

<sup>4</sup>Il lettore avrà notato la seguente convenzione: le quantità aleatorie sono denotate con lettere *maiuscole* mentre le quantità deterministiche con lettere *minuscole*. Avrà inoltre notato che la media, così come la varianza e i momenti di una v.a. sono quantità deterministiche.

Ora, tale quantità è ancora aleatoria. Un modo per ovviare a questo problema è quello di considerarne la media. Definiamo allora

$$\psi(c) = \mathbb{E}\left((X - c)^2\right), \quad c \in \mathbb{R}.$$

Dunque, la funzione  $\psi$  è strettamente legata ad una forma di distanza<sup>5</sup> tra  $X$  e  $c$  ed inoltre è una quantità deterministica. Allora, un modo di rispondere al problema dell'approssimazione di una v.a.  $X$  con una costante  $c$  è quello di trovare la costante migliore  $c^*$  che rende minima la funzione  $\psi$ :

$$\psi(c^*) = \min_{c \in \mathbb{R}} \psi(c) \equiv \min_{c \in \mathbb{R}} \mathbb{E}\left((X - c)^2\right).$$

Il problema della determinazione di  $c^*$  è davvero molto semplice<sup>6</sup> e si ha

$$c^* = \mathbb{E}(X) \quad \text{e} \quad \psi(c^*) = \text{Var}(X).$$

Allora, la media  $\mathbb{E}(X)$  di una v.a. è la costante che approssima meglio una v.a.  $X$  (nel senso sopra descritto) e la sua varianza  $\text{Var}(X)$  dà in qualche senso l'errore che si fa se  $X$  è approssimata con  $\mathbb{E}(X)$ . Tra l'altro, questa interpretazione di  $\text{Var}(X)$  motiva anche perché spesso si dice che la varianza misura la *dispersione* di  $X$  intorno alla sua media  $\mathbb{E}(X)$ .

Uno strumento che invece dà indicazioni della relazione che sussiste tra due v.a. è la covarianza, definita come segue: se  $X$  e  $Y$  sono due v.a. con momento secondo<sup>7</sup>, la **covarianza tra  $X$  e  $Y$**  è data da

$$\text{Cov}(X, Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right) = \mathbb{E}(X Y) - \mathbb{E}(X) \mathbb{E}(Y)$$

(ancora una volta, la linearità e l'omogeneità della media garantiscono la seconda uguaglianza dell'espressione sopra scritta). Si noti che la covarianza è simmetrica ed anzi, è un'applicazione bilineare<sup>8</sup>.

La covarianza è uno strumento piuttosto importante nel Calcolo delle Probabilità. Ad esempio, è usata per il calcolo della varianza della somma di v.a.: se  $X_1, X_2, \dots, X_n$  sono  $n$  v.a. con momento secondo allora (si veda la Proposizione 1.2.2) anche  $X_1 + X_2 + \dots + X_n$  ha momento secondo, dunque ha media e varianza date da

$$\mathbb{E}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \mathbb{E}(X_i)$$

<sup>5</sup>Per i più grandi, la funzione  $\psi(c)$  dà (il quadrato del)la distanza in  $L^2$  tra  $X$  e  $c$ .

<sup>6</sup>Infatti, dalla linearità ed omogeneità della media, segue immediatamente che  $\psi(c) = \mathbb{E}(X^2) - 2c\mathbb{E}(X) + c^2$ .

<sup>7</sup>Questa richiesta garantisce l'esistenza di quanto si sta per definire, ma non entreremo nei dettagli, per i quali si rimanda a [1], §2, Cap. 2.

<sup>8</sup>Cioè,  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$  e per ogni v.a.  $X_1, X_2, Y_1, Y_2$  con momento secondo e per ogni costante  $a_1, a_2, b_1, b_2$ , si ha:  $\text{Cov}(a_1 X_1 + a_2 X_2, b_1 Y_1 + b_2 Y_2) = a_1 b_1 \text{Cov}(X_1, Y_1) + a_1 b_2 \text{Cov}(X_1, Y_2) + a_2 b_1 \text{Cov}(X_2, Y_1) + a_2 b_2 \text{Cov}(X_2, Y_2)$ .

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}(X_i, X_j)$$

In particolare, se  $\text{Cov}(X, Y) = 0$  si dice che le v.a.  $X$  e  $Y$  sono **non correlate**. Prese quindi  $n$  v.a. a due a due non correlate  $X_1, X_2, \dots, X_n$ , la seconda sommatoria che compare nella formula sopra scritta è fatta di termini tutti uguali a 0, quindi in tal caso

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

In particolare, **due v.a.  $X$  e  $Y$  indipendenti sono non correlate**: è facile dimostrare che se  $X$  e  $Y$  sono indipendenti allora (cfr. Proposizione 2.33 di [1])

$$\mathbb{E}(X Y) = \mathbb{E}(X) \mathbb{E}(Y)$$

dunque  $\text{Cov}(X, Y) = 0$ . Ma allora:

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i), \quad \text{se } X_1, X_2, \dots, X_n \text{ sono indipendenti.}$$

Ricordiamo (e sottolineiamo) che

$$\text{indipendenza} \not\Rightarrow \text{covarianza}=0$$

Anche la covarianza ha un'interessante interpretazione probabilistica, che è legata alla **retta di regressione** e che esprime il concetto di **dipendenza (positiva o negativa)** tra v.a. Gli interessati possono trovare una breve descrizione di tali oggetti nell'Appendice 1.6.

Concludiamo con alcune utili disuguaglianze.

**Proposizione 1.2.3.** *Siano  $X$  e  $Y$  due v.a. tali che tutte le medie coinvolte nel seguito esistono.*

- **(Jensen)** *Se  $g : \mathbb{R} \rightarrow \mathbb{R}$  è una funzione convessa<sup>9</sup>, allora*

$$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X)).$$

- **(Hölder)** *Siano  $p, q > 0$  tali che  $\frac{1}{p} + \frac{1}{q} = 1$ . Allora*

$$|\mathbb{E}(X Y)| \leq \mathbb{E}(|X Y|) \leq \mathbb{E}(|X|^p)^{1/p} \mathbb{E}(|Y|^q)^{1/q}.$$

---

<sup>9</sup>Cioè,  $g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y)$ , per ogni  $x, y \in \mathbb{R}$  e  $\alpha \in [0, 1]$ .

In particolare (**Cauchy-Schwarz**)

$$\mathbb{E}(|X - Y|) \leq \mathbb{E}(|X|^2)^{1/2} \mathbb{E}(|Y|^2)^{1/2}$$

e quindi

$$|\text{Cov}(X, Y)| \leq \left( \text{Var}(X) \text{Var}(Y) \right)^{1/2}$$

- (**Minkowski**) Sia  $p \geq 1$ . Allora

$$\mathbb{E}(|X + Y|^p)^{1/p} \leq \mathbb{E}(|X|^p)^{1/p} + \mathbb{E}(|Y|^p)^{1/p}.$$

Concludiamo questo paragrafo con una tabella contenente distribuzione, media e varianza di alcune v.a. discrete largamente utilizzate.

NOME	SIMBOLO	LEGGE	MEDIA	VARIANZA
Bernoulli	$\text{Be}(p), p \in [0, 1]$	$p_0 = p, p_1 = 1 - p$	$p$	$p(1 - p)$
Binomiale	$\text{Bi}(n, p),$ $n \in \mathbb{N}, p \in [0, 1]$	$p_k = \binom{n}{k} p^k (1 - p)^{n-k},$ $k = 0, 1, \dots, n$	$np$	$np(1 - p)$
Geometrica	$\text{Ge}(p), p \in (0, 1)$	$p_k = p(1 - p)^k, k = 0, 1, \dots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson	$\text{Po}(\lambda), \lambda > 0$	$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \dots$	$\lambda$	$\lambda$

### 1.3 La Legge dei Grandi Numeri

Cominciamo questo paragrafo presentando una disuguaglianza di cui faremo largo uso, la disuguaglianza di Chebycev:

**Proposizione 1.3.1.** (i) Sia  $Y$  una v.a. non negativa con momento  $k$ -esimo. Vale allora la **disuguaglianza di Markov**: per ogni  $\delta > 0$  si ha

$$\mathbb{P}(Y > \delta) \leq \mathbb{P}(Y \geq \delta) \leq \frac{\mathbb{E}(Y^k)}{\delta^k}. \quad (1.3)$$

(ii) Sia  $Z$  una v.a. con momento secondo. Vale allora la **disuguaglianza di Chebycev**: per ogni  $\delta > 0$  si ha

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > \delta) \leq \mathbb{P}(|Z - \mathbb{E}(Z)| \geq \delta) \leq \frac{\text{Var}(Z)}{\delta^2}. \quad (1.4)$$



**Dimostrazione.** (i) Sia  $\delta > 0$ . Allora<sup>10</sup>,

$$Y^k = Y^k \mathbb{I}_{\{Y \geq \delta\}} + Y^k \mathbb{I}_{\{Y < \delta\}} \geq Y^k \mathbb{I}_{\{Y \geq \delta\}} \geq \delta^k \mathbb{I}_{\{Y \geq \delta\}},$$

dove si è usato che  $Y \geq 0$  (cioè  $E_Y \subset [0, +\infty)$ ). Quindi, dal punto 4 della Proposizione 1.2.2 (monotonia della media), otteniamo

$$\mathbb{E}(Y^k) \geq \mathbb{E}(\delta^k \mathbb{I}_{\{Y \geq \delta\}}) \geq \delta^k \mathbb{E}(\mathbb{I}_{\{Y \geq \delta\}}) = \delta^k \mathbb{P}(Y \geq \delta)$$

da cui segue immediatamente la disuguaglianza di Markov.

(ii) Applicando la disuguaglianza di Markov alla v.a.  $Y = |Z - \mathbb{E}(Z)|$ , con  $k = 2$ , otteniamo

$$\mathbb{P}(|Z - \mathbb{E}(Z)| \geq \delta) \leq \frac{\mathbb{E}\left((Z - \mathbb{E}(Z))^2\right)}{\delta^2} = \frac{\text{Var}(Z)}{\delta^2}$$

□

Usando la disuguaglianza di Chebycev, è possibile dimostrare la Legge dei Grandi Numeri (nella sua versione più semplice), che sarà uno degli strumenti fondamentali per le simulazioni che andremo a fare.

**Teorema 1.3.2.** *Siano  $X_1, X_2, \dots$  v.a. indipendenti ed identicamente distribuite (i.i.d.<sup>11</sup>) e dotate di momento secondo. Siano  $m$  e  $\sigma^2$  rispettivamente la media e la varianza comune. Sia  $\delta > 0$ . Allora: per ogni  $n$  si ha*

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| \geq \delta\right) \leq \frac{\sigma^2}{n}.$$

In particolare quindi,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| \geq \delta\right) = 0.$$

o, equivalentemente,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| < \delta\right) = 1.$$

<sup>10</sup>La funzione indicatrice di  $A$  è la funzione

$$\mathbb{I}_{x \in A} = \mathbb{I}_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$$

(in analisi è comunemente detta “funzione caratteristica”, e si denota con  $\chi_A$ ). Se  $A$  è un evento, cioè  $A \in \mathcal{F}$ , allora  $\mathbb{I}_A \equiv \mathbb{I}_A(\omega)$  è una variabile aleatoria, con media  $\mathbb{E}(\mathbb{I}_A) = \mathbb{P}(A)$ . In particolare quindi è una v.a. bernoulliana, di parametro  $p = \mathbb{P}(A)$ .

<sup>11</sup>Vuol dire che, oltre ad essere indipendenti, hanno tutte la stessa distribuzione, in particolare quindi stessi momenti.

**Dimostrazione.** Le ipotesi fatte consentono di applicare la disuguaglianza di Chebycev alla v.a.  $Z = \frac{1}{n} \sum_{k=1}^n X_k$ . Poiché

$$\mathbb{E}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k) = \frac{1}{n} \sum_{k=1}^n m = m$$

e (ricordando che le  $X_k$  sono indipendenti)

$$\text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \text{Var}\left(\sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{1}{n^2} \sum_{k=1}^n \sigma^2 = \frac{\sigma^2}{n},$$

dalla (1.4) segue che, per ogni  $\delta > 0$ ,

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| \geq \delta\right) = \mathbb{P}(|Z - \mathbb{E}(Z)| \geq \delta) \leq \frac{\text{Var}(Z)}{\delta^2} = \frac{\sigma^2}{n},$$

per ogni fissato  $n$ . Mandando  $n$  all'infinito, segue immediatamente che

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| \geq \delta\right) = 0$$

e quindi, passando al complementare,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| < \delta\right) = 1.$$

□

Il Teorema 1.3.2 è noto in letteratura con il nome di **Legge (Debole) dei Grandi Numeri (LGN)** e in qualche modo afferma che la successione di v.a. definita dalla media aritmetica delle prime  $n$  v.a.

$$\frac{1}{n} \sum_{k=1}^n X_k,$$

converge a  $m$  quando  $n \rightarrow \infty$ . Tale forma di convergenza non è ovviamente l'usuale convergenza cui si è abituati a lavorare al primo anno di Matematica e, in qualche senso, è una forma di convergenza più debole di quella classica. Vediamo perché.

Fissiamo un numero  $\delta > 0$ , immaginandolo piccolo. L'evento

$$A_n^\delta = \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - m \right| > \delta \right\}$$

si può tradurre come l'evento "la media aritmetica dista dalla media teorica ( $m$ ) per più di  $\delta$ ". Tale evento avrà quindi una probabilità, stimata nella prima parte del Teorema 1.3.2:

$$\mathbb{P}(A_n^\delta) \leq \frac{\sigma^2}{n}.$$

All'aumentare di  $n$ , tale probabilità è sempre più piccola e converge a 0 per  $n \rightarrow \infty$ , così come la probabilità del complementare di  $A_n^\delta$  tende a 1. Il che tradotto diviene: per  $n \rightarrow \infty$ , l'evento "la media aritmetica dista dalla media teorica per più di  $\delta$ " tende a diventare un evento trascurabile. Oppure, usando il complementare: per  $n \rightarrow \infty$ , l'evento "la media aritmetica dista dalla media teorica per meno di  $\delta$ " tende a diventare un evento certo. Ora, se immaginiamo  $\delta$  piccolo, è chiaro che tale proprietà esprime una forma di convergenza della media empirica verso la media teorica.

Vediamo un esempio elementare della Legge dei Grandi Numeri, che mostra anche come in realtà essa sia usata nella pratica.

Supponiamo di avere una moneta di cui non sappiamo nulla, cioè non sappiamo se è truccata o no. Per avere indicazioni sulla probabilità che esca testa o croce, tipicamente quello che si fa è tirare tante volte la moneta, segnare quante volte esce testa e quante volte esce croce e stimare la probabilità che dia testa (ad esempio) come il rapporto tra il numero di volte in cui si è osservato testa sul numero totale dei tiri. In altre parole, si usa la Legge dei Grandi Numeri. Vediamo perché.

Indichiamo con  $X_k$  il risultato del  $k$ -esimo lancio, con la convenzione (ad esempio)

$$X_k = \begin{cases} 1 & \text{se al } k\text{-esimo lancio si è osservato TESTA} \\ 0 & \text{se al } k\text{-esimo lancio si è osservato CROCE} \end{cases}$$

Le v.a.  $X_k$  hanno ovviamente tutte la stessa distribuzione (la moneta è sempre la stessa!) ed inoltre è abbastanza intuitivo supporre che i lanci avvengano indipendentemente l'uno dall'altro. Quindi, possiamo assumere che le v.a.  $X_k$  siano i.i.d. Allora, la v.a.

$$\sum_{k=1}^n X_k$$

conta il numero di volte in cui è uscita testa sugli  $n$  lanci effettuati, quindi la v.a.

$$\frac{1}{n} \sum_{k=1}^n X_k$$

dà la proporzione delle volte in cui è uscita testa sugli  $n$  lanci totali. Ci troviamo quindi nella condizione di usare la Legge dei Grandi Numeri, che

garantisce che (nel senso prima specificato)

$$\frac{1}{n} \sum_{k=1}^n X_k \text{ "converge" per } n \rightarrow \infty \text{ alla media teorica } m.$$

Ma chi è  $m$ ? Si ha

$$\begin{aligned} m &= \mathbb{E}(X_k) = \mathbb{E}(\mathbb{I}_{\{\text{al } k\text{-esimo lancio esce testa}\}}) \\ &= \mathbb{P}(\{\text{al } k\text{-esimo lancio esce testa}\}) \end{aligned}$$

cioè  $m$  è proprio la probabilità che la moneta dia testa.

Vedremo nel seguito numerosi altri esempi di applicazione della Legge dei Grandi Numeri, che è alla base dei Metodi Monte Carlo di cui ci occuperemo in questo corso.

## 1.4 Variabili aleatorie assolutamente continue

Sia  $X$  una v.a. Dalla Definizione 1.1.1 segue che per ogni  $t \in \mathbb{R}$ , l'insieme  $\{X \leq t\}$  è un evento, cioè appartiene alla  $\sigma$ -algebra  $\mathcal{F}$  su cui è definita la probabilità di riferimento  $\mathbb{R}$ . In realtà, si può dire di più.

**Lemma 1.4.1.** *Sia  $X$  una v.a. su  $(\Omega, \mathcal{F}, \mathbb{P})$ . Allora per ogni intervallo  $I$  di  $\mathbb{R}$ ,  $\{X \in I\}$  è un evento, cioè*

$$\{X \in I\} \in \mathcal{F}.$$

**Dimostrazione.** Mostriamo che l'asserzione è vera per gli intervalli della forma: (i)  $I = (t, +\infty)$ , (ii)  $I = (-\infty, t)$ , (iii)  $I = [t, +\infty)$ .

(i) Poiché  $\{X \in (t, +\infty)\} = \{X \leq t\}^c \in \mathcal{F}$  e  $\mathcal{F}$  è chiusa sotto operazione di complementare,  $\{X \in (t, +\infty)\} = \{X > t\} \in \mathcal{F}$ , per ogni  $t \in \mathbb{R}$ .

(ii)  $(-\infty, t) = \bigcup_{n=1}^{\infty} (-\infty, t - 1/n]$ , quindi  $\{X \in (-\infty, t)\} = \bigcup_{n=1}^{\infty} \{X \in (-\infty, t - 1/n]\} = \bigcup_{n=1}^{\infty} \{X \leq t - 1/n\}$ . Poiché  $\mathcal{F}$  è chiusa sotto operazione di unione numerabile,  $\{X \in (-\infty, t)\} = \{X < t\} \in \mathcal{F}$ , per ogni  $t \in \mathbb{R}$ .

(iii) Poiché  $\{X \in [t, +\infty)\} = \{X \geq t\} = \{X < t\}^c \in \mathcal{F}$  e  $\mathcal{F}$  è chiusa sotto operazione di complementare,  $\{X \in [t, +\infty)\} = \{X > t\} \in \mathcal{F}$ , per ogni  $t \in \mathbb{R}$ .

Da (i), (ii) e (iii) segue che  $\{X \in I\} \in \mathcal{F}$  per ogni intervallo  $I$  di estremi  $a$  e  $b$ , , aperto o chiuso a destra e/o a sinistra. Infatti, si può sempre scrivere  $I = I_1 \cap I_2$ , con  $I_1$  e  $I_2$  intervallo non limitato, aperto o chiuso (ad esempio,  $[a, b) = [a, +\infty) \cap (-\infty, b)$ ), quindi

$$\{X \in I\} = \{X \in I_1\} \cap \{X \in I_2\} \in \mathcal{F}$$

perché, come visto sopra,  $\{X \in I_1\}, \{X \in I_2\} \in \mathcal{F}$  e  $\mathcal{F}$  è chiusa sotto operazione di intersezione finita o numerabile.

□

Il Lemma 1.4.1 garantisce quindi che per ogni  $I$  intervallo di  $\mathbb{R}$ , è ben posta  $\mathbb{P}(X \in I)$ , dunque nel seguito potremo sempre scrivere  $\mathbb{P}(X \in I)$ .

In particolare, è ben posta la funzione

$$F_X(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

La funzione  $F_X$  prende il nome di **funzione di distribuzione (o funzione di ripartizione)** di  $X$ . Riassumiamo nella seguente proposizione alcune delle principali proprietà di  $F_X$  (cfr. pp. 70-71 e Proposizione 3.1 di [1])

**Proposizione 1.4.2.** 1. Per ogni  $x \in \mathbb{R}$ ,  $0 \leq F_X(x) \leq 1$ .

2.  $F_X$  è una funzione monotona non decrescente.

3.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  e  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

4.  $F_X$  è continua a destra e con limite da sinistra ed inoltre  $F_X(x_0^-) \leq F_X(x_0^+) = F_X(x_0)$  ( $F_X(x_0^-)$ =limite da sinistra;  $F_X(x_0^+)$ =limite da destra).

5. Per ogni  $-\infty \leq a \leq b \leq +\infty$ ,

$$\begin{aligned} \mathbb{P}(a < X \leq b) &= F_X(b) - F_X(a) \\ \mathbb{P}(a < X < b) &= F_X(b^-) - F_X(a) \\ \mathbb{P}(a \leq X \leq b) &= F_X(b) - F_X(a^-) \\ \mathbb{P}(a \leq X < b) &= F_X(b^-) - F_X(a^-) \end{aligned}$$

(dove si è posto  $F_X(-\infty) = 0$ ,  $F_X(+\infty) = 1$ ).

6. Per ogni  $x \in \mathbb{R}$ ,  $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$ .

Per le v.a. discrete, è facile vedere  $F_X$  è sempre una *funzione costante a tratti*, che nei punti di salto assume sempre il valore a destra.

Se invece  $F_X$  è continua, si dice che  $X$  è **una v.a. continua**. In tal caso quindi

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-) = 0$$

per ogni  $x$ , cioè  $X$  assume con probabilità nulla un qualsiasi numero prefissato.

Una v.a. si dice **assolutamente continua** se esiste una funzione  $p_X : \mathbb{R} \rightarrow [0, +\infty)$  tale che per ogni intervallo  $I$  di estremi  $a$  e  $b$  si ha

$$\mathbb{P}(X \in I) = \int_a^b p_X(x) dx. \quad (1.5)$$

La funzione  $p_X$  prende il nome di **densità di probabilità** di  $X$ . Scelto  $I = (-\infty, x]$ , si ottiene<sup>12</sup>

$$F_X(x) = \int_{-\infty}^x p_X(x) dx, \quad (1.6)$$

che dà la rappresentazione in forma integrale per la funzione di distribuzione di una v.a. assolutamente continua. Si noti che  $F_X$  è continua (Teorema Fondamentale del Calcolo Integrale), dunque le v.a. assolutamente continue sono in particolare continue. Inoltre, scelto un intervallo  $I$  (aperto o chiuso, a destra e/o a sinistra) di estremi  $a$  e  $b$ , usando la (1.6) si ottiene

$$\mathbb{P}(X \in I) = F_X(b) - F_X(a) = \int_a^b p_X(x) dx$$

---

<sup>12</sup>Chiarimento qual è il senso che si vuole dare agli integrali, qui e nel seguito. Intanto, poiché nei primi anni di Matematica si studia l'integrale di Riemann, l'ipotesi che faremo sempre sulle funzioni  $g$  di cui studieremo l'integrale è che  $g$  sia una funzione continua a tratti. In generale, data una funzione  $g \geq 0$ , porremo:

$$\int_{-\infty}^b g(t) dt \stackrel{def}{=} \lim_{a \rightarrow -\infty} \int_a^b g(t) dt \quad \text{e} \quad \int_a^{+\infty} g(t) dt \stackrel{def}{=} \lim_{b \rightarrow +\infty} \int_a^b g(t) dt$$

purché tali limiti siano *finiti*. In tal caso, definiamo anche

$$\int_{-\infty}^{+\infty} g(t) dt \stackrel{def}{=} \int_{-\infty}^x g(t) dt + \int_x^{+\infty} g(t) dt = \lim_{\alpha \rightarrow +\infty} \int_{-\alpha}^{\alpha} g(t) dt.$$

Osserviamo che poiché  $\int_a^b g(t) dt = \int_a^x g(t) dt + \int_x^b g(t) dt$  per ogni  $a \leq x \leq b$ , cioè l'integrale è un operatore additivo, si verifica facilmente che tale definizione è ben posta (cioè: indipendente dalla scelta di  $x$ ), come peraltro si evince dalla seconda uguaglianza. Per funzioni  $\psi$  qualsiasi (cioè: non necessariamente non-negative), diremo che  $\psi$  è integrabile se esiste finito  $\int_{-\infty}^{+\infty} |\psi(t)| dt$ , cioè se esiste l'integrale di  $g(t) = |\psi(t)|$ . In tal caso, si definiscono

$$\begin{aligned} \int_{-\infty}^b \psi(t) dt &\stackrel{def}{=} \lim_{a \rightarrow -\infty} \int_a^b \psi(t) dt \quad \text{e} \quad \int_a^{+\infty} \psi(t) dt \stackrel{def}{=} \lim_{b \rightarrow +\infty} \int_a^b \psi(t) dt \\ \int_{-\infty}^{+\infty} \psi(t) dt &\stackrel{def}{=} \int_{-\infty}^x \psi(t) dt + \int_x^{+\infty} \psi(t) dt = \lim_{\alpha \rightarrow +\infty} \int_{-\alpha}^{\alpha} \psi(t) dt. \end{aligned}$$

(si può dimostrare che tali definizioni sono in effetti ben poste).

La teoria sottostante alle definizioni precedenti è, nel caso dei primi due anni di Matematica, la teoria degli integrali impropri. Non diremo molto di più in proposito; gli studenti potranno acquisire i risultati teorici sottostanti negli opportuni corsi di Analisi. Faremo però uso di questa forma di integrazione, perché in realtà dal punto di vista computazionale non cambia (quasi) nulla rispetto all'usuale integrale di Riemann.

(dove si è usata la proprietà di additività dell'integrale), da cui segue che le (1.5) e (1.6) sono due modi equivalenti di definire le v.a. assolutamente continue. Usando la Proposizione 1.4.2, da quest'ultima uguaglianza segue che

$$1 = \lim_{b \rightarrow +\infty} F_X(b) - \lim_{a \rightarrow -\infty} F_X(a) = \int_{-\infty}^{+\infty} p_X(x) dx = \mathbb{P}(X \in \mathbb{R}),$$

cioè: una densità dev'essere una funzione integrabile, non negativa e tale che  $\int_{-\infty}^{+\infty} p_X(x) dx = 1$ . In realtà, si può dimostrare che vale anche il viceversa: presa  $f$  non negativa, integrabile e tale che  $\int_{-\infty}^{+\infty} f(x) dx = 1$ , allora esiste una v.a.  $X$  assolutamente continua che ha  $p_X = f$  come densità<sup>13</sup>.

Ora, nota la densità  $p_X$ , usando la (1.6) si può calcolare la funzione di distribuzione  $F_X$ . Ma la (1.6) consente anche di calcolare la densità di probabilità, una volta nota la funzione di distribuzione. Infatti, la (1.6) dice che  $F_X$  è la funzione integrale di  $p_X$ . Quindi, per il Teorema Fondamentale del Calcolo Integrale, se  $F_X$  è una funzione derivabile con derivata continua su tutto  $\mathbb{R}$  (tranne al più un numero finito di punti) allora  $F_X$  è una primitiva di  $F'_X$ :

$$F_X(x) = \int_{-\infty}^x F'_X(x) dx.$$

Quindi  $p_X = F'_X$  può essere scelta come densità<sup>14</sup>.

Proponiamo qui un esempio, di cui faremo largo uso nel capitolo successivo. Fissati due numeri  $a < b$ , una v.a.  $X$  si dice **uniforme su**  $E = (a, b)$ , scriveremo brevemente  $\text{Un}(a, b)$ , se la sua densità è

$$p_X(x) = \frac{1}{b-a} \mathbb{I}_{(a,b)}(x).$$

Calcoliamo la funzione di distribuzione associata: per ogni  $x \in \mathbb{R}$ ,

$$F_X(x) = \int_{-\infty}^x \frac{1}{b-a} \mathbb{I}_{(a,b)}(t) dt.$$

Ora, se  $x < a$ , la funzione integranda è nulla nell'intervallo di integrazione, quindi  $F_X(x) = 0$ . Se  $a \leq x < b$ , allora, tenendo conto che  $p_X(t) = 0$  per  $t < a$ ,

$$F_X(x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}.$$

Infine, se  $x \geq b$  rimane

$$F_X(x) = \int_a^b \frac{1}{b-a} dt = 1.$$

<sup>13</sup>Confrontare questa affermazione con quanto riportato nella nota 2.

<sup>14</sup>In generale, la densità di probabilità non è univocamente determinata. Ad esempio, se si modifica  $p_X$  su un numero finito di punti, il risultato rimane ancora una densità. Ciò che rimane univocamente determinata è la sua primitiva, cioè  $F_X$ .

Riassumendo,

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

Quindi, per ogni intervallo  $I$  di estremi  $\alpha \leq \beta$ ,

$$\mathbb{P}(X \in I) = F_X(\beta) - F_X(\alpha) = \frac{\beta \wedge b - \alpha \vee a}{b - a},$$

dove  $\wedge$ =minimo e  $\vee$ =massimo. Particolare importanza rivestono le v.a. **uniformi su**  $(0, 1)$ , denotate con  $\text{Un}(0, 1)$ , quando cioè si scelgono  $a = 0$  e  $b = 1$ . In tal caso,

$$p_X(x) = \mathbb{I}_{(0,1)}(x) \quad \text{e} \quad F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1. \end{cases}$$

In Figura 1.1 sono visualizzate le funzioni  $p_X$  e  $F_X$  del caso uniforme.



**Figura 1.1** Grafico della densità (a sinistra) e della funzione di distribuzione (a destra) della legge uniforme.

Anche nel caso discreto esistono le v.a. uniformi. Ricordiamo che dato un insieme finito  $E = \{x_1, x_2, \dots, x_n\}$ , per qualche  $x_1, x_2, \dots, x_n \in \mathbb{R}$  distinti, una v.a. (discreta)  $X$  si dice **uniforme su**  $E$  se assume punti di  $E$  con uguale probabilità: posto per semplicità di notazione  $p_i = p_X(x_i)$ , dev'essere

$$p_i = p, \quad \text{per ogni } i.$$

Ora, poiché  $\sum_i p_i = 1$ , si ottiene  $1 = np$ , ovvero

$$p_i = \frac{1}{n}, \quad \text{per ogni } i.$$

Per ogni sottoinsieme  $A$  di  $\mathbb{R}$ , sappiamo che  $\mathbb{P}(X \in A) = \sum_{x \in A \cap E} p_i$ , che in tal caso diventa  $\mathbb{P}(X \in A) = \frac{\#A \cap E}{n}$ . In particolare quindi  $X$  è una v.a. discreta uniforme su  $E$  se e solo se<sup>15</sup>

$$\mathbb{P}(X \in A) = \frac{\#A \cap E}{\#E}. \quad (1.7)$$

<sup>15</sup>Osserviamo che tale uguaglianza si può interpretare come: la probabilità che  $X$  appartenga ad  $A$  è data dal rapporto tra i *casi favorevoli ad*  $A$  ( $\#A \cap E$  [si noti che da  $E$  non si può uscire!]) e tutti i *casi possibili* ( $\#E$ ).



Si dice anche che  $X$  è **un punto scelto a caso in  $E$** . Vediamo come le v.a. uniformi assolutamente continue siano la naturale generalizzazione al caso continuo.

Fissato un intervallo limitato  $I$  di estremi  $\alpha$  e  $\beta$ , sia  $\lambda(I)$  la lunghezza di  $I$ :  $\lambda(I) = \beta - \alpha$ . È facile verificare che

$$\beta \wedge b - \alpha \vee a = \lambda(I \cap (a, b)).$$

Quindi, per quanto visto sopra, se  $X \sim \text{Un}(a, b)$  allora per ogni intervallo  $I$

$$\mathbb{P}(X \in I) = \frac{\lambda(I \cap (a, b))}{\lambda(a, b)} \quad (1.8)$$

Tra le (1.7) e (1.8) ci sono molte somiglianze. Poiché la cardinalità di un intervallo non è finita (in generale), il rapporto in (1.7) non è definito se  $A$  ed  $E$  sono intervalli. Allora, nel caso di v.a. uniformi su un intervallo, quindi su un insieme che non è né finito né numerabile ma continuo, l'operazione  $\#$ =cardinalità è sostituita con  $\lambda$ =lunghezza (o misura).

Sia data una v.a. assolutamente continua, con funzione di distribuzione  $F_X$ . Se  $F_X$  è invertibile, il che accade ad esempio quando la densità  $p_X$  è positiva, per ogni  $\alpha \in (0, 1)$  esiste ed è unico  $q_\alpha \in \mathbb{R}$  tale che

$$\alpha = F_X(q_\alpha) = \mathbb{P}(X \leq q_\alpha).$$

$q_\alpha$  prende il nome di **quantile di ordine  $\alpha$  di  $X$** . In generale però  $F_X$  non è invertibile. Nel caso generale, si definisce quantile di ordine  $\alpha$  il più grande numero  $q_\alpha$  tale che

$$\alpha \geq F_X(q_\alpha) = \mathbb{P}(X \leq q_\alpha).$$

Si noti che se  $F_X$  è continua, o  $X$  è una v.a. continua, allora dalla proprietà 3. della Proposizione 1.4.2 si ha che  $F_X$  assume tutti i valori di  $(0, 1)$ , quindi esiste sempre una soluzione dell'equazione  $F_X(x) = \alpha$ . Ora, poiché  $F_X$  è non decrescente, segue che il quantile  $q_\alpha$  è la più grande di queste soluzioni. Se  $F_X$  è invertibile, tale soluzione è unica: ecco perché in questo caso  $q_\alpha = F_X^{-1}(\alpha)$ , come infatti avevamo detto all'inizio.

Vediamo ora brevemente come si estendono i concetti di media e varianza al caso di v.a. assolutamente continue.

Sia  $X$  una v.a. assolutamente continua, con densità  $p_X$ . Si dice che  $X$  ha media se

$$\text{esiste} \quad \int_{-\infty}^{+\infty} |x| p_X(x) dx \quad \text{finito}$$

e in tal caso si definisce **media di  $X$**  la quantità<sup>16</sup>

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

Si dice che  $X$  ha momento secondo se

$$\text{esiste } \int_{-\infty}^{+\infty} x^2 p_X(x) dx \quad \text{finito}$$

e in tal caso si definisce il **momento secondo di  $X$**  la quantità

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 p_X(x) dx.$$

Se  $X$  ha momento secondo, si mostra facilmente che anche la v.a.  $X - \mathbb{E}(X)$  ha momento secondo. In tal caso, si definisce la **varianza di  $X$**  la quantità

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Vale anche l'analogo risultato contenuto nel Teorema 1.2.1: se  $Y$  è una funzione di  $X$ ,  $Y = g(X)$ , allora  $Y = g(X)$  ha media se e solo se

$$\int_{-\infty}^{+\infty} |g(x)| p_X(x) dx < \infty$$

e in tal caso la media di  $Y = g(X)$  è

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) p_X(x) dx.$$

Va detto che tutte le proprietà della media e della varianza viste nel Paragrafo 1.2 rimangono valide anche per v.a. assolutamente continue. Si noti infatti (almeno da queste prime proprietà introdotte) che nel caso assolutamente continuo, le sommatorie presenti nel Paragrafo 1.2 vengono semplicemente sostituite da integrali: nei corsi più avanzati, si vedrà che non ci sono troppe differenze teoriche tra sommatorie ed integrali, nel senso che le sommatorie si possono semplicemente vedere come integrali opportuni (o meglio, integrali rispetto ad opportune misure, dette *di conteggio*).

La seguente tabella riassume densità, media e varianza di alcune v.a. assolutamente continue largamente utilizzate.

---

<sup>16</sup>In virtù della nota 12, possiamo brevemente dire che  $X$  ha media se la funzione  $\psi(x) = x p_X(x)$  è integrabile e in tal caso la media di  $X$  è l'integrale di  $\psi$ .

NOME	SIMBOLO	LEGGE	MEDIA	VARIANZA
Uniforme	$\text{Un}(a, b),$ $a < b, a, b \in \mathbb{R}$	$\frac{1}{b-a} \mathbb{I}_{(a,b)}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Esponenziale	$\text{Exp}(\lambda), \lambda > 0$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$\Gamma(\alpha, \lambda),$ $\alpha, \lambda > 0$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{I}_{x>0}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Normale	$\text{N}(m, \sigma^2),$ $m \in \mathbb{R}, \sigma^2 > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$m$	$\sigma^2$
Cauchy	–	$\frac{1}{\pi(1+x^2)}$	non esiste	non esiste

Ovviamente anche nel caso di v.a. continue è possibile introdurre il concetto di indipendenza, la cui definizione in effetti non dipende dal “tipo” di v.a. Sappiamo infatti che due v.a.  $X$  e  $Y$  sono indipendenti se

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

per ogni sottoinsieme “buono”  $A$  e  $B$  di  $\mathbb{R}$ , ad esempio  $A$  e  $B$  intervalli. Ciò che è meno banale è caratterizzare l’indipendenza in termini della densità (se esiste) delle due v.a. (analogamente a quanto fatto nel caso discreto nel Paragrafo 1.1). Sarà questo uno degli argomenti del prossimo corso di Probabilità, dove si dimostrerà anche il seguente risultato, qui solo enunciato (perché ne sarà fatto uso):

**Proposizione 1.4.3.** *Siano  $X$  ed  $Y$  due v.a. assolutamente continue, con densità  $p_X$  e  $p_Y$  rispettivamente. Se  $X$  ed  $Y$  sono indipendenti allora anche la v.a.  $X + Y$  ha densità data da:*

$$p_{X+Y}(z) = \int_{-\infty}^{+\infty} p_X(x) p_Y(z-x) dx = \int_{-\infty}^{+\infty} p_X(z-y) p_Y(y) dy.$$

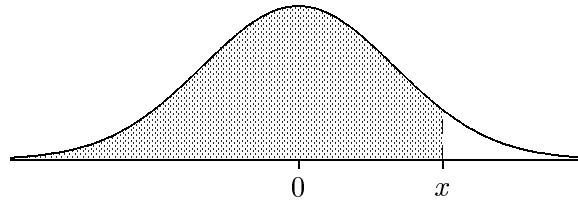
L’operazione sopra scritta è detta “prodotto di convoluzione”. Si confronti tale risultato con quanto avviene nel caso discreto, in particolare con la (1.2).

## 1.5 Il Teorema del Limite Centrale

Il Teorema del Limite Centrale è un altro caposaldo del Calcolo delle Probabilità. Prima di parlare di questo teorema, introduciamo brevemente la legge gaussiana.

Sia

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}, \quad t \in \mathbb{R},$$



**Figura 1.2** Grafico della densità gaussiana standard; l'area ombreggiata rappresenta la funzione di distribuzione  $\Phi(x) = \int_{-\infty}^x \phi(t) dt$ .

detta **densità gaussiana (o normale) standard**. Si noti che  $\phi > 0$  ed inoltre si può dimostrare che

$$\int_{-\infty}^{+\infty} \phi(t) dt = 1$$

(per la dimostrazione, occorrono alcune nozioni di integrazione [anche multipla] che si acquisiscono almeno al secondo anno; gli studenti più grandi, e comunque tutti gli interessati, troveranno una dimostrazione in Appendice 1.7). Dunque, esiste una v.a. assolutamente continua  $X$  con densità  $p_X = \phi$ .  $X$  prende il nome di **v.a. gaussiana (o normale) standard**, e si denota con  $N(0, 1)$ . La sua funzione di distribuzione si denota con il simbolo  $\Phi$ :

$$\Phi(x) \stackrel{\text{def}}{=} \int_{-\infty}^x \phi(t) dt,$$

nota in letteratura come la **distribuzione gaussiana (o normale) standard**. Purtroppo non è possibile determinare analiticamente una primitiva della funzione  $t \mapsto e^{-t^2/2}$ , quindi non è possibile scrivere in modo esplicito la funzione  $\Phi$ . In ogni caso, ogni libro di probabilità o statistica (a volte anche di analisi) contiene una tavola che dà molti valori di  $\Phi$  e da cui si possono leggere molti quantili. In Figura 1.2 è evidenziata la tipica forma a campana della densità gaussiana (“campana di Gauss”).

Valgono le seguenti proprietà:

**Proposizione 1.5.1.** *Sia  $X \sim N(0, 1)$ . Allora  $X$  ha media e varianza date da:*

$$\mathbb{E}(X) = 0 \quad \text{Var}(X) = 1.$$

**Dimostrazione.** Cominciamo a studiare la media. Per quanto visto nel paragrafo precedente, occorre dapprima dimostrare che la funzione  $\psi(t) = t\phi(t)$  è integrabile, cioè esiste finito

$$\int_{-\infty}^{+\infty} |\psi(t)| dt.$$

Poiché  $|\psi(t)| = |t| e^{-t^2/2} / \sqrt{2\pi}$  è una funzione pari, basta far vedere che esiste finito

$$\int_0^{+\infty} |\psi(t)| dt = \lim_{b \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_0^b t e^{-t^2/2} dt.$$

Ora,  $\int_0^b t e^{-t^2/2} dt = -e^{-t^2/2} \Big|_0^b = 1 - e^{-b^2/2}$ , dunque,

$$\lim_{b \rightarrow \infty} \int_0^b |\psi(t)| dt = \lim_{b \rightarrow \infty} \frac{1}{\sqrt{2\pi}} (1 - e^{-b^2/2}) = \frac{1}{\sqrt{2\pi}}.$$

Quindi, possiamo dire che la media esiste e vale

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t e^{-t^2/2} dt = \int_{-\infty}^0 t e^{-t^2/2} dt + \int_0^{+\infty} t e^{-t^2/2} dt.$$

Ora,  $\int_{-\infty}^0 t e^{-t^2/2} dt = -\int_0^{+\infty} t e^{-t^2/2} dt$  perché la funzione  $t \mapsto t e^{-t^2/2}$  è dispari, quindi

$$\mathbb{E}(X) = 0.$$

Passiamo ora alla varianza. Poiché la media è nulla,  $\text{Var}(X) = \mathbb{E}(X^2)$ , se esiste. Dunque, basta dimostrare che  $t \mapsto t^2 e^{-t^2/2} / \sqrt{2\pi}$  è integrabile e calcolarne l'integrale. Poiché trattasi di funzione pari, basta calcolare

$$\int_0^{+\infty} t^2 e^{-t^2/2} dt = \lim_{b \rightarrow \infty} \int_0^b t^2 e^{-t^2/2} dt.$$

Integrando per parti, si ottiene

$$\begin{aligned} \int_0^b t^2 e^{-t^2/2} dt &= \int_0^b t \cdot \frac{d}{dt}(-e^{-t^2/2}) dt \\ &= -t e^{-t^2/2} \Big|_0^b + \int_0^b e^{-t^2/2} dt = -b e^{-b^2/2} + \int_0^b e^{-t^2/2} dt. \end{aligned}$$

Ma allora,

$$\int_0^{+\infty} t^2 e^{-t^2/2} dt = \lim_{b \rightarrow \infty} \left( -b e^{-b^2/2} + \int_0^b e^{-t^2/2} dt \right) = \int_0^{+\infty} e^{-t^2/2} dt$$

quindi, tenendo conto che  $t \mapsto t^2 e^{-t^2/2}$  e  $t \mapsto e^{-t^2/2}$  sono funzioni pari,

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} t^2 e^{-t^2/2} dt &= 2 \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} t^2 e^{-t^2/2} dt \\ &= 2 \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \int_{-\infty}^{+\infty} \phi(t) dt = 1. \end{aligned}$$

Quindi,  $\mathbb{E}(X^2)$  esiste e  $\mathbb{E}(X^2) = 1$ , da cui segue che  $\text{Var}(X) = 1$

□

La proposizione precedente dà quindi il significato dei parametri 0 e 1.

In generale, la densità gaussiana (non-standard) è definita a partire da due numeri  $m \in \mathbb{R}$  e  $\sigma^2 > 0$ :

$$\phi_{m,\sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-m)^2/(2\sigma^2)}, \quad t \in \mathbb{R}$$

e si denota con il simbolo  $N(m, \sigma^2)$ . Si noti che per  $m = 0$  e  $\sigma^2 = 1$  torna il caso normale-standard. Vediamo qual è la relazione con le gaussiane standard e qual è il significato dei parametri  $m$  e  $\sigma^2$ :

**Proposizione 1.5.2.** *Sia  $Z \sim N(0, 1)$  e, per  $m \in \mathbb{R}$  e  $\sigma \neq 0$ ,*

$$X = m + \sigma Z.$$

Allora  $X \sim N(m, \sigma^2)$  ed inoltre

$$\mathbb{E}(X) = m \quad \text{Var}(X) = \sigma^2.$$

**Dimostrazione.** Ricordiamo che se esiste  $F'_X$  allora  $p_X = F'_X$ : useremo questa proprietà per calcolare la densità di  $X$ . Si ha

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(m + \sigma Z \leq x).$$

Supponiamo  $\sigma > 0$ : in tal caso

$$F_X(x) = \mathbb{P}\left(Z \leq \frac{x-m}{\sigma}\right) = \Phi\left(\frac{x-m}{\sigma}\right);$$

se invece  $\sigma < 0$ , allora

$$F_X(x) = \mathbb{P}\left(Z \geq \frac{x-m}{\sigma}\right) = 1 - \mathbb{P}\left(Z < \frac{x-m}{\sigma}\right) = 1 - \Phi\left(\frac{x-m}{\sigma}\right).$$

Usando le proprietà di derivabilità delle funzioni composte, possiamo compattare i due casi scrivendo:

$$\frac{d}{dx}F_X(x) = \Phi'\left(\frac{x-m}{\sigma}\right) \cdot \frac{1}{|\sigma|},$$

dove  $\Phi'$  denota la funzione derivata di  $\Phi$ , quindi  $\Phi' = \phi$ . Allora

$$p_X(x) = \frac{1}{|\sigma|} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \Big|_{t=(x-m)/\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)},$$

quindi  $X \sim N(m, \sigma^2)$ . Per quanto riguarda media e varianza, poiché  $X$  è una funzione lineare della v.a.  $Z$ , che ha media e varianza, ha essa stessa media e varianza, date da:

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(m + \sigma Z) = m + \sigma \mathbb{E}(Z) = m, \\ \text{Var}(X) &= \text{Var}(m + \sigma Z) = \text{Var}(\sigma Z) = \sigma^2 \text{Var}(Z) = \sigma^2. \end{aligned}$$

□

Dunque, i parametri  $m$  e  $\sigma^2$  di una v.a.  $X \sim N(m, \sigma^2)$  rappresentano, rispettivamente, la media e la varianza di  $X$ .

La legge gaussiana riveste un ruolo fondamentale nel Calcolo delle Probabilità. Ad esempio, vale il seguente teorema, noto con il nome di **Teorema del Limite Centrale (TLC)**:

**Teorema 1.5.3.** *Siano  $X_1, X_2, \dots$  v.a. indipendenti ed identicamente distribuite, dotate di momento secondo. Siano  $m$  e  $\sigma^2$  rispettivamente la media e la varianza comune. Sia  $\delta > 0$ . Allora: per ogni intervallo  $I$  di  $\mathbb{R}$  di estremi  $a$  e  $b$ , si ha*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{\sum_{k=1}^n X_k - nm}{\sqrt{n\sigma^2}} \in I \right) = \Phi(b) - \Phi(a) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

*In altre parole, si dice che  $\frac{\sum_{k=1}^n X_k - nm}{\sqrt{n\sigma^2}}$  “converge in legge” ad una v.a. gaussiana standard, nel senso che*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{\sum_{k=1}^n X_k - nm}{\sqrt{n\sigma^2}} \in I \right) = \mathbb{P}(Z \in I), \quad \text{con } Z \sim N(0, 1).$$

La dimostrazione del Teorema del Limite Centrale è assolutamente non banale e prevede la conoscenza di alcuni strumenti di probabilità (ed analisi) piuttosto avanzati. Per tale ragione la tralasciamo (sarà però senz'altro proposta nel secondo corso di probabilità; rimandiamo comunque al Teorema 4.14 in [1]).

Il Teorema del Limite Centrale ha numerose applicazioni nel Calcolo delle Probabilità e nella Statistica. Discuteremo qui l'aspetto che ci interessa di più: come il Teorema del Limite Centrale dia indicazioni sulla velocità di convergenza nella Legge dei Grandi Numeri

Supponiamo di avere una successione numerica  $\{z_n\}_n \subset \mathbb{R}$  convergente a 0. Potrebbe essere (lo è!) interessante capire con quale velocità  $z_n$  tende a 0. Ma: che significato dare alla nozione di “velocità di convergenza”? Un modo potrebbe essere quello di confrontare  $\{z_n\}_n$  con una successione ben nota: ad esempio, con la successione  $1/n^\gamma$ , che sappiamo convergere a 0 se  $\gamma > 0$ . Allora, con  $\gamma > 0$ , diremo che

1.  $z_n$  va a 0 più velocemente di  $1/n^\gamma$ , e scriveremo  $z_n = o(1/n^\gamma)$ , se

$$\limsup_{n \rightarrow \infty} n^\gamma z_n = 0;$$

2.  $z_n$  va a 0 come  $1/n^\gamma$ , e scriveremo  $z_n = O(1/n^\gamma)$ , se esiste (finito)

$$\lim_{n \rightarrow \infty} n^\gamma z_n \neq 0;$$

questo limite però potrebbe non esistere e in tal caso, generalizzando, si dice che  $z_n = O(1/n^\gamma)$  se esistono due costanti positive  $c_\gamma$  e  $C_\gamma$  tali che

$$c_\gamma \leq \liminf_{n \rightarrow \infty} n^\gamma z_n \leq \limsup_{n \rightarrow \infty} n^\gamma z_n \leq C_\gamma,$$

(il che, in effetti, è vero se in particolare esiste finito e non nullo  $\lim_{n \rightarrow \infty} n^\gamma z_n$ );

3.  $z_n$  va a 0 più lentamente di  $1/n^\gamma$  se

$$\liminf_{n \rightarrow \infty} n^\gamma z_n = +\infty.$$

Ad esempio,  $z_n = e^{-n} = o(1/n^\gamma)$  per ogni  $\gamma > 0$ ;  $z_n = \sin^2(1/n) = o(1/n)$  ed anzi,  $\sin^2(1/n) = O(1/n^2)$ , così come  $\sin^2(1/n)$  va a 0 più lentamente di  $1/n^3$ .

Questo accade per le successioni numeriche. Vediamo ora come dare un senso al concetto di “velocità di convergenza” nella Legge dei Grandi Numeri, in cui abbiamo visto che la convergenza coinvolta è diversa da quella numerica. In questo caso, infatti, la successione di cui si studia la convergenza è data dalla media aritmetica  $\frac{1}{n} \sum_{k=0}^n X_k$ , dove le  $X_k$  sono v.a. dunque anche la media aritmetica è una v.a.

Intanto, cominciamo con l’osservare che la v.a. coinvolta nel Teorema del Limite Centrale

$$Y_n = \frac{\sum_{k=1}^n X_k - nm}{\sqrt{n\sigma^2}}$$

si può riscrivere:

$$Y_n = \sqrt{\frac{n}{\sigma^2}} \cdot \left( \frac{1}{n} \sum_{k=1}^n X_k - m \right).$$

Osserviamo che, usando anche l’invarianza per traslazioni delle varianze,

$$\text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k - m\right) = \text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{\sigma^2}{n},$$

quindi posto  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ , si ha

$$Y_n = \frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sqrt{\text{Var}(\bar{X}_n)}},$$



cioè  $Y_n$  è la v.a. che si ottiene **standardizzando**<sup>17</sup> la media empirica  $\bar{X}_n$ . Ora, la LGN assicura che (in qualche senso)

$$Z_n = \frac{1}{n} \sum_{k=1}^n X_k - m \rightarrow 0, \quad \text{se } n \rightarrow \infty$$

e il TLC che

$$\mathbb{P}\left(\sqrt{\frac{n}{\sigma^2}} Z_n \in (a, b)\right)$$

converge per  $n \rightarrow \infty$ , mantenendosi quindi limitata. Ci troviamo quindi in una condizione analoga a quella vista nel punto 2. della definizione precedente: in questo caso ciò che è limitato non è esattamente  $\sqrt{\frac{n}{\sigma^2}} Z_n$  ma è la probabilità che tale v.a. appartenga ad un qualsiasi intervallo  $(a, b)$ . Per tale ragione, possiamo dire che

$$Z_n = \frac{1}{n} \sum_{k=1}^n X_k - m \text{ "converge" a } 0 \text{ per } n \rightarrow \infty \text{ con "velocità" } \sqrt{\frac{\sigma^2}{n}},$$

e quindi:

$$Z_n = \frac{1}{n} \sum_{k=1}^n X_k \text{ "converge" a } m \text{ per } n \rightarrow \infty \text{ con "velocità" } \sqrt{\frac{\sigma^2}{n}}.$$

È anche evidente che al diminuire della varianza  $\sigma^2$  tale forma di convergenza diviene più veloce.

Vedremo poi come utilizzare tali nozioni per le simulazioni e nei **metodi Monte Carlo**.

---

<sup>17</sup>In generale, data una v.a.  $Z$  con media  $\mu$  e varianza  $s^2$ , la v.a. standard associata a  $Z$  è  $(Z - \mu)/\sqrt{s^2}$ . Si noti che tale v.a. ha media 0 e varianza 1.

## 1.6 Appendice: retta di regressione ed interpretazione della covarianza

Siano  $X$  e  $Y$  due v.a. sullo stesso spazio di probabilità  $(\Omega, \mathcal{F}, \mathbb{P})$ . Si vuole trovare la migliore approssimazione possibile della  $Y$  tramite una funzione lineare della  $X$ . In altre parole, si cercano due numeri  $a, b \in \mathbb{R}$  tali che  $aX + b$  approssima  $Y$  nel miglior modo possibile. Un problema di questo tipo appare, ad esempio, quando si vuole stimare una quantità  $Y$  che non è osservabile direttamente. Si consideri, a titolo di esempio, il caso in cui  $Y$  è un segnale e  $X$  è un'osservazione "disturbata" (vale a dire,  $X$  è uguale a  $Y$  più un errore dovuto all'osservazione).

Il primo problema consiste nel definire un criterio di ottimalità: che significato si vuole dare a "migliore approssimazione possibile"? Ovviamente, occorre partire dalla distanza tra  $aX + b$  e  $Y$ , ovvero da  $|aX + b - Y|$ . Poiché però lavorare con i moduli può essere oneroso in termini di calcoli, si preferisce considerare  $(aX + b - Y)^2$ . Ora, tale quantità è aleatoria e il problema che è stato posto è "deterministico": i numeri  $a$  e  $b$  che si cercano *non* si vogliono aleatori. Per tale ragione, si considera il valore atteso di  $(aX + b - Y)^2$ . Dunque, il criterio di ottimalità scelto è il seguente: si vogliono determinare due numeri  $a$  e  $b$  tali da minimizzare<sup>18</sup>

$$(a, b) \mapsto g(a, b) \stackrel{def}{=} \mathbb{E}[(aX + b - Y)^2].$$

Ora, è facile dimostrare che<sup>19</sup> tale minimo è assunto in  $(a^*, b^*)$ , dove

$$a^* = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \quad \text{e} \quad b^* = \mathbb{E}[Y] - a^* \mathbb{E}[X]$$

La retta di equazione  $y = a^*x + b^*$ , ovvero

$$y = \varphi(x) \equiv \frac{\text{Cov}(X, Y)}{\text{Var}(X)}(x - \mathbb{E}[X]) + \mathbb{E}[Y],$$

è detta **retta di regressione**. Dunque, in generale non è detto che  $Y = \varphi(X)$  ma ciò che è vero è che la funzione  $\varphi$  è la funzione lineare che minimizza la distanza (vista sopra) tra  $Y$  e  $\varphi(X)$ .

Si noti che la retta di regressione non è "simmetrica, ovvero scambiando ruolo a  $X$  e  $Y$  cambia anche la retta di regressione. Inoltre, il coefficiente angolare della retta di regressione ha lo stesso segno della covarianza (perché

---

<sup>18</sup>Va detto che esistono altri criteri di ottimizzazione, che qui non saranno presi in considerazione. Si noti che ciò che si vuole minimizzare è la distanza in  $L^2$  tra la v.a.  $Y$  e la v.a.  $aX + b$ .

<sup>19</sup>Occorrono alcune conoscenze di base di calcolo di minimo e/o massimo di funzioni di più variabili.

$\text{Var}(X)$  è sempre positiva). Ciò indica che, se la covarianza è positiva,  $Y$  tende ad assumere valori grandi in corrispondenza di valori grandi della  $X$ . Tale proprietà è detta di *dipendenza positiva*. Al contrario, se la covarianza è negativa,  $Y$  tende ad assumere valori piccoli in corrispondenza di valori grandi della  $X$  e in tal caso si dice che c'è *dipendenza negativa* tra le due v.a. Ad esempio, se si volesse creare un modello che legghi la relazione tra il peso  $X$  e l'altezza  $Y$  di un individuo preso a caso da una popolazione, ci si aspetta una covarianza positiva tra queste due v.a. perché individui alti hanno la tendenza ad essere più pesanti (ciò non toglie che possano esistere individui alti e leggeri oppure bassi e molto pesanti, eventi che, in generale, hanno probabilità di verificarsi più bassa). Ci si attende una covarianza negativa quando si considerino quantità aleatorie che hanno un effetto "antitetico l'una rispetto all'altra.

Infine, se la covarianza è nulla (come avviene quando le due v.a. sono indipendenti), la retta di correlazione è costante, ovvero l'unica funzione lineare  $\varphi$  che minimizza la distanza tra  $Y$  e  $\varphi(X)$  non dipende da  $X$ . Siamo quindi in presenza di una "forma debole di indipendenza, e in tal caso le v.a.  $X$  e  $Y$  sono dette *non correlate* (ma attenzione: non è detto che siano indipendenti nel senso classico!).

Abbiamo appena visto come la covarianza possa essere utilizzata per misurare la dipendenza di due v.a. In realtà ciò non è del tutto vero perché la covarianza presenta un difetto:

$$\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$$

Ciò significa che la covarianza tra due v.a. è strettamente legata all'unità di misura: la covarianza cambia se si cambia l'unità di misura con la quale le v.a.  $X$  e  $Y$  sono state misurate. Per tale ragione, si introduce la seguente quantità:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

$\rho_{X,Y}$  è detto **coefficiente di correlazione di  $X$  e  $Y$**  e non dipende dai cambiamenti di misurazione, cioè (esercizio!)

$$\rho_{aX,bY} = \rho_{X,Y}$$

per ogni<sup>20</sup>  $a, b$  tali che  $ab > 0$ . Inoltre, abbiamo visto che  $\text{Cov}(X, Y)^2 \leq \text{Var}(X)\text{Var}(Y)$ , quindi  $-1 \leq \rho_{X,Y} \leq 1$ .

Riassumendo, la condizione di non correlazione è in realtà molto più debole di quella di indipendenza, ma è anche più facile da verificare per cui è abbastanza usata nella pratica come forma debole di indipendenza.

<sup>20</sup>Se  $ab < 0$ , il coefficiente di correlazione cambia segno: esercizio! Tale proprietà è intuitiva se si ricorda la relazione che intercorre tra il segno del coefficiente di correlazione e la proprietà di dipendenza positiva o negativa tra le due v.a. in questione.

## 1.7 Appendice: complementi sulla legge gaussiana.

**Lemma 1.7.1.** *Si ha:*

$$\text{esiste } \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1.$$

**Dimostrazione.** Mostriamo che  $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$  è integrabile su  $\mathbb{R}$ . Osserviamo che se  $|t| > 1$  allora  $t^2 > |t|$ , quindi possiamo scrivere

$$0 \leq \phi(t) \leq \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathbb{I}_{|t| \leq 1} + \frac{1}{\sqrt{2\pi}} e^{-|t|/2} \mathbb{I}_{|t| > 1}.$$

Allora, basta dimostrare che la funzione a destra è integrabile su  $\mathbb{R}$ , anzi: che la funzione  $\psi(t) = \frac{1}{\sqrt{2\pi}} e^{-|t|/2} \mathbb{I}_{|t| > 1}$  è integrabile su  $\mathbb{R}$ . Poniamo allora

$$\psi_n(t) = \frac{1}{\sqrt{2\pi}} e^{-|t|/2} \mathbb{I}_{1 < |t| < n}.$$

$\psi_n$  è integrabile su  $\mathbb{R}$  e  $\psi_n \uparrow \psi$  per  $n \rightarrow \infty$ , quindi per il teorema di convergenza monotona possiamo scrivere

$$\int_{\mathbb{R}} \psi(t) dt = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \psi_n(t) dt = \lim_{n \rightarrow \infty} 2 \int_1^n \frac{1}{\sqrt{2\pi}} e^{-t/2} dt = \frac{4}{\sqrt{2\pi}} e^{-1/2}.$$

Segue allora che  $\psi$  è integrabile su  $\mathbb{R}$  e quindi, per quanto visto sopra, anche  $\phi$  è integrabile su  $\mathbb{R}$ . Calcoliamo ora  $\int_{\mathbb{R}} \phi(t) dt$ . Dal teorema di Fubini possiamo scrivere

$$\begin{aligned} \left( \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \right)^2 &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds \\ &= \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{-(t^2+s^2)/2} dt ds. \end{aligned}$$

Passando in coordinate polari ( $t = \rho \cos \theta$ ,  $s = \rho \sin \theta$ ) ed usando il teorema del cambio di variabile, si ha<sup>21</sup>

$$\begin{aligned} \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{-(t^2+s^2)/2} dt ds &= \frac{1}{2\pi} \int_0^{+\infty} d\rho \int_0^{2\pi} d\theta \rho e^{-\rho^2/2} \\ &= \frac{1}{2\pi} 2\pi \left[ -e^{-\rho^2/2} \right]_0^{+\infty} = 1. \end{aligned}$$

Quindi  $(\int_{\mathbb{R}} \phi(t) dt)^2 = 1$  e poiché  $\phi > 0$ , dev'essere  $\int_{\mathbb{R}} \phi(t) dt = 1$ .

□

---

<sup>21</sup>L'integrabilità della funzione  $\rho \mapsto \rho e^{-\rho^2/2}$  su  $[0, +\infty)$  si dimostra con argomenti analoghi a quelli usati sopra.

## Capitolo 2

# Simulazione di variabili aleatorie discrete

In questo capitolo studieremo la simulazione di alcune delle principali variabili aleatorie discrete, con applicazioni al calcolo numerico di quantità (deterministiche) con metodi di simulazione, ad esempio per il calcolo numerico di aree o integrali. Tali metodi prendono il nome di **metodi Monte Carlo**.

Fondamentalmente in questo capitolo saranno proposti problemi guidati ed alcuni esercizi. I primi mostreranno (e dimostreranno) anche gli algoritmi di simulazione delle variabili aleatorie di cui ci stiamo per occupare; i secondi sono lasciati al lettore, e ovviamente richiedono la conoscenza e la comprensione degli altri (un esercizio qui si intende risolto se è anche implementato al calcolatore). Anticipiamo che gli esempi di codici proposti in queste note sono scritti in linguaggio Pascal.

Per la simulazione, vedremo che uno strumento fondamentale è la possibilità di generare variabili aleatorie su  $(0, 1)$ . Molti linguaggi di programmazione hanno in libreria una procedura che genera v.a.  $Un(0, 1)$ . Ad esempio, la procedura Pascal si chiama **Random**: il comando

```
Z:=Random
```

assegna in Z un numero casuale su  $(0, 1)$ . Qui e nel seguito, indicheremo sempre con **nrandom** una procedura che genera un numero a caso in  $(0, 1)$ : con il comando

```
Z:=nrandom
```

intenderemo sempre assegnare in Z un numero casuale in  $(0, 1)$ . Ad esempio, una possibile procedura per simulare un numero (in doppia precisione) a caso in  $(0, 1)$  in linguaggio C è:

```
double(r);
srand48( (unsigned)time(NULL) );
r=drand48( );
```

(senza dimenticare di includere le librerie necessarie, ad esempio `<TIME.H>`).

Sono possibili ovviamente anche altre procedure di generazione (per eventuali suggerimenti, si rimanda al sito <http://random.mat.sbg.ac.at/links>). Infine, è anche possibile disegnare grafici, ad esempio usando GNUPLLOT (per informazioni: <http://www.duke.edu/~hpgavin/gnuplot.html>).

## 2.1 Schema di Bernoulli e conseguenze

### Problema 2.1.1. Generazione di uno schema di Bernoulli.

**a)** Sia  $Z \sim \text{Un}(0, 1)$  e  $p \in [0, 1]$  fissato. Dimostrare che  $X = \mathbb{I}_{[0,p]}(Z) \sim \text{Be}(p)$ .

**b)** Usando **a)**, scrivere una procedura per simulare una v.a.  $X \sim \text{Be}(p)$ , con  $p \in [0, 1]$  dato in input.

**c)** Sia  $N$  un intero positivo fissato (ad es.  $N = 10000$ ). Scrivere un programma che, generando  $N$  v.a. i.i.d. bernoulliane (schema di Bernoulli) calcoli la media aritmetica delle  $N$  simulazioni, ovvero

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$$

e si confronti il valore osservato con il valor medio  $\mathbb{E}[X] = p$  previsto teoricamente calcolando

$$\text{errmedia} = |\bar{X}_N - \mathbb{E}[X]|.$$

**Soluzione.** **a)** Fissato  $p \in [0, 1]$ , se  $Z \sim \text{Un}(0, 1)$  allora

$$\begin{aligned} \mathbb{P}(X = 1) &= \mathbb{P}(Z < p) = \lambda([0, p]) = p \\ \mathbb{P}(X = 0) &= \mathbb{P}(Z \geq p) = \lambda([p, 1]) = 1 - p, \end{aligned}$$

quindi  $X \sim \text{Be}(p)$ .

**b)** Per simulare una v.a.  $X$  bernoulliana di parametro  $p$ , generiamo una v.a.  $Z \sim \text{Un}(0, 1)$  e poi poniamo  $X = \mathbb{I}_{Z < p}$ :

```
Z:=nrandom
if Z < p then X:=1 else X:=0
```

dove `nrandom` definisce un generatore di numeri casuali su  $(0, 1)$ . Indicheremo con `bernoulli(p)` una funzione che prendendo in *input* il valore  $p \in (0, 1)$ , genera una v.a. bernoulliana di parametro  $p = \mathbf{p}$ .

**c)** Per generare uno schema di Bernoulli di lunghezza  $N$  basta richiamare la procedura `bernoulli(p)`  $N$  volte. Se indichiamo con `Xmedio` la media aritmetica, la quantità `errmedia` si può calcolare come segue:

```
Xmedio:=0;
for i:=1 to N do Xmedio=Xmedio+bernoulli(p);
Xmedio:=Xmedio/N;
errmedia:=abs(Xmedio-p)
```

dove `abs` denota una funzione che restituisce il valore assoluto dell'argomento.

◇

**Osservazioni.** Dalla LGN, sappiamo che  $\bar{X}_n$  converge a  $\mathbb{E}(X_1) = p$  per  $n \rightarrow \infty$ , quindi ci si deve aspettare che la quantità *errmedia* nel problema precedente (ma anche nei successivi) sia piuttosto piccola (questo fatto si può prendere come *test* per verificare che il programma dia un risultato corretto).

### Problema 2.1.2. Generazione di v.a. binomiali.

**a)** Siano  $Z_1, \dots, Z_n$  indipendenti, tutte di legge  $\text{Be}(p)$ . Dimostrare che  $X = \sum_{i=1}^n Z_i$  è una v.a. di legge  $\text{Bi}(n, p)$  (una binomiale è il numero di successi su  $n$  prove bernoulliane).

**b)** Scrivere una procedura per simulare una v.a.  $X \sim \text{Bi}(n, p)$ , con  $n$  e  $p$  dati in *input*.

**c)** Fissato un intero positivo  $N$  (ad es.  $N = 10000$ ), scrivere un programma che, generando  $N$  v.a.  $X_1, \dots, X_N$  indipendenti di legge  $\text{Bi}(n, p)$  (senza memorizzarle tutte!), confronti le frequenze relative con cui si presentano i valori  $\{0, 1, \dots, n\}$  nelle  $N$  simulazioni con le rispettive probabilità teoriche<sup>1</sup>  $\{p_0, p_1, \dots, p_n\}$  calcolando

$$errmax = \max_{0 \leq i \leq n} |f_i - p_i|$$

---

<sup>1</sup>Si ricorda che  $p_i = \binom{n}{i} p^i (1-p)^{n-i} \equiv \binom{n}{i} \exp(i \ln p + (n-i) \ln(1-p))$ ; per il calcolo dei coefficienti binomiali, potrebbe essere utile osservare che

$$\binom{n}{0} = 1, \quad \binom{n}{i} = \binom{n}{i-1} \frac{n-i+1}{i}, \quad i = 1, 2, \dots, n.$$

dove  $f_i$  è la frequenza relativa di  $i$ .

**d)** Aggiungere al programma il calcolo della media aritmetica delle  $N$  simulazioni, ovvero

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$$

e confrontare il valore osservato con il valor medio  $\mathbb{E}[X] = np$  previsto teoricamente calcolando

$$errmedia = |\bar{X}_N - \mathbb{E}[X]|.$$

**e)** Determinare un valore di  $N$  affinché con probabilità maggiore di 0.99,  $\bar{X}_N$  disti dalla media teorica  $np$  per meno di 0.01. Procedere al calcolo di  $errmedia$  con questo valore per  $N$ .

**Soluzione.** **a)** Siano  $Z_1, \dots, Z_n$   $n$  v.a. i.i.d.  $\text{Be}(p)$ . Allora

$$X = \sum_{i=1}^n Z_i$$

“conta” il numero di v.a.  $Z_i$  che assumono il valore 1 (se si dà ad 1 il significato di *successo*,  $X$  è il *numero di successi su  $n$  prove*). Quindi  $E_X = \{0, 1, \dots, n\}$ . Fissato  $k \in E_X$ , per il momento  $0 < k < n$ ,

$$\begin{aligned} p_k &= \mathbb{P}(X = k) = \mathbb{P}(\#\{Z_i : Z_i = 1\} = k) \\ &= \mathbb{P}(\cup_{I \in \mathcal{I}_k} \{Z_i = 1, i \in I, Z_j = 0, j \notin I\}) \end{aligned}$$

dove  $\mathcal{I}_k$  denota tutti i sottoinsiemi di  $\{1, \dots, n\}$  di cardinalità  $k$ . Poiché l’unione sopra scritta è di insiemi disgiunti, si ha

$$p_k = \sum_{I \in \mathcal{I}_k} \mathbb{P}(Z_i = 1, i \in I, Z_j = 0, j \notin I)$$

e poiché le v.a.  $Z_i$  sono identicamente distribuite ed indipendenti,

$$\begin{aligned} p_k &= \sum_{I \in \mathcal{I}_k} \mathbb{P}(Z_1 = 1, \dots, Z_k = 1, Z_{k+1} = 0, \dots, Z_n = 0) \\ &= \#\mathcal{I}_k \cdot \prod_{i=1}^k \mathbb{P}(Z_i = 1) \prod_{i=k+1}^n \mathbb{P}(Z_i = 0) = \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Ovviamente,  $p_0 = \mathbb{P}(Z_1 = 0, \dots, Z_n = 0) = \prod_{i=1}^n \mathbb{P}(Z_i = 0) = (1-p)^n$  e  $p_n = \mathbb{P}(Z_1 = 1, \dots, Z_n = 1) = \prod_{i=1}^n \mathbb{P}(Z_i = 1) = p^n$ , da cui segue che  $X \sim \text{Bi}(n, p)$ .

**b)** Se  $X \sim \text{Bi}(n, p)$  allora  $X$  si può rappresentare come la somma di  $n$  v.a. i.i.d. bernoulliane di parametro  $p$ :



```
X:=0;
for i:=1 to n do X:=X+bernoulli(p)
```

Indicheremo con `binomiale(n,p)` una funzione che prendendo in *input* un naturale  $n$  e un numero  $p \in (0, 1)$ , genera una v.a. binomiale di parametri  $n = n$  e  $p = p$ .

c) Indichiamo con `fr` un vettore  $(n + 1)$ -dimensionale la cui generica componente `fr[k]` denoterà la frequenza relativa osservata del valore  $k$ , per  $k = 0, 1, \dots, n$ . Inizialmente `fr` dà il vettore delle frequenze assolute; normalizzando si ottengono le frequenze relative. Nel seguito, la variabile `NN` denota<sup>2</sup> il numero di simulazioni  $N$ :

```
for k:=0 to n do fr[k]:=0;
for i:=1 to NN do
  begin
    X:=binomiale(n,p);
    fr[X]:=fr[X]+1
  end;
for k:=0 to n do fr[k]:=fr[k]/NN
```

(si noti che il comando `fr[X]:=fr[X]+1` ha senso se la variabile `X` è dichiarata di tipo `Integer`, quindi l'*output* della funzione `binomiale(n,p)` dev'essere di tipo intero.)

Ora `fr` dà il vettore delle frequenze (relative) osservate, che devono essere confrontate con il vettore delle probabilità teoriche. Tale vettore va quindi calcolato, ad esempio come segue: sia `pr` un vettore  $(n + 1)$ -dimensionale, tale che `pr[k]` dà la probabilità  $p_k$ . Ora, alcuni linguaggi di programmazione non consentono di effettuare elevamenti a potenza ma possono calcolare esponenziali (`exp`) e logaritmi naturali (`ln`), quindi usiamo la proprietà

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \exp(k \ln p + (n-k) \ln(1-p)).$$

Inoltre, il coefficiente binomiale può essere facilmente calcolato in modo iterativo usando la proprietà

$$\binom{n}{0} = 1, \quad \binom{n}{k} = \binom{n}{k-1} \frac{n-k+1}{k}.$$

Denotando con `cbin` il coefficiente binomiale, possiamo scrivere:

---

<sup>2</sup>Alcuni linguaggi di programmazione non distinguono le minuscole dalle maiuscole; poiché la variabile `n` è già impegnata (e denota il parametro  $n$  della binomiale da studiare), per evitare problemi meglio non usare `N` per denotare il numero  $N$  di simulazioni da effettuare.

```

cbin:=1;
for k:=0 to n do
  begin
    pr[k]:=cbin*exp(k*ln(p)+(n-k)*ln(1-p));
    cbin:=cbin*(n-k+1)/k
  end;

```

Indichiamo con `distrbin(n,p)` una funzione che prendendo in *input* un naturale  $n$  e un numero  $p \in (0, 1)$ , restituisce la distribuzione `pr` di una v.a. binomiale di parametri  $n = n$  e  $p = p$ .

Siamo ora pronti per calcolare *errmax*:

```

pr:=distrbin(n,p);
errmax:=abs(pr[0]-fr[0]);
for k:=1 to n do
  if (abs(pr[k]-fr[k])>errmax) then errmax:=abs(pr[k]-fr[k])

```

c) Basta inserire nelle righe del programma in cui vengono simulate le  $N$  v.a. anche il calcolo della somma dei risultati: il programma in questione diviene

```

for k:=0 to n do fr[k]:=0;

```

```

Xmedio:=0;

```

```

for i:=1 to NN do
  begin
    X:=binomiale(n,p);
    fr[X]:=fr[X]+1;

```

```

Xmedio:=Xmedio+X

```

```

end;
for k:=0 to n do fr[k]:=fr[k]/NN;
pr:=distrbin(n,p);
errmax:=abs(pr[0]-fr[0]);
for k:=1 to n do
  if (abs(pr[k]-fr[k])>errmax) then errmax:=abs(pr[k]-fr[k]);

```

```

Xmedio:=Xmedio/NN;
errmedia:=abs(Xmedio-n*p)

```

(i riquadri sottolineano i punti in cui sono state effettuate le aggiunte che riguardano  $\bar{X}_N$ ).

e) Ricordando che se  $X \sim \text{Bi}(n, p)$  allora  $\text{Var}(X) = np(1 - p)$ , la disuguaglianza di Chebycev assicura che

$$\mathbb{P}(|\bar{X}_N - np| \geq \delta) \leq \frac{np(1-p)}{\delta^2 N}$$

o equivalentemente

$$\mathbb{P}(|\bar{X}_N - np| < \delta) \geq 1 - \frac{np(1-p)}{\delta^2 N}.$$

Quindi, scelto  $\delta = 0.01$ ,  $N$  verifica la condizione richiesta se

$$1 - \frac{np(1-p)}{0.01^2 N} \geq 0.99,$$

il che dà

$$N \geq np(1-p) \cdot 10^6.$$

Scegliamo quindi il più piccolo  $N$  che verifica questa proprietà:  $N_0 = \lceil np(1-p) \cdot 10^4 \rceil + 1$ . Simuliamo allora  $N_0$  v.a. binomiali, seguendo la procedura utilizzata nel punto precedente con NN sostituito da

`NN0:=Int(n*p*(1-p)*10000)+1`

dove `Int` denota una funzione che calcola la parte intera di un numero reale.

◇

**Osservazioni.** Così come osservato nel Problema 2.1.1, anche qui *errmedia* dev'essere piccolo. Ma ci si deve aspettare che anche *errmax* produca un numero piccolo. La motivazione segue sempre dalla LGN. Infatti, fissiamo  $i \in E_X$ . La frequenza relativa di  $i$  nelle  $N$  osservazioni è la v.a.

$$f_i^{(N)} = \frac{1}{N} \sum_{k=1}^N \mathbb{I}_{X_k=i}.$$

Ora, le v.a.  $\{\mathbb{I}_{X_k=i}\}_k$  sono i.i.d., di legge bernoulliana con parametro dato da

$$\mathbb{E}(\mathbb{I}_{X_k=i}) = \mathbb{P}(X = i) = p_i.$$

Ma allora, usando la LGN possiamo dire che, per  $N \rightarrow \infty$ ,  $f_i^{(N)}$  converge a  $p_i$ . Quindi,  $|f_i^{(N)} - p_i|$  converge a 0 per ogni  $i \in E_X$ . Ora, essendo  $E_X$  finito, è abbastanza semplice verificare che anche  $\text{errmax}^{(N)} = \max_{i \in E_X} |f_i^{(N)} - p_i|$  tende a 0 quando  $N \rightarrow \infty$ , il che spiega perché se il numero  $N$  di simulazioni è alto ci si deve aspettare un *errmax* piccolo.

**Problema 2.1.3. Generazione di v.a. geometriche.**

a) Siano  $Z_1, \dots, Z_n$  indipendenti, tutte di legge  $\text{Be}(p)$ ,  $p \in (0, 1)$ . Posto

$$T = \inf\{n \geq 1 ; Z_n = 1\} - 1,$$

dimostrare che  $T \sim \text{Ge}(p)$  (una geometrica è l'istante di primo successo -1 in prove bernoulliane ripetute).

b) Scrivere una procedura per simulare una v.a.  $T \sim \text{Ge}(p)$ .

c) Fissato un intero positivo  $N$  (ad es.  $N = 10000$ ), scrivere un programma che, generando  $N$  v.a.  $T_1, \dots, T_N$  indipendenti e geometriche di parametro  $p$  (senza memorizzarle!), calcoli la media aritmetica delle  $N$  simulazioni, ovvero

$$\bar{T}_N = \frac{1}{N} \sum_{i=1}^N T_i.$$

Confrontare il valore osservato con il valor medio  $\mathbb{E}[T] = (1-p)/p$  previsto teoricamente calcolando

$$\text{errmedia} = |\bar{T}_N - \mathbb{E}[T]|.$$

d) Determinare analiticamente un valore  $N_0$  per  $N$  affinché

$$\mathbb{P}(\text{errmedia} < 0.1) \geq 0.99$$

e verificare empiricamente la validità di questa stima.

e) Determinare numericamente il primo  $N$  tale che

$$\text{errmedia} < 0.1$$

e confrontarlo con il valore di  $N$  ottenuto nel punto d).

**Soluzione.** a) Siano  $Z_1, Z_2 \dots$  v.a. i.i.d.  $\text{Be}(p)$ . Allora, preso  $j \geq 1$  intero, si ha

$$\{\inf\{n \geq 1 ; Z_n = 1\} = j\} = \{Z_1 = 0, \dots, Z_{j-1} = 0, Z_j = 1\}$$

(con la condizione di porre  $\{Z_1 = 0, \dots, Z_{j-1} = 0, Z_j = 1\} = \{Z_1 = 1\}$  quando  $j = 1$ ). Quindi per  $k = 0, 1, \dots$

$$\begin{aligned} p_k &= \mathbb{P}(T = k) = \mathbb{P}(\inf\{n \geq 1 ; Z_n = 1\} = k + 1) \\ &= \mathbb{P}(Z_1 = 0, \dots, Z_k = 0, Z_{k+1} = 1) \\ &= \prod_{i=1}^k \mathbb{P}(Z_i = 0) \cdot \mathbb{P}(Z_{k+1} = 1) = (1-p)^k p, \end{aligned}$$

cioè  $T \sim \text{Ge}(p)$ .

b) Usiamo la rappresentazione  $T = (\text{tempo di primo successo}) - 1$  in uno schema di Bernoulli:

```

T:=0;
X:=0;
while (X=0) do
  begin
    X:=bernoulli(p);
    T:=T+1
  end;
T:=T-1

```

(si noti che, così scritto, il ciclo `while` restituisce nella variabile `T` il tempo di primo successo).

Indicheremo con `geometrica(p)` una funzione che prendendo in *input* il valore  $p \in (0, 1)$ , genera una v.a. geometrica di parametro  $p = p$ .

c) Basta effettuare un ciclo per simulare  $N$  v.a. geometriche e sommare man mano i valori osservati:

```

Tmedio:=0;
for i:=1 to N do Tmedio:=Tmedio+geometrica(p);
Tmedio:=Tmedio/N;
errmedia:=abs(Tmedio-(1-p)/p)

```

d) Poiché  $\mathbb{E}(\bar{T}_N) = \mathbb{E}(T)$ , usando la disuguaglianza di Chebycev,

$$\begin{aligned} \mathbb{P}(errmedia \geq 0.1) &= \mathbb{P}(|\bar{T}_N - \mathbb{E}[T]| \geq 0.1) \\ &\leq \frac{\text{Var}(\bar{T}_N)}{0.1^2} = 100 \cdot \frac{1}{N} \text{Var}(T) = \frac{100(1-p)}{Np^2}, \end{aligned}$$

quindi  $\mathbb{P}(errmedia < 0.1) > 0.99$  se  $\frac{100(1-p)}{Np^2} < 0.01$ , cioè  $N > 10^4 \frac{(1-p)}{p^2}$ . Per visualizzare questo valore, basta aggiungere al programma il comando

```

N0:=Int(10000*(1-p)/sqr(p))+1

```

dove, ricordiamo, `Int` denota una funzione che dà la parte intera di un numero reale e `sqr` denota una funzione che restituisce il quadrato dell'argomento. Per verificare che questa stima sia valida, si potrebbe procedere come segue. Abbiamo visto che `N0` è tale che se si simulano `N0` geometriche di parametro  $p$  allora l'evento  $\{errmedia < 0.1\}$  si verifica con probabilità alta. Allora, stimiamo la probabilità dell'evento  $\{errmedia < 0.1\}$ , quando  $errmedia$  è valutato su `N0` geometriche, e verifichiamo che è  $> 0.99$ . La stima di  $\mathbb{P}(errmedia < 0.1)$  si può fare nel solito modo, ovvero con la LGN: valutiamo  $errmedia$  su un numero alto, ad esempio 1000, prove e stimiamo

$$\mathbb{P}(errmedia < 0.1) \simeq \frac{1}{1000} \sum_{k=1}^{1000} \mathbb{I}_{errmedia_k < 0.1},$$

dove  $errmedia_k$  è il valore di  $errmedia$  determinato alla  $k$ -esima prova. In codice Pascal, diventa:

```
proba:=0;
for k:=1 to 1000 do
  begin
    Tmedio:=0;
    for i:=1 to N0 do Tmedio:=Tmedio+geometrica(p);
    Tmedio:=Tmedio/N0;
    errmedia:=abs(Tmedio-(1-p)/p);
    if (errmedia<0.1) then proba:=proba+1
  end;
proba:=proba/1000;
if (proba>0.99) then write('la stima di N0 va bene!')
```

e) In questo caso, occorre simulare fin tanto che  $errmedia$  risulta maggiore o uguale a 0.1. Quindi, non è a priori noto il numero di simulazioni necessarie e il ciclo `for` non si può più usare. Indicando questa volta con  $M$  il numero totale di simulazioni, possiamo modificare il programma come segue:

```
M:=0;
errmedia:=1;
while (errmedia >= 0.1) do
  begin
    M:=M+1;
    Tmedio:=0;
    for i:=1 to M do Tmedio:=Tmedio+geometrica(p);
    Tmedio:=Tmedio/M;
    errmedia:=abs(Tmedio-(1-p)/p)
  end;
```

Si può ora procedere al confronto tra  $N0$  e  $M$ , ad esempio calcolando la distanza tra questi due numeri:

```
dist:=abs(N0-M)
```

Va però detto che tale confronto non è molto indicativo:  $M$  è una quantità aleatoria e in tal modo ne viene simulato un solo valore.

Il confronto diviene più interessante se  $N0$  viene paragonato con la media aritmetica  $N1$  ottenuta da tanti valori di  $M$ , ad esempio 1000 valori:

```
N1:=0;
for k:=1 to 1000 do
```

```

begin
  M:=0;
  errmedia:=1;
  while (errmedia >= 0.1) do
    begin
      M:=M+1;
      Tmedio:=0;
      for i:=1 to M do Tmedio:=Tmedio+geometrica(p);
      Tmedio:=Tmedio/M;
      errmedia:=abs(Tmedio-(1-p)/p)
    end;
  N1:=N1+M
end;
N1:=Int(N1/1000)

```

Ora, confrontando  $N_1$  con  $N_0$  otteniamo una indicazione più precisa della bontà della stima  $N_0$  ottenuta tramite la disuguaglianza di Chebycev.

Dal risultato che ottenete confrontando  $N_0$  e  $N_1$  potete dedurre che la stima  $N_0$  ottenuta con Chebycev sia una buona stima?

◇

**Osservazioni.** In questo problema ci sono un paio di cose più sofisticate di quelle proposte nei problemi precedenti: la verifica sperimentale della stima  $N_0$  e il calcolo numerico di  $N_1$ , cioè del più piccolo  $N$  affinché  $errmedia < 0.1$ . Nel primo caso,  $N_0$  è una quantità deterministica, qui valutata sulla base della disuguaglianza di Chebycev. Nel secondo caso, invece,  $N_1$  è una quantità aleatoria, e quello che viene fatto è calcolare la media empirica di tanti valori di  $N_1$  osservati su 1000 simulazioni. Osserviamo che il valore  $N_1$  che viene dato in *output* dà quindi una stima della media della v.a.  $N_1$ , come segue dalla LGN.

Ricapitolando, nei problemi precedenti abbiamo visto come simulare v.a. bernoulliane indipendenti (schema di Bernoulli: successo-insuccesso), quindi binomiali (numero di successi su  $n$  prove) e geometriche (tempo di primo successo - 1). Inoltre, abbiamo di fatto visto che, grazie alla LGN, per calcolare numericamente una probabilità del tipo  $p_A = \mathbb{P}(X \in A)$ , basta simulare un numero (grande)  $N$  di v.a.  $X_1, X_2, \dots, X_N$  i.i.d. di legge uguale a  $X$  e stimare

$$p_A \simeq \frac{1}{N} \sum_{k=1}^N \mathbb{I}_{X_k \in A} = \frac{\#\{k : X_k \in A\}}{N}.$$

Usando la disuguaglianza di Chebycev, è possibile determinare analiticamente un valore di  $N$  (= numero simulazioni) affinché la stima sopra scritta

abbia una “precisione” prefissata, ad esempio: con probabilità grande ( $> q$ , con  $q \simeq 1$ ), la distanza tra  $p_A$  e la sua stima sia piccola ( $< \delta$ , con  $\delta \simeq 0$ ). Abbiamo anche visto che tale stima è strettamente legata alla varianza della v.a.  $\mathbb{I}_{X_k \in A}$ , che vale  $p_A(1 - p_A)$  (verificare!). Nei problemi precedenti  $p_A$  è sempre nota, in generale però potrebbe non esserlo (come nell’Esercizio 2.1.4). In tal caso, occorrerà procedere ad una stima (maggiorazione) della varianza, altrimenti non sarà possibile determinare esplicitamente un valore per  $N$ .

Quanto detto sopra è legato al fatto che

$$p_A = \mathbb{P}(X \in A) = \mathbb{E}(\mathbb{I}_{X \in A}).$$

In generale quindi la procedura sopra descritta si può applicare al caso della stima della media  $\mathbb{E}(Y)$  di una v.a.  $Y$ :

$$\mathbb{E}(Y) \simeq \frac{1}{N} \sum_{k=1}^N Y_k,$$

dove le  $Y_k$  sono  $N$  v.a. i.i.d. simulate con la stessa legge di  $Y$ , e la disuguaglianza di Chebycev dà anche un valore di  $N$  legato alla precisione della stima. Questo metodo di stima, stocastico perché legato alla simulazione (che abbiamo usato per risolvere il punto e) del Problema 2.1.3), è un esempio di **metodo Monte Carlo**.

Proponiamo ora alcuni esercizi, la cui soluzione è strettamente legata alle considerazioni appena fatte.

**Esercizio 2.1.4.** *Una moneta dà testa con probabilità  $p$ , che non si conosce e che potrebbe essere un qualsiasi numero di  $(0, 1)$ . Simulando lanci ripetuti della moneta, scrivere un programma per stimare  $p$  in modo tale che la probabilità che la stima disti dal valore vero per meno di 0.01 sia maggiore di 0.99.*

**Esercizio 2.1.5.** *Scrivere un programma per calcolare numericamente la probabilità che, lanciando 100 volte una moneta equilibrata, escano esattamente 6 teste consecutive.*

**Esercizio 2.1.6.** *Una v.a. di Poisson assume i valori  $k = 0, 1, \dots$  con probabilità  $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$ . Dimostrare che presa  $Z_n \sim \text{Bi}(n, \lambda/n)$  allora per  $n$  grande,*

$$\mathbb{P}(Z_n = k) \simeq p_k.$$

*Usando quest’approssimazione, scrivere una procedura per simulare una v.a. di Poisson.*



## 2.2 Simulazione di variabili aleatorie finite

### Problema 2.2.1. Generazione di variabili aleatorie finite.

**a)** Sia  $Z \sim \text{Un}(0, 1)$ . Fissati  $x_1, \dots, x_n \in \mathbb{R}$  e  $p_1, \dots, p_n \in \mathbb{R}_+$  tali che  $p_1 + \dots + p_n = 1$ , si ponga

$$X = x_1 \mathbb{I}_{A_1}(Z) + x_2 \mathbb{I}_{A_2}(Z) + x_3 \mathbb{I}_{A_3}(Z) + \dots + x_n \mathbb{I}_{A_n}(Z)$$

dove

$$A_1 = [0, p_1), \quad A_2 = [p_1, p_1 + p_2), \\ A_3 = [p_1 + p_2, p_1 + p_2 + p_3), \dots, A_n = [p_1 + p_2 + \dots + p_{n-1}, 1],$$

cioè:  $X = x_i$  se e solo se  $Z \in A_i$ . Dimostrare che  $X$  è una v.a. finita, con  $E_X = \{x_1, \dots, x_n\}$  e distribuzione  $p_X(x_i) = p_i$ ,  $i = 1, \dots, n$ .

**b)** Sia  $X$  una variabile aleatoria finita, che può assumere i valori  $x_1, \dots, x_n$  con probabilità  $p_1, \dots, p_n$ . Usando **a)**, scrivere una procedura per simulare la v.a.  $X$ , assegnati  $n$ ,  $\{x_1, \dots, x_n\}$  e  $\{p_1, \dots, p_n\}$ .

**c)** Fissato un intero positivo  $N$  (ad es.  $N = 10000$ ), scrivere un programma che, generando  $N$  v.a.  $X_1, \dots, X_N$  indipendenti tutte con la stessa legge di  $X$ ,

**c1)** confronti le frequenze relative con cui si presentano i valori  $x_1, \dots, x_n$  nelle  $N$  simulazioni con le rispettive probabilità  $p_1, \dots, p_n$  tramite

$$\text{errmax} = \max_{1 \leq i \leq n} |f_i - p_i|$$

dove  $f_i$  è la frequenza relativa di  $x_i$ ;

**c2)** calcoli la media aritmetica delle  $N$  simulazioni  $\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i$  e confronti il valore osservato con il valor medio  $\mathbb{E}[X] = \sum_{j=1}^n x_j p_j$  previsto teoricamente tramite

$$\text{errmedia} = |\bar{X}_N - \mathbb{E}[X]|.$$

**Soluzione.** **a)** Mostriamo la validità della rappresentazione

$$X = x_i \quad \text{se e solo se} \quad Z \in A_i$$

dove  $A_1 = [0, p_1)$ ,  $A_i = [p_1 + \dots + p_{i-1}, p_1 + \dots + p_{i-1} + p_i)$  per  $2 \leq i \leq n-1$ ,  $A_n = [p_1 + \dots + p_{n-1}, 1]$  e  $Z$  è una uniforme su  $[0, 1]$ . Detti  $a_i$  e  $b_i$  gli estremi dell'intervallo  $A_i$ ,

$$\mathbb{P}(X = x_i) = \mathbb{P}(Z \in A_i) = b_i - a_i = p_1 + \dots + p_{i-1} + p_i - (p_1 + \dots + p_{i-1}) = p_i.$$

Quindi  $X$  è in effetti una v.a. che assume i valori  $x_1, \dots, x_n$  con rispettive probabilità  $p_1, \dots, p_n$ .

**b)** Indichiamo con **val** e **pr** rispettivamente il vettore dei valori ammissibili e della distribuzione di  $X$ : **val**[**k**] =  $x_k$  e **pr**[**k**] =  $p_k$ , per  $k = 1, \dots, n$ . Immaginiamo ovviamente che tali vettori vengano assegnati in *input*.

Per simulare  $X$ , usiamo la rappresentazione di cui sopra: generiamo una v.a. casuale  $Z$  su  $[0, 1]$  e la confrontiamo dapprima con  $p_1$ , poi con  $p_1 + p_2$  e così via. Nel frattempo aggiorniamo un contatore **kappa** (di tipo **Integer** e inizializzato a 1) che consente di calcolare il valore **sommparz** =  $p_1 + \dots + p_k$  con cui effettuiamo il confronto “  $Z < \text{sommparz}$  “. La prima volta che la disuguaglianza “  $Z < \text{sommparz}$  “ risulta vera determina il valore  $k = \text{kappa}$  tale che  $Z \in A_k$  e quindi  $X = x_k$ .

Infine, usiamo una variabile di controllo **flag** che, inizializzata a 0, viene posta uguale a 1 quando il valore  $x_k$  è stato determinato:

```
Z:=nrandom;
sommparz:=pr[1];
kappa:=1;
flag:=0;
while (flag=0) do
  begin
    if (Z < sommparz) then
      begin
        flag:=1;
        X:=val[kappa]
      end
    else
      begin
        kappa:=kappa+1;
        sommparz:=sommparz+pr[kappa]
      end
    end
  end
```

Indichiamo con **va\_finita**(**val**,**pr**) una funzione che prendendo in *input* i vettori **val**  $\in \mathbb{R}^n$  e **pr**  $\in \{p = (p_1, \dots, p_n) \in \mathbb{R}_+^n : p_1 + \dots + p_n = 1\}$ , genera una v.a. che assume i valori  $x_k = \text{val}[\mathbf{k}]$  con probabilità  $p_k = \text{pr}[\mathbf{k}]$ , per  $k = 1, \dots, n$ : l'*output* è dato dal risultato **X** ottenuto sopra.

Come vedremo nel seguito, sarà utile non tanto avere il valore simulato  $x_k$  di  $X$  quanto conoscere l'indice  $k$ . Pertanto, sia **genera\_indice**(**val**,**pr**)

una funzione che prendendo in *input* i vettori **val** e **pr**, restituisce l'indice **kappa** tale che il valore simulato di  $X$  è **val[kappa]**. Tale funzione avrà come *output* il valore **kappa** ottenuto tramite il programma sopra scritto, così com'è oppure (meglio) sostituendo la parte in riquadro semplicemente con

```
if (Z < sommeparz) then flag:=1;
```

c) Poiché la variabile **n** è già impegnata, indichiamo con **NN** il numero  $N$  di simulazioni. Possiamo riassumere quanto richiesto in **c1)** e **c2)** come segue:

- calcolo del vettore delle frequenze relative **fr** e della media aritmetica **Xmedio**:

```
for k:=1 to n do fr[k]:=0;
Xmedio:=0;
for i:=1 to NN do
  begin
    k:=genera_indice(val,pr);
    fr[k]:=fr[k]+1;
    Xmedio:=Xmedio+val[k]
  end;
for k:=1 to n do fr[k]:=fr[k]/NN;
Xmedio:=Xmedio/NN
```

- calcolo di *errmax*:

```
errmax:=abs(fr[1]-pr[1]);
for k:=2 to n do
  if (errmax > abs(fr[k]-pr[k]))
    then errmax:=abs(fr[k]-pr[k])
```

- calcolo della media teorica **mediavera** =  $\sum_{k=1}^n x_k p_k$  e di *errmedia*:

```
mediavera:=0;
for k:=1 to n do mediavera:=mediavera+val[k]*pr[k];
errmedia:=abs(Xmedio-mediavera)
```

◇

**Esercizio 2.2.2.** *Fissato  $m > 1$  intero, dimostrare che se  $Z \sim \text{Un}(1, m+1)$  allora  $X = [Z]$  ( $[ \cdot ]$  = parte intera) è una v.a. uniforme su  $\{1, 2, \dots, m\}$ . Scrivere una procedura per generare un numero a caso<sup>3</sup> su  $\{1, 2, \dots, m\}$  e più in generale su  $\{x_1, x_2, \dots, x_m\}$ , con  $x_1, \dots, x_m \in \mathbb{R}$  dati in input.*

<sup>3</sup>Si veda anche il prossimo Problema 2.3.1.

**Esercizio 2.2.3.** Scrivere un programma che simula il lancio di un dado nei casi in cui:

a) il dado è equilibrato;

b) il dado non è equilibrato e sono note le probabilità  $p_1, \dots, p_6$  con le quali si presentano le sei facce.

## 2.3 Metodo Monte Carlo per il calcolo numerico di integrali ed aree

### Problema 2.3.1. Calcolo numerico di integrali, I

a) Siano  $a, b \in \mathbb{R}$ , tali che  $a < b$ . Sia  $Z \sim \text{Un}(0, 1)$  e  $U = a + (b - a)Z$ . Mostrare che  $U \sim \text{Un}(a, b)$ . Scrivere una procedura per generare una v.a.  $\text{Un}(a, b)$ .

b) Sia  $f : [a, b] \rightarrow \mathbb{R}$  una funzione continua (a tratti). Siano  $U_1, \dots, U_N$  v.a. i.i.d. di legge  $\text{Un}(a, b)$ . Mostrare che, per  $N \rightarrow \infty$ ,

$$\frac{b-a}{N} \sum_{i=1}^N f(U_i) \simeq \int_a^b f(x) dx$$

nel senso che: per ogni  $\delta > 0$ ,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \left| \frac{b-a}{N} \sum_{i=1}^N f(U_i) - \int_a^b f(x) dx \right| > \delta \right) = 0.$$

c) Siano  $M$  e  $m$  rispettivamente il massimo ed il minimo di  $f$  su  $[a, b]$ . Dimostrare che<sup>4</sup>

$$\text{Var}(f(U_i)) \leq (M - m)^2.$$

Usare questa stima e la disuguaglianza di Chebycev per determinare un  $N^*$  tale che con probabilità maggiore di 0.9 l'approssimazione di  $\int_a^b f(x) dx$  data in b) sia precisa alla seconda cifra decimale

d) È noto che

$$\int_0^1 \frac{4}{1+x^2} dx = 4 \arctan x \Big|_0^1 = \pi = 3,14159265358\dots$$

Usare questa rappresentazione di  $\pi$  per scrivere un programma che calcoli (numericamente)  $\pi$ , simulando  $N^*$  v.a. (con  $N^*$  come in c)). Verificare empiricamente che la stima è corretta alla seconda cifra decimale con probabilità maggiore di 0.9.

---

<sup>4</sup>Sugg.: poiché  $m \leq f(U_i) \leq M$ , allora  $m \leq \mathbb{E}(f(U_i)) \leq M$ , quindi  $|f(U_i) - \mathbb{E}(f(U_i))| \leq M - m$ . Si usi poi la rappresentazione  $\text{Var}(f(U_i)) = \mathbb{E}(|f(U_i) - \mathbb{E}(f(U_i))|^2)$ .

Presumibilmente le stime corrette alla seconda cifra decimale saranno ben più del 90% perché è possibile dimostrare che l' $N^*$  può essere scelto più piccolo di quello determinato in **c**). Verificare quindi numericamente con quale probabilità questo valore di  $N^*$  dà una stima corretta alla terza cifra decimale.

e) Scrivere un programma per calcolare numericamente  $\mathbb{P}(|W| \leq 2)$ , con  $W \sim N(0, 1)$ , con precisione alla seconda cifra decimale<sup>5</sup>.

**Soluzione.** a) Si ha

$$F_U(u) = \mathbb{P}(U \leq u) = \mathbb{P}(a + (b - a)Z \leq u) = \mathbb{P}\left(Z \leq \frac{u - a}{b - a}\right) = F_Z\left(\frac{u - a}{b - a}\right)$$

quindi

$$f_U(u) = \frac{d}{du}F_U(u) = \frac{d}{du}F_Z\left(\frac{u - a}{b - a}\right) = f_Z\left(\frac{u - a}{b - a}\right) \cdot \frac{1}{b - a}.$$

Ricordando che  $f_Z(z) = \mathbb{I}_{z \in (0,1)}$  otteniamo

$$f_U(u) = \frac{1}{b - a} \mathbb{I}_{\frac{u - a}{b - a} \in (0,1)} = \frac{1}{b - a} \mathbb{I}_{u \in (a,b)}$$

che è la densità di una v.a. uniforme su  $(a, b)$ , quindi  $U \sim \text{Un}(a, b)$ . Per generare un numero  $U$  uniforme su  $(a, b)$ , basta quindi scrivere

```
Z:=nrandom;
U:=a+(b-a)*Z
```

Chiameremo **uniforme(a,b)** tale funzione.

b) La LGN assicura che

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\left|\frac{1}{N} \sum_{i=1}^N X_i - \mu\right| > \delta\right) = 0, \quad \text{per ogni } \delta > 0 \quad (2.1)$$

dove  $X_1, X_2, \dots$  sono v.a. i.i.d. con (momento secondo e) media  $\mu$ . Quindi, per  $N$  grande possiamo dire  $\frac{1}{N} \sum_{i=1}^N X_i \simeq \mu$ , nel senso dato da (2.1).

Consideriamo ora  $U_1, U_2, \dots$  v.a. i.i.d. di legge  $\text{Un}(a, b)$  e poniamo, per ogni  $i$ ,  $X_i = f(U_i)$ . Le  $X_i$  rimangono indipendenti. Inoltre, hanno momento secondo se e solo se

$$\mathbb{E}(|f(U_i)|^2) = \int_a^b |f(u)|^2 p_{U_i}(u) du < \infty$$

---

<sup>5</sup>Si noti che  $\mathbb{P}(|W| \leq 2) = \int_{-2}^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^2 e^{-x^2/2} dx$ . Si fa presente che le tavole della distribuzione gaussiana danno  $\mathbb{P}(|W| \leq 2) = 0.9545$ .

dove  $p_{U_i}(u)$  denota la densità di  $U_i$ :  $p_{U_i}(u) = \frac{1}{b-a} \mathbb{1}_{u \in (a,b)}$ . Quindi

$$\mathbb{E}(|f(U_i)|^2) = \frac{1}{b-a} \int_a^b |f(u)|^2 du$$

che esiste finito se  $f$  è continua (o anche continua a tratti). Ciò significa che la LGN vale in questo caso: per  $N$  grande,

$$\frac{1}{N} \sum_{i=1}^N f(U_i) \simeq \mu = \mathbb{E}(f(U_1)) = \int_a^b f(u) p_{U_1}(u) du = \frac{1}{b-a} \int_a^b f(u) du$$

cioè

$$\frac{b-a}{N} \sum_{i=1}^N f(U_i) \simeq \int_a^b f(u) du.$$

**c)** Poiché  $m \leq f(U_i) \leq M$ , si ha  $m \leq \mathbb{E}(f(U_i)) \leq M$ , cioè  $f(U_i), \mathbb{E}(f(U_i)) \in [m, M]$ . Allora la distanza tra  $f(U_i)$  e  $\mathbb{E}(f(U_i))$  non supera l'ampiezza totale dell'intervallo  $[m, M]$ , quindi  $|f(U_i) - \mathbb{E}(f(U_i))| \leq M - m$  e  $|f(U_i) - \mathbb{E}(f(U_i))|^2 \leq (M - m)^2$ . Infine,

$$\text{Var}(f(U_i)) = \mathbb{E}\left(|f(U_i) - \mathbb{E}(f(U_i))|^2\right) \leq (M - m)^2.$$

Ora, l'approssimazione è corretta alla seconda cifra decimale se

$$\left| \int_a^b f(x) dx - \frac{(b-a)}{N} \sum_{i=1}^N f(U_i) \right| < 0.01.$$

Si chiede quindi  $N$  affinché

$$\mathbb{P}\left(\left| \int_a^b f(x) dx - \frac{(b-a)}{N} \sum_{i=1}^N f(U_i) \right| < 0.01\right) > 0.9,$$

o equivalentemente

$$\mathbb{P}\left(\left| \int_a^b f(x) dx - \frac{(b-a)}{N} \sum_{i=1}^N f(U_i) \right| \geq 0.01\right) < 0.1.$$

Poiché  $\mathbb{E}\left(\frac{(b-a)}{N} \sum_{i=1}^N f(U_i)\right) = \int_a^b f(x) dx$ , usando la disuguaglianza di Chebycev otteniamo

$$\mathbb{P}\left(\left| \int_a^b f(x) dx - \frac{(b-a)}{N} \sum_{i=1}^N f(U_i) \right| \geq 0.01\right)$$

$$\leq \frac{\text{Var}\left(\frac{(b-a)}{N} \sum_{i=1}^N f(U_i)\right)}{0.01^2} = 10^4 \frac{(b-a)^2}{N^2} \sum_{i=1}^N \text{Var}(f(U_i)).$$

Usando la stima  $\text{Var}(f(U_i)) \leq (M-m)^2$ , otteniamo

$$\mathbb{P}\left(\left|\int_a^b f(x)dx - \frac{(b-a)}{N} \sum_{i=1}^N f(U_i)\right| \geq 0.01\right) \leq 10^4 \frac{(b-a)^2 (M-m)^2}{N}$$

Quindi, basta scegliere  $N^*$  tra gli  $N$  tali che

$$10^4 \frac{(b-a)^2 (M-m)^2}{N} < 0.1, \text{ cioè } N > (b-a)^2 (M-m)^2 10^5.$$

È opportuno prendere  $N^*$  il più piccolo possibile, quindi uguale alla parte intera di  $(b-a)^2 (M-m)^2 10^5$  più 1.

**d)** Qui  $a = 0$ ,  $b = 1$ ,  $M = 4$ ,  $m = 2$ . Inoltre,  $N^* = 4 \times 10^5 + 1$ . Per calcolare numericamente  $\pi$  con buona precisione fino alla seconda cifra decimale, basta seguire gli algoritmi descritti prima: generiamo  $N^*$  v.a.  $U_1, \dots, U_{N^*}$  di legge  $\text{Un}(0, 1)$  (quindi qui basta usare la funzione `nrandom`) e calcoliamo la media empirica delle  $f(U_i)$ , con  $f(x) = 4/(1+x^2)$ . Ad esempio,

```

pigreco:=0;
for k:=1 to 400001 do
  begin
    U:=nrandom;
    pigreco:=pigreco+4/(1+sqr(U));
  end;
pigreco:=pigreco/400001

```

Per verificare che  $N^*$  dà la stima corretta alla seconda cifra decimale con probabilità maggiore di 0.9, scriviamo un programma che calcoli numericamente  $\pi$   $M$  volte (ad esempio,  $M = 1000$ ) e contiamo quante volte il valore ottenuto per  $\pi$  dista da 3.14 per meno di 0.01: basta considerare la variabile `cont_2` che conta quante volte il risultato è corretto alla seconda cifra decimale. Analogamente, la variabile `cont_3` conta quante volte il risultato è corretto alla terza cifra decimale.

```

cont_2:=0;
cont_3:=0;
for j:=1 to M do
  begin
    pigreco:=0;
    for k:=1 to 400001 do

```

```

begin
  U:=nrandom;
  pigreco:=pigreco+4/(1+sqr(U));
end;
pigreco:=pigreco/400001;
if (abs(pigreco-3.14)<0.01) then cont_2:=cont_2+1;
if (abs(pigreco-3.141)<0.001) then cont_3:=cont_3+1
end

```

In questo modo la variabile `cont_2` dà quante volte il valore di `pigreco` è della forma 3.14... e `cont_3` dà quante volte il valore di `pigreco` è della forma 3.141.... Ci aspettiamo dunque che `cont_2`  $> \frac{9}{10} M$  (ad esempio, su 1000 prove ci aspettiamo che `cont_2` sia maggiore di 900: torna?). Inoltre, la probabilità `proba_3` con la quale il risultato è corretto alla terza cifra decimale si può stimare con `cont_3` diviso per il numero `M` di risultati ottenuti: basterà aggiungere al programma il comando

```
proba_3:=cont_3/M
```

e) Poiché  $\mathbb{P}(|W| \leq 2) = \int_{-2}^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ , la probabilità richiesta si può approssimare stimando  $\int_{-2}^2 e^{-x^2/2} dx$  con gli algoritmi sopra descritti (qui  $a = -2$  e  $b = 2$ ):

$$\int_{-2}^2 e^{-x^2/2} dx \simeq \frac{4}{N^*} \sum_{i=1}^{N^*} e^{-U_i^2/2}.$$

Ora,  $M = 1$ ,  $m = e^{-2}$ , quindi  $(b-a)^2 (M-m)^2 10^5 = 16 \cdot (1-0.1353)^2 \cdot 10^5 = 11.96329744 \cdot 10^5$ , quindi  $N^* = 1196330$ . Possiamo scrivere:

```

proba:=0;
for k:=1 to 299059 do
  begin
    U:=uniforme(-2,2);
    proba:=proba+exp(-sqr(U)/2);
  end;
proba:=1/sqrt(2*Pi)*4*proba/299059

```

Si fa presente che le tavole della distribuzione gaussiana danno  $\mathbb{P}(|W| \leq 2) = 0.9545$ .

◇

**Problema 2.3.2. Approssimazione di aree (Metodo del rigetto)**



**a)** Siano  $a, b, m, M \in \mathbb{R}$  tali che  $a < b$  e  $m < M$  e sia  $R = [a, b] \times [m, M]$ . Si può dire che un punto  $(X, Y)$  è “a caso nel rettangolo  $R$  se le sue coordinate  $X$  e  $Y$  sono indipendenti, ciascuna a caso nell’intervallo di appartenenza. Scrivere un programma per simulare un punto  $(X, Y)$  “a caso nel rettangolo  $R$ .

**b)** Sia  $(X, Y)$  un punto “a caso nel rettangolo  $R = [a, b] \times [m, M]$ . Siano  $g_1, g_2 : [a, b] \rightarrow [m, M]$  due funzioni continue a tratti tali che  $g_1(x) \leq g_2(x)$ , per ogni  $x \in [a, b]$ . Definiamo

$$A = \{(x, y) \in R : g_1(x) \leq y \leq g_2(x)\} \subset R,$$

$$W = \begin{cases} 1 & \text{se } (X, Y) \in A \\ 0 & \text{altrimenti.} \end{cases}$$

Scrivere una procedura per simulare  $W$ , a partire dalla simulazione di una coppia  $(X, Y)$  come in **a)**, supponendo  $g_1$  e  $g_2$  date in input. Verificare inoltre che  $W \sim \text{Be}(p_A)$ , con  $p_A = \mathbb{P}((X, Y) \in A)$ <sup>6</sup>.

**c)** Se  $(X, Y)$  è scelto a caso in  $R = [a, b] \times [m, M]$ , si può dimostrare che

$$\mathbb{P}((X, Y) \in A) = \frac{\text{area di } A}{\text{area di } R}$$

per ogni  $A$  sottoinsieme “buono di  $R$ , in particolare per  $A$  scelto come in **b)**. Utilizzare questa proprietà per descrivere una procedura per calcolare numericamente l’area di  $A$ <sup>7</sup>.

Ad esempio, scrivere un programma per calcolare numericamente l’area racchiusa dall’ellisse di equazione  $\lambda(x - x_0)^2 + (y - y_0)^2 = \mu$ , dove  $\lambda, \mu > 0$  e  $x_0, y_0$  sono dati in input (che succede se  $x_0 = y_0 = 0$  e  $\lambda = \mu = 1$ ?).

**Soluzione.** **a)** Sia `uniforme(alpha, beta)` una funzione che genera una v.a.  $\text{Un}(\alpha, \beta)$  (cfr. Problema 2.3.1). Per generare una v.a. uniforme (o a caso) nel rettangolo  $R = [a, b] \times [m, M]$  (con  $a, b, m, M$  dati in input), basta generare un punto  $X \sim \text{Un}(a, b)$  ed un punto  $Y \sim \text{Un}(m, M)$ , indipendenti, e poi considerare il punto  $(X, Y)$  di  $[a, b] \times [m, M]$ . In sostanza, basta richiamare la procedura `uniforme(alpha, beta)` opportunamente:

<sup>6</sup>Sebbene  $W \sim \text{Be}(p_A)$ , l’algoritmo per la generazione di questa bernoulliana è diverso da quello proposto nel Problema 2.1.1, dove è richiesto il parametro. Si noti che qui il parametro  $p_A$  **non** è noto.

<sup>7</sup>Se  $W$  denota la v.a. definita in **b)**, si noti che  $\mathbb{E}(W) = p_A = \mathbb{P}((X, Y) \in A)$ , quindi

$$\text{area di } A = (\text{area di } R) \times p_A = (b - a)(M - m)p_A.$$

Per approssimare  $p_A$  si può usare la LGN: prese  $N$  v.a.  $W_1, \dots, W_N$  indipendenti, tutte con la stessa legge di  $W$ , allora per  $N$  grande,

$$p_A = \mathbb{E}(W) \simeq \frac{1}{N} \sum_{k=0}^N W_k.$$

```
X:=uniforme(a,b);
Y:=uniforme(m_min,M_max)
```

b) Indichiamo con  $g_1(x)$  e  $g_2(x)$  due funzioni che danno, rispettivamente, il valore di  $g_1$  e  $g_2$  nel punto  $x = x$ . Allora, per simulare  $W$  basterà scrivere

```
X:=uniforme(a,b);
Y:=uniforme(m_min,M_max);
if ( (g1(X) <= Y) and (Y <= g2(X)) ) then W:=1
    else W:=0
```

Inoltre, ovviamente  $W$  può assumere solo i valori 1 e 0, quindi  $W$  è una bernoulliana di parametro  $\mathbb{P}(W = 1) = \mathbb{P}((X, Y) \in A)$ .

c) Seguiremo l'idea suggerita: simuliamo  $N$  (ad esempio,  $N = 10000$ ) v.a. indipendenti di legge  $W$  e approssimiamo  $p_A$  con la media empirica dei risultati ottenuti:

```
NA:=0;
for k:=1 to N do
    begin
        X:=uniforme(a,b);
        Y:=uniforme(m_min,M_max);
        if ( (g1(X) <= Y) and (Y <= g2(X)) ) then NA:=NA+1
    end;
area_di_A:=(b-a)*(M_max-m_min)*NA/N
```

L'insieme  $A$  racchiuso dall'ellisse di equazione  $\lambda(x - x_0)^2 + (y - y_0)^2 = \mu$  è dato dalle coppie  $(x, y)$  tali che  $\lambda(x - x_0)^2 + (y - y_0)^2 \leq \mu$ , cioè  $|y - y_0| \leq \sqrt{\mu - \lambda(x - x_0)^2}$ . Ora, evidentemente dev'essere  $\mu - \lambda(x - x_0)^2 \geq 0$ , cioè  $|x - x_0| \leq \sqrt{\mu/\lambda}$  e quindi  $x \in [x_0 - \sqrt{\mu/\lambda}, x_0 + \sqrt{\mu/\lambda}]$ . Inoltre,  $|y - y_0| \leq \sqrt{\mu - \lambda x^2} \leq \sqrt{\mu}$ , quindi  $y \in [y_0 - \sqrt{\mu}, y_0 + \sqrt{\mu}]$ . Allora

$$A = \left\{ (x, y) \in [x_0 - \sqrt{\mu/\lambda}, x_0 + \sqrt{\mu/\lambda}] \times [y_0 - \sqrt{\mu}, y_0 + \sqrt{\mu}] : \right. \\ \left. y_0 - \sqrt{\mu - \lambda(x - x_0)^2} \leq y \leq y_0 + \sqrt{\mu - \lambda(x - x_0)^2} \right\}.$$

cioè  $a = x_0 - \sqrt{\mu/\lambda}$ ,  $b = x_0 + \sqrt{\mu/\lambda}$ ,  $m = y_0 - \sqrt{\mu}$ ,  $M = y_0 + \sqrt{\mu}$ ,  $g_1(x) = y_0 - \sqrt{\mu - \lambda(x - x_0)^2}$ ,  $g_2(x) = y_0 + \sqrt{\mu - \lambda(x - x_0)^2}$ :

```
(* AREA RACCHIUSA DALL'ELLISSE DI EQUAZIONE *)
(*      lambda*(x-x0)^2+(y-y0)^2=mu      *)
```

```

N_ellisse:=0;
a:=x0-sqrt(mu/lambda);
b:=x0+sqrt(mu/lambda);
m_min:=y0-sqrt(mu);
M_max:=y0+sqrt(mu);
for k:=1 to N do
  begin
    X:=uniforme(a,b);
    Y:=uniforme(m_min,M_max);
    g1_X:=y0-sqrt(mu-lambda*sqr(X-x0));
    g2_X:=y0+sqrt(mu-lambda*sqr(X-x0));
    if ( (g1_X <= Y) and (Y <= g2_X) )
      then N_ellisse:=N_ellisse+1
    end;
area_ellisse:=(b-a)*(M_max-m_min)*N_ellisse/N

```

Per verificare che il programma dà il risultato corretto, si può scegliere, ad esempio,  $x_0 = y_0 = 0$  e  $\lambda = \mu = 1$ : in tal caso, l'ellisse diviene la circonferenza di raggio unitario, centrata nell'origine, la cui area è  $\pi$ . Il risultato quindi dev'essere vicino a  $\pi$ . Si noti infine che, con questa scelta dei coefficienti, abbiamo determinato una ulteriore procedura di approssimazione di  $\pi$  (cfr. Problema 2.3.1).

◇

### Problema 2.3.3. Calcolo numerico di integrali, II

**a)** Sia  $f : [a, b] \rightarrow [0, +\infty)$  continua. Posto  $A = \{(x, y) \in [a, b] \times \mathbb{R} : 0 \leq y \leq f(x)\}$  e ricordando che  $\int_a^b f(x)dx = \text{area di } A$ , descrivere un algoritmo per il calcolo numerico di  $\int_a^b f(x)dx$  facendo uso del Problema 2.3.2. In particolare, calcolare numericamente  $\pi$ , usando la funzione  $f(x) = 4(1+x^2)^{-1}$ .

**b)** Abbiamo visto due metodi per il calcolo numerico dell'integrale. Qui,

$$\int_a^b f(x)dx \simeq \frac{1}{N} \sum_{k=1}^N (b-a) M W_k,$$

dove  $M = \max_{x \in [a,b]} f(x)$  e  $W_k = \mathbb{I}_{Y_k \leq f(X_k)}$ , con  $X_k \sim \text{Un}(a, b)$ ,  $Y_k \sim \text{Un}(0, M)$  indipendenti. Oppure, come nell'Esercizio 2.3.1:

$$\int_a^b f(x)dx \simeq \frac{1}{N} \sum_{k=1}^N (b-a)f(U_k),$$

dove  $U_k \sim \text{Un}(a, b)$  indipendenti.

Mostrare che  $\text{Var}\left((b-a)M W_k\right) \leq \text{Var}\left((b-a)f(U_k)\right)$  e dedurre che l'algoritmo qui proposto è preferibile a quello proposto nell'Esercizio 2.3.1 (perché più veloce).

**Soluzione.** a) Per calcolare numericamente  $\int_a^b f(x)dx$  basta usare la procedura in c) del Problema 2.3.2, con  $m = 0$ ,  $g_1(x) \equiv 0$  e  $g_2(x) = f(x)$ . Indicando con  $\mathbf{f}(\mathbf{x})$  una funzione che dà il valore della funzione  $f$  in  $x = \mathbf{x}$ , possiamo scrivere

```
(*   CALCOLO DELL'INTEGRALE DI f(x) TRA a E b   *)
(*   (M_max denota il max di f(x) per x in [a,b]) *)
```

```
N_int:=0;
for k:=1 to N do
  begin
    X:=uniforme(a,b);
    Y:=uniforme(0,M_max);
    if (Y <= f(X)) ) then N_int:=N_int+1
  end;
integrale_a_b:=(b-a)*M_max*N_int/N
```

Per calcolare  $\pi$ , usiamo  $\int_0^1 4/(1+x^2)dx = \pi$ : qui  $a = 0$ ,  $b = 1$ ,  $M = 4$ .

```
(* CALCOLO DI pi_greco *)
```

```
N_pigreco:=0;
for k:=1 to N do
  begin
    X:=uniforme(0,1);
    Y:=uniforme(0,4);
    if (Y <= 4/(1+X*X)) ) ) then N_pigreco:=N_pigreco+1
  end;
pi_greco:=4*N_pigreco/N
```

b) Calcoliamo le due varianze richieste. Cominciamo dalla seconda: ricordando che  $U_k \sim \text{Un}(a, b)$ ,

$$\begin{aligned} \text{Var}\left((b-a)f(U_k)\right) &= (b-a)^2 \text{Var}\left(f(U_k)\right) = (b-a)^2 \left(\mathbb{E}(f^2(U_k)) - \mathbb{E}^2(f(U_k))\right) \\ &= (b-a)^2 \left(\int_a^b f^2(x) \frac{1}{b-a} dx - \left(\int_a^b f(x) \frac{1}{b-a} dx\right)^2\right) \\ &= (b-a) \int_a^b f^2(x) dx - \left(\int_a^b f(x) dx\right)^2. \end{aligned}$$

Inoltre, ricordando che  $W_k \sim \text{Be}(p_A)$ , con

$$p_A = \frac{\text{area di } A}{\text{area di } R} = \frac{\int_a^b f(x) dx}{(b-a)M},$$

si ha

$$\begin{aligned} \text{Var}\left((b-a)M W_k\right) &= (b-a)^2 M^2 \text{Var}(W_k) = (b-a)^2 M^2 p_A(1-p_A) \\ &= (b-a)^2 M^2 \frac{1}{(b-a)M} \int_a^b f(x) dx \left(1 - \frac{1}{(b-a)M} \int_a^b f(x) dx\right) \\ &= (b-a) \int_a^b M f(x) dx - \left(\int_a^b f(x) dx\right)^2. \end{aligned}$$

Ora,  $M = \max_{x \in [a,b]} f(x)$ , quindi  $M \geq f(x) \geq 0$  per ogni  $x \in [a, b]$  e  $M f(x) \geq f^2(x)$  per ogni  $x \in [a, b]$ . Allora,  $\int_a^b M f(x) dx \geq \int_a^b f^2(x) dx$  e

$$\begin{aligned} \text{Var}\left((b-a)M W_k\right) &= (b-a) \int_a^b M f(x) dx - \left(\int_a^b f(x) dx\right)^2 \\ &\geq (b-a) \int_a^b f^2(x) dx - \left(\int_a^b f(x) dx\right)^2 = \text{Var}\left((b-a)f(U_k)\right). \end{aligned}$$

Per rispondere alla seconda domanda, ricordiamo (cfr. Paragrafo 1.5) che il TLC assicura che l'approssimazione

$$\bar{X}_N \equiv \frac{1}{N} \sum_{k=1}^N X_k \simeq \mu = \mathbb{E}(X_k).$$

è tale che, per  $N$  grande,

$$\bar{X}_N - \mu \simeq \frac{\sigma}{\sqrt{N}} Z,$$

dove  $Z \sim N(0, 1)$  e  $\sigma^2 = \text{Var}(X_k)$ . Quindi, la convergenza di  $\bar{X}_N$  a  $\mu$  ha velocità  $\sigma/\sqrt{N}$ , in particolare è direttamente proporzionale a  $\sigma$ : più  $\sigma$  è grande, più la convergenza si rallenta.

◇

**Osservazioni e complementi.** Nel caso del calcolo numerico di un integrale, la LGN è stata usata in due modi:

$$\int_a^b f(x) dx \simeq \bar{X}_N$$

dove

$$\begin{aligned} \text{primo caso (Problema 2.3.1):} \quad X_k &= (b-a) f(U_k) \\ \text{secondo caso (Problema 2.3.3):} \quad X_k &= (b-a) M W_k \end{aligned}$$

Abbiamo appena visto che nel primo caso la varianza delle  $X_k$  è minore o uguale alla varianza che si ha nel secondo: possiamo quindi dedurre che il primo metodo suggerito è senz'altro preferibile al secondo.

Esiste un terzo modo per calcolare numericamente l'integrale (**metodo delle variabili antitetiche**): usando sempre la LGN, si verifica facilmente che (supponiamo, ma solo per semplicità, che  $a = 0$  e  $b = 1$ )

$$\int_0^1 f(x) dx \simeq \frac{1}{N} \sum_{k=1}^N \frac{f(V_k) + f(1 - V_k)}{2}$$

dove le  $V_k$  sono v.a. indipendenti uniformi su  $(0, 1)$ , cioè qui  $X_k = (f(V_k) + f(1 - V_k))/2$ . Infatti,

$$\mathbb{E}\left(\frac{f(U_k) + f(1 - U_k)}{2}\right) = \int_0^1 \left(\frac{f(x) + f(1 - x)}{2}\right) dx;$$

effettuando il cambio di variabile  $t = 1 - x$ ,  $\int_0^1 f(1 - x) dx = \int_0^1 f(t) dt$  e sostituendo,

$$\mathbb{E}\left(\frac{f(V_k) + f(1 - V_k)}{2}\right) = \int_0^1 f(x) dx.$$

Quindi la LGN garantisce che  $\bar{X}_N$  converge a  $\int_0^1 f(x) dx$  con la scelta  $X_k = (f(V_k) + f(1 - V_k))/2$ .

Si può anche dimostrare che la varianza qui è più piccola di quella vista con le precedenti scelte di  $X_k$ . Infatti, ponendo  $\mu = \int_0^1 f(x) dx$ ,

$$\begin{aligned} \text{Var}\left(\frac{f(V_k) + f(1 - V_k)}{2}\right) &= \mathbb{E}\left(\left(\frac{f(V_k) + f(1 - V_k)}{2}\right)^2\right) - \mu^2 \\ &= \int_0^1 \left(\frac{f(x) + f(1 - x)}{2}\right)^2 dx - \frac{1}{4} \int_0^1 (f^2(x) + f^2(1 - x) + 2f(x)f(1 - x)) dx \\ &= \frac{1}{2} \left( \int_0^1 f^2(x) dx + \int_0^1 f(x)f(1 - x) dx \right) \end{aligned}$$

dove si è usata l'uguaglianza  $\int_0^1 f^2(1 - x) dx = \int_0^1 f^2(x) dx$  (immediata conseguenza del cambio di variabile  $x \mapsto 1 - x$ ). Ora, usando la disuguaglianza di Cauchy-Schwarz per gli integrali<sup>8</sup>, si ottiene

$$\int_0^1 f(x)f(1 - x) dx \leq \left(\int_a^b f(x)^2 dx\right)^{1/2} \left(\int_a^b f(1 - x)^2 dx\right)^{1/2} = \int_a^b f(x)^2 dx.$$

<sup>8</sup>Cioè.  $\int_a^b |g_1(x)g_2(x)| dx \leq \left(\int_a^b |g_1(x)|^2 dx\right)^{1/2} \left(\int_a^b |g_2(x)|^2 dx\right)^{1/2}$ .

Allora,

$$\begin{aligned}\text{Var}\left(\frac{f(V_k) + f(1 - V_k)}{2}\right) &= \frac{1}{2}\left(\int_0^1 f^2(x)dx + \int_0^1 f(x)f(1-x)dx\right) - \mu^2 \\ &\leq \int_0^1 f^2(x)dx - \mu^2 = \text{Var}(f(U_k))\end{aligned}$$

dove  $U_k$  denota una v.a.  $\text{Un}(0, 1)$ . Quindi quest'ultimo metodo è più veloce di quelli già descritti, e infatti nella pratica è quello che solitamente viene usato.

**Esercizio 2.3.4.** *Implementare il metodo delle variabili antitetiche, ad esempio per calcolare numericamente  $\pi$ .*

## 2.4 Lista degli esercizi/problemi relativi al Capitolo 2

- **Problema 2.1.1** Generazione di uno schema di Bernoulli.
- **Problema 2.1.2** Generazione di v.a. binomiali.
- **Problema 2.1.3** Generazione di v.a. geometriche.
- **Esercizio 2.1.4** Lancio di una moneta.
- **Esercizio 2.1.5** Calcolo numerico della probabilità che escano 6 teste consecutive su 100 lanci di una moneta.
- **Esercizio 2.1.6** Generazione di v.a. di Poisson, I: approssimazione di binomiali.
- **Problema 2.2.1** Generazione di v.a. finite.
- **Esercizio 2.2.2** Generazione di v.a. uniformi su un insieme finito.
- **Esercizio 2.2.3** Simulazione del lancio di un dado.
- **Problema 2.3.1** Calcolo numerico di integrali, I: metodo “standard”.
- **Problema 2.3.2** Approssimazione di aree: metodo del rigetto.
- **Problema 2.3.3** Calcolo numerico di integrali, II: metodo del rigetto.
- **Esercizio 2.3.4** Calcolo numerico di integrali, III: metodo delle variabili antitetiche.



## Capitolo 3

# Simulazione di variabili aleatorie continue

In questo capitolo studiamo alcune v.a. assolutamente continue e, in particolare, presentiamo gli algoritmi di simulazione per v.a. esponenziali, con molte conseguenze (v.a. di Weibull, Gamma, di Poisson, geometriche), per v.a. di Cauchy e gaussiane.

Cominciamo una proposizione che si rivela particolarmente utile nell'ambito della simulazione di v.a. assolutamente continue.

**Proposizione 3.0.1.** *Sia  $X$  una v.a. continua con funzione di distribuzione  $F_X$ . Si supponga l'esistenza di due numeri  $a, b$ , con  $-\infty \leq a < b \leq +\infty$ , tali che:*

- i) se  $a > -\infty$  allora  $F_X(x) = 0$ , per ogni  $x \leq a$ ; se  $b < +\infty$  allora  $F_X(x) = 1$ , per ogni  $x \geq b$ ;*
- ii) esiste una funzione  $G : (0, 1) \rightarrow (a, b)$  tale che  $F_X \circ G(y) = y$  per ogni  $y \in (0, 1)$  e  $G \circ F_X(x) = x$  per ogni  $x \in (a, b)$  (cioè  $F_X$  è invertibile su  $(a, b)$  e  $G = F_X^{-1}$ ).*

*Sia  $Z \sim \text{Un}(0, 1)$ . Allora la v.a.*

$$Y = G(Z)$$

*si distribuisce come  $X$ , cioè  $F_Y(\xi) = F_X(\xi)$ , per ogni  $\xi \in \mathbb{R}$ .*

**Dimostrazione.**  $F_X(x)$  è la funzione di distribuzione di una v.a. continua, quindi  $F_X$  è una funzione continua, non decrescente, che tende a 0 per  $x \rightarrow -\infty$  e tende a 1 per  $x \rightarrow +\infty$ . Le ipotesi *i)* e *ii)* garantiscono quindi che l'immagine di  $F_X$  su  $(a, b)$  è l'intervallo  $(0, 1)$  e che almeno su  $(a, b)$  di  $\mathbb{R}$ ,

$F_X$  è invertibile, dunque strettamente crescente, con inversa  $G$ , definita su  $(0, 1)$ , che a sua volta è continua e strettamente crescente. Allora, tenendo presente che  $F_X \circ G = \text{Id}$ ,

$$\begin{aligned} F_Y(\xi) &= \mathbb{P}(Y \leq \xi) = \mathbb{P}(G(Z) \leq \xi) = \mathbb{P}(F_X \circ G(Z) \leq F_X(\xi)) \\ &= \mathbb{P}(Z \leq F_X(\xi)) = F_Z(F_X(\xi)) = F_X(\xi) \end{aligned}$$

perché  $Z \sim \text{Un}(0, 1)$  e  $F_X(\xi) \in [0, 1]$ , per ogni  $\xi \in \mathbb{R}$ . Quindi  $F_Y(\xi) = F_X(\xi)$ , per ogni  $\xi$ . Inoltre, si noti che  $f_Y(\xi) = \frac{d}{d\xi} F_Y(\xi) = \frac{d}{d\xi} F_X(\xi) = p_X(\xi)$ , per ogni  $\xi$  per cui tali derivate esistono, dunque  $X$  e  $Y$  hanno ovviamente anche la stessa densità.

□

Dunque, se  $X$  denota una v.a. con funzione di distribuzione  $F_X$  invertibile su un intervallo  $(a, b)$  di  $\mathbb{R}$  e se  $G$  denota l'inversa, la Proposizione 3.0.1 garantisce che  $G(Z)$  ha la stessa legge di  $X$  se  $Z \sim \text{Un}(0, 1)$ : ciò significa che per simulare  $X$  basterà simulare una  $Z \sim \text{Un}(0, 1)$  e considerare  $G(Z)$  come possibile realizzazione di  $X$ . Ovviamente  $G$  deve potersi scrivere in forma esplicita, il che, purtroppo, non è sempre vero (come ad esempio nel caso gaussiano, si veda il Paragrafo 3.2).

### 3.1 Simulazione di esponenziali e conseguenze

**Problema 3.1.1. a)** Usando la Proposizione 3.0.1, scrivere una procedura per simulare una v.a. esponenziale di parametro  $\lambda$ , con  $\lambda$  dato in input.

**b)** A partire dalla simulazione di esponenziali è possibile generare un certo numero di v.a. di interesse.

**b1) Legge di Weibull.** Una v.a.  $X$  si dice di Weibull con parametri  $\alpha, \lambda > 0$  se la sua densità è<sup>1</sup>

$$p(x) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha} \mathbb{I}_{x>0}.$$

Dimostrare che se  $T \sim \text{Exp}(\lambda)$  allora  $T^{1/\alpha}$  è una v.a. di Weibull di parametri  $\alpha$  e  $\lambda$ ; usare tale risultato per scrivere una procedura di simulazione per una v.a. di Weibull, con parametri  $\alpha$  e  $\lambda$  dati in input.

**b2) Legge Gamma.** Una v.a.  $X$  si dice con legge Gamma di parametri  $k = 1, 2, \dots$  e  $\lambda > 0$  (in simboli:  $\Gamma(k, \lambda)$ ) se la sua densità è

$$p(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} \mathbb{I}_{x>0}.$$

---

<sup>1</sup>La legge di Weibull, così come la legge esponenziale, si rivelano particolarmente utili in teoria dell'affidabilità. Rimandiamo all'Appendice 3.4 alcuni complementi sulle leggi di Weibull ed esponenziale.

Dimostrare che se  $T_1, \dots, T_k$  sono i.i.d. con legge  $\text{Exp}(\lambda)$  allora  $T_1 + \dots + T_k \sim \Gamma(k, \lambda)$ ; usare tale risultato per scrivere una procedura di simulazione per una v.a. Gamma, con parametri  $k = 1, 2, \dots$  e  $\lambda > 0$  dati in input.

**b3) Legge di Poisson.** Una v.a.  $X$  si dice di Poisson con parametro  $\lambda$  se è (discreta) con distribuzione (discreta)

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Dimostrare che se  $T_1, \dots, T_n, \dots$  sono i.i.d. con legge  $\text{Exp}(\lambda)$  allora

$$N_t = \sum_{n=0}^{\infty} \mathbb{I}_{T_1 + \dots + T_n \leq t} = \begin{cases} 0 & \text{se } T_1 > t \\ \max\{n \geq 1 : T_1 + \dots + T_n \leq t\} & \text{altrimenti} \end{cases}$$

è una v.a. di Poisson di parametro  $\lambda t$  (detto processo di Poisson<sup>2</sup>); usare tale risultato per scrivere una procedura di simulazione per una v.a. di Poisson, con parametro  $\lambda > 0$  dato in input. Confrontare quest'algoritmo con quello proposto nell'Esercizio 2.1.6.

**b4) Legge Geometrica.** Dimostrare che se  $T \sim \text{Exp}(\lambda)$  allora  $[T] \sim \text{Ge}(p_\lambda)$  ( $[\cdot]$  denota la parte intera), con  $p_\lambda = 1 - e^{-\lambda}$ ; usare tale risultato per scrivere una procedura di simulazione per una v.a. geometrica, con parametro  $p \in (0, 1)$  dato in input. Confrontare questo algoritmo con quello suggerito nel Problema 2.1.3.

**Soluzione. a)** La f.d. di una v.a. esponenziale è  $F(x) = (1 - e^{-\lambda x}) \mathbb{I}_{x > 0}$ , quindi su  $(0, +\infty)$   $F$  è invertibile con inversa  $G(y) = -\frac{1}{\lambda} \ln(1 - y)$ ,  $y \in (0, 1)$ . Usiamo allora la Proposizione 3.0.1: se  $Z \sim \text{Un}(0, 1)$  allora  $G(Z) \sim \text{Exp}(\lambda)$ , il che dà un modo di generare una v.a. esponenziale:

```
Z:=nrandom;
T:=-ln(1-Z)/lambda
```

Chiameremo `esponenziale(lambda)` una funzione che genera una v.a. di legge esponenziale, con parametro `lambda`.

**b1)** Poiché  $T \sim \text{Exp}(\lambda)$ , in particolare  $T > 0$ , dunque  $X = T^{1/\alpha}$  è ben posta. Inoltre  $X > 0$ , dunque  $F_X(x) = \mathbb{P}(X \leq x) = 0$  per  $x \leq 0$ . Se invece  $x > 0$ , si ha

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(T^{1/\alpha} \leq x) = \mathbb{P}(T \leq x^\alpha) = F_T(x^\alpha)$$

---

<sup>2</sup>Il processo di Poisson è particolarmente importante perché utilizzato come modello in grado di descrivere, ad esempio, processi di arrivo, quindi *file d'attesa* e *code*. Per chi ne volesse sapere di più, si rimanda a Billingsley [3] o anche Feller [4].

quindi

$$f_X(x) = \frac{d}{dx} F_T(x^\alpha) = f_T(x^\alpha) \cdot \alpha x^{\alpha-1} = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha} \mathbb{I}_{x>0}$$

cioè  $X$  è una v.a. di Weibull di parametri  $\alpha$  e  $\lambda$ .

Possiamo quindi scrivere

(\* GENERAZIONE DI UNA v.a. X DI WEIBULL \*)  
 (\* CON PARAMETRI alpha E lambda \*)

T:=esponenziale(lambda);

X:=exp(ln(T)/alpha)

**b2)** Dimostriamo, per induzione su  $k$ , che se  $T_1, \dots, T_k$  sono i.i.d. con legge  $\text{Exp}(\lambda)$  allora  $T_1 + \dots + T_k \sim \Gamma(k, \lambda)$ .

Ricordiamo che la densità di una  $\Gamma(k, \lambda)$  è  $p(x) = \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} \mathbb{I}_{x>0}$ . Se  $k = 1$ , è vera perché  $\Gamma(1, \lambda) = \text{Exp}(\lambda)$ . Supponiamo quindi sia vera per  $k > 1$  e dimostriamo che rimane vera per  $k + 1$ . Si noti che  $T_1 + \dots + T_k + T_{k+1} = (T_1 + \dots + T_k) + T_{k+1}$ . Ora,  $T_1 + \dots + T_k \sim \Gamma(k, \lambda)$  per ipotesi induttiva,  $T_{k+1} \sim \text{Exp}(\lambda)$  per ipotesi ed inoltre  $T_1 + \dots + T_k$  e  $T_{k+1}$  sono indipendenti: la densità di  $T_1 + \dots + T_k + T_{k+1}$  è data dalla convoluzione della densità di  $T_1 + \dots + T_k$  e della densità di  $T_{k+1}$  (cfr. Proposizione 1.4.3):

$$\begin{aligned} p_{T_1+\dots+T_k+T_{k+1}}(t) &= \int_{\mathbb{R}} p_{T_1+\dots+T_k}(s) p_{T_{k+1}}(t-s) ds \\ &= \int_{\mathbb{R}} \frac{\lambda^k}{(k-1)!} s^{k-1} e^{-\lambda s} \mathbb{I}_{s>0} \lambda e^{-\lambda(t-s)} \mathbb{I}_{t-s>0} ds \\ &= \mathbb{I}_{t>0} \cdot \frac{\lambda^{k+1}}{(k-1)!} e^{-\lambda t} \int_0^t s^{k-1} ds = \mathbb{I}_{t>0} \cdot \frac{\lambda^{k+1}}{(k-1)!} e^{-\lambda t} \left. \frac{s^k}{k} \right|_0^t \\ &= \frac{\lambda^{k+1}}{k!} t^k e^{-\lambda t} \mathbb{I}_{t>0} \end{aligned}$$

il che prova la tesi.

Possiamo quindi scrivere

(\* GENERAZIONE DI UNA v.a. X Gamma(k,lambda) \*)

X:=0;

for i:=1 to k do X:=X+esponenziale(lambda)

**b3)** Dimostriamo che se  $T_1, \dots, T_n, \dots$  sono i.i.d. con legge  $\text{Exp}(\lambda)$  allora

$$N_t = \sum_{n=0}^{\infty} \mathbb{I}_{T_1+\dots+T_n \leq t} = \begin{cases} 0 & \text{se } T_1 > t \\ \max\{n \geq 1 : T_1 + \dots + T_n \leq t\} & \text{altrimenti} \end{cases}$$

è un *processo di Poisson*: per ogni  $t > 0$  fissato,  $N_t \sim \text{Po}(\lambda t)$ .

Poiché le  $T_i$  sono esponenziali e quindi v.a. positive,

$$\{N_t = 0\} = \{T_1 > t\} \quad \{N_t \geq j\} = \{T_1 + \dots + T_j \leq t\}, \quad j \geq 1.$$

Quindi

$$\begin{aligned} \mathbb{P}(N_t = 0) &= \mathbb{P}(T_1 > t) = e^{-\lambda t} \quad e \\ \mathbb{P}(N_t = k) &= \mathbb{P}(N_t \geq k) - \mathbb{P}(N_t \geq k + 1) \\ &= \mathbb{P}(T_1 + \dots + T_k \leq t) - \mathbb{P}(T_1 + \dots + T_{k+1} \leq t). \end{aligned}$$

Ora, per  $k \geq 1$ , ricordando che  $T_1 + \dots + T_j \sim \Gamma(j, \lambda)$  e integrando per parti, si ha

$$\begin{aligned} \mathbb{P}(T_1 + \dots + T_k \leq t) &= \frac{\lambda^k}{(k-1)!} \int_0^t s^{k-1} e^{-\lambda s} ds \\ &= \frac{\lambda^k}{(k-1)!} \left( \frac{s^k}{k} e^{-\lambda s} \Big|_0^t - \int_0^t \frac{s^k}{k} (-\lambda) e^{-\lambda s} ds \right) \\ &= \frac{(\lambda t)^k}{k!} e^{-\lambda t} + \frac{\lambda^{k+1}}{k!} \int_0^t s^k e^{-\lambda s} ds = \frac{(\lambda t)^k}{k!} e^{-\lambda t} + \mathbb{P}(T_1 + \dots + T_{k+1} \leq t), \end{aligned}$$

da cui segue che  $\mathbb{P}(T_1 + \dots + T_{k+1} \leq t) - \mathbb{P}(T_1 + \dots + T_k \leq t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$ , quindi  $\mathbb{P}(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$  per ogni  $k \geq 0$  intero, cioè  $N_t \sim \text{Po}(\lambda t)$ .

Quindi, per generare una v.a. di legge  $\text{Po}(\lambda)$  basta scegliere  $t = 1$  e scrivere, ad esempio,

(\* GENERAZIONE DI UNA v.a. X, DI LEGGE  $\text{Po}(\lambda)$  \*)

```
max_indice:=0;
flag:=0;
somma:=0;
while (flag=0) do
  begin
    somma:=somma+esponenziale(lambda);
    if (somma>1) then flag=1;
    max_indice:=max_indice+1
  end;
X:=max_indice-1
```

Ovviamente, il confronto con l'Esercizio 2.1.6 non regge: questa procedura simula esattamente una v.a. di Poisson, mentre il precedente algoritmo simula solo un'approssimazione.

**b4)** Dimostriamo che se  $T \sim \text{Exp}(\lambda)$  allora  $[T] \sim \text{Ge}(p_\lambda)$  ( $[\cdot]$  denota la parte intera), con  $p_\lambda = 1 - e^{-\lambda}$ .

Per  $k \geq 0$  intero, si ha:

$$\begin{aligned} \mathbb{P}([T] = k) &= \mathbb{P}(k \leq T < k + 1) = \int_k^{k+1} \lambda e^{-\lambda x} dx = -e^{-\lambda(k+1)} + e^{-\lambda k} \\ &= (1 - e^{-\lambda}) \left( e^{-\lambda} \right)^k = p_\lambda (1 - p_\lambda)^k, \end{aligned}$$

cioè  $[T] \sim \text{Ge}(p_\lambda)$ .

Quindi, per generare una v.a. geometrica di parametro  $p \in (0, 1)$  basta considerare una v.a. esponenziale di parametro  $\lambda$  tale che  $p_\lambda = p$ , cioè  $\lambda = -\ln(1 - p)$ :

(\* GENERAZIONE DI UNA v.a. X, DI LEGGE Ge(p) \*)

```
lambda:=-ln(1-p);
T:=esponenziale(lambda);
X:=Int(T)
```

dove **Int** denota una funzione che restituisce la parte intera dell'argomento.

La velocità di esecuzione del programma costruito a partire dall'algoritmo suggerito nel Problema 2.1.3 dipende strettamente da  $p$ . In quel caso infatti la v.a. è generata come il primo successo (-1) in una serie di prove bernoulliane con probabilità di successo pari a  $p$ . Quindi, più  $p$  è piccolo, più è lento; più  $p$  è grande, più è veloce. Invece, la velocità di esecuzione del programma basato sull'algoritmo suggerito qui non dipende da  $p$ .

◇

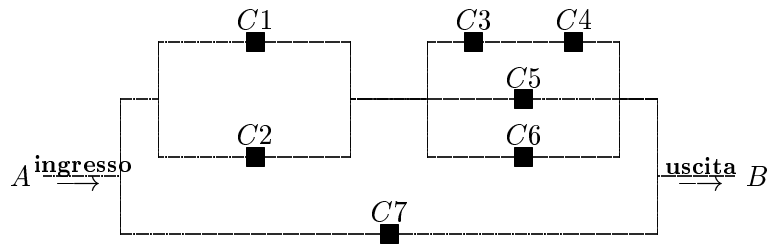
Lasciamo per esercizio la simulazione di un *sistema di vita*.

**Esercizio 3.1.2. Simulazione del tempo di vita di un sistema a più componenti.**

*Un sistema in rete è costituito da 7 componenti  $C_1, C_2, \dots, C_7$ , posti come in Figura 3.1. Diremo che il sistema è in funzione se esiste un cammino da  $A$  a  $B$  che non incontra alcun componente guasto.*

*Siano  $T_1, T_2, \dots, T_7$  i tempi di vita (aleatori!) delle unità  $C_1, C_2, \dots, C_7$  e  $S$  il tempo di vita del sistema:*

$$S = \left[ (T_1 \vee T_2) \wedge \left( (T_3 \wedge T_4) \vee T_5 \vee T_6 \right) \right] \vee T_7 \quad (3.1)$$



**Figura 3.1** Schema del sistema di vita dell'Esercizio 3.1.2 (si immagini, ad esempio, un sistema elettrico, con la corrente che entra in A ed esce in B, e si pensi ai 7 componenti come a 7 lampadine: il sistema è *in funzione* se questo strano lampadario dà luce).

È noto che  $T_1, T_2, T_3, T_5$  sono indipendenti e che  $T_4 = T_1$ ,  $T_6 = T_2/2$ ,  $T_7 = 3T_3$ .

Si suppone inoltre che

- (i)  $T_1, T_2 \sim \text{Ge}(p)$  e  $T_3, T_5 \sim \text{Ge}(q)$ , con  $p, q > 0$  dati in input, oppure che
- (ii)  $T_1, T_2 \sim \text{Exp}(\lambda)$  e  $T_3, T_5 \sim \text{Exp}(\mu)$ , con  $\lambda, \mu > 0$  dati in input.

- a) Scrivere una procedura per generare il tempo di vita  $S$  del sistema<sup>3</sup>.
- b) Scrivere un un programma per stimare la durata del sistema.

### 3.2 Simulazione di variabili aleatorie di Cauchy e gaussiane

#### Problema 3.2.1. Distribuzione di Cauchy

Una v.a. di Cauchy ha densità

$$p(x) = \frac{1}{\pi(1+x^2)}.$$

Usando la Proposizione 3.0.1, scrivere una procedura per simulare una v.a. di Cauchy. Verificare, teoricamente ed empiricamente, che una v.a. di Cauchy non ha media.

**Soluzione.** Calcoliamo la funzione di distribuzione di una v.a. di Cauchy:

$$F(x) = \int_{-\infty}^x \frac{1}{\pi(1+x^2)} dx = \lim_{a \rightarrow -\infty} \frac{1}{\pi} \arctan x \Big|_a^x = \frac{1}{\pi} \left( \arctan x + \frac{\pi}{2} \right).$$

<sup>3</sup>Basta generare  $X_1, X_2, X_3, X_4$  indipendenti, tali che (i)  $X_1, X_2 \sim \text{Ge}(3/4)$  e  $X_3, X_4 \sim \text{Ge}(2/3)$  oppure (ii)  $X_1, X_2 \sim \text{Exp}(\lambda)$  e  $X_3, X_4 \sim \text{Exp}(\mu)$ , e poi porre:  $T_1 = X_1, T_2 = X_2, T_3 = X_3, T_5 = X_4$  e di conseguenza  $T_4, T_6, T_7$ . Calcolare poi i massimi ed i minimi coinvolti nella formula (3.1) per ottenere una realizzazione di  $S$ .

In particolare,  $F$  è invertibile, con inversa  $G(y) = \tan(\pi y - \pi/2)$ ,  $y \in (0, 1)$ . Quindi, per la simulazione basta usare la Proposizione 3.0.1:

(\* GENERAZIONE DI UNA v.a.  $X$  di Cauchy \*)

```
Z:=nrandom;
X:=tan(Pi*Z-Pi/2)
```

dove `tan` denota una funzione che dà la tangente dell'argomento e `Pi` è la funzione che dà  $\pi$ .

Mostriamo che una v.a. di Cauchy non ha media (teorica), mostriamo cioè che la funzione  $|x|p(x)$  non è integrabile. Poiché si tratta di una funzione pari, basta mostrare che non esiste finito il limite per  $b \rightarrow \infty$  di  $\int_0^b |x|p(x) dx$ . Infatti,

$$\lim_{b \rightarrow +\infty} \int_0^b |x|p(x) dx = \lim_{b \rightarrow +\infty} \frac{1}{\pi} \int_0^b \frac{x}{1+x^2} dx = \frac{1}{2\pi} \lim_{b \rightarrow +\infty} \ln(1+x^2) \Big|_0^b = +\infty$$

quindi una v.a. di Cauchy non ha media.

Per la verifica empirica, generare più volte (ad esempio, 10 volte) la media empirica su  $N$  (grande) v.a. di Cauchy indipendenti ed osservare che queste stime sono molto instabili.

◇

### Problema 3.2.2. Distribuzione gaussiana: algoritmo di Box-Muller

*Il procedimento per la simulazione di v.a. continue, usato nei precedenti problemi 3.1.1 e 3.2.1 (basato sulla Proposizione 3.0.1) è di difficile realizzazione pratica nel caso in cui la funzione di distribuzione non si riesce a scrivere in modo esplicito. Un esempio è dato dal caso gaussiano, dov'è noto che non è possibile scrivere esplicitamente né la funzione di distribuzione né la sua inversa. Tuttavia, date due variabili aleatorie  $Z_1, Z_2$  distribuite uniformemente in  $(0, 1)$ , si può dimostrare che le variabili aleatorie*

$$\begin{aligned} X_1 &= \sqrt{-2 \log Z_1} \cos(2\pi Z_2) \\ X_2 &= \sqrt{-2 \log Z_1} \sin(2\pi Z_2) \end{aligned}$$

*sono variabili normali standard indipendenti. Usare questo risultato per scrivere una procedura per generare una v.a.  $N(0, 1)$  e più in generale  $N(m, \sigma^2)$ ,  $m \in \mathbb{R}$ ,  $\sigma \neq 0$ .*

**Soluzione.** Basta trascrivere quanto suggerito da Box-Muller. Se la gaussiana è  $N(m, \sigma^2)$ , ricordiamo (cfr. Proposizione 1.5.2) che  $X = m + \sigma X_{st} \sim N(m, \sigma^2)$  se  $X_{st} \sim N(0, 1)$ :



```

(* generazione di una v.a. *)
(* - X_st di legge N(0,1) *)
(* - X di legge N(m,sigma2) *)

Z1:=nrandom;
Z2:=nrandom;
X_st:=sqrt(-2*ln(Z1))*cos(2*Pi*Z2);
X:=m+sqrt(sigma2)*X_st

```

dove `cos` e `sqrt` denotano le funzioni coseno e radice quadrata rispettivamente. Indicheremo con `normale(m,sigma2)` una funzione che, prendendo in *input* i parametri `m` e `sigma2`, restituisce una gaussiana di media  $m = m$  e varianza  $\sigma^2 = \text{sigma2}$ .

◇

Nel problema che segue, usiamo l'algoritmo di simulazione delle gaussiane per studiare la probabilità condizionata.

**Problema 3.2.3. a)** *Sia  $U$  una v.a. e  $\Lambda$  un insieme fissato tale che  $\mathbb{P}(U \in \Lambda) \in (0, 1)$ . Siano  $U_1, U_2, \dots$  delle copie indipendenti di  $U$  e*

$$\tau = \inf\{k \geq 1 : U_k \in \Lambda\}.$$

*Posto  $W = U_\tau$ , dimostrare che<sup>4</sup> per ogni  $A$ ,*

$$\mathbb{P}(W \in A) = \mathbb{P}(U \in A | U \in \Lambda),$$

*(cioè: la  $\tau$ -esima copia di  $U$ , ovvero la copia di  $U$  che per la prima volta si osserva in  $\Lambda$ , si distribuisce seguendo la distribuzione condizionata di  $U$  dato  $\{U \in \Lambda\}$ ).*

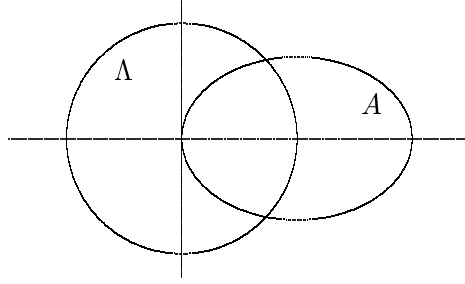
**b)** *Siano  $X$  e  $Y$  due v.a. indipendenti, gaussiane standard. Sia  $A$  il dominio di  $\mathbb{R}^2$  delimitato dall'ellisse di equazione  $(x - 1)^2 + 2y^2 = 1$  e sia  $\Lambda$  il disco unitario centrato nell'origine (si veda la Figura 3.2). Calcolare numericamente  $\mathbb{P}((X, Y) \in A | (X, Y) \in \Lambda)$ .*

**Soluzione. a)** Osserviamo intanto che  $\tau$  è il tempo di primo successo in prove bernoulliane ripetute, dove qui il successo equivale all'osservazione  $U_k \in \Lambda$ . Ora, usando la formula delle probabilità totali,

$$\mathbb{P}(W \in A) = \mathbb{P}(U_\tau \in A) = \sum_{k=1}^{\infty} \mathbb{P}(U_\tau \in A, \tau = k)$$

---

<sup>4</sup>Potrebbe essere utile ricordare che  $\sum_{k=0}^{\infty} \rho^k = 1/(1 - \rho)$ , per ogni  $\rho \in (-1, 1)$ .



**Figura 3.2** Gli insiemi  $\Lambda$  e  $A$  del Problema 3.2.3, parte **b**).

$$\begin{aligned}
&= \sum_{k=1}^{\infty} \mathbb{P}(U_k \in A, \tau = k) = \sum_{k=1}^{\infty} \mathbb{P}(U_k \in A, U_1 \notin \Lambda, \dots, U_{k-1} \notin \Lambda, U_k \in \Lambda) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(U_1 \notin \Lambda, \dots, U_{k-1} \notin \Lambda, U_k \in A \cap \Lambda) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(U_1 \notin \Lambda) \cdots \mathbb{P}(U_{k-1} \notin \Lambda) \mathbb{P}(U_k \in A \cap \Lambda) \\
&= \sum_{k=1}^{\infty} \mathbb{P}(U \notin \Lambda)^{k-1} \mathbb{P}(U \in A \cap \Lambda)
\end{aligned}$$

Posto momentaneamente  $\rho = \mathbb{P}(U \notin \Lambda) \in (0, 1)$ , effettuando il cambio di variabile  $j = k - 1$  nell'indice della sommatoria si ottiene  $\sum_{k=1}^{\infty} \rho^{k-1} = \sum_{j=0}^{\infty} \rho^j = 1/(1 - \rho)$ . Allora,

$$\begin{aligned}
\mathbb{P}(W \in A) &= \frac{1}{1 - \mathbb{P}(U \notin \Lambda)} \mathbb{P}(U \in A \cap \Lambda) \\
&= \frac{\mathbb{P}(U \in A \cap \Lambda)}{\mathbb{P}(U \in \Lambda)} = \mathbb{P}(U \in A | U \in \Lambda).
\end{aligned}$$

**b)** Usiamo la procedura suggerita nel punto **a**): qui la v.a. di interesse  $U$  equivale alla coppia  $(X, Y)$  di gaussiane indipendenti. Simuliamo quindi tante v.a. indipendenti  $X_k$  e  $Y_k$  gaussiane standard fino a che osserviamo che  $(X_k, Y_k) \in \Lambda$ , il che è vero se  $X_k^2 + Y_k^2 \leq 1$ . Sia  $\tau$  il primo valore di  $k$  per il quale si osserva che la coppia appartiene a  $\Lambda$ . A questo punto, osserviamo se  $(X_\tau, Y_\tau) \in A$ , cioè se  $(X_\tau - 1)^2 + 2Y_\tau^2 \leq 1$ , o meglio consideriamo la v.a.  $\mathbb{I}_{(X_\tau, Y_\tau) \in A}$ . Poiché

$$\mathbb{E}(\mathbb{I}_{(X_\tau, Y_\tau) \in A}) = \mathbb{P}((X_\tau, Y_\tau) \in A) = \mathbb{P}(X \in A | X \in \Lambda),$$

approssimiamo la probabilità richiesta con la media aritmetica valutata su  $N$  generazioni di  $\mathbb{I}_{(X_\tau, Y_\tau) \in A}$ :

proba:=0;

```

for i:=1 to N do
  begin
    flag:=0;
    while (flag=0) do
      begin
        X:=normale(0,1);
        Y:=normale(0,1);
        if (sqr(X)+sqr(Y)<=1) then flag:=1
      end;
      if (sqr(X-1)+2*sqr(Y)<=1) then proba:=proba+1
    end;
  end;
proba:=proba/N

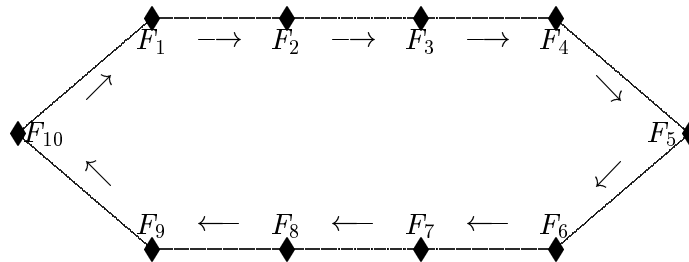
```

Va detto che se anche fossimo in possesso dei prerequisiti necessari per calcolare questo tipo di probabilità, ci accorgeremmo che non è possibile scriverla in modo esplicito, dunque non si può fare a meno di usare un metodo numerico per calcolarla.

◇

### 3.3 Simulazione di una linea tramviaria

**Problema 3.3.1.** *Una linea tramviaria ha la forma di un circuito nel quale si trovano 10 fermate  $F_1, F_2, \dots, F_{10}$ , come in Figura 3.3.*



**Figura 3.3** Schema della linea tramviaria dell'Esercizio 3.3.1.

*Sulla linea circola un unico tram da 50 posti. Si suppone che, a regime,*

- a) *i passeggeri arrivano alle 10 fermate in modo indipendente e il numero dei passeggeri che arrivano alla fermata  $F_i$  tra gli istanti  $s$  e  $t$ , con  $s < t$ , è una v.a. di Poisson di parametro  $(t - s)/30$  (dove  $t$  e  $s$  sono espressi in minuti);*
- b) *quando il tram arriva alla fermata  $F_i$ , mediamente un passeggero su cinque scende, il numero di persone che sale è il massimo possibile delle*

*persone in attesa all'istante in cui arriva il tram, tenuto ovviamente conto dei 50 posti disponibili; tra le persone che eventualmente non riescono a salire, tre su quattro decidono di attendere il tram alla corsa successiva, gli altri invece se ne vanno via;*

- c) ad ogni fermata  $F_i$ , il tram riparte dopo un tempo uguale, in minuti, al massimo delle persone che sono salite e quelle che sono scese diviso per 10;*
- d) il tempo di percorrenza tra due fermate consecutive (ovvero, tra la partenza alla fermata  $F_i$  e l'arrivo alla fermata  $F_{i+1}$ ) è un tempo aleatorio uniformemente distribuito tra gli 8 e i 12 minuti.*

*Scrivere un programma per calcolare il tempo medio di una corsa del tram.*

**Soluzione.** Basterà simulare un numero elevato `Num_corse` di corse complete, calcolare quanto tempo occorre per effettuare le corse e dividerlo per il totale delle corse effettuate.

Ora, fissiamo una corsa  $j$  e supponiamo che il tram sia arrivato alla fermata  $F_i$ . Due variabili indispensabili sono:

- l'istante corrente `T_corrente` che indica l'istante di arrivo alla stazione  $F_i$ ;
- il numero corrente `N_corrente` di passeggeri presenti sul tram all'arrivo alla fermata  $F_i$ .

Per simulare le persone che attendono alla fermata  $F_i$ , quelle che salgono e quelle che scendono, serve conoscere sia l'ultimo istante di passaggio del tram alla stazione  $F_i$  sia quante persone sono rimaste ad attendere il tram alla corsa precedente. Indichiamo dunque con `T_ultimo_passaggio` e `N_ultimo_passaggio` due vettori, con 10 componenti, tali che

- `T_ultimo_passaggio[i]` dà l'istante di ultimo passaggio (cioè alla  $(j - 1)$ -esima corsa) del tram alla stazione  $F_i$ ;
- `N_ultimo_passaggio[i]` dà il numero di persone che sono rimaste in attesa alla corsa precedente  $((j - 1)$ -esima), sempre alla fermata  $F_i$ .

Per la simulazione dell'arresto del tram alla fermata  $F_i$ , quindi del movimento dei passeggeri e del tempo che occorre, le quantità che vanno generate sono:

- il numero `N_in_arrivo` dei clienti che arrivano: è una quantità aleatoria con distribuzione di Poisson di parametro che dipende da quanto tempo è passato dall'ultimo passaggio, cioè da `T_corrente - T_ultimo_passaggio[i]`;

- il numero `N_in_attesa` dei clienti che attendono il tram: è pari al numero `N_ultimo_passaggio[i]` delle persone che alla corsa precedente non sono riuscite a salire e che hanno deciso di attendere il tram più `N_in_arrivo`;
- il numero `N_in_discesa` dei clienti che scendono tram: è una quantità aleatoria con distribuzione binomiale di parametri dati dal numero di persone presenti sul tram, cioè `N_corrente`, e  $1/5$ .

Note le quantità di cui sopra, possiamo aggiornare tutte le variabili:

- `N_corrente` = numero di passeggeri sul tram + quelli che sono saliti – quelli che sono scesi;
- `N_ultimo_passaggio[i]` = v.a. binomiale di parametri  $n$  = numero di passeggeri che non riescono a salire e  $p = 3/4$ , purché  $n > 0$ ;
- `T_corrente` e `T_ultimo_passaggio[i]`: occorre aggiungere il tempo necessario al tram per la salita/discesa dei passeggeri.

Indicando con `Poisson(lambda)` e `binomiale(n,p)` due funzioni che generano, rispettivamente, una v.a. di Poisson di parametro `lambda` e una v.a. binomiale di parametri `n` e `p`, possiamo scrivere la procedura `ARRESTO` ad esempio come segue:

```
(*          ARRESTO all'i-esima fermata          *)

(* calcolo del numero degli arrivi e quindi del numero *)
(*          effettivo di clienti in attesa:          *)

lambda:=(T_corrente-T_ultimo_passaggio[i])/30;
N_in_arrivo:=Poisson(lambda);
N_in_attesa:=N_in_arrivo+N_ultimo_passaggio[i];

(* calcolo del numero dei passeggeri che scendono: *)

N_in_discesa:=binomiale(N_corrente, 0.2);

(* calcolo del numero dei clienti che salgono, *)
(* tenendo conto che il tram ha 50 posti: *)

if (N_corrente-N_in_discesa+N_in_attesa<=50) then
    N_in_salita:=N_in_attesa
else
```

```

        N_in_salita:=50-(N_corrente-N_in_discesa);

(* aggiornamento del numero di passeggeri sul tram *)
(* e di quelli che non sono riusciti a salire: *)

N_corrente:=N_corrente-N_in_discesa+N_in_salita;
N_rimasti:=N_in_attesa-N_in_salita;

(* calcolo delle persone che rimangono in attesa *)
(* al passaggio successivo *)

if (N_rimasti>0) then
    N_ultimo_passaggio[i]:=binomiale(N_rimasti, 0.75);

(* calcolo del tempo necessario per la *)
(* salita/discesa dei passeggeri: *)

if (N_in_salita>N_in_discesa) then
    tempo:=N_in_salita/10
else
    tempo:=N_in_discesa/10;

(* aggiornamento del tempo corrente e *)
(* dell'ultimo istante di passaggio alla fermata: *)

T_corrente:=T_corrente+tempo;
T_ultimo_passaggio[i]:=T_corrente

```

Quindi, la procedura ARRESTO prende in *input* le variabili

$$\begin{array}{ll}
 T_{\text{corrente}}, & N_{\text{corrente}} \\
 T_{\text{ultimo\_passaggio}[i]}, & N_{\text{ultimo\_passaggio}[i]}
 \end{array} \quad (3.2)$$

le elabora e dà in *output* l'aggiornamento delle stesse.

Ora, la simulazione dello spostamento del tram dalla fermata  $F_i$  alla fermata  $F_{i+1}$  è molto semplice e si può scrivere ad esempio

```

tragitto:=uniforme(8,12);
T_corrente:=T_corrente+tragitto

```

dove `uniforme(a,b)` denota una funzione che genera una v.a. uniforme nell'intervallo di estremi  $a$  e  $b$ .

Per poter scrivere il programma principale, occorre inizializzare le variabili in (3.2). Il problema non dà alcuna indicazione, quindi si lascia allo studente la scelta.

Ad esempio, una scelta possibile potrebbe essere:

```
T_corrente:=uniforme(8,12);
N_corrente:=0;
for i:=1 to 10 do
  begin
    T_ultimo_passaggio[i]:=0;
    N_ultimo_passaggio[i]:=0
  end
```

In questo caso, si immagina che il tram parte vuoto da un ipotetico capolinea (che coincide con l'ultima fermata  $F_{10}$ ), per cominciare il suo giro da  $F_1$ ; nessun cliente è (ovviamente) rimasto ad aspettare il tram e, alla prima corsa, gli arrivi sono poissoniani con parametro direttamente proporzionale al tempo necessario per arrivare a ciascuna fermata.

Sia

```
INIZIALIZZA(T_corrente, N_corrente,
            T_ultimo_passaggio, N_ultimo_passaggio)}
```

una procedura che inizializza le variabili (ad esempio, come quella sopra scritta).

Ricapitolando, con le considerazioni fatte finora il programma principale sarebbe:

```
INIZIALIZZA(T_corrente, N_corrente,
            T_ultimo_passaggio, N_ultimo_passaggio);

for j:=1 to Num_corse do      (* <----- corse *)
  begin
    for i:=1 to 10 do          (* <----- fermate *)
      begin
        ARRESTO(T_corrente, N_corrente,
                T_ultimo_passaggio[i], N_ultimo_passaggio[i]);
        tragitto:=uniforme(8,12);
        T_corrente:=T_corrente+tragitto
```

```
        end
    end

Tempo_medio_di_una_corso:=T_corrente/Num_corse
```



### 3.4 Appendice: complementi sulla legge esponenziale e la legge di Weibull

Una v.a. di Weibull di parametri  $\alpha, \lambda > 0$  è una v.a. assolutamente continua  $T$  con densità

$$p(t) = \alpha \lambda t^{\alpha-1} e^{-\lambda t^\alpha} \mathbb{I}_{t>0}.$$

Si noti che se si sceglie  $\alpha = 1$  si ottiene la densità esponenziale.

Le v.a. di Weibull vengono utilizzate per descrivere *tempi di vita (aleatori)*, ad esempio di apparecchiature, o di popolazioni biologiche. Infatti si ha

$$\mathbb{P}(T > 0) = \int_0^{+\infty} p(t) dt = \int_{\mathbb{R}} p(t) dt = 1$$

perché  $p(t) \equiv 0$  per  $t < 0$ . Quindi  $T$  è una v.a. positiva e può dunque essere usata per descrivere dei tempi di vita.

Fissato  $u \in \mathbb{R}$ , è possibile calcolare esplicitamente  $\mathbb{P}(T > u)$  e quindi la funzione di distribuzione  $F_T$  di  $T$ : per  $u \leq 0$ , si ha

$$\mathbb{P}(T > u) = \int_u^{+\infty} p(t) dt = \int_0^{+\infty} p(t) dt = 1$$

perché  $p(t) \equiv 0$  per  $t < 0$ ; se invece  $u > 0$ ,

$$\begin{aligned} \mathbb{P}(T > u) &= \int_u^{+\infty} p(t) dt = \int_u^{+\infty} \alpha \lambda t^{\alpha-1} e^{-\lambda t^\alpha} \mathbb{I}_{t>0} dt \\ &= - \int_u^{+\infty} \frac{d}{dt} \left( e^{-\lambda t^\alpha} \right) dt = - e^{-\lambda t^\alpha} \Big|_{t=u}^{t \uparrow +\infty} = e^{-\lambda u^\alpha}. \end{aligned}$$

Quindi la funzione di distribuzione di  $T$  è

$$F_T(u) = \begin{cases} 0 & \text{se } u \leq 0 \\ 1 - e^{-\lambda u^\alpha} & \text{se } u > 0 \end{cases}$$

Per capire meglio come utilizzare le v.a. di Weibull nella modellistica, fissato  $t > 0$  consideriamo la funzione

$$[0, +\infty) \ni s \mapsto g_t(s) = \mathbb{P}(T > t + s | T > s) = \mathbb{P}(T - s > t | T > s) \in [0, 1].$$

Si noti che

$$\mathbb{P}(T > t + s | T > s) = \frac{\mathbb{P}(T > t + s, T > s)}{\mathbb{P}(T > s)} = \frac{\mathbb{P}(T > t + s)}{\mathbb{P}(T > s)}$$

perché  $\{T > t + s\} \subset \{T > s\}$ . Quindi, usando l'espressione della funzione di distribuzione,

$$g_t(s) = \exp \left\{ -\lambda \left( (t + s)^\alpha - s^\alpha \right) \right\}.$$

Studiamo il segno della derivata di  $g_t$ , al variare di  $s > 0$ :

$$g'_t(s) = -\lambda\alpha g_t(s) \left[ (t+s)^{\alpha-1} - s^{\alpha-1} \right].$$

Allora,  $g'_t(s) \geq 0$  se e solo se  $(t+s)^{\alpha-1} - s^{\alpha-1} \leq 0$ , ovvero

$$\left( \frac{t+s}{s} \right)^{\alpha-1} \leq 1$$

che equivale (si ricordi che  $t, s > 0$ ) a richiedere

$$\alpha \leq 1.$$

Possiamo quindi riassumere come segue:  $g_t(s)$  è

- monotona crescente in  $s > 0$  se e solo se  $\alpha < 1$ ;
- costante in  $s > 0$  se e solo se  $\alpha = 1$ ;
- monotona decrescente in  $s > 0$  se e solo se  $\alpha > 1$ .

Ora, fissato  $t > 0$ ,  $g_t(s) = \mathbb{P}(T-s > t | T > s)$  è la probabilità che l'unità cui  $T$  si riferisce rimanga in vita per un ulteriore tempo  $t$  noto che all'istante  $s$  è funzionante. Se è noto che  $T > s$ , la v.a.  $T-s$  viene usualmente interpretata come il *tempo di vita residuo*, quindi  $g_t(s)$  è la probabilità che il tempo di vita residuo a  $s$  sia maggiore di  $t$ . Se quindi  $g_t$  è decrescente ( $\alpha < 1$ ), tale apparecchiatura dev'essere soggetta ad un fenomeno di tipo usura, ovvero tende a deteriorarsi con il tempo: all'aumentare del tempo ( $s$ ), il tempo di vita residuo tende a diminuire (ad esempio,  $T$  è la durata di una lampadina). Se invece  $g_t$  è crescente ( $\alpha > 1$ ), l'unità in questione tende a migliorare con l'età: all'aumentare del tempo ( $s$ ), il tempo di vita residuo tende ad aumentare (come accade, ad esempio, per certe popolazioni biologiche). Qualora invece  $g_t$  fosse costante ( $\alpha = 1$ ), significherebbe che l'età non influenza il tempo di vita residuo dell'apparecchiatura ("perdita di memoria, che con buona approssimazione si verifica per certe apparecchiature elettroniche).

Si noti infine che per  $\alpha = 1$  allora  $p$  è la densità di una  $\text{Exp}(\lambda)$ : le v.a. esponenziali soddisfano quindi la proprietà di "perdita di memoria, che nel caso discreto è tipica delle v.a. geometriche.

### 3.5 Lista degli esercizi/problemi relativi al Capitolo 3

- **Problema 3.1.1** Simulazione di v.a. esponenziali e di conseguenza con legge: di Weibull, Gamma, Poisson, geometrica.
- **Esercizio 3.1.2** Simulazione del tempo di vita di un sistema a più componenti.
- **Problema 3.2.1** Generazione di v.a. di Cauchy.
- **Problema 3.2.2** Generazione di v.a. gaussiane: algoritmo di Box-Muller.
- **Problema 3.2.3** Studio della legge condizionata.
- **Problema 3.3.1** Simulazione di una linea tramviaria.

# Bibliografia

- [1] P. Baldi *Calcolo delle probabilità e statistica*. Seconda Edizione. McGraw-Hill, 1998.
- [2] P. Baldi, R. Giuliano, L. Ladelli. *Laboratorio di probabilità e statistica*. McGraw-Hill, 1995.
- [3] P. Billingsley *Probability and measure*. Second edition. Wiley series in Probability and Statistics, John Wiley & Sons, 1986.
- [4] W. Feller *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons, Inc., New York-London-Sydney 1968.
- [5] Y.G. Sinai *Probability theory. An introductory course*. Springer, 1992.