un campione normale con varianza nota. Dalle formule (6.12) e (6.11) si ottiene perciò

$$\text{test unilaterale:} \quad c = \frac{\sigma}{\sqrt{n}} q^{\mathcal{N}}_{1-\alpha}, \qquad \text{test bilaterale:} \quad c = \frac{\sigma}{\sqrt{n}} q^{\mathcal{N}}_{1-\alpha/2}. \tag{7.7}$$

Un test di questo tipo è anche detto *Test-Z* (in quanto coinvolge i quantili di una v.a. normale standard, che spesso è indicata con Z).

Vediamo adesso come si modificano le formule precedenti nel caso, molto più comune, in cui la varianza sia incognita. Le ipotesi nulle, le regole di decisione e le regioni di accettazione rimangono le stesse. Quello che cambia sono le formule per i valori critici. Infatti, essendo la varianza incognita, non si può usare il parametro sconosciuto σ^2 ma dobbiamo ricorrere al suo stimatore S^2. Dunque, fissato il livello α, tutto si svolge come nel caso di varianza nota con la differenza che stavolta i valori critici corripondono ai δ degli IdC per la media diun campione normale con varianza incognita e sono perciò dati dalle formule (6.14) e (6.13):

$$\text{test unilaterale:} \quad c = \frac{S}{\sqrt{n}} q^{t(n-1)}_{1-\alpha}, \qquad \text{test bilaterale:} \quad c = \frac{S}{\sqrt{n}} q^{t(n-1)}_{1-\alpha/2}. \tag{7.8}$$

Test di questo tipo prendono il nome di *Test-t*, in quanto coinvolgono la distribuzione t di Student.

7.2.2 Confronto fra medie di due campioni indipendenti

Una situazione molto frequente è quella in cui si vogliono confrontare le medie di due diversi campioni. Consideriamo ad esempio la Tabella 6.1: è lecito domandarsi se la differenza fra le medie di due campioni, raccolti in siti differenti è dovuta soltanto a una fluttuazione statistica o se c'è piuttosto una differenza "sostanziale" tra i due campioni.
Consideriamo quindi, in generale, due campioni normali indipendenti

$$\begin{aligned} \boldsymbol{X} &= (X_1, X_2, \ldots, X_{n_1}), & X_i &\sim \mathcal{N}(\mu_1, \sigma_1^2), \\ \boldsymbol{Y} &= (Y_1, Y_2, \ldots, Y_{n_2}), & Y_i &\sim \mathcal{N}(\mu_2, \sigma_2^2), \end{aligned}$$

con medie incognite. Volendo mettere in luce le differenze fra le due medie, considereremo un'ipotesi nulla di tipo bilaterale:

$$\mathrm{H}_0 : \mu_1 = \mu_2.$$

Se invece ci aspettiamo che una media sia predominante sull'altra, potremmo testare un'ipotesi nulla di tipo unilaterale $\mathrm{H}_0 : \mu_1 < \mu_2$ (lasciamo al lettore il compito di adattare a questo caso le formule che troveremo). L'ipotesi nulla bilaterale è dunque quella che *non* ci siano differenze fra i due campioni. I due parametri μ_1 e μ_2 saranno stimati dalle due medie campionarie che indicheremo, rispettivamente con $\overline{X}$ e $\overline{Y}$. La differenza fra le due stime è data dalla v.a.

$$W = \overline{X} - \overline{Y}$$

ed ha senso utilizzare una regola di decisione del tipo

$$\textit{si rifiuta } \mathrm{H}_0 \textit{ se } |W| > c,$$

ovvero, si conclude che c'e' un'effettiva differenza se la distanza fra le due stime è sufficientemente grande, maggiore di un valore critico c da fissare in base al livello α.
Consideriamo innanzitutto il caso in cui le varianze delle due popolazioni, σ_1^2 e σ_2^2, siano note. Poiché

$$\overline{X} \sim \mathcal{N}\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \qquad \overline{Y} \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2^2}{n_2}\right),$$

allora (si veda la proposizione 4.23)

$$W \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) = \mathcal{N}\left(\mu_1 - \mu_2, \frac{n_2\sigma_1^2 + n_1\sigma_2^2}{n_1 n_2}\right).$$

In particolare, se assumiamo vera l'ipotesi H_0, si avrà che W ha media nulla. Dunque, fissato il livello α, la condizione

$$\alpha = P^{\mu_1=\mu_2}\left(|W| > c\right)$$

ci dice che c è il δ di un IdC bilaterale di livello α per la media di un campione normale con varianza nota $\frac{n_2\sigma_1^2+n_1\sigma_2^2}{n_1 n_2}$. Mediante la solita procedura di standardizzazione si ottiene perciò

$$c = \sqrt{\frac{n_2\sigma_1^2 + n_1\sigma_2^2}{n_1 n_2}}\, q^{\mathcal{N}}_{1-\alpha/2}. \tag{7.9}$$

Se le varianze non sono note le cose diventano più difficili. Se ci aspettiamo che le due varianze (incognite) siano all'incirca uguali,[2] allora un buono stimatore per il valore comune della varianza è la cosiddetta *varianza mediata*:

$$\overline{S}^2 := \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{i=1}^{n_2}(X_i - \overline{Y})^2}{n_1+n_2-2}, \tag{7.10}$$

dove con S_1^2 e S_2^2 abbiamo indicato le varianze campionarie dei due campioni. Poiché si può dimostrare che

$$\frac{W}{\overline{S}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2), \tag{7.11}$$

dalla condizione $\alpha = P^{\mu_1=\mu_2}\left(|W| > c\right)$ si ottiene la formula per il valore critico:

$$c = \overline{S}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\, q^{t(n_1+n_2-2)}_{1-\alpha/2}. \tag{7.12}$$

Consideriamo ad esempio un test di confronto fra le medie dei campioni *C1* e *C4* della tabella (6.1). Le stime di medie e varianze sono riportate nella tabella 6.2: si può notare che le varianze stimate sono simili, per cui si può pensare

[2] Questa è un'ulteriore ipotesi che può essere sottoposta al test-F, studiato nel paragrafo 7.2.3

di applicare il procedimento sopra descritto (verificheremo l'ipotesi di uguale varianza nel paragrafo 7.2.3). Si ha in questo caso

$$\begin{aligned}
&n_1 = n_2 = 20;\\
&\overline{X} = 8.90; \quad \overline{Y} = 7.26; \quad W = 1.64;\\
&S_1^2 = 33.11; \quad S_2^2 = 38.91;\\
&\overline{S}^2 = \frac{19 \times (33.11 + 38.91)}{38} = 36.01\\
&c = \sqrt{\frac{1}{10}} \times \sqrt{36.01} \times q_{0.975}^{t(38)} \approx 1.90 \times q_{0.975}^{\mathcal{N}} = 1.90 \times 1.96 = 3.72
\end{aligned}$$

(dove si è considerato il livello $\alpha = 0.05$ e il fatto che $t(n)$ si può approssimare con $\mathcal{N}$ per $n > 30$). Poiché risulta $|W| < c$, possiamo accettare l'ipotesi nulla $\mu_1 = \mu_2$ con un grado di certezza del 95%.

7.2.3 Confronto fra varianze di due campioni indipendenti

Talvolta è di un certo interesse confrontare le varianze di due diversi campioni. Una situazione di questio tipo l'abbiamo già incontrata nel precedente paragrafo 7.2.2, in cui era necessario verificare l'ipotesi che due campioni avessero la stessa varianza per poter utilizzare lo stimatore "varianza mediata" (7.10).
Consideriamo quindi, di nuovo, due campioni normali indipendenti

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_{n_1}), \qquad \boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_{n_2}),$$

con varianze incognite σ_1^2 e σ_2^2 (e medie incognite). L'ipotesi nulla adatta a mettere in luce differenze fra le due varianze è quella di tipo bilaterale:

$$\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2.$$

Anche qui, se ci aspettiamo che una varianza sia predominante sull'altra, potremmo testare un'ipotesi nulla di tipo unilaterale $\mathrm{H}_0 : \sigma_1^2 < \sigma_2^2$ (come al solito lasciamo al lettore lo svolgimento di questo caso). Come stimatori di σ_1^2 e σ_2^2 utilizziamo le varianze campionarie

$$S_1^2 := \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(X_i - \overline{X}\right)^2, \qquad S_2^2 := \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} \left(Y_i - \overline{Y}\right)^2.$$

Per capire quale regola decisionale è più adatta a questo test utilizziamo il fatto che, *nell'ipotesi nulla*, si ha $\sigma_2^2/\sigma_1^2 = 1$ e quindi possiamo scrivere

$$\frac{S_1^2}{S_2^2} = \frac{n_2 - 1}{n_1 - 1} \frac{\frac{(n_1-1)}{\sigma_1^2} S_1^2}{\frac{(n_2-1)}{\sigma_2^2} S_2^2}.$$

Osserviamo che, per il Teorema di Cochran 6.13,

$$\frac{(n_1 - 1)}{\sigma_1^2} S_1^2 \sim \mathcal{X}^2(n_1 - 1), \qquad \frac{(n_2 - 1)}{\sigma_2^2} S_2^2 \sim \mathcal{X}^2(n_2 - 1),$$

e quindi, per quanto osservato nell'Esempio 4.19, si ha

$$\frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1} \,,$$

dove F_{n_1-1,n_2-1} è la distribuzione F di Fischer con gradi di libertà $n_1 - 1$ e $n_2 - 1$ (si veda il paragrafo 3.7). Conviene perciò utilizzare il rapporto S_1^2/S_2^2 per stimare la diversità delle due varianze e adottare di conseguenza la regola decisionale:

$$\textit{supponendo } S_1^2 \geq S_2^2, \textit{ si rifiuta } \mathrm{H}_0 \textit{ se } \frac{S_1^2}{S_2^2} > c,$$

dove, al solito, c è un valore critico da calcolare in funzione del livello del test α. Ovviamente, se S_2^2 fosse la varianza maggiore, si deve utilizzare $\frac{S_2^2}{S_1^2}$ e la distribuzione F_{n_2-1,n_1-1}. Fissato dunque α, per calcolare c scriveremo

$$P^{\sigma_1^2=\sigma_2^2}\left(\frac{S_1^2}{S_2^2} > c\right)$$

dove sappiamo che, nell'ipotesi $\sigma_1^2 = \sigma_2^2$, la v.a. S_1^2/S_2^2 ha distribuzione F_{n_1-1,n_2-1}. Si ha perciò direttamente

$$c = q_{1-\alpha}^{F_{n_1-1,n_2-1}}. \tag{7.13}$$

Riprendendo l'esempio del paragrafo 7.2.2, se vogliamo testare l'ipotesi $\sigma_1^2 = \sigma_2^2$ sui campioni *C1* e *C4* della tabella 6.1 dobbiamo utilizzare (per il livello $\alpha = 0.05$)

$$\begin{aligned} &n_1 = n_2 = 20; \\ &S_1^2 = 33.11; \quad S_2^2 = 38.91; \quad S_2^2 > S_1^2; \quad S_2^2/S_1^2 = 1.17 \\ &c = q_{0.95}^{F_{19,19}} \approx 2.1\,. \end{aligned}$$

Poiché $S_2^2/S_1^2 < c$, l'ipotesi nulla $\sigma_1^2 = \sigma_2^2$ non può essere respinta. Confrontando, invece, *C1* e *C2* si ha

$$S_1^2 = 33.11; \quad S_2^2 = 1.82; \quad S_1^2 > S_2^2; \quad S_1^2/S_2^2 = 18.19$$

per cui, in questo caso, l'ipotesi nulla può essere rifiutata.
Test di questo tipo, che coinvolgono la distribuzione di Fisher, sono chiamati *Test-F*.

7.3 Test del $\mathcal{X}^2$

Il cosiddetto *test del* $\mathcal{X}^2$ è un utilissimo strumento per confrontare frequenze teoriche e frequenze misurate e verificare così l'aderenza alla realtà di un modello probabilistico con distribuzione discreta. Ad esempio, lanciando la moneta il nostro tipico modello probabilistico è basato sulla v.a. Bernoulliana

$$P(X = 1) = 1/2 \;\; (\text{“testa”}), \qquad P(X = 0) = 1/2 \;\; (\text{“croce”}).$$

Se volessimo testare statisticamente questo nostro modello, dovremmo lanciare tante volte la moneta, ottenendo una sequenza $X_1, X_2, \ldots X_n$ di v.a. Bernoulliane, e confrontare le frequenze relative sperimentali

$$F(\textit{testa}) = \frac{\textit{numero teste}}{\textit{numero lanci}} = \frac{X_1 + X_2 + \cdots + X_n}{n};$$

$$F(\textit{croce}) = \frac{\textit{numero croci}}{\textit{numero lanci}} = 1 - \frac{X_1 + X_2 + \cdots + X_n}{n};$$

con le frequenze relative teoriche che, come ci si aspetta dalla legge dei grandi numeri (Teorema 4.26), sono date da

$$F(\textit{testa}) \approx 1/2; \qquad F(\textit{croce}) \approx 1/2;$$

se il numero di prove è sufficientemente alto.

Consideriamo quindi in generale un campione di rango n, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, con distribuzione discreta. Questo significa che ciascuna X_i assume m possibili valori $x_1, x_2, \ldots, x_m$ con probabilità $p_1, p_2, \ldots, p_m$:

$$P(X_i = x_k) = p_k, \qquad \text{per} \;\; k = 1, 2, , \ldots, m \;\; \text{e} \;\; i = 1, 2, , \ldots, n,$$

con la condizione $\sum_{k=1}^{m} p_k = 1$. Nel contesto del test del $\mathcal{X}^2$ le probabilità p_k sono i *parametri incogniti* della distribuzione. Consideriamo le m variabili aleatorie

$$N_k = \textit{numero delle } X_i \textit{ uguali a } x_k, \qquad k = 1, 2, , \ldots, m, \tag{7.14}$$

dette *frequenze campionarie* e le m variabili aleatorie

$$F_k = \frac{N_k}{n}, \qquad k = 1, 2, , \ldots, m, \tag{7.15}$$

dette *frequenze campionarie relative* (sono esattamente le frequenze assolute e relative introdotte nel paragrafo 5.2 ma stavolta viste come variabili aleatorie). Si può verificare molto facilmente che F_k *è uno stimatore corretto di* p_k.

Osservazione 7.2. Notiamo per inciso che F_k non solo è uno stimatore corretto di p_k ma per di più tende a p_k per $n \to \infty$. Questo lo si può vedere scrivendo

$$N_k = Y_1^{(k)} + Y_2^{(k)} + \cdots + Y_n^{(k)}$$

dove, per ogni $i = 1, 2, , \ldots, n$, si definisce

$$Y_i^{(k)} := \begin{cases} 1, & \text{se } X_i = x_k; \\ 0, & \text{se } X_i \neq x_k. \end{cases}$$

Poiché $E[Y_i^{(k)}] = p_k$, dalla legge dei grandi numeri (Teorema 4.26) si ottiene che $F_k \to p_k$ quando $n \to \infty$. □

Consideriamo ora la v.a.

$$T_n := n \sum_{k=1}^{m} \frac{(F_k - p_k)^2}{p_k} = \sum_{k=1}^{m} \frac{(N_k - np_k)^2}{np_k}, \tag{7.16}$$

che dà una misura dello scarto quadratico delle frequenze teoriche da quelle campionarie. L'indice n serve a ricordare la numerosità del campione. Si ha il seguente risultato, che enunciamo senza dismostrazione.

Teorema 7.3. (Teorema di Pearson)
Se $P(X_i = x_k) = p_k$, la successione di variabili aleatorie $T_1, T_2, T_3, \ldots$ definite dalla (7.16) *tende ad avere distribuzione $\mathcal{X}^2(m-1)$.*[3] *Sinteticamente:*

$$\lim_{n\to\infty} T_n \sim \mathcal{X}^2(m-1). \tag{7.17}$$

[3]Il fatto che i gradi di libertà siano $m-1$ è dovuto al vincolo $\sum_{k=0}^{m} p_k = 1$.

Questo risultato ci permette di concepire un test dell'ipotesi nulla

$$\mathrm{H}_0 \,:\, p_1 = p_1^0,\ p_2 = p_2^0, \ldots,\ p_m = p_m^0,$$

che esprime il fatto che ci aspettiamo certe probabilità discrete (abbiamo indicato con $p_1^0, p_2^0, \ldots, p_m^0$ i valori che ci aspettiamo per $p_1, p_2, \ldots, p_m$). Poiché la v.a. T_n ci fornisce una misura della distanza fra probabilità teoriche e probabilità misurate, possiamo adottare la regola decisionale

$$\textit{si accetta}\ \mathrm{H}_0\ \textit{se}\ \ T_n \leq c.$$

Per stabilire il valore critico c si osserva che, se H_0 è vera, la v.a. T_n calcolata usando i nostri p_k^0 ha approssimativamente distribuzione $\mathcal{X}^2(m-1)$. Ovviamente, questo è vero *solo se n è abbastanza grande* (generalmente si accetta $n \geq 30$). Dunque, fissato il livello α, la condizione

$$\alpha = P^{p_1=p_1^0, \ldots,\ p_m=p_m^0}\,(T_n > c)$$

implica (approssimativamente)

$$c = q_{1-\alpha}^{\mathcal{X}^2(m-1)}. \tag{7.18}$$

Ricapitolando: il cosiddetto *Test del $\mathcal{X}^2$* consiste nel calcolare l'espressione

$$T_n = \sum_{k=1}^{m} \frac{(N_k - np_k^0)^2}{np_k^0}$$

(o l'equivalente in termini delle frequenze relative F_k) e confrontarne il valore con il numero $c = q_{1-\alpha}^{\mathcal{X}^2(m-1)}$. Se T_n è minore di c si può concludere che le nostre probabilità teoriche $p_1^0, p_2^0, \ldots, p_m^0$ sono corrette.

Esempio 7.4. Non ci possiamo sottarre dal fare l'esempio più classico di utilizzo del Test del $\mathcal{X}^2$: il famoso esperimento di Mendel. Mendel osservò l'occorrenza di quattro possibili coppie, c_1, c_2, c_3, c_4, di caratteri genetici in semi di pisello. Tali coppie sono definite nella seguente tabella:

aspetto:

		LISCIO $_{(D)}$	GRINZOSO
colore:	GIALLO $_{(D)}$	c_1	c_3
	VERDE	c_2	c_4

dove il pedice (D) sta ad indicare il carattere dominante (l'altro è detto recessivo). In base alle osservazioni, Mendel enunciò le sue famose leggi della genetica che qui riassumiamo in una forma adatta ai nostri scopi:

(i) il rapporto dominante/recessivo è di 3 a 1;

(ii) i due caratteri genetici (aspetto e colore) sono indipendenti.

La prima legge si può tradurre, da un punto di vista probabilistico, nel modo seguente:

$$P(\textit{dominante}) = 3/4, \qquad P(\textit{recessivo}) = 1/4.$$

La seconda legge ci aiuta a calcolare, analogamente a quanto fatto nel paragrafo 1.6, la probabilità di occorrenza delle quattro coppie:

$$P(c_1) = \frac{9}{16}, \quad P(c_2) = \frac{3}{16}, \quad P(c_3) = \frac{3}{16}, \quad P(c_4) = \frac{1}{16}.$$

Vogliamo perciò testare la validità della seguente ipotesi:

$$\mathrm{H}_0 \ : \ p_1 = \frac{9}{16}, \quad p_2 = p_3 = \frac{3}{16}, \quad p_4 = \frac{1}{16}$$

basandoci sulle osservazioni dello stesso Mendel:

$$n = 556, \quad N_1 = 315, \quad N_2 = 101, \quad N_3 = 108, \quad N_4 = 32.$$

Calcoliamo perciò T_n usando le nostre probabilità teoriche:

$$\begin{aligned} T_n = {} & \frac{(315 - 556 \times 9/16)^2}{556 \times 9/16} + \frac{(101 - 556 \times 3/16)^2}{556 \times 3/16} \\ & + \frac{(108 - 556 \times 3/16)^2}{556 \times 3/16} + \frac{(32 - 556 \times 1/16)^2}{556 \times 1/16} \approx 0.47\,. \end{aligned}$$

In base alla (7.18), se scegliamo $\alpha = 0.05$, dobbiamo confrontare il valore trovato per T_n con

$$q_{1-\alpha}^{\mathcal{X}^2(m-1)} = q_{0.95}^{\mathcal{X}^2(3)} \approx 7.81,$$

da cui risulta chiaramente che H_0 può essere accettata. □

Il test del $\mathcal{X}^2$ è molto utile anche per testare l'appartenenza di un campione a una certa distribuzione teorica (anche continua). L'idea è quella di fissare un certo numero m di intervalli, $I_1, I_2, \ldots, I_m$, e di calcolare (in base alla distribuzione ipotizzata $\mathcal{D}$) le probabilità teoriche $p_1, p_2, \ldots, p_m$ che il campione ha di cadere in quegli intervalli. Così possiamo saggiare l'ipotesi

$$\mathrm{H}_0 : X_i \sim \mathcal{D}$$

con un test del $\mathcal{X}^2$ sulle probabilità discrete $P(X_i \in I_k) = p_k$. Tuttavia il problema può essere un po' più complicato se $\mathcal{D}$ dipende da parametri stimati col campione stesso (che poi è il caso più frequente). Se ad esempio vogliamo testare la normalità di un campione di cui ignoriamo media e varianza, per calcolare le probabilità teoriche $p_k = P(X_i \in I_k)$ dobbiamo utilizzare una distribuzione normale in cui μ e σ^2 sono stimati da $\overline{X}$ e S^2. In questi casi ci viene in aiuto la seguente versione raffinata del Teorema di Pearson.

Teorema 7.5. (Teorema di Pearson migliorato)
Se le probabilità teoriche $p_k = P(X_i = x_k)$ sono calcolate usando ℓ stimatori, allora

$$\lim_{n\to\infty} T_n \sim \mathcal{X}^2(m - \ell - 1). \tag{7.19}$$

Esempio 7.6.: Test di normalità. In base alle considerazioni sopra esposte, si può costruire un *test di normalità* del campione $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, ovvero un test dell'ipotesi

$$\mathrm{H}_0 : X_i \sim \mathcal{N}(\mu, \sigma^2),$$

usando il test $\mathcal{X}^2$ nel modo seguente:

1. stimo media μ e varianza σ^2 sul campione utilizzando $\overline{X}$ e S^2;
2. standardizzo il campione ponendo $Z_i = \frac{X_i - \mu}{\sigma}$;
3. suddivido la retta reale in m intervalli $I_1, I_2, \ldots, I_m$ (meglio se simmetrici rispetto all'origine);
4. utilizzando la distribuzione $\mathcal{N}$, calcolo le probabilità teoriche, $p_k^0 = P(Z_i \in I_k)$, di appartenenza agli intervalli (figura 7.2);
5. conteggio le frequenze campionarie: $N_k =$ *numero di* Z_i *che cadono in* I_k;
6. eseguo il test del $\mathcal{X}^2$.

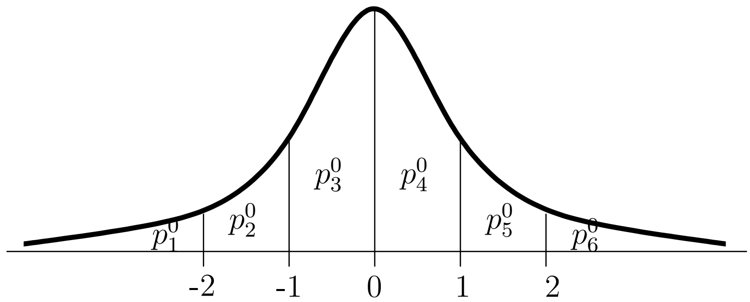

Figura 7.2: Suddivisione della retta reale in m intervalli. In questo esempio si è scelto $m = 6$, $I_1 = (-\infty, -2]$, $I_2 = [-2, -1]$, $I_3 = [-1, 0]$, $I_4 = [0, 1]$, $I_5 = [1, 2]$, $I_6 = [2, +\infty)$ e le corrispondenti probabilità discretizzate sono $p_1^0 = p_6^0 = 0.023$, $p_2^0 = p_4^0 = 0.136$, $p_3^0 = p_4^0 = 0.341$.

Sottoponiamo a questo test i campioni della concentrazione di cromo nelle piante di *Alyssum*, di cui ci siamo già occupati nell'esempio 6.17. Suddividendo la retta reale in $m = 6$ intervalli come in figura 7.2 e standardizzando i nove campioni della tabella 6.1, si trovano le frequenze campionarie e teoriche riportate nella tabella 7.1. Se calcoliamo T_{20} nei nove casi otteniamo i risultati riportati nell'ultima riga della tabella. Scegliendo $\alpha = 0.05$, si ottiene un valore critico del test $\mathcal{X}^2$ pari a

$$q_{1-\alpha}^{\mathcal{X}^2(m-\ell-1)} = q_{0.95}^{\mathcal{X}^2(3)} \approx 7.81$$

che dunque ci porta a non rifiutare l'ipotesi di normalità solo per i campioni 4 e 6. Può essere interessante visualizzare la situazione mediante istogrammi: nella figura 7.3 sono riportati in istogramma le frequenze relative dei campioni (standardizzati) *C6* e *C9* che hanno, rispettivamente, il valore di T_{20} più basso e più alto. Confrontando con la distribuzione normale standard riportata nei due grafici, si nota come, in effetti, *C6* è abbastanza compatibile con la distribuzione normale mentre il campione *C9* si discosta notevolmente da essa. □

	C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	*C9*	np_k^0
N_1	0	0	0	0	0	0	0	0	0	0.46
N_2	3	0	0	2	1	2	0	0	0	2.72
N_3	8	13	12	10	13	9	14	13	16	6.82
N_4	7	4	5	5	3	6	5	4	2	6.82
N_5	0	1	2	2	2	2	0	1	1	2.72
N_6	2	2	1	1	1	1	1	2	1	0.46
T_{20}	8.6	16.2	8.4	3.4	10.1	2.3	14.6	16.2	20.7	

Tabella 7.1: Frequenze campionarie N_k e teoriche np_k^0 relative ai campione della tabella 6.1 e alla discretizzazione della figura 7.2. Nell'ultima riga sono riportati i corrispondenti valori dello stimatore T_{20}.

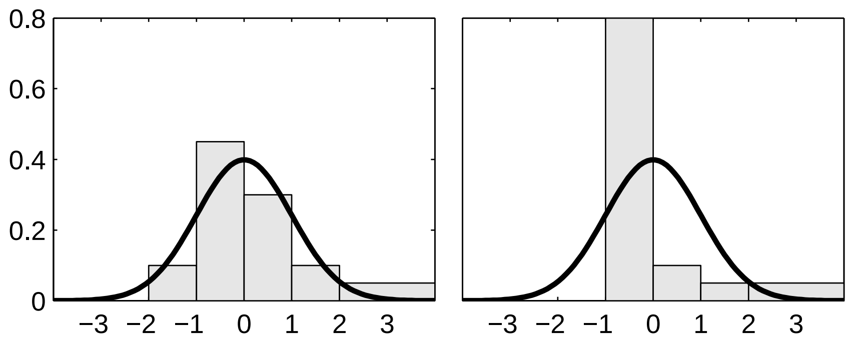

Figura 7.3: Confronto fra le frequenze relative campionarie (istogramma) e la distribuzione normale standard (linea continua). Il due grafici si riferiscono, rispettivamente, ai campioni *C6* e *C9* della tabella 7.1.