INTRODUZIONE AL CALCOLO DELLE PROBABILITÀ E ALL'INFERENZA STATISTICA

Note del corso di Statistica

Francesco Grotto
Università di Pisa
A. A. 2023-2024

Contents

I Statistica Descrittiva e Indici Statistici 3
- A Campioni Statistici . . . 3
- B Indici statistici . . . 3
- C Quantili . . . 5
- D Dati Multivariati . . . 5
- Esercizi . . . 7

II Probabilità e (In)dipendenza . . . 9
- A Spazi di Probabilità . . . 9
- B Probabilità Discreta . . . 10
- C Probabilità Condizionata e formula di Bayes 11
- D Indipendenza . . . 11
- E Entropia di Shannon . . . 13
- F Densità di Probabilità . . . 13
- Esercizi . . . 14

III Variabili Aleatorie - I . . . 15
- A La legge di una Variabile Aleatoria . . . 15
- B Funzione di Ripartizione e Quantili . . . 16
- C Variabili Aleatorie Notevoli Discrete . . . 17
- Esercizi . . . 18

IV Variabili Aleatorie - II . . . 20
- A Variabili Aleatorie Notevoli con Densità . . 20
- Trasformazioni di Variabili con Densità . . . 21
- B Valore Atteso, Varianza e Momenti . . . 21
- C Momenti di Variabili Aleatorie Notevoli . . . 23
- Esercizi . . . 24

V Distribuzioni Multivariate . . . 26
- A Variabili Doppie . . . 26
- B Indipendenza di Variabili Aleatorie . . . 27
- C Funzioni di variabili indipendenti . . . 27
- D Covarianza e Correlazione . . . 28
- Esercizi . . . 29

VI Variabili Indipendenti e Teoremi Limite 31
- A Teoremi Limite . . . 31
- B Variabili Chi-Quadro e di Student . . . 32
- Esercizi . . . 34

VII Campioni Statistici e Stimatori . . . 36
- A Campioni con legge dipendente da un parametro . . . 36
- B Stima parametrica . . . 36
- C Massima Verosimiglianza e Metodo dei Momenti . . . 37
- Esercizi . . . 38

VIII Intervalli di Fiducia . . . 39
- A Intervalli di fiducia per la media di un campione Gaussiano . . . 39
- B Intervalli di fiducia per la media di un campione Bernoulli . . . 41
- C Intervalli di fiducia per la media di un campione di taglia grande . . . 41
- D Intervalli di fiducia per la varianza di un campione Gaussiano . . . 42
- E Esercizi . . . 42

IX Test Statistici e Z-test . . . 44
- A Verifica di ipotesi . . . 44
- B Z-test . . . 46
- Un esempio svolto . . . 46
- C Z-test unilatero . . . 47
- D Test Z approssimato su un campione di Bernoulli . . . 48
- E Esercizi . . . 48

X T-test . . . 49
- A Test di Student . . . 49
- B Test approssimato sulla media di un campione di taglia grande . . . 49
- C Confronto tra campioni . . . 50
- D T-test di regressione . . . 51
- E Esercizi . . . 52

XI Test Chi-Quadro . . . 53
- A Test sulla varianza di un campione Gaussiano 53
- B Test Chi-Quadro di Pearson . . . 53
- C Confronto di Distribuzioni Generiche . . . 54
- D Test per l'indipendenza . . . 55
- E Esercizi . . . 55

XII Esercizi di Verifica . . . 56
- A Quesiti VERO/FALSO . . . 56
- B Soluzioni dei Quesiti VERO/FALSO . . . 57

I. STATISTICA DESCRITTIVA E INDICI STATISTICI

La statistica si occupa dello studio di dati, ovvero della loro raccolta –questione di cui queste note non si occupano– e della loro analisi ed interpretazione. Le risposte di tale analisi, effettuata con i metodi matematici di cui queste note danno una introduzione elementare, non dipendono solo dai dati, ma anche dalla conoscenza pregressa del problema studiato attraverso le eventuali ipotesi ed assunzioni che entrano a far parte del modello matematico usato.

Si parla di **statistica descrittiva** quando i dati vengono analizzati senza fare assunzioni esterne all'insieme di dati considerati. Lo scopo è la sola organizzazione dei dati in modo da evidenziarne la struttura e di rappresentarli in modo efficace. Di questo ci occuperemo brevemente in questa prima parte delle note.

L'**inferenza statistica** invece studia i dati utilizzando un modello probabilistico, cioè suppone che i dati siano valori assunti da **variabili aleatorie** aventi una certa **distribuzione di probabilità** dipendente da dei parametri da scegliere opportunamente. L'analisi statistica ha come scopo la stima di questi parametri, ed il modello matematico che risulta da tale stima può essere utilizzato per fare delle previsioni ed informare delle decisioni. La definizione matematicamente rigorosa di questi concetti è il tema principale di queste note. Conclusioni più forti dell'analisi dei dati seguono a prezzo di assunzioni più stringenti e dettagliate, che quindi rischiano di essere meno affidabili e non sempre verificate.

A. Campioni Statistici

I dati che ci proponiamo di analizzare sono risultati numerici di misure ripetute. Si considera cioè una **popolazione**, l'insieme di oggetti o fenomeni che si vuole studiare, su ognuno dei quali è possibile effettuare la stessa misura, ovvero considerarne un **carattere**. La popolazione può essere reale, ad esempio la popolazione italiana, un'urna con 10 biglie, oppure può essere ideale: ad esempio per un esperimento fisico ripetuto la popolazione è idealmente data da "le infinite possibili ripetizioni dell'esperimento". Un **campione statistico** (*statistical sample*) è un sottoinsieme della popolazione scelto per rappresentarla. I **dati** consistono nelle misure effettuate su di esso, spesso nell'impossibilità di effettuare la misura sulla totalità della popolazione.

Dato un possibile esito della misura, la sua *frequenza (assoluta)* è il numero di volte in cui questo esito compare nei dati, mentre la *frequenza relativa* è la frazione di volte in cui questo esito compare sul totale dei dati. Le frequenze e quindi anche la loro rappresentazione dipendono dai dati del campione scelto e non coincidono in generale con le frequenze su tutta la popolazione.

Esempio I.1. Un *exit poll* è un sondaggio sul voto espresso condotto tra votanti che hanno appena lasciato il seggio elettorale. Gli esiti possibili sono i partiti o i candidati in corsa, e il campione è una parte di solito molto ridotta della popolazione (che è l'insieme di tutti i votanti), tanto che spesso se ne discute la rappresentatività.

Alle *midterm elections* statunitensi, l'8 Novembre 2022, un *exit poll* della CNN ha intervistato 18571 persone, che alla Camera hanno votato i Democratici per il 47.47% oppure i Repubblicani per il 50.53% (il 2% altro). Possiamo in questo caso confrontare una misura su tutta la popolazione votante: il risultato generale ha visto i Democratici al 47.7% e i Repubblicani al 50.7% (1.6% altro), la frequenza del carattere sul campione non coincide con quella su tutta la popolazione.

Come si sceglie il campione in modo che sia rappresentativo? È una questione molto importante e complessa nelle applicazioni, ma non rientra negli obbietivi di queste note. Nella teoria della statistica inferenziale è semplicemente sottointesa l'assunzione di avere a disposizione un campione "scelto a caso uniformemente" nella popolazione.

La rappresentazione grafica più immediata di un vettore di dati numerici è l'**istogramma** (parola grecizzante coniata da Karl Pearson per indicare questi "grafici a bastoni"). Quest'ultimo consiste in una serie di colonne, ognuna delle quali ha per base un intervallo numerico e per *area* la frequenza relativa dei dati contenuti in quell'intervallo: la scelta delle ampiezze degli intervalli di base è cruciale. Nell'esempio di Fig. 1 scegliere 2 intervalli farebbe perdere molta informazione, mentre scegliere 100 intervalli renderebbe l'istogramma troppo dipendente da piccole variazioni dei dati. Ampliando gli intervalli alla base delle colonne si può dare una rappresentazione significativa in casi in cui i dati sono tutti diversi tra loro, e per cui ha poco senso considerare la frequenza di un singolo esito. In generale, un buon compromesso deve essere individuato empiricamente sulla base della numerosità dei dati e della loro specifica distribuzione.

La forma dell'istogramma può essere descritta con dei termini specifici (di cui riportiamo degli esempi in Fig. 2):

- "normale" se la forma assomiglia a quella di una campana simmetrica;
- "unimodale" o "bimodale" a seconda che l'istogramma si concentri rispettivamente attorno a una singola o a due colonne più alte;
- nel caso unimodale, asimmetrica (*skewed*) a destra o a sinistra se i dati sono più concentrati a destra o a sinistra del picco;
- "platicurtica" o "leptocurtica" a seconda che, rispettivamente, i dati siano concentrati quasi totalmente in un certo intervallo (anche grande), oppure siano composti da un gruppo centrale e da molti *outliers* (dati isolati distanti dal corpo della distribuzione).

B. Indici statistici

Consideriamo un vettore $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ di dati numerici. Gli indici statistici sono quantità numeriche che riassumono alcune proprietà significative della distribuzione dei dati $(x_1, \dots, x_n)$.

La media campionaria e la mediana campionaria forniscono una stima della posizione del centro geometrico della distribuzione dei dati:

Definizione I.2. *La* ***media campionaria*** *(*sample average*) è la media aritmetica dei dati,*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Altezza	Frequenza Assoluta
1.55-1.58	2
1.58-1.61	5
1.61-1.64	12
1.64-1.67	25
1.67-1.70	58
1.70-1.73	41
1.73-1.76	43
1.76-1.79	7
1.79-1.82	6
1.82-1.85	1

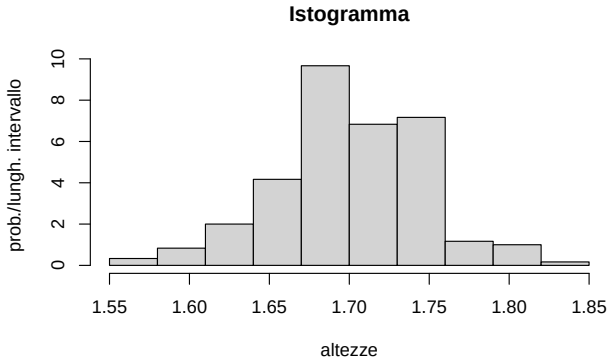


Figura 1: Frequenze assolute delle altezze (in metri) di 200 persone e relativo istogramma.

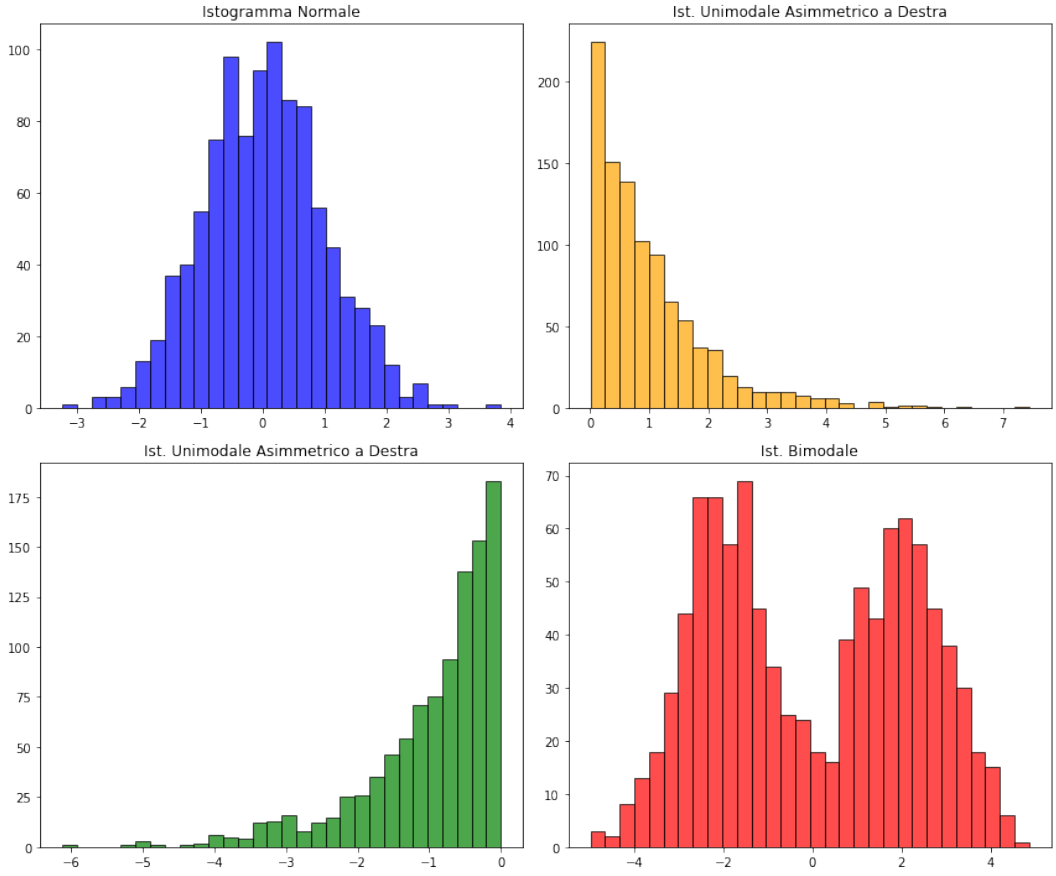


Figura 2: Quattro istogrammi, rispettivamente normale, asimmetrico a sinistra, a destra, e bimodale.

*La **mediana campionaria** (*sample median*) è il dato x_i tale che metà degli altri valori è minore o uguale a x_i e l'altra metà maggiore o uguale (nel caso n sia pari si può prendere la media aritmetica dei due valori centrali in questo senso).*

La mediana è un indice utile nel caso di dati molto asimmetrici (*skewed*) e soprattutto è robusta rispetto alle code della distribuzione: la media campionaria viene facilmente spostata da singoli dati molto piccoli o molto grandi, e questo non succede alla mediana.

La varianza campionaria si usa per misurare la dispersione dei dati, cioè quanto i dati sono "sparsi" attorno al "centro" della distribuzione. Essa è infatti la media degli scarti quadratici dalla media campionaria $\bar{x}$:

Definizione I.3. *Si dice **varianza campionaria** (*sample variance*) l'indice statistico*

$$\mathrm{var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

La radice della varianza campionaria è chiamata scarto quadratico medio o anche deviazione standard: la indichiamo $\sigma(x)$, e si ha dunque $\sigma(x) = \sqrt{\mathrm{var}(x)}$.

La scelta del denominatore $n-1$ è importante in inferenza statistica, e sarà chiara in seguito. La varianza è evidentemente nulla se e solo se i dati sono tutti uguali, e più in generale è una misura della dispersione dei dati attorno alla media campionaria:

Proposizione I.4. *Preso un campione di dati x ed un numero positivo d si ha*

$$\frac{\#\{i : |x_i - \bar{x}| > d\}}{n-1} \leq \frac{\mathrm{var}(x)}{d^2}$$

dove $\#A$ indica il numero degli elementi di A.

Dimostrazione. Si ha che

$$\begin{aligned}\sum_{i=1}^{n} (x_i - \bar{x})^2 &\geq \sum_{i:|x_i-\bar{x}|>d} (x_i - \bar{x})^2 \\ &\geq \sum_{i:|x_i-\bar{x}|>d} d^2 = d^2 \cdot \#\{i : |x_i - \bar{x}| > d\},\end{aligned}$$

da cui

$$\#\{i : |x_i - \bar{x}| > d\} \leq \frac{\sum_i (x_i - \bar{x})^2}{d^2},$$

e dividendo per $n-1$ si ha la tesi. □

Il termine di sinistra è la frazione di dati che differiscono da $\bar{x}$ più di d, ed è quindi evidente l'idea di varianza come misura della dispersione dei dati.

Se la varianza fa uso della funzione $x \mapsto x^2$ per misurare la distanza media dei punti dalla media campionaria, tramite la funzione $x \mapsto x^3$, che assume valori grandi man mano che x si allontana dall'origine *ma mantenendo il segno di x*, possiamo misurare l'asimmetria della distribuzione. Questo perché se ci sono molti dati a sinistra della media campionaria, nella somma di termini $(\bar{x} - x_i)^3$ prevarranno quelli con segno negativo, e viceversa.

Definizione I.5. *La* sample skewness *(misura campionaria di asimmetria) è*

$$b = \frac{1}{\sigma^3} \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3.$$

Usando la funzione $x \mapsto x^4$ si può misurare in modo simile a sopra "quanto piatta" (o anche "quanto normale") è la distribuzione dei dati, definendo la **curtosi** (*kurtosis*). Nel caso di $x \mapsto x^4$ consideriamo una funzione che resituisce numeri di segno sempre positivo, che crescono molto velocemente all'allontanarsi dall'origine ("pesano le code" della distribuzione).

C. Quantili

Un altro modo per visualizzare le frequenze relative dei dati è considerare la seguente funzione:

Definizione I.6. *Dato $x = (x_1, \dots x_n) \in \mathbb{R}^n$, la funzione di ripartizione empirica (*empirical cumulative distribution function*, in breve e.c.d.f.) $F_e : \mathbb{R} \to \mathbb{R}$ è definita da*

$$F_e(t) = \frac{\#\{i \mid x_i \leq t\}}{n}.$$

In altre parole, per ogni $t \in \mathbb{R}$, $F_e(t)$ restituisce la frequenza relativa dei dati minori o uguali a t. La funzione F_e è sempre non-decrescente, e $F_e(-\infty) = 0$, $F(+\infty) = 1$. La Fig. 3 riporta un esempio.

Dato un numero $\beta \in (0, 1)$ (equivalentemente, una percentuale) è naturale chiedersi come scegliere una soglia che divida in due parti i dati, in modo che una di queste contenga dati più piccoli di ogni dato dell'altra, e al contempo contenga una frazione β del totale dei dati. Ad esempio, se i dati misurano la qualità di un prodotto o l'esito di un test, questa è la procedura con cui si sceglie una porzione predeterminata del campione in modo da selezionare gli esiti migliori, o peggiori. Chiameremo questa soglia β-quantile.

In effetti, abbiamo già affrontato la questione definendo la mediana: in quel caso $\beta = 0.5 = 50\%$, ovvero la mediana divide il campione in due parti con la stessa numerosità. Come in quel caso, definire in generale il β-quantile come il più piccolo dato che supera una frazione β dei dati ordinati non è una buona scelta, perché non sempre uno degli x_i soddisfa esattamente la condizione. Si dà quindi una definizione più elaborata:

Definizione I.7. *Si chiama β-quantile il dato x_i tale che:*

- *almeno βn dati siano $\leq x_i$,*
- *almeno $(1-\beta)n$ siano $\geq x_i$,*

e se due dati soddisfano questa condizione, si prende come β-quantile la media aritmetica tra il più piccolo e il più grande di questi ultimi. Lo 0.25*-quantile è chiamato primo quartile, il* 0.5*-quantile (la mediana) è chiamato secondo quartile, il* 0.75*-quantile è detto terzo quartile.*

A livello operativo: detti n la numerosità del campione, $x_{(1)}, \dots x_{(n)}$ i dati ordinati in senso crescente,

- se βn non è intero, il β-quantile è $x_{(\lceil \beta n \rceil)}$, dove $\lceil \beta n \rceil$ è l'arrotondamento di βn all'intero successivo;
- se βn è intero, il β-quantile è la media aritmetica tra $x_{(\beta n)}$ e $x_{(\beta n+1)}$.

Il *box-plot* dei dati è un grafico che evidenzia l'intervallo in cui sono concentrati i dati: si ottiene sovrapponendo a una linea che va dal minimo al massimo dei dati un rettangolo che va dal primo al terzo quartile, con una linea che lo divide al livello della mediana. Fig. 4 ne riporta un esempio, che continua quello di Fig. 3. Il rettangolo centrale contiene il 50% dei dati centrali e viene detto *interquartile range*.

D. Dati Multivariati

Ci limitiamo al caso di coppie di dati (dati bivariati) $(x, y) = ((x_1, y_1), \dots, (x_n, y_n)) \in \mathbb{R}^{2\times n}$; il caso generale è del tutto analogo. Indichiamo con $\bar{x}$ e $\bar{y}$ le medie campionarie di x e y, e con $\sigma(x)$, $\sigma(y)$ le deviazioni standard campionarie di x e y.

Definizione I.8. *Si chiama covarianza campionaria la quantità*

$$\mathrm{cov}(x, y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}.$$

Definizione I.9. *Supponiamo $\sigma(x) \neq 0$ e $\sigma(y) \neq 0$: si chiama coefficiente di correlazione tra x e y il numero*

$$\begin{aligned} r(x, y) &= \frac{\mathrm{cov}(x, y)}{\sigma(x)\sigma(y)} \\ &= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}. \end{aligned}$$

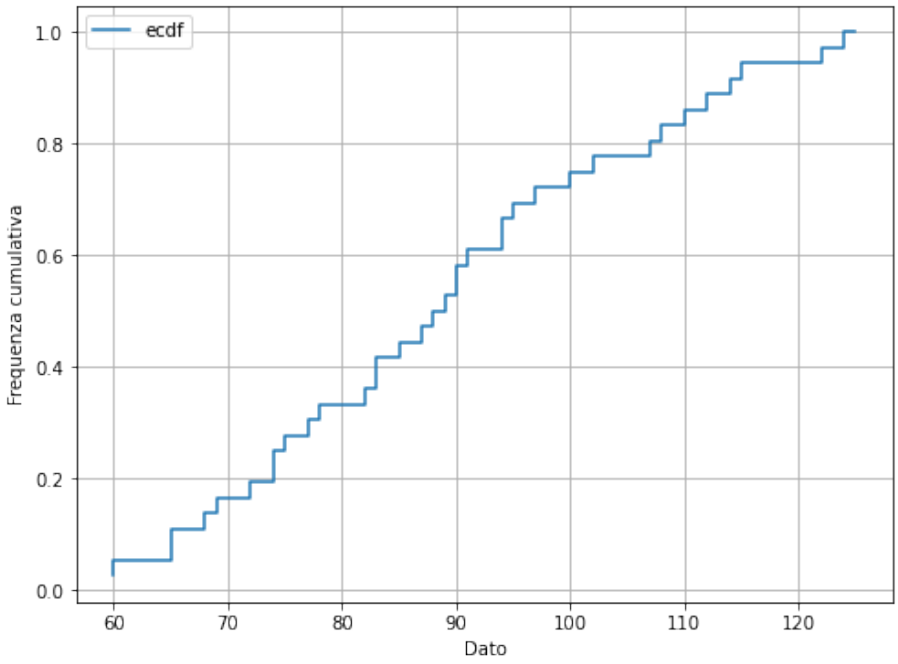


Figura 3: All'esterno della Great Central Station viene misurato il rumore (in dB, in giorni diversi alla stessa ora) con esiti: 82, 89, 94, 110, 74, 122, 112, 95, 100, 78, 65, 60, 90, 83, 87, 75, 114, 85, 69, 94, 124, 115, 107, 88, 97, 74, 72, 68, 83, 91, 90, 102, 77, 125, 108, 65. La relativa e.c.d.f. è riportata nel grafico.

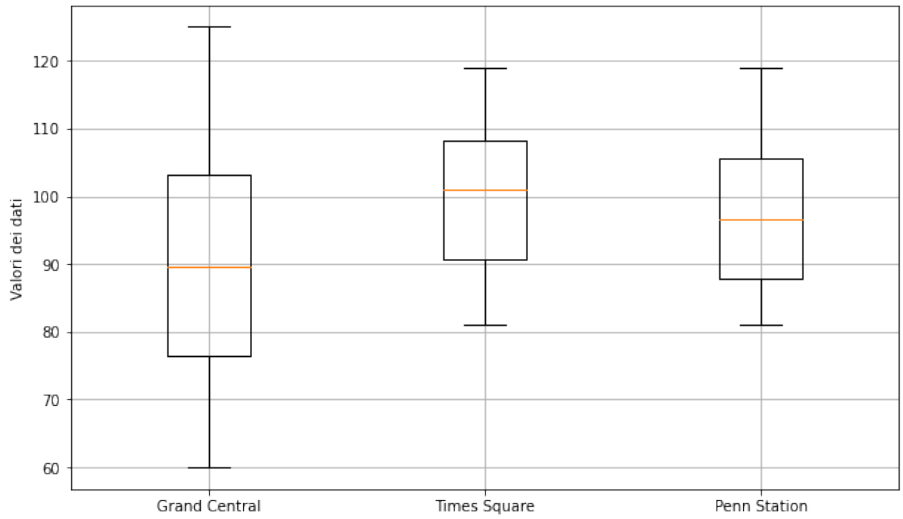


Figura 4: In Fig. 3 i dati ordinati sono: 60, 65, 65, 68, 69, 72, 74, 74, 75, 77, 78, 82, 83, 83, 85, 87, 88, 89, 90, 90, 91, 94, 94, 95, 97, 100, 102, 107, 108, 110, 112, 114, 115, 122, 124, 125. Il primo, secondo e terzo quartile sono quindi: $Q_1 = \frac{75+77}{2} = 76$, $Q_2 = \frac{89+90}{2} = 89.5$, $Q_3 = \frac{102+107}{2} = 104.5$. Il relativo boxplot è riportato a sinistra, gli altri due sono relativi ad altre stazioni di cui non riportiamo i dati.

Dalla diseguaglianza di Cauchy-Schwarz,

$$\left|\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right| \leq \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

e di conseguenza $|r(x,y)| \leq 1$.

Il coefficiente di correlazione misura la presenza di una relazione lineare tra i dati x e y, quantificata dalla *retta di regressione*. L'idea è di approssimare nel modo migliore i dati y_i con una combinazione lineare affine $a + bx_i$, per farlo misuriamo la distanza dei dati da tale retta con i quadrati degli scarti: siamo dunque condotti a cercare i parametri a, b calcolando

$$\inf_{a,b\in\mathbb{R}} \sum_{i=1}^{n}(y_i - a - bx_i)^2 .$$

Teorema I.10. *Se $\sigma(x) \neq 0$, esiste un unico minimo al variare di $a, b \in \mathbb{R}$ della quantità $\sum_{i=1}^{n}(y_i - a - bx_i)^2$, dato da*

$$b^* = \frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)}, \quad a^* = -b^*\bar{x} + \bar{y},$$

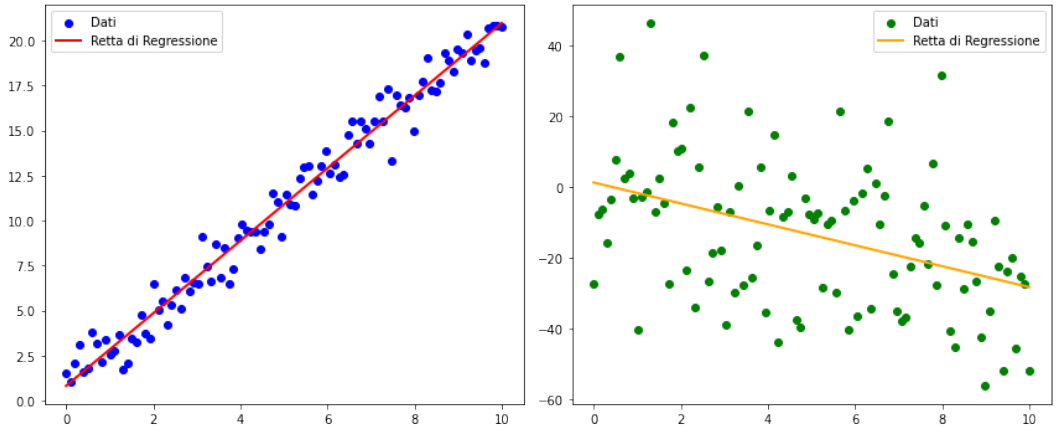


Figura 5: Due insiemi di coppie di dati rappresentati come punti sul piano, con relativa retta di regressione. In quello di sinistra il coefficiente di correlazione è vicino a 1, e i dati sono raggruppati attorno alla retta di regressione, mentre nel secondo caso sono molto più sparpagliati e il coefficiente di correlazione è vicino a 0.

e se inoltre $\sigma(y) \neq 0$ *vale*

$$\min_{a,b\in\mathbb{R}} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = \left(1 - r(x,y)^2\right) \sum_{i=1}^{n} (y_i - \bar{y})^2 .$$

La retta $y = a^* + b^* x$ è chiamata retta di regressione: quanto più $r(x,y)$ è vicino a 1 in valore assoluto, tanto più i punti tendono ad essere allineati e vicino a tale retta. In particolare $|r(x,y)| = 1$ se e solo se i punti si trovano tutti su una stessa retta, e $r(x,y)$ è positivo o negativo se il coefficiente angolare della retta è rispettivamente positivo o negativo. Se $r(x,y)$ è prossimo a zero non è possibile dare una buona approssimazione dei dati con una retta, nel senso che per ogni retta considerata esisteranno dei dati distanti da essa. Questo non esclude che possano esistere altre relazioni (approssimate) tra x e y, ad esempio polinomiali.

Dimostrazione. Consideriamo la funzione continua (e infinitamente differenziabile)

$$Q(a,b) = \sum_{i=1}^{n} (y_i - a - bx_i)^2 .$$

La funzione Q tende a $+\infty$ quando $|a|, |b| \to \infty$ e pertanto ha un minimo. Nel punto di minimo deve valere $\frac{\partial Q}{\partial a} = \frac{\partial Q}{\partial b} = 0$, condizione da cui otteniamo

$$\begin{cases} \sum_i y_i - na - b\sum_i x_i = 0, \\ \sum_i x_i y_i - a\sum_i x_i - b\sum_i x_i^2 = 0. \end{cases}$$

Dividendo per n si ottiene

$$\begin{cases} a + b\bar{x} = \bar{y}, \\ a\bar{x} + b\sum_i \frac{x_i^2}{n} = \sum_i \frac{x_i y_i}{n}, \end{cases}$$

che è un sistema di due equazioni nelle due incognite a, b. L'unica soluzione è quella nell'enunciato, ed essendo l'unico punto stazionario di Q essa coincide con il suo minimo, che quindi è unico. Il valore di Q nel minimo si ottiene sostituendo i valori trovati e sviluppando i calcoli. □

Esercizi

Esercizio I.11. Si mostri che vale la seguente formula per la varianza campionaria:

$$\mathrm{var}(x) = \frac{n}{n-1}\left(\sum_{i=1}^{n} \frac{x_i^2}{n} - \bar{x}^2\right)$$

(sugg.: si mostri che $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$); si mostri poi che per la covarianza campionaria vale:

$$\mathrm{cov}(x,y) = \frac{n}{n-1}\left(\sum_{i=1}^{n} \frac{x_i y_i}{n} - \bar{x}\bar{y}\right).$$

Esercizio I.12. Si mostri che modificando dei dati $x_1, \ldots, x_n$ aggiungendo uno stesso numero $c \in \mathbb{R}$, ovvero considerando invece $y_1 = x_1 + x, \ldots, y_n = x_n + c$, la varianza campionaria non cambia, ovvero $\sigma(x) = \sigma(y)$. Si mostri poi che, moltiplicando i dati per uno stesso numero $a \neq 0$, ovvero invece $z_1 = ax_1, \ldots, z_n = ax_n$, la varianza campionaria risulta $\sigma(ax) = a\sigma(x)$.

Esercizio I.13. Assumiamo che in un vettore di dati $x_1, \ldots, x_n$ vengano assunti m esiti $a_1, \ldots, a_m$, ovvero ogni singolo x_i è uguale a uno degli a_j. Avendo definito la frequenza relativa di un esito a_j nel vettore di dati x come

$$p(a_j, x) = \frac{\#\{i : x_i = a_j\}}{n},$$

mostrare che valgono le seguenti formule per media e varianza campionaria:

$$\bar{x} = \sum_{j=1}^{M} a_j p(a_j, x)$$

$$\mathrm{var}_e(x) = \sum_{j=1}^{M} a_j^2 p(a_j, x) - \bar{x}^2.$$

Esercizio I.14. Nella regressione lineare si definisce il *coefficiente di determinazione* come

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - a^* - b^* x_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

(la frazione a secondo membro è detta *varianza non spiegata*, *unexplained variance*, del modello lineare). Mostrare che R^2 coincide con il coefficiente di correlazione $r(x, y)$.

Nota: il coefficiente di determinazione si definisce in modelli di regressione più generali di quello considerato, e solo nel caso lineare risulta coincidere con quello di correlazione.

Esercizio I.15. Si mostri che il coefficiente di correlazione r è invariante per cambiamenti lineari di scala, cioè $r(x, y) = r(\alpha x + \beta, \gamma y + \delta)$ per ogni $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ con α, β non nulli.

Esercizio I.16. Si mostri che, in regressione lineare, la somma dei residui

$$\sum_{i=1}^n (y_i - a^* - b^* x_i) = 0$$

è sempre nulla.

Esercizio I.17. Si consideri un insieme di dati tali che $x_2 = -x_1$, $x_4 = -x_3$, ... (ovvero ogni dato di indice pari è l'opposto del precedente), per un totale di n dati con n pari. Si mostri che la misura campionaria di asimmetria è in questo caso nulla.

Esercizio I.18. Si consideri lo *scarto medio assoluto*,

$$\delta = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

di un insieme di dati $x_1, \dots x_n$. Si dimostri la seguente disuguaglianza analoga a quella della Proposizione I.4,

$$\frac{1}{n} \#\left(i : |x_i - \bar{x}| > d\right) \leq \delta / n.$$

Si mostri poi la seguente disuguaglianza che la mette in relazione alla deviazione standard

$$\delta \leq \sqrt{\frac{n-1}{n}} \sigma.$$

II. PROBABILITÀ E (IN)DIPENDENZA

La teoria della Probabilità nasce per tentare di quantificare l'incertezza, misurando la fiducia che un evento possa accadere. La definizione matematica di probabilità risale ai lavori di Kolmogorov ed è assiomatica, cioè definisce la probabilità in modo rigoroso tramite le sue proprietà elementari.

A. Spazi di Probabilità

Si vuole descrivere con oggetti matematici l'incertezza di una misura numerica in un determinato fenomeno. Per fissare le idee consideriamo un esempio: la produzione di un pezzo meccanico. Un esito del nostro fenomeno è un singolo pezzo uscito dalla linea di produzione, la misura numerica effettuata è –ad esempio– la lunghezza del pezzo. Un evento è una affermazione sulla misura, come "la misura non eccede 1 metro". Due pezzi diversi (esiti diversi) possono avere lunghezze diverse, ma essere entrambi più corti di un metro (stesso evento).

Formalmente, rappresentiamo tutti gli esiti possibili dell' esperimento con gli elementi ω di un insieme astratto Ω, detto *spazio campionario* (*sample space*). Ogni affermazione su misure effettuate sull'esperimento corrisponde a un sottoinsieme $A \subset \Omega$ degli esiti che la soddisfa, e combinazioni logiche di affermazioni corrispondono dunque a operazioni insiemistiche secondo la usuale algebra Booleana richiamata in Sezione II A.

Il singolo esito $\omega \in \Omega$ costituisce un "evento elementare" se identificato con l'insieme singoletto $\{\omega\}$, cosa che spesso faremo tacitamente. L'affermazione sempre vera (`true`) si identifica con l'intero Ω, mentre quella sempre falsa (`false`) con il vuoto $\emptyset$. Se due eventi sono disgiunti $A \cap B = \emptyset$ (A `and` B restituisce `false`) li diciamo anche *incompatibili*.

Esempio II.1. Lo spazio campionario Ω associato a due lanci (ordinati) di un dado è dato da

$$\{(i,j) \mid i,j \in \{1,2,\dots 6\}\} = \{1,2,\dots 6\}^2,$$

ovvero dall'insieme di tutte le coppie di possibili esiti. Più in generale, se un esperimento è costituito da una successione (ordinata) di n sotto-esperimenti, allora il suo spazio campionario è

$$\Omega = \{(\omega_1, \omega_2, \dots \omega_n) \mid \omega_1 \in \Omega_1, \dots \omega_n \in \Omega_n\}$$

dove Ω_i è l'insieme degli esiti dell'i-simo sotto-esperimento.

Per descrivere le nascite del prossimo anno in Italia dovremo considerare l'insieme infinito $\Omega = \mathbb{N}_0 = \{1,2,\dots\}$, perché sebbene l'esito sia certamente finito esso può assumere valori arbitrariamente grandi, anche se più grande è l'esito considerato, meno esso sarà probabile.

Il grado di fiducia che un evento si realizzi, chiamato probabilità, è rappresentato da un numero compreso tra 0 e 1. In generale è necessario restringere la possibilità di definire la probabilità a un sottoinsieme di $\mathcal{P}(\Omega)$, l'insieme di tutti i sottoinsiemi di Ω, che sia chiuso per le operazioni logiche che ci servono, in particolare **unioni e intersezioni infinite**: classi di questo tipo si chiamano *σ-algebre*. **Assumeremo sempre che tutte le operazioni insiemistiche tra eventi siano possibili**, ovvero che restituiscano a loro volta eventi su cui possiamo definire la probabilità, sebbene questo presenti delle difficoltà matematiche –che ignoreremo sistematicamente– nel caso in cui Ω è infinito.

Operazione logica	operazione insiemistica
A `and` B	$A \cap B$
A `or` B	$A \cup B$
A `xor` B	$A \triangle B$
`not` A	$A^c = \Omega \setminus A$
A `imply` B	$A^c \cup B$
A `notimply` B	$A \setminus B$

Tabella I: Operazioni logiche e insiemistiche. Ricordiamo che $A \triangle B = (A \setminus B) \cup (B \setminus A)$ e $A \setminus B = A \cap B^c$ è la differenza insiemistica.

È intuitivo assumere che se due eventi sono incompatibili la probabilità che si realizzi uno qualsiasi dei due debba essere la somma delle probabilità dei singoli eventi. La definizione generale tiene debitamente conto del caso in cui dobbiamo trattare infiniti eventi.

Definizione II.2. *Dato Ω un insieme e $\mathcal{F}$ una σ-algebra di parti di Ω (che ai nostri fini possiamo sempre supporre essere $\mathcal{P}(\Omega)$) una* misura di probabilità *(anche semplicemente "una probabilità") è una funzione $\mathbb{P} : \mathcal{F} \to [0,1]$ tale che*

- *l'evento certo ha probabilità unitaria, $\mathbb{P}(\Omega) = 1$;*
- *(σ-additività) se $(A_n)_{n=1,2,\dots}$ è una successione di eventi a due a due disgiunti, si ha*

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

Si dice trascurabile *un evento A tale che $\mathbb{P}(A) = 0$, e si dice* quasi certo *(*almost sure*) un evento A tale che $\mathbb{P}(A) = 1$.*

Naturalmente la σ-additività implica l'additività su finiti sottoinsiemi disgiunti, cioè, dati $A_1, \dots A_N$ a due a due disgiunti,

$$\mathbb{P}\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mathbb{P}(A_n);$$

infatti, è sufficiente scegliere $A_n = \emptyset$ da un certo n in poi per specializzare la definizione generale al caso finito.

Uno spazio campionario Ω Ogni misura di probabilità verifica le seguenti proprietà, la cui dimostrazione è lasciata per esercizio:

Proposizione II.3. *Valgono:*

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *e di conseguenza* $\mathbb{P}(\emptyset) = 0$*;*
- *se* $B \subset A$, $\mathbb{P}(A \backslash B) = \mathbb{P}(A) - \mathbb{P}(B)$*;*
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$*;*
- $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)$.

La σ-additività permette di "passare al limite", nel seguente senso:

Proposizione II.4. *Data una successione di eventi $A_1, \dots, A_n, \dots$, assumiamo (alternativamente) che*

- *la successione sia crescente, ovvero* $A_n \subseteq A_{n+1}$*, e poniamo* $A = \bigcup_{n=1}^{+\infty} A_n = \lim_{n\to\infty} A_n$*;*
- *la successione sia decrescente, ovvero* $A_n \supseteq A_{n+1}$*, e poniamo* $A = \bigcap_{n=1}^{+\infty} A_n = \lim_{n\to\infty} A_n$*.*

In entrambi i casi vale

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Dimostrazione. Consideriamo dapprima il caso in cui la successione è crescente. Poniamo

$$B_1 = A_1, B_n = A_n \backslash A_{n-1}, \quad n > 1.$$

Gli insiemi $(B_n)_{n\geq 1}$ sono a due a due disgiunti e per l'additività finita si ha $\mathbb{P}(B_n) = \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})$. Poiché $\bigcup_{n\geq 1} A_n = \bigcup_{n\geq 1} B_n$,

$$\mathbb{P}(A) = \sum_{n=1}^{+\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \sum_{h=1}^{n} \mathbb{P}(B_h) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Nel caso in cui la successione è decrescente la tesi segue dal caso crescente semplicemente passando al complementare. □

B. Probabilità Discreta

Se Ω è numerabile, ovvero se si possono indicizzare i suoi elementi con (un sottoinsieme possibilmente finito de)i numeri naturali $\mathbb{N}$,

$$\Omega = (\omega_1, \omega_2, \dots, \omega_n, \dots),$$

una misura di probabilità è univocamente determinata dai numeri

$$p_i = \mathbb{P}(\omega_i) = \mathbb{P}(\{\omega_i\}) \in [0, 1]$$

(ometteremo a volte le doppie parentesi nelle probabilità di insiemi con un singolo elemento) e per ogni evento $A \subset \Omega$ si ha $\mathbb{P}(A) = \sum_{\omega_i \in A} p_i$. In questo caso diciamo che $\mathbb{P}$ è una **probabilità discreta**.

Un esempio significativo di spazio di probabilità è la *probabilità uniforme su un insieme finito* Ω,

$$p_1 = p_2 = \cdots = p_N.$$

Ne sono esempi i modelli associati all'estrazione da un'urna e al lancio di un dado (non truccato). In questo caso specifico vale la formula

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega} = \frac{\text{“casi favorevoli”}}{\text{“casi possibili”}}, \quad A \subseteq \Omega,$$

dove con $\#A$ si indica la cardinalità (o numero degli elementi) dell'insieme A.

Nel caso di probabilità discreta i problemi diventano molto spesso calcoli combinatori: riportiamo solo alcune formule notevoli. Anziché dire “un insieme di n elementi”, scriveremo per brevità $\{1, \dots, n\}$.

Proposizione II.5. *Siano* k *ed* n *due interi.*

- *il numero di sequenze ordinate, possibilmente con ripetizione, di* k *numeri da* 1 *a* n*, cioè il numero di funzioni da* $\{1, \dots, k\}$ *a* $\{1, \dots, n\}$ *è* n^k*;*
- *il numero di modi in cui si può ordinare* $\{1, \dots, n\}$ *(ovvero il numero di funzioni biiettive dall'insieme a se stesso, o di permutazioni di* n *elementi) è*

$$n! = 1 \cdot 2 \cdots (n-1) \cdot n;$$

- *se* $0 \leq k \leq n$*, il numero di sequenze ordinate senza ripetizione di* k *numeri di* $\{1, \dots, n\}$ *(ovvero il numero di funzioni iniettive da* $\{1, \dots, k\}$ *in* $\{1, \dots, n\}$*) è*

$$\frac{n!}{(n-k)!} = n \cdot (n-1) \cdots (n-k+1);$$

- *se* $0 \leq k \leq n$*, il numero di sottinsiemi di* $\{1, \dots, n\}$ *formati da* k *elementi (coefficiente binomiale) è*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

dove $n! = n \cdot (n-1) \cdot \ldots 2 \cdot 1$ *e* $0! = 1$*.*

Questi enunciati si dimostrano facilmente per induzione su n. Il coefficiente binomiale è così detto perché interviene nella formula del binomio di Newton:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Esempio II.6. Consideriamo l'esperimento dato da k estrazioni ordinate, con rimpiazzo, da una collezione di n oggetti, ad esempio un'urna di n biglie (per rimpiazzo si intende che l'oggetto estratto viene nuovamente inserito nell'urna e quindi può essere selezionato nelle estrazioni successive). Questo esperimento è descritto dallo spazio campionario $\Omega = \{1, \dots, n\}^k$ e probabilità $\mathbb{P}$ uniforme: infatti tutte le possibili sequenze di k oggetti estratti sono equiprobabili. In particolare $P(A) = \#A/\#\Omega$ con $\#\Omega = n^k$.

Lo stesso modello si può usare per k lanci di un dado equilibrato, $\Omega = \{1, \dots, 6\}^k$, e k lanci di moneta equilibrata, $\Omega = \{\text{testa, croce}\}^k$.

Esempio II.7. Consideriamo l'esperimento dato da k estrazioni ordinate senza rimpiazzo da una collezione S di n oggetti, con $k \leq n$. La probabilità è ancora uniforme, ma con un spazio campionario: $\Omega = \{(x_1, \dots x_k) \mid x_i \in S, x_i \text{ tutti distinti}\}$, con $\#\Omega = n!/(n-k)!$.

Invitiamo il lettore a descrivere nel formalismo degli spazi di probabilità estrazioni senza ordine e senza rimpiazzo, o senza ordine e con rimpiazzo.

Un caso speciale di probabilità discreta si ha se $\Omega = \{x_1, x_2, \dots\} \subset \mathbb{R}$ è un sottoinsieme numerabile (si noti che sebbene sia sempre possibile ordinare i punti x_i essi possono contenere successioni che vanno sia a $+\infty$ che $-\infty$). Si definisce in questo caso la *funzione di massa*

$$\Omega \ni x_i \mapsto p(x_i) = \mathbb{P}(\{x_i\}) \in [0, 1],$$

ponendo $p(x) = \mathbb{P}(\{x\}) = 0$ se $x \neq x_i$ per ogni i possiamo estendere la funzione a tutto $\mathbb{R}$. Vale quindi:

$$\mathbb{P}(A) = \sum_{i : x_i \in A} p(x_i), \quad \forall A \subseteq \mathbb{R}. \tag{1}$$

Vale $p(x_i) \geq 0$ e $\sum_{i=1,2,\dots} p(x_i) = 1$. Date una successione di punti $x_1, x_2, \dots$ in $\mathbb{R}$ e una funzione $p(x_i)$ di tali punti che soddisfa $p(x_i) \geq 0$ e $\sum_{i=1,2,\dots} p(x_i) = 1$, la formula

(1) definisce un'unica probabilità discreta $\mathbb{P}$ su $\Omega = \mathbb{R}$ avente $p(x_i)$ per funzione di massa.

C. Probabilità Condizionata e formula di Bayes

Quando si è a conoscenza della realizzazione di un evento, cambia la valutazione di probabilità di ogni altro evento. Ad esempio se si sa che il numero uscito dal lancio di un dado è pari, la probabilità che sia uscito il numero 6 non è più $\frac{1}{6}$, ma $\frac{1}{3}$. Infatti, se si è realizzato l'evento $B = \{2, 4, 6\} \subset \Omega = \{1, \dots, 6\}$ (cioè è uscito un numero pari) sono rimasti 3 casi possibili dei quali uno è favorevole: se indichiamo con $A = \{6\}$, notiamo che la nuova probabilità che è stata attribuita ad A verifica la formula $\mathbb{P}(A \cap B)/\mathbb{P}(B)$.

Si possono fornire diversi esempi simili che verificano sempre la formula sopra riportata, e si dà dunque la seguente definizione:

Definizione II.8. *Dati due eventi A, B, con B non trascurabile, si chiama probabilità condizionata di A rispetto a B il numero*

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

Si osserva facilmente che, fissato B non trascurabile, la probabilità condizionata

$$A \mapsto \mathbb{P}(A \mid B)$$

è una probabilità: ad esempio vale $\mathbb{P}(A^c \mid B) = 1 - \mathbb{P}(A \mid B)$. Dati due eventi A e B non trascurabili,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B) = \mathbb{P}(B \mid A) \cdot \mathbb{P}(A),$$

e più in generale vale la seguente:

Proposizione II.9 (Condizionamento ripetuto)**.** *Se l'intersezione di eventi $A_1 \cap \ldots \cap A_{n-1}$ non è trascurabile vale*

$$\begin{aligned}&\mathbb{P}(A_1 \cap \ldots \cap A_n)\\ &\quad = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1) \cdots \mathbb{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1}).\end{aligned}$$

La dimostrazione si ottiene (per induzione) esplicitando i vari termini; si noti che, se $1 \le k < n-1$, anche $A_1 \cap \ldots \cap A_k$ è non trascurabile.

Definizione II.10. *Una partizione di Ω è una collezione di n eventi $B_1, \dots, B_n$ a due a due disgiunti, tali che $B_1 \cup \cdots \cup B_n = \Omega$. Un sistema di alternative è una partizione di Ω in eventi non trascurabili.*

Teorema II.11 (Formula della probabilità totale o formula di fattorizzazione)**.** *Sia $B_1, \dots, B_n$ un sistema di alternative. Allora, per un qualunque evento A, vale*

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i).$$

Dimostrazione. Si noti che $A = (A \cap B_1) \cup \ldots \cup (A \cap B_n)$, come unione di eventi a due a due disgiunti. Si ha pertanto

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i). \qquad \square$$

Teorema II.12 (Formula di Bayes)**.** *Siano A e B due eventi non trascurabili. Allora vale la formula*

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

Se $B_1, \dots, B_n$ è un sistema di alternative e A è un evento non trascurabile, vale

$$\begin{aligned}\mathbb{P}(B_i \mid A) &= \frac{\mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}{\mathbb{P}(A)}\\ &= \frac{\mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}{\sum_{j=1}^{n} \mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j)}.\end{aligned}$$

Dimostrazione. La prima formula segue da:

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

La seconda formula segue dalla prima e dalla formula della probabilità totale. $\square$

La formula della probabilità totale e il teorema di Bayes sono valide anche se il sistema di alternative anziché essere finito è numerabile, sostituendo alle somme finite le somme di una serie. Il condizionamento ripetuto, la formula di fattorizzazione e quella di Bayes permettono di affrontare tutti i calcoli di base relativi al condizionamento, l'esempio che segue è paradigmatico.

Esempio II.13. Un test diagnostico per una data malattia ha indice di sensibilità 0.99: se una persona è malata il test è positivo con probabilità 0.99. Inoltre il test ha indice di specificità 0.97: se se una persona è sana il test è negativo con probabilità 0.97. Supponiamo che l'1% della popolazione soffra di tale malattia.

Un individuo scelto a caso è sottoposto al test. La situazione può essere descritta con due coppie di eventi alternativi: S: "individuo sano" e il complementare S^c; P: "individuo positivo" e il complementare P^c. Si può rappresentarli come sottoinsiemi di uno spazio Ω con quattro elementi, che rappresentano le quattro possibilità "sano e positivo", "sano e negativo", "non sano e positivo", "non sano e negativo". I dati sono quindi $\mathbb{P}(P \mid S^c) = 0.99$, $\mathbb{P}(P^c \mid S) = 0.97$, $\mathbb{P}(S^c) = 0.01$.

Calcoliamo la probabilità che il test sia positivo: applicando la formula di fattorizzazione,

$$\begin{aligned}\mathbb{P}(P) &= \mathbb{P}(P \mid S)\mathbb{P}(S) + \mathbb{P}(P \mid S^c)\mathbb{P}(S^c)\\ &= (1 - \mathbb{P}(P^c \mid S))(1 - \mathbb{P}(S^c)) + \mathbb{P}(P \mid S^c)\mathbb{P}(S^c)\\ &= 0.03 \cdot 0.99 + 0.99 \cdot 0.01 = 0.0396.\end{aligned}$$

Calcoliamo ora la probabilità che il paziente non sia sano sapendo che il test è positivo: applicando la formula di Bayes,

$$\mathbb{P}(S^c \mid P) = \frac{\mathbb{P}(P \mid S^c)\mathbb{P}(S^c)}{\mathbb{P}(P)} = \frac{0.99 \cdot 0.01}{0.0396} = 0.25.$$

D. Indipendenza

Vogliamo codificare nel formalismo matematico l'idea che la conoscenza che si è realizzato un certo evento non modifica la valutazione di probabilità di un altro evento,

e viceversa. Consideriamo quindi due eventi A e B (non trascurabili) e imponiamo che

$$\mathbb{P}(A) = \mathbb{P}(A \mid B), \qquad \mathbb{P}(B) = \mathbb{P}(B \mid A).$$

Queste richieste sono equivalenti tra loro, ed equivalenti all'eguaglianza $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. A differenza delle due precedenti, quest'ultima è simmetrica rispetto ai due eventi ed ha senso anche per eventi trascurabili: viene scelta dunque come definizione di indipendenza.

Definizione II.14. *Due eventi A e B sono indipendenti se vale*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Proposizione II.15. *Valgono:*

- *Se A e B sono indipendenti, lo sono anche A^c e B, A e B^c, A^c e B^c;*
- *se $\mathbb{P}(A) = 0$ oppure $\mathbb{P}(A) = 1$, A è indipendente da qualsiasi altro evento;*
- *due eventi disgiunti non possono essere indipendenti, a meno che almeno uno dei due sia trascurabile.*

La verifica di queste proprietà è un facile esercizio.

Vediamo ora come si estende la nozione di indipendenza al caso di più eventi, cominciando con tre. Affinché A, B e C siano indipendenti, occorre anzitutto che siano soddisfatte le tre eguaglianze

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B),\\ \mathbb{P}(A \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(C),\\ \mathbb{P}(B \cap C) &= \mathbb{P}(B) \cdot \mathbb{P}(C),\end{aligned}$$

cioè che siano a due a due indipendenti, ma questo non basta! Deve anche esse soddisfatta l'eguaglianza

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C),$$

come rende chiaro l'esempio seguente.

Esempio II.16. Sull'insieme $\Omega = \{1, 2, 3, 4\}$ munito della distribuzione uniforme di probabilità, gli eventi $A = \{1, 2\}, B = \{1, 3\}$ e $C = \{2, 3\}$ sono a due a due indipendenti. Tuttavia, valgono anche

$$\begin{aligned}\mathbb{P}(A \cap B \mid C) = 0 \neq \frac{1}{4} = \mathbb{P}(A \cap B),\\ \mathbb{P}(C \mid A \cap B) = 0 \neq \frac{1}{2} = \mathbb{P}(C),\end{aligned}$$

ovvero conoscere l'evento C fornisce informazione su $A \cap B$ e viceversa. Non è ragionevole dunque dire che A, B, C sono globalmente indipendenti.

Definizione II.17. *Assegnati n eventi $A_1, \dots, A_n$, questi si dicono indipendenti se per ogni intero k con $2 \leq k \leq n$ e per ogni scelta di interi $1 \leq i_1 < i_2 < \dots < i_k \leq n$, vale l'eguaglianza*

$$\mathbb{P}\left(A_{i_1} \cap \dots \cap A_{i_k}\right) = \mathbb{P}\left(A_{i_1}\right) \dots \mathbb{P}\left(A_{i_k}\right).$$

Quando il numero di eventi cresce, il numero di eguaglianze da verificare sale notevolmente: per n eventi le eguaglianze da verificare sono $2^n - n - 1$. Questo sembra un problema teorico soprattutto nella descrizione probabilistica di un esperimento ripetuto più volte nelle medesime condizioni, che sarà uno dei nostri obiettivi principali: ci attendiamo infatti che gli eventi associati a ripetizioni distinte siano indipendenti. La complicazione è però solo apparente: eventi (ed in seguito variabili aleatorie) indipendenti sono facilmente costruiti considerando spazi prodotto. Ne vediamo un esempio paradigmatico con il risultato che segue.

Proposizione II.18. *Si consideri*

$$\Omega = \{a = (a_1, \dots, a_n) \mid a_i = 0, 1\} = \{0, 1\}^n,$$

su cui definiamo, per ogni singolo $a = (a_1, \dots, a_n) \in \Omega$,

$$\begin{aligned}\mathbb{P}(\{a\}) &= p^{\#\{i:a_i=1\}}(1-p)^{\#\{i:a_i=0\}}\\ &= p^{\sum_{i=1}^n a_i}(1-p)^{n-\sum_{i=1}^n a_i}.\end{aligned}$$

Ciò definisce una probabilità, e gli eventi

$$A_i = \{a \in \Omega : a_i = 1\}, \quad i = 1, \dots, n,$$

sono tutti indipendenti tra di loro, così come i loro complementari A_i^c.

Lo spazio di probabilità suddetto descrive ad esempio n lanci di una moneta, in generale non equilibrata se $p \neq 1/2$. Nel caso di una moneta equilibrata, $p = 1/2$, siccome i due esiti di un lancio hanno la stessa probabilità, ogni stringa di risultati ha la stessa probabilità di accadere, e $\mathbb{P}$ è la distribuzione uniforme su $\Omega = \{0, 1\}^n$.

Dimostrazione. È sufficiente mostrare che, per ogni $k = 0, 1, \dots, n$, posto

$$B_k = \{a \in \Omega : a_{n-k+1} = \dots = a_n = 1\}$$

e $B_0 = \Omega$ (sono gli eventi che fissano le ultime k coordinate), vale

$$\mathbb{P}\left(B_k\right) = p^k;$$

se vale questa relazione scegliendo $k = 0$ si mostra che $\mathbb{P}(\Omega) = 1$, quindi lo spazio è di probabilità. Vale:

$$\begin{aligned}\mathbb{P}(B_k) &= \sum_{a \in B_k} \mathbb{P}(\{a\})\\ &= \sum_{a_1, \dots a_{n-k}=0,1} \mathbb{P}(\{(1, \dots 1, a_{k+1}, \dots a_n)\})\\ &= \sum_{a_1, \dots a_{n-k}=0,1} p^k p^{\sum_{i=1}^{n-k} a_i}(1-p)^{n-\sum_{i=1}^{n-k} a_i}\\ &= p^k \sum_{h=0}^{n-k} \binom{n-k}{h} p^h (1-p)^{n-k-h} = p^k,\end{aligned}$$

in cui nell'ultimo passaggio abbiamo usato il fatto che ci sono $\binom{n-k}{h}$ scelte delle entrate $a_1, \dots a_{n-k}$ in cui esattamente h sono uguali a 1.

Inoltre, per verificare l'indipendenza degli eventi A_i dobbiamo considerare le intersezioni

$$A_{i_1} \cap \dots \cap A_{i_k} = \{a \in \Omega : a_{i_1} = \dots = a_{i_k} = 1\}$$

per ogni scelta della famiglia di indici $i_1, \dots, i_k$. Non è difficile convincersi che riordinando gli indici (questo non cambia le probabilità degli eventi come si vede dalla definizione di $\mathbb{P}$) si ottiene $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(B_k)$. Vale quindi

$$\begin{aligned}\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(B_k) = p^k\\ p^k = \mathbb{P}(B_1)^k = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}),\end{aligned}$$

(usando a destra il ragionamento per $k = 1$) da cui la tesi. □

Esempio II.19. Si lanciano due monete equilibrate. Gli eventi $A = \{\text{testa al primo lancio}\}$, $B = \{\text{testa al secondo lancio}\}$, $C = \{\text{è uscita solo una testa}\}$ sono a due a due indipendenti ma non sono globalmente indipendenti.

Osservazione II.20. Si può *erroneamente* essere indotti a pensare che l'indipendenza probabilistica sia legata al concetto di causalità, ma questo, nell'ambito del calcolo delle Probabilità, non ha una definizione univoca nè una generalmente accettata.

Come mostra l'esercizio precedente, due eventi possono essere indipendenti anche in presenza di una qualche relazione causale (il primo evento determina assieme al secondo l'esito del terzo). Viceversa, due eventi possono essere dipendenti anche in assenza di una relazione causale: i dinosauri sono estinti e, per quanto ne sappiamo, nessuno di loro sapeva leggere; dato un certo fossile di dinosauro gli eventi "appartiene a una specie estinta" e "appartiene a una creatura che non sapeva leggere" si verificano sempre assieme, ovvero non sono indipendenti, e tuttavia l'analfabetismo non si annovera tra le cause dell'estinzione dei dinosauri.

E. Entropia di Shannon

Una misura di probabilità può essere uno strumento per codificare (e quantificare) informazione. L'esempio di base consiste nel considerare la frequenza relativa delle lettere dell'alfabeto nella lingua scritta: tali frequenze possono essere viste come una misura di probabilità sull'insieme Ω delle 26 lettere (dell'alfabeto Inglese). Conoscere tale distribuzione risulta utile ad esempio per ottimizzare la codifica dei messaggi scritti, perché si possono usare stringhe brevi per le lettere più comuni (come la E), e riservare stringhe più lunge a lettere rare (come la Q). Per quantificare tale informazione, o meglio la sua mancanza ed ovvero il "disordine", Claude Shannon diede la seguente definizione:

Definizione II.21. *Data una misura di probabilità discreta $\mathbb{P}$ su $\Omega = \{x_1, \dots, x_n\}$, $p_i = \mathbb{P}(\{x_i\})$, la sua entropia è data dalla funzione*

$$H^{(n)}(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i \log(p_i).$$

La scelta del logaritmo naturale, in base 2 o 10 non è rilevante per la nostra discussione. Si noti come, per chiarezza, nella notazione $H^{(n)}$ sottolineiamo la dipendenza da $n = \#\Omega$. Anzitutto, alcuni fatti del tutto intuitivi:

1. è una funzione simmetrica (scambiando p_i e p_j l'entropia non cambia);
2. $H^{(n)}(1, 0, \dots, 0) = 0$;
3. è coerente tra n diversi,

$$H^{(n)}(p_1 = 0, p_2, \dots, p_n) = H^{(n-1)}(p_2, \dots, p_n).$$

Vediamo ora due proprietà caratterizzanti:

4. $H^{(n)}(p_1, \dots, p_n) \leq H^{(n)}\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$.

ovvero la massima entropia (disordine) è data dalla distribuzione uniforme di probabilità. Sebbene intuitivo, questo non è un fatto banale e discende dalla convessità della funzione $x \mapsto x \log x$, che viene scelta appositamente. Infine, vale:

5. data una probabilità su $n \times m$ oggetti, $\Omega = \{x_{11}, \dots, x_{ij}, \dots, x_{nm}\}$, $\mathbb{P}(\{x_{ij}\}) = q_{ij}$, considerando gli eventi $A_i = \{x_{i,1}, \dots, x_{i,m}\}$, $\mathbb{P}(A_i) = p_i$,

$$H^{(nm)}(q_{11}, \dots, q_{ij}, q_{nm}) = H^{(n)}(p_1, \dots, p_n) + \sum_{i=1}^{n} p_i H^{(m)}\left(\frac{q_{i1}}{p_i}, \dots, \frac{q_{im}}{p_i}\right)$$

Questo equivale a dire che l'entropia è data da quella relativa al sistema di alternative A_i, *meno fine* rispetto a quello dei singoli punti $B_{ij} = \{x_{ij}\}$, più la media pesata delle *entropie relative* nei blocchi A_i, calcolate con la probabilità condizionata $q_{ij}/p_i = \mathbb{P}(B_{ij} \mid A_i)$.

Teorema II.22 (Shannon)**.** *Una funzione continua che soddisfa queste 5 proprietà deve avere la forma $cH^{(n)}$, $c > 0$.*

F. Densità di Probabilità

Anzitutto, **una premessa importante**. Nel seguito faremo diffusamente uso di integrali di funzioni sulla retta reale: il rigore matematico necessario per le nostre applicazioni va oltre lo scopo di queste note. Ogni integrale e l'integrabilità delle funzioni possono essere intesi nel senso di Riemann. Ometteremo dettagli tecnici facendone però menzione: essi non sono trascurabili, e il lettore sia conscio che tali omissioni comportano precise limitazioni nei nostri argomenti.

Definizione II.23. *Si chiama densità di probabilità una funzione non-negativa $f : \mathbb{R} \to [0, +\infty)$, integrabile e tale che $\int_{-\infty}^{+\infty} f(x)dx = 1$. Ad ogni densità di probabilità si associa un'unica probabilità definita da:*

$$\mathbb{P}(A) = \int_A f(x)dx, \quad A \subseteq \Omega.$$

Perché questa sia una buona definizione va controllato che $\mathbb{P}$ così definita sia una probabilità nel senso della Definizione II.2. Anzitutto, $\mathbb{P}(\mathbb{R}) = \int_{\mathbb{R}} f(x)dx = 1$ per ipotesi, ed inoltre se $A \cap B = \emptyset$ si ha

$$\mathbb{P}(A \cup B) = \int_{A\cup B} f(x)dx = \int_A f(x)dx + \int_B f(x)dx = \mathbb{P}(A) + \mathbb{P}(B).$$

Omettiamo il controllo della σ-additività. Osserviamo però che per una probabilità così definita **la probabilità di ogni singolo punto è nulla**,

$$\mathbb{P}(\{t\}) = \int_{\{t\}} f(x)dx = 0,$$

e vale più in generale $\mathbb{P}(A) = 0$ per ogni $A \subset \mathbb{R}$ con al più numerabili elementi.

Uno degli esempi di base è la scelta casuale di un numero nell'intervallo $[0, 1]$. Se tale scelta va effettuata in modo "uniforme", non è difficile convinversi che la

probabilità di ogni singolo punto $\mathbb{P}(\{x\})$, $x \in [0,1]$, **non può essere positiva**: in effetti altrimenti calcolando $\mathbb{P}(\Omega)$ ci ritroveremmo a sommare una quantità più che numerabile di numeri positivi, il che rende impossibile soddisfare l'assioma $\mathbb{P}(\Omega) = 1$. Dobbiamo ricorrere invece a una probabilità definita tramite la *densità uniforme* su $[0,1]$,

$$f(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{altrove.} \end{cases}$$

Dato un insieme $S \subset \mathbb{R}$ la sua *funzione indicatrice* è

$$\chi : \mathbb{R} \to \{0,1\}, \quad \chi_S(x) = \begin{cases} 1 & x \in S, \\ 0 & x \notin S. \end{cases}$$

Si consideri lo spazio di probabilità $\Omega = [0,1]$ con

$$\mathbb{P}(A) = \frac{1}{2}|A| + \frac{1}{2}\chi_{\{2\}}(A).$$

Questa è una misura di probabilità: non è né discreta, né è definita da una densità.

Esercizi

Esercizio II.24. Da un mazzo di 40 carte (10 per seme) vengono estratte tre carte, senza rimpiazzo. Qual è la probabilità che le prime due siano di denaro e la terza di coppe? Qual è la probabilità che escano esattamente due carte di denaro (in qualunque ordine)?

Esercizio II.25. Tre turisti arrivano in un paese con cinque alberghi, avendo prenotato ciascuno un albergo in modo del tutto casuale. a) Qual è la probabilità che si trovino tutti in alberghi differenti? b) Qual è la probabilità che si trovino tutti nello stesso albergo?

Esercizio II.26. Un club di tennis molto esclusivo in cui volete entrare vi iscriverà se vincerete due partite *consecutive* su tre contro due suoi campioni, Alfonso e Bertrando. Potete scegliere di affrontarli in uno dei due ordini seguenti: Alfonso, Bertrando, Alfonso oppure Bertrando, Alfonso, Bertrando. Vi sentite più sicuri nel singolo contro Alfonso, quale dei due ordini scegliete?

Esercizio II.27. Se tre eventi A, B, C hanno la stessa probabilità, si determini il valore massimo di $\mathbb{P}(A)$ e si esibisca un esempio di spazio di probabilità con eventi A, B, C aventi le proprietà suddette per cui $\mathbb{P}(A)$ è massima.

Esercizio II.28. In una famiglia ci sono due figli. Se almeno uno dei due è femmina, qual è la probabilità che siano entrambi femmine?

Esercizio II.29. Nella città di Urbopoli, si stima che, tra coloro che abitualmente utilizzano gli autobus pubblici, il 20% lo faccia sistematicamente senza pagare il biglietto. Tra coloro che non pagano il biglietto, il 70% sono individui di età inferiore a 15 anni, mentre tra coloro che lo pagano chi ha meno di 15 anni rappresenta il 40%.

1. Qual è la probabilità che un utente dell'autobus abbia meno di 15 anni?
2. Se un utente ha meno di 15 anni, qual è la probabilità che egli paghi abitualmente il biglietto?
3. Se invece ha 15 anni o più, qual è la probabilità che egli abitualmente non paghi il biglietto?

Esercizio II.30. Un furto è stato commesso a Urbopoli. Sulla base degli indizi raccolti, il commissario pensa che il sig. Rossi ne sia colpevole con probabilità del 40% (cioè, su 100 casi simili, il sig. Rossi sarebbe colpevole in 40 casi). Il commissario viene ora a conoscenza di un nuovo indizio: il colpevole ha i capelli biondi, come il sig. Rossi. Da una statistica, emerge che il 20% degli abitanti di Urbopoli ha i capelli biondi. Come cambia la valutazione del commissario circa la probabilità di colpevolezza del sig. Rossi?

Esercizio II.31. In due lanci di un dado equilibrato, gli eventi "1 al primo lancio" e "somma degli esiti dei due lanci = 3" sono indipendenti?

Esercizio II.32. Siano A, B due eventi, e indichiamo con $p_1 = \mathbb{P}(A)$, $p_2 = \mathbb{P}(A^c)$, e $q_1 = \mathbb{P}(B)$, $q_2 = \mathbb{P}(B^c)$. Verificare che

$$\begin{aligned} H^{(4)}(\mathbb{P}(A \cap B), \mathbb{P}(A \smallsetminus B), \mathbb{P}(B \smallsetminus A), \mathbb{P}((A \cap B)^c)) \\ \le H^{(2)}(\mathbb{P}(A), \mathbb{P}(A^c)) + \le H^{(2)}(\mathbb{P}(B), \mathbb{P}(B^c)), \end{aligned}$$

ovvero l'entropia associata al sistema di alternative generato da A e B è inferiore alla somma di quelle associate ai singoli eventi. Mostrare poi che l'uguaglianza è realizzata se e solo se A, B sono indipendenti.

III. VARIABILI ALEATORIE - I

Consideriamo il lancio di n monete. In un esperimento di questo tipo ha spesso poco interesse chiedersi l'esito di un particolare lancio: l'esito del lancio della quinta moneta non ha per noi nulla di speciale, mentre quasi tutte le domande rilevanti coinvolgono il numero delle teste (o delle croci). In termini più astratti, l'esito dell'esperimento è soggetto a una grande variabilità –gli esiti dei singoli lanci delle monete– ma a noi interessa una quantità che da questa variabilità dipende senza però contenerne tutta l'informazione.

Siamo condotti quindi a studiare una *funzione* dello spazio di probabilità associato all'esperimento: a ogni esito degli n lanci la funzione assocerà il numero di teste (o croci) uscite, ed esiti diversi possono avere chiaramente lo stesso numero di teste. Funzioni come questa, definite su uno spazio di probabilità, sono dette *Variabili Aleatorie* (*Random Variables*).

Questa idea diventa fondamentale per descrivere *osservazioni diverse fatte su uno stesso spazio di probabilità* $(\Omega, \mathcal{F}, \mathbb{P})$, corrispondenti a diverse funzioni di Ω, ovvero variabili aleatorie diverse. In effetti, è proprio nel caso in cui vogliamo confrontare diverse misurazioni di uno stesso esperimento che il semplice formalismo degli spazi di probabilità diventa limitante, mentre introducendo la nozione di variabile aleatoria una gran quantità di domande e operazioni rientrano in modo naturale nel formalismo matematico.

A. La legge di una Variabile Aleatoria

Chiamiamo *Variabile Aleatoria* (v.a.) una funzione $X : \Omega \to \mathbb{R}$ definita su uno spazio di probabilità. Il rigore matematico può richiedere a volte che la probabilità $\mathbb{P}$ sia definita solo su una classe ristretta (σ-algebra) di sottoinsiemi $A \subseteq \Omega$, e di conseguenza bisogna imporre condizioni restrittive nella definizione di variabile aleatoria. Essendo una questione poco rilevante nei successivi argomenti, ne omettiamo la discussione.

Ad una v.a. sono naturalmente associati gli eventi della forma "X prende valori in un insieme $A \subseteq \mathbb{R}$", per i quali introduciamo la seguente notazione:

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\},$$
$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$$

(in cui l'ultima espressione definisce le precedenti).

Esempio III.1. Come visto nella Proposizione II.18, n lanci indipendenti di una moneta non truccata si descrivono con lo spazio di probabilità

$$\Omega = \{0,1\}^n, \quad \mathbb{P}(a) = (1/2)^n.$$

Indicando con 0 l'esito "croce" e con 1 "testa", le variabili aleatorie

$$X = X(a) = a_1 + \cdots + a_n,$$
$$Y = Y(a) = n - X(a) = n - a_1 - \cdots - a_n$$

rappresentano rispettivamente il numero di teste e di croci. Se $A = \{0,1,2\}$, l'evento $\{X \in A\}$ corrisponde a "testa compare al massimo 2 volte (negli n lanci)". Si possono naturalmente definire v. a. più complicate: ad esempio la variabile

$$Z = a_1a_2 + a_2a_3 + \cdots + a_{n-1}a_n$$

conta il numero di coppie di lanci consecutivi con esito testa.

Osservazione III.2. In Probabilità, come abbiamo fatto sopra, una volta specificato qual è lo spazio Ω sottostante si è soliti omettere la dipendenza di una variabile aleatoria $X : \Omega \to \mathbb{R}$ dal suo argomento: questo non causa confusione grazie a una serie di altri accorgimenti notazionali, invitiamo quindi il lettore ad osservare attentamente i simboli che usiamo e ad attenersi alle convenzioni stabilite.

Proposizione III.3. *La funzione di insiemi $\mathbb{P}_X$ è una probabilità su $\mathbb{R}$ detta* legge (di probabilità) *di X.*

Dimostrazione. La parte non banale è il controllo della σ-additività: se $(A_n)_{n\geq 1}$ è una successione di sottoinsiemi di $\mathbb{R}$ a due a due disgiunti, anche le immagini inverse sono disgiunte e si ha

$$\mathbb{P}_X\left(\bigcup_{n=1}^{+\infty} A_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{n=1}^{+\infty} A_n\right)\right)$$
$$= \sum_{n=1}^{+\infty} \mathbb{P}(X^{-1}(A_n)) = \sum_{n=1}^{+\infty} \mathbb{P}_X(A_n).$$

Omettiamo i restanti dettagli. □

Quando due v. a. hanno la stessa legge di probabilità sono dette *equidistribuite*. Non a caso indichiamo le v. a. con lettere maiuscole $X, Y, T...$ e le variabili reali (come quelle di integrazione) con lettere minuscole: questo è importante per evitare confusioni!

Osservazione III.4. Nelle applicazioni, spesso non vengono definiti lo spazio di probabilità Ω e la variabile $X : \Omega \to \mathbb{R}$, ma solo la legge di probabilità $\mathbb{P}_X$. Si può mostrare che, assegnata una probabilità $\mathbb{Q}$ sui sottinsiemi di $\mathbb{R}$, è possibile costruire un insieme Ω, una probabilità $\mathbb{P}$ sui sottinsiemi di Ω ed una variabile $X : \Omega \to \mathbb{R}$ tale che $\mathbb{P}_X = \mathbb{Q}$.

In particolare, quando si sta considerando una singola variabile aleatoria, $\mathbb{P}_X$ riassume tutte l'informazione in gioco: è equivalente avere una variabile aleatoria (cioè una funzione $X : \Omega \to \mathbb{R}$) oppure una distribuzione di probabilità sulla retta reale $\mathbb{R}$. In effetti la costruzione stessa delle v. a. come funzioni di uno spazio Ω sottostante diventa rilevante solo quando ce n'è più di una e si è interessati alle relazioni tra di loro.

Ai due esempi di probabilità su $\mathbb{R}$ (discreta e con densità) introdotti nel capitolo precedente corrispondono due tipi di variabili aleatorie. **Non tutte le v.a. appartengono a uno di questi due tipi**, ma sono i casi che appaiono più frequentemente.

Definizione III.5. *Una variabile aleatoria è detta discreta se la sua immagine $X(\Omega) \subset \mathbb{R}$ è un sottinsieme al più numerabile di $\mathbb{R}$, o equivalentemente se la sua legge di probabilità è discreta.*

Conoscere la legge di probabilità di una v.a. discreta equivale a conoscere la funzione di massa $p_X(x) = \mathbb{P}_X(\{x\})$. Se $A \subseteq \mathbb{R}$ vale

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \sum_{x_i \in A} p_X(x_i),$$

in cui la somma si estende agli x_i tali che $\mathbb{P}(X = x_i) \neq 0$.

Definizione III.6. *Una variabile aleatoria è detta con densità (o* assolutamente continua*) se la sua legge di probabilità è definita da una densità f, cioè se esiste una densità di probabilità f tale che valga*

$$\mathbb{P}_X(A) = \mathbb{P}\{X \in A\} = \int_A f(x)dx.$$

In particolare, se $A = [a, b]$ è un segmento, per una variabile X con densità f vale

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

B. Funzione di Ripartizione e Quantili

Data una generica variabile aleatoria $X : \Omega \to \mathbb{R}$, al fine di studiare la sua legge $\mathbb{P}_X$ è conveniente introdurre un oggetto più familiare (una funzione su $\mathbb{R}$) che comunque continua a codificare tutte le proprietà di $\mathbb{P}_X$.

Definizione III.7. *Si chiama Funzione di ripartizione (*Cumulative Distribution Function*, per brevità* c.d.f.*) della v.a. X la funzione*

$$F_X : \mathbb{R} \to [0, 1], \quad F_X(x) = \mathbb{P}\{X \leq x\}.$$

È evidente dalla definizione che in realtà F_X dipende solo dalla legge di probabilità della v.a. X. Le seguenti proprietà di base delle funzioni di ripartizione costituiscono un esercizio istruttivo.

Esercizio III.8. Sia $F = F_X$ la c.d.f di una variabile aleatoria X. Si mostri che

- F è non decrescente, ovvero se $x < y$ allora $F(x) \leq F(y)$;
- valgono
$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1;$$
- F è continua a destra, ossia per ogni $x \in \mathbb{R}$ vale $F(x_n) \to F(x)$ per ogni successione $x_n \to x$ con $x_n \geq x$.

In effetti, queste proprietà caratterizzano le funzioni di ripartizione, come chiarisce il seguente risultato che riportiamo senza dimostrazione.

Proposizione III.9. *Data una funzione $F : \mathbb{R} \to [0, 1]$ con le proprietà sopra elencate, esiste una ed una sola probabilità $\mathbb{P}$ sui sottinsiemi misurabili di $\mathbb{R}$ tale che $F(t) = \mathbb{P}((-\infty, t])$, ovvero che F sia la c.d.f. $F = F_X$ di una variabile X la cui legge di probabilità sia $\mathbb{P}_X = \mathbb{P}$.*

Dalla funzione di ripartizione F di una v.a. X, si ricava la probabilità che X cada in un dato intervallo (anche infinito): per $a < b$ vale

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

come si verifica facilmente a partire da $\{a < X \leq b\} = \{X \leq b\} \setminus \{X \leq a\}$.

La funzione di ripartizione è ben definita per qualsiasi variabile aleatoria, e possiamo usarla per distinguere vari tipi di variabili aleatorie. La c.d.f. di una variabile aleatoria X discreta che assume valori $x_1, x_2, \ldots$ è una funzione costante a tratti ("a gradini"),

$$F_X(t) = \sum_{x_i \leq t} p(x_i).$$

La funzione esibisce un salto in corrispondenza di ogni punto x tale che $\mathbb{P}(X = x) > 0$ e la probabilità di quel punto è proprio l'ampiezza del salto: si ha pertanto

$$\mathbb{P}\{X = x\} = F(x) - F_-(x),$$

dove con $F_-(x) = \lim_{y \nearrow x} F(y)$ si intende il limite sinistro di F nel punto x.

Quando la variabile ha densità f —e solamente in quel caso!— la sua funzione di ripartizione diventa $F(x) = \int_{-\infty}^x f(t)dt$ ed è ovviamente continua. Nel caso in cui la densità f sia continua a tratti (come sarà in tutti gli esempi che vedremo), o equivalentemente la c.d.f. F sia continua su $\mathbb{R}$ e differenziabile a tratti, partendo dalla formula $F(x) = \int_{-\infty}^x f(t)dt$, si ottiene (in tutti i punti in cui f è continua)

$$f(x) = \frac{dF(x)}{dx};$$

quindi f può essere ottenuta derivando F.

Il fatto che la c.d.f. descriva completamente una singola variabile aleatoria X (nel caso di due o più v. a. la situazione non è così semplice) suggerisce che tra tutti gli eventi associati ad una variabile aleatoria i più interessanti siano quelli della forma $\{X \leq r\}$.

È quindi naturale chiedersi, dato $0 < \beta < 1$, quale sia il il numero r_β tale che $\mathbb{P}(X \leq r_\beta) = F(r_\beta) = \beta$, ovvero qual è la semiretta di valori di X che ha una probabilità assegnata. Questa **non** può essere una definizione per r_β, per due motivi: non sempre esiste un numero r tale che $F(r) = \beta$ e non è detto che tale numero sia univocamente determinato. Il problema è risolto dando la seguente definizione:

Definizione III.10. *Assegnata una v.a. X ed un numero β con $0 < \beta < 1$, si chiama β-quantile il numero*

$$r_\beta = \inf\{r \in \mathbb{R} : F(r) \geq \beta\}, \quad \beta \in (0, 1).$$

La funzione $F^{\leftarrow} : (0, 1) \to \mathbb{R}$,

$$F^{\leftarrow}(t) = \inf\{r \in \mathbb{R} : F(r) \geq t\},$$

si chiama inversa generalizzata *di F.*

Le seguenti proprietà discendono direttamente dalla definizione:

- se F è strettamente crescente, l'inversa generalizzata coincide con l'inversa, $F^{\leftarrow} = F^{-1}$;
- $F^{\leftarrow}$ è sempre non decrescente;
- $F^{\leftarrow}(F(t)) \leq t$ per ogni $t \in \mathbb{R}$;
- $F(F^{\leftarrow}(t)) \geq t$ per ogni $t \in \mathbb{R}$;
- $F^{\leftarrow}(t) \leq s$ se e solo se $F(s) \geq t$.

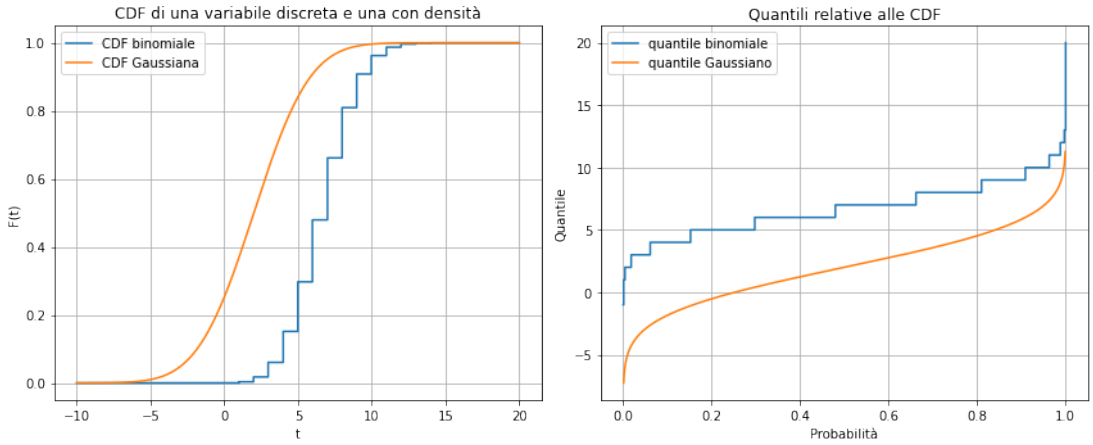


Figura 6: Due funzioni di ripartizione e relative inverse generalizzate. In blu quella di una variabile discreta (binomiale di parametri $n = 20$ e $p = 1/3$) e in arancione quella di una variabile continua (Gaussiana, vedi il capitolo successivo).

C. Variabili Aleatorie Notevoli Discrete

Variabili Binomiali $B(n,p)$. Consideriamo n prove ripetute di un esperimento che ha solo due esiti (o di cui ci interessano solo due esiti), chiamiamo successo uno di questi e sia p con $0 < p < 1$ la probabilità del successo: ricordiamo che un esempio di spazio di probabilità che modellizza questa situazione è stato dato nella Proposizione II.18.

Sia X la variabile che conta il numero dei successi. I valori possibili sono $0, 1, \dots, n$ e vale, per $0 \le h \le n$,

$$\mathbb{P}(X = h) = \binom{n}{h} p^h (1-p)^{n-h}.$$

Infatti, per l'indipendenza delle prove, ogni sequenza di risultati con h successi ed $(n-h)$ insuccessi ha probabilità $p^h \cdot (1-p)^{n-h}$, e il numero di tali sequenze, cioè il numero di modi di disporre gli h successi, è $\binom{n}{h}$. Si noti che la probabilità totale è 1 come conseguenza della formula del binomio di Newton. In generale, una variabile aleatoria $X : \Omega \to \{1, \dots, n\}$ per cui $\mathbb{P}(X = h)$ è data dalla suddetta formula per un qualche $p \in (0,1)$ è detta *variabile Binomiale* $B(n,p)$.

Il seguente esempio mostra ancora una volta come l'informazione codificata da una variabile aleatoria $X : \Omega \to \mathbb{R}$ non sia contenuta nel particolare spazio di probabilità Ω sottostante, quanto nella legge $\mathbb{P}_X$.

Esempio III.11. Dati $n \in \mathbb{N}_0$ e $p \in [0,1]$, si considerino due spazi di probabilità:

$$\Omega = \{a = (a_1, \dots, a_n) \mid a_i = 0, 1\} = \{0,1\}^n,$$
$$\mathbb{P}(\{a\}) = p^{\sum_{i=1}^n a_i} (1-p)^{n - \sum_{i=1}^n a_i},$$

(definito nella Proposizione II.18) e

$$\tilde{\Omega} = \{0, 1, \dots, n\}, \ \tilde{\mathbb{P}}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Definiamo le variabili aleatorie

$$X : \Omega \to \{0, 1, \dots, n\}, \quad X(a) = \sum_{i=1}^n a_i,$$
$$Y : \tilde{\Omega} \to \{0, 1, \dots, n\}, \quad Y(k) = k.$$

Sappiamo già che la legge di X è binomiale $B(n,p)$, ma si verifica immediatamente che anche Y ha legge binomiale $B(n,p)$. In altre parole, $\mathbb{P}_X = \mathbb{P}_Y$ (come uguaglianza tra probabilità sulla retta $\mathbb{R}$) anche se le due variabili sono costruite su spazi di probabilità distinti e alquanto diversi tra di loro.

Osservazione III.12. Più in generale, data una successione $x_1, x_2, \dots \in \mathbb{R}$ e una successione di numeri non-negativi $p_1, p_2, \dots \in [0, \infty)$ tali che $\sum_{i=1}^\infty p_i = 1$, possiamo definire una variabile aleatoria discreta tramite

$$\Omega = \mathbb{N}, \quad \mathbb{P}(\{k\}) = p_k, \quad X(k) = x_k.$$

Questo è *un* modo di costruire una variabile aleatoria X con una data legge di probabilità discreta $\mathbb{P}_X(k) = \mathbb{P}(X = x_k) = p_k$, ma come appena visto di gran lunga non l'unico.

Variabili Geometriche $G(p)$. Consideriamo nuovamente un esperimento che ha solo due esiti, chiamiamo successo uno di questi e sia p con $0 < p < 1$ la probabilità del successo, e ripetiamo l'esperimento fino a quando non avviene un successo. La variabile X che conta l'istante del primo successo (cioè il numero h tale che alla prova h-sima si verifichi il primo successo) è detta *geometrica* di parametro p: i valori possibili di X sono i naturali positivi $1, 2, 3, \dots$ e si ha

$$\mathbb{P}(X = h) = (1-p)^{h-1} p, \quad h \in \mathbb{N}_0.$$

Infatti, detto A_i l'evento "successo alla prova i-sima", per l'indipendenza delle prove abbiamo

$$\begin{aligned}\mathbb{P}(X = h) &= \mathbb{P}(A_1^c \cap A_2^c \cap \dots \cap A_{h-1}^c \cap A_h) \\ &= \mathbb{P}(A_1^c) \cdot \mathbb{P}(A_2^c) \cdots \mathbb{P}(A_{h-1}^c) \cdot \mathbb{P}(A_h) \\ &= (1-p)^{h-1} p.\end{aligned}$$

Le variabili geometriche sono caratterizzate da una interessante proprietà detta *assenza di memoria*, la cui discussione viene lasciata all'Esercizio III.13 qui di seguito.

Variabili Ipergeometriche $I(n,h,r)$. In un'urna ci sono n biglie, di cui $0 \le h \le n$ bianche e $n-h$ nere, e si estraggono $0 \le r \le n$ biglie (senza reinserimento): la variabile aleatoria X che conta quante delle biglie estratte sono bianche si dice *ipergeometrica* di parametri n, h, r.

Contando i casi in cui vengono estratte k biglie bianche e quelli totali si determina la funzione di massa

$$\mathbb{P}(X=k)=\frac{\binom{h}{k}\binom{n-h}{r-k}}{\binom{n}{r}},\quad k=0,\dots,h.$$

Basta infatti osservare che dato k il numero di scelte possibili di k biglie bianche e $r-k$ nere è dato da $\binom{h}{k}\binom{n-h}{r-k}$, mentre il numero totale di scelte di r biglie è $\binom{n}{r}$. Sommando su k il numero di scelte di k biglie bianche e $r-k$ nere si deve ottenere il numero totale delle scelte,

$$\sum_{k=0}^{h}\binom{h}{k}\binom{n-h}{r-k}=\binom{n}{r}$$

(*identità di Vandermonde*) e questo mostra che $\sum_{k=0}^{h}\mathbb{P}(X=k)=1$.

Variabili di Poisson $P(\lambda)$. Dato $\lambda>0$, si dice che una variabile $X:\Omega\to\mathbb{N}$ a valori naturali è una variabile di Poisson di parametro $\lambda>0$ se

$$\mathbb{P}(X=h)=e^{-\lambda}\frac{\lambda^h}{h!},\quad h\in\mathbb{N}.$$

Il fatto che $\sum_{h=0}^{+\infty}\mathbb{P}(X=h)=1$ (ossia la probabilità totale è 1) è una conseguenza dello sviluppo esponenziale $e^{\lambda}=\sum_{h=0}^{+\infty}\lambda^h/h!$.

Si può mostrare (si veda l'Esercizio III.14) che la distribuzione di Poisson di parametro λ è una buona approssimazione della distribuzione binomiale di parametri n e p, quando n è grande, p è piccolo e np vale circa λ. In altre parole, la distribuzione di Poisson conta il numero di successi in prove ripetute, quando il numero n di prove è elevato e la probabilità p di successo è bassa. Per questo, la distribuzione di Poisson viene anche detta distribuzione degli eventi rari, e si usa per modellizzare, ad esempio:

- il numero di particelle α emesse da una sorgente radioattiva in un dato intervallo di tempo;
- il numero di telefonate a un operatore del servizio clienti in un'ora;
- il numero di eruzioni di un certo vulcano in un secolo.

Esercizi

Esercizio III.13. Data una variabile geometrica $X\sim G(p)$, mostrare che per ogni coppia di interi positivi n,h si ha

$$\mathbb{P}\{X=n+h\mid X>n\}=\mathbb{P}\{X=h\}.$$

Viceversa, assumiamo ora sia data una variabile discreta $X:\Omega\to\mathbb{N}_0$ che soddisfa questa ultima equazione (si dice appunto che X è una variabile senza memoria, perché sapere che siamo oltre l'n-esimo passo è come ripartire da quest'ultimo). Si mostri che X è geometrica per un certo parametro p.

Suggerimento: si usi l'equazione dell'assenza di memoria con $n=1$ per ottenere una formula ricorsiva per $p(h)=\mathbb{P}\{X=h\}$.

Esercizio III.14. Consideriamo una legge binomiale $B(n,p_n)$ con parametro p legato a n dalla relazione $p_n=\lambda/n$ per un $\lambda>0$ fissato. In altre parole, la funzione di massa è

$$p^{(n)}(k)=\binom{n}{k}p_n^k(1-p_n)^{n-k}$$

Mostrare che per ogni $k\in\mathbb{N}$,

$$\lim_{n\to\infty}p^{(n)}(k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

ovvero la legge $B(n,p_n)$ converge alla legge di Poisson $P(\lambda)$ per $n\to\infty$.

Suggerimento: si ricordi l'approssimazione di Stirling,

$$\lim_{n\to\infty}\frac{n!}{\sqrt{2\pi n}n^ne^{-n}}=1.$$

Esercizio III.15. Sullo spazio di probabilità $\Omega=\{0,1\}^{\mathbb{N}}$ (successi o insuccessi di infinite prove), definire una probabilità e una variabile aleatoria $X:\Omega\to\mathbb{R}$ avente legge geometrica $G(p)$. Si faccia lo stesso per la legge di Poisson $P(\lambda)$ (sullo stesso Ω).

Esercizio III.16. Si consideri una successione di prove indipendenti di un esperimento, in cui la singola prova ha probabilità di successo $p\in[0,1]$. Osserviamo dette prove fino a quando una quantità predeterminata $r\geq 0$ di esse ha avuto successo, e definiamo la variabile aleatoria X come il numero dei fallimenti osservati (così che $r+X$ è il numero totale di prove osservate). Mostrare che la legge di X (detta legge *binomiale negativa* di parametri r,p) è data da

$$\mathbb{P}(X=k)=\binom{k+r-1}{k}(1-p)^kp^r,\ k\in\mathbb{N},$$

e si verifichi che $\sum_{k=0}^{\infty}\mathbb{P}(X=k)=1$ usando la serie nota

$$\frac{1}{p^r}=\sum_{k=0}^{\infty}\binom{k+r-1}{k}(1-p)^k.$$

Esercizio III.17. Xavier e Yannick hanno un dado a sei facce a testa, e li lanciano finchè non esce il 6. Calcolare la probabilità che impieghino lo stesso numero di lanci.

Esercizio III.18. Il problema di generare (con una macchina) numeri *veramente* casuali è complesso, esula dallo scopo di queste note e richiederebbe una discussione approfondita sul significato di "esito casuale" di un esperimento. Assumendo però di poter generare un numero casuale con distribuzione uniforme sull'intervallo $[0,1]$, possiamo mostrare come generare un numero casuale la cui legge di probabilità è data da una c.d.f. F assegnata.

Sia data una funzione di ripartizione $F:\mathbb{R}\to\mathbb{R}$, ovvero F soddisfa le proprietà che caratterizzano le c.d.f. (che abbiamo preso come ipotesi nell'esercizio precedente). Sia poi $U:\Omega\to[0,1]$ una variabile con densità uniforme definita su uno spazio di probabilità $(\Omega,\mathcal{F},\mathbb{P})$. Si mostri che

$$X:\Omega\to\mathbb{R},\quad X=F^{\leftarrow}(U),$$

è una variabile aleatoria avente c.d.f. $F_X=F$.

Esercizio III.19. Si definisca uno spazio di probabilità $\Omega, \mathbb{P}$ e una variabile aleatoria $X : \Omega \to \mathbb{R}$ che descrivano "il massimo tra i valori di due lanci di un dado a sei facce". Si calcoli la legge di X e si dica se corrisponde a uno degli esempi di questo capitolo.

IV. VARIABILI ALEATORIE - II

A. Variabili Aleatorie Notevoli con Densità

Continuiamo il catalogo di v. a. notevoli, considerando ora casi in cui esiste una funzione di densità f_X (non-negativa, di integrale unitario su tutto $\mathbb{R}$) tale che

$$\mathbb{P}(X \in [a,b]) = \mathbb{P}(X \in (a,b)) = \int_a^b f_X(t)dt$$

(ricordiamo che in tal caso $\mathbb{P}(X = x) = 0$ per ogni $x \in \mathbb{R}$, dunque gli estremi dell'intervallo che definisce l'evento possono essere inclusi o meno senza che cambi la probabilità).

Variabili Uniformi su Intervalli Dati due numeri reali $a < b$, la densità uniforme sull'intervallo $[a,b]$ è costante sull'intervallo e nulla fuori da esso, ovvero

$$f(t) = \begin{cases} \frac{1}{b-a} & a < t < b \\ 0 & \text{altrove} \end{cases}.$$

La relativa funzione di ripartizione è:

$$F(t) = \begin{cases} 0 & t \le a \\ \frac{t}{b-a} & 0 < t \le b \\ 1 & t > b. \end{cases}$$

L'esempio precedentemente menzionato del numero preso a caso tra 0 e 1 corrisponde alla densità uniforme su $[0,1]$.

Variabili Esponenziali La densità esponenziale di parametro $\lambda (\lambda > 0)$ è così definita:

$$f(x) = \begin{cases} \lambda e^{-\lambda t} & t > 0, \\ 0 & t \le 0. \end{cases}$$

Poiché la densità è diversa da 0 solo per x positivo, la variabile X prende solo valori positivi (nel senso che $\mathbb{P}\{X \le 0\} = 0$). Si controlla facilmente che è una densità di probabilità:

$$\int_0^{+\infty} \lambda e^{-\lambda t} dt = -\, e^{-\lambda t}\big|_0^{+\infty} = 1.$$

Con lo stesso calcolo si mostra che la c.d.f. è:

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & t > 0, \\ 0 & t \le 0. \end{cases}$$

La v.a. esponenziale descrive ad esempio il tempo di attesa tra due eventi aleatori, come ad esempio due chiamate ad un call center.

Variabili di Pareto La densità di Pareto di parametri $x_m, \alpha > 0$ è così definita:

$$f(t) = \begin{cases} \alpha x_m^\alpha t^{-1-\alpha} & t > x_m, \\ 0 & t \le x_m. \end{cases}$$

La densità è non nulla solo dopo la soglia x_m, ovvero una variabile di Pareto prende valori solo nella semiretta (x_m, ∞), e al diminuire del parametro $\alpha > 0$ ha una "coda" sempre più "pesante" (*fat tail*). La c.d.f. è

$$F(t) = \begin{cases} 1 & t < x_m \\ 1 - (x_m/t)^\alpha & t \ge x_m. \end{cases}$$

Le variabili di Pareto, anche dette con distribuzione a legge di potenza (*power law*) sono fondamentali per descrivere fenomeni in cui eventi estremi hanno una probabilità cospicua di avvenire. Ad esempio, Vilfredo Pareto le propose per primo come descrizione della distribuzione della ricchezza nella società.

Variabili Gaussiane Standard La funzione $f(x) = e^{-x^2/2}$ tende a 0 molto velocemente per $|x| \to \infty$ e quindi è integrabile, ma non è possibile scrivere la sua primitiva in termini di funzioni elementari. Non possiamo chiarire qui il significato preciso di questa affermazione: per il lettore può significare semplicemente che l'unico modo di rappresentare il valore di $\int_0^t e^{-x^2/2}dx$ per un generico t è ricorrere ad approssimazioni numeriche. Tuttavia, per alcuni particolari valori di t tale integrale assume valori semplici: ad esempio vale

$$\int_{-\infty}^{+\infty} e^{-x^2/2}dx = \sqrt{2\pi}.$$

Di conseguenza, dividendo la funzione $e^{-x^2/2}$ per $\sqrt{2\pi}$ si ottiene una densità di probabilità: si chiama *densità Gaussiana (o Normale) standard*, indicata $N(0,1)$, la funzione

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

La sua funzione di ripartizione (di cui come detto non possiamo dare una rappresentazione più esplicita) è

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}dt.$$

Queste funzioni sono talmente centrali in Probabilità e Statistica che i simboli φ, Φ e q_α sono riservati ad indicare rispettivamente la densità, la c.d.f. e lo α-quantile della variabile $N(0,1)$.

La densità φ è una funzione pari, ovvero $\varphi(x) = \varphi(-x)$, e di conseguenza dati $x \in \mathbb{R}$ e $0 < \alpha < 1$, si ha

$$\Phi(-x) = 1 - \Phi(x) \quad q_{1-\alpha} = -q_\alpha.$$

(Questa uguaglianza non è specifica della Gaussiana standard, ma conseguenza del solo fatto che la densità è una funzione pari). Di conseguenza, se X è una v.a. $N(0,1)$, allora

$$\mathbb{P}\{-t \le X \le t\} = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1$$

per ogni $t \in \mathbb{R}$. Inoltre

$$\Phi(0) = \mathbb{P}\{X \ge 0\} = \mathbb{P}\{X \le 0\} = 1/2.$$

Variabili Gaussiane (non Standard) Data X una v.a. $N(0,1)$, siano $\sigma > 0$ e $m \in \mathbb{R}$ e consideriamo la v.a. $Y = \sigma X + m$. La c.d.f. di Y è data da

$$\begin{aligned} F_Y(t) = \mathbb{P}\{Y \le t\} &= \mathbb{P}\{\sigma X + m \le t\} \\ &= \mathbb{P}\left(X \le \frac{t-m}{\sigma}\right) = \Phi\left(\frac{t-m}{\sigma}\right), \end{aligned}$$

e la sua densità (ottenuta derivando F_Y) è

$$f_Y(t) = \frac{1}{\sigma} f_X\left(\frac{t-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-m)^2}{2\sigma^2}}.$$

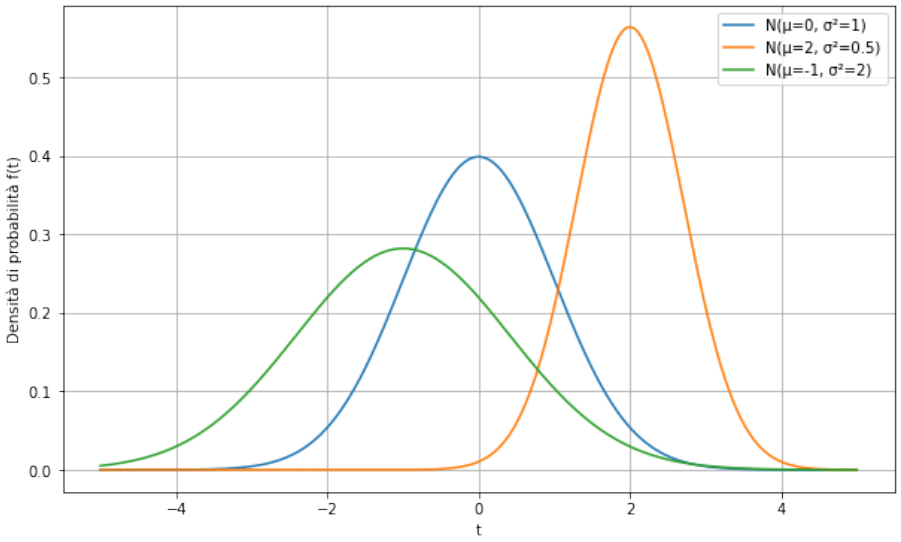


Figura 7: Grafici della densità Gaussiana con diversi parametri. Si noti che gli assi hanno scale diverse, in proporzione 1:1 le campane risultano molto più schiacciate.

Chiamiamo *densità Gaussiana (o normale)* $N(m, \sigma^2)$ quest'ultima funzione, ovvero diciamo che Y è Gaussiana $N(m, \sigma^2)$ se la sua densità ha questa forma.

Si possono ricondurre tutti i calcoli relativi alla c.d.f. di una variabile Gaussiana generica alla c.d.f. standard Φ: è sufficiente sostituire a Y Gaussiana $N(m, \sigma^2)$ la rappresentazione $\sigma X + m$, con X Gaussiana standard. Ad esempio, in particolare,

$$\mathbb{P}\{a < Y < b\} = \mathbb{P}\{(a-m)/\sigma < X < (b-m)/\sigma\}.$$

Riportiamo alcune approssimazioni notevoli: se Y è Gaussiana $N(m, \sigma^2)$,

$$\begin{aligned}\mathbb{P}\{m - \sigma \le Y \le m + \sigma\} &\approx 0.68,\\ \mathbb{P}\{m - 2\sigma \le Y \le m + 2\sigma\} &\approx 0.94,\\ \mathbb{P}\{m - 3\sigma \le Y \le m + 3\sigma\} &\approx 0.997.\end{aligned}$$

Trasformazioni di Variabili con Densità

Supponiamo che $X : \Omega \to \mathbb{R}$ abbia densità f, e data una funzione $h : \mathbb{R} \to \mathbb{R}$ ci domandiamo se la variabile aleatoria

$$Y : \Omega \to \mathbb{R}, \quad Y = h \circ X$$

abbia densità e come calcolarla. Non è detto che Y abbia densità per ogni funzione h, e anche quando Y ha densità non esiste una regola generale per calcolarla. Se è possibile calcolare la c.d.f. di Y,

$$F_Y(y) = \mathbb{P}\{Y \le y\} = \mathbb{P}\{h(X) \le y\},$$

ed essa risulta essere continua su tutto $\mathbb{R}$ e differenziabile (almeno a tratti), derivando si ottiene la densità di Y. Questo è possibile ad esempio sotto le ipotesi del seguente risultato. Nel seguito, col termine generico di intervallo aperto intendiamo sia un intervallo limitato del tipo (a, b) sia una semiretta $(-\infty, b)$ oppure $(a, +\infty)$, sia tutta la retta $(-\infty, +\infty)$.

Proposizione IV.1 (Cambio di Variabile)**.** *Supponiamo che X abbia densità f_X che sia supportata su un intervallo aperto A (cioè tale che f_X sia nulla su A^c), sia $h : A \to B$, con B un altro intervallo aperto, e assumiamo che h sia biunivoca, differenziabile e con inversa differenziabile. La variabile $Y = h \circ X$ ha densità f_Y data dalla formula*

$$f_Y(y) = \begin{cases} f_X\left(h^{-1}(y)\right) \cdot \left|\frac{dh^{-1}(y)}{dy}\right| & y \in B,\\ 0 & y \notin B.\end{cases}$$

Dimostrazione. Diamo una dimostrazione nel caso in cui f sia continua a tratti (cioè F_X sia continua su $\mathbb{R}$ e differenziabile a tratti). Osserviamo che sotto le ipotesi assunte su h, essa può essere solo strettamente crescente o strettamente decrescente. Se h è crescente, allora

$$\begin{aligned}F_Y(y) = \mathbb{P}(Y \le y) &= \mathbb{P}(h(X) \le y)\\ &= \mathbb{P}\left(X \le h^{-1}(y)\right) = F_X\left(h^{-1}(y)\right)\end{aligned}$$

e derivando si ottiene

$$f_Y(y) = f_X\left(h^{-1}(y)\right) \cdot \frac{dh^{-1}(y)}{dy}.$$

Se invece h è decrescente,

$$\mathbb{P}(h(X) \le y) = \mathbb{P}\left(X \ge h^{-1}(y)\right) = 1 - F_X\left(h^{-1}(y)\right),$$

per cui ora derivando appare un segno meno, che però è compensato dal fatto che la derivata di h^{-1} è negativa: per tenere assieme i due casi basta prendere il valore assoluto come nell'enunciato. □

La formula proposta sopra è utile ma non sempre applicabile: è istruttivo svolgere l'esercizio che segue.

Esercizio IV.2. Sia X una variabile con densità uniforme su $[-1, 2]$, si determini se $Z = X^2$ ha densità e se sì la si calcoli.

Sia poi Y una variabile con densità esponenziale di parametro 2, si determini se $W = Y^2$ ha densità e se sì la si calcoli.

B. Valore Atteso, Varianza e Momenti

Abbiamo definito la media campionaria di un insieme di dati $x = (x_1, \dots, x_n)$ come la loro media aritmetica, $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$, per dare una misura del "centro" della

distribuzione dei dati, e di seguito la varianza campionaria per misurare quanto i dati sono "dispersi". Le stesse idee sono alla base delle definizioni di valore atteso e varianza di una variabile aleatoria.

Definizione IV.3 (Valore Atteso)**.** *Sia X una variabile discreta con funzione di massa p_X: Si dice che X ha valore atteso se $\sum_i |x_i| p_X(x_i) < +\infty$, e in tal caso si chiama valore atteso il numero*

$$\mathbb{E}[X] = \sum_i x_i p_X(x_i).$$

Sia poi X con densità f_X: essa ha valore atteso se $\int_{-\infty}^{+\infty} |x| f_X(x)dx < +\infty$ e in tal caso si chiama valore atteso il numero

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

Una definizione più generale valida per tutte le variabili aleatorie, che contiene quelle sopra scritte come casi particolari, necessita di elementi di teoria dell'integrazione generale. Il valore atteso (*expectation*) è chiamato anche speranza matematica o momento primo.

Il valore atteso dipende solo dalla funzione di massa (nel caso discreto) o dalla densità (nel caso con densità). In particolare $\mathbb{E}[X]$ dipende solo dalla legge $\mathbb{P}_X$ di X, e dunque se due variabili sono equidistribuite esse hanno anche lo stesso valore atteso.

Se X prende solo valori positivi, ha sempre senso scrivere $\mathbb{E}[X]$ amettendo che possa assumere il valore $+\infty$. Per le variabili discrete questo significa che i valori $x_1, x_2, \dots$ sono tutti positivi e quindi ha comunque senso $\sum_{i=1}^{+\infty} x_i p(x_i)$; per le variabili con densità questo significa $f(x) = 0$ per $x < 0$ e quindi ha senso $\mathbb{E}[X] = \int_0^{+\infty} x f(x)dx \in [0, +\infty]$. Di conseguenza, presa una generica v.a. X questa ha valore atteso se $\mathbb{E}[|X|] < +\infty$.

Vediamo ora come calcolare il valore atteso di trasformazioni di una variabile aleatoria X, ossia di una nuova variabile Y della forma $g(X)$, $g : \mathbb{R} \to \mathbb{R}$.

Proposizione IV.4. *Sia X discreta: la variabile $g(X)$ ammette valore atteso se $\sum_i |g(x_i)| p(x_i) < +\infty$, e in tal caso*

$$\mathbb{E}[g(X)] = \sum_i g(x_i) p(x_i).$$

Sia X con densità f: la variabile $g(X)$ ha valore atteso se $\int_{-\infty}^{+\infty} |g(x)| f(x)dx < +\infty$, e in tal caso

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

Omettiamo la dimostrazione, e ci limitiamo a proporre degli esempi:

Esercizio IV.5. Calcolare il valore atteso del quadrato dell'esito del lancio di un dado equo tramite la prima definizione di valore atteso e usando la Proposizione appena vista.

Esercizio IV.6. Sia X con densità uniforme su $[0, 1]$ e $Y = X^2$: mostrare che la densità di Y è data da $f_Y(y) = \frac{1}{2\sqrt{y}}$ per $y \in [0, 1]$ e nulla altrove, e che le due formule

$$\mathbb{E}\left[X^2\right] = \int_0^1 x^2 dx = 1/3,$$
$$\mathbb{E}[Y] = \int_0^1 \frac{y}{2\sqrt{y}} dy = 1/3,$$

(la prima dalla Proposizione appena illustrata, la seconda dalla definizione di valore atteso) danno lo stesso risultato.

Proposizione IV.7. *Se la v.a. X (discreta o con densità) ha valore atteso, valgono le seguenti proprietà:*

- *per ogni $a, b \in \mathbb{R}$, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$; in particolare, $\mathbb{E}[b] = b$;*
- *$|\mathbb{E}[X]| \le \mathbb{E}[|X|]$;*
- *se $\mathbb{P}(X \ge 0) = 1$, allora $\mathbb{E}[X] \ge 0$.*

Sono tutte conseguenze abbastanza dirette della definizione, che lasciamo per esercizio.

Definizione IV.8. *Si dice che la v.a. X (discreta o con densità) ammette momento di ordine $n = 1, 2, \dots$ se $\mathbb{E}[|X|^n] < +\infty$ e in tal caso si chiama momento di ordine n il numero $\mathbb{E}[X^n]$.*

Il momento primo è il valore atteso. Se una v.a. discreta assume solo finiti valori, tutti i suoi momenti sono finiti; lo stesso vale per una variabile X con densità f_X diversa da 0 solo su un intervallo limitato. Il risultato che segue mostra che se una v.a. possiede momento n-simo, possiede anche tutti i momenti di ordine inferiore.

Proposizione IV.9. *Siano $1 \le m < n$: se $\mathbb{E}[|X|^n] < +\infty$, anche $\mathbb{E}[|X|^m] < +\infty$.*

Dimostrazione. Limitandoci al caso discreto (quello con densità è del tutto analogo), poiché per ogni $t \in \mathbb{R}$ vale $|t|^m \le |t|^n + 1$, si ha

$$\begin{aligned}\sum_{x_i} |x_i|^m p(x_i) &\le \sum_{x_i} (|x_i|^n + 1) p(x_i) \\ &= \sum_{x_i} |x_i|^n p(x_i) + \sum_{x_i} p(x_i) < +\infty,\end{aligned}$$

da cui la tesi. □

La dimostrazione più generale (valida per ogni tipo di variabile aleatoria) segue dalla importante *disuguaglianza di Jensen*, dalla quale si ricava un risultato più preciso,

$$\mathbb{E}[|X|^m]^{1/m} \le \mathbb{E}[|X|^n]^{1/n}.$$

Proposizione IV.10 (Disuguaglianza di Markov)**.** *Se X è una variabile aleatoria (discreta o con densità) a valori positivi e $a > 0$ vale*

$$a\mathbb{P}\{X \ge a\} \le \mathbb{E}[X].$$

Dimostrazione. Nel caso discreto,

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i) \\ &\geq \sum_{i=1}^{\infty} \mathbb{P}(X = x_i) \cdot \begin{cases} a & \text{se } x_i \geq a \\ 0 & \text{se } x_i < a \end{cases} \\ &= \sum_{i: x_i \geq a}^{\infty} a\mathbb{P}(X = x_i) = a\mathbb{P}(X \geq a).\end{aligned}$$

Nel caso con densità, similmente

$$\begin{aligned}\mathbb{E}[X] = \int_0^{\infty} t f_X(t) dt &\geq \int_a^{\infty} t f_X(t) dt \\ &\geq \int_a^{\infty} a f_X(t) dt = a\mathbb{P}(X \geq a),\end{aligned}$$

usando la stessa idea del caso discreto. □

Nell'ultima Proposizione non si fa alcuna ipotesi sull'esistenza dei momenti: le diseguaglianze hanno senso e sono verificate anche se qualcuno dei valori attesi è infinito.

Definiamo ora la varianza, che indica quanto è dispersa (attorno al valore atteso) una variabile aleatoria.

Definizione IV.11. *Se X ammette momento secondo, la sua varianza è il numero*

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

Si chiama scarto quadratico medio o anche deviazione standard la radice quadrata della varianza di X, cioè

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

Come per il valore atteso, la varianza di X dipende solo dalla legge $\mathbb{P}_X$ di X.

Proposizione IV.12 (Disuguaglianza di Chebyshev)**.** *Se X è una variabile aleatoria e $d > 0$ vale*

$$\mathbb{P}\{|X - \mathbb{E}[X]| > d\} \leq \frac{\mathrm{Var}(X)}{d^2}.$$

Dimostrazione. Questa diseguaglianza deriva immediatamente dalla disuguaglianza di Markov prendendo $Y = (X - \mathbb{E}[X])^2, a = d^2$ e tenendo presente che

$$\mathbb{P}\{|X - \mathbb{E}[X]| > d\} = \mathbb{P}\left((X - \mathbb{E}[X])^2 > d^2\right)$$

(i due membri sono la probabilità dello stesso evento). □

Il caso limite si ha quando la varianza è 0 : la varianza della variabile X è eguale a 0 se e solo se X è costante eccetto che su un insieme trascurabile, cioè $\mathbb{P}\{|X - \mathbb{E}[X]| \neq 0\} = 0$.

C. Momenti di Variabili Aleatorie Notevoli

Concludiamo questa sezione tornando alle v. a. notevoli introdotte precedentemente, per discuterne i momenti.

Variabili Binomiali $B(n, p)$. La variabile di Bernoulli, $n = 1$, assume il valore 1 con probabilità p ed il valore 0 con probabilità $(1-p)$, ed è dunque immediato constatare che per $k \geq 1$

$$\mathbb{E}[X^k] = p, \quad \mathrm{Var}(X) = p - p^2 = p(1-p).$$

Vedremo in seguito che una variabile X Binomiale di parametri n, p può essere vista come somma di n variabili di Bernoulli di parametro p indipendenti: da questo discenderà che

$$\mathbb{E}[X] = np, \quad \mathrm{Var}(X) = np(1-p).$$

Il calcolo diretto si può effettuare applicando (in modo non banale) il binomio di Newton, lo lasciamo come esercizio.

Variabili di Poisson $P(\lambda)$. Poiché assumono solo valori positivi ha senso scrivere

$$\begin{aligned}\mathbb{E}[X] &= \sum_{h=0}^{+\infty} h e^{-\lambda} \frac{\lambda^h}{h!} \\ &= \lambda e^{-\lambda} \sum_{h=1}^{+\infty} \frac{\lambda^{h-1}}{(h-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = \lambda.\end{aligned}$$

Un calcolo simile mostra che $\mathbb{E}\left[X^2\right] = \lambda + \lambda^2$ e quindi $\mathrm{Var}(X) = \lambda$.

Variabili uniformi su intervalli finiti. Se X ha densità uniforme su $[a, b]$:

$$\begin{aligned}\mathbb{E}[X] &= \int_a^b \frac{x}{b-a} dx = \left.\frac{x^2}{2(b-a)}\right|_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}, \\ \mathbb{E}\left[X^2\right] &= \int_a^b \frac{x^2}{b-a} dx = \left.\frac{x^3}{3(b-a)}\right|_a^b \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}.\end{aligned}$$

e quindi $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$.

Variabili esponenziali $E(\lambda)$**.** Vale

$$\mathbb{E}[X] = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda},$$

in particlare il momento primo è finito. Con calcoli simili si trova $\mathbb{E}\left[X^n\right] = \frac{n!}{\lambda^n}$ e quindi $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

Variabili Gaussiane. Se X è Gaussiana standard $N(0, 1)$, poiché l'esponenziale cresce più velocemente di ogni potenza, per ogni n

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x|^n e^{-\frac{x^2}{2}} dx < +\infty,$$

quindi la variabile possiede tutti i momenti. Inoltre se $n = 2h + 1$ è dispari vale

$$\begin{aligned}\mathbb{E}\left[X^n\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-\frac{x^2}{2}} dx \\ &= \lim_{M \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-M}^{+M} x^n e^{-\frac{x^2}{2}} dx = 0.\end{aligned}$$

Rimangono da calcolare i momenti pari. Integrando per parti, poiché $-xe^{-\frac{x^2}{2}}$ è la derivata di $e^{-\frac{x^2}{2}}$, si ha

$$\mathbb{E}\left[X^2\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx$$

$$= \frac{-xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\Bigg|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,dx = 1$$

e dunque $\mathrm{Var}(X) = 1$. Un calcolo simile mostra che

$$\mathbb{E}\left[X^{2h+2}\right] = (2h+1)\mathbb{E}\left[X^{2h}\right],$$

e quindi $\mathbb{E}\left[X^4\right] = 3, \mathbb{E}\left[X^6\right] = 15$ e così via.

Una v.a. Y Gaussiana $N\left(m, \sigma^2\right)$ si scrive nella forma $\sigma X + m$ con X standard, e per la linearità del valore atteso segue quindi:

$$\mathbb{E}[Y] = \mathbb{E}[\sigma X + m] = m$$
$$\mathrm{Var}(Y) = \mathrm{Var}(\sigma X + m) = \sigma^2 \mathrm{Var}(X) = \sigma^2.$$

Lo stesso si può ottenere direttamente dalla definizione con un opportuno cambio di variabili. Per i momenti di ordine superiore si procede analogamente a sopra.

Esercizi

Esercizio IV.13. Mostrare che una variabile esponenziale X di parametro λ è senza memoria, ovvero vale l'eguagliana (per s, t positivi)

$$\mathbb{P}\{X \le s + t \mid X > s\} = \mathbb{P}\{X \le t\}.$$

Viceversa, data una variabile X con densità che assume solo valori positivi e che soddisfa l'equazione suddetta (è senza memoria), si mostri che è una variabile esponenziale per un certo parametro λ.

Suggerimento: usare la formula dell'assenza di memoria per ottenere l'equazione

$$S(t+s) = S(t)S(s), \quad S(t) = \mathbb{P}\{X > s\},$$

e usare il fatto (che diamo per noto) che le uniche soluzioni continue di questa equazione funzionale sono della forma $S(t) = e^{-\lambda t}$ con $\lambda \ge 0$.

Esercizio IV.14. Determinare il β-quantile di una variabile aleatoria con densità esponenziale di parametro $\lambda > 0$.

Esercizio IV.15. Se Y è Gaussiana $N(m, \sigma^2)$, sia $V = \alpha Y + \beta$, con $\alpha, \beta \in \mathbb{R}$, $\alpha \neq 0$. Allora V è Gaussiana $N(\alpha m + \beta, \alpha^2\sigma^2)$.

Esercizio IV.16. Una fabbrica produce viti la cui lunghezza ha distribuzione Gaussiana di media 1.7cm e deviazione standard 0.6cm.

- Calcolare la probabilità che una vite abbia lunghezza inferiore a 1.6cm.
- Quanto deve valere x (in cm) in modo che il 95% delle viti abbia lunghezza almeno x?
- Qual è la probabilità che, estratte 3 viti a caso, esattamente 2 abbiano lunghezza inferiore a 1.6cm?

Esercizio IV.17. Sia $c \in \mathbb{R}$ un parametro e sia $p : \mathbb{N} \to \mathbb{R}$ data da $p(k) = c2^{-k}$. Determinare l'unico valore di c per cui p sia una funzione di massa. Prendiamo ora c pari a tale valore e sia X una v.a. discreta con funzione di massa p. Dire per quali valori di $\alpha \in \mathbb{R}$ la v.a. $e^{\alpha X}$ ammette momento primo.

Esercizio IV.18. Data una variabile X con densità esponenziale di parametro λ, mostrare che

$$Y = \lfloor X \rfloor = \max\{n \in \mathbb{N} : n \le X\}$$

(la parte intera di X, ovvero $Y = n$ se $X \in [n, n+1)$) è una variabile discreta con distribuzione geometrica, e determinare il parametro p di tale distribuzione.

Esercizio IV.19. I 42 studenti di un corso di calligrafia, dei quali 13 sono mancini, sono divisi per le esercitazioni in due gruppi di eguale numero per sorteggio. Si indichino rispettivamente con X e Y le v.a. il cui valore è il numero di mancini presenti nel primo e nel secondo gruppo.

- Calcolare la funzione di probabilità (o densità discreta) della v.a. X.
- Provare che le v.a. X e Y sono equidistribuite, ma non sono indipendenti.
- È possibile determinare i valori attesi $\mathbb{E}[X]$ e $\mathbb{E}[Y]$ senza un calcolo che usa esplicitamente le leggi di X e Y ?

Esercizio IV.20. Sia $p \in [0, 1]$, e si consideri la funzione

$$f(x) = \begin{cases} p, & 0 < x \le 1 \\ (1-p), & 1 < x \le 2 \\ 0, & x \le 0, x > 2. \end{cases}$$

- Dopo aver riconosciuto che la funzione sopra scritta è una densità di probabilità, scrivere l'espressione della funzione di ripartizione e la formula per il β-quantile (con $0 < \beta < 1$) per una v.a. X che abbia quella densità.
- Calcolare i momenti primo e secondo di una v.a. X che abbia densità $f(x)$. Esaminare quando, al variare di p in $[0, 1]$, la varianza è massima e quando è minima.

Esercizio IV.21. Sia $c \in \mathbb{R}$ un parametro, si consideri la funzione

$$f(x) = \begin{cases} cx^2 & -1 < x < 1 \\ 0 & |x| \ge 1 \end{cases}.$$

Dimostrare che f è densità se e solo se $c = 3/2$. Poniamo ora $c = 3/2$, sia X una v.a. con densità f. Determinare se esistono i momenti primo e secondo di X, e calcolarli. Determinare se $Y = X^3$ ha densità.

Esercizio IV.22. Dati $a, b \in \mathbb{R}$ parametri reali, si consideri la funzione

$$f(x) = \begin{cases} -x & -1 \le x < 0 \\ ax + b & 0 \le x \le 1 \\ 0 & |x| > 1 \end{cases},$$

- Trovare tutti i valori di a e b per i quali f è una densità di probabilità.
- Sia X una v.a. avente densità f (con a e b che soddisfano alle condizioni trovate nel punto precedente): calcolare a e b in modo che si abbia $E[X] = 0$.
- In funzione di a e b, calcolare $\mathbf{P}\left(|X| \ge \frac{1}{2} \mid X \le 0\right)$.

Esercizio IV.23. Dati $a, b \in \mathbb{R}$ parametri reali, si consideri la funzione

$$F_{a,b}(x) = \begin{cases} 0, & \text{se } x \leq 0 \\ a - e^{-x}, & \text{se } 0 < x < 1 \\ a - e^{-1} + b(x-1), & \text{se } 1 \leq x \leq 2 \\ 1, & \text{se } x > 2 \end{cases}$$

- Si determinino i valori di a e b affinché $F_{a,b}$ sia la funzione di ripartizione di una variabile aleatoria X con densità.
- Usando i valori di a e b trovati in (i), scrivere la formula per il calcolo dei β-quantili della variabile aleatoria X per ogni $\beta \in (0, 1)$.
- Sia X la variabile aleatoria al primo punto, scrivere la densità della variabile aleatoria $Y = (X+1)^2$, e calcolare il valore atteso di Y.

Esercizio IV.24. Mostrare che la seguente funzione è la densità di probabilità di una variabile aleatoria:

$$f(x) = \begin{cases} 3x^{-4}, & \text{se } x > 1 \\ 0, & \text{se } x \leq 1 \end{cases}.$$

Se X ha densità $f_X = f$, si determini quali momenti possiede X e li si calcoli. Si determini se la variabile $Y = \log(X)$ ha densità e momento primo (ed in tal caso li si calcoli).

Esercizio IV.25. La temperatura corporea in una persona sana ha distribuzione Gaussiana di media 98.2 gradi Fahrenheit (F) e deviazione standard 0.62 gradi F. Un certo ospedale indica 100.6 come temperatura minima per la febbre.

- Quale percentuale di persone sane sono considerate con la febbre dall'ospedale?
- In quale intervallo centrale si colloca il 60% delle temperature delle persone sane?
- Se misuriamo la temperatura a 10 persone scelte a caso, con quale probabilità la loro temperatura media sarà superiore a 98.7 gradi F?
- Se misuriamo la temperatura a 10 persone scelte a caso, con quale probabilità almeno 6 di loro avranno temperatura superiore ai 98.5 gradi F?

V. DISTRIBUZIONI MULTIVARIATE

A. Variabili Doppie

Sia $\Omega, \mathbb{P}$ uno spazio di probabilità, e consideriamo due v. a. $X, Y : \Omega \to \mathbb{R}$. Il vettore (X, Y) può essere dunque visto come una funzione

$$\Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in \mathbb{R}^2.$$

In analogia al caso univariato definiamo la legge della variabile aleatoria doppia (X, Y) come una probabilità sui sottoinsiemi $A \subseteq \mathbb{R}^2$,

$$\begin{aligned}\mathbb{P}_{(X,Y)}(A) &= \mathbb{P}((X, Y) \in A) \\ &= \mathbb{P}\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\}.\end{aligned}$$

Se $A = A_1 \times A_2$ è un sottoinsieme rettangolare,

$$\{(X, Y) \in A\} = \{X \in A_1, Y \in A_2\}.$$

Al fine di evitare confusioni, precisiamo che la virgola indica l'intersezione di due condizioni:

$$\begin{aligned}\{X \in A_1, Y \in A_2\} &= \{X \in A_1\} \cap \{Y \in A_2\} \\ &= X^{-1}(A_1) \cap Y^{-1}(A_2) = (X, Y)^{-1}(A_1 \times A_2).\end{aligned}$$

L'evento $\{X \in A_1, Y \in A_2\}$ corrisponde a "si verificano *congiuntamente* $X \in A_1$ e $Y \in A_2$".

Data una variabile doppia (X, Y) possiamo considerare separatamente le leggi $\mathbb{P}_X$ e $\mathbb{P}_Y$ (su $\mathbb{R}$) delle due componenti: tali leggi sono dette *distribuzioni marginali* della variabile doppia. Se $I \subseteq \mathbb{R}$,

$$\begin{aligned}\mathbb{P}_X(I) &= \mathbb{P}_{(X,Y)}(I \times \mathbb{R}), \\ \mathbb{P}_Y(I) &= \mathbb{P}_{(X,Y)}(\mathbb{R} \times I).\end{aligned}$$

In particolare la legge congiunta $\mathbb{P}_{(X,Y)}$ determina univocamente le leggi marginali $\mathbb{P}_X$ e $\mathbb{P}_Y$.

Osservazione V.1. **Le distribuzioni marginali non contengono tutta l'informazione della legge** $\mathbb{P}_{(X,Y)}$, ossia non si può ricostruire univocamente quest'ultima a partire dalle marginali, come risulterà evidente nella discussione sulla (in)dipendenza di v. a. a seguire. L'idea intuitiva è che $\mathbb{P}_{(X,Y)}$ codifica anche le relazioni tra le v.a. X e Y, cosa che invece non fanno le leggi marginali $\mathbb{P}_X$ e $\mathbb{P}_Y$.

Definizione V.2. *Una variabile doppia (X, Y) è discreta se la sua immagine è concentrata in un insieme finito o numerabile di punti (x_i, y_j). In questo caso la sua distribuzione di probabilità $\mathbb{P}_{(X,Y)}$ è identificata dalla funzione di massa*

$$p(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$$

e se $A \subseteq \mathbb{R}^2$,

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\{(X, Y) \in A\} = \sum_{(x_i,y_j)\in A} p(x_i, y_j).$$

Proposizione V.3. *Se la variabile doppia (X, Y) è discreta con funzione di massa $p(x_i, y_j)$, le sue componenti sono v. a. discrete con funzioni di massa*

$$p_X(x_i) = \sum_{j=1}^{\infty} p(x_i, y_j),$$

$$p_Y(y_j) = \sum_{i=1}^{\infty} p(x_i, y_j).$$

Dimostrazione. Notiamo che $(X = x_i) = \bigcup_j (X = x_i, Y = y_j)$ e questi ultimi insiemi sono a due a due disgiunti. Si ha pertanto

$$\begin{aligned}p_X(x_i) &= \mathbb{P}(X = x_i) \\ &= \sum_{y_j} \mathbb{P}(X = x_i, Y = y_j) = \sum_{y_j} p(x_i, y_j),\end{aligned}$$

l'altra uguaglianza è analoga. □

Esercizio V.4. Se X e Y sono v.a. reali discrete, allora (X, Y) è una v.a. doppia discreta.

Definizione V.5. *Una variabile doppia (X, Y) ha densità se esiste una funzione $f : \mathbb{R}^2 \to [0, \infty)$ integrabile e con $\iint_{\mathbb{R}^2} f(x, y)dxdy = 1$ tale che valga, per $A \subseteq \mathbb{R}^2$*

$$\mathbb{P}_{(X,Y)}(A) = \mathbb{P}\{(X, Y) \in A\} = \iint_A f(x, y)dxdy.$$

Osservazione V.6. Non ci addentriamo nei dettagli della definizione degli integrali doppi: ai fini di queste note basta assumere di poter integrare funzioni di due variabili su sottoinsiemi (misurabili) di $\mathbb{R}^2$, e che per funzioni non-negative integrate su insiemi rettangolari $A = A_1 \times A_2$ vale:

$$\begin{aligned}\iint_A f(x, y)dxdy &= \int_{A_1} \left(\int_{A_2} f(x, y)dy \right) dx \\ &= \int_{A_2} \left(\int_{A_1} f(x, y)dx \right) dy\end{aligned}$$

(formula nota come Teorema di Fubini-Tonelli), cioè l'integrale doppio coincide con un integrale iterato e l'ordine in cui si integrano le variabili non conta.

Proposizione V.7. *Se (X, Y) ha densità $f(x, y)$, anche X e Y hanno densità rispettivamente date da*

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{+\infty} f(x, y)dy, \\ f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y)dx.\end{aligned}$$

Dimostrazione. Notiamo che per ogni $A \subseteq \mathbb{R}$, $\{X \in A\} = \{X \in A, Y \in \mathbb{R}\}$, quindi

$$\begin{aligned}\mathbb{P}\{X \in A\} &= \mathbb{P}\{X \in A, Y \in \mathbb{R}\} \\ &= \int_A \left(\int_{\mathbb{R}} f(x, y)dy \right) dx.\end{aligned}$$

Dunque $f_X(x) = \int_{\mathbb{R}} f(x, y)dy$ è la densità della v.a. X. Analogamente si ottiene la formula per la densità f_Y. □

A differenza del caso discreto, se X e Y sono v.a. con densità, non è detto che (X, Y) abbia densità. Ad esempio, si può dimostrare che (X, X) non ha densità.

Le definizioni e proprietà di questo paragrafo si estendono in modo del tutto analogo a vettori di n v. a.,

$$\Omega \ni \omega \mapsto (X_1(\omega), \ldots, X_n(\omega)) \in \mathbb{R}^n.$$

B. Indipendenza di Variabili Aleatorie

Definizione V.8. *Due v. a. X e Y sono dette indipendenti se, presi comunque A, B sottinsiemi di $\mathbb{R}$, gli eventi $\{X \in A\}$ e $\{Y \in B\}$ sono indipendenti, cioè se vale*

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

Più in generale, le v. a. $X_1, \dots, X_n : \Omega \to \mathbb{R}$ si dicono indipendenti se presi comunque $A_1, \dots, A_n \subseteq \mathbb{R}$ vale

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n).$$

Esempio V.9. Consideriamo n prove ripetute, in cui ciascuna prova abbia come esiti il successo con probabilità p e l'insuccesso con probabilità $1-p$, definendo lo spazio di probabilità

$$\Omega = \{0,1\}^n, \quad \mathbb{P}(a) = p^{\sum_{i=1}^n a_i}(1-p)^{n-\sum_{i=1}^n a_i}.$$

Per $i = 1, \dots n$, sia X_i la v.a. che rappresenta l'esito dell'i-simo esperimento, cioè $X_i(a) = a_i$. Le v.a. X_i sono tutte indipendenti, come si può dimostrare osservando che $\{X_i = 1\} = A_i$, l'evento "successo all'i-sima prova". Osserviamo che le X_i hanno tutte la stessa legge di Bernoulli di parametro p. Osserviamo inoltre che $X = X_1 + \dots X_n$ conta il numero di successi in n lanci ed è quindi v.a. binomiale; in altre parole, la somma di n v.a. di Bernoulli di parametro p indipendenti è una v.a. binomiale di parametri n e p.
Supponendo n pari, le v.a. Y_1 = numero di teste nei primi $n/2$ lanci, Y_2 = numero di teste negli ultimi $n/2$ lanci, Y_1 e Y_2 sono indipendenti, in quanto $Y_1 = X_1 + \dots + X_{n/2}$, $Y_2 = X_{n/2+1} + \dots + X_n$ sono funzioni di gruppi disgiunti di v.a. indipendenti.

Esempio V.10. Data una variabile X (non costante), le variabili X e $-X$ *non* sono indipendenti: intuitivamente se conosciamo una abbiamo immediatamente l'altra cambiando di segno, e infatti scegliendo ad esempio $B = -A$ (lo stesso sottoinsieme di $\mathbb{R}$ ma con i segni cambiati) nella definizione vale

$$\mathbb{P}\{X \in A, -X \in -A\} = \mathbb{P}\{X \in A\},$$
$$\mathbb{P}\{X \in A\} \cdot \mathbb{P}\{-X \in -A\} = \mathbb{P}\{X \in A\}^2,$$

che sono due numeri diversi se scegliamo un qualsiasi A tale che $\mathbb{P}\{X \in A\} \neq 0, 1$ (questo è sempre possibile se X non è costante).

Come per gli eventi, l'indipendenza a due a due di v. a. $X_1, \dots X_n$ non implica l'indipendenza della famiglia $X_1, \dots X_n$. È importante avere una caratterizzazione pratica dell'indipendenza:

Proposizione V.11. *Date due variabili discrete X e Y, con immagine rispettivamente nei punti x_i, y_j, queste sono indipendenti se e solo se vale l'eguaglianza tra le funzioni di massa*

$$p(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j), \quad \forall (x_i, y_j).$$

Date due variabili X, Y tali che (X, Y) abbia densità, le variabili sono indipendenti se e solo se vale l'eguaglianza tra densità

$$f(x,y) = f_X(x) \cdot f_Y(y), \quad \forall (x,y).$$

Dimostrazione. Ci limitiamo al caso di variabili discrete. Da una parte, prendendo $A = \{x_i\}$ e $B = \{y_j\}$, si ha

$$\begin{aligned} p(x_i, y_j) &= \mathbb{P}(X = x_i, Y = y_j) \\ &= \mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j) = p_X(x_i)\, p_Y(y_j). \end{aligned}$$

Viceversa, presi $A, B \subseteq \mathbb{R}$ vale:

$$\begin{aligned} \mathbb{P}\{X \in A, Y \in B\} &= \sum_{x_i \in A, y_j \in B} p(x_i, y_j) \\ &= \sum_{x_i \in A} \sum_{y_j \in B} p_X(x_i)\, p_Y(y_j) \\ &= \left(\sum_{x_i \in A} p_X(x_i) \right) \left(\sum_{y_j \in B} p_Y(y_j) \right) \\ &= \mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\}. \quad \square \end{aligned}$$

I prossimi due esempi mostrano che due v. a. doppie possono avere le stesse distribuzioni marginali ma essere diverse: ad esempio perché in un caso le componenti sono indipendenti e nell'altro no.

Esempio V.12. Consideriamo la variabile doppia discreta (X, Y) con funzione di massa

$$p(1,1) = p(1,-1) = p(-1,1) = p(-1,-1) = \frac{1}{4}.$$

Per quanto visto sopra, le due componenti X e Y hanno la stessa funzione di massa, $p_X(1) = p_X(-1) = 1/2$, $p_X = p_Y$. Le variabili X, Y sono inoltre indipendenti: si verifica facilmente che $p(x,y) = p_X(x)p_Y(y)$.

Esempio V.13. Consideriamo una variabile doppia discreta (X, Y) con una funzione di massa leggermente diversa dall'esempio precedente:

$$p(1,1) = p(-1,-1) = \frac{1}{2}.$$

Le due componenti X e Y hanno ancora la stessa funzione di massa $p_X(1) = p_X(-1) = 1/2$, $p_X = p_Y$. Tuttavia X e Y *non sono indipendenti*:

$$p(1,-1) = 0 \neq p_X(1)p_Y(-1) = \frac{1}{4},$$
$$p(-1,1) = 0 \neq p_X(-1)p_Y(1) = \frac{1}{4}.$$

C. Funzioni di variabili indipendenti

La seguente osservazione, che lasciamo come ulteriore ed istruttivo esercizio, è spesso utile per mostrare che due variabili sono indipendenti.

Esercizio V.14. Date X e Y variabili indipendenti, e due funzioni $h, k : \mathbb{R} \to \mathbb{R}$, mostrare che la variabili $h(X)$ e $k(Y)$ sono indipendenti.

Più in generale, funzioni di più variabili indipendenti sono indipendenti se la stessa variabile non compare in due funzioni diverse: ad esempio se X, Y e Z sono indipendenti, lo sono anche $\sqrt{X^2 + Y^2}$ e Z^3 mentre non sono in generale indipendenti $\sqrt{X^2 + Y^2}$ e $\sqrt{X^2 + Z^2}$ (questo ultimo fatto è più difficile da dimostrare).

Data una variabile doppia (X, Y) e una funzione $h : \mathbb{R}^2 \to \mathbb{R}$, è in generale difficile calcolare la legge della variabile aleatoria $h(X, Y)$. In alcuni casi specifici l'ipotesi di indipendenza rende possibile un calcolo esplicito.

Proposizione V.15. *Se X e Y sono rispettivamente Binomiali $B(n,p)$ e $B(m,p)$ e sono indipendenti, $Z = X+Y$ è binomiale $B(n+m,p)$.*

Dimostrazione. Supponiamo dapprima che Y sia di Bernoulli : in tal caso

$$\begin{aligned}\mathbb{P}\{Z=h\} &= \mathbb{P}\{X=h, Y=0\} \\ &\quad + \mathbb{P}\{X=h-1, Y=1\} \\ &= \binom{n}{h} p^h (1-p)^{n-h}(1-p) \\ &\quad + \binom{n}{h-1} p^{h-1}(1-p)^{n-h+1} p \\ &= \binom{n+1}{h} p^h (1-p)^{n+1-h},\end{aligned}$$

ovvero Z è di tipo $B(n+1,p)$. La dimostrazione si completa poi per induzione su m. □

Questo risultato suggerisce una formula generale nel caso in cui le variabili X, Y di cui consideriamo la somma sono indipendenti.

Proposizione V.16. *Supponiamo che $X, Y : \Omega \to \mathbb{N}$ siano variabili discrete a valori naturali e indipendenti, e sia $Z = X+Y$: dette p_X, p_Y e p_Z le funzioni di massa rispettivamente di X, Y e Z, si ha*

$$p_Z(n) = \sum_{h=0}^{n} p_X(h) \cdot p_Y(n-h).$$

Dimostrazione. Vale l'uguaglianza insiemistica

$$\{Z=n\} = \bigcup_{h=0}^{n} \{X=h, Y=n-h\}$$

e di conseguenza

$$\begin{aligned}p_Z(n) &= \mathbb{P}(Z=n) \\ &= \sum_{h=0}^{n} \mathbb{P}(X=h)\mathbb{P}(Y=n-h) \\ &= \sum_{h=0}^{n} p_X(h) p_Y(n-h),\end{aligned}$$

in cui il secondo passaggio fa uso essenziale dell'indipendenza. □

Vale una analoga formula nel caso di variabili con densità.

Proposizione V.17 (Formula della convoluzione)**.** *Siano X e Y indipendenti, con densità rispettive f_X e f_Y, e sia $Z = X+Y$: la variabile Z ha densità data da*

$$\begin{aligned}f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)dx \\ &= \int_{-\infty}^{+\infty} f_Y(y) f_X(z-y)dy.\end{aligned}$$

Dimostrazione. Osserviamo che vale

$$\begin{aligned}A_z &= \{(x,y) \in \mathbb{R}^2 : x+y \le z\} \\ &= \{(x,y) \in \mathbb{R}^2 : x \le z-y\},\end{aligned}$$

e dunque la funzione di ripartizione di Z è data da (usando la sostituzione $x' = x+y$ nell'ultimo passaggio)

$$\begin{aligned}F_Z(z) &= \mathbb{P}(X+Y \le z) \\ &= \iint_{A_z} f_X(x) f_Y(y) dxdy \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-y} f_X(x)dx \right) f_Y(y)dy \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z} f_X(x'-y)dx' \right) f_Y(y)dy,\end{aligned}$$

da cui la tesi discende immediatamente derivando in z sotto il segno di integrale (o applicando il teorema fondamentale del calcolo). □

Usando quest'ultimo risultato (e svolgendo pazientemente i calcoli) si ottiene il seguente fatto, che incoraggiamo il lettore a dimostrare.

Esercizio V.18. Se X e Y sono indipendenti e Gaussiane rispettivamente $N(m_1, \sigma_1^2)$ e $N(m_2, \sigma_2^2)$, allora $Z = X+Y$ è Gaussiana $N(m_1+m_2, \sigma_1^2+\sigma_2^2)$.

D. Covarianza e Correlazione

Proposizione V.19. *Supponiamo che X ed Y abbiano valore atteso, allora $X+Y$ ha valore atteso e valgono le seguenti proprietà:*

- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$;
- *se $X \ge Y$, allora $\mathbb{E}[X] \ge \mathbb{E}[Y]$.*

Dimostrazione. Riportiamo la dimostrazione della prima proprietà solo nel caso delle variabili discrete. Anzitutto $X+Y$ ha valore atteso se $\sum_{x_i,y_j} |x_i+y_j|\, p(x_i,y_j) < +\infty$. Ma

$$\begin{aligned}&\sum_{x_i,y_j} |x_i+y_j|\, p(x_i,y_j) \\ &\le \sum_{x_i,y_j} (|x_i|+|y_j|)\, p(x_i,y_j) \\ &= \sum_{x_i} |x_i| \sum_{y_j} p(x_i,y_j) + \sum_{y_j} |y_j| \sum_{x_i} p(x_i,y_j) \\ &= \sum_{x_i} |x_i|\, p_X(x_i) + \sum_{y_j} |y_j|\, p_Y(y_j) < +\infty.\end{aligned}$$

Poiché le serie convergono assolutamente, possiamo ripetere i passaggi togliendo i valori assoluti (la diseguaglianza diventa eguaglianza) trovando esattamente $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

La seconda proprietà si mostra considerando la variabile $X-Y$, e applicando una proprietà del valore atteso vista nel capitolo precedente. □

Proposizione V.20. *Supponiamo che X e Y abbiano valore atteso e che siano indipendenti: anche XY ha valore atteso e vale la formula*

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Dimostrazione. Di nuovo ci limitiamo al caso discreto: vale

$$\begin{aligned}\sum_{x_i,y_j} |x_i y_j|\, p(x_i,y_j) &= \sum_{x_i,y_j} |x_i|\,|y_j|\, p_X(x_i)\, p_Y(y_j)\\ &= \left(\sum_{x_i} |x_i|\, p_X(x_i)\right)\cdot\left(\sum_{y_j} |y_j|\, p_Y(y_j)\right),\end{aligned}$$

in cui il membro destro è finito per ipotesi: è così provato che XY ha valore atteso. Ripetendo gli stessi passaggi senza valore assoluto, si ottiene la formula $\mathbb{E}[XY] = \mathbb{E}[X]\cdot\mathbb{E}[Y]$. □

Se X e Y sono v.a. indipendenti, allora, per tutte le funzioni $h, k : \mathbb{R} \to \mathbb{R}$, sono anche $h(X)$ e $k(Y)$ sono indipendenti, in particolare, se $h(X)$ e $k(Y)$ hanno valore atteso, vale

$$\mathbb{E}[h(X)k(Y)] = \mathbb{E}[h(X)]\cdot\mathbb{E}[k(Y)].$$

Proposizione V.21 (Disuguaglianza di Schwartz)**.** *Siano X e Y due variabili aleatorie: vale la disuguaglianza*

$$\mathbb{E}[|XY|] \le \sqrt{\mathbb{E}[X^2]}\cdot\sqrt{\mathbb{E}[Y^2]},$$

(in cui, se una delle due variabili non ha momento secondo, il membro destro è infinito).

Omettiamo la dimostrazione, ma sottolineiamo una conseguenza. Se X e Y hanno valore atteso non è detto che il prodotto XY abbia valore atteso (a parte il caso in cui sono indipendenti), tuttavia se X e Y hanno momento secondo, il prodotto XY ha valore atteso.

Definizione V.22. *La covarianza tra due v. a. X e Y aventi momento secondo finito è definita da*

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]\\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

Se $\mathrm{Var}(X) \neq 0$ e $\mathrm{Var}(Y) \neq 0$, il coefficiente di correlazione è

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Quando $\mathrm{Cov}(X,Y) = 0$, quindi $\rho(X,Y)$, le variabili sono dette scorrelate.

Lasciamo la verifica delle proprietà di base della covarianza come esercizio, si veda l'Esercizio V.31. L'Esercizio V.32 invece spiega come il coefficiente di correlazione sia una misura della presenza di una relazione linare tra X e Y, mentre l'Esercizio V.24 chiarisce che la correlazione *non* è una misura di indipendenza.

Esercizi

Esercizio V.23. Se in generale date le distribuzioni marginali $\mathbb{P}_X, \mathbb{P}_Y$ non conosciamo la legge della coppia $\mathbb{P}_{(X,Y)}$, possiamo però sempre costruire artificialmente uno spazio di probabilità in cui due variabili X, Y hanno marginali assegnate $\mathbb{P}_X, \mathbb{P}_Y$ e sono indipendenti. Siano date due leggi di probabilità $\mathbb{P}_1$ e $\mathbb{P}_2$ su $\mathbb{R}$ e si definisca sugli insiemi rettangolari

$$\Omega = \mathbb{R}^2, \quad \mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2),$$

con $A_1, A_2 \subseteq \mathbb{R}$; si assuma di poter estendere la definizione agli insiemi misurabili di $\mathbb{R}^2$, ovvero in questo caso quelli per i quali è definita l'area. Si introducano poi due v. a.

$$X : \Omega = \mathbb{R}^2 \to \mathbb{R}, X(x,y) = x,$$
$$Y : \Omega = \mathbb{R}^2 \to \mathbb{R}, Y(x,y) = y.$$

Si verifichi che $\mathbb{P}_X = \mathbb{P}_1$, $\mathbb{P}_Y = \mathbb{P}_2$ e che le variabili X, Y sono indipendenti.

Esercizio V.24. Due variabili indipendenti sono scorrelate, ma *non* è vero il viceversa: si usi come controesempio una coppia di variabili X, Y in cui X è di tipo uniforme sull'intervallo $[-1, 1]$ e $Y = X^2$.

Esercizio V.25. Mostrare che se X è di Poisson di parametro λ, Y di Poisson di parametro μ, e X, Y sono indipendenti, allora $X + Y$ è di Poisson di parametro $(\lambda + \mu)$.

Esercizio V.26. Determinare la legge di $Z = X + Y$, dove X, Y sono due v.a. indipendenti con densità esponenziale di parametro $\lambda > 0$.

Esercizio V.27. Sappiamo dalla letteratura scientifica che un certo farmaco è efficace contro una data malattia nell'80% dei pazienti. Presi 10 pazienti a caso, calcolare:

- la probabilità che il farmaco sia efficace su almeno 9 di loro;
- il valore atteso e la deviazione standard del numero di pazienti (tra i 10 scelti) su cui il farmaco è efficace.

Le risposte ai punti precedenti cambiano se sappiamo che i pazienti scelti sono parenti tra loro?

Esercizio V.28. In una fabbrica, il numero di guasti in un giorno ha distribuzione di Poisson di parametro 1.5; inoltre supponiamo che il numero di guasti in un giorno non influenzi il numero di guasti in altri giorni.

- Qual è la probabilità che, in un dato giorno, ci siano almeno 2 guasti?
- Qual è la probabilità che, in 2 giorni consecutivi, ci siano almeno 3 guasti?
- Qual è la probabilità che, in una data settimana lavorativa (cioè 5 giorni), in almeno 4 giorni si verifichi al massimo un guasto?

Esercizio V.29. Il numero giornaliero di clienti a uno sportello segue una distribuzione di Poisson di parametro 3. Supponiamo che il numero di clienti in un dato giorno sia indipendente dal numero di clienti in altri giorni.

- Qual è la distribuzione del numero di clienti in una settimana lavorativa (= 5 giorni)?
- Stimare, tramite la disuguaglianza di Chebyshev, la probabilità che in una settimana ci siano almeno 20 clienti.

Esercizio V.30. Siano (X, Y) una coppia di variabili indipendenti ed equidistribuite con densità uniforme su $[0, 1]$. Mostrare che la variabile $Z = \max(X, Y)$ (il massimo tra X e Y) ha la stessa distribuzione della variabile $W = \sqrt{X}$.

Esercizio V.31. Verificare che:

- $\mathrm{Cov}(aX + bY + c, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z)$;
- $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$;
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$;
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$;
- sono equivalenti le seguenti tre proprietà:
  1. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$,
  2. $\mathrm{Cov}(X, Y) = 0$,
  3. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$;
- $|\rho(X, Y)| \leq 1$.

Esercizio V.32. Date due v. a. X, Y vale

$$\min_{a,b\in\mathbb{R}^2} \mathbb{E}\left[(Y - a - bX)^2\right] = \mathrm{Var}(Y) \cdot \left(1 - \rho(X, Y)^2\right).$$

Osservazione: presi a_*, b_* che realizzano il minimo $\min_{a,b\in\mathbb{R}^2} \mathbb{E}[(Y - a - bX)^2]$, la retta $y = a_* + b_* x$ è la migliore approssimazione lineare tra X e Y. Poiché il valore del minimo è proporzionale a $1 - \rho^2$, la relazione tra X e Y è tanto meglio approssimata da una retta, quanto più $|\rho|$ è vicino a 1: in questo senso, il coefficiente di correlazione è una misura della dipendenza *lineare* tra X e Y.

VI. VARIABILI INDIPENDENTI E TEOREMI LIMITE

Lanciamo una moneta 1000 volte: benché non sia possibile fare una previsione esatta del numero di teste, è esperienza comune attendersi che esso si aggiri intorno a 500; in altre parole, la frequenza relativa dell'evento "testa" è vicina a 1/2. I teoremi limite si occupano proprio di verificare e quantificare questo tipo di fenomeni. L'esempio appena proposto ad esempio si formalizza con la legge dei grandi numeri, mentre il teorema centrale del limite quantifica quanto possa oscillare la frequenza dell'esito testa attorno al valore atteso 1/2.

I teoremi limite che discuteremo sono relativi a v. a. che soddisfano le seguenti ipotesi, del tutto naturali ogni volta che si vuole descrivere prove ripetute di uno stesso esperimento.

Definizione VI.1. *Consideriamo una famiglia di v. a. $X_1, \dots, X_n, \dots$ possibilmente infinita. Scriviamo per brevità che esse sono* i.i.d. *(da* independent and identically distributed*) se sono indipendenti e equidistribuite (nel caso di famiglia infinita, diciamo che $X_1, \dots, X_n, \dots$ sono indipendenti se, per ogni n, $X_1, \dots X_n$ sono indipendenti). Equivalentemente, $X_1, \dots, X_n, \dots$ sono i.i.d. se hanno tutte la stessa funzione di ripartizione,*

$$F_{X_n}(t) = \mathbb{P}(X_n \leq t) = F(t),$$

con $t \in \mathbb{R}, n = 1, 2 \dots,$ e se vale

$$\mathbb{P}(X_{k_1} \leq t_1, \dots, X_{k_n} \leq t_n) = F_{X_1}(t_1) \cdots F_{X_n}(t_n)$$

per ogni scelta di indici $k_1, \dots, k_n \in \mathbb{N}$ e valori $t_1, \dots, t_n \in \mathbb{R}$.

A. Teoremi Limite

Indichiamo con

$$\bar{X}_n := \frac{X_1 + \dots X_n}{n}$$

la media aritmetica di n variabili aleatorie, e notiamo che $\bar{X}_n$ è essa stessa una variabile aleatoria. I teoremi limite si interessano del comportamento al limite $n \to \infty$ di funzioni di v.a. i.i.d. quali $\bar{X}_n$. È quindi necessario definire il limite di una successione di variabili aleatorie.

Definizione VI.2. *Date v. a. X e $X_1, X_2, \dots$ sullo stesso spazio di probabilità, si dice che X_n* converge in probabilità *a X per $n \to \infty$ se*

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0.$$

Il limite può anche essere una costante $X = c$.

Intuitivamente, la successione X_n tende in probabilità a X se, per n grande, X_n è vicina a X con probabilità alta. Il seguente risultato offre una condizione sufficiente per la convergenza in probabilità a una costante.

Proposizione VI.3. *Se valgono*

$$\lim_{n\to\infty} \mathbb{E}\left[X_n\right] = c \in \mathbb{R}, \quad \lim_{n\to\infty} \mathrm{Var}\left(X_n\right) = 0,$$

allora la successione $(X_n)_{n\geq 1}$ converge in probabilità alla costante c.

Dimostrazione. Dalla disuguaglianza di Markov applicata alla variabile $(X_n - c)^2$ si ha

$$\mathbb{P}\left(|X_n - c| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[\left(X_n - c\right)^2\right],$$

di cui espandiamo il membro destro:

$$\begin{aligned}
&\mathbb{E}\left[\left(X_n - c\right)^2\right] \\
&\quad = \mathbb{E}\left[\left(X_n - \mathbb{E}\left[X_n\right] + \mathbb{E}\left[X_n\right] - c\right)^2\right] \\
&\quad = \mathbb{E}\left[\left(X_n - \mathbb{E}\left[X_n\right]\right)^2\right] \\
&\qquad + 2\mathbb{E}\left[X_n - \mathbb{E}\left[X_n\right]\right] \cdot \left(\mathbb{E}\left[X_n\right] - c\right) \\
&\qquad + \left(\mathbb{E}\left[X_n\right] - c\right)^2 \\
&\quad = \mathrm{Var}\left(X_n\right) + \left(\mathbb{E}\left[X_n\right] - c\right)^2.
\end{aligned}$$

Di conseguenza $\lim_{n\to\infty} \mathbb{E}\left[\left(X_n - c\right)^2\right] = 0$, il che conclude la dimostrazione. □

Teorema VI.4 (Legge debole dei Grandi Numeri)**.** *Sia $X_1, X_2, \dots$ una successione di variabili i.i.d. dotate di momento secondo finito, e sia $\mu = \mathbb{E}\left[X_i\right]$ il loro valore atteso. Allora $\bar{X}_n$ converge in probabilità a μ per $n \to \infty$, cioè, per ogni $\varepsilon > 0$ si ha*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

Dimostrazione. Posto $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$, si ha

$$\begin{aligned}
\mathbb{E}\left[\bar{X}_n\right] &= \frac{\mathbb{E}\left[X_1\right] + \dots + \mathbb{E}\left[X_n\right]}{n} = \mu, \\
\mathrm{Var}\left(\bar{X}_n\right) &= \frac{\mathrm{Var}\left(X_1\right) + \dots + \mathrm{Var}\left(X_n\right)}{n^2} \\
&= \frac{\mathrm{Var}\left(X_1\right)}{n} \xrightarrow{n\to\infty} 0,
\end{aligned}$$

dove abbiamo usato

$$\mathrm{Var}(X_1 + \dots + X_n) = \mathrm{Var}(X_1) + \dots \mathrm{Var}(X_n)$$

per l'indipendenza delle X_i. La tesi segue grazie alla Proposizione precedente. □

Esempio VI.5. Il lancio ripetuto di una moneta è descritto da una successione $X_1, X_2, \dots$ di variabili i.i.d. di Bernoulli di parametro $p = 1/2$. La frequenza dell'esito testa (l'esito $X_i = 1$) è data dalla variabile $\bar{X}_n$, che per la legge dei grandi numeri converge in probabilità al valore atteso $\mathbb{E}[X_1] = 1/2$, ovvero alla probabilità dell'esito testa in un singolo lancio.

Una conseguenza della legge dei grandi numeri è il comportamento, per n grande, della variabile aleatoria

$$S_n^2 := \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1},$$

che corrisponde alla varianza campionaria degli esiti delle v.a. X_i.

Proposizione VI.6. *Sia $X_1, X_2, \dots$ una successione di variabili i.i.d. dotate di momento quarto, e sia $\sigma^2 = \mathrm{Var}(X_i)$ la loro varianza. Per $n \to \infty$, S_n converge in probabilità a σ, ovvero*

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}\left(\left|S_n^2 - \sigma^2\right| > \varepsilon\right) = 0.$$

Dimostrazione. Premettiamo questo fatto (senza dimostrazione), che useremo nel seguito: se $(Y_n)_n$ tende a Y in probabilità e $(Z_n)_n$ tende a Z in probabilità, allora $(Y_nZ_n)_n$, rispettivamente $(Y_n+Z_n)_n$, tende in probabilità a YZ, rispettivamente $Y+Z$. Sviluppando i quadrati nella definizione di S_n^2, si vede che

$$S_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - n\bar{X}_n^2\right).$$

Per $n\to\infty$, $n/(n-1)$ tende a 1. Poiché le X_i sono i.i.d. (con momento quarto), le X_i^2 sono i.i.d. (con momento secondo), quindi per la L.G.N. abbiamo $\frac{1}{n}\sum_{i=1}^n X_i^2 \to \mathbb{E}[X_1^2]$, dove la convergenza è in probabilità. Sempre per la L.G.N. abbiamo $\bar{X}_n \to \mathbb{E}[X_1]$ e quindi $\bar{X}_n \to \mathbb{E}[X_1]$, sempre con convergenza in probabilità. Quindi vale la convergenza in probabilità

$$S_n^2 \to \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \sigma^2. \qquad \square$$

Il Teorema Centrale del Limite richiede una diversa nozione di convergenza, la cui definizione in generale non è elementare: ci restringiamo quindi al caso in cui la v.a. limite ha c.d.f. continua, che include tutti i casi rilevanti ai fini di queste note.

Definizione VI.7. *Siano $(X_n)_{n\geq 1}$ un successione di v.a. ed X una variabile aleatoria, siano rispettivamente F_n ed F le funzioni di ripartizione di X_n ed X e supponiamo che F sia continua. La successione $(X_n)_{n\geq 1}$ converge ad X in distribuzione se per ogni t si ha*

$$\lim_{n\to\infty} F_n(t) = F(t).$$

Si noti che la convergenza in distribuzione riguarda solo le leggi delle v. a. coinvolte.

Teorema VI.8 (Teorema Centrale Del Limite)**.** *Sia $X_1, X_2, \ldots$ una successione di variabili i.i.d. dotate di momento secondo finito, con valore atteso $\mathbb{E}[X_i] = \mu$ e varianza $\sigma^2(X_i) = \sigma^2 > 0$. Presi $-\infty \leq a < b \leq +\infty$ si ha*

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx = \Phi(b)-\Phi(a).$$

Equivalentemente, la v.a.

$$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$$

converge in distribuzione a una v.a. Gaussiana standard.

Il nome più comune dato in italiano (e anche in francese) a questo risultato è "teorema del limite centrale". In realtà tale nome sembra essere frutto di un'errata traduzione, dal tedesco "zentraler Grenzwertsatz", termine introdotto da George Pólya: l'aggettivo "centrale" (zentraler) si riferisce al teorema (Satz), non al limite (Grenzwert).

Agli effetti pratici, possiamo dire che nelle ipotesi suddette la variabile

$$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} = \sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}$$

quando n è grande, è approssimativamente Gaussiana standard. Naturalmente n grande non è un termine preciso: una approssimazione ragionevole si ha per almeno $n \geq 50$.

Uno degli usi più frequenti è l'approssimazione di variabili Binomiali con n grande: poiché possiamo scrivere una variabile X Binomiale di parametri n e p come somma di n variabili indipendenti di Bernoulli di parametro p, la variabile $\frac{X-np}{\sqrt{np(1-p)}}$ è approssimativamente $N(0,1)$. In questo caso si possono dare stime più precise sulla numerosità necessaria a una buona approssimazione, che si ottiene con almeno $np(1-p) \geq 15$.

Esempio VI.9. Gli aerei della compagnia Airfly hanno 180 posti, ma la compagnia sa che in media si presentano all'imbarco solo nove passeggeri su dieci, e per questo mette in vendita 195 biglietti (*overbooking*). Ci domandiamo qual è la probabilità che qualche cliente rimanga a terra e di conseguenza la compagnia sia costretta a un risarcimento.

Il numero aleatorio X di clienti che si presentano è una variabile Binomiale di parametri 195 e 0.9, vogliamo calcolare $\mathbb{P}(X \geq 181)$: per il T.C.L. la variabile

$$\frac{X-195\times 0.9}{\sqrt{195\times 0.1\times 0.9}} = \frac{X-175.5}{4.18}$$

è approssimativamente Gaussiana standard. Di conseguenza

$$\begin{aligned}\mathbb{P}(X\geq 181) &= \mathbb{P}\left(\frac{X-175.5}{4.18} \geq \frac{181-175.5}{4.18}\right)\\ &= \mathbb{P}\left(\frac{X-175.5}{4.18} \geq 1.31\right)\\ &\approx 1-\Phi(1.31) \approx 0.095\end{aligned}$$

(il risultato esatto è circa 0.070).

La seguente conseguenza del T.C.L. mostra come si possa rimpiazzare la varianza teorica σ^2 con la varianza campionaria S_n^2:

Proposizione VI.10. *Sia $X_1, X_2, \ldots$ una successione di variabili i.i.d. con valore atteso $\mathbb{E}[X_i] = \mu$ e varianza $Var(X_i) = \sigma^2 > 0$. Presi $-\infty \leq a < b \leq +\infty$ si ha*

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \sqrt{n}\frac{\bar{X}_n-\mu}{S_n} \leq b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx = \Phi(b)-\Phi(a).$$

B. Variabili Chi-Quadro e di Student

Raccogliamo in quest'ultima sezione alcuni esempi di v. a. che emergono in modo naturale nello studio di famiglie di

variabili Gaussiane indipendenti. Introduciamo dapprima le relative definizioni tramite densità di probabilità, e per farlo anzitutto definiamo una delle cosiddette *funzioni speciali*: la Gamma di Eulero,

$$\Gamma(r) = \int_0^{+\infty} x^{r-1}e^{-x}dx, \quad r > 0.$$

Per un generico r questo integrale non ha una rappresentazione in termini di funzioni elementari. Vale $\Gamma(1) = 1$, e se $r > 1$

$$\Gamma(r) = \int_0^{+\infty} x^{r-1}e^{-x}dx = \\ -x^{r-1}e^{-x}\big|_0^{+\infty} + (r-1)\int_0^{+\infty} x^{r-2}e^{-x}dx \\ (r-1)\Gamma(r-1),$$

da cui segue che per n intero positivo si ha $\Gamma(n) = (n-1)!$ e dunque la funzione Gamma estende a numeri non interi il fattoriale.

Definizione VI.11 (Densità Gamma)**.** *Si chiama densità Gamma di parametri $r, \lambda > 0$, indicata con $\Gamma(r, \lambda)$, la funzione così definita:*

$$f(x) = \begin{cases} \frac{1}{\Gamma(r)}\lambda^r x^{r-1}e^{-\lambda x} & x > 0, \\ 0 & x \le 0. \end{cases}$$

Si tratta effettivamente di una densità, come si vede con un cambio di variabili $\lambda x = t$,

$$\int_0^{+\infty} (\lambda x)^{r-1}e^{-\lambda x}\lambda dx = \int_0^{+\infty} t^{r-1}e^{-t}dt = \Gamma(r).$$

Si può osservare che la densità esponenziale di parametro λ corrisponde alla densità $\Gamma(1, \lambda)$.

Proposizione VI.12. *Una variabile X con densità $\Gamma(r, \lambda)$ ha tutti i momenti, e per ogni $\beta > 0$ vale*

$$\mathbb{E}\left[X^\beta\right] = \frac{\Gamma(r+\beta)}{\Gamma(r)\lambda^\beta}.$$

In particolare,

$$\mathbb{E}[X] = \frac{\Gamma(r+1)}{\Gamma(r)\lambda} = \frac{r}{\lambda},$$
$$\mathbb{E}\left[X^2\right] = \frac{\Gamma(r+2)}{\Gamma(r)\lambda} = \frac{(r+1)r}{\lambda^2},$$
$$\mathrm{Var}(X) = \frac{r}{\lambda^2}.$$

Dimostrazione. Poiché X prende solo valori positivi, vale

$$\mathbb{E}\left[X^\beta\right] = \frac{1}{\Gamma(r)}\int_0^{+\infty} x^\beta \lambda^r x^{r-1}e^{-\lambda x}dx \\ = \frac{1}{\Gamma(r)\lambda^\beta}\int_0^{+\infty} \lambda^{r+\beta}x^{r+\beta-1}e^{-\lambda x}dx \\ = \frac{\Gamma(r+\beta)}{\Gamma(r)\lambda^\beta},$$

da cui segue l'esistenza di tutti i momenti e la formula proposta. □

Osservazione VI.13. Ci si riferisce a r, λ rispettivamente come *parametri di forma e tasso di decadimento* (*shape* e *(decay) rate*). Alcuni autori preferiscono mantenere il parametro *shape* r, ma usare come secondo parametro $s = 1/\lambda$, detto *di scala* (*scale*). Con questa parametrizzazione la densità diventa

$$f(x) = \frac{1}{\Gamma(r)}\frac{1}{s}\left(\frac{x}{s}\right)^{r-1}e^{-\frac{x}{s}}, \quad x > 0$$

(e nulla per $x \le 0$).

Proposizione VI.14. *Date due variabili indipendenti X, Y con densità rispettivamente $\Gamma(r, \lambda)$ e $\Gamma(s, \lambda)$ (con **lo stesso** parametro λ), allora la variabile $(X + Y)$ ha densità $\Gamma(r + s, \lambda)$.*

Omettiamo la dimostrazione, che consiste in una applicazione laboriosa della formula della convoluzione vista nella Proposizione V.17.

Proposizione VI.15 (Densità Chi-quadro χ^2)**.** *Se $X_1, \dots, X_n$ sono variabili $N(0, 1)$ indipendenti la variabile $\left(X_1^2 + \cdots + X_n^2\right)$ ha densità $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. Tale densità è anche detta densità Chi-quadro con n gradi di libertà ($n \in \mathbb{N}_0$), e indicata con $\chi^2(n)$.*

Dimostrazione. Alla luce della Proposizione VI.14 è sufficiente provare che, se X è Gaussiana standard, X^2 è $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$, si procede poi per induzione su n. Sembra naturale cercare di applicare la formula di cambio di variabile della Proposizione IV.1 con $h : \mathbb{R} \to [0, \infty)$, $h(x) = x^2$, ma questa funzione *non soddisfa le ipotesi necessarie perché non è bigettiva.*

Determiniamo invece direttamente la c.d.f. di $Z = X^2$,

$$F_Z(t) = \mathbb{P}(X^2 \le t) \\ = \begin{cases} 0 & t < 0, \\ \mathbb{P}(-\sqrt{t} \le X \le \sqrt{t}) = 2F_X(\sqrt{t}) - 1 & t \ge 0, \end{cases}$$

dunque la densità di probabilità si ricava calcolando $f_Z(t) = \frac{d}{dt}F_Z(t)$, ed poiché per $t > 0$

$$2\frac{d}{dt}F_X(\sqrt{t}) = \frac{2}{\sqrt{2\pi}}\frac{d}{dt}\int_{-\infty}^{\sqrt{t}} e^{-x^2/2}dx \\ = \frac{2}{\sqrt{2\pi}}\cdot\frac{e^{-t/2}}{2\sqrt{t}} = \frac{e^{-t/2}(t)^{-1/2}}{2^{1/2}\Gamma(1/2)},$$

dove abbiamo usato che $\Gamma(1/2) = \sqrt{\pi}$, si ha la tesi. □

La motivazione per la scelta del termine *gradi di libertà* diventerà chiara con le applicazioni statistiche. Faremo uso frequente dello α-quantile della variabile chi-quadro, che indichiamo con $\chi^2_{\alpha,n}$, e per i valori dei quali si ricorre ad approssimazioni numeriche come già nel caso delle v.a. Gaussiane.

Poiché la variabile $\chi^2(n)$ è in realtà una $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, ne conosciamo già i momenti e sappiamo che due variabili indipendenti X con densità $\chi^2(n)$ e Y con densità $\chi^2(m)$ hanno somma $(X + Y)$ con densità $\chi^2(n + m)$.

Vediamo ora utili approssimazioni della variabile $\chi^2(n)$ per n è grande, numericamente valide indicativamente a partire da $n \ge 80$. Una variabile C_n con densità $\chi^2(n)$ si può ottenere definendo $C_n = X_1^2 + \cdots + X_n^2$ con X_i Gaussiane standard indipendenti, per cui dai teoremi limite visti otteniamo che se $n \to \infty$:

- per la legge dei Grandi Numeri C_n converge a 1 in probabilità, quindi $\frac{C_n}{n} \approx 1$;

- per il teorema Centrale del Limite $\frac{C_n-n}{\sqrt{2n}}$ converge in distribuzione alla variabile $N(0,1)$, quindi $\frac{C_n-n}{\sqrt{2n}}$ è approssimativamente Gaussiana standard.

Proposizione VI.16 (Densità di Student)**.** *Siano X, C_n due variabili indipendenti con densità rispettivamente $N(0,1)$ e $\chi^2(n)$, $n \geq 1$. La variabile*

$$T_n = \frac{X}{\sqrt{\frac{C_n}{n}}} = \sqrt{n}\frac{X}{\sqrt{C_n}}$$

ha densità

$$f_{T_n}(t) = \frac{\Gamma(n/2+1/2)}{\sqrt{n\pi}\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-n/2-1/2},$$

detta densità di Student a n gradi di libertà.

Omettiamo la dimostrazione, e ci limitiamo a esporre alcune proprietà elementari di questo tipo di v.a. che saranno necessarie alle nostre applicazioni.

Proposizione VI.17. *La densità di T_n è una funzione pari. Di conseguenza indicati con $F_n(x)$ e $\tau_{(\alpha,n)}$ rispettivamente la c.d.f. e lo α-quantile della variabile T_n, valgono*

$$F_n(-x) = 1 - F_n(x), \quad \tau_{(\alpha,n)} = -\tau_{(1-\alpha,n)}.$$

La variabile T_n ha momenti fino all'ordine $(n-1)$, e i momenti di ordine dispari, quando esistono, sono nulli.

Anche per i quantili di T_n si ricorre ad approssimazioni numeriche. Se n è grande, sappiamo che una variabile C_n con densità $\chi^2(n)$ si approssima con $\frac{C_n}{n} \approx 1$, dunque ci aspettiamo che T_n sia approssimativamente Gaussiana standard:

Teorema VI.18. *Consideriamo, per ogni n, una variabile T_n di Student a n gradi di libertà: la successione $(T_n)_{n\geq 1}$ converge in distribuzione alla variabile $N(0,1)$.*

Anche di questo fatto omettiamo la dimostrazione: per quello che riguarda le applicazioni pratiche, per n grande una variabile di Student si può tranquillamente rimpiazzare con una variabile Gaussiana standard. Va però notata una differenza importante tra le due: rispetto alla densità $N(0,1)$, la densità di Student va a 0 più lentamente per $|x| \to \infty$ (si parla di "code pesanti" o "fat tails"). In particolare, a differenza delle variabili Gaussiane e Chi-Quadro, le variabili di Student *non hanno tutti i momenti finiti.*

Concludiamo questo paragrafo riassumendo come variabili chi-quadro e di Student emergono naturalmente nello studio di variabili Gaussiane indipendenti.

Teorema VI.19 (di Cochran)**.** *Siano $X_1, \ldots, X_n$ variabili Gaussiane i.i.d. $N(m, \sigma^2)$; valgono i seguenti risultati:*

1. *le variabili $\bar{X}_n$ e S_n^2 sono indipendenti;*
2. *la variabile $\bar{X}$ ha densità $N(m, \sigma^2/n)$;*
3. *la variabile $\frac{n-1}{\sigma^2}S_n^2$ ha densità $\chi^2(n-1)$;*
4. *la variabile $T = \sqrt{n}\frac{(\bar{X}_n - m)}{S}$ ha densità di Student $T(n-1)$.*

Ad eccezione del primo punto (di cui omettiamo la dimostrazione) questi fatti seguono direttamente dalla discussione precedente.

Esercizi

Esercizio VI.20. Calcolare in modo approssimato la probabilità che, su 1000 lanci di moneta equilibrata, ci siano almeno 480 teste. Calcolare poi, sempre in modo approssimato, il valore k tale che, con probabilità del 95%, testa compaia almeno k volte (su 1000).

Esercizio VI.21. Data una variabile Gaussiana standard X (ovvero $N(0,1)$) e una funzione $f : \mathbb{R} \to \mathbb{R}$ pari (ovvero $f(x) = f(-x)$) si mostri che le v. a. $X, f(X)$ sono scorrelate ma (ovviamente) non indipendenti.

Si consideri poi una coppia di v. a. X, Y con densità congiunta data da

$$f(x,y) = f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{1-r^2}}e^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}}$$

con $r \in [-1,1]$. Si mostri che le variabili X, Y sono entrambe Gaussiane $N(0,1)$, e che le seguenti affermazioni sono equivalenti:

- X, Y sono scorrelate;
- X, Y sono indipendenti;
- $r = 0$.

(ovvero ognuna implica ambo le altre).

Esercizio VI.22. Una ditta produce certi componenti elettronici dei quali circa il 20% sono difettosi: questi componenti sono esportati in scatole da 400 pezzi e la ditta si impegna a sostituire integralmente la scatola se il numero di pezzi difettosi è superiore a 90. (i) Qual è (approssimativamente) la probabilità che la ditta debba sostituire una scatola? (ii) Se si vuole che la probabilità di dover sostituire una scatola sia inferiore a 0.05, come deve migliorare la produzione (cioè di quanto deve -approssimativamente- scendere la percentuale di pezzi difettosi)?

Esercizio VI.23. Sia X una variabile di Student a $n \geq 1$ gradi di libertà. Determinare, al variare di n, quali sono i momenti $\mathbb{E}[X^p]$ che sono ben definiti, e calcolare valore atteso e varianza di X (quando esistono). *Suggerimento: sfruttare la c.d.f. riportata sopra, e usare un opportuno cambio di variabili nell'integrale per la varianza per ricondursi alla definizione della funzione Gamma.*

Esercizio VI.24. L'altezza degli italiani adulti segue una distribuzione Gaussiana di media 1.77 cm e deviazione standard 10.6cm .

- Qual è la probabilità che l'altezza di un italiano (adulto) sia maggiore di 180 cm?
- Qual è la probabilità che, scelti due italiani a caso, la media delle loro altezze sia maggiore di 180 cm?
- Qual è la probabilità che, scelti 100 italiani a caso, la media delle loro altezze sia maggiore di 180 cm?
- Come cambia la risposta al punto precedente, se la distribuzione dell'altezza ha sempre media 1.77 e deviazione standard 10.6 ma non è necessariamente Gaussiana?

Esercizio VI.25. Il numero giornaliero di errori commessi da un server segue una distribuzione di Poisson di parametro 2.5. Supponiamo che il numero di errori commessi in un dato giorno sia indipendenti dal numero di errori in altri giorni.

- Qual è la probabilità che, in due dati giorni consecutivi, avvengano almeno 3 errori?

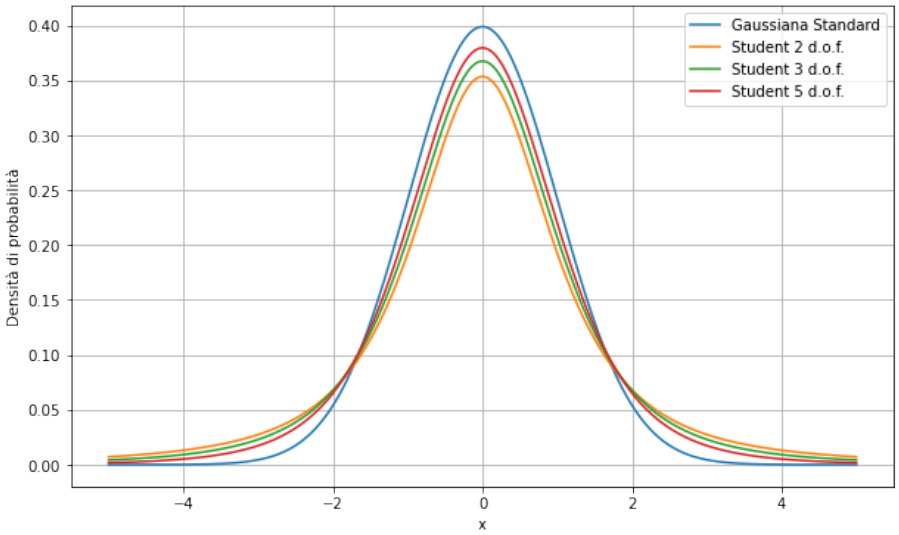


Figura 8: Densità di Student con diversi gradi di libertà, comparate con la densità Gaussiana standard.

- Qual è la probabilità che, in un anno (= 365 giorni), avvengano almeno 900 errori?
- Qual è la probabilità che, in almeno 2 giorni di una data settimana, il server commetta almeno 3 errori?

VII. CAMPIONI STATISTICI E STIMATORI

Lo scopo dell'inferenza statistica è l'analisi di un campione per ricavare informazioni su (un carattere di) una intera popolazione. Ad esempio, un sondaggio sulle intenzioni di voto raccoglie le risposte di un certo numero di intervistati per ricostruire gli orientamenti elettorali dell'intera popolazione, in questo caso l'intero corpo elettorale).

L'assunzione di base della statistica inferenziale è che si possa descrivere la misura del carattere desiderato su un individuo scelto casualmente con una variabile aleatoria, di cui vogliamo determinare la legge, o informazioni su di essa. La casualità risiede quindi nell'estrazione dell'individuo dalla popolazione, dunque le ipotesi che faremo sul modello probabilistico corrispondono ad assunzioni sul modo in cui il campione viene scelto.

Assumiamo che:

- ogni nuova estrazione non sia condizionata dalla precedente (ad esempio la popolazione deve essere molto grande, in modo che scegliere un individuo non cambi la statistica su quelli rimanenti): questo corrisponderà all'ipotesi di *indipendenza*;
- l'estrazione di ogni nuovo individuo per il campione sia effettuata in modo uguale alle precedenti: questo corrisponderà all'ipotesi di *equidistribuzione*.

In termini matematici, il nostro studio della statistica inizia quindi dall'analisi di famiglie finite di v.a. i.i.d., di cui vogliamo determinare informazioni sulla legge a partire dagli esiti.

A. Campioni con legge dipendente da un parametro

Assumiamo nel seguito di avere fissato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ su cui sono definite tutte le v.a. che consideriamo.

Definizione VII.1. *Data $F = F_X$ c.d.f. di una v.a. X, una famiglia finita $X_1, \dots X_n$ di variabile aleatorie i.i.d. con legge data dalla c.d.f. F si dice* campione statistico *o* campione aleatorio *della v.a. X di numerosità (o* taglia*) n.*

Osservazione VII.2. È doveroso insistere sull'importanza di usare *lettere minuscole* $x_1, x_2, ...$ per indicare dati numerici (ad esempio risultati di alcune misurazioni) e *lettere maiuscole* $X_1, X_2, ..$ per indicare *variabili aleatorie*. Se i numeri $x_1, \dots, x_n$ sono gli esiti di n misurazioni, questi possono essere *interpretati* come gli esiti $X_1(\omega), \dots, X_n(\omega)$ di n variabili aleatorie, ma *non sono variabili aleatorie essi stessi*.

Sia $(X_1, \dots X_n)$ un campione i.i.d. di una v.a. X. Senza ulteriori informazioni sulla distribuzione $\mathbb{P}_X$ di X (equivalentemente sulla sua c.d.f.) non si può dire molto di più di quello che è stato fatto, a meno di addentrarsi nella *statistica non parametrica*, che esula dagli obbiettivi di queste note.

Assumeremo invece che la distribuzione di probabilità $\mathbb{P}_X$ sia *parzialmente specificata*, ovvero di volerla identificare all'interno di una famiglia di probabilità (equivalentemente c.d.f.) dipendenti da un opportuno *parametro* $\theta \in \Theta \subseteq \mathbb{R}$, o da più parametri $\theta_1, \dots \theta_m$. Per evidenziare la dipendenza da θ della legge di un campione statistico, la indicheremo con $\mathbb{P}_\theta$, e di conseguenza indicheremo con

- $F_\theta(t) = \mathbb{P}_\theta(X \le t)$ la c.d.f. del campione,
- f_θ e p_θ rispettivamente la densità di probabilità o la funzione di massa quando esse sono definite,
- $\mathbb{E}_\theta$ il valore atteso rispetto $\mathbb{P}_\theta$, da calcolarsi con la definizione usuale utilizzando f_θ o p_θ rispettivamente nei casi con densità o discreto.

Esempio VII.3. Un campione di variabili di legge Bernoulli di parametro p può descrivere un esperimento che può avere successo o fallire, in questo caso si cerca di stimare il parametro $\theta = p$ che può assumere valori $\Theta = [0, 1]$. In questo caso il parametro da stimare coincide con il valore atteso della distribuzione.

Un campione di variabili Gaussiane $N(m, \sigma^2)$ può descrivere i risultati di una misura fisica, di cui vogliamo stimare *due* parametri, ovvero il valor medio m e la varianza σ^2.

L'obiettivo della **stima parametrica** è determinare i parametri incogniti a partire dalle osservazioni, cioè in funzione del campione, in modo che il modello probabilistico descriva gli esiti dell'esperimento nel modo migliore.

B. Stima parametrica

Una funzione $g(X_1, \dots, X_n)$ di un campione statistico è chiamata *statistica campionaria*, o semplicemente *statistica*. Tali sono ad esempio la media e varianza campionarie,

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}, \quad S_n^2 = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n-1}.$$

Uno *stimatore* di un parametro θ della distribuzione è una statistica che approssima il valore di θ; la bontà di tale approssimazione è l'oggetto di questo paragrafo. Poiché uno stimatore è una funzione del campione, anch'esso è una variabile aleatoria: realizzazioni diverse $x_1, \dots x_n$ del campione $X_1, \dots X_n$ portano a valori diversi dello stimatore.

Definizione VII.4. *Dato un campione $X_1, \dots X_n$ con legge dipendente da un parametro θ, una statistica $g(X_1, \dots X_n)$ si dice stimatore* corretto *(o* non distorto*,* unbiased*) del parametro θ se ammette momento primo e*

$$\mathbb{E}_\theta[g(X_1, \dots, X_n)] = \theta,$$

cioè la media dello stimatore è il parametro θ.

Quando il parametro coincide con il valore atteso, $\mathbb{E}_\theta[X] = \theta$, come ad esempio nel caso di un campione di variabili Bernoulli di parametro p, si usa solitamente come stimatore la media campionaria $\bar{X}_n$. La varianza campionaria S^2 è lo stimatore abituale di parametri che coincidono con la varianza del campione $\mathrm{Var}_\theta(X)$, come nel caso di un campione Gaussiano di legge $N(m, \sigma^2)$ in cui m è già nota e si desidera stimare σ^2. La seguente proposizione garantisce che media e varianza campionarie sono stimatori corretti.

Proposizione VII.5. *Sia $X_1, \dots, X_n$ un campione statistico, supponiamo che le variabili possiedano momento secondo e siano $\mu = \mathbb{E}[X_i]$ e $\sigma^2 = Var(X_i)$. Valgono:*

$$\mathbb{E}[\bar{X}] = \mu,\ Var(\bar{X}) = \sigma^2/n,\ \mathbb{E}\left[S_n^2\right] = \sigma^2.$$

Dimostrazione. Le prime due uguaglianze sono immediate dalla proprietà di linearità del valore atteso e dall'ipotesi di indipendenza. Per la terza,

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2,$$

e per ogni X_i si ha $\mathbb{E}[X_i^2]=Var(X_i)+\mathbb{E}[X_i]^2=\sigma^2+\mu^2$, per cui ricordando anche le prime due formule si conclude

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right]&=\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right]-n\mathbb{E}\left[\bar{X}^2\right]\\&=n(\sigma^2+\mu^2)-n(\sigma^2/n+\mu^2)\\&=(n-1)\sigma^2.\quad\square\end{aligned}$$

Osservazione VII.6. Il calcolo implica anche

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right]=\frac{n-1}{n}\sigma^2,$$

ovvero $\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ *non* è uno stimatore corretto della varianza. La necessità di usare il denominatore $n-1$ (invece che quello forse più intuitivo n) per ottenere uno stimatore corretto è nota come *correzione di Bessel*.

Definizione VII.7. *Dato un campione $X_1,\dots X_n,\dots$ di infinite v.a. i.i.d. con legge dipendente da un parametro θ, la successione di statistiche $g_n(X_1,\dots X_n)$ si dice uno stimatore* consistente *di θ se, per $n\to\infty$, $g_n(X_1,\dots X_n)$ tende a θ in probabilità, cioè se, per ogni $\varepsilon>0$,*

$$\lim_{n\to\infty}\mathbb{P}_\theta\{|g_n(X_1,\dots X_n)-\theta|>\varepsilon\}=0.$$

In altre parole, quando la taglia del campione diventa molto grande, uno stimatore $g_n(X_1,\dots X_n)$ consistente si avvicina, con alta probabilità, al valore del parametro θ.

La L.G.N. e il suo corollario ci dicono che media campionaria e varianza campionaria sono stimatori *consistenti* di valore atteso e varianza, rispettivamente.

Definizione VII.8. *Dato un campione $X_1,\dots X_{n_0}$ con legge dipendente da un parametro θ, e dati due stimatori corretti $g(X_1,\dots X_n)$ e $h(X_1,\dots X_m)$, $n_0\geq n,m$, che ammettono momento secondo, diciamo che $g(X_1,\dots X_n)$ è più efficiente di $h(X_1,\dots X_m)$ se*

$$\mathrm{Var}_\theta(g(X_1,\dots X_n))\leq\mathrm{Var}_\theta(h(X_1,\dots X_m)).$$

In altre parole, la dispersione di $g(X_1,\dots X_n)$ attorno al valor medio θ è minore o uguale alla dispersione di $h(X_1,\dots X_m)$ attorno a θ.

Dato un campione i.i.d. $X_1,\dots X_n$ con momento secondo finito, la media campionaria $\bar{X}_n$ è sempre più efficiente al crescere di n: infatti $\mathrm{Var}(\bar{X}_n)=\mathrm{Var}(X_1)/n$ è una funzione decrescente di n. Si può dimostrare che anche la varianza campionaria S_n^2 è sempre più efficiente al crescere di n.

C. Massima Verosimiglianza e Metodo dei Momenti

Ci occupiamo ora di come scegliere uno stimatore. Supponiamo di avere un campione statistico la cui legge di probabilità dipende da un parametro $\theta\in\Theta$, nel quale le variabili possono essere discrete con funzione di massa $p_\theta(x)$ oppure aventi densità $f_\theta(x)$.

Definizione VII.9. *Si chiama* ***funzione di verosimiglianza*** *(*likelihood*) la funzione*

$$L:\Theta\times\mathbb{R}^n\to[0,1],$$

definita nel caso discreto da

$$L\left(\theta;x_1,\dots,x_n\right)=\prod_{i=1}^{n}p_\theta(x_i)$$

e nel caso con densità da

$$L(\theta;x_1,\dots,x_n)=\prod_{i=1}^{n}f_\theta(x_i).$$

La verosimiglianza altro non è che, rispettivamente, la funzione di massa o la densità congiunta delle variabili $X_1,\dots,X_n$.

Definizione VII.10. *Si chiama* ***stima di massima verosimiglianza*** *(* maximum likelihood estimation*), se esiste, una statistica campionaria $\widehat{\theta}=\widehat{\theta}(x_1,\dots,x_n)$ tale che*

$$L\left(\widehat{\theta};x_1,\dots,x_n\right)=\max_{\theta\in\Theta}L\left(\theta;x_1,\dots,x_n\right)$$

per ogni scelta $x_1,\dots,x_n$.

La stima di massima verosimiglianza sceglie il (o un) parametro $\hat{\theta}$ che massimizza la probabilità (o la densità, nel caso relativo) dell'esito $x_1,\dots x_n$ effettivamente ottenuto.

Presentiamo anche un secondo modo per scegliere uno stimatore. L'idea del metodo dei momenti è di confrontare i momenti teorici (i valori attesi di X^k)

$$m_k(\theta)=\mathbb{E}_\theta\left[X^k\right]$$

con i *momenti empirici*

$$\sum_{i=1}^{n}\frac{x_i^k}{n}.$$

Poiché la media campionaria è un buon stimatore del valore atteso, è ragionevole prendere, come stima di θ, un valore $\tilde{\theta}$ che realizzi l'uguaglianza tra i momenti teorici e quelli empirici.

Definizione VII.11. *si chiama* ***stima col metodo dei momenti****, se esiste, una statistica campionaria $\tilde{\theta}=\tilde{\theta}(x_1,\dots x_n)$ che permetta di eguagliare alcuni momenti teorici con i corrispondenti momenti empirici, cioè di scrivere, per uno o più interi positivi k,*

$$\mathbb{E}_{\tilde{\theta}}\left[X^k\right]=\frac{1}{n}\sum_{i=1}^{n}x_i^k,\qquad\textit{per ogni }(x_1,\dots x_n).$$

Esempio VII.12. La coppia di parametri media campionaria e varianza empirica è la stima della coppia media e varianza col metodo dei momenti, per $k=1,2$.

Esempio VII.13 (Campione esponenziale). Supponiamo che $X_1,\dots,X_n$ abbiano densità esponenziale di parametro θ con $0<\theta<+\infty$.

Consideriamo osservazioni $x_1,\dots,x_n$ *positive*: non possiamo descrivere dati negativi con variabili esponenziali. Poiché $\mathbb{E}_\theta[X_i]=1/\theta$, la stima col metodo dei momenti si ottiene imponendo $\frac{1}{\theta}=\sum_i\frac{x_i}{n}$, cioè $\tilde{\theta}=\frac{1}{\bar{x}}$.

Allo stesso modo (sempre per dati x_i positivi) la verosimiglianza è

$$L(\theta; x_1, \dots, x_n) = \theta^n e^{-\theta \sum_i x_i}.$$

Questa funzione tende a 0 per $\theta \to 0$ e per $\theta \to \infty$ (quindi ha massimo) e annullando la derivata si ottiene di nuovo $\widehat{\theta} = \frac{n}{\sum_i x_i} = \frac{1}{\overline{x}}$.

Esempio VII.14 (Densità uniformi su un intervallo variabile). Supponiamo dunque che, per $0 < \theta < +\infty$, la densità sia

$$f_\theta(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{altrove} \end{cases}.$$

Partiamo dunque dalle osservazioni $x_1, \dots, x_n$ che di nuovo supponiamo positive. Poiché $\mathbb{E}_\theta[X_i] = \theta/2$, anche questa volta la stima col metodo dei momenti è facile e si ottiene $\widetilde{\theta} = \frac{2}{n} \sum_i x_i = 2\overline{x}$.

Riguardo la stima di massima verosimiglianza: la verosimiglianza è

$$L(\theta; x_1, \dots, x_n) = \begin{cases} \theta^{-n} & x_i \le \theta \quad \forall i \\ 0 & \text{altrimenti} \end{cases},$$

cioè la verosimiglianza è diversa da 0 se e solo se $\theta \ge \max(x_1, \dots, x_n)$, oltre il quale valore è una funzione decrescente. Ne segue $\hat{\theta}(x_1, \dots, x_n) = \max(x_1, \dots, x_n)$.

Esercizi

Esercizio VII.15. Si consideri un campione di variabili di Poisson di parametro θ con $0 < \theta < +\infty$. Si determinino lo stimatore di massima verosimiglianza e la stima con il metodo del momento primo per θ. Si osservi che coincidono e che sono stimatori corretti e consistenti.

Esercizio VII.16. Si consideri un campione di variabili Geometriche di parametro $p \in [0, 1]$. Si determinino lo stimatore di massima verosimiglianza e la stima con il metodo del momento primo per p.

Esercizio VII.17. Si consideri un campione $X_1, \dots, X_n$ con momento primo finito. Si mostri che per ogni scelta di numeri $a_1, \dots, a_n \in \mathbb{R}$ tali che $a_1 + \cdots + a_n = 1$ lo stimatore $a_1 X_1 + \cdots + a_n X_n$ del valore atteso $\mu = \mathbb{E}[X_i]$ è corretto. (Questo mostra in particolare che possono esistere infiniti stimatori corretti di un parametro).

Esercizio VII.18. Si consideri una (singola) variabile aleatoria discreta X a valori in $\{1, \dots, n\}$, la cui legge dipende da un parametro $\theta \in \Theta = \{\theta_1, \dots, \theta_m\}$ che può assumere un numero finito di valori. Uno stimatore è in questo caso una funzione $f(X)$ che si può rappresentare come il vettore $f(1), \dots, f(n)$ dei valori assunti sugli n esiti possibili. Mostrare che se i vettori $v_k = (p_{\theta_1}(k), \dots, p_{\theta_m}(k))$, $k = 1, \dots, n$, sono linearmente indipendenti allora esiste un unico stimatore corretto del parametro θ.

VIII. INTERVALLI DI FIDUCIA

La quotidianità ci ha abituato all'uso degli intervalli di fiducia, che una cattiva traduzione dall'Inglese *confidence interval* spesso chiama "intervalli di confidenza". Ad esempio le proiezioni dei risultati di un referendum vengono annunciate come nel seguente esempio: dopo un'ora di spoglio il SÌ viene dato al 26.2 ± 2.3, dopo due ore e mezza al 25.9 ± 1.2 e infine dopo quattro ore al 26.1 ± 0.4. All'aumentare della numerosità del campione (che in questo caso si avvicina durante lo spoglio a quella della popolazione totale) la misurazione diventa più precisa, viene cioè proposto un intervallo sempre più stretto.

Gli estremi di questi intervalli sono determinati a partire dagli esiti della misurazione del campione, e dal punto di vista probabilistico dobbiamo definirli come v.a. date da funzioni del campione statistico. Un intervallo di fiducia è dunque un intervallo i cui estremi sono calcolati a partire dai valori assunti da $X_1, \dots, X_n$, quindi un *intervallo casuale*, nel quale ci aspettiamo sia contenuto il parametro θ. Proprio perché l'intervallo è casuale, *non è detto che θ appartenga all'intervallo*: la definizione matematica chiede che questo avvenga con probabilità alta.

Definizione VIII.1. *Dato un campione statistico $X_1, \dots, X_n$ di legge $\mathbb{P}_\theta$, $\theta \in \Theta \subseteq \mathbb{R}$, e un numero $\alpha \in (0,1)$, un* intervallo aleatorio

$$I = [a(X_1, \dots, X_n), b(X_1, \dots, X_n)]$$

*è un **intervallo di fiducia** per θ al livello $(1-\alpha)$ se per ogni valore ammesso del parametro $\theta \in \Theta$ vale*

$$\begin{aligned} 1 - \alpha &\leq \mathbb{P}_\theta(\theta \in I) \\ &= \mathbb{P}_\theta(a(X_1, \dots, X_n) \leq \theta \leq b(X_1, \dots, X_n)). \end{aligned}$$

Tipicamente α è un numero piccolo, ad esempio 0.05 o 0.02. La scelta dell'intervallo di fiducia è oggetto di un compromesso:

- da un lato vogliamo intervalli di fiducia più piccoli possibile per identificare θ con precisione, ma al ridursi dell'intervallo deve diminuire il livello di fiducia $1-\alpha$;
- dall'altro vogliamo un livello di fiducia alto, ovvero vogliamo sia molto probabile che il parametro θ stia nell'intervallo (altrimenti la nostra predizione è fuorviante), ma questo richiede intervalli grandi oppure estremi a, b molto precisi.

Nel seguito analizziamo esempi particolari in cui il campione ha leggi di probabilità dei tipi più comuni esposti nei capitoli precedenti, con particolare attenzione ai campioni Gaussiani.

A. Intervalli di fiducia per la media di un campione Gaussiano

Sia $X_1, \dots, X_n$ un campione statistico con legge Gaussiana $N(m, \sigma^2)$. Tale legge dipende da *due* parametri, corrispondenti a media e varianza: in questo paragrafo cerchiamo intervalli di fiducia per il parametro $\theta = m \in \mathbb{R}$.

Abbiamo visto che $\bar{X}_n$ è uno stimatore corretto e consistente del parametro m, per cui è naturale cercare un intervallo di fiducia della forma

$$I = [\bar{X}_n - d, \bar{X}_n + d] = [\bar{X}_n \pm d]$$

(l'ultima espressione è una abbreviazione definita da quella precedente), in cui $d > 0$ è un numero da determinare, in modo che il livello di fiducia sia alto ma senza che d sia troppo grande.

Definizione VIII.2. *Se $[\bar{X}_n \pm d]$, con $d > 0$ una v. a. possibilmente costante, è un intervallo di fiducia per la media m, d è detta **precisione della stima** e $d/\bar{X}_n$ è detta **precisione relativa della stima**.*

Proposizione VIII.3 (Intervallo di fiducia per la media, varianza nota)**.** *Dato $\alpha \in (0,1)$ e assumendo sia nota $\sigma > 0$, l'intervallo aleatorio*

$$\left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]$$

è un intervallo di fiducia per m con livello di fiducia $1-\alpha$.

Dimostrazione. Dobbiamo imporre la condizione

$$\begin{aligned} 1 - \alpha &\leq \mathbb{P}_m(m \in I) = \mathbb{P}_\theta\left(|\bar{X}_n - m| \leq d\right) \\ &= \mathbb{P}_\theta\left(\frac{\sqrt{n}}{\sigma}\left|\overline{X} - m\right| \leq \frac{d\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right) - 1, \end{aligned}$$

in cui l'ultimo passaggio segue ricordando che la variabile $\sqrt{n}(\bar{X}_n - m)/\sigma$ è Gaussiana standard. Stiamo cercando il valore di d più piccolo possibile, per cui scegliamo quello per cui è verificata l'uguaglianza, il che corrisponde a scegliere esattamente il quantile $\frac{d\sqrt{n}}{\sigma} = q_{1-\alpha/2}$. □

Osserviamo che la precisione d della stima (cioè la semi-ampiezza dell'intervallo):

- cresce al crescere del livello $1-\alpha$;
- cresce al crescere di σ^2;
- decresce al crescere di n.

In particolare, dalla formula per d è possibile determinare la numerosità del campione n per cui l'intervallo abbia una data precisione e un dato livello di fiducia.

Esempio VIII.4. Il peso (in kg) di una specie di salmoni segue una distribuzione Gaussiana di media m non nota e deviazione standard $\sigma = 0.4$. Su un campione di 10 salmoni, la media aritmetica dei pesi risulta essere 3.58 kg.

Seguendo la procedura sopra descritta, determiniamo un intervallo di fiducia per la media al livello $95\% = 1-\alpha$:

$$\left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{0.975}\right] = \left[\bar{X}_{10} \pm \frac{0.4}{\sqrt{10}} \cdot 1.96\right]$$

Inserendo il dato $\bar{X}_n = 3.58$, otteniamo l'intervallo numerico $[3.332, 3.828]$.

Mantenendo il livello di fiducia al 95%, quanto grande deve essere n affinché la precisione della stima sia 0.1? Per trovare il minimo n che soddisfi $d \leq 0.1$, $1-\alpha = 0.95$, imponiamo

$$d = \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2} \leq 0.1$$

da cui $n \geq \left(\frac{\sigma^2 q_{0.975}^2}{0.1}\right)^2 \approx 61.46$, quindi il minimo n è 62.

Osservazione VIII.5. La corretta interpretazione della nozione di intervallo di fiducia è una questione sottile. Come osservato già da Jerzy Neyman nell'articolo originale in cui li introdusse nel 1937, il livello dell'intervallo quantifica l'affidabilità con cui la procedura identifica un

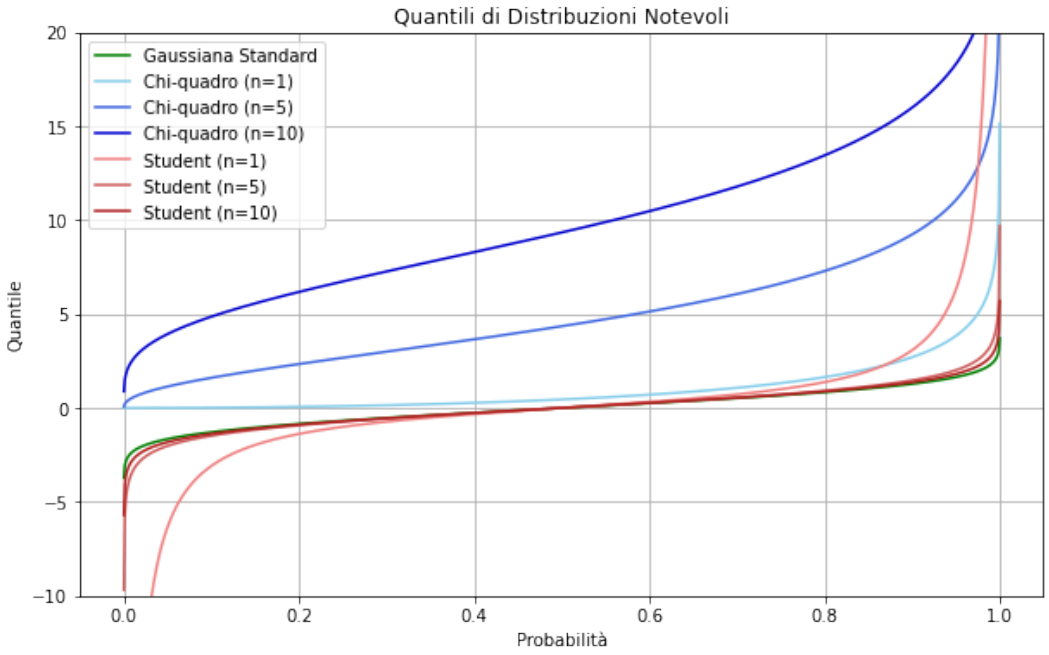


Figura 9: Quantili di variabili Student, Chi-quadro e Gaussiana Standard utilizzati negli intervalli di fiducia per parametri di campioni Gaussiani.

intervallo di possibili valori per il parametro ignoto, ma *non è la probabilità che un intervallo numerico ottenuto a partire dai dati contenga detto parametro.*

Vediamo ora il caso più realistico in cui la varianza σ^2 è sconosciuta. Non possiamo in questo caso usare il numero σ per definire gli estremi dell'intervallo, ma questo problema si aggira sostituendo a σ^2 lo stimatore

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

In questo caso, la variabile $\sqrt{n}(\bar{X}_n - m)/S_n^2$ è di Student $T(n-1)$, per cui per determinare il miglior parametro d dobbiamo considerare il β-quantile $\tau_{\beta,n-1}$ della variabile $T(n-1)$. Con passaggi analoghi a quelli visti sopra si ottiene:

Proposizione VIII.6 (Intervallo di fiducia per la media, varianza non nota)**.** *Dato $\alpha \in (0,1)$, l'intervallo aleatorio*

$$\left[\bar{X}_n \pm \frac{S_n}{\sqrt{n}}\tau_{1-\frac{\alpha}{2},n-1}\right]$$

è un intervallo di fiducia per la media m del campione $X_1,\dots,X_n$ con livello di fiducia $1-\alpha$.

Quando n è grande ($n \geq 60$) si può approssimare il quantile della variabile di Student con quello della Gaussiana standard. Student è il *nom de plume* scelto dallo statistico William Sealy Gosset, che ha introdotto l'idea sopra esposta per determinare intervalli di fiducia.

Esempio VIII.7. Nell'esempio precedente, supponiamo ora che anche σ sia non nota e che la deviazione standard sul campione sia $S_n = 0.4$.

In questo caso, l'intervallo di fiducia di livello $1-\alpha = 0.95$ è dato da

$$\left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}}t_{0.975,9}\right]$$

$$= \left[\bar{X}_{10} \pm \frac{S_n}{\sqrt{10}}\cdot 2.262\right] = \left[\bar{X}_{10} \pm 0.715\cdot S_n\right].$$

Inserendo i dati $\bar{X}_n = 3.58$, $S = 0.4$ otteniamo la precisione $d = 0.715\cdot 0.4 = 0.286$, e l'intervallo numerico $[3.294, 3.866]$. Notiamo che anche se il valore numerico assunto da S_n è lo stesso di σ, la precisione è minore: questo perché la distribution T di Student ha code più pesanti rispetto alla Gaussiana standard, cioè $t_{\beta,n} > q_\beta$ per ogni $\beta > 1/2$.
Per quanto riguarda il minimo n che soddisfi $d \leq 0.1$, $1-\alpha = 0.05$, purtroppo in questo caso non è possibile dare una risposta a priori, poiché d dipende da S_n, che dipende dai dati e in particolare dalla loro taglia. Tuttavia si può dare una stima ragionevole su S_n, ad esempio non ci attendiamo che essa sia superiore a $2\cdot 0.4$, quindi calcolare n per tale stima, infine, dopo aver raccolto i dati su un campione di taglia n, verificare che il valore di S_n fornito dai dati rispetti la stima che avevamo imposto, ad esempio $S_n \leq 2\cdot 0.4$.

Gli intervalli di fiducia appena descritti sono detti *bilateri*, perché entrambi gli estremi sono variabili aleatorie. A volte può essere interessante considerare un intervallo *unilatero*: ad esempio se ci chiediamo soltanto se la media non è troppo alta si considera un intervallo *unilatero sinistro* della forma $I = (-\infty, \bar{X}_n + d]$. Con passaggi simili a quelli visti sopra si ottiene:

Proposizione VIII.8 (Intervalli di fiducia unilateri per la media, varianza nota)**.** *Dato $\alpha \in (0,1)$, gli intervalli aleatori*

$$\left(-\infty, \bar{X}_n + \frac{\sigma}{\sqrt{n}}q_{1-\alpha}\right],$$
$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}}q_{1-\alpha} + \infty\right),$$

sono intervalli di fiducia per la media m del campione $X_1,\dots,X_n$ con livello di fiducia $1-\alpha$.

Gli intervalli di fiducia unilateri con varianza sconosciuta sono perfettamente analoghi, sostituendo a σ la variabile S_n e ai quantili della variabile Gaussiana standard quelli della distribuzione di Student.

B. Intervalli di fiducia per la media di un campione Bernoulli

Consideriamo un campione $X_1, \dots, X_n$ di variabili di Bernoulli di parametro $p \in (0,1)$, corrispondente al valore atteso delle singole variabili. Cerchiamo quindi intervalli di fiducia per il parametro p nella forma $[\bar{X}_n \pm d]$, essendo $\bar{X}_n$ uno stimatore per p.

La determinazione di intervalli di fiducia per la media nel caso Gaussiano si basava sul fatto che la legge della variabile $\sqrt{n}(\bar{X}_n - m)/\sigma$ è Gaussiana standard. Nel caso delle variabili di Bernoulli sappiamo che $X_1 + \cdots X_n$ è di tipo binomiale $B(n,p)$, e dunque i relativi quantili, che appariranno nella determinazione dell'intervallo di fiducia, sono complicati e dipendono anche da n.

Quando n è grande possiamo fare affidamento sul Teorema Centrale del Limite, che afferma che la variabile

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}} = \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$$

è approssimativamente Gaussiana standard. Questo ancora non ci permette di concludere, perché la varianza $\sigma^2 = p(1-p)$ non è nota. Tuttavia σ^2 è funzione del parametro incognito p e quindi è ragionevole stimarla con $\bar{X}_n(1-\bar{X}_n)$. Precisamente, come conseguenza del Teorema Centrale del Limite, si dimostra che

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{n\bar{X}_n(1-\bar{X}_n)}} = \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}$$

converge in distribuzione a una Gaussiana standard, per $n \to \infty$. Quindi, con passaggi analoghi a quelli del paragrafo precedente si ottiene:

Proposizione VIII.9 (Intervallo di fiducia per la media)**.** *Dato $\alpha \in (0,1)$, l'intervallo aleatorio*

$$\left[\bar{X}_n \pm \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} q_{1-\alpha/2}\right]$$

è un intervallo di fiducia per la media p del campione $X_1, \dots, X_n$ con livello di fiducia ***approssimativamente*** *$1-\alpha$. Precisamente, si ha*

$$\lim_{n\to\infty} \mathbb{P}\left(p \in \left[\bar{X}_n \pm \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} q_{1-\alpha/2}\right]\right) = 1-\alpha.$$

Il termine "approssimativamente" significa che ad α fissato non è detto che il livello di fiducia sia $1-\alpha$, ma per n grande abbastanza l'errore è trascurabile. Anche la precisione della stima,

$$d = \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} q_{1-\alpha/2},$$

è di conseguenza approssimata.

Esempio VIII.10. Si vuole condurre un sondaggio telefonico per determinare la percentuale di gradimento del governo. Qual è il numero minimo di telefonate che bisogna effettuare per avere una precisione della stima inferiore a 1% con fiducia al 95 % ?

Assumiamo che la risposta al sondaggio possa essere solo SÌ oppure NO, dunque stiamo considerando un campione aleatorio di variabili Bernoulli di cui vogliamo stimare il parametro p. Dalla definizione di livello di fiducia abbiamo $0.95 = 1-\alpha$, dunque il quantile Gaussiano che ci interessa è $q_{1-\alpha/2} = q_{0.975} \approx 1.96$. La condizione sulla precisione (approssimata) risulta essere

$$\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} q_{0.975} \le 0.01,$$

ovvero

$$\sqrt{n} \ge \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)} \times 1.96}{0.01}.$$

Non conosciamo a priori le risposte alle telefonate: vogliamo che n sia il più piccolo possibile tale che la condizione sia verificata per qualsiasi esito possibile. Scegliamo dunque l'esito che massimizza il membro di destra nell'ultima disuguaglianza: siccome $\max_{0<x<1} x(1-x) = \frac{1}{4}$, ovvero il caso peggiore è quello in cui $\bar{X}_n = 0.5$, otteniamo la condizione

$$n \ge \left(\frac{0.5 \cdot 1.96}{0.01}\right)^2 = 9604.$$

Osserviamo che per n così grande, l'approssimazione del livello di fiducia dell'intervallo è molto buona.

C. Intervalli di fiducia per la media di un campione di taglia grande

La procedura sopra esposta per individuare intervalli di fiducia approssimati per la media usando il Teorema Centrale del Limite si può applicare esattamente nello stesso modo a campioni di v. a. di taglia grande (e di momento secondo finito):

Proposizione VIII.11 (Intervallo di fiducia per la media, campioni grandi, varianza nota)**.** *Sia $X_1, \dots X_n$ un campione i.i.d., la cui legge ha momento secondo finito, con n grande. Dato $\alpha \in (0,1)$, l'intervallo aleatorio*

$$\left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]$$

è un intervallo di fiducia per la media m del campione con livello di fiducia ***approssimativamente*** *$1-\alpha$. Precisamente, si ha*

$$\lim_{n\to\infty} \mathbb{P}\left(m \in \left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]\right) = 1-\alpha.$$

Proposizione VIII.12 (Intervallo di fiducia per la media, campioni grandi, varianza non nota)**.** *Sia $X_1, \dots X_n$ un campione i.i.d., la cui legge ha momento secondo finito, con n grande. Dato $\alpha \in (0,1)$, l'intervallo aleatorio*

$$\left[\bar{X}_n \pm \frac{S_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]$$

è un intervallo di fiducia per la media m del campio-

ne con livello di fiducia ***approssimativamente*** $1-\alpha$. *Precisamente, si ha*

$$\lim_{n\to\infty} \mathbb{P}\left(m \in \left[\bar{X}_n \pm \frac{S_n}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right]\right) = 1-\alpha.$$

D. Intervalli di fiducia per la varianza di un campione Gaussiano

Se vogliamo trovare intervalli di fiducia per la *varianza* di un campione Gaussiano $N(m,\sigma^2)$, non c'è sostanziale differenza tra i casi in cui m è nota oppure no: consideriamo direttamente il secondo.

Il calcolo di intervalli di fiducia bilateri in questo caso non è agevole, ci concentriamo su intervalli unilateri. Del resto, questi ultimi sono sono di maggiore interesse, essendo la varianza una quantità positiva di cui ci interessa capire se supera o meno una certa soglia. Ad esempio, in una misurazione di prodotti industriali standardizzati, siamo interessati a controllare che la varianza delle misure nel controllo qualità sia piccola, ovvero inferiore a una soglia fissata, ad indicare che la produzione è di buona qualità perché tutti i pezzi hanno misura simile.

Proposizione VIII.13 (Intervalli di fiducia per la varianza)**.** *Dato $\alpha \in (0,1)$, gli intervalli aleatori*

$$\left(0, \frac{(n-1)S_n^2}{\chi^2_{\alpha,n-1}}\right], \quad \left[\frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}, +\infty\right),$$

sono intervalli di fiducia per σ^2 (rispettivamente unilateri sinistri e destri) con livello di fiducia $1-\alpha$.

Ricordiamo che con $\chi^2_{\beta,n-1}$ indichiamo il β-quantile di una variabile di tipo $\chi^2(n-1)$. Se $X_1,\dots,X_n$ è un campione di variabili $N(m,\sigma^2)$, ricordiamo che la variabile

$$\sum_{i=1}^{n} \frac{\left(X_i - \bar{X}_n\right)^2}{\sigma^2} = (n-1)\frac{S_n^2}{\sigma^2}$$

è di tipo $\chi^2(n-1)$. La dimostrazione è ora una facile verifica della definizione.

E. Esercizi

Esercizio VIII.14. Un certo metodo per misurare il pH di una soluzione fornisce un risultato distribuito come una Gaussiana, di media il valore autentico del pH della soluzione e deviazione standard 0.1. Vengono effettuate 50 misurazioni di una soluzione; la media degli esiti di tali misurazioni risulta 8.19.

- Fornire un intervallo di fiducia di livello 0.95 per il valore autentico del pH della soluzione.
- Quante misurazioni è necessario eseguire affinché, con livello di fiducia del 95%, la precisione dell'intervallo sia 0.01?

Esercizio VIII.15. Un server che elabora un certo tipo di richiesta può commettere un errore. Su 1000 richieste, sono stati osservati 50 errori.

- Fornire un intervallo di fiducia di livello 0.99 per la probabilità di errore su una singola richiesta.
- Quante richieste dobbiamo analizzare, per ridurre la precisione a 0.003 (mantenendo lo stesso livello di fiducia)?

Esercizio VIII.16. Si misurano i livelli di emoglobina di 80 persone adulte di sesso femminile, ottenendo una media di 14.1 g/dl e una deviazione standard di 0.7 g/dl. Fornire un intervallo di fiducia del 99% per il livello medio di emoglobina (in una persona adulta di sesso femminile).

Esercizio VIII.17. Su un campione di 80 individui di una certa popolazione, 48 presentano un certo gene. (i) Trovare un intervallo di fiducia di livello 95% per la frequenza del gene nella popolazione. (ii) Quanto grande deve essere la taglia del campione, affinché la precisione della stima sia inferiore al 5%, mantenendo lo stesso livello 95%?

Esercizio VIII.18. Per ogni anno dal 2016 al 2021 viene misurata la temperatura media (in gradi centigradi) a Milano nel mese di settembre; la media e la deviazione standard di tali dati risultano rispettivamente 20.83 e 1.35. Supponendo che la temperatura media di settembre sia distribuita come una Gaussiana, fornire un intervallo di fiducia a livello 95% per la media delle temperature medie di settembre a Milano.

Esercizio VIII.19. Si consideri un campione $X_1,\dots,X_n$ di variabili con densità esponenziale di parametro $\lambda > 0$. Si mostri che la variabile

$$\frac{\bar{X}_n}{\lambda} = \frac{X_1 + \cdots + X_n}{n\lambda}$$

ha densità $\Gamma(n,1/n)$. Si usi questo fatto per determinare intervalli di fiducia bilateri per λ, con estremi che dipendono dal β-quantile $\gamma_{\beta,n}$ della densità $\Gamma(n,1/n)$.

Esercizio VIII.20. Utilizzando la stessa procedura vista per campioni statistici di Bernoulli, si determini un intervallo di fiducia bilatero per il parametro λ di un campione $X_1,\dots,X_n$ di v. a. di Poisson.

Esercizio VIII.21. Siano X_1, X_2 un campione di variabili con densità uniforme sull'intervallo $[\theta-1/2, \theta+1/2]$, con $\theta \in \mathbb{R}$ il parametro da stimare. Posto $\bar{X} = \frac{X_1+X_2}{2}$, si considerino gli intervalli aleatori

$$I_1 = \left[\bar{X} \pm \frac{\min\{|X_1-X_2|, 1-|X_1-X_2|\}}{2}\right],$$
$$I_2 = \left[\bar{X} \pm \frac{1-|X_1-X_2|}{4}\right].$$

Si mostri che:

- sono entrambi intervalli di fiducia per θ con livello di affidabilità 50%;
- l'intervallo I_1 è mediamente più piccolo di I_2, nel senso che se $\theta' \neq \theta$ vale
$$\mathbb{P}_\theta(\theta' \in I_1) \leq \mathbb{P}_\theta(\theta' \in I_2);$$
- se $|X_1 - X_2| \geq 1/2$ l'intervallo I_1 contiene *sempre* il vero parametro θ, ma questo non è vero per I_2.

Si osservi infine che se $|X_1 - X_2|$ è molto piccolo, I_1 è quasi ridotto a un punto, ed è praticamente impossibile che contenga θ, per questo il livello di affidabilità dell'intervallo risulta essere così basso.

Esercizio VIII.22. Si consideri un monitoraggio sulla presenza di sostanze tossiche nell'aria, effettuato in 10 stazioni di monitoraggio vicine. I valori ottenuti restituiscono una concentrazione media di 4.8mg/dm^3 con una varianza campionaria di 0.49mg/dm^3. Supponiamo che la distribuzione della concentrazione sia Gaussiana. Fornire una

stima della concentrazione delle sostanze tossiche con una fiducia del 90% mediante un intervallo bilatero. Con quale fiducia si ottiene una precisione relativa di $5 \cdot 10^{-2}$? Dire se l'ipotesi che la concentrazione non sia superiore a 4.3mg/dm^3 è plausibile.

IX. TEST STATISTICI E Z-TEST

Un test statistico è una procedura per verificare ipotesi su uno o più parametri incogniti della distribuzione di probabilità con cui vogliamo descrivere un esperimento ripetuto di cui conosciamo gli esiti.

A. Verifica di ipotesi

Una ipotesi statistica è una affermazione sul parametro θ che governa la legge di un campione statistico in considerazione:

Definizione IX.1 (Ipotesi Nulla e Alternativa)**.** *Sia $X_1, \dots, X_n$ un campione statistico la cui legge (ovvero la funzione di ripartizione di ogni singola variabile $F_\theta = F_{X_i}$) dipende da un parametro $\theta \in \Theta \subseteq \mathbb{R}$.*

*Data una partizione $\Theta = \Theta_0 \cup \Theta_1$ in due sottoinsiemi disgiunti, l'*ipotesi nulla *H_0 è l'affermazione logica $\theta \in \Theta_0$, e Θ_0 è detto l'insieme dei parametri dell'ipotesi nulla. L'*ipotesi alternativa *H_1 è l'affermazione complementare $\theta \in \Theta_1$, e Θ_1 è l'insieme dei parametri dell'ipotesi alternativa.*

A questo livello non vi è differenza tra i ruoli delle ipotesi H_0 e H_1, ma vi è una distizione fondamentale contenuta invece nella definizione di test statistico, che introduciamo dapprima in termini euristici. Un **test statistico** è una procedura per decidere se accettare o rifiutare l'ipotesi nulla H_0 a partire dai valori assunti dal campione $X_1, \dots X_n$:

- si accetta l'ipotesi nulla H_0 se i valori assunti dal campione sono con essa compatibili (in un senso al momento da rendere preciso): in questo caso si parla di **esito negativo** del test;
- si rifiuta H_0 in favore di H_1 se, con alto grado di fiducia, i valori non sono compatibili con H_0 (e sono in favore di H_1), cioè se c'è evidenza statistica contro H_0 (e per H_1): in questo caso si parla di **esito positivo** del test.

Notiamo qui l'asimmetria tra le ipotesi: rifiutare H_0 vuol dire che i dati forniscono una evidenza contro l'ipotesi nulla H_0; accettare H_0 invece non implica un'evidenza per H_0, ma ci dice solo che i dati non forniscono evidenza contro H_0.

Esempio IX.2. Per verificare se è in corso una infezione di un virus si preleva un campione di materiale biologico, lo si sottopone a una procedura che trasforma il RNA in DNA (ed amplifica il numero delle molecole) e si introduce una sostanza che rende fluorescenti le parti di DNA virali. Si contano poi queste parti e si determina che l'infezione è in corso (ovvero il test è positivo) se tale numero supera una certa soglia, che a titolo esemplificativo diciamo essere l'1%.

Il modello probabilistico è un campione di variabili di Bernoulli di parametro $\theta \in \Theta = (0, 1)$, a rappresentare il fatto che singole sezioni di DNA siano fluorescenti (quindi virale) o meno. L'ipotesi nulla "la percentuale di sezioni fluorescenti di DNA non supera il 1%" (corrispondente all'esito negativo del test) è rappresentata dalla scelta $\Theta_0 = (0, 0.01]$ e $\Theta_1 = (0.01, 1)$.

Esempio IX.3. In un controllo di qualità si desidera valutare la percentuale di pezzi difettosi in una produzione. In modo del tutto analogo all'esempio precedente, il modello probabilistico è un campione di variabili di Bernoulli di parametro $\theta \in \Theta = (0, 1)$, e l'ipotesi nulla "la percentuale di pezzi difettosi non supera il 2%" corrisponde a considerare $\Theta_0 = (0, 0.02]$ e $\Theta_1 = (0.02, 1)$.

Consideriamo invece la situazione di due fabbriche concorrenti, con l'obbiettivo di valutare se una produca più pezzi difettosi dell'altra o viceversa. Possiamo a tal fine considerare come modello due campioni indipendenti $X_1, \dots, X_n$ e $Y_1, \dots, Y_n$ di variabili di Bernoulli di parametri rispettivamente $\theta_1, \theta_2 \in (0, 1)$. Essendo però interessati al confronto, possiamo in effetti ridurci a considerare il singolo parametro $\theta = \theta_1 - \theta_2 \in (-1, 1) = \Theta$. L'ipotesi nulla "non c'è differenza tra le due produzioni" corrisponde quindi a considerare $\Theta_0 = \{0\}$.

Osservazione IX.4. La scelta del termine "ipotesi nulla" (*null hypothesis*) si riferisce al fatto che solitamente H_0 **è l'ipotesi sotto cui la variabilità degli esiti dei dati è spiegata dalla sola casualità dell'estrazione** e non invece da una cattiva scelta dei parametri del modello. Nello specifico caso in cui il parametro θ misuri la differenza tra due situazioni, come nell'ultimo esempio proposto, H_0 corrisponde all'ipotesi in cui tale differenza è appunto "nulla", e le piccole differenze nei dati sono dovute all'aleatorietà inerente alla misurazione.

Fissata l'ipotesi, dobbiamo quindi determinare un insieme di risultati che portano a rifiutarla: in altre parole, all'interno del modello probabilistico, dobbiamo scegliere un evento che se si realizza conduce a rifiutare H_0. Questo corrisponde a scegliere un sottoinsieme dello spazio di probabilità: vogliamo un insieme di elementi $\omega \in \Omega$ tale che l'osservazione di esiti $X_1(\omega), \dots, X_n(\omega)$ conduca a rifiutare H_0.

Definizione IX.5 (Regione Critica e Test)**.** *La* regione critica *o* regione di rifiuto *per l'ipotesi nulla H_0 è un evento $C \subset \Omega$; il complementare $A = \Omega \setminus C$ è detto invece* regione di accettazione.

Data una realizzazione $(x_1, \dots x_n)$ del campione $(X_1, \dots X_n)$, l'ipotesi nulla viene:

- respinta *se il vettore di v. a. $X_1(\omega), \dots, X_n(\omega)$ assume valore $(x_1, \dots x_n)$ per un qualche $\omega \in C$, ovvero* se si verifica l'evento *C;*
- accettata *se invece all'interno della regione critica $\omega \in C$ il vettore $X_1(\omega), \dots, X_n(\omega)$ non coincide mai con i dati $(x_1, \dots x_n)$, ovvero* se non si verifica l'evento *C.*

Osservazione IX.6. Si potrebbe pensare che sia più intuitivo e comodo definire la regione critica come un sottoinsieme dello spazio dei valori del campione $X_1, \dots, X_n$, ovvero di $\mathbb{R}^n$: date delle osservazioni $(x_1, \dots, x_n) \in \mathbb{R}^n$ vorremmo decidere se sono coerenti o meno con l'ipotesi nulla. Nella pratica, tuttavia, la regione critica è praticamente sempre definita come un evento associato a una statistica campionaria, come ad esempio $\{\bar{X}_n \in I\}$ per qualche regione $I \subset \mathbb{R}$. In effetti, trattando dati numerosi non è ragionevole nè conveniente considerare valori di singole misure e si opta invece per indici statistici: la definizione matematica rispecchia questa esigenza.

Osservazione IX.7. Come già osservato nel caso degli intervalli di fiducia, le conclusioni delle procedure della statistica parametrica quali i test statistici *non devono*

essere considerate come verità concernenti il fenomeno osservato. Questa sarebbe infatti una interpretazione scorretta del metodo, che invece restituisce (internamente al modello matematico) una certa *approssimazione* degli esiti che sarà buona con un opportuno *grado di fiducia*.

Il risultato del test (accettare o rifiutare H_0) è soggetto a due tipi di errore:

- un *errore di prima specie* consiste nel rifiutare l'ipotesi H_0 quando questa è soddisfatta (*falso positivo*): se $\theta \in \Theta_0$, la probabilità di commettere un errore di prima specie è dunque $\mathbb{P}_\theta(C)$;
- un *errore di seconda specie* consiste nell'accettare l'ipotesi H_0 quando questa non è soddisfatta (*falso negativo*): se $\theta \in \Theta_1$ la probabilità di commettere un errore di seconda specie è $\mathbb{P}_\theta(A)$.

Esempio IX.8. Tornando all'esempio del controllo di qualità in una singola produzione, si rifiuterà l'ipotesi se la percentuale di pezzi difettosi riscontrata è superiore a una certa soglia d. Scriveremo sinteticamente $C = \{\omega \in \Omega \big| \bar{X}_n(\omega) > d\} = \{\bar{X}_n > d\}$. Minimizzando l'errore di prima specie, la scelta di d dovrà essere tale che la probabilità $\mathbb{P}_\theta(C) = \mathbb{P}_\theta\{\bar{X} > d\}$ sia piccola per ogni $\theta \leq 0.02$.

Definizione IX.9 (Livello e potenza di un test)**.** *Dato $0 < \alpha < 1$, si dice che il test è di livello α se*

$$\forall \theta \in \Theta_0, \quad \mathbb{P}_\theta(C) \leq \alpha.$$

*Si chiama potenza del test (*statistical power*) la funzione, definita sui parametri dell'alternativa,*

$$\Theta_1 \ni \theta \mapsto \mathbb{P}_\theta(C) \in [0, 1].$$

Fissare un livello significa pertanto fissare *un limite superiore per la probabilità dell'errore di prima specie*, scegliendo opportunamente la regione critica. I valori tipici del livello sono piccoli, ad esempio 0.05 o 0.02. Il livello del test è legato al grado di fiducia che chiediamo per rifiutare: più piccolo è il livello α, maggiore è l'evidenza che chiediamo per rifiutare H_0.

La potenza è la probabilità $\mathbb{P}_\theta$ di rifiutare correttamente H_0 quando questa è falsa, quindi rappresenta la *capacità di accorgersi che l'ipotesi H_0 non è soddisfatta*.

Osservazione IX.10. È desiderabile avere test con livello basso (bassa probabilità di errore di prima specie) e potenza alta (bassa probabilità di errore di seconda specie). In genere, tuttavia, più basso è il livello, più bassa è la potenza, per cui le due richieste sono in contrapposizione. Questo è intuitivo: più evidenza chiediamo per rifiutare, più aumenta il rischio di accettare H_0 quando questa non è soddisfatta. Per avere sia livello basso sia potenza alta, si può agire sulla taglia del campione: come vedremo, aumentare la taglia del campione diminuisce il livello e aumenta la potenza.

Il metodo classico di impostazione di un test prevede di fissare un livello α a priori e di scegliere la regione critica C in modo da soddisfare $\mathbb{P}_\theta(C) \leq \alpha$ per ogni $\theta \in \Theta_0$ (cosicché il livello del test sia appunto α). La scelta del livello α influisce sulla scelta della regione critica C e quindi sul risultato del test: più piccolo è α (cioè maggiore è l'evidenza che chiediamo per rifiutare), più piccola sarà la regione critica C. In particolare, se rifiutiamo H_0 ad un livello α, allora rifiutiamo H_0 a ogni livello $\alpha' \geq \alpha$, mentre potremmo accettare H_0 per un qualche livello minore di α.

Per ovviare alla dipendenza del risultato del test da α, nelle applicazioni si fa spesso uso di un'altra formulazione, centrata sulla nozione di *p-value*. A differenza della regione critica, che è fissata a priori indipendentemente dai valori del campione, il p-value dipende dalla realizzazione $(x_1, \dots x_n)$ del campione $(X_1, \dots X_n)$. Informalmente, si dice che

> *il p-value dei dati $(x_1, \dots x_n)$ è la probabilità, calcolata accettando l'ipotesi nulla H_0, di ottenere "dati più estremi", rispetto all'ipotesi H_0, di quelli osservati.*

Naturalmente questa non è una definizione matematica: anzitutto non è chiaro il significato di "estremo", ed inoltre in generale l'ipotesi nulla è data da un insieme di valori Θ_1, ma per calcolare una probabilità dobbiamo scegliere un singolo $\theta \in \Theta$. In effetti, la scarsa conoscenza dei fondamenti della statistica porta spesso nelle applicazioni ad applicare male il metodo del p-value, basandosi solo sulla nozione intuitiva e portando a invalidare i risultati dell'analisi statistica.

Esempio IX.11. Torniamo al controllo di qualità con ipotesi nulla "la percentuale di pezzi difettosi non supera il 2%". Se su 1000 pezzi ne risultano 30 difettosi, il p-value corrisponde alla probabilità di trovare *almeno* 30 pezzi difettosi su 1000, nell'ipotesi H_0 che θ sia 2%. Si osservi che l'ipotesi nulla è data da un intervallo, $\Theta_0 = (0, 0.02)$, ma per calcolare il p-value noi scegliamo l'estremo $\theta = 0.02$.

Sebbene la nozione informale di p-value sopra descritta possa essere un utile strumento mnemonico, dobbiamo ora dare una definizione matematicamente rigorosa.

Definizione IX.12 (p-value)**.** *Si assuma data una famiglia di regioni critiche $\{C(\alpha)\}_{\alpha \in (0,1)}$ tali che il test (per il parametro θ con ipotesi nulla H_0: $\theta \in \Theta_0$) con regione critica $C(\alpha)$ abbia livello α.*

Data una realizzazione $(x_1, \dots x_n)$ del campione $(X_1, \dots X_n)$, il p-value *è il numero $\bar{\alpha} = \bar{\alpha}(x_1, \dots x_n)$ tale che:*

- *se $\alpha < \bar{\alpha}$, l'ipotesi nulla viene accettata dal test di regione critica $C(\alpha)$ (diremo che* viene accettata a livello α*),*
- *se invece $\alpha > \overline{\alpha}$, l'ipotesi viene rifiutata dal test di regione critica $C(\alpha)$.*

Il calcolo del *p-value* è molto importante sia perché sintetizza in un solo numero la plausibilità di una ipotesi, sia perché, come esito quantitativo tra 0 e 1, fornisce un'informazione più ricca della semplice dicotomia accettazione o rifiuto, peraltro indipendente dal livello α. In pratica, se il p-value è molto basso, ad esempio inferiore a 0.001, l'ipotesi H_0 è decisamente poco plausibile, se il p-value è alto, ad esempio superiore a 0.3, l'ipotesi H_0 è molto plausibile, mentre per valori intermedi il p-value ci dice quanto forte è l'indicazione contro H_0 (più il p-value è basso, più forte è l'indicazione contro H_0).

Proprio perché si rinuncia a fissare il livello α, non è possibile fissare un valore standard del p-value per cui si rifiuta l'ipotesi nulla. La scelta di α o equivalentemente della soglia del p-value sotto cui rifiutare l'ipotesi nulla dipende dalla situazione. Ad esempio, se modificare la produzione per avere meno pezzi difettosi comporta un grande costo per l'azienda (non supportato dai maggiori

ricavi), sceglieremo un livello α molto basso, per cui rifiuteremo l'ipotesi H_0 (e conseguentemente modificheremo la produzione) solo in presenza di una forte evidenza di più del 2% di pezzi difettosi.

B. Z-test

Si chiama **Z-test** il test sulla media di un campione Gaussiano con varianza nota. Iniziamo discutendo il caso particolare in cui l'ipotesi nulla è data da un singolo valore per la media.

Consideriamo un campione $X_1, \dots, X_n$ di legge $N(m, \sigma^2)$, e supponiamo di conoscere il valore di $\sigma > 0$. Vogliamo effettuare un test per decidere se la media m coincide o meno con un certo valore m_0, ovvero scegliamo $m = \theta \in \Theta = \mathbb{R}$, e $\Theta_0 = \{m_0\}$, $\Theta_1 = \mathbb{R} \setminus \{m_0\}$,

$$H_0)\, m = m_0 \qquad \text{contro} \qquad H_1)\, m \neq m_0$$

Come per la determinazione dell'intervallo di fiducia, in presenza di un campione Gaussiano con varianza nota i calcoli sono basati sul fatto che la variabile $Z = \sqrt{n}(\bar{X}_n - m)/\sigma$ è Gaussiana standard.

Poiché $\bar{X}_n$ è uno stimatore corretto e consistente di m, l'intuizione ci porta a rifiutare l'ipotesi se $\bar{X}_n$ si scosta troppo da m_0, cioè a scegliere una regione critica della forma

$$C = \{|\bar{X}_n - m_0| > d\},$$

dove il numero $d > 0$ deve essere determinato in funzione del livello α scelto. Si deve cioè avere $\mathbb{P}_{m_0}\{|\bar{X}_n - m_0| > d\} \leq \alpha$. Per ottenere una regione critica più grande possibile (allo scopo di aumentare la potenza del test) richiederemo che valga

$$\begin{aligned}\alpha &= \mathbb{P}_{m_0}\left(\left|\bar{X}_n - m_0\right| > d\right) \\ &= \mathbb{P}_{m_0}\left(\frac{\sqrt{n}}{\sigma}\left|\bar{X}_n - m_0\right| > \frac{d\sqrt{n}}{\sigma}\right) \\ &= \mathbb{P}\left(|Z| > \frac{d\sqrt{n}}{\sigma}\right),\end{aligned}$$

e questo ci porta a definire la regione critica di livello α scegliendo $\frac{d\sqrt{n}}{\sigma} = q_{1-\frac{\alpha}{2}}$,

$$\begin{aligned}C &= \left\{\sqrt{n}\frac{\left|\bar{X}_n - m_0\right|}{\sigma} > q_{1-\frac{\alpha}{2}}\right\} \\ &= \left\{|\bar{X}_n - m_0| > \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right\}.\end{aligned}$$

Dati esiti $x_1, \dots, x_n$ dell'esperimento, si calcola la media empirica $\bar{x}_n$ e si rifiuta l'ipotesi H_0 se $|\bar{X}_n - m_0| > \sigma q_{1-\alpha/2}/\sqrt{n}$, si accetta H_0 in caso contrario.

Osservazione IX.13. **L'ipotesi $H_0) m = m_0$ è accettata al livello α se e solo se m_0 appartiene all'intervallo di fiducia per la media con livello di fiducia** $(1-\alpha)$. Questa è una proprietà generale: si può dimostrare che è equivalente identificare un intervallo di fiducia al livello $(1-\alpha)$ oppure un test nel quale l'ipotesi è semplice, cioè ridotta a un solo valore del parametro.

Come discusso nel caso degli intervalli di fiducia, l'ampiezza della regione di accettazione C^c, cioè $2\sigma q_{1-\alpha/2}/\sqrt{n}$:

- cresce al crescere del livello del test $1-\alpha$;
- cresce al crescere di σ^2;
- decresce al crescere di n.

Calcoliamo ora il p-value dei dati $(x_1, \dots x_n)$. Procediamo prima usando la definizione informale di p-value come probabilità, sotto H_0, di ottenere "dati più estremi", rispetto all'ipotesi H_0, di $(x_1, \dots x_n)$. Poiché $\bar{X}$ è la stima del valore atteso m e l'ipotesi nulla H_0 è $m = m_0$, è intuitivo considerare come "dati più estremi" quelli $(y_1, \dots y_n)$ che verificano $|\bar{y}_n - m_0| > |\bar{x}_n - m_0|$ (dove $\bar{x}_n, \bar{y}_n$ sono le medie campionarie di $(x_1, \dots x_n)$, $(y_1, \dots y_n)$ rispettivamente). Ci aspettiamo quindi che il p-value di $(x_1, \dots x_n)$ sia

$$\begin{aligned}\overline{\alpha} &= \overline{\alpha}(x_1, \dots x_n) \\ &= \mathbb{P}_{m_0}\left(\sqrt{n}\frac{\left|\bar{X}_n - m_0\right|}{\sigma} > \frac{\sqrt{n}}{\sigma}|\bar{X}_n - m_0|\right) \\ &= 2\left[1 - \Phi\left(\frac{\sqrt{n}}{\sigma}|\bar{X}_n - m_0|\right)\right],\end{aligned}$$

in cui l'ultimo passaggio segue ricordando che $Z = \sqrt{n}(\bar{X} - m_0)/\sigma$ è una Gaussiana standard. Verifichiamo ora che il valore $\overline{\alpha}$ trovato sopra soddisfa la definizione rigorosa di p-value.

Dobbiamo verificare che, nel test di livello α, si rifiuta H_0 se e solo se $\alpha > \overline{\alpha}$. Abbiamo visto che H_0 è rifiutata al livello α se e solo se $\frac{\sqrt{n}}{\sigma}|\bar{X}_n - m_0| > q_{1-\frac{\alpha}{2}}$. Quindi, in caso di rifiuto, abbiamo

$$\begin{aligned}\overline{\alpha} &= \mathbb{P}_{m_0}\left(\sqrt{n}\frac{\left|\bar{X}_n - m_0\right|}{\sigma} > \frac{\sqrt{n}}{\sigma}|\bar{X}_n - m_0|\right) \\ &< \mathbb{P}_{m_0}\left(\sqrt{n}\frac{\left|\bar{X}_n - m_0\right|}{\sigma} > q_{1-\frac{\alpha}{2}}\right) = \alpha,\end{aligned}$$

mentre, in caso di accettazione, vale la disuguaglianza opposta. Dunque $\overline{\alpha}$ soddisfa la definizione rigorosa di p-value.

Un esempio svolto

Il peso (in kg) di una popolazione di salmoni segue una distribuzione Gaussiana di media m non nota e deviazione standard 0.5. Viene rilevato il peso su un campione di $n = 16$ esemplari. Si richiede di:

- formulare un test di livello 5% per l'ipotesi H_0 "peso medio pari a 3", e applicarlo nel caso in cui la media sul campione sia 3.2;
- determinare il p-value per la media campionaria 3.2;
- calcolare la probabilità dell'errore di seconda specie (accettare H_0 quando questa è falsa) assumendo $m = 3.4$;
- determinare la minima taglia n del campione per cui il test abbia potenza almeno 95% se $m = 3.4$.

L'ipotesi è $H_0 : m = 3$ contro $H_1 : m \neq 3$, il livello α è 0.05 ($q_{0.975} = 1.96$) e la regione critica C è

$$C = \left\{|\bar{X} - 3| > 0.245\right\}.$$

Se la media campionaria vale $\bar{x} = 3.2$, i dati cadono nella regione di accettazione C^c e quindi l'ipotesi H_0: "peso medio pari a 3" viene accettata a livello $\alpha = 0.5$.

Il p-value corrispondente a $\bar{x}_n = 3.2$ è

$$\bar{\alpha} = \mathbb{P}\{|Z| > \frac{\sqrt{n}}{\sigma}|\bar{x} - 3|\} = 2\left(1 - \Phi\left(\frac{\sqrt{16}}{0.5}|3.2 - 3|\right)\right) = 0.1096,$$

dunque, per $\bar{x} = 3.2$, rifiutiamo per $\alpha > 0.1096$ e accettiamo per $\alpha < 0.1096$.
La probabilità dell'errore di seconda specie per $m = 3.4$ è

$$\mathbb{P}_{m=3.4}\left(|\bar{X}_n - m_0| < \frac{\sigma}{\sqrt{n}}q_{1-\alpha/2}\right) = \Phi\left(\sqrt{n}\frac{|m_0 - m|}{\sigma} + q_{1-\frac{\alpha}{2}}\right) - \Phi\left(\sqrt{n}\frac{|m_0 - m|}{\sigma} - q_{1-\frac{\alpha}{2}}\right)$$
$$= \Phi\left(\sqrt{16}\frac{|3 - 3.4|}{0.5} + 1.96\right) - \Phi\left(\sqrt{16}\frac{|3 - 3.4|}{0.5} - 1.96\right) = \Phi(5.16) - \Phi(1.24) = 0.1075.$$

In altre parole, se il peso medio è 3.4, c'è una probabilità di quasi l'11% di accettare l'ipotesi "peso medio pari a 3". Perché il test abbia potenza almeno 0.95, imponiamo

$$1 - \Phi\left(\sqrt{n}\frac{|m_0 - m|}{\sigma} - q_{1-\frac{\alpha}{2}}\right) = 1 - \Phi(\sqrt{n} \cdot 0.8 - 1.96) \leq 1 - 0.95 = 0.05,$$

cioè $\sqrt{n} \cdot 0.8 - 1.96 \geq q_{0.95} = 1.64$, cioè $n \geq 20.25$, quindi 21 è la minima taglia del campione richiesta.

C. Z-test unilatero

Descriviamo ora lo Z-test per la media di un campione Gaussiano a varianza nota con ipotesi alternative

$$H_0) m \leq m_0 \qquad \text{contro} \qquad H_1) m > m_0.$$

Lo sviluppo del test è analogo al caso bilatero, e ometteremo i dettagli.

L'intuizione ci spinge a rifiutare l'ipotesi se $(\bar{X}_n - m_0)$ è troppo grande, cioè a considerare una regione critica della forma

$$C = \left((\bar{X}_n - m_0) > d\right),$$

e la condizione sul livello diventa

$$\forall m \leq m_0, \quad \mathbb{P}_m\left((\bar{X}_n - m_0) > d\right) \leq \alpha.$$

La probabilità sopra scritta cresce al crescere di m e si arriva dunque richiedendo potenza maggiore possibile a

$$\mathbb{P}_{m_0}\left((\bar{X}_n - m_0) > d\right) = \mathbb{P}_{m_0}\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m_0) > \frac{d\sqrt{n}}{\sigma}\right) = \alpha,$$

da cui segue $\frac{d\sqrt{n}}{\sigma} = q_{1-\alpha}$, e dunque la *regione critica* al livello α è

$$C = \left\{\sqrt{n}\frac{(\bar{X}_n - m_0)}{\sigma} > q_{1-\alpha}\right\} = \left\{(\bar{X}_n - m_0) > \frac{\sigma}{\sqrt{n}}q_{1-\alpha}\right\}.$$

La dipendenza dell'ampiezza della regione di accettazione C^c dai parametri α, σ, n è la stessa del caso bilatero.

Calcoliamo il p-value dei dati $(x_1, \dots x_n)$. In questo caso, i dati $(y_1, \dots y_n)$ sono "più estremi" di $(x_1, \dots x_n)$ rispetto ad $H_0 : m \leq m_0$ se $\bar{y}_n - m_0 > \bar{x}_n - m_0$. La probabilità, per $m \leq m_0$ di avere "dati più estremi" di $(x_1, \dots x_n)$ è

$$\mathbb{P}_m\left(\sqrt{n}\frac{(\bar{X}_n - m_0)}{\sigma} > \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m_0)\right).$$

Come p-value si prende il massimo di tali probabilità per $m \in H_0 : m \leq m_0$, che si realizza in m_0: il p-value diventa

$$\overline{\alpha}(x_1, \dots x_n) = \mathbb{P}_{m_0}\left(\sqrt{n}\frac{(\bar{X}_n - m_0)}{\sigma} > \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m_0)\right) = 1 - \Phi\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m_0)\right)$$

e si può verificare che tale valore verifica la definizione rigorosa di p-value.

Se l'ipotesi è della forma $H_0) m \geq m_0$, si invertono tutte le diseguaglianze e si sostituisce $q_{1-\alpha}$ con q_α: la regione critica è

$$C = \left\{\sqrt{n}\frac{(\bar{X}_n - m_0)}{\sigma} < q_\alpha\right\},$$

ed il p-value è

$$\overline{\alpha}(x_1, \dots x_n) = \mathbb{P}_{m_0}\left(\sqrt{n}\frac{(\bar{X}_n - m_0)}{\sigma} < \frac{\sqrt{n}}{\sigma}(\bar{X}_n - m_0)\right) = \Phi\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - m_0)\right).$$

Esercizio IX.14. Ripetere l'esercizio svolto di cui sopra (peso di un campione di salmone), nel caso del test unilatero con H_0 : "peso medio non superiore a 3".

D. Test Z approssimato su un campione di Bernoulli

Sia $X_1, \dots, X_n$ un campione di variabili di Bernoulli di cui vogliamo determinare il parametro $p \in (0,1)$. Come per gli intervalli di fiducia sfruttiamo il fatto che *per n grande* la variabile $\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$ è approssimativamente Gaussiana standard.

Consideriamo ad esempio il test **bilatero** dell'ipotesi

$$H_0) p = p_0 \qquad \text{contro} \qquad H_1) p \neq p_0$$

al livello α, assumendo che *n sia grande*; la regione critica approssimata è

$$C = \left\{ \frac{\sqrt{n}|\bar{X}_n - p_0|}{\sqrt{p_0(1-p_0)}} > q_{1-\frac{\alpha}{2}} \right\},$$

e il p-value è

$$\mathbb{P}_{p_0}\left(\frac{\sqrt{n}|\bar{X}_n - p_0|}{\sqrt{p_0(1-p_0)}} > \frac{\sqrt{n}|\bar{x} - p_0|}{\sqrt{p_0(1-p_0)}} \right) \approx 2\left[1 - \Phi\left(\frac{\sqrt{n}|\bar{x} - p_0|}{\sqrt{p_0(1-p_0)}} \right) \right].$$

Anche per quanto riguarda i test *unilateri* si procede in analogia con il caso Gaussiano, lasciamo i dettagli per esercizio.

Esempio IX.15. In un controllo qualità su un certo tipo di pezzi, vengono estratti e verificati 1000 pezzi. Vogliamo formulare un test, di livello 5%, per decidere se l'ipotesi H_0 "percentuale di pezzi difettosi non superiore al 2%" è plausibile o no, e applicarlo nel caso in cui vengono rilevati 25 pezzi difettosi (tra i 1000 estratti).

Consideriamo un campione aleatorio $X_1, \dots, X_{1000}$ di variabili Bernoulli di parametro p (con esito 1 che indica un pezzo estratto difettoso, 0 altrimenti). L'ipotesi è $H_0 : p \leq 0.02$ contro $H_1 : p > 0.02$, il livello α è 0.05 ($q_{0.95} = 1.64$). La regione critica approssimata è

$$C = \left\{ \sqrt{n}\frac{\bar{X} - 0.02}{\sqrt{0.02(1-0.02)}} > q_{1-\alpha} \right\} = \{\bar{X} - 0.02 > 0.0073\}.$$

Se il numero di pezzi difettosi trovati è 25, allora la media campionaria $\bar{x}$ (frequenza relativa campionaria) vale $25/1000 = 0.025$ e quindi i dati cadono nella regione (approssimata) di accettazione C^c: l'ipotesi $H_0 : p \leq 0.02$ viene accettata (è plausibile) a livello $\alpha = 0.05$. Per $\bar{x} = 0.025$,

$$\mathbb{P}_{0.02}\left(\frac{\sqrt{n}(\bar{X} - 0.02)}{\sqrt{0.02 \cdot 0.98}} > \frac{\sqrt{n}(\bar{x} - 0.02)}{\sqrt{0.02 \cdot 0.98}} \right) \approx 1 - \Phi\left(\frac{\sqrt{n}\bar{x} - 0.02}{\sqrt{0.02 \cdot 0.98}} \right) = 1 - \Phi(1.13) = 0.129$$

restituisce il p-value approssimato.

E. Esercizi

Esercizio IX.16. Una ditta produce aste e dichiara una lunghezza media di 2.3m. Viene rilevata la lunghezza per 100 aste, ottenendo una media campionaria di 2.317m e deviazione standard campionaria di 0.1m. L'affermazione della ditta sulla lunghezza media è plausibile o no? a quale livello?

Esercizio IX.17. Un certo tipo di misurazioni del pH di una sostanza produce misure distribuite in modo Gaussiano, con media pari al valore autentico del pH e deviazione standard $\sigma = 0.02$. Dieci misurazioni forniscono una media campionaria di 8.179. (i) Formulare un test di livello 0.05 per decidere se l'ipotesi "valore del pH pari o superiore a 8.2" sia plausibile o no e applicarlo al valore 8.179 assunto dalla media campionaria. (ii) Qual è il p-value corrispondente? (iii) Qual è la potenza del test in 8.1?

Esercizio IX.18. Una ditta produce una lozione per la ricrescita dei capelli ed afferma che si nota una ricrescita in almeno il 60% dei casi: tuttavia l'unione consumatori ha effettuato un'indagine ed ha rilevato che su 137 persone che hanno usato quella lozione solo 70 hanno notato una ricrescita dei capelli ed afferma che questa indagine contraddice l'affermazione della ditta. In termini di un modello statistico, se p è la percentuale (sconosciuta) delle persone sulle quali la lozione ha effetti positivi, l'affermazione della ditta si traduce nell'ipotesi

$$H_0)\, p \geq 0.6 \text{ contro } H_1)\, p < 0.6.$$

(i) Si può accettare, al livello 0.05, l'ipotesi sopra indicata (cioè l'affermazione della ditta)? (ii) A quale livello (approssimativamente) si può accettare l'affermazione della ditta?

Esercizio IX.19. Si hanno a disposizione 16 osservazioni indipendenti di una v.a. Gaussiana con media μ sconosciuta e varianza nota eguale a 36 ; con queste oservazioni si vuole effettuare il test dell'ipotesi $H_0)\, \mu = 30$ contro $H_1)\, \mu \neq 30$.

Si decide di respingere l'ipotesi $H_0)$ se la media campionaria delle 16 osservazioni cade al di fuori dell'intervallo $(26.91, 33.09)$. A quale livello viene effettuato il test? Se le osservazioni fossero 25 e il test venisse effettuato ancora allo stesso livello della domanda precedente, quale sarebbe la regione critica?

X. T-TEST

Si chiamano T-test quei test statistici che sono basati su una statistica del campione che ha distribuzione di Student. L'esempio di base è il test per la media m di un campione Gaussiano $N(m,\sigma^2)$ con varianza non nota, noto appunto come *T-test di Student*.

A. Test di Student

Consideriamo un campione $X_1,\dots X_n$ Gaussiano $N(m,\sigma^2)$ di cui non conosciamo la deviazione standard. Ricordando le notazioni

$$\bar X_n=\frac{\sum_{i=1}^n X_i}{n},\ S_n^2=\frac{\sum_{i=1}^n (X_i-\bar X_n)^2}{n-1},$$

determiniamo un test statistico per il parametro m procediamo similmente a quanto fatto per lo Z-test, ma considerando la statistica

$$T_n=\sqrt{n}\frac{\bar X_n-m}{S_n},$$

che sappiamo avere legge di student $T(n-1)$ ($n-1$ gradi di libertà) e che dipende dal parametro da testare m ma non dal parametro ignoto σ.

Consideriamo il test bilatero di ipotesi

$$H_0)m=m_0,\qquad H_1)m\neq m_0,$$

e cerchiamo una regione critica della forma $C=\{|\bar X_n-m_0|>d\}$. Imponiamo livello α (e massima potenza) risolvendo:

$$\begin{aligned}\alpha&=\mathbb{P}_{m_0}\left(\left|\bar X_n-m_0\right|>d\right)\\&=\mathbb{P}_{m_0}\left(\frac{\sqrt{n}}{S}\left|\bar X_n-m_0\right|>\frac{d\sqrt{n}}{S}\right)\\&=\mathbb{P}\left(|T_n|>\frac{d\sqrt{n}}{S}\right),\end{aligned}$$

da cui $d\sqrt{n}/S_n=\tau_{1-\frac{\alpha}{2},n-1}$, dove $\tau_{\beta,k}$ è il β-quantile della densità di Student $T(k)$. La regione critica a livello α è quindi

$$C=\left\{\sqrt{n}\frac{\left|\bar X_n-m_0\right|}{S}>\tau_{1-\frac{\alpha}{2},n-1}\right\}.$$

Il p-value dei dati $(x_1,\dots x_n)$ è in questo caso:

$$\begin{aligned}&\mathbb{P}_{m_0}\left(\sqrt{n}\frac{\left|\bar X_n-m_0\right|}{S}>\frac{\sqrt{n}}{s}|\bar X_n-m_0|\right)\\&\qquad=2\left[1-F_{T_{n-1}}\left(\frac{\sqrt{n}}{s}|\bar X_n-m_0|\right)\right],\end{aligned}$$

dove $\bar x_n$ e $\bar s_n$ sono rispettivamente la media campionaria e la varianza campionaria di $(x_1,\dots x_n)$ e dove $F_{T_{n-1}}$ è la c.d.f. della variabile $T(n-1)$ (da approssimare numericamente).

In modo simile si ottengono le formule per il test unilatero

$$H_0)m\leq m_0,\qquad H_1)m>m_0,$$

per il quale la regione critica a livello α è

$$C=\left\{\sqrt{n}\frac{(\bar X_n-m_0)}{S_n}>\tau_{1-\alpha,n-1}\right\},$$

e il p-value è

$$\begin{aligned}&\bar\alpha(x_1,\dots,x_n)\\&\quad=\mathbb{P}_{m_0}\left(\sqrt{n}\frac{(\bar X_n-m_0)}{S_n}>\frac{\sqrt{n}}{s}(\bar X_n-m_0)\right)\\&\qquad=1-F_{T_{n-1}}\left(\frac{\sqrt{n}}{s}(\bar X_n-m_0)\right).\end{aligned}$$

Esempio X.1. Riprendiamo l'esempio del peso (in kg) di una popolazione di salmoni, supponendo sempre una distribuzione Gaussiana ma questa volta di varianza non nota (e anche di media m non nota). Viene rilevato il peso su un campione di 16 esemplari. Vogliamo formulare un test, di livello 5%, per decidere se l'ipotesi H_0 "peso medio pari a 3" è plausibile oppure no, e applicarlo nel caso in cui le misurazioni abbiano media campionaria 3.2 e deviazione standard campionaria 0.5.

L'ipotesi è $H_0: m=3$ contro $H_1: m\neq 3$, il livello α è 0.05 ($\tau_{0.975,15}=2.13$) e la regione critica C è

$$\begin{aligned}C&=\left\{\sqrt{n}\frac{|\bar X-3|}{S}>\tau_{1-\alpha/2,n-1}\right\}\\&=\left\{\frac{|\bar X-3|}{S}>0.5325\right\}.\end{aligned}$$

Dunque, se media campionaria e varianza campionaria valgono $\bar x=3.2$ e $s=0.5$, anche in questo caso i dati cadono nella regione di accettazione C^c e quindi l'ipotesi H_0: "peso medio pari a 3" viene accettata a livello $\alpha=0.5$. Il p-value vale:

$$\begin{aligned}&\mathbb{P}_3\left(\sqrt{n}\frac{\left|\bar X_n-3\right|}{S}>\frac{\sqrt{16}}{0.5}|3.2-3|\right)\\&\qquad=2\left[1-F_{T_{15}}\left(\frac{\sqrt{16}}{0.5}|3.2-3|\right)\right]\\&\qquad=2(1-0.9347)=0.1306.\end{aligned}$$

Anche se il valore di s è uguale al valore $\sigma=0.5$ dell'esempio con varianza nota, il p-value è più alto nel caso nel caso varianza non nota: l'incertezza sulla varianza si traduce in un'ampiezza maggiore della regione di accettazione.

B. Test approssimato sulla media di un campione di taglia grande

Come per gli intervalli di fiducia, l'approssimazione tramite il Teorema Centrale del Limite si può usare in generale per formulare test sulla media di una popolazione non necessariamente Gaussiana, se il campione ha *taglia n grande*.

Sia dunque $X_1,\dots X_n$ un campione i.i.d. con media m e varianza σ^2, con n grande. Per il Teorema Centrale del Limite, le statistiche

$$Z_n=\sqrt{n}\frac{\bar X_n-m}{\sigma},\quad T_n=\sqrt{n}\frac{\bar X_n-m}{S_n},$$

hanno approssimativamente legge Gaussiana standard $N(0,1)$ e di Student $T(n-1)$. **Si possono dunque usare** (in modo approssimato) **i test sopra descritti anche per la media di questo campione non Gaussiano**.

Ad esempio, per il test bilatero

$$H_0) m = m_0, \qquad \text{contro} \qquad H_1) m \neq m_0$$

al livello α, una regione critica approssimata è

$$C = \left\{ \sqrt{n} \frac{|\bar{X}_n - m_0|}{S_n} > \tau_{n-1,1-\frac{\alpha}{2}} \right\}$$

con p-value

$$\mathbb{P}\left(|T_n| > \frac{\sqrt{n}}{\sigma} |\bar{X}_n - m_0| \right) \approx 2\left[1 - F_{n-1}\left(\frac{\sqrt{n}}{\sigma} |\bar{X}_n - m_0| \right) \right].$$

Se σ è nota si può usare invece lo Z-test. Considerazioni del tutto analoghe valgono per test unilateri.

C. Confronto tra campioni

Consideriamo due esempi di test statistici apparentemente simili:

Esempio X.2. Viene propagandata una cura dimagrante che promette *7 chili in 7 giorni*: vengono misurati *"prima e dopo"* 7 pazienti ottenendo i valori che seguono:

- "prima": 72 89 94 77 86 91 83,
- "dopo": 68 83 90 71 80 88 74.

Si può accettare l'affermazione dell'istituto che offre la cura? Quale sarebbe stata la conclusione se invece di promettere 7 chili in 7 giorni avessero promesso solo 6 chili in 7 giorni?

Esempio X.3. Vengono misurate le lunghezze delle tibie di uomini adulti provenienti da reperti di tombe etrusche: dal sito di Cerveteri vengono effettuate 13 misurazioni ottenendo un valore medio di 47.2 ed una varianza campionaria di 7.98 , mentre dal sito di Ladispoli si ottengono 8 misurazioni con un valore medio di 44.9 ed una varianza campionaria di 8.85. Si può affermare che la differenza sia una semplice fluttuazione statistica oppure si deve concludere che gli abitanti di Cerveteri erano veramente più alti?

Fermo restando che questi campioni sono troppo poco numerosi per trarre delle conclusioni veramente significative, si tratta di due situazioni *radicalmente diverse*.

Nel primo caso i campioni sono *accoppiati*: se chiamiamo X_i, Y_i il peso della i-sima persona del campione rispettivamente prima e dopo la cura, X_i e Y_i non sono indipendenti (se sappiamo ad esempio che X_i è molto alto, la probabilità di avere anche Y_i alto sarà più alta). Nel secondo caso i campioni sono invece *indipendenti*: se chiamiamo X_i la lunghezza dell'i-simo elemento del campione del sito di Cerveteri, Y_j la lunghezza del j-simo elemento del campione del sito di Ladispoli, informazioni su X_i non cambiano le probabilità relative agli Y_j.

Nel caso dei campioni *accoppiati* non c'è sostanzialmente nulla di nuovo da aggiungere dal punto di vista teorico: si considera il campione $V_i = X_i - Y_i$ dato dalle differenze delle coppie e si esegue il T-test per il parametro $m = \mathbb{E}[V_i]$. L'ipotesi nulla è $H_0) m \geq 7$ contro l'ipotesi alternativa $H_1) m < 7$. Usiamo quindi un T-test unilatero per il campione Gaussiano V_i con varianza sconosciuta. Prendiamo la statistica $T = \sqrt{7}\frac{(\overline{V}-7)}{S}$, dove $\overline{V}$ e S sono rispettivamente media campionaria e varianza campionaria. Il p-value del test è

$$\mathbb{P}\left(T \leq \sqrt{7}\frac{(\overline{v}-7)}{s} \right) = F_6\left(\sqrt{7}\frac{(\overline{v}-7)}{s} \right),$$

con $\bar{v}$, s media e varianza campionaria dei dati v_i del campione V_i, e valgono

$$\sqrt{7}\frac{(\overline{v}-7)}{s} = -2.09, \quad F_6\left(\sqrt{7}\frac{(\overline{v}-7)}{s} \right) = 0.04,$$

cioè l'ipotesi H_0 è poco credibile. Se si considera l'ipotesi $H_0) m \geq 6$, risulta

$$\sqrt{7}\frac{(\overline{v}-6)}{s} = -0.76, \quad F_6\left(\sqrt{7}\frac{(\overline{v}-6)}{s} \right) = 0.23,$$

e questa volta l'ipotesi è molto più verosimile.

Quando invece i due campioni sono *indipendenti* (tra l'altro in genere di numerosità diverse) la situazione è più complicata e rientra nell'ambito dell'inferenza statistica detto *analisi della varianza*, argomento ampio in cui il T-test è solo il primo passo, a cui ci limitiamo. Ci occupiamo qui del caso di campioni Gaussiani indipendenti in cui le varianze teoriche dei due campioni sono sconosciute ma uguali, e facciamo uso del seguente risultato:

Teorema X.4. *Siano $X_1, \dots, X_n$ un campione con legge $N(m_1, \sigma^2)$ e $Y_1, \dots, Y_k$ un campione indipendente dal primo con legge $N(m_2, \sigma^2)$, con m_1, m_2, σ parametri non noti. La variabile*

$$T_{n,k} = \frac{\sqrt{n+k-2}}{\sqrt{\frac{1}{n}+\frac{1}{k}}} \cdot \frac{(\bar{X}_n - \overline{Y}) - (m_1 - m_2)}{\sqrt{\sum_i (X_i - \bar{X}_n)^2 + \sum_j (Y_j - \overline{Y})^2}}$$

ha densità di Student $T(n+k-2)$ con $(n+k-2)$ gradi di libertà.

Si può quindi formulare un T-test su $m_1 - m_2$, la differenza delle medie nelle due popolazioni. Sia X_i la lunghezza della tibia dell'i-simo elemento del campione di Cerveteri, $i = 1, \dots n = 13$, e sia Y_j la lunghezza della tibia del j-simo elemento del campione di Ladispoli, $j = 1, \dots k = 8$. Supponiamo che i due campioni siano Gaussiani e che le loro varianze siano uguali (le varianze campionarie sono abbastanza simili). Chiamiamo m_1, m_2 i valori attesi per la lunghezza della tibia per gli uomini di Cerveteri e di Ladispoli rispettivamente, poniamo $m = m_1 - m_2$. Consideriamo il test dell'ipotesi $H_0) m = 0$ (cioè stessa media) contro $H_1) m > 0$ (cioè la media a Cerveteri è superiore alla media a Ladispoli).

Per il teorema precedente, sotto l'ipotesi nulla la variabile

$$T_{n,k} = \frac{\sqrt{19}}{\sqrt{\frac{1}{13}+\frac{1}{8}}} \cdot \frac{\bar{X}_n - \overline{Y}}{\sqrt{\sum_i (X_i - \bar{X}_n)^2 + \sum_j (Y_j - \overline{Y})^2}}$$

ha densità di Student $T(19)$. Usiamo quindi un T-test unilatero: la regione critica di livello α è

$$C = \{T > \tau_{1-\alpha,19}\}.$$

Il valore assunto da $T_{n,k}$ per i dati raccolti è

$$t_{n,k} = \frac{47.2 - 44.9}{\sqrt{7.98 \cdot 12 + 8.85 \cdot 7}} \frac{\sqrt{19}}{\sqrt{\frac{1}{13} + \frac{1}{8}}} = 1.776,$$

e il *p-value* che risulta è quindi

$$\mathbb{P}(T > t) = 1 - F_{19}(1.776) = 0.045.$$

Si tratta di un valore piuttosto basso (anche se non trascurabile): ad esempio a livello $\alpha = 0.07$ si conclude che gli abitanti di Cerveteri erano effettivamente più alti, mentre non si può avallare tale conclusione per $\alpha = 0.01$.

Naturalmente rimane aperto il problema di decidere se è verosimile che i due campioni abbiano la stessa varianza. Questa restrizione può essere superata con test più raffinati, come il T-test di Welch o lo F-test di Fisher, che però non illustriamo.

D. T-test di regressione

Consideriamo un esperimento in cui ad ogni ripetizione misuriamo due diverse quantità (dati bivariati), tra cui ipotizziamo valere una relazione lineare perturbata da un errore (Gaussiano) dovuto alla procedura di misurazione: vogliamo determinare un test statistico per il coefficiente angolare di tale relazione lineare. Questo problema, detto della *regressione lineare*, è il punto di partenza dell'*analisi della regressione*, che a sua volta è tra i fondamentali della contemporanea teoria dell'*apprendimento statistico*. Per fissare le idee, diamo un esempio classico.

Esempio X.5. In Econometria, la *legge di Okun* prevede che ci sia una relazione lineare tra il tasso di disoccupazione (variazione in percentuale dell'occupazione della popolazione, ad esempio quadrimestrale) e il tasso di crescita del prodotto interno lordo di un Paese (variazione in percentuale del PIL sullo stesso periodo). Arthur Melvin Okun propose l'ipotesi che a ogni punto percentuale di disoccupazione in più corrisponda la perdita di due punti percentuali di PIL.

Consideriamo quindi n dati bivariati $(x_1, y_1), \dots, (x_n, y_n)$. Vogliamo descrivere gli esiti y_i con delle variabili aleatorie

$$Y_i = a + bx_i + E_i,$$

in cui $E_1, \dots, E_n$ sono un campione Gaussiano $N(0, \sigma^2)$ (σ ignota) che descrive gli errori di misurazione, e a, b, σ sono quindi parametri liberi da stimare. I coefficienti della retta di regressione introdotti all'inizio delle note forniscono buoni stimatori per a, b, in particolare:

Teorema X.6. *Le statistiche*

$$B_n = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}$$
$$A_n = \bar{Y}_n - B_n \bar{x}_n,$$

sono stimatori corretti di b, a rispettivamente. Inoltre B_n ha densità Gaussiana $N(b, \sigma^2 / \sum_{i=1}^n (x_i - \bar{x}_n)^2)$, la statistica

$$Q_n = \sum_{i=1}^n (Y_i - A_n - B_n x_i)^2$$

ha densità Chi-Quadro $\chi^2(n-2)$ ed è indipendente da B_n, e la statistica

$$T_n = \sqrt{\frac{(n-2)\sum_{i=1}^n (x_i - \bar{x}_n)^2}{Q_n}} \cdot (B_n - b)$$

ha densità di Student $T(n-2)$.

Possiamo quindi pianificare un T-test per il parametro b, con a, σ non noti: una regione critica di livello α per le ipotesi $H_0)\, b = b_0$ contro $H_1)\, b \neq b_0$ è

$$C = \left\{ |B_n - b_0|^2 \geq \frac{\tau_{n-2,1-\frac{\alpha}{2}} \sum_{i=1}^n E_i^2}{(n-2)\sum_{i=1}^n (x_i - \bar{x}_n)^2} \right\},$$

e il relativo p-value è $\bar{\alpha} = 2\,(1 - F_{n-2}\,(t_n))$, dove

$$t_n = (b_n - b_0)\sqrt{\frac{(n-2)\sum_{i=1}^n (x_i - \bar{x}_n)^2}{\sum_{i=1}^n (y_i - a_n - b_n x_i)^2}}$$

Equivalentemente, ricordando la definizione di coefficiente di correlazione empirico $r_n = r(x, y)$ tra i dati x_i e y_i data nella prima parte delle note, e la relazione

$$\sum_{i=1}^n (y_i - a_n - b_n x_i)^2 = (1 - r_n^2) \sum_{i=1}^n (y_i - \bar{y}_n)^2,$$

si può esprimere il p-value in termini dei soli a_n, b_n e $r = r(x, y)$:

$$\bar{\alpha} = 2\left(1 - F_{n-2}\left((b_n - b_0)\frac{b_n}{r_n}\sqrt{\frac{n-2}{1-r_n^2}}\right)\right).$$

Osservazione X.7. Grazie al Teorema Centrale del Limite, il test sopra descritto si può applicare anche quando gli errori E_i non hanno distribuzione Gaussiana, ma sono dotati di momento secondo e la numerosità n del campione è grande. Osserviamo inoltre che grazie al Teorema sopra enunciato è possibile costruire intervalli di fiducia per il parametro b.

Esempio X.8. Nell'articolo originale in cui Okun propose la sua legge (1962) egli considerava $n = 55$ osservazioni quadrimestrali $(x_i, y_i) \in [0, 1]^2$ in cui x_i è la variazione percentuale del PIL e y_i il tasso di occupazione (negli USA). Dai dati egli otteneva:

$$b_n = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} = -0.30,$$
$$a_n = \bar{y}_n - b_n \bar{x}_n = 0.30,$$

ma con un coefficiente di correlazione non molto vicino a 1, ovvero

$$r(x, y) = b_n \cdot \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_n)^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}} = 0.79\,.$$

Si vuole testare l'ipotesi $H_0) b = -0.5$, e quindi dai dati proposti ricaviamo il p-value (alto)

$$\bar{\alpha} = 2(1 - F_{53}(0.90)) = 2(1 - 0.81) = 0.38\,.$$

E. Esercizi

Esercizio X.9. Una ditta produce chiodi di lunghezza dichiarata uguale a 5 cm, e il proprietario della ditta afferma che la deviazione standard delle lunghezze dei chiodi prodotti non supera 0.2 cm. Analizzando la lunghezza di un campione di 16 pezzi si trova media campionaria di 4.935 cm e varianza campionaria 0.06 cm. Supponendo che le lunghezze dei chiodi possano essere rappresentate con v. a. Gaussiane, si può accettare al livello 0.05 l'affermazione del proprietario della ditta sulla deviazione standard della lunghezza dei chiodi prodotti? Descrivere un test da utilizzare per verificare l'ipotesi che la lunghezza dei chiodi prodotti sia di 5 cm, e usare i dati del campione analizzato per determinare la plausibilità di questa ipotesi.

Esercizio X.10. Il responsabile di una ditta petrolifera afferma che il contenuto medio di zolfo per litro, nella benzina prodotta da quella ditta, non supera 0.15mg/l; tuttavia l'unione consumatori contesta questa affermazione perché sono stati prelevati 41 campioni che hanno dato valori $x_1, \dots, x_{41}$ dai quali si ottiene

$$\bar{x} = \frac{x_1 + \cdots + x_{41}}{41} = 0.2.$$

Il responsabile afferma che questo dato non è significativo poiché la variabilità era alta: si è infatti ottenuto il valore $\sum_{i \leq 41} (x_i - \bar{x})^2 = 1$. (Si interpretino i valori delle misurazioni come variabili Gaussiane con media e varianza ignote.) Si può accettare l'affermazione del responsabile della ditta (cioè l'ipotesi che il contenuto medio di zolfo non superi 0.15mg/l) ? Impostare un opportuno test e calcolare il relativo p-value. Scrivere poi l'intervallo di fiducia unilatero destro (della forma $[a, +\infty)$) per il contenuto medio di zolfo al livello del 95%.

Esercizio X.11. Vengono effettuate 25 misurazioni del peso di un piccolo mammifero africano, e si ottengono i risultati $(x_1, \dots, x_{25})$ (espressi in grammi) dai quali si ricava una media campionaria eguale a 648.6 ed una varanza campionaria eguale a 457.66.

Si vuole verificare l'ipotesi che il peso medio di questo mammifero sia di 640 grammi, e a tal fine si suppone che i dati possano essere rappresentati con 25 variabili Gaussiane indipendenti. Effettuare il test dell'ipotesi

$$H_0)\, m = 640 \quad \text{contro} \quad H_1)\, m \neq 640$$

supponendo che la varianza sia nota ed eguale a 400: dopo aver calcolato il p-value, che cosa si conclude? Effettuare lo stesso test sopra scritto, supponendo però che la varianza sia sconosciuta: che cosa si conclude?

Esercizio X.12. Un certo farmaco viene testato per controllare possibili effetti collaterali. Su 863 pazienti cui viene somministrato il farmaco, 19 manifestano sintomi influenzali. Sappiamo (da precedenti studi) che il tasso di insorgenza di sintomi influenzali, in pazienti non trattati con il farmaco in esame, è 1.9%. Vi è evidenza a livello 0.05 che il farmaco induca sintomi influenzali? Calcolare il p-value dei dati.

Esercizio X.13. Si vuole testare un nuovo farmaco contro il colesterolo alto. Per questo, il farmaco viene somministrato a 100 persone con un alto livello di colesterolo. Al termine del periodo di somministrazione, si registra una riduzione media di 7.8, con una deviazione standard campionaria di 6.4. Che conclusioni possiamo trarne sull'efficacia del farmaco?

Esercizio X.14. Due campioni di due soluzioni vengono analizzati per confrontarne il pH, con lo stesso metodo di misurazione. L'errore di misurazione ha distribuzione Gaussiana. Vengono effettuate 10 misurazioni (indipendenti) per ciascuna soluzione, ottenendo per la prima soluzione una media di 6.267 e una deviazione standard di 0.0295, mentre per la seconda soluzione una media di 6.285 e una deviazione standard di 0.0327. Formulare un test di livello 0.01 per decidere se l'ipotesi "le due soluzioni hanno lo stesso pH" è plausibile o no e applicarlo ai valori assunti dai campioni. Calcolare il p-value dei dati.

XI. TEST CHI-QUADRO

Se un campione $X_1, \dots, X_n$ è formato da variabili Gaussiane con varianza σ^2, la variabile

$$Q_n := \frac{n-1}{\sigma^2} S_n^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2$$

ha densità $\chi^2(n-1)$. Combinato al Teorema Centrale del Limite, questo fatto fa sì che la distribuzione Chi-Quadro sia al centro di una varietà di test per parametri che quantificano l'incertezza di una misurazione.

A. Test sulla varianza di un campione Gaussiano

Come per gli intervalli di fiducia per la varianza, ci occupiamo solo del caso unilatero

$$H_0)\sigma^2 \le \sigma_0^2 \qquad \text{contro} \qquad H_1)\sigma^2 > \sigma_0^2.$$

Poiché $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$ è uno stimatore corretto e consistente di σ^2, risulta naturale considerare una regione critica della forma

$$C = \left\{ \sum_i \left(X_i - \bar{X}_n\right)^2 > d \right\},$$

in cui d viene scelto imponendo livello α, ovvero chiedendo che per $\sigma \le \sigma_0$,

$$\begin{aligned}\alpha &\ge \mathbb{P}_\sigma\left(\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 > d\right) \\ &= \mathbb{P}_\sigma\left(\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}{\sigma^2} > \frac{d}{\sigma^2}\right) \\ &= \mathbb{P}(Q_n > d/\sigma^2),\end{aligned}$$

in cui la legge di Q_n non dipende da σ, per cui la probabilità sopra scritta cresce con σ. Dunque è sufficiente imporre la condizione per $\sigma = \sigma_0$, e massimizzando il livello chiediamo quindi $\mathbb{P}(Q_n > d/\sigma_0^2) = \alpha$, e questo ci porta a scegliere $\frac{d}{\sigma_0^2} = \chi^2_{1-\alpha,n-1}$. Si ottiene quindi la regione critica di livello α

$$\begin{aligned}C &= \left\{ \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}{\sigma_0^2} > \chi^2_{1-\alpha,n-1} \right\} \\ &= \left\{ \frac{n-1}{\sigma_0^2} S_n^2 > \chi^2_{1-\alpha,n-1} \right\}.\end{aligned}$$

Il p-value dei dati $(x_1, \dots x_n)$ è

$$\begin{aligned}&\mathbb{P}_{\sigma_0}\left(\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}{\sigma_0^2} > \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2}{\sigma_0^2}\right) \\ &\qquad = 1 - G_{n-1}\left(\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2}{\sigma_0^2}\right),\end{aligned}$$

dove con G_n si indica la c.d.f. della distribuzione $\chi^2(n)$, per i cui valori si ricorre ad approssimazioni numeriche.

Se si considera un test unilatero destro, cioè dell'ipotesi $H_0)\sigma^2 \ge \sigma_0^2$, la discussione è adattata invertendo tutte le diseguaglianze e cambiando il quantile $\chi^2_{1-\alpha,n-1}$ con $\chi^2_{\alpha,n-1}$.

B. Test Chi-Quadro di Pearson

Il Teorema di Pearson è un teorema limite (che si può ricondurre a una opportuna generalizzazione del Teorema Centrale del Limite) che permette di effettuare test statistici basati sui quantili della distribuzione chi-quadro, per la verifica di ipotesi in un grande varietà di situazioni. Ne vediamo due applicazioni: la prima a test per verificare se un campione ha legge compatibile con una distribuzione di probabilità assegnata, la seconda per verificare l'indipendenza di due (campioni di) variabili aleatorie.

Consideriamo dapprima il caso in cui $X_1, \dots, X_n$ sia un campione di v.a. discrete che assumono un insieme finito di valori, ovvero la funzione di massa è non-nulla solo su un sottoinsieme finito che possiamo supporre senza perdita di generalità essere $\{1, \dots, r\}$,

$$q_j = \mathbb{P}(X_i = j), \quad j = 1, \dots, r,$$

in cui naturalmente $\sum_{j=1}^{r} q_j = 1$. In altre parole, la legge del campione si può rappresentare con il *vettore di probabilità* $q = (q_1, \dots, q_r)$.

Vogliamo definire un test statistico per verificare se la legge del campione coincide con una legge assegnata rappresentata da un vettore di probabilità $p = (p_1, \dots, p_r)$, con $p_j \ge 0$ e $\sum_{j=1}^{r} p_j = 1$. L'ipotesi nulla è dunque $H_0)q = p$, mentre l'alternativa è $H_1)q \ne p$. Per determinare una regione critica, facciamo affidamento alla statistica di Pearson,

$$T_n = \sum_{j=1}^{r} \frac{\left(O_{n,j} - np_j\right)^2}{np_j},$$

in cui la v.a.

$$O_{n,j} = \sum_{i=1}^{n} 1_{\{X_i = j\}} = \#\left\{i : X_i = x_j\right\},$$

detta *effettivo empirico* indica il numero di osservazioni X_i che hanno dato valore j. Dalla Legge dei Grandi Numeri segue che $O_{n,j}/n$ converge in probabilità a q_j, e per analogia chiameremo allora effettivo teorico del valore j la quantità np_j.

Teorema XI.1 (di Pearson)**.** *Se la legge comune delle X_j è diversa da p, allora*

$$\mathbb{P}(\lim_{n\to\infty} T_n = +\infty) = 1.$$

Se invece la legge comune delle X_j coincide con p, allora T_n converge in legge ad una $\chi^2(r-1)$.

In altre parole, per n grande, sotto l'ipotesi alternativa T_n assumerà valori molto grandi: cerchiamo quindi una regione critica della forma $\{T_n > c\}$, con c da determinare. Per imporre livello α (massimizzando la potenza) chiediamo che valga $\mathbb{P}(T_n > c) = \alpha$ *sotto l'ipotesi nulla.* Qui interviene la seconda parte del Teorema, che ci permette di approssimare la c.d.f di T_n con quella di una variabile chi-quadro, e dunque di ottenere la condizione

$$\mathbb{P}_{H_0}(T_n > c) \approx 1 - G_{r-1}(c) = \alpha,$$

da cui $c = \chi^2_{r-1,1-\alpha}$, per cui la regione critica è

$$C_\alpha = \left\{T_n > \chi^2_{r-1,1-\alpha}\right\},$$

e il relativo p-value è

$$\bar{\alpha} = 1 - G_{r-1}\left(\sum_{j=1}^{r} \frac{(o_{n,j} - np_j)^2}{np_j}\right),$$

in cui $o_{n,i}$ indica il valore assunto da $O_{n,i}$ nei dati. Una condizione pratica sotto la quale l'approssimazione fatta, e dunque il test del chi-quadro, viene considerata attendibile è che la numerosità n del campione sia tale che $np_j \geq 5$, per ogni $j = 1, \ldots, r$.

Vale la pena di osservare che i *gradi di libertà* della variabile χ^2 che stiamo utilizzando contano i parametri che stiamo stimando: in un vettore di probabilità $p = (p_1, \ldots, p_r)$ ci sono in effetti solo $r-1$ parametri liberi, perché se sono note le prime $r-1$ componenti del vettore l'ultima è determinata dal vincolo $\sum_{j=1}^{k} p_j = 1$.

Esempio XI.2. La fabbrica di caramelle N&N's produce caramelle di 5 colori, e le imbusta in pacchetti da 100, cercando di mescolare i colori in modo che essi compaiano in egual quantità in ogni sacchetto: si conduce un test statistico per verificarlo. Avendo aperto 10 sacchetti (n =1000 caramelle) ci sono rispettivamente 180, 250, 120, 225, 225 caramelle per i 5 colori possibili (questi sono le osservazioni $o_{n,j}$). La statistica test risulta essere $t_n = 52.75$, che conduce (considerando la c.d.f. G_4 con 4 gradi di libertà) a un $p-$value praticamente nullo. A livello $\alpha = 0.05$ infatti il quantile è $\chi^2_{4,0.95} \simeq 9.488$. Siamo quindi condotti a rigettare l'ipotesi nulla che il mescolamento delle caramelle sia efficiente.

C. Confronto di Distribuzioni Generiche

Il test chi-quadro può essere applicato per confrontare la legge di un campione $X_1, \ldots, X_n$ con quella di una variabile aleatoria nota X generica. Se tutte le variabili coinvolte assumono solo finiti valori si applica la discussione precedente, altrimenti ci si può ricondurre a quel caso nel seguente modo. Si partiziona $\mathbb{R}$ in un numero finito di intervalli,

$$(-\infty = a_0, a_1], (a_1, a_2], \ldots, (a_{k-2}, a_{k-1}], (a_{r-1}, a_r = +\infty)$$

e si applica il test al campione di variabili $Y_i : \Omega \to \{1, \ldots, r\}$ che prendono valore j se X_i è cade in $(a_{j-1}, a_j]$. In simboli, si definisce

$$h(x) = \sum_{j=1}^{r} j \cdot \begin{cases} 1 & x \in (a_{j-1}, a_j] \\ 0 & x \notin (a_{j-1}, a_j] \end{cases}$$

e $Y_i = h(X_i)$. Si verifica facilmente che le Y_i sono variabili i.i.d. che prendono valori in $\{1, \ldots, r\}$. Il vettore di probabilità dell'ipotesi nulla si determina analogamente:

$$p_j = \mathbb{P}(X \in (a_{j-1}, a_j]),$$

in cui X è la variabile nota con cui vogliamo confrontare il campione.

La scelta della partizione in intervalli $(a_{j-1}, a_j]$ *è arbitraria e deve essere effettuata sulla base di considerazioni adeguate al problema specifico.* Il problema è analogo a scegliere adeguatamente gli intervalli per disegnare un istogramma significativo per dati $x_1, \ldots, x_n$.

Esempio XI.3. Si vuole testare l'ipotesi che il contatto con una persona infetta da una certa malattia determini il contagio con probabilità 50%. Potendo testare su un singolo individuo un alto numero di contatti potremmo usare uno Z-test per la distribuzione di Bernoulli, ma in realtà il dato a disposizione è il numero di contatti con infetti prima dell'infezione dei pazienti, per cui usiamo un test chi-quadro di adattamento alla distribuzione geometrica. I dati registrati per $n = 100$ pazienti sono:

numero contatti infetti	1	2	3	4	≥5
numero contagiati	58	25	14	2	1

Considerando gli $r = 5$ esiti della misurazione effettuata, determiniamo con la distribuzione geometrica di parametro $p = 0.5$ le probabilità teoriche

$$p_1 = 0.5,\ p_2 = 0.25,\ p_3 = 0.125,\ p_4 \simeq 0.062$$

mentre scegliamo $p_5 = 1 - p_1 - p_2 - p_3 - p_4 \simeq 0.063$ per l'ultimo esito che raggruppa tutti quelli della variabile geometrica dal 5 in poi. Calcolando il p-value per il test chi-quadro con $r - 1 = 4$ gradi di libertà si ottiene 0.067, che è decisamente troppo basso per ritenere plausibile l'ipotesi nulla.

Nell'esempio, rimane il dubbio che i dati possano in effetti essere descritti da una distribuzione geometrica ma con un diverso parametro. Una procedura più ragionevole è stimare il parametro con un opportuno stimatore e quindi applicare il test chi-quadro (in effetti lo stimatore di massima verosimiglianza nell'esempio è $\hat{p} \simeq 0.613$). Quando la distribuzione teorica ha a sua volta uno o più parametri stimati a partire dai dati è però necessario sottrarre il numero di tali parametri dal numero dai gradi di libertà del test. Una giustificazione teorica di questo fatto va al di là dei nostri scopi: ci limitiamo a proporre il seguente esempio.

Esempio XI.4. Al fine di stabilire il premio assicurativo, una compagnia di assicurazioni si chiede se il numero di sinistri in un mese per gli abitanti di una certa zona segue una distribuzione di Poisson. Nel corso degli ultimi $n = 50$ mesi si è registrato:

numero sinistri	0	1	2	3	4
numero mesi	7	8	13	10	12

mentre in nessun mese ci sono stati 5 o più incidenti. La la media campionaria è uno stimatore corretto e consistente del valore atteso: poiché la media campionaria sui dati è 2.24, ipotizzando una distribuzione Poisson di parametro λ stimiamo dunque un valore teorico del parametro $\lambda_0 = 2.24$. Possiamo ora applicare il test chi-quadro: consideriamo la partizione in $r = 5$ intervalli

$$(-\infty, 1), [1, 2), [2, 3), [3, 4), [5, \infty)$$

le cui probabilità teoriche sono

$$p_1 = \frac{\lambda_0^0}{0!} e^{-\lambda_0} \simeq 0.106, p_2 = \frac{\lambda_0^1}{1!} e^{-\lambda_0} \simeq 0.238,$$

similmente si calcolano $p_3 \simeq 0.267$, $p_4 \simeq 0.199$ e $p_5 = 1 - p_1 - \cdots - p_4 \simeq 0.188$. Siamo quindi ricondotti a un test chi-quadro per variabili che possono assumere 5 valori distinti, **ma la statistica di Pearson ha in questo caso solo 3 gradi di libertà** e non $r - 1 = 4$: questo perché va sottratto ai gradi di libertà il numero di parametri addizionali stimati, in questo caso 1. Calcolando

la statistica di Pearson come descritto sopra si ottiene il p-value

$$\bar{\alpha} = 1 - G_3(2.530) \simeq 0.470,$$

per cui accettiamo l'ipotesi nulla (distribuzione dei dati Poisson con parametro $\lambda = 2.24$) ad ogni ragionevole livello.

D. Test per l'indipendenza

Siano X e Y due variabili aleatorie che assumono rispettivamente valori $x_1, \dots, x_r$ e $y_1, \dots, y_s$. La variabile aleatoria bivariata (X, Y) assume quindi valori (x_j, y_k), con $j = 1, \dots, r; k = y_1, \dots, y_s$, e i parametri non noti sono le probabilità $\mathbb{P}(X = x_i, Y = y_j)$. Si vuole verificare l'ipotesi nulla

$$H_0) X \text{ e } Y \text{ sono indipendenti,}$$

il che è equivalente a

$$P(X = x_j, Y = y_k) = P(X = x_j)P(Y = y_k).$$

Consideriamo quindi stimatori delle probabilità congiunte e marginali,

$$O_{n,j,k} = \sum_{i=1}^{n} 1_{\{X_i = x_j, Y_i = y_k\}},$$

$$O^X_{n,j} = \sum_{k=1}^{s} O_{n,j,k}, \quad O^Y_{n,k} = \sum_{j=1}^{s} O_{n,j,k},$$

ovvero gli effettivi empirici delle coppie di valori e dei valori singoli. Più precisamente abbiamo le coppie:

$P(X = x_j, Y = y_k)$	$O_{n,j,k}/n$
$P(X = x_j)$	$O^X_{n,j}/n$
$P(Y = y_k)$	$O^Y_{n,k}/n$

di parametri e loro stimatori corretti e consistenti.

Teorema XI.5. *La statistica di Pearson*

$$T = \sum_{i=j}^{r} \sum_{k=1}^{s} \frac{\left(O_{n,j,k} - O^X_{n,j} O^Y_{n,k}/n\right)^2}{O^X_{n,j} O^Y_{n,k}/n}$$

sotto l'ipotesi nulla di indipendenza converge in legge a una variabile con densità $\chi^2((r-1)(s-1))$, mentre sotto la relativa ipotesi alternativa soddisfa

$$\mathbb{P}(\lim_{n\to\infty} T_n = +\infty) = 1.$$

Osserviamo che al prodotto degli stimatori delle probabilità marginali $O^X_{n,j}$ e $O^Y_{n,k}$ viene applicato un fattore $1/n$ per tenere conto del fatto che i prodotti $P(X = x_j)P(Y = y_k)$ hanno come stimatore $O^X_{n,j} O^Y_{n,k}/n^2$, il cui denominatore ha un n in più rispetto a quelli delle probabilità congiunte.

Similmente a quanto discusso nei paragrafi precedenti, la regione critica di livello α è

$$C_\alpha = \left\{ T_n > \chi^2_{(r-1)(s-1),1-\alpha} \right\},$$

e il relativo p-value è $\bar{\alpha} = 1 - G_{(r-1)(s-1)}(t_n)$ con

$$t_n = \sum_{i=j}^{r} \sum_{k=1}^{s} \frac{\left(o_{n,j,k} - o^X_{n,j} o^Y_{n,k}/n\right)^2}{o^X_{n,j} o^Y_{n,k}/n},$$

in cui le o minuscole indicano i valori assunti dalle relative O maiuscole nei dati.

E. Esercizi

Esercizio XI.6. Le fluttuazioni giornaliere di un titolo in borsa vengono descritte con una variabile aleatoria Gaussiana $N(m, \sigma^2)$, di cui è fondamentale avere una buona stima per la varianza σ^2 (in questo contesto detta *volatilità*). Nel corso della fase di negoziazione continua (9:00-17:30) di una giornata in borsa viene registrato il prezzo di una certa azione ogni mezz'ora ($n = 17$ volte) e la deviazione standard campionaria dei prezzi è stata 8.50\$. L'ipotesi "deviazione standard inferiore a 7\$" è plausibile a livello 0.05? La risposta cambia se questo fosse l'esito di misurazioni effettuate ogni quarto d'ora ($n = 34$) ?

Esercizio XI.7. Un botanico prevede sulla base di un modello genetico che il colore dell'infiorescenza di una pianta sia distribuita come segue: 20% rosso, 10% azzurro, 25% bianco, 45% rosa. In un campione di 200 piante dalla coltura in serra se ne contano 40 rosse, 28 azzurre, 43 bianche, 89 rosa. Il modello del botanico è da considerarsi compatibile con le osservazioni?

Esercizio XI.8. Si vuole verificare l'effetto di un farmaco su pazienti affetti da una certa malattia. Su un campione di $n = 200$ persone, ad alcune delle quali viene somministrato un placebo, si registrano i seguenti effetti:

miglioramento	nessuno	lieve	netto
farmaco	29	31	20
placebo	52	57	11

Si effettui un test per l'indipendenza tra l'assunzione del farmaco e il miglioramento del paziente: è plausibile affermare che il farmaco non ha effetto?

Esercizio XI.9. Sono riportati di seguito i voti di laurea degli studenti del corso di laurea in Ingegneria dello scorso anno, aggregati nei quattro curriculum a cui ci si può iscrivere:

	curr. A	curr. B	curr. C	curr. D
tra 90 e 100	31	37	40	25
tra 101 e 105	40	33	40	30
tra 106 e 110	19	20	10	25

Nel relativo report di ateneo, il presidente del corso di studi afferma che la scelta del curriculum non influenza il voto finale. È plausibile questa affermazione?

XII. ESERCIZI DI VERIFICA

A. Quesiti VERO/FALSO

Si determini se le seguenti affermazioni sono vere o false, dimostrandole nel primo caso o producendo un controesempio nel secondo.

1. In uno spazio di probabilità, due eventi disgiunti sono indipendenti.
2. Una variabile aleatoria che assume infiniti valori è una variabile con densità.
3. I momenti $\mathbb{E}[X^n]$ di una variabile aleatoria X di Bernoulli di parametro p sono tutti uguali.
4. Dato un campione aleatorio $X_1, \dots X_n$, il valore atteso delle X_i coincide con la media campionaria del campione $X_1, \dots X_n$.
5. Se due eventi A, B sono indipendenti, lo sono anche i loro complementari A^c, B^c.
6. Dato un campione aleatorio $X_1, \dots, X_n$ la cui legge dipende da un solo parametro $\theta \in \Theta \subseteq \mathbb{R}$, un intervallo di fiducia per il parametro θ è dato da una qualsiasi coppia di numeri $a, b \in \mathbb{R}$ in modo che $I = (a, b)$ contenga θ.
7. Una variabile aleatoria X ammette funzione di densità $f : \mathbb{R} \to \mathbb{R}$ se la probabilità che X assuma il valore $x \in \mathbb{R}$ è dato da $f(x)$.
8. Per effettuare il test di Student sulla media di una popolazione è necessario conoscere la deviazione standard della popolazione.
9. Due variabili aleatorie indipendenti di varianza finita sono sempre scorrelate.
10. In un test statistico, il p-value è la probabilità che l'ipotesi nulla sia vera.
11. Una variabile aleatoria è discreta se e solo se assume una quantità finita di esiti.
12. Una variabile aleatoria X con legge Gaussiana $N(0, \sigma^2)$ prende valori nell'intervallo $[-\sigma, \sigma]$ con probabilità di circa il 95%.
13. In un test statistico, fissare un livello α è equivalente a decidere se il p-value è abbastanza alto da ritenere accettabile H_0).
14. Se un campione statistico $X_1, \dots, X_n$ ha legge con momento primo finito, allora la media campionaria è uno stimatore corretto del valore atteso.
15. Lo stimatore $\frac{1}{n-1}\sum_i (X_i - \bar{X})^2$ ha denominatore $n-1$ per far sì che esso sia consistente.
16. Due eventi A e B di probabilità positiva sono indipendenti se e solo se $P(A \mid B) = P(B)$.
17. Per una variabile aleatoria X con funzione di ripartizione $F(x) = P(X \leq x)$, abbiamo $P(a < X \leq b) = \int_a^b F(x)dx$.
18. Per un campione aleatorio $X_1, \dots, X_n$ di variabili con densità uniforme sull'intervallo $[0, L]$ gli stimatori del parametro L ottenuti con il metodo del momento primo e quello di massima verosimiglianza coincidono.
19. Un intervallo di fiducia ha come estremi due variabili aleatorie e può non contenere il parametro stimando.
20. Imporre il livello di un test statistico fissa un limite alla probabilità di commettere errori di prima specie, ma non dà vincoli su quelli di seconda.
21. La determinazione del p-value non dipende dalla scelta della regione critica del test.
22. Considerando per fissare le idee lo Z-test, all'aumentare della numerosità del campione la regione di accettazione si riduce, ossia è più facile che l'ipotesi nulla sia rifiutata.
23. Se due variabili aleatorie X, Y hanno la stessa distribuzione di probabilità, allora non sono indipendenti.
24. Due variabili aleatorie X, Y soddisfano una relazione lineare $Y = aX + b$ se e solo se la loro correlazione vale $\rho(X, Y) = \pm 1$.
25. Data una generica variabile aleatoria X, la sua funzione di ripartizione è data dall'integrale $F_X(t) = \int_0^t f(x)dx$ di una certa funzione f.
26. Scambiare l'ipotesi nulla e quella alternativa può cambiare l'esito di un test statistico.
27. In generale, più basso è il livello imposto a un test statistico, inferiore è la potenza.
28. In un test statistico una forte evidenza a favore dell'ipotesi nulla può produrre un p-value superiore a 1.
29. Considerando il lancio di una moneta equilibrata e la variabile X che prende valore 1 se esce testa e 0 se croce, e la variabile Y che invece prende valore 0 se esce testa e 1 se croce, le variabili sono di Bernoulli $p = 1/2$ e dunque la loro somma $X + Y$ è di tipo binomiale con $n = 2$ e $p = 1/2$.
30. Una variabile aleatoria X con varianza $\mathrm{Var}(X) = 0$ è tale che esiste un numero $c \in \mathbb{R}$ tale che $\mathbb{P}(X = c) = 1$.
31. Considerando una successione $X_1 = X_2 = X_3 = \cdots$ di variabili uguali tra di loro e con momento secondo finito, per la Legge dei Grandi Numeri la quantità $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ converge in probabilità a $\mathbb{E}[X]$.
32. Una variabile Gaussiana $N(m, \sigma^2)$ ha tutti i momenti finiti.
33. Dato un campione statistico $X_1, X_2, \dots, X_n$ di v.a. indipendenti con momento secondo finito, la funzione $\mathrm{Var}(\bar{X}_n)$ è decrescente in n.
34. Una variabile aleatoria $X : \Omega \to \mathbb{R}$ costante, ovvero $X(\omega) = c$ per ogni $\omega \in \Omega$, è indipendente da qualsiasi altra variabile aleatoria $Y : \Omega \to \mathbb{R}$.
35. Il β-quantile di una variabile aleatoria con densità esponenziale di parametro $\lambda > 0$ è $-\log(1-\beta)/\lambda$.
36. La media campionaria $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ di un campione di variabili aleatorie con momento secondo finito è uno stimatore corretto e consistente del parametro $m = \mathbb{E}[X_i]$.
37. Un intervallo di fiducia ha livello 95% se esso contiene almeno il 95% dei dati del campione.
38. Se X è una v.a. discreta con funzione di massa p_X, la funzione di ripartizione F_X di X è $F_X(x) = \int_{-\infty}^x p_X(y)dy$.

39. Due variabili aleatorie X, Y con momento secondo finito sono scorrelate se e solo se $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)$.
40. Una variabile aleatoria Binomiale (per qualunque scelta dei parametri n, p) possiede tutti i momenti.
41. Il coefficiente di correlazione $\rho(X, Y)$ tra due variabili aleatorie, quando è ben definito, è un numero compreso tra 0 e 1.
42. La densità di Student (con qualsiasi numero di gradi di libertà) è una funzione pari.
43. Un intervallo di fiducia di livello $1-\alpha$ per un parametro θ non contiene il parametro con una probabilità che al massimo è α.
44. La funzione di ripartizione di una variabile aleatoria di Poisson è una funzione continua.
45. Se X è una variabile Gaussiana standard, la variabile aleatoria $Y = \frac{X+|X|}{2|X|}$ è Bernoulli di parametro $p = 1/2$.
46. Una variabile aleatoria con densità chi-quadro non assume mai valori negativi, a prescindere dal numero di gradi di libertà.
47. Per effettuare il test chi-quadro sulla varianza σ^2 di un campione Gaussiano $N(m, \sigma^2)$ è necessario conoscere il valore di m.
48. La regione critica dello Z-test bilatero corrisponde all'evento che un intervallo di fiducia per la media m del campione Gaussiano in considerazione contenga il parametro m.
49. La funzione di ripartizione
$$F(t) = \begin{cases} 0 & t < 0 \\ 1/2 & 0 \le t < 1 \\ 1 & t \ge 1 \end{cases},$$
essendo differenziabile ovunque ad eccezione dei punti $t = 0, 1$, è la funzione di ripartizione di una variabile con densità data da $F'(t)$.
50. Se due variabili aleatorie di Poisson X, Y sono equidistribuite (cioè hanno la stessa legge), hanno lo stesso parametro λ.
51. Data una variabile aleatoria non-negativa X, la successione di variabili aleatorie $X_n = X/n$ converge a 0 in probabilità.
52. La somma di due variabili aleatorie di Bernoulli con lo stesso parametro p è una variabile Binomiale di paramentri $n = 2$ e p.
53. Una variabile aleatoria geometrica di parametro $p \in (0, 1)$ ha tutti i momenti finiti.
54. Per un campione statistico con legge dipendente da un parametro θ, un intervallo di fiducia è un intervallo numerico $[\theta-d, \theta+d]$ attorno al parametro in cui ci si aspetta di trovare gli esiti del campione con probabilità alta.
55. Il prodotto XY di due variabili aleatorie di Bernoulli X e Y è anch'essa una variabile di Bernoulli.

B. Soluzioni dei Quesiti VERO/FALSO

1. FALSO: se A, B sono disgiunti e indipendenti,
$$\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$$
è soddisfatta se e solo se uno dei due eventi è esso stesso vuoto, per cui l'affermazione non è verificata in generale.
2. FALSO: ad esempio una variabile di Poisson può assumere qualsiasi valore naturale $(0, 1, 2, \dots)$ ma non è una variabile con densità.
3. VERO: infatti vale sempre $X^n = X$, perché se $X = 0$ allora $X^n = 0^n = 0$, e similmente se $X = 1$ allora $X^n = 1^n = 1$, di conseguenza
$$\begin{aligned}\mathbb{E}[X^n] &= 1^n\mathbb{P}(X^n = 1) + 0^n\mathbb{P}(X^n = 0) \\ &= 1\mathbb{P}(X = 1) + 0\mathbb{P}(X = 0) = \mathbb{E}[X].\end{aligned}$$
4. FALSO: Ad esempio, per l'esperimento lancio di una moneta equilibrata, e $X = 1$ se esce testa, $X = 0$ se esce croce, il valore atteso di X è $1/2$, ma la media campionaria del campione (testa, testa) è 1.
5. VERO: se A, B sono eventi indipendenti, dalla legge di De Morgan e la definizione di indipendenza si deduce che
$$\begin{aligned}P\left(A^c \cap B^c\right) &= P\left((A \cup B)^c\right) = 1 - P(A \cup B) \\ &= 1 - P(A) - P(B) + P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= (1 - P(A))(1 - P(B)) = P\left(A^c\right) P\left(B^c\right).\end{aligned}$$
6. FALSO: un intervallo di fiducia è dato da una coppia di variabili aleatorie, $I = [a(\omega), b(\omega)]$, e contiene il parametro θ con una certa probabilità, determinata dal livello di fiducia.
7. FALSO: se $f(x)$ è una funzione tale che
$$\mathbb{P}(X = x) = f(x),$$
allora la condizione $\mathbb{P}(\Omega) = 1$ impone che f sia non nulla su al più numerabili valori, e X è dunque una variabile discreta.
8. FALSO: Il test di Student usa la deviazione standard campionaria.
9. VERO: l'indipendenza di X, Y implica che $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, che è il caso con valori attesi nulli, da cui quello generale segue con passaggi elementari.
10. FALSO: per definizione il p-value è la probabilità di esiti di coda della statistica su cui il test si basa, sotto l'ipotesi nulla. NON si può determinare "la probabilità che l'ipotesi nulla sia vera" nell'approccio frequentista alla statistica.
11. FALSO: è falsa la parte "solo se", perché esistono variabili discrete con infiniti esiti di probabilità positiva, ad esempio le variabili di Poisson.
12. FALSO: La probabilità cercata è $\Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6826$.
13. VERO: il p-value è per definizione il livello α sotto cui, fissati gli esiti dell'esperimento, si accetta H_0).

14. VERO: infatti per linearità del valore atteso, e usando che le variabili del campione hanno lo stesso valore atteso,

$$\mathbb{E}[\frac{1}{n}(X_1 + \cdots + X_n)] = \frac{1}{n}(\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = \mathbb{E}[X_1].$$

15. FALSO: il denominatore è scelto in modo che lo stimatore sia corretto, e quest'ultimo sarebbe consistente anche con denominatore n per la Legge dei Grandi numeri e il fatto che $\frac{n-1}{n} \to 1$ per $n \to \infty$.
16. FALSO: A e B sono indipendenti se e solo se $P(A \mid B) = P(A)$.
17. FALSO: $P(a < X \leq b) = F(b) - F(a)$
18. FALSO: il metodo del momento primo restituisce $2\bar{X}$, mentre il metodo di massima verosimiglianza $\max_i X_i$.
19. VERO: un intervallo di fiducia è dato da una coppia di variabili aleatorie $I = [a(\omega), b(\omega)]$ date da opportune funzioni del campione aleatorio, e contiene il parametro θ con una certa probabilità, determinata dal livello di fiducia.
20. VERO: la prima parte dell'enunciato è la definizione di livello, mentre va ricordato che è la potenza del test a quantificare la probabilità di commettere errori di seconda specie.
21. FALSO: per definizione il p-value è la soglia (in α) oltre cui una regione critica C_α rifiuta l'ipotesi alla luce dei dati dell'esperimento, dunque la forma della regione critica è imprescindibile per determinarlo.
22. VERO: l'ampiezza della regione di accettazione è $2\sigma q_{1-\alpha/2}/\sqrt{n}$, dunque è decrescente nella numerosità n.
23. FALSO: le singole componenti di una scelta uniforme su $\{(0,0),(0,1),(1,0),(1,1)\}$ sono Bernoulli($\frac{1}{2}$) indipendenti.
24. VERO: vale

$$\min_{a,b} \mathbb{E}[(Y - aX - b)^2] = \mathrm{Var}(Y)(1 - \rho(X,Y)^2),$$

per cui quando il minimo a sinistra è nullo deve valere $\rho(X,Y) = 1$.

25. FALSO: ciò è vero —per definizione— solo per le variabili con densità, e per cui la densità f sia zero per $x < 0$; ad esempio non esiste una funzione f che soddisfa la relazione proposta se X è una variabile di Bernoulli.
26. VERO: le condizioni su livello e potenza non sono simmetriche nelle due ipotesi. Ad esempio se si è misurata l'altezza di 100 persone ottenendo media campionaria 175.5cm, supponendo deviazione standard 5cm, consideriamo due test con ipotesi nulle $H_0) m \leq 175$ e $H_0') m > 175$. In entrambi i casi $Z = \sqrt{N}(\bar{X} - m)/\sigma = 1$, e i quantili da confrontare sono $q_{0.95} = 1.64$ e $q_{0.05} = -1.64$. In particolare, in entrambi i casi siamo nella regione critica, quindi scambiare un ipotesi nulla che sta venendo rigettata con la corrispondente alternativa non conduce ad accettazione.
27. VERO: il livello è la probabilità sotto H_0 della regione critica, ed imporlo più basso significa ridurre la regione critica; a sua volta la potenza è la probabilità sotto l'ipotesi alternativa della regione critica, ma se quest'ultima si è ridotta la potenza non può aumentare.
28. FALSO: il p-value è una probabilità e non può mai assumere valori superiori a 1.
29. FALSO: la variabile $X + Y$ prende sempre valore 1.
30. VERO: in effetti $c = \mathbb{E}[X]$ che è finito perché X ammette momento secondo essendo la varianza finita (anzi, nulla), e questo perché dalla disuguaglianza di Chebychev

$$\mathbb{P}(|X - \mathbb{E}[X]| > \delta) = \frac{1}{\delta^2}\,\mathrm{Var}(X) = 0,$$

da cui mandando $\delta \to 0$ si ha $\mathbb{P}(X \neq \mathbb{E}[X]) = 1$.

31. FALSO: non è verificata l'ipotesi di indipendenza, e in questo caso $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) = X_1$.
32. VERO: la condizione di esistenza del momento n-esimo è

$$\int_{-\infty}^{\infty} |t|^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-m)^2/(2\sigma^2)} dt < \infty,$$

che è verificata per ogni n poiché l'esponenziale negativo decresce più velocemente di ogni potenza, e nello specifico

$$|t|^n e^{-(t-m)^2/(2\sigma^2)} \leq e^{-c(t-m)^2/(2\sigma^2)}$$

per $c > 0$ abbastanza piccola.

33. VERO: infatti per indipendenza $\mathrm{Var}(\bar{X}_n) = \mathrm{Var}(X_1)/n$.
34. VERO: l'evento $\{X \in A\}$ è sempre uno tra $\emptyset$ e Ω, ma questi ultimi sono indipendenti da ogni evento $\{Y \in B\}$.
35. VERO: $F(x) = 1 - e^{-\lambda x} = \beta$ se e solo se $x = -\log(1-\beta)/\lambda$.
36. VERO: la correttezza discende dalla linearità del valore atteso, mentre la consistenza dalla Legge Debole dei Grandi Numeri.
37. FALSO: non ci sono legami tra intervallo di fiducia e numero di dati in esso contenuti.
38. FALSO: La funzione di ripartizione è $F_X(x) = \sum_{x_i \leq x} p_X(x_i)$.
39. VERO: infatti $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$.
40. VERO: la v.a. binomiale è limitata, quindi possiede tutti i momenti.
41. FALSO: $\rho(X,Y)$ può prendere valori nell'intervallo $[-1,1]$, ad esempio se $Y = -X$ (e $\mathrm{var}(X) \neq 0$) vale $\rho(X,Y) = -1$.
42. VERO: è immediata conseguenza della relazione $F(t) + F(-t) = 1$. Un modo per mostrarla è ricordare che la densità di Student corrisponde alla densità di una variabile

$$T = \frac{X_0}{\sqrt{X_1^2 + \cdots + X_n^2}}$$

in cui $(X_0, X_1, \ldots, X_n)$ sono Gaussiane standard indipendenti, ma poiché $-X_0$ ha la stessa legge

(Gaussiana standard) di X_0 ed è indipendente dalle altre $X_1, \dots, X_n$, allora $(-X_0, X_1, \dots, X_n)$ ha la stessa legge di $(X_0, X_1, \dots, X_n)$ e quindi

$$-T = \frac{-X_0}{\sqrt{X_1^2 + \cdots + X_n^2}}$$

ha la stessa legge di T.

43. VERO: infatti per definizione di intervallo di fiducia $\mathbb{P}(\theta \in I) \geq 1 - \alpha$, e quindi $\mathbb{P}(\theta \notin I) = 1 - \mathbb{P}(\theta \in I) \leq 1 - (1 - \alpha) = \alpha$.

44. FALSO: la c.d.f. di una variabile discreta è una funzione a salti, con i salti corrispondenti ai punti in cui la funzione di massa non si annulla.

45. VERO: anzitutto osserviamo che siccome $\mathbb{P}(X = 0) = 0$ possiamo ignorare il problema della divisione per 0; detto questo se $X > 0$ vale $Y = 1$ e $\mathbb{P}(X > 0) = 1/2$, mentre se $X > 0$ vale $Y = 0$ e $\mathbb{P}(X < 0) = 1/2$.

46. VERO: la densità chi-quadro è la densità della somma di quadrati di variabili Gaussiane, dunque è un numero non negativo.

47. FALSO: la statistica test è $\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$ e non fa uso del parametro m.

48. FALSO: La regione di accettazione si può riscrivere come segue:

$$\begin{aligned} C^c &= \left\{ |\bar{X}_n - m_0| \leq \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right\} \\ &= \left\{ m_0 \in \left[\bar{X}_n \pm \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right] \right\}, \end{aligned}$$

quindi essa corrisponde all'evento che un intervallo di fiducia *non* contenga il parametro.

49. FALSO: le funzioni di ripartizione di variabili con densità sono continue, mentre la funzione F proposta ha due discontinuità di salto in $t = 0, 1$.

50. VERO: indicando con λ, λ' i parametri di X, Y, per definizione di equidistribuzione, per ogni intero $k \geq 0$,

$$e^{-\lambda} \frac{\lambda^k}{k!} = \mathbb{P}(X = k) = \mathbb{P}(Y = k) = e^{-\lambda'} \frac{(\lambda')^k}{k!},$$

da cui $e^{\lambda - \lambda'} = (\lambda/\lambda')^k$, che per $k = 0$ implica $e^{\lambda - \lambda'} = 1$ e dunque $\lambda = \lambda'$. Alternativamente si può ricordare che se due variabili sono equidistribuite hanno lo stesso valore atteso, ma per una variabile di Poisson quest'ultimo coincide con λ.

51. VERO: Poiché la successione di eventi $A_n = \{X > n\varepsilon\}$ è decrescente (se $n \geq m$ vale $A_n \subseteq A_m$),

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(X > n\varepsilon) &= \lim_{n\to\infty} \mathbb{P}(A_n) \\ &= \mathbb{P}(\cap_{n=1}^{\infty} A_n) = \mathbb{P}(\emptyset) = 0, \end{aligned}$$

(si può giungere alla stessa conclusione dalle proprietà della funzione di ripartizione). Da ciò si ottiene la convergenza cercata, ossia per ogni $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - 0| > \varepsilon) = 0.$$

52. FALSO: ad esempio data X variabile di Bernoulli di parametro p, posto $Y = X$ (ancora Bernoulli di parametro p) si ha che $X + Y = 2X$ assume solo due valori, 0 e 2, ma se fosse Binomiale con $n = 2$ potrebbe assumere anche valore 1.

53. VERO: infatti il momento n-esimo è finito se converge la serie

$$\sum_{k=0}^{\infty} k^n \mathbb{P}(X = k) = p \sum_{k=0}^{\infty} k^n (1-p)^k = p \sum_{k=0}^{\infty} k^n e^{\log(1-p)k},$$

che è convergente perché $\log(1 - p) < 0$ essendo $0 < 1 - p < 1$, e dunque l'esponenziale negativo prevale su k^n.

54. FALSO: il parametro θ non è noto, quindi non può essere usato per definire un intervallo di fiducia, che solitamente viene invece centrato in uno stimatore del parametro.

55. VERO: che siano indipendenti o meno, il prodotto XY può assumere solo valori $0, 1$ perché lo stesso vale per entrambi i fattori, e una variabile aleatoria che assume valori $0, 1$ è di Bernoulli.