

Distribuzioni di Probabilità e Statistica Inferenziale

Massimiliano Ferrara, Francesco Strati

massimiliano.ferrara@unirc.it - francesco.strati@unirc.it

a.a. 2016/2017

Indice

1	RICHIAMI SULLE VARIABILI ALEATORIE	1
1.1	Variabili aleatorie	1
1.2	Variabili aleatorie discrete	3
1.3	Il valore atteso di una v.a. discreta	6
1.3.1	Momenti	7
1.4	Variabili aleatorie continue	8
1.5	Il valore atteso di una v.a. continua	12
1.6	Teoremi limite	13
1.6.1	La legge dei grandi numeri	13
1.6.2	Legge debole e legge forte	14
2	DISTRIBUZIONI DI PROBABILITÀ	17
2.1	Distribuzione uniforme	17
2.2	Distribuzione normale	18
2.2.1	Distribuzione normale standard	21
2.2.2	Teorema del limite centrale	25
2.3	Distribuzione binomiale	29
2.4	Distribuzione di Poisson	32
2.5	Distribuzione esponenziale	33
2.6	Distribuzione geometrica	35
3	L'INFERENZA STATISTICA	37
3.1	Cosa è l'inferenza statistica?	37
3.2	Campionamento	37
3.3	Parametri, Statistiche e Probabilità	38
3.4	Distribuzioni campionarie delle statistiche	39

3.5	Problemi Inferenziali	41
3.6	Stima puntuale dei parametri	41
3.6.1	Stimatore non distorto	42
3.6.2	Stimatore consistente	42
3.6.3	Stimatore efficiente	43
3.6.4	Il metodo della massima verosimiglianza	43
3.7	Stima intervallare di un parametro	43
3.7.1	Il margine di errore	46
3.8	Verifica di Ipotesi	47
3.8.1	Il p-value	50
4	INFERENZA SU MEDIE	53
4.1	Intervallo di confidenza per la media della popolazione	54
4.1.1	a) Varianza nota	54
4.1.2	b) Varianza incognita	56
4.2	Dimensionamento del Campione	57
4.3	Verifica delle ipotesi sul valore di μ	59
4.3.1	a) Varianza nota	59
4.3.2	a) Varianza incognita	62
	Appendices	65
	Appendice	65
C	Probabilità: Assiomi	65

Capitolo 1

RICHIAMI SULLE VARIABILI ALEATORIE

1.1 Variabili aleatorie

Una variabile aleatoria (o casuale) è una funzione numerica che associa valori numerici in corrispondenza ai risultati di un esperimento aleatorio. In genere, indicheremo con la lettera maiuscola, es. X , la variabile aleatoria, mentre indicheremo con la lettera minuscola, es. x , una sua realizzazione. Indicando con Ω lo spazio degli eventi, si consideri il lancio di una moneta non truccata dalla quale possono verificarsi due eventi: testa (T) e croce (C). In questo caso $\Omega = \{T, C\}$. Definiamo con A l'insieme di tutti i sottoinsiemi di Ω e con $P(T) = 1/2$, $P(C) = 1/2$, le probabilità che assumano un valore determinato. Si supponga che noi abbiamo pagato 100 euro per una scommessa: se esce testa T vinciamo 150 euro, se esce croce C ne vinciamo 50. Il guadagno G è definito come $G(T) = 150 - 100 = 50$ e $G(C) = 50 - 100 = -50$. In questo senso la variabile G è aleatoria in quanto assume un valore preciso solo dopo che l'esperimento aleatorio sia stato effettuato.

Una **variabile aleatoria discreta** assume al più un insieme numerabile di valori.

La v.a. discreta può assumere solo un numero finito di valori. Una v.a. è discreta anche se i possibili risultati sono infiniti, ma devono essere numerabili.

Una **variabile aleatoria continua** può assumere un qualunque valore in un intervallo.

Si pensi alla temperatura che è sempre misurata in un intervallo continuo. In particolare le realizzazioni sono una infinità più che numerabile alla quale non può logicamente essere attribuito un singolo valore.

Ulteriore discussione

Un insieme è definito *numerabile* se e solo se i suoi elementi possono essere messi in corrispondenza biunivoca con quelli appartenenti all'insieme dei numeri naturali \mathbb{N} .

L'insieme dei numeri pari P è numerabile, cioè

P	\mathbb{N}
0	0
2	1
4	2
6	3
\vdots	\vdots
$2n$	n

dove ad ogni elemento di $n \in \mathbb{N}$ si associa il suo doppio $2n \in P$. La funzione è molto semplice, $f: \mathbb{N} \rightarrow P$ tale che $f(x) = 2x$. Con la stessa procedura si capisce che anche l'insieme dei numeri dispari è numerabile in quanto $n \leftrightarrow 2n + 1$, in cui ogni numero non negativo $n \geq 0$ è associato a un numero dispari. L'insieme $2, 4, 8, \dots, 2^n$ di potenza 2 è numerabile, in quanto $n \leftrightarrow 2^n$, e così via.

Gli insiemi non numerabili sono un po' più difficili da testare, in questa sede basti sapere che l'insieme dei punti sulla retta reale \mathbb{R} è non numerabile. Si noti che questo deriva dal fatto che l'intervallo $[0, 1]$ è non numerabile, cosa che può essere provata tramite il procedimento cardinale di Cantor.

Ovviamente si possono avere insiemi *finiti*, come ad esempio gli studenti di una scuola. Per capire la cardinalità di questi insiemi, basta confrontarli fra loro. Ad esempio: l'insieme delle sedie dell'auditorium della scuola hanno la stessa cardinalità dell'insieme degli studenti se nessuno di loro rimarrà alzato o se nessuna sedia rimarrà libera. Se, per assurdo, il numero di sedie e studenti è infinito, allora non si potranno confrontare in maniera esplicita. In questo caso si deve testare se vi sia o meno una corrispondenza biunivoca fra i due insiemi. Negli insiemi non numerabili, questa corrispondenza non c'è. Per farsi una idea degli insiemi *infiniti*, rimuoviamo un elemento da ogni insieme, poi un altro, un altro e un altro ancora all'infinito: ogni volta rimarranno sempre degli elementi lasciati nell'insieme.

1.2 Variabili aleatorie discrete

Come al solito, definiamo X come la variabile aleatoria discreta e x una sua realizzazione. La probabilità che X assuma il valore x è definita dalla probabilità $P(x) = P(X = x)$. La distribuzione di tutti i valori che può assumere la variabile è chiamata *distribuzione di probabilità*. In generale

$$- 0 \leq P(x) \leq 1 \quad \forall x;$$

$$- \sum_x P(x) = 1.$$

Questo significa che le probabilità di accadimento di ogni singolo evento non possono essere negative, né superare 1, visto che uno dei possibili risultati dell'esperimento aleatorio si verificherà certamente. Inoltre gli eventi $X = x$ devono essere mutuamente esclusivi e collettivamente esaustivi. Gli eventi sono mutuamente esclusivi quando non possono essere contemporaneamente veri $X_i \cap X_j = \emptyset$ per ogni i, j . Sono collettivamente esaustivi perché, come detto, almeno uno deve essere vero $X_i \cup X_j = \Omega$.

La distribuzione di una probabilità di una variabile aleatoria può essere descritta tramite la c.f *funzione di ripartizione*

Una **funzione di ripartizione** (o funzione cumulativa delle frequenze) per una variabile aleatoria X , è la funzione definita da

$$F(x_0) = P(X \leq x_0), \quad (1.1)$$

dove $F(x_0)$ esprime la probabilità che X non superi il valore x_0 .

Dicendo che $F(x_0) = 0.8$, asseriamo che l'80% della massa unitaria di probabilità che si riferisce a Ω , riguarda realizzazioni della variabile aleatoria non superiori a x_0 .

Esempio

Il proprietario di un negozio di scarpe, osservando le vendite precedenti, sa che ogni giorno il numero di scarpe vendute di una certa marca, varia da 0 a 5. Il proprietario vuole capire come pianificare le scorte.

X	$P(x)$	$F(x)$
0	0.10	0.10
1	0.20	0.30
2	0.25	0.55
3	0.25	0.80
4	0.15	0.95
5	0.05	1.00

Tenendo in negozio 4 paia di scarpe di quella marca, si potrà soddisfare il 95% di domanda. Se il proprietario opererà per 3 paia di scarpe, si soddisferà l'80% e così via.

La relazione tra la funzione di ripartizione e la funzione di probabilità, è definita come segue

$$F(x_0) = \sum_{x \leq x_0} P(x).$$

La funzione di ripartizione varia al variare di x , ma solo nei punti dove x è la realizzazione della v.a. X . In particolare, seguendo gli assiomi della probabilità richiamati nella sezione C, sappiamo che l'unione di eventi mutuamente esclusivi è uguale alla sommatoria di questi eventi. In questo caso l'unione è dei possibili valori di x minori o uguali a x_0 .

La funzione di ripartizione presenta importanti proprietà:

1. $0 \leq F(x) \leq 1$;
2. $F(x_1) \leq F(x_2)$ se $x_1 < x_2$;
3. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$;
4. $\lim_{x \rightarrow x_0^+} F(x) = F(x_0) = F(x_0)$ dove $x_0 = \lim_{\epsilon \rightarrow 0^+} x_0 + \epsilon$;
5. $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$, da cui la funzione di ripartizione consente di stabilire la probabilità che la v.a. X assuma valori nell'intervallo $(x_1, x_2]$ per $x_1 < x_2$ con $x_1, x_2 \in \mathbb{R}$.

La proprietà 1 ci dice semplicemente che la probabilità non può essere maggiore di uno né minore di zero. La proprietà 2 afferma che la funzione di ripartizione è monotona non decrescente. Se $x_1 < x_2$, allora $(X \leq x_1) \subset (X \leq x_2)$. Dalle proprietà della probabilità (in particolare la 3) ne

segue che $P(X \leq x_1) \leq P(X \leq x_2)$, cioè $F(x_1) \leq F(x_2)$. La proprietà 3 deriva dall'assioma 2 e dalla proprietà della probabilità 2

$$\lim_{x \rightarrow \infty} P(X = x) = P(X \leq \infty) = P(\Omega) = 1,$$

$$\lim_{x \rightarrow -\infty} P(X = x) = P(X \leq -\infty) = P(\emptyset) = 0.$$

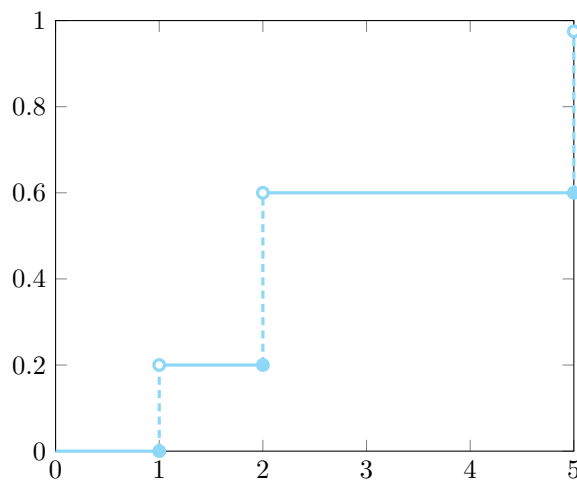
La proprietà 4 ci informa che la funzione di ripartizione è continua a destra. Si noti come

$$\lim_{x \rightarrow x_0^-} F(x) \neq F(x_0),$$

da cui deriva che la funzione di ripartizione presenta dei punti di discontinuità di prima specie (o salto). Il salto può essere descritto in questo modo

$$\lim_{x \rightarrow x_0^+} F(x) - \lim_{x \rightarrow x_0^-} F(x) = F(x_0) - F(x_0^-) = P(X = x_0),$$

da cui deriva il fatto che la funzione di ripartizione cambia solo nei punti in cui x_0 rappresentano realizzazioni di X . La seguente è la figura di una funzione di ripartizione discreta



Questa è una funzione costante a tratti e non decrescente, ad ogni punto x_i con $i = 1, 2, 5$, corrisponde un salto $P(x_i)$ esplicitato dalla linea tratteggiata. Alle ordinate troviamo la misurazione della probabilità, infatti l'ammontare del salto è la distanza verticale fra il tondino vuoto e quello pieno. Si noti come il salto, ad esempio al punto 2 del grafico, è descritto dai limiti

$$P(X = 2) = f(2^+) - F(2^-) = 0.4$$

1.3 Il valore atteso di una v.a. discreta

La misura di tendenza centrale di una variabile aleatoria è chiamata valore atteso (come la media campionaria per i dati di un campione).

Il **valore atteso**, $\mathbb{E}(X)$, di una v.a. discreta X è la media pesata di tutti i suoi possibili valori x , dove i pesi sono le corrispondenti probabilità. Formalmente

$$\mathbb{E}(X) = \sum_x xP(x)$$

Una impresa produttrice di biscotti rileva che il 90% dei pacchi venduti non conteneva biscotti con tracce di nocciole, l' 8% ne conteneva uno, mentre il 2% due biscotti. Fare la media degli errori non tiene conto del fatto che il 90% dei pacchi venduti non presenta nessun biscotto con tracce di nocciole. Infatti

$$P(0) = 0.9 \qquad P(1) = 0.08 \qquad P(2) = 0.02$$

avrebbe un valore centrale di 0.33 biscotti con tracce di nocciola per pacco. Il numero medio, o valore atteso, sarà invece

$$\mathbb{E}(X) = \sum_x xP(x) = (0)(0.90) + 1(0.08) + 2(0.02) = 0.12$$

da cui ci aspetteremo una media di 0.12 biscotti con tracce di nocciole per pacco.

Proprietà del valore atteso:

P1 Se $X = c$ dove c è una costante, allora

$$\mathbb{E}(X) = c;$$

P2 Se $Y = aX + b$, allora

$$\mathbb{E}(Y) = a\mathbb{E}(X) + b;$$

P3 Se $Y = a_1X_1 + a_2X_2 + \dots + a_nX_n + b$, allora

$$\mathbb{E}(Y) = a_1\mathbb{E}(X_1) + a_2\mathbb{E}(X_2) + \dots + a_n\mathbb{E}(X_n)$$

P4 Se X e Y sono indipendenti, allora $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Si noti che la proprietà *P2* è sempre valida solo per funzioni lineari della v.a. X . Non è così per funzioni non lineari in cui si usa la disuguaglianza di Jensen (vedi la sezione ?).

1.3.1 Momenti

Di una certa rilevanza è la trattazione delle funzioni di variabili aleatorie $Y = f(X)$. Dove Y è ancora una variabile aleatoria e $X : \Omega \rightarrow \mathbb{R}$. Dato che $f : \mathbb{R} \rightarrow \mathbb{R}$, la composizione di f e X è ancora una funzione da Ω alla retta reale

$$Y = f \circ X : \Omega \rightarrow \mathbb{R}$$

Tra le funzioni di variabili aleatorie, sono molto importanti i *momenti*.

Sia X una variabile aleatoria. Si consideri inoltre un numero naturale $k = 1, 2, \dots$ tramite il quale X viene considerata alla sua k -esima potenza X^k .

Supponendo l'esistenza del valore atteso di X^k , si chiama **momento** di ordine k di X rispettivamente all'origine

$$\mu_k = \mathbb{E}(X^k).$$

Inoltre, dato un valore θ , supponendo l'esistenza del valore atteso di $(X - \theta)^k$, allora questo è chiamato **momento** di ordine k di X rispettivamente al polo θ

$$\mu_k^\theta = \mathbb{E}[(X - \theta)^k].$$

Se $\theta = \mu_1 = \mathbb{E}(X)$ allora si parla di momenti centrali

$$\bar{\mu}_k = \mathbb{E}\{[X - \mathbb{E}(X)]^k\} = \mathbb{E}\left[(X - \mu_1)^k\right].$$

Si noti come la media è una misura della *posizione*. La *varianza* è invece una misura della *dispersione* da quella posizione, è il momento centrale di ordine $k = 2$.

La **varianza** si denota come $\text{Var}(X)$ o $\sigma^2(X)$:

$$\text{Var}(X) = \bar{\mu}_k = \mathbb{E}\left[(X - \mu_1)^2\right] = \mathbb{E}\{[X - \mathbb{E}(X)]^2\} = \sigma^2(X).$$

La radice quadrata della varianza

$$\sigma(X) = \sqrt{\mathbb{E}\left[(X - \mu_1)^2\right]}$$

viene chiamata **deviazione standard** della variabile aleatoria.

Le proprietà della varianza:

P1 $\text{Var}(X) = 0$ se e soltanto se $P(X = c) = 1$ per qualche costante c ;

P2 Se $Y = aX + b$, allora $\text{Var}(Y) = a^2\text{Var}(X)$;

$$P3 \quad \text{Var}(X) = \text{Var}(X^2) - \text{Var}(X)^2;$$

P4 Per v.a. indipendenti X_1, X_2, \dots, X_n , se $Y = a_1X_1 + a_2X_2 + \dots + a_nX_n + b$, allora

$$\text{Var}(Y) = a_1^2\text{Var}(X_1) + a_2^2\text{Var}(X_2) + \dots + a_n^2\text{Var}(X_n).$$

Da notare che la proprietà P3 della varianza ci dice che questa è uguale alla differenza fra il momento secondo rispetto a 0 e il quadrato del valore atteso.

La varianza è molto importante quando si vuole confrontare la variabilità di diverse distribuzioni di probabilità. Se due investimenti hanno un rendimento atteso uguale, ma variabilità diversa, allora una varianza più alta significa che in media sono più probabili rendimenti che si discostano dalla media, rispetto all'investimento con variabilità più bassa. Si può già intuire come la varianza sia collegata al rischio.

1.4 Variabili aleatorie continue

Una **funzione di ripartizione** (o funzione cumulativa delle frequenze) per una variabile aleatoria continua X , è la funzione definita da

$$F(x_0) = P(X \leq x_0), \quad -\infty < x_0 < +\infty \quad (1.2)$$

dove $F(x_0)$ esprime la probabilità che X non superi il valore x_0 .

Si supponga ci sia una vasca che può essere riempita contenendo un massimo di 1000 litri. Questa vasca può essere riempita e svuotata tutti i giorni. Dalla analisi storica, la variabile aleatoria X può quindi assumere qualsiasi valore che si trovi nell'intervallo $[0, 1000]$ con uguale probabilità. Ne segue che

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.001x & 0 \leq x < 1000 \\ 1 & x \geq 1000 \end{cases}$$

Quindi la probabilità di venire riempita in un intervallo tra 0 e 500 litri è

$$P(X \leq 500) = F(500) = (0.001)(500) = 0.50.$$

Dalla proprietà 5 della funzione di ripartizione, sappiamo che la probabilità che X assuma valori in un certo intervallo è data dalla differenza tra i valori di $F(x)$ nel limite superiore e inferiore

dell'intervallo. Infatti per due possibili valori di X , $a < b$, la probabilità che X assuma valori fra a e b è

$$P(a < X < b) = F(b) - F(a). \quad (1.3)$$

Così, la probabilità di riempire oggi la vasca in un intervallo fra 400 e 800 litri è

$$P(400 < X < 800) = 0.001(800 - 400) = 0.8 - 0.4 = 0.40$$

Anche qui, la funzione di ripartizione contiene tutte le informazioni necessarie sulla struttura probabilistica della variabile aleatoria. Il legame tra funzioni di ripartizione per variabili aleatorie discrete, e la funzione di probabilità che esprime la probabilità che la v.a. discreta assuma un determinato valore, è descritta in termini dei salti $P(X_i)$. Nel caso di v.a. continue, la funzione di probabilità viene chiamata **funzione di densità di probabilità**. Visto che la probabilità di assumere un determinato valore nel caso continuo è zero, allora la funzione di densità esprime la probabilità che X appartenga a un determinato intervallo invece che assuma uno specifico valore. La relazione che intercorre tra funzione di ripartizione e funzione di densità è che quest'ultima è la derivata in x (la realizzazione) della funzione di ripartizione. Di contro, la funzione di ripartizione è l'integrale da $-\infty$ a x della funzione di densità.

La **funzione di densità di probabilità** $f(x)$ di una v.a. continua X dove x è un punto appartenente a un intervallo di valori che X può assumere, gode di alcune proprietà

1. $f(x) > 0$ per qualunque x ammissibile nell'intervallo e $f(x) = 0$ altrove;
2. si ha che

$$\int_{-\infty}^{\infty} f(x)dx = 1;$$

3. per due possibili realizzazioni di X , definite come x_1 e x_2 dove $x_1 < x_2$, la probabilità che X assuma valori entro l'area sottesa all'intervallo $[x_1, x_2]$ è definita come

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx;$$

4. la funzione di ripartizione $F(x)$ è l'area sottesa alla funzione di densità di probabilità sino ad un determinato punto x_0

$$F(x_0) = \int_{-\infty}^{x_0} f(x)dx.$$

Dal teorema fondamentale del calcolo integrale abbiamo inoltre che

$$F'(x) = \frac{d}{dx}F(x) = f(x).$$

Nelle immagini che seguono è ben descritto il fondamentale rapporto tra la funzione di ripartizione e la funzione di densità della probabilità. In particolare si vede graficamente, per quanto non rigorosamente e con qualche grado di imprecisione, come la derivata al punto x_2 della funzione di ripartizione $F(x)$ ci aiuta a conoscere la funzione di densità $f(x)$ fino a quel punto x_2 .

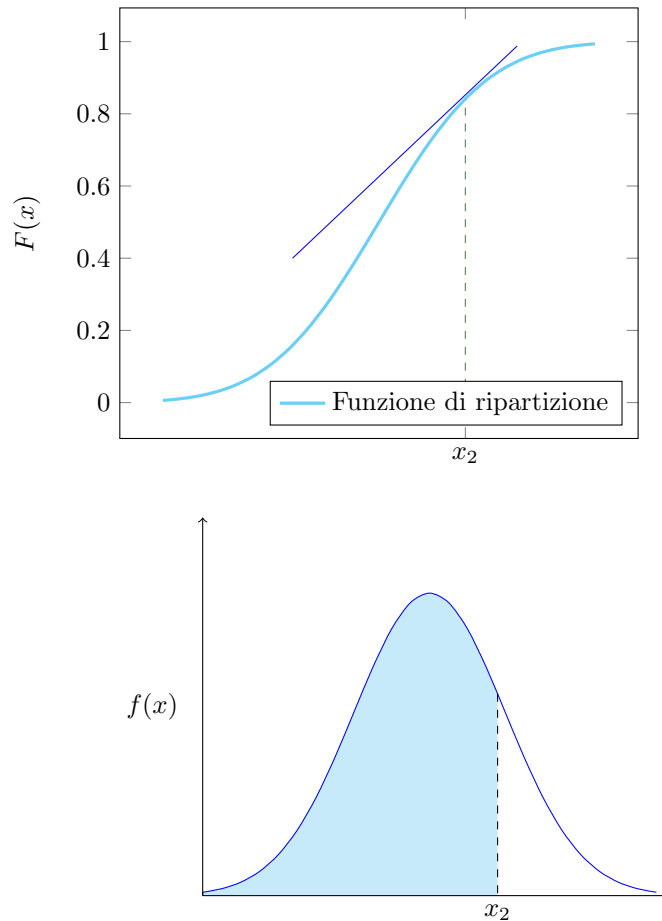


Figura 1.1: Rapporto tra $F(x)$ e $f(x)$.

Esempio Provare se la seguente è una funzione di ripartizione.

Sia

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{1+x} & x \geq 0 \end{cases}$$

Dalla definizione di funzione di ripartizione, questa deve soddisfare almeno queste tre proprietà

1. $\lim_{x \rightarrow -\infty} \frac{x}{1+x} = 0$
2. $\lim_{x \rightarrow \infty} \frac{x}{1+x} = 1$
3. $F(x)$ è non decrescente

$$f(x) = F'(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{(1+x)^2} & x \geq 0 \end{cases}$$

Vediamo quindi se $f(x)$ è una funzione di densità della probabilità. Dalle sue proprietà, $f(x) \geq 0$ e deve contenere l'evento che si concretizzerà

$$\int_{-\infty}^{\infty} f(t) dt = \lim_{x \rightarrow \infty} F(x) - \lim_{x \rightarrow -\infty} F(x) = 1 - 0 = 1$$

È dunque una funzione di ripartizione.

Esempio 2 Provare se la seguente è una funzione di ripartizione.

Sia

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

Anche qui

1. $\lim_{x \rightarrow -\infty} 1 - e^{-x} = 0$
2. $\lim_{x \rightarrow \infty} 1 - e^{-x} = 1$
3. $F(x)$ è non decrescente

$$f(x) = F'(x) = \begin{cases} 0 & x < 0 \\ e^{-x} & x \geq 0 \end{cases}$$

da cui deriva che $f(x)$ è la funzione di densità della probabilità.

$F(x)$ è una funzione di ripartizione.

1.5 Il valore atteso di una v.a. continua

Considerando quello che si è già detto nel caso della variabile casuale discreta, iniziamo con la definizione formale di valore atteso quando la v.a. è continua.

Se si effettuano n ripetizioni indipendenti in un esperimento aleatorio, il **valore atteso**, $\mathbb{E}(X)$, di una v.a. continua X è la media ponderata dei valori x assunti quando il numero delle ripetizioni tende a infinito. Formalmente

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

Se $g(x)$ è una qualunque funzione della variabile aleatoria X , allora

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

Sappiamo che la media e la varianza sono due importanti misure di sintesi di una distribuzione di probabilità.

La **media** di X è definita come il suo valore atteso

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

Vedremo che questo è vero grazie alla legge *forte* dei grandi numeri.

La **varianza** σ_X^2 , è definita come

$$\text{Var}(X) = \mathbb{E}[(X - \mu_X)^2] = \int_{\mathbb{R}} (x - \mu_X)^2 f(x)dx.$$

Esattamente come nel caso discreto, la varianza esiste se e solo se μ_X e $\mathbb{E}(X^2)$ esistono

$$\text{Var}(X) = \mathbb{E}(X^2) - \mu_X^2 = \int_{\mathbb{R}} x^2 f(x)dx - \mu^2.$$

La **deviazione standard** è definita come $\sigma = \sqrt{\text{Var}(X)}$

Esempio X è una v.a. dalla quale

$$f(x) = \begin{cases} \frac{1}{2}x & 0 \leq x \leq 2 \\ 0 & \text{altrove.} \end{cases}$$

Il valore atteso è definito come

$$\mathbb{E}(X) = \int_{\mathbb{R}} xf(x)dx = \int_0^2 \frac{1}{2}x^2 dx = \left[\frac{x^3}{6} \right]_0^2 = \frac{4}{3}.$$

Tenendo presente i vari modi in cui la varianza può essere calcolata, dalle informazioni disponibili sappiamo che

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 f(x) dx = \int_0^2 \frac{1}{2} x^3 dx = \left[\frac{x^4}{8} \right]_0^2 = 2$$

da cui si ottiene la varianza

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 2 - \frac{16}{9} = \frac{2}{9}$$

e la deviazione standard

$$\sigma_X = \sqrt{\frac{2}{9}} = \frac{1}{3}\sqrt{2}.$$

1.6 Teoremi limite

Di grande rilevanza nella teoria della probabilità sono i *teoremi limite*. I più importanti sono quelli afferenti alle leggi dei grandi numeri e i teoremi del limite centrale (che vedremo nella sezione 2.2.2). Sono classificati “leggi dei grandi numeri” tutti quei teoremi per i quali una sequenza di v.a. converge in media al valore atteso. Mentre i teoremi del limite centrale si hanno quando un grande numero di v.a. ha una distribuzione di probabilità che si approssima a quella normale.

1.6.1 La legge dei grandi numeri

Prima di addentrarci nella discussione dei teoremi principali, abbiamo bisogno di conoscere delle importanti disuguaglianze che ci permettono di individuare la convergenza in probabilità quando soltanto la media, oppure anche la varianza sono conosciuti.

Disuguaglianza di Markov

Sia X una v.a. a valori non negativi, con qualsiasi valore $a > 0$, allora

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Dimostrazione. Dalla definizione $X : \Omega \rightarrow \mathbb{R}_+$ dove Ω è, come al solito, lo spazio campionario.

Per ogni $a > 0$ si ha che

$$I = \begin{cases} 1 & X \geq a \\ 0 & \text{altrimenti} \end{cases}$$

da cui deriva l'intervallo $0 \leq I \leq X/a$. Il valore atteso della v.a. I è

$$\mathbb{E}(I) = 0 \cdot P(I = 0) + 1 \cdot P(I = 1) = P(I = 1).$$

Si ha quindi che

$$\mathbb{E}(I) \leq \frac{\mathbb{E}(X)}{a}$$

dato che il valore atteso di I è uguale alla probabilità che il valore atteso di X sia maggiore di a , allora $\mathbb{E}(I) = P(X \geq a)$ che conclude la prova. \square

Disuguaglianza di Chebyshev

Sia X una v.a. con media μ e varianza σ^2 , allora per qualsiasi valore k

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Dimostrazione. Si associ a $X : \Omega \rightarrow \mathbb{R}$, una variabile aleatoria non negativa $Y = (X - \mu)^2$. Si può così applicare la disuguaglianza di Markov con $a = k^2$, si ottiene così

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{\mathbb{E}(X - \mu)^2}{k^2}$$

ma visto che $(X - \mu)^2 \geq k^2$ se e solo se $|X - \mu| \geq k$, ne segue che

$$P\{|X - \mu| \geq k\} \leq \frac{\mathbb{E}(X - \mu)^2}{k^2} = \frac{\sigma^2}{k^2}$$

completando la prova. \square

1.6.2 Legge debole e legge forte

Legge debole dei grandi numeri

Sia X_1, X_2, \dots una sequenza di v.a. indipendenti e distribuite identicamente (*iid*, aventi ognuna la stessa media finita $\mathbb{E}(X_i) = \mu$). Per qualsiasi $\epsilon > 0$

$$P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \epsilon\right\} \rightarrow 0 \text{ come } n \rightarrow \infty$$

Dimostrazione. Assumendo una varianza finita si ha che

$$\mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \mu,$$

e che

$$\text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma^2}{n}.$$

Dalla diseguaglianza di Chebyshev si ottiene che

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2},$$

e il risultato è dimostrato. \square

Legge forte dei grandi numeri

Sia X_1, X_2, \dots una sequenza iid di v.a., aventi ognuna la stessa media finita $\mathbb{E}(X_i) = \mu$.

Allora **con probabilità 1** si ha che

$$\frac{X_1 + \cdots + X_n}{n} \rightarrow \mu \text{ come } n \rightarrow \infty.$$

Quel “con probabilità 1” è da intendersi

$$P\left(\lim_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1.$$

La prova di questo teorema è più complicata; non la dimostreremo in questa sede. È però importante capire bene la differenza fra la legge debole e la legge forte dei grandi numeri. La legge debole dice che per ogni valore specificato \hat{n} , allora $(X_1 + \cdots + X_{\hat{n}}/\hat{n})$ è limitato nell’intorno di μ . Questo teorema quindi non ci dice nulla sui valori $n > \hat{n}$. Infatti rimane aperta la possibilità che $|(X_1 + \cdots + X_n/n) - \mu| > \epsilon$ per un numero *infinito* di volte. Qui entra in gioco la legge forte dei grandi numeri che impedisce che questo accada. Questa implica che

$$\left|\sum_{i=1}^n \frac{X_i}{n} - \mu\right|$$

sia maggiore di ϵ solo per un numero *finito* di volte.

Capitolo 2

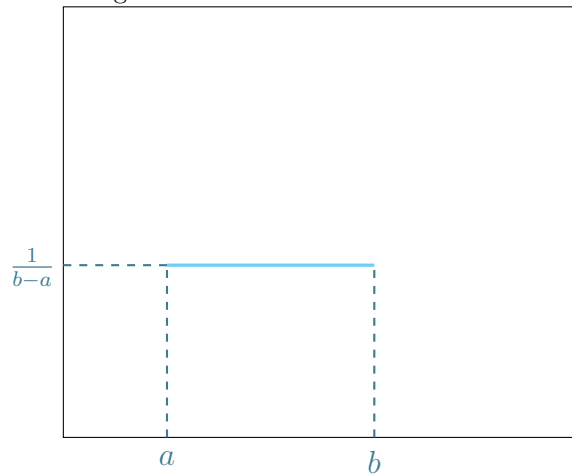
DISTRIBUZIONI DI PROBABILITÀ

2.1 Distribuzione uniforme

Una v.a. X si definisce come *uniforme* su un intervallo (a, b) , assumendo che due sottointervalli della stessa lunghezza hanno uguale probabilità, se la sua funzione di densità di probabilità è data da

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

si presenta graficamente come segue



Visto che $F(x) = \int_{-\infty}^x f(x)dx$, allora la funzione di ripartizione di X è definita come

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

Il valore costante della funzione di densità in (a, b) è dovuto alla proprietà della normalizzazione

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_a^b \frac{1}{b-a} dx.$$

La varianza e la media sono definite come segue

$$\mu = \mathbb{E}(X) = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}.$$

2.2 Distribuzione normale

La distribuzione di probabilità normale è la variabile aleatoria continua più utilizzata nelle applicazioni economiche e non. Ad esempio sarà la distribuzione fondamentale per studiare la teoria delle decisioni. Inoltre, si pensi al fatto che la normale approssima le distribuzioni di probabilità di un grande numero di variabili aleatorie. Nel capitolo 3 si studieranno le medie campionarie, e si vedrà che queste medie, per grandi campioni, approssimano la distribuzione normale. Inoltre questo tipo di distribuzione è semplice da gestire nei calcoli.

Vediamo subito la funzione di densità (vedi figura 2.2a).

La **funzione di densità** della variabile aleatoria normale X è definita come

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{per } -\infty < x < +\infty.$$

Si noti che $0 < \sigma^2 < \infty$; $-\infty < \mu < +\infty$. Inoltre, vi sono due costanti note $\pi \approx 3.141$ ed $e \approx 2.718$.

È importante notare subito che la distribuzione di probabilità normale rappresenta diverse funzioni di densità normali al variare di μ e σ come si vede nella figura 2.2b,c.

Come si nota, una diversa media o la varianza hanno effetti diversi sulla forma della funzione di densità. Variando la media, la forma rimane praticamente identica (figura 2.2b), l'effetto è invece quello di traslare l'intera curva (da μ_1 a μ_2 in fig. 2.2b). Quando vi sono varianze diverse e medie uguali, la varianza più alta, nella figura 2.2c σ_2^1 , determina una distribuzione più ampia. Si noti che ognuna è di forma campanulare e simmetrica, centrata rispetto alla media comune. In generale, quando una variabile casuale X segue una distribuzione normale con parametri μ e σ^2 , si ha che la media della v.a. è $\mathbb{E}(X) = \mu$, la varianza è $\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \sigma^2$, e si identifica con la notazione $X \sim \mathcal{N}(\mu, \sigma^2)$.

Come abbiamo visto per le variabili aleatorie in generale, per calcolare le probabilità associate ad

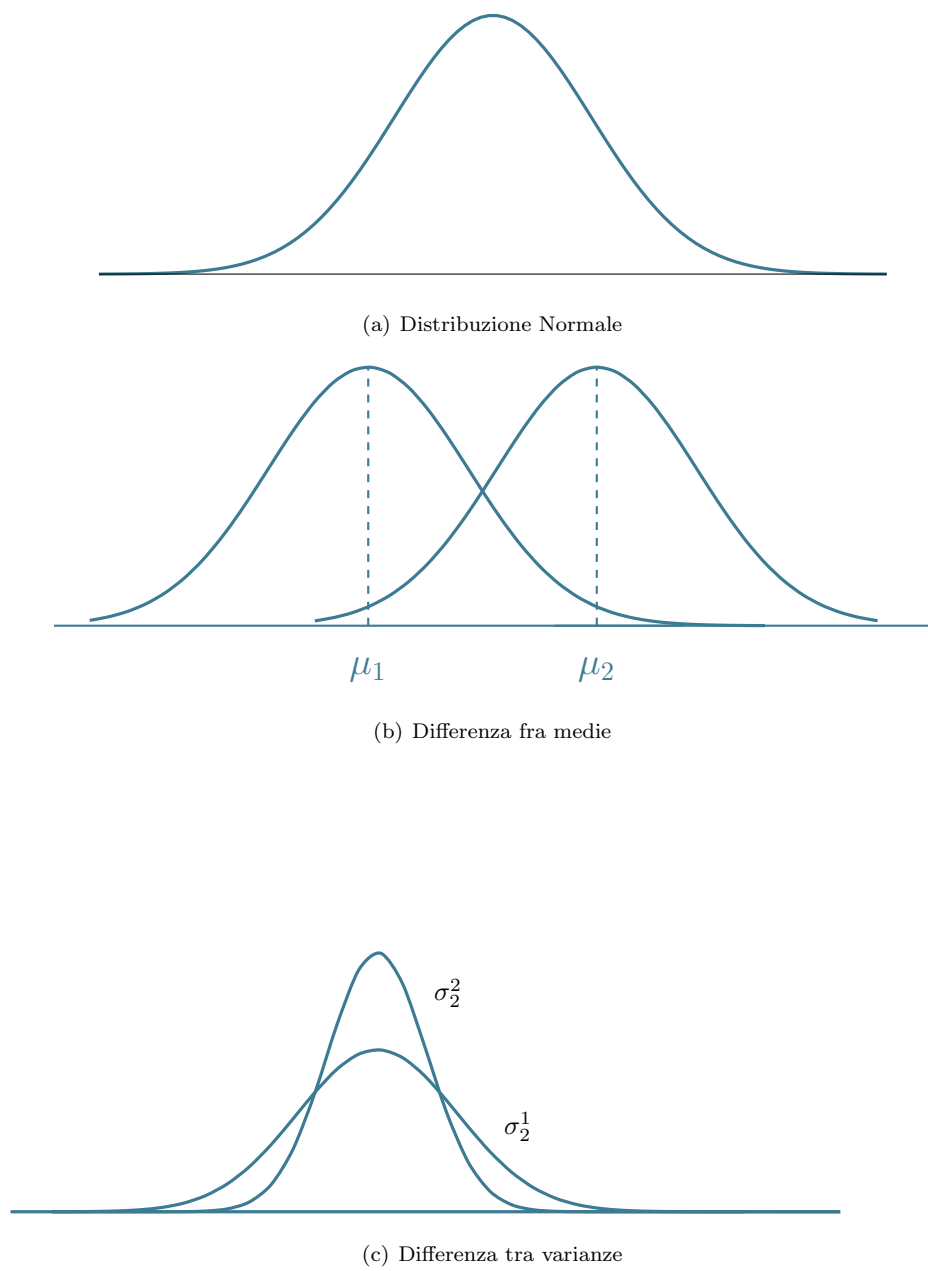
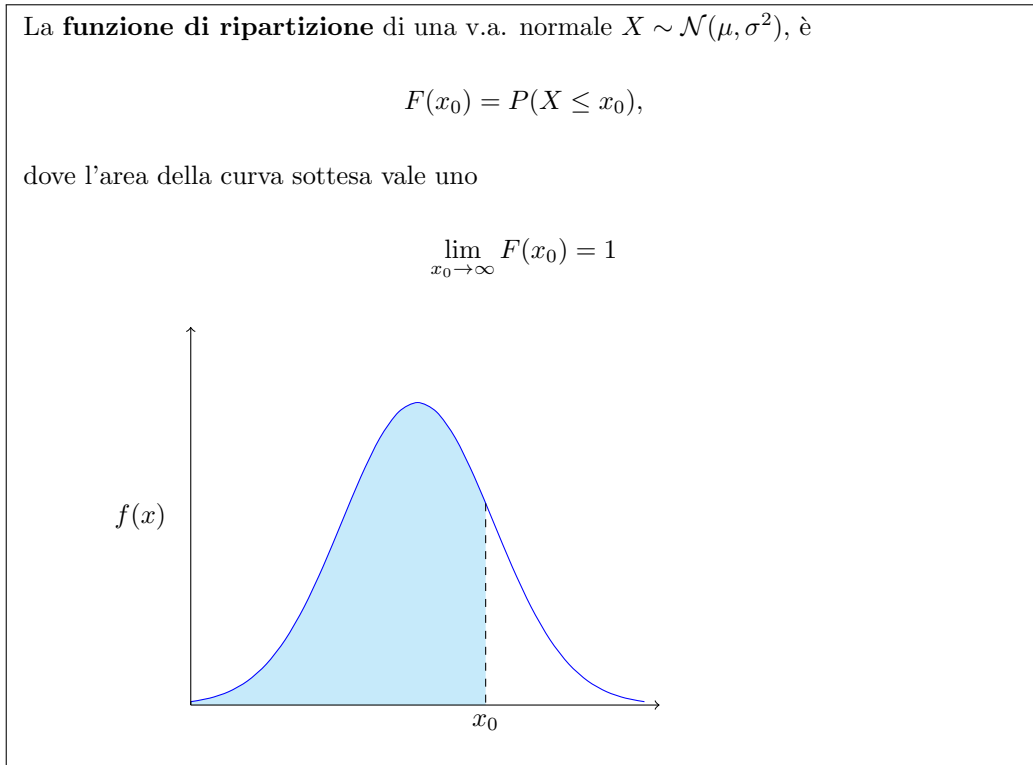
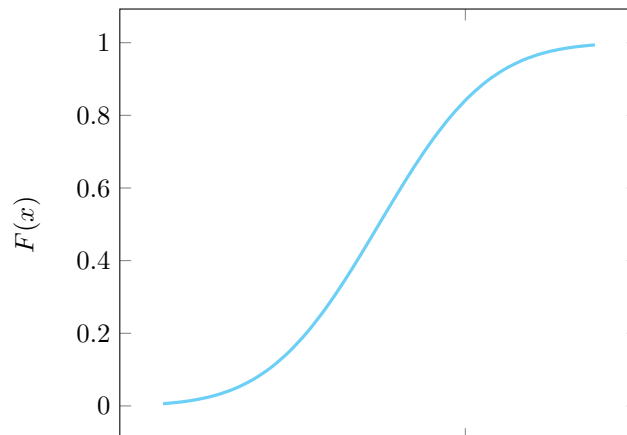


Figura 2.1: Distribuzioni di probabilità normale

una generica distribuzione normale si deve costruire una *funzione di ripartizione*. Come si nota dalla figura 1.1, abbiamo definito il rapporto tra funzione di ripartizione e funzione di densità, ancora non era chiaro, ma nella figura 1.1 abbiamo considerato una funzione di ripartizione di una variabile $X \sim \mathcal{N}(\mu, \sigma^2)$.



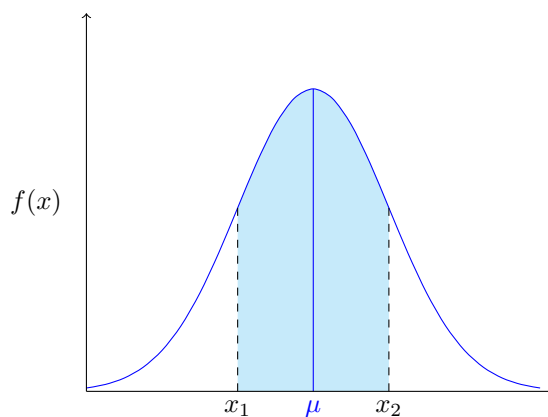
La forma generale di una funzione di ripartizione della variabile normale è quella già vista, cioè



Come abbiamo visto in generale nella equazione (1.3), data questa volta una variabile aleatoria normale $X \sim \mathcal{N}(\mu, \sigma^2)$ con funzione di ripartizione $F(x)$, e definendo due punti x_1 e x_2 , con $x_1 < x_2$ come due possibili realizzazioni di X , allora

$$P(x_1 < X < x_2) = F(x_2) - F(x_1).$$

La probabilità da definire è quindi quell'area sottesa alla densità di probabilità fra x_1 e x_2 , graficamente significa



Nella pratica risulta spesso complicato calcolare la probabilità relativa a un intervallo ottenuto tramite la funzione di ripartizione, per una generica distribuzione normale con media e varianza assegnate. Per poter studiare qualsiasi distribuzione si utilizza quindi la *standardizzazione* della distribuzione normale, comunemente chiamata come *distribuzione normale standard* in cui si ha $\mu = 0$ e $\sigma = 1$.

2.2.1 Distribuzione normale standard

Prima di dare una definizione formale, cerchiamo di capire bene cosa questa standardizzazione significa. Abbiamo detto che la media deve essere zero, mentre la varianza uno. Questo vuol dire che trasformeremo i valori x_i in z_i , cioè li standardizzeremo, così da usare tali valori come *distanza dalla media in termini di deviazioni standard*, cioè usare σ come *unità di misura*. Se z è positivo, allora sta sopra la media che è zero, di contro, se negativo sta sotto. Infatti, visto che z_i dipende da quanto l' i -esima realizzazione x_i della v.a. normale X si discosta dalla media, calcolare questa distanza risulta abbastanza semplice tramite la formula

$$z_i = \frac{x_i - \mu}{\sigma},$$

o in generale

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

Esempio

Da un test generico, risulta una media di 21.125 e deviazione standard 8.57. Estraendo 4 soggetti, vogliamo trasformare i punti x_i in punti z_i con $i = 1, 2, 3, 4$. In particolare

Soggetti	1	2	3	4
x_i	7	13	18	25

La standardizzazione avviene come segue

$$z_1 = \frac{7 - 21.125}{8.57} = -1.65, \quad z_2 = \frac{13 - 21.125}{8.57} = -0.94$$

$$z_3 = \frac{18 - 21.125}{8.57} = -0.36, \quad z_4 = \frac{25 - 21.125}{8.57} = 0.45$$

Si noti che

x_i	7	13	18	21.125	25
z_i	-1.65	-0.94	-0.36	0	0.45

da cui risulta chiaro come la *misura* della deviazione standard riesca a riassumere la distanza dell' i -esima realizzazione x dalla media.

Quindi, trasformando i punti da x_i a z_i si ottiene una distribuzione normale standardizzata con media $\mu = 0$ e deviazione standard $\sigma = 1$. La funzione di densità si formalizza come segue

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2},$$

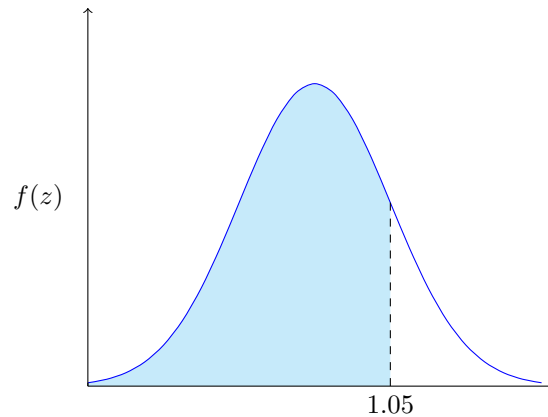
e si denota come $X \sim \mathcal{N}(0, 1)$. Sapendo che dalla equazione (1.3)

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) dx,$$

per avere sempre la possibilità di trovare l'area sottesa, si potrà ricorrere alla tavola della normale standard ogni volta che si dovranno calcolare le probabilità di una v.a. normale. Infatti, i valori degli integrali per tutti i valori di z sono stati opportunamente tabulati.

Le tavole sono redatte dipendentemente sia dalla variabile aleatoria specifica a cui si riferiscono, sia alla precisa area sottesa a cui i risultati sono collegati. Ovviamente c'è differenza fra una distribuzione t -student e una χ^2 , ma bisogna anche stare attenti a quale porzione della funzione di densità quei risultati degli integrali si riferiscono. Infatti ci sono diverse tabelle per la stessa distribuzione (ad esempio la normale standardizzata) che fanno riferimento a diverse aree della

Per trovare tutti i risultati, a volte, può essere utile ricorrere alla *simmetria* della distribuzione normale. Questa volta non ci interessa sapere l'area dalla media al valore z_1 , ma in generale $F(z) = P(Z \leq z_1)$. Non si potrà usare la tavola precedente, ma quella che si riferisce all'area che va ben oltre la media 0. Consideriamo ancora che $z_1 = 1.05$, il valore dell'integrale sarà quindi diverso perché l'area sottesa è diversa rispetto all'esempio precedente. Il valore dell'integrale



questa volta sarà

$$P(Z < 1.05) = \int_{-\infty}^{1.05} f(z) dz = 0.85314.$$

Se ad esempio, si vuole calcolare $F(-1) = P(Z \leq -1)$, allora la proprietà di simmetria gioca un ruolo importante. Infatti, si noti che $P(Z \leq -1)$ è semplicemente il complemento di $P(Z \leq 1)$, e per la simmetria equivale a $P(Z > 1)$

$$F(-1) = 1 - P(Z \leq +1) = 1 - F(1).$$

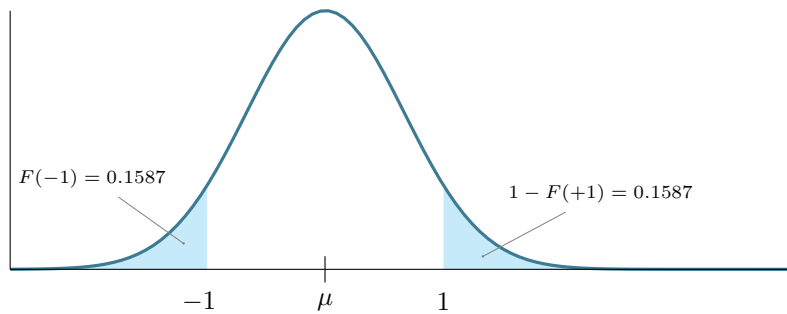
Osservando la tavola, si ha che $P(Z \leq 1) = 0.8413$, che applicando la formula

$$F(-1) = 1 - 0.8413 = 0.1587.$$

Si noti però che che l'area sottesa alla curva di sinistra $Z = -1$ (detta *coda sinistra*) è congruente con l'area sottesa a destra $Z = +1$ (detta *coda di destra*).

Facciamo adesso un semplice esempio. In investitore vuole investire in un titolo che ha media 100 e deviazione standard 10. Prima di investire il denaro, l'investitore vuole sapere quel è la probabilità che il reale valore del titolo (la realizzazione della v.a.) sarà compreso fra 90 e 120. Si passa subito a standardizzare gli x_i , cioè

$$z_1 = \frac{90 - 100}{10} = -1, \quad z_2 = \frac{120 - 100}{10} = 2.$$



Si è così passati da un intervallo $P(98 \leq X \leq 120)$ a uno standardizzato $P(-1 \leq Z \leq 2)$ basandoci sulla misura della deviazione standard. Si ottiene che

$$P(98 \leq X \leq 120) = P(-1 \leq Z \leq 2) = 1 - P(Z \leq +1) - P(Z \geq 2)$$

da cui, leggendo i valori nella tavola e grazie alla simmetria, si ottiene

$$P(-1 \leq Z \leq 2) = 1 - 0.1587 - 0.0228 = \int_{-1}^2 f(z) dz = 0.8185$$

che è la probabilità cercata.

Formalizzando, si avrà così una formula generale per calcolare la probabilità di una distribuzione normale standard. In particolare, sarà molto importante quando si andranno a costruire gli intervalli di confidenza, nel capitolo 3.7.

Sia $X \sim \mathcal{N}(\mu, \sigma^2)$, la variabile aleatoria $Z = (X - \mu)/\sigma$ segue la distribuzione normale standard $Z \sim \mathcal{N}(0, 1)$. Date due possibili realizzazioni x_1 e x_2

$$\begin{aligned} P(x_1 < X < x_2) &= P\left(\frac{x_1 - \mu}{\sigma} < Z < \frac{x_2 - \mu}{\sigma}\right) \\ &= F\left(\frac{x_2 - \mu}{\sigma}\right) - F\left(\frac{x_1 - \mu}{\sigma}\right) \end{aligned}$$

in cui $F(z)$ è la funzione di ripartizione della v.a. $Z \sim \mathcal{N}(0, 1)$.

2.2.2 Teorema del limite centrale

Questo è uno dei più importanti risultati nella teoria della probabilità. Il teorema del limite centrale ci dice semplicemente che la somma di un grande numero di variabili aleatorie indipendenti ha una distribuzione *approssimativamente* normale. Si capisce subito l'importanza che ricopre

la distribuzione gaussiana (o normale) nella statistica e nella teoria della probabilità.

Nei teoremi limite della probabilità, si guarda alla somma delle variabili aleatorie e si vogliono estrarre informazioni dalla distribuzione di questa somma. Qui si vuole dimostrare che questa è approssimativamente distribuita normalmente. Per la legge debole dei grandi numeri, la distribuzione della media campionaria (del singolo campione e non della intera popolazione, vedi capitolo 3), che definiamo \bar{m} , è concentrata attorno alla media μ . Il problema però generale, è che se n tende a infinito, la distribuzione della somma di v.a. si disperde. Infatti, definiamo la somma $H_n = X_1 + X_2 + \dots + X_n$ come la somma delle prime n variabili aleatorie iid. I teoremi limite sappiamo essere collegati alle proprietà di H_n e delle v.a. X_i per n che diventa sempre più grande. Data l'indipendenza delle v.a., ogni X_i avrà media e varianza uguale alle altre n v.a., cioè $\text{Var}(H_n) = n\sigma^2$ facendo disperdere sempre più H_n al crescere di n . Definendo invece la media del campione (cioè delle prime n v.a.) come $\bar{m} = H_n/n$, allora abbiamo visto che si ottiene $\mathbb{E}(\bar{m}) = \mu$ e $\text{Var}(\bar{m}) = \sigma^2/n$. Si è visto, con la legge dei grandi numeri, che $\text{Var}(\bar{m})$ converge a zero per n che cresce, mentre la distribuzione di \bar{m} converge attorno alla media della popolazione μ .

Abbiamo visto, inoltre, come una distribuzione normale con media μ e varianza σ^2 può essere standardizzata così da ottenere una distribuzione con media $\mu = 0$ e varianza $\sigma^2 = 1$. Si è visto che

$$Z_n = \frac{H_n - n\mu}{\sigma\sqrt{n}},$$

dal quale $\mathbb{E}(Z) = 0$ e $\text{Var}(Z) = 1$.

La distribuzione di Z_n , all'aumentare di n non si disperderà né si contrarrà a un punto, perché la varianza è sempre 1 attorno alla media centrata 0. Il teorema del limite centrale si occupa della forma asintotica che la distribuzione di Z_n assumerà, che si proverà essere una distribuzione normale standardizzata.

La varianza della somma sarà nH_n , tendendo ad infinito all'aumentare di n . Tramite la standardizzazione, possiamo mantenere la varianza fissa, così

$$Z_n = \frac{H_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}},$$

da cui

$$\mathbb{E}(Z) = \frac{\mathbb{E}[X_1 + X_2 + \dots + X_n] - n\mu}{\sigma\sqrt{n}} = 0,$$

e

$$\text{Var}(Z_n) = \frac{\text{Var}(Z_1) + \text{Var}(Z_2) + \dots + \text{Var}(Z_n)}{n\sigma^2} = \frac{n\sigma^2}{n\sigma^2} = 1.$$

Il teorema del limite centrale.

Siano X_1, X_2, \dots variabili aleatorie iid, ognuna avente media μ e varianza σ^2 . Per ogni $-\infty < a < \infty$ e per $n \rightarrow \infty$, si ottiene

$$P\left\{\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}z^2} dx$$

cioè, con

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}x^2} dx,$$

allora

$$\lim_{n \rightarrow \infty} P(Z_n \leq a) = \phi(z).$$

Interessante è sicuramente notare che per il teorema del limite centrale, la distribuzione di ogni X_i deve essere indipendente, ma non deve soddisfare nessun altro requisito (continua, discreta etc). C'è invece bisogno che media e varianza siano finite. Ci sarebbe anche un altro sottile requisito, molto spesso considerato soltanto implicitamente: le distribuzioni di X_i , per soddisfare il teorema, devono essere mantenute *fisse*, nel senso che implicitamente assumiamo che all'aumentare di n , la statistica di ogni X_i non cambia. Ci sono però casi in cui variando n , variano anche le distribuzioni (media e varianza) delle variabili aleatorie. Si noti come il teorema è spesso utilizzato nei modelli dei mercati finanziari, infatti movimenti nei prezzi sono spesso considerati come distribuiti normalmente perché causati da piccoli eventi e piccole decisioni individuali prese da un numero tendente ad infinito di diversi agenti economici nel mercato, questo crea le basi per un mercato efficiente dove il prezzo riflette le informazioni degli agenti economici. Il ragionamento è simile al moto causato da piccolissime molecole in un fluido, che spostano una particella causando una distribuzione di v.a. chiamata movimento Brawniano.

Tramite il teorema del limite centrale tratteremo Z_n *come se fosse* normalmente distribuita. Questo perché la condizione di normalità è preservata per trasformazioni lineari del tipo della standardizzazione da H_n a Z_n , cioè H_n può essere trattata *come se fosse* normale con media $n\mu$ e varianza $n\sigma^2$.

La normalità è preservata per trasformazioni lineari.

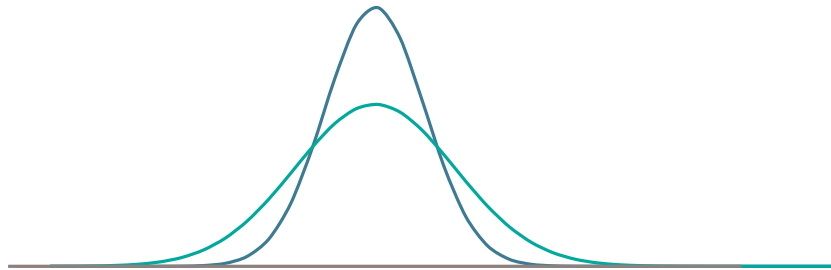
Sia X una v.a. con media μ e varianza σ^2 . Si denotino due scalari, $a \neq 0$ e b , allora

$$Y = aX + b$$

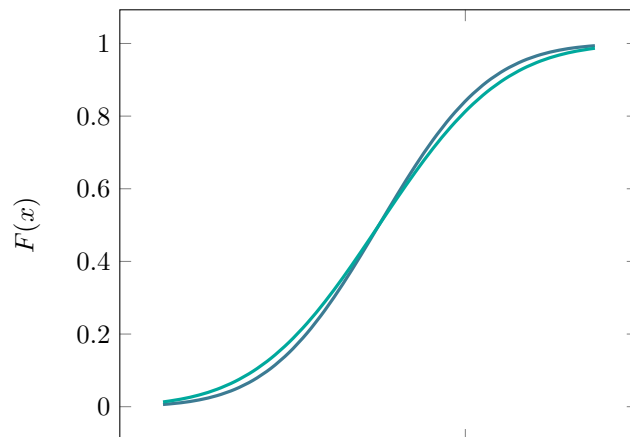
è ancora normale, con media e varianza

$$\mathbb{E}(Y) = a\mu + b \quad \text{Var}(Y) = a^2\sigma^2.$$

Si noti inoltre che il teorema del limite centrale ci informa che la funzione di ripartizione in Z_n converge alla funzione di ripartizione normale. Questo significa che non è un teorema sulla convergenza di funzioni di densità di probabilità, ma delle funzioni di ripartizione. Infatti le densità possono assumere ampiezze diverse, mantenendo però tra loro le funzioni di ripartizione molto simili.



(a) Funzioni di densità



(b) Funzioni di ripartizione

2.3 Distribuzione binomiale

Si consideri un esperimento casuale che può presentare due soli risultati che siano mutuamente esclusivi e collettivamente esaustivi, ad esempio successo e insuccesso. Si immagini un lancio di dadi non truccati n volte, dove l'evento testa può accadere con probabilità p , mentre l'evento croce con probabilità *di coda* $1 - p$. Si noti che ogni realizzazione del lancio è indipendente dalle realizzazioni precedenti, come richiesto. Chiameremo X variabile aleatoria **binomiale** con parametri n e p . La funzione di probabilità (quindi nel discreto) è calcolata come

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

Si noti come ogni evento è indipendente dall'altro, si deve quindi applicare la regola delle probabilità indipendenti, cioè come precedentemente visto nel capitolo C, $P(A \cap B) = P(A)P(B)$. Nel caso in questione si assuma $n = 4$ lanci. La probabilità che esca 4 volte testa è p^4 , mentre la probabilità che esca 4 volte croce è $(1 - p)^4$. Invece, la probabilità che esca 3 volte testa e 1 croce è $p^3(1 - p)$. Ancora, la probabilità che esca 2 volte testa e 2 volte croce è $p^2(1 - p)^2$ e così via.

La probabilità $P(x)$ si considera come la probabilità che esca testa x volte in n lanci. Mentre definiamo coefficiente binomiale

$$\binom{n}{x} = \text{il numero di differenti } n \text{ lanci che contiene } x \text{ teste.}$$

In generale, il numero di sequenze con x successi in n prove indipendenti e mutuamente esclusive è la combinazione

$$B_x^n = \binom{n}{x} = \frac{n!}{x!(n-x)!},$$

con $n! = n \times (n - 1) \times (n - 2) \times \dots \times 1$, con la convenzione $0! = 1$. In poche parole, stiamo dicendo che l'evento " x successi in n prove" può manifestarsi in B_x^n modi mutualmente esclusivi, ognuno con probabilità $p^x(1 - p)^{n-x}$.

Per una variabile aleatoria le cui realizzazioni possono essere successo $P(1) = p$ o insuccesso $P(0) = 1 - p$, la distribuzione sarà chiamata di Bernoulli, per cui

$$P(x) = P(X = x) = p^x (1 - p)^{1-x},$$

si noti che qui non si considerano gli n lanci come nel caso della distribuzione binomiale, infatti la distribuzione di Bernoulli ne è un suo caso semplice. Inoltre, si definisca la media come

$$\mu = \sum_x xP(x) = (0)(1 - p) + (1)p = p,$$

e la varianza come

$$\sigma^2 = \sum_x (x - \mu)^2 P(x) = \sum_x x^2 P(x) - \mu^2 = 0^2(1-p) + 1^2p - p^2 = p(1-p).$$

Per la distribuzione binomiale, avremo che l'esperimento verrà ripetuto n volte in modo indipendente e mutuamente esclusivo. Si consideri X come il numero di successi in n prove indipendenti con, ognuna, probabilità di successo p . Allora X segue una distribuzione binomiale con media

$$\mu = np$$

e varianza

$$\sigma^2 = np(1-p).$$

Si noti che la probabilità della realizzazione di una prova è costante ad ogni ripetizione e che quindi passate realizzazioni non possono influire sulla probabilità delle realizzazioni future. Per l'applicazione di tale distribuzione è importante che sia applicata a casi concreti che possano presentare due soli possibili risultati: successo o insuccesso (acceso o spento) come testa o croce etc. Infine, è da sottolineare come la funzione di probabilità rispetti il requisito

$$\sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = 1.$$

Esempio (Statistica - Newbold, Carlson, Throne)

Un agente assicurativo ipotizza che, per un particolare tipo di polizza, la probabilità di una sottoscrizione al primo contatto sia 0.4. Se la v.a. X assume il valore 1 in caso di sottoscrizione e 0 altrimenti, si tratta di una variabile bernoulliana, con probabilità di successo $p = 0.4$. Trovare la media e la varianza della distribuzione.

La funzione di probabilità di X è $P(0) = 0.6$ e $P(1) = 0.4$. Applicando le formule, la media della distribuzione è $\mu = np = 1.6$ e la varianza $\sigma^2 = np(1-p) = (0.4)(0.6) = 0.24$.

Supponiamo adesso che lo stesso agente assicurativo abbia cinque polizze e che, per ognuna di esse, la probabilità di sottoscrizione al primo contatto sia ancora 0.4. Usando la formula della distribuzione binomiale:

- a) Calcolare la probabilità che l'assicuratore riesca a far sottoscrivere al massimo una polizza;

b) Calcolare la probabilità che le polizze sottoscritte siano nell'intervallo $[2, 4]$;

c) illustrare graficamente la funzione di probabilità.

a): Si consideri che

$$P(\text{nessuna polizza}) = P(0) = \frac{5!}{0!5!}(0.4)^0(0.6)^5 = 0.078$$

$$P(\text{una polizza}) = P(1) = \frac{5!}{1!4!}(0.4)^1(0.6)^4 = 5(0.4)(0.6)^4 = 0.259$$

ne segue che

$$\begin{aligned} P(\text{al massimo una polizza}) &= P(X \leq 1) = P(X = 0) + P(X = 1) \\ &= 0.078 + 0.259 \\ &= 0.337 \end{aligned}$$

b): Si consideri che

$$P(2) = \frac{5!}{2!3!}(0.4)^2(0.6)^3 = 10(0.4)^2(0.6)^3 = 0.346$$

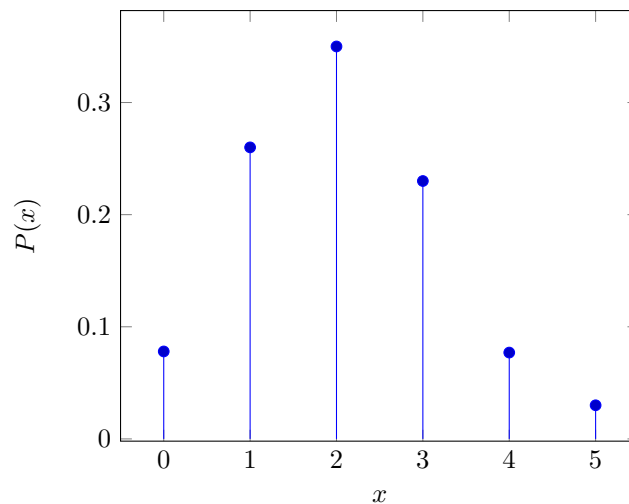
$$P(3) = \frac{5!}{3!2!}(0.4)^3(0.6)^2 = 10(0.4)^3(0.6)^2 = 0.230$$

$$P(4) = \frac{5!}{4!1!}(0.4)^4(0.6)^1 = 5(0.4)^4(0.6)^1 = 0.077$$

da cui

$$\begin{aligned} P(2 \leq X \leq 4) &= P(2) + P(3) + P(4) \\ &= 0.346 + 0.230 + 0.077 \\ &= 0.653 \end{aligned}$$

c): La funzione di probabilità si rappresenta come segue



2.4 Distribuzione di Poisson

Una v.a. X si dice di Poisson con parametro $\lambda > 0$ se la sua funzione di probabilità è data da

$$P(x) = P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

Questa è una funzione di probabilità in quanto

$$\sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right) = e^{-\lambda} e^{\lambda} = 1. \quad (2.4)$$

La distribuzione di Poisson si usa quando si ha una variabile aleatoria del tipo: il numero di clienti che arriva alla cassa di un negozio in un particolare intervallo di tempo, oppure il numero di passeggeri che arrivano in stazione in ogni intervallo di 30 minuti dalle 15 alle 20, il numero delle chiamate telefoniche che arrivano in un ufficio in un giorno, oppure ancora il numero di automobili coinvolte in un incidente in un determinato giorno etc. Si nota però una certa vicinanza con la distribuzione binomiale, infatti se ad esempio si assume che la v.a. X sia il numero di errori in un libro che ha un totale di n parole, si è portati a pensare di applicare proprio la distribuzione binomiale. Si usa la distribuzione di Poisson ogni volta che p è bassa ed n è grande. Nel caso dell'esempio del libro, la probabilità che ogni parola sia errata è molto bassa, mentre le parole n sono tante, ed è proprio un caso in cui si potrebbe applicare la distribuzione di Poisson. In poche parole ogni volta che un intervallo è diviso in un grande numero di sottointervalli n , e la probabilità p del verificarsi di un evento in ogni sottointervallo è molto piccola, deve far pensare che X sia distribuita secondo Poisson. Inoltre p deve essere costante in ogni sottointervallo, la probabilità del verificarsi di un evento nello stesso sottointervallo è trascurabile e gli eventi accaduti in intervalli disgiunti sono indipendenti. La distribuzione di Poisson può essere derivata dalla distribuzione binomiale dove i limiti soddisfano $n \rightarrow \infty$ e $p \rightarrow 0$ così da ottenere $\lambda = np$, come costante che ci informa circa il numero medio di successi per un particolare intervallo. In particolare

$$e^{-\lambda} \frac{\lambda^x}{x!} \approx \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

La media e la varianza della distribuzione di Poisson sono

$$\mu = \lambda, \quad \sigma = \lambda.$$

Esempio In un centralino, le chiamate telefoniche che arrivano in ogni intervallo di 10 minuti è distribuito come una v.a. di Poisson con $\lambda = 2$.

- a) Si calcoli la probabilità che più di tre chiamate arrivino in un intervallo di 10 minuti.

b) Si calcoli la probabilità che nessuna chiamata arrivi nello stesso intervallo.

Risolviamo il punto a). La funzione di probabilità di X è

$$P(x) = P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} = e^{-2} \frac{2^x}{x!}, \quad x = 0, 1, \dots$$

da cui, sfruttando l'equazione (2.4)

$$\begin{aligned} P(X > 3) &= -P(X \leq 3) = 1 - \sum_{x=0}^3 e^{-2} \frac{2^x}{x!} \\ &= 1 - e^{-2} \left(1 + 2 + \frac{4}{2} + \frac{8}{6} \right) \approx 0.143 \end{aligned}$$

Risolvendo b)

$$P(X = 0) = e^{-2} \approx 0.135$$

2.5 Distribuzione esponenziale

Si chiama *distribuzione esponenziale* una v.a. continua X con parametro $\lambda > 0$ con funzione di densità

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

La funzione di ripartizione si definisce

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

con media $\mu = \frac{1}{\lambda}$ e varianza $\sigma^2 = \frac{1}{\lambda^2}$.

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ &= (-x e^{-\lambda x}) \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\ &= 0 - \frac{e^{-\lambda x}}{\lambda} \\ &= \frac{1}{\lambda}. \end{aligned}$$

Per la varianza si calcoli prima il valore atteso seguente

$$\begin{aligned}\mathbb{E}(X^2) &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \\ &= (-x^2 e^{-\lambda x}) \Big|_0^{\infty} + \int_0^{\infty} 2x e^{-\lambda x} dx \\ &= 0 - \frac{2}{\lambda} \mathbb{E}(X) \\ &= \frac{2}{\lambda^2}.\end{aligned}$$

Dalle proprietà della varianza, sappiamo che $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$, cioè

$$\text{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

La distribuzione esponenziale è utile nei problemi delle code di attesa, ad esempio, è un buon modello per l'ammontare di tempo fino a che un evento di interesse si verifichi: un messaggio che arrivi a un computer, che si verifichi un incidente di interesse, una lampadina che si fulmini etc. Si vedrà, inoltre, che la distribuzione esponenziale è molto vicina alla distribuzione geometrica che è invece una distribuzione simile, ma nel discreto. Vi è anche una certa relazione con la distribuzione di Poisson. Abbiamo visto che λ ci informa sul numero di volte che si manifesta un evento nell'intervallo di tempo, la probabilità di un successo durante l'intervallo di tempo è invece fornita dalla distribuzione esponenziale. Infatti λ nella distribuzione di Poisson esprime il numero medio di eventi nell'unità di tempo, mentre $1/\lambda$ rappresenta il tempo medio di attesa affinché si verifichi il primo evento.

Esempio *La durata di una chiamata telefonica in minuti è una v.a. X esponenziale con parametro $\lambda = 1/10$. Se qualcuno arriva ad una cabina telefonica poco prima di voi, si calcoli la probabilità che avrete di attendere*

- a) meno di 5 minuti;
- b) tra $[5, 10]$ minuti.

a): La funzione di densità di X è definita come

$$f(x) = \begin{cases} \frac{1}{10} e^{-\frac{x}{10}} & x > 0 \\ 0 & x < 0. \end{cases}$$

da cui

$$P(X < 5) = \int_0^5 \frac{1}{10} e^{-\frac{x}{10}} dx = -e^{-\frac{x}{10}} \Big|_0^5 = 1 - e^{-0.5} \approx 0.393$$

b) : Con lo stesso procedimento

$$P(5 < X < 10) = \int_5^{10} \frac{1}{10} e^{-\frac{x}{10}} dx = e^{-0.5} - e^{-1} \approx 0.239$$

2.6 Distribuzione geometrica

Si supponga di effettuare ripetuti e indipendenti lanci di una moneta, $0 < p < 1$ è la probabilità che esca testa. La variabile aleatoria geometrica è il numero X di lanci che vi è bisogno di fare affinché si verifichi un evento testa per la prima volta. La funzione di probabilità è data da

$$P(x) = P(X = x) = (1 - p)^{x-1} p.$$

Per avere $X = x$, è necessario e sufficiente che i primi $x - 1$ lanci siano croce, mentre l' x -esimo lancio sia x . Si noti che questo risultato dipende anche dal fatto che i prossimi lanci siano indipendenti da quelli passati. Inoltre

$$\begin{aligned} \sum_{x=0}^{\infty} P(x) &= \sum_{x=0}^{\infty} (1 - p)^{x-1} p \\ &= p \sum_{x=0}^{\infty} (1 - p)^{x-1} \\ &= p \cdot \frac{1}{1 - (1 - p)} = 1 \end{aligned} \tag{2.5}$$

da cui, l'evento testa, si verificherà con probabilità 1 (prima o poi). In generale la distribuzione geometrica può essere interpretata come una v.a. discreta in termini di ripetuti e indipendenti esperimenti finché il primo evento si verifica. Si può quindi intuire la similarità con il caso continuo della distribuzione esponenziale.

Esempio *Si consideri l'esperimento che consiste nel lancio di una coppia di dadi non truccati*

a) Calcolare la probabilità che occorranza meno di sei lanci per ottenere un 7.

b) Calcolare la probabilità che occorranza più di sei lanci per ottenere un 7.

a): Se la probabilità di ottenere un 7 in ogni lancio è $1/6$, sia X il numero di lanci richiesti per il primo successo nell'ottenere un 7. Quindi X è una v.a. geometrica con parametro $p = 1/6$. L'equazione (2.5) è ottenuta sfruttando la proprietà delle serie geometriche secondo cui

$$\sum_{a=0}^{\infty} ar^n = \sum_{a=1}^{\infty} ar^{n-1} = \frac{a}{1 - r} \quad |r| < 1.$$

Quindi

$$P(X > x) = \sum_{i=x+1}^{\infty} (1-p)^{i-1} p = \frac{(1-p)^x p}{1 - (1-p)} = (1-p)^x.$$

Dalla simmetria

$$F(x) = P(X \leq x) = 1 - (1-p)^x$$

che applicata al problema a), si ottiene

$$P(X < 6) = P(X \leq 5) = F(5) = 1 - \left(\frac{5}{6}\right)^5 \approx 0.598$$

Analogamente b):

$$\begin{aligned} P(X > 6) &= 1 - P(X \leq 6) \\ &= 1 - F(6) \\ &= 1 - \left[1 - \left(\frac{5}{6}\right)^6\right] \\ &= \left(\frac{5}{6}\right)^6 \approx 0.335 \end{aligned}$$

Capitolo 3

L'INFERENZA STATISTICA

3.1 Cosa è l'inferenza statistica?

Per motivi di contenimento del “costo” di una indagine statistica, vi è la necessità di limitare l'indagine ad un *campione*, cioè una parte rappresentativa dell'intero collettivo da indagare o *popolazione*. Per esempio, come intervistare tutti gli italiani per conoscerne le preferenze di voto? Come chiedere a tutti i consumatori di un certo prodotto il loro grado di soddisfazione? È evidente che analizzare tutta la popolazione di riferimento (es. gli italiani, i consumatori etc. . .) è praticamente troppo lungo e costoso. Si deve quindi trovare un metodo con il quale trovare campioni opportunamente selezionati per ottenere stime caratteristiche della popolazione che siano più vicine possibili ai suoi veri valori registrati. Il metodo che consente il passaggio dal campione osservato alla popolazione generale si chiama *inferenza statistica*, cioè permette di “inferire” alcune caratteristiche della popolazione, osservandone una sua parte. Atteso che una indagine campionaria non può che fornire risultati che sono approssimativamente validi per l'intera popolazione, i metodi dell'inferenza statistica consentono di precisare a posteriori i margini di tale approssimazione, oppure di programmare a priori l'articolazione ed il dimensionamento ottimale dell'indagine campionaria che consentano di contenere i suddetti margini di approssimazione entro limiti accettabili sotto il profilo pratico.

3.2 Campionamento

Si denomi *popolazione* il collettivo statistico oggetto di inferenza statistica. La parte del collettivo statistico sottoposta all'osservazione viene chiamata *campione*. Questo per rivelarsi utile, deve essere *rappresentativo*, ossia il campione deve possedere per quanto possibile le stesse carat-

teristiche della popolazione. Se, ad esempio, per capire il grado di soddisfazione della popolazione dei consumatori di un determinato prodotto, si scelgono solo quei consumatori che hanno dato un feedback positivo sul sito aziendale, allora questo campione *non* è rappresentativo della intera popolazione. Gli elementi che formeranno il campione, dovranno essere *indipendenti* dalle caratteristiche che si vogliono esaminare. Questo è importante perché, in fin dei conti, non si conosce con certezza quali siano queste caratteristiche in grado di soddisfare il requisito di indipendenza. La rappresentatività del campione impone che si seguano delle tecniche che diano a ciascuna unità della popolazione la stessa possibilità di essere inserita nel campione, senza scegliere soggettivamente unità portatrici di talune caratteristiche. In questo caso si otterrà un campione *non distorto*.

La tecnica principale per ottenere questo campione è il “campionamento casuale”, dove gli elementi scelti sono per definizione indipendenti dalle caratteristiche dei soggetti selezionati. La formazione di un campione casuale, può essere assimilata alla estrazione di n palline da un'urna che ne contenga N :

- i) Estrazione senza ripetizione o esaustiva*, ogni pallina può essere estratta al più una volta e il contenuto dell'urna N si modifica riducendosi dopo ogni ripetizione. Ciò accade perché la pallina estratta non viene rimessa nell'urna.
- ii) Estrazione con ripetizione o bernulliana*, ciascuna pallina può essere estratta più volte mantenendo costante il contenuto dell'urna N . In tal caso la pallina estratta viene sempre rimessa dentro l'urna prima dell'estrazione successiva.

I campioni senza ripetizione al punto *i*), sono quelli più precisi, anche se per grossi campioni, l'uso del metodo con ripetizione permette di utilizzare metodi più semplici senza alterare il risultato. Infatti, l'estrazione di una pallina da un'urna contenente numerosissime palline, non varia in maniera sensibile la sua composizione. In questo caso la estrazione senza ripetizione e quella con ripetizione coincidono.

3.3 Parametri, Statistiche e Probabilità

Avendo un campione, si vogliono ricavare delle informazioni sufficientemente approssimate sulla distribuzione di un carattere della popolazione. I valori caratteristici di detta distribuzione si denominano *parametri*, che noi indicheremo con lettere greche, quali θ . Solitamente, la media del carattere nella popolazione si indica con μ , mentre la sua varianza con σ^2 .

La corrispondenza fra caratteristiche del campione e della popolazione è soggetta alle leggi del calcolo delle probabilità. Prima di formare il campione, ogni osservazione campionaria può essere definita come variabile casuale suscettibile di assumere ciascuna delle modalità della popolazione con probabilità pari alla frequenza relativa di tale modalità nella popolazione medesima. Se la popolazione di un comune è costituita per la frazione π da uomini e $1 - \pi$ da donne, allora un abitante da scegliere a caso da questa popolazione sarà con probabilità π uomo e con probabilità $1 - \pi$ donna. L'osservazione campionaria a priori, cioè dalla quale si formerà il campione osservato, è una variabile casuale (v.c.) che ha distribuzione di probabilità pari alla distribuzione relativa del carattere nella popolazione. Indicheremo la v.c. con X_i dove $i = 1, 2, \dots, n$. Si assuma che in un'urna N ci siano 10000 palline, di cui θ (parametro della popolazione) sono rosse. Vi sono state 100 estrazione con $x = 10$ palline rosse estratte. Il problema è quello di capire, in termini probabilistici, la relazione fra θ (la vera e ignota proporzione di palline rosse in N) e x (l'osservazione). Il campione è quindi formato da osservazioni campionarie x_i che sono realizzazione della v.c. X_i . La distribuzione di X_i è determinata dalle caratteristiche della popolazione: θ , e il tipo di campionamento utilizzato. Ad esempio, se il campionamento è del tipo i), senza ripetizione, allora la X_i sarà distribuita come una v.c. binomiale. Se il campionamento è del tipo ii), con ripetizione, la X_i sarà distribuita come v.c. ipergeometrica e così via.

Ogni funzione delle osservazioni campionarie è chiamata *statistica*. Queste sono caratteristiche del campione, così come i parametri sono caratteristiche della popolazione. La media del campione, la varianza del campione eccetera, sono delle statistiche. Queste sono v.c. funzioni delle v.c. X_i prima della formazione del campione, e quindi delle osservazioni x_i . Una volta che il campione è formato, le statistiche diventano valori delle osservazioni x_i . Si utilizzeranno lettere latine maiuscole se le statistiche sono v.c., minuscole se sono valori. In generale \bar{X}, \bar{x} per la media campionaria, S^2, s^2 per la varianza campionaria Π, π per la frequenza relativa campionaria, ecc.

3.4 Distribuzioni campionarie delle statistiche

La distribuzione di probabilità di una statistica è detta *distribuzione campionaria della statistica*, in quanto esprime la distribuzione dei valori che la statistica può assumere in relazione a ciascuno dei campioni che possono scaturire dalla estrazione. Nel capitolo 2 la analizzeremo più in dettaglio, in questa introduzione però è importante capirne il significato. Si consideri una caratteristica di una popolazione, ad esempio la media μ . Formato un campione rappresentativo, si sceglie una statistica campionaria, in questo caso la media campionaria \bar{X} . Ogni campione

casuale determinerà un diverso valore, o realizzazione \bar{x} della media \bar{X} . La distribuzione campionaria di questa statistica è quindi la distribuzione delle medie campionarie ottenute su tutti i possibili campioni di pari ampiezza, estratti dalla popolazione. Si immagini un'urna N formata da palline numerate come segue $N = 1, 2, 4, 4, 5, 8$. La media (aritmetica) della popolazione sarà

$$\mu = \frac{1 + 2 + 4 + 4 + 5 + 8}{6} = 4.$$

Si considerino dieci estrazioni

Campione	Media Campionaria
1, 5	3.0
1, 8	4.5
2, 4	3.0
2, 8	5.0
4, 4	4.0
4, 5	4.5
4, 8	6.0
5, 2	2.3
5, 8	6.5
5, 1	3.0

Ogni campione ha la probabilità di essere estratto pari a $1/10$. Si noti però, come alcune medie campionarie siano ripetute più volte, infatti la probabilità di ottenere una media campionaria pari a 3 è $3/10$. Cioè ci sono delle medie campionarie la cui realizzazione è collegata a una probabilità più alta:

\bar{X}	Probabilità
3.0	$3/10$
4.5	$2/10$

Si noti che all'aumentare della ampiezza del campione, la media campionaria \bar{X} diventa più concentrata attorno alla media della popolazione μ . Si tenga a mente però che per popolazioni molto grandi, sarebbe poco ragionevole costruire la distribuzione di tutti i possibili campioni di pari ampiezza, di questo si parlerà più avanti.

3.5 Problemi Inferenziali

In questa sede, si possono sottolineare tre principali problemi inferenziali:

- a) stima dei parametri;
- b) verifica di ipotesi sui parametri;
- c) verifica non parametrica.

In questo scritto ci concentreremo sui primi due problemi a) e b), chiamati problemi *parametrici*: sulla base delle risultanze campionarie, questi tendono a valutare i parametri che caratterizzano la distribuzione del carattere della popolazione, o a vagliare delle ipotesi formulate a priori sul valore di tali parametri.

I problemi di tipo c) attengono ad aspetti della distribuzione del carattere nella popolazione non suscettibili di essere espressi dai parametri che compaiono nella forma funzionale di tale distribuzione.

Della stima dei parametri si suole distinguere due aspetti:

- a_1 stima puntuale dei parametri;
- a_2 stima intervallare dei parametri.

Un problema analogo a quello intervallare a_2 è quello del *dimensionamento ottimale del campione*.

3.6 Stima puntuale dei parametri

Si indichi con θ un parametro incognito della distribuzione di un carattere in una determinata popolazione. Il problema della stima puntuale consiste nella migliore valutazione di θ ottenibile sulla base delle osservazioni campionarie, sintetizzate da un'opportuna statistica. Questa statistica, riguardata prima della formazione del campione, è una v.c. denominata *stimatore* del parametro θ , che indicheremo con

$$T_n = h(X_1, X_2, \dots, X_n). \quad (3.6)$$

Una volta formato il campione, il valore assunto da tale statistica si chiama *stima* del parametro θ che indicheremo come

$$t_n = h(x_1, x_2, \dots, x_n) \quad (3.7)$$

Uno **stimatore** per un parametro della popolazione è una v.c. funzione delle variabili del campione: la funzione che raccoglie ogni possibile valore che il parametro della popolazione da stimare può assumere. La **stima** è invece ogni singolo valore che viene assegnato al parametro della popolazione, è il valore di questa v.c.

Si considerino le vendite mensili di un particolare prodotto. Uno *stimatore* della media della popolazione μ , può essere la media campionaria (come anche la mediana campionaria ecc). Rilevando la media di un campione casuale di vendite mensili si scopre che questo è 8950 pezzi, questo valore è la *stima* delle vendite medie mensili nella popolazione. Quando bisogna valutare la stima di un parametro incognito θ , se si usa un singolo valore ottenuto dal campione come valore rappresentativo del parametro incognito, allora questo valore viene definito *stima puntuale*. Lo *stimatore puntuale* determina quindi un unico valore chiamato stima puntuale. La media campionaria \bar{X} è uno stimatore puntuale della media della popolazione μ , e il valore che \bar{X} assume come sua realizzazione campionaria viene detto stima puntuale \bar{x} .

In presenza di più stimatori di θ è opportuno utilizzare lo stimatore migliore. Questo deve soddisfare tre proprietà: deve essere *non distorto*, *efficiente* e *consistente*.

3.6.1 Stimatore non distorto

Uno stimatore puntuale T_n viene definito come **non distorto** (o corretto) per il parametro della popolazione θ se il suo valore medio è pari al parametro da stimare:

$$\mathbb{E}(T_n) = \theta \quad (3.8)$$

Si noti che uno stimatore non distorto si intende capace di stimare “in media” un parametro della popolazione, non significa che $T_n = \theta$. Si ricordi che il valore atteso $\mathbb{E}(\cdot)$ è una media di valori di T_n in un elevato numero di campioni ripetuti. Questa proprietà assicura che lo stimatore non sia errato sistematicamente per eccesso o per difetto: per più stime del parametro, gli errori dovrebbero tendere a compensarsi.

La **distorsione** di T_n è la differenza fra la sua media e θ :

$$D(T_n) = \mathbb{E}(T_n) - \theta.$$

3.6.2 Stimatore consistente

Si consideri la seconda proprietà: la *non distorsione asintotica*

Uno stimatore puntuale T_n viene definito **consistente** per il parametro θ della popolazione se al crescere della dimensione della popolazione, lo stimatore converge in probabilità verso il parametro.

$$\lim_{n \rightarrow \infty} \mathbb{P}\{|T_n - \theta| < \epsilon\} = 1$$

Al crescere della ampiezza del campione n , la distorsione $D(T_n)$ è inferiore ad un valore ϵ arbitrariamente piccolo tendendo a 0. Questo significa che è certo che l'errore non supererà ϵ .

Si tenga a mente che tutti gli stimatori non distorti sono asintoticamente non distorti (o consistenti), mentre non tutti gli stimatori consistenti sono non distorti (es. la varianza campionaria, perché?).

3.6.3 Stimatore efficiente

In presenza di più stimatori non distorti per uno stesso parametro, si definisce **stimatore più efficiente** quello che possiede una varianza minore rispetto ad altri stimatori corretti.

La varianza dello stimatore

$$\sigma^2 = D^2(T_n) = \mathbb{E}[(T_n - \theta)^2],$$

è una media degli errori di stima $T - \theta$, cioè la misura di imprecisione del campione. Lo stimatore più efficiente deve possedere l'imprecisione minore fra tutti gli stimatori non distorti.

3.6.4 Il metodo della massima verosimiglianza

3.7 Stima intervallare di un parametro

Si è detto che con la stima puntuale si vuole valutare un parametro incognito θ tramite un singolo valore. Con la *stima intervallare* si valuta il parametro incognito θ cercando un intervallo (detto *intervallo di confidenza*¹) che con una probabilità assegnata (detta *livello di confidenza*) dovrebbe contenere il parametro. In poche parole, la stima per intervalli fornisce un determinato campo di valori all'interno del quale possa cadere il valore reale (sconosciuto) della popolazione. Dal *teorema del limite centrale*, sappiamo che all'aumentare della ampiezza dei campioni, la distribuzione della media campionaria sarà sempre più vicina ad essere *normale*: il campionamento

¹Dall'inglese *confidence* che significa sicurezza, fiducia.

di una popolazione diventa più preciso e tendente verso il vero valore incognito. Questo però non è rispettato dalla stima puntuale. La stima puntuale della proporzione di palline rosse sarebbe uguale osservando 1 pallina rossa in un campione da 10 palline, che osservando 100 palline rosse in un campione formato da 1000 palline. Invece la stima intervallare rispetta il teorema del limite centrale, infatti, a parità di altre condizioni, campioni più grandi determinano stime per intervallo più precise. Si indichi con $(1 - \alpha)$ il parametro del livello di confidenza, il livello prefissato di probabilità dell'intervallo. Si indichi con $\alpha/2$ la probabilità a di non rientrare in questo intervallo. Il ragionamento risulta chiaro osservando la figura 3.1.

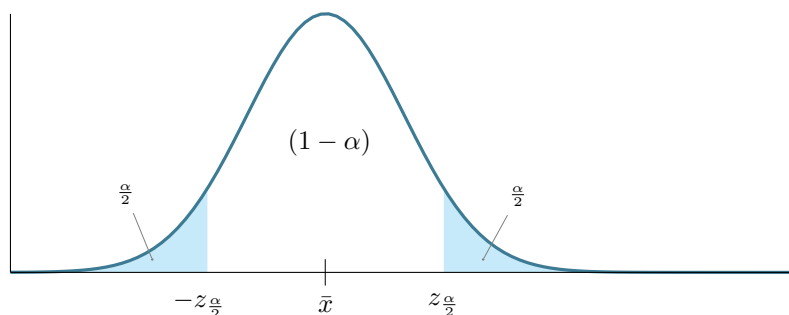
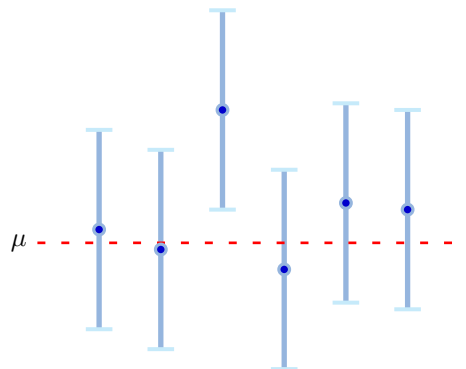


Figura 3.1: Distribuzione di probabilità Normale. Esempio Intervallo di Confidenza

Il coefficiente di confidenza è denotato come $z_{\alpha/2}$. In genere ci informa sulla distanza dalla media in termini di deviazione standard nella normale standardizzata: è il valore della distribuzione che lascia nella coda destra (e sinistra) una probabilità di $\alpha/2$. Vedremo meglio questo concetto nei prossimi capitoli. Dalla figura 3.1 è chiaro che l'ampiezza dell'intervallo costituisce una misura della *precisione* della stima: più stretto è l'intervallo, più precisa è la stima. L'ampiezza dell'intervallo è legata alla probabilità che questo includa il parametro: se ci si accontenta di una probabilità più bassa, l'intervallo è più piccolo e la stima più precisa; se si desidera una probabilità più elevata l'intervallo è più ampio e quindi la stima meno precisa. Nota bene che qui non si parla della probabilità che il parametro della popolazione faccia parte dell'intervallo, ma riguarda la possibilità di estrarre un campione che contenga il parametro della popolazione. Una espressione del tipo: «un livello di confidenza al 95%» significa che esiste il 95% di probabilità che ogni intervallo di confidenza calcolato per i campioni casuali estratti da una stessa popolazione, contenga l'effettivo valore della media della popolazione. Nella seguente figura si osserva come il 95% degli intervalli include la reale media della popolazione, soltanto

uno sembra essere troppo esterno per comprenderla.



Ne segue che è errato dire che «un livello di confidenza al 95%» significa che c'è il 95% di probabilità che la popolazione sia compresa nell'intervallo di confidenza.

Più avanti si affronteranno vari problemi di inferenza dove si tratterà approfonditamente l'argomento. In questo capitolo introduttivo, daremo una definizione generale dell'intervallo di confidenza. Nella stima intervallare viene fissata a priori l'attendibilità (livello di confidenza), e in funzione di essa si determinano gli estremi dell'intervallo di confidenza. Per la sua determinazione è importante conoscere la distribuzione campionaria di uno stimatore corretto ed efficiente del parametro incognito. Indichiamo con $F(t; \theta)$ la funzione di ripartizione dello stimatore T del parametro θ . In base ad essa è possibile ricavare due valori t , espressi in funzione del parametro incognito θ , tali che

$$\Pr\{T < t|\theta\} = F(t; \theta) = \frac{\alpha}{2}$$

$$1 - \Pr\{T < t|\theta\} = 1 - F(t; \theta) = \frac{\alpha}{2}$$

Si indichino tali valori con $t_1(\theta)$ e $t_2(\theta)$. Si otterrà così un intervallo di probabilità pari al livello di confidenza

$$\Pr\{t_1(\theta) \leq T \leq t_2(\theta)\} = 1 - \alpha.$$

L'intervallo di probabilità va interpretato nel senso che al variare del campione, la frazione $1 - \alpha$ delle stime ottenibili sarà compresa in esso. La probabilità di ottenere un campione che fornisca la stima contenuta in detto intervallo è pari a $1 - \alpha$.

Risolvendo la doppia disuguaglianza dell'intervallo di probabilità in funzione di θ , si ottiene l'intervallo di confidenza

$$\Pr\{t_1^{-1}(T) \leq \theta \leq t_2^{-1}(T)\} = 1 - \alpha.$$

Si noti che l'intervallo $[t_1^{-1}(T), t_2^{-1}(T)]$ è un *intervallo casuale*: al variare del campione, lo stimatore T assume valori diversi e quindi anche gli estremi assumono diversi valori. Come abbiamo ampiamente dimostrato, l'intervallo di confidenza va interpretato nel senso che al variare del campione, la frazione $1 - \alpha$ degli intervalli ottenibili includerà il parametro θ , ossia che la probabilità di ottenere un campione che fornisca un intervallo che includa il parametro θ è pari a $1 - \alpha$.

3.7.1 Il margine di errore

Si è visto che l'imprecisione di una stima campionaria (espressa dall'intervallo di confidenza) è una funzione decrescente della attendibilità della stima stessa, misurata dal livello di confidenza. In realtà gli elementi che lo influenzano sono tre: la deviazione standard della popolazione (la sua variabilità), l'ampiezza campionaria e il livello di confidenza (il grado di precisione e livello di attendibilità della stima).

A parità di altri fattori, più si riduce la deviazione standard, più si riduce il margine di errore. La diminuzione della variabilità non è però sempre possibile, è spesso uno degli obiettivi principali delle imprese produttive. Visto che non si può agire direttamente su questa, allora fissando gli altri due termini, si può avere come risultato il terzo. Costruendo l'intervallo di confidenza, abbiamo fissato l'ampiezza del campione n e l'attendibilità della stima o livello di confidenza $1 - \alpha$. Il risultato è l'imprecisione della stima costituita dall'ampiezza dell'intervallo di confidenza. Vogliamo adesso una stima che, con un livello di confidenza $1 - \alpha$ presenti un errore non superiore ad ϵ . Per far questo si può agire sull'aumento della ampiezza del campione. Imponiamo quindi il vincolo ϵ alla semiampiezza dell'intervallo di confidenza

$$\frac{t_2^{-1}(T) - t_1^{-1}(T)}{2} < \epsilon.$$

L'ampiezza del semintervallo di confidenza è approssimativamente pari a $k\sigma/\sqrt{n}$ dove k è una funzione crescente del livello di confidenza e σ^2 è la varianza del carattere della popolazione. Approssimando al semintervallo, si ha quindi che

$$\frac{k\sigma}{\sqrt{n}} < \epsilon,$$

che risolta per n , diventa

$$n > \frac{k^2\sigma^2}{\epsilon^2}. \quad (3.9)$$

La dimensione campionaria necessaria per ottenere nella stima campionaria almeno una determinata precisione, ossia un errore inferiore al margine prefissato ϵ , è una funzione crescente della

variabilità del fenomeno e di una misura della attendibilità desiderata, è invece funzione decrescente del margine di errore massimo desiderato. Tali relazioni in (3.9) sono tutte quadratiche: se lo scarto quadratico medio di un carattere nella popolazione si raddoppia, desiderando mantenere alla stima la stessa attendibilità e la stessa precisione, occorre quadruplicare l'ampiezza campionaria. Se, ancora, volendo ridurre alla decima parte l'errore massimo della stima, ferme restando la variabilità del carattere e l'attendibilità della stima stessa, occorre centuplicare la dimensione campionaria, etc.

Sia X una variabile casuale $X \sim N(\mu, \sigma^2)$ dove σ^2 assumiamo sia nota.

Esempio.

L'intervallo di confidenza per μ al 95% con ampiezza campionaria pari a $n = 64$, è

$$\left(\bar{x} - 1.96\sqrt{\frac{\sigma}{64}}, \bar{x} + 1.96\sqrt{\frac{\sigma}{64}} \right).$$

L'ampiezza dell'intervallo è

$$2 \cdot 1.96 \frac{\sigma}{8}$$

Passiamo adesso ad un intervallo con livello di confidenza 99%, dove $z_{0.995} = 2.58$. In generale avremo

$$2 \cdot 2.58 \frac{\sigma}{\sqrt{n}}.$$

Il problema è mantenere la stessa ampiezza dei due intervalli al variare del livello di confidenza

$$2 \cdot 1.96 \frac{\sigma}{8} = 2 \cdot 2.58 \frac{\sigma}{\sqrt{n}},$$

che, risolvendo, si può ottenere aumentando l'ampiezza campionaria a $n = 111$.

3.8 Verifica di Ipotesi

Partendo dallo studio del campione (casuale) rappresentativo della popolazione, abbiamo stimato la precisione della stima tramite la costruzione degli intervalli di confidenza. In questa sezione introdurremo invece una procedura chiamata *test*, la quale, sulla base dei risultati campionari, suggerisce se accettare o respingere una ipotesi. In particolare, una *ipotesi statistica* è una congettura, una assunzione formulata su un aspetto della popolazione in base a considerazioni teoriche e/o ad informazioni relative a popolazioni analoghe o collegate a quella oggetto di indagine. L'ipotesi di partenza viene chiamata **ipotesi nulla**, indicata con H_0 , questa verrà mantenuta a meno che non venga fuori una prova che evidenzia la sua non veridicità. Se l'ipotesi nulla viene

rigettata, allora viene logicamente accettata l'**ipotesi alternativa** H_1 . L'ipotesi nulla è una opposizione logica della alternativa, come «non nero» è logicamente opposto a «nero». C'è però da dire che una volta formulata l'ipotesi di partenza, la alternativa è selezionato i dati campionari, si può: rifiutare l'ipotesi nulla accettando la alternativa, oppure non rifiutare l'ipotesi nulla. Quando si opta per quest'ultima decisione, si tengono aperti due casi: o l'ipotesi nulla è vera, oppure la procedura di verifica non è stata abbastanza forte da rigettarla causando un errore. Infatti, accanto alle decisioni di accettare l'ipotesi quando è vera, e rifiutarla quando è falsa, possono essere prese, in base al test, le decisioni errate di rifiutare l'ipotesi quando essa è vera, in tal caso si suol dire che si commette un *errore di prima specie*, o di accettarla quando è falsa, nel qual caso si suol dire che si commette un *errore di seconda specie*. Ad esempio, si supponga, in generale, una ipotesi nulla «in media gli studenti al primo anno dell'istituto J pesano 70 chilogrammi», cioè

$$H_0 : \mu = 70.$$

In questo caso si parlerà di **ipotesi semplice**: l'ipotesi nulla afferma che il parametro μ della popolazione è uguale a 70. L'ipotesi alternativa che considera che il peso degli studenti sia superiore 70 chilogrammi

$$H_0 : \mu > 70,$$

viene definita come **ipotesi alternativa unilaterale** e **composta**. Un'altra possibilità è quella della **ipotesi alternativa bidirezionale**, che è l'unione delle due possibili ipotesi direzionali

$$H_0 : \mu \neq 70 \equiv (\mu > 70) \cup (\mu < 70).$$

In genere si usa l'ipotesi alternativa bidirezionale quando il rigetto della ipotesi nulla è dovuto sia a piccoli valori che a valori grandi del test statistico. In questo senso, il processo decisionale usa una **statistica test** che viene costruita considerando gli stimatori puntuali del parametro ed ha una distribuzione campionaria nota. Da questa distribuzione si determinano i valori della statistica test che avrebbero una bassa probabilità di verificarsi nel caso che l'ipotesi nulla fosse vera. Se la statistica test dovesse assumere uno di questi valori, allora l'ipotesi nulla viene rifiutata e accettiamo l'ipotesi alternativa, oppure non rifiutiamo l'ipotesi nulla. Risulta chiaro che il campo di variazione di una statistica test viene suddiviso in due regioni complementari: la *regione di accettazione* A e la *regione di rifiuto* R . La regola di decisione consiste quindi nell'accettare H_0 se la statistica test $u = f(x_1, \dots, x_n)$, calcolata per le osservazioni campionarie, cade nella regione di accettazione $u \in A$, oppure se rifiutarla proponendo H_1 , se il test conduce ad un valore che cade nella regione critica $u \in R$. Il problema dell'errore si ha perché, fissata

una statistica test, bisogna individuare la regione critica (si vedrà approfonditamente nei capitoli successivi): R è costituita da valori più piccoli, più grandi di u o da entrambi? La probabilità dell'errore di prima specie, chiamata **livello di significatività** α , consente l'individuazione del punto (o punti) di confine tra la regione critica e quella di accettazione della ipotesi nulla. Inoltre $(1 - \alpha)$ è la probabilità di non rifiutare l'ipotesi nulla quando è vera.

Più grande è R , maggiore è la probabilità di commettere un errore di prima specie: il livello di significatività α stabilisce la dimensione della regione critica R , e quindi anche quella di accettazione A che ne è complementare. L'esigenza di evitare di incorrere nell'errore e di rifiutare una ipotesi vera, potrebbe suggerire di fissare $\alpha = 0$ annullando la probabilità di commettere tale errore. Così facendo però si accresce la probabilità di accettazione della ipotesi nulla anche quando essa è falsa, incorrendo nell'errore di seconda specie β : diminuendo α si accresce β e viceversa. Il compromesso tra la minimizzazione dei due errori, si ottiene fissando a priori il livello di significatività ad un valore sufficientemente basso (in genere $\alpha = 0.05$, oppure $\alpha = 0.01$); la probabilità di commettere un errore di seconda specie β risulta anch'essa stabilita in via residuale. Infine, si chiama **potenza del test** la probabilità $(1 - \beta)$ di rifiutare una ipotesi nulla quando è falsa.

In generale, la formula della statistica test è

$$\text{statistica test} = \left[\frac{\text{statistica di interesse} - \text{parametro ipotizzato}}{\text{errore standard della statistica di interesse}} \right].$$

Ovviamente, come nel problema di stima in cui si possono scegliere più stimatori, per la verifica di una ipotesi parametrica possono essere usate diverse statistiche test. Quella da scegliere, a parità di livello di significatività e degli altri elementi, è quella che rende minima la probabilità di commettere un errore di seconda specie β , massimizzando la potenza del test. Il test dotato di tale caratteristica viene chiamato *test più potente* (similmente all'individuazione dello stimatore più efficiente). Chiameremo *consistenti* quei test per i quali, al divergere della dimensione del campione, la potenza tende all'unità. Questo è importante perché, a parità di livello di significatività, all'aumentare del campione, la potenza test aumenterà.

Verifica delle Ipotesi e Processo Penale C'è una certa analogia fra il *test* statistico di verifica delle ipotesi e il processo penale:

- nel processo penale vige la presunzione di innocenza, che sarebbe l'ipotesi nulla: l'imputato non ha commesso reato. Vi è anche l'ipotesi alternativa: è colpevole, ed è per questo che

il presunto criminale viene arrestato. L'ipotesi alternativa è proprio quella che si vuole andare a testare.

- Esiste in entrambi i casi una regola di decisione. La sua definizione è molto importante, perché nel processo penale, come in statistica, non disponiamo di prove assolute. C'è quindi bisogno di un processo decisionale generale tramite cui poter decidere sul rigetto della ipotesi nulla. Nel sistema penale questo è il *ragionevole dubbio*, nel senso che l'ipotesi nulla deve essere rigettata a meno che ci sia una evidenza di colpevolezza che va oltre ogni ragionevole dubbio. Anche in statistica, l'ipotesi nulla è quella che deve essere considerata vera a meno che i dati contengano una forte evidenza empirica per rifiutarla.
- Il processo di decisione si basa su un insieme di dati. Per entrambi esistono metodi generali per la loro raccolta. Come in statistica, anche nel processo penale si lavora su dati *campionari* visto che nel modo reale si potrà utilizzare solo una informazione parziale.

3.8.1 Il p-value

Sappiamo che bisogna predeterminare il livello di significatività α per cui l'ipotesi nulla veniva rifiutata. Capita che H_0 viene rifiutata a livello di significatività 0.05, ma non a un livello 0.005. Riducendo livelli via via inferiori ridurremmo la probabilità di rifiutare una ipotesi nulla vera, aumentando però la probabilità dell'errore di secondo tipo. Il p -value è il livello di significatività osservato, ossia la probabilità di ottenere un valore della statistica test uguale o più estremo del valore osservato, calcolata pensando vera l'ipotesi nulla. Si rifiuta l'ipotesi nulla se il p -value $< \alpha$. Il p -value fornisce l'esatta probabilità di rifiuto della ipotesi nulla dato il valore della statistica campionaria osservata, infatti, essendo una probabilità osservata, esso è diverso per ogni valore della statistica campionaria. Il p -value è dato da

$$P(\bar{X} \geq \bar{x} \mid H_0 : \mu = \mu_0)$$

cioè $P(Z \geq z_p)$, dove

$$z_p = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

A differenza di α che è predeterminato, il p -value fornisce l'esatta probabilità di rifiuto della ipotesi nulla derivante dai dati campionari. Se in un test con un livello di significatività $\alpha = 0.05$, l'ipotesi nulla viene rifiutata, allora ottenendo un p -value più piccolo significa che dai risultati campionari avremmo rifiutato l'ipotesi nulla ad un livello di significatività, ad esempio, $\alpha_p = 0.006$. Questo ci fa capire che il p -value ci informa anche quanto vicini o lontani ci troviamo

dal rifiuto dell'ipotesi nulla. Infatti con un p -value 0.07, non avremmo rifiutato l'ipotesi nulla, ma avremmo saputo di essere molto vicini alla regione di rifiuto.

Capitolo 4

INFERENZA SU MEDIE

In questo capitolo si applicheranno le definizioni discusse nel capitolo precedente, in modo da risolvere problemi di inferenza sul parametro *media della popolazione* μ . Si supponga un campione casuale di n osservazioni estratte da una grande popolazione con media μ e varianza σ^2 . Indicheremo i singoli elementi del campione con $X_1 \dots X_n$. Per generare una stima sulla media della popolazione μ , sappiamo che bisogna utilizzare la media campionaria

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Questa statistica è uno stimatore corretto in quanto il valore medio dello stimatore \bar{X} coincide con il parametro μ , cioè

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{\sum X_i}{n}\right) = \frac{\sum \mathbb{E}(X_i)}{n} = \frac{n\mu}{n} = \mu.$$

Avendo anche una varianza finita, la media campionaria sarà anche uno stimatore consistente di μ : aumentando l'ampiezza campionaria, la varianza dello stimatore media campionaria tende a zero, formalmente

$$\begin{aligned} \sigma_{\bar{X}}^2 &= \mathbb{E}[(\bar{X} - \mu)^2] = \mathbb{E}\left[\left(\frac{\sum X_i}{n} - \mu\right)^2\right] = \\ &= \mathbb{E}\left\{\frac{[\sum (X_i - \mu)]^2}{n^2}\right\} = \\ &= \mathbb{E}\left[\frac{\sum (X_i - \mu)^2 + \sum \sum_{i \neq j} (X_i - \mu)(X_j - \mu)}{n^2}\right] = \\ &= \frac{\sum \mathbb{E}[(X_i - \mu)^2] + \sum \sum_{i \neq j} \mathbb{E}(X_i - \mu)(X_j - \mu)}{n^2} = \\ &= \frac{n\sigma^2 + n(n-1) \cdot 0}{n^2} = \frac{\sigma^2}{n} \end{aligned} \tag{4.10}$$

4.1 Intervallo di confidenza per la media della popolazione

Come abbiamo già visto nel capitolo precedente, una volta sintetizzate le osservazioni campionarie tramite una statistica (in questo capitolo la media campionaria), bisogna individuare un intervallo (di confidenza), il quale con un alto grado di attendibilità (livello di confidenza) potrebbe essere uno dei possibili intervalli casuali che includono la media incognita μ . In questa sede consideriamo due semplici casi in cui il carattere della popolazione è distribuito normalmente con media μ incognita. Si considereranno inoltre due sottocasi: *a*) la varianza σ^2 è nota, *b*) la varianza è incognita come la media.

4.1.1 a) Varianza nota

Come abbiamo appena visto, la statistica media campionaria \bar{X} è una v.a. normale di media μ e varianza σ^2/n . La standardizzazione segue quindi

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}. \quad (4.11)$$

Per la variabile normale standardizzata, fissata una probabilità $1 - \alpha$, tramite le tavole, è possibile ricavare gli estremi $-z_\alpha$, $+z_\alpha$ dell'intervallo simmetrico intorno a zero di probabilità $1 - \alpha$

$$P(-z_\alpha < Z < +z_\alpha) = 1 - \alpha,$$

da cui si ricava l'evento complementare

$$P(Z < -z_\alpha) + P(Z > z_\alpha) = \alpha.$$

Data la simmetria della curva normale standardizzata

$$P(Z < -z_\alpha) = P(Z > z_\alpha) = \frac{\alpha}{2},$$

che, come abbiamo visto, consente di ricavare z_α usando la tavola dell'integrale della curva normale standardizzata. Ad esempio per $1 - \alpha = 0.95$, $\alpha = 0.05$ e $\alpha/2 = 0.025$, la tavola fornisce $z_{0.025} = 1.96$. Solitamente, i tre livelli di confidenza più utilizzati sono

Livelli di confidenza $1 - \alpha$	Estremi $\pm z_\alpha$
0.95	± 1.96
0.99	± 2.58
0.999	± 3.29

Utilizzando l'equazione (4.11)

$$P\left\{-z_\alpha < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < +z_\alpha\right\} = 1 - \alpha,$$

da cui, con semplicissimi passaggi algebrici, si ottiene

$$P\left(\mu - z_\alpha \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Questo intervallo è però $\mu \pm z_\alpha \sigma / \sqrt{n}$, che ci informa che con probabilità $1 - \alpha$, l'*intervallo fisso* potrà comprendere il *punto casuale* media campionaria \bar{X} . Risolvendo la doppia disuguaglianza abbiamo dunque

$$P\left(\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha, \quad (4.12)$$

che è l'*intervallo casuale* $\bar{X} \pm z_\alpha \sigma / \sqrt{n}$, che con probabilità $1 - \alpha$ potrà comprendere il *punto fisso* media della popolazione μ .

A posteriori, in presenza delle osservazioni campionarie x_1, \dots, x_n , si può avere fiducia che l'intervallo di confidenza $\bar{x} \pm z_\alpha \sigma / \sqrt{n}$, sia uno di quelli che includono la media μ della popolazione. L'affermazione che questo intervallo include μ sarà corretta nella frazione $1 - \alpha$ delle volte, ed errata nella frazione α delle volte.

Esempio Si supponga di voler determinare l'intervallo di confidenza al livello 95% per il peso medio delle barrette prodotte da una determinata macchina. Assumendo che gli errori siano accidentali, è ragionevole supporre che la distribuzione del peso in grammi (g) delle barrette sarà normale. Non conosciamo la media ma, teoricamente (in pratica è raro), conosciamo la varianza $\sigma = (0.09g)^2$. Si ha un campione casuale di $n = 20$ barrette scelte casualmente, e si scopre che la media campionaria è $\bar{x} = 4.95g$. A questo punto abbiamo tutti gli elementi per stimare l'intervallo di confidenza. Infatti $\bar{x} = 4.95g$, $1 - \alpha = 95\%$, e quindi la soglia sarà $z_\alpha = z_{0.05} = 1.96$, inoltre lo scarto quadratico medio sarà $\sigma = \sqrt{0.09} = 0.3g$, e l'ampiezza campionaria sarà $n = 20$. Utilizzando l'equazione (4.12)

$$P\left(4.95 - 1.96 \cdot \frac{0.3}{\sqrt{20}} < \mu < 4.95 + 1.96 \cdot \frac{0.3}{\sqrt{20}}\right) = 0.95$$

cioè

$$P(4.82 < \mu < 5.08) = 0.95.$$

Tale intervallo, con la probabilità del 95%, potrebbe essere uno di quelli che includono il peso medio di tutte le barrette prodotte da quella macchina.

4.1.2 b) Varianza incognita

Quando la varianza non è nota, non si può utilizzare la distribuzione normale standard della popolazione, ma una distribuzione normale standard definita nel campione. In questo caso la varianza della popolazione σ^2 verrà stimata in base alla varianza del campione s^2 . Ma c'è un problema: abbiamo visto che per determinare la probabilità che la media della popolazione sia compresa in un intervallo attorno la media del campione, si assume che il campione sia distribuito normalmente, con deviazione standard s/\sqrt{n} , dove s è la deviazione standard del campione, utilizzando poi le tavole statistiche. Ma come rilevato da Gosset (che pubblicava con lo pseudonimo di Student), al diminuire del numero degli esperimenti, il valore della deviazione standard del campione diventa esso stesso soggetto ad un errore crescente. Si potrebbe risolvere il problema ripetendo moltissimi esempi, ma questo non sempre è possibile. Vediamo come risolvere il problema.

Dalla statistica descrittiva, sappiamo che la varianza campionaria è data da

$$s^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}.$$

Troviamo quindi quel $n - 1$ al denominatore e non n come nella varianza della popolazione. Si è dimostrato che la varianza campionaria è uno stimatore non distorto della varianza della popolazione per $n - 1$. Si tenga presente che qui si sta misurando la variabilità del campione. In generale, geometricamente la variabilità di un punto è data dalla sua distanza rispetto alla “origine” del sistema di riferimento. Se il sistema di riferimento è unidimensionale, bidimensionale o tridimensionale etc, si deve tenere conto della distanza rispetto alla origine che sarà diversa per il caso bidimensionale rispetto a quello tridimensionale e così via. Si calcolano così le somme dei quadrati delle distanze normalizzandoli per le dimensioni dello spazio di riferimento. In questo modo la variabilità terrà conto delle diverse coordinate. Quel $n - 1$ fa esattamente la stessa cosa, sono i “gradi di libertà”, cioè si deve tenere conto dello spazio entro il quale le osservazioni possono variare. La distribuzione che considera una varianza campionaria è la t -Student, che identificheremo con T , la quale è simmetrica e con ampiezza dipendente dal numero di osservazioni del campione. Per campioni molto grandi, la v.a. T tende ad essere una normale standard Z , dove

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}.$$

Il ragionamento è analogo alla normale standardizzata con varianza nota, infatti tramite le tavole è possibile ricavare gli estremi $-t_{n-1,\alpha}$ e $+t_{n-1,\alpha}$ dell'intervallo simmetrico intorno allo zero di

probabilità $1 - \alpha$

$$P\left(\bar{X} - t_{n-1,\alpha} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,\alpha} \frac{s}{\sqrt{n}}\right) = 1 - \alpha,$$

che esprime che l'intervallo casuale $\bar{X} \pm t_{n-1,\alpha} s/\sqrt{n}$, con probabilità $1 - \alpha$, potrà comprendere la media della popolazione μ . L'intervallo di confidenza è

$$\bar{x} \pm t_{n-1,\alpha} \frac{s}{\sqrt{n}},$$

dove \bar{x} ed s sono i valori della media campionaria e della varianza campionaria corretta, calcolati sui valori osservati del campione x_1, x_2, \dots, x_n . Con probabilità α potrebbe essere un campione di quelli che includono il parametro incognito μ .

Esempio Determiniamo l'intervallo di confidenza per la quantità media annuale di precipitazioni (in mm) sulla città di Reggio Calabria. Estratto un campione casuale di 15 osservazioni (che non riporteremo tutte per questioni di brevità) si calcoleranno la media e la varianza campionaria corretta

$$\bar{x} = \frac{442.8 + 767.6 + \dots + 671.2}{15} = 584.1$$

$$s^2 = \frac{(442.8 - 584.1)^2 + (767.6 - 584.1)^2 + \dots + (671.2 - 584.1)^2}{14} = 15221.11 = (123.4\text{mm})^2$$

dove l'intervallo di confidenza richiesto, atteso che dalle tavole della distribuzione T per $1 - \alpha = 0.95$ e per $n - 1 = 14$ gradi di libertà si legge $t_{14,0.05} = 2.14$, risulta

$$584 \pm 2.14 \times \frac{123.4}{14}$$

cioè l'intervallo da 513.5 a 654.7 con probabilità al 95%, potrebbe essere uno di quelli che includono il valor medio della v.a. "precipitazione annua sulla città di Reggio Calabria", ossia la media di tale fenomeno meteorologico in una lunghissima serie di anni. Infine, si ricordi che in virtù della convergenza in legge della v.a. T alla v.a. Z , per campioni di ampiezza maggiori di 30 unità, con buona approssimazione, è possibile utilizzare le soglie della curva normale standardizzata.

4.2 Dimensionamento del Campione

Abbiamo visto che la precisione di una indagine campionaria tendente a stimare, attraverso la media del campione \bar{x} , la media di un carattere della popolazione μ , è data dalla ampiezza dell'intervallo di confidenza

$$\bar{x} + z_\alpha \frac{s}{\sqrt{n}} - \left(\bar{x} - z_\alpha \frac{s}{\sqrt{n}}\right) = 2z_\alpha \frac{s}{\sqrt{n}}$$

dove la semiampiezza dell'intervallo è

$$z_\alpha \frac{s}{\sqrt{n}}.$$

La semiampiezza esprime l'errore per eccesso o per difetto, cioè lo scostamento positivo o negativo tra \bar{x} e μ , che, ad un prefissato livello di confidenza $1 - \alpha$, si deve attendere dall'indagine campionaria. Tale errore dipende dalla variabilità del fenomeno misurata da s sulla quale non c'è modo di influire, dal livello di confidenza $1 - \alpha$, che viene prefissato in relazione alle esigenze pratiche di attendibilità del problema inferenziale in esame e dalla ampiezza campionaria n . Effettuando una indagine e misurandone a posteriori la precisione si corre il rischio di avere risultati poco precisi e quindi inutilizzabili sotto il profilo pratico o di avere risultati di precisione eccessiva, ottenuti con una ampiezza campionaria superiore alle necessità, con aggravio di costi e tempi.

Si ha quindi l'esigenza di fissare, in sede di pianificazione della indagine campionaria, l'ampiezza del campione in relazione alla precisione desiderata. Indicando con ϵ l'errore massimo che si è disposti a commettere, occorre scegliere l'ampiezza campionaria n in maniera tale che l'errore della indagine, al livello di confidenza stabilito, risulti inferiore al margine prefissato ϵ , cioè

$$z_\alpha \frac{s}{\sqrt{n}} z_\alpha \epsilon,$$

dal quale si ottiene

$$n > \frac{z_\alpha^2 s^2}{\epsilon^2}. \quad (4.13)$$

Per l'applicazione della precedente formula, occorre preliminarmente stimare la varianza campionaria corretta attraverso una indagine pilota condotta su poche unità.

Esempio Si vuole dimensionare una indagine tendente a stimare il numero medio di ore giornaliere passate a studiare dagli studenti del primo anno in Scienze Economiche del Di.Gi.Ec. al livello di confidenza del 95%, con un errore massimo di mezzora $\epsilon = 0.5$. Una indagine preliminare condotta su 20 studenti ha fornito una stima $\sigma = (4.8)^2$ ore. Applicando l'equazione (4.13)

$$n = \frac{1.96^2 \cdot 4.8^2}{0.5} = 354$$

la quale ci permette di accertare che un campione di 354 studenti dovrebbe consentirci di valutare il numero medio di ore al giorno passate a studiare dalla popolazione di giovani considerata, con la precisione e l'attendibilità desiderate.

4.3 Verifica delle ipotesi sul valore di μ

Formulata l'ipotesi che la media della popolazione sia pari al valore μ_0 , cioè formulata l'ipotesi nulla

$$H_0 : \mu = \mu_0$$

e scelta una ipotesi alternativa fra

$$H_1 : \mu > \mu_0$$

$$H'_1 : \mu < \mu_0$$

o ancora

$$H''_1 : \mu \neq \mu_0$$

che sia più confacente in relazione al problema in esame, occorre individuare la funzione test, ossia la statistica sulla base della quale costruire la regola di decisione di accettare o rifiutare H_0 . La funzione test che sceglieremo è la media del campione \bar{X} .

Se l'ipotesi alternativa è H_1 , la regione critica del test è costituita dalla coda destra della distribuzione della media campionaria \bar{X} , come si può agevolmente intuire tenendo conto del fatto che valori grandi di \bar{X} appaiono più verosimili se è vera $H_1 : \mu > \mu_0$ piuttosto che se fosse vera $H_0 : \mu = \mu_0$.

Se l'ipotesi alternativa è H'_1 , la regione critica è costituita dalla coda sinistra della distribuzione della media campionaria \bar{X} , atteso che valori piccoli di \bar{X} appaiono più verosimili se fosse vera $H'_1 : \mu < \mu_0$ piuttosto che $H_0 : \mu = \mu_0$.

Infine, se l'ipotesi alternativa è H''_1 , la regione critica è costituita sia dalla coda di destra, che dalla coda di sinistra della distribuzione della media campionaria \bar{X} , atteso che tanto i valori piccoli che i valori grandi di \bar{X} appaiono più verosimili nel caso fosse vera $H''_1 : \mu \neq \mu_0$.

4.3.1 a) Varianza nota

Se il carattere nella popolazione è normale e se è nota la varianza σ^2 , si può operare sulla funzione test

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},$$

se è vera l'ipotesi nulla $H_0 : \mu = \mu_0$, è una v.a. normale standardizzata.

Nel caso della ipotesi alternativa unidirezionale $H_1 : \mu > \mu_0$, la regione critica è costituita dai valori di Z che superano il valore z_α . Si rifiuta H_0 se

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha,$$

o equivalentemente se

$$\bar{x} > \bar{x}_c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}},$$

dove \bar{x}_c è il valore critico. Vediamo adesso un esempio.

Esempio *Vi è un processo produttivo che produce una media di 80 unità orarie con $\sigma = 8$. Non si passerà a un nuovo processo produttivo a meno di una forte evidenza empirica che questo determini un aumento della produzione media. (Newbold - Carlson - Throne, Statistica)*

Si determini

$$H_0 : \mu \leq 80$$

$$H_1 : \mu > 80.$$

Si fissi un livello di significatività $\alpha = 0.05$, cioè che il nuovo processo ha una produttività più elevata, la probabilità di errore sarà minore o uguale a 0.05. Si estrae un campione di $n = 25$ ore di produzione con il nuovo processo e si calcola la media campionaria \bar{x} delle unità prodotte in ogni ora selezionata. Con un livello di significatività $\alpha = 0.05$, la regola di decisione risulta essere: si rifiuta H_0 se

$$\frac{\bar{x} - 80}{8/\sqrt{25}} > 1.645,$$

con $z_{0-05} = 1.645$ ottenuto dalla tavola della distribuzione normale standardizzata. Alternativamente, si rifiuta H_0 se

$$\bar{x} > 80 + 1.645 \times \frac{8}{\sqrt{25}} = 82.63,$$

cioè $\bar{x}_c = 82.63$.

Si supponga che la media campionaria $\bar{x} = 83$. In questo caso $\bar{x} > 82.63$ e

$$z = \frac{83 - 80}{8/\sqrt{25}} = 1.875 > 1.645$$

rifutando H_0 avendo evidenza empirica che avvalora l'ipotesi che il nuovo processo produttivo aumenterà la produttività. Calcolando il p -value e osservando la tavola della normale standardizzata

$$p\text{-value} = P(\bar{X} \geq 83 \mid H_0 : \mu = 80) = P(Z \geq 1.875) = 0.03,$$

e quindi la probabilità di commettere un errore del primo tipo è pari a 0.03. Si ricordi che dalla tavola si ha che $P(Z \leq 1.875) = 0.9693$, da cui $P(Z \geq 1.875) = 1 - 0.9693 \approx 0.03$.

Nel caso invece della ipotesi alternativa $H'_1 : \mu < \mu_0$, la regione critica del test è costituita da valori di Z inferiori a $-z_\alpha$, e la regione di accettazione da quelli maggiori o uguali a $-z_\alpha$. Bisognerà quindi seguire lo stesso ragionamento precedente, ma rifiutando H_0 se

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z_\alpha$$

o equivalentemente se

$$\bar{x} < \bar{x}_c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Nel caso infine della ipotesi alternativa $H''_1 : \mu \neq \mu_0$, la regione critica del test è costituita dai valori di Z inferiori a $-z_\alpha$ e da quelli superiori a $+z_\alpha$, mentre la regione di accettazione è racchiusa nell'intervallo $-z_\alpha, +z_\alpha$. L'ipotesi nulla è quindi messa in dubbio se la media campionaria è molto più grande o molto più piccola di μ_0 . Con più precisione, diciamo che se la distribuzione è normale con deviazione standard σ , allora si ha un livello di significatività α dove

$$P(Z > z_{\alpha/2}) = \frac{\alpha}{2} \text{ e } P(Z < -z_{\alpha/2}) = -\frac{\alpha}{2}$$

dove si intuisce che il livello di significatività α è egualmente diviso fra le due code. Infatti la probabilità che Z superi $z_{\alpha/2}$ o sia inferiore $-z_{\alpha/2}$ vale α . La regola di decisione sarà quindi: si rifiuta H_0 se

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2} \quad \text{oppure} \quad \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2}$$

o analogamente se

$$\bar{x} < \mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{oppure} \quad \bar{x} > \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Per calcolare il p -value, la probabilità di ottenere valori estremi si riferisce alla somma delle probabilità nelle code di destra e sinistra della distribuzione, e sarà perciò ottenuta raddoppiando i valori dei test unilaterali, cioè

$$p\text{-value} = 2P(Z \geq z_{p/2}) \quad \text{dove} \quad z_{p/2} = \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}.$$

Questa probabilità rappresenta il livello di significatività al quale l'ipotesi nulla può essere rifiutata, dato il valore della statistica campionaria osservata.

Esempio *Si assuma che per uno studio particolare si ha una distribuzione normale con media 2, e deviazione standard 0.06. La media campionaria di un campione casuale di nove misurazioni è invece $\bar{x} = 1.95$. Si assuma inoltre un livello di significatività $\alpha = 0.05$*

In questo caso

$$H_0 : \mu = 2$$

mentre

$$H_1 : \mu \neq 2.$$

Seguendo le formule relative alla verifica delle ipotesi bilaterale, con un livello di significatività $\alpha = 0.05$ dalle tavole segue che $z_{0.05/2} = 1.96$. Dato che

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.95 - 2}{0.06/\sqrt{9}} = -2.50,$$

allora l'ipotesi nulla viene rifiutata, infatti $-2.50 < -1.96$. Per calcolare il p -value, dalla tavola risulta che la probabilità di ottenere valori $P(Z < -2.50) = 0.0062$. Visto che il test è bilaterale, allora il p -value deve considerare le due code, cioè il

$$p\text{-value} = 0.0062 + 0.0062 = 0.0124$$

cioè, un livello di significatività maggiore dell'1.24% avrebbe determinato il rifiuto della ipotesi nulla.

4.3.2 a) Varianza incognita

Il ragionamento è praticamente identico al caso di varianza conosciuta, si usa però in questo caso la varianza campionaria.

Se il carattere nella popolazione è normale ma non si conosce la varianza della popolazione σ^2 , che deve essere stimata in base alla varianza campionaria corretta s^2 , si può far uso della funzione test

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

che, se vera l'ipotesi nulla $H_0 : \mu = \mu_0$, è una variabile casuale t -Student con $n - 1$ gradi di libertà.

Nel caso della ipotesi alternativa unidirezionale, $H_1 : \mu > \mu_0$ la regione critica del test è costituita dai valori di T che superano $+t_{n-1,\alpha}$ e la regione di accettazione da quelli che sono inferiori o uguali a $+t_{n-1,\alpha}$.

Nel caso dell'ipotesi alternativa $H_1' : \mu < \mu_0$, la regione critica del test è costituita dai valori di T inferiori a $-t_{n-1,\alpha}$ e la regione di accettazione da quelli maggiori o uguali a $-t_{n-1,\alpha}$.

Nel caso, infine, della ipotesi alternativa $H_1'' : \mu \neq \mu_0$ la regione critica del test è costituita dai valori di T inferiori a $-t_{n-1,\alpha/2}$ e da quelli superiori a $+t_{n-1,\alpha/2}$. La regione di accettazione da

quelli che rientrano nell'intervallo $-t_{n-1,\alpha}, +t_{n-1,\alpha}$.

Per campioni di ampiezza $n > 30$ con buona approssimazione è possibile sostituire alle soglie della T di Student, quelle della variabile casuale normale standardizzata Z .

Appendice

C Probabilità: Assiomi

La probabilità fa un grande uso delle operazioni fra insiemi. Un **insieme** è una collezione di oggetti chiamati **elementi**, se A è un insieme e a un suo elemento, allora $a \in A$ (in caso contrario $a \notin A$). Se non vi sono elementi in A , allora è vuoto, $A = \emptyset$. Se ogni elemento di A_1 è un elemento di A_2 , allora $A_1 \subset A_2$. Se $A_1 \subset A_2$ e $A_2 \subset A_1$, allora i due insiemi sono uguali $A_1 = A_2$. Si può definire un insieme universale Ω di cui fanno parte gli oggetti di interesse. In questo caso, considereremo tutti gli insiemi A_i con $i = 1, 2, \dots$ che sono sottospazi di Ω . Si definisca **esperimento aleatorio**, un processo le cui possibili realizzazioni sono due o più risultati di cui non si può prevedere con certezza quale si realizzerà. Sia Ω uno spazio campionario, cioè l'insieme di tutte le possibili realizzazioni di un esperimento. La legge di probabilità assegna all'insieme A di possibili realizzazioni (chiamati eventi) un numero non negativo $P(A)$ (probabilità di A). In particolare, un **evento** si verifica quando il risultato dell'esperimento aleatorio è uno degli eventi che lo costituiscono.

Un insieme è definito come **complementare** A^c , quando $\{a \in \Omega \mid a \notin A\}$. Si noti che $\Omega^c = \emptyset$. L'**unione** di due insiemi A_1 e A_2 , definita con $A_1 \cup A_2$ è l'insieme di tutti gli eventi di Ω che appartengono ad almeno uno dei due eventi, $A_1 \cup A_2$ si verifica se e solo se A_1 o A_2 , oppure entrambi si verificano. L'**intersezione** fra due insiemi A_1 e A_2 , definita come $A_1 \cap A_2$, è l'insieme di elementi che appartengono a entrambi gli insiemi. Due insiemi si dicono **disgiunti** se fra loro non ci sono elementi in comune $A_1 \cap A_2 = \emptyset$. In questo caso i due insiemi si dicono **mutuamente esclusivi**. Mentre se dati n eventi A_1, \dots, A_n dello spazio campionario Ω , se la loro unione $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$, allora questi n eventi sono detti **collettivamente esaustivi**. Nella figura 1 si possono osservare alcune importanti operazioni fra insiemi.

Sia Ω uno spazio campionario finito e A un suo evento. La probabilità dell'evento A è così un numero reale che segue tre assiomi:

Assioma 1 $P(A) \geq 0$;

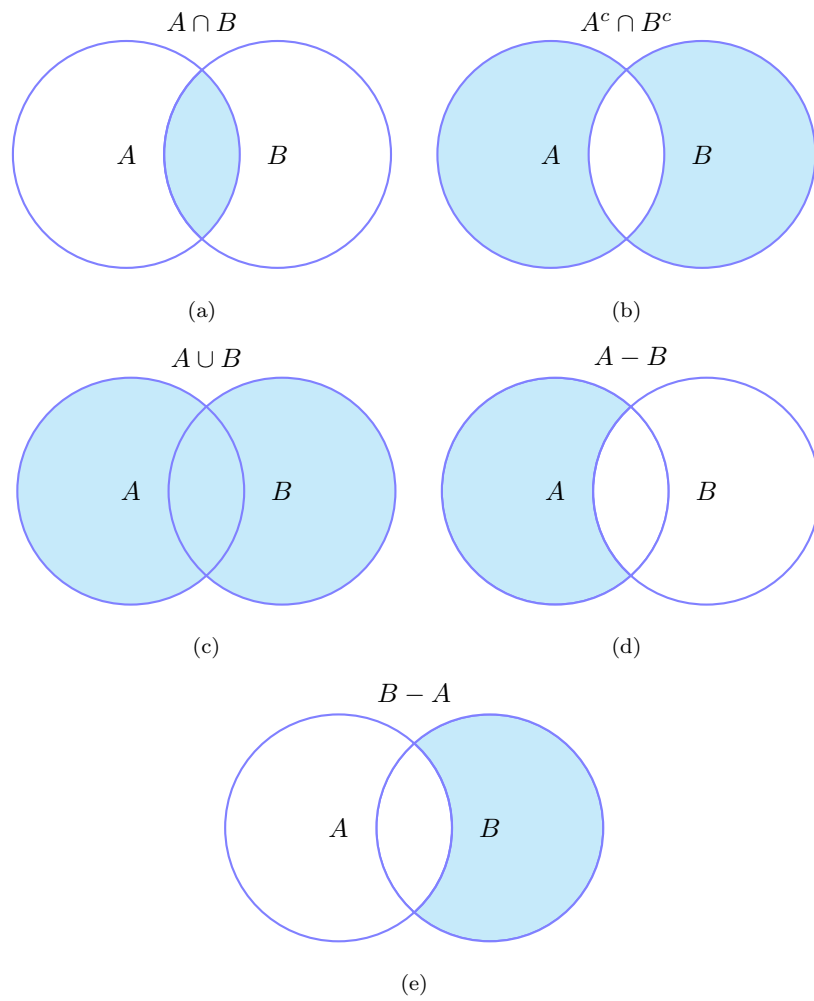


Figura 1: Alcuni diagrammi di Venn per due insiemi $A, B \in \Omega$.

Assioma 2 $P(\Omega) = 1$ (normalizzazione);

Assioma 3 $P(A \cup B) = P(A) + P(B)$ se $A \cap B = \emptyset$.

Dagli assiomi derivano le seguenti proprietà

1. $P(A^c) = 1 - P(A)$;
2. $P(\emptyset) = 0$;
3. $P(A) \leq P(B)$ se $A \subset B$;
4. $P(A) \leq 1$;

5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

6. Se A_1, A_2, \dots, A_n sono n eventi di Ω , allora

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \sum_{i \neq j \neq h} P(A_i \cap A_j \cap A_h) - \dots (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n);$$

7. Se gli eventi sono tutti disgiunti ($A_i \cap A_j = \emptyset, \forall i \neq j$), allora

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$