Statistica II
Appunti del corso

Claudio Ortelli

Semestre Autunnale
2011

Indice

Capitolo 1

Introduzione

1.1 Ricapitolazione

Nel corso introduttivo di Statistica I sono stati trattati i concetti fondamentali dell'analisi statistica e della teoria della probabilità. In particolare avete visto
Capitolo 1: introduzione

1. Popolazione di riferimento di un'analisi statistica.
2. Il campione.

Capitolo 2: elementi di statistica descrittiva

1. Le distribuzioni di frequenze (assolute o relative), istogrammi.

Capitolo 3: misure empiriche di centralità e variabilità

1. Indicatori di centralità (moda, media e mediana) e variabilità (range, intervallo interquartile, deviazione standard). Regole di sommatoria e relativa notazione.

$$\overline{x},\ \sum_{i\ \text{pari}} x_i,\ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Capitolo 4: Elementi di probabilità I

1. Spazio campionario Ω: insieme di tutti gli esiti di un esperimento statistico. Evento $E \subset \Omega$: un particolare sottoinsieme di Ω. E^c: il complemento dell'evento E. Prime regole di calcolo:

$$P(A) + P(A^c) = 1.$$

Capitolo 5: Elementi di probabilità II

1. Intersezione di eventi. $A \cap B$: l'insieme degli esiti dell'esperimento che appartengono sia all'evento A che all'evento B.

2. Eventi indipendenti: A e B sono indipendenti se vale la seguente condizione
$$P(A \cap B) = P(A)P(B).$$

3. Probabilità condizionata
$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}.$$

4. Unione di eventi. $A \cup B$: l'insieme degli esiti dell'esperimento che appartengono all'evento A o all'evento B o ad entrambi.

5. Eventi mutualmente esclusivi. A e B sono mutualmente esclusivi se non hanno esiti in comune, ovvero se $A \cap B = \emptyset$.

6. Probabilità di unioni di eventi A e B mutualmente esclusivi:
$$P(A \cup B) = P(A) + P(B)\,.$$

7. Probabilità di unioni di eventi:
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,.$$

8. Diagrammi ad albero per la rappresentazione dei possibili esiti di un esperimento e per il calcolo delle probabilità corrispondenti.

9. Teorema di Bayes.

Capitolo 6: Introduzione alle variabili aleatorie

1. Esperimento statistico: procedimento di osservazione di un dato fenomeno. Variabile aleatoria (*V.A.*) X. Definizione di variabili aleatorie discrete e continue. Distribuzioni e funzioni di probabilità.

2. Valore atteso e varianza di una variabile *V.A* discreta X:
$$\begin{aligned} \mu &= \sum_{i=1}^{n} x_i p(x_i) \\ \sigma^2 &= \sum_{i=1}^{n} (x_i - \mu)^2 \, p(x_i) \end{aligned}$$
dove p è la funzione di probabilità di X e $x_1, \ldots, x_n$ sono le possibili osservazioni (realizzazioni) di X.

3. Regole di calcolo:

$$\begin{aligned} \mu_{X+bY} &= \mu_X + b\mu_Y \\ \sigma^2_{a+bX} &= b^2\sigma^2_X \\ \sigma^2_X &= \mu_{X^2} + (\mu_X)^2 \end{aligned}$$

4. Distribuzione di Bernoulli e distribuzione uniforme.

Capitolo 7: alcune variabili aleatorie discrete

1. Tecniche combinatorie: fattoriali, permutazioni e combinazioni.
2. Esperimenti binomiali e distribuzioni binomiali. Valore atteso e varianza di distribuzioni binomiali.
3. Esperimenti e distribuzioni di Poisson. Valore atteso e varianza di distribuzioni di Poisson.

Capitolo 8: variabili aleatorie continue I

1. Variabili aleatorie continue e funzione di densità.
2. Distribuzione uniforme e triangolare.

Capitolo 9: variabili aleatorie continue II

1. Densità di una distribuzione normale.
2. Valore atteso e varianza di distribuzioni normali.
3. Aree e probabilità di densità normali.

Capitolo 10: variabili aleatorie continue III

1. Unità standard e distribuzione normale standard.
2. Trasformazioni di variabili distribuite in modo normale.
3. Aree sotto la densità normale standard.
4. Aree sotto qualsiasi densità normale.

Capitoli 11 e 12: tecnica di calcolo integrale
Capitolo 13: variabili aleatorie continue (continuazione)

1. Funzione di ripartizione di una *V.A.*

2. Funzione di densità di una *V.A.* continua.

3. Proprietà della funzione di ripartizione.

4. Valore atteso e varianza.

5. Densità di funzioni biiettive e derivabili di *V.A.* continue.

Capitolo 14: Teorema del Limite Centrale
Capitolo 15: Distribuzioni Multivariate Discrete

1. Variabili aleatorie multivariate discrete.

2. Funzioni di probabilità congiunte, marginali, condizionate, indipendenza stocastica.

3. Valore atteso e varianza condizionata.

4. Valore atteso di funzioni di variabili aleatorie multivariate, covarianza e correlazione.

$$Cov(X, Y) = E\left[(X - E(X))\,(Y - E(Y))\right]$$

5. Covarianze e indipendenza, varianze di somme di *V.A.*.

Seguono infine i Capitoli 16, 17 e 18: distribuzioni multivariate continue I e II, la distribuzione normale bivariata.

1.2 Fondamenti di probabilità

Nella prima parte del corso (Statistica I) sono stati trattati quegli elementi di statistica descrittiva che hanno in seguito permesso di affrontare i temi propri della teoria della probabilità quali ad esempio

- Lo spazio campionario
- La funzione di probabilità
- Le variabili aleatorie di cui avete visto
  - la funzione di ripartizione
  - la funzione di densità
  - alcuni momenti teorici (valore atteso, varianza)

In questa sezione desideriamo rivedere alcuni concetti e definizioni ritenuti fondamentali per la comprensione dei prossimi capitoli. Il primo tema che vogliamo affrontare riguarda il concetto di *esperimento aleatorio*[1].

1.2.1 Esperimento Aleatorio

Definizione 1. Un esperimento aleatorio consiste nell'osservazione di un processo o procedimento aleatorio i cui esiti sono definiti ma non prevedibili.

Definizione 2. L'insieme degli esiti di un esperimento aleatorio è chiamato *spazio campionario* ed è notato Ω.

Definizione 3. Un qualsiasi sottoinsieme E di Ω è chiamato evento.

Esempio 1. Lancio di un dado. $\Omega = \{1, 2, 3, 4, 5, 6\}$. $E = \{$osservo un numero pari$\} = \{2, 4, 6\}$.

Esempio 2. Lancio indipendente di due dadi. In questo caso avremo

$$\Omega = \{(1,1), (1,2), ..., (1,6), ..., (6,1), ..., (6,6)\}.$$

Un possibile evento: $E = \{$Osservo una loro somma superiore a $9\} = \{(5,5), (5,6), (6,5), (6,6)\}$.

Tipicamente siamo interessati a calcolare la probabilità di eventi tramite una legge o distribuzione o misura[2] di probabilità P. Ad esempio, quando lo spazio campionario è discreto e finito (cioè Ω contiene un numero finito di esiti) e gli esiti sono equiprobabili avete imparato a calcolare la probabilità di un evento E tramite la formula

$$P(E) = \frac{\#\text{ esiti in } E}{\#\text{ esiti in } \Omega} = \frac{\#\text{ esiti favorevoli}}{\#\text{ esiti totali}}. \tag{1.1}$$

Osservazioni:

[1] Rinominiamo il concetto di *esperimento statistico* in *esperimento aleatorio*. Come avremo modo di discutere più avanti, all'espressione "esperimento statistico" verrà assegnato un significato diverso.

[2] Utilizzeremo questi tre termini quali sinonimi.

1. La formula (1.1) consente di assegnare una probabilità ad uno qualsiasi dei $2^6 = 64$ diversi eventi[3] di Ω.

2. La formula (1.1) ha validità limitata: essa è applicabile solo al caso di uno spazio campionario finito i cui esiti sono equiprobabili.

3. Quando lo spazio campionario Ω ha un numero infinito di esiti le cose si complicano: in generale non è più possibile calcolare la probabilità di un qualsiasi sottoinsieme di Ω.

4. Quando gli esiti non sono equiprobabili la formula (1.1) è inutilizzabile. Ad esempio, nell'esperimento aleatorio definito dal lancio di un dado la probabilità dell'evento $E = \{\text{l'esito è un numero pari}\}$ non è calcolabile tramite la formula 1.1 quando il dado è truccato!

La definizione formale dell'insieme degli eventi $\mathcal{E}$ è la seguente:

Definizione 4. L'insieme degli eventi $\mathcal{E}$ è un insieme i cui elementi sono sottoinsiemi di Ω. Gli elementi $\mathcal{E}$ sono chiamati *eventi*. Devono valere le seguenti proprietà:

1. Lo spazio campionario è elemento di $\mathcal{E}$, cioè $\Omega \in \mathcal{E}$.

2. Dato un qualsiasi elemento E di $\mathcal{E}$ allora anche il suo complemento deve appartenere a $\mathcal{E}$.

3. Se $E_1, E_2, \dots$ è una successione di elementi qualsiasi di $\mathcal{E}$, allora anche la loro unione deve appartenere ad $\mathcal{E}$:

$$E_i \in \mathcal{E}\ \forall i \in \{1, 2, \dots\} \Rightarrow \bigcup_{i=1}^{\infty} E_i \in \mathcal{E}.$$

Esempio 3. Lancio un dado. Lo spazio campionario è $\Omega = \{1, 2, 3, 4, 5, 6\}$. I seguenti insiemi sono tutti esempi validi di insieme degli eventi

1. $\mathcal{E} = \{\Omega, \emptyset\}$.

2. $\mathcal{E} = \{\Omega, \emptyset, \{1\}, \{2, 3, 4, 5, 6\}\}$.

[3] Il numero di eventi diversi fra loro di uno spazio campionario finito Ω è dato dalla formula

$$2^{\#\text{ esiti in }\Omega}.$$

3. $\mathcal{E} = \{\Omega, \emptyset, \{1,3,5\}, \{2,4,6\}\}$.

4. $\mathcal{E} = \{\Omega, \emptyset, \{1,2,3\}, \{4,5\}, \{6\}, \{4,5,6\}, \{1,2,3,6\}, \{1,2,3,4,5\}\}$.

I seguenti insiemi non sono esempi validi di insiemi degli eventi (perché?):

1. $\mathcal{E} = \{\Omega, \emptyset, \{1\}, \{2,3\}, \{1,2,3\}, \{2,3,4,5,6\}\}$.

2. $\mathcal{E} = \{\Omega, \emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$.

3. $\mathcal{E} = \{\Omega, \emptyset, \{1\}, \{1,2\}, \{1,2,3\}, \{1,2,3,4\}, \{1,2,3,4,5\}\}$.

Lasciando i dettagli tecnici ai matematici indicheremo semplicemente con $\mathcal{E}$ l'insieme degli eventi e senza preoccuparci oltre misura della sua struttura daremo per scontato che $\mathcal{E}$ soddisfa le condizioni enunciate nella Definizione 4. Se tale condizioni sono verificate diremo che l'insieme degli eventi $\mathcal{E}$ è misurabile, nel senso che è possibile assegnare una probabilità a ciascun evento in maniera "consistente". La proprietà importante da ricordare è che l'insieme degli eventi $\mathcal{E}$ contiene dei *sottoinsiemi* di Ω: ogni elemento E di $\mathcal{E}$ è un sottoinsieme di Ω, vedi Figura 1.1.

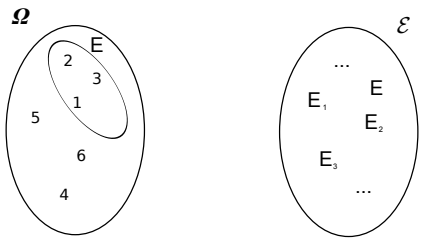

Figura 1.1: Lancio di un dado. Spazio campionario e insieme degli eventi.

Il terzo ed ultimo oggetto fondamentale che vogliamo discutere è la legge di probabilità P per la quale utilizziamo la seguente definizione.

Definizione 5. La legge di probabilità P è una funzione il cui dominio è $\mathcal{E}$, l'insieme degli eventi, con le seguenti proprietà:

1. $P(E) \geq 0$ per ogni evento $E \in \mathcal{E}$.

2. $P(\Omega) = 1$.

3. Se $E_1, E_2, \ldots$ è una successione di eventi di $\mathcal{E}$ fra loro incompatibili, cioè $E_i \cap E_j = \varnothing$ per $i \neq j$, allora

$$P\left(E_1 \cup E_2 \cup \ldots\right) = \sum_{i=1}^{\infty} P\left(E_i\right).$$

Della Definizione 5 è importante notare una cosa: gli argomenti da inserire in P non sono esiti (elementi di Ω) ma eventi (elementi di $\mathcal{E}$)!

Per concludere: un esperimento aleatorio è caratterizzato da tre oggetti: lo spazio campionario Ω, l'insieme degli eventi $\mathcal{E}$ e la legge di probabilità P. La tripla $(\Omega, \mathcal{E}, P)$ è chiamata *spazio di probabilità*.

Esempio 4. Lancio di un dado non truccato. Abbiamo visto in precedenza che

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{E}$ è l'insieme che contiene i $2^6 = 64$ possibili sottoinsiemi (eventi) di Ω. Il loro elenco è lasciato come esercizio.
- P è la misura o legge di probabilità definita su $\mathcal{E}$. Per qualsiasi elemento (evento) E di $\mathcal{E}$ avremo

$$P(E) = \frac{\#\text{ esiti in } E}{6}.$$

La tripla $(\Omega, \mathcal{E}, P)$ è lo spazio di probabilità definito dall'esperimento aleatorio "lancio di un dato non truccato".

Esempio 5. Lancio di un dado truccato. Gli esiti dell'esperimento non cambiano rispetto al lancio di un dado non truccato. Avremo quindi

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Anche gli eventi rimangono immutati. Come in precedenza varrà che

- $\mathcal{E}$ è l'insieme che contiene i $2^6 = 64$ possibili sottoinsiemi (eventi) di Ω.

La legge di probabilità di questo esperimento aleatorio, notata $\tilde{P}$, è diversa dalla precedente. Infatti, mentre in precedenza un qualsiasi evento elementare quale ad esempio $E = \{1\}$ aveva probabilità uguale a $\frac{1}{6}$, ora a causa della "manomissione" del dado i sei eventi elementari $E_i = \{i\}$, $i = 1, \ldots, 6$ avranno probabilità $\tilde{p}_i$ delle quali almeno due saranno diverse da $\frac{1}{6}$. Dato un qualsiasi evento E di $\mathcal{E}$, la sua probabilità sarà calcolata come la

somma delle probabilità dei singoli esiti in esso contenuti. Per capire ciò, prendiamo l'evento $E = \{\text{osservo un numero pari}\} = \{2, 4, 6\}$. Notiamo che $E = E_2 \cup E_4 \cup E_6$. Ma i tre eventi elementari sono disgiunti e quindi per la proprietà 3 di una legge di probabilità avremo che

$$\tilde{P}(E) = \tilde{P}(E_2) + \tilde{P}(E_4) + \tilde{P}(E_6) = \tilde{p}_2 + \tilde{p}_4 + \tilde{p}_6.$$

Per un qualsiasi evento E avremo quindi che

$$\tilde{P}(E) = \sum_{i \in E} \tilde{P}(\{i\}) = \sum_{i \in E} \tilde{p}_i.$$

La tripla $(\Omega, \mathcal{E}, \tilde{P})$ è lo spazio di probabilità che corrisponde al "lancio di un dato truccato". Rispetto all'esempio precedente i primi due termini della tripla sono identici, è cambiata unicamente la legge di probabilità.

1.2.2 Funzione di ripartizione e di densità

Abbiamo visto che per ottenere un esperimento aleatorio è necessario definire lo spazio campionario Ω nonché l'insieme degli eventi $\mathcal{E}$ sul quale sarà definta la legge di probabilità P. Consideriamo ora un esperimento aleatorio $(\Omega, \mathcal{E}, P)$ in cui lo spazio campionario Ω è sottoinsieme di $\mathbb{R}$. Questo significa che gli esiti dell'esperimento saranno dei numeri[4]. In tal caso è possibile definire la funzione di ripartizione della legge di probabilità P.

Definizione 6. (Funzione di ripartizione di un esperimento aleatorio). La funzione di ripartizione $F : \mathbb{R} \to [0, 1]$ della legge di probabilità P è definita come

$$F(x) := P((-\infty, x]) \text{ con } x \in \mathbb{R}. \tag{1.2}$$

La funzione di ripartizione $F(x)$ è dunque la probabilità dell'evento $E = (-\infty, x]$ di osservare un esito inferiore o al massimo uguale al valore x.

Osservazione 1. Facciamo notare che P ed F sono funzioni diverse. F è una funzione definita su $\mathbb{R}$ mentre P è una funzione definita su $\mathcal{E}$, l'insieme degli eventi. È possibile dimostrare come P ed F siano "le due facce della stessa moneta" nel senso che da P segue univocamente F e viceversa.

Esempio 6. Distribuzione uniforme discreta con $\Omega = \{0.2, 0.4, 0.6, 0.8, 1\}$.

[4]Non tutti gli esperimenti aleatori soddisfano questa condizione. Pensate all'esperimento aleatorio di osservare il colore della prima autovettura che circola su una determinata tratta stradale a partire da un certo orario. L'esito non è numerico ma è un colore. Per contro, l'esempio del lancio di un dado soddisfa questa condizione in quanto lo spazio campionario $\Omega = \{1, 2, 3, 4, 5, 6\} \subset \mathbb{R}$.

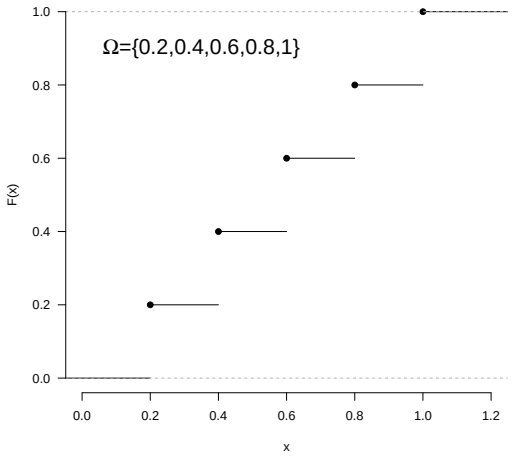

Figura 1.2: Distribuzione uniforme discreta

Esempio 7. Distribuzione uniforme continua $U[0,1]$.

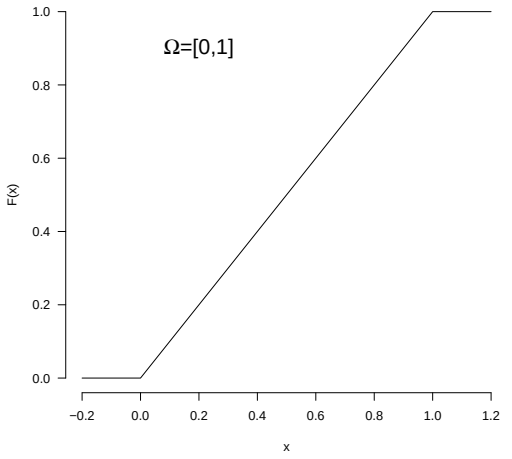

Figura 1.3: Distribuzione uniforme continua

Definizione 7. Diciamo che F è assolutamente continua se esiste una funzione reale f non negativa tale per cui

$$F(x) = \int_{-\infty}^{x} f(u)du \text{ per ogni } x \in \mathbb{R}.$$

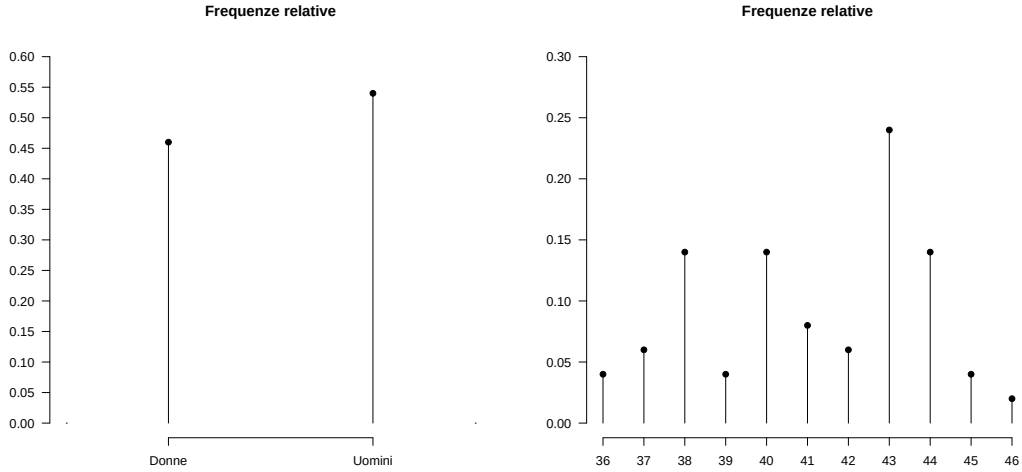

Figura 1.4: Frequenze relative popolazione studenti Statistica II - 2009

La funzione f è chiamata funzione di densità di F. Vale inoltre la relazione

$$F'(x) = f(x).$$

1.3 Statistica descrittiva e teoria della probabilità

Nel corso di Statistica I avete studiato che il punto di partenza di qualunque analisi statistica è la definizione di una popolazione obiettivo (o popolazione di riferimento) nonché della caratteristica della popolazione a cui si è interessati. A titolo di esempio prendiamo quale popolazione di riferimento l'insieme di studenti che lo scorso anno ha seguito il corso di Statistica II. Quali caratteristiche sotto esame scegliamo il sesso e il numero di scarpe di ciascun studente. La Figura 1.4 riporta il grafico delle frequenze relative per le due caratteristiche.
Notiamo che la seconda caratteristica è numerica. In questo semplice esempio non abbiamo definito delle classi di valore: tutti i numeri di scarpe osservati figurano nell'istogramma. Tuttavia se avessimo studiato un'altra caratteristica quale ad esempio il peso o l'altezza ecco che molto probabilmente avremmo costruito delle classi in cui inserire ciascun individuo della popolazione. La costruzione di classi può però costituire un problema. Innanzi tutto è necessario trovare una regola che a partire dalle singole osservazioni indichi quante e quali classi costruire. Secondariamente il passaggio dall'insieme della popo-

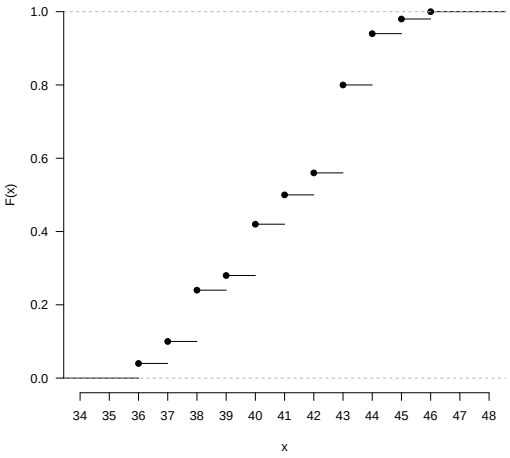

Figura 1.5: Funzione di ripartizione del numero di scarpe della popolazione studenti Statistica II - 2009

lazione al suo istogramma di frequenze relative o assolute genera una perdita di informazione. Ad esempio, anche se sapessimo che 23 individui hanno un peso compreso tra 70 e 75 chilogrammi non saremmo in grado di stabilire quanti di essi hanno un peso superiore a 72 chilogrammi. Per tale motivo quando abbiamo a che fare con caratteristiche numeriche preferiamo studiare la funzione di ripartizione della popolazione obiettivo.

Definizione 8. (Funzione di ripartizione della popolazione obiettivo). La funzione di ripartizione, notata F, di una caratteristica x della popolazione obiettivo è una funzione di variabile reale definita su tutto $\mathbb{R}$ e tale che per ogni $x \in \mathbb{R}$

$$F(x) = \frac{\#\text{ unità con caratteristica } \leq x}{\#\text{ unità della popolazione}}. \tag{1.3}$$

L'interpretazione è semplice ed è molto simile all'interpretazione della funzione di ripartizione di un esperimento aleatorio. Infatti, tornando al nostro esempio iniziale della popolazione di studenti che hanno seguito il corso di Statistica II, $F(38) = 0.24$ significa che il 24% della popolazione possiede un numero di scarpa inferiore o uguale a 38. Come per un esperimento aleatorio, la funzione di ripartizione racchiude tutta l'informazione disponibile sulla popolazione (confronta la Definizione (1.2)). Nel caso della funzione di ripartizione di una popolazione obiettivo non si parla di probabilità ma di frequenze relative in quanto non c'è nulla di aleatorio nella popolazione. La popolazione esiste e l'attributo è osservabile. F descrive come l'attributo

in esame è distribuito all'interno della popolazione. La conoscenza di F è equivalente all'osservazione dell'attributo su tutta la popolazione obiettivo.

Lo scopo di molti studi empirici in economia (micro- e macroeconomia, marketing ma anche in altre scienze sociali e non) è quello di determinare la funzione di ripartizione di un determinato attributo o caratteristica di una popolazione obiettivo[5] ed in seguito utilizzare questa informazione per giungere a delle conclusioni di carattere generale. Per tale motivo uno dei temi dei prossimi capitoli sarà proprio quello di studiare come ottenere un'approssimazione della distribuzione dell'intera popolazione partendo da un sottoinsieme di osservazioni della stessa (campione).

Le definizioni (1.2) e (1.3) di funzione di ripartizione caratterizzano rispettivamente l'aspetto probabilistico di un esperimento aleatorio e la struttura della popolazione obiettivo. Le due definizioni, anche se concettualmente molto simili, non devono essere confuse.

1.4 Famiglie parametriche di distribuzioni

Nel corso di Statistica I sono state presentate alcune particolari distribuzioni di probabilità con le relative funzioni di ripartizione, di densità o di probabilità. La distribuzione Normale (caso continuo) e la distribuzione di Poisson (caso discreto) sono due esempi a voi noti. Per quanto riguarda la distribuzione Normale, essa è caratterizzata da due *parametri* che come già sapete corrispondono al valore atteso ed alla varianza della distribuzione.

Osservazione 2. Attenzione a non generalizzare questa caratteristica della distribuzione Normale. Durante questo corso incontreremo diverse nuove distribuzioni di probabilità. Come nel caso della distribuzione Normale o di Poisson la loro "forma" dipenderà da uno o più parametri la cui interpretazione varierà da distribuzione a distribuzione.

La Figura 1.6 esplicita quanto affermato mostrando la funzione di densità della distribuzione Normale per due diversi valori di (μ, σ^2). Nella Figura 1.6 la formula della funzione di densità è invariata. Cambiano i valori dei due parametri μ e σ^2. Per ogni possibile valore dei parametri si ottiene una diversa distribuzione della famiglia Normale. Si è soliti indicare col termine famiglia parametrica Normale l'insieme di tutte le distribuzioni che si possono ottenere facendo variare i due parametri μ e σ^2. In generale si parlerà di

[5] Parleremo semplicemente di *distribuzione della popolazione* quando è chiaro quali siano l'attributo e la popolazione obiettivo in esame.

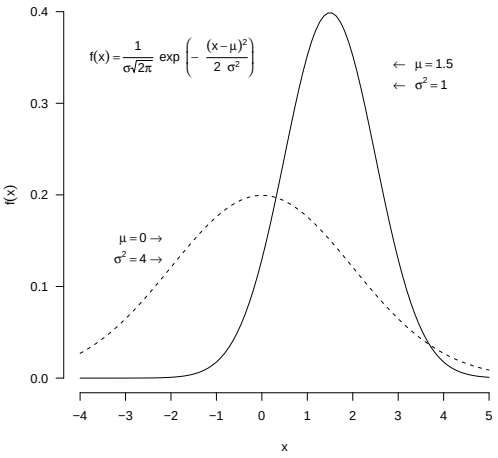

Figura 1.6: Due funzioni di densità della famiglia parametrica Normale

famiglia parametrica di distribuzioni per indicare un preciso insieme di distribuzioni la cui funzione di probabilità, funzione di ripartizione o funzione di densità è identica a meno di un numero finito di parametri. I parametri assumono valori in $\mathbb{R}$ o in sottoinsiemi di $\mathbb{R}$. Nel caso della famiglia parametrica Normale $\mu \in \mathbb{R}$ mentre $\sigma^2 > 0$. L'insieme dei valori che i parametri di una famiglia parametrica di distribuzioni possono assumere è chiamato spazio parametrico ed è indicato con il simbolo Θ. Per la famiglia parametrica Normale $\Theta = \mathbb{R} \times \mathbb{R}^+$. Riportiamo alcuni esempi di famiglie parametriche di distribuzioni discrete e continue. Nella loro definizione appare la *funzione indicatrice* $I_A(x)$, dove A rappresenta un sottoinsieme di $\mathbb{R}$.

Ricordiamo che per qualsiasi sottoinsieme A di $\mathbb{R}$ la funzione $I_A : \mathbb{R} \to \{0, 1\}$ è definita nel seguente modo

$$I_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{altrimenti} \end{cases} .$$

Prendendo quale esempio la distribuzione *discreta* uniforme, l'insieme A corrisponde all'insieme dei numeri interi compresi da 1 a n, ovvero $\{1, ..., n\}$. Se poniamo $n = 5$ avremo che $I_{\{1,...,5\}}(20) = 0$, $I_{\{1,...,5\}}(2.6) = 0$, $I_{\{1,...,5\}}(4) = 1$.

	Uniforme	**Bernoulli**
$f(x) =$	$\frac{1}{n} I_{\{1,\dots,n\}}(x)$	$p^x(1-p)^{1-x} I_{\{0,1\}}(x)$
$F(x) =$	$\min(\frac{\lfloor x \rfloor}{n}, 1) I_{[1,\infty)}(x)$	$(1-p) I_{[0,\infty)}(x) + p I_{[1,\infty)}(x)$
Spazio parametrico Θ	$n = 1, 2, \dots$	$0 \le p \le 1$
Valore atteso	$\frac{n+1}{2}$	p
Varianza	$\frac{n^2-1}{12}$	$p(1-p)$

	Binomiale	**Binomiale negativa**
$f(x) =$	$\binom{n}{x} p^x (1-p)^{n-x} I_{\{0,\dots,n\}}(x)$	$\binom{r+x-1}{x} p^r (1-p)^x I_{\{0,1,\dots\}}(x)$
$F(x) =$	$\sum_{i=0}^{\lfloor x \rfloor} f(i)$	$\sum_{i=0}^{\lfloor x \rfloor} f(i)$
Spazio parametrico Θ	$0 \le p \le 1 \ ; \ n = 1, 2, \dots$	$0 < p \le 1 \ ; \ r = 1, 2, \dots$
Valore atteso	np	$\frac{r(1-p)}{p}$
Varianza	$np(1-p)$	$\frac{r(1-p)}{p^2}$

	Geometrica	**Ipergeometrica**
$f(x) =$	$p(1-p)^x I_{\{0,\dots,n\}}(x)$	$\frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} I_{\{0,1,\dots,n\}}(x)$
$F(x) =$	$\left(1 - (1-p)^{\lfloor x+1 \rfloor}\right) I_{[0,\infty)}(x)$	$\sum_{i=0}^{\lfloor x \rfloor} f(i)$
Spazio parametrico Θ	$0 < p \le 1$	$M = 1, 2, \dots \ ; \ K = 0, \dots, M \ ; \ n = 1, 2, \dots$
Valore atteso	$\frac{1-p}{p}$	$n\frac{K}{M}$
Varianza	$\frac{1-p}{p^2}$	$n\frac{K}{M}\frac{M-K}{M}\frac{M-n}{M-1}$

	Uniforme	**Normale**
$f(x) =$	$\frac{1}{b-a} I_{[a,b]}(x)$	$\frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2)$
$F(x) =$	$\min(\frac{x-a}{b-a}, 1) I_{[a,\infty)}(x)$	$\int_{-\infty}^{x} f(u)du$
Spazio parametrico Θ	$-\infty < a < b < \infty$	$-\infty < \mu < \infty \ ; \ \sigma > 0$
Valore atteso	$(a+b)/2$	μ
Varianza	$(b-a)^2/12$	σ^2

	Logistica	**Pareto**
$f(x) =$		$f(x) = \frac{\theta k^\theta}{x^{\theta+1}} I_{(k,\infty)}(x)$
$F(x) =$	$[1 + \exp(-(x-\alpha)/\beta]^{-1}$	
Spazio parametrico Θ	$\beta > 0 \ ; \ -\infty < \alpha < \infty$	$k > 0 \ ; \ \theta > 0$
Valore atteso	$\alpha + \beta\gamma$ con $\gamma \approx 0.577216$	$\frac{\theta k}{\theta - 1}$ quando $\theta > 1$
Varianza	$\frac{\pi^2 \beta^2}{6}$	$\frac{\theta k^2}{(\theta-1)^2(\theta-2)}$ quando $\theta > 2$

Le famiglie appena presentate sono solo alcune delle numerose famiglie parametriche di distribuzioni. Come potete verificare voi stessi il numero di parametri e la loro interpretazione varia da famiglia a famiglia.

1.5 Contenuto del corso

Nella seconda parte del corso (Statistica II) verranno affrontati i temi propri dell'induzione statistica, ovvero

- La teoria del campionamento

  La teoria del campionamento trae la sua ragion d'essere dalla necessità di raccogliere in maniera appropriata le informazioni (dati) necessari allo studio del fenomeno a cui siamo interessati.

- La teoria della stima

  La teoria della stima utilizza le informazioni così raccolte (più eventuali altre informazioni) per trarre delle conclusioni o dare delle risposte alle domande relative al fonomeno sotto esame.

- La verifica d'ipotesi

  A causa dell'incompleta informazione e alla natura aleatoria di molti fenomeni studiati in economia e nelle scienze sociali tali conclusioni non saranno vere in assoluto. Tuttavia, se lo studio è stato condotto seguendo certi principi, il grado di incertezza potrà essere quantificato. La verifica d'ipotesi tramite dei test statistici e la construzione di intervalli di confidenza relativi ai parametri stimati permettono di misurare l'attendibilità dei risultati.

Esempio 8. Pino, il responsabile del settore marketing di una grossa azienda, desidera conoscere il grado di visibilità sul mercato di un proprio prodotto. Il motivo di questo interesse consiste nel fatto che si vuole capire se lo scarso successo commerciale sia dovuto alla scarsa qualità del prodotto o al fatto

che il prodotto non sia sufficientemente conosciuto. La dirigenza dell'azienda decide di quantificare il grado di visibilità tramite la percentuale p di consumatori che conoscono (ma non necessariamente acquistano o hanno acquistato in passato) il prodotto in questione. Essa valuta il livello attuale di p secondo la seguente tabella

$p < 30\%$	$30\% \leq p < 60\%$	$60\% \leq p$
insufficiente	discreto	buono

La quantità in esame è dunque p. Poiché è praticamente impossibile (e troppo oneroso) intervistare tutta la popolazione dei consumatori sarà necessario selezionare ed intervistare un sottoinsieme di tale popolazione che chiameremo campione.

- Come si dovrà costruire il campione da intervistare?
- Sulla base di quali criteri dovranno essere selezionate le persone (o unità) da intervistare: l'età, il reddito, la provenienza geografica, il sesso?
- Quante persone dovranno essere intervistate?

La teoria del campionamento si occupa di trovare una risposta a queste domande.

Una volta effettuato il campionamento si potrà stimare

$$0 \leq p \leq 1.$$

La teoria della stima consente di effettuare due tipi di stime: una stima puntuale di p ed una stima per intervallo. La stima puntuale di p, notata semplicemente $\widehat{p}$, fornisce quella che noi riteniamo essere la migliore approssimazione o alternativa al valore sconosciuto p. Poiché il calcolo è effettuato sulla base di un campione di consumatori e non sull'intera popolazione (l'informazione è incompleta), il valore di $\widehat{p}$ sarà molto probabilmente diverso da p. Si è dunque confrontati con un errore di stima.

La stima per intervallo affronta il problema di stima in maniera diversa. Essa infatti fornisce un intervallo di valori nel quale noi confidiamo che p sia incluso. Tale intervallo è chiamato *intervallo di confidenza*. La forza o intensità con cui noi crediamo nella nostra affermazione è chiamata *livello di confidenza*. Il livello di confidenza corrisponde alla probabilità ex-ante che

l'intervallo costruito sulla base delle osservazioni disponibili contenga p. Un esempio di una realizzazione di intervallo di confidenza potrebbe essere:

$$[0.2, 0.4].$$

La teoria della stima si occupa dunque di come calcolare $\widehat{p}$ o come costruire l'intervallo di confidenza utilizzando le informazioni disponibili (dati).

Esempio 9. Continuiamo l'Esempio 8. Supponiamo che, nel caso in cui p fosse insufficiente, Pino abbia intenzione di condurre una vasta campagna pubblicitaria e che il valore stimato di p sia $\widehat{p} = 29\%$. A questo punto dobbiamo chiederci se la campagna pubblicitaria (ed i relativi costi) sia veramente necessaria. Infatti, potrebbe essere che il vero valore di p sia superiore al 30% ma per semplice "sfortuna" nella scelta del campione la nostra stima risulti inferiore a questa soglia. Il grado di visibilità è realmente insufficiente? Ci troviamo qui confrontati col terzo tema dell'inferenza statistica: la verifica d'ipotesi. Valori di $\widehat{p}$ inferiori al 30% sono evidenza a favore dell'ipotesi di un p insufficiente e contro l'ipotesi di un p discreto o buono. Tenendo conto dell'aleatorietà nella stima di p e del relativo errore, quanto bassa deve essere la stima $\widehat{p}$ per decidere di effettuare la campagna pubblicitaria e quindi rifiutare l'ipotesi che p sia discreto o buono? Per dare una risposta a questa domanda si costruirà un test statistico. Sarà innanzi tutto necessario una formalizzazione matematica della nostra ipotesi. Per Pino, che non ha seguito nessun corso di statistica ed è uomo d'azione, $\widehat{p} = 29\%$ è un valore più che sufficiente per eseguire la campagna pubblicitaria e spendere un sacco di soldi.

Sappiamo che tanto più la campagna pubblicitaria sarà efficace, tanto più aumenterà il grado di visibilità. Il manager dell'azienda non è soddisfatto del modo come Pino ha condotto la campagna pubblicitaria e desidera verificarne immediatamente l'efficacia. Ordina quindi ad una società indipendente e specializzata in sondaggi di intervistare un nuovo campione da cui stimare il nuovo grado di visibilità del prodotto che indicheremo ora con

$$0 \leq p_{new} \leq 1.$$

p_{new} è il grado di visibilità *dopo* la campagna pubblicitaria. Il manager utilizzerà la seguente regola per decidere la sorte di Pino:

$p_{new} = p$	$p_{new} \geq p$
licenzio Pino	non licenzio pino

La stima puntuale di p_{new}, notata $\widehat{p}_{new}$ risulta essere del 34%, un valore di poco superiore al precedente 29% calcolato prima della campagna pubblicitaria. Pino è felice. Il suo capo gli fa però notare che l'esiguo aumento del 5%

rispetto alla precedente stima potrebbe essere semplicemente dovuto al caso. Pino vi chiede allora di quantificare sulla base dei dati a disposizione il livello di evidenza per cui l'ipotesi $p = p_{new}$ (licenzio Pino) possa essere rigettata. È compito della verifica d'ipotesi fornire una risposta a questa domanda.

Esempio 10. Un produttore di mele deve fissare anticipatamente la quantità di mele da consegnare al suo acquirente. Sappiamo che la quantità q prodotta dipende tra le altre cose da fattori climatici, difficilmente prevedibili ed assai variabili. Da ricerche effettuate si sa che la produzione di mele q è distribuita secondo la distribuzione Normale[6] che sapete essere univocamente identificata dai parametri μ e σ^2. Purtroppo i due parametri μ e σ^2 in questo caso sono sconosciuti e dovranno essere stimati sulla base dei valori della produzione osservati negli ultimi 20 anni. È questo un problema legato alla teoria della stima. Infine, potremmo chiederci se l'assunzione della distribuzione normale si addice alla variabile aleatoria q o se tale assunzione è fondamentalmente sbagliata. In questo frangente si tratta quindi di verificare l'ipotesi concernente la distribuzione di una variabile aleatoria per la quale si osservano un certo numero di realizzazioni.

1.6 Domande di fine capitolo

Domanda 1. Cosa rappresenta il simbolo Ω? Esplicitate il suo contenuto facendo un nuovo esempio non visto in classe.

Domanda 2. Che relazione sussiste tra Ω, $\mathcal{E}$ e E?

Domanda 3. Quali sono i tre elementi fondamentali che costituiscono un esperimento aleatorio?

Domanda 4. Costruite un esperimento aleatorio per cui $\Omega \nsubseteq \mathbb{R}$ (Ω *non* è l'insieme dei numeri reali o un suo qualsiasi sottoinsieme).

Domanda 5. Cos'è la funzione di ripartizione? Dato un qualsiasi esperimento aleatorio $(\Omega, \mathcal{E}, P)$ è sempre possibile definire la funzione di ripartizione della legge di probabilità P? Se non fosse sempre possibile costruite un semplice esempio di esperimento aleatorio per cui non è possibile definire la funzione di ripartizione.

Domanda 6. Che differenza c'è tra P ed F?

Domanda 7. Che differenza c'è tra la funzione di probabilità e la funzione di densità?

[6] Per la precisione la distribuzione Normale assegna probabilità positive all'evento di produrre una quantità negativa di mele. Tuttavia tale probabilità è trascurabile per dati valori di sigma.

Domanda 8. Cos'è una famiglia parametrica di distribuzioni? Cosa rappresenta Θ?

Domanda 9. Per due famiglie parametriche di distribuzioni discrete e due famiglie parametriche di distribuzioni continue a scelta indicate

1. il numero di parametri
2. la funzione di probabilità o di densità o di ripartizione
3. lo spazio parametrico

che le contraddistingue. Scegliete in seguito dei valori appropriati per i parametri ed eseguite il grafico della funzione di probabilità o di densità o di ripartizione a dipendenza dalla scelta effettuata.

Domanda 10. Ripetete lo stesso esercizio della domanda precedente con altre due famiglie parametriche di distribuzioni (una discreta e l'altra continua) che non siano già elencate in questo capitolo.