

Lezioni dell'insegnamento

CALCOLO DELLE PROBABILITA'
E STATISTICA

a.a. 2020/21

Corso di Laurea in Informatica
Università degli Studi di Udine

Docente: Luigi Pace, DIES

Premessa

Questa dispensa si rivolge agli studenti del secondo anno del corso di laurea in Informatica dell'Università di Udine. Contiene le lezioni dell'insegnamento Calcolo delle Probabilità e Statistica, 6 CFU, settore scientifico-disciplinare SECS-S/01. È frutto di un'esperienza più che ventennale di didattica di probabilità e statistica in corsi di laurea in Informatica e Scienze dell'Informazione.

L'obiettivo dell'insegnamento è trasmettere i concetti fondamentali della modellazione matematica dell'incertezza tramite modelli probabilistici e statistici. I modelli probabilistici descrivono la produzione di dati potenzialmente osservabili. I modelli statistici permettono, partendo da dati effettivamente osservati, di ottenere una ricostruzione plausibile di un modello probabilistico adatto a descrivere il fenomeno osservato.

Da un modello probabilistico supposto noto si deducono probabilità di eventi, che permettono di prevedere alcuni aspetti degli ulteriori dati che saranno osservati continuando la sperimentazione. In pratica, però, il modello probabilistico ha certe caratteristiche che sono ignote, descritte tipicamente da parametri, su cui i dati osservati sono chiamati a fornire informazione. Questo aspetto induttivo, che generalizza ciò che è stato osservato su un numero limitato di casi, dal campione alla popolazione come si dice, e permette la previsione di ulteriori istanze del fenomeno, è di pertinenza della Statistica ed è detto inferenza statistica.

Il contenuto dell'insegnamento si può suddividere in sei parti. Ciascuna corrisponde, grosso modo, a un CFU.

La prima parte, Lezioni 1–5, introduce gli assiomi del Calcolo delle Probabilità, i primi teoremi, le nozioni di eventi indipendenti e di probabilità condizionale, le probabilità ipergeometriche e binomiali, il teorema di Bayes.

La seconda parte, Lezioni 6–10, tratta le variabili casuali e le loro leggi di probabilità. Si introducono le variabili casuali univariate, bivariate e multivariate (queste ultime con componenti indipendenti), sottolineando la distinzione fra leggi discrete e leggi continue, fra componenti indipendenti e componenti dipendenti (nel caso bivariato), e il ruolo della funzione di ripartizione nel caratterizzare le leggi di probabilità univariate.

La terza parte, Lezioni 11–13, illustra in dettaglio alcune leggi univariate notevoli, discrete (leggi di Poisson e geometriche) come pure continue (esponenziali, di Weibull, gamma). Si conclude la parte con alcune idee sul reperimento

della legge di probabilità di funzioni di variabili casuali (variabili casuali trasformate), con particolare attenzione alle leggi del massimo e del minimo di variabili casuali multivariate con componenti indipendenti.

La quarta parte, Lezioni 14–16, tratta gli indici di posizione e di variabilità dando grande risalto alla funzione generatrice dei momenti. Si tratta di uno strumento che può non solo agevolare il calcolo di valori attesi e varianze, ma soprattutto che può caratterizzare la legge di probabilità, e risulta assai utile per reperire la legge della somma di variabili casuali indipendenti.

La quinta parte, Lezioni 17–20, introduce le leggi normali e, sotto campionamento casuale semplice da legge normale, studia le distribuzioni campionarie delle principali statistiche, media campionaria e varianza campionaria corretta. Si motiva, anche più in generale, la selezione delle statistiche informative sui parametri del modello. Si studia inoltre il comportamento sotto campionamento casuale semplice della media campionaria quando il campionamento non è da popolazione normale, introducendo la legge dei grandi numeri e il teorema centrale del limite. Questi teoremi mostrano che, quando la numerosità di campionamento è sufficientemente grande, la media campionaria varia, al variare dei campioni di quella numerosità, secondo una legge non molto diversa da quella che vale quando il campionamento è da una popolazione normale.

La sesta e ultima parte, Lezioni 21–23, offre un'introduzione all'inferenza statistica frequentista, con la classica suddivisione fra problemi di stima puntuale, problemi di costruzione di intervalli di confidenza e problemi di costruzione di test statistici (sia nella versione decisionale di accettazione-rifiuto sia nella versione non decisionale: P -value).

Ogni lezione è corredata da numerosi esercizi di natura sia teorica sia applicativa. Per assimilare i contenuti del volume bisogna mettersi alla prova con una adeguata parte di essi. A conclusione di parecchie Lezioni è proposta una guida ad esercitazioni con il computer utilizzando il software statistico R (R Core Team, 2019). Lo scopo è ovviare alla mancanza di una parte laboratoriale, proponendo agli studenti più interessati un minilaboratorio in autoapprendimento.

Segnalazioni di errori, imprecisioni, punti oscuri saranno accolte con gratitudine.

Il manoscritto è stato composto con L^AT_EX.

Indice

1	Esperimenti casuali, spazi campionari, eventi	1
1.1	Incertezza e variabilità	1
1.2	Esperimenti casuali	2
1.3	Spazi campionari	3
1.4	Eventi	4
1.4.1	Eventi notevoli	5
1.4.2	Eventi: esempi	5
1.4.3	Eventi: realizzati e non realizzati	5
1.5	Esercizi	6
1.6	Appendice: introduzione all'ambiente R	8
2	Probabilità classica, frequentista, assiomatica	11
2.1	Premessa	11
2.2	Operazioni su eventi	11
2.3	Eventi incompatibili	12
2.4	La classe degli eventi e le probabilità	13
2.5	Probabilità classica	13
2.6	Probabilità frequentista	14
2.7	Probabilità assiomatica	15
2.7.1	La classe degli eventi come σ -algebra	15
2.7.2	Spazi probabilizzabili	16
2.7.3	Misure di probabilità	16
2.7.4	Spazi probabilizzati	17
2.7.5	Le variabili casuali (v.c.) e le σ -algebre di Borel	17
2.8	Esercizi	17
2.9	Appendice: la probabilità soggettiva	19
2.10	Appendice: i vettori in R	20
3	Teoremi elementari e calcolo combinatorio	23
3.1	I primi cinque teoremi elementari	23
3.2	Partizioni dello spazio campionario: due teoremi	24
3.3	La probabilità classica e il calcolo combinatorio	25
3.3.1	Principi del calcolo combinatorio	26
3.3.2	Terminologia del calcolo combinatorio	27

3.3.3	Esempi	28
3.4	Le probabilità ipergeometriche	29
3.5	Esercizi	30
3.6	Appendice: tabulare probabilità con R	32
4	Condizionamento, indipendenza, leggi binomiali	33
4.1	Coefficienti binomiali e teorema del binomio: richiami	33
4.2	Le probabilità condizionali	35
4.2.1	Definizione	35
4.2.2	Formula della probabilità composta	37
4.3	Formula della probabilità totale	37
4.4	Eventi indipendenti	38
4.5	Le probabilità binomiali	39
4.6	Esercizi	42
4.7	Appendice: probabilità binomiali e ipergeometriche con R	43
5	Il teorema di Bayes	45
5.1	Un problema introduttivo	45
5.2	Il teorema: enunciato, dimostrazione, commenti	46
5.3	Esempi di applicazione	47
5.4	Esercizi	49
5.5	Appendice: il teorema di Bayes con R	51
6	Le v.c. e la loro legge di probabilità	53
6.1	Premessa	53
6.2	Visione euristica	53
6.3	Le v.c. come spazi probabilizzati	54
6.4	Le v.c. come applicazioni e la loro legge di probabilità	54
6.5	Classi di equivalenza di v.c.: l'identica distribuzione	55
6.6	Prime leggi notevoli di tipo discreto	56
6.7	Le leggi discrete in generale	57
6.8	Rappresentazioni in tabella di leggi discrete	57
6.9	Le leggi uniformi discrete	58
6.10	Attesa e varianza di v.c. univariate discrete	58
6.11	Esercizi	60
6.12	Appendice: R e le v.c. discrete con supporto finito	61
7	Le v.c. bivariate con legge discreta	63
7.1	Definizione	63
7.2	Le leggi marginali	64
7.3	Le leggi condizionali	65
7.4	La specificazione gerarchica	66
7.5	V.c. con legge discreta, componenti indipendenti	68
7.6	La covarianza	68
7.7	Esercizi risolti	69
7.8	Esercizi	72

7.9	Appendice: v.c. bivariate discrete con R	73
8	V.c. con legge continua	75
8.1	Leggi univariate di tipo continuo	75
8.2	Valore atteso e varianza	77
8.3	Due leggi univariate notevoli di tipo continuo	77
8.4	Il supporto di una v.c. univariata con legge continua	80
8.5	V.c. bivariate e multivariate con legge continua	80
8.6	V.c. multivariate con componenti indipendenti	83
8.7	Esercizi	84
8.8	Appendice: leggi uniformi continue ed esponenziali con R	85
9	La funzione di ripartizione	87
9.1	Definizione	87
9.2	Caso univariato: calcolo di probabilità di intervalli	88
9.3	Caso univariato: dalla $p_X(x)$ alla $F_X(x)$	89
9.4	Caso univariato discreto: dalla f.r. alla f.m.p.	89
9.5	Caso univariato continuo: dalla f.r. alla f.d.p.	90
9.6	Esercizi risolti	90
9.7	Esercizi	92
9.8	Appendice: alcune funzioni di ripartizione con R	94
10	F.r. univariate: proprietà caratterizzanti	95
10.1	Teorema sulle proprietà strutturali	95
10.2	Il teorema di caratterizzazione	96
10.3	Leggi univariate di tipo mistura	97
10.4	Esercizi	98
10.5	Appendice: simulare con R da una legge mistura	99
11	Leggi di Poisson e leggi geometriche	101
11.1	Il teorema di Poisson	101
11.2	Le leggi di Poisson	103
11.3	Le leggi geometriche	104
11.4	Assenza di memoria delle leggi geometriche	106
11.5	Esercizi	107
11.6	Appendice: v.c. con legge Poisson e geometrica con R	108
12	Tempi d'attesa nel tempo continuo	111
12.1	Assenza di memoria e leggi esponenziali	111
12.2	La funzione tasso di guasto	111
12.3	Le leggi di Weibull	113
12.4	Le leggi gamma	114
12.5	Esercizi	115
12.6	Appendice: v.c. con legge Weibull e gamma con R	116

13 Leggi di v.c. trasformate	117
13.1 Trasformazioni di v.c.	117
13.2 Il supporto di una v.c. trasformata	118
13.3 La f.m.p. di una v.c. trasformata	118
13.4 La f.r. di trasformate monotone	119
13.5 La f.d.p. di trasformate monotone	120
13.6 Le leggi del massimo e del minimo	122
13.7 Esercizi risolti	125
13.8 Esercizi	127
13.9 Appendice: v.c. trasformate con R	129
14 Indici di posizione e proprietà di $E(X)$	131
14.1 Indici di posizione di v.c. univariate	131
14.1.1 La moda	131
14.1.2 La mediana	132
14.1.3 Il quantile- p	133
14.2 Il valore atteso	134
14.3 Proprietà del valore atteso	134
14.3.1 Prime proprietà, utili per il calcolo	134
14.3.2 Ulteriori proprietà	135
14.4 Esercizi	138
14.5 Appendice: quantili con R	139
15 Indici di variabilità e proprietà di $\text{Var}(X)$	141
15.1 Indici di variabilità	141
15.2 Varianza e scarto quadratico medio	141
15.3 Altri indici di variabilità	142
15.4 Proprietà immediate della varianza	143
15.5 V.c. univariate standardizzate	144
15.6 Le disuguaglianze di Markov e di Čebyshev	144
15.7 Indici di posizione e variabilità per v.c. multivariate	146
15.8 La varianza di una combinazione lineare	147
15.9 Il coefficiente di correlazione lineare	148
15.10 Esercizi	149
15.11 Appendice: modelli di regressione lineare con R	150
16 La funzione generatrice dei momenti	153
16.1 Definizioni	153
16.2 Ottenimento dei momenti	154
16.3 Caratterizzazione della legge di probabilità	156
16.4 La f.g.m. di leggi notevoli	157
16.5 La legge della somma di v.c. indipendenti	160
16.6 Proprietà additive	161
16.6.1 Binomiali indipendenti con lo stesso p	161
16.6.2 Poisson indipendenti	161
16.6.3 Gamma indipendenti con lo stesso λ	162

16.7 Esercizi risolti	162
16.8 Esercizi	163
17 Le leggi normali	165
17.1 Genesi e definizione	165
17.2 Chiusura sotto trasformazioni affini	167
17.3 F.g.m. e momenti di una normale	167
17.4 Proprietà additiva delle normali indipendenti	169
17.5 Esercizi	169
18 Probabilità e quantili normali	171
18.1 La f.r. di una v.c. normale	171
18.2 Uso delle tavole della f.r. normale standard	172
18.2.1 Uso diretto: calcolo di probabilità normali	173
18.2.2 Uso inverso: calcolo di quantili normali	174
18.3 Esercizi svolti	175
18.4 Esercizi	176
18.5 Appendice: probabilità e quantili normali con R	177
19 Campioni e statistiche riassuntive	179
19.1 Leggi di campionamento casuale semplice	179
19.2 Statistiche riassuntive	181
19.3 Campionamento da normale con σ_0^2 noto	183
19.4 Le leggi chi-quadrato	184
19.5 La varianza campionaria da normale con μ noto	186
19.6 C.c.s. da normale con μ e σ^2 ignoti	187
19.7 Le leggi t di Student	189
19.8 Esercizi svolti	190
19.9 Esercizi	191
20 Media campionaria e popolazioni non normali	197
20.1 Attesa e varianza della media campionaria	197
20.2 Modi di convergenza di successioni di v.c.	198
20.3 Due semplici condizioni sufficienti	199
20.4 La legge dei grandi numeri	200
20.5 Il teorema centrale del limite	201
20.6 Probabilità normali per probabilità non normali	202
20.7 Esercizi	204
21 Inferenza statistica: stima puntuale	205
21.1 La variabilità campionaria	205
21.2 La distribuzione empirica	206
21.3 Stime e stimatori	207
21.4 Proprietà campionarie degli stimatori	209
21.5 Come reperire stimatori	210
21.6 Lo standard error di uno stimatore	212

21.7 Esercizi	213
22 Inferenza statistica: intervalli di confidenza	217
22.1 Dalla stima puntuale alla stima intervallare	217
22.2 Regioni di confidenza con livello di confidenza $1 - \alpha$	218
22.3 Normale con varianza nota: IC per μ	219
22.4 Campionamento da normale con valore atteso noto: IC per σ^2 . .	220
22.5 IC per σ^2	220
22.6 IC per μ se σ^2 è ignoto	221
22.7 Campionamento da legge con varianza finita	222
22.8 Esercizi	223
23 I test statistici	227
23.1 I dati e l'ipotesi	227
23.2 Ipotesi statistiche, nulla e alternativa	228
23.3 Test decisionali con livello di significatività α	229
23.4 Test non decisionali: il P -value	232
23.5 Interpretazione del P -value	234
23.6 Test con livello di significatività α dal P -value	234
23.7 Calcolo del P -value in modelli normali di c.c.s.	235
23.8 Esercizi	237

Lezione 1

Esperimenti casuali, spazi campionari, eventi

1.1 Incertezza e variabilità

Probabilità e Statistica hanno lo scopo di fornire, tramite modellazioni matematiche, valutazioni quantitative sull'incertezza e la variabilità di fenomeni aleatori, ossia dall'esito non certo. Le valutazioni ottenute condurranno a formulare predizioni ben calibrate sull'esito del fenomeno oggetto di studio. Una procedura di predizione è ben calibrata quando la frazione che corrisponde alla realtà dei fatti delle sue predizioni etichettate al livello del 95% è circa il 95% (e analogamente per gli altri livelli).

Incertezza e variabilità sono dunque alla base di Probabilità e Statistica. L'incertezza è anzitutto uno stato interno della mente di un soggetto particolare. (La mia incertezza non è la tua incertezza). La variabilità invece si constata da parte di tutti i soggetti quando si osservano particolari fenomeni aleatori, quelli che si possono ripresentare più e più volte, senza memoria del passato, sotto condizioni essenzialmente identiche. La variabilità porta con sé una forma stabile della valutazione dell'incertezza, condivisibile da soggetti singoli o da comunità che indagano su quel fenomeno aleatorio. Si possono così distinguere due tipi estremi di incertezze, le incertezze oggettivamente valutabili, o pubbliche, e le incertezze puramente soggettive, o private.

Sono oggettivamente valutabili le incertezze legate alla variabilità dei fenomeni, come viene idealizzata nella scienza e nella tecnologia. Si pensi al tempo di corretto funzionamento di un certo prodotto industriale: si può essere interessati a stabilire se è raro che il prodotto cessi di essere utilizzabile nell'anno che segue la fine della garanzia. Se sia effettivamente raro o no è un fatto

2 LEZIONE 1. ESPERIMENTI CASUALI, SPAZI CAMPIONARI, EVENTI

che può essere accertato, con consenso generale, acquisendo dati su un numero, sufficientemente elevato, di elementi sottoposti a prova.

Le incertezze puramente soggettive fanno riferimento a situazioni di ignoranza su fenomeni aleatori per le quali non è praticabile una valutazione largamente condivisa. Si pensi, ad esempio, in ambito finanziario, a quale sarà fra un anno il prezzo di una specifica azione. Sulle incertezze private si possono solo fare scommesse.

Le incertezze oggettivamente valutabili sono legate in modo diretto a dati potenzialmente producibili. Ad esempio, possiamo osservare effettivamente le durate di buon funzionamento di tanti esemplari del prodotto industriale d'interesse. Le incertezze private non hanno un tale appoggio: il prezzo dell'azione fra un anno è un fenomeno unico che non trova collocazione in modo univoco su uno sfondo di fenomeni essenzialmente equivalenti.

In concreto, il Calcolo delle Probabilità propone, per le situazioni di incertezza, modelli matematicamente convenienti tramite i quali valutare l'attendibilità di particolari predizioni. Tipicamente, tali modelli dipendono da parametri il cui valore numerico è di solito ignoto. La Statistica rende operativa la modellazione probabilistica, usando dati generati nella situazione di incertezza che si desidera studiare per 'stimare' il valore assunto, nella particolare applicazione, dai parametri del modello adottato.

Il Calcolo delle Probabilità modella ogni tipo di incertezza (oggettiva o soggettiva). È stato strutturato come teoria matematica da A.N. Kolmogorov (1903–1987) con la sua assiomatizzazione del 1933. Quindi prende le mosse da nozioni primitive e assiomi, da cui si deducono teoremi; poi appropriate definizioni permettono di enunciare e dimostrare ulteriori teoremi, e così via.

Anche la Statistica è utile a fronte di ogni tipo di incertezza. Non è però strutturata come teoria matematica. È una scienza matematica: in essa non si parte da assiomi, ma si fa uso con buon senso di strumenti tratti dal Calcolo delle Probabilità e da altre aree della Matematica per fornire orientamenti utili a risolvere problemi di 'stima' dei parametri ignoti del modello probabilistico adottato.

Si inizia dunque il percorso di questo insegnamento dal Calcolo delle Probabilità, in una forma adatta a trattare le incertezze pubbliche. Anche gli elementi di Statistica che si introdurranno a completamento del percorso saranno indirizzati a trattare incertezze pubbliche. Lo scopo dei primi passi sarà associare un contenuto di intuizione agli oggetti astratti della teoria assiomatica della Probabilità.

1.2 Esperimenti casuali

Il punto di partenza per acquisire l'intuizione di incertezze indagabili in forme largamente condivise da una comunità di soggetti è la nozione di esperimento casuale.

Definizione 1.1 *Un esperimento casuale, indicato con \mathcal{E} , è ogni attività, ripetibile essenzialmente nelle stesse condizioni secondo preassegnate regole, che, se eseguita, produce un risultato. Generalmente il risultato di \mathcal{E} (anch'esso rilevato secondo regole predeterminate) è incerto, perché varia al ripetersi delle prove.*

Un generico risultato di \mathcal{E} sarà indicato con s . Se, effettuato \mathcal{E} , si presenta s , si dice che \mathcal{E} realizza s . Si scriverà allora in breve $\mathcal{E} \longrightarrow s$.

Esempi di esperimenti casuali e di loro risultati:

- lanciare una moneta (per vedere che faccia esca);
 $s = \text{'esce testa'} = T$;
- lanciare due monete (per vedere che coppia di facce esca);
 $s = (\text{'esce testa'}, \text{'esce testa'}) = (T, T)$, dove il primo elemento della coppia ordinata descrive l'esito della prima moneta, il secondo elemento l'esito della seconda moneta: le due monete sono ovviamente distinguibili;
- estrarre una carta da un mazzo (per vedere che carta si presenta);
 $s = \heartsuit$ (asso di cuori);
- far girare un programma su un server remoto (per misurare il tempo di acquisizione dei risultati);
 $s = 6.28 \text{ ms}$ (millisecondi).

1.3 Spazi campionari

La prima descrizione dell'incertezza in gioco in un esperimento casuale \mathcal{E} si fa stabilendo quali risultati s sono possibili. Prima ancora di effettuare \mathcal{E} , occorre dunque dedurre dalle regole di \mathcal{E} che cosa sarà possibile osservare.

Definizione 1.2 *Lo spazio campionario di un esperimento casuale \mathcal{E} , indicato con S , è l'insieme di tutti i possibili risultati macroscopici s di \mathcal{E} , $S = \{s\}$.*

Quale s in S si realizzerà è incerto: $\mathcal{E} \longrightarrow ?$. È tuttavia certo che $? \in S$: S comprende tutti i possibili risultati dell'esperimento.

Conviene dare subito alcuni esempi di esperimenti casuali accompagnati dal proprio spazio campionario:

1. L'esperimento casuale $\mathcal{E}_1 = \text{'lancio di una moneta'}$ ha spazio campionario $S_1 = \{T, C\}$ dove $T = \text{'esce testa'}$ e $C = \text{'esce croce'}$.
2. L'esperimento casuale $\mathcal{E}_2 = \text{'lancio di due monete'}$ ha spazio campionario $S_2 = \{(T, T), (T, C), (C, T), (C, C)\} = \{T, C\} \times \{T, C\} = S_1 \times S_1$.

3. $\mathcal{E}_3 = \text{'estrazione di una carta da un mazzo (di 52 carte)'} ha spazio campionario $S_3 = \{s_1, s_2, \dots, s_{52}\}$, dove $s_1 = \text{'asso di cuori'}$, $s_2 = \text{'due di cuori'}$, \dots , $s_{52} = \text{'re di picche'}$.$
4. Si immagini di ispezionare un lotto di circuiti integrati fino a trovare un *chip* difettoso. Questa attività è un esperimento casuale, \mathcal{E}_4 . Abbreviato con d difettoso e con c conforme (non difettoso), lo spazio campionario di \mathcal{E}_4 è

$$S_4 = \{d, cd, ccd, cccd, \dots\}$$

dove d va interpretato *'trovato chip difettoso alla prima ispezione'*, dc va interpretato *'trovato chip conforme alla prima ispezione e difettoso alla seconda'*, eccetera.

5. Far girare un programma su un server remoto per misurare il tempo di acquisizione dei risultati corrisponde a un esperimento casuale \mathcal{E} con spazio campionario

$$S = \mathbb{R}^+ = \{t \in \mathbb{R} : t > 0\}$$

con il tempo misurato in millisecondi.

Uno spazio campionario può essere:

- finito (con cardinalità finita), cfr. esempi 1, 2, 3 sopra;
- infinito numerabile (con cardinalità del numerabile, ossia di \mathbb{N}), cfr. 4 sopra;
- infinito continuo (con la cardinalità del continuo, ossia di \mathbb{R}), cfr. 5.

L'incertezza nel discreto (spazio campionario finito o numerabile) e l'incertezza nel continuo richiedono descrizioni matematiche differenti (cfr. la suddivisione fra matematica discreta e analisi matematica). Perciò la distinzione fra caso discreto e caso continuo è fondamentale e va sempre tenuta presente. Le descrizioni possibili, come si vedrà, devono essere compatibili con assiomi molto semplici, e sfruttano le proprietà di \mathbb{N} e \mathbb{R} .

1.4 Eventi

Per un esperimento casuale \mathcal{E} con associato spazio campionario S l'incertezza da valutare è relativa a proposizioni che predicono il risultato di \mathcal{E} .

Definizione 1.3 Si dice **evento**, indicato con E (ma anche con A , B , A_1 , A_2 , eccetera) un sottoinsieme di S , $E \subseteq S$.

Ogni evento è dunque una collezione di possibili risultati di \mathcal{E} (da tutti a nessuno). Tradurre in un sottoinsieme di S una predizione espressa in forma verbale richiede sovente molta attenzione.

1.4.1 Eventi notevoli

Alcuni eventi sono particolarmente importanti e meritano un nome.

- L'intero spazio campionario S è detto **evento certo**.
- L'unico sottoinsieme di S privo di elementi, indicato con \emptyset , è detto **evento impossibile**.
- Un evento costituito da un solo elemento di S , $\{s\}$ con $s \in S$, è detto **evento elementare**.
- Se $S = \{s_1, s_2, s_3\}$, un **evento non elementare** è, ad esempio, $A = \{s_1, s_2\}$.

Una nota di terminologia: S è detto anche **spazio degli eventi elementari** oppure **spazio fondamentale**.

1.4.2 Eventi: esempi

Esempio 1.1 Si consideri l'esperimento $\mathcal{E} = \text{'lancio di due monete'}$, con spazio campionario

$$S = \{(T, T), (T, C), (C, T), (C, C)\}.$$

Qui (T, T) è un evento elementare, *'ogni moneta produce o testa o croce'* è un evento certo, *'le due monete, lanciate, diventano colombe bianche e volano via'* è un evento impossibile. Si noti che mentre l'insieme vuoto è unico, l'evento impossibile ha infinite espressioni linguistiche. \diamond

Esempio 1.2 Per l'esperimento \mathcal{E} che consiste nel far eseguire un programma a un *server* remoto e misurare il tempo di risposta, con spazio campionario

$$S = [0, +\infty) = \{t \in \mathbb{R} : t \geq 0\},$$

dove il tempo è misurato in millisecondi, sono eventi elementari $\{\pi\}$ e $\{e\}$ ed eventi non elementari

$$\begin{aligned} A &= \{t : t > 3.14 \text{ ms}\} = (3.14, \infty) \\ B &= \{t : 0 < t \leq 5 \text{ ms}\} = (0, 5]. \end{aligned}$$

Invece $C = (-10, -5]$ è chiaramente un evento impossibile. \diamond

1.4.3 Eventi: realizzati e non realizzati

Effettuato \mathcal{E} con spazio campionario S , uno e un solo $s \in S$ sarà osservato. L'esito s osservato è detto **realizzazione** di \mathcal{E} .

Definizione 1.4 Dato un evento $E \subseteq S$, si dice che E è **realizzato** se è osservato un $s \in E$, altrimenti si dice che E **non è realizzato**.

6 LEZIONE 1. ESPERIMENTI CASUALI, SPAZI CAMPIONARI, EVENTI

Osservazione: si capisce ora meglio perché S è certo, \emptyset impossibile. L'evento certo si realizza sempre, l'evento impossibile mai.

Esempio 1.3 Si consideri l'esperimento $\mathcal{E} = \text{'lancio di un dado'}$, con spazio campionario

$$S = \{s_1, s_2, s_3, s_4, s_5, s_6\},$$

dove $s_i = \text{'esce la faccia } i\text{'}$. Si fissi l'attenzione sull'evento

$$E = \{s_2, s_4, s_6\} = \text{'esce una faccia pari'}.$$

Se $\mathcal{E} \longrightarrow s_4$, allora E è realizzato. Se $\mathcal{E} \longrightarrow s_1$, allora E non è realizzato. \diamond

1.5 Esercizi

Esercizio 1.1 Si stabilisca se il prossimo derby fra le due note squadre di calcio di Milano costituisce un esperimento casuale.

Esercizio 1.2 Si stabilisca se le prossime elezioni presidenziali americane costituiscono un esperimento casuale.

Esercizio 1.3 Si consideri il lancio di un dado seguito dal lancio di tante monete quanto è il punteggio del dado. Si dica se è un esperimento casuale.

Esercizio 1.4 Con riferimento all'esperimento casuale *'lancio di tre monete'*, si indichi un evento elementare e un evento impossibile. Si calcoli la cardinalità dello spazio campionario.

Esercizio 1.5 Si spieghi perché 1 e $\{1\}$ sono diversi e $\{a\}$ e $\{a, a\}$ sono uguali.

Esercizio 1.6 Si dica perché $\{a, b\}$ e $\{b, a\}$ sono uguali e \emptyset e $\{\emptyset\}$ sono diversi.

Esercizio 1.7 Il significato del simbolo (a, b) dipende dal contesto. Se si sta parlando di intervalli e $a < b$, allora (a, b) rappresenta l'intervallo aperto $\{x \in \mathbb{R} : a < x < b\}$. Se non si sta parlando di intervalli, (a, b) rappresenta una **coppia ordinata**, ossia $(a, b) = \{\{a\}, \{a, b\}\}$ (diversamente da $\{a, b\}$, in (a, b) si distingue ciò che sta al primo posto da ciò che sta al secondo posto). Si definisca la terna ordinata (a, b, c) , dove $a \in A, b \in B, c \in C$.

Esercizio 1.8 Il **prodotto cartesiano** di due insiemi A e B è

$$A \times B = \{(a, b), a \in A, b \in B\}.$$

Due esperimenti casuali, \mathcal{E}_1 con spazio campionario S_1 ed \mathcal{E}_2 con spazio campionario S_2 , privi di comunicazione fra loro, formano un esperimento casuale congiunto $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2)$ con spazio campionario $S = S_1 \times S_2$. Si diano tre esempi di esperimento il cui spazio campionario ha questa struttura.

Esercizio 1.9 Siano A e B due insiemi non vuoti. Una **applicazione** o **funzione** f da A a valori in B , indicata con $f : A \rightarrow B$, fa corrispondere ad ogni $a \in A$ un unico $b \in B$, detto immagine di a tramite f e scritto $f(a)$. Ha come *input* $a \in A$ e come *output* $f(a) \in B$. Si mostri che $\{(a, f(a)), a \in A\} \subseteq A \times B$.

Esercizio 1.10 Siano $f : [0, 1] \rightarrow \mathbb{R}$ e $g : [0, 1] \rightarrow \mathbb{R}$ definite da $f(x) = 1$ e $g(x) = |x| + |x - 1|$. Si mostri che f e g sono uguali.

Esercizio 1.11 Si dica quali dei seguenti insiemi ha cardinalità finita e quali del numerabile: $\{0, 1, 2, 3\}$, $\{n \in \mathbb{N} : n = 2m, m \in \mathbb{N}\}$, $\{n \in \mathbb{N} : n \text{ divide } 2^{10}\}$, $\{n \in \mathbb{N} : n \text{ primo}\}$.

Esercizio 1.12 Si dica quali dei seguenti insiemi ha cardinalità finita o numerabile e quali del continuo: $\{n \in \mathbb{N} : n = 2m + 1, m \in \mathbb{N}\}$, $(0, 1)$, $\{x \in \mathbb{R} : x^2 + 1 = 0\}$, \mathbb{R}^2 .

Esercizio 1.13 Si spieghi perché $\{0, 1\}$, $[0, 1]$, $[0, 1)$, $(0, 1)$ sono sottoinsiemi di \mathbb{R} diversi.

Esercizio 1.14 Siano $a_i \in \mathbb{R}$, $i \in \{0, 1, \dots, n\}$. La **sommatoria** di a_i per i che va da 1 a n è definita ricorsivamente da $\sum_{i=0}^0 a_i = a_0$ e $\sum_{i=0}^n a_i = a_n + \sum_{i=0}^{n-1} a_i$. Si giustifichino le seguenti proprietà:

$$\begin{aligned} \sum_{i=0}^n a_i &= \sum_{j=0}^n a_j, \\ \sum_{i=0}^n a_i &= \sum_{i=0}^m a_i + \sum_{i=m+1}^n a_i \end{aligned}$$

per ogni m tale che $0 \leq m \leq n - 1$,

$$\sum_{i=1}^n c = nc,$$

ove $c \in \mathbb{R}$. Infine si giustifichi, per $b_i \in \mathbb{R}$, $i \in \{0, 1, \dots, n\}$,

$$\begin{aligned} \sum_{i=0}^n (a_i + b_i) &= \sum_{i=0}^n a_i + \sum_{i=0}^n b_i, \\ \sum_{i=0}^n ca_i &= c \sum_{i=0}^n a_i. \end{aligned}$$

Esercizio 1.15 Siano $a_i \in \mathbb{R}$, $i \in \{0, 1, \dots, n\}$. La **produttoria** di a_i per i che va da 1 a n è definita ricorsivamente da $\prod_{i=0}^0 a_i = a_0$ e $\prod_{i=0}^n a_i = a_n \prod_{i=0}^{n-1} a_i$. Si giustifichino le seguenti proprietà:

$$\begin{aligned} \prod_{i=0}^n a_i &= \prod_{j=0}^n a_j, \\ \prod_{i=0}^n a_i &= \left(\prod_{i=0}^m a_i \right) \left(\prod_{i=m+1}^n a_i \right) \end{aligned}$$

per ogni m tale che $0 \leq m \leq n-1$. Si giustifichi infine

$$\prod_{i=1}^n c = c^n$$

ove $c \in \mathbb{R}$.

Esercizio 1.16 Siano $b, b_1, b_2 > 0$ e $x, y \in \mathbb{R}$. Si richiama che valgono le seguenti **proprietà delle potenze**

$$\begin{aligned} b^x b^y &= b^{x+y} && \text{prodotto di potenze con la stessa base} \\ b_1^x b_2^x &= (b_1 b_2)^x && \text{prodotto di potenze con lo stesso esponente.} \end{aligned}$$

Si mostri che per ogni $n \geq 2$, a_i e $x \in \mathbb{R}$, $b_i > 0$ si ha

$$\prod_{i=1}^n b^{a_i} = b^{\sum_{i=1}^n a_i}$$

e

$$\prod_{i=1}^n b_i^x = \left(\prod_{i=1}^n b_i \right)^x.$$

Esercizio 1.17 Si mostri per induzione matematica che $\sum_{i=1}^n i = n(n+1)/2$.

Esercizio 1.18 Si mostri per induzione matematica (o più direttamente considerando una somma telescopica) che se $x \neq 1$ allora $1 + x + x^2 + \dots + x^n$ è uguale a $(1 - x^{n+1})/(1 - x)$, ossia

$$\sum_{i=0}^n x^i = \frac{1 - x^{n+1}}{1 - x}.$$

Se invece $x = 1$, $\sum_{i=0}^n x^i = n + 1$.

1.6 Appendice: introduzione all'ambiente R

R è un ambiente per il calcolo e la grafica statistica, molto utilizzato, da statistici e non statistici, per lo sviluppo di software statistico e di *data science*. Caratteristiche primarie di R sono:

- è un progetto GNU (quindi *open-source*)
- il sito principale è all'indirizzo <http://www.r-project.org>
- un mirror in Italia da cui scaricare il codice sorgente è all'indirizzo <http://cran.stat.unipd.it>

- è disponibile per Linux, (Mac) OS X, Windows.

Installato il software, si può iniziare una sessione di lavoro. Un doppio click sulla icona di R apre la finestra di comando che presenta il prompt

```
>
```

Nel suo uso più scontato, R funziona come una calcolatrice. Ad esempio, scrivendo dopo il prompt l'espressione aritmetica 2^{10} e premendo ENTER per l'esecuzione, si ottiene la valutazione

```
> 2^10
[1] 1024
```

In modo analogo, espressioni logiche definite da operatori relazionali danno

```
> 1 >= 0
[1] TRUE
> 2/4 == 4/8
[1] TRUE
> 5 != 25/5
[1] FALSE
```

Se un comando non è completo, la risposta è il prompt + (invito ad aggiungere). Ovviamente

```
> 0/0
[1] NaN
> -1/0
[1] -Inf
```

L'operatore = consente l'assegnazione. Una assegnazione valuta una espressione salvandone il risultato in un oggetto individuato da un nome. Per visualizzare il risultato basta scrivere il nome dell'oggetto. Ad esempio

```
> n = 0:10
> n
[1] 0 1 2 3 4 5 6 7 8 9 10
> y = 2^n
> y
[1] 1 2 4 8 16 32 64 128 256 512 1024
> x = 3*y+1
> x
[1] 4 7 13 25 49 97 193 385 769 1537 3073
```

Nota: i comandi `a = 1`, `a <- 1` e `1 -> a` sono equivalenti.

In R i nomi degli oggetti (variabili numeriche, vettori, matrici, stringhe, funzioni, eccetera) sono *case sensitive*, per cui `a` e `A` possono rappresentare oggetti diversi. Conviene che il nome di un oggetto inizi con una lettera; come caratteri successivi al primo sono ammessi lettere, cifre, punto e trattino basso.

10LEZIONE 1. ESPERIMENTI CASUALI, SPAZI CAMPIONARI, EVENTI

Anche le funzioni elementari sono disponibili in R. Ad esempio il valore assoluto, **abs**, la radice quadrata, **sqrt**, esponenziale e logaritmo naturale, **exp** e **log**, le funzioni trigonometriche dirette e inverse, **sin cos tan asin acos atan**. Il nome **pi** ha preassegnato il valore π . Meglio evitare di ridefinire oggetti già impegnati da R (l'assegnazione **pi=0.5** può creare errori). Esempi d'uso:

```
> pi
[1] 3.141593
> sin(pi/2)
[1] 1
> sin(pi)
[1] 1.224647e-16
> sqrt(4)
[1] 2
> sqrt(atan(1))
[1] 0.886227
> exp(1)
[1] 2.718282
> exp(-2)
[1] 0.1353353
```

R salva gli oggetti per nome in un'area dedicata detta area di lavoro (*workspace*). Per controllare quali oggetti vi siano presenti si usa

```
> ls()
[1] "n" "x" "y"
```

La funzione **rm()** rimuove uno o più oggetti dall'area di lavoro. Ad esempio

```
> rm(x,y)
> ls()
[1] "n"
```

Quando si inizia da zero una sessione di lavoro conviene che l'area di lavoro non contenga oggetti. Per esserne certi basta dare il comando **rm(list = ls())**. Quindi

```
> rm(list=ls())
> ls()
character(0)
> q()
Save workspace image? [y/n/c]:
```

Il comando **q()** chiude una sessione di lavoro. Se si risponde **y** alla domanda, l'area di lavoro viene salvata e ripresentata all'inizio della sessione successiva.

Lezione 2

Probabilità classica, frequentista, assiomatica

2.1 Premessa

Dato un evento $E \subseteq S$, dove S è lo spazio campionario di un esperimento casuale \mathcal{E} , si modella matematicamente l'incertezza dando una valutazione numerica di quanto è facile che E risulti realizzato, effettuato \mathcal{E} . La valutazione sarà la probabilità di E , indicata con $P(E)$. Un evento che non si realizza mai, un evento impossibile, avrà probabilità 0. Uno che si realizza sempre, un evento certo, avrà probabilità 1. Per un evento che in alcune prove si realizza e in altre no, $P(E)$ assumerà un valore intermedio fra 0 e 1. La probabilità di un evento può essere espressa anche in percentuale, come $100P(E)\%$. Quindi, se $P(E) = 0.5$, si può anche scrivere $P(E) = 50\%$.

La valutazione dell'incertezza non viene fatta isolatamente per un dato evento E , ma collettivamente per tutti gli eventi di una classe di eventi, il dominio dell'applicazione P . Le relazioni che gli eventi hanno fra loro vincolano in effetti le valutazioni. Si considerino ad esempio due eventi E ed F per cui vale la relazione $E \subseteq F$, ossia E **implica** F (la realizzazione di E necessariamente comporta la realizzazione di F). Sarà più facile che l'esperimento realizzi F piuttosto che E , e quindi dovrà essere $P(E) \leq P(F)$.

Le relazioni fra eventi, oltre che dall'implicazione, possono sorgere dalle operazioni logiche su eventi, che saranno chiamate in breve operazioni su eventi.

2.2 Operazioni su eventi

Le operazioni su eventi costruiscono nuovi eventi a partire da eventi dati. Le principali sono le tre seguenti.

- Passaggio all'**evento contrario**.

Sia A un evento assegnato. L'evento contrario di A , detto anche '*non A*', è l'evento che si realizza se e solo se A non si realizza. È indicato con \bar{A} . Ad esempio $\bar{\bar{S}} = \bar{\emptyset}$, $\bar{\emptyset} = S$. Chiaramente, $\bar{\bar{A}} = A$.

- **Unione** di due o più eventi.

Siano A e B due eventi assegnati. L'evento unione di A e B è l'evento che si realizza se e solo se si realizza A oppure B (o entrambi). È indicato con $A \cup B$. Per una successione finita o numerabile di eventi assegnati, A_i , $i \in I \subseteq \mathbb{N}$, l'evento unione è l'evento che si realizza se e solo se almeno uno degli A_i si realizza. Si indica con $\cup_{i \in I} A_i$.

- **Intersezione** di due o più eventi.

Siano A e B due eventi assegnati. L'evento intersezione di A e B è l'evento che si realizza se e solo se si realizzano sia A sia B . È indicato con $A \cap B$. Per una successione finita o numerabile di eventi assegnati, A_i , $i \in I \subseteq \mathbb{N}$, l'evento intersezione è l'evento che si realizza se e solo se tutti gli A_i si realizzano. Si indica con $\cap_{i \in I} A_i$.

Si considera inoltre abbastanza spesso **l'evento differenza**, definito come

$$A \setminus B = A \cap \bar{B}.$$

Essendo definita tramite altre operazioni, la differenza non è un'operazione su eventi fondamentale. Si possono assumere come fondamentali passaggio all'evento contrario e **unione**. Infatti l'intersezione è definibile tramite contrario e unione: $A \cap B = \overline{\bar{A} \cup \bar{B}}$ (formula di De Morgan).

È utile disegnare i diagrammi di Venn corrispondenti alle operazioni su eventi, rappresentando lo spazio campionario S come un quadrato unitario.

Si noti che, sopra, unione e intersezione di eventi sono considerate per un insieme di indici I finito o numerabile. Quindi sono considerabili come eventi anche limiti di successioni di eventi. Nel Calcolo delle Probabilità non si considerano invece unioni o intersezioni di collezioni di eventi A_i per cui I ha cardinalità la cardinalità del continuo.

2.3 Eventi incompatibili

Eventi che non possono realizzarsi simultaneamente sono di particolare interesse per la valutazione quantitativa dell'incertezza.

Definizione 2.1 Due eventi A e B per cui $A \cap B = \emptyset$ sono detti **incompatibili**, o anche **mutuamente esclusivi**.

Se A e B non sono incompatibili, si diranno **compatibili**. Per essi $A \cap B \supset \emptyset$.

Definizione 2.2 Una successione di eventi $A_i, i \in I \subseteq \mathbb{N}$, si dice costituita da **eventi a due a due incompatibili** (o anche **mutuamente esclusivi**), se

$$i, j \in I \text{ con } i \neq j \implies A_i \cap A_j = \emptyset.$$

Esempio 2.1 Si consideri l'esperimento $\mathcal{E} = \text{'lancio di un dado'}$, con spazio campionario $S = \{s_i, i = 1 \dots, 6\}$, dove $s_i = \text{'esce la faccia } i\text{'}$. Allora $A = \text{'esce una faccia pari'}$ e $B = \text{'esce una faccia dispari'}$ sono eventi incompatibili: $\{s_2, s_4, s_6\} \cap \{s_1, s_3, s_5\} = \emptyset$. Anche gli eventi elementari sono eventi a due a due incompatibili, ad esempio $\{s_1\} \cap \{s_2\} = \emptyset$.

2.4 La classe degli eventi e le probabilità

È bene, dati \mathcal{E} e S , considerare gli eventi A, B , eccetera, nel loro complesso. Si introduce perciò la **classe degli eventi**, che è la collezione di tutti gli eventi di cui ha senso parlare nel contesto del dato esperimento casuale.

La classe degli eventi è indicata con \mathcal{F} (o \mathcal{B}). Ovviamente $\mathcal{F} \subseteq \mathcal{P}(S)$, dove $\mathcal{P}(S)$ è l'insieme delle parti di S . Se $\mathcal{F} \subset \mathcal{P}(S)$, \mathcal{F} dovrà essere 'simile' a $\mathcal{P}(S)$: le usuali operazioni fatte su eventi in \mathcal{F} dovranno restituire ancora eventi in \mathcal{F} .

Costruiti S e \mathcal{F} , si modella infine l'incertezza presente nell'esperimento casuale \mathcal{E} attribuendo agli eventi una valutazione della facilità con cui risultano realizzati (la prossima volta che si effettua \mathcal{E}).

A ogni $E \in \mathcal{F}$, si associa la **probabilità** di E , indicata con $P(E)$. Quindi P è un'opportuna applicazione $P : \mathcal{F} \rightarrow \mathbb{R}$, per la cui esplicitazione sono utili alcuni schemi concettuali. Probabilità classica, frequentista, soggettivista, assiomatica sono i più rilevanti. La probabilità soggettivista tratta anche le incertezze private, meno rilevanti per questo corso. Un veloce sguardo ad essa sarà offerto in Appendice.

2.5 Probabilità classica

Quando lo spazio campionario S ha cardinalità finita $|S|$ e la classe degli eventi è $\mathcal{F} = \mathcal{P}(S)$, pure con cardinalità finita $2^{|S|}$, un modo semplice per valutare $P(E)$ è seguire la cosiddetta **definizione classica** della probabilità (Laplace, 1812). Per essa, la **probabilità** $P(E)$ di un evento E si ottiene **rapportando il numero di risultati dell'esperimento che comportano che E sia realizzato (casi favorevoli) al numero complessivo dei risultati possibili (casi possibili)**, supponendo che tutti i risultati siano egualmente possibili. Quindi

$$P(E) = \frac{|E|}{|S|}. \quad (2.1)$$

Ad esempio, per l'esperimento casuale $\mathcal{E} = \text{'si lanciano due monete'}$ l'evento $E = \text{'si osserva almeno una testa'}$ ha probabilità $P(E) = 3/4$ in base alla (2.1).

La (2.1) misura l'incertezza sulla realizzazione di E in accordo con la concezione secondo cui gli 'effetti' sono proporzionali alle 'cause'.

Il calcolo $P(E) = |E|/|S|$ si basa sull'assunto che gli eventi elementari siano **equiprobabili**: per ogni $s \in S$

$$P(\{s\}) = \frac{|\{s\}|}{|S|} = \frac{1}{|S|}$$

(secondo la terminologia di Laplace, 'egualmente possibili'). Per i giochi di sorte con monete, dadi, carte, roulette, e simili, per i quali la probabilità classica fu introdotta, l'equiprobabilità degli eventi elementari fornisce generalmente una buona approssimazione dell'incertezza in azione (salvo trucchi ...). Non è però la situazione generale. Anche assumere $|S|$ finita è assai limitativo.

Dalla (2.1) si vede subito che $P(\emptyset) = 0$, $P(S) = 1$, in accordo con il fatto che S è una predizione vera e \emptyset una predizione falsa dell'esito di \mathcal{E} . Si vede inoltre che $0 \leq P(E) \leq 1$, $P(\bar{E}) = 1 - P(E)$, $E \subseteq F \implies P(E) \leq P(F)$, $E \cap F = \emptyset \implies P(E \cup F) = P(E) + P(F)$. Si ottengono facilmente altri risultati che, nella probabilità assiomatica, saranno o adottati come assiomi o dimostrati come teoremi.

2.6 Probabilità frequentista

Per la concezione frequentista, dati un esperimento casuale \mathcal{E} , descritto da spazio campionario S e classe degli eventi \mathcal{F} , e un evento $E \in \mathcal{F}$, conviene pensare a $P(E)$ come la propensione a produrre la realizzazione di E che l'esperimento \mathcal{E} ha, in base alle sue preassegnate regole. Quindi $P(E)$ è una misura della facilità di vedere l'evento realizzato in future repliche dell'esperimento.

Assumendo omogeneità nel tempo delle ripetizioni dell'esperimento, il valore $P(E)$, relativo a ciò che si deve ancora sperimentare, può venire empiricamente approssimato da un dato di esperienza: la frequenza relativa di realizzazione di E in un insieme di repliche di E già effettuate. Così sarà l'esperimento stesso, replicato tante volte, a permettere di valutare $P(E)$, superando le limitazioni della probabilità classica.

Si supponga dunque di replicare \mathcal{E} un numero grande di volte, R , e di prendere nota di ciò che accade relativamente alla realizzazione di E nelle varie repliche, secondo lo schema della Tabella 2.1. Il valore $P(E)$ sarà approssimato dalla proporzione di esiti s_i al quesito se E risulta realizzato. Codificando s_i con 1, no con 0, si ha quindi

$$P(E) \approx \frac{1}{R} \sum_{i=1}^R y_i,$$

dove $y_i \in \{0, 1\}$ indica la codificazione dell'esito della i -esima replicazione e \approx una valutazione approssimata.

Tabella 2.1: Diario di laboratorio di una sperimentazione frequentista.

replicazione di \mathcal{E}	E realizzato?	codificazione numerica y_i
1	sì	1
2	sì	1
3	no	0
\vdots	\vdots	\vdots
R	no	0

La $\sum_{i=1}^R y_i$ è la **frequenza assoluta** di realizzazioni di E nelle R repliche. Dividere una frequenza assoluta per il numero totale di casi dà la corrispondente **frequenza relativa**. Quindi la probabilità è una idealizzazione (per R molto grande) di una frequenza relativa. Da qui l'aggettivo frequentista.

2.7 Probabilità assiomatica

L'assiomatizzazione del Calcolo delle Probabilità offre le regole per calcolare correttamente nuove probabilità a partire da probabilità assegnate. Seguendo le regole, i calcoli funzionano qualunque sia lo scopo dell'analisi o la fonte delle probabilità preassegnate, e qualunque sia la nostra concezione di probabilità.

Sia dato dunque \mathcal{E} , e di conseguenza S , descrizione dei possibili esiti fisici di \mathcal{E} . Il primo passo dell'assiomatizzazione è descrivere bene $\mathcal{F} \subseteq \mathcal{P}(S)$. La classe degli eventi è descritta bene se le ordinarie operazioni su eventi, iterate o composte fino al numerabile, non portano fuori di \mathcal{F} .

In pratica, la scelta di $\mathcal{F} \subseteq \mathcal{P}(S)$ è guidata dalla cardinalità di S .

Se S ha cardinalità finita o numerabile, si farà sempre la scelta $\mathcal{F} = \mathcal{P}(S)$. Con questa scelta le operazioni su eventi di \mathcal{F} non possono fare uscire da \mathcal{F} .

Se S ha la cardinalità del continuo, accade che $\mathcal{P}(S)$ risulta 'troppo grande' e si deve scegliere una $\mathcal{F} \subset \mathcal{P}(S)$ 'somigliante' a $\mathcal{P}(S)$, cioè con analoghe proprietà di chiusura alle operazioni su eventi.

2.7.1 La classe degli eventi come σ -algebra

Si postula che \mathcal{F} sia sempre una σ -algebra di parti di S , ossia una collezione di parti di S tale che

- i) $S \in \mathcal{F}$
- ii) $A \in \mathcal{F} \implies \bar{A} \in \mathcal{F}$

iii) $A_i \in \mathcal{F}$ per ogni $i \in I \subseteq \mathbb{N} \implies \cup_{i \in I} A_i \in \mathcal{F}$.

Dunque i) richiede che S sia un evento e ii) postula che se A è un evento anche \bar{A} sia sempre un evento. La più articolata condizione iii) esige che, data una successione finita o numerabile di eventi, anche la loro unione sia un evento, ossia che si possa sempre parlare della realizzazione di almeno uno degli eventi di una successione finita o numerabile.

Si noti subito che anche \emptyset è un evento: $\emptyset = \bar{S} \in \mathcal{F}$ per i) e ii). Anche le intersezioni fino al numerabile di eventi assegnati costituiscono un evento. Infatti una formula di De Morgan (cfr. Esercizio 2.4) mostra che $\cap_{i \in I} A_i = \overline{\cup_{i \in I} \bar{A}_i}$ e \mathcal{F} è chiusa rispetto a negazione e unione per ii) e iii). Anche le differenze di due eventi assegnati sono un evento: $A, B \in \mathcal{F} \implies A \setminus B = A \cap \bar{B} \in \mathcal{F}$ per la chiusura di \mathcal{F} rispetto a negazione e intersezione.

2.7.2 Spazi probabilizzabili

Ad un esperimento casuale \mathcal{E} si associa dunque

- 1 lo spazio campionario S (un insieme non vuoto di natura qualunque)
- 2 la classe degli eventi \mathcal{F} , σ -algebra di parti di \mathcal{F} .

Definizione 2.3 La coppia ordinata (S, \mathcal{F}) , dove S è non vuoto e \mathcal{F} è una σ -algebra di parti di \mathcal{F} , è detta **spazio probabilizzabile**.

2.7.3 Misure di probabilità

Definizione 2.4 Dato lo spazio probabilizzabile (S, \mathcal{F}) , si dice **misura di probabilità** su \mathcal{F} una applicazione

$$P : \mathcal{F} \rightarrow [0, 1]$$

che soddisfi le tre proprietà, dette **assiomi di Kolmogorov**,

- A1** $P(A) \geq 0$ per ogni $A \in \mathcal{F}$ (assioma di non-negatività)
- A2** $P(S) = 1$ (assioma di normalizzazione)
- A3** Per ogni successione A_i , $i \in I \subseteq \mathbb{N}$, di eventi a due a due incompatibili

$$P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i) \quad (\text{assioma di } \sigma\text{-additività}).$$

Il Calcolo delle Probabilità ha dunque solo tre assiomi. I primi due sono molto semplici ed evidenti sia per la probabilità classica sia per quella frequentista. Il terzo ovviamente è il più informativo. Anch'esso è evidente, almeno quando I ha cardinalità finita. Si noti subito che copre anche il caso $|I| = |\mathbb{N}|$. Ciò comporta che limiti di successioni di eventi hanno probabilità che è il limite della corrispondente successione di probabilità.

2.7.4 Spazi probabilizzati

Definizione 2.5 Sia (S, \mathcal{F}) uno spazio probabilizzabile e P una misura di probabilità su \mathcal{F} . La terna ordinata (S, \mathcal{F}, P) è detta **spazio probabilizzato**.

Le incertezze in gioco nell'esperimento casuale \mathcal{E} sono completamente descritte dalla modellazione con un particolare spazio probabilizzato. Quindi il processo di realizzazione $\mathcal{E} \longrightarrow ?$ è ora descritto da (S, \mathcal{F}, P) per opportuni S , \mathcal{F} , P . La modellazione più delicata è ovviamente quella di P .

2.7.5 Le variabili casuali (v.c.) e le σ -algebre di Borel

Le variabili casuali (in sigla v.c.), cui sarà dedicata buona parte delle successive lezioni, sono spazi probabilizzati dove S è numerico. Sono indicate il simbolo X (o Y o simile), che sta per un oggetto numerico il cui valore sarà realizzato da un esperimento casuale, e quindi varia per effetto del caso replicando l'esperimento.

Se $S = \mathbb{R}$ (o un intervallo, anche illimitato, quindi anche una semiretta) la classe degli eventi è la σ -algebra di Borel associata ad \mathbb{R} , indicata con \mathcal{B}_1 . Questa è la più piccola σ -algebra di parti di \mathbb{R} che contiene tutti gli intervalli $(a, b]$ con $a < b$. Quindi \mathcal{B}_1 contiene non solo gli intervalli ma anche tutte le parti di \mathbb{R} che sono limiti di successioni di parti di costruite con le usuali operazioni insiemistiche a partire da un numero finito di intervalli.

Se $S = \mathbb{R}^2$, allora $\mathcal{F} = \mathcal{B}_2$, la σ -algebra di Borel su \mathbb{R}^2 . Questa è la più piccola σ -algebra di parti di \mathbb{R}^2 che contiene tutti gli intervalli di \mathbb{R}^2 di forma $(a_1, b_1] \times (a_2, b_2]$ con $a_1 < b_1$, $a_2 < b_2$.

Analogamente, in dimensione superiore, $S = \mathbb{R}^d$, $d > 2$, si avrà $\mathcal{F} = \mathcal{B}_d$, la σ -algebra di Borel su \mathbb{R}^d . Questa è la più piccola σ -algebra di parti di \mathbb{R}^d che contiene tutti gli intervalli di \mathbb{R}^d di forma $(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_d, b_d]$ con $a_i < b_i$ per ogni $i \in \{1, \dots, d\}$.

Una nota a margine per i più curiosi. Si dimostra che \mathcal{B}_1 (e quindi \mathcal{B}_d , $d \geq 2$) ha la cardinalità del continuo, mentre $\mathcal{P}(\mathbb{R})$ ha una cardinalità superiore a quella del continuo. In pratica gli elementi di \mathcal{B}_1 sono le parti di \mathbb{R} approssimabili arbitrariamente bene con operazioni insiemistiche su un numero finito (ma sempre più grande) di intervalli. Non tutte le parti di \mathbb{R} hanno questa approssimabilità. Tuttavia, quelle che non la hanno non ci interessano perché non si saprebbe dire sempre se, come evento, risultano realizzate oppure no, effettuato \mathcal{E} . Analogo discorso si può fare per le parti di \mathbb{R}^d non in \mathcal{B}_d .

2.8 Esercizi

Esercizio 2.1 Sia A un evento. Si argomenti che l'evento $\bar{\bar{A}}$ si realizza se e solo se A si realizza, e quindi $\bar{\bar{A}} = A$.

Esercizio 2.2 Sia A un evento. Si scriva una formula che rappresenta l'affermazione che o si realizza A o si realizza \bar{A} (principio del terzo escluso).

Esercizio 2.3 Siano E, F eventi. Si mostri che la relazione E implica F equivale a $E = E \cap F$.

Esercizio 2.4 Un evento intersezione non si realizza se e solo se almeno uno degli eventi coinvolti nell'intersezione non si realizza. Si scriva una formula che rappresenta tale affermazione. Si mostri che è equivalente alla **formula di De Morgan** $\cap_{i \in I} A_i = \overline{\cup_{i \in I} \bar{A}_i}$.

Esercizio 2.5 Si mostri che per eventi A e B non impossibili non vale la proprietà commutativa della differenza: $(A \setminus B) \neq (B \setminus A)$.

Esercizio 2.6 Si consideri l'esperimento casuale che consiste nel lancio di un dado seguito dal lancio di tante monete quante è indicato dal punteggio del dado. Si individuino tre eventi a due a due incompatibili ma non elementari.

Esercizio 2.7 Sia A un evento. Si mostri che A e \bar{A} sono incompatibili.

Esercizio 2.8 Siano A, B, C eventi a due a due incompatibili. Si mostri che la loro realizzazione simultanea è impossibile: $A \cap B \cap C = \emptyset$.

Esercizio 2.9 Sia $S = \{e_1, e_2, e_3\}$ lo spazio campionario di un esperimento casuale. Si scriva $\mathcal{P}(S)$ elencandone gli elementi. Si controlli che la cardinalità di $\mathcal{P}(S)$ è $2^3 = 8$.

Esercizio 2.10 Una **successione** è una applicazione $f : I \rightarrow \mathbb{R}$ dove $I \subseteq \mathbb{N}$, spesso \mathbb{N} o $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. Il suo generico termine si indica usualmente con a_n , quindi $a_n = f(n)$, $n \in I$. Si dice che una successione con infiniti termini ($|I| = |\mathbb{N}|$) ha **limite** $a \in \mathbb{R}$, e si scrive

$$\lim_{n \rightarrow \infty} a_n = a,$$

se a approssima a_n arbitrariamente bene per n sufficientemente grande, ossia se per ogni assegnata tolleranza $\varepsilon > 0$ vi è un naturale \bar{n} (che in generale dipende da ε : è tanto più grande quanto più ε si avvicina a 0) a partire dal quale la tolleranza è sempre soddisfatta: $|a_n - a| < \varepsilon$ per ogni $n > \bar{n}$. Si dimostri che per una successione costante, $a_n = c$, $\lim_{n \rightarrow \infty} a_n = c$. Si dimostri che per $a_n = 1/n$ $\lim_{n \rightarrow \infty} a_n = 0$.

Esercizio 2.11 Il **fattoriale** di un valore $n \in \mathbb{N}$, indicato con $n!$, è definito per ricorrenza come

$$n! = \begin{cases} 1 & \text{se } n = 0 \\ n(n-1)! & \text{se } n > 0. \end{cases}$$

Si mostri per induzione matematica che $n! < n^n$ se $n > 1$.

Esercizio 2.12 Si giustifichi il fatto che una successione a_n monotona non decrescente ($a_n \leq a_{n+1}$ per ogni $n \in \mathbb{N}$) o converge ad un limite $a \in \mathbb{R}$ o diverge a $+\infty$ (ossia, per ogni $K > 0$ esiste \bar{n} tale che per ogni $n > \bar{n}$ valga $a_n > K$; si scriverà allora $\lim_{n \rightarrow \infty} a_n = +\infty$).

Esercizio 2.13 Sia a_n , $n \in \mathbb{N}$, una successione con valori non negativi. Si definisce **somma della serie** di tali valori, indicata con $\sum_{i=0}^{\infty} a_n$, il limite al divergere di n della successione delle somme parziali $\sum_{i=0}^n a_n$, ossia

$$\sum_{i=0}^{\infty} a_n = \lim_{n \rightarrow \infty} \sum_{i=0}^n a_n .$$

Si mostri che, per $-1 < x < 1$,

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x} .$$

Esercizio 2.14 Tenendo conto del risultato dell'esercizio 2.12 e del fatto che per ogni evento $P(E) \leq P(S) = 1$, si mostri che se A_i è una successione di eventi a due a due incompatibili allora $\sum_{i \in I} P(A_i)$ converge a un valore in $[0, 1]$.

Esercizio 2.15 Sia A un evento e A_i , $i \in I$, una partizione di A in eventi. Si mostri che, se $|I| = |\mathbb{N}|$, allora $\sum_{i \in I} P(A_i)$ converge. A quale valore converge?

2.9 Appendice: la probabilità soggettiva

Si consideri un evento E , la cui realizzazione è incerta. Un soggetto, Tu, è interessato a valutare la propria incertezza rispetto alla realizzazione di E anche al di fuori dell'ambito particolare degli esperimenti casuali.

Per rendere operativa la valutazione, ed esplicitare le proprie aspettative e informazioni sull'evento, Tu effettua un esperimento mentale. Immagina che gli sia offerto un contratto in cui, contro un prezzo, la controparte si obbliga a pagare 1 € se E si realizza. Allora per Tu la probabilità di E , $P(E)$, è il massimo prezzo che egli è disposto a pagare per acquistare il contratto descritto. Ad ogni prezzo più grande di $P(E)$ Tu sarà disposto a vendere il contratto, se è già in suo possesso.

Poiché Tu è razionale, non accetterà contratti che gli impongano una perdita certa. Quindi per Tu sarà vero che $0 \leq P(E) \leq 1$ con $P(S) = 1$. Inoltre se A e B sono incompatibili, per Tu sarà vero che $P(A \cup B) = P(A) + P(B)$. La probabilità soggettiva rispetta dunque essenzialmente gli stessi assiomi della probabilità frequentista.

La probabilità soggettiva ha limiti evidenti. Anzitutto, se non vi è un esperimento casuale a cui ancorare le opinioni, $P(E)$ varia da soggetto a soggetto, anche di molto. Inoltre, un soggetto può avere un forte pregiudizio o distorsione nel giudicare la probabilità di un evento. Va tuttavia riconosciuto che è meglio avere un'opinione, sia pure un po' azzardata, che nessuna idea su ciò che il futuro ci riserva. Un chiaro vantaggio delle probabilità soggettive è che sono ottenute senza un apparato matematico o sperimentale impegnativo.

2.10 Appendice: i vettori in R

In R i vettori possono essere assegnati in vari modi. Anzitutto grazie alle funzioni `c()` (concatenazione) e `scan()`:

```
> x=c(0,-1,1)
> x
[1] 0 -1 1
> y=scan()
1: 0
2: -1
3: 1
4:
Read 3 items
> y
[1] 0 -1 1
```

La scansione funziona anche immettendo in una posizione più valori numerici separati da spazi

```
> y=scan()
1: 0
2: -1 1
4:
Read 3 items
> y
[1] 0 -1 1
```

Vettori con elementi ripetuti possono essere ottenuti con la funzione `rep()`

```
> a=rep(1, times=7)
> a
[1] 1 1 1 1 1 1 1
> b=c(rep(1,times=3), rep(0, times=4))
> b
[1] 1 1 1 0 0 0 0
```

Vettori corrispondenti a successioni in progressione aritmetica possono ottenuti nei modi seguenti:

```
> c=seq(from=0.5, to=5.5, by=0.5)
> c
[1] 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5
> d=seq(from=0.5, to=5.5)
> d
[1] 0.5 1.5 2.5 3.5 4.5 5.5
> e=0.5:5.5
> e
[1] 0.5 1.5 2.5 3.5 4.5 5.5
```

Da un assegnato vettore si possono estrarre uno o più elementi. Si deve specificare tra parentesi quadre la posizione degli elementi da selezionare, o quella degli elementi da scartare (con segno -), o anche una condizione logica da soddisfare. Ad esempio:

```
> e[4]
[1] 3.5
> e[c(3,4)]
[1] 2.5 3.5
> e[1:3]
[1] 0.5 1.5 2.5
> e[-1]
[1] 1.5 2.5 3.5 4.5 5.5
> e[-c(1,2)]
[1] 2.5 3.5 4.5 5.5
> e[e>2]
[1] 2.5 3.5 4.5 5.5
```

Ai vettori si applicano, elemento per elemento, gli operatori aritmetici e relazionali, come pure le funzioni reali di variabile reale. Quindi

```
> 2*e
[1] 1 3 5 7 9 11
> 2*e+4*d
[1] 3 9 15 21 27 33
> e>2
[1] FALSE FALSE TRUE TRUE TRUE TRUE
> e^2
[1] 0.25 2.25 6.25 12.25 20.25 30.25
```

Sono funzioni importanti per analizzare vettori `length` `max` `min` `sum` `prod`. Ad esempio:

```
> f=e[e>2]
> length(f)
[1] 4
> max(f)
[1] 5.5
> min(f)
[1] 2.5
> sum(f)
[1] 16
> prod(f)
[1] 216.5625
```

Per avere informazioni su una funzione conviene digitare il comando `help()` con la particolare funzione come argomento. Potrebbe essere interessante `help(prod)`.

Per verificare numericamente la formula della somma di una serie geometrica (cfr. Esercizio 2.13) si può impostare un calcolo del tipo seguente:

22LEZIONE 2. PROBABILITÀ CLASSICA, FREQUENTISTA, ASSIOMATICA

```
> x=seq(-0.6,0.6,0.01)
> f8=1+x+x^2+x^3+x^4+x^5+x^6+x^7+x^8
> f12=1+x+x^2+x^3+x^4+x^5+x^6+x^7+x^8+x^9+x^10+x^11+x^12
> g=1/(1-x)
> max(abs(f8-g))
[1] 0.02519424
> max(abs(f12-g))
[1] 0.003265174
```

Per calcolare $n!$ (cfr. Esercizio 2.11) in R si dispone della funzione `factorial()`. Si possono tabulare i fattoriali dei primi 11 numeri naturali:

```
> n=0:10
> fn=factorial(n)
> A=rbind(n,fn)
> print(A)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]
n	0	1	2	3	4	5	6	7	8	9	10
fn	1	1	2	6	24	120	720	5040	40320	362880	3628800

A è una matrice con 2 righe e 12 colonne. Si provi l'effetto del comando `cbind(n,fn)`.

Alternativamente le matrici possono essere assegnate tramite la funzione `matrix`. Di *default* la matrice è popolata per colonne. Ad esempio

```
> m=0:7
> v=(-1)^m
> v
[1] 1 -1 1 -1 1 -1 1 -1
> B=matrix(v)
> B
```

	[,1]
[1,]	1
[2,]	-1
[3,]	1
[4,]	-1
[5,]	1
[6,]	-1
[7,]	1
[8,]	-1

```
> C=matrix(v,ncol=2)
> C
```

	[,1]	[,2]
[1,]	1	1
[2,]	-1	-1
[3,]	1	1
[4,]	-1	-1

Si provi l'effetto dei comandi `dim(A)` `ncol(A)` `nrow(A)` `C[,1]` `C[,2]` e `C[1,2]`.

Lezione 3

Teoremi elementari e richiami di calcolo combinatorio

3.1 I primi cinque teoremi elementari

Sia (S, \mathcal{F}, P) uno spazio probabilizzato e A, B, C eventi, ossia $A, B, C \in \mathcal{F}$. Per qualunque P che soddisfa gli assiomi di Kolmogorov **A1**, **A2**, **A3**, cfr. Definizione 2.4, (e non non solo per la probabilità classica per cui sono ovvi) valgono i seguenti teoremi.

Teorema 3.1 $P(\emptyset) = 0$.

Dimostrazione. Poiché $S = S \cup \emptyset$, con S e \emptyset eventi incompatibili ($S \cap \emptyset = \emptyset$), segue da **A3** che $P(S) = P(S) + P(\emptyset)$, dove $P(S) = 1$ per **A2**. \square

Teorema 3.2 Per ogni evento A , $P(\bar{A}) = 1 - P(A)$.

Dimostrazione. Poiché $S = A \cup \bar{A}$, con A e \bar{A} eventi incompatibili, segue da **A3** che $P(S) = P(A) + P(\bar{A})$, dove $P(S) = 1$ per **A2**. \square

Teorema 3.3 Per ogni evento A , $0 \leq P(A) \leq 1$.

Dimostrazione. Per il Teorema 3.2 si ha $P(A) = 1 - P(\bar{A})$. Per **A1** valgono sia $P(A) \geq 0$ sia $P(\bar{A}) \geq 0$. Da quest'ultima disuguaglianza si ottiene $P(A) \leq 1$. \square

Teorema 3.4 Se A e B sono eventi con $A \subseteq B$, allora $P(A) \leq P(B)$.

Dimostrazione. Poiché $S = A \cup \bar{A}$, con A e \bar{A} eventi incompatibili, segue da **A3** che $P(B) = P(A) + P(B \setminus A)$, dove $P(B \setminus A) \geq 0$ per **A1**. Quindi si ha $P(B) \geq P(A)$. \square

Corollario 3.1 Se A e B sono eventi con $A \subseteq B$, allora

$$P(B \setminus A) = P(B) - P(A).$$

Teorema 3.5 Se A e B sono eventi, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Dimostrazione. Si può rappresentare $A \cup B$ come unione di tre eventi mutuamente esclusivi:

$$A \cup B = \{A \setminus (A \cap B)\} \cup (A \cap B) \cup \{B \setminus (A \cap B)\}.$$

Segue da **A3** che

$$P(A \cup B) = P(A \setminus (A \cap B)) + P(A \cap B) + P(B \setminus (A \cap B)),$$

da dove applicando il Corollario 3.1 segue la tesi. \square

Sono facili conseguenze del Teorema 3.5 due diseuguaglianze:

la diseuguaglianza di Boole: $P(A \cup B) \leq P(A) + P(B)$

la diseuguaglianza di Bonferroni: $P(A \cap B) \geq P(A) + P(B) - 1$.

Quest'ultima si dimostra partendo da $P(A \cap B) = P(A) + P(B) - P(A \cup B)$.

Sono generalizzazioni del Teorema 3.5 le **formule di Poincaré**, di cui il primo esempio è

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

dove A , B e C sono eventi. Queste formule sono utili perché si possono applicare a eventi non mutuamente esclusivi.

3.2 Partizioni di S : due teoremi

Definizione 3.1 Una successione di eventi A_i , $i \in I \subseteq \mathbb{N}$, è detta **partizione** dello spazio campionario S se

i) è costituita da eventi mutuamente esclusivi:

$$i, j \in I, i \neq j \implies A_i \cap A_j = \emptyset$$

ii) è costituita da eventi necessari:

$$\cup_{i \in I} A_i = S \text{ (necessariamente almeno uno degli } A_i \text{ si realizza).}$$

La partizione di S minimale (minimamente dettagliata) è costituita dai soli due eventi S e \emptyset . La partizione massimale (massimamente dettagliata) è costituita dagli eventi elementari $\{s\}$, $s \in S$ (se S è finito o numerabile). Partizioni intermedie sono A , \overline{A} , ($A \in \mathcal{F}$) come pure, se $A, B \in \mathcal{F}$ con $A \cap B = \emptyset$, A , B , $\overline{A \cup B}$. Data una partizione, è certo che esattamente uno dei suoi elementi si realizzerà.

Valgono i seguenti risultati che coinvolgono le partizioni di S .

Teorema 3.6 Se $A_i, i \in I \subseteq \mathbb{N}$, è una partizione di S , allora $\sum_{i \in I} P(A_i) = 1$.

Dimostrazione. Applicando prima **A2** e poi **A3**, si ha $1 = P(S) = P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$. \square

Teorema 3.7 (Formula di addizione) Se $A_i, i \in I \subseteq \mathbb{N}$, è una partizione di S , allora per ogni $E \in \mathcal{F}$ vale

$$P(E) = \sum_{i \in I} P(E \cap A_i).$$

Dimostrazione. La tesi segue da $P(E) = P(E \cap S) = P(E \cap (\cup_{i \in I} A_i)) = P(\cup_{i \in I} (E \cap A_i)) = \sum_{i \in I} P(E \cap A_i)$ applicando il terzo assioma. Si noti che l'assioma è applicabile poiché $A_i \cap A_j = \emptyset$ dà $(E \cap A_i) \cap (E \cap A_j) = \emptyset$. In altri termini, se $A_i, i \in I$ è una partizione di S , allora $E \cap A_i, i \in I$, è una partizione di E . \square

Esempio 3.1 Si lancia un dado e successivamente si lancia una moneta tante volte quanto è il punteggio indicato dal dado. Si calcoli la probabilità di vedere solo teste. Sia E l'evento 'solo teste'. Gli eventi

$$A_i = \text{'il dado mostra il punteggio } i\text{'},$$

$i = 1, \dots, 6$, sono una partizione di S (sono a due a due incompatibili e la loro unione è certa). Sono anche equiprobabili, per cui $P(A_i) = \frac{1}{6}, i = 1, \dots, 6$. Fissato i , gli eventi elementari che appartengono ad A_i sono in numero di 2^i e si possono supporre equiprobabili. Ora $E \cap A_i$ è un evento elementare. Quindi $P(E \cap A_i) = \frac{1}{6} \frac{1}{2^i}$. Si ha dunque

$$P(E) = \sum_{i=1}^6 P(E \cap A_i) = \sum_{i=1}^6 \frac{1}{6} \frac{1}{2^i} = \frac{1}{6} \frac{1}{2} \frac{1 - (1/2)^6}{1 - (1/2)} = \frac{21}{128} = 0.1640625 \doteq 0.164,$$

dove \doteq indica una approssimazione numerica. Si rinvia all'Esercizio 1.18 per la valutazione di $\sum_{i=0}^5 (1/2)^i$. \diamond

3.3 La probabilità classica e il calcolo combinatorio

Come si è visto, quando $|S|$ è finita, si pone $\mathcal{F} = \mathcal{P}(S)$. Se si assume che gli eventi elementari siano equiprobabili, i calcoli della probabilità classica sono tutti della forma

$$P(E) = \frac{|E|}{|S|},$$

con $E \in \mathcal{F}$. Quindi calcolare probabilità è in effetti computare cardinalità di insiemi finiti, ossia fare calcoli combinatori.

3.3.1 Principi del calcolo combinatorio

Siano A e B insiemi di cardinalità finita. Conviene avere ben presenti le seguenti due regole:

1. **regola di addizione**

$$A \cap B = \emptyset \implies |A \cup B| = |A| + |B|$$

2. **regola di moltiplicazione**

$$|A \times B| = |A| |B|.$$

Esempio 3.2 Se una *password* consiste di 8 lettere, tratte da un alfabeto di 26, si calcoli la probabilità che in una *password* scelta a caso nessuna lettera sia ripetuta. Si indichi con A l'alfabeto, $A = \{a, b, c, \dots, x, y, z\}$, con $|A|=26$. Lo spazio campionario è costituito da tutte le possibili *password* di 8 lettere (ottuple ordinate),

$$S = A \times A \times \dots \times A = A^8.$$

L'evento di interesse è $E = \text{'password senza ripetizione di lettera'}$. Interpretiamo la scelta 'a caso' come un'asserzione che gli eventi elementari sono equiprobabili. Siamo nel campo della probabilità classica, per cui

$$P(E) = \frac{|E|}{|S|}.$$

Iterando la regola di moltiplicazione si vede che

$$|S| = |A|^8 = 26^8 = 208\,827\,064\,576 \text{ (quasi 209 miliardi).}$$

Solo un po' più complesso è il calcolo di $|E|$. Se una *password* fosse costituita da due lettere, le *password* senza ripetizione di scelta sarebbero $26^2 - 26 = 26 * 25$. Analogamente con tre lettere sarebbero $26 * 25 * 24$ (dopo aver fatto due scelte, restano 24 possibilità, diverse dalle prime due, per la terza scelta). Si ha quindi

$$|E| = 26 * 25 * 24 * 23 * 22 * 21 * 20 * 19 = 62\,990\,928\,000 \text{ (quasi 63 miliardi),}$$

per cui $P(E) = 0.301641591$, ossia circa il 30%. \diamond

Esempio 3.3 (*Il problema dei compleanni*) Vi sono n persone in una stanza. Si chiede qual è la probabilità che almeno due di esse festeggino il compleanno nello stesso giorno (evento E_n). Con l'assunzione di equiprobabilità delle successioni dei giorni di nascita (e anni di 365 giorni) si ha

$$P(E_n) = 1 - P(\overline{E}_n) = 1 - \frac{365 * 364 * \dots * (365 - n + 1)}{365^n}.$$

Si ha $P(E_{24}) \doteq 0.538$ e $P(E_{60}) \doteq 0.994$. Il minimo n per cui $P(E_n) > 0.5$ è $n = 23$. \diamond

3.3.2 Terminologia del calcolo combinatorio

In quanti modi si possono ‘schierare’ k simboli tratti da un alfabeto di n simboli? La risposta dipende dalla regola di schieramento. Si possono considerare quattro regole di schieramento fondamentali, a seconda che l’ordine di schieramento conti o non conti e che la ripetizione della scelta del simbolo sia ammessa o non ammessa.

Le successioni di k simboli tratti da un alfabeto di n (successioni non insiemi: quindi l’ordine conta) sono dette nel calcolo combinatorio **disposizioni con ripetizione** di ordine n, k . Il loro numero è indicato con $D'_{n,k}$. Argomentando come nell’Esempio 3.2, si ha che

$$D'_{n,k} = n^k.$$

Si noti che $D'_{n,k} = |\{f : U \rightarrow A\}|$, dove $|U| = k$ e $|A| = n$. Gli elementi di U possono essere pensati come ubicazioni o posti dove vengono collocati i simboli presi dall’alfabeto A . Ogni ubicazione riceve esattamente un simbolo, e ogni simbolo è disponibile in almeno k copie.

Le successioni di k simboli diversi (senza ripetizione di simbolo) tratti da un alfabeto di $n \geq k$ simboli (ancora una volta, successioni non insiemi: quindi l’ordine conta) sono dette nel calcolo combinatorio **disposizioni senza ripetizione o semplici** di ordine n, k . Il loro numero è indicato con $D_{n,k}$. Ragionando come nell’Esempio 3.2, si ottiene che

$$D_{n,k} = n(n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!} = \binom{n}{k} k!.$$

Si noti che $D_{n,k} = |\{f : U \rightarrow A : f \text{ iniettiva}\}|$, dove $|U| = k$ e $|A| = n$. Gli elementi di U possono essere pensati ancora come le ubicazioni dove si collocano i simboli. Si ricorda che una applicazione iniettiva associa ad elementi distinti del dominio immagini distinte nel codominio. Quindi ogni ubicazione riceve esattamente un simbolo, ma ora ogni simbolo è disponibile in una sola copia.

Come caso particolare notevole delle disposizioni senza ripetizione, si considerino le successioni di n simboli diversi (quindi senza ripetizione di simbolo) tratti da un alfabeto di n simboli. Esse sono dette **permutazioni di ordine n** . Il loro numero è indicato con P_n . Poiché $P_n = D_{n,n}$ e $0!=1$ si ha

$$P_n = n!.$$

Si noti che $P_n = |\{f : A \rightarrow A : f \text{ iniettiva}\}|$, dove $|A| = n$. Conviene pensare alle permutazioni degli elementi di A come all’insieme di tutti i possibili ordinamenti degli elementi di A .

I sottoinsiemi di k simboli diversi tratti da un alfabeto di n (insiemi non successioni: quindi l’ordine di schieramento non conta) sono detti **combinazioni senza ripetizione o semplici** di k oggetti da n . Il loro numero è indicato con $C_{n,k}$. Facendo agire su ogni combinazione di k oggetti da n tutte le permutazioni possibili di k oggetti, si ottengono tutte le disposizioni senza ripetizione di

ordine n, k . Dunque $C_{n,k} P_k = D_{n,k}$ e si ottiene

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1) \cdots (n-k+1)}{k(k-1) \cdots 1} = \binom{n}{k}.$$

I multinsiemi di k simboli non necessariamente diversi tratti da un alfabeto di n (insiemi non successioni: quindi l'ordine di schieramento non conta; multinsiemi perché ogni elemento è accompagnato dalla sua molteplicità) sono detti **combinazioni con ripetizione** di k oggetti da n . Il loro numero è indicato con $C'_{n,k}$ e risulta

$$C'_{n,k} = \binom{n+k-1}{k}.$$

Infatti la possibilità di ripetizione equivale ad avere nell'alfabeto ulteriori $k-1$ simboli *jolly* da schierare, ciascuno da identificare con una ripetizione di scelta.

La Tabella 3.1 riassume la situazione (salvo che per le permutazioni, caso particolare delle disposizioni semplici).

Tabella 3.1: Numero di 'schieramenti' di k simboli tratti da un alfabeto di n simboli.

LA RIPETIZIONE E' AMMESSA?				
	SI'		NO	
L'				
O				
R				
D	SI'	n^k	$n(n-1) \cdots (n-k+1)$	DISPOSIZIONI
I				
N				
E				
C				
O	NO	$\binom{n+k-1}{k}$	$\binom{n}{k}$	COMBINAZIONI
N				
T				
A				
?				
	CON RIPETIZIONE		SEMPLICI	

3.3.3 Esempi

Consideriamo due esempi di quesiti la cui soluzione è facilitata dai richiami di calcolo combinatorio appena esposti.

Esempio 3.4 Un lotto contiene 80 *chip* conformi e 20 difettosi. Interessa calcolare la probabilità che, estraendone a caso (in blocco) 10, risultino tutti conformi (evento E). Occorre anzitutto determinare S e la sua cardinalità. Per l'estrazione in blocco, l'ordine non conta e la ripetizione non è ammessa. Quindi

$$|S| = \binom{100}{10} \quad |E| = \binom{80}{10}$$

per cui

$$P(E) = \frac{|E|}{|S|} = \frac{\binom{80}{10}}{\binom{100}{10}} = \frac{80 * 79 * \dots * 71}{100 * 99 * \dots * 91} \doteq 0.09512.$$

◇

Esempio 3.5 Si stabilisca se, in una mano di 5 carte estratte da un mazzo di carte francesi, è più probabile che si realizzi $A_1 = \text{'un asso'}$ (uno esattamente!) o $F_2 = \text{'due carte di fiori'}$. Assumendo che le $\binom{52}{5} = 2\,598\,960$ mani estraibili dal mazzo di 52 carte siano equiprobabili, si calcola anzitutto

$$|A_1| = \binom{4}{1} \binom{52-4}{4} = 778\,320,$$

dove $\binom{4}{1} = 4$ è il numero di possibili scelte dell'asso e $\binom{48}{4} = \frac{48*47*46*45}{4!} = 194\,580$ è il numero delle possibili scelte delle quattro carte diverse dall'asso; si applica poi la regola di moltiplicazione. Analogamente, si ha

$$|F_2| = \binom{13}{2} \binom{52-13}{3} = \frac{13*12}{2!} \frac{39*38*37}{3!} = 712\,842.$$

In conclusione,

$$P(A_1) = \frac{778\,320}{2\,598\,960} \doteq 0.29947 \quad P(F_2) = \frac{712\,842}{2\,598\,960} \doteq 0.27428$$

per cui è più facile vedere realizzato A_1 piuttosto che F_2 .

◇

3.4 Le probabilità ipergeometriche

Un lotto contiene N pezzi, di cui D difettosi e $N - D$ conformi. Se si estraggono in blocco n pezzi, ci chiediamo quanti difettosi si possono manifestare un tale campione e con quale probabilità si manifestano.

Per evitare banalità, sia $0 < D < N$ e $0 < n < N$. Si consideri l'evento

$$E_x = \text{'si presentano } x \text{ difettosi nel campione'}.$$

Affinché E_x non sia impossibile, occorre che siano estraibili x difettosi e $n - x$ conformi, ossia che $x \leq \min(n, D)$ e che $n - x \leq \min(n, N - D)$. L'ultima

relazione è equivalente a $x \geq \max(0, n - N + D)$. Quindi i valori $x \in \mathbb{N}$ osservabili nel campione sono quelli che soddisfano

$$\max(0, n - N + D) \leq x \leq \min(n, D). \quad (3.1)$$

I corrispondenti eventi E_x formano una partizione di S e la probabilità di ciascuno di loro è

$$P(E_x) = \frac{|E_x|}{|S|} = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}}. \quad (3.2)$$

Infatti, l'estrazione in blocco produce sottoinsiemi del lotto, per cui la valutazione di $|S|$ è immediata. La valutazione di $|E_x|$ poggia sulla regola di moltiplicazione. Le probabilità (3.2) sono dette **probabilità ipergeometriche**.

Come si è visto nel paragrafo 2.7.5, valori numerici determinati da un esperimento casuale sono detti **variabili casuali** (v.c.) e indicate con simboli quali X . Se l'esperimento fa sì che X realizzi x , si scrive $X \rightarrow x$.

Una v.c. X con possibili realizzazioni nell'insieme $S_X = \{x \in \mathbb{N} : \max(0, n - N + D) \leq x \leq \min(n, D)\}$, come (3.1), tale che ogni realizzazione ha probabilità $P(X = x) = P(E_x)$ come in (3.2), è detta con **legge ipergeometrica** con indice n e parametri D e N . Si scrive in breve $X \sim IG(n; D, N)$. Se $X \sim IG(n; D, N)$ si ha $S_X = \{0, 1, \dots, n\}$ se e solo se $n \leq \min(D, N - D)$.

3.5 Esercizi

Esercizio 3.1 Dopo calcoli di un qualche impegno, per un certo evento E si ottiene $P(E) = -0.25$. Che cosa si deduce? E se per un certo evento F si ottiene $P(F) = 3.14$?

Esercizio 3.2 Se, per $\epsilon \in (0, 1)$, un evento E ha $P(E) \leq \epsilon$, allora $P(\bar{E}) \geq 1 - \epsilon$.

Esercizio 3.3 Si dimostri che se A_i , $i \in I \subseteq \mathbb{N}$, sono eventi, allora

$$P(\cup_{i \in I} A_i) \leq \sum_{i \in I} P(A_i),$$

formula che generalizza la disuguaglianza di Boole.

Esercizio 3.4 Siano A, B eventi con $P(A) = 0.5$, $P(B) = 0.7$, $A \subseteq B$. Si dica quali valori può assumere $P(A \cap B)$.

Esercizio 3.5 Siano A, B, C, D eventi. Si scriva la formula di Poincaré per $P(A \cup B \cup C \cup D)$.

Esercizio 3.6 Da un mazzo di carte trevisane, si estrae a caso una carta. Si calcoli la probabilità che sia o un re o una carta di denari.

Esercizio 3.7 Allo stadio di San Siro in occasione di una certa partita in tribuna nove persone su dieci sono tifose del Milan e sempre nove persone su dieci guadagnano più di 50 000 € l'anno. Cosa si può dedurre sulla probabilità che una persona scelta a caso tra quelle in tribuna sia tifosa del Milan e guadagni almeno 50 000 € l'anno?

Esercizio 3.8 Allo stadio di San Siro in occasione di una certa partita in tribuna due persone su dieci sono tifose del Milan e una persona su dieci guadagna meno di 50 000 € l'anno. Cosa si può dedurre sulla probabilità che una persona scelta a caso tra quelle in tribuna sia o una persona tifosa del Milan oppure guadagni meno di 50 000 € l'anno?

Esercizio 3.9 Un caso particolare interessante delle disposizioni con ripetizione si ha quando l'alfabeto A contiene due soli simboli (0 e 1) e l'insieme delle ubicazioni U ha k elementi. Si mostri che $D'_{2,k} = 2^k = |\mathcal{P}(U)|$.

Esercizio 3.10 Si calcoli il numero di targhe distinte formate dalla successione due lettere tre cifre due lettere (le cifre vanno da 0 a 9, le lettere sono 26).

Esercizio 3.11 Si mostri che $D_{n,k} < D'_{n,k}$ se $1 < k \leq n$.

Esercizio 3.12 Si calcoli in quanti modi distinti si possono disporre su uno scaffale 7 libri diversi.

Esercizio 3.13 Si valuti in quanti modi distinti dieci persone attive in politica possono essere scelte come ministri/e di cinque ministeri diversi.

Esercizio 3.14 Si considerino 5 esami senza propedeuticità. Si classificano gli studenti e le studentesse che li hanno superati tutti a seconda della sequenza di registrazione dei cinque esami. Quante sono le classi osservabili?

Esercizio 3.15 Si considerino 5 esami senza propedeuticità. Si classificano gli studenti e le studentesse che li hanno in piano di studi a seconda di quali di essi hanno superato. Quante sono le classi osservabili?

Esercizio 3.16 Si considerino 6 esami senza propedeuticità. Si classificano gli studenti e le studentesse che ne hanno superati esattamente 3 a seconda di quali di essi hanno superato. Quante sono le classi osservabili?

Esercizio 3.17 Un'urna contiene 20 palline nere e 80 palline bianche. Si estraggono in blocco 5 palline. Si calcoli la probabilità di vedere nel campione estratto 3 palline nere (e 2 bianche).

Esercizio 3.18 Si lanciano $2n$ monete. Fatta l'opportuna assunzione, si calcoli la probabilità che il numero delle teste sia uguale a quello delle croci.

3.6 Appendice: tabulare probabilità con R

Con riferimento all'Esempio 3.3, si desidera tabulare $P(E_n)$ per n da 21 a 25.

```
> n=1:30                                # QUESTO E' UN COMMENTO
> pn=rep(NA, times=length(n))          # NA "not available"
> for (i in n) {
+   pn[i]=1-prod(365:(365-i+1))/365^i
+ }
> R=rbind(n[21:25],pn[21:25])
> R
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	21.0000000	22.0000000	23.0000000	24.0000000	25.0000000
[2,]	0.4436883	0.4756953	0.5072972	0.5383443	0.5686997

In un'urna vi si sono 100 palline, di cui 60 sono nere. Si estraggono in blocco 10 palline. Si desidera tabulare le probabilità di osservare nel campione x palline nere, $x = 0, \dots, 10$. Le probabilità richieste sono ipergeometriche, cfr. (3.2). Il coefficiente binomiale $\binom{n}{k}$ si calcola con il comando `choose(n,k)`. Si ha

```
> x=0:10
> px=choose(60,x)*choose(40,10-x)/choose(100,10)
> R=rbind(x,px)
> R[,1:5]
```

	[,1]	[,2]	[,3]	[,4]	[,5]
x	0.000000e+00	1.000000e+00	2.000000e+00	3.000000e+00	4.000000e+00
px	4.896854e-05	0.0009477781	0.007863597	0.03685565	0.1081280

```
> R[,6:10]
```

	[,1]	[,2]	[,3]	[,4]	[,5]
x	5.000000e+00	6.000000e+00	7.000000e+00	8.000000e+00	9.000000e+00
px	0.2076057	0.2643128	0.2204307	0.1152911	0.03416032

Si vede che $x = 6$ ha la probabilità più grande. Per ottenere un grafico di queste probabilità ipergeometriche si usi il comando `plot(x,px,type="h")`. Che risultato dà `sum(px)`?

Lezione 4

Condizionamento, indipendenza, leggi binomiali

4.1 Coefficienti binomiali e teorema del binomio: richiami

Per completare i richiami di calcolo combinatorio, si fissi l'attenzione sui valori che esprimono il numero di combinazioni semplici. Dal significato combinatorio (numero di sottoinsiemi) è ovvio che

$$\binom{n}{0} = \binom{n}{n} = 1 \quad \text{per ogni } n \in \mathbb{N} \quad (4.1)$$

ed è immediato vedere che, per $k \in \{0, 1, \dots, n\}$, vale la relazione di simmetria

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

Inoltre, la somma delle probabilità ipergeometriche dà 1 (gli eventi E_x formano una partizione di S), e quindi

$$\binom{N}{n} = \sum_{x=\max(0, n-N+D)}^{\min(n, D)} \binom{D}{x} \binom{N-D}{n-x}.$$

Quando $D = 1$ e $N > n \geq 1$, x può assumere solo i valori 0 e 1 e si ottiene

$$\binom{N}{n} = \binom{1}{0} \binom{N-1}{n} + \binom{1}{1} \binom{N-1}{n-1}$$

ossia

$$\binom{N}{n} = \binom{N-1}{n} + \binom{N-1}{n-1}.$$

Ritornando alla notazione con n e k , si ha dunque la ricorrenza

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}, \quad (4.2)$$

valida per $n \geq k \geq 1$. Le relazioni (4.1) e (4.2) sono utili per calcolare $\binom{n}{k}$ per valori piccoli di n . L'algoritmo produce il famoso triangolo aritmetico, detto anche di Tartaglia o di Pascal (o di altri), cfr. Tabella 4.1.

Tabella 4.1: *Il triangolo aritmetico. I coefficienti binomiali esterni valgono 1, quelli interni sono somma dei due coefficienti superiori adiacenti.*

$$\begin{array}{ccccccc}
 & & & & \binom{0}{0} & & \\
 & & & & & & \\
 & & & \binom{1}{0} & & \binom{1}{1} & \\
 & & & & & & \\
 & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} \\
 & & & & & & \\
 & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & \binom{3}{3} \\
 & & & & & & \\
 \binom{4}{0} & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & \binom{4}{4} \\
 & & & & & & \\
 & & \dots & \dots & \dots & & \\
 & & & 1 & & & \\
 & & & & & & \\
 & & 1 & & 1 & & \\
 & & & & & & \\
 & 1 & & 2 & & 1 & \\
 & & & & & & \\
 & & 1 & & 3 & & 3 & & 1 \\
 & & & & & & & & \\
 1 & & 4 & & 6 & & 4 & & 1 \\
 & & & & & & & & \\
 & & \dots & \dots & \dots & & & &
 \end{array}$$

Un valore $\binom{n}{k}$ è anche detto **coefficiente binomiale**. Il nome è dovuto al seguente teorema.

Teorema 4.1 (teorema del binomio, di Newton) Per ogni $a, b \in \mathbb{R}$ e $n \in \mathbb{N}$ vale

$$(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}. \quad (4.3)$$

Ad esempio, $(a + b)^5 = b^5 + 5b^4a + 10b^3a^2 + 10b^2a^3 + 5ba^4 + a^5$.

4.2 Le probabilità condizionali

Un evento A è detto **trascurabile** se $P(A) = 0$, **non trascurabile** se $P(A) > 0$. Non ci aspettiamo proprio di veder realizzato un predesignato evento trascurabile, nemmeno se l'esperimento venisse ripetuto un grandissimo numero di volte. Un evento non trascurabile invece sarà prima o poi realizzato. Considerato un altro evento E , è naturale cercare di quantificare quanto è facile vedere E realizzato, sotto la condizione che l'evento non trascurabile A sia realizzato.

4.2.1 Definizione

Sia A un evento non trascurabile, detto condizione, E un altro evento.

Definizione 4.1 Si dice **probabilità di E condizionale ad A** , e si indica con $P(E|A)$, il valore

$$P(E | A) = \frac{P(E \cap A)}{P(A)}. \quad (4.4)$$

Nota: non conviene memorizzare la definizione con nomi specifici degli eventi, meglio *probabilità dell'intersezione fratto probabilità della condizione*.

Se A è realizzato, o pensato realizzato, la probabilità di E pertinente alla nuova situazione (reale o ipotizzata) non è più $P(E)$ ma $P(E | A)$. Infatti il condizionamento fa passare dall'esperimento \mathcal{E} all'esperimento $\mathcal{E}|_A$ che ha le stesse regole di \mathcal{E} ma in cui si 'squalificano' tutte le repliche che non realizzano A . Per l'esperimento $\mathcal{E}|_A$ l'evento A assume il ruolo di evento certo. In tale situazione, E si realizza se e solo se si realizza un s in $E \cap A$. La definizione (4.4) rinormalizza $P(E \cap A)$ semplicemente riproporzionandola a $P(A)$, di modo che sia $P(A | A) = 1$, come pure, per $E \supseteq A$, anche $P(E | A) = 1$.

Dal punto di vista frequentista, $P(E | A)$ relativa all'esperimento $\mathcal{E}|_A$, fatte R repliche di \mathcal{E} , è approssimata da

$$\frac{n^0 \text{ realizzazioni in } E \cap A}{n^0 \text{ realizzazioni in } A} = \frac{(n^0 \text{ realizzazioni in } E \cap A)/R}{(n^0 \text{ realizzazioni in } A)/R} \doteq \frac{P(E \cap A)}{P(A)}.$$

Esempio 4.1 Cinque informatici, $\alpha, \beta, \gamma, \delta, \epsilon$, sono in gara per aggiudicarsi un certo contratto. Si considerino gli eventi $A = \text{'vince } \alpha\text{'}$, $B = \text{'vince } \beta\text{'}$, $C = \text{'vince } \gamma\text{'}$, $D = \text{'vince } \delta\text{'}$, $E = \text{'vince } \epsilon\text{'}$, che formano una partizione

36LEZIONE 4. CONDIZIONAMENTO, INDIPENDENZA, LEGGI BINOMIALI

dello spazio campionario. Sulla base dell'esperienza passata, le probabilità di aggiudicazione del contratto sono

$$P(A) = 0.45, \quad P(B) = 0.25, \quad P(C) = 0.15, \quad P(D) = 0.15, \quad P(E) = 0.$$

Se A viene assunto da un'azienda californiana e si ritira dalla gara, le 'nuove' probabilità, pertinenti alla nuova situazione, sono

$$\begin{aligned} P(A | \bar{A}) &= \frac{P(A \cap \bar{A})}{P(\bar{A})} = \frac{P(\emptyset)}{0.55} = \frac{0}{55} = 0 \\ P(B | \bar{A}) &= \frac{P(B \cap \bar{A})}{P(\bar{A})} = \frac{P(B)}{0.55} = \frac{25}{55} = 0.4545455 \\ P(C | \bar{A}) &= \frac{P(C \cap \bar{A})}{P(\bar{A})} = \frac{P(C)}{0.55} = \frac{15}{55} = 0.2727273 \\ P(D | \bar{A}) &= \frac{P(D \cap \bar{A})}{P(\bar{A})} = \frac{P(D)}{0.55} = \frac{15}{55} = 0.2727273 \\ P(E | \bar{A}) &= \frac{P(E \cap \bar{A})}{P(\bar{A})} = \frac{P(E)}{0.55} = \frac{0}{55} = 0. \end{aligned}$$

◇

Il teorema seguente mostra che la probabilità condizionale ad A è anch'essa una misura di probabilità su (S, \mathcal{F}) . Valgono dunque per le probabilità condizionali gli stessi teoremi che si sono enunciati per le probabilità non condizionali. Ad esempio $P(\bar{E} | A) = 1 - P(E | A)$, $P(E \cup F | A) = P(E | A) + P(F | A) - P(E \cap F | A)$, eccetera.

Teorema 4.2 *Sia (S, \mathcal{F}, P) uno spazio probabilizzato e A un assegnato evento non trascurabile. Allora $P(E | A)$, al variare di $E \in \mathcal{F}$, è una misura di probabilità per lo spazio probabilizzabile (S, \mathcal{F}) .*

Dimostrazione. Occorre verificare che i tre assiomi di Kolmogorov sono soddisfatti. La non-negatività di $P(E | A)$ è ovvia dalla definizione (4.4). La normalizzazione segue da $P(S|A) = 1$. La verifica della σ -additività richiede, per una successione finita o numerabile di eventi $E_i \in \mathcal{F}$ a due a due incompatibili, il calcolo

$$\begin{aligned} P(\cup_{i \in I} E_i | A) &= \frac{P((\cup_{i \in I} E_i) \cap A)}{P(A)} = \frac{P(\cup_{i \in I} (E_i \cap A))}{P(A)} = \frac{\sum_{i \in I} P(E_i \cap A)}{P(A)} \\ &= \sum_{i \in I} \frac{P(E_i \cap A)}{P(A)} = \sum_{i \in I} P(E_i | A), \end{aligned}$$

dove si sfrutta il fatto che se gli eventi E_i sono a due a due incompatibili anche gli eventi $E_i \cap A$ sono a due a due incompatibili. □

4.2.2 Formula della probabilità composta

La formula della probabilità composta e le sue generalizzazioni consentono di calcolare probabilità di eventi intersezione tramite opportune probabilità condizionali.

Teorema 4.3 (Formula della probabilità composta) *Se E è un evento e A è un evento non trascurabile, allora*

$$P(E \cap A) = P(A) P(E | A). \quad (4.5)$$

In molte applicazioni, è più immediato calcolare $P(A)$ e $P(E | A)$ invece che direttamente $P(E \cap A)$.

Esempio 4.2 Si estraggono in sequenza due carte da un mazzo di carte trevisane (senza rimettere nel mazzo la prima estratta). Sia E_i l'evento 'la i -esima carta è punti', $i = 1, 2$. Sono punti le carte fante, cavallo, re, tre, asso dei quattro semi. Per calcolare la probabilità che entrambe le carte siano punti, ossia $P(E_1 \cap E_2)$, conviene considerare che $P(E_1) = 20/40 = 1/2$ e $P(E_2 | E_1) = 19/39$, per cui

$$P(E_1 \cap E_2) = P(E_1) P(E_2 | E_1) = \frac{1}{2} \frac{19}{39} = \frac{19}{78} = 0.2435897.$$

◇

Con tre eventi, A, B, C tali che $A \cap B$ sia non trascurabile, la formula della probabilità composta si generalizza nel modo seguente

$$P(A \cap B \cap C) = P(A) P(B | A) P(C | A \cap B).$$

La formula è ben definita perché $A \supseteq A \cap B$ e quindi anche A è non trascurabile. Si lascia al lettore di considerare la formula per $P(\cap_{i=1}^n A_i)$, dove gli A_i sono eventi e $\cap_{i=1}^{n-1} A_i$ è non trascurabile.

4.3 Formula della probabilità totale

La formula della probabilità totale consente di calcolare la probabilità (non condizionale) di un evento come media pesata di probabilità dell'evento stesso condizionali agli elementi di una partizione.

Teorema 4.4 (Formula della probabilità totale) *Se E è un evento e A_i , $i \in I \subseteq \mathbb{N}$, è una partizione dello spazio campionario in eventi non trascurabili, allora*

$$P(E) = \sum_{i \in I} P(A_i) P(E | A_i). \quad (4.6)$$

Dimostrazione. La tesi segue da

$$P(E) = \sum_{i \in I} P(E \cap A_i) = \sum_{i \in I} P(A_i) P(E | A_i).$$

La prima eguaglianza è il Teorema 3.7 (formula di addizione), l'ultima usa la formula della probabilità composta, (4.5). \square

Esempio 4.3 Si considera l'esperimento di estrazione senza reinserimento di due palline da un'urna che contiene N palline nere e B palline bianche. L'estrazione è sequenziale: vi è una prima pallina estratta e una seconda pallina estratta. Si definiscono gli eventi $B_i = \text{'bianca alla } i\text{-esima estrazione'}$ e $N_i = \text{'nera alla } i\text{-esima estrazione'}$, $i = 1, 2$. L'evento di cui interessa calcolare la probabilità è N_2 , nera alla seconda estrazione. Si calcola facilmente $P(N_2)$ considerando che gli eventi B_1 e N_1 formano una partizione dello spazio campionario, con probabilità rispettive

$$P(B_1) = \frac{B}{N+B}, \quad P(N_1) = \frac{N}{N+B}.$$

Seguendo le modificazioni della composizione dell'urna, si vede facilmente che, condizionatamente ai due casi, le probabilità di N_2 sono

$$P(N_2 | B_1) = \frac{N}{N+B-1}, \quad P(N_2 | N_1) = \frac{N-1}{N+B-1}.$$

In conclusione,

$$\begin{aligned} P(N_2) &= P(B_1) P(N_2 | B_1) + P(N_1) P(N_2 | N_1) \\ &= \frac{B}{N+B} \frac{N}{N+B-1} + \frac{N}{N+B} \frac{N-1}{N+B-1} \\ &= \frac{N(B+N-1)}{(N+B)(N+B-1)} = \frac{N}{B+N}. \end{aligned}$$

\diamond

4.4 Eventi indipendenti

Sia A un evento non trascurabile e E un altro evento. Possono darsi tre casi:

$$i) P(E | A) > P(E) \quad ii) P(E | A) < P(E) \quad iii) P(E | A) = P(E).$$

Nei primi due casi, la condizione che A sia realizzato facilita, o rende più difficile, la realizzazione di E . L'ultimo caso, una situazione di indifferenza del condizionamento, è molto interessante. Chiaramente,

$$P(E | A) = \frac{P(E \cap A)}{P(A)} = P(E) \implies P(E \cap A) = P(E)P(A)$$

e dunque, se pure E è non trascurabile, vale anche

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

Due eventi indifferenti al reciproco condizionamento sono detti indipendenti. Per essere svincolati da assunzioni di non trascurabilità, si pone la seguente definizione.

Definizione 4.2 (*Indipendenza di due eventi*) A e B in \mathcal{F} sono detti **eventi indipendenti** se $P(A \cap B) = P(A)P(B)$.

Si noti che l'evento impossibile è indipendente da qualunque evento. Anche ogni evento trascurabile è indipendente da qualunque evento.

Definizione 4.3 (*Indipendenza degli eventi di una successione*) A_i , $i \in I$ con $|I| > 2$, è una **successione di eventi indipendenti** se per ogni $J \subseteq I$, J di cardinalità finita e pari almeno a 2, vale

$$P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

In altri termini, la probabilità di realizzazione simultanea di un numero finito di eventi distinti della successione si calcola sempre come prodotto delle probabilità individuali degli eventi considerati.

Assumere l'indipendenza di certi eventi è un potente strumento di modellazione probabilistica degli esperimenti casuali. Si considerino ad esempio due esperimenti,

- \mathcal{E}_1 descritto dallo spazio probabilizzato $(S_1, \mathcal{F}_1, P_1)$
- \mathcal{E}_2 descritto dallo spazio probabilizzato $(S_2, \mathcal{F}_2, P_2)$.

Se i due esperimenti non hanno possibilità di interagire l'uno con l'altro, gli eventi relativi al primo esperimento sono da ritenere indipendenti dagli eventi relativi al secondo. L'esperimento complessivo $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2)$ avrà spazio campionario $S = S_1 \times S_2$. La classe degli eventi \mathcal{F} da associare a S sarà la più piccola σ -algebra che contiene tutte le parti di S di forma $\{E \subseteq S : E = A \times B, A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$. La misura di probabilità P che dà la descrizione (S, \mathcal{F}, P) di \mathcal{E} attribuirà ad ogni $E \in \mathcal{F}$ di forma $A \times B$, dove $A \in \mathcal{F}_1$ e $B \in \mathcal{F}_2$, probabilità

$$P(E) = P_1(A)P_2(B).$$

4.5 Una applicazione dell'indipendenza: le probabilità binomiali

Per l'esperimento casuale $\mathcal{E} = \text{'si lancia una moneta } n \text{ volte'}$ un evento elementare $s \in S$ è ad esempio $s = (T, T, C, \dots, T) \in \{T, C\}^n$. Si vede immediatamente che $|S| = 2^n$. Gli eventi

$$E_{n,x} = \text{'}x \text{ teste negli } n \text{ lanci'}, \quad x = 0, 1, \dots, n,$$

formano una partizione di S . Si calcola subito che $|E_{n,x}| = \binom{n}{x}$. Se gli eventi elementari sono equiprobabili,

$$P(E_{n,x}) = \frac{|E_{n,x}|}{|S|} = \binom{n}{x} \frac{1}{2^n}.$$

Supporre che gli eventi elementari siano equiprobabili equivale ad assumere:

1. che la moneta sia equilibrata: nel singolo lancio, $P(T) = P(C) = 0.5$;
2. che vi sia indipendenza degli esiti (testa/croce) dei vari lanci distinti.

Per verificare l'affermazione, conviene introdurre gli eventi

$$T_i = \text{'testa all'}i\text{'-esimo lancio'}$$

e

$$C_i = \text{'croce all'}i\text{'-esimo lancio'}.$$

Si ha allora, ad esempio con $n = 4$, che un particolare evento elementare è $\{s\} = T_1 \cap T_2 \cap C_3 \cap T_4$ e, per l'indipendenza,

$$P(T_1 \cap T_2 \cap C_3 \cap T_4) = P(T_1)P(T_2)P(C_3)P(T_4) = \frac{1}{2^4},$$

(l'ultima eguaglianza deriva dall'equilibrio della moneta) e così per tutti gli altri eventi elementari.

Se la moneta può essere truccata, con $P(T_i) = p$ e $P(C_i) = 1-p$, $i = 1, \dots, n$, $p \in (0, 1)$, e si continua ad assumere l'indipendenza dei lanci, gli eventi elementari non sono più equiprobabili. Hanno in effetti probabilità che in generale dipendono dal numero di teste e croci che li definiscono. Indicando con $T(s)$ il numero di teste in s e con $C(s)$ il numero di croci in s , si avrà per l'indipendenza

$$P(\{s\}) = p^{T(s)}(1-p)^{C(s)}.$$

Tuttavia gli eventi elementari la cui unione dà $E_{n,x}$ hanno tutti lo stesso numero di teste, x , e quindi la stessa probabilità, $p^x(1-p)^{n-x}$. Tali eventi elementari, a due a due incompatibili, sono in numero di $\binom{n}{x}$. Quindi nell'esperimento di n lanci indipendenti di una moneta truccata, con probabilità p di testa in ogni singolo lancio, si ottiene

$$P(E_{n,x}) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n. \quad (4.7)$$

Le probabilità $\binom{n}{x}p^x(1-p)^{n-x}$ ($x = 0, 1, \dots, n$) sono dette **probabilità binomiali**. Infatti la loro normalizzazione si controlla subito grazie al teorema del binomio:

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

La variabile casuale X con possibili realizzazioni in

$$S_X = \{x \in \mathbb{N} : 0 \leq x \leq n\},$$

tale che ogni realizzazione ha probabilità $P(X = x) = P(E_{n,x})$ come in (4.7), viene detta con **legge binomiale** con parametri n e p . Si scrive in breve $X \sim Bi(n, p)$, dove $n \in \mathbb{N}^+$ e $0 < p < 1$.

Generalizzando l'esempio dei lanci indipendenti di una moneta, si vede che le probabilità binomiali sono descrittive della probabilità di ottenere x successi in n prove indipendenti con costante probabilità p ($p \in (0, 1)$) di successo nella singola prova, $x = 0, 1, \dots, n$. La Tabella 4.2 mostra le probabilità binomiali come funzione di p per valori piccoli di n .

Tabella 4.2: *Le probabilità binomiali per valori piccoli di n .*

		n^0 successi, x					
		0	1	2	3	4	5
n^0	1	$1 - p$	p				
p	2	$(1 - p)^2$	$2p(1 - p)$	p^2			
r							
o	3	$(1 - p)^3$	$3p(1 - p)^2$	$3p^2(1 - p)$	p^3		
v							
e	4	$(1 - p)^4$	$4p(1 - p)^3$	$6p^2(1 - p)^2$	$3p^3(1 - p)$	p^4	
n	5	$(1 - p)^5$	$5p(1 - p)^4$	$10p^2(1 - p)^3$	$10p^3(1 - p)^2$	$5p^4(1 - p)$	p^5

Esempio 4.4 Si supponga che, considerando un gran numero di parti, i nati siano per il 51.5% maschi e per il 48.5% femmine. In un punto nascita sono nati nelle ultime 24 ore 10 bambini. Si desidera calcolare la probabilità che vi siano fra i 10 nati più maschi che femmine come pure la probabilità che vi siano più femmine che maschi. Sia X il numero casuale di maschi in 10 nascite. Per l'indipendenza della attribuzione del sesso ai vari nati e il fatto che la probabilità di osservare un maschio è costante in ogni nascita e pari a 0.515, la legge di probabilità di X è $Bi(10, 0.515)$. Dunque il 'pareggio' ha probabilità

$$P(X = 5) = P(E_{10,5}) = \binom{10}{5} 0.515^5 0.485^5 \doteq 0.24499$$

mentre la prima probabilità richiesta è

$$\begin{aligned} P(X \geq 6) &= P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\ &= \sum_{x=6}^{10} \binom{10}{x} 0.515^x 0.485^{10-x} \doteq 0.41438. \end{aligned}$$

Chiaramente, $P(X \leq 4) = 1 - P(X = 5) - P(X \geq 6) \doteq 0.34063$. \diamond

4.6 Esercizi

Esercizio 4.1 Si mostri che se $k < n$ vale $\binom{2n}{k} < \binom{2n}{n}$, per cui per n pari $\binom{n}{k}$ è massimo per $k = n/2$. Cosa accade quando n è dispari?

Esercizio 4.2 Si mostri che $\sum_{i=0}^n \binom{n}{i} = 2^n$.

Esercizio 4.3 Si mostri che, per $n \geq 1$, $\binom{2n}{n} < 2^{2n}$.

Esercizio 4.4 Si calcoli $\sum_{i=0}^n (-1)^i \binom{n}{i}$.

Esercizio 4.5 Da un mazzo di carte trevisane si estraggono due carte, lasciandole coperte. Si scopre la prima carta, è un asso (evento A_1). Si calcoli la probabilità che la seconda carta sia un asso. È una probabilità condizionale?

Esercizio 4.6 Si lanciano in sequenza due dadi, il primo rosso e il secondo verde. Si calcoli la probabilità che la somma dei punteggi dei due dadi sia 12
a) senza condizioni ii) sotto la condizione che il dado rosso dia 4 iii) sotto la condizione che il dado rosso dia 6.

Esercizio 4.7 Siano A e B due eventi con $P(A) = 0.8$, $P(B) = 0.7$, $P(A \cap B) = 0.3$. Si calcoli $P(A \cap B \mid A \cup B)$.

Esercizio 4.8 Siano A , B , C e D eventi e $A \cap B \cap C$ sia non trascurabile. Si scriva la formula della probabilità composta per $P(A \cap B \cap C \cap D)$.

Esercizio 4.9 Un evento E è indipendente da sé stesso se

$$P(E) = P(E \cap E) = P(E)P(E) = P(E)^2.$$

Si dica quali sono gli eventi indipendenti da sé stessi.

Esercizio 4.10 Si mostri che, se A e B sono una coppia di eventi indipendenti, allora anche A e \bar{B} , \bar{A} e B , \bar{A} e \bar{B} sono coppie di eventi indipendenti.

Esercizio 4.11 Si lancia un dado equilibrato. Il punteggio ottenuto dal dado determina il numero di lanci indipendenti di una moneta equa: se il punteggio è pari, la moneta è lanciata due volte; se è dispari, la moneta è lanciata tre volte. Si calcoli la probabilità di ottenere due teste.

4.7. APPENDICE: PROBABILITÀ BINOMIALI E IPERGEOMETRICHE CON R43

Esercizio 4.12 Si lanciano un dado e una moneta. Se gli eventi elementari sono equiprobabili, gli eventi $A = \text{'testa dalla moneta'}$ e $B = \text{'punteggio} \geq 3 \text{ dal dado}'$ sono indipendenti?

Esercizio 4.13 Siano A e B due eventi incompatibili. Se sono non trascurabili, sono anche indipendenti?

Esercizio 4.14 Sia A un evento e B un evento quasi certo, ossia tale che $P(B) = 1$. Si mostri che $P(A \cap B) = P(A)$.

Esercizio 4.15 Si considerino le famiglie con 5 figli. Se la probabilità di maschio alla nascita è 0.51 e le attribuzioni del sesso sono indipendenti, è più probabile incontrare una famiglia con una sola femmina o con un solo maschio?

Esercizio 4.16 Sia A un evento che ha probabilità 0.5. Condizionatamente alla realizzazione di A , un conteggio ha legge $Bi(3, 0.5)$, quindi $X | A \sim Bi(3, 0.5)$. Se A non si realizza, il conteggio ha legge $Bi(2, 0.5)$, quindi $X | \bar{A} \sim Bi(2, 0.5)$. Se non si sa se A è realizzato o no, si stabilisca quali sono i possibili valori di X e con quale probabilità si presentano.

Esercizio 4.17 Si riconsiderino i quesiti dell'Esempio 4.4 supponendo che: a) i nati siano per il 51% maschi e per il 49% femmine; b) i nati siano per il 50% maschi e per il 50% femmine.

4.7 Appendice: probabilità binomiali e ipergeometriche con R

Si consideri un lotto contenente $N = 100$ pezzi, di cui $D = 5$ sono difettosi. Dal lotto è estratto casualmente un campione di $n = 10$ pezzi. Interessa descrivere il numero aleatorio X di pezzi difettosi nel campione. La legge di X dipenderà da come avvengono le estrazioni. Se sono con reinserimento, si ha $X \sim Bi(n, p)$ dove $n = 10$ e $p = D/N = 0.05$. Se invece le estrazioni sono in blocco o senza reinserimento, si ha $X \sim IG(n; D, N)$, quindi $X \sim IG(10, 5, 100)$. Per calcolare la probabilità binomiale $P(X = x)$ si usa il comando `dbinom(x, size=n, prob=p)` o anche semplicemente `dbinom(x,n,p)`. Per calcolare la probabilità ipergeometrica $P(X = x)$ si usa invece il comando `dhyper(x,D,N-D,n)`. Le variabili `x,n,p,D,N` vanno inizializzate. Si ha

```
> x=0:10
> pbin=dbinom(x,size=10,prob=0.05)
> phyp=dhyper(x,5,95,10)
> x[1:7]
[1] 0 1 2 3 4 5 6
> round(pbin[1:7],digits=6)
[1] 0.598737 0.315125 0.074635 0.010475 0.000965 0.000061 0.000003
> round(phyp[1:7],digits=6)
[1] 0.583752 0.339391 0.070219 0.006384 0.000251 0.000003 0.000000
```

44LEZIONE 4. CONDIZIONAMENTO, INDIPENDENZA, LEGGI BINOMIALI

per cui, ad esempio, $P(X = 6)$ è valutata 3×10^{-6} con la legge binomiale (estrazioni con reinserimento) mentre è valutata 0 con la legge ipergeometrica (estrazioni senza reinserimento). Le probabilità binomiali approssimano bene le probabilità ipergeometriche, più delicate da calcolare numericamente, se $n/N \leq 0.1$, ossia quando, come in questo caso, la numerosità del campione è una frazione non maggiore del 10% della numerosità della popolazione.

R dà anche la possibilità di simulare l'estrazione di elementi da un lotto. Ad esempio, con le codifiche "d" 'difettoso' e "c" 'conforme', estrazioni con reinserimento (`replace=TRUE`) e probabilità di difettoso 0.05 e di conforme 0.95 si ha

```
> set.seed(12345)
> y = sample(c("d","c"), 10, replace=TRUE, prob=c(0.05,0.95))
> y
[1] "c" "c" "c" "c" "c" "c" "c" "c" "c" "d"
```

Il comando `set.seed(12345)` rende replicabile il risultato della simulazione. Il lettore è invitato a generare altri cinque campioni con la stessa numerosità di y.

Ecco due esempi di estrazione in blocco di elementi da un lotto.

```
> set.seed(12345)
> items=c(rep("d",5),rep("c",5))
> sample(items,4,replace=FALSE)
[1] "c" "c" "c" "c"
> sample(items,4,replace=FALSE)
[1] "d" "d" "d" "d"
```

L'esito delle due estrazioni è sospetto. Conviene controllare che, in molte prove, le frequenze assolute osservate di X difettosi siano vicine alle frequenze attese ipergeometriche (probabilità per numero delle prove). Il risultato è rassicurante.

```
> set.seed(12345)
> nrep=10^4
> x=rep(NA,nrep)
> for (ir in 1:nrep){
+   campione=sample(items,4,replace=FALSE)
+   x[ir]=length(campione[campione=="d"])
+ }
> table(x)
x
 0    1    2    3    4      # numero pezzi difettosi
232 2352 4745 2416 255      # freq. assolute osservate
> print(nrep*round(dhyper(xx,5,5,4), digits=4))
[1] 238 2381 4762 2381 238      # freq. attese ipergeometriche
```

Lezione 5

Il teorema di Bayes

5.1 Un problema introduttivo

Si consideri un canale di comunicazione binaria, in cui il segnale può essere 0 o 1. Vi è del rumore nel canale: 0 trasmesso può a volte essere ricevuto 1, e 1 trasmesso può a volte essere ricevuto 0. Per un dato canale, si sa che

- 0 viene ricevuto correttamente con probabilità 0.95
- 1 viene ricevuto correttamente con probabilità 0.80
- il 60% dei segnali trasmessi è 0.

Quesito: se un segnale ricevuto è 0, qual è la probabilità che sia stato anche trasmesso 0?

Il segnale trasmesso è, assieme al rumore, la ‘causa’ del segnale ricevuto. Il segnale ricevuto è osservato, e ci si chiede qual è la probabilità di una causa visto un effetto. Si definiscano gli eventi $T_i = \text{‘trasmesso } i\text{’}$ e $R_i = \text{‘ricevuto } i\text{’}$, $i = 1, 2$. Evidentemente, la coppia di eventi T_0 e T_1 è una partizione di S e i dati del problema permettono le seguenti valutazioni:

$$\begin{aligned} P(T_0) &= 0.6 & P(T_1) &= 0.4 \\ P(R_0 | T_0) &= 0.95 & P(R_1 | T_1) &= 0.8. \end{aligned}$$

L’incognita richiesta è $P(T_0 | R_0)$. Si ha

$$\begin{aligned} P(T_0 | R_0) &= \frac{P(T_0 \cap R_0)}{P(R_0)} = \frac{P(T_0)P(R_0 | T_0)}{P(T_0)P(R_0 | T_0) + P(T_1)P(R_0 | T_1)} \\ &= \frac{0.6 * 0.95}{0.6 * 0.95 + 0.4 * (1 - 0.8)} = \frac{0.57}{0.65} = 0.876923. \end{aligned}$$

Si è usata la relazione $P(R_0 | T_1) = 1 - P(R_1 | T_1)$, che sorge dal fatto che anche R_0, R_1 è una partizione di S . Si ha subito per la probabilità dell’altra causa è $P(T_1 | R_0) = 1 - P(T_0 | R_0) = 0.123077$.

Con facili calcoli si può ottenere anche $P(T_1 | R_1)$:

$$P(T_1 | R_1) = \frac{P(T_1 \cap R_1)}{P(R_1)} = \frac{P(T_1)P(R_1 | T_1)}{1 - P(R_0)} = \frac{0.4 * 0.8}{1 - 0.65} = 0.9142857.$$

5.2 Il teorema: enunciato, dimostrazione, commenti

Sia E un ‘effetto’ osservato, la cui manifestazione possa essere attribuita a ‘cause’ o alternative A_i , mutuamente esclusive e necessarie, ossia che formano una partizione di S . Prima dell’osservazione di E , *a priori*, gli eventi A_i hanno probabilità $P(A_i)$ di essere la causa attiva. Queste sono dette **probabilità iniziali** o *a priori*. Descrivono in effetti l’incertezza sulla causa prima che sia constatato l’effetto. Osservato E , *a posteriori*, le probabilità pertinenti diventano invece le $P(A_i | E)$. Queste sono chiamate **probabilità aggiornate** o *a posteriori*. La formula di Bayes, da Thomas Bayes (1702–1761), detta anche formula della probabilità delle cause, permette di calcolare $P(A_i | E)$ noti i valori $P(A_i)$ e $P(E | A_i)$, $i \in I$.

Teorema 5.1 (Formula di Bayes) *Sia A_i , $i \in I \subseteq \mathbb{N}$, una partizione di S in eventi non trascurabili, e sia E un evento non trascurabile. Allora per ogni $i \in I$*

$$P(A_i | E) = \frac{P(A_i) P(E | A_i)}{\sum_{j \in I} P(A_j) P(E | A_j)}. \quad (5.1)$$

Dimostrazione. Per definizione, $P(A_i | E) = P(A_i \cap E)/P(E)$. Per la formula della probabilità composta, Teorema 4.3, $P(A_i \cap E) = P(A_i) P(E | A_i)$. Per la formula della probabilità totale, Teorema 4.4, $P(E) = \sum_{j \in I} P(A_j) P(E | A_j)$. □

Oltre alle probabilità iniziali e alle probabilità aggiornate, compaiono nella formula i valori $P(E | A_i)$, $i \in I$. Nel contesto della formula di Bayes E è supposto realizzato, per cui i valori $P(E | A_i)$ appaiono come delle probabilità solo per la notazione. Concettualmente non sono delle probabilità, come 0.5 non è la probabilità di ottenere testa da una moneta equilibrata in una realizzazione che ha manifestato testa. I valori $P(E | A_i)$, $i \in I$, sono detti **verosimiglianze** che l’osservazione di E attribuisce alle cause A_i . Rappresentano la propensione che l’esperimento aveva di produrre l’effetto osservato E sotto le varie possibili cause attive A_i , $i \in I$.

Una lettura molto semplice della formula di Bayes è

$$P(A_i | E) \propto P(A_i) \times P(E | A_i), \quad i \in I,$$

dove \propto indica **proporzionale a** e la costante di proporzionalità è $c = 1/P(E) = 1/(\sum_{j \in I} P(A_j) P(E | A_j))$. Quindi: le probabilità aggiornate sono proporzionali a probabilità iniziali per verosimiglianze.

La formula di Bayes è molto utile nella classificazione automatica. Si pensi a una situazione in cui ogni unità di osservazione appartiene a una sola classe fra più classi possibili, A_i , $i \in I$. Un paziente può essere sano o malato. Un debitore può essere un buon pagatore o un cattivo pagatore. Una *e-mail* può essere *spam* o *ham*. Un cliente può essere, rispetto all'acquisto di un prodotto, molto interessato, interessato, poco interessato. Il *data base* contiene molte storie già interamente osservate che permettono di valutare le $P(A_i)$. Per predire il comportamento non ancora osservato di una nuova unità spesso si possono acquisire dal *data base* ulteriori caratteristiche già osservate dell'unità, che permettono una classificazione più pertinente all'unità (profilazione). Ad esempio, per un cliente, sesso, professione, tipologia del contatto, tempo trascorso dall'ultimo acquisto, entità dell'ultimo acquisto. Indicando con E il complesso delle caratteristiche dell'unità già osservate, si è dunque interessati a valutare $P(A_i | E)$, che per la formula di Bayes richiede i valori $P(E | A_i)$, pure ottenibili dal *data base*, spesso tramite modelli statistici. Si noti che la classificazione non viene effettuata sulla base delle sole verosimiglianze $P(E | A_i)$, ma sulla base simultaneamente di probabilità iniziali e verosimiglianze. Una classificazione basata sulle verosimiglianze sarebbe attendibile solo se le probabilità iniziali fossero all'incirca costanti. E naturalmente una classificazione basata sulle probabilità iniziali sarebbe valida solo se E fosse indipendente dagli A_i .

5.3 Esempi di applicazione

Esempio 5.1 Un esame diagnostico per una certa malattia dà esito positivo, ossia segnala la presenza della malattia, nel 90% dei malati. Dà invece esito negativo, ossia non segnala la presenza della malattia, nel 99% dei sani. La prevalenza della malattia nella popolazione, cioè la frazione di malati, è del 2%. Si vuole indagare sulla probabilità di ottenere un falso positivo da quell'esame diagnostico. Un falso positivo è un paziente che ha esito positivo dell'esame ma non ha la malattia.

Le alternative d'interesse sono $M = \text{'paziente malato'}$ e \bar{M} e formano una partizione di S , con probabilità iniziali che risultano dai dati del problema

$$P(M) = 0.02, \quad P(\bar{M}) = 0.98.$$

Si suppone realizzato l'evento $E = \text{'esame positivo'}$, che determina le verosimiglianze

$$P(E | M) = 0.90, \quad P(E | \bar{M}) = 1 - P(\bar{E} | \bar{M}) = 1 - 0.99 = 0.01.$$

La probabilità richiesta è, per la formula di Bayes,

$$\begin{aligned} P(\bar{M} | E) &= \frac{P(\bar{M}) P(E | \bar{M})}{P(\bar{M}) P(E | \bar{M}) + P(M) P(E | M)} \\ &= \frac{0.98 * 0.01}{0.98 * 0.01 + 0.02 * 0.90} = 0.352518. \end{aligned}$$

La relativa rarità dei malati fa sì che anche un esito positivo al test non sia decisivo. Se la prevalenza della malattia fosse ancora più piccola, ad esempio del 2 per mille, si avrebbe il valore ancora meno decisivo $P(\bar{M} | E) = 0.8448276$.

◇

Esempio 5.2 In un certo ufficio il 60% dei documenti è scritto in Word, il 30% in HTML, il 10% in \LaTeX . Si sa che superano le 12 pagine il 50% dei documenti scritti in Word, il 10% dei documenti scritti in HTML, il 20% dei documenti scritti in \LaTeX . Si estrae a caso un documento e si constata che supera le 12 pagine. Con quale probabilità risulta scritto in \LaTeX ? Si considerino gli eventi

$$\begin{aligned} E &= \text{'il documento supera le 12 pagine'}, \\ W &= \text{'il documento risulta scritto in Word'}, \\ H &= \text{'il documento risulta scritto in HTML'}, \\ L &= \text{'il documento risulta scritto in \LaTeX'}. \end{aligned}$$

L'incognita richiesta è $P(L | E)$. I dati del problema sono:

- W, H, L formano una partizione dello spazio campionario
- le probabilità iniziali sono $P(W) = 0.6, P(H) = 0.3, P(L) = 0.1$
- le verosimiglianze sono $P(E | W) = 0.5, P(E | H) = 0.1, P(E | L) = 0.2$.

La probabilità richiesta è, per la formula di Bayes,

$$\begin{aligned} P(L | E) &= \frac{P(L) P(E | L)}{P(L) P(E | L) + P(W) P(E | W) + P(H) P(E | H)} \\ &= \frac{0.1 * 0.2}{0.1 * 0.2 + 0.6 * 0.5 + 0.3 * 0.1} = \frac{2}{35} = 0.05714286. \end{aligned}$$

In modo analogo si trova $P(W | E) = 30/35$ e $P(H | E) = 3/35$.

◇

Esempio 5.3 Un'urna contiene 10 palline nere e 90 bianche. Una seconda urna contiene 50 palline nere e 50 bianche. Una terza urna contiene 90 palline nere e 10 bianche. Uno sperimentatore sceglie a caso un'urna fra le tre con equiprobabilità, poi estrae a caso, con reinserimento, due palline dall'urna scelta. Si determini la probabilità che l'urna scelta sia stata quella con 50 nere, se le palline estratte risultano, senza tener conto dell'ordine di estrazione, una nera e una bianca.

Le alternative $A_i = \text{'scelta l'urna } i\text{'}$, $i = 1, 2, 3$, formano una partizione dello spazio campionario. Sono tre eventi equiprobabili, quindi le probabilità iniziali sono $P(A_i) = 1/3, i = 1, 2, 3$. Sia E l'evento osservato,

$$E = \text{'una pallina bianca e una nera'}$$

in due estrazioni con reinserimento, ossia due prove indipendenti con costante probabilità di successo data l'urna. Le verosimiglianze sono quindi determinate dalle probabilità binomiali

$$\begin{aligned} P(E | A_1) &= \binom{2}{1} \frac{10}{100} \frac{90}{100} = 2 \frac{900}{10000} = 0.18 \\ P(E | A_2) &= \binom{2}{1} \frac{50}{100} \frac{50}{100} = 2 \frac{2500}{10000} = 0.50 \\ P(E | A_3) &= \binom{2}{1} \frac{90}{100} \frac{10}{100} = 2 \frac{900}{10000} = 0.18. \end{aligned}$$

(Se le estrazioni fossero in blocco o senza reinserimento, le verosimiglianze sarebbero determinate da probabilità ipergeometriche). La probabilità richiesta è, per la formula di Bayes,

$$\begin{aligned} P(A_2 | E) &= \frac{P(A_2) P(E | A_2)}{P(A_1) P(E | A_1) + P(A_2) P(E | A_2) + P(A_3) P(E | A_3)} \\ &= \frac{\frac{1}{3} * 0.5}{\frac{1}{3} * (0.18 + 0.50 + 0.18)} = \frac{50}{86} = \frac{25}{43} = 0.5813953 \doteq 58\%. \end{aligned}$$

◇

5.4 Esercizi

Esercizio 5.1 Dagli archivi di un'azienda di credito al consumo risulta che il 95% dei clienti è un buon pagatore. Tra i buoni pagatori, il 75% è coniugato. Tra i cattivi pagatori, il 10% è coniugato. Si calcoli la probabilità che un nuovo cliente coniugato sia un buon pagatore.

Esercizio 5.2 Tre urne, all'apparenza indistinguibili, contengono 10 palline ciascuna. Si sa che due urne contengono 6 palline bianche e 4 nere e la terza contiene 1 pallina bianca e 9 nere. Viene scelta a caso un'urna e da essa si estraggono due palline senza reinserimento. Si determini la probabilità che le due palline siano nere. Nell'ipotesi che le due palline estratte siano nere, si ottenga la probabilità che si sia selezionata l'urna con 9 palline nere.

Esercizio 5.3 Un'azienda che produce monitor per PC possiede due linee di produzione. Il 20% dei monitor proviene dalla prima linea di produzione, il restante 80% dalla seconda. Per un controllo di qualità viene scelto a caso un monitor, senza sapere da quale linea provenga. Esso risulta difettoso. In base allo storico, risulta che la prima linea di produzione produce 8 difettosi su 1000 e la seconda 5 su 1000. Con quale probabilità il monitor difettoso proviene dalla prima linea di produzione?

Esercizio 5.4 La proporzione di elementi non conformi agli standard di qualità, prodotti da una certa azienda, è pari a 1 su 1000. L'azienda usa una

procedura di controllo della qualità con le seguenti caratteristiche: un elemento non conforme viene scartato con probabilità 0.8 e un elemento conforme viene scartato erroneamente con probabilità 0.01. Qual è la probabilità che un elemento scelto a caso venga scartato dal sistema? Se l'elemento viene scartato dal sistema, qual è la probabilità che sia effettivamente non conforme agli standard di qualità?

Esercizio 5.5 (*compito del 31/01/18*) Un'urna contiene 30 palline nere e 70 bianche. Una seconda urna contiene 50 palline nere e 50 bianche. Una terza urna contiene 70 palline nere e 30 bianche. Uno sperimentatore sceglie a caso un'urna fra le tre con equiprobabilità, poi estrae a caso, con reinserimento, otto palline dall'urna scelta. Si determini la probabilità che l'urna scelta sia stata quella con 50 nere, se le palline estratte risultano, senza tener conto dell'ordine di estrazione, quattro nere e quattro bianche.

Esercizio 5.6 (*compito del 12/02/18*) Un'urna contiene 5 palline nere e 95 bianche. Una seconda urna contiene 40 palline nere e 60 bianche. Una terza urna contiene 70 palline nere e 30 bianche. Uno sperimentatore sceglie a caso un'urna fra le tre con equiprobabilità, poi estrae a caso, con reinserimento, quattro palline dall'urna scelta. Si determini la probabilità che l'urna scelta sia stata quella con 5 nere, se le palline estratte risultano, senza tener conto dell'ordine di estrazione, una nera e tre bianche.

Esercizio 5.7 (*compito del 18/06/18*) Un'urna contiene 25 palline nere e 75 bianche. Una seconda urna contiene 50 palline nere e 50 bianche. Una terza urna contiene 75 palline nere e 25 bianche. Uno sperimentatore sceglie a caso un'urna fra le tre con equiprobabilità, poi estrae a caso, con reinserimento, quattro palline dall'urna scelta. Si determini la probabilità che l'urna scelta sia stata quella con 50 nere, se le palline estratte risultano, senza tener conto dell'ordine di estrazione, due nere e due bianche.

Esercizio 5.8 (*compito del 23/07/18*) Un'urna contiene 20 palline nere e 80 bianche. Una seconda urna contiene 50 palline nere e 50 bianche. Una terza urna contiene 60 palline nere e 40 bianche. Una quarta urna contiene 80 palline nere e 20 bianche. Uno sperimentatore sceglie a caso un'urna fra le quattro con equiprobabilità, poi estrae a caso, con reinserimento, sei palline dall'urna scelta. Si determini la probabilità che l'urna scelta sia stata quella con 80 nere, se le palline estratte risultano, senza tener conto dell'ordine di estrazione, cinque nere e una bianca.

Esercizio 5.9 (*compito del 11/09/18*) Un'urna contiene 10 palline nere e 90 bianche. Una seconda urna contiene 80 palline nere e 20 bianche. Una terza urna contiene 100 palline nere e 0 bianche. Uno sperimentatore sceglie a caso un'urna fra le tre con equiprobabilità, poi estrae a caso, con reinserimento, cinque palline dall'urna scelta. Si determini la probabilità che l'urna scelta sia stata quella con 10 nere, se le palline estratte risultano, senza tener conto dell'ordine di estrazione, una nera e quattro bianche.

5.5 Appendice: il teorema di Bayes con R

Si desidera affrontare con R l'esercizio 5.5. Si ha

```
> partizione=c("urna1","urna2","urna3")
> p_iniziali=rep(1/3, times=3)
> p_successo_data_urna=c(0.3,0.5,0.7)
> prove=8
> successi=4
> verosimiglianze=
+ dbinom(successi,size=prove,prob=p_successo_data_urna)
> p_aggiornate=
+ p_iniziali*verosimiglianze/sum(p_iniziali*verosimiglianze)
> p_aggiornate=round(p_aggiornate, digits=5)
> print(rbind(partizione,p_aggiornate))
      [,1]      [,2]      [,3]
partizione  "urna1"  "urna2"  "urna3"
p_aggiornate "0.24947" "0.50107" "0.24947"
```

per cui la probabilità richiesta è circa 0.50107.

Si supponga che che l'urna scelta, ignota allo sperimentatore, sia stata la 3 e che si eseguano da essa 100 estrazioni con reinserimento. Il numero di successi sarà una osservazione di $X \sim Bi(100, 0.6)$. R dà il valore simulato

```
> set.seed(12345)
> x=rbinom(1,100,0.7)
> x
[1] 74
```

(*r* in *rbinom* sta per *random*; *rbinom(rep,n,p)* restituisce un vettore di *rep* realizzazioni indipendenti di $X \sim Bi(n,p)$). Usando la formula di Bayes, lo sperimentatore ricostruisce la situazione nel modo seguente

```
> partizione=c("urna1","urna2","urna3")
> p_iniziali=rep(1/3, times=3)
> p_successo_data_urna=c(0.3,0.5,0.7)
> prove=100
> successi=74
> verosimiglianze=
+ dbinom(successi,size=prove,prob=p_successo_data_urna)
> p_aggiornate=
+ p_iniziali*verosimiglianze/sum(p_iniziali*verosimiglianze)
> p_aggiornate=round(p_aggiornate, digits=6)
> print(rbind(partizione,p_aggiornate))
      [,1]      [,2]      [,3]
partizione  "urna1" "urna2"  "urna3"
p_aggiornate "0"    "9e-06" "0.999991"
```

ossia è ragionevolmente convinto che l'urna scelta sia stata la terza.

Per fare altre sperimentazioni conviene definire una *function* di R che esprima il teorema di Bayes quando le verosimiglianze sono date da probabilità binomiali. La sintassi del comando **function** e della valutazione della funzione si desume dall'esempio

```
> somma_quadrati <- function(x,y){
+ x^2+y^2
+ }
> somma_quadrati(3,4)
[1] 25
```

Quindi, passando da 3 a un numero finito, **casipartizione**, di casi nella partizione, con associate le probabilità iniziali **p_iniziali**, e le probabilità di successo in una singola prova dato l'elemento della partizione, **p_successo_data_urna**, si definisce la *function*

```
> Bayes.verosim.binom <- function
+ (casipartizione,p_iniziali,p_successo_data_urna,prove,successi)
+ {
+ verosimiglianze=
+ dbinom(successi,size=prove,prob=p_successo_data_urna)
+ p_aggiornate=
+ p_iniziali*verosimiglianze/sum(p_iniziali*verosimiglianze)
+ # TEOREMA DI BAYES
+ p_aggiornate=round(p_aggiornate, digits=6)
+ casipartizione=1:casipartizione
+ print(rbind(casipartizione,p_aggiornate))
+ }
```

Ecco due esempi di chiamata.

```
> set.seed(12345)
> x=rbinom(1,400,0.6)
> x
[1] 229
> Bayes.verosim.binom(3,rep(1/3,times=3),c(0.4,0.5,0.6),400, 229)
      [,1]    [,2]    [,3]
casipartizione    1 2.00000 3.00000
p_aggiornate      0 0.02675 0.97325
> set.seed(12345)
> x=rbinom(1,1000,0.6)
> x
[1] 584
> Bayes.verosim.binom(3,rep(1/3,times=3),c(0.4,0.5,0.6),1000, 584)
      [,1]    [,2]    [,3]
casipartizione    1 2e+00 3.000000
p_aggiornate      0 1e-06 0.999999
```

Lezione 6

Le variabili casuali e la loro legge di probabilità

6.1 Premessa

Fino a questo punto si sono considerati esperimenti casuali descritti da spazi probabilizzati (S, \mathcal{F}, P) dove S ha natura qualsiasi. Quindi $\mathcal{E} \rightarrow s$, con s qualunque tipo di oggetto. Nelle applicazioni, e anche nella teoria matematica, è importante il caso in cui s è numerico. Se $s \in \mathbb{R}^d$ si dice che \mathcal{E} produce una realizzazione di una **variabile casuale** (v.c.). Se invece s è una più generale funzione numerica determinata dal caso, si dice che s è realizzazione di un **processo stocastico**. Lo studio delle v.c. è l'oggetto principale di questo insegnamento. Si sono già incontrati due esempi, le v.c. con legge ipergeometrica e binomiale. Conviene introdurre il concetto generale di v.c. anzitutto in modo non tecnico.

6.2 Visione euristica

Per acquisire una prima intuizione, si può pensare una variabile casuale come una successione finita di valori reali che saranno determinati dal caso. Ad esempio, il numero complessivo di teste in n lanci di una moneta è un singolo valore x determinato dal caso in \mathbb{R} . Se invece si considerano i punteggi di un primo e di un secondo dado, si ottiene un punto (x_1, x_2) in \mathbb{R}^2 determinato dal caso. Se si osservano i valori dei tempi di corretto funzionamento di d esemplari di un certo prodotto industriale, si otterrà un punto (x_1, x_2, \dots, x_d) determinato dal caso in \mathbb{R}^d . Quando $d = 1$, la v.c. è detta **univariata**, per $d = 2$ è detta **bivariata**, per $d > 2$ è detta **multivariata**.

Conviene anche considerare le v.c. come particolari digitalizzazioni del risultato materiale di un esperimento casuale \mathcal{E} , descritto dallo spazio probabilizzato

(S, \mathcal{F}, P) , secondo lo schema

$$\mathcal{E} \longrightarrow s \xrightarrow{X} X(s) \in \mathbb{R}^d$$

dove s si è un oggetto materiale, $X(s)$ la sua digitalizzazione, l'applicazione X la regola di digitalizzazione. Una v.c. sarà indicata con X , o all'occorrenza Y , Z o un'altra lettera maiuscola tra le ultime dell'alfabeto latino, come pure con X_1, X_2 eccetera.

6.3 Le v.c. come spazi probabilizzati

La realizzazione di una v.c. $X = (X_1, \dots, X_d)$ è un oggetto numerico $x = (x_1, \dots, x_d) \in \mathbb{R}^d$. Il valore x è prodotto dal caso non del tutto capricciosamente, ma secondo una 'regolarità di produzione' nel complesso di moltissime replicazioni, ossia secondo una particolare misura di probabilità con spazio campionario \mathbb{R}^d (o un suo sottoinsieme).

Gli eventi che si considerano in \mathbb{R}^d sono di base gli intervalli di \mathbb{R}^d

$$(a_1, b_1] \times (a_2, b_2] \times \dots \times (a_d, b_d]$$

dove $a_i \leq b_i$, $i = 1, \dots, d$. Ulteriori eventi si possono definire tramite complementazione, unione, intersezione di eventi dati (operazioni su eventi) fatte un numero finito o anche un'infinità numerabile di volte. La classe di tutti tali eventi è detta la σ -algebra di Borel su \mathbb{R}^d e indicata con \mathcal{B}_d .

Una v.c. X può essere pensata come una misura di probabilità P_X associata allo spazio probabilizzabile $(\mathbb{R}^d, \mathcal{B}_d)$, quindi in definitiva come un particolare spazio probabilizzato $(\mathbb{R}^d, \mathcal{B}_d, P_X)$. In concreto, questo significa che X produrrà, ogni volta che è osservato, un valore tipicamente diverso, ma la distribuzione di tali valori è descritta da P_X .

6.4 Le v.c. come applicazioni e la loro legge di probabilità

In termini matematici, la digitalizzazione del risultato materiale conduce a definire una v.c. come una particolare applicazione dallo spazio probabilizzato (S, \mathcal{F}, P) allo spazio probabilizzabile $(\mathbb{R}^d, \mathcal{B}_d)$. È questa la base di partenza per descrivere la nozione precisa di v.c.. Si danno le seguenti definizioni.

Definizione 6.1 Sia (S, \mathcal{F}, P) uno spazio probabilizzato. Una variabile casuale d -variata è una applicazione

$$X : S \rightarrow \mathbb{R}^p$$

misurabile, ossia tale che per ogni $B \in \mathcal{B}_d$ si ha $X^{-1}(B) \in \mathcal{F}$.

Definizione 6.2 *La legge di probabilità di una v.c. d -variata X è una applicazione $P_X : \mathcal{B}_d \rightarrow [0, 1]$ definita, per ogni $B \in \mathcal{B}_d$, da*

$$P_X(B) = P(X^{-1}(B)). \quad (6.1)$$

Osservazione 1: la misurabilità di X è richiesta affinché la legge di probabilità sia ben definita: le due definizioni sono ‘ad incastro’.

Osservazione 2: si può interpretare la misurabilità come una descrizione astratta della costruibilità pratica della applicazione, tramite approssimazioni sempre più precise.

Osservazione 3: spesso, invece di $P_X(B)$, si usa la scrittura più semplice e intuitiva $P(X \in B)$.

Osservazione 4: la regola $P_X(B) = P(X^{-1}(B))$ trasferisce la ‘vecchia’ misura di probabilità P , pertinente agli eventi ‘fisici’, agli eventi numerici $B \in \mathcal{B}_d$.

Teorema 6.1 *La legge di probabilità di una v.c. d -variata X è una misura di probabilità per lo spazio probabilizzabile $(\mathbb{R}^d, \mathcal{B}_d)$.*

Dimostrazione. Va verificato che P_X soddisfa i tre assiomi di Kolmogorov. La non-negatività è ovvia dalla (6.1). La normalizzazione segue da $P_X(\mathbb{R}^d) = P(X^{-1}(\mathbb{R}^d)) = P(S) = 1$. Verificare la σ -additività per una successione finita o numerabile di eventi $B_i \in \mathcal{B}_d$ a due a due incompatibili richiede il calcolo

$$\begin{aligned} P_X(\cup_{i \in I} B_i) &= P(X^{-1}(\cup_{i \in I} B_i)) = P(\cup_{i \in I} X^{-1}(B_i)) = \sum_{i \in I} P(X^{-1}(B_i)) \\ &= \sum_{i \in I} P_X(B_i), \end{aligned}$$

dove si sfrutta il fatto che se gli eventi $B_i \in \mathcal{B}_d$ sono a due a due incompatibili anche le loro immagini inverse $X^{-1}(B_i)$ sono eventi in \mathcal{F} a due a due incompatibili. \square

6.5 Classi di equivalenza di v.c.: l'identica distribuzione

Definizione 6.3 *Due v.c. X e Y si dicono **identicamente distribuite**, e si scrive $X \sim Y$, se $P_X = P_Y$, ossia per ogni $B \in \mathcal{B}_d$ vale $P_X(B) = P_Y(B)$.*

Nel Calcolo delle Probabilità le v.c. identicamente distribuite sono identificate. Quindi non importa come è definita una v.c. come applicazione, ma solo importa la sua legge di probabilità. Ciò è in pieno accordo con il considerare le v.c. spazi probabilizzati $(\mathbb{R}^d, \mathcal{B}_d, P_X)$, cfr. paragrafo 6.3.

6.6 Prime leggi notevoli di tipo discreto

Le leggi degeneri

Si dice che una v.c. d -variata X ha legge degenerare in $x_0 \in \mathbb{R}^d$, valore prefissato, e si scrive $X \sim \mathcal{D}(x_0)$, se per $B \in \mathcal{B}_d$ la legge di probabilità di X è

$$P_X(B) = P(X \in B) = \begin{cases} 1 & \text{se } x_0 \in B \\ 0 & \text{se } x_0 \notin B. \end{cases}$$

Poiché $P(X = x_0) = P_X(\{x_0\}) = 1$, se si realizza l'esperimento si otterrà senz'altro il punto x_0 . Quindi $X \sim \mathcal{D}(x_0) \implies X \rightarrow x_0$.

Sia $d = 2$, $X = (X_1, X_2) \sim \mathcal{D}((x_{01}, x_{02}))$. Allora $X_1 \sim \mathcal{D}(x_{01})$ e $X_2 \sim \mathcal{D}(x_{02})$. Più in generale, se $X \sim \mathcal{D}(x_0)$ e g è una funzione numerica definita in x_0 , si ha $g(X) \sim \mathcal{D}(g(x_0))$.

Le leggi binomiali

Si dice che la v.c. univariata X ha legge binomiale con indice $n \in \mathbb{N}^+$ e parametro $p \in (0, 1)$, e si scrive $X \sim Bi(n, p)$, se per ogni $B \in \mathcal{B}_1$ vale

$$P_X(B) = P(X \in B) = \sum_{x \in B \cap \{0, 1, \dots, n\}} \binom{n}{x} p^x (1-p)^{n-x}.$$

Quando $n = 1$, le leggi binomiali sono dette **di Bernoulli** o **binomiali elementari**. Un esempio semplice è $X \sim Bi(1, 0.5)$ per cui $P_X(\{0\}) = P_X(\{1\}) = 0.5$ ossia $P(X = 0) = P(X = 1) = 0.5$ e per tutti gli altri valori si ha $P(X \notin \{0, 1\}) = 0$, ossia non si osservano.

Le leggi ipergeometriche

Si dice che la v.c. univariata X ha legge ipergeometrica con indice $n \in \mathbb{N}^+$ e parametri N e D , dove $n \leq N \in \mathbb{N}^+$ e $D \in \mathbb{N}^+$ con $D \leq N$, e si scrive in breve $X \sim IG(n; D, N)$, se vale

$$P_X(\{x\}) = P(X = x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}}$$

per tutti i valori x per cui hanno senso i coefficienti binomiali indicati, e altrimenti $P_X(\{x\}) = P(X = x) = 0$. Usando il simbolo S_X per i valori x per cui $P(X = x) > 0$, le leggi P_X ipergeometriche sono tali che per ogni $B \in \mathcal{B}_1$

$$P_X(B) = P(X \in B) = \sum_{x \in B \cap S_X} \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}}.$$

6.7 Le leggi discrete in generale

Le tre precedenti famiglie di leggi di probabilità sono esempi particolari di leggi di probabilità di tipo discreto. Le leggi discrete sono in generale individuate da due ingredienti:

1. un insieme

$$S_X = \cup_{i \in I} \{x_i\}, \quad x_i \in \mathbb{R}^d, \quad i \in I \subseteq \mathbb{N},$$

successione di valori distinti senza punti di accumulazione al finito, detto **supporto** della v.c.

2. una applicazione

$$p_X : S_X \rightarrow [0, 1]$$

detta **funzione di massa di probabilità** (f.m.p.) che soddisfa le condizioni

- $p_X(x) > 0$ per ogni $x \in S_X$ (positività sul supporto)
- $\sum_{x \in S_X} p_X(x) = 1$ (normalizzazione).

Dato supporto e f.m.p., la legge discreta corrispondente è definita da

$$P_X(B) = P(X \in B) = \sum_{x \in B \cap S_X} p_X(x).$$

È facile infatti verificare che P_X così definita soddisfa i tre assiomi di Kolmogorov.

6.8 Rappresentazioni in tabella di leggi discrete

Con pochi punti nel supporto e quindi poche masse di probabilità da specificare, si possono descrivere le leggi discrete univariate in forma di tabella. Ad esempio una v.c. univariata X può essere rappresentata come

valori	0	1	2	Totale
probabilità	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1

o anche, più succintamente, come

Dalla tabella si ricava che il supporto è $S_X = \{0, 1, 2\}$ e che la f.m.p. è

$$p_X(0) = \frac{1}{2}, \quad p_X(1) = \frac{1}{4}, \quad p_X(2) = \frac{1}{4}.$$

Rappresentiamo in forma di tabella $X \sim Bi(3, 0.5)$. Si ha $S_X = \{0, 1, 2, 3\}$ e la probabilità binomiale dà $p_X(x) = \binom{3}{x}(\frac{1}{2})^x(1 - \frac{1}{2})^{3-x} = \binom{3}{x}(\frac{1}{2})^3$. Si ottiene

x	0	1	2	Totale
$p_X(x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1

x	0	1	2	3	Totale
$p_X(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

6.9 Le leggi uniformi discrete

La v.c. d -variata X ha legge uniforme discreta in $D = \cup_{i=1}^k \{x_i\}$, $x_i \in \mathbb{R}^d$, dove gli x_i sono k punti distinti, e si scrive in breve $X \sim Ud(x_1, \dots, x_k)$, se

- $S_X = D$
- $p_X(x) = \frac{1}{k}$ per ogni $x \in S_X$.

Un esempio univariato con $k = 3$ è $X \sim Ud(0, 1, 2)$ che ha la rappresentazione in tabella

x	0	1	2	Totale
$p_X(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Si vede facilmente che $Ud(0, 1) \sim Bi(1, 0.5)$. Infatti la rappresentazione in tabella è, per entrambe,

6.10 Valore atteso e varianza di v.c. univariate con legge discreta

Sia X una v.c. univariata con legge discreta e supporto finito. Si possono cogliere aspetti importanti della distribuzione di X attraverso indici sintetici. I principali sono gli indici sintetici

- di posizione: il **valore atteso**

$$E(X) = \sum_{x \in S_X} xp_X(x)$$

x	0	1	Totale
$p_X(x)$	$\frac{1}{2}$	$\frac{1}{2}$	1

- di variabilità: la **varianza**

$$\text{Var}(X) = \sum_{x \in S_X} (x - E(X))^2 p_X(x) = E((X - E(X))^2).$$

Se la legge di X è discreta ma il supporto ha la cardinalità del numerabile, $E(X)$ e $\text{Var}(X)$ si definiscono come somma di serie. Viene richiesta però la convergenza assoluta della serie. Quindi

$$E(X) = \sum_{x \in S_X} x p_X(x) \quad \text{purché} \quad \sum_{x \in S_X} |x| p_X(x) < +\infty.$$

Il valore atteso è semplicemente la media aritmetica ponderata dei valori assumibili dalla v.c., con pesi dati dalle masse di probabilità. Il valore $E(X)$ è indicativo di un centro della distribuzione dei potenziali valori assumibili da X .

La varianza è la media aritmetica ponderata del quadrato degli scarti di X dal proprio valore atteso, $(X - E(X))^2$. Più grande è la varianza, più si ha indicazione che i valori $(x - E(X))^2$ per $x \in S_X$ sono tipicamente grandi.

Esempio 6.1 Sia $X \sim \mathcal{D}(x_0)$ con $x_0 \in \mathbb{R}$. Si ha $E(X) = x_0$ e $\text{Var}(X) = 0$. \diamond

Esempio 6.2 Sia $X \sim Bi(1, p)$ con $p \in (0, 1)$. Si ha $S_X = \{0, 1\}$ e $p_X(x) = p^x(1-p)^{1-x}$ per cui $p_X(0) = 1-p$ e $p_X(1) = p$. Quindi

$$E(X) = 0 p_X(0) + 1 p_X(1) = p$$

e

$$\begin{aligned} \text{Var}(X) &= (0-p)^2 p_X(0) + (1-p)^2 p_X(1) = p^2(1-p) + (1-p)^2 p \\ &= p(1-p)(p+1-p) = p(1-p). \end{aligned}$$

Per p prossimi a 0 o 1 la varianza di $X \sim Bi(1, p)$ è prossima a 0; il suo valore massimo, pari a 1/4, si ha quando $p = 1/2$. Questa è la situazione di massima incertezza sull'esito di X . \diamond

Esempio 6.3 Sia $X \sim Bi(2, p)$ con $p \in (0, 1)$. Si ha $S_X = \{0, 1, 2\}$ e $p_X(x) = \binom{2}{x} p^x (1-p)^{2-x}$ per cui $p_X(0) = (1-p)^2$, $p_X(1) = 2p(1-p)$, $p_X(2) = p^2$. Quindi

$$\begin{aligned} E(X) &= 0 p_X(0) + 1 p_X(1) + 2 p_X(2) = 2p(1-p) + 2p^2 = 2p(1-p+p) = 2p, \\ \text{Var}(X) &= (0-2p)^2 p_X(0) + (1-2p)^2 p_X(1) + (2-2p)^2 p_X(2) \\ &= 4p^2(1-p)^2 + (1-2p)^2 2p(1-p) + 4(1-p)^2 p^2 \\ &= 2p(1-p) \{2p(1-p) + (1-2p)^2 + 2p(1-p)\} = 2p(1-p). \end{aligned}$$

Per valori di p prossimi a 0 o 1 la varianza di $X \sim Bi(2, p)$ è prossima a 0; il suo valore massimo, pari a $1/2$, si ha quando $p = 1/2$. Questa è la situazione di massima incertezza sull'esito di X . \diamond

Generalizzando il risultato degli Esempi 6.2 e 6.3, si dimostrerà nell'Esempio 16.3 che per $X \sim Bi(n, p)$ si ha $E(X) = np$ e $\text{Var}(X) = np(1 - p)$.

6.11 Esercizi

Esercizio 6.1 Sia \mathcal{E} l'esperimento casuale che consiste nel lancio di un dado equilibrato. Dopo aver descritto gli ingredienti dello spazio probabilizzato (S, \mathcal{F}, P) che modella \mathcal{E} , si consideri la v.c. univariata $X : S \rightarrow \mathbb{R}$ definita da $X(s) = \text{'punteggio mostrato dal dado'}$. Si determini che X ha legge discreta e si calcolino supporto e f.m.p. di X .

Esercizio 6.2 Sia \mathcal{E} l'esperimento casuale che consiste nel lancio di due dadi equilibrati. Dopo aver descritto gli ingredienti dello spazio probabilizzato (S, \mathcal{F}, P) che modella \mathcal{E} , si consideri la v.c. univariata $X : S \rightarrow \mathbb{R}$ definita da $X(s) = \text{'somma dei punteggi mostrati dai due dadi'}$. Si determini che X ha legge discreta e si calcolino supporto e f.m.p. di X .

Esercizio 6.3 Si indagli se le v.c. univariate $X \sim Ud(2)$ e $Y \sim \mathcal{D}(2)$ sono identicamente distribuite.

Esercizio 6.4 Si indagli se le v.c. univariate $X \sim Bi(2, 0.5)$ e $Y \sim Ud(0, 1, 2)$ sono identicamente distribuite.

Esercizio 6.5 Sia $X \sim Ud(x_1, \dots, x_k)$. Si calcolino $E(X)$ e $\text{Var}(X)$.

Esercizio 6.6 Sia $X \sim Bi(3, p)$ con $p \in (0, 1)$. Si calcoli $E(X)$.

Esercizio 6.7 Sia $X \sim Ig(2; 2, 4)$. Si rappresenti X in forma di tabella e si calcoli $E(X)$.

Esercizio 6.8 (*La legge di Benford*)

La cifra più significativa di $\pi \doteq 3.141593$ è 3, di $\pi^4 \doteq 97.4091$ è 9. Contrariamente all'intuizione, in molti *dataset* la cifra più significativa, X , non è uniformemente distribuita su $S_X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, ma le frequenze sono decrescenti. In particolare 1 si presenta circa il 30% delle volte, 9 solo circa il 5% delle volte. La legge di Benford è un modello per la distribuzione della prima cifra significativa. Ha f.m.p. espressa tramite il logaritmo in base 10

$$p_X(x) = \log_{10} \left(1 + \frac{1}{x} \right) = \log_{10} \left(\frac{1+x}{x} \right), \quad x \in S_X.$$

Si dimostri che $p_X(x)$ è effettivamente una f.m.p. e se ne tabulino i valori.

6.12 Appendice: R e le v.c. discrete con supporto finito

Si consideri una v.c. univariata X con legge discreta, il cui supporto è $S_X = \{0, 1, 2, 3, 4\}$ e la f.m.p. è $p_X(0) = 0.5$, $p_X(1) = p_X(2) = 0.2$, $p_X(3) = p_X(4) = 0.05$. Una illustrazione grafica di S_X e $p_X(\cdot)$ può essere ottenuta con i comandi

```
> rm(list=ls())
> x=c(0,1,2,3,4)
> p=c(0.5,0.20,0.20,0.05,0.05)
> plot(x,p, type="h" , xlab="valori", ylab="probabilita'")
> title(main="Una f.m.p.", sub="V.c. discreta")
> axis(2, at=c(0.05,0.15,0.25))
```

L'opzione `type="h"` fa sì che i valori in ordinata siano rappresentati da segmenti, `xlab` e `ylab` danno una *label* ad ascissa e ordinata. Il comando `title` permette di inserire un titolo principale e un eventuale sottotitolo. Il comando `axis(1,at=c(0.4,2.5))` permette di inserire valori sull'asse 1 (ascisse).

Per calcolare valore atteso e varianza di X si possono creare due *function*

```
> attesa <- function(x,p){
+ sum(x*p)
+ }
>
> varianza <- function(x,p){
+ sum((x-attesa(x,p))^2*p)
+ }
>
> attesa(x,p)
[1] 0.95
>
> varianza(x,p)
[1] 1.3475
```

Per simulare la generazione di `nrep` (numero repliche) realizzazioni da X si usa il comando `sample(x,nrep,replace=TRUE, prob=p)`. Si consideri

```
> set.seed(12345)
> y=sample(x,10^4,replace=TRUE, prob=p)
> y[1:30]
[1] 1 1 1 1 0 0 0 2 1 3 0 0 1 0 0 0 0 0 3 0 0 3 1 2 0 2 2 0 0
> mean(y)
[1] 0.9497
> var(y)
[1] 1.326503
```

Il comando `mean(y)` calcola la media campionaria dei valori empirici y , ossia $\bar{y}_n = \sum_{i=1}^n y_i/n$ dove n è pari a `nrep`. Il comando `var` calcola la varianza campionaria corretta dei valori y , ossia

$$\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2.$$

Con n grande, come in questo caso, la varianza campionaria corretta è quasi uguale alla varianza (non corretta) dei valori empirici y , che ha denominatore n invece che $n-1$.

Media e varianza campionarie risultano qui molto vicine a $E(X)$ e $\text{Var}(X)$. Infatti le frequenze relative dei valori $0, 1, \dots, 4$ nelle 10^4 repliche approssimano bene le corrispondenti probabilità:

```
> table(y)/10^4
y
  0      1      2      3      4
0.4959 0.2033 0.2049 0.0470 0.0489
```

Si desidera infine verificare con R che per $X \sim \text{Bi}(n, p)$ si ha $E(X) = np$. Si fissa ad esempio $n = 100$ e si scelgono alcuni valori di p . La verifica è fatta sia per simulazione stocastica sia numericamente.

```
> rm(list=ls())
> p=1:99/100
> n=100
> emp_expec=rep(NA,length(p))
> x=0:100
> attesa=rep(NA,length(p))
> for (i in 1:length(p)){
+   emp_expec[i]=mean(rbinom(10^4,size=n,prob=p[i]))
+   attesa[i]=sum(x*dbinom(x,size=n, prob=p[i]))
+ }
> max(abs(emp_expec-n*p))
[1] 0.1178
> max(abs(attesa-n*p))
[1] 4.263256e-14
```

La precisione della verifica numerica è molto maggiore di quella dell'approcio di simulazione stocastica. Ma si provino i comandi `plot(p,emp_expec)` e `lines(p,n*p,col=2)`.

Lezione 7

Le v.c. bivariate con legge discreta

7.1 Definizione

Si adotta spesso per una v.c. bivariata, cioè per un punto in \mathbb{R}^2 determinato dal caso, la notazione (X, Y) , ispirata alla notazione (x, y) per le coordinate di un generico punto del piano cartesiano. Si parla di supporto congiunto, legge congiunta, eccetera, della v.c. bivariata per sottolineare il fatto che viene modellata la variabilità simultanea delle v.c. univariate X e Y , componenti di (X, Y) .

Una v.c. bivariata con legge discreta è definita da

- il supporto congiunto $S_{X,Y} = \{(x_i, y_i) \in \mathbb{R}^2, i \in I \subseteq \mathbb{N}\}$, che è una successione finita o numerabile di punti distinti in \mathbb{R}^2 (senza punti di accumulazione al finito)
- la f.m.p. congiunta $p_{X,Y} : S_{X,Y} \rightarrow [0, 1]$ positiva e normalizzata: $p_{X,Y}(x, y) > 0$ per ogni $(x, y) \in S_{X,Y}$ e $\sum_{(x,y) \in S_{X,Y}} p_{X,Y}(x, y) = 1$.

La legge congiunta della v.c. (X, Y) è allora, per ogni $B \in \mathcal{B}_2$, data da

$$P_{X,Y}(B) = P((X, Y) \in B) = \sum_{(x,y) \in B \cap S_{X,Y}} p_{X,Y}(x, y). \quad (7.1)$$

Si presenta subito un esempio molto semplice di v.c. bivariata con legge discreta. Si prescrive come supporto congiunto, quindi come insieme dei valori realizzabili dall'esperimento di osservazione di (X, Y) ,

$$S_{X,Y} = \{(0, 0), (0, 1), (1, 0)\}.$$

Per ognuno dei punti di $S_{X,Y}$ la probabilità di realizzazione è data dalla f.m.p.

$$p_{X,Y}(0, 0) = p_{X,Y}(0, 1) = p_{X,Y}(1, 0) = \frac{1}{3}.$$

Quindi $(X, Y) \sim Ud((0, 0), (0, 1), (1, 0))$. Si vede subito che

$$P(X \geq 0, Y \geq 0) = 1, \quad P(X > 0, Y > 0) = 0,$$

e la probabilità di qualunque altro evento è calcolabile usando la (7.1).

7.2 Le leggi marginali

Da una assegnata v.c. bivariata con legge discreta (X, Y) , specificata da $S_{X,Y}$ e $p_{X,Y}(x, y)$, si possono ‘estrarre’ varie leggi di v.c. univariate, tutte di tipo discreto. Anzitutto conviene considerare le leggi delle componenti X e Y , dette **leggi marginali**.

La legge marginale di X ha

- supporto marginale:
 $S_X = \{x \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ per qualche } y \in \mathbb{R}\}$
- f.m.p. marginale: per $x \in S_X$
 $p_X(x) = P(X = x) = \sum_{y: (x,y) \in S_{X,Y}} p_{X,Y}(x, y).$

A parole: il supporto marginale di X è costituito dal complesso dei punti di \mathbb{R} che sono ascisse di punti visibili in \mathbb{R}^2 , ossia ascisse di punti in $S_{X,Y}$. Ogni punto fissato x di S_X si presenterà con una probabilità pari alla somma delle probabilità dei punti di $S_{X,Y}$ la cui ascissa è x . Per i punti con ascissa x ciò che varia è l'ordinata. Costituiscono l'insieme $S_{Y|X=x} = \{y \in \mathbb{R} : (x, y) \in S_{X,Y}\}$.

Analogamente, la legge marginale di Y ha

- supporto marginale:
 $S_Y = \{y \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ per qualche } x \in \mathbb{R}\}$
- f.m.p. marginale: per $y \in S_Y$
 $p_Y(y) = P(Y = y) = \sum_{x: (x,y) \in S_{X,Y}} p_{X,Y}(x, y).$

A parole: il supporto marginale di Y è costituito dal complesso dei punti di \mathbb{R} che sono ordinate di punti visibili in \mathbb{R}^2 , ossia ordinate di punti in $S_{X,Y}$. Ogni punto fissato y di S_Y si presenterà con una probabilità pari alla somma delle probabilità dei punti di $S_{X,Y}$ la cui ordinata è y . Per i punti con ordinata y varia solo l'ascissa. Costituiscono l'insieme $S_{X|Y=y} = \{x \in \mathbb{R} : (x, y) \in S_{X,Y}\}$.

Osservazione. Si supponga di poter ottenere, con R esperimenti indipendenti, da (X, Y) realizzazioni $(x_1, y_1), \dots, (x_R, y_R)$. Allora i valori $x_r, r = 1, \dots, R$, (che ignorano le y_r) sono realizzazioni indipendenti di X con la sua legge marginale, e analogamente i valori $y_r, r = 1, \dots, R$ (che ignorano le x_r) sono realizzazioni indipendenti di Y con la sua legge marginale.

Esempio 7.1 Sia $(X, Y) \sim Ud((0, 0), (0, 1), (1, 1))$, per cui il supporto è $S_{X,Y} = \{(0, 0), (0, 1), (1, 1)\}$, con i tre punti equiprobabili. Allora le leggi marginali

hanno supporti

$$\begin{aligned} S_X &= \{x \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ per qualche } y\} = \{0, 1\} \\ S_Y &= \{y \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ per qualche } x\} = \{0, 1\} \end{aligned}$$

mentre le f.m.p. marginali sono date da

$$\begin{aligned} p_X(0) &= p_{X,Y}(0, 0) + p_{X,Y}(0, 1) = \frac{2}{3} \\ p_X(1) &= p_{X,Y}(1, 1) = \frac{1}{3} \end{aligned}$$

e da

$$\begin{aligned} p_Y(0) &= p_{X,Y}(0, 0) = \frac{1}{3} \\ p_Y(1) &= p_{X,Y}(0, 1) + p_{X,Y}(1, 1) = \frac{2}{3}. \end{aligned}$$

◇

7.3 Le leggi condizionali

Da una assegnata v.c. bivariata con legge discreta (X, Y) si possono ‘estrarre’ anche due famiglie di leggi condizionali, tutte di tipo discreto.

La **legge condizionale di Y dato un valore osservabile di X** , indicata con $Y|X = x$, $x \in S_X$, ha

- supporto condizionale:
 $S_{Y|X=x} = \{y \in \mathbb{R} : (x, y) \in S_{X,Y}\}$
- f.m.p. condizionale: per $y \in S_{Y|X=x}$

$$p_{Y|X=x}(y) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

Analogamente, la **legge condizionale di X dato un valore osservabile di Y** , indicata con $X|Y = y$, $y \in S_Y$, ha

- supporto condizionale:
 $S_{X|Y=y} = \{x \in \mathbb{R} : (x, y) \in S_{X,Y}\}$
- f.m.p. condizionale: per $x \in S_{X|Y=y}$

$$p_{X|Y=y}(x) = P(X = x|Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

Osservazione. Si supponga di ottenere, con esperimenti indipendenti, da (X, Y) realizzazioni $(x_1, y_1), \dots, (x_R, y_R)$. Se si selezionano i valori y_r per gli r tali che $x_r = x$, si ottengono realizzazioni indipendenti di una v.c. univariata che ha la legge di $Y|X = x$. Analogamente, i valori x_r per gli r tali che $y_r = y$ sono realizzazioni indipendenti di una v.c. univariata che ha la legge di $X|Y = y$.

Esempio 7.2 Sia $(X, Y) \sim Ud((0, 0), (0, 1), (1, 1))$, con leggi marginali di X e di Y come nell'Esempio 7.1. Le leggi condizionali di Y dato un valore osservabile di X sono due, $Y|X=0$ e $Y|X=1$. Si ha che

$$\begin{aligned} S_{Y|X=0} &= \{y \in \mathbb{R} : (0, y) \in S_{X,Y}\} = \{0, 1\} \\ S_{Y|X=1} &= \{y \in \mathbb{R} : (1, y) \in S_{X,Y}\} = \{1\}, \end{aligned}$$

per cui $Y|X=0 \sim Bi(1, P(Y=1|X=0))$ e $Y|X=1 \sim \mathcal{D}(1)$. Si vede subito che

$$P(Y=1|X=0) = \frac{P(X=0, Y=1)}{P(X=0)} = \frac{1/3}{2/3} = \frac{1}{2}.$$

◇

7.4 La specificazione gerarchica

Si prescrive una legge congiunta (X, Y) di tipo discreto tramite specificazione gerarchica assegnando:

- la legge marginale di X
ossia, equivalentemente, supporto S_X e f.m.p. $p_X(x)$
- le leggi condizionali di $Y|X=x$ per ogni $x \in S_X$
ossia, equivalentemente, per tutti gli $x \in S_X$
i supporti $S_{Y|X=x}$ e le f.m.p. $p_{Y|X=x}(y)$.

Si determina così una v.c. bivariata (X, Y) con legge discreta. Essa ha supporto congiunto

$$S_{X,Y} = \{(x, y) \in \mathbb{R}^2 : x \in S_X, y \in S_{Y|X=x}\},$$

esprimibile anche come

$$S_{X,Y} = \cup_{x \in S_X} \{(x, y) : y \in S_{Y|X=x}\} = \cup_{x \in S_X} \{x\} \times S_{Y|X=x},$$

e, per $(x, y) \in S_{X,Y}$, f.m.p. congiunta

$$\begin{aligned} p_{X,Y}(x, y) &= P(X=x, Y=y) = \text{(per il teorema della probabilità composta)} \\ &= P(X=x)P(Y=y|X=x) = \text{(per definizione di f.m.p.)} \\ &= p_X(x)p_{Y|X=x}(y). \end{aligned}$$

Esempio 7.3 Sia (X, Y) una v.c. bivariata tale che $X \sim Bi(2, 0.5)$ e, per ogni $x \in S_X$, sia $Y|X=x \sim Bi(1, 0.5)$. Si desidera reperire $S_{X,Y}$ e $p_{X,Y}(x, y)$. Conviene analizzare i dati del problema. Marginalmente si ha

$$X \sim Bi(2, 0.5) \iff S_X = \{0, 1, 2\} \text{ e } p_X(0) = \frac{1}{4}, p_X(1) = \frac{1}{2}, p_X(2) = \frac{1}{4}.$$

Si devono quindi considerare le tre leggi condizionali $Y|X = 0$, $Y|X = 1$, $Y|X = 2$, identicamente distribuite:

$$Y|X = x \sim Bi(1, 0.5) \iff S_{Y|X=x} = \{0, 1\} \text{ e } p_{Y|X=x}(0) = \frac{1}{2}, p_{Y|X=x}(1) = \frac{1}{2},$$

per $x = 0, 1, 2$.

Quindi

$$\begin{aligned} S_{X,Y} &= \{(x, y) \in \mathbb{R}^2 : x \in S_X, y \in S_{Y|X=x}\} \\ &= \{(x, y) \in \mathbb{R}^2 : x \in \{0, 1, 2\}, y \in \{0, 1\}\} \\ &= \{0, 1, 2\} \times \{0, 1\} = \{(0, 0), (0, 1), (1, 0), (1, 1), (2, 0), (2, 1)\} \end{aligned}$$

e inoltre

$$\begin{aligned} p_{X,Y}(0, 0) &= p_X(0) p_{Y|X=0}(0) = \frac{1}{4} \frac{1}{2} = \frac{1}{8} \\ p_{X,Y}(0, 1) &= p_X(0) p_{Y|X=0}(1) = \frac{1}{4} \frac{1}{2} = \frac{1}{8} \\ p_{X,Y}(1, 0) &= p_X(1) p_{Y|X=1}(0) = \frac{1}{2} \frac{1}{2} = \frac{1}{4} \\ p_{X,Y}(1, 1) &= p_X(1) p_{Y|X=1}(1) = \frac{1}{2} \frac{1}{2} = \frac{1}{4} \\ p_{X,Y}(2, 0) &= p_X(2) p_{Y|X=2}(0) = \frac{1}{4} \frac{1}{2} = \frac{1}{8} \\ p_{X,Y}(2, 1) &= p_X(2) p_{Y|X=2}(1) = \frac{1}{4} \frac{1}{2} = \frac{1}{8}. \end{aligned}$$

La legge bivariata discreta di (X, Y) può essere rappresentata in forma di tabella come una **tabella a doppia entrata**

$Y \quad X$	0	1	2	Totale
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
Totale	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

◇

7.5 V.c. con legge discreta, componenti indipendenti

Come nella probabilità elementare, il più importante strumento di modellazione della legge congiunta di una variabile casuale è assumere l'indipendenza delle componenti. Per v.c. bivariate con legge di tipo discreto si ha la seguente definizione.

Definizione 7.1 *La v.c. con legge discreta (X, Y) si dice **con componenti indipendenti** se*

$$S_{X,Y} = S_X \times S_Y \quad (7.2)$$

e se inoltre per ogni $(x, y) \in S_{X,Y}$ si ha

$$p_{X,Y}(x, y) = p_X(x)p_Y(y). \quad (7.3)$$

La condizione (7.3) è la solita fattorizzazione della probabilità dell'intersezione, $P(X = x, Y = y) = P(X = x)P(Y = y)$.

Se $Y|X = x \sim Y$ per ogni $x \in S_X$ allora le componenti X e Y di (X, Y) sono indipendenti. Infatti per la specificazione gerarchica si ha

$$\begin{aligned} S_{X,Y} &= \{(x, y) \in \mathbb{R}^2 : x \in S_X, y \in S_{Y|X=x}\} \\ &= \{(x, y) \in \mathbb{R}^2 : x \in S_X, y \in S_Y\} = S_X \times S_Y \end{aligned}$$

e inoltre

$$p_{X,Y}(x, y) = p_X(x)p_{Y|X=x}(y) = p_X(x)p_Y(y)$$

per ogni $(x, y) \in S_{X,Y}$.

Anche se $Y|X = x \sim Y|X = x'$ per ogni $x, x' \in S_X$, allora le componenti X e Y di (X, Y) sono indipendenti, e ovviamente $Y|X = x \sim Y$ per ogni $x \in S_X$.

Si noti che la condizione (7.2), molto facile da verificare, è una condizione necessaria, ma non sufficiente, per l'indipendenza di X e Y . Se ad esempio (X, Y) ha $S_{X,Y} = \{(0, 0), (0, 1), (1, 0)\}$, si deduce subito che X e Y non sono indipendenti. Invece per una v.c. bivariata con legge discreta (X, Y) con

$$S_{X,Y} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

nulla si può dedurre senza esaminare se vale o no anche la fattorizzazione (7.3).

Una v.c. multivariata con legge discreta ha componenti indipendenti se vale la naturale generalizzazione delle fattorizzazioni (7.2) e (7.3).

7.6 La covarianza

Se le componenti X e Y di (X, Y) non sono indipendenti, si dice che la v.c. bivariata (X, Y) ha **componenti dipendenti**. Un indice sintetico di dipendenza delle componenti di una v.c. bivariata è costituito dalla covarianza. Conviene

indicare valori e probabilità in forma di tabella (eventualmente infinita). A tale scopo, si rappresenta il supporto di (X, Y) come successione dipendente da due indici,

$$S_{X,Y} = \{(x_i, y_j), i \in I, j \in J\} \quad \text{dove } I \text{ e } J \text{ sono finiti o numerabili}$$

e si esprime la f.m.p. in forma abbreviata come applicazione

$$S_{X,Y} \ni (x_i, y_j) \rightarrow p_{ij} = P(X = x_i, Y = y_j).$$

La **covarianza tra X e Y** , indicata con il simbolo $\text{Cov}(X, Y)$, è allora definita come media ponderata del prodotto di scarti delle variabili dalla loro media

$$\text{Cov}(X, Y) = \sum_{i \in I} \sum_{j \in J} (x_i - E(X))(y_j - E(Y))p_{ij} = E((X - E(X))(Y - E(Y))),$$

purché la serie converga assolutamente. Se X e Y sono indipendenti, si ha

$$\begin{aligned} \text{Cov}(X, Y) &= \sum_{i \in I} \sum_{j \in J} (x_i - E(X))(y_j - E(Y))p_i p_j \\ &= \sum_{i \in I} (x_i - E(X))p_i \sum_{j \in J} (y_j - E(Y))p_j \end{aligned}$$

per cui

$$\text{Cov}(X, Y) = 0$$

dal momento che

$$\sum_{i \in I} (x_i - E(X))p_i = \sum_{i \in I} x_i p_i - \sum_{i \in I} E(X)p_i = E(X) - E(X) \sum_{i \in I} p_i = 0$$

valendo $\sum_{i \in I} p_i = 1$ per la normalizzazione.

7.7 Esercizi risolti

Esempio 7.4 Sia (X, Y) una variabile casuale bivariata con componente marginale $X \sim Bi(1, 1/2)$ (legge binomiale con indice $n = 1$ e parametro $p = 1/2$) e distribuzioni condizionate binomiali $Y|X = x \sim Bi(1, 1/2)$, per $x \in S_X$. Si determinino il supporto congiunto di (X, Y) , la funzione di probabilità congiunta di (X, Y) , il supporto marginale di Y , la funzione di probabilità marginale di Y . Si dica, motivando, se (X, Y) ha componenti indipendenti. Si calcoli infine $P(X + Y = 1)$.

Conviene iniziare analizzando i dati del problema.

$$X \sim Bi(1, 0.5) \iff S_X = \{0, 1\} \text{ e } p_X(0) = \frac{1}{2}, p_X(1) = \frac{1}{2}$$

per cui vi sono due distribuzioni condizionali, $Y|X = 0$ e $Y|X = 1$, identicamente distribuite, con legge $Bi(1, 0.5)$. Quindi per $x = 0, 1$

$$Y|X = x \sim Bi(1, 0.5) \iff S_{Y|X=x} = \{0, 1\} \text{ e } p_{Y|X=x}(0) = \frac{1}{2}, p_{Y|X=x}(1) = \frac{1}{2}.$$

Si ha allora

$$\begin{aligned} S_{X,Y} &= \{(x, y) \in \mathbb{R}^2 : x \in S_X, y \in S_{Y|X=x}\} \\ &= \{(x, y) \in \mathbb{R}^2 : x \in \{0, 1\}, y \in \{0, 1\}\} = \{(0, 0), (0, 1), (1, 0), (1, 1)\} \end{aligned}$$

e inoltre

$$\begin{aligned} p_{X,Y}(0, 0) &= p_X(0) p_{Y|X=0}(0) = \frac{1}{2} \frac{1}{2} = \frac{1}{4} \\ p_{X,Y}(0, 1) &= p_X(0) p_{Y|X=0}(1) = \frac{1}{2} \frac{1}{2} = \frac{1}{4} \\ p_{X,Y}(1, 0) &= p_X(1) p_{Y|X=1}(0) = \frac{1}{2} \frac{1}{2} = \frac{1}{4} \\ p_{X,Y}(1, 1) &= p_X(1) p_{Y|X=1}(1) = \frac{1}{2} \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

Per la legge marginale di Y si ha

$$S_Y = \{y \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ per qualche } x\} = \cup_{x \in S_X} S_{Y|X=x} = \{0, 1\}$$

e

$$p_Y(0) = p_{X,Y}(0, 0) + p_{X,Y}(1, 0) = \frac{1}{2} \quad p_Y(1) = p_{X,Y}(0, 1) + p_{X,Y}(1, 1) = \frac{1}{2}.$$

La v.c. bivariata (X, Y) ha componenti indipendenti poiché

$$S_{X,Y} = \{(0, 0), (1, 0), (0, 1), (1, 1)\} = \{0, 1\} \times \{0, 1\} = S_X \times S_Y$$

e, per ogni $(x, y) \in S_{X,Y}$, vale

$$p_{X,Y}(x, y) = \frac{1}{4} = \frac{1}{2} \frac{1}{2} = p_X(x) p_Y(y).$$

Infine,

$$P(X + Y = 1) = p_{X,Y}(0, 1) + p_{X,Y}(1, 0) = \frac{1}{2}.$$

◇

Esempio 7.5 Sia (X, Y) una variabile casuale bivariata con componente marginale $X \sim Bi(2, 1/2)$ (legge binomiale con indice $n = 2$ e parametro $p = 1/2$) e distribuzioni condizionate binomiali $Y|X = x \sim Bi(1, 1/2)$, per $x \in S_X$. Si determinino il supporto congiunto di (X, Y) , la funzione di probabilità congiunta di (X, Y) , il supporto marginale di Y , la funzione di probabilità marginale di Y . Si dica, motivando, se (X, Y) ha componenti indipendenti. Si calcoli infine

$P(X + Y > 0)$.

Conviene iniziare analizzando i dati del problema.

$$X \sim Bi(2, 0.5) \iff S_X = \{0, 1, 2\} \text{ e } p_X(0) = \frac{1}{4}, p_X(1) = \frac{1}{2}, p_X(2) = \frac{1}{4}$$

per cui vi sono tre distribuzioni condizionali, $Y|X = 0$, $Y|X = 1$, $Y|X = 2$, identicamente distribuite, tutte con legge $Bi(1, 0.5)$,

$$Y|X = x \sim Bi(1, 0.5) \iff S_{Y|X=x} = \{0, 1\} \text{ e } p_{Y|X=x}(0) = \frac{1}{2}, p_{Y|X=x}(1) = \frac{1}{2}$$

per $x = 0, 1, 2$. Si ha allora

$$\begin{aligned} S_{X,Y} &= \{(x, y) \in \mathbb{R}^2 : x \in S_X, y \in S_{Y|X=x}\} \\ &= \{(x, y) \in \mathbb{R}^2 : x \in \{0, 1, 2\}, y \in \{0, 1\}\} \\ &= \{(0, 0), (0, 1), (1, 0), (1, 1), (2, 0), (2, 1)\} \end{aligned}$$

e inoltre

$$\begin{aligned} p_{X,Y}(0, 0) &= p_X(0) p_{Y|X=0}(0) = \frac{1}{4} \frac{1}{2} = \frac{1}{8} \\ p_{X,Y}(0, 1) &= p_X(0) p_{Y|X=0}(1) = \frac{1}{4} \frac{1}{2} = \frac{1}{8} \\ p_{X,Y}(1, 0) &= p_X(1) p_{Y|X=1}(0) = \frac{1}{2} \frac{1}{2} = \frac{1}{4} \\ p_{X,Y}(1, 1) &= p_X(1) p_{Y|X=1}(1) = \frac{1}{2} \frac{1}{2} = \frac{1}{4} \\ p_{X,Y}(2, 0) &= p_X(2) p_{Y|X=2}(0) = \frac{1}{4} \frac{1}{2} = \frac{1}{8} \\ p_{X,Y}(2, 1) &= p_X(2) p_{Y|X=2}(1) = \frac{1}{4} \frac{1}{2} = \frac{1}{8}. \end{aligned}$$

Per la legge marginale di Y si ha

$$S_Y = \{y \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ per qualche } x\} = \cup_{x \in S_X} S_{Y|X=x} = \{0, 1\}$$

e

$$\begin{aligned} p_Y(0) &= p_{X,Y}(0, 0) + p_{X,Y}(1, 0) + p_{X,Y}(2, 0) = \frac{1}{2} \\ p_Y(1) &= p_{X,Y}(0, 1) + p_{X,Y}(1, 1) + p_{X,Y}(2, 1) = \frac{1}{2}. \end{aligned}$$

La v.c. bivariata (X, Y) ha componenti indipendenti poiché

$$S_{X,Y} = \{(0, 0), (1, 0), (0, 1), (1, 1), (2, 0), (2, 1)\} = \{0, 1, 2\} \times \{0, 1\} = S_X \times S_Y$$

e, per ogni $(x, y) \in S_{X,Y}$, vale

$$p_{X,Y}(x, y) = p_X(x) p_Y(y).$$

Infine,

$$P(X + Y > 0) = 1 - P(X + Y \leq 0) = 1 - p_{X,Y}(0, 0) = 1 - \frac{1}{8} = \frac{7}{8}.$$

◇

7.8 Esercizi

Esercizio 7.1 Sia (X, Y) una v.c. bivariata con supporto congiunto $S_{X,Y} = \{(0, 0), (5, 0), (0, 3), (5, 4)\}$ e f.m.p. congiunta $p_{X,Y}(0, 0) = 1/8$, $p_{X,Y}(5, 0) = 1/8$, $p_{X,Y}(0, 3) = 1/4$, $p_{X,Y}(5, 4) = 1/2$. Si ottengano supporto e f.m.p. marginali di X e di Y . Si dica se (X, Y) ha componenti i) identicamente distribuite ii) indipendenti.

Esercizio 7.2 Sia (X, Y) una v.c. bivariata con supporto congiunto $S_{X,Y} = \{(0, 0), (1, 0), (0, 2), (1, 1), (2, 2)\}$ e f.m.p. congiunta data da $p_{X,Y}(0, 0) = 1/10$, $p_{X,Y}(1, 0) = 1/10$, $p_{X,Y}(0, 2) = 1/5$, $p_{X,Y}(1, 1) = 1/5$, $p_{X,Y}(2, 2) = 2/5$. Si ottengano supporto e f.m.p. marginali di X e di Y . Si dica se (X, Y) ha componenti i) identicamente distribuite ii) indipendenti. Si calcoli $\text{Cov}(X, Y)$.

Esercizio 7.3 Sia (X, Y) una v.c. bivariata con supporto congiunto $S_{X,Y} = \{(0, 0), (5, 0), (0, 5), (5, 5)\}$ e f.m.p. congiunta $p_{X,Y}(0, 0) = 1/4$, $p_{X,Y}(5, 0) = 1/4$, $p_{X,Y}(0, 5) = 1/4$, $p_{X,Y}(5, 5) = 1/4$. Si ottengano supporto e f.m.p. marginali di X e di Y . Si dica se (X, Y) ha componenti i) identicamente distribuite ii) indipendenti. Si ottenga $\text{Cov}(X, Y)$.

Esercizio 7.4 (compito del 01/02/17) Sia (X, Y) una variabile casuale bivariata con componente marginale $X \sim \text{Bi}(1, 2/3)$ (legge binomiale con indice $n = 1$ e parametro $p = 2/3$) e distribuzioni condizionate binomiali $Y|X = x \sim \text{Bi}(1 + x, 1/2)$, per $x \in S_X$. Si determinino il supporto congiunto di (X, Y) , la funzione di probabilità congiunta di (X, Y) , il supporto marginale di Y , la funzione di probabilità marginale di Y . Si dica, motivando, se (X, Y) ha componenti indipendenti. Si calcoli infine $P(XY > 1)$.

Esercizio 7.5 (compito del 10/02/17) Sia (X, Y) una variabile casuale bivariata con componente marginale $X \sim \text{Bi}(1, 1/2)$ (legge binomiale con indice $n = 1$ e parametro $p = 1/2$) e distribuzioni condizionate binomiali $Y|X = x \sim \text{Bi}(1, (1 + x)/3)$, per $x \in S_X$. Si determinino il supporto congiunto di (X, Y) , la funzione di probabilità congiunta di (X, Y) , il supporto marginale di Y , la funzione di probabilità marginale di Y . Si dica, motivando, se (X, Y) ha componenti indipendenti. Si calcoli infine $P(X > Y)$.

Esercizio 7.6 (compito del 18/06/18) Sia (X, Y) una variabile casuale bivariata con componente marginale $X \sim \text{Bi}(2, 1/2)$ (legge binomiale con indice $n = 2$ e parametro $p = 1/2$) e distribuzioni condizionate binomiali $Y|X = x \sim \text{Bi}(1, 1/2)$, per $x \in S_X$. Si determinino il supporto congiunto di (X, Y) , la funzione di probabilità congiunta di (X, Y) , il supporto marginale di Y , la funzione di probabilità marginale di Y . Si dica, motivando, se (X, Y) ha componenti indipendenti. Si calcoli infine $P(X + Y = 2)$.

Esercizio 7.7 (compito del 17/07/17) Sia (X, Y) una variabile casuale bivariata con componente marginale $X \sim \text{Bi}(1, 1/2)$ (legge binomiale con indice $n = 1$ e parametro $p = 1/2$) e distribuzioni condizionate binomiali $Y|X =$

$x \sim Bi(1, (1+x)/3)$, per $x \in S_x$. Si determinino il supporto congiunto di (X, Y) , la funzione di probabilità congiunta di (X, Y) , il supporto marginale di Y , la funzione di probabilità marginale di Y . Si dica, motivando, se (X, Y) ha componenti indipendenti. Si calcoli infine $P(X > Y)$.

Esercizio 7.8 (*compito del 11/09/17*) Sia (X, Y) una variabile casuale bivarata con componente marginale $X \sim Bi(1, 1/2)$ (legge binomiale con indice $n = 1$ e parametro $p = 1/2$) e distribuzioni condizionate binomiali $Y|X = x \sim Bi(1, (1+2x)/4)$, per $x \in S_x$. Si determinino il supporto congiunto di (X, Y) , la funzione di probabilità congiunta di (X, Y) , il supporto marginale di Y , la funzione di probabilità marginale di Y . Si dica, motivando, se (X, Y) ha componenti indipendenti. Si calcoli infine $P(X > Y)$.

7.9 Appendice: v.c. bivariate discrete con R

Sia $X \sim Bi(m, p)$ e $Y|X = x \sim Bi(n(x), q(x))$, dove $m, n(x) \in \mathbb{N}^+$ e $p, q(x) \in (0, 1)$. Si desidera calcolare con R il supporto congiunto $S_{X,Y}$, la f.m.p. congiunta $p_{X,Y}(x, y)$, supporto e f.m.p. marginale di Y , risolvendo l'Esercizio 7.4. Si inizia con l'inserimento dei dati del problema e il calcolo del supporto congiunto, S_{xy} :

```
> m=1
> p=2/3
> nx <- function(x) 1+x
> qx <- function(x) 0.5
> suppxy <- function(m){
+ Sxy=c()
+ for (x in 0:m){
+   for (y in 0:nx(x)){
+     Sxy=rbind(Sxy,c(x,y))
+   }
+ }
+ return(Sxy)
+ }
> Sxy=suppxy(m)
> colnames(Sxy)=c("x", "y")
>
> Sxy
      x y
[1,] 0 0
[2,] 0 1
[3,] 1 0
[4,] 1 1
[5,] 1 2
```

L'inserimento di $n(x)$ e $q(x)$ come `function` permette di adattare facilmente il programma ai problemi dello stesso tipo.

La f.m.p. congiunta come applicazione $p_{X,Y} : S_{X,Y} \rightarrow (0, 1]$ è pure rappresentabile come `function`:

```
> pxy <- function(Sxy){
+   pxy=rep(NA, length(Sxy)/2)
+   for (i in 1:length(Sxy)/2) {
+     x=Sxy[i,1]
+     y=Sxy[i,2]
+     pxy[i]=dbinom(x,m,p)*dbinom(y,nx(x),qx(x))
+   }
+   return(pxy)
+ }
> pxy=pxy(Sxy)
> ris=cbind(Sxy,pxy)
> print(ris)
      x y      pxy
[1,] 0 0 0.1666667
[2,] 0 1 0.1666667
[3,] 1 0 0.1666667
[4,] 1 1 0.3333333
[5,] 1 2 0.1666667
```

Per ottenere supporto e f.m.p. marginale di Y si può usare il comando `tapply`, che crea tabulazioni (da qui il `t`) applicando al primo argomento, selezionato secondo la partizione definita dai valori distinti del secondo argomento, la funzione data come terzo argomento:

```
> tapply(ris[,3],ris[,2],sum)
      0      1      2
0.3333333 0.5000000 0.1666667
```

Per controllo, si calcola anche la marginale X :

```
> tapply(ris[,3],ris[,1],sum)
      0      1
0.3333333 0.6666667
```

Come ultimo passo, si desidera calcolare $\text{Cov}(X, Y)$.

```
> covarianza <- function(Sxy,pxy){
+   EX = sum(Sxy[,1]*pxy)
+   EY = sum(Sxy[,2]*pxy)
+   return(sum((Sxy[,1]-EX)*(Sxy[,2]-EY)*pxy))
+ }
> covarianza(Sxy,pxy)
[1] 0.1111111
```

Poiché $\text{Cov}(X, Y) \neq 0$, X e Y non sono indipendenti. La dipendenza è positiva: $E(Y|X=0) = 0.5$ e $E(Y|X=1) = 1$. Se X cresce, in media cresce anche Y .

Lezione 8

V.c. con legge continua

La distinzione fra continuo e discreto è fondamentale in Matematica e nelle scienze matematiche. Dopo aver delineato nelle due lezioni precedenti le nozioni di base per trattare le v.c. con legge discreta, conviene presentare adesso le v.c. con legge continua. Sono v.c. che si manifestano in un continuo di potenziali valori, quale ad esempio l'intervallo $[0, 1]$, la semiretta $[0, +\infty)$, l'intera retta \mathbb{R} . La probabilità di osservare un punto prespecificato sarà allora 0. Una funzione di massa di probabilità non può modellare tale comportamento, che sarà descritto invece da una funzione di densità di probabilità.

8.1 Leggi univariate di tipo continuo

Si dice che X è una **v.c. univariata con legge di tipo continuo** (in breve: con legge continua; ancora più in breve: una v.c. univariata continua) se per ogni $B \in \mathcal{B}_1$ si può esprimere $P_X(B)$ in forma integrale come

$$P_X(B) = P(X \in B) = \int_B p_X(x) dx = \int_a^b p_X(x) dx \quad (8.1)$$

se $B = [a, b]$, $a < b$, dove la funzione $p_X(x)$, detta **funzione di densità di probabilità** (f.d.p.), soddisfa le condizioni

- i) $p_X(x) \geq 0$ per ogni $x \in \mathbb{R}$ (non negatività)
- ii) $\int_{\mathbb{R}} p_X(x) dx = \int_{-\infty}^{+\infty} p_X(x) dx = 1$ (normalizzazione).

Una P_X definita come in (8.1) tramite una f.d.p. soddisfa gli assiomi di Kolmogorov. Innanzi tutto è evidentemente non negativa e normalizzata. Dalle proprietà dell'integrale ordinario (integrale di Riemann) si ha subito l'additività.

Ad esempio per $a < b < c < d$, $B_1 = [a, b]$ e $B_2 = [c, d]$ sono incompatibili e vale

$$\begin{aligned} P_X(B_1 \cup B_2) &= \int_{[a,b] \cup [c,d]} p_X(x) dx = \int_a^b p_X(x) dx + \int_c^d p_X(x) dx \\ &= P_X(B_1) + P_X(B_2). \end{aligned}$$

Una particolare legge univariata continua è assegnata prescrivendo una particolare f.d.p.. Per modellare opportunamente la legge risulta importante attribuire un significato al valore di $p_X(x)$ in un particolare punto $x_0 \in \mathbb{R}$. Questo è agevole se $p_X(x)$ è continua in un intervallo $(x_0 - \varepsilon_0, x_0 + \varepsilon_0)$, con $\varepsilon_0 > 0$. Si ha allora, per il teorema del valor medio dell'integrale, che per $0 < \varepsilon \leq \varepsilon_0$

$$P(x_0 - \varepsilon \leq X \leq x_0 + \varepsilon) = \int_{x_0 - \varepsilon}^{x_0 + \varepsilon} p_X(x) dx = 2\varepsilon p_X(\tilde{x}_\varepsilon)$$

dove \tilde{x}_ε è un punto in $(x_0 - \varepsilon, x_0 + \varepsilon)$. Ne consegue che

$$\lim_{\varepsilon \rightarrow 0^+} \frac{P(x_0 - \varepsilon \leq X \leq x_0 + \varepsilon)}{2\varepsilon} = p_X(x_0)$$

perché $\lim_{\varepsilon \rightarrow 0^+} p_X(\tilde{x}_\varepsilon) = p_X(x_0)$ per la continuità di $p_X(x)$ in $x = x_0$. Quindi, per valori sufficientemente piccoli di ε ,

$$P(x_0 - \varepsilon \leq X \leq x_0 + \varepsilon) \doteq 2\varepsilon p_X(x_0).$$

In altri termini, in un punto dove è continua, la densità ha un valore proporzionale alla probabilità di un intorno del punto. La densità dunque non è una probabilità ma rappresenta la probabilità per unità di lunghezza (o più in generale di misura di estensione spaziale).

La probabilità che la v.c. continua X assuma la sua realizzazione coincidente con un punto predesignato x è

$$P(X = x) = \lim_{\varepsilon \rightarrow 0} P(x - \varepsilon \leq X \leq x + \varepsilon) = \int_x^x p_X(t) dt = 0,$$

e questo per ogni $x \in \mathbb{R}$. Le v.c. con legge discreta distribuiscono la probabilità su una successione di punti ognuno dei quali ha probabilità strettamente positiva. Al contrario, le v.c. con legge continua diffondono la probabilità sugli intervalli. Infatti, nel continuo vi sono ‘troppi’ punti e la probabilità che si presenti un punto prefissato è sempre 0.

Osservazione. In queste lezioni si usa per la f.d.p. del caso continuo la stessa notazione che si usa per la f.m.p. del caso discreto, ossia $p_X(x)$. Sarà il supporto a distinguere fra caso continuo e caso discreto. In alcuni testi le f.d.p. sono indicate con un simbolo distinto, tipicamente $f_X(x)$.

8.2 Valore atteso e varianza

Per una v.c. univariata con legge continua X l'indice di posizione valore atteso, $E(X)$, è definito sempre come media aritmetica ponderata dei valori assumibili dalla v.c.. La ponderazione è data dalla funzione di densità di probabilità e la somma viene interpretata in senso continuo, cioè come un integrale. Come nel caso discreto, viene richiesto che l'integrale converga assolutamente. Quindi nel caso continuo

$$E(X) = \int_{-\infty}^{+\infty} x p_X(x) dx \quad \text{purché} \quad \int_{-\infty}^{+\infty} |x| p_X(x) dx < +\infty.$$

L'indice di variabilità principale, la varianza di X , è sempre il valore atteso del quadrato degli scarti dal valore atteso di X , quindi

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

La varianza è di solito più facilmente calcolata con la formula per il calcolo $\text{Var}(X) = E(X^2) - (E(X))^2$, come sarà dimostrato nel paragrafo 15.4.

8.3 Due leggi univariate notevoli di tipo continuo

Conviene iniziare a familiarizzare con le leggi di v.c. univariate di tipo continuo considerando due esempi notevoli, importanti anche per le applicazioni, le leggi uniformi continue e le leggi esponenziali.

Le leggi uniformi continue

Si dice che X ha legge uniforme continua in (a, b) , dove $a < b$, e si scrive in breve $X \sim U(a, b)$, se la f.d.p. di X è

$$p_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in (a, b) \\ 0 & \text{se } x \notin (a, b). \end{cases}$$

Si verifica facilmente che $p_X(x)$ è una f.d.p.. Anzitutto, la condizione di non negatività è soddisfatta in modo ovvio. Per indagare la normalizzazione, si calcoli

$$\int_{-\infty}^{+\infty} p_X(x) dx = \int_a^b p_X(x) dx = \int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} [x]_a^b = \frac{b-a}{b-a} = 1.$$

Se $X \sim U(a, b)$, valore atteso e varianza di X sono

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

Infatti

$$E(X) = \int_{-\infty}^{+\infty} x p_X(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{1}{2} x^2 \right]_a^b = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{a+b}{2}$$

e

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 p_X(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{1}{3} x^3 \right]_a^b \\ &= \frac{1}{3} \frac{b^3 - a^3}{b-a} = \frac{a^2 + ab + b^2}{3} \end{aligned}$$

da cui

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{a^2 + b^2 - 2ab}{12} = \frac{(b-a)^2}{12}.$$

Le leggi esponenziali

Le leggi esponenziali sono strettamente collegate alla funzione esponenziale, $e^x = \exp(x)$ (due notazioni equivalenti), ben nota dall'Analisi. Conviene fare anzitutto alcuni veloci richiami sulla funzione esponenziale.

- Definizione:

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n, \quad x \in \mathbb{R}. \quad (8.2)$$

- Proprietà delle potenze:

$$\begin{aligned} e^0 &= 1 \\ e^{a+b} &= e^a e^b \\ (e^a)^b &= e^{ab} \end{aligned}$$

per ogni $a, b \in \mathbb{R}$.

- Proprietà analitiche:

$$\begin{aligned} e^x &> 0 \text{ per ogni } x \in \mathbb{R} \\ \frac{d}{dx} e^x &= e^x \\ e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \dots = \sum_{i=0}^{+\infty} \frac{x^i}{i!}. \end{aligned} \quad (8.3)$$

La formula (8.3) è lo **sviluppo in serie di potenze** di e^x . Si dimostra che $\lim_{n \rightarrow +\infty} \sum_{i=0}^n \frac{x^i}{i!} = e^x$, ossia che la serie converge a e^x per ogni $x \in \mathbb{R}$. Si è ora pronti ad esaminare la definizione delle leggi esponenziali. Si dice che X ha legge esponenziale con parametro $\lambda > 0$, e si scrive in breve $X \sim \text{Esp}(\lambda)$, se la f.d.p. di X è

$$p_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0. \end{cases}$$

Si verifica facilmente che $p_X(x)$ è una f.d.p.. La condizione di non negatività è soddisfatta perché $\lambda > 0$ e $e^{-\lambda x} > 0$. Per verificare la normalizzazione, si ha

$$\begin{aligned} \int_{-\infty}^{+\infty} p_X(x) dx &= \int_0^{+\infty} p_X(x) dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^{+\infty} = \\ &= -\lim_{x \rightarrow +\infty} e^{-\lambda x} - (-e^0) = 1. \end{aligned}$$

Alcuni calcoli di probabilità esponenziali:

$$x > 0 \implies P(X > x) = \int_x^{+\infty} p_X(t) dt = \int_x^{+\infty} \lambda e^{-\lambda t} dx = [-e^{-\lambda t}]_x^{+\infty} = e^{-\lambda x}$$

(la funzione $P(X > x)$, $x \in \mathbb{R}$, è detta funzione di sopravvivenza)

$$x > 0 \implies P(X \leq x) = 1 - P(X > x) = 1 - e^{-\lambda x}$$

(la funzione $P(X \leq x)$, $x \in \mathbb{R}$, è detta funzione di ripartizione)

$$P(X = x) = 0 \quad \text{per ogni } x \in \mathbb{R}.$$

Il valore atteso di $X \sim \text{Esp}(\lambda)$ è $E(X) = 1/\lambda$. Infatti

$$E(X) = \int_{-\infty}^{+\infty} x p_X(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{+\infty} \lambda x e^{-\lambda x} d(\lambda x),$$

dove l'ultimo integrale vale 1. Infatti il cambiamento di variabile $\lambda x = t$ porta a un integrale facilmente calcolabile per parti:

$$\int_0^{+\infty} t e^{-t} dt = [-t e^{-t}]_0^{+\infty} + \int_0^{+\infty} e^{-t} dt = 0 + 1 = 1.$$

Poiché $P(X > 0) = 1$ per ogni $\lambda > 0$, una v.c. con legge esponenziale è adatta a modellare fenomeni che si manifestano nel continuo su valori positivi non limitati superiormente. È per esempio la modellazione di *default* per tempi d'attesa (non misurati nel tempo discreto). Il risultato sul valore atteso dà un significato al parametro λ . Per un tempo d'attesa esponenziale $1/\lambda$ è il valor medio dell'attesa. Quindi

$$\lambda = \frac{1}{\text{media dell'attesa}} = \frac{1}{E(X)}.$$

Esempio 8.1 Si supponga che il tempo di corretto funzionamento di un componente abbia legge esponenziale con valore atteso 5 anni. Si calcoli $P(1 < X < 3)$. Poiché $\lambda = 1/E(X) = 1/5$, si ha $X \sim \text{Esp}(1/5)$. Quindi

$$\begin{aligned} P(1 < X < 3) &= P(X < 3) - P(X \leq 1) = P(X \leq 3) - P(X \leq 1) = \\ &= \left\{ 1 - \exp\left(-\frac{1}{5} 3\right) \right\} - \left\{ 1 - \exp\left(-\frac{1}{5} 1\right) \right\} = \\ &= e^{-1/5} - e^{-3/5} \doteq 0.2699. \end{aligned}$$

◇

8.4 Il supporto di una v.c. univariata con legge continua

Il supporto S_X di una v.c. univariata X con legge continua è il più piccolo sottoinsieme chiuso di \mathbb{R} al quale P_X dà probabilità 1. Si ricorda che un insieme è chiuso se contiene tutti i suoi punti di accumulazione. Ad esempio, $[0, 1]$ è chiuso ma $[0, 1)$ non è chiuso: 1 è un punto di accumulazione dell'insieme che non appartiene all'insieme.

Quindi $S_X \in \mathcal{B}_1$ è tale che:

- i) S_X è chiuso
- ii) $S_X \subseteq C$ per ogni $C \subseteq \mathbb{R}$ chiuso con $P_X(C) = P(X \in C) = 1$.

In pratica, S_X è la chiusura dell'insieme $\{x \in \mathbb{R} : p_X(x) > 0\}$. (La chiusura di un insieme numerico è l'unione fra l'insieme stesso e i punti non dell'insieme che sono però suoi punti di accumulazione).

È facile controllare che

- se $X \sim U(a, b)$, con $a < b$, $S_X = [a, b]$
- se $X \sim \text{Esp}(\lambda)$, con $\lambda > 0$, $S_X = [0, +\infty)$.

8.5 V.c. bivariate e multivariate con legge continua

La v.c. bivariata (X, Y) ha legge continua se per ogni $B \in \mathcal{B}_2$, in particolare per i rettangoli $B = [a_1, b_1] \times [a_2, b_2]$,

$$\begin{aligned} P_{(X,Y)}(B) &= P((X, Y) \in B) = \\ &= P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \\ &= \iint_B p_{X,Y}(x, y) dx dy, \end{aligned}$$

dove $p_{X,Y}(x, y)$, la funzione di densità di probabilità congiunta, ha le proprietà

- i) $p_{X,Y}(x, y) \geq 0$ per ogni $(x, y) \in \mathbb{R}^2$ (non negatività)
- ii) $\iint_{\mathbb{R}^2} p_{X,Y}(x, y) dx dy = 1$ (normalizzazione).

Gli integrali doppi sui rettangoli vanno intesi come integrali semplici iterati:

$$\iint_{[a_1, b_1] \times [a_2, b_2]} p_{X,Y}(x, y) dx dy = \int_{a_1}^{b_1} \left\{ \int_{a_2}^{b_2} p_{X,Y}(x, y) dy \right\} dx.$$

Si dimostra infatti che

$$\int_{a_1}^{b_1} \left\{ \int_{a_2}^{b_2} p_{X,Y}(x, y) dy \right\} dx = \int_{a_2}^{b_2} \left\{ \int_{a_1}^{b_1} p_{X,Y}(x, y) dx \right\} dy$$

e il loro comune valore definisce $\iint_{[a_1, b_1] \times [a_2, b_2]} p_{X,Y}(x, y) dx dy$.

Anche per una v.c. bivariata con legge continua vale che $P(X = x, Y = y) = 0$ per ogni punto $(x, y) \in \mathbb{R}^2$. Inoltre $S_{X,Y}$ è la chiusura dell'insieme $\{(x, y) \in \mathbb{R}^2 : p_{X,Y}(x, y) > 0\}$.

Da (X, Y) con legge continua si possono ottenere leggi univariate indotte, che saranno ancora di tipo continuo. In particolare, le leggi

- marginale X , con f.d.p. $p_X(x) = \int_{-\infty}^{+\infty} p_{X,Y}(x, y) dy$
- marginale Y , con f.d.p. $p_Y(y) = \int_{-\infty}^{+\infty} p_{X,Y}(x, y) dx$
- condizionali $Y|X = x$, $x \in S_X$, con f.d.p. $p_{Y|X=x}(y) = \frac{p_{X,Y}(x, y)}{p_X(x)}$
supposto $p_X(x) > 0$ analogamente al caso discreto
- condizionali $X|Y = y$, $y \in S_Y$, con f.d.p. $p_{X|Y=y}(x) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$
supposto $p_Y(y) > 0$ analogamente al caso discreto.

I supporti di tali leggi si ottengono dal supporto congiunto secondo le stesse formule viste nel caso discreto, ovvero

$$\begin{aligned} S_X &= \{x \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ per qualche } y\} \\ S_Y &= \{y \in \mathbb{R} : (x, y) \in S_{X,Y} \text{ per qualche } x\} \\ S_{Y|X=x} &= \{y \in \mathbb{R} : (x, y) \in S_{X,Y}\} \\ S_{X|Y=y} &= \{x \in \mathbb{R} : (x, y) \in S_{X,Y}\}. \end{aligned}$$

Viceversa, si può definire una v.c. bivariata con legge continua (X, Y) tramite specificazione gerarchica. Ad esempio,

- si prescrive X con f.d.p. $p_X(x)$ e supporto S_X

- per ogni $x \in S_X$ si prescrive la legge condizionale $Y|X = x$ con f.d.p. $p_{Y|X=x}(y)$ e supporto $S_{Y|X=x}$.

Allora, con le stesse formule del caso discreto,

$$S_{X,Y} = \{(x, y) \in \mathbb{R}^2 : x \in S_X, y \in S_{Y|X=x}\}$$

e

$$p_{X,Y}(x, y) = p_X(x) p_{Y|X=x}(y).$$

Esempio 8.2 Si supponga che $X \sim U(0, 1)$ e che per ogni $x \geq 0$ si abbia $Y|X = x \sim Esp(x + 1)$. Il supporto congiunto di (X, Y) è

$$S_{X,Y} = \{(x, y) \in \mathbb{R}^2 : x \in S_X, y \in S_{Y|X=x}\} = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, y \geq 0\}.$$

Inoltre,

$$p_{X,Y}(x, y) = p_X(x) p_{Y|X=x}(y) = \begin{cases} (x+1)e^{-(x+1)y} & \text{se } (x, y) \in S_{X,Y} \\ 0 & \text{altrimenti.} \end{cases}$$

◇

Una v.c. multivariata $X = (X_1, \dots, X_d)$ ha legge P_X di tipo continuo se esiste una funzione non negativa

$$p_X(x) = p_{X_1, \dots, X_d}(x_1, \dots, x_d),$$

normalizzata, ossia tale che $\int_{\mathbb{R}^d} p_X(x) dx = 1$, per cui per ogni $B \in \mathcal{B}_d$, in particolare per i ‘rettangoli’ di forma $B = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d]$, valga

$$P_X(B) = P(X \in B) = \int_B p_X(x) dx.$$

Gli integrali multipli sui rettangoli di \mathbb{R}^d si calcolano, in analogia con gli integrali doppi, come integrali semplici iterati. Quindi ad esempio, con $d = 3$, $B = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$, si ha

$$\begin{aligned} \int_B p_X(x) dx &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \int_{a_3}^{b_3} p_X(x) dx \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \int_{a_3}^{b_3} p_{X_1, X_2, X_3}(x_1, x_2, x_3) dx_1 dx_2 dx_3 \\ &= \int_{a_1}^{b_1} \left\{ \int_{a_2}^{b_2} \left\{ \int_{a_3}^{b_3} p_{X_1, X_2, X_3}(x_1, x_2, x_3) dx_3 \right\} dx_2 \right\} dx_1. \end{aligned}$$

8.6 V.c. bivariate e multivariate con componenti indipendenti

Una v.c. biviata (X, Y) con legge continua ha componenti indipendenti se per ogni $(x, y) \in \mathbb{R}^2$ vale

$$p_{X,Y}(x, y) = p_X(x) p_Y(y).$$

Vale allora anche che $S_{X,Y} = S_X \times S_Y$. Inoltre si ha

$$\begin{aligned} P_{X,Y}(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} p_X(x) p_Y(y) dx dy \\ &= \int_{a_1}^{b_1} p_X(x) dx \int_{a_2}^{b_2} p_Y(y) dy \\ &= P(a_1 \leq X \leq b_1) P(a_2 \leq Y \leq b_2). \end{aligned}$$

Analogamente, una v.c. multivariata $X = (X_1, \dots, X_d)$ ha componenti indipendenti se per ogni $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ vale

$$p_X(x) = p_{X_1, \dots, X_d}(x_1, \dots, x_d) = p_{X_1}(x_1) \cdots p_{X_d}(x_d),$$

per cui

$$S_X = S_{X_1} \times \cdots \times S_{X_d}$$

e

$$\begin{aligned} P(a_1 \leq X_1 \leq b_1, \dots, a_d \leq X_d \leq b_d) &= \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} p_X(x) dx \\ &= \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} \prod_{i=1}^d p_{X_i}(x_i) dx_i \\ &= \prod_{i=1}^d \int_{a_i}^{b_i} p_{X_i}(x_i) dx_i \\ &= \prod_{i=1}^d P(a_i \leq X_i \leq b_i). \end{aligned}$$

In pratica quindi una v.c. multivariata ha componenti indipendenti quando eventi congiunti che dipendono separatamente dalle singole componenti sono eventi indipendenti.

Spesso si assume per semplicità di modellazione che le componenti di una v.c. multivariata siano indipendenti. Grande importanza in Statistica hanno le v.c. con componenti indipendenti e identicamente distribuite. Queste modellano osservazioni ripetute e indipendenti dello stesso fenomeno. Si pensi ad un fenomeno descritto in ogni singola occasione o unità da una v.c. univariata con legge

$Bi(1, p)$. Se questa è osservata in un certo numero di n occasioni indipendenti, o su un certo numero n di unità indipendenti, le osservazioni saranno realizzazione $y = (y_1, \dots, y_n)$ di una v.c. multivariata $Y = (Y_1, \dots, Y_n)$ con componenti indipendenti e identicamente distribuite, $Y_i \sim Bi(1, p)$, $i = 1, \dots, n$.

8.7 Esercizi

Esercizio 8.1 Sia $X \sim U(0, 1)$. Si calcolino $P(X > 0)$, $P(X < 1)$, $P(X > 1/2)$, $P(1/4 < X < 3/4)$, $E(X)$.

Esercizio 8.2 La durata X di corretto funzionamento di un certo componente ha legge esponenziale con valore atteso 6 anni. Si calcolino $P(X > 6)$, $P(X > 12)$, $P(X > 18)$.

Esercizio 8.3 Sia X una v.c. continua con densità di forma $p_X(x) = c(1 - x)$, se $x \in [0, 1]$, e zero altrove. Si determini il valore della costante c di modo che $p_X(x)$ sia una funzione di densità di probabilità. Si calcolino $P(0 \leq X \leq 2)$, $P(-1 \leq X \leq 0.5)$ e $P(X = 0.5)$. Infine, si ottengano valore atteso e varianza di X .

Esercizio 8.4 Sia X una v.c. continua con densità di forma $p_X(x) = ce^x$, se $x \in (-\infty, 0]$, e zero altrove. Si determini il valore della costante c di modo che $p_X(x)$ sia effettivamente una funzione di densità di probabilità. Si calcolino $P(0 \leq X \leq 2)$, $P(-1 \leq X \leq 0.5)$ e $P(X = -0.5)$. Infine, si ottenga il valore atteso di X .

Esercizio 8.5 Sia X una variabile casuale continua con densità di forma

$$p_X(x) = \begin{cases} c + x & \text{se } x \in [-1, 0] \\ c - x & \text{se } x \in (0, 1] \\ 0 & \text{altrove.} \end{cases}$$

Si determini il valore della costante c di modo che $p_X(x)$ sia effettivamente una funzione di densità di probabilità. Si calcolino $P(-0.5 \leq X \leq 0)$, $P(|X| \leq 0.5)$ e $P(X = 0)$. Infine, si ottenga il valore atteso di X .

Esercizio 8.6 Sia X una variabile casuale continua con supporto $S_X = [0, 2]$ e, per $x \in S_X$, densità di forma

$$p_X(x) = \begin{cases} cx & \text{se } x \in [0, 1) \\ 2 - cx & \text{se } x \in [1, 2]. \end{cases}$$

Si determini il valore della costante c di modo che $p_X(x)$ sia effettivamente una funzione di densità di probabilità. Si calcolino $P(0 \leq X \leq 0.5)$ e $P(X = 0.5)$. Infine, si ottengano valore atteso e varianza di X .

Esercizio 8.7 Sia X una variabile casuale continua con densità di forma

$$p_X(x) = \begin{cases} c(4 - x^2) & \text{se } x \in [0, 2] \\ 0 & \text{altrove.} \end{cases}$$

Si determini il valore della costante c di modo che $p_X(x)$ sia effettivamente una funzione di densità di probabilità. Si calcolino $P(X > 1)$, $P(X > 1|X > 0.5)$. Infine, si ottengano valore atteso e varianza di X .

Esercizio 8.8 Sia X una variabile casuale bivariata continua con densità di forma

$$p_{X_1, X_2}(x_1, x_2) = \begin{cases} e^{-x_1 - x_2} & \text{se } x_1, x_2 \geq 0 \\ 0 & \text{altrove.} \end{cases}$$

Si stabilisca se (X_1, X_2) ha componenti indipendenti. Si calcolino $P(X_2 > 1)$, $P(X_2 > 1|X_1 = 0.5)$.

8.8 Appendice: leggi uniformi continue ed esponenziali con R

Il comando `dunif(x,min=a, max=b)` calcola in x il valore della funzione di densità di $X \sim U(a, b)$. Il comando `runif(n,min=a,max=b)` simula la generazione di n realizzazioni indipendenti da $U(a, b)$, ossia una realizzazione di una v.c. n -variata con componenti indipendenti e leggi marginali univariate $U(a, b)$.

```
> rm(list=ls())
> set.seed(12345)
> x=seq(-1,2,0.001)
> plot(x,dunif(x,min=0,max=1),type="l")
> y=runif(10^4,min=0,max=1)
> y[1:5]
[1] 0.7209039 0.8757732 0.7609823 0.8861246 0.4564810
> mean(y)
[1] 0.5006812
> var(y)
[1] 0.08240007
> yy=runif(10^4,min=0,max=1)
> yy[1:5]
[1] 0.2443204 0.6894012 0.8696410 0.9812336 0.5692775
> mean(yy)
[1] 0.4966774
> var(yy)
[1] 0.08341291
```

Da notare che i valori di `yy` sono diversi da quelli di `y` perché non si è ripetuto il comando `set.seed(12345)`. Medie e varianze empiriche dei due campioni sono assai vicini a valore atteso e varianza di X , pari rispettivamente a $1/2$ e $1/12 \doteq 0.08333$.

Se $X \sim \text{Esp}(\lambda)$ i comandi per densità e generazione sono `dexp(x,rate=lambda)` e `rexp(n,rate=lambda)`. Si dà sotto un esempio del loro uso.

```
> x=seq(-1,8,0.001)
> plot(x,dexp(x,rate=1),type="l")
> lines(x,dexp(x,rate=2),type="l",col="2")
> lines(x,dexp(x,rate=1/2),type="l",col="3")
> y=rexp(10^4,rate=1/2)
> y[1:5]
[1] 0.2562503 0.1346927 0.2227418 0.2292324 2.3876483
> mean(y)
[1] 1.993117
> var(y)
[1] 3.869911
```

Cosa si congettura per $\text{Var}(X)$? Si provi anche il comando `hist(y)`.

Lezione 9

La funzione di ripartizione

9.1 Definizione

In questa lezione ci si occuperà principalmente di funzioni di ripartizione (f.r.) di v.c. univariate. Poco costa tuttavia dare la definizione generale, per una v.c. multivariata.

Definizione 9.1 Si dice **funzione di ripartizione** di $X = (X_1, \dots, X_d)$ la funzione

$$F_X : \mathbb{R}^d \rightarrow [0, 1]$$

che a ciascun punto di \mathbb{R}^d , $x = (x_1, \dots, x_d)$, fa corrispondere il valore d'immagine

$$\begin{aligned} F_X(x) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d) \\ &= P_X((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d]) \\ &= P\left(\bigcap_{i=1}^d \{s \in S : X_i(s) \leq x_i\}\right). \end{aligned}$$

Osservazione: a ogni punto $x \in \mathbb{R}^d$ corrisponde un angoloide con vertice x , A_x , definito da $A_x = (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_d]$, e la funzione di ripartizione è $F_X(x) = P(X \in A_x)$. Per $d = 1$ si ha $F_X(x) = P(X \leq x)$, per $d = 2$ si ha $F_{X_1, X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$.

Risultato 9.1 Se X ha componenti indipendenti, per ogni $x \in \mathbb{R}^d$ vale la relazione

$$F_X(x) = F_{X_1, \dots, X_d}(x_1, \dots, x_d) = \prod_{i=1}^d F_{X_i}(x_i). \quad (9.1)$$

La relazione (9.1) riduce i calcoli di probabilità su v.c. multivariate a calcoli sulle singole componenti univariate. Ad esempio, $P(X_1 > x_1, \dots, X_d > x_d) = \prod_{i=1}^d P(X_i > x_i)$.

9.2 Caso univariato: calcolo di probabilità di intervalli

Se X è una v.c. univariata, la sua f.r. $F_X(x)$ permette di calcolare agevolmente le probabilità degli intervalli, e quindi di ogni evento definito a partire dagli intervalli con operazioni che non portano fuori dalla classe degli eventi.

Risultato 9.2 Se $a < b$, $P(a < X \leq b) = P_X((a, b]) = F_X(b) - F_X(a)$.

Dimostrazione. Poiché

$$(a, b] = (-\infty, b] \setminus (-\infty, a]$$

e

$$(-\infty, a] \subseteq (-\infty, b]$$

per il Corollario 3.1 si ha

$$\begin{aligned} P_X((a, b]) &= P_X((-\infty, b] \setminus (-\infty, a]) \\ &= P_X((-\infty, b]) - P_X((-\infty, a]) \\ &= F_X(b) - F_X(a). \end{aligned}$$

□

Risultato 9.3 Per ogni $x \in \mathbb{R}$, $P(X = x) = F_X(x) - \lim_{\varepsilon \rightarrow 0^+} F_X(x - \varepsilon)$.

Dimostrazione. Se A_n , $n \in \mathbb{N}$, è una successione di eventi monotona rispetto all'inclusione, si dimostra dagli assiomi che $P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$. Da $\{X = x\} = \lim_{\varepsilon \rightarrow 0^+} \{x - \varepsilon < X \leq x\}$ si ha allora

$$P(X = x) = \lim_{\varepsilon \rightarrow 0^+} P(x - \varepsilon < X \leq x) = \lim_{\varepsilon \rightarrow 0^+} \{F_X(x) - F_X(x - \varepsilon)\}$$

da cui segue la tesi. □

Osservazione 1: se X ha legge continua, per ogni punto x si ha $P(X = x) = 0$, e dunque $F_X(x) - \lim_{\varepsilon \rightarrow 0^+} F_X(x - \varepsilon) = 0$, ossia $F_X(x) = \lim_{\varepsilon \rightarrow 0^+} F_X(x - \varepsilon)$, continuità da sinistra di $F_X(\cdot)$.

Osservazione 2: se X ha legge discreta e $x \in S_X$, si avrà un salto della f.r. in x : $F_X(x) > \lim_{\varepsilon \rightarrow 0^+} F_X(x - \varepsilon)$. L'entità del salto è

$$F_X(x) - \lim_{\varepsilon \rightarrow 0^+} F_X(x - \varepsilon) = P(X = x) = p_X(x).$$

Ulteriori calcoli immediati:

$$\begin{aligned} P(a \leq X \leq b) &= P(X = a) + P(a < X \leq b) = F_X(b) - F_X(a) + P(X = a) \\ P(a < X < b) &= P(a < X \leq b) - P(X = b) = F_X(b) - F_X(a) - P(X = b) \\ P(X < x) &= P(X \leq x) - P(X = x) = F_X(x) - P(X = x) \\ P(X > x) &= 1 - P(X \leq x) = 1 - F_X(x) \text{ (funz. di sopravvivenza di } X) \\ P(X \geq x) &= P(X = x) + P(X > x) = 1 - F_X(x) + P(X = x). \end{aligned}$$

9.3 Caso univariato: dalla $p_X(x)$ alla $F_X(x)$

Sia X una v.c. univariata. Si ha subito

$$F_X(x) = P(X \leq x) = \begin{cases} \sum_{t \in S_X : t \leq x} p_X(t) & \text{se } X \text{ ha legge discreta} \\ \int_{-\infty}^x p_X(t) dt & \text{se } X \text{ ha legge continua.} \end{cases} \quad (9.2)$$

Alcuni esempi:

- $X \sim \mathcal{D}(x_0)$, $x_0 \in \mathbb{R}$, ha f.r.

$$F_X(x) = \begin{cases} 0 & \text{se } x < x_0 \\ 1 & \text{se } x \geq x_0; \end{cases}$$

- $X \sim Bi(1, p)$, $p \in (0, 1)$, ha f.r.

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - p & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1; \end{cases}$$

- $X \sim Esp(\lambda)$, $\lambda > 0$, ha f.r.

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-\lambda x} & \text{se } x \geq 0. \end{cases}$$

Dalla (9.2) e dagli esempi si conclude che:

- se X ha legge discreta univariata, la $F_X(x)$ è costante a tratti, con punti di salto posizionati ai punti del supporto, e il valore del salto è pari alla massa di probabilità posta sul punto;
- se X ha legge continua, la $F_X(x)$, in quanto primitiva di una funzione integrabile, è una funzione continua di $x \in \mathbb{R}$.

9.4 Caso univariato discreto: dalla f.r. alla f.m.p.

Se è data $F_X(x)$, f.r. di una v.c. discreta X , quindi funzione costante a tratti, si recupera il supporto di X come insieme dei punti di salto di $F_X(x)$:

$$S_X = \{x \in \mathbb{R} : F_X(x) - \lim_{\varepsilon \rightarrow 0^+} F_X(x - \varepsilon) > 0\}.$$

Si avrà inoltre, per $x \in S_X$, che la f.m.p. è

$$p_X(x) = P(X = x) = \lim_{\varepsilon \rightarrow 0^+} P(x - \varepsilon < X \leq x) = F_X(x) - \lim_{\varepsilon \rightarrow 0^+} F_X(x - \varepsilon),$$

entità del salto della f.r. nell'intorno sinistro di x .

9.5 Caso univariato continuo: dalla f.r. alla f.d.p.

Nei punti x in cui $p_X(x)$ è continua la f.r. $F_X(x)$ è derivabile e vale

$$p_X(x) = \frac{d}{dx} F_X(x). \quad (9.3)$$

Infatti, la derivata è il limite del rapporto incrementale, per cui

$$\begin{aligned} \frac{d}{dx} F_X(x) &= \lim_{\varepsilon \rightarrow 0^+} \frac{F_X(x + \varepsilon) - F_X(x - \varepsilon)}{2\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{\int_{x-\varepsilon}^{x+\varepsilon} p_X(t) dt}{2\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{2\varepsilon p_X(\tilde{x}_\varepsilon)}{2\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0^+} p_X(\tilde{x}_\varepsilon) \\ &= p_X(x). \end{aligned}$$

Per giustificare gli ultimi passaggi si richiama che, per il teorema del valor medio dell'integrale, $\int_{x-\varepsilon}^{x+\varepsilon} p_X(t) dt = 2\varepsilon p_X(\tilde{x}_\varepsilon)$ con $\tilde{x}_\varepsilon \in (x - \varepsilon, x + \varepsilon)$. Quindi $\lim_{\varepsilon \rightarrow 0^+} p_X(\tilde{x}_\varepsilon) = p_X(x)$ nei punti x in cui la funzione $p_X(x)$ è continua.

9.6 Esercizi risolti

Esempio 9.1 Sia X una variabile casuale con supporto $S_X = [-1, 0]$ e funzione di densità di probabilità di forma $p_X(x) = cx$ per $x \in S_X$ e 0 altrove. Si completi la definizione della funzione di densità di X , determinando il valore della costante di normalizzazione c . Si calcoli la funzione di ripartizione di X , esplicitandola in tutti i suoi tratti. Si ottenga il valore atteso e la varianza di X .

La costante di normalizzazione è determinata dalla condizione di normalizzazione:

$$1 = \int_{-\infty}^{+\infty} p_X(x) dx = \int_{-1}^0 p_X(x) dx = \int_{-1}^0 cx dx = c \left[\frac{x^2}{2} \right]_{-1}^0 = -\frac{c}{2}$$

per cui $c = -2$. Quindi

$$p_X(x) = \begin{cases} -2x & \text{se } x \in [-1, 0] \\ 0 & \text{se } x \notin [-1, 0]. \end{cases}$$

Poiché il supporto è un intervallo, la f.r. di X richiede lo studio di tre casi:

i) $x < -1$:

$$F_X(x) = \int_{-\infty}^x p_X(t) dt = \int_{-\infty}^x 0 dt = 0;$$

ii) $-1 \leq x \leq 0$:

$$F_X(x) = \int_{-\infty}^x p_X(t) dt = \int_{-\infty}^{-1} 0 dt + \int_{-1}^x (-2t) dt = 0 - [t^2]_{-1}^x = 1 - x^2;$$

iii) $x > 0$:

$$F_X(x) = \int_{-\infty}^x p_X(t) dt = \int_{-\infty}^{-1} 0 dt + \int_{-1}^0 (-2t) dt + \int_0^x 0 dt = 0 + 1 + 0 = 1.$$

Si ha dunque

$$F_X(x) = \begin{cases} 0 & \text{se } x \in (-\infty, -1) \\ 1 - x^2 & \text{se } x \in [-1, 0] \\ 1 & \text{se } x > 0. \end{cases}$$

Il valore atteso di X è

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x p_X(x) dx = \int_{-1}^0 x(-2x) dx \\ &= -2 \int_{-1}^0 x^2 dx = -2 \left[\frac{x^3}{3} \right]_{-1}^0 \\ &= -2 \left(\frac{0^3}{3} - \frac{(-1)^3}{3} \right) = -\frac{2}{3}. \end{aligned}$$

Analogamente si calcola $E(X^2)$:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 p_X(x) dx = \int_{-1}^0 x^2(-2x) dx \\ &= -2 \int_{-1}^0 x^3 dx = -2 \left[\frac{x^4}{4} \right]_{-1}^0 \\ &= -2 \left(\frac{0^4}{4} - \frac{(-1)^4}{4} \right) = \frac{1}{2}. \end{aligned}$$

In conclusione la varianza di X risulta

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{1}{2} - \left(-\frac{2}{3} \right)^2 = \frac{1}{18}.$$

◇

Esempio 9.2 Sia X una variabile casuale con supporto $S_X = [0, 1]$ e funzione di densità di probabilità di forma $p_X(x) = ce^{-x}$ per $x \in S_X$ e 0 altrove. Si completi la definizione della funzione di densità di X , determinando il valore della costante di normalizzazione c . Si calcoli la funzione di ripartizione di X , esplicitandola in tutti i suoi tratti.

La costante di normalizzazione è determinata dalla condizione di normalizzazione:

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} p_X(x) dx = \int_0^1 p_X(x) dx = \int_0^1 c e^{-x} dx = c [-e^{-x}]_0^1 \\ &= c(-e^{-1} + e^0) = c(1 - e^{-1}) \end{aligned}$$

per cui $c = 1/(1 - e^{-1})$. Quindi

$$p_X(x) = \begin{cases} \frac{e^{-x}}{1 - e^{-1}} & \text{se } x \in [0, 1] \\ 0 & \text{se } x \notin [0, 1]. \end{cases}$$

Poiché il supporto è un intervallo, la f.r. di X richiede lo studio di tre casi:

i) $x < 0$:

$$F_X(x) = \int_{-\infty}^x p_X(t) dt = \int_{-\infty}^x 0 dt = 0;$$

ii) $0 \leq x \leq 1$:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x p_X(t) dt = \int_{-\infty}^0 0 dt + \int_0^x \frac{e^{-t}}{1 - e^{-1}} dt \\ &= 0 + \frac{1}{1 - e^{-1}} [-e^{-t}]_0^x = \frac{1 - e^{-x}}{1 - e^{-1}}; \end{aligned}$$

iii) $x > 1$:

$$F_X(x) = \int_{-\infty}^x p_X(t) dt = \int_{-\infty}^0 0 dt + \int_0^1 p_X(t) dt + \int_1^x 0 dt = 0 + 1 + 0 = 1,$$

valendo per qualunque densità che $\int_{S_X} p_X(t) dt = 1$. Si ha dunque

$$F_X(x) = \begin{cases} 0 & \text{se } x \in (-\infty, 0) \\ \frac{1 - e^{-x}}{1 - e^{-1}} & \text{se } x \in [0, 1] \\ 1 & \text{se } x > 1. \end{cases}$$

◇

9.7 Esercizi

Esercizio 9.1 Sia X una v.c. univariata con f.r.

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ x^2 & \text{se } 0 \leq x \leq 1 \\ 1 & \text{se } x > 1. \end{cases}$$

Si reperiscano la funzione di densità di probabilità di X e il supporto di X . Si calcolino $P(X \leq 1/3)$, $P(X = 1/3)$, $P(1/3 \leq X \leq 1/2)$, $P(X > 1/2 \mid X > 1/3)$.

Esercizio 9.2 Sia X una v.c. univariata con f.r.

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \sin(x) & \text{se } 0 \leq x \leq \frac{\pi}{2} \\ 1 & \text{se } x > \frac{\pi}{2}. \end{cases}$$

Si reperiscano la funzione di densità di probabilità di X e il supporto di X . Si calcolino $P(X \leq 0)$, $P(X > 1)$, $P(X > 1/2 \mid X > 0)$.

Esercizio 9.3 Sia X una v.c. univariata con f.r.

$$F_X(x) = \begin{cases} 0 & \text{se } x < -1 \\ x + \frac{1}{2}x^2 + \frac{1}{2} & \text{se } -1 \leq x \leq 0 \\ x - \frac{1}{2}x^2 + \frac{1}{2} & \text{se } 0 < x \leq 1 \\ 1 & \text{se } x > 1. \end{cases}$$

Si reperiscano la funzione di densità di probabilità di X e il supporto di X . Si calcolino $P(X \leq 0)$, $P(X = -1/3)$, $P(-1/3 \leq X \leq 1/3)$, $P(X > 1/2 \mid X > 0)$.

Esercizio 9.4 Sia X una variabile casuale con supporto $S_X = [-1, 1]$ e funzione di densità di probabilità di forma $p_X(x) = 0.5 + cx^2$ per $x \in S_X$ e 0 altrove. Si completi la definizione della funzione di densità di X , determinando il valore della costante di normalizzazione c . Si calcoli la funzione di ripartizione di X , esplicitandola in tutti i suoi tratti. Si ottenga il valore atteso di X . Si calcoli $P(X = 0)$.

Esercizio 9.5 Sia X una variabile casuale con supporto $S_X = [0, 1]$ e funzione di densità di probabilità di forma $p_X(x) = cx$ per $x \in S_X$ e 0 altrove. Si completi la definizione della funzione di densità di X , determinando il valore della costante di normalizzazione c . Si calcoli la funzione di ripartizione di X , esplicitandola in tutti i suoi tratti. Si calcoli $P(0.4 < X < 0.6)$. Si ottengano valore atteso e la varianza di X .

Esercizio 9.6 Sia X una variabile casuale con supporto $S_X = [-1, 1]$ e funzione di densità di probabilità di forma $p_X(x) = c|x|$ per $x \in S_X$ e 0 altrove. Si completi la definizione della funzione di densità di X , determinando il valore della costante di normalizzazione c . Si calcoli la funzione di ripartizione di X , esplicitandola in tutti i suoi tratti. Si calcoli poi $P(X > 0.5)$. Si ottenga il valore atteso di X .

Esercizio 9.7 Sia X una variabile casuale con supporto $S_X = [0, 1]$ e funzione di densità di probabilità di forma $p_X(x) = ce^x$ per $x \in S_X$ e 0 altrove. Si completi la definizione della funzione di densità di X , determinando il valore della costante di normalizzazione c . Si calcoli la funzione di ripartizione di X , esplicitandola in tutti i suoi tratti. Si ottenga $P(X < 2)$.

Esercizio 9.8 Sia X una variabile casuale con supporto $S_X = [-1, 1]$ e funzione di densità di probabilità di forma $p_X(x) = cx^2$ per $x \in S_X$ e 0 altrove. Si completi la definizione della funzione di densità di X , determinando il valore della costante di normalizzazione c . Si calcoli la funzione di ripartizione di X , esplicitandola in tutti i suoi tratti. Si calcoli poi $P(-0.5 < X < 0.5)$. Si ottenga il valore atteso di X .

Esercizio 9.9 Sia X una v.c. con supporto $S_X = [0, +\infty)$ e f.r. $F_X(x) = 1 - e^{-x}$ per $x \geq 0$, $F_X(x) = 0$ per $x < 0$. Si ottengano supporto e f.r. della legge di $X \mid X > x$, dove $x > 0$. Si ottenga poi la f.d.p. della legge di $X \mid X > x$.

Esercizio 9.10 Sia X una v.c. con legge $Esp(1)$. Si ottengano supporto e f.r. della legge di $X \mid X \leq 1$. Si ottenga poi la f.d.p. di tale legge.

Esercizio 9.11 (*La legge di Benford*)

Si supponga che la cifra più significativa, X , con supporto

$$S_X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\},$$

segua la legge di Benford, quindi abbia f.m.p.

$$p_X(x) = \log_{10} \left(1 + \frac{1}{x} \right) = \log_{10} \left(\frac{1+x}{x} \right), \quad x \in S_X.$$

Si mostri che per $x \in S_X$ la f.r. di X è $F_X(x) = \log_{10}(1+x)$.

9.8 Appendice: alcune funzioni di ripartizione con R

Sia $X \sim Bi(n, p)$. Il comando `pbinom(x,size=n,prob=p)` calcola $F_X(x)$. Esempio d'uso:

```
> n=10 ; p=0.5 ; x=0:n
> fr=pbinom(x, size=n,prob=p)
> fr=round(fr,digits=3)
> tabulazione=rbind(x, fr)
> print(tabulazione)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
x  0.000 1.000 2.000 3.000 4.000 5.000 6.000 7.000 8.000 9.000    10
fr 0.001 0.011 0.055 0.172 0.377 0.623 0.828 0.945 0.989 0.999     1
```

Analogamente le f.r. di $IG(n; N, D)$, $U(a, b)$ e $Esp(\lambda)$ si ottengono con `phyper(x,D,N-D,n)`, `punif(x, min=a, max=b)`, `pexp(x,rate=lambda)`, rispettivamente.

Lezione 10

F.r. univariate: proprietà caratterizzanti

10.1 Teorema sulle proprietà strutturali

Nella lezione precedente si è vista la definizione della funzione di ripartizione di una v.c.. Si è anche iniziato ad usare la f.r. per calcolare probabilità di eventi relativi a una v.c. univariata. In questa lezione si studieranno le proprietà che distinguono le funzioni di ripartizione delle v.c. univariate dalle altre funzioni reali di variabile reale. Si enuncia subito il teorema sulle proprietà che ogni funzione di ripartizione necessariamente ha, che si possono chiamare le proprietà strutturali delle funzioni di ripartizione delle v.c. univariate.

Teorema 10.1 *Sia X una v.c. univariata con legge qualsiasi. La funzione di ripartizione di X , $F_X(x)$, è una applicazione $F_X : \mathbb{R} \rightarrow [0, 1]$ con le seguenti proprietà:*

i) è monotona non decrescente:

$$x_1 < x_2 \implies F_X(x_1) \leq F_X(x_2);$$

ii) è continua da destra in ogni punto $x \in \mathbb{R}$:

$$\text{per ogni } x \in \mathbb{R} \text{ si ha } \lim_{\varepsilon \rightarrow 0^+} F_X(x + \varepsilon) = F_X(x);$$

iii) valgono i limiti:

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

Dimostrazione.

i) $x_1 < x_2$ implica $(-\infty, x_1] \subseteq (-\infty, x_2]$; quindi per la monotonicità della

probabilità (Teorema 3.4) si ha $P_X((-\infty, x_1]) \leq P_X((-\infty, x_2])$ ossia $F_X(x_1) \leq F_X(x_2)$.

ii) $\varepsilon > 0$ implica $(\infty, x] \subseteq (-\infty, x + \varepsilon]$ per ogni $x \in \mathbb{R}$; inoltre

$$\lim_{\varepsilon \rightarrow 0^+} (x, x + \varepsilon] = \emptyset$$

poiché x non è un elemento di $(x, x + \varepsilon]$ e ogni valore maggiore di x viene escluso dall'insieme prendendo ε sufficientemente piccolo. Pertanto

$$\begin{aligned} 0 &= P_X(\emptyset) \\ &= \lim_{\varepsilon \rightarrow 0^+} P_X((x, x + \varepsilon]) \\ &= \lim_{\varepsilon \rightarrow 0^+} P_X((-\infty, x + \varepsilon] \setminus (-\infty, x]) \\ &= \lim_{\varepsilon \rightarrow 0^+} (F_X(x + \varepsilon) - F_X(x)) \end{aligned}$$

da cui la proprietà.

iii) $\emptyset = \lim_{x \rightarrow -\infty} (-\infty, x]$ e $\mathbb{R} = \lim_{x \rightarrow +\infty} (-\infty, x]$, per cui

$$0 = P_X(\emptyset) = \lim_{x \rightarrow -\infty} P_X((-\infty, x]) = \lim_{x \rightarrow -\infty} F_X(x)$$

e

$$1 = P_X(\mathbb{R}) = \lim_{x \rightarrow +\infty} P_X((-\infty, x]) = \lim_{x \rightarrow +\infty} F_X(x).$$

□

10.2 Il teorema di caratterizzazione

Le proprietà i), ii) e iii) del Teorema 10.1 sono godute unicamente dalle funzioni reali di variabile reale che sono funzioni di ripartizione di una qualche legge di probabilità. È questa la tesi del teorema di caratterizzazione, di cui si dà solo l'enunciato.

Teorema 10.2 *Se $F : \mathbb{R} \rightarrow [0, 1]$ ha le proprietà i), ii) e iii) del Teorema 10.1, allora esiste una v.c. univariata X con legge di probabilità P_X di cui F è funzione di ripartizione, ossia tale che*

$$P_X((-\infty, x]) = F(x)$$

per ogni $x \in \mathbb{R}$. Inoltre la f.r. di X caratterizza la legge di probabilità: data F_X , per ogni $B \in \mathcal{B}_1$ si può sempre calcolare $P_X(B)$ con procedimenti di limite a partire da probabilità di intervalli.

10.3 Leggi univariate di tipo mistura

Le leggi di tipo discreto e le leggi di tipo continuo non esauriscono le possibili leggi di probabilità. Ad esempio si può facilmente costruire la f.r. di una v.c. con legge mista, con una parte discreta e una parte continua. Si consideri ad esempio una v.c. univariata X con f.r.

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{2} & \text{se } x = 0 \\ \frac{1}{2} + \frac{1}{2}(1 - e^{-\lambda x}) & \text{se } x > 0 \end{cases}$$

dove $\lambda > 0$. Per $F_X(x)$ l'origine è un punto di salto, e $P(X = 0) = 0.5$. Se $x \neq 0$ si ha $P(X = x) = 0$. La parte discreta della legge si può vedere come $\mathcal{D}(0)$, osservata con probabilità 0.5, la parte continua come $Exp(\lambda)$, pure osservata con probabilità 0.5. Si dice che X ha una legge mista (di altre leggi di probabilità).

In generale, gli ingredienti delle **leggi mistura** univariate sono

- le **leggi misturate**

v.c. X_i univariate con f.r. $F_{X_i}(x)$, $i \in I \subseteq \mathbb{N}$,

- i **pesi della mistura**

$$p_i > 0 \text{ per ogni } i \in I \subseteq \mathbb{N} \text{ con } \sum_{i \in I} p_i = 1.$$

Allora la mistura è

$$X = \begin{cases} X_1 \text{ con probabilità } p_1 \\ X_2 \text{ con probabilità } p_2 \\ \vdots \\ X_i \text{ con probabilità } p_i \\ \vdots \end{cases}$$

e la legge di probabilità di X è espressa tramite la sua f.r. data da

$$F_X(x) = \sum_{i \in I} p_i F_{X_i}(x). \quad (10.1)$$

Osservazione 1. Le X_i sono associabili ad una partizione A_i , $i \in I$, di S in eventi non trascurabili con $P(A_i) = p_i$ e $X = X_i$ se si realizza A_i , ossia $X|A_i \sim X_i$. Con tale identificazione la (10.1) è la formula della probabilità totale per $P(X \leq x)$.

Osservazione 2. Se gli eventi A_i sono definiti come nella precedente Osservazione, per ottenere realizzazioni x_r^* , $r = 1, \dots, R$ da X con legge mistura basta, per ogni fissato r , osservare quale evento A_i si realizza nella r -esima ripetizione dell'esperimento e porre $x_r^* = x_i$ dove x_i è realizzazione di X_i .

Osservazione 3. Tutte le v.c. X con legge discreta sono misture di degeneri, $X_i \sim \mathcal{D}(x_i)$ dove $x_i \in S_X$, con pesi $p_i = P(X = x_i)$.

10.4 Esercizi

Esercizio 10.1 Si dica, motivando, se la funzione $G(x) = e^x$ è f.r. di una qualche variabile casuale univariata X .

Esercizio 10.2 Si dica, motivando, se la funzione $G(x) = 1/(1 + x^2)$ è f.r. di una qualche variabile casuale univariata X .

Esercizio 10.3 Si dica, motivando, se la funzione $G(x) = e^x/(1 + e^x)$ è f.r. di una qualche variabile casuale univariata X . In caso affermativo, si dica se X ha legge discreta.

Esercizio 10.4 Si dica, motivando, se la funzione

$$G(x) = \begin{cases} 0 & \text{se } x < 1 \\ 1 - \frac{1}{x} & \text{se } x \geq 1 \end{cases}$$

è f.r. di una qualche variabile casuale univariata X . In caso affermativo, si dica se X ha legge continua.

Esercizio 10.5 Si dica, motivando, se la funzione

$$G(x) = \begin{cases} 0 & \text{se } x < 1 \\ \frac{1}{3} & \text{se } 1 \leq x \leq 2 \\ 1 & \text{se } x > 2 \end{cases}$$

è f.r. di una qualche variabile casuale univariata X . In caso affermativo, si dica se X ha legge discreta.

Esercizio 10.6 Si dica, motivando, se la funzione

$$G(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{2} & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

è f.r. di una qualche variabile casuale univariata X . In caso affermativo, si dica se X ha legge continua.

Esercizio 10.7 Sia X mistura, con pesi della mistura p_i , di un numero finito k di v.c. X_i tutte con legge continua. Si ottenga l'espressione della funzione di densità di probabilità di X .

Esercizio 10.8 Sia X mistura, con pesi della mistura p_i , di un numero finito k di v.c. X_i tutte con legge continua. Si mostri che $E(X) = \sum_{i=1}^k p_i E(X_i)$.

10.5 Appendice: simulare con R da una legge mistura

Si considerino i tempi di corretto funzionamento di un prodotto industriale. Si sa che il prodotto è fabbricato in tre siti che operano a livelli di qualità differenti. Dal sito A viene la metà della produzione, con distribuzione dei tempi di corretto funzionamento ben descritta da una esponenziale con valore atteso 2 (in una opportuna unità di misura). Dal sito B viene un quarto della produzione, con distribuzione esponenziale con valore atteso 3. Da C viene il resto della produzione, e la distribuzione dei tempi è esponenziale con valore atteso 1.

Si desidera simulare tempi di corretto funzionamento per prodotti di cui non è noto il sito produttivo, e che si possono considerare scelti a caso dal complesso della produzione. Siamo di fronte alla mistura X delle v.c. $X_A \sim Esp(1/2)$, $X_B \sim Esp(1/3)$, $X_C \sim Esp(1)$, con pesi della mistura $p_A = 0.5$, $p_B = 0.25$, $p_C = 0.25$. Lo scopo della simulazione è paragonare le quattro probabilità condizionali $P(X > 1 \mid X > 0)$, $P(X > 2 \mid X > 1)$, $P(X > 3 \mid X > 2)$ e $P(X > 4 \mid X > 3)$. La simulazione permette di osservare frequenze relative che valutano empiricamente le probabilità condizionali oggetto d'interesse, in accordo con la concezione frequentista della probabilità.

Si sceglie di effettuare 10^5 repliche. Si ottiene

```
> set.seed(12345)
> nrep=10^5
> k=3
> pesi=c(0.5,0.25,0.25)
> lambda=c(1/2, 1/3, 1)
> sito=sample(1:k, prob=pesi, size=nrep, replace=TRUE)
> mistura=rexp(nrep,rate=lambda[sito])
> length(mistura[mistura>1])/length(mistura[mistura>0])
[1] 0.57535
> length(mistura[mistura>2])/length(mistura[mistura>1])
[1] 0.6034935
> length(mistura[mistura>3])/length(mistura[mistura>2])
[1] 0.6241
> length(mistura[mistura>4])/length(mistura[mistura>3])
[1] 0.6356253
```

Si osserva che all'aumentare della soglia di condizionamento, aumenta anche la probabilità che il prodotto funzioni una ulteriore unità di tempo. Il risultato sembra paradossale. Tuttavia ha una chiara spiegazione: più a lungo un prodotto funziona, più è facile che provenga dalla fabbrica che opera a un livello di qualità elevata.

Lezione 11

Leggi di Poisson e leggi geometriche

Si presentano in questa lezione due leggi univariate notevoli di tipo discreto, le leggi di Poisson e le leggi geometriche.

11.1 Il teorema di Poisson

Le leggi di Poisson trovano la loro origine nella possibilità di approssimare in modo semplice certe probabilità binomiali, come nell'esempio seguente.

Esempio 11.1 Alla *roulette*, il giocatore Caio punta 37 volte una *fiche* sul suo numero preferito, il 13. Per le regole del gioco, quando esce il numero su cui si punta, si recupera la *fiche* assieme a 35 altre *fiche*. Quando non esce il numero su cui si punta, si perde la *fiche*. I numeri sono 37, da 1 a 36 più lo 0. Sia X il numero delle vittorie di Caio nelle 37 puntate. Se la *roulette* non è truccata,

$$X \sim Bi(37, 1/37).$$

Si ha allora

$$\begin{aligned} P(X=0) &= \binom{37}{0} \left(\frac{1}{37}\right)^0 \left(1 - \frac{1}{37}\right)^{37-0} = \left(1 - \frac{1}{37}\right)^{37} \doteq 0.3628513 \\ P(X=1) &= \binom{37}{1} \left(\frac{1}{37}\right)^1 \left(1 - \frac{1}{37}\right)^{37-1} = \left(1 - \frac{1}{37}\right)^{36} \doteq 0.3729305 \\ P(X=2) &= \binom{37}{2} \left(\frac{1}{37}\right)^2 \left(1 - \frac{1}{37}\right)^{37-2} = \frac{36}{2 * 37} \left(1 - \frac{1}{37}\right)^{35} \doteq 0.1864653 \end{aligned}$$

eccetera.

Dal limite notevole $\lim_{n \rightarrow \infty} (1 - \frac{1}{n})^n = e^{-1}$, si ottengono le approssimazioni

$$\begin{aligned} P(X = 0) &\doteq e^{-1} \doteq 0.3678794 \\ P(X = 1) &\doteq e^{-1} \doteq 0.3678794 \\ P(X = 2) &\doteq \frac{1}{2}e^{-1} \doteq 0.1839397 \end{aligned}$$

eccetera, dove si è usata per $P(X = 2)$ anche l'approssimazione implicata dal limite $\lim_{n \rightarrow \infty} (n-1)/n = 1$. Le approssimazioni ottenute usando i limiti sono in decoroso accordo con i valori esatti $P(X = x)$ almeno per x non troppo lontano da 1. \diamond

Più in generale si consideri, per $n > \lambda > 0$, la successione di v.c.

$$X_n \sim Bi(n, \lambda/n)$$

dove la legge binomiale ha n , numero di prove indipendenti, grande, e la probabilità costante di successo nelle singole prove $p = \lambda/n$, di conseguenza, piccola. Se si indica con X la legge limite della successione X_n al divergere di n , è facile vedere che X ha supporto

$$S_X = \lim_{n \rightarrow \infty} S_{X_n} = \cup_{n=1}^{\infty} S_{X_n} = \cup_{n=1}^{\infty} \{0, 1, \dots, n\} = \mathbb{N}.$$

Inoltre, per $x \in S_X$, la f.m.p. di X , $p_X(x)$, è determinata dal seguente **teorema di Poisson**.

Teorema 11.1 *Se $X_n \sim Bi(n, \lambda/n)$, per $x \in S_X = \mathbb{N}$,*

$$p_X(x) = \lim_{n \rightarrow \infty} P(X_n = x) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

Dimostrazione. Per $n > \lambda$ e $x \leq n$

$$\begin{aligned} P(X_n = x) &= \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x n(n-1) \cdots (n-x+1)}{x! n^x} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}. \end{aligned}$$

Il risultato segue dai limiti

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-x+1)}{n^x} &= 1 \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= e^{-\lambda} \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-x} &= 1. \end{aligned}$$

□

11.2 Le leggi di Poisson

Le considerazioni delineate nel paragrafo precedente conducono alla seguente definizione.

Definizione 11.1 Si dice che X ha **legge di Poisson con parametro λ** , $\lambda > 0$, e in breve si scriverà $X \sim P(\lambda)$, se è una v.c. univariata con legge discreta che ha supporto $S_X = \mathbb{N}$ e f.m.p., per $x \in S_X$, pari a

$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}. \quad (11.1)$$

Per verificare che si tratta di una buona definizione occorre controllare due condizioni:

- positività di $p_X(x)$ sul supporto $S_X = \mathbb{N}$:
banale perché $p_X(x)$ è prodotto di tre fattori positivi;
- normalizzazione:

$$\begin{aligned} \sum_{x \in S_X} p_X(x) &= \sum_{x=0}^{+\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} \\ &= e^{-\lambda+\lambda} = e^0 = 1. \end{aligned}$$

Sopra, si è usato lo sviluppo in serie di potenze della funzione esponenziale, cfr. formula (8.3),

$$e^z = \sum_{i=0}^{+\infty} \frac{z^i}{i!}.$$

Con la giustificazione data dal teorema di Poisson, si usa tipicamente una legge di Poisson $P(\lambda)$, $\lambda > 0$, per modellare la distribuzione di una variabile casuale che esprime un **conteggio** con supporto illimitato superiormente, o con limitazione superiore n poco rilevante perché n è molto più grande dei valori tipicamente assunti dal conteggio. È questo il caso di $Bi(n, p)$ per $p = \lambda/n$ e n sufficientemente grande.

Esempi di conteggi modellabili con una legge di Poisson sono il numero di

- telefonate che arrivano a un *call center* nell'unità di tempo
- batteri nel campo visivo di un microscopio
- bolle d'aria in una provetta
- *chip* difettosi in un *wafer*
- difetti in un prodotto
- mutazioni in un gene esposto ai raggi X

- accessi a un *web server* in un giorno.

Per una legge di Poisson il parametro λ rappresenta il valore atteso del conteggio. Infatti

$$E(X) = \sum_{x \in S_X} x p_X(x) = \sum_{x=0}^{+\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda \sum_{x=1}^{+\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} = \lambda \sum_{i=0}^{+\infty} e^{-\lambda} \frac{\lambda^i}{i!} = \lambda,$$

dove si è usato il cambiamento di indice $i = x - 1$. Quindi, se la legge è Poisson, la sola conoscenza del conteggio medio determina tutte le probabilità dei vari valori osservabili del conteggio. Analogamente, se $X \sim Bi(n, p)$ con n grande e $p = \lambda/n$, l'unica cosa che conta e che determina la distribuzione, sia pure in via approssimata, è il valore atteso della binomiale, $np = \lambda$.

Esempio 11.2 A un *server* arrivano in media 2 chiamate al minuto. Qual è la probabilità che in un determinato minuto non vi siano chiamate? E che vi siano più di 4 chiamate?

Si è di fronte a un conteggio X , numero di chiamate nel minuto in questione. Il conteggio è senza estremo superiore obbligato. Conviene modellare X con una legge di Poisson, quindi $X \sim P(2)$ perché $E(X) = 2$. Si ha allora

$$P(X = 0) = p_X(0) = e^{-2} \frac{2^0}{0!} = e^{-2} \doteq 0.1353353$$

e

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) \\ &= 1 - e^{-2} - e^{-2} 2 - e^{-2} \frac{2^2}{2!} - e^{-2} \frac{2^3}{3!} - e^{-2} \frac{2^4}{4!} \\ &\doteq 0.05265. \end{aligned}$$

◇

11.3 Le leggi geometriche

Una moneta con probabilità di dare testa pari a p , $p \in (0, 1)$, viene lanciata ripetutamente, con lanci indipendenti, finché si manifesta testa per la prima volta. Indicato $T_i = \text{'testa all'}i\text{'-esimo lancio'}$ e $C_i = \text{'croce all'}i\text{'-esimo lancio'}$, gli eventi elementari in S , spazio campionario dell'esperimento descritto, sono quindi

$$\begin{aligned} s_1 &= T_1 \\ s_2 &= C_1 \cap T_2 \\ s_3 &= C_1 \cap C_2 \cap T_3 \\ &\vdots \end{aligned}$$

Sia X il numero aleatorio di lanci che saranno necessari per vedere testa per la prima volta:

$$X(s_1) = 1, X(s_2) = 2, X(s_3) = 3, \text{ eccetera.}$$

È ovvio che

$$S_X = \{1, 2, 3, \dots\} = \mathbb{N}^+$$

per cui X ha legge discreta. La f.m.p. sarà, per l'indipendenza dei lanci, e quindi di eventi relativi a lanci distinti,

$$\begin{aligned} p_X(1) &= P(X=1) = P(\{s_1\}) = P(T_1) = p \\ p_X(2) &= P(X=2) = P(\{s_2\}) = P(C_1 \cap T_2) = (1-p)p \\ p_X(3) &= P(X=3) = P(\{s_3\}) = P(C_1 \cap C_2 \cap T_3) = (1-p)^2 p \\ &\vdots \\ p_X(x) &= P(X=x) = P(\{s_x\}) = P(C_1 \cap C_2 \cap \dots \cap C_{x-1} \cap T_x) = (1-p)^{x-1} p. \end{aligned}$$

Definizione 11.2 Si dice che ha **legge geometrica con parametro p** , $p \in (0, 1)$, e in breve si scriverà $X \sim Ge(p)$, una v.c. univariata X con legge discreta che ha supporto $S_X = \mathbb{N}^+$ e f.m.p. per $x \in S_X$ pari a

$$p_X(x) = p(1-p)^{x-1}. \quad (11.2)$$

Per verificare che si tratta di una buona definizione occorre controllare due condizioni:

- positività di $p_X(x)$ sul supporto $S_X = \mathbb{N}^+$:
banale perché $p_X(x)$ è prodotto di fattori positivi;
- normalizzazione:

$$\begin{aligned} \sum_{x \in S_X} p_X(x) &= \sum_{x=1}^{+\infty} p(1-p)^{x-1} \\ &= p \sum_{x=1}^{+\infty} (1-p)^{x-1} \\ &= p \sum_{i=0}^{+\infty} (1-p)^i \\ &= p \frac{1}{1-(1-p)} = 1. \end{aligned}$$

Sopra, si è usato il cambiamento di indice $x - 1 = i$ e il risultato sulla somma della serie geometrica con ragione x , $|x| < 1$:

$$\begin{aligned}\sum_{i=0}^{\infty} x^i &= \lim_{n \rightarrow \infty} \sum_{i=0}^n x^i \\ &= \lim_{n \rightarrow \infty} \frac{1 - x^{n+1}}{1 - x} \\ &= \frac{1}{1 - x} \quad \text{per } |x| < 1.\end{aligned}$$

Per una legge geometrica, $X \sim Ge(p)$, il parametro p rappresenta $P(X = 1)$. Quindi, se la legge è geometrica, la sola conoscenza di $P(X = 1)$ determina tutte le probabilità dei vari valori osservabili del conteggio. Si dimostra che $E(X) = 1/p$, cfr. paragrafo 16.4.

Esempio 11.3 Il 20% dei clienti dà disdetta entro il tempo legale all'ordine di un certo prodotto fatto via *web*. Sia X il numero di ordini che arriveranno da adesso fino al primo ordine che sarà disdetto compreso, ossia il numero di ordini occorrenti per osservare la disdetta per la prima volta. Evidentemente $S_X = \mathbb{N}^+$ e $P(X = 1) = 0.2$. Quindi $P(X = 2) = 0.2 * 0.8 = 0.16$, $P(X = 3) = 0.2 * 0.8^2 = 0.128$, $P(X = 4) = 0.2 * 0.8^3 = 0.1024$, $P(X = 5) = 0.2 * 0.8^4 = 0.08192$, eccetera. \diamond

11.4 Assenza di memoria delle leggi geometriche

La funzione di sopravvivenza di $X \sim Ge(p)$ è, per $x \in \mathbb{N}^+$,

$$P(X > x) = P(C_1 \cap C_2 \cap \dots \cap C_x) = P(C_1)P(C_2) \dots P(C_x) = (1 - p)^x$$

per cui la funzione di ripartizione è, sempre per $x \in \mathbb{N}^+$,

$$F_X(x) = P(X \leq x) = 1 - P(X > x) = 1 - (1 - p)^x.$$

Per $x \in \mathbb{R}$ si ha

$$F_X(x) = \begin{cases} 0 & \text{se } x < 1 \\ 1 - (1 - p)^{\lfloor x \rfloor} & \text{se } x \geq 1 \end{cases}$$

dove $\lfloor x \rfloor$ indica la parte intera di x , ossia il più grande intero minore o uguale a x .

Sia $X \sim Ge(p)$ con $0 < p < 1$. La v.c. X può essere vista come un tempo d'attesa nel tempo discreto che gode della **proprietà di assenza di memoria**: per ogni $s, t \in \mathbb{N}^+$

$$P(X > s + t \mid X > s) \quad \text{non dipende da } s.$$

Infatti

$$P(X > s + t \mid X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{(1 - p)^{s+t}}{(1 - p)^s} = (1 - p)^t = P(X > t).$$

Se vi è assenza di memoria, il tempo d'attesa X realizzerà un ulteriore ritardo di almeno t unità di tempo con una probabilità indipendente da quante unità di tempo, s , sono già trascorse dall'inizio della prova senza che l'attesa finisse.

11.5 Esercizi

Esercizio 11.1 Un vetro è troppo fragile per la posa in opera se ha più di due bolle d'aria. Si calcoli la probabilità che sia troppo fragile se la probabilità che non abbia bolle d'aria è 0.90.

Esercizio 11.2 Un prodotto fabbricato in grande serie ha in media un difetto. Sia X il numero di difetti di un esemplare estratto a caso dal flusso di produzione. Si calcoli $P(X \geq 2)$.

Esercizio 11.3 A un certo server arrivano in media 20 chiamate al minuto. Si calcoli la probabilità che nel prossimo secondo non arrivino chiamate.

Esercizio 11.4 Su una autostrada il numero medio annuo di incidenti è 1 ogni cento chilometri. Si calcoli la probabilità che su un tratto di 200 chilometri avvengano più di due incidenti in un anno.

Esercizio 11.5 Due macchine, α e β , hanno rispettivamente, in media settimanale, 1 e 2 blocchi per errore. Le due macchine operano indipendentemente e si suppone che i blocchi siano riparati istantaneamente. Si calcoli la probabilità che la prossima settimana il numero totale di blocchi delle due macchine sia 2.

Esercizio 11.6 Due squadre di calcio, A e B , segnano un numero di goal per partita X e Y , rispettivamente, con distribuzione di Poisson $X \sim P(\lambda_A)$ e $Y \sim P(\lambda_B)$. Supponendo che X e Y siano indipendenti, si calcolino $P(X = 1, Y = 2)$, $P(X = 2, Y = 1)$, $P(X + Y = 0)$, $P(X + Y = 1)$.

Esercizio 11.7 Il numero richieste di risarcimento presentate in un anno da un assicurato è $X \sim P(0.1)$ per danni all'abitazione, $Y \sim P(0.2)$ per danni a mobili ed elettrodomestici, $Z \sim P(0.3)$ per danni all'autoveicolo. Supponendo che X , Y e Z siano indipendenti, si calcoli $P(X + Y + Z = 0)$.

Esercizio 11.8 Una prova d'esame costituita da un test a crocette viene superata da uno studente impreparato con probabilità 0.05. Con quale probabilità l'esame sarà superato entro i prossimi tre appelli, se lo studente impreparato ha fallito l'esame già due volte?

Esercizio 11.9 A una determinata ruota del lotto il numero 57 non esce da 99 estrazioni. Sia X il numero di estrazioni necessarie per rivedere il 57 a quella ruota del lotto dopo l'ultima sua uscita. Si calcoli $P(X > 110 \mid X > 99)$. Si commenti il risultato $P(X > s + t \mid X > s) = P(X > t)$, per ogni $s, t \in \mathbb{N}^+$.

Esercizio 11.10 Sia $X \sim Ge(0.1)$, si tabulino funzione di massa di probabilità e funzione di ripartizione di X per $x = 1, 2, 3, 4, 5$.

11.6 Appendice: v.c. con legge Poisson e geometrica con R

Grazie a R si possono rifare con maggior dettaglio i calcoli dell'Esempio 11.1. Il confronto fra probabilità binomiali e probabilità Poisson dà

```
> x=0:6
> prob_bin=dbinom(x,size=37,prob=1/37)
> prob_pois=dpois(x,lambda=1)          # lambda=size*prob
> tb=rbind(x,round(prob_bin,digits=5),round(prob_pois,digits=5))
> tb
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
x	0.00000	1.00000	2.00000	3.00000	4.00000	5.00000	6.00000
	0.36285	0.37293	0.18647	0.06043	0.01427	0.00262	0.00039
	0.36788	0.36788	0.18394	0.06131	0.01533	0.00307	0.00051

Se invece di puntare 37 volte sul numero fortunato si punta $37*2=74$ volte, l'approssimazione di Poisson per la legge del conteggio è ancora adeguata:

```
> x=0:6
> prob_bin=dbinom(x,size=37*2,prob=1/37)
> prob_pois=dpois(x,lambda=2)          # lambda=size*prob
> tb=rbind(x,round(prob_bin,digits=5),round(prob_pois,digits=5))
> print(tb)
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
x	0.00000	1.00000	2.00000	3.00000	4.00000	5.00000	6.00000
	0.13166	0.27064	0.27440	0.18293	0.09019	0.03508	0.01120
	0.13534	0.27067	0.27067	0.18045	0.09022	0.03609	0.01203

Il comando `ppois(x,lambda=lambda)` calcola i valori della f.r. di $X \sim P(\lambda)$ per un vettore di valori x . Per $n = 37$ e $\lambda = 1$, la f.r. di $X_n \sim Bi(n, \lambda/n)$ è vicina a quella di $X \sim P(\lambda)$ e che le due f.r. sono ancora più vicine se $n = 100$:

```
> x=0:37
> frbin=pbinom(x,size=37,prob=1/37)
> frpois=ppois(x,lambda=1)
> max(frbin-frpois)
[1] 0.002548567
> min(frbin-frpois)
[1] -0.005028096
> x=0:100
> frbin=pbinom(x,size=100,prob=1/100)
> frpois=ppois(x,lambda=1)
```

```
> max(frbn-frpois)
[1] 0.0009281948
> min(frbn-frpois)
[1] -0.0018471
```

Il comando `rpois(nrep,lambda=lambda)` permette di generare `nrep` realizzazioni da una $P(\lambda)$. Ad esempio, il codice

```
> set.seed(12345)
> rpois(10,lambda=3)
[1] 4 5 4 5 3 1 2 3 4 8
> y=rpois(10^6,lambda=3)
> mean(y)
[1] 3.000456
> var(y)
[1] 3.008265
> y=rpois(10^6,lambda=4)
> mean(y)
[1] 4.000447
> var(y)
[1] 4.009081
```

permette di congetturare quanto vale la varianza di una $P(\lambda)$.

I comandi principali per trattare v.c. con legge geometrica sono i seguenti: `dgeom(x,prob=prob)`, `pgeom(x,prob=prob)` e `rgeom(n,prob=prob)`. R considera come geometrica con parametro `prob` non la legge del numero di prove occorrenti per realizzare il primo successo ma la legge del conteggio del numero di insuccessi realizzati nelle prove indipendenti occorse per realizzare il primo successo. Le prove hanno sempre costante probabilità di successo `prob`. Con questa definizione il supporto è traslato di una unità verso sinistra, non è \mathbb{N}^+ ma \mathbb{N} . Esempi d'uso:

```
> set.seed(12345)
> rgeom(10,prob=1/2)
[1] 1 1 1 3 1 0 1 1 4 1
> round(dgeom(0:6,prob=1/2),digits=5)
[1] 0.50000 0.25000 0.12500 0.06250 0.03125 0.01562 0.00781
> round(pgeom(0:6,prob=1/2),digits=5)
[1] 0.50000 0.75000 0.87500 0.93750 0.96875 0.98438 0.99219
```

Anche traslate, le leggi geometriche hanno sempre la proprietà di assenza di memoria. Una verifica per alcuni tempi particolari è presto fatta. Si considerino le quattro probabilità $P(X > 1|X > 0)$, $P(X > 2|X > 1)$, $P(X > 3|X > 2)$, $P(X > 4|X > 3)$. Seguendo la concezione frequentista, queste probabilità condizionali sono valutate in modo approssimato dalle corrispondenti frequenze relative in uno studio di simulazione con un numero di repliche grande. Ad esempio, con $p = 1/2$ e un milione di repliche, si ha

```
> set.seed(12345)
> y=rgeom(10^6,prob=1/2)
> length(y[y>1])/length(y[y>0])
[1] 0.4998543
> length(y[y>2])/length(y[y>1])
[1] 0.5024724
> length(y[y>3])/length(y[y>2])
[1] 0.5012758
> length(y[y>4])/length(y[y>3])
[1] 0.5003806
```

che mostra chiaramente l'assenza di memoria, fatta la tara delle fluttuazioni campionarie.

Lezione 12

Modelli per tempi d'attesa nel tempo continuo

12.1 Assenza di memoria e leggi esponenziali

Nel tempo continuo, sono le leggi esponenziali che possiedono la proprietà di assenza di memoria. Infatti

$$X \sim \text{Esp}(\lambda), \lambda > 0, \quad \Longleftrightarrow \quad P(X > x) = e^{-\lambda x} \text{ per } x > 0,$$

per cui, per $s, t > 0$,

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s} e^{-\lambda t}}{e^{-\lambda s}} \\ &= e^{-\lambda t} = P(X > t). \end{aligned}$$

Quindi $X \sim \text{Esp}(\lambda)$ esprime un tempo d'attesa (nel tempo continuo), come il tempo al guasto o il tempo al primo 'successo', per il quale la fine dell'attesa non è né favorita né procrastinata dal protrarsi dell'attesa. Se l'attesa non finisce entro il tempo s , è come ripartire da 0 nell'attesa. Esempi: tempo fra due chiamate telefoniche, fra due accessi a un *server*, eccetera.

12.2 La funzione tasso di guasto

Sia T un tempo d'attesa, ossia una v.c. univariata con $P(T \geq 0) = 1$, con l'attesa misurata nel tempo continuo, per cui T ha legge continua. In diversi contesti applicativi l'assunzione di assenza di memoria non sembra appropriata e quindi si richiedono modellazioni alternative all'esponenziale.

Si indichino con

- $F_T(t) = P(T \leq t)$ la funzione di ripartizione, come al solito;
- $\bar{F}_T(t) = 1 - F_T(t) = P(T > t)$ la funzione di sopravvivenza;
- $p_T(t) = \frac{d}{dt} F_T(t)$ la funzione di densità supposta continua (ovunque, salvo che in un numero finito di punti).

Definizione 12.1 Si dice funzione **tasso di guasto** di T (hazard rate, failure rate) la funzione $r_T(\cdot)$, definita per i valori di t per cui $F_T(t) < 1$, da

$$r_T(t) = \frac{p_T(t)}{\bar{F}_T(t)} = -\frac{d}{dt} \log \bar{F}_T(t).$$

La funzione tasso di guasto ha la seguente interpretazione. Per i punti t dove per $\varepsilon > 0$ sufficientemente piccolo vale

$$P(t < T \leq t + \varepsilon) \doteq p_T(t)\varepsilon$$

vale anche

$$P(T \in (t, t + \varepsilon] \mid T > t) = \frac{P(t < T \leq t + \varepsilon)}{P(T > t)} \doteq \frac{p_T(t)\varepsilon}{\bar{F}_T(t)} = r_T(t)\varepsilon.$$

Quindi, nei punti in cui $r_T(t)$ è continua, $r_T(t)$ è proporzionale alla probabilità che l'attesa, ancora viva al tempo t , termini entro il tempo $t + \varepsilon$, ossia nel tempuscolo immediatamente successivo a t .

La funzione tasso di guasto determina la f.r. e la f.d.p. di T (e di conseguenza la legge di probabilità del tempo d'attesa). Infatti, integrando $r_T(t) = -\frac{d}{dt} \log \bar{F}_T(t)$ si ha

$$\int_0^t r_T(u) du = [-\log \bar{F}_T(u)]_0^t = -\log \bar{F}_T(t)$$

da cui si ottiene che per $t > 0$

$$F_T(t) = 1 - \exp \left\{ -\int_0^t r_T(u) du \right\}$$

e

$$p_T(t) = \frac{d}{dt} F_T(t) = r_T(t) \exp \left\{ -\int_0^t r_T(u) du \right\}.$$

Si vede subito che una funzione reale, definita su un intervallo, eventualmente illimitato, S_T , è funzione tasso di guasto di un tempo d'attesa T se e solo se

- $r_T(t) \geq 0$ per ogni $t \in S_T$
- $\int_{S_T} r_T(t) = +\infty$.

Se si esamina l'andamento qualitativo delle funzioni tasso di guasto dove $F_T(t) < 1$, si distinguono tre situazioni notevoli:

1. $r_T(t) = \lambda$ per ogni $t > 0$: tasso di guasto costante;
2. $r_T(t)$ è una funzione crescente: tasso di guasto crescente;
3. $r_T(t)$ è una funzione decrescente di t : tasso di guasto decrescente.

È immediato constatare che T ha tasso di guasto costante se e solo se ha legge esponenziale, ossia ha la proprietà di assenza di memoria nel tempo continuo. Se $r_T(t) = \lambda > 0$ per ogni $t > 0$ si ha infatti

$$F_T(t) = 1 - \exp \left\{ - \int_0^t \lambda dt \right\} = 1 - e^{-\lambda t}.$$

Pertanto il parametro λ delle leggi esponenziali esprime il tasso di guasto costante delle stesse.

Se il tempo d'attesa T ha tasso di guasto crescente, al passare del tempo nell'attesa, risulta via via più facile che la fine dell'attesa avvenga nel tempuscolo immediatamente successivo al presente. Se invece T ha tasso di guasto decrescente, al passare del tempo nell'attesa risulta via via più difficile che la fine dell'attesa avvenga nel tempuscolo immediatamente successivo al presente.

Se il tempo d'attesa è un tempo al guasto, $r_T(t)$ crescente modella guasti derivanti da usura o invecchiamento. Il tasso di guasto può essere decrescente in situazioni di selezione iniziale delle unità meno ben costruite (cfr. la mistura considerata nel paragrafo 10.5). Il tasso di guasto costante corrisponde infine a situazioni in cui il guasto è dovuto a uno *shock* esterno puramente casuale.

12.3 Le leggi di Weibull

Per modellare tempi d'attesa nel continuo con tasso di guasto monotono sono assai utili e semplici da descrivere le leggi di Weibull.

Un tempo d'attesa T_0 ha **legge di Weibull monoparametrica**, con parametro di forma $c > 0$, se per $t > 0$ il suo tasso di guasto è

$$r_{T_0}(t) = ct^{c-1}.$$

Si scrive allora in breve $T_0 \sim W(c)$.

Si vede subito che $r_{T_0}(t)$ è decrescente se $0 < c < 1$, costante se $c = 1$, crescente se $c > 1$. Poiché

$$\int_0^t r_{T_0}(u) du = \int_0^t cu^{c-1} du = [u^c]_0^t = t^c$$

si ha, per $t > 0$,

$$F_{T_0}(t) = 1 - e^{-t^c}$$

e, sempre per $t > 0$,

$$p_{T_0}(t) = ct^{c-1}e^{-t^c}.$$

Poiché solo $Esp(1)$ fa parte delle leggi di Weibull monoparametriche, conviene introdurre un parametro $\lambda > 0$ come nell'esponenziale, considerando un tempo d'attesa T con tasso di guasto per $t > 0$

$$r_T(t) = \lambda c(\lambda t)^{c-1},$$

ancora decrescente se $0 < c < 1$, costante e pari a λ se $c = 1$, crescente se $c > 1$, per ogni valore di λ . Si dice allora che T ha **legge di Weibull biparametrica**, con parametro di forma c e di scala λ , e in breve si scrive $T \sim W(c, \lambda)$.

Poiché $\int_0^t r_T(u) du = (\lambda t)^c$ la f.r. di $T \sim W(c, \lambda)$ è, per $t > 0$,

$$F_T(t) = 1 - \exp\{-(\lambda t)^c\}$$

e la corrispondente f.d.p. è, per $t > 0$

$$p_T(t) = \lambda c(\lambda t)^{c-1} \exp\{-(\lambda t)^c\}.$$

Le leggi esponenziali sono un caso particolare delle leggi di Weibull biparametriche, $Esp(\lambda) \sim W(1, \lambda)$. La ragione per cui c è detto parametro di forma è che al variare di c cambia fortemente la forma della funzione di densità di probabilità, da monotona decrescente in $(0, \infty)$ con asintoto verticale in 0^+ se $c \in (0, 1)$ ad esponenziale se $c = 1$ a circa campanulare se $c > 1$.

Il parametro λ è detto di scala perché fra $T \sim W(c, \lambda)$ e $T_0 \sim W(c)$ sussiste la relazione

$$T \sim \frac{T_0}{\lambda}.$$

Si ha infatti per $t > 0$

$$P(T_0/\lambda \leq t) = P(T_0 \leq \lambda t) = F_{T_0}(\lambda t) = 1 - \exp\{-(\lambda t)^c\} = P(T \leq t)$$

e due v.c. hanno la stessa legge di probabilità se hanno la stessa funzione di ripartizione.

12.4 Le leggi gamma

Tempi d'attesa nel continuo con tasso di guasto monotono possono essere modellati anche dalle leggi gamma, il cui supporto è $[0, +\infty)$. Conviene anche per esse distinguere fra leggi monoparametriche e leggi biparametriche.

Un tempo d'attesa T_0 ha **legge gamma monoparametrica**, con parametro di forma $\alpha > 0$, se per $t > 0$ la sua funzione di densità di probabilità è

$$p_{T_0}(t) = \frac{1}{\Gamma(\alpha)} t^{\alpha-1} e^{-t}.$$

La funzione $\Gamma(\alpha)$, detta **funzione gamma di Eulero**, è definita dall'integrale, convergente per $\alpha > 0$,

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt.$$

La proprietà saliente di $\Gamma(\alpha)$ è che estende a un argomento reale positivo la nozione di fattoriale di un numero naturale n . Valgono infatti le eguaglianze

- $\Gamma(1) = 1$
- $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$

da cui $\Gamma(n + 1) = n!$.

Per valori non interi di α calcolare la f.r. di $T_0 \sim Ga(\alpha)$ richiede un'integrazione numerica, e pertanto anche la funzione tasso di guasto di $Ga(\alpha)$ non è esprimibile in generale in forma chiusa. Si dimostra comunque che $r_{T_0}(t)$ è decrescente se $0 < \alpha < 1$, costante se $\alpha = 1$, crescente se $\alpha > 1$.

Conviene introdurre un parametro $\lambda > 0$ come nell'esponenziale. La v.c. T è detta con **legge gamma biparametrica**, con parametro di forma α e di scala λ , in breve si scrive $T \sim Ga(\alpha, \lambda)$, se la f.d.p. di T è, per $t > 0$,

$$p_T(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}.$$

Le leggi esponenziali sono dunque un caso particolare delle leggi gamma con due parametri, $Exp(\lambda) \sim Ga(1, \lambda)$. La ragione per cui α è detto parametro di forma è che al variare di α cambia fortemente la forma della funzione di densità di probabilità, da monotona decrescente in $(0, \infty)$ con asintoto verticale in 0^+ se $\alpha \in (0, 1)$, ad esponenziale se $\alpha = 1$, a circa campanulare se $\alpha > 1$.

Il parametro λ è detto di scala perché $T \sim T_0/\lambda$. Si ha infatti per $t > 0$

$$F_{T_0/\lambda}(t) = P(T_0/\lambda \leq t) = P(T_0 \leq \lambda t) = F_{T_0}(\lambda t)$$

per cui, derivando rispetto a t ,

$$p_{T_0/\lambda}(t) = \lambda p_{T_0}(\lambda t) = \lambda \frac{1}{\Gamma(\alpha)} (\lambda t)^{\alpha-1} e^{-\lambda t} = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} = p_T(t)$$

e due v.c. hanno la stessa legge di probabilità di tipo continuo se hanno la stessa funzione di densità di probabilità.

12.5 Esercizi

Esercizio 12.1 Sia T una v.c. univariata con legge di Weibull, $T \sim W(2)$. Si ottengano $p_T(t)$ e $F_T(t)$. Si calcolino $P(T > 2 \mid T > 1)$ e $P(T > 3 \mid T > 2)$.

Esercizio 12.2 Sia T una v.c. univariata con legge di Weibull, $T \sim W(2, 2)$. Si ottengano $p_T(t)$ e $F_T(t)$. Si calcoli $P(T > 2 \mid T > 1)$ e $P(T > 3 \mid T > 2)$.

Esercizio 12.3 Sia T una v.c. univariata con legge gamma, $T \sim Ga(2)$. Si scriva $p_T(t)$ e si ottenga $F_T(t)$ (integrando per parti). Si calcolino le probabilità $P(T > 2 \mid T > 1)$ e $P(T > 3 \mid T > 2)$.

Esercizio 12.4 Sia T una v.c. univariata con legge gamma, $T \sim Ga(1, 0.5)$. Si scriva $p_T(t)$ e si ottenga $F_T(t)$. Si calcolino $P(T > 2 \mid T > 1)$ e $P(T > 3 \mid T > 2)$.

Esercizio 12.5 Sia T una v.c. univariata con supporto $S_T = [1, +\infty)$ e funzione tasso di guasto data, per $t \geq 1$, da

$$r_T(t) = \alpha/t.$$

dove $\alpha > 0$. Si ottengano $p_T(t)$ e $F_T(t)$. Si calcoli $P(T > 3 \mid T > 2)$.

Esercizio 12.6 Si mostri che se T è una v.c. univariata con legge continua e $P(T \geq 0) = 1$ si ha $E(T) = \int_0^{+\infty} (1 - F_T(t)) dt$.
Sugg.: si integri per parti e si consideri che $\lim_{t \rightarrow \infty} t(1 - F_T(t)) = 0$ perché $0 \leq t(1 - F_T(t)) \leq \int_t^{+\infty} u p_T(u) du$.

Esercizio 12.7 Sia $T \sim Esp(\lambda)$, $\lambda > 0$. Si mostri che $E(T) = 1/\lambda$ usando $E(T) = \int_0^{+\infty} (1 - F_T(t)) dt$.

Esercizio 12.8 Sia T una v.c. univariata con supporto $S_T = [0, +\infty)$ e funzione tasso di guasto lineare, ossia, per $t \geq 0$ della forma

$$r_T(t) = \lambda + \alpha t$$

dove $\lambda > 0$ e $\alpha \geq 0$. Si ottengano $p_T(t)$ e $F_T(t)$. Si verifichi che T non ha, per $\alpha > 0$, la proprietà di assenza di memoria, ossia che, per $s, t > 0$ la probabilità condizionale $P(X > s + t \mid X > s)$ dipende da s .

12.6 Appendice: v.c. con legge Weibull e gamma con R

Sia $X \sim W(c, \lambda)$, con $c = 3$, $\lambda = 2$. La densità di X per un vettore \mathbf{x} di valori reali si calcola con il comando `dweibull(x, shape=3, scale=1/2)`. Si noti che `scale = 1/λ`. Il valore della f.r. di X si ottiene con il comando `pweibull(x, shape=3, scale=1/2)`. Si generano `nrep` realizzazioni di X con il comando `rweibull(nrep, shape=3, scale=1/2)`.

Sia $X \sim Ga(\alpha, \lambda)$, con $\alpha = 3$, $\lambda = 2$. La densità di X in corrispondenza con un vettore \mathbf{x} di valori reali si calcola con il comando `dgamma(x, shape=3, rate=2)`. Il valore della f.r. di X si ottiene con il comando `pgamma(x, shape=3, rate=2)`. Si generano `nrep` realizzazioni di X con `rgamma(nrep, shape=3, rate=2)`.

Lezione 13

Leggi di v.c. trasformate

13.1 Trasformazioni di v.c.

Come si può operare sui numeri reali trasformandoli mediante funzioni, allo stesso modo anche le variabili casuali sono trasformabili mediante funzioni. Si ottengono così nuove v.c., dette v.c. **trasformate**.

Ad esempio, se un conteggio di successi è $X \sim Bi(n, p)$ ci si può chiedere quale sia la legge del conteggio degli insuccessi, ossia la legge di $Y = n - X$. Analogamente, se $X \sim Esp(\lambda)$ esprime un tempo d'attesa misurato in minuti, ci si può chiedere quale sia la legge di probabilità se il tempo d'attesa è misurato invece in secondi, ovvero quale sia la legge di $Y = 60X$. Ancora, se si ha $X = (X_1, X_2, X_3)$, una v.c. trasformata di X , indicata con Y , è la v.c. bivariata $Y = (Y_1, Y_2)$ dove $Y_1 = \min(X_1, X_2, X_3)$ e $Y_2 = \max(X_1, X_2, X_3)$. Un'altra trasformata, questa volta univariata, è $Y = X_1 + X_2 + X_3$.

Sia $X = (X_1, \dots, X_d)$ una v.c. d -variata. Si considerino poi un dominio di definizione $D \in \mathcal{B}_d$ tale che $P(X \in D) = 1$ e una funzione

$$g : D \rightarrow \mathbb{R}^{d'} \quad \text{dove } d' \leq d$$

misurabile, ossia tale che per ogni $B' \in \mathcal{B}_{d'}$ si abbia $g^{-1}(B') \in \mathcal{B}_d$, dove $\mathcal{B}_{d'}$ e \mathcal{B}_d sono le σ -algebre di Borel associate a $\mathbb{R}^{d'}$ e a \mathbb{R}^d , rispettivamente. In pratica, tutte le funzioni che si riesce a scrivere esplicitamente sono misurabili. Allora la v.c. trasformata $Y = g(X)$ è una v.c. d' -variata la cui legge di probabilità è data, per ogni $B' \in \mathcal{B}_{d'}$, da

$$P_Y(B') = P_X(g^{-1}(B')), \quad (13.1)$$

o equivalentemente da $P(Y \in B') = P(X \in g^{-1}(B'))$.

In linea di principio, la formula (13.1) determina la legge di probabilità della v.c. trasformata Y per ogni assegnata legge di probabilità della X di partenza. Tuttavia, molto spesso la legge di probabilità di X è definita in modo indiretto tramite strumenti semplici quali supporto, S_X , e f.m.p./f.d.p., $p_X(x)$, o anche

tramite la f.r. $F_X(x)$. È allora richiesto di esprimere anche la legge di Y tramite un analogo strumento semplice.

13.2 Il supporto di una v.c. trasformata

Il supporto di $Y = g(X)$ si determina facilmente. Se X ha legge discreta, $D = S_X$, e $g(D)$ non ha punti di accumulazione al finito, allora

$$S_Y = g(S_X) = \{y \in \mathbb{R}^{d'} : y = g(x) \text{ per qualche } x \in S_X\}. \quad (13.2)$$

Quindi, se X ha legge discreta, anche $Y = g(X)$ ha legge discreta: S_Y è anch'esso finito o numerabile perché $|S_Y| = |g(S_X)| \leq |S_X|$. Se X ha legge continua, $D = S_X$ e g è continua in D , si ottiene per S_Y la stessa espressione (13.2).

Più in generale S_Y sarà la chiusura di $g(D)$, ossia l'unione di $g(D)$ con l'insieme dei punti di accumulazione di $g(D)$, dove $D \subseteq S_X$ tale che $P_X(D) = 1$. Ad esempio, se $X \sim \text{Esp}(1)$ e $Y = \log X = g(X)$, si ha $S_X = [0, +\infty)$ ma $g(0)$ non è definita. Allora basta considerare $D = (0, +\infty)$ come dominio di definizione della trasformazione. Si ha $g(D) = g((0, +\infty)) = \mathbb{R}$ che è chiuso (ha come elementi tutti i suoi punti di accumulazione), quindi $S_Y = \mathbb{R}$.

13.3 La f.m.p. di una v.c. trasformata

Sia X con legge discreta, supporto S_X e f.m.p. $p_X(x)$, definita per ogni $x \in S_X$. Data g si determina il supporto di $Y = g(X)$ come nel paragrafo precedente. Per ogni $y \in S_Y$ la f.m.p. di Y vale allora

$$p_Y(y) = \sum_{x \in S_X : g(x)=y} p_X(x).$$

Infatti

$$\begin{aligned} p_Y(y) &= P(Y = y) = P(g(X) = y) \\ &= P(X \in g^{-1}(y)) \\ &= \sum_{x \in g^{-1}(y)} p_X(x) \\ &= \sum_{x \in S_X : g(x)=y} p_X(x), \end{aligned}$$

poiché $g^{-1}(y) = \{x \in S_X : g(x) = y\}$.

Esempio 13.1 Si mostra che, se $X \sim \text{Bi}(n, p)$ allora la trasformata $Y = g(X) = n - X \sim \text{Bi}(n, 1 - p)$. Poiché $S_X = \{0, 1, \dots, n\}$ e $y = g(x) = n - x$, si

vede che $S_Y = S_X$ e, dal momento che $x = g^{-1}(y) = n - y$, per $y \in S_Y$,

$$\begin{aligned} p_Y(y) &= p_X(g^{-1}(y)) \\ &= p_X(n - y) \\ &= \binom{n}{n-y} p^{n-y} (1-p)^y \\ &= \binom{n}{y} (1-p)^y (1 - (1-p))^{n-y}, \end{aligned}$$

quindi $Y \sim Bi(n, 1-p)$. \diamond

Esempio 13.2 Si lanciano, con lanci indipendenti, un dado rosso e un dado verde. Sia X_1 il punteggio del dado rosso, X_2 il punteggio del dado verde. Poiché (X_1, X_2) ha componenti indipendenti,

$$S_{X_1, X_2} = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$

e, per $(x_1, x_2) \in S_{X_1, X_2}$,

$$p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2) = \frac{1}{6} \frac{1}{6} = \frac{1}{36}$$

per cui $(X_1, X_2) \sim Ud(S_{X_1, X_2})$. Si consideri la trasformata

$$Y = g(X_1, X_2) = \max(X_1, X_2).$$

Si ha subito

$$S_Y = \{1, 2, 3, 4, 5, 6\}$$

e la f.m.p. per $y \in S_Y$ risulta

$$\begin{aligned} p_Y(1) &= p_{X_1, X_2}(g^{-1}(1)) = p_{X_1, X_2}(1, 1) = \frac{1}{36} \\ p_Y(2) &= p_{X_1, X_2}(g^{-1}(2)) = p_{X_1, X_2}(1, 2) + p_{X_1, X_2}(2, 2) + p_{X_1, X_2}(2, 1) = \frac{3}{36} \\ p_Y(3) &= \dots = \frac{5}{36} \\ &\vdots \\ p_Y(6) &= \dots = \frac{11}{36}. \end{aligned}$$

Quindi per $y \in S_Y$ si ha $p_Y(y) = (2y - 1)/36$. \diamond

13.4 La f.r. di trasformate monotone

Se X è univariata e $g(x)$ è monotona con inversa $g^{-1}(y)$ è facile trovare la f.r. di Y , $F_Y(y)$. Vanno distinti due casi:

i) se g è monotona crescente

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)); \end{aligned}$$

ii) se g è monotona decrescente

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \geq g^{-1}(y)) \\ &= P(X > g^{-1}(y)) + P(X = g^{-1}(y)) \\ &= 1 - F_X(g^{-1}(y)) + P(X = g^{-1}(y)). \end{aligned}$$

I risultati ottenuti sopra valgono sia per leggi discrete sia per leggi continue, anzi per qualunque tipo di legge di X . Si osservi che se X ha legge continua $P(X = g^{-1}(y)) = 0$ e con g monotona decrescente si ottiene la più semplice formula $F_Y(y) = 1 - F_X(g^{-1}(y))$.

13.5 La f.d.p. di trasformate monotone

Se X ha legge continua con f.d.p.

$$p_X(x) = \frac{d}{dx} F_X(x)$$

e inoltre g è monotona e derivabile assieme alla sua inversa con derivata diversa da 0, allora anche Y ha legge continua con f.d.p. $p_Y(y)$ espressa, nei due casi da considerare, da

i) se g è monotona crescente

$$\begin{aligned} p_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} F_X(g^{-1}(y)) \\ &= p_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y); \end{aligned}$$

ii) se g è monotona decrescente

$$\begin{aligned} p_Y(y) &= \frac{d}{dy} (1 - F_X(g^{-1}(y))) \\ &= -p_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y). \end{aligned}$$

Va osservato che se g è monotona crescente anche g^{-1} è crescente e la derivata di $g^{-1}(y)$ è positiva. Se invece g è monotona decrescente anche g^{-1} è decrescente e la derivata di $g^{-1}(y)$ è negativa. Le due formule trovate si possono quindi unificare grazie al valore assoluto.

Nelle assunzioni fatte, se $y \in S_Y$ (inutile fare calcoli altrove), in definitiva si ha

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|. \quad (13.3)$$

Esempio 13.3 Sia $X \sim U(0, 1)$ e si consideri $Y = -\log X$, ben definita in $D = (0, 1]$, con $P_X(D) = 1$. Si vede subito che $S_Y = -\log(0, 1] = [0, +\infty)$. Dal momento che

$$\begin{aligned} y &= g(x) = -\log x \\ x &= g^{-1}(y) = e^{-y} \\ \frac{d}{dy} g^{-1}(y) &= -e^{-y} \end{aligned}$$

si ottiene che, per $y \in S_Y$,

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = 1 \cdot | -e^{-y} | = e^{-y},$$

per cui $Y \sim Esp(1)$. ◇

Esempio 13.4 Sia $X \sim Esp(1)$ e si consideri, per $\lambda > 0$,

$$Y = \frac{1}{\lambda} X,$$

ben definita in $D = S_X = [0, +\infty)$, con $P_X(D) = 1$. Si vede subito che $S_Y = [0, +\infty)$. Dal momento che

$$\begin{aligned} y &= g(x) = \frac{1}{\lambda} x \\ x &= g^{-1}(y) = \lambda y \\ \frac{d}{dy} g^{-1}(y) &= \lambda, \end{aligned}$$

si ottiene che, per $y \in S_Y$,

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = e^{-\lambda y} \lambda = \lambda e^{-\lambda y},$$

per cui $Y \sim Esp(\lambda)$. ◇

Esempio 13.5 Sia X con supporto S_X , f.r. $F_X(x)$ e, per $x \in S_X$, f.d.p.

$$p_X(x) = \frac{d}{dx} F_X(x).$$

Si consideri, per $a > 0$ e $b \in \mathbb{R}$,

$$Y = aX + b$$

ben definita in $D = S_X$, con $P_X(D) = 1$. Si vede subito che

$$S_Y = \{y \in \mathbb{R} : y = ax + b \text{ per } x \in S_X\}.$$

Dal momento che

$$\begin{aligned} y &= g(x) = ax + b \\ x &= g^{-1}(y) = \frac{y - b}{a} \\ \frac{d}{dy} g^{-1}(y) &= \frac{1}{a} \end{aligned}$$

si ottiene che, per $y \in S_Y$,

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = p_X\left(\frac{y - b}{a}\right) \frac{1}{a} = \frac{1}{a} p_X\left(\frac{y - b}{a}\right).$$

La f.r. di Y è

$$F_Y(y) = F_X\left(\frac{y - b}{a}\right).$$

◇

13.6 Le leggi del massimo e del minimo

Si richiama che la v.c. multivariata $X = (X_1, X_2, \dots, X_d)$ con f.r. congiunta $F_X(x) = F_{X_1, \dots, X_d}(x_1, \dots, x_d)$ ha componenti indipendenti se per ogni $x \in \mathbb{R}^d$

$$\begin{aligned} F_{X_1, \dots, X_d}(x_1, \dots, x_d) &= P(X_1 \leq x_1, \dots, X_d \leq x_d) \\ &= \prod_{i=1}^d P(X_i \leq x_i) \\ &= \prod_{i=1}^d F_{X_i}(x_i). \end{aligned}$$

È quindi facile specificare v.c. multivariate con componenti indipendenti. Basta prescrivere le leggi marginali delle componenti X_i , esplicitando le $F_{X_i}(x_i)$ per $i = 1, \dots, d$, e assumere che valga la fattorizzazione $F_{X_1, \dots, X_d}(x_1, \dots, x_d) = \prod_{i=1}^d F_{X_i}(x_i)$.

Ad esempio, nel bivariato, supposto $X_1 \sim \text{Esp}(\lambda)$ e $X_2 \sim \text{Esp}(\lambda)$, con X_1 e X_2 indipendenti, per cui $F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$, si ottiene

$$F_{X_1, X_2}(x_1, x_2) = \begin{cases} (1 - e^{-\lambda x_1})(1 - e^{-\lambda x_2}) & \text{se } x_1 > 0, x_2 > 0 \\ 0 & \text{altrimenti.} \end{cases}$$

Sia ora $X = (X_1, \dots, X_d)$ con componenti indipendenti e si consideri la trasformata **massimo**, $T = \max_{i=1, \dots, d} X_i = \max(X_1, \dots, X_d)$. Si ha

$$\begin{aligned} T \leq t &\iff X_i \leq t \quad \forall i \in \{1, \dots, d\} \\ &\iff X_1 \leq t, \dots, X_d \leq t \end{aligned}$$

quindi

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= P(X_1 \leq t, \dots, X_d \leq t) \\ &= P(X_1 \leq t)P(X_2 \leq t) \cdots P(X_d \leq t) \quad \text{per l'indipendenza} \\ &= \prod_{i=1}^d F_{X_i}(t). \end{aligned}$$

Se le componenti X_i della v.c. d -variata X sono indipendenti e identicamente distribuite (i.i.d.) si ha dunque

$$F_T(t) = F_{X_1}(t)^d.$$

Esempio 13.6 Sia $X = (X_1, \dots, X_n)$ con componenti i.i.d. $X_i \sim U(0, 1)$ e si consideri $T_n = \max_{i=1, \dots, n} X_i = \max(X_1, \dots, X_n)$. Si ha

$$F_{X_1}(x) = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } x \in [0, 1] \\ 1 & \text{se } x > 1 \end{cases}$$

per cui

$$F_{T_n}(t) = \begin{cases} 0 & \text{se } t < 0 \\ t^n & \text{se } t \in [0, 1] \\ 1 & \text{se } t > 1. \end{cases}$$

Quindi T_n ha supporto $[0, 1]$ e f.d.p.

$$p_{T_1}(t) = \begin{cases} nt^{n-1} & \text{se } t \in [0, 1] \\ 0 & \text{altrimenti.} \end{cases}$$

Si calcola subito che

$$\begin{aligned} E(T_n) &= \frac{n}{n+1} \\ E(T_n^2) &= \frac{n}{n+2} \\ \text{Var}(T_n) &= \frac{n}{(n+1)^2(n+2)} \doteq \frac{1}{n^2} \end{aligned}$$

per n grande. ◇

Sia poi $X = (X_1, \dots, X_d)$ con componenti indipendenti e si consideri la trasformata **minimo**, $T = \min_{i=1, \dots, d} X_i = \min(X_1, \dots, X_d)$. Si ha ora

$$\begin{aligned} T > t &\iff X_i > t \quad \forall i \in \{1, \dots, d\} \\ &\iff X_1 > t, \dots, X_d > t \end{aligned}$$

quindi, ragionando sulla funzione di sopravvivenza,

$$\begin{aligned} \bar{F}_T(t) &= P(T > t) \\ &= P(X_1 > t, \dots, X_d > t) \\ &= P(X_1 > t)P(X_2 > t) \cdots P(X_d > t) \quad \text{per l'indipendenza} \\ &= \prod_{i=1}^d P(X_i > t) \\ &= \prod_{i=1}^d \bar{F}_{X_i}(t) \end{aligned}$$

ossia

$$1 - F_T(t) = \prod_{i=1}^d (1 - F_{X_i}(t))$$

che dà, in termini di funzioni di ripartizione, la relazione

$$F_T(t) = 1 - \prod_{i=1}^d (1 - F_{X_i}(t)).$$

Se le componenti X_i della v.c. d -variata X sono i.i.d. (indipendenti e identicamente distribuite), si ottiene dunque per il minimo

$$F_T(t) = 1 - (1 - F_{X_1}(t))^d.$$

Teorema 13.1 *Il minimo di esponenziali indipendenti $X_i \sim \text{Esp}(\lambda_i)$, $i = 1, \dots, d$, ha ancora legge esponenziale, precisamente $\text{Esp}(\sum_{i=1}^d \lambda_i)$.*

Dimostrazione. Sia $X = (X_1, \dots, X_d)$ una v.c. multivariata con componenti indipendenti aventi legge marginale $X_i \sim \text{Esp}(\lambda_i)$, dove $\lambda_i > 0$, $i = 1, \dots, d$. Allora $T = \min_{i=1, \dots, d} X_i$ ha supporto

$$S_T = [0, +\infty)$$

in quanto il minimo di valori positivi è un valore positivo.
La f.r. di T , per $t > 0$, è

$$\begin{aligned} F_T(t) &= 1 - \prod_{i=1}^d (1 - F_{X_i}(t)) \\ &= 1 - \prod_{i=1}^d e^{-\lambda_i t} \\ &= 1 - \exp \left\{ - \sum_{i=1}^d \lambda_i t \right\} \\ &= 1 - \exp \left\{ - \left(\sum_{i=1}^d \lambda_i \right) t \right\}. \end{aligned}$$

In conclusione,

$$T \sim \text{Esp} \left(\sum_{i=1}^d \lambda_i \right),$$

risultato che riflette il fatto che il tasso di guasto del minimo di v.c. indipendenti è la somma dei tassi di guasto individuali delle v.c. in questione. \square

13.7 Esercizi risolti

Esempio 13.7 Una apparecchiatura dispone di due resistenze. La vita operativa X_i ($i = 1, 2$) di ciascuna di esse ha distribuzione esponenziale con valore atteso pari a 2 anni, indipendentemente dalla durata di corretto funzionamento dell'altra resistenza. Quando **almeno una** delle due resistenze è guasta, l'apparecchiatura non è più operativa. Si supponga, per semplicità, che le resistenze siano le uniche componenti soggette a guasto. Sia T il tempo di corretto funzionamento dell'apparecchiatura. Si esprima T come funzione di X_1, X_2 . Si dica qual è il supporto di T . Si ottengano poi la funzione di ripartizione e la funzione di densità di probabilità di T , esplicitandole in tutti i loro tratti. Si calcolino il primo quartile di T (è il quantile- p con $p = 1/4$) e la probabilità condizionale $P(T > 3 | T > 2)$.

L'apparecchiatura si guasta quando si guasta la resistenza più debole, ossia quella che ha la minima durata di corretto funzionamento. Quindi

$$T = \min(X_1, X_2).$$

Il supporto di un'esponenziale è $[0, +\infty)$. Poiché il minimo di due valori in $[0, +\infty)$ è ancora un valore in $[0, +\infty)$, si ha che $S_T = [0, +\infty)$.

Il valore atteso di un'esponenziale con tasso di guasto λ è $1/\lambda$. Qui $E(X_i) = 2$, $i = 1, 2$, quindi

$$\lambda = \frac{1}{E(X_i)} = \frac{1}{2}.$$

La funzione di ripartizione di T è, per $t > 0$,

$$\begin{aligned}
 F_T(t) &= P(T \leq t) = 1 - P(T > t) \\
 &= 1 - P(X_1 > t, X_2 > t) \\
 &= 1 - P(X_1 > t) P(X_2 > t) \quad \text{per l'indipendenza di } X_1 \text{ e } X_2 \\
 &= 1 - e^{-\frac{1}{2}t} e^{-\frac{1}{2}t} \\
 &= 1 - e^{-\frac{1}{2}t - \frac{1}{2}t} \\
 &= 1 - e^{-t}.
 \end{aligned}$$

Quindi $T \sim \text{Esp}(1)$. F.r. e f.d.p. esplicitate in tutti i tratti, sono

$$F_T(t) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-t} & \text{se } x \geq 0 \end{cases}$$

e

$$p_T(t) = \frac{d}{dt} F_T(t) = \begin{cases} 0 & \text{se } x < 0 \\ e^{-t} & \text{se } x > 0 \end{cases}$$

con $p_T(0) = 1$, ad esempio. Il primo quartile di T è soluzione dell'equazione $F_T(t) = 0.25$, ossia

$$\begin{aligned}
 1 - e^{-t} &= 0.25 \\
 e^{-t} &= 0.75 \\
 -t &= \log(0.75) \quad \text{passando ai logaritmi} \\
 \Rightarrow t_{0.25} &= -\log(0.75) \doteq 0.2876821.
 \end{aligned}$$

Infine,

$$\begin{aligned}
 P(T > 3 \mid T > 2) &= \frac{P(T > 3)}{P(T > 2)} \\
 &= \frac{e^{-3}}{e^{-2}} = e^{-1} \doteq 0.3678794.
 \end{aligned}$$

◇

Esempio 13.8 Una apparecchiatura dispone di tre resistenze. La vita operativa X_i ($i = 1, 2, 3$) di ciascuna di esse ha distribuzione uniforme continua in $(0, b)$, dove $b > 0$, con valore atteso pari a 5 anni, indipendentemente dalla durata di corretto funzionamento delle altre resistenze. Quando **tutte** le tre resistenze sono guaste, l'apparecchiatura non è più operativa. Si supponga, per semplicità, che le resistenze siano le uniche componenti soggette a guasto. Sia T il tempo di corretto funzionamento dell'apparecchiatura. Si esprima T come funzione di X_1, X_2, X_3 . Si dica qual è il supporto di T . Si ottenga poi la funzione di ripartizione di T , esplicitandola in tutti i suoi tratti. Si calcolino le probabilità $P(0 \leq T \leq 3)$ e $P(T > 8)$.

Il valore atteso di $X \sim U(a, b)$ è $(a + b)/2$, dunque, con i dati dell'esempio,

$5 = b/2$, per cui $b = 10$. L'apparecchiatura si guasta quando si guasta la resistenza più forte, ossia quella che ha la massima durata di corretto funzionamento. Quindi

$$T = \max(X_1, X_2, X_3).$$

Il supporto di $U(0, 10)$ è $[0, 10]$. Poiché il massimo di tre valori in $[0, 10]$ è ancora un valore in $[0, 10]$, si ha che $S_T = [0, 10]$.

La funzione di ripartizione di T è, per $t \in S_T$,

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= P(X_1 \leq t, X_2 \leq t, X_3 \leq t) \\ &= P(X_1 \leq t) P(X_2 \leq t) P(X_3 \leq t) && \text{per l'indipendenza delle } X_i \\ &= P(X_1 \leq t)^3 && \text{per l'identica distribuzione} \\ &= \left(\frac{t}{10}\right)^3. \end{aligned}$$

Quindi f.r. e f.d.p. esplicitate in tutti i loro tratti, sono

$$F_T(t) = \begin{cases} 0 & \text{se } x < 0 \\ \left(\frac{t}{10}\right)^3 & \text{se } 0 \leq x \leq 10 \\ 1 & \text{se } x > 10 \end{cases}$$

e

$$p_T(t) = \begin{cases} 3\frac{t^2}{1000} & \text{se } x \in [0, 10] \\ 0 & \text{se } x \notin [0, 10] \end{cases}$$

Infine

$$\begin{aligned} P(0 \leq T \leq 3) &= P(0 < T \leq 3) \\ &= F_T(3) - F_T(0) \\ &= F_T(3) = 0.3^3 = 0.027 \end{aligned}$$

e

$$\begin{aligned} P(T > 8) &= 1 - P(T \leq 8) \\ &= 1 - 0.8^3 = 0.488. \end{aligned}$$

◇

13.8 Esercizi

Esercizio 13.1 Sia $X \sim Bi(1, p)$ Si trovi il supporto e la f.m.p. di $T = X^2$.

Esercizio 13.2 Sia $X \sim Bi(2, 0.5)$ Si trovi il supporto e la f.m.p. di $T = 2X$.

Esercizio 13.3 Sia $X \sim Esp(1)$. Si ottengano supporto e f.r. di $T = X^2$.

Esercizio 13.4 Sia $X \sim Esp(\lambda)$, $\lambda > 0$. Per $T = \sqrt{X}$ si ottengano S_T e $F_T(t)$.

Esercizio 13.5 Si mostri che se $X \sim Esp(1)$ allora $T = X^{1/c} \sim W(c)$.

Esercizio 13.6 Si mostri che se $X \sim Ga(\alpha)$, dove $\alpha > 0$, allora, per $\lambda > 0$, $T = X/\lambda \sim Ga(\alpha, \lambda)$.

Esercizio 13.7 Si lanciano, con lanci indipendenti, un dado rosso e un dado verde. Sia X_1 il punteggio del dado rosso, X_2 il punteggio del dado verde. Si consideri la trasformata della v.c. con componenti indipendenti (X_1, X_2)

$$Y = g(X_1, X_2) = \min(X_1, X_2).$$

Si ottengano S_Y e $p_Y(y)$.

Esercizio 13.8 Siano $X_1 \sim Bi(1, 0.5)$ e $X_2 \sim Bi(1, 0.5)$ due v.c. indipendenti. Si trovi il supporto e la f.m.p. di $S = X_1 + X_2$.

Esercizio 13.9 Siano $X_1 \sim Bi(1, p)$ e $X_2 \sim Bi(1, p)$ due v.c. indipendenti. Si trovi il supporto e la f.m.p. di $S = X_1 + X_2$.

Esercizio 13.10 Siano X_1 e X_2 due tempi d'attesa (quindi $P(X_1 \geq 0) = 1$ e $P(X_2 \geq 0) = 1$) con legge continua, f.d.p. $p_{X_1}(x_1)$ e $p_{X_2}(x_2)$, rispettivamente. Si mostri che $T = \min(X_1, X_2)$ ha funzione tasso di guasto $r_T(t) = r_{X_1}(t) + r_{X_2}(t)$.

Esercizio 13.11 (*compito del 01/02/17*) Una apparecchiatura ha solo tre componenti che si possono guastare. La vita operativa X_i ($i = 1, 2, 3$) di ciascuna di esse ha distribuzione esponenziale con valore atteso pari a 12 anni, indipendentemente dalla durata di corretto funzionamento delle altre. Quando **almeno una** delle tre è guasta, l'apparecchiatura non è più operativa. Sia T il tempo di corretto funzionamento dell'apparecchiatura. Si esprima T come funzione di X_1, X_2, X_3 . Si dica qual è il supporto di T . Si ottengano poi la funzione di ripartizione e la funzione di densità di probabilità di T , esplicitandole in tutti i loro tratti. Si calcolino la mediana di T (è il quantile- p con $p = 50/100$) e la probabilità condizionale $P(T > 8 | T > 4)$.

Esercizio 13.12 (*compito del 10/02/17*) Una apparecchiatura ha solo due componenti che si possono guastare. La vita operativa X_i ($i = 1, 2$) di ciascuna di esse ha distribuzione esponenziale con valore atteso pari a 10 anni, indipendentemente dalla durata di corretto funzionamento dell'altra. Quando **almeno una** delle due è guasta, l'apparecchiatura non è più operativa. Sia T il tempo di corretto funzionamento dell'apparecchiatura. Si esprima T come funzione di X_1, X_2 . Si dica qual è il supporto di T . Si ottengano poi la funzione di ripartizione e la funzione di densità di probabilità di T , esplicitandole in tutti i loro tratti. Si calcolino il terzo quartile di T (è il quantile- p con $p = 75/100$) e la probabilità condizionale $P(T > 10 | T > 5)$.

Esercizio 13.13 (*compito del 16/06/17*) Una apparecchiatura ha solo due componenti che si possono guastare. La vita operativa X_i ($i = 1, 2$) di ciascuna di esse ha distribuzione esponenziale con valore atteso pari a 10 anni, indipendentemente dalla durata di corretto funzionamento dell'altra. Quando **entrambe** sono guaste, l'apparecchiatura non è più operativa. Sia T il tempo di corretto funzionamento dell'apparecchiatura. Si esprima T come funzione di X_1, X_2 . Si dica qual è il supporto di T . Si ottengano poi la funzione di ripartizione e la funzione di densità di probabilità di T , esplicitandole in tutti i loro tratti. Si calcolino la mediana di T (è il quantile- p con $p = 50/100$) e $P(T > 10)$.

Esercizio 13.14 (*compito del 17/07/17*) Una apparecchiatura ha solo tre componenti che si possono guastare. La vita operativa X_i ($i = 1, 2, 3$) di ciascuna di esse ha distribuzione esponenziale con valore atteso pari a 9 anni, indipendentemente dalla durata di corretto funzionamento dell'altra. Quando **almeno una** delle tre è guasta, l'apparecchiatura non è più operativa. Sia T il tempo di corretto funzionamento dell'apparecchiatura. Si esprima T come funzione di X_1, X_2, X_3 . Si dica qual è il supporto di T . Si ottengano poi la funzione di ripartizione e la funzione di densità di probabilità di T , esplicitandole in tutti i loro tratti. Si calcolino il novantesimo percentile di T (è il quantile- p con $p = 90/100$) e la probabilità condizionale $P(T > 9 | T > 6)$.

Esercizio 13.15 (*compito del 11/09/17*) Una apparecchiatura ha solo due componenti che si possono guastare. La vita operativa X_i ($i = 1, 2$) di ciascuna di esse ha distribuzione esponenziale con valore atteso pari a 8 anni, indipendentemente dalla durata di corretto funzionamento dell'altra. Quando **almeno una** delle due è guasta, l'apparecchiatura non è più operativa. Sia T il tempo di corretto funzionamento dell'apparecchiatura. Si esprima T come funzione di X_1, X_2 . Si dica qual è il supporto di T . Si ottengano poi la funzione di ripartizione e la funzione di densità di probabilità di T , esplicitandole in tutti i loro tratti. Si calcolino il primo quartile di T (è il quantile- p con $p = 25/100$) e la probabilità condizionale $P(T > 2 | T > 1)$.

13.9 Appendice: v.c. trasformate con R

Grazie alla simulazione, R permette di approssimare la legge di una variabile casuale trasformata $T = g(X)$, semplicemente trasformando tramite $g(\cdot)$ i valori realizzati di X , x_r^* , $r = 1, \dots, R$. Si ottengono così i valori $t_r^* = g(x_r^*)$, $r = 1, \dots, R$, realizzazioni di T . La frequenza relativa di tali realizzazioni che cade in un dato intervallo (a, b) approssima $P(T \in (a, b))$ se R è molto grande, seguendo la concezione frequentista della probabilità.

Come primo esempio, si consideri la v.c. bivariata (X, Y) con componenti indipendenti e leggi marginali poissoniane, $X \sim P(1)$ e $Y \sim P(2)$. Le v.c. X e Y rappresentano conteggi senza limitazione superiore, dove le probabilità dei vari conteggi osservabili sono determinate dal valore atteso del conteggio. Anche la v.c. trasformata $T = X + Y$ rappresenta un conteggio senza limitazione

superiore. Poiché $E(T) = E(X+Y) = E(X)+E(Y) = 3$, sembra plausibile che T abbia ancora legge di Poisson, con parametro 3. Per verificare numericamente la congettura basta confrontare le frequenze relative dei valori di T ottenute dalla simulazione con le probabilità degli stessi valori offerte dalla Poisson $P(3)$.

```
> set.seed(12345)
> nrep=10^4
> x=rpois(nrep, 1)
> y=rpois(nrep, 2)
> sum=x+y
> table(sum)/nrep
sum
 0      1      2      3      4      5      6      7      8
0.0473 0.1474 0.2289 0.2293 0.1664 0.1005 0.0460 0.0220 0.0090
> s=0:8
> p_poisson_3=round(exp(-3)*3^ss/factorial(ss), digits=4)
> p_poisson_3
[1] 0.0498 0.1494 0.2240 0.2240 0.1680 0.1008 0.0504 0.0216 0.0081
```

L'accordo tra le due distribuzioni sembra molto buono. Si dimostrerà nella Lezione 16 che la somma di due v.c. indipendenti con legge di Poisson ha effettivamente ancora legge di Poisson.

Come secondo esempio, si desidera dare una verifica tramite simulazione del fatto che il minimo di esponenziali indipendenti ha ancora legge esponenziale, con tasso di guasto che è la somma dei tassi di guasto delle v.c. esponenziali considerate. In particolare, si verifica la proprietà per $T = \min(X_1, X_2, X_3)$, dove le componenti sono indipendenti e $X_i \sim \text{Esp}(i)$, $i = 1, 2, 3$, confrontando la f.r. dei valori simulati da T (funzione di ripartizione empirica: è la f.r. di $Ud(t_1, \dots, t_R)$) con la f.r. di $\text{Esp}(6)$.

```
> set.seed(12345); nrep=10^6
> x1=rexp(nrep,rate=1); x2=rexp(nrep,rate=2); x3=rexp(nrep,rate=3)
> min=pmin(x1,x2,x3); min=sort(min) # pmin: parallel minimum
> t=seq(0,6,0.001); funz_rip_t=rep(NA,length(t))
> funz_rip_esp_6=rep(NA,length(t))
> for (i in 1:length(t)){
+ funz_rip_t[i] = length(min[min<=t[i]])/nrep
+ funz_rip_esp_6[i] = 1-exp(-6*t[i])
+ }
> max(abs(funz_rip_t - funz_rip_esp_6))
[1] 0.0006633658 # accordo eccellente
```

Lezione 14

Indici di posizione e proprietà di $E(X)$

14.1 Indici di posizione di v.c. univariate

Un indice di posizione di una v.c. X è una predizione ‘puntuale’ del valore futuro di X . Tramite un indice di posizione, si predice dunque un punto, la realizzazione futura x di X , con un punto, il valore dell’indice. Il principale indice di posizione per una v.c. univariata X è il valore atteso, $E(X)$. Altri indici di posizione importanti, sempre per v.c. univariate, sono la moda, la mediana e i quantili- p .

14.1.1 La moda

Definizione 14.1 Sia X una v.c. con legge discreta o continua, con f.m.p./f.d.p. $p_X(x)$. Si dice **moda** di X , indicata con x_{mod} , un valore del supporto di X per cui

$$p_X(x_{mod}) \geq p_X(x) \quad \text{per ogni } x \in S_X.$$

Nel caso continuo, si richiede inoltre che la densità di X sia continua almeno da destra o da sinistra in x_{mod} .

Si osservi che la definizione, con opportune modifiche, funziona anche se X è multivariata. La predizione del valore futuro di X insita in x_{mod} è ottenuta con il criterio del valore ‘più probabile’. Nel caso continuo, la massima probabilità è riferita ad intervalli di lunghezza ε contenenti x_{mod} . Conviene considerare alcuni esempi.

- $X \sim Bi(2, 0.5)$ ha supporto $S_X = \{0, 1, 2\}$ e f.m.p. $p_X(0) = 1/4$, $p_X(1) = 1/2$, $p_X(2) = 1/4$, per cui $x_{mod} = 1$.

- $X \sim Bi(3, 0.5)$ ha supporto $S_X = \{0, 1, 2, 3\}$ e f.m.p. $p_X(0) = 1/8$, $p_X(1) = 3/8$, $p_X(2) = 3/8$, $p_X(3) = 1/8$, per cui $x_{mod} = 1$ e $x_{mod} = 2$ (la moda non è necessariamente unica: possono esserci più mode di X).
- $X \sim Ge(p)$, $p \in (0, 1)$, ha supporto $S_X = \{1, 2, \dots\}$ e f.m.p. $p_X(1) = p$, $p_X(2) = p(1-p) < p$, $p_X(3) = p(1-p)^2 < p$, eccetera, per cui $x_{mod} = 1$.
- $X \sim Esp(\lambda)$, $\lambda > 0$, ha supporto $S_X = [0, +\infty)$ e f.d.p. $p_X(x) = \lambda e^{-\lambda x}$ per $x \geq 0$. Poiché la funzione e^{-x} per $x \geq 0$ è monotona decrescente, $x_{mod} = 0$, intendendo ovviamente 0^+ .

14.1.2 La mediana

Definizione 14.2 Sia X una v.c. univariata con f.r. $F_X(x)$. Si dice **mediana** di X , indicata con $x_{0.5}$, un valore reale tale che valgano simultaneamente

$$P(X \leq x_{0.5}) \geq 0.5 \quad e \quad P(X \geq x_{0.5}) \geq 0.5.$$

La predizione del valore futuro di X data da $x_{0.5}$ è ottenuta con il criterio di spezzare in due le possibilità, essendo le probabilità di previsione in eccesso ($X \leq x_{0.5}$) e di predizione in difetto ($X \geq x_{0.5}$) grosso modo bilanciate, 0.5 o più. Si osservi che nel caso discreto può essere che $P(X = x_{0.5}) > 0$, e ciò spiega le disequaglianze nella definizione.

Anche la mediana non è necessariamente unica. Tutte le soluzioni dell'equazione $F_X(x) = 0.5$ sono mediana di X . D'altro canto, se l'equazione $F_X(x) = 0.5$ non ha soluzioni, la mediana di X è unica ed è il più piccolo valore di x per cui $F_X(x) \geq 0.5$. Conviene considerare alcuni esempi.

- $X \sim Bi(2, 0.5)$ ha supporto $S_X = \{0, 1, 2\}$, f.m.p. $p_X(0) = 1/4$, $p_X(1) = 1/2$, $p_X(2) = 1/4$, e f.r. nei punti del supporto $F_X(0) = 1/4$, $F_X(1) = 3/4$, $F_X(2) = 1$, per cui $x_{0.5} = 1$.
- $X \sim Bi(3, 0.5)$ ha supporto $S_X = \{0, 1, 2, 3\}$, f.m.p. $p_X(0) = 1/8$, $p_X(1) = 3/8$, $p_X(2) = 3/8$, $p_X(3) = 1/8$, f.r. nei punti del supporto $F_X(0) = 1/8$, $F_X(1) = 1/2$, $F_X(2) = 7/8$, $F_X(3) = 1$, per cui $x_{0.5} \in [1, 2)$.
- X con f.d.p. $p_X(x) = 0.5e^{-|x|}$ ha f.r. $F_X(x) = 0.5e^x$ se $x \leq 0$ e $F_X(x) = 1 - 0.5e^{-x}$ se $x > 0$, per cui l'equazione $F_X(x) = 0.5$ ha un'unica soluzione, $x_{0.5} = 0$. Più in generale se la densità $p_X(x)$ è simmetrica attorno a \tilde{x} , ossia $p_X(\tilde{x} + \Delta) = p_X(\tilde{x} - \Delta)$ per ogni $\Delta > 0$, allora la mediana è unica, $x_{0.5} = \tilde{x}$.
- $X \sim Esp(\lambda)$, $\lambda > 0$, ha supporto $S_X = [0, +\infty)$ e f.r. per $x \in S_X$ pari a $F_X(x) = 1 - e^{-\lambda x}$. Poiché $F_X(x)$ è continua e monotona crescente per $x \geq 0$, la mediana di X è l'unica soluzione dell'equazione $F_X(x) = 0.5$. Ora $1 - e^{-\lambda x} = 0.5$ equivale a $e^{-\lambda x} = 0.5$ ossia, passando ai logaritmi, $-\lambda x = \log 0.5 = -\log 2 \doteq 0.69$, che dà $x_{0.5} \doteq 0.69/\lambda$.

14.1.3 Il quantile- p

La bipartizione bilanciata 0.5 e 0.5 della probabilità totale è effettuata dalla mediana e generalizzata dalla bipartizione sbilanciata p e $1-p$. Così si perviene ai quantili.

Definizione 14.3 Sia X una v.c. univariata con f.r. $F_X(x)$. Per $p \in (0, 1)$ si dice **quantile- p** di X , indicato con x_p , un valore reale tale che valgano simultaneamente

$$P(X \leq x_p) \geq p \quad e \quad P(X \geq x_p) \geq 1 - p.$$

Tra i quantili- p spiccano alcuni quantili che per il loro uso frequente hanno meritato un nome specifico. Per

- $p = 1/2$, il quantile-0.5 è detto **mediana**
- $p = i/4$, $i = 1, 2, 3$, il quantile- $i/4$ è detto i -esimo **quartile**
- $p = i/10$, $i = 1, 2, \dots, 9$, il quantile- $i/10$ è detto i -esimo **decile**
- $p = i/100$, $i = 1, 2, \dots, 99$, il quantile- $i/100$ è detto i -esimo **percentile**.

Come la mediana, anche il quantile- p non è necessariamente unico. Tutte le soluzioni dell'equazione $F_X(x) = p$ sono quantile- p di X . D'altro canto, se l'equazione $F_X(x) = p$ non ha soluzioni, il quantile- p di X è unico ed è il più piccolo valore di x per cui $F_X(x) \geq p$.

Esempio 14.1 Sia $X \sim Esp(\lambda)$, $\lambda > 0$, con supporto $S_X = [0, +\infty)$ e f.r. per $x \in S_X$ pari a $F_X(x) = 1 - e^{-\lambda x}$. Poiché $F_X(x)$ è continua e monotona crescente per $x \geq 0$, il quantile- p di X è l'unica soluzione dell'equazione $F_X(x) = p$. Quindi, con facili passaggi,

$$\begin{aligned} 1 - e^{-\lambda x} &= p \\ 1 - p &= e^{-\lambda x} \\ \log(1 - p) &= -\lambda x \end{aligned}$$

che dà

$$x_p = -\frac{\log(1 - p)}{\lambda}.$$

I quartili di $Exp(\lambda)$ sono dunque

- primo quartile $\log(4/3)/\lambda \doteq 0.29/\lambda$
- mediana $\log(2)/\lambda \doteq 0.69/\lambda$
- terzo quartile $\log(4)/\lambda \doteq 1.39/\lambda$.

◇

14.2 Il valore atteso

Conviene anzitutto richiamare la definizione.

Definizione 14.4 Sia X una v.c. univariata con legge discreta o continua, con supporto S_X e f.m.p./f.d.p. $p_X(x)$. Si dice **valore atteso** di X , indicato con $E(X)$, il valore reale

$$E(X) = \begin{cases} \sum_{x \in S_X} x p_X(x) & \text{se } X \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} x p_X(x) dx & \text{se } X \text{ ha legge continua} \end{cases}$$

purché la serie o l'integrale convergano assolutamente, ossia

$$\begin{cases} \sum_{x \in S_X} |x| p_X(x) < +\infty & \text{se } X \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} |x| p_X(x) dx < +\infty & \text{se } X \text{ ha legge continua.} \end{cases}$$

Si osservi che, con opportune modificazioni, la definizione funziona anche se X è bivariata o multivariata. Diversamente da moda, mediana e quantili, l'esistenza del valore atteso non è garantita. In compenso, il valore atteso di una v.c., se esiste finito, è unico.

14.3 Proprietà del valore atteso

Si considerano nel seguito unicamente v.c. con valore atteso finito. Quindi se si scrive $E(X)$ si sottintende che la v.c. X ammette valore atteso finito.

14.3.1 Prime proprietà, utili per il calcolo

a) Il valore atteso di trasformate

Siano X e Y v.c. univariate con $Y = g(X)$. Allora si ha, seguendo la definizione,

$$E(Y) = \begin{cases} \sum_{y \in S_Y} y p_Y(y) & \text{se } Y \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} y p_Y(y) dy & \text{se } Y \text{ ha legge continua} \end{cases}$$

ma vale anche

$$E(Y) = E(g(X)) = \begin{cases} \sum_{x \in S_X} g(x) p_X(x) & \text{se } X \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} g(x) p_X(x) dx & \text{se } X \text{ ha legge continua.} \end{cases}$$

In pratica, per calcolare $E(g(X))$ non si deve necessariamente reperire prima la legge di $Y = g(X)$. Si osservi che la proprietà, con le opportune modificazioni, ha senso (e si dimostra che vale) anche se X è bivariata o multivariata, e così pure se Y è bivariata o multivariata. È facile verificarla per X discreta univariata:

$$\sum_{x \in S_X} g(x) p_X(x) = \sum_{y \in S_Y} y \sum_{x \in S_X : g(x)=y} p_X(x) = \sum_{y \in S_Y} y p_Y(y).$$

b) Il valore atteso del prodotto di v.c. indipendenti

Sia $X = (X_1, \dots, X_d)$ una v.c. con componenti indipendenti, ossia con

$$S_X = S_{X_1, X_2, \dots, X_d} = S_{X_1} \times S_{X_2} \times \dots \times S_{X_d}$$

e con

$$p_X(x) = p_{X_1, \dots, X_d}(x_1, \dots, x_d) = \prod_{i=1}^d p_{X_i}(x_i).$$

Allora

$$E(X_1 X_2 \dots X_d) = \prod_{i=1}^d E(X_i)$$

e, più in generale, per g_i funzioni reali misurabili, $i = 1, \dots, d$,

$$E(g_1(X_1)g_2(X_2) \dots g_d(X_d)) = \prod_{i=1}^d E(g_i(X_i)).$$

Motto per memorizzare: sotto indipendenza la media del prodotto è il prodotto delle medie. La proprietà è di verifica immediata. Supponendo X bivariata con legge discreta, si ha

$$\begin{aligned} E(g_1(X_1)g_2(X_2)) &= \sum_{(x_1, x_2) \in S_{X_1, X_2}} g_1(x_1)g_2(x_2)p_{X_1, X_2}(x_1, x_2) \\ &= \sum_{(x_1, x_2) \in S_{X_1} \times S_{X_2}} g_1(x_1)g_2(x_2)p_{X_1}(x_1)p_{X_2}(x_2) \text{ per l'indip.} \\ &= \sum_{x_1 \in S_{X_1}} \sum_{x_2 \in S_{X_2}} g_1(x_1)g_2(x_2)p_{X_1}(x_1)p_{X_2}(x_2) \\ &= \sum_{x_1 \in S_{X_1}} g_1(x_1)p_{X_1}(x_1) \sum_{x_2 \in S_{X_2}} g_2(x_2)p_{X_2}(x_2) \text{ raccogliendo} \\ &= E(g_1(X_1))E(g_2(X_2)). \end{aligned}$$

14.3.2 Ulteriori proprietà

1. La proprietà di Cauchy

Quando esiste finito, il valore atteso della v.c. univariata X può non essere un punto del supporto di X , ma è sempre intermedio fra punti del supporto:

$$\inf(S_X) \leq E(X) \leq \sup(S_X). \quad (14.1)$$

Ci si limita a verificare la (14.1) per X con legge discreta e supporto finito $S_X = \{x_1, \dots, x_k\}$. Supponendo, senza perdita di generalità,

$$x_1 < x_2 < \dots < x_k$$

si ha

$$x_1 \leq x_i \leq x_k \text{ per ogni } i \in \{1, 2, \dots, k\}$$

e quindi

$$x_1 p_X(x_i) \leq x_i p_X(x_i) \leq x_k p_X(x_i) \text{ per ogni } i \in \{1, 2, \dots, k\}.$$

Sommando le disuguaglianze si ottiene

$$x_1 \sum_{i=1}^k p_X(x_i) \leq \sum_{i=1}^k x_i p_X(x_i) \leq x_k \sum_{i=1}^k p_X(x_i)$$

da cui, poiché $\sum_{i=1}^k p_X(x_i) = 1$ per la normalizzazione,

$$x_1 \leq E(X) \leq x_k.$$

Per concludere basta osservare che $\inf(S_X) = \min(S_X) = x_1$ e analogamente per x_k .

2. La proprietà di linearità

Siano X e Y v.c. univariate con $Y = aX + b$. Allora

$$E(Y) = E(aX + b) = aE(X) + b. \quad (14.2)$$

Per la dimostrazione, si supponga senza perdita di generalità che X abbia legge continua. Allora

$$\begin{aligned} E(aX + b) &= \int_{-\infty}^{+\infty} (ax + b) p_X(x) dx \\ &= \int_{-\infty}^{+\infty} ax p_X(x) dx + \int_{-\infty}^{+\infty} b p_X(x) dx \\ &= a \int_{-\infty}^{+\infty} x p_X(x) dx + b \int_{-\infty}^{+\infty} p_X(x) dx \\ &= aE(X) + b. \end{aligned}$$

3. La proprietà di baricentro

Si tratta di un caso particolare della (14.2), relativo alla variabile scarto $Y = X - E(X)$:

$$E(X - E(X)) = 0. \quad (14.3)$$

Il motto per la memorizzazione è: la media degli scarti di X dalla propria media è zero. Poiché gli scarti da $E(X)$ in media si compensano, $E(X)$ funge da baricentro della distribuzione di X .

4. La proprietà di linearità (generalizzazione)

Se (X, Y) è una v.c. bivariata e $T = aX + bY$ è una combinazione lineare delle componenti di (X, Y) con $a, b \in \mathbb{R}$, allora

$$E(T) = E(aX + bY) = aE(X) + bE(Y). \quad (14.4)$$

Più in generale, se $X = (X_1, X_2, \dots, X_d)$ è una v.c. multivariata, il valore atteso della combinazione lineare è la combinazione lineare dei valori attesi, ossia per ogni $a = (a_1, \dots, a_d) \in \mathbb{R}^d$

$$E\left(\sum_{i=1}^d a_i X_i\right) = \sum_{i=1}^d a_i E(X_i). \quad (14.5)$$

Per dimostrare la (14.4), si supponga senza perdita di generalità che (X, Y) abbia legge discreta. Allora

$$\begin{aligned} E(aX + bY) &= \sum_{(x,y) \in S_{X,Y}} (ax + by) p_{X,Y}(x, y) \\ &= a \sum_{(x,y) \in S_{X,Y}} x p_{X,Y}(x, y) + b \sum_{(x,y) \in S_{X,Y}} y p_{X,Y}(x, y) \\ &= a \sum_{(x,y) \in S_{X,Y}} x p_X(x) p_{Y|X=x}(y) + b \sum_{(x,y) \in S_{X,Y}} y p_Y(y) p_{X|Y=y}(x) \end{aligned}$$

e

$$\begin{aligned} \sum_{(x,y) \in S_{X,Y}} x p_X(x) p_{Y|X=x}(y) &= \sum_{x \in S_X} \sum_{y \in S_{Y|X=x}} x p_X(x) p_{Y|X=x}(y) \\ &= \sum_{x \in S_X} x p_X(x) \sum_{y \in S_{Y|X=x}} p_{Y|X=x}(y) \\ &= E(X), \end{aligned}$$

poiché $\sum_{y \in S_{Y|X=x}} p_{Y|X=x}(y) = 1$ per la normalizzazione, e analogamente

$$\sum_{(x,y) \in S_{X,Y}} y p_Y(y) p_{X|Y=y}(x) = E(Y).$$

4. La proprietà dei minimi quadrati

Se X è una v.c. univariata e i valori attesi indicati esistono, allora per ogni $c \in \mathbb{R}$

$$E((X - c)^2) \geq E((X - E(X))^2) \quad (14.6)$$

dove vale l'eguaglianza solo per $c = E(X)$.

Per valutare l'importanza della (14.6), conviene considerare che

- c è una predizione puntuale della realizzazione futura di X
- $X - c$ è l'errore di predizione
- $(X - c)^2$ è la perdita quadratica dovuta all'errore di predizione
- $E((X - c)^2)$ è la perdita quadratica media, detta rischio quadratico

per cui il rischio quadratico della predizione di X è minimizzato dalla predizione $c = E(X)$. La (14.6) è quindi la proprietà del valore atteso più concettuale: dà una ragione per scegliere la media come indice di posizione.

Per dimostrare la (14.6), si si consideri lo sviluppo del quadrato del binomio

$$\begin{aligned}(X - c)^2 &= (X - E(X) + E(X) - c)^2 \\ &= (X - E(X))^2 + (E(X) - c)^2 + 2(E(X) - c)(X - E(X)).\end{aligned}$$

Per la proprietà di linearità del valore atteso

$$\begin{aligned}E((X - c)^2) &= E((X - E(X))^2) + (E(X) - c)^2 + 2(E(X) - c)E((X - E(X))) \\ &= E((X - E(X))^2) + (E(X) - c)^2\end{aligned}$$

poiché $E((X - E(X))) = 0$ per la proprietà di baricentro. La tesi si consegue osservando che $(E(X) - c)^2 \geq 0$ per ogni $c \in \mathbb{R}$ con eguaglianza a 0 solo per $E(X) - c = 0$ ossia solo per $c = E(X)$.

14.4 Esercizi

Esercizio 14.1 Si calcolino moda, mediana e valore atteso di $X \sim Bi(4, 0.5)$.

Esercizio 14.2 Si calcolino moda, mediana e valore atteso della v.c. X con f.d.p.

$$p_X(x) = \begin{cases} 1 + x & \text{se } x \in (-1, 0] \\ 1 - x & \text{se } x \in (0, 1) \\ 0 & \text{altrove.} \end{cases}$$

Esercizio 14.3 Si calcolino moda, mediana e valore atteso della v.c. X con f.d.p.

$$p_X(x) = \frac{1}{2}|x|e^{-|x|}.$$

Esercizio 14.4 Sia $T = 3X + 5$. Si calcoli il valore atteso di T sapendo che il valore atteso di X è 2.

Esercizio 14.5 Sia (X, Y) una v.c. bivariata con $E(X) = E(Y) = 1$. Si calcoli il valore atteso di $T = 3X - 2Y$.

Esercizio 14.6 Sia (X, Y) una v.c. bivariata con componenti indipendenti e leggi marginali $X \sim Esp(2)$ e $Y \sim Esp(0.5)$. Si calcoli il valore atteso di $T = XY$.

14.5 Appendice: quantili con R

Per una v.c. X con legge data `data distr`, il comando per calcolare valori della funzione di densità o di massa di probabilità è `ddistr`. Si calcolano valori della funzione di ripartizione con `pdistr`. Il comando `rdistr` fornisce realizzazioni da X . Come si può intuire, quantili di X sono ottenuti con il comando `qdist`. Eccone alcuni esempi d'uso.

```
> qbinom(c(0.25,0.50,0.75), size=30, p=0.4)      # quartili
[1]  7 14 22
> qpois(c(0.25,0.50,0.75), lambda=5)
[1]  3 5 6
> qpois(c(0.25,0.50,0.75), lambda=50)
[1] 45 50 55
> qpois(seq(0.1,0.9,0.1), lambda=10)             # decili
[1]  6  7  8  9 10 11 12 13 14
> round(qexp(seq(0.1,0.9,0.1), rate=1), digits=4)
[1] 0.1054 0.2231 0.3567 0.5108 0.6931 0.9163 1.2040 1.6094 2.3026
> perc=qgamma(seq(0.90,0.99,0.01), shape=5,rate=1) # percentili
> round(perc, digits=2)
[1]  7.99  8.18  8.38  8.60  8.86  9.15  9.51  9.96 10.58 11.60
```

Come la media e la varianza, mediana e quantili si possono calcolare anche per un vettore di dati y . Si parla allora di quantili empirici, ossia quantili della distribuzione empirica. Il comando per calcolare quantili empirici è `quantile`. Il suo uso è subito illustrato.

```
> set.seed(12345)
> y=rexp(10^3,rate=1)
> quantile(y, probs=c(0.01, 0.25,0.5,0.75,0.99))
          1%          25%          50%          75%          99%
0.009985524 0.286127719 0.630665346 1.260517887 4.371488621
```

Collegato ai quartili, e a quantili estremi sulle code della distribuzione empirica, è un grafico noto come *box and whiskers plot*, o semplicemente *boxplot*. La scatola è definita da primo e terzo quartile e tagliata in corrispondenza della mediana. I baffi si estendono fino ai valori minimo e massimo delle osservazioni, o a quantili estremi della distribuzione. Il grafico viene realizzato dal comando `boxplot`. Ecco un esempio in cui si confrontano graficamente le distribuzioni di dati ottenuti da due v.c. esponenziali, la prima con valore atteso $1/2$, la seconda con valore atteso 1.

```
> set.seed(12345)
> x=rexp(10^3,rate=2)
> y=rexp(10^3,rate=1)
> boxplot(x,y)
```


Lezione 15

Indici di variabilità e proprietà di $\text{Var}(X)$

15.1 Indici di variabilità

Un indice di variabilità di una v.c. univariata X è una predizione ‘puntuale’ di quanto disterà il valore futuro di X da una sua particolare predizione puntuale, quindi in sostanza un indice di prevedibilità di X . Il principale indice di variabilità per una v.c. univariata X è la varianza, $\text{Var}(X)$, assieme alla sua trasformazione espressa nella stessa unità di misura di X , lo scarto quadratico medio $\sqrt{\text{Var}(X)}$. Altri indici di variabilità, sempre per v.c. univariate, sono il *range* e lo scarto interquartilico.

Anche per v.c. bivariate e multivariate si considerano indici di variabilità. Il principale è una generalizzazione multidimensionale della varianza, la matrice delle varianze e covarianze.

15.2 Varianza e scarto quadratico medio

Conviene iniziare richiamando la definizione della varianza e presentando l'indice di variabilità ad essa strettamente collegato, lo scarto quadratico medio. La proprietà dei minimi quadrati del valore atteso, formula (14.6), mostra che il rischio quadratico (media della perdita quadratica) associato a prevedere X con $E(X)$,

$$E((X - E(X))^2),$$

è minimo fra quelli in cui si incorre prevedendo X con una costante c . Il rischio quadratico minimo $E((X - E(X))^2)$ è dunque un naturale indice di prevedibilità del valore futuro di X , detto **varianza** di X e indicato con $\text{Var}(X)$, o anche

$V(X)$, σ_X^2 o semplicemente σ^2 . Quindi, a parole, la varianza di X è la media del quadrato dello scarto di X dalla propria media. In simboli,

$$\text{Var}(X) = E((X - E(X))^2). \quad (15.1)$$

Esplicitando la definizione (15.1) si ha:

$$\text{Var}(X) = \begin{cases} \sum_{x \in S_X} (x - E(X))^2 p_X(x) & \text{se } X \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} (x - E(X))^2 p_X(x) dx & \text{se } X \text{ ha legge continua.} \end{cases} \quad (15.2)$$

La varianza è il principale indice di variabilità per una v.c. univariata, dal momento che è l'indice di variabilità naturalmente associato al principale indice di posizione, il valore atteso.

Equivalente alla varianza come indice di variabilità è lo **scarto quadratico medio** o **deviazione standard**, definito come la radice quadrata aritmetica della varianza (l'unica non negativa), indicato con σ_X o semplicemente σ . In simboli,

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

Data la varianza si può calcolare lo scarto quadratico medio e, viceversa, dato lo scarto quadratico medio si può calcolare la varianza. Per questo i due indici sono equivalenti. D'altra parte va sottolineato che $\text{Var}(X)$ è espressa nel quadrato dell'unità di misura di X , mentre σ_X , come $\mu_X = E(X)$, è nell'unità di misura di X . Questa omogeneità è il principale vantaggio dello scarto quadratico medio rispetto alla varianza. Ad esempio i punti $\mu \pm \sigma_X$ indicano la dispersione media di X attorno a μ_X , mentre i punti $\mu \pm \sigma_X^2$ non hanno interpretazione.

15.3 Altri indici di variabilità

Il **range** di una v.c. univariata X , indicato con R_X , è definito per v.c. con supporto limitato come

$$R_X = \max(S_X) - \min(S_X).$$

È quindi molto semplice da calcolare. Ad esempio, per $X \sim U(a, b)$ si ha $R_X = b - a$.

Lo **scarto interquartilico** (*interquartile range*) di una v.c. univariata X , indicato con IQR_X , è definito come

$$IQR_X = x_{0.75} - x_{0.25}.$$

È anch'esso di calcolo facile. Ad esempio, per $X \sim \text{Esp}(\lambda)$, $\lambda > 0$, si ottiene facilmente $IQR_X = \log 3/\lambda$.

Sia il *range* che lo scarto interquartilico sono espressi nella stessa unità di misura di X .

15.4 Proprietà immediate della varianza

La varianza, o equivalentemente lo scarto quadratico medio, ha alcune proprietà notevoli.

0. Non negatività

Per ogni X con varianza finita $\text{Var}(X) \geq 0$ e $\text{Var}(X) = 0$ solo per $X \sim \mathcal{D}(x_0)$.

La dimostrazione della non-negatività è immediata dalla proprietà di Cauchy del valore atteso. Infatti $\text{Var}(X)$ è il valore atteso della v.c. $(X - \text{E}(X))^2$ il cui supporto è un sottoinsieme di $[0, +\infty)$.

1. Formula per il calcolo

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

La dimostrazione parte dalla definizione della varianza:

$$\begin{aligned} \text{Var}(X) &= \text{E}((X - \text{E}(X))^2) \\ &= \text{E}\{X^2 + (\text{E}(X))^2 - 2\text{E}(X)X\} && \text{sviluppando il quadrato} \\ &= \text{E}(X^2) + (\text{E}(X))^2 - 2\text{E}(X)\text{E}(X) && \text{linearità del valore atteso} \\ &= \text{E}(X^2) - (\text{E}(X))^2. \end{aligned}$$

2. Invarianza rispetto a traslazioni

$$\text{Var}(X + b) = \text{Var}(X).$$

La dimostrazione parte dalla definizione della varianza:

$$\begin{aligned} \text{Var}(X + b) &= \text{E}((X + b - \text{E}(X + b))^2) \\ &= \text{E}((X + b - \text{E}(X) - b)^2) && \text{perché } \text{E}(X + b) = \text{E}(X) + b \\ &= \text{E}((X - \text{E}(X))^2) && \text{semplificando} \\ &= \text{Var}(X). \end{aligned}$$

3. Omogeneità di secondo grado

$$\text{Var}(aX) = a^2 \text{Var}(X).$$

La dimostrazione parte ancora una volta dalla definizione della varianza:

$$\begin{aligned}
 \text{Var}(aX) &= E((aX - E(aX))^2) \\
 &= E((aX - aE(X))^2) && \text{perché } E(aX) = aE(X) \\
 &= E((a(X - E(X)))^2) && \text{raccolgendo} \\
 &= E(a^2(X - E(X))^2) && \text{proprietà delle potenze} \\
 &= a^2 E((X - E(X))^2) && \text{linearità del valore atteso} \\
 &= a^2 \text{Var}(X).
 \end{aligned}$$

Dalle due proprietà precedenti si ricava che

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

15.5 V.c. univariate standardizzate

Una v.c. univariata con $E(X) = 0$ è detta **centrata**.

Una v.c. univariata con $E(X) = 0$ e $\text{Var}(X) = 1$ è detta **standardizzata**.

Sia X una v.c. univariata non degenera con valore atteso μ_X e varianza σ_X^2 finite (ovviamente per la proprietà **0** della varianza $\sigma_X^2 > 0$). La trasformazione affine di X in una v.c. con valore atteso 0 e varianza 1 è detta **standardizzazione**. In concreto, è la trasformazione

$$X \longrightarrow Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}} = \frac{X - \mu_X}{\sigma_X}.$$

È facile verificare che Z è standardizzata. Si osservi che Z è espressa adimensionalmente, è un numero puro (aleatorio).

Da Z standardizzata si può riottenere X ancora per trasformazione affine:

$$X = \mu_X + \sigma_X Z.$$

15.6 Le disuguaglianze di Markov e di Čebyshev

In questo paragrafo si approfondisce il tema dell'informazione che gli indici di posizione $E(X)$ e di variabilità $\text{Var}(X)$ portano, quando esistono finiti, sulla distribuzione di X . Si vede che l'informazione è espressa da maggiorazioni per la probabilità che X assuma il suo valore lontano dal centro della distribuzione, sulle code.

Conviene partire dal considerare l'informazione che il valore atteso $E(X)$ porta su una v.c. X non negativa, ossia tale che $P(X \geq 0) = 1$. Per fissare le idee, si pensi a un tempo d'attesa. Se $E(X) = 0$, $X \sim \mathcal{D}(0)$. Se $E(X) > 0$ si ha il seguente risultato.

Teorema 15.1 (*Diseguaglianza di Markov*) Sia X una v.c. non negativa con valore atteso $\mu = E(X) > 0$ finito. Allora per ogni $c > 0$ vale

$$P(X \geq c\mu) \leq \frac{1}{c}. \quad (15.3)$$

Dimostrazione. Senza perdita di generalità, supponiamo che X abbia legge continua. Allora

$$\begin{aligned} \mu &= \int_0^{+\infty} x p_X(x) dx \\ &= \int_0^{c\mu} x p_X(x) dx + \int_{c\mu}^{+\infty} x p_X(x) dx \\ &\geq \int_{c\mu}^{+\infty} x p_X(x) dx && \text{perché } \int_0^{c\mu} x p_X(x) dx \geq 0 \\ &\geq c\mu \int_{c\mu}^{+\infty} p_X(x) dx && \text{perché } x \geq c\mu \text{ in } (c\mu, +\infty) \\ &= c\mu P(X \geq c\mu) \end{aligned}$$

da cui, semplificato $\mu > 0$,

$$1 \geq cP(X \geq c\mu)$$

che dà la tesi essendo $c > 0$. \square

Per $c \in (0, 1]$ la diseguaglianza di Markov asserisce che la probabilità di un certo evento è minore o uguale a $1/c$, ossia a un valore maggiore o uguale a 1. Quindi la diseguaglianza è vera ma non è informativa. Per $c > 1$ è invece informativa. Ad esempio, si ricavano da (15.3) le diseguaglianze

$$P(X \geq 2\mu) \leq 1/2,$$

$$P(X \geq 3\mu) \leq 1/3,$$

eccetera. Quindi per X tale che $P(X \geq 0) = 1$ la (15.3) migliora la probabilità che X cada in una coda destra formata da valori lontani da μ , quale $[c\mu, +\infty)$.

Se X non degenerare ha valore atteso e varianza finite, come corollario della diseguaglianza di Markov si ottiene subito un'altra diseguaglianza. Questa volta si migliora la probabilità di ottenere una realizzazione lontana da $E(X)$ su entrambe le code formate dai valori lontani da μ .

Teorema 15.2 (*Diseguaglianza di Čebyshev*) Sia X una v.c. univariata con valore atteso $\mu = E(X)$ finito e varianza pure finita $\sigma^2 = \text{Var}(X) > 0$. Allora per ogni $k > 0$ vale

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \quad (15.4)$$

Dimostrazione. Sia $T = (X - \mu)^2$. Si ha $P(T \geq 0) = 1$ e vale anche $E(T) = E((X - E(X))^2) = \text{Var}(X) = \sigma^2 > 0$. Dunque

$$\begin{aligned} P(|X - \mu| \geq k\sigma) &= P((X - \mu)^2 \geq k^2\sigma^2) \\ &= P(T \geq k^2 E(T)) \\ &\leq \frac{1}{k^2} \end{aligned}$$

per la disuguaglianza di Markov applicata a T . \square

Per $k \in (0, 1]$ la disuguaglianza di Čebyshev è vera ma non è informativa. Per $k > 1$ è invece informativa. Ad esempio, si ricavano da (15.4) le disuguaglianze

$$P(X \in (-\infty, \mu - 2\sigma] \cup [\mu + 2\sigma, +\infty)) \leq 1/4$$

$$P(X \in (-\infty, \mu - 3\sigma] \cup [\mu + 3\sigma, +\infty)) \leq 1/9,$$

eccetera. La disuguaglianza (15.4) maggiore con $1/k^2$ la probabilità dell'unione di entrambe le code, sinistra (verso $-\infty$) e destra (verso $+\infty$), formate dai valori lontani da μ almeno k unità σ , $(-\infty, \mu - k\sigma]$ e $[\mu + k\sigma, +\infty)$.

Per l'evento contrario di $|X - \mu| \geq k\sigma$, la disuguaglianza di Čebyshev (15.2) implica

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

ossia

$$P(X \in (\mu - k\sigma, \mu + k\sigma)) \geq 1 - 1/k^2,$$

risultato che è informativo ancora una volta per $k > 1$.

Esempio 15.1 Se X è una v.c. con $E(X) = 1000$, $\text{Var}(X) = 100 = \sigma^2$ per cui $\sigma = 10$, si deduce che

$$P(1000 - 2 * 10 < X < 1000 + 2 * 10) = P(980 < X < 1020) \geq 1 - 1/4 = 0.75$$

come pure

$$P(1000 - 3 * 10 < X < 1000 + 3 * 10) = P(970 < X < 1030) \geq 1 - 1/9 = 0.8\overline{8}.$$

15.7 Indici di posizione e variabilità per v.c. multivariate

Una v.c. multivariata X può essere considerata come un vettore aleatorio $d \times 1$, $X = (X_1, \dots, X_d)^\top$. Il principale indice di posizione per X è il punto baricentrico della sua distribuzione, il **valore atteso** di X ,

$$\mu_X = E(X) = (E(X_1), \dots, E(X_d))^\top.$$

Anche μ_X è un vettore $d \times 1$.

Il principale indice di variabilità delle componenti e di dipendenza fra le componenti di X è la **matrice di varianze e covarianze**

$$\Sigma_X = \text{Var}(X) = [\text{Cov}(X_i, X_j)] \quad i, j = 1, \dots, d$$

dove $\text{Cov}(X_i, X_j) = E((X_i - E(X_i))(X_j - E(X_j)))$ e quindi $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$. La matrice Σ_X è $d \times d$, risulta simmetrica poiché $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$. Si dimostra che Σ_X è definita non negativa.

Scritta per esteso, la matrice di varianze e covarianze risulta

$$\Sigma_X = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_d) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_d, X_1) & \text{Cov}(X_d, X_2) & \cdots & \text{Var}(X_d) \end{bmatrix}.$$

Formula per il calcolo della covarianza

Per calcolare una covarianza, risulta spesso utile una formula per il calcolo analoga a quella che si usa per la varianza, ossia, per una v.c. bivariata (X, Y) ,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y). \quad (15.5)$$

Infatti

$$\begin{aligned} \text{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= E\{XY - XE(Y) - YE(X) + E(X)E(Y)\} \\ &= E(XY) - E(X)E(Y) \quad \text{per la linearità di } E(\cdot). \end{aligned}$$

È interessante notare che la (15.5) permette di mostrare subito che se (X, Y) ha componenti indipendenti e i valori attesi indicati esistono allora $\text{Cov}(X, Y) = 0$. Infatti, per l'indipendenza, $E(XY) = E(X)E(Y)$.

15.8 La varianza di una combinazione lineare

Sia (X, Y) una v.c. bivariata e $T = aX + bY$ una combinazione lineare delle componenti di (X, Y) . Allora

$$\text{Var}(T) = \text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y). \quad (15.6)$$

Infatti

$$\begin{aligned} \text{Var}(T) &= E((T - E(T))^2) \\ &= E\{(aX + bY - E(aX + bY))^2\} \\ &= E\{(aX + bY - aE(X) - bE(Y))^2\} \\ &= E\{(a(X - E(X)) + b(Y - E(Y)))^2\} \\ &= a^2E\{(X - E(X))^2\} + b^2E\{(Y - E(Y))^2\} \\ &\quad + 2abE\{(X - E(X))(Y - E(Y))\} \\ &= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y). \end{aligned}$$

Se (X, Y) ha componenti indipendenti la formula (15.6) si semplifica e diventa

$$\text{Var}(T) = \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y).$$

La varianza della combinazione lineare di più di due v.c. indipendenti è pure data da una formula semplice. Se $X = (X_1, \dots, X_d)$ ha componenti indipendenti e $T = \sum_{i=1}^d a_i X_i$, combinazione lineare delle X_i con coefficienti reali a_i , si ha

$$\text{Var}(T) = \text{Var}\left(\sum_{i=1}^d a_i X_i\right) = \sum_{i=1}^d a_i^2 \text{Var}(X_i). \quad (15.7)$$

15.9 Il coefficiente di correlazione lineare

Campo di variazione della covarianza

Sia (X, Y) una v.c. bivariata. La covarianza fra X e Y può assumere sia valori positivi sia valori negativi, ma vale

$$-\sigma_X \sigma_Y \leq \text{Cov}(X, Y) \leq \sigma_X \sigma_Y \quad (15.8)$$

dove $\sigma_X = \sqrt{\text{Var}(X)}$ e $\sigma_Y = \sqrt{\text{Var}(Y)}$.

Per la dimostrazione, si ponga $\sigma_{XY} = \text{Cov}(X, Y)$ e si consideri la combinazione lineare $T = \sigma_X^2 Y - \sigma_{XY} X$. Essa ha varianza

$$\begin{aligned} \text{Var}(T) &= \sigma_X^4 \text{Var}(Y) + \sigma_{XY}^2 \text{Var}(X) - 2\sigma_X^2 \sigma_{XY} \text{Cov}(X, Y) \\ &= \sigma_X^4 \sigma_Y^2 + \sigma_{XY}^2 \sigma_X^2 - 2\sigma_X^2 \sigma_{XY}^2 \\ &= \sigma_X^4 \sigma_Y^2 - \sigma_X^2 \sigma_{XY}^2 \\ &= \sigma_X^2 (\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2) \geq 0 \end{aligned}$$

per cui

$$\sigma_X^2 \sigma_Y^2 \geq \sigma_{XY}^2$$

ossia

$$\sigma_{XY}^2 \leq \sigma_X^2 \sigma_Y^2$$

che dà

$$|\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y$$

equivalente alla (15.8).

Se (X, Y) è una v.c. bivariata con $\text{Cov}(X, Y) \neq 0$ allora le componenti X e Y sono dipendenti. Quindi la covarianza è un indice di dipendenza. Spesso l'indice viene presentato in forma adimensionale come **coefficiente di correlazione**

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \quad (15.9)$$

Per la (15.8), il campo di variazione di $\rho(X, Y)$ è

$$-1 \leq \rho(X, Y) \leq 1.$$

Inoltre, quanto più $|\rho(X, Y)|$ è prossimo a 1, tanto più c'è dipendenza fra X e Y . La dipendenza è negativa, ossia al crescere di una variabile tendenzialmente l'altra decresce, se $\rho(X, Y) < 0$. La dipendenza è positiva se $\rho(X, Y) > 0$, ossia al crescere di una variabile tendenzialmente cresce anche l'altra.

15.10 Esercizi

Esercizio 15.1 Sia $X \sim U(a, b)$. Si calcoli lo scarto interquartilico di X .

Esercizio 15.2 Sia $X \sim \text{Esp}(\lambda)$. Si calcoli lo scarto interquartilico di X .

Esercizio 15.3 Sia X una variabile casuale con varianza 16. Si calcoli la varianza di $T = X - 3$.

Esercizio 15.4 Sia X una variabile casuale standardizzata. Si calcolino valore atteso e varianza di $T = 2 + X$.

Esercizio 15.5 Sia X una variabile casuale standardizzata. Si calcolino valore atteso e varianza di $T = 5 + 7X$.

Esercizio 15.6 Sia X una v.c. con distribuzione simmetrica, $E(X) = 100$ e $\text{Var}(X) = 100$. Si dia una maggiorazione della probabilità che X sia maggiore di 130.

Esercizio 15.7 Sia X una v.c. con $E(X) = 100$ e $\text{Var}(X) = 100$. Si dia una maggiorazione della probabilità che X sia maggiore di 130.

Esercizio 15.8 Sia $X \sim \text{Esp}(1)$. Si calcoli $P(X > 5)$ e si confronti il valore ottenuto con la maggiorazione data dalla disuguaglianza di Markov.

Esercizio 15.9 Sia (X, Y) una variabile casuale bivariata con componenti centrate. Si sa che $\text{Var}(X) = \text{Var}(Y) = 9$ e che $\text{Cov}(X, Y) = -1$. Si calcolino valore atteso e varianza di $S = X + Y$.

Esercizio 15.10 Sia (X, Y) una variabile casuale bivariata con componenti centrate. Si sa che $\text{Var}(X) = \text{Var}(Y) = 4$ e che $\text{Cov}(X, Y) = 1$. Si calcolino valore atteso e varianza di $T = 3X - 2Y$.

Esercizio 15.11 Sia (X_1, X_2) una variabile casuale bivariata con componenti indipendenti e standardizzate. Si calcoli il vettore dei valori attesi e la matrice di varianze e covarianze di $Y = (Y_1, Y_2)$, dove $Y_1 = X_1$, $Y_2 = X_1 + X_2$. Si ottenga il valore di $\rho(Y_1, Y_2)$.

Esercizio 15.12 Sia (X_1, X_2, X_3) una variabile casuale bivariata con componenti indipendenti e standardizzate. Si calcoli il vettore dei valori attesi e la matrice di varianze e covarianze di $Y = (Y_1, Y_2, Y_3)$, dove $Y_1 = X_1$, $Y_2 = X_1 + X_2$, $Y_3 = X_1 + X_2 + X_3$. Si ottenga infine il valore di $\rho(Y_1, Y_3)$.

15.11 Appendice: modelli di regressione lineare con R

Sia (X, Y) una v.c. bivariata con $E(X)$, $E(Y)$, $\text{Var}(X) > 0$, $\text{Var}(Y) > 0$, $\text{Cov}(X, Y)$ finiti. Si desidera predire Y conoscendo X , più accessibile all'osservazione. La predizione \hat{Y} di forma più semplice è $\hat{Y} = a + bX$ per opportuni a e b , che corrisponde a un modello di regressione lineare. L'errore di predizione è $Y - \hat{Y}$. Si desidera che $E\{(Y - \hat{Y})^2\}$ sia minima (criterio dei minimi quadrati).

Poiché

$$E\{(Y - \hat{Y})^2\} = \text{Var}(Y - \hat{Y}) + \{E(Y - \hat{Y})\}^2,$$

il minimo di $E\{(Y - \hat{Y})^2\}$ si ottiene imponendo che $E(Y - \hat{Y})$ sia uguale a zero e che $\text{Var}(Y - \hat{Y})$ sia minima. La prima richiesta dà $a = E(Y) - bE(X)$. Aggiungendo anche la seconda richiesta si ha che $\text{Var}(Y - a - bX)$ è pari a

$$\text{Var}\{Y - E(Y) - b(X - E(X))\} = \text{Var}(Y) + b^2\text{Var}(X) - 2b\text{Cov}(X, Y),$$

quantità che è minima quando $2b\text{Var}(X) - 2\text{Cov}(X, Y) = 0$, ossia per $b = \text{Cov}(X, Y)/\text{Var}(X)$. In definitiva, la previsione di Y dei minimi quadrati è

$$\hat{Y} = E(Y) + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}(X - E(X)).$$

È importante accompagnare all'algoritmo di previsione una misura di bontà della previsione, data dal **rapporto di determinazione** $\rho^2(X, Y)$. Si ha infatti

$$\text{Var}(\hat{Y}) = \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)} = \rho^2(X, Y)\text{Var}(Y),$$

per cui $\rho^2(X, Y)$ è la frazione di varianza di Y spiegata da \hat{Y} , ossia da X . La bontà della previsione è massima quando $\rho^2(X, Y) = 1$. Poiché

$$\text{Var}(Y - \hat{Y}) = (1 - \rho^2(X, Y))\text{Var}(Y),$$

si ha allora $\text{Var}(Y - \hat{Y}) = 0$ ossia $P(Y = \hat{Y}) = 1$.

Spesso il modello di regressione lineare viene presentato nella forma $Y = a + bX + \varepsilon$, dove ε è un errore aleatorio indipendente da X con $E(\varepsilon) = 0$. Il modello di regressione lineare multipla è la generalizzazione del modello di regressione lineare con $X = (X_1, \dots, X_d)^\top$ v.c. multivariata, $Y = a + X^\top \beta + \varepsilon$, dove $\beta = (\beta_1, \dots, \beta_d)^\top$ è un vettore di parametri. Applicando il criterio dei minimi quadrati si ottiene che il valore ottimo dei parametri è $\beta = \Sigma_X^{-1}[\text{Cov}(Y, X_i)]$ e $a = E(Y) - E(X)^\top \beta$.

Con R è immediato ottenere modelli predittivi secondo il criterio dei minimi quadrati. Per iniziare si generano i dati dal modello di regressione e si ottengono le stime dei parametri in base ai dati.


```

> set.seed(12345)
> x=rexp(10^2)
> epsilon=rexp(10^2)-1
> y=1+x+epsilon
> b=cov(x,y)/var(x)
> a=mean(y)-b*mean(x)
> print(c(a,b)) # stime dei parametri (vero valore di entrambi 1)
[1] 0.9180817 1.0341412
> print(cor(x,y)^2) # indice  $\rho^2(X,Y)$  stimato
[1] 0.6539704
> plot(x,y)
> lines(x,mean(y)+(cov(x,y)/var(x))*(x-mean(x)))

```

I risultati si possono ottenere più affidabilmente con il comando `lm`. Questo permette di stimare anche i parametri di un modello di regressione lineare multipla. Si ha

```

> model1=lm(formula = y~x)
> model1

```

```

Call:
lm(formula = y ~ x)

```

```

Coefficients:
(Intercept)          x
      0.9181         1.0341

```

```

> summary(model1)

```

```

Call:
lm(formula = y ~ x)

```

```

Residuals:
      Min       1Q   Median       3Q      Max
-1.0546 -0.6844 -0.2886  0.5549  2.7316

```

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.91808    0.11405    8.05 1.98e-12 ***
x            1.03414    0.07599   13.61 < 2e-16 ***
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 0.8506 on 98 degrees of freedom
Multiple R-squared:  0.654, Adjusted R-squared:  0.6504
F-statistic: 185.2 on 1 and 98 DF, p-value: < 2.2e-16

```

152 LEZIONE 15. INDICI DI VARIABILITÀ E PROPRIETÀ DI $\text{VAR}(X)$

```
> x2=x^2
> model2=lm(formula= y~ x+x2)
> model2

Call:
lm(formula = y ~ x + x2)

Coefficients:
(Intercept)          x          x2
    0.88586      1.09703     -0.01361

> summary(model2)

Call:
lm(formula = y ~ x + x2)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0374 -0.6825 -0.2695  0.5895  2.7071

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.88586    0.14657   6.044 2.79e-08 ***
x            1.09703    0.19408   5.653 1.59e-07 ***
x2          -0.01361    0.03861  -0.352  0.725
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8545 on 97 degrees of freedom
Multiple R-squared:  0.6544, Adjusted R-squared:  0.6473
F-statistic: 91.84 on 2 and 97 DF,  p-value: < 2.2e-16
```

Risulta che il secondo modello, che cerca di migliorare le previsioni sfruttando anche X^2 oltre che X , in sostanza non produce miglioramenti. L'indice di bontà R-squared (frazione di varianza spiegata) passa da 0.654 a 0.6544 solamente.

Lezione 16

La funzione generatrice dei momenti

16.1 Definizioni

Si dicono **momenti** di una v.c. univariata X i valori

$$\mu_r = E(X^r), \quad r = 1, 2, \dots \quad (16.1)$$

In particolare, μ_r è detto momento r -esimo di X . Si ha subito

$$\begin{aligned} \mu_1 &= \mu = E(X) \\ \mu_2 - \mu_1^2 &= E(X^2) - (E(X))^2 = \text{Var}(X). \end{aligned}$$

Per la disuguaglianza di Čebyshev, i valori di μ_1 e μ_2 vincolano la legge di probabilità di X , in particolare le probabilità sulle code. Si può intuire che, sotto opportune condizioni, l'intera successione dei momenti, μ_r , $r \in \mathbb{N}^+$, determini univocamente la legge di X . Per tenere sotto controllo la successione dei momenti, si introduce una funzione che li esprime tutti, permettendone il calcolo tramite facili passaggi.

Definizione 16.1 Sia X una v.c. univariata con f.m.p. o f.d.p. $p_X(x)$. La **funzione generatrice dei momenti** di X , indicata con $M_X(t)$, è una funzione reale di variabile reale definita da

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_{x \in S_X} e^{tx} p_X(x) & \text{se } X \text{ ha legge discreta} \\ \int_{-\infty}^{+\infty} e^{tx} p_X(x) dx & \text{se } X \text{ ha legge continua.} \end{cases}$$

Una funzione generatrice dei momenti (f.g.m.) ha le seguenti proprietà fondamentali:

- nell'origine vale sempre 1: $M_X(0) = 1$ per ogni X ;
- il dominio di finitezza di $M_X(t)$, ovvero $D_X = \{t \in \mathbb{R} : M_X(t) < +\infty\}$, è convesso (ossia è un intervallo, o una semiretta, o l'intera retta reale);
- $M_X(t) > 0$ per ogni $t \in D_X$.

La f.g.m. ha ulteriori proprietà molto utili se il suo dominio di finitezza è sufficientemente 'grande', precisamente se include un intervallo aperto che contiene l'origine. Questa condizione merita una etichetta particolare.

Definizione 16.2 *Sia X una v.c. univariata. Si dice che la v.c. univariata X ha funzione generatrice dei momenti propria se il dominio di finitezza di $M_X(t)$ include l'origine come punto interno, ossia se esiste $\varepsilon > 0$ tale che $(-\varepsilon, \varepsilon) \subseteq D_X$, e pertanto per ogni $t \in (-\varepsilon, \varepsilon)$ vale $M_X(t) < +\infty$.*

Esempio 16.1 Se X ha supporto limitato, X ha f.g.m. propria. Infatti, se $S_X \subseteq [a, b]$ per $a, b \in \mathbb{R}$ con $a < b$ (supporto limitato) si ha anche

$$\begin{aligned} \text{per } t > 0 \quad & S_{e^{tX}} \subseteq [e^{ta}, e^{tb}] \\ \text{per } t < 0 \quad & S_{e^{tX}} \subseteq [e^{tb}, e^{ta}] \end{aligned}$$

per cui in entrambi i casi $M_X(t)$, cioè $E(e^{tX})$, è finita per la proprietà di Cauchy del valore atteso. Il dominio di finitezza D_X è quindi \mathbb{R} e la f.g.m. è propria. \diamond

Esempio 16.2 Sia $X \sim Bi(n, p)$, $n \in \mathbb{N}^+$, $p \in (0, 1)$. Si ha $S_X = \{0, 1, \dots, n\} \subseteq [0, n]$, quindi X ha f.g.m. propria, con dominio di finitezza \mathbb{R} . È facile calcolare $M_X(t)$:

$$\begin{aligned} M_X(t) &= \sum_{x \in S_X} e^{tx} p_X(x) \\ &= \sum_{x=0}^n (e^t)^x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= (1-p + pe^t)^n \end{aligned}$$

per il teorema del binomio, cfr. teorema 4.1. \diamond

16.2 Ottenimento dei momenti

Una f.g.m. propria ha nell'origine derivate di ogni ordine, i cui valori 'generano' i momenti di X .

Teorema 16.1 *Se X ha f.g.m. propria, tutti i momenti di X sono finiti e vale*

$$\mu_r = E(X^r) = \left. \frac{d^r}{dt^r} M_X(t) \right|_{t=0} = M_X^{(r)}(0), \quad (16.2)$$

per $r = 1, 2, \dots$.

Per giustificare il risultato (16.2), si supponga che X abbia legge discreta con supporto costituito da un numero finito di punti. Si ha allora

$$M_X(t) = \sum_{x \in S_X} e^{tx} p_X(x)$$

dove la somma ha un numero finito di addendi. Quindi, essendo

$$\frac{d}{dt} e^{tx} = x e^{tx},$$

si ha

$$\begin{aligned} M'_X(t) &= \sum_{x \in S_X} x e^{tx} p_X(x) = E(X e^{tX}) \\ M''_X(t) &= \sum_{x \in S_X} x^2 e^{tx} p_X(x) = E(X^2 e^{tX}) \end{aligned}$$

e in generale

$$\frac{d^r}{dt^r} M_X(t) = E(X^r e^{tX}).$$

Basta valutare in $t = 0$ l'espressione ottenuta per stabilire la (16.2).

Il vantaggio di calcolare i momenti, in particolare valore atteso e varianza, passando attraverso la f.g.m. è presto detto. Se si dispone di $M_X(t)$ in forma chiusa, come composizione di funzioni elementari, calcolare le derivate di $M_X(t)$ è sempre possibile. Quindi è sufficiente un solo 'trucco' per sommare o integrare l'espressione che definisce $M_X(t)$. Non serve un accorgimento particolare per ogni r per sommare o integrare l'espressione che definisce $E(X^r)$.

Esempio 16.3 La f.g.m. di $X \sim Bi(n, p)$, $n \in \mathbb{N}^+$, $p \in (0, 1)$, propria, è

$$M_X(t) = (1 - p + p e^t)^n.$$

Si desidera calcolare $E(X)$ e $\text{Var}(X)$. Derivando $M_X(t)$ rispetto a t si ha

$$\begin{aligned} M'_X(t) &= n(1 - p + p e^t)^{n-1} p e^t \\ M''_X(t) &= n(n-1)(1 - p + p e^t)^{n-2} p^2 e^{2t} + M'_X(t) \end{aligned}$$

quindi, poiché $e^0 = 1$,

$$\begin{aligned} E(X) &= M'_X(0) = n(1 - p + p e^0)^{n-1} p e^0 = np \\ E(X^2) &= M''_X(0) = n(n-1)(1 - p + p e^0)^{n-2} p^2 e^0 + M'_X(0) \\ &= n(n-1)p^2 + np \end{aligned}$$

da cui

$$\text{Var}(X) = E(X^2) - (E(X))^2 = n(n-1)p^2 + np - n^2p^2 = -np^2 + np = np(1-p).$$

◇

16.3 Caratterizzazione della legge di probabilità

Quando la f.g.m. di X è propria, la successione di tutti i momenti di X determina univocamente P_X , la legge di probabilità di X . Vale infatti il seguente risultato, che si enuncia senza dimostrazione.

Teorema 16.2 *Siano X e Y due v.c. univariate con f.g.m. propria, $M_X(t)$ e $M_Y(t)$, rispettivamente. Se per qualche $\varepsilon > 0$ si ha $M_X(t) = M_Y(t)$ per ogni $t \in (-\varepsilon, \varepsilon)$, allora X e Y hanno la stessa legge di probabilità, $X \sim Y$.*

Quindi $X \sim Y \iff M_X(t) = M_Y(t)$ (se $M_X(t)$ e $M_Y(t)$ sono proprie).

Esempio 16.4 Sia X una v.c. con f.g.m.

$$M_X(t) = \frac{e^{-t} + e^t}{2}, \quad t \in \mathbb{R},$$

quindi propria. In linea di principio $M_X(t)$ determina la legge di X . In questo caso è facile ricostruire P_X . Dal confronto

$$M_X(t) = e^{(-1)t} \frac{1}{2} + e^{(1)t} \frac{1}{2} = \sum_{x \in S_X} e^{tx} p_X(x)$$

si ricava $S_X = \{-1, 1\}$ e $p_X(-1) = 1/2$, $p_X(1) = 1/2$.

◇

Esempio 16.5 In modo simile

$$\begin{aligned} M_X(t) &= \frac{1 + e^{-2t} + e^{3t}}{3} \\ &= (1) \frac{1}{3} + e^{-2t} \frac{1}{3} + e^{3t} \frac{1}{3} \\ &= e^{(0)t} \frac{1}{3} + e^{(-2)t} \frac{1}{3} + e^{3t} \frac{1}{3} \end{aligned}$$

dà

$$S_X = \{-2, 0, 3\}$$

e

$$p_X(x) = \frac{1}{3}, \quad x \in S_X,$$

in breve $X \sim Ud(-2, 0, 3)$.

◇

Con l'eccezione di esempi molto semplici, come quelli appena esaminati, determinare la legge di X conoscendo $M_X(t)$ è possibile solo se si dispone già di $M_X(t)$ nella collezione di f.g.m. calcolate a partire da S_X e $p_X(x)$. (Il problema inverso è sempre più difficile del problema diretto). Risulta quindi importante calcolare le f.g.m. delle principali leggi di probabilità.

16.4 La f.g.m. di leggi notevoli

Molte fra le principali leggi di probabilità univariate hanno f.g.m. propria.

- **F.g.m. di una degenere**

Sia $X \sim \mathcal{D}(x_0)$, $x_0 \in \mathbb{R}$, per cui $S_X = \{x_0\}$ e $p_X(x_0) = 1$. Allora

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \sum_{x \in S_X} e^{tx} p_X(x) = e^{tx_0} (1) = e^{tx_0} \\ M'_X(t) &= x_0 e^{tx_0} \implies E(X) = M'_X(0) = x_0 \\ M''_X(t) &= x_0^2 e^{tx_0} \implies E(X^2) = M''_X(0) = x_0^2 \end{aligned}$$

da cui $\text{Var}(X) = E(X^2) - (E(X))^2 = x_0^2 - (x_0)^2 = 0$.

- **F.g.m. di una Poisson**

Sia $X \sim P(\lambda)$, $\lambda > 0$, per cui $S_X = \mathbb{N}$ e, per $x \in S_X$, $p_X(x) = e^{-\lambda} \lambda^x / x!$. Allora

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \sum_{x \in S_X} e^{tx} p_X(x) \\ &= \sum_{x=0}^{+\infty} (e^t)^x e^{-\lambda} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{+\infty} \frac{(\lambda e^t)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} \quad \text{per la (8.3)} \\ &= e^{\lambda(e^t - 1)} \end{aligned}$$

e pertanto

$$\begin{aligned} M'_X(t) &= e^{\lambda(e^t - 1)} \lambda e^t \implies E(X) = M'_X(0) = \lambda \\ M''_X(t) &= e^{\lambda(e^t - 1)} \lambda^2 e^{2t} + e^{\lambda(e^t - 1)} \lambda e^t \implies E(X^2) = M''_X(0) = \lambda^2 + \lambda \end{aligned}$$

da cui $\text{Var}(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda$.

Osservazione: se $X \sim P(\lambda)$ allora $E(X) = \text{Var}(X)$.

- **F.g.m. di una geometrica**

Sia $X \sim Ge(p)$, $p \in (0, 1)$, per cui $S_X = \mathbb{N}^+$ e, per $x \in S_X$, $p_X(x) =$

$p(1-p)^{x-1}$. Allora

$$\begin{aligned}
 M_X(t) &= E(e^{tX}) = \sum_{x \in S_X} e^{tx} p_X(x) \\
 &= \sum_{x=1}^{+\infty} e^{t(x-1+1)} p(1-p)^{x-1} \\
 &= pe^t \sum_{x=1}^{+\infty} e^{t(x-1)} (1-p)^{x-1} \\
 &= pe^t \sum_{i=0}^{+\infty} \{e^t(1-p)\}^i \quad \text{posto } x-1=i \\
 &= \frac{pe^t}{1-(1-p)e^t}
 \end{aligned}$$

purché la ragione della serie geometrica sia minore di 1, quindi $(1-p)e^t < 1$ ossia $t < -\log(1-p)$. Poiché $-\log(1-p) > 0$, la f.g.m. di $Ge(p)$ è propria. Si ha

$$\begin{aligned}
 M'_X(t) &= p \frac{e^t(1-(1-p)e^t) - e^t(-(1-p)e^t)}{(1-(1-p)e^t)^2} \\
 &= \frac{pe^t}{(1-(1-p)e^t)^2}
 \end{aligned}$$

da cui

$$E(X) = M'_X(0) = \frac{p}{(1-(1-p))^2} = \frac{1}{p}.$$

Con calcoli che si lasciano come esercizio si trova che

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

• **F.g.m. di una esponenziale**

Sia $X \sim Esp(\lambda)$, $\lambda > 0$, per cui $S_X = [0, +\infty)$ e, per $x \in S_X$, $p_X(x) = \lambda e^{-\lambda x}$. Allora

$$\begin{aligned}
 M_X(t) &= E(e^{tX}) = \int_0^{+\infty} e^{tx} \lambda e^{-\lambda x} dx \\
 &= \lambda \int_0^{+\infty} e^{-(\lambda-t)x} dx \\
 &= \frac{\lambda}{\lambda-t} \int_0^{+\infty} (\lambda-t) e^{-(\lambda-t)x} dx \\
 &= \frac{\lambda}{\lambda-t}
 \end{aligned}$$

purché $\lambda - t > 0$ ossia $t < \lambda$. Poiché $\lambda > 0$, $X \sim Esp(\lambda)$ ha f.g.m. propria. Pertanto

$$X \sim Esp(\lambda) \iff M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}, \quad t < \lambda.$$

Poiché

$$\begin{aligned} M'_X(t) &= -\left(1 - \frac{t}{\lambda}\right)^{-2} \left(-\frac{1}{\lambda}\right) = \frac{1}{\lambda} \left(1 - \frac{t}{\lambda}\right)^{-2} \\ M''_X(t) &= \frac{-2}{\lambda} \left(1 - \frac{t}{\lambda}\right)^{-3} \left(-\frac{1}{\lambda}\right) = \frac{2}{\lambda^2} \left(1 - \frac{t}{\lambda}\right)^{-3} \end{aligned}$$

si ottiene

$$\begin{aligned} E(X) &= M'_X(0) = \frac{1}{\lambda} \\ E(X^2) &= M''_X(0) = \frac{2}{\lambda^2} \end{aligned}$$

da cui

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

Osservazione: se $X \sim Esp(\lambda)$ allora $\text{Var}(X) = (E(X))^2$.

• **F.g.m. di una gamma**

Sia $X \sim Ga(\alpha, \lambda)$, $\alpha, \lambda > 0$, per cui $S_X = [0, +\infty)$ e, per $x \in S_X$,

$$p_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}.$$

Allora

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \int_0^{+\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \lambda^\alpha \int_0^{+\infty} \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-(\lambda-t)x} dx \\ &= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^{+\infty} \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} e^{-(\lambda-t)x} dx \\ &= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \end{aligned}$$

purché $\lambda - t > 0$ ossia $t < \lambda$. Poiché $\lambda > 0$, $X \sim Ga(\alpha, \lambda)$ ha f.g.m. propria. Pertanto

$$X \sim Ga(\alpha, \lambda) \iff M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-\alpha}, \quad t < \lambda.$$

Poiché

$$\begin{aligned} M'_X(t) &= -\alpha \left(1 - \frac{t}{\lambda}\right)^{-\alpha-1} \left(-\frac{1}{\lambda}\right) = \frac{\alpha}{\lambda} \left(1 - \frac{t}{\lambda}\right)^{-\alpha-1} \\ M''_X(t) &= -\frac{\alpha(\alpha+1)}{\lambda} \left(1 - \frac{t}{\lambda}\right)^{-\alpha-2} \left(-\frac{1}{\lambda}\right) = \frac{\alpha(\alpha+1)}{\lambda^2} \left(1 - \frac{t}{\lambda}\right)^{-\alpha-2} \end{aligned}$$

si ottiene

$$\begin{aligned} E(X) &= M'_X(0) = \frac{\alpha}{\lambda} \\ E(X^2) &= M''_X(0) = \frac{\alpha(\alpha+1)}{\lambda^2} \end{aligned}$$

da cui

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

16.5 La legge della somma di v.c. indipendenti

Il prodotto di f.g.m. proprie resta una f.g.m. propria. Precisamente, è la f.g.m. della v.c. somma quando gli addendi aleatori X_i sono v.c. indipendenti con funzione generatrice dei momenti propria.

Teorema 16.3 *Sia $X = (X_1, \dots, X_d)$ una v.c. multivariata con componenti X_i indipendenti che hanno funzione generatrice dei momenti propria, $M_{X_i}(t)$. Allora $S = \sum_{i=1}^d X_i$ ha f.g.m. propria data da $M_S(t) = \prod_{i=1}^d M_{X_i}(t)$.*

Dimostrazione. Si deve calcolare la f.g.m. di S :

$$\begin{aligned} M_S(t) &= E(e^{tS}) && \text{definizione di f.g.m.} \\ &= E\left(e^{t \sum_{i=1}^d X_i}\right) && \text{definizione di } S \\ &= E\left(\prod_{i=1}^d e^{tX_i}\right) && \text{proprietà di exp} \\ &= \prod_{i=1}^d E(e^{tX_i}) && \text{per l'indipendenza delle } X_i \\ &= \prod_{i=1}^d M_{X_i}(t) && \text{definizione di f.g.m.} \end{aligned}$$

□

Il risultato permette di calcolare $M_S(t)$ molto facilmente. Tuttavia di solito ciò che interessa è la legge di probabilità di S . Per individuarla, si dovrà controllare se $M_S(t)$ appare nella lista delle f.g.m. già disponibili.

16.6 Proprietà additive

Talvolta le v.c. indipendenti che si addizionano hanno tutte legge di un certo tipo e anche la loro somma ha legge di quello stesso tipo. Si dice allora che vale la proprietà additiva per quella legge. Si descrivono in dettaglio tre risultati notevoli.

16.6.1 Binomiali indipendenti con lo stesso p

Sia $X = (X_1, \dots, X_d)$ una v.c. multivariata con componenti X_i indipendenti e legge marginale $X_i \sim Bi(n_i, p)$, dove $n_i \in \mathbb{N}^+$ e $p \in (0, 1)$, per cui

$$M_{X_i}(t) = (1 - p + pe^t)^{n_i}.$$

Allora $S = \sum_{i=1}^d X_i$ ha f.g.m. propria

$$\begin{aligned} M_S(t) &= \prod_{i=1}^d M_{X_i}(t) \\ &= \prod_{i=1}^d (1 - p + pe^t)^{n_i} \\ &= (1 - p + pe^t)^{\sum_{i=1}^d n_i} \end{aligned}$$

per cui, per il teorema di caratterizzazione,

$$S \sim Bi\left(\sum_{i=1}^d n_i, p\right).$$

16.6.2 Poisson indipendenti

Sia $X = (X_1, \dots, X_d)$ una v.c. multivariata con componenti X_i indipendenti e legge marginale $X_i \sim P(\lambda_i)$, dove $\lambda_i > 0$, $i = 1, \dots, d$, per cui

$$M_{X_i}(t) = e^{\lambda_i(e^t - 1)}.$$

Allora $S = \sum_{i=1}^d X_i$ ha f.g.m. propria

$$\begin{aligned} M_S(t) &= \prod_{i=1}^d M_{X_i}(t) \\ &= \prod_{i=1}^d e^{\lambda_i(e^t - 1)} \\ &= \exp\left\{\left(\sum_{i=1}^d \lambda_i\right)(e^t - 1)\right\} \end{aligned}$$

e in conclusione, per il teorema di caratterizzazione,

$$S \sim P\left(\sum_{i=1}^d \lambda_i\right).$$

16.6.3 Gamma indipendenti con lo stesso λ

Sia $X = (X_1, \dots, X_d)$ una v.c. multivariata con componenti X_i indipendenti e legge marginale $X_i \sim Ga(\alpha_i, \lambda)$, dove $\alpha_i > 0$, $i = 1, \dots, n$, e $\lambda > 0$, per cui

$$M_{X_i}(t) = \left(1 - \frac{t}{\lambda}\right)^{-\alpha_i}.$$

Allora $S = \sum_{i=1}^d X_i$ ha f.g.m. propria

$$\begin{aligned} M_S(t) &= \prod_{i=1}^d M_{X_i}(t) \\ &= \prod_{i=1}^d \left(1 - \frac{t}{\lambda}\right)^{-\alpha_i} \\ &= \left(1 - \frac{t}{\lambda}\right)^{-\sum_{i=1}^d \alpha_i} \end{aligned}$$

e in conclusione, per il teorema di caratterizzazione,

$$S \sim Ga\left(\sum_{i=1}^d \alpha_i, \lambda\right).$$

16.7 Esercizi risolti

Esempio 16.6 Sia Y una variabile casuale univariata avente quale funzione generatrice dei momenti $M_Y(t) = \exp(t^2/2)$, dove $\exp(z) = e^z$. Siano poi Y_1, Y_2 copie indipendenti di Y e si ponga $W = -Y_1 - Y_2$. Si calcoli la funzione generatrice dei momenti di W . Si ottengano valore atteso e varianza di W .

La f.g.m. di W è

$$\begin{aligned} M_W(t) &= E(e^{tW}) && \text{definizione di f.g.m.} \\ &= E(e^{t(-Y_1 - Y_2)}) && \text{definizione di } W \\ &= E(e^{-tY_1} e^{-tY_2}) && \text{proprietà di } \exp \\ &= E(e^{-tY_1}) E(e^{-tY_2}) && Y_1 \text{ e } Y_2 \text{ indipendenti} \\ &= M_{Y_1}(-t) M_{Y_2}(-t) && \text{definizione di f.g.m.} \\ &= e^{(-t)^2/2} e^{(-t)^2/2} && \text{f.g.m. di } Y \sim Y_1 \sim Y_2 \\ &= e^{t^2} && \text{proprietà di } \exp. \end{aligned}$$

Poiché le derivate di $M_W(t)$ sono

$$M'_W(t) = 2te^{t^2} \quad \text{e} \quad M''_W(t) = 2e^{t^2} + (2t)^2 e^{t^2}$$

si ha

$$E(W) = M'_W(0) = 0 \quad \text{e} \quad E(W^2) = M''_W(0) = 2$$

per cui

$$\text{Var}(W) = E(W^2) - (E(W))^2 = 2 - 0^2 = 2.$$

◇

Esempio 16.7 Sia Y una variabile casuale univariata avente quale funzione generatrice dei momenti $M_Y(t) = (1 + \exp(t))/2$, dove $\exp(z) = e^z$. Siano poi Y_1, Y_2, Y_3 copie indipendenti di Y e si ponga $W = Y_1 + Y_2 + Y_3$. Si calcoli la funzione generatrice dei momenti di W . Si ottengano valore atteso e varianza di W .

La f.g.m. di W è

$$\begin{aligned} M_W(t) &= E(e^{tW}) && \text{definizione di f.g.m.} \\ &= E(e^{t(Y_1+Y_2+Y_3)}) && \text{definizione di } W \\ &= E(e^{tY_1} e^{tY_2} e^{tY_3}) && \text{proprietà di } \exp \\ &= E(e^{tY_1}) E(e^{tY_2}) E(e^{tY_3}) && Y_1, Y_2, Y_3 \text{ indipendenti} \\ &= M_{Y_1}(t) M_{Y_2}(t) M_{Y_3}(t) && \text{definizione di f.g.m.} \\ &= M_Y(t)^3 && Y \sim Y_1 \sim Y_2 \sim Y_3 \\ &= \frac{(1 + e^t)^3}{8}. \end{aligned}$$

Poiché le derivate di $M_W(t)$ sono

$$M'_W(t) = \frac{3}{8}(1 + e^t)^2 e^t \quad \text{e} \quad M''_W(t) = \frac{3}{8}2(1 + e^t)e^{2t} + M'_W(t)$$

si ha

$$E(W) = M'_W(0) = \frac{3}{2} \quad \text{e} \quad E(W^2) = M''_W(0) = \frac{3}{2} + \frac{3}{2} = 3$$

per cui

$$\text{Var}(W) = E(W^2) - (E(W))^2 = 3 - \frac{9}{4} = \frac{3}{4}.$$

◇

16.8 Esercizi

Esercizio 16.1 Una v.c. X ha, per $t \in \mathbb{R}$, funzione generatrice dei momenti $M_X(t) = (1 + 3e^{2t})/4$. Si dica se X ha funzione generatrice dei momenti propria.

Esercizio 16.2 Si calcoli la funzione generatrice dei momenti di $X \sim Ud(1, 2, 3)$ e si dica se X ha f.g.m. propria.

Esercizio 16.3 Una v.c. X ha, per $t \in \mathbb{R}$, funzione generatrice dei momenti $M_X(t) = (1 + e^{2t})/2$. Si dica qual è la legge di probabilità di X .

Esercizio 16.4 Una v.c. X ha, per $t \in \mathbb{R}$, funzione generatrice dei momenti $M_X(t) = e^{e^t - 1}$. Si dica qual è la legge di probabilità di X .

Esercizio 16.5 Una v.c. X ha, per $t \in \mathbb{R}$, funzione generatrice dei momenti $M_X(t) = (1 + 3e^{2t})/4$. Si calcolino $E(X)$ e $\text{Var}(X)$.

Esercizio 16.6 Una v.c. X ha funzione generatrice dei momenti $M_X(t) = e^t/(2 - e^t)$ per $t < \log 2$. Si calcolino $E(X)$ e $\text{Var}(X)$.

Esercizio 16.7 Siano $Y_1 \sim P(\lambda_1)$, $Y_2 \sim P(\lambda_2)$ e $Y_3 \sim P(\lambda_3)$ v.c. univariate indipendenti, con $\lambda_i > 0$, $i = 1, 2, 3$. Si reperisca la legge di probabilità di $S = Y_1 + Y_2 + Y_3$.

Esercizio 16.8 (*Compito del 01/02/2017*) Sia Y una variabile casuale univariata avente quale funzione generatrice dei momenti $M_Y(t) = \exp(0.5t^2 + 2t)$, dove $\exp(z) = e^z$. Siano poi Y_1, Y_2 copie indipendenti di Y e si ponga $W = 2Y_1 - Y_2$. Si calcoli la funzione generatrice dei momenti di W . Si ottengano valore atteso e varianza di W .

Esercizio 16.9 (*Compito del 10/02/2017*) Sia Y una variabile casuale univariata avente quale funzione generatrice dei momenti $M_Y(t) = \exp(0.5t^2 + t)$, dove $\exp(z) = e^z$. Siano poi Y_1, Y_2 copie indipendenti di Y e si ponga $W = Y_1 - Y_2$. Si calcoli la funzione generatrice dei momenti di W . Si ottengano valore atteso e varianza di W .

Esercizio 16.10 (*Compito del 16/06/2017*) Sia Y una variabile casuale univariata avente quale funzione generatrice dei momenti $M_Y(t) = \exp(0.5t^2 - 3t)$, dove $\exp(z) = e^z$. Siano poi Y_1, Y_2 copie indipendenti di Y e si ponga $W = Y_1 - Y_2$. Si calcoli la funzione generatrice dei momenti di W . Si ottengano valore atteso e varianza di W .

Esercizio 16.11 (*Compito del 17/07/2017*) Sia Y una variabile casuale univariata avente quale funzione generatrice dei momenti $M_Y(t) = \exp(0.5t^2 + 2t)$, dove $\exp(z) = e^z$. Siano poi Y_1, Y_2 copie indipendenti di Y e si ponga $W = Y_1 - Y_2$. Si calcoli la funzione generatrice dei momenti di W . Si ottengano valore atteso e varianza di W .

Esercizio 16.12 (*Compito del 11/09/2017*) Sia Y una variabile casuale univariata avente quale funzione generatrice dei momenti $M_Y(t) = \exp(0.5t^2)$, dove $\exp(z) = e^z$. Siano poi Y_1, Y_2 copie indipendenti di Y e si ponga $W = Y_1 + 2Y_2$. Si calcoli la funzione generatrice dei momenti di W . Si ottengano valore atteso e varianza di W .

Lezione 17

Le leggi normali

17.1 Genesi e definizione

Si considerino misure ripetute, effettuate con lo stesso strumento di misura, di una quantità μ . Lo strumento di misura è affetto da errore, per cui le misure x_i saranno realizzazioni di una v.c. X , $X \rightarrow x_i$. Siano z_i gli errori di misura, espressi in una scala standard, realizzazioni di una v.c. Z , quindi $Z \rightarrow z_i$. La relazione fra misure ed errori di misura espressi su una scala standard è

$$x_i = \mu + \sigma z_i, \quad i = 1, \dots, n.$$

Analoga relazione vale per le v.c. X e Z ,

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

Per modellare la legge di probabilità di Z , Gauss considerò che siano possibili errori positivi e negativi di qualunque entità, quindi che $S_Z = \mathbb{R}$. Poiché Z ha supporto continuo, si può assumere che abbia legge continua, con f.d.p.

$$p_Z(z) > 0 \quad \text{per ogni } z \in \mathbb{R}.$$

La simmetria delle misure attorno a μ , empiricamente riscontrata, comporta che l'errore di misura Z abbia densità simmetrica attorno all'origine, quindi

$$p_Z(-z) = p_Z(z) \quad \text{per ogni } z \in \mathbb{R},$$

per cui la mediana di Z è 0. Si riscontra anche empiricamente che 'grandi' errori sono rari mentre 'piccoli' errori sono frequenti. Ciò comporta che

$$p_Z(|z|) \quad \text{sia monotona decrescente in } |z|,$$

per cui la moda degli errori di misura risulta 0.

Per Gauss, l'andamento qualitativo della $p_Z(z)$ appena delineato è modellato nel modo matematicamente più semplice dall'equazione differenziale

$$\frac{d}{dz}p_Z(z) = -z p_Z(z) \quad (17.1)$$

che cattura l'andamento di simmetria attorno all'origine e la moda in 0 di $p_Z(z)$. Infatti la (17.1) implica che

- $p'_Z(z) > 0$ per $z < 0$, ossia $p_Z(z)$ è crescente in $(-\infty, 0)$;
- $p'_Z(z) < 0$ per $z > 0$, ossia $p_Z(z)$ è decrescente in $(0, \infty)$;
- quindi $p_Z(z)$ ha un massimo assoluto in $z = 0$, fatto confermato da $p'_Z(z) = 0$ per $z = 0$.

Derivando la relazione $p'_Z(z) = -z p_Z(z)$ si ha

$$\begin{aligned} p''_Z(z) &= -p_Z(z) - z(-z p_Z(z)) \\ &= (z^2 - 1)p_Z(z). \end{aligned}$$

Quindi $p_Z(z)$ ha due punti di flesso, in $z = \pm 1$, posti simmetricamente rispetto alla moda in 0. La scala standard per l'errore di misura è precisamente quella per cui i flessi di $p_Z(z)$ sono a ± 1 . La densità di Z risulta concava in $(-1, 1)$, e convessa esternamente all'intervallo $(-1, 1)$. Qualitativamente, il grafico di $p_Z(z)$ si conforma quindi ad un andamento a campana, la famosa campana di Gauss.

Per ricavare la forma analitica di $p_Z(z)$, basta risolvere l'equazione differenziale

$$\frac{p'_Z(z)}{p_Z(z)} = -z$$

che è equivalente a

$$\frac{d}{dz} \log p_Z(z) = -z.$$

Integrando si ottiene

$$\log p_Z(z) = -\frac{1}{2}z^2 + c$$

che dà

$$p_Z(z) = k e^{-\frac{1}{2}z^2}.$$

Si dimostra che la costante di normalizzazione $k > 0$ risulta

$$k = \frac{1}{\sqrt{2\pi}}.$$

Definizione 17.1 Una v.c. univariata Z con supporto $S_Z = \mathbb{R}$ e f.d.p.

$$p_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad (17.2)$$

è detta con legge **normale standard**, in breve $Z \sim N(0, 1)$. Inoltre, la legge di $X = \mu + \sigma Z$, dove $Z \sim N(0, 1)$, è detta **normale con parametri μ e σ^2** , in breve $X \sim N(\mu, \sigma^2)$.

La f.d.p. di $X = \mu + \sigma Z$ è (cfr. formula (13.3) ed Esempio 13.5)

$$p_X(x) = p_Z(g^{-1}(x)) \left| \frac{dg^{-1}(x)}{dx} \right|$$

dove $x = g(z) = \mu + \sigma z$, $z = g^{-1}(x) = \frac{x-\mu}{\sigma}$, $\frac{dg^{-1}(x)}{dx} = \frac{1}{\sigma}$. Quindi X con legge normale generale ha f.d.p.

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\} \quad (17.3)$$

e il supporto è ovviamente $S_X = \mathbb{R}$.

L'andamento qualitativo di (17.3) è ancora di tipo campanulare, con moda in $x = \mu$ e flessi ai punti $\mu \pm \sigma$.

17.2 Chiusura sotto trasformazioni affini

Sia $X \sim N(\mu, \sigma^2)$. Poiché $X = \mu + \sigma Z$, dove $Z \sim N(0, 1)$, il parametro $\mu \in \mathbb{R}$ è detto **parametro di posizione** e il parametro $\sigma > 0$ è detto **parametro di scala**. Per come è costruita, una X con legge normale è una trasformazione affine di $Z \sim N(0, 1)$. Non sorprende che se si applica a X normale una trasformazione affine la v.c. trasformata rimanga con legge normale.

Teorema 17.1 Se $X \sim N(\mu, \sigma^2)$ e $T = a + bX$ con $b \neq 0$, allora

$$T \sim N(a + b\mu, b^2\sigma^2).$$

Dimostrazione. Con $Z \sim N(0, 1)$, si ha $Z \sim -Z$ per la simmetria attorno all'origine di $p_Z(z)$. Inoltre, $X \sim \mu + \sigma Z$. Quindi

$$\begin{aligned} T &= a + bX \\ &\sim a + b(\mu + \sigma Z) && \text{definizione di } X \\ &\sim a + b\mu + b\sigma Z \\ &\sim a + b\mu + |b|\sigma Z && Z \sim -Z \\ &\sim N(a + b\mu, b^2\sigma^2). \end{aligned}$$

□

17.3 F.g.m. e momenti di una normale

Sia $Z \sim N(0, 1)$ e $X \sim N(\mu, \sigma^2)$, per cui $X \sim \mu + \sigma Z$. Allora

$$\begin{aligned}
 M_X(t) &= E(e^{tX}) && \text{definizione di f.g.m.} \\
 &= E(e^{t(\mu + \sigma Z)}) && \text{definizione di } X \\
 &= E(e^{t\mu} e^{t\sigma Z}) && \text{proprietà di } \exp(\cdot) \\
 &= e^{t\mu} E(e^{t\sigma Z}) && \text{linearità del valore atteso} \\
 &= e^{t\mu} M_Z(t\sigma) && \text{definizione di f.g.m.}
 \end{aligned}$$

Si ha poi

$$\begin{aligned}
 M_Z(t) &= E(e^{tZ}) \\
 &= \int_{-\infty}^{+\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
 &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2tz + t^2 - t^2)} dz \\
 &= e^{\frac{1}{2}t^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz \\
 &= e^{\frac{1}{2}t^2}
 \end{aligned}$$

perché l'ultimo integrale vale 1. Viene infatti integrata su \mathbb{R} la f.d.p. di $N(t, 1)$. Essendo $M_Z(t)$ finita per ogni $t \in \mathbb{R}$, Z ha f.g.m. propria.

In conclusione, per $X \sim N(\mu, \sigma^2)$, si ha

$$M_X(t) = e^{t\mu} e^{\frac{1}{2}(t\sigma)^2} = e^{t\mu + \frac{1}{2}t^2\sigma^2} \quad (17.4)$$

e quindi anche $X \sim N(\mu, \sigma^2)$ ha f.g.m. propria. Pertanto

$$X \sim N(\mu, \sigma^2) \iff M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

per il risultato di caratterizzazione, cfr. Teorema 16.2.

Derivando la (17.4) si ottiene

$$\begin{aligned}
 M'_X(t) &= (\mu + t\sigma^2)M_X(t) \\
 M''_X(t) &= \sigma^2 M_X(t) + (\mu + t\sigma^2)^2 M_X(t)
 \end{aligned}$$

per cui i momenti primo e secondo di $X \sim N(\mu, \sigma^2)$ sono dati da

$$\begin{aligned}
 E(X) &= M'_X(t) \Big|_{t=0} = \mu \\
 E(X^2) &= M''_X(t) \Big|_{t=0} = \sigma^2 + \mu^2.
 \end{aligned}$$

Quindi μ , oltre che mediana e moda, è anche valore atteso di X . Inoltre,

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \sigma^2 + \mu^2 - (\mu)^2 = \sigma^2.$$

Per le leggi normali, il parametro σ^2 ha il significato di varianza, e σ ha il significato di scarto quadratico medio, come la notazione lascia intendere.

17.4 Proprietà additiva delle normali indipendenti

Teorema 17.2 Se $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$ sono indipendenti, allora

$$S = X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

Dimostrazione. Una f.g.m. propria determina la legge di probabilità, per cui

$$S \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) \iff M_S(t) = e^{t(\mu_X + \mu_Y) + \frac{1}{2}t^2(\sigma_X^2 + \sigma_Y^2)}.$$

Conviene quindi calcolare la f.g.m. di S . Per l'indipendenza, la f.g.m. di S è il prodotto delle f.g.m. di X e di Y . In dettaglio:

$$\begin{aligned} M_S(t) &= E(e^{tS}) && \text{definizione di f.g.m.} \\ &= E(e^{t(X+Y)}) && \text{definizione di } S \\ &= E(e^{tX} e^{tY}) && \text{proprietà di } \exp(\cdot) \\ &= E(e^{tX}) E(e^{tY}) && X \text{ e } Y \text{ indipendenti} \\ &= M_X(t) M_Y(t) && \text{definizione di f.g.m.} \\ &= e^{t\mu_X + \frac{1}{2}t^2\sigma_X^2} e^{t\mu_Y + \frac{1}{2}t^2\sigma_Y^2} && X \text{ e } Y \text{ normali} \\ &= e^{t(\mu_X + \mu_Y) + \frac{1}{2}t^2(\sigma_X^2 + \sigma_Y^2)} && \text{proprietà di } \exp(\cdot) \text{ e raccoglimenti.} \end{aligned}$$

□

La proprietà additiva delle normali indipendenti ha facili generalizzazioni:

- se $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, X e Y sono indipendenti, allora
 $T = aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$ purché $(a, b) \neq (0, 0)$;
- se X_i , $i = 1, \dots, n$, sono v.c. indipendenti con $X_i \sim N(\mu_i, \sigma_i^2)$ allora
 $T = \sum_{i=1}^d a_i X_i \sim N(\sum_{i=1}^d a_i \mu_i, \sum_{i=1}^d a_i^2 \sigma_i^2)$ purché $a_i \neq 0$ per almeno un $i \in \{1, \dots, d\}$.

17.5 Esercizi

Esercizio 17.1 Si mostri che se

$$p'_X(x) = -\frac{x - \mu}{\sigma^2} p_X(x)$$

allora $X \sim N(\mu, \sigma^2)$.

Esercizio 17.2 Si mostri che se $X \sim N(\mu, \sigma^2)$ allora X ha moda μ e $p_X(x)$ ha due punti di flesso, in $x = \mu \pm \sigma$.

Esercizio 17.3 Siano $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$ due v.c. indipendenti. Si mostri che, se a e b non sono entrambi zero,

$$T = aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2).$$

Esercizio 17.4 Siano X_i , $i = 1, \dots, n$, v.c. indipendenti con $X_i \sim N(\mu_i, \sigma_i^2)$. Si mostri che

$$T = \sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right).$$

Esercizio 17.5 Sia $X = (X_1, \dots, X_n)$ una v.c. multivariata con componenti indipendenti $X_i \sim N(\mu_i, \sigma_i^2)$. Si mostri che la v.c. somma, $S_n = \sum_{i=1}^n X_i$, ha legge

$$S_n \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Esercizio 17.6 Sia $X = (X_1, \dots, X_n)$ una v.c. multivariata con componenti indipendenti e identicamente distribuite $X_i \sim N(\mu, \sigma^2)$. Si mostri che la v.c. somma, $S_n = \sum_{i=1}^n X_i$, ha legge

$$S_n \sim N(n\mu, n\sigma^2).$$

Esercizio 17.7 Sia $X = (X_1, \dots, X_n)$ una v.c. multivariata con componenti indipendenti $X_i \sim N(\mu_i, \sigma_i^2)$. Si mostri che la v.c. media aritmetica delle componenti, $\bar{X}_n = \sum_{i=1}^n X_i/n$, ha legge

$$\bar{X}_n \sim N\left(\sum_{i=1}^n \mu_i/n, \sum_{i=1}^n \sigma_i^2/n^2\right).$$

Esercizio 17.8 Sia $X = (X_1, \dots, X_n)$ una v.c. multivariata con componenti indipendenti e identicamente distribuite $X_i \sim N(\mu, \sigma^2)$. Si mostri che la v.c. media aritmetica delle componenti, $\bar{X}_n = \sum_{i=1}^n X_i/n$, ha legge

$$\bar{X}_n \sim N(\mu, \sigma^2/n).$$

Lezione 18

Probabilità e quantili normali

18.1 La f.r. di una v.c. normale

Sia $Z \sim N(0, 1)$ e $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$. Si può ricondurre la f.r. della v.c. normale generale X alla f.r. della normale standard:

$$\begin{aligned} F_X(x) &= P(X \leq x) && \text{definizione di f.r.} \\ &= P(\mu + \sigma Z \leq x) && \text{definizione di } X \\ &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) && \text{algebra} \\ &= F_Z\left(\frac{x - \mu}{\sigma}\right) && \text{definizione di f.r.} \end{aligned}$$

La trasformazione

$$\text{da } x \text{ a } \frac{x - \mu}{\sigma}$$

è detta **standardizzazione**. Esprime x come deviazione da μ in unità di scarto quadratico medio σ .

Per calcolare $F_X(x)$ occorre dunque calcolare la f.r. di $Z \sim N(0, 1)$. Questa è indicata con $\Phi(z)$ ed è definita da

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du. \quad (18.1)$$

La $\Phi(z)$ è una funzione speciale dell'Analisi Matematica, non esprimibile in termini finiti tramite le altre funzioni elementari (polinomi, funzioni razionali, funzioni trigonometriche ed iperboliche, funzioni esponenziali, loro inverse e composizioni delle precedenti funzioni). Studiando l'andamento di $\Phi(z)$, si vede

che risulta strettamente crescente in ogni punto perché

$$\frac{d}{dz}\Phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} > 0$$

con i limiti all'infinito

$$\lim_{z \rightarrow -\infty} \Phi(z) = 0 \quad \text{e} \quad \lim_{z \rightarrow +\infty} \Phi(z) = 1.$$

Inoltre, per la simmetria della densità $N(0, 1)$ attorno all'origine, si ha

$$\Phi(0) = \frac{1}{2} \tag{18.2}$$

e, più in generale,

$$\Phi(-z) = 1 - \Phi(z). \tag{18.3}$$

Infatti, per $z \geq 0$, il cambio di variabile $v = -u$ dà

$$\int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = \int_z^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv.$$

Si noti che, essendo $P(Z = z) = 0$, si ha $P(Z < z) = P(Z \leq z) = \Phi(z)$ e $P(Z \geq z) = P(Z > z) = 1 - \Phi(z)$.

La f.r. di $X \sim N(\mu, \sigma^2)$ è in conclusione

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right). \tag{18.4}$$

L'unico valore della funzione $F_X(\cdot)$ noto esattamente è $F_X(\mu) = \Phi(0) = 0.5$. Tutti gli altri valori saranno da approssimare numericamente.

18.2 Uso delle tavole della f.r. normale standard

Per poter calcolare in via approssimata le probabilità e i quantili normali senza dover disporre di un computer o di una calcolatrice programmabile, è conveniente anche ai nostri giorni far ricorso alla vecchia tecnologia, la tabulazione della funzione $\Phi(\cdot)$.

La relazione (18.3) consente di restringere la tavola a valori positivi di z . Spesso $\Phi(z)$ è tabulata per z che va da 0.00 a 3.99 con passo 0.01. Per contenere la tabulazione in un solo foglio, l'intero e il primo decimale del valore z intestano le righe della tavola, il secondo decimale intesta le colonne della tavola. Si veda la Tabella 18.1. Se ad esempio si desidera ottenere $\Phi(1.05)$, si cerca nella tavola il valore riportato all'incrocio fra la riga intestata 1.0 e la colonna intestata 0.05. In tale posizione si legge 0.85314, per cui $\Phi(1.05) = 0.85314$.

In modo analogo si trova che $\Phi(1) \doteq 0.84134$, per cui

$$\begin{aligned} P(|Z| < 1) &= P(-1 < Z < 1) \\ &= P(-1 < Z \leq 1) \\ &= \Phi(1) - \Phi(-1) \\ &= \Phi(1) - (1 - \Phi(1)) \\ &= 2\Phi(1) - 1 \\ &\doteq 0.68268 \approx 68\%. \end{aligned}$$

Poiché $\Phi(2) \doteq 0.97725$,

$$\begin{aligned} P(|Z| < 2) &= \Phi(2) - \Phi(-2) \\ &= 2\Phi(2) - 1 \\ &\doteq 0.95450 \approx 95.5\%, \end{aligned}$$

valore da confrontare con la minorazione ottenuta dalla diseguaglianza di Čebyshev

$$P(|Z| < 2) \geq 1 - \frac{1}{4} = \frac{3}{4} = 0.75.$$

Similmente, $\Phi(3) \doteq 0.99865$ comporta che

$$\begin{aligned} P(|Z| < 3) &= \Phi(3) - \Phi(-3) \\ &= 2\Phi(3) - 1 \\ &\doteq 0.99730 \approx 99.7\% \end{aligned}$$

e quindi

$$P(|Z| > 3) = 1 - P(|Z| < 3) \doteq 0.00270 \approx 0.3\%,$$

valore molto più piccolo della maggiorazione $1/9 = 0.\bar{1}$ data dalla diseguaglianza di Čebyshev. Per una normale standard, quasi tutta la probabilità è assorbita dall'intervallo $(-3, 3)$. Solo molto raramente si osserveranno realizzazioni di Z che cadono fuori dell'intervallo $(-3, 3)$.

Le tavole consentono un duplice uso, diretto e inverso. Nell'uso diretto, data la v.c. $X \sim N(\mu, \sigma^2)$ e un valore x particolare, si calcola $F_X(x)$. Nell'uso inverso, dato $p \in (0, 1)$, si calcola il quantile- p di X , $x_p = F_X^{-1}(p)$.

18.2.1 Uso diretto: calcolo di probabilità normali

Esempio 18.1 Un impianto di imbottigliamento può essere considerato un particolare strumento di misura affetto da errore. La variabilità del contenuto, X , espresso in millilitri, di bottiglie di latte dal contenuto nominale di un litro (1000 millilitri) è descritta da una legge normale. Si supponga che sia $X \sim N(1010, 100)$. Interessa calcolare la probabilità che il contenuto di una bottiglia estratta a caso dal flusso di produzione non superi il contenuto nominale.

Si ha

$$\begin{aligned}
 P(X \leq 1000) &= F_X(1000) \\
 &= \Phi\left(\frac{1000 - 1010}{10}\right) && \text{standardizzazione} \\
 &= \Phi(-1) \\
 &= 1 - \Phi(1) \\
 &\doteq 1 - 0.84134 = 0.15866.
 \end{aligned}$$

◇

18.2.2 Uso inverso: calcolo di quantili normali

Poiché se $Z \sim N(0, 1)$ si ha $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$, il quantile- p di X è legato al quantile- p di Z dalla relazione

$$x_p = \mu + \sigma z_p. \quad (18.5)$$

Infatti,

$$F_X(x_p) = \Phi\left(\frac{x_p - \mu}{\sigma}\right) = \Phi\left(\frac{\mu + \sigma z_p - \mu}{\sigma}\right) = \Phi(z_p) = p.$$

Solo quantili- p con $p \geq 0.5$ si ottengono dalla tavola di $\Phi(z)$ (lettura inversa: si cerca p entro la tavola e si legge $z_p = \Phi^{-1}(p)$ ai bordi della tavola). Ma questo è quanto serve tenuto conto della relazione

$$z_p = -z_{1-p} \quad (18.6)$$

dovuta alla simmetria attorno all'origine della densità di Z . Infatti,

$$\Phi(-z_{1-p}) = 1 - \Phi(z_{1-p}) = 1 - (1 - p) = p = \Phi(z_p),$$

da cui si giunge alla conclusione perché $\Phi(\cdot)$ è strettamente monotona.

Esempio 18.2 Calcolare il primo percentile di $X \sim N(1010, 100)$, $x_{0.01}$. Si ha

$$\begin{aligned}
 x_{0.01} &= 1010 + 10 z_{0.01} \\
 &= 1010 + 10 (-z_{0.99}) \\
 &= 1010 - 10 z_{0.99} \\
 &\doteq 1010 - 10 * 2.33 = 986.7.
 \end{aligned}$$

Infatti dalla tavola si legge $\Phi(2.32) \doteq 0.98983$ e $\Phi(2.33) \doteq 0.99010$, per cui $z_{0.99} \doteq 2.33$. ◇

18.3 Esercizi svolti

Saper calcolare probabilità e quantili normali è fondamentale. Si presentano due esempi di esercizi tipici.

Esempio 18.3 Se la variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, e in particolare $Y_1 \sim N(3, 2)$, la variabile casuale media campionaria $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ ha legge normale, $\bar{Y}_n \sim N(3, 2/n)$. Sia $n = 50$. Si calcolino $P(\bar{Y}_{50} > 3.4)$ e $P(\bar{Y}_{50} < 2.8)$. Si ottenga infine il cinquantesimo percentile di \bar{Y}_{50} (è il quantile- p con $p = 50/100$).

Con $n = 50$ si ha $2/n = 2/50 = 1/25$, quindi

$$\bar{Y}_{50} \sim N\left(3, \frac{1}{25}\right) \quad \sigma^2 = \frac{1}{25} \implies \sigma = \frac{1}{5} = 0.2$$

per cui

$$P(\bar{Y}_{50} > 3.4) = 1 - P(\bar{Y}_{50} \leq 3.4) = 1 - F_{\bar{Y}_{50}}(3.4) = 1 - \Phi\left(\frac{3.4 - 3}{0.2}\right) = 1 - \Phi(2)$$

che dà

$$P(\bar{Y}_{50} > 3.4) \doteq 1 - 0.97725 = 0.02275.$$

Analogamente

$$P(\bar{Y}_{50} < 2.8) = P(\bar{Y}_{50} \leq 2.8) = F_{\bar{Y}_{50}}(2.8) = \Phi\left(\frac{2.8 - 3}{0.2}\right) = \Phi(-1)$$

che dà

$$P(\bar{Y}_{50} < 2.8) = 1 - \Phi(1) \doteq 1 - 0.84134 = 0.15866.$$

Infine, la mediana di \bar{Y}_{50} è pari a $3 + 0.2z_{0.5} = 3$. ◇

Esempio 18.4 Se la variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, e in particolare $Y_1 \sim N(2, 0.25)$, la variabile casuale media campionaria $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ ha legge normale, $\bar{Y}_n \sim N(2, 0.25/n)$. Sia $n = 25$. Si calcolino $P(\bar{Y}_{25} > 2.3)$ e $P(\bar{Y}_{25} < 1.8)$. Si ottenga infine il decimo percentile di \bar{Y}_{25} (è il quantile- p con $p = 10/100$).

Con $n = 25$ si ha $0.25/n = 0.25/25 = 1/100$, quindi

$$\bar{Y}_{25} \sim N\left(2, \frac{1}{100}\right) \quad \sigma^2 = \frac{1}{100} \implies \sigma = \frac{1}{10} = 0.1$$

per cui

$$P(\bar{Y}_{25} > 2.3) = 1 - P(\bar{Y}_{25} \leq 2.3) = 1 - F_{\bar{Y}_{25}}(2.3) = 1 - \Phi\left(\frac{2.3 - 2}{0.1}\right) = 1 - \Phi(3)$$

che dà

$$P(\bar{Y}_{25} > 2.3) \doteq 1 - 0.99865 = 0.00135.$$

Analogamente

$$P(\bar{Y}_{25} \leq 1.8) = F_{\bar{Y}_{25}}(2.8) = \Phi\left(\frac{1.8 - 2}{0.1}\right) = \Phi(-2)$$

che dà

$$P(\bar{Y}_{25} \leq 1.8) = 1 - \Phi(2) \doteq 1 - 0.97725 = 0.02275.$$

Infine, il decimo percentile di \bar{Y}_{25} è pari a

$$\bar{y}_{250.10} = 2 + 0.1 z_{0.10}$$

e

$$z_{0.10} = -z_{0.90} \doteq -1.28$$

e quindi in conclusione

$$\bar{y}_{250.10} \doteq 2 + 0.1(-1.28) = 2 - 0.128 = 1.872.$$

◇

18.4 Esercizi

Esercizio 18.1 Sia $Z \sim N(0, 1)$. Si calcolino $P(-1 < Z < 2)$ e $P(Z > -2.5)$.

Esercizio 18.2 Sia $Z \sim N(0, 1)$. Si calcoli $P(Z^2 \leq 3.84)$.

Esercizio 18.3 Sia $Z \sim N(0, 1)$. Si calcolino $z_{0.001}$, $z_{0.0025}$, $z_{0.025}$.

Esercizio 18.4 Sia $X \sim N(20, 25)$. Si calcolino $P(X > 35)$, $P(X \geq 36)$, $P(X \geq 37)$.

Esercizio 18.5 Sia $X \sim N(100, 144)$. Si calcolino $P(X \geq 90)$ e $P(X \geq 100)$.

Esercizio 18.6 Siano $X \sim N(10, 16)$ e $Y \sim N(12, 9)$ due v.c. indipendenti, e sia $S = X + Y$. Si calcolino $P(S > 18)$ e $P(S < 25)$. Si ottengano infine il primo e il novantanovesimo percentile di S .

Esercizio 18.7 Siano $X \sim N(1, 1)$ e $Y \sim N(2, 1)$ due v.c. indipendenti, e sia $T = 3X + 4Y$. Si calcolino $P(6 < T < 16)$ e $P(T > 21)$. Si ottengano infine il quinto e il novantacinquesimo percentile di T .

Esercizio 18.8 Siano $X_1 \sim N(1, 1)$ e $X_2 \sim N(1, 1)$ due v.c. indipendenti, e sia $T = \max(X_1, X_2)$. Si calcolino $P(T > 3)$ e $P(T > 5)$. Si ottenga infine la mediana di T .

18.5 Appendice: probabilità e quantili normali con R

Per calcolare la f.d.p. e la f.r. di $X \sim N(\mu, \sigma^2)$ nell'ambiente R si dispone delle funzioni `dnorm` e `pnorm`. Se ad esempio $\mu = 1000$ e lo scarto quadratico medio (*standard deviation* in inglese) è $\sigma = 10$, si ottiene un grafico della f.d.p. con i comandi

```
> x=seq(960,1040,0.01)
> plot(x,dnorm(x, mean=1000, sd=10))
> abline(h=0)
```

Un grafico della f.r. della $N(1000, 100)$ si ottiene con gli analoghi comandi

```
> x=seq(960,1040,0.01)
> plot(x,pnorm(x, mean=1000, sd=10))
> abline(h=0)
> abline(h=1)
```

Ulteriori esempi di calcoli:

```
> pnorm(1000, mean=1010, sd=10)
[1] 0.1586553
> pnorm(1000, mean=1020, sd=10)
[1] 0.02275013
> pnorm(1000, mean=1030, sd=10)
[1] 0.001349898
> qnorm(0.01, mean=1020, sd=10)
[1] 996.7365
```

La funzione `qnorm` calcola il quantile- p della normale con valore atteso `mean` e scarto quadratico medio `sd`.

Tabella 18.1: Funzione di ripartizione della normale standard, $\Phi(z)$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Lezione 19

Campioni e statistiche riassuntive

19.1 Leggi di campionamento casuale semplice

Un esperimento di **campionamento casuale semplice** (c.c.s.) è l'osservazione ripetuta, in circostanze indipendenti, o su unità indipendenti, di un dato fenomeno espresso in forma quantitativa. Il numero di repliche dell'osservazione si dice **numerosità** dell'esperimento di c.c.s.. Tipicamente, di ripetizione in ripetizione, il fenomeno manifesterà una apprezzabile variabilità. Effettuato un esperimento di c.c.s., si ottengono dati numerici $y = (y_1, \dots, y_n)$. I dati y sono interpretati come realizzazione di una v.c. n -variata $Y = (Y_1, \dots, Y_n)$ con componenti Y_i , $i = 1, \dots, n$, indipendenti e identicamente distribuite (i.i.d.). L'interpretazione spiega la variabilità dei valori y_i ma introduce anche l'idea che siano prodotti secondo una distribuzione da indagare.

In sintesi si ha

$$Y = (Y_1, \dots, Y_n) \longrightarrow y = (y_1, \dots, y_n)$$

dove y è il campione di n valori da una popolazione di valori. La legge di probabilità di Y_1 (e anche delle altre componenti, $Y_1 \sim \dots \sim Y_n$) rappresenta la distribuzione nella popolazione dei valori oggetto di indagine. Il supporto di Y , detto anche **spazio campionario** e indicato con \mathcal{Y} , è

$$\mathcal{Y} = S_{Y_1} \times \dots \times S_{Y_n} = S_{Y_1}^n.$$

Se Y_1 ha legge discreta con f.m.p. $p_{Y_1}(y_1; \theta)$ dipendente da un parametro θ , la legge della v.c. Y che esprime un esperimento di c.c.s. con numerosità n ha f.m.p.

$$p_Y(y; \theta) = p_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \theta) = \prod_{i=1}^n p_{Y_1}(y_i; \theta)$$

dove $y \in \mathcal{Y} = S_{Y_1}^n$.

Se Y_1 ha legge continua con f.d.p. $p_{Y_1}(y_1; \theta)$ dipendente da un parametro θ , la legge della v.c. Y che esprime un esperimento di c.c.s. con numerosità n ha f.d.p.

$$p_Y(y; \theta) = p_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \theta) = \prod_{i=1}^n p_{Y_1}(y_i; \theta)$$

sempre per $y \in \mathcal{Y} = S_{Y_1}^n$.

Esempio 19.1 (*C.c.s. con numerosità n da $P(\lambda)$, con valore atteso $\lambda > 0$*) Si supponga di osservare dei conteggi indipendenti su unità di osservazione equivalenti, ad esempio il numero di errori di stampa ogni 1000 parole di un testo. I dati y apparterranno allo spazio campionario $\mathcal{Y} = \mathbb{N}^n$. Con dati di conteggio la distribuzione di riferimento è la Poisson. La v.c. Y generatrice dei dati ha allora supporto $\mathcal{Y} = \mathbb{N}^n$ e f.m.p.

$$p_Y(y; \lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{y_i}}{y_i!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!}$$

per $y \in \mathcal{Y}$. Il campione osservato è $y = (4, 3, 2, 1, 1, 2, 3, 2, 3, 3)$, con $n = 10$. Interessa valutare la distribuzione di popolazione, ossia la legge di Y_1 . Poiché questa è determinata da $\lambda = E(Y_1)$, in definitiva interessa valutare, alla luce del campione osservato, il valore di λ . \diamond

Esempio 19.2 (*C.c.s. con numerosità n da $Exp(\lambda)$, con tasso di guasto $\lambda > 0$*) Si supponga di osservare dei tempi al guasto su unità equivalenti, ad esempio compressori per frigorifero di un certo modello. I dati y apparterranno allo spazio campionario $\mathcal{Y} = [0, +\infty)^n$. Supponendo che l'usura per attrito sia trascurabile, si può assumere come distribuzione di riferimento la esponenziale. La v.c. Y generatrice dei dati ha supporto $\mathcal{Y} = [0, +\infty)^n$ e f.d.p.

$$p_Y(y; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda y_i} = \lambda^n e^{-\lambda \sum_{i=1}^n y_i}$$

per $y \in \mathcal{Y}$. Osservati $n = 15$ compressori, si ottengono i tempi al guasto $y = (1.15, 4.83, 0.59, 0.31, 0.49, 3.35, 0.15, 0.67, 0.29, 0.57, 0.46, 0.43, 0.28, 0.43, 0.13)$ espressi in una opportuna unità di misura (e arrotondati alla seconda cifra decimale). Interessa valutare la distribuzione di popolazione, ossia la legge di Y_1 . Poiché questa è determinata dal tasso di guasto $\lambda = 1/E(Y_1)$, in definitiva interessa valutare, alla luce del campione osservato, il valore di λ , o, equivalentemente, il valore atteso di popolazione $E(Y_1) = 1/\lambda$. \diamond

19.2 Statistiche riassuntive

L'esperimento di c.c.s. con numerosità n viene di solito effettuato per acquisire dai dati informazione sul parametro θ della f.m.p./f.d.p. di Y_1 , quindi della popolazione da cui il campione è tratto. Infatti, nelle situazioni concrete di analisi di dati ottenuti con esperimenti casuali, quasi sempre il valore di θ è ignoto. Il campione osservato permetterà allora di illuminare aspetti altrimenti ignoti della legge di popolazione (la legge della v.c. Y_1 ripetutamente osservata con repliche indipendenti).

Ottenuto un campione y , allo scopo di estrarre da y le informazioni cercate sugli aspetti ignoti di θ , si procederà ad una sintesi dei dati. In pratica, si calcola una opportuna statistica riassuntiva $t = t(y)$, scalare o vettoriale, tipicamente della stessa dimensione di θ . Statistiche opportune sono spesso suggerite dalla forma di $p_Y(y; \theta)$.

Esempio 19.3 (*Un'indagine campionaria*) In una indagine sui comportamenti degli studenti dell'Università di Udine, si è interessati a valutare la frazione di fumatori. Si intervistano allo scopo $n = 400$ studenti (unità statistiche), estratti a caso dalla lista degli iscritti. Si ottengono i dati $y = (y_1, \dots, y_{400})$ dove $y_i = 1$ se l'unità statistica i -esima risponde di essere fumatore, $y_i = 0$ se risponde di non essere fumatore. Poiché $n = 400$ è una piccola frazione degli N iscritti ($n/M < 0.1$), si può trascurare il fatto che le estrazioni sono senza reinserimento e utilizzare come modello di riferimento per questi dati 0/1 leggi di Bernoulli indipendenti e identicamente distribuite. I dati sono pertanto considerati realizzazione di una v.c. $Y = (Y_1, \dots, Y_{400})$ dove le componenti Y_i sono i.i.d. con legge $Bi(1, p)$. Il valore p è la ignota frazione di fumatori nella popolazione (gli studenti dell'Università di Udine). Dunque p è l'ignoto θ in questo esempio. Si ha $y \in \mathcal{Y} = \{0, 1\}^{400}$ e per $y \in \mathcal{Y}$ la f.m.p. di Y è

$$p_Y(y; p) = \prod_{i=1}^{400} p^{y_i} (1-p)^{1-y_i} = p^{\sum_{i=1}^{400} y_i} (1-p)^{400 - \sum_{i=1}^{400} y_i}.$$

La funzione $p_Y(y; p)$, applicata ai dati osservati y , mette in evidenza come sintesi dei dati la statistica $s_{400} = \sum_{i=1}^{400} y_i$. Nel valore scalare s_{400} è contenuta l'informazione che i dati y portano sull'ignoto p . Infatti la dipendenza da p di $p_Y(y; p)$ è tramite la statistica $\sum_{i=1}^{400} y_i$. Due campioni y e y' con lo stesso valore della somma mostrano valori di $p_Y(\cdot; p)$ che dipendono da p nello stesso modo. Essendo equivalente che si osservi l'uno o l'altro, y e y' con la stessa somma delle componenti portano lo stesso messaggio sull'ignoto p . \diamond

Calcolata una statistica riassuntiva $t = t(y)$ si deve capire quale informazione essa porta su θ . È chiaro che una statistica non può indicare con certezza il valore di θ , perché il valore di $t = t(y)$ dipende dal campione estratto y , e quest'ultimo è realizzato dal caso: quindi anche t è realizzato dal caso. Chiaramente, t è realizzazione della v.c. trasformata $T = t(Y)$. L'informazione su θ portata da t è contenuta nella legge di probabilità della v.c. T che genera t .

Definizione 19.1 La **distribuzione campionaria** di una statistica $t = t(y)$ è la legge di probabilità della v.c. trasformata $T = t(Y)$ quando la legge di Y è descritta da $p_Y(y; \theta)$.

Conviene considerare subito un esempio.

Esempio 19.4 Si supponga di aver osservato $y = (y_1, \dots, y_n)$, realizzazione di $Y = (Y_1, \dots, Y_n)$ dove le componenti Y_i sono i.i.d. con legge $Bi(1, p)$, per cui y è espressione di un esperimento di c.c.s. con numerosità n , con $p \in (0, 1)$. Per la proprietà additiva delle binomiali indipendenti con lo stesso p , cfr. paragrafo 16.6.1, la statistica $s_n = \sum_{i=1}^n y_i$ ha distribuzione campionaria

$$S_n = \sum_{i=1}^n Y_i \sim Bi(n, p).$$

Pertanto la v.c. media campionaria $\bar{Y}_n = S_n/n$ ha valore atteso e varianza

$$\begin{aligned} E(\bar{Y}_n) &= E\left(\frac{S_n}{n}\right) = \frac{1}{n}E(S_n) = \frac{1}{n}np = p \\ \text{Var}(\bar{Y}_n) &= \text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\text{Var}(S_n) = \frac{1}{n}np(1-p) = \frac{p(1-p)}{n} \leq \frac{1}{4n}. \end{aligned}$$

La varianza di \bar{Y}_n è molto piccola se n è grande (è non superiore a $1/1600$ se $n = 400$; lo scarto quadratico medio è allora maggiorato da $1/40 = 0.025$). Si deduce che quando n è sufficientemente grande la gran parte dei campioni estraibili produce un valore della statistica $\bar{y}_n = \sum_{i=1}^n y_i/n$ che non si discosta molto da p . Dunque \bar{y}_n è una ragionevole congettura sull'ignoto valore p basata sul campione osservato y . \diamond

Esempio 19.5 Si supponga di aver osservato $y = (y_1, \dots, y_n)$, realizzazione di $Y = (Y_1, \dots, Y_n)$ dove le componenti Y_i sono i.i.d. con legge $P(\lambda)$, con $\lambda > 0$. Per la proprietà additiva delle Poisson indipendenti, cfr. paragrafo 16.6.2, la statistica $s_n = \sum_{i=1}^n y_i$ ha distribuzione campionaria

$$S_n = \sum_{i=1}^n Y_i \sim P(n\lambda).$$

Pertanto la v.c. media campionaria $\bar{Y}_n = S_n/n$ ha valore atteso e varianza

$$\begin{aligned} E(\bar{Y}_n) &= E\left(\frac{S_n}{n}\right) = \frac{1}{n}E(S_n) = \frac{1}{n}n\lambda = \lambda \\ \text{Var}(\bar{Y}_n) &= \text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\text{Var}(S_n) = \frac{1}{n}n\lambda = \frac{\lambda}{n}. \end{aligned}$$

La varianza di \bar{Y}_n è molto piccola se n è grande, per cui quando n è grande la gran parte dei campioni estraibili produce un valore della statistica $\bar{y}_n = \sum_{i=1}^n y_i/n$ che non si discosta molto da λ . Dunque \bar{y}_n è una ragionevole congettura sull'ignoto valore λ basata sul campione osservato y . \diamond

19.3 Campionamento da normale con σ_0^2 noto

Sia $Y = (Y_1, \dots, Y_n)$ la v.c. che descrive un esperimento di campionamento casuale semplice da $N(\mu, \sigma_0^2)$ dove solo μ è ignoto, mentre $\sigma_0^2 > 0$ è noto. Si pensi a n misure ripetute di una quantità μ effettuate utilizzando uno strumento di misura con precisione nota, quindi misure affette da errore con scarto quadratico medio noto σ_0 . Si osserverà allora $Y \rightarrow y \in \mathcal{Y} = \mathbb{R}^n$. La v.c. Y ha supporto $S_Y = \mathbb{R}^n$ e f.d.p., dipendente dall'ignoto μ ,

$$\begin{aligned} p_Y(y; \mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma_0^2}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (y_i - \mu)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (y_i^2 + \mu^2 - 2\mu y_i)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n y_i^2 - \frac{n\mu^2}{2\sigma_0^2} + \frac{\mu}{\sigma_0^2} \sum_{i=1}^n y_i\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n y_i^2\right\} \exp\left\{-\frac{n\mu^2}{2\sigma_0^2} + \frac{\mu}{\sigma_0^2} \sum_{i=1}^n y_i\right\} \\ &= h(y) g\left(\sum_{i=1}^n y_i; \mu\right), \end{aligned}$$

dove il fattore

$$h(y) = \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n y_i^2\right\}$$

non dipende da μ .

L'espressione di $p_Y(y; \mu)$ mette in evidenza come informativa sul parametro ignoto $\theta = \mu$ la statistica $\sum_{i=1}^n y_i$. Infatti la dipendenza da μ di $p_Y(y; \mu)$ è tramite $\sum_{i=1}^n y_i$. Due campioni y e y' con lo stesso valore della somma mostrano valori di $p_Y(\cdot; \mu)$ che dipendono da μ nello stesso modo, essendo $g(\sum_{i=1}^n y_i; \mu) = g(\sum_{i=1}^n y'_i; \mu)$. Per estrarre dai dati l'informazione su μ si deve indagare sulla distribuzione campionaria di $\sum_{i=1}^n y_i$ o di statistiche equivalenti come sintesi dei dati.

Nel c.c.s. da popolazione normale la proprietà additiva delle normali indipendenti, cfr. paragrafo 17.4, comporta che la distribuzione campionaria della statistica somma $s_n = \sum_{i=1}^n y_i$ sia

$$\sum_{i=1}^n Y_i \sim N(n\mu, n\sigma_0^2).$$

L'informazione su μ portata dalla statistica somma è equivalentemente portata dalla statistica media campionaria, $\bar{y}_n = s_n/n$, analogo empirico di $\mu = E(Y)$.

Per la chiusura delle leggi normali sotto trasformazioni affini, cfr. paragrafo 17.2, la distribuzione campionaria della statistica \bar{y}_n è

$$\bar{Y}_n \sim N\left(\mu, \frac{\sigma_0^2}{n}\right). \quad (19.1)$$

Poiché la varianza di \bar{Y}_n è piccola se n è grande, nella gran parte dei campioni estraibili con, ad esempio, $n = 100$, la statistica $\bar{y}_n = \sum_{i=1}^n y_i/n$ non si discosta molto da μ (se σ_0 è dell'ordine di 1) e dunque \bar{y}_n è una ragionevole congettura sull'ignoto valore μ basata sui dati y .

Per l'importanza del risultato, si dà una dimostrazione diretta della (19.1) basata sulla funzione generatrice dei momenti. Si ha

$$\begin{aligned} M_{\bar{Y}_n}(t) &= E\left(e^{t\bar{Y}_n}\right) && \text{def. di f.g.m.} \\ &= E\left(e^{\frac{t}{n} \sum_{i=1}^n Y_i}\right) && \text{def. di } \bar{Y}_n \\ &= E\left(\prod_{i=1}^n e^{\frac{t}{n} Y_i}\right) && \text{proprietà di } \exp(\cdot) \\ &= \prod_{i=1}^n E\left(e^{\frac{t}{n} Y_i}\right) && \text{indip. delle } Y_i \\ &= \left(E\left(e^{\frac{t}{n} Y_1}\right)\right)^n && \text{id. distrib. delle } Y_i \\ &= \left(M_{Y_1}\left(\frac{t}{n}\right)\right)^n && \text{def. di f.g.m.} \\ &= \left(e^{\frac{t}{n}\mu + \frac{1}{2}\left(\frac{t}{n}\right)^2 \sigma_0^2}\right)^n && Y_1 \sim N(\mu, \sigma_0^2) \\ &= e^{n\left(\frac{t}{n}\mu + \frac{1}{2}\frac{t^2}{n^2}\sigma_0^2\right)} && (e^a)^b = e^{ab} \\ &= e^{t\mu + \frac{1}{2}t^2 \frac{\sigma_0^2}{n}} && \text{semplificazione} \end{aligned}$$

da cui si conclude che $\bar{Y}_n \sim N\left(\mu, \frac{\sigma_0^2}{n}\right)$ perché una f.g.m. propria caratterizza la legge di probabilità.

Conviene sottolineare che se è ignota anche la varianza, quindi nel c.c.s. con numerosità n da normale $N(\mu, \sigma^2)$ con entrambi i parametri ignoti, la distribuzione campionaria di \bar{Y}_n risulta ancora normale, e precisamente

$$\bar{Y}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

19.4 Le leggi chi-quadrato

Le leggi chi-quadrato sono leggi associate al c.c.s. da popolazione normale, in particolare a statistiche del tipo somme di quadrati di normali indipendenti con valore atteso zero.

Definizione 19.2 Siano Z_1, Z_2, \dots, Z_ν v.c. i.i.d. con legge $N(0, 1)$. Si dice che la v.c. T_ν ha legge chi-quadrato con ν gradi di libertà, e si scrive $T_\nu \sim \chi_\nu^2$, se T_ν ha la legge di probabilità di $\sum_{i=1}^\nu Z_i^2$.

Conviene indagare in primo luogo la legge chi-quadrato con un grado di libertà, χ_1^2 . È la legge di $T = Z^2$ dove $Z \sim N(0, 1)$. Il supporto di T è $S_T = [0, +\infty)$. Per $t > 0$ la f.r. di T è

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= P(Z^2 \leq t) \\ &= P(-\sqrt{t} \leq Z \leq \sqrt{t}) \\ &= \Phi(\sqrt{t}) - \Phi(-\sqrt{t}) \\ &= \Phi(\sqrt{t}) - (1 - \Phi(\sqrt{t})) \\ &= 2\Phi(\sqrt{t}) - 1. \end{aligned}$$

Poiché $\frac{d}{dz}\Phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$, per $t > 0$ la f.d.p. di T è

$$\begin{aligned} p_T(t) &= \frac{d}{dt}F_T(t) \\ &= 2 \frac{1}{2\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{t})^2} \\ &= c t^{-\frac{1}{2}} e^{-\frac{1}{2}t} \\ &= c t^{\frac{1}{2}-1} e^{-\frac{1}{2}t} \end{aligned}$$

da confrontare con la f.d.p. di $X \sim Ga(\alpha, \lambda)$, che, per $x > 0$, è

$$p_X(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}.$$

Dal confronto si ricava che la legge χ_1^2 è una particolare legge gamma, con parametro di forma $\alpha = \frac{1}{2}$ e parametro di scala $\lambda = \frac{1}{2}$. Quindi

$$\chi_1^2 \sim Ga\left(\frac{1}{2}, \frac{1}{2}\right).$$

Il risultato precedente fa emergere che $\sum_{i=1}^\nu Z_i^2$ è somma di ν v.c. indipendenti con legge $Ga(1/2, 1/2)$. Per la proprietà additiva delle gamma indipendenti con comune parametro di scala, cfr. paragrafo 16.6.3, si conclude che

$$\chi_\nu^2 \sim Ga\left(\frac{\nu}{2}, \frac{1}{2}\right).$$

Poiché se $V \sim Ga(\alpha, \lambda)$ si ha

$$\begin{aligned} E(V) &= \frac{\alpha}{\lambda} \\ \text{Var}(V) &= \frac{\alpha}{\lambda^2} \end{aligned}$$

si ottiene per $T_\nu \sim \chi_\nu^2$

$$\begin{aligned} E(T_\nu) &= \frac{\nu/2}{1/2} = \nu \\ \text{Var}(T_\nu) &= \frac{\nu/2}{(1/2)^2} = 2\nu. \end{aligned} \tag{19.2}$$

Una v.c. con legge χ_ν^2 ha valore atteso pari al numero dei gradi di libertà, ν , e varianza pari al doppio dei gradi di libertà, risultato che si memorizza facilmente.

La Tabella 19.1 dà alcuni quantili delle distribuzioni chi-quadrato in corrispondenza a valori piccoli e moderati dei gradi di libertà, ν . Per la lettura della tavola, si consideri che, ad esempio, la mediana di una v.c. con legge χ_{30}^2 è 29.34.

19.5 La varianza campionaria da normale con μ noto

Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti $Y_i \sim N(\mu_0, \sigma^2)$, dove μ_0 è noto mentre $\sigma^2 > 0$ è ignoto. Si tratta dunque della v.c. descrittiva di un esperimento di c.c.s. con numerosità n da una popolazione normale con solo la varianza ignota. Essendo noto che la media di popolazione è μ_0 , la f.d.p. di Y risulta

$$\begin{aligned} p_Y(y; \mu_0, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \frac{(y_i - \mu_0)^2}{\sigma^2}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_0)^2\right\} \\ &= h(y) g\left(\sum_{i=1}^n (y_i - \mu_0)^2; \sigma^2\right), \end{aligned}$$

scrittura che mette in evidenza come informativa su σ^2 la statistica scalare $\sum_{i=1}^n (y_i - \mu_0)^2$. Dividendo per n si ottiene la statistica **varianza campionaria**

$$\hat{\sigma}_{\mu_0}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_0)^2,$$

analogo empirico di $\sigma^2 = E((Y - \mu_0)^2)$.

La distribuzione campionaria della statistica $\hat{\sigma}_{\mu_0}^2$ è la legge della v.c. trasformata

$$\hat{\sigma}_{\mu_0}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu_0)^2,$$

indicata con lo stesso simbolo della statistica per convenienza di notazione. La legge di probabilità di $\hat{\sigma}_{\mu_0}^2$ si trova facilmente sostituendo nella sua definizione Y_i con $\mu_0 + \sigma Z_i$, dove le Z_i sono i.i.d. $N(0, 1)$. Si ottiene

$$\begin{aligned}\hat{\sigma}_{\mu_0}^2 &= \frac{1}{n} \sum_{i=1}^n (\mu_0 + \sigma Z_i - \mu_0)^2 \\ &= \frac{\sigma^2}{n} \sum_{i=1}^n Z_i^2 \\ &\sim \frac{\sigma^2}{n} \chi_n^2.\end{aligned}$$

Per il risultato (19.2) su valore atteso e varianza di χ_ν^2 ,

$$\begin{aligned}E(\hat{\sigma}_{\mu_0}^2) &= \frac{\sigma^2}{n} n = \sigma^2 \\ \text{Var}(\hat{\sigma}_{\mu_0}^2) &= \frac{\sigma^4}{n^2} 2n = \frac{2\sigma^4}{n}.\end{aligned}$$

La varianza della distribuzione campionaria di $\hat{\sigma}_{\mu_0}^2$ diviene prossima a zero quando n è grande. Quindi il valore assunto dalla statistica $\hat{\sigma}_{\mu_0}^2$ per lo specifico campione sorteggiato sarà facilmente vicino all'ignoto σ^2 .

19.6 C.c.s. da normale con μ e σ^2 ignoti: la varianza campionaria corretta

Si consideri un esperimento di campionamento casuale semplice da una popolazione normale con media e varianza entrambe ignote. La v.c. multivariata descrittiva dell'esperimento è $Y = (Y_1, \dots, Y_n)$ con componenti Y_i i.i.d. con legge $N(\mu, \sigma^2)$, dove i valori $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ sono ignoti.

L'espressione di $p_Y(y; \mu, \sigma^2)$ mette in evidenza come informative sul parametro $\theta = (\mu, \sigma^2)$ le statistiche $\sum_{i=1}^n y_i$ e $\sum_{i=1}^n y_i^2$. Infatti

$$\begin{aligned}p_Y(y; \mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^2 + \mu^2 - 2\mu y_i)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 - \frac{n\mu^2}{2\sigma^2} + \frac{\mu}{\sigma^2} \sum_{i=1}^n y_i\right\} \\ &= h(y) g\left(\sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2; \mu, \sigma^2\right),\end{aligned}$$

dove $h(y) = (1/\sqrt{2\pi})^n$. Per estrarre dai dati l'informazione su μ e σ^2 si deve indagare sulla distribuzione campionaria di $\sum_{i=1}^n y_i$ e $\sum_{i=1}^n y_i^2$ e di statistiche loro collegate.

Chiaramente, se $n = 1$ si ha solo informazione su μ . Si supponga dunque $n \geq 2$. La statistica **varianza campionaria corretta**, indicata con s_n^2 , è una funzione di $\sum_{i=1}^n y_i$ e $\sum_{i=1}^n y_i^2$ informativa su σ^2 . La sua definizione è

$$s_n^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n y_i^2 - \left(\frac{1}{n} \sum_{i=1}^n y_i \right)^2 \right) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2$$

dove $\bar{y}_n = \sum_{i=1}^n y_i/n$ è la media campionaria.

La distribuzione campionaria della statistica s_n^2 è la legge della v.c. trasformata

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2,$$

dove $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ è la v.c. media campionaria, quando $Y = (Y_1, \dots, Y_n)$ ha componenti i.i.d. con legge $N(\mu, \sigma^2)$. Sostituendo Y_i con $\mu + \sigma Z_i$ e \bar{Y}_n con $\mu + \sigma \bar{Z}_n$ dove $\bar{Z}_n = \sum_{i=1}^n Z_i/n$ e le Z_i sono i.i.d. $N(0, 1)$, si ottiene

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (\mu + \sigma Z_i - \mu - \sigma \bar{Z}_n)^2 \\ &= \frac{\sigma^2}{n-1} \sum_{i=1}^n (Z_i - \bar{Z}_n)^2. \end{aligned}$$

Si dimostra che quando le Z_i sono i.i.d. $N(0, 1)$ anche la variabile casuale trasformata $\sum_{i=1}^n (Z_i - \bar{Z}_n)^2$ ha legge chi-quadrato, precisamente

$$\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 \sim \chi_{n-1}^2.$$

Inoltre si dimostra che le v.c. $\sum_{i=1}^n (Z_i - \bar{Z}_n)^2$ e \bar{Z}_n sono indipendenti. Si ha allora per $n \geq 2$

$$S_n^2 \sim \frac{\sigma^2}{n-1} \chi_{n-1}^2,$$

indipendente da

$$\bar{Y}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Se si confronta la distribuzione campionaria di S_n^2 con quella di $\hat{\sigma}_{\mu_0}^2$, si vede che l'effetto di dover stimare μ ignoto è che si perde un grado di libertà. Poiché

$$\begin{aligned} E(S_n^2) &= \frac{\sigma^2}{n-1} (n-1) = \sigma^2 \\ \text{Var}(S_n^2) &= \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}, \end{aligned}$$

per n sufficientemente grande il valore s_n^2 , realizzazione campionaria di S_n^2 , sarà tipicamente vicino al valore di popolazione σ^2 .

Nel c.c.s. da popolazione normale \bar{Y}_n , la quantità stocastica informativa su $\mu = E(Y_1)$, è indipendente da quella informativa su $\sigma^2 = \text{Var}(Y_1)$, S_n^2 . Anche se la popolazione non è normale \bar{Y}_n e S_n^2 stimano $E(Y_1)$ e $\text{Var}(Y_1)$. Tuttavia l'indipendenza fra le v.c. \bar{Y}_n e S_n^2 vale solo nel c.c.s. da una popolazione normale.

19.7 Le leggi t di Student

Sia $Y = (Y_1, \dots, Y_n)$ un c.c.s. da $N(\mu, \sigma_0^2)$ con $\mu \in \mathbb{R}$ ignoto e $\sigma_0^2 > 0$ noto. Poiché $\bar{Y}_n \sim N\left(\mu, \frac{\sigma_0^2}{n}\right)$ la **media campionaria standardizzata**

$$Z = \frac{\bar{Y}_n - \mu}{\frac{\sigma_0}{\sqrt{n}}} = \frac{\sqrt{n}}{\sigma_0}(\bar{Y}_n - \mu) \quad (19.3)$$

ha distribuzione $N(0, 1)$. Se invece $Y = (Y_1, \dots, Y_n)$ è un c.c.s. da $N(\mu, \sigma^2)$ con $\mu \in \mathbb{R}$ ignoto e $\sigma^2 > 0$ pure ignoto, $\bar{Y}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ e non si conosce il denominatore della standardizzazione, σ/\sqrt{n} . Lo si può stimare come $\sqrt{S_n^2/n}$. Sostituire la stima dà luogo alla **media campionaria studentizzata**

$$T_n = \frac{\bar{Y}_n - \mu}{S_n/\sqrt{n}} = \frac{\sqrt{n}}{S_n}(\bar{Y}_n - \mu). \quad (19.4)$$

Per studiare la distribuzione campionaria di T_n è utile la seguente definizione.

Definizione 19.3 Siano $Z \sim N(0, 1)$ e $W_\nu \sim \chi_\nu^2$ due v.c. indipendenti. Allora la legge di probabilità di

$$S_\nu = \frac{Z}{\sqrt{\frac{W_\nu}{\nu}}}$$

è detta t di Student con ν gradi di libertà. In breve si scrive $S_\nu \sim t_\nu$.

Poiché $Z \sim -Z$ anche $S_\nu \sim -S_\nu$ per cui t_ν ha distribuzione simmetrica attorno a zero. La Tabella 19.2 mostra alcuni quantili di una v.c. con legge t_ν per valori piccoli dei gradi di libertà. La simmetria attorno all'origine comporta che per $p < 0.5$ si abbia $t_{\nu, p} = -t_{\nu, 1-p}$.

Tornando alla statistica (19.4), si vede che la distribuzione campionaria di T_n è t di Student con $n-1$ gradi di libertà. Infatti, sostituendo Y_i con $\mu + \sigma Z_i$ dove le Z_i sono i.i.d. $N(0, 1)$, $i = 1, \dots, n$, si ha

$$\bar{Y}_n \sim \mu + \frac{\sigma}{n} \sum_{i=1}^n Z_i \quad \text{e} \quad S_n^2 \sim \frac{\sigma^2}{(n-1)} \sum_{i=1}^n (Z_i - \bar{Z}_n)^2$$

per cui

$$T_n \sim \frac{\sqrt{n} \left(\mu + \frac{\sigma}{n} \sum_{i=1}^n Z_i - \mu \right)}{\sqrt{\frac{\sigma^2}{(n-1)} \sum_{i=1}^n (Z_i - \bar{Z}_n)^2}} \sim \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 / (n-1)}} \sim \frac{Z}{\sqrt{W_{n-1} / (n-1)}}$$

dove $Z = \sum_{i=1}^n Z_i / \sqrt{n} \sim N(0, 1)$ e $W_{n-1} = \sum_{i=1}^n (Z_i - \bar{Z}_n)^2 / (n-1) \sim \chi_{n-1}^2$ sono indipendenti, per cui $T_n \sim t_{n-1}$.

19.8 Esercizi svolti

Esempio 19.6 La variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, in particolare $Y_1 \sim N(3, 2)$. Si mostri che la variabile casuale $\bar{Y}_n = \sum_{i=1}^n Y_i / n$ ha legge normale, $\bar{Y}_n \sim N(3, 2/n)$.

Poiché $X \sim N(\mu, \sigma^2) \iff M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$, occorre mostrare che $M_{\bar{Y}_n}(t) = e^{3t + \frac{1}{2}t^2 \frac{2}{n}}$. Si ha

$$\begin{aligned} M_{\bar{Y}_n}(t) &= E\left(e^{t\bar{Y}_n}\right) \\ &= E\left(e^{t \sum_{i=1}^n Y_i / n}\right) \\ &= E\left(\prod_{i=1}^n e^{tY_i / n}\right) \\ &= \prod_{i=1}^n E\left(e^{\frac{t}{n}Y_i}\right) && \text{per l'indipendenza delle } Y_i \\ &= \prod_{i=1}^n M_{Y_i}\left(\frac{t}{n}\right) \\ &= \left(M_{Y_1}\left(\frac{t}{n}\right)\right)^n && \text{per l'identica distribuzione delle } Y_i \\ &= \left(e^{3\frac{t}{n} + \frac{1}{2}\left(\frac{t}{n}\right)^2 2}\right)^n \\ &= e^{n\left(3\frac{t}{n} + \frac{1}{2}\frac{t^2}{n^2} 2\right)} \\ &= e^{3t + \frac{1}{2}t^2 \frac{2}{n}}. \end{aligned}$$

◇

Esempio 19.7 La variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, in particolare $Y_1 \sim N(2, 0.25)$. Si mostri che la variabile casuale $\bar{Y}_n = \sum_{i=1}^n Y_i / n$ ha legge normale, $\bar{Y}_n \sim N(2, 0.25/n)$.

Poiché $X \sim N(\mu, \sigma^2) \iff M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$, occorre mostrare che $M_{\bar{Y}_n}(t) = e^{2t + \frac{1}{2}t^2 \frac{0.25}{n}}$. Si ha

$$\begin{aligned}
 M_{\bar{Y}_n}(t) &= E\left(e^{t\bar{Y}_n}\right) \\
 &= E\left(e^{t\sum_{i=1}^n Y_i/n}\right) \\
 &= E\left(\prod_{i=1}^n e^{tY_i/n}\right) \\
 &= \prod_{i=1}^n E\left(e^{\frac{t}{n}Y_i}\right) && \text{per l'indipendenza delle } Y_i \\
 &= \prod_{i=1}^n M_{Y_i}\left(\frac{t}{n}\right) \\
 &= \left(M_{Y_1}\left(\frac{t}{n}\right)\right)^n && \text{per l'identica distribuzione delle } Y_i \\
 &= \left(e^{2\frac{t}{n} + \frac{1}{2}\left(\frac{t}{n}\right)^2 0.25}\right)^n \\
 &= e^{n\left(2\frac{t}{n} + \frac{1}{2}\frac{t^2}{n^2} 0.25\right)} \\
 &= e^{2t + \frac{1}{2}t^2 \frac{0.25}{n}}.
 \end{aligned}$$

◇

19.9 Esercizi

Esercizio 19.1 Sia $T \sim \chi_{10}^2$. Si calcolino $E(T)$ e $\text{Var}(T)$. Si ottengano il primo e il novantanovesimo percentile di T .

Esercizio 19.2 Sia $T \sim \chi_{\nu}^2$. Si discuta se per la mediana di T è ragionevole l'approssimazione $t_{0.5} \doteq \nu - 2/3$.

Esercizio 19.3 Sia $T \sim t_{\nu}$. Si ottenga la mediana di T .

Esercizio 19.4 Sia $T \sim t_{12}$. Si ottengano il novantesimo e il novantacinquesimo percentile di T e li si confronti con quelli di $Z \sim N(0, 1)$.

Esercizio 19.5 Sia $Y = (Y_1, \dots, Y_{21})$ una v.c. con componenti i.i.d. aventi legge $N(3, 16)$. Si ottengano il primo e il novantanovesimo percentile di $S_{21}^2 = \sum_{i=1}^{21} (Y_i - \bar{Y}_{21})^2/20$.

Esercizio 19.6 Sia $Y = (Y_1, \dots, Y_{21})$ una v.c. con componenti i.i.d. aventi legge $N(3, 25)$. Si ottengano il primo e il novantanovesimo percentile di $S_{21}^2 = \sum_{i=1}^{21} (Y_i - \bar{Y}_{21})^2/20$.

Esercizio 19.7 (*Compito del 01/02/2017*) La variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, in particolare $Y_1 \sim N(15, 4)$. Si mostri che la variabile casuale $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ ha legge normale, $\bar{Y}_n \sim N(15, 4/n)$. Sia $n = 100$. Si calcolino $P(\bar{Y}_{100} > 16)$ e $P(\bar{Y}_{100} < 14)$. Si ottenga infine il novantacinquesimo percentile di \bar{Y}_{100} (è il quantile- p con $p = 95/100$).

Esercizio 19.8 (*Compito del 10/02/2017*) La variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, in particolare $Y_1 \sim N(5, 9)$. Si mostri che la variabile casuale $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ ha legge normale, $\bar{Y}_n \sim N(5, 9/n)$. Sia $n = 900$. Si calcolino $P(\bar{Y}_{900} > 5.2)$ e $P(\bar{Y}_{900} < 4.9)$. Si ottenga infine il novantanovesimo percentile di \bar{Y}_{900} (è il quantile- p con $p = 99/100$).

Esercizio 19.9 (*Compito del 16/06/2017*) La variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, in particolare $Y_1 \sim N(1, 4)$. Si mostri che la variabile casuale $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ ha legge normale, $\bar{Y}_n \sim N(1, 4/n)$. Sia $n = 400$. Si calcolino $P(\bar{Y}_{400} > 1.25)$ e $P(\bar{Y}_{400} < 1.8)$. Si ottenga infine il novantacinquesimo percentile di \bar{Y}_{400} (è il quantile- p con $p = 95/100$).

Esercizio 19.10 (*Compito del 17/07/2017*) La variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, in particolare $Y_1 \sim N(0, 1)$. Si mostri che la variabile casuale $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ ha legge normale, $\bar{Y}_n \sim N(0, 1/n)$. Sia $n = 25$. Si calcolino $P(\bar{Y}_{25} > 0.4)$ e $P(\bar{Y}_{25} < -0.6)$. Si ottenga infine il novantanovesimo percentile di \bar{Y}_{25} (è il quantile- p con $p = 99/100$).

Esercizio 19.11 (*Compito del 11/09/2017*) La variabile casuale multivariata (Y_1, \dots, Y_n) ha componenti indipendenti e identicamente distribuite con legge marginale normale, in particolare $Y_1 \sim N(-2, 1)$. Si mostri che la variabile casuale $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ ha legge normale, $\bar{Y}_n \sim N(-2, 1/n)$. Sia $n = 9$. Si calcolino $P(\bar{Y}_9 > -2)$ e $P(\bar{Y}_9 < -2.5)$. Si ottenga infine il novantesimo percentile di \bar{Y}_9 (è il quantile- p con $p = 90/100$).

Tabella 19.1: Quantili- p di χ^2_ν .

ν	0.01	0.025	0.05	0.10	0.20	0.50	0.80	0.90	0.95	0.975	0.99
1	0.00	0.00	0.00	0.02	0.06	0.45	1.64	2.71	3.84	5.02	6.63
2	0.02	0.05	0.10	0.21	0.45	1.39	3.22	4.61	5.99	7.38	9.21
3	0.11	0.22	0.35	0.58	1.01	2.37	4.64	6.25	7.81	9.35	11.34
4	0.30	0.48	0.71	1.06	1.65	3.36	5.99	7.78	9.49	11.14	13.28
5	0.55	0.83	1.15	1.61	2.34	4.35	7.29	9.24	11.07	12.83	15.09
6	0.87	1.24	1.64	2.20	3.07	5.35	8.56	10.64	12.59	14.45	16.81
7	1.24	1.69	2.17	2.83	3.82	6.35	9.80	12.02	14.07	16.01	18.48
8	1.65	2.18	2.73	3.49	4.59	7.34	11.03	13.36	15.51	17.53	20.09
9	2.09	2.70	3.33	4.17	5.38	8.34	12.24	14.68	16.92	19.02	21.67
10	2.56	3.25	3.94	4.87	6.18	9.34	13.44	15.99	18.31	20.48	23.21
11	3.05	3.82	4.57	5.58	6.99	10.34	14.63	17.28	19.68	21.92	24.72
12	3.57	4.40	5.23	6.30	7.81	11.34	15.81	18.55	21.03	23.34	26.22
13	4.11	5.01	5.89	7.04	8.63	12.34	16.98	19.81	22.36	24.74	27.69
14	4.66	5.63	6.57	7.79	9.47	13.34	18.15	21.06	23.68	26.12	29.14
15	5.23	6.26	7.26	8.55	10.31	14.34	19.31	22.31	25.00	27.49	30.58
16	5.81	6.91	7.96	9.31	11.15	15.34	20.47	23.54	26.30	28.85	32.00
17	6.41	7.56	8.67	10.09	12.00	16.34	21.61	24.77	27.59	30.19	33.41
18	7.01	8.23	9.39	10.86	12.86	17.34	22.76	25.99	28.87	31.53	34.81
19	7.63	8.91	10.12	11.65	13.72	18.34	23.90	27.20	30.14	32.85	36.19
20	8.26	9.59	10.85	12.44	14.58	19.34	25.04	28.41	31.41	34.17	37.57
21	8.90	10.28	11.59	13.24	15.44	20.34	26.17	29.62	32.67	35.48	38.93
22	9.54	10.98	12.34	14.04	16.31	21.34	27.30	30.81	33.92	36.78	40.29
23	10.20	11.69	13.09	14.85	17.19	22.34	28.43	32.01	35.17	38.08	41.64
24	10.86	12.40	13.85	15.66	18.06	23.34	29.55	33.20	36.42	39.36	42.98
25	11.52	13.12	14.61	16.47	18.94	24.34	30.68	34.38	37.65	40.65	44.31
26	12.20	13.84	15.38	17.29	19.82	25.34	31.79	35.56	38.89	41.92	45.64
27	12.88	14.57	16.15	18.11	20.70	26.34	32.91	36.74	40.11	43.19	46.96
28	13.56	15.31	16.93	18.94	21.59	27.34	34.03	37.92	41.34	44.46	48.28
29	14.26	16.05	17.71	19.77	22.48	28.34	35.14	39.09	42.56	45.72	49.59
30	14.95	16.79	18.49	20.60	23.36	29.34	36.25	40.26	43.77	46.98	50.89
40	22.16	24.43	26.51	29.05	32.34	39.34	47.27	51.81	55.76	59.34	63.69
50	29.71	32.36	34.76	37.69	41.45	49.33	58.16	63.17	67.50	71.42	76.15
60	37.48	40.48	43.19	46.46	50.64	59.33	68.97	74.40	79.08	83.30	88.38
80	53.54	57.15	60.39	64.28	69.21	79.33	90.41	96.58	101.88	106.63	112.33
100	70.06	74.22	77.93	82.36	87.95	99.33	111.67	118.50	124.34	129.56	135.81

Tabella 19.2: Quantili- p di t_ν .

ν	0.80	0.90	0.95	0.975	0.995	0.9975	0.999
1	1.376	3.078	6.314	12.706	63.657	127.321	318.309
2	1.061	1.886	2.920	4.303	9.925	14.089	22.327
3	0.978	1.638	2.353	3.182	5.841	7.453	10.215
4	0.941	1.533	2.132	2.776	4.604	5.598	7.173
5	0.920	1.476	2.015	2.571	4.032	4.773	5.893
6	0.906	1.440	1.943	2.447	3.707	4.317	5.208
7	0.896	1.415	1.895	2.365	3.499	4.029	4.785
8	0.889	1.397	1.860	2.306	3.355	3.833	4.501
9	0.883	1.383	1.833	2.262	3.250	3.690	4.297
10	0.879	1.372	1.812	2.228	3.169	3.581	4.144
11	0.876	1.363	1.796	2.201	3.106	3.497	4.025
12	0.873	1.356	1.782	2.179	3.055	3.428	3.930
13	0.870	1.350	1.771	2.160	3.012	3.372	3.852
14	0.868	1.345	1.761	2.145	2.977	3.326	3.787
15	0.866	1.341	1.753	2.131	2.947	3.286	3.733
16	0.865	1.337	1.746	2.120	2.921	3.252	3.686
17	0.863	1.333	1.740	2.110	2.898	3.222	3.646
18	0.862	1.330	1.734	2.101	2.878	3.197	3.610
19	0.861	1.328	1.729	2.093	2.861	3.174	3.579
20	0.860	1.325	1.725	2.086	2.845	3.153	3.552
21	0.859	1.323	1.721	2.080	2.831	3.135	3.527
22	0.858	1.321	1.717	2.074	2.819	3.119	3.505
23	0.858	1.319	1.714	2.069	2.807	3.104	3.485
24	0.857	1.318	1.711	2.064	2.797	3.091	3.467
25	0.856	1.316	1.708	2.060	2.787	3.078	3.450
26	0.856	1.315	1.706	2.056	2.779	3.067	3.435
27	0.855	1.314	1.703	2.052	2.771	3.057	3.421
28	0.855	1.313	1.701	2.048	2.763	3.047	3.408
29	0.854	1.311	1.699	2.045	2.756	3.038	3.396
30	0.854	1.310	1.697	2.042	2.750	3.030	3.385
\vdots							
∞	0.842	1.282	1.645	1.960	2.576	2.807	3.090

Lezione 20

Media campionaria e popolazioni non normali

20.1 Attesa e varianza della media campionaria

Sotto c.c.s. da popolazione normale con valore atteso μ e varianza σ^2 , per la media campionaria vale $\bar{Y}_n \sim N(\mu, \sigma^2/n)$, si è visto nel paragrafo 19.3. Anche quando l'esperimento di campionamento casuale semplice con numerosità n è da una popolazione non normale, la media campionaria è informativa sul valore atteso della popolazione, sotto opportune condizioni.

Teorema 20.1 *Sotto c.c.s. con numerosità n da Y_1 con valore atteso $E(Y_1) = \mu$ e varianza $\text{Var}(Y_1) = \sigma^2$ finiti, la v.c. media campionaria, $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$, ha valore atteso e varianza che risultano*

$$E(\bar{Y}_n) = \mu \quad (20.1)$$

$$\text{Var}(\bar{Y}_n) = \frac{\sigma^2}{n}. \quad (20.2)$$

Dimostrazione. Si ha

$$\begin{aligned} E(\bar{Y}_n) &= E\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n E(Y_i) && \text{linearità del v.a.} \\ &= \frac{1}{n} \sum_{i=1}^n \mu && \text{identica distribuzione} \\ &= \frac{1}{n} n\mu = \mu \end{aligned}$$

e

$$\begin{aligned}
\text{Var}(\bar{Y}_n) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n Y_i\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(Y_i) && \text{indipendenza delle } Y_i \\
&= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 && \text{identica distribuzione} \\
&= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}. \quad \square
\end{aligned}$$

Se la popolazione non è normale, quindi Y_1 non ha legge normale, si sa dunque che

$$\bar{Y}_n \sim \mathcal{L}_n\left(\mu, \frac{\sigma^2}{n}\right),$$

dove $X \sim \mathcal{L}(\mu, \sigma^2)$ indica che X ha una legge di probabilità con $E(X) = \mu$ e $\text{Var}(X) = \sigma^2$. Di solito non è facile determinare la forma precisa della legge \mathcal{L}_n , che dipende da n . Alcuni risultati di limite permettono però di approssimare \mathcal{L}_n quando n è sufficientemente grande.

20.2 Modi di convergenza di successioni di v.c.

Si richiama che si dice che una successione di reali a_n , $n = 1, 2, \dots$, ha limite $a \in \mathbb{R}$ se per ogni $\varepsilon > 0$ esiste un naturale \bar{n}_ε tale che

$$n > \bar{n}_\varepsilon \implies |a_n - a| < \varepsilon.$$

Se si considera, invece di una successione di numeri reali, una successione di v.c. univariate X_n , $n = 1, 2, \dots$, si possono descrivere due principali modi di convergenza. Per $n > \bar{n}$:

- i) il valore realizzato da X_n ‘cambia poco’ al variare di n perché con elevata probabilità è assai vicino a un particolare valore non stocastico;
- ii) il valore realizzato da X_n può continuare a cambiare al variare di n ma la legge di probabilità di X_n ‘varia poco’ al variare di n .

Si contrappongono quindi stabilità dei valori e stabilità delle leggi. Si danno le seguenti definizioni.

Definizione 20.1 Si dice che la successione di v.c. univariate X_n **converge in probabilità** a $c \in \mathbb{R}$ se per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P(|X_n - c| > \varepsilon) = 0.$$

Si scrive allora $X_n \xrightarrow{p} c$.

Definizione 20.2 Si dice che la successione di v.c. univariate X_n **converge in distribuzione** alla v.c. univariata X , e si scrive $X_n \xrightarrow{d} X$, se vale

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

per ogni $x \in \mathbb{R}$ in cui $F_X(x)$ è continua.

Esempio 20.1 Sia $Y = (Y_1, \dots, Y_n)$ con Y_i i.i.d. con legge marginale $U(0, 1)$, $i = 1, \dots, n$. Si consideri $T_n = \min_{i=1, \dots, n} Y_i$. La v.c. univariata T_n ha f.r.

$$F_{T_n}(t) = 1 - (1 - t)^n, \quad t \in [0, 1].$$

Per ogni $\varepsilon \in (0, 1)$ si ha $P(|T_n| > \varepsilon) = 1 - F_{T_n}(\varepsilon) = (1 - \varepsilon)^n$, per cui $\lim_{n \rightarrow +\infty} P(|T_n| > \varepsilon) = 0$. Pertanto $T_n \xrightarrow{p} 0$. Per n molto molto grande le realizzazioni di T_n variano poco al variare di n perché devono essere assai vicine a 0. \diamond

Esempio 20.2 Nelle assunzioni dell'Esempio 20.1, si consideri la successione $X_n = nT_n$. Per $0 < x < n$ si ha

$$F_{X_n}(x) = P(X_n \leq x) = P(nT_n \leq x) = P(T_n \leq \frac{x}{n}) = 1 - \left(1 - \frac{x}{n}\right)^n$$

per cui per $x > 0$, grazie al limite notevole (8.2), si ha

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = 1 - e^{-x} = F_X(x),$$

e si conclude che $X_n \xrightarrow{d} X \sim \text{Esp}(1)$. Quindi per n sufficientemente grande non si può prevedere in modo accurato la particolare realizzazione di X_n , ma si possono fare affermazioni quale $P(2 < X_n < 5) \doteq e^{-2} - e^{-5} \doteq 0.1286$. \diamond

20.3 Due semplici condizioni sufficienti

Una condizione sufficiente per la convergenza in probabilità molto utile nelle applicazioni e con assunzioni assai naturali è la seguente.

Teorema 20.2 Condizione sufficiente affinché $X_n \xrightarrow{p} c$ è che

$$\lim_{n \rightarrow +\infty} E(X_n) = c$$

e che

$$\lim_{n \rightarrow +\infty} \text{Var}(X_n) = 0.$$

Dimostrazione. Fissato $\varepsilon > 0$ si ha

$$\begin{aligned} 0 &\leq P(|X_n - c| \geq \varepsilon) = P((X_n - c)^2 \geq \varepsilon^2) \\ &= P\left((X_n - c)^2 \geq \frac{\varepsilon^2}{E((X_n - c)^2)} E((X_n - c)^2)\right) \\ &\leq \frac{E((X_n - c)^2)}{\varepsilon^2} \quad \text{per la disuguaglianza di Markov} \\ &= \frac{\text{Var}(X_n) + (E(X_n) - c)^2}{\varepsilon^2}. \end{aligned}$$

Per le assunzioni fatte, al divergere di n il numeratore dell'ultima maggiorazione tende a 0, e dunque anche $\lim_{n \rightarrow \infty} P(|X_n - c| > \varepsilon) = 0$, per ogni $\varepsilon > 0$. \square

Si dà, senza dimostrazione, anche una condizione sufficiente per la convergenza in distribuzione. Il risultato completa il quadro delle proprietà della funzione generatrice dei momenti.

Teorema 20.3 *Condizione sufficiente affinché $X_n \xrightarrow{d} X$, con X_n e X v.c. univariate aventi funzione generatrice dei momenti propria, è che esista $\varepsilon > 0$ tale che per ogni $t \in (-\varepsilon, \varepsilon)$ valga*

$$\lim_{n \rightarrow +\infty} M_{X_n}(t) = M_X(t).$$

Esempio 20.3 Sia $X_n \sim Bi(n, \lambda/n)$, dove $0 < \lambda < n$. La funzione generatrice dei momenti di X_n è propria, e risulta

$$M_{X_n}(t) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n} e^t\right)^n = \left(1 + \frac{\lambda(e^t - 1)}{n}\right)^n.$$

Per la (8.2), per ogni $t \in \mathbb{R}$ si ha

$$\lim_{n \rightarrow \infty} M_{X_n}(t) = e^{\lambda(e^t - 1)} = M_X(t)$$

dove $X \sim P(\lambda)$. Si conclude che $X_n \xrightarrow{d} P(\lambda)$. \diamond

20.4 La legge dei grandi numeri

Una legge dei grandi numeri è un risultato che mostra la convergenza di una successione di medie campionarie a un valore deterministico. Il seguente Teorema enuncia la legge dei grandi numeri nella sua forma più semplice.

Teorema 20.4 *Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti Y_i indipendenti e identicamente distribuite con attesa $E(Y_1) = \mu$ e varianza $\text{Var}(Y_1) = \sigma^2$ finite. Allora per la successione delle medie campionarie, $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$, vale*

$$\bar{Y}_n \xrightarrow{p} \mu.$$

Dimostrazione. Per il Teorema 20.1, $E(\bar{Y}_n) = \mu$ e $\text{Var}(\bar{Y}_n) = \sigma^2/n$. Quindi $\lim_{n \rightarrow +\infty} E(\bar{Y}_n) = \mu$ (banalmente) e $\lim_{n \rightarrow +\infty} \text{Var}(\bar{Y}_n) = 0$. Si applica dunque la condizione sufficiente per la convergenza in probabilità, Teorema 20.2. \square

In pratica, disponendo di un campione con numerosità n molto, molto grande (spesso dell'ordine del milione o più)

$$\bar{Y}_n \sim D(\mu),$$

ossia nelle condizioni dette la legge \mathcal{L}_n di \bar{Y}_n è approssimata sufficientemente bene da una legge degenere in μ . Se μ è ignoto, con $n \gg 1$, dove il simbolo \gg indica molto, molto più grande, qualunque campione y sarà estratto avrà una media campionaria osservata molto vicina all'ignoto μ , e dunque μ sarà agli effetti pratici surrogabile da \bar{y}_n . Schematicamente,

$$\bar{Y}_n \longrightarrow \bar{y}_n \doteq \mu.$$

20.5 Il teorema centrale del limite

Un teorema centrale del limite è un risultato che mostra la convergenza in distribuzione di una successione di medie campionarie centrate ad una legge normale. L'enunciato più semplice da dimostrare è il seguente.

Teorema 20.5 *Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti Y_i indipendenti e identicamente distribuite con attesa $E(Y_1) = \mu$ e varianza $\text{Var}(Y_1) = \sigma^2 > 0$ finite, e con funzione generatrice dei momenti propria. Allora*

$$\sqrt{n}(\bar{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Dimostrazione. Senza perdita di generalità si supponga $\mu = 0$. La successione di v.c. da considerare è dunque $\sqrt{n}\bar{Y}_n$. Per il Teorema 20.3, basta mostrare che $\lim_{n \rightarrow +\infty} M_{\sqrt{n}\bar{Y}_n}(t) = e^{\frac{1}{2}t^2\sigma^2}$. Poiché $\bar{Y}_n = \frac{1}{n}S_n$, dove $S_n = \sum_{i=1}^n Y_i$, si ha

$$\begin{aligned} M_{\sqrt{n}\bar{Y}_n}(t) &= E(e^{t\sqrt{n}\frac{1}{n}S_n}) \\ &= E(e^{\frac{t}{\sqrt{n}}S_n}) \\ &= M_{S_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n \\ &= \left(1 + \frac{t}{\sqrt{n}}E(Y_1) + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 E(Y_1^2) + \dots\right)^n \\ &= \left(1 + \frac{\frac{1}{2}t^2\sigma^2}{n} + \dots\right)^n \end{aligned}$$

dove con \dots si indicano termini trascurabili al divergere di n . Si è usata l'assunzione $E(Y_1) = 0$ e il fatto che allora $E(Y_1^2) = (E(Y_1))^2 + \text{Var}(Y_1) = \sigma^2$. In conclusione

$$\lim_{n \rightarrow +\infty} M_{\sqrt{n}\bar{Y}_n}(t) = \lim_{n \rightarrow +\infty} \left(1 + \frac{\frac{1}{2}t^2\sigma^2}{n} + \dots \right)^n = e^{\frac{1}{2}t^2\sigma^2}.$$

□

Osservazione 1. L'assunzione che Y_1 abbia funzione generatrice propria abbrevia la dimostrazione ma il risultato vale in generale per v.c. i.i.d. con attesa e varianza finite.

Osservazione 2. Se $\sqrt{n}(\bar{Y}_n - \mu)$ ha legge approssimata normale, anche le trasformazioni affini di $\sqrt{n}(\bar{Y}_n - \mu)$ sono in modo approssimato normalmente distribuite. In particolare, dal teorema centrale del limite si ottiene una approssimazione normale anche per la legge della media campionaria \bar{Y}_n e per la legge della somma $S_n = \sum_{i=1}^n Y_i$. La legge normale approssimante ha lo stesso valore atteso e la stessa varianza della statistica considerata.

Osservazione 3. In pratica, disponendo di un campione con numerosità n sufficientemente grande (tipicamente dell'ordine di qualche decina, non milioni come nella legge dei grandi numeri) si ha, per la distribuzione della media campionaria, l'approssimazione

$$\bar{Y}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

indipendentemente dalla particolare legge di popolazione (la legge di Y_1).

20.6 Probabilità normali per probabilità non normali

Sia $Y = (Y_1, \dots, Y_n)$ con Y_i i.i.d. $Bi(1, p)$, $i = 1, \dots, n$, dove $p \in (0, 1)$. Si consideri $S_n = \sum_{i=1}^n Y_i$. Per la proprietà additiva delle binomiali indipendenti con lo stesso p si ha che $S_n \sim Bi(n, p)$. D'altra parte Y_1 ha valore atteso finito p e varianza pure finita $p(1-p)$, per cui valgono le assunzioni del teorema centrale del limite. Pertanto

$$\sqrt{n}(\bar{Y}_n - p) \xrightarrow{d} N(0, p(1-p)).$$

Per n sufficientemente grande valgono dunque le relazioni

$$\begin{aligned}\sqrt{n}(\bar{Y}_n - p) &\sim N(0, p(1-p)) \\ \bar{Y}_n &\sim N\left(p, \frac{p(1-p)}{n}\right) \\ S_n &\sim N(np, np(1-p)).\end{aligned}$$

Dall'ultima si ottiene l'approssimazione delle probabilità binomiali con le probabilità normali

$$Bi(n, p) \dot{\sim} N(np, np(1-p))$$

per n convenientemente grande. Indicativamente, n può essere considerato sufficientemente grande perché l'approssimazione sia decente se sia np sia $n(1-p)$ sono maggiori di 5. Ad esempio, si consideri il numero di teste in $n = 400$ lanci di una moneta equilibrata, S_{400} . Si ha $p = 0.5$ e sia np sia $n(1-p)$ valgono 200, per cui con buona approssimazione

$$S_{400} \dot{\sim} N(200, 100)$$

e quindi

$$P(180 \leq S_{400} \leq 220) \doteq \Phi(2) - \Phi(-2) \doteq 0.95449.$$

In modo analogo, se $Y = (Y_1, \dots, Y_n)$ con Y_i i.i.d. $P(\lambda)$, $i = 1, \dots, n$, dove $\lambda > 0$, per n sufficientemente grande si ha

$$\begin{aligned} \bar{Y}_n &\dot{\sim} N\left(E(Y_1), \frac{\text{Var}(Y_1)}{n}\right) \\ &\dot{\sim} N\left(\lambda, \frac{\lambda}{n}\right) \end{aligned}$$

$$\begin{aligned} S_n &\dot{\sim} N(E(S_n), \text{Var}(S_n)) \\ &\dot{\sim} N(n\lambda, n\lambda). \end{aligned}$$

Per la proprietà additiva delle Poisson indipendenti si ha anche $S_n \sim P(n\lambda)$, per cui per λ grande vale l'approssimazione delle probabilità Poisson con le probabilità normali

$$P(\lambda) \dot{\sim} N(\lambda, \lambda).$$

L'approssimazione può essere considerata proponibile per $\lambda > 10$.

Come terzo esempio, se $Y = (Y_1, \dots, Y_n)$ con Y_i i.i.d. $Exp(\lambda)$, $i = 1, \dots, n$, dove $\lambda > 0$, per n sufficientemente grande si ha

$$\begin{aligned} \bar{Y}_n &\dot{\sim} N\left(E(Y_1), \frac{\text{Var}(Y_1)}{n}\right) \\ &\dot{\sim} N\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right) \end{aligned}$$

$$\begin{aligned} S_n &\dot{\sim} N(E(S_n), \text{Var}(S_n)) \\ &\dot{\sim} N\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right). \end{aligned}$$

Per la proprietà additiva delle gamma indipendenti con lo stesso λ si ha $S_n \sim Ga(n, \lambda)$, per cui per α grande vale anche l'approssimazione delle probabilità gamma con le probabilità normali

$$Ga(\alpha, \lambda) \dot{\sim} N\left(\frac{\alpha}{\lambda}, \frac{\alpha}{\lambda^2}\right).$$

L'approssimazione può essere considerata proponibile per $\alpha > 15$.

20.7 Esercizi

Esercizio 20.1 Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti i.i.d. con legge marginale $P(\lambda)$, $\lambda > 0$. Si mostri che la successione $\bar{Y}_n = \sum_{i=1}^n Y_i/n$ converge in probabilità a λ .

Esercizio 20.2 Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti i.i.d. con legge marginale $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Si mostri che la successione di v.c. $S_n^2 = \sum_{i=1}^n (Y_i - \bar{Y}_n)^2/(n-1)$, dove $\bar{Y}_n = \sum_{i=1}^n Y_i/n$, converge in probabilità a σ^2 .

Esercizio 20.3 Sia $X \sim Bi(10^4, 0.5)$. Si calcolino in via approssimata $P(X > 4900)$, $P(X \leq 5200)$, il novantanovesimo percentile di X .

Esercizio 20.4 Sia $X \sim Bi(20, 0.5)$. Si calcoli in via approssimata $P(X \leq 15)$, usando la correzione di continuità, ossia come fosse da calcolare $P(X < 15.5)$ tramite l'approssimazione normale.

Esercizio 20.5 Sia $X \sim P(400)$. Si calcolino in via approssimata $P(X > 440)$, $P(X \leq 480)$, il primo percentile di X .

Esercizio 20.6 Sia $X \sim P(20)$. Si calcoli in via approssimata $P(X = 20)$, usando la correzione di continuità, ossia come fosse da calcolare $P(19.5 < X < 20.5)$ tramite l'approssimazione normale.

Esercizio 20.7 Si mostri che quando ν è sufficientemente grande

$$\chi_\nu^2 \dot{\sim} N(\nu, 2\nu).$$

Esercizio 20.8 Sia $W \sim \chi_{200}^2$. Sfruttando il risultato dell'esercizio precedente si calcolino in via approssimata $P(W > 260)$ e il novantacinquesimo percentile di W .

Lezione 21

Inferenza statistica: stima puntuale

21.1 La variabilità campionaria

Nell'inferenza statistica si vuole studiare una popolazione tramite un campione casuale di unità statistiche tratte da quella popolazione. Le conclusioni, o inferenze, che si possono trarre su una popolazione grazie a un campione casuale, sono affette da un errore dovuto al caso. Mentre l'errore indotto da una scelta 'di buon senso' di unità statistiche rappresentative della popolazione non è descrivibile, l'errore dovuto al caso è descrivibile con gli strumenti forniti dal Calcolo delle Probabilità. Quantificare l'incertezza delle conclusioni dovuta alla natura casuale dei dati è dunque l'obiettivo fondamentale dell'inferenza statistica.

Nel seguito si studieranno procedure di inferenza statistica in presenza di dati campionari (y_1, \dots, y_n) visti come realizzazione di una v.c. di campionamento casuale semplice con numerosità n . Quindi lo schema di base è:

(dati) $y = (y_1, \dots, y_n) \leftarrow Y = (Y_1, \dots, Y_n)$ dove le Y_i sono i.i.d. (modello).

Il campione effettivamente estratto è solo uno dei campioni potenzialmente estraibili. Se si rifacesse l'esperimento di campionamento casuale semplice con numerosità n si otterrebbero dati y' diversi da y . Quindi si otterrebbero conclusioni o inferenze sulla popolazione in qualche misura diverse. Occorre saper valutare quanto diverse potranno ragionevolmente essere.

21.2 La distribuzione empirica

Un **modello statistico** è una collezione di leggi di probabilità per la v.c. Y generatrice dei dati y . Quale particolare collezione sia utile per analizzare i dati y dipenderà dalle conoscenze sul fenomeno oggetto di studio.

Il più generale modello di c.c.s. con numerosità n è il seguente:

- $Y = (Y_1, \dots, Y_n)$ ha componenti i.i.d.
- la legge di Y_1 è determinata da una ignota funzione di ripartizione $F_{Y_1}(y_1)$.

Secondo questo modello, tutte le f.r. di una v.c. univariata sono ritenute possibili per Y_1 . Non si dispone di informazioni ulteriori sulla popolazione che permettano di descrivere più dettagliatamente la forma di $F_{Y_1}(y_1)$.

I dati $y = (y_1, \dots, y_n)$, realizzazione di $Y = (Y_1, \dots, Y_n)$ che segue il modello appena descritto, possono dare informazione sulla funzione ignota $F_{Y_1}(\cdot)$ tramite la **distribuzione empirica**. La distribuzione empirica pone peso $1/n$ su ciascun y_i osservato, $i = 1, \dots, n$. Il peso è cumulato se vi sono più unità statistiche che danno lo stesso valore y_i come osservazione, se vi sono, come si dice in inglese, *ties*. La distribuzione empirica è dunque $\hat{Y} \sim Ud(y_1, y_2, \dots, y_n)$, con possibilità di ripetizione dei valori y_i .

La **funzione di ripartizione empirica** è la f.r. della distribuzione empirica, ossia

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(y_i), \quad x \in \mathbb{R}. \quad (21.1)$$

Sopra, $I_{(-\infty, x]}(\cdot)$ è la funzione indicatrice

$$I_{(-\infty, x]}(y_i) = \begin{cases} 1 & \text{se } y_i \leq x \\ 0 & \text{se } y_i > x. \end{cases}$$

La f.r. empirica associa dunque a ogni $x \in \mathbb{R}$ la frequenza relativa di valori osservati y_i che sono minori o uguali a x . Il grafico della f.r. empirica sarà quello di una f.r. di tipo discreto con supporto sui valori osservati, quindi $\hat{F}_n(x)$ ha un salto di $1/n$ sopra ogni dato (salvo *ties*).

Il valore atteso della distribuzione empirica è

$$E(\hat{Y}) = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}_n,$$

la statistica media campionaria. La varianza della distribuzione empirica è

$$\text{Var}(\hat{Y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)^2,$$

ossia la statistica varianza campionaria non corretta. Analogamente si possono calcolare momenti e quantili di \hat{Y} , indicati con $\hat{\mu}_r$ e \hat{y}_p , $r \in \mathbb{N}^+$ e $p \in (0, 1)$.

Si fissi ora l'attenzione su come varia la f.r. empirica al variare dei potenziali campioni. Così pensata, $\hat{F}_n(x)$ diventa una funzione non più di y ma di Y , quindi un oggetto determinato dal caso. In particolare, per ogni fissato x , $\hat{F}_n(x)$ è la v.c. univariata

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(Y_i).$$

Le indicatrici $I_{(-\infty, x]}(Y_i)$ sono v.c. i.i.d. che assumono valori in $\{0, 1\}$, quindi con legge marginale di Bernoulli con $p = P(I_{(-\infty, x]}(Y_i) = 1)$, ossia

$$I_{(-\infty, x]}(Y_i) \sim Bi(1, P(Y_i \leq x)) \sim Bi(1, F_{Y_1}(x)).$$

Pertanto

$$\sum_{i=1}^n I_{(-\infty, x]}(Y_i) \sim Bi(n, F_{Y_1}(x))$$

per la proprietà additiva delle binomiali indipendenti con lo stesso p , cfr. paragrafo 16.6.1. Si ha allora, per il risultato su valore atteso e varianza di una binomiale,

$$\begin{aligned} E(\hat{F}_n(x)) &= \frac{1}{n} n F_{Y_1}(x) = F_{Y_1}(x) \\ \text{Var}(\hat{F}_n(x)) &= \frac{1}{n^2} n F_{Y_1}(x)(1 - F_{Y_1}(x)) = \frac{F_{Y_1}(x)(1 - F_{Y_1}(x))}{n}. \end{aligned}$$

Si intuisce che per n molto grande l'ignota f.r. $F_{Y_1}(x)$ è ricostruibile dal campione $y = (y_1, \dots, y_n)$ con elevata precisione. Valgono infatti il risultato di convergenza in probabilità

$$\hat{F}_n(x) \xrightarrow{p} F_{Y_1}(x) \quad \text{per ogni } x \in \mathbb{R},$$

per la condizione sufficiente (Teorema 20.2) evidentemente soddisfatta, e il risultato di convergenza in distribuzione

$$\sqrt{n}(\hat{F}_n(x) - F_{Y_1}(x)) \xrightarrow{d} N(0, F_{Y_1}(x)(1 - F_{Y_1}(x)))$$

dal teorema centrale del limite per le binomiali elementari i.i.d..

Se qualche aspetto della distribuzione di popolazione, la legge di Y_1 , è noto, il compito di ricostruire la legge di probabilità di Y_1 sarà ulteriormente facilitato.

21.3 Stime e stimatori

In presenza di qualche elemento di informazione sulla distribuzione di popolazione, un modello statistico di campionamento casuale semplice con numerosità n ha di solito la forma di una collezione di f.m.p./f.d.p., quindi

$$\mathcal{F} = \left\{ p_Y(y; \theta) = \prod_{i=1}^n p_{Y_1}(y_i; \theta), \quad y \in \mathcal{Y}, \theta \in \Theta \right\}.$$

Quando $\Theta \subseteq \mathbb{R}^p$, il modello \mathcal{F} è detto **modello statistico parametrico**.

Si suppone che il modello statistico \mathcal{F} contenga tra i suoi elementi, almeno con una ragionevole approssimazione, anche la vera legge di probabilità di $Y = (Y_1, \dots, Y_n)$. In altre parole, si suppone che fra i valori di θ in Θ ce ne sia uno, indicato con θ^0 , che è il vero valore del parametro.

L'inferenza statistica è un insieme di procedure utili per localizzare θ^0 entro Θ . La forma più semplice di localizzazione è tramite una stima puntuale.

Si dice che $\hat{\theta}$ è una **procedura di stima puntuale** di θ (intendendo di θ^0) se è una applicazione

$$\hat{\theta} : \mathcal{Y} \longrightarrow \Theta.$$

Quindi una stima puntuale, identificata con una procedura di stima puntuale, è una statistica $\hat{\theta} : \mathcal{Y} \ni y \rightarrow \hat{\theta}(y) \in \Theta$ definita per ogni $y \in \mathcal{Y}$. In corrispondenza al valore y osservato, la statistica produce la **stima** $\hat{\theta}(y)$.

Spesso sarà interessante seguire il comportamento di una procedura di stima puntuale al crescere della numerosità campionaria n . Si userà allora la notazione $\hat{\theta}_n$ che sottolinea la dipendenza da n dei risultati.

Per valutare l'efficacia di una procedura di stima puntuale è vano calcolare per un dato y la stima $\hat{\theta}(y)$ e confrontarla con θ^0 , il valore di θ che ha generato i dati y come realizzazione di Y . Il campione sorteggiato è solo uno dei campioni potenzialmente sorteggiabili.

Occorre, come si è fatto per la funzione di ripartizione empirica, guardare il comportamento della v.c. trasformata $\hat{\theta} = \hat{\theta}(Y)$, ossia la distribuzione campionaria della statistica $\hat{\theta}(Y)$.

Definizione 21.1 La v.c. $\hat{\theta} = \hat{\theta}(Y)$ è detta **stimatore**.

Definizione 21.2 La **distribuzione campionaria dello stimatore** $\hat{\theta}$ è la legge di probabilità della v.c. trasformata $\hat{\theta}(Y)$ sotto θ , ossia quando la legge di Y è descritta da $p_Y(y; \theta)$.

Esempio 21.1 (C.c.s. con numerosità n da $P(\lambda)$)

La f.m.p. congiunta della v.c. di c.c.s. $Y = (Y_1, \dots, Y_n)$ per $y \in \mathcal{Y} = \mathbb{N}^n$,

$$p_Y(y; \lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{y_i}}{y_i!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!},$$

pone in evidenza la statistica $\sum_{i=1}^n y_i$ come informativa su λ . D'altro lato, il valore atteso $E(Y_1) = \lambda$ ha come controparte empirica $E(\hat{Y}) = \bar{y}_n$, che è funzione di $\sum_{i=1}^n y_i$. Quindi conviene considerare la procedura di stima puntuale

$$\hat{\lambda}_n = \hat{\lambda}_n(y) = \bar{y}_n.$$

Lo stimatore è

$$\hat{\lambda}_n = \bar{Y}_n,$$

la v.c. media campionaria. Sulla distribuzione di $\hat{\lambda}_n$ sotto λ il Calcolo delle Probabilità fornisce molte informazioni. Ad esempio le relazioni

$$E(\hat{\lambda}_n) = \lambda \quad \text{e} \quad \text{Var}(\hat{\lambda}_n) = \frac{\lambda}{n}$$

valgono sotto ogni $\lambda > 0$, per cui quando n è grande la stima varia poco attorno al vero valore di λ , la media di popolazione, al variare del campione sorteggiato. \diamond

21.4 Proprietà campionarie degli stimatori

Uno stimatore ha l'obiettivo di fornire stime di θ di solito non lontane dal vero valore del parametro. Affinché ciò accada, è bene che la legge di probabilità dello stimatore sotto θ abbia alcune proprietà desiderabili. Si supponga per semplicità che θ sia scalare. Il modello statistico è detto allora monoparametrico: $\theta \in \Theta \subseteq \mathbb{R}$. Le principali proprietà campionarie desiderabili per gli stimatori di un parametro scalare sono espresse dalle seguenti definizioni.

Definizione 21.3 *Uno stimatore $\hat{\theta}_n$ è detto **non distorto** per θ se, sotto θ ,*

$$E(\hat{\theta}_n) = \theta \quad \text{per ogni } \theta \in \Theta.$$

Uno stimatore non distorto pone il vero valore del parametro al centro della propria distribuzione di probabilità, dovunque θ^0 stia in Θ . Ad esempio, nel modello di c.c.s. con numerosità n da una normale con varianza nota, \bar{Y}_n è stimatore non distorto di μ .

Definizione 21.4 *Uno stimatore $\hat{\theta}_n$ è detto **asintoticamente non distorto** per θ se, sotto θ ,*

$$\lim_{n \rightarrow +\infty} E(\hat{\theta}_n) = \theta \quad \text{per ogni } \theta \in \Theta.$$

Quando n è sufficientemente grande, uno stimatore asintoticamente non distorto pone il vero valore del parametro al centro della propria distribuzione di probabilità, dovunque θ^0 stia in Θ . Ad esempio, nel modello di c.c.s. con numerosità n da una normale con media nota, $\hat{\sigma}_n^2 = \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 / n$ è stimatore asintoticamente non distorto di σ^2 . Infatti $E(\hat{\sigma}_n^2) = \sigma^2(n-1)/n$.

Definizione 21.5 *Uno stimatore $\hat{\theta}_n$ è detto **consistente** per θ se, sotto θ ,*

$$\hat{\theta}_n \xrightarrow{p} \theta$$

per ogni $\theta \in \Theta$.

Se lo stimatore è consistente, quando vi è moltissima informazione nel campione perché n è molto, molto grande, praticamente ogni campione sorteggiabile dà una stima del parametro molto vicina al vero valore del parametro, qualunque esso sia in Θ .

Definizione 21.6 Uno stimatore $\hat{\theta}_n$ è detto **asintoticamente normale** se, sotto θ ,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$$

per ogni $\theta \in \Theta$. La funzione $\sigma^2(\theta)$ è detta **varianza asintotica** dello stimatore.

Se lo stimatore è asintoticamente normale, quando vi è una decente informazione nel campione perché n è sufficientemente grande, l'errore di stima ha una distribuzione approssimata normale centrata, qualunque sia il vero valore del parametro in Θ . In pratica, sotto θ , vale l'approssimazione $\hat{\theta}_n \sim N(\theta, \sigma^2(\theta)/n)$ quando n è sufficientemente grande.

Uno stimatore asintoticamente normale con la più piccola varianza asintotica $\sigma^2(\theta)$ è asintoticamente ottimo o, come si dice, **asintoticamente efficiente**.

21.5 Come reperire stimatori

Per reperire stimatori, si seguono certi accorgimenti applicabili ad una vasta platea di modelli statistici, noti come **metodi di stima**.

Il metodo di stima più elementare è il **metodo dell'analogia**. È adatto a trattare modelli di campionamento casuale semplice. Poiché la distribuzione empirica è in grado di 'localizzare' la vera legge di Y_1 , basta attribuire a θ un significato come momento o quantile di Y_1 e, per analogia, stimare θ applicando lo stesso significato alla distribuzione empirica.

Ad esempio si stima

$$\theta = E(Y_1)$$

con

$$\hat{\theta}_n = E(\hat{Y}) = \bar{y}_n.$$

Per sottolineare che il calcolo del valore atteso è fatto sotto θ , si può usare la notazione esplicita $E_\theta(Y_1)$ invece della consueta notazione implicita, $E(Y_1)$.

Più in generale, si consideri un modello statistico parametrico di campionamento casuale semplice con un parametro p -dimensionale $\theta = (\theta_1, \dots, \theta_p)$. Se le componenti di θ sono espresse in funzione dei momenti $E_\theta(Y_1), E_\theta(Y_1^2), \dots, E_\theta(Y_1^p)$ tramite le equazioni

$$\theta_j = g_j(E_\theta(Y_1), E_\theta(Y_1^2), \dots, E_\theta(Y_1^p)), \quad j = 1, \dots, p,$$

dove le g_j sono funzioni continue, allora usando i momenti della distribuzione empirica si stimerà θ_j con

$$\hat{\theta}_j = g_j\left(\frac{1}{n} \sum y_i, \frac{1}{n} \sum y_i^2, \dots, \frac{1}{n} \sum y_i^p\right), \quad j = 1, \dots, p.$$

Il metodo di stima appena descritto è noto come **metodo dei momenti**.

Un ulteriore metodo di stima molto semplice è il **metodo di sostituzione**. Se si dispone per θ di una stima $\hat{\theta}_n$ e si è interessati a stimare la funzione di θ

$$\tau = g(\theta),$$

dove $g(\cdot)$ è una funzione continua, si usa come stima di τ , per sostituzione,

$$\hat{\tau}_n = g(\hat{\theta}_n).$$

Esempio 21.2 (*C.c.s. con num. n da $Exp(\lambda)$, stima del tasso di guasto λ*)

Sia $Y = (Y_1, \dots, Y_n)$ con componenti i.i.d. aventi legge $Exp(\lambda)$ dove $\lambda > 0$ è il tasso di guasto. Poiché sotto λ

$$E(Y_1) = \frac{1}{\lambda}$$

per il metodo dei momenti $\theta = 1/\lambda$ è stimato da $\hat{\theta}_n = \bar{y}_n$. Allora $\lambda = 1/\theta$ sarà stimato per sostituzione da

$$\hat{\lambda}_n = \frac{1}{\hat{\theta}_n} = \frac{1}{\bar{y}_n}.$$

◇

Il metodo di stima adatto a trattare i modelli statistici parametrici in generale (non solo quelli di campionamento casuale semplice) e in grado di produrre stimatori asintoticamente efficienti è il **metodo della massima verosimiglianza**. La funzione di verosimiglianza (*likelihood function*) per i dati y e il modello statistico parametrico $\mathcal{F} = \{p_Y(y; \theta), y \in \mathcal{Y}, \theta \in \Theta \subseteq \Theta\}$ è

$$L(\theta) = L(\theta; y) = p_Y(y; \theta) \quad \text{dove } \theta \in \Theta.$$

È quindi la f.m.p./f.d.p. del modello calcolata ai dati. Per un modello di c.c.s. la funzione di verosimiglianza è semplicemente $L(\theta) = \prod_{i=1}^n p_{Y_1}(y_i; \theta)$.

Definizione 21.7 *Data una funzione di verosimiglianza $L(\theta) = L(\theta; y)$, si dice stima di massima verosimiglianza di θ un valore $\hat{\theta} \in \Theta$ tale che*

$$L(\hat{\theta}) \geq L(\theta) \quad \text{per ogni } \theta \in \Theta.$$

Si noti che la costruzione di una stima di massima verosimiglianza di θ è automatica una volta che si disponga dei dati y e del modello statistico parametrico \mathcal{F} . Può essere affidata a un algoritmo di ottimizzazione. In alcuni modelli particolarmente semplici la stima di massima verosimiglianza si calcola anche esplicitamente e non solo numericamente.

Esempio 21.3 (*C.c.s. con num. n da $Exp(\lambda)$, stima del tasso di guasto λ*)

Sia $Y = (Y_1, \dots, Y_n)$ con componenti i.i.d. aventi legge $Exp(\lambda)$ dove $\lambda > 0$ è il tasso di guasto. Con i dati y la funzione di verosimiglianza è

$$L(\lambda) = L(\lambda; y) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{y_i}}{y_i!} = \frac{1}{\prod y_i!} e^{-n\lambda} \lambda^{\sum y_i} \quad \text{per } \lambda > 0.$$

Anziché massimizzare direttamente $L(\lambda)$, conviene osservare che la funzione $\ell(\lambda) = \log L(\lambda)$ ha il suo massimo nello stesso punto (il logaritmo è una funzione monotona crescente), ma ha espressione molto più semplice da studiare, risultando

$$\ell(\lambda) = \log \left(\frac{1}{\prod y_i!} \right) - n\lambda + \sum_{i=1}^n y_i \log \lambda.$$

L'equazione

$$\frac{d}{d\lambda} \ell(\lambda) = -n + \frac{\sum_{i=1}^n y_i}{\lambda} = 0$$

mostra che l'unico punto stazionario è

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n y_i} = \frac{1}{\bar{y}_n}$$

che risulta un massimo perché

$$\frac{d^2}{d\lambda^2} \ell(\lambda) = -\frac{\sum_{i=1}^n y_i}{\lambda^2} < 0.$$

◇

21.6 Lo standard error di uno stimatore

Sia $\tau = g(\theta)$ un parametro d'interesse scalare. Come misura della precisione dello stimatore $\hat{\tau} = g(\hat{\theta})$ si considera l'errore quadratico medio di stima sotto θ ,

$$EQM_{\theta}(\hat{\tau}) = E_{\theta} \left\{ \left(g(\hat{\theta}) - g(\theta) \right)^2 \right\}.$$

Se $g(\hat{\theta})$ è stimatore non distorto di $g(\theta)$, ossia $E_{\theta} \left(g(\hat{\theta}) \right) = g(\theta)$ per ogni $\theta \in \Theta$, allora

$$EQM_{\theta}(\hat{\tau}) = \text{Var}_{\theta} \left(g(\hat{\theta}) \right)$$

altrimenti in generale

$$EQM_{\theta}(\hat{\tau}) = \text{Var}_{\theta} \left(g(\hat{\theta}) \right) + \left(E_{\theta} \left(g(\hat{\theta}) \right) - g(\theta) \right)^2,$$

dove $E_{\theta} \left(g(\hat{\theta}) \right) - g(\theta)$ è la **distorsione** di $g(\hat{\theta})$.

Lo **standard error** di $\hat{\tau}$ è la radice quadrata positiva di $EQM_{\theta}(\hat{\tau})$,

$$se_{\theta}(\hat{\tau}) = \sqrt{E_{\theta} \left\{ \left(g(\hat{\theta}) - g(\theta) \right)^2 \right\}}.$$

Esso dà una misura della tipica grandezza dell'errore di stima. Poiché $se_\theta(\hat{\tau})$ dipende da θ , viene generalmente stimato per sostituzione. Si ottiene così l'**estimated standard error**

$$\hat{se}(\hat{\tau}) = se_{\hat{\theta}}(\hat{\tau}) = \sqrt{EQM_{\hat{\theta}}(\hat{\tau})}$$

che corrisponde alla stima di $se_\theta(\hat{\tau})$ con i dati del campione.

È un fatto notevole che lo stesso campione che permette di stimare $\tau = g(\theta)$ fornisca anche una stima della accuratezza della procedura di stima di τ , tramite $\hat{se}(\hat{\tau})$.

È buona pratica nelle applicazioni riportare non solo il valore di stima dell'ignoto parametro oggetto di studio, ma completare l'informazione fornendo anche il corrispondente *estimated standard error*.

Esempio 21.4 (*C.c.s. con numerosità n da $P(\lambda)$*)

Si è già ottenuta la stima

$$\hat{\lambda}_n = \hat{\lambda}_n(y) = \bar{y}_n.$$

Lo stimatore è la media campionaria

$$\hat{\lambda}_n = \bar{Y}_n,$$

e risulta non distorto. Quindi

$$EQM_\lambda(\hat{\lambda}_n) = \text{Var}(\hat{\lambda}_n) = \frac{\lambda}{n}$$

per cui

$$se_\lambda(\hat{\lambda}_n) = \frac{\sqrt{\lambda}}{\sqrt{n}}$$

e infine

$$\hat{se}(\hat{\lambda}_n) = \frac{\sqrt{\hat{\lambda}_n}}{\sqrt{n}}.$$

Se $n = 100$ e $\bar{y}_{100} = 25$, λ è stimato da $\hat{\lambda}_{100} = 25$ con *estimated standard error* $\hat{se}(\hat{\lambda}_{100}) = \frac{\sqrt{25}}{\sqrt{100}} = 0.5$. Sembra ragionevole congetturare che il vero valore di λ stia nell'intervallo (24, 26). \diamond

21.7 Esercizi

Esercizio 21.1 Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti i.i.d. con legge marginale $N(\mu, \sigma_0^2)$, dove $\mu \in \mathbb{R}$ è ignoto e $\sigma_0^2 > 0$ è noto. Si mostri che lo stimatore di μ , $\bar{Y}_n = \sum_{i=1}^n Y_i/n$, è non distorto, consistente e asintoticamente normale.

Esercizio 21.2 Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti i.i.d. con legge marginale $N(\mu_0, \sigma^2)$, dove $\mu_0 \in \mathbb{R}$ è noto e $\sigma^2 > 0$ è ignoto. Si mostri che lo stimatore di σ^2 , $\hat{\sigma}_{\mu_0}^2 = \sum_{i=1}^n (Y_i - \mu_0)^2/n$, è non distorto, consistente e asintoticamente normale.

Esercizio 21.3 Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti i.i.d. con legge marginale $P(\lambda)$, $\lambda > 0$ ignoto. Si mostri che lo stimatore $\hat{\lambda}_n = \bar{Y}_n = \sum_{i=1}^n Y_i/n$ è non distorto e consistente per λ .

Esercizio 21.4 Sia $Y = (Y_1, \dots, Y_n)$ una v.c. multivariata con componenti i.i.d. con legge marginale $Bi(1, p)$, $p \in (0, 1) > 0$ ignoto. Si mostri che $\hat{p}_n = \bar{Y}_n = \sum_{i=1}^n Y_i/n$ è uno stimatore di p non distorto, consistente e asintoticamente normale.

Esercizio 21.5 Il responsabile del controllo di qualità di una certa azienda meccanica vuole valutare la proporzione p di pezzi prodotti che non risultano idonei alla vendita. Viene estratto un campione casuale di 200 pezzi prodotti in un certo giorno e si osserva che 35 di essi non sono idonei. Si definisca un modello statistico parametrico utile per descrivere l'esperimento effettuato. Si reperisca uno stimatore per la proporzione in esame, si determini l'associato *standard error* e si calcolino le corrispondenti stime. Si indagino le proprietà campionarie dello stimatore di p .

Esercizio 21.6 Dato un campione costituito da variabili casuali Y_1, \dots, Y_n , $n > 1$, indipendenti con legge normale con media μ e varianza σ^2 , entrambe ignote, si reperisca uno stimatore di σ^2 e indagino le proprietà campionarie dello stimatore considerato.

Esercizio 21.7 Un'azienda manifatturiera vuole studiare la resistenza media alla rottura di un determinato materiale isolante. A tale scopo esegue 30 prove di resistenza alla rottura valutando il peso in *kg* necessario a rompere l'isolante. L'esperimento ha fornito i seguenti risultati: $\sum_{i=1}^{30} y_i = 51702$ *kg* e $\sum_{i=1}^{30} y_i^2 = 89335770$ *kg*². Nell'ipotesi che il modello statistico di riferimento sia normale, si determini una stima per la resistenza media μ , con l'associato *standard error* stimato. Si indagino le proprietà campionarie dello stimatore di μ considerato.

Esercizio 21.8 Siano Y_1, \dots, Y_n v.c. i.i.d. con legge $N(\mu, \sigma^2)$, dove $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ sono ignoti. Si mostri che $S = \sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2 / (n-1)}$, dove \bar{Y}_n è la media campionaria, è uno stimatore distorto di σ , in particolare $E_{\mu, \sigma^2}(S) \leq \sigma$. Sugg.: $\text{Var}_{\mu, \sigma^2}(S) = E_{\mu, \sigma^2}(S^2) - \{E_{\mu, \sigma^2}(S)\}^2 \geq 0$.

Esercizio 21.9 (*Compito del 01/02/2017*) Il responsabile qualità di una fabbrica vuole valutare la proporzione p di pezzi prodotti che non sono idonei alla vendita. Viene estratto dalla produzione del giorno un campione casuale di 400 pezzi e si osserva che 17 di essi non sono idonei. Si definisca un modello statistico parametrico utile per descrivere l'esperimento effettuato. Si reperisca uno stimatore per p , si determini l'associato *standard error*, e si calcolino le corrispondenti stime.

Esercizio 21.10 (*Compito del 10/02/2017*) Dato un campione y_1, \dots, y_n , realizzazione di variabili casuali Y_1, \dots, Y_n , $n > 1$, indipendenti con legge di Poisson di media $\lambda > 0$ ignota, si reperisca una stima di λ e si indagino le proprietà campionarie dello stimatore considerato.

Esercizio 21.11 (*Compito del 16/06/2017*) Dato un campione y_1, \dots, y_n , realizzazione di variabili casuali Y_1, \dots, Y_n , $n > 1$, indipendenti con legge normale di media $\mu \in \mathbb{R}$ ignota e varianza nota pari a 1, si reperisca uno stima di μ e si indaghino le proprietà campionarie dello stimatore corrispondente.

Esercizio 21.12 (*Compito del 17/07/2017*) Dato un campione y_1, \dots, y_n , realizzazione di variabili casuali Y_1, \dots, Y_n , $n > 1$, indipendenti con legge normale di media nota pari a zero e varianza $\sigma^2 > 0$ ignota, si reperisca uno stima di σ^2 e si indaghino le proprietà campionarie dello stimatore corrispondente.

Esercizio 21.13 (*Compito del 11/09/2017*) Dato un campione y_1, \dots, y_n , realizzazione di variabili casuali Y_1, \dots, Y_n , $n > 1$, indipendenti con legge esponenziale con tasso di guasto $\lambda > 0$ ignoto, si reperisca una stima di λ .

Lezione 22

Inferenza statistica: intervalli di confidenza (IC)

22.1 Dalla stima puntuale alla stima intervallare

Come le previsioni puntuali, anche le stime puntuali sono fatte per essere smentite. Di fatto, quando un parametro d'interesse $\tau = g(\theta)$ ha stima $\hat{\tau}_n = 82.39$ non ci si aspetta che il vero valore di τ sia 82.39. Per questo è opportuno accompagnare al valore di stima puntuale $\hat{\tau}_n$ una indicazione della precisione con cui la procedura di stima opera, calcolando una stima dell'errore medio di stima, lo *estimated standard error*, $\hat{se}(\hat{\tau}_n)$. La stima puntuale potrà così essere 'dilatata' per incorporare un adeguato margine d'errore.

L'incertezza con cui $\hat{\tau}_n$ stima τ è ben rappresentata da un intervallo di confidenza (più in generale sarà una regione di confidenza se τ non è scalare), quale ad esempio **l'intervallo classico**

$$\hat{\tau}_n \pm 1.96 \hat{se}(\hat{\tau}_n).$$

All'intervallo classico è associato un livello di confidenza nominale $1 - \alpha = 0.95$. Il livello di confidenza nominale è indicativo della fiducia da riporre nella affermazione 'il vero valore di τ è contenuto nell'intervallo ottenuto dai dati' quando lo stimatore $\hat{\tau}_n$ è asintoticamente normale e n è sufficientemente grande. Infatti per $Z \sim N(0, 1)$ si ha $P(-1.96 < Z < 1.96) \doteq 0.95$. Aumentando l'ampiezza dell'intervallo di confidenza, aumenterà il livello di confidenza. Ad esempio **l'intervallo prudente**

$$\hat{\tau}_n \pm 2.81 \hat{se}(\hat{\tau}_n).$$

ha livello di confidenza nominale $1 - \alpha \doteq 0.995$.

L'intuizione sottostante alla costruzione degli intervalli di confidenza è che la distribuzione campionaria di $\hat{\tau}_n$ sotto θ , indicata con $\mathcal{L}(\hat{\tau}_n(Y); \theta)$, sia stimata

da $\mathcal{L}(\hat{\tau}_n(Y); \hat{\theta})$. Quando $\hat{\theta}$ è vicino al vero valore di θ , $\mathcal{L}(\hat{\tau}_n(Y); \hat{\theta})$ descrive accuratamente come le stime di τ varino al variare dei campioni sorteggiabili. In particolare, se $\mathcal{L}(\hat{\tau}_n(Y); \theta) \dot{\sim} N(\tau, se_{\theta}(\hat{\tau}_n)^2)$ si avrà

$$\mathcal{L}(\hat{\tau}_n(Y); \hat{\theta}) \dot{\sim} N(\hat{\tau}, \hat{se}(\hat{\tau}_n)^2),$$

da cui si ricava il margine d'errore $\pm z_{1-\alpha/2} \hat{se}(\hat{\tau}_n)$ per il livello di confidenza nominale $1 - \alpha$. L'errore su una scala standard infatti ha legge approssimata $Z \sim N(0, 1)$.

22.2 Regioni di confidenza con livello di confidenza $1 - \alpha$

Sia

$$\mathcal{F} = \{p_Y(y; \theta), \quad y \in \mathcal{Y}, \quad \theta \in \Theta \subseteq \mathbb{R}^p\}$$

un modello statistico parametrico, sovente di c.c.s. con $p_Y(y; \theta) = \prod_{i=1}^n p_{Y_1}(y_i; \theta)$.

Definizione 22.1 Si dice **regione di confidenza per θ** una applicazione con dominio \mathcal{Y} che mappa $y \rightarrow \hat{\Theta}(y) \subseteq \Theta$.

Quando $p = 1$ e la regione risulta un intervallo si parlerà di intervallo di confidenza. Spesso sarà interessante seguire il comportamento di una regione di confidenza al crescere della numerosità campionaria n . Si userà allora la notazione $\hat{\Theta}_n(y)$.

Una proprietà campionaria desiderabile per una regione di confidenza è che facilmente contenga il vero valore del parametro. L'obiettivo si consegue richiedendo che la **probabilità di copertura** sotto θ

$$P_{\theta}(\theta \in \hat{\Theta}(Y))$$

sia elevata per ogni $\theta \in \Theta$, e quindi anche per il vero valore di θ . Un valore elevato di $P_{\theta}(\theta \in \hat{\Theta}(Y))$ garantisce infatti che per gran parte dei campioni sorteggiabili sotto θ valga che la regione di confidenza 'copra' il bersaglio θ . Si osservi che per sottolineare che il calcolo della probabilità di copertura è fatto sotto θ , si è usata la notazione esplicita $P_{\theta}(\theta \in \hat{\Theta}(Y))$ invece della consueta notazione implicita, che qui sarebbe $P(\theta \in \hat{\Theta}(Y))$.

Definizione 22.2 Si dice che $\hat{\Theta}(y)$ è una **regione di confidenza per θ con livello di confidenza $1 - \alpha$** se

$$P_{\theta}(\theta \in \hat{\Theta}(Y)) = 1 - \alpha, \quad \text{per ogni } \theta \in \Theta.$$

Nella definizione data il livello di confidenza è esatto. Di solito si otterranno regioni di confidenza con livello di confidenza $1 - \alpha$ approssimato, ad esempio in un senso asintotico:

$$\lim_{n \rightarrow \infty} P_{\theta}(\theta \in \hat{\Theta}_n(Y)) = 1 - \alpha, \quad \text{per ogni } \theta \in \Theta.$$

Il livello di confidenza $1 - \alpha$ è fissato dall'analista dei dati. La scelta classica è $1 - \alpha = 0.95$. Una scelta prudente è $1 - \alpha = 0.995$.

Osservato y , l'inferenza su θ viene fatta riferendo la regione $\hat{\Theta}(y)$ e il livello di confidenza $1 - \alpha$.

22.3 Campionamento da normale con varianza nota: intervalli di confidenza per μ

Si consideri l'esperimento di c.c.s. con numerosità n da una popolazione normale, $N(\mu, \sigma_0^2)$, con $\mu \in \mathbb{R}$ ignoto e $\sigma_0^2 > 0$ noto. Si osserva dunque $Y = (Y_1, \dots, Y_n)$ con componenti Y_i che sono i.i.d. e hanno legge marginale $N(\mu, \sigma^2)$.

L'informazione sull'ignoto μ è raccolta dallo stimatore non distorto e consistente

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \sim N\left(\mu, \frac{\sigma_0^2}{n}\right).$$

Lo *standard error* dello stimatore qui è noto e pari a σ_0/\sqrt{n} .

Per costruire intervalli di confidenza per μ con livello di confidenza $1 - \alpha$ conviene considerare che la media campionaria standardizzata

$$Z = \frac{\bar{Y}_n - \mu}{\sigma_0/\sqrt{n}}$$

ha legge $N(0, 1)$, e che $Z \sim -Z$ per la simmetria attorno all'origine della f.d.p. di Z . Si ha allora

$$\begin{aligned} 1 - \alpha &= P(-z_{1-\alpha/2} \leq -Z \leq z_{1-\alpha/2}) \\ &= P\left(-z_{1-\alpha/2} \leq \frac{\mu - \bar{Y}_n}{\sigma_0/\sqrt{n}} \leq z_{1-\alpha/2}\right) \\ &= P\left(\bar{Y}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2} \leq \mu \leq \bar{Y}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}\right). \end{aligned}$$

Disponendo della realizzazione di Y , $y = (y_1, \dots, y_n)$, si riferisce che

$$\bar{y}_n \pm \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}$$

è un intervallo di confidenza per μ con livello di confidenza esatto $1 - \alpha$. Quando si sceglie $1 - \alpha = 0.95$ si ha $z_{0.975} \doteq 1.96$ e si ritrova un caso particolare dell'intervallo classico.

Esempio 22.1 Si supponga che molte misurazioni fatte in precedenza con uno strumento di misura affetto da errore facciano ritenere che lo scarto quadratico medio delle misure fatte con quello strumento sia $\sigma_0 = 2.5$. Un campione di $n = 9$ misure di un'ignota quantità μ è

$$y = (8.6, 9.1, 11.7, 14.2, 12.9, 9.0, 8.9, 12.3, 9.6).$$

Si ha $\bar{y}_9 = 10.7$. L'intervallo di confidenza per μ con livello di confidenza $1 - \alpha = 0.95$ è $\bar{y}_9 \pm (\sigma_0/\sqrt{9})z_{0.975} = 10.7 \pm (2.5/3) * 1.96 = [9.06, 12.3]$. Si ha fiducia che l'intervallo ottenuto copra μ , perché ciò accade nel 95% dei campioni sorteggiabili. Si noti però che sarebbe errato, dato il campione effettivamente osservato, affermare che $\mu \in [9.06, 12.3]$ con probabilità 0.95. In effetti μ è un valore ignoto ma fissato, sta o non sta nell'intervallo calcolato con i dati. \diamond

22.4 Campionamento da normale con valore atteso noto: IC per σ^2

Si consideri l'esperimento di c.c.s. con numerosità n da una popolazione normale, $N(\mu_0, \sigma^2)$, con $\mu_0 \in \mathbb{R}$ noto e $\sigma^2 > 0$ ignoto, descritto dalla v.c. $Y = (Y_1, \dots, Y_n)$.

L'informazione sull'ignoto σ^2 è raccolta dallo stimatore non distorto e consistente

$$\hat{\sigma}_{\mu_0}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu_0)^2 \sim \frac{\sigma^2}{n} \chi_n^2.$$

Quindi $(n/\sigma^2)\hat{\sigma}_{\mu_0}^2 \sim \chi_n^2$ e indicando con $\chi_{\nu,p}^2$ il quantile- p di una v.c. con legge χ_ν^2 si ha

$$\begin{aligned} 1 - \alpha &= P\left(\chi_{n,\alpha/2}^2 \leq \frac{n}{\sigma^2} \hat{\sigma}_{\mu_0}^2 \leq \chi_{n,1-\alpha/2}^2\right) \\ &= P\left(\frac{1}{\chi_{n,1-\alpha/2}^2} \leq \frac{\sigma^2}{n} \frac{1}{\hat{\sigma}_{\mu_0}^2} \leq \frac{1}{\chi_{n,\alpha/2}^2}\right) \\ &= P\left(\frac{n\hat{\sigma}_{\mu_0}^2}{\chi_{n,1-\alpha/2}^2} \leq \sigma^2 \leq \frac{n\hat{\sigma}_{\mu_0}^2}{\chi_{n,\alpha/2}^2}\right) \end{aligned}$$

per cui, disponendo della realizzazione $y = (y_1, \dots, y_n)$ di Y , da cui si calcola la statistica $n\hat{\sigma}_{\mu_0}^2 = \sum_{i=1}^n (y_i - \mu_0)^2$, un intervallo di confidenza per σ^2 con livello di confidenza esatto $1 - \alpha$ è

$$\left[\frac{n\hat{\sigma}_{\mu_0}^2}{\chi_{n,1-\alpha/2}^2}, \frac{n\hat{\sigma}_{\mu_0}^2}{\chi_{n,\alpha/2}^2} \right] = \left[\frac{\sum_{i=1}^n (y_i - \mu_0)^2}{\chi_{n,1-\alpha/2}^2}, \frac{\sum_{i=1}^n (y_i - \mu_0)^2}{\chi_{n,\alpha/2}^2} \right].$$

22.5 Campionamento da normale con valore atteso ignoto: intervalli di confidenza per σ^2

Si consideri l'esperimento di c.c.s. con numerosità n da una popolazione normale, $N(\mu, \sigma^2)$, con $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ entrambi ignoti, descritto dalla v.c. $Y = (Y_1, \dots, Y_n)$.

L'informazione sull'ignoto σ^2 è raccolta dallo stimatore non distorto e consistente

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \sim \frac{\sigma^2}{n-1} \chi_{n-1}^2.$$

Si ha dunque che $((n-1)/\sigma^2)S_n^2 \sim \chi_{n-1}^2$ e ripetendo i calcoli fatti nel precedente paragrafo si ottiene che

$$\left[\frac{(n-1)S_n^2}{\chi_{n-1,1-\alpha/2}^2}, \frac{(n-1)S_n^2}{\chi_{n-1,\alpha/2}^2} \right] = \left[\frac{\sum_{i=1}^n (Y_i - \hat{Y}_n)^2}{\chi_{n-1,1-\alpha/2}^2}, \frac{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}{\chi_{n-1,\alpha/2}^2} \right]$$

è un intervallo aleatorio che ‘copre’ l'ignoto σ^2 , qualunque esso sia, con probabilità $1 - \alpha$.

Disponendo della realizzazione $y = (y_1, \dots, y_n)$ di Y , da cui si calcola la statistica $\sum_{i=1}^n (y_i - \bar{y}_n)^2$, un intervallo di confidenza per σ^2 con livello di confidenza esatto $1 - \alpha$ è allora

$$\left[\frac{\sum_{i=1}^n (y_i - \bar{y}_n)^2}{\chi_{n-1,1-\alpha/2}^2}, \frac{\sum_{i=1}^n (y_i - \bar{y}_n)^2}{\chi_{n-1,\alpha/2}^2} \right].$$

22.6 Campionamento da normale con varianza ignota: intervalli di confidenza per μ

Si consideri l'esperimento di c.c.s. con numerosità n da una popolazione normale, $N(\mu, \sigma^2)$, con $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ entrambi ignoti. Si osserva dunque $Y = (Y_1, \dots, Y_n)$ con componenti Y_i i.i.d. e legge marginale $N(\mu, \sigma^2)$.

L'informazione sull'ignoto μ è raccolta dallo stimatore non distorto e consistente

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

L'informazione sull'ignoto σ^2 è raccolta dallo stimatore non distorto e consistente

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \sim \frac{\sigma^2}{n-1} \chi_{n-1}^2.$$

Le v.c. \bar{Y}_n e S_n^2 sono indipendenti.

Per costruire intervalli di confidenza per μ con livello di confidenza $1 - \alpha$ non si può standardizzare la media campionaria perché lo *standard error* σ/\sqrt{n} è ignoto. Si passa allora allo *estimated standard error* S_n/\sqrt{n} e si opera la ‘studentizzazione’ della media campionaria. La **media campionaria studentizzata** è

$$T_n = \frac{\bar{Y}_n - \mu}{S_n/\sqrt{n}}.$$

La legge di T_n è t di Student con $n-1$ gradi di libertà. Per la simmetria attorno all'origine della distribuzione t di Student si ha $T_n \sim -T_n$. Pertanto, indicato con $t_{n-1,p}$ il quantile- p della t di Student con $n-1$ gradi di libertà, si ha

$$\begin{aligned} 1 - \alpha &= P(-t_{n-1,1-\alpha/2} \leq -T_n \leq t_{n-1,1-\alpha/2}) \\ &= P\left(-t_{n-1,1-\alpha/2} \leq \frac{\mu - \bar{Y}_n}{S_n/\sqrt{n}} \leq t_{n-1,1-\alpha/2}\right) \\ &= P\left(\bar{Y}_n - \frac{S_n}{\sqrt{n}} t_{n-1,1-\alpha/2} \leq \mu \leq \bar{Y}_n + \frac{S_n}{\sqrt{n}} t_{n-1,1-\alpha/2}\right). \end{aligned}$$

Disponendo della realizzazione $y = (y_1, \dots, y_n)$ di Y , da cui si calcolano le statistiche $\bar{y} = \sum_{i=1}^n y_i/n$ e $s_n^2 = \sum_{i=1}^n (y_i - \bar{y}_n)^2/(n-1)$, un intervallo di confidenza per μ con livello di confidenza esatto $1 - \alpha$ è dunque

$$\bar{y}_n \pm \frac{s_n}{\sqrt{n}} t_{n-1,1-\alpha/2}.$$

Poiché $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$ per ogni $\alpha \in (0, 1)$, l'intervallo di confidenza con σ^2 ignoto tende ad essere più ampio di quello con σ^2 noto. Quando, con n sufficientemente grande (orientativamente, maggiore di 30), si sceglie $1 - \alpha = 0.95$ si ha $t_{n-1,0.975} \doteq z_{0.975} \doteq 1.96$ e si ritrova un caso particolare dell'intervallo classico.

22.7 Campionamento da legge con varianza finita: IC per il valore atteso

Si consideri l'esperimento di c.c.s. con numerosità n da una v.c. Y_1 con $E(Y_1) = \mu \in \mathbb{R}$ e $\text{Var}(Y_1) = \sigma^2 > 0$ entrambi ignoti. Si osserva dunque $Y = (Y_1, \dots, Y_n)$ con componenti Y_i i.i.d. da una popolazione non normale.

Si desidera determinare un intervallo di confidenza per μ con livello di confidenza circa pari a $1 - \alpha$ quando n è sufficientemente grande. L'informazione sull'ignoto μ è raccolta dallo stimatore non distorto e consistente

$$\hat{\mu}_n = \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \sim \mathcal{L}_n\left(\mu, \frac{\sigma^2}{n}\right),$$

il cui *standard error* σ/\sqrt{n} è ignoto e va stimato. Informativo su σ^2 , l'ignota varianza di Y_1 , è in generale lo stimatore non distorto e consistente

$$\hat{\sigma}_n^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2.$$

Tuttavia nei modelli statistici monparametrici, quali quelli di c.c.s. da Poisson o binomiale elementare, la varianza è una funzione della media, $\sigma^2 = \sigma^2(\mu)$, per

cui sarà più conveniente stimare σ^2 con $\hat{\sigma}_n^2 = \sigma^2(\hat{\mu}_n)$ (metodo di sostituzione). Quindi lo *estimated standard error* di $\hat{\mu}_n$, $\hat{\sigma}_n/\sqrt{n}$, è in generale s_n/\sqrt{n} , ma nei modelli monoparametrici appena richiamati è $\sigma(\hat{\mu}_n)/\sqrt{n}$.

Per il teorema centrale del limite quando n è sufficientemente grande

$$\hat{\mu}_n = \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

ossia

$$\frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) .$$

Sostituendo l'ignoto *standard error* σ/\sqrt{n} con l'*estimated standard error* $\hat{\sigma}_n/\sqrt{n}$ si ha ancora, per la consistenza di $\hat{\sigma}_n$,

$$\frac{\bar{Y}_n - \mu}{\hat{\sigma}_n/\sqrt{n}} \sim N(0, 1) .$$

Pertanto, come nel campionamento da popolazione normale quando n è sufficientemente grande,

$$\bar{y}_n \pm \frac{\hat{\sigma}_n}{\sqrt{n}} z_{1-\alpha/2}$$

è l'intervallo di confidenza per μ cercato.

In particolare, però,

- quando $y = (y_1, \dots, y_n)$ è un c.c.s. da una $Bi(1, p)$, $p \in (0, 1)$ ignoto, la stima $\hat{p}_n = \sum_{i=1}^n y_i/n$ ha *estimated standard error* $\sqrt{\hat{p}_n(1 - \hat{p}_n)/n}$ per cui, per n sufficientemente grande, l'intervallo di confidenza per p con livello approssimato $1 - \alpha$ è

$$\hat{p}_n \pm \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} z_{1-\alpha/2} ;$$

- quando $y = (y_1, \dots, y_n)$ è un c.c.s. da una $P(\lambda)$, $\lambda > 0$ ignoto, la stima $\hat{\lambda}_n = \sum_{i=1}^n y_i/n$ ha *estimated standard error* $\sqrt{\hat{\lambda}_n/n}$ per cui, per n sufficientemente grande, l'intervallo di confidenza per λ con livello approssimato $1 - \alpha$ è

$$\hat{\lambda}_n \pm \sqrt{\frac{\hat{\lambda}_n}{n}} z_{1-\alpha/2} .$$

22.8 Esercizi

Esercizio 22.1 Dato un campione costituito da variabili casuali Y_1, \dots, Y_n indipendenti con legge di Bernoulli $Bi(1, p)$, $p \in (0, 1)$ ignoto, si reperisca un intervallo di confidenza per p con livello approssimato $1 - \alpha$.

Esercizio 22.2 Si vuole stimare la porzione $p \in (0, 1)$ di telespettatori che seguono una certa serie televisiva. Si acquisisce un campione casuale di 300 telespettatori e si rileva che 69 dichiarano di seguire la serie televisiva. Si definisca un modello statistico adatto a descrivere l'esperimento in esame. Si reperisca un opportuno stimatore di p , si determini l'associato *standard error* e si calcolino i corrispondenti valori di stima. Infine, con i risultati dell'indagine campionaria, si calcoli un intervallo di confidenza per p di livello approssimato $1 - \alpha = 0.95$.

Esercizio 22.3 Dato un campione casuale costituito da un numero $n \in \mathbb{N}^+$, sufficientemente elevato, di variabili casuali Y_1, \dots, Y_n , indipendenti con legge di Poisson avente parametro $\lambda > 0$ ignoto, si definisca un intervallo di confidenza per λ con livello approssimato $1 - \alpha$.

Esercizio 22.4 Dato un campione costituito da variabili casuali Y_1, \dots, Y_n indipendenti con legge marginale normale avente media ignota μ e varianza nota $\sigma_0^2 > 0$, si definisca un intervallo di confidenza per μ con livello $1 - \alpha$.

Esercizio 22.5 Dato un campione costituito da variabili casuali indipendenti Y_1, \dots, Y_n , $n > 1$, tutte con distribuzione di probabilità normale con media μ e varianza σ^2 , entrambe ignote, si definisca un intervallo di confidenza per μ con livello $1 - \alpha$.

Esercizio 22.6 Un'azienda manifatturiera vuole studiare la resistenza media alla rottura di un determinato materiale isolante. A tale scopo esegue 30 prove di resistenza alla rottura valutando il peso in *kg* necessario a rompere l'isolante. L'esperimento ha fornito i seguenti risultati: $\sum_{i=1}^{30} y_i = 51\,702$ *kg* e $\sum_{i=1}^{30} y_i^2 = 89\,335\,770$ *kg*². Nell'ipotesi che il modello statistico di riferimento sia normale, si determini una stima per la resistenza media μ , con l'associato *estimated standard error*. Inoltre, si fornisca un intervallo di confidenza per μ con livello $1 - \alpha = 0.95$.

Esercizio 22.7 Su 532 visite a un sito, si sono registrati 79 click su un *banner* pubblicitario. Si calcoli un intervallo di confidenza per la probabilità che un visitatore clicchi il *banner*, ad un livello di confidenza approssimato 0.995.

Esercizio 22.8 Il numero settimanale di reclami che arrivano allo *help desk* di un'azienda segue una distribuzione di Poisson con parametro λ . Nelle ultime 12 settimane le chiamate osservate sono state $y = (2, 3, 5, 4, 8, 4, 8, 2, 9, 6, 5, 6)$. Si calcoli un intervallo di confidenza per λ con livello di confidenza approssimato 0.95. Si dia una stima puntuale e un intervallo di confidenza con livello di confidenza approssimato 0.95 per la probabilità di non ricevere reclami in una settimana.

Esercizio 22.9 Sia $y = (26.26, 34.5, 27.69, 32.4, 19.78, 23.02, 21.57, 16.49, 32.61)$ realizzazione di un esperimento di c.c.s. con numerosità $n = 9$ da $N(\mu, 25)$, dove $\mu \in \mathbb{R}$ è ignoto. Si calcoli un intervallo di confidenza per μ con livello di confidenza 0.995.

Esercizio 22.10 Sia $y = (12.14, 5.90, 10.47, 14.39, 5.25, 6.50, 15.92, 0.92, 15.51)$ realizzazione di un esperimento di c.c.s. con numerosità $n = 9$ da $N(10, \sigma^2)$, dove $\sigma^2 > 0$ è ignoto. Si calcoli un intervallo di confidenza per σ^2 con livello di confidenza 0.95.

Esercizio 22.11 Sia $y = (5.9, 10.2, 8.7, 14.9, 1.8, 14.1, 9.6, 5.1, 13.3)$ realizzazione di un esperimento di c.c.s. con numerosità 9 da $N(\mu, \sigma^2)$, dove $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ sono ignoti. Si calcoli un intervallo di confidenza per μ con livello di confidenza 0.95.

Esercizio 22.12 Sia $y = (5.9, 10.2, 8.7, 14.9, 1.8, 14.1, 9.6, 5.1, 13.3)$ realizzazione di un esperimento di c.c.s. con numerosità 9 da $N(\mu, \sigma^2)$, dove $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ sono ignoti. Si calcoli un intervallo di confidenza per σ^2 con livello di confidenza 0.95.

Lezione 23

I test statistici

23.1 I dati e l'ipotesi

Affianca le due tipologie di procedure di inferenza statistica già esaminate, ossia l'ottenimento di stime puntuali e la costruzione di regioni di confidenza, una terza tipologia di procedure, che riguarda la verifica di ipotesi statistiche. Prima di considerare i dettagli tecnici, conviene avere una visione d'insieme dei concetti in gioco.

Nell'inferenza i dati y sono visti come una realizzazione di una v.c. Y con legge almeno in parte ignota. È allora naturale che il ricercatore formuli congetture su ciò che è ignoto della legge di Y . Una particolare congettura sulla legge di Y viene detta **ipotesi statistica**. Si presuppone che la congettura sia formulata prima di disporre dei dati, o comunque indipendentemente dai dati. I dati sono acquisiti proprio per gettare luce sulla congettura in modo imparziale rispetto alle preferenze del ricercatore.

Per ricavare qualche indicazione empirica sulla validità della congettura si effettua un **test statistico**. Si calcola dai dati una opportuna statistica, detta statistica test, e si paragona il suo valore osservato ai valori usualmente assunti dalla statistica quando l'ipotesi sotto indagine è vera.

Come esempio guida, si supponga che la congettura su Y comporti che la distribuzione campionaria di una certa statistica $z(y)$ sia $Z = z(Y) \sim N(0, 1)$. Sia $z = z(y)$ il valore osservato di Z . Evidentemente, sono in accordo con la congettura non solo $z = 0$ ma anche tutti i valori di z non lontani da 0. Un valore osservato $z = 4.18$ è invece in marcato disaccordo con la congettura, specialmente se, quando la congettura non vale, la statistica z assume valori usualmente più grandi di quelli predetti da $N(0, 1)$.

Vi sono due generi di test statistici. Infatti, a un test statistico in alcuni casi si richiede una indicazione netta, se i dati sono o no conformi all'ipotesi (test decisionale). In altri casi invece si richiede una indicazione più sfumata, una misura di quanto i dati sono conformi all'ipotesi (test non decisionale).

Continuando l'esempio guida, se la statistica z assume valori grandi quando la congettura non vale, un test decisionale è:

- la statistica test $z = z(y)$ dà indicazione che i dati sono conformi all'ipotesi se si osserva $z \leq 2.33$;
- dà invece indicazione che i dati non sono conformi all'ipotesi se si osserva $z > 2.33$.

Un valore osservato $z = 4.18$ porta dunque al rifiuto dell'ipotesi. Seguendo la regola di condotta sopra enunciata, raramente si prenderà una decisione errata quando vale l'ipotesi. Infatti, quando $Z \sim N(0, 1)$, $P(Z > 2.33) = 1 - \Phi(2.33) \doteq 0.01$. Affinché i dati possano gettare luce sulla congettura in modo indipendente dalle preferenze del ricercatore, si presuppone che la regola di decisione sia formulata prima di disporre dei dati, o comunque indipendentemente dai dati,

Un test non decisionale conduce al calcolo di un valore di probabilità o *P-value*, indicato con P . Il valore P rappresenta quanto è probabile osservare, valendo l'ipotesi, dati ancora più in disaccordo con l'ipotesi di quelli effettivamente osservati. Quindi, nell'esempio guida,

$$P = P(Z > z) = 1 - \Phi(z).$$

Più piccolo è il *P-value*, minore è l'accordo fra dati ed ipotesi, e dunque più i dati danno indicazione contraria alla validità dell'ipotesi.

Un valore osservato $z = 4.18$ conduce al *P-value*

$$P = 1 - \Phi(4.18) \doteq 1.5 \times 10^{-5},$$

indicazione che i dati costituiscono fortissima testimonianza contro l'ipotesi.

23.2 Ipotesi statistiche, nulla e alternativa

Come si è anticipato, una ipotesi statistica è una congettura sulla legge della v.c. Y di cui i dati y sono una realizzazione. Quindi una congettura sul modello probabilistico generatore dei dati.

L'ipotesi è detta **semplice** se propone un solo modello probabilistico come possibile generatore dei dati. L'ipotesi è detta **composita** se propone più di un modello probabilistico come possibile generatore dei dati.

Spesso l'ipotesi è formulata nell'ambito di un modello statistico parametrico

$$\mathcal{F} = \{p_Y(y; \theta), \quad y \in \mathcal{Y}, \quad \theta \in \Theta \subseteq \mathbb{R}^p\}$$

come **ipotesi nulla**

$$H_0 : \theta \in \Theta_0 \subset \Theta.$$

Un'ipotesi nulla esprime una congettura plausibile di semplificazione del modello, per esempio la mancanza di un certo effetto. È detta nulla perché sovente

attribuisce valore zero a uno o più parametri. In generale H_0 afferma che il vero valore del parametro θ è un punto della parte Θ_0 di Θ . Come si è sottolineato, la parte Θ_0 deve essere indicata prima di acquisire i dati. Questo requisito distingue nettamente Θ_0 da una regione di confidenza, che invece è funzione dei dati.

Ad esempio con i dati $y = (y_1, \dots, y_n)$ dove $y_i \in \{0, 1\}$ e modello di c.c.s. con numerosità n da $Bi(1, p)$, può essere interessante indagare sull'ipotesi nulla semplice

$$H_0 : p = \frac{1}{2}.$$

Nell'esperimento che consiste in n lanci indipendenti di una moneta, H_0 esprime l'assenza di squilibrio della moneta, un'ipotesi molto naturale da formulare ancor prima di vedere in azione la moneta.

In contrapposizione all'ipotesi nulla, il modello statistico \mathcal{F} individua una **ipotesi alternativa**, negazione di H_0 nel modello statistico \mathcal{F} ,

$$H_1 : \theta \in \Theta \setminus \Theta_0.$$

Quindi H_1 afferma che il vero valore del parametro θ è un punto di $\Theta \setminus \Theta_0$, il complementare di Θ_0 .

In un modello statistico monoparametrico, con $\Theta = \mathbb{R}$ oppure Θ un intervallo (eventualmente illimitato), un'ipotesi nulla semplice, $H_0 : \theta = \theta_0$, può essere contrapposta a tre tipi di alternative:

- **alternativa bilaterale** $H_1 : \theta \neq \theta_0$
- **alternativa unilaterale destra** $H_1 : \theta > \theta_0$
- **alternativa unilaterale sinistra** $H_1 : \theta < \theta_0$.

Quando l'alternativa è bilaterale, Θ è o \mathbb{R} oppure un intervallo di cui θ_0 è un punto interno. Quando l'alternativa è unilaterale destra, $\Theta = [\theta_0, +\infty)$ o un intervallo di cui θ_0 è il minimo. Quando l'alternativa è unilaterale sinistra, $\Theta = (-\infty, \theta_0]$ o un intervallo di cui θ_0 è il massimo.

23.3 Test decisionali con livello di significatività

α

In certe situazioni un problema di verifica d'ipotesi viene formulato per poter prendere una decisione che tenga conto dell'informazione portata dai dati stessi. Se alla luce dei dati l'ipotesi nulla sembra valere, la decisione conveniente potrebbe essere ad esempio non commercializzare un prodotto, o non risettare una macchina. Se invece l'ipotesi nulla non sembra valere, la decisione conveniente potrebbe essere commercializzare un prodotto, o risettare una macchina.

Per un problema di verifica d'ipotesi su un parametro scalare con ipotesi nulla semplice ed alternativa, ad esempio, bilaterale,

$$H_0 : \theta = \theta_0 \quad \text{contro} \quad H_1 : \theta \neq \theta_0 ,$$

si consideri una statistica, detta **statistica test**, $t_n = t_n(y)$, che conduce a decidere di rifiutare H_0 per certi valori di y , detti **critici**, mentre conduce ad accettare H_0 per i valori complementari in \mathcal{Y} . Si può convenire che il test decisionale abbia regola di decisione

$$d_n(y) = \begin{cases} 0 & \text{se } t_n \text{ indica di accettare } H_0 \\ 1 & \text{se } t_n \text{ indica di rifiutare } H_0 . \end{cases}$$

Le regioni

$$\begin{aligned} A &= \{y \in \mathcal{Y} : d_n(y) = 0\} \\ R &= \{y \in \mathcal{Y} : d_n(y) = 1\} \end{aligned}$$

sono dette, rispettivamente, **regione di accettazione** e **regione di rifiuto**. Vale

$$A \cap R = \emptyset, \quad A \cup R = \mathcal{Y} .$$

Il test $t_n = t_n(y)$ con regione di rifiuto R è detto con **livello di significatività** α se

$$P_{\theta_0}(Y \in R) = \alpha ,$$

esattamente o, sovente, come approssimazione asintotica.

Quando sotto H_0 , quindi sotto $\theta = \theta_0$, si realizza $y \in R$, il test induce a commettere un errore, rifiutare H_0 quando H_0 è vera, detto **errore del primo tipo**. Quindi α è la probabilità di commettere un errore del primo tipo.

Il test può inoltre indurre a commettere un **errore del secondo tipo**, accettare H_0 quando H_0 è falsa. Questo avviene quando, sotto $\theta \neq \theta_0$, si realizza $y \in A$. L'errore di secondo tipo ha probabilità

$$\beta = \beta(\theta) = P_\theta(Y \in A) , \quad \theta \neq \theta_0 .$$

Idealmente, si vorrebbe reperire un test per il quale le probabilità dei due tipi d'errore, α e β , siano entrambe piccole. Tuttavia, data la numerosità campionaria n , se si riduce R , e di conseguenza si diminuisce α , inevitabilmente si espande A , e di conseguenza si aumenta β . Si ricorre dunque a un compromesso. Si fissa un livello di significatività α convenientemente piccolo, quale $\alpha = 0.05$ o $\alpha = 0.005$, e si cerca un test con livello di significatività α che induca probabilità di commettere errore del secondo tipo piccole, per quanto è possibile.

Il compromesso permette di controllare direttamente la probabilità d'errore del primo tipo. La probabilità d'errore del secondo tipo è invece subita, e può essere vicina a $1 - \alpha$ se θ è vicino a θ_0 e la funzione $P_\theta(Y \in A)$ è continua in un intorno di θ_0 .

Test per $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$ con livello di significatività α si possono ottenere a partire da regioni di confidenza $\hat{\Theta}_n(y)$ con livello di confidenza $1 - \alpha$. La regola di costruzione è

$$\begin{aligned} d_n(y) = 0, \text{ ossia si accetta } H_0, & \iff \hat{\Theta}_n(y) \text{ copre } \theta_0 \\ d_n(y) = 1, \text{ ossia si rifiuta } H_0, & \iff \hat{\Theta}_n(y) \text{ non copre } \theta_0. \end{aligned}$$

Esempio 23.1 Test di $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$ nel campionamento casuale semplice con numerosità n da $N(\mu, \sigma_0^2)$, $\sigma_0^2 > 0$ noto.

L'intervallo di confidenza per μ con livello di confidenza $1 - \alpha$ è

$$\bar{y}_n \pm \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}.$$

Il test corrispondente è

$$\begin{aligned} \text{se } \mu_0 \in \left[\bar{y}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}, \bar{y}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2} \right] & \text{ allora } d_n(y) = 0, \text{ ossia si accetta } H_0 \\ \text{se } \mu_0 \notin \left[\bar{y}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2}, \bar{y}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2} \right] & \text{ allora } d_n(y) = 1, \text{ ossia si rifiuta } H_0. \end{aligned}$$

Il livello di significatività del test è

$$\begin{aligned} P_{\mu_0}(d_n(Y) = 1) &= 1 - P_{\mu_0}(d_n(Y) = 0) \\ &= 1 - P_{\mu_0} \left(\bar{Y}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2} \leq \mu_0 \leq \bar{Y}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha/2} \right) \\ &= 1 - (1 - \alpha) = \alpha. \end{aligned}$$

Equivalentemente, si può considerare che la statistica test media campionaria standardizzata sotto H_0

$$Z = \frac{\bar{Y}_n - \mu_0}{\sigma_0/\sqrt{n}}$$

ha, se vale H_0 , distribuzione $N(0, 1)$ e sono critici contro H_0 i valori di Z lontani da 0 su entrambe le code. Quindi, considerando in corrispondenza al livello di significatività α i punti critici $\pm z_{1-\alpha/2}$, la regola di decisione espressa tramite la statistica test Z è

$$\begin{aligned} \text{se } \left| \frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}} \right| &\geq z_{1-\alpha/2} \quad \text{allora } d_n(y) = 1 \\ \text{se } \left| \frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}} \right| &< z_{1-\alpha/2} \quad \text{allora } d_n(y) = 0. \end{aligned}$$

◇

La statistica test media campionaria standardizzata sotto H_0 risulta utile anche quando l'ipotesi alternativa è direzionale. Solo la coda di Z nella direzione prevista dall'alternativa sarà però critica contro H_0 . Si spenderà quindi tutta la probabilità α di commettere errore del primo tipo sul lato previsto

dall'alternativa. Di conseguenza, nelle assunzioni dell'Esempio 23.1, per il problema di verifica d'ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu > \mu_0$, il test con livello di significatività α da utilizzare è

$$\begin{aligned} \text{se } \frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}} &\geq z_{1-\alpha} & \text{allora } d_n(y) &= 1 \\ \text{se } \frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}} &< z_{1-\alpha} & \text{allora } d_n(y) &= 0. \end{aligned} \quad (23.1)$$

Per un problema di verifica d'ipotesi su un parametro p -dimensionale,

$$H_0 : \theta \in \Theta_0 \quad \text{contro} \quad H_1 : \theta \in \Theta \setminus \Theta_0,$$

un test decisionale con regione di accettazione A e regione di rifiuto R ha livello di significatività α se

$$\sup_{\theta \in \Theta_0} P_\theta(Y \in R) = \alpha,$$

esattamente o come approssimazione asintotica. Quindi in generale per un test con livello di significatività α la più grande probabilità d'errore del primo tipo è α o circa α .

23.4 Test non decisionali: il P -value

Un test statistico decisionale è una procedura che valuta se vi è o no ragionevole accordo fra i dati e una preassegnata ipotesi, H_0 . Se non vi è l'urgenza di prendere una decisione, il grado di accordo fra dati e ipotesi può essere espresso, anziché in termini dicotomici (accettazione/rifiuto di H_0), su una scala con molte sfumature intermedie. Lo strumento di valutazione usato a tale scopo è il calcolo di un P -value.

Come dice il nome, un P -value è un valore di probabilità, precisamente la probabilità che valendo H_0 si osservino valori almeno altrettanto critici contro H_0 di quelli effettivamente osservati.

Se i dati y sono perfettamente compatibili con H_0 (il grado di accordo è elevato), il P -value varrà 1 o un valore prossimo a 1: sotto H_0 , quasi tutti i valori osservabili sono più critici contro H_0 di quello effettivamente osservato. Se i dati y sono poco compatibili con H_0 (il grado di accordo è quasi nullo), il P -value varrà 0 o un valore molto prossimo a 0: sotto H_0 , valori ancora più critici contro H_0 di quello osservato sono inesistenti o sporadici.

Per calcolare un P -value si suppone che, data l'ipotesi nulla

$$H_0 : \theta \in \Theta_0,$$

per ogni dato osservabile $y \in \mathcal{Y}$ si sia in grado di costruire l'insieme E_y dei valori y^* almeno altrettanto critici contro H_0 di y , o più difficilmente compatibili con

H_0 di y . Allora il P -value associato ai dati y è

$$P = \sup_{\theta \in \Theta_0} P_{\theta}(Y \in E_y).$$

Se H_0 è semplice, di forma $\theta = \theta_0$, si ha $P = P_{\theta_0}(Y \in E_y)$. Se H_0 è composta, il sup fa valutare il P -value nel caso più favorevole ad H_0 . In entrambe le circostanze, un valore piccolo di P è interpretato come evidenza contro H_0 , evidenza tanto più forte quanto più P è vicino a zero.

Esempio 23.2 P -value per $H_0 : p = 0.5$ nel campionamento casuale semplice con numerosità $n = 100$ da $Bi(1, p)$, $p \in (0, 1)$.

Si lancia 100 volte una moneta e si prende nota se l'esito è testa ($y_i = 1$) o croce ($y_i = 0$). Lo spazio campionario è allora

$$\mathcal{Y} = \{0, 1\}^{100},$$

e il modello statistico è $Y = (Y_1, \dots, Y_{100})$ con Y_i i.i.d. $Bi(1, p)$, $p \in (0, 1)$ ignoto. Si considera come ipotesi nulla la congettura che la moneta sia equilibrata, quindi

$$H_0 : p = 0.5.$$

Come si è visto, la statistica informativa su p è $s = s(y) = \sum_{i=1}^{100} y_i$. L'ipotesi nulla prevede che $E(\sum_{i=1}^{100} Y_i) = 50$. Sarà però facilmente $\sum_{i=1}^{100} y_i$ diverso da 50. Più s è lontano da 50, in entrambe le direzioni, più vi è nei dati evidenza contro H_0 . Quindi si ha

$$E_y = \{y^* \in \mathcal{Y} : |s(y^*) - 50| \geq |s(y) - 50|\}.$$

Con $s(y) = 62$ si ottiene

$$\begin{aligned} P &= P_{H_0}(Y \in E_y) \\ &= P_{H_0}(|s(Y) - 50| \geq 12) \\ &= P_{H_0}(s(Y) \leq 38) + P_{H_0}(s(Y) \geq 62) \\ &= 2P_{H_0}(s(Y) \leq 38) \\ &\doteq 2 * 0.01049 = 0.02098 \end{aligned}$$

perché sotto H_0 si ha $s(Y) \sim Bi(100, 0.5)$. Con l'approssimazione normale, $Bi(100, 0.5) \sim N(50, 25)$, si otterrebbe $P \doteq 0.01639$. Vi è dunque nei dati qualche evidenza contro H_0 . \diamond

Come nell'esempio, il P -value è di solito calcolato considerando una **statistica test** scalare, $t = t(y)$, e un insieme di valori della statistica test almeno altrettanto estremo di $t = t(y)$, E_t . Usando le leggi sotto H_0 di $T = t(Y)$, si ha

$$P = \sup_{\theta \in \Theta_0} P_{\theta}(T \in E_t).$$

Se solo valori grandi di t sono evidenza contro H_0 sarà

$$E_t = \{\tilde{t} \in \mathcal{T} : \tilde{t} \geq t\}.$$

Se solo valori piccoli di t sono evidenza contro H_0 sarà

$$E_t = \{\tilde{t} \in \mathcal{T} : \tilde{t} \leq t\}.$$

Se valori sia grandi sia piccoli di t sono evidenza contro H_0 sarà

$$E_t = \{\tilde{t} \in \mathcal{T} : |\tilde{t}| \geq |t|\}.$$

23.5 Interpretazione del P -value

Calcolato un P -value, sia esso P , si possono trarre le conclusioni dell'inferenza statistica secondo lo schema orientativo:

- se $P \leq 0.005$, allora i dati sono difficilmente compatibili con H_0 : in sostanza si rifiuta H_0 ;
- se $0.005 < P \leq 0.05$, allora i dati sono poco compatibili con H_0 : si guarda H_0 con sospetto, forse è meglio acquisire nuovi dati;
- se $0.05 < P \leq 0.10$, allora i dati sono sufficientemente compatibili con H_0 : c'è però incertezza; se bisogna dissipare l'incertezza forse è meglio acquisire nuovi dati;
- se $P > 0.10$, allora i dati sono pienamente compatibili con H_0 : in sostanza non si rifiuta H_0 .

Tradizionalmente la soglia più severa, 0.005, viene mitigata a 0.01 e si indica la forza dell'indicazione ottenuta con degli asterischi:

*** se il P -value è fortemente critico contro H_0 ($P \leq 0.01$)

** se il P -value è critico contro H_0 ($0.01 < P \leq 0.05$)

* se il P -value è debolmente critico contro H_0 ($0.05 < P \leq 0.10$).

Se $P > 0.10$ non si segnano asterischi.

23.6 Test con livello di significatività α dal P -value

Si desidera un test decisionale per il problema di verifica d'ipotesi

$$H_0 : \theta \in \Theta_0 \text{ contro } H_1 : \theta \in \Theta \setminus \Theta_0.$$

Si supponga che sia stato calcolato il *P-value*

$$P = P(y) = \sup_{\theta \in \Theta_0} P_\theta(Y \in E_y).$$

Allora la regola di decisione:

$$\begin{aligned} \text{se } P(y) \leq \alpha & \text{ allora } d_n(y) = 1 \\ \text{se } P(y) > \alpha & \text{ allora } d_n(y) = 0 \end{aligned}$$

costituisce un test decisionale con livello di significatività almeno approssimato α . In altri termini, si basa la decisione sul paragone tra il *P-value* ottenuto con i dati y e il **prefissato** livello di significatività α .

In pratica, se un software statistico calcola il *P-value*, è assai facile effettuare il corrispondente test decisionale con livello di significatività α .

Esempio 23.3 Test di $H_0 : \mu = \mu_0$ contro $H_1 : \mu > \mu_0$ nel campionamento casuale semplice con numerosità n da $N(\mu, \sigma_0^2)$, $\sigma_0^2 > 0$ noto.

La statistica test è la media campionaria $\bar{y}_n = \sum_{i=1}^n y_i/n$, con solo valori grandi critici contro H_0 . Il *P-value* è

$$P(y) = P_{\mu_0}(\bar{Y}_n \geq \bar{y}_n) = 1 - \Phi\left(\frac{\bar{y}_n - \mu_0}{\sigma/\sqrt{n}}\right).$$

Si rifiuta H_0 al livello di significatività α se $P(y) \leq \alpha$, ossia se

$$\begin{aligned} 1 - \Phi\left(\frac{\bar{y}_n - \mu_0}{\sigma/\sqrt{n}}\right) &\leq \alpha \\ \iff 1 - \alpha &\leq \Phi\left(\frac{\bar{y}_n - \mu_0}{\sigma/\sqrt{n}}\right) \\ \iff \Phi^{-1}(1 - \alpha) &\leq \frac{\bar{y}_n - \mu_0}{\sigma/\sqrt{n}} \\ \iff z_{1-\alpha} &\leq \frac{\bar{y}_n - \mu_0}{\sigma/\sqrt{n}} \\ \iff \bar{y}_n &\geq \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha}, \end{aligned}$$

ritrovando così la regione di rifiuto del test (23.1). ◇

23.7 Calcolo del *P-value* in modelli normali di c.c.s.

Nel modello di c.c.s. con numerosità n da $N(\mu, \sigma_0^2)$ con $\sigma_0^2 > 0$ noto si consideri l'ipotesi nulla $H_0 : \mu = \mu_0$. L'informazione su μ presente nel campione è concentrata nella statistica \bar{y}_n . La statistica media campionaria standardizzata sotto H_0

$$Z = \frac{\bar{Y}_n - \mu_0}{\sigma_0/\sqrt{n}}$$

ha, sotto H_0 , distribuzione campionaria $N(0, 1)$. Il P -value associato all'osservazione

$$z = \frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}}$$

dipende da quali valori sono considerati più estremi, più critici contro H_0 . Si ha

- se $H_1 : \mu > \mu_0$ sono più critici contro H_0 valori grandi di \bar{y}_n , e quindi di z , per cui il P -value è

$$P = P(Z \geq z) = 1 - \Phi\left(\frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}}\right)$$

- se $H_1 : \mu < \mu_0$ sono più critici contro H_0 valori piccoli di \bar{y}_n , e quindi di z , per cui il P -value è

$$P = P(Z \leq z) = \Phi\left(\frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}}\right)$$

- se $H_1 : \mu \neq \mu_0$ sono critici contro H_0 valori sia grandi sia piccoli di \bar{y}_n , e quindi di z , per cui il P -value è

$$\begin{aligned} P &= P(|Z| \geq |z|) = 2P(Z \geq |z|) = 2\min(P(Z \geq z), P(Z \leq z)) \\ &= 2\min\left(\Phi\left(\frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}}\right), 1 - \Phi\left(\frac{\bar{y}_n - \mu_0}{\sigma_0/\sqrt{n}}\right)\right). \end{aligned}$$

Analogamente, nel modello di c.c.s. con numerosità n da $N(\mu, \sigma^2)$ con $\sigma^2 > 0$ ignoto si consideri l'ipotesi nulla $H_0 : \mu = \mu_0$. L'informazione su μ presente nel campione è concentrata nella statistica \bar{y}_n . La statistica media campionaria studentizzata

$$T_n = \frac{\bar{Y}_n - \mu_0}{S_n/\sqrt{n}}$$

sotto H_0 ha distribuzione campionaria t_{n-1} . Il P -value associato all'osservazione

$$t_n = \frac{\bar{y}_n - \mu_0}{s_n/\sqrt{n}}$$

dipende da quali valori sono considerati più estremi, più critici contro H_0 . Con $T_n \sim t_{n-1}$ si ha

- se $H_1 : \mu > \mu_0$

$$P = P(T_n \geq t_n)$$

- se $H_1 : \mu < \mu_0$

$$P = P(T_n \leq t_n)$$

- se $H_1 : \mu \neq \mu_0$

$$P = P(|T_n| \geq |t_n|) = 2P(T_n \geq |t_n|).$$

23.8 Esercizi

Esercizio 23.1 Dato un campione casuale costituito da un numero $n \in \mathbb{N}^+$, sufficientemente elevato, di variabili casuali Y_1, \dots, Y_n indipendenti con legge di Poisson avente parametro $\lambda > 0$ ignoto, si definisca la regione di rifiuto di un test unilaterale destro per λ di livello approssimato α e si espliciti la formula per il calcolo approssimato del *p-value*.

Esercizio 23.2 Il peso, espresso in milligrammi, dei granelli di polvere rilevati su una lastra di silicio è distribuito come una variabile casuale normale con media μ e varianza σ^2 . Si osservano 12 granelli il cui peso è 0.39, 0.68, 0.82, 1.35, 1.38, 1.62, 1.70, 1.71, 1.85, 2.14, 2.89, 3.69. Si individui uno stimatore per μ , si determini l'associato *standard error* e si calcolino i corrispondenti valori di stima. Si verichi l'ipotesi $H_0 : \mu = 2$ contro un'alternativa unilaterale sinistra con un livello di significatività $\alpha = 0.01$ supponendo σ^2 ignoto.

Esercizio 23.3 Il peso, espresso in chilogrammi, della *Lagenaria vulgaris* (una zucca ornamentale) è distribuito come una variabile casuale normale con media μ e varianza σ^2 . Si osservano i pesi di 12 zucche, che sono 0.39, 0.68, 0.82, 1.35, 1.38, 1.62, 1.70, 1.71, 1.85, 2.14, 2.89, 3.69. Si individuino stimatori di μ e σ^2 e si calcolino i corrispondenti valori di stima. Si determini lo *standard error* stimato associato allo stimatore per μ . Si verichi l'ipotesi $H_0 : \sigma^2 = 1$ contro un'alternativa unilaterale sinistra con un livello di significatività $\alpha = 0.01$.

Esercizio 23.4 Per valutare se una bilancia è ben calibrata si effettuano 60 pesate di un oggetto dal peso di 1000 g. Dai dati si ricava che $\sum_{i=1}^{60} y_i = 60\,036$ e $\sum_{i=1}^{60} y_i^2 = 60\,072\,262$. Nell'ipotesi che il peso sia distribuito come una variabile casuale normale con media μ e varianza σ^2 , si individuino opportuni stimatori per μ e per σ^2 e si calcolino i corrispondenti valori di stima. Si determini lo *standard error* stimato associato allo stimatore per μ . Si verifichi l'ipotesi $H_0 : \mu = 1\,000$ contro un'alternativa bilaterale, con un livello di significatività $\alpha = 0.01$.

Esercizio 23.5 Si esegua il comando R

```
> binom.test(n=100, x=50, p=0.5)
```

e si interpretino i risultati.

Esercizio 23.6 Si esegua il comando R

```
> binom.test(n=100, x=35, p=0.5)
```

e si interpretino i risultati.

Esercizio 23.7 Si esegua il comando R

```
poisson.test(x=12, r=5, alternative="two.sided", conf.level=0.95)
```

e si interpretino i risultati.