Probabilità e incertezza di misura

G. D'Agostini
Dipartimento di Fisica, Università "La Sapienza", Roma

30 ottobre 2001

Indice

I Dal concetto di probabilità ai problemi di probabilità inversa 1

1 Incertezza e probabilità 3
1.1 Determinismo e probabilismo nei metodi di indagine scientifica 3
1.2 Incertezze in Fisica e nelle altre scienze naturali 4
1.3 Limiti all'accuratezza delle misure - un esempio 6
1.4 Imparare dagli esperimenti: il problema dell'induzione 7
1.5 ❄Limiti del metodo di falsificazione 9
1.6 Decisioni in condizioni di incertezza 10
1.7 Concetto di probabilità 10
1.8 Semplici valutazioni di probabilità 13
1.9 Ricapitolando 15
1.10 Problemi 17

2 Valutazioni e interpretazioni della probabilità 19
2.1 Primi interessi in stime quantitative di probabilità 19
2.2 Valutazione combinatoria 20
2.3 Probabilità e frequenza 21
2.4 Legge empirica del caso e "definizione" frequentista 23
2.5 Interpretazione oggettivista e soggettivista della probabilità . . 25
2.6 Concetto di probabilità condizionata 26
2.7 Eventi di probabilità nulla 28
2.8 Probabilità e scommesse eque 29
2.9 ⭮ Probabilità e quote di scommessa 30
2.10 Definizione soggettiva di probabilità 31
2.11 ❄ La "definizione ISO" 32
2.12 ❄Note sul termine "soggettivo" 33
2.13 ❄Ruolo virtuale della scommessa, valore dei soldi e ordini di grandezza non intuitivamente percepibili 34
2.14 ⭮ Speranza matematica e previsione di vincita 36
2.15 ❄Previsione di guadagno e decisioni 37
2.16 ❄Decisioni vantaggiose e etica della ricerca 39
2.17 ❄Regola di penalizzazione - il bastone e la carota 40
2.18 Ricapitolando 41
2.19 Problemi 43

3 Elementi di calcolo combinatorio 47
3.1 Problemi elementari tipici 47
3.2 Disposizioni e combinazioni 48
3.2.1 Regola fondamentale del calcolo combinatorio 48
3.2.2 Numero di r-disposizioni di n oggetti 48
3.2.3 Numero di r-disposizioni semplici di n oggetti 48
3.2.4 Numero di permutazioni di n oggetti 50
3.2.5 Combinazioni 50
3.2.6 Coefficienti binomiali 52
3.2.7 Note su nomenclatura e simbologia 53
3.3 Note sul calcolo dei grandi numeri 53
3.4 Ordinamenti, occupazioni ed estrazioni 55
3.5 Alcuni esempi classici 57
3.6 Ricapitolando 59
3.7 Problemi 60

4 Regole della probabilità 61
4.1 Probabilità della somma logica di due eventi incompatibili . . 61
4.2 Eventi e insiemi 62
4.3 probabilità come misura 67
4.4 Evento condizionato 67
4.5 Regole di base della probabilità - assiomi 69
4.5.1 ❄Dimostrazioni delle proprietà della probabilità 71
4.6 Relazione fra probabilità condizionata e congiunta 72
4.7 Condizionamento da eventi di probabilità nulla 74
4.8 Indipendenza stocastica (o in probabilità) 75
4.9 Altre proprietà della probabilità condizionata 76
4.9.1 Legge della moltiplicazione 76
4.9.2 Legge delle alternative 78
4.10 Indipendenza logica e indipendenza stocastica 78
4.11 Ricapitolando 78
4.12 Problemi 81

5 Probabilità delle cause e meccanismo di aggiornamento delle probabilità 85
5.1 Inferenza probabilistica 85
5.2 Teorema di Bayes 87
5.3 Chiavi di lettura del teorema di Bayes 89
5.4 Visione combinatoria del teorema di Bayes 91
5.5 Esempi tipici di applicazione 92
5.5.1 Classificazione di eventi e rapporto segnale rumore . . 92
5.5.2 Uso iterativo del teorema di Bayes 94
5.6 ⭮ Statistica bayesiana: imparare dall'esperienza 95
5.7 ⭮ Il caso del sospetto baro 96
5.7.1 I "fatti" 96
5.7.2 Riaggiornamento della probabilità 96
5.7.3 Confronto fra inferenza diretta e inferenza iterativa . . 97
5.7.4 Dipendenza dalla probabilità iniziale 98
5.7.5 Pregiudizio, indizi e conclusioni 98

© G. D'Agostini 2001

5.7.6 Probabilità e decisione 98
5.8 ❄Recupero e superamento del metodo di falsificazione 99
5.9 ⟳ Osservazioni indipendenti e prodotto delle verosimiglianze . 100
5.10 ❄Fattore di Bayes e incremento logaritmico delle quote di scommessa 100
5.11 ⟳ Indifferenza iniziale e massima verosimiglianza 101
5.12 ❄Problema della verificabilità ed estensione del concetto di evento 101
5.13 Ricapitolando 102
5.14 Problemi 104

II Variabili casuali - I **109**

6 Variabili casuali e distribuzioni di probabilità di variabili discrete 111

6.1 Numeri aleatori 111
6.2 Distribuzione di probabilità 112
6.3 ⟳ Distribuzione di probabilità e distribuzioni statistiche 113
6.4 Esempi di costruzione di distribuzioni di variabili casuali . . . 115
6.5 Proprietà delle distribuzioni di probabilità discrete 118
6.6 Distribuzioni elementari notevoli 119
6.6.1 Distribuzione uniforme discreta 119
6.6.2 ❄Distribuzione uniforme discreta - caso generale . . . 119
6.6.3 Processo di Bernoulli 120
6.6.4 Combinazione di molti processi di Bernoulli indipendenti e di uguale probabilità 121
6.7 Distribuzione geometrica 122
6.8 Sintesi di una distribuzione di probabilità: previsione e incertezza di previsione 123
6.9 Previsione (o valore atteso) come baricentro della distribuzione 126
6.9.1 Osservazioni su terminologia e notazioni 127
6.9.2 Valore atteso di una funzione di una variabile casuale . 128
6.10 Valore atteso di distribuzioni elementari 128
6.10.1 Distribuzione uniforme discreta 129
6.10.2 Processo di Bernoulli 129
6.10.3 Distribuzione geometrica 130
6.11 Incertezza “standard” di previsione 130
Varianza e deviazione standard 130
6.12 Proprietà formali di varianza e deviazione standard 132
6.13 ❄ Momenti di una distribuzione e altri indicatori di forma . . . 133
6.14 ❄ Entropia come misura dello stato di incertezza 134
6.15 Deviazione standard delle distribuzioni elementari 134
6.15.1 Distribuzione uniforme fra 1 e n 135
6.15.2 ❄ Distribuzione uniforme di n valori fra a e b 135
6.15.3 Processo di Bernoulli 135
6.15.4 Distribuzione geometrica 136
6.16 ⟳ Processo di Bernoulli e percezione di probabilità prossime a 0 o a 1 137

© G. D’Agostini 2001

6.17 ❄ Previsione e incertezza di previsione di vincita in giochi d'azzardo 137
6.17.1 Gioco della roulette 137
6.17.2 I sistemi "per vincere" al lotto 139
6.18 ⟳ Misure di centralità e di dispersione di distribuzioni statistiche 141
6.19 Ricapitolando 143
6.20 Problemi 145

7 Distribuzioni di probabilità di variabili discrete - II 147
7.1 Distribuzione binomiale 147
7.2 ⟳ Distribuzione binomiale – da capo 149
7.3 Proprietà della distribuzione binomiale e note sul suo uso . . . 151
7.3.1 Valore atteso e deviazione standard 151
7.3.2 Usi tipici della distribuzione binomiale 154
7.4 Distribuzione di Poisson 154
7.5 ⟳ Processo di Poisson - prima parte 156
7.6 ❄ Formule ricorsive per la distribuzione binomiale e di Poisson 161
7.7 ⟳ Proprietà riproduttiva delle distribuzioni di probabilità binomiale e di Poisson 161
7.8 ❄ Altre distribuzioni di interesse 162
Distribuzione di Pascal 162
Binomiale negativa 164
Distribuzione ipergeometrica 165
7.9 ❄Cammino casuale e problema della rovina del giocatore . . . 166
7.10 Quanto credere in "$X = \mu \pm \sigma$"? 168
7.10.1 Alcuni esempi numerici 168
7.10.2 Disuguaglianza di Markov 170
7.10.3 Disuguaglianza di Cebicev 170
7.11 Intervalli di probabilità, o di credibilità 171
7.12 ❄Previsione, penalizzazione e valore sul quale scommettere . . 172
7.13 ⟳ Previsione di frequenza relativa e legge dei grandi numeri . 173
7.14 ⟳ Previsione di una distribuzione statistica 174
7.14.1 Introduzione al concetto di correlazione fra variabili casuali 175
7.15 ⟳ Un esempio storico di distribuzione di Poisson come introduzione al problema della verifica delle leggi statistiche 176
7.15.1 Previsione del tipo di distribuzione 176
7.15.2 Stima "puntuale" del parametro della distribuzione . . 176
7.15.3 Previsione quantitativa della distribuzione statistica, subordinata a $\lambda = \overline{d}$, e confronto con le osservazioni . . 177
Inferenza probabilistica su λ 178
Previsione della distribuzione statistica subordinata all'incerteza su λ 179
7.16 ⟳ Estensione dei teoremi sulla probabilità alle funzioni di probabilità discrete 179
7.17 Ricapitolando 181
7.18 Problemi 184

8 Distribuzioni di probabilità di variabili continue 187
8.1 Variabili casuali continue e densità di probabilità 187
8.1.1 Probabilità nulle con diversi gradi di fiducia 187
8.1.2 Dal grado di fiducia alla probabilità finita 188
8.1.3 Funzione densità di probabilità 189
8.1.4 Proprietà della funzione densità di probabilità e della funzione di ripartizione 190
8.1.5 Valori attesi 190
8.2 Distribuzione uniforme continua 192
8.3 ❄ Simulazione al computer di processi stocastici 193
8.3.1 Costruzioni di altre semplici variabili casuali 194
Generica distribuzione uniforme fra a e b 194
Processo di Bernoulli e distribuzione binomiale 194
Distribuzione uniforme discreta 194
Marcia a caso 194
8.3.2 Scelta pesata con $f(x)$ 195
8.3.3 Scelta uniforme lungo $F(x)$ 195
8.4 ↻ Distribuzioni triangolari 196
8.5 Distribuzione esponenziale 197
8.6 ❄ Distribuzione esponenziale doppia 198
8.7 Distribuzione normale 199
8.8 Distribuzione normale standardizzata 202
8.9 Uso delle tabelle dell'integrale della distribuzione normale standardizzata 204
8.10 ❄ Derivazione della gaussiana come limite di funzione binomiale o poissoniana 208
8.11 ↻ Proprietà riproduttiva della distribuzione normale 209
8.12 ↻ Processo di Poisson - Seconda parte 210
8.12.1 Distribuzione del tempo di attesa del primo successo . 210
8.12.2 Relazione fra esponenziale e poissoniana 211
8.12.3 Relazione fra esponenziale e geometrica 212
8.12.4 Tempo di attesa del k-mo successo 213
8.12.5 Intensità di più processi di Poisson indipendenti 214
8.12.6 Vita media di decadimento 215
8.13 ❄ Funzione generatrice dei momenti 215
Binomiale 217
Poissoniana 217
Gaussiana 217
Altre proprietà e applicazioni 218
8.14 ↻ Altre distribuzioni di interesse 219
8.14.1 Beta 219
8.14.2 Gamma 221
8.14.3 Chi^2 221
8.14.4 t di Student 224
8.14.5 F 225
8.15 Ricapitolando 226
8.16 Problemi 227

© G. D'Agostini 2001

III Variabili casuali - II **229**

9 Variabili casuali multiple **231**
9.1 Vettori aleatori . . . 231
9.1.1 Variabili casuali doppie discrete . . . 232
9.1.2 Variabili casuali doppie continue . . . 233
9.2 Distribuzioni marginali . . . 234
9.3 Estensione dei teoremi sulla probabilità alle distribuzioni di probabilità . . . 236
9.3.1 Distribuzioni condizionate . . . 236
9.3.2 Variabili casuali indipendenti . . . 237
9.3.3 Formula delle alternative e teorema di Bayes . . . 237
9.4 Previsione e incertezza di previsione . . . 238
9.5 ⏱ Covarianza e coefficiente di correlazione . . . 239
9.5.1 Variabili correlate e misura della correlazione . . . 239
9.5.2 Proprietà formali di covarianza e coefficiente di correlazione . . . 242
9.6 ⏱ Matrice di covarianza e matrice di correlazione . . . 244
9.7 ⏱ Esempi di variabili doppie discrete . . . 244
9.8 ⏱ Esempi di distribuzione bidimensionale continua . . . 249
9.8.1 Distribuzione uniforme in un rettangolo . . . 249
9.8.2 Distribuzione uniforme in un triangolo . . . 250
9.9 ❄ Distribuzione multinomiale . . . 251
9.10 ❄ Distribuzione normale bivariata . . . 256
9.11 ❄ Caso generale di distribuzione multivariata . . . 261
Derivate di Q^2 rispetto alle variabili casuali . . . 263
9.12 ⏱ Distribuzioni statistiche multivariate . . . 263
9.13 varie . . . 264

10 Funzioni di variabili casuali e teoremi limite **265**
10.1 Propagazione delle incertezze . . . 265
10.2 Soluzione generale per variabili discrete . . . 266
10.2.1 Regola generale . . . 266
10.2.2 ❄ Convoluzione di due funzioni di probabilità . . . 267
10.2.3 Trasformazione di una variabile distribuita uniformemente . . . 269
10.3 ❄ Soluzione generale per variabili continue . . . 271
10.3.1 Cambiamento di variabile . . . 271
Trasformazioni di una distribuzione uniforme . . . 272
Applicazioni alle simulazioni di variabili casuali . . . 272
Trasformazione lineare di una variabile distribuita normalmente . . . 274
10.3.2 Caso di funzioni non monotone . . . 274
10.3.3 Somma di due variabili . . . 275
Somma di due variabili distribuite uniformemente . . . 275
Somma di due variabili distribuite normalmente . . . 276
10.4 ❄ Uso della funzione generatrice dei momenti . . . 277
10.4.1 $Z = X + Y$, con X e Y poissoniane . . . 277
10.4.2 $Z = aX + bY + c$, con X e Y gaussiane . . . 278

© G. D'Agostini 2001

10.5 ❄ Stime a bruta forza: metodi di Monte Carlo 278
10.6 Riepilogo di alcune proprietà delle funzioni di variabili casuali 280
10.7 Valore atteso e varianza di combinazioni lineari 280
Valore atteso e varianza della distribuzione binomiale . 283
Valore atteso e varianza della distribuzione di Erlang . 283
Previsione di una media aritmetica di variabili aleatorie analoghe 283
10.8 ⟳ Correlazione fra diverse combinazioni lineari di variabili casuali 284
Covarianza di due medie aritmetiche 286
Correlazione fra una variabile e una combinazione lineare che la contiene 287
10.9 Legge dei grandi numeri 287
10.9.1 Limite della media aritmetica 288
10.9.2 Teorema di Bernoulli 289
Lancio di una moneta 290
Sul recupero dei numeri ritardatari 290
10.10 Teorema del limite centrale 292
10.10.1 Distribuzione della media aritmetica 295
10.10.2 Convergenza in distribuzione della binomiale e della poissoniana 295
10.11 Estensione del teorema del limite centrale a variabili non indipendenti 297
10.12 ❄ Simulazione di numeri aleatori distribuiti secondo una distribuzione normale 297
10.13 ⟳ Linearizzazione 298
10.14 ⟳ Esempio di applicazione alle incertezze di misure 299
10.15 ⟳ Moto browniano, "pallinometro" ed errori di misura 301
10.16 ❄ Distribuzione di velocità delle molecole di un gas perfetto . 304
10.17 Problemi 307

IV Applicazioni di statistica inferenziale 309

11 Impostazione del problema. Caso di verosimiglianza gaussiana 311
11.1 Introduzione 311
11.2 Verosimiglianza normale con σ nota 313
11.3 Effetto di una prior rilevante: combinazione di risultati 316
11.4 ❄ Derivazione di Gauss della gaussiana 318
11.5 ❄ Caso di forte vincolo dato dalla prior 320
11.6 Caso di σ ignota 322
11.6.1 Ragionamento intuitivo 323
11.6.2 Possibili dubbi sul modello normale 324
11.6.3 ❄ Inferenza simultanea su μ e σ 324
Prior uniforme in σ 325
Prior uniforme in $\log\sigma$ 327
Incertezza su σ 328
11.6.4 Distribuzione di $1/\sigma^2$ 331
11.6.5 Conclusioni e raccomandazioni 333

© G. D'Agostini 2001

11.7 Distribuzione predittiva 333
11.8 Combinazione scettica 335
11.9 Problemi 338

12 Verosimiglianza binomiale e poissoniana. Approssimazioni 339
12.1 Misure di conteggi, di proporzioni e di efficienze 339
12.2 Inferenza su p e λ (o r) in condizioni di normalità. 339
12.2.1 Caso poissoniano 340
12.2.2 Caso binomiale 340
12.3 ⥁ Caso generale di inferenza con verosimiglianza binomiale . 341
12.3.1 Caso di routine 342
12.3.2 Casi critici 343
12.3.3 Combinazione di misure indipendenti 344
12.3.4 ❄ Uso della prior coniugata Beta 344
12.4 ⥁ Caso generale di inferenza con verosimiglianza poissoniana 346
12.4.1 Caso di routine 346
12.4.2 Caso di $x = 0$ con prior uniforme 347
12.4.3 Combinazione di risultati 348
12.4.4 ❄ Uso della prior coniugata Gamma 348
12.4.5 Inferenza sull'intensità del processo di Poisson da osservazioni effettuate con diversi tempi di osservazione 349

13 Sufficienza statistica, limite a normale e metodi frequentistici 351

14 Effetti sistematici e di rumore 353
14.1 Considerazioni generali 353
14.2 Soluzioni esatte sotto ipotesi di normalità 353
14.2.1 Incertezza sullo zero dello strumento 353
14.2.2 Correzione per errori sistematici noti 355
14.2.3 Correlazione fra i risultati introdotta dalla non perfetta conoscenza dello zero dello strumento 356
14.3 Effetto del background nella misura dell'intensità di un processo di Poisson 358
14.4 Propagazioni di incertezza, approssimazioni e linearizzazioni . 361
14.5 Matrice di covarianza di dati correlati 361
Offset uncertainty 361
Normalization uncertainty 362
General case 363

15 Adattamento di curve ai dati sperimentali e stima dei parametri 365
15.1 Inferenza sui parametri di una legge 365
15.2 ❄ Come tener conto anche di possibili incertezze sulle X . . . 367
15.3 Formule dei minimi quadrati 368
15.3.1 σ_Y nota e costante 368
15.3.2 σ_{Y_i} ignote e supposte costanti 369
15.3.3 σ_{Y_i} diverse e note a priori 369

© G. D'Agostini 2001

16 Test di ipotesi **371**
16.1 Riepilogo dell'approccio probabilistico 371
16.2 Schema di test di ipotesi nell'approccio frequentista 371
16.3 Conclusioni 371

V Soluzione dei problemi **373**

© G. D'Agostini 2001

Parte III

Variabili casuali - II

Capitolo 9

Variabili casuali multiple

9.1 Vettori aleatori

Nei capitoli precedenti abbiamo considerato un numero aleatorio alla volta, introducendo il formalismo generale per descrivere lo stato di incertezza rispetto. Come si può bene immaginare, nei casi reali ci si trova spesso a dover considerare più grandezze dai valori incerti. Facciamo degli esempi.

- Consideriamo 100 lancio consecutivi di una moneta. Si può essere interessati al numero di teste, al numero di croci, al numero più elevato di teste consecutive che si è verificato e al numero di cambiamenti di esito (ogni volta che dopo un certo numero di teste è uscita croce e viceversa). Quali risultati sarà possibile ottenere? $\{0, 0, 0, 0\}$, $\{1, 0, 0, 0\}$, $\ldots\{50, 50, 1, 49\}$, $\ldots\{100, 100, 100, 100\}$? (Con "$\{\ldots\}$" è indicata la lista ordinata dei valori numerici di ciascuna delle variabili di interesse. Molte di queste combinazioni risultano impossibili.)

- Scelgo uno studente a caso che abbia terminato tutti gli esami del corso. Che voti ha riportato in ciascuno degli esami? $\{18, 18, 18, \ldots, 18\}$, $\{19, 18, 18, \ldots, 18\}$, $\ldots$, $\{30, 30, 30, \ldots, 30\}$? (Con "$\{\ldots\}$" è indicata la lista dei possibili risultati, ordinata per ordine alfabetico degli esami.)

- Si lancia una pallina su un tavolo di $120\,\text{cm} \times 80\,\text{cm}$? In quale punto del piano si fermerà? $\{0.0, 0.0\}$, $\{0.0, 0.1\}$, $\{0.1, 0.1\}$, $\ldots\{120.0, 80.0\}$, (discretizzando al millimetro le coordinate, prese a partire da uno spigolo)

- La legge fisica che lega la grandezza X (misurata in unità arbitrarie) al tempo è del tipo $X = m\,t + c$. Se m e c valgono rispettivamente 3 e di 1 (unità arbitrarie) e la misura è affetta da errori casuali, quali valori di X_1 e di X_2 (misurati con "strumenti reali") si osserveranno ai tempi $t_1 = 1$ e $t_2 = 2$? $\{3.7, 6.7\}$, $\{3.8, 6.7\}$, $\ldots\{4.0, 7.0\}$, $\ldots\{4.5, 7.5\}$?

- Caso inverso del precedente: avendo osservato $X_1 = 3.9$ e $X_2 = 7.2$ ai tempi $t_1 = 1$ e $t_2 = 2$, cosa si può dire sui "valori veri" dei coefficienti m e c? $\{3.1, 0.4\}$, $\ldots\{3.3, 0.6\}$, $\ldots\{3.4, 0.5\}$?

Ricapitolando, in generale si è in stato di incertezza rispetto alla realizzazione di n variabili casuali, o , come si dice, rispetto a un *vettore aleatorio* di n componenti, anche chiamato n-*tupla* (leggi "entupla") di valori.

Come al solito, lo stato di incertezza su ciascun possibile vettore aleatorio sarà quantificato da un grado di fiducia. Vediamo separatamente il caso discreto e quello continuo. Successivamente, molte delle proprietà saranno descritte in modo generale per i due casi, eventualmente scegliendo variabili discrete o continue per gli esempi, a seconda di quella che meglio si presta al caso.

9.1.1 Variabili casuali doppie discrete

Il concetto di distribuzione di probabilità si estende in modo naturale alle variabili multiple. Cominciamo anche in questo caso con le variabili discrete e, per semplicità, con il caso bidimensionale. L'estensione al caso generale è immediato. La funzione di probabilità è[1]:

$$f(x,y) = P(X = x, Y = y)\,, \tag{9.1}$$

la quale è definita non negativa e compresa fra 0 e 1. Le possibili coppie di valori di X e Y (o il vettore, o n-tupla, di dimensione 2), ciascuna con il suo grado di fiducia $f(x,y)$, danno luogo alla *distribuzione congiunta* delle variabili.

La condizione di normalizzazione richiede che la somma su tutte le coppie sia pari a 1. Questo può essere scritto in diversi modi.

- Si può pensare a ciascuna coppia come un punto $\mathbf{R}$ del piano (X,Y) e ordinata mediante l'indice i. La somma delle probabilità sommare su tutti gli $n_R = n_X \times n_Y$ punti deve dare 1:

$$\sum_{i=1}^{n_R} P(\mathbf{R_i}) = \sum_{i=1}^{n_R} P(X = x_i, Y = y_i) \;\; = \;\; \sum_{i=1}^{n_R} f(x_i, y_i) = 1\,.$$

- Si può indicare la variabile X con l'indice i, la Y con j e sommare su tutte le possibilità degli indici:

$$\sum_{i,j} P(X = x_i, Y = y_j) \;\; = \;\; \sum_{i,j} f(x_i, y_j) = 1\,.$$

  Facciamo un esempio che chiarisca l'uso del simbolo $\sum_{i,j}$. Se la variabile X può assumere 3 valori e la variabile Y 2 valori si ha che

$$\begin{aligned}\sum_{i,j} f(x_i, y_j) \;\; &= \;\; f(x_1,y_1) + f(x_1,y_2) + f(x_2,y_1) \\ &\quad + f(x_2,y_2) + f(x_3,y_1) + f(x_3,y_2)\end{aligned}$$

[1] $P(X = x, Y = y)$ sta per $P(X = x \cap Y = y)$. Si ricordi inoltre quanto detto nel capitolo 6 sulla flessibilità dell'uso di simboli.

© G. D'Agostini 2001

- si può infine scrivere più semplicemente
$$\sum_{x,y} f(x,y) \;=\; 1\,,$$
intendendo semplicemente che la sommatoria corre su tutti i possibili valori di x e di y.

Anche in questo caso si definisce una funzione di ripartizione, in perfetta analogia al caso unidimensionale:

$$F(x_k, y_l) = P(X \leq x_k,\, Y \leq y_l) = \sum_{\substack{x_i \leq x_k \\ x_j \leq x_l}} f(x_i, y_l),$$

Essa soddisfa le seguenti proprietà:

1. $F(x,y)$ è non decrescente, ossia $F(x_1,y_1) \leq F(x_2,y_2)$ se $x_1 < x_2$ e $y_1 < y_2$;
2. $F(x,y)$ tende a 0 per $(x \to -\infty, y \to -\infty)$;
   $F(x,y)$ tende a 1 per $(x \to +\infty, y \to +\infty)$;
3. $F(x,y)$ è continua a destra;
4. $f(x_i,y_i) = F(x_i,y_i) - F(x_{i-1},y_{i-1})$.

La tabella 9.1 mostra un esempio di distribuzione doppia di variabili discrete.

9.1.2 Variabili casuali doppie continue

Per introdurre le variabili continue partiamo dalla funzione di ripartizione:

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x',y')\,dx'dy'.$$

In questo caso l'elemento infinitesimo di probabilità è pari a:

$$dF(x,y) = f(x,y)dxdy\,.$$

Esso ha il significato di

$$P[(x \leq X \leq x+dx) \cap (y \leq Y \leq y+dy)]\,.$$

La funzione densità di probabilità, ottenuta dalla funzione di ripartizione come

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}\,,$$

rappresenta invece il grado di fiducia del punto (x,y), come discusso al momento di introdurre le variabili continue. Le proprietà della densità di probabilità congiunta $f(x,y)$ sono analoghe al caso unidimensionale visto nel capitolo precedente. Ricordiamo soltanto che $f(x,y)$ ha le dimensioni inverse di $X \cdot Y$ e che la condizione di normalizzazione diventa:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1\,.$$

© G. D'Agostini 2001

	$f(x,y)$				
	x_1	x_2	x_3	x_4	$f(y)$ $\downarrow$
y_1	0.08	0.05	0.01	0.10	0.24
y_2	0.11	0.07	0.12	0.02	0.32
y_3	0.10	0.08	0.14	0.12	0.44
$f(x) \rightarrow$	0.29	0.20	0.17	0.24	

	$F(x,y)$				
	x_1	x_2	x_3	x_4	$F(y)$ $\downarrow$
y_1	0.08	0.13	0.14	0.24	0.24
y_2	0.19	0.31	0.44	0.56	0.56
y_3	0.29	0.49	0.76	1.00	1.00
$F(x) \rightarrow$	0.29	0.49	0.76	1.00	

Tabella 9.1: Esempio di distribuzione congiunta e di marginalizzazione.

9.2 Distribuzioni marginali

Dalla distribuzione di probabilità congiunta è possibile ottenere la distribuzione di probabilità di una delle variabili, "sommando" (in senso lato, che per il caso continuo equivale ad integrare) le probabilità di tutte le possibilità delle altre variabili per un certo valore della variabile di interesse. Nel caso discreto si ha, ad esempio rispetto alla X:

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j)\,.$$

Questa operazione è detta di *marginalizzazione* e la tabella 9.1 ne mostra un esempio (si noti come le distribuzioni di ciascuna delle variabili sono ottenute sommando "a margine" i valori della tabella della probabilità congiunta). Nel caso di variabili continue si ha[2]:

$$f(x) \;=\; \int f(x,y)dy\,, \tag{9.2}$$

e analogalmente per $f(y)$, integrando su x. Le distribuzioni ottenute con questa operazioni sono chiamate *distribuzioni marginali*. Sia chiaro che, in principio, tale appellativo dovrebbe essere superfluo, in quanto ogni distribuzione

[2]Chiaramente $f(y)$ e $f(y)$ sono in genere funzioni diverse, anche se entrambe indicate con lo stesso simbolo $f(\cdot)$, cambiando solo la variabile nell'argomento. A volte, per ricordare che si tratta di funzioni diverse, si $f_X(x)$ e $f_Y(y)$. In modo analogo dovremmo indicare $f(x\,|\,y)$ con $f_{X\,|\,Y}(x\,|\,y)$, e così via. Eviteremo questo modo più preciso (e pesante) in quanto non ci sono ragioni per temere ambiguità. Allo stesso modo gli estremi degli integrali saranno omessi a sottintendere che essi si estendono su tutti i possibili valori di X:

$$f(x) = \int f(x,y)\,\mathrm{d}y \longleftrightarrow f(x) = \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y\,.$$

© G. D'Agostini 2001

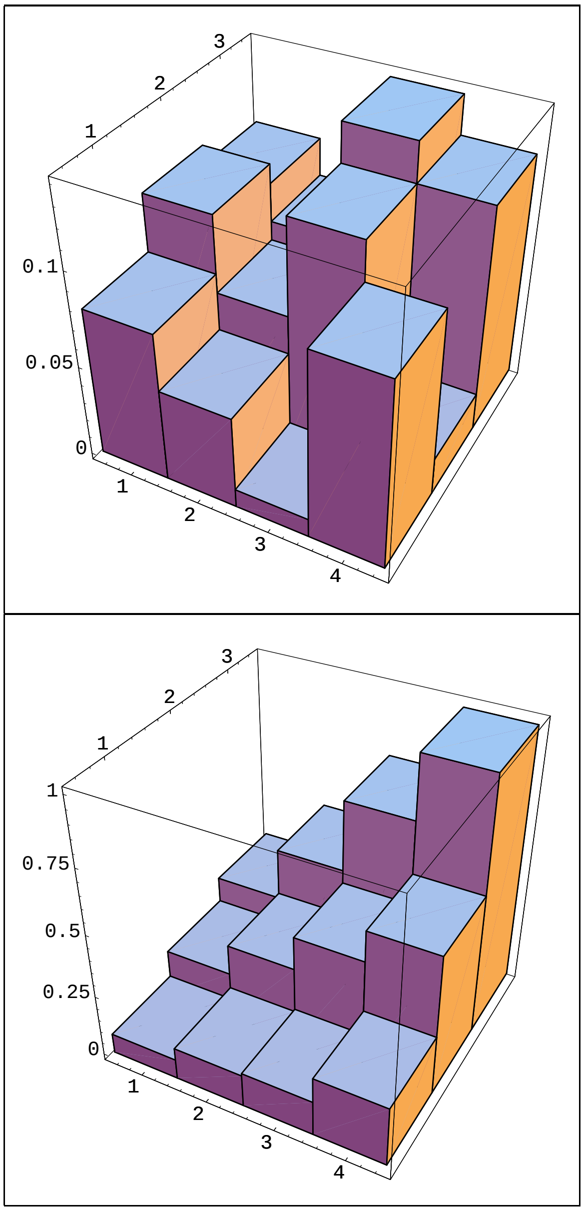

Figura 9.1: Visualizzazione qualitativa, mediante “lego plot”, della distribuzione di probabilità di tabella 9.1: in alto $f(x, y)$ e in basso $F(x, y)$. Si faccia attenzione al fatto che la rappresentazione grafica non corrisponde esattamente alla distribuzione di partenza: perche?

© G. D’Agostini 2001

di una sola variabile può essere pensata come una marginale rispetto ad altre infinite variabili che non interessano. In sostanza, quando si parla esplicitamente di distribuzione marginale, significa semplicemente che si sta facendo riferimento ad una distribuzione congiunta di partenza. Si noti inoltre l'interpretazione della marginale come, per così dire, "proiezione" della congiunta (non da prendere in modo rigoroso nel senso geometrico).

9.3 Estensione dei teoremi sulla probabilità alle distribuzioni di probabilità

Questo argomento è già stato introdotto nel paragrafo 7.16, limitatamente alle variabili distrete. L'estensione alle variabili continue è abbastanza immediata, essendo sufficiente sostituire alle probabilità gli elementi infinitesimi di probabilità. Riprendiamo ora il discorso in maniera più generale, aggiungendo anche delle osservazioni sull'interpretazioni delle distribuzioni condizionate.

9.3.1 Distribuzioni condizionate

In analogia alla probabilità condizionata esistono le distribuzioni di probabilità condizionate, nel senso che $f(x)$ può essere subordinata alla conoscenza che Y sia uguale a y, ovvero $f(x\,|\,Y=y)$. Essa è indicata con

$$f(x\,|\,y)$$

ed è funzione soltanto di x, mentre y svolge, per così dire, il ruolo di parametro. Quindi esistono tante diverse $f(x\,|\,y)$ quanti sono i possibili valori di Y, infiniti nel caso che Y sia una variabile continua.

Come detto nel paragrafo 7.16, l'estensione al caso discreto è immediata. Vediamo come ci si comporta per il caso continuo. Ricordiamo la formula che lega probabilità condizionata alla congiunta (vedi paragrafo 4.6):

$$P(A\,|\,B) = \frac{P(A\cap B)}{P(B)} \qquad (\,P(B)\neq 0\,) \tag{9.3}$$

Nel caso di variabile casuale continua possiamo scrivere in perfetta analogia:

$$P(x\le X\le x{+}\mathrm{d}x\,|\,Y=y) = \frac{P[(x\le X\le x+\mathrm{d}x)\cap(y\le Y\le y+\mathrm{d}y)]}{P(y\le Y\le y+\mathrm{d}y)}\,, \tag{9.4}$$

con $P(y\le Y\le y+\mathrm{d}y)\neq 0$. Passando alle funzioni densità di probabilità si ha

$$f(x\,|\,y)\,\mathrm{d}x \quad = \quad \frac{f(x,y)\,\mathrm{d}x\mathrm{d}y}{f(y)\,\mathrm{d}y}\,,$$

da cui

$$f(x\,|\,y) \quad = \quad \frac{f(x,y)}{f(y)} \tag{9.5}$$

con $f(y)\neq 0$. Si ricorda che, come visto per la (9.3), neanche la (9.5) definisce la distribuzione condizionata. In particolare, $f(x\,|\,y)$ può essere valutata

© G. D'Agostini 2001

senza che sia nota (o abbia importanza) la funzione congiunta $f(x, y)$ (vedi discussioni ai paragrafi 4.4 4.6). È comunque chiaro che, qualora invece la $f(x, y)$ sia assegnata, essa contiene l'informazione più completa sullo stato di incertezza di Y e di Y, in quanto da essa è possibile ricavarsi le marginali e le condizionate, ad esempio:

$$f(x \mid y) = \frac{f(x, y)}{\int f(x, y) dx} .$$

Si noti come la formula (9.5) non dipenda dal fatto che le $f()$ siano probabilità o densità di probabilità, ovvero la formula è indipendentemente dal tipo di variabili. Inoltre, per simmetria, si ha

$$f(x, y) = f(x \mid y) \cdot f(y) = f(y \mid x) \cdot f(x) . \tag{9.6}$$

Così pure, in modo analogo alla (4.22), si ha, nel caso di molte variabili:

$$f(x, y, z, \ldots) = f(y) \cdot f(x \mid y) \cdot f(z \mid x, y) \ldots$$

9.3.2 Variabili casuali indipendenti

Se la distribuzione di una variabile, subordinata ad un certo valore di un'altra variabile, è uguale alla distribuzione marginale si dice che le due variabili sono (stocasticamente) indipendenti. Questa condizione equivale a richiedere

$$\begin{aligned} f(x) &= f(x \mid y) \\ f(y) &= f(y \mid x) \end{aligned} \tag{9.7}$$

o

$$f(x, y) = f(x) \cdot f(y) , \tag{9.8}$$

analoga della

$$P(A \cap B) = P(A) \cdot P(B) .$$

Nel caso di molte variabili, la condizione di indipendenza richiede le (9.7) per ciascuna delle coppie, o, equivalentemente e più succintamente:

$$f(x, y, z, \cdots) = f(x) \cdot f(y) \cdot f(z) \cdots . \tag{9.9}$$

9.3.3 Formula delle alternative e teorema di Bayes

La versione generale del teorema di Bayes discende dalle espressioni della probabilità congiunta in funzione delle condizionate e delle marginali. Infatti dalla (9.6) segue

$$f(x \mid y) = \frac{f(y \mid x) \, f(x)}{f(y)} , \tag{9.10}$$

analoga della (5.3). Ricordando che $f(y)$ è la marginale rispetto a Y di $f(x, y)$ si ottiene l'analoga della formula delle alternative (mostriamo il solo caso continuo, quello discreto è assolutamente simile)

$$f(y) = \int f(x, y) \, \mathrm{d}x = \int f(y \mid x) f(x) \, \mathrm{d}x , \tag{9.11}$$

© G. D'Agostini 2001

che, sostituita nella (9.10), dà luogo alla formula più consueta del teorema di Bayes per variabili continue, del quale faremo largo uso nell'inferenza su valori di grandezze fisiche:

$$f(x\,|\,y) = \frac{f(y\,|\,x)\,f(x)}{\int f(y\,|\,x) f(x)\,\mathrm{d}x}\,. \tag{9.12}$$

9.4 Previsione e incertezza di previsione

La valutazione dei valori attese segue in modo naturale dal caso di una sola variabile. Questo può essere visto più chiaramente nel caso di variabili discrete: il valore atteso di una funzione di variabili casuali si calcola "facendo la somma, estesa su tutto il campo di definizione, del prodotto della funzione, calcolata in ogni punto, per la probabilità di quel punto". Come si vede, questa definizione non dipende dal numero di variabili casuali (ovvero dalle dimensioni dello spazio che contiene il vettore aleatorio), quindi nel caso più generale:

$$\mathrm{E}[g(X,Y,Z,\cdots)] = \sum_x\sum_y\sum_z \cdots g(x,y,z,\cdots)\,f(x,y,z,\cdots)\,. \tag{9.13}$$

Nel caso di variabili continue, la "probabilità di un punto" viene sostituita dal concetto di elemento infinitesimo di probabilità di trovare le variabili casuali nell'intorno di quel punto e le sommatorie diventano integrali su tutte le variabili:

$$\mathrm{E}[g(X,Y,Z,\cdots)] = \int\cdots\int g(x,y,z,\cdots)\cdot f(x,y,z,\cdots)\,\mathrm{d}z\mathrm{d}y\mathrm{d}z\cdots \tag{9.14}$$

Dalla formula generale ci possiamo calcolare il valore atteso e la varianza di ciacuna delle variabili. Prendendo, ad esempio, la X e considerando, per semplicità, ma senza perdere in generalità, due sole variabili continue, si tratta di calcolare $\mathrm{E}(X)$ e $\mathrm{E}[(X-\mu_X)^2]$. Il fatto che queste due funzioni, di cui si è interessati al valore atteso, dipendano soltanto da una variabile semplifica i calcoli, in quanto, in generale

$$\begin{aligned}
\mathrm{E}[g(X)] &= \iint g(x)\,f(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \int g(x)\,\left[\int f(x,y)\,\mathrm{d}y\right]\,\mathrm{d}x \\
&= \int g(x)\cdot f(x)\,\mathrm{d}x,
\end{aligned} \tag{9.15}$$

ottenendo lo stesso risultato che si sarebbe ottenuto utilizzando la funzione marginale della variabile di interesse. Abbiamo quindi ricondotto il problema del calcolo di media e di deviazione standard nel caso molte variabili (*multivariato*) a quello di una sola variabile (*univariato*). Questo risultato è "confortante". Difatti abbiamo già detto che per ogni evento è possibile definire un numero arbitrario di variabili casuali. Quando se ne considera una sola si ignorano tutte le altre: se il valore atteso di una variabile (scelta senza nessuna condizione restrittiva) dovesse dipendere da tutte le altre si porrebbero problemi di definizione.

© G. D'Agostini 2001

	t_1						
c_1	0	1	2	3	4	5	
0	0	0	0	0	0	3.1	
1	0	0	0	0	15.6	0	
2	0	0	0	31.3	0	0	$\Lleftarrow f_A(t_1, c_1)$
3	0	0	31.3	0	0	0	
4	0	15.6	0	0	0	0	
5	3.1	0	0	0	0	0	
c_2							
0	0.1	0.5	1.0	1.0	0.5	0.1	
1	0.5	2.4	4.9	4.9	2.4	0.5	
2	1.0	4.9	9.8	9.8	4.9	1.0	$\Lleftarrow f_B(t_1, c_2)$
3	1.0	4.9	9.8	9.8	4.9	1.0	
4	0.5	2.4	4.9	4.9	2.4	0.5	
5	0.1	0.5	1.0	1.0	0.5	0.1	

Tabella 9.2: Distribuzioni congiunta del numero di teste e numero di croci nel lancio di 5 monente (T_1, C_1) confrontata con quella del numero di teste e di croci relative a due diversi lanci di monete. In valori sono in percentuale.

9.5 ⟳ Covarianza e coefficiente di correlazione

Introducendo i concetti di previsione e di incertezza di previsione abbiamo detto che queste rappresentano le informazioni minime nelle quali riassumere lo stato di incertezza quantificato dalla distribuzione di probabilità. Ulteriori informazioni sulla forma sono date dai valori attesi, opportunamente scalati, di potenze più elevate degli scarti (skewness e curtosi, vedi paragrafo 6.13). Purtroppo, nel riassumere una distribuzione di molte variabili in previsione e incertezza di previsione si perde la struttura multidimensionale della distribuzione. Questa perdita di informazione non è dovuta al solo passaggio dalle distribuzioni marginali ai valori attesi, ma è già implicita nel passaggio dalla congiunta alle marginali. (È un po' come la perdita di informazioni spaziali che si ha eseguendo due radiografie, su piani ortogonali, di un corpo.) Considerando, ad esempio, la distribuzione di tabella 9.1, si capisce come sia impossibile di risalire a $f(x, y)$ a partire da $f(x)$ e $f(y)$, mentre è univoco il passaggio inverso.

9.5.1 Variabili correlate e misura della correlazione

Facciamo un esempio numerico per capire meglio il problema. Immaginiamo di dover lanciare 5 monete e di interessarci alle variabili casuali T_1 e C_1, numeri di teste e di croce. Consideriamo anche un altro lancio di 5 monete e chiamiamo T_2 e C_2 i rispettivi esiti. Possiamo costruire le distribuzioni doppie $f_A(t_1, c_1)$ e $f_B(t_1, c_2)$. Sebbene tutte le marginali siano uguali e quindi tutte la variabili abbiano un valore atteso 2.5 e una deviazione standard di 1.1, le due distribuzioni sono completamente diverse (vedi tabella 9.2). Ad esempio $f_A(3, 4) = 0$ (evento impossibile), mentre $f_B(3, 4) = 0.049$.

© G. D'Agostini 2001

La differenza di rilievo fra le due distribuzioni dell'esempio è che mentre nella prima ad ogni valore di T_1 può essere associato un solo valore di C_1, nella seconda la conoscenza di T_1 non modifica lo stato di incertezza rispetto a C_2: T_1 e C_1 sono *dipendenti* (o *correlate*); T_1 e C_2 sono *indipendenti* (o *scorrelate*). Questo fatto si riflette sulle distribuzioni condizionate, le quali differiscono nei due casi. Ad esempio $f(t_1 = 0 \,|\, c_1 = 5) = 1$, mentre $f(t_1 = 0 \,|\, c_2 = 5) = 0.031$, e così via

Dovendo quantificare il grado di correlazione con un solo numero si utilizza il valore atteso del prodotto degli scarti rispetto alle previsioni:

$$\mathrm{E}\left[(X - \mathrm{E}(X))\,(Y - \mathrm{E}(Y))\right] .$$

Esso è chiamato *covarianza* ed è indicato con $\mathrm{Cov}(\cdot,\cdot)$:

$$\mathrm{Cov}(X,Y) \equiv \mathrm{E}\left[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right] . \tag{9.16}$$

Per capire come mai essa possa essere adeguata[3] allo scopo si pensi che

- se in corrispondenza di scarti positivi di X (rispetto alla previsione) si attendono (sempre in senso probabilistico) scarti positivi anche per Y, ne segue che la previsione del prodotto degli scarti è un numero positivo; lo stesso è vero se, in corrispondenza di scarti negativi di X, si attendono scarti negativi anche per Y;
- se invece si prevedono scarti di segno opposto la covarianza è negativa;
- infine, la covarianza è nulla se si attendono scarti mediamente incoerenti.

Ne segue che ci aspettiamo covarianza negativa fra T_1 e C_1 dell'esempio di tabella 9.2, covarianza nulla fra T_1 e C_2.

Per quanto riguarda il valore assoluto della covarianza, esso non indica in maniera intuitiva quanto due variabili sono correlate, in quanto la covarianza non è una grandezza omogenea con le due variabili casuali e dipende anche dall'unità di misura scelta per le variabili. Si preferisce rendere adimensionale la misura di correlazione, dividendo per le scale naturali degli scarti di ciascuna variabile, ovvero le due deviazioni standard. Si definisce allora il *coefficiente di correlazione*, definito come

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\,\sigma(Y)} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} . \tag{9.17}$$

Come detto a proposito della previsione e della sua incertezza, in principio ci potrebbero essere modi alternativi per riassumere con un numero una caratteristica di una distribuzione. Volendo giustificare questa scelta per la misura di correlazione, possiamo fare le seguenti considerazioni:

- La covarianza è formalmente una estensione della definizione operativa della varianza a due variabili e quindi la covarianza di una variabile con sé stessa è uguale alla varianza:

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X) .$$

[3] Per un altro modo di capire come mai la covarianza definita come (9.16) tenda a comparire nella teoria della probabilità si veda la (10.27) a proposito della somma di variabili casuali.

© G. D'Agostini 2001

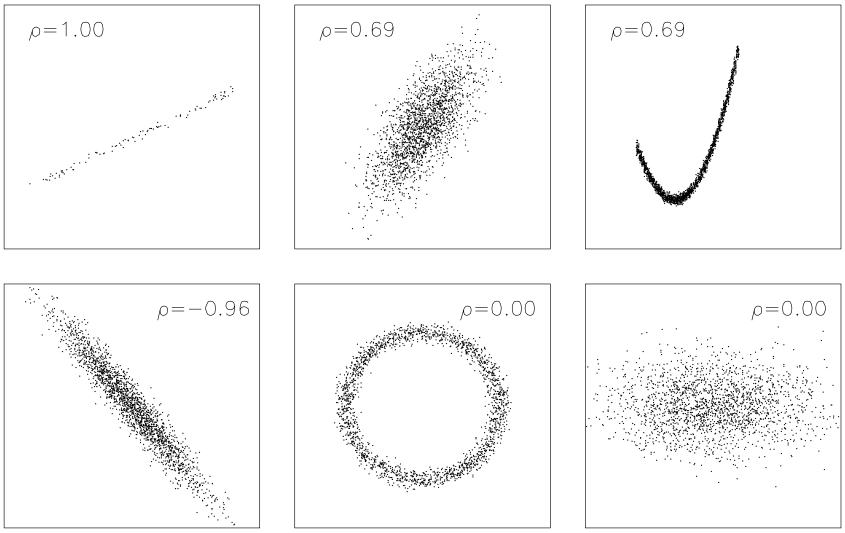

Figura 9.2: Esempi di correlazione fra variabili casuali.

Questa analogia suggerisce di indicare la covarianza con un simbolo compatto simile a quello di varianza:

$$\sigma_{XY} \equiv \mathrm{Cov}(X, Y) \tag{9.18}$$

con la convenzione che $\sigma_{XX} = \sigma_X^2$ (questa notazione sarà molto utile nelle applicazioni).

- Ne segue che il coefficiente di correlazione di una variabile con sé stessa è uguale a 1 ($\rho(X, X) = 1$): si dice che una variabile è correlata al 100 % con sé stessa.

- La covarianza, e quindi il coefficiente di correlazione, si annulla se due variabili sono fra loro indipendenti, come abbiamo giustificato intuitivamente sopra e come vedremo formalmente fra breve.

- Vedremo come la covarianza entra in modo naturale nel calcolo della varianza di una combinazione lineare di variabili casuali (paragrafo 9.5.2).

- Se due variabili sono legate (in modo deterministico) da una relazione lineare, il grado di correlazione misurato da ρ è massimo in modulo ($|\rho| = 1$) e il suo segno dipende dal fatto che una variabile cresca o diminuisca al crescere dell'altra, come intuitivamente comprensibile.

La figura 9.2 mostra alcuni esempi di variabili doppie discrete in cui la $f(x, y)$ è proporzionali all'intensità dei punti. Si faccia attenzione come correlazioni complicate possano dare $\rho = 0$.

© G. D'Agostini 2001

9.5.2 Proprietà formali di covarianza e coefficiente di correlazione

Avendo introdotto il concetto di covarianza, vediamo il modo con cui essa viene calcolata a partire dalla definizione.

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{E}\left[(X-\mathrm{E}(X))\ (Y-\mathrm{E}(Y))\right] \\
& \left[= \iint \left[(x-\mathrm{E}(X))\ (y-\mathrm{E}(Y))\right] f(x,y)\,\mathrm{d}x\mathrm{d}y\right] \\
&= \mathrm{E}\left[X\cdot Y - X\cdot \mathrm{E}(Y) - Y\cdot\mathrm{E}(X) + \mathrm{E}(X)\cdot\mathrm{E}(Y)\right] \\
&= \mathrm{E}(XY) - \mathrm{E}(X)\cdot\mathrm{E}(Y) \qquad (9.19)
\end{aligned}$$

Il secondo passaggio, fra parentesi, mostra il modo di calcolare la covarianza dalla definizione operativa (avendo preso come esempio il caso di variabili continue). In realtà la (9.19) rappresenta il modo più semplice per calcolare la covarianza, in analogia alla formula che esprime la varianza come "media dei quadrati meno il quadrato della media". Infatti, anche la (9.19) può essere letta come "media del prodotto meno il prodotto delle medie".

Per quanto riguarda $\mathrm{E}(XY)$, esso non è in genere risolvibile in termine di altre grandezze note e va calcolato dalla definizione operativa, che, per variabili continue, è

$$\mathrm{E}(X\,Y) = \iint x\,y\,f(x,y)dxdy\,. \qquad (9.20)$$

Esiste un solo caso notevole in cui è possibile evitare di fare l'integrale doppio. Se

$$f(x,y) = f(x)\cdot f(y)\,,$$

ovvero Y e Y sono *fra loro indipendenti*, possiano riscrivere l'integrale doppio come prodotto di due integrali:

$$\begin{aligned}
\mathrm{E}(X\,Y) &= \int x\,f(x)dx\cdot\int y\,f(y)dy \\
&= \mathrm{E}(X)\cdot\mathrm{E}(Y)\,. \qquad (9.21)
\end{aligned}$$

Ne segue, allora, che $\mathrm{Cov}(X,Y) = \mathrm{E}(X)\cdot\mathrm{E}(Y) - \mathrm{E}(X)\cdot\mathrm{E}(Y) = 0$:

> *se due variabili casuali sono indipendenti, la loro covarianza è nulla.*

Si faccia attenzione a questa affermazione. Essa si basa sulla definizione di indipendenza stocastica espressa dalla $f(x,y) = f(x)\,f(y)$ e non significa che non esista alcun legame fra i possibili valori delle due variabili. La covarianza nulla implica soltanto l'assenza di una *correlazione* di tipo *lineare*, come si vedrà fra breve quando vedremo i casi che massimizzano $|\rho|$. Due variabili possono essere fortemente dipendenti pur avendo covarianza nulla. Un caso "clamoroso" è quello di punti nel piano X,Y distribuiti lungo una circonferenza.[4] Se per semplicità poniamo il centro del cerchio nell'origine è facile

[4]Per mostrare come le variabili X e Y del caso del cerchio siano non indipendenti, nonostante la distribuzione congiunta dia covarianza nulla, si pensi alla pdf condizionata $f(y\,|\,x)$, ottenuta "affettando" il "cerchio" per un certo valore di $X = x$. Chiaramente la forma di $f(y\,|\,x)$ dipenderà dal valore di x scelto e, in ogni caso, sarà diversa dalla "proiezione" $f(y)$.

© G. D'Agostini 2001

verificare che sono nulli $\mathrm{E}(X)$, $\mathrm{E}(Y)$ e $\mathrm{E}(X\,Y)$ e quindi anche la covarianza (vedi figura 9.2).

Calcoliamo ora covarianza e coefficiente di correlazione fra due variabili *linearmente dipendenti*, ovvero legate da una relazione del tipo

$$Y = a\,X + b\,.$$

Utilizzando i simboli compatti $\mu_X = \mathrm{E}(X)$ e $\mu_Y = \mathrm{E}(Y)$, abbiamo:

$$\begin{aligned}
\mu_Y &= a\,\mu_X + b && (9.22)\\
\sigma_Y^2 &= a^2\,\sigma_X^2 && (9.23)\\
\mathrm{Cov}(X,Y) &= \mathrm{E}[(X-\mu_X)\,(Y-\mu_Y)] \\
&= \mathrm{E}[(X-\mu_X)\,(a\,X+b-a\,\mu_X-b)] \\
&= \mathrm{E}[a\,(X-\mu_X)\,(X-\mu_X)] \\
&= a\,\sigma_X^2 && (9.24)
\end{aligned}$$

da cui

$$\rho = \frac{a\,\sigma_X^2}{\sqrt{\sigma_X^2\,a^2\,\sigma_X^2}} = \frac{a}{|a|}\,. \tag{9.25}$$

Quindi, in caso di dipendenza lineare, il coefficiente di correlazione vale ± 1 a seconda che il coefficiente angolare a sia positivo o negativo. Si può dimostrare che, in generale,

$$-1 \le \rho \le +1 \tag{9.26}$$

e il grado di correlazione lineare fra due variabili è misurato da $|\rho|$.

Dimostrazione della (9.26): consideriamo la variabile casuale $W = X + \alpha\,Y$. La varianza di W è, per definizione, non negativa per qualsiasi valore di α:

$$\sigma_W^2(\alpha) = \sigma_X^2 + \alpha^2\sigma_Y^2 + 2\,\alpha\,\sigma_{XY} \ge 0\,. \tag{9.27}$$

La dipendenza di $\sigma_W^2(\alpha)$ da α è parabolico. Essendo $\sigma_W^2(\alpha) \ge 0$, questo sarà vero anche quando $\sigma_W^2(\alpha)$ assume il valore minimo, in corrispondenza di

$$\alpha_{min} = -\frac{\sigma_{XY}}{\sigma_Y^2}\,.$$

Sostituendo nella (9.27) si ottiene

$$\frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \le 1\,, \tag{9.28}$$

da cui segue la (9.26). ■

© G. D'Agostini 2001

9.6 ○ Matrice di covarianza e matrice di correlazione

Siamo giunti a sintetizzare l'incertezza su una coppia di variabili con tre valori: due varianze e una covarianza, o, equivalentemente, due deviazioni standard e un coefficiente di correlazione. Nel caso di n variabili (X_1, X_2, $\ldots X_n$) servono

$$N = n + \frac{n\,(n-1)}{2} = \frac{n\,(n+1)}{2} \tag{9.29}$$

informazioni indipendenti.

Talvolta è preferibile utilizzare una rappresentazione compatta di queste grandezze mediante la cosidetta *matrice di correlazione*, una matrice quadrata i cui termini sono dati da

$$V_{ij} = \mathrm{E}\left[(X_i - \mathrm{E}(X_i))(X_j - \mathrm{E}(X_j))\right] \,. \tag{9.30}$$

Infatti, i termini diagonali (quando $i = j$) danno le varianze, mentre gli altri (quando $i \neq j$) danno le covarianze. In forma matriciale, abbiamo:

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Var}(X_1, X_n) \\ \mathrm{Cov}(X_1, X_2) & \sigma_2^2 & \cdots & \mathrm{Var}(X_2, X_n) \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{Var}(X_1, X_n) & \mathrm{Var}(X_2, X_n) & \cdots & \sigma_n^2 \end{pmatrix} \tag{9.31}$$

La matrice è simmetrica, come è facile verificare, e quindi i termini indipendenti sono esattamente $n\,(n+1)/2)$ come richiesto dalla (9.29).

Alternativamente, a volte può essere conveniente presentare l'incertezza dei dati mediante la *matrice di correlazione*, i cui elementi sono dati da

$$\rho(X_i, X_j) = \frac{V_{ij}}{\sigma_i \sigma_j} \,,$$

ovvero

$$\mathbf{R} = \begin{pmatrix} 1 & \rho(X_1, X_2) & \cdots & \rho(X_1, X_n) \\ \rho(X_1, X_2) & 1 & \cdots & \rho(X_2, X_n) \\ \cdots & \cdots & \cdots & \cdots \\ \rho(X_1, X_n) & \rho(X_2, X_n) & \cdots & 1 \end{pmatrix} \tag{9.32}$$

Come si vede, nel passaggio dalla matrice di covarianza alla matrice di correlazione si perdono le informazioni sulle n deviazioni standard (gli elementi della diagonale sono tutti unitari). Quindi, qualora si preferisca questo modo di condensare la distribuzione, le deviazione standard devono essere fornite separatamente.

9.7 ○ Esempi di variabili doppie discrete

Per chiarire alcuni dei concetti appena illustrati e per introduzione altre problematiche che seguiranno, facciamo l'esempio di costruzione di variabili doppie discrete a partire da eventi elementari.

Immaginiamo di dover effettuare una serie di esperimenti che consistono nel lancio di tre monete e nell'estrazione di una carta scelta a caso fra due. Per

© G. D'Agostini 2001

e_i	X	Y	Z	(X, Z)	(Y, Z)	(Ξ, X)	(Ξ, Z)
ATTT	3	0	2	(3,2)	(0,2)	(1,3)	(1,2)
ATTC	2	1	1	(2,1)	(1,1)	(1,2)	(1,1)
ATCT	2	1	0	(2,0)	(1,0)	(1,2)	(1,0)
ATCC	1	2	0	(1,0)	(2,0)	(1,1)	(1,0)
ACTT	2	1	1	(2,1)	(1,1)	(1,2)	(1,1)
ACTC	1	2	0	(1,0)	(2,0)	(1,1)	(1,0)
ACCT	1	2	0	(1,0)	(2,0)	(1,1)	(1,0)
ACCC	0	3	0	(0,0)	(3,0)	(1,0)	(1,0)
RTTT	3	0	2	(3,2)	(0,2)	(0,3)	(0,2)
RTTC	2	1	1	(2,1)	(1,1)	(0,2)	(0,1)
RTCT	2	1	0	(2,0)	(1,0)	(0,2)	(0,0)
RTCC	1	2	0	(1,0)	(2,0)	(0,1)	(0,0)
RCTT	2	1	1	(2,1)	(1,1)	(0,2)	(0,1)
RCTC	1	2	0	(1,0)	(2,0)	(0,1)	(0,0)
RCCT	1	2	0	(1,0)	(2,0)	(0,1)	(0,0)
RCCC	0	3	0	(0,0)	(3,0)	(0,0)	(0,0)

Tabella 9.3: Variabili casuali costruite sugli eventi elementari costituiti dai possibili esiti di tre lanci di monete e dell'estrazione di una carta da gioco: X: numero di teste; Y: numero di croci; Z: numero di teste consecutive; Ξ: numero di assi.

ogni moneta si può verificare Testa o Croce, mentre supponiamo che i valori delle due carte siano Asso (A) e Re (R). Per ogni evento possiamo costruire diverse variabili casuali, per esempio:

- X: numero di teste;
- Y: numero di croci;
- Z: numero di teste consecutive;
- Ξ: numero di assi;

Inoltre possiamo costruire a partire da queste tutte le possibili combinazioni di variabili multiple. Ci limitiamo alle variabili doppie (X, Z), (Y, Z), (Ξ, X) e (Ξ, Z). Riportiamo i risultati nella tabella 9.3

Da questa tabella ci possiamo calcolare tutte le distribuzioni di probabilità di interesse applicando le relazioni viste nei paragrafi precedenti. Cominciamo con le variabili (X, Z)

(x, z)	$f(x, z)$
(0,0)	2/16
(1,0)	6/16
(2,0)	2/16
(2,1)	4/16
(3,2)	2/16

© G. D'Agostini 2001

Da questa possiamo ricavare le distribuzioni marginali di X e Z. Per esempio, per ottenere $f(x)$ bisogna , per ogni valore di X, “integrare” su tutti i valori di Z:

$$f(x) = \sum_{z=-\infty}^{+\infty} f(x,z).$$

x	$f(x)$
0	$2/16 = 1/8$
1	$6/16 = 3/8$
2	$(2/16 + 4/16) = 3/8$
3	$2/16 = 1/8$

z	$f(z)$
0	$(2/16 + 6/16 + 2/16) = 5/8$
1	$4/16 = 2/8$
2	$2/16 = 1/8$

Come si vede dalla tabella, la distribuzione marginale di una certa variabile calcolata dalla formula è esattamente quella che si otterrebbe esaminando la tabella ove sono riportati gli eventi elementari e ignorando tutte le altre variabili. Dalla distribuzione di probabilità di (X, Z) ci possiamo costruire anche le distribuzioni condizionali $f(x \mid z)$. Come detto, di queste distribuzioni ne esiste una per ogni valore di Z. È interessante, in questo semplice caso, confrontare il risultato che si ottiene contando il numero di casi favorevoli e il numero di casi possibili con quello che si otterrebbe applicando le formule delle variabili condizionali viste sopra:

x	$f(x\vert z=0)$	$\frac{f(x,z=0)}{f(z=0)}$
0	$2/10 = 1/5$	$\frac{2/16}{5/8} = 1/5$
1	$6/10 = 3/5$	$\frac{6/16}{5/8} = 3/5$
2	$2/10 = 1/5$	$\frac{2/16}{5/8} = 1/5$

x	$f(x\vert z=1)$	$\frac{f(x,z=1)}{f(z=1)}$
2	1	$\frac{4/16}{2/8} = 1$

x	$f(x\vert z=2)$	$\frac{f(x,z=2)}{f(z=2)}$
3	1	$\frac{2/16}{1/8} = 1$

Queste sono le altre distribuzioni di variabili doppie ottenibili dalla tabella degli eventi elementari[5]:

[5]Si ricorda che la lettera minuscola di Ξ è ξ.

© G. D’Agostini 2001

(y,z)	$f(y,z)$
(0,2)	2/16
(1,0)	2/16
(1,1)	4/16
(2,0)	6/16
(3,0)	2/16

(ξ,x)	$f(\xi,x)$
(0,0)	1/16
(0,1)	3/16
(0,2)	3/16
(0,3)	1/16
(1,0)	1/16
(1,1)	3/16
(1,2)	3/16
(1,3)	1/16

(ξ,z)	$f(\xi,z)$
(0,0)	5/16
(0,1)	2/16
(0,2)	1/16
(1,0)	5/16
(1,1)	2/16
(1,2)	1/16

La distribuzione marginale di Ξ è

ξ	$f(\xi)$
0	1/2
1	1/2

Come ultimo esempio ci costruiamo la distribuzione condizionale di x sotto la condizione $\Xi = 0$.

x	$f(x\|\xi=0)$	$\frac{f(x,\xi=0)}{f(\xi=0)}$
0	1/8	$\frac{1/16}{1/2}=1/8$
1	3/8	$\frac{3/16}{1/2}=3/8$
2	3/8	$\frac{3/16}{1/2}=3/8$
3	1/8	$\frac{1/16}{1/2}=1/8$

Da quest'ultima tabella si vede che che la distribuzione di X sotto la condizione $\Xi = 0$ è uguale a quella marginale di X, contrariamente a quanto accadeva quando la variabile condizionante era Z. Questo corrisponde al fatto che, mentre lo stato di conoscenza del numero di teste consecutive condiziona il grado di fiducia del numero di teste, questo non è influenzato dal sapere se si è verificato un Asso o un Re.

Calcoliamo ora valori attesi di ciascuna variabile:

$$\begin{aligned}
\mathrm{E}(X) &= 0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+2\cdot\frac{3}{8}+3\cdot\frac{1}{8}=\frac{12}{8}\\
\mathrm{E}(Z) &= 0\cdot\frac{5}{8}+1\cdot\frac{2}{8}+2\cdot\frac{1}{8}=\frac{4}{8}\\
\mathrm{E}(Y) &= 0\cdot\frac{1}{8}+1\cdot\frac{3}{8}+2\cdot\frac{3}{8}+3\cdot\frac{1}{8}=\frac{12}{8}\\
\mathrm{E}(\Xi) &= 0\cdot\frac{8}{16}+1\cdot\frac{8}{16}=\frac{1}{2}
\end{aligned} \tag{9.33}$$

© G. D'Agostini 2001

Per quanto riguarda la deviazione standard, svolgiamo in dettaglio il calcolo di $\sigma(X)$, lasciando gli altri conti come esercizio:

$$\begin{aligned}
\mathrm{E}(X^2) &= 0^2\cdot\frac{1}{8}+1^2\cdot\frac{3}{8}+2^2\cdot\frac{3}{8}+3^2\cdot\frac{1}{8}=\frac{24}{8}\\
\mathrm{Var}(X) &= \frac{24}{8}-\left(\frac{12}{8}\right)^2=\frac{3}{4}\\
\sigma(X) &= \sqrt{\frac{3}{4}}=0.87\\
\sigma(Y) &= \sqrt{\frac{3}{4}}=0.87\\
\sigma(Z) &= \sqrt{\frac{1}{2}}=0.71\\
\sigma(\Xi) &= \sqrt{\frac{1}{4}}=0.5
\end{aligned} \tag{9.34}$$

Per quanto riguarda le covarianze e coefficienti di correlazione abbiamo:

1. (X, Z):

$$\mathrm{E}(X\,Z) = 0\cdot\frac{5}{8}+2\cdot\frac{2}{8}+6\cdot\frac{1}{8}=\frac{10}{8}$$
$$\mathrm{Cov}(X,Z) = \frac{10}{8}-\frac{12}{8}\cdot\frac{4}{8}=\frac{1}{2} \tag{9.35}$$
$$\rho(X,Z) = 0.81 \quad > \mathbf{0} \tag{9.36}$$

2. (Y, Z):

$$\mathrm{E}(Y\,Z) = 0\cdot\frac{6}{8}+1\cdot\frac{2}{8}=\frac{2}{8}$$
$$\mathrm{Cov}(Y,Z) = \frac{2}{8}-\frac{12}{8}\frac{4}{8}=-\frac{1}{2} \tag{9.37}$$
$$\rho(Y,Z) = -0.81 \quad < \mathbf{0} \tag{9.38}$$

3. (Ξ, Z):

$$\mathrm{E}(\Xi\,Z) = 0\cdot\frac{13}{16}+1\cdot\frac{2}{16}+2\cdot\frac{1}{16}=\frac{4}{16}=\frac{1}{4}$$
$$\mathrm{Cov}(\Xi,Z) = \frac{1}{4}-\frac{1}{2}\frac{1}{2} \quad =0 \tag{9.39}$$
$$\rho(\Xi,Z) = \mathbf{0} \tag{9.40}$$

Il coefficiente di correlazione fra il numero di teste consecutive e il numero di teste è positivo; fra il numero di teste consecutive e il numero di croci è negativo; fra il numero di teste consecutive e il numero di assi è nullo. Questo è in accordo con quanto ci si aspetta da una grandezza atta a quantificare il grado

© G. D'Agostini 2001

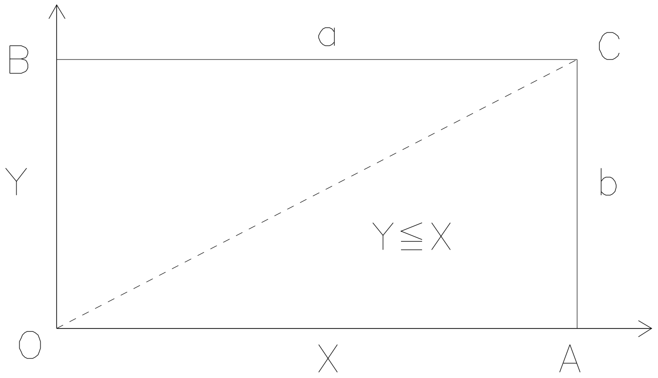

Figura 9.3: Distribuzione bidimensionale uniforme

di correlazione. È infatti chiaro che, sotto l'ipotesi che un evento contenga un alto numero di teste, cresce la probabilità di trovare due teste consecutive. L'effetto opposto avrebbe l'ipotesi che si verichino una grande numero di croci. L'ipotesi dell'uscita o meno di un asso non altera invece le probabilità degli esiti dei lanci delle monete, e viceversa.

9.8 ↻ Esempi di distribuzione bidimensionale continua

Entrambi gli esempi che facciamo hanno una funzione densità di probabilità congiunta costante. Nel primo caso le variabili sono definite in un rettangolo con i lati paralleli agli assi, nel secondo sono definite nel triangolo ottenuto dividendo il rettangolo precedente lungo la diagonale. Questi esempi permettono di prendere familiarità con alcuni aspetti delle distribuzioni multiple, pur mantenendo al livello minimo le difficoltà di calcolo,

9.8.1 Distribuzione uniforme in un rettangolo

Come primo e semplice esempio di distribuzione multipla di variabili continue consideriamo due variabili X e Y negli intervalli $[x_1, x_2]$ e $[y_1, y_2]$ e tali che la loro densità di probabilità congiunta $f(x, y)$ sia costante nella porzione di piano definita da tali limiti. Per semplificare i conti, ma senza perdere in generalità, facciamo coincidere gli estremi inferiori x_1 e y_1 con l'origine degli assi, e chiamiamo a e b le lunghezze dei due segmenti (vedi figura 9.3). La condizione di normalizzazione permette di ricavare il valore della costante:

$$\int_0^a \int_0^b f(x, y)\, \mathrm{d}x\mathrm{d}y = \int_0^a \int_0^b k\, \mathrm{d}x\mathrm{d}y = ab \tag{9.41}$$

da cui segue

$$f(x, y) = \frac{1}{ab} \qquad \left\{ \begin{array}{l} 0 \le x \le a \\ 0 \le y \le b \end{array} \right. \tag{9.42}$$

© G. D'Agostini 2001

Possiamo quindi calcolare le distribuzioni marginali delle due variabili:

$$f(x) = \int_0^b f(x,y)\,\mathrm{d}y = \frac{1}{a}\,; \tag{9.43}$$

$$f(y) = \int_0^a f(x,y)\,\mathrm{d}x = \frac{1}{b}\,. \tag{9.44}$$

Come ovvio, la densità è uniforme in ciascuna delle due distribuzioni. Le medie e le varianze valgono quindi: $\mathrm{E}(X) = a/2$; $\mathrm{E}(Y) = b/2$; $\sigma_X^2 = a^2/12$; $\sigma_Y^2 = b^2/12$.

Per il calcolo della covarianza serve il valore atteso del prodotto:

$$\begin{aligned} \mathrm{E}(X\,Y) &= \int_0^a \int_0^b x\,y\,f(x,y)\,\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{ab}\int_0^a \int_0^b x\,y\,\mathrm{d}x\mathrm{d}y = \frac{ab}{4}\,, \end{aligned} \tag{9.45}$$

da cui

$$\mathrm{Cov}(X,Y) = \mathrm{E}(X\,Y) - \mathrm{E}(X)\,\mathrm{E}(Y) = \frac{a\,b}{4} - \frac{a}{2}\frac{b}{2} = 0\,. \tag{9.46}$$

e quindi $\rho(X,Y) = 0$. In effetti, anche se si venisse a sapere che X vale $a/10$, $a/2$ o a, questa informazione non cambia le nostre aspettative sui possibili valori di Y.

Il calcolo della densità di probabilità di x, condizionata dall'occorrenza di un valore di $Y = y$, è altresì semplice:

$$f(x\,|\,y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{1}{a\,b}}{\frac{1}{b}} = \frac{1}{a}\,. \tag{9.47}$$

Essa è identica alla densità marginale e quindi Y e Y sono indipendenti.

9.8.2 Distribuzione uniforme in un triangolo

Se le variabili casuali possono assumere valori, sempre in modo uniforme, all'interno del triangolo $AOCA$ di figura 9.3, la funzione densità di probabilità congiunta è data da

$$f(x,y) = \frac{2}{ab} \qquad \left\{ \begin{array}{l} 0 \le x \le a \\ 0 \le y \le x \end{array} \right. \tag{9.48}$$

È possibile ottenere le distribuzioni marginali e condizionate senza fare alcun conto:

- $f(x)$ è ottenuta integrando $f(x,y)$ lungo y e quindi è proporzionale all'altezza del triangolo rettangolo $OXC'O$, ove C' indica il punto lungo la diagonale $\overline{OC}$ in corrispondenza di X. la distribuzione marginale di X è quindi una triangolare asimmetrica (vedi paragrafo 8.4) avente $x_\circ = a$, $\Delta_+ = 0$ e $\Delta_- = a$;

© G. D'Agostini 2001

- per simmetria $f(y)$ è data da una triangolare avente $x_\circ = 0$, $\Delta_+ = b$ e $\Delta_- = 0$;

- fissando $X = x$ la distribuzione di Y è una uniforme fra 0 e $y_{max} = (b/a)x.$;

- analogamente, fissando $Y = y$ la distribuzione di X è uniforme fra $x_{min} = (a/b)y$ e a.

Come esercizio facciamo i conti:

$$
\begin{aligned}
f(x) &= \int f(x,y)\,\mathrm{d}y = \int_0^{\frac{b}{a}x} \frac{2}{a\,b}\,\mathrm{d}y = \frac{2x}{a^2} && (0 \le x \le a) \\
f(y) &= \int f(x,y)\,\mathrm{d}x = \int_{\frac{a}{b}y}^{a} \frac{2}{a\,b}\,\mathrm{d}x = \frac{2}{b^2}(b-y) && (0 \le y \le b) \\
f(x\,|\,y) &= \frac{f(x,y)}{f(y)} = \frac{1}{a}\,\frac{b}{b-y} && (\frac{a}{b}y \le x \le a) \\
f(y\,|\,x) &= \frac{f(x,y)}{f(x)} = \frac{a}{b\,x} && (0 \le y \le \frac{b}{a}x)
\end{aligned}
$$

Si ricordi che le nelle funzioni densità di probabilità condizionate il condizionante ha il ruolo di parametro. A differenza del caso precedente, le funzioni le condizionate e le marginali sono diverse: sapere che X vale 0, $a/2$ o a cambia lo stato di incertezza su Y. Ad esempio, la previsione (± incertezza standard) vale nei tre casi: 0 ± 0 (certezza!), $(0.25 \pm 0.14)a$ e $(0.50 \pm 0.29)a$.

9.9 ❄ Distribuzione multinomiale

Dopo aver introdotto i concetti di base delle distribuzioni multivariate e alcuni esempi elementari, mostriamo ora una una distribuzione di variabile discreta che descrive i possibili istogrammi che possono essere osservati effettuando un certo esperimento. Questo argomento era già stato introdotto nel paragrafo 7.14. In tale occasione ci eravamo interessati soltanto della previsione della distribuzione statistica, ma era stato tralasciato il problema di calcolare la probabilità di ciascuna delle possibili configurazioni.

Cominciamo con un esempio numerico. Immaginiamo di ripetere 20 volte un certo esperimento, in ciascuno crediamo che possa uscire, con uguale probabilità, un numero intero i compreso fra 0 e 10. Indichiamo con X_i il numero di occorrenze dell'esito i-mo. Riassumiamo le cose che già sappiamo fare:

- ciascuna variabile X_i è distribuita secondo una binomiale di $n = 20$ e $p = 1/11$;

- prevediamo una distribuzione statistica con $p\,n \pm \sqrt{p\,q\,n}$ occorrenze per ciascun esito; (*** **att** ***)

- prevediamo che le frequenze relative siano $p \pm \sqrt{p\,q}/\sqrt{n}$.

© G. D'Agostini 2001

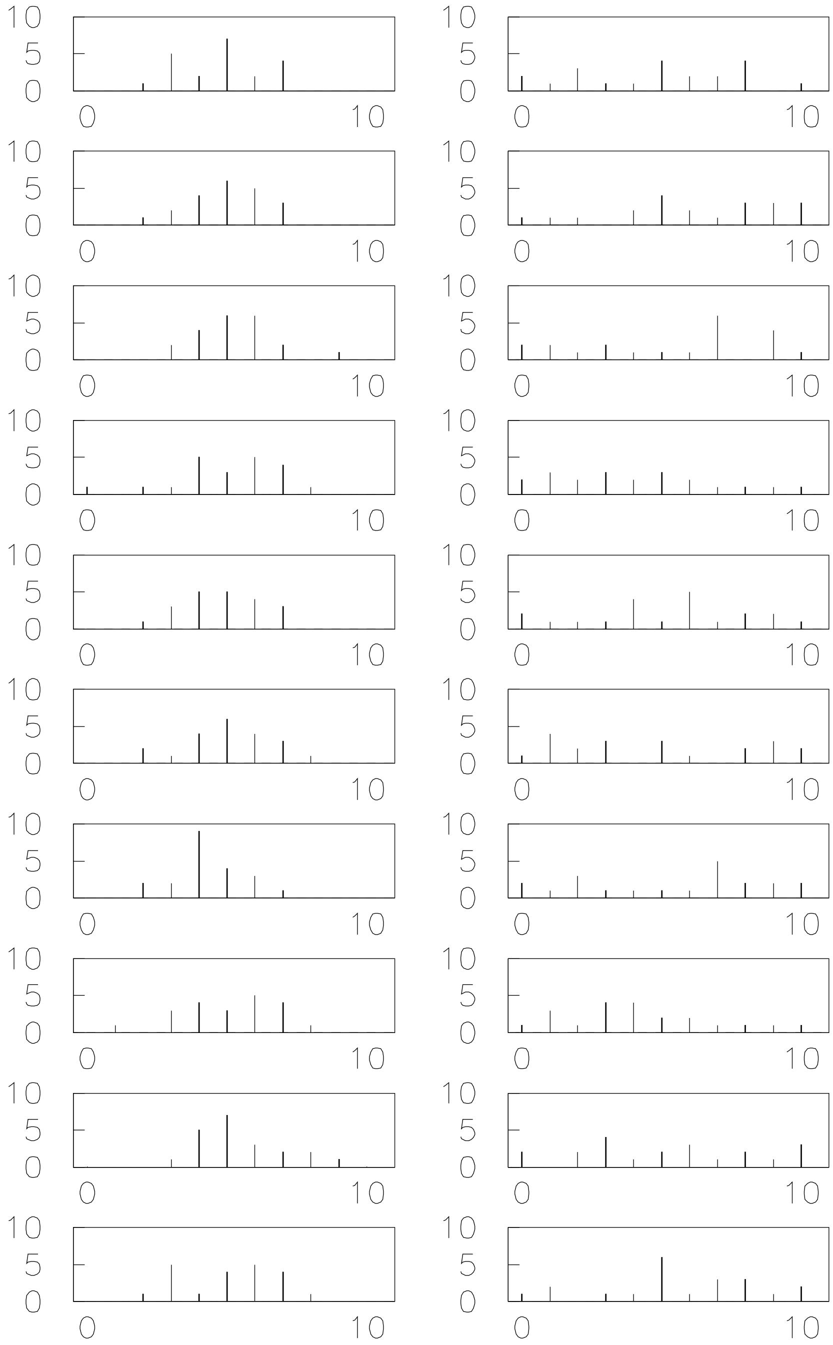

Figura 9.4: Esempi di eventi multinomiali generati casualmente. Tutte le distribuzioni hanno il parametro n pari a 20, $m = 11$, $x_1 = 0$ e $x_m = 10$. Negli istogrammi a destra p_i è costante e pari $1/11$, mentre in quelli a sinistra vale invece $p_i = f(x_i \,|\, \mathcal{B}_{0.5,10})$.

Interessiamoci ora delle possibili configurazioni di occorrenze che è possibile osservare. La figura 9.4 (parte a destra) ne mostra qualcuna (istogramma ottenuto da simulazione al calcolatore). Cominciamo con dei casi particolari e poi passeremo alla trattazione generale.

- Può capitare 20 volte di seguito lo stesso valore di i, ad esempio $i = 0$, dando luogo a $X_\circ = 20$ e $X_1 = X_2 = \cdots = X_{10} = 0$, con probabilità 1.5×10^{-21} per ciascuna possibilità. (La probabilità che si verifichi uno qualsiasi di questo tipo di eventi è 1.6×10^{-20}.)

- Può capitare che esca la metà delle volte un valore, l'altra metà un altro e nessuna volta gli altri 9. Concentriamoci su $X_\circ = 10$, $X_1 = 10$ e gli altri zero. Questi valori possono essere ottenuti da $\binom{20}{10} = 184756$ possibili sequenze, ciascuna avente probabilità 1.5×10^{-21}. Ne seque che la probabilità dell'evento $X_\circ = 10$, $X_1 = 10$ è uguale a 2.8×10^{-16}. (Per calcolare la probabilità di osservare una qualsiasi equipartizione dei 20 successi in due classi bisogna considerare il numero combinazioni di 11 esiti presi 2 a 2, pari a 55, ottenendo 1.5×10^{-14}.)

- Consideriamo ora un caso in cui i risultati sono "più equilibrati", ad esempio i primi 9 esiti che si verificano 2 volte e gli ultimi due una sola volta: $X_0 = x_1 = \cdots = X_8 = 2$; $X_9 = X_{10} = 1$. Questo sembra essere, intuitivamente, un risultato molto più probabile dei due precedenti. Calcoliamo il numero di sequenze (tutte di probabilità 1.5×10^{-21}):

  – ci sono $\binom{20}{2}$ combinazioni che danno $i = 0$;
  – per ciascuna di esse ce ne sono $\binom{20-2}{2} = \binom{18}{2}$ che danno $i = 1$ e così via;
  – otteniamo un totale di
  $$\binom{20}{2}\binom{18}{2}\cdots\binom{1}{1}$$
  combinazioni, ovvero:
  $$\frac{20!}{18!\,2!}\,\frac{18!}{16!\,2!}\,\frac{16!}{14!\,2!}\ldots 1 = \frac{20!}{(2!)^9} = 4.8 \times 10^{15}\,.$$

  Questa configurazione ha quindi una probabilità di 7.1×10^{-6}, enormemente più alta delle altre, anche se di per sé ancora molto piccola. Se poi consideriamo che ci sono 55 configurazioni distinte in cui 9 esiti possono uscire 2 volte e gli altri 1 volta, otteniamo una probabilità di circa lo 0.4 per mille di osservare una qualsiasi di queste configurazioni "equilibrate".

Passiamo ora al caso generale di m possibili modalitè per l'esito di ciascuna prova e sia p_i la probabilità che assegniamo a ciscuna modalità. Naturalmente deva valere la condizione
$$\sum_i^m p_i = 1\,.$$

© G. D'Agostini 2001

Dovendo effettuare n prove nelle stesse condizioni possiamo osservare X_i volte ciascun esito, con

$$\sum_i^n X_i = n\,.$$

Le n prove possono dar luogo a $n!$ sequenze possibili, in genere non tutte equiprobabili (a meno che le p_i siano tutte uguali). Ciascuna sequenza che produce la configurazione di variabili casuali $X_1 = x_1$, $X_2 = x_2$, $\ldots X_m = x_m$ ha probabilità

$$p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m}\,. \tag{9.49}$$

Il numero delle sequenze che produce la stessa configurazione $\{x_1, x_2, \ldots, x_m\}$ è dato dal coefficiente multinomiale

$$\frac{n!}{x_1!\, x_2! \cdots x_m!}\,. \tag{9.50}$$

Infatti, si hanno $\binom{n}{x_1}$ combinazioni per la prima modalità; per ciascuna di esse ce ne sono $\binom{n-x_1}{x_2}$ la seconda, e così via. Quindi il numero di combinazioni è

$$\frac{n!}{(n-x_1)!\, x_1!} \cdot \frac{(n-x_1)!}{(n-x_1-x_2)!\, x_2!} \cdot \frac{(n-x_1-x_2)!}{(n-x_1-x_2-x_3)!\, x_3!} \cdot \ldots\,.$$

Semplificando otteniamo la (9.50). Moltiplicando la probabilità (9.49) della sequenza per il numero di sequanze abbiamo finalmente la *distribuzione multinomiale*

$$f(\underline{x} \,|\, \mathcal{M}_{n\underline{p}}^{m}) = \frac{n!}{x_1!\, x_2! \cdots x_m!}\, p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m}\,, \tag{9.51}$$

dove con $\underline{x}$ è stato indicato l'insieme dei valori $\{x_1, x_2, \ldots, x_m\}$ e con $\underline{p}$ l'insieme delle probabilità $\{p_1, p_2, \ldots, p_m\}$. Quando $m = 2$ si riottiene la binomiale.

Il calcolo del valore atteso e della varianza di ciascuna variabile X_i è immediato, in quanto è sufficiente pensare la distribuzione marginale di ogni variabile pari ad una distribuzione binomiale di parametro p_i. Ne segue

$$\mathrm{E}(X_i) = n\, p_i \tag{9.52}$$

$$\mathrm{Var}(X_i) = n\, p_i\, (1-p_i)\,. \tag{9.53}$$

Per quanto riguarda covarianza e coefficiente di correlazione, invece del conto esatto utilizziamo un metodo euristico (più importante del conto esatto per l'impostazione di questo corso).

Riprendiamo la funzione di probabilità della distribuzione binomiale

$$f(x \,|\, \mathcal{B}_{np}) = \frac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x}$$

ed esplicitiamo il fatto ci troviamo di fronte a due classi di eventi (favorevoli e sfavorevoli) scrivendo

$$f(x_1, x_2 \,|\, \mathcal{B}_{np}) = \frac{n!}{x_1!\, x_2!}\, p_1^{x_1}\, p_2^{x_2}\,, \tag{9.54}$$

© G. D'Agostini 2001

dove è stata operata la seguente trasformazione di simboli:

$$\begin{aligned} x &\rightarrow x_1\,; \\ n-x &\rightarrow x_2\,; \\ p &\rightarrow p_1\,; \\ 1-p &\rightarrow p_2\,. \end{aligned}$$

Il fatto che adesso $f(x_1, x_2)$ dipenda apparentemente da due variabili non è in contraddizione con il fatto che la formula originale delle binomiale fosse funzione di una sola variabile. Infatti x_1 e x_2 sono linermente (anti-)correlate in quanto devono soddisfare la condizione $x_1 + x_2 = n$. Quindi $f(x_1, x_2)$ dipende in realtà soltanto da una variabile e il coefficiente di correlazione fra le due variabili vale $\rho(X_1, X_2) = -1$, come è intuitivo pensare e come risulta dalla (9.25)

Anche nel caso generale della multinomiale le variabili X_i sono fra loro correlate, in quanto vale la relazione $\sum_{i=1}^{m} x_i = n$. Ma il grado di correlazione è, per $m > 2$, minore che nella binomiale. Infatti, mentre nel caso $m = 2$ il non verificarsi di un evento nella classe 1 implica il suo verificarsi nella classe 2, nel caso di molte classi si può al più affermare che esso si sia verificato in una delle restanti $m-1$ classi, ciascuna con probabilità proporzionale a p_i. Quindi il grado di correlazione diminuisce in valore assoluto ("si diluisce") al crescere del numero di classi e, per m abbastanza grande, la multinomiale può essere essere vista come la probabilità congiunta di tante binomiali indipendenti.

Abbiamo detto che nel caso binomiale $\rho(X_1, X_2)$ vale -1. Ne segue che

$$\text{Cov}(X_1, X_2) = \rho(X_1, X_2)\,\sigma_1\,\sigma_2 = -n\,p_1\,p_2\,, \tag{9.55}$$

in quanto $\sigma_1 = \sigma_2 = \sqrt{n\,p_1\,p_2}$. In effetti si può verificare che la (9.55) è valida anche nel caso generale. Si ottiene quindi:

$$\begin{aligned} \text{Cov}(X_i, X_j) &= -n\,p_i\,p_j \qquad (9.56) \\ \rho(X_i, X_j) &= \frac{-n\,p_i\,p_j}{\sqrt{n\,p_i\,(1-p_i)\,n\,p_j\,(1-p_j)}} \qquad (9.57) \\ &= -\sqrt{\frac{p_i\,p_j}{(1-p_i)\,(1-p_j)}}\,. \qquad (9.58) \end{aligned}$$

Come previsto qualitativamente, la correlazione è trascurabile se le probabilità delle due classi sono abbastanza piccole[6].

L'importanza della distribuzione multinomiale risiede nel fatto che essa descrive il numero di eventi di un istogramma o di un diagramma a barre, indipendentemente dalla distribuzione di probabilità che segue la variabile casuale associata alla grandezza di cui si costruisce l'istogramma o il diagramma a barre in questione. In Fig 9.4 sono riportati degli esempi di eventi casuali generati secondo due distribuzioni multinomiali di $n = 20$, in un caso (grafici a destra) con p_i tutte uguali e nell'altro con p_i generato secondo una binomiale:

[6]È da notare che il caso in cui la formula non funziona ($p_i = 1$) è quello in cui un solo una sola modalità è certa e le altre sono impossibili e quindi non ha più senso parlare di distribuzione di probabilità

© G. D'Agostini 2001

$p_i = f(x_i \,|\, \mathcal{B}_{0.5,10})$. In questo caso il numero di eventi per ciascun x_i è distribuito secondo una binomiale di parametri p_i ed $n = 20$, e quindi ha un valore medio $20p_i$. Quindi le realizzazioni della multinomiale dell'esempio fluttuano intorno ad un andamento medio che ha la stessa forma della distribuzione di p_i, riscalato ("normalizzato") di un fattore 20. Si confrontino questi grafici con quello della distribuzione $\mathcal{B}_{0.5,10}$ di Fig. 7.1.

9.10 ❄ Distribuzione normale bivariata

Una funzione di distribuzione di variabili multiple particolarmente importante per le applicazione è quella in cui tutte le distribuzioni marginali sono normali. Consideriamo per ora il caso di due sole variabili, rimandando nei prossimi paragrafi il caso generale.

La situazione più semplice si verifica se le due variabili casuali sono *indipendenti*. Nel tale caso la densità congiunta è data semplicemente dal prodotto delle densità marginali. Chiamando X e Y le variabili e μ_x, μ_y, σ_x e σ_y i parametri delle gaussiane otteniamo

$$f(x,y) \;=\; \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right].$$

Il caso in cui le variabili sono correlate è decisamente più complicato. Rinunciamo ad una trattazione rigorosa e cerchiamo di capire la forma della distribuzione ragionando sulla dipendenza della densità condizionata $f_Y(y\,|\,x)$ dai parametri delle gaussiane e dal coefficiente di correlazione ρ (nel seguito indichiamo $\rho_{X,Y}$ semplicemente come ρ).

- Nel limite $\rho \to 0$ si deve riottenere la distribuzione marginale $\mathcal{N}(\mu_y, \sigma_y)$.
- Se ρ è diverso da zero la $f_Y(y\,|\,x)$ è distribuita intorno ad un valore medio diverso da μ_y e avente una diversa varianza. In particolare, chiamando $\mu_{y\,|\,x}$ e $\sigma_{y\,|\,x}$ i parametri della distribuzione condizionata:
  - $\mu_{y\,|\,x} = \mu_y$ se $\rho = 0$
  - $\mu_{y\,|\,x} > \mu_y$ se $\rho > 0$ e $x > \mu_x$, oppure se $\rho < 0$ e $x < \mu_x$;
  - $\mu_{y\,|\,x} < \mu_y$ se $\rho < 0$ e $x > \mu_x$, oppure se $\rho > 0$ e $x < \mu_x$;
  - l'effetto di spostamento da μ_y deve dipendere, oltre che dal segno, anche dall'entità della correlazione;
  - esso deve dipendere dai rapporti fra le larghezze delle due gaussiane: se, ad esempio, σ_y è maggiore di σ_x, una piccola deviazione di x da μ_x produrrà una grande deviazione di $\mu_{y\,|\,x}$ da μ_y.
  - anche la deviazione standard della distribuzione condizionata $\sigma_{y\,|\,x}$ deve dipendere da ρ, nel senso che si deve riottenere σ_y per variabili scorrelate, mentre essa si deve annullare nel caso di correlazione lineare ($|\rho| = 1$). In quest'ultimo caso infatti ogni valore di X determina univocamente un valore di Y. Inoltre questo effetto di

strizzamento della distribuzione non deve dipendere dal segno di ρ. La forma esatta della dipendenza funzionale è meno intuitiva, ma il risultato esatto è in accordo con le aspettative.

Queste semplici considerazioni sono in accordo con i risultati esatti dati dalle seguenti relazioni:

$$\begin{aligned} \mu_{y\,|\,x} &= \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x) \\ \sigma_{y\,|\,x} &= \sigma_y\sqrt{1-\rho^2}\,. \end{aligned}$$

- Assumiamo inoltre che anche la distribuzione condizionata sia normale (fatta eccezione del caso degenere quando $\rho = 0$ e $\sigma_{y\,|\,x} = 0$).

Mettendo insieme i vari ingredienti otteniamo:

$$f(y\,|\,x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{\left(y-\left[\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]\right)^2}{2\sigma_y^2(1-\rho^2)}\right]$$

Utilizzando la formula della densità condizionata si giunge finalmente alla distribuzione congiunta $f(x,y)$:

$$\begin{aligned} f(x,y) &= f(y\,|\,x)f(x) \\ &= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{\left(y-\left[\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]\right)^2}{2\sigma_y^2(1-\rho^2)}\right]\cdot \\ &\quad \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right] \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{Q^2}{2}\right], \end{aligned} \tag{9.59}$$

dove con Q^2 si è indicata la forma quadratica che compare nell'argomento dell'esponenziale, che dopo le opportune semplificazioni, ha la seguente forma:

$$Q^2 = \frac{1}{1-\rho^2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right] \tag{9.60}$$

È facile convincersi che, per definizione, Q^2 è una quantità non negativa, la quale si annulla nel solo punto $x = \mu_x$ e $y = \mu_y$, dove ha chiaramente un minimo.

La probabilità congiunta $f(x,y)$ descritta dalla (9.59) è nota come la *normale bivariata*, o *binormale* e la possiamo indicare con $\mathcal{N}_2(\mu_x, \sigma_x, \mu_y, \sigma_y, \rho)$.

Si può verificare facilmente che, nel caso di correlazione nulla, si riottiene la densità congiunta ottenuta precedentemente dal prodotto delle due gaussiane.

Terminiamo elencando alcune sue proprietà, alcune delle quali già incontrate, e facendo riferimento alle Figg. 9.5 e 9.6.

© G. D'Agostini 2001

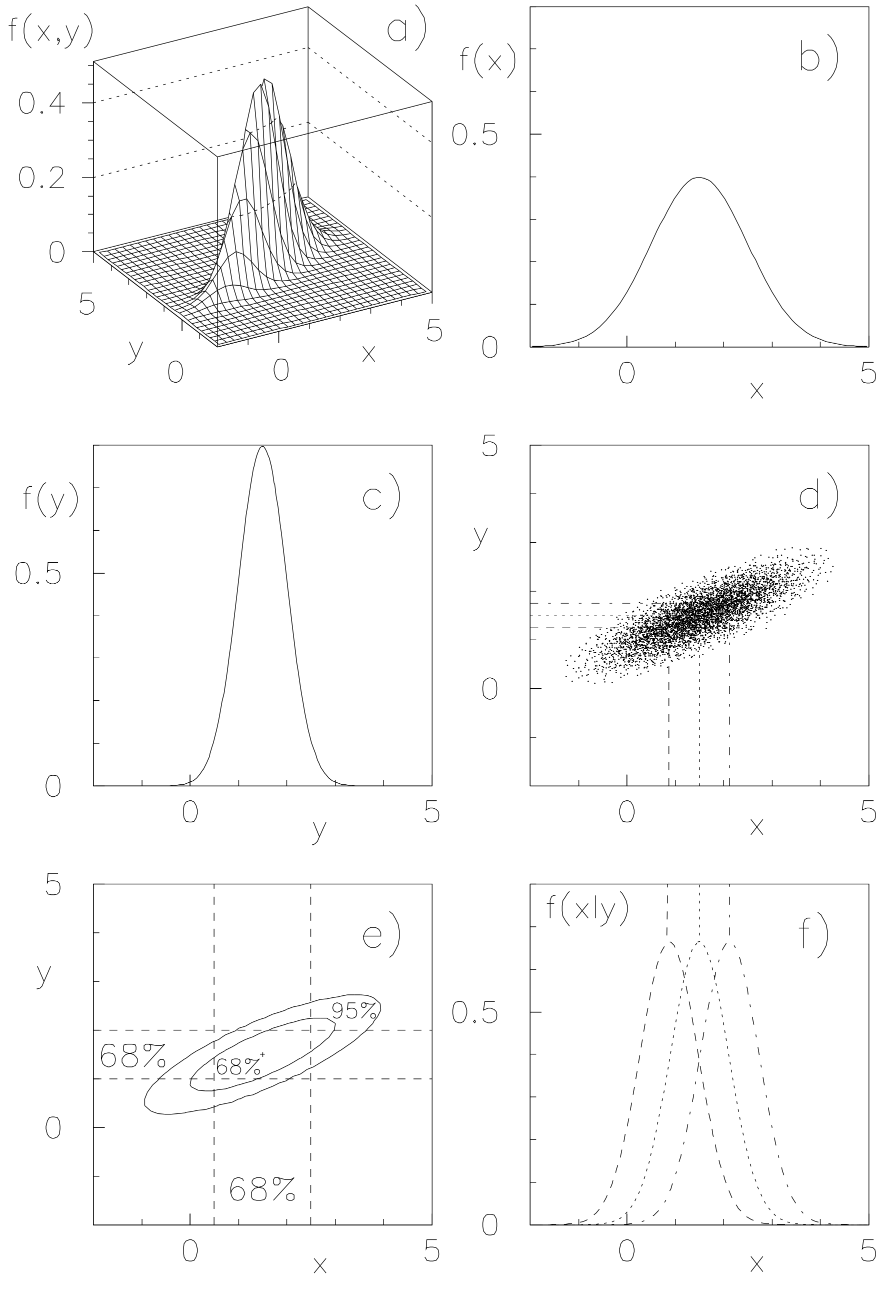

Figura 9.5: Esempio di distribuzione normale bivariata.

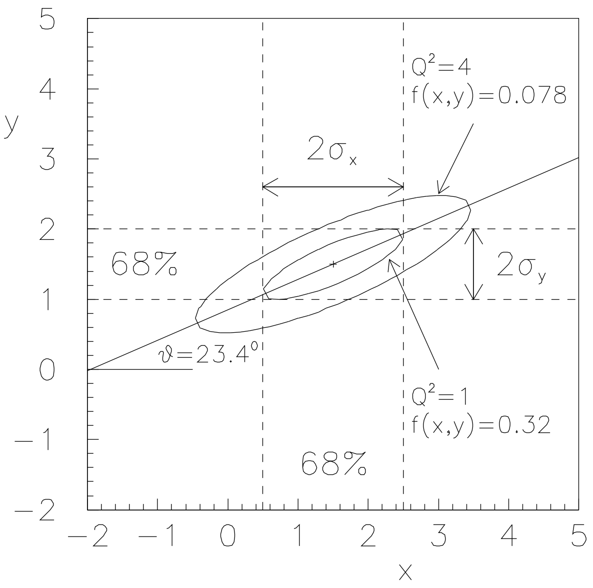

Figura 9.6: Distribuzione normale bivariata: ellissi di equidensità e parametri delle distribuzioni marginali. I valori numerici di $f(x,y)$ e di θ dipendono dai parametri dell'esempio.

- Rappresentata nello spazio, essa ha una tipica forma a campana, più o meno schiacciata e più o meno inclinata, a seconda dei valori dei parametri.

- Il valore massimo di $f(x,y)$ è situato in corrispondenza di $X = \mu_x$ e $Y = \mu_y$ e vale
$$f(\mu_x, \mu_y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\,. \tag{9.61}$$
Esso dipende, oltre che dall'inverso delle deviazioni standard, anche dal coeffiente di correlazione.

- La distribuzione marginale di ciascuna delle variabili è normale come si può verificare integrando[7] la (9.59).

- I valori attesi delle variabili X e Y sono μ_x e μ_y e le loro deviazioni standard σ_x e σ_y.

- La covarianza fra le variabili vale $\text{Cov}(X,Y) = \rho\,\sigma_x\sigma_y$.

- I valori di X e Y per i quali la densità di probabilità è costante sono quelli per i quali è costante Q^2. Si riconosce facilmente come tali punti, per

[7] Per i calcoli si usi l'integrale indefinito:
$$\int_{-\infty}^{+\infty} \exp\left[bx - \frac{x^2}{a^2}\right] dx = \sqrt{a^2\pi}\exp\left[\frac{a^2b^2}{4}\right].$$

© G. D'Agostini 2001

k	1 (1.65)	2 (1.96)	3 (2.57)
Q^2	1 (2.71)	4.0 (3.84)	9 (6.63)

Tabella 9.4: Valori di Q^2 per calcolare le ellissi contenute esattamente entro i rettangoli di ampiezza $2k\sigma_x$ e $2k\sigma_y$ intorno ai valori medi.

una dato valore di $f(x, y)$, definiscono un'ellisse nel piano (X, Y), chiamata *ellisse di equidensità*; di particolare interesse sono due famiglie di ellissi:

- ellissi esattamente contenute entro rettangoli di lati $2k\sigma_x$ e $2k\sigma_y$ centrati intorno a μ_x e μ_y (Fig 9.6);
- ellissi tali da avere un certo livello probabilità P che entrambe le variabili cadano all'interno di esse (Fig 9.5.E).

Le prime ellissi possono essere ricavate direttamente da considerazioni geometriche, mentre per le altre è necessario calcolare l'integrale $\int_{x,y\in ellisse} f(x, y)dxdy = P$. È interessante sapere che queste ultime ellissi dipendono da Q^2 in modo indipendente dagli altri parametri, mentre invece i valori di $f(x, y)$ sono irrilevati in quanto dipendono da σ_x, σ_y e ρ. Le tabelle 9.4 e 9.5 forniscono alcuni valori di interesse di Q^2.

- L'angolo dell'asse maggiore dell'ellisse rispetto all'asse delle ascisse è dato dalla relazione:

$$\tan 2\theta = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2} \tag{9.62}$$

La figura 9.5 mostra un esempio di distribuzione bivariata normale di parametri $\mu_x = 1.5$, $\mu_y = 1.5$, $\sigma_x = 1.0$, $\sigma_y = 0.5$ e $\rho = 0.8$. La distribuzione congiunta è rappresentata sia come superficie nello spazio (9.5.a), che densità di punti proporzionali alla p.d.f. (9.5.d). Le funzioni di densità di probabilità marginali sono mostrate in 9.5.b e 9.5.c. Tre esempi di densità condizionate cono mostrati in 9.5.d e 9.5.f, per valori di Y pari rispettivamente a 1.25, 1.5 e 1.75. La Fig. 9.5.e mostra infine le curve di equidensità tali che la probabilità congiunta di trovare i valori di x e di y al loro interno sia pari al 68 e al 95 %. Per confronto sono anche riportati gli intervalli di probabilità al 68 % delle distribuzioni marginali.

P	50 %	68.3 %	95 %	99 %
Q^2	1.39	2.30	5.99	9.21

Tabella 9.5: Valori di Q^2 per calcolare le ellissi che racchiudono le variabili X e Y con probabilità P.

© G. D'Agostini 2001

9.11 ❄ Caso generale di distribuzione multivariata

È semplice dimostrare che Q^2 può essere riscritto come il prodotto di un vettore riga per un vettore colonna

$$Q^2 = \frac{1}{1-\rho^2}\left(\left(\frac{x-\mu_x}{\sigma_x^2} - \frac{\rho(y-\mu_y)}{\sigma_x\sigma_x}\right) \quad \left(-\frac{\rho(x-\mu_x)}{\sigma_x\sigma_y} + \frac{y-\mu_y}{\sigma_y^2}\right) \right) \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} .$$

Il vettore riga può essere ulteriormente scomposto come il prodotto di un vettore riga per una matrice 2×2, ottenendo così la seguente decomposizione:

$$\begin{pmatrix} x-\mu_x & y-\mu_y \end{pmatrix} \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{pmatrix} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} . \quad (9.63)$$

Chiamando $\underline{\Delta}$ il vettore colonna della differenza fra le variabili casuali e il loro valore atteso, $\underline{\Delta^T}$ il vettore riga, in quanto trasposto di $\underline{\Delta}$, e $\mathbf{A}$ la matrice, Q^2 acquista la seguente forma:

$$Q^2 = \underline{\Delta}^T \mathbf{A} \underline{\Delta} ,$$

con

$$\mathbf{A} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{pmatrix} . \quad (9.64)$$

È possibile riscrivere la matrice A in una forma più semplice se si passa alla sua inversa[8], ottenendo

$$\mathbf{\Sigma} = \mathbf{A}^{-1} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} . \quad (9.65)$$

La matrice $\mathbf{\Sigma}$ ha quindi un significato più immediato di $\mathbf{A}$, in quanto i suoi elementi sono le varianze e le covarianze delle variabili casuali (si ricordi che $\rho\sigma_x\sigma_y = \text{Cov}(X,Y)$). Utilizzando la forma compatta di scrivere le varianze e covarianze per un numero arbitrario di variabili casuali (vedi par sec:covarianza) si riconosce che

$$\Sigma_{i,j} = \sigma_{i,j} . \quad (9.66)$$

Inoltre, si può dimostrare che l'espressione di Q^2 scritta nella forma

$$Q^2 = \underline{\Delta}^T \mathbf{\Sigma}^{-1} \underline{\Delta} \quad (9.67)$$

può essere estesa ad un numero qualsiasi di variabili. Questo ci permette di ottenere la formula generale della multinomiale, se si riesce a scrivere in un modo compatto il denominatore del termine di fronte all'esponenziale. Per derivarne intuitivamente la forma si pensi possono fare le seguenti considerazioni.

[8] Si ricorda che $A_{11}^{-1} = A_{22}/|\mathbf{A}|$, $A_{22}^{-1} = A_{11}/|\mathbf{A}|$, $A_{12}^{-1} = -A_{12}/|\mathbf{A}|$ e $A_{21}^{-1} = -A_{12}/|\mathbf{A}|$, dove $|\mathbf{A}|$ è il determinante della matrice, il quale vale nel nostro caso $(1-\rho^2)/(\sigma_x^2\sigma_y^2)$.

© G. D'Agostini 2001

- Il fattore 2π, differisce dall'originario $\sqrt{2\pi}$ della gaussiana in quanto abbiamo due variabili; ne segue che la sua naturale estensione ad n variabili è:
$$(2\pi)^{n/2}\,.$$
- Il denominatore contiene il prodotto delle deviazioni standard per questioni di dimensioni. Infatti l'elemento di probabilità infinitesimo $f(x,y)\cdot \mathrm{d}x\mathrm{d}y$ deve essere adimensionale per il suo significato di probabilità e quindi la densità deve avere le dimensioni inverse di $\mathrm{d}x\mathrm{d}y$. Il termine $\sqrt{1-\rho^2}$ è dovuto invece all'effetto di schiacciamento della forma a campana della distribuzione binormale, dovuto alla correllazione fra le variabili. Se il numero di variabili aumenta esso è destinato a complicarsi. Per formulare $\sigma_x\sigma_y\sqrt{1-\rho^2}$ in forma compatta estendibile ad un numero qualsiasi di variabili riscriviamola come
$$\sqrt{\sigma_x^2\sigma_y^2(1-\rho^2)} = \sqrt{\sigma_x^2\sigma_y^2 - (\rho\sigma_x\sigma_y)^2}\,.$$
Si riconosce quindi come questo termine sia pari alla radice quadrata del determinante di $\mathbf{\Sigma}$, che indicheremo con $|\mathbf{\Sigma}|$.

Poiché si può dimostrare che questa formula è valida anche nel caso generale, abbiamo finalmente ottenuto l'espressione della densità di probabilità di una distribuzione multinomiale di n variabili comunque correlate fra di loro:

$$f(\underline{x}\,|\,\mathcal{N}(\underline{\mu},\mathbf{\Sigma})) = (2\pi)^{-\frac{n}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\underline{\Delta}^T\mathbf{\Sigma}^{-1}\underline{\Delta}\right], \qquad (9.68)$$

ovvero

$$f(\underline{x}\,|\,\mathcal{N}(\underline{\mu},\mathbf{\Sigma})) = (2\pi)^{-\frac{n}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})^T\mathbf{\Sigma}^{-1}(\underline{x}-\underline{\mu})\right], \quad (9.69)$$

dove, ripetiamo ancora una volta, $\underline{x}$ rappresenta il vettore aleatorio $\{x_1, x_2, \ldots, x_n\}$. La notazione $\mathcal{N}(\underline{\mu},\mathbf{\Sigma})$ è simile a quella della semplice gaussiana, con le dovute differenze:

- al posto di una sola media μ ne abbiamo n, compattate in un vettore $\underline{\mu} = \{\mu_1, \mu_2, \ldots, \mu_n\}$;
- la deviazione standard è sostituita da una matrice di varianze e covarianze Σ, chiamata *matrice di covarianza*. Poiché vale che $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$, la matrice è simmetrica e i parametri indipendenti sono $n(n+1)/2$, ovvero le n varianze e gli $n(n-1)/2$ coefficienti di correlazione.

La densità di probabilità a molte dimensioni non è rappresentabile graficamente. Si ricordi comunque che

- la distribuzione marginale di ciascuna delle variabili X_i è una normale $\mathcal{N}(\mu_i, \sigma_i)$;
- la distribuzione congiunta di due variabili qualsiasi X_i e X_j è una normale bivariata $\mathcal{N}_2(\mu_i, \sigma_i, \mu_j, \sigma_j, \rho_{ij})$ descritta dall'equazione (9.59).

© G. D'Agostini 2001

Si può verificare facilmente che, se i coefficienti di correlazione fra le variabili sono tutti nulli e quindi nella matrice di covarianza sono diversi da zero soltanto i termini diagonali, la (9.69) si riduce al prodotto di n normali univariate:

$$f(\underline{x}\,|\,\mathcal{N}(\underline{\mu},\underline{\sigma})) = \frac{1}{(2\pi)^{\frac{n}{2}}\Pi_i\sigma_i}\exp\left[-\frac{1}{2}\sum_i\frac{(x_i-\mu_i)^2}{\sigma_i^2}\right]. \tag{9.70}$$

Derivate di Q^2 rispetto alle variabili casuali

Come ultimo punto su questo argomento osserviamo come le derivate seconde della forma quadratica che compare nell'esponenziale, effettuate rispetto alle variabili casuale, siano legate agli elementi dell'inversa della matrice di covarianza:

$$\frac{1}{2}\frac{\partial^2 Q^2}{\partial x_i \partial x_j} = -\frac{\partial^2}{\partial x_i \partial x_j}\ln f(\underline{x}\,|\,\mathcal{N}(\underline{\mu},\underline{\sigma})) = A_{ij} \quad = \quad \left(\Sigma^{-1}\right)_{ij}\,. \tag{9.71}$$

Nel caso unidimensianale questa proprietà si riduce alla banale constatazione che

$$\frac{\partial^2}{\partial x^2}\ln f(x\,|\,\mathcal{N}(\mu,\sigma)) = -\frac{1}{\sigma^2}\,. \tag{9.72}$$

9.12 ↻ Distribuzioni statistiche multivariate

Il concetto di distribuzione multivariata si estende anche al caso delle distribuzioni statistiche, prestando attenzione a tenere ben divisi i concetti, che per comodità ripetiamo: nella distribuzione di probabilità si associa un grado di fiducia ad ogni vettore che descrive un possibile esito; nel caso di distribuzione statistica si associa ad esso un peso statistico dato dalla frequenza relativa con cui i vettori di esiti si sono verificati. Nel caso discreto le due distribuzioni sono quindi formalmente equivalente:

$$\begin{array}{rl} \text{variabili casuali:} & \{x_i, y_i\} \leftrightarrow f(x_i, y_i) \\ \text{variabili statistiche:} & \{x_i, y_i\} \leftrightarrow w_i \end{array}$$

Nel caso di distribuzioni statistiche non ha molto senso parlare di variabili continue, in quanto i numeri che si registrano hanno sempre una risoluzione finita. Ne segue che gli analoghi degli elementi infinitesimi di probabilità sono i pesi statistici di una celletta di dimensione finita:

$$\begin{array}{rl} \text{variabili casuali:} & \{x, y\} \leftrightarrow f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ \text{variabili statistiche:} & \{(x_i, \Delta X), (y_i, \Delta Y)\} \leftrightarrow w_i\,. \end{array}$$

Anche nel caso di distribuzioni statistiche si può fare uso di covarianza e deviazione standard per riassumere la correlazione fra variabili statistiche, tenendo conto delle solite note sulle diversità di interpretazione delle grandezze

© G. D'Agostini 2001

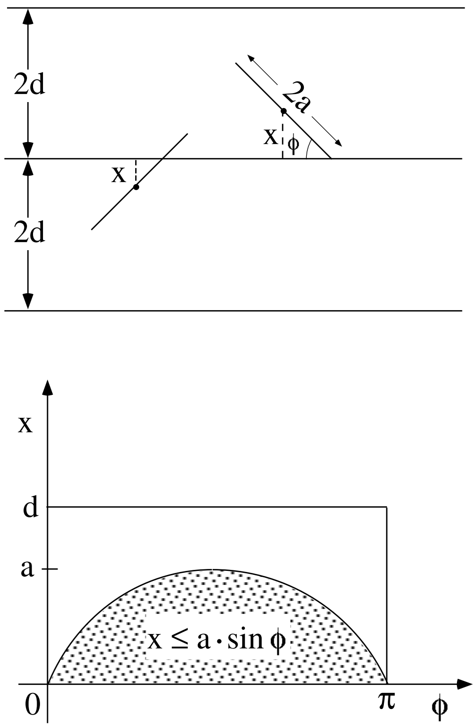

Figura 9.7: Ago di Buffon

di nome analogo. Se abbiamo n coppie di valori, ciascuna di peso statistico w_i, abbiamo

$$\mathrm{Cov}(x, y) = \sum_i w_i \, (x_i - \overline{x}) \, (y_i - \overline{y}) \,, \tag{9.73}$$

che nel caso di pesi tutti uguali abbiamo

$$\mathrm{Cov}(x, y) = \frac{1}{n} \sum_i (x_i - \overline{x}) \, (y_i - \overline{y}) \,. \tag{9.74}$$

9.13 varie

© G. D'Agostini 2001

Capitolo 10

Funzioni di variabili casuali e teoremi limite

10.1 Propagazione delle incertezze

La condizione di incertezza su alcune grandezze si riflette in genere su ogni altra grandezza che sia funzione di esse. Ad esempio, immaginiamo di essere interessati all'area (A) e al perimetro (p) di un rettangolo (idealizzazione del piano di un tavolo), di cui sono stati misurati i due lati (a e b). L'inevitabile incertezza sul valore di a e di b (a cui andrebbe aggiunta quella legata al fatto che il quadrilatero potrebbe essere non perfettamente retto) si riflette su A e su p. Il termine con cui questo processo è noto è *propagazione delle incertezze*.

Come sappiamo, i gradi di fiducia sui possibili valori di a e di b sono espressi da $f(a)$ e $f(b)$. Ma, nel caso generale, i valori[1] di a e di b non sono indipendenti, come può succedere nel caso che i lati siano stati misurati con lo stesso strumento, non perfettamente calibrato[2]. Quindi, in genere bisognerà considerare la funzione congiunta $f(a, b)$. Per poter quantificare nel modo più generale l'incertezza su area e perimetro, bisogna imparare a valutare $f(A)$ e $f(p)$ partendo da $f(a, b)$ (o da $f(a)$ e $f(b)$ nel caso di indipendenza). In realtà, anche in questo caso, la soluzione più generale al problema si ottiene mediante il calcolo di $f(A, p)$. Infatti ci aspettiamo che A e p abbiano un certo grado di correlazione, in quanto sono calcolate dalle stesse informazioni di partenza. Riepilogando, il problema consiste nel valutare $f(A, p)$ a partire da $f(a, b)$:

$$f(a, b) \Rightarrow f(A, p)\,. \tag{10.1}$$

Come vedremo, il problema generale può diventare abbastanza complesso dal punto di vista del calcolo. Ci accontenteremo di calcolare soltanto previsione e incertezza di previsione delle varie grandezze e, qualora esistano correlazioni, del loro coefficiente di correlazione. Nella maggior parte dei casi pratici (e

[1] Si noti l'uso della stessa lettera sia per il nome della variabile che per i possibili valori. Quando si passa alle applicazioni può essere più importante usare gli stessi simboli sia per le grandezze che per le possibili realizzazioni. Quindi si raccomanda di abituarsi ad una certa flessibilità.

[2] Ritorneremo in dettaglio su tale effetto (vedi paragrafo 14.1) , ma si capisce che, ad esempio, se la misura sottostima a sottostimerà anche b.

"tranquilli") queste approssimazioni sono più che ragionevole e vedremo come esse ci permetteranno, sotto certe ipotesi spesso soddisfatte, di effettuare affermazioni probabilistiche sui valori delle grandezza. È comunque importante ricordarsi che, all'occorrenza, bisogna affrontare il problema nel modo più rigoroso, descritto da una generalizzazione della (10.1), che scriviamo con

$$f(x_1, x_2, \ldots, x_n) \xrightarrow[Y_j = Y_j(X_1, X_2, \ldots, X_n)]{} f(y_1, y_2, \ldots, y_m)\,, \tag{10.2}$$

ove con $Y_j = Y_j(\cdots)$ è indicata la j-ma funzione che lega le n X_i alle m Y_j. La soluzione minimale che affronteremo nei dettagli sarà invece il solo caso di propagazione di previsione ed incertezza di combinazioni lineari di variabili:

$$\left\{\begin{array}{l} \mathrm{E}(X_i) \\ \sigma(X_i) \\ \rho(X_i, X_{i'}) \end{array}\right. \xrightarrow[Y_j = c_{j0} + c_{j1}X_1 + c_{j2}X_2 + \cdots + c_{jn}X_n]{} \left\{\begin{array}{l} \mathrm{E}(Y_j) \\ \sigma(Y_j) \\ \rho(Y_j, Y_{j'}) \end{array}\right. \tag{10.3}$$

A questo caso ci ridurremo, previe *linearizzazioni*, nel caso di funzioni qualsiasi (e nei limiti in cui le linearizzazioni siano ragionevoli).

I prossimi paragrafi, dedicati alla valutazione della distribuzione di probabilità di funzioni di variabili casuali, possono essere saltati da chi non è interessato a tale argomento. In tale caso si vada direttamente al paragrafo 10.6.

10.2 Soluzione generale per variabili discrete

Il caso di variabili discrete, benché di scarso interesse per le applicazioni di questo corso, è molto semplice, almeno dal punto di vista concettuale.

10.2.1 Regola generale

Consideriamo due variabili casuali di partenza, X e Y, e soltanto una variabile, Z, funzione di esse:

$$Z = g(X, Y)\,.$$

Cominciamo dal considerare il caso in cui Z assuma un certo valore z per la sola combinazione della variabili $X = x$ e $Y = y$. Chiaramente, la probabilità di $Z = z$ sarà uguale alla probabilità che si verifichino esattamente questi valori di X e di Y:

$$f(z) = P(Z = z) = P(X = x, Y = y) = f(x, y) \qquad \text{con } g(x, y) = z \tag{10.4}$$

© G. D'Agostini 2001

Quando ci sono invece più coppie che contribuiscono al valore $Z = z$, bisogna sommare su tutte le loro probabilità:[3]

$$f(z) = \sum_{\substack{x,y \\ g(x,y)=z}} f(x,y) \tag{10.5}$$

Come esempio, si veda la somma degli esiti nel lancio di due dadi, già mostrato nel paragrafo 6.4 (vedi anche figura 6.2). Si noti come l'alta probabilità per valori centrali della somma sia dovuto soltanto al fatto che questi valori possono essere causati da più combinazioni di quelli estremi.

La (10.5) si estende facilmente al caso di più funzioni. Ad esempi, considerando anche la variabile $V = f(X,Y)$, la congiunta è data da:

$$f(z,v) = \sum_{\substack{x,y \\ g(x,y)=z \\ h(x,y)=v}} f(x,y) \tag{10.6}$$

Anche se l'impostazione della (10.6) è semplice, il calcolo che permette di ottenere delle formule analitiche può essere molto complicato, ben al di là dello scopo di questo corso. Vediamo soltanto nei prossimi paragrafi alcuni sottocasi molto istruttivi.

10.2.2 ❄ Convoluzione di due funzioni di probabilità

Consideriamo la variabile Z, somma di due variabili indipendenti X e Y. Applicando la regola generale otteniamo che

$$f(z) = \sum_{\substack{x,y \\ x+y=z}} f(x,y) \tag{10.7}$$

$$= \sum_x f(x, z-y)\,. \tag{10.8}$$

Nel caso di variabili indipendenti la (10.8) si riduce a

$$= \sum_x f(x) f(z-y)\,, \tag{10.9}$$

che rappresenta il caso discreto della *convoluzione* di due funzioni di probabilità indipendenti. La figura 10.1 mostra la distribuzione della somma di due variabili casuali uniformi associate agli esiti del lancio di due dadi. È anche mostrata la somma di tre dadi, la quale può essere interpretata come la somma di due dadi più un terzo dado. Si noti l'alta probabilità dei valori centrali e la bassa probabilità di quelli laterali.

[3] Dal punto di vista formale, la (10.5) equivale, per variabili a valori interi (ricordiamo che discreto non implica necessariamente intero) a

$$f(z) = \sum_{x,y} \delta_{z,g(x,y)} f(x,y)\,,$$

ove $\delta_{i,j}$ è la delta di Kronecker che vale 1 se $i = j$ e zero altrimenti.

© G. D'Agostini 2001

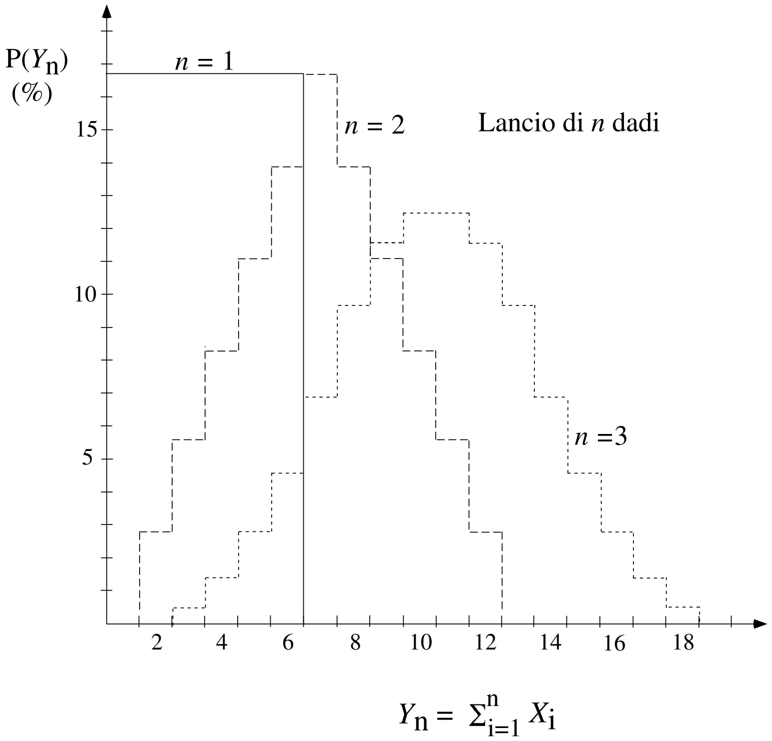

Figura 10.1: Distribuzione della somma dei risultati ottenuti dal lancio di n dadi. La concentrazione della probabilità al centro della distribuzione è dovuta all'elevato numero di combinazioni i quali producono risultati intermedi e giustifica qualitativamente il teorema del limite centrale.

Come esempio di applicazione della (10.9) consideriamo la somma di due variabili ciascuna delle quali è una distribuzione di Poisson, con parametri λ_1 e λ_2. Abbiamo già detto, basandoci su argomenti legati all'interpretazione fisica del processo di Poisson, che la distribuzione della somma è ancora una poissoniana di parametro $\lambda_1 + \lambda_2$. Utilizzando la (10.9) otteniamo:

$$f(z) = \sum_{x=0}^{z} f(x \,|\, \mathcal{P}_{\lambda_1}) f(z-x \,|\, \mathcal{P}_{\lambda_2}) \,,$$

ove la sommatoria va da 0 a z in quanto $y = z - x$ non può assumere valori negativi. Esplicitando le funzioni si ottiene:

$$\begin{aligned} f(z) &= \sum_{x=0}^{z} \frac{e^{-\lambda_1}\lambda^x}{x!} \frac{e^{-\lambda_2}\lambda_2^{z-x}}{(z-x)!z!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^z}{z!} \sum_{x=0}^{z} \frac{z!}{(z-x)!z!} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{z-x} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^z}{z!} = f(z \,|\, \mathcal{P}_{\lambda_1+\lambda_2}) \,, \end{aligned} \tag{10.10}$$

in cui l'ultimo passaggio è dovuto al fatto che la sommatoria è uguale allo sviluppo del binomio

$$\left(\frac{\lambda_1}{\lambda_1+\lambda_2}+\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^z = 1^z = 1\,.$$

10.2.3 Trasformazione di una variabile distribuita uniformemente

Consideriamo il caso di una variabile (Y) funzione di un'altra (X) che sia distribuita uniformemente fra 1 e $n = 100$ ($X \sim \mathcal{K}_{1,n}$). Esso ci aiuterà a capire il comportamento della trasformazione delle variabili continue. Scegliamo le tre diverse funzioni

$$\begin{aligned} Y_1 &= 100\,X \\ Y_2 &= X^2 \\ Y_3 &= 1000\,\sqrt{X}\,. \end{aligned}$$

Alcuni valori delle variabili sono mostrati nella prima parte della tabella 10.1.

Il calcolo delle funzioni di probabilità è banale. Essa vale 1/100 per ciascuno dei valori considerati. La differenza fra Y_1 e le altre due variabili è che queste non hanno valori equidistanziati, pur essendo la distribuzione uniforme nel senso di $f(\cdot) = k$:

- la variabile Y_2, pari a X^2, ha valori fitti all'inizio che poi si diradano[4] al crescere di i;
- la variabile Y_3, proporzionale a $\sqrt{X}$, ha invece l'andamento opposto: i valori sono molto distanziati all'inizio e diventano più fitti alla fine;
- come consequenza di questi diversi comportamenti la distribuzione di Y_2 ha il baricentro più basso di Y_1, Y_3 più alto;
- la diversa "densità di punti per intervallo della variabile" è quantificata nella tabella dalla probabilità che si verifichi la variabile in un intervallo di ampiezza 1000 (semplicemente proporzionale al numero di valori in tale intervallo, in quanto la probabilità di ciascuno di essi è costante);
- sebbene, a diversità delle funzioni continue, non si possa parlare di una funzione continua densità, possiamo considerare ugualmente la densità media di probabilità:

$$\begin{aligned} \text{densità media di prob. di Y} &= \frac{\text{Prob in }\Delta y}{\Delta Y} \\ &\propto \frac{\#\text{ valori}}{\Delta Y} \propto \frac{\Delta X}{\Delta Y}\frac{1}{\frac{\Delta Y}{\Delta X}}\,; \end{aligned}$$

  questo mostra che, quando i valori di Y sono abbastanza fitti entro l'unità di scala variabilità della grandezza ($\approx \sigma$)

$$\frac{\text{Prob in }\Delta y}{\Delta Y} \to \frac{1}{\frac{\mathrm{d}Y}{\mathrm{d}X}}\,.$$

[4] Si noti che la differenza fra il quadrato di un intero i e quello dell'intero precedente è pari a $2\,i - 1$.

© G. D'Agostini 2001

i	$Y_1 = 100\,X$	$Y_2 = X^2$	$Y_3 = 1000\,\sqrt{X}$
1	100.	1.	1000.
2	200.	4.	1414.
3	300.	9.	1732.
4	400.	16.	2000.
5	500.	25.	2236.
6	600.	36.	2449.
7	700.	49.	2646.
8	800.	64.	2828.
...	...	...	...
93	9300.	8649.	9644.
94	9400.	8836.	9695.
95	9500.	9025.	9747.
96	9600.	9216.	9798.
97	9700.	9409.	9849.
98	9800.	9604.	9899.
99	9900.	9801.	9950.
100	10000.	10000.	10000.
$P(0 \le Y_j \le 1000)$	10	31	1
$P(1000 < Y_j \le 2000)$	10	13	3
$P(2000 < Y_j \le 3000)$	10	10	5
$P(3000 < Y_j \le 4000)$	10	9	7
$P(4000 < Y_j \le 5000)$	10	7	9
$P(5000 < Y_j \le 6000)$	10	7	11
$P(6000 < Y_j \le 7000)$	10	6	13
$P(7000 < Y_j \le 8000)$	10	6	15
$P(8000 < Y_j \le 9000)$	10	5	17
$P(9000 < Y_j \le 10000)$	10	6	19
$\mathrm{E}(Y_j)$	5050	3384	6715
$\sigma(Y_j)$	2887	3009	2327
$v(Y_j)$	0.57	0.89	0.35
$Y_j[\mathrm{E}(X)]$	5050	2550	7106

Tabella 10.1: Tre trasformazioni della distribuzione uniforme discreta $\mathcal{K}_{1,100}$: $Y_1 = 100\,X$; $Y_2 = X^2$; $Y_3 = 1000\,\sqrt{X}$. Sono mostrati: i primi e gli ultimi valori assunti dalle variabili; la probabilità (in %) che le variabili siano comprese nei dieci intervalli di ampiezza 1000; il valore atteso e la deviazione standard.

© G. D'Agostini 2001

10.3 ❄Soluzione generale per variabili continue

Anche nel caso di variabili continue il problema è *in principio* semplice[5], ma i conti possono diventare immediatamente complicati. Per analogia con la (10.6), si può dire che l'elemento infinitesimo di probabilità $f(z,v)dzdv$ è dato dalla somma di tutti elementi infinitesimi $f(x,y)dxdy$, tali che $g(x,y) = z$ e $h(x,y) = v$.

Consideriamo nel seguito soltanto il caso di cambiamento di una singola variabile e di somma di due variabili indipendenti.

10.3.1 Cambiamento di variabile

Prendiamo la variabile X, con funzione densità di probabilità $f(x)$, e consideriamo una funzione crescente $Y = g(X)$. Applicando quanto appena espresso a parole per il caso generale, abbiamo che

$$P(y \leq Y \leq y + dy) = P(x \leq X \leq x + dx)\,, \qquad \text{con } g(x) = y\,,$$

ovvero

$$f_Y(y)\,\mathrm{d}y = f_X(x)\,\mathrm{d}x\,, \qquad \text{con } g(x) = y\,,$$

in cui è stato esplicitato il fatto (generalmente sottinteso in questo testo) che $f_Y(\cdot)$ è una funzione di y mentre $f_X(\cdot)$ è funzione di x. Ne segue che

$$f_Y(y) = \frac{f_X(x)}{\mathrm{d}y/\mathrm{d}x} = \frac{f_X(x)}{g'(x)}\,, \tag{10.11}$$

dove con $g'(x)$ è stata indicata la derivata di $Y = g(X)$ calcolata in corrispondenza di Y tale che $g(x) = y$, ovvero di $x = g^{-1}(y)$, ove - chiariamo - $(g^{-1}(\cdot)$ sta per la funzione inversa di $g(\cdot)$. Quindi il modo più corretto per esprimere la (10.11), funzione solo di y, è

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}\,. \tag{10.12}$$

Nel caso che $g(X)$ sia decrescente si segue un ragionamento analogo, ma nel risultato finale cambia il segno al secondo membro. Quindi la formula generale, per funzioni sempre crescenti o sempre decrescenti è:

$$f_Y(y) = \frac{f_X(x)}{|\mathrm{d}y/\mathrm{d}x|} = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}\,. \tag{10.13}$$

[5]Un modo concettualmente semplice, analogo al caso generale delle variabili discrete, consiste nel "sommare" tutti gli infiniti elementi di probabilità che danno luogo allo stesso valore della variabile finale. Quindi, l'estensione al continuo della forma della nota 10.2.1 è

$$f(z) = \int \delta(z - g(x,y) \cdot f(x,y)\,\mathrm{d}x\mathrm{d}y\,,$$

ove $\delta()$ sta per la δ di Dirac. Chi è familiare con questa "funzione speciale" può trovare di agevole impostare i conti in questo modo e, facendo uso delle sue proprietà, ricavare facilmente le formule che incontreremo nel seguito.

© G. D'Agostini 2001

$f(x)$	$g(X) \to Y$	$g^{-1}(y)$	$f(y)$	dominio di y
1	$\frac{X}{2}+\frac{1}{4}$	$2\,y-\frac{1}{2}$	2	$0.25 \le Y \le 0.75$
1	$\sqrt{X}$	y^2	$2y$	$0 \le Y \le 1$
1	X^2	$\sqrt{y}$	$\frac{1}{2\sqrt{y}}$	$0 \le Y \le 1$
1	X^4	$y^{1/4}$	$\frac{1}{4\,y^{3/4}}$	$0 \le Y \le 1$

Tabella 10.2: Funzioni densità di probabilità ottenute da diverse trasformazioni di una variabile distribuita uniformemente fra 0 e 1.
.

Ad esempio, se la trasformazione è del tipo $k\,X$, si ottiene

$$f_Y(y) = \frac{1}{|k|} f_X(\frac{Y}{k})\,.$$

Vediamo in dettaglio cosa succede quando si applica una trasformazione ad una variabile distribuita uniformemente.

Trasformazioni di una distribuzione uniforme

Il caso della distribuzione di partenza uniforme è interessante in quanto esso mostra chiaramente la distorsione delle funzioni densità di probabilità operate dal cambiamento di metrica. Questo si capisce meglio se si osserva attentamente la figura 10.2, in cui le diverse densità dei puntini danno un'idea della funzione densità di probabilità. È anche interessante confrontare quanto ottenuto con la discussione fatta a proposito delle trasformazioni uniformi variabili discrete (vedi anche tabella 10.1).

Le funzioni $g(X)$ utilizzate e la funzione densità di probabilità delle trasformate sono date in tabella 10.2. I calcoli vengono lasciati per esercizio.

Applicazioni alle simulazioni di variabili casuali

È interessante il caso in cui la funzione di trasformazione $g(X)$ abbia la stessa espressione della funzione di distribuzione $F(x)$, ovvero:

$$g(X) = F(X)\,.$$

(Si noti l'uso di maiuscole e minuscole per distinguere due funzioni matematematicamente uguali in questo caso, ma concettualmente diverse.) Facendo uso della 10.12, in quanto $F(x)$ è crescente, otteniamo:

$$f_y(y) = \frac{f_X(x)}{\frac{\mathrm{d}F(x)}{\mathrm{d}x}} = \frac{f_X(x)}{f_X(x)} = 1\,. \tag{10.14}$$

© G. D'Agostini 2001

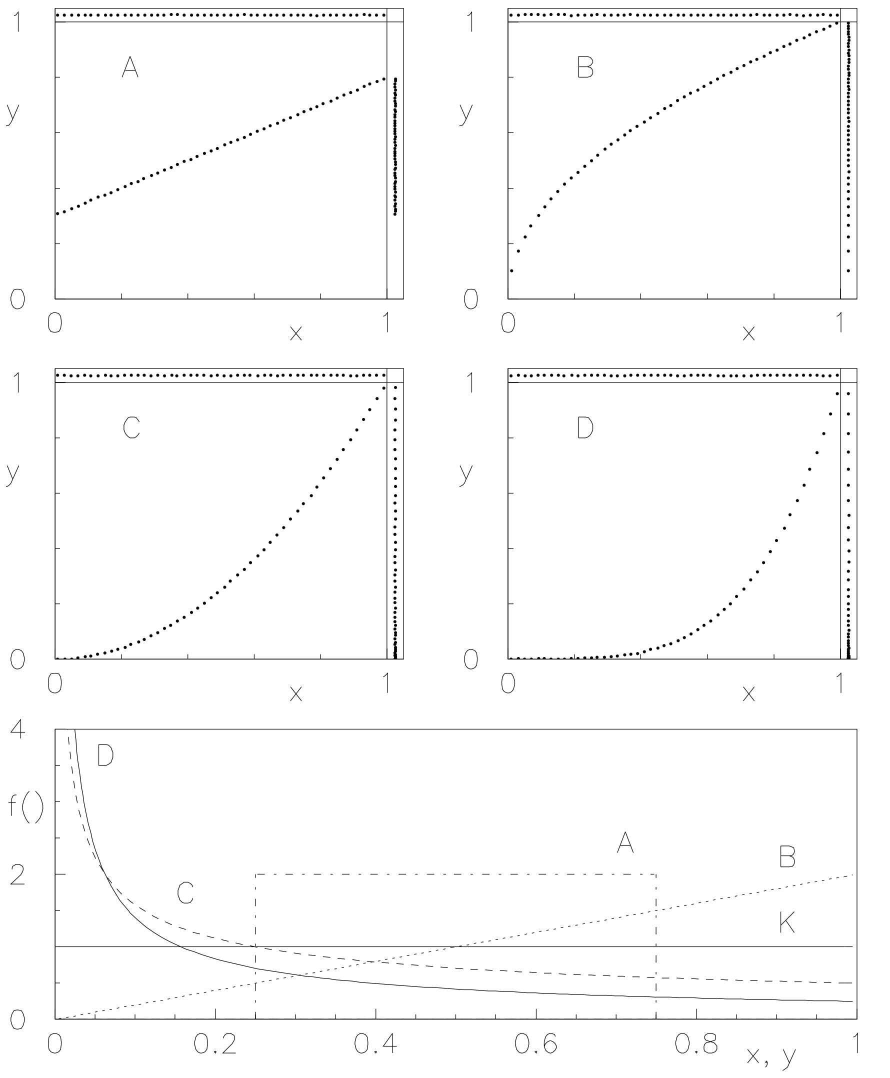

Figura 10.2: Esempi di trasformazione di variabile; A) $Y = 0.5\,X + 0.25$; B) $Y = \sqrt{X}$; C) $Y = X^2$; D) $Y = X^4$.

© G. D'Agostini 2001

Quindi la trasformazione di variabile data da una funzione uguale alla funzione di ripartizione produce una variabile uniforme fra 0 e 1, per qualunque $f(x)$ di partenza. Questa osservazione può essere utilizzata come un modo alternativo per giustificare il metodo di simulazione di variabili casuali qualsiasi discusso nel paragrafo 8.3

Trasformazione lineare di una variabile distribuita normalmente

Consideriamo il caso

$$\begin{cases} X \sim \mathcal{N}(\mu_X, \sigma_X) \\ Y = a\,X + b \end{cases} .$$

Facendo uso della 10.13 si ha:

$$\begin{aligned} f_Y(y) &= \frac{f_X\left(\frac{y-b}{a}\right)^2}{|a|} \\ &= \frac{1}{\sqrt{2\,\pi}|a|\sigma} \exp\left[-\frac{y-(a\mu_X+b))^2}{2\,a^2\,\sigma_X^2}\right] , \end{aligned} \tag{10.15}$$

ovvero

$$\begin{cases} Y \sim \mathcal{N}(\mu_Y, \sigma_Y) \\ \mu_Y = a\,\mu_X + b \\ \sigma_Y = |a|\sigma_X \end{cases} .$$

Una trasformazione lineare di una normale dà ancora luogo ad una normale. Per quanto riguarda la trasformazione di valore atteso e deviazione standard, la legge di trasformazione segue una regola generale che non dipende dal tipo di distribuzione.

10.3.2 Caso di funzioni non monotone

Se $g(X)$ è una funzione che acquista lo stesso valore di y per più valori di x la procedura mostrata può essere utilizzata in ciascuno dei tratti in cui la funzione è monotona. Quindi si sommano i diversi termini che contribuiscono allo stesso y. Mostriamo il modo di operare con un esempio particolare: il quadrato di una variabile distribuita normalmente.

$$\begin{cases} X \sim \mathcal{N}(0, 1) \\ Y = X^2 \end{cases}$$

Per definizione abbiamo

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \mathrm{d}x \\ &= \int_0^{y} \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}t} \mathrm{d}t \\ f(y) &= \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y} . \end{aligned}$$

© G. D'Agostini 2001

Confrontando questo risultato con la (8.45) si riconosce nella Y una variabile di chi^2 con un grado di libertà. Questa identificazione deriva da un proprietà più generale, secondo cui la somma dei quadrati di n variabili normali standardizzate indipendenti è descritta da una chi^2 con $\nu = n$.

10.3.3 Somma di due variabili

La distribuzione della somma $Z = X + Y$ è analoga al caso discreto in cui al posto della sommatoria compare l'integrale su x. Le (10.8) e (10.9) diventano

$$f(z) = \int_{-\infty}^{+\infty} f(x, z-x)\,\mathrm{d}x \qquad \text{(caso generale)} \qquad (10.16)$$

$$f(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x \qquad (X \text{ e } Y \text{ indip.}) \qquad (10.17)$$

La forma (10.17) è nota come *convoluzione* ed è indicata con $f_X * f_Y$.

Somma di due variabili distribuite uniformemente

L'analogo continuo della somma dei due dadi di figura 10.1 è dado dalla somma di due variabili indipendenti distribuite uniformemente, per semplicità nell'intervallo [0, a]. Sia $f_X(\cdot)$ che $f_Y(\cdot)$ valgono $1/a$. Applicando la (10.17) si ha:

$$\begin{aligned} f(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x \\ &= \frac{1}{a}\int_0^a f_Y(z-x)\,\mathrm{d}x \\ &= \frac{1}{a}\int_{z-a}^{z} f(u)\,\mathrm{d}u \qquad (10.18) \\ &= \left(\text{con } f(u) = f_Y(u) = \frac{1}{a} \text{ per } 0 \le u \le a\right), \end{aligned}$$

avendo eseguito, nell'ultimo passaggio, la seguente trasformazione di variabili:

$$\begin{aligned} u &= z - x \\ du &= -dt \\ x = 0 &\to u = z \\ x = a &\to u = z - a\,. \end{aligned}$$

Facendo attenzione a come gli estremi di integrazione della (10.18) sono legati al dominio della $f_Y(\cdot)$, si ottiene finalmente:

$$f(z) = \begin{cases} \frac{1}{a}\int_0^z \frac{1}{a}\,\mathrm{d}u = \frac{1}{a^2} z & 0 \le x \le a \\ \frac{1}{a}\int_{z-a}^{a} \frac{1}{a}\,\mathrm{d}u = \frac{1}{a^2}(2a - z) & a \le x \le 2\,a \\ 0 & \text{altrove} \end{cases} \qquad (10.19)$$

© G. D'Agostini 2001

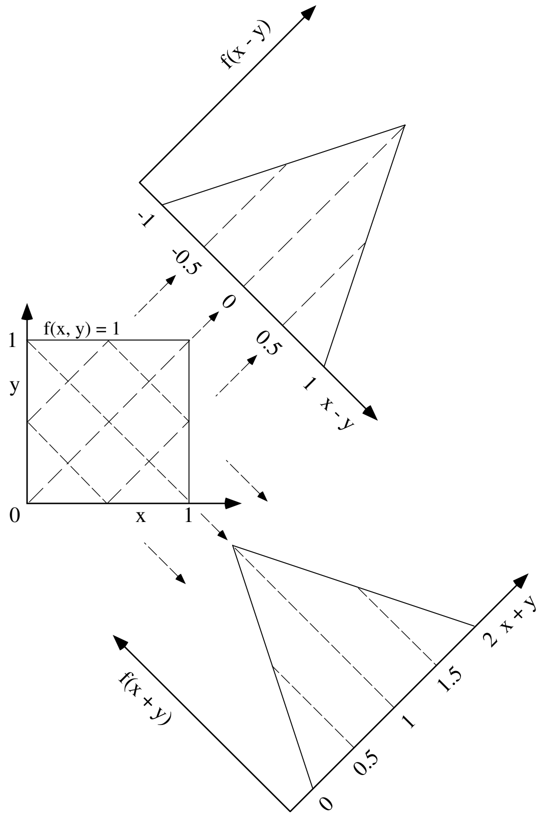

Figura 10.3: Costruzione delle distribuzioni di $X+Y$ e di $X-Y$ a partire da X e Y indipendenti e distribuite uniformemente fra 0 e 1.

Si riconosce una distribuzione triangolare (vedi paragrafo 8.4) centrata in a e di semiampiezza $\Delta = a$. Per capire in modo intuitivo questo andamento si può confrontare questo risultato con quanto ottenuto nella somma di due dadi (vedi figura 10.1). Si veda anche la figura 10.3 che mostra, per una distribuzione uniforme fra 0 e 1 la distribuzione della somma e della differenza. Si vede che, a parte la posizione del centro, tale combinazione dà luogo alla stessa forma, quindi con uguale incertezza.

Somma di due variabili distribuite normalmente

Un caso molto interessante è quello della somma di due variabili indipendenti distribuite normalmente. Applicando la formula di convoluzione si ha

$$f(z) = \int_{-\infty}^{+\infty} \frac{1}{2\,\pi\,\sigma_X \sigma_Y} \exp\left[-\frac{(x-\mu_X)^2}{2\,\sigma_X^2} - \frac{(z-x-\mu_Y)^2}{2\,\sigma_Y^2}\right] \mathrm{d}x\,.$$

© G. D'Agostini 2001

Effettuando l'integrale[6] si ottiene

$$f(z) = \frac{1}{\sqrt{2\,\pi}\sqrt{\sigma_X^2+\sigma_Y^2}}\exp\left[-\frac{[z-(\mu_X+\mu_Y)]^2}{2(\sigma_X^2+\sigma_Y^2)}\right]. \qquad (10.20)$$

La somma è ancora distribuita normalmente con

$$\mu_Z = \mu_X+\mu_Y \qquad (10.21)$$
$$\sigma_Z^2 = \sigma_X^2+\sigma_Y^2 \qquad (10.22)$$

Utilizzando il risultato precedente sulla trasformazione lineare di una variabile (vedi (10.15)), possiamo affermare che la combinazione lineare di variabili casuali indipendenti distribuite normalmente è ancora distribuita normalmente. In altre parole, la gaussiana gode di una proprità riproduttiva più generale di quella di cui godono binomiale e poissoniana.

10.4 ❄Uso della funzione generatrice dei momenti

Un altro semplice e utile metodo per ricavare la distribuzione di una combinazione lineare di variabili casuali indipendenti fa uso dlle proprietà della funzione generatrice dei momenti (vedi 8.13). In effetti, già introducendo tale strumento di calcolo era stato mostrato che una trasformazione lineare di una gaussiana dà ancora luogo ad una gaussiana (vedi (8.36)).

Come esempio di applicazione, ricaviamoci con tale metodo quanto già dimostrato nei paragrafi precedenti a proposito della somma di due variabili indipendente che seguono una distribuzione di Poisson e della combinazione linare di più variabili distribuite normalmente. [7]

10.4.1 $Z = X+Y$, con X e Y poissoniane

Se $X \sim \mathcal{P}_{\lambda_1}$ e $Y \sim \mathcal{P}_{\lambda_2}$, segue

$$G_X(t) = e^{-\lambda_1}e^{\lambda_1 e^t}$$
$$G_Y(t) = e^{-\lambda_2}e^{\lambda_2 e^t}.$$

Ne segue che

$$\begin{aligned} G_{X+Y} &= G_X(t)G_Y(t) \\ &= e^{-\lambda_1-\lambda_2}e^{(\lambda_1+\lambda_2)e^t} \\ &= e^{-\lambda}e^{\lambda e^t}. \end{aligned} \qquad (10.23)$$

[6] Può far comodo, ad un certo punto, l'uso del seguente integrale indefinito:

$$\int_{-\infty}^{+\infty}\exp\left[bx-\frac{x^2}{a^2}\right]\mathrm{d}x = \sqrt{a^2\pi}\exp\left[\frac{a^2b^2}{4}\right].$$

Per il resto serve solo molta attenzione nei passaggi.

[7] Facciamo notare come la proprietà 3) del paragrafo 8.13, presentata in anticipo rispetto alle variabili multiple e che riportiamo per comodità

3) $\quad G_{X+Y}(t) = G_X(t)G_Y(t) \qquad$ con X e Y indipendenti .

segue dal valore atteso di e^{X+Y} sulla funzione congiunta $f(x,y)$, che si riduce a $f(x)f(y)$ in quanto X e Y indipendenti.

© G. D'Agostini 2001

G_{X+Y} ha quindi la forma della funzione generatrice di una distribuzione di Poisson, con $\lambda = \lambda_1 + \lambda_2$. Siamo quindi arrivati alle stesse conclusioni della (10.10), ma in un modo molto più semplice.

Per curiosità consideriamo anche la trasformazione lineare $Y = aX + b$, sempre a partire da Y poissoniana:

$$\begin{aligned} G_{aX+b} &= e^{b\,t} G_X(at) \\ &= e^{b\,t} e^{-\lambda} e^{\lambda e^{a\,t}} \\ &= e^{b\,t-\lambda} e^{\lambda e^{a\,t}} . \end{aligned}$$

Questa funzione non è assolutamente riconducibile alla funzioone generatrice di una poissoniana e quindi una trasformazione lineare di una poissoniana non dà una poissoniana.

10.4.2 $Z = aX + bY + c$, con X e Y gaussiane

. Consideriamo la semplice lineare di due variabili distribuite normalmente. L'estensione al casop generale è immediata. Siano quandi

$$\begin{cases} X \sim \mathcal{N}(\mu_X, \sigma_X) \\ Y \sim \mathcal{N}(\mu_Y, \sigma_Y) \\ Z = aX + bY + c \end{cases}$$

Indichiamo con $V = aX + c$. Facendo uso delle tre proprietà del paragrafo 8.13 si ha

$$G_Z(t) = \exp\left[(a\,\mu_X + c)t + b\mu_Y + (a^2\,\sigma_X^2 + b^2\,\sigma_Y^2)\,t^2/2\right] ,$$

nella quale si riconosce ancora una distribuzione normale di parametri

$$\begin{cases} \mu_Z = a\,\mu_X + b\,\mu_Y + c \\ \sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 . \end{cases}$$

10.5 ❄ Stime a bruta forza: metodi di Monte Carlo

Un metodo pratico molto usato per stimare le distribuzioni di probabilità di funzioni complicate di variabili casuali consiste nell'utilizzare delle tecniche di simulazione al computer. Tenendo conto che in pratica anche le funzioni continue sono discretizzate per eseguire i calcoli, non si fa altro che applicare che applicare la regola generale incontrata nel paragrafo 10.2.1 (formula (10.6), (10.5), o loro generalizzazione a molte dimensioni). Il caso più semplice consiste nello stimare $f(y)$ a partire da $f(x_1)$, $f(x_2)$,... $f(x_n)$, sapendo che $Y = Y(X_1, X_2, \ldots, X_n)$ e assumendo che le variabili di partenza X_i siano fra loro indipendenti. I passi richiesti sono i seguenti:

1. estrarre le variabili X_i usando, ad esempio, una delle tecniche descritte nel paragrafo 8.3;
2. valutare la funzione Y in corrispondenza dei valori X_i;
3. istogrammare la variabile Y e ripetere il processo un numero di volte n.

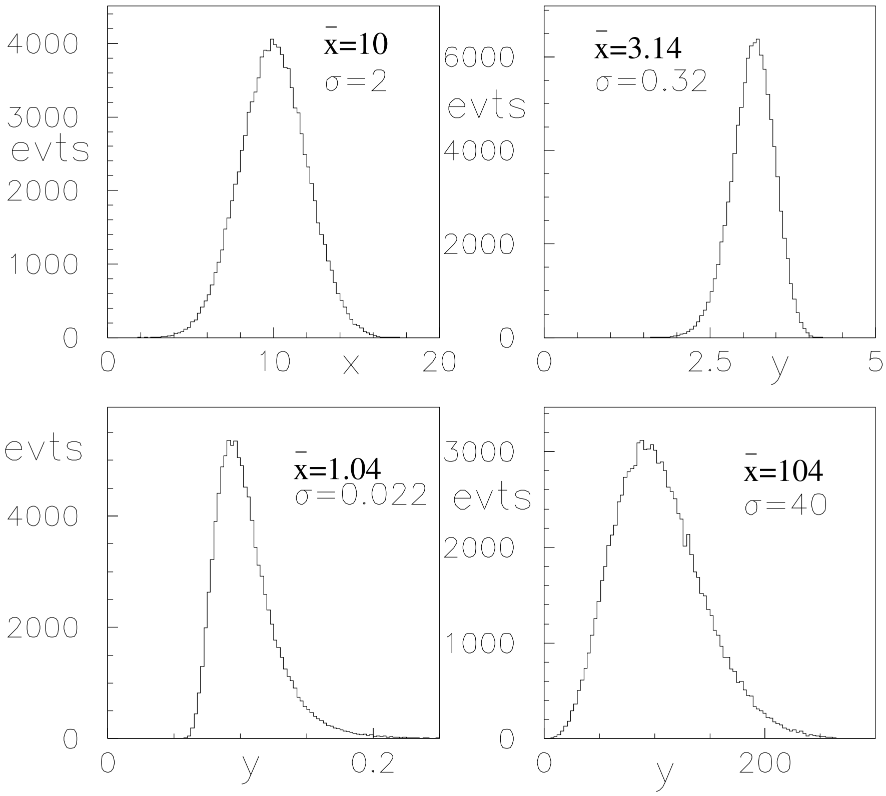

Figura 10.4: Esempi di trasformazione di una variabile normale: $X \sim \mathcal{N}(10, 2)$; $\sqrt{X}$; $1/X$; X^2.

Si calcola infine la frequenza relativa w_{Y_i} con cui il valore Y compare nelle diverse classi dell'istogramma. Poiché al crescere di n si è sempre più sicuri che la distribuzione statistica delle frequenze che si otterrà sarà molto vicina alla distribuzione di probabilità (vedi paragrafi 7.13 e 7.14, nonché prossimo paragrafo 10.9.2) la distribuzione di probabilità è stimata nel seguente modo:

A) caso discreto

$$f(y_i) \approx w_i(Y)\,;$$

B) caso continuo

$$f(y_i) \approx \frac{w_i(Y)}{\Delta Y_i}\,,$$

ove $w_i(Y)$ indica la frequenza relativa di occorrenze del valore $Y = y_i$ (caso discreto) o delle volte in cui il valore della Y è risultato compreso fra y_i e $y_i + \Delta Y_i$ (caso continuo). Un esempio di applicazione verrà mostrato in figura 10.5 a proposito del teorema del limite centrale. L'estensione al caso di più variabili di arrivo Y_j o a quello in cui le variabili di partenza sono correlate è semplice dal punto di vista concettuale, anche se l'implementazione al computer può essere non proprio banale.

© G. D'Agostini 2001

Un esempio di questa tecnica è mostrato in figura 10.4, ove una variabile casuale X distribuita secondo una distribuzione normale di $\mu = 10$ e $\sigma = 2$ è generata 100˙000 volte con la tecnica di Monte Carlo (vedi paragrafo 10.12) e successivamente trasformata in $Y = \sqrt{X}$, $1/X$ e X^2 (nell'ordine e come facilmente riconoscibile dai valori medi e deviazioni standard delle distribuzioni statistiche risultanti dalla simulazione).

10.6 Riepilogo di alcune proprietà delle funzioni di variabili casuali

Prima di proseguire è opportuno fare brevemente il punto della situazione riguardo le funzioni di variabili casuali. Questo dovrebbe permettere una agevole lettura dei paragrafi che seguono anche a coloro che hanno saltato i paragrafi precedenti.

- Il calcolo della forma della distribuzione di una funzione qualsiasi di variabili casuali è un problema complicato.

- Le applicazioni specifiche che incontremo, legate all'incertezza di misura, non richiedono tale abilità.

- Alcune distribuzioni godono della cosiddetta proprietà riproduttiva rispetto alla somma. In particolare, questo è vero per
  - la binomiale (vedi paragrafo 7.7);
  - la distribuzione di Poisson (ibidem);
  - la normale (vedi punto seguente);

- La distribuzione normale è anche riproduttiva sotto una qualsiasi combinazione lineare:

  > se n variabili indipendenti X_i sono distribuite normalmente con $\mathrm{E}(X_i) = \mu_i$ e $\mathrm{Var}(X_i) = \sigma_i^2$, la variabile casuale $Y = \sum_i c_i X_i$ è ancora distribuita normalmente con $\mathrm{E}(Y) = \mu_Y = \sum_i c_i \mu_i$ e $\mathrm{Var}(Y) \sum_i c_i^2 \sigma_i^2$ (si noti il quadrato dei coefficienti).

- La somma di alcune distribuzioni può dar luogo ad altre distribuzioni notevoli. Il caso più semplice è quello che conduce alla binomiale partendo da tanti processi di Bernoulli indipendenti.

- La tabella 10.3 mostra alcuni esempi notevoli di somma di variabili indipendenti.

10.7 Valore atteso e varianza di combinazioni lineari

Sviluppiamo ora il programma illustrato all'inizio di questo capitolo e schematizzato nella (10.3). Calcoliamo quindi il valore atteso di una combinazione

© G. D'Agostini 2001

distribuzione degli addendi	distribuzione della somma
n Bernoulli, $\mathcal{B}_p$	binomiale, $\mathcal{B}_{n,p}$
binomiale, $\mathcal{B}_{n_i,p}$	binomiale, $\mathcal{B}_{\sum_i n_i,p}$
Poisson, $\mathcal{P}_{\lambda_i}$	Poisson, $\mathcal{P}_{\sum_i \lambda_i}$
normale, $\mathcal{N}(\mu_i, \sigma_i)$	normale, $\mathcal{N}(\sum \mu_i, (\sum_i \sigma_i^2)^{\frac{1}{2}})$
k geometriche, $\mathcal{G}_p$	binomiale negativa, $\mathcal{B}^-_{k,p}$ (***)
k esponenziali, $\mathcal{E}_{r=1/\tau}$	Erlang(k,r) (***)

Tabella 10.3: Distribuzione della somma di variabili casuali indipendenti. Esempi notevoli.

lineare di variabili casuali, cominciando dal caso più semplice, quello della somma algebrica:

$$\begin{aligned}
\mathrm{E}(X \pm Y) &= \iint (x \pm y)\, f(x,y) dx dy \\
&= \iint x\, f(x,y) dx dy \pm \iint y\, f(x,y) dx dy \\
&= \mathrm{E}(X) \pm \mathrm{E}(Y)\,. \qquad (10.24)
\end{aligned}$$

Nel caso generale

$$\mathrm{E}(\alpha\, X + \beta\, Y) = \alpha\, \mathrm{E}(X) + \beta\, \mathrm{E}(Y)\,. \qquad (10.25)$$

Avendo ottenuto il valore della previsione, interessiamoci all'incertezza di previsione, data dalla deviazione standard. Anche in questo caso cominciamo dalla somma algebrica:

$$\begin{aligned}
\mathrm{Var}(X \pm Y) &= \mathrm{E}\left[((X \pm Y) - \mathrm{E}(X \pm Y))^2\right] \\
&= \mathrm{E}\left[(X - \mathrm{E}(X)) \pm (Y - \mathrm{E}(Y))^2\right] \\
&= \mathrm{E}\left[(X - \mathrm{E}(X))^2 + (Y - \mathrm{E}(Y))^2\right. \qquad (10.26) \\
&\qquad \left.\pm 2\ (X - \mathrm{E}(X))\ (Y - E[Y])\right] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\, \mathrm{Cov}(X,Y)\,. \qquad (10.27)
\end{aligned}$$

Nel caso in cui la covarianza si annulla troviamo che la varianza di una somma o differenza di due variabili casuali è uguale alla somma delle varianze.

Utilizzando i simboli compatti di σ^2 per la varianza e σ_{XY} per la covarianza, possiamo riscrivere la 10.27 come

$$\sigma^2(X \pm Y) = \sigma_X^2 + \sigma_Y^2 \pm 2\, \sigma_{XY}\,. \qquad (10.28)$$

Nel caso di una somma di n variabili $Y = \sum_i X_i$ otteniamo:

$$\begin{aligned}
\mathrm{E}(Y) &= \sum_i \mathrm{E}(X_i) \qquad (10.29) \\
\mathrm{Var}(Y) &= \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) \qquad (10.30)
\end{aligned}$$

© G. D'Agostini 2001

e, nella notazione simmetrica appena introdotta:

$$\sigma_Y^2 = \sum_{ij} \sigma_{ij} \,, \tag{10.31}$$

ove $\sigma_{ii} = \sigma_{x_i}^2$ e $\sigma_{ij} = \text{Cov}(x_i, x_j)$. Consideriamo infine una combinazione lineare di due variabili $Y = a\,X_1 + b\,X_2 + c$:

$$\begin{aligned}
\text{Var}(Y) &= \text{E}\left[(aX_1 + bX_2 + c - \text{E}[a\,X_1 + b\,X_2 + c])^2\right] \\
&= \text{E}\left[(a\,(X_1 - \text{E}(X_1)) + b\,(X_2 - \text{E}(X_2)))^2\right] \\
&= a^2\,\text{Var}(X) + b^2\,\text{Var}(X_2) + 2\,a\,b\,\text{Cov}(X_1, X_2)\,.
\end{aligned} \tag{10.32}$$

Questo risultato può essere esteso a un numero qualsiasi di variabili:

$$\begin{aligned}
Y &= \sum_i \alpha_i\,X_i \\
\text{E}(Y) &= \sum_i \alpha_i\ \text{E}(X_i) && (10.33) \\
\text{Var(Y)} &= \sum_i \alpha_i^2\,\text{Var}(X_i) + 2\sum_{i<j} \alpha_i\,\alpha_j\,\text{Cov}(X_i, X_j) && (10.34)
\end{aligned}$$

Possiamo riscrivere in modo più compatto l'espressione della covarianza, osservando innanzitutto che

$$2\sum_{i<j} \alpha_i\,\alpha_j\,\text{Cov}(X_i, X_j) = \sum_{i\neq j} \alpha_i\,\alpha_j\,\text{Cov}(X_i, X_j)\,,$$

in quanto il fattore 2 è dovuto a considerare sia $i < j$ che $i > j$. Utilizzando inoltre la notazione compatta, questa può essere riscritta come

$$\sum_{i\neq j} \alpha_i\,\alpha_j \sigma_{ij}\,.$$

Anche i termini dovuti alle varianze possono essere riscritti come

$$\sum_i \alpha_i^2 \text{Var}(X_i) = \sum_i \alpha_i \alpha_i \sigma_i^2 = \sum_i \alpha_i \alpha_i \sigma_{ii}\,,$$

ove σ_{ii} sta per σ_i^2. Per concludere, il modo più compatto di scrivere la varianza di una combinazione lineare[8] è:

$$\sigma_Y^2 = \sum_{ij} \alpha_i\,\alpha_j\,\sigma_{ij}\,. \tag{10.35}$$

[8]Siccome questa notazione presenta a volte difficoltà di interpretazione, facciamo un esempio nel caso di 2 variabili. La sommatoria $\sum_{ij}$ indica che bisogna considerare le $2 \times 2 = 4$ possibili combinazioni:

$$\{1,1\},\ \{1,2\},\ \{2,1\},\ \{2,2\}\,.$$

Quindi:

$$\begin{aligned}
\sigma_Y^2 &= \alpha_1\alpha_1\sigma_{1\,1} + \alpha_1\alpha_2\sigma_{1\,2} + \alpha_2\alpha_1\sigma_{2\,1} + \alpha_2\alpha_2\sigma_{2\,2} \\
&= \alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\alpha_1\alpha_2\sigma_{1\,2}\,.
\end{aligned}$$

(ricordiamo che $\sigma_{2\,1} = \sigma_{1\,2}$).

Vediamo due applicazioni di quanto abbiamo mostrato in questo paragrafo.

Valore atteso e varianza della distribuzione binomiale

Come sappiamo, la binomiale è data dalla somma di n variabili casuali indipendenti, ciascuna legata ad un processo di Bernoulli di probabilità p, uguale per tutte. Poiché valore atteso e varianza di una distribuzione di Bernoulli valgono rispettivamente p e $p\,q$, con $q = 1 - p$, otteniamo

$$\mathrm{E}(X \mid \mathcal{B}_{n,p}) = \sum_i p = n\,p \tag{10.36}$$

$$\mathrm{Var}(X \mid \mathcal{B}_{n,p}) = \sum_i p\,q = n\,p\,q\,. \tag{10.37}$$

Valore atteso e varianza della distribuzione di Erlang

Anche il valore atteso e la previsione del tempo in cui si verifichi il k-mo successo in un processo di Poisson (distribuzione di Erlang) può essere valutato dalla proprietà delle combinazioni lineari di variabili casuali indipendenti. Poiché il tempo di attesa fra un evento e l'altro segue una distribuzione di esponenziale, il tempo di attesa di k successi è pari alla somma di k variabili casuali ciascuna di valore atteso e deviazione standard pari $\tau = 1/r$. Ne segue che

$$\mathrm{E}(T \mid \mathrm{Erlang}(k,r)) = k\,\tau = \frac{k}{r} \tag{10.38}$$

$$\mathrm{Var}(T \mid \mathrm{Erlang}(k,r)) = k\,\tau^2 = \frac{k}{r^2} \tag{10.39}$$

$$\sigma(T \mid \mathrm{Erlang}(k,r)) = \sqrt{k}\,\tau = \frac{\sqrt{k}}{r} \tag{10.40}$$

Previsione di una media aritmetica di variabili aleatorie analoghe

Una applicazione di particolare interesse della proprietà della varianza è quella relativa alla media aritmetica effettuata n variabili casuali analoghe, aventi tutte la stessa distribuzione e indipendenti l'una dall'altra (ad esempio l'i-mo esito di un esperimento condotto in condizioni apparentemente identiche):

$$\overline{X}_n = \frac{1}{n}\sum_i X_i \tag{10.41}$$

$$\mathrm{E}\left(\overline{X}_n\right) = \frac{1}{n}\sum_i \mathrm{E}(X_i) = \mu \tag{10.42}$$

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_i \mathrm{Var}(X_i) = \frac{\mathrm{Var}(X)}{n} \tag{10.43}$$

$$\sigma(\overline{X}_n) = \frac{1}{\sqrt{n}}\,\sigma(X)\,. \tag{10.44}$$

La media aritmetica ha la stessa previsione della singola variabile, ma un'incertezza $\sqrt{n}$ minore. Questo è un risultato molto importante per quanto riguarda

© G. D'Agostini 2001

le misure e sarà ripreso nel seguito. Detto alla buona, esso giustifica l'idea intuitiva che il valore ottenuto da una media di molte misure è più affidabile di quello ottenuto da una singola misura. L'espressione di $\sigma(\overline{X}_n)$ fornisce una espressione quantitativa dell'incremento della "qualità" della media aritmetica in funzione del numero di prove.

10.8 ⥁ Correlazione fra diverse combinazioni lineari di variabili casuali

Immaginiamo di avere due diverse combinazioni lineari, costruite sulle stesse variabili casuali di partenza:

$$\begin{cases} Y = \sum_i \alpha_i X_i \\ Z = \sum_i \beta_i X_i \end{cases} .$$

Dipendendo Y e Z dalle stesse variabili X_i, esse non sono in genere indipendenti. Ci aspettiamo quindi un coefficiente di correlazione non nullo. Vediamo quindi come calcolare la covarianza. Cominciamo con due sole variabili di partenza, X_1 e X_2, ovvero

$$\begin{cases} Y = \alpha_1 X_1 + \alpha_2 X_2 \\ Z = \beta_1 X_1 + \beta_2 X_2 \end{cases} .$$

Per definizione

$$\mathrm{Cov}(Y, Z) \;=\; \mathrm{E}[(Y - \mu_Y)(Z - \mu_Z)] \,.$$

Essendo

$$\begin{cases} \mu_Y = \mathrm{E}(Y) = \alpha_1 \mu_1 + \alpha_2 \mu_2 \\ \mu_Z = \mathrm{E}(Z) = \beta_1 \mu_1 + \beta_2 \mu_2 \end{cases}$$

si ha

$$\begin{aligned} Y - \mu_Y &= \alpha_1(X_1 - \mu_1) + \alpha_2(X_2 - \mu_2) \\ Z - \mu_Z &= \beta_1(X_1 - \mu_1) + \beta_2(X_2 - \mu_2) \end{aligned}$$

e quindi la covarianza è pari al valore atteso di

$$[\alpha_1(X_1 - \mu_1) + \alpha_2(X_2 - \mu_2)]\,[\beta_1(X_1 - \mu_1) + \beta_2(X_2 - \mu_2)]$$

Sviluppando il prodotto del binomio e prendendo il valore atteso di ciascuno di esso si hanno questi quattro termini:

$$\begin{aligned} \mathrm{E}\left[\alpha_1\beta_1(X_1 - \mu_1)^2\right] &= \alpha_1\beta_1 \mathrm{E}\left[(X_1 - \mu_1)^2\right] = \alpha_1\beta_1 \mathrm{Var}(X_1) \\ \mathrm{E}\left[\alpha_2\beta_2(X_2 - \mu_2)^2\right] &= \alpha_2\beta_2 \mathrm{E}\left[(X_2 - \mu_2)^2\right] = \alpha_2\beta_2 \mathrm{Var}(X_2) \\ \mathrm{E}\left[\alpha_1\beta_2(X_1 - \mu_1)(X_2 - \mu_2)\right] &= \alpha_1\beta_2 \mathrm{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] \\ &= \alpha_1\beta_2 \mathrm{Cov}(X_1, X_2) \\ \mathrm{E}\left[\alpha_2\beta_1(X_2 - \mu_2)(X_1 - \mu_1)\right] &= \alpha_2\beta_1 \mathrm{E}\left[(X_2 - \mu_2)(X_1 - \mu_1)\right] \\ &= \alpha_2\beta_1 \mathrm{Cov}(X_1, X_2) \end{aligned}$$

© G. D'Agostini 2001

(Si ricordi che

$$\mathrm{E}\left[(X_2-\mu_2)(X_1-\mu_1)\right]=\mathrm{E}\left[(X_1-\mu_1)(X_2-\mu_2)\right]=\mathrm{Cov}(X_1,X_2)\,.\,)$$

Possiamo quindi scrivere la covarianza fra Y e Z come la somma di tre contributi: quello dovuto alla varianza di X_1, quello dovuto alla varianza di X_2 e quello dovuto alla loro covarianza:

$$\begin{aligned}\mathrm{Cov}(Y,Z) &= \alpha_1\beta_1\mathrm{Var}(X_1)+\alpha_2\beta_2\mathrm{Var}(X_2)\\ &\quad +(\alpha_1\beta_2+\alpha_2\beta_1)\mathrm{Cov}(X_1,X_2)\,. \end{aligned}\tag{10.45}$$

Questa espressione mostra come, in effetti, le variabili finali sono correlate anche se quelle iniziali non lo sono. Inoltre, si noti la covarianza dipende dal prodotto dei coefficienti della stessa variabile X_i, o da quelli relativi a variabili legate da una covarianza non nulla. Quindi, la condizione affinché due combinazioni lineari siano fra loro correlate è che abbiano in comune almeno una variabile, oppure, per dirlo in modo figurato, "si parlino" attraverso due variabili fra loro correlate. Questa condizione è necessaria ma non sufficiente, in quanto il valore della covarianza dipende dai diversi contributi e dai segni di coefficienti. Si possono verificare quindi delle compensazioni fra le diverse variabili tali da annullare la covarianza.

Il caso limite di correlazione fra Y e Z si verifica quando si ha una sola variabile.

$$\left\{\begin{array}{l} Y=\alpha X\\ Z=\beta X\end{array}\right. .$$

Si ha infatti

$$\rho(Y,Z)=\frac{\mathrm{Cov}(Y,Z)}{\sigma_Y\sigma_Z}=\frac{\alpha\beta\sigma_X^2}{|\alpha|\sigma_X\cdot|\beta|\sigma_X}=\frac{\alpha\beta}{|\alpha\beta|}\,.$$

La correlazione vale ±1 a seconda che i segni di α e β siano concordi o discordi.

In analogia a quanto fatto nel paragrafo precedente a proposito della varianza, è facile estendere l'espressione della covarianza nel caso di molte variabili di partenza X_i.

$$\begin{aligned}\mathrm{Cov}(Y,Z) &= \sum_i \alpha_i\beta_i\mathrm{Var}(X_i)+\\ &\quad \sum_{i<j}(\alpha_i\beta_j+\alpha_j\beta_i)\mathrm{Cov}(X_i,X_j) \qquad (10.46)\\ &= \sum_i \alpha_i\beta_i\sigma_i^2+\sum_{i<j}(\alpha_i\beta_j+\alpha_j\beta_i)\sigma_{ij}\\ &= \sum_i \alpha_i\beta_i\sigma_i^2+\sum_{i\neq j}\alpha_i\beta_j\sigma_{ij}\\ \sigma_{YZ} &= \sum_{i,j}\alpha_i\beta_j\sigma_{ij} \qquad (10.47)\end{aligned}$$

Si noti come, se $Z=Y$, si ritorna all'espressione (10.35) della varianza di Y.

© G. D'Agostini 2001

Covarianza di due medie aritmetiche

Immaginiamo di ripetere molte volte un esperimento, sotto condizioni apparentemente identiche, e che per l'i-mo esperimento si misurino due grandezze X_i e Y_i. I possibili esiti degli esperimenti saranno n coppie di variabili casuali analoghe, cioè tutte aventi la stessa distribuzione di probabilità. Nel caso generale X_i e Y_i non sono in genere indipendenti e avremo quindi una covarianza $\text{Cov}(X, Y)$ (è stato omesso l'indice i in quanto la covarianza è la stessa per tutte le coppie. Applichiamo ora i risultati del paragrafo precedente per calcolare la covarianza fra le due medie aritmetiche $\overline{X}_n = \frac{1}{n}\sum_i X_i$ e $\overline{Y}_n = \frac{1}{n}\sum_i Y_i$. Al fine di utilizzare le formule precedenti si consideri $2\,n$ variabili W_i, tali che

$$\begin{aligned} W_i &= X_i \qquad (1 \le i \le n) \\ W_{n+i} &= Y_i \qquad (1 \le i \le n) \end{aligned}$$

Le due medie possono essere riscritte come

$$\begin{aligned} \overline{X}_n &= \sum_i^{2n} \alpha_i W_i \\ \overline{Y}_n &= \sum_i^{2n} \beta_i W_i \end{aligned}$$

I coefficienti α_i valgono $1/n$ per i compreso fra 1 e n, mentre sono nulli per $i > n$. Viceversa β_i sono nulli per $i < n$ e valgono $1/n$ per $n + 1 \le i \le n$. Osservando la (10.46) (riscritta in termini delle W_i), notiamo che i soli contributi non nulli sono quelli dovuti alla covarianza fra W_i e W_{n+i}, corrispondenti alla coppia $\{X_i, Y_i\}$[9]. Ne segue che

$$\begin{aligned} \text{Cov}(\overline{X}_n, \overline{Y}_n) &= \sum_i^n \alpha_i \beta_{n+i} \text{Cov}(W_i, W_{n+i}) \\ &= n \cdot \left(\frac{1}{n}\frac{1}{n}\right) \text{Cov}(X, Y) \\ &= \frac{1}{n}\text{Cov}(X, Y) \end{aligned} \tag{10.48}$$

Come la varianza, anche la covarianza delle medie è $1/n$ volte inferiore alla covarianza di ciascuna coppia di variabili. Rimane invece invariato il coefficiente di correlazione:

$$\rho(\overline{X}_n, \overline{Y}_n) = \frac{\text{Cov}(\overline{X}_n, \overline{Y}_n)}{\sigma(\overline{X}_n)\sigma(\overline{Y}_n)} = \rho(X, Y)\,. \tag{10.49}$$

[9] Le variabili X_i e Y_j con $i \neq j$ non sono invece correlate in quanto, per dirlo in modo semplice, appartengono a diversi esperimenti. Ad esempio, la conoscenza del valore di X della prima prova non modifica il grado di fiducia dei valori di Y che possono accadere nella seconda prova.

© G. D'Agostini 2001

Correlazione fra una variabile e una combinazione lineare che la contiene

C'è un altro caso che può a tornare utile per le applicazioni di laboratorio. Riguarda la valutazione del coefficiente di correlazione fra una variabile casuale, ad esempio X_1, e una combinazione lineare di variabili casuali che contiene anche questa, ad esempio $Y = \sum_i \alpha_i X_i$. Ovviamente il risultato segue da quento trattato in questo paragrafo se si considera una seconda combinazione lineare Z che contenga soltanto X_1 con coefficiente $\beta_1 = 1$. Facciamo soltanto il caso semplice ed istruttivo di variabili x_i fra loro indipendenti. Otteniamo allora

$$\text{Cov}(Y, X_1) = \alpha_1 Var(X_1)$$

e quindi

$$\rho(Y, X_1) = \frac{\alpha_1\sigma_1}{\sqrt{\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + \cdots + \alpha_n^2\sigma_n^2}}\,.$$

Il coefficiente di correlazione è dato dal rapporto fra il contributo a σ_Y della sola X_1 e σ_Y stessa. Quindi tende a ± 1 quando X_1 domina l'incertezza su Y, mentre tende a zero quando il suo contributo è trascurabile.

10.9 Legge dei grandi numeri

In questo paragrafo approfondiamo la discussione su due risultati che abbiamo già incontrato.

1. Parlando delle previsioni di distribuzioni statistiche (paragrafo 7.13) abbiamo visto come la previsione della frequenza relativa tenda alla probabilità, con una incertezza di previsione che decresce con la radice quadrata del numero di prove.

2. Analogalmente, abbiamo visto nelle pagine precedenti come la previsione della media aritmetica di variabili casuali aventi la stessa distribuzione di probabilità è uguale alla previsione della variabile casuale, con una incertezza di previsione che decresce con la radice quadrata del numero di prove.

Questi sono due modi di formulare la *legge dei grandi numeri*, indubbiamente la legge più nota e più fraintesa[10] del calcolo delle probabilità. L'espressione semplificata con la quala essa è nota è che *"con l'aumentare del numero di prove effettuate nelle stesse condizioni, la frequenza relativa tende alla probabilità e la media sperimentale tende alla media teorica"*, expressione che potrebbe anche essere corretta, una volta chiarito il significato dei termini. Vedremo quali sono invece le interpretazioni assolutamente scorrette di tale legge. Riformuliamo prima correttamente le due "leggi", che sono due semplici teoremi del calcolo delle probabilità.

[10] Un volta una persona mi confidò: *"non credo al calcolo delle probabilità, perché mi sembra una grande stupidaggine che se un numero non è uscito da molte settimane debba avere più probabilità degli altri di uscire"...*

© G. D'Agostini 2001

10.9.1 Limite della media aritmetica

Prendiamo n variabili casuali X_i i cui valori sono descritti da una distribuzione di probabilità che sia la stessa per tutte le variabili. Si può pensare quindi ad un esperimento condotto n volte sotto le stesse condizioni. Tale distribuzione ha valore atteso μ e deviazione standard σ.

Interessiamoci alla media aritmetica

$$\overline{X}_n = \frac{1}{n}\sum_i X_i$$

effettuata sui valori di X_i che si verificheranno. Come abbiamo già visto, anche $\overline{X}_n$ è una variabile casuale, in quanto funzione di variabili casuali. Abbiamo già visto che

$$\begin{aligned} \mu_{\overline{X}_n} \equiv \mathrm{E}(\overline{X}) &= \mu \\ \sigma_{\overline{X}_n} = \sigma(\overline{X}_n) &= \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Per quanto detto nel paragrafo 7.10 ci aspettiamo che al crescere di n diminuisce la probabilità di trovare valori di $\overline{X}_n$ "molto distanti" da μ. Ripetiamo il ragionamento, facendo uso della disuguaglianza di Cebicev, applicata alla variabile casuale $\overline{X}_n$ (in realtà si può fare anche uso della forma esatta della distribuzione di probabilità di $\overline{X}_n$, che vedremo fra breve parlando del teorema del limite centrale):

$$P(|\overline{X}_n - \mu_{\overline{X}_n}| \geq k\sigma\overline{X}_n) \leq \frac{1}{k^2}\,.$$

Inserendo i valori di $\mu_{\overline{X}_n}$ e $\sigma\overline{X}_n$ si ottiene

$$P(|\overline{X} - \mu| \geq k\,\frac{\sigma}{\sqrt{n}}) \leq \frac{1}{k^2}\,. \qquad (10.50)$$

L'interpretazioone di questa disuguaglianza è che, data una certa σ, fissato un certo livello di probabilità $1/k^2$ e un valore dello scarto Δ piccoli a piacere, è sufficiente effettuare un numero di prove corrispondentemente grande affinche' la probabiltà di osservare uno scarto assoluto maggiore di Δ sia inferiore al valore di probabilità prescelto.

Facciamo un esempio numerico: prendiamo $\sigma = 1$, il valore di probabilità dell'1 % e uno scarto $\Delta = 10^{-3}$. Riformuliamo il problema: quante prove devo fare per essere sicuro al 99 % che $\overline{X}$ non differirà per più di 10^{-3} da μ? Dalla 10.50 abbiamo che

$$\begin{aligned} k &= \sqrt{\frac{1}{p}} = 10 \\ n &= \left(\frac{k\,\sigma}{\Delta}\right)^2 = 10\times 10^6\,. \end{aligned}$$

Se scegliamo $\Delta = 10^{-6}$ abbiamo bisogno di $n = 10^{13}$ per avere lo stesso livello di sicurezza. Mantenendo $\Delta = 10^{-6}$, possiamo richiedere un livello di

© G. D'Agostini 2001

sicurezza del 99.99 % che lo scarto sia contenuto entro Δ, ma allora avremo bisogno di $n = 10^{15}$ prove. E così via. Quindi, quando il numero di prove è grandissimo siamo "praticamente" sicuri di trovare $\overline{X}$ prossimo a μ. Possiamo esprimere il risultato dicendo che

$$\text{“}\lim_{n\to\infty}\text{”}\overline{X} = \mu\,,$$

dove le virgolette stanno a ricordare che tale concetto è ben diverso dal limite di funzioni. In particolare, esso non garantisce che, pur effettuando un numero di prove arbitrariamente grande, ci sia la certezza di trovare $\overline{X}$ entro Δ da μ (anche se Δ è relativamente grande, purché non superiore all'intervallo di definizione di X).

10.9.2 Teorema di Bernoulli

Riprendiamo quanto detto nel paragrafo 7.13. Considerando n prove indipendenti, la previsione dellla frequenza relativa di successo, che per comodità chiamiamo f_n (uguale a W del paragrafo 7.13), abbiamo

$$\mathrm{E}(f_n) \equiv \frac{1}{n}\mathrm{E}(X \mid \mathcal{B}_{n,p}) = \frac{n\,p}{n} = p \tag{10.51}$$

$$\sigma(f_n) \equiv \frac{1}{n}\sigma((X \mid \mathcal{B}_{n,p}) = \frac{\sqrt{n\,p\,q}}{n} = \frac{\sqrt{p\,q}}{\sqrt{n}} \tag{10.52}$$

Procedendo per analogia al caso precedente otteniamo:

$$P(|f_n - \mathrm{E}(f_n)| \geq k\,\sigma(f_n)) \leq \frac{1}{k^2}$$

$$P(|f_n - p| \geq k\,\frac{\sqrt{p\,q}}{\sqrt{n}}) \leq \frac{1}{k^2} \tag{10.53}$$

L'interpretazione di questa disuguaglianza, nota come *teorema di Bernoulli*, è analoga a quanto visto per $\overline{X}_n$:

> *al crescere di n diventa sempre meno probabile osservare valori di frequenze relative che differiscono molto da p.*

Essendo questo uno dei teoremi più intriganti del calcolo delle probabilità, merita una serie di osservazioni.

- La formulazione va intesa in termini di probabilità e non di certezza.
- Il teorema non implica assolutamente che, se per un certo N lo scarto $|f_n - p|$ è grande, allora per $n > N$ la frequanza relativa f_n "debba recuperare" per "mettersi in regola con la legge". (Nel seguito mostreremo, come esercizio, la fallacia dell'interpretazione della legge dei grandi numeri per aspettarsi che i numeri ritardatari al lotto recuperino.)
- Esso non giustifica la "definizione" frequentista di probabilità. Affermando infatti che "è molto probabile che la frequenza non differisca molto dalla probabilità" si sta assumendo il concetto di probabilità. Inoltre:

© G. D'Agostini 2001

- Non si dimentichi che il teorema di Bernoulli ...è un teorema, basato sulle regole di base dalla probabilità e su tutte le proprietà che ne derivano. Quindi non può definire il concetto di probabilità.
- Su tale argomento è molto convincente de Finetti

  > *"Per quanti tendono a ricollegare il concetto stesso di probabilità alla nozione di frequenza, tali risultati [*che f_n "tenda a p"*] vengono ad assumere un ruolo di cerniera per convalidare tale avvicinamento o identificazione di nozioni. Logicamente non si sfugge però al dilemma che la stessa cosa non si può assumere prima per definizione e poi dimostrare come teorema, né alla contraddizione di una definizione che assumerebbe una cosa certa mentre il teorema afferma che è soltanto molto probabile.*

- Si noti inoltre che la condizione di p costante implica che essa sia prefissata a priori e che anche le valutazioni sui possibili esiti di f_n siano fatte prima di iniziare le prove (o in condizione di incertezza sul loro esito).
- Ciò nonostante, vedremo come la frequenza relativa di prove effettuate possa essere utilizzata per valutare p, ma questo è un problema inferenziale che c'entra poco con il teorema di Bernoulli (il quale ha a che vedere, per dirla alla buona, soltanto con "affermazioni probabilistiche su frequenza relative future, o comunque ignote").

Facciamo degli esempi pratici

Lancio di una moneta

Prima di lanciare una moneta regolare 1000 (o se so di qualcuno che lo sta facendo, o che lo abbia fatto, ma prima che conosca i risultati), la legge dei grandi numeri mi dice che io sono abbastanza sicuro di avere circa il 50% di teste; se i lanci saranno un 1˙000˙000 sono molto più sicuro dell'affermazione, e così via. Il teorema di Bernoulli non dice niente di più!

E se a metà dei 1000 lanci ho osservato 200 teste? Chiaramente si tratta di un altro problema, di tipo inferenziale. Posso, ad esempio, soppesare la mia fiducia sulla regolarità della moneta con l'"eccezionalità" dell'osservazione sperimentale per riagggiornare la probabilità che esca testa in ciascuno dei prossimi lanci. Quindi si può far di nuovo uso della legge dei grandi numeri per fare delle considerazioni sui prossimi 500 lanci.

Sul recupero dei numeri ritardatari

Assumiamo ora di avere un meccanismo di estrazione talmente simmetrico da non avere nessun dubbio sulla probabilità di ciascuna estrazione (ad esempio la ruota del lotto). Sia p tale probabilità (ad esempio 1/90 per il primo estratto su una ruota). Immaginiamo di interessarci ad N estrazioni e che dopo N_1 estrazioni ($N_1 < N$) la frequenza relativa sia diversa da p (come in genere accade...):

$$f_{N_1} \neq p\,.$$

Cosa ci aspettiamo sul totale delle N estrazioni da eseguire?

- ci sarà un meccanismo di compensazione tale che $f_N \approx p$?
- Quanto vale il valore atteso della frequenza relativa nelle restanti $N_2 = N - N_1$ estrazioni e sul totale delle N estrazioni?

Queste sono le considerazioni probabilisticamente corrette all'istante $t_\circ$ prima di iniziare le estrazioni e all'istante t_1 dopo che sono noti i risultati delle prime N_1 estrazioni:

a) $t_\circ$**:** siamo in stato di incertezza su ciascuno degli esiti. Ne segue che

$$\begin{aligned} \mathrm{E}(f_{N_1}) &= \mathrm{E}(f_{N_2}) = \mathrm{E}(f_N) = p\,. && (10.54) \\ \sigma(f_{N_1}) &= \frac{\sqrt{p\,(1-p)}}{\sqrt{N_1}} && (10.55) \\ \sigma(f_{N_2}) &= \frac{\sqrt{p\,(1-p)}}{\sqrt{N_2}} && (10.56) \\ \sigma(f_N) &= \frac{\sqrt{p\,(1-p)}}{\sqrt{N}} && (10.57) \end{aligned}$$

B) t_1**:** il risultato f_{N_1} è acquisito (certo!). Esso corrisponde a $X_1 = N_1\, f_{N_1}$ successi. Quindi:

$$\begin{aligned} \mathrm{E}(f_{N_1}) &= f_{N_1} \qquad \text{con certezza} \\ \mathrm{E}(f_{N_2}) &= p \qquad \text{con } \sigma(f_{N_2}) = \frac{\sqrt{p\,(1-p)}}{\sqrt{N_2}} \\ \mathrm{E}(f_N) &= \frac{N_1\, f_{N_1} + N_2\, p}{N_1 + N_2} \neq p\ (\mathbf{!}) \\ \sigma(f_N) &= \frac{\sqrt{N_2 p\,(1-p)}}{N_1 + N_2} = \sqrt{\frac{N - N_1}{N}}\sqrt{p\,(1-p)}\,. \end{aligned}$$

La previsione della frequenza è la media pesata fra p e la frequenza nei primi N_1 eventi. L'incertezza di previsione è massima all'inizio (per $N_1 = 0$) e va a zero per $N_1 \to N$, in quanto rimane poco su cui essere incerti.

Lasciando al lettore tutte le conclusioni legate ai numeri ritardatari e le raccomandazioni ai perdendi che "contano" di rifarsi nelle partite successive, concludiamo con delle raccomandazioni. I ragionamenti probabilistici vanno effettuati quando si è in condizioni di incertezza. Si possono usare (vanno usate!) le informazioni sulle frequenze osservate per valutare le probabilità di altri eventi ignoti nei modi che saranno visti quando affronteremo in modo sistematico il problema dell'inferenza statistica. Utilizzare argomentazioni probabilistiche su eventi noti non ha alcun senso. Utilizzare poi tali argomentazioni allo stesso tempo sia su eventi incerti che su eventi certi conduce ad una grossa confusione, una di questa è l'interpretazione distorta del teorema di Bernoulli.

© G. D'Agostini 2001

10.10 Teorema del limite centrale

Siamo ora giunti a quello che è il più importante teorema sulle distribuzioni di probabilità.

Nei paragrafi precedenti abbiamo imparato a valutare previsione e incertezza di previsione di combinazioni lineari di variabili casuali, indipendentemente dalla loro distribuzione. Ma per esprimere affermazioni probabilistiche su tali funzioni serve conoscere la distribuzione di probabilità e questo è un problema non banale (vedi paragrafi facoltativi 10.2 e 10.3). Comunque, l'esperienza su distribuzioni tipiche (vedi paragrafo 7.10) ci insegna che in genere c'è alta probabilità che in valore della variabile aleatoria cada entro alcune deviazioni standard dal valore atteso. Quando poi ci sono servite della considerazioni probabilistiche estreme, indipendentemente dalla distribuzione, abbiamo fatto uso della disuguaglianza di Cebicev.

In realtà le combinazioni lineari tendono ad avere una distribuzione di probabilità universale, indipendentemente dalle distribuzioni di partenza, come consequenza del *teorema del limite centrale*, che formularemo fra breve. Siccome consideriamo assolutamente facoltativa la dimostrazione del teorema e molto più importante una sua modellizzazione e rappresentazione visiva, cominciamo con degli esempi pratici.

Abbiamo visto nel capitolo 6 come costruire una variabile casuale sulla somma degli esiti di due dadi. Mentre la distribuzione dell'esito di un solo dado è uniforme, la distribuzione della somma mostra un addenzamento della probabilià al centro, ed un massimo in corrispondenza di $X = 7$. La figura 6.2 mostra chiaramente l'origine del cambiamento di forma: ci sono molte combinazioni che possono dare valori centrali e poche che possono dare valori laterali. Se si somma la variabile "due dadi" con un'altra variabile indipendente legata ad un solo dado, di nuovo sarà molto probabile formare combinazioni intorno al centro della distribuzione (vedi ad esempio figura 10.1).

Anche partendo da una distribuzione continua uniforme si ha lo stesso effetto: la somma di due variabili dà luogo ad una distribuzione triangolare (chi è interessato alla dimostrazione può consultare il paragrafo 10.3.3, ma anche la sola figura 10.3 è autoesplicativa e mostra l'analogia con il lancio dei dadi).

Osserviamo ora la figura 10.5. Essa mostra le distribuzioni simulate ottenute estraendo un certo numero di variabili casuali ($n = 1$, 2, 3, 5, 10, 20 e 100) e sommandole fra di loro. Questo processo è ripetuto 10000 volte. Si ottiene quindi, per ciascun caso, una distribuzione statistica che somiglia alla corrispondente distribuzione di probabilità[11]. Le distribuzioni di base sono una uniforme fra 0 e 1 e una distribuzione "strana" uniforme fra 0 e 0.25, fra 0.75 e 1, nulla altrove. La distribuzione "strana" è particolarmente istruttiva per capire l'effetto di addensamento al centro della probabilità dovuto alle combinazione dei diversi valori. Ad esempio, per $n = 2$, si nota un triangolo centrale doppio di ciascuno dei triangoli laterali. Esso è dovuto alla probabilità che un valore grande di una variabile si combini con un valore grande dell'altra variabile.

[11] Si ricorda che la previsione di una distribuzione statistica è uguale alla distribuzione di probabilità, con una incertezza che decresce con il numero di estrazioni. Per l'uso delle simulazioni per stimare distribuzioni di probabilità si veda il paragrafo 10.5

© G. D'Agostini 2001

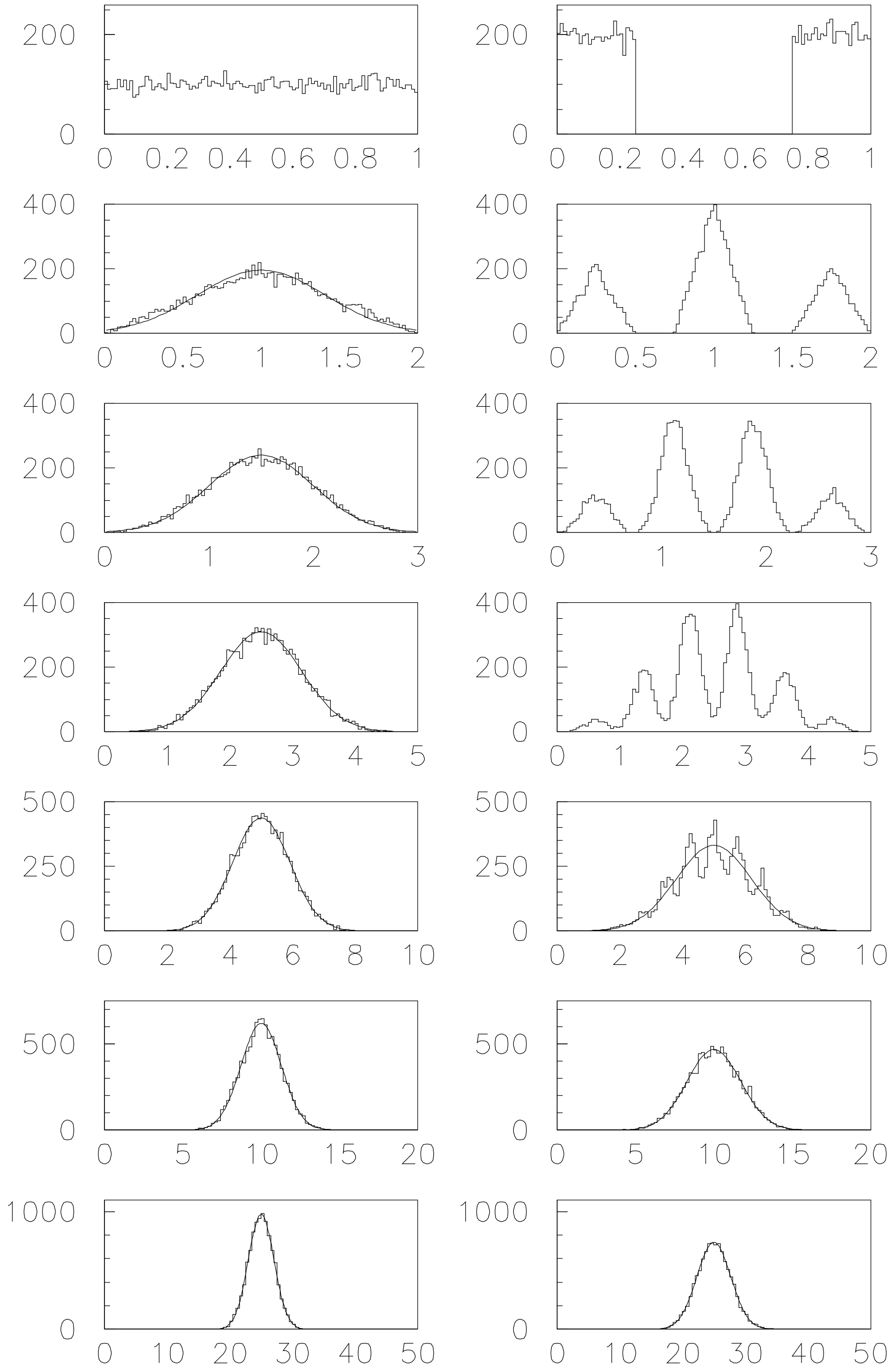

Figura 10.5: Teorema del limite centrale all'opera. Le due figure in alto mostrano la distribuzione 10000 eventi generati secondo una distribuzione uniforme (sinistra) e una distribuzione a "onda quadra", diversa da zero e costante negli intervalli fra 0 e 0.25 e fra 0.75 e 1. Successivamente (dall'alto verso il basso) sono mostrate le somme di n variabili casuali estratte dalle due distribuzioni. n vale, nell'ordine, 1, 2, 3, 5, 10, 20 e 100. In alcuni casi sono anche mostrate le gaussiane aventi la stessa media e varianza delle distribuzioni e normalizzate a 10000 eventi.

© G. D'Agostini 2001

Si nota che, al crescere dei termini della sommatoria, la distribuzione risultante ha una forma regolare a campana indipendentemente dalla forma iniziale. Si noti come le distribuzioni asintotiche ($n = 100$) sono centrate sullo stesso valore in entrambi i casi, mentre sono diverse le larghezze. Questa è conseguenza che le due distribuzioni iniziali avessero stessa media, ma diversa deviazione standard (quella "strana" indica una previsione del quadrato degli scarti dalla media più grande dell'altro caso). La curva riportata sui vari istogrammi è una gaussiaana. Si vede come, da un certo n in poi (diverso per i due casi!) la distribuzione è ben approssimata da una normale.

Questo comportamento è dovuto al *teorema del limite centrale*:

> *date n variabili casuali indipendenti X_i, anche descritte da distribuzioni di probabilità diverse (purché aventi valore medio e varianza finiti), al crescere di n ("nel limite di $n \rightarrow \infty$") la combinazione lineare*
>
> $$Y = \sum_{i=1}^{n} \alpha_i X_i$$
>
> *ha distribuzione di probabilità normale di parametri*
>
> $$\left\{ \begin{array}{l} \mu = \sum_i \alpha_i \mu_i \\ \sigma^2 = \sum_i \alpha_i^2 \sigma_i^2 \end{array} \right.$$
>
> *purché sia*
>
> $$\alpha_i^2 \sigma_i^2 \ll \sigma^2$$
>
> *per ogni variabile X_i non distribuita normalmente.*

Data l'importanza del teorema, sono necessari alcuni chiarimenti sulle condizioni di validità.

- La condizione $\alpha_i^2 \sigma_i^2 \ll \sigma^2$ può essere espressa dicendo che nessun termine deve dominare le fluttuazioni, altrimenti esso sarà determinante ai fini della distribuzione.

- La deroga a tale condizioni per le variabili distribuite normalmente è dovuto al fatto che le combinazioni lineari di gaussiane sono sempre gaussiane indipendentemente dal loro numero e dai valori di σ.

- Non c'è invece nessuna condizione sul valori attesi, in quanto questo teorema descrive le fluttuazioni, mentre le posizioni possono essere variate a piacere (purché i valori siano finiti, naturalmente).

- Il teorema non dice da quale valore di n l'approssimazione è valida. Dipende dal tipo di distribuzione. Dalla figura 10.5 si vede che, a partire da distribuzioni uniformi di larghezza confrontabile, già per $n = 4$ o 5 l'approssimazione è ragionevole. Partendo da una distribuzione triangolare (equivalente a 2 uniformi) sono sufficienti già 2 o 3 termini. In effetti per i casi pratici, con distribuzioni piccate al centro la convergenza è rapidissima.

© G. D'Agostini 2001

- Per non dare l'illusione che la convergenza a normale sia sempre rapida, facciamo un controesempio notevole. Per la proprietà riproduttiva della poissoniana, la somma di n poissoniane indipendenti danno ancora luogo ad una poissoniana. Quindi, sotto certe condizioni, essa tenderà a normale (vedi paragrafo successivo). Se però consideriamo un miliardo di variabili, ciascuna avente $\lambda = 10^{-9}$ la distribuzione finale sarà ancora poissoniana con $\lambda = 1$, ben diversa da una normale! (In questo caso particolare la convergenza a normale si ha per $n \gg 10^9$.)

10.10.1 Distribuzione della media aritmetica

Diamo qui una prima applicazione del teorema alla distribuzione della media aritmetica. La variabile casuale media aritmetica è pari, a meno di un fattore di proporzionalità, ad una sommatoria e quindi, al crescere del numero di osservazioni sulle quali essa è effettuata, sarà descritta sempre meglio da normale:

$$\overline{X} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})\,. \tag{10.58}$$

10.10.2 Convergenza in distribuzione della binomiale e della poissoniana

Sia la binomiale che la poissoniana godono della proprietà riproduttiva. Questa proprietà permette di pensare una distribuzione caratterizzata da un valore medio elevato come una somma di tante variabili casuali provenienti dallo stesso tipo di distribuzione ma caratterizzate da valori medi più piccoli. Siccome per il teorema del limite centrale una somma di variabili casuali tende ad una distribuzione normale, la distribuzione binomiale e quella di Poisson tendono alla distribuzione normale al crescere, rispettivamente, di np e di λ:

- distribuzione binomiale: per valori di np e di nq "abbastanza grandi" la distribuzione binomiale tende ad una distribuzione normale di $\mu = np$ e $\sigma = \sqrt{pqn}$;

- distribuzione di Poisson: per valori di λ "abbastanza grandi" la distribuzione di Poisson tende ad una distribuzione normale di $\mu = \lambda$ e $\sigma = \sqrt{\lambda}$.

La condizione di valore "abbastanza grande" dipende dal grado di accuratezza con cui si vuole calcolare la probabilità. Per la maggior parte delle applicazioni che si incontrano nella trattazione degli errori e per la determinazione degli intervalli di fiducia la condizione è soddisfatta per valori di np, nq e λ maggiori di 10-15.

Bisogna precisare cosa si intende per limite di una distribuzione discreta a una distribuzione continua. Infatti in un caso la funzione di distribuzione ha il significato di una probabilità e nell'altro di una densità di probabilità. Il limite va allora inteso per la funzione di ripartizione:

$$\lim_{n\to\infty} F(x_i) = \int_{-\infty}^{x_i+1/2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

© G. D'Agostini 2001

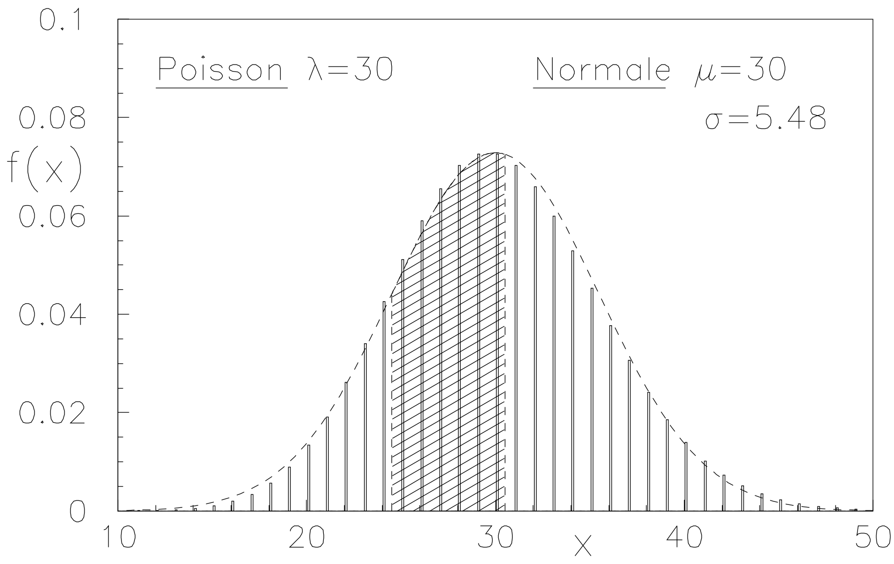

Figura 10.6: Approssimazione normale della distribuzione di Poisson.

dove l'integrale è stato esteso fino a $x_i + 1/2$ per tener conto della natura discreta della variabile casuale. Otteniamo quindi per la probabilità di un determinato valore della variabile casuale:

$$
\begin{aligned}
P(X = x_i) = f(x_i) &= \lim_{n\to\infty}(F(x_i) - F(x_i - 1)) \\
&= \int_{x_i-1/2}^{x_i+1/2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\end{aligned}
$$

dove abbiamo utilizzato il fatto $x_i - x_{i-1} = 1$ per le distribuzioni discrete di interesse.

A volte, se si è interessati alla probabilità di un solo valore x_i, l'integrale viene ignorato e la distribuzione normale viene utilizzata come se fosse una distribuzione discreta. Questo deriva dall'approssimazione

$$
\begin{aligned}
f(x_i) &= \int_{x_i-1/2}^{x_i+1/2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&\approx \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \Delta x \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}},
\end{aligned}
$$

essendo $\Delta x = 1$.

© G. D'Agostini 2001

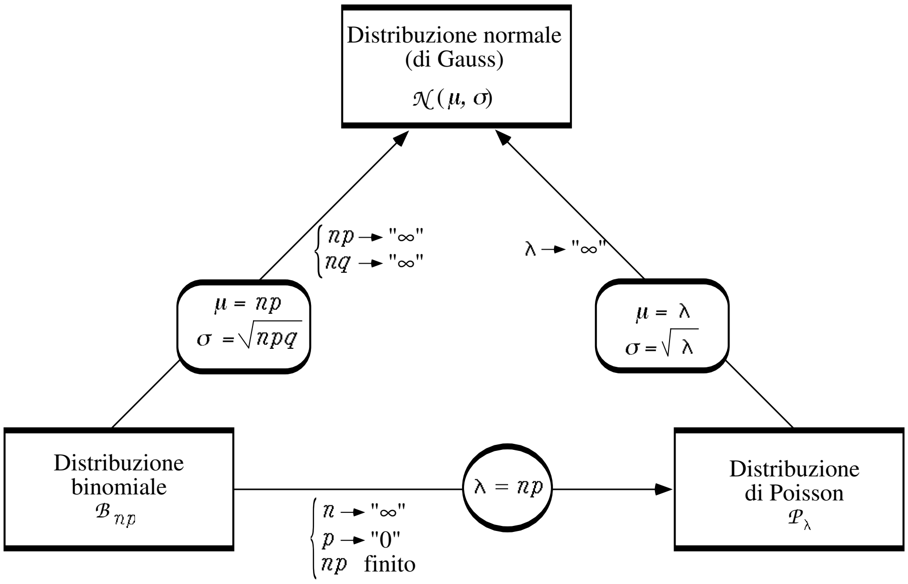

Figura 10.7: Proprietà asintotiche della distribuzione binomiale e di quella di Poisson.

10.11 Estensione del teorema del limite centrale a variabili non indipendenti

$\rightarrow$ considerare anche i termini di covarianza
$\rightarrow$ è possibile pensare ad una trasformazione di variabili linare tali che porti il nuovo insieme di variabili sia indipendenti e la varianza di Y è uguale a quella delle variabili originali calcolata anche con i termini di covarianza.

10.12 ❄ Simulazione di numeri aleatori distribuiti secondo una distribuzione normale

Abbiamo visto nel paragrafo 8.3 alcune tecniche per simulare numeri aleatori. Vediamone un'altra, di semplice implementazione che permette di generare dei numeri secondo una distribuzione normale.

Consideriamo, come al solito in questi casi, la variabile R distribuita uniformemente fra 0 e 1. Consideriamo 12 variabili indipendenti R_i e costruiamo la nuova variabile casuale

$$X = \sum_{i=1}^{12} R_i - 6$$

di valore atteso e varianza pari a

$$\mathrm{E}(X) = 12\,\mathrm{E}(R_i) - 6 = 0 \qquad (10.59)$$

$$\mathrm{Var}(X) = 12\,\mathrm{Var}(R_i) = 12\,\frac{1}{12} = 1\,. \qquad (10.60)$$

© G. D'Agostini 2001

Quindi, poiché come abbiamo visto, a partire da distribuzioni uniformi simili la convergenza è molto rapida, per il teorema del limite centrale

$$X \sim \mathcal{N}(0,1)\,.$$

Per ottenere qualsiasi altra distribuzione normale Y di parametri μ e σ è sufficiente effettuare la trasformazione

$$Y = \sigma\, X + \mu\,.$$

Un esempio di istogramma ottenuto con questo metodo è mostrato in figura 10.4 per $\mu = 10$ e $\sigma = 2$.

Si noti come, ovviamente, la tecnica non riproduce bene il comportamento delle code estreme, ma la si può migliorare aumentando il numero di termini della somma (e modificando opportunamente le formule).

10.13 ○ Linearizzazione

Una conseguenza importante delle proprietà della combinazione lineare e del teorema del limite centrale è la seguente: se si hanno delle variabili $X_1, X_2, \ldots X_n$ e una variabile Y funzione arbitraria di esse, ovvero

$$Y = Y(X_1, X_2, \ldots, X_n)$$

che sia abbastanza lineare nell'intorno del valore atteso di ciascuna delle X_i (ovvero μ_i) ove si addensa il grosso della probabilità (ovvero entro qualche σ_i da μ_i) allora, espandendo Y in serie di Taylor otteniamo

$$\begin{aligned} Y &\approx Y(\mu_1, \mu_2, \ldots, \mu_n) + \sum_i \frac{\partial Y}{\partial X_i}(X_i - \mu_i) && (10.61) \\ &\approx k + \sum_i \frac{\partial Y}{\partial X_i} X_i\,, && (10.62) \end{aligned}$$

ove le derivate si intendono calcolate per $X_i = \mu_i$. Il secondo modo di scrivere l'espansione, in cui sono stati inglobati in k tutti i termini non dipendenti da X_i, mostra chiaramente che la Y è una combinazione lineare delle X_i con coefficienti pari alle derivate calcolate nella previsione del vettore aleatorio $\{X_1, X_2, \ldots X_n\}$. Si possono applicare quindi a questa combinazione lineare ed ad altre combinazioni lineari costruite sulle X_i tutti i teoremi incontrati in questo capitolo. In particolare, abbiamo:

- la previsione di Y è circa uguale alla funzione Y calcolata sulle previsioni delle X_i:
$$\mathrm{E}(Y) \approx Y(\mu_1, \mu_2, \ldots, \mu_n)\,, \qquad (10.63)$$
come si può valutare rapidamente dalla (10.61) tenendo conto che $\mathrm{E}(X_i - \mu_i) = 0$;

© G. D'Agostini 2001

- la varianza di Y è pari a
$$\sigma_Y^2 = \frac{\partial Y}{\partial X_i}\frac{\partial Y}{\partial X_j}\sigma_{ij}\,, \tag{10.64}$$
la quale si riduce, in caso di variabili X_i fra loro indipendenti a
$$\sigma_Y^2 = \left(\frac{\partial Y}{\partial X_i}\right)^2\sigma_i^2; \tag{10.65}$$
- la distribuzione di Y è circa normale in virtù del teorema del limite centrale e se sono valide le condizioni generali;
- nel caso di tante combinazioni lineari $Y_1(), Y_2(), \ldots Y_m()$ le covarianze fra le varie Y_l sono date da
$$\mathrm{Cov}(Y_l, Y_m) = \sum_{ij}\frac{\partial Y_l}{\partial X_i}\frac{\partial Y_m}{\partial X_j}\sigma_{ij}\,. \tag{10.66}$$

10.14 ⟳ Esempio di applicazione alle incertezze di misure

Facciamo subito un esempio di come sia possibile applicare le proprietà delle combinazioni lineari e della linearizzazione alle incertezze di misure. Anche se non è stato ancora affrontato il discorso dell'inferenza statistica e di come si valutino le incertezze delle misure dirette e le correlazioni introdotte da possibili errori sistematici di entità incerta, supponiamo che siano stati misurati i due lati di un rettangolo e che i risultati siano presentati in termini di previsione e di incertezza di previsione:

$$\begin{aligned} a &= 29.71 \pm 0.03\text{cm} \\ b &= 21.44 \pm 0.03\text{cm}\,. \end{aligned}$$

Ripetendo quanto detto al momento di introdurre previsione e incertezza e incertezza di previsione: $\mathrm{E}(a) = 29.71\,\mathrm{cm}$ rappresenta il valore intorno al quale riteniamo che, ragionevolmente, si trovi a; $\sigma(a) = 0.03\,\mathrm{cm}$ indica la dispersione dei valori che possono essere assunti da a intorno ad $\mathrm{E}(a)$.

Immaginiamo di essere interessati ora alle seguenti grandezze derivate da a e b: perimetro ($p = 2\,a + 2\,b$); somma di lati ($s = a + b$); differenza dei lati ($d = a - b$); area del rettangolo ($A = a\,b$). Supponiamo inoltre i due casi: a) le misure di a e b sono assolutamente indipendenti, per cui $\rho(a, b) = 0$; b) parziale correlazione, con $\rho(a, b) = 0.5$; c) correlazione totale (positiva), con $\rho(a, b) = 1$.

Si riconoscono in p, s e d correlazioni lineari di a e b, mentre il caso dell'area è non lineare e va linearizzato. Riportiamo nel seguito i risultati dettagliati per s, d e A nei tre casi di correlazione considerati. Innanzitutto, le previsioni di s, d e A non dipendono da $\rho(a, b)$ e valgono

$$\mathrm{E}(s) = \mathrm{E}(a) + \mathrm{E}(b) \tag{10.67}$$
$$\mathrm{E}(d) = \mathrm{E}(a) - \mathrm{E}(b) \tag{10.68}$$
$$\mathrm{E}(A) \approx \mathrm{E}(a)\cdot\mathrm{E}(b)\,. \tag{10.69}$$

© G. D'Agostini 2001

Per quanto riguarda le varianze, ricordando che Cov(a, b) è pari a $\rho(a, b)\,\sigma(a)\,\sigma(b)$, abbiamo

$$\begin{aligned}
\sigma^2(s) &= \sigma^2(a) + \sigma^2(b) + 2\,\rho(a, b)\,\sigma(a)\,\sigma(b) && (10.70)\\
\sigma^2(d) &= \sigma^2(a) + \sigma^2(b) - 2\,\rho(a, b)\,\sigma(a)\,\sigma(b) && (10.71)\\
\sigma^2(A) &\approx b^2\sigma^2(a) + a^2\sigma^2(b) + 2\,a\,b\,\rho(a, b)\,\sigma(a)\,\sigma(b)\,. && (10.72)
\end{aligned}$$

Si noti il simbolo "$\approx$" al posto di "$=$" nel caso di A, per ricordare che si tratta di un risultato approssimato basato sulla linearizzazione. I risultati numerici sono:

$$\begin{aligned}
s/\text{cm} &= 51.15 \pm 0.04 && [\rho(a, b) = 0]\\
& 51.15 \pm 0.05 && [\rho(a, b) = 0.5]\\
& 51.15 \pm 0.06 && [\rho(a, b) = 1]\\
d/\text{cm} &= 51.15 \pm 0.04 && [\rho(a, b) = 0]\\
& 51.15 \pm 0.03 && [\rho(a, b) = 0.5]\\
& 51.15 \pm 0 && [\rho(a, b) = 1]\\
A/\text{cm}^2 &= 636.7 \pm 1.1 && [\rho(a, b) = 0]\\
& 636.7 \pm 1.3 && [\rho(a, b) = 0.5]\\
& 636.7 \pm 1.5 && [\rho(a, b) = 1]
\end{aligned}$$

Troviamo infine i coefficienti di correlazione fra s, d e A, come applicazione di quanto visto nel paragrafo 10.8. Utilizzando la formula (10.45), svolgendo in dettaglio il primo passaggio, otteniamo

$$\begin{aligned}
\text{Cov}(s, d) &= (+1)\,(+1)\,\sigma^2(a) + (+1)\,(-1)\,\sigma^2(b) +\\
&\quad + [(+1)\,(-1) + (+1)\,(+1)]\;\rho(a, b)\,\sigma(a)\,\sigma(b)\\
\text{Cov}(s, A) &\approx (+1)\,(+b)\,\sigma^2(a) + (+1)\,(+a)\,\sigma^2(b) +\\
&\quad [(+1)\,(+a) + (+1)\,(+b)]\;\rho(a, b)\,\sigma(a)\,\sigma(b)\\
\text{Cov}(d, A) &\approx (+1)\,(+b)\,\sigma^2(a) + (-1)\,(+a)\,\sigma^2(b) +\\
&\quad [(+1)\,(+a) + (-1)\,(+b)]\;\rho(a, b)\,\sigma(a)\,\sigma(b)\,,
\end{aligned}$$

ove, di nuovo, "$\approx$" sta a ricordare che il risultato è basato sulla linearizzazione. Inserendo i valori e dividendo per le deviazione standard, otteniamo direttamente i coefficienti di correlazione in funzione di $\rho(a, b)$:

$$\begin{aligned}
\rho(s, d) &= 0 && \forall\,\rho(a, b)\\
\rho(s, A) &= 0.9871 && \rho(a, b) = 0\\
&= 0.9957 && \rho(a, b) = 0.5\\
&= 0.9999 && \rho(a, b) = 1.0\\
\rho(d, A) &= -0.160 && \rho(a, b) = 0\\
&= -0.093 && \rho(a, b) = 0.5\\
&= 0 && \rho(a, b) = 1\,.
\end{aligned}$$

Si noti:

© G. D'Agostini 2001

- il coefficiente di correlazione fra s e d è nullo per puro accidente, dovuto all'uguaglianza di $\sigma(a)$ e $\sigma(b)$);
- per calcolare i coefficienti di correlazione per valori molto prossimi ad 1, bisogna utilizzare molte cifre significative nel calcoli, altrimenti si possono ottenere risultati assurdi ($|\rho| > 1$) a causa degli arrotondamenti;
- quando le funzioni diventano complicate non è facile farsi un'idea intuitiva del grado di correlazione e addirittura nemmeno del suo segno.

10.15 ↻ Moto browniano, "pallinometro" ed errori di misura

Concludiamo questo capitolo con delle interessanti applicazioni del teorema del limite centrale.

Nel 1905 Albert Einstein pubblicò un articolo dal titolo "Il moto delle particelle in sospensione nei fluidi in quiete, come previsto dalla teoria cinetico-molecolare del calore". In esso diede una spiegazione in termini di teoria cinetica dei gas del *moto browniano*, osservato per primo dal biologo Robert Brown nel 1828 durante lo studio del comportamento di polline in soluzione acquosa. È molto interessante leggere l'inizio dell'articolo[12]:

> "In questo lavoro faremo vedere come, secondo la teoria cinetico-molecolare del calore, particelle in sospensione in una soluzione compiano, in conseguenza del moto termico delle molecole, movimenti di ampiezza tale che li si può agevolmente osservare al microscopio, purché, beninteso, la dimensione delle particelle stesse sia accessibile allo strumento. Può darsi che i moti che qui saranno considerati coincidano con il cosiddetto moto molecolare browniano; tuttavia i dati che ho potuto ottenere su quest'ultimo sono così imprecisi che non mi è stato possibile formulare alcun giudizio in merito.
>
> Se il moto in questione si potrà effettivamente osservare, con tutte le regolarità che per esso si possono attendere, la termodinamica classica non sarà più da considerare esattamente valida già per spazi microscopicamente distinguibili, e quindi una determinazione esatta della vera grandezza degli atomi sarà possibile. Se viceversa la nostra previsione si dimostrasse inesatta, ciò fornirebbe un serio argomento contro la concezione cinetico-molecolare del calore".

Il moto browniano viene quindi interpretato come una conseguenza del moto disordinato delle particelle a seguito degli urti che ricevono dalle molecole sottoposte ad agitazione termica. Nella trattazione ogni particella viene considerata indipendentemente dalle altre. Il risultato di Einstein fu - prendendo per semplicità il caso unidimensionale - che dopo un certo intervallo di tempo t la distribuzione di probabilità delle distanze X dalla posizione iniziale $X = 0$ per ($t = 0$) è una gaussiana centrata in $X = 0$ la cui varianza è funzione

[12] Albert Einstein, "Opere scelte", Bollati Boringhieri editore, Torino 1988, pag. 136

© G. D'Agostini 2001

lineare del tempo:

$$f(\Delta x, t) = \frac{1}{\sqrt{2\pi}\,\sigma(t)} \exp\left[-\frac{x^2}{2\,\sigma^2(t)}\right] \qquad (10.73)$$

$$\sigma(t) = \sqrt{2\,\delta\,t}\,, \qquad (10.74)$$

dove δ è il *coefficiente di diffusione*. Commenta Einstein (si noti il "come ci si poteva attendere").

> "La distribuzione di frequenza delle variazioni di posizione avvenute in un tempo t arbitrario coincide dunque con quella degli errori casuali, come ci si poteva attendere."

Ci possiamo ricavare lo stesso risultato seguendo un approccio basato sul cosiddetto *moto casuale*: schematizziamo il moto browniano come un moto casuale di una pedina che, partendo all'istante iniziale dalla posizione iniziale $X = 0$, ad ogni mossa possa avanzare o retrocedere di un passo con uguali probabilità. È esattamente quello che si verifica nella classica esperienza del "quinconce di Galton"[13] Ad ogni chiodo la pallina può andare a destra o a sinistra con la stessa probabilità. Il numero di mosse corrisponde alle file di chiodi che la pallina incontra e la posizione $X = 0$ è quella in corrispondenza del primo chiodo che su cui cade la pallina.

Per utilizzare la stessa notazione del paragrafo 2.9 della seconda parte di queste dispense consideriamo degli intervalli di tempo ΔT di durata sufficientemente più grande del tempo tipico di interazione, tale tale che i versi di spostamenti successivi osservabili dopo intervalli ΔT siano fra di loro indipendenti. Assumiamo inoltre, per semplificare il problema, che gli spostamenti dopo il tempo ΔT siano costanti in modulo (ΔX). Quindi siamo interessati al calcolo dei numeri di passi S orientati lungo la direzione dell'asse X, ciascuno di lunghezza ΔX (vedi figura 10.8). Dopo ogni intervallo i-mo ΔT $S_{\Delta T_i}$ può valere, +1 o -1, con pari probabilità. Ne segue che

$$\begin{aligned} \mathrm{E}(S_{\Delta T_i}) &= 0 \\ \mathrm{Var}(S_{\Delta T_i}) &= 1\,. \end{aligned}$$

[13] È quello che gli studenti romani chiamano familiarmente "Pallinometro", composto da una tavola con chiodi disposti a "quinconce" fra i quali scende un pallina. Per averne un'idea, si immagini di ruotare la figura 10.8 di 90 gradi in senso orario, immaginando i puntini come chiodi e il punto 0 il punto di immisione della pallina.

© G. D'Agostini 2001

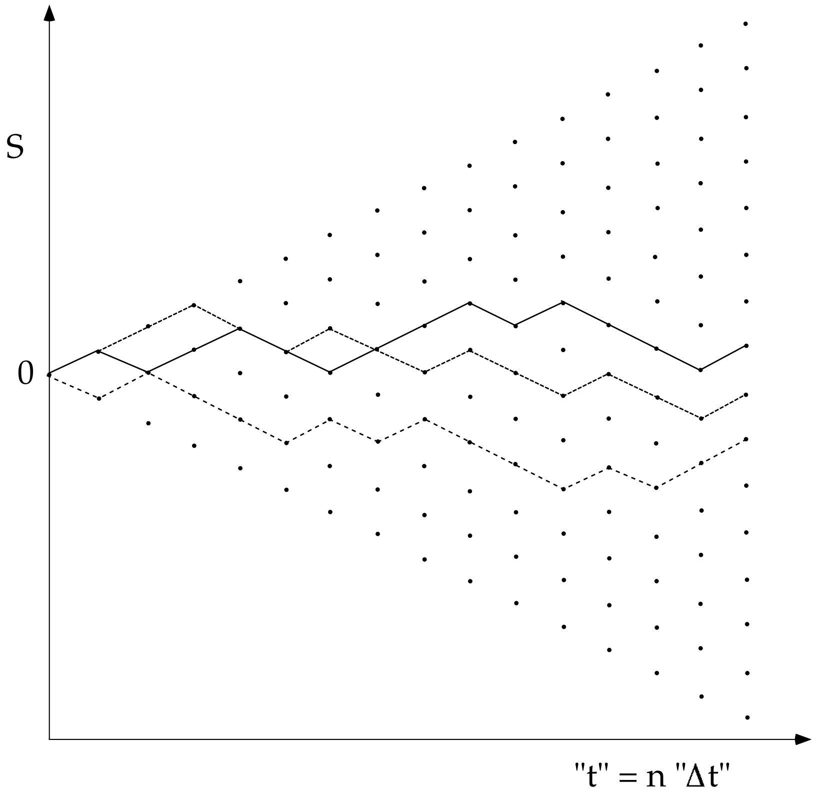

Figura 10.8: Cammino casuale su una sola dimensione.

Quindi, essendo gli spostamenti successivi indipendenti, abbiamo che

$$\begin{aligned} S_n &= \sum_{i=1}^{n} S_{\Delta T_i} \\ X_n &= \Delta X \sum_{i=1}^{n} S_{\Delta T_i} \\ \mathrm{E}(S_n) &= 0 \\ \mathrm{Var}(S_n) &= n \qquad (10.75) \end{aligned}$$

$$\begin{aligned} \mathrm{E}(X_n) &= 0 \qquad (10.76) \\ \mathrm{Var}(X_n) &= n\,\Delta X\,. \qquad (10.77) \end{aligned}$$

Passando alla variabile $X(t)$ in funzione del tempo, dobbiamo soltanto sostituire $n = t/\Delta T$ nelle formule precedenti, ottenendo

$$\begin{aligned} \mathrm{E}(X(t)) &= 0 \qquad (10.78) \\ \mathrm{Var}(X(t)) &= \frac{\Delta X}{\Delta T}\,t \qquad (10.79) \\ \sigma(X(t)) &= \sqrt{\frac{\Delta X}{\Delta T}\,t}\,. \qquad (10.80) \end{aligned}$$

© G. D'Agostini 2001

Applicando inoltre il teorema del limite centrale, otteniamo finalmente che

$$X(t) \sim \mathcal{N}(0, \sqrt{\alpha\, t})$$

ove α, pari a $\Delta X/\Delta T$ e che ha un significato analogo a quello del coefficiente di diffusione (il fattore 2 rispeto alla (10.74) dipende dall'esatta definizione di δ).

A questo punto vale la pena di sottolineare la relazione fra teorema del limite centrale, moto browniano e distribuzione degli errori casuali di misura. Il risultato di ogni singola misura può essere pensato come dovuto a tanti piccoli contributi incoerenti e non controllabili dallo sperimentatore, ognuno dei quali tende a modificare il risultato, in modo del tutto casuale, in un verso o nell'altro. Si pensi ad esempio a vibrazioni, attriti, variazioni termiche, rumore elettronico etc. I vari contributi modificano il valore ideale della grandezza da misurare, il quale viene a compiere una specie di moto browniano in uno spazio astratto. In virtù di questa variazione casuale della grandezza fisica, la singola misura di uno strumento ha la tendenza ad avere errori di tipo normale. In genere essa viene mediata con altri risultati. Infine i risultati delle misure di più grandezze possono venire combinati per ottenere, attraverso relazioni funzionali, altre grandezze fisiche. Si comprende quindi il motivo per il quale all'aumentare del numero di passaggi gli errori tendono sempre più ad assumere una distribuzione normale.

10.16 ❄ Distribuzione di velocità delle molecole di un gas perfetto

Come altro esempio di applicazione della distribuzione normale, ricaviamo la distribuzione di velocità di un gas perfetto, a partire da un moto browniano nello spazio dei momenti. Supponiamo di avere una molecola inizialmente immobile in un recipiente pieno di gas e consideriamo la componente lungo l'asse x della velocità.

- se supponiamo che ad ogni urto la particella acquisti una certa velocità $\pm\Delta v_x$ possiamo immaginare di ottenere un moto browniano in v_x la cui soluzione sarà, come abbiamo visto, una gaussiana di media nulla e varianza $\sigma^2_{v_x}(t)$;
- $\sigma^2_{v_x}(t)$ non può andare all'infinito. Superata una certa velocità media (in modulo) la probabilità che la molecola perda energia è maggiore di quella che ne acquisti. La condizione di equilibrio termodinamico è che, passato un tempo sufficientemente grande, tutte le molecole sono descritte dalla stessa distribuzione di probabilità delle velocità, e in particolare:

  $$\lim_{t\to\infty} \sigma^2_{v_x}(t) = \overline{v_x^2},$$

  dove $\overline{v_x^2}$ è la velocità quadratica media che si ottiene dalla teoria cinetica dei gas:

  $$\overline{v_x^2} = \frac{k\,T}{M}.$$

© G. D'Agostini 2001

Il risultato è:

$$
\begin{aligned}
f(v_x) &= \frac{1}{\sqrt{2\,\pi\,\overline{v_x^2}}}\exp\left[-\frac{v_x^2}{2\,\overline{v_x^2}}\right] \\
&= \frac{1}{\sqrt{2\pi\,\frac{k\,T}{m}}}\exp\left[-\frac{v_x^2}{2\,\frac{k\,T}{m}}\right] \\
&= \sqrt{\frac{m}{2\pi\,k\,T}}\exp\left[-\frac{m\,v_x^2}{2\,k\,T}\right]. \qquad (10.81)
\end{aligned}
$$

$f(v_x)\,\mathrm{d}v_x$ dà la probabilità di trovare la molecola con velocità compresa fra v_x e $v_x+\mathrm{d}v_x$. Essendo i movimenti della molecola sui tre assi indipendenti l'uno dall'altro otteniamo che la probabilità congiunta è pari al prodotto delle probabilità:

$$
\begin{aligned}
f(v_x,v_y,v_z)\,\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z &= f(v_x)\,f(v_y)\,f(v_z)\,\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_x \\
&= \left(\frac{m}{2\pi\,k\,T}\right)^{\frac{3}{2}}\exp\left[-\frac{m\,(v_x^2+v_y^2+v_z^2)}{2\,k\,T}\right]\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_x \\
&= \left(\frac{m}{2\pi\,k\,T}\right)^{\frac{3}{2}}\exp\left[-\frac{m\,v^2}{2\,k\,T}\right]\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_x \\
f(v)\,\mathrm{d}v &= 4\pi\left(\frac{m}{2\pi\,k\,T}\right)^{\frac{3}{2}}\exp\left[-\frac{m\,v^2}{2\,k\,T}\right]v^2\,\mathrm{d}v \\
f(v) &= 4\pi\left(\frac{m}{2\pi\,k\,T}\right)^{\frac{3}{2}}\exp\left[-\frac{m\,v^2}{2\,k\,T}\right]v^2 \\
&\left(=4\pi\left(\frac{m}{2\pi\,k\,T}\right)^{\frac{3}{2}}\exp\left[-\frac{E}{k\,T}\right]v^2\right).
\end{aligned}
$$

Negli ultimi passaggi siamo passati al modulo delle velocità v attraverso $\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z = 4\pi\,v^2\mathrm{d}v$ (passaggio da coordinate cartesiane a coordinate polari). Abbiamo così riottenuto in un modo euristico la distribuzione di Maxwell delle velocità, verificando in tal modo che ciascuna delle componenti della velocità ha una distribuzione di probabilità normale. La Fig. 10.9 mostra le distribuzioni di probabilità del modulo della velocità e di una delle sua componenti per alcune molecole a temperatura ambiente.

© G. D'Agostini 2001

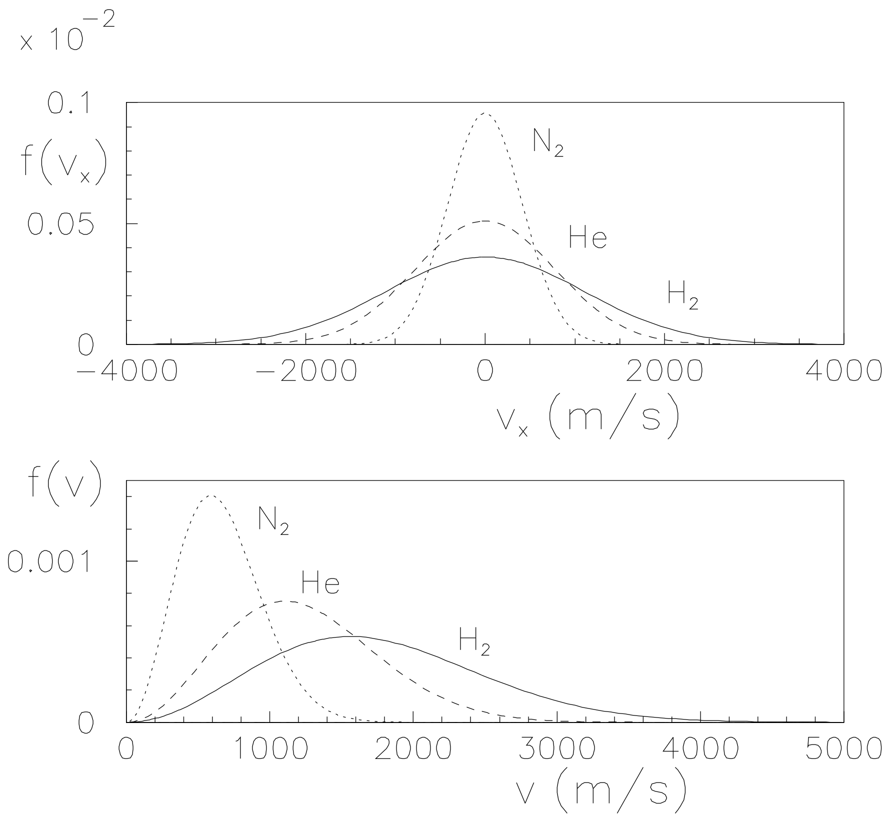

Figura 10.9: Distribuzione di velocità delle molecole di un gas perfetto

© G. D'Agostini 2001

10.17 Problemi

1. Una variabile casuale segue una distribuzione uniforme nell'intervallo da 0 a 1. Dodici realizzazioni della variabile casuale sono sommate fra di loro e il risultato viene diminuito di 6. Calcolare la probabilità che questa nuova variabile cada nell'intervallo compreso fra -1 e 1.
2. Un campione contiene 300 realizzazioni di una variabile casuale, ciascuna delle quali segue una distribuzione uniforme di parametri $a = 0$ e $b = 1$. Quanto vale la probabilità che la media aritmetica dei 300 valori sia maggiore di 0.55 ?
3. In un autogrill sull'autostrada si fermano in media fra le 12:00 e le 13:00 85 auto. Supponendo che il numero di macchine segua una distribuzione di Poisson, quanto vale la probabilità che un giorno se ne fermino 100 o più?
4. Si lancia 1500 volte un dado. Quanto vale la probabiltà che il "sei" si verifichi esattamente 250 volte?
5. Una moneta viene lanciata rispettivamente 10, 100, 1000 e 10000 volte. Calcolare la probabilità che l'occorrenza dell'evento Testa sia maggiore o uguale rispettivamente di 6, 60, 600 e 6000.
6. Si ritiene che due campioni di n risultati sperimentali indipendenti siano descritti da distribuzione di valore medio μ e deviazione standard σ. Quanto vale la probabilità che le due medie differiscano fra di loro più di $\sigma/\sqrt{n}$?
7. Le variabili casuali X_i, con $i = 1, 2, \ldots 20$, seguono una distribuzione normale di parametri $\mu_i = 5$ e $\sigma = 2$; le variabili casuali X_i, con $i = 21, 22, \ldots, 30$, seguono una distribuzione normale di parametri $\mu_i = 1000$ e $\sigma = 1$; le variabili casuali X_i, con $i = 31, 32, \ldots 60$, seguono una distribuzione uniforme nell'intervallo [-10, -2]. Definendo la variabile $Y = \sum_{i=1}^{60} X_i$, quanto vale la probabilità che Y sia minore di 9900?
8. Le variabili casuali X_i, con $i = 1, 2, \ldots 20$, seguono una distribuzione normale di parametri $\mu_i = 120$ e $\sigma = 0.1$; le variabili casuali X_i, con $i = 21, 22, \ldots 30$, seguono una distribuzione normale di parametri $\mu_i = -240$ e $\sigma = 0.02$; la variabile casuale X_{31} segue una distribuzione uniforme nell'intervallo [-100, 100]. Definendo la variabile $Y = \sum_{i=1}^{31} X_i$, quanto vale la probabilità che Y sia maggiore di -150?
9. Sul problema 31(**check**) del capitolo 7: quanto vale la probabilità di ottenere una frequenza relativa minore di 0.6 o maggiore di 0.6? Che senso ha tale "dimostrazione"?

© G. D'Agostini 2001

© G. D'Agostini 2001

22. $\mathrm{E}(X) = (\pi/4) \times n = 7854$, $\sigma(X) = 41$, $v = 0.0052$.
$v = 0.0001 \rightarrow n = 10000 \times 52^2 \approx 27 \times 10^6$.

23. Essendoci 6.02×10^{23} molecole in 24.4 litri, la previsione è pari a 2.5×10^{16} molecole. L'incertezza di previsione dovuta alla natura aleatoria del fenomeno è pari a $\sqrt{2.5 \times 10^{16}} = 1.6 \times 10^8$, con una incertezza relativa di previsione di 6×10^{-9}. Chiaramente diventano dominanti le incertezze dovute alla temperatura, pressione e determinazione del volume.

    Affinché $f(0\,|\,\mathcal{P}_\lambda) = 0.01$, λ (numero medio di molecole nel volumetto) deve essere pari a 2.3 . Ne segue $V_\circ = 9.3 \times 10^{-26}$ m^3 = $(4.5 \times 10^{-9})^3$, ovvero un cubettino di lato pari a 4.5 nm, il quale potrebbe contenere $\approx 10^5$ oggetti di 0.1 nm (tipiche dimensioni atomiche).

24. $P(X \neq 0) = 1 - f(0\,|\,\mathcal{P}_{3.6}) = 97.3\,\%$.

25. 156 $\pm$ 89 volte (distribuzione di Pascal con $p = 1/52$ e $k = 3$).

26. $\leq 2.3\,\%$ (disuguaglianza di Cebicev).

27. No: $1 - F(1\,|\,\mathcal{P}_{1.22}) = 34.5\,\%$, contro $1 - F(0\,|\,\mathcal{P}_{0.61}) = 45.7\,\%$; no.

28. 0. (Non confondere con $P(X \geq 6\,|\,\mathcal{P}_{0.61})$ = 4.2×10^{-5} !).

29. $\lambda_1 = n/365$;
$p_\circ = 1 - F(1\,|\,\mathcal{P}_{\lambda_1})$;
$p(n) = 1 - f(0\,|\,\mathcal{B}_{365,p_\circ})$.
Per esempio per $n = 23$ si ottiene $p(n) = 50.1\,\%$ invece di $50.7\,\%$ della formula esatta.

30. $f(x\,|\,\mathcal{H}_{7,3,3})$, con $x = 0, \ldots, 3$: 0.114,0.514, 0.343, 0.29 .
$v = (1/2)/\sqrt{n} < 1/50$: $\rightarrow n > 625$.

Capitolo 8

1. 48.4 %.
2. $f(x) = 2x$; $\mu = 2/3$; $\sigma = 0.24$.
3. $\mathrm{E}[(X-\mu)/\sigma] = (\mathrm{E}(X) - \mu)/\sigma = 0$.
$Var[(X-\mu)/\sigma] = \mathrm{Var}(X)/\sigma^2 = 1$.
4. $F(x)$ non può essere decrescente!
5. $k = -1$; $1 \leq x \leq 4$;
$f(x) = 1/(2\sqrt{x})$;
$\mathrm{E}(X) = 7/3 = 2.33$; $\mathrm{Var}(X) = 34/45 = 0.75$; $\sigma = 0.87$.
Come mai $\sigma \approx 3/\sqrt{12}$?
6. 0.30854; 0.09185; 0.69146; 0.34134; 0.47725; 0.9973.
7. $5.74 \cdot 10^{-7}$
8. $G(t) = 1/(1 - \lambda\, t)$; $\mathrm{E}(X) = \lambda$; $\sigma = \lambda$.
9. $\leq 2.3\,\%$
10. 20.6 %.
11. $r_{tot} = 0.25\,\mathrm{s}^{-1}$; $r_{atomo} = 0.25 \times 10^{-20}\,\mathrm{s}^{-1}$: $\tau = 4 \times 10^{20}$ s.
12. Dati del problema: $\tau = 3.14 \times 10^{17}$ s; $r = 3.17 \times 10^{-18}$; nucleoni in 70 kg: 4.1×10^{28}. Ne segue: $r_{tot} = 1.5 \times 10^{11}\,\mathrm{s}^{-1}$. Il fatto che non ci autodistruggiamo per radioattività indica che la vita media dei nucleoni è molto maggiore dell'età dell'Universo.

Capitolo 9

1. Per entrambe 4.3 e 12.3 %.
2. 7.0 (e non 6.1).
3. $\mathrm{E}(T) = \mathrm{E}(C) = n/2$; $\mathrm{Var}(T) = \mathrm{Var}(C) = n/4$; $\mathrm{Cov}(T, C) = -n/4$;
$\mathrm{E}(T - C) = 0$; $\mathrm{Var}(T - C) = n$; $\sigma(T - C) = \sqrt{n}$.

© G. D'Agostini 2001

Capitolo 10

1. E$(X) = 0$; Var$(X) = 1$; $X \sim \mathcal{N}(0,1)$:
   $P(-1 \leq X \leq 1) = 68.3\,\%$.
2. $0.14\,\%$.
3. $5.8\,\%$.
4. $2.8\,\%$.
5. 0.377, 0.029 , $\ll 10^{-7}$; $P(X \geq 6000 \,|\, X \sim \mathcal{B}_{10000,0.5})$ è talmente piccola che non è calcolabile con l'uso delle usuali tabelle.
6. $52\,\%$.
7. $10.3\,\%$.
8. 25%.

Capitolo 11

1. e la soluzione è: ....

Capitolo 12

1. e la soluzione è: ....

Capitolo 13

1. e la soluzione è: ....

© G. D'Agostini 2001