Indice

Parte II

Variabili casuali - I

Capitolo 6

Variabili casuali e distribuzioni di probabilità di variabili discrete

Questo capitolo introduce la maggior parte dei concetti relativi alle cosiddette variabili casuali, pur limitandosi al caso discreto. Le distribuzioni trattate qui, pur avendo poca rilevanza per le applicazioni pratiche legate alle incertezze di misura, hanno una grande importanza concettuale. In particolare, il processo di Bernoulli rappresenta l'elemento unificatore di molte distribuzioni, fra cui la binomiale e la poissoniana che incontreremo nel prossimo capitolo. La distribuzione geometrica ha poi molte proprietà interessanti che ne fanno una distribuzione di riferimento per scopi didattici.

6.1 Numeri aleatori

Abbiamo visto nei capitoli precedenti come il concetto di probabilità sia legato allo stato di incertezza rispetto ad eventi di cui è sconosciuto il contenuto di verità. A volte gli eventi sono associati a valori numerici, legati ad essi da relazioni fisiche o di convenienza. Lo stato di incertezza sull'evento si riflette sullo stato di incertezza sul valore della variabile, o grandezza, di interesse. L'esempio più banale è quello relativo al lancio di un dado, se, ad esempio, associamo all'evento "faccia con il numero n" il valore numerico n. Facciamo altri esempi.

- Se si lancia una moneta più volte, dopo quanti lanci si verifica "Testa" la prima volta? 1, 2, . . . ?
- Scelgo uno studente "a caso" (per esempio il 16^o di una certa lista). Quanti esami ha sostenuto? 0, 1, . . . ?
- Sapendo che lo studente ha superato l'esame di Fisica Generale, che voto ha riportato? 18, 19, . . . ?
- Si immerge un termometro digitale (con indicazione del decimo di grado) in un liquido a temperatura ambiente. Quale valore numerico leggerò? 19.0, 19.1, . . . 21.1 . . . 22.0?

- Un termometro digitale "perfettamente calibrato" e in grado di fornire il valore della temperatura al grado indica 28°C. Se avessi a disposizione un termometro con indicazione al decimo di grado, che temperatura leggerei? 27.5, 27.6,... 28.4?

- Pongo un chilogrammo campione su una bilancia di laboratorio con indicazione (digitale) dei centesimi. Che valore leggerò (in grammi)? 1000.00, 999.95, 1000.03 ...?

- Leggo su una bilancia di laboratorio 3.415 g. Quanto leggerei se ponessi lo stesso corpo su una bilancia di riferimento "di altissima precisione e perfettamente calibrata" che operi sotto vuoto?

Si evince da questi esempi che, in analogia agli eventi, una *variabile casuale* (anche nota con gli appellativi *numero casuale, variabile aleatoria*, *numero incerto* e *numero aleatorio*[1]) non è altro che *un numero rispetto al quale si è in stato di incertezza.* Anche in questo caso la maggiore o minore fiducia del verificarsi del numero verrà misurata dalla probabilità.

6.2 Distribuzione di probabilità

Se indichiamo con X (lettera maiuscola) la variabile e con x i valori[2] che può assumere, associamo, all'evento $X = x$ la probabilità $P(X = x)$ sulla base del nostro stato di conoscenza. La variabile casuale X è detta *discreta* se può assumere un numero finito (o una infinità numerabile) di valori[3]. Poiché in molti problemi è possibile trovare delle funzioni matematiche più o meno semplici con le quali descrivere le probabilità di tutte le occorrenze di X, è usuale indicare[4] tale probabilità con $f(x)$. In sintesi:

- X indica la variabile casuale (esempio "valore che verrà indicato sul display della bilancia");

- x è il valore numerico che può assumere la variabile, con grado di fiducia $P(X = x)$;

- $f(x) = P(X = x)$ indica la funziona matematica mediante la quale assegnamo la probabilità di tutti i possibili valori x di X.

[1] Questo è forse il nome che rende meglio l'idea. In questo testo si è preferito utilizzare come nome standard "variabile casuale" in quanto è la denominazione più usuale fra i fisici. Esso però rischia di prestarsi ad interpretazione troppo legate ai risultati di esperimenti ripetuti e non, più in generale, allo stato di incertezza. Cercheremo comunque di usare i due termini come sinonimi, insistendo su "numero aleatorio", o 'numero incerto"..

[2] In altri testi le variabili discrete che assumono valori interi vengono di preferenza indicate con i, j, k o n, mentre la lettera x è usata solo per variabili continue. In questo testo viene generalmente usato lo stesso simbolo x sia per variabili discrete che continue.

[3] Si noti quindi che discreto non significa necessariamente valore intero, anche se questo sarà il caso più frequente nelle applicazioni semplici che tratteremo. Ad esempio, se si associa ad ogni faccia del dado la variabile casuale X = "radice quadrata del numero impreso sulla faccia", si ottiene una variabile discreta a valori reali.

[4] In altri testi si preferisce $p(\cdot)$ a $f(\cdot)$, a ricordare che essa ha il significato di probabilità. Altre volte ancora si trova, ad esempio, p_k al posto di $p(k)$.

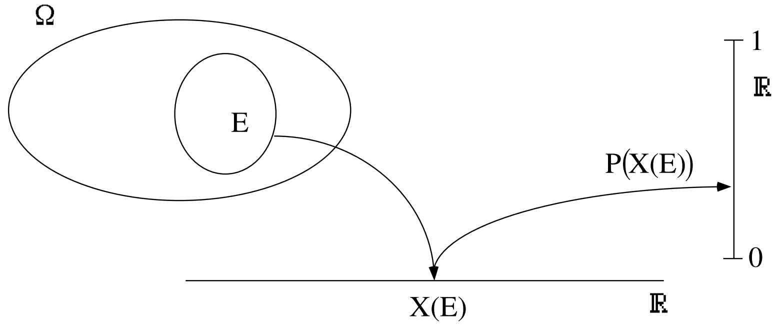


Figura 6.1: Costruzione di una variabile casuale.

- quando sono specificati i valori x che la variabile X può assumere, insieme alla *funzione di probabilità* $f(x)$, si parla di *distribuzione di probabilità*.

Nel seguito vedremo alcune distribuzioni di probabilità. Anche se talvolta si arriverà a formule che possono sembrare complicate, bisogna tenere conto che lo schema di valutazione della probabilità è lo stesso presentato nei capitoli precedenti:

- in base allo stato di conoscenze si assegna un certo valore di probabilità a eventi elementari;
- usando le regole della probabilità si calcolano le probabilità di eventi più complicati ottenuti da operazioni logiche sui primi;
- la sola condizione importante per la costruzione della variabile casuale è che ad ogni evento considerato sia associata sempre e univocamente (ma non necessariamente biunivocamente) una variabile numerica, come mostrato in figura 6.1. Questo sta ad indicare che l'insieme degli eventi "X assume il valore x" costituisce, quando si considerano tutti i possibili valori di x, una classe completa.

Gli esempi dei prossimi paragrafi dovrebbero chiarire gli eventuali dubbi.

6.3 ○ Distribuzione di probabilità e distribuzioni statistiche

Prima di intraprendere lo studio sistematico delle distribuzioni di probabilità, è opportuno chiarire la differenza fra queste distribuzioni e le *distribuzioni statistiche*. Esse sono infatti legate da molti aspetti comuni, come una analoga terminologia per gli indicatori di centralità e di dispersione (concetti che saranno introdotti nei paragrafi 6.8, 6.9 e 6.11) e analoghe rappresentazioni grafiche.

Per chiarire meglio cosa si intende con i due tipi di distribuzioni, è opportuno fare un accenno alla differenza fra *statistica descrittiva* e *statistica inferenziale*. Infatti il termine stesso *statistica* viene usato in vari contesti e a volte, non del tutto propriamente, anche come sinonimo di probabilità.

Senza voler entrare nei dettagli, diciamo che bisogna distinguere le *statistiche*, di cui si parla in continuazione, dalla *statistica*. Le "statistiche" stanno ad indicare sintesi di dati su aspetti sociali, economici, politici, geografici, e così via (*"le statistiche dicono"* che "il ... % della popolazione è ultrasessantenne", che "questa è l'estate più calda degli ultimi n anni", che "il ... % delle coppie divorzia nei primi 5 anni di matrimonio", etc.).

Con *statistica* si intende invece la disciplica che, in senso lato, si interessa della raccolta e l'analisi dei dati e dell'interpretazione dei risultati. In particolare, la *statistica descrittiva* si occupa di descrivere la massa dei dati sperimentali con pochi numeri o grafici significativi. Quindi, per così dire si occupa di fotografare una data situazione e di sintetizzarne le caratteristiche salienti. La *statistica inferenziale* utilizza i dati statistici, generalmente sintetizzati dalla statistica descrittiva, per fare previsioni di tipo probabilistico su situazioni future o comunque incerte. Ad esempio esaminando un piccolo campione estratto da una grande popolazione si può cercare di valutare la frazione della popolazione che possiede una certa caratteristica, ha un certo reddito, voterà per un certo candidato.

Per quello che riguarda la teoria e la pratica delle misure, indubbiamente quella più interessante è la statistica inferenziale in quanto scopo delle misure è quello di fare affermazioni sul valore di una grandezza o sulla validità di una teoria a partire da un numero limitato di misure, effettuate con strumenti non ideali, con parametri e disturbi ambientali della cui entità non si è assolutamente certi.

Anche la statistica descrittiva ha una sua importanza, in quanto nella maggior parte dei casi non è necessario conoscere il dettaglio di tutti i dati sperimentali raccolti per inferire qualcosa, ma spesso sono sufficienti pochi numeri nei quali i dati sono stati precedentemente sintetizzati. Quindi, lo schema di massima che si usa nella statistica inferenziale è formato dai seguenti passi:

1. raccolta dei dati sperimentali;

2. sintesi statistiche (statistica descrittiva);

3. inferenza (affermazioni probabilistiche);

Come si può immaginare, questa classificazione è artificiosa ed è difficile separare i tre stadi. Ad esempio, è difficile raccogliere dati statistici su un campione della popolazione se non si ha nessuna idea della caratteristiche della popolazione stessa (si pensi agli "exit poll"), oppure fare delle misure i cui risultati siano utilizzabili se non si conosce la fenomenologia sulla quale si sta indagando con tutti gli effetti sistematici. Infatti, il primo punto racchiude tutta l'arte della sperimentazione, a partire dalla conoscenza della fenomenologia e degli strumenti, alla progettazione, realizzazione e conduzione dell'esperimento. Così pure, alcune grandezza di sintesi di dati statistici sono costruite già pensando ad un successivo uso inferenziale (si pensi alla deviazione standard di una distribuzione statistica calcolata dividendo la somma dei quadrati degli scarti per $n-1$ invece di n). Tornando alle distribuzioni, possiamo dire che la differenza sostanziale fra i due tipi è che, mentre le distribuzioni di probabilità, fanno riferimento a variabili casuali, ovvero a numeri rispetto ai quali siamo in stato di incertezza, le distribuzioni statistiche descrivono *variabili statistiche*, ovvero occorrenze certe nel passato di determinati valori (o classi di valori). In sintesi:

- *le distribuzioni di probabilità associano ad ogni valore una funzione che esprime il grado di fiducia sul suo realizzarsi*

- *le distribuzioni statistiche associano ad ogni valore (o classe di valori) un peso statistico pari alla frequenza relativa con cui esso si è verificato nel passato.*

Ovviamente, come le frequenze di eventi giocano un ruolo importante nella valutazione della probabilità, così le distribuzioni statistiche hanno una analoga importanza nella valutazione delle distribuzioni di probabilità, anche se, come vedremo (già a partire dal paragrafo 7.15), a nessuna persona ragionevole dovrebbe venire in mente di affermare che la distribuzione di probabilità è data esattamente dalla distribuzione statistica osservata.

Altri commenti su differenze e analogie fra i due tipi di distribuzione verranno fatti nel paragrafo 6.18.

6.4 Esempi di costruzione di distribuzioni di variabili casuali

Facciamo alcuni esempi di costruzione di una variabile casuale a partire da eventi elementari.

- Lancio di **un dado**, con X = "numero impresso sulla faccia superiore":

$$f(x) = 1/6 \qquad x = 1, 2, \ldots, 6. \tag{6.1}$$

- Lancio di **due dadi**, con X = "somma dei due risultati": $f(2) = 1/36$, $f(3) = 2/36$, etc (vedi figura 6.2). Volendo un'espressione sintetica ("forma chiusa", nel gergo matematico) per le probabilità (non è assolutamente necessario e non sempre fattibile) si può scrivere

$$f(x) = \frac{6 - |7 - x|}{36} \qquad x = 2, 3, \ldots, 12\,. \tag{6.2}$$

- Consideriamo un esperimento consistente nel lancio di **tre monete** regolari. Gli eventi elementari e_i di Ω sono costituiti da tutte le possibili sequenze Testa/Croce: Ω = { TTT, TTC, ..., CCC }. Ognuno di questi eventi ha la stessa probabilità di verificarsi. A partire da questi eventi possiamo costruire più variabili casuali, ad esempio il numero di teste (X), il numero di croci (Y), il numero di teste consecutive (Z), etc. Riportiamo nella seguente tabella gli eventi elementari, insieme al valore che acquistano le variabili casuali costruite su di essi.

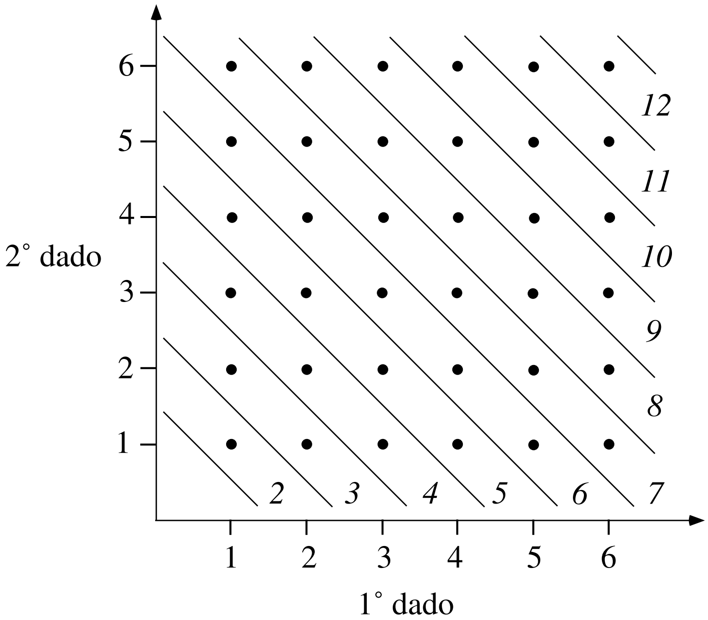


Figura 6.2: Spazio delle possibilità relativo al lancio di due dadi e costruzione della variabile casuale "somma".

e_i	$P(e_i)$	X	Y	Z
TTT	1/8	3	0	2
TTC	1/8	2	1	1
TCT	1/8	2	1	0
TCC	1/8	1	2	0
CTT	1/8	2	1	1
CTC	1/8	1	2	0
CCT	1/8	1	2	0
CCC	1/8	0	3	0

Le distribuzioni di probabilità per X, Y e Z sono le seguenti[5]:

[5]Dal punto di vista formale, la probabilità, per esempio, di Z è valutata dalla legge delle alternative (cfr. par. 4.9.2):

$$\begin{aligned} P(Z=z) &= \sum_i P(Z=z\,|\,e_i)\cdot P(e_i) \\ &= \sum_{\{Z(e_i)=z\}} P(e_i)\,, \end{aligned}$$

ovvero la somma delle probabilità di tutti gli eventi di una classe completa per i quali la variabile Z vale z. Il passaggio dalla (6.3) alla (6.3) si basa sul fatto che essendo la regola di costruzione della variabile casuale univoca, $P(Z=z\,|\,e_i)$ può valere soltanto 0 o 1.

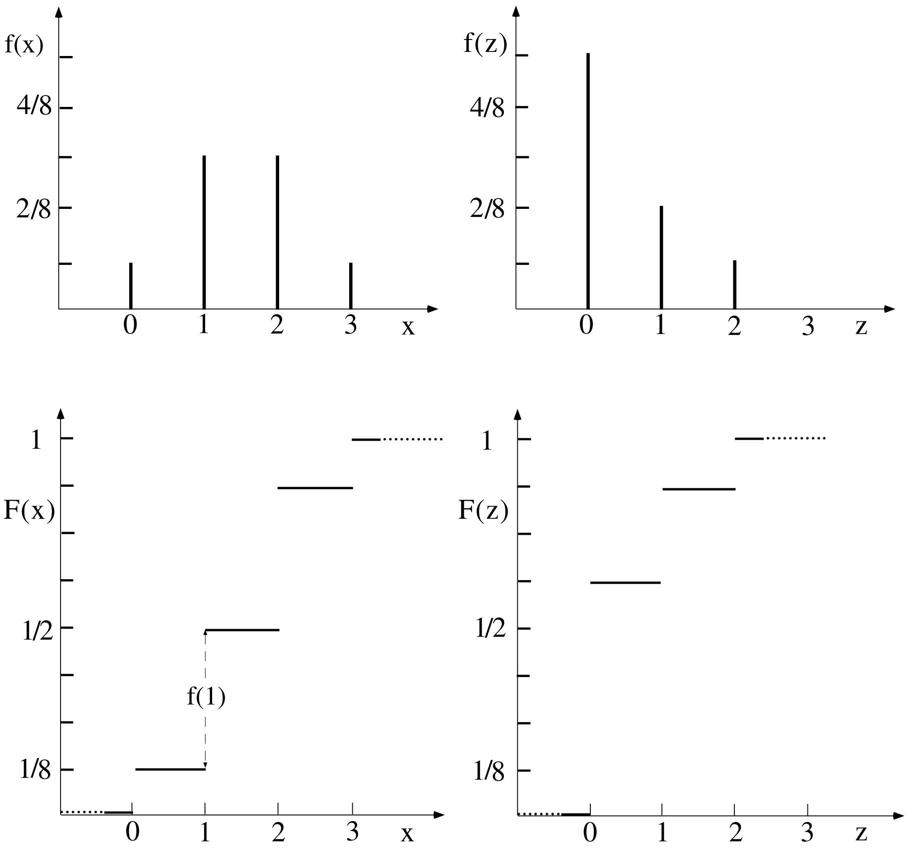


Figura 6.3: Rappresentazione grafica di probabilità, della funzione di probabilità discreta e della funzione di ripartizione. Per $x = 1$ è mostrato esplicitamete come il salto della $F(x)$ sia pari al valore di $f(x)$ in quel punto.

x	$f(x)$
0	1/8
1	3/8
2	3/8
3	1/8

y	$f(y)$
0	1/8
1	3/8
2	3/8
3	1/8

z	$f(z)$
0	5/8
1	2/8
2	1/8

La figura 6.3 mostra la rappresentazione grafica, mediante diagrammi a barre, delle distribuzioni di probabilità delle variabili X e Z.

6.5 Proprietà delle distribuzioni di probabilità discrete

La funzione di probabilità $f(x)$, avendo il significato di probabilità di eventi di una classe completa, deve soddisfare le seguenti condizioni:

$$1)\qquad 0 \le f(x) \le 1\,; \tag{6.3}$$

$$2)\qquad P(X = x_i \cup X = x_j) = f(x_i) + f(x_j)\,; \tag{6.4}$$

$$3)\qquad \sum_i f(x_i) = 1\,. \tag{6.5}$$

(Si faccia attenzione all'uso flessibile degli indici. A volte, per alleggerire la notazione verranno omessi. Altre volte, quando la variabile X può assumere valori che differiscono fra loro di una unità, verrà utilizzato lo stesso simbolo x come indice delle sommatorie. Inoltre, gli estremi delle sommatorie sono spesso omessi per indicare che, implicitamente, sono considerati tutti i valori possibili. Ad esempio, la 3) potrebbe essere scritta più sinteticamente come $\sum_x f(x) = 1$.)

La proprietà 2) deriva dal fatto che valori diversi della realizzazione di una variabile casuale sono incompatibili. La condizione 3) è anche chiamata *condizione di normalizzazione* (si dice che "la distribuzione di probabilità è normalizzata ad 1"). La somma è da intendersi estesa a tutti i possibili valori che può assumere X.

In alcuni casi può avere interesse trovare la probabilità che la variabile casuale X assuma un valore uguale o minore di un certo x_k. Si introduce allora il concetto di *probabilità cumulativa*, descritta dalla *funzione di ripartizione* $F(x_k)$:

$$F(x_k) \equiv P(X \le x_k) \;=\; f(x_1) + f(x_2) + ... + f(x_k) = \sum_{x_i \le x_k} f(x_i)\,. \tag{6.6}$$

È comodo poter estendere la somma a tutti i valori reali di x, sottintendendo che, al di fuori del campo di definizione di $f(x)$, $F(x)$ è pari al valore che essa assume per x_i immediatamente inferiore ad x e per il quale la $f(x_i)$ sia definita[6].

La funzione di ripartizione gode delle seguenti proprietà che discendono direttamente dalla definizione:

$$1)\qquad 0 \le F(x) \le 1 \tag{6.7}$$

$$2)\qquad \lim_{x\to-\infty} F(x) = 0 \tag{6.8}$$

$$3)\qquad \lim_{x\to+\infty} F(x) = 1 \tag{6.9}$$

$$4)\qquad F(x_i) - F(x_{i-1}) = f(x_i) \tag{6.10}$$

$$5)\qquad \lim_{\epsilon\to o} F(x+\epsilon) = F(x) \qquad \text{continuità a destra).} \tag{6.11}$$

La Fig. 6.3 mostra anche le funzioni di ripartizione delle variabili X e Z definite nel paragrafo precedente.

[6] Ad esempio, nel caso del valore ottenuto nel lancio di un dado, $F(3.5) = P(X \le 3.5) = P(X \le 3) = 0.5$. Anche se a qualcuno potrà sembrare strano che ci si possa interessare di $P(X \le 3.5)$ nel lancio dei dadi è fuori di dubbio l'espressione probabilistica sia corretta, come lo sarebbe $P(X \le 100) = 1$ o $P(X < 0) = 0$.

6.6 Distribuzioni elementari notevoli

6.6.1 Distribuzione uniforme discreta

La più semplice distribuzione di interesse generale è quella in cui si assegna lo stesso grado di fiducia a tutte le possibili realizzazioni di X. Essa è chiamata *distribuzione uniforme* (discreta). Prendiamo per semplicità una variabile casuale che può assumere con uguale probabilità i primi n interi positivi. Otteniamo la funzione

$$f(x\,|\,\mathcal{K}_{1,n}) = \frac{1}{n} \qquad x = 1, 2, \ldots, n\,. \tag{6.12}$$

Introduciamo la notazione $f(x\,|\,\cdot)$ per indicare che la funzione di probabilità è condizionata da una certa distribuzione. Qui $\mathcal{K}$ sta per *"costante"* e "$1, n$" sono i *parametri* della distribuzione. In genere, per indicare sinteticamente che una variabile segue una certa distribuzione di probabilità, useremo il simbolo "$\sim$". Ad esempio, in questo caso:

$$X \sim \mathcal{K}_{1,n} \Longleftrightarrow f(x\,|\,\mathcal{K}_{1,n})\,.$$

Valori diversi da 1, 2, ..., n sono impossili ed è quindi pari a zero la funzione di probabilità in corrispondenza di essi. A volte, soprattutto in trattazioni di impostazione più matematico, questo viene esplicitato nella definizione, scrivendo, ad esempio:

$$f(x\,|\,\mathcal{K}_{1,n}) \;=\; \begin{cases} \frac{1}{n} & x = 1, 2, \ldots, n \\ 0 & \text{altrimenti} \end{cases} \quad .$$

La funzione di ripartizione è "a gradini equidistanziati" e, per x interi compresi fra 1 e n, assume i valori

$$F(x\,|\,\mathcal{K}_{1,n}) = \frac{x}{n}\,, \tag{6.13}$$

mentre di altri valori sono ottenuti dalla definizione di $F(x)$.

La figura 6.4 mostra la rappresentazione a diagramma a barre di questo tipo di distribuzione, per $n = 6$. I simboli μ e σ stanno ad indicare grandezze atte a misurare in modo convenzionale ("standard") la posizione e la larghezza della distribuzione e saranno descritte a partire dal paragrafo 6.8. Esempi banali di questa distribuzione sono le variabili associate alle faccie di un dado (figura 6.4) o ai numeri della tombola. Un esempio più vicino alla problematica della misura è quello della prima cifra decimale che verrà indicata da un termometro digitale se si misura un fluido "a temperatura ambiente".

6.6.2 ❄Distribuzione uniforme discreta - caso generale

La distribuzione definita sui prini n interi positivi può essere estesa a n valori compresi fra a e b (anche reali) e distanziati di $\Delta = (b-a)/(n-1)$:

$$f(x\,|\,\mathcal{K}_{a,b,n}) = \frac{1}{n}\,, \qquad x = a, a+\Delta, a+2\Delta, \ldots, b\,. \tag{6.14}$$

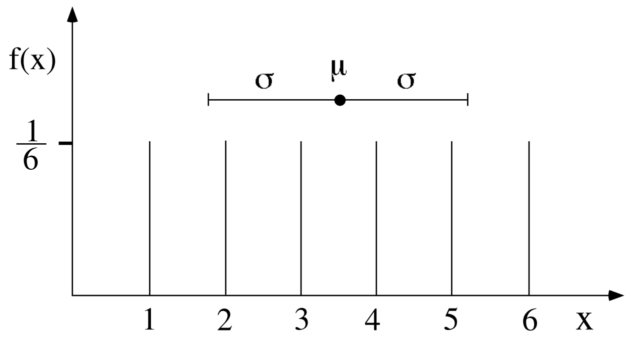


Figura 6.4: Distribuzione uniforme discreta ottenuta associando dalla variabile casuale X = "esito del lancio di un dado". Il significato di μ e σ sarà illustrato nel seguito.

La funzione di ripartizione vale

$$F(x \,|\, \mathcal{K}_{a,b,n}) = \frac{1}{n}\left[1 + \frac{x-a}{\Delta}\right] \qquad (6.15)$$

nei punti in cui è definita $f(x)$ (e i valori negli altri punti sono ottenuti come al solito).

6.6.3 Processo di Bernoulli

Possiamo pensare al valore di verità di un evento, definito come 1 se è vero e 0 se è falso, come una variabile casuale (nel paragrafo 2.17 avevamo introdotto a tale scopo l'indicatore dell'evento $|E|$). La distribuzione che ne segue è molto semplice, ma essa è di grande importanza perché illustra una schematizzazione che descrive molti casi di interesse, come vedremo fra breve. Per questa ragione tale schematizzazione è nota sotto nome proprio: si parla di *processo di Bernoulli*. A partire da tanti processi "elementari" di questo tipo si possono costruire altre distribuzioni largamente usate.

Schematizzando, il processo di Bernoulli consiste nell'effettuare una prova nella quale

- si valuta in p la probabilità di un certo evento, chiamato convenzionalmente *successo*; di consequenza, $q \equiv 1 - p$ è la probabilità di *insuccesso*; il ruolo di successo e insuccesso sono assolutamente arbitrari e quindi tutti i ragionamenti saranno simmetrici; è però importante fare attenzione alla convenzione utilizzata;

- la variabile assume il valore 1 se l'evento si verifica e 0 se esso non si verifica.

Ne segue che

$$\begin{aligned} f(1) &= p & (6.16)\\ f(0) &= 1 - p = q\,, & (6.17)\end{aligned}$$

o, in generale,

$$f(x \,|\, \mathcal{B}_p) = p^x q^{1-x} \qquad \left\{ \begin{array}{l} 0 \le p \le 1 \\ x = 0, 1 \end{array} \right. \qquad (6.18)$$

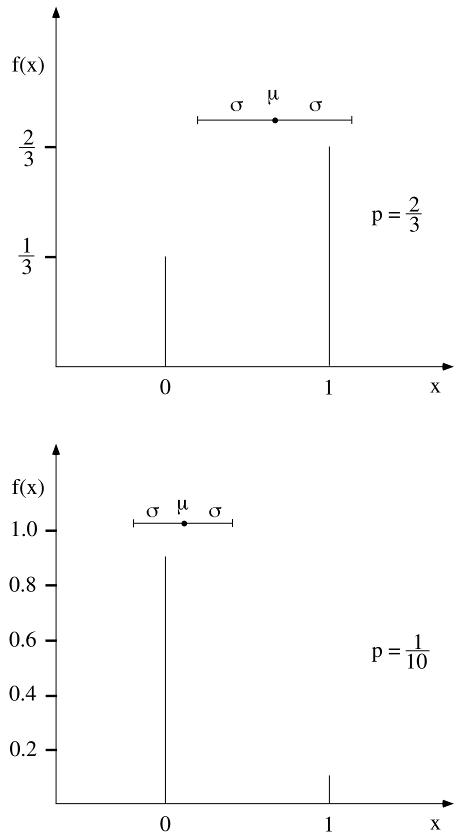


Figura 6.5: Distribuzione di Bernoulli per $p = 2/3$ e $p = 1/10$.

6.6.4 Combinazione di molti processi di Bernoulli indipendenti e di uguale probabilità

Molto spesso si è interessati a eventi "dello stesso tipo"[7], ai quali assegnamo la stessa probabilità e che consideriamo stocasticamente indipendenti, ad esempio E_i = "testa all'i-mo lancio di una moneta" (o, equivalentemente, "…per l'i-ma moneta numerata").

$$P(E_i) = p\ \forall i\,. \tag{6.19}$$

Considerando processi di Bernoulli indipendenti, si può essere interessati a due domande tipiche:

1. quante prove bisogna effettuare affinché si verifichi per la prima volta un successo? (Più esattamente: "quanto vale la probabilità che il successo si verifichi per la prima volta alla prova X?")
2. se si effettuano n prove, quanto vale la probabilità che si verifichino X successi?

[7]Ma non è corretto parlare dello stesso evento. Infatti, ogni evento, come affermazione sul verificarsi di un qualche accadimento è unico e irripetibile. Quindi se si lancia 100 volte una moneta si possono considerare gli eventi "testa al primo lancio", "testa all'i-mo lancio", "nessuna testa nei primi 5 lanci", eccetera, ma non "l'evento testa".

Le due schematizzazioni corrispondono rispettivamente alle cosiddette distribuzioni *geometrica* e *binomiale*. Un altro problema che può avere un certo interesse, ma di minore rilevanza rispetto agli altri per le applicazioni che presenteremo in questo testo, è:

3. quanto vale la probabilità che il k-mo successo si verifichi esattamente alla prova X?

Ad esso è associata la distribuzione di *Pascal* e, in modo complementare, la *binomiale negativa*.

6.7 Distribuzione geometrica

Trattiamo prima la distribuzione geometrica, con un esempio che mette in luce alcuni aspetti intuitivi e altri aspetti controintuitivi legati alle distribuzioni di probabilità.

Esempio Un ubriaco deve aprire la porta di casa al buio e ha un mazzo di 8 chiavi "indistinguibili" (o che sembrano tali a lui). Ammettiamo che, nel caso che egli non ci riesca in un tentativo, la concitazione e lo stato di annebbiamento gli impediscano di ricordare con quale chiave abbia provato e quindi si ritrovi nelle stesse condizioni nel tentativo successivo. Cerchiamo di rispondere a queste domande:

- quanti tentativi si prevede che gli serviranno affinché riesca ad aprire la porta?
- se si dovesse fare una scommessa alla pari per vincere ("non coerente") su quale tentativo bisognerebbe puntare?

Esempi analoghi Altri esempi schematizzabili nello stesso modo sono: la prima volta che viene testa nel lancio di una moneta ($p = 1/2$); la prima volta che esce il 5 lanciando un dado ($p = 1/6$) e la prima volta che esce un numero su una certa ruota del lotto ($p = 1/18$).

È importante a confrontare fra loro i problemi appena proposti prima di provare a rispondere intuitivamente alle domande formulate a proposito del problema dell'ubriaco. Le risposte *intuitive* possono essere del tipo:

- "passando, in ordine, dal problema della moneta per terminare a quello del lotto, bisogna considerare più tentativi prima di sperare ragionevolmente in un successo";
- "scommetterei intorno al 2° tentativo per la moneta, intorno all' 8° per l'ubriaco e intorno al 18° per il numero al lotto".

Ricaviamoci la funzione di probabilità e confrontiamola con le risposte intuitive:

$$\begin{aligned}
f(1) &= P(E_1) = p \\
f(2) &= P(\overline{E}_1) \cdot P(E_2 \,|\, \overline{E}_1) = (1-p)\,p \\
f(3) &= P(\overline{E}_1) \cdot P(\overline{E}_2 \,|\, \overline{E}_1) \cdot P(E_3 \,|\, \overline{E}_1 \cap \overline{E}_2) = (1-p)^2 p \\
\ldots & \quad \ldots \\
f(x \,|\, \mathcal{G}_p) &= (1-p)^{x-1} p \qquad \left\{ \begin{array}{l} 0 \le p \le 1 \\ x = 1, 2, \ldots \infty \end{array} \right.
\end{aligned} \tag{6.20}$$

Questa distribuzione è chiamata *geometrica* in quanto la funzione di probabilità è caratterizzata da tale progressione. Verifichiamo, come esercizio, che la funzione di probabilità soddisfa alla condizione di normalizzazione. Infatti:

$$\begin{aligned}
\sum_{x=1}^{\infty} f(x \,|\, \mathcal{G}_p) &= \sum_{x=1}^{\infty} (1-p)^{x-1}\, p \\
&= p \sum_{x=1}^{\infty} q^{x-1} \\
&= p \, \frac{1}{1-q} = 1 \, .
\end{aligned} \tag{6.21}$$

La funzione cumulativa $F(x)$ può essere ottenuta direttamente dalla definizione:

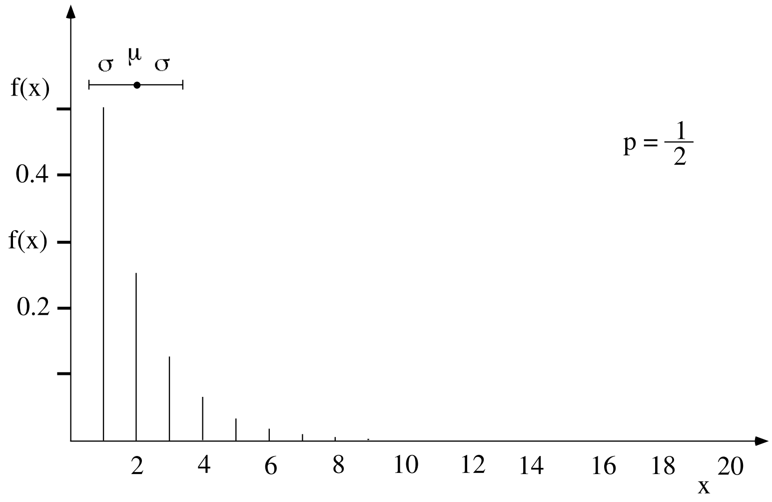
$$\begin{aligned}
F(x \,|\, \mathcal{G}_p) \equiv P(X \le x) &= 1 - P(X > x) \\
&= 1 - P(\overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_x) \\
&= 1 - (1-p)^x \qquad \text{per } x = 1, 2, \ldots
\end{aligned} \tag{6.22}$$

Essa vale 0 per $x < 1$, mentre per i reali > 1 ha il valore che assume in corrispondenza del numero intero immediatamente inferiore. Si vede che, come deve essere, per $x \to \infty$, $F(x)$ tende a 1. La figura 6.6 mostra la distribuzione geometrica per $p = 1/2$ e $p = 1/8$.

Per molti potrà essere una sorpresa scoprire che il massimo di probabilità è in corrispondenza di $x = 1$, indipendentemente da p. Quindi la seconda domanda posta riguardo al problema dell'ubriaco è in un certo senso controintuitiva. Questo è dovuto ad una confusione fra "il valore che ci aspettiamo" e "il valore più probabile". In effetti, anche se si accetta il fatto che la prova alla quale si crede di più che si verifichi il successo sia la prima, e che, per avere l'assoluta certezza, bisogna considerare un infinito numero di prove, permane ancora l'idea che il successo *è atteso* prima nel caso di lancio di una moneta che in quello di singolo estratto al lotto. E in effetti questa volta l'intuizione è corretta, a parte quantificare meglio cosa si intende per *previsione* di un numero aleatorio.

6.8 Sintesi di una distribuzione di probabilità: previsione e incertezza di previsione

La distribuzione di probabilità contiene tutte le informazioni dettagliate sullo stato di incertezza rispetto al numero aleatorio di interesse, contenendo, infatti,



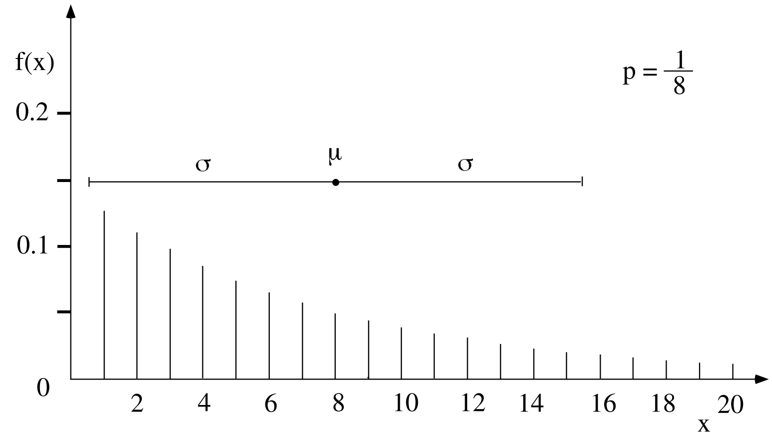


Figura 6.6: Distribuzione geometrica per p uguale a 1/2 e a 1/8.

il nostro grado di fiducia su ciascuno dei valori che tale numero può assumere. Cerchiamo ora di definire delle grandezze che abbiano la capacità di riassumere in modo immediato e sintetico alcune caratteristiche della distribuzione. Detto alla buona, esse rispondono alla domanda: "quali valori ci aspettiamo che si verifichino?" Prima di mostrare i criteri per definire operativamente tali grandezze, aiutiamoci con l'intuito su alcuni esempi, al fine di capire meglio a quale domanda "naturale" stiamo cercando di rispondere.

Abbiamo visto come, nel caso della distribuzione geometrica, il valore più probabile non dia l'idea dei valori che tipicamente si verificano. Lo stesso vale per l'incertezza di previsione, ovvero per la dispersione dei valori che si possono ragionevolmente presentare. Il fatto che per avere la certezza assoluta bisogna considerare un numero infinito di lanci non induce ad affermare che "l'incertezza" è infinita e indipendente da p. Come noto, nel caso del lancio di una moneta, una fluttuazione di tre o quattro estrazioni, rispetto alle aspettative di un circa successo ogni due, tende già ad essere considerata rara. Nel caso del singolo estratto al lotto invece, si tendono a considerare le fluttuazioni "anomale" quando il numero di estrazioni senza successo supera le 50-60. Ciò sta ad indicare che l'incertezza di previsione è ritenuta, nel secondo caso, dell'ordine delle decine di estrazioni.

Facciamo un esempio con più lanci di una moneta:

- Se lanciamo una moneta 2 volte, ci aspettiamo che il numero di volte in cui si verifica testa sarà intorno a 1. Nessuno si stupisce se esso sarà 0 o 2. Se lanciamo la moneta 1000 volte ci aspettiamo circa 500 teste, ma riteniamo molto improbabili i valori estremi di 0 e di 1000. Quindi mentre qualitativamente il primo caso può essere riassunto in una previsione di $1\pm\approx 1$, nel secondo caso una previsione di $500\pm\approx 500$ è decisamente pessimistica.

- Nel caso poi che la moneta venga lanciata invece un numero dispari di volte, ci "aspettiamo" un numero frazionario di successi. Ad esempio una previsione di 2.5 teste su 5 lanci rende bene l'idea, anche se nessuno penserà mai di osservare esattamente 2.5 successi!

Ancora degli esempi ispirati alle immancabili urne:

- Considerando un'urna contente 1000 palline di cui 500 bianche e le restanti nere. Se estraiamo, e successivamente reintroduciamo, 5 palline bianche, le nostre aspettative sul numero di palline bianche sono analoghe al caso precedente del lancio di una moneta. In particolare ci aspettiamo che le combinazioni più asimmetriche (5 bianche o 5 nere) siano le meno probabili. Per ragioni di simmetria ci aspettiamo che la probabilità massima sia per $X = 2$ e $X = 3$. Quindi un modo alternativo di fornire la previsione potrebbe essere quello di dire "i valori che mi aspetto di più sono 2 e 3". In particolare, la probabilità che X sia ≤ 2 è uguale a quella di $X \geq 3$ (ne segue che il punto intermedio fra 2 e 3 è quello che divide i possibili valori di X in due regioni di pari probabilità). Quindi anche questi criteri indicherebbero una previsione dei valori "fra 2 e 3".

- Si può però immaginare facilmente (il modo standard di fare i conti sarà mostrato parlando della binomiale) che cambiando leggermente la composizione dell'urna (ad esempio 499 bianche, poi 498, 490, etc) dapprima ci sarà una variazione brusca dell'indicazione fornita da questa definizione operativa dell'incertezza (nell'esempio fatto passerà a 2), per diventare poi largamente insensibile all'esatta composizione dell'urna, per poi subire un'altra variazione brusca verso il valore 1, e così via.

Riassumendo:

- volendo definire qualitativamente i concetti di previsione e di incertezza di previsione si può dire che[8]:

[8]❄ Questa definizione del concetto di incertezza è ottenuta parafrasando la definizione ISO di incertezza di misura (vedi anche paragrafo 2.11.

la previsione è un parametro che indica il valore intorno al quale possa ragionevolmente verificarsi il numero aleatorio;

l'incertezza di previsione è un parametro che caratterizza la dispersione dei valori che possono essere ragionevolmente assunti dal numero casuale;

- per quanto concerne la definizione operativa di previsione, gli esempi hanno mostrato che né il valore più probabile (chiamato *moda*) né quello che divide i possibili valori della variabili in due classi ordinate di pari probabilità (chiamato *mediana*) si prestano a caratterizzare il concetto espresso;

- analogalmente, associare l'incertezza all'ampiezza dell'intervallo di valori che il numero aleatorio può assumere conduce a sovrastime dell'incertezza e ad una insensibilità dal tipo di distribuzione, sia per variabili definite su un'intervallo finito che infinito.

Inoltre:

- in analogia a parametri che caratterizzano le cosidette *posizione* e *dispersione* dei numeri aleatori, ce ne sono altre che misurano la forma della distribuzione in modo più dettagliato di quanto non possa fare l'incertezza di previsione (legata alla larghezza). Ad esempio, si può essere interessati al grado di asimmetria della distribuzione rispetto al suo centro (vedi paragrafo 6.13).

6.9 Previsione (o valore atteso) come baricentro della distribuzione

Abbiamo visto nel paragrafo 2.14 come il concetto di previsione (o speranza matematica, o valore atteso) di guadagno riesca a caratterizzare un problema di decisione (ad esempio un gioco d'azzardo) senza conoscere i dettagli del problema (le regole del gioco). Possiamo estendere questo concetto alle variabili casuali e definire, in analogia della (2.17), la *previsione di una variabile casuale* come la somma dei valori della variabile casuale moltiplicati per la loro probabilità[9]:

$$\mathrm{P}(X) \;=\; \sum_i x_i f(x_i)\,. \tag{6.23}$$

In effetti, non si tratta soltanto di una analogia formale in quanto anche nel caso della (2.17) si poteva parlare di numeri aleatori[10] S_{N_i}, ciascuno con il

[9]Si noti il diverso uso delle lettere maiuscole e minuscole, consistenti con le definizioni introdotte.

[10]❄Si noti che la richiesta di previsione nulla in caso di valutazioni coerenti di probabilità è valida soltanto se i numeri casuali hanno il significato di guadagni netti (con segno) associati ad ogni scommessa. Ad esempio, scommettendo alla pari 1000 lire nel lancio di una moneta, la previsione di guadagno è nulla, mentre, associando all'evento testa il valore $X = 1$ e all'evento croce $X = 0$, la previsione della variabile casuale X è 1/2.

suo grado di fiducia $P(E_i)$. Nel seguito preferiremo indicare la previsione con il simbolo $\mathrm{E}(\cdot)$, che ricorda il nome "valore atteso" (inglese *expected*) e che, trattando ora di variabili casuali, non si confonde più con il generico simbolo di evento[11]:

$$\mathrm{E}(X) \equiv \mathbb{P}(X) \quad = \quad \sum_i x_i f(x_i)\,. \tag{6.24}$$

Per capire meglio il significato di $\mathrm{E}(X)$, riscriviamo la (6.24) come

$$\mathrm{E}(X) \quad = \quad \frac{\sum_i x_i f(x_i)}{\sum_i f(x_i)} \tag{6.25}$$

(operazione consentita in quanto $\sum_i f(x_i) = 1$), rendendo esplicito il fatto che la previsione di una variabile casuale rappresenta la *media pesata* dei valori delle variabili casuali, con peso pari alla sua probabilità.

Con una espressione mutuata dalla meccanica, possiamo affermare che $\mathrm{E}(X)$ rappresenta il *baricentro* della distribuzione di probabilità. Si riconosce infatti nella (6.25) la coordinata del centro di massa di un sistema di punti, ciascuno avente una "massa di probabilità" $f(x_i)$. Questa constatazione è una ulteriore giustificazione dell'uso di tale definizione per quantificare il valore intorno al quale ci aspettiamo che la variabile casuale assuma il valore. Altre motivazioni che giustificano l'adozione di tale definizione operativa di previsione verranno indicate nel seguito (ad esempio paragrafi 6.5 e 10.7).

Prima di passare alle applicazioni, ricordiamo ancora una volta che il valore atteso non corrisponde, in generale, né al valore più probabile, né a uno dei possibili valori che la variabile casuale può assumere.

6.9.1 Osservazioni su terminologia e notazioni

Gli esempi del paragrafo precedente dovrebbero aver chiarito il concetto da associare alla previsione, o valore atteso, chiamato anche *valore aspettato* o (in un pessimo italiano che dovrebbe rendere l'inglese *expectation*) *valore di aspettazione*. A volte viene utilizzato semplicemente il termine *media*, ma questo può indurre a equivoci in quanto tale nome può essere associato a tre concetti diversi:

- $\mathrm{E}(X)$ che stiamo trattando ora;

- medie (in genere aritmetiche) di variabili casuali ottenute da eventi analoghi (quindi essa stessa una variabile casuale); come esempio di quest'ultimo si pensi alla media aritmetica del numero di successi ottenuti in m esperienti costituiti dal lancio di n monete.

- medie di valori osservati (e in questo caso non si tratta di una variabile casuale ma di un numero certo);

[11] ❄Questa scelta è motivata dal fatto che il simbolo $\mathrm{E}(\cdot)$ è quello più largamente usato nella letteratura scientifica. Tuttavia preferiamo usare il termine "previsione" per designarlo, in quanto è quello che rende meglio l'idea del concetto (questo risulterà più chiaro quando ad esso sarà affiancato il concetto di "incertezza di previsione").

Oltre alle notazioni $\mathbb{P}(X)$ ed $\mathrm{E}(X)$ è spesso usato il simbolo μ (o anche μ_X ad indicare che si riferisce alla variabile X), particolarmente conveniente per alleggerire le formule. Quindi, volendo indicare in modo più compatto il valore atteso, faremo uso di

$$\mu \equiv \mathrm{E}(X) \equiv \mathbb{P}(X)\,.$$

Si noti infine che, dato il significato di media (pesata), si incontrano in letteratura anche le notazioni $\overline{X}$ e $< X >$ per tale grandezza.

6.9.2 Valore atteso di una funzione di una variabile casuale

Se X è una variabile casuale, anche la generica funzione $g(X)$ è una variabile casuale, in quanto, se siamo in stato di incertezza rispetto a X, siamo in genere[12] in stato di incertezza rispetto ad una sua funzione.

Nel caso di variabile discreta, la probabilità che la funzione assuma il valore $g(x_i)$ è uguale[13] alla probabilità che X assuma il valore x_i, ovvero $f(x_i)$:

$$P(g(X) = g(x_i)) = P(X = x_i) = f(x_i)\,. \tag{6.26}$$

Ne segue l'espressione del valore atteso di una generica funzione:

$$\mathrm{E}[g(X)] = \sum_i g(x_i)\, f(x_i)\,. \tag{6.27}$$

Nel caso semplice di variabile casuale $g(X)$ che dipende linearmente da X, ovvero $g(X) = a\,X + b$, si ha il seguente risultato:

$$\begin{aligned}
\mathrm{E}(a\,X + b) &= \sum_i (a\,x_i + b)\, f(x_i) \\
&= \sum_i a\,x_i\, f(x_i) + b\sum_i f(x_i) \\
&= a\,\mathrm{E}(X) + b\,,
\end{aligned} \tag{6.28}$$

ovvero $\mathrm{E}(\cdot)$ si comporta formalmente come un *operatore lineare*. Nel caso generale invece $\mathrm{E}[g(X)] \neq g(\mathrm{E}[X])$.

6.10 Valore atteso di distribuzioni elementari

Applichiamo la definizione operativa di valore atteso alle prime distribuzioni che abbiamo incontrato.

[12]Un controesempio è $g(X) = X^0$, nota con certezza se $X > 0$.

[13]Si presti attenzione al fatto che, se la funzione non è monotona e sia $y = g(x_i)$, non vuol dire che, in generale, $P(Y = y) = P(X = x_i)$. Infatti, bisogna sommare le probabilità di tutti i valori di X per i quali g(X)=y. Ma questo è un altro argomento e verrà trattato nel capitolo 10. Le formule che seguono sono valide anche nel caso di funzioni non monotone.

6.10.1 Distribuzione uniforme discreta

Nel caso dei primi n interi positivi abbiamo:

$$\mathrm{E}(X) \;=\; \sum_{x=1}^{n} \frac{x}{n} = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}\,, \qquad (6.29)$$

in accordo con quanto ci si poteva attendere intuitivamente. Come già visto a proposito delle funzioni di probabilità anche per previsioni ed incertezze conviene talvolta esplicitare il condizionante del valore atteso, ovvero il tipo di distribuzione di probabilità. Nel nostro caso potremmo scrivere quindi:

$$\mathrm{E}(X \mid \mathcal{K}_{1,n}) \;=\; \frac{n+1}{2}\,.$$

Comunque, come anche già detto a proposito di $f(x)$, cercheremo di utilizzare semplicemente $\mathrm{E}(X)$ se dal contesto si evince a quale distribuzione ci si riferisce.

Passiamo ora al caso più generale di distribuzione uniforme di passo arbitrario:

$$\begin{aligned}
\mathrm{E}(X \mid \mathcal{K}_{a,b,n}) &= \frac{1}{n}\,[a + (a+\Delta) + (a+2\Delta) + \cdots + (a+(n-1)\,\Delta)] \\
&= \frac{1}{n}\left[n\,a + \frac{n\,(n-1)}{2}\Delta\right] \\
&= \frac{a+b}{2}\,, \qquad (6.30)
\end{aligned}$$

altro risultato intuitivo. In entrambi i casi il risultato coincide da quanto ci aspettavamo dal significato di baricentro della distribuzione e con l'intuizione che suggerisce che, se una variabile puà assumere con pari probabilità valori equidistanziati, questi saranno distribuiti intorno al centro dell'intervallo.

6.10.2 Processo di Bernoulli

Il caso è particolarmente semplice:

$$\mathrm{E}(X \mid \mathcal{B}_p) = 1 \cdot p + 0 \cdot (1-p) = p\,. \qquad (6.31)$$

Questo risultato può anche essere letto al contrario, nel senso che *la probabilità è pari alla previsione del valore di verità di un evento*.

6.10.3 Distribuzione geometrica

Questa volta il calcolo è meno banale:

$$
\begin{aligned}
\mathrm{E}(X \mid \mathcal{G}_p) &= \sum_{x=1}^{\infty} x\, f(x) = \sum_{x=1}^{\infty} x\, p\,(1-p)^{x-1} \\
&= p \sum_{x=1}^{\infty} x\, q^{x-1} \\
&= p \frac{d}{dq} \sum_{x=1}^{\infty} q^x = p \frac{d}{dq}\left(\frac{1}{1-q}\right) = p \frac{1}{(1-q)^2} \\
&= \frac{1}{p}\,. \qquad (6.32)
\end{aligned}
$$

Per i casi della moneta, del dado, dell'ubriaco con le 8 chiavi e del singolo estratto al lotto otteniamo rispettivamente delle previsioni di 2, 6, 8 e 18 prove. Questo sta ad indicare che la risposta intuitiva "mi aspetto un successo ogni 2 (o 6, 8, 18) estrazioni" faceva riferimento a questo concetto.

6.11 Incertezza "standard" di previsione

Abbiamo visto come la previsione, o valore atteso, riassuma in un solo numero l'informazione relativa al valore intorno al quale crediamo che si realizzerà la variabile casuale. Ma questa previsione non ha alcun carattere di certezza e non va confusa con una *predizione*. Ne segue che alla previsione è necessariamente associata una incertezza di previsione, in quanto, nel caso contrario, non si potrebbe parlare di numeri aleatori. Intuitivamente possiamo dire che l'incertezza è "grande" o "piccola" se "ci aspettiamo ragionevolmente" valori che possono differire rispettivamente "molto" o "poco" dal valore atteso. In termini geometrici, l'incertezza è legata alla *"larghezza"* della distribuzione, misurata in qualche modo convenzionale.

Varianza e deviazione standard

Per quantificare il concetto di incertezza di previsione riprendiamo l'esempio della distribuzione geometrica, che anche in questo caso si presta a confrontare il concetto intuitivo con la definizione operativa. Abbiamo visto che il valore atteso vale $1/p$ e che la variabile può assumere un qualsiasi valore intero positivo. Questo ci mostra che, se associassimo all'incertezza di previsione l'ampiezza dell'intervallo nel quale la variabile può verificarsi, questo sarebbe lo stesso (e per lo più di ampiezza infinita) per qualsiasi p. Ciò contrasta con l'idea intuitiva che la previsione del numero di prove per avere testa nel lancio di una moneta sia meno incerta del numero di settimane che bisogna attendere affinché esca un prefissato numero in una certa ruota del lotto.

Per arrivare alla definizione operativa, utilizziamo il concetto intuitivo secondo il quale se la previsione è "buona" ("poco incerta") ci aspettiamo piccoli scarti fra il valore che si verificherà e la previsione stessa, ove per *scarto* intendiamo

$$\Delta = X - \mathrm{E}(X) = X - \mu\,.$$

Se invece la previsione è "cattiva" ("molto incerta") ci aspettiamo grandi scarti. Quindi una possibile misura dell'incertezza di previsione potrebbe essere il valore atteso dei possibili scarti dalla previsione stessa[14], $\mathrm{E}(X-\mu)$. Si può facilmente dimostrare come questa grandezza non sia adatta a misurare l'incertezza di previsione, essendo identicamente nulla per qualunque distribuzione:

$$\mathrm{E}(X-\mu) = \mathrm{E}(X) - \mu = 0\,.$$

La ragione è dovuta al fatto che gli scarti positivi ($X - \mu > 0$) sono compensati - in peso - da quelli negativi ($X - \mu < 0$), anche se essi differiscono in numerosità (la distribuzione geometrica ne è un buon esempio).

Si potrebbe quindi provare con il valore atteso del modulo degli scarti. Tale quantità è in linea di principio accettabile, ma in pratica si preferisce il *valore atteso del quadrato degli scarti* perché, per dirla alla buona, è più comodo lavorare con i quadrati che con i moduli. Inoltre la quantità che ne risulta è - come vedremo fra breve - formalmente coniugata al valore atteso come il momento di inerzia lo è rispetto al baricentro. Infine, la grandezza risultante gode di proprietà generali molto interessanti e molto interessanti ai fini delle applicazioni (vedi paragrafo successivo e capitolo 10.)

Il valore atteso dei quadrati degli scarti è più spesso chiamato *varianza* ed è indicato dal simbolo $\sigma^2(X)$ (o semplicemente σ^2 se non ci sono ambiguità):

$$\sigma^2(X) \equiv \mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2]\,. \tag{6.33}$$

Come si vede dalla definizione, la varianza è pari alla media dei quadrati degli scarti, ciascuno pesato con la probabilità che ad esso si attribuisce (si ricorda che la probabilità di $\Delta = \delta_i = x_i - \mu$ è uguale a quella di $X = x_i$). Esplicitando l'operatore valore atteso otteniamo la seguente definizione operativa:

$$\sigma^2 = \sum_i (x_i - \mu)^2 f(x_i)\,. \tag{6.34}$$

Dividendo il secondo membro per $\sum_i f(x_i) = 1$, otteniamo

$$\sigma^2 = \frac{\sum_i (x_i - \mu)^2 f(x_i)}{\sum_i f(x_i)}$$

dalla quale è evidente l'analogia meccanica con il momento di inerzia del sistema di punti ciascuno avente una "massa di probabilità" $f(x_i)$.

Questa grandezza caratterizza la dispersione dei valori che possono verificarsi intorno a quello di previsione, ma ha l'inconveniente di non essere

[14] Si noti come si considerino gli scarti previsti e non quelli osservati. Insistiamo nel ripetere che l'incertezza di previsione, così come la probabilità, è un concetto che si applica sui numeri rispetto ai quali si è in stato di incertezza e non ai numeri che si sono verificati. A questi è associato invece il concetto di distribuzione statistica, come già indicato nel paragrafo 6.3.

facilmente percepibile a livello intuitivo, non essendo omogenea alla previsione stessa. Si preferisce allora introdurre la *deviazione standard*, o *scarto quadratico medio*, definita come la radice quadrata (positiva) della varianza:

$$\sigma = \sqrt{\sigma^2}\,. \tag{6.35}$$

Quindi nel seguito associeremo il concetto qualitativo di "incertezza", quello operativo di *incertezza standard*, legato alla deviazione standard della distribuzione di probabilità.

In conclusione, il modo di riassumere sinteticamente lo stato di incertezza su un numero aleatorio consisterà in una affermazione del tipo:

$$\text{“\textit{previsione} } \pm \text{ \textit{incertezza standard}”}\,, \tag{6.36}$$

o, in simboli:

$$\text{“}X \;\;=\;\; \mathrm{E}(X) \pm \sigma(X)\text{”}$$

Come è noto nella vita quotidiana, quello che spesso interessa non è tanto l'entità dell'incertezza di previsione, quanto il suo valore rapportato a quello della previsione stessa. Ad esempio un'incertezza di 10 cm è enorme se riferita alla lunghezza di tavolo, piccolissima se riferita alla distanza fra due specchi distanti 10 km. La qualità della previsione è quantificata quindi dall'*incertezza relativa* (spesso espressa come *percentuale*). Essa è quantificata dal *coefficiente di variazione*, definito come

$$v(X) = \frac{\sigma(X)}{|\mathrm{E}(X)|}\,, \tag{6.37}$$

dove il modulo serve a poter utilizzare la stessa definizione indipendentemente dal segno della previsione.

A volte può essere opportuno aggiungere qualche altra grandezza che quantifichi in qualche modo convenzionale la forma della distribuzione, come può essere il grado di asimmetria fra le aspettative di scarti positivi rispetto a scarti negativi (vedi paragrafo 6.13).

6.12 Proprietà formali di varianza e deviazione standard

È interessante osservare che la varianza può essere riscritta come:

$$\begin{aligned}\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] &= \mathrm{E}[X^2 + \mu^2 - 2\,\mu\,X] \\ &= \mathrm{E}(X^2) - 2\,\mu\,\mathrm{E}(X) + \mu^2 \\ &= \mathrm{E}(X^2) - 2\,\mu^2 + \mu^2 \\ &= \mathrm{E}(X^2) - \mu^2\,. \end{aligned} \tag{6.38}$$

Questa espressione permette di memorizzare l'espressione della varianza come "valore atteso del quadrato meno il quadrato del valore atteso" (o "media dei quadrati meno il quadrato della media" - intendendo come al solito "media pesata”). Essa rappresenta inoltre il modo più comodo e più usato per calcolare la varianza, come vedremo nel seguito.

Abbiamo visto come si comporta il valore atteso della variabile casuale sotto una trasformazione lineare. Per la varianza abbiamo:

$$\begin{aligned}
\mathrm{Var}(a\,X+b) &= \mathrm{E}\left([(a\,X+b)-\mathrm{E}(a\,X+b)]^2\right) \\
&= \mathrm{E}\left([a\,X-a\,\mathrm{E}(X)]^2\right) \\
&= a^2\,\mathrm{E}\left([X-\mathrm{E}(X)]^2\right) \\
&= a^2\,\mathrm{Var}(X)\,; \qquad (6.39)
\end{aligned}$$

$$\sigma(a\,X+b) = |a|\,\sigma(X)\,. \qquad (6.40)$$

Come era da aspettarsi, la varianza non è un operatore lineare, ovvero $\mathrm{Var}(a\,X+b) \neq a\,\mathrm{Var}(X)+b$. In particolare:

- se si effettua una semplice traslazione degli assi ($a=1$, $b\neq 0$) la deviazione standard rimane invariata;
- la deviazione standard si comporta linearmente in caso di cambiamento di scala.

Si noti inoltre come la (6.38) sia l’analoga del teorema di Huygens-Steiner della meccanica, avendo $\mathrm{E}(X^2)$ il significato di *momento di inerzia rispetto all’origine* ed essendo unitaria la massa totale del sistema.

Per completezza, anche se non ne faremo uso nelle applicazione, accenniamo al fatto che la radice quadrata positiva di $\mathrm{E}(X^2)$ è chiamata *valore quadratico medio*, in inglese *root mean square*, simbolo *r.m.s.* Difatti, del tutto in generale, il nome “quadratico medio” indica la radice quadrata di medie di quadrati (si faccia attenzione a non confondere con “media dei quadrati”). Quindi

$$r.m.s. = \sqrt{\mathrm{E}(X^2)}\,.$$

Ne segue che

$$r.m.s.^2 = \sigma^2+\mu^2\,.$$

6.13 ❄ Momenti di una distribuzione e altri indicatori di forma

Abbiamo visto come valore atteso e varianza di una distribuzione di probabilità hanno una analogia formale con il baricentro e il momento di inerzia di una distribuzione di punti materiali. Allo stesso modo, Si può definire nel modo più generale

$$\mathrm{E}\left[(X-c)^r\right]\,,$$

il *momento di ordine* r *di* X *rispetto a* c.

Se c è pari a zero si parla semplicemente di *momento di ordine* r. In termini di momenti il valore atteso e la varianza hanno le sequenti definizioni:

- il valore atteso è pari al momento primo di X;
- la varianza è pari al momento secondo di X rispetto a μ.

Accenniamo ad altri due momenti che possono essere utili per quantificare la forma della distribuzione, anche se non ne faremo alcun uso per quanto riguarda le applicazioni. Per convenienza essi sono divisi per una potenza di ordine r della deviazione standard, scala tipica degli scarti.

- la *skewness*, definita come
$$\frac{\mathrm{E}[(X-\mu)^3]}{\sigma^3},$$
misura il grado di asimmetria della distribuzione rispetto alla media. Infatti, nel caso di distribuzione simmetrica la previsione degli scarti (con segno) è nulla.

- La *curtosi*, definita come
$$\frac{\mathrm{E}[(X-\mu)^4]}{\sigma^4},$$
indica quanto una distribuzione è "aguzza". Un valore grande sta ad indicare che è molto probabile trovare la variabile casuale entro circa $\pm\sigma$ dalla media e, nello stesso tempo, sono possibili grandi valori dello scarto (i quali danno un contributo a $\mathrm{E}[(X-m)^4]$ maggiore di quanto lo diano a σ^2).

Per capire le informazioni complementari fornite dai vari momenti, si possono considerare queste due semplici distribuzioni:

1. $\mathcal{K}_9$: $f(x_i) = 1/9\ \forall x_i$;
2. $f(x_i) = 1/6$, con $x_i = 0.65$, 4, 5, 5, 6 e 9.35.

Esse hanno stesso valore atteso ($\mu = 5$), deviazione standard ($\sigma = 2.58$) e skewness (uguale a 0, in quanto entrambe simmetriche rispetto al baricentro). Differiscono solo per la curtosi, uguale a 1.8 nella prima (uniforme) e 2.8 nella seconda (costituita da quattro valori valori raggruppati e da due "lontani").

6.14 ❄ Entropia come misura dello stato di incertezza

6.15 Deviazione standard delle distribuzioni elementari

Applichiamo le definizioni di varianza alle distribuzioni che abbiamo incontrato nei paragrafi precedenti.

6.15.1 Distribuzione uniforme fra 1 e n

$$\begin{aligned}
\mathrm{E}(X^2) &= \sum_{x=1}^{n} x^2 \frac{1}{n} = \frac{1}{n}\sum_{x=1}^{n} x^2 \\
&= \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right] \\
&= \frac{(n+1)(2n+1)}{6}
\end{aligned}$$

$$\begin{aligned}
\sigma^2(X) &= \mathrm{E}(X^2) - \mathrm{E}^2(X) \\
&= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\
&= \frac{n^2-1}{12} \qquad (6.41)
\end{aligned}$$

$$\sigma(X) = \frac{\sqrt{n^2-1}}{\sqrt{12}} \xrightarrow[n\gg 1]{}> \frac{n}{\sqrt{12}}\,. \qquad (6.42)$$

Quando n è abbastanza grande la deviazione standard vale circa $n/\sqrt{12}$, ovvero circa il 30 % della larghezza dell'intervallo. Questo è un risultato che incontreremo di nuovo nel caso di distribuzioni continue.

6.15.2 ❄ Distribuzione uniforme di n valori fra a e b

In questo caso i conti diventano più complessi. È preferibile allora utilizzare le proprietà di valore atteso e varianza sotto trasformazione lineare. Infatti, se consideriamo la distribuzione uniforme del paragrafo precedente, definita fra 1 ed n, si vede come si possa ottenere la distribuzione uniforme fra a e b di valori distanziati di Δ, mediante la trasformazione

$$X' = X \cdot \Delta + (a - \Delta)\,.$$

Dalle (6.29), (6.41) e (6.40) segue:

$$\begin{aligned}
\mathrm{E}(X') &= \frac{n+1}{2} + a - \Delta = \frac{a+b}{2} \qquad (6.43) \\
\sigma(X') &= \sqrt{\frac{n^2-1}{12}}\,\Delta \\
&= \frac{b-a}{\sqrt{12}}\sqrt{\frac{n+1}{n-1}} \xrightarrow[n\gg 1]{}> \frac{b-a}{\sqrt{12}}\,, \qquad (6.44)
\end{aligned}$$

riottenendo sia il risultato precedente per la media (vedi (6.30)) che, come visto nel caso analogo, una deviazione standard pari a circa al 30 % dell'ampiezza dell'intervallo appena n è sensibilmente maggiore di 1.

6.15.3 Processo di Bernoulli

La semplicità della distribuzione di Bernoulli permette di calcolare la varianza in due modi diversi, come media (pesata) dei quadrati degli scarti o come media dei quadrati meno il quadrato della media:

$$\begin{aligned}
1.\quad \mathrm{Var}(X) &= \mathrm{E}[(X-\mu)^2] \\
&= (0-p)^2(1-p) + (1-p)^2 p \\
&= p\,(1-p) \qquad (6.45)
\end{aligned}$$

$$\begin{aligned}
2.\quad \mathrm{Var}(X) &= \mathrm{E}(X^2) - \mu^2 \\
&= 0^2\cdot(1-p) + 1^2\cdot p - p^2 = p - p^2 \\
&= p\,(1-p)\,, \qquad (6.46)
\end{aligned}$$

ovvero

$$\sigma(X) = \sqrt{p\,(1-p)} = \sqrt{p\,q}. \qquad (6.47)$$

L'incertezza di previsione è massima quando $p = 0.5$ e, in tale caso, assume il valore della previsione stessa. Per p che tende a 0 o a 1 la deviazione standard tende a 0. Per quanto riguarda il coefficiente di variazione abbiamo:

$$v = \frac{\sqrt{p\,(1-p)}}{p} = \sqrt{\frac{1-p}{p}} = \sqrt{\frac{q}{p}}\,. \qquad (6.48)$$

Esso indica che l'incertezza di previsione relativa tende a zero per p prossimo a 1, mentre diverge come $1/\sqrt{p}$ per p molto piccolo.

6.15.4 Distribuzione geometrica

In questo caso il calcolo della varianza è meno semplice di quelli precedenti. Diamo direttamente il risultato:

$$\mathrm{Var}(X) = \frac{1-p}{p^2} = \frac{q}{p^2} \qquad (6.49)$$

$$\sigma(X) = \frac{\sqrt{q}}{p} \xrightarrow[p\to 0]{} \frac{1}{p} = \mathrm{E}(X)\,. \qquad (6.50)$$

Quindi, se p è abbastanza piccola, l'incertezza di previsione è circa pari alla previsione stessa. Il coefficiente di variazione vale

$$v = \frac{\sqrt{q/p}}{1/p} = \sqrt{q} \xrightarrow[p\to 0]{} 1\,, \qquad (6.51)$$

indicando un'incertezza percentuale di previsione che tende al $100\,\%$ per p piccolo.

Nel caso del singolo estratto al lotto ($p = 1/18$) l'incertezza standard di previsione vale 17.5 settimane a fronte di una previsione di 18. Tenendo conto inoltre della forma fortemente asimmetrica della distribuzione, ci capisce come enormi ritardi rispetto alla frequenza media siano da ritenere niente affatto "sorprendenti". Nel caso del lancio di una moneta regolare ci si aspetta invece di ottenere un certo esito per la prima volta dopo 2 lanci, con un'incertezza di 1.4. Si vede quindi come, in effetti, la deviazione standard fornisca una misura dell'incertezza di previsione migliore di quella ottenibile facendo uso dell'intervallo di variabilità della variabile, così come il baricentro della distribuzione rende meglio l'idea di previsione del valore di massima probabilità.

6.16 ↻ Processo di Bernoulli e percezione di probabilità prossime a 0 o a 1

Nel paragrafo 2.13, a proposito del significato dell'interpretazione soggettiva della probabilità e dei possibili fraintendimenti, si è fatto cenno ai valori di probabilità non facilmente percepibili. Può essere allora di aiuto pensare ad una distribuzione geometrica ottenuta da tanti processi di Bernoulli di pari probabilità. Ad esempio una probabilità di 10^{-20} indica una previsione di $10^{20} \pm 10^{20}$ tentativi prima che si verifichi un successo. Tale numero di tentativi, rapportato all'ordine di grandezza dell'età dell'Universo espressa in secondi, dà un'idea della piccolissima probabilità di quella classe di eventi:

$$(\approx 15 \times 10^9 \text{ anni}) \times (\approx \pi \times 10^7 \text{ s/anno}) = \approx 5 \times 10^{17}\text{s}$$

.

Simmetricamente, per probabilità molto prossime ad 1 può essere utile ragionare per complemento e pensare al numero di tentativi per osservare un insuccesso.

6.17 ❄ Previsione e incertezza di previsione di vincita in giochi d'azzardo

Parlando della speranza matematica di una vincita nei giochi d'azzardo avevamo visto che nella roulette tutte le combinazioni di gioco davano la stessa speranza matematica per unità di puntata. Possiamo utilizzare il concetto di incertezza di previsione, quantificato dalla deviazione standard, per caratterizzare la variabilità attesa degli esiti del gioco.

6.17.1 Gioco della roulette

Esaminiamo soltanto i caso di: rosso/nero e puntata su un numero. V indica la variabile "vincita" e S, come nel paragrafo 2.14, la puntata:

- rosso/nero:

$$\begin{aligned}
\mathrm{E}(V) &= \frac{18}{37} \times 2\,S + \frac{19}{37} \times 0 = \frac{36}{37}\,S\,; \\
\mathrm{E}(V^2) &= \frac{18}{37} \times (2\,S)^2 + \frac{19}{37} \times 0^2 = \frac{72}{37}\,S^2\,; \\
\mathrm{Var}(V) &= 0.999\,S^2\,; \\
\sigma(V) &\approx S\,.
\end{aligned}$$

- numero singolo:

$$\begin{aligned}
\mathrm{E}(V) &= \frac{1}{37} \times 36\,S + \frac{36}{37} \times 0 = \frac{36}{37}\,S\,; \\
\mathrm{E}(V^2) &= \frac{1}{37} \times (36\,S)^2 + \frac{36}{37} \times 0^2 = \frac{1296}{37}\,S^2\,; \\
\mathrm{Var}(V) &= 34\,S^2\,; \\
\sigma(V) &\approx 6\,S\,.
\end{aligned}$$

Si vede ancora una volta come speranza matematica e varianza della vincita caratterizzano sinteticamente la strategia del gioco. Giocare a rosso/nero conduce a un gioco più "tranquillo".

A volte si sente parlare di "sistemi per vincere" alla roulette o al lotto giocando aumentando la puntata ad ogni scommessa successiva. Una "strategia classica" per la roulette è quella di puntare sempre sullo stesso colore, raddoppiando la puntata ad ogni giocata perché "è estremamente probabile che l'altro colore esca tante volte di seguito" (il ché è vero a priori, come è vero che ciascuna altra sequenza ha la stessa probabilità, ma non ha niente a che vedere con la probabilità che un colore esca alla $n+1$-ma giocata se esso è già uscito nelle n giocate precedenti).

Cerchiamo di capire cosa non va in questa strategia[15]. Innanzitutto è chiaro che in caso di vincita non si fa altro che recuperare la posta iniziale, mentre c'è un rischio minimo (valutato a priori) di perdere il totale delle puntate se ad un certo punto si è costretti a smettere di giocare. Immaginiamo di puntare inizialmente $A = 1000$ lire e di avere disponibilità finanziarie (o limiti imposti dal banco) per giocare n volte. Per semplificare i conti assumiamo che la roulette non abbia lo "zero" (il quale essendo né pari né dispari, né rosso né nero, favorisce il banco). Cerchiamo di analizzare il problema in termini di variabile casuale guadagno netto G:

- se il colore giusto esce entro n estrazioni si vince A con probabilità
$$P(G = A) = 1 - \left(\frac{1}{2}\right)^n \;;$$
- se invece il colore non esce ci si deve ritirare perdendo tutto quanto era stato puntato fino a quel momento, per un ammontare di
$$A + A \times 2 + \cdots + A \times 2^{n-1} = (2^n - 1)\, A \,.$$
Quindi:
$$P(G = -(2^n - 1) \cdot A) = \left(\frac{1}{2}\right)^n \,.$$

La previsione di guadagno è

$$\begin{aligned} \mathrm{E}(G) &= A \cdot P(A) - (2^n - 1)\, A \cdot P\left(-(2^n - 1) \cdot A\right) \\ &= A \cdot \left(1 - \left(\frac{1}{2}\right)^n\right) - (2^n - 1)\, A \cdot \left(\frac{1}{2}\right)^n = 0 \,, \end{aligned}$$

in quanto avevamo escluso lo zero che favorirebbe il banco. Naturalmente, consistendo il gioco in altissima probabilità di bassissimi guadagni e di bassissima probabilità di altissime perdite la deviazione standard è molto alta e cresce con n (il conto viene lasciato per esercizio). Questo è anche il motivo per cui è possibile incontrare molte persone disposte a giurare che la strategia "funziona (e che quelli che hanno perso il massimo perdibile erano semplicemente sfortunati o non hanno saputo cominciare ad applicare la strategia quando il colore "era maturo"...).

[15] Si ricorda che la ragione profonda dell'errore consiste nel "dimenticare" - o più propriamente nel rifiutarsi di credere - all'equiprobabilità delle singole prove, espresso anche con l'"assenza di memoria" della distribuzione geometrica (vedi paragrafo 8.12.3).

giocata vincente	Punt. sett.	Punt. tot.	Vincita lorda	Vincita netta	G	$P(G)$ (%)
1^a	2	2	22.470	20.470	19.8	5.6
2^a	2	4	22.470	18.470	17.8	5.2
3^a	2	6	22.470	16.470	15.8	5.0
4^a	3	9	33.700	24.700	23.7	4.7
5^a	3	12	33.700	21.700	20.7	4.4
6^a	4	16	44.900	28.900	27.6	4.2
7^a	5	21	56.150	35.150	33.5	3.9
8^a	6	27	67.410	40.410	38.4	3.7
9^a	7	34	78.700	44.410	42.2	3.5
10^a	9	43	101.000	58.000	55.0	3.3
11^a	11	54	123.470	69.470	65.8	3.1
12^a	13	67	145.940	78.940	74.6	3.0
13^a	16	83	179.700	96.700	91.3	2.8
14^a	19	102	213.470	111.467	105.1	2.6
15^a	23	125	258.300	133.300	125.6	2.5
16^a	28	153	314.580	161.580	152.1	2.4
17^a	35	188	393.100	205.100	193.3	2.2
18^a	45	233	505.675	272.575	257.4	2.1
19^a	55	288	617.925	329.925	311.4	2.0
20^a	70	358	786.200	428.200	404.6	1.9
21^a	90	448	1010.800	562.800	532.5	1.8
22^a	110	558	1235.850	677.850	640.8	1.7
23^a	140	698	1572.900	874.900	827.7	1.6
24^a	180	878	2022.300	1144.300	1083.6	1.5
altrimenti	–	878	0	-878	-878.0	25.4

Tabella 6.1: "Piano di gioco indicativo speculativo" raccomandato da "Il manuale del Lotto". Nella tabella originale si precisa che "dalla vincita va detratto il 3 %". Le ultime due colonne mostrano la variabile casuale G = "guadagno netto" (al netto del 3 %) con la rispettiva probabilità. Tutti gli importi sono in migliaia di lire.

6.17.2 I sistemi "per vincere" al lotto

Facciamo un altro esempio di strategia un poco più complicato, relativa al gioco del lotto, come raccomandata da un libro "specializzato" del settore[16]. Il piano di gioco consigliato è mostrato in tabella 6.1. Le prime 5 colonne sono quelle riportate dal libro. Le altre sono la rielaborazione in termini di variabile casuale di "guadagno netto",[17] con le rispettive probabilità, calcolate secondo la formula della distribuzione geometrica di $p = 1/18$ e riportate anche in figura 6.7.

Dalla tabella si evince come sia molto probabile vincere cifre basse e poco probabile vincere somme dell'ordine delle diverse centinaia di migliaia lire

[16]"Il manuale del lotto", Mariotti Publishing, Milano, 1996.

[17]Con l'aiuto di un tabaccaio, è stato possibile accertarsi che per ogni lira giocata si ricevono 11.235 lire, ovvero si vince 10.235 volte quanto viene puntato.

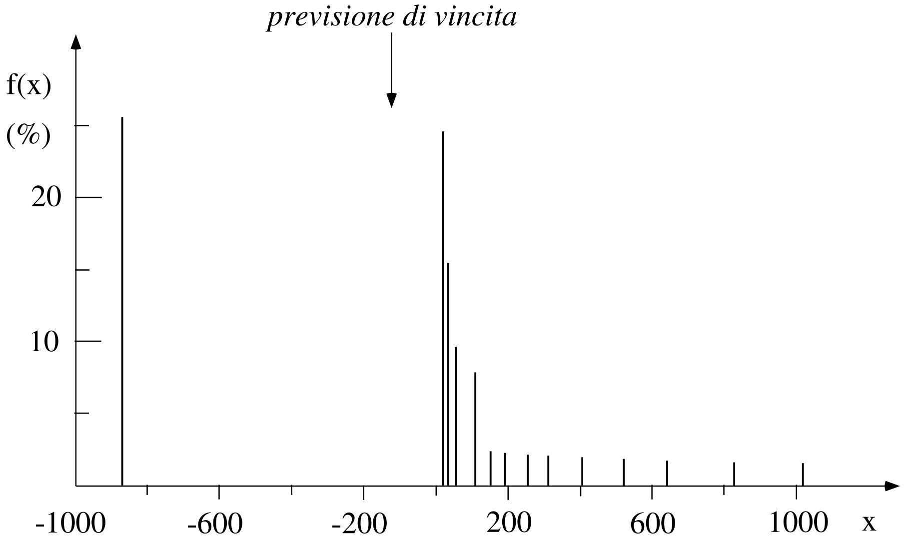


Figura 6.7: Distribuzione di probabilità della variabile casuale "guadagno" ottenuta seguendo la strategia di "vincita" al lotto riportata nella tabella 6.1.

fino al milione. Inoltre, la probabilità di vincere "qualche cosa" è del 75 % circa e questo può trarre in inganno gli sprovveduti. Il trucco sta chiaramente nel 25 % di probabilità di perdere quasi 900 mila lire se non si vince entro le prime $n = 24$ giocate. Si noti come ogni tentativo di prolungare il gioco non fa altro che rimandare il rischio di perdite sempre maggiori.

Per caratterizzare il gioco calcoliamo valore atteso di guadagno e sua deviazione standard (in migliaia di lire):

$$\left\{ \begin{array}{rcl} \mathrm{E}(G\,|\,n=24) & = & -121 \\ \sigma(G\,|\,n=24) & = & 480\,. \end{array} \right.$$

Il confronto fra deviazione standard e valore atteso mostra che questa strategia è senza dubbio "vivace", ma purtroppo mediamente perdente. Per capire la dipendenza di questi parametri dal numero di giocate previste dal piano di gioco, facciamo i conti anche per $n = 10, 15$ e 20.

$$\left\{ \begin{array}{rcl} \mathrm{E}(G\,|\,n=10) & = & -12 \\ \sigma(G\,|\,n=10) & = & 36 \end{array} \right.$$
$$\left\{ \begin{array}{rcl} \mathrm{E}(G\,|\,n=15) & = & -28 \\ \sigma(G\,|\,n=15) & = & 86 \end{array} \right.$$
$$\left\{ \begin{array}{rcl} \mathrm{E}(G\,|\,n=20) & = & -62 \\ \sigma(G\,|\,n=20) & = & 215 \end{array} \right.$$

Praticamente la previsione di "guadagno" è una perdita che, grosso modo, raddoppia ogni cinque giocate del piano.

6.18 ↻ Misure di centralità e di dispersione di distribuzioni statistiche

Nel paragrafo 6.3 abbiamo chiarito quali sono le differenze e le analogie fra distribuzioni statistiche e distribuzioni di probabilità. Se con x_i indichiamo il generico valore della variabile (sia casuale che statistica) e, con riferimento alle distribuzioni statistiche, con n_i il numero di volte che si è verificato x_i, con n il numero totale di occorrenze e con $w_i = n_i/n$ il *peso statistico* di x_i, si nota l'analogia

$$f(x_i) \leftrightarrow w_i \ :$$

- come per $f(x_i)$, vale $\sum_i w_i = 1$;
- $f(x_1) > f(x_2)$ indica che si crede più al verificarsi di x_1 che al verificarsi di x_2; analogalmente $w_1 > w_2$ indica che x_1 <u>si è</u> verificato più di x_2;
- come analoghe misure di posizione e di dispersione si possono prendere

$$\overline{x} = \sum_i x_i w_i \tag{6.52}$$

$$\text{Var}(x) = \sum_i (x_i - \overline{x})^2 w_i = \overline{x_2} - \overline{x}^2 \, , \tag{6.53}$$

  da cui segue la deviazione standard.

  Quando ciascuna delle variabili statistiche compaiono ciascuna una volta il peso statistico è uguale per tutte $w_i = 1/n$ e si ottiene:

$$\overline{x} = \frac{1}{n}\sum_i x_i \tag{6.54}$$

$$\text{Var}(x) = \frac{1}{n}\sum_i (x_i - \overline{x})^2 \tag{6.55}$$

Analogalmente a quanto visto nel paragrafo 6.13 si possono definire momenti e altri indicatori di forma.

Siccome le strette analogie formali possono portare a confondere è opportuno fare delle osservazioni su simboli e terminologia.

- Il termine *media* viene usato per entrambe le distribuzioni, ma i termini che hanno un contenuto prettamente probabilistico o inferenziale, come valore atteso, previsione o speranza matematica, hanno solo senso se riferiti a distribuzioni di probabilità; a volte ci si riferisce alla media di una distribuzione statistica con il termine *media campionaria*[18].
- La media di distribuzioni statistiche è indicata generalmente con il simbolo $\overline{x}$;

[18] Qualcuno vorrebbe anche separare i concetti mediante i termini *media* e *valore medio*, ma mi sembra una battaglia persa, in quanto intuitivamente essi non fanno riferimento a cose diverse.

- I termini varianza e deviazione standard sono usati pressoché indistintamente per entrambe le distribuzioni, anche se qualcuno tende ad usare il nome scarto quadratico medio preferibilmente per le distribuzioni statistiche, o insistere sull'aggettivo *campionario*, o *sperimentale* (o *empirica*), quando si fa riferimento a queste ultime.

- Anche i simboli $\text{Var}(\cdot)$, σ^2 e σ sono usati quasi indifferendemente per entrambe le distribuzioni; molto spesso le quantità statistiche sono indicate usando il simbolo s a posto di σ. Inoltre, per motivi di interferenza fra statistica descrittiva, s^2 è calcolata dividendo i quadrati degli scarti per $n-1$ anziché per n (si noti come molti calcolatorini tascabili calcolano sia "σ_n" che "σ_{n-1}", oppure "σ" e "s". Siccome questo testo non ha la velleità di imporre degli standard di notazione, ma piuttosto vuole incoraggiare il lettore ad una certa flessibilità e ad evitare la pedanteria, invitiamo soltanto a prestare attenzione a specificare chiaramente cosa si sta facendo e ad interpretare correttamente il lavoro di altri. Siccome ci occupiamo principalmente di questioni probabilistiche, per σ intendiamo la deviazione standard di una distribuzione di probabilità, oppure il "parametro σ" della gaussiana. Quando interverranno deviazioni standard su dati sperimentali o incertezze standard sul parametro σ della gaussiana, cercheremo chiarire nel contesto di ogni discussione il significato dei simboli.

6.19 Ricapitolando

- Un numero aleatorio (o variabile casuale) rappresenta un numero ben definito, ma rispetto al quale si è in stato di incertezza.

- Quando a ciascun possibile valore del numero aleatoria si associa un grado di fiducia si costruisce una distribuzione di probabilità. Nel caso di variabili discrete la funzione di probabilità $f(x)$ ha il significato di probabilità che la variabile X assuma il valore x.

- Le proprietà delle funzioni di probabilità discendono direttamente dalle proprietà della probabilità.

- Quando si hanno invece valori numerici associati alla frequenza con il quale essi si sono verificati nel passato si parla di distribuzione statistica.

- La distribuzione che assegna pari probabilità a tutti i possibili valori di una variabile casuale è la distribuzione uniforme. In questo capitolo sono state mostrate soltanto quelle più semplici con valori della variabile casuale equidistanziati.

- Nel processo di Bernoulli si definisce una variabile pari all'indicatore del contenuto di verità di un evento. Considerando più processi di Bernoulli indipendenti e definiti su eventi analoghi aventi tutti la stessa probabilità si ottengono altre distribuzioni di notevole interesse. In particolare, la distribuzione geometrica è legata numero di processo di Bernoulli nel quale si verifica per la prima volta un evento favorevole.

- Sebbene la distribuzione di probabilità descriva completamente lo stato di incertezza ripetto al numero aleatorio, è molto pratico poter riassumere le aspettative sul verificarsi del numero in termini di previsione (il valore intorno al quale si verificherà ragionevolmente la variabile casuale) e di incertezza di previsione (caratterizzata dalla dispersione dei possibili valori intorno alla previsione).

- È conveniente utilizzare, come definizione operativa di previsione, la media dei valori che il numero aleatorio può assumere, ciascuno pesato con grado di fiducia che gli si attribuisce (baricentro della distribuzione).

- Valore atteso, valore aspettato, speranza matematica, (valore di aspettazione) e media sono da considerare sinonimi di previsione. Si noti inoltre che il concetto di previsione non va confuso né con quello di valore di massima probabilità (detto moda) né con quello centrale (detto mediana) che divide i valori della variabile casuale in due classi ordinate di pari probabilità.

- Il modo "standard" di quantificare l'incertezza di previsione consiste nel fare uso della previsione dei quadrati degli scarti rispetto alla previsione stessa. Questo indicatore di dispersione è chiamato varianza. Per convenienza si usa la la radice quadrata della varianza, detta deviazione standard, grandezza omogena con la variabile di interesse e con la previsione e quindi più facilmente percepibile.

- Anche le distribuzioni statistiche possono essere sintetizzate con misure di centralità e di dispersione analoghe a quelle delle distribuzioni di probabilità. Spesso nella letteratura scientifica alcuni termini e simboli vengono spesso usati pressoché indistintamente per le due distribuzioni. È opportuno abituarsi ad una certa flessibilità di simbologia, a capire dal contesto qual'è il significato esatto da attribuire alle grandezze e, infine, a chiarire cosa a cosa ci si voglia riferire se non è univoco dal contesto.

6.20 Problemi

1. Ad un gioco di società si lancia un dado. Se il numero è pari si retrocede di un numero di caselle pari alla metà del valore indicato; se il numero è dispari di avanza del doppio del numero indicato. Ricavarsi la distribuzione di probabilità della variabile casuale "numero di caselle delle quali si avanza".

2. Calcolare previone e incertezza di previsione del numero incerto definito nel problema precedente.

3. Le variabili X e Y possono assumere ciascuna cinque valori. Come distribuzione di probabilità vengono assegnate, rispettivamente: $f(x_1) = 0.2$, $f(x_2) = 0.001$, $f(x_3) = 0.399$, $f(x_4) = -0.1$, $f(x_5) = 0.5$;
   $f(y_1) = 0.2$, $f(y_2) = 0.001$, $f(y_3) = 0.399$, $f(y_4) = 0.1$, $f(y_5) = 0.5$.
   Cos'è che non va in ciascuna delle distribuzioni?

4. Una variabile casuale discreta ha una funzione di ripartizione $F(x)$ discontinua in corrispondenza dei primi 6 interi positivi. In tali punti essa assume valori: 0.10, 0.30, 0.60, 0.90, 0.95, 1.00. Calcolare le seguenti probabilità:
   - $P(X = 2)$;
   - $P(X = 3.5)$;
   - $P(X < \pi)$;
   - $P(X = 1 \cup X = 6)$;
   - $P(X < 4)$;
   - $P(X > 1.5)$;
   - $P(\frac{2}{3}\,\pi \leq X \leq \frac{4}{3}\,\pi)$.

5. Si estrae una carta da un mazzo di 40 carte. Quanto vale la distribuzione di probabilità dei valori delle carte? Calcolare media e deviazione standard della distribuzione.

6. Si lanciano due dadi. Associare alla variabile X_1 la somma dei valori che appaiono sulle facce rivolte verso l'alto e alla variabile X_2 il modulo della loro differenza. Ricavarsi la distribuzione di probabilità delle due variabili casuali con i rispettivi valori di previsione e incertezza di previsione.

7. ↻ Uno sperimentatore misura due tensioni di un circuito con un voltmetro digitale in grado di indicare fino al decimo di Volt. Lo sperimentatore sa che la misura non è affetta da errori sistematici, che lo strumento è ben calibrato e che l'indicazione è effettuata arrotondando le cifre successive a quella indicata secondo le regole "delle calcolatrici tascabili" (fino a 4 approssima per difetto, dal 5 - compreso - in poi per eccesso). Se lo sperimentatore legge $V_{L_1} = 2.3$ V e $V_{L_2} = 2.1$ V, quanto vale la previsione (con la sua incertezza) della differenza fra i valori veri delle tensioni? (Si costruiscano le distribuzioni delle variabili associate ai valori veri delle grandezze fisiche di interesse, al centesimo di Volt, dato questo stato di informazione).

8. ↻ Sul problema precedente. Quanto vale la probabilità che, nelle condizioni precedentemente illustrate, nei due punti del circuito la tensione "vera" sia la stessa?

9. ↻ Vengono misurate due temperature con due termometri simili aventi un display digitale al grado. Si sa che quei termometri hanno una probabilità del 70 % di indicare la temperature giusta, del 15 % di indicare un grado in più e del 15 % di indicare un grado in meno. I due termometri danno informazioni indipendenti in quanto non è stato possibile intercalibrarli. I due termometri indicano 85 e 81°C, temperature per le quali il giudizio a priori dello sperimentatore è molto più vago dell'informazione fornita dallo strumento. Cosa si può dire sulla differenza di temperatura?

10. Un appassionato di lotto decide di "inseguire" un certo numero su una certa ruota. Quanto vale la probabilità che egli riesca a vincere entro, rispettivamente, 10, 18, 30, 50 e 100 estrazioni? Quanto vale la probabilità che il numero esca alla 101-ma estrazione nei casi che si sia verificato o che non si sia non verificato precedenetemente?

11. Una persona gioca sempre le stesse due colonne all'enalotto, sperando che prima o poi possano uscire. Calcolare previsione e incertezza di previsione del numero di volte che dovrà giocare per vincere.

12. La deviazione standard di una distribuzione di Bernoulli si annulla per $p = 0$ o $p = 1$ (consistentemente con il fatto che tali valori corrispondono a certezze, e quindi l'incertezze di previsione deve essere nulla. Per quale valore è massima tale incertezza?

13. ❄ Si immagini una variabile casuale definita su un intervallo finito, per semplicità $X \in [0, 1]$. Mostrare che la distribuzione di massima varianza è quella che concentra la probabilità agli estremi dell'intervallo. Si verifichi, ad esempio, la varianza diminuisce se: a) le probabilità dei due estremi diventano, rispettivamente, $1/2 - \epsilon$ e $1/2 + \epsilon$; b) due punti molto prossimi agli estremi (Δ e $1 - \Delta$) acquistano una probabilità diversa da zero.

14. Una variabile casuale è definita nell'intervallo compreso fra 100 e 1000. È possibile immaginare una distribuzione di probabilità tale che la deviazione standard valga 500 o più?
15. Si rianalizzi il problema 13 del capitolo 4 alla luce delle distribuzioni di probabilità. Quanto vale la previsione (con la sua incertezza) del tentativo in cui la pistola fa fuoco?
16. In un gioco di società uno dei partecipanti è "finito in prigione" e può riprendere il gioco soltanto quando riuscirà ad ottenere il numero 1 con il lancio di un dado. Quanto vale le probabilità che ci riesca entro il terzo colpo?
17. Una persona gioca alla roulette a rosso e nero praticando la strategia di raddoppio descritta nel paragrafo 6.17, cominciando con una puntata iniziale di 10˙000. Calcolare la probabilità che egli perda oltre 10 milioni la prima serie di giocate (per semplificare i conti si trascuri l'effetto dello zero).
18. Un banco di roulette accetta puntate massime di 100 milioni. Un miliardario comincia a giocare 100˙000 lire, raddoppiando la puntata dopo ogni perdita. Calcolare previsione e incertezza di previsione di vincita tenendo conto dell'effetto dello zero.
19. Paradosso di San Pietroburgo (baclawski, pag. 174)

Capitolo 7

Distribuzioni di probabilità di variabili discrete - II

Vengono qui introdotte le distribuzioni di variabile discreta più interessanti per le applicazioni di laboratorio: la binomiale e la poissoniana. Quest'ultima è presentata sia come particolare limite formale della binomiale che dal punto di vista del cosiddetto processo di Poisson, che lega le distribuzioni del numero di conteggi a quelle dei tempi di attesa (esponenziale e gamma) che saranno trattate nel prossimo capitolo. Vengono inoltre illustrati altri aspetti generali, quali la traduzione dei teoremi sulle probabilità nel linguaggio delle funzioni di probabilità, il significato probabilistico dell'incertezza standard e il concetto di intervallo di credibilità.

Gli esempi di applicazione di binomiale e poissoniana sono finalizzati ad una introduzione ai problemi inferenziali riguardo variabili aleatorie e alla verifica delle leggi statistiche, argomenti che saranno ripresi sistematicamente nella terza parte del testo.

Viene anche presentata una formulazione della legge dei grandi numeri, ottenuta da considerazioni sulla distribuzione binomiale. L'argomento sarà ripreso nel capitolo 10.

7.1 Distribuzione binomiale

Consideriamo ora la seconda schematizzazione di eventi legati al processo di Bernoulli, descritta nel paragrafo 6.6.4. Ovvero, interessandoci al numero di successi che possono verificarsi in un certo numero di tentativi effettuati nelle stesse condizioni.

Se analizziamo n prove indipendenti, per ciascuna delle quali la probabilità di successo è p, la variabile casuale X = "numero totale di successi" può andare da 0 a n (da nessuno a tutti).

Ricaviamoci la funzione di probabilità, a partire dai valori "più facili".

$\boxed{x = 0 \textbf{ e } x = n}$

$$\begin{aligned} f(n) &= P(E_1 \cap E_2 \cap \cdots \cap E_n) = p^n \\ f(0) &= P(\overline{E}_1 \cap \overline{E}_2 \cap \cdots \cap \overline{E}_n) = (1-p)^n = q^n\,. \end{aligned}$$

$\boxed{x = 1 \textbf{ e } x = n-1}$ Per $x = 1$ si deve verificare un solo successo e $n-1$ insuccessi. Quindi sembrerebbe, ad esempio, che

$$f(1) \stackrel{?}{=} p\,(1-p)^{n-1}\,. \tag{7.1}$$

Ma in realtà questa espressione dà la probabilità che il successo si verifichi ad un certo tentativo (ad esempio al primo) e gli insuccessi nei rimanenti $n-1$ tentativi. Poiché abbiamo n possibili tentativi fra loro incompatibili nei quali si può verificare il successo, dobbiamo moltiplicare l'espressione precedente per n. Per $f(n-1)$ otteniamo un analogo risultato. Quindi:

$$\begin{aligned} f(1) &= n\,p\,(1-p)^{n-1} \\ f(n-1) &= n\,p^{n-1}\,(1-p)\,. \end{aligned}$$

$\boxed{x \textbf{ generica}}$ Ne segue che l'espressione generale della funzione di probabilità è data dalla probabilità che si verifichino x successi e $n-x$ insuccessi, pari a

$$p^x\,(1-p)^{n-x}\,, \tag{7.2}$$

moltiplicata per il numero di volte che, indipendentemente dall'ordine, si possono ottenere gli x successi in n prove. Questo numero è pari a quello delle combinazioni semplici di n elementi presi x a x, che - ricordiamo - sono date dai coefficienti binomiali, indicati con

$$\binom{c}{n}$$

(vedi paragrafo 3.2.5). La formula generale della funzione di probabilità è quindi

$$f(x\,|\,\mathcal{B}_{n,p}) = \binom{n}{x} p^x\,(1-p)^{n-x} \qquad \left\{ \begin{array}{l} n = 1, 2, \ldots, \infty \\ 0 \le p \le 1 \\ x = 0, 1, \ldots, n \end{array} \right. \tag{7.3}$$

Esplicitando l'espressione dei coefficienti binomiali, la (7.3) può essere riscritta in modo più pratico come

$$f(x\,|\,\mathcal{B}_{n,p}) = \frac{n!}{(n-x)!\,x!}\,p^x(1-p)^{n-x}\,, \tag{7.4}$$

ovvero

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

avendo indicato con q la probabilità di insuccesso $1-p$. Per quanto riguarda la funzione di ripartizione $F(x)$, in questo caso essa non ha una espressione matematica semplice e va calcolata dalla definizione stessa, sommando tutti i valori di $f(x)$ fino al numero intero positivo immediatamente precedente ad x (ricordiamo ancora una volta che $F(x)$ è convenzionalmente definita su tutto l'asse reale).

7.2 ↻ Distribuzione binomiale – da capo

Reintroduciamo la distribuzione binomiale in modo alternativo, ad uso di chi è poco disinvolto con il calcolo combinatorio, o per coloro che abbiano saltato il capitolo 3

Supponiamo di lanciare un dado e di essere interessati all'uscita di un certo valore, per esempio il "4". Definiamo *successo* ogni esito del lancio in cui si verifica il numero "4". Pensiamo di dover eseguire un certo numero n di lanci ed esaminiamo la variabile X "numero di successi" ($0 \leq X \leq n$). Vediamo quanto vale la probabilità della variabile casuale X all'aumentare del numero dei lanci (S sta per successo e N sta per insuccesso):

<u>n = 1</u>

e_i	$P(e_i)$	X
S	1/6	1
N	5/6	0

da cui $f(0) = P(X = 0) = \frac{5}{6}$ e $f(1) = P(X = 1) = \frac{1}{6}$. È immediato provare che $f(0) + f(1) = 1$.

<u>n = 2</u>

e_i	$P(e_i)$	X
SS	$(1/6)^2$	2
SN	$(1/6) \times (5/6)$	1
NS	$(5/6) \times (1/6)$	1
NN	$(5/6)^2$	0

da cui $f(0) = (\frac{5}{6})^2$, $f(1) = 2 \times \frac{1}{6} \times \frac{5}{6}$ e $f(2) = (\frac{1}{6})^2$ e chiaramente anche in questo caso $\sum_i f(x_i) = 1$.

<u>n = 3</u>

e_i	$P(e_i)$	X
SSS	$(1/6)^3$	3
SSN	$(1/6)^2 \times (5/6)$	2
SNS	$(1/6)^2 \times (5/6)$	2
SNN	$(1/6) \times (5/6)^2$	1
NSS	$(1/6)^2 \times (5/6)$	2
NSN	$(1/6) \times (5/6)^2$	1
NNS	$(1/6) \times (5/6)^2$	1
NNN	$(5/6)^3$	0

Otteniamo quindi per la variabile X:

x	$f(x)$
0	$(5/6)^3$
1	$3 \times (1/6) \times (5/6)^2$
2	$3 \times (1/6)^2 \times (5/6)$
3	$(1/6)^3$

Notiamo l'analogia con lo sviluppo del binomio di Newton. Infatti chiamando p la probabilità di un successo nel singolo lancio ($p = \frac{1}{6}$ nel nostro esempio) e $q = 1 - p$ la probabilità di un insuccesso, le $f(x_i)$ corrispondono ai termini dello sviluppo di $(p+q)^n$. Questo è dovuto al fatto che l'evento x successi e $(n-x)$ insuccessi si può presentare in tanti modi diversi, ed esattamente quante sono le possibilità di costituire, partendo da n elementi, dei gruppetti di x elementi ciacuno, indipendentemente dall'ordine con cui gli x elementi sono scelti. In termini matematici esse sono le combinazioni di n elementi presi "x a x". La probabilità di ciascuno dei modi è pari a $p^x q^{n-x}$. Quindi, in generale, possiamo scrivere questa distribuzione di probabilità, chiamata *binomiale*, come:

$$f(x|\mathcal{B}_{n,p}) = \binom{n}{x} p^x \, q^{n-x} \, . \tag{7.5}$$

Resta da calcolar quanto vale il coefficiente binomiale

$$\binom{n}{x} .$$

Ricordiamo qui brevemente i concetti di *permutazioni* e *combinazioni*:
<u>Permutazioni</u>: il numero totale dei possibili modi di disporre ("ordinare","mettere in fila") n oggetti si calcola considerando che ci sono n possibilità per il primo, $n-1$ per il secondo, $n-2$ per il terzo e così via. Cioè il numero di permutazioni di n oggetti è pari a $n!$.
<u>Combinazioni</u>: supponiamo ora di dover scegliere un x elementi da un insieme che ne contiene un numero n senza curarsi dell'ordine con cui essi sono scelti. Ad esempio in una classe di 25 persone si vogliono formare delle squadre di pallavolo da 6 persone. Quante squadre diverse si possono fare? Abbiamo 25 modi per scegliere la prima persona, 24 per la seconda e così via, cioè $25 \times 24 \times \cdots \times \cdot (25 - 6 + 1)$. Così facendo abbiamo contato $6! = 720$ volte (ovvero il numero di permutazioni di 6 elementi) ogni squadra composta dagli stessi giocatori ma estratti in modi diversi. Quindi, in generale, bisogna devidere l'espressione precedentemente trovata per $x!$, ottenendo

$$\frac{n\,(n-1)\,\cdots\,(n-x+1)}{x!}$$

. Moltiplicando numeratore e denominatore per $(n-x)!$ otteniamo:

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

e quindi

$$f(x \,|\, \mathcal{B}_{n,p}) = \frac{n!}{(n-x)!\,x!} \, p^x \, q^{n-x} \, , \qquad \begin{cases} n = 1, 2, \ldots, \infty \\ 0 \le p \le 1 \\ x = 0, 1, \ldots, n \end{cases} \tag{7.6}$$

7.3 Proprietà della distribuzione binomiale e note sul suo uso

La figura 7.1 mostra degli esempi di distribuzione binomiale per alcuni valori di n e di p. Si noti come:

- all'aumentare di n e di p; la distribuzione si sposta a valori sempre più elevati;
- all'aumentare di n la distribuzione diventa relativamente più stretta (ovvero se la larghezza viene rapportata al valore centrale);
- all'aumentare di n la distribuzione di infittisce e, per valori di p distanti da 0 e da 1, acquista una forma regolare "a campana".

Per provare che $\sum_{x=0}^{n} f(x) = 1$ è sufficiente ricordare che la sommatoria corrisponde allo sviluppo di $(p+q)^n$ (vedi anche par. 3.2.6), pari a 1 essendo $p+q=1$.

7.3.1 Valore atteso e deviazione standard

È abbastanza naturale pensare che se si lancia 1000 volte una moneta ci si aspetta un numero di teste intorno a 500. Nello stesso modo è ovvio aspettarsi che se si lancia un dado 20 volte si prevede che una certa faccia uscirà circa un sesto del numero di lanci, anche se chiaramente non potrà verificarsi 3.33 volte. Verifichiamo come la definizione operativa di valore atteso sia in accordo con tale previsione intuitiva:

$$
\begin{aligned}
\mathrm{E}(X \mid \mathcal{B}_{n,p}) &= \sum_{x=0}^{n} x\, f(x \mid \mathcal{B}_{n,p}) \\
&= \sum_{x=0}^{n} x\, \frac{n!}{(n-x)!\,x!}\, p^x q^{n-x} \\
&= \sum_{x=1}^{n} x\, \frac{n!}{(n-x)!\,x!}\, p^x\, q^{n-x} \\
&= n\,p \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!\,(x-1)!} p^{x-1} q^{n-x} \\
&= n\,p \sum_{y=0}^{n-1} \frac{(n-1)!}{(n-1-y)!\,y!} p^y q^{n-1-y} \\
&= n\,p\,(p+q)^{n-1} \\
&= n\,p
\end{aligned}
$$

avendo chiamato, per comodità $y = x - 1$ e utilizzando le proprietà dei coefficienti binomiali (vedi paragrafo 3.2.6).

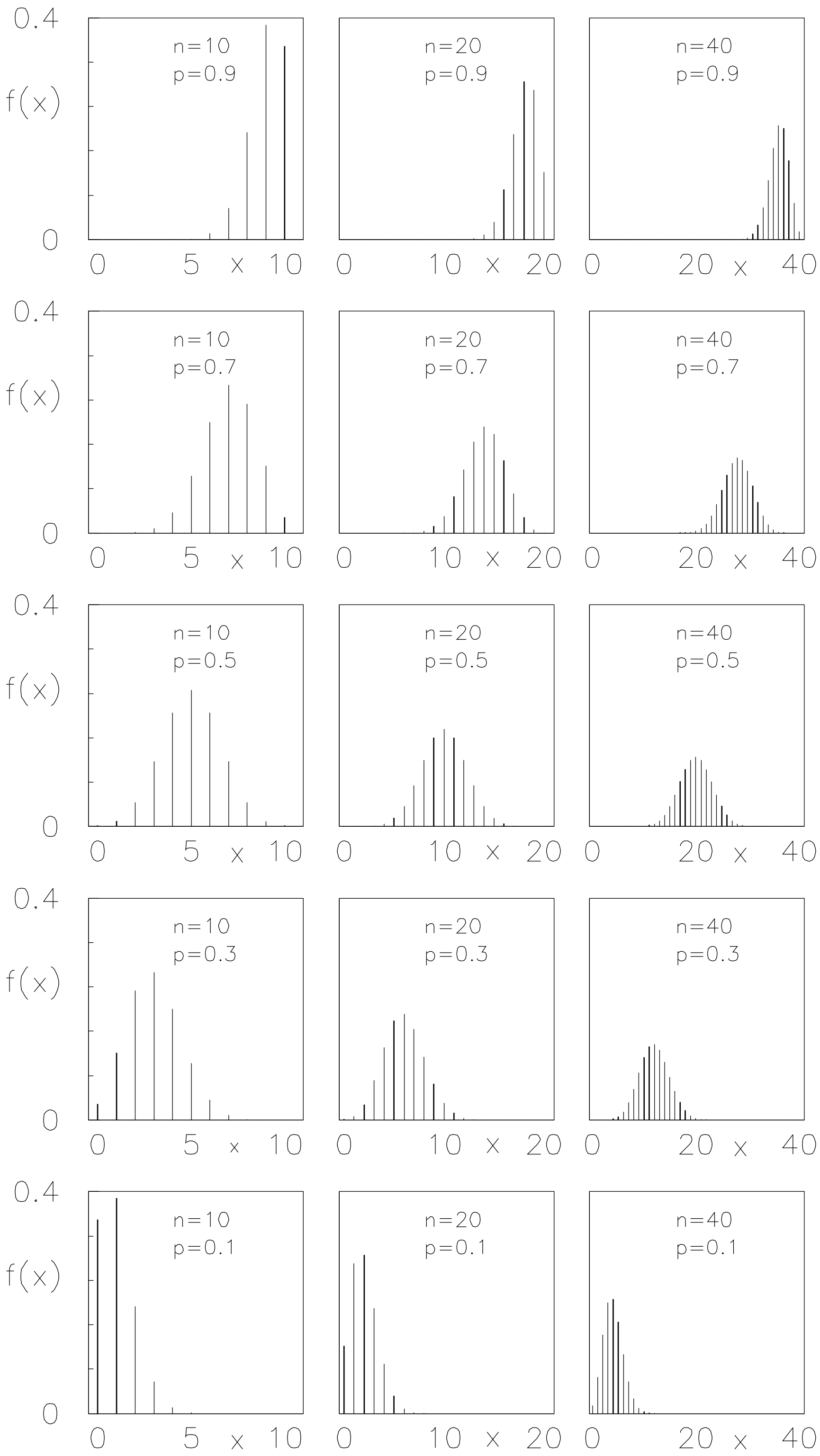


Figura 7.1: Esempi di distribuzione binomiale in funzione dei parametri. Dall'alto verso il basso viene fissato n e diminuisce p. Da sinistra a destra aumenta n per p costante.

Per calcolare la varianza della distribuzione binomiale è conveniente, come al solito, partire dal valore atteso di X^2:

$$\begin{aligned}
\mathrm{E}(X^2 \,|\, \mathcal{B}_{n,p}) &= \sum_{x=0}^{n} x^2\, f(x \,|\, \mathcal{B}_{n,p}) \\
&= \sum_{x=0}^{n} x^2 \frac{n!}{(n-x)!\,x!}\, p^x\, q^{n-x} \\
&= n\,p \sum_{x=1}^{n} \frac{x\,(n-1)!}{(n-x)!\,(x-1)!} p^{x-1} q^{n-x} \\
&= n\,p \sum_{y=0}^{n-1} \frac{(1+y)\,(n-1)!}{(n-1-y)!\,y!} p^{y} q^{n-1-y} \\
&= n\,p\,(1+(n-1)\,p) \\
&= n\,p + n\,(n-1)\,p^2 .
\end{aligned}$$

Da cui:

$$\begin{aligned}
\sigma^2 &= p\,(1-p)\,n = p\,q\,n \\
\sigma &= \sqrt{p\,(1-p)\,n} = \sqrt{p\,q\,n}
\end{aligned}$$

Nella seguente tabella riportiamo nel caso di lanci di una moneta i valori attesi e la varianza del numero di teste.

n	μ	σ	σ/μ	$P(n/2)$
10	5	1.6	0.32	0.24
100	50	5.0	0.10	0.08
1000	500	16	0.032	0.025
10000	5000	50	0.010	0.010
1000000	500000	500	0.001	0.0008

Si può osservare come al crescere di n ci attendiamo sempre una maggiore dispersione di valori della variabile casuale intorno al valore atteso[1]. In particolare, si noti come al crescere di n il valore atteso diventi sempre meno probabile. Diminuisce invece la dispersione relativa, essendo il coefficiente di variazione pari a

$$v = \frac{\sigma}{\mu} = \frac{\sqrt{n\,p\,q}}{n\,p} \propto \frac{1}{\sqrt{n}} \,. \tag{7.7}$$

[1]In particolare, non è vero che "il numero di teste tende al numero di croci", come si sente dire talvolta. Questa affermazione, oltre che essere assolutamente fuorviante quando si pensa ad un effetto di "recupero" di un esito sull'altro per "mantenersi in pari", è anche errata quando si pensa ad una previsione di piccoli scarti fra i due esiti. Questo discorso sarà ripreso quando si parlerà delle cattive interpretazioni della legge dei grandi numeri.

$m = n\,p = 1$						
x	$f(x)$					
	$\mathcal{B}_{2,\frac{1}{2}}$	$\mathcal{B}_{4,\frac{1}{4}}$	$\mathcal{B}_{10,\frac{1}{10}}$	$\mathcal{B}_{50,\frac{1}{50}}$	$\mathcal{B}_{1000,\frac{1}{1000}}$	$\mathcal{B}_{10^6,10^{-6}}$
0	0.25	0.316	0.349	0.364	0.368	0.368
1	0.50	0.422	0.387	0.372	0.368	0.368
2	0.25	0.211	0.194	0.186	0.184	0.184
3		0.047	0.057	0.061	0.061	0.061
4		0.004	0.011	0.015	0.015	0.015
5			0.001	0.003	0.003	0.003
6			...	...	0.001	...
...			...	...	...	...
10			10^{-10}	...	...	...
...				...	...	...
50				$\approx 10^{-85}$	...	...
...					...	...
1000					10^{-3000}	...
...						...
1000000						10^{-10^6}

Tabella 7.1: Valori della funzione di probabilità di una distribuzione binomiale al variare di n e p, con il vincolo $n\,p = 1$. Al di sotto di probabilità dello 0.1 % è riportato il solo valore minimo (per $x = n$).

Riassumendo, otteniamo che *la previsione dei valori della variabile è uguale ad una frazione p del numero di prove da effettuare, con una incertezza relativa di previsione inversamente proporzionale alla radice quadrata del numero di prove*.

7.3.2 Usi tipici della distribuzione binomiale

La binomiale è utilizzata ogni qualvolta gli esiti degli esperimenti indipendenti e sotto le stesse condizioni (p costante) possono essere classificati in successo/insuccesso. Per le applicazioni di laboratorio essa descrive, detto in modo generico, problemi di *efficienze*. Ad esempio, si può essere interessati a quante volte un apparato riveli una particella, o diagnosi una certa malattia, ritenendo uguale a p la probabilità che esso ci riesca ad ogni prova.

7.4 Distribuzione di Poisson

Consideriamo numeri aleatori che seguono distribuzioni binomiali di parametri p e n diversi, ma tali che il prodotto di p per n sia lo stesso, ad esempio $np = 1$. Nella tabella 7.1 sono riportate le distribuzioni di probabilità al variare di p e n.

Si nota innanzitutto che, nonostante n cresca, i valori per i quali la probabilità è ragionevolmente diversa da zero, sono soltanto quelli intorno a qualche unità. Inoltre, la distribuzione sembra stabilizzarsi intorno a valori di probabilità che non dipendono dall'esatto valore di n. Se si facesse una seconda tabella con un diverso valore di μ si troverebbe nuovamente un comportamento asintotico al crescere di n, ovviamente *non* alla stessa distribuzione, visto che il valore atteso è diverso. Quindi al crescere di n, la distribuzione sembra dipendere soltanto da μ.

Questa è una proprietà molto interessante, in quanto ci sono fenomeni di interesse che sono descritti da leggi di tipo binomiale, ma con n talmente grande per cui la formula della distribuzione binomiale sarebbe di scarso uso pratico. Per esempio supponiamo di avere un campione di sostanza radioattiva. Il numero di nuclei contenuti può essere dell'ordine di 10^{20}, mentre la probabilità di osservare un decadimento in un piccolo intervallo di tempo opportunamente scelto può essere dell'ordine di 10^{-20}. Il valore atteso del numero di decadimenti è $np = 1$, e quindi ci aspettiamo che la funzione di probabilità sia la stessa della tabella.

Oltre al vantaggio computazionale appena descritto, è molto interessante ai fini applicativi che la distribuzione non dipenda dagli esatti valori di n e di p, ma solo dal prodotto. Sempre nell'esempio dei decadimenti radioattivi, i dati sperimentali possono suggerire che il valore atteso di decadimenti in quell'intervallo di tempo sia 1. È allora possibile valutare la distribuzione di probabilità senza conoscere né n né p, stante la sola ragionevolissima condizione che n sia "grande".

Per dimostrare che quanto mostrato nella tabella è una proprietà generale, facciamo il limite della distribuzione binomiale per $n \to \infty$ e $p \to 0$, con la condizione che $n\,p$ resti finito e molto minore di n. Indichiamo[2] il prodotto $n\,p$ con λ. Riscriviamo la formula della distribuzione binomiale:

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

Sostituendo a p

$$p = \frac{\lambda}{n}$$

si vede che fare il limite $n \to \infty$ e $p \to 0$ corrisponde al solo limite $n \to \infty$, ovvero:

$$\lim_{n\to\infty} \frac{n!}{(n-x)!\,x!} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}\,.$$

Mettendo a fattore i termini non dipendenti da n e riscrivendo il rapporto fra $n!$ e $(n-x)!$ nella sua forma originaria, cioè $n\,(n-1)\,\cdots\,(n-x-1)$ riscriviamo il limite di $f(x)$ come:

$$\lim_{n\to\infty} \frac{\lambda^x}{x!} \cdot \frac{\overbrace{n\,(n-1)\,(n-2)\,\cdots(n-x+1)}^{\to n^x}}{n^x} \cdot \frac{\overbrace{\left(1-\frac{\lambda}{n}\right)^n}^{\to e^{-\lambda}}}{\underbrace{\left(1-\frac{\lambda}{n}\right)^x}_{\to 1}}\,.$$

Notiamo che, per tale limite:

[2] Lo si potrebbe anche chiamare μ, avendo esso il significato di valore atteso, e così si trova difatti in alcuni testi. Qui preferiamo dare un simbolo proprio al parametro della poissoniana.

- il numeratore della seconda frazione è uguale a x fattori, ciascuno circa uguale a n, in quanto $x \ll n$; esso è quindi uguale al denominatore e si semplifica;
- il numeratore della terza frazione tende, come noto, a $e^{-\lambda}$;
- il denominatore della terza frazione tende a 1.

In conclusione abbiamo:

$$f(x \,|\, \mathcal{B}_{n,p}) \xrightarrow[\substack{p \to 0 \\ n \to \infty \\ n\,p = \lambda \text{ (finito)}}]{} \frac{\lambda^x}{x!}\, e^{-\lambda} \,.$$

Otteniamo quindi una distribuzione di probabilità caratterizzata dal solo parametro λ, numero reale positivo, chiamata *distribuzione di Poisson* o (*poissoniana*), che indichiamo nel seguente modo

$$f(x \,|\, \mathcal{P}_\lambda) = \frac{\lambda^x}{x!}\, e^{-\lambda} \qquad \left\{ \begin{array}{l} 0 \le x < \infty \\ 0 < \lambda < \infty \end{array} \right. . \tag{7.8}$$

Non c'è bisogno di calcolare il valore atteso e la varianza della distribuzione,[3] in quanto possono essere ottenute dalla binomiale effettuando il limite per $p \to 0$. Ne segue:

$$\begin{aligned} \mathrm{E}(X) &= \lambda \\ \sigma^2 &= \lambda \\ \sigma &= \sqrt{\lambda} \\ v &= \frac{1}{\sqrt{\lambda}} \end{aligned}$$

Riportiamo nella tabella 7.2 $f(x \,|\, \mathcal{P}_\lambda)$ per alcuni valori (piccoli) di λ. La figura 7.2 mostra alcuni esempi di rappresentazione grafica. Si noti come la probabilità di osservare $X = 0$ sia data da $e^{-\lambda}$. Per $\lambda < 1$ essa è chiaramente dominante e diventa trascurabile soltanto per λ abbastanza grande. Al crescere di λ la distribuzione comincia a diventare simmetrica intorno al valor medio, il quale si avvicina anche al valore di massima probabilità.

7.5 ○ Processo di Poisson - prima parte

La distribuzione di Poisson è stata introdotta nel paragrafo precedente come limite della distribuzione binomiale. Per quanto riguarda l'interpretazione fisica dei parametri n e p, suggerita anche dall'esempio sui decadimenti radioattivi, si può pensare che:

- n sia il numero di "oggetti" ai quali può succedere qualcosa;
- p sia la probabilità che a ciascuno di quelli eventi succeda "quella cosa". Essa è la stessa per tutti gli n oggetti.

[3]*** Mettere in nota? ***

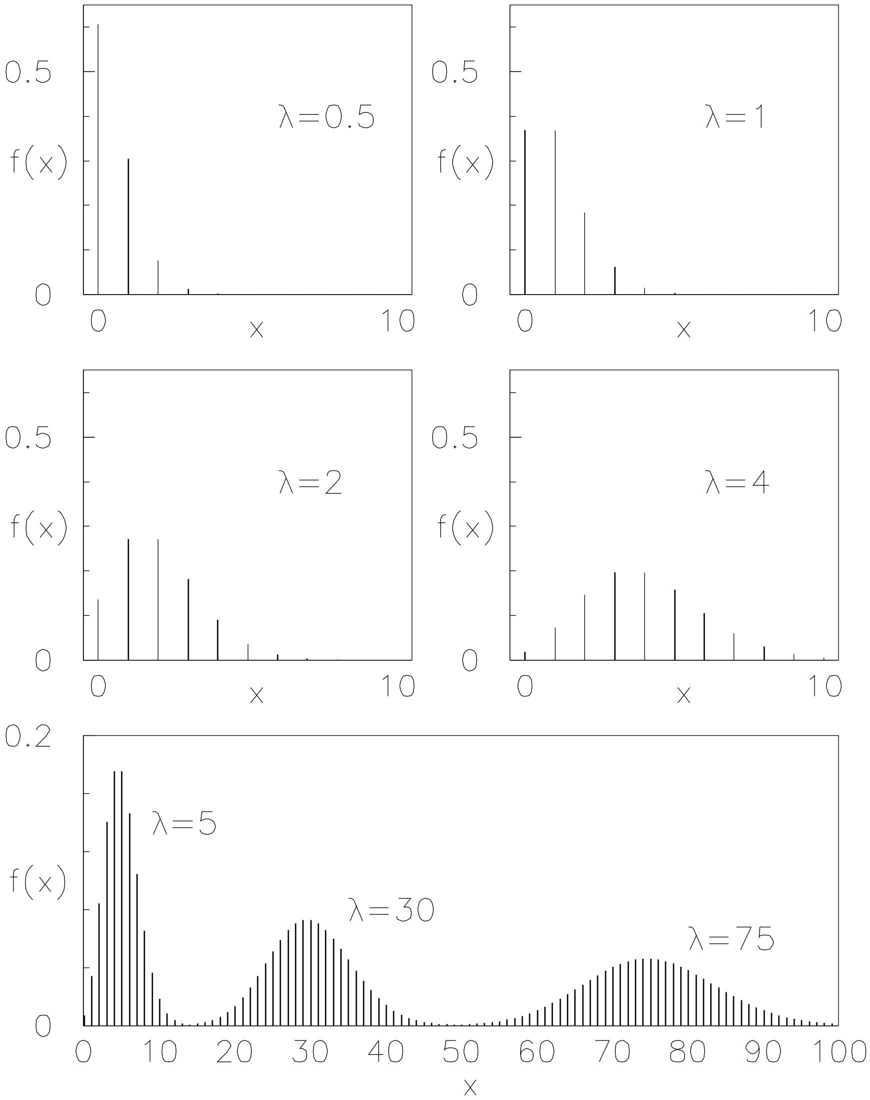


Figura 7.2: Esempi di distribuzione di Poisson con λ uguale a 0.5, 1, 2, 4, 5, 30 e 75 (le ultime tre sono sovrapposte nel grafico in basso).

x	$f(x)$					
	$\mathcal{P}_{0.01}$	$\mathcal{P}_{0.1}$	$\mathcal{P}_{0.5}$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_4$
0	**0.990**	**0.905**	**0.607**	**0.368**	0.135	0.018
1	0.010	0.090	0.303	**0.368**	**0.271**	0.073
2	$5 \cdot 10^{-5}$	0.005	0.076	0.184	**0.271**	0.147
3	$2 \cdot 10^{-7}$	$2 \cdot 10^{-4}$	0.013	0.061	0.180	**0.195**
4	$4 \cdot 10^{-10}$	$4 \cdot 10^{-6}$	0.002	0.015	0.090	**0.195**
5	$8 \cdot 10^{-13}$	$8 \cdot 10^{-13}$	$2 \cdot 10^{-4}$	0.003	0.036	0.156
6	$1 \cdot 10^{-15}$	$1 \cdot 10^{-9}$	$1 \cdot 10^{-5}$	$5 \cdot 10^{-4}$	0.012	0.104
7	$2 \cdot 10^{-18}$	$2 \cdot 10^{-11}$	$9 \cdot 10^{-7}$	$7 \cdot 10^{-5}$	0.003	0.060
8	$2 \cdot 10^{-21}$	$2 \cdot 10^{-13}$	$6 \cdot 10^{-8}$	$9 \cdot 10^{-6}$	0.001	0.030
9	$3 \cdot 10^{-24}$	$2 \cdot 10^{-15}$	$3 \cdot 10^{-9}$	$1 \cdot 10^{-6}$	$2 \cdot 10^{-4}$	0.015
10	$3 \cdot 10^{-27}$	$2 \cdot 10^{-17}$	$2 \cdot 10^{-10}$	$1 \cdot 10^{-7}$	$4 \cdot 10^{-5}$	0.005
11	$2 \cdot 10^{-30}$	$2 \cdot 10^{-19}$	$7 \cdot 10^{-12}$	$9 \cdot 10^{-9}$	$7 \cdot 10^{-6}$	0.002

Tabella 7.2: Distribuzione di Poisson per alcuni valori di λ. Gli eventi di massima probabilità sono indicati in grassetto.

Questo punto di vista può essere per alcuni aspetti limitativo, in quanto non sempre è possibile o ha senso tale schematizzazione. Se ad esempio pensiamo alla probabilità che una macchina rossa percorra un tratto di strada in un certo intervallo di tempo, una trattazione secondo la distribuzione di Poisson implicherebbe una precedente schematizzazione in termini binomiali, con n pari al numero di macchine e p la probabilità che ciascuna delle macchine transiti a quell'ora. Ma bisognerà considerare solo le macchine di quella città o "tutte" le macchine? E poi anche p varia da macchina a macchina! Volendo si può anche risolvere il problema insistendo a voler riferire n alle macchine e intendendo p una probabilità condizionata dalla sola conoscenza di "macchina rossa" (una sorta di p media).

È interessante mostrare lo stesso problema da un altro punto di vista, quello degli *atti elementari di osservazione*. Questo modo alternativo di ragionare è molto più generale del precedente ed inoltre collega il numero aleatorio "numero di osservazioni" al numero aleatorio reale "tempo fra due osservazioni successive" (questo secondo aspetto verrà ripreso nel paragrafo **??**)

Consideriamo fenomeni che si manifestano nel tempo o nello spazio e di cui siamo interessanti al numero di occorrenze, indipendentemente dall'ordine. Si parla in generale di *misure di conteggio*. Esempi tipici sono

- telefonate che arrivano ad un centralino;
- errori di stampa in un libro;
- decadimenti radioattivi in un tempo molto inferiore a quello di dimezzamento;
- numero di globuli bianchi osservati al microscopio in un campo ottico;
- difetti di fabbricazione di un cavo;
- numero di molecole in un piccolo volume di gas;

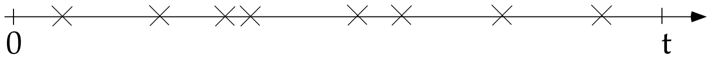


Figura 7.3: Processo di Poisson nel dominio del tempo. Le crocette indicano gli istanti delle occorrenze delle osservazioni.

Ciascuno di questi fenomeni può manifestarsi, indipendentemente dagli altri, in un certo intervallo o elemento molto piccolo, sia esso di tempo, lunghezza, superficie o volume (rispettivamente Δt, Δl, ΔS e ΔV).

Nel seguito, per comodità ma senza perdere di generalità, prenderemo in considerazione problemi nel dominio del tempo (vedi figura 7.3) Interessiamoci quindi al *numero di conteggi* registrati in un certo intervallo finito di tempo t, ovvero il numero aleatorio è definito come X = "numero di conteggi fra 0 e t". Supponiamo ora che

1. La probabilità che che si verifichi esattamente un conteggio in un intervallino Δt sia proporzionale a Δt:

   $$p = P(\text{“1 conteggio in } \Delta t\text{”}) = r\Delta t\,,$$

   con r costante nell'intervallo finito t, in modo tale che p non dipenda dall'intervallino preso in considerazione, ma soltanto dalla sua sola durata;

2. La probabilità che in Δt si verifichino più di 1 eventi sia trascurabile in confronto a quella che se ne verifichi esattamente 1;

3. il numero di conteggi in un intervallo finito sia indipendente dal numero di conteggi che si verificano in un altro intervallo, se i due intervalli sono disgiunti .

Queste ipotesi definiscono i cosidetti *processi di Poisson*.

Consideriamo n intervallini disgiunti, ciascuno di durata Δt, tali che $t = n\Delta t$, ovvero

$$\Delta t = \frac{t}{n}\,.$$

Quando n tende ad infinito ne segue che $\Delta t \to 0$ e, di conseguenza, $p \to 0$. Consideriamo inoltre che:

- in ogni intervallino l'evento "accade un conteggio" può essere considerato come un processo di Bernoulli indipendente dagli altri;

- essendo r costante e quindi p costante, gli n processi dei quali ci interessiamo soltanto al numero di successi danno luogo ad una binomiale;

- la condizione $n \to \infty$ (e consequente $p \to 0$) rende la binomiale approssimabile da una poissoniana, ovvero

  $$f(x\,|\,\mathcal{B}_{np}) \longrightarrow f(x\,|\,\mathcal{P}_\lambda)\,,$$

con
$$\lambda = n\,p = \frac{t}{\Delta t}\,r\,\Delta t = r\,t\,.$$
Essendo il valore atteso della distribuzione di Poisson uguale a λ, e quest'ultima pari a $r\,t$, si vede quindi che r ha il significato di *numero atteso di conteggi per unità di tempo*, ovvero quantifica l'*intensità del processo*. Il simbolo r dovrebbe ricordare il "rateo" (tasso), in inglese "rate"(e, in una tipica applicazione di tale processo, anche la *radioattività*).

Come già detto, alcuni problemi possono essere considerati o dal punto di vista degli n oggetti o dal punto di vista degli n atti di osservazione. Consideriamo i due casi per mostrare che il numero aleatorio "globuli nel sangue osservati al microscopio" segue una distribuzione di Poisson:

A) n_g**:** numero di globuli di una persona;

p_g**:** probabilità che un certo globulo venga estratto e che si trovi nel campo ottico di quella osservazione.

Essendo n_g molto grande e p_g molto piccolo, ne segue che
$$X \sim \mathcal{P}_{\lambda_g = p_g\, n_g}\,.$$

B) n_V**:** numero di volumetti ($V/\Delta V$) di cui è costituito il sangue nel campo del microscopio;

p_V**:** probabilità di trovare un globulo in un volumetto ΔV;

Anche in questo caso n_V è molto grande, in quanto ΔV può essere pensato dell'ordine di grandezza del globulo stesso. Ne segue che
$$X \sim \mathcal{P}_{\lambda_V = p_V\, n_V}\,.$$

È da notare come n e p siano diversi nei due casi (e di conseguenza sono stati designati con simboli diversi), ma la distribuzione risultante è la stessa ($\lambda_g = \lambda_V = \lambda$), in quanto essa dipende soltanto dal valore atteso di conteggi e non da n e da p separatamente.

Terminiamo con due osservazioni, una relativa all'uso della distribuzione di Poisson, l'altra sui cosidetti "eventi rari".

Innanzitutto è importante ricordare che gli oggetti da contare debbano apparire indipendentemente uno dall'altro. Per esempio, se alcuni oggetti preferiscono manifestarsi a coppie o a gruppi più numerosi (per esempio i turisti giapponesi su un autobus) non si può applicare la distribuzione di Poisson sui singoli elementi, ma eventualmente sui gruppi, se si crede che essi sono siano loro indipendenti (sicuramente non vale per i gruppi di turisti, schedulati dalle agenzie di viaggio…).

Talvolta la distribuzione di Poisson è chiamata anche *"distribuzione degli eventi rari"*. Questo può essere giustificato dal fatto che nell'intervallino dell'atto di elementare osservazione la probabilità è effettivamente bassa (o simmetricamente che sia molto bassa la probabilità che a ciascuno degli oggetti in questione possa succedere qualcosa), oppure perché in molti casi macroscopici la "rarità" del fenomeno è richiesta dalla condizione di indipendenza degli eventi (gli affollamenenti creano inevitabilmente delle correlazioni: ad esempio il numero di macchine che transitano per una strada di campagna fra le 10 e le 11 del mattino può essere descritto da un processo di Poisson, ma sicuramente tale schematizzazione non può andare bene per descrivere il traffico urbano nelle ore di punta). Ma queste condizioni non implicano che tutti gli eventi debbano essere "rari" su scala umana. Ad esempio un materiale radiattivo potrebbe emettere un milione di particelle al secondo, oppure si può essere interessati al numero di molecole contenute in un cm^3 di aria, ottenendo previsioni tutt'altro che piccole pur essendo tali numeri aleatori ben descritti da distribuzioni di Poisson.

7.6 ❄ Formule ricorsive per la distribuzione binomiale e di Poisson

Nel calcolare i valori della distribuzione di probabilità di Poisson può essere utile far uso di una formula ricorsiva al fine di semplificare i calcoli. Infatti

$$\begin{aligned}
f(0\,|\,\mathcal{P}_\lambda) &= e^{-\lambda}\,, \\
f(1\,|\,\mathcal{P}_\lambda) &= \frac{\lambda\, e^{-\lambda}}{1} = \frac{\lambda}{1}\, f(0\,|\,\mathcal{P}_\lambda) \\
f(2\,|\,\mathcal{P}_\lambda) &= \frac{\lambda^2\, e^{-\lambda}}{2} = \frac{\lambda}{2}\, f(1\,|\,\mathcal{P}_\lambda) \\
\cdots & \quad \cdots \\
f(x\,|\,\mathcal{P}_\lambda) &= \frac{\lambda}{x}\, f(x-1\,|\,\mathcal{P}_\lambda)\,. \qquad (7.9)
\end{aligned}$$

Dalla formula ricorsiva si può facilmente vedere che $f(x\,|\,\mathcal{P}_\lambda)$ è maggiore di $f(x-1\,|\,\mathcal{P}_\lambda)$ finché λ/x è maggiore o uguale dell'unità. Ciò significa che $f(x\,|\,\mathcal{P}_\lambda)$ ha un massimo in prossimità di $x=\lambda$. Più esattamente, se λ è intero il fattore di riaggiornamento della (7.9) è pari a 1 in corrispondenza di $x=\lambda$, ovvero $f(x)=f(x-1)$. Quindi, per λ intero, la distribuzione assume il valore massimo in corrispondenza di $x=\lambda$ e di $x=\lambda-1$, come anche mostrato in figura 7.2 e in tabella 7.2.

Nel caso della distribuzione binomiale la formula ricorsiva non è semplice come per la poissoniana, ma può tornare utile, specie con l'ausilio di un computer, per il calcolo di grandi valori di n:

$$\begin{aligned}
f(0\,|\,\mathcal{B}_{n,p}) &= q^n\,, \\
f(x\,|\,\mathcal{B}_{n,p}) &= \frac{n-x+1}{x}\,\frac{p}{q}\, f(x-1\,|\,\mathcal{B}_{n,p})\,.
\end{aligned}$$

7.7 ⟳ Proprietà riproduttiva delle distribuzioni di probabilità binomiale e di Poisson

Una distribuzione di probabilità gode della *proprietà riproduttiva* rispetto alla somma se una variabile casuale costruita come somma di altre variabili casuali, ognuna delle quali è descritta da una certa distribuzione, obbedisce alla stessa distribuzione. Sia la binomiale che la poissoniana godono di tale proprietà. Più esattamente:

- binomiale: Se $X_1, X_2, ..., X_m$ sono m variabili casuali indipendenti, ciascuna descritta da una distribuzione di probabilità binomiale di parametri p e n_i, la variabile casuale $Y = \sum_i X_i$ segue ancora una distribuzione di probabilità binomiale con parametri p e $n = \sum_i n_i$.

- poissoniana: Se X_1, X_2, ..., X_m sono m variabili casuali indipendenti, ciascuna descritta da una distribuzione di probabilità di Poisson di parametro λ_i, la variabile casuale $Y = \sum_i X_i$ segue ancora una distribuzione di Poisson con parametro $\lambda = \sum_i \lambda_i$.

Queste proprietà possono essere dimostrate matematicamente. È più istruttivo dimostrarle invece ragionando sui fenomeni descritti da queste distribuzioni.

Per il caso della binomiale, la proprietà riproduttiva segue dal fatto che ciascuna distribuzione è dovuta a n_i processi di Bernoulli indipendenti. Ne segue che gli $n = \sum_i n_i$ processi di Bernoulli indipendenti e aventi la stessa p danno luogo ad una binomiale di parametri p e n.

Ovviamente questo non richiede che i processi di Bernoulli in questione siano legati allo stesso fenomeno.

Per quanto riguarda la distribuzione di Poisson è sufficiente pensare ai conteggi effettuati in un certo intervallo (temporale o spaziale, o entrambi, a seconda dei casi) di osservazione come se fossero dovuti alla somma dei conteggi effettuati in intervalli più piccoli. Se la distribuzione è poissoniana nell'intervallo prescelto a maggior ragione lo sarà per ciascuno degli intervalli in cui è suddiviso e, siccome il valore atteso del numero di conteggi è proporzionale all'ampiezza dell'intervallo, la proprietà riproduttiva è dimostrata per variabili casuali legate allo stesso fenomeno.

7.8 ❄ Altre distribuzioni di interesse

Distribuzione di Pascal

Abbiamo studiato la distribuzione geometrica, legata al numero aleatorio "tentativo per il quale si registra il primo successo", quando si considerano tanti processi di Bernoulli di uguale p. Il caso più generale è quello che descrive la distribuzione di probabilità del "tentativo per il quale si registrano esattamente k successi". Essa è nota come distribuzione di Pascal. La derivazione è abbastanza semplice:

- negli $x-1$ tentativi precedenti si devono essere verificati $k-1$ successi e $x-k$ insuccessi, indipendentemente dall'ordine. La probabilità di questo evento si ottiene dalla distribuzione binomiale di parametri $n = x - 1$ e p
$$f(k-1\,|\,\mathcal{B}_{x-1}, p) = \binom{x-1}{k-1} p^{k-1}(1-p)^{x-k}\,;$$

- nel tentativo x si deve verificare il successo, con probabilità p.

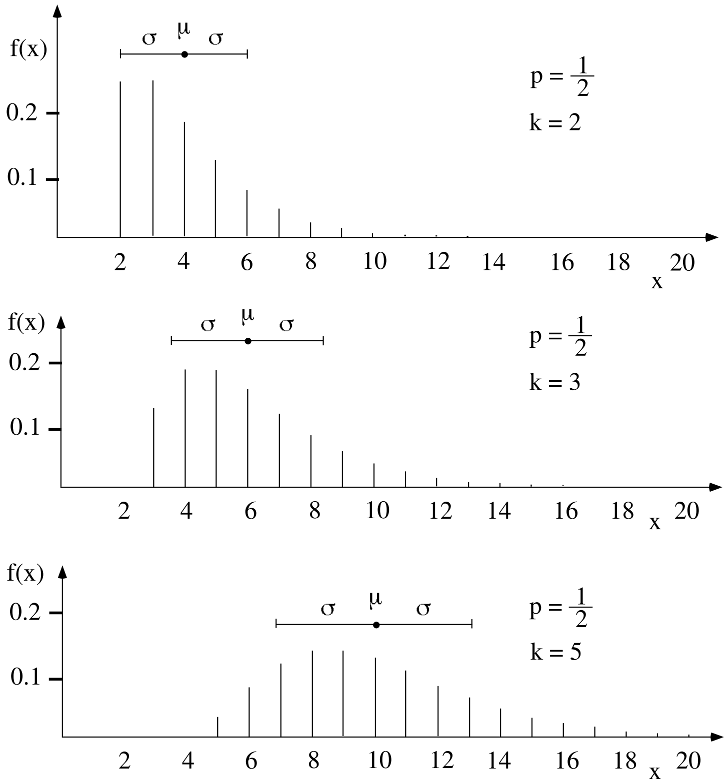


Figura 7.4: Esempi di distribuzione di Pascal: probabilità che, lanciando una moneta, si ottengano $k = 2$, 3 e 5 teste al tentativo x.

Essendo i due eventi indipendenti, si ottiene la distribuzione di probabilità

$$f(x\,|\,\mathcal{P}a_{k,p}) \;=\; \binom{x-1}{k-1} p^k (1-p)^{x-k} \qquad \left\{ \begin{array}{l} 0 \le p \le 1 \\ k = 1, 2, \ldots \infty \\ x = k, k+1, \ldots \infty \end{array} \right. \tag{7.10}$$

Naturalmente, per $k = 1$, si riottiene la distribuzione geometrica, ovvero

$$f(x\,|\,\mathcal{P}a_{1,p}) = f(x\,|\,\mathcal{G}_p)\,.$$

Diamo direttamente previsione e incertezza di previsione del numero aleatorio, descritto da questa distribuzione:

$$\mathrm{E}(X) = \frac{k}{p}\,, \tag{7.11}$$

$$\mathrm{Var}(X) = k\,\frac{q}{p^2} \tag{7.12}$$

$$\sigma(X) = \sqrt{k}\,\frac{\sqrt{q}}{p} \xrightarrow[\rightarrow 0]{} \sqrt{k}\,\frac{1}{p} \tag{7.13}$$

$$v = \frac{1}{\sqrt{k\,q}} \xrightarrow[p\rightarrow 0]{} \frac{1}{\sqrt{k}}\,. \tag{7.14}$$

Come si capisce intuitivamente, la previsione del numero di tentativi per avere k successi è proporzionale al numero di successi richiesto. Quello che è meno intuitivo, ma che risulterà essere una proprietà generale della varianza, è che non è la deviazione standard, bensì la varianza, ad essere proporzionale a k. Ne segue che l'incertezza relativa decresce all'aumentare di k.

Binomiale negativa

Il problema precedente può essere trattato usando, al posto della variabile X precedente definita la variabile complementare Y "numero di insuccessi al momento in cui si verificano esattamente k successi", ovvero

$$Y = X - k\,.$$

La distribuzione di probabilità di Y si ricava direttamente dalla (7.10):

$$f(x\,|\,\mathcal{B}^-_{k,p}) \;=\; \binom{y+k-1}{k-1} p^k (1-p)^y \qquad \begin{cases} 0 \le p \le 1 \\ k = 1, 2, \ldots \infty \\ x = k, k+1, \ldots \infty \end{cases} \tag{7.15}$$

Questa distribuzione è chiamata *binomiale negativa* in quanto è possibile riscrivere la sua espressione in modo da far comparire dei cosidetti coefficienti binomiali negativi, scritti, in generale, come

$$\binom{-n}{r},$$

formalmente analoghi dei normali coefficienti binomiali.

Per calcolare previsione e incertezza di previsione si possono applicare direttamente le proprietà degli operatori $\mathrm{E}(\cdot)$ e $\mathrm{Var}(\cdot)$ alla trasformazione $Y = X - k$. Ne segue:

$$\mathrm{E}(Y) = k\,\frac{q}{p}\,, \tag{7.16}$$

$$\mathrm{Var}(Y) = k\,\frac{q}{p^2} \tag{7.17}$$

$$\sigma(Y) = \sqrt{k}\,\frac{\sqrt{q}}{p} \xrightarrow[p\to 0]{} \sqrt{k}\,\frac{1}{p} \tag{7.18}$$

$$v = \frac{1}{\sqrt{k\,q}} \xrightarrow[p\to 0]{} \frac{1}{\sqrt{k}}\,. \tag{7.19}$$

Come ultima osservazione su questa distribuzione, che non avremo più modo di incontrare nel seguito, è che la sua complementarità con la distribuzione di Pascal fa sì che sia possibile utilizzare la binomiale negativa per risolvere dei problemi per la quale sarebbe più naturale utilizzare quella. Inoltre, a volte in alcuni testi è la (7.15) ad essere chiamata anche distribuzione di Pascal. Si presti quindi attenzione.

Distribuzione ipergeometrica

Vediamo ora una distribuzione non legata al processo di Bernoulli e che rappresenta un modo alternativo per arrivare alla distribuzione di binomiale.

Supponiamo che in una popolazione di N persone ce ne siano M aventi una certa caratteristica. Se viene scelta una persona "a caso" la probabilità che essa abbia quella caratteristica è uguale a $p = M/N$. Se n osservatori estraggono a caso ciascuno una persona, in modo che l'estrazione di un osservatore non sia influenzato da quella degli altri (e quindi la stessa persona può essere scelta più volte), la variabile casuale X = "numero di persone che presentano quella caratteristica" ha probabilità secondo una binomiale di parametri p e n.

Se invece vengono scelte contemporaneamente n persone il numero aleatorio X è descritto da una diversa distribuzione di probabilità. Ad esempio, nel caso limite in cui vengano prese tutte le N persone, la variabile X può assumere soltanto il valore M, e quindi essa non è più una variabile aleatoria, bensì un numero certo.

I due casi sono schematizzati con i classici problemi di estrazioni da urne di palline bianche e nere, *con reintroduzione* (o *reimbussolamento*) e *senza reintroduzione*. Nel primo caso vengono repristinate le condizioni iniziale dopo ogni estrazioni e quindi si ha la condizioni di indipendenza della probabilità che sta a base della distribuzione binomiale. Nel secondo caso la probabilità di estrarre, ad esempio, una pallina bianca dipende dal numero di palline bianche e nere estratte precedentemente.

Riformuliamo quindi il problema con lo schema dell'urna:

- un'urna contiene N palline, di cui M bianche e le $N - M$ restanti nere;
- vengono estratte a caso e senza reintroduzione n palline;
- ci interessiamo alla variabile casuale X = "x delle n palline estratte sono bianche".

Bisogna considerare i modi di scegliere x palline bianche fra le M totali, indipendentemente dal loro ordine. Essi sono dati dalle combinazioni semplici di M elementi a gruppi di x:

$$\binom{M}{x}\,.$$

Per ciascuno di questi modi ci sono, con ragionamento analogo,

$$\binom{N-M}{n-x}$$

modi di estrarre le $n - x$ palline nere. Quindi il numero di casi favorevoli è dato dal prodotto dei due coefficienti binomiali.

Il numero dei casi possibili è dato dai numero di scelte di n palline fra le N. Quindi, assumendo l'equiprobabilità, si ottiene

$$\begin{aligned} f(x\,|\,\mathcal{H}_{N,M,n}) &= \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad (7.22)\\ \text{con} &= \max(0, n-(N-M)) \le x \le \min(n, M)\,. \end{aligned}$$

	Ipergeometrica per $n=4$ e $M=\frac{1}{2}N$				$\mathcal{B}_{n=4,p=\frac{1}{2}}$
x	$N=4$	$N=6$	$N=10$	$N=100$	
0	-	-	0.0238	0.0587	0.0625
1	-	$1/5=0.2$	0.2381	0.2499	0.2500
2	1	$3/5=0.6$	0.4762	0.3827	0.3750
3	-	$1/5=0.2$	0.2381	0.2499	0.2500
4	-	-	0.0238	0.0587	0.0625
E(X)	2	2	2	2	2
$\sigma(X)$	0	0.63	0.82	0.98	1.00

Tabella 7.3: Confronto fra alcune distribuzioni ipergeometriche e la binomiale aventi lo stesso numero di estrazioni (n) e con $r=p$ (r indica M/N). Si noti il limite alla binomiale per $N \gg n$.

Al di fuori di tali ovvii limiti la probabilità è pari a zero.

Questa distribuzione è chiamata *distribuzione ipergeometrica*. Diamo direttamente valore atteso e varianza:

$$
\begin{aligned}
\mathrm{E}(X) &= n\,\frac{M}{N} = n\,r && (7.23)\\
\mathrm{Var}(X) &= n\,\frac{M}{N}\,\frac{N-M}{N}\,\frac{N-n}{N-1} && \\
&= n\,r\,(1-r)\,\frac{N-n}{N-1} && (7.24)\\
\sigma(X) &= \sigma_B\sqrt{\frac{N-n}{N-1}} && (7.25)
\end{aligned}
$$

ove con r è stato indicato il rapporto fra il numero di palline bianche e il numero totale di palline e con σ_B il prodotto $n\,r\,(1-r)$. Si noti che, se $n=N$, la varianza si annulla, in quanto diventa certo il solo esito $X=M$.

Se N è molto più grande di n, l'estrazione non cambia di molto le proporzioni di palline all'interno della scatola. Quindi ci si aspetta che quando $N/n \to \infty$ la distribuzione ipergeometrica tenda alla binomiale di parametri n e $p=r=M/N$. Questo è in effetti il caso, anche se non lo dimostriamo. È invece immediato vedere come valore atteso e varianza tendano rispettivamente a $n\,p$ e a $n\,p\,(1-p)$ (questa è la ragione del simbolo σ_B nella (7.25)). La tabella 7.3 mostra alcune distribuzioni ipergeometriche di $M=N/2$ e $n=4$ confrontate con la binomiale di $n=4$ e $p=1/2$.

7.9 ❄Cammino casuale e problema della rovina del giocatore

Come applicazione della distribuzione binomiale, immaginiamo il sequente processo. Una persona lancia una moneta: se esce testa fa un passo avanti; se esce croce fa un passo indietro (vedi figura 7.5). Dove si troverà dopo n lanci? Questo esempio illustra un importante modello di fenomeni casuali, chiamato *cammino* (o *marcia*) *casuale* o (in inglese *random walk*). Con esso è possibile

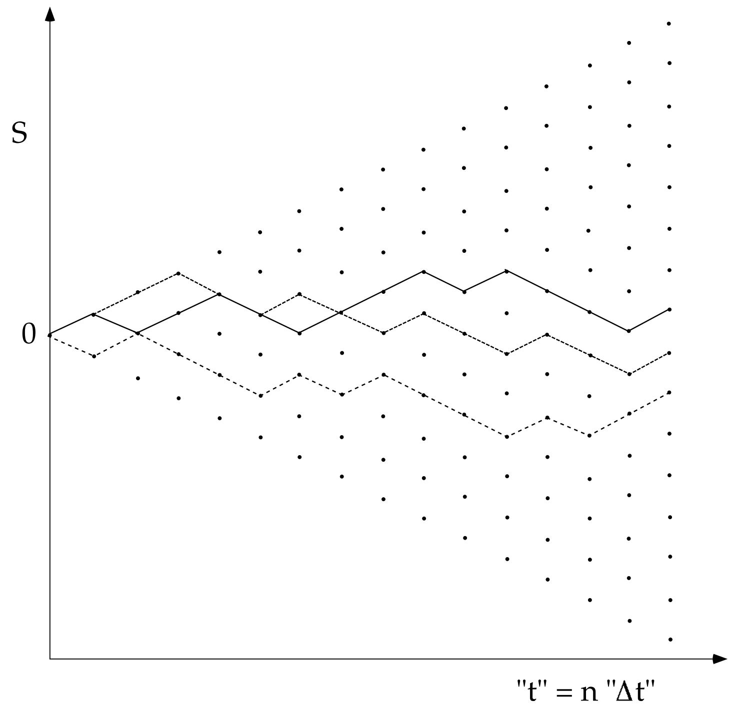


Figura 7.5: Processo di Bernoulli nel dominio del tempo.

descrivere fenomeni interessanti di diffusione come il moto browniano e la combinazione degli errori di misura, e anche la distribuzione di velocità delle molecole.

Per simmetria, il valore atteso della posizione dopo ogni passo è pari a zero (punto iniziale). Siamo quindi interessati a calcolare la sola varianza. Notiamo che se chiamiamo con S la variabile casuale "spostamento in avanti", con X il "numero di teste" e con Y il "numero di croci", abbiamo che $S = X - Y$. Ma poiché $Y = n - X$, abbiamo

$$S = 2X - n \tag{7.26}$$
$$\mathrm{E}(S) = 2\,\mathrm{E}(X) - n = 0 \tag{7.27}$$
$$\mathrm{Var}(S) = \mathrm{Var}(X) = \frac{n}{2} \tag{7.28}$$
$$\sigma(S) = \sqrt{\frac{n}{2}}\,. \tag{7.29}$$

Quindi la deviazione standard cresce come la radice quadrata del numero di passi compiuti. Per ricavare la distribuzione di probabilità di S, notiamo che $P(S = s)$ è uguale a $P(X = (s + n)/2)$. Essendo X distribuita secondo una binomiale, ne segue:

$$\begin{aligned} f(s) &\equiv P(S = s) = P(X = \frac{s+n}{2} \,|\, \mathcal{B}_{n,p}) \\ &= \binom{n}{\frac{s+n}{2}} p^{\frac{n+s}{2}} (1-p)^{\frac{n-s}{2}}\,, \end{aligned} \tag{7.30}$$

che nel nostro caso ($p = 1/2$) diventa

$$f(s) \;=\; \binom{n}{\frac{s+n}{2}}\frac{1}{2^n}\,, \tag{7.31}$$

La marcia a caso serve anche a descrivere un classico problema dei giochi d'azzardo, quello della *rovina del giocatore*. Per risolverlo si valuta la probabilità che un giocatore, iniziando a giocare con una certa somma iniziale $s_\circ$, si trovi senza soldi ad un certo punto del gioco e quindi non possa più tentare la fortuna per rifarsi. In termini di marcia casuale è equivalente è equivalente a cominciare a $s_\circ$ passi da un burrone. Per risolvere il problema, si parte dalle formule che abbiamo appena visto e si calcola la funzione di probabilità della variabile casuale K "arriva in $s_\circ$ allo spostamento k" (arrivare a $s_\circ$ partendo da zero, è equivalente ad arrivare a zero partendo da $s_\circ$).

Random walk, moto browniano ..

7.10 Quanto credere in "$X = \mu \pm \sigma$"?

A questo punto, dopo che sono state incontrate diversi tipi di distribuzioni di probabilità, è opportuno affrontare in modo unitario una questione che nei paragrafi precedenti era stata volutamente tralasciata. Le definizioni di previsione e di incertezza implicano un "certo grado di fiducia" che il numero aleatorio risulti essere nell'intervallo $\mu \pm \sigma$, ma quanto esattamente? Ovvero quanto vale

$$P(\mu - \sigma \leq X \leq \mu + \sigma)\,?$$

Questa probabilità si calcola in modo immediato dalla conoscenza della funzione di distribuzione. Senza dare una dimostrazione generale, alcuni esempi ci convinceranno che, mentre la definizione di incertezza standard mediante la deviazione standard è universale, non è universale il grado di fiducia che la variabile casuale possa cadere fra $\mu - \sigma$ e $\mu + \sigma$.

7.10.1 Alcuni esempi numerici

La tabella 7.4 mostra, per alcuni esempi di distribuzioni, i valori di probabilità legati agli intervalli $\mu \pm \sigma$ e $\mu \pm 2\sigma$. Si noti che, trattandosi di distribuzioni discrete definite per valori interi ed essendo invece μ e σ reali, potrebbe succedere per puro caso che, a seconda dei parametri, l'intervallo $\mu \pm \sigma$ comprenda o no uno o due valori della variabile, producendo una variazione discontinua nel valore della probabilità. Siccome la tabella 7.4 è mostrata a scopo indicativo, gli intervalli sono stati arrotondati ai valori interi.

Si noti come, nonostante le variazione da caso a caso, si possa tranquillamente affermare che c'è "buona" probabilità di trovare la variabile casuale *"entro un sigma"* dal valore atteso, mentre si è abbastanza sicuri che questo si verifichi se si sceglie un *"intervallo di due sigma"*.

Cosa si può dire invece su tale probabilità si conoscono soltanto i valori si m e σ, ma non è dato di sapere il tipo di distribuzione? Poiché valore atteso e varianza sono legati alla distribuzione, ovvero, per dirlo in modo figurato, alla configurazione dei valori della variabile casuale, in qualche modo dovrebbe essere possibile affermare qualcosa di generale su tale configurazione,

distr.	μ	σ	$P(\text{“}X \approx \mu \pm \sigma\text{”})$ (%)	$P(\text{“}X \approx \mu \pm 2\sigma\text{”})$ (%)
$\mathcal{K}_{1,6}$	3.5	1.7	≈ 67	100
$\mathcal{K}_{1,90}$	45.5	26	≈ 58	100
$\mathcal{B}_{\frac{1}{2}}$	0.5	0.5	1	1
$\mathcal{B}_{\frac{1}{10}}$	0.1	0.3	90	90
$\mathcal{B}_{0.99}$	0.99	0.10	99	99
$\mathcal{G}_{\frac{1}{2}}$	2.0	1.4	≈ 88	≈ 97
$\mathcal{G}_{\frac{1}{18}}$	18.0	17.5	≈ 86	≈ 97
$\mathcal{B}_{5,\frac{1}{2}}$	2.5	1.1	≈ 97	≈ 100
$\mathcal{B}_{10,\frac{1}{2}}$	5.0	1.6	≈ 89	≈ 97
$\mathcal{B}_{20,\frac{1}{2}}$	10.0	2.2	≈ 74	≈ 96
$\mathcal{B}_{5,0.8}$	4.0	0.9	≈ 94	≈ 99
$\mathcal{B}_{10,0.8}$	8.0	1.3	≈ 77	≈ 99
$\mathcal{B}_{20,0.8}$	16.0	1.8	≈ 84	≈ 99
$\mathcal{P}_1$	1.0	1	≈ 92	≈ 98
$\mathcal{P}_5$	5.0	2.2	≈ 74	≈ 97
$\mathcal{P}_{20}$	20.0	4.5	≈ 69	≈ 97
$\mathcal{P}a_{2,\frac{1}{2}}$	4.0	2.0	≈ 89	≈ 96
$\mathcal{P}a_{5,\frac{1}{2}}$	10.0	3.2	≈ 76	≈ 96
$\mathcal{P}a_{10,\frac{1}{2}}$	20.0	4.5	≈ 70	≈ 97
“Cebicev”	-	-	≥ 0	≥ 75

Tabella 7.4: Probabilità che il numero aleatorio X sia compreso nell'intervallo di una o due σ intorno alla sua previsione, valutata per alcune distribuzioni di probabilità. $P(\text{“}X \approx \mu \pm k\sigma\text{”})$ sta per $P(\mu - k\,\sigma \lessapprox X \lessapprox \mu + k\,\sigma)$, dove “$\approx$” indica che le valutazioni sono in genere approssimate arrotondando le ampiezze degli intervalli a valori interi. Nell'ultima riga (“Cebicev”) è anche riportato il limite ottenuto dall'uguaglianza di Cebicev (vedi paragrafo 7.10.3).

che si rifletta sui valori di probabilità di interesse. Questo è quanto affermato dalla *disuguaglianza di Cebicev*, dimostrata a partire dalla *disuguaglianza di Markov*

7.10.2 Disuguaglianza di Markov

Questa disuguaglianza permette di stabilire un limite superiore al valore di probabilità dalla sola conoscenza del valore atteso μ, a condizione che la variabile casuale sia definita non negativa. Dato un valore valore $\alpha > \mu$ si ha che

$$P(X \geq \alpha) \leq \frac{\mu}{\alpha}\,. \tag{7.32}$$

Difatti, dalla definizione operativa di valore atteso e scegliendo un valore di k tale che per $i \geq k$ X sia maggiore di α, segue

$$\begin{aligned} \mu = \sum_{i=1}^{n} x_i f(x_i) &\geq \sum_{i=k}^{n} x_i f(x_i) \\ &\geq \sum_{i=k}^{n} \alpha f(x_i) = \alpha \sum_{i=k}^{n} f(x_i) = \alpha\, P(X \geq \alpha)\,, \end{aligned}$$

da cui segue la (7.32). Per esempio, la sola conoscenza di $\mu = 2$ implica $P(X \geq 6) \leq 1/3$. Se però si venisse a sapere che la distribuzione è di Poisson il valore di probabilità sarebbe dell'1.7 %; se fosse una binomiale di $p = 0.2$ esso varrebbe 0.7; se una geometrica 1.6%. Tutti i valori sono compresi entro il limite dato dalla (7.32).

7.10.3 Disuguaglianza di Cebicev

La disuguaglianza di Cebicev afferma che *la probabilità che lo scarto fra il valore della variabile casuale e la previsione di essa ecceda* (o sia uguale) *in modulo k volte σ non è maggiore di* $1/k^2$ (con $k \geq 1$) . In formule:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}\,. \tag{7.33}$$

Difatti il quadrato dello scarto $Y = (X - \mu)^2$ è una variabile casuale non negativa, che - per definizione di varianza - ha valore atteso $\mu_Y = \sigma^2$. Scegliendo un valore $\alpha = (k\,\sigma)^2$, con $k \geq 1$ in modo tale che $\alpha \geq \sigma^2$, possiamo applicare la disuguaglianza ad Y di Markov:

$$\begin{aligned} P(Y \geq \alpha) &\leq \frac{\mu_Y}{\alpha} \\ P((X-\mu)^2 \geq (k\,\sigma)^2) &\leq \frac{\sigma^2}{(k\,\sigma)^2}, \end{aligned}$$

da cui segue la (7.33) in quanto $(X - \mu)^2 \geq (k\,\sigma)^2$ è equivalente a $|X - \mu| \geq k\sigma$.

Queste disuguaglianze sono di scarso valore pratico, in quanto è veramente raro il caso di non sapere assolutamente niente sul tipo di distribuzione, e la regola normativa della scommessa coerente che sta dietro le affermazioni di probabilità richiede di prendere in considerazione ogni informazione sugli eventi (invece di accontentarsi soltanto dei limiti "di sicurezza"). Le utilizzazioni della disuguaglianza sono più di carattere teorico che applicativo. Infatti, è raro essere nelle condizioni di conoscenza per le quali vale tale teorema. Invece, le disuguaglinze permettono di dimostrare dei teoremi limite in modo indipendente dalla distribuzione di probabilità (vedi ad esempio paragrafo 10.9.2).

Come esempio, riprendiamo quello già visto per illustrare la disuguaglianza di Markov. L'ulteriore conoscenza di $\sigma = 1.4$ (quanto si avrebbe per $\mathcal{P}_2$ o $\mathcal{G}_{\frac{1}{2}}$) o 1.3 (il caso di $\mathcal{B}_{10,0.2}$) modifica la probabilità di $P(X \geq 6)$ rispettivamente in $\leq 12.3\,\%$ e $\leq 10.6\,\%$, limiti meno laschi di quelli ottenuti dalla disuguaglianza precedente, ma ancora lontani dai valori che l'esatta conoscenza delle distrubuzioni forniscono.

7.11 Intervalli di probabilità, o di credibilità

Un problema simmetrico a quello mostrato nel paragrafo precedente consiste nel definire un intervallo entro cui c'è una certa probabilità di trovare la variabile casuale. Quindi si possono trovare intervalli al 50 % al 90 % e così via. Da quanto abbiamo appena visto segue che l'ampiezza di tale intervallo dipende dal tipo di distribuzione e dai suoi parametri.

Rispetto al problema precedente sorgono due complicazioni:

- la prima è peculiare delle variabili discrete e scompare per quelle continue: fissata una certo livello di probabilità non è sempre possibile fissare un intervallo in cui la variabile casuale abbia esattamente tale probabilità di verificarvisi; in tale caso i livelli di probabilità sono da intendere a livello approssimativo;

- la seconda è più generale e dipende dal fatto che, fissato un certo livello di probabilità ci sono infiniti intervalli che soddisfano tale condizione (eventualmente da intendersi in modo approssimativo, vedi punto precedente). Spesso si fa riferimento a *intervalli centrali*, intendendo intervalli centrati intorno al valore atteso, o intervalli asimmetrici intorno al valore atteso, tale che sia a destra che a sinistra del valore atteso ci sia la metà del valore di probabilità richiesto.

Dal punto di vista di linguaggio naturale, gli intervalli di credibilità pottebbero essere chiamanti tranquillamente anche *intervalli confidenza*. Purtroppo, l'espressione "intervallo di confidenza" viene usata ad indicare un concetto che, in principio non dovrebbe significare intervallo di probabilità, ma che in pratica è usato usato come tale, provocando tanta confusione, a cominciare dagli statistici sostenitori della validità di tale concetto. Questo è un argomento che riguarderà l'inferenza statistica e sarà trattato nella sede adatta. Per ora si vuole soltanto mettere in guardia il lettore.

p	media	moda	mediana
0.16	1.6	1	1
0.18	1.8	1	2
0.20	2.0	2	2
0.22	2.2	2	2
0.24	2.4	2	2
0.26	2.6	2	3
0.28	2.8	3	3

Tabella 7.5: Confronto fra media, moda e mediana per una distribuzione binomiale avente $n = 10$ e per diversi valori di p.

7.12 ❄Previsione, penalizzazione e valore sul quale scommettere

Abbiamo parlato per esteso del valore atteso come media della distribuzione e abbiamo accennato alla moda e alla mediana, possibili alternative per quantificare sinteticamente la ditribuzione, definite rispettivamente come il valore di massima probabilità e quello centrale (nel senso di 50 di probabilità a sinistra e 50 % a destra).

Nel capitolo precedente abbiamo discusso dei diversi vantaggi della media. Anche parlando della distribuzione binomiale, abbiamo visto come la media dipenda linearmente da p e riproduca il valore intuitivo $n\,p$ per il valore atteso. La tabella 7.5 mostra come cambiano i valori di media, moda e mediana al variare di p, avendo fissato n.

C'è un altro aspetto interessante di queste tre grandezze legato al *valore sul quale scommettere*, relativamente alle condizioni di vincita, o meglio alle condizioni di penalizzazione (vedi paragrafo 2.17). È evidente che

- nel caso si vinca soltanto se si verifica esattamente il valore sul quale si è scommesso conviene puntare sul valore più probabile, ovvero sulla moda.

Se invece la possibilità di vincita è legata all'avvicinarsi al valore che si verificherà, è in genere opportuno non scommettere sul valore più probabile, almeno che la distribuzione di probabilità non sia simmetrica intorno alla moda. Il valore esatto sul quale scommettere dipendere da quanto ci si rimette (ovvero da quanto poco si vince) a commettere errori di valutazione. In particolare, si può dimostrare che

- se la penale (o il mancato guadagno) dipende dal modulo dello scarto rispetto al valore che si verificherà, la previsione di guadagno è massima quando si scommette sulla mediana;

- se la penale va invece come il quadrato dello scarto, o vince realizza il minor quadrato dello scarto, conviene puntare sulla media aritmetica.

La dimostrazione di questo secondo caso è immediata, in quanto la previsione di guadagno (negativo, in quanto si tratta di una penale!) non è altro che l'opposto della varianza. E poich´e la media aritmetica minimizza la varianza, essa minimizza anche la previsione di penale, ovvero massimizza la previone di guadagno.

Un ultimo caso in cui moda e mediana sono senz'altro utili è quando la media aritmetica non esiste (nel senso che la formula che la definisce non converge a un valore finito), pur avendo la distribuzione di probabilità una forma abbastanza regolare (o addirittura simmetrica) tale che moda e/o mediana possano ancora essere usati come indici di centralità.

7.13 ○ Previsione di frequenza relativa e legge dei grandi numeri

Consideriamo la distribuzione binomiale degli X successi su n prove. Dividendo il numero di successi per il numero di prove otteniamo una nuova variabile casuale

$$W = \frac{X}{n}$$

associata alla *frequenza relativa* di successi. La distribuzione di W può essere ricavata direttamente dalla binomiale in quanto

$$P(W = \frac{x}{n}) = P(X = x)\,.$$

L'espressione della distribuzione di probabilità di W si ottiene direttamente da quella di X:

$$f(w\,|\,\mathcal{B}_{n,p}) = \binom{n}{nw} p^{nw}\,(1-p)^{n(1-w)} \qquad \begin{cases} n = 1, 2, \ldots, \infty \\ 0 \le p \le 1 \\ w = 0, \frac{1}{n}, \frac{2}{n} \ldots, 1 \end{cases} \tag{7.34}$$

Per il valore atteso e la sua incertezza abbiamo:

$$\mathrm{E}(W) = \mathrm{E}\left[\frac{X}{n}\right] = \frac{\mathrm{E}(X)}{n} = \frac{n\,p}{n} = p \tag{7.35}$$

$$\sigma(W) = \frac{\sigma(X)}{n} = \frac{\sqrt{n\,p\,q}}{n} = \frac{\sqrt{p\,q}}{\sqrt{n}} \tag{7.36}$$

La previsione della frequenza relativa è pari alla probabilità di ciascuno dei processi elementati di Bernoulli. Inoltre l'incertezza di previsione decresce come $1/\sqrt{n}$, ovvero *al crescere di n diventa sempre meno probabile trovare valori di frequenza che differiscono molto dalla probabilità*. Questo è uno dei modi di esprimere la *legge dei grandi numeri*, sulla quale ritorneremo con maggiore dettaglio nel capitolo 10.

Per ora utilizzeremo la variabile casuale frequenza relativa per parlare delle previsioni di distribuzioni statistiche e per introdurre la problematica della verifica delle leggi statistiche.

7.14 ⟳ Previsione di una distribuzione statistica

Prendiamo ancora una volta in considerazione un numero aleatorio ai cui valori assegniamo una probabilità secondo la distribuzione binomiale, ad esempio il numero di teste ottenibili dal lancio di $m = 5$ monete regolari[4]. Supponiamo ora di voler ripetere l'esperimento $n = 1000$ volte. Il nostro interesse si rivolge ora al numero di volte che osserveremo 0 teste, quello di 1 testa, e così via, fino al numero di 5 teste. Indichiamo ora con X_i il numero aleatorio "numero di occorrenze di i teste in n esperimenti". In ciascun esperimento la probabilità di i teste è data dalla binomiale

$$\mathcal{B}_{m=5,\frac{1}{2}}\,,$$

ovvero

$$p_i = P(\text{“}i\text{ teste”}) = f(i\,|\,\mathcal{B}_{m=5,\frac{1}{2}})\,.$$

Ne segue che

$$X_i \sim \mathcal{B}_{n=1000,p_i}\,.$$

Quindi, per ciascun numero di teste, abbiamo una previsione di

$$X_i = p_i\,n \pm \sqrt{p_i\,q_i\,n}\ \text{occorrenze}\,. \tag{7.37}$$

Analogalmente, per ciascuna delle frequenze relative W_i abbiamo una previsione di

$$W_i = p_i \pm \frac{\sqrt{p_i\,q_i}}{\sqrt{n}} \tag{7.38}$$

Poiché le variabili i, ciascuna associata al numero di occorrenze (o alla frequenza relativa), riferite ad eventi osservati costituiscono una distribuzione statistica, le (7.37) e (7.38) possono essere interpretate come *previsione della distribuzione statistica che potrà essere osservata* ripetendo n volte un certo esperimento sotto le stesse ipotesi. La figura 7.6 mostra la previsione della distribuzione statistica appena discussa.

Con queste considerazioni comincia l'interessante discorso sulla relazione fra distribuzione statistiche e distribuzioni di probabilità che si presenta sotto due aspetti

- previsione di distribuzione statistiche a partire da distribuzioni di probabilità (il caso appena visto);

- inferenza di una distribuzione di probabilità a partire da una distribuzione statistica (un primo esempio sarà mostrato nel prossimo paragrafo, mentre il discorso sarà affrontato in modo più organico a partire dal capitolo 11).

[4]Si faccia attenzione ai diversi simboli per le varie binomiali che entrano in gioco.

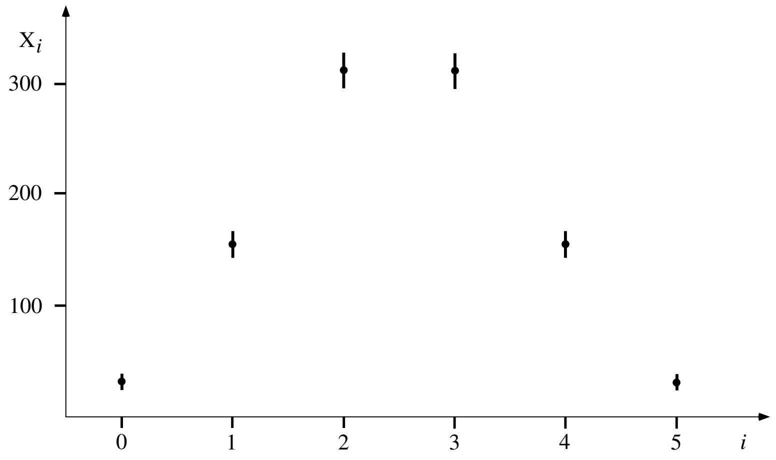


Figura 7.6: Previsione della distribuzione statistica del "numero di teste" nel lancio di 5 monete ottenibile ripetendo tale esperimento 1000 volte. Le barre rappresentano le incertezze di previsione, quantificate in $\pm\, 1\,\sigma$ intorno alla previsione stessa.

7.14.1 Introduzione al concetto di correlazione fra variabili casuali

Terminiamo con una breve introduzione su un argomento che tratteremo più in dettaglio nel capitolo ***correlazione***. Fare una previsione di una distribuzione statistica comporta inevitabilmente una trattazione simultanea di molte variabili casuali (le X_i dell'esempio precedente). Abbiamo visto come la valutazione di ciascuno dei valori attesi non richieda nuovi concetti. Si presenta invece un problema nuovo quando vogliamo fare una previsione globale della distribuzione. Ad esempio, mentre riteniamo ragionevole che la variabile X_i cada fra $\mathrm{E}(X_i)$ e ∞, è impossibile che <u>ciascuna</u> delle variabili sia compresa in quell'intervallo, in quanto implicherebbe che

$$\sum_i X_i > \sum_i \mathrm{E}(X_i) \quad = \quad \sum_i n\, p_i = n \sum_i p_i = n\,,$$

cosa impossibile in quanto la somma di ciascuna delle possibili frequenze deve essere uguale al numero totale di esperimenti. Ciò significa che alcune di queste possono essere valori maggiori delle loro previsioni se sono compensate da altre che assumono valori inferiori. Si dice che queste variabili casuali sono *correlate* (per ora nel senso che questo termine può significare nel linguaggio comune). Un caso estremo è quando si hanno soltanto due variabili, ad esempio gli esperimenti precedentemente descritti consistano nel lancio di una sola moneta ($m = 1$). X_1 sarà il numero di teste e X_0 il numero di croci. Su $n = 1000$ esperimenti ci aspettiamo 500 ± 16 teste e 500 ± 16 croci, ma la loro somma non è un numero aleatorio (deve dare con certezza 1000) e quindi le due variabili sono *linearmente dipendenti* ($X_1 = n - X_\circ$) e *completamente anticorrelate*.

# di morti	0	1	2	3	4	≥ 5
# di occorrenze	109	65	22	3	1	0

Tabella 7.6: Distribuzione statistica del numero di morti l'anno, per reggimento, dovuti a calcio di cavallo. I dati si riferiscono a 20 anni di osservazioni di 10 reggimenti dell'esercito prussiano (1875-1894).

7.15 ⥁ Un esempio storico di distribuzione di Poisson come introduzione al problema della verifica delle leggi statistiche

Nel 1898 ** Bortkiewicz pubblicò uno studio sui decessi di soldati dell'esercito prussiano in seguito a calcio di cavallo. Analizzando i verbali di 20 anni di 10 reggimenti constatò che c'erano stati in totale 122 morti dovuti a quel tipo di incidenti. La tabella 7.6 mostra la frequenza annuale per reggimento del numero di morti. Supponendo che i reggimenti siano equivalenti e che fenomeno non dipenda dal tempo, questi dati possono essere pensati come 200 osservazioni di un anno su un reggimento tipo.

7.15.1 Previsione del tipo di distribuzione

Cerchiamo di capire che tipo di distribuzione di probabilità ci aspettiamo.

- Possiamo dividere l'anno di osservazione in tanti piccoli intervalli di tempo (ad esempio ogni giorno), in cui la probabilità che ci sia uno di tali incidenti mortali sia abbastanza piccola e quella di più di un incidente trascurabile.

- In alternativa si può pensare al gran numero di soldati di un reggimento, ciascuno dei quali ha una bassissima probabilità di morire da calcio di cavallo.

Quindi ci aspettiamo che il numero di morti l'anno per reggimento segua una distribuzione di Poisson.

7.15.2 Stima "puntuale" del parametro della distribuzione

Per ottenere previsioni quantitative servono delle ipotesi sul parametro parametro λ, o, in ultima analisi, su r, il numero di morti per reggimento per unità di tempo. Questo può essere stimato *assumendo la regolarità del fenomeno* rispetto al tempo e rispetto ai reggimenti e ritenendo quindi il tasso atteso circa uguale al tasso osservato nel passato[5]:

$$r \approx \overline{d} = \frac{\sum_i d_i\, w_i}{\sum_i w_i}\,,$$

da cui

$$\begin{aligned} r &\approx 0.61 \frac{\text{morti}}{\text{reggimento} \times \text{anno}} \\ &\approx 1.67 \times 10^{-3} \frac{\text{morti}}{\text{reggimento} \times \text{giorno}}\,. \end{aligned}$$

Ne segue che λ su un anno vale 0.61. Da questo valore possiamo calcolare tutti i gradi di fiducia sul possibile numero di incidenti. Ad esempio:

$$p_\circ = P(D=0) \;=\; f(0\,|\,\mathcal{P}_{0.61}) = 0.543\,,$$

ovvero crediamo al 54.3 % che un reggimento non avrà nessun incidente di quel tipo in un anno (abbiamo indicato con D la variabile casuale “numero di morti” e abbiamo chiamato p_0 la probabilità che D valga 0).

7.15.3 Previsione quantitativa della distribuzione statistica, subordinata a $\lambda = \overline{d}$, e confronto con le osservazioni

Estendiamo ora le previsioni a $n = 200$ ipotetici reggimenti reggimenti che si comportino in quel modo (ma diversi dai 200 osservati). In quanti ci aspettiamo che non si verifica nessun morto? Se indichiamo con $X_\circ$ la variabile casuale “numeri di volte in cui si registra nessun morto”, questa sarà distribuita secondo una binomiale:

$$X_\circ \;\sim\; \mathcal{B}_{n,p_\circ} = \mathcal{B}_{200,\,0.543}\,.$$

Ne segue che

$$\begin{aligned} \mathrm{E}(X_\circ) &= p_\circ\, n = 108.6 \\ \sigma(X_\circ) &= 7.0\; : \end{aligned}$$

ci aspettiamo di trovare 109 ± 7 reggimenti su 200 in non si verifichi alcun morto in un anno. La tabella 7.6 riporta un valore di 109. Un caso di accordo perfetto al quale non va attribuito nessun significato particolare (se qualche ombra di dubbio che i dati siano stati truccati…). Riportiamo nella tabella 7.7 i dati originali, accompagnati dalle previsioni e incertezze di previsioni subordinate ad un processo di Poisson di $\lambda = 0.61$. Si noti l’ottimo accordo fra dati e previsioni (entro le incertezze di queste ultime).

Questo esempio, oltre alla funzione propedeutica alla problematica della “verifica”[6] delle *leggi statistiche*, rappresenta anche un interessante caso di una quantità calcolata mediante una distribuzione di probabilità (p_i) che viene successivamente usata come parametro di un’altra distribuzione.

[5]Si noti il salto logico rispetto alla semplice valutazione della probabilità dalla frequenza di ciascuna classe di eventi, ovvero, la probabilità di $D = 0$, ad esempio, non è valutata dalla frequenza di zero morti nel passato. Quindi, anche se si stanno utilizzando informazioni statistiche la probabilità non è valutata secondo il paradigma frequentista.

[6]Come dicevamo nel primo capitolo, la natura probabilistica di tali leggi preclude ogni verificabilità oggettiva. Sono i nostri pregiudizi sulla regolarità delle leggi della natura a convincerci che ragionevolmente il processo si sia svolto in quel modo.

	dati sperimentali		previsioni basate su $\mathcal{P}_{\lambda=\overline{d}=0.61}$				
			su d_i	sulle frequenze di d_i			
d_i	x_{sp_i}	w_i (%)	p_i (%)	$\mathrm{E}(X_i)$	$\sigma(X_i)$	$\mathrm{E}(W_i)$ (%)	$\sigma(W_i)$ (%)
0	109	54.5	54.3	108.6	7.0	54.3	3.5
1	65	32.5	33.1	66.2	6.7	33.1	3.4
2	22	11.0	10.1	20.2	4.3	10.1	2.1
3	3	1.5	2.1	4.2	2.0	2.1	1.0
4	1	0.5	0.3	0.6	0.8	0.3	0.4
≥ 5	0	0	0.04	0.1	0.3	0.04	0.2

Tabella 7.7: Confronto fra dati sperimentali e previsioni calcolate assumendo un processo di Poisson con $\lambda = 0.61$. d_i indica il numero di morti in un anno, x_{sp_i} il numero di occorrenze osservate sperimentalmente, w_i la frequenza relativa delle occorrenze; p_i è la probabilità di d_i subordinata ad un processo di Poisson con $\lambda = 0.61$; X_i rappresenta il numero di occorrenze della i-ma possibilità e W_i la rispettiva frequenza relativa.

Inferenza probabilistica su λ

Un ultimo commento sull'uso inferenziale (e non di soltanto verifica) delle distribuzioni statistiche. Immaginiamo di voler calcolare la probabilità che si verifichino esattamente 2 morti in una futura osservazione (ad esempio un ipotetico 201^{mo} reggimento[7]). Dalla tabella 7.7 si vede che la probabilità non è stata calcolata semplicemente dalla frequenza, bensì dall'ipotesi di un processo di Poisson con $\lambda = 0.61$. Ma come si capirà bene, non possiamo essere assolutamente certi di tale valore di λ. Ad esempio, l'eventualità di un morto in più o in meno, ragionevolissima alla luce del tipo di "esperimento", avrebbero suggerito valori di λ di 0.605 e 0.615. Anche se λ può assumere valori reali positivi con continuità, immaginiamo per un momento di poterne considerare un certo numero discreto n_λ e indichiamo le possibilità con λ_j (ad esempio 0.605, 0.610, 0.615, etc.), ciascuna con grado di fiducia $P(\lambda = \lambda_j) = f(\lambda_j \,|\, \text{dati}, I_\circ)$ (avendo esplicitato il fatto che le probabilità sono condizionate dall'osservazione di certi dati sperimentali e da un certo stato di informazione iniziale $I_\circ$, vedi capitolo **??**).

Sorvoliamo sul modo con il quale viene stimata la $f(\lambda_j \,|\, \text{dati}, I_\circ)$, argomento che riprenderemo dai prossimi paragrafi, e interessiamoci soltanto alle consequenze sulla previsione della distribuzione statistica.

[7]Si noti l'importanza di considerare un altro reggimento, invece di estrarre a caso uno dei duecento dai quali è stata ricavata la distribuzione statistica. In questo caso la probabilità è data dal numero dei "casi" favorevoli (22) diviso il numero dei casi possibili (200) (si pensi all'estrazione di uno dei duecento resoconti dei reggimenti, ciascuno indicante il numero di morti da calcio di cavallo).

Previsione della distribuzione statistica subordinata all'incerteza su λ

Vediamo ora come l'incertezza su λ si propaga sull'incertezza sui futuri esiti, e quindi sulla possibile distribuzione statistica che sarà osservata. Concentriamoci, tanto per fare un esempio, sul valore $D = 2$ (due morti). Subordinatamente all'ipotesi $\lambda = \lambda_j$ si calcola:

$$P(D = 2 \,|\, \mathcal{P}_{\lambda_i}) \;\; = \;\; f(2 \,|\, \mathcal{P}_{\lambda_j}) = \frac{e^{-\lambda_i}\,\lambda_i^2}{2}\,.$$

Pesando le varie ipotesi con le loro probabilità (legge delle alternative, vedi paragrafo 4.9.2):

$$P(D = 2) \;\; = \;\; \sum_{j=1}^{n_\lambda} P(D = 2 \,|\, \mathcal{P}_{\lambda_j}) \cdot P(\lambda = \lambda_j)\,.$$

Scrivendo le probabilità in termini delle funzioni di probabilità otteniamo

$$p_2 \equiv P(D = 2 \,|\, \text{dati}, I_\circ) \;\; = \;\; \sum_{j=1}^{n_\lambda} f(2 \,|\, \mathcal{P}_{\lambda_j}) \cdot f(\lambda_j \,|\, \text{dati}, I_\circ)\,.$$

L'espressione generale della distribuzione di probabilità di D diventa:

$$p_i = f(d_i \,|\, \text{dati}, I_\circ) = \sum_{j=1}^{n_\lambda} \frac{e^{-\lambda_j}\,\lambda_j^i}{i!} f(\lambda_j \,|\, \text{dati}, I_\circ)\,.$$

*** per fare le cose bene si dovrebbe introdurre la multinomiale ***

7.16 ↻ Estensione dei teoremi sulla probabilità alle funzioni di probabilità discrete

Avendo la funzione di probabilità di variabili casuali discrete il significato di probabilità, si possono applicare ad esse tutte le proprietà delle probabilità incontrate nel capitolo chap:RegoleProb. Basta sostituire al simbolo $P(\cdot)$ il simbolo $f(\cdot)$ e ai generici eventi E_i i valori assunti dalle variabili casuali. Ad esempio,

$$\begin{aligned}
P(X = x \,|\, Y = y) \;\; &\rightarrow \;\; f(x \,|\, y)\,, \\
P(X = x) = \sum_y P(X = x \,|\, Y = y) \cdot P(Y = y) \;\; &\rightarrow \;\; f(x) = \sum_y f(x \,|\, y) f(y) \\
P(X = x \cap Y = y) \;\; &\rightarrow \;\; f(x, y)\,,
\end{aligned}$$

e così via. In particolare l'ultimo esempio mostra come si costruisce una variabile doppia, argomento sul quale ritorneremo nel capitolo 9 (non perché entrino in gioco concetti particolari, ma solo per seguire un certo ordine di esposizione del matriale). Vediamo ora come si estende il teorema di Bayes, facendo degli esempi di inferenza che ci serviranno di preparazione ai problemi di inferenza sui valori di grandezze fisiche.

Se abbiamo due variabili casuali X e Y, la probabilità che Y assuma il valore y, subordinatamente all'informazione che X assuma il valore x è data da:

$$P(Y = y \,|\, X = x) \propto P(X = x \,|\, Y = y) \cdot P(Y = y)\,.$$

In termini di funzioni di probabilità essa può essere scritta come

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

che, opportunamente normalizzata, diventa

$$f(y \,|\, x) \quad = \quad \frac{f(x \,|\, y)\, f(y)}{\sum_y f(x \,|\, y)\, f(y)} \tag{7.39}$$

Questa è l'espressione del teorema di Bayes per variabili casuali discrete.

nome	simbolo	funzione	m	σ
uniforme	$\mathcal{K}_{1,n}$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{\sqrt{n^2-1}}{12}$
Bernoulli	$\mathcal{B}_p$	$p^x q^{1-x}$	p	$\sqrt{p\,q}$
geometrica	$\mathcal{G}_p$	$p\,q^{x-1}$	$\frac{1}{p}$	$\frac{\sqrt{q}}{p}$
binomiale	$\mathcal{B}_{n,p}$	$\binom{n}{x}\,p^x\,q^{n-x}$	$n\,p$	$\sqrt{n\,p\,q}$
Poisson	$\mathcal{P}_\lambda$	$\frac{\lambda^x}{x!}\,e^{-\lambda}$	λ	$\sqrt{\lambda}$
Pascal	$\mathcal{P}a_{k,p}$	$\binom{x-1}{k-1}\,p^k(1-p)^{x-k}$	$\frac{k}{p}$	$\sqrt{k}\,\frac{\sqrt{q}}{p}$
binomiale negativa	$\mathcal{B}^-_{k,p}$	$\binom{x+k-1}{k-1}\,p^k(1-p)^x$	$\frac{k\,q}{p}$	$\sqrt{k}\,\frac{\sqrt{q}}{p}$
ipergeometrica	$\mathcal{H}_{N,M,n}$	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$n\,r$	$\sigma_B\sqrt{\frac{N-n}{N-1}}$

Tabella 7.8: Sommario delle varie distribuzioni incontrate, con relativi valori attesi e deviazioni standard. In genere, q sta per $1-p$. Nella ipergeometrica r sta per M/M e "σ_B" sta per la deviazione standard della binomiale di $p = r$ e stesso n. Per i domini delle funzioni e il tipo di parametri si rimanda al testo.

7.17 Ricapitolando

- La distribuzione binomiale e quella di Poisson sono le distribuzioni di variabile discreta di maggiore interesse per le applicazioni pratiche concernenti i problemio di misura. La prima è legata a misure di proporzioni (e quindi efficienze di rivelazione o di analisi e piccoli campionamenti di popolazioni). La seconda è legata a molti fenomeni di conteggio.

- È importante il ruolo unificatore del processo di Bernoulli, il quale dà origine alla binomiale e, qualora si considerino un numero di possibili processi virtualmente infinito e di probabilità infinitesima, alla distribuzione di Poisson.

- È particolarmente interessante il processo di Poisson considerato rispetto alla variabile tempo (gli aspetti legati ai tempi di attesa saranno descritti nel prossimo capitolo).

- La distribuzione di Pascal rappresenta la generalizzazione della distribuzione geometrica nel caso che ci si interessi all'attesa di più eventi. La binomiale negativa è una distribuzione complementare a quella di Pascal.

- La distribuzione ipergeometrica è legata alle estrazioni senza reintroduzione. La corretta applicazione della distribuzione è importante quando si hanno campioni di numerosità confrontabile con quella della popolazione stessa. Quando la grandezza della popolazione è molto maggiore di quella del campione si riottiene la distribuzione binomiale, in quanto le ripetute estrazioni, sebbene senza reintroduzione, non modificano apprezzabilmente la composizione relativa della popolazione e quindi nemmeno la probabilità di estrazione del carattere di interesse.

- La tabella 7.8 riassume le principali distribuzioni di probabilità incontrate in questo capitolo e nel precedente. La figura **??** dà una visione d'insieme delle relazioni fra le diverse distribuzioni. *** far partire tutto dal modello di estrazione: reintroduzione: Bernoulli; non reintroduzione: ipergeopetrica;

- La distribuzione binomiale (e quindi il sottostante processo di Bernoulli) serve anche a schematizzare il cammino casuale, modello importante di comportamenti stocastici, fra i quali il moto browniano e la combinazione di errori di misura.

- L'espressione della previsione di una variabile casuale, "più o meno" la sua incertertezza standard, ha un contenuto probabilistico che dipende dal tipo e dai parametri della distribuzione. C'è comunque una probabilità abbastanza alta (tipicamente $\gtrapprox$ 60-70 %) di trovare la variabile in tale intervallo. Considerando intervalli di 2 o 3 deviazioni standard intorno al valore atteso si arriva, per molte distribuzioni di utilità applicativa, a livelli di "certezza pratica".

- Nei casi (di irrilevanza pratica) di assoluta ignoranza sul tipo di distribuzione, la disuguaglianza di Cebicev, consequenza del teorema di Markov, permette di valutare limiti inferiori alla probabilità di trovare la variabile entro un certo numero di deviazioni standard dal valore atteso.

- Per i processi di Bernoulli, è interessante anche la variabile casuale "frequenza relativa di successi". Si mostra facilmente che la sua previsione è pari alla probabilità di ciascuna delle prove, mentre l'incertezza di previsione decresce come la radice quadrata del numero di prove (legge dei grandi numeri).

- La previsione di una distribuzione statistica è data dai valori che la variabile casuale può assumere, per ciascuno dei quali viene data la previsione (con relativa incertezza) delle frequenze che si osserverano se l'esperiemento viene ripetuto molte volte sotto condizioni ipoteticamente identiche.

  Si noti il ruolo di variabile casuale prima legato a "quanto si andrà ad osservare nel singolo esperimento" e, successivamente, al "numero di occorrenze con le quali si verificherà ciascuno dei possibili esiti dell'esperimento". Qualsiasi sia la distribuzione che descrive "quanto si andrà ad osservare", la distribuzione del "numero di occorrenze è sempre una binomiale, qualora si considerino uno alla volta i possibili esiti.

Quando invece vengono considerati globalmente tutti i possibili “numeri di occorrenze” le cose si complicano per effetto di correlazioni. Questi effetti saranno trattati nel capitolo ***.

- L’esempio storico dei “morti da calcio di cavallo” viene utilizzato sia come applicazione della previsione di una distribuzione statistica, che come introduzione alla verifica delle leggi statistiche e all’inferenza di una distribuzione di probabilità a partire da una distribuzione statistica.

- I teoremi sulle probabilità si applicano “tout court” alle funzioni di probabilità di variabili discrete. In particolare, il teorema di Bayes permette di fare inferenze sui possibili valori dei parametri delle distribuzioni che possono aver prodotto le distribuzioni statistiche osservate.

- L’esempio dei “morti da calcio di cavallo” mostra in modo chiaro come la probabilità non venga stimata soltanto dalla frequenza con cui l’evento di interesse si è verificato nel passato, ma anche dalle frequenze di altri eventi legati allo stesso fenomeno (l’intera distribuzione statistica) e dal modello assunto per descrivere il fenomeno stesso.

7.18 Problemi

1. La Ferrero dichiara che 1 ovetto Kinder su 5 contiene una sorpresa "interessante". Un bambino compra 6 ovetti. Quanto vale la probabilità che trovi esattamente 2 sorprese? Che ce ne trovi almeno una?
2. Seguito del problema precedente: 8 bambini acquistano ciascuno 6 ovetti. Quanto vale la probabilità che esattamente 3 bambini trovino due sorprese?
3. Sempre sui problemi precedenti: un bambino acquista ogni giorno, mentre va a scuola, due ovetti. Calcolare la previsione di giorni che devono passare affinché gli capiti di vincere due sorprese la stessa mattina.
4. Una variabile casuale X segue una distribuzione binomiale di valor medio 4 e varianza 3. Trovare la probabilità che la variabile casuale assuma un valore minore di 5.
5. Un esperimento consiste in 2 lanci di 5 monete diverse, lanciate simultaneamente. Le monete sono: 100 Lit, 500 Lit, 1 DM, 1 FS e 10 FF. Quante vale la probabilità che in totale si verifichino esattamente 3 teste?
6. Due monete vengono saldate fra di loro in modo tale che risultino affiancate e che la testa dell'una sia dalla stessa parte della croce dell'altra. Le due monete sono lanciate per cinque volte. Quanto vale la probabilità di osservare in totale esattamente 3 teste?
7. Un esperimento consiste in: lancio di una moneta, lancio di un dado, estrazione di un numero della tombola, estrazione di un numero alla roulette ed estrazione di una carta da gioco da un mazzo di 40 carte. Si definisce evento favorevole nei diversi giochi il verificarsi, rispettivamente, di: testa; numero 6; numero maggiore di 30; numero 8; denari. Quanto vale la probabilità di avere un numero totale di successi minore o uguale ad 1? Quanto vale la probabilità di ottenere 5 eventi favorevoli?
8. Un ingegnere della motorizzazione sa per sua esperienza che il 60 % dei candidati riescano a passare l'esame di guida. Se in un giorno dovrà esaminare 10 candidati, in quanto stimerà la probabilità che almeno 8 siano promossi, senza nessun'altra informazione (nemmeno guardandoli in faccia o sapendo età, sesso, etc.)?
9. Risolvere il problema sulla suddivisione della posta in caso di interruzione del gioco (n. 31 del capitolo 2) considerando la distribuzione di probabilità del numero di vittorie dei due giocatori nelle partite residue.
10. Variazione sul tema del problema delle sei scatole (nr. 18 e 19 del capitolo 5). Immaginiamo che la scatola scelta sia stata preparata scegliendo a caso 5 palline da un grande sacco che conteneva in proporzione uguale palline bianche e nere. Si calcoli $P(H_i \,|\, I_i)$ con $j = 0, \ldots 3$, ove I_0 è lo stato di informazione iniziale e I_1, I_2 e I_3 sono gli esiti delle estrazioni del problema 18 del capitolo 5. Si calcoli anche $P(B \,|\, I_i)$, ove B sta per la pallina bianca.
11. In una razza di cani il mantello nero (N) domina su quello rosso (n). Una cagna nera dal genotipo eterozigote (N, n) viene incrociata con un maschio omozigote recessivo, ovvero (n, n). Calcolare le probabilità che su 6 cuccioli ne nascano rispettivamente 0, 1, 2, 3, 4, 5 e 6 di colore rosso.
12. Un campione contiene 2.3×10^{24} nuclei radioattivi. La probabilità che uno dei nuclei possa nell'intervallo di tempo di un secondo vale 3.8×10^{-25}. Calcolare la probabilità di registrare, durante un secondo di osservazione, un numero di decadimenti pari a 0, 1, 2, 3, 4 e 5.
13. È noto che un certi fenomeni si verificano casualmente con una frequenza media di 50 l'ora. Calcolare la probabilità di osservarne almeno uno in un intervallo di osservazione di 30 secondi.
14. Sul problema precedente. Ammettiamo di ripetere per 10 volte le osservazioni da 30 secondi precedentemente descritte. Quanto vale la probabilità che in 5 delle 10 osservazioni non si verichi il fenomeno di interesse?
15. Uno sperimentatore ha buoni ragioni per ritenere che il numero di conteggi di un certo fenomeno segua la distribuzione di Poisson. Valutare la probabilità che si verifichino almeno 4 conteggi, sapendo che lo sperimenattore ritiene che la probabilità di non osservarne nessuno sia del 5 %.
16. Un campione di materia contiene 10^{22} nuclei. Un contatore di radioattività registra una frequenza di decadimenti di 100 al minuto. Quanto vale la probabilità che uno preciso nucleo decada in un intervallo di tempo di un secondo?
17. Una dattilografa scrive un resoconto costituito da 15 cartelle. Ciascuna cartella ha 20 linee di 60 caratteri ciascuna. La probabilità che la dattilografa sbagli un carattere è dello 0.1 %. Quanto vale la probabilità che in una cartella non ci sia nemmeno un errore? Quanto vale la probabilità che sull'intero resoconto ci siano al più 10 errori?

18. Una variabile casuale segue una distribuzione di Poisson. Sapendo che la probabilità che si verifichi un valore diverso da 0 vale $98.47\,\%$, determinare il coefficiente di variazione della distribuzione.

19. Una variabile casuale segue una distribuzione di Poisson. Sapendo che il valore medio della distribuzione è pari a due volte il valore della deviazione standard, trovare la probabilità che la variabile casuale assuma un valore più grande di 2.

20. Un certo giorno nel reparto di ostetria di un ospedale sono disponibili 10 posti letto. La media giornaliera del numero di donne che si presentano al reparto per partorire è pari a 6.5. Quanto vale la probabilità che una o più donne possa essere non accettata?

21. Fra le 16:00 e le 17:00 in un tratto di strada di una grande città transitano in media 20 taxi. Il $50\,\%$ è occupato e, dei restanti, il $70\,\%$ è prenotato. In base a queste informazioni:

    (a) Quanto vale la probabilità che un giorno passino al più 10 taxi in quell'ora? E la probabilità che ne passino esattamente 10?

    (b) Nel caso ne passino esattamente 10: quanto vale la probabilità che di essi almeno 2 siano liberi? E nel caso ne passano 20?

    (c) Quanto vale la probabilità che, uscendo dall'ufficio alle 16:30, si riesca a prendere un taxi al volo entro un quarto d'ora?

    (d) Quanto vale la probabilità che in una settimana lavorativa di 5 giorni si riesca almeno 2 volte a prendere un taxi entro un quarto d'ora?

22. Uno strumento registra dei di fenomeni casuali che seguono la distribuzione di Poisson e si verificano con una frequenza di 50 volte al secondo. Per quanto tempo bisognerà tenere lo strumento in funzione affinché l'incertezza di previsione relativa sul numero di conteggi che saranno registrati sia dell'$1\,\%$?

23. Riprendiamo il problema 9 del capitolo 2. Supponiamo di disporre di un dispositivo (tipicamente un programma di simulazione al computer) che permetta di distribuire in modo casuale dei punti all'interno del quadrato. Supponiamo di generare in totale 10000 punti. Trovare il valore atteso, la deviazione standard e il coefficiente di variazione del numero di punti che cade all'interno del quarto di cerchio. È possibile utilizzare questo dispositivo per stimare empiricamante il valore di π dalla frequenza osservata? Sapendo che $\pi \approx 3.14$, quanti punti bisogna generare affinché il coefficiente di variazione di $\widehat{\pi}$ ("valore misurato di π") sia pari a 0.0001?

24. Calcolare la previsione del numero di molecole in un millimetro cubico di gas perfetto a pressione atmosferica e ad una temperatura di 20°C.

25. Sui dati del problema precedente. Calcolare il lato del cubo tale che sia del $10\,\%$la probabilità che, ad un certo istante, non vi sia alcuna molecola al suo interno. Per avere un'idea di quanto poco compatta sia la materia in tali condizioni, stimare l'ordine di grandezza del numero di molecole che sarebbe possibile compattare all'interno di tale volumetto, assumendo per le molecole delle dimensioni lineari dell'ordine di 0.1 nm.

26. Un rivelatore risponde al passaggio di una particella emettendo in media 1000 fotoni. In media l'$1\,\%$ dei fotoni riesce ad arrivare su un *fotocatodo* dove ciascuno dei fotoni ha una probabilità del $36\,\%$ di emettere un elettrone. L'elettrone entra in un dispositivo opportuno (fotomoltiplicatore) che produce, con probabilità 1, un impulso elettrico di ampiezza maggiore di $0.1\,V$. Nel caso che, dal passaggio di una particella siano prodotti più impulsi, questi vengono sommati. Infine, un dispositivo elettronico fornisce un segnale percepibile dallo sperimentatore se esso riceve un impulso di almeno $0.05\,V$. Calcolare l'efficienza del rivelatore, ovvero la probabilità che lo sperimentatore registri un segnale al passaggio di una particella. (Assumere che fra l'arrivo di due particelle passi abbastanza tempo in modo tale che esse non diano luogo a segnali che si confonderebbero fra di loro.)

27. Una persona fa il seguente solitario: estrae a caso una carta da un mazzo da 52, la guarda e la rimette nel mazzo, rimischia e cosi' via. Vince quando ha trovato per 3 volte la regina dui cuori. Calcolare la previsione (con relativa incertezza) del numero di volte che deve provare.

28. In un paese il reddito medio per ogni persona è pari a 10000 \$ l'anno con una deviazione standard di 3000 \$. Qual'è la probabilità che scelta una persona a caso essa abbia un reddito superiore a 30000\$?

29. Con riferimento ai dati sperimentali dei morti da calcio di cavallo di tabella 7.6: scelto a caso un "ipotetico" reggimento e scelti a caso due anni, è giusto dire che la probabilità che ci siano almeno due morti nel periodo dei due anni sia il doppio della probabilità che ci sia almeno un morto in un

anno? La risposta dipende dal fatto che i due anni scelti siano consecutivi o no?

30. Sempre sugli stessi dati sperimentali del problema precedente: immaginiamo di estrarre a sorte uno dei 200 reggimenti che sono serviti a ricavare i dati di mortalità. Quanto vale la probabilità che uno dei reggimenti registri un numero di incidenti maggiore o uguale a 6?

31. Risolvere il problema 10 del capitolo 4 facendo uso della distribuzione ipergeometrica.

32. Un insegnante vuole "dimostrare" che, lanciando un grande numero di monete, la frequenza relativa di teste "è" del 50%. Quante monete deve far lanciare dagli studenti affinché la sua incertezza di previsione sia inferiore a 0.01?

Capitolo 8

Distribuzioni di probabilità di variabili continue

Avendo introdotto nei capitoli precedenti gli aspetti fondamentali delle distribuzione di variabili casuali, il concetto nuovo che viene presentato in questo capitolo è quello della funzione densità di probabilità. La distribuzione più importante presentata in questo capitolo è senz'altro la gaussiana, sia per le implicazioni teoriche che pratiche. La uniforme e le distribuzioni triangolari sono interessanti sia perché permettono di applicare concetti generali su funzioni matematiche elementari, sia perché rappresentano schemi abbastanza realistici per modellizzare l'incertezza su grandezze fisiche. Il processo di Poisson sotto l'aspetto dei tempi di attesa, da cui seguono esponenziale e gamma.

8.1 Variabili casuali continue e densità di probabilità

Nel capitolo precedente ci siamo occupati di variabili casuali che possono assumere soltanto un numero discreto di valori. Nella pratica scientifica è molto più frequente il caso di grandezze che, virtualmente, possono assumere valori con continuità in un certo intervallo. La precisazione "virtualmente" serve a ricordare che le variabili continue sono un concetto puramente matematico, astrazione di grandezze che possono assumere valori contigui la cui distanza è sufficientemente piccola rispetto alla scala tipica che descrive la loro variabilità (ad esempio una deviazione standard).

8.1.1 Probabilità nulle con diversi gradi di fiducia

Nel caso di variabili continue non ha senso parlare della probabilità di un determinato valore della variabile casuale. Infatti, essendoci, per ogni intervallo arbitrariamente scelto, un numero infinito di punti, la probabilità di ognuno di essi è nulla:

$$P(X = x_i) = 0\,.$$

Ma questo non significa che l'evento $X = x_i$ è impossibile, altrimenti sarebbe impossibile ottenere un risultato qualsiasi. Quindi, come già detto nel paragrafo 2.7, mentre l'evento impossibile implica che la sua probabilità sia nulla, non è vero il contrario.

L'osservazione interessante è che, sebbene tutti i possibili valori di X abbiano probabilità nulla, si può credere che alcuni di essi possano verificarsi più facilmente di altri. Si consideri il seguente *esperimento concettuale*. Si immagini di lasciar cadere una pallina "puntiforme" su un tavolo, da un'altezza di circa un metro. Piccole perturbazioni (vibrazioni della mano, urti con le molecole dell'aria e con il tavolo) fanno sì che il punto sul tavolo in cui la pallina si fermerà non sia univocamente determinato. Consideriamo la proiezione del punto di impatto lungo un arbitrario asse X giacente sul piano del tavolo. Per convenienza, scegliamo l'origine in corrispondenza della proiezione sul piano del tavolo del punto di rilascio e misuriamo la posizione in centimetri (scala naturale dell'esperimento). Ipotizziamo inoltre che la pallina si fermi dove cada, o al più possa fare dei piccoli rimbalzi ("piccoli" nella scala tipica dell'esperimento). Stanti queste ipotesi, non si può non convenire che

- la probabilità di ognuno dei possibili punti 'e uguale a zero (ad esempio $P(X = 0.000\ldots) = 0$, $P(X = \pi) = 0$, $P(X = 10^6) = 0$);
- ciò nonostante, si può crederà più a valori intorno a $X = 0$ che a valori intorno a $X = 10$ o a $X = 100$; in particolare si escluderanno completamente valori intorno a $X = 10^6$.

Anche in questo caso il diverso grado di fiducia verrà quantificato fa una funzione continua (o con al più un numero finito di punti di discontinuità) $f(x)$, che però, a differenza del caso discreto <u>non</u> ha il significato immediato di probabilità. Scrivere, ad esempio,

$$f(x_1) > f(x_2)$$

implica che il grado di fiducia in $X = x_1$ sia maggiore di quello in $X = x_2$. Scrivere invece

$$f(x) = 0$$

sta and indicare che l'evento è *impossibile*.

8.1.2 Dal grado di fiducia alla probabilità finita

È invece finita la probabilità $P(a \le X \le b)$ che la variabile sia compresa in un certo intervallo. Se la distanza fra i punti a e b diventa infinitesima anche la probabilità sarà infinitesima.

Se tutti i valori di X hanno lo stesso grado di fiducia (e - si noti bene! - non soltanto la stessa probabilità, $P(X = x) = 0$), ovvero

$$f(x) = k \qquad \forall\, x\,,$$

la probabilità è proporzionale all'ampiezza dell'intervallo e non dipende dal valore particolare di X:

$$\Delta P \propto \Delta x\,.$$

Quando l'intervallo diventa infinitesimo

$$\mathrm{d}P \propto \mathrm{d}x\,.$$

Nel caso generale ($f(x) \neq k$) e considerando due punti in corrispondenza dei quali $f(x)$ è continua (almeno da una parte) si ha che il rapporto fra le probabilità infinitesime intorno a tali punti è proporzionale ai loro gradi di fiducia:

$$\frac{\mathrm{d}P(x_1 \leq X \leq x_1 + \mathrm{d}x)}{\mathrm{d}P(x_2 \leq X \leq x_2 + \mathrm{d}x)} = \frac{f(x_1)}{f(x_2)}\,,$$

ovvero

$$\mathrm{d}P \propto f(x)\,\mathrm{d}x\,. \tag{8.1}$$

Essendo i diversi valori di X a due a due incompatibili, la probabilità su un intervallo finito è data dalla somma degli infiniti *elementi di probabilità* infinitesimi definiti dalla (8.1)

$$P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x\,, \tag{8.2}$$

avendo incluso il fattore di proporzionalità della (8.1) nella definizione di $f(x)$, ovvero l'elemento infinitesimo di probabilità è definito essere esattamente $f(x)\,\mathrm{d}x$:

$$\mathrm{d}P = f(x)\,\mathrm{d}x\,.$$

Avendo definito l'elemento infinitesimo di probabilità, si ricavano tutte le altre proprietà delle distribuzioni di variabili continue da quelle discrete mediante le seguenti sostituzioni:

$$\begin{array}{ccc} f(x_i) & \longrightarrow & f(x)\,\mathrm{d}x \\ \sum_i & \longrightarrow & \int_X\,. \end{array}$$

8.1.3 Funzione densità di probabilità

È immediato calcolare la probabilità cumulativa, descritta dalla funzione di ripartizione (chiamata talvolta semplicemente *integrale della funzione di probabilità* o *funzione integrale*):

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(z)\,\mathrm{d}z, \tag{8.3}$$

(analogamente al caso delle funzioni di distribuzione discrete, $f(x)$ vale 0 al di fuori del campo di definizione della variabile casuale).

Dalla (8.3) segue che

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}\,.$$

Siccome $F(x)$ <u>ha</u> il significato di probabilità, $f(x)$ è una "probabilità per unità di X", da cui il nome di *funzione densità di probabilità*, talvolta indicata con *p.d.f.* (*"probability density function"*). Inoltre si vede come l'elemento infinitesimo di probabilità, prima chiamato $\mathrm{d}P$, può essere anche indicato con $\mathrm{d}F(x) = f(x)\,\mathrm{d}x$.

Si noti anche come $f(x)$ non sia una grandezza adimensionale. Essendo $F(x)$ un numero puro, in quanto probabilità, ne segue che $f(x)$ ha le dimensioni inverse di quelle di X (ad esempio cm^{-1} nell'esempio precedente della caduta della pallina puntiforme).

8.1.4 Proprietà della funzione densità di probabilità e della funzione di ripartizione

Dalle regole di base della probabilità seguono le seguenti proprietà che la densità di probabilità e la funzione di ripartizione devono soddisfare (si veda anche il paragrafo 6.5):

- $f(x) \geq 0$, in quanto $f(x)\,\mathrm{d}x \geq 0$: la densità di probabilità è definita non negativa;
- $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$: la probabilità totale vale 1; si dice anche che $f(x)$ è *normalizzata* a 1;
- $0 \leq F(x) \leq 1$, dato il significato probabilistico che possiede; non c'è invece nessun vincolo al limite superiore di $f(x)$ purché essa assuma valori molto maggiori di 1 in intervalli molto piccoli, in modo tale che l'integrale nell'intervallo sia minore o uguale all'unità;
- $F(x)$ è una funzione non decrescente di x, essendo la sua derivata definita non negativa;
- $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to+\infty} F(x) = 1$;
- Il valore della probabilità in un intervallo è dato da:

$$
\begin{aligned}
P(a \leq X \leq b) &= \int_a^b f(x)\,\mathrm{d}x \\
&= \int_{-\infty}^b f(x)\,\mathrm{d}x - \int_{-\infty}^a f(x)\,\mathrm{d}x \\
&= F(b) - F(a);
\end{aligned}
$$

La Fig. 8.1 mostra la rappresentazione grafica di una generica distribuzione di variabile continua. Si noti l'interpretazione geometrica di probabilità come l'area sotto $f(x)$.

8.1.5 Valori attesi

Anche il calcolo del valore atteso di una funzione qualsiasi della variabile casuale si estende in modo naturale alle variabili casuali continue:

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

Ne segue che

$$
\begin{aligned}
\mu = \mathrm{E}(X) &= \int_{-\infty}^{+\infty} x\, f(x)\,\mathrm{d}x \\
\mathrm{Var}(X) &= \int_{-\infty}^{+\infty} (x-\mu)^2\, f(x)\,\mathrm{d}x
\end{aligned}
$$

Anche le proprietà degli operatori valore atteso e varianza, dimostrate sulle distribuzioni discrete (vedi paragrafi 6.9.2 e 6.12), sono valide anche per le distribuzioni continue.

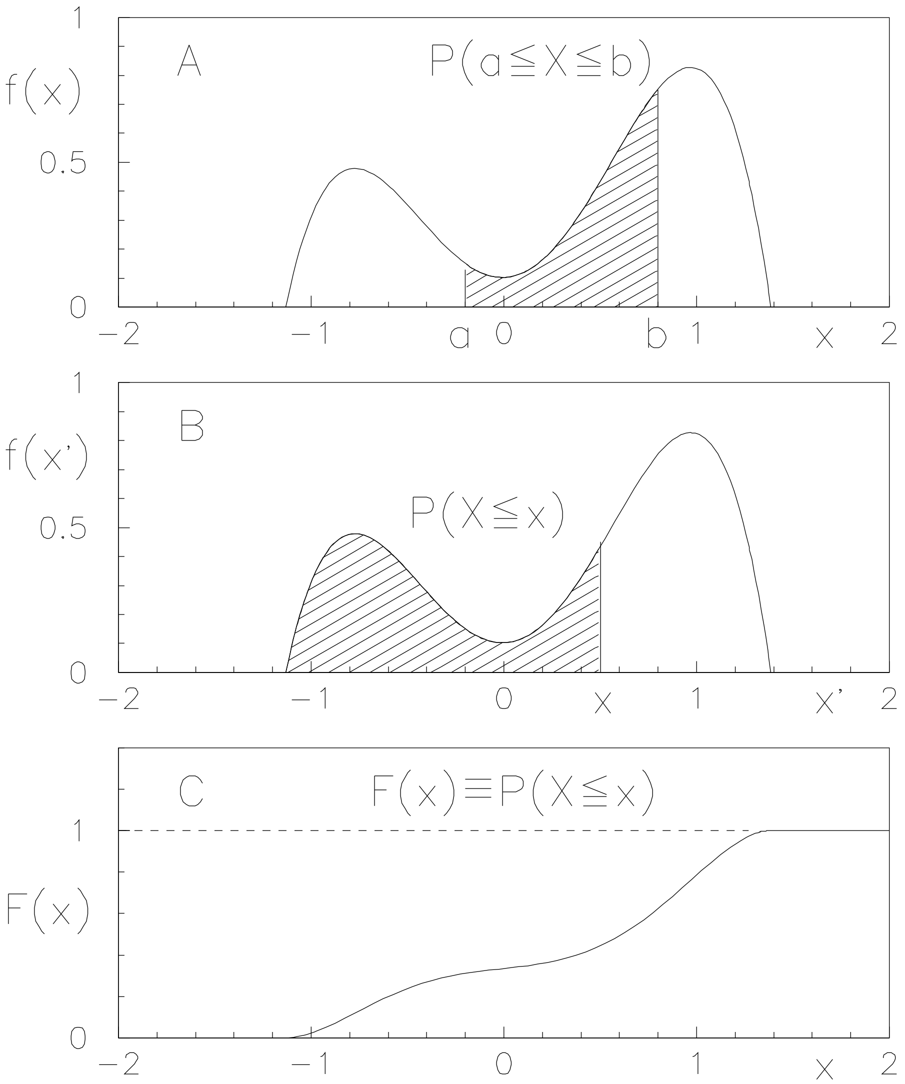


Figura 8.1: Distribuzioni di probabilità di variabili casuali continue. Le A e B illustrano l'interpretazione geometrica della probabilità. La C mostra la funzione di ripartizione relativa alla $f(x)$ delle A e B.

8.2 Distribuzione uniforme continua

La più semplice delle funzioni di distribuzione di probabilità di variabile continua è quella in cui si assegna lo stesso grado di fiducia a tutti i possibili valori di una variabile definita in un certo intervallo. Essa è detta *distribuzione uniforme*. Nell'esempio precedente della pallina puntiforme si può ottenere una distribuzione uniforme se si considerano i punti un piccolo intervallo finito (ad esempio da $a = -1\,\mu$m a $b = +1\,\mu$m), subordinatamente alla condizione che la pallina cada in quell'intervallo. Avendo tutti i punti lo stesso grado di fiducia, segue che

$$f(x) = k \qquad a \le x \le b$$

Il valore della costante k viene ricavato dalla condizione di normalizzazione:

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x = k\,(b-a) = 1.$$

Quindi:

$$f(x\,|\,\mathcal{K}(a,b)) \;=\; \frac{1}{b-a} \qquad \begin{cases} a \le x \le b \\ a < b \end{cases} \tag{8.4}$$

Si noti come i parametri di $\mathcal{K}$ siano, diversamente da quanto avveniva nel caso di variabile discreta, fra parentesi invece che a pedice. Questo è semplicemente un modo convenzionale per indicare che la distribuzione è a variabili continue.

Essendo la funzione densità di probabilità costante, la funzione di ripartizione è lineare nell'intervallo $[a,b]$, come si verifica facilmente:

$$F(x\,|\,\mathcal{K}(a,b)) \;=\; \int_{-\infty}^{x} f(z)\,\mathrm{d}z = \frac{1}{b-a}\int_a^x \mathrm{d}z = \frac{x-a}{b-a}\,.$$

La figura 8.2 mostra la funzione di densità di probabilità e la funzione di ripartizione della distribuzione uniforme.

Calcoliamo il valore valore atteso e la varianza di X (previo il calcolo di $\mathrm{E}(X^2)$):

$$\begin{aligned} \mu = \mathrm{E}(X) &= \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \frac{a+b}{2} \\ \mathrm{E}(X^2) &= \int_a^b \frac{x^2}{b-a}\,\mathrm{d}x = \frac{1}{3}\,\frac{b^3-a^3}{b-a} = \frac{1}{3}\,(b^2+a\,b+a^2)\,. \end{aligned}$$

Quindi:

$$\sigma^2 = \frac{1}{3}\,(b^2+a\,b+a^2) - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}\,,$$

da cui:

$$\sigma = \frac{b-a}{\sqrt{12}}\,. \tag{8.5}$$

Come ci si poteva attendere, la previsione della variabile casuale coincide con il centro dell'intervallo. La deviazione standard è pari a circa il 30 % dell'intervallo.

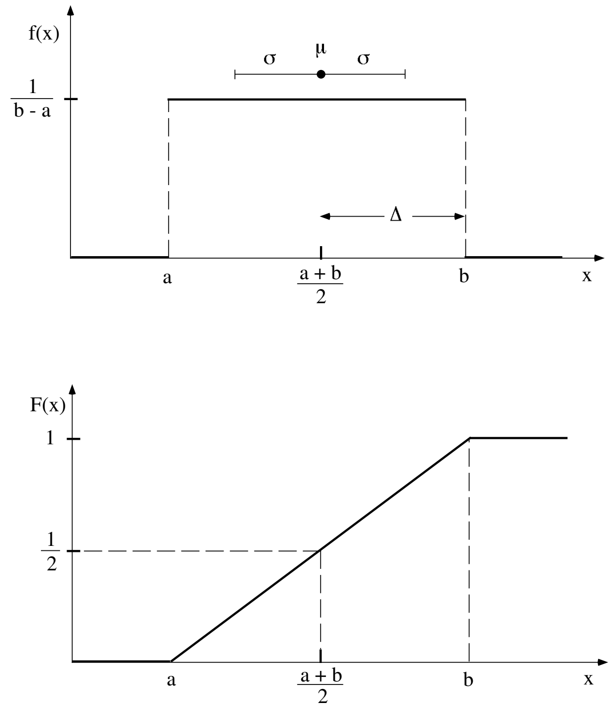


Figura 8.2: Distribuzione uniforme

La distribuzione uniforme continua può essere pensata come il limite della distribuzione discreta (vedi paragrafo 6.6.2) quando il numero di punti tende a infinito e la spaziatura a zero. Si riottengono infatti gli stessi valori di media e deviazione standard.

La distribuzione uniforme viene impiegata nella trattazione degli errori di misura ogni qual volta si sa con sicurezza che una certa variabile è contenuta in un certo intervallo, ma non si ha alcun motivo per ritenere alcuni valori più plausibili di altri. Per queste applicazioni è conveniente chiamare la larghezza dell'intervallo $b - a = 2\,\Delta$, in quanto spesso si usa dire che la variabile può verificarsi uniformemente nell'intervallo $x_c \pm \Delta$, dove x_c sta per il centro dell'intervallo. Ne segue che

$$\sigma = \frac{\Delta}{\sqrt{3}} = 0.58\,\Delta \approx 0.6\,\Delta\,. \tag{8.6}$$

Un altro uso della uniforme continua è nelle *simulazioni* al calcolatore. Infatti, mediante opportune tecniche è possibile, a partire da una variabile distribuita uniformemente, costruire altre variabili distribuite a piacere. Il prossimo paragrafo è da considerare come una breve e interessante parentesi su tale argomento.

8.3 ❄ Simulazione al computer di processi stocastici

La distribuzione uniforme è molto utilizzata e qualsiasi calcolatore è in grado di produrre ("generare") un numero aleatorio con probabilità uniforme nell'intervallo $[0, 1]$ (anche nei calcolatorini c'è in genere un tasto "RAN#" che fornisce un numero a caso in tale intervallo). Tali numeri sono generalmente chiamati *numeri casuali*, o *pseudocasuali*, ad indicare che in realtà si tratta di sequenze generate con algoritmi numerici e riproducibili, ma che si comportano in modo indistinguibile da numeri casuali che abbiano le proprietà richieste. Inoltre, essendo finito il numero di cifre dei computer, questi numeri sono a rigore numeri aleatori discreti (con 10^n possibilità, dove n è il numero di cifre decimali) ma dal punto di vista pratico possono essere considerati a tutti gli effetti continui.

8.3.1 Costruzioni di altre semplici variabili casuali

Mostriamo delle semplici tecniche per generare numeri aleatori che seguano una certa distribuzione. Chi ha un minimo di familiarità con linguaggi di progammazione può utilizzare questi suggerimenti per sviluppare programmi di simulazione le cui applicazioni e il cui grado di complicazione dipendono soltanto dalla fantasia dell'interessato. Indichiamo con R il numero casuale di partenza, uniformemente distribuito fra 0 e 1.

Generica distribuzione uniforme fra a e b

Si moltiplica R per l'ampiezza dell'intervallo e al risultato si aggiunge l'estremo inferiore:

$$X = a + (b - a)\,R$$

Processo di Bernoulli e distribuzione binomiale

Si scelga un valore p di R. Si dichiara l'evento favorevole se $R \leq p$. Ripetendo n volte la procedura e contando il numero di successi si ottiene il numero aleatorio di interesse (il metodo non eccelle in efficienza e, per n molto grandi, chiaramente bisogna cambiare tecnica).

Distribuzione uniforme discreta

Definendo n intervallini di uguale ampiezza nell'intervallo fra 0 e 1, il numero aleatorio discreto assume il valore i se R capita nell'i-mo intervallino (in pratica si prende la parte intera di $n \times R$ aumentata di 1). Con questa tecnica si simulano facilmente lanci di dadi ed estrazioni del lotto.

Marcia a caso

Se $R \geq 0.5$ si va a destra (incremento della posizione di $+1$), altrimenti si va a sinistra (-1). Si noti come la asimmetria destra/sinistra indotta dal "$\geq$" sia in genere trascurabile (ad essere proprio pignoli si può invalidare l'estrazione nel caso di $R = 0.5\ldots$).

8.3.2 Scelta pesata con $f(x)$

Vediamo ora un metodo più generale per generare numeri casuali discreti o continui che seguono una distribuzione qualsiasi, purché definita in un intervallo finito:

1. Si estraggono due numeri casuali R_1 e R_2;

2. da R_1 si costruisce il valore della variabile di interesse nell'intervallo in cui essa è definita (vedi sopra il caso di generica distribuzione uniforme continua o discreta): sia esso x_1;

3. da R_2 si costruisce un valore f_R distribuito uniformemente fra 0 e il massimo di $f(x)$ (o un valore ad esso superiore):

   - se $f_R \leq f(x_1)$ si considera $X = x_1$ come valore estratto;
   - altrimenti si annulla l'estrazione;

   in entrambi i casi si ricomincia dal punto 1.

È facile convincersi che la probabilità che un valore di X sia accettato è proporzionale a $f(x)$, sia che questa abbia il significato di funzione probabilità (variabile discreta) che di funzione densità di probabilità (variabile continua). Sebbene questo metodi sia uno dei più facili da implementare può avere seri problemi di efficienza (nel senso che in troppi casi l'estrazione va a monte) se la $f(x)$ assume alti valori in un piccolo intervallo e bassi valori altrove. Questo metodo è chiamato *"hit or miss"* (inglese per *colpito o mancato*) per il modo peculiare di estrazione.

8.3.3 Scelta uniforme lungo $F(x)$

Uno dei metodi più universali di estrazione, valido anche per variabili casuali definite in un dominio infinito, fa uso della funzione di ripartizione.

Cominciamo con variabili discrete, ricordandoci della rappresentazione grafica "a gradini" della $F(x)$ (vedi figura 6.3). Per ogni valore x_i in cui X è definita, il "salto" da un gradino all'altro è pari a $f(x_i)$. Se si estrae un numero casuale R fra 0 e 1 e lo si posiziona lungo l'asse delle ordinate del grafico di $F(x)$ si ha che

- la probabilità che R cada nel salto i-mo è proporzionale all'altezza del salto stesso, ovvero a $f(x_i)$;

- se il valore R è compreso nell'intervallo

  $$F(x_{i-1}) < R \leq F(x_i)$$

  il gradino è quello in corrisponenza del valore $X = x_i$ (si ricordi infatti che $F(x_i) - F(X_{i-1} = f(x_i)\,)$.

Quindi questa procedura permette di ottenere valori di X la cui probabilità di verificarsi è data dalla distribuzione di interesse.

Per passare alle variabili continue basta pensare al limite di una funzione a gradini, con infiniti gradini talmente ravvicinati e di salto infinitesimo $\mathrm{d}F =$

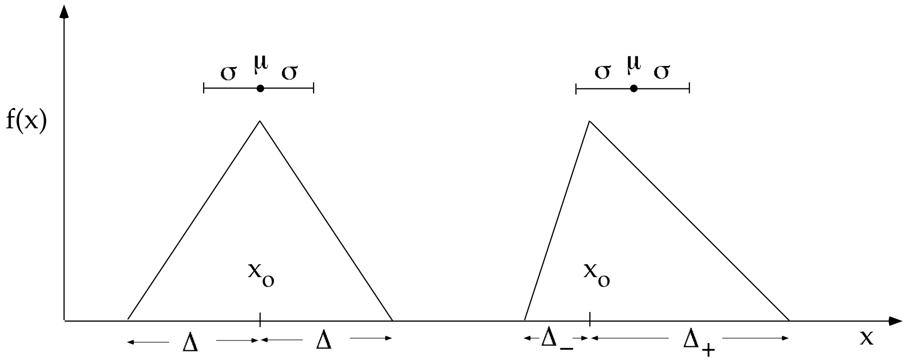


Figura 8.3: Esempio di distribuzione triangolare simmetrica e asimmetrica.

$f(x)\,\mathrm{d}x$. Ancora una volta si vede come il grado di fiducia con cui i possibili valori di X possono essere sorteggiati è proporzionale a $f(x)$. Così pure, in analogia al caso discreto, si può dire che, se

$$F(x_R) \leq R < F(x_R) + \mathrm{d}F\,,$$

ovvero se

$$F(x_R) = R\,,$$

il gradino infinitesimo è quello in corrispondenza di x_R. Questo metodo, applicato alle variabili continue la cui funzione densità di probabilità è integrabile analiticamente e la $F(x)$ facilmente invertibile, è senz'altro il più rapido, in quanto:

$$x_R = F^{-1}(R)\,,$$

ove $F^{-1}(\cdot)$ sta per l'inversa di $F(\cdot)$.

8.4 ○ Distribuzioni triangolari

Quando la variabile casuale è definita in un certo intervallo, ma ci sono delle ragioni per ritenere che i gradi di fiducia decrescano linearmente dal centro ($x_\circ$) verso gli estremi si ha la cosiddetta *distribuzione triangolare* (o di *Simpson*). Anch'essa è molto utile per il calcolo delle incertezza di misura, in quanto in molte circostanze, questo tipo di modello può essere più realistico di quello uniforme.

Se la variabile può verificarsi con sicurezza nell'intervallo $x_\circ \pm \Delta$, il valore atteso è $x_\circ$ e la deviazione standard vale

$$\sigma = \frac{\Delta}{\sqrt{6}} = 0.41\,\Delta \approx 0.4\,\Delta\,. \tag{8.7}$$

I conti vengono lasciati per esercizio.

C'è un'altra distribuzione triangolare che può avere interessi pratici: il valore al quale si crede di più è $x_\circ$ e i gradi di fiducia decrescono linearmente verso gli estremi a e b, ma $x_\circ$ non corrisponde con il centro dell'intervallo. Chiamando

$$\begin{aligned} \Delta_+ &= b - x_\circ & (8.8)\\ \Delta_- &= x_\circ - a\,, & (8.9) \end{aligned}$$

si ottengono i seguenti valori di valore atteso e varianza

$$\begin{aligned} \mathrm{E}(X) &= x_\circ + \frac{\Delta_+ - \Delta_-}{3}\,. & (8.10)\\ \sigma^2 &= \frac{\Delta_+^2 + \Delta_-^2 + \Delta_+\,\Delta_-}{18}\,, & (8.11) \end{aligned}$$

le cui dimostrazioni vengono lasciate come esercizio.

Quando $\Delta_+ = \Delta_- = \Delta$, si riottengono le formule del caso precedente. È interessante inoltre notare che, se la differenza fra Δ_+ e Δ_- è piccola si ottiene una deviazione standard circa pari a quella ottenibile con un valore intermedio fra i due:

$$\sigma \approx \frac{\overline{\Delta}}{\sqrt{6}} = \frac{(\Delta_+ + \Delta_-)/2}{\sqrt{6}}\,,$$

come può essere verificato mediante una espansione in serie della (8.11).

Un sottocaso particolare della triangolare asimmetrica (non isoscele) è quando uno dei due Δ è nullo ed il triangolo diventa rettangolo. Questa distribuzione può modellizzare gradi di fiducia che decrescono linearmente in un certo intervallo. Ad esempio, ci possono essere delle ragioni per ritenere che una grandezza definita non negativa valga molto verosimilmente 0 e che comunque non debba eccedere un certo valore a: si ottiene una previsione di $a/3 \pm a/(3\sqrt{2})$.

8.5 Distribuzione esponenziale

Abbiamo visto finora variabili continue definite in un intervallo finito. Quando l'intervallo diventa infinito, la funzione densità di probabilità deve avere un andamento opportunamente decrescente in modo tale che il suo integrale sia uguale a 1. Un caso molto interessante di distribuzione è quando si ha una grandezza definita fra 0 e infinito, con i gradi di fiducia che decrescono esponenzialmente con il suo valore:

$$f(x) \propto e^{-\alpha\, x} \qquad \left\{ \begin{array}{l} 0 \le x < \infty \\ \alpha > 0 \end{array} \right.$$

Il fattore di proporzionalità si ricava dalla condizione di normalizzazione, da cui:

$$f(x) = \alpha\, e^{-\alpha\, x}\,.$$

Per ragioni di convenienza legate alle applicazioni che incontreremo, consideriamo una variabile casuale che abbia il significato di tempo T. È anche opportuno utilizzare un parametro omogeneo con T indicato con τ $(= 1/\alpha)$.

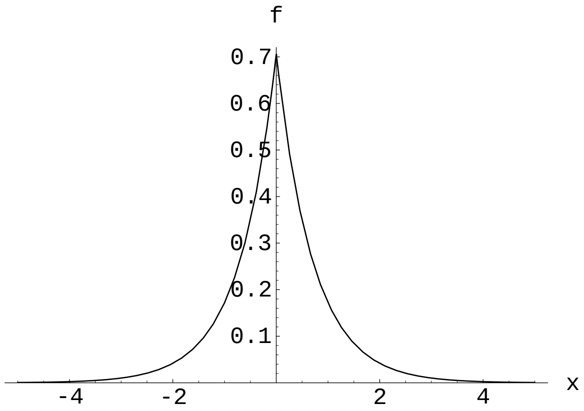


Figura 8.4: Distribuzione esponenziale doppia (o di Laplace) avente valore atteso nullo e deviazione standard unitaria ($\beta = 1/\sqrt{2}$).

Si ottiene allora la seguente espressione per la *distribuzione esponenziale* nel dominio del tempo:

$$f(t\,|\,\mathcal{E}(\tau)) = \frac{1}{\tau}\,e^{-t/\tau} \qquad \left\{\begin{array}{l} 0 \le t < \infty \\ \tau > 0 \end{array}\right. \quad (0 \le t < \infty) \tag{8.12}$$

$$F(t\,|\,\mathcal{E}(\tau)) = 1 - e^{-t/\tau} \qquad (0 \le t < \infty) \tag{8.13}$$

Calcoliamoci il valore atteso e deviazione standard di T:

$$\begin{aligned}
\mathrm{E}(T) &= \int_0^\infty \frac{t}{\tau}\,e^{-t/\tau}\mathrm{d}t \\
&= -t\,e^{-t/\tau}\Big|_0^\infty + \int_0^\infty e^{-t/\tau}\mathrm{d}t \\
&= \tau \\
\mathrm{E}(T^2) &= \int_0^\infty \frac{t^2}{\tau}\,e^{-t/\tau}\mathrm{d}t \\
&= -t^2\,e^{-t/\tau}\Big|_0^\infty + 2\int_0^\infty t\,e^{-t/\tau}\mathrm{d}t \\
&= 2\,\tau^2 \\
\sigma(T) &= \tau
\end{aligned}$$

La distribuzione esponenziale può essere atta a descrivere una situazione di incertezza in cui anche valori molto grandi della grandezza possono essere ammissibili, ma con gradi di fiducia tali che dopo alcune deviazioni standard si è “praticamente certi” che essi non si verifichino. Un caso interessantissimo in cui essa entra in gioco è nei tempi di attesa di conteggi in fenomeni descritti da processi di Poisson (vedi paragrafo 8.12 nel prossimo paragrafo.

8.6 ❄ Distribuzione esponenziale doppia

La distribuzione esponenziale può essere estesa per simmetria intorno al valore di massimo grado di fiducia x_m, ottenendo una funzione densità di probabilità del tipo

$$f(x) \propto e^{-\alpha|x-x_m|}\,,$$

con α definito positivo. Normalizzando opportunamente la funzione e facendo comparire il parametro β omogeneo a X, si ottiene la *distribuzione esponenziale doppia* o *di Laplace*:

$$f(x) = \frac{1}{2\beta} e^{-\frac{|x-x_m|}{\beta}}\,, \tag{8.14}$$

avente

$$\begin{aligned} \mathrm{E}(X) &= x_m \\ \sigma &= \sqrt{2}\,\beta\,. \end{aligned}$$

La figura 8.4 mostra un esempio di tale distribuzione. Sebbene questa distribuzione sembri formalmente atta a descrivere variabili casuali che possono assumere, sebbene con probabilità piccolissime, grandissimi scarti dal valor medio, è in realtà di scarsa utilità per le applicazioni.

8.7 Distribuzione normale

Una distribuzione, in principio simile alla esponenziale doppia per quanto riguarda la simmetria rispetto al valore centrale e l'estensione a grandissimi scarti, ma che meglio si presta a descrivere moltissimi casi di interesse è quella in cui i gradi di fiducia vanno come

$$f(x) \propto e^{-h\,(x-x_m)^2}\,,$$

ove h è una costante positiva. Questa funzione, opportunamente normalizzata, è nota come *funzione di Gauss*, o *gaussiana*. Essa deve il nome a Karl Friederick Gauss, che la propose per la descrizione delle deviazioni delle misure astronomiche rispetto al loro andamento medio. Egli ipotizzò infatti che tali deviazioni fossero dovute ad *errori casuali di misura* e, in base ad argomenti abbastanza generali, derivò una funzione densità di probabilità del tipo appena mostrato (vedi paragrafo 11.4. Stanti i forti argomenti teorici per ritenere che gli errori casuali debbano seguire tale distribuzione (vedi paragrafo 10.15 e 11.2) e la effettiva compatibilità dei dati sperimentali con tale ipotesi, viene comunemente detto che gli errori casuali "seguono normalmente" tale distribuzione e la distribuzione stessa è perciò chiamata anche *distribuzione normale*.

Imponendo la condizione di normalizzazione e ridefinendo opportunamente i parametri in modo tale da far apparire esplicitamente valore atteso e deviazione standard della distribuzione otteniamo la forma nella quale essa è comunemente conosciuta:

$$f(x\,|\,\mathcal{N}(\mu,\sigma)) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \begin{cases} -\infty < x < +\infty \\ -\infty < \mu < +\infty \\ 0 < \sigma < \infty \end{cases} \tag{8.15}$$

Quindi (anche se non lo dimostriamo):

$$\begin{aligned}\mathrm{E}(X) &= \mu \\ \mathrm{Var}(X) &= \sigma^2 .\end{aligned}$$

Questa distribuzione ricopre un ruolo notevole non soltanto per la descrizione degli errori casuali, ma anche perchè essa risulta essere la distribuzione a cui tendono, sotto condizioni generali che descriveremo, molte altre distribuzioni, comprese la binomiale e la poissoniana.

Elenchiamo le sue proprietà principali:

- come detto, ha due parametri, μ e σ, che corrispondono al valore atteso e alla deviazione standard[1];
- presenta una tipica forma a campana;
- è simmetrica intorno alla media, ovvero $f(\mu + x) = f(\mu - x)$; la media coincide con il massimo della distribuzione (ricordiamo, moda) e con il punto che i valori della variabile casuale in due regioni di uguale probabilità (ricordiamo, mediana).
- il valore massimo della distribuzione (nel punto $X = \mu$) è inversamente proporzionale alla deviazione standard[2]:

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma} ;$$

- presenta due punti di flesso in corrispondenza di $\mu - \sigma$ e $\mu + \sigma$;
- è definita per tutti i valori di X, ma assume valori trascurabili per valori distanti più di “alcune” σ dalla media.

La figura 8.5a mostra degli esempi di distribuzione normale, per alcuni valori di μ e di σ. Nella figura 8.5b è anche riportata, per ciascuna $f(x)$, la relativa funzione di ripartizione[3] $F(x)$, la cui espressione matematica è data, per definizione::

$$F(x \,|\, \mathcal{N}(\mu, \sigma)) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \mathrm{d}z \qquad (8.16)$$

Purtroppo l’integrale non ha una forma semplice. Vedremo nel prossimo paragrafo come valutarla mediante opportune tabelle. Anche senza l’espressione analitica, possiamo elencare alcune proprietà della funzione di ripartizione, ottenibili direttamente da quelle della $f(x)$:

[1] I matematici preferiscono considerare come secondo parametro la varianza invece della deviazione standard. Anche se in principio la scelta è equivalente, bisogna fare attenzione ad interpretare correttamente notazioni sintetiche del tipo $\mathcal{N})50, 100)$. In questo testo utilizziamo la deviazione standard (e quindi la notazione precedente stava a significare $\sigma = 100$) in quanto omogenea alla grandezza di interessa e legata al concetto di incertezza standard di previsione.

[2] Come giustificazione intuitiva si pensi che: dovendo essere costante l’area sotto la curva, al diminuire di σ deve aumentare il massimo; $f(x)$ ha le dimensioni inverse di X e quindi il suo denominatore deve dipendere linearmente da σ.

[3] A volte la funzione di ripartizione della normale è indicata con $\Phi(x)$, ovvero:

$$\Phi(x) = F(x \,|\, \mathcal{N})$$

.

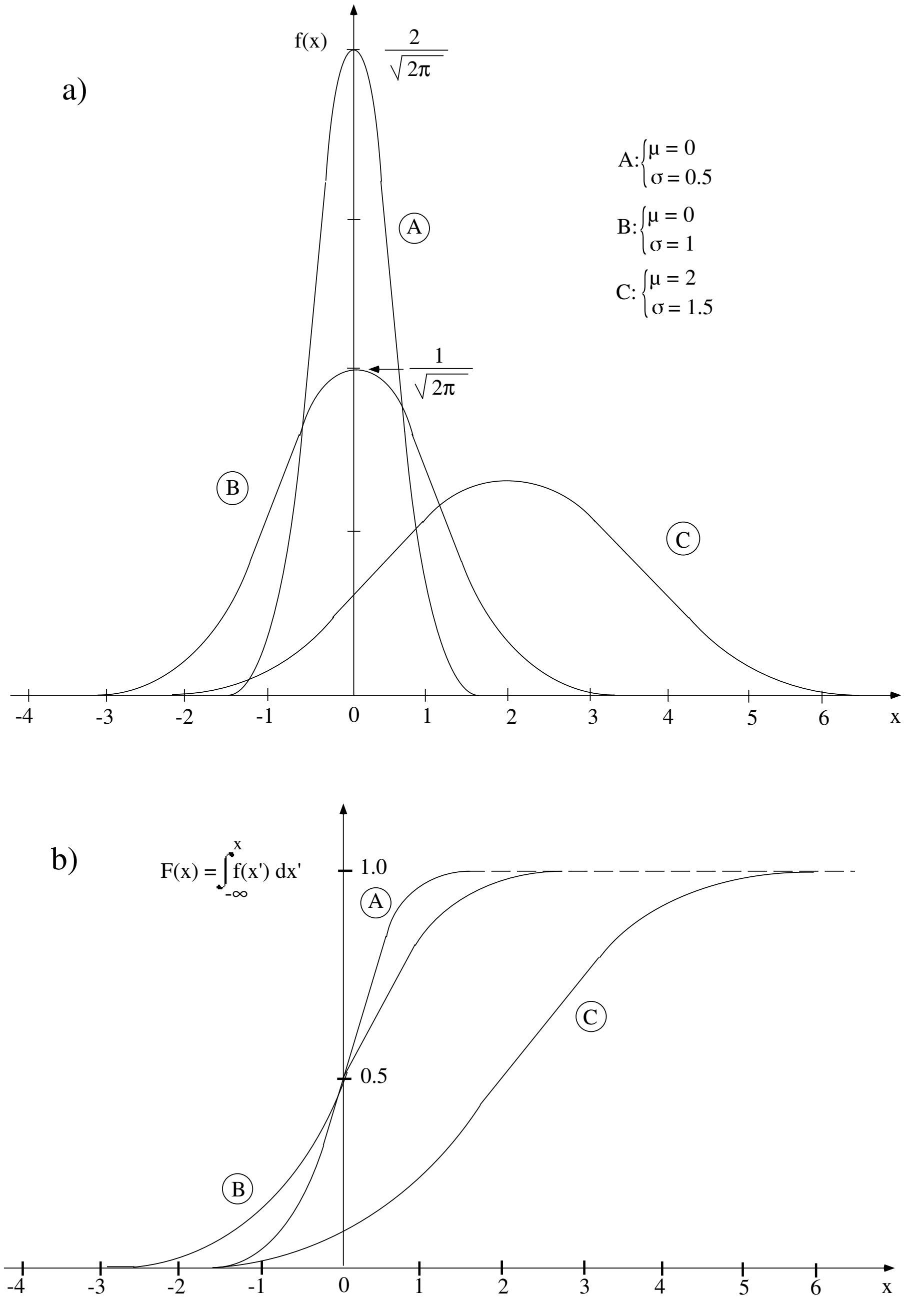


Figura 8.5: Esempi di distribuzione normale.

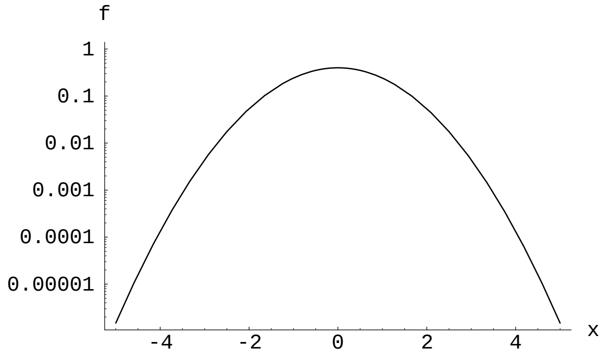


Figura 8.6: Rappresentazione su scala delle ordinate logaritmica della distribuzione normale standardizzata ($\mathcal{N}(0,1)$).

- ha una forma di "s allungata" (*sigmoide*) con asintoti orizzontali per $x \to \pm\infty$;
- assume il valore 0.5 per $x = \mu$, in quanto $F(\mu) = P(X \leq \mu) = 0.5$;
- ha un andamento pressoché lineare intorno a μ, con pendenza inversamente proporzionale a σ;
- ha un punto di flesso, ovvero cambia curvatura, in corrispondenza di μ.

Per fare apprezzare meglio gli andamenti delle code della distribuzione la figura 8.6 mostra su *su scala logaritmica* la distribuzione normale avente $\mu = 0$ e $\sigma = 1$.

8.8 Distribuzione normale standardizzata

Essendo la distribuzione normale largamente utilizzata e non avendo il suo integrale indefinito una forma analitica, per il calcolo delle probabilità vengono usati valori tabulati. Si capisce bene che sarebbe impossibile avere tabelle per tutte le possibili coppie di valori dei parametri. È quindi conveniente rendere il calcolo della funzione di ripartizione $F(x)$ indipendente dai parametri. Eseguendo la seguente trasformazione di variabili:

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

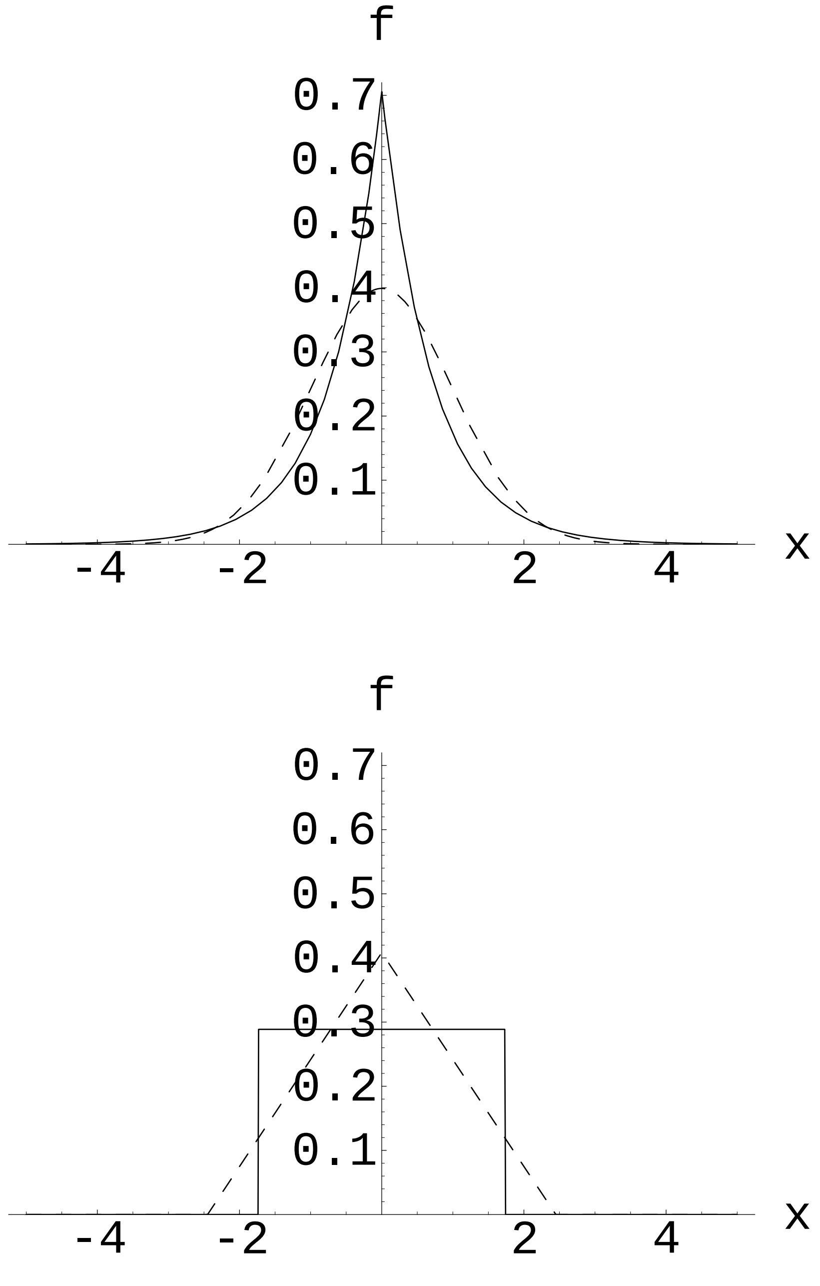


Figura 8.7: Distribuzioni di Laplace, di Gauss, triangolare e uniforme aventi $\mathrm{E}(X) = 0$ e $\sigma(X) = 1$.

otteniamo quindi che la probabilità che la variabile sia compresa fra a e b vale

$$\begin{aligned}
P(a \leq X \leq b) &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x'-\mu)^2}{2\sigma^2}} \mathrm{d}x' \\
&= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \mathrm{d}z \\
&= \int_{z_a}^{z_b} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \mathrm{d}z \\
&= P(z_a \leq Z \leq z_b) \\
&= P(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma})
\end{aligned}$$

La variabile Z è chiamata variabile *normale standardizzata* e la funzione di probabilità

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{8.17}$$

è detta *distribuzione normale standardizzata*. La variabile Z corrisponde ad una "trasformazione di coordinate lungo l'asse X tale da misurare la variabile in unità di σ a partire dal punto $X = \mu$" (vedi figura 8.9). Come si vede facilmente, la distribuzione normale standardizzata è una particolare normale di valor medio nullo e varianza unitaria:

$$Z \sim \mathcal{N}(0, 1) \tag{8.18}$$

8.9 Uso delle tabelle dell'integrale della distribuzione normale standardizzata

Esistono tabelle dell'integrale della (8.17) espresso in genere come [4]

$$T(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{z'^2}{2}} \mathrm{d}z' \,. \tag{8.19}$$

Da queste tabelle è possibile calcolare qualsiasi altro integrale facendo uso delle proprietà di simmetria della funzione e dei valori notevoli di $F(x)$. Un esempio è riportato in tabella 8.1. Essa va letta nel seguente modo:

- il valore di z, fino alla prima cifra decimale, è riportato nella prima colonna; la seconda cifra decimale è indicata nella prima riga delle altre colonne e, in corrispondenza di essa, è riportato il valore dell'integrale (8.19), sottindendendo "0.";

[4] Questo integrale è legato alla funzione matematica "erf(z)", il cui nome ricorda "error function", definita come

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} \mathrm{d}t \,,$$

dalla relazione

$$T(z) = \frac{1}{2} \mathrm{erf}\left(\frac{z}{\sqrt{2}}\right) \qquad (z \geq 0) \,.$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0.1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0.2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0.3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0.4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0.5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0.6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0.7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0.8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0.9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1.0	34134	34375	34614	34849	35083	35314	35543	35769	35993	36214
1.1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1.2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1.3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1.4	41924	42073	42220	42364	42507	42647	42785	42922	43056	43189
1.5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1.6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1.7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1.8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1.9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2.0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2.1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2.2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2.3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2.4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2.5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2.6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2.7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2.8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2.9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3.0	135-02	131-02	126-02	122-02	118-02	114-02	111-02	107-02	104-02	100-02
3.1	968-03	935-03	904-03	874-03	845-03	816-03	789-03	762-03	736-03	711-03
3.2	687-03	664-03	641-03	619-03	598-03	577-03	557-03	538-03	519-03	501-03
3.3	483-03	466-03	450-03	434-03	419-03	404-03	390-03	376-03	362-03	349-03
3.4	337-03	325-03	313-03	302-03	291-03	280-03	270-03	260-03	251-03	242-03
3.5	233-03	224-03	216-03	208-03	200-03	193-03	185-03	178-03	172-03	165-03
3.6	159-03	153-03	147-03	142-03	136-03	131-03	126-03	121-03	117-03	112-03
3.7	108-03	104-03	996-04	957-04	920-04	884-04	850-04	816-04	784-04	753-04
3.8	723-04	695-04	667-04	641-04	615-04	591-04	567-04	544-04	522-04	501-04
3.9	481-04	461-04	443-04	425-04	407-04	391-04	375-04	359-04	345-04	330-04
4.0	317-04	304-04	291-04	279-04	267-04	256-04	245-04	235-04	225-04	216-04
4.1	207-04	198-04	189-04	181-04	174-04	166-04	159-04	152-04	146-04	139-04
4.2	133-04	128-04	122-04	117-04	112-04	107-04	102-04	977-05	934-05	893-05
4.3	854-05	816-05	780-05	746-05	712-05	681-05	650-05	621-05	593-05	567-05
4.4	541-05	517-05	494-05	471-05	450-05	429-05	410-05	391-05	373-05	356-05
4.5	340-05	324-05	309-05	295-05	281-05	268-05	256-05	244-05	232-05	222-05
4.6	211-05	201-05	192-05	183-05	174-05	166-05	158-05	151-05	143-05	137-05
4.7	130-05	124-05	118-05	112-05	107-05	102-05	968-06	921-06	876-06	834-06
4.8	793-06	755-06	718-06	683-06	649-06	617-06	587-06	558-06	530-06	504-06
4.9	479-06	455-06	433-06	411-06	391-06	371-06	352-06	335-06	318-06	302-06

Tabella 8.1: Tabella per il calcolo della funzione cumulativa della distribuzione della normale.

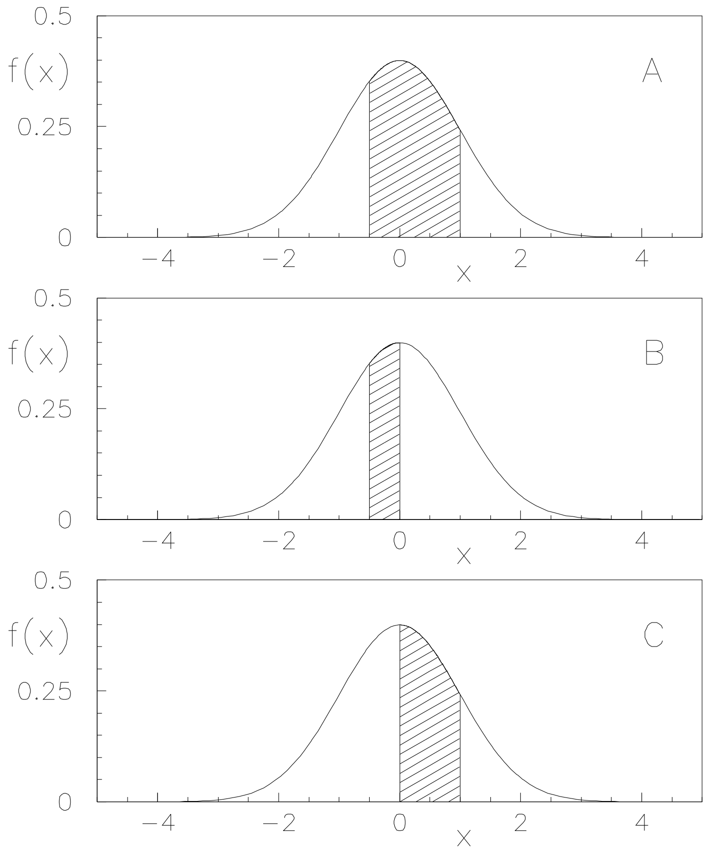


Figura 8.8: Esempio di calcolo dell'integrale della funzione normale standardizzata. L'integrale della figura A è pari alla somma di quelli di B e C, leggibili dalle tabelle.

- per i valori di z superiore a 3 è riportato il complemento a 0.5 dell'integrale (8.19) nella notazione $.xxx \cdot 10^{-yy}$, ove xxx e yy sono separate dal segno "$-$";

- esempi:

$$\begin{aligned} P(0 \le Z \le 1.25) &= 0.39435; \\ P(0 \le Z \le 4.33) &= 0.5 - 0.746 \cdot 10^{-5} = 0.49999254. \end{aligned}$$

La simmetria della distribuzione normale permette di valutare dalle stesse tavole anche l'integrale su un intervallo qualsiasi. Facciamo alcuni esempi di integrali calcolati fra z_1 e z_2, con $z_1 < z_2$:

- z_1 e z_2 positivi:

$$P(z_1 \le Z \le z_2) = T(z_2) - T(z_1);$$

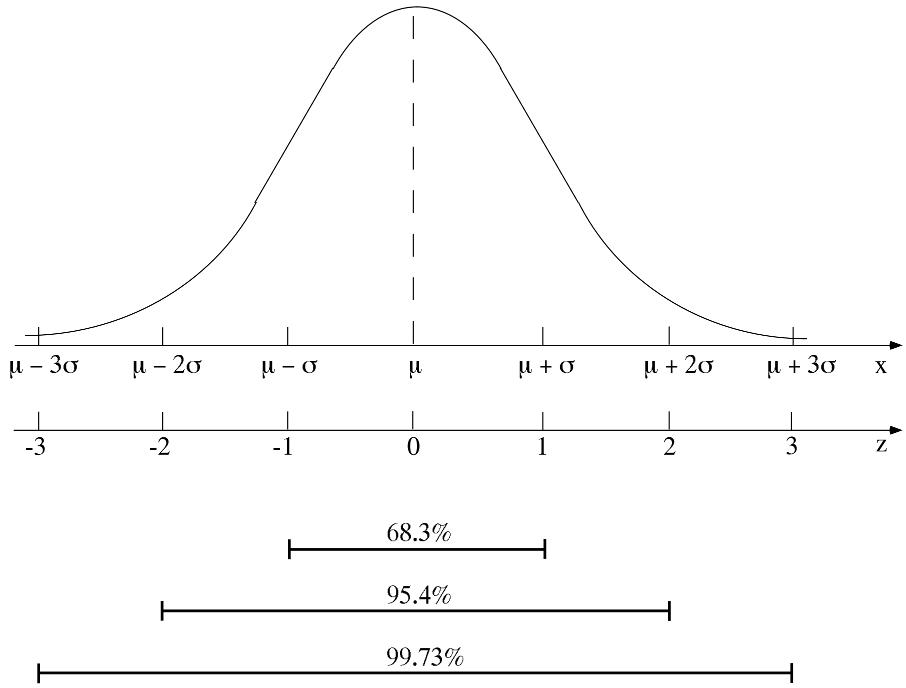


Figura 8.9: Distribuzione normale standardizzata e intervalli di probabilità.

- z_1 e z_2 negativi:

$$P(z_1 \leq Z \leq z_2) = T(-z_2) - T(-z_1);$$

- z_1 negativo e z_2 positivo:

$$P(z_1 \leq Z \leq z_2) = T(z_2) + T(-z_1)$$

Alcuni esempi sono mostrati in figura 8.8. Riportiamo la soluzione numerica ottenute mediante la tabella 8.1:

$$P(-0.5 \leq Z \leq 1.0) = 0.19146 + 0.34134 = 0.53280\,.$$

Terminiamo con un'ultima osservazione, implicita in quanto già detto: la probabilità di trovare la variabile casuale entro un certo numero di deviazioni standard non dipende dai valori di μ e di σ. Diamo alcuni valori notevoli di probabilità (vedi anche figura 8.9):

$$\begin{aligned} P(\mu - 0.675\,\sigma \leq X \leq \mu + 0.675\,\sigma) &= 50.0\,\% \\ P(\mu - \sigma \leq X \leq \mu + \sigma) &= 68.3\,\% \\ P(\mu - 1.64\,\sigma \leq X \leq \mu + 1.64\,\sigma) &= 90.0\,\% \\ P(\mu - 1.96\,\sigma \leq X \leq \mu + 1.96\,\sigma) &= 95.0\,\% \\ P(\mu - 2\,\sigma \leq X \leq \mu + 2\,\sigma) &= 95.4\,\% \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 99.73\,\% \end{aligned} \tag{8.20}$$

È interessante confrontare questo si ottiene entro 1 e 2 σ con quanto visto nel paragrafo 7.10.

Per tornare ancora una volta sul fatto che la probabilità che un numero aleatorio sia compreso nell'intervallo di ± 1 deviazione standard dal valore atteso, mostriamo in figura 8.7 quattro diverse distribuzioni aventi tutte stesso valore atteso e stessa deviazione standard.

8.10 ❄ Derivazione della gaussiana come limite di funzione binomiale o poissoniana

La figura 7.1 mostra come la distribuzione binomiale ha una forma a campana che somiglia alla gaussiana, quando n, $n\,p$ e $n\,q$ sono abbastanza grandi. È quindi interessante trovare l'epressione asintotoca della binomiale. Questo problema portò alla funzione che ora conosciamo come gaussiano quasi un secolo prima della derivazione di Gauss, sulla quale ritorneremo nel paragrafo 11.4. Riscriviamo la funzione di probabilità binomiale riscrivendo i fattoriali nella loro espressione asintotica data dalla formula di Stirling (De Moivre e Stirling, più precisamente), che ricordiamo qui:

$$n! \xrightarrow[n\to\infty]{} n^n e^{-n}\sqrt{2\,\pi\,n} \propto n^{n+1/2}e^{-n}\,.$$

Chiamando $\Delta = x - n\,p$, da cui

$$\begin{aligned} x &= n\,p - \Delta \\ n - x &= n\,q - \Delta\,, \end{aligned}$$

ne segue, omettendo di volta in volta i fattori moltiplicativi che non dipendono da x (ovvero da Δ) e che possiamo inglobare nella costante di normalizzazione:

$$\begin{aligned} f(x\,|\,\mathcal{B}_{n,p}) &\propto \frac{n^{n+1/2}}{x^{x+1/2}(n-x)^{n-x+1/2}}\,p^x q^{n-x} \\ &\propto \frac{p^x q^{n-x}}{p^x\left(1+\frac{\Delta}{n\,p}\right)^{n\,p+\Delta+1/2} q^{n-x}\left(1-\frac{\Delta}{n\,q}\right)^{n\,p-\Delta+1/2}} \\ &\propto \underbrace{\left(1+\frac{\Delta}{n\,p}\right)^{-(n\,p+\Delta+1/2)}}_{D_p}\underbrace{\left(1-\frac{\Delta}{n\,q}\right)^{-(n\,q-\Delta+1/2)}}_{D_q} \end{aligned}$$

Per studiare il comportamento asintotico di D_p e D_q è preferibile passare ai logaritmi:

$$\begin{aligned}
\ln D_p &= -(n\,p+\Delta+\frac{1}{2})\,\ln\left(1+\frac{\Delta}{n\,p}\right) \xrightarrow[\Delta\ll n\,p]{} -\Delta-\frac{\Delta^2}{2\,n\,p} \\
\ln D_q &= -(n\,q-\Delta+\frac{1}{2})\,\ln\left(1-\frac{\Delta}{n\,p}\right) \xrightarrow[\Delta\ll n\,p]{} +\Delta-\frac{\Delta^2}{2\,n\,q}
\end{aligned}$$

$$\ln D_p+\ln D_q \quad \xrightarrow[\left\{\begin{array}{l}\Delta\ll n\,p\\ \Delta\ll n\,q\end{array}\right.]{} -\frac{\Delta^2}{2\,n\,p\,q}$$

$$D_p\,D_q \quad \xrightarrow[\left\{\begin{array}{l}\Delta\ll n\,p\\ \Delta\ll n\,q\end{array}\right.]{} e^{-\Delta^2/(2\,n\,p\,q)}$$

$$X \quad \xrightarrow[\left\{\begin{array}{l}\Delta\ll n\,p\\ \Delta\ll n\,q\end{array}\right.]{} \sim \mathcal{N}(n\,p,\sqrt{n\,p\,(1-p)}.$$

Per completezza, vediamo anche come si può arrivare alla gaussiana come limite della poissoniana per $\lambda\to\infty$, ovvero quando i valori x sui quali si addensa la massa di probabilità sono relativamente prossimi a λ e lontani da zero. Chiamando, in analogia al caso precedente $\Delta = x-\lambda$, prendiamo il logaritmo della funzione di distribuzione poissoniana e, con l'ausilio dell'approssimazione di Stirling, facciamo il limite per $\Delta/\lambda\to 0$ (ovvero siamo interessati a piccole deviazioni dal valore atteso). Omettendo, come nel caso precedente, i termini che di volta in volta non dipendono da Δ e sviluppando il logaritmo di $1+\Delta/\lambda$ al secondo ordine intorno a 1, abbiamo:

$$\begin{aligned}
\ln f(x\,|\,\mathcal{P}_\lambda) &\approx x\,\ln\lambda-(x+1/2)\,\ln x+x \\
&= x\,\ln\lambda-(x+1/2)\,\ln\left[\lambda\left(1+\frac{\Delta}{\lambda}\right)\right]+x \\
&\approx -\left(\lambda+\Delta+\frac{1}{2}\right)\left[\frac{\Delta}{\lambda}-\frac{1}{2}\left(\frac{\Delta}{\lambda}\right)^2\right]+\Delta \\
&\approx -\frac{\Delta^2}{2\,\lambda},
\end{aligned}$$

da cui segue

$$X \quad \xrightarrow[\left\{\begin{array}{l}\Delta\ll \lambda\\ \lambda\gg 0\end{array}\right.]{} \mathcal{N}(\lambda,\sqrt{\lambda})\,. \qquad (8.21)$$

8.11 ↻ Proprietà riproduttiva della distribuzione normale

La distribuzione normale gode della proprietà riproduttiva rispetto alla combinazione lineare (vedi dimostrazione nei paragrafi 8.13 e 10.3.3): Se $X_1\sim$

$\mathcal{N}(\mu_1, \sigma_1)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$, la variabile $Y = c_\circ + c_1\, X_1 + c_2\, X_2$ segue una distribuzione normale di parametri $\mu = c_\circ + c_1\, \mu_1 + c_2\, \mu_2$ e $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

8.12 ⥁ Processo di Poisson - Seconda parte

Riprendiamo la trattazione del processo di Poisson. Nel paragrafo 7.5 ci eravamo interessati del numero di conteggi avendo fissato il tempo di osservazione. Analizziamo ora il suo aspetto complementare: fissiamo il numero di conteggi e ci interessiamo all'istante di tempo nel quale tale numero viene raggiunto. Come appare chiaro, si tratta della stessa complementarità già discussa fra distribuzione binomiale e distribuzione geometrica (e di Pascal per il caso generale).

8.12.1 Distribuzione del tempo di attesa del primo successo

Calcoliamo innanzitutto la probabilità che non si verifichi nessun *"evento"*[5] durante un certo tempo finito t. Lo facciamo in modo generale considerando il numero aleatorio reale T, definito come "tempo di attesa per registrare il primo conteggio, a partire da un certo istante arbitrario". Se, come fatto precedentemente, immaginiamo di suddividere il tempo finito t in n intervallini, otteniamo (vedi distribuzione geometrica)

$$\begin{aligned} P(T > t) &= (1-p)^n \\ &= (1 - r\,\frac{t}{n})^n \xrightarrow[n\to\infty]{} e^{-rt}\,, \end{aligned} \tag{8.22}$$

da cui segue la probabilità cumulativa

$$F(t) \equiv P(T \le t) = 1 - P(T > t) = 1 - e^{-r\,t}\,. \tag{8.23}$$

Quindi la funzione cumulativa dei tempi di attesa affinché si verifichi un evento descritto da un processo di Poisson è data da una *esponenziale*. Per ottenere la funzione densità densità di probabilità deriviamo la $F(t)$, ottenendo:

$$f(t) = \frac{\mathrm{d}\,F(t)}{\mathrm{d}\,t} = r\,e^{-r\,t}\,.$$

Introducendo la grandezza $\tau = 1/r$ omogenea a t, otteniamo l'espressione usuale della distribuzione esponenziale nel dominio tempo (vedi (8.12). Il parametro τ ha quindi il significato di *previsione del tempo di attesa prima che si verifichi il primo conteggio*, a partire da un istante arbitrario. Essendo $\tau = 1/r$, tale previsione è pari all'*inverso della previsione del numero di eventi per unità di tempo*.

Si noti che essendo arbitrario l'istante da cui parte l'osservazione, esso può anche essere uno dei possibili conteggi. Quindi la distribuzione trovata descrive il *tempo di arrivo fra due conteggi successivi*.

[5]Si faccia attenzione al diverso significato che acquista qui il termine "evento". Secondo la prassi scientifica esso è anche utilizzato con l'accezione di "occorrenza" (ad esempio, "l'esperimento ha registrato 1000 eventi di interazione da neutrino"). È opportuno abituarsi a convivere a queste ambiguità di linguaggio, quando dal contesto si evince il significato corretto del termine.

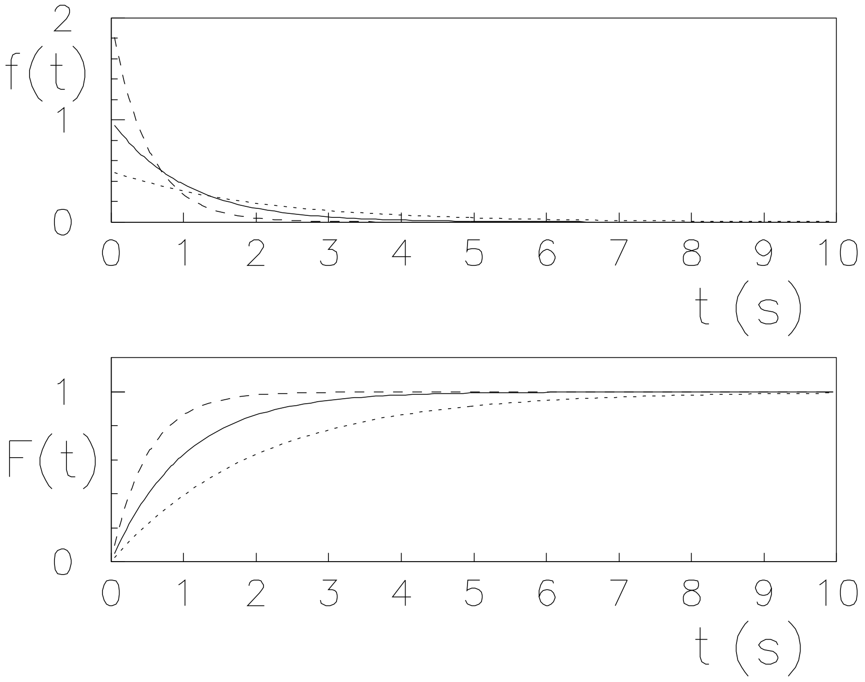


Figura 8.10: Esempi di distribuzioni esponenziali nel dominio del tempo.

8.12.2 Relazione fra esponenziale e poissoniana

Abbiamo visto come la distribuzione esponenziale e quella di Poisson rappresentano due modi di descrivere lo stesso processo. I parametri delle due distribuzioni dipendono l'uno dall'altro:

- se si ha un processo di Poisson con un tasso di conteggi nell'unità di tempo di r, ne segue che:
  - fissando un tempo finito T di osservazione, il numero di eventi che sarà possibile osservare segue una distribuzione di Poisson con $\lambda = r\,T$;
  - l'intervallo di tempo che bisogna attendere per osservare il primo evento, a partire da un certo istante arbitrario, è un numero aleatorio reale descritto da una esponenziale di parametro $\tau = 1/r$;
  - fra i parametri delle due distribuzioni vale quindi la relazione

$$\tau = \frac{T}{\lambda}\,. \tag{8.24}$$

La figura 8.11 illustra bene la complementarità fra esponenziale e poissoniana.

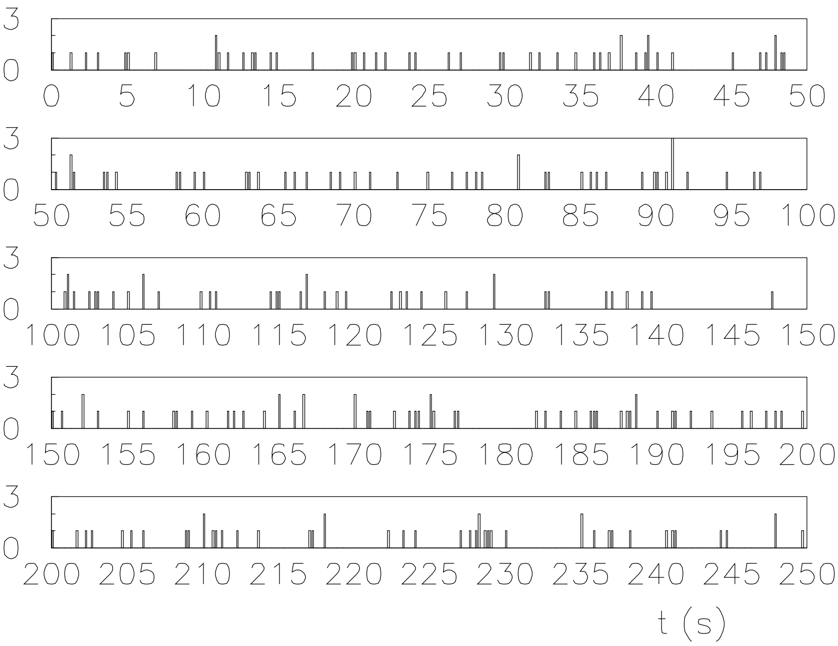


Figura 8.11: Simulazione al computer dell'istante di occorrenza di eventi casuali descritti da un processo di Poisson di intensità r uguale a 1 conteggio al secondo. La risoluzione temporale dell'istogrammaè pari a $0.2\,s$.

8.12.3 Relazione fra esponenziale e geometrica

Che l'esponenziale non sia altro che il limite della geometrica per $p \rightarrow 0$ dovrebbe essere già chiaro da come essa è stata introdotta. Facciamo altri chiarimenti:

- quando $p \rightarrow 0$ non ha più senso parlare del numero di prova, in quanto le "prove" (ciascuna in un intervallino di tempo $\Delta t \rightarrow 0$) sono necessariamente infinite;

- così pure perde di significato p per "ciascuna" prova (essa è pari a zero indipendentemente dall'"intensità" del processo) e si può parlare soltanto di τ o di r;

- in particolare, ricordando le espressioni del valore atteso le due distribuzioni, è possibile esprimere in modo analogo p e r delle due distribuzioni:

  geometrica: p è pari al *numero medio di successi per estrazione*;

  esponenziale r è pari al *numero medio di successi per unità di tempo* (r ha pertanto il significato di "intensità" del processo di Bernoulli).

Un altro aspetto comune delle due distribuzioni è la cosiddetta proprietà di *assenza di memoria* delle due distribuzioni, espressa in generale come

$$P(X > x + x_\circ \,|\, X > x_\circ) \;=\; P(X > x) \qquad (8.25)$$

(per l'esponenziale si sostituisca X a T), facilmente dimostrabile dalla formula della probabilità condizionata. Vedremo una applicazione di tale proprietà nell'esempio dei decadimenti radioattivi. Si noti comunque come questa proprità di assenza di memoria caratterizzi le due distribuzioni. Difatti è possibile dimostrare come la richiesta della (8.25) conduca alla geometrica o all'esponenziale a seconda che si tratti di variabile discreta o continua.

8.12.4 Tempo di attesa del k-mo successo

Considerando un processo di Poisson nel dominio del tempo di intensità r, si può essere, in generale, interessati alla variabile casuale T "tempo di attesa affinché si verifichi il k-mo successo". Il caso $k = 1$ è quello descritto dall'esponenziale. Consideriamo l'evento E "il successo k-mo si verifica ad un tempo $T > t$". La richiesta di questo evento corrisponde alla richiesta che entro il tempo $T = t$ si siano verificati al più $k - 1$ successi. Quindi, poiché il numero di successi nel tempo $T = t$ è dato da una distribuzione di Poisson di parametro $\lambda = r\,t$, abbiamo:

$$P(E) = f(0\,|\,\mathcal{P}_{r\,t}) + f(1\,|\,\mathcal{P}_{r\,t}) + \cdots + f(k-1|\mathcal{P}_{r\,t}) = F(k-1\,|\,\mathcal{P}_{r\,t})\,.$$

La probabilità che l'ennesimo successo si verifichi entro il tempo $T = t$ è uguale a

$$\begin{aligned} P_k(T \leq t) &= 1 - P(E) = 1 - F(k-1\,|\,\mathcal{P}_{r\,t} \\ &= 1 - \sum_{x=0}^{k-1} \frac{e^{-r\,t}\cdot(r\,t)^x}{x!}\,. \end{aligned}$$

Si può verificare - integrando iterativamente per parti - che quest'ultima espressione è soluzione del seguente integrale:

$$\int_0^x \frac{t^{k-1}}{(k-1)!}\cdot r^k \cdot e^{-r\,t}\mathrm{d}t\,.$$

Otteniamo finalmente la funzione di partizione e quindi la distribuzione di probabilità cercate, nota come distribuzione *di Erlang*:

$$\begin{aligned} F(t\,|\,\mathrm{Erlang}(k,r)) &= \int_0^t \frac{t'^{k-1}}{(k-1)!}\cdot r^k\cdot e^{-r\,t'}\mathrm{d}t' \\ f(t\,|\,\mathrm{Erlang}(k,r)) &= \frac{t^{k-1}}{(k-1)!}\cdot r^k\cdot e^{-r\,t}\,. \qquad (8.26) \end{aligned}$$

Valore atteso, varianza e deviazione standard valgono rispettivamente

$$\begin{aligned} \mathrm{E}(T) &= \frac{k}{r} = k\,\tau & (8.27) \\ \mathrm{Var}(T) &= \frac{k}{r^2} & (8.28) \\ \sigma(T) &= \frac{\sqrt{k}}{r} = \sqrt{k}\,\tau\,, & (8.29) \end{aligned}$$

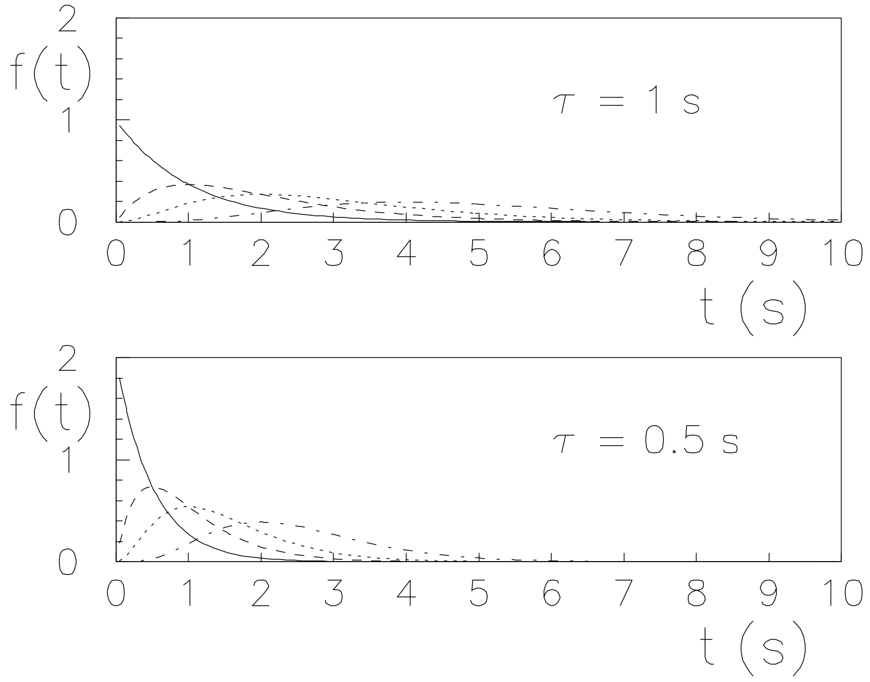


Figura 8.12: Distribuzione del tempo di attesa del k-mo successo ("Erlang") in un processo di Poisson di intensità r tale che $\tau = 1/r$ sia pari a 1 e 0.5 secondi, con k uguale a 1 (—), 2 (- - - -), 3 (· · · · · ·) e 5 (— · — · —).

ove si è fatto uso anche di $\tau = 1/r$ per esprimere il risultato in modo più intuitivo: la previsione del tempo di attesa è pari a k volte quello di attesa del primo successo, mentre la deviazione standard aumenta come la radice quadrata del numeri di successi. Valore atteso e varianza possono essere valutati in modo molto semplice pensando a k distribuzioni esponenziali (vedi paragrafo 10.7). La distribuzione di Erlang sta alla distribuzione esponenziale come la distribuzione di Pascal sta alla geometrica.

La distribuzione può essere estesa anche a valori di k non interi, ottenendo la distribuzione Gamma (vedi par. 12.4).

8.12.5 Intensità di più processi di Poisson indipendenti

Per quanto detto a proposito del processo di Poisson, isulta evidente che, se si hanno diversi processi di Poisson indipendenti, ciascuno avente intensità r_i e che producono eventi fra loro indistinguibili (nei limiti della nostra risoluzione o del nostro interesse), essi possono essere considerati come un unico evento di Poisson di intensità pari alle somma delle intensità:

$$r_{tot} = \sum_i r_i. \tag{8.30}$$

8.12.6 Vita media di decadimento

Supponiamo di avere un nucleo radioattivo (o una particella subnucleare instabile) per il quale la probabilità di decadimento sia indipendente dal tempo, ovvero $dp = r\,dt$. Per quanto detto a proposito del processo di Poisson, l'istante di decadimento del nucleo a partire da un certo istante scelto arbitrariamente è descritto da una distribuzione esponenziale di parametro r. Il valore atteso del tempo di decadimento è pari a $\tau = 1/r$, chiamato anche *vita media di decadimento*. È interessante calcolare il tempo tale che ci sia il 50 % di probabilità che la particella sia già decaduta (ovvero la mediana della distribuzione):

$$P(T \leq t_{1/2}) = 1 - e^{-t/\tau} = 0.5\,,$$

ovvero

$$t_{1/2} = \tau\,\ln 2\,.$$

La mediana è chiamata, in questa applicazione, anche *tempo di dimezzamento*. Per capire meglio il suo significato, consideriamo $N_\circ$ nuclei all'istante $T = 0$, inizio delle nostre osservazioni, e calcoliamoci previsione e incertezza di previsione del numero di nuclei che sono rimasti non decaduti all'istante t. Dobbiamo considerare una distribuzione binomiale avente $n = N_\circ$ e $p = e^{-t/\tau}$ e, quindi, di valore atteso e deviazione standard

$$\begin{aligned}
\mathrm{E}(N) &= N_\circ e^{-t/\tau} \\
\sigma(N) &= \sqrt{N_\circ}\sqrt{e^{-t/\tau}\,(1 - e^{-t/\tau})} \xrightarrow[t \ll \tau]{} \sqrt{\mathrm{E}(N)}\,.
\end{aligned}$$

Dopo un tempo $t = t_{1/2}$ il numero iniziale di nuclei si è dimezzato. Si noti inoltre come, per $N_\circ$ molto grandi (tipicamente si considerano numeri dell'ordine di grandezza del numero di Avogadro) $\sigma(N)/\mathrm{E}(N)$ è molto minore di 1 anche per t abbastanza maggiore della vita media. Quindi la previsione del numero di nuclei non decaduti può essere considerata con ottima approssimazione una predizione deterministica che obbedisce ad una legge del tipo

$$N(t) = N_\circ e^{-t/\tau}\,,$$

soluzione dell'equazione differenziale

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -r\,N = -\frac{1}{\tau}N\,.$$

Essa può essere espressa dicendo che "il numero di decadimenti nell'unità di tempo è proporzionale al numero di nuclei, con un fattore di proporzionalità pari all'inverso della vita media di decadimento".

La figura 8.13 illustra la legge esponenziale del decadimento e la proprietà di assenza di memoria della distribuzione esponenziale.

8.13 ❄ Funzione generatrice dei momenti

Questo paragrafo, decisamente tecnico, mostra come sia possibile calcolare in modo relativamente semplice i momenti delle distribuzioni (e da questi, ad

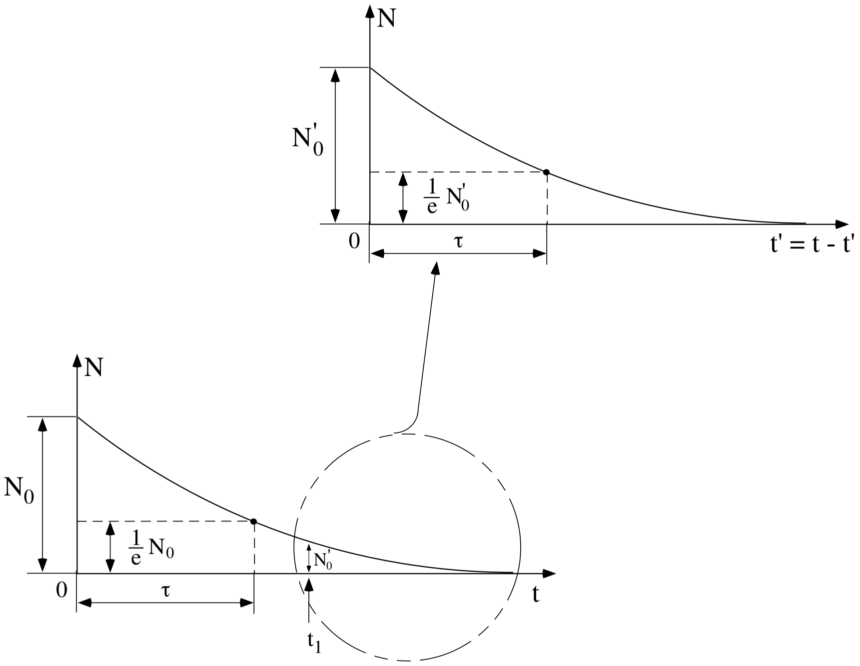


Figura 8.13: Proprietà di "assenza di memoria" di una distribuzione esponenziale (con applicazione ai decadimenti radioattivi).

esempio, media e deviazione standard), facendo uso di proprietà formali dei valori attesi.

Consideriamo il valore atteso della funzione $e^{X\,t}$, dove t rappresenta un parametro che può assumere con continuità valori reali, ed indichiamolo con $G(t)$:

$$G(t) = \mathrm{E}\left(e^{X\,t}\right). \tag{8.31}$$

L'interesse di questo valore atteso risiede nella proprietà matematica di cui esso gode. Espandendo $G(t)$ in serie di Taylor intorno a $t = 0$ si ha infatti:

$$\begin{aligned} G(t) &= \mathrm{E}\left(1 + X\,t + \frac{1}{2}\,X^2\,t^2 + \cdots + \frac{1}{x!}\,X^n\,t^n\right) \\ &= 1 + \mu\,t + \frac{1}{2}\,\mu_2\,t^2 + \cdots + \frac{1}{n!}\,\mu_n\,t^n\,. \end{aligned}$$

Otteniamo che i coefficienti dell'espansione sono proporzionali ai momenti intorno a zero, essendo

$$\mu^r = \mathrm{E}(X^r)\,.$$

(si noti che t è un parametro e non un numero aleatorio).

Se effettuiamo la derivata di ordine r di $G(t)$ rispetto a t, il primo termine non nullo è μ_r, mentre tutti i termini successivi sono proporzionali ai momenti di ordine superiore a r, moltiplicati per potenze di t. Prendendo il valore della derivata r-esima di $G(t)$, calcolata in $t = 0$, possiamo allora isolare quindi il momento di ordine r:

$$\mu_r = \left.\frac{\mathrm{d}^r}{\mathrm{d}t^r}G(t)\right|_{t=0}\,. \tag{8.32}$$

A causa della notevole proprietà di cui $G(t)$, essa è nota come *funzione generatrice dei momenti*. La sua utilità risiede nel fatto che a volte è più semplice ricavarsi $G(t)$ una volta per tutte e ottenere i momenti mediante derivate che effettuare gli integrali o le sommatorie necessarie per il loro calcolo diretto.

Applichiamo questa tecnica ad alcune della distribuzioni incontrate, lasciando come esercizio il caso della distribuzione esponenziale:

Binomiale

$$\begin{aligned} G(t\,|\,\mathcal{B}_{n,p}) &= \sum_{x=0}^{n} e^{t\,x}\binom{n}{x} p^x\, q^{n-x} \\ &= \sum_{x=0}^{n}\binom{n}{x}\left(e^t\,p\right)^x\, q^{n-x} \\ &= \left(e^t\,p+q\right)^n . \end{aligned}$$

Riotteniamo quindi rapidamente:

$$\begin{aligned} \mathrm{E}(X) &= n\,p\,e^t\,\left(e^t\,p+q\right)^{n-1}\Big|_{t=0} = n\,p \\ \mathrm{E}(X^2) &= \Big[n\,(n-1)\,\left(p\,e^t\right)^2\,\left(e^t\,p+q\right)^{n-2} \\ &\quad +n\,p\,e^t\,\left(e^t\,p+q\right)^{n-1}\Big]\Big|_{t=0} \\ &= n\,(n-1)\,p^2+n\,p \\ \mathrm{Var}(X) &= n\,p\,q\,. \end{aligned}$$

Poissoniana

$$\begin{aligned} G(t\,|\,\mathcal{P}_{n,p}) &= \sum_{x=0}^{\infty}\frac{e^{t\,x}\,e^{-\lambda}\,\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\left(\lambda\,e^t\right)^x}{x!} \\ &= e^{-\lambda}e^{\lambda e^t}\,, \end{aligned}$$

da cui

$$\begin{aligned} \mathrm{E}(X) &= \lambda\,e^{-\lambda}\,e^{\lambda\,e^t}\,e^t\Big|_{t=0} = \lambda \\ \mathrm{E}(X^2) &= \lambda\,e^{-\lambda}\left(\lambda\,e^{\lambda\,e^t}\,e^t + e^{\lambda\,e^t}\right)\Big|_{t=0} = \lambda\,(\lambda+1) \\ \mathrm{Var}(X) &= \lambda\,. \end{aligned}$$

Gaussiana

$$\begin{aligned} G(t\,|\,\mathcal{N}(\mu,\sigma)) &= \int_{-\infty}^{+\infty}\frac{e^{tx}}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \\ &= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x^2+\mu^2-2\,x\,\mu-2\,\sigma^2\,t\,x)/2\sigma^2}\mathrm{d}x\,. \end{aligned}$$

Riscriviamo il numeratore del termine all'esponente nella seguente forma:

$$x^2 - 2\,(\mu + \sigma^2\,t)x + (\mu + \sigma^2\,t)^2 - (\mu + \sigma^2\,t)^2 + \mu^2\,.$$

Introducendo il nuovo parametro $\mu' = \mu + \sigma^2\,t$ otteniamo:

$$G(t\,|\,\mathcal{N}(\mu,\sigma)) \;\;=\;\; e^{\frac{(\mu+\sigma^2\,t)^2-\mu^2}{2\sigma^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu')^2}{2\sigma^2}}\mathrm{d}x\,,$$

da cui, finalmente:

$$\begin{aligned}
G(t\,|\,\mathcal{N}(\mu,\sigma)) &= e^{\mu\,t+\sigma^2\,t^2/2} \\
\\
\mathrm{E}(X) &= (\mu+\sigma^2\,t)\,e^{\mu\,t+\sigma^2\,t^2/2}\Big|_{t=0} = \mu \\
\mathrm{E}(X^2) &= \left[(\mu+\sigma^2\,t)^2+\sigma^2\right]\,e^{\mu\,t+\sigma^2\,t^2/2}\Big|_{t=0} = \mu^2+\sigma^2 \\
\mathrm{Var}(X) &= \sigma^2\,.
\end{aligned}$$

Altre proprietà e applicazioni

Oltre a poter ricalcolare agevolmente i momenti delle distribuzioni più note, la funzione generatrice è molto utile nel caso di distribuzioni più complicate, ottenute come trasformazioni da variabili di cui sia nota la funzione generatrice. Valgono infatti le seguenti proprietà:

$$\begin{aligned}
&1) \qquad G_{a\,X}(t) = G_X(a\,t) &\quad (8.33)\\
&2) \qquad G_{X+b}(t) = e^{b\,t}G_X(t) &\quad (8.34)\\
&3) \qquad G_{X+Y}(t) = G_X(t)G_Y(t) &\quad (8.35)\\
&\qquad\quad (\text{con } X \text{ e } Y \text{ indipendenti})\,.
\end{aligned}$$

Le prime due proprità si ricavano facilmente applicando la definizione di $G(t)$. La terza invece richiederebbe, in principio, la conoscenza delle distribuzioni di molte variabili (vedi capitolo 9), ma si dovrebbe capire intuitivamente di cosa si tratta.

Come esercizio su queste proprità, ricaviamoci la funzione generatrice di una variabile Y ottenuta da una trasformazione lineare di un'altra variabile, X, distribuita normalmente:

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

Dalla prime due proprietà si ottiene:

$$\begin{aligned}
G_Y(t) &= e^{b\,t}G_X(a\,t\,|\,\mathcal{N}(\mu_X,\sigma_X)) \\
&= e^{b\,t}e^{\mu_X\,(a\,t)+\sigma_X^2\,(a\,t)^2/2} \\
&= e^{(a\,\mu_X+b)t+(a^2\,\sigma_X^2)\,t^2/2} \\
&= e^{\mu_Y\,t+\sigma_Y^2\,t^2/2}\,, \qquad (8.36)
\end{aligned}$$

con

$$\left\{ \begin{array}{l} \mu_Y = a\,\mu_X + b \\ \sigma_Y^2 = a^2\,\sigma_X^2 \end{array} \right. .$$

Ne segue che una trasformazione lineare trasforma una distribuzione normale in un'altra distribuzione normale.

Vedremo le applicazioni della terza proprietà nel paragrafo 10.4.

8.14 ↻ Altre distribuzioni di interesse

In questo paragrafo vengono elencate delle distribuzioni di interesse generale. Nonostante siano messe in questo capitolo tutte insieme, è consigliabile evitarne lo studio sistematiche ed affrontarle soltanto quando si presenta il problema specifico di interesse.

8.14.1 Beta

$$f(x\,|\,\text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \left\{ \begin{array}{l} 0 \le x \le 1 \\ r,\, s > 0 \end{array} \right. \tag{8.37}$$

Il denominatore ha il ruolo di costante di normalizzazione, ovvero

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\,\mathrm{d}x\,.$$

Questa funzione speciale, denominata "beta" e che dà il nome alla distribuzione è calcolabile dalle Gamma di Eulero mediante la relazione

$$\beta(r,s) = \frac{\Gamma(r)\,\Gamma(s)}{\Gamma(r+s)}$$

ove, ricordiamo,

$$\Gamma(c) = \int_0^\infty x^{c-1}e^{-x}dx\,,$$

che per argomento n intero vale

$$\Gamma(n+1) = n!$$

Valore atteso e varianza sono:

$$\mathrm{E}(X) = \frac{r}{r+s} \tag{8.38}$$

$$\mathrm{Var}(X) = \frac{r\,s}{(r+s+1)(r+s)^2} \tag{8.39}$$

$$= \mathrm{E}(X)\,\frac{s}{(r+s+1)(r+s)} \tag{8.40}$$

Se $r > 1$ e $s > 1$ la moda è unica e vale $(r-1)/(r+s-2)$. La figura 8.14 mostra come la distribuzione beta possa assumere una ricca varietà di forme e quindi si presta a modellizzare bene un certo numero di problemi. In particolare, per $r = s = 1$ la distribuzione si riduce ad una distribuzione uniforme. Si noti la somiglianza formale della (8.37), a meno del fattore di normalizzazione, con l'espressione della distribuzione binomiale, a meno del coefficiente binomiale. Vedremo infatti come utilizzare tale proprietà formale per semplificare un classico problema infernziale (vedi paragrafo 12.3).

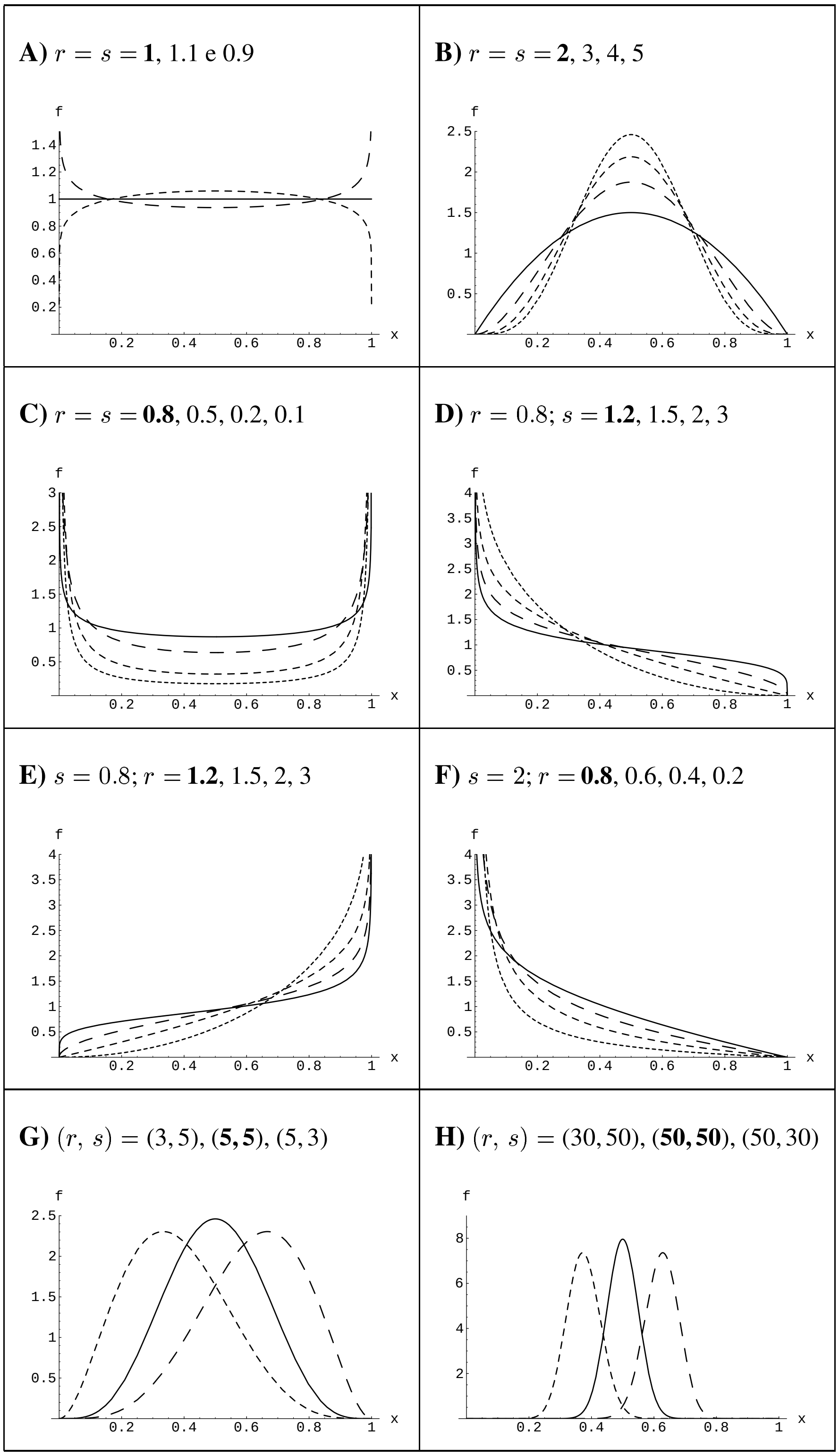


Figura 8.14: Esempi di distribuzioni Beta per vari valori di r e s. I numeri in grassetto si riferiscono alle curve continue.

8.14.2 Gamma

$$f(x \,|\, \text{Gamma}(c,r)) = \frac{r^c}{\Gamma(c)} x^{c-1} e^{-r\,x} \qquad \left\{ \begin{array}{l} x \geq 0 \\ r,\, c > 0 \end{array} \right. , \qquad (8.41)$$

Valore atteso, varianza e moda sono:

$$\text{E}(X) = \frac{c}{r} \qquad (8.42)$$

$$\text{Var}(X) = \frac{c}{r^2} \qquad (8.43)$$

$$\text{moda}(X) = \left\{ \begin{array}{ll} 0 & \text{se } c < 1 \\ \frac{c-1}{r} & \text{se } c \geq 1 \end{array} \right. . \qquad (8.44)$$

Se c è intero si parla della distribuzione di *Erlang*, la quale descrive il tempo di attesa per osservare c eventi in un processo di Poisson evente un tasso di conteggi per unità di tempo pari ad r. Alcuni esempi sono mostrati in figura 8.15. Chiaramente, quando $c = 1$ essa si riduce alla distribuzione esponenziale negativa.

La funzione generatrice dei momenti vale

$$G(t) = \left(1 - \frac{t}{r}\right)^{-c} .$$

8.14.3 Chi^2

La distribuzione di Chi^2 (si legge "chi quadro") è data da

$$f(x \,|\, \text{chi}^2(\nu)) = \left(2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)\right)^{-1} x^{\frac{\nu}{2}-1} e^{-x/2} \qquad \left\{ \begin{array}{l} x \geq 1 \\ \nu > 0 \end{array} \right. \qquad (8.45)$$

La variabile che segue tale distribuzione è usualmente indicata con χ^2_ν. Il parametro ν, non necessariamente intero è detto *numero di gradi di libertà*. Valore atteso, varianza e moda sono:

$$\text{E}(X) = \nu \qquad (8.46)$$

$$\text{Var}(X) = 2\,\nu\,, \qquad (8.47)$$

$$\text{moda}(X) = \left\{ \begin{array}{ll} 0 & \text{se}\nu \leq 2 \\ \nu - 2 & \text{se}\nu > 2 \end{array} \right. \qquad (8.48)$$

Alcuni esempi sono mostrati in figura 8.16.

La funzione generatrice dei momenti vale $G(t) = (1 - 2\,t)^{\frac{\nu}{2}}$.

La distribuzione di Chi^2 è legata alla distribuzione Gamma da

$$\text{chi}^2(\nu) = \text{Gamma}\left(\frac{\nu}{2}, \frac{1}{2}\right) .$$

Una importante importante proprietà che la rende utile per le applicazioni è la seguente: se n variabili Z_i indipendenti seguono una distribuzione normale standardizzata, la loro somma segue una distribuzione di Chi^2 di parametro $\nu = n$. Inoltre, se n variabili indipendenti X_i seguono ciascuna una distribuzione di Chi^2 di parametro ν_i, la loro somma segue una distribuzione di Chi^2 evente $\nu = \sum_i \nu_i$ (proprietà riproduttiva rispetto alla somma).

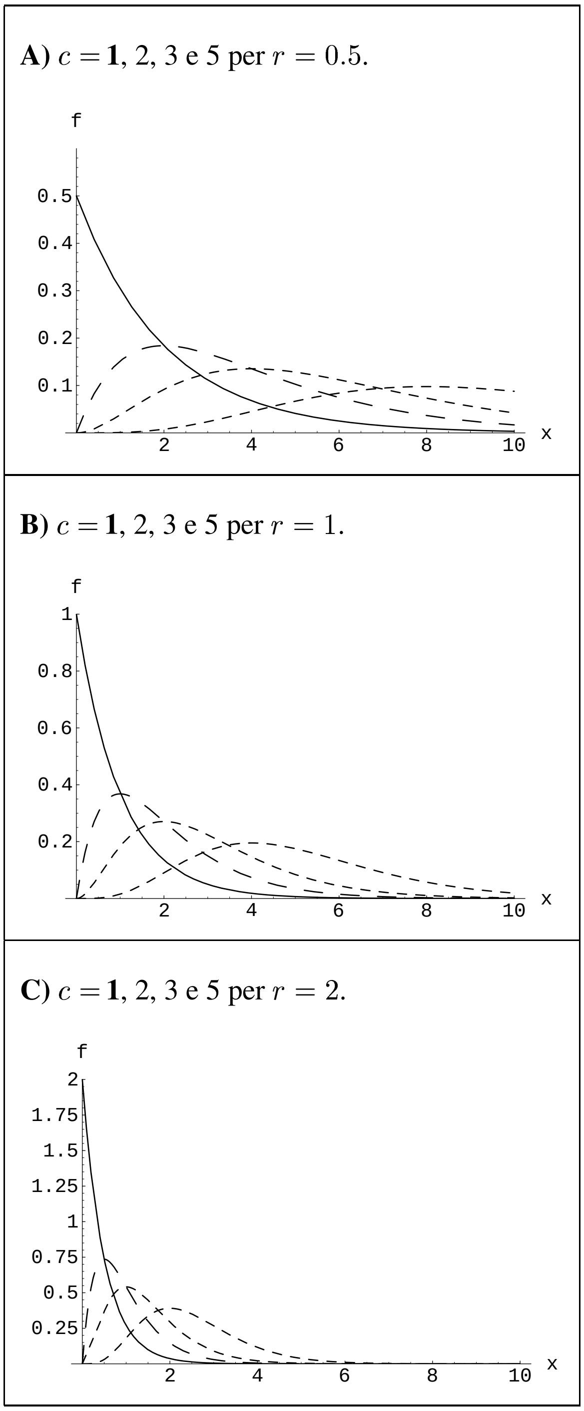


Figura 8.15: Esempi di distribuzioni Gamma per diversi valori dei parametri. I numeri in grassetto si riferiscono alle curve continue.

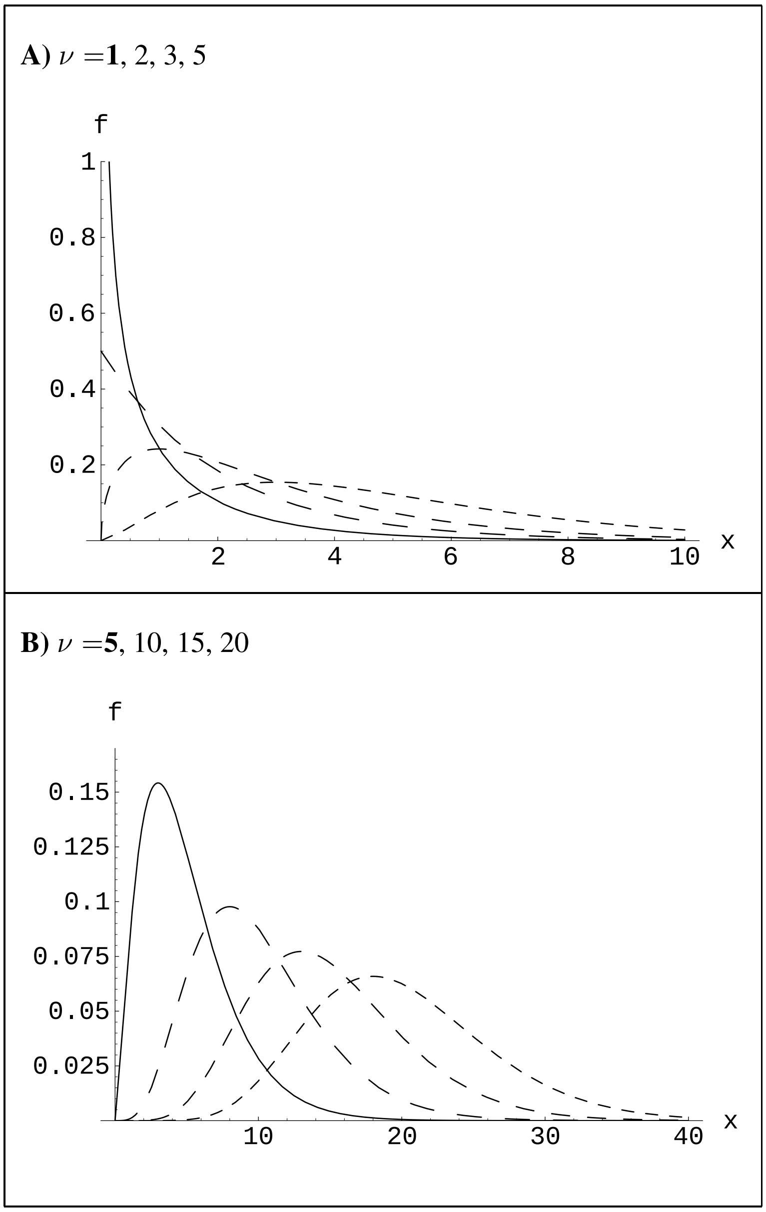


Figura 8.16: Esempi di distribuzioni di Chi^2. I numeri in grassetto si riferiscono alle curve continue.

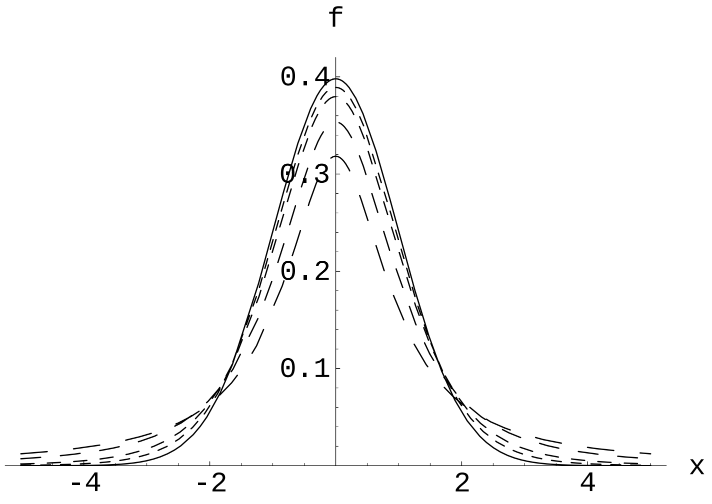


Figura 8.17: Esempi di distribuzioni di t di Student per ν uguale a 1 (curva più larga), 2, 5, 10 e 100 ($\approx$ "∞").

8.14.4 t di Student

$$f(x\,|\,\mathrm{Student}(\nu)) = \frac{1}{\beta\left(\frac{1}{2},\frac{\nu}{2}\right)\sqrt{\nu}}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \begin{cases} -\infty < x < +\infty \\ \nu > 0 \end{cases} \tag{8.49}$$

La variabile che segue la distribuzione di Student è indicata usualmente con il simbolo t (o t_ν) e chiamata "t di Student". Come per il caso della distribuzione di Chi^2, il parametro ν, non necessariamente intero è detto *numero di gradi di libertà*. Valore atteso e varianza sono:

$$\mathrm{E}(X) = 0 \qquad (\text{se } \nu > 1) \tag{8.50}$$

$$\mathrm{Var}(X) = \frac{\nu}{\nu-2} \qquad (\text{se } \nu > 2) \tag{8.51}$$

mentre la moda vale 0. La funzione generatrice dei momenti non esiste. Una proprietà importante per l'uso è la seguente: se si hanno una variabile Z che segue una normale standardizzata e una variabile χ^2_ν che segue una distribuzione di Chi^2 con ν gradi di libertà e le due variabili sono indipendenti, allora la nuova variabile $Z/\sqrt{\chi^2/\nu}$ segue una distribuzione di Student di parametro ν.

La t di Student ha una forma a campana come la gaussiana, ma con code più pronunciate per piccoli valori di ν. Per ν grande (virtualmente per $\nu \to \infty$, ma in pratica quando ν supera la decina) la distribuzione tende ad una gaussiana standardizzata.

8.14.5 F

La distribuzione di Fisher (o di Fisher-Snedecor) chiamata anche semplicemente F è data da

$$f(x\,|\,\mathrm{F}(\nu_1,\nu_2)) = \frac{\nu_1^{\frac{\nu_1}{2}}\,\nu_2^{\frac{\nu_2}{2}}}{\beta\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}\,\frac{x^{\frac{\nu_1}{2}-1}}{(\nu_1\,x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}} \qquad \begin{array}{l} x \geq 0 \\ \nu_1,\,\nu_2 > 0 \end{array}\,, \tag{8.52}$$

Talvolta la variabile che segue la distribuzione F è indicata anch'essa con il simbolo F_{ν_1,ν_2}. I parametri ν_1 e ν_2, non necessariamente intero è detti *numeri di gradi di libertà*. Valore atteso e varianza sono:

$$\mathrm{E}(X) = \frac{\nu_2}{\nu_2-2} \quad (\text{se } \nu > 2) \tag{8.53}$$

$$\mathrm{Var}(X) = \frac{2\nu^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)} \quad (\text{se } \nu_2 > 4)\,. \tag{8.54}$$

La moda vale $\nu_2(\nu_1-2)/[\nu_1\,(\nu_2+2)]$ se $\nu_1 > 2$, altrimenti essa vale 0.

La proprietà che rende interessante la distribuzione di Fisher è la seguente: se due variabili $chi^2_{\nu_1}$ e $\chi^2_{\nu_2}$ sono indipendenti, allora il rapporto fra le due variabili, ciascuna divisa per il proprio numero di gradi di libertà, è distribuita secondo una F:

$$\frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2} \sim F(\nu_1,\nu_2)\,. \tag{8.55}$$

8.15 Ricapitolando

- In nuovo concetto che interviene nella trattazione delle variabili casuali continue è quello della funzione densità di probabilità, $f(x)$, probabilità per unità di X. La probabilità che la variabile casuale sia compresa nell'intervallo compreso fra a e b è data dall'integrale di $f(x)$ in dx nell'intervallo (area sotto la curva).

- Quello che gioca il ruolo di probabilità è l'elemento infinitesimo di probabilità $f(x)\,\mathrm{d}x$. Dalle usuali regole della probabilità seguono le proprietà di $f(x)$, di $F(x)$ e le formule per il calcolo dei valori attesi.

- La distribuzione uniforme e le distribuzioni triangolari rappresentano sia semplici casi didattici che utili strumenti per modellizzare l'ignoranza sul valore di grandezze fisiche.

- La distribuzione esponenziale, introdotta come primo esempio di distribuzione definita su un intervallo infinito, è legata al tempo di attesa di eventi descritti dal processo di Poisson (vedi capitolo successivo).

- La distribuzione di Gauss o normale è indubbiamente la più importante e la più conosciuta fra le distribuzioni per la varietà di fenomeni che descrive. Questo è dovuto al fatto che molte altre distribuzioni tendono ad essa sotto ipotesi che si presentano frequentemente in casi pratici (vedi teorema del limite centrale nel capitolo ***).

- L'intenso uso della distribuzione uniforme nei programmi di simulazione al calcolatore ha suggerito di includere una parentesi sui vari metodi per costruire variabili casuali descritte da una funzione arbitraria a partire dalla uniforme fra 0 e 1.

- Il processo di Poisson, incontrato nel capitolo precedente, è stato ripreso e studiato in dettaglio dal punto di vista della variabile tempo di attesa. Esso dà origine alla distribuzione gamma, che si riduce alla esponenziale se si prende in considerazione un solo evento.

  Confrontando con quanto visto a proposito delle variabili discrete, abbiamo visto l'analogia fra esponenziale e geometrica e, più in generale, fra gamma e Pascal.

8.16 Problemi

1. Una variabile casuale, definita nell'intervallo [0,4] segue una distribuzione uniforme. Quanto vale la probabilità che su 3 realizzazioni della variabile casuale nessuna sia maggiore di π?

2. Una variabile casuale, definita fra 0 e 1 ha una densità di probabilità proporzionale al valore della variabile. Trovare la forma della distribuzione, il valore medio e la deviazione standard.

3. Dalla definizione di variabile normale standardizzata $Z = (X - \mu)/\sigma$ e dalle proprietà del valore atteso e della varianza dimostrare che $\mathrm{E}(Z) = 0$ e $\mathrm{Var}(Z) = 1$.

4. Una variabile casuale continua ha la seguente funzione di ripartizione $F(x) = 1/\sqrt{x}$. Sapendo che il valore minimo che la variabile può assumere è 1, quanto vale il valore massimo? Quanto vale il valor medio e la deviazione standard della distribuzione?

5. Risolvere il problema precedente nel caso in cui sia $F(x) = \sqrt{x} + k$.

6. Una variabile casuale segue una distribuzione normale di parametri $\mu = 1$ e $\sigma = 2$. Calcolare: $P(X < 0)$; $P(-2 < X < -1)$; $P(0 < X < 20)$; $P(1 \le X \le 3)$; $P(1 \le X \le 5)$; $P(-5 \le X \le 7)$.

7. Una ditta produce resistori da $100\,\Omega$. Da campionamenti effettuati in periodi diversi e sotto diverse condizioni di lavoro delle macchine e di forniture dei materiali risulta che i valori delle resistenza dei singoli campioni sono centrati intorno al valore nominale con distribuzione gaussiana di deviazione standard $0.20\,\Omega$. Qual'è la probabilità che un valore di resistenza si discosti di più dell'1% dal valore atteso?

8. Calcolare media e deviazione standard di una distribuzione esponenziale mediante la funzione generatrice dei momenti.

9. In un paese il reddito per ogni persona è pari a 10000 \$ l'anno con una deviazione standard di 3000 \$. Qual'è la percentuale di persone che hanno un reddito superiore a 30000\$?

10. Un numero aleatorio è descritto da una gaussiana centrata intorno a 30. Sapendo che la probabilità che la variabile ecceda il valore di 50 è del $5\,\%$, quanto vale la probabilità di avere un valore inferiore a 20?

11. La radioattività dovuta ad un campione di di materiale produce su un opportuno strumento, in grado di rivelare le particelle dovute a ciascun decadimento con il $100\,\%$ di probabilità, 15 conteggi al minuto. Sapendo che il campione contiene 10^{20} atomi, determinare la vita media del nucleo di tale sostanza.

12. Supponiamo che i protoni e i nucleoni[6] abbiano una vita media di 10 miliardi di anni, ovvero dell'ordine di grandezza dell'età dell'Universo. Stimare il valore della radiattività, misurata in conteggi al secondo, emessa da una persona di 70 kg. (Si ricorda che protoni circa la stessa massa, pari a $1.7 \cdot 10^{-27}\,Kg$, e che la massa dovuta agli elettroni è trascurabile.).

[6] I neutroni hanno una vita media di circa 20 minuti quando sono liberi (ovvero non all'interno di un nucleo) e decadono in protone, elettrone e (anti-)neutrino. Questo è solo modo di decadimento osservato. Altri tipi di processi legati a nuove teorie unificatrici delle forze fondamentali, permetterebbero sia al protone e al neutrone di decadere - anche all'interno dei nuclei - in modi più complessi, ma con vite medie non inferiori a 10^{23} volte di quelle supposte in questo esercizio.

schematicamente e magari si è ancora disposti a sbagliarsi (“forse ho caldo perché sono agitato”). Nel caso di lettura pari a 5 °C il termometro viene dichiarato “oggettivamente” rotto, in quanto tutti, freddolosi e calorosi, esperti o principianti, converranno che “non può essere”.

28. Obiezioni del tutto lecite. Il fatto è che nella ricerca di frontiera bisogna basare le conclusioni sui rivelatori che si hanno e che già rappresentano quanto di più sofisticato si possa costruire. Vedremo che le cose si semplificano enormemente nei “casi tranquilli” di laboratorio. In ogni caso, questo modo di procedere è quello degli sperimentali esperti, i quali, in base alle loro aspettative e alla conoscenze degli strumenti, “riprovano”, “ricalibrano”, “cambiano strumento”, etc.

29. Il padre è eterozigote (Rr) in quanto ha avuto almeno un figlio che non arrotola la lingua (rr) da una donna (rr).
   La madre del bambino può essere (RR) (Rr) o (rr). Poiché in quella regione i due alleli R e r sono equiprobabili, le probabilità iniziali per la madre sono $P_\circ(RR) = P_\circ(rr) = \frac{1}{2}P_\circ(Rr) = 1/4$ (consideriamo indistinguibili (Rr) e (rR)). Applicando le leggi di Mendel otteniamo le probabilità che il bambino sia *roller* condizionate dal genotipo della madre, e da queste le probabilità finali dei genotipi della madre:

$$\begin{aligned}
P(\text{roller}\,|\,RR) &= 1;\\
P(\text{roller}\,|\,Rr) &= 3/4;\\
P(\text{roller}\,|\,rr) &= 1/2;\\
P(RR\,|\,\text{roller}) &= \frac{P(\text{roller}\,|\,RR)\cdot P_\circ(RR)}{P(\text{roller}\,|\,RR)\cdot P_\circ(RR)+\ldots}\\
&= \frac{1\times\frac{1}{4}}{1\times\frac{1}{4}+\frac{3}{4}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{4}} = \frac{1}{3}\\
P(Rr\,|\,\text{roller}) &= \frac{1}{2}\\
P(rr\,|\,\text{roller}) &= \frac{1}{6}
\end{aligned}$$

   Ne segue una probabilità di $5/6$ che la mamma fosse stata roller.

30. I dati del problema indicano che il nonno può avere i geni (Rr) o (RR) con probabilità che stanno fra di loro in proporzione 2:1. La nonna può essere invece (RR) (Rr) o (rr), con proporzioni 1:2:1.
   Delle 48 possibilità con cui la madre del bambino può ereditare i geni dei genitori ce ne sono 16 che danno luogo a (RR), 8 a (rr) e 24 a (Rr).
   Le probabilità iniziali dei geni della madre del bambino diventano ora: $P_\circ(RR) = 1/3$, $P_\circ(Rr) = 1/2$ e $P_\circ(rr) = 1/6$.
   Seguono: $P(RR|roller) = 0.42$, $P(Rr|roller) = 0.47$ e $P(rr|roller) = 0.11$.
   Come ci si poteva attendere, sapere che la madre del bambino avesse un genitore roller aumenta le probabilità che anche lei lo fosse (dall’83 % all’89 %). Come cambierebbe ulteriormente la probabilità, se in quella regione la frequenza degli alleli di tipo R fosse il doppio di quelli del tipo r?

31. Dopo il primo esperimento la probabilità sale al 50 % (lo scetticismo della commissione comincia a vacillare). L’osservazione del prestigiatore riporta essenzialmente la probabilità al livello iniziale (1.01×10^{-8}, ad essere precisi…).
   Il rifiuto a presentarsi manda definivamente a zero la probabilità, se la commissione assume, ragionevolmente, che P(“non si presenta” | “imbroglione”) $= 1$, mentre P(“non si presenta” | “onesto”) $= 0$.

32. Ne risulterebbe che: una persona dichiarata positiva sia praticamente infetta; una persona operata di chirurgia plastica sia un artista; chi riesce a ingannare una commissione di scienziati sia un sensitivo.

Capitolo 6

1. La variabile casuale può assumere con uguale probabilità i valori x_i = -1, -2, -3, 2, 6, 10: $f(x_i) = 1/6$.

2. $\mathrm{E}(X) = ***$, $\sigma(X) = ***$.
3. $f(x_4) < 0$;
   $\sum_i f(y_i) \neq 1$.
4. $f(x)$, $x = 1, 2, \ldots, 6$: $\{0.1, 0.2, 0.3, 0.3, 0.05, 0.05\}$. Ne seguono i sette valori di probabilità richiesti: 0.2; 0; 0.6; 0.15; 0.6; 0.9; 0.6.
5. Distribuzione discreta uniforme: $f(x) = 1/10$;
   $\overline{x} = \sum_i x_i\, f(x_i) = 5.5$;
   $\overline{x^2} = \sum_i x_i^2\, f(x_i) = 38.5$;
   $\sigma = 2.87$.
6. $2 \leq X_1 \leq 12$: $f(2) = 1/36$, $f(3) = 2/36$, $f(4) = 3/36$, $f(5) = 4/36$, $f(6) = 5/36$, $f(7) = 6/36$, $f(8) = 5/36$, $f(9) = 4/36$, $f(10) = 3/36$, $f(11) = 2/36$, $f(12) = 1/36$; $F(2) = 1/36$, $F(3) = 3/36$, $F(4) = 6/36$, $F(5) = 10/36$, $F(6) = 15/36$, $F(7) = 21/36$, $F(8) = 26/36$, $F(9) = 30/36$, $F(10) = 33/36$, $F(11) = 35/36$, $F(12) = 36/36$.
   $\mathrm{E}(X_1) = 7$, $\mathrm{Var}(X_1) = 5.75$; $\sigma_{X_1} = 2.40$
   $0 \leq X_2 \leq 5$: $f(0) = 6/36$, $f(1) = 10/36$, $f(2) = 8/36$, $f(3) = 6/36$, $f(4) = 4/36$, $f(5) = 2/36$. $F(0) = 6/36$, $F(1) = 16/36$, $F(2) = 24/36$, $F(3) = 30/36$, $F(4) = 34/36$, $F(5) = 34/36$.
   $\mathrm{E}(X_2) = 1.94$, $\mathrm{Var}(X_2) = 1.55$, $\sigma_{X_2} = 1.25$.
7. La prima tensione può valere 2.25, 2.26, …2.34 V con distribuzione uniforme di probabilità. Analogalmente, per la seconda si ha una distribuzione uniforme fra 2.05 e 2.14. Previsione e incertezza di previsione delle due grandezza in Volt sono rispettivamente[1]: $\mathrm{E}(V_1) = 2.295$, $\sigma(V_1) = 0.029$; $\mathrm{E}(V_1) = 2.095$, $\sigma(V_2) = 0.029$. Costruendo una tabellina di tutte le possibilità, in analogia con il caso dei dadi di figura 6.2, si vede che la differenza dei valori $X = V_1 - V_2$ può assumere valori distanziati 0.01 V e compresi fra 0.11 e 0.29 V, con probabilità massima al centro $P(X = 0.20) = f(0.20) = 1/10$ e decrescente verso il valori estremi, rappresentabile matematicamente da:
   $$f(x) = 0.1 - |x - 0.20|\,.$$
   Facendo i conti si ottiene che la differenza di tensione ha una previsione di 0.20 V con una incertezza di 0.04 V (si noti come l'incertezza sia al di sotto della cifra meno significativa della lettura!)
8. Con i simboli precedentemente introdotti: $P(X = 0) = f(0) = 0$.
9. La differenza di temperatura (in °C) è una variabile aleatoria con distribuzione di probabilità: $f(2) = 0.0225$, $f(3) = 0.21$, $f(4) = 0.535$, $f(5) = 0.21$, $f(6) = 0.0225$. Ne deriva un valore atteso di 4.0 °C, con incertezza standard di 0.8 °C.
10. 43.5, 64.3, 82.0, 94.3 e 99.7 %. Ovviamente la probabilità che il numero si verifichi alla 101 estrazione è sempre 1/18.
11. Previsione 265720.5 giocate, con una incertezza uguale alla previsione stessa.
12. Lavorando con la varianza: $\mathrm{d}\sigma^2/\mathrm{d} = 0$, segue $1 - 2\,p = 0$, da cui $p = 1/2$.
13. Chiamando ϵ e δ due numeri reali positivi e minori di 1/2:
    a) Se $f(0) = 1/2 - \epsilon$ e $f(1) = 1/2 + \epsilon$ segue: $\mathrm{E}(X) = 1/2 + \epsilon$, $Var[X] = 1/4 - \epsilon^2 \leq 1/4$;
    b) Se $f(0) = 1/2 - \epsilon$, $f(0+\delta) = \epsilon$, $f(1-\delta) = \epsilon$ e $f(1) = 1/2 - \epsilon$ segue: $\mathrm{E}(X) = 1/2$, $Var[X] = 1/4 - 2\epsilon(\delta - \delta^2) \leq 1/4$
14. No. Se a e b sono gli estremi dell'intervallo, la deviazione standard non può eccedere $\sigma_{max} = (b - a)/2$ (vedi problema precedente).
15. $\mathrm{E}(X) = 6$, $\sigma(X) = 5.5$.
16. $P(X \leq 3) = 91/216 = 42.1\,\%$; in media deve attendere 6 turni prima di poter rientrare in gioco.

[1] Le piccolissime differenze rispetto alle previsioni intuitive di 2.3 e 2.1 V sono dovute alla rozza schematizzazione dell'incertezza, curabili modellizzando meglio il meccanismo di arrotondamento o passando alle variabili continue.

17. $1/1024 = 0.098\%$.
18. Indicando con G la vincita netta: $P(G = -102\,300\,000) = 1.3 \times 10^{-3}$; $P(G = 100\,000) = 0.9987$. $\mathrm{E}(G) = -33\,000$; $\sigma(G) = 3\,690\,000$.
19. Alla prossima versione...

Capitolo 7

Mettere in tutti i problemi esplicitamente fa funzione utilizzata (es $f(0\,|\,\mathcal{P}_3)$).

1. $f(2\,|\,\mathcal{B}_{6,1/5}) = 24.6\,\%$; $1 - f(0\,|\,\mathcal{B}_{6,1/5}) = 73.8\,\%$.
2. $f(3\,|\,\mathcal{B}_{8,0.246}) = 15.3\,\%$.
3. Indicando con $p_2 = f(2\,|\,\mathcal{B}_{2,1/5}) = 0.04$, siamo interessati a $X \sim \mathcal{G}_{p_2}$. Quindi: 25 ± 25 (± 24.5, ad essere precisi). .
4. $p = 1/4$, $n = 16$: $P(X < 5) = F(4\,|\,\mathcal{B}_{16,1/4}) = 63.0\,\%$.
5. $f(3\,|\,\mathcal{B}_{10,1/2}) = 0.172$. (Il fatto che si tratti di diversi tipi di monete è irrilevante.)
6. Il totale di monete lanciate è pari a 10, ciascuna ha probabilità $1/2$ di mostrare testa, ma gli eventi non sono indipendenti. Si vede verifica facilmente che $P(X = 5) = 1$ e $P(X \neq 5) = 0$. La distribuzione non è binomiale. Chiamando n il numero totale di monete lanciate, si ha $f(x = n/2) = 1$ e $f(x \neq n/2) = 0$, da cui $\mathrm{E}(X) = n/2$, $\mathrm{Var}(X) = 0$.
7. Il processo non è binomiale in quanto ogni estrazione ha diversa probabilità (1/2, 1/6, 2/3, 1/37 e 1/4). $P(X = 0) = 0.101 = 15/148$, $P(X = 1) = 0.361$: $F(1) = 0.462$. $P(X = 5) = 3.75 \cdot 10^{-4} = 1/2664$.
8. $1 - F(7\,|\,\mathcal{B}_{10,0.6}) = 16.7\,\%$ Per sveltire i conti è preferibile considerare il processo complementare con p e q scambiate: $F(2\,|\,\mathcal{B}_{10,0.4})) = 16.7\%$
9. Se il giocatore A si aggiudica meno di 2 delle residue 4 partite allora ha vinto il giocatore B: $P(B) = F_A(1) F(1\,|\,\mathcal{B}_{4,1/4}) = f(0) + f(1) = 5/16$; $P(A) = 11/16$.
10. $P(H_i\,|\,I_0) = \{0.03,\ 0.16,\ 0.31,\ 0.31, 0.16,\ 0.03\}$.
    $H_1 = N$: ***
    $H_2 = N, B$: ***
    $H_3 = N, B, N$: ***
11. $\mathcal{B}_{6,1/2}$: 0.016, 0.094, 0.234, 0.313, 0.234, 0.094, 0.016.
    $\mathcal{P}_{0.874}$: 0.417, 0.365, 0.159, 0.046, 0.010, 0.0018.
12. $P(X > 0\,|\,\mathcal{P}_{0.417}) = 1 - f(0\,|\,\mathcal{P}_{0.417}) = 34.1\,\%$.
13. $f(5\,|\,\mathcal{B}_{10,0.659}) = 14.4\,\%$.
14. $f(0\,|\,\mathcal{P}_\lambda) = 0.05 \Rightarrow \lambda = 3.00$. Da cui $P(X \geq 4\,|\,\mathcal{P}_\lambda) = 1 - F(3\,|\,\mathcal{P}_\lambda) = 35.3\,\%$.
15. $r = 1.67$ cont/s. In un secondo $\lambda = 1.67$, da cui $p = 1.67 \times 10^{-22}$.
16. $f(0\,|\,\mathcal{P}_{1.2}) = 30\,\%$;
    $F(10\,|\,\mathcal{P}_{18}) = 3.0\,\%$.
17. $f(0\,|\,\mathcal{P}_\lambda) = 0.0153 \rightarrow \lambda = 4.18 \rightarrow v = \sigma/\mu = 1/\sqrt{\lambda} = 0.489$.
18. $1 - F(2\,|\,\mathcal{P}_4) = 76.2\,\%$.
19. $1 - F(10\,|\,\mathcal{P}_{6.5}) = 6.7\,\%$.
20. Chiamando X_i le opportune variabili casuali di interesse in ciascuna delle domande:
    (a) $P(X_1 \leq 10) = F(10\,|\,\mathcal{P}_{20}) = 0.011$; $P(X_1 = 10) = f(10\,|\,\mathcal{P}_{20}) = 0.0058$
    (b) $P(X_2 \geq 2) = 1 - F(1\,|\,\mathcal{B}_{10,0.15}) = 0.456$; $P(X_2 \geq 2) = 1 - F(1\,|\,\mathcal{B}_{20,0.15}) = 0.824$ (si noti la dipendenza non lineare da n: si provi anche $n = 30$).
    (c) $P(X_3 \neq 0) = 1 - f(0\,|\,\mathcal{P}_{3/4}) = 0.528$;
    (d) $P(X_4 \geq 2) = 1 - F(1\,|\,\mathcal{B}_{5,0.528}) = 0.846$.
21. $v = 0.01 \Rightarrow \lambda = 10\,000 \Rightarrow t = 2\,000$ s.

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 $P(X \geq 6\,|\,\mathcal{P}_{0.61}) = 4.2 \times 10^{-5}$!).
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. $\mathrm{E}(T) = \mathrm{E}(C) = n/2$; $\mathrm{Var}(T) = \mathrm{Var}(C) = n/4$; $\mathrm{Cov}(T, C) = -n/4$;
$\mathrm{E}(T - C) = 0$; $\mathrm{Var}(T - C) = n$; $\sigma(T - C) = \sqrt{n}$.