B Concetti di probabilità

SOMMARIO

Obiettivi d'apprendimento

Lo studio di questo capitolo dovrebbe insegnarvi come:

1. Spiegare e illustrare con un esempio la differenza fra una variabile casuale e i valori che essa può assumere.
2. Spiegare e illustrare con alcuni esempi la differenza fra variabili casuali discrete e continue.
3. Spiegare e illustrare con alcuni esempi le caratteristiche delle funzioni di probabilità (*fdp*) di variabili casuali discrete e delle funzioni di densità (*fdd*) di variabili casuali continue.
4. Calcolare probabilità di eventi utilizzando funzioni di probabilità per variabili casuali discrete e di densità per variabili casuali continue.
5. Illustrare geometricamente e algebricamente, usando un integrale, il calcolo di probabilità usando la *fdd* di una variabile casuale continua.
6. Usare le definizioni di valore atteso di una variabile casuale discreta o continua per calcolare il valore atteso di una funzione $g(x)$, data la funzione di probabilità o densità $f(x)$.
7. Definire la varianza di una variabile casuale e spiegare in che senso la dispersione dei suoi valori è maggiore se la varianza è elevata.
8. Usare la *fdd* congiunta di due variabili casuali continue per calcolare probabilità di eventi congiunti e per ricavare le *fdd* (marginali) delle singole variabili.
9. Ricavare la *fdd* condizionale di una variabile casuale dato il valore di un'altra variabile e la loro *fdd* congiunta e usare il risultato per calcolare probabilità, media e varianza condizionali.
10. Definire covarianza e correlazione fra due variabili casuali e calcolarne i valori data una funzione di probabilità congiunta.
11. Spiegare e applicare la legge dei valori attesi iterati.
12. Ricavare la distribuzione della variabile casuale $Y = g(X)$, dove $g(X)$ è una funzione monotona crescente o decrescente, a partire dalla funzione di densità $f(x)$ della variabile casuale X.
13. Generare un numero casuale da una funzione di densità $f(x)$ quando la funzione di ripartizione corrispondente $F(x)$ è invertibile.
14. Spiegare in che senso i numeri casuali generati da un computer sono realmente casuali e in che senso non lo sono.

Parole chiave

correlazione
covarianza
distribuzione chi quadro
distribuzione di Poisson
distribuzione F
distribuzione marginale
distribuzione normale
distribuzione normale standardizzata
distribuzione t
distribuzione uniforme
esperimento
fdd o *fdp*
fdd o *fdp* condizionale
fdr
funzione di densità o di probabilità
funzione di densità o di probabilità congiunta
funzione di ripartizione
gradi di libertà
indipendenza statistica
jacobiano
media
mediana
metodo di inversione
modulo
monotona
numeri pseudo-casuali
numero casuale
probabilità
probabilità condizionale
scarto quadratico medio
seme dei numeri casuali
tecnica del cambiamento di variabili
trasformazione monotona in senso stretto
valore atteso

valore atteso iterato	variabile casuale binomiale	varianza
variabile binaria	variabile casuale continua	
variabile casuale	variabile casuale discreta	

In questa appendice daremo per scontato che abbiate già una certa familiarità con i principi introduttivi di probabilità e statistica e che abbiate letto il Piccolo manuale di probabilità che precede il capitolo 2. Se non lo avete già fatto, fatelo ora.

Inizieremo questa appendice riassumendo la definizione e le proprietà di valori attesi e varianze di variabili casuali discrete, in modo che siano facilmente reperibili in caso di necessità. Successivamente svilupperemo regole simili valide nel caso di variabili casuali continue ma per le quali è necessario applicare i concetti di integrale introdotti nell'appendice A.4. Ricorderemo le proprietà di alcune importanti variabili casuali discrete e continue, compresa le distribuzioni t, chi quadro e F. Per finire introdurremo alcuni concetti relativi ai numeri casuali generati mediante un computer.

B.1. Variabili casuali discrete

In questo paragrafo riassumiamo la definizione e le principali proprietà delle variabili casuali discrete. Si veda il Piccolo manuale di probabilità per alcuni esempi e una discussione introduttiva generale.

Una **variabile casuale** è una variabile il cui valori sono ignoti fino al momento in cui vengono osservati; in altre parole, è una variabile non perfettamente osservabile. Una **variabile casuale discreta** può assumere solo un numero limitato, o numerabile, di valori. Un esempio di variabile casuale discreta è il numero di rimborsi di spese sostenute con carta di credito effettuati in ritardo durante l'anno precedente da un individuo scelto casualmente. Un caso particolare importante è quello di una variabile casuale che può assumere solo due possibili valori; per esempio, un rimborso può essere effettuato in ritardo o meno. Un esito come questo può essere caratterizzato usando una **variabile binaria**, indicata per esempio con RIT, che assume valore 1 per i rimborsi avvenuti in ritardo e 0 per quelli avvenuti entro la data stabilita. Variabili di questo tipo sono chiamate anche **variabili indicatrici** o **variabili dummy**.

Per descrivere le probabilità dei possibili esiti useremo una **funzione di probabilità** (***fdp***). La *fdp* di una variabile casuale discreta indica la probabilità che si verifichi ciascuno dei valori possibili. Per una variabile casuale discreta X il valore della funzione di probabilità $f(x)$ rappresenta la probabilità che X assuma il valore x, $f(x) = P(X = x)$. Dato che $f(x)$ è una probabilità, deve necessariamente soddisfare la condizione $0 \leq f(x) \leq 1$ e, se X può assumere n possibili valori $x_1, \ldots, x_n$, la somma delle loro probabilità deve essere pari a 1:

$$P(X = x_1) + P(X = x_2) + \ldots + P(X = x_n) = f(x_1) + f(x_2) + \ldots + f(x_n) = 1$$

La **funzione di ripartizione** (***fdr***) è un modo alternativo per rappresentare le probabilità. La *fdr* della variabile casuale X, indicata con $F(x)$, descrive la probabilità che X sia inferiore o uguale a uno specifico valore. In altre parole:

$$F(x) = P(X \leq x) \tag{B.1}$$

Due caratteristiche fondamentali di una distribuzione di probabilità sono il suo centro (posizione) e la sua ampiezza (dispersione). La **media**, o **valore atteso**, è una misura del centro; per misurare la dispersione useremo la **varianza** e la sua radice quadrata, lo **scarto quadratico medio**.

B.1.1. Valore atteso di una variabile casuale discreta

La **media** di una variabile casuale è data dalla sua **speranza matematica**. Se X è una variabile casuale discreta che assume i valori $x_1, \ldots, x_n$, la speranza matematica, detta **valore atteso**, di X è data da:

$$\mu_X = E(X) = x_1 P(X = x_1) + x_2 P(X = x_2) + \ldots + x_n P(X = x_n) \tag{B.2a}$$

Il valore atteso o media di X è la somma dei suoi possibili valori ponderati con le rispettive probabilità di essere osservati. La media è spesso indicata con il simbolo μ o μ_X. Il valore atteso rappresenta la media della variabile casuale calcolata su un numero infinito di realizzazioni dell'esperimento sottostante. Dato che la probabilità che la variabile casuale discreta X assuma valore x è data dalla *fdp* $f(x)$, $P(X = x) = f(x)$, il valore atteso in (B.2a) può essere riformulato in maniera equivalente come:

$$\begin{aligned} \mu_X &= E(X) = x_1 f(x_1) + x_2 f(x_2) + \ldots + x_n f(x_n) \\ &= \sum_{i=1}^{n} x_i f(x_i) = \sum_x x f(x) \end{aligned} \tag{B.2b}$$

Funzioni di variabili casuali sono anch'esse casuali. I loro valori attesi possono essere calcolati con passaggi simili a quelli in (B.2). Se X è una variabile casuale discreta e $g(X)$ è una sua funzione:

$$E[g(X)] = \sum_x g(x) f(x) \tag{B.3}$$

Usando la (B.3) possiamo ricavare alcune proprietà spesso molto utili. Se a è una costante:

$$E(aX) = aE(X) \tag{B.4}$$

Allo stesso modo, se a e b sono due costanti è possibile dimostrare che:

$$E(aX + b) = aE(X) + b \tag{B.5}$$

Per osservare come può essere ricavato questo risultato applichiamo la definizione fornita in (B.3) alla funzione $g(X) = aX + b$:

$$\begin{aligned} E[g(X)] &= \sum g(x) f(x) = \sum (ax + b) f(x) = \sum [axf(x) + bf(x)] \\ &= \sum [axf(x)] + \sum [bf(x)] = a \sum x f(x) + b \sum f(x) \\ &= aE(X) + b \end{aligned}$$

Nell'ultimo passaggio abbiamo usato la definizione (B.2) di $E(X)$ e il fatto che $\sum f(x) = 1$.

Se $g_1(X), g_2(X), \ldots, g_M(X)$ sono funzioni di X:

$$E[g_1(X) + g_2(X) + \ldots + g_M(X)] = E[g_1(X)] + E[g_2(X)] + \ldots + E[g_M(X)] \tag{B.6}$$

Questa proprietà vale per un numero qualsiasi di funzioni. **Il valore atteso di una somma è sempre pari alla somma dei valori attesi**.

Questa regola non è in generale valida per funzioni non lineari. In altre parole, $E[g(X)] \neq g[E(X)]$. Per esempio, $E(X^2) \neq [E(X)]^2$.

B.1.2. Varianza di una variabile casuale discreta

La **varianza** di una variabile casuale discreta X è il valore atteso di:

$$g(X) = [X - E(X)]^2$$

L'importanza della varianza di una variabile casuale sta nel fatto che essa caratterizza la scala di misura e la dispersione della sua distribuzione di probabilità. La varianza è di solito indicata con il simbolo σ^2, che si legge "sigma quadro", o σ_X^2. Da un punto di vista algebrico, se indichiamo il valore atteso $E(X)$ con μ_X:

$$\text{(B.7)} \qquad \operatorname{Var}(X) = \sigma_X^2 = E\left[(X - \mu_X)^2\right] = E\left(X^2\right) - \mu_X^2$$

La varianza di una variabile casuale è la *media* del quadrato dello scarto fra la variabile casuale X e il suo valore atteso μ_X. Quanto maggiore è la varianza di una variabile casuale, tanto più dispersi possono essere i suoi valori. La radice quadrata della varianza è chiamata **scarto quadratico medio** ed è indicata con σ o σ_X. Lo scarto quadratico medio misura la dispersione di una distribuzione e ha il vantaggio di essere espresso nella stessa unità di misura della variabile casuale.

La varianza ha diverse proprietà importanti; una particolarmente utile è la seguente. Se a e b sono due costanti:

$$\text{(B.8)} \qquad \operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$$

Questo risultato è dimostrato nel Piccolo manuale di probabilità al paragrafo P.5.4.

L'**asimmetria** e la **curtosi** sono altre due caratteristiche di una distribuzione di probabilità. La loro definizione è la seguente:

$$\text{(B.9)} \qquad \text{asimmetria} = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}$$

e

$$\text{(B.10)} \qquad \text{curtosi} = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}$$

L'asimmetria misura la mancanza di simmetria di una distribuzione. Se la distribuzione è simmetrica, il suo indice di asimmetria è nullo. Distribuzioni con coda sinistra molto pronunciata sono negativamente asimmetriche e il loro indice di asimmetria è negativo; viceversa, l'indice è positivo per distribuzioni positivamente asimmetriche, con coda destra molto pronunciata. La curtosi misura la velocità con cui le code della distribuzione vanno a zero. Una distribuzione con curtosi elevata ha code che scendono a zero lentamente e, viceversa, una distribuzione con curtosi bassa ha code che scendono a zero velocemente. Nel caso della curtosi il valore di riferimento è 3, che corrisponde alla curtosi della distribuzione *normale* che sarà discussa nel paragrafo B.3.5 di questa appendice.

B.1.3. Distribuzione congiunta, marginale e condizionale

Se X e Y sono due variabili casuali discrete, la probabilità congiunta che $X = a$ e $Y = b$ è data dalla *fdp* congiunta di X e Y, indicata con $f(x, y)$ e tale che $P(X = a, Y = b) = f(a, b)$. La somma delle probabilità congiunte è 1, $\sum_x \sum_y f(x, y) = 1$.

Data la funzione di probabilità congiunta, possiamo ricavare le distribuzioni di probabilità delle singole variabili casuali, chiamate anche **distribuzioni marginali**. Se X e Y sono due variabili casuali discrete:

(B.11) $$f_X(x) = \sum_y f(x,y) \text{ per tutti i possibili valori di } X$$

Nel caso di variabili casuali discrete, la probabilità che la variabile casuale Y assuma il valore y *sapendo* che $X = x$ si indica con $P(Y = y|X = x)$. Questa probabilità condizionale è data dalla ***fdp* condizionale** $f(y|x)$:

(B.12) $$f(y|x) = P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)} = \frac{f(x,y)}{f_X(x)}$$

Due variabili casuali sono **statisticamente indipendenti** se la probabilità condizionale che $Y = y$ sapendo che $X = x$ coincide con la probabilità non condizionale che $Y = y$. In questo caso la conoscenza del valore di X non altera la distribuzione di probabilità di Y. Se X e Y sono variabili casuali indipendenti:

(B.13) $$P(Y = y|X = x) = P(Y = y)$$

Equivalentemente, se X e Y sono indipendenti, la *fdp* condizionale di Y data $X = x$ coincide con la *fdp* non condizionale, o marginale, di Y:

(B.14) $$f(y|x) = \frac{f(x,y)}{f_X(x)} = f_Y(y)$$

Essendo valida anche l'implicazione opposta, è possibile affermare che se (B.13) o (B.14) sono vere *per qualsiasi possibile coppia di valori x e y*, allora X e Y sono statisticamente indipendenti.

Risolvendo la (B.14) rispetto alla *fdp* congiunta possiamo anche verificare che X e Y sono statisticamente indipendenti se la loro *fdp* congiunta può essere espressa come il prodotto delle *fdp* marginali:

(B.15) $$f(x,y) = f_X(x)f_Y(y)$$

Se la (B.15) è vera per qualunque coppia di valori x e y, X e Y sono statisticamente indipendenti. Questo risultato può essere esteso a un numero di variabili maggiore di 2. Se X, Y e Z sono statisticamente indipendenti, la loro funzione di probabilità congiunta può essere fattorizzata ed espressa come $f(x,y,z) = f_X(x) \cdot f_Y(y) \cdot f_Z(z)$.

B.1.4. Valori attesi di più variabili casuali

Esiste una regola simile alla (B.3) valida nel caso di più variabili casuali. Supponiamo che X e Y siano variabili casuali discrete con *fdp* congiunta $f(x,y)$. Se $g(X,Y)$ è una funzione di X e Y:

(B.16) $$E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y)$$

Usando la (B.16) è possibile dimostrare che:

(B.17) $$E(X+Y) = E(X) + E(Y)$$

Questo risultato deriva dalla (B.16) e dal definire $g(X,Y)$ come $X+Y$. In questo caso:

$$\begin{aligned}
E(X+Y) &= \sum_x \sum_y g(x,y)f(x,y) \\
&= \sum_x \sum_y (x+y)f(x,y) \\
&= \sum_x \sum_y xf(x,y) + \sum_x \sum_y yf(x,y) \\
&= \sum_x x \sum_y f(x,y) + \sum_y y \sum_x f(x,y) \\
&= \sum_x xf(x) + \sum_y yf(y) \\
&= E(X) + E(Y)
\end{aligned}$$

Per passare dalla quarta alla quinta riga abbiamo usato la (B.11) per sostituire le distribuzioni marginali di X e Y, e il fatto che l'ordine delle sommatorie è irrilevante. Usando la stessa logica possiamo dimostrare che:

(B.18) $$E(aX + bY + c) = aE(X) + bE(Y) + c$$

In generale, $E[g(X,Y)] \neq g[E(X), E(Y)]$. Per esempio, $E(XY) \neq E(X)E(Y)$. Se X e Y sono statisticamente indipendenti, tuttavia, usando (B.16) possiamo anche dimostrare che $E(XY) = E(X)E(Y)$. Per verificarlo, si ricordi che se X e Y sono indipendenti la loro *fdp* congiunta si fattorizza nel prodotto delle *fdp* marginali, $f(x,y) = f(x)f(y)$. Definendo $g(X,Y) = XY$, otteniamo:

$$\begin{aligned}
E(XY) &= E[g(X,Y)] = \sum_x \sum_y xyf(x,y) = \sum_x \sum_y xyf(x)f(y) \\
&= \sum_x xf(x) \sum_y yf(y) = E(X)E(Y)
\end{aligned}$$

Questa proprietà può essere estesa al caso di più variabili casuali indipendenti.

B.1.5. Covarianza e correlazione

Una particolare applicazione di (B.16) è la derivazione della **covarianza** fra X e Y. Definiamo la funzione $g(X,Y)$ come il prodotto degli scarti di X e Y dalla rispettiva media :

(B.19) $$g(X,Y) = (X - \mu_X)(Y - \mu_Y)$$

La covarianza è il valore atteso di (B.19):

(B.20) $$\text{Cov}(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y$$

Una covarianza σ_{XY} fra le variabili positiva implica che quando X è al di sopra della propria media anche Y tende a esserlo e che quando X è al di sotto della propria media anche Y tende a esserlo. In questo caso le variabili casuali X e Y sono definite **associate positivamente** o **direttamente**. Se $\sigma_{XY} < 0$ l'associazione è negativa o inversa. Se $\sigma_{XY} = 0$ non esiste alcun tipo di associazione, né negativa né positiva.

L'interpretazione del valore di σ_{XY} è complicata dal fatto che X e Y potrebbero essere espresse in unità di misura diverse. Per eliminare le unità di misura possiamo scalare la covarianza con gli scarti quadratici medi, definendo così la **correlazione** fra X e Y:

$$\rho = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{B.21}$$

Come nel caso della covarianza, la correlazione ρ fra due variabili casuali misura il loro grado di associazione *lineare*. A differenza della covarianza, tuttavia, la correlazione deve stare fra -1 e 1. La correlazione fra X e Y è 1 se fra le due variabili esiste una relazione lineare positiva perfetta e a -1 se esiste una relazione lineare negativa, o inversa, perfetta. Se fra X e Y non esiste alcuna associazione *lineare*, $\text{Cov}(X,Y) = 0$ e $\rho = 0$. Per valori diversi della correlazione, la grandezza del valore assoluto $|\rho|$ indica la "forza" dell'associazione lineare fra i valori delle variabili casuali.

Se X e Y sono variabili casuali indipendenti, la loro covarianza e la loro correlazione sono nulle. L'affermazione inversa, tuttavia, *non* è vera. Due variabili casuali X e Y indipendenti hanno covarianza nulla perché fra loro non esiste alcuna associazione lineare. Il fatto che la covarianza e la correlazione fra due variabili sia nulla *non significa* che esse debbano necessariamente essere indipendenti. Potrebbero esistere associazioni non lineari più complesse, per esempio $X^2 + Y^2 = 1$.

Nella (B.17) abbiamo derivato il valore atteso di una somma di variabili casuali. Esistono proprietà simili per le varianze. Se a e b sono costanti:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y) \tag{B.22}$$

Per verificare questo risultato è conveniente definire una nuova variabile casuale discreta $Z = aX + bY$. Questa variabile casuale ha valore atteso:

$$\mu_Z = E(Z) = E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y$$

La varianza di Z è data da:

$$\begin{aligned}
\text{Var}(Z) &= E[(Z - \mu_Z)^2] = E\left\{[(aX + bY) - (a\mu_X + b\mu_Y)]^2\right\} \\
&= E\left\{[(aX - a\mu_X) + (bY - b\mu_Y)]^2\right\} \\
&= E\left\{[a(X - \mu_X) + b(Y - \mu_Y)]^2\right\} \\
&= E[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)] \\
&= E[a^2(X - \mu_X)^2] + E[b^2(Y - \mu_Y)^2] + E[2ab(X - \mu_X)(Y - \mu_Y)] \\
&= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)
\end{aligned}$$

Questa regola può essere estesa al caso di più variabili casuali. Per esempio, se X, Y e Z sono tre variabili casuali:
(B.23)

$$\begin{aligned}
\text{Var}(aX + bY + cZ) = {} & a^2\text{Var}(X) + b^2\text{Var}(Y) + c^2\text{Var}(Z) + 2ab\text{Cov}(X,Y) \\
& + 2bc\text{Cov}(Y,Z) + 2ac\text{Cov}(X,Z)
\end{aligned}$$

B.1.6. Valori attesi condizionali

Se X e Y sono due variabili casuali con distribuzione di probabilità congiunta $f(x, y)$, la distribuzione di probabilità di Y condizionale a X è data da $f(y|x)$. Possiamo usare questa *fdp* condizionale per calcolare la **media condizionale** di Y data X; in altre parole, il valore atteso di Y se $X = x$. Il valore atteso condizionale $E(Y|X = x)$ è il valore medio di Y sapendo che X assume valore x. Nel caso discreto $E(Y|X = x)$ è dato da:

$$E(Y|X = x) = \sum_y yP(Y = y|X = x) = \sum_y yf(y|x) \tag{B.24}$$

Allo stesso modo possiamo definire la **varianza condizionale** di Y data X, la varianza della distribuzione condizionale di Y sapendo che $X = x$. Nel caso discreto questa varianza è data da:

$$\text{Var}(Y|X = x) = \sum_y [y - E(Y|X = x)]^2 f(y|x) \tag{B.25}$$

B.1.7. Valori attesi iterati

La **legge dei valori attesi iterati** afferma che il valore atteso di Y è pari al valore atteso del valore atteso di Y data X. In altre parole:

$$E(Y) = E_X[E(Y|X)] \tag{B.26}$$

Il significato di questa formula è spiegato dalla dimostrazione seguente, valida nel caso discreto. Per svilupparla useremo due proprietà delle distribuzioni di probabilità. In primo luogo, la *fdp* marginale di Y è $f(y) = \sum_x f(x, y)$ e, secondo, la *fdp* congiunta di X e Y può essere espressa come $f(x, y) = f(y|x)f(x)$.

$$\begin{aligned}
E(Y) &= \sum_y yf(y) = \sum_y y \left[\sum_x f(x, y)\right] \\
&= \sum_y y \left[\sum_x f(y|x)f(x)\right] \\
&= \sum_x \left[\sum_y yf(y|x)\right] f(x) \qquad \text{[invertendo l'ordine delle sommatorie]} \\
&= \sum_x E(Y|X = x)f(x) \\
&= E_X[E(Y|X)]
\end{aligned}$$

Nell'ultima espressione il simbolo $E_X[\ ]$ significa che il valore atteso del termine fra parentesi quadra è calcolato considerando X casuale. Il valore atteso di Y può dunque essere ottenuto trovando il suo valore atteso condizionale a X e calcolandone il valore atteso rispetto a X.

Nello stesso modo possiamo dimostrare altre due proprietà:

$$E(XY) = E_X[XE(Y|X)] \tag{B.27}$$

e

$$\text{Cov}(X, Y) = E_X[(X - \mu_X)E(Y|X)] \tag{B.28}$$

B.2. Variabili casuali continue

Le variabili casuali continue possono assumere qualsiasi valore all'interno di un intervallo. In economia variabili come il reddito e i prezzi di mercato sono considerate variabili casuali continue. Nella figura P.2 del Piccolo manuale di probabilità abbiamo illustrato la funzione di densità di una variabile continua che assume valori da zero all'infinito, $x \geq 0$. Dato che le variabili continue possono assumere valori all'interno di un insieme non numerabile, la probabilità di ognuno dei valori che possono essere osservati in un esperimento casuale è zero. Per esempio, $P(X = 100) = 0$ o $P(X = 200) = 0$. Qualsiasi affermazione a proposito di probabilità relative a variabili casuali continue ha senso solo se si riferisce a esiti rappresentabili come intervalli di valori. Possiamo per esempio chiederci: qual è la probabilità che X assuma un valore fra 100 e 200? Questi concetti sono stati introdotti nei paragrafi P.1 e P.2 del Piccolo manuale di probabilità. In quella sede abbiamo osservato che probabilità come queste corrispondono ad aree al di sotto della curva che rappresenta la funzione di densità. Se questi concetti non vi sono del tutto familiari, è opportuno rileggere quei paragrafi. Ciò che non abbiamo discusso nel Piccolo manuale di probabilità è come sono calcolate esattamente queste probabilità. Questa parte della discussione è stata rinviata fino a ora perché per dare una risposta sono necessari gli strumenti del calcolo integrale.

In questo paragrafo discuteremo come lavorare con variabili casuali continue. L'**interpretazione** di probabilità, valori attesi e varianze resta immutata rispetto a quanto avete appreso per il caso di variabili casuali discrete. Ciò che cambia sono i passaggi algebrici – i segni di sommatoria diventano integrali e per abituarsi a questo cambiamento serve un po' di tempo. Se non lo avete ancora fatto, è il momento di rileggere la discussione degli integrali nell'appendice A.4.

B.2.1. Calcolare probabilità

Se X è una variabile casuale continua, la sua funzione di densità (*fdd*) $f(x)$ deve soddisfare alcune condizioni:

(B.29) $$f(x) \geq 0$$

(B.30) $$\int_{-\infty}^{\infty} f(x)dx = 1$$

(B.31a) $$P(a \leq X \leq b) = \int_a^b f(x)dx$$

La proprietà (B.29) afferma che la *fdd* non può assumere valori negativi. La proprietà (B.30) stabilisce che l'area al di sotto della *fdd*, che corrisponde alla probabilità che X assuma un valore compreso fra $-\infty$ e ∞, sia pari a 1. La proprietà (B.31a) afferma che la probabilità che X cada all'interno dell'intervallo $[a, b]$ è l'area sotto la curva $f(x)$ compresa fra questi due valori. Dato che un singolo punto ha probabilità nulla, è anche vero che:

(B.31b)
$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f(x)dx$$

La **funzione di ripartizione** (*fdr*) di una variabile casuale continua è data da $F(x) = P(X \le x)$. Usando la *fdr* possiamo calcolare:

$$P(X \le a) = \int_{-\infty}^{a} f(x)dx = F(a) \tag{B.32a}$$

La *fdr* è ricavata integrando la *fdd*. Dato che l'integrale è un'"antiderivata", possiamo ottenere la *fdd* differenziando la *fdr* $F(x)$. In altre parole:

$$f(x) = \frac{dF(x)}{dx} = F'(x) \tag{B.32b}$$

Il concetto di *fdr* è utile per molte ragioni. Una di queste è che tutti i software econometrici offrono semplici comandi per calcolare le *fdr* di molte variabili casuali, grazie ai quali è possibile ottenere facilmente il valore di una probabilità.

Esempio B.1
Indichiamo con X una variabile casuale continua con *fdd* $f(x) = 2(1-x)$ per $0 \le x \le 1$. Questa *fdd* è rappresentata nella figura B.1.

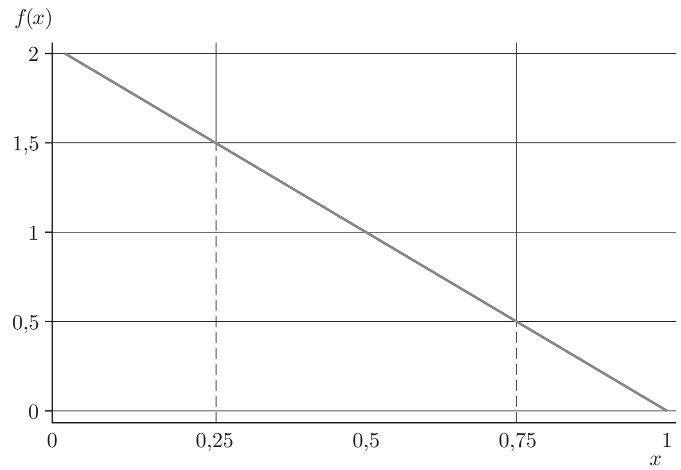

Figura B.1
Funzione di densità $f(x) = 2(1-x)$.

La proprietà (B.29) è valida per valori di x compresi fra 0 e 1. Anche la proprietà (B.30) è valida, dato che:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} 2(1-x)dx = \int_{0}^{1} 2dx - \int_{0}^{1} 2xdx = 2x\Big|_{0}^{1} - x^2\Big|_{0}^{1} = 2 - 1 = 1$$

Usando la figura B.1 possiamo calcolare $P(\frac{1}{4} \le X \le \frac{3}{4}) = \frac{1}{2}$ usando un semplice ragionamento geometrico. Usando l'integrazione arriviamo allo stesso risultato:

$$\begin{aligned} P(\tfrac{1}{4} \le X \le \tfrac{3}{4}) &= \int_{1/4}^{3/4} f(x)dx = \int_{1/4}^{3/4} 2(1-x)dx \\ &= \int_{1/4}^{3/4} 2dx - \int_{1/4}^{3/4} 2xdx = 2x\Big|_{1/4}^{3/4} - x^2\Big|_{1/4}^{3/4} = 1 - \left(\frac{9}{16} - \frac{1}{16}\right) = \frac{1}{2} \end{aligned}$$

La funzione di ripartizione è data da $F(x) = 2x - x^2$ per valori di x nell'intervallo $[0, 1]$. La probabilità richiesta può dunque essere calcolata anche come:

$$P(\tfrac{1}{4} \le X \le \tfrac{3}{4}) = F(\tfrac{3}{4}) - F(\tfrac{1}{4})$$

Esempio B.2
Indichiamo con X una variabile casuale continua con *fdd* $f(x) = 3x^2$ per x appartenente all'intervallo $[0, 1]$. Le proprietà (B.29) e (B.30) sono valide. Dato che la *fdd* è quadratica non possiamo usare argomenti geometrici semplici per calcolare $P(\frac{1}{4} \leq X \leq \frac{3}{4})$; possiamo però usare l'integrazione, ottenendo:

$$P(\tfrac{1}{4} \leq X \leq \tfrac{3}{4}) = \int_{1/4}^{3/4} f(x)dx = \int_{1/4}^{3/4} 3x^2 dx = x^3 \Big|_{1/4}^{3/4} = \frac{27}{64} - \frac{1}{64} = \frac{13}{32}$$

B.2.2. Proprietà di variabili casuali continue

Se X è una variabile casuale continua con funzione di densità $f(x)$, il suo valore atteso è dato da:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x)dx \tag{B.33}$$

Confrontate quest'espressione con quella del valore atteso di una variabile casuale discreta definito dalla (B.2): la sommatoria è stata sostituita da un integrale. L'interpretazione di $E(X)$ è esattamente la stessa vista nel caso discreto: il valore atteso è la media dei valori di X calcolata su un numero infinito di sue realizzazioni.

Esempio B.1 (continua)
Il valore atteso della variabile casuale dell'esempio B.1 è:

$$\int_{-\infty}^{\infty} x f(x)dx = \int_0^1 x \cdot 2(1-x)dx = \int_0^1 (2x - 2x^2)dx = x^2\Big|_0^1 - \frac{2}{3}x^3\Big|_0^1 = 1 - \frac{2}{3} = \frac{1}{3}$$

La varianza di una variabile casuale X è definita come $\sigma_X^2 = E[(X - \mu_X)^2]$. Questa definizione vale sia per variabili casuali discrete sia per variabili continue. Per calcolare la varianza usiamo una versione della (B.3) adatta al caso di variabili continue:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)dx \tag{B.34}$$

Definendo $g(x) = (X - \mu_X)^2$, otteniamo:

$$\begin{aligned}
\sigma_X^2 &= E\left[(X - \mu_X)^2\right] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx \\
&= \int_{-\infty}^{\infty} (x^2 + \mu_X^2 - 2x\mu_X) f(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx + \mu_X^2 \int_{-\infty}^{\infty} f(x)f(x) - 2\mu_X \int_{-\infty}^{\infty} x f(x)dx \\
&= E(X^2) + \mu_X^2 - 2\mu_X^2 \\
&= E(X^2) - \mu_X^2
\end{aligned} \tag{B.35}$$

Per passare dalla terza alla quarta riga abbiamo usato la proprietà (B.30) e la definizione (B.33) di valore atteso. Il risultato finale è $\sigma_X^2 = E[(X - \mu_X)^2] = E(X^2) - \mu_X^2$, come nel caso discreto.

Per calcolare la varianza della variabile casuale descritta nell'esempio B.1 calcoliamo innanzitutto:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 x^2 \cdot 2(1-x)dx = \int_0^1 (2x^2 - 2x^3)dx \\ &= \frac{2}{3}x^3\Big|_0^1 - \frac{2}{4}x^4\Big|_0^1 = \frac{2}{3} - \frac{1}{2} = \frac{1}{6} \end{aligned}$$

Di conseguenza:

$$\operatorname{Var}(X) = \sigma_X^2 = E(X^2) - \mu_X^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$$

B.2.3. Distribuzioni congiunte, marginali e condizionali

Per calcolare probabilità relative a più di una variabile casuale continua abbiamo bisogno della **funzione di densità congiunta** delle variabili in questione. Per esempio, consideriamo le due variabili casuali continue U (la disoccupazione) e P (il tasso d'inflazione). Supponiamo che la *fdd* congiunta sia quella rappresentata nella figura B.2.

Una *fdd* congiunta è una superficie e le probabilità sono volumi al di sotto della superficie. Se le due variabili casuali sono non negative, potremmo chiederci: qual è la probabilità che l'inflazione sia minore del 5% e contemporaneamente il tasso di disoccupazione sia inferiore al 6%? In altre parole, a quanto ammonta $P(U \leq 6, P \leq 5)$? Da un punto di vista geometrico la risposta è data dal volume al di sotto della superficie e sopra il rettangolo (la base della figura) che definisce l'evento. Proprio come un integrale può essere usato per calcolare l'area sotto una curva, un integrale doppio viene utilizzato per calcolare volumi come quello

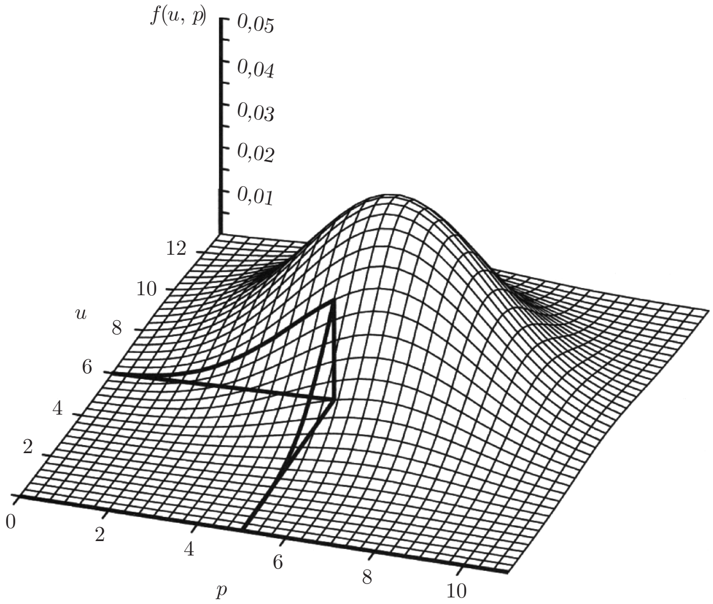

Figura B.2
Una funzione di densità congiunta.

illustrato nella figura B.2. Data la *fdd* congiunta $f(u,p)$, possiamo calcolare la probabilità come:

$$P(U \leq 6, P \leq 5) = \int_{u=0}^{6} \int_{p=0}^{5} f(u,p)dpdu$$

Come possiamo ricavare la *fdd* marginale di una delle variabili casuali a partire dalla *fdd* congiunta? La *fdd* marginale ci consente di rispondere immediatamente a domande del tipo: qual è la probabilità che la disoccupazione sia compresa fra il 2% e il 5%? Esattamente come per la (B.11), dobbiamo integrare rispetto alla variabile che non ci interessa. In altre parole, la **funzione di densità marginale** di U è data da:

(B.36)
$$f(u) = \int_{-\infty}^{\infty} f(u,p)dp$$

Una volta ottenuta questa *fdd*, possiamo calcolare la probabilità richiesta come $P(2 \leq U \leq 5) = \int_2^5 f(u)du$.

Potremmo anche chiederci: qual è la probabilità che il tasso di disoccupazione si trovi fra il 2% e il 5% **se** possiamo usare la politica monetaria per mantenere il tasso d'inflazione al 2%? Questa domanda riguarda una **probabilità condizionale**: dato che $P = 2$, qual è la probabilità che $2 \leq U \leq 5$? Più formalmente, qual è $P(2 \leq U \leq 5 | P = 2)$? Per rispondere a domande come questa nel caso di variabili casuali continue abbiamo bisogno della **funzione di densità condizionale** $f(u|p)$, definita da:

(B.37)
$$f(u|p) = \frac{f(u,p)}{f(p)}$$

A differenza del risultato descritto dalla (B.12) nel caso di variabili casuali discrete, questo rapporto non fornisce direttamente una probabilità ma una funzione di densità che può essere usata per calcolare probabilità. Usando $f(u|p)$ non otteniamo solo probabilità: possiamo anche calcolare il **valore atteso** o la **media condizionale**:

(B.38)
$$E(U|P = p) = \int_{-\infty}^{\infty} uf(u|p)du$$

Analogamente, la **varianza condizionale** è data da:

(B.39)
$$\text{Var}(U|P = p) = \int_{-\infty}^{\infty} [u - E(U|P = p)]^2 f(u|p)du$$

I problemi legati a inflazione e disoccupazione sono di enorme importanza sociale e sono molti gli economisti ed econometrici che li studiano; ne avete avuto alcuni brevi esempi nel corso di questo volume. Questi problemi economici tuttavia sono troppo complessi per permettere di illustrare in maniera semplice i concetti che abbiamo definito in questo paragrafo. Per questo motivo considereremo un esempio astratto ma più semplice.

Esempio B.3

Indichiamo con X e Y due variabili casuali continue con *fdd* $f(x,y) = x + y$ per x in $[0,1]$ e y in $[0,1]$. Potete mettere alla prova le vostre capacità geometriche provando a tracciare il grafico a tre dimensioni di questa funzione di densità congiunta. A questo proposito, chiediamoci se $f(x,y)$ descriva una funzione di densità valida. Ovviamente $f(x,y)$ soddisfa una versione più generale della (B.29), dato che $f(x,y) \geq 0$ per tutti i punti $x \in [0,1]$ e $y \in [0,1]$. La probabilità totale, inoltre, corrisponde al volume sotto la superficie ed è data da:

$$\begin{aligned}
\int_{y=0}^{1}\int_{x=0}^{1} f(x,y)dx\,dy &= \int_{y=0}^{1}\int_{x=0}^{1}(x+y)dx\,dy \\
&= \int_{y=0}^{1}\int_{x=0}^{1} xdx\,dy + \int_{y=0}^{1}\int_{x=0}^{1} ydx\,dy \\
&= \int_{y=0}^{1}\left[\int_{x=0}^{1} xdx\right]dy + \int_{x=0}^{1}\left[\int_{y=0}^{1} ydy\right]dx \\
&= \int_{y=0}^{1}\left[\frac{1}{2}x^2\Big|_0^1\right]dy + \int_{x=0}^{1}\left[\frac{1}{2}y^2\Big|_0^1\right]dx \\
&= \int_{y=0}^{1}\frac{1}{2}dy + \int_{x=0}^{1}\frac{1}{2}dx = \frac{1}{2}+\frac{1}{2} = 1
\end{aligned}$$

Nella terza riga abbiamo usato una proprietà degli integrali multipli. Nel Piccolo manuale di probabilità, al paragrafo P.4, la regola "Somma 9" afferma che l'ordine con cui vengono calcolate più sommatorie è irrilevante. Allo stesso modo, se i limiti di integrazione di una variabile non dipendono dal valore dell'altra, l'ordine di integrazione di un integrale multiplo è irrilevante. È fondamentale tuttavia mantenere il giusto accoppiamento fra il simbolo di integrale, con i suoi limiti inferiore e superiore, e la variabile di integrazione, rappresentata da dx o dy. Nel primo termine della terza riga dei passaggi precedenti abbiamo isolato l'integrale che riguarda x all'interno di quello rispetto a y. Gli integrali multipli sono calcolati procedendo "dall'interno verso l'esterno". Per prima cosa si risolve l'integrale rispetto a x e successivamente quello più esterno rispetto a y.

Per capire meglio i procedimenti necessari per calcolare un integrale multiplo consideriamo la probabilità che X sia compreso fra zero e $\frac{1}{2}$, mentre Y si trova fra $\frac{1}{4}$ e $\frac{3}{4}$. La quantità che cerchiamo è una probabilità congiunta ed è data da:

$$\begin{aligned}
P(0 \leq X \leq \tfrac{1}{2}, \tfrac{1}{4} \leq Y \leq \tfrac{3}{4}) &= \int_{y=1/4}^{3/4}\int_{x=0}^{1/2} f(x,y)dx\,dy \\
&= \int_{y=1/4}^{3/4}\int_{x=0}^{1/2}(x+y)dx\,dy \\
&= \int_{y=1/4}^{3/4}\int_{x=0}^{1/2} xdx\,dy + \int_{y=1/4}^{3/4}\left[\int_{x=0}^{1/2} ydx\right]dy \\
&= \int_{y=1/4}^{3/4}\left[\int_{x=0}^{1/2} xdx\right]dy + \int_{y=1/4}^{3/4} y\left[\int_{x=0}^{1/2} dx\right]dy \\
&= \int_{y=1/4}^{3/4}\left[\frac{1}{2}x^2\Big|_0^{1/2}\right]dy + \int_{y=1/4}^{3/4} y\left[x\Big|_0^{1/2}\right]dy \\
&= \frac{1}{8}\int_{y=1/4}^{3/4} dy + \frac{1}{2}\int_{y=1/4}^{3/4} ydy \\
&= \frac{1}{8}\left[y\Big|_{1/4}^{3/4}\right] + \frac{1}{2}\left[\frac{1}{2}y^2\Big|_{1/4}^{3/4}\right] = \frac{1}{8}\times\frac{1}{2} + \frac{1}{2}\times\frac{1}{4} = \frac{3}{16}
\end{aligned}$$

Nella terza uguaglianza di questo esempio non abbiamo cambiato l'ordine di integrazione del secondo termine. Questo fatto illustra un'altra caratteristica tipica delle operazioni con integrali multipli: quando calcoliamo l'integrale "interno" rispetto a x il valore di y resta fisso e può quindi essere portato fuori dall'integrale come nella quarta riga, semplificando il calcolo dell'integrale interno.

La *fdd* marginale di X, per $x \in [0,1]$, è data da:

$$\begin{aligned} f(x) &= \int_{y=0}^{1} f(x,y)dy = \int_{y=0}^{1} (x+y)dy = \int_{y=0}^{1} xdy + \int_{y=0}^{1} ydy = x \cdot y\Big|_0^1 + \frac{1}{2}y^2\Big|_0^1 \\ &= x + \frac{1}{2} \end{aligned}$$

Per essere completamente corretti dovremmo anche specificare che $f(x) = 0$ per $x \notin [0,1]$, ma in generale questa informazione addizionale non viene esplicitata. Usando passaggi analoghi possiamo dimostrare che la *fdd* di Y è $f(y) = y + \frac{1}{2}$ per valori di y nell'intervallo $[0,1]$. La *fdd* marginale di X può essere usata per calcolare la probabilità che X appartenga a un particolare sottointervallo del dominio di X, $x \in [0,1]$. Per esempio:

$$\begin{aligned} P(\tfrac{1}{2} < X < \tfrac{3}{4}) &= \int_{1/2}^{3/4} \left(x + \frac{1}{2}\right) dx = \int_{1/2}^{3/4} xdx + \frac{1}{2}\int_{1/4}^{3/4} dx \\ &= \frac{1}{2}x^2\Big|_{1/2}^{3/4} + \frac{1}{2}x\Big|_{1/2}^{3/4} \\ &= \frac{1}{2}\left(\frac{9}{16} - \frac{1}{4}\right) + \frac{1}{2}\left(\frac{3}{4} - \frac{1}{2}\right) \\ &= \frac{1}{2} \times \frac{5}{16} + \frac{1}{2} \times \frac{1}{4} = \frac{9}{32} \end{aligned}$$

Usando la *fdd* marginale di X possiamo calcolare il suo valore atteso:

$$\begin{aligned} \mu_X = E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x\Big(x + \frac{1}{2}\Big)dx \\ &= \int_0^1 x^2dx + \int_0^1 \frac{1}{2}xdx \\ &= \frac{1}{3}x^3\Big|_0^1 + \frac{1}{4}x^2\Big|_0^1 = \frac{1}{3} + \frac{1}{4} = \frac{7}{12} \end{aligned}$$

Nella prima riga i limiti di integrazione cambiano da $(-\infty, \infty)$ a $[0,1]$ perché per $x \notin [0,1]$, $f(x) = 0$ e l'area (probabilità) sotto $f(x) = 0$ è nulla.

Per calcolare la varianza di X dobbiamo per prima cosa ottenere:

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 f(x)dx = \int_0^1 x^2 \left(x + \frac{1}{2}\right) dx \\ &= \int_0^1 x^3dx + \int_0^1 \frac{1}{2}x^2dx \\ &= \frac{1}{4}x^4\Big|_0^1 + \frac{1}{6}x^3\Big|_0^1 = \frac{1}{4} + \frac{1}{6} = \frac{5}{12} \end{aligned}$$

Usando questo risultato otteniamo:

$$\sigma_X^2 = \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$$

La *fdd* condizionale di Y data $X = x$ è:

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

Nell'esempio B.3, la *fdd* condizionale è data da:

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{x+y}{x+\frac{1}{2}} \quad \text{per } y \in [0,1]$$

Un esempio specifico:

$$f\left(y \middle| X = \frac{1}{3}\right) = \frac{y+\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}} = \frac{1}{5}(6y+2) \quad \text{per } y \in [0,1]$$

La *fdd* condizionale può essere usata per calcolare la probabilità che Y appartenga a uno specifico intervallo. Possiamo inoltre calcolare la **media condizionale** di Y sapendo che $X = 1/3$:

$$\begin{aligned}
\mu_{Y|X=1/3} &= E\left(Y|X = \frac{1}{3}\right) = \int_{y=0}^{1} y f\left(y|X = \frac{1}{3}\right) dy \\
&= \int_{y=0}^{1} y \cdot \frac{1}{5}(6y+2) dy \\
&= \int_{y=0}^{1} \frac{6}{5} y^2 dy + \int_{y=0}^{1} \frac{2}{5} y \, dy \\
&= \frac{6}{5}\left(\frac{1}{3} y^3 \Big|_0^1\right) + \frac{2}{5}\left(\frac{1}{2} y^2 \Big|_0^1\right) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5}
\end{aligned}$$

Si noti che il valore atteso condizionale non coincide con il valore atteso **non condizionale** $\mu_Y = E(Y) = \frac{7}{12}$.

Per calcolare la **varianza condizionale** iniziamo da:

$$E\left(Y^2 \middle| X = \frac{1}{3}\right) = \int_0^1 y^2 f\left(y \middle| X = \frac{1}{3}\right) dy = \int_{y=0}^{1} y^2 \frac{1}{5}(6y+2)\, dy = \frac{13}{30}$$

La varianza condizionale è allora data da:

$$\mathrm{Var}\left(Y \middle| X = \frac{1}{3}\right) = E\left(Y^2 \middle| X = \frac{1}{3}\right) - \left[E\left(Y \middle| X = \frac{1}{3}\right)\right]^2 = \frac{11}{150} = 0{,}07333$$

La varianza non condizionale è $\sigma_Y^2 = \mathrm{Var}(Y) = \frac{11}{144} = 0{,}07639$. In questo particolare esempio la varianza condizionale è più piccola della varianza non condizionale.

La **correlazione** fra X e Y è data da:

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

La covarianza fra X e Y può essere calcolata usando la proprietà $\mathrm{Cov}(X,Y) = E(XY) - \mu_X \mu_Y$. Per calcolare il valore atteso di XY consideriamo l'integrale doppio seguente:

$$
\begin{aligned}
E(XY) &= \int_{y=0}^{1} \int_{x=0}^{1} xyf(x,y)\,dx\,dy = \int_{y=0}^{1} \int_{x=0}^{1} xy(x+y)\,dx\,dy \\
&= \int_{y=0}^{1} \int_{x=0}^{1} x^2y\,dx\,dy + \int_{y=0}^{1} \int_{x=0}^{1} xy^2\,dx\,dy \\
&= \int_{y=0}^{1} y\left[\int_{x=0}^{1} x^2\,dx\right] dy + \int_{y=0}^{1} y^2\left[\int_{x=0}^{1} x\,dx\right] dy = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}
\end{aligned}
$$

Di conseguenza:

$$
\mathrm{Cov}(X,Y) = E(XY) - \mu_X\mu_Y = \frac{1}{3} - \left(\frac{7}{12}\right)\left(\frac{7}{12}\right) = -\frac{1}{144}
$$

Infine, la correlazione fra X e Y è data da:

$$
\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{-1/144}{\sqrt{11/144}\sqrt{11/144}} = -\frac{1}{11} = -0{,}09091
$$

B.2.4. Valori attesi iterati

La **legge dei valori attesi iterati**, dimostrata nel paragrafo B.1.7 nel caso di variabili casuali discrete, è un risultato molto utile in diverse situazioni. Se X e Y sono variabili casuali continue con *fdd* congiunta $f(x,y)$, il valore atteso di Y può essere calcolato come:

$$
E(Y) = E_X[E(Y|X)]
$$

Questo risultato coincide con quello espresso dalla (B.26) nel caso discreto. Per capire meglio il significato esatto di questa espressione conviene per prima cosa dimostrarlo e successivamente illustrarlo con un esempio numerico. Per mostrare che l'uguaglianza precedente è corretta procediamo nel modo seguente:

$$
\begin{aligned}
E(Y) &= \int_{y=-\infty}^{\infty} yf(y)\,dy \\
&= \int_{y=-\infty}^{\infty} y\left[\int_{x=-\infty}^{\infty} f(x,y)\,dx\right] dy && \text{sostituendo la } \textit{fdd} \text{ marginale} \\
&= \int_y \int_x yf(x,y)\,dx\,dy && \text{semplificando l'integrale} \\
&= \int_y \int_x y[f(y|x)f(x)]\,dx\,dy && \text{sostituendo la } \textit{fdd} \text{ congiunta} \\
&= \int_x \left[\int_y yf(y|x)\,dy\right] f(x)\,dx && \text{invertendo l'ordine di integrazione} \\
&= \int_x [E(Y|X)]\; f(x)\,dx && \text{sostituendo } E(Y|X) \\
&= E_X\,[E(Y|X)] && \text{sostituendo il valore atteso rispetto a } X
\end{aligned}
$$

Nell'ultima riga di questa espressione il simbolo $E_X[\ ]$ indica il valore atteso del termine fra parentesi quadra rispetto a X. Si noti che nella terza riga abbiamo anche sostituito gli integrali di estremi $(-\infty, \infty)$ con una notazione compatta, $\int_y$ e $\int_x$, il cui significato è che l'integrale deve essere calcolato "su tutti i valori" della variabile di integrazione.

Per comprendere meglio l'espressione del valore atteso iterato calcoliamo il **valore atteso di Y condizionale a $X = x$** nell'esempio B.3, ma senza specificare un valore numerico per x:

$$E(Y|X = x) = \int_{y=0}^{1} yf(y|x)\,dy = \int_{y=0}^{1} y\left[\frac{x+y}{x+\frac{1}{2}}\right]dy = \frac{2+3x}{3(2x+1)}$$

Si noti che il calcolo dell'integrale rispetto ai valori di Y e considerando x fissa produce una funzione di x. Se teniamo conto che x può essere un valore qualsiasi, in altre parole che x rappresenta una variabile casuale, possiamo calcolare il valore atteso della funzione:

$$g(X) = \frac{2+3X}{3(2X+1)}$$

La legge dei valori attesi iterati afferma che il valore atteso di $g(X)$, calcolato considerando X casuale, è pari a $E(Y)$. Come esercizio, proviamo a verificare che questo è esattamente ciò che accade:

$$\begin{aligned}
E[g(X)] &= \int_{x=0}^{1} \frac{2+3x}{3(2x+1)} f(x)\,dx \\
&= \int_{x=0}^{1} \frac{2+3x}{3(2x+1)}\left(x+\frac{1}{2}\right)dx \\
&= \int_{x=0}^{1} \frac{2+3x}{3(2x+1)}\frac{1}{2}(2x+1)\,dx = \int_{x=0}^{1} \frac{1}{6}(2+3x)\,dx \\
&= \int_{x=0}^{1} \frac{1}{3}\,dx + \int_{x=0}^{1} \frac{1}{2}x dx = \frac{1}{3}x\Big|_0^1 + \frac{1}{4}x^2\Big|_0^1 = \frac{1}{3} + \frac{1}{4} = \frac{7}{12} = E(Y)
\end{aligned}$$

Funziona!
Oltre a rappresentare una proprietà interessante, la legge dei valori attesi iterati ha un paio di implicazioni importanti. Primo, sfruttando $E(Y) = E_X[E(Y|X)]$ possiamo verificare che se $E(Y|X) = 0$, $E(Y) = E_X[E(Y|X)] = E_X(0) = 0$. Se il valore atteso condizionale di Y è nullo, anche il valore atteso non condizionale di Y è nullo.

Secondo, se $E(Y|X) = E(Y)$, allora $\mathrm{Cov}(X, Y) = 0$. Per verificare questo risultato iniziamo riscrivendo $E(XY)$ come:

$$\begin{aligned}
E(XY) &= \int_x \int_y xyf(x,y)\,dy\,dx \\
&= \int_x \int_y xyf(y|x)f(x)\,dy\,dx \\
&= \int_x x\left[\int_y yf(y|x)\,dy\right] f(x)\,dx \\
&= \int_x x[E(Y|X)]\,f(x)\,dx
\end{aligned} \tag{B.40}$$

Se $E(Y|X) = E(Y)$, l'ultima riga di (B.40) diventa:

$$E(XY) = \int_x x[E(Y)]\, f(x)\, dx = E(Y) \int_x x\, f(x)\, dx = E(Y)E(X) = \mu_Y \mu_X$$

La covarianza fra Y e X in questo caso è pari a:

$$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y = \mu_X \mu_Y - \mu_X \mu_Y = 0$$

Un caso particolare estremamente importante di questi due risultati riguarda le conseguenze di $E(Y|X) = 0$. Abbiamo già visto che $E(Y|X) = 0$ implica $E(Y) = 0$. Possiamo ora osservare anche che se $E(Y|X) = E(Y) = 0$, $\text{Cov}(X, Y) = 0$. Questo risultato svolge un ruolo importante nell'ipotesi A.10.3* del paragrafo 10.1.3.

B.2.5. Distribuzioni di funzioni di variabili casuali

Tabella B.1
Cambiamento di variabile: il caso discreto

x	$P(X = x) = P(Y = y)$	y
1	0,1	5
2	0,2	8
3	0,3	11
4	0,4	14

In più occasioni abbiamo osservato che una funzione di una variabile casuale è essa stessa casuale. Il problema che affrontiamo in questo paragrafo è il seguente: 'qual è la funzione di densità della nuova variabile casuale? Nel caso di una variabile discreta questo problema non è troppo difficile. Consideriamo per esempio la variabile casuale discreta X che può assumere i valori 1, 2, 3 o 4 con probabilità rispettivamente pari a 0,1, 0,2, 0,3 e 0,4. Sia $Y = 2 + 3X = g(X)$: qual è la *fdp* di Y? In questo caso la risposta è chiara: la probabilità che $Y = 5$, 8, 11 o 14 corrisponde esattamente alla probabilità che X sia rispettivamente pari a 1, 2, 3 o 4, come indicato nella tabella B.1.

Ciò che rende possibile un risultato così semplice è il fatto che a ciascun valore di y corrisponde un unico valore di x e che a ogni valore di x corrisponde un unico valore di y. Un altro modo di definire questa situazione consiste nell'osservare che la trasformazione da X a Y è "biunivoca". Questa proprietà è senz'altro valida se la funzione $g(X)$ che collega Y con X è strettamente crescente o strettamente decrescente. Funzioni di questo tipo sono dette **monotone in senso stretto**. La nostra funzione $Y = 2 + 3X = g(X)$ è strettamente (o monotona) crescente; ciò garantisce che se $x_2 > x_1$, $y_2 = g(x_2) > y_1 = g(x_1)$. Si noti in particolare che stiamo escludendo la possibilità che $y_1 = y_2$.

Determinare la distribuzione di $Y = g(X)$ nel caso continuo è un po' più complicato. Nell'esempio seguente presentiamo la tecnica del **cambiamento di variabili** che si applica quando la funzione $g(X)$ è strettamente crescente o decrescente.

Esempio B.4
Indichiamo con X una variabile casuale continua con *fdd* $f(x) = 2x$ per $0 < x < 1$. Sia $Y = g(X) = 2X$ un'altra variabile casuale. Vogliamo calcolare la probabilità che Y appartenga a un certo intervallo. Una possibile soluzione consiste nel calcolare la probabilità per Y utilizzando la probabilità dell'evento corrispondente per X. Per esempio:

$$P(0 < Y < 1) = P\left(0 < X < \frac{1}{2}\right) = \int_0^{1/2} 2x\, dx = x^2 \Big|_0^{1/2} = \frac{1}{4}$$

Anche se questa strategia è ragionevole e in questo caso relativamente semplice, le cose non sono sempre così immediate. Un'alternativa preferibile consiste nel determinare la *fdd* di Y, indicata con $h(y)$, e usare quest'ultima per calcolare probabilità relative a Y. Dato che $X = Y/2$, potremmo immaginare di sostituire questa espressione nella *fdd* $f(x)$ e ottenere $h(y) = 2(y/2) = y$ per $0 < y < 2$. Sfortunatamente questa sostituzione non fornisce il risultato desiderato, dato che:

$$\int_{-\infty}^{\infty} h(y)\,dy = \int_0^2 y\,dy = \frac{1}{2}y^2\Big|_0^2 = 2$$

Questo risultato viola la proprietà (B.30) di una funzione di densità. Se usiamo $h(y)$ per calcolare la probabilità che Y appartenga all'intervallo $(0, 1)$, inoltre, otteniamo come risultato 0,5, che sappiamo essere errato.

Il problema è che l'altezza di $h(y)$ deve essere modificata per tenere conto del fatto che Y può assumere valori nell'intervallo $(0, 2)$, mentre X può assumere valori solo in $(0, 1)$. In effetti, una variazione di un'unità di Y corrisponde a una variazione di X di mezza unità. Se aggiustiamo $h(y)$ per questo fattore, otteniamo:

$$h(y) = 2(y/2)\left(\frac{1}{2}\right) = y/2, \quad 0 < y < 2$$

Usando questa *fdd* modificata la proprietà (B.30) è soddisfatta:

$$\int_{-\infty}^{\infty} h(y)\,dy = \int_0^2 \frac{1}{2}\,y\,dy = \frac{1}{4}y^2\Big|_0^2 = 1$$

Possiamo anche ottenere la probabilità corretta che Y cada nell'intervallo $(0, 1)$:

$$P(0 < Y < 1) = \int_0^1 \frac{1}{2}\,y\,dy = \frac{1}{4}y^2\Big|_0^1 = \frac{1}{4}$$

Possiamo osservare da un punto di vista diverso la tecnica del cambiamento di variabili esaminando la rappresentazione integrale della probabilità che Y appartenga all'intervallo $(0, 1)$:

$$P(0 < Y < 1) = \int_0^1 h(y)\,dy$$

La rappresentazione integrale dell'evento equivalente espresso in termini di X, evidenziando esplicitamente l'estremo inferiore e quello superiore dell'integrale, è data da:

$$P(0 < Y < 1) = P\left(0 < X < \frac{1}{2}\right) = \int_{x=0}^{x=1/2} f(x)\,dx = \int_{x=0}^{x=1/2} 2x\,dx$$

Se interpretiamo dx come una piccola variazione di X, la relazione $x = y/2$ implica $dx = dy/2$. Sostituendo questa espressione nell'integrale precedente, otteniamo:

$$P(0 < Y < 1) = \int_{y=0}^{y/2=1/2} 2\left(\frac{1}{2}\,y\right)\left(\frac{1}{2}\,dy\right) = \int_{y=0}^{y=1} \frac{1}{2}\,y\,dy$$

Il fattore di aggiustamento 1/2 in precedenza ottenuto intuitivamente compare in questa espressione attraverso la relazione fra dx e dy. Il nome matematico di questo fattore di aggiustamento è **jacobiano della trasformazione** (in realtà si tratta del suo valore assoluto, come vedremo ben presto). Il suo scopo è quello di rendere l'espressione dell'integrale in termini di x equivalente a quella in termini di y.

Siamo ora pronti a descrivere in maniera più accurata la tecnica del cambiamento di variabili.

Sia X una variabile casuale continua con *fdd* $f(x)$ e sia $Y = g(X)$ una funzione strettamente crescente o strettamente decrescente. Questa condizione assicura che la funzione sia biunivoca, in modo che esista esattamente un solo Y per ogni valore di X ed esattamente un solo X per ogni valore di Y. Il vantaggio di questa condizione su $g(X)$ è quello di poter risolvere $Y = g(X)$ rispetto a X; in altre parole, possiamo trovare una **funzione inversa** $X = w(Y)$. In questo caso la *fdd* di Y è data da:

$$h(y) = f[w(y)] \cdot \left| \frac{dw(y)}{dy} \right| \tag{B.41}$$

dove $|\cdot|$ indica il valore assoluto.

Tecnica del cambiamento di variabili per calcolare la *fdd* di Y:

1. Risolvere $y = g(x)$ rispetto a x in funzione di y;
2. Sostituire il risultato così ottenuto al posto di x in $f(x)$;
3. Moltiplicare il risultato per il valore assoluto della derivata $dw(y)/dy$, chiamata jacobiano della trasformazione.

Il fattore di scala $|dw(y)/dy|$ rappresenta il fattore di aggiustamento che consente di calcolare in maniera corretta le probabilità (in altri termini, gli integrali). Nell'esempio precedente la funzione inversa è $X = w(Y) = Y/2$. Il termine jacobiano è $dw(y)/dy = d(y/2)/dy = \frac{1}{2}$ e $|dw(y)/dy| = |\frac{1}{2}| = \frac{1}{2}$.

Esempio B.5

Sia X una variabile casuale continua di *fdd* $f(x) = 2x$ per $0 < x < 1$ e sia $Y = g(X) = 8X^3$ la funzione di X cui siamo interessati. La funzione $Y = g(X) = 8X^3$ è strettamente crescente sull'intervallo di valori ammissibili di X, $0 < x < 1$, cui corrisponde l'insieme di valori ammissibili di Y, dato da $0 < y < 8$. Dato che la funzione è strettamente crescente possiamo ricavare la funzione inversa:

$$x = w(y) = \left(\frac{1}{8}\,y\right)^{1/3} = \frac{1}{2}\,y^{1/3}$$

e

$$\frac{dw(y)}{dy} = \frac{1}{6}y^{-2/3}$$

Applicando la formula (B.41) del cambiamento di variabili, otteniamo:

$$\begin{aligned} h(y) &= f[w(y)] \cdot \left| \frac{dw(y)}{dy} \right| \\ &= 2\left(\frac{1}{2}\,y^{1/3}\right) \cdot \left| \frac{1}{6}y^{-2/3} \right| \\ &= \frac{1}{6}y^{-1/3}, \quad 0 < y < 8 \end{aligned}$$

La tecnica del cambiamento di variabili può essere adattata al caso di più variabili casuali, X_1, X_2, trasformate in Y_1, Y_2. Per una descrizione di questo metodo, che richiede una certa familiarità con l'algebra delle matrici, si vedano pp. 1004-1005 di William Greene, *Econometric Analysis*, 6a edizione, Pearson Prentice Hall, 2008.

B.3. Alcune importanti distribuzioni di probabilità

In questo paragrafo descriveremo e riassumeremo in maniera sintetica le proprietà delle distribuzioni di probabilità usate in questo volume.

B.3.1. Distribuzione di Bernoulli

Indichiamo con X la variabile casuale che descrive il risultato di un esperimento con due soli possibili esiti, A o B. Indichiamo con $X = 1$ l'esito A e con $X = 0$ l'esito B, e con $P(X = 1) = p$ e $P(X = 0) = 1 - p$ le probabilità dei due risultati, dove $0 \leq p \leq 1$. In questo caso X ha una **distribuzione di Bernoulli**. La *fdp* di una variabile casuale di Bernoulli è data da:

$$f(x|p) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1 \\ 0 & \text{altrimenti} \end{cases} \tag{B.42}$$

Il valore atteso di X è $E(X) = p$ e la sua varianza è $\text{Var}(X) = p(1 - p)$. Questa variabile casuale è molto utilizzata nei **modelli di scelta**, come il modello di probabilità lineare (capitoli 7, 8 e 16) e nei modelli di scelta binaria e multinomiale (capitolo 16).

B.3.2. Distribuzione binomiale

Supponiamo che $X_1, X_2, \ldots, X_n$ siano variabili casuali indipendenti, tutte con distribuzione di Bernoulli di parametro p; in questo caso $X = X_1 + X_2 + \ldots + X_n$ è una variabile casuale discreta che conta il numero di successi (in altre parole, il numero di esperimenti di Bernoulli nei quali $X_i = 1$) su n replicazioni dell'esperimento. La variabile casuale X ha **distribuzione binomiale**. La *fdp* della variabile casuale è:

$$P(X = x|n, p) = f(x|n, p) = \binom{n}{x} p^x(1-p)^{n-x} \quad \text{per } x = 0, 1, \ldots, n \tag{B.43}$$

dove:

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

è il numero di combinazioni possibili di n oggetti considerati a gruppi di x. Questa distribuzione ha due parametri, n e p, dove n è un intero positivo che indica il numero di replicazioni dell'esperimento e $0 \leq p \leq 1$. Queste probabilità sono abbastanza noiose da calcolare a mano, ma tutti i software econometrici sono in grado di fornirle automaticamente. Le probabilità dei valori 0, 1, 2, …, 10 sono illustrate nella figura B.3 nel caso $n = 10$.

Il valore atteso e la varianza di X sono dati da:

$$E(X) = \sum_{i=1}^{n} E(X_i) = np$$

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) = np(1-p)$$

Una variabile casuale collegata alla binomiale è $Y = X/n$, la quota di successi su n replicazioni dell'esperimento. La sua media e varianza sono rispettivamente date da $E(Y) = p$ e $\text{Var}(Y) = p(1 - p)/n$.

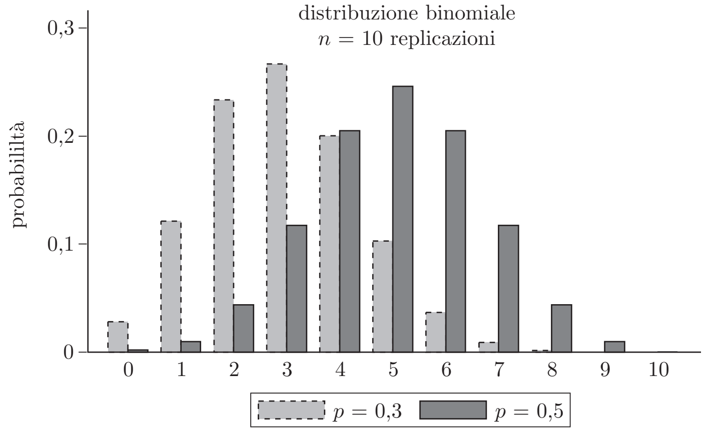

Figura B.3
Distribuzioni binomiali per $n = 10$.

B.3.3. Distribuzione di Poisson

Mentre la variabile casuale binomiale conta il numero di volte in cui si verifica un evento su n replicazioni dell'esperimento, la variabile casuale di Poisson conta il numero di realizzazioni di un evento in un certo intervallo di tempo o in una certa area. La funzione di probabilità di questa variabile casuale discreta X è data da:

$$P(X = x|\mu) = f(x|\mu) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{per } x = 0, 1, 2, 3, \ldots \tag{B.44}$$

Le probabilità dipendono dal parametro μ ed $e \approx 2{,}71828$ è la base dei logaritmi naturali. Il valore atteso e la varianza di X sono $E(X) = \mu$ e $\mathrm{Var}(X) = \mu$. La distribuzione di Poisson è usata in modelli per variabili di conteggio (capitolo 16), per esempio per descrivere il numero di visite mediche effettuate da un individuo in un anno. La figura B.4 illustra le probabilità per x compreso fra 0 e 10 quando $\mu = 3$ e $\mu = 4$.

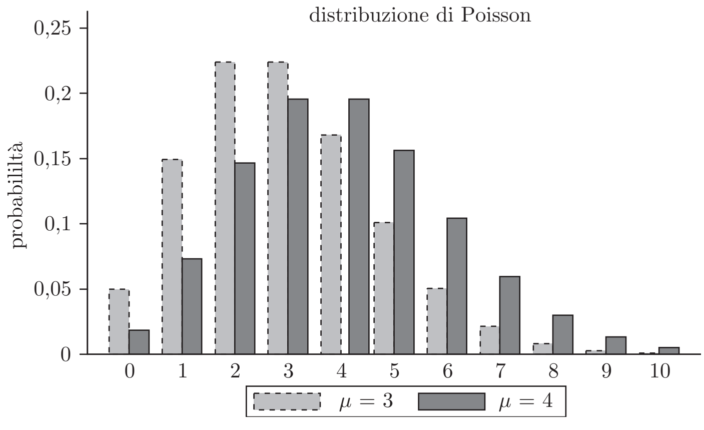

Figura B.4
Distribuzioni di Poisson.

B.3.4. Distribuzione uniforme

Una distribuzione continua molto importante da un punto di vista teorico è la **distribuzione uniforme**. La variabile casuale X con valori $a \leq X \leq b$ ha distribuzione uniforme se la sua *fdd* è data da:

$$f(x|a, b) = \frac{1}{b - a} \quad \text{per } a \leq x \leq b \tag{B.45}$$

La figura B.5 illustra il grafico di questa funzione di densità.

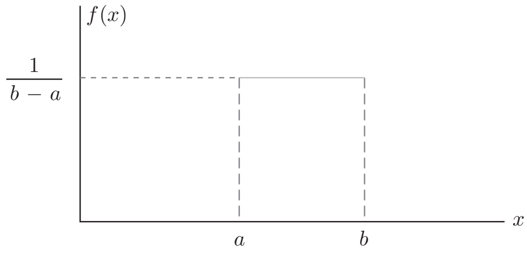

Figura B.5
Una distribuzione uniforme.

L'area sotto $f(x)$ compresa fra a e b vale 1, come richiesto per qualsiasi funzione di densità di una variabile casuale continua. Il valore atteso di X è il punto centrale dell'intervallo $[a, b]$, $E(X) = (a + b)/2$. Questo risultato può essere dedotto dalla simmetria della distribuzione. La varianza di X è $\mathrm{Var}(X) = E(X^2) - \mu^2 = (b - a)^2/12$.

Un caso particolare interessante si verifica quando $a = 0$ e $b = 1$; in questa situazione $f(x) = 1$ per $0 \leq x \leq 1$ e la distribuzione, illustrata nella figura B.6, descrive ciò che normalmente si intende per "numero casuale compreso fra 0 e 1".

La distribuzione uniforme ha la proprietà che due intervalli qualsiasi di pari ampiezza hanno la stessa probabilità di realizzarsi. In altre parole:

$$P(0{,}1 \leq X \leq 0{,}6) = P(0{,}3 \leq X \leq 0{,}8) = P(0{,}21131 \leq X \leq 0{,}71131) = 0{,}5$$

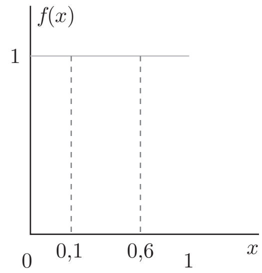

Figura B.6
Una distribuzione uniforme sull'intervallo $[0, 1]$.

Da un punto di vista concettuale, l'estrazione di un numero casuale compreso fra 0 e 1 è un'operazione complessa a causa del fatto che l'intervallo contiene un'infinità non numerabile di valori e che la probabilità di osservare ognuno di essi è nulla. Ciò che realmente si intende con la proprietà precedente è che tutti gli intervalli di uguale ampiezza hanno la stessa probabilità di essere osservati, indipendentemente da quanto possano essere stretti. È esattamente questa la proprietà che caratterizza la natura della distribuzione uniforme.

B.3.5. Distribuzione normale

La distribuzione normale è già stata descritta nel paragrafo P.6 del Piccolo manuale di probabilità. Un punto che non abbiamo discusso in quella sede è il motivo per il quale per calcolare le probabilità normali è necessario consultare delle tavole statistiche come la tabella 1 dell'appendice D. Per esempio, sappiamo che, per la variabile casuale continua X di distribuzione normale con media μ e varianza σ^2, la probabilità di appartenere all'intervallo $[a, b]$ è data da:

$$\int_a^b f(x)\, dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-\mu)^2 / 2\sigma^2\right] dx$$

Sfortunatamente questo integrale non ha una soluzione algebrica in forma chiusa. Di conseguenza siamo costretti a ricorrere a tavole di valori che contengono approssimazioni numeriche delle aree al di sotto della distribuzione normale standardizzata oppure a usare un software statistico in grado di fornire la stessa informazione.

La distribuzione normale è collegata alle distribuzioni chi quadro, t ed F che discuteremo nei prossimi paragrafi.

B.3.6. Distribuzione chi quadro

Le variabili casuali **chi quadro** si ottengono elevando al quadrato variabili casuali normali standardizzate. Se $Z_1, Z_2, \ldots, Z_m$ sono m variabili casuali $\mathcal{N}(0,1)$ *indipendenti fra loro*:

$$V = Z_1^2 + Z_2^2 + \ldots + Z_m^2 \sim \chi^2_{(m)} \tag{B.46}$$

La notazione $V \sim \chi^2_{(m)}$ va letta nel modo seguente: la variabile casuale V ha distribuzione chi quadro con m **gradi di libertà**. Il numero di gradi di libertà è un parametro che indica il numero di variabili casuali $\mathcal{N}(0,1)$ *indipendenti* che vengono elevate al quadrato e sommate fra loro per ottenere V. Il valore di m determina l'intera forma della distribuzione chi quadro, comprese media e varianza:

$$\begin{aligned} E(V) &= E\left[\chi^2_{(m)}\right] = m \\ \mathrm{Var}(V) &= \mathrm{Var}\left[\chi^2_{(m)}\right] = 2m \end{aligned} \tag{B.47}$$

La figura B.7 illustra la distribuzione chi quadro per diversi valori del numero di gradi di libertà m. I valori di V sono non negativi, $v \geq 0$, dato che V è ottenuta elevando al quadrato e sommando m variabili casuali normali standardizzate $\mathcal{N}(0,1)$. La distribuzione è *asimmetrica* a destra, la sua coda destra è molto pronunciata. Al crescere del numero di gradi di libertà m, tuttavia, la distribuzione diventa più simmetrica e "a campana". In effetti al crescere di m la distribuzione chi quadro converge a una distribuzione normale.

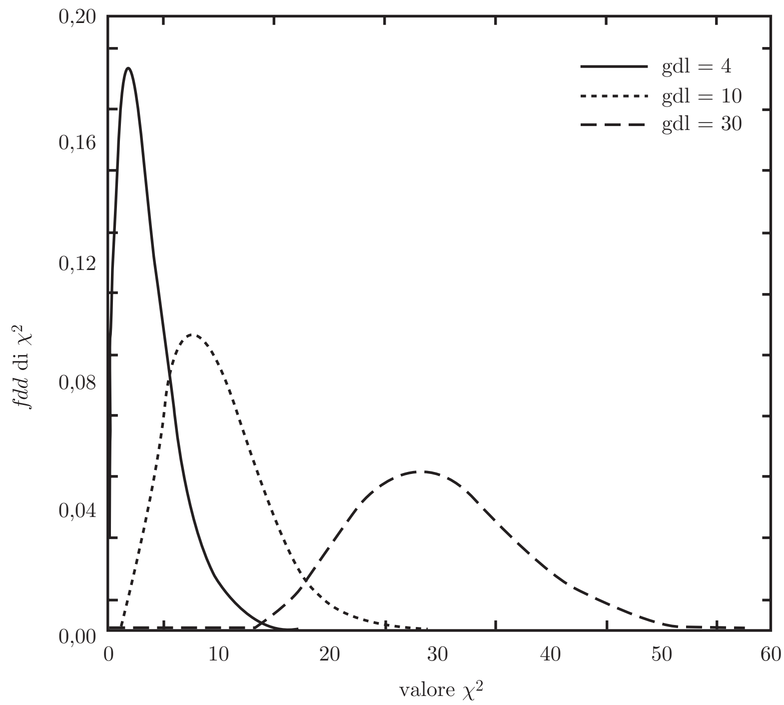

Figura B.7
La distribuzione chi quadro.

La tabella 3 dell'appendice D riporta il 90-esimo, 95-esimo e 99-esimo percentile della distribuzione chi quadro per alcuni valori del numero di gradi di libertà. Questi valori critici sono spesso utilizzati nelle verifiche d'ipotesi.

B.3.7. Distribuzione t

Una variabile casuale t (in lettera minuscola) è ottenuta dividendo una variabile casuale normale standardizzata, $Z \sim \mathcal{N}(0, 1)$, per la radice quadrata di una variabile casuale chi quadro *indipendente da* Z, $V \sim \chi^2_{(m)}$, divisa per il numero dei suoi gradi di libertà m. Se $Z \sim \mathcal{N}(0, 1)$ e $V \sim \chi^2_{(m)}$, e se Z e V sono indipendenti:

$$t = \frac{Z}{\sqrt{V/m}} \sim t_{(m)} \tag{B.48}$$

La forma della distribuzione t è interamente determinata dal suo parametro, il numero di gradi di libertà m; la distribuzione inoltre è di solito indicata con $t_{(m)}$.

La figura B.8 illustra un grafico della distribuzione t con $m = 3$ gradi di libertà e la confronta con la distribuzione $\mathcal{N}(0, 1)$. Si noti che la distribuzione t è meno "a punta" e più dispersa della normale. La distribuzione t è simmetrica, con media $E[t_{(m)}] = 0$ e varianza $\text{Var}[t_{(m)}] = m/(m - 2)$. Se facciamo tendere all'infinito il numero m dei gradi di libertà, la distribuzione $t_{(m)}$ converge alla normale standardizzata $\mathcal{N}(0, 1)$.

I software statistici ed econometrici offrono comandi che consentono di calcolare immediatamente la *fdr* di variabili casuali t e che sono particolarmente utili per calcolare delle probabilità. Dato che alcune di queste sono molto utilizzate in diversi contesti, la tabella 2 dell'appendice D contiene alcuni percentili molto utilizzati delle distribuzioni t chiamati **valori critici** della distribuzione. Per esempio, il 95-esimo percentile di una distribuzione t con 20 gradi di libertà è $t_{(0,95;\,20)} = 1{,}725$. Dato che la distribuzione t è simmetrica, la tabella 2 contiene solo i valori critici relativi alla coda destra della distribuzione.

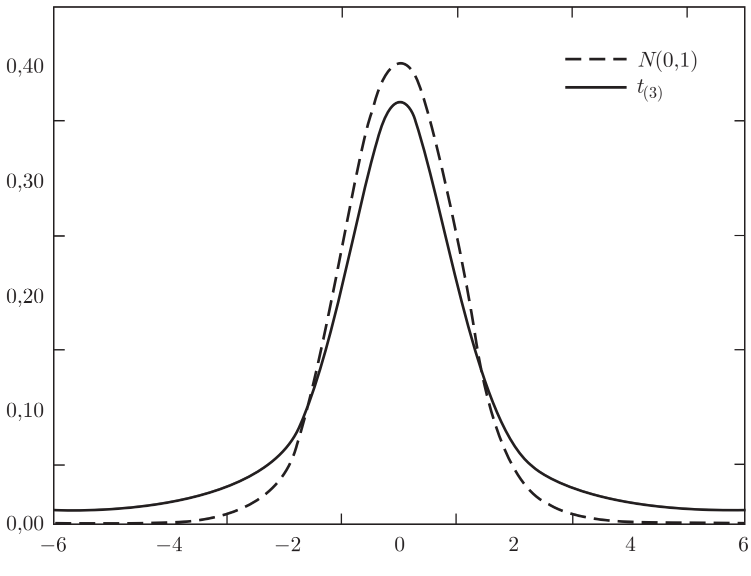

Figura B.8
Funzioni di densità delle distribuzioni normale standardizzata e $t_{(3)}$.

B.3.8. Distribuzione F

Una variabili casuale F è definita dal rapporto fra due variabili casuali chi quadro indipendenti, ciascuna divisa per il proprio numero di gradi di libertà. Se $V_1 \sim \chi^2_{(m_1)}$ e $V_2 \sim \chi^2_{(m_2)}$, e se V_1 e V_2 sono indipendenti:

$$F = \frac{V_1/m_1}{V_2/m_2} \sim F_{(m_1;\, m_2)} \tag{B.49}$$

Una distribuzione $F_{(m_1;\, m_2)}$ ha m_1 *gradi di libertà al numeratore* e m_2 *gradi di libertà al denominatore*. I valori di m_1 e m_2 determinano la forma della distribuzione, che in generale è simile a quella illustrata nella figura B.9. Questa distribuzione è definita sull'insieme di valori ammissibili $(0, \infty)$ e la sua coda destra è molto pronunciata. Per esempio, il 95-esimo percentile di una distribuzione F con $m_1 = 8$ gradi di libertà al numeratore e $m_2 = 20$ al denominatore è $F_{(0{,}95;\, 8;\, 20)} = 2{,}45$. I valori critici della distribuzione F sono riportati nelle tabelle 4 (95-esimo percentile) e 5 (99-esimo percentile) dell'appendice D.

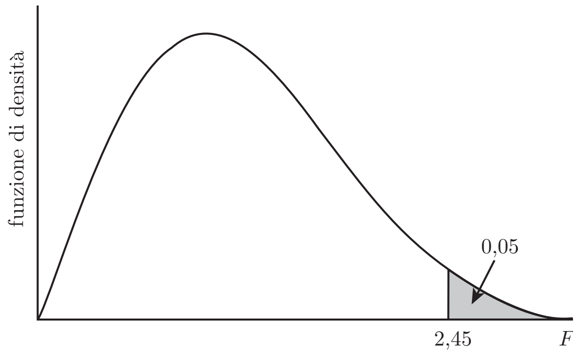

Figura B.9
Funzione di densità di una variabile casuale $F_{(8;\, 20)}$.

B.4. Numeri casuali

In molti capitoli di questo volume abbiamo svolto simulazioni Monte Carlo per illustrare le proprietà campionarie degli stimatori; si vedano per esempio i capitoli 3, 4, 5, 10 e 11. Per usare le simulazioni Monte Carlo dobbiamo essere in grado di generare **numeri casuali** da specifiche distribuzioni di probabilità, come per esempio l'uniforme e la normale. L'uso del computer per compiere esperimenti di simulazione è diffuso in tutte le scienze. In questo paragrafo vi forniremo un'introduzione a questo genere di applicazioni numeriche[1]. Per prima cosa è importante capire che la sola idea di generare numeri casuali con l'ausilio del computer è paradossale, perché per definizione i numeri casuali che vengono "generati" non possono essere autenticamente casuali. Quelli generati da un computer sono **numeri pseudo-casuali**, nel senso che "si comportano come se fossero casuali". Presenteremo una tecnica utilizzata per generare numeri pseudo-casuali chiamata **metodo della trasformazione inversa**, o **metodo di inversione**. Questa tecnica assume di essere in grado di generare numeri pseudo-casuali dalla **distribuzione uniforme** (si vedano i paragrafi B.3.4 e B.4.1) sull'intervallo $(0, 1)$. Le variabili casuali con distribuzione uniforme vengono poi trasformate in variabili casuali provenienti da un'altra distribuzione.

[1] Un buon testo di riferimento su questo argomento è quello di James E. Gentle, *Random Number Generation and Monte Carlo Methods*, Springer, New York, 2003.

Esempio B.6
Indichiamo con U una variabile casuale di distribuzione uniforme; U è una variabile continua con *fdd* $h(u) = 1$ per $u \in (0, 1)$ (si veda la figura B.6 per una rappresentazione grafica). Se consideriamo la trasformazione $Y = U^{1/2}$, l'insieme dei suoi valori ammissibili è $0 < y < 1$; dato che la radice quadrata è strettamente crescente, inoltre, possiamo applicare la tecnica del cambiamento di variabile per calcolare la *fdd* di Y. La funzione inversa è data da $U = w(Y) = Y^2$ e lo jacobiano della trasformazione è $dw(y)/dy = d(y^2)/dy = 2y$. La *fdd* di Y è dunque data da:

$$f(y) = h[w(y)] \cdot \left|\frac{dw(y)}{dy}\right| = 1 \cdot |2y| = 2y, \quad 0 < y < 1 \tag{B.50}$$

Abbiamo già usato questa distribuzione negli esempi B.4 e B.5. L'importanza di questa derivazione sta nel fatto che essa mostra che per ottenere un numero casuale a partire dalla distribuzione descritta dalla (B.50) è sufficiente calcolare la radice quadrata di un numero casuale generato da una distribuzione uniforme.

L'esempio B.6 suggerisce una tecnica generale, il metodo di inversione, per estrarre numeri casuali da una distribuzione qualsiasi. Supponiamo che desideriate generare un numero casuale da una specifica distribuzione di probabilità con *fdd* $f(y)$ e *fdr* $F(y)$.

Metodo di inversione:

1. Generare un numero casuale uniforme u_1 nell'intervallo $(0, 1)$.
2. Considerare l'equazione $u_1 = F(y_1)$.
3. Risolvere l'equazione al passo 2 rispetto a y_1.
4. Il valore y_1 è un numero casuale tratto dalla *fdd* $f(y)$.

Il metodo di inversione può essere usato per generare numeri casuali in accordo con qualsiasi distribuzione che consenta di portare a termine il passo 3. La soluzione è spesso indicata con $y_1 = F^{-1}(u_1)$, dove F^{-1} è detta **funzione di ripartizione inversa**. Affinché la F^{-1} esista, la *fdr* F deve essere **invertibile**.

Supponiamo che la distribuzione dalla quale ci interessa generare numeri casuali sia descritta da $f(y) = 2y$, $0 < y < 1$. La *fdr* di Y è $P(Y \leq y) = F(y) = y^2$, $0 < y < 1$. Le due funzioni sono illustrare nella figura B.10. Dato un numero casuale tratto dalla distribuzione uniforme su $(0, 1)$, consideriamo l'equazione $u_1 = F(y_1) = y_1^2$ e risolviamola rispetto a y_1, ottenendo $y_1 = F^{-1}(u_1) = (u_1)^{1/2}$. Il valore y_1 calcolato in questo modo è un'**estrazione casuale** dalla distribuzione di probabilità descritta da $f(y) = 2y$, $0 < y < 1$. Questa procedura si accorda perfettamente con il risultato ottenuto nell'esempio B.6 nel quale abbiamo mostrato che la radice quadrata di una variabile casuale uniforme ha proprio la distribuzione dalla quale vogliamo simulare.

La figura B.10a assume che il numero casuale generato dalla distribuzione uniforme sia $u_1 = 0{,}16$. Questo numero cade fra 0 e 1 lungo l'asse verticale sul quale viene misurata la *fdr* $F(x)$. Il valore $u_1 = 0{,}16$ corrisponde a $y_1 = 0{,}4 = (u_1)^{1/2} = (0{,}16)^{1/2}$ sull'asse orizzontale. Nel grafico inferiore osserviamo il collegamento fra la *fdd* e la *fdr* . L'area sotto la *fdd* a sinistra di $y_1 = 0{,}4$ rappresenta la probabilità $P(0 < y < 0{,}4) = 0{,}16$. A ogni numero casuale u_i tratto da una distribuzione uniforme corrisponde un unico y_i tratto dalla distribuzione $f(y) = 2y$, $0 < y < 1$.

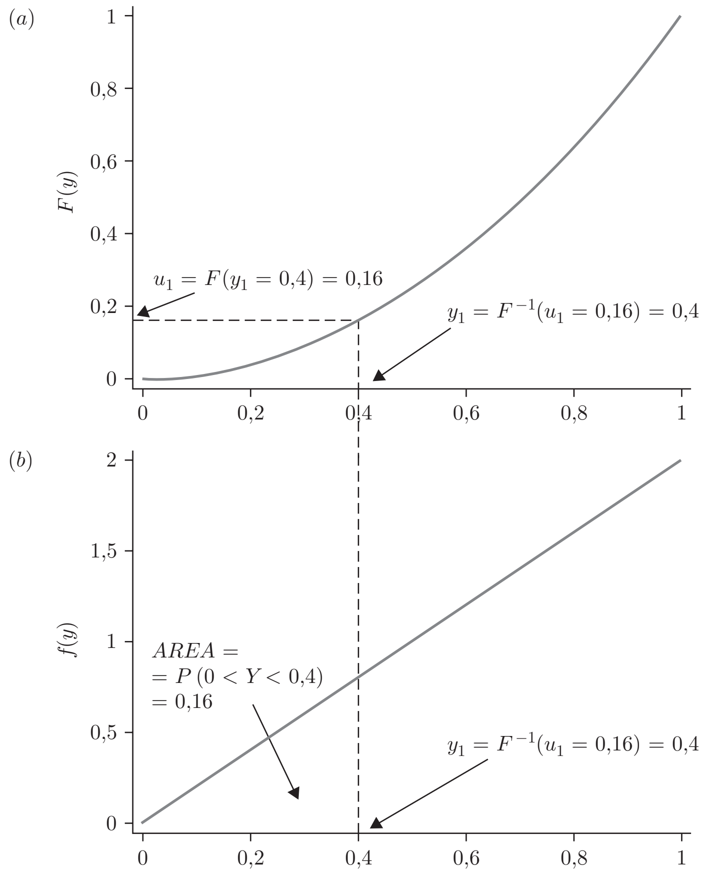

Figura B.10
(*a*) Funzione di ripartizione.
(*b*) Funzione di densità.

Come esempio, consideriamo le 1000 osservazioni relative a due variabili casuali uniformi indipendenti *U1* e *U2* contenute nel file *uniform1.dat*[2]. La figura B.11 illustra l'istogramma di *U1*; in ognuno dei 10 intervalli è contenuto il 10% circa delle osservazioni, proprio quello che ci aspetteremmo per valori generati da una distribuzione uniforme.

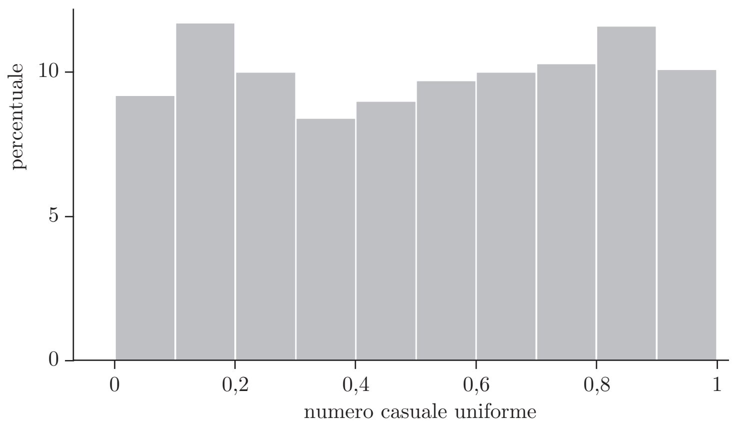

Figura B.11
Istogramma di 1000 valori casuali uniformi.

[2]Se preferite lavorare con un campione più numeroso, il file *uniform2.dat* contiene 10 000 osservazioni.

Indichiamo con *Y1* le radici quadrate dei valori di *U1*. La figura B.12 illustra l'istogramma di questi valori; non sembra anche a voi che assomigli molto a un triangolo? Proprio come la densità $f(y) = 2y$, $0 < y < 1$.

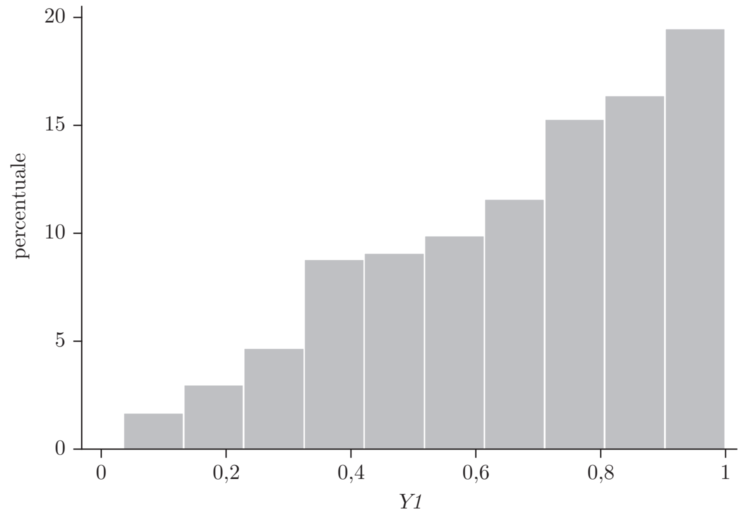

Figura B.12
Istogramma di 1000 radici quadrate di numeri casuali uniformi.

Come secondo esempio consideriamo una distribuzione leggermente più esotica. La **distribuzione a valore estremo** è alla base dei modelli di scelta logit discussi nel capitolo 16. Questa distribuzione ha funzione di densità $f(v) = \exp(-v) \cdot \exp[-\exp(-v)]$, illustrata nella figura B.13. La *fdr* a valore estremo è $F(v) = \exp[-\exp(-v)]$. Nonostante questa espressione apparentemente complessa, possiamo generare valori da questa distribuzione usando $v = F^{-1}(u) = -\log[-\log(u)]$. Usando i 1000 valori di *U1* contenuti in *uniform1.dat* otteniamo l'istogramma dei valori generati dalla distribuzione a valore estremo illustrato nella figura B.14[3].

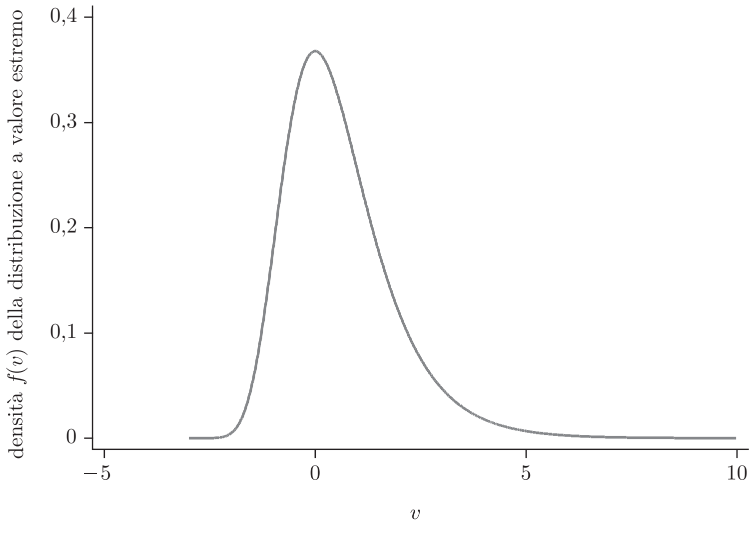

Figura B.13
Distribuzione a valore estremo.

[3]La curva continua è una stima della funzione di densità ottenuta sul campione simulato applicando un metodo non parametrico basato su un kernel gaussiano. Si veda l'appendice C.10 per una presentazione di questo tipo di stime.

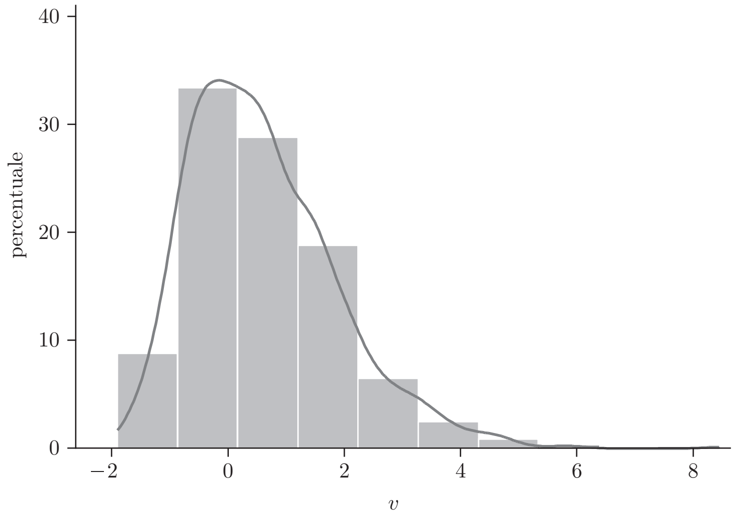

Figura B.14
Istogramma di estrazioni casuali dalla distribuzione a valore estremo.

La curva continua sovrimpressa all'istogramma assomiglia molto alla funzione di densità a valore estremo rappresentata nella figura B.13.

Riepilogando, il **metodo di inversione** usato per generare numeri casuali in accordo con una specifica distribuzione dipende da (1) la disponibilità di un certo numero di numeri casuali uniformi e (2) il fatto che la densità da cui vogliamo simulare abbia una *fdr* invertibile. La procedura non può essere usata per distribuzioni congiunte.

Grazie al metodo di inversione potete generare variabili casuali da distribuzioni non uniformi a partire da un generatore di numeri casuali uniformi. I testi dedicati all'esame delle distribuzioni di probabilità[4] sono ricchi di suggerimenti utili per trasformare numeri casuali uniformi in numeri casuali generati in accordo con un'ampia varietà di distribuzioni. L'esercizio B.8 illustra un metodo particolare per generare numeri casuali da una distribuzione normale.

B.4.1. Numeri casuali uniformi

Per poter utilizzare il metodo di inversione è necessario saper generare numeri casuali da una distribuzione uniforme. Per generazione di "numeri casuali", senza ulteriori qualificazioni, si intende normalmente quella da una distribuzione uniforme, che di per se stessa è un argomento molto vasto di ricerca e di studio. In precedenza abbiamo osservato che la nozione di numeri casuali generati da un computer è intrinsecamente illogica: i computer usano algoritmi per fare il loro lavoro e un algoritmo è una formula costruita in modo che il risultato non sia "casuale", anche se può apparire tale. I computer generano **numeri pseudo-casuali**. Provate a inserire questo termine in un motore di ricerca e vedrete apparire molti, moltissimi collegamenti.

Un termine matematico che appare spesso in questi documenti è quello di **modulo**, indicato con $a \bmod b$. In breve, $a \bmod b$ è il resto della divisione di a per b.

[4] Si veda per esempio Catherine Forbes, Merran Evans, Nicholas Hastings e Brian Peacock, *Statistical Distributions*, 4a edizione, John Wiley and Sons, 2010.

Un metodo per calcolare il modulo è[5]:

$$n \bmod m = n - m\,\mathrm{int}(n/m) \tag{B.51}$$

dove **int** è l'abbreviazione di **parte intera**, una funzione che arrotonda all'intero più piccolo[6]. Per osservare questa definizione all'opera consideriamo:

$$7 \bmod 3 = 7 - 3\,\mathrm{int}(7/3) = 7 - 3\,\mathrm{int}(2{,}3333) = 7 - 3 \cdot 2 = 1$$

Un metodo molto usato per generare un numero casuale uniforme è costituito dal **generatore lineare congruente**[7]. Consideriamo la relazione ricorsiva:

$$X_n = (aX_{n-1} + c) \bmod m \tag{B.52}$$

dove a, c e m sono costanti arbitrarie. La relazione (B.52) assegna a X_n un valore pari al resto della divisione intera di $aX_{n-1} + c$ per m ed è ricorsiva, dato che l'n-esimo valore dipende dall'$n - 1$-esimo. Questa proprietà implica che per inizializzare la sequenza è necessario scegliere un valore di partenza X_0, chiamato **seme** della sequenza di numeri casuali. Chiunque usi lo stesso seme e gli stessi valori di a, c e m genererà la stessa sequenza di numeri casuali. m è il divisore usato nella (B.52) e determina il periodo massimo dei valori generati ricorsivamente. Per ottenere numeri casuali uniformi nell'intervallo $(0, 1)$ considereremo $U_n = X_n/m$. Nei computer con architettura a 32 bit il valore di m è spesso fissato a 2^{32}. I valori di a e c hanno un'importanza cruciale per il successo del generatore casuale: una scelta non adeguata produce sequenze di numeri chiaramente non casuali. Provate per esempio a digitare RANDU in un motore di ricerca. Questo nome indica un generatore di numeri casuali molto popolare negli anni Sessanta del secolo scorso (anche noi lo abbiamo usato!), ma di cui sono stati in seguito scoperti alcuni gravi difetti, fra i quali l'incapacità di superare con successo i più semplici test di casualità[8].

Per illustrare il modo in cui il processo definito dalla (B.52) può generare numeri apparentemente casuali scegliamo $X_0 = 1234567$, $a = 1664525$, $c = 1013904223$ e $m = 2^{32}$ e generiamo 10 000 valori, indicati con *U1* nel file *uniform3.dat*[9]. Usando un istogramma con 20 intervalli, ci aspetteremmo che ognuno di essi contenga il 5% dei numeri casuali; la figura B.15 conferma che questo è approssimativamente proprio quello che accade.

I 10 000 valori di *U1* hanno media campionaria 0,4987197 e varianza 0,0820758, da confrontare con i veri valori di media e varianza per una distribuzione uniforme, rispettivamente pari a 0,5 e 0,08333. Il numero casuale minimo e quello massimo sono rispettivamente 0,0000327 e 0,9998433.

Ciò che questi esperimenti ci dicono è che i numeri casuali non sono veramente casuali e che alcuni generatori di numeri casuali sono migliori di altri. Alcuni di quelli citati più spesso sono il *Marsenne twister* (implementato in SAS 9.1) e l'algoritmo *KISS+Monster* (usato da Gauss 10). Nuovi generatori vengono sviluppati

[5] `www.functions.wolfram.com/IntegerFunctions/Mod/27/01/03/01/0001/`.

[6] $\mathrm{int}(x)$ è il più grande intero non superiore a x.

[7] Per una descrizione e alcuni riferimenti bibliografici si veda: `www.en.wikipedia.org/wiki/Linear_congruential_generator`.

[8] George Marsaglia ha sviluppato una serie di test di casualità molto utilizzati. Questi strumenti sono disponibili presso `www.stat.fsu.edu/pub/diehard/`.

[9] La variabile *U2* nello stesso file usa come seme 987654321.

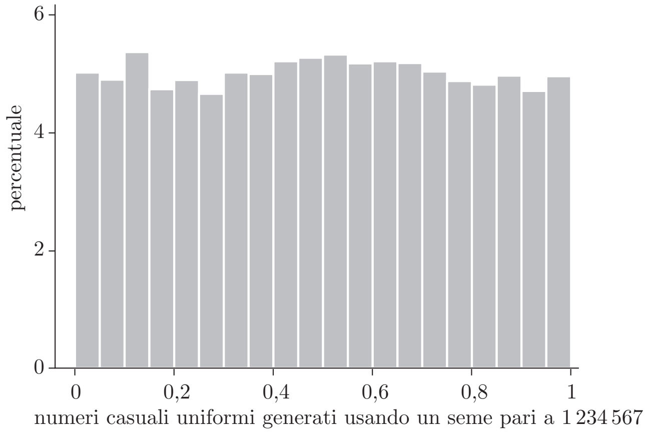

Figura B.15
Istogramma di 10 000 numeri casuali.

di continuo e ogni produttore di software usa un algoritmo diverso, mantenendo gelosamente il segreto sulle sue caratteristiche, che sono in ogni caso molto difficili da ricostruire.

La terza lezione è che con ogni probabilità **non dovreste mai** tentare di sviluppare voi stessi generatori di numeri casuali. Il professor Ken Train, un econometrico che ha dedicato gran parte delle sue ricerche allo studio di metodi numerici, afferma[10] che "da un punto di vista pratico, il mio consiglio è il seguente: a meno di non essere disposti a investire una grande quantità di tempo nello studio e nella ri-soluzione [...]" dei problemi associati alla progettazione di algoritmi di generazione di numeri casuali, "[...] è probabilmente preferibile usare i comandi software già disponibili anziché crearne di nuovi". A nostro avviso la cosa migliore da fare è usare i generatori di numeri casuali disponibili, ma documentare adeguatamente il lavoro specificando il software utilizzato e la sua versione, dato che aggiornamenti del software possono far cambiare i risultati da una versione alla successiva.

B.5. Esercizi

Alla pagina web `http://online.universita.zanichelli.it/hillecon` sono disponibili le risposte agli esercizi indicati con un asterisco.

B.1* Siano $X_1, X_2, \ldots, X_n$ variabili casuali indipendenti con la stessa distribuzione di probabilità di media μ e varianza σ^2. Considerate:

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

(a) Usate le proprietà del valore atteso per dimostrare che $E(\overline{X}) = \mu$.
(b) Usate le proprietà della varianza per dimostrare che $\mathrm{Var}(\overline{X}) = \sigma^2/n$. Come avete usato l'ipotesi di indipendenza?

B.2 Supponiamo che Y_1, Y_2, Y_3 sia un campione di osservazioni *non indipendenti fra loro* tratte da una popolazione $\mathcal{N}(\mu, \sigma^2)$. In particolare, supponiamo che:

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{Cov}(Y_1, Y_3) = \mathrm{Cov}(Y_2, Y_3) = \frac{\sigma^2}{2}$$

[10] *Discrete Choice Methods with Simulation*, Cambridge University Press, 209, 2003.

Considerate $\overline{Y} = (Y_1 + Y_2 + Y_3)/3$.
(a) Calcolate $E(\overline{Y})$.
(b) Calcolate $\text{Var}(\overline{Y})$.

B.3 Supponiamo che X sia una variabile casuale continua con funzione di densità data da:

$$f(x) = -\frac{1}{2}x + 1 \qquad 0 \leq x \leq 2$$

(a) Tracciate il grafico della funzione di densità $f(x)$.
(b) Calcolate l'area complessiva sotto $f(x)$ per $0 \leq x \leq 2$.
(c) Calcolate $P(X \geq 1)$ usando sia un argomento geometrico sia il calcolo di un integrale.
(d) Calcolate $P(X \leq \frac{1}{2})$.
(e) Calcolate $P(X = \frac{3}{2})$.
(f) Calcolate il valore atteso e la varianza di X.
(g) Calcolate la funzione di ripartizione di X.

B.4 Supponiamo che X sia una variabile casuale uniforme sull'intervallo (a, b).
(a) Calcolate con due integrali media e varianza di X.
(b) Calcolate la funzione di ripartizione di X.

B.5* Usate la relazione ricorsiva descritta da (B.52) con $X_0 = 79$, $m = 100$, $a = 263$ e $c = 71$ per generare 40 valori $X_1, X_2, \ldots, X_{40}$. Secondo voi, questi numeri sembrano casuali? Quello proposto è o non è un buon generatore di numeri casuali?

B.6 Supponiamo che X abbia distribuzione normale di media μ e varianza σ^2. Usate la tecnica del cambiamento di variabili per calcolare la funzione di densità di $Y = aX + b$.

B.7* Mostrate che se $E(Y|X) = E(Y)$, allora $\text{Cov}[Y, g(X)] = 0$ per qualsiasi funzione $g(X)$.

B.8 I numeri casuali normali sono molto utili nelle simulazioni Monte Carlo. Un modo per generarli consiste nell'usare la trasformazione di Box e Muller, che trasforma due numeri casuali uniformi *U1* e *U2* in due nuove variabili casuali, *Z1* e *Z2*, indipendenti fra loro e con distribuzione $\mathcal{N}(0, 1)$. La trasformazione è definita da:

$$Z1 = \sqrt{-2\log(U1)}\,\cos(2\pi U2) \qquad Z2 = \sqrt{-2\log(U1)}\,\text{sen}(2\pi U2)$$

(a) Costruite un istogramma delle osservazioni di *Z1* e *Z2* ottenute usando i 1000 numeri casuali uniformi *U1* e *U2* contenuti nel file *uniform1.dat* (oppure i 10 000 numeri casuali uniformi in *uniform2.dat*). Vi sembra di osservare una forma "a campana"?
(b) Calcolate le statistiche descrittive di *Z1* e *Z2*. Medie e varianze campionarie sono vicine rispettivamente a 0 e 1?
(c) Costruite un diagramma a dispersione di *Z1* rispetto a *Z2*; in altre parole, tracciate il grafico delle coppie di numeri casuali misurando *Z1* sull'asse verticale e *Z2* su quello orizzontale di un piano cartesiano. Vi sembra di individuare qualche traccia di correlazione positiva o negativa fra le osservazioni di *Z1* e quelle di *Z2*?

B.9* Sia X una variabile casuale continua con *fdd* $f(x) = 3x^2/8$ per $0 < x < 2$. Calcolate:
(a) $P(0 < X < \frac{1}{2})$.
(b) $P(1 < X < 2)$.

B.10 Una variabile casuale continua X ha distribuzione esponenziale se la sua *fdd* è $f(x) = e^{-x}$, $x \geq 0$.
(a) Tracciate il grafico della funzione di densità per $0 \leq x \leq 10$.
(b) La funzione di ripartizione di X è $F(x) = 1 - e^{-x}$. Tracciate il grafico di questa funzione nell'intervallo $0 \leq x \leq 10$. Vi sembra strettamente crescente, strettamente decrescente, o con andamento ambiguo?
(c) Usate il metodo della trasformazione inversa per generare osservazioni della variabile *X1* tratte da questa distribuzione. A questo scopo, usate i 1000 valori di *U1* nel file *uniform1.dat* oppure i 10 000 valori della stessa variabile in *uniform2.dat*. Costruite un istogramma dei valori che avete ottenuto. Vi sembra che questo istogramma assomigli al grafico ottenuto al punto (a)?
(d) La vera media e varianza di X sono $\mu = 1$ e $\sigma^2 = 1$. Quanto sono vicine ai veri valori la media e la varianza campionarie?

B.11 Usate la relazione ricorsiva (B.52) con $X_0 = 1\,234\,567$, $m = 2^{32}$, $a = 1\,103\,515\,245$ e $c = 12\,345$ per generare 1000 numeri casuali indicati con *U1*. Vi sembra che questi valori siano casuali? Quello proposto è o non è un buon generatore? Scegliete un altro seme e generate altri 1000 valori, indicandoli con *U2*. Calcolate le statistiche descrittive di *U1* e *U2*. Vi sembra che questi valori si comportino come atteso?

B.12* Supponete che la *fdd* congiunta delle variabili casuali continue X e Y sia data da $f(x, y) = 6x^2y$ per $0 \leq x \leq 1, 0 \leq y \leq 1$.
(a) Questa funzione soddisfa le condizioni necessarie per poter essere una *fdd* valida?
(b) Calcolate la *fdd* marginale di X, nonché la sua media e varianza.
(c) Calcolate la *fdd* marginale di Y.
(d) Calcolate la *fdd* di X condizionale a $Y = \frac{1}{2}$.
(e) Calcolate la media e la varianza di X condizionali a $Y = \frac{1}{2}$.
(f) Vi sembra che X e Y siano indipendenti? Giustificate la vostra risposta.

B.13 Supponete che X e Y siano variabili casuali continue con *fdd* congiunta $f(x, y) = \frac{1}{2}$ per $0 \leq x \leq y \leq 2$ e $f(x, y) = 0$ altrimenti. Si noti che i valori di X sono sempre inferiori o uguali a quelli di Y.
(a) Verificate che il volume al di sotto della funzione di densità vale 1.
(b) Calcolate le *fdd* marginali di X e Y.
(c) Calcolate $P(X < \frac{1}{2})$.
(d) Calcolate la *fdr* marginale di Y.
(e) Calcolate la probabilità condizionale $P(X < \frac{1}{2} | Y = 1{,}5)$. Secondo voi, X e Y sono indipendenti?
(f) Calcolate il valore atteso e la varianza di Y.
(g) Usate la legge dei valori attesi iterati per calcolare $E(X)$.