

Appendice di matematica e statistica

1 Appendice matematica	257
1.1 Concetto di derivata	257
1.2 Funzioni continue e funzioni derivabili.....	257
1.3 Regole generali di derivazione.....	258
1.3.1 Regola di derivazione di una costante.....	258
1.3.2 Regola di derivazione della funzione potenza.....	258
1.3.3 Linearità della derivazione	259
1.3.4 Regola di derivazione di un prodotto di funzioni	259
1.3.5 Regola di derivazione di un quoziente di funzioni.....	260
1.3.6 Regola della catena di derivate	260
1.3.7 Regola di derivazione di una funzione inversa	260
1.3.8 Regola di derivazione di una funzione logaritmica	261
1.3.9 Regola di derivazione di una funzione esponenziale	261
1.4 Derivata parziale	261
1.5 Differenziale parziale e totale.....	262
1.6 Derivata totale	263
1.7.2 Derivata di una funzione implicita.....	264
1.8 Derivata seconda, funzioni concave e funzioni convesse.....	264
1.9 Ottimizzazione statica	266
1.8.1 Condizione di primo ordine	266
1.8.2 Condizione di secondo ordine	267
1.9 Espansione in serie di Taylor	268
1.10 Il caso di più variabili di scelta.....	269
1.11 Ottimizzazione vincolata.....	271
2 Appendice statistica.....	274
2.1 Variabile casuale discreta	274
2.2 Valore atteso di una variabile casuale discreta.....	276
Regole del valore atteso.....	277
2.3 Varianza di una variabile casuale discreta.....	277
Regole della varianza.....	278
2.4 Covarianza tra due variabili casuali	279
Regole della covarianza	280
2.5 Coefficiente di correlazione.....	280
2.6 Aspettativa condizionata.....	280

1 Appendice matematica

1.1 Concetto di derivata

Data una generica funzione $y = f(x)$, la *derivata* misura la variazione della variabile dipendente y derivante da una variazione infinitesimale della variabile indipendente x (detta anche *argomento* della funzione).

Per definire più precisamente il concetto di derivata è necessario partire dal concetto di *rapporto incrementale*. Una variazione discreta della variabile indipendente da un valore iniziale x_0 a un valore finale x comporta una variazione della variabile dipendente da $y_0 = f(x_0)$ a $y = f(x)$. Il rapporto incrementale è pari al rapporto tra la variazione della variabile dipendente $\Delta y \equiv y - y_0 = f(x) - f(x_0)$ e quella della variabile indipendente $\Delta x \equiv x - x_0$:

$$\frac{\Delta y}{\Delta x} \equiv \frac{f(x) - f(x_0)}{x - x_0},$$

e misura il tasso di variazione *medio* della variabile dipendente per unità di variazione della variabile dipendente.

La derivata di y rispetto a x è il *limite* del rapporto incrementale a seguito di una variazione infinitesimale di x , ovvero:

$$f'(x) \equiv \frac{dy}{dx} \equiv \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \equiv \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0},$$

e misura il tasso di variazione istantaneo della variabile dipendente per effetto di una variazione della variabile indipendente. Da un punto di vista geometrico, invece, la derivata rappresenta la *pendenza* della funzione primitiva.

1.2 Funzioni continue e funzioni derivabili

La funzione $y = f(x)$ si dice *continua* in un punto x_0 se valgono le seguenti due condizioni:

1. x_0 è nel dominio della funzione, ovvero è uno dei possibili valori che può assumere la x ;
2. il limite di y per $x \rightarrow x_0$ esiste ed è esattamente uguale a $f(x_0)$.

La funzione è continua nel suo dominio X se essa è continua in ogni suo elemento x_0 . Più semplicemente, una funzione è continua se è possibile disegnarla senza mai sollevare la matita dal foglio.

La funzione $y = f(x)$ si dice *derivabile* un punto x_0 se valgono le seguenti due condizioni:

1. è *continua*;
2. è *liscia* (ovvero non ha punti d'angolo).

La funzione $y = f(x)$ si chiama *funzione primitiva*, e la derivata $dy/dx = f'(x)$ è anch'essa funzione della variabile indipendente x . Si noti, dalla precedente definizione, che la derivabilità implica la continuità, ma non viceversa. Cioè, la continuità (in un punto) è condizione necessaria ma non sufficiente per la derivabilità.

In logica, una *condizione necessaria* è un presupposto (ipotesi) con il quale una affermazione o proprietà (tesi) potrebbe sussistere, ma senza il quale l'affermazione non può valere ("solo se"). Una *condizione sufficiente* invece è un presupposto (ipotesi) con il quale una affermazione (tesi) sussiste, ma senza il quale l'affermazione (tesi) potrebbe comunque sussistere ("se"). Ne consegue che una *condizione necessaria e sufficiente* è un'equivalenza logica: il presupposto (ipotesi) e l'affermazione/proprietà (tesi) coincidono dal punto di vista logico ("se e solo se")¹.

1.3 Regole generali di derivazione

1.3.1 Regola di derivazione di una costante

Data una funzione del tipo $y = c$ dove c è una costante, cioè un qualunque numero reale, se deriviamo y rispetto a x otteniamo:

$$\frac{dy}{dx} = 0$$

1.3.2 Regola di derivazione della funzione potenza

Se la funzione primitiva è del tipo $y = ax^n$, dove n è un qualsiasi numero reale, vale la seguente regola:

¹ Formalmente, A è condizione necessaria per B se B implica A; A è condizione sufficiente per B se A implica B; A è condizione necessaria e sufficiente per B se A implica B e B implica A. Esempio: A: "T è un triangolo equilatero"; B: "T ha due angoli uguali". Chiaramente A è condizione sufficiente per B (se T è equilatero, allora esso ha due angoli uguali), ma non vale il viceversa (che T abbia due angoli uguali non implica che esso sia equilatero perché potrebbe essere soltanto isoscele).

$$\frac{d}{dx}ax^n = anx^{n-1}.$$

Un caso speciale di questa regola è quello della funzione lineare $y = ax$, che corrisponde a $n = 1$:

$$\frac{d}{dx}ax = ax^{1-1} = a,$$

per cui la derivata di una funzione lineare è semplicemente il coefficiente angolare della x , cioè a .

Altri casi speciali sono quello della funzione radice quadrata $y = a\sqrt{x}$, che corrisponde a $n = 1/2$:

$$\frac{d}{dx}a\sqrt{x} = \frac{d}{dx}ax^{\frac{1}{2}} = a \cdot \frac{1}{2} \cdot x^{\frac{1}{2}-1} = \frac{a \cdot x^{-\frac{1}{2}}}{2} = \frac{a}{2\sqrt{x}},$$

e quello dell'iperbole equilatera $y = \frac{a}{x}$, che corrisponde a $n = -1$:

$$\frac{d}{dx}\frac{a}{x} = \frac{d}{dx}ax^{-1} = a(-1)x^{-2} = -\frac{a}{x^2}.$$

1.3.3 Linearità della derivazione

Date due funzioni derivabili $f(x)$ e $g(x)$ e dati due numeri reali a e b , vale la seguente proprietà:

$$\frac{d}{dx}[af(x) + bg(x)] = a\frac{df(x)}{dx} + b\frac{dg(x)}{dx} \equiv af'(x) + bg'(x).$$

Questa regola incorpora la regola di derivazione di una costante moltiplicativa, ponendo $b = 0$:

$$\frac{d}{dx}af(x) = af'(x).$$

Inoltre casi speciali si hanno per la somma di due funzioni, che si ha per $a = b = 1$:

$$\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x),$$

e per la differenza tra due funzioni, che si ha per $a = 1$ e $b = -1$:

$$\frac{d}{dx}[f(x) - g(x)] = f'(x) - g'(x)$$

1.3.4 Regola di derivazione di un prodotto di funzioni

Date due funzioni derivabili $f(x)$ e $g(x)$, vale la seguente proprietà:

$$\frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$$

1.3.5 Regola di derivazione di un quoziente di funzioni

Data la funzione $y = \frac{f(x)}{g(x)}$, vale la seguente regola di derivazione:

$$\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2},$$

da cui si può ottenere il seguente caso particolare, noto come regola di derivazione del reciproco di una funzione:

$$\frac{d}{dx} \frac{1}{g(x)} = -\frac{g'(x)}{[g(x)]^2}.$$

Un caso particolare di quest'ultimo è a sua volta quello dell'iperbole equilatera, in cui $g(x) = x$.

1.3.6 Regola della catena di derivate

Se l'argomento di una funzione $f(x)$ è a sua volta un'altra funzione $g(x)$, vale la seguente regola:

$$\frac{d}{dx} f(g(x)) = f'(g(x))g'(x) \equiv \frac{df(g(x))}{dg(x)} \frac{dg(x)}{dx}.$$

1.3.7 Regola di derivazione di una funzione inversa

Se la funzione $y = f(x)$ è tale per cui a diversi valori di x corrispondono sempre valori diversi di y allora la funzione ammette un'inversa, ovvero esiste $x = f^{-1}(y) = g(y)$ tale per cui:

$$g(f(x)) = x \quad \Leftrightarrow \quad f(g(y)) = y,$$

cioè, non solo dato un valore di x ne risulterà determinato un unico valore di y , ma anche dato un valore di y ne risulterà determinato un unico valore di x .

La regola di derivazione di una funzione inversa è la seguente:

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}$$

In particolare, se f e g sono rette, questa regola implica che i loro coefficienti angolari siano uno l'inverso dell'altro.

L'invertibilità è propria delle funzioni *monotòne*. Una funzione $y = f(x)$ si dice:

- *monotòna crescente* se, dati qualunque due numeri reali del suo dominio x_1 e x_2 tali che $x_1 > x_2$, allora $f(x_1) > f(x_2)$;
- *monotòna non decrescente* se, dati x_1 e x_2 tali che $x_1 > x_2$, allora $f(x_1) \geq f(x_2)$;
- *monotòna decrescente* se, dati x_1 e x_2 tali che $x_1 > x_2$, allora $f(x_1) < f(x_2)$.
- *monotòna non crescente* se, dati x_1 e x_2 tali che $x_1 > x_2$, allora $f(x_1) \leq f(x_2)$;

Il modo più immediato per vedere se una funzione è monotona crescente o decrescente è calcolare la sua derivata e controllare che mantenga sempre lo stesso segno e che non sia mai uguale a zero, il che a sua volta implica che la funzione ha sempre pendenza positiva o negativa.

1.3.8 Regola di derivazione di una funzione logaritmica

Data la funzione $y = \log(x)$, dove per \log si intende logaritmo naturale (ovvero in base e), la sua derivata si ottiene applicando la seguente regola:

$$\frac{dy}{dx} \equiv \frac{d \log(x)}{dx} = \frac{1}{x}$$

cioè, la derivata della funzione logaritmica è uguale al reciproco dell'argomento.

1.3.9 Regola di derivazione di una funzione esponenziale

Data la funzione $y = e^x$, la sua derivata rispetto a x è data da:

$$\frac{dy}{dx} \equiv \frac{d}{dx} e^x = e^x.$$

Si ricordi che la funzione logaritmica e la funzione esponenziale sono una l'inversa dell'altra.

1.4 Derivata parziale

Per una generica funzione derivabile del tipo $y = f(x_1, \dots, x_n)$ dove x_1, \dots, x_n sono tutte indipendenti tra loro, la *derivata parziale* di y rispetto a un suo generico argomento x_i è:

$$\frac{\partial y}{\partial x_i} \equiv \lim_{\Delta x_i \rightarrow 0} \frac{\Delta y}{\Delta x_i} \equiv \lim_{\Delta x_i \rightarrow 0} \frac{f(x_1, \dots, x_i + \Delta x_i, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{\Delta x_i}$$

ovvero, nella derivazione parziale di y rispetto a x_i , tutte le altre $n-1$ variabili indipendenti sono trattate come se fossero costanti.

1.5 Differenziale parziale e totale

Si consideri una funzione derivabile del tipo $y = f(x_1, \dots, x_n)$, i cui argomenti, però, non sono tutti indipendenti tra loro. Per sapere come y varia al variare di x_i non sarà più sufficiente calcolare la derivata parziale $\partial y / \partial x_i$, perché essa coglierà solo l'effetto diretto di una variazione di x_i sulla variabile dipendente y . Infatti, poichè gli argomenti della funzione non sono tutti indipendenti tra loro, al variare di x_i anche altre variabili indipendenti potrebbero variare, provocando un ulteriore effetto su y di tipo "indiretto", che la derivata parziale di cui sopra non riesce a cogliere. Ecco perché, in tal caso, è opportuno ricorrere alla derivazione totale, concetto basato su quello di differenziale totale e che ci condurrà al calcolo della derivata totale.

Se, data una funzione del tipo $y = f(x)$, consideriamo solo incrementi infinitesimali di x e y (cioè per $\Delta x \rightarrow 0$ e $\Delta y \rightarrow 0$), parleremo allora di *differenziali* (rispettivamente dx e dy), e avremo che

$$dy \equiv f'(x)dx.$$

Tornando alla funzione $y = f(x_1, \dots, x_n)$, il *differenziale totale* di questa funzione sarà:

$$dy = \frac{\partial y}{\partial x_1} dx_1 + \dots + \frac{\partial y}{\partial x_n} dx_n.$$

Intuitivamente, esso indica la variazione di y dovuta a incrementi infinitesimali di tutte le variabili indipendenti x_1, \dots, x_n . Se invece si lascia variare, ad esempio, solo x_1 e si mantengono costanti tutte le altre variabili indipendenti, otteniamo il *differenziale parziale*:

$$dy = \frac{\partial y}{\partial x_1} dx_1 \quad \Leftrightarrow \quad \frac{dy}{dx_1} = \frac{\partial y}{\partial x_1}.$$

Si noti che i due rapporti *non* esprimono la stessa cosa: infatti, dy e dx_1 sono due variabili distinte perché rappresentano due diverse variazioni (esse sono, infatti, i differenziali di y e x_1) mentre $\partial y / \partial x_1$ è una derivata parziale e rappresenta quindi un'unica entità.

Il calcolo del differenziale totale di una funzione è semplice: basta determinare i coefficienti delle dx_i , ovvero bisogna calcolare le n derivate parziali di y rispetto alle x_i .

1.6 Derivata totale

Torniamo dunque al problema iniziale: cosa succede quando x_1, \dots, x_n non sono indipendenti?

Consideriamo un modello molto semplice, con due sole variabili indipendenti:

$$y = f(x, w),$$

dove $x = g(w)$. Possiamo allora riscrivere:

$$y = f(g(w), w).$$

Ora si noti che una variazione di w avrà due diversi effetti: l'*effetto diretto*, misurato dalla derivata parziale $\partial y / \partial w$, e l'*effetto indiretto*, dove la variazione di w farà variare x , che però a sua volta farà variare y . Per vedere ciò, si consideri il differenziale totale di y :

$$dy = \frac{\partial y}{\partial x} dx + \frac{\partial y}{\partial w} dw,$$

da cui, dividendo ambo i membri per dw otteniamo la *derivata totale* di y :

$$\frac{dy}{dw} = \frac{\partial y}{\partial x} \frac{dx}{dw} + \frac{\partial y}{\partial w}$$

Si noti che il membro di destra è costituito dalla somma dell'effetto indiretto e dell'effetto diretto di w su y .

1.7 Funzione implicita

La funzione $y = f(x)$ definisce sempre una funzione implicita del tipo $F(x, y) = 0$: basta riscrivere la funzione come $y - f(x) = 0$. Tuttavia, non si può essere altrettanto certi che $F(y, x_1, \dots, x_n) = 0$ definisca la funzione $y = f(x_1, \dots, x_n)$. Come esserne sicuri?

1.7.1 Teorema della funzione implicita

Data una funzione del tipo $F(y, x_1, \dots, x_n) = 0$ a valori reali e differenziabile, se:

- F ha derivate parziali continue nel punto $(y_0, x_{10}, \dots, x_{n0})$ tale che $F(y_0, x_{10}, \dots, x_{n0}) = 0$
- $F_y \neq 0$ nel punto $(y_0, x_{10}, \dots, x_{n0})$,

allora esiste un intorno di dimensione n di (x_0, \dots, x_n) in cui y è una funzione definita implicitamente dalle variabili (x_1, \dots, x_n) del tipo $y = f(x_1, \dots, x_n)$; inoltre, tale funzione è derivabile (reale) e ha derivate parziali continue.

1.7.2 Derivata di una funzione implicita

Data una funzione implicita del tipo $F(y, x_1, \dots, x_n) = 0$, se essa definisce una funzione del tipo $y = f(x_1, \dots, x_n)$ (vedi Teorema della funzione implicita), allora le sue derivate parziali sono:

$$\frac{\partial y}{\partial x_i} = - \frac{\frac{\partial F(y, x_1, \dots, x_n)}{\partial x_i}}{\frac{\partial F(y, x_1, \dots, x_n)}{\partial y}} \equiv - \frac{F_{x_i}}{F_y}$$

per $i = 1, \dots, n$ (a patto ovviamente che $F_y \neq 0$).

1.8 Derivata seconda, funzioni concave e funzioni convesse

La derivata di una funzione ci permette di stabilire se la funzione sia crescente o decrescente in un dato punto e di determinare la pendenza della tangente alla funzione in quel punto. Tuttavia è importante anche determinare se e come cambi tale pendenza, cioè caratterizzare la curvatura della funzione in ciascun punto.

Consideriamo per esempio una funzione derivabile $y = f(x)$ che sia monotona crescente, cosicché $f'(x) > 0$ per ogni valore di x nel dominio della funzione. Se la sua pendenza è a sua volta crescente, la sua derivata $f'(x)$ aumenta al crescere di x , per cui la derivata della sua derivata – detta derivata seconda – sarà anch'essa positiva: $df'(x)/dx \equiv f''(x) > 0$. In tal caso, si tratta di una funzione crescente e convessa, poiché il suo grafico è una curva che presenta una convessità verso il basso. Se invece la funzione ha pendenza via via decrescente, la sua derivata $f'(x)$ diminuisce al crescere di x , per cui la sua derivata seconda sarà negativa: $f''(x) < 0$, e la funzione sarà crescente e concava, poiché il suo grafico è una curva che presenta una concavità verso il basso.

Consideriamo ora una funzione derivabile $y = f(x)$ che sia monotona decrescente, cosicché $f'(x) < 0$ per ogni x nel suo dominio. Se la pendenza della funzione è decrescente in valore assoluto (cioè la curva corrispondente diventa via via meno ripida al crescere di x), il valore algebrico della sua derivata $f'(x)$ aumenta al crescere di x (essendo un numero negativo con valore

assoluto via via minore), per cui la sua derivata seconda sarà positiva: $f''(x) > 0$. In tal caso, si tratta di una funzione decrescente e *convessa*. Se la sua pendenza è invece crescente in valore assoluto (cioè la curva corrispondente diventa via via più ripida al crescere di x), il valore algebrico della sua derivata $f'(x)$ diminuisce al crescere di x (essendo un numero negativo con valore assoluto via via maggiore), per cui la sua derivata seconda sarà negativa: $f''(x) < 0$. In tal caso, si tratta di una funzione decrescente e concava.

Riassumendo, possiamo caratterizzare una funzione derivabile utilizzando le due derivate prime e seconde, come indicato dalla seguente tabella:

Derivata prima	Derivata seconda	Forma della funzione $f(x)$
$f'(x) > 0$	$f''(x) > 0$	crescente e convessa
$f'(x) > 0$	$f''(x) < 0$	crescente e concava
$f'(x) < 0$	$f''(x) > 0$	decrescente e convessa
$f'(x) < 0$	$f''(x) < 0$	decrescente e concava

Dalla tabella si vede chiaramente che, indipendentemente dalla pendenza (e quindi dalla derivata prima), la convessità o concavità della funzione dipende dal segno della sua derivata seconda. Le funzioni convesse hanno $f''(x) > 0$, mentre quelle concave hanno $f''(x) < 0$. Per la precisione, queste condizioni caratterizzano funzioni *strettamente* concave o convesse, poiché le disuguaglianze non includono il caso in cui la derivata seconda è pari a zero. Con $f''(x) \geq 0$, la funzione sarebbe detta *debolmente* convessa, e con $f''(x) \leq 0$, *debolmente* concava.

Il segno della derivata seconda fin qui è stato riferito all'intero dominio della funzione, cioè considerato valido per ogni valore di x : in tal caso, si dice che la funzione $y = f(x)$ è *globalmente* convessa o concava. Tuttavia esso può essere riferito anche a intervalli del dominio della funzione: per esempio una funzione monotona crescente può essere prima *localmente* concava per $x \in (x_0, x_1)$ e poi *localmente* convessa per $x \in (x_1, x_2)$, cioè il segno della derivata seconda può essere negativo nel primo intervallo e positivo nel secondo. In tal caso, per continuità la derivata seconda sarà pari a zero nel punto x_1 , cioè $f''(x_1) = 0$. Tale punto è detto "punto di flesso" della funzione.

Infine, va detto che è possibile caratterizzare la concavità o convessità di una funzione anche se essa non è derivabile: per esempio, se si tratta di una funzione lineare spezzata, che presenta almeno un punto d'angolo in cui non se ne può calcolare la derivata. In generale, una funzione è *strettamente concava* se il segmento rettilineo che si ottiene congiungendo due punti qualsiasi del grafico che rappresenta la funzione giace interamente al di sotto del grafico stesso, cioè se:

$$\lambda f(x_0) + (1 - \lambda)f(x_1) < f(\lambda x_0 + (1 - \lambda)x_1).$$

Similmente, una funzione è *strettamente convessa* se il segmento rettilineo che si ottiene congiungendo due punti qualsiasi del grafico che rappresenta la funzione giace interamente al di sopra del grafico stesso, cioè se:

$$\lambda f(x_0) + (1 - \lambda)f(x_1) > f(\lambda x_0 + (1 - \lambda)x_1).$$

Anche in questo caso, se le condizioni avessero i segni di disuguaglianza debole, si parlerebbe rispettivamente di funzione debolmente concava e di funzione debolmente convessa.

1.9 Ottimizzazione statica

Il problema di ottimizzazione si riferisce alla scelta del miglior elemento in un insieme di possibili alternative. Matematicamente, si tratta di massimizzare o minimizzare la cosiddetta funzione obiettivo attraverso la ricerca dei suoi valori estremi (massimi o minimi), che indichiamo come valori "ottimi". Possiamo trovarci di fronte a due tipi di problemi di ottimizzazione: libera o vincolata. Entrambe si basano sul controllo di due condizioni: la condizione di primo ordine e la condizione di secondo ordine.

1.8.1 Condizione di primo ordine

La condizione di primo ordine consiste semplicemente nel calcolare la derivata prima della nostra funzione obiettivo², porla uguale a zero e trovare il valore della variabile indipendente che risolve l'equazione che viene fuori. Per esempio se vogliamo massimizzare/minimizzare $y = f(x)$, procediamo nel seguente modo: 1) calcoliamo $f'(x)$; 2) poniamo $f'(x) = 0$; 3) troviamo il valore di x che risolve questa equazione. Questo sarà il valore 'critico' della nostra variabile indipendente, che chiameremo x^* . Ad esso corrisponderà il valore $f(x^*)$ della variabile dipendente y , valore che

² Se definita su un insieme aperto (cioè che non include i punti di frontiera), oppure derivata solo sui punti interni del dominio. In effetti, la condizione del primo ordine deve valere (condizione necessaria) solo per questi ultimi.

è detto un *valore stazionario* di y . La coppia di valori $(x^*, f(x^*))$ rappresenta un *punto stazionario*.

Perché “stazionario”? Perché x^* soddisfa $f'(x^*) = 0$, quindi giace su un tratto della curva $y = f(x)$ che non è inclinato né positivamente né negativamente. (Si ricordi infatti che la derivata rappresenta la pendenza della funzione primitiva.). Si noti che: (i) La condizione del primo ordine è una *condizione necessaria* (per un ottimo interno): se $f(x^*)$ è un ottimo (massimo o minimo) interno, allora deve valere $f'(x^*) = 0$. Il viceversa non è sempre vero: possono esistere valori della variabile indipendente che annullano la derivata prima della funzione (cioè sono punti stazionari), eppure non sono ottimi per la funzione. (ii) Il procedimento appena descritto viene eseguito allo stesso modo sia per i problemi di massimizzazione che per quelli di minimizzazione. Ma allora a cosa corrisponde il valore x^* che ne viene fuori, un minimo o un massimo della funzione $f(x)$? Per rispondere a questa seconda domanda dobbiamo controllare la condizione di secondo ordine.

1.8.2 Condizione di secondo ordine

Sia ora x^* un valore critico per la nostra funzione, ossia sia $f'(x^*) = 0$. La condizione di secondo ordine si basa sul calcolo della *derivata seconda*, che consiste semplicemente nella derivata della derivata. Ovvero:

$$\frac{df'(x)}{dx} = f''(x)$$

Si noti che mentre la derivata prima, cioè $f'(x)$, descrive come varia $f(x)$ al variare di x (si ricordi il significato concettuale della derivata), la derivata seconda invece, cioè $f''(x)$, descrive la variazione della *pendenza* di $f(x)$ al variare di x , cioè misura la curvatura del grafico di $y = f(x)$.

Quindi:

- se $f'(x) > 0$ il *valore* della funzione aumenta all'aumentare di x (curva con pendenza positiva);
- se $f'(x) < 0$ il *valore* della funzione diminuisce all'aumentare di x (curva con pendenza negativa);
- se $f''(x) > 0$ la *pendenza* della funzione aumenta all'aumentare di x (curva crescente con incrementi crescenti, che implica convessità);

- se $f''(x) < 0$ la *pendenza* della funzione diminuisce all'aumentare di x (curva crescente a incrementi decrescenti, che implica concavità).

Allora lo studio della condizione di secondo ordine produce uno dei seguenti casi:

- se $f''(x^*) > 0$ la funzione $f(x)$ ha un *minimo locale* in x^* ;
- se $f''(x^*) < 0$ la funzione $f(x)$ ha un *massimo locale* in x^* ;
- se $f''(x^*) = 0$ non possiamo pronunciarsi su cosa sia x^* , a meno che non si controllino derivate di ordine superiore al secondo (ma ciò non sarà oggetto di questo corso).

Vale la pena di evidenziare due casi speciali:

- se $f''(x) < 0$ per tutti i valori della variabile indipendente x , allora $f(x)$ si dice strettamente concava ed ha un *massimo globale* in x^* ;
- se $f''(x) > 0$ per tutti i valori della variabile indipendente x , allora $f(x)$ si dice strettamente convessa ed ha un *minimo globale* in x^* .

La differenza tra un minimo/massimo *locale* ed un minimo/massimo *globale* sta nel fatto che nel secondo caso il segno della derivata seconda si mantiene costante per qualunque valore della x , mentre nel primo caso tale segno varia al variare di x .

1.9 Espansione in serie di Taylor

A volte è utile trasformare una funzione in un'altra forma più conveniente. Qui analizzeremo una trasformazione particolare: l'*espansione* di una funzione del tipo $y = f(x)$ intorno ad un certo punto x_0 , il che vuol dire trasformare la funzione in un polinomio i cui coefficienti sono espressi in termini delle derivate di x valutate nel punto x_0 . Il risultato dell'espansione viene anche chiamato *serie di potenze* perché, essendo un polinomio, consiste in una somma di funzioni potenza.

Si consideri un polinomio di ordine n :

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n .$$

È possibile dimostrare che il polinomio può essere espresso nel modo seguente, attraverso un'*espansione in serie di Taylor* intorno al punto x_0 della variabile indipendente x :

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3 + \dots + \frac{f^n(x_0)}{n!}(x-x_0)^n.$$

Intuitivamente, il termine costante $f(x_0)$ rappresenta l'“altezza” della funzione in corrispondenza del valore x_0 della variabile indipendente x . Il secondo termine, che è lineare nella deviazione di x dal valore di riferimento x_0 , appare moltiplicato per la derivata $f'(x_0)$, che rappresenta la “pendenza” della funzione in corrispondenza del valore x_0 della variabile indipendente. Il terzo termine, che è quadratico nella deviazione di x dal valore di riferimento x_0 , appare moltiplicato per la derivata seconda $f''(x_0)$, che rappresenta la “curvatura” della funzione in corrispondenza del valore x_0 della variabile indipendente. E così via.

Tuttavia, in base al teorema di Taylor, la stessa trasformazione può essere effettuata su qualunque funzione arbitraria $\phi(x)$, non necessariamente polinomiale, purché $\phi(x)$ abbia derivate continue fino all'ordine desiderato nel punto di espansione x_0 . L'unica differenza rispetto al caso della funzione polinomiale è che la trasformazione non è esatta, bensì funziona soltanto come un'approssimazione e quindi genererà un resto:

$$\phi(x) = \underbrace{\left[\phi(x_0) + \frac{\phi'(x_0)}{1!}(x-x_0) + \frac{\phi''(x_0)}{2!}(x-x_0)^2 + \frac{\phi'''(x_0)}{3!}(x-x_0)^3 + \dots + \frac{\phi^n(x_0)}{n!}(x-x_0)^n \right]}_{=P_n} + R_n$$

ovvero:

$$\phi(x) = P_n + R_n$$

Intuitivamente, questa viene definita come *serie di Taylor con resto*. Si noti che il resto rappresenta la differenza tra la funzione originaria $\phi(x)$ e il polinomio $P(x)$: $R_n = \phi(x) - P_n$. Il resto dipende da n : quanto maggiore è il numero di termini e quindi di derivate che si includono nel polinomio, tanto migliore sarà l'approssimazione e di conseguenza tanto più piccolo sarà il resto, purché la serie sia convergente, ovvero per $n \rightarrow \infty$ si verifica che $R_n \rightarrow 0$ e $P_n \rightarrow \phi(x)$.

1.10 Il caso di più variabili di scelta

Finora abbiamo preso in esame il problema dell'ottimizzazione di una funzione obiettivo dipendente da una sola variabile, $y = f(x)$. In quel caso, bastava scegliere il valore della sola funzione x per individuare il massimo o il minimo di y . Nel caso di variabili che dipendono da due o

più variabili decisionali, è necessario adottare un metodo diverso per individuare i valori estremi della funzione, perché in questo caso occorre ovviamente scegliere i valori di tutte le variabili indipendenti in corrispondenza dei quali la variabile dipendente raggiunge il suo valore massimo o minimo.

Si consideri una funzione $z = f(x, y)$ continua e differenziabile: la variabile z quindi varia in funzione delle due variabili indipendenti x e y . Per individuare i punti di massimo o minimo relativo di z , bisogna ancora guardare ai segni della derivata prima e della derivata seconda, ma il nuovo contesto (il caso di più variabili indipendenti) richiede l'introduzione delle derivate parziali di primo e secondo ordine. La funzione $z = f(x, y)$ avrà due derivate parziali di primo ordine, una per ciascuna variabile indipendente:

$$f_x \equiv \frac{\partial z}{\partial x} \text{ e } f_y \equiv \frac{\partial z}{\partial y}$$

e due derivate parziali di secondo ordine:

$$f_{xx} \equiv \frac{\partial}{\partial x} f_x \equiv \frac{\partial^2 z}{\partial x^2} \text{ e } f_{yy} \equiv \frac{\partial}{\partial y} f_y \equiv \frac{\partial^2 z}{\partial y^2},$$

dove il doppio indice sta a indicare che la funzione primitiva è stata derivata parzialmente rispetto all'una o all'altra variabile due volte. Avremo, però, anche le cosiddette derivate miste, dal momento che le derivate parziali di primo ordine, così come la funzione primitiva, sono anch'esse funzioni sia di x che di y ³:

$$f_{xy} \equiv \frac{\partial}{\partial x} f_y = f_{yx} \equiv \frac{\partial}{\partial y} f_x \equiv \frac{\partial^2 z}{\partial x \partial y}.$$

In questo caso, la condizione di secondo ordine è più complessa rispetto al caso di una singola variabile decisionale. Per garantire la concavità della funzione, e quindi l'individuazione di un massimo, non basta che le derivate seconde sia entrambe negative. Similmente, per garantire la convessità della funzione, e quindi l'individuazione di un minimo, non basta che le derivate seconde sia entrambe positive. In entrambi i casi, occorre anche che la derivata mista “non sia troppo grande”. Più precisamente, la tabella seguente riassume le condizioni di primo e secondo ordine per la ricerca dei punti di massimo o minimo relativo, quando la funzione è del tipo $z = f(x, y)$ ⁴:

³ Le derivate miste sono uguali per ogni funzione derivabile (almeno) due volte e a derivate continue (almeno) fino al secondo ordine.

⁴ Per semplicità, consideriamo solo i segni “stretti” delle derivate parziali seconde, nella condizione del secondo ordine. Esistono tuttavia semplici generalizzazioni del risultato secondo cui un punto stazionario può generare un massimo o un minimo della funzione anche quando alcune di tali derivate sono uguali a zero. Ciò tuttavia non sarà oggetto del presente corso.

	Massimo	Minimo
Condizione di 1° ordine: le derivate parziali di primo ordine entrambe uguali a zero	$f_x = f_y = 0$	$f_x = f_y = 0$
Condizione di 2° ordine: si considerano <ul style="list-style-type: none"> • le derivate parziali di secondo ordine e • la derivata mista (elevata al quadrato) 	$f_{xx} < 0, f_{yy} < 0$	$f_{xx} > 0, f_{yy} > 0$
	$f_{xx} \cdot f_{yy} > f_{xy}^2$	

1.11 Ottimizzazione vincolata

Se la funzione che si vuole massimizzare o minimizzare deve sottostare a un vincolo, l'ottimizzazione (intesa come ricerca dei valori ottimi) non sarà più "libera" ma "vincolata". Il vincolo stabilisce una relazione tra le variabili della funzione obiettivo, che risultano dipendenti le une dalle altre in quanto variabili decisionali. In breve, un vincolo indica la presenza di determinati fattori limitativi nel problema di ottimizzazione (ad esempio, il vincolo di bilancio tiene conto del potere d'acquisto del consumatore e della sua capacità di spesa nell'ambito della massimizzazione della sua utilità). Il risultato finale sarà quello di un ottimo condizionato.

Si supponga che il problema sia quello di scegliere i valori di due variabili x e y che rendono massimo il valore della funzione $z = f(x, y)$ sotto al vincolo $g(x, y) = c$, dove c è una costante, ovvero

$$\max_{x,y} z = f(x, y) \text{ sub } g(x, y) = c .$$

Esistono due modi equivalenti di risolvere tale problema:

- 1) il **metodo di sostituzione**,
- 2) il **metodo di Lagrange**.

Il primo metodo consiste nella sostituzione del vincolo all'interno della funzione obiettivo, allo scopo di eliminare variabili decisionali e riportare il problema di ottimizzazione vincolata a quello di ottimizzazione libera (quando, in sostanza, c'è una sola variabile decisionali rispetto alla quale massimizzare o minimizzare la funzione obiettivo). La **sostituzione** consisterà nell'esplicitare il vincolo rispetto a una delle due variabili decisionali (laddove valgano le condizioni del teorema

della funzione implicita), riscrivendo per esempio $g(x, y) = c$ come $x = h(y; c)$, e sostituirlo nella funzione obiettivo, ottenendo così una nuova funzione obiettivo la cui sola variabile di scelta è y :

$$\max_y z = f(h(y; c), y) \text{ sub } g(x, y) = c .$$

La soluzione può a questo punto essere ottenuta con le stesse tecniche di massimizzazione (o minimizzazione) di una funzione non vincolata, già spiegate più sopra. Specificamente, la condizione di primo ordine sarà:

$$\frac{\partial f(h(y; c), y)}{\partial x} \frac{dh(y; c)}{dy} + \frac{\partial f(h(y; c), y)}{\partial y} = 0 . \quad (1)$$

Poiché, in base al teorema della funzione implicita,

$$\frac{dh(y; c)}{dy} = - \frac{\frac{\partial g(x, y)}{\partial y}}{\frac{\partial g(x, y)}{\partial x}} ,$$

questa condizione può essere riscritta come:

$$\frac{\frac{\partial f(x, y)}{\partial x}}{\frac{\partial f(x, y)}{\partial y}} = \frac{\frac{\partial g(x, y)}{\partial x}}{\frac{\partial g(x, y)}{\partial y}} . \quad (1')$$

Tale metodo, però, può rivelarsi inefficace in presenza di vincoli che non possono essere trasformati in una funzione esplicita, in cui una delle variabili decisionali appare come variabile dipendente e le altre come variabili indipendenti. In alternativa si ricorre, quindi, al **metodo di Lagrange**, con il quale si trasforma il problema di ottimizzazione vincolata in una forma tale da consentire l'applicazione delle condizioni di ottimo tipiche dell'ottimizzazione libera. In pratica, data una funzione obiettivo $z = f(x, y)$ e un vincolo $g(x, y) = c$, dove c è una costante, si trasforma la funzione obiettivo in una nuova funzione L , detta funzione Lagrangiana, che conterrà z (la funzione obiettivo originaria) più il vincolo g moltiplicato per un nuovo parametro, λ , detto moltiplicatore di Lagrange.

La funzione Lagrangiana sarà:

$$L(x, y, \lambda) = f(x, y) - \lambda[g(x, y) - c]$$

oppure

$$L(x, y, \lambda) = f(x, y) + \lambda[c - g(x, y)],$$

una nuova funzione L in tre variabili (x , y e λ), alla quale applicare la condizione di primo ordine per un estremo libero. Basterà porre, infatti, le tre derivate parziali di L rispetto a queste tre variabili pari a zero e risolvere il sistema di tre equazioni in tre variabili per calcolare il punto di ottimo, ottenendo i tre valori (x^*, y^*, λ^*) che risolvono il seguente sistema di tre equazioni:

$$\begin{cases} \frac{\partial L(x, y, \lambda)}{\partial x} = 0, \\ \frac{\partial L(x, y, \lambda)}{\partial y} = 0, \\ \frac{\partial L(x, y, \lambda)}{\partial \lambda} = 0, \end{cases} \quad (2)$$

ovvero:

$$\begin{cases} \frac{\partial f(x, y)}{\partial x} = \lambda \frac{\partial g(x, y)}{\partial x}, \\ \frac{\partial f(x, y)}{\partial y} = \lambda \frac{\partial g(x, y)}{\partial y}, \\ g(x, y) = c. \end{cases}$$

Mentre la terza condizione non fa altro che riprodurre il vincolo, le prime due condizioni di primo ordine possono essere riscritte eliminando il moltiplicatore di Lagrange λ , nel modo seguente:

$$\frac{\frac{\partial f(x, y)}{\partial x}}{\frac{\partial f(x, y)}{\partial y}} = \frac{\frac{\partial g(x, y)}{\partial x}}{\frac{\partial g(x, y)}{\partial y}}, \quad (2')$$

che è identica alla condizione (1') ottenuta più sopra. Ciò dimostra che, quando anche il metodo della sostituzione del vincolo può essere utilizzato (poiché per il vincolo valgono le condizioni del teorema della funzione implicita), la condizione di primo ordine ottenuta con i due metodi è identica.

2 Appendice statistica

2.1 Variabile casuale discreta

Una *variabile casuale* (o aleatoria) è una variabile i cui valori non possono essere previsti con esattezza. Essa può essere discreta o continua. Una variabile casuale *discreta* può assumere un numero *finito* di valori x_1, \dots, x_n , in corrispondenza degli eventi e_1, \dots, e_n , che si verificano con le rispettive probabilità p_1, \dots, p_n . Un esempio è il punteggio totale del tiro di due dadi. Invece un esempio di una variabile casuale *continua* è la temperatura in una stanza, che può assumere infiniti valori all'interno di un certo intervallo. L'intervallo dei valori che può assumere la variabile casuale è detto *supporto* della sua distribuzione di probabilità, che può essere finito (come nel caso di una variabile casuale uniforme) o infinito (nel caso di una variabile casuale normale).

Una variabile casuale può risultare dalla somma di altre variabili casuali. Continuando con l'esempio dei due dadi, supponiamo che uno di essi sia verde e che l'altro sia rosso. Si noti che tali dadi possono dar vita a 36 possibili *combinazioni*, in quanto ciascun dado può assumere valori da 1 a 6, ciascuna con probabilità di verificarsi pari ad $1/36$ (assumendo che i dadi non siano truccati!). Tuttavia, il numero delle combinazioni non coincide necessariamente con il numero di *valori* di una variabile casuale. Ad esempio, se definiamo la variabile casuale "X = somma dei due dadi", questa può assumere valori soltanto da 2 a 12 per un totale di 11 valori (o *esiti*) diversi, nonostante le combinazioni possibili dei due dadi siano ben 36: l'esito $X = 2$ ha probabilità di verificarsi pari ad $1/36$ in quanto consiste nell'*intersezione* tra gli eventi rosso = 1 e verde = 1, mentre l'esito $X = 7$ può accadere nelle seguenti combinazioni: (rosso = 1, verde = 6), (rosso = 2, verde = 5), (rosso = 3, verde = 4), (rosso = 4, verde = 3), (rosso = 5, verde = 2), (rosso = 6, verde = 1); di conseguenza, $X = 7$ ha probabilità di verificarsi pari all'*unione* di questi 6 eventi mutuamente esclusivi, e cioè pari a $6/36$. Si noti infine che la somma delle probabilità di tutti gli esiti è esattamente pari ad 1, ovvero un esito tra 2 e 12 deve necessariamente verificarsi.

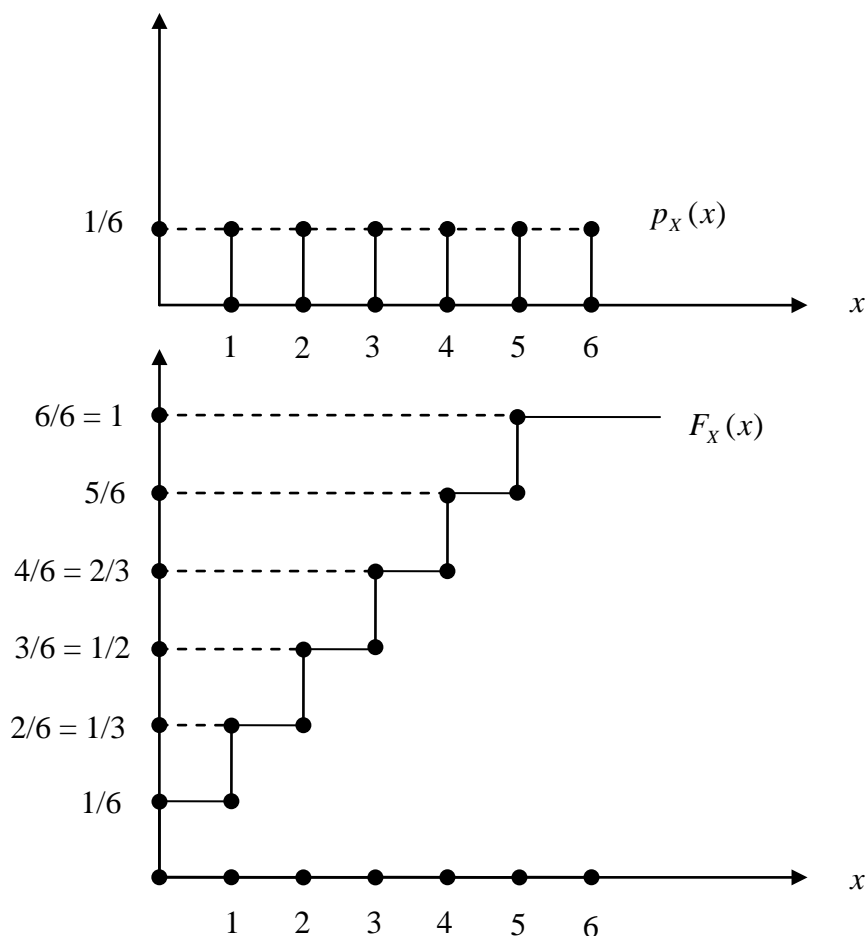
La funzione di probabilità $p_X(x)$, o funzione di massa di probabilità, o di *densità* di una variabile casuale discreta X è una funzione che assegna ad ogni valore possibile di X la probabilità dell'evento elementare ($X = x$). Per esempio, la probabilità che il risultato del lancio di un dado sia 3 è $p_X(3) = 1/6$. Alternativamente, possiamo scrivere $\Pr(x = 3) = 1/6$. In realtà, anche la probabilità che il risultato del lancio del dado sia pari a 1, 2, 4, 5 o 6 è pari a $1/6$. In altri termini, nell'esempio del dado la funzione di densità in questo caso è una funzione uniforme discreta.

La funzione di ripartizione $F_X(x)$, o funzione di *distribuzione* di probabilità di una variabile casuale discreta X è invece una funzione che indica la probabilità che la variabile X assuma un valore x inferiore o pari a un certo valore z (cioè che $x \leq z$):

$$F_X(z) = \sum_{x \leq z} p_X(x) = \Pr(x \leq z).$$

In altri termini, nel caso di variabile casuale discreta, la funzione di *distribuzione* di probabilità è pari alla sommatoria della funzione di densità per valori inferiori o pari al valore z . (Corrispondentemente, nel caso di una variabile casuale continua, essa è pari all'integrale della funzione di densità.) Per esempio, nel caso del lancio di un dado, la probabilità che il risultato sia pari o inferiore a 3 è pari a $\Pr(x=1) + \Pr(x=2) + \Pr(x=3) = 1/6 + 1/6 + 1/6 = 1/2$. La figura seguente mostra la funzione di densità e di distribuzione relativa al risultato del lancio del dado:

Figura 1. Esempio di variabile discreta con funzione uniforme di densità



Per definizione, se calcolata in corrispondenza del limite superiore del supporto della distribuzione, questa funzione è pari a 1: per esempio, nel caso del lancio di un dado, il dado mostrerà certamente una faccia con valore pari o inferiore a 6.

2.2 Valore atteso di una variabile casuale discreta

In generale, il *valore atteso* (o *aspettativa*) di una variabile casuale discreta è pari alla media ponderata di tutti i suoi possibili valori, usando la probabilità di ciascun esito come peso. In altre parole, si moltiplica ciascun possibile valore della variabile casuale per la sua probabilità e si somma tutti i valori così ottenuti. In termini matematici, il valore atteso della variabile casuale X viene indicato con $E(X)$ o μ_X e, nel caso di una variabile casuale discreta, viene definito come:

$$E(X) \equiv \mu_X = x_1 p_1 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i$$

Supponiamo che X possa assumere solo due valori, x_1 e x_2 , rispettivamente con probabilità p e $1-p$, indicato come segue:

$$X = \begin{cases} x_1 & \text{con probabilità } p \\ x_2 & \text{con probabilità } 1-p \end{cases}$$

Il valore atteso di X sarà pari a

$$E(X) \equiv \mu_X = px_1 + (1-p)x_2.$$

La variabile casuale X può ovviamente assumere anche più di due valori. Per esempio, se definiamo X come il punteggio risultante dal tiro di un dado, il suo valore atteso sarà:

$$E(X) = \sum_{i=1}^6 x_i p_i = \frac{1}{6}(1+2+3+4+5+6) = 3,5$$

Questa formula in realtà si può generalizzare a qualunque funzione della variabile casuale X , ovvero a $g(X)$, il cui valore atteso è:

$$E[g(X)] = g(x_1)p_1 + \dots + g(x_n)p_n = \sum_{i=1}^n g(x_i)p_i$$

Si noti che, per indicare una variabile casuale si usa talvolta il simbolo “ \sim ” sopra la lettera che la indica (per esempio, essa può essere indicata con il simbolo \tilde{X} , usando invece la lettera X per indicare una sua particolare realizzazione, cioè uno dei valori che essa può assumere).

Regole del valore atteso

Ci sono quattro regole principali da ricordare:

- Regola 1. Il valore atteso di una somma di variabili casuali è uguale alla somma dei loro valori attesi (cioè, il valore atteso è un operatore lineare):

$$E(X + Y + Z) = E(X) + E(Y) + E(Z)$$

- Regola 2. Se si moltiplica una variabile casuale per una costante, il suo valore atteso si moltiplica per la stessa costante:

$$E(bX) = bE(X)$$

- Regola 3. Il valore atteso di una costante è pari alla costante stessa:

$$E(b) = b$$

- Regola 4. La derivata rispetto a Z del valore atteso di (aXZ) , in cui X è una variabile casuale e Z è una variabile deterministica (cioè non casuale), sarà:

$$\frac{d}{dZ} E[aXZ] = aE(X).$$

Unendo queste 3 regole, si può calcolare il valore atteso di espressioni più complesse come ad esempio $Y = b_1 + b_2 X$:

$$E(Y) = E(b_1 + b_2 X) = E(b_1) + E(b_2 X) = b_1 + b_2 E(X).$$

2.3 Varianza di una variabile casuale discreta

Indicata con $\text{var}(X)$ o σ_X^2 , è definita come il valore atteso del quadrato della differenza tra X e la sua media:

$$\text{var}(X) \equiv \sigma_X^2 = E\left[(X - \mu_X)^2\right]$$

Per una variabile casuale discreta:

$$\text{var}(X) \equiv p_1(x_1 - \mu_X)^2 + \dots + p_n(x_n - \mu_X)^2 = \sum_{i=1}^n p_i(x_i - \mu_X)^2$$

La radice quadrata della varianza invece, che indichiamo con σ_X , rappresenta lo *scarto quadratico medio* o *deviazione standard* ed è definita nella stessa unità di misura della media (contrariamente alla varianza, che invece è definita nel quadrato dell'unità di misura della media).

Regole della varianza

Ci sono quattro regole principali da ricordare:

- Regola 1. La varianza di una somma di variabile casuale è data da:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$$

dove $\text{cov}(X, Y)$ rappresenta la covarianza tra X e Y (vedremo tra breve di che si tratta).

- Regola 2. La varianza del prodotto tra una variabile casuale ed una costante è pari a:

$$\text{var}(bX) = b^2 \text{var}(X)$$

- Regola 3. La varianza di una costante (o di una variabile deterministica) è sempre nulla:

$$\text{var}(b) = 0$$

- Regola 4. La varianza della somma di una variabile casuale ed una costante è pari a:

$$\text{var}(X + b) = \text{var}(X) + 0 = \text{var}(X)$$

Si noti infine che esiste un'utile formula alternativa per calcolare la varianza:

$$\text{var}(X) = E(X^2) - [E(X)]^2 \equiv E(X^2) - \mu_X^2$$

Tale formula può essere ottenuta partendo dalla formula originale presentata precedentemente nel modo riportato qui di seguito:

$$\begin{aligned} \text{var}(X) &\equiv \sigma_X^2 = E[(X - \mu_X)^2] = \\ &= E(X^2 + \mu_X^2 + 2X\mu_X) = \\ &= E(X^2) + \mu_X^2 - 2\mu_X E(X) \end{aligned}$$

Poiché $E(X) = \mu_X$, si ottiene:

$$\sigma_X^2 = E(X^2) + \mu_X^2 - 2\mu_X^2 = E(X^2) - \mu_X^2.$$

2.4 Covarianza tra due variabili casuali

Indicata con $\text{cov}(X, Y)$ o σ_{XY} , viene definita come il valore atteso del prodotto delle deviazioni di X e Y dalle rispettive medie:

$$\text{cov}(X, Y) \equiv \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)],$$

che nel caso di due variabili casuali discrete è

$$\text{cov}(X, Y) = \sum_{i=1}^n p_i (x_i - \mu_X)(y_i - \mu_Y)$$

Si noti che la covarianza rappresenta una misura di *associazione* tra due variabili. La sua funzione principale è di esprimere il segno di tale associazione, ma risente della scala di tali variabili, cioè della loro unità di misura. Per esempio, supponiamo che la variabile X sia il peso della quantità di grano prodotta su un certo terreno, misurata in tonnellate, e la Y sia la quantità di pioggia caduta in un anno su quel terreno, e immaginiamo di cambiare l'unità di misura della X da tonnellate a quintali, per cui ad ogni valore iniziale della X (in tonnellate) ora corrisponde un valore $1000X$ (in quintali). Anche la covarianza ne risulterà moltiplicata per 1000:

$$\text{cov}(1000X, Y) = E[(1000X - 1000\mu_X)(Y - \mu_Y)] = E[1000(X - \mu_X)(Y - \mu_Y)] = 1000 \text{cov}(X, Y).$$

Proprio perché la covarianza tra due variabili dipende dalla loro unità di misura, può esser utile normalizzarla trasformandola nel coefficiente di correlazione lineare, che può assumere solo valori compresi tra -1 e 1 (come vedremo tra breve).

Si noti inoltre che se le due variabili sono indipendenti, la loro covarianza è pari a zero. Questo risultato deriva dalla definizione stessa di indipendenza: due variabili casuali X e Y si dicono indipendenti se e solo se $E[f(X)g(Y)] = E(f(X))E(g(Y))$ per qualsiasi funzione f e g . Allora, se consideriamo $f(X) = X - \mu_X$ e $g(Y) = Y - \mu_Y$, otteniamo:

$$\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(X - \mu_X)E(Y - \mu_Y) = (\mu_X - \mu_X)(\mu_Y - \mu_Y) = 0.$$

Tuttavia non vale il contrario, ovvero una covarianza nulla non implica necessariamente indipendenza. Ciò accade se, pur essendo la variabile X non correlata con la Y , una funzione della variabile X è correlata con una funzione della Y .

Regole della covarianza

Ci sono tre regole principali da ricordare:

- Regola 1. La covarianza tra una variabile casuale e una somma di variabili casuali è data da:

$$\text{cov}(X, Y + Z) = \text{cov}(X, Y) + \text{cov}(X, Z)$$

- Regola 2. La covarianza tra una variabile casuale e il prodotto di una variabile casuale per una costante è data da:

$$\text{cov}(X, bY) = b \text{cov}(X, Y)$$

- Regola 3. La covarianza tra una variabile casuale ed una costante è pari a zero:

$$\text{cov}(X, b) = 0$$

Si noti infine che esiste un'utile formula alternativa anche per il calcolo della covarianza:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

2.5 Coefficiente di correlazione

Come si è detto più sopra, la covarianza dipende dalle unità di misura delle variabili a cui si riferisce, ovvero una variazione dell'unità di misura utilizzata cambia anche il valore della covarianza. Una misura migliore è pertanto rappresentata dal *coefficiente di correlazione* perché è indipendente dalle unità di misura utilizzate. Tale coefficiente si indica con ρ ed è definito da:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}} \equiv \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

ovvero si calcola come il rapporto tra la covarianza ed il suo massimo teorico, dato appunto dal prodotto dei due scarti quadratici medi. Di conseguenza, i valori che può assumere ρ sono compresi tra -1 e 1 . Si noti che nel caso in cui le due variabili casuali siano indipendenti, ρ è pari a zero esattamente come succede per la covarianza (che rappresenta infatti il suo numeratore).

2.6 Aspettativa condizionata

La distribuzione di una variabile casuale \tilde{Y} condizionatamente al fatto che un'altra variabile casuale \tilde{X} assuma uno specifico valore è detta "distribuzione condizionata di Y data X ". La probabilità condizionata che Y assuma il valore y quando $X = x$ si indica con $\Pr(Y = y | X = x)$.

In generale, l'aspettativa condizionata di Y data X è uguale al valore atteso della variabile Y , calcolato utilizzando la sua distribuzione condizionata rispetto a X . Se Y può assumere uno di k valori (y_1, \dots, y_k) in corrispondenza di $X = x$, allora la media condizionata di Y data $X = x$ sarà

$$E(Y | X = x) = \sum_{i=1}^k y_i \Pr(Y = y_i | X = x).$$

Un esempio potrà chiarire la differenza tra distribuzione condizionata e non condizionata. Supponiamo di dover utilizzare uno dei computer della biblioteca di dipartimento per scrivere una tesina. Il bibliotecario ci assegna casualmente un computer tra quelli disponibili, informandoci che potrebbero verificarsi dei blocchi dovuti al fatto che non tutti i computers della biblioteca sono nuovi. L'informazione sull'età del computer rappresenta un'informazione aggiuntiva che ci consente di condizionare la probabilità che il computer si blocchi mentre stiamo lavorando al fatto che esso sia nuovo oppure vecchio. Supponiamo che la metà dei computer sia nuova e l'altra vecchia, e che ciò sia noto. Questo è il nostro "insieme informativo", rispetto al quale possiamo condizionare la probabilità che il computer a noi assegnato si blocchi.

Formalmente:

- indichiamo la variabile "età del computer" con la variabile dicotomica A , che può assumere due valori: 1 se il computer è nuovo e 0 se il computer è vecchio;
- indichiamo con N la variabile "numero di blocchi" e supponiamo che possa assumere valori da 0 a 4, avremo una distribuzione condizionata di N data A , rappresentata dalla seguente tabella:

	$N = 0$	$N = 1$	$N = 2$	$N = 3$	$N = 4$	Totale
$\Pr(N A = 0)$	0,70	0,13	0,10	0,05	0,02	1
$\Pr(N A = 1)$	0,90	0,07	0,02	0,01	0	1

Dalla tabella si può notare che la probabilità di avere zero blocchi ($N = 0$) quando il computer è nuovo è maggiore di quella condizionata a un computer vecchio, così come la probabilità di $N = 4$ è pari a 0 nel caso di un computer nuovo e positiva nel caso di uno vecchio.

La media condizionata della variabile "blocchi" dato $A = 0$ sarà dunque:

$$E(N | A = 0) = (0 \cdot 0,70) + (1 \cdot 0,13) + (2 \cdot 0,10) + (3 \cdot 0,05) + (4 \cdot 0,02) = 0,56$$

mentre per $A = 1$ sarà:

$$E(N | A = 1) = (0 \cdot 0,90) + (1 \cdot 0,07) + (2 \cdot 0,02) + (3 \cdot 0,01) + (4 \cdot 0) = 0,14$$

Risulta evidente che il numero atteso di blocchi dato un computer vecchio è maggiore di quello atteso dato un computer nuovo.

Ma se volessimo conoscere il valore atteso di N non condizionato, ovvero il numero di blocchi che posso aspettarmi indipendentemente dall'età del computer, cioè $E(N)$? Secondo la **legge delle aspettative iterate**:

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

la media di Y è pari alla media ponderata delle medie condizionate. Tornando al nostro esempio, l'aspettativa di M è la media delle distribuzioni condizionate di M data A , ponderate per le probabilità che il computer sia vecchio ($A = 0$) o nuovo ($A = 1$), che nell'esempio in questione sono entrambe pari a $1/2$:

$$\begin{aligned} E(M) &= E[E(M | A)] \\ &= E(M | A = 0) \Pr(A = 0) + E(M | A = 1) \Pr(A = 1) \\ &= \left(0,56 \cdot \frac{1}{2}\right) + \left(0,14 \cdot \frac{1}{2}\right) = 0,35. \end{aligned}$$